

## **1. OVERVIEW OF THE 2009 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, with grades 4, 6, and 7 being added to the program in 2004. 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 obtained 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).

It should be noted that in 2007, the MSA-Math was administered using a new vendor and applying a different IRT method (e.g., the Rasch model); therefore, a transformation of scale scores using the equipercentile method was conducted with the 2006 population data. Detailed information on the scale score transformation and its results can be found in Appendix C, *Year 2006 MSA-Math Recalibration Results from 3-PL IRT to the Rasch Model Using the Equipercentile Method* in the *2007 MSA-Math Technical Report*.

In 2007, MSDE implemented an important action plan on the MSA-Math test: dropping all of the SAT10 items from the 2008 assessment. Consequently, several SAT10 items which contributed to the 2007 criterion-referenced test (CRT) were replaced by Maryland-specific items in 2008.

For the purposes of year-to-year linking and equating, we first constructed 2009 a linking pool: only operational selected-response (SR) items (i.e., multiple-choice items) were included in the linking pool. It should be noted that these SR items appeared both in current and previous years’ assessments and were used as either core or core link items in previous years’ assessments (i.e., in any assessment before 2009). After setting up the linking pool, we conducted a stability check of linking items and decided which items should be excluded from or which items remain in the linking pool. During the calibration and equating processes, we kept and fixed the original operational Rasch item difficulty parameters of any linking items that remained through the stability check to put the 2009 assessment on a common scale. Accordingly, scale scores of the 2009 assessment were linked back to the 2006 assessment and all the scale scores of different years were comparable within each content and grade.

### **1.1 Purposes/Uses of the 2009 MSA-Math**

By measuring students’ achievement against the new academic standards, the 2009 MSA-Math fulfills 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.

## 1.2 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.

The objectives assessed by the MSA at each grade level are embedded in the VSC. In addition, they are identified with the notation, ***assessment limit***. Assessment limits provide clarification about the specific skills and content that students are expected to have learned for each assessed objective. Even though some objectives in the VSC may not have an Assessment limit at a given grade-level, these non-assessed objectives still must be included in instruction. They introduce important concepts in preparation for assessed skills and content at subsequent grade levels.

The following provides one example of assessment limit of Grade 3 MSA-Math:

### STANDARD 1.0

#### Knowledge of Algebra, Patterns, and Functions

##### TOPIC:

##### A. [PATTERNS](#) AND FUNCTIONS

##### INDICATOR:

1. Identify, describe, extend, and create numeric [patterns](#) and functions

##### OBJECTIVES:

- a. Represent and analyze numeric [patterns](#) using skip counting

##### Assessment limits:

Use 2, 5, 10, or 100 starting with any whole number (0 – 1000)

It should be noted that it was not the case that every indicator would necessarily be tested each year even if 100% of the standards should be tested. Consequently, the VSC specified curricular indicators and objectives that contributed directly to measuring content standards, which were aligned to the MSA. More information on assessment limits and standards can be found in Appendix D, *The 2009 MSA-Math Blueprint*.

### **1.3 Development and Review of the 2009 MSA-Math Items and Test**

As seen in Table 1.1, the development of the 2009 MSA-Math test required the involvement of four groups in addition to MSDE and Pearson. These groups are as follows:

#### **National Psychometric Council**

The National Psychometric Council (NPC) took a major role in reviewing and making recommendations to MSDE on the development and implementation of the 2009 MSA-Math program. For example, they made recommendations to MSDE on issues, such as test blueprints, operational form construction, 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 approach 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 Vision Review Committee reviewed the items and any associated art for bias to the visually impaired. The committee makes their recommendations to NOT put any item they had a concern with on Form A.

**Table 1.1 The 2009 MSA-Math Responsibility for Test Development**

Development of the 2009 MSA-Math	Primary Responsibility
Development of Preliminary Blueprints and Item Specifications	Pearson; MSDE; NPC
Development of Operational Form Requirement and Session Blueprint	MSDE
Item Writing	MSDE; Pearson
Item Review	Pearson; MSDE; NPC; Content Review Committee
Bias Review	Pearson; MSDE; Bias Review Committee
Vision Review	Pearson; MSDE; Vision Review Committee
Modification of Special Forms	Pearson; MSDE
Review of Special Forms	MSDE
Construction of Operational Test Forms	Pearson; MSDE; NPC
Construction of Field Test Forms	Pearson; MSDE
Review of Operational Test Forms	MSDE
Final Construction of Test Forms	Pearson; MSDE

## 1.4 Test Form Design, Specifications, Item Types, and Item Roles

### Test Form Design

The MSA-Math test had two forms of operational items at each grade. Field test items were embedded within the operational items resulting in a total of 10 test forms at each grade. As can be seen from Table 1.2, Forms A, B, C, D and E are identical with respect to operational items (designated as operational Form A) and differ only with respect to field test items. This is also true for Forms F, G, H, J, and K (designated as operational Form F).

**Table 1.2 The 2009 MSA-Math Test Form Design: 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.

### Test Form Specifications and Reporting Category

Tables 1.3, 1.4, and 1.5 provide information on the total number of operational items included in each operational test form and how these items were broken down based on each content standard. It should be noted that the test specifications in these tables represent the targeted test design for each grade and show the targeted distribution of each content standard.

Mathematics has a total of seven content standards (Algebra, Geometry, Measurement, Statistics, Probability, Numbers and Computation, and Process). It should be acknowledged that some standards were combined for purposes of reporting subscale. Specifically, Geometry and Measurement standards and Statistics and Probability standards were combined to produce a total of five subscale reporting categories. Tables 1.6 through 1.23 provide information on the actual distribution of score points by standard and reporting category. The number of items and score points for each reporting standard were identical across forms within each grade.

### Item Types

The 2009 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 2009 MSA-Math, students selected an answer from four options. Each *SR* item was scored dichotomously (i.e., 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 2009 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 a 0-1 score point range from two independent scorers; all *BCR* Step B items received a 0-2 score point range; all *ECR* Step B items received a 0-3 score point range from two independent scorers. The score given was the higher of the first and the second Reader's scores, provided they were adjacent. A resolution reader's score was used if two non-adjacent initial scores were received. That is, the resolution reader's score was used in place of both the first and second Reader's scores. It should be noted that grade 3 and 4 tests did not include *ECR* items.

### **The Role of Operational SR Item**

Most SR items were used for both form-to-form and year-to-year calibration and linking. As a result, operational SR items fell into one of the following four categories: unique core, common core, unique core linking, and common core linking items. First of all, it should be noted that form-to-form linking was conducted with both the common core and the common core linking items. Form-to-form calibration and linking procedures can be found in section of chapter 1.9, *Form-to-Form Linking Procedures*. More importantly, however, year-to-year linking was conducted with only the core linking items and year-to-year calibration and linking procedures can be found in section of chapter 1.9, *Year-to-Year Linking Procedures*.

While unique core items appeared on either operational form A or F, common core items appeared on both forms. As a result, only the common core items were used for form-to-form linking. Because the core items were not included into the possible 2009 linking pool, on the other hand, item parameters of these items were recalibrated with the 2009 live, operational data (i.e., stratified random sample) and then reserved in the 2009 Maryland item bank for the possible use as core linking items in the future. Classical and Rasch analyses on these core items can be found in section of chapter 1.8, *Validation Check with the 2009 Core Items*.

While a few core linking items appeared only on either operational form A or F (i.e., unique core linking), most core linking items (i.e., common core linking) appeared on both operational forms. As a result, the common core linking items appearing on both forms were used for both form-to-form and year-to-year linking. The unique core linking items were used only for year-to-year linking.

The role of the core linking items was to place the 2009 scale on the 2006 scale. Because these core linking items carried their operational item parameters on the 2006 scale, they were included in the 2009 year-to-year linking pool. Classical analysis on these items can be found in section of chapter 1.8, *P-Value Check with Year-to-Year Core Linking Items*, and calibration, linking and equating procedures on these core linking items can be found in chapter 1.9, *Linking, Equating, and Scaling Procedures of the 2009 MSA-Math*.

### **The Role of Operational SPR, BCR, and ECR Items**

SPR, BCR, and ECR items were divided into one of the following two categories: unique core or common core items. Only the common core items appearing on both operational forms were used for form-to-form calibration and linking. Because these items were not included in the 2009 year-to-year linking pool, new Rasch item and step difficulty parameters were estimated with the 2009 live, operational data set (i.e., stratified random sample). These new item and step difficulty parameters were used to produce each student's theta estimate. More detailed information about how much these items changed across years in terms of classical and Rasch item difficulty can be found in section of chapter 1.8, *Validation Check with the 2009 Core Items*.

**Table 1.3 Item Type of Content Standard for the 2009 MSA-Math: Grades 3 and 4**

Grade	Standard	Item Type	No. of Items of Each Form	
			A	F
3			65	65
	Algebra	<i>SR, BCR</i>	13	13
	Geometry	<i>SR, BCR</i>	8	8
	Measurement	<i>SR, BCR</i>	7	7
	Statistics	<i>SR, BCR</i>	12	12
	Probability	<i>SR</i>	2	2
	Number Computation	<i>SR, BCR</i>	16	16
	Process	<i>BCR</i>	7	7
4			64	64
	Algebra	<i>SR, BCR</i>	14	14
	Geometry	<i>SR, BCR</i>	7	7
	Measurement	<i>SR, BCR</i>	7	7
	Statistics	<i>SR, BCR</i>	8	8
	Probability	<i>SR, BCR</i>	7	7
	Number Computation	<i>SR, BCR</i>	14	14
	Process	<i>BCR</i>	7	7

*Note.* *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 Item Type of Content Standard for the 2009 MSA-Math: Grades 5 and 6**

Grade	Standard	Item Type	No. of Items of Each Form	
			A	F
5			65	65
	Algebra	<i>SR, BCR, ECR</i>	15	15
	Geometry	<i>SR, BCR</i>	6	6
	Measurement	<i>SR, BCR</i>	8	8
	Statistics	<i>SR, BCR</i>	9	9
	Probability	<i>SR, BCR</i>	4	4
	Number Computation	<i>SR, BCR</i>	15	15
	Process	<i>BCR, ECR</i>	8	8
6			62	62
	Algebra	<i>SR, BCR, ECR</i>	14	14
	Geometry	<i>SR, BCR</i>	8	8
	Measurement	<i>SR, BCR</i>	6	6
	Statistics	<i>SR, BCR</i>	9	9
	Probability	<i>SR, BCR</i>	4	4
	Number Computation	<i>SR, BCR</i>	14	14
	Process	<i>BCR, ECR</i>	7	7

*Note.* *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 Item Type of Content Standard for the 2009 MSA-Math: Grades 7 and 8**

Grade	Standard	Item Type	No. of Items of Each Form	
			A	F
7			62	62
	Algebra	SR, SPR, BCR, ECR	14	14
	Geometry	SR, SPR, ECR	7	7
	Measurement	SR, SPR, BCR	6	6
	Statistics	SR, SPR, BCR, ECR	9	9
	Probability	SR, SPR, BCR	5	5
	Number Computation	SR, SPR	14	14
	Process	BCR, ECR	7	7
8			61	60
	Algebra	SR, SPR, BCR, ECR	15	15
	Geometry	SR, SPR, ECR	8	8
	Measurement	SR, SPR, BCR	5	5
	Statistics	SR, SPR, BCR, ECR	9	8*
	Probability	SR, SPR, BCR	4*	5
	Number Computation	SR, SPR	12	12
	Process	BCR, ECR	8	7*

*Note.* 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.

*Note.* 62 items were originally developed for each operational form in grade 8. However, due to item development issues that affected two BCR items (i.e., scatter plot) on operational Form F caused MSDE and NPC to Do Not Score (DNS) the items. One probability SR item on operational Form A was deemed DNS.

**Table 1.6 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 3**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	13	8	7	12	2	16	7	65
F	13	8	7	12	2	16	7	65

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

**Table 1.7 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 3**

Form	Total and Reporting Standard Scores					
	1	2&3	4&5	6	7	Total Score
A	13	15	14	16	14	72
F	13	15	14	16	14	72

**Table 1.8 Item Type and Score Point Distribution for the 2009 MSA-Math: Grade 3**

Form	# of SR Item	# of BCR Item		Total # of Item	Scores of SR	Scores of BCR		Total Score
		Step A	Step B			Step A	Step B	
A	51	7	7	65	51	7	14	72
F	51	7	7	65	51	7	14	72

**Table 1.9 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 4**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	14	7	7	8	7	14	7	64
F	14	7	7	8	7	14	7	64

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

**Table 1.10 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 4**

Form	Total and Reporting Standard 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 2009 MSA-Math: Grade 4**

Form	# of SR Item	# of BCR item		Total # of Item	Scores of SR Item	Scores of BCR		Total Score
		Step A	Step B			Step A	Step B	
A	50	7	7	64	50	7	14	71
F	50	7	7	64	50	7	14	71

**Table 1.12 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 5**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	15	6	8	9	4	15	8	<b>65</b>
F	15	6	8	9	4	15	8	<b>65</b>

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

**Table 1.13 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 5**

Form	Total and Reporting Standard Scores					Total Score
	1	2&3	4&5	6	7	
A	15	14	13	15	17	<b>74</b>
F	15	14	13	15	17	<b>74</b>

**Table 1.14 Item Type and Score Point Distribution for the 2009 MSA-Math: Grade 5**

Form	# of SR Item	# of BCR Item		# of ECR Item		Total # of Item	Scores of SR	Scores of BCR		Scores of ECR		Total Score
		Step A	Step B	Step A	Step B			Step A	Step B	Step A	Step B	
A	49	7	7	1	1	<b>65</b>	49	7	14	1	3	<b>74</b>
F	49	7	7	1	1	<b>65</b>	49	7	14	1	3	<b>74</b>

**Table 1.15 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 6**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	14	8	6	9	4	14	7	<b>62</b>
F	14	8	6	9	4	14	7	<b>62</b>

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

**Table 1.16 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 6**

Form	Total and Reporting Standard Scores					
	1	2&3	4&5	6	7	Total Score
A	14	14	13	14	15	<b>70</b>
F	14	14	13	14	15	<b>70</b>

**Table 1.17 Item Type and Score Point Distribution for the 2009 MSA-Math: Grade 6**

Form	# of SR Item	# of BCR Item		# of ECR Item		Total # of Item	Scores of SR	Scores of BCR		Scores of ECR		Total Score
		Step A	Step B	Step A	Step B			Step A	Step B			
A	48	6	6	1	1	<b>62</b>	48	6	12	1	3	<b>70</b>
F	48	6	6	1	1	<b>62</b>	48	6	12	1	3	<b>70</b>

**Table 1.18 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 7**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	14	7	6	9	5	14	7	<b>62</b>
F	14	7	6	9	5	14	7	<b>62</b>

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

**Table 1.19 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 7**

Form	Total and Reporting Standard Scores					
	1	2&3	4&5	6	7	Total Score
A	14	13	14	14	17	<b>72</b>
F	14	13	14	14	17	<b>72</b>

**Table 1.20 Item Type and Score Point Distribution for the 2009 MSA-Math: Grade 7**

Form	# of SR Item	# of SPR Item	# of BCR Item		# of ECR Item		Total # of Item	Scores of SR	Scores of SPR	Scores of BCR		Scores of ECR		Total Score
			Step A	Step B	Step A	Step B				Step A	Step B	Step A	Step B	
A	36	12	4	4	3	3	<b>62</b>	36	12	4	8	3	9	<b>72</b>
F	36	12	4	4	3	3	<b>62</b>	36	12	4	8	3	9	<b>72</b>

**Table 1.21 Item Distribution of Each Content Standard for the 2009 MSA-Math: Grade 8**

Form	Total Item Number of Each Standard							Total # of Item
	1*	2*	3*	4*	5*	6*	7*	
A	15	8	5	9	4*	12	8	61
F	15	8	5	8*	5	12	7*	60

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

Note. 62 items were originally developed for each operational form in grade 8. However, due to item development issues that affected two BCR items (i.e., scatter plot) on operational Form F caused MSDE and NPC to Do Not Score (DNS) the items. One probability SR item on operational Form A was deemed DNS.

**Table 1.22 Total and Reporting Content Standard Scores for the 2009 MSA-Math: Grade 8**

Form	Total and Reporting Standard Scores					Total Score
	1	2&3	4&5	6	7	
A	15	13	13	12	19	72
F	15	13	13	12	17	70

**Table 1.23 Item Type and Score Point Distribution for the 2009 MSA-Math: Grade 8**

Form	# of SR Item	# of SPR Item	# of BCR Item		# of ECR Item		Total # of Item	Scores of SR	Scores of SPR	Scores of BCR		Scores of ECR		Total Score
			Step A	Step B	Step A	Step B				Step A	Step B	Step A	Step B	
A	33	12	5	5	3	3	61	33	12	5	10	3	9	72
F	34	12	4	4	3	3	60	34	12	4	8	3	9	70



## 1.5 Operational Test Form Construction Using the Rasch Model

The selection of items to be included in the final operational test forms of the 2009 MSA-Math required a careful consideration based on test blueprints, classical item analyses, *DIF* analyses, and IRT analyses. Specifically, the Rasch model (i.e., 1-Parameter Logistic IRT) played a major role in constructing the 2009 operational forms. First, Pearson suggested the following guidelines:

- Do not include items that are too easy or too hard.
- Do not include *BCR* items with score distributions that do not elicit the full range of rubric scores.
- Do not include items with *DIF* classifications “C” for the *SR* items and “CC” for the *BCR* items *unless* they have been deemed acceptable by the external review of content experts.
- Finally, do not include items which have Rasch *Infit* and *Outfit* mean-squares lower than .5 or higher than 1.5. More specific information on Rasch *Infit* and *Outfit* mean-squares can be found in the third part of the 2009 technical report, *Overview of Statistical Summaries*.

A procedure for using IRT methods to build tests that meet any desired set of test specifications was outlined by Lord (1977). The procedure utilizes an item bank with item parameter estimates available for the IRT model of choice, with accompanying information functions. The steps in the procedure suggested by Lord (1977) are as follows:

- First, the shape of desired test information needs to be decided. This was termed as the “target information function” by Lord (1977).
- Second, specific items need to be selected from the item bank with item information functions that will fill up hard-to-fill areas under the target information function.
- Third, the test information function after test items are added needs to be recalculated.
- Fourth, until the test information function approximates the target information function to a satisfactory degree, test items need to keep on being selected.

It should be noted that these steps were implemented within a framework defined by the content specification of the test. In addition, math content specialists from MSDE reviewed the final test forms of the 2009 MSA-Math. The following table and figure show an example of the 2009 MSA-Math operational test form construction using the Rasch (i.e., 1-PL IRT) method. Detailed information about constructing operational forms using the Rasch method can be obtained from either MSDE or Pearson.

**Table 1.24 The 2009 Math Operational Test Construction Using the Rasch Model: Grade 4 Form A**

Item Type	P-value	A	$D_{i1}$	$D_{i2}$
BCR_A	0.74	1.00	-0.0713	
BCR_B	0.73	1.00	-0.2255	0.4783
BCR_A	0.87	1.00	-1.1338	
BCR_B	0.45	1.00	-0.6131	4.3963
BCR_A	0.37	1.00	2.1316	
BCR_B	0.47	1.00	0.4623	2.7745
BCR_A	0.44	1.00	1.6468	
BCR_B	0.40	1.00	0.3771	3.7263
BCR_A	0.44	1.00	-0.2409	4.0341
BCR_B	0.82	1.00	-0.5527	
BCR_A	0.50	1.00	1.3917	
BCR_B	0.43	1.00	-0.5064	4.821
BCR_A	0.70	1.00	0.1552	
BCR_B	0.48	1.00	-0.7421	3.9335
SR	0.84	1.00	-0.7274	
SR	0.55	1.00	0.9747	
SR	0.94	1.00	-2.1097	
SR	0.80	1.00	-0.7990	
SR	0.83	1.00	-1.0550	
SR	0.68	1.00	-0.0118	
SR	0.79	1.00	-0.4437	
SR	0.70	1.00	-0.1077	
SR	0.66	1.00	0.1763	
SR	0.94	1.00	-2.7781	
SR	0.69	1.00	0.1230	
SR	0.79	1.00	-0.3698	
SR	0.76	1.00	-0.3619	
SR	0.65	1.00	0.0796	
SR	0.98	1.00	-3.5593	
SR	0.89	1.00	-1.2509	
SR	0.84	1.00	-0.7925	
SR	0.79	1.00	-0.3646	
SR	0.83	1.00	-0.7461	
SR	0.51	1.00	0.9291	
SR	0.53	1.00	1.1949	
SR	0.35	1.00	1.7570	
SR	0.50	1.00	0.6281	
SR	0.48	1.00	1.0327	
SR	0.50	1.00	0.9009	

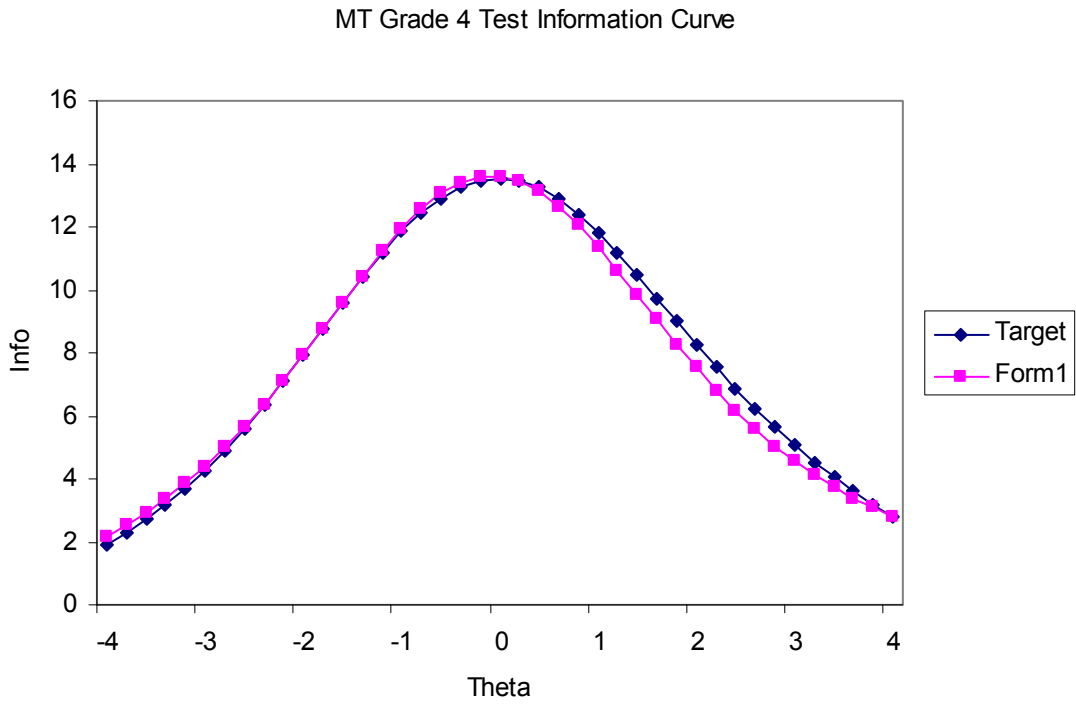
Table 1.24 (Continued)

Item Type	P-value	A	$D_{i1}$	$D_{i2}$
SR	0.71	1.00	0.1774	
SR	0.40	1.00	1.4979	
SR	0.85	1.00	-1.2169	
SR	0.60	1.00	0.3940	
SR	0.66	1.00	0.4883	
SR	0.80	1.00	-0.4685	
SR	0.75	1.00	-0.2435	
SR	0.55	1.00	0.6901	
SR	0.81	1.00	-0.8522	
SR	0.93	1.00	-2.3000	
SR	0.81	1.00	-0.5617	
SR	0.79	1.00	-0.2781	
SR	0.81	1.00	-0.8156	
SR	0.75	1.00	-0.4674	
SR	0.97	1.00	-2.8436	
SR	0.73	1.00	-0.1831	
SR	0.63	1.00	0.3118	
SR	0.84	1.00	-0.9014	
SR	0.68	1.00	0.3863	
SR	0.90	1.00	-1.3494	
SR	0.68	1.00	0.1618	
SR	0.51	1.00	1.2932	
SR	0.70	1.00	-0.1060	
SR	0.83	1.00	-0.9767	
SR	0.60	1.00	0.4796	

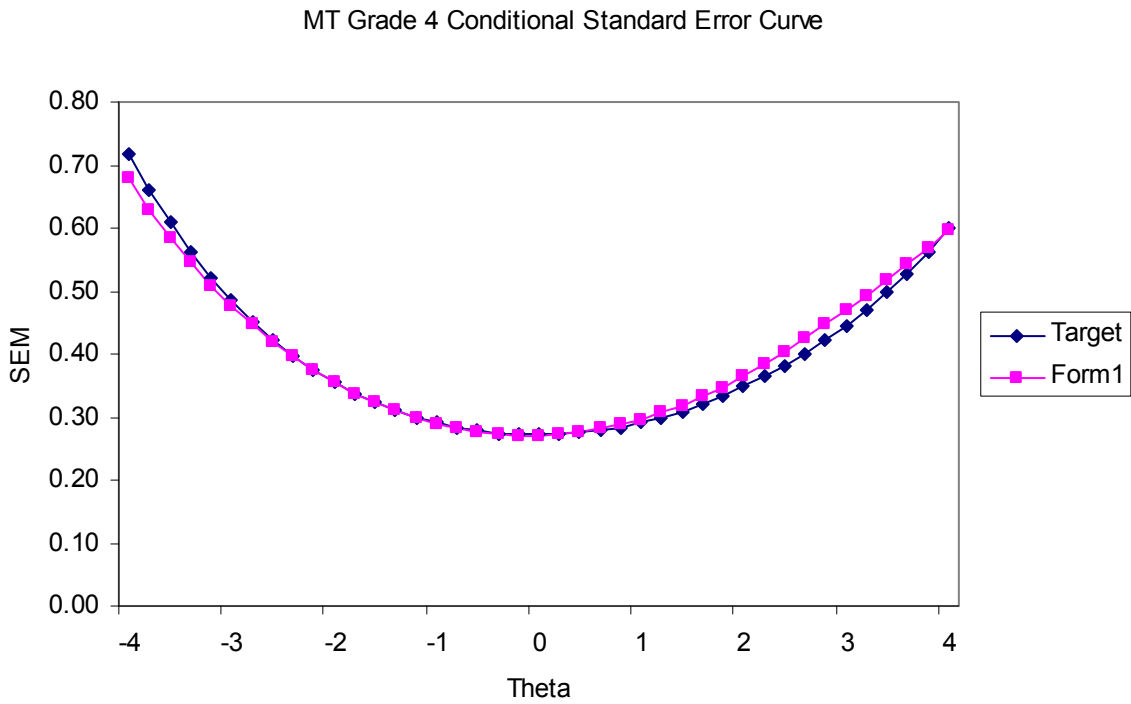
Note. A: item discrimination;  $D_{i1}$ : first structure measure estimate;  $D_{i2}$ : second structure measure estimate.

Note. Please refer to section 3.3 of this technical report to get detailed information about how to estimate structure measure estimate ( $D_{ij} = D_i + F_{ij}$ )

Note. BCR\_A: Step A item; BCR\_B: Step B item



**Figure 1.1 Test Information Curves of Target Form vs. Current Year’s Math Operational Test Form**



**Figure 1.2 Standard Errors of Target Form vs. Current Year’s Math Operational Test Form**

## 1.6 Test Administration of the 2009 MSA-Math

The 2009 MSA-Math test was administered to all students in grades 3 through 8 except for students taking the Alt-MSA-Math or the Mod-MSA-Math. Pearson coordinated the test administration procedures with MSDE prior to implementation. This chapter was prepared to provide general information about the 2009 test administration. Detailed information about the 2009 test administration can be obtained from the 2009 Test Administration and Coordination Manual (TACM) and Examiners Manual (EM) which are available from either MSDE or Pearson.

### 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. The Daily Testing Materials Tracking Record (or an equivalent form designed by the LEA) was used to track the distribution and return of Test Books.

Before testing began, the Test Examiners (TEs) carefully inventoried all test materials given to them, as they were accountable for the return of all secure materials at the end of testing. The TEs checked to ensure they have all the materials they needed for testing.

For the Test Examiner, Pearson provided the following materials:

- MSA Examiner's Manual for grades 3 through 8- Math
- Pre-printed and generic labels
- Scoring Service Identification (SSID) sheets

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

- 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
- Classroom Calculator for Day 1 (all grades)
- Classroom ruler with both U.S. customary and metric measurements (all grades)
- Classroom protractor for grades 5 through 8
- Classroom compass for grades 7 and 8 only

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

- A sign for the door that reads "Testing: Do not Disturb"
- A digital clock or a watch, or clock with a second hand

Two test-related Examiners Manuals (EMs) were developed for the 2009 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 Pearson for scoring. The TACM was distributed and reviewed during a workshop in January for STCs and LACs, with duplicates sent to each school along with its testing materials.

### **Test Administration Schedule**

The primary test window for MSA was established by MSDE (March 16-25, 2009, with make-up testing held March 26-31, 2009). However, each LEA (Local Education Agency) set a specific schedule for administration of the MSA within that window for their district. For a given test, grade, content area, and test format, all testing (with the exception of the make-up administration) had to take place on the same schedule. Each LEA schedule was submitted to MSDE in advance and approved for each district by the state. For example, all Grade 3 MSA-Math must be administered on the same days throughout the LEA. In addition, each content area in each grade was tested on two days during the window. In any given grade, one content area's primary testing window was completed before beginning the second content area's primary testing window.

The MSA-Math testing schedule allowed approximately 2 hours and 30 minutes for testing on Day 1 and 1 hour and 45 minutes on Day 2 (including preparation time and breaks).

For the 2009 MSA-Math, the primary testing days were as follows:

- Test materials delivered to schools (Examiner's Manuals, Test/Answer Books, and Test Coordinator's Kits) On or Before March 2, 2009
- Mathematics Primary Testing Window March 16 – March 25, 2009
- Make-up Testing Window March 26 – March 31, 2009

If a student was absent on the testing days, a make-up test was administered on any two consecutive days within the 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 by LACs to determine the testing schedule to be followed.

During the administration of the 2009 MSA-Math, MSDE had testing monitors in selected schools observing administration procedures and testing conditions. All monitors had identification cards for security purposes. There was 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**

MSDE calculates actual participation of students who took the test. This means that the schools are held accountable not only for student achievement on MSA or Mod-MSA testing, but also they are accountable to ensure that at least 95% of students participate in testing. Accordingly, schools should do all they can to test all students on MSA, Mod-MSA, or Alt-MSA, as applicable.

All students in grades 3 through 5 had to participate in the 2009 MSA-Math, and all students in grades 6 through 8 had to participate in either the 2009 MSA-Math or Mod-MSA-Math. All students in grade 6 through 8 had to participate in the 2009 Mod-MSA-Math, if determined to be eligible by the student's IEP. 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 or Mod-MSA-Math. The criteria that students need in order to be tested in the Alt-MSA program instead of the MSA-Math can be viewed in section 5, Appendix A of the TACM.

## **Participation of English Language Learners (ELLs) in the MSA-Math or the Mod-MSA-Math**

There are special rules that apply to the participation of English Language Learners (ELLs) in the MSA-Math and the Mod-MSA-Math, as follows:

For the MSA-Math and Mod-MSA-Math, ELL students must participate in MSA-Math or Mod-MSA-Math regardless of how recently they entered the U.S. educational system. For ELL students in their first year of enrollment in a U.S. school, "participation" in the MSA-Math or the Mod-MSA-Math is defined as allowing the student to attempt the test for at least 20 minutes. If, after 20 minutes, the TE determines in his or her professional judgment that the student does not possess sufficient English fluency to be able to continue testing, the test administration for that student may be concluded at that time.

## **Accommodations for Assessment**

Accommodations for assessment of 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 could 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 2009 TACM for further information regarding testing accommodations.

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

The 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 a verbatim reading

accommodation. 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 student's name, LEA number, and school number were written on the large-print Test/Answer Book for proper transcription into the the standard-size Test/Answer Book.

The pre-printed student ID label was affixed to the standard-size Test/Answer Book containing the transcribed responses, and not to the large-print Test/Answer Book or Braille books. The bubbles on the demographic page of the standard-size Test/Answer Book were not filled in if there was a pre-printed student ID label for the student.

A certified Test Examiner (TE) transcribed the student responses into a standard-size Test/Answer Book exactly as given by the student. The standard-size Test/Answer Book with the pre-printed or general label attached was returned to Pearson with all other Test/Answer Books.

Large-Print Test/Answer Books and Braille Test/Answer Books containing the original student responses prior to transcription are to be returned with Non-Scorable materials. Any Test/Answer Books which were used as source documents for transcription were 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 2009 TACM in sections 2 and 3.

### **Verbatim Reading Accommodation and Kurzweil™ Test Form on CD**

Students who had a verbatim reading accommodation documented in their Individual Education Plan (IEP), ELL Plan, or Section 504 Plan, and who received that accommodation in regular instruction, received the accommodation on the 2009 MSA-Math. The accommodation was provided by a live reader or through technology. Appendix L of the 2009 TACM provided information on verbatim reading instruction. Technology used to provide the verbatim reading accommodation was Kurzweil™ reading software. Official, secure electronic copies of the test were ordered through the LAC. MSDE encouraged (but did not require) the use of the Kurzweil™ software to ensure uniformity in the delivery of the verbatim reading accommodation throughout the state.

