

## 1. OVERVIEW OF THE 2009 MARYLAND SCHOOL ASSESSMENT-READING

---

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 what students were expected to learn in schools.

In 2003, the MSA-Reading 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 (for grades 3, 5, and 8) or 2004 (for grades 4, 6, and 7).

Until 2007 the MSA-Reading was administered along with the *Stanford Achievement Test Series, Tenth Edition (SAT10)*, and the SAT10 common items aligned to the Maryland curriculum were used exclusively for the purpose of form-to-form and year-to-year linking. In 2007, however, MSDE implemented an important action plan on MSA-Reading test: dropping all of the SAT10 items from the 2008 assessment. Due to this decision, MSDE and Pearson team members examined options to replace the SAT10 items removed from the test. The minimum requirement was to develop enough items to cover the same total and subtotal score points that SAT10 common items contributed in previous years (for grade 5, for example, 45 total score points with 15 points each for general reading, literary, and informational reading). In addition, it was decided that only one operational form would be developed for the 2008 administration. More detailed information about the test and equating design changes of the 2008 administration can be found in section 1.11 of the 2008 MSA-Reading technical report, *Constructing the 2008 MSA-Reading Operational Forms*.

For the 2009 reading assessment, MSDE decided to develop and administer two operational test forms in each grade to maintain a high level of test security. To implement this plan, MSDE and Pearson team members decided to place two sets of literary and informational passages in sessions 2 and 3 of the first day of the reading test. Detailed information about the test sessions and timing can be found in the 2009 MSA-Reading Examiners Manual (EM) which is available from either MSDE or Pearson.

For the purposes of year-to-year linking and equating, we first constructed a 2009 linking pool which included only operational selected-response items (i.e., multiple-choice items). These items appeared both in 2009 and in 2007. After setting up the linking pool, we then 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 processes, we kept and fixed the original field test Rasch item difficulty parameters of any linking items (i.e., 2007 assessment) that remained through the stability check to put the 2009 assessment on a common scale. Accordingly, all scale scores of the 2009 assessment were comparable within each grade

since all the scale scores were linked back to the 2003 (for grades 3, 5, and 8) and 2004 (for grades 4, 6, and 7).

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

By measuring students' achievement against the new academic standards, the 2009 MSA-Reading fulfills two main purposes. First, the MSA-Reading 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-Reading 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-Reading:

#### **STANDARD 1.0**

##### **General Reading Process**

##### **TOPIC:**

B. VOCABULARY: Students will apply their knowledge of letter/sound relationships and word structure to **decode** unfamiliar words

##### **INDICATOR:**

1. Use a variety of phonetic skills to read unfamiliar words

##### **OBJECTIVES:**

a. Apply phonics skills

**Assessment limits:**

- Hard and soft consonants
- Initial consonant blends (2 letters)
- Open and closed syllables
- Digraphs

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-Reading Blueprint*.

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

As seen in Table 1.1, the development of the 2009 MSA-Reading 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-Reading program. For example, they made recommendations to MSDE on issues, such as test blueprints, field test design, item analysis, item selection for scoring purposes, linking, equating and scaling issues, standard setting, and other relevant statistical and psychometric issues. MSDE adopted their guidelines and recommendations.

#### **Content Review Committee**

Content Review Committee members ensured that the MSA-Reading was appropriately difficult and fair. Committee members were either specialists in reading 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, passage and art on reading 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 passages, art, and items for bias to the visually impaired. The committee makes their recommendations to NOT put any item they had a concern with on Form 1.

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

Development of the 2009 MSA-Reading	Primary Responsibility
Development of Preliminary Blueprints and Item Specifications	Pearson; MSDE; NPC
Development of Preliminary Brief Constructed Response Rubrics	MSDE; NPC
Item Writing	Pearson; MSDE
Item Review	Pearson; MSDE; Content Review Committee
Bias Review	Pearson; MSDE; Bias Review Committee
Vision Review	Pearson; MSDE; Vision Review Committee
Construction of Field Test Forms	Pearson; MSDE
Modification of Special Forms	Pearson; MSDE
Review of Special Forms	MSDE
Pre-Field Test Training Workshops	Pearson; MSDE; LEAs
Field Test Administrations	MSDE; LEAs
Construction of Operational Test Forms	Pearson; MSDE; NPC
Review of Operational Test Forms	MSDE
Final Construction of Operational Test Forms	Pearson; MSDE

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

The MSA-Reading 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 in Table 1.2, Forms 1, 3, 5, 7 and 9 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 2, 4, 6, 8, and 10 (designated as operational Form B).

### Test Form Specifications and Reporting Category

Tables 1.3 through 1.8 provide information on the total number of operational items included in the 2009 operational test forms 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.

Specifically, each standard was used for reporting purposes (i.e., reporting subscale scores). That is, there were three reporting standards for reading across grades: General Reading, Literary, and Informational Processes. The number of raw score points for each reporting standard was identical (i.e., 15) for all grades except for grades 3 and 8.

**Table 1.2 The 2009 MSA-Reading Test Form Design: Grades 3 through 8**

	Operational Item Sets		Field Test Item Sets									
	A	B	1	2	3	4	5	6	7	8	9	10
Form 1	X		X									
Form 2		X		X								
Form 3	X				X							
Form 4		X				X						
Form 5	X						X					
Form 6		X						X				
Form 7	X								X			
Form 8		X								X		
Form 9	X										X	
Form 10		X										X

*Note.* Forms 1, 3, 5, 7, and 9 (Form A) are identical, and Forms 2, 4, 6, 8, and 10 (Form B) are identical in terms of operational test items.

**Item Types**

The 2009 MSA-Reading contains two types of items: *selected response (SR)* and *brief constructed response (BCR)* items. *SR* items required students to select a correct answer from several alternatives. For the 2009 MSA-Reading, students selected an answer from four alternatives. Each *SR* item was scored as right or wrong.

*BCR* items required students to answer a question with a couple of words, a sentence, or a more elaborate way. For the 2009 MSA-Reading, these items were scored using a general rubric with maximum values between 0 and 3.

**The Role of Operational SR Items**

All the *SR* items except for those in sessions 2 (Literary Reading) and 3 (Informational Reading) were used for both form-to-form and year-to-year linking. The session 2 and 3 items were used only for the purpose of year-to-year linking since they are unique items.

Detailed information about form-to-form and year-to-year linking procedures can be found in section 1.9, *Form-to-Form Linking Procedures* and *Year-to-Year Linking Procedures*.

**Table 1.3 The 2009 MSA-Reading Item Distribution of Each Standard: Grade 3 and 8**

Form	General Reading			Literary Reading			Informational reading			Total Number of Items
	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	
A	16	0	16	8	2	10	9	2	11	<b>37</b>
B	16	0	16	8	2	10	9	2	11	<b>37</b>

**Table 1.4 The 2009 MSA-Reading Item Distribution of Each Standard: Grade 5**

Form	General Reading			Literary Reading			Informational reading			Total Number of Items
	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	
A	15	0	15	9	2	11	9	2	11	<b>37</b>
B	15	0	15	9	2	11	9	2	11	<b>37</b>

**Table 1.5 The 2009 MSA-Reading Item Distribution of Each Standard: Grade 4, 6, and 7**

Form	General Reading			Literary Reading			Informational reading			Total Number of Items
	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	No. of SR	No. of BCR	No. of Items	
A	15	0	15	9	2	11	9	2	11	<b>37</b>
B	15	0	15	9	2	11	9	2	11	<b>37</b>

**Table 1.6 The 2009 MSA-Reading Total and Standard Scores: Grade 3 and 8**

Form	Total and Each Cluster Scores			
	General Reading	Literary Reading	Informational Reading	Total Score
A	16 (16 MC)	14 (8 MC + 6 BCR)	15 (9 MC + 6 BCR)	45
B	16 (16 MC)	14 (8 MC + 6 BCR)	15 (9 MC + 6 BCR)	45

**Table 1.7 The 2009 MSA-Reading Total and Standard Scores: Grade 5**

Form	Total and Each Cluster Scores			
	General Reading	Literary Reading	Informational Reading	Total Score
A	15 (15 MC)	15 (9 MC + 6 BCR)	15 (9 MC + 6 BCR)	45
B	15 (15 MC)	15 (9 MC + 6 BCR)	15 (9 MC + 6 BCR)	45

**Table 1.8 The 2009 MSA-Reading Total and Standard Scores: Grade 4, 6, and 7**

Form	Total and Each Cluster Scores			
	General Reading	Literary Reading	Informational Reading	Total Score
A	15 (15 MC)	15 (9 MC + 6 BCR)	15 (9 MC + 6 BCR)	45
B	15 (15 MC)	15 (9 MC + 6 BCR)	15 (9 MC + 6 BCR)	45



## 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-Reading 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, reading content specialists from MSDE reviewed the final test forms of the 2009 MSA-Reading. The following table and figure show an example of the 2009 MSA-Reading 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 1.9 The 2009 Reading Operational Test Construction Using the Rasch Model: Grade 4 Form A**

Item Type	P-value	A	$D_{i1}$	$D_{i2}$	$D_{i3}$
BCR	0.55	1.00	-2.8865	0.1713	4.6908
BCR	0.36	1.00	-0.2820	1.6397	4.9886
BCR	0.47	1.00	-1.9819	1.1775	6.3237
BCR	0.33	1.00	-0.4489	2.3407	6.2535
SR	0.86	1.00	-0.6371		
SR	0.81	1.00	-0.2093		
SR	0.91	1.00	-1.2263		
SR	0.93	1.00	-1.4827		
SR	0.92	1.00	-1.3213		
SR	0.95	1.00	-1.8707		
SR	0.89	1.00	-1.0118		
SR	0.50	1.00	1.4012		
SR	0.93	1.00	-1.7932		
SR	0.63	1.00	0.7230		
SR	0.71	1.00	0.3088		
SR	0.75	1.00	0.0330		
SR	0.76	1.00	-0.0371		
SR	0.79	1.00	-0.1849		
SR	0.52	1.00	1.3123		
SR	0.55	1.00	1.1575		
SR	0.69	1.00	0.4436		
SR	0.62	1.00	0.8269		
SR	0.62	1.00	0.7861		
SR	0.81	1.00	-0.3227		
SR	0.55	1.00	1.1761		
SR	0.72	1.00	0.1781		
SR	0.65	1.00	0.5588		
SR	0.47	1.00	1.4551		
SR	0.69	1.00	0.3030		
SR	0.82	1.00	-0.5030		
SR	0.49	1.00	1.3460		
SR	0.72	1.00	0.2006		
SR	0.64	1.00	0.6870		
SR	0.65	1.00	0.5988		
SR	0.59	1.00	0.9160		
SR	0.62	1.00	0.7593		
SR	0.58	1.00	0.9692		
SR	0.82	1.00	-0.8817		
SR	0.65	1.00	0.1471		

**Table 1.9 (Continued)**

Item Type	P-value	A	$D_{i1}$	$D_{i2}$	$D_{i3}$
SR	0.66	1.00	0.1540		
SR	0.59	1.00	0.1906		
SR	0.73	1.00	-0.1963		
SR	0.86	1.00	-1.2429		
SR	0.67	1.00	0.1071		
SR	0.63	1.00	0.1229		

*Note.* A: item discrimination;  $D_{i1}$ : first structure measure estimate;  $D_{i2}$ : second structure measure estimate;  $D_{i3}$ : 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 ( $D_{ij} = D_i + F_{ij}$ )

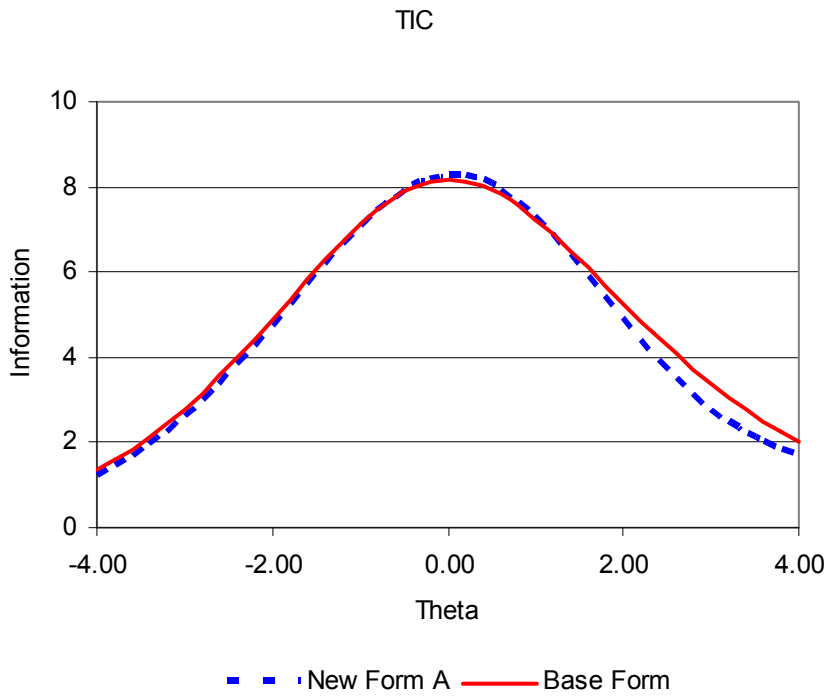


Figure 1.1 Test Information Curves of Base Form vs. Current Year's Reading Operational Test Form

