1. OVERVIEW OF THE 2007 *MARYLAND SCHOOL ASSESSMENT-MATHEMATICS*

In 2002, the Maryland State Department of Education (MSDE), in order to conform to the requirements of the new Federal program "No Child Left Behind," retired its award-winning *Maryland School Performance Assessment Program* and adopted a testing program known as the *Maryland School Assessment* (*MSA*). The new program, like its predecessor, was based on the *Voluntary State Curriculum*, which set reasonable academic standards for what teachers were expected to teach and for what students were expected to learn in schools.

In 2003, the MSA-Math was introduced in grades 3, 5, and 8, and grades 4, 6, and 7 were added to the program in 2004. In addition, another vendor has administered the MSA-Math until 2006, and the MSA-Math was award to Harcourt Assessment, Inc. starting spring 2007. Because of different equating procedures between two vendors, a transformation of scale scores was conducted in 2006 using equipercentile method. Detailed information on scale score transformation can be found in Appendix C, Year 2006 MSA-Math Recalibration Results from 3PL IRT to the Rasch Model Using Equipercentile Method.

In 2007, SAT10 Core Form A or B was administered at grades 3 through 8. Several of the SAT10 items aligned to Maryland math curriculum were included into criterion-referenced scores. In addition, all of the SAT10 items were used to report norm-referenced scores. Some of the Maryland-specific items that appeared both in 2007 and in previous years were used to conduct year-to-year calibration and equating. It should be noted that the Rasch difficulty estimates of the Maryland-specific items which were generated by recalibration in 2006 were kept as fixed parameters in 2007 calibration and equating procedures. So all scale scores in 2006 were on the same scale in 2007 within each content and grade.

A Bookmark standard setting was conducted in 2003 to set proficiency level cut scores for grades 3, 5, and 8. Because 2004 was the first testing year for grades 4, 6, and 7, a second Bookmark standard setting was held in summer 2004 to set cut scores for these additional grades. The performance level cut scores were used to assign students to three proficiency levels (Basic, Proficient, and Advanced) for AYP reporting under the "No Child Left Behind" act. Information about the Bookmark procedures and results can be found from MSDE. It should be noted that these cut scores have been applied since 2003 (grades 3, 5, and 8) and 2004 (grades 4, 6, and 7).

From March 12 to March 21, 2007, students in grades 3 through 8 took the 2007 *MSA* in mathematics.

1.1 General Overview of the 2007 MSA-Math

The 2007 MSA-Math was designed to provide two types of information. First, *norm-referenced* information was provided by the items from the abbreviated form of the *Stanford Achievement Test Series, Tenth Edition (SAT10)*. For example, the *SAT10* consisted of *Problem Solving and Procedures*. Second, to produce *criterion-referenced* information, additional items, called augmented items, were written for the *Maryland Mathematics Standards* in grades 3 through 8 and were organized under the seven math content standards: *Algebra, Geometry, Measurement, Statistics, Probability, Numbers and Computation,* and *Process*. However, it should be noted that some standards were combined for reporting purposes, and the reporting strands can be found in Tables 1.6 through 1.23.

The 2007 MSA-Math produced both norm-referenced and criterion-referenced scores for each student. While norm-referenced scores included only the *SAT10* items, both items selected from the *SAT10* and augmented items created for Maryland comprised criterion-referenced scores. Figure 1.1 shows a schematic of the *SAT10* and augmented items that produced these test scores.

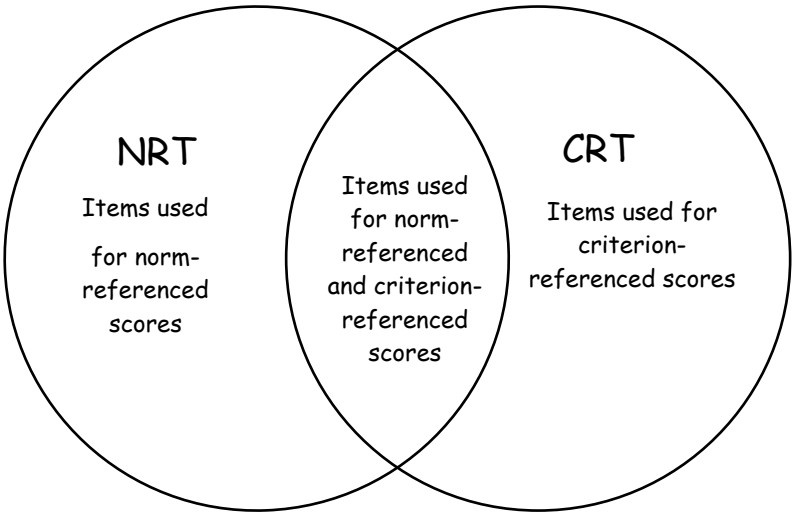


Figure 1.1 Schematic of the 2007 MSA-Math

1.2 Purposes/Uses of the 2007 MSA-Math

By measuring students' achievement against the new academic standards, the 2007 MSA-Math provides two main purposes. First, the MSA-Math was designed to inform parents, teachers, and educators of what students actually learned in schools by providing specific feedback that can be used to improve the quality of schools, classrooms, and individualized instructional programs and to model effective assessment approaches that can be used in classrooms. Second, the MSA-Math serves as an accountability tool to measure performance levels of individual students, schools, and districts against the new academic standards.

1.3 *The Voluntary State Curriculum*

Federal law requires that states align their tests with their state content standards. MSDE worked carefully and rigorously to construct new tests to provide a strong alignment as defined by the U.S. Department of Education.

The *Voluntary State Curriculum* (*VSC*), which defined what students should know and be able to do at each grade level, helped schools understand the standards more clearly, and included more specificity with indicators and objectives. The format of the *VSC* specified standards statements, indicators, and objectives. Standards are broad, measurable statements of what students should know and be able to do. Indicators and objectives provide more specific content knowledge and skills that are unique at each grade level.

While 100% of the standards should be tested, it was not the case that every indicator would necessarily be tested each year. Consequently, the *VSC* specified curricular indicators and objectives that contributed directly to measuring content standards, which were aligned to the *Maryland School Assessment (MSA).*

1.4 Development and Review of the 2007 MSA-Math

Developing the 2007 MSA-Math was a complex process. It required a great deal of involvement from MSDE, Harcourt, and local school systems. In addition, teachers, administrators, and content specialists from all over Maryland were recruited for different test development committees. These individuals reviewed test forms and items to ensure that they measured students' knowledge and skills fairly and without bias. Table 1.1 identifies which groups were responsible for developing the 2007 MSA-Math.

Table 1.1 The 2007 MSA-Math Responsibility for Test Development

Development of the 2007 MSA-Math	Primary Responsibility
Development of Preliminary Blueprints and Item Specifications	Harcourt; MSDE; NPC
Development of Operational Form Requirement and Session Blueprint	MSDE
Item Writing	Harcourt
Item Review	Harcourt; MSDE; NPC; Content Review Committee
Bias Review	Harcourt; MSDE; Bias Review Committee
Vision Review	MSDE; Harcourt; Vision Review Committee
Modification of Special Forms	Harcourt; MSDE
Review of Special Forms	MSDE
Construction of Operational Test Forms	Harcourt; MSDE; NPC
Construction of Field Test Forms	Harcourt; MSDE
Review of Operational Test Forms	MSDE
Final Construction of Test Forms	Harcourt; MSDE

National Psychometric Council

The National Psychometric Council (NPC) took a major role in reviewing and recommending to MSDE on the development and implementation of the 2007 MSA-Math program. For example, they made recommendations to MSDE on issues, such as test blueprints, field test design, item analysis, item selection for scoring purposes, linking, equating and scaling issues, and other relevant statistical and psychometric issues.

Content Review Committee

Content Review Committee members ensured that the MSA-Math was appropriately difficult and fair. Committee members were either specialists in math for test items, or experts in test construction and measurement. They represented all levels of education as well as the ethnic and social diversity of Maryland students. Committee members were from different areas of the state.

The educators' understanding of Maryland curriculum and extensive classroom experience made them a valuable source of information. They reviewed test items and forms and took a holistic view to ensure that tests were fair and balanced across reporting categories.

Bias Review Committee

In addition to the Content Review Committee, a separate Bias Review Committee examined each item on math tests. They looked for indications of bias that would impact the performance of an identifiable group of students. Committee members discussed and, if necessary, rejected items based on gender, ethnic, religious, or geographical bias.

Vision Review Committee

A separate Vision Review Committee examined each item on math tests. They looked for indications of bias that would impact the performance of an identifiable group of students. Committee members discussed and suggested edits, based on ethnic, religious, disability, or geographical bias.

1.5 Test Structure of the 2007 MSA-Math

2007 MSA-Math Test Structure

The 2007 MSA-Math was composed of the *SAT10* items and augmented (Maryland-specific) operational items. In addition, the uniqueness of the MSA-Math was to spiral a relatively large number of Maryland field test items into multiple test forms (10 forms) for each grade in test administration.

As can be seen from Table 1.2, the 2007 MSA-Math produced 10 test forms for each grade, and there were 2 operational test forms within each grade. This means that Forms A, B, C, D and E (Form A) are identical, and Forms F, G, H, J, and K (Form F) are identical with respect to operational item sets.

Tables 1.3, 1.4, and 1.5 provide information concerning the test design of NRT and CRT and the number of operational and field test items included for each test form. Tables 1.6 through 1.23 provide information concerning the number of items that contribute to each strand (e.g., *Algebra, Geometry, Measurement, Statistics, Probability, Numbers and Computation,* and *Process*).

Table1.2 The 2007 MSA-Math Test Structure: Grades 3 through 8

	Operational Item Sets		Field Test Item Sets									
	A	F	A	B	C	D	E	F	G	H	J	K
Form A	X		X									
Form B	X			X								
Form C	X				X							
Form D	X					X						
Form E	X						X					
Form F		X						X				
Form G		X							X			
Form H		X								X		
Form J		X									X	
Form K		X										X

Note. Forms A, B, C, D, and E (Form A) are identical, and Forms F, G, H, J, and K (Form F) are identical in terms of operational test items.

2007 MSA-Math Item Types

The 2007 MSA-Math included four types of items: *selected response (SR), student-produced response (SPR), brief constructed response (BCR), and extended constructed response (ECR)*.

SR items require students to select a correct answer from several alternatives. For the 2007 MSA-Math, students selected an answer from four or five alternatives. Each *SR* item was scored dichotomously (e.g., 0 or 1).

SPR items require students to record their answers on a grid by shading in circles corresponding to the numbers in their answer. For the 2007 MSA-Math, only grade 7 and 8 tests included *SPR* items. Each *SPR* item was scored dichotomously.

BCR items require students to provide a short answer using words, numbers, and/or symbols while *ECR* items require students to write an answer that consists of more information than is required for a brief constructed response item.

Both *BCR* and *ECR* items consist of Step A and Step B. Step A contributes to the content score while Step B contributes to the process score. Each step was considered as an independent item and separately scored;

All *BCR* and *ECR* Step A items received 0-1 score point range from two independent scorers; all *BCR* Step B items received a 0-2 score point range and all *ECR* Step B items received a 0-3 score point range from two independent scorers. Score was the higher of the first and the second readers' scores provided they were adjacent. A resolution reader's score was used of two non-adjacent initial scores were received. That is, the resolution reader's score was used in place of both the first and second readers' scores. It should be noted that grade 3 and 4 tests did not include *ECR* items.

Table 1.3 The 2007 MSA-Math Test Design: Grades 3 and 4

Grade	Strand Title	*SAT10* / Augmented	Item Type	No. of Items of Each Form	
				FA	FF
3	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR*, *BCR*	61	61
	Algebra	Augmented	*SR*, *BCR*	13	13
	Geometry	Augmented	*SR*, *BCR*	7	7
	Measurement	Augmented	*SR*, *BCR*	7	7
	Statistics	Augmented	*SR*, *BCR*	12	12
	Probability	Augmented	*SR*	2	2
	Number Computation	Augmented	*SR*, *BCR*	13	13
	Process	Augmented	*BCR*	7	7
4	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR*, *BCR*	64	64
	Algebra	Augmented	*SR*, *BCR*	14	14
	Geometry	Augmented	*SR*, *BCR*	7	7
	Measurement	Augmented	*SR*, *BCR*	7	7
	Statistics	Augmented	*SR*, *BCR*	8	8
	Probability	Augmented	*SR*, *BCR*	7	7
	Number Computation	Augmented	*SR*, *BCR*	14	14
	Process	Augmented	*BCR*	7	7

Note. CRT contains *SAT10* items. *SR* items are selected response items, and *BCR* items are brief constructed response items. Form A designates the forms A, B, C, D, and E. Form F designates the forms F, G, H, J, and K.

Table 1.4 The 2007 MSA-Math Test Design: Grades 5 and 6

Grade	Strand Title	*SAT10* / Augmented	Item Type	No. of Items of Each Form	
				FA	FF
5	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR*, *BCR, ECR*	61	59
	Algebra	Augmented	*SR*, *BCR, ECR*	14	14
	Geometry	Augmented	*SR*, *BCR*	5	5
	Measurement	Augmented	*SR*, *BCR*	7	7
	Statistics	Augmented	*SR*, *BCR*	9	8
	Probability	Augmented	*SR*, *BCR*	4	4
	Number Computation	Augmented	*SR*, *BCR*	14	14
	Process	Augmented	*BCR, ECR*	8	7
6	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR*, *BCR, ECR*	59	59
	Algebra	Augmented	*SR*, *BCR, ECR*	13	13
	Geometry	Augmented	*SR*, *BCR*	8	8
	Measurement	Augmented	*SR*, *BCR*	6	6
	Statistics	Augmented	*SR*, *BCR*	9	9
	Probability	Augmented	*SR*	4	4
	Number Computation	Augmented	*SR*, *BCR*	12	12
	Process	Augmented	*BCR, ECR*	7	7

Note. CRT contains *SAT10* items. *SR* items are selected response items, *BCR* items are brief constructed response items, and *ECR* items are Extended Constructed Response. Form A designates the forms A, B, C, D, and E. Form F designates the forms F, G, H, J, and K.

Table 1.5 The 2007 MSA-Math Test Design: Grades 7 and 8

Grade	Strand Title	*SAT10* / Augmented	Item Type	No. of Items of Each Form	
				FA	FF
7	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR, SPR, BCR, ECR*	61	61
	Algebra	Augmented	*SR,SPR, BCR, ECR*	14	14
	Geometry	Augmented	*SR, SPR, ECR*	7	7
	Measurement	Augmented	*SR, SPR, BCR*	6	6
	Statistics	Augmented	*SR, SPR, BCR, ECR*	9	9
	Probability	Augmented	*SR, SPR, BCR*	5	5
	Number Computation	Augmented	*SR, SPR*	13	13
	Process	Augmented	*BCR, ECR*	7	7
8	Total NRT	*SAT10*	*SR*	40	40
	Problem Solving	*SAT10*	*SR*	20	20
	Procedures	*SAT10*	*SR*	20	20
	Total CRT	Augmented	*SR, SPR, BCR, ECR*	59	59
	Algebra	Augmented	*SR, SPR, BCR, ECR*	15	15
	Geometry	Augmented	*SR, SPR, ECR*	7	7
	Measurement	Augmented	*SR, SPR, BCR*	4	4
	Statistics	Augmented	*SR, SPR, BCR, ECR*	9	9
	Probability	Augmented	*SR, SPR, BCR*	5	5
	Number Computation	Augmented	*SR, SPR*	11	11
	Process	Augmented	*BCR, ECR*	8	8

Note. CRT contains *SAT10* items. *SR* items are selected response items, *SPR* items are student-produced response, *BCR* items are brief constructed response items, and *ECR* items are extended constructed response. Form A designates the forms A, B, C, D, and E. Form F designates the forms F, G, H, J, and K.

Table 1.6 Test Design and Item Distribution for the 2007 MSA-Math: Grades 3

Form	*SAT10* / Maryland			Augmented Maryland Item								**Total # of Item**
	2*	6*	Total	1*	2*	3*	4*	5*	6*	7*	Total	
A	1	2	3	13	7	7	12	2	13	7	61	**64**
F	1	2	3	13	7	7	12	2	13	7	61	**64**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process

Table 1.7 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 3

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	13	15	14	15	14	**71**
F	13	15	14	15	14	**71**

Table 1.8 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 3

Form	# of *SAT 10* SR Item	# of Aug. SR Item	# of Augmented BCR Item		**Total # of Item**	Scores of *SAT10*	Scores of Augmented SR	Scores of Aug. BCR		**Total Score**
			Step A	Step B				Step A	Step B	
A	3	47	7	7	**64**	3	47	7	14	**71**
F	3	47	7	7	**64**	3	47	7	14	**71**

Table 1.9 Test Design and Item Distribution for the 2007 MSA-Math: Grades 4

Form	*SAT10* / Maryland	Augmented Maryland Item								**Total**
		1*	2*	3*	4*	5*	6*	7*	Total	**# of Items**
A	None of them	14	7	7	8	7	14	7	64	**64**
F	None of them	14	7	7	8	7	14	7	64	**64**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process

Table 1.10 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 4

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	14	14	15	14	14	**71**
F	14	14	15	14	14	**71**

Table 1.11 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 4

Form	# of *SAT 10* SR Item	# of Aug. SR Item	# of Augmented BCR Item		**Total # of Item**	Scores of *SAT10*	Scores of Augmented SR	Scores of Aug. BCR		**Total Score**
			Step A	Step B				Step A	Step B	
A	0	50	7	7	**64**	0	50	7	14	**71**
F	0	50	7	7	**64**	0	50	7	14	**71**

Table 1.12 Test Design and Item Distribution for the 2007 MSA-Math: Grades 5

Form	*SAT10* / Maryland					Augmented Maryland Item								**Total**
	1*	2*	3*	6*	Total	1*	2*	3*	4*	5*	6*	7*	Total	**# of Items**
A	1	1	1	1	4	14	5	7	9	4	14	8	61	**65**
F	1	1	1	1	4	14	5	7	**8***	4	14	**7***	59	**63**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process
Note. One *BCR* item (both Step A and Step B) was dropped based on MSDE recommendation.

Table 1.13 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 5

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	15	14	13	15	17	**74**
F	15	14	**12***	15	**15***	**71**

Note. One *BCR* item was dropped based on MSDE recommendation.

Table 1.14 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 5

Form	# of *SAT 10* SR Item	# of CRT SR Item	# of Augmented of BCR Item		# of Augmented of ECR Item		**Total # of Item**	Scores of *SAT10*	Scores of Aug. SR	Scores of Aug. BCR		Scores of Aug. ECR		**Total Score**
			Step A	Step B	Step A	Step B				Step A	Step B	Step A	Step B	
A	4	45	7	7	1	1	**65**	4	45	7	14	1	3	**74**
F	4	45	**6***	**6***	1	1	**63***	4	45	**6***	**12***	1	3	**71***

Note. One *BCR* item was dropped based on MSDE recommendation.

Table 1.15 Test Design and Item Distribution for the 2007 MSA-Math: Grades 6

Form	*SAT10* / Maryland			Augmented Maryland Item								**Total # of Item**
	1*	6*	Total	1*	2*	3*	4*	5*	6*	7*	Total	
A	1	2	3	13	8	6	9	4	12	7	59	**62**
F	1	2	3	13	8	6	9	4	12	7	59	**62**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process

Table 1.16 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 6

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	14	14	13	14	15	**70**
F	14	14	13	14	15	**70**

Table 1.17 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 6

Form	# of *SAT 10* SR Item	# of CRT SR Item	# of Augmented of BCR Item		# of Augmented of ECR Item		**Total # of Item**	Scores of *SAT10*	Scores of Aug. SR	Scores of Aug. BCR		Scores of Aug. ECR		**Total Score**
			Step A	Step B	Step A	Step B				Step A	Step B	Step A	Step B	
A	3	45	6	6	1	1	**62**	3	45	6	12	1	3	**70**
F	3	45	6	6	1	1	**62**	3	45	6	12	1	3	**70**

Table 1.18 Test Design and Item Distribution for the 2007 MSA-Math: Grades 7

Form	*SAT10* / Maryland		Augmented Maryland Item								**Total # of Item**
	6	Total	1*	2*	3*	4*	5*	6*	7*	Total	
A	1	1	14	7	6	9	5	13	7	61	**62**
F	1	1	14	7	6	9	5	13	7	61	**62**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process

Table 1.19 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 7

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	14	13	14	14	17	**72**
F	14	13	14	14	17	**72**

Table 1.20 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 7

Form	# of *SAT 10*	# of Aug. SR Item	# of Aug. SPR Item	# of Aug. BCR Item		# of Aug. ECR Item		**Total # of Item**	Scores of *SAT 10*	Scores of Aug. SR	Scores of Aug. SPR	Scores of Aug. BCR		Scores of Aug. ECR		**Total Score**
				Step A	Step B	Step A	Step B					Step A	Step B	Step A	Step B	
A	1	35	12	4	4	3	3	**62**	1	35	12	4	8	3	9	**72**
F	1	35	12	4	4	3	3	**62**	1	35	12	4	8	3	9	**72**

Table 1.21 Test Design and Item Distribution for the 2007 MSA-Math: Grades 8

Form	*SAT10* / Maryland				Augmented Maryland Item								**Total # of Item**
	2*	3*	6*	Total	1*	2*	3*	4*	5*	6*	7*	Total	
A	1	1	1	3	15	7	4	9	5	11	8	59	**62**
F	1	1	1	3	15	7	4	9	5	11	8	59	**62**

Note. 1*. Algebra; 2*. Geometry; 3*. Measurement; 4*. Statistics; 5*. Probability; 6*. Numbers and Computation; 7*. Process

Table 1.22 Total and Reporting Cluster Scores for the 2007 MSA-Math: Grades 8

Form	Total and Reporting Cluster Scores					
	1	2&3	4&5	6	7	**Total Score**
A	15	13	14	12	19	**73**
F	15	13	14	12	19	**73**

Table 1.23 Item Type and Score Point Distribution for the 2007 MSA-Math: Grades 8

Form	# of *SAT 10*	# of Aug. SR Item	# of Aug. SPR Item	# of Aug. BCR Item		# of Aug. ECR Item		**Total # of Item**	Scores of *SAT 10*	Scores of Aug. SR	Scores of Aug. SPR	Scores of Aug. BCR		Scores of Aug. ECR		**Total Score**
				Step A	Step B	Step A	Step B					Step A	Step B	Step A	Step B	
A	3	31	12	5	5	3	3	**62**	3	31	12	5	10	3	9	**73**
F	3	31	12	5	5	3	3	**62**	3	31	12	5	10	3	9	**73**

1.6 Test Administration

Test Materials

All test materials had to be stored in a secure location prior to test administration. The School Test Coordinator (STC) provided test administration training and test materials to the test examiners. Pre-test workshops were held in Baltimore for all Local Accountability Coordinators in Maryland. These workshops provided the representatives of all the local school divisions with an overview of the test's content, security expectations, and procedures for completing the answer documents. They also considered the receipt, distribution, and return of test materials.

For the test examiner, Harcourt provided the following materials:

- Examiner's Manuals
- Preprinted and generic labels, which were applied to the Test/Answer Books by or under the direct supervision of the STC.
- Scoring Service Identification sheets
- Student Roster

For each student, the following materials were provided by Harcourt:

- Test/Answer Book
- Special accommodations testing materials, if necessary

For each student, the following additional materials were provided by school or student:

- Two No. 2 pencils with erasers
- Blank scratch paper for mathematical computations
- Classroom ruler(s) for both U.S. customary and metric measurements and a classroom calculator for all grade levels
- Classroom protractor(s) for grades 5 through 8
- Classroom compass(es) for grades 7 through 8

Each classroom used for the assessment also needed the following additional materials:

- A sign for the door, "Testing: Do not Disturb"
- A digital clock or a watch, or clock with a second hand
- Copies of the STOP and GO ON sample pages

Two test related examiners manuals (EM) were developed for the 2007 MSA; one version for reading and the other for mathematics for use in all grades 3-8. Developed in partnership with MSDE, the EMs contained instructions for preparation and administration of the test. In addition to the EMs, one Test Administration and Coordination Manual (TACM) was developed for use by the Local Accountability Coordinators (LAC) and building-level School Test Coordinators (STC). Included in this manual were instructions for preparation of materials for testing, monitoring of

testing, and packaging of materials for return to Harcourt for scoring. The TACM was distributed and reviewed during a workshop in January for STCs and LACs with duplicates sent to each school with its testing materials.

Test Administration Schedule

The overall test window for MSA was established by MSDE (March 12-21, 2007, with make-up testing held March 22-27, 2007). However, each Local Education Agency (LEA) set a specific schedule for administration of the MSA within that window for their district. Each LEA schedule was submitted to MSDE in advance and proved for each district by the State. For a given grade and content area, all testing had to take place on the same schedule. In addition, each content area at each grade was tested on two days during the window. For the 2007 MSA-Math, the primary testing days were as follows:

- Test materials delivered to schools (Examiner's Manuals, Test/Answer Books, and Test Coordinator's Kit) — On or Before February 26, 2007
- Mathematics Primary Testing Window — March 12 - March 21, 2007
- Make-up Testing Window — March 22 - March 27, 2007

Students and parents should be reminded of the importance of students attending school during the administration of the MSA and the importance of student participation in MSA testing. Maryland was held to the 95% participation requirement under NCLB by the US Department of Education, and schools should do all they can to test all students on MSA or Alt-MSA (as applicable).

If a student was absent on the testing days, a make-up test was administered on any two consecutive days within testing window. If a school had an unscheduled closing or delayed opening that prohibited the administration from occurring on the scheduled testing dates, the STCs were consulted with LACs to determine the testing schedule to be followed.

During the administration of the 2007 MSA-Math, MSDE had testing monitors in selected schools observing administration procedures and testing conditions. All monitors had identification cards for security purposes. There were no prior notification of which schools would be monitored, but monitors followed local procedures for reporting to the school's main office and giving proper notification that an MSDE monitor was in the building.

Student Participation

All students in grades 3 through 8 had to participate in the 2007 MSA-Math. The only exception was that students with severe cognitive disabilities were assessed by the *Alternate Maryland School Assessment* (ALT-MSA) instead of the regular MSA-Math. The criteria that students should need to be tested in the Alt-MSA program instead of the MSA-Math could be viewed in section 2, Appendix C of the TACM.

The U.S. Department of Education was developing specific guidance related to Modified Assessment, but that guidance, as yet, had not been issued. Students might have been identified through the Individualized Education Program (IEP) process in the current school year as takers of the Mod-MSA. However, since the Mod-MSA was not available, those students had to be assessed using the regular MSA-Math.

Testing Accommodations

Testing accommodations for students with disabilities (i.e., students having an Individualized Education Program or a Section 504 Plan) and students for English Language Learners (ELL) had to be approved and documented according to the procedures and requirements outlined in the document entitled "*Maryland Accommodations Manual: A Guide to Selecting, Administrating, and Evaluating the Use of Accommodations for Instruction and Assessment*," (MAM). A copy of the most recent edition of this document is available electronically on the LAC and STC web pages at https://docushare.msde.state.md.us/docushare.

No accommodations might be made for students merely because they were members of an instructional group. Any accommodation had to be based on individual needs and not on a category of disability area, level of instruction, environment, or other group characteristics. Responsibility for confirming the need and appropriateness of an accommodation rested with the LAC and school-based staff involved with each student's instructional program. A master list of all students and their accommodations had to be maintained by the principal and submitted to the LAC, who provided a copy to MSDE upon request. Please refer to Section 1 of the 2007 TACM for further information regarding testing accommodations.

Large-Print and Braille Test Books and Kurzweil™ Test Forms on CD

MSA-Math was administered to those requiring (1) large-print Student Test/Answer Books or (2) Braille Test Books, or (3) Kurzweil™ Test Forms on CD. For large-print Test/Answer Books, Braille Test Books, and Kurzweil™ Test Forms on CD, student responses were transcribed into the standard-size Test/Answer Book following testing.

The pre-printed student ID label was affixed to the standard-size Test/Answer Book containing the transcribed responses, not to the large-print Test/Answer Book or Braille books.

An eligible Test Examiner (TE) transcribed the student responses into a standard-size Test/Answer Book exactly as given by the student. Any original student Test/Answer Books which were used as source documents for transcription was invalidated by drawing a large slash across the student demographic page with a black permanent marker.

Once the student responses had been transcribed, the transcribed Test/Answer Book was returned for scoring with the standard-size materials. Specific packing instructions are provided in the TACM in section 3 and 4.

Security of Test Materials

The following code of ethnics conforms to the Standards for Educational and Psychological Testing developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (Harcourt, 2007):

> It is breach of professional ethics for school personnel to provide verbal or nonverbal clues or answers, teach items on the test, share writing prompts, coach, hint, or in any way influence a student's performance during the testing situation. A breach of ethics may result in invalidation of test results and local education agency or MSDE disciplinary action. (p. 9)

The Test/Answer Books for the 2007 MSA-Math were confidential and kept secure at all times. Unauthorized use, duplication, or reproduction of any or all portions of the assessment was prohibited, which is reflected by the following statement (Harcourt, 2007):

> Violation of security can result in prosecution and/or penalties as imposed by the Maryland State Board of Education and/or State Superintendent of Schools in accordance with the COMAR 13A.03.04 and 13A.12.05. (p. 9)

All materials were treated as confidential and placed in locked areas. Secure and non-secure test materials were as follows:

- Secure materials: Test/Answer Books (including large-print and Braille), Kurzweil™ test forms on CD, and used scratch paper
- Non-secure materials: TACM, Examiner's Manuals, unused pre-printed student and generic ID labels, unused FedEx return shipping labels, and unused green/orange shipping labels.

Test Format

In 2007, there were 10 forms of MSA-Math. Different test forms were administered to students in each classroom participating in math tests, and each test form was identified by color and form number/letter. All forms of the MSA Test/Answer Books for each grade had the same grade designation and picture on the front cover.

The Test/Answer Books were spiraled within a classroom, and each student used a combined Test/Answer Book. Since the Test/Answer Books were scanned for scoring, students were encouraged not to use highlights in any part of the book. Although students might be accustomed to using highlighters in daily instruction, highlighting in the Test/Answer Book could obliterate information in a student's book when it was scanned for scoring. As an alternative to highlighting, students were allowed to lightly circle or underline information in test items or perform calculations to help them in responding, as long as markings do not interfere with the bubbled answer choice area and/or the track marks along the outside margins of each page.

1.7 MSA-Math Scoring Procedures

Students' responses to *SR* and *SPR* items were machine-scored, and their responses to *BCR* and *ECR* items were individually read and scored by Harcourt.

Once received by Harcourt, Test/Answer Books were scanned into an electronic imaging system so that the information necessary to score responses was captured and converted into an electronic format. Students' identification and demographic information, school information, and answers to *SR* items were converted to alphanumeric format; hand-written responses were captured in digital image format.

Machine-Scored Items

After students' responses to *SR* and *SPR* items were converted to text format, the scoring key was applied to the captured item responses. Correct answers were assigned a score of one point. Incorrect answers, blank responses (omits), and responses with multiple marks were also assigned a score of zero.

Hand-Scored Items

Test/Answer Books were scanned into the electronic imaging system, allowing scorers to score these responses online at all scoring sites while maintaining the live documents at the contractor's facility. The imaging system randomly distributed responses, ensuring no one scorer scored a disproportionate number of responses from any one school. This online scoring system maintained a database of actual student responses and the scores associated with those responses. An off-site backup of all images and scores was maintained as well to guard against potential loss of data and images due to system failure. The system also provided continuous, up-to-date monitoring of all scoring activities. Detailed information on MSA scoring specification can be obtained in a document, Performance Assessment Scoring Center: Spring 2007 Scoring Specification for MSA-Reading and Math which is available from MSDE.