Students using Kurzweil™ software had to familiarize themselves with its operation prior to the test administration. When there were technical difficulties with Kurzweil™ a certified staff member was used instead. Kurzweil™ Test Form CDs were shipped by Pearson. After testing, schools returned the CDs to Pearson with the non-scorable secure materials.

### **Administration Procedures for Students with IEP, 504 Plan, or ELL Plan Permitting a Dictated Responses or Use of Word Processor**

A student whose IEP, 504 Plan, or ELL Plan permitted a dictated response had his/her responses transcribed at the school level by an eligible TE, or by a staff member working under the direct supervision of a certified TE, into the student's Test/Answer Book with a pre-printed or generic ID label attached.

A student whose IEP, 504 Plan, or ELL plan permitted the use of a word processor had his/her responses transcribed by hand or under the direct supervision of an eligible TE or STC exactly as the student entered his/her responses on the word processor. The student's responses were always



transcribed at the school level into the student's Test/Answer Book with the pre-printed or generic ID label attached. After the student's responses had been transcribed, the memory of the word processor was cleared. The original word-processed print-out was returned to Pearson with the non-scorable materials.

### **Test Format**

All grade levels of the MSA-Math used a Test Book format in which students wrote their answers directly in the Test Book. 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 did not interfere with the bubbled answer choice area and/or the track marks along the outside margins of each page.

### **Security of Test Materials**

The following code of ethics 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 (Pearson, 2009):

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 (LEA) or MSDE disciplinary action. (p. 11)

The Test/Answer Books for the 2009 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 (Person, 2009):

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. 11)

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

## 1.7 Hand Scoring Procedures of the 2009 MSA-Math

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

### Hand Scoring Staff

The PSC Project Manager (PSC PM), Content Specialist (CS), and Scoring Directors (SD) participated in the rangefinding sessions in Maryland. (Detailed information about rangefinding procedures can be found in the following portion of this section: *Development Procedures for Rangefinding*.) The SD 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 SD responsible for training the Scoring Supervisors and Scorers for the respective Maryland prompts.

#### 1) Scorer

A graduate of a four-year accredited college or university who had completed the Maryland-specific domain training. The scorers were eligible to score items for which they had been trained and successfully qualified.

#### 2) Scoring Supervisor

A reader who directly monitored the scoring of a team of Scorers and retrained as needed. The reader had successfully completed the PSC Scoring Supervisor training.

#### 3) Scoring Director (SD)

An experienced and knowledgeable PSC team leader who was responsible for selecting a wide variety of student responses for such activities as rangefinding and building training materials. Selected papers were then submitted to MSDE for comment and approval. Scoring directors remained on the project as rangefinding participants and trainers. Scoring directors worked with scoring supervisors and the Content Specialist to oversee the scoring of several items. An SD's main duty during scoring was to rule on validity of questionable papers and to maintain consistency in scoring decisions.

#### 4) Content Specialist (CS)

Experienced content/training personnel who had served as SDs and were selected by the Scoring Resources staff and Project Manager to train and support Scoring Directors for Maryland.

### Scorer Recruitment and Qualifications

All Scorers for MSDE had to provide Pearson their résumé and documentation of a four-year college degree. Human Resources made every effort to recruit Scorers with a teaching background and to match Scorers to projects which suited their educational background and previous scoring experience. Regardless of educational background, applicants then participated in a one-day general introductory training workshop presented by a PSC staff member. These workshops allowed Pearson to introduce potential Scorers and Scoring Supervisors to large-scale scoring in general and to the Maryland rubric specifically. The PSC staff member who presented the workshop evaluated potential Supervisors and submitted these evaluations to the PSC Site Manager with his/her recommendations. Those who successfully completed the workshop were

added to Pearson's general pool of potential Scorers and Supervisors of MSA Math. This addition to the scoring pool did not qualify these Scorers for scoring the MSDE program.

### **Scoring Supervisor Selection**

The training for new Scoring Supervisors consisted of a two-day course focusing on the duties and responsibilities necessary to successfully manage a team of Scorers. The workshop was led by the PSC Site Manager and Scoring Directors. The instruction included a review of PSC policies and procedures, sessions on use of ePEN and the monitoring reports to track a Scorer's speed and accuracy, role playing activities which explored various situations that could occur with Scorers during the scoring of a project, and Scorer counseling and retraining guidelines. Upon completion of the workshop, the PSC Site Manager and Scoring Directors in conjunction with the Content Specialist reviewed each participant's performance, making sure that each had a complete understanding of the Scoring Supervisor role and its responsibilities. Any participant they found who did not perform to their satisfaction was not added to the qualified Supervisor list.

### **Scoring Supervisor Project Training and Qualification**

Project-specific Supervisor training for MSDE was conducted in the days immediately preceding Scorer training. This training began with the SD reading the rubrics aloud and answering any questions the Supervisor might have regarding the rubric. The SD then read each anchor paper aloud to the Supervisors. Each response in the anchor set was thoroughly explained, including the notes and comments of the rangefinding committee. Practice Set 1 was reviewed next. The Supervisors scored the practice set individually in the electronic scoring system (ePEN) as well as recorded their scores on a paper copy of the practice set, and then waited for all Supervisors to complete scoring the set. When everyone had completed scoring the training set, the SD discussed the responses one by one, focusing on why each received that score and not another. The SD reviewed with the group the reason for assigning each score point and discussed each paper in its entirety. The Supervisors were then ready to score Practice Set 2. Practice Set 2 was scored and reviewed exactly as Practice Set 1.

Having thoroughly discussed both practice sets with the group, the SD explained that in order for a participant to qualify as a Scoring Supervisor, it was required that the Supervisor should score at least 90% perfect agreement on Step A and 80% perfect agreement on Step B on two of three qualifying sets or one of two qualifying sets, depending on the number of sets available for each item (Qualification Rules, Attachment M). The Supervisors scored the first qualifying set individually and recorded their scores in ePEN. As each Supervisor finished scoring, the SD reviewed the qualifying reports before allowing the Supervisor to proceed to the next qualifying set. Each response was reviewed and any questions the Supervisor had were addressed before the Supervisor attempted the next qualifying set. The Supervisor followed the same procedure with Qualifying set 2 (and set 3 if available). Supervisors had to pass one of two or two of three sets (depending on the number of qualifying sets available per item) with 80% agreement 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.

### **Scoring Supervisor Duties**

Scoring Supervisors were responsible for monitoring the training and qualifying of the Scorers assigned to their team. The Supervisors assisted the SD, if requested, during the training of the Scorers. The Supervisor was responsible for monitoring Scorers' progress through the qualifying sets. The Supervisor was also responsible for monitoring each Scorer's assignment of scores to

the responses. Additionally, the Supervisor reviewed the statistical reports with each individual on the team. The Supervisor consulted the SD regarding variations by the team members from the acceptable standards (95% for Math Step A, and 85% for Math Step B). The Supervisor had the initial responsibility to see that the Scorer maintained the set standards through individual retraining. The SD monitored the Supervisor by reviewing team statistics and working one-on-one with the Supervisor.

### **Scoring Director Selection and Qualification**

The candidates for Scoring Director had been recommended by the Content Specialist, PSC Resource Staffing Managers or Site Manager. The recommendations were based upon the evaluations the candidates received as Scorers and Supervisors and were part of their personnel file. The candidates generally had been Supervisors on large-scale projects for multiple teams, and/or they had served as Supervisors on small-scale projects where Supervisors trained their individual teams. They had been evaluated on their ability to train Scorers as well as their ability to monitor the scoring accuracy and consistency of Scorers. These evaluations were submitted in writing at the end of each scoring project by the Site Managers and SDs that had observed the work of the SD candidates.

### **Scoring Director Project Training**

The SDs familiarized themselves with the rubric. Any questions regarding the rubric were addressed by the PSC Content Specialist or MSDE. The next step was for the SD to become familiar with all their items and all training materials and scoring decisions/issues associated with their items prior to Supervisor training.

### **Scoring Director Duties**

The SD's job was to conduct the training of the Supervisors and Scorers, oversee the actual scoring of the papers, monitor the work of the Supervisor, and act as the decision-maker for situations or questions that may arise during the scoring process. For example, all condition code (foreign language, off-topic, off-mode, etc.) responses were reviewed by the SD, who had to confirm any such decision and ensure consistency of decisions. (Blank condition codes were assigned at the Scorer level and did not require SD confirmation.) Additionally the SD and Supervisor conducted all resolution readings. The resolution score became the reported score.

The SD also reviewed any potential questionable content responses and forwarded those to the Content Specialist to consult with MSDE before processing.

The SD was also responsible for daily statistical review and analysis of all monitoring reports to ensure the quality of the scoring. Review of the data allowed the SD not only to monitor the Scorer but also to provide the Supervisor with additional input. Available data included 1) individual Scorer agreement rates between two independent scorings; 2) score point distributions by Scorer and trend review; 3) prompt statistics for agreement rates and score point distributions; 4) Resolution data; 5) scorer-level and item-level agreement on validity papers pre-scored by MSDE.

### **Scorer Training**

Scorer training was led by the SD, and each SD was responsible for training the items he/she monitored throughout scoring. After sufficient student responses were scored for equating purposes for the first item, the SD 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 item, 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 Scorers' understanding of how the responses differed from one another in incremental quality, 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 Scorers had all their questions answered and the discussion of the anchor set was finished, the Scorers began to assign scores to the first practice set. Each Scorer independently read and scored the responses in the practice set in the electronic scoring system (ePEN). The correct scores were then read to the group when everyone had completed the scoring. In addition, each practice paper was discussed as to reasons for applying each given score. At this point, Scorers interacted with the SD in discussing the characteristics of each response that earned the assigned score point. The same format was followed for each practice set. During this process, the job of the Scorer was to internalize the scoring scale and adjust his or her individual scoring to conform to that scale. Once all practice papers had been scored and fully discussed, Scorers began the qualifying process.

For MSA Math, there were two or three qualifying sets, depending on the particular item. MSDE informed PSC in writing for each specific administration how many qualifying sets were approved and were available to the Scorers. Scorers had to achieve at least 90% perfect agreement on Step A and 80% perfect agreement on Step B on two of three qualifying sets or one of two qualifying sets, depending on the number of sets available for each item.

### Scoring Rules for MSA-Math

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
- Scores given were the higher of the 1st and 2nd Scorer's scores provided they were adjacent.
- For example:

1 <sup>st</sup> Scorer	2 <sup>nd</sup> Scorer	Final Score
1	2	2
2	3	3

- A resolution scorer was used if two non-adjacent initial scores were received.

- The resolution scorer's score was used in place of both the 1st and 2nd Scorers' scores.
- For example:

1 <sup>st</sup> Scorer	2 <sup>nd</sup> Scorer	Resolution Scorer	Final Score
0	2	1	1
0	3	2	2
1	3	3	3
2	0	1	1
3	0	2	2

### Inter-Rater Agreement

Pearson's scoring system generated many kinds of internal monitoring reports that enabled the project leadership to monitor the accuracy and consistency of scoring. These reports were compiled by prompt, listed the entire prompt's Scorers, and provided the results of their scoring for each day. Information on these reports included the number of responses read by the Scorers during the period, the number and percent of condition code 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 Scorer was a point lower than the second Scorer; the number and percent of responses on which the first Scorer was a point higher than the second Scorer (Adjacent); and the number and percent of responses differing by more than one score point (Non-Adjacent). The Scoring Director 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*.

### Scorer Retraining

When a Scorer's performance fell below acceptable parameters for a project, the Scorer was retrained. Retraining was the process by which the SD or Supervisor 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 Scorer back on track with the guidelines provided by a specific program. Group retraining was conducted by the SD every Monday (or following any extended break) during the scoring project. In addition, daily retraining occurred as deemed necessary by the MSDE representative and CS.

### Backreading

Pearson's ePEN system allowed Supervisors and/or SDs to conduct backreads as an additional monitoring method. When conducting backreads, the Supervisor or SD received images of student responses and the scores assigned by the Scorer. Responses selected for backreads might be randomly selected or might be targeted backreads (e.g., responses receiving specific scores, etc.). These backreads were very useful in tracking specific areas of confusion for a given Scorer or group of Scorers and assisted the Supervisor and SD in knowing just how to direct retraining activities for individual Scorers or teams. The initial backreading percentage was set at 3%. This

percentage might be adjusted either higher or lower by the Supervisor based upon the performance of the Scorer.

### **Development Procedures for Rangefinding**

Scoring Directors were selected by the PSC Scoring Resource Manager and Content Specialist to prepare sets of papers for client approval. These experienced SDs were judged by the CS for their ability to recognize and assemble a wide variety of responses. The SD also participated with the clients as a facilitator during the rangefinding session in order to make notes and be prepared to assemble the finished sets to the client's specifications. For a given math prompt, the SD had the following responsibilities:

- 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, or Low.
  - Marked High, Medium, and Low papers—marked especially good ones that might potentially receive top scores.
- 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 rangefinding
  - Decided which particular papers from the sorted piles should go into sets for rangefinding. Each paper selected went into a rangefinding set arranged in performance from low to high performance.

### **Rangefinding Procedures**

The objective of rangefinding sessions 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 PSC Project Manager, Content Specialists, and Scoring Directors, 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, Practice 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 rangefinding, the procedure for assigning scores to the papers in each set was as follows:

- The item was reviewed by the committee and criteria were discussed for receiving full credit.
- Selected “grounding” papers that represented the full range of scores were read aloud and discussed by the rangefinding 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.
- All subsequent responses were read and scored by each panel member independently, 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 after each committee member had already recorded tentative scoring decisions. There might be frequent reference to previous responses to make sure that decisions on score points were consistent.

This iterative process of reading, charting, and discussing successive responses had three results:

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

During this process, the tentative scores assigned to earlier responses became firm.



## 1.8 The 2009 Operational Item Analyses

### Classical Analysis with Common Items Used for Form-to-Form Linking

As mentioned in chapter 1.4, two operational forms were randomly distributed to students and linked using common items appearing on both forms (i.e., operational forms A and F). As a result, classical analysis of these common items was conducted to check if the two groups taking different operational forms were equivalent. The following descriptive statistics were calculated based on a raw, number-right score of the common items: mean ( $M$ ) and standard deviation ( $SD$ ). The results indicated that the students taking the two operational forms were statistically close and equivalent across all grades, as seen from Table 1.25.

**Table 1.25 Descriptive Statistics of Form-to-Form Common Items**

Grade	Form	No. of Items	$N$	$M$	$SD$
3	A	38	30,174	31.11	6.51
	F	38	29,789	31.15	6.36
4	A	37	29,532	28.48	7.15
	F	37	29,476	28.66	7.00
5	A	37	30,344	28.60	7.50
	F	37	30,103	28.91	7.33
6	A	31	29,789	21.03	7.06
	F	31	29,240	20.96	6.98
7	A	40	30,318	26.91	9.57
	F	40	29,596	27.21	9.54
8	A	28	30,760	16.47	7.23
	F	28	30,282	16.65	7.25

*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.* Analysis was conducted with a statewide population.

### P-Value Check with Year-to-Year Core Linking Items

As mentioned in chapter 1.4, different years' assessments were linked using core linking items. This section was prepared to provide information about how much p-values (i.e., classical item difficulty) of the 2009 core linking items varied from previous years.

First, only SR items were used for the purpose of year-to-year linking. Second, classical analysis (e.g., p-value) on these items was conducted with a statewide population, and item sequence numbers on the tables were assigned based on the 2009 assessment. Finally it should be noted that detailed information about Rasch analysis on these core linking items can be found in chapter 1.9, *Linking, Equating, Scaling Procedures of the 2009 MSA-Math*.

As seen in Tables 1.26 through 1.37, we could conclude that most of the 2009 p-values were almost the same or slightly increased compared to those of previous years across all grades.

**Table 1.26 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 3 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
1	3509931	0.69	0.70	49	3509961	0.91	0.93
2	3548055	0.93	0.98	50	10000044158	0.86	0.87
5	3510009	0.84	0.86	51	3510018	0.78	0.80
7	3548054	0.93	0.96	52	3510035	0.88	0.89
14	3510017	0.91	0.93	55	3510055	0.62	0.62
15	3510006	0.59	0.58	56	3510058	0.88	0.90
16	3509960	0.78	0.80	62	3510347	0.74	0.72
17	3509964	0.79	0.78	63	3510053	0.84	0.85
21	3509983	0.91	0.94	64	3510041	0.92	0.94
22	3510022	0.51	0.56	65	3510051	0.57	0.57
23	3509927	0.80	0.79	66	3509929	0.53	0.56
29	3510062	0.85	0.85	67	3510329	0.55	0.66
32	3509988	0.73	0.72	68	3510033	0.82	0.85
33	3510070	0.97	0.98	69	3510043	0.77	0.77
41	3510063	0.78	0.80	72	3509962	0.90	0.91
45	3509926	0.39	0.51	82	3510036	0.85	0.87
46	3548507	0.85	0.88				
48	3510065	0.94	0.95				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

### Descriptive Statistics of Year-to-Year Core Linking Items: Grade 3 Form A

Form	Year	No. of Items	<i>M</i>	<i>SD</i>
A	Previous Year	34	0.78	0.15
	Year 2009	34	0.80	0.14

**Table 1.27 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 3 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
1	3509931	0.69	0.70	49	3509961	0.91	0.93
2	3548055	0.93	0.98	51	3510018	0.78	0.81
5	3510009	0.84	0.86	52	3510035	0.88	0.89
7	3548054	0.93	0.96	55	3510055	0.62	0.63
8	3509979	0.84	0.89	56	3510058	0.88	0.90
14	3510017	0.91	0.92	62	3510347	0.74	0.73
15	3510006	0.59	0.58	63	3510053	0.84	0.87
16	3509960	0.78	0.80	64	3510041	0.92	0.95
17	3509964	0.79	0.78	65	3510051	0.57	0.56
21	3509983	0.91	0.94	66	3509929	0.53	0.56
22	3510022	0.51	0.57	67	3510329	0.55	0.64
23	3509927	0.80	0.73	68	3510033	0.82	0.86
24	3509928	0.88	0.95	69	3510043	0.77	0.77
32	3509988	0.73	0.74	70	3510012	0.80	0.80
33	3510070	0.97	0.98	72	3509962	0.90	0.92
41	3510063	0.78	0.80	80	3509950	0.72	0.78
45	3509926	0.39	0.48	82	3510036	0.85	0.87
47	10000044163	0.76	0.78				
48	3510065	0.94	0.95				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 3 Form F**

Form	Year	No. of Items	<i>M</i>	<i>SD</i>
F	Previous Year	36	0.78	0.14
	Year 2009	36	0.80	0.14

**Table 1.28 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 4 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
1	3488052	0.61	0.65	47	3515575	0.71	0.90
2	3515407	0.86	0.87	49	3515471	0.86	0.86
3	100000044146	0.91	0.92	50	3515630	0.52	0.56
6	3515408	0.76	0.79	54	3515533	0.84	0.84
7	3515641	0.79	0.81	55	3515631	0.78	0.79
8	3515410	0.87	0.89	62	100000201857	0.45	0.48
10	3515605	0.61	0.63	63	3515543	0.80	0.80
18	3488159	0.87	0.89	64	3515853	0.80	0.79
19	3515447	0.52	0.53	65	3497869	0.81	0.82
22	3515604	0.69	0.69	66	3548078	0.50	0.51
23	3515737	0.83	0.83	67	3515933	0.76	0.79
24	3515576	0.65	0.67	68	3515519	0.86	0.86
25	3515470	0.73	0.73	69	3515795	0.65	0.64
26	3515643	0.42	0.47	71	3548086	0.81	0.80
32	3515571	0.80	0.84	78	3515506	0.90	0.90
33	100000044144	0.94	0.96	80	3515632	0.69	0.70
34	3515421	0.85	0.85	81	3548088	0.75	0.77

Note. Analysis was conducted with a statewide population.

Note. Item sequence numbers were assigned based on the 2009 assessment.

#### Descriptive Statistics of Year-to-Year Core Linking Items: Grade 4 Form A

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
A	Previous Year	34	0.74	0.14
	Year 2009	34	0.76	0.13

**Table 1.29 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 4 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
2	3515407	0.86	0.87	36	3548767	0.71	0.70
3	10000044146	0.91	0.92	47	3515575	0.71	0.91
6	3515408	0.76	0.79	49	3515471	0.86	0.86
7	3515641	0.79	0.82	50	3515630	0.52	0.57
8	3515410	0.87	0.89	54	3515533	0.84	0.85
10	3515605	0.61	0.63	55	3515631	0.78	0.78
11	3488056	0.51	0.51	62	100000201857	0.45	0.45
18	3488159	0.87	0.88	63	3515543	0.80	0.81
19	3515447	0.52	0.54	64	3515853	0.80	0.82
22	3515604	0.69	0.69	65	3515785	0.71	0.78
23	3515737	0.83	0.84	66	3548078	0.50	0.54
25	3515470	0.73	0.73	67	3515933	0.76	0.80
26	3515643	0.42	0.47	68	3515519	0.86	0.87
27	3497882	0.77	0.79	69	3515795	0.65	0.64
31	100000201937	0.82	0.91	71	3548086	0.81	0.81
32	3515571	0.80	0.85	78	3515506	0.90	0.91
33	10000044144	0.94	0.96	80	3515632	0.69	0.69
34	3515421	0.85	0.86	81	3548088	0.75	0.79

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 4 Form F**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
F	Previous Year	36	0.74	0.14
	Year 2009	36	0.76	0.14

**Table 1.30 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 5 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
2	3511269	0.88	0.88	43	3511513	0.85	0.86
8	3511203	0.86	0.92	44	3488272	0.56	0.53
16	3511196	0.58	0.57	47	3511266	0.70	0.70
18	3488373	0.66	0.68	48	3488431	0.74	0.74
19	3511467	0.82	0.82	49	3511470	0.86	0.85
20	3512529	0.58	0.58	55	3512595	0.80	0.78
21	3511339	0.66	0.65	56	3488241	0.91	0.92
23	100000043853	0.67	0.70	58	1000000438	0.82	0.81
26	3511216	0.71	0.69	60	3511396	0.88	0.85
27	3512638	0.74	0.70	61	3511429	0.77	0.78
28	3511499	0.63	0.63	64	3511626	0.81	0.86
34	3488506	0.40	0.40	70	3511631	0.78	0.79
37	3488324	0.75	0.78	71	3488251	0.61	0.64
38	3511246	0.76	0.79	72	3511439	0.77	0.78
39	3511458	0.87	0.90	82	3488328	0.71	0.71
42	3511566	0.66	0.68	83	3511448	0.77	0.76

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 5 Form A**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
A	Previous Year	32	0.74	0.11
	Year 2009	32	0.74	0.12

**Table 1.31 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 5 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
2	3511269	0.88	0.89	44	100000209182	0.21	0.20
8	3511203	0.86	0.94	47	3511266	0.70	0.70
16	3511196	0.58	0.57	48	3488431	0.74	0.75
18	3488373	0.66	0.69	49	3511470	0.86	0.86
19	3511467	0.82	0.82	55	3512595	0.80	0.79
20	3512529	0.58	0.60	56	3488241	0.91	0.92
21	3511339	0.66	0.66	58	100000043857	0.82	0.83
23	100000043853	0.67	0.70	60	3511396	0.88	0.86
26	3511216	0.71	0.64	61	3511429	0.77	0.77
27	3512638	0.74	0.76	64	3511626	0.81	0.88
37	3488324	0.75	0.78	70	3511631	0.78	0.80
38	3511246	0.76	0.77	71	3488251	0.61	0.64
39	3511458	0.87	0.93	72	3511439	0.77	0.79
40	3512616	0.44	0.48	82	3488328	0.71	0.72
42	3511566	0.66	0.67	83	3511448	0.77	0.77
43	3511513	0.85	0.87				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

#### Descriptive Statistics of Year-to-Year Core Linking Items: Grade 5 Form F

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
F	Previous Year	31	0.73	0.14
	Year 2009	31	0.74	0.15

**Table 1.32 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 6 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
1	3516257	0.88	0.90	34	3516331	0.49	0.52
2	100000078832	0.55	0.82	35	3516241	0.84	0.86
3	3516291	0.53	0.54	36	3516247	0.60	0.63
6	3516243	0.72	0.73	37	3516329	0.60	0.71
9	3516248	0.74	0.85	38	3516355	0.70	0.71
10	3516559	0.91	0.92	45	3492095	0.80	0.79
11	3516255	0.77	0.77	50	3516929	0.65	0.72
12	3516258	0.61	0.64	54	3516906	0.60	0.60
19	3516240	0.64	0.67	55	3516332	0.52	0.53
20	3516909	0.59	0.58	56	3516256	0.61	0.63
25	3516351	0.52	0.50	57	3516302	0.69	0.71
26	3516290	0.75	0.69	61	3516375	0.61	0.63
27	100000043862	0.61	0.62	68	3516613	0.54	0.57
29	3517010	0.48	0.54	80	3516303	0.55	0.55
30	100000043865	0.53	0.55				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 6 Form A**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
A	Previous Year	29	0.64	0.12
	Year 2009	29	0.67	0.12



**Table 1.33 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 6 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
1	3516257	0.88	0.89	34	3516331	0.49	0.52
2	3488263	0.79	0.87	35	3516241	0.84	0.86
3	3516291	0.53	0.55	36	3516247	0.60	0.63
4	3492143	0.77	0.79	37	3516329	0.60	0.65
6	3516243	0.72	0.74	38	3516355	0.70	0.71
9	3516248	0.74	0.84	45	3492095	0.80	0.78
10	3516559	0.91	0.92	50	3516929	0.65	0.70
11	3516255	0.77	0.78	54	3516906	0.60	0.66
12	3516258	0.61	0.65	55	3516332	0.52	0.53
19	3516240	0.64	0.66	56	3516256	0.61	0.63
20	3516909	0.59	0.58	57	3516302	0.69	0.71
25	3516351	0.52	0.50	61	3516375	0.61	0.62
26	3516290	0.75	0.69	68	3516613	0.54	0.56
27	100000043862	0.61	0.61	80	3516303	0.55	0.56
30	100000043865	0.53	0.55				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

#### Descriptive Statistics of Year-to-Year Core Linking Items: Grade 6 Form F

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
F	Previous Year	29	0.66	0.12
	Year 2009	29	0.68	0.12

**Table 1.34 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 7 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
1	3517604	0.34	0.36	50	100000433	0.34	0.35
2	3517601	0.51	0.55	51	3517687	0.57	0.61
3	3517609	0.58	0.60	52	3517692	0.83	0.84
7	3517616	0.63	0.67	63	3517712	0.46	0.47
8	3517634	0.67	0.71	64	3517714	0.55	0.58
10	3517638	0.77	0.78	65	3517716	0.68	0.69
12	3517650	0.66	0.69	66	3517718	0.70	0.72
18	3517652	0.69	0.74	69	3517721	0.52	0.53
19	3547473	0.80	0.83	70	3517691	0.61	0.66
20	3517663	0.32	0.35	72	3555858	0.45	0.46
30	3517667	0.53	0.60	79	3555859	0.74	0.77
31	3517678	0.92	0.94	80	3517752	0.64	0.66
32	3517742	0.59	0.60	81	3488830	0.58	0.60
42	10000043349	0.36	0.34				
43	3517656	0.65	0.67				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

#### Descriptive Statistics of Year-to-Year Core Linking Items: Grade 7 Form A

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
A	Previous Year	28	0.60	0.15
	Year 2009	28	0.62	0.16

**Table 1.35 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 7 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
1	3517604	0.34	0.36	44	3491634	0.29	0.32
2	3517601	0.51	0.56	50	10000043338	0.34	0.34
3	3517609	0.58	0.61	51	3517687	0.57	0.61
7	3517616	0.63	0.67	52	3517692	0.83	0.84
8	3517634	0.67	0.72	63	3517712	0.46	0.50
10	3517638	0.77	0.78	64	3517714	0.55	0.62
12	3517650	0.66	0.70	65	3517716	0.68	0.69
18	3517652	0.69	0.77	66	3517718	0.70	0.73
19	3547473	0.80	0.84	69	3517721	0.52	0.55
20	3517663	0.32	0.35	70	3517691	0.61	0.69
30	3517667	0.53	0.56	72	3555858	0.45	0.47
31	3517678	0.92	0.95	79	3555859	0.74	0.77
32	3517742	0.59	0.60	80	3517752	0.64	0.67
42	10000043349	0.36	0.33	81	3488830	0.58	0.57
43	3517656	0.65	0.67				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 7 Form F**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
F	Previous Year	29	0.59	0.16
	Year 2009	29	0.62	0.17

**Table 1.36 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 8 Form A**

Item Seq. No.	Item CID	Previous Year	Y09 FA	Item Seq. No.	Item CID	Previous Year	Y09 FA
1	3514015	0.28	0.27	47	3514052	0.53	0.53
2	3514014	0.57	0.57	48	3487539	0.63	0.64
5	3514016	0.78	0.81	49	100000043311	0.36	0.37
7	3514053	0.73	0.77	50	3487525	0.50	0.53
8	100000043330	0.45	0.47	51	3487540	0.65	0.67
14	3500150	0.47	0.48	52	3514074	0.42	0.46
22	3514595	0.68	0.71	53	3514075	0.65	0.67
27	100000043320	0.47	0.46	62	3514095	0.31	0.33
32	3514058	0.33	0.35	63	3487568	0.19	0.20
33	3514062	0.43	0.45	65	3514103	0.68	0.70
38	3514156	0.73	0.77	66	100000043304	0.28	0.32
41	100000043323	0.49	0.51	78	3487912	0.53	0.53
42	3514291	0.75	0.81	79	3514710	0.54	0.56
46	3514055	0.56	0.59				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 8 Form A**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
A	Previous Year	27	0.52	0.16
	Year 2009	27	0.54	0.17

**Table 1.37 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2009: Grade 8 Form F**

Item Seq. No.	Item CID	Previous Year	Y09 FF	Item Seq. No.	Item CID	Previous Year	Y09 FF
1	3514015	0.28	0.28	47	3514052	0.53	0.54
2	3514014	0.57	0.58	48	3487539	0.63	0.65
5	3514016	0.78	0.82	50	3487525	0.50	0.54
7	3514053	0.73	0.78	51	3487540	0.65	0.68
8	100000043330	0.45	0.49	52	3514074	0.42	0.48
14	3500150	0.47	0.49	53	3514075	0.65	0.67
32	3514058	0.33	0.35	62	3514095	0.31	0.34
33	3514062	0.43	0.46	65	3514103	0.68	0.72
38	3514156	0.73	0.74	73	3492047	0.34	0.30
41	100000043323	0.49	0.52	78	3487912	0.53	0.56
42	3514291	0.75	0.80	79	3514710	0.54	0.56
46	3514055	0.56	0.58				

*Note.* Analysis was conducted with a statewide population.

*Note.* Item sequence numbers were assigned based on the 2009 assessment.

**Descriptive Statistics of Year-to-Year Core Linking Items: Grade 8 Form F**

Form	Year	<i>N</i>	<i>M</i>	<i>SD</i>
F	Previous Year	23	0.54	0.15
	Year 2009	23	0.56	0.16

### **Validation Check with the 2009 MSA-Math Core Items**

As mentioned in chapter 1.4, operational items fell into one of two categories: core and core linking items. Because the core items were not included into the 2009 year-to-year linking pool, Rasch item and step difficulty parameters of the core items were reestimated with the 2009 stratified random samples during calibration and equating. (Please see section 1.9 and Appendix A for stratified random sampling procedures) As a result, this section was prepared to provide detailed information about how much the core items changed in terms of item difficulty, both classical item p-value and Rasch item difficulty. Detailed information about the roles of the 2009 core and core linking items can be found in section 1.4, *Test Form Design, Specifications, Item Type, and Item Roles*.

A smaller number of cases (approximately 2,500) in the table indicates that it is a field-test item. P-values of both BCR and ECR items were calculated by dividing the item mean score by the item score range (i.e., score point 2 for BCR and 3 for ECR). The percentage of “Omits” for each CR item was low and indicated that a small number of students did not respond at all. In general, item p-value analysis results indicated that most of the 2009 p-values were almost the same or somewhat increased compared to those in previous years across all grades.