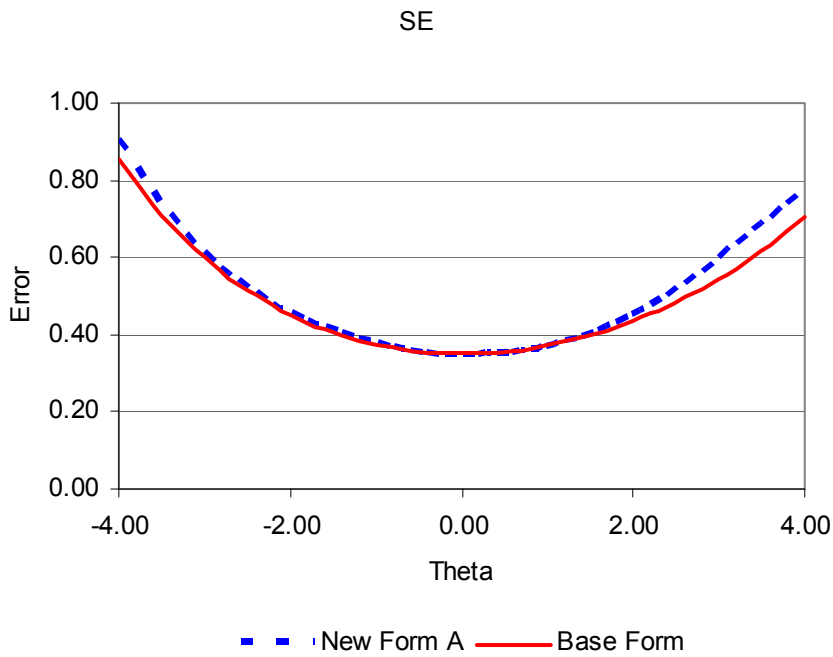


Figure 1.2 Standard Errors of Base Form vs. Current Year's Reading Operational Test Form

## 1.6 Test Administration of the 2009 MSA-Reading

The 2009 MSA-Reading test was administered to all students in grades 3 through 8 except for students taking the Alt-MSA-Reading or the Mod-MSA- Reading. 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. TEs checked to ensure they had 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- Reading
- 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

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 (EM) 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-Reading 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-Reading 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-Reading, 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
- Reading Primary Testing Window March 16 – March 25, 2009
- Make-up Testing Window March 26 – March 31, 2009

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

If a student was absent on the testing days, a make-up test was administered on any two consecutive days within 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-Reading, 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-Reading, and all students in grades 6 through 8 had to participate in either the 2009 MSA-Reading or Mod-MSA-Reading. All students in grade 6 through 8 had to participate in the 2009 Mod-MSA-Reading, 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-Reading or Mod-MSA-Reading. The criteria that students should need to meet in order to be tested in the Alt-MSA program instead of the MSA-Reading can be viewed in section 5, Appendix A of the TACM.

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

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

ELL students in their first year of enrollment in a U.S. school may substitute their score on the English Language Proficiency Test for the MSA-Reading or the Mod-MSA-Reading test. ELL students must participate in the MSA-Reading or the Mod-MSA-Reading test starting in their second year of enrollment in a U.S. school.

## **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 who are 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, Administering, 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-Reading 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 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 were 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 section 4.

### **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-Reading. 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 were 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-Reading used a Test Book format in which students wrote their answers directly in the Test Book. There were 10 forms of MSA-Reading. Different test forms were administered to students in each classroom participating in reading 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 highlighters 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, creating problems 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-Reading 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 (Pearson, 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<sup>TM</sup> 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 Scoring Procedures of the 2009 MSA-Reading

Students' responses to *SR* items were machine-scored, and their responses to *BCR* 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. 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 80% perfect agreement 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 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 (i.e., 80%). 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-Reading, 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 80% perfect agreement 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-Reading

The following scoring rules were applied to MSA-Reading BCR items:

- Reading BCR items were scored:  
0, 1, 2, or 3 with two readings
- Scores given were the higher of the 1st and 2nd Reader's scores provided they were adjacent.
- For example:

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

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

- For example:

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

### 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 reading 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 MSA-Reading 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 B). 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 in Table 1.10.

**Table 1.10 Descriptive Statistics for the 2009 MSA-Reading Form-to-Form Linking Common Items**

Grade	Form	No. of Items	<i>N</i>	<i>M</i>	<i>SD</i>
3	A	25	26,900	19.06	4.42
	B	25	26,903	19.11	4.37
4	A	25	29,449	17.71	4.51
	B	25	29,266	17.62	4.49
5	A	25	30,193	18.59	4.03
	B	25	29,986	18.61	4.00
6	A	25	29,751	18.23	4.17
	B	25	29,070	18.22	4.17
7	A	25	30,046	19.32	4.24
	B	25	29,541	19.37	4.15
8	A	25	30,717	18.12	3.87
	B	25	30,235	18.13	3.80

*Note.* Form A designates the identical operational portion of Forms 1, 3, 5, 7, and 9. Form B designates the identical operational portion of Forms 2, 4, 6, 8, and 10.

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

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

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

It should be noted that only SR items were used for the purpose of year-to-year linking. Second, the item sequence numbers on the tables were assigned based on the 2009 assessment. The statistics of the previous year's assessment (i.e., 2007) were calculated based on a smaller field-test sample while the 2009 statistics are based on the current year's statewide population. Finally, it should be noted that detailed information about the Rasch analysis on these core linking items can be found in section 1.9, *Calibration, Equating, Scaling*.

In general, we can conclude that most of the 2009 p-values were slightly increased compared to the 2007 p-values across all grades.

**Table 1.11 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 3 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3399914	SR	0.93	0.94	24	3588154	SR	0.65	0.71
2	3588051	SR	0.98	0.98	25	3588158	SR	0.49	0.55
3	3588052	SR	0.74	0.75	26	3592457	SR	0.71	0.76
4	3588010	SR	0.87	0.90	27	3592456	SR	0.49	0.51
5	3588015	SR	0.89	0.92	28	3592460	SR	0.81	0.83
6	3588020	SR	0.94	0.96	29	3592458	SR	0.65	0.70
7	3588023	SR	0.68	0.72	36	3592482	SR	0.60	0.62
8	3588035	SR	0.64	0.71	37	3592483	SR	0.89	0.89
9	3588039	SR	0.77	0.79	38	3592481	SR	0.84	0.85
<b>10</b>	<b>3492376</b>	<b>SR</b>	<b>0.85</b>	<b>0.91</b>	39	3592485	SR	0.80	0.78
<b>12</b>	<b>3492383</b>	<b>SR</b>	<b>0.77</b>	<b>0.82</b>	40	3592477	SR	0.55	0.55
<b>13</b>	<b>3492387</b>	<b>SR</b>	<b>0.74</b>	<b>0.81</b>	41	3592478	SR	0.79	0.79
<b>15</b>	<b>3492385</b>	<b>SR</b>	<b>0.63</b>	<b>0.69</b>	42	3592473	SR	0.64	0.64
<b>16</b>	<b>3471500</b>	<b>SR</b>	<b>0.54</b>	<b>0.67</b>	43	3592474	SR	0.60	0.60
<b>18</b>	<b>3471502</b>	<b>SR</b>	<b>0.73</b>	<b>0.81</b>					
<b>19</b>	<b>3471497</b>	<b>SR</b>	<b>0.68</b>	<b>0.75</b>					
<b>21</b>	<b>3471496</b>	<b>SR</b>	<b>0.64</b>	<b>0.77</b>					
22	3588155	SR	0.76	0.82					
23	3588156	SR	0.70	0.78					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
3	2007	33	0.73	0.13
	2009	33	0.77	0.12

**Table 1.12 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 3 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3399914	SR	0.93	0.94		3588154	SR	0.65	0.72
2	3588051	SR	0.98	0.98		3588158	SR	0.49	0.55
3	3588052	SR	0.74	0.75		3592457	SR	0.71	0.76
4	3588010	SR	0.87	0.90		3592456	SR	0.49	0.50
5	3588015	SR	0.89	0.92		3592460	SR	0.81	0.83
6	3588020	SR	0.94	0.96		3592458	SR	0.65	0.70
7	3588023	SR	0.68	0.73		3592482	SR	0.60	0.60
8	3588035	SR	0.64	0.71		3592483	SR	0.89	0.89
9	3588039	SR	0.77	0.79		3592481	SR	0.84	0.85
<b>10</b>	<b>3497775</b>	<b>SR</b>	<b>0.69</b>	<b>0.80</b>		3592485	SR	0.80	0.80
<b>12</b>	<b>3497781</b>	<b>SR</b>	<b>0.79</b>	<b>0.80</b>		3592477	SR	0.55	0.56
<b>13</b>	<b>3497779</b>	<b>SR</b>	<b>0.70</b>	<b>0.77</b>		3592478	SR	0.79	0.79
<b>15</b>	<b>3497774</b>	<b>SR</b>	<b>0.72</b>	<b>0.78</b>		3592473	SR	0.64	0.64
<b>16</b>	<b>3490488</b>	<b>SR</b>	<b>0.52</b>	<b>0.59</b>		3592474	SR	0.60	0.59
<b>18</b>	<b>3490487</b>	<b>SR</b>	<b>0.65</b>	<b>0.69</b>					
<b>19</b>	<b>3490490</b>	<b>SR</b>	<b>0.77</b>	<b>0.80</b>					
<b>21</b>	<b>3490489</b>	<b>SR</b>	<b>0.69</b>	<b>0.75</b>					
22	3588155	SR	0.76	0.83					
23	3588156	SR	0.70	0.80					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
3	2007	33	0.73	0.13
	2009	33	0.76	0.12

**Table 1.13 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 4 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3588095	SR	0.65	0.65	24	3592889	SR	0.65	0.68
2	3399931	SR	0.79	0.79	25	3592891	SR	0.53	0.46
3	3588096	SR	0.78	0.78	26	3592925	SR	0.62	0.61
4	3595149	SR	0.94	0.94	27	3592933	SR	0.51	0.55
5	3399943	SR	0.95	0.96	28	3592930	SR	0.55	0.54
6	3399944	SR	0.90	0.97	29	3592934	SR	0.76	0.77
7	3588105	SR	0.55	0.56	36	3592902	SR	0.67	0.67
8	3588111	SR	0.94	0.94	37	3592903	SR	0.62	0.57
9	3588114	SR	0.89	0.89	38	3592909	SR	0.65	0.65
<b>10</b>	<b>3497923</b>	<b>SR</b>	<b>0.73</b>	<b>0.80</b>	39	3592905	SR	0.78	0.78
<b>12</b>	<b>3497925</b>	<b>SR</b>	<b>0.72</b>	<b>0.79</b>	40	3588222	SR	0.59	0.58
<b>13</b>	<b>3497924</b>	<b>SR</b>	<b>0.72</b>	<b>0.78</b>	41	3588220	SR	0.78	0.74
<b>15</b>	<b>3497922</b>	<b>SR</b>	<b>0.68</b>	<b>0.62</b>	42	3588217	SR	0.76	0.74
<b>16</b>	<b>3470326</b>	<b>SR</b>	<b>0.43</b>	<b>0.47</b>	43	3588218	SR	0.51	0.48
<b>18</b>	<b>3470320</b>	<b>SR</b>	<b>0.60</b>	<b>0.63</b>					
<b>19</b>	<b>3470319</b>	<b>SR</b>	<b>0.65</b>	<b>0.66</b>					
<b>21</b>	<b>3470324</b>	<b>SR</b>	<b>0.72</b>	<b>0.78</b>					
22	3592890	SR	0.59	0.64					
23	3592893	SR	0.69	0.77					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
4	2007	33	0.69	0.13
	2009	33	0.70	0.14