Scoring Staff

The MSDE had one Room Director (RD) dedicated for each grade level, domain (Math) and site. The RD worked closely with the PASC Training Supervisor and the PASC Math Specialists. The PASC Training Supervisor, Math Specialist, and RDs participated in the anchor-pulling sessions in Maryland. The Room Director/Training Team Leader was responsible for maintaining annotations and meeting minutes from all sessions. These notes were a record of the comments and decisions made by the MSDE personnel and members of the Maryland teacher committee. These notes were utilized by the RD responsible for training the Team Leaders (TLs) and Readers for the respective Maryland prompts. For MSDE scoring projects, PASC had qualified alternate RDs available at the beginning of the project to ensure a timely start of training in the event that the primary RD was unavailable to start as scheduled. The alternate RD acted as a TL unless the RD couldn't fulfill his/her duties.

1) Reader/Scorer

A graduate of a four-year accredited college or university who had successfully passed the PASC new reader exam and new reader training. The readers were eligible to score custom programs for which they have been trained and successfully qualify.

2) Team Leader (TL)

An experienced reader who directly monitored the scoring of a team of readers and retrains as needed. The reader had successfully completed the PASC TL training program.

3) Room Director (RD)

A knowledgeable team leader who had been selected to work with team leaders and the training supervisor to oversee the scoring of several teams. An RD's main duty was to rule on validity of questionable papers and to maintain consistency in scoring decisions. RDs also served as trainers.

4) Reader's Aide (RA)

PASC storeroom personnel whose main responsibilities during scoring were to do copying and printing for the PASC materials center. During anchor pulling, RA responsibility might include duplicating student papers. They might also be assigned a variety of clerical duties.

5) Developers

An experienced PASC reader that was responsible for selecting a wide variety of student responses for such activities as benchmarking, anchor pulling range finding, and training materials. Selected papers were then submitted to MSDE for comment and approval. Developers remained on the project as anchor pulling participants and trainers whenever possible.

6) Trainers

Experienced personnel who were TLs or RDs and selected by the Training Supervisor to train and qualify readers for Maryland. Additionally these experienced personnel might also train new readers and do domain specific training.

Reader Recruitment and Qualifications

All Readers for MSDE had to provide Harcourt's staffing vendor their resume and documentation of a four-year, college degree. As part of the initial screening process for recruiting Readers into Harcourt's general pool, applicants had to respond to an open-ended prompt. This writing sample ensured that all applicants were able to perform the kinds of tasks they would assess. The writing sample was intended to screen out those who couldn't write standard, idiomatically correct English or who couldn't organize their thoughts clearly. The writing prompt was scored by a qualified PASC staff member. If successful on the preliminary screening, applicants then participated in a one-day general introductory training workshop presented by a PASC staff member. These workshops allowed Harcourt to eliminate potential Readers who might seem qualified according to their educational and professional experience but who couldn't learn to score to a scale consistently or who were otherwise unsuitable for assignment to large-scale scoring projects. The PASC staff member who presented the workshop evaluated each potential Reader and submitted these evaluations to the Training Supervisor/Site Supervisor with his/her recommendations. Those who successfully completed the workshop were to Harcourt's general pool of Readers who were potential scorers of Math assessments. This addition to the general pool did not necessarily qualify these Readers for scoring the MSDE program.

Team Leader Selection and Qualification

The training for new TLs consisted of a two day course focusing on the duties and responsibilities necessary to successfully manage a team of Readers. The workshop was led by two PASC Training Supervisors. The instruction included a review of PASC policies and procedures, sessions on use of the Reader monitoring reports to track a Reader's speed and accuracy, practice annotating anchors and simulated training of the annotated papers, role playing activities which explored various situations that could occur with Readers during the scoring of a project, and Reader counseling and retraining guidelines. Hands-on training on the various TL computer applications were also covered in the work shop. Upon completion of the workshop, the two PASC Training Supervisors reviewed each participant's performance making sure that each had a complete understanding of the TL role and its responsibilities. Any participant they found who had not performed to their satisfaction was not added to the qualified TL list.

Team Leader Project Training

Project-specific TL training for MSDE was conducted in the days immediately preceding scoring and Reader training. This training begun with the RD reading the rubrics aloud and answering any questions the TL or assistant RD might have regarding the rubric. The RD then read each anchor paper aloud to the TLs. Each response in the anchor set was thoroughly explained including the notes and comments of the anchor-pulling committee. Training set A was reviewed next. The TLs scored the training set individually, recorded the scores on the answer sheet and then waited for all TLs to complete the scoring. When everyone had completed scoring the training set, the RD discussed the answers one-by-one, focusing on why it was that score and not another. The RD reviewed with the group the reason for assigning each score point and discussed each paper in its entirety. The TLs were then ready to score Training set B. Training set B was scored and reviewed exactly as Training set A.

Having thoroughly discussed both training sets with the group, the RD explained that in order for a participant to qualify as a TL, it was required that the TL should score at least an 80% perfect match on both of the qualifying sets (Qualification Rules, Attachment M). The TLs scored the first qualifying set individually and recorded their scores on the appropriate answer sheet. As each TL finished scoring, he/she brought the answer sheet to the RD for grading. Each answer was reviewed and any questions the TL had were addressed before the TL attempted the next qualifying set; the TL followed the same procedure with Qualifying set 2. Upon completing the second qualifying set, the TL submitted the answer sheet to the RD for grading. The TL had to achieve both sets for Math Step B and 90% in Math Step A as specified in the qualification rules or they would be released from the MSDE project.

After the qualification process, the RD continued the training process with the decision set. This set was read aloud and each paper thoroughly explained and discussed. By following these procedures, the RD ensured that the anchor-pulling committees' notes and comments were completely understood.

Team Leader Duties

TLs were responsible for monitoring the training and qualifying of the Readers assigned to their team. The TLs assisted the RD, if requested, during the training of the Readers. The TL was responsible for grading the Readers' qualifying sets and discussing the results with the Readers so everyone received the same direction. The TL certified to the RD and Training Supervisor that the Reader was qualified and recorded the scores under Qualification scores on the Reader evaluation

form. The TL was also responsible for monitoring each Reader's assignment of scores to the responses. Additionally, the TL reviewed the daily Reader statistical reports with each individual on the team. The TL consulted the RD regarding variations by the team members from the acceptable standards (95% for Math Step A, and 85% for Math Step B). The TL had the initial responsibility to see that the Reader maintained the set standards through individual retraining. The RD monitored the TL by reviewing team statistics and working one on one with the TL.

Room Director Selection and Qualification

The candidates for RD had been recommended by the PASC Managers or Training Supervisors. The recommendations were based upon the evaluations the candidates received as Readers and TLs and were part of their personnel file. The Training Supervisors met as a group to discuss who might be considered for the position of RD. The Training Supervisor group reviewed the evaluations and the duties that the potential RDs had performed. The candidates generally had been TLs on large-scale projects for multiple teams, and/or they had served as TLs on small-scale projects where TLs trained their individual teams. They had been evaluated on their ability to train Readers as well as their ability to monitor the scoring accuracy and consistency of Readers. These evaluations were submitted in writing at the end of each scoring project by the Readers and RDs that had observed the work of the RD candidates.

Room Director Project Training

The RDs familiarized themselves with the rubric. Any questions regarding the rubric were addressed by the PASC Language Arts and Math Specialists, or MSDE. The next step was for the RD/TTL to prepare the anchors by annotating each response to all score points in the Anchor Set utilizing the notes from the anchor-pulling session. The MSDE approved the anchor-pulling notes and the Training Supervisor confirmed that the RD had accurately added the anchor-pulling notes to the training materials. The RD continued the process by annotating the training sets and decision sets with all notes and comments from the anchor-pulling session. Additionally, the RDs became familiar with the wording of all of the other prompts for the administration on which they are assigned.

Room Director Duties

The RD's job was to conduct the training of the TLs and Readers, oversee the actual scoring of the papers, monitor the work of the TL, and act as the decision maker for situations or questions that may arise during the scoring process. For example, all invalid (foreign language, off-topic, off-mode, etc.) responses were reviewed by the RD, who had to confirm any such decision and ensure consistency of decisions (Blanks were confirmed at the TL level and did not require RD confirmation). Additionally the RD and TL (after approval of Training Supervisor) conducted all resolution readings. Responses for which scores were non-matching or non-adjacent were automatically routed to the RD for an independent resolution scoring. The resolution score became the reported score.

The RD was familiar with all prompts and trained the TLs and Readers to recognize these alternate prompts. Thus, should the student place his/her answer in the wrong place, the answer was recognized by the RD, who could electronically move the response to the appropriate space for scoring by a Reader qualified on the appropriate prompt. The RD also reviewed any potential questionable content responses and forwarded those to the Training Supervisor to consult with the MSDE before processing.

The RD was also responsible for daily statistical review and analysis of all monitoring reports to ensure the quality of the scoring within the room. Review of the data allowed the RD not only to monitor the Reader but also to provide the TL with additional input. Available data included 1) individual Reader agreement rates between two independent scorings; 2) score point distributions by Reader and trend review; 3) prompt statistics for agreement rates and score point distributions; 4) Resolution data.

Project Scoring Parameters

MSDE had a long-standing history of implementing assessments that were composed of multiple item types: selected response (SR), brief constructed response (BCR), extended constructed response (ECR), and gridded or student-produced response (SPR). The MSA contained all such item types for operational scoring and each of the 10 forms per grade/subject also contained field-test items of each of these types. Open-ended items were scored using a generic rubric as follows:

- Mathematics BCR items: Step A 0-1 scale, Step B 0-2 scale
- ECR items Step A 0-1 scale, Step B 0-3 scale

All MSA response documents were image scanned at Harcourt's scoring center in San Antonio, Texas. The image scanner captured document identification (ID), demographic information, SR responses, and creates a bi-tonal image of the entire document, allowing images of the BCR and ECR responses to be distributed to readers for human scoring while images of the SR, SPR and all other data were made available to Scoring Editing for human review.

All constructed responses were scored by Harcourt's Performance Assessment Scoring Center (PASC). The PASC mission was to provide accurate, reliable, on-time scores for all student responses entrusted to our care. PASC maintained large pools of qualified, trained, professional readers who were well-experienced in scoring a wide range of writing assessments and open-ended assessments in reading, mathematics, science, social science, and other subjects, at each of our scoring sites.

Reader Project Training

Reader training was lead by the RD/TTL and was conducted utilizing our central scoring model. There was one RD responsible for each site, grade and Domain (Math). After all student responses were scored for the first item, the RD reconvened the group and trained the second item. Training began with the definition and an overview of holistic scoring. Training continued with a reading and discussion of the generic rubric and then the student responses in the anchor set were read and discussed. In the anchor set the scores had been recorded on the student responses and were arranged in ascending point-scale order. Each annotated anchor response was read aloud and discussed thoroughly. Emphasis was placed on the Readers' understanding of how the responses differed from one another in incremental quality and how each response reflected the description of its score point as generalized in the scoring rubric and how each reflected the MSDE's standard for application of each score point.

Once Readers had all their questions answered and the discussion of the anchor set was finished, the Readers began to score the first training set. Each Reader independently read and scored the responses in the training set. The trainer scored and recorded each reader's responses on a training record form. The correct scores were then read to the group when everyone had completed the scoring. In addition, each training paper was discussed as to reasons for applying each given score. At this point, Readers interacted with the RD in discussing the characteristics of each

response that earned the assigned score point. The same format was followed for each training set. During this process, the job of the Reader was to internalize the scoring scale and adjust his or her individual scoring to conform to that scale. Once all training papers had been scored and fully discussed, Readers began the qualifying process.

For MSDE, there were three qualifying sets. MSDE informed PASC in writing for each specific administration how many qualifying sets were approved and were available to the Readers. Readers must score an 80% on at least one of two for Math.

Inter-Rater Agreement

Harcourt's scoring system generated many kinds of internal monitoring reports that enabled the project leadership to monitor the accuracy and consistency of MSDE scoring. These reports were compiled by prompt, listed the entire prompt's Readers and provided the results of their scoring for each day. Information on these reports included the number of responses read by the Readers during the period, the number and percent of invalid responses and the number of responses for which there had been a second reading. The number of responses with second readings provided data that allowed for reporting of the number and percent of responses with perfect agreement; the number and percent of responses on which the first Reader was a point lower than the second Reader; the number and percent of responses on which the first Reader was a point higher than the second Reader (Adjacent) and the number and percent of responses differing by more than one score point (Non-Adjacent/Non-Perfect). The Training Supervisor also reviewed the daily statistical reports to identify individuals or teams who might need retraining in order to provide continuous scoring consistency on the project. MSDE received data summary reports. Statistical summaries of inter-rater reliability can be found in section 3.4, *inter-rater reliability*.

Reader Retraining

When a Reader's performance fell below acceptable parameters for a project, the Reader was retrained. Retraining was the process by which the RD or TL utilized a number of methods such as individual tutoring on problem score points, individual review of selected responses and anchor and rubric review to get a Reader back on track with the guidelines provided by a specific program. Group retraining was conducted by the RD every Monday (or following any extended break) during the scoring project. In addition, daily retraining occurred as deemed necessary by the MSDE representative and Training Supervisor.

Read Behinds

Harcourt's system allowed TLs and/or RDs to conduct read behinds as an additional monitoring method. When conducting read behinds, the TL or RD received images of student responses and the scores assigned by the Reader. Responses selected for read behinds might be randomly selected or might be targeted read behinds (i.e., responses receiving specific scores, etc.). These read behinds were very useful in tracking specific areas of confusion for a given Reader or group of Readers and assisted the TL and RD in knowing just how to direct retraining activities for individual Readers or teams. The initial read behind percentage was set at 50%. This percentage might be adjusted either higher or lower by the TL based upon the performance of the Reader.

Retrain Readers with <80% Agreement rates

It was the responsibility of the Team Leader ("TL") to not only address questions and provide guidance to the Readers, but to also monitor and manage performance; this included Calibrations, Read Behinds, Agreement rates and Resolution rates. At times, TLs could become easily side-

tracked and spend more time acting as a resource for Readers more so than managing performance. PASC had identified this issue and planed to allocate additional TLs whose primary job responsibility was to manage/monitor performance. This level of staffing allowed us to monitor each Reader daily and provided retraining when the level of acceptable performance had not been met.

Pre-"Live" training on Field Test prompts

For 2007, PASC used scored student responses from the appropriate field test administration. This allowed the Readers to build familiarity with the program prior to live scoring.

Trainers Earlier and Longer

In addition to increasing the number of TLs dedicated to the program, PASC also felt it more effective to expedite and extend the time the Trainers were onsite. PASC trained a qualified individual at each site to act as the remote Trainer once the primary left. This individual was responsible for re-training Readers as needed.

Technology

PASC utilized the Student Response Window ("SRW") application supplemented with the PASC Performance Monitoring ("PPM") system that provided the Reader and/or client a "real-time" look into the scoring of each item. This system allowed the viewer to filter the information to provide detail down to the prompt, item, domain, site, Reader, etc. level. This helped in reporting results and creating a custom view of the program. The most important attribute of the application was its security features. Even though Readers in the same room could access the SRW application, each Reader could be setup to view different information within a program. This allowed segregation per domain or even grade within a partitioned room. This system greatly enhanced the quality and timeliness of reporting.

Scoring rules for MSA

The following scoring rules were applied to MSA-Math BCR and ECR items:

- Math BCR (Brief Constructed Response) items were scored:
  - Step A: 0, 1 with two readings
  - Step B: 0, 1, 2 with two readings
- Math ECR (Extended Constructed Response) items were scored:
  - Step A: 0, 1 with two readings
  - Step B: 0,1,2,3 with two readings
- Score were the higher of the 1st and 2nd Readers' scores provided they were adjacent. If they are equal that was the score.
- The score result from adjacent reads was a decimal numeric; round this up to the nearest whole number.
- For example:

1st Reader	2nd Reader	Final Score
1	2	2
2	3	3

- A resolution reader was used if two non-adjacent initial scores were received.
- The resolution reader's score was used in place of both the 1st and 2nd Readers' scores.
- For example:

1st Reader	2nd Reader	Resolution Reader	Final Score
0	2	1	1
0	3	2	2
1	3	3	3
2	0	1	1
3	0	2	2

Development Procedures for Anchor Pulling

For a given math prompt, the PASC Developers had the following responsibilities (A developer was a PASC Reader who was selected by the PASC Training Supervisor to prepare sets of papers for client approval. These experienced Readers were judged by the Training Supervisor for their ability to recognize and assemble a wide variety of responses. A Material Development Evaluation was completed by the Language Arts Specialist for review by the Training Supervisor. This evaluation was part of the developer's personnel file. The developer also participated with the clients as a facilitator during the anchor-pulling session in order to make notes and be prepared to assemble the finished sets to the client's specifications. In the case of the MSDE, the developer was also the RD):

1) To know the prompt and the rubric thoroughly

2) To read responses

- Looked for responses that seemed to represent the full range of quality as described in the rubric.
- Searched all orders for responses, with particular emphasis on the state's high performing districts.
- Included not only papers that were homogeneous in their level of quality but also papers that differed in quality from variable to variable but which could be given an overall classification of High, Medium, and Low.
- Marked High, Medium, and Low papers—marked especially good ones that might be the potentially top scores.
- Identified and flagged problem papers—off-topic, off-task, verbatim copying, strange, potential teacher interference, etc.
- Marked the flag with score range or the nature of the problem and paper ID.

3) To sort copies

- Copies were sorted into piles, reflecting the nature of the flag—all potential high papers were together, all potential medium papers were together, etc., with all problem papers grouped together.
- For problem or decision papers, duplicates of types of problems were culled. The best example of each problem type was retained; the rest were set aside for possible future use.

4) To develop sets for anchor pulling

- Decided which particular papers from the sorted piles should go into which set for anchor pulling. Each paper selected went into only one set.
- Used the following guidelines in deciding for which set a paper was most appropriate.

  A. ***Anchor set***: At least three examples of each score point, depending upon the score scale (no invalids). These had to be clean papers but should illustrate different types of the same score point, if there were such clear differences. Once completed, this set was submitted to the Training Supervisor and to MSDE for review and approval.

B. ***Decision set***: This had to be a set of whatever size necessary to illustrate the various kinds of problems that might arise with this prompt or item. If the number of such responses was small, these might be incorporated into the first training set instead of being grouped into a separate additional set.

C. ***Training sets***: These were at least two sets of up to 20 papers each (again, this varied according to the score point scale). They had to contain a range of responses including clean papers, line papers, and problem papers. The responses had to be in random order of quality and unmarked.

D. ***Qualifying sets***: There were three sets of these. Generally there were 10 responses per set, but could be fewer, depending upon the score scale. These had to consist heavily of clean papers but not exclusively so. One of the sets might include an example of an invalid response, but it should be clearly so.

E. ***Calibration sets (validity sets)***: These were composed of five responses of mixed quality, arranged in random order. Harcourt created as many different sets as there were expected to be scoring days on a single prompt or group of items—minus one or two for the training day and the initial scoring day.

Comprehensive notes concerning the specific problems presented in these papers (and the solutions as decided by the committee during the anchor-pulling session) were to be recorded by the Harcourt representatives (developers and Training Specialists) and were to be discussed with the Readers during training. Any subsequent notes or communication from MSDE were incorporated into the training material as well.

Anchor Pulling Procedures

The objective of anchor pulling was for the team members to arrive at a consensus as to the score of each paper in the proposed training materials. These sessions were attended by Maryland educators, MSDE and from PASC the Language Arts and Math Specialists, Manager, Training Supervisor and the developers who selected and prepared all of the papers that would be reviewed. These papers and their corresponding scores formed the basis of selecting final Anchor Sets, Decision Sets Training Sets and Qualifying Sets. Discussions among the team members were important, as they revealed what kinds of qualities characterized certain score points. The most difficult aspects involved balancing widely discrepant qualities found in the same paper and defining the line between adjacent scores.

During formal anchor pulling, the procedure for assigning scores to the papers in each set was as follows:

- Papers were read aloud and discussed by the anchor-pulling panel. Reading aloud focused attention on the ideas presented—or what the student had to say—allowing the panel members to divorce themselves from how the paper looked or how well it had been edited.
- After each response was read, each panel member independently assigned a score. An overall tentative score was assigned to each response on which there seemed to be consensus. However, all assigned scores at this point, even those on responses for which there were complete agreement, were provisional and subject to change based on later considerations.

- Each subsequent set was read and scored by each panel member, using the tentative scores on the previous sets as guidelines. After each set had been read, the results were recorded on a consensus sheet and discussed.

The responses in which score points were not in perfect agreement were discussed starting with the lowest, but least controversial, score point. The papers that had the widest discrepancies of assigned scores around this lowest score point were discussed next before moving to the papers whose assigned scores were in the next higher range. There might be frequent reference to previous sets to make sure that decisions on score points were consistent.

This iterative process of reading, charting, and discussing successive sets had three goals:

- It established scores on papers for which there was virtual agreement.
- It identified papers that were on the line between two adjacent scores, forcing the clarification of that line.
- It contributed to understanding the rationale behind scoring decisions.

During this process, the tentative scores assigned to papers in earlier sets became firm.

1.8 Operational Test Analyses

Classical Analyses with Form-to-Form Linking Common Items

The main purpose of this analysis was to check that the groups taking the two operational forms were essentially equivalent. Descriptive statistics, such as mean (*M*), standard deviation (*SD*) were calculated with the common items that appeared on both operational test forms. The statistical results of the two test forms were almost identical across all grades, as can be seen from Table 1.24.

Table 1.24 Descriptive Statistics for the 2007 MSA-Math Form-to-Form Linking Common Items

Grade	Form	No. of Items	*N*	*M*	*SD*
3	A	33	29,897	25.40	6.41
	F	33	29,858	25.47	6.34
4	A	26	30,402	18.93	5.61
	F	26	30,103	19.05	5.63
5	A	42	31,083	27.73	9.63
	F	42	30,875	28.10	9.58
6	A	38	31,558	24.89	9.12
	F	38	31,258	24.95	9.03
7	A	25	32,264	13.71	6.47
	F	25	32,000	13.85	6.40
8	A	24	32,836	12.24	5.53
	F	24	32,480	12.29	5.46

Note. Form A designates the identical operational portion of Forms A, B, C, D, and E. Form F designates the identical operational portion of Forms F, G, H, J, and K.
Note. Analyses were conducted with a whole population.

P-Value Check with Year-to-Year Linking Common Items

Tables 1.25 through 1.36 provide information about how much the p-values of the items designated as a year-to-year linking item changed in Year 2007 from Year 2006. It should be noted that these analyses were conducted with a whole population. In general, we could conclude that most of the p-values in Year 2007 were almost the same or slightly increased compared to those in Year 2006 across all grades.

Table 1.25 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 3 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3509918	0.80	0.76	it82	3510060	0.83	0.84
it42	3564076	0.44	0.50	it83	3564078	0.50	0.53
it43	3509931	0.66	0.65	it84	3510346	0.77	0.85
it44	3510022	0.52	0.47	it85	3510033	0.78	0.79
it45	3510009	0.79	0.79	it86	3510012	0.73	0.78
it46	3509953	0.94	0.94	it87	3510062	0.84	0.85
it47	3548054	0.91	0.93	it88	3510063	0.73	0.78
it48	3509955	0.49	0.57	it89	3509983	0.92	0.91
it53	3509964	0.80	0.74	it90	3510065	0.96	0.96
it54	3509966	0.88	0.90	it91	3510066	0.80	0.80
it55	3509974	0.64	0.66	it92	3509936	0.74	0.74
it56	3509979	0.81	0.84	it93	3564079	0.43	0.48
it57	3509919	0.65	0.64	it97	3510071	0.66	0.64
it58	3564077	0.60	0.57	it98	3510072	0.85	0.85
it59	3509987	0.65	0.66	it99	3564080	0.54	0.58
it60	3510017	0.92	0.91	it100	3510126	0.75	0.78
it61	3510003	0.84	0.84	it101	3509945	0.91	0.91
it62	3510006	0.60	0.61	it103	3509957	0.77	0.77
it63	3548055	0.97	0.93	it104	3564081	0.37	0.43
it64	3510011	0.65	0.63	it105	3509958	0.84	0.87
it65	3510125	0.60	0.52	it106	3509961	0.92	0.92
it66	3510018	0.76	0.77	it107	3510068	0.84	0.81
it68	3510023	0.50	0.50	it108	3510069	0.32	0.35
it69	3510027	0.87	0.87	it109	3510070	0.97	0.97
it70	3510029	0.92	0.94	it111	3510034	0.32	0.30
it71	3510032	0.91	0.88	it112	3564082	0.25	0.32
it72	3510035	0.87	0.87	it113	3510041	0.95	0.92
it78	3510051	0.56	0.54	it114	3510043	0.80	0.76
it79	3510053	0.83	0.84	it117	3510044	0.85	0.86
it80	3510055	0.59	0.62	it119	3510329	0.52	0.55
it81	3510058	0.87	0.86				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 3 Form A

Form	Year	No. of Items	*M*	*SD*
A	Year 2006	61	0.73	0.18
	Year 2007	61	0.73	0.17

Table 1.26 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 3 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it41	3509918	0.80	0.76	it82	3510060	0.83	0.83
it42	3564076	0.44	0.51	it83	3564078	0.50	0.53
it43	3509980	0.42	0.45	it84	3510052	0.85	0.75
it44	3510022	0.52	0.49	it85	3510347	0.68	0.68
it45	3548059	0.59	0.71	it86	3510036	0.83	0.85
it46	3510071	0.66	0.64	it87	3510062	0.84	0.85
it47	3548057	0.65	0.73	it88	3510063	0.73	0.79
it48	3509955	0.49	0.56	it89	3509945	0.91	0.92
it53	3509964	0.80	0.75	it90	3509935	0.62	0.67
it54	3509966	0.88	0.90	it91	3510067	0.82	0.82
it55	3509923	0.80	0.82	it92	3564083	0.65	0.73
it56	3509959	0.68	0.70	it93	3510006	0.60	0.64
it57	3509919	0.65	0.65	it97	3509956	0.68	0.64
it58	3564077	0.60	0.57	it98	3509963	0.76	0.74
it59	3509926	0.40	0.36	it99	3564084	0.47	0.47
it60	3509960	0.79	0.76	it100	3548063	0.89	0.93
it61	3509927	0.75	0.78	it101	3509965	0.96	0.94
it62	3509928	0.86	0.88	it103	3509922	0.68	0.65
it63	3510009	0.78	0.77	it104	3564085	0.32	0.34
it64	3510069	0.32	0.33	it105	3509958	0.84	0.88
it65	3509988	0.74	0.73	it106	3509961	0.92	0.92
it66	3509929	0.52	0.54	it107	3510066	0.82	0.82
it68	3509930	0.95	0.95	it108	3509938	0.94	0.93
it69	3510018	0.77	0.78	it109	3510070	0.97	0.98
it70	3510027	0.87	0.87	it111	3509932	0.98	0.98
it71	3510029	0.92	0.95	it112	3564086	0.37	0.39
it72	3510035	0.87	0.87	it113	3510041	0.95	0.93
it78	3510053	0.83	0.85	it114	3510043	0.80	0.77
it79	3509933	0.92	0.91	it117	3510044	0.85	0.84
it80	3510051	0.56	0.51	it119	3510013	0.49	0.50
it81	3509962	0.87	0.88				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 3 Form F

Form	Year	No. of Items	*N*	*M*	*SD*
F	Year 2006	61	29,897	0.73	0.18
	Year 2007	61	29,858	0.73	0.17

Table 1.27 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 4 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3515405	0.78	0.80	it81	3515886	0.41	0.45
it42	3564160	0.51	0.57	it82	3564162	0.46	0.52
it43	3515406	0.58	0.60	it83	3515909	0.47	0.49
it44	3515407	0.81	0.85	it84	3548085	0.45	0.55
it45	3515408	0.66	0.68	it85	3548086	0.66	0.76
it46	3515410	0.84	0.81	it86	3515787	0.55	0.51
it47	3515411	0.80	0.84	it87	3515557	0.70	0.69
it48	3515421	0.79	0.82	it88	3515558	0.29	0.36
it53	3515425	0.59	0.64	it89	3515648	0.45	0.50
it54	3515426	0.39	0.44	it90	3564163	0.49	0.56
it55	3515428	0.90	0.94	it91	3515559	0.66	0.72
it56	3515447	0.41	0.45	it92	3515570	0.53	0.52
it57	3515451	0.59	0.70	it93	3515571	0.82	0.85
it58	3564161	0.51	0.67	it94	3515573	0.44	0.50
it59	3515604	0.60	0.64	it95	3515574	0.83	0.85
it60	3515456	0.75	0.81	it98	3515577	0.69	0.73
it61	3515467	0.89	0.95	it99	3564164	0.44	0.50
it62	3515840	0.64	0.65	it100	3548081	0.41	0.58
it63	3515470	0.66	0.69	it101	3515807	0.79	0.79
it64	3515705	0.71	0.75	it102	3564165	0.39	0.37
it65	3515471	0.83	0.86	it103	3515423	0.73	0.90
it66	3515479	0.69	0.73	it104	3515424	0.55	0.57
it67	3515484	0.90	0.92	it109	3515575	0.71	0.77
it68	3515486	0.56	0.57	it110	3515576	0.59	0.61
it69	3515630	0.51	0.50	it114	3515585	0.22	0.19
it70	3515631	0.76	0.77	it115	3564166	0.48	0.43
it71	3515490	0.86	0.92	it116	3515500	0.86	0.71
it76	3515514	0.88	0.89	it117	3515506	0.88	0.89
it77	3515519	0.79	0.82	it118	3548083	0.87	0.88
it78	3515533	0.80	0.85	it119	3515832	0.64	0.66
it79	3515543	0.73	0.79	it120	3548088	0.66	0.74
it80	3515545	0.81	0.86	it121	3515853	0.75	0.71

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 4 Form A

Form	Year	*N*	*M*	*SD*
A	Year 2006	64	0.65	0.17
	Year 2007	64	0.68	0.17

Table 1.28 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 4 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it41	3515595	0.73	0.77	it81	3515638	0.63	0.62
it42	3564167	0.38	0.47	it82	3564169	0.42	0.46
it43	3515407	0.81	0.85	it83	3515791	0.71	0.75
it44	3515596	0.77	0.78	it84	3515795	0.54	0.60
it45	3515447	0.41	0.46	it85	3515869	0.54	0.56
it46	3515408	0.66	0.69	it86	3515836	0.57	0.58
it47	3515599	0.64	0.71	it87	3515557	0.70	0.71
it48	3515410	0.84	0.82	it88	3515640	0.35	0.42
it53	3515600	0.74	0.75	it89	3515648	0.45	0.50
it54	3515601	0.69	0.68	it90	3564163	0.49	0.55
it55	3515602	0.49	0.53	it91	3515641	0.81	0.83
it56	3515428	0.90	0.94	it92	3515570	0.53	0.50
it57	3515603	0.54	0.55	it93	3515571	0.82	0.85
it58	3564168	0.36	0.38	it94	3515643	0.36	0.38
it59	3515604	0.60	0.64	it95	3515645	0.70	0.71
it60	3515605	0.50	0.53	it98	3515862	0.47	0.49
it61	3515456	0.75	0.80	it99	3564170	0.51	0.59
it62	3515467	0.89	0.94	it100	3548081	0.41	0.57
it63	3515606	0.87	0.91	it101	3515807	0.79	0.78
it64	3515652	0.63	0.68	it102	3564165	0.39	0.37
it65	3515471	0.83	0.86	it103	3515424	0.55	0.59
it66	3515936	0.80	0.86	it104	3515425	0.59	0.71
it67	3515486	0.56	0.57	it109	3515575	0.71	0.79
it68	3548078	0.52	0.50	it110	3515576	0.59	0.60
it69	3515630	0.51	0.50	it114	3515830	0.94	0.95
it70	3515631	0.76	0.77	it115	3564171	0.74	0.71
it71	3515632	0.69	0.71	it116	3515933	0.74	0.76
it76	3515634	0.72	0.75	it117	3515506	0.88	0.89
it77	3515635	0.56	0.60	it118	3515592	0.77	0.82
it78	3548079	0.80	0.94	it119	3515931	0.68	0.67
it79	3515636	0.53	0.54	it120	3515880	0.67	0.69
it80	3515545	0.81	0.86	it121	3515887	0.88	0.89