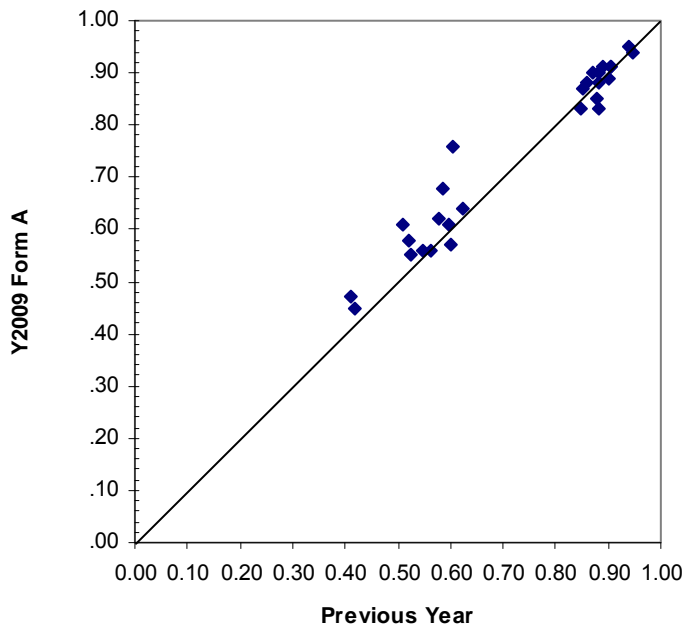
With respect to the Rasch item calibration and equating, it should be noted that we coded “Omit” of each item as “missing” before we ran the data with the Rasch model. In general, the level of the 2009 item difficulties stayed almost the same or became a little lower compared to that of previous years across all grades. It should be noted that all of the Rasch item and step difficulty parameters were on a common scale (i.e., linked to the 2006 assessment).

In conclusion, both p-value and Rasch item difficulty results reflected the same phenomenon, indicating that the level of item difficulty stayed the same or became a little lower.

**Table 1.38 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 3 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
<b>10000011184</b>	<b>0.60</b>	<b>0.57</b>	<b>3488139</b>	<b>0.41</b>	<b>0.47</b>
<b>3595527</b>	<b>0.56</b>	<b>0.56</b>	<b>3564095</b>	<b>0.42</b>	<b>0.45</b>
100000210424	0.52	0.58	<b>3510072</b>	<b>0.85</b>	<b>0.87</b>
100000025225	0.88	0.83	<b>3564080</b>	<b>0.60</b>	<b>0.61</b>
<b>3509941</b>	<b>0.58</b>	<b>0.62</b>	3509967	0.90	0.91
<b>3595501</b>	<b>0.55</b>	<b>0.56</b>	<b>3509949</b>	<b>0.60</b>	<b>0.76</b>
100000025207	0.94	0.95	<b>3985609</b>	<b>0.58</b>	<b>0.68</b>
3510005	0.51	0.61	3497888	0.90	0.91
<b>100000025196</b>	<b>0.87</b>	<b>0.90</b>	<b>3547998</b>	<b>0.88</b>	<b>0.88</b>
<b>3595519</b>	<b>0.63</b>	<b>0.64</b>	<b>3564094</b>	<b>0.52</b>	<b>0.55</b>
100000025199	0.95	0.94	100000011195	0.88	0.90
3488065	0.86	0.88	100000018397	0.88	0.85
100000004258	0.90	0.89	100000011207	0.85	0.83
			3490569	0.89	0.91

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



**Table 1.39 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 3 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)			
						0	1	2	Omit
2008	100000011184	BCR	2,571	0.60	0.49	39.32	60.21		0.47
2008	3595527	BCR	2,571	1.13	0.52	6.65	71.61	20.50	1.24
2008	3509941	BCR	29,364	0.58	0.49	41.23	57.86		0.91
2008	3595501	BCR	29,364	1.10	0.61	12.19	62.16	23.74	1.91
2008	100000025196	BCR	2,556	0.87	0.33	12.56	87.21		0.23
2008	3595519	BCR	2,556	1.25	0.78	19.72	33.61	45.70	0.98
2007	3488139	BCR	2,150	0.41	0.49	57.95	41.07		0.98
2007	3564095	BCR	2,150	0.84	0.62	26.70	59.35	12.23	1.72
2008	3510072	BCR	29,364	0.85	0.36	14.45	84.99		0.56
2008	3564080	BCR	29,364	1.19	0.79	22.14	33.94	42.60	1.32
2006	3509949	BCR	2,845	0.60	0.49	37.50	60.35		2.14
2006	3985609	BCR	2,845	1.17	0.57	24.00	32.40	42.20	1.34
2007	3547998	BCR	2,189	0.88	0.32	10.83	88.31		0.87
2007	3564094	BCR	2,189	1.05	0.57	12.75	67.29	18.82	1.14
2009	100000011184	BCR	30,174	0.57	0.49	42.47	57.18		0.35
2009	3595527	BCR	30,174	1.11	0.51	7.28	73.07	19.15	0.50
2009	3509941	BCR	30,174	0.62	0.48	37.39	62.16		0.44
2009	3595501	BCR	30,174	1.12	0.57	10.59	65.66	22.94	0.81
2009	100000025196	BCR	30,174	0.9	0.3	9.86	89.69		0.45
2009	3595519	BCR	30,174	1.28	0.74	16.33	37.70	45.16	0.81
2009	3488139	BCR	30,174	0.47	0.5	52.11	47.33		0.56
2009	3564095	BCR	30,174	0.9	0.61	23.12	61.70	14.29	0.88
2009	3510072	BCR	30,174	0.87	0.34	12.76	86.92		0.32
2009	3564080	BCR	30,174	1.23	0.75	18.69	38.28	42.26	0.77
2009	3509949	BCR	30,174	0.76	0.43	23.06	75.81		1.14
2009	3985609	BCR	30,174	1.36	0.71	12.65	37.33	49.22	0.80
2009	3547998	BCR	30,174	0.88	0.33	11.66	87.83		0.51
2009	3564094	BCR	30,174	1.1	0.56	10.53	67.07	21.60	0.80



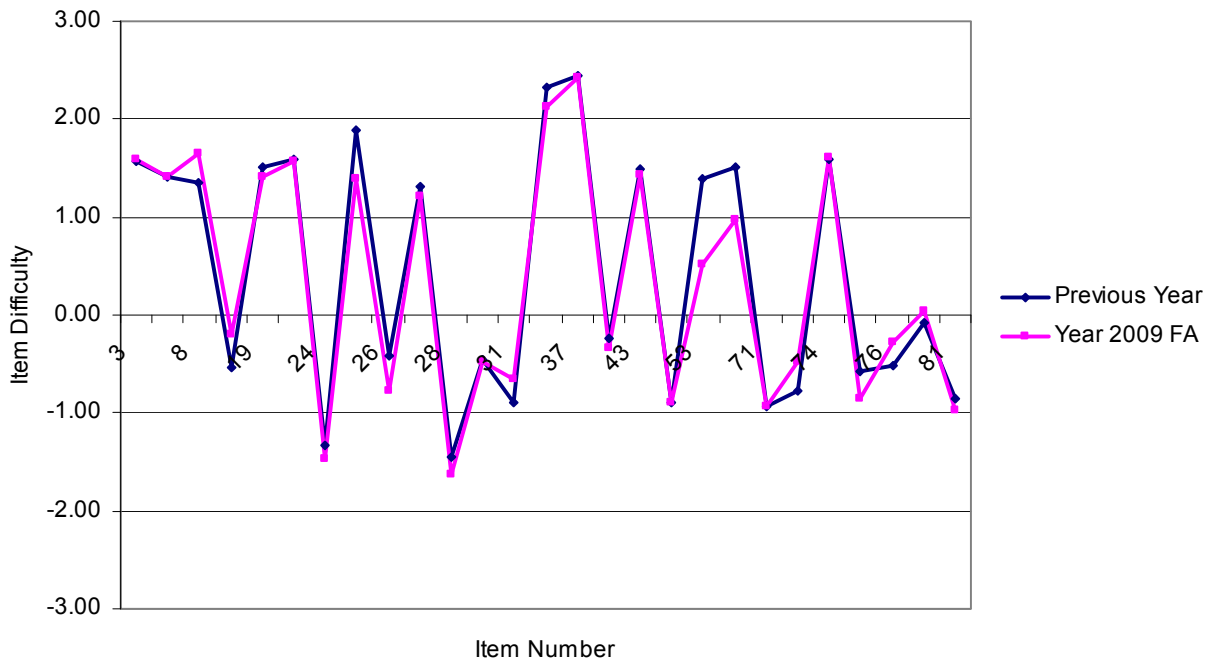
**Table 1.40 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 3 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2008	3	10000011184	BCR_A	1.5687		
2008	4	3595527	BCR_B	1.4040	-2.3770	2.3770
2004	8	10000210424	SR	1.3435		
2008	18	10000025225	SR	-0.5433		
2008	19	3509941	BCR_A	1.5122		
2008	20	3595501	BCR_B	1.5891	-1.8002	1.8002
2008	24	10000025207	SR	-1.3387		
2006	25	3510005	SR	1.8874		
2008	26	10000025196	BCR_A	-0.4187		
2008	27	3595519	BCR_B	1.3106	-0.5282	0.5282
2008	28	10000025199	SR	-1.4434		
2007	30	3488065	SR	-0.4640		
2008	31	10000004258	SR	-0.8902		
2007	36	3488139	BCR_A	2.3340		
2007	37	3564095	BCR_B	2.4437	-1.7282	1.7282
2008	42	3510072	BCR_A	-0.2447		
2008	43	3564080	BCR_B	1.5000	-0.5243	0.5243
2006	47	3509967	SR	-0.9029		
2006	53	3509949	BCR_A	1.3843		
2006	54	3985609	BCR_B	1.5179	-0.4973	0.4973
2007	71	3497888	SR	-0.9392		
2007	73	3547998	BCR_A	-0.7716		
2007	74	3564094	BCR_B	1.5840	-2.0306	2.0306
2008	75	10000011195	SR	-0.5799		
2008	76	10000018397	SR	-0.5179		
2008	80	10000011207	SR	-0.0709		
2007	81	3490569	SR	-0.8519		
2009	3	10000011184	BCR_A	1.5806		
2009	4	3595527	BCR_B	1.4153	-2.5617	2.5617
2009	8	10000210424	SR	1.6564		
2009	18	10000025225	SR	-0.2038		
2009	19	3509941	BCR_A	1.4160		
2009	20	3595501	BCR_B	1.5639	-1.9549	1.9549
2009	24	10000025207	SR	-1.4723		
2009	25	3510005	SR	1.3897		
2009	26	10000025196	BCR_A	-0.7697		
2009	27	3595519	BCR_B	1.2125	-0.9262	0.9262
2009	28	10000025199	SR	-1.6379		

**Table 1.40 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2009	30	3488065	SR	-0.4834		
2009	31	100000004258	SR	-0.6636		
2009	36	3488139	BCR_A	2.1303		
2009	37	3564095	BCR_B	2.4239	-2.0904	2.0904
2009	42	3510072	BCR_A	-0.3424		
2009	43	3564080	BCR_B	1.4331	-0.8085	0.8085
2009	47	3509967	SR	-0.8856		
2009	53	3509949	BCR_A	0.5146		
2009	54	3985609	BCR_B	0.9641	-0.9820	0.9820
2009	71	3497888	SR	-0.9265		
2009	73	3547998	BCR_A	-0.4669		
2009	74	3564094	BCR_B	1.6096	-2.008	2.008
2009	75	100000011195	SR	-0.8620		
2009	76	100000018397	SR	-0.2798		
2009	80	100000011207	SR	0.0479		
2009	81	3490569	SR	-0.9778		

Note. Rasch item and step difficulties are on a common scale.

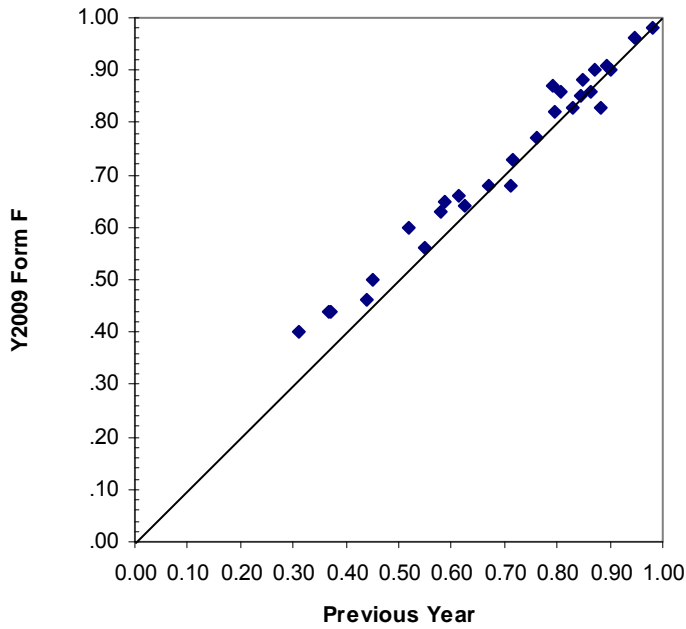


**Figure 1.3 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 3 Form A**

**Table 1.41 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 3 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
<b>100000011186</b>	<b>0.72</b>	<b>0.73</b>	<b>3564083</b>	<b>0.79</b>	<b>0.82</b>
<b>3595529</b>	<b>0.52</b>	<b>0.60</b>	<b>3488087</b>	<b>0.37</b>	<b>0.44</b>
100000011211	0.67	0.68	<b>3564099</b>	<b>0.31</b>	<b>0.40</b>
100000025225	0.88	0.83	100000004263	0.86	0.86
<b>3509941</b>	<b>0.58</b>	<b>0.63</b>	100000213058	0.79	0.87
<b>3595501</b>	<b>0.55</b>	<b>0.56</b>	3511729	0.71	0.68
100000004275	0.37	0.44	<b>3509978</b>	<b>0.61</b>	<b>0.66</b>
<b>100000025196</b>	<b>0.87</b>	<b>0.90</b>	<b>3985610</b>	<b>0.45</b>	<b>0.50</b>
<b>3595519</b>	<b>0.63</b>	<b>0.64</b>	100000025211	0.76	0.77
100000025210	0.95	0.96	<b>3509932</b>	<b>0.98</b>	<b>0.98</b>
100000004270	0.85	0.88	<b>3564086</b>	<b>0.44</b>	<b>0.46</b>
100000025202	0.81	0.86	100000011196	0.83	0.83
100000004258	0.90	0.90	100000018395	0.90	0.91
<b>3510067</b>	<b>0.85</b>	<b>0.85</b>	3510176	0.59	0.65

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



**Table 1.42 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 3 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)			
						0	1	2	Omit
2008	100000011186	BCR	2,585	0.72	0.45	27.66	71.76		0.58
2008	3595529	BCR	2,585	1.03	0.43	6.38	81.55	10.95	1.12
2008	3509941	BCR	29,364	0.58	0.49	41.23	57.86		0.91
2008	3595501	BCR	29,364	1.10	0.61	12.19	62.16	23.74	1.91
2008	100000025196	BCR	2,556	0.87	0.33	12.56	87.21		0.23
2008	3595519	BCR	2,556	1.25	0.78	19.72	33.61	45.70	0.98
2008	3510067	BCR	29,253	0.85	0.36	14.75	84.62		0.62
2008	3564083	BCR	29,253	1.59	0.61	5.26	28.70	65.01	1.03
2007	3488087	BCR	2,073	0.37	0.48	60.83	37.34		1.83
2007	3564099	BCR	2,073	0.62	0.69	47.13	38.54	11.63	2.70
2006	3509978	BCR	2,818	0.61	0.49	38.04	61.18		0.78
2006	3985610	BCR	2,818	0.91	0.38	18.50	70.10	10.10	1.10
2008	3509932	BCR	29,253	0.98	0.14	1.31	98.07		0.63
2008	3564086	BCR	29,253	0.88	0.62	24.86	59.60	14.23	1.32
2009	100000011186	BCR	29,789	0.73	0.44	26.18	73.44		0.38
2009	3595529	BCR	29,789	1.21	0.52	5.05	68.22	26.18	0.55
2009	3509941	BCR	29,789	0.63	0.48	36.18	63.38		0.44
2009	3595501	BCR	29,789	1.12	0.57	10.04	66.55	22.62	0.79
2009	100000025196	BCR	29,789	0.90	0.3	9.41	90.16		0.44
2009	3595519	BCR	29,789	1.28	0.73	16.12	38.23	44.89	0.76
2009	3510067	BCR	29,789	0.85	0.36	15.10	84.52		0.39
2009	3564083	BCR	29,789	1.65	0.61	6.31	21.38	71.69	0.61
2009	3488087	BCR	29,789	0.44	0.5	55.20	43.86		0.94
2009	3564099	BCR	29,789	0.79	0.64	31.53	54.22	12.57	1.69
2009	3509978	BCR	29,789	0.66	0.47	33.40	66.18		0.43
2009	3985610	BCR	29,789	1.00	0.45	9.57	79.39	10.34	0.69
2009	3509932	BCR	29,789	0.98	0.13	1.28	98.31		0.41
2009	3564086	BCR	29,789	0.92	0.65	24.46	57.37	17.39	0.79

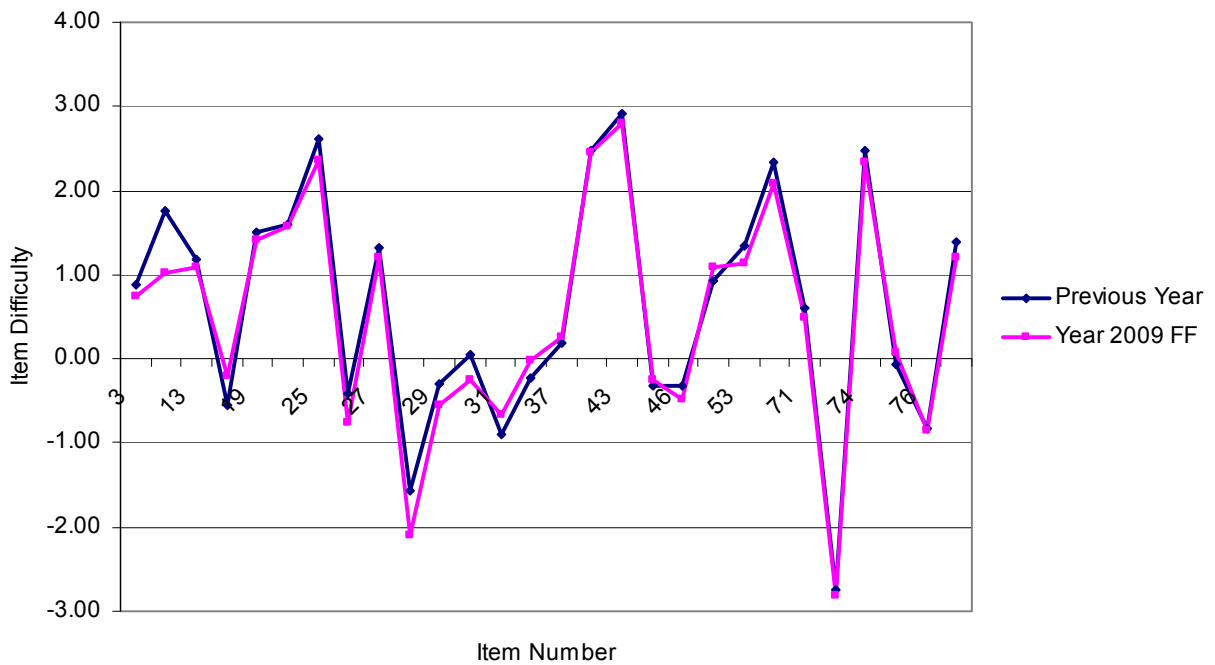
**Table 1.43 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 3 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2008	3	100000011186	BCR_A	0.8721		
2008	4	3595529	BCR_B	1.7686	-2.8580	2.8580
2008	13	100000011211	SR	1.1742		
2008	18	100000025225	SR	-0.5433		
2008	19	3509941	BCR_A	1.5122		
2008	20	3595501	BCR_B	1.5891	-1.8002	1.8002
2008	25	100000004275	SR	2.6241		
2008	26	100000025196	BCR_A	-0.4187		
2008	27	3595519	BCR_B	1.3106	-0.5282	0.5282
2008	28	100000025210	SR	-1.5572		
2008	29	100000004270	SR	-0.3022		
2008	30	100000025202	SR	0.0473		
2008	31	100000004258	SR	-0.8902		
2008	36	3510067	BCR_A	-0.2338		
2008	37	3564083	BCR_B	0.1995	-0.9490	0.9490
2007	42	3488087	BCR_A	2.4701		
2007	43	3564099	BCR_B	2.9223	-0.9692	0.9692
2008	44	100000004263	SR	-0.3298		
2004	46	100000213058	SR	-0.3189		
2008	50	3511729	SR	0.9162		
2006	53	3509978	BCR_A	1.3364		
2006	54	3985610	BCR_B	2.3317	-2.1612	2.1612
2008	71	100000025211	SR	0.5946		
2008	73	3509932	BCR_A	-2.7407		
2008	74	3564086	BCR_B	2.4652	-1.6902	1.6902
2008	75	100000011196	SR	-0.0568		
2008	76	100000018395	SR	-0.8257		
2006	81	3510176	SR	1.3847		
2009	3	100000011186	BCR_A	0.7532		
2009	4	3595529	BCR_B	1.0151	-2.4072	2.4072
2009	13	100000011211	SR	1.0793		
2009	18	100000025225	SR	-0.2038		
2009	19	3509941	BCR_A	1.4160		
2009	20	3595501	BCR_B	1.5639	-1.9549	1.9549
2009	25	100000004275	SR	2.3631		
2009	26	100000025196	BCR_A	-0.7697		
2009	27	3595519	BCR_B	1.2125	-0.9262	0.9262
2009	28	100000025210	SR	-2.1068		
2009	29	100000004270	SR	-0.5454		
2009	30	100000025202	SR	-0.2431		
2009	31	100000004258	SR	-0.6636		
2009	36	3510067	BCR_A	-0.0129		

**Table 1.43 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2009	37	3564083	BCR_B	0.2571	-0.4253	0.4253
2009	42	3488087	BCR_A	2.4484		
2009	43	3564099	BCR_B	2.8035	-1.5459	1.5459
2009	44	10000004263	SR	-0.2543		
2009	46	100000213058	SR	-0.4736		
2009	50	3511729	SR	1.0809		
2009	53	3509978	BCR_A	1.1411		
2009	54	3985610	BCR_B	2.0758	-2.6995	2.6995
2009	71	100000025211	SR	0.4822		
2009	73	3509932	BCR_A	-2.8170		
2009	74	3564086	BCR_B	2.3423	-1.5697	1.5697
2009	75	100000011196	SR	0.0623		
2009	76	100000018395	SR	-0.8495		
2009	81	3510176	SR	1.2021		

Note. Rasch item and step difficulties are on a common scale.

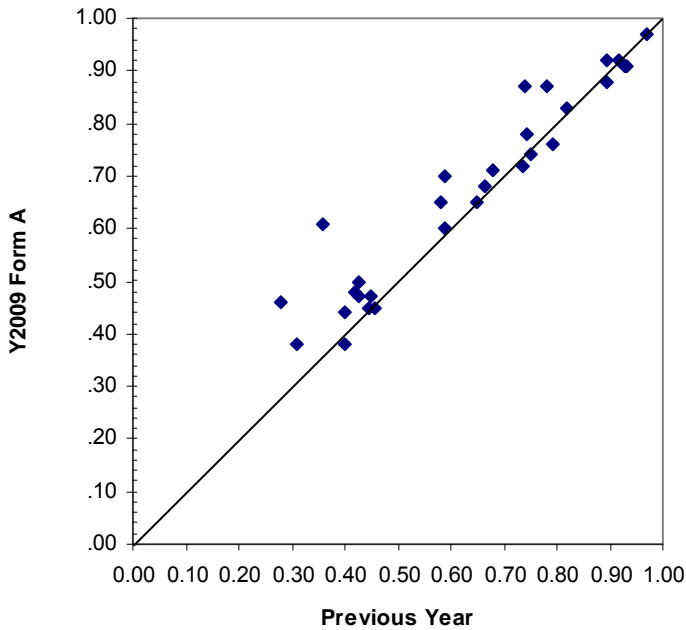


**Figure 1.4 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 3 Form F**

**Table 1.44 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
<b>3487819</b>	<b>0.43</b>	<b>0.50</b>	<b>3488150</b>	<b>0.31</b>	<b>0.38</b>
<b>3564186</b>	<b>0.40</b>	<b>0.44</b>	<b>3564176</b>	<b>0.43</b>	<b>0.47</b>
3515827	0.58	0.65	10000007115	0.89	0.88
10000025162	0.93	0.91	10000025157	0.92	0.92
<b>10000044142</b>	<b>0.82</b>	<b>0.83</b>	<b>3488145</b>	<b>0.59</b>	<b>0.70</b>
<b>3595499</b>	<b>0.45</b>	<b>0.47</b>	<b>3564189</b>	<b>0.42</b>	<b>0.48</b>
10000007124	0.65	0.65	10000012183	0.59	0.6
<b>10000025172</b>	<b>0.36</b>	<b>0.61</b>	10000011509	0.79	0.76
<b>3985613</b>	<b>0.28</b>	<b>0.46</b>	10000011489	0.97	0.97
10000018336	0.75	0.78	100000201842	0.74	0.87
10000007112	0.46	0.45	<b>3515783</b>	<b>0.75</b>	<b>0.74</b>
10000025188	0.66	0.68	<b>3595560</b>	<b>0.74</b>	<b>0.72</b>
3487993	0.89	0.92	3488164	0.93	0.91
10000007125	0.68	0.71	100000201853	0.78	0.87
<b>3515823</b>	<b>0.45</b>	<b>0.45</b>			
<b>3595532</b>	<b>0.40</b>	<b>0.38</b>			

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



**Table 1.45 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 4 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)			
						0	1	2	Omit
2007	3487819	BCR	2,173	0.43	0.50	55.82	42.75		1.43
2007	3564186	BCR	2,173	0.80	0.46	20.29	75.38	2.49	1.84
2008	100000044142	BCR	30,101	0.82	0.38	16.89	82.01		1.10
2008	3595499	BCR	30,101	0.90	0.56	19.62	67.90	10.80	1.67
2004	100000025172	BCR	10,401	0.37	0.48	61.37	36.92		1.48
2004	3985613	BCR	10,401	0.56	0.61	45.57	44.12	6.18	3.26
2008	3515823	BCR	30,101	0.45	0.50	54.09	44.61		1.30
2008	3595532	BCR	30,101	0.80	0.62	28.76	57.76	11.15	2.33
2007	3488150	BCR	2,179	0.31	0.46	68.20	31.07		0.73
2007	3564176	BCR	2,179	0.86	0.69	30.89	50.11	17.76	1.24
2007	3488145	BCR	2,130	0.59	0.49	39.95	58.87		1.17
2007	3564189	BCR	2,130	0.84	0.63	27.37	57.84	13.05	1.74
2008	3515783	BCR	29,933	0.75	0.43	24.51	74.97		0.51
2008	3595560	BCR	29,933	1.47	0.73	13.22	24.30	61.52	0.95
2009	3487819	BCR	29,532	0.50	0.50	48.90	50.38		0.72
2009	3564186	BCR	29,532	0.87	0.48	17.96	74.87	6.12	1.04
2009	100000044142	BCR	29,532	0.83	0.37	15.97	83.19		0.83
2009	3595499	BCR	29,532	0.93	0.55	17.64	69.16	11.98	1.23
2009	100000025172	BCR	29,532	0.61	0.49	38.31	61.13		0.56
2009	3985613	BCR	29,532	0.92	0.63	22.83	59.65	16.42	1.10
2009	3515823	BCR	29,532	0.45	0.50	54.31	44.64		1.05
2009	3595532	BCR	29,532	0.76	0.58	30.03	60.40	7.84	1.74
2009	3488150	BCR	29,532	0.38	0.48	61.60	37.51		0.89
2009	3564176	BCR	29,532	0.95	0.72	27.53	47.94	23.30	1.24
2009	3488145	BCR	29,532	0.70	0.46	29.29	70.10		0.61
2009	3564189	BCR	29,532	0.95	0.53	15.00	72.15	11.65	1.20
2009	3515783	BCR	29,532	0.74	0.44	25.25	74.11		0.64
2009	3595560	BCR	29,532	1.43	0.76	15.36	23.47	59.89	1.28



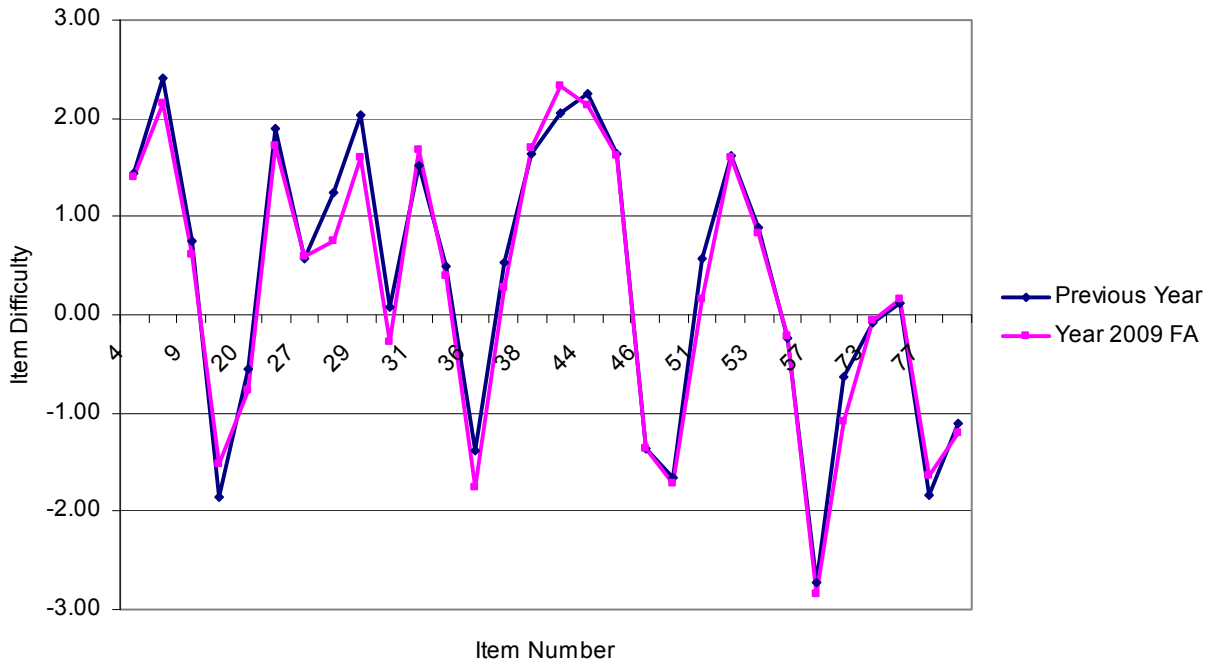
**Table 1.46 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2007	4	3487819	BCR_A	1.4461		
2007	5	3564186	BCR_B	2.4073	-2.9945	2.9945
2006	9	3515827	SR	0.7444		
2008	11	10000025162	SR	-1.8646		
2008	20	10000044142	BCR_A	-0.5527		
2008	21	3595499	BCR_B	1.8966	-2.1375	2.1375
2008	27	10000007124	SR	0.5634		
2004	28	10000025172	BCR_A	1.2405		
2004	29	3985613	BCR_B	2.0371	-1.4842	1.4842
2008	30	10000018336	SR	0.0884		
2008	31	10000007112	SR	1.5155		
2008	35	10000025188	SR	0.4849		
2007	36	3487993	SR	-1.3829		
2008	37	10000007125	SR	0.5246		
2008	38	3515823	BCR_A	1.6468		
2008	39	3595532	BCR_B	2.0517	-1.6746	1.6746
2007	44	3488150	BCR_A	2.2404		
2007	45	3564176	BCR_B	1.6411	-1.2952	1.2952
2008	46	10000007115	SR	-1.3641		
2008	48	10000025157	SR	-1.6646		
2007	51	3488145	BCR_A	0.5813		
2007	52	3564189	BCR_B	1.6234	-1.6600	1.6600
2008	53	10000012183	SR	0.8859		
2008	56	10000011509	SR	-0.2393		
2008	57	10000011489	SR	-2.7177		
2005	70	100000201842	SR	-0.6388		
2008	73	3515783	BCR_A	-0.0713		
2008	74	3595560	BCR_B	0.1264	-0.3519	0.3519
2008	77	3488164	SR	-1.8417		
2004	79	100000201853	SR	-1.1133		
2009	4	3487819	BCR_A	1.3917		
2009	5	3564186	BCR_B	2.1573	-2.6637	2.6637
2009	9	3515827	SR	0.6152		
2009	11	10000025162	SR	-1.5223		
2009	20	10000044142	BCR_A	-0.7600		
2009	21	3595499	BCR_B	1.7168	-2.1933	2.1933
2009	27	10000007124	SR	0.5869		
2009	28	10000025172	BCR_A	0.7480		
2009	29	3985613	BCR_B	1.5908	-1.7399	1.7399
2009	30	10000018336	SR	-0.2781		
2009	31	10000007112	SR	1.6779		
2009	35	10000025188	SR	0.3956		

**Table 1.46 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2009	36	3487993	SR	-1.7595		
2009	37	10000007125	SR	0.2837		
2009	38	3515823	BCR_A	1.6944		
2009	39	3595532	BCR_B	2.3297	-1.8612	1.8612
2009	44	3488150	BCR_A	2.1316		
2009	45	3564176	BCR_B	1.6184	-1.1561	1.1561
2009	46	10000007115	SR	-1.3705		
2009	48	100000025157	SR	-1.7237		
2009	51	3488145	BCR_A	0.1552		
2009	52	3564189	BCR_B	1.5957	-2.3378	2.3378
2009	53	100000012183	SR	0.8211		
2009	56	100000011509	SR	-0.2239		
2009	57	100000011489	SR	-2.8436		
2009	70	100000201842	SR	-1.0863		
2009	73	3515783	BCR_A	-0.0626		
2009	74	3595560	BCR_B	0.1566	-0.3225	0.3225
2009	77	3488164	SR	-1.6322		
2009	79	100000201853	SR	-1.2081		

Note. Rasch item and step difficulties are on a common scale.