**Table 1.14 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 4 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3588095	SR	0.65	0.65	24	3592889	SR	0.65	0.68
2	3399931	SR	0.79	0.80	25	3592891	SR	0.53	0.45
3	3588096	SR	0.78	0.78	26	3592925	SR	0.62	0.60
4	3595149	SR	0.94	0.94	27	3592933	SR	0.51	0.55
5	3399943	SR	0.95	0.96	28	3592930	SR	0.55	0.55
6	3399944	SR	0.90	0.97	29	3592934	SR	0.76	0.78
7	3588105	SR	0.55	0.56	36	3592902	SR	0.67	0.65
8	3588111	SR	0.94	0.94	37	3592903	SR	0.62	0.56
9	3588114	SR	0.89	0.89	38	3592909	SR	0.65	0.63
<b>10</b>	<b>3488811</b>	<b>SR</b>	<b>0.71</b>	<b>0.76</b>	39	3592905	SR	0.78	0.78
<b>12</b>	<b>3488810</b>	<b>SR</b>	<b>0.70</b>	<b>0.74</b>	40	3588222	SR	0.59	0.56
<b>13</b>	<b>3488812</b>	<b>SR</b>	<b>0.80</b>	<b>0.83</b>	41	3588220	SR	0.78	0.74
<b>15</b>	<b>3488809</b>	<b>SR</b>	<b>0.52</b>	<b>0.57</b>	42	3588217	SR	0.76	0.73
<b>16</b>	<b>3497913</b>	<b>SR</b>	<b>0.57</b>	<b>0.62</b>	43	3588218	SR	0.51	0.48
<b>18</b>	<b>3497910</b>	<b>SR</b>	<b>0.52</b>	<b>0.53</b>					
<b>19</b>	<b>3497912</b>	<b>SR</b>	<b>0.69</b>	<b>0.72</b>					
<b>21</b>	<b>3497914</b>	<b>SR</b>	<b>0.85</b>	<b>0.82</b>					
22	3592890	SR	0.59	0.65					
23	3592893	SR	0.69	0.76					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
4	2007	33	0.70	0.14
	2009	33	0.70	0.14

**Table 1.15 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 5 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3400077	SR	0.86	0.87	24	3588245	SR	0.69	0.76
2	3400080	SR	0.81	0.80	25	3588248	SR	0.82	0.84
3	3400086	SR	0.91	0.91	26	3588250	SR	0.72	0.71
4	3400088	SR	0.94	0.93	27	3588252	SR	0.64	0.62
5	3451551	SR	0.92	0.93	28	3588251	SR	0.65	0.69
6	3451440	SR	0.95	0.96	29	3588254	SR	0.62	0.71
7	3451552	SR	0.89	0.91	36	3588240	SR	0.69	0.66
8	3588453	SR	0.50	0.49	37	3588239	SR	0.62	0.68
9	3588454	SR	0.93	0.93	38	3588242	SR	0.62	0.63
<b>10</b>	<b>3486361</b>	<b>SR</b>	<b>0.82</b>	<b>0.90</b>	39	3595121	SR	0.55	0.54
<b>12</b>	<b>3486359</b>	<b>SR</b>	<b>0.65</b>	<b>0.74</b>	40	3588233	SR	0.63	0.60
<b>13</b>	<b>3486364</b>	<b>SR</b>	<b>0.66</b>	<b>0.67</b>	41	3588236	SR	0.71	0.66
<b>15</b>	<b>3486363</b>	<b>SR</b>	<b>0.59</b>	<b>0.75</b>	42	3588234	SR	0.76	0.74
<b>16</b>	<b>3468076</b>	<b>SR</b>	<b>0.73</b>	<b>0.75</b>	43	3588235	SR	0.79	0.74
<b>18</b>	<b>3468079</b>	<b>SR</b>	<b>0.86</b>	<b>0.91</b>					
<b>19</b>	<b>3468075</b>	<b>SR</b>	<b>0.67</b>	<b>0.74</b>					
<b>21</b>	<b>3468080</b>	<b>SR</b>	<b>0.63</b>	<b>0.72</b>					
22	3588243	SR	0.72	0.72					
23	3588244	SR	0.47	0.55					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
5	2007	33	0.73	0.13
	2009	33	0.75	0.13

**Table 1.16 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 5 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3400077	SR	0.86	0.87	24	3588245	SR	0.69	0.76
2	3400080	SR	0.81	0.80	25	3588248	SR	0.82	0.86
3	3400086	SR	0.91	0.92	26	3588250	SR	0.72	0.72
4	3400088	SR	0.94	0.94	27	3588252	SR	0.64	0.63
5	3451551	SR	0.92	0.93	28	3588251	SR	0.65	0.69
6	3451440	SR	0.95	0.96	29	3588254	SR	0.62	0.71
7	3451552	SR	0.89	0.91	36	3588240	SR	0.69	0.66
8	3588453	SR	0.50	0.49	37	3588239	SR	0.62	0.68
9	3588454	SR	0.93	0.93	38	3588242	SR	0.62	0.63
<b>10</b>	<b>3296480</b>	<b>SR</b>	<b>0.87</b>	<b>0.95</b>	39	3595121	SR	0.55	0.54
<b>12</b>	<b>3296482</b>	<b>SR</b>	<b>0.74</b>	<b>0.74</b>	40	3588233	SR	0.63	0.61
<b>13</b>	<b>3296476</b>	<b>SR</b>	<b>0.69</b>	<b>0.63</b>	41	3588236	SR	0.71	0.67
<b>15</b>	<b>3296479</b>	<b>SR</b>	<b>0.65</b>	<b>0.69</b>	42	3588234	SR	0.76	0.72
<b>16</b>	<b>3486184</b>	<b>SR</b>	<b>0.74</b>	<b>0.74</b>	43	3588235	SR	0.79	0.74
<b>18</b>	<b>3486187</b>	<b>SR</b>	<b>0.59</b>	<b>0.60</b>					
<b>19</b>	<b>3486183</b>	<b>SR</b>	<b>0.46</b>	<b>0.53</b>					
<b>21</b>	<b>3486185</b>	<b>SR</b>	<b>0.72</b>	<b>0.80</b>					
22	3588243	SR	0.72	0.74					
23	3588244	SR	0.47	0.53					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
5	2007	33	0.72	0.14
	2009	33	0.74	0.14

**Table 1.17 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 6 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3400102	SR	0.92	0.93	24	3588279	SR	0.84	0.86
2	3400104	SR	0.93	0.97	25	3588281	SR	0.78	0.82
3	3595144	SR	0.85	0.85	26	3588289	SR	0.55	0.56
4	3400107	SR	0.88	0.88	27	3588290	SR	0.67	0.72
5	3588412	SR	0.81	0.81	28	3588292	SR	0.70	0.75
6	3451451	SR	0.92	0.93	29	3588288	SR	0.42	0.45
7	3451452	SR	0.52	0.50	36	3594634	SR	0.65	0.66
8	3451553	SR	0.80	0.80	37	3594633	SR	0.51	0.50
9	3451453	SR	0.93	0.93	38	3594638	SR	0.62	0.62
<b>10</b>	<b>3470024</b>	<b>SR</b>	<b>0.77</b>	<b>0.85</b>	39	3594637	SR	0.76	0.76
<b>12</b>	<b>3470020</b>	<b>SR</b>	<b>0.67</b>	<b>0.79</b>	40	3588272	SR	0.86	0.85
<b>13</b>	<b>3470018</b>	<b>SR</b>	<b>0.72</b>	<b>0.76</b>	41	3588275	SR	0.66	0.64
<b>15</b>	<b>3470021</b>	<b>SR</b>	<b>0.64</b>	<b>0.73</b>	42	3588274	SR	0.50	0.52
<b>16</b>	<b>3498430</b>	<b>SR</b>	<b>0.68</b>	<b>0.68</b>	43	3588273	SR	0.59	0.56
<b>18</b>	<b>3498424</b>	<b>SR</b>	<b>0.73</b>	<b>0.80</b>					
<b>19</b>	<b>3498425</b>	<b>SR</b>	<b>0.70</b>	<b>0.78</b>					
<b>21</b>	<b>3498426</b>	<b>SR</b>	<b>0.51</b>	<b>0.58</b>					
22	3588278	SR	0.50	0.51					
23	3588277	SR	0.88	0.88					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
6	2007	33	0.71	0.15
	2009	33	0.73	0.15



**Table 1.18 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 6 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3400102	SR	0.92	0.93	24	3588279	SR	0.84	0.86
2	3400104	SR	0.93	0.97	25	3588281	SR	0.78	0.81
3	3595144	SR	0.85	0.85	26	3588289	SR	0.55	0.56
4	3400107	SR	0.88	0.88	27	3588290	SR	0.67	0.71
5	3588412	SR	0.81	0.81	28	3588292	SR	0.70	0.75
6	3451451	SR	0.92	0.93	29	3588288	SR	0.42	0.44
7	3451452	SR	0.52	0.50	36	3594634	SR	0.65	0.66
8	3451553	SR	0.80	0.80	37	3594633	SR	0.51	0.49
9	3451453	SR	0.93	0.94	38	3594638	SR	0.62	0.60
<b>10</b>	<b>3470033</b>	<b>SR</b>	<b>0.55</b>	<b>0.59</b>	39	3594637	SR	0.76	0.77
<b>12</b>	<b>3470035</b>	<b>SR</b>	<b>0.66</b>	<b>0.69</b>	40	3588272	SR	0.86	0.85
<b>13</b>	<b>3470034</b>	<b>SR</b>	<b>0.69</b>	<b>0.81</b>	41	3588275	SR	0.66	0.63
<b>15</b>	<b>3470032</b>	<b>SR</b>	<b>0.59</b>	<b>0.61</b>	42	3588274	SR	0.50	0.51
<b>16</b>	<b>3489686</b>	<b>SR</b>	<b>0.76</b>	<b>0.84</b>	43	3588273	SR	0.59	0.56
<b>18</b>	<b>3489689</b>	<b>SR</b>	<b>0.77</b>	<b>0.82</b>					
<b>19</b>	<b>3489692</b>	<b>SR</b>	<b>0.61</b>	<b>0.68</b>					
<b>21</b>	<b>3489691</b>	<b>SR</b>	<b>0.69</b>	<b>0.76</b>					
22	3588278	SR	0.50	0.53					
23	3588277	SR	0.88	0.87					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
6	2007	33	0.71	0.15
	2009	33	0.73	0.15

**Table 1.19 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 7 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3400135	SR	0.95	0.96	24	3588331	SR	0.60	0.70
2	3400120	SR	0.92	0.93	25	3588333	SR	0.88	0.95
3	3400132	SR	0.79	0.81	32	3588330	SR	0.87	0.88
4	3451470	SR	0.90	0.91	33	3588326	SR	0.64	0.61
5	3451556	SR	0.94	0.95	34	3588327	SR	0.73	0.75
<b>6</b>	<b>3470045</b>	<b>SR</b>	<b>0.75</b>	<b>0.85</b>	35	3595060	SR	0.72	0.75
<b>8</b>	<b>3470047</b>	<b>SR</b>	<b>0.65</b>	<b>0.77</b>	36	3588283	SR	0.56	0.54
<b>9</b>	<b>3470044</b>	<b>SR</b>	<b>0.81</b>	<b>0.87</b>	37	3588284	SR	0.71	0.71
<b>11</b>	<b>3470048</b>	<b>SR</b>	<b>0.58</b>	<b>0.68</b>	38	3588286	SR	0.84	0.83
<b>12</b>	<b>3468871</b>	<b>SR</b>	<b>0.59</b>	<b>0.69</b>	39	3594600	SR	0.80	0.82
<b>14</b>	<b>3468868</b>	<b>SR</b>	<b>0.53</b>	<b>0.62</b>	40	3588316	SR	0.75	0.73
<b>15</b>	<b>3468869</b>	<b>SR</b>	<b>0.64</b>	<b>0.73</b>	41	3588317	SR	0.81	0.80
<b>17</b>	<b>3468875</b>	<b>SR</b>	<b>0.47</b>	<b>0.57</b>	42	3588318	SR	0.78	0.74
18	3595069	SR	0.63	0.72	43	3595059	SR	0.47	0.50
19	3595067	SR	0.81	0.86					
20	3595071	SR	0.65	0.66					
21	3595072	SR	0.61	0.70					
22	3588332	SR	0.80	0.85					
23	3595058	SR	0.60	0.67					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
7	2007	33	0.72	0.13
	2009	33	0.76	0.12