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 4 Form F

Form	Year	*N*	*M*	*SD*
F	Year 2006	64	0.64	0.16
	Year 2007	64	0.68	0.16

Table 1.29 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 5 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3511531	0.61	0.68	it81	3563989	0.35	0.34
it42	3563986	0.45	0.55	it82	3511458	0.89	0.92
it43	3511196	0.56	0.55	it83	3511467	0.81	0.85
it44	3511203	0.86	0.87	it84	3512627	0.86	0.88
it45	3511216	0.62	0.67	it85	3511470	0.78	0.81
it46	3512606	0.59	0.63	it86	3511479	0.56	0.51
it47	3511246	0.73	0.78	it91	3511504	0.83	0.90
it48	3512632	0.37	0.39	it92	3511513	0.83	0.85
it52	3512702	0.60	0.54	it93	3511521	0.62	0.67
it53	3511307	0.38	0.41	it94	3556476	0.54	0.49
it54	3511312	0.38	0.39	it95	3563990	0.49	0.46
it57	3511336	0.33	0.33	it96	3511563	0.70	0.62
it58	3563987	0.30	0.34	it97	3511258	0.79	0.81
it59	3511339	0.57	0.62	it98	3563991	0.41	0.49
it60	3511345	0.84	0.92	it99	3511266	0.67	0.71
it61	3511348	0.49	0.57	it100	3511320	0.86	0.91
it62	3511626	0.85	0.81	it101	3512595	0.78	0.79
it63	3511371	0.48	0.53	it102	3511483	0.35	0.38
it64	3511376	0.82	0.81	it103	3563992	0.29	0.34
it65	3512638	0.66	0.64	it104	3511499	0.64	0.63
it66	3511396	0.82	0.84	it105	3511330	0.55	0.63
it67	3511410	0.68	0.67	it107	3511269	0.82	0.81
it68	3512618	0.35	0.45	it108	3511566	0.64	0.66
it69	3563988	0.40	0.52	it109	3511455	0.75	0.79
it70	3511429	0.75	0.75	it110	3563993	0.61	0.67
it71	3511433	0.94	0.97	it111	3511442	0.55	0.61
it72	3511439	0.76	0.79	it112	3512710	0.57	0.59
it77	3512616	0.33	0.44	it113	3512687	0.54	0.52
it78	3512625	0.85	0.88	it114	3512628	0.71	0.77
it79	3512714	0.92	0.91	it115	3511448	0.78	0.76
it80	3512649	0.27	0.27				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 5 Form A

Form	Year	*N*	*M*	*SD*
A	Year 2006	61	0.63	0.19
	Year 2007	61	0.65	0.18

Table 1.30 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 5 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it43	3512527	0.68	0.63	it82	3512578	0.86	0.88
it44	3512528	0.84	0.87	it83	3511467	0.81	0.86
it45	3512606	0.59	0.65	it84	3512605	0.89	0.94
it46	3511216	0.62	0.68	it85	3511470	0.78	0.82
it47	3511246	0.73	0.79	it86	3511479	0.56	0.58
it48	3512529	0.59	0.56	it91	3511504	0.83	0.89
it52	3512702	0.60	0.53	it92	3511513	0.83	0.86
it53	3511307	0.38	0.41	it93	3511521	0.62	0.69
it54	3511312	0.38	0.40	it94	3556476	0.54	0.51
it57	3511336	0.33	0.33	it95	3563990	0.49	0.47
it58	3563987	0.30	0.34	it96	3511563	0.70	0.64
it59	3512534	0.64	0.69	it97	3512530	0.61	0.64
it60	3511345	0.84	0.92	it98	3563999	0.46	0.45
it61	3511348	0.49	0.56	it99	3511266	0.67	0.70
it62	3512540	0.55	0.57	it100	3511320	0.86	0.93
it63	3511371	0.48	0.54	it101	3512595	0.78	0.77
it64	3512543	0.73	0.72	it102	3511483	0.35	0.38
it65	3512546	0.83	0.84	it103	3563992	0.29	0.35
it66	3512638	0.66	0.63	it104	3511499	0.64	0.64
it67	3512553	0.58	0.59	it105	3511330	0.55	0.63
it68	3512618	0.35	0.45	it107	3511269	0.82	0.81
it69	3563988	0.40	0.53	it108	3512637	0.79	0.80
it70	3511439	0.76	0.78	it109	3512559	0.83	0.84
it71	3511410	0.68	0.70	it110	3564001	0.55	0.59
it72	3511396	0.82	0.85	it111	3511442	0.55	0.61
it77	3512612	0.38	0.38	it112	3512648	0.45	0.48
it78	3512696	0.83	0.87	it113	3512688	0.41	0.42
it79	3512691	0.51	0.52	it114	3511631	0.74	0.76
it80	3512649	0.27	0.29	it115	3511448	0.78	0.80
it81	3563989	0.35	0.36				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 5 Form F

Form	Year	*N*	*M*	*SD*
F	Year 2006	59	0.61	0.18
	Year 2007	59	0.64	0.18

Table 1.31 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 6 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3516240	0.55	0.56	it75	3564004	0.43	0.57
it42	3516241	0.82	0.84	it76	3516313	0.80	0.83
it43	3516243	0.64	0.69	it77	3516318	0.86	0.88
it44	3516242	0.35	0.38	it78	3516327	0.41	0.44
it45	3516912	0.52	0.54	it79	3564005	0.53	0.59
it46	3516248	0.72	0.75	it84	3517000	0.45	0.51
it47	3516247	0.52	0.55	it85	3517010	0.44	0.48
it48	3516249	0.66	0.67	it86	3516328	0.71	0.75
it49	3516452	0.64	0.64	it87	3516293	0.39	0.45
it50	3564002	0.44	0.47	it88	3516330	0.65	0.79
it51	3516255	0.67	0.70	it89	3516331	0.38	0.41
it52	3516256	0.56	0.60	it94	3516352	0.73	0.77
it53	3516257	0.79	0.83	it95	3516353	0.57	0.58
it54	3516258	0.53	0.54	it96	3516354	0.72	0.72
it55	3516279	0.71	0.73	it97	3516355	0.62	0.66
it56	3516280	0.46	0.50	it98	3516627	0.56	0.52
it57	3516281	0.43	0.44	it99	3564006	0.38	0.42
it58	3516283	0.42	0.43	it100	3516284	0.47	0.52
it61	3516285	0.52	0.54	it101	3564007	0.35	0.42
it62	3516290	0.60	0.64	it102	3516351	0.50	0.51
it63	3516291	0.43	0.47	it103	3516332	0.48	0.51
it66	3516298	0.26	0.29	it104	3516329	0.50	0.62
it67	3516573	0.62	0.67	it105	3516295	0.55	0.65
it68	3516301	0.60	0.67	it106	3516333	0.54	0.60
it69	3516302	0.66	0.69	it107	3564008	0.46	0.61
it70	3516303	0.47	0.53	it112	3516326	0.74	0.77
it71	3516305	0.64	0.68	it113	3564009	0.56	0.58
it72	3516307	0.54	0.61	it114	3516320	0.83	0.90
it73	3516310	0.65	0.69	it115	3516323	0.60	0.67
it74	3517013	0.30	0.35				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 6 Form A

Form	Year	*N*	*M*	*SD*
A	Year 2006	59	0.56	0.14
	Year 2007	59	0.60	0.14

Table 1.32 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 6 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it41	3516240	0.55	0.55	it75	3564004	0.43	0.56
it42	3516429	0.89	0.92	it76	3516305	0.64	0.68
it43	3516242	0.35	0.39	it77	3516320	0.83	0.92
it44	3516912	0.52	0.54	it78	3516327	0.41	0.42
it45	3516243	0.64	0.69	it79	3564005	0.53	0.59
it46	3516247	0.52	0.56	it84	3517000	0.45	0.49
it47	3516248	0.72	0.75	it85	3516907	0.55	0.61
it48	3516451	0.72	0.74	it86	3516328	0.71	0.74
it49	3517004	0.80	0.87	it87	3516293	0.39	0.46
it50	3564010	0.50	0.58	it88	3516618	0.34	0.34
it51	3516255	0.67	0.72	it89	3516621	0.79	0.71
it52	3516256	0.56	0.61	it94	3516623	0.72	0.74
it53	3516280	0.46	0.51	it95	3516624	0.25	0.23
it54	3516453	0.73	0.76	it96	3516625	0.79	0.84
it55	3516454	0.78	0.80	it97	3516354	0.72	0.72
it56	3516455	0.49	0.49	it98	3516627	0.56	0.52
it57	3517002	0.72	0.74	it99	3564006	0.38	0.41
it58	3516517	0.32	0.32	it100	3516284	0.47	0.53
it61	3516559	0.81	0.84	it101	3564007	0.35	0.43
it62	3516565	0.42	0.44	it102	3516332	0.48	0.55
it63	3516571	0.34	0.35	it103	3516351	0.50	0.51
it66	3516291	0.43	0.46	it104	3516329	0.50	0.65
it67	3516573	0.62	0.69	it105	3516295	0.55	0.65
it68	3516301	0.60	0.69	it106	3516622	0.38	0.42
it69	3516302	0.66	0.69	it107	3564011	0.43	0.49
it70	3516303	0.47	0.53	it112	3516616	0.39	0.40
it71	3516594	0.76	0.80	it113	3564012	0.44	0.49
it72	3516313	0.80	0.83	it114	3516318	0.86	0.84
it73	3516613	0.51	0.55	it115	3516323	0.60	0.64
it74	3517013	0.30	0.36				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 6 Form F

Form	Year	*N*	*M*	*SD*
F	Year 2006	59	0.56	0.14
	Year 2007	59	0.60	0.14

Table 1.33 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 7 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3517744	0.42	0.35	it79	3564020	0.38	0.40
it42	3564018	0.21	0.24	it80	3517757	0.28	0.35
it43	3517604	0.33	0.32	it81	3517704	0.38	0.43
it44	3517601	0.44	0.45	it82	3517759	0.45	0.43
it45	3517609	0.50	0.50	it83	3517719	0.21	0.26
it46	3517613	0.63	0.62	it84	3564021	0.36	0.41
it47	3517616	0.55	0.55	it85	3517709	0.64	0.64
it48	3517634	0.61	0.63	it86	3517712	0.41	0.45
it49	3517642	0.45	0.42	it87	3517714	0.50	0.54
it50	3517638	0.72	0.69	it88	3517716	0.57	0.61
it51	3517647	0.64	0.65	it89	3517718	0.55	0.61
it52	3517643	0.63	0.66	it90	3517721	0.41	0.42
it53	3517650	0.60	0.60	it91	3517723	0.39	0.39
it54	3517652	0.63	0.66	it92	3555858	0.33	0.39
it59	3547473	0.68	0.77	it93	3547477	0.39	0.49
it60	3517663	0.26	0.27	it94	3517725	0.25	0.26
it61	3517665	0.35	0.35	it95	3564022	0.35	0.40
it62	3517667	0.61	0.57	it98	3517730	0.57	0.58
it63	3517670	0.27	0.30	it99	3517732	0.29	0.31
it64	3564019	0.12	0.15	it102	3517656	0.59	0.63
it65	3517675	0.64	0.68	it103	3517736	0.49	0.51
it66	3555857	0.34	0.36	it104	3517818	0.31	0.33
it67	3517681	0.53	0.56	it105	3564023	0.30	0.38
it68	3517683	0.43	0.46	it107	3517876	0.10	0.14
it69	3517678	0.89	0.88	it108	3547482	0.15	0.17
it70	3517710	0.68	0.61	it109	3564024	0.30	0.35
it71	3517742	0.47	0.50	it110	3517779	0.45	0.64
it72	3517687	0.54	0.56	it111	3517697	0.31	0.37
it73	3517692	0.77	0.79	it112	3517733	0.50	0.53
it74	3517694	0.73	0.73	it113	3555859	0.72	0.74
it78	3517673	0.62	0.65				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 7 Form A

Form	Year	*N*	*M*	*SD*
A	Year 2006	61	0.46	0.17
	Year 2007	61	0.49	0.17

Table 1.34 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 7 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it41	3517706	0.54	0.47	it79	3564027	0.57	0.58
it42	3564025	0.29	0.28	it80	3517695	0.36	0.35
it43	3517613	0.63	0.62	it81	3517729	0.64	0.68
it44	3555861	0.64	0.72	it82	3517757	0.28	0.33
it45	3517604	0.33	0.32	it83	3517693	0.13	0.16
it46	3517602	0.39	0.45	it84	3564028	0.40	0.45
it47	3517638	0.72	0.69	it85	3517709	0.64	0.66
it48	3517679	0.41	0.49	it86	3517712	0.41	0.45
it49	3517609	0.50	0.49	it87	3517714	0.50	0.56
it50	3517643	0.66	0.66	it88	3517716	0.57	0.63
it51	3517740	0.49	0.53	it89	3517662	0.45	0.47
it52	3517631	0.72	0.71	it90	3517721	0.41	0.44
it53	3517634	0.61	0.62	it91	3517664	0.73	0.80
it54	3517665	0.35	0.34	it92	3517752	0.60	0.62
it59	3517635	0.68	0.67	it93	3517885	0.34	0.35
it60	3517615	0.66	0.68	it94	3517666	0.25	0.27
it61	3517637	0.71	0.74	it95	3564029	0.36	0.40
it62	3517639	0.28	0.28	it98	3517668	0.32	0.34
it63	3517670	0.27	0.29	it99	3517671	0.32	0.34
it64	3564019	0.12	0.16	it102	3517650	0.60	0.60
it65	3517675	0.64	0.69	it103	3517652	0.62	0.67
it66	3555864	0.20	0.22	it104	3517715	0.74	0.81
it67	3517683	0.44	0.45	it105	3564030	0.44	0.50
it68	3517645	0.64	0.69	it107	3517758	0.23	0.24
it69	3517741	0.90	0.91	it108	3547487	0.66	0.77
it70	3517812	0.49	0.54	it109	3564031	0.23	0.31
it71	3547535	0.59	0.76	it110	3555865	0.27	0.34
it72	3517687	0.54	0.57	it111	3517718	0.55	0.64
it73	3517692	0.77	0.79	it112	3517756	0.40	0.44
it74	3517694	0.73	0.75	it113	3555859	0.72	0.76
it78	3517648	0.62	0.63				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 7 Form F

Form	Year	*N*	*M*	*SD*
F	Year 2006	61	0.50	0.18
	Year 2007	61	0.53	0.19

Table 1.35 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 8 Form A

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FA	Y07 FA
it41	3514015	0.23	0.23	it80	3514093	0.32	0.33
it42	3514014	0.52	0.56	it81	3514107	0.13	0.12
it43	3514016	0.72	0.75	it82	3514103	0.54	0.60
it44	3514046	0.51	0.52	it83	3514608	0.43	0.41
it45	3514013	0.43	0.44	it84	3514287	0.58	0.62
it46	3564107	0.58	0.64	it85	3514267	0.29	0.35
it47	3547550	0.67	0.57	it86	3564110	0.43	0.62
it53	3514056	0.74	0.79	it87	3514113	0.65	0.65
it54	3514053	0.70	0.71	it88	3514275	0.74	0.72
it55	3514058	0.30	0.30	it92	3514117	0.28	0.32
it56	3514059	0.59	0.63	it93	3564111	0.32	0.39
it57	3514062	0.39	0.41	it94	3514279	0.20	0.21
it58	3514702	0.28	0.28	it96	3514131	0.32	0.39
it59	3564108	0.29	0.34	it97	3514057	0.65	0.65
it60	3514064	0.16	0.14	it98	3514607	0.23	0.26
it61	3514276	0.43	0.45	it99	3564112	0.21	0.24
it62	3514127	0.24	0.22	it100	3514055	0.53	0.57
it63	3514125	0.62	0.60	it101	3514052	0.50	0.50
it64	3514121	0.68	0.69	it102	3514118	0.08	0.09
it65	3514139	0.64	0.73	it103	3564113	0.22	0.40
it66	3514073	0.56	0.55	it104	3514291	0.67	0.73
it67	3514074	0.42	0.42	it105	3514606	0.69	0.69
it68	3514075	0.60	0.63	it106	3514076	0.45	0.46
it69	3514078	0.21	0.22	it107	3514100	0.63	0.75
it70	3564109	0.24	0.31	it111	3514080	0.53	0.52
it73	3514611	0.61	0.65	it112	3514079	0.27	0.31
it74	3514083	0.18	0.24	it113	3514669	0.53	0.51
it76	3514092	0.42	0.42	it114	3564114	0.64	0.63
it77	3514102	0.54	0.62	it116	3514710	0.54	0.53
it78	3514095	0.29	0.31				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 8 Form A

Form	Year	*N*	*M*	*SD*
A	Year 2006	59	0.45	0.18
	Year 2007	59	0.47	0.18

Table 1.36 Year 2006 vs. Y2007 Linking Common Item P-Value Comparison: Grade 8 Form F

Item Seq. No.	Item CID	Y06 FF	Y07 FF	Item Seq. No.	Item CID	Y06 FF	Y07 FF
it41	3514015	0.23	0.22	it80	3514100	0.63	0.69
it42	3514014	0.52	0.56	it81	3514138	0.60	0.58
it43	3514016	0.72	0.77	it82	3514213	0.56	0.63
it44	3514055	0.53	0.58	it83	3514103	0.54	0.58
it45	3514147	0.36	0.38	it84	3547555	0.51	0.51
it46	3564115	0.25	0.33	it85	3514108	0.12	0.13
it47	3514052	0.50	0.51	it86	3564118	0.19	0.23
it53	3514058	0.30	0.30	it87	3514263	0.54	0.51
it54	3514062	0.39	0.42	it88	3514111	0.40	0.42
it55	3514059	0.59	0.63	it92	3514117	0.28	0.32
it56	3514156	0.74	0.73	it93	3564111	0.32	0.40
it57	3514056	0.74	0.75	it94	3514708	0.59	0.57
it58	3514283	0.33	0.34	it96	3514114	0.43	0.41
it59	3564116	0.43	0.49	it97	3514046	0.51	0.52
it60	3514159	0.63	0.62	it98	3514152	0.33	0.33
it61	3514161	0.19	0.19	it99	3564119	0.32	0.34
it62	3514162	0.37	0.37	it100	3547550	0.67	0.64
it63	3514163	0.54	0.52	it101	3547551	0.83	0.85
it64	3514122	0.58	0.60	it102	3514266	0.29	0.29
it65	3514092	0.42	0.42	it103	3564120	0.43	0.49
it66	3514075	0.60	0.64	it104	3547547	0.46	0.54
it67	3514073	0.56	0.54	it105	3514288	0.59	0.58
it68	3514076	0.45	0.48	it106	3514074	0.42	0.41
it69	3514164	0.52	0.55	it107	3514102	0.54	0.64
it70	3564117	0.37	0.45	it111	3514083	0.18	0.26
it73	3514090	0.54	0.63	it112	3514611	0.61	0.62
it74	3514281	0.36	0.21	it113	3514133	0.30	0.37
it76	3514173	0.47	0.49	it114	3564121	0.41	0.49
it77	3514095	0.29	0.29	it116	3547536	0.48	0.49
it78	3514174	0.44	0.58				

Note. Analyses were conducted with a whole population.

Descriptive Statistics for Year-to-Year Linking Common Items: Grade 8 Form F

Form	Year	*N*	*M*	*SD*
F	Year 2006	59	0.46	0.15
	Year 2007	59	0.48	0.16

Validation Check with Year 2007 Operational Items

To collect information about how much the same items that appeared on the test forms in consecutive years changed in terms of item difficulty, difficulty indices such as p-value and Rasch difficulty were calculated.

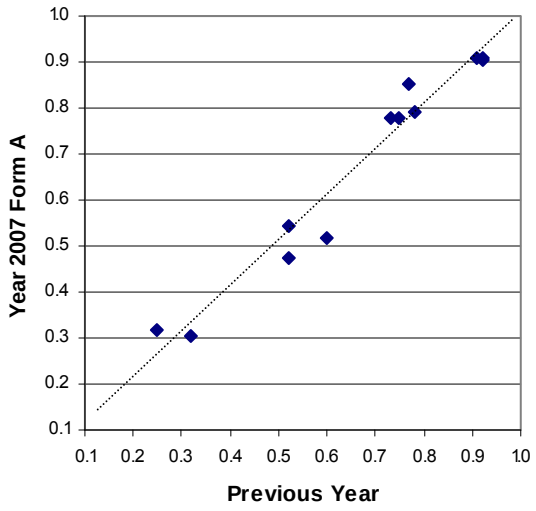
First, it should be noted these items were at first augmented as field test items in previous years and then appeared as operational test items in Year 2007. Second, Year 2007 Forms A, B, C, E, and E are the same, and Year 2007 Forms F, G, H, J, and K are the same except for the field test portion. More detailed information about the specific test design and construction of Year 2007 can be obtained from section 1.5, Test Structure of the 2007 MSA-Math.

First of all, it should be noted that p-value in previous years was calculated with a field-tested sample and Year 2007 p-value was calculated with a whole population. P-value of BCR item was the item mean score divided by the item score range. In addition, the numbers in "Omits" in each table were very substantial and included students who did not responded at all. Item p-value (easiness) results indicated that in general, most of the p-values in Year 2007 were almost the same or somewhat increased compared to those in previous years across all grades.

With respect to Rasch difficulty analysis, most of the items in Year 2007 were almost the same or somewhat easier compared to those in previous years across all grades. It should be noted that Rasch difficulties were based on the same scale (e.g., linked to Year 2006).

In conclusion, both p-value and Rasch difficulty results reflected the same phenomenon, indicating that most of the items became easier.

Table 1.37 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 3 Form A



Item CID	Previous Year	Year 07 Form A
3510022	0.52	0.47
3510017	0.92	0.91
3510125	0.60	0.52
3510346	0.77	0.85
3510033	0.78	0.79
3510012	0.73	0.78
3509983	0.92	0.91
3510126	0.75	0.78
3509945	0.91	0.91
3510034	**0.32**	**0.30**
3564082	**0.25**	**0.32**
3510329	0.52	0.55

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.38 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 3 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3510034	BCR	2,733	0.32	0.47	66.92	32.38	N/A	N/A	0.70
2006	3564082	BCR	2,733	0.50	0.47	58.76	30.85	9.37	N/A	1.02
2007	3510034	BCR	29,897	0.30	0.46	68.51	30.47	N/A	N/A	1.02
2007	3564082	BCR	29,897	0.63	0.62	42.80	47.44	7.89	N/A	1.87

Table 1.39 Augmented IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 3 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	44	3510022	SR	1.7764		
2006	60	3510017	SR	-1.2060		
2006	65	3510125	SR	1.3096		
2006	84	3510346	SR	0.4046		
2006	85	3510033	SR	0.2523		
2006	86	3510012	SR	0.6353		
2006	89	3509983	SR	-1.1855		
2006	100	3510126	SR	0.3985		
2006	101	3509945	SR	-1.0471		
2006	111	3510034	BCR	2.8051		
2006	112	3564082	BCR	3.2825	-0.7346	0.7346
2006	119	3510329	SR	1.8416		
2007	44	3510022	SR	2.0077		
2007	60	3510017	SR	-1.1315		
2007	65	3510125	SR	1.6971		
2007	84	3510346	SR	-0.5185		
2007	85	3510033	SR	0.0473		
2007	86	3510012	SR	0.0993		
2007	89	3509983	SR	-1.1470		
2007	100	3510126	SR	0.1797		
2007	101	3509945	SR	-1.1209		
2007	111	3510034	BCR	2.8934		
2007	112	3564082	BCR	3.0491	-1.5541	1.5541
2007	119	3510329	SR	1.5719		

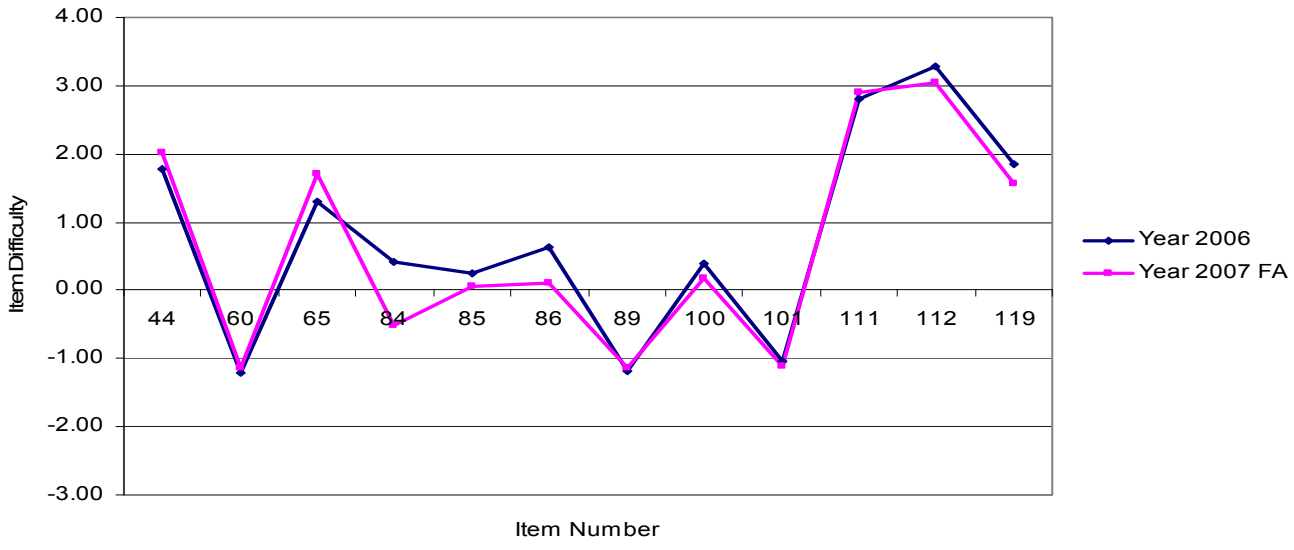
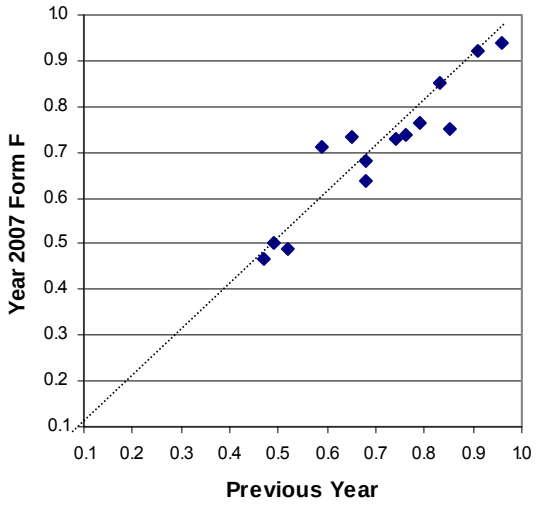


Figure 1.2 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 3 Form A

Table 1.40 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 3 Form F



Item CID	Previous Year	Year 07 Form F
3510022	0.52	0.49
3548059	0.59	0.71
3548057	0.65	0.73
3509960	0.79	0.76
3509988	0.74	0.73
3510052	0.85	0.75
3510347	0.68	0.68
3510036	0.83	0.85
3509945	0.91	0.92
3509956	0.68	0.64
3509963	**0.76**	**0.74**
3564084	**0.47**	**0.47**
3509965	0.96	0.94
3510013	0.49	0.50

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.41 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 3 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3509963	BCR	2,754	0.76	0.43	16.99	76.11	N/A	N/A	6.90
2006	3564084	BCR	2,754	0.94	0.29	10.57	82.57	5.63	N/A	1.23
2007	3509963	BCR	29,858	0.74	0.44	17.49	73.73	N/A	N/A	8.78
2007	3564084	BCR	29,858	0.93	0.36	8.54	86.94	3.27	N/A	1.25

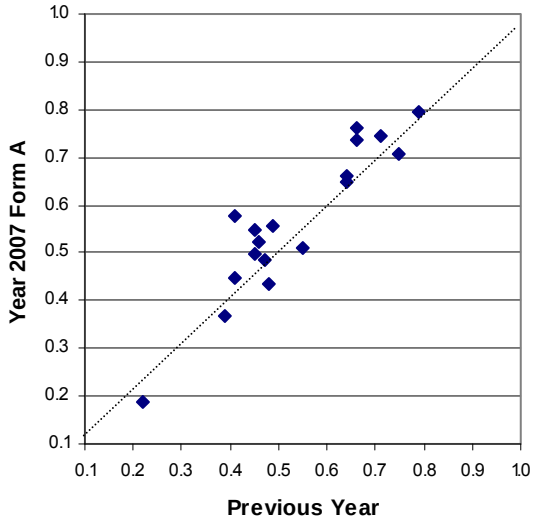
Table 1.42 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 3 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	44	3510022	SR	1.7764		
2004	45	3548059	SR	0.9222		
2005	47	3548057	SR	0.8586		
2006	60	3509960	SR	0.2320		
2006	65	3509988	SR	0.5483		
2006	84	3510052	SR	-0.4111		
2006	85	3510347	SR	0.9718		
2006	86	3510036	SR	-0.1787		
2006	89	3509945	SR	-1.0471		
2006	97	3509956	SR	0.9445		
2006	98	3509963	BCR	0.1860		
2006	99	3564084	BCR	2.3419	-3.0004	3.0004
2006	101	3509965	SR	-1.8621		
2006	119	3510013	SR	1.9699		
2007	44	3510022	SR	2.0077		
2007	45	3548059	SR	0.6288		
2007	47	3548057	SR	0.5502		
2007	60	3509960	SR	0.3981		
2007	65	3509988	SR	0.5005		
2007	84	3510052	SR	0.5884		
2007	85	3510347	SR	0.9229		
2007	86	3510036	SR	-0.5397		
2007	89	3509945	SR	-1.1209		
2007	97	3509956	SR	1.1953		
2007	98	3509963	BCR	0.1298		
2007	99	3564084	BCR	1.9318	-3.8671	3.8671
2007	101	3509965	SR	-1.6657		
2007	119	3510013	SR	1.8364		

Note. These Rasch difficulties were based on a common scale.