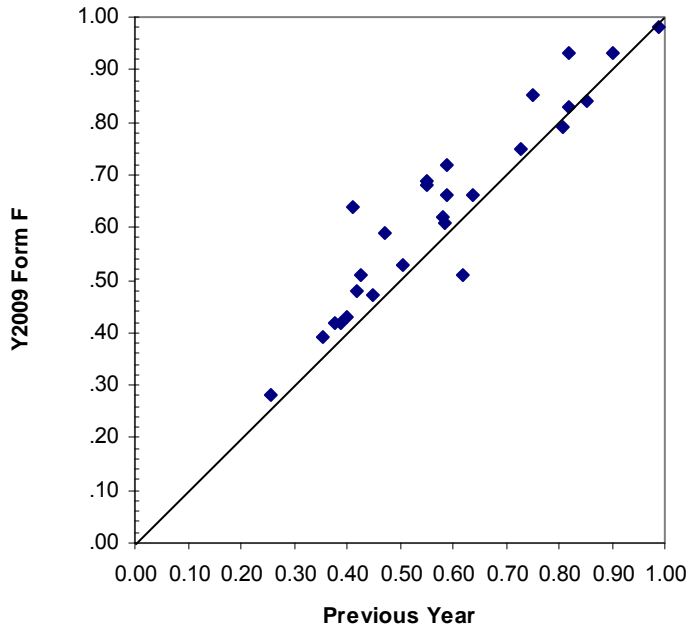


**Figure 1.5 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form A**

**Table 1.47 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
3488053	0.81	0.79	<b>3515807</b>	<b>0.73</b>	<b>0.75</b>
<b>3487819</b>	<b>0.43</b>	<b>0.51</b>	<b>3564165</b>	<b>0.35</b>	<b>0.39</b>
<b>3564186</b>	<b>0.40</b>	<b>0.43</b>	3497865	0.64	0.66
10000025153	0.99	0.98	10000025156	0.85	0.84
<b>10000044142</b>	<b>0.82</b>	<b>0.83</b>	<b>3488145</b>	<b>0.59</b>	<b>0.72</b>
<b>3595499</b>	<b>0.45</b>	<b>0.47</b>	<b>3564189</b>	<b>0.42</b>	<b>0.48</b>
10000007116	0.58	0.62	3551477	0.50	0.53
<b>3515642</b>	<b>0.26</b>	<b>0.28</b>	100000201942	0.55	0.69
<b>3985619</b>	<b>0.38</b>	<b>0.42</b>	100000201852	0.75	0.85
3498610	0.90	0.93	100000201939	0.82	0.93
10000025158	0.62	0.51	<b>100000201940</b>	<b>0.55</b>	<b>0.68</b>
<b>100000201938</b>	<b>0.59</b>	<b>0.66</b>	<b>3985623</b>	<b>0.41</b>	<b>0.64</b>
<b>3985620</b>	<b>0.39</b>	<b>0.42</b>	100000212986	0.47	0.59
			3502602	0.59	0.61

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



**Table 1.48 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 4 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)			
						0	1	2	Omit
2007	3487819	BCR	2,173	0.43	0.50	55.82	42.75		1.43
2007	3564186	BCR	2,173	0.80	0.46	20.29	75.38	2.49	1.84
2008	10000044142	BCR	30,101	0.82	0.38	16.89	82.01		1.10
2008	3595499	BCR	30,101	0.90	0.56	19.62	67.90	10.80	1.67
2006	3515642	BCR	24,774	0.26	0.44	72.37	25.63		1.84
2006	3985619	BCR	24,774	0.75	0.51	38.7	41.2	16.9	2.25
2005	100000201938	BCR	25,326	0.59	0.49	38.68	59.11		1.89
2005	3985620	BCR	25,326	0.78	0.41	27.0	61.3	8.53	2.25
2008	3515807	BCR	30,101	0.73	0.45	24.02	72.76		3.23
2008	3564165	BCR	30,101	0.71	0.65	38.63	48.61	11.01	1.75
2007	3488145	BCR	2,130	0.59	0.49	39.95	58.87		1.17
2007	3564189	BCR	2,130	0.84	0.63	27.37	57.84	13.05	1.74
2005	100000201940	BCR	12,555	0.56	0.50	43.03	55.56		1.41
2005	3985623	BCR	12,555	0.83	0.63	48.60	16.80	32.8	1.62
2009	3487819	BCR	29,476	0.51	0.5	48.50	50.77		0.73
2009	3564186	BCR	29,476	0.87	0.49	18.36	74.41	6.18	1.04
2009	10000044142	BCR	29,476	0.83	0.37	15.90	83.23		0.87
2009	3595499	BCR	29,476	0.93	0.55	17.41	69.37	11.96	1.26
2009	3515642	BCR	29,476	0.28	0.45	70.80	28.03		1.16
2009	3985619	BCR	29,476	0.84	0.72	32.78	46.11	18.80	2.31
2009	100000201938	BCR	29,476	0.66	0.47	33.17	66.19		0.64
2009	3985620	BCR	29,476	0.85	0.46	18.39	76.43	4.12	1.06
2009	3515807	BCR	29,476	0.75	0.43	20.96	75.02		4.02
2009	3564165	BCR	29,476	0.79	0.64	32.17	54.02	12.42	1.39
2009	3488145	BCR	29,476	0.72	0.45	27.65	71.70		0.65
2009	3564189	BCR	29,476	0.96	0.52	14.61	72.43	11.77	1.19
2009	100000201940	BCR	29,476	0.68	0.47	30.92	68.13		0.95
2009	3985623	BCR	29,476	1.28	0.84	23.61	22.50	52.59	1.30

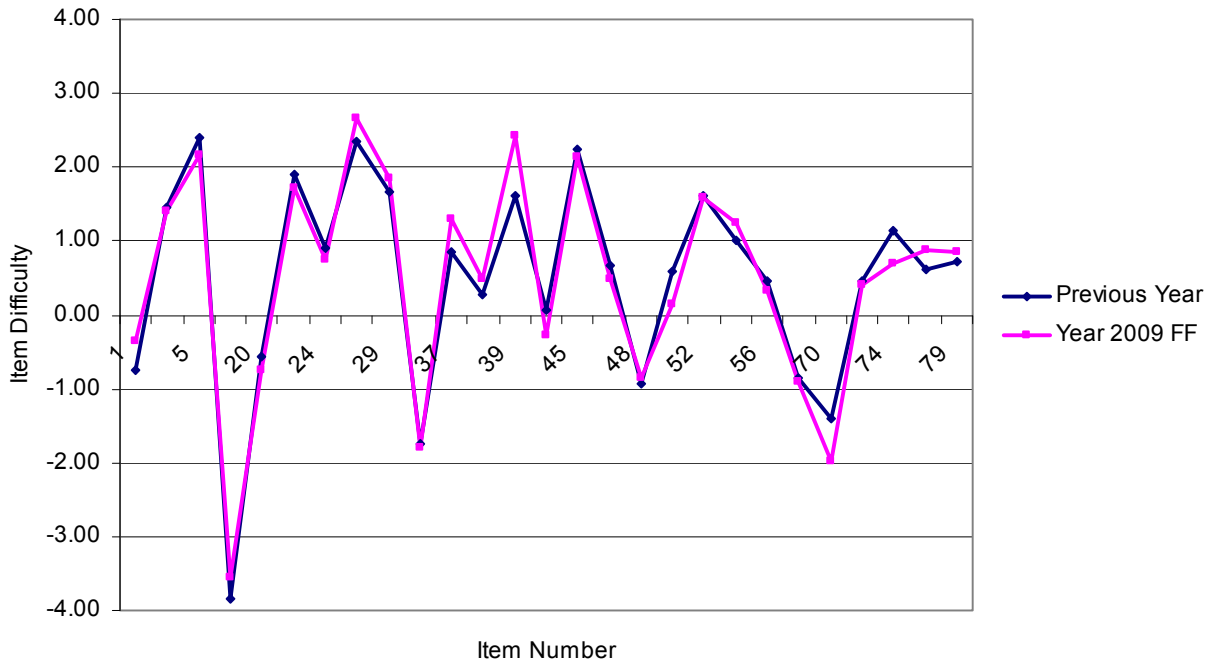
**Table 1.49 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2007	1	3488053	SR	-0.7442		
2007	4	3487819	BCR_A	1.4461		
2007	5	3564186	BCR_B	2.4073	-2.9945	2.9945
2008	9	10000025153	SR	-3.8367		
2008	20	10000044142	BCR_A	-0.5527		
2008	21	3595499	BCR_B	1.8966	-2.1375	2.1375
2008	24	10000007116	SR	0.9058		
2006	28	3515642	BCR_A	2.3491		
2006	29	3985619	BCR_B	1.6714	-0.9459	0.9459
2007	35	3498610	SR	-1.7458		
2008	37	10000025158	SR	0.8480		
2005	38	100000201938	BCR_A	0.2673		
2005	39	3985620	BCR_B	1.6258	-1.9431	1.9431
2008	44	3515807	BCR_A	0.0553		
2008	45	3564165	BCR_B	2.2474	-1.2842	1.2842
2008	46	3497865	SR	0.6767		
2008	48	10000025156	SR	-0.9340		
2007	51	3488145	BCR_A	0.5813		
2007	52	3564189	BCR_B	1.6234	-1.6600	1.6600
2007	53	3551477	SR	1.0150		
2005	56	100000201942	SR	0.4552		
2004	57	100000201852	SR	-0.8584		
2004	70	100000201939	SR	-1.4026		
2005	73	100000201940	BCR_A	0.4695		
2005	74	3985623	BCR_B	1.1279	0.4578	-0.4578
2004	77	100000212986	SR	0.6052		
2007	79	3502602	SR	0.7239		
2009	1	3488053	SR	-0.3452		
2009	4	3487819	BCR_A	1.3917		
2009	5	3564186	BCR_B	2.1573	-2.6637	2.6637
2009	9	10000025153	SR	-3.5593		
2009	20	10000044142	BCR_A	-0.7600		
2009	21	3595499	BCR_B	1.7168	-2.1933	2.1933
2009	24	10000007116	SR	0.7504		
2009	28	3515642	BCR_A	2.6593		
2009	29	3985619	BCR_B	1.8365	-1.1225	1.1225
2009	35	3498610	SR	-1.7896		
2009	37	10000025158	SR	1.2922		
2009	38	100000201938	BCR_A	0.4797		

**Table 1.49 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2
2009	39	3985620	BCR_B	2.4328	-2.8528	2.8528
2009	44	3515807	BCR_A	-0.2644		
2009	45	3564165	BCR_B	2.1333	-1.6175	1.6175
2009	46	3497865	SR	0.4883		
2009	48	10000025156	SR	-0.8555		
2009	51	3488145	BCR_A	0.1552		
2009	52	3564189	BCR_B	1.5957	-2.3378	2.3378
2009	53	3551477	SR	1.2590		
2009	56	100000201942	SR	0.3220		
2009	57	100000201852	SR	-0.9014		
2009	70	100000201939	SR	-1.9724		
2009	73	100000201940	BCR_A	0.4107		
2009	74	3985623	BCR_B	0.6834	0.0238	-0.0238
2009	77	100000212986	SR	0.8914		
2009	79	3502602	SR	0.8642		

Note. Rasch item and step difficulties are on a common scale.

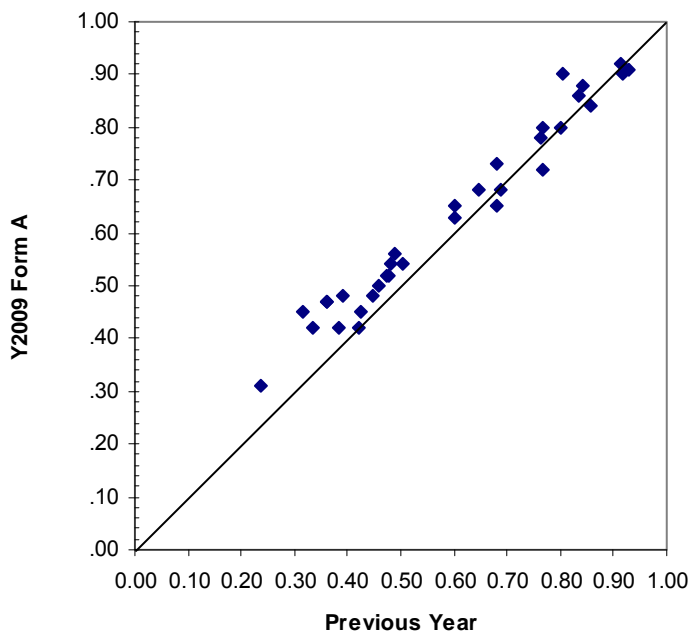


**Figure 1.6 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 4 Form F**

**Table 1.50 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 5 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
3488443	0.42	0.42	<b>3548429</b>	<b>0.69</b>	<b>0.68</b>
3492139	0.48	0.54	<b>3564047</b>	<b>0.60</b>	<b>0.63</b>
<b>3488471</b>	<b>0.24</b>	<b>0.31</b>	3488375	0.92	0.90
<b>3564052</b>	<b>0.34</b>	<b>0.42</b>	100000028274	0.83	0.86
3488391	0.84	0.88	<b>3488347</b>	<b>0.42</b>	<b>0.45</b>
100000022548	0.48	0.52	<b>3564046</b>	<b>0.38</b>	<b>0.42</b>
3488331	0.80	0.80	3488393	0.93	0.91
3488507	0.76	0.78	3511572	0.39	0.48
3492130	0.51	0.54	3512712	0.81	0.9
3512698	0.91	0.92	<b>3488277</b>	<b>0.36</b>	<b>0.47</b>
<b>3488461</b>	<b>0.46</b>	<b>0.50</b>	<b>3564193</b>	<b>0.32</b>	<b>0.45</b>
<b>3564055</b>	<b>0.68</b>	<b>0.73</b>	100000022555	0.45	0.48
<b>3488522</b>	<b>0.60</b>	<b>0.65</b>	<b>3488406</b>	<b>0.49</b>	<b>0.56</b>
<b>3564059</b>	<b>0.77</b>	<b>0.80</b>	<b>3563998</b>	<b>0.48</b>	<b>0.52</b>
3488419	0.77	0.72	3488326	0.68	0.65
100000028276	0.86	0.84	<b>3488348</b>	<b>0.65</b>	<b>0.68</b>
			<b>3464056</b>	<b>0.36</b>	<b>0.47</b>

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



**Table 1.51 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 5 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2007	3488471	BCR	2,171	0.24	0.43	72.78	23.77			3.45
2007	3564052	BCR	2,171	0.67	0.65	38.09	46.98	10.00		4.93
2007	3488461	BCR	2,125	0.46	0.50	52.52	45.98			1.51
2007	3564055	BCR	2,125	1.36	0.75	14.54	31.62	52.09		1.74
2007	3488522	BCR	2,161	0.60	0.49	38.87	60.02			1.11
2007	3564059	BCR	2,161	1.53	0.65	7.27	29.20	62.05		1.48
2007	3548429	ECR	2,125	0.69	0.46	28.52	68.75			2.73
2007	3564047	ECR	2,125	1.81	1.06	13.50	18.2	33.4	31.9	2.82
2007	3488347	BCR	2,164	0.42	0.49	56.61	42.28			1.11
2007	3564046	BCR	2,164	0.77	0.90	52.73	14.33	31.15		1.80
2007	3488277	BCR	2,164	0.36	0.48	56.01	36.18			7.81
2007	3564193	BCR	2,164	0.63	0.72	42.24	35.07	14.09		8.60
2007	3488406	BCR	2,188	0.49	0.50	47.94	48.67			3.38
2007	3563998	BCR	2,188	0.95	0.68	22.12	54.11	20.43		3.34
2007	3488348	BCR	2,178	0.65	0.48	30.44	64.51			5.05
2007	3464056	BCR	2,178	0.72	0.82	45.30	24.90	23.5		6.20
2009	3488471	BCR	30,344	0.31	0.46	66.91	31.44			1.66
2009	3564052	BCR	30,344	0.84	0.59	23.93	63.09	10.55		2.44
2009	3488461	BCR	30,344	0.50	0.50	49.36	49.61			1.03
2009	3564055	BCR	30,344	1.46	0.70	10.79	30.01	57.95		1.25
2009	3488522	BCR	30,344	0.65	0.48	34.01	65.10			0.90
2009	3564059	BCR	30,344	1.59	0.60	4.84	28.22	65.60		1.34
2009	3548429	ECR	30,344	0.68	0.47	30.00	68.26			1.74
2009	3564047	ECR	30,344	1.89	1.06	11.92	20.92	27.54	37.53	2.09
2009	3488347	BCR	30,344	0.45	0.50	53.65	44.93			1.42
2009	3564046	BCR	30,344	0.84	0.92	49.26	12.77	35.74		2.23
2009	3488277	BCR	30,344	0.47	0.50	48.79	46.81			4.40
2009	3564193	BCR	30,344	0.90	0.75	28.93	42.89	23.71		4.47
2009	3488406	BCR	30,344	0.56	0.50	42.02	56.21			1.77
2009	3563998	BCR	30,344	1.05	0.68	19.15	53.39	25.66		1.80
2009	3488348	BCR	30,344	0.68	0.47	28.58	67.81			3.61
2009	3464056	BCR	30,344	0.95	0.78	28.77	38.57	28.08		4.58



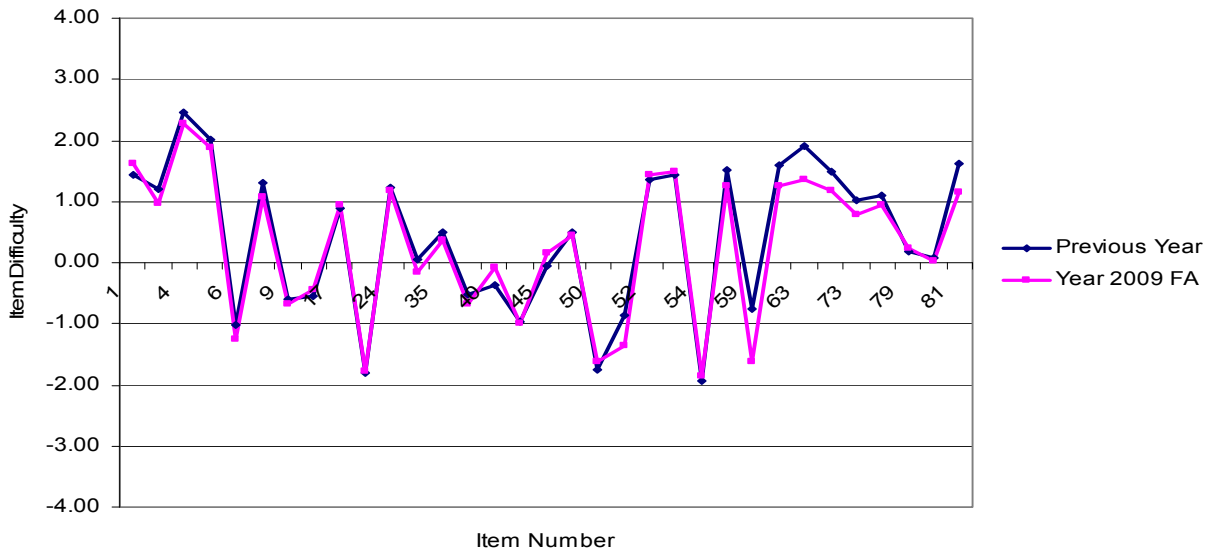
**Table 1.52 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 5 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	1	3488443	SR	1.4475			
2007	3	3492139	SR	1.1965			
2007	4	3488471	BCR_A	2.4670			
2007	5	3564052	BCR_B	2.0114	-1.4070	1.4070	
2007	6	3488391	SR	-1.0187			
2008	7	100000022548	SR	1.2992			
2008	9	3488331	SR	-0.5933			
2007	10	3488507	SR	-0.5534			
2007	17	3492130	SR	0.8800			
2006	22	3512698	SR	-1.8140			
2007	24	3488461	BCR_A	1.2389			
2007	25	3564055	BCR_B	0.0582	-0.5915	0.5915	
2007	35	3488522	BCR_A	0.4908			
2007	36	3564059	BCR_B	-0.5136	-0.7535	0.7535	
2008	40	3488419	SR	-0.3660			
2008	41	100000028276	SR	-0.9757			
2007	45	3548429	ECR_A	-0.0549			
2007	46	3564047	ECR_B	0.4862	-0.7464	-0.3761	1.1225
2008	50	3488375	SR	-1.7634			
2008	51	100000028274	SR	-0.8644			
2007	52	3488347	BCR_A	1.3491			
2007	53	3564046	BCR_B	1.4474	0.6192	-0.6192	
2007	54	3488393	SR	-1.9476			
2006	57	3511572	SR	1.5079			
2006	59	3512712	SR	-0.7681			
2007	62	3488277	BCR_A	1.5848			
2007	63	3564193	BCR_B	1.9023	-0.8832	0.8832	
2008	65	100000022555	SR	1.4943			
2007	73	3488406	BCR_A	1.0165			
2007	74	3563998	BCR_B	1.1030	-1.4333	1.4333	
2007	79	3488326	SR	0.1850			
2007	80	3488348	BCR_A	0.0886			
2007	81	3464056	BCR_B	1.6146	-0.2041	0.2041	
2009	1	3488443	SR	1.6277			
2009	3	3492139	SR	0.9589			
2009	4	3488471	BCR_A	2.2862			
2009	5	3564052	BCR_B	1.8914	-2.0216	2.0216	
2009	6	3488391	SR	-1.2591			
2009	7	100000022548	SR	1.0761			
2009	9	3488331	SR	-0.6848			
2009	10	3488507	SR	-0.4448			
2009	17	3492130	SR	0.9322			
2009	22	3512698	SR	-1.7862			

**Table 1.52 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	24	3488461	BCR_A	1.1765			
2009	25	3564055	BCR_B	-0.1442	-0.4881	0.4881	
2009	35	3488522	BCR_A	0.3599			
2009	36	3564059	BCR_B	-0.6893	-0.8147	0.8147	
2009	40	3488419	SR	-0.0825			
2009	41	10000028276	SR	-0.9996			
2009	45	3548429	ECR_A	0.1442			
2009	46	3564047	ECR_B	0.4382	-0.9888	0.1966	0.7922
2009	50	3488375	SR	-1.6090			
2009	51	10000028274	SR	-1.3725			
2009	52	3488347	BCR_A	1.4298			
2009	53	3564046	BCR_B	1.5002	0.8247	-0.8247	
2009	54	3488393	SR	-1.8559			
2009	57	3511572	SR	1.2621			
2009	59	3512712	SR	-1.6307			
2009	62	3488277	BCR_A	1.2616			
2009	63	3564193	BCR_B	1.3546	-1.0498	1.0498	
2009	65	10000022555	SR	1.1820			
2009	73	3488406	BCR_A	0.7860			
2009	74	3563998	BCR_B	0.9340	-1.4162	1.4162	
2009	79	3488326	SR	0.2463			
2009	80	3488348	BCR_A	0.0277			
2009	81	3464056	BCR_B	1.1386	-0.8486	0.8486	

Note. Rasch item and step difficulties are on a common scale.

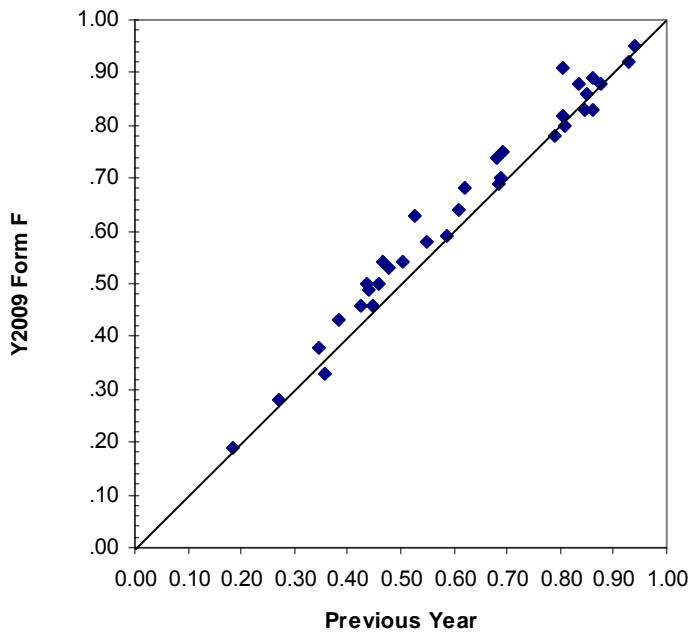


**Figure 1.7 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 5 Form A**

**Table 1.53 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 5 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
3492123	0.47	0.54	<b>3488525</b>	<b>0.80</b>	<b>0.82</b>
3492117	0.94	0.95	<b>3564053</b>	<b>0.44</b>	<b>0.50</b>
<b>3511531</b>	<b>0.69</b>	<b>0.70</b>	3492126	0.85	0.86
<b>3563986</b>	<b>0.59</b>	<b>0.59</b>	100000028274	0.83	0.88
3492128	0.50	0.54	<b>3488347</b>	<b>0.42</b>	<b>0.46</b>
100000022548	0.48	0.53	<b>3564046</b>	<b>0.38</b>	<b>0.43</b>
100000022547	0.84	0.83	3488393	0.93	0.92
100000028251	0.53	0.63	3488509	0.81	0.8
100000028253	0.45	0.46	3512712	0.81	0.91
3488515	0.86	0.83	<b>3548459</b>	<b>0.69</b>	<b>0.75</b>
<b>3488461</b>	<b>0.46</b>	<b>0.50</b>	<b>3564051</b>	<b>0.62</b>	<b>0.68</b>
<b>3564055</b>	<b>0.68</b>	<b>0.74</b>	3488275	0.44	0.49
3488495	0.88	0.88	<b>3512649</b>	<b>0.27</b>	<b>0.28</b>
3492140	0.86	0.89	<b>3563989</b>	<b>0.34</b>	<b>0.38</b>
<b>3512615</b>	<b>0.79</b>	<b>0.78</b>	3492134	0.61	0.64
<b>3595439</b>	<b>0.55</b>	<b>0.58</b>	<b>3488259</b>	<b>0.18</b>	<b>0.19</b>
3488240	0.68	0.69	<b>3564048</b>	<b>0.36</b>	<b>0.33</b>

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



**Table 1.54 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 5 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2008	3511531	BCR	30,537	0.69	0.46	30.23	68.94			0.84
2008	3563986	BCR	30,537	1.18	0.68	13.87	51.27	33.19		1.68
2007	3488461	BCR	2,125	0.46	0.50	52.52	45.98			1.51
2007	3564055	BCR	2,125	1.36	0.75	14.54	31.62	52.09		1.74
2008	3512615	BCR	30,537	0.79	0.41	19.92	78.87			1.21
2008	3595439	BCR	30,537	1.10	0.68	16.68	53.00	28.44		1.88
2007	3488525	ECR	2,315	0.80	0.40	18.88	80.26			0.86
2007	3564053	ECR	2,315	1.31	0.83	12.30	48.60	27.60	8.85	2.38
2007	3488347	BCR	2,164	0.42	0.49	56.61	42.28			1.11
2007	3564046	BCR	2,164	0.77	0.90	52.73	14.33	31.15		1.80
2007	3548459	BCR	2,171	0.69	0.46	25.70	69.00			5.30
2007	3564051	BCR	2,171	1.24	0.72	10.78	42.56	40.90		5.76
2007	3512649	BCR	31,083	0.27	0.44	66.35	27.13			6.52
2007	3563989	BCR	31,083	0.69	0.89	52.00	10.88	29.04		8.08
2007	3488259	BCR	2,315	0.18	0.39	78.49	18.40			3.11
2007	3564048	BCR	2,315	0.71	0.85	50.32	19.27	25.96		4.45
2009	3511531	BCR	30,103	0.7	0.46	29.59	69.82			0.58
2009	3563986	BCR	30,103	1.18	0.64	11.89	55.75	31.12		1.24
2009	3488461	BCR	30,103	0.50	0.50	48.47	50.27			1.26
2009	3564055	BCR	30,103	1.48	0.69	9.97	29.35	59.28		1.41
2009	3512615	BCR	30,103	0.78	0.41	20.34	77.95			1.71
2009	3595439	BCR	30,103	1.16	0.63	10.76	57.99	28.97		2.28
2009	3488525	ECR	30,103	0.82	0.38	17.33	82.17			0.50
2009	3564053	ECR	30,103	1.51	0.86	9.91	39.93	35.24	13.67	1.26
2009	3488347	BCR	30,103	0.46	0.5	52.97	45.86			1.17
2009	3564046	BCR	30,103	0.86	0.92	48.34	13.75	36.11		1.80
2009	3548459	BCR	30,103	0.75	0.43	22.18	75.45			2.37
2009	3564051	BCR	30,103	1.36	0.67	8.54	41.70	47.33		2.42
2009	3512649	BCR	30,103	0.28	0.45	69.53	27.73			2.73
2009	3563989	BCR	30,103	0.76	0.88	49.94	17.41	29.40		3.25
2009	3488259	BCR	30,103	0.19	0.39	78.83	18.82			2.36
2009	3564048	BCR	30,103	0.66	0.85	55.41	15.61	25.19		3.79

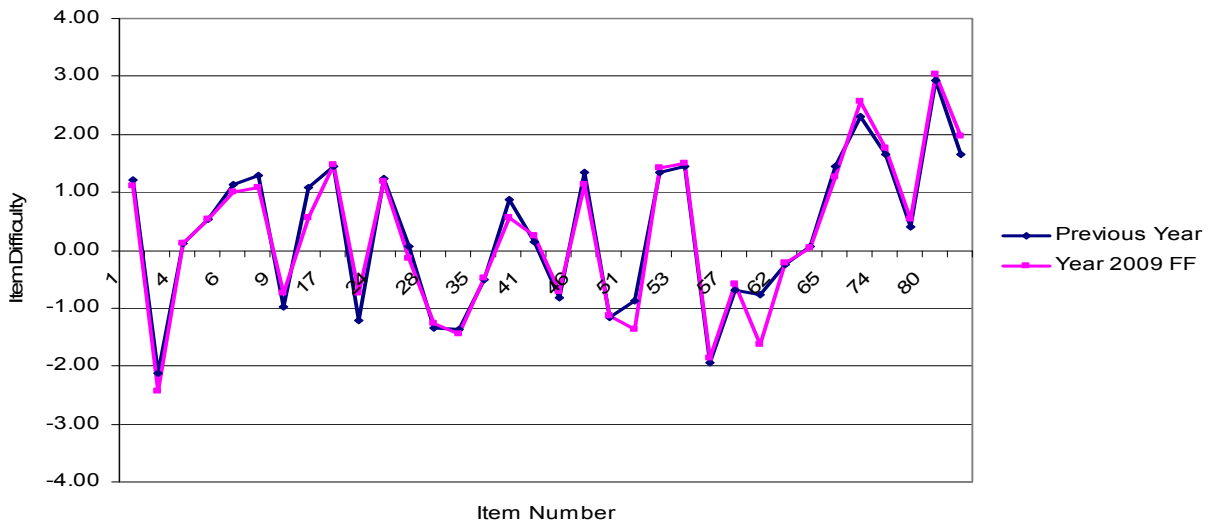
**Table 1.55 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 5 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	1	3492123	SR	1.2035			
2008	3	3492117	SR	-2.1343			
2008	4	3511531	BCR_A	0.1259			
2008	5	3563986	BCR_B	0.5335	-1.3908	1.3908	
2008	6	3492128	SR	1.1464			
2008	7	10000022548	SR	1.2992			
2008	9	10000022547	SR	-0.9649			
2008	10	10000028251	SR	1.0867			
2008	17	10000028253	SR	1.4534			
2007	22	3488515	SR	-1.2030			
2007	24	3488461	BCR_A	1.2389			
2007	25	3564055	BCR_B	0.0582	-0.5915	0.5915	
2008	28	3488495	SR	-1.3420			
2007	34	3492140	SR	-1.3783			
2008	35	3512615	BCR_A	-0.5151			
2008	36	3595439	BCR_B	0.8697	-1.4537	1.4537	
2008	41	3488240	SR	0.1335			
2007	45	3488525	ECR_A	-0.8205			
2007	46	3564053	ECR_B	1.3543	-2.4417	0.4281	2.0136
2007	50	3492126	SR	-1.1530			
2008	51	10000028274	SR	-0.8644			
2007	52	3488347	BCR_A	1.3491			
2007	53	3564046	BCR_B	1.4474	0.6192	-0.6192	
2007	54	3488393	SR	-1.9476			
2008	57	3488509	SR	-0.6778			
2006	59	3512712	SR	-0.7681			
2007	62	3548459	BCR_A	-0.2421			
2007	63	3564051	BCR_B	0.0647	-1.1926	1.1926	
2008	65	3488275	SR	1.4453			
2007	73	3512649	BCR_A	2.3175			
2007	74	3563989	BCR_B	1.6549	0.7655	-0.7655	
2007	79	3492134	SR	0.4131			
2007	80	3488259	BCR_A	2.9215			
2007	81	3564048	BCR_B	1.6508	0.14900	-0.14900	
2009	1	3492123	SR	1.1181			
2009	3	3492117	SR	-2.4397			
2009	4	3511531	BCR_A	0.1282			
2009	5	3563986	BCR_B	0.5377	-1.5615	1.5615	
2009	6	3492128	SR	1.0153			
2009	7	10000022548	SR	1.0761			
2009	9	10000022547	SR	-0.7480			
2009	10	10000028251	SR	0.5516			
2009	17	10000028253	SR	1.4773			
2009	22	3488515	SR	-0.7556			

**Table 1.55 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	24	3488461	BCR_A	1.1765			
2009	25	3564055	BCR_B	-0.1442	-0.4881	0.4881	
2009	28	3488495	SR	-1.2540			
2009	34	3492140	SR	-1.4589			
2009	35	3512615	BCR_A	-0.4877			
2009	36	3595439	BCR_B	0.5673	-1.6948	1.6948	
2009	41	3488240	SR	0.2496			
2009	45	3488525	ECR_A	-0.7202			
2009	46	3564053	ECR_B	1.1388	-2.2589	0.246	2.0128
2009	50	3492126	SR	-1.1318			
2009	51	10000028274	SR	-1.3725			
2009	52	3488347	BCR_A	1.4298			
2009	53	3564046	BCR_B	1.5002	0.8247	-0.8247	
2009	54	3488393	SR	-1.8559			
2009	57	3488509	SR	-0.5908			
2009	59	3512712	SR	-1.6307			
2009	62	3548459	BCR_A	-0.2124			
2009	63	3564051	BCR_B	0.0385	-1.2689	1.2689	
2009	65	3488275	SR	1.2554			
2009	73	3512649	BCR_A	2.5751			
2009	74	3563989	BCR_B	1.7474	0.2559	-0.2559	
2009	79	3492134	SR	0.5423			
2009	80	3488259	BCR_A	3.0327			
2009	81	3564048	BCR_B	1.9710	0.4950	-0.4950	

Note. Rasch item and step difficulties are on a common scale.