**Table 1.20 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 7 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3400135	SR	0.95	0.96	24	3588331	SR	0.60	0.70
2	3400120	SR	0.92	0.93	25	3588333	SR	0.88	0.95
3	3400132	SR	0.79	0.82	32	3588330	SR	0.87	0.89
4	3451470	SR	0.90	0.91	33	3588326	SR	0.64	0.61
5	3451556	SR	0.94	0.95	34	3588327	SR	0.73	0.76
<b>6</b>	<b>3497792</b>	<b>SR</b>	<b>0.59</b>	<b>0.73</b>	35	3595060	SR	0.72	0.75
<b>8</b>	<b>3497790</b>	<b>SR</b>	<b>0.75</b>	<b>0.83</b>	36	3588283	SR	0.56	0.54
<b>9</b>	<b>3497793</b>	<b>SR</b>	<b>0.75</b>	<b>0.81</b>	37	3588284	SR	0.71	0.72
<b>11</b>	<b>3497789</b>	<b>SR</b>	<b>0.69</b>	<b>0.73</b>	38	3588286	SR	0.84	0.84
<b>12</b>	<b>3468857</b>	<b>SR</b>	<b>0.59</b>	<b>0.68</b>	39	3594600	SR	0.80	0.82
<b>14</b>	<b>3468860</b>	<b>SR</b>	<b>0.73</b>	<b>0.79</b>	40	3588316	SR	0.75	0.73
<b>15</b>	<b>3468856</b>	<b>SR</b>	<b>0.55</b>	<b>0.60</b>	41	3588317	SR	0.81	0.80
<b>17</b>	<b>3468863</b>	<b>SR</b>	<b>0.61</b>	<b>0.64</b>	42	3588318	SR	0.78	0.74
18	3595069	SR	0.63	0.69	43	3595059	SR	0.47	0.50
19	3595067	SR	0.81	0.87					
20	3595071	SR	0.65	0.66					
21	3595072	SR	0.61	0.71					
22	3588332	SR	0.80	0.86					
23	3595058	SR	0.60	0.66					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
7	2007	33	0.73	0.13
	2009	33	0.76	0.12

**Table 1.21 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 8 Form A**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FA
1	3400154	SR	0.91	0.95	24	3595097	SR	0.55	0.57
2	3400158	SR	0.93	0.93	25	3595095	SR	0.63	0.63
3	3451476	SR	0.84	0.85	32	3588376	SR	0.69	0.72
4	3451557	SR	0.83	0.84	33	3588377	SR	0.48	0.52
5	3451558	SR	0.89	0.91	34	3588378	SR	0.72	0.72
<b>6</b>	<b>3514207</b>	<b>SR</b>	<b>0.65</b>	<b>0.76</b>	35	3595087	SR	0.87	0.87
<b>8</b>	<b>3514204</b>	<b>SR</b>	<b>0.64</b>	<b>0.59</b>	36	3588367	SR	0.61	0.63
<b>9</b>	<b>3514208</b>	<b>SR</b>	<b>0.72</b>	<b>0.80</b>	37	3588366	SR	0.82	0.83
<b>11</b>	<b>3514206</b>	<b>SR</b>	<b>0.62</b>	<b>0.73</b>	38	3588369	SR	0.53	0.49
<b>12</b>	<b>3327512</b>	<b>SR</b>	<b>0.86</b>	<b>0.91</b>	39	3588368	SR	0.87	0.88
<b>14</b>	<b>3327514</b>	<b>SR</b>	<b>0.64</b>	<b>0.74</b>	40	3588392	SR	0.79	0.80
<b>15</b>	<b>3327513</b>	<b>SR</b>	<b>0.70</b>	<b>0.77</b>	41	3588396	SR	0.34	0.27
<b>17</b>	<b>3327516</b>	<b>SR</b>	<b>0.65</b>	<b>0.72</b>	42	3588393	SR	0.64	0.59
18	3588360	SR	0.64	0.75	43	3588395	SR	0.50	0.54
19	3588361	SR	0.62	0.72					
20	3595083	SR	0.76	0.83					
21	3588362	SR	0.63	0.69					
22	3595094	SR	0.66	0.71					
23	3595099	SR	0.81	0.89					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
8	2007	33	0.70	0.14
	2009	33	0.73	0.15

**Table 1.22 P-Value Comparison of Linking Common Items for Year 2007 vs. Year 2009: Grade 8 Form B**

Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB	Item Seq. No.	Item CID	Item Type	Previous Year	Y09 FB
1	3400154	SR	0.91	0.96	24	3595097	SR	0.55	0.56
2	3400158	SR	0.93	0.93	25	3595095	SR	0.63	0.63
3	3451476	SR	0.84	0.85	32	3588376	SR	0.69	0.72
4	3451557	SR	0.83	0.84	33	3588377	SR	0.48	0.51
5	3451558	SR	0.89	0.91	34	3588378	SR	0.72	0.73
<b>6</b>	<b>3470061</b>	<b>SR</b>	<b>0.65</b>	<b>0.76</b>	35	3595087	SR	0.87	0.87
<b>8</b>	<b>3470060</b>	<b>SR</b>	<b>0.56</b>	<b>0.60</b>	36	3588367	SR	0.61	0.62
<b>9</b>	<b>3470054</b>	<b>SR</b>	<b>0.69</b>	<b>0.77</b>	37	3588366	SR	0.82	0.84
<b>11</b>	<b>3470058</b>	<b>SR</b>	<b>0.78</b>	<b>0.87</b>	38	3588369	SR	0.53	0.48
<b>12</b>	<b>3489335</b>	<b>SR</b>	<b>0.88</b>	<b>0.94</b>	39	3588368	SR	0.87	0.88
<b>14</b>	<b>3489334</b>	<b>SR</b>	<b>0.68</b>	<b>0.75</b>	40	3588392	SR	0.79	0.80
<b>15</b>	<b>3489336</b>	<b>SR</b>	<b>0.60</b>	<b>0.66</b>	41	3588396	SR	0.34	0.27
<b>17</b>	<b>3489337</b>	<b>SR</b>	<b>0.61</b>	<b>0.65</b>	42	3588393	SR	0.64	0.59
18	3588360	SR	0.64	0.74	43	3588395	SR	0.50	0.54
19	3588361	SR	0.62	0.73					
20	3595083	SR	0.76	0.83					
21	3588362	SR	0.63	0.69					
22	3595094	SR	0.66	0.72					
23	3595099	SR	0.81	0.89					

Note. Bold-faced items are sessions 2 (Literary) and 3 (Informational) items.

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

Grade	Year	No. of Items	<i>M</i>	<i>SD</i>
8	2007	33	0.70	0.14
	2009	33	0.73	0.16

### Validation Check with the 2009 Operational BCR Items

To collect information about how much the same BCR items that appeared in both 2007 and 2009 changed in terms of item difficulty, indices such as the classical p-value and Rasch item difficulty were calculated.

These items were first field-tested on the 2007 assessment and appeared as operational test items on the 2009 assessment, as shown in Table 1.23. The item numbers in Tables 1.24 through 1.59 were assigned based on the 2009 assessment. Detailed information about the specific test design and construction of Year 2009 can be obtained from section 1.4, *Test Structure of the 2009 MSA-Reading*.

While the 2007 p-value was calculated with a field test sample, the 2009 p-value was calculated with a statewide population. The p-value of a BCR item was the mean item score divided by the item score range. The percentage of “Omits” response to each CR item was low and indicated that a small number of students did not respond at all. In general, the 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 the grades except for grade 6 form B.

With respect to 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 the 2007 assessment across all the grades except for grade 6 form B. It should be noted that all of the Rasch item and step difficulty parameters were on a common scale (i.e., linked to the 2003 or 2004 assessment).

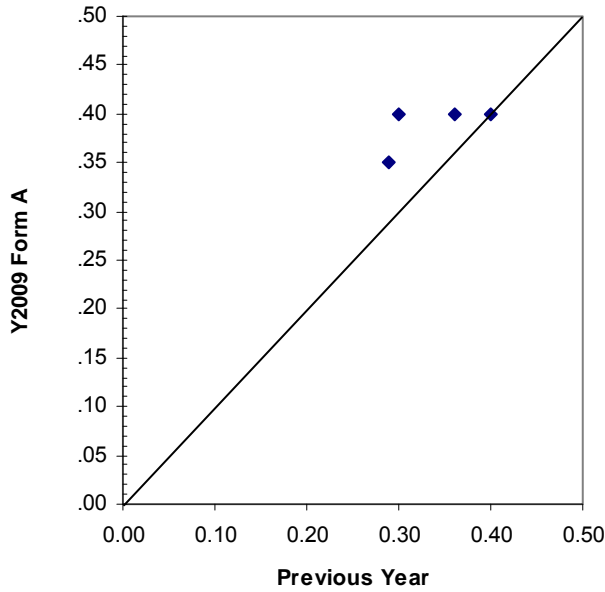
In conclusion, both p-value and Rasch item difficulty results reflected the same phenomenon, indicating that the level of item difficulty stayed the same or became a little lower across all the grades except for grade 6 form B.

**Table 1.23 Form Identification for Items Appearing in both 2007 and 2009: Grades 3 through 8**

Grade	Year 2007	Year 2009
3	Form 1, 2	Form A
	Form 7, 10	Form B
4	Form 1, 2	Form A
	Form 8, 9	Form B
5	Form 4, 7	Form A
	Form 8, 9	Form B
6	Form 1, 10	Form A
	Form 2, 3	Form B
7	Form 1, 4	Form A
	Form 9, 10	Form B
8	Form 1, 6	Form A
	Form 9, 10	Form B

**Table 1.24 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form A**

Item Number	CID	Item Type	Year 07	Year 09
11	3492389	BCR	0.40	0.40
14	3492393	BCR	0.36	0.40
17	3471505	BCR	0.29	0.35
20	3471506	BCR	0.30	0.40



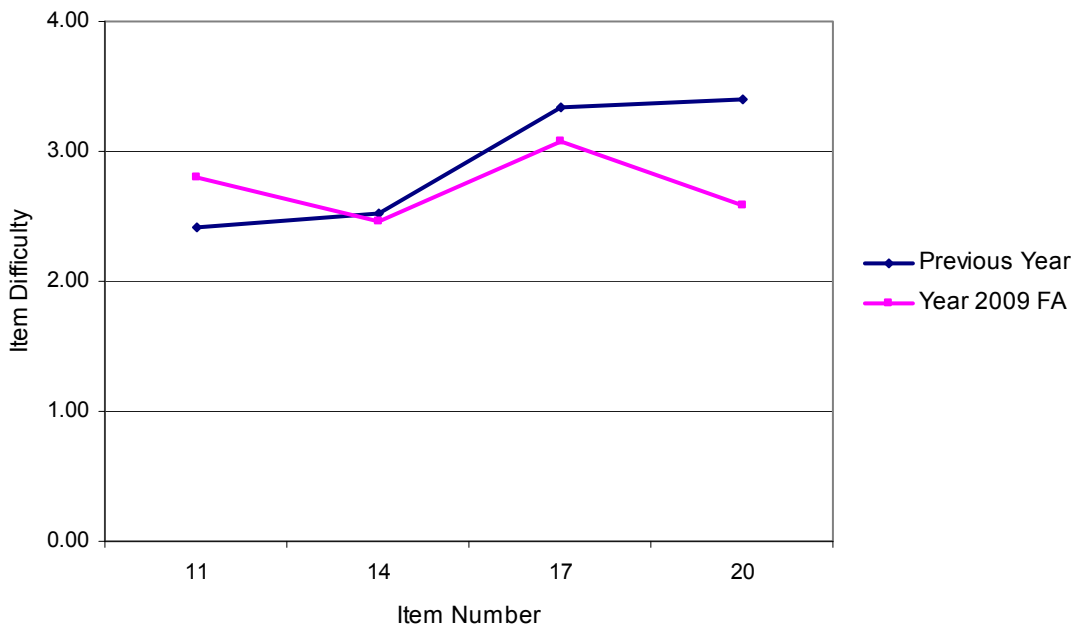
**Table 1.25 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3492389	BCR	2,341	1.19	0.63	10.04	58.65	29.35	0.64	1.32
2007	14	3492393	BCR	2,341	1.07	0.63	13.58	63.05	20.55	0.94	1.88
2007	17	3471505	BCR	2,292	0.88	0.66	26.79	55.54	15.75	0.22	1.70
2007	20	3471506	BCR	2,292	0.89	0.66	24.91	55.54	16.54	0.17	2.84
2009	11	3492389	BCR	26,900	1.19	0.66	12.87	55.20	30.75	0.62	0.56
2009	14	3492393	BCR	26,900	1.20	0.62	8.88	61.51	27.54	1.22	0.86
2009	17	3471505	BCR	26,900	1.05	0.66	18.57	57.41	23.05	0.50	0.47
2009	20	3471506	BCR	26,900	1.19	0.67	13.07	55.16	30.04	1.15	0.58

**Table 1.26 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 3 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3492389	BCR	2.4226	-3.8304	-0.2899	4.1204
2007	14	3492393	BCR	2.5163	-3.5826	0.1865	3.3961
2007	17	3471505	BCR	3.3416	-3.2847	-0.4582	3.7429
2007	20	3471506	BCR	3.3992	-3.4335	-0.5794	4.0129
2009	11	3492389	BCR	2.8042	-3.5604	-0.5136	4.0741
2009	14	3492393	BCR	2.4570	-3.5613	-0.0615	3.6228
2009	17	3471505	BCR	3.0745	-3.3108	-0.2529	3.5638
2009	20	3471506	BCR	2.5785	-3.2384	-0.3404	3.5788

Note. Rasch item and step difficulties were placed on a common scale.