Figure 1.3 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 3 Form F

Table 1.43 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 4 Form A



Item CID	Previous Year	Year 07 Form A
3515840	0.64	0.65
3515705	0.71	0.75
3515886	**0.41**	**0.45**
3564162	**0.46**	**0.52**
3515909	0.47	0.49
3548085	0.45	0.55
3548086	0.66	0.76
3515787	0.55	0.51
3515648	**0.45**	**0.50**
3564163	**0.49**	**0.56**
3548081	0.41	0.58
3515807	**0.79**	**0.79**
3564165	**0.39**	**0.37**
3515585	**0.22**	**0.19**
3564166	**0.48**	**0.43**
3515832	0.64	0.66
3548088	0.66	0.74
3515853	0.75	0.71

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.44 Item Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 4 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3515886	BCR	2,799	0.41	0.49	55.88	41.01	N/A	N/A	3.11
2006	3564162	BCR	2,799	0.92	0.45	21.22	58.81	16.36	N/A	3.61
2006	3515648	BCR	2,875	0.45	0.50	53.95	44.94	N/A	N/A	1.11
2006	3564163	BCR	2,875	0.97	0.56	31.34	36.80	30.16	N/A	1.70
2006	3515807	BCR	2,858	0.79	0.40	17.56	79.39	N/A	N/A	3.04
2006	3564165	BCR	2,858	0.78	0.48	34.81	48.64	14.77	N/A	1.78
2006	3515585	BCR	2,940	0.22	0.42	75.31	22.11	N/A	N/A	2.59
2006	3564166	BCR	2,940	0.96	0.58	33.23	32.14	32.11	N/A	2.52
2007	3515886	BCR	30,402	0.45	0.50	51.99	44.94	N/A	N/A	3.07
2007	3564162	BCR	30,402	1.05	0.61	11.50	62.43	21.10	N/A	4.97
2007	3515648	BCR	30,402	0.50	0.50	49.35	49.60	N/A	N/A	1.05
2007	3564163	BCR	30,402	1.11	0.75	21.50	42.22	34.60	N/A	1.68
2007	3515807	BCR	30,402	0.79	0.41	16.33	79.31	N/A	N/A	4.36
2007	3564165	BCR	30,402	0.73	0.62	34.05	54.85	9.32	N/A	1.77
2007	3515585	BCR	30,402	0.19	0.39	78.65	18.63	N/A	N/A	2.72
2007	3564166	BCR	30,402	0.87	0.73	31.17	44.49	21.05	N/A	3.29

Table 1.45 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 4 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	62	3515840	SR	0.2870		
2006	64	3515705	SR	-0.1011		
2006	81	3515886	BCR	1.5380		
2006	82	3564162	BCR	1.3256	-1.6892	1.6892
2006	83	3515909	SR	1.2766		
2006	84	3548085	SR	0.7940		
2006	85	3548086	SR	-0.4484		
2006	86	3515787	SR	0.8057		
2006	89	3515648	BCR	1.3302		
2006	90	3564163	BCR	1.1295	-0.6464	0.6464
2006	100	3548081	SR	0.9106		
2006	101	3515807	BCR	-0.6634		
2006	102	3564165	BCR	1.8360	-1.2560	1.2560
2006	114	3515585	BCR	2.6800		
2006	115	3564166	BCR	1.1258	-0.4681	0.4681
2006	119	3515832	SR	0.4035		
2006	120	3548088	SR	-0.3474		
2006	121	3515853	SR	-0.2637		
2007	62	3515840	SR	0.4436		
2007	64	3515705	SR	-0.2051		
2007	81	3515886	BCR	1.4586		
2007	82	3564162	BCR	0.8111	-1.9929	1.9929
2007	83	3515909	SR	1.2871		
2007	84	3548085	SR	0.9256		
2007	85	3548086	SR	-0.2943		
2007	86	3515787	SR	1.1443		
2007	89	3515648	BCR	1.2409		
2007	90	3564163	BCR	0.8470	-0.9809	0.9809
2007	100	3548081	SR	0.8669		
2007	101	3515807	BCR	-0.7079		
2007	102	3564165	BCR	2.1566	-1.7285	1.7285
2007	114	3515585	BCR	3.2248		
2007	115	3564166	BCR	1.6067	-1.1609	1.1609
2007	119	3515832	SR	0.3339		
2007	120	3548088	SR	-0.1831		
2007	121	3515853	SR	-0.1060		

Note. These Rasch difficulties were based on a common scale.

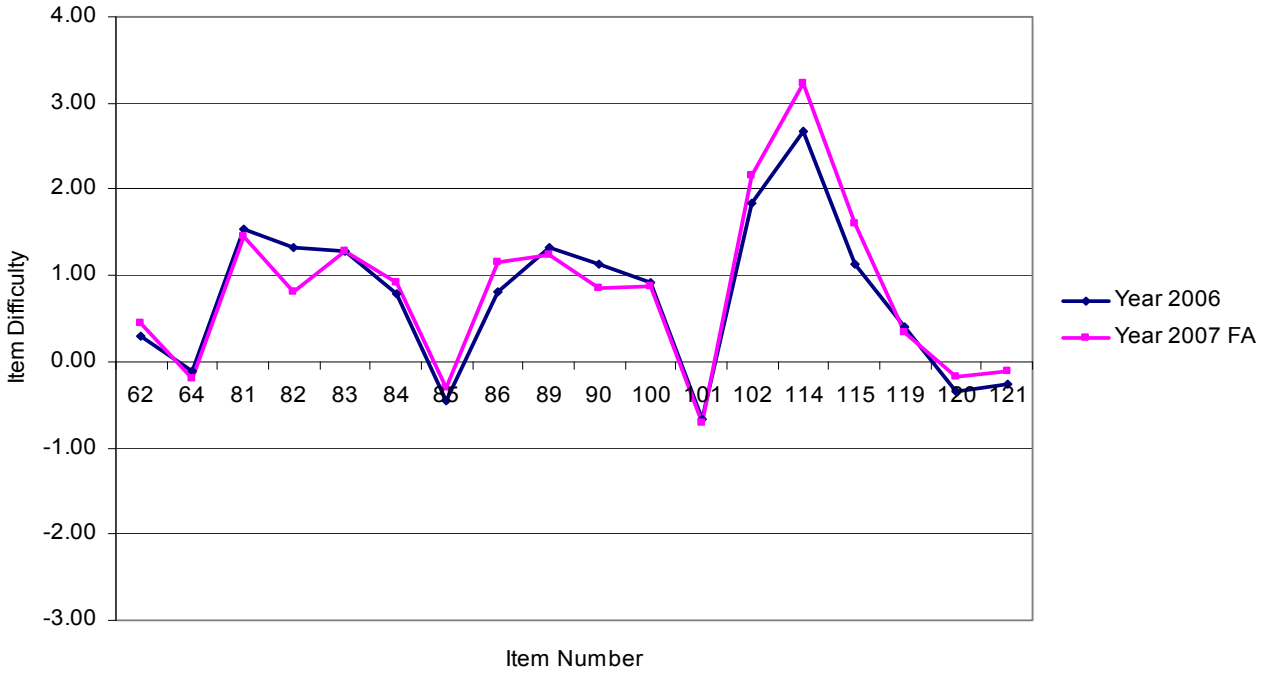
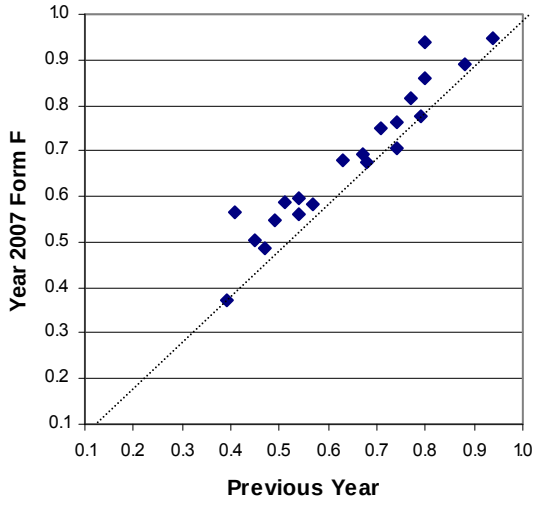


Figure 1.4 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 4 Form A

Table 1.46 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 4 Form F



Item CID	Previous Year	Year 07 Form F
3515652	0.63	0.68
3515936	0.80	0.86
3548079	0.80	0.94
3515791	0.71	0.75
3515795	0.54	0.60
3515869	0.54	0.56
3515836	0.57	0.58
3515648	**0.45**	**0.50**
3564163	**0.49**	**0.55**
3515862	**0.47**	**0.49**
3564170	**0.51**	**0.59**
3548081	0.41	0.57
3515807	**0.79**	**0.78**
3564165	**0.39**	**0.37**
3515830	**0.94**	**0.95**
3564171	**0.74**	**0.71**
3515933	0.74	0.76
3515592	0.77	0.82
3515931	0.68	0.67
3515880	0.67	0.69
3515887	0.88	0.89

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.47 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 4 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3515648	BCR	2,875	0.45	0.50	53.95	44.94	N/A	N/A	1.11
2006	3564163	BCR	2,875	0.97	0.56	31.34	36.80	30.16	N/A	1.70
2006	3515862	BCR	2,799	0.47	0.50	50.63	46.77	N/A	N/A	2.61
2006	3564170	BCR	2,799	1.02	0.58	29.87	33.33	34.16	N/A	2.64
2006	3515807	BCR	2,858	0.79	0.40	17.56	79.39	N/A	N/A	3.04
2006	3564165	BCR	2,858	0.78	0.48	34.81	48.64	14.77	N/A	1.78
2006	3515830	BCR	2,858	0.95	0.23	4.72	94.44	N/A	N/A	0.84
2006	3564171	BCR	2,858	1.49	0.46	7.17	34.81	56.93	N/A	1.08
2007	3515648	BCR	30,103	0.50	0.50	48.42	50.46	N/A	N/A	1.12
2007	3564163	BCR	30,103	1.10	0.76	22.73	40.88	34.54	N/A	1.85
2007	3515862	BCR	30,103	0.49	0.50	49.21	48.76	N/A	N/A	2.02
2007	3564170	BCR	30,103	1.17	0.79	21.52	34.56	41.26	N/A	2.66
2007	3515807	BCR	30,103	0.78	0.42	19.28	77.54	N/A	N/A	3.18
2007	3564165	BCR	30,103	0.75	0.62	33.08	55.81	9.51	N/A	1.60
2007	3515830	BCR	30,103	0.95	0.22	4.40	94.72	N/A	N/A	0.87
2007	3564171	BCR	30,103	1.41	0.59	4.28	47.72	46.68	N/A	1.32

Table 1.48 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 4 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	64	3515652	SR	0.3611		
2006	66	3515936	SR	-0.7328		
2004	78	3548079	SR	-1.2412		
2006	83	3515791	SR	-0.0774		
2006	84	3515795	SR	0.8572		
2006	85	3515869	SR	0.9029		
2006	86	3515836	SR	0.7894		
2006	89	3515648	BCR	1.3302		
2006	90	3564163	BCR	1.1295	-0.6464	0.6464
2006	98	3515862	BCR	1.2450		
2006	99	3564170	BCR	1.0292	-0.5080	0.5080
2004	100	3548081	SR	0.9106		
2006	101	3515807	BCR	-0.6634		
2006	102	3564165	BCR	1.8360	-1.2560	1.2560
2006	114	3515830	BCR	-2.3704		
2006	115	3564171	BCR	-0.3428	-1.0181	1.0181
2006	116	3515933	SR	-0.2666		
2006	118	3515592	SR	-0.5349		
2006	119	3515931	SR	0.1335		
2006	120	3515880	SR	0.1498		
2006	121	3515887	SR	-1.3678		
2007	64	3515652	SR	0.1694		
2007	66	3515936	SR	-1.1494		
2007	78	3548079	SR	-2.3000		
2007	83	3515791	SR	-0.2628		
2007	84	3515795	SR	0.5626		
2007	85	3515869	SR	0.8080		
2007	86	3515836	SR	0.6346		
2007	89	3515648	BCR	1.2409		
2007	90	3564163	BCR	0.8470	-0.9809	0.9809
2007	98	3515862	BCR	1.1734		
2007	99	3564170	BCR	0.6541	-0.6701	0.6701
2007	100	3548081	SR	0.8669		
2007	101	3515807	BCR	-0.7079		
2007	102	3564165	BCR	2.1566	-1.7285	1.7285
2007	114	3515830	BCR	-2.4304		
2007	115	3564171	BCR	-0.5629	-1.5858	1.5858
2007	116	3515933	SR	-0.3619		
2007	118	3515592	SR	-0.8156		
2007	119	3515931	SR	0.1924		
2007	120	3515880	SR	0.0211		
2007	121	3515887	SR	-1.4589		

Note. These Rasch difficulties were based on a common scale.

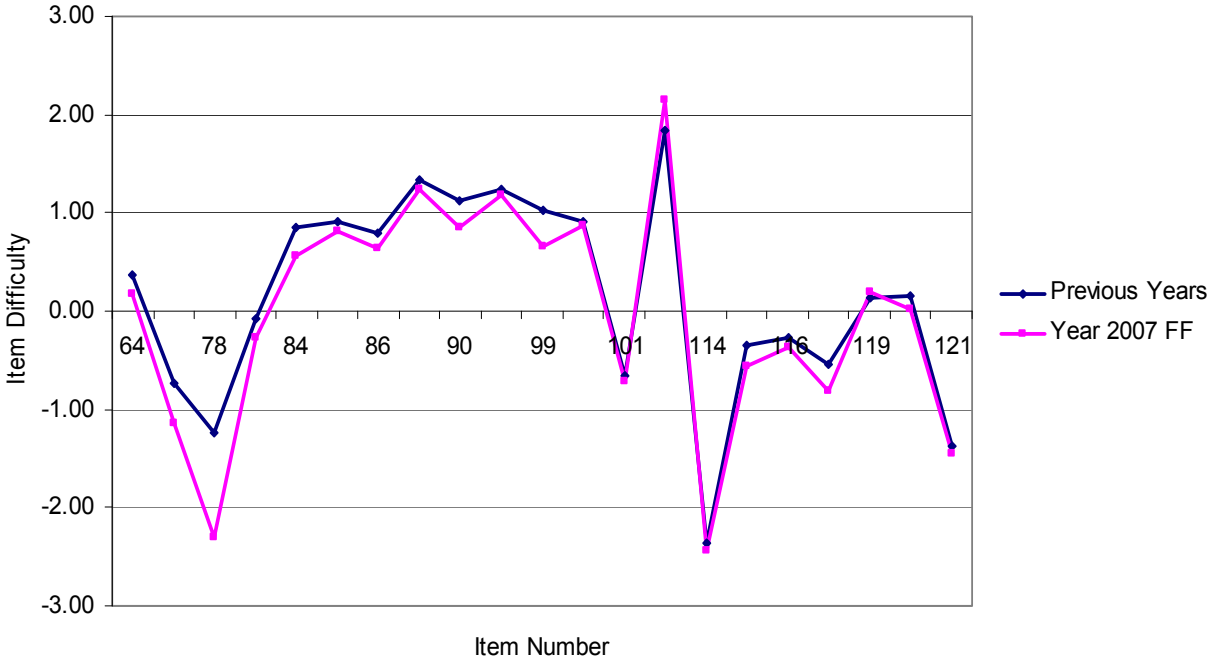
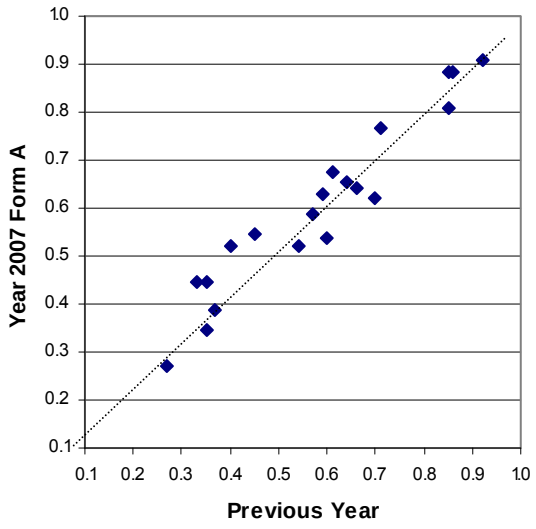


Figure 1.5 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 4 Form F

Table 1.49 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 5 Form A



Item CID	Previous Year	Year 07 Form A
3511531	**0.61**	**0.68**
3563986	**0.45**	**0.55**
3512606	0.59	0.63
3512632	0.37	0.39
3512702	0.60	0.54
3511626	0.85	0.81
3512638	0.66	0.64
3512618	**0.35**	**0.45**
3563988	**0.40**	**0.52**
3512616	0.33	0.44
3512625	0.85	0.88
3512714	0.92	0.91
3512649	**0.27**	**0.27**
3563989	**0.35**	**0.34**
3512627	0.86	0.88
3511563	0.70	0.62
3511566	0.64	0.66
3512710	0.57	0.59
3512687	0.54	0.52
3512628	0.71	0.77

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.50 Item Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 5 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3511531	BCR	2,975	0.61	0.49	37.08	61.21	N/A	N/A	1.71
2006	3563986	BCR	2,975	0.89	0.51	29.68	46.76	21.28	N/A	2.29
2006	3512618	BCR	2,949	0.35	0.48	61.65	35.44	N/A	N/A	2.92
2006	3563988	BCR	2,949	0.80	0.47	30.25	52.42	13.77	N/A	3.56
2006	3512649	BCR	2,873	0.27	0.44	70.66	26.84	N/A	N/A	2.51
2006	3563989	BCR	2,873	0.70	0.63	56.46	11.63	28.96	N/A	2.96
2007	3511531	BCR	31,083	0.68	0.47	31.32	67.50	N/A	N/A	1.18
2007	3563986	BCR	31,083	1.09	0.65	15.05	56.26	26.61	N/A	2.08
2007	3512618	BCR	31,083	0.45	0.50	52.54	44.54	N/A	N/A	2.92
2007	3563988	BCR	31,083	1.05	0.52	7.46	72.60	15.99	N/A	3.95
2007	3512649	BCR	31,083	0.27	0.44	66.35	27.13	N/A	N/A	6.52
2007	3563989	BCR	31,083	0.69	0.89	52.00	10.88	29.04	N/A	8.08

Table 1.51 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 5 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	41	3511531	BCR	0.3047		
2006	42	3563986	BCR	1.2297	-1.1443	1.1443
2006	46	3512606	SR	0.5058		
2006	48	3512632	SR	1.5474		
2006	52	3512702	SR	0.5524		
2006	62	3511626	SR	-1.2487		
2006	65	3512638	SR	0.1748		
2006	68	3512618	BCR	1.7011		
2006	69	3563988	BCR	1.5613	-1.4527	1.4527
2006	77	3512616	SR	1.8493		
2006	78	3512625	SR	-1.2385		
2006	79	3512714	SR	-2.0481		
2006	80	3512649	BCR	2.2683		
2006	81	3563989	BCR	1.6142	0.8261	-0.8261
2006	84	3512627	SR	-1.1851		
2006	96	3511563	SR	-0.1377		
2006	108	3511566	SR	0.1830		
2006	112	3512710	SR	0.7118		
2006	113	3512687	SR	0.8104		
2006	114	3512628	SR	-0.2779		
2007	41	3511531	BCR	0.0868		
2007	42	3563986	BCR	0.6862	-1.6106	1.6106
2007	46	3512606	SR	0.3045		
2007	48	3512632	SR	1.6552		
2007	52	3512702	SR	0.8431		
2007	62	3511626	SR	-0.9260		
2007	65	3512638	SR	0.2606		
2007	68	3512618	BCR	1.2891		
2007	69	3563988	BCR	0.6654	-2.4487	2.4487
2007	77	3512616	SR	1.2809		
2007	78	3512625	SR	-1.6381		
2007	79	3512714	SR	-1.9727		
2007	80	3512649	BCR	2.3175		
2007	81	3563989	BCR	1.6549	0.7655	-0.7655
2007	84	3512627	SR	-1.3981		
2007	96	3511563	SR	0.3760		
2007	108	3511566	SR	0.1548		
2007	112	3512710	SR	0.5954		
2007	113	3512687	SR	0.9241		
2007	114	3512628	SR	-0.5862		

Note. These Rasch difficulties were based on a common scale.

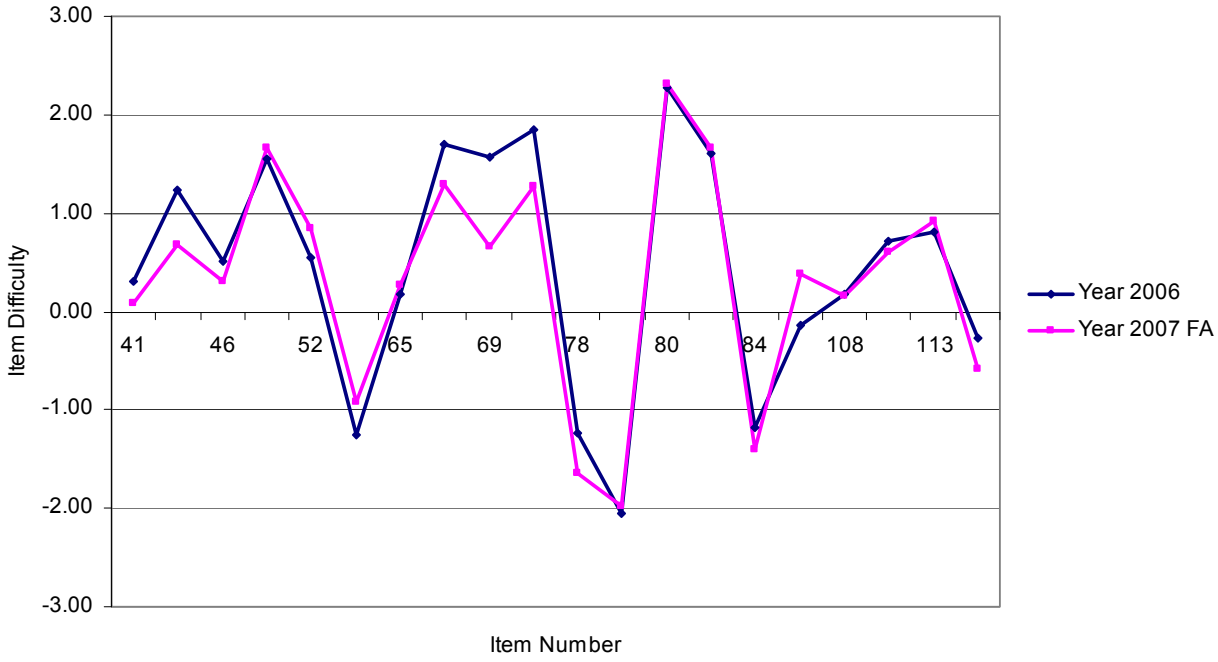
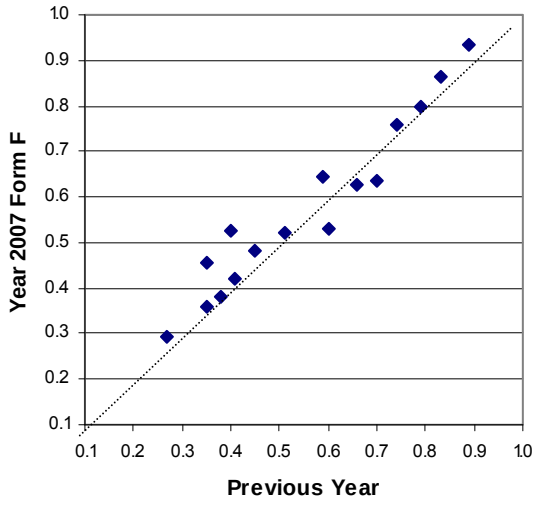


Figure 1.6 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 5 Form A

Table 1.52 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 5 Form F



Item CID	Previous Year	Year 07 Form F
3512606	0.59	0.65
3512702	0.60	0.53
3512638	0.66	0.63
3512618	**0.35**	**0.45**
3563988	**0.40**	**0.53**
3512612	0.38	0.38
3512696	0.83	0.87
3512691	0.51	0.52
3512649	**0.27**	**0.29**
3563989	**0.35**	**0.36**
3512605	0.89	0.94
3511563	0.70	0.64
3512637	0.79	0.80
3512648	0.45	0.48
3512688	0.41	0.42
3511631	0.74	0.76

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.53 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 5 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3512618	BCR	2,949	0.35	0.48	61.65	35.44	N/A	N/A	2.92
2006	3563988	BCR	2,949	0.80	0.47	30.25	52.42	13.77	N/A	3.56
2006	3512649	BCR	2,873	0.27	0.44	70.66	26.84	N/A	N/A	2.51
2006	3563989	BCR	2,873	0.70	0.63	56.46	11.63	28.96	N/A	2.96
2007	3512618	BCR	30,875	0.45	0.50	51.61	45.37	N/A	N/A	3.01
2007	3563988	BCR	30,875	1.05	0.54	7.81	71.05	17.18	N/A	3.96
2007	3512649	BCR	30,875	0.29	0.46	62.10	29.43	N/A	N/A	8.47
2007	3563989	BCR	30,875	0.72	0.91	48.65	9.55	31.26	N/A	10.54

Table 1.54 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 5 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	45	3512606	SR	0.5058		
2006	52	3512702	SR	0.5524		
2006	66	3512638	SR	0.1748		
2006	68	3512618	BCR	1.7011		
2006	69	3563988	BCR	1.5613	-1.4527	1.4527
2006	77	3512612	SR	1.6049		
2006	78	3512696	SR	-1.2411		
2006	79	3512691	SR	0.9245		
2006	80	3512649	BCR	2.2683		
2006	81	3563989	BCR	1.6142	0.8261	-0.8261
2006	84	3512605	SR	-1.6432		
2006	96	3511563	SR	-0.1377		
2006	108	3512637	SR	-0.6144		
2006	112	3512648	SR	1.3551		
2006	113	3512688	SR	1.4904		
2006	114	3511631	SR	-0.3956		
2007	45	3512606	SR	0.3045		
2007	52	3512702	SR	0.8431		
2007	66	3512638	SR	0.2606		
2007	68	3512618	BCR	1.2891		
2007	69	3563988	BCR	0.6654	-2.4487	2.4487
2007	77	3512612	SR	1.6747		
2007	78	3512696	SR	-1.4768		
2007	79	3512691	SR	1.0014		
2007	80	3512649	BCR	2.3175		
2007	81	3563989	BCR	1.6549	0.7655	-0.7655
2007	84	3512605	SR	-2.0760		
2007	96	3511563	SR	0.3760		
2007	108	3512637	SR	-0.6947		
2007	112	3512648	SR	1.2138		
2007	113	3512688	SR	1.4200		
2007	114	3511631	SR	-0.3862		

Note. These Rasch difficulties were based on a common scale.

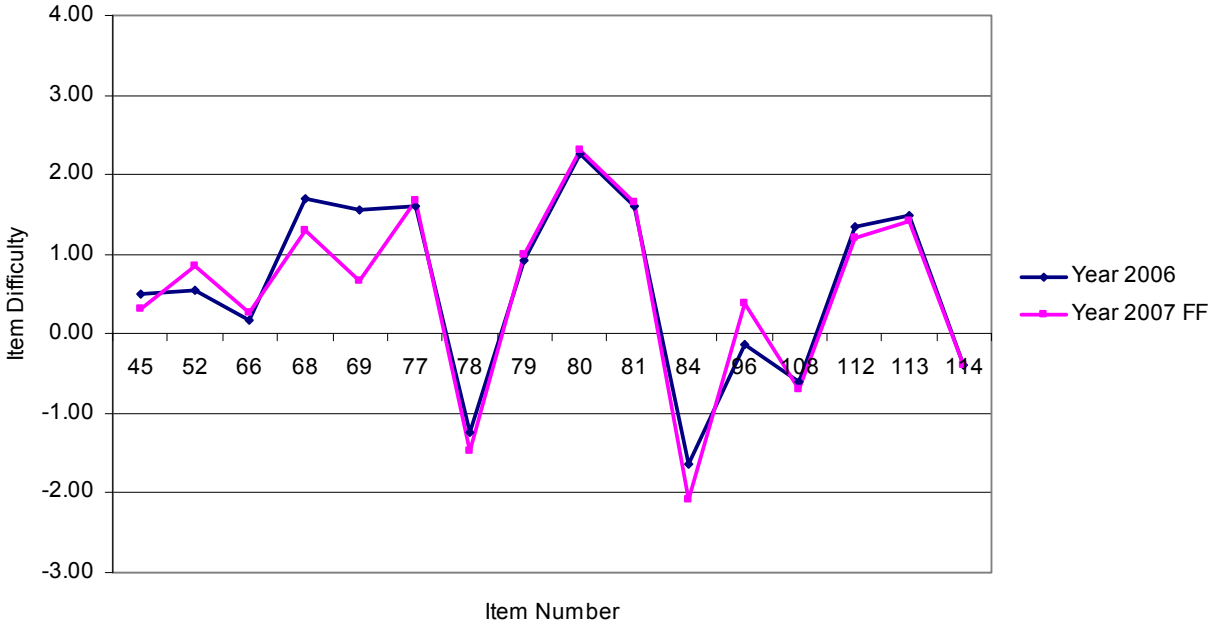
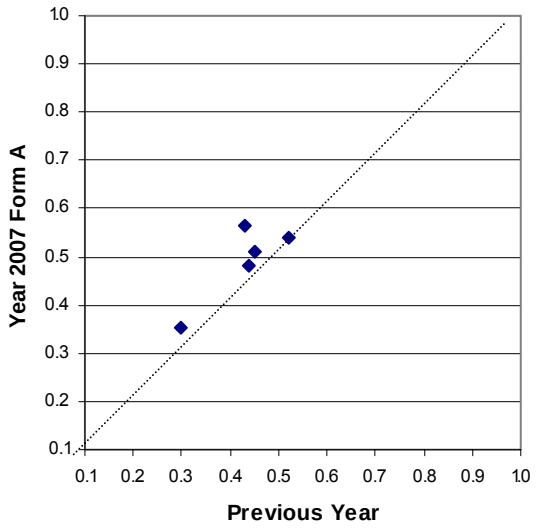


Figure 1.7 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 5 Form F

Table 1.55 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 6 Form A



Item CID	Previous Year	Year 07 Form A
3516912	0.52	0.54
3517013	**0.30**	**0.35**
3564004	**0.43**	**0.57**
3517000	0.45	0.51
3517010	0.44	0.48

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item.

Table 1.56 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 6 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3517013	BCR	3,219	0.30	0.50	65.49	30.20	N/A	N/A	4.32
2006	3564004	BCR	3,219	0.87	0.50	27.71	48.68	18.95	N/A	4.66
2007	3517013	BCR	31,558	0.35	0.48	61.55	35.38	N/A	N/A	3.07
2007	3564004	BCR	31,558	1.13	0.63	10.29	59.17	27.04	N/A	3.50

Table 1.57 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 6 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2006	45	3516912	SR	0.3519		
2006	74	3517013	BCR	1.5654		
2006	75	3564004	BCR	0.8553	-1.2512	1.2512
2006	84	3517000	SR	0.7559		
2006	85	3517010	SR	0.8241		
2007	45	3516912	SR	0.5174		
2007	74	3517013	BCR	1.4674		
2007	75	3564004	BCR	0.0865	-1.7954	1.7954
2007	84	3517000	SR	0.6588		
2007	85	3517010	SR	0.8496		

Note. These Rasch difficulties were based on a common scale.