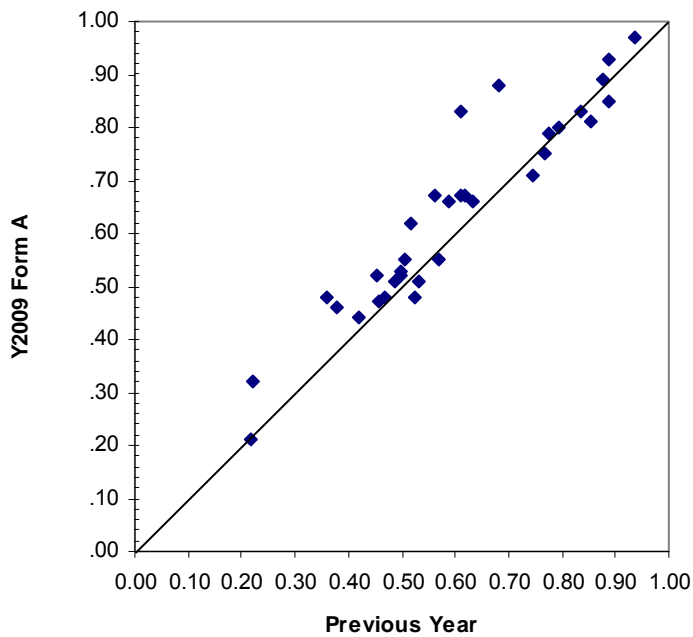


**Figure 1.8 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 5 Form F**

**Table 1.56 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 6 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
3492091	0.77	0.75	3488398	0.94	0.97
100000028397	0.79	0.80	3488358	0.61	0.67
<b>3517004</b>	<b>0.89</b>	<b>0.93</b>	3488302	0.57	0.55
<b>3564010</b>	<b>0.63</b>	<b>0.66</b>	100000012859	0.74	0.71
3488422	0.52	0.48	100000208906	0.68	0.88
100000012866	0.22	0.21	<b>100000208909</b>	<b>0.38</b>	<b>0.46</b>
<b>3548404</b>	<b>0.50</b>	<b>0.53</b>	<b>3985730</b>	<b>0.36</b>	<b>0.48</b>
<b>3564013</b>	<b>0.46</b>	<b>0.47</b>	3503954	0.83	0.83
3488296	0.85	0.81	<b>3516616</b>	<b>0.42</b>	<b>0.44</b>
100000208908	0.22	0.32	<b>3564012</b>	<b>0.50</b>	<b>0.52</b>
<b>3488462</b>	<b>0.52</b>	<b>0.62</b>	100000022470	0.49	0.51
<b>3564075</b>	<b>0.45</b>	<b>0.52</b>	100000208907	0.61	0.83
3503966	0.53	0.51	3488306	0.88	0.89
3516359	0.56	0.67	<b>3488411</b>	<b>0.47</b>	<b>0.48</b>
100000028408	0.89	0.85	<b>3564014</b>	<b>0.51</b>	<b>0.55</b>
<b>3516333</b>	<b>0.62</b>	<b>0.67</b>	3488258	0.77	0.79
<b>3564008</b>	<b>0.59</b>	<b>0.66</b>			

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



**Table 1.57 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 6 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2008	3517004	ECR	31,060	0.89	0.32	10.64	88.75			0.61
2008	3564010	ECR	31,060	1.90	0.94	6.38	25.6	34.0	32.1	1.70
2007	3548404	BCR	2,049	0.50	0.50	48.80	49.68			1.51
2007	3564013	BCR	2,049	0.91	0.39	10.74	83.75	3.76		1.76
2007	3488462	BCR	2,051	0.52	0.50	45.69	51.78			2.54
2007	3564075	BCR	2,051	0.90	0.82	34.42	32.13	29.16		4.29
2008	3516333	BCR	31,060	0.62	0.49	36.21	61.65			2.14
2008	3564008	BCR	31,060	1.18	0.77	18.90	37.50	40.0		3.48
2004	100000208909	BCR	11,242	0.39	0.49	54.88	38.47			6.17
2004	3985730	BCR	11,242	0.72	0.85	45.71	19.76	26.12		7.34
2008	3516616	BCR	31,060	0.42	0.49	51.83	42.07			6.11
2008	3564012	BCR	31,060	0.99	0.62	12.77	61.83	18.73		6.67
2007	3488411	BCR	2,049	0.47	0.50	50.61	47.05			2.34
2007	3564014	BCR	2,049	1.01	0.71	21.57	49.73	25.77		2.93
2009	3517004	ECR	29,789	0.93	0.25	6.08	93.46			0.47
2009	3564010	ECR	29,789	1.99	0.88	4.04	23.74	37.97	33.14	1.10
2009	3548404	BCR	29,789	0.53	0.50	44.82	52.67			2.51
2009	3564013	BCR	29,789	0.95	0.42	8.21	82.27	6.31		3.21
2009	3488462	BCR	29,789	0.62	0.48	36.29	62.29			1.42
2009	3564075	BCR	29,789	1.04	0.78	25.82	38.86	32.46		2.85
2009	3516333	BCR	29,789	0.67	0.47	31.90	66.90			1.20
2009	3564008	BCR	29,789	1.32	0.80	19.61	25.45	53.10		1.84
2009	100000208909	BCR	29,789	0.46	0.50	51.68	45.65			2.67
2009	3985730	BCR	29,789	0.97	0.88	37.33	21.99	37.47		3.21
2009	3516616	BCR	29,789	0.44	0.5	52.26	43.62			4.12
2009	3564012	BCR	29,789	1.05	0.62	12.30	61.48	21.59		4.63
2009	3488411	BCR	29,789	0.48	0.50	50.50	48.10			1.40
2009	3564014	BCR	29,789	1.09	0.67	16.44	54.77	27.22		1.56



**Table 1.58 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 6 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	4	3492091	SR	-0.7205			
2008	5	10000028397	SR	-0.7086			
2008	7	3517004	ECR_A	-1.7891			
2008	8	3564010	ECR_B	0.1623	-1.8832	0.1382	1.745
2008	13	3488422	SR	0.9143			
2008	21	10000012866	SR	2.6975			
2007	22	3548404	BCR_A	0.6845			
2007	23	3564013	BCR_B	1.3882	-3.24110	3.24110	
2008	24	3488296	SR	-1.1250			
2004	28	100000208908	SR	1.5224			
2007	31	3488462	BCR_A	0.6027			
2007	32	3564075	BCR_B	0.9052	-0.4417	0.4417	
2008	33	3503966	SR	0.8698			
2006	44	3516359	SR	0.1259			
2008	46	10000028408	SR	-1.6132			
2008	47	3516333	BCR_A	0.1802			
2008	48	3564008	BCR_B	0.3670	-0.8656	0.8656	
2007	49	3488398	SR	-2.7376			
2007	51	3488358	SR	0.0051			
2007	52	3488302	SR	0.3811			
2008	53	10000012859	SR	-0.2946			
2004	58	100000208906	SR	-0.9310			
2004	59	100000208909	BCR_A	0.6056			
2004	60	3985730	BCR_B	0.6260	0.1263	-0.1263	
2008	62	3503954	SR	-0.9176			
2008	66	3516616	BCR_A	1.2017			
2008	67	3564012	BCR_B	0.6973	-2.0205	2.0205	
2008	69	10000022470	SR	0.9473			
2004	70	100000208907	SR	-0.5845			
2008	71	3488306	SR	-1.4753			
2007	77	3488411	BCR_A	0.8189			
2007	78	3564014	BCR_B	0.5966	-1.3083	1.3083	
2007	79	3488258	SR	-1.0397			
2009	4	3492091	SR	-0.4928			
2009	5	10000028397	SR	-0.9101			
2009	7	3517004	ECR_A	-2.4767			
2009	8	3564010	ECR_B	-0.2026	-1.9106	0.2374	1.6732
2009	13	3488422	SR	1.0056			
2009	21	10000012866	SR	2.8196			
2009	22	3548404	BCR_A	0.7098			
2009	23	3564013	BCR_B	1.2037	-3.0576	3.0576	
2009	24	3488296	SR	-1.0460			
2009	28	100000208908	SR	1.8806			

**Table 1.58 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	31	3488462	BCR_A	0.2562			
2009	32	3564075	BCR_B	0.7929	-0.803	0.803	
2009	33	3503966	SR	0.8279			
2009	44	3516359	SR	0.0835			
2009	46	10000028408	SR	-1.4442			
2009	47	3516333	BCR_A	-0.0954			
2009	48	3564008	BCR_B	0.0294	-0.1519	0.1519	
2009	49	3488398	SR	-3.1672			
2009	51	3488358	SR	-0.0834			
2009	52	3488302	SR	0.6636			
2009	53	10000012859	SR	-0.2466			
2009	58	10000208906	SR	-1.5247			
2009	59	10000208909	BCR_A	1.1385			
2009	60	3985730	BCR_B	0.9849	0.1811	-0.1811	
2009	62	3503954	SR	-1.1222			
2009	66	3516616	BCR_A	1.1920			
2009	67	3564012	BCR_B	0.5835	-1.8963	1.8963	
2009	69	10000022470	SR	0.8821			
2009	70	10000208907	SR	-1.1677			
2009	71	3488306	SR	-1.7554			
2009	77	3488411	BCR_A	1.1054			
2009	78	3564014	BCR_B	0.6090	-1.4986	1.4986	
2009	79	3488258	SR	-0.8592			

Note. Rasch item and step difficulties are on a common scale.

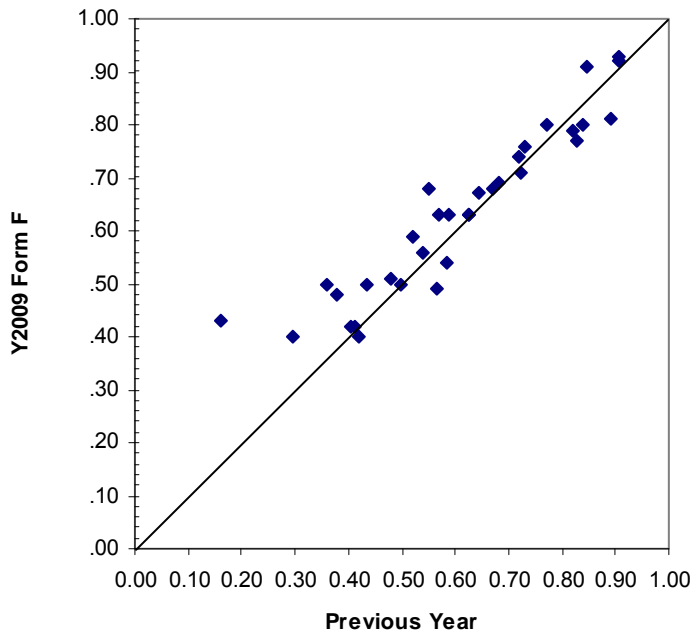


**Figure 1.9 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 6 Form A**

**Table 1.59 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 6 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
3488262	0.85	0.91	3492093	0.84	0.80
<b>3548350</b>	<b>0.52</b>	<b>0.59</b>	100000028419	0.73	0.76
<b>3564015</b>	<b>0.43</b>	<b>0.50</b>	100000004450	0.63	0.63
3488440	0.55	0.68	100000028416	0.56	0.49
3488307	0.67	0.68	3488489	0.72	0.71
<b>3516627</b>	<b>0.48</b>	<b>0.51</b>	<b>100000208909</b>	<b>0.38</b>	<b>0.48</b>
<b>3564006</b>	<b>0.41</b>	<b>0.42</b>	<b>3985730</b>	<b>0.36</b>	<b>0.5</b>
100000022483	0.58	0.54	3492087	0.77	0.8
3488424	0.89	0.81	<b>3516616</b>	<b>0.42</b>	<b>0.40</b>
100000028430	0.64	0.67	<b>3564012</b>	<b>0.50</b>	<b>0.50</b>
<b>3488469</b>	<b>0.68</b>	<b>0.69</b>	100000028409	0.57	0.63
<b>3564071</b>	<b>0.72</b>	<b>0.74</b>	3503961	0.91	0.93
3492099	0.90	0.92	3492120	0.62	0.63
3488260	0.59	0.63	<b>3516913</b>	<b>0.40</b>	<b>0.42</b>
100000004462	0.82	0.79	<b>3985725</b>	<b>0.54</b>	<b>0.56</b>
<b>3516358</b>	<b>0.16</b>	<b>0.43</b>	3488385	0.83	0.77
<b>3985729</b>	<b>0.30</b>	<b>0.40</b>			

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



**Table 1.60 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 6 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2007	3548350	ECR	2,049	0.52	0.50	43.63	51.93			4.44
2007	3564015	ECR	2,049	1.30	1.08	24.74	27.18	25.33	17.33	5.42
2008	3516627	BCR	31,060	0.48	0.50	45.52	47.79			6.69
2008	3564006	BCR	31,060	0.82	0.67	24.69	52.36	14.89		8.06
2007	3488469	BCR	2,024	0.68	0.47	31.32	68.03			0.64
2007	3564071	BCR	2,024	1.44	0.83	21.64	11.36	66.25		0.74
2006	3516358	BCR	3,289	0.16	0.37	78.38	16.11			5.50
2006	3985729	BCR	3,289	0.60	0.39	37.70	52.40	3.53		6.23
2004	100000208909	BCR	11,242	0.39	0.49	54.88	38.47			6.17
2004	3985730	BCR	11,242	0.72	0.85	45.71	19.76	26.12		7.34
2008	3516616	BCR	31,060	0.42	0.49	51.83	42.07			6.11
2008	3564012	BCR	31,060	0.99	0.62	12.77	61.83	18.73		6.67
2008	3516913	BCR	31,060	0.40	0.49	57.57	40.49			1.94
2008	3985725	BCR	31,060	1.08	0.66	15.50	55.50	26.20		2.59
2009	3548350	ECR	29,240	0.59	0.49	36.06	58.68			5.26
2009	3564015	ECR	29,240	1.49	1.13	20.13	21.99	27.08	24.36	6.43
2009	3516627	BCR	29,240	0.51	0.50	43.06	51.24			5.69
2009	3564006	BCR	29,240	0.83	0.60	21.27	61.48	10.83		6.42
2009	3488469	BCR	29,240	0.69	0.46	29.03	69.05			1.92
2009	3564071	BCR	29,240	1.49	0.82	18.61	9.55	69.53		2.31
2009	3516358	BCR	29,240	0.43	0.50	54.09	43.06			2.85
2009	3985729	BCR	29,240	0.79	0.70	33.67	47.15	16.04		3.14
2009	100000208909	BCR	29,240	0.48	0.5	48.62	47.99			3.39
2009	3985730	BCR	29,240	1.00	0.88	35.27	22.01	38.86		3.85
2009	3516616	BCR	29,240	0.4	0.49	53.72	39.87			6.42
2009	3564012	BCR	29,240	0.99	0.61	11.57	63.13	18.07		7.23
2009	3516913	BCR	29,240	0.42	0.49	56.78	41.57			1.66
2009	3985725	BCR	29,240	1.12	0.64	13.15	57.26	27.51		2.09

\*Note. Bold type was used to indicate a visually significant outlier in both classical and Rasch item difficulty graphs.

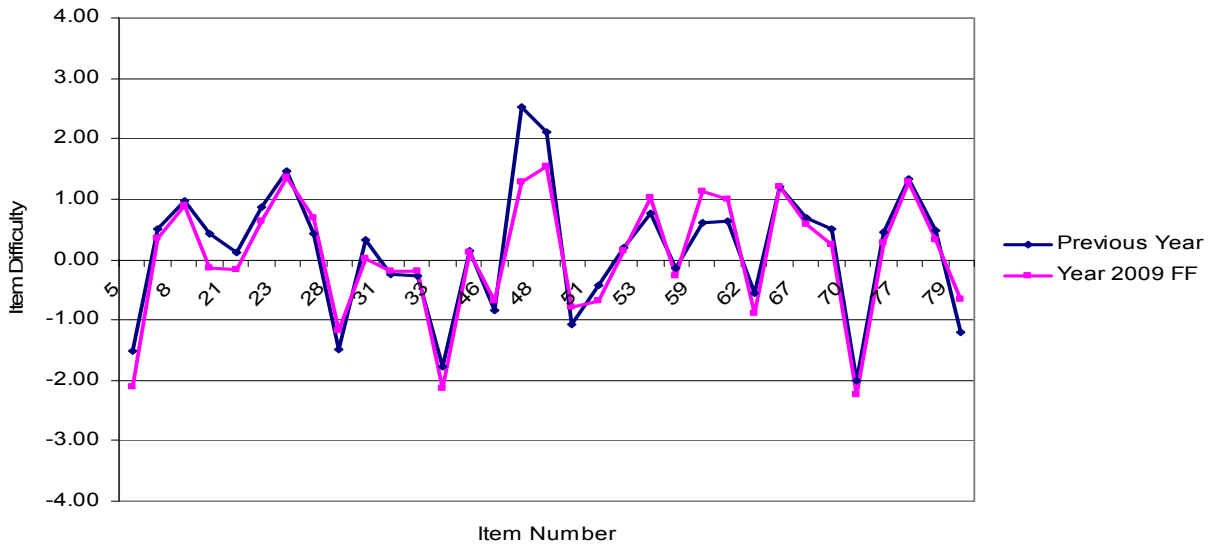
**Table 1.61 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 6 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	5	3488262	SR	-1.5035			
2007	7	3548350	ECR_A	0.5054			
2007	8	3564015	ECR_B	0.9710	-1.0270	-0.0908	1.1178
2007	13	3488440	SR	0.4219			
2008	21	3488307	SR	0.1101			
2008	22	3516627	BCR_A	0.8724			
2008	23	3564006	BCR_B	1.4503	-1.6293	1.6293	
2008	24	100000022483	SR	0.4254			
2008	28	3488424	SR	-1.4785			
2008	29	100000028430	SR	0.3362			
2007	31	3488469	BCR_A	-0.2447			
2007	32	3564071	BCR_B	-0.2617	0.7667	-0.7667	
2008	33	3492099	SR	-1.7828			
2007	44	3488260	SR	0.1509			
2008	46	100000004462	SR	-0.8451			
2006	47	3516358	BCR_A	2.5213			
2006	48	3985729	BCR_B	2.1193	-2.0983	2.0983	
2008	49	3492093	SR	-1.0765			
2008	51	100000028419	SR	-0.4274			
2008	52	100000004450	SR	0.1826			
2008	53	100000028416	SR	0.7594			
2008	58	3488489	SR	-0.1343			
2004	59	100000208909	BCR_A	0.6056			
2004	60	3985730	BCR_B	0.6260	0.1263	-0.1263	
2008	62	3492087	SR	-0.5596			
2008	66	3516616	BCR_A	1.2017			
2008	67	3564012	BCR_B	0.6973	-2.0205	2.0205	
2008	69	100000028409	SR	0.4978			
2008	70	3503961	SR	-2.0081			
2008	71	3492120	SR	0.4607			
2008	77	3516913	BCR_A	1.3284			
2008	78	3985725	BCR_B	0.4845	-1.4754	1.4754	
2007	79	3488385	SR	-1.2026			
2009	5	3488262	SR	-2.1168			
2009	7	3548350	ECR_A	0.3475			
2009	8	3564015	ECR_B	0.8817	-0.7206	-0.1891	0.9097
2009	13	3488440	SR	-0.1323			
2009	21	3488307	SR	-0.1634			
2009	22	3516627	BCR_A	0.6268			
2009	23	3564006	BCR_B	1.3578	-1.9164	1.9164	
2009	24	100000022483	SR	0.6958			
2009	28	3488424	SR	-1.1816			
2009	29	100000028430	SR	0.0122			

**Table 1.61 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	31	3488469	BCR_A	-0.1844			
2009	32	3564071	BCR_B	-0.2008	0.9385	-0.9385	
2009	33	3492099	SR	-2.1347			
2009	44	3488260	SR	0.1206			
2009	46	10000004462	SR	-0.6746			
2009	47	3516358	BCR_A	1.2893			
2009	48	3985729	BCR_B	1.5499	-1.2584	1.2584	
2009	49	3492093	SR	-0.7772			
2009	51	100000028419	SR	-0.6904			
2009	52	10000004450	SR	0.1537			
2009	53	100000028416	SR	1.0327			
2009	58	3488489	SR	-0.2639			
2009	59	100000208909	BCR_A	1.1385			
2009	60	3985730	BCR_B	0.9849	0.1811	-0.1811	
2009	62	3492087	SR	-0.8918			
2009	66	3516616	BCR_A	1.1920			
2009	67	3564012	BCR_B	0.5835	-1.8963	1.8963	
2009	69	100000028409	SR	0.2401			
2009	70	3503961	SR	-2.2456			
2009	71	3492120	SR	0.2830			
2009	77	3516913	BCR_A	1.2807			
2009	78	3985725	BCR_B	0.3129	-1.6336	1.6336	
2009	79	3488385	SR	-0.6632			

\*Note. Bold type was used to indicate a visually significant outlier in both classical and Rasch item difficulty graphs. Note. These Rasch difficulties were based on a common scale.

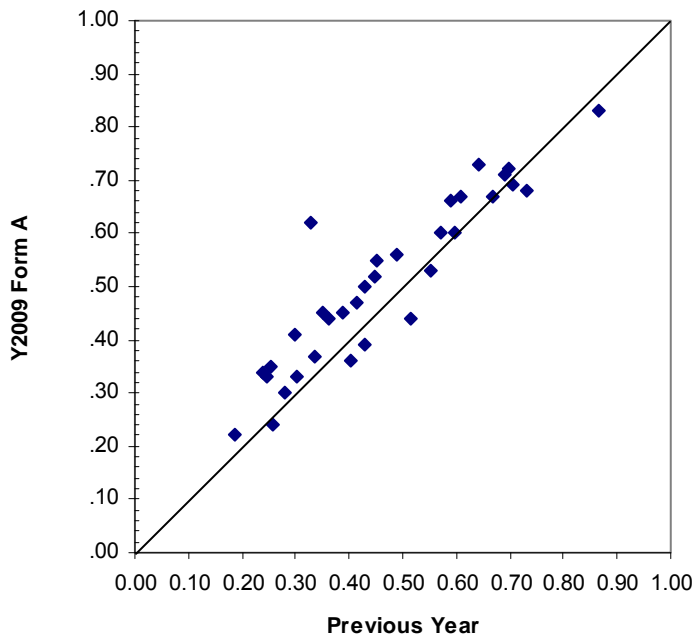


**Figure 1.10 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 6 Form F**

**Table 1.62 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 7 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
3487664	0.26	0.24	<b>3517725</b>	<b>0.30</b>	<b>0.33</b>
<b>3517744</b>	<b>0.35</b>	<b>0.45</b>	<b>3564022</b>	<b>0.45</b>	<b>0.52</b>
<b>3564018</b>	<b>0.24</b>	<b>0.34</b>	100000012796	0.49	0.56
3517677	0.61	0.67	<b>100000043347</b>	<b>0.69</b>	<b>0.71</b>
100000026796	0.87	0.83	<b>3595366</b>	<b>0.30</b>	<b>0.41</b>
3487922	0.73	0.68	<b>3547777</b>	<b>0.25</b>	<b>0.33</b>
<b>100000018130</b>	<b>0.55</b>	<b>0.53</b>	<b>100000207793</b>	<b>0.33</b>	<b>0.62</b>
<b>3491692</b>	<b>0.39</b>	<b>0.45</b>	<b>3517648</b>	<b>0.67</b>	<b>0.67</b>
<b>3564159</b>	<b>0.41</b>	<b>0.47</b>	<b>3564027</b>	<b>0.64</b>	<b>0.73</b>
<b>3517668</b>	<b>0.34</b>	<b>0.37</b>	<b>3500155</b>	<b>0.19</b>	<b>0.22</b>
<b>100000043354</b>	<b>0.40</b>	<b>0.36</b>	<b>3517704</b>	<b>0.43</b>	<b>0.39</b>
<b>100000043350</b>	<b>0.57</b>	<b>0.6</b>	<b>100000004171</b>	<b>0.51</b>	<b>0.44</b>
<b>3487925</b>	<b>0.59</b>	<b>0.66</b>	<b>3487678</b>	<b>0.43</b>	<b>0.5</b>
<b>3564151</b>	<b>0.71</b>	<b>0.69</b>	<b>3564153</b>	<b>0.36</b>	<b>0.44</b>
100000208466	0.45	0.55	100000018106	0.60	0.60
<b>3547642</b>	<b>0.70</b>	<b>0.72</b>	<b>3492167</b>	<b>0.25</b>	<b>0.35</b>
<b>3487560</b>	<b>0.28</b>	<b>0.3</b>			

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



**Table 1.63 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 7 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2007	3517744	BCR	32,264	0.35	0.48	57.42	35.11			7.48
2007	3564018	BCR	32,264	0.48	0.66	52.89	29.28	9.33		8.51
2007	3491692	ECR	2,173	0.39	0.49	58.21	38.84			2.95
2007	3564159	ECR	2,173	1.24	1.12	27.75	28.72	18.04	19.88	5.61
2007	3487925	BCR	2,214	0.59	0.49	38.08	58.76			3.16
2007	3564151	BCR	2,214	1.41	0.71	8.67	33.33	53.84		4.16
2008	3517725	BCR	31,804	0.30	0.46	66.31	30.25			3.44
2008	3564022	BCR	31,804	0.89	0.84	36.49	28.55	30.46		4.50
2008	100000043347	ECR	31,804	0.69	0.46	27.92	69.10			2.99
2008	3595366	ECR	31,804	0.89	0.55	15.77	69.82	9.23	0.35	4.82
2008	3517648	ECR	31,048	0.67	0.47	31.60	66.88			1.52
2008	3564027	ECR	31,048	1.93	0.85	8.05	9.71	57.19	22.85	2.19
2007	3487678	BCR	2,214	0.43	0.50	53.48	42.82			3.70
2007	3564153	BCR	2,214	0.73	0.62	31.48	54.29	9.12		5.10
2009	3517744	BCR	30,318	0.45	0.5	50.92	44.84			4.23
2009	3564018	BCR	30,318	0.69	0.72	41.68	37.99	15.41		4.91
2009	3491692	ECR	30,318	0.45	0.5	51.43	44.91			3.65
2009	3564159	ECR	30,318	1.40	1.11	19.65	30.48	19.98	23.25	6.64
2009	3487925	BCR	30,318	0.66	0.47	27.03	66.02			6.95
2009	3564151	BCR	30,318	1.37	0.76	9.42	28.05	54.60		7.93
2009	3517725	BCR	30,318	0.33	0.47	64.80	32.61			2.58
2009	3564022	BCR	30,318	1.05	0.78	25.01	39.58	32.67		2.74
2009	100000043347	ECR	30,318	0.71	0.45	26.79	71.03			2.19
2009	3595366	ECR	30,318	1.24	0.74	14.30	41.58	39.73	0.91	3.49
2009	3517648	ECR	30,318	0.67	0.47	32.00	66.76			1.24
2009	3564027	ECR	30,318	2.19	0.89	5.34	10.21	39.04	43.42	1.99
2009	3487678	BCR	30,318	0.50	0.50	47.36	50.27			2.36
2009	3564153	BCR	30,318	0.88	0.57	20.11	66.33	10.70		2.86



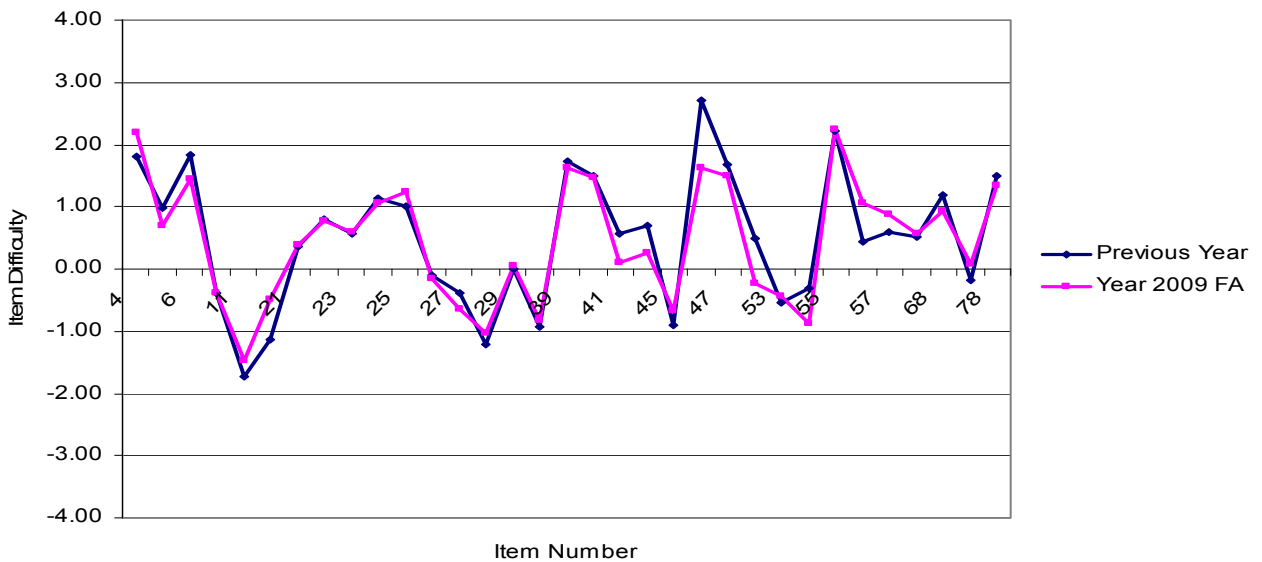
**Table 1.64 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 7 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	4	3487664	SR	1.8127			
2007	5	3517744	BCR_A	0.9733			
2007	6	3564018	BCR_B	1.8283	-0.88100	0.88100	
2006	9	3517677	SR	-0.3983			
2008	11	100000026796	SR	-1.7325			
2007	17	3487922	SR	-1.1436			
2008	21	100000018130	SPR	0.3646			
2007	22	3491692	ECR_A	0.7934			
2007	23	3564159	ECR_B	0.5566	-1.114	0.3367	0.7773
2007	24	3517668	SPR	1.1331			
2008	25	100000043354	SPR	1.0066			
2008	26	100000043350	SPR	-0.0987			
2007	27	3487925	BCR_A	-0.3798			
2007	28	3564151	BCR_B	-1.2247	-0.9516	0.9516	
2004	29	100000208466	SR	0.0047			
2008	38	3547642	SPR	-0.9166			
2008	39	3487560	SPR	1.7398			
2008	40	3517725	BCR_A	1.5045			
2008	41	3564022	BCR_B	0.5585	-0.4548	0.4548	
2008	44	100000012796	SR	0.6843			
2008	45	100000043347	ECR_A	-0.8969			
2008	46	3595366	ECR_B	2.6970	-4.3825	0.7102	3.6723
2007	47	3547777	SPR	1.6862			
2004	48	100000207793	SPR	0.4839			
2008	53	3517648	ECR_A	-0.536			
2008	54	3564027	ECR_B	-0.3195	-0.8801	-1.5639	2.444
2007	55	3500155	SPR	2.2178			
2007	56	3517704	SPR	0.4293			
2008	57	100000004171	SPR	0.6033			
2007	67	3487678	BCR_A	0.5190			
2007	68	3564153	BCR_B	1.1805	-1.8410	1.8410	
2008	71	100000018106	SR	-0.1775			
2007	78	3492167	SPR	1.5035			
2009	4	3487664	SR	2.2060			
2009	5	3517744	BCR_A	0.6921			
2009	6	3564018	BCR_B	1.4348	-1.1447	1.1447	
2009	9	3517677	SR	-0.3998			
2009	11	100000026796	SR	-1.4762			
2009	17	3487922	SR	-0.4981			
2009	21	100000018130	SPR	0.3886			
2009	22	3491692	ECR_A	0.7615			
2009	23	3564159	ECR_B	0.5864	-1.186	0.4071	0.7788
2009	24	3517668	SPR	1.0492			