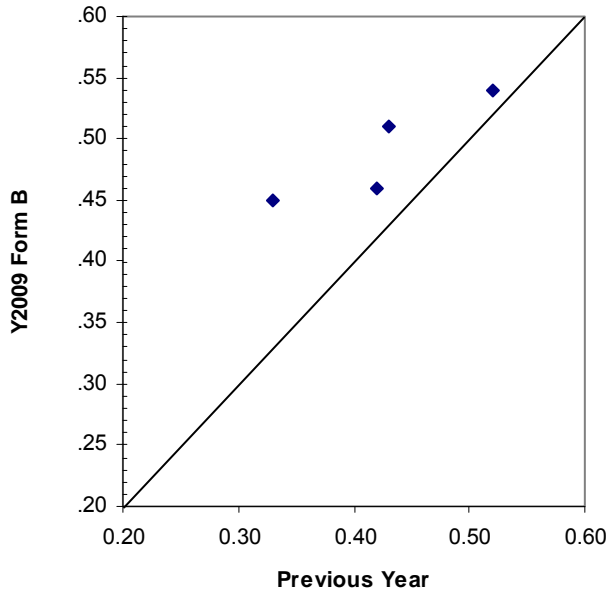


**Figure 1.3 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form A**



**Table 1.27 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form B**

Item Number	CID	Item Type	Year 07	Year 09
11	3497784	BCR	0.42	0.46
14	3497783	BCR	0.52	0.54
17	3490497	BCR	0.33	0.45
20	3490494	BCR	0.43	0.51



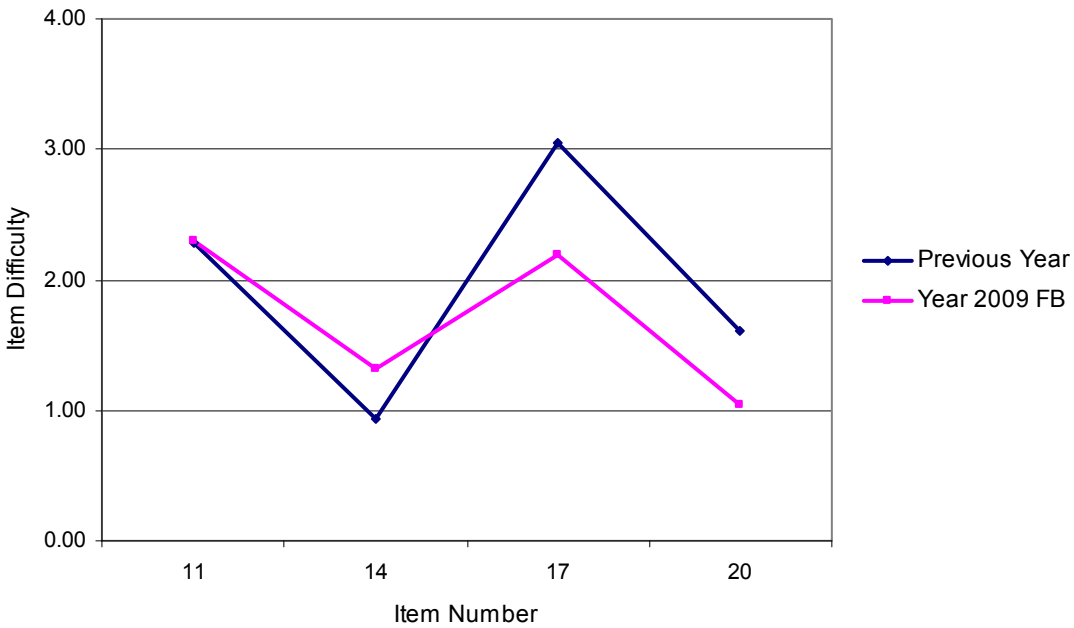
**Table 1.28 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3497784	BCR	2,166	1.25	0.65	9.28	53.83	34.53	0.65	1.71
2007	14	3497783	BCR	2,166	1.57	0.76	5.68	37.26	46.31	9.10	1.66
2007	17	3490497	BCR	2,138	1.00	0.68	19.97	54.21	22.73	0.23	2.85
2007	20	3490494	BCR	2,138	1.28	0.75	11.74	50.42	31.24	5.00	1.59
2009	11	3497784	BCR	26,903	1.38	0.65	8.39	45.52	45.20	0.59	0.30
2009	14	3497783	BCR	26,903	1.62	0.62	3.23	33.48	59.07	3.32	0.90
2009	17	3490497	BCR	26,903	1.35	0.65	7.94	50.11	40.19	1.56	0.19
2009	20	3490494	BCR	26,903	1.54	0.70	3.55	43.44	44.55	7.17	1.30

**Table 1.29 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 3 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3497784	BCR	2.2807	-3.6737	-0.4378	4.1115
2007	14	3497783	BCR	0.9306	-2.6508	-0.0901	2.7409
2007	17	3490497	BCR	3.0534	-3.3480	-0.6442	3.9921
2007	20	3490494	BCR	1.6020	-2.5315	0.1732	2.3582
2009	11	3497784	BCR	2.3012	-3.2529	-0.9743	4.2273
2009	14	3497783	BCR	1.3125	-3.1596	-0.6624	3.8220
2009	17	3490497	BCR	2.1899	-3.3766	-0.4784	3.8550
2009	20	3490494	BCR	1.0425	-3.1669	0.2284	2.9385

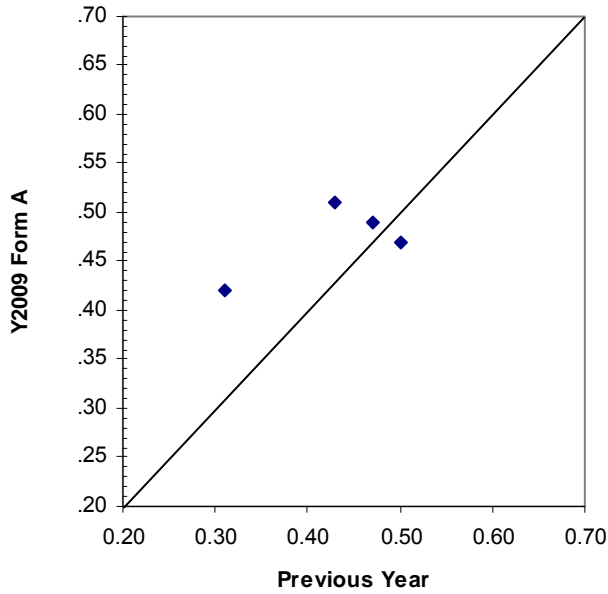
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.4 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 3 Form B**

**Table 1.30 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form A**

Item Number	CID	Item Type	Year 07	Year 09
11	3497929	BCR	0.43	0.51
14	3497931	BCR	0.31	0.42
17	3470329	BCR	0.47	0.49
20	3470328	BCR	0.50	0.47



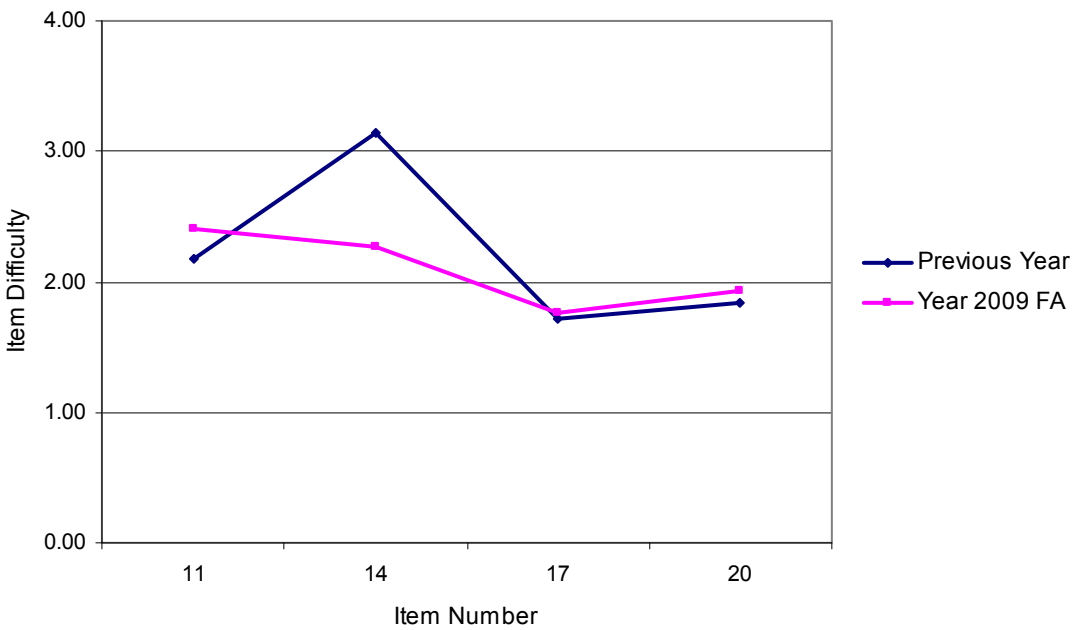
**Table 1.31 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3497929	BCR	2,358	1.29	0.59	4.71	60.22	32.65	1.02	1.40
2007	14	3497931	BCR	2,358	0.93	0.60	19.25	64.93	12.93	0.64	2.25
2007	17	3470329	BCR	2,340	1.42	0.66	5.26	48.38	42.39	2.99	0.98
2007	20	3470328	BCR	2,340	1.50	0.71	7.69	31.75	55.68	2.35	2.52
2009	11	3497929	BCR	29,449	1.53	0.64	6.51	33.89	58.11	0.89	0.60
2009	14	3497931	BCR	29,449	1.26	0.64	6.95	58.24	31.18	1.90	1.73
2009	17	3470329	BCR	29,449	1.46	0.77	11.39	35.28	47.74	5.02	0.58
2009	20	3470328	BCR	29,449	1.42	0.73	9.31	39.42	46.10	3.37	1.81

**Table 1.32 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 4 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3497929	BCR	2.1815	-4.0390	0.0636	3.9754
2007	14	3497931	BCR	3.1493	-3.3472	0.4526	2.8946
2007	17	3470329	BCR	1.7142	-3.3129	-0.0713	3.3842
2007	20	3470328	BCR	1.8388	-2.6916	-0.9951	3.6867
2009	11	3497929	BCR	2.4003	-3.5824	-1.4379	5.0203
2009	14	3497931	BCR	2.2747	-3.4363	0.0236	3.4127
2009	17	3470329	BCR	1.7568	-2.0543	-0.7715	2.8258
2009	20	3470328	BCR	1.9270	-2.5007	-0.6058	3.1064

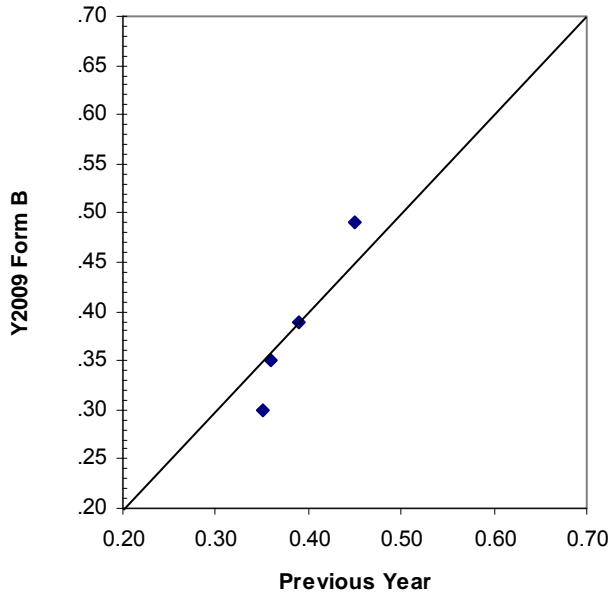
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.5 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form A**