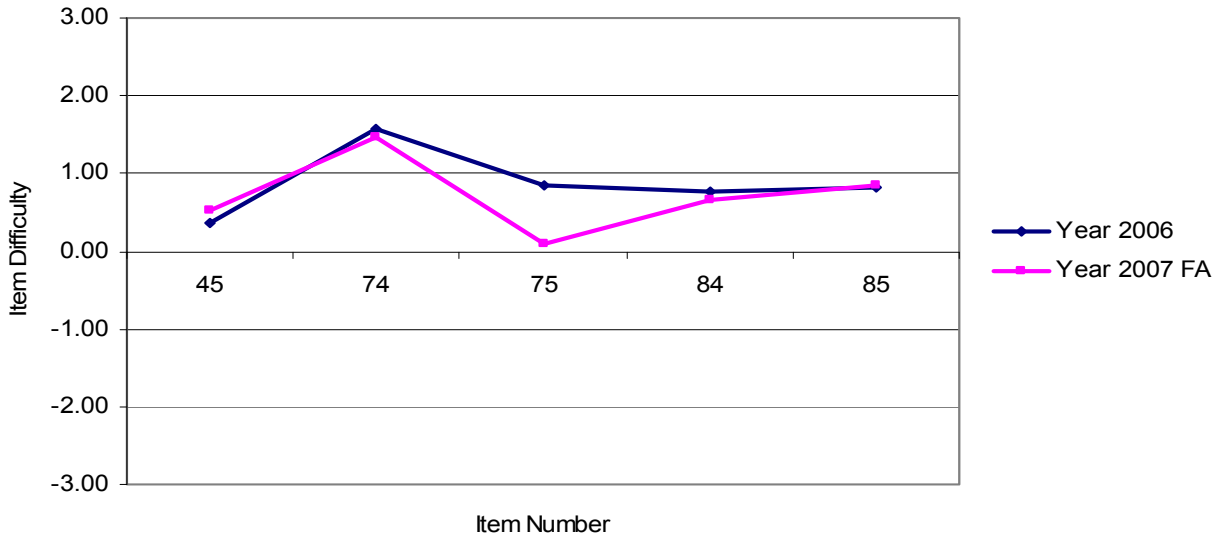
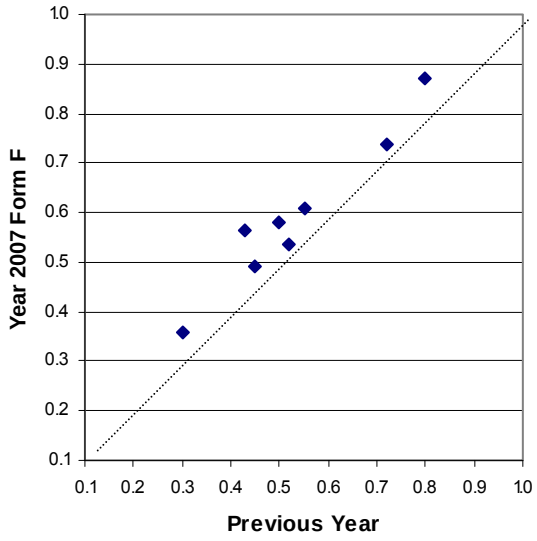


Figure 1.8 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 6 Form A

Table 1.58 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 6 Form F



Item CID	Previous Year	Year 07 Form F
3516912	0.52	0.54
3517004	**0.80**	**0.87**
3564010	**0.50**	**0.58**
3517002	0.72	0.74
3517013	**0.30**	**0.36**
3564004	**0.43**	**0.56**
3517000	0.45	0.49
3516907	0.55	0.61

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item or Extended Constructed Response (ECR) item.

Table 1.59 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 6 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3517004	ECR	3,219	0.80	0.40	16.78	79.96	N/A	N/A	3.26
2006	3564010	ECR	3,219	1.51	0.59	15.04	32.34	28.02	20.78	3.82
2006	3517013	BCR	3,219	0.30	0.46	65.49	30.20	N/A	N/A	4.32
2006	3564004	BCR	3,219	0.87	0.50	27.71	48.68	18.95	N/A	4.66
2007	3517004	ECR	31,258	0.87	0.34	11.36	87.09	N/A	N/A	1.55
2007	3564010	ECR	31,258	1.74	0.96	8.71	29.21	34.34	25.26	2.49
2007	3517013	BCR	31,258	0.36	0.48	61.55	35.68	N/A	N/A	2.78
2007	3564004	BCR	31,258	1.13	0.63	11.05	58.70	27.09	N/A	3.16

Table 1.60 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 6 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2006	44	3516912	SR	0.3519			
2006	49	3517004	ECR	-1.3826			
2006	50	3564010	ECR	0.4105	-1.4436	0.2291	1.2145
2006	57	3517002	SR	-0.7658			
2006	74	3517013	BCR	1.5654			
2006	75	3564004	BCR	0.8553	-1.2512	1.2512	
2006	84	3517000	SR	0.7559			
2006	85	3516907	SR	0.2435			
2007	44	3516912	SR	0.5174			
2007	49	3517004	ECR	-1.7238			
2007	50	3564010	ECR	0.2493	-1.6097	0.1701	1.4396
2007	57	3517002	SR	-0.7729			
2007	74	3517013	BCR	1.4674			
2007	75	3564004	BCR	0.0865	-1.7954	1.7954	
2007	84	3517000	SR	0.6588			
2007	85	3516907	SR	0.1598			

Note. These Rasch difficulties were based on a common scale.

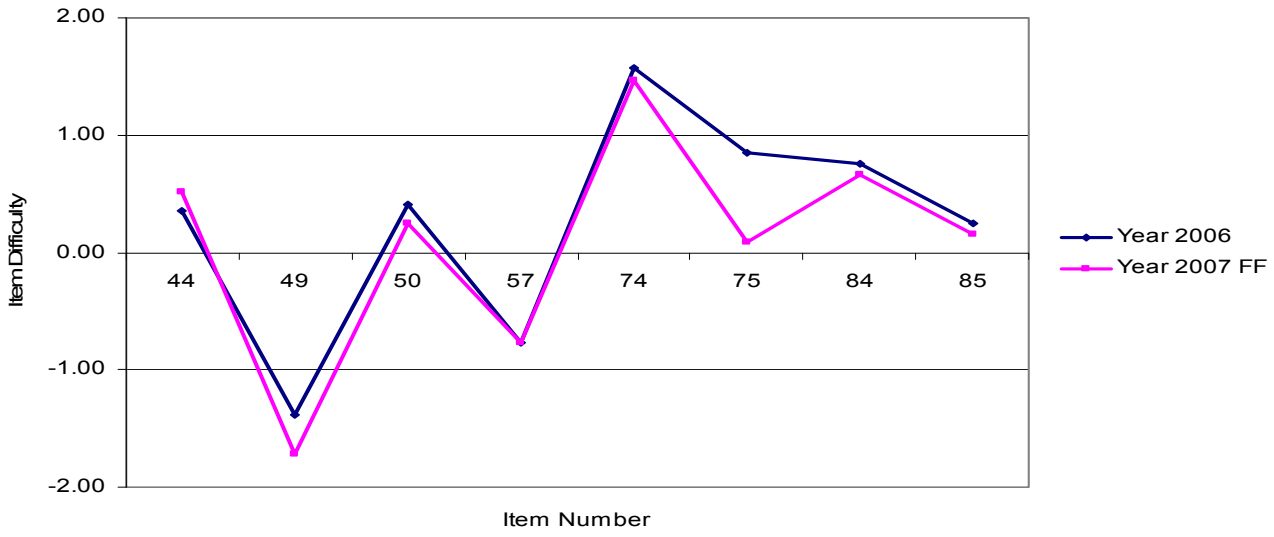
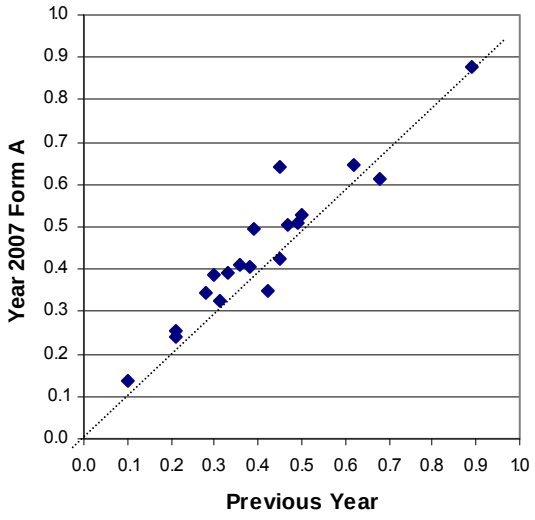


Figure 1.9 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 6 Form F

Table 1.61 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 7 Form A



Item CID	Previous Year	Year 07 Form A
3517744	**0.42**	**0.35**
3564018	**0.21**	**0.24**
3517678	0.89	0.88
3517710	0.68	0.61
3517742	0.47	0.50
3517673	**0.62**	**0.65**
3564020	**0.38**	**0.40**
3517757	0.28	0.35
3517759	0.45	0.43
3517719	**0.21**	**0.26**
3564021	**0.36**	**0.41**
3555858	0.33	0.39
3547477	0.39	0.49
3517736	0.49	0.51
3517818	**0.31**	**0.33**
3564023	**0.30**	**0.38**
3517876	0.10	0.14
3517779	0.45	0.64
3517733	0.50	0.53

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item, Extended Constructed Response (ECR) item or Student Produced Response (SPR) item.

Table 1.62 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 7 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3517744	BCR	3,398	0.42	0.49	50.21	41.85	N/A	N/A	7.95
2006	3564018	BCR	3,398	0.43	0.43	55.18	29.49	6.59	N/A	8.74
2006	3517673	ECR	3,438	0.62	0.49	34.58	61.93	N/A	N/A	3.49
2006	3564020	ECR	3,438	1.14	0.31	3.66	73.04	18.96	1.05	3.29
2006	3517719	BCR	3,422	0.21	0.41	65.23	21.04	N/A	N/A	13.73
2006	3564021	BCR	3,422	0.71	0.39	19.73	61.08	5.08	N/A	14.11
2006	3517818	BCR	3,382	0.31	0.46	56.15	30.90	N/A	N/A	12.95
2006	3564023	BCR	3,382	0.61	0.37	26.91	57.27	1.71	N/A	14.10
2007	3517744	BCR	32,264	0.35	0.48	57.42	35.11	N/A	N/A	7.48
2007	3564018	BCR	32,264	0.48	0.66	52.89	29.28	9.33	N/A	8.51
2007	3517673	ECR	32,264	0.65	0.48	31.06	64.74	N/A	N/A	4.20
2007	3564020	ECR	32,264	1.21	0.60	3.39	66.72	23.79	2.30	3.80
2007	3517719	BCR	32,264	0.26	0.44	62.91	25.69	N/A	N/A	11.40
2007	3564021	BCR	32,264	0.82	0.55	13.67	67.11	7.58	N/A	11.64
2007	3517818	BCR	32,264	0.33	0.47	61.95	32.55	N/A	N/A	5.50
2007	3564023	BCR	32,264	0.77	0.58	23.33	61.20	7.86	N/A	7.61

Table 1.63 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 7 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2006	41	3517744	BCR	0.5220			
2006	42	3564018	BCR	1.9858	-1.0098	1.0098	
2006	69	3517678	SR	-2.5560			
2006	70	3517710	SR	-0.7950			
2006	71	3517742	SR	0.3521			
2006	78	3517673	ECR	-0.5260			
2006	79	3564020	ECR	1.0845	-4.6256	0.9338	3.6918
2006	80	3517757	SPR	1.3453			
2006	82	3517759	SPR	0.3521			
2006	83	3517719	BCR	1.7788			
2006	84	3564021	BCR	1.1889	-2.4446	2.4446	
2006	92	3555858	SR	0.7163			
2006	93	3547477	SR	0.3369			
2006	103	3517736	SR	0.2112			
2006	104	3517818	BCR	1.1113			
2006	105	3564023	BCR	2.0754	-2.7931	2.7931	
2006	107	3517876	SPR	2.8747			
2006	110	3517779	SPR	0.3583			
2006	112	3517733	SPR	-0.0557			
2007	41	3517744	BCR	0.9733			
2007	42	3564018	BCR	1.8283	-0.8810	0.8810	
2007	69	3517678	SR	-2.6820			
2007	70	3517710	SR	-0.6119			
2007	71	3517742	SR	0.0227			
2007	78	3517673	ECR	-0.8144			
2007	79	3564020	ECR	0.8436	-4.4403	0.7733	3.6670
2007	80	3517757	SPR	1.0880			
2007	82	3517759	SPR	0.5501			
2007	83	3517719	BCR	1.5471			
2007	84	3564021	BCR	0.7817	-2.6186	2.6186	
2007	92	3555858	SR	0.6673			
2007	93	3547477	SR	0.0680			
2007	103	3517736	SR	0.0790			
2007	104	3517818	BCR	1.1456			
2007	105	3564023	BCR	1.1416	-2.2500	2.2500	
2007	107	3517876	SPR	2.7529			
2007	110	3517779	SPR	-0.7021			
2007	112	3517733	SPR	-0.0370			

Note. These Rasch difficulties were based on a common scale.

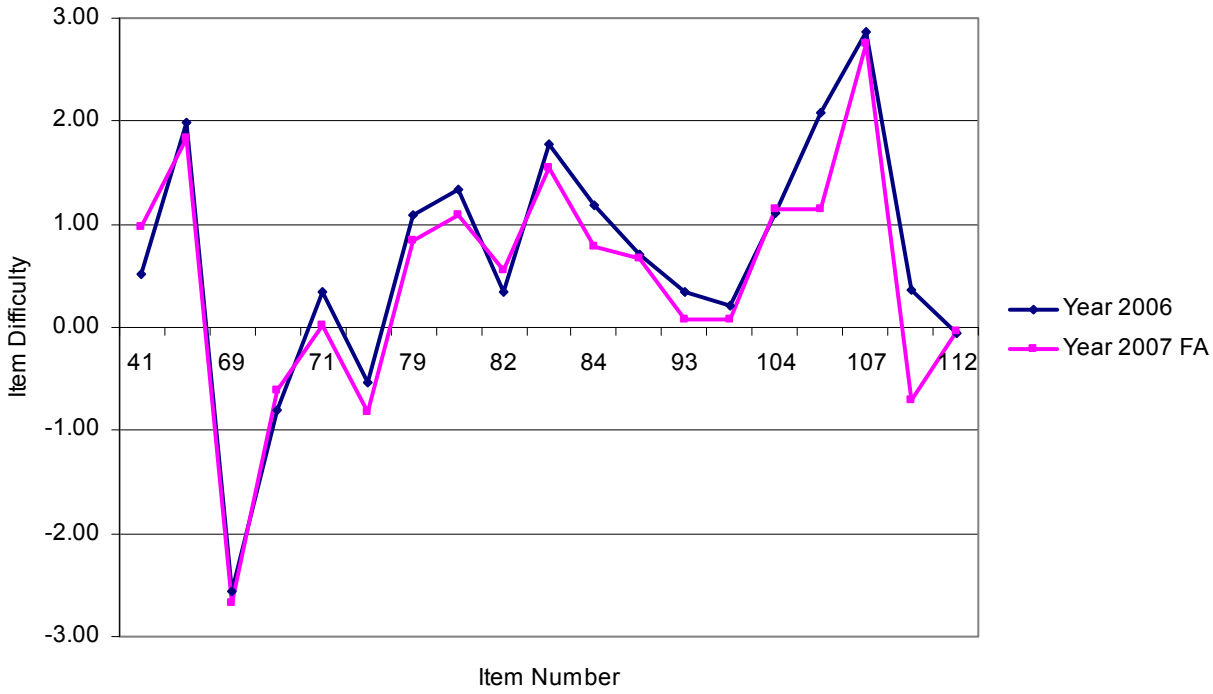
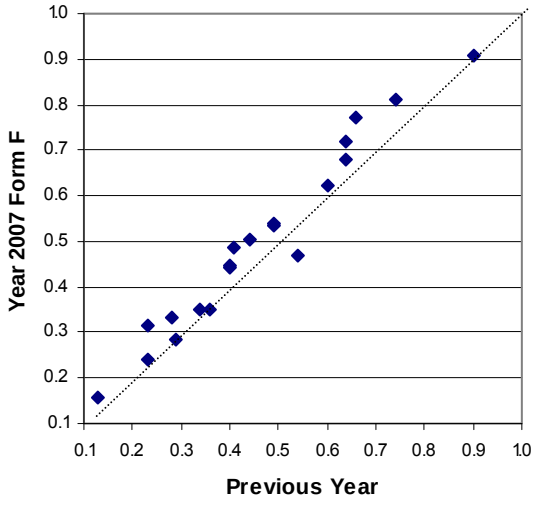


Figure 1.10 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 7 Form A

Table 1.64 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 7 Form F



Item CID	Previous Year	Year 07 Form F
3517706	**0.54**	**0.47**
3564025	**0.29**	**0.28**
3555861	0.64	0.72
3517679	0.41	0.49
3517740	0.49	0.53
3517741	0.90	0.91
3517812	0.49	0.54
3517695	0.36	0.35
3517729	0.64	0.68
3517757	0.28	0.33
3517693	**0.13**	**0.16**
3564028	**0.40**	**0.45**
3517752	0.60	0.62
3517885	0.34	0.35
3517715	**0.74**	**0.81**
3564030	**0.44**	**0.50**
3517758	0.23	0.24
3547487	**0.66**	**0.77**
3564031	**0.23**	**0.31**
3517756	0.40	0.44

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item, Extended Constructed Response (ECR) item or Student-Produced Response (SPR) item.

Table 1.65 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 7 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3517706	BCR	3,422	0.54	0.50	36.09	54.47	N/A	N/A	9.44
2006	3564025	BCR	3,422	0.58	0.49	44.24	34.86	11.43	N/A	9.47
2006	3517693	BCR	3,438	0.13	0.34	76.61	12.91	N/A	N/A	10.47
2006	3564028	BCR	3,438	0.80	0.49	24.20	48.78	15.47	N/A	11.55
2006	3517715	BCR	3,422	0.74	0.44	19.46	73.73	N/A	N/A	6.81
2006	3564030	BCR	3,422	0.88	0.36	7.69	11.81	72.82	N/A	7.69
2006	3547487	ECR	13,123	0.66	0.47	26.34	65.77	N/A	N/A	7.89
2006	3564031	ECR	13,123	0.70	0.33	26.59	59.00	5.49	0.14	8.78
2007	3517706	BCR	32,000	0.47	0.50	45.48	46.67	N/A	N/A	7.85
2007	3564025	BCR	32,000	0.57	0.71	48.26	30.38	13.08	N/A	8.28
2007	3517693	BCR	32,000	0.16	0.36	76.92	15.57	N/A	N/A	7.51
2007	3564028	BCR	32,000	0.90	0.70	21.54	49.89	19.95	N/A	8.62
2007	3517715	BCR	32,000	0.81	0.39	17.02	80.97	N/A	N/A	2.01
2007	3564030	BCR	32,000	1.01	0.46	6.94	78.89	10.88	N/A	3.29
2007	3547487	ECR	32,000	0.77	0.42	19.42	77.24	N/A	N/A	3.34
2007	3564031	ECR	32,000	0.94	0.62	16.25	62.26	15.38	0.48	5.63

Table 1.66 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 7 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2006	41	3517706	BCR	-0.2503			
2006	42	3564025	BCR	1.3767	-0.9842	0.9842	
2006	44	3555861	SR	-0.9715			
2006	48	3517679	SR	0.6718			
2006	51	3517740	SR	0.1924			
2006	69	3517741	SR	-2.7212			
2006	70	3517812	SR	0.2492			
2006	80	3517695	SPR	0.8441			
2006	81	3517729	SPR	-0.6657			
2006	82	3517757	SPR	1.3453			
2006	83	3517693	BCR	2.6009			
2006	84	3564028	BCR	0.6356	-1.4998	1.4998	
2006	92	3517752	SR	-0.3369			
2006	93	3517885	SR	1.0752			
2006	104	3517715	BCR	-1.4639			
2006	105	3564030	BCR	0.5617	-2.6859	2.6859	
2006	107	3517758	SPR	1.8098			
2006	108	3547487	ECR	-1.0830			
2006	109	3564031	ECR	2.8049	-3.8213	0.685	3.1363
2006	112	3517756	SPR	0.6904			
2007	41	3517706	BCR	0.2566			
2007	42	3564025	BCR	1.4706	-0.8554	0.8554	
2007	44	3555861	SR	-1.1968			
2007	48	3517679	SR	0.2781			
2007	51	3517740	SR	-0.0121			
2007	69	3517741	SR	-3.1184			
2007	70	3517812	SR	-0.1150			
2007	80	3517695	SPR	0.9921			
2007	81	3517729	SPR	-1.0032			
2007	82	3517757	SPR	1.0880			
2007	83	3517693	BCR	2.4839			
2007	84	3564028	BCR	0.4116	-1.5038	1.5038	
2007	92	3517752	SR	-0.5723			
2007	93	3517885	SR	1.0861			
2007	104	3517715	BCR	-1.8006			
2007	105	3564030	BCR	0.1441	-2.9105	2.9105	
2007	107	3517758	SPR	1.7915			
2007	108	3547487	ECR	-1.5658			
2007	109	3564031	ECR	2.1233	-3.8495	0.2592	3.5903
2007	112	3517756	SPR	0.4675			

Note. These Rasch difficulties were based on a common scale.

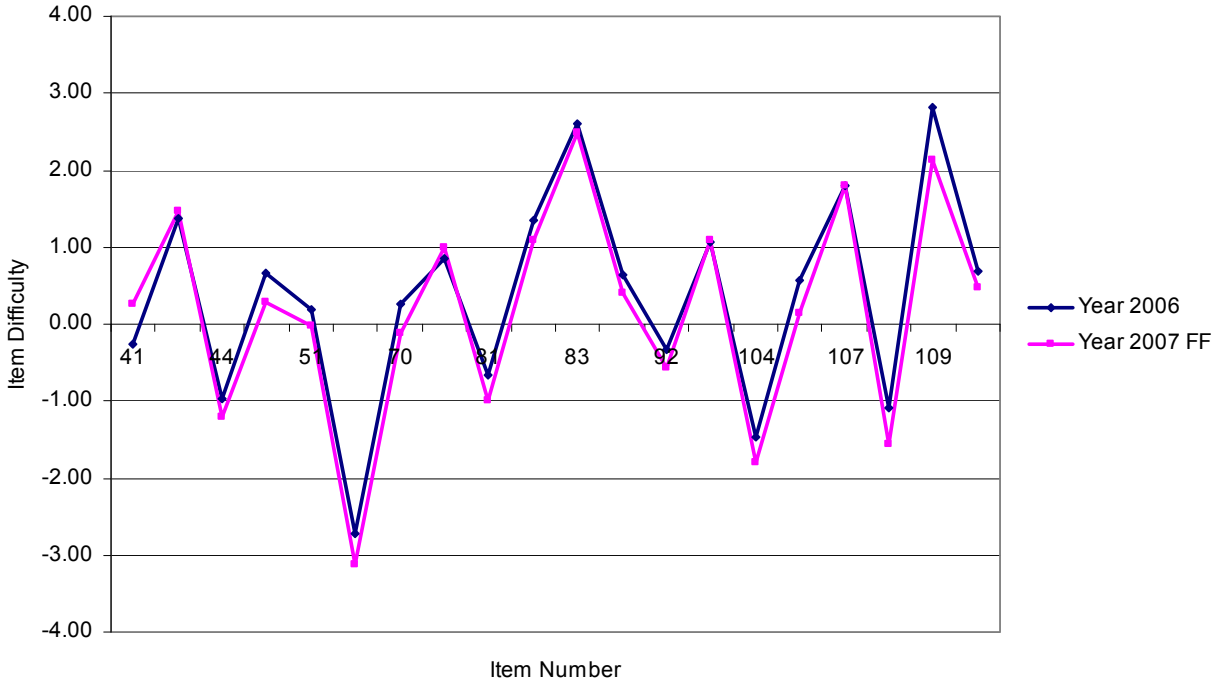
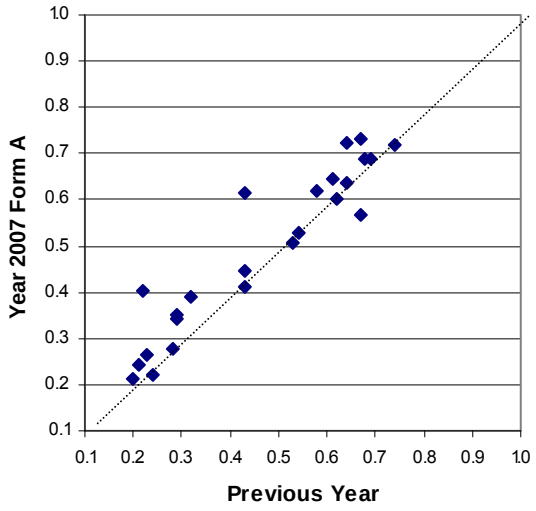


Figure 1.11 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 7 Form F

Table 1.67 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 8 Form A



Item CID	Previous Year	Year 07 Form A
3547550	0.67	0.57
3514702	**0.28**	**0.28**
3564108	**0.29**	**0.34**
3514276	**0.43**	**0.45**
3514127	**0.24**	**0.22**
3514125	**0.62**	**0.60**
3514121	0.68	0.69
3514139	0.64	0.73
3514611	**0.61**	**0.65**
3514608	0.43	0.41
3514287	0.58	0.62
3514267	**0.29**	**0.35**
3564110	**0.43**	**0.62**
3514275	**0.74**	**0.72**
3514279	**0.20**	**0.21**
3514131	0.32	0.39
3514607	**0.23**	**0.26**
3564112	**0.21**	**0.24**
3514118	**0.08**	**0.09**
3564113	**0.22**	**0.40**
3514291	0.67	0.73
3514606	0.69	0.69
3514669	**0.53**	**0.51**
3564114	**0.64**	**0.63**
3514710	0.54	0.53

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item, Extended Constructed Response (ECR) item or Student-Produced Response (SPR) item.

Table 1.68 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 8 Form A

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3514702	ECR	3,524	0.28	0.45	60.36	28.04	N/A	N/A	11.61
2006	3564108	ECR	3,524	0.86	0.63	39.87	22.50	9.85	14.76	13.02
2006	3514267	BCR	3,520	0.29	0.46	55.99	29.49	N/A	N/A	14.52
2006	3564110	BCR	3,520	0.86	0.56	24.23	34.57	25.88	N/A	15.31
2006	3514607	ECR	3,478	0.23	0.42	60.93	22.97	N/A	N/A	16.10
2006	3564112	ECR	3,478	0.62	0.57	49.80	12.56	12.77	7.88	16.99
2006	3514118	BCR	3,573	0.08	0.27	83.18	7.70	N/A	N/A	9.12
2006	3564113	BCR	3,573	0.45	0.38	47.52	41.06	1.79	N/A	9.63
2006	3514669	BCR	3,478	0.53	0.50	42.38	52.65	N/A	N/A	4.97
2006	3564114	BCR	3,478	1.27	0.58	17.94	24.93	51.24	N/A	5.89
2007	3514702	ECR	32,836	0.28	0.45	65.83	27.65	N/A	N/A	6.51
2007	3564108	ECR	32,836	1.03	1.13	34.62	26.39	11.53	17.81	9.66
2007	3514267	BCR	32,836	0.35	0.48	61.06	34.98	N/A	N/A	3.96
2007	3564110	BCR	32,836	1.23	0.67	8.73	49.35	36.87	N/A	5.05
2007	3514607	ECR	32,836	0.26	0.44	64.57	26.32	N/A	N/A	9.12
2007	3564112	ECR	32,836	0.73	1.05	49.17	12.90	15.01	10.02	12.89
2007	3514118	BCR	32,836	0.09	0.29	86.62	9.37	N/A	N/A	4.00
2007	3564113	BCR	32,836	0.80	0.49	18.42	72.25	3.99	N/A	5.34
2007	3514669	BCR	32,836	0.51	0.50	41.96	50.51	N/A	N/A	7.53
2007	3564114	BCR	32,836	1.27	0.81	14.37	26.39	50.27	N/A	8.97

Table 1.69 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 8 Form A

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2006	47	3547550	SR	-1.0816			
2006	58	3514702	ECR	1.0288			
2006	59	3564108	ECR	0.7810	-0.5187	0.4650	0.0537
2006	61	3514276	SPR	0.2970			
2006	62	3514127	SPR	1.3774			
2006	63	3514125	SPR	-0.8524			
2006	64	3514121	SR	-1.0964			
2006	65	3514139	SR	-0.8722			
2006	73	3514611	SPR	-0.6597			
2006	83	3514608	SR	0.3077			
2006	84	3514287	SR	-0.4312			
2006	85	3514267	BCR	0.9172			
2006	86	3564110	BCR	0.0297	-0.8109	0.8109	
2006	88	3514275	SPR	-1.3908			
2006	94	3514279	SPR	1.8130			
2006	96	3514131	SR	0.8165			
2006	98	3514607	ECR	1.3483			
2006	99	3564112	ECR	1.3596	-0.1059	-0.6317	0.7376
2006	102	3514118	BCR	2.9738			
2006	103	3564113	BCR	2.2349	-2.1441	2.1441	
2006	104	3514291	SR	-0.9785			
2006	105	3514606	SR	-1.0884			
2006	113	3514669	BCR	-0.2451			
2006	114	3564114	BCR	-0.8169	-0.2221	0.2221	
2006	116	3514710	SR	-0.2307			
2007	47	3547550	SR	-0.4694			
2007	58	3514702	ECR	1.2761			
2007	59	3564108	ECR	0.6901	-0.7491	0.5272	0.2219
2007	61	3514276	SPR	0.2809			
2007	62	3514127	SPR	1.6861			
2007	63	3514125	SPR	-0.6461			
2007	64	3514121	SR	-1.0563			
2007	65	3514139	SR	-1.3743			
2007	73	3514611	SPR	-0.9362			
2007	83	3514608	SR	0.3422			
2007	84	3514287	SR	-0.8039			
2007	85	3514267	BCR	0.8169			
2007	86	3564110	BCR	-0.9309	-1.4936	1.4936	
2007	88	3514275	SPR	-1.3750			
2007	94	3514279	SPR	1.6979			
2007	96	3514131	SR	0.5831			
2007	98	3514607	ECR	1.2953			
2007	99	3564112	ECR	1.2629	0.1082	-0.8532	0.7450
2007	102	3514118	BCR	2.8471			
2007	103	3564113	BCR	1.0451	-2.7281	2.7281	
2007	104	3514291	SR	-1.4001			
2007	105	3514606	SR	-1.1060			
2007	113	3514669	BCR	-0.1522			
2007	114	3564114	BCR	-0.8897	-0.4608	0.4608	
2007	116	3514710	SR	-0.1424			

Note. These Rasch difficulties were based on a common scale.