**Table 1.64 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	25	100000043354	SPR	1.2412			
2009	26	100000043350	SPR	-0.1607			
2009	27	3487925	BCR_A	-0.6335			
2009	28	3564151	BCR_B	-1.0291	-0.8589	0.8589	
2009	29	100000208466	SR	0.0546			
2009	38	3547642	SPR	-0.8342			
2009	39	3487560	SPR	1.6314			
2009	40	3517725	BCR_A	1.4747			
2009	41	3564022	BCR_B	0.1132	-1.1539	1.1539	
2009	44	100000012796	SR	0.2502			
2009	45	100000043347	ECR_A	-0.6646			
2009	46	3595366	ECR_B	1.6353	-3.0103	-0.6940	3.7043
2009	47	3547777	SPR	1.5096			
2009	48	100000207793	SPR	-0.2274			
2009	53	3517648	ECR_A	-0.4422			
2009	54	3564027	ECR_B	-0.8657	-0.9282	-0.5852	1.5135
2009	55	3500155	SPR	2.2489			
2009	56	3517704	SPR	1.0690			
2009	57	100000004171	SPR	0.8686			
2009	67	3487678	BCR_A	0.5631			
2009	68	3564153	BCR_B	0.9376	-2.0991	2.0991	
2009	71	100000018106	SR	0.0819			
2009	78	3492167	SPR	1.3391			

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

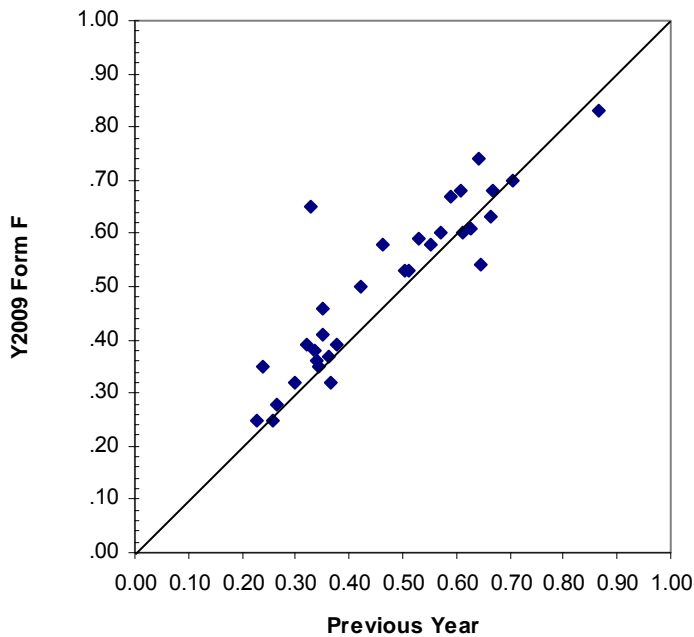


**Figure 1.11 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 7 Form A**

**Table 1.65 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 7 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
3487664	0.26	0.25	<b>100000048821</b>	<b>0.64</b>	<b>0.54</b>
<b>3517744</b>	<b>0.35</b>	<b>0.46</b>	<b>3595371</b>	<b>0.37</b>	<b>0.32</b>
<b>3564018</b>	<b>0.24</b>	<b>0.35</b>	<b>100000012779</b>	<b>0.51</b>	<b>0.53</b>
3517677	0.61	0.68	<b>3595378</b>	<b>0.34</b>	<b>0.36</b>
100000026796	0.87	0.83	<b>100000207795</b>	<b>0.33</b>	<b>0.65</b>
3487605	0.63	0.61	<b>3517683</b>	<b>0.46</b>	<b>0.58</b>
<b>100000043344</b>	<b>0.36</b>	<b>0.37</b>	<b>3517648</b>	<b>0.67</b>	<b>0.68</b>
<b>3487765</b>	<b>0.42</b>	<b>0.50</b>	<b>3564027</b>	<b>0.64</b>	<b>0.74</b>
<b>3564141</b>	<b>0.50</b>	<b>0.53</b>	<b>3492169</b>	<b>0.38</b>	<b>0.39</b>
<b>3517668</b>	<b>0.34</b>	<b>0.38</b>	<b>100000043342</b>	<b>0.67</b>	<b>0.63</b>
<b>3547893</b>	<b>0.23</b>	<b>0.25</b>	<b>3492156</b>	<b>0.35</b>	<b>0.41</b>
<b>100000043350</b>	<b>0.57</b>	<b>0.6</b>	<b>100000012810</b>	<b>0.27</b>	<b>0.28</b>
<b>3487925</b>	<b>0.59</b>	<b>0.67</b>	<b>3595375</b>	<b>0.32</b>	<b>0.39</b>
<b>3564151</b>	<b>0.71</b>	<b>0.70</b>	3487748	0.61	0.60
100000208467	0.53	0.59	<b>100000018133</b>	<b>0.30</b>	<b>0.32</b>
<b>3555865</b>	<b>0.34</b>	<b>0.35</b>			
<b>100000043360</b>	<b>0.55</b>	<b>0.58</b>			

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



**Table 1.66 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 7 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2007	3517744	BCR	32,264	0.35	0.48	57.42	35.11			7.48
2007	3564018	BCR	32,264	0.48	0.66	52.89	29.28	9.33		8.51
2008	3487765	ECR	31,048	0.42	0.49	54.91	42.35			2.75
2008	3564141	ECR	31,048	1.51	0.80	8.29	29.32	50.91	6.52	4.96
2007	3487925	BCR	2,214	0.59	0.49	38.08	58.76			3.16
2007	3564151	BCR	2,214	1.41	0.71	8.67	33.33	53.84		4.16
2008	100000048821	BCR	31,048	0.64	0.48	32.78	64.45			2.78
2008	3595371	BCR	31,048	0.73	0.52	25.16	66.09	3.70		5.05
2008	100000012779	ECR	2,631	0.51	0.50	46.37	51.12			2.51
2008	3595378	ECR	2,631	1.02	0.77	20.90	49.87	21.28	3.15	4.79
2008	3517648	ECR	31,048	0.67	0.47	31.60	66.88			1.52
2008	3564027	ECR	31,048	1.93	0.85	8.05	9.71	57.19	22.85	2.19
2008	100000012810	BCR	2,635	0.27	0.44	60.53	26.64			12.83
2008	3595375	BCR	2,635	0.64	0.56	25.31	56.09	4.14		14.46
2009	3517744	BCR	29,596	0.46	0.5	49.83	46.17			4.00
2009	3564018	BCR	29,596	0.71	0.73	40.57	38.41	16.11		4.92
2009	3487765	ECR	29,596	0.50	0.50	47.94	49.86			2.20
2009	3564141	ECR	29,596	1.59	0.79	7.66	23.81	57.03	7.13	4.37
2009	3487925	BCR	29,596	0.67	0.47	26.30	67.22			6.47
2009	3564151	BCR	29,596	1.40	0.75	9.04	27.61	55.95		7.41
2009	100000048821	BCR	29,596	0.54	0.5	44.92	54.33			0.74
2009	3595371	BCR	29,596	0.64	0.53	38.01	58.63	2.62		0.74
2009	100000012779	ECR	29,596	0.53	0.5	44.53	53.03			2.44
2009	3595378	ECR	29,596	1.08	0.79	20.85	49.25	23.48	4.02	2.40
2009	3517648	ECR	29,596	0.68	0.47	31.09	67.74			1.17
2009	3564027	ECR	29,596	2.21	0.88	4.88	10.09	38.57	44.61	1.85
2009	100000012810	BCR	29,596	0.28	0.45	63.90	27.65			8.45
2009	3595375	BCR	29,596	0.77	0.52	18.57	67.43	4.83		9.16

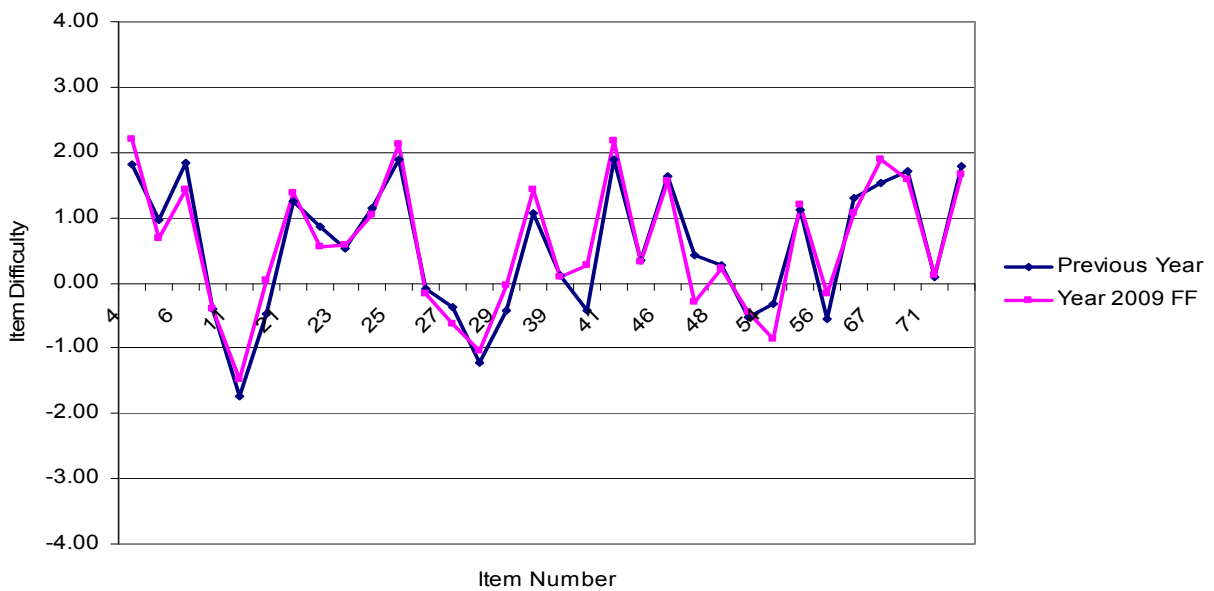
**Table 1.67 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 7 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	4	3487664	SR	1.8127			
2007	5	3517744	BCR_A	0.9733			
2007	6	3564018	BCR_B	1.8283	-0.8810	0.8810	
2006	9	3517677	SR	-0.3983			
2008	11	100000026796	SR	-1.7325			
2007	17	3487605	SR	-0.4674			
2008	21	100000043344	SPR	1.2419			
2008	22	3487765	ECR_A	0.8535			
2008	23	3564141	ECR_B	0.5148	-2.3799	-0.926	3.3059
2007	24	3517668	SPR	1.1331			
2007	25	3547893	SPR	1.9012			
2008	26	100000043350	SPR	-0.0987			
2007	27	3487925	BCR_A	-0.3798			
2007	28	3564151	BCR_B	-1.2247	-0.9516	0.9516	
2004	29	100000208467	SR	-0.4135			
2007	38	3555865	SPR	1.0579			
2008	39	100000043360	SPR	0.1275			
2008	40	100000048821	BCR_A	-0.4202			
2008	41	3595371	BCR_B	1.8783	-2.7673	2.7673	
2008	45	100000012779	ECR_A	0.3568			
2008	46	3595378	ECR_B	1.6372	-2.6078	0.2235	2.3843
2004	47	100000207795	SPR	0.4116			
2007	48	3517683	SPR	0.2733			
2008	53	3517648	ECR_A	-0.5360			
2008	54	3564027	ECR_B	-0.3195	-0.8801	-1.5639	2.444
2008	55	3492169	SPR	1.1079			
2008	56	100000043342	SPR	-0.5518			
2008	57	3492156	SPR	1.2944			
2008	67	100000012810	BCR_A	1.5293			
2008	68	3595375	BCR_B	1.7097	-2.4467	2.4467	
2008	71	3487748	SR	0.0836			
2008	78	100000018133	SPR	1.7903			
2009	4	3487664	SR	2.206			
2009	5	3517744	BCR_A	0.6921			
2009	6	3564018	BCR_B	1.4348	-1.1447	1.1447	
2009	9	3517677	SR	-0.3998			
2009	11	100000026796	SR	-1.4762			
2009	17	3487605	SR	0.0261			
2009	21	100000043344	SPR	1.3744			
2009	22	3487765	ECR_A	0.5452			
2009	23	3564141	ECR_B	0.5905	-2.1342	-1.1123	3.2465
2009	24	3517668	SPR	1.0492			

**Table 1.67 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	25	3547893	SPR	2.1319			
2009	26	100000043350	SPR	-0.1607			
2009	27	3487925	BCR_A	-0.6335			
2009	28	3564151	BCR_B	-1.0291	-0.8589	0.8589	
2009	29	100000208467	SR	-0.0362			
2009	38	3555865	SPR	1.4240			
2009	39	100000043360	SPR	0.0968			
2009	40	100000048821	BCR_A	0.2601			
2009	41	3595371	BCR_B	2.1656	-2.2964	2.2964	
2009	45	100000012779	ECR_A	0.3215			
2009	46	3595378	ECR_B	1.5606	-2.5425	0.2773	2.2652
2009	47	100000207795	SPR	-0.2841			
2009	48	3517683	SPR	0.2289			
2009	53	3517648	ECR_A	-0.4422			
2009	54	3564027	ECR_B	-0.8657	-0.9282	-0.5852	1.5135
2009	55	3492169	SPR	1.1920			
2009	56	100000043342	SPR	-0.1596			
2009	57	3492156	SPR	1.0662			
2009	67	100000012810	BCR_A	1.8858			
2009	68	3595375	BCR_B	1.5793	-2.5264	2.5264	
2009	71	3487748	SR	0.1268			
2009	78	100000018133	SPR	1.6464			

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

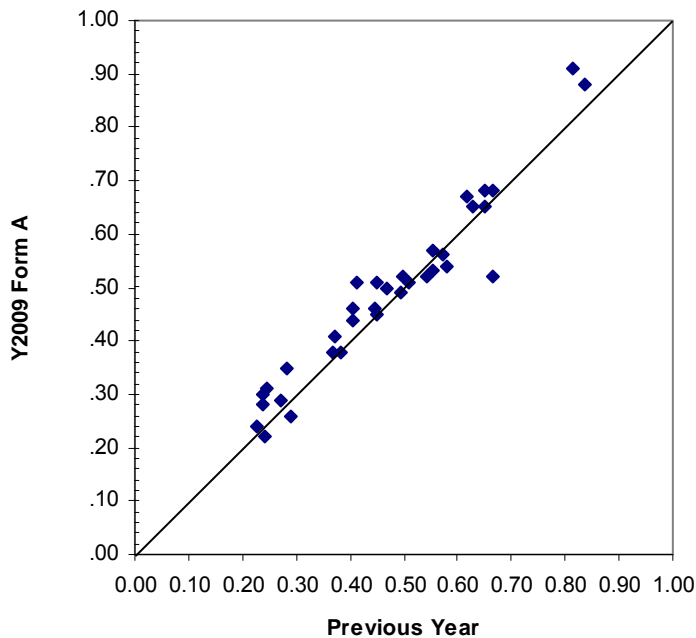


**Figure 1.12 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 7 Form F**

**Table 1.68 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 8 Form A**

Item CID	Previous Year	Year 09 Form A	Item CID	Previous Year	Year 09 Form A
<b>3514013</b>	<b>0.47</b>	<b>0.50</b>	<b>3487934</b>	<b>0.51</b>	<b>0.51</b>
<b>3564107</b>	<b>0.65</b>	<b>0.68</b>	<b>3564122</b>	<b>0.38</b>	<b>0.38</b>
100000018156	0.67	0.68	100000049037	0.65	0.65
<b>100000018174</b>	<b>0.54</b>	<b>0.52</b>	<b>3487633</b>	<b>0.62</b>	<b>0.67</b>
<b>3487680</b>	<b>0.25</b>	<b>0.31</b>	<b>3564123</b>	<b>0.49</b>	<b>0.49</b>
<b>3564133</b>	<b>0.28</b>	<b>0.35</b>	<b>3514164</b>	<b>0.55</b>	<b>0.57</b>
<b>3514065</b>	<b>0.82</b>	<b>0.91</b>	<b>3564117</b>	<b>0.45</b>	<b>0.51</b>
<b>100000026780</b>	<b>0.37</b>	<b>0.38</b>	<b>100000043326</b>	<b>0.45</b>	<b>0.45</b>
<b>100000043325</b>	<b>0.57</b>	<b>0.56</b>	<b>3514083</b>	<b>0.24</b>	<b>0.28</b>
<b>100000004108</b>	<b>0.56</b>	<b>0.53</b>	3519734	0.23	0.24
100000012754	0.63	0.65	100000004118	0.58	0.54
<b>3487759</b>	<b>0.24</b>	<b>0.30</b>	<b>3487906</b>	<b>0.24</b>	<b>0.22</b>
<b>3564128</b>	<b>0.41</b>	<b>0.51</b>	<b>3500160</b>	<b>0.29</b>	<b>0.26</b>
<b>100000004078</b>	<b>0.50</b>	<b>0.52</b>	<b>3487939</b>	<b>0.41</b>	<b>0.44</b>
<b>3514117</b>	<b>0.37</b>	<b>0.41</b>	<b>3564124</b>	<b>0.27</b>	<b>0.29</b>
<b>3564111</b>	<b>0.40</b>	<b>0.46</b>	100000004091	0.67	0.52
<b>3513646</b>	<b>0.45</b>	<b>0.46</b>			
<b>3514597</b>	<b>0.84</b>	<b>0.88</b>			

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



**Table 1.69 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 8 Form A**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2008	3514013	BCR	32,318	0.47	0.50	50.68	47.01			2.31
2008	3564107	BCR	32,318	1.30	0.69	9.99	43.80	43.23		2.99
2007	3487680	ECR	2,430	0.25	0.43	67.28	24.61			8.11
2007	3564133	ECR	2,430	0.85	1.18	48.89	13.17	8.48	18.27	11.19
2007	3487759	BCR	2,157	0.24	0.43	71.86	23.69			4.45
2007	3564128	BCR	2,157	0.83	0.76	33.10	39.73	21.42		5.75
2008	3514117	BCR	32,318	0.37	0.48	55.99	36.99			7.02
2008	3564111	BCR	32,318	0.81	0.79	33.68	34.31	23.22		8.79
2007	3487934	ECR	2,277	0.51	0.50	43.17	51.12			5.71
2007	3564122	ECR	2,277	1.15	1.02	26.39	20.11	34.34	8.83	10.32
2007	3487633	BCR	2,277	0.62	0.49	34.74	61.66			3.60
2007	3564123	BCR	2,277	0.99	0.83	30.13	30.43	34.08		5.36
2007	3514164	ECR	32,480	0.55	0.50	39.18	55.39			5.43
2007	3564117	ECR	32,480	1.35	0.88	9.54	45.66	26.64	11.97	6.18
2007	3487939	BCR	2,277	0.41	0.49	52.48	40.71			6.81
2007	3564124	BCR	2,277	0.54	0.64	46.07	37.64	8.26		8.04
2009	3514013	BCR	30,760	0.5	0.5	47.48	50.44			2.08
2009	3564107	BCR	30,760	1.36	0.67	8.24	42.04	47.04		2.68
2009	3487680	ECR	30,760	0.31	0.46	64.81	30.59			4.61
2009	3564133	ECR	30,760	1.05	1.21	43.46	18.94	10.04	21.90	5.67
2009	3487759	BCR	30,760	0.30	0.46	65.48	30.18			4.34
2009	3564128	BCR	30,760	1.01	0.77	24.44	40.05	30.52		4.98
2009	3514117	BCR	30,760	0.41	0.49	53.15	40.69			6.16
2009	3564111	BCR	30,760	0.93	0.76	23.29	42.42	25.29		9.01
2009	3487934	ECR	30,760	0.51	0.5	43.69	51.46			4.84
2009	3564122	ECR	30,760	1.13	1.01	29.64	19.20	35.55	7.69	7.93
2009	3487633	BCR	30,760	0.67	0.47	29.05	67.41			3.54
2009	3564123	BCR	30,760	0.98	0.88	34.06	23.29	37.38		5.26
2009	3514164	ECR	30,760	0.57	0.49	39.20	57.44			3.36
2009	3564117	ECR	30,760	1.53	0.93	10.09	37.57	31.24	17.51	3.59
2009	3487939	BCR	30,760	0.44	0.5	45.97	43.69			10.34
2009	3564124	BCR	30,760	0.58	0.63	38.42	43.16	7.57		10.85



**Table 1.70 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 8 Form A**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	3	3514013	BCR_A	0.2481			
2008	4	3564107	BCR_B	-0.9490	-1.2905	1.2905	
2008	6	100000018156	SR	-0.6970			
2008	15	100000018174	SPR	0.0706			
2007	16	3487680	ECR_A	1.4000			
2007	17	3564133	ECR_B	0.8595	0.1923	0.0792	-0.2715
2006	18	3514065	SPR	-2.2105			
2008	19	100000026780	SPR	0.7914			
2008	20	100000043325	SPR	-0.4508			
2008	21	100000004108	SPR	-0.1193			
2008	23	100000012754	SR	-0.4903			
2007	24	3487759	BCR_A	1.5680			
2007	25	3564128	BCR_B	0.4439	-0.9715	0.9715	
2008	26	100000004078	SPR	0.1312			
2008	34	3514117	BCR_A	0.7094			
2008	35	3564111	BCR_B	0.5525	-0.7657	0.7657	
2007	36	3513646	SPR	0.2873			
2006	37	3514597	SPR	-2.2009			
2007	39	3487934	ECR_A	-0.2738			
2007	40	3564122	ECR_B	0.5911	-1.0085	-1.0017	2.0102
2008	43	100000049037	SR	-0.4956			
2007	44	3487633	BCR_A	-0.8219			
2007	45	3564123	BCR_B	-0.1502	-0.4588	0.4588	
2007	54	3514164	ECR_A	-0.4257			
2007	55	3564117	ECR_B	0.1001	-2.4484	0.5623	1.8861
2008	56	100000043326	SPR	0.3297			
2007	57	3514083	SPR	1.4236			
2007	72	3519734	SR	1.6876			
2008	73	100000004118	SR	-0.1435			
2007	74	3487906	SPR	1.4345			
2008	75	3500160	SPR	1.3250			
2007	76	3487939	BCR_A	0.3023			
2007	77	3564124	BCR_B	1.3855	-1.3070	1.3070	
2008	80	100000004091	SR	-0.5899			
2009	3	3514013	BCR_A	0.1810			
2009	4	3564107	BCR_B	-0.9848	-1.1835	1.1835	
2009	6	100000018156	SR	-0.7550			
2009	15	100000018174	SPR	0.0671			
2009	16	3487680	ECR_A	1.2610			
2009	17	3564133	ECR_B	0.8697	-0.111	0.3415	-0.2305
2009	18	3514065	SPR	-2.9220			
2009	19	100000026780	SPR	0.7842			

**Table 1.70 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	20	100000043325	SPR	-0.2713			
2009	21	100000004108	SPR	0.0206			
2009	23	100000012754	SR	-0.6096			
2009	24	3487759	BCR_A	1.2982			
2009	25	3564128	BCR_B	0.0944	-1.0009	1.0009	
2009	26	100000004078	SPR	0.0720			
2009	34	3514117	BCR_A	0.7346			
2009	35	3564111	BCR_B	0.2987	-1.0752	1.0752	
2009	36	3513646	SPR	0.4579			
2009	37	3514597	SPR	-2.3874			
2009	39	3487934	ECR_A	0.0273			
2009	40	3564122	ECR_B	0.9667	-0.9274	-1.0424	1.9698
2009	43	100000049037	SR	-0.5927			
2009	44	3487633	BCR_A	-0.8046			
2009	45	3564123	BCR_B	0.2204	0.0018	-0.0018	
2009	54	3514164	ECR_A	-0.2697			
2009	55	3564117	ECR_B	0.0215	-2.0897	0.3751	1.7146
2009	56	100000043326	SPR	0.3680			
2009	57	3514083	SPR	1.2566			
2009	72	3519734	SR	1.7460			
2009	73	100000004118	SR	0.0905			
2009	74	3487906	SPR	1.7189			
2009	75	3500160	SPR	1.5068			
2009	76	3487939	BCR_A	0.3153			
2009	77	3564124	BCR_B	1.5079	-1.5771	1.5771	
2009	80	100000004091	SR	0.0638			

Note. Rasch item and step difficulties are on a common scale.

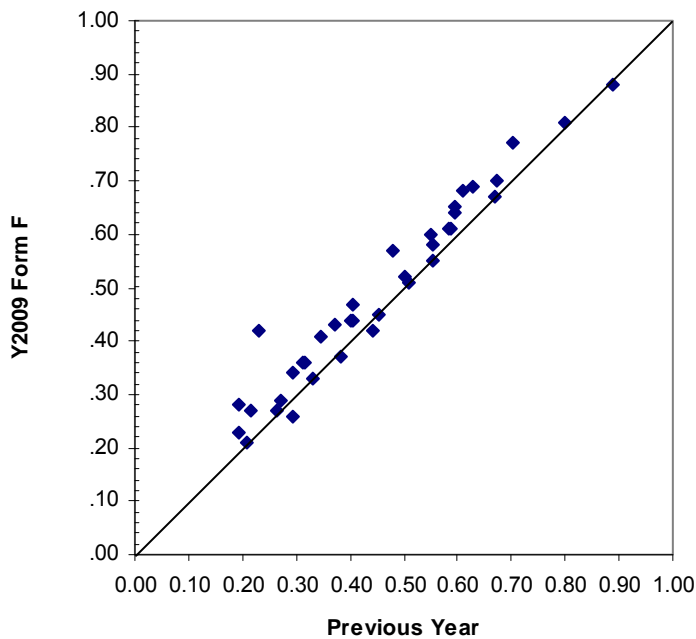


Figure 1.13 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 8 Form A

**Table 1.71 P-Value Comparisons of Core Items for Previous Year vs. Year 2009: Grade 8 Form F**

Item CID	Previous Year	Year 09 Form F	Item CID	Previous Year	Year 09 Form F
100000018154	0.63	0.69	<b>3487934</b>	<b>0.51</b>	<b>0.51</b>
<b>3500167</b>	<b>0.21</b>	<b>0.21</b>	<b>3564122</b>	<b>0.38</b>	<b>0.37</b>
<b>3514283</b>	<b>0.40</b>	<b>0.44</b>	3487545	0.67	0.7
<b>3564116</b>	<b>0.55</b>	<b>0.58</b>	<b>3519815</b>	<b>0.48</b>	<b>0.57</b>
<b>3492049</b>	<b>0.59</b>	<b>0.61</b>	<b>3564138</b>	<b>0.35</b>	<b>0.41</b>
<b>100000004107</b>	<b>0.50</b>	<b>0.52</b>	3514705	0.30	0.26
<b>100000199104</b>	<b>0.23</b>	<b>0.42</b>	<b>100000043313</b>	<b>0.59</b>	<b>0.64</b>
<b>3514161</b>	<b>0.19</b>	<b>0.23</b>	<b>3595407</b>	<b>0.70</b>	<b>0.77</b>
100000018159	0.89	0.88	<b>100000012732</b>	<b>0.26</b>	<b>0.27</b>
3488841	0.61	0.68	<b>3514167</b>	<b>0.55</b>	<b>0.60</b>
<b>3491681</b>	<b>0.19</b>	<b>0.28</b>	100000018153	0.60	0.65
<b>3564126</b>	<b>0.31</b>	<b>0.36</b>	100000004114	0.33	0.33
<b>3513650</b>	<b>0.32</b>	<b>0.36</b>	100000018151	0.55	0.55
100000026754	0.67	0.67	3487712	0.58	0.61
<b>3514117</b>	<b>0.37</b>	<b>0.43</b>	<b>100000018179</b>	<b>0.29</b>	<b>0.34</b>
<b>3564111</b>	<b>0.40</b>	<b>0.47</b>	<b>3500164</b>	<b>0.45</b>	<b>0.45</b>
<b>3492059</b>	<b>0.44</b>	<b>0.42</b>	<b>3487939</b>	<b>0.41</b>	<b>0.44</b>
<b>3514279</b>	<b>0.21</b>	<b>0.27</b>	<b>3564124</b>	<b>0.27</b>	<b>0.29</b>
			3487902	0.80	0.81

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



**Table 1.72 Score-Point Distribution Comparisons of Constructed Response Core Items for Previous Year vs. Year 2009: Grade 8 Form F**

Year	Item CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
						0	1	2	3	Omit
2008	3514283	ECR	31,743	0.40	0.49	57.00	40.29			2.71
2008	3564116	ECR	31,743	1.66	0.98	4.88	44.63	18.27	28.15	4.07
2007	3491681	BCR	2,206	0.19	0.39	74.21	19.27			6.53
2007	3564126	BCR	2,206	0.62	0.62	40.71	46.92	7.62		4.76
2008	3514117	BCR	32,318	0.37	0.48	55.99	36.99			7.02
2008	3564111	BCR	32,318	0.81	0.79	33.68	34.31	23.22		8.79
2007	3487934	ECR	2,277	0.51	0.50	43.17	51.12			5.71
2007	3564122	ECR	2,277	1.15	1.02	26.39	20.11	34.34	8.83	10.32
2007	3519815	BCR	2,223	0.48	0.50	47.86	48.13			4.00
2007	3564138	BCR	2,223	0.69	0.70	39.27	41.07	14.13		5.53
2008	100000043313	ECR	32,318	0.59	0.49	37.84	59.42			2.74
2008	3595407	ECR	32,318	2.11	1.04	6.87	17.2	22.0	49.9	3.86
2007	3487939	BCR	2,277	0.41	0.49	52.48	40.71			6.81
2007	3564124	BCR	2,277	0.54	0.64	46.07	37.64	8.26		8.04
2009	3514283	ECR	30,282	0.44	0.50	53.72	44.32			1.97
2009	3564116	ECR	30,282	1.73	0.98	4.35	44.10	18.02	31.03	2.51
2009	3491681	BCR	30,282	0.28	0.45	68.20	27.86			3.94
2009	3564126	BCR	30,282	0.73	0.62	33.57	54.31	9.35		2.77
2009	3514117	BCR	30,282	0.43	0.49	51.62	42.78			5.59
2009	3564111	BCR	30,282	0.95	0.77	23.94	41.07	26.81		8.18
2009	3487934	ECR	30,282	0.51	0.50	43.61	51.35			5.04
2009	3564122	ECR	30,282	1.12	1.00	30.07	18.72	36.11	7.01	8.08
2009	3519815	BCR	30,282	0.57	0.49	39.26	57.35			3.38
2009	3564138	BCR	30,282	0.82	0.68	29.23	50.42	15.74		4.61
2009	100000043313	ECR	30,282	0.64	0.48	33.66	63.87			2.47
2009	3595407	ECR	30,282	2.30	0.96	4.77	12.47	21.87	57.91	2.99
2009	3487939	BCR	30,282	0.44	0.5	47.28	43.97			8.74
2009	3564124	BCR	30,282	0.59	0.65	40.98	40.95	8.86		9.20

**Table 1.73 Rasch Item and Step Difficulty Comparisons of Core Items for Previous Year vs. Year 2009:  
Grade 8 Form F**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	6	100000018154	SR	-0.5902			
2007	15	3500167	SPR	1.6931			
2008	16	3514283	ECR_A	0.7105			
2008	17	3564116	ECR_B	-0.5004	-2.4388	1.4114	1.0274
2008	18	3492049	SPR	-0.3910			
2008	19	100000004107	SPR	0.0548			
2004	20	100000199104	SPR	1.1363			
2007	21	3514161	SPR	1.9454			
2008	22	100000018159	SR	-2.3300			
2007	23	3488841	SR	-0.5225			
2007	24	3491681	BCR_A	1.9208			
2007	25	3564126	BCR_B	1.3892	-1.6181	1.6181	
2008	26	3513650	SPR	1.1148			
2008	27	100000026754	SR	-0.6449			
2008	34	3514117	BCR_A	0.7094			
2008	35	3564111	BCR_B	0.5525	-0.7657	0.7657	
2008	36	3492059	SPR	0.3553			
2007	37	3514279	SPR	1.6979			
2007	39	3487934	ECR_A	-0.2738			
2007	40	3564122	ECR_B	0.5911	-1.0085	-1.0017	2.0102
2007	43	3487545	SR	-0.9210			
2007	44	3519815	BCR_A	0.0451			
2007	45	3564138	BCR_B	0.8852	-1.1130	1.1130	
2006	49	3514705	SR	1.1032			
2008	54	100000043313	ECR_A	-0.5034			
2008	55	3595407	ECR_B	-1.1346	-0.8613	0.1993	0.6621
2008	56	100000012732	SPR	1.5001			
2008	57	3514167	SPR	-0.2011			
2008	63	100000018153	SR	-0.4406			
2008	64	100000004114	SR	1.1309			
2008	66	100000018151	SR	0.0194			
2007	72	3487712	SR	-0.4925			
2008	74	100000018179	SPR	1.2452			
2007	75	3500164	SPR	0.2389			
2007	76	3487939	BCR_A	0.3023			
2007	77	3564124	BCR_B	1.3855	-1.3070	1.3070	
2007	80	3487902	SR	-1.8380			
2009	6	100000018154	SR	-0.6627			
2009	15	3500167	SPR	1.8723			
2009	16	3514283	ECR_A	0.5874			
2009	17	3564116	ECR_B	-0.5358	-2.4564	1.4718	0.9846
2009	18	3492049	SPR	-0.4824			
2009	19	100000004107	SPR	0.1624			

**Table 1.73 (continued)**

Year	Item Seq. No.	Item CID	Item Type	Item Difficulty	Step 0-1	Step 1-2	Step 2-3
2009	20	100000199104	SPR	0.6409			
2009	21	3514161	SPR	1.7775			
2009	22	100000018159	SR	-2.1690			
2009	23	3488841	SR	-0.7247			
2009	24	3491681	BCR_A	1.4238			
2009	25	3564126	BCR_B	1.1318	-1.6837	1.6837	
2009	26	3513650	SPR	1.0321			
2009	27	100000026754	SR	-0.5765			
2009	34	3514117	BCR_A	0.7346			
2009	35	3564111	BCR_B	0.2987	-1.0752	1.0752	
2009	36	3492059	SPR	0.5743			
2009	37	3514279	SPR	1.5561			
2009	39	3487934	ECR_A	0.0273			
2009	40	3564122	ECR_B	0.9667	-0.9274	-1.0424	1.9698
2009	43	3487545	SR	-0.8828			
2009	44	3519815	BCR_A	-0.2045			
2009	45	3564138	BCR_B	0.7170	-1.3017	1.3017	
2009	49	3514705	SR	1.5905			
2009	54	100000043313	ECR_A	-0.5323			
2009	55	3595407	ECR_B	-1.2176	-0.7351	0.1474	0.5877
2009	56	100000012732	SPR	1.5351			
2009	57	3514167	SPR	-0.3043			
2009	63	100000018153	SR	-0.5150			
2009	64	100000004114	SR	1.2076			
2009	66	100000018151	SR	0.0346			
2009	72	3487712	SR	-0.3805			
2009	74	100000018179	SPR	1.0936			
2009	75	3500164	SPR	0.5224			
2009	76	3487939	BCR_A	0.3153			
2009	77	3564124	BCR_B	1.5079	-1.5771	1.5771	
2009	80	3487902	SR	-1.8764			

Note. Rasch item and step difficulties are on a common scale.