**Table 1.33 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form B**

Item Number	CID	Item Type	Year 07	Year 09
11	3488817	BCR	0.36	0.35
14	3488816	BCR	0.45	0.49
17	3497919	BCR	0.35	0.30
20	3497917	BCR	0.39	0.39



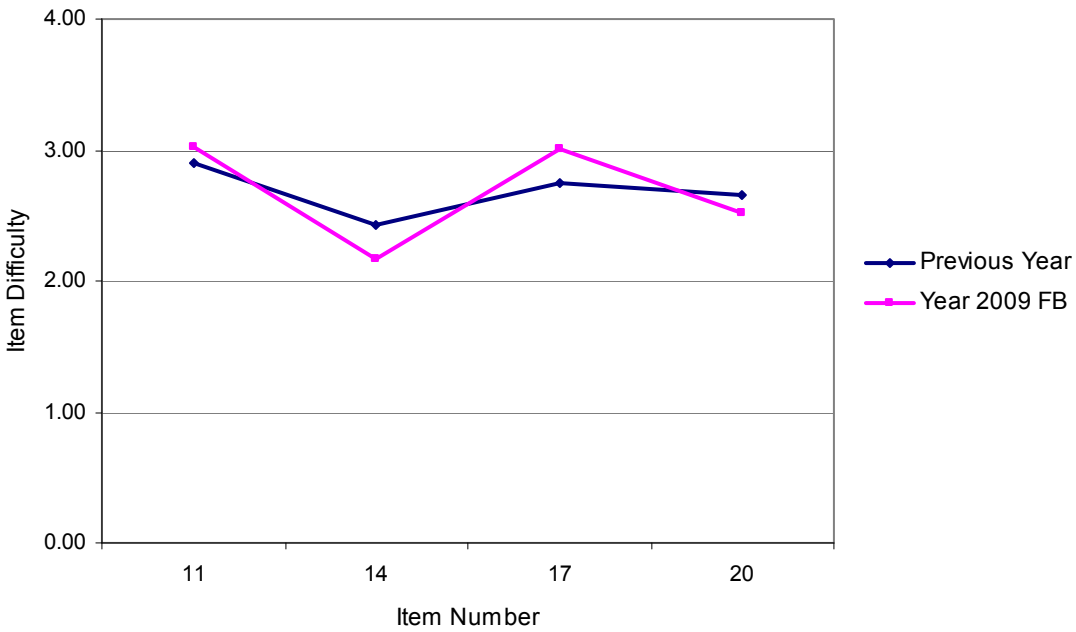
**Table 1.34 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3488817	BCR	2,155	1.07	0.70	19.30	51.97	26.73	0.60	1.39
2007	14	3488816	BCR	2,155	1.34	0.61	6.36	52.58	39.77	0.46	0.84
2007	17	3497919	BCR	2,140	1.05	0.74	22.15	50.33	24.44	1.82	1.26
2007	20	3497917	BCR	2,140	1.16	0.67	14.16	56.12	27.99	1.36	0.37
2009	11	3488817	BCR	29,266	1.04	0.77	25.10	43.70	28.61	0.99	1.60
2009	14	3488816	BCR	29,266	1.46	0.61	4.04	44.65	48.98	1.12	1.22
2009	17	3497919	BCR	29,266	0.90	0.83	36.73	36.69	23.74	2.00	0.85
2009	20	3497917	BCR	29,266	1.17	0.69	14.51	54.45	28.87	1.75	0.42

**Table 1.35 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 4 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3488817	BCR	2.9076	-2.9471	-0.4764	3.4236
2007	14	3488816	BCR	2.4251	-3.8636	-0.5811	4.4446
2007	17	3497919	BCR	2.7550	-2.5443	-0.1159	2.6603
2007	20	3497917	BCR	2.6612	-3.1431	-0.1371	3.2802
2009	11	3488817	BCR	3.0282	-2.4597	-0.6250	3.0847
2009	14	3488816	BCR	2.1630	-3.7566	-0.6292	4.3858
2009	17	3497919	BCR	3.0116	-1.8045	-0.5460	2.3505
2009	20	3497917	BCR	2.5184	-2.8116	-0.0477	2.8593

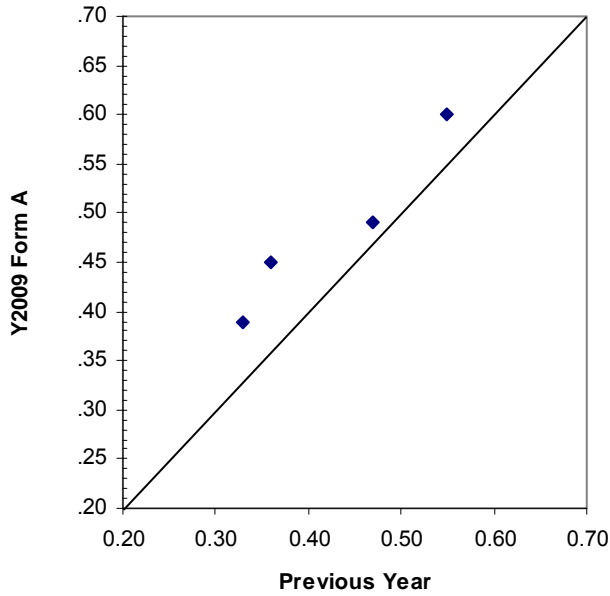
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.6 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 4 Form B**

**Table 1.36 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form A**

Item Number	CID	Item Type	Year 07	Year 09
11	3486368	BCR	0.55	0.60
14	3486369	BCR	0.36	0.45
17	3468082	BCR	0.47	0.49
20	3468083	BCR	0.33	0.39



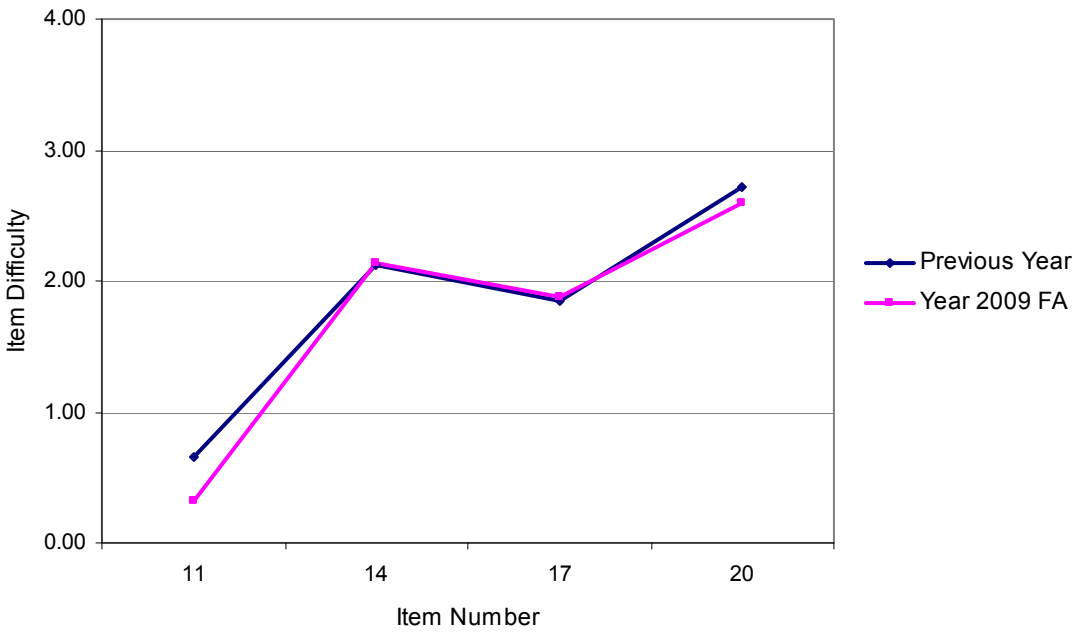
**Table 1.37 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3486368	BCR	2,184	1.64	0.58	1.79	33.61	60.99	2.66	0.96
2007	14	3486369	BCR	2,184	1.08	0.76	21.66	45.92	29.21	1.33	1.88
2007	17	3468082	BCR	2,240	1.40	0.61	5.58	49.24	44.38	0.58	0.22
2007	20	3468083	BCR	2,240	1.00	0.68	20.00	54.96	21.70	0.45	2.90
2009	11	3486368	BCR	30,193	1.80	0.51	0.37	23.62	71.26	4.46	0.28
2009	14	3486369	BCR	30,193	1.34	0.80	16.90	33.20	46.17	2.95	0.77
2009	17	3468082	BCR	30,193	1.48	0.59	3.55	46.13	48.68	1.39	0.25
2009	20	3468083	BCR	30,193	1.16	0.65	12.31	59.46	26.23	1.23	0.77

**Table 1.38 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 5 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3486368	BCR	0.6585	-3.5450	-0.4872	4.0323
2007	14	3486369	BCR	2.1154	-2.3974	-0.4757	2.8732
2007	17	3468082	BCR	1.8398	-3.8217	-0.6623	4.4839
2007	20	3468083	BCR	2.7151	-3.1640	-0.3744	3.5384
2009	11	3486368	BCR	0.3206	-4.3169	-0.3328	4.6497
2009	14	3486369	BCR	2.1384	-2.1325	-0.9141	3.0466
2009	17	3468082	BCR	1.8843	-3.6346	-0.6491	4.2837
2009	20	3468083	BCR	2.5891	-3.1967	-0.0013	3.1980

Note. Rasch item and step difficulties were placed on a common scale.

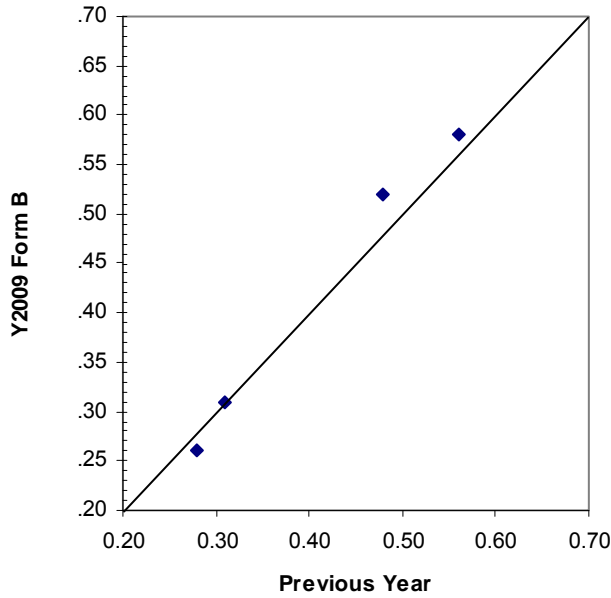


**Figure 1.7 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form A**



**Table 1.39 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form B**

Item Number	CID	Item Type	Year 07	Year 09
11	3296564	BCR	0.48	0.52
14	3296565	BCR	0.56	0.58
17	3486188	BCR	0.31	0.31
20	3486190	BCR	0.28	0.26