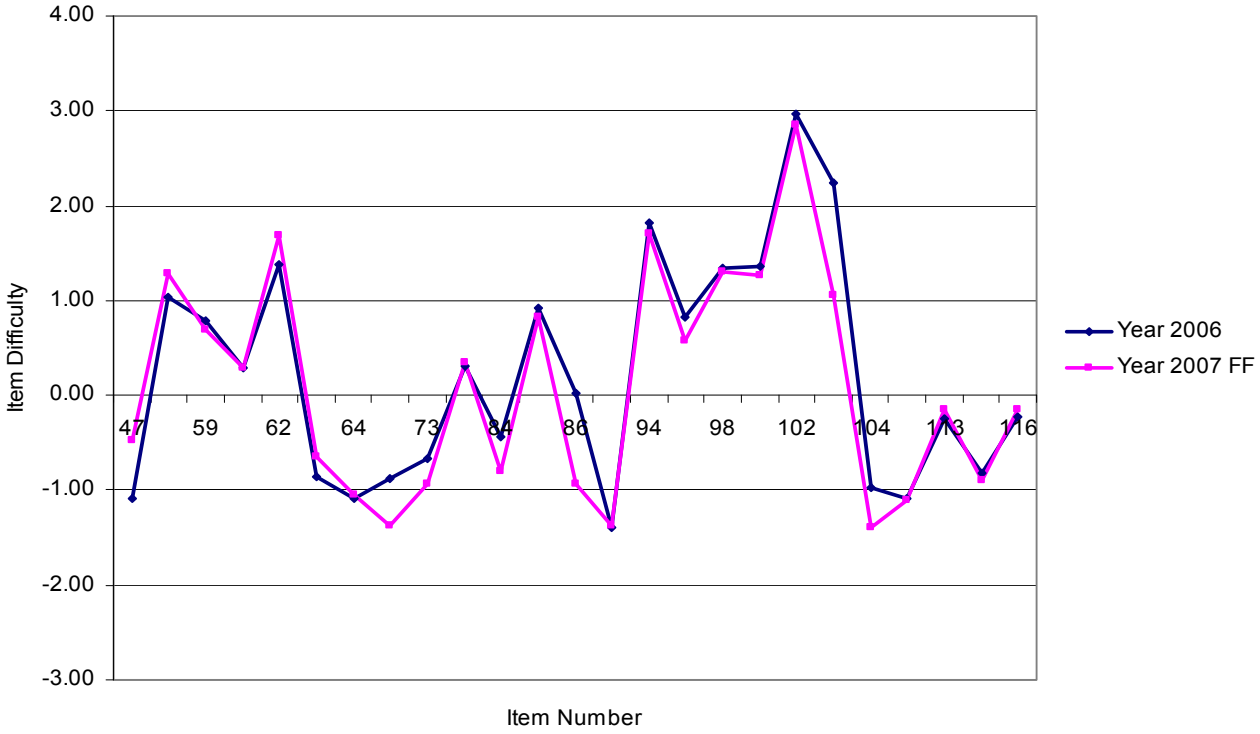
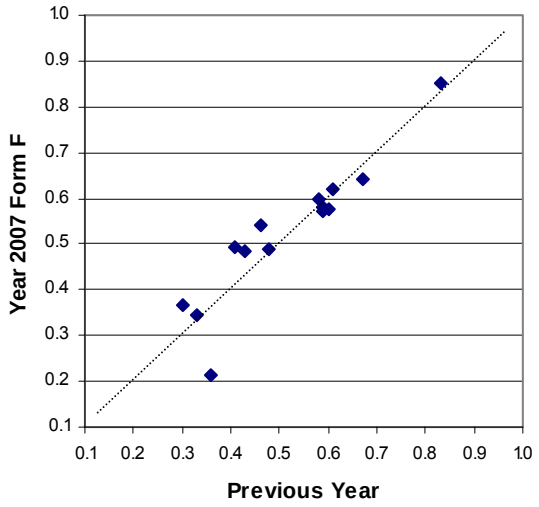


Figure 1.12 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 8 Form A

Table 1.70 Augmented Item P-Value Comparison for Previous Year vs. Year 2007: Grade 8 Form F



Item CID	Previous Year	Year 07 Form F
3514283	**0.33**	**0.34**
3564116	**0.43**	**0.49**
3514122	0.58	0.60
3514281	**0.36**	**0.21**
3514138	0.60	0.58
3514708	**0.59**	**0.57**
3547550	0.67	0.64
3547551	0.83	0.85
3547547	0.46	0.54
3514288	0.59	0.58
3514611	**0.61**	**0.62**
3514133	**0.30**	**0.37**
3564121	**0.41**	**0.49**
3547536	0.48	0.49

*Bold-faced number indicates that it is Brief Constructed Response (BCR) item, Extended Constructed Response (ECR) item or Student-Produced Response (SPR) item.

Table 1.71 Score-Point Distribution Comparison for Previous Year vs. Year 2007: Grade 8 Form F

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2006	3514283	ECR	3,525	0.33	0.47	55.77	33.42	N/A	N/A	10.81
2006	3564116	ECR	3,525	1.30	0.61	13.73	39.04	16.03	19.63	11.57
2006	3514133	BCR	3,573	0.30	0.46	54.46	29.95	N/A	N/A	15.59
2006	3564121	BCR	3,573	0.83	0.51	19.70	45.51	18.50	N/A	16.29
2007	3514283	ECR	32,480	0.34	0.48	59.90	34.44	N/A	N/A	5.66
2007	3564116	ECR	32,480	1.46	1.00	7.40	46.89	15.97	22.24	7.52
2007	3514133	BCR	32,480	0.37	0.48	54.78	36.53	N/A	N/A	8.70
2007	3564121	BCR	32,480	0.99	0.65	11.68	57.95	20.33	N/A	10.03

Table 1.72 Augment IRT Item Difficulty Comparison for Previous Year vs. Year 2007: Grade 8 Form F

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2006	58	3514283	ECR	0.6798			
2006	59	3564116	ECR	-0.0487	-1.6928	0.9468	0.746
2006	64	3514122	SR	-0.5723			
2006	74	3514281	SPR	0.7098			
2006	81	3514138	SR	-0.6746			
2006	94	3514708	SPR	-0.5929			
2006	100	3547550	SR	-1.0816			
2006	101	3547551	SR	-2.1845			
2006	104	3547547	SR	-0.0354			
2006	105	3514288	SR	-0.4589			
2006	112	3514611	SPR	-0.6597			
2006	113	3514133	BCR	0.7426			
2006	114	3564121	BCR	0.0346	-1.3926	1.3926	
2006	116	3547536	SR	-0.1400			
2007	58	3514283	ECR	0.8146			
2007	59	3564116	ECR	-0.2444	-2.2962	1.2817	1.0145
2007	64	3514122	SR	-0.5576			
2007	74	3514281	SPR	1.7603			
2007	81	3514138	SR	-0.4768			
2007	94	3514708	SPR	-0.3665			
2007	100	3547550	SR	-0.4694			
2007	101	3547551	SR	-2.2204			
2007	104	3547547	SR	-0.2056			
2007	105	3514288	SR	-0.4203			
2007	112	3514611	SPR	-0.9362			
2007	113	3514133	BCR	0.6858			
2007	114	3564121	BCR	-0.1650	-1.8706	-1.8706	
2007	116	3547536	SR	-0.0487			

Note. These Rasch difficulties were based on a common scale.

Figure 1.13 Augmented IRT Item Difficulty Comparison Plot for Previous Year vs. Year 2007: Grade 8 Form F

1.9 Field Test Analyses

All field test items embedded in operational forms were subjected to rigorous analyses for their properties because these analyses will provide information about which items would be included as operational items in the future. All statistical results concerning field test items were reserved in the 2007 item bank. Information on item bank can be found in section 1.14, Item Bank Construction. The following field test analyses were conducted:

- Classical item analyses for *SR*, *SPR*, *BCR*, and *ECR* items
- *Differential item functioning* (*DIF*) analyses
- *IRT* analyses

Classical Item Analyses for *SR, SPR, BCR, and ECR* items

Classical item analyses for *SR*, *SPR*, *BCR*, and *ECR* items were conducted within each field test form.

SR items were flagged for further scrutiny if:

- An item distractor was not selected by all students (i.e., nonfunctional distractor), or selected by a large number of high ability students, with low selection from other ability groupings (i.e., ambiguous distractor).
- An item *p*-value was less than .20 or greater than .90.
- An item point-biserial was less than .10 (i.e., poorly discriminating). If an item point-biserial was close to zero or negative, the item was checked for a miskeyed answer.

SPR items were flagged for further scrutiny if:

- An item *p*-value was less than .20 or greater than .90.
- An item point-biserial was less than .10 (i.e., poorly discriminating). If an item point-biserial was close to zero or negative, the item was checked for a wrong answer.

BCR and *ECR* were flagged for further scrutiny if:

- An item did not elicit the full range of rubric scores.
- The ratio of mean item score to maximum score was less than .20 or greater than .90.
- An item-total correlation was less than .10.

Any items needed a careful decision. For example, an item that was flagged as being difficult (*p*-value less than .20) and poorly discriminating (point-biserial less than .10) was considered for dropping as a possible operational item. If the item represented important content that had not been extensively taught, however, it would be justified to be included in operational test form.

Differential Item Functioning Analyses

Analyses of *Differential item functioning* (*DIF*) are intended to compare the performance of different subgroups of the population on specific items, when the group have been statistically matched on their tested proficiency.

In present analyses, the gender reference group was males, and the ethnic group was Caucasians. The gender focal group was females and the ethic focal group was African-Americans. Because

the 2007 MSA-Math included both the *SAT10* items and the "Maryland-specific" items on each field test form, the total score as the matching variable consisted of selected SAT items and Maryland-specific items.

Any *SR*, *SPR*, *BCR*, and *ECR* items that were flagged as showing *DIF* were subjected to further examination. For each of these items, for example, math experts judged if the differential difficulty of the item was unfairly related to group membership:

- If the difficulty of the item is unfairly related to group membership, then the item should not be used at all.
- If the difficulty of the item is related to group membership, then the item should only be used if there is no other item matching the test blueprint.

For further information about the *DIF* procedures used for the 2007 MSA-Math, please see the section 3.7, *Differential Item Functioning*.

Item Response Theory (IRT) Analyses

To put field test items on the same scale of the operational test items, field test items were calibrated by fixing the parameters of the operational test items within each test form. Then, item difficulties, step difficulties, and fit statistics were stored in the 2007 item bank.

1.10 Linking, Equating, and Scaling Procedures

The 2007 MSA-Math was calibrated, equated, and scaled using Rasch fixed method. It should be noted that only SR items were considered as potential year-to-year linking items.

Stratified Random Sampling Procedures

To select equating samples to conduct linking and equating with, stratified random sampling procedures were used in 2007. To verify that the sample was representative of the statewide examinee population in terms of gender and ethnicity, the distributions of gender and ethnicity in the 2007 sample were compared with the total 2007 MSA population distributions. Appendix A, The 2007 MSA-Math Stratified Random Sampling provides the results of sampling. The results indicated that the calibration sample were representative of the statewide examinee population in terms of gender and ethnicity.

Robust Z Procedures

Robust z values were calculated by the following calculations (South Carolina Department of Education, 2001):

- The mean and standard deviation of the linking pool's item difficulties for each form
- The ratio of the standard deviations between form 1 and the rest of the forms
- The correlation between test form 1 and other test form item difficulties
- The difference between test form 1 and other test form item difficulties for each item in the linking pool
- The mean of the differences calculated above
- The median of the differences
- The interquartile range of the differences
- The robust z for each item in the linking pool where the robust z is defined as (the difference between the test form1 and other test form item difficulty minus the median of the differences) / (interquartile range multiplied by 0.74).

Guidelines for Possible Linking Items

Once the above calculations were made, the following guidelines were taken in determining possible sets of common items to be used for the Rasch equating (SCDE, 2001):

- Do not include those items with an absolute value of robust z exceeding 1.645. In addition, if one difficulty or step from a *SR* item is eliminated from the pool based on robust z, all other difficulties are also removed.
- Do not eliminate more than 20 percent of the pool linking items.
- Consider that the ratio of the standard deviations of the test form 1 and other test form item difficulties should be in the 90 to 110 percent range.
- It is assumed that the correlation of the test form 1 and other test form item difficulties is greater than .95.

The reason to apply these guidelines was to exclude items that changed in difficulty more than the other items.

Form-to-Form Linking Procedures

The stability of Maryland-specific common items appearing on both form A and form B was verified at each grade level:

- Calibrate the two operational test forms separately
- Calculate robust z with Rasch difficulties for form A and form B
- Correlate Rasch difficulties for form A and form B

After examining the robust z and correlations from the two separate calibration, it was determined that the common item difficulties were consistent across the two forms for all items and could be included as form-to-form linking items in the fixed calibration of the two forms.

Year-to-Year Linking Procedures

Year 2007 operational form A and form B contained a set of Maryland-specific common items that appeared in previous years including Year 2006. It should be note that the Rasch fixed method was applied to all items to put them on a same scale within each grade.

The stability of the equating common items was evaluated using robust z, correlation coefficients, and standard deviations.

Tables 1.56 through 1.61 included Rasch item difficulties used for calculating robust z values, correlation coefficients, and standard deviations. Figures 1.14 through 1.37 depicts common item difficulty between the previous years and either 2007 form A or B. It should be noted that the item difficulties in 2007 form A or B were obtained from independent calibration, and those in previous years were on a common scale (e.g., linked to 2006 item parameters).

Table 1.73 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 3

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
43	3509931	0.9627	0.6837	46	3510071	0.9317	0.7153
45	3510009	0.0690	-0.1842	48	3509955	1.8411	1.2869
46	3509953	-1.6881	-1.9647	53	3509964	-0.0360	-0.0539
47	3548054	-1.7100	-1.6588	54	3509966	-0.7332	-1.1353
48	3509955	1.8411	1.1119	55	3509923	0.0147	-0.5749
53	3509964	-0.0360	0.0899	56	3509959	0.8740	0.4752
54	3509966	-0.7332	-1.3466	59	3509926	2.4187	2.2610
55	3509974	1.0359	0.6713	61	3509927	0.4123	-0.1538
56	3509979	-0.0209	-0.6789	62	3509928	-0.6271	-1.2844
57	3509987	0.9682	0.6222	63	3510009	0.0690	-0.1055
61	3510003	-0.4214	-0.7608	64	3510069	2.8084	2.5339
62	3510006	1.2257	0.8755	66	3509929	1.8021	1.3531
63	3548055	-2.4386	-1.9474	67	3509930	-1.9318	-2.2234
64	3510011	0.9634	0.7961	68	3510018	0.2953	-0.1660
66	3510018	0.2953	-0.1075	69	3510027	-0.5906	-1.0539
68	3510023	1.8271	1.5209	70	3510029	-1.3693	-2.4609
69	3510027	-0.5906	-0.9837	72	3510035	-0.6165	-0.7741
70	3510029	-1.3693	-2.1718	78	3510053	-0.2691	-0.6839
71	3510032	-1.0976	-0.9166	79	3509933	-1.2635	-1.2463
72	3510035	-0.6165	-0.9137	80	3510051	1.4814	1.4787
78	3510051	1.4814	1.2881	81	3509962	-0.6247	-1.1088
79	3510053	-0.2691	-0.7478	87	3510062	-0.3652	-0.7752
80	3510055	1.2952	0.8052	88	3510063	0.4861	-0.1525
81	3510058	-0.6059	-0.9048	90	3509935	1.2515	0.5663
87	3510062	-0.3652	-0.7452	93	3510006	1.2257	0.6824
88	3510063	0.4861	-0.1890	100	3548063	-1.4037	-2.0495
90	3510065	-2.1822	-2.2558	105	3509958	-0.3242	-1.0244
91	3510066	0.0425	-0.3184	106	3509961	-1.3667	-1.6120
97	3510071	0.9317	0.6059	107	3510066	0.0425	-0.5789
105	3509958	-0.3242	-0.9357	108	3509938	-1.6759	-1.8554
106	3509961	-1.3667	-1.8030	109	3510070	-2.6459	-3.0487
107	3510068	-0.3305	-0.4760	113	3510041	-1.8190	-1.7124
108	3510069	2.8084	2.2797	114	3510043	0.0444	-0.0284
109	3510070	-2.6459	-3.1687	117	3510044	-0.5231	-0.6076
113	3510041	-1.8190	-1.6254				
114	3510043	0.0444	0.0422				
117	3510044	-0.5231	-0.8977				

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.132	-.441	-.064	-.445
SD	1.276	1.209	1.284	1.304
Comparison of Each Form with Base Form				
Correlation with Base	1.000	.977	1.000	.980
SD Ratio	100%	95%	100%	102%
Mean Diff	N/A	-.309	N/A	-.380
Median Diff	N/A	-.346	N/A	-.406
IQR Diff	N/A	.285	N/A	.387

Based on robust z and item difficulty plot, none of items was dropped from the year-to-year linking item pool.

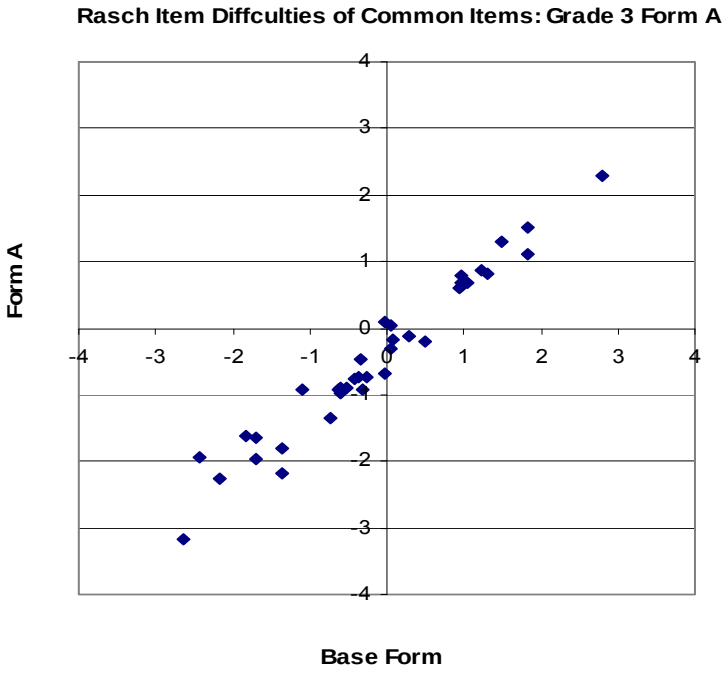


Figure 1.14 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 3 Form A

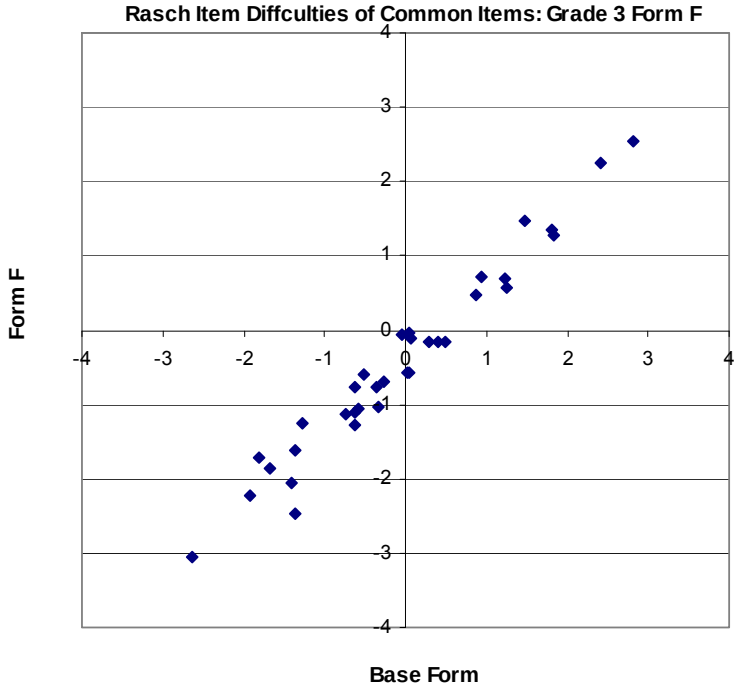


Figure 1.15 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 3 Form F

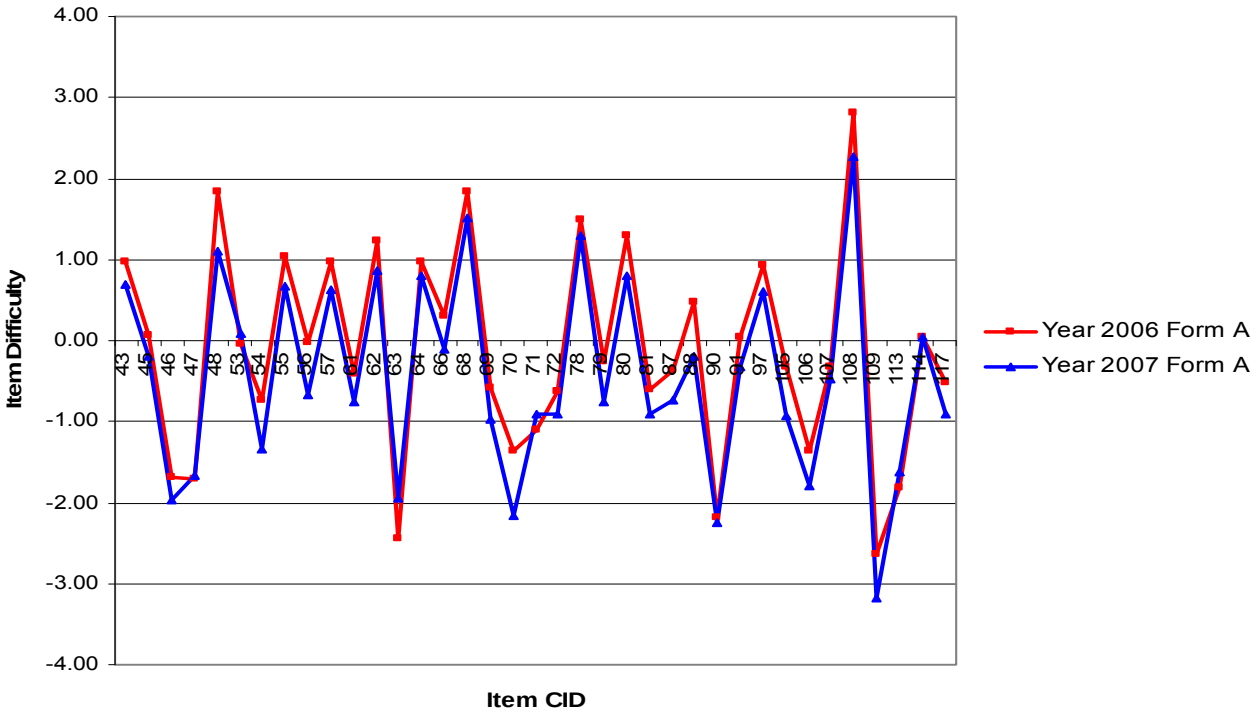


Figure 1.16 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 3 Form A

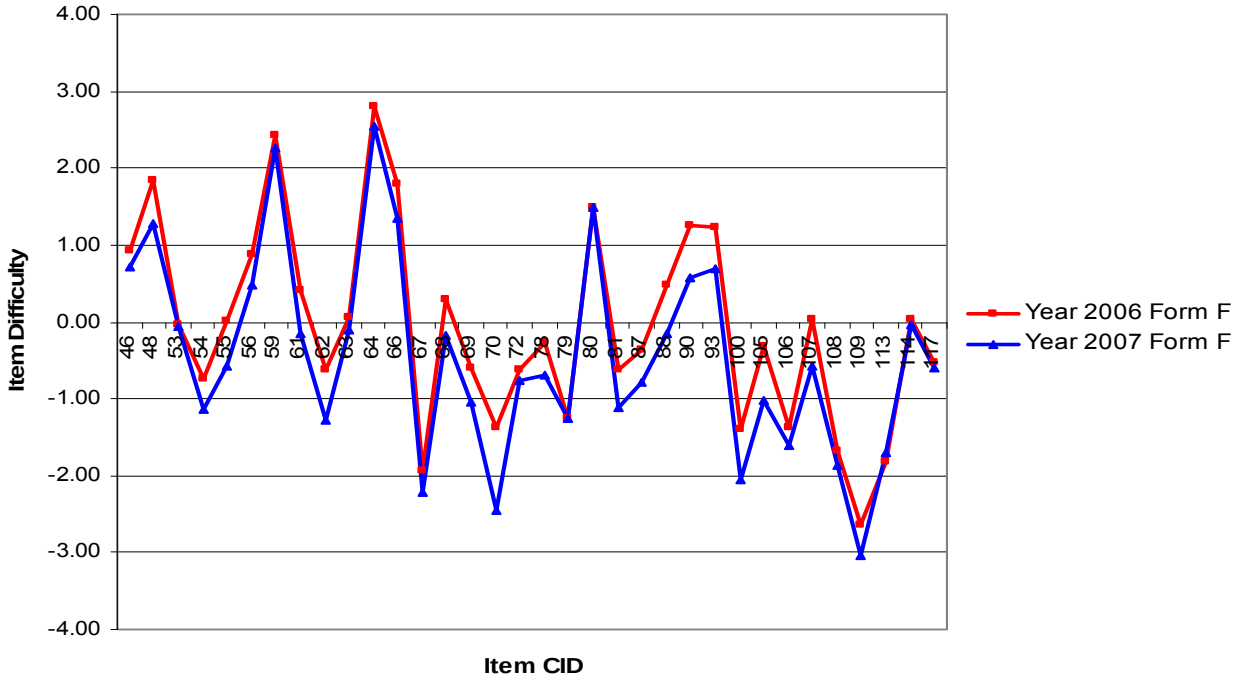


Figure 1.17 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 3 Form F

Table 1.74 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 4

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
43	3515406	0.6241	0.6808	43	3515407	-0.799	-0.9891
44	3515407	-0.799	-1.1104	44	3515596	-0.5201	-0.4596
45	3515408	0.1763	0.1875	45	3515447	1.4979	1.4064
46	3515410	-1.055	-0.7743	46	3515408	0.1763	0.0715
47	3515411	-0.6969	-0.915	47	3515599	0.223	0.0209
48	3515421	-0.6701	-0.8543	48	3515410	-1.055	-0.8837
53*	3515425	0.5403	0.4166	53	3515600	-0.3709	-0.2448
54	3515426	1.6228	1.4429	54	3515601	-0.0452	0.0123
55	3515428	-1.7288	-2.5015	55	3515602	1.0231	0.982
56	3515447	1.4979	1.4194	56	3515428	-1.7288	-2.3651
59	3515604	0.394	0.4531	59	3515604	0.394	0.3731
60	3515456	-0.7475	-0.6217	60	3515605	0.9009	0.9469
61*	3515467	-2.1248	-2.6576	61	3515456	-0.7475	-0.718
63	3515470	0.0797	0.0716	**62***	3515467	-2.1248	-2.77
65	3515471	-0.9767	-1.0736	63	3515606	-1.7067	-1.8384
66	3515479	0.0054	-0.0497	65	3515471	-0.9767	-1.2539
67	3515484	-1.7626	-1.8937	67	3515486	0.7468	0.7526
68	3515486	0.7468	0.8056	68	3548078	0.6281	1.1101
69	3515630	0.9291	1.1992	69	3515630	0.9291	1.2086
70	3515631	-0.4674	-0.4493	70	3515631	-0.4674	-0.4745
71	3515490	-1.2672	-1.9834	71	3515632	-0.0118	-0.0831
76	3515514	-1.4725	-1.5693	76	3515634	-0.2435	-0.2938
77	3515519	-0.6898	-0.7244	77	3515635	0.6901	0.6652
78	3515533	-0.7839	-1.0635	79	3515636	0.8456	0.8808
79	3515543	-0.2743	-0.6122	80	3515545	-0.8464	-1.1328
80	3515545	-0.8464	-1.2133	87	3515557	-0.0497	-0.0331
87	3515557	-0.0497	0.1906	88	3515640	1.772	1.3988
88	3515558	2.1761	1.8836	91	3515641	-0.8522	-0.9445
91	3515559	0.1734	-0.0701	92	3515570	0.8666	1.1331
92	3515570	0.8666	1.1297	93	3515571	-0.9395	-1.136
93	3515571	-0.9395	-1.1122	94	3515643	1.757	1.7974
94	3515573	1.3716	1.1774	95	3515645	-0.1355	0.0122
95	3515574	-0.9677	-1.0717	103	3515424	0.7782	0.7214
103*	3515423	-0.2585	-1.6655	**104***	3515425	0.5403	-0.1808
104	3515424	0.7782	0.7266	109	3515575	-0.1077	-0.5578
109	3515575	-0.1077	-0.4825	110	3515576	0.5508	0.3789
110	3515576	0.5508	0.5856	117	3515506	-1.2169	-1.4875
116	3515500	-0.3554	0.0348				
117	3515506	-1.2169	-1.5423				
118	3548083	-1.6445	-1.2604				

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.234	-.372	-.017	-.107
SD	1.015	1.132	.968	1.084

Comparison of Each Form with Base Form				
Correlation with Base	1.000	.958	1.000	.976
SD Ratio	100%	112%	100%	112%

Mean Diff	N/A	-.137	N/A	-.091
Median Diff	N/A	-.114	N/A	-.050
IQR Diff	N/A	.305	N/A	.237

Based on robust z and item difficulty plot, item 61 on Form A and item 62 on Form F were dropped from the year-to-year linking item pool because this item appeared on both operational forms. Second, item 103 on Form A was dropped from the year-to-year linking item pool. Finally, item 104 and item 53 were dropped from the year-to-year linking item pool because this item appeared on both operational forms.