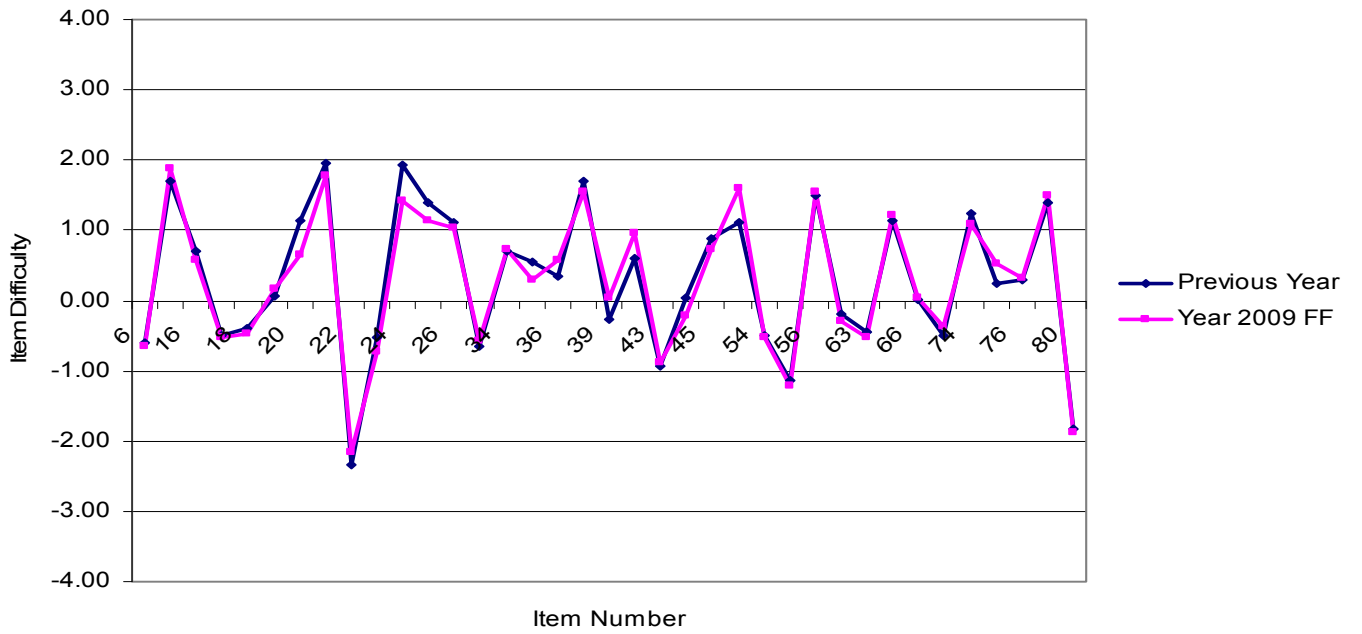


Figure 1.14 Rasch Item Difficulty Comparisons of Core Items for Previous Year vs. Year 2009: Grade 8 Form F



## 1.9 Linking, Equating, and Scaling Procedures of the 2009 MSA-Math

For the purpose of year-to-year linking and equating, we constructed a 2009 linking pool: only operational selected-response (SR) items (i.e., multiple-choice items) were included in the linking pool. It should be noted that these SR items appeared both in current and previous years' assessments and were used as either core or core link item in previous years' assessments (i.e., in any assessment before 2009). After setting up the linking pool, we conducted a stability check of linking items and decided which items should be excluded from or which items should remain in the linking pool. During the calibration and equating process, we kept and fixed the original operational Rasch item difficulty parameters of any linking items that remained through the stability check to put the 2009 assessment on a common scale. Accordingly, scale scores of the 2009 assessment were linked back to the 2006 assessment and all the scale scores of different years were comparable within each content and grade. It should be noted that Rasch recalibration was conducted using the 2006 MSA-Math data in 2007 due to the IRT model transition (i.e., from 3-PL to the Rasch). Detailed information on the 2006 Rasch recalibration and results can be obtained in the 2007 MSA-Math technical report.

### Stratified Random Sampling Procedures

To select equating samples, a stratified random sampling method was applied to the 2009 state examinee population. To verify that the sample was representative of the statewide examinee population, the distributions of LEA, gender, and ethnicity of the 2009 sample were compared with those of the 2009 population. Appendix A, *The 2009 MSA-Math Stratified Random Sampling*, provides the results of 2009 sampling. The results indicated that the equating samples were well representative of the statewide examinee population in terms of LEA, gender, and ethnicity.

### Robust Z Procedures

After selecting equating samples, each operational form was independently calibrated to estimate Rasch item difficulty of each item. Then Robust z values of all anchor items were calculated using the following calculations (South Carolina Department of Education, 2001):

- The mean and standard deviation of the linking pool's item difficulties for each operational form
- The ratio of the standard deviations between operational form A and form F
- The correlation between operational form A and F item difficulties
- The difference between operational form A and F for each item in the linking pool
- The mean of the differences calculated above
- The median of the differences calculated above
- The interquartile range of the differences calculated above
- 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 Selecting Form-to-Form or Year-to-Year Linking Items**

Once the above calculations were made, the following guidelines were followed in determining form-to-form or year-to-year common items used for Rasch linking and equating:

- Conform to the following “Protocol Criteria:” A correlation greater than 0.95 and a standard deviation ratio between 0.9 and 1.1. For example, use all the possible linking items as anchors if an original set of linking items meets these two criteria.
- Try not to include items with an absolute value of robust  $z$  exceeding 1.645.
- If one item difficulty on one form of the current year is eliminated from the linking pool, other item difficulties of the other forms should not be included.
- Should not eliminate more than 20 percent of the linking pool items.

Figure 1.15 depicts how we applied the anchor stability guidelines into the 2009 MSA-Math equating.

### **Form-to-Form Linking Procedures**

The stability of the common items appearing on both operational forms was verified at each grade level:

- Calibrate the two operational test forms separately
- Calculate robust  $z$  values of Rasch item difficulties for forms A and F
- Correlate Rasch item difficulties between form A and form F
- Calculate standard deviation ratio between two forms

After examining the robust  $z$  values, correlation coefficient, and standard deviation ratio between the two separate calibrations, 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**

The two 2009 operational forms included a set of year-to-year linking common items that appeared on both current and previous operational forms. We utilized the Rasch item fixed equating method for all of the operational items to be placed on a common scale within each grade.

The stability of the linking common items was evaluated using robust  $z$  values, correlation coefficients, and standard deviation ratios.

Tables 1.74 through 1.79 include Rasch item difficulties used for calculating robust  $z$  values, correlation coefficients, and standard deviations. Figures 1.16 through 1.27 depict item difficulty plots between current and previous years. It should be noted that the item difficulties of the 2009 operational forms were obtained from independent calibration, and those of previous assessments were on a common scale (i.e., linked to the 2006 assessment).

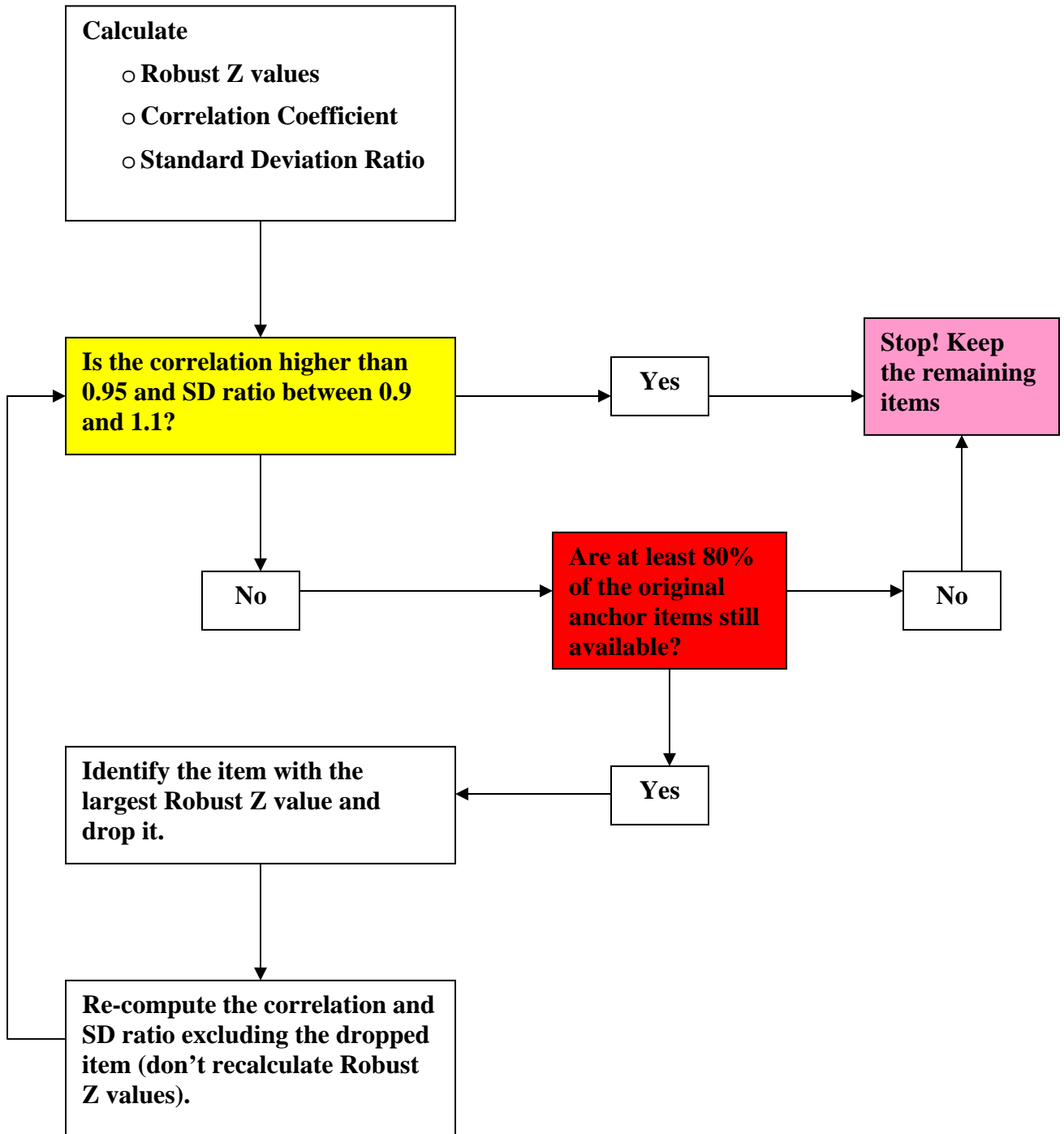


Figure 1.15 Anchor Evaluation Steps Chart for MSA-Math

**Table 1.74 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 3**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
1	0.9627	0.8909	.1368	1	0.9627	0.8673	.5730
2	-2.4386	-2.5066	.1517	2	-2.4386	-2.7836	-.1652
5	0.0690	-0.3436	-1.2010	5	0.0690	-0.3587	-.4098
7	-1.7100	-1.8678	-.2008	7	-1.7100	-2.3004	-.8909
14	-1.1315	-1.5370	-1.1731	14	-1.1315	-1.4469	-.0776
15	1.2257	1.5659	1.7541	15	1.2257	1.4639	1.5596
16	0.3981	0.1324	-.6243	16	0.3981	0.1382	.0865
17	-0.0360	0.2301	1.4632	17	-0.0360	0.1873	1.5156
21	-1.1470	-1.4508	-.7739	21	-1.1470	-1.4345	.0049
22	2.0077	1.6140	-1.1268	22	2.0077	1.5829	-.4012
23	0.4123	0.2607	-.1765	23	0.4123	0.6682	1.6120
32	0.5005	0.7292	1.3164	32	0.5005	0.5203	.9137
33	-2.6459	-3.0984	-1.3576	33	-2.6459	-3.1202	-.5476
41	0.4861	0.1676	-.8316	41	0.4861	0.0941	-.3042
45	2.4187	1.8902	-1.6560	45	2.4187	1.9468	-.5405
48	-2.1822	-1.7810	1.9936	48	-2.1822	-1.9312	1.5975
49	-1.3667	-1.3268	.5753	49	-1.3667	-1.6754	-.0578
51	0.2953	0.1403	-.1898	51	0.2953	-0.0511	-.1693
52	-0.6165	-0.5658	.6177	52	-0.6165	-0.7707	.3991
55	1.2952	1.3208	.5191	55	1.2952	1.1661	.4733
56	-0.6059	-0.7473	-.1364	56	-0.6059	-0.8967	-.0049
62	0.9229	0.7393	-.3021	62	0.9229	0.5466	-.2577
63	-0.2691	-0.3309	.1761	63	-0.2691	-0.4972	.1806
64	-1.8190	-1.4738	1.7737	64	-1.8190	-1.6845	1.2529
65	1.4814	1.6108	.9266	65	1.4814	1.5589	1.0844
66	1.8021	1.7302	.1364	66	1.8021	1.5638	.1504
67	1.5719	1.3010	-.6448	67	1.5719	1.1515	-.3882
68	0.0473	-0.2789	-.8618	68	0.0473	-0.4505	-.6171
69	0.0444	0.4099	1.8534	69	0.0444	0.1354	1.1243
72	-0.6247	-0.8357	-.4096	72	-0.6247	-1.2487	-.9903
82	-0.5397	-0.4788	.6577	82	-0.5397	-0.5998	.6774
29A	-0.3652	-.2406	.9078	8F	-0.0209	-0.6868	-1.1142
46A	-0.2784	-0.5693	-.7233	24F	-0.6271	-1.7911	-2.5874
50A	-0.4990	-0.4578	.5804	47F	-0.4817	0.3005	3.1685
				70F	0.0993	0.1655	1.0509
				80F	0.8475	0.3453	-.6301

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 students) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2009 items were independently calibrated with the 2009 stratified random sample.

**Form Statistics**

	Previous	2009	Previous	2009
Form Statistics	Base Form	Form A	Base Form	Form F
Mean	-.069	-.152	-.038	-.259
SD	1.278	1.264	1.252	1.304

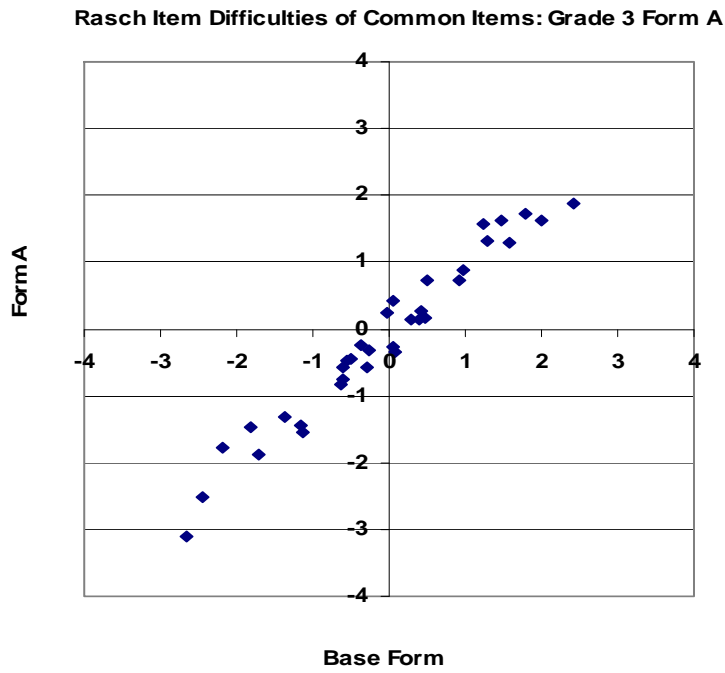
**Correlation and Standard Deviation Ratio**

	2009	2009
With Base Form	Form A	Form F
Correlation	<b>.980</b>	<b>.963</b>
SD Ratio	<b>99%</b>	<b>104%</b>

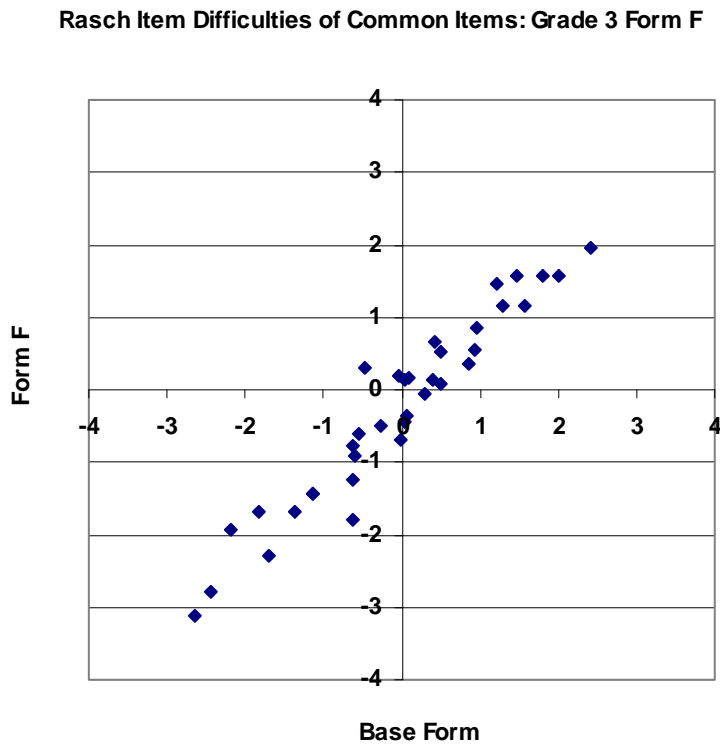
**Values Used for Robust Z Statistics**

	2009	2009
With Base Form	Form A	Form F
Mean Diff	-.083	-.221
Median Diff	-.107	-.289
IQR Diff	.344	.457

Based on correlation coefficients and SD ratios, none of the linking common items were dropped from the linking pool.



**Figure 1.16 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 3 Form A**



**Figure 1.17 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 3 Form F**

**Table 1.75 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 4**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
2	-0.7990	-0.9394	-.5298	2	-0.7990	-1.1520	-.7858
3	-1.8595	-1.8166	.2708	3	-1.8595	-1.7505	.8079
6	0.1763	-0.3630	-2.2722	6	0.1763	-0.3432	-1.3601
7	-0.8522	-0.4720	1.7441	7	-0.8522	-0.6672	1.0700
8	-1.0550	-1.2807	-.9024	8	-1.0550	-1.3842	-.7037
10	0.9009	0.7322	-.6535	10	0.9009	0.6678	-.3722
18	-1.1317	-1.2300	-.3459	18	-1.1317	-1.2499	.0241
19	1.4979	1.3332	-.6360	19	1.4979	1.1892	-.6330
22	0.3940	0.3282	-.2040	22	0.3940	0.2909	.0762
23	-0.7461	-0.5709	.8487	23	-0.7461	-0.6706	.6923
25	0.0797	0.1991	.6050	25	0.0797	0.0586	.3591
26	1.7570	1.6901	-.2088	26	1.7570	1.5140	-.4064
32	-0.9395	-0.7697	.8251	32	-0.9395	-0.9757	.3070
33	-2.7781	-2.4959	1.3161	33	-2.7781	-2.6854	.7517
34	-0.6701	-0.8879	-.8679	34	-0.6701	-1.0409	-.8472
<b>47*</b>	<b>-0.1077</b>	<b>-1.4578</b>	<b>-5.8138</b>	<b>47*</b>	<b>-0.1077</b>	<b>-1.6107</b>	<b>-4.7528</b>
49	-0.9767	-1.0012	-.0236	49	-0.9767	-0.9677	.4629
50	0.9291	1.0590	.6508	50	0.9291	1.0143	.7258
54	-0.7839	-0.7475	.2424	54	-0.7839	-0.9161	-.0241
55	-0.4674	-0.2193	1.1671	55	-0.4674	-0.2638	1.1342
62	1.0327	1.4872	2.0687	62	1.0327	1.7511	2.9100
63	-0.2743	-0.4293	-.5936	63	-0.2743	-0.5587	-.5492
64	-0.1060	-0.3060	-.7902	64	-0.1060	-0.6616	-1.4847
66	0.6282	1.3886	3.4049	66	0.6282	1.1709	2.3039
67	-0.3619	-0.2163	.7194	67	-0.3619	-0.5128	-.0887
68	-0.6898	-0.9813	-1.1898	68	-0.6898	-1.1114	-1.0224
69	0.5626	0.6502	.4661	69	0.5626	0.5354	.3381
71	-0.2943	-0.3755	-.2713	71	-0.2943	-0.5384	-.4101
78	-1.2169	-1.4517	-.9422	78	-1.2169	-1.5823	-.8286
80	-0.0118	0.2772	1.3458	80	-0.0118	0.2180	1.2246
81	-0.1831	-0.1844	.0778	81	-0.1831	-0.3494	-.1418
1A	0.7943	0.5864	-.8247	11F	1.3222	1.2926	.3298
24A	0.5508	0.5465	.0646	27F	-0.1920	-0.3944	-.2663
65A	-0.5937	-0.6074	.0236	31F	-1.6415	-1.5985	.5802
				36F	0.0868	0.2692	1.0611
				65F	0.1771	-0.3287	-1.3129

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 cases) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2008 items were independently calibrated with the 2009 stratified random sample.

**Form Statistics**

	Previous	2009	Previous	2009
Form Statistics	Base Form	Form A	Base Form	Form F
Mean	-.223	-.251	-.239	-.371
SD	.945	.998	.962	1.034

**Correlation and Standard Deviation Ratio**

	2009	2009
With Base Form	Form A	Form F
Correlation	<b>.940</b>	<b>.936</b>
SD Ratio	106%	107%

**Values Used for Robust Z Statistics**

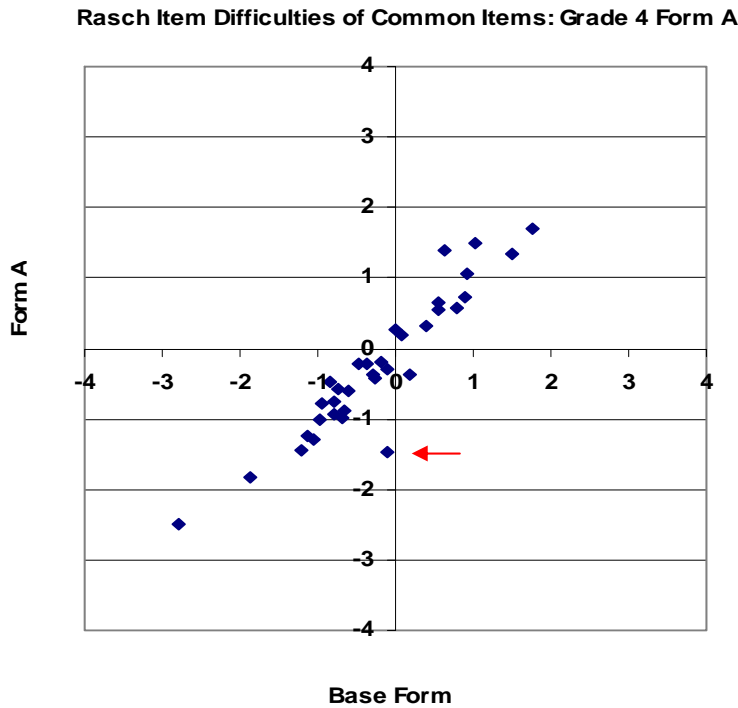
	2009	2009
With Base Form	Form A	Form F
Mean Diff	-.027	-.132
Median Diff	-.019	-.125
IQR Diff	.309	.392

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 47 appearing on both forms was dropped from the linking pool.

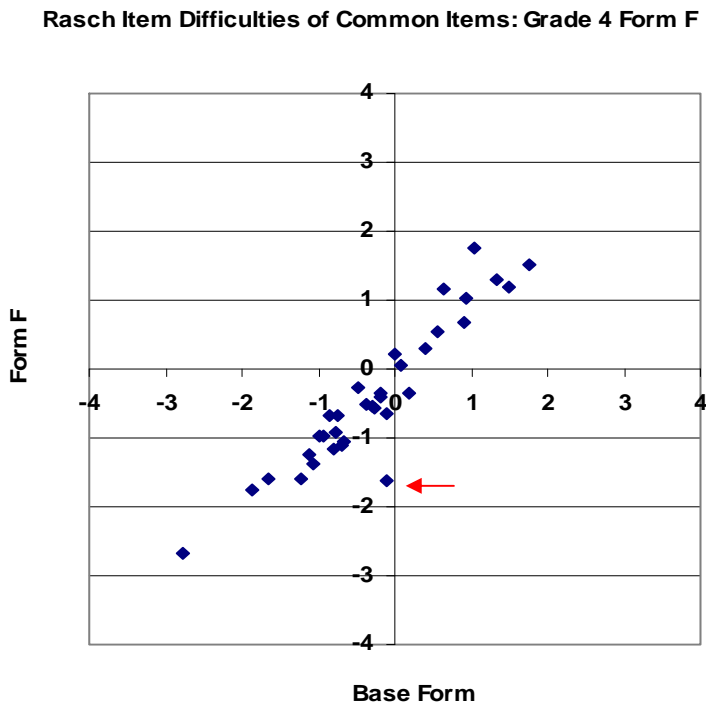
The following correlation coefficients and SD ratios were calculated after dropping that item:

	2009	2009
With Base Form	Form A	Form F
Correlation	<b>.967</b>	<b>.961</b>
SD Ratio	<b>103%</b>	<b>105%</b>





**Figure 1.18 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 4 Form A**



**Figure 1.19 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 4 Form F**

**Table 1.76 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 5**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
2	-1.0850	-1.3600	-1.1707	2	-1.0845	-1.2391	-.6121
<b>8*</b>	<b>-1.3086</b>	<b>-1.8631</b>	<b>-2.4574</b>	<b>8*</b>	<b>-1.3086</b>	<b>-2.1339</b>	<b>-3.0660</b>
16	0.6094	0.7629	.8077	16	0.6094	0.8464	.8207
18	0.1790	0.0949	-.2880	18	0.1790	0.0446	-.5382
19	-0.9090	-1.0086	-.3581	19	-0.9093	-0.8966	.0000
20	0.4459	0.7249	1.3865	20	0.4459	0.6939	.8609
21	0.4633	0.3471	-.4360	21	0.4633	0.4054	-.2583
23	0.3350	0.1300	-.8456	23	0.3350	0.0919	-.9359
26	0.2030	0.1878	.0297	26	0.2030	0.5211	1.1174
27	0.2606	0.0602	-.8243	27	0.2606	-0.1852	-1.6775
37	-0.2850	-0.4292	-.5647	37	-0.2851	-0.5397	-.9780
38	-0.3310	-0.5894	-1.0918	38	-0.3310	-0.4252	-.3911
<b>39*</b>	<b>-1.7040</b>	<b>-1.7267</b>	<b>-.0039</b>	<b>39*</b>	<b>-1.7042</b>	<b>-2.3770</b>	<b>-2.5081</b>
42	0.1548	0.2672	.6182	42	0.1548	0.4121	.8949
43	-1.1290	-1.0875	.2926	43	-1.1293	-1.1610	-.1624
47	0.0148	0.1865	.8917	47	0.0148	0.1432	.4233
48	-0.2130	-0.2227	.0551	48	-0.2130	-0.1631	.1361
49	-0.6900	-0.9724	-1.2034	49	-0.6898	-1.0759	-1.4591
55	-0.6830	-0.4623	1.1167	55	-0.6828	-0.4448	.8243
56	-1.7930	-1.9641	-.6901	56	-1.7928	-1.8245	-.1624
58	-0.9440	-0.8536	.5163	58	-0.9439	-0.9003	.1131
60	-1.1520	-1.1159	.2645	60	-1.1516	-0.9917	.5386
61	-0.5030	-0.4589	.3009	61	-0.5025	-0.2812	.7632
64	-0.9260	-1.1646	-1.0005	64	-0.9260	-1.2516	-1.2377
70	-0.3860	-0.5802	-.7948	70	-0.3862	-0.5892	-.7892
71	0.5581	0.4044	-.6090	71	0.5581	0.5058	-.2378
72	-0.5780	-0.4509	.6855	72	-0.5779	-0.3567	.7628
82	-0.0720	-0.0925	.0039	82	-0.0717	-0.0411	.0655
83	-0.6840	-0.4320	1.2615	83	-0.6839	-0.5336	.5034
28A	0.1746	0.4756	1.4880	40F	1.2809	1.3850	.3344
34A	1.7536	1.6853	-.21510	<b>44F*</b>	<b>2.4411</b>	<b>3.1433</b>	<b>2.5227</b>
44A	0.8010	1.0378	1.1919				

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 cases) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2009 items were independently calibrated with the 2009 stratified random sample.