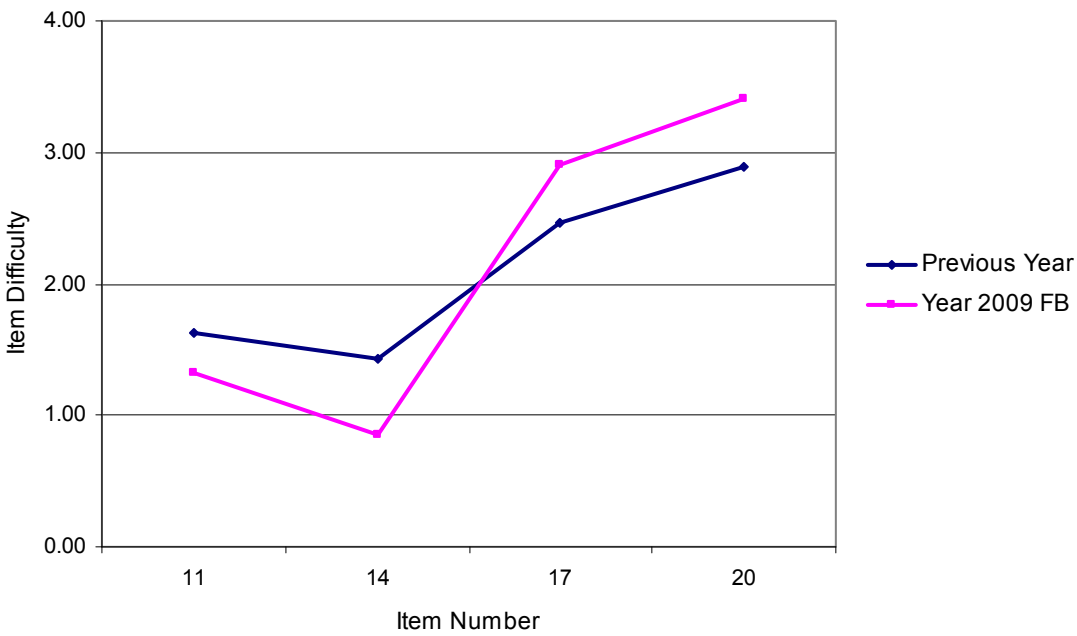
**Table 1.40 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3296564	BCR	2,196	1.43	0.53	1.37	54.64	43.44	0.32	0.23
2007	14	3296565	BCR	2,196	1.69	0.51	1.14	26.96	70.77	0.23	0.91
2007	17	3486188	BCR	2,176	0.92	0.76	30.42	45.17	21.23	1.33	1.84
2007	20	3486190	BCR	2,176	0.85	0.67	27.34	54.00	14.66	0.46	3.54
2009	11	3296564	BCR	29,986	1.56	0.55	1.22	42.86	54.39	1.39	0.14
2009	14	3296565	BCR	29,986	1.74	0.48	0.35	26.28	71.66	1.53	0.19
2009	17	3486188	BCR	29,986	0.92	0.75	30.30	46.36	21.05	1.24	1.06
2009	20	3486190	BCR	29,986	0.79	0.69	35.08	50.20	13.21	0.74	0.77

**Table 1.41 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 5 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3296564	BCR	1.6248	-4.9990	-0.3491	5.3481
2007	14	3296565	BCR	1.4227	-4.5535	-1.5788	6.1323
2007	17	3486188	BCR	2.4624	-2.2193	-0.3325	2.5518
2007	20	3486190	BCR	2.8860	-2.8858	-0.1152	3.0010
2009	11	3296564	BCR	1.3236	-4.3522	-0.4184	4.7706
2009	14	3296565	BCR	0.8486	-4.5717	-0.7313	5.3030
2009	17	3486188	BCR	2.9082	-2.2503	-0.2926	2.5429
2009	20	3486190	BCR	3.4047	-2.5754	-0.0174	2.5928

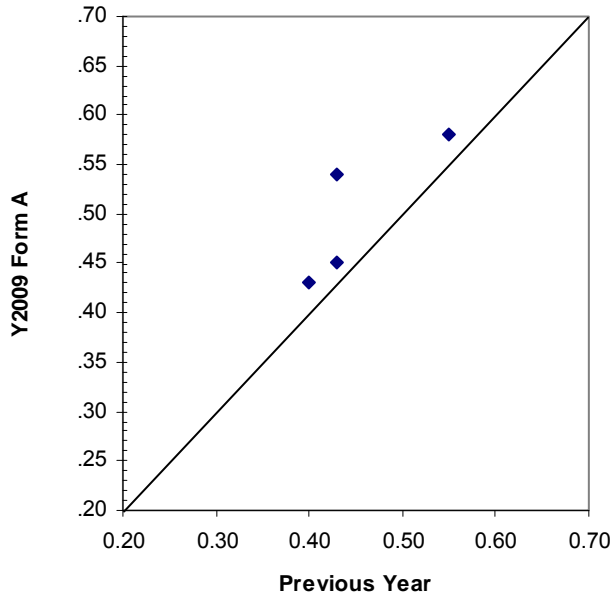
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.8 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 5 Form B**

**Table 1.42 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form A**

Item Number	CID	Item Type	Year 07	Year 09
11	3470028	BCR	0.55	0.58
14	3470029	BCR	0.43	0.54
17	3498432	BCR	0.43	0.45
20	3498435	BCR	0.40	0.43



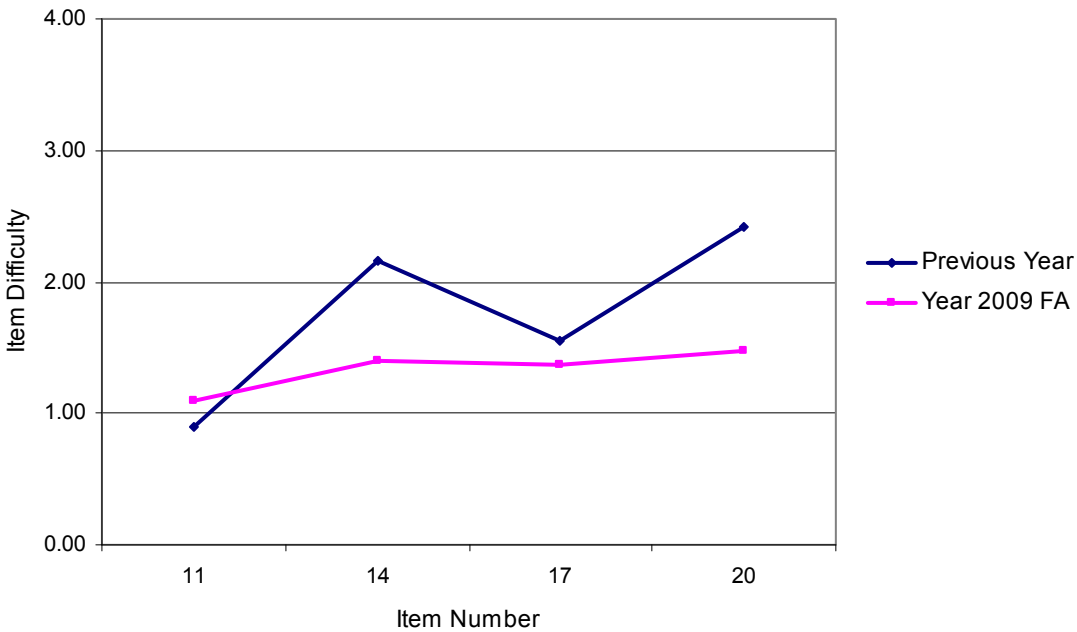
**Table 1.43 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3470028	BCR	2,421	1.64	0.69	5.04	28.58	59.11	5.58	1.69
2007	14	3470029	BCR	2,421	1.29	0.74	12.80	40.15	42.75	0.95	3.35
2007	17	3498432	BCR	2,166	1.29	0.57	3.74	62.33	31.53	1.29	1.11
2007	20	3498435	BCR	2,166	1.19	0.62	8.73	58.40	30.01	0.28	2.59
2009	11	3470028	BCR	29,751	1.73	0.58	3.24	22.57	70.54	3.26	0.39
2009	14	3470029	BCR	29,751	1.62	0.62	3.98	30.31	62.31	2.42	0.97
2009	17	3498432	BCR	29,751	1.36	0.57	1.70	62.17	33.29	2.36	0.47
2009	20	3498435	BCR	29,751	1.29	0.55	1.83	67.64	27.34	2.16	1.03

**Table 1.44 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 6 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3470028	BCR	0.8972	-2.6134	-0.7275	3.3408
2007	14	3470029	BCR	2.1533	-2.9458	-0.9954	3.9412
2007	17	3498432	BCR	1.5488	-3.9655	0.3399	3.6256
2007	20	3498435	BCR	2.4180	-3.8242	-0.4813	4.3056
2009	11	3470028	BCR	1.1024	-2.6691	-1.1491	3.8183
2009	14	3470029	BCR	1.4032	-2.7761	-1.1685	3.9446
2009	17	3498432	BCR	1.3735	-4.4358	0.7564	3.6794
2009	20	3498435	BCR	1.4715	-4.5049	0.9811	3.5239

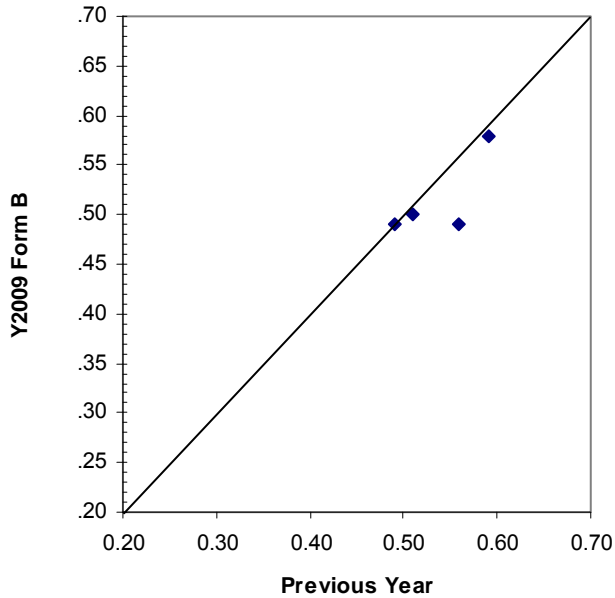
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.9 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form A**

**Table 1.45 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form B**

Item Number	CID	Item Type	Year 07	Year 09
11	3470041	BCR	0.49	0.49
14	3470039	BCR	0.59	0.58
17	3489694	BCR	0.56	0.49
20	3489696	BCR	0.51	0.50



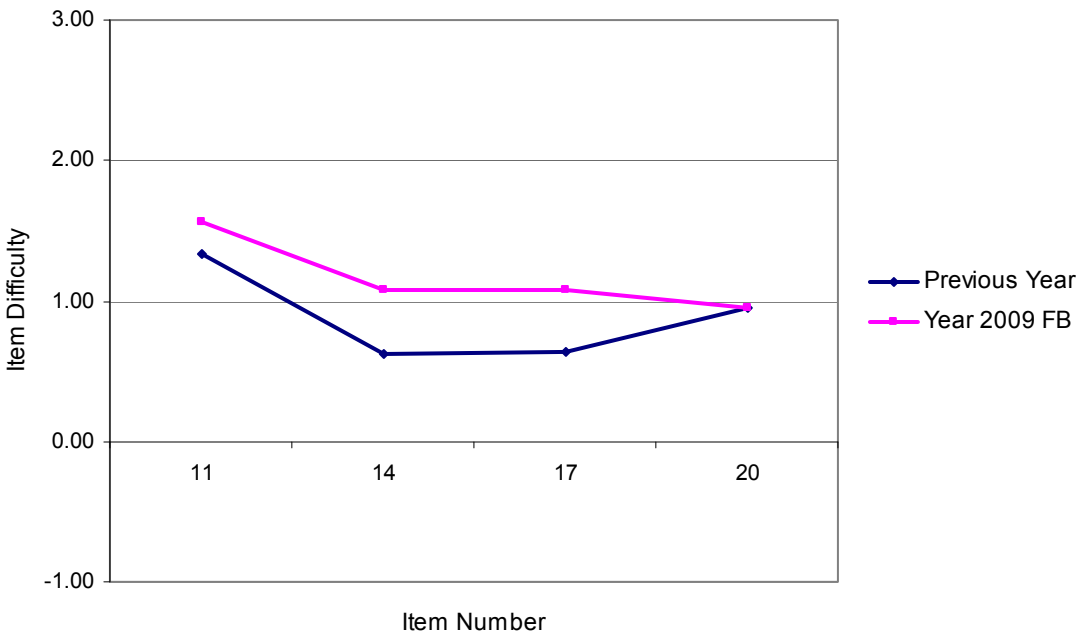
**Table 1.46 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	11	3470041	BCR	2,377	1.46	0.77	10.14	35.34	47.67	5.26	1.60
2007	14	3470039	BCR	2,377	1.78	0.59	2.57	18.93	72.07	4.96	1.47
2007	17	3489694	BCR	2,179	1.68	0.56	1.47	30.56	64.53	2.85	0.60
2007	20	3489696	BCR	2,179	1.52	0.62	2.25	43.41	49.75	2.89	1.70
2009	11	3470041	BCR	29,070	1.47	0.70	8.36	38.98	48.73	3.43	0.50
2009	14	3470039	BCR	29,070	1.74	0.57	3.21	21.61	71.80	2.90	0.49
2009	17	3489694	BCR	29,070	1.48	0.55	0.81	51.83	45.42	1.65	0.29
2009	20	3489696	BCR	29,070	1.49	0.58	1.04	51.35	44.46	2.75	0.41