The following correlation and SD ratio were calculated after dropping those items:

Comparison of Each Form with Base Form				
Correlation with Base	1.000	.972	1.000	.980
SD Ratio	100%	108%	100%	110%

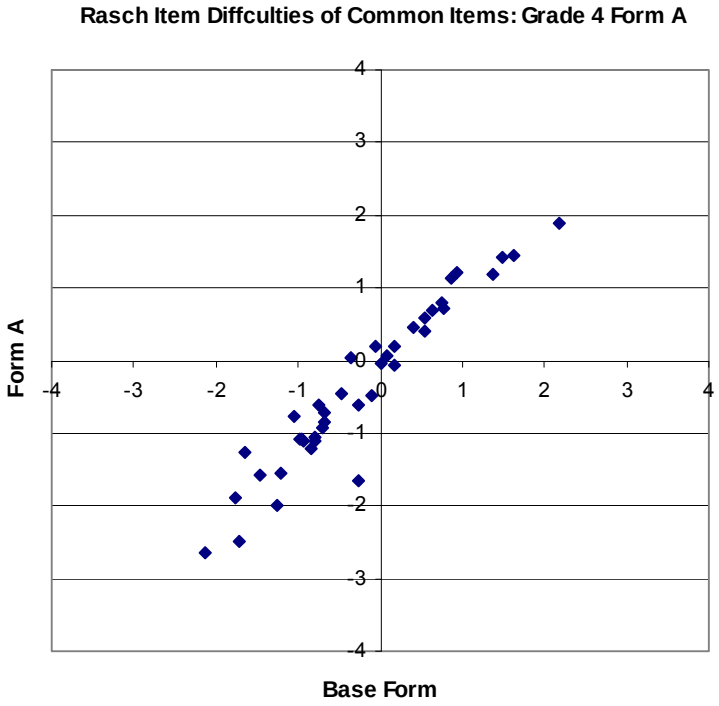


Figure 1.18 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 4 Form A

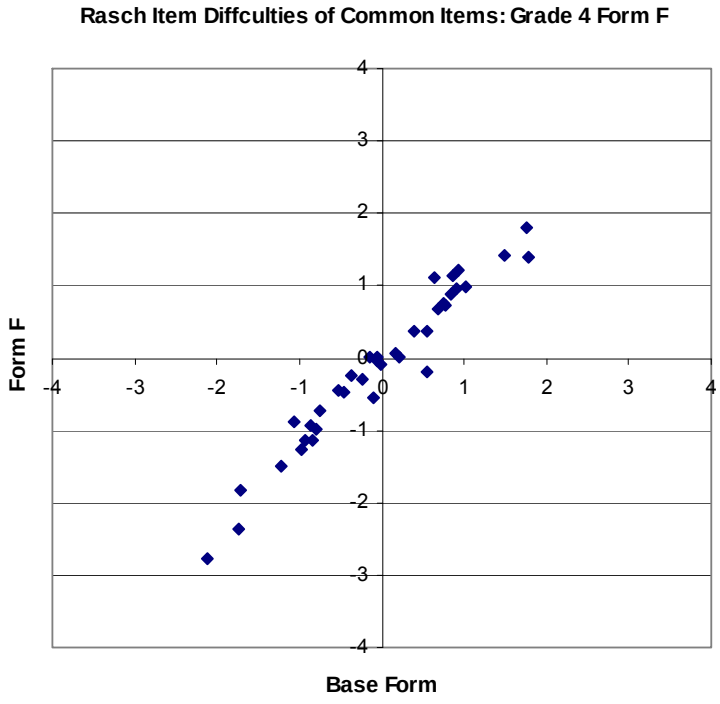


Figure 1.19 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 4 Form F

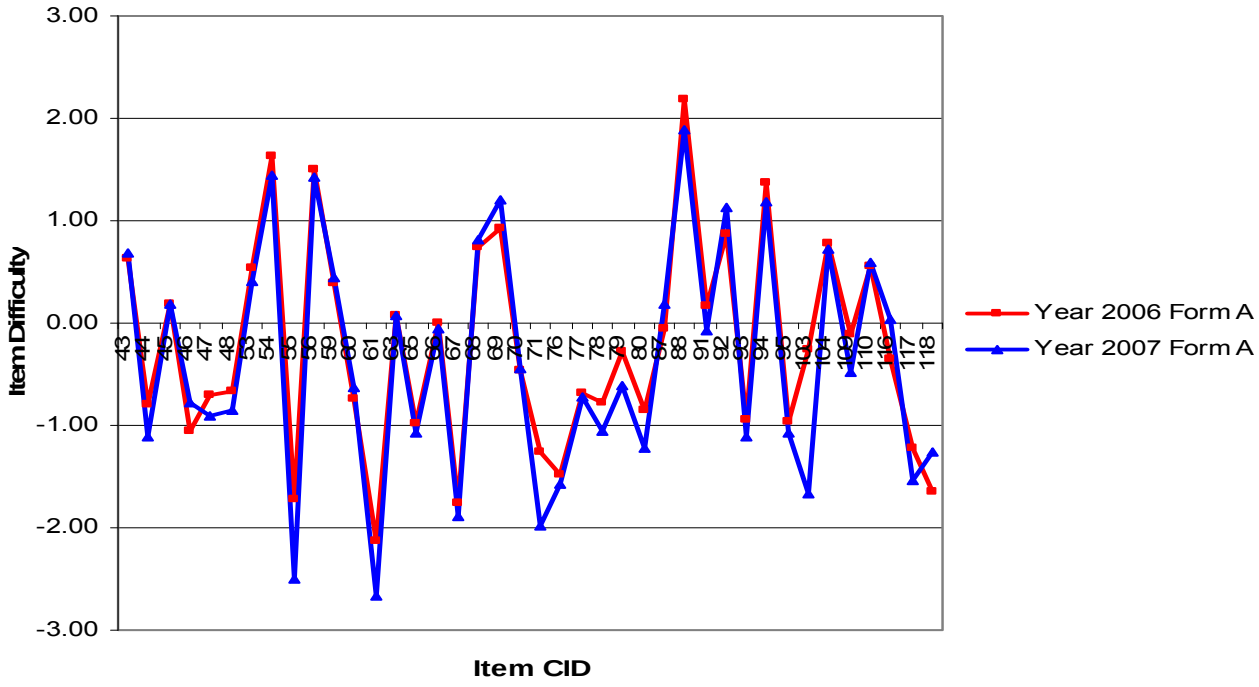


Figure 1.20 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 4 Form A

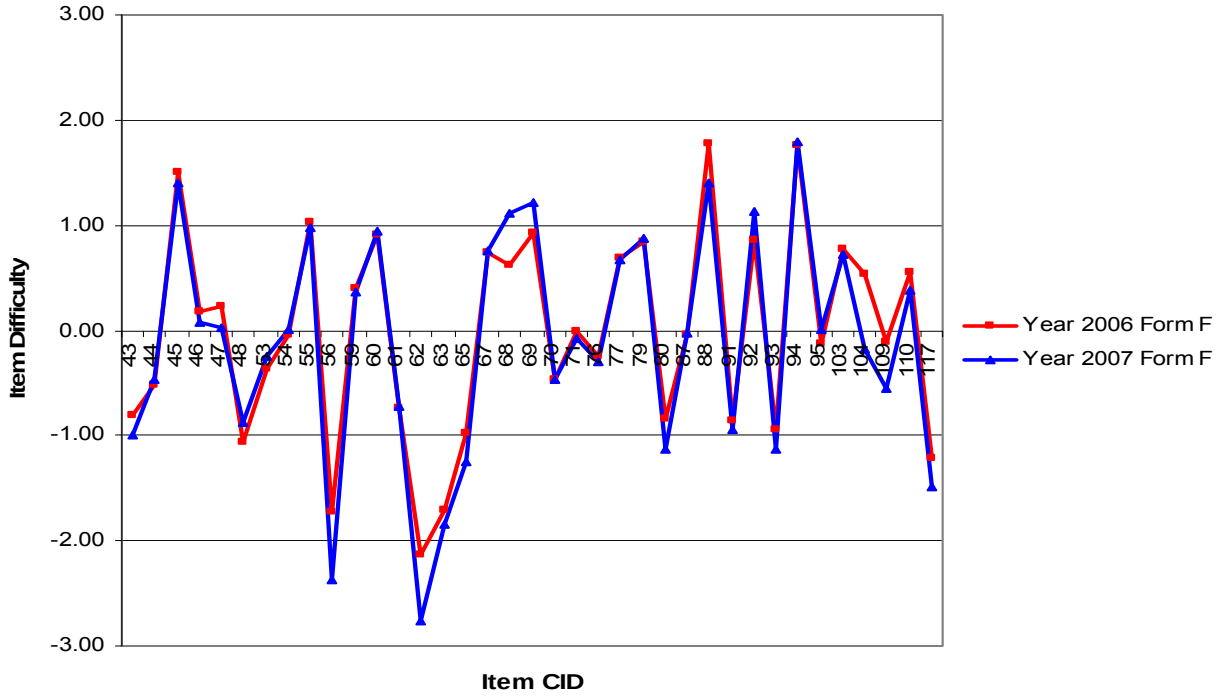


Figure 1.21 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 4 Form F

Table 1.75 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 5

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
43	3511196	0.6094	0.7018	43	3512527	-0.05	0.212
44	3511203	-1.3086	-1.4551	44	3512528	-1.1721	-1.401
45	3511216	0.203	0.0172	46	3511216	0.203	-0.0382
47	3511246	-0.331	-0.6265	47	3511246	-0.331	-0.7657
53	3511307	1.5483	1.473	48	3512529	0.4459	0.5238
54	3511312	1.5795	1.5909	53	3511307	1.5483	1.2787
59	3511339	0.4633	0.2294	54	3511312	1.5795	1.381
60	3511345	-1.6886	-2.1475	59	3512534	0.0331	-0.2011
61	3511348	0.8118	0.5502	60	3511345	-1.6886	-2.3045
63	3511371	0.8516	0.7913	61	3511348	0.8118	0.488
64	3511376	-0.9892	-0.8515	62	3512540	0.4738	0.5593
66	3511396	-1.1516	-1.1272	63	3511371	0.8516	0.5825
67	3511410	-0.0507	0.1043	64	3512543	-0.342	-0.2709
70	3511429	-0.5025	-0.3514	65	3512546	-1.0402	-1.2427
71*	3511433	-2.4669	-3.4859	67	3512553	0.4855	0.4293
72	3511439	-0.5779	-0.7847	70	3511439	-0.5779	-0.6998
82	3511458	-1.7042	-1.9448	71	3511410	-0.0507	-0.2172
83	3511467	-0.9093	-1.3621	72	3511396	-1.1516	-1.1562
85	3511470	-0.6898	-0.9032	82	3512578	-1.362	-1.695
86	3511479	0.6218	0.9952	83	3511467	-0.9093	-1.2808
91	3511504	-1.255	-1.9101	85	3511470	-0.6898	-1.0443
92	3511513	-1.1293	-1.1966	86	3511479	0.6218	0.3728
93	3511521	0.2895	-0.0159	91	3511504	-1.255	-1.6049
99	3511266	0.0148	-0.1144	92	3511513	-1.1293	-1.3747
100	3511320	-1.4191	-1.9041	93	3511521	0.2895	-0.2296
101	3512595	-0.6828	-0.7769	99	3511266	0.0148	-0.2322
104	3511499	0.1746	0.2666	100	3511320	-1.4191	-2.2311
105	3511330	0.6342	0.2373	101	3512595	-0.6828	-0.6063
107	3511269	-1.0845	-0.8993	104	3511499	0.1746	0.1529
111	3511442	0.5383	0.4517	105	3511330	0.6342	0.1631
115	3511448	-0.6839	-0.5046	107	3511269	-1.0845	-0.9209
				111	3511442	0.5383	0.269
				115	3511448	-0.6839	-0.8143

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.332	-.482	-.210	-.422
SD	.992	1.142	.867	.931
Comparison of Each Form with Base Form				
Correlation with Base	1.000	.974	1.000	.970
SD Ratio	100%	115%	100%	107%
Mean Diff	N/A	.264	N/A	-.212
Median Diff	N/A	.186	N/A	-.241
IQR Diff	N/A	1.993	N/A	.277

Item 71 on Form A was dropped from the year-to-year linking item pool based on robust z and item difficulty plot.

The following correlation and SD ratio are based on dropping that item:

Comparison of Each Form with Base Form				
Correlation with Base	1.000	.974	1.000	.970
SD Ratio	100%	110%	100%	107%

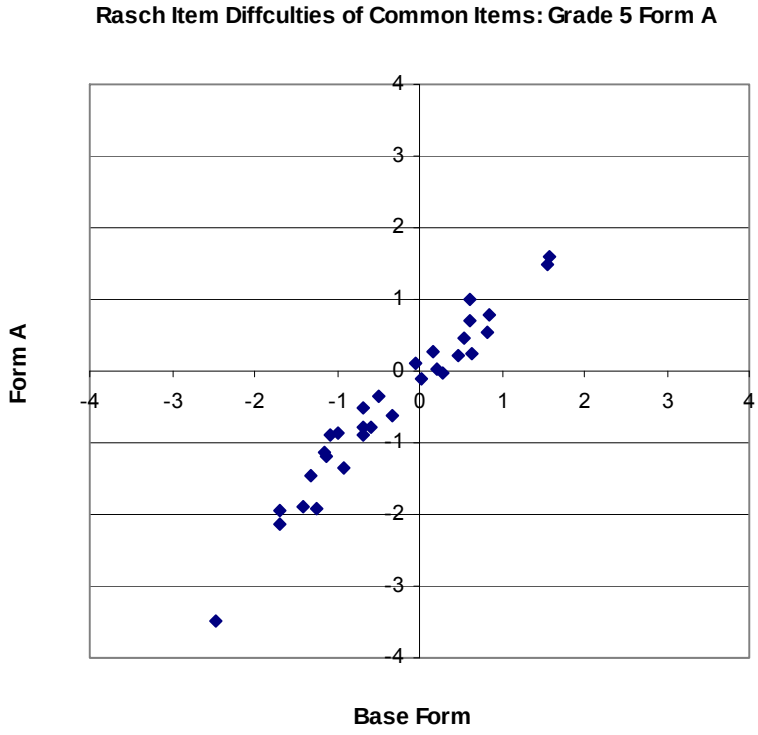


Figure 1.22 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 5 Form A

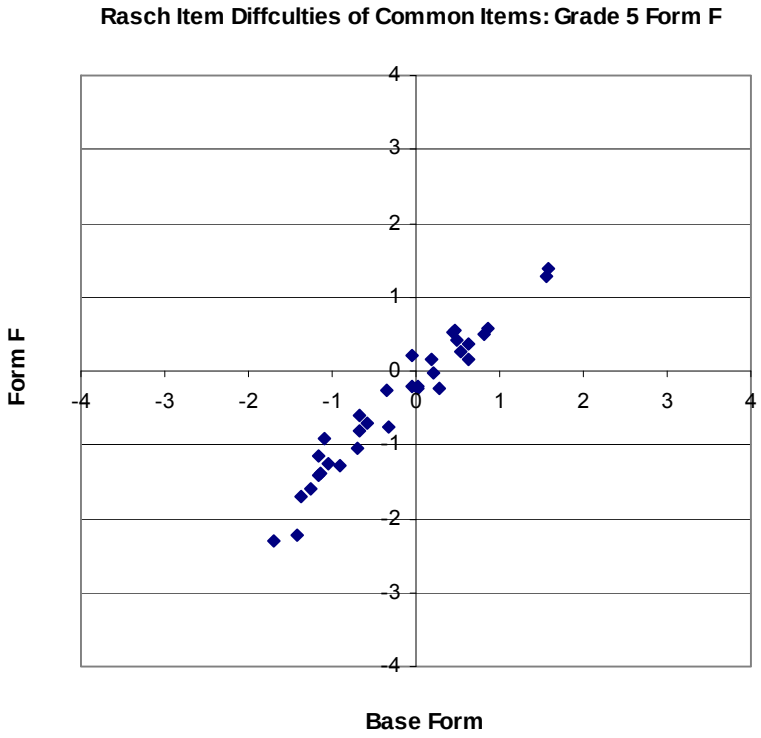


Figure 1.23 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 5 Form F

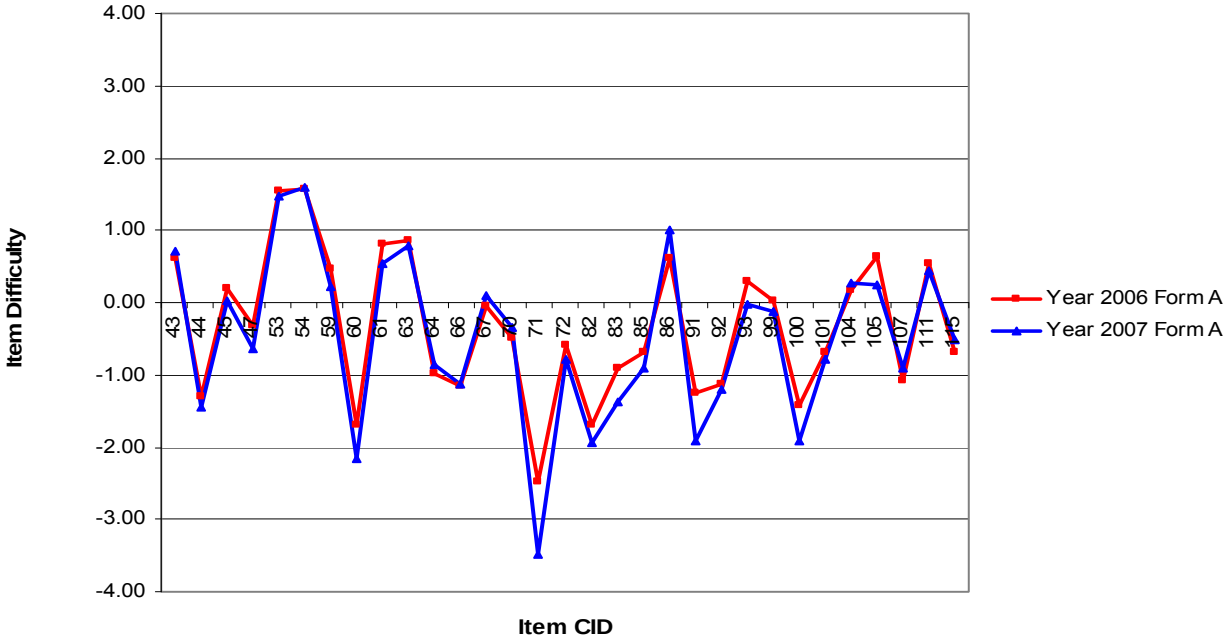


Figure 1.24 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 5 Form A

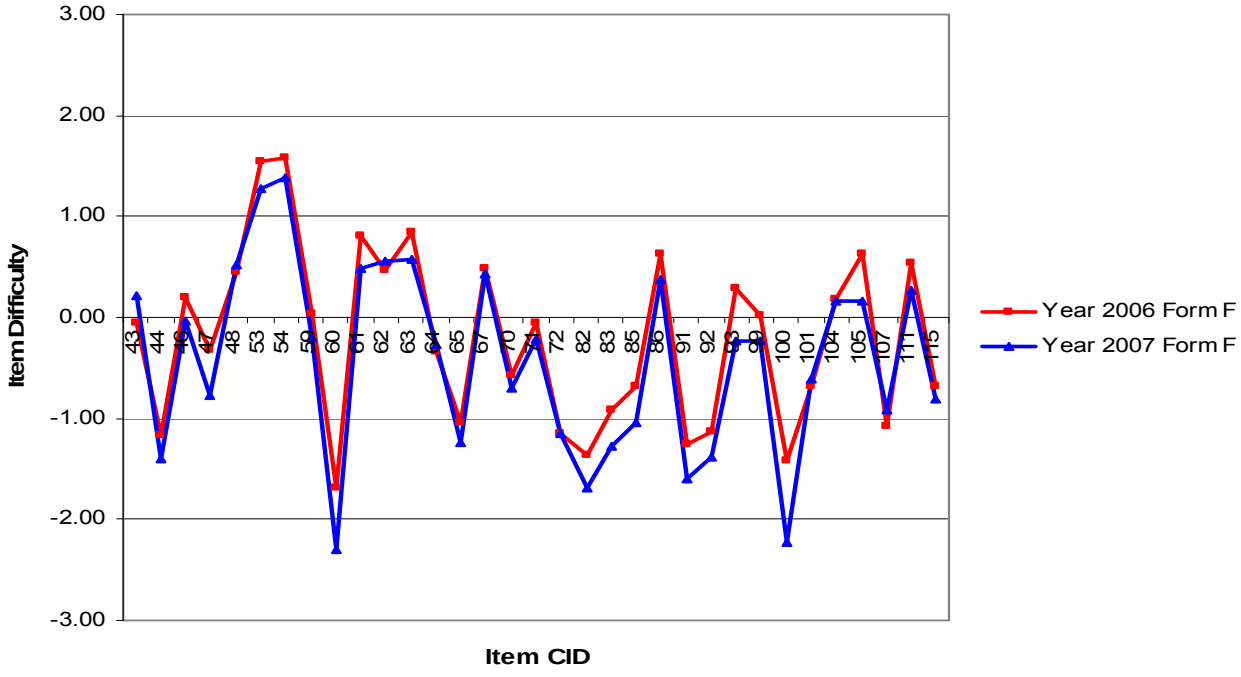


Figure 1.25 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 5 Form F

Table 1.76 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 6

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
41	3516240	0.2409	0.3261	41	3516240	0.2409	0.3147
42	3516241	-1.47	-1.4286	42	3516429	-2.1599	-2.2951
43	3516243	-0.284	-0.4469	43	3516242	1.2969	1.2199
44	3516242	1.2969	1.2646	45	3516243	-0.2844	-0.3867
46	3516248	-0.728	-0.7815	46	3516247	0.3674	0.2915
47	3516247	0.3674	0.3193	47	3516248	-0.7278	-0.6479
48	3516249	-0.409	-0.4048	48	3516451	-0.7288	-0.7314
51	3516255	-0.47	-0.4865	51	3516255	-0.4703	-0.5764
52	3516256	0.135	0.0774	52	3516256	0.135	0.0709
53	3516257	-1.205	-1.3342	53	3516280	0.6666	0.6641
54	3516258	0.3254	0.3466	54	3516453	-0.816	-0.8582
55	3516279	-0.664	-0.6689	55	3516454	-1.1295	-1.1634
56	3516280	0.6666	0.6066	56	3516455	0.5348	0.7063
57	3516281	0.8563	0.9184	58	3516517	1.4277	1.6545
58	3516283	0.9203	0.9967	61	3516559	-1.4432	-1.4389
61	3516285	0.3104	0.3886	62	3516565	0.8786	0.9057
62	3516290	-0.14	-0.2788	63	3516571	1.3093	1.4981
63	3516291	0.6406	0.7515	66	3516291	0.6406	0.8261
66	3516298	1.7544	1.7661	67	3516573	-0.3209	-0.4822
67	3516573	-0.321	-0.5068	68	3516301	-0.2182	-0.5785
68	3516301	-0.218	-0.3874	69	3516302	-0.4092	-0.3939
69	3516302	-0.409	-0.2799	70	3516303	0.658	0.4906
70	3516303	0.658	0.4878	71	3516594	-1.0547	-1.0921
71	3516305	-0.258	-0.305	72	3516313	-1.3362	-1.4039
72	3516307	0.2577	0.0152	73	3516613	0.4071	0.3893
73	3516310	-0.35	-0.4757	76	3516305	-0.2581	-0.3954
76	3516313	-1.336	-1.3814	77	3516320	-1.6189	-2.3998
77	3516318	-1.83	-1.786	86	3516328	-0.7001	-0.6845
86	3516328	-0.7	-0.7878	87	3516293	1.0083	0.844
87	3516293	1.0083	0.8819	88	3516618	1.3357	1.5791
88	3516330	-0.311	-1.082	89	3516621	-1.2026	-0.4359
89	3516331	1.1378	1.0437	94	3516623	-0.7259	-0.5891
94	3516352	-0.834	-0.8714	95	3516624	1.9065	2.1527
95	3516353	0.0893	0.2476	96	3516625	-1.2475	-1.4366
96	3516354	-0.784	-0.6552	97	3516354	-0.7843	-0.7022
97	3516355	-0.185	-0.2654	102	3516332	0.5885	0.4543
102	3516351	0.4777	0.5648	103	3516351	0.4777	0.5405
103	3516332	0.5885	0.5398	104	3516329	0.5144	-0.2431
104	3516329	0.5144	0.0445	105	3516295	0.1004	-0.057
105	3516295	0.1004	-0.0852	114	3516318	-1.8302	-1.5518
114	3516320	-1.619	-2.1421	115	3516323	-0.0894	-0.2161
115	3516323	-0.089	-0.2875				

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.054	-.132	-.123	-.150
SD	.803	.844	.984	1.045
Comparison of Each Form with Base Form				
Correlation with Base	.978	.978	1.000	.971
SD Ratio	100%	105%	100%	106%
Mean Diff	N/A	-.075	N/A	-.024
Median Diff	N/A	-.048	N/A	-.026
IQR Diff	N/A	.171	N/A	.215

Based on robust z and item difficulty plot, item, none of items was dropped from the year-to-year linking item pool.

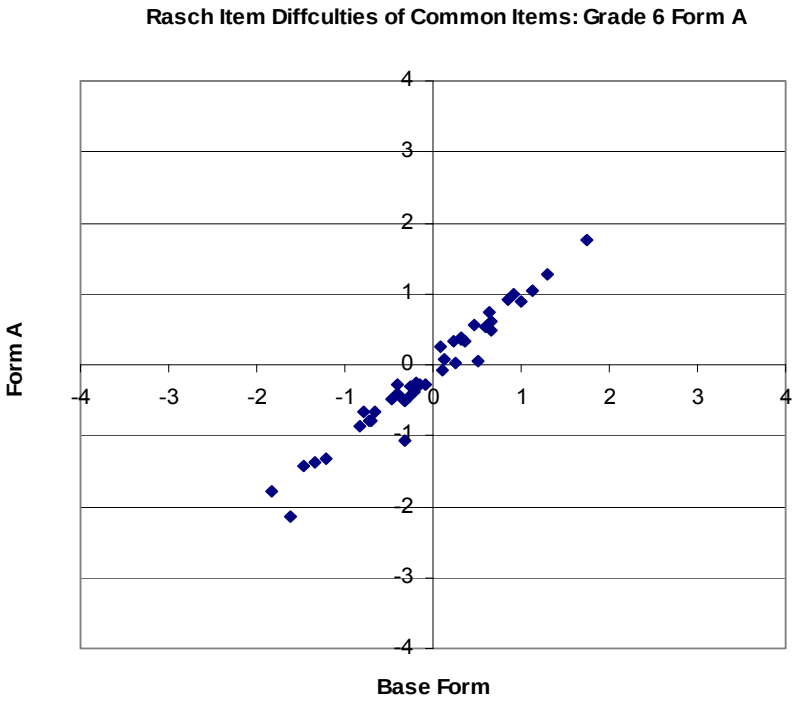


Figure 1.26 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 6 Form A

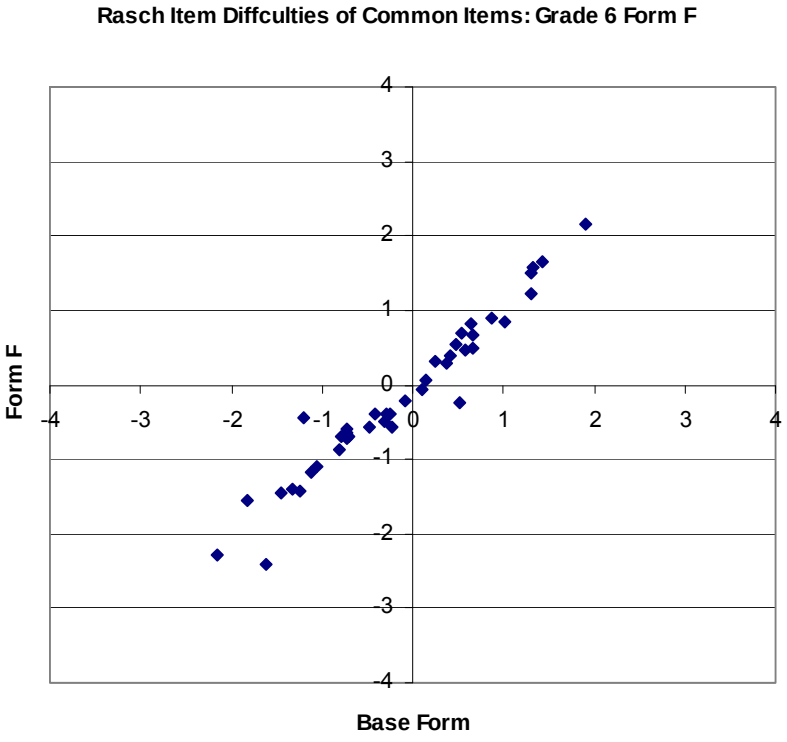


Figure 1.27 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 6 Form F

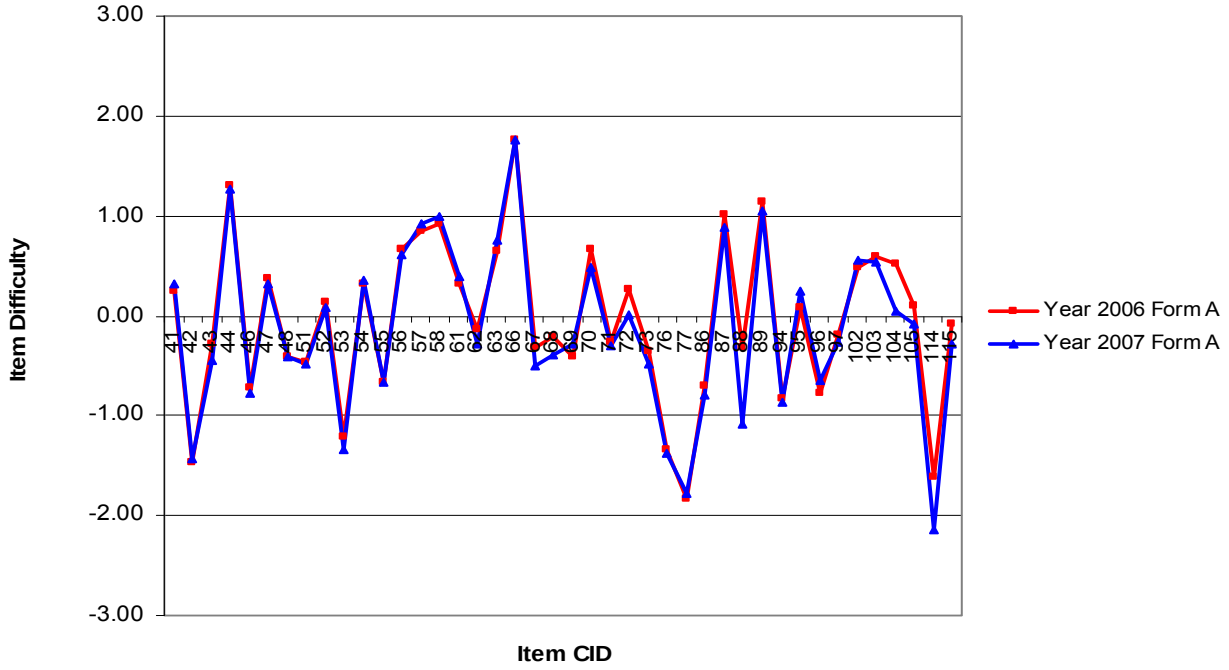


Figure 1.28 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 6 Form A

Figure 1.29 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 6 Form F

Table 1.77 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 7

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
43	3517604	1.0539	0.8998	43	3517613	-0.642	-0.4785
44	3517601	0.4455	0.1906	45	3517604	1.0539	1.2557
45	3517609	0.1508	-0.0066	46	3517602	0.7735	0.5138
46	3517613	-0.642	-0.7191	47	3517638	-1.1551	-0.9644
47	3517616	-0.14	-0.3203	49	3517609	0.1508	0.2495
48	3517634	-0.471	-0.7756	50	3517643	-0.6035	-0.7369
49	3517642	0.3982	0.3973	52	3517631	-1.1518	-1.1993
50	3517638	-1.155	-1.1761	53	3517634	-0.4706	-0.5127
51	3517647	-0.662	-1.0126	54	3517665	0.9745	1.1563
52	3517643	-0.604	-1.029	59	3517635	-0.9569	-0.9516
53	3517650	-0.468	-0.5976	60	3517615	-0.7793	-0.9715
54	3517652	-0.636	-1.0265	61	3517637	-1.1275	-1.2431
59	3547473	-1.124	-1.8105	62	3517639	1.4497	1.5914
60	3517663	1.5825	1.3037	**71***	3547535	-0.7473	-1.6587
61	3517665	0.9745	0.8688	72	3517687	-0.0583	-0.1884
62	3517667	-0.515	-0.479	73	3517692	-1.4991	-1.6751
72	3517687	-0.058	-0.4005	74	3517694	-1.2172	-1.3918
73	3517692	-1.499	-1.9087	85	3517709	-0.7302	-0.762
74	3517694	-1.217	-1.4742	86	3517712	0.5663	0.459
85	3517709	-0.73	-0.9238	87	3517714	0.0092	-0.1811
86	3517712	0.5663	0.1605	88	3517716	-0.4333	-0.5728
87	3517714	0.0092	-0.3799	89	3517662	0.3081	0.256
88	3517716	-0.433	-0.8327	90	3517721	0.5231	0.5546
89	3517718	-0.296	-0.8678	91	3517664	-1.621	-1.9417
90	3517721	0.5231	0.3243	102	3517650	-0.4683	-0.391
91	3517723	0.6126	0.4626	103	3517652	-0.6359	-0.8357
102	3517656	-0.409	-0.8084	111	3517718	-0.2963	-0.5278
111	3517697	1.2314	0.7277	113	3555859	-1.4603	-1.4769
113	3555859	-1.46	-1.5				

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.171	-.438	-.366	-.451
SD	.810	.839	.831	.923
Comparison of Each Form with Base Form				
Correlation with Base	1.000	.977	1.000	.974
SD Ratio	100%	104%	100%	111%
Mean Diff	N/A	-.267	N/A	-.085
Median Diff	N/A	-.257	N/A	-.080
IQR Diff	N/A	.249	N/A	.223

Based on robust z and item difficulty plot, item 71 on Form F was dropped from the year-to-year linking item pool.