**Form Statistics**

Form Statistics	Previous Base Form	2009 Form A	Previous Base Form	2009 Form F
Mean	-.294	-.327	-.272	-.297
SD	.781	.855	.880	1.072

**Correlation and Standard Deviation Ratio**

	2009 Form A	2009 Form F
with Base Form		
Correlation	.975	.970
SD Ratio	109%	<b>122%</b>

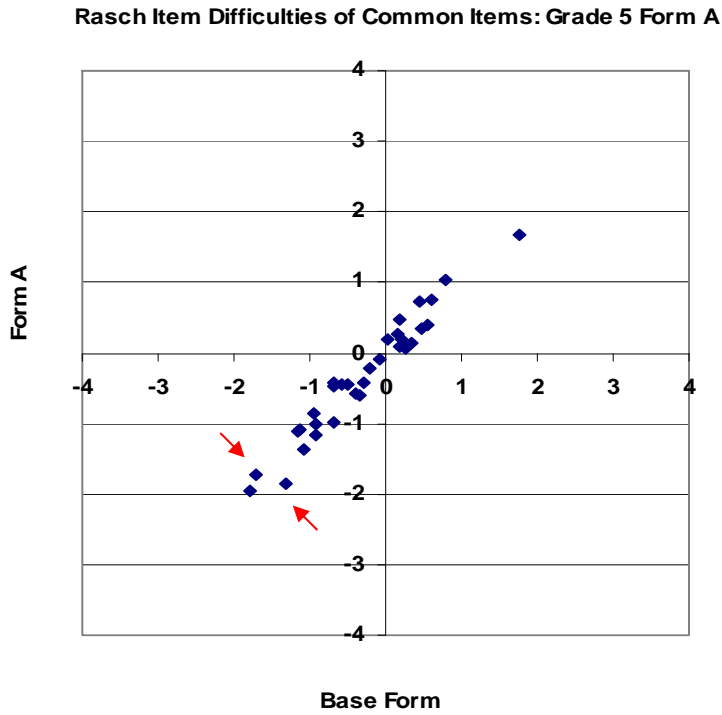
**Values Used for Robust Z Statistics**

	2009 Form A	2009 Form F
With Base Form		
Mean Diff	-.033	-.026
Median Diff	-.022	.013
IQR Diff	.293	.369

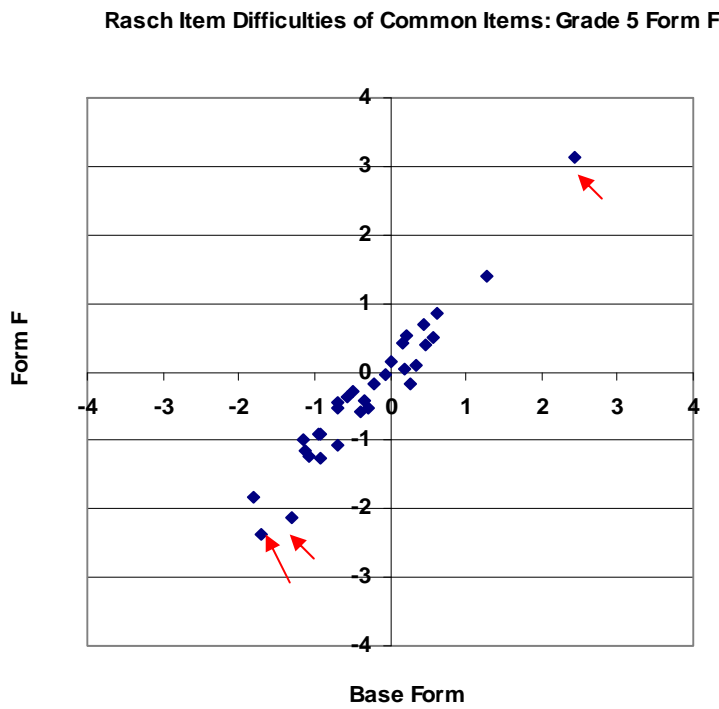
Item number 8 and 39 on both forms were dropped from the linking pool based on correlation coefficients, SR ratios, robust z, and item difficulty plot. In addition, item number 44 appearing only on Form F was dropped from the linking pool.

The following correlation coefficients and SD ratios are based on dropping those items:

	2009 Form A	2009 Form F
With Base Form		
Correlation	<b>.975</b>	<b>.960</b>
SD Ratio	<b>107%</b>	<b>108%</b>



**Figure 1.20 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 5 Form A**



**Figure 1.21 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 5 Form F**

**Table 1.77 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 6**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
<b>1*</b>	<b>-1.2053</b>	<b>-1.7693</b>	<b>-2.3677</b>	<b>1*</b>	<b>-1.2053</b>	<b>-1.6311</b>	<b>-1.3638</b>
3	0.6406	0.7691	.6569	3	0.6406	0.7010	.4715
6	-0.2844	-0.2783	.1223	6	-0.2844	-0.4437	-.3578
<b>9*</b>	<b>-0.7278</b>	<b>-1.2723</b>	<b>-2.2825</b>	<b>9*</b>	<b>-0.7278</b>	<b>-1.0840</b>	<b>-1.1011</b>
<b>10*</b>	<b>-1.4432</b>	<b>-2.0707</b>	<b>-2.6450</b>	<b>10*</b>	<b>-1.4432</b>	<b>-2.0368</b>	<b>-1.9972</b>
11	-0.4703	-0.5566	-.2813	11	-0.4703	-0.5429	-.0306
12	0.3254	0.2112	-.4031	12	0.3254	0.1825	-.2959
19	0.2409	0.0116	-.9059	19	0.2409	-0.0208	-.7444
20	0.4042	0.4881	.4621	20	0.4042	0.5394	.7538
25	0.4777	0.9367	2.1004	25	0.4777	0.9356	1.9719
26	-0.1396	-0.1893	-.12140	26	-0.1396	-0.2041	.0000
27	0.2101	0.2381	.2179	27	0.2101	0.2458	.3782
30	0.7262	0.6486	-.2433	30	0.7262	0.7374	.2857
34	1.1378	0.9123	-.8893	34	1.1378	0.9081	-.6236
35	-1.4702	-1.3927	.4341	35	-1.4702	-1.5113	.0883
36	0.3674	0.3455	.0000	36	0.3674	0.2637	-.1480
<b>37*</b>	<b>0.5144</b>	<b>-0.2472</b>	<b>-3.2308</b>	<b>37*</b>	<b>0.5144</b>	<b>0.1394</b>	<b>-1.1721</b>
38	-0.1849	-0.2188	-.0524	38	-0.1849	-0.2098	.1495
45	-0.9261	-0.8525	.4171	45	-0.9261	-0.7479	.9161
50	0.0810	-0.2742	-1.4557	50	0.0810	-0.2682	-1.0747
54	0.2864	0.4210	.6835	54	0.2864	0.1854	-.1378
55	0.5885	0.8930	1.4256	55	0.5885	0.7904	1.0056
56	0.1350	0.2684	.6783	56	0.1350	0.3146	.9214
57	-0.4092	-0.2116	.9587	57	-0.4092	-0.1570	1.1955
61	0.2607	0.3220	.3634	61	0.2607	0.3890	.7278
68	0.4071	0.5731	.8207	68	0.4071	0.5706	.8606
80	0.6580	0.7211	.3713	80	0.6580	0.6756	.3099
<b>2A*</b>	<b>-0.3185</b>	<b>-1.0116</b>	<b>-2.9316</b>	<b>2F*</b>	<b>-0.7946</b>	<b>-1.4436</b>	<b>-2.2063</b>
29A	0.8496	0.7664	-.2677	4F	-0.6658	-0.7634	-.1249

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 cases) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2009 items were independently calibrated with the 2009 stratified random sample.

**Form Statistics**

Form Statistics	Previous Base Form	2009 Form A	Previous Base Form	2009 Form F
Mean	.025	-.063	-.043	-.120
SD	.675	.824	.681	.816

**Correlation and Standard Deviation Ratio**

With Base Form	2009 Form A	2009 Form F
Correlation	<b>.938</b>	.958
SD Ratio	<b>122%</b>	<b>120%</b>

**Values Calculated for Robust Z Statistics**

With Base Form	2009 Form A	2009 Form F
Mean Diff	-.088	-.077
Median Diff	-.022	-.065
IQR Diff	.309	.358

Based on correlation coefficients, SD ratio, robust z, and item difficulty plot, item numbers 1, 9, 10, and 37 appearing on both forms were dropped from the linking pool. In addition, the unique core linking item in position on 2 on Form A and the unique core linking item in position 2 on Form F were dropped from the linking pool (although these linking items appeared in the same position on each form they are unique items).

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

With Base Form	2009 Form A	2009 Form F
Correlation	<b>.956</b>	<b>.954</b>
SD Ratio	<b>103%</b>	<b>105%</b>

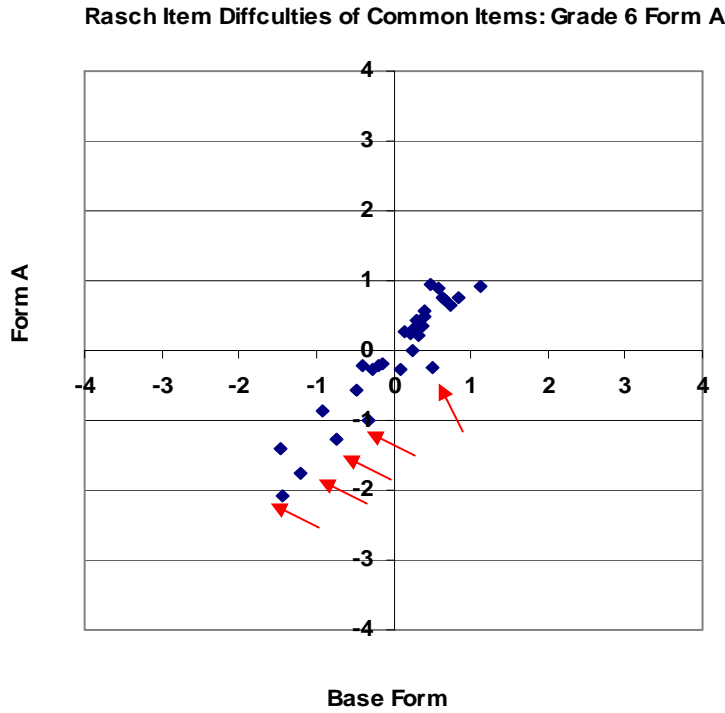


Figure 1.22 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 6 Form A

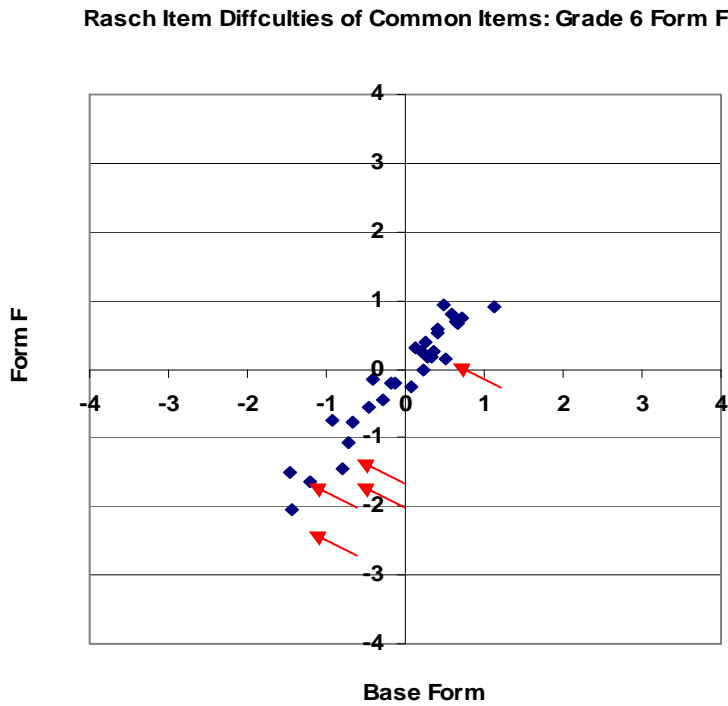


Figure 1.23 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 6 Form F

**Table 1.78 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 7**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
1	1.0539	1.3179	1.9392	1	1.0539	1.2073	2.8523
2	0.4455	0.1965	-.4505	2	0.4455	0.1233	-.4268
3	0.1508	-0.1525	-.7034	3	0.1508	-0.1628	-.3675
7	-0.1398	-0.5871	-1.3742	7	-0.1398	-0.5428	-.9839
8	-0.4706	-0.7600	-.6386	8	-0.4706	-0.7637	-.2261
10	-1.1551	-1.3149	-.0349	10	-1.1551	-1.3943	.1455
12	-0.4683	-0.6507	-.1402	12	-0.4683	-0.7150	.0938
18	-0.6359	-1.0764	-1.3425	18	-0.6359	-1.2132	-2.1856
19	-1.1243	-1.7113	-2.0249	19	-1.1243	-1.8325	-3.0881
20	1.5825	1.2198	-.9801	20	1.5825	1.2638	-.4026
30	-0.5147	-0.1199	2.5485	30	-0.5147	-0.0428	5.0483
31	-2.6820	-3.1414	-1.4305	31	-2.6820	-3.2279	-1.9691
32	0.0227	-0.1181	.0536	32	0.0227	-0.2249	.0876
42	1.3415	1.3879	.9256	42	1.3415	1.3091	1.5713
43	-0.4094	-0.5108	.2371	43	-0.4094	-0.6188	.3509
50	1.4392	1.4118	.5818	50	1.4392	1.2160	.2558
51	-0.0583	-0.1570	.2497	51	-0.0583	-0.2511	.4654
52	-1.4991	-1.8211	-.7905	52	-1.4991	-1.8879	-.8860
63	0.5663	0.5770	.7593	63	0.5663	0.373	.4619
64	0.0092	-0.1324	.0498	64	0.0092	-0.2843	-.2289
65	-0.4333	-0.6919	-.4952	65	-0.4333	-0.6936	.0000
66	-0.2963	-0.8359	-1.8041	66	-0.2963	-0.9359	-2.6151
69	0.5231	0.3113	-.2772	69	0.5231	0.1077	-1.0694
70	-0.2784	-0.4232	.0349	70	-0.2784	-0.7068	-1.1590
72	0.6673	0.7031	.8762	72	0.6673	0.4886	.5626
79	-1.4603	-1.3173	1.3756	79	-1.4603	-1.3477	2.5710
80	-0.5723	-0.5045	1.0253	80	-0.5723	-0.656	1.2176
81	-0.0385	-0.1187	.3359	81	-0.0385	-0.0600	1.6464
				44F	1.7720	1.4161	-.6591

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 cases) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2009 items were independently calibrated with the 2009 stratified random sample.



**Form Statistics**

Form Statistics	Previous Base Form	2009 Form A	Previous Base Form	2009 Base Form
Mean	-.158	-.322	-.092	-.347
SD	.947	1.041	.996	1.070

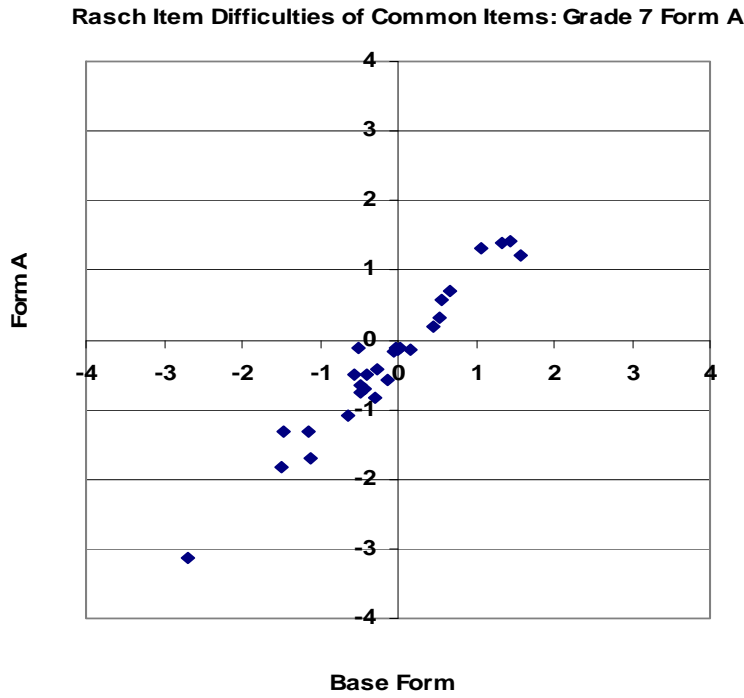
**Correlation and Standard Deviation Ratio**

With Base Form	2009 Form A	2009 Form F
Correlation	<b>.977</b>	<b>.975</b>
SD Ratio	<b>110%</b>	<b>107%</b>

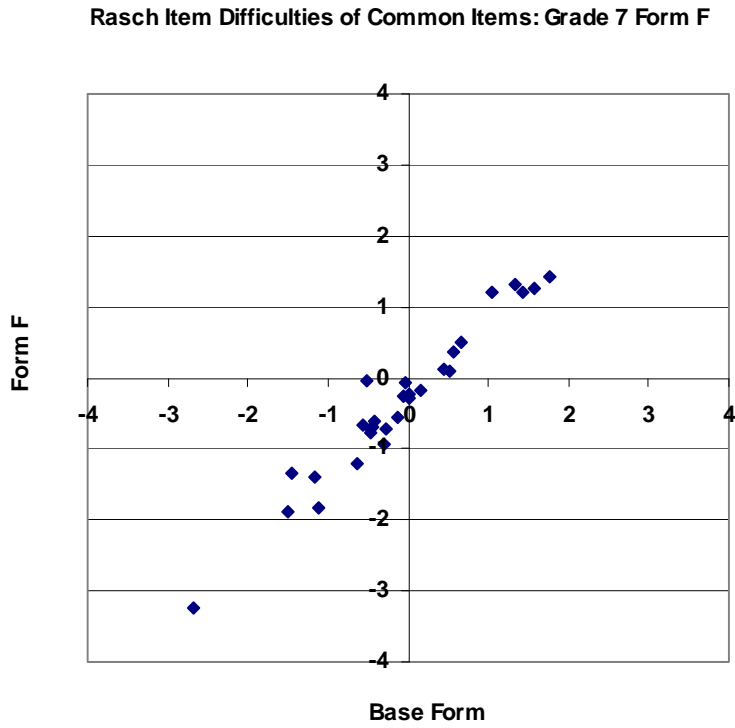
**Values Used for Robust Z Statistics**

With Base Form	2009 Form A	2009 Form F
Mean Diff	-.164	-.255
Median Diff	-.152	-.260
IQR Diff	.290	.196

Based on correlation coefficients and SD ratios, none of the linking common items were dropped from the linking pool.



**Figure 1.24 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 7 Form A**



**Figure 1.25 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 7 Form F**

**Table 1.79 Rasch Item Difficulties and Robust Z values for Previous Year vs. Year 2009: Grade 8**

Item Seq No.	Previous Form A	Y2009 Form A	Robust Z	Item Seq No.	Previous Form F	Y2009 Form F	Robust Z
1	1.4965	1.4366	.3019	1	1.4965	1.4069	.4767
2	-0.2177	-0.2601	.4482	2	-0.2177	-0.3646	-.2079
5	-1.3613	-1.6916	-1.9593	5	-1.3613	-1.8797	-4.6467
7	-1.2003	-1.4748	-1.4927	7	-1.2003	-1.4356	-1.2641
8	0.3621	0.2940	.2333	8	0.3621	0.1906	-.5018
14	0.3158	0.1330	-.7258	14	0.3158	0.1960	.1159
32	1.0306	0.9528	.1522	32	1.0306	0.9998	1.1793
33	0.5139	0.3732	-.3738	33	0.5139	0.3521	-.3859
38	-1.4579	-1.3942	1.3355	38	-1.4579	-1.3088	3.3288
41	0.5661	0.0594	-3.4344	41	0.5661	0.1094	-3.9095
42	-1.4001	-1.7531	-2.1491	42	-1.4001	-1.6853	-1.8603
46	-0.2581	-0.3520	.0176	46	-0.2581	-0.3602	.3274
47	-0.1085	-0.0427	1.3530	47	-0.1085	-0.0781	1.9105
48	-0.6178	-0.6275	.7217	48	-0.6178	-0.6889	.6978
50	0.0551	-0.0417	-.0067	50	0.0551	-0.0415	.3931
51	-0.7102	-0.7862	.1672	51	-0.7102	-0.7696	.8376
52	0.3257	0.3439	.9550	52	0.3257	0.2765	.9594
53	-0.6275	-0.8368	-.9475	53	-0.6275	-0.7604	-.0406
62	1.2102	1.0490	-.5452	62	1.2102	1.0250	-.6655
65	-0.5330	-0.9743	-2.8875	65	-0.5330	-1.1330	-5.6217
78	-0.0934	-0.1124	.6439	78	-0.0934	-0.2229	.0000
79	-0.1424	-0.3442	-.8847	79	-0.1424	-0.3016	-.3549
22A	-0.8540	-1.0332	-.6957	73F	1.0119	1.3186	5.2118
27A	0.2581	0.3247	1.3597				
49A	0.8435	0.7475	.0000				
63A	2.0087	2.0240	.9307				
66A	1.3321	1.1916	-.3721				

*Note.* The 2009 item sequence number was used to indicate that it was the same item appearing across years.

*Note.* Each item parameter was generated with a live, stratified random sample (i.e., about 3,000 cases) of the year.

*Note.* Item parameters of previous years were on a common scale.

*Note.* The 2009 items were independently calibrated with the 2009 stratified random sample.

**Form Statistics**

	Previous	2009	Previous	2009
Form Statistics	Base Form	Form A	Base Form	Form F
Mean	.027	-.104	-.080	-.224
SD	.922	.969	.844	.914

**Correlation and Standard Deviation Ratio**

	2009	2009
With Base Form	Form A	Form F
Correlation	<b>.989</b>	<b>.978</b>
SD Ratio	<b>105%</b>	<b>108%</b>

**Values Used for Robust Z Statistics**

	2009	2009
With Base Form	Form A	Form F
Mean Diff	-.131	-.144
Median Diff	-.096	-.130
IQR Diff	.162	.113

Based on correlation coefficients and SD ratios, none of the linking common items were dropped from the linking pool.

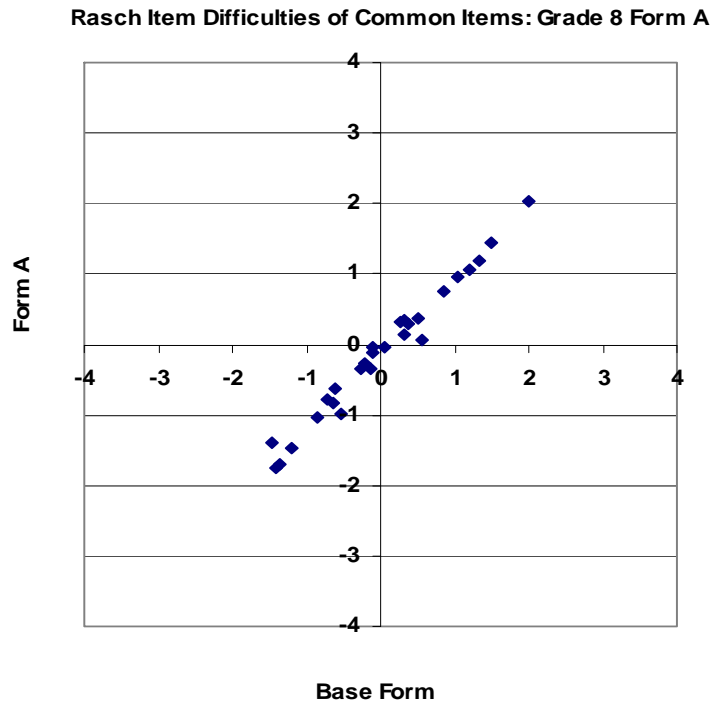


Figure 1.26 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 8 Form A

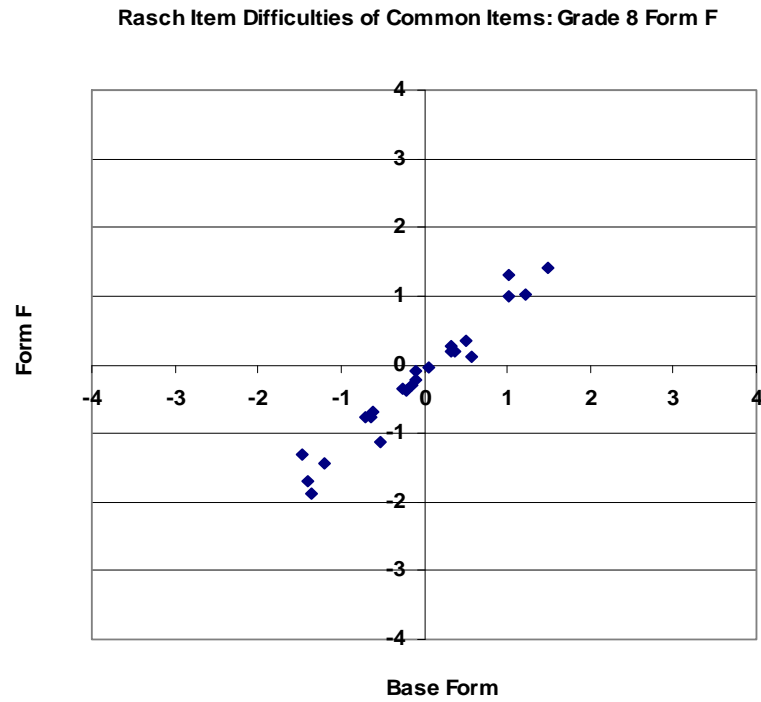


Figure 1.27 Item Difficulty Plot of Previous Year Form (Base Form) vs. Current Year (2009) Form: Grade 8 Form F

### Reporting Scale Scores

In order to facilitate the use and interpretation of the results of the 2009 MSA-Math, the following formula was used to convert each student's ability or theta to the reporting scale score:

$$\text{ReportingAbilityScaleScore} = 32.8398 \cdot \text{theta} + 380.2954$$

$$\text{ReportingSE} = 32.8398 \cdot \text{SE}$$

where

theta = the Rasch (i.e., 1-PL IRT) ability estimate, and

SE = the conditional standard error of the ability estimate.

The following table contains information about the slopes and intercepts used to generate the 2009 scale scores. First of all, it should be noted that the slopes and intercepts were obtained during the 2006 recalibration. The same slopes and intercepts have been used since the 2006 assessment.

**Table 1.80 The 2009 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

## 1.10 Score Interpretation

To help provide appropriate interpretation of the 2009 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 1.9, *Linking, Equating, and Scaling Procedures*, the 2009 scale scores were placed on a common scale (i.e., 2006 assessment) within the same grade and ranged from 240 to 260. As a result, these scale scores have the same meaning and are comparable across different years' assessments. However, it should be noted that they are not comparable across grade levels.

For scale scores, a higher score simply means a higher performance on 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 2009 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.marylandpublicschools.org](http://www.marylandpublicschools.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.

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 the MSA-Math standard. This level is considered a realistic and rigorous level of achievement. Further information about the 2009 MSA-Math score interpretation can be obtained from MSDE.

## 1.11 Test Validity of the 2009 MSA-Math

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 ongoing 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 2009 MSA-Math, content-related evidence, item development procedures, differential item functioning (DIF) analysis on 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 2009 MSA-Math blueprints provide extensive evidence regarding the alignment between the content in the 2009 MSA-Math and the *VSC*. It should be noted that the 2009 MSA-Math operational test forms were built exclusively using a Maryland item bank program which contained both content and statistical information about both operational and field-tested items. Information on the item composition of the operational test forms can be obtained from section 1.4, *Test Form Design, Specifications, Item Type, and Item Roles*. In addition, the 2009 MSA-Math blueprints are presented in Appendix D.

### Item Development

Test development for MSA-Math is ongoing and continuous. Content specialists, teachers from across Maryland, Pearson, and MSDE were greatly involved in developing and reviewing items. Committees such as content review, bias review, and vision review reviewed all of the items, which were finally stored in a Maryland item bank. Specifically, an internal review by MSDE and Pearson staff for content 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 1.3, *Development and Review of the 2009 MSA-Math Items and Test*.



Field test items were embedded and administered in one of ten test forms. Once these items were scored, MSDE and Pearson conducted additional item analysis and content review. Any field test items that exhibited statistical results that suggested potential problems were carefully reviewed by both MSDE and Pearson content specialists. A determination was then made as to whether an item should be eliminated, revised, or field-tested again. Information on statistical analyses for field test items can be obtained in section 1.13, *Field Test Analyses and Item Bank Construction*.

### ***Differential Item Functioning (DIF)***

#### 1) Bias Review of Items

A separate Bias Review Committee examined each math item, with looking for indications of bias that could 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 according to either reference or focal groups. For the 2009 MSA-Math, males and whites were assigned to the reference group and females and African-Americans were assigned to the focal group.

While the Mantel-Haenszel procedure was used for SR and SPR items, the standardized mean difference (SMD) and the standard deviation (SD), along with the Mantel statistic, were calculated for BCR and ECR items. All of the items were classified based on Educational Testing Service (ETS) guidelines. All *DIF* results were kept in the 2009 Maryland item bank. More information on *DIF* analyses can be obtained in section 3.7, *Differential Item Functioning*.

### **Evidence from Internal Structure**

The 2009 MSA-Math has five reporting math standards: *Algebra, Geometry and Measurement, Statistics and Probability, Numbers and Computations, and Process*. Tables 4.3 through 4.8 show the correlations among the math standards.

## 1.12 Unidimensionality Analyses of the 2009 MSA-Math

Measurement implies order and magnitude along a single dimension (Andrich, 1989). Consequently, in the case of scholastic achievement, one-dimensional scale is required to reflect this idea of measurement (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 2009 MSA-Math, we examined the relative sizes of the eigenvalues associated with a principal component analysis of the item set. First, polychoric correlation coefficients were computed with *LISREL 8.5* (Jöreskog & Sörbom, 1993) because they were polytomously scored on math items. Principal component analysis was then applied to produce eigenvalues. The first and the second principal component eigenvalues were compared *without rotation*. Table 1.81 summarizes the results of the first and second principal component eigenvalues of the 2009 MSA-Math.

A general rule of thumb in exploratory factor analysis suggests that a set of items may represent as many factors as there are eigenvalues greater than 1 in this analysis because there is one unit of information per item and the eigenvalues sum to the total number of items. However, a set of items may have multiple eigenvalues greater than 1 and still be sufficiently unidimensional for analysis with IRT (Loehlin, 1987; Orlando, 2004). As seen from the following table, the first component extracted substantially larger eigenvalues across all grades: the size of the eigenvalue of the first component was over ten times greater than the second eigenvalue for each form at each grade. As a result, we could conclude that the assumption of unidimensionality for the 2009 MSA-Math was met.

**Table 1.81 The 2009 MSA-Math Eigenvalues between the First and Second Components**

Grade	Form	Number of Items	First Eigenvalue	Second Eigenvalue
3	A	65	23.95	2.34
	F	65	23.31	2.47
4	A	64	23.18	2.07
	F	64	23.07	1.92
5	A	65	23.11	1.99
	F	65	23.05	2.28
6	A	62	22.28	2.05
	F	62	21.46	1.74
7	A	62	25.12	2.08
	F	62	25.46	1.93
8	A	61	24.18	2.25
	F	60	21.97	2.06

*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.* Analysis was conducted with a statewide population.

### 1.13 Field Test Analyses and Item Bank Construction

All field test items embedded in operational forms were subjected to rigorous statistical analyses for their properties in order to provide information about which items may be included as operational items in the future. All statistical results concerning field test items were preserved in the 2009 Maryland item bank. 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 any students (i.e., nonfunctional distractor)
- An item was selected by a high proportion of high-ability students while being selected by a low proportion of low-ability students (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 miskeyed answer.

*BCR* and *ECR* items 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 (i.e., adjusted *p*-value) was less than .20 or greater than .90.
- An item-total correlation was less than .10.

All items required 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 being dropped as a possible operational item. However, if the item represented important content that had not been extensively taught, a justification could have been made for including it in an 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 reference group was Caucasians. The gender focal group was females and the ethnic focal group was African-Americans. For each operational form, the student's total score was used as the matching variable.

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 whether the differential difficulty of the item was unfairly related to group membership using the following criteria:

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

It should be noted that *DIF* analysis results for all the field test items were archived in the 2009 Maryland item bank. Detailed information about the *DIF* procedures can be found in section 3.7, *Differential Item Functioning*.

### ***Item Response Theory (IRT) Analyses***

To put the 2009 field test items on a common scale (i.e., the 2006 scale), each field test item was freely calibrated after fixing Rasch item and step difficulty parameters of the 2009 operational items that had been already placed on the base scale during the 2009 operational calibration and equating. For example, each unique field test item appearing on one of five math test forms (i.e., A, B, C, D, and E) was independently calibrated after fixing the same operational items appearing across the field test forms with the same Rasch item and step difficulties because these unique field test forms all correspond to the same operational form (i.e., operational form A).

It should be noted that all the Rasch item difficulties, step difficulties, and fit statistics (i.e., Rasch Infit and Outfit indices) of the field test items were archived in the 2009 Maryland item bank. These field test items are eligible to be used as operational items in subsequent years.

### ***Item Bank Construction***

The number of test forms constructed each year and the need to replace items that are released to the public necessitates the availability of a large pool of items. The 2009 MSA-Math item bank continues to be maintained by Pearson in the form of computer files and paper copies. This enables the test items to be readily available to both Pearson and MSDE staff for reference, test construction, test book design, and printing.

## 1.14 Quality Control Procedures

A standard quality procedure at Pearson 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:

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

- Verified data descriptions for accuracy against data file
- Created 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 could not 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 Pearson.

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 Pearson staff experts prior to release of the information to MSDE.

Upon completion of the processing of all district-level data, Pearson Scoring Operations provided the Quality Assurance Department with one or more state-level data files, along with state data for review and approval. Pearson Quality Assurance programmers duplicated all data independently to ensure accurate interpretation of the expected results. A series of SAS programs were run on these files to ensure 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 were 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.