The following correlation and SD ratio were calculated after dropping the item:

Comparison of Each Form with Base Form				
Correlation with Base	1.000	.977	1.000	.988
SD Ratio	100%	104%	100%	108%

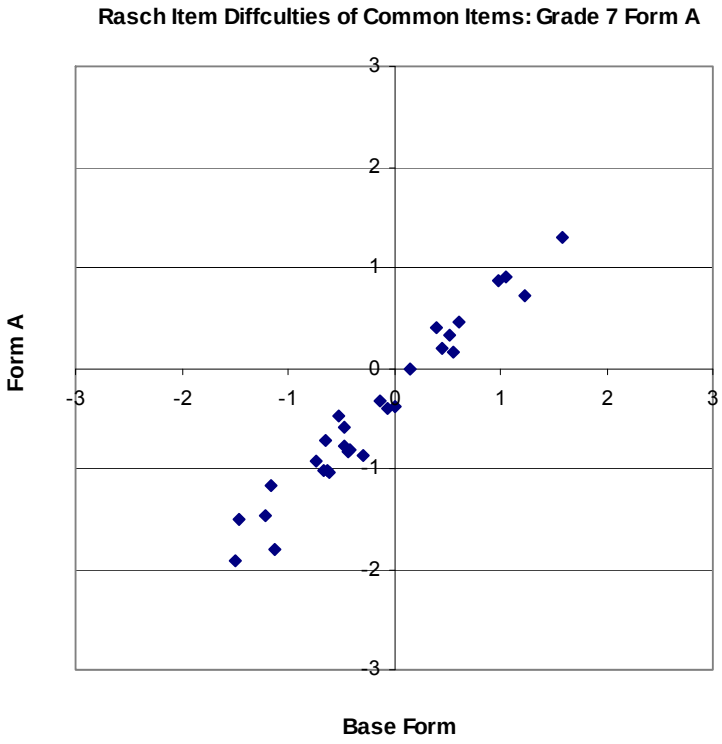


Figure 1.30 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 7 Form A

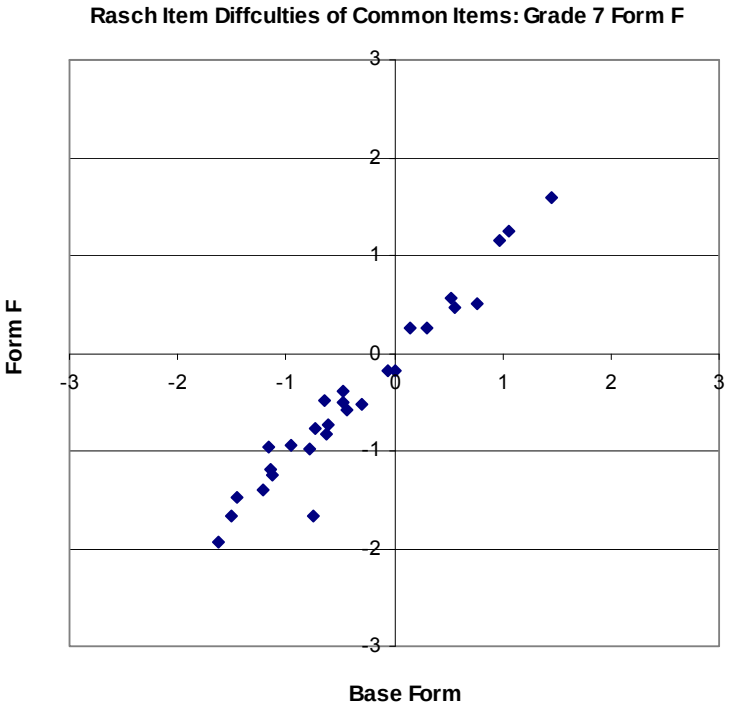


Figure 1.31 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 7 Form F

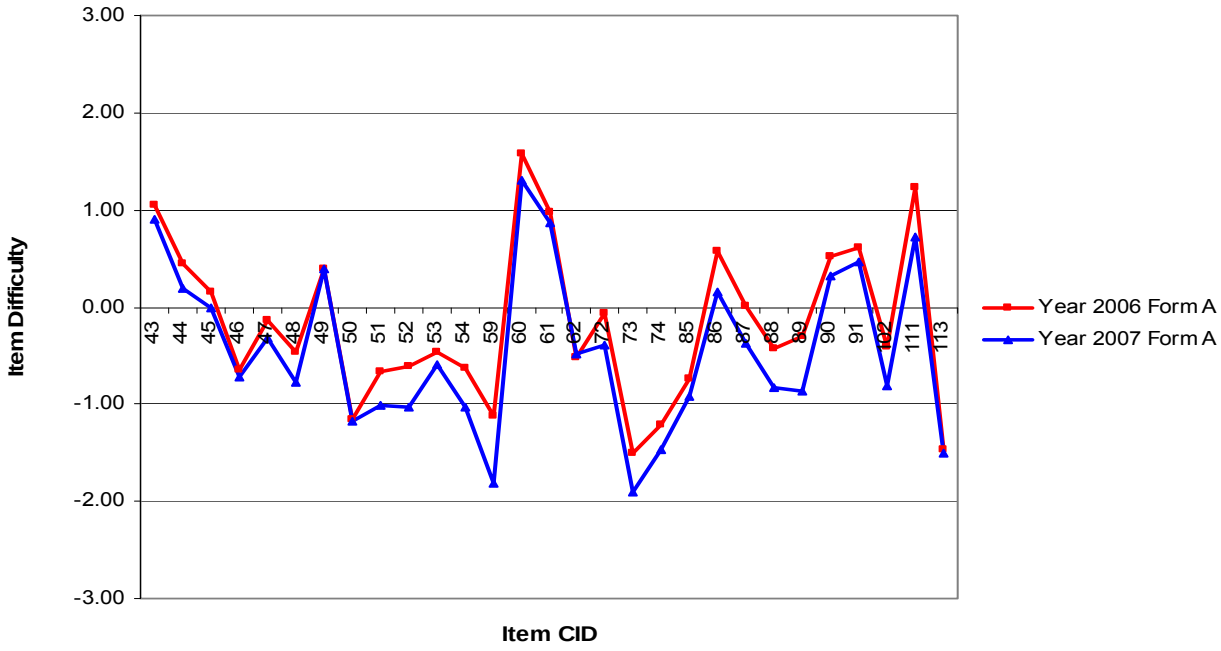


Figure 1.32 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 7 Form A

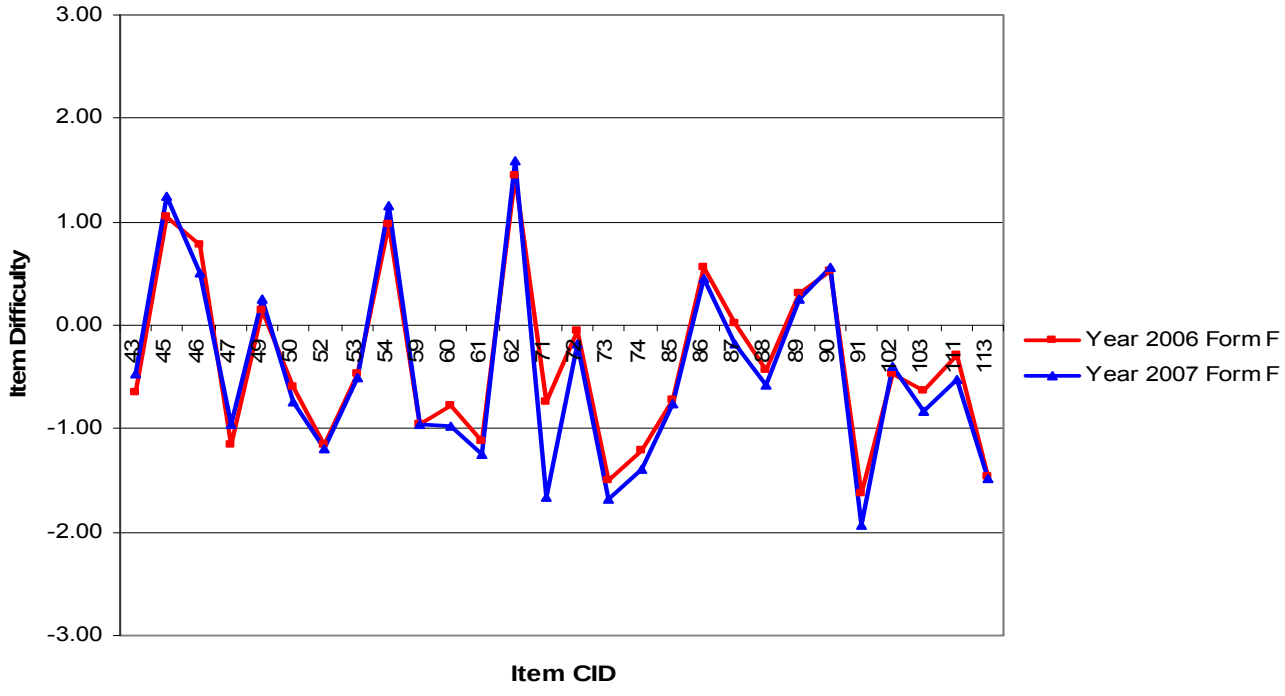


Figure 1.33 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 7 Form F

Table 1.78 Common Linking Item Difficulties of Year 2006 vs. Year 2007 MSA-Math: Grade 8

Item Seq. No.	Item CID	Y06 FA	Y07 FA	Item Seq. No.	Item CID	Y06 FF	Y07 FF
41	3514015	1.4965	1.5748	41	3514015	1.4965	1.6124
42	3514014	-0.2177	-0.4332	42	3514014	-0.2177	-0.434
43	3514016	-1.3613	-1.525	43	3514016	-1.3613	-1.6123
44	3514046	-0.1452	-0.266	44	3514055	-0.2581	-0.5373
53*	3514056	-1.4852	-2.1083	47	3514052	-0.1085	-0.1131
54	3514053	-1.2003	-1.3134	53	3514058	1.0306	1.0909
55	3514058	1.0306	1.0853	54	3514062	0.5139	0.37
56	3514059	-0.5815	-0.7848	55	3514059	-0.5815	-0.7641
57	3514062	0.5139	0.3858	56	3514156	-1.4579	-1.4167
66	3514073	-0.4061	-0.412	**57***	3514056	-1.4852	-1.3956
67	3514074	0.3257	0.2186	65	3514092	0.2379	0.3826
68	3514075	-0.6275	-0.8553	66	3514075	-0.6275	-0.8298
76	3514092	0.2379	0.2284	67	3514073	-0.4061	-0.273
77	3514102	-0.4851	-0.7808	68	3514076	0.1649	0.0838
78	3514095	1.2102	0.9362	76	3514173	0.0215	-0.1
80	3514093	0.8718	0.7389	77	3514095	1.2102	1.1593
81	3514107	2.3547	2.4035	78	3514174	0.1391	-0.5804
82	3514103	-0.533	-0.8017	80	3514100	-1.0918	-1.2854
97	3514057	-0.938	-0.9027	82	3514213	-0.6097	-0.9327
100	3514055	-0.2581	-0.4687	83	3514103	-0.533	-0.628
101	3514052	-0.1085	-0.2249	84	3547555	-0.5164	-0.2558
106	3514076	0.1649	0.0528	97	3514046	-0.1452	-0.238
107	3514100	-1.0918	-1.6524	106	3514074	0.3257	0.4216
				107	3514102	-0.4851	-0.8236

Form Statistics	Y06 FA	Y07 FA	Y06 FF	Y07 FF
Mean	-.054	-.213	-.198	-.296
SD	.968	1.075	.777	.833
Comparison of Each Form with Base Form				
Correlation with Base	1.000	.991	1.000	.969
SD Ratio	100%	111%	100%	107%
Mean Diff	N/A	-.160	N/A	-.098
Median Diff	N/A	-.128	N/A	-.094
IQR Diff	N/A	.163	N/A	.273

Based on robust z and item difficulty plot, item 53 on Form A was dropped from the year-to-year linking item pool, and item 57 on Form F was also dropped because this item appeared on both operational forms.

The following correlation and SD ratio were calculated after dropping this item:

Comparison of Each Form with Base Form				
Correlation with Base	1.000	.992	1.000	.968
SD Ratio	100%	108%	100%	110%

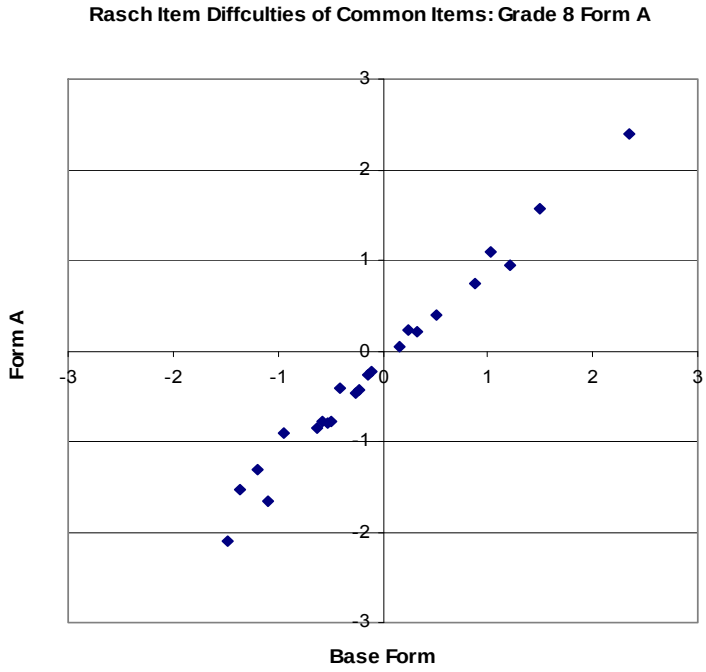


Figure 1.34 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 8 Form A

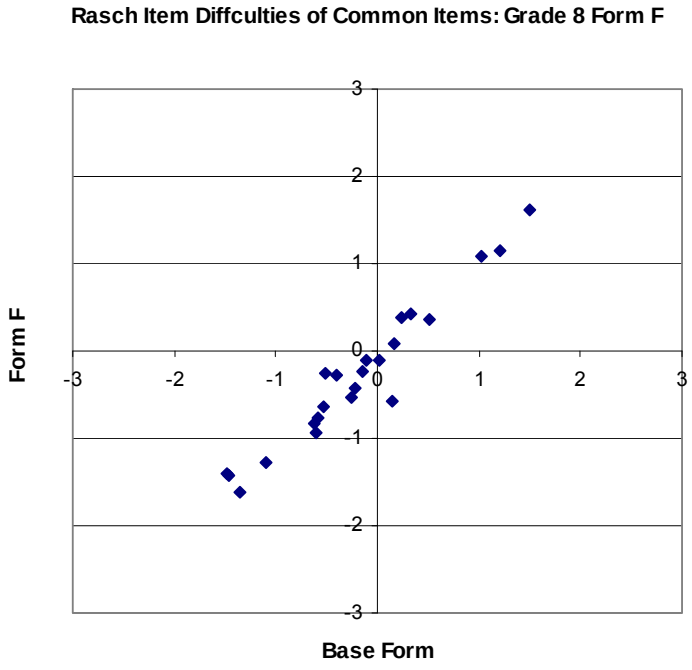


Figure 1.35 Item Difficulty Plot of Base Year Form vs. Current Year Form: Grade 8 Form F

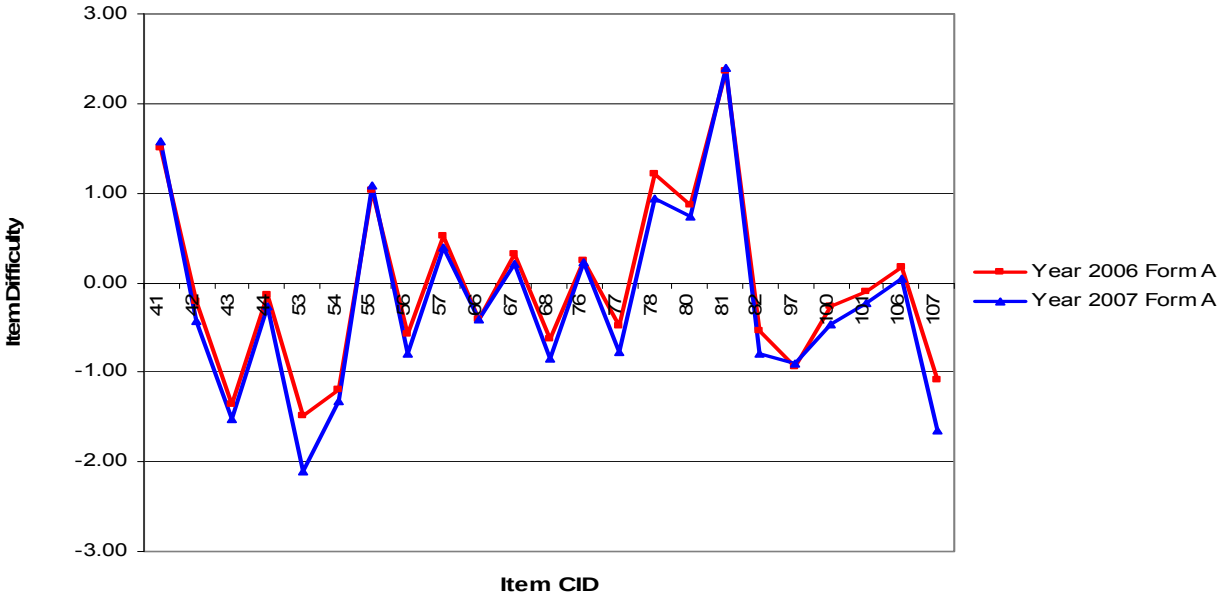


Figure 1.36 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 8 Form A

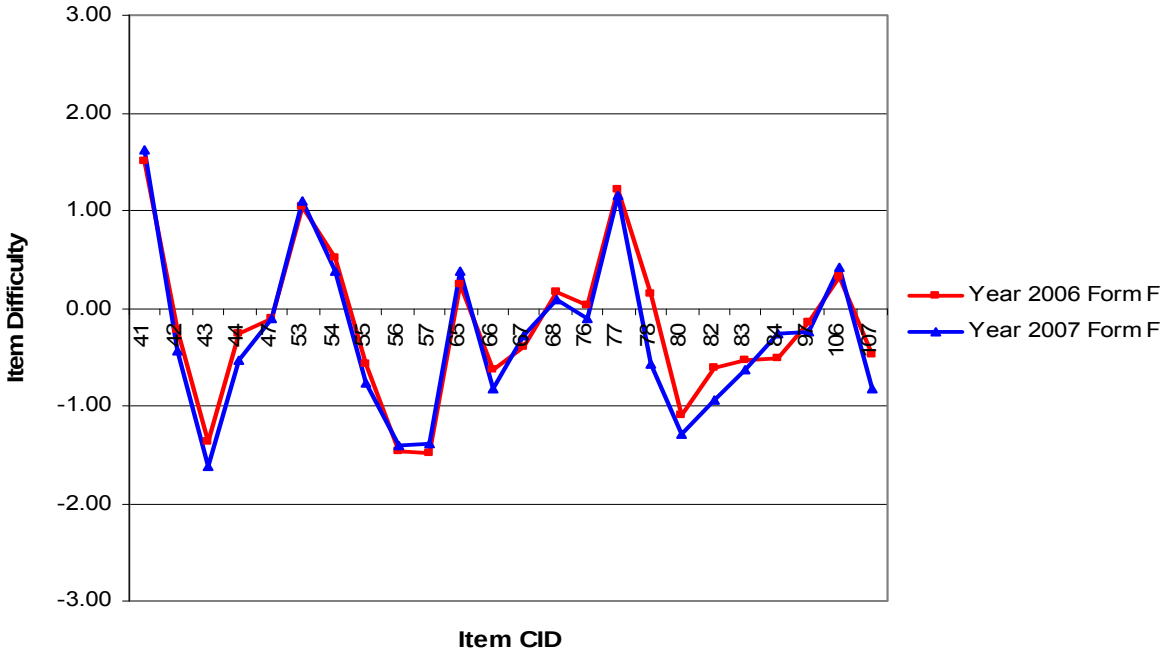


Figure 1.37 Free Calibration Item Difficulty Comparison of Year 2006 vs. Year 2007: Grade 8 Form F

Reporting Scale Scores

In order to facilitate the use and interpretation of the results of the 2007 MSA-Math, MSDE provided Harcourt with specifications about the score scale (Mean = 400, SD = 40, LOSS = 240, HOSS = 650). For grade 4, for example, the following is the formula to convert each student' ability or theta to scale score:

$$ReportingAbilityScaleScore = 32.8398 \cdot theta + 380.2954$$

$$ReportingSEM = 32.8398 \cdot SEM$$

where

theta = the *IRT* ability estimate, and

SEM = the conditional *SEM* of the ability estimate.

Table 1.79 depicts the slope and intercept to use for each grade. It should be noted that these same slops and intercepts were used for both Year 2006 recalibration and scaling, and Year 2007 scaling.

Table 1.79 The 2007 MSA-Mathematic Slope and Intercept: Grades 3 through 8

Grade	Slope	Intercept
3	32.6935	352.2959
4	32.8398	380.2954
5	30.7057	390.2866
6	29.6236	398.5595
7	28.1690	405.9549
8	28.3634	418.4843

1.11 Score Interpretation

To help provide appropriate interpretation of the 2007 MSA-Math test scores, two types of scores were created: 240-650 scale scores, and performance levels and descriptions.

240-650 Scale Scores

As explained in section 1.10, *Linking, Equating, and Scaling*, the 2007 MSA-Math produced scale scores that ranged between 240 and 650. Those scale scores have the same meaning within the same grade, but those scores are not comparable across grade levels.

It should be noted that those scale scores have only simple meaning that higher scale scores represent higher performance in math tests. Thus, performance levels and descriptions can give a specific interpretation other than a simple interpretation because they were developed to bring meaning to those scale scores.

Performance Level Descriptors

As previously explained, performance level descriptors provide specific information about students' performance levels and help interpret the 2007 MSA-Math scale scores. They describe what students at a particular level generally know and can be applicable to all students within each grade level.

Maryland standards are divided into three levels of achievement (*www.marylandpublicshools.org*):

- Advanced is a highly challenging and exemplary level of achievement indicating outstanding accomplishment in meeting the needs of students.
- Proficient is a realistic and rigorous level of achievement indicating proficiency in meeting the needs of students.
- Basic is a level of achievement indicating that more work is needed to attain proficiency in meeting the needs of students.

As Table 2.1 shows a range of scale scores at each performance level, for example, grade 4 math scale scores from 374 to 432 indicate the level of *Proficient*. Students in this level passed MSA-Math standard. This level is considered a realistic and rigorous level of achievement. Further information about the 2007 MSA-Math score interpretation can be obtained from MSDE.

1.12 Test Validity

As noted in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), "validity is the most important consideration in test evaluation."

Messick (1989) defined validity as follows:

> Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (p.5)

This definition implies that test validation is the process of accumulating evidence to support intended use of test scores. Consequently, test validation is a series of on-going and independent processes that are essential investigations of the appropriate use or interpretation of test scores from a particular measurement procedure (Suen, 1990).

In addition, test validation embraces all of the experimental, statistical, and philosophical means by which hypotheses and scientific theories can be evaluated. This is the reason that validity is now recognized as a unitary concept (Messick, 1989).

To investigate the validity evidence of the 2007 MSA-Math, content-related evidence, item development procedures, information on different item functioning (DIF) with respect to gender and ethnicity, and evidence from internal structure were collected.

Content-Related Evidence

Content validity is frequently defined in terms of the sampling adequacy of test items. That is, content validity is the extent to which the items in a test adequately represent the domain of items or the construct of interest (Suen, 1990). Consequently, content validity provides judgmental evidence in support of the domain relevance and representativeness of the content in the test (Messick, 1989).

The 2007 MSA-Math blueprints provide extensive evidence regarding the alignment between the content in the 2007 MSA-Math and the *VSC*. The 2007 MSA-Math operational test forms were created from the item bank which contained both operational and field-test items which had been administered as operational items or field-test items in 2006 and before. Information on the item composition of these tests can be obtained from section 1.5, Test Structure of the 2007 MSA-Math. In addition, 2007 MSA-Math blueprints are presented in Appendix E.

Item Development

Test development for MSA-Math is ongoing and continuous. Content specialists, teachers all over Maryland, Harcourt, and MSDE were greatly involved in developing and reviewing test items. Committees such as content review, bias review, and vision review reviewed all of the items which were finally stored in the item bank. Specifically, an internal review by MSDE and Harcourt staff for alignment and quality required a great deal of time and energy. More specific information on item (test) development and review can be obtained in section 1.4, Development and Review of the 2007 MSA-Math.

Field testing was conducted within a test window scheduled. Once field-test items were scored, MSDE and Harcourt conducted additional item analysis and content review. Any field-test items

that exhibited statistics that suggested potential problems were carefully reviewed by content specialists within MSDE and Harcourt. A determination was then made as to whether the item should be eliminated or revised and field-tested again. Information on statistical analyses for field test items can be obtained in section 1.9, Field Test Analyses.

Differential Item Functioning (DIF)

1) Bias Review of Field Test Items

A separate Bias Review Committee examined each item on math tests looking for indications of bias that would impact the performance of an identifiable group of students. They discussed or rejected items biased on gender, ethnic, religious, or geographical bias.

2) DIF Statistics

For DIF analyses, subgroups were first identified to either reference or focal groups. For 2007 MSA-Math, males and whites were assigned to the reference group and females and African-Americans were assigned to the focal group.

For SR and SPR items, Harcourt applied Mantel-Haenszel procedure, and standardized mean difference (SMD) and standard deviation (SD) were calculated for BCR and ECR DIF analyses. All items were placed in severity classifications base don Educational Testing Service (ETS) guidelines. More information on DIF analyses can be obtained in section 3.7, Differential Item Functioning.

Evidence from Internal Structure

The 2007 MSA-Math has five math reporting strands: *Algebra, Geometry and Measurement, Statistics and Probability, Numbers and Computations,* and *Process*. Tables 4.3 through 4.8 show the correlations among the math strands.

1.13 Unidimensionality Analyses

Measurement implies order and magnitude on a single dimension (Andrich, 1989). Consequently, in the case of scholastic achievement, this requires a linear scale to reflect this idea of measurement. Such a test is considered to be unidimensional (Andrich, 1988, 1989). However, unidimensionality cannot be strictly met in a real testing situation because students' cognitive, personality, and test-taking factors usually have a unique influence on their test performance to some level (Andrich, 1988; Hambleton, Swaminathan, & Rogers, 1991). Consequently, what is required for unidimensionality to be met is an investigation of the presence of a dominant factor that influences test performance. This dominant factor is considered as the ability measured by the test (Andrich, 1988; Hambleton et al., 1991; Ryan, 1983).

To check the unidimensionality of the 2007 MSA-Math, polychoric correlation coefficients were computed with *LISREL 8.5* (Jöreskog & Sörbom, 1993) because they were polytomously scored on math tests. Principal component analysis was then applied to produce eigenvalues. The first and the second principal component eigenvalues were compared without rotation. Table 1.80 summarizes the results of the first and second principal component eigenvalues of the 2007 MSA-Math.

In general, the first factor extracted somewhat large amount of eigenvalues across all grades. With regard to factor analysis and eigenvalues, there is one unit of information per item so that the eigenvalues sum to the number of items. The rule of thumb to determine the unidimensionality of a test requires that the eigenvalue of the first component or factor should be at least three times larger than the second one. As can be seen, the size of the eigenvalue of the first component meets the criterion for the unidimensionality. Thus, we can conclude that the assumption of unidimensionality for the 2007 MSA-Math was met.

Table 1.80 The 2007 MSA-Math Eigenvalues between the First and Second Components

Grade	Form	Number of Items	First Eigenvalue	Second Eigenvalue
3	A	64	23.44	2.37
	F	64	23.06	2.62
4	A	64	22.75	2.17
	F	64	22.73	2.21
5	A	65	23.03	2.29
	F	63	20.97	2.17
6	A	62	23.00	2.12
	F	62	21.75	2.34
7	A	62	27.48	2.34
	F	62	26.82	2.38
8	A	62	25.67	2.18
	F	62	24.70	1.93

Note. Form A designates the operational portion of Forms A, B, C, D, and E, which is identical. Form F designates the operational portion of Forms F, G, H, J, and K, which is identical.
Note. Analyses were conducted based on a whole population.

1.14 Item Bank Construction

The number of test forms to be constructed each year and the need to replace items that would be released to the public necessitated the availability of a large pool of items. The 2007 MSA-Math item bank continues to be maintained by Harcourt as computer files and paper copies. This enables test items to be readily available to both Harcourt and MSDE staff for reference, test construction, test book design, and printing.

Harcourt maintains a computerized statistical item bank to store supporting and identification information on each item. The information stored in this item bank for each item is as follows:

- CID
- Test administration year and season
- Test form
- Grade level
- Item type
- Item stem and options
- Passage code and title
- Subject code and description
- Process code and description
- Standard code and description
- Indicator code and description
- Objective code and description
- Item status
- Item statistics

In terms of Rasch item statistics data, all field test items were calibrated by fixing the parameters of the operational test items within each operational test form. For example, each unique field test items of math test forms A, B, C, D, and E were independently calibrated after fixing the same operational items appearing across the field test forms with the same Rasch difficulties because these field test forms belonged to the same operational form A (e.g., contained the same operational items on each field test form). Then, item difficulties, step difficulties, and fit statistics were stored in the 2007 item bank.

1.15 Quality Control Procedures

A standard quality procedure at Harcourt Assessment, Inc. was to create a test deck for MSA programs. The test deck began when Quality Assurance entered mock data into the enrollment system, which was transferred to the materials requisition system; the order was packaged by our Distribution Center, and shipped to the Quality Assurance Department. We then reviewed the packing list against the data entered, the materials algorithms applied, the materials packaged against the packing list, and the actual packaging of the documents. These documents were then used to create a test deck of mock data along with advance copies of documents that were received from the printer. Advance printer copies were inclusive of documents throughout the print run to assure we were randomly testing printed documents. The Maryland test deck was a comprehensive set of all documents that:

- Verify all scan positions for item responses and demographics to verify scanning setup and scan densities
- Verify all constructed response score points, zoning of image, reader scoring, reader resolution, and reader check scores
- Verify the handling of blank documents through the system
- Test all demographic and item edits
- Verify pre-id bar code read, match and no-match
- Verify attemptedness rules applied by subtest
- Verify duplicate student handling (same test duplicate, different test duplicate)
- Verify duplicate student with different demographics rules applied
- Verify the document counts to the enrollment, pre-id and actual document receipt
- Verify pre-id matching and application to student record
- Verify various raw score points and access to dummy and live scoring tables
- Verify cut scores applied
- Verify valid score on one subtest and invalid score on other subtest
- Verify scoring applied to Braille and Large Print
- Verify valid multiple choice and invalid constructed response
- Verify valid constructed response and invalid multiple choice
- Verify all special scoring rules
- Verify all summary programs for rounding
- Verify summary inclusion and exclusion (Braille, standard and non-standard student summarization)
- Verify each scoring level for group reporting
- Verify all reporting programs for accuracy in all text and data presented
- Verify class, school, district, and state summary data on home reports
- Verify all data file programs to assure valid information in every field

- Verify data descriptions for accuracy against data file
- Create compare programs to allow for update of files

The Maryland test deck was the first order processed through the Maryland system to verify all aspects of the materials packaging, scanning, editing, scoring, summary, and reporting. Pre-determined conditions were included in the test deck to assure the programs were processing all data to meet the requirements of the program with zero defects. Processing of live orders couldn't proceed until each phase of the test deck had been approved by our Quality Assurance Department. An Issues Log with sign-off approvals was utilized to assure we were addressing any issues that arose in the review of the test deck data across all functional groups at Harcourt.

Prior to release of any order for reporting we received a preliminary file from Scoring Operations to run a key check TRIAN to assure that all scoring keys had been determined and applied accurately. Any item that was not performing as expected was flagged and reviewed by our content specialist and psychometrician. Upon completion of the key check, we proceeded to run the pilot level reports.

We ran the pilot district utilizing live data. The pilot district included multiple buildings, all grades, and any unique accommodations. A formal pilot review process was conducted with expert Harcourt staff prior to release of the information to MSDE.

Upon completion of the processing of all district level data, Harcourt Scoring Operations provided the Quality Assurance Department with a state level data file(s) and state data for review and approval. Harcourt Quality Assurance programmers duplicated all data independently to assure accurate interpretation of the expected results. A series of SAS programs were run on these files to assure 100% accuracy. These included but were not limited to:

- Statewide Duplicate Student
- Statewide FD of Demographic Variables
- District/Building/N-Count
- Statewide RS/SS/Cut Score tables
- Proc Means to verify summary statistics
- Item Response listing to verify all constructed responses are scored and within the valid range
- Normative data check for all raw scores
- Reader Resolution report to verify all readings and resolution combinations

Upon complete review and approval by Quality Assurance, we posted the statewide student files to a secure FTP site for review by MSDE.