**Table 1.47 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	11	3470041	BCR	1.3304	-2.2681	-0.5478	2.8159
2007	14	3470039	BCR	0.6282	-2.6404	-1.1747	3.8150
2007	17	3489694	BCR	0.6439	-3.7005	-0.5408	4.2413
2007	20	3489696	BCR	0.9453	-3.7393	-0.0646	3.8039
2009	11	3470041	BCR	1.5577	-2.4112	-0.6607	3.0719
2009	14	3470039	BCR	1.0770	-2.8849	-1.1457	4.0306
2009	17	3489694	BCR	1.0755	-4.5144	0.1215	4.3928
2009	20	3489696	BCR	0.9476	-4.4842	0.4649	4.0192

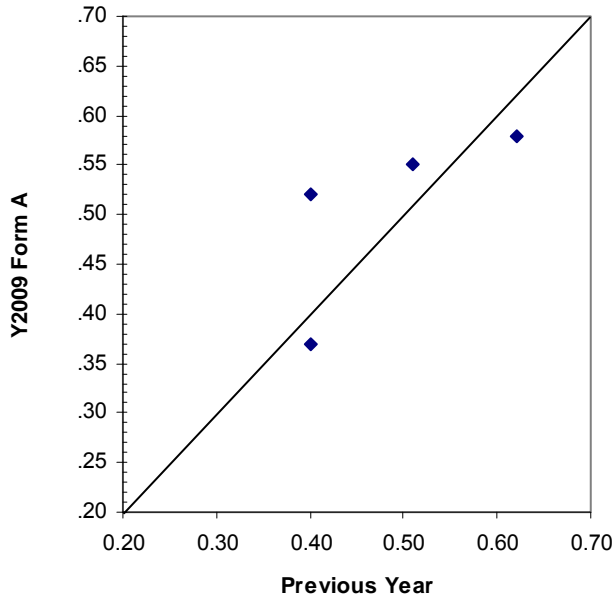
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.10 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 6 Form B**

**Table 1.48 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form A**

Item Number	CID	Item Type	Year 07	Year 09
7	3470050	BCR	0.62	0.58
10	3470051	BCR	0.40	0.52
13	3468879	BCR	0.40	0.37
16	3468877	BCR	0.51	0.55



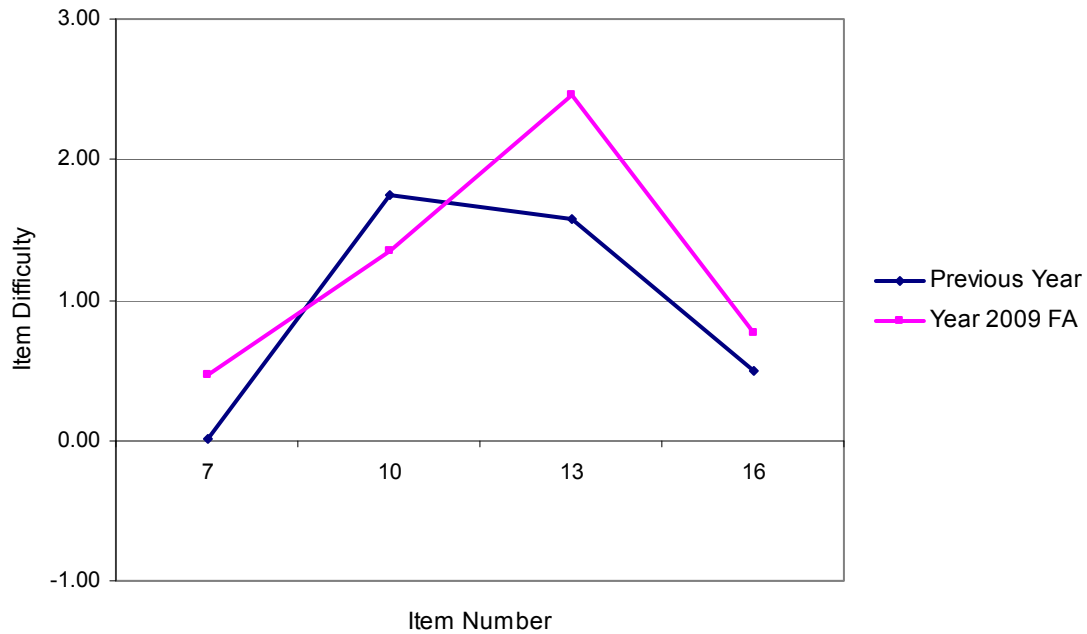
**Table 1.49 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	7	3470050	BCR	2,436	1.87	0.66	2.59	18.27	65.93	12.19	1.03
2007	10	3470051	BCR	2,436	1.21	0.87	21.67	29.76	41.05	2.96	4.56
2007	13	3468879	BCR	2,187	1.20	0.80	15.50	47.01	28.94	5.03	3.52
2007	16	3468877	BCR	2,187	1.53	0.68	1.51	46.91	42.16	7.13	2.29
2009	7	3470050	BCR	30,046	1.74	0.56	0.63	29.12	65.02	4.83	0.40
2009	10	3470051	BCR	30,046	1.56	0.68	4.94	34.57	54.77	4.05	1.67
2009	13	3468879	BCR	30,046	1.10	0.80	23.26	43.95	28.59	2.90	1.31
2009	16	3468877	BCR	30,046	1.64	0.62	1.23	37.27	54.79	5.85	0.86

**Table 1.50 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 7 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	7	3470050	BCR	0.0144	-2.2692	-0.8359	3.1051
2007	10	3470051	BCR	1.7408	-1.7946	-1.0423	2.8368
2007	13	3468879	BCR	1.5702	-2.3257	0.0583	2.2674
2007	16	3468877	BCR	0.4958	-3.6492	0.5098	3.1394
2009	7	3470050	BCR	0.4728	-4.1484	-0.2496	4.3980
2009	10	3470051	BCR	1.3428	-3.1612	-0.6117	3.7729
2009	13	3468879	BCR	2.4622	-2.2283	-0.2021	2.4304
2009	16	3468877	BCR	0.7640	-4.1250	0.2255	3.8995

Note. Rasch item and step difficulties were placed on a common scale.

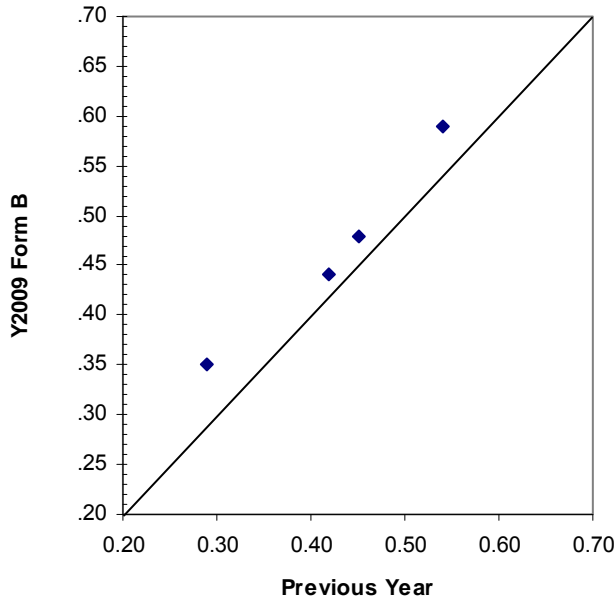


**Figure 1.11 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form A**



**Table 1.51 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form B**

Item Number	CID	Item Type	Year 07	Year 09
7	3497797	BCR	0.45	0.48
10	3497798	BCR	0.42	0.44
13	3468867	BCR	0.29	0.35
16	3468866	BCR	0.54	0.59



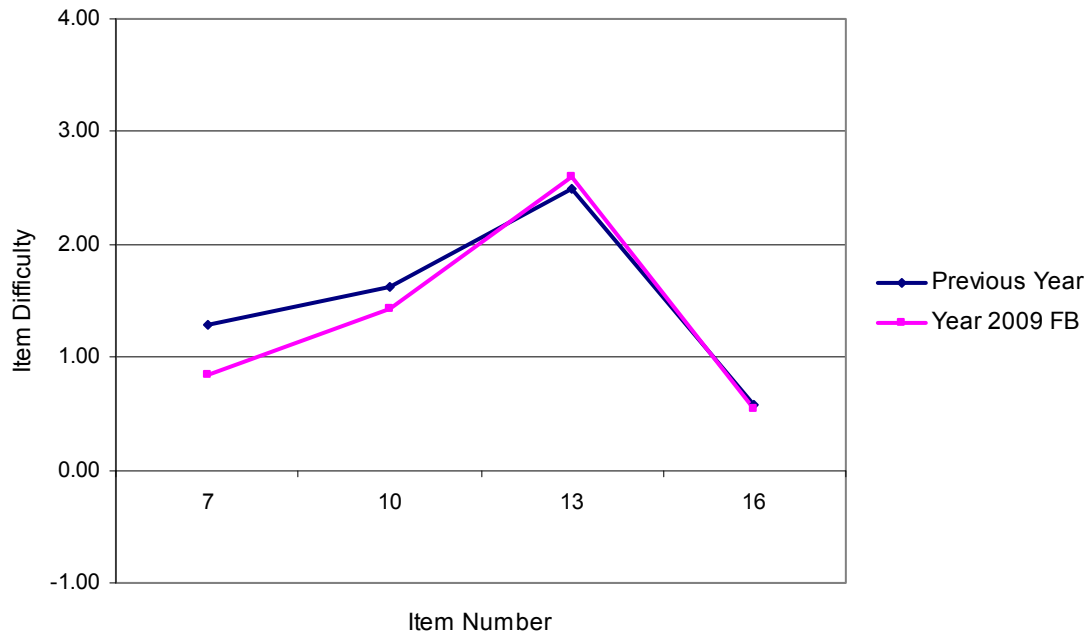
**Table 1.52 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	7	3497797	BCR	2,217	1.34	0.57	1.89	60.22	34.82	1.26	1.80
2007	10	3497798	BCR	2,217	1.27	0.62	5.19	58.37	32.43	1.22	2.80
2007	13	3468867	BCR	2,191	0.87	0.82	36.42	35.78	23.46	1.51	2.83
2007	16	3468866	BCR	2,191	1.61	0.83	6.25	33.36	43.95	13.28	3.15
2009	7	3497797	BCR	29,541	1.45	0.56	0.42	55.09	41.41	2.47	0.61
2009	10	3497798	BCR	29,541	1.31	0.53	1.00	66.32	30.66	1.28	0.74
2009	13	3468867	BCR	29,541	1.06	0.87	31.63	30.14	34.82	2.12	1.29
2009	16	3468866	BCR	29,541	1.76	0.77	2.26	34.22	45.10	17.09	1.32

**Table 1.53 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 7 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	7	3497797	BCR	1.2870	-4.4535	0.2650	4.1885
2007	10	3497798	BCR	1.6206	-3.8434	0.0369	3.8065
2007	13	3468867	BCR	2.4933	-1.8749	-0.697	2.5719
2007	16	3468866	BCR	0.5711	-2.1601	-0.0661	2.2261
2009	7	3497797	BCR	0.8441	-5.4570	0.8736	4.5834
2009	10	3497798	BCR	1.4321	-5.1787	0.9201	4.2586
2009	13	3468867	BCR	2.5970	-1.6856	-0.9921	2.6777
2009	16	3468866	BCR	0.5434	-3.0723	0.5310	2.5413

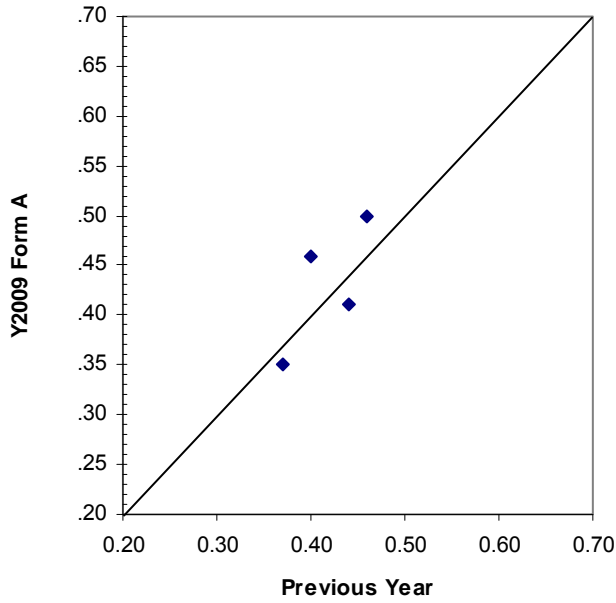
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.12 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 7 Form B**

**Table 1.54 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form A**

Item Number	CID	Item Type	Year 07	Year 09
7	3514210	BCR	0.46	0.50
10	3514209	BCR	0.40	0.46
13	3327522	BCR	0.44	0.41
16	3327523	BCR	0.37	0.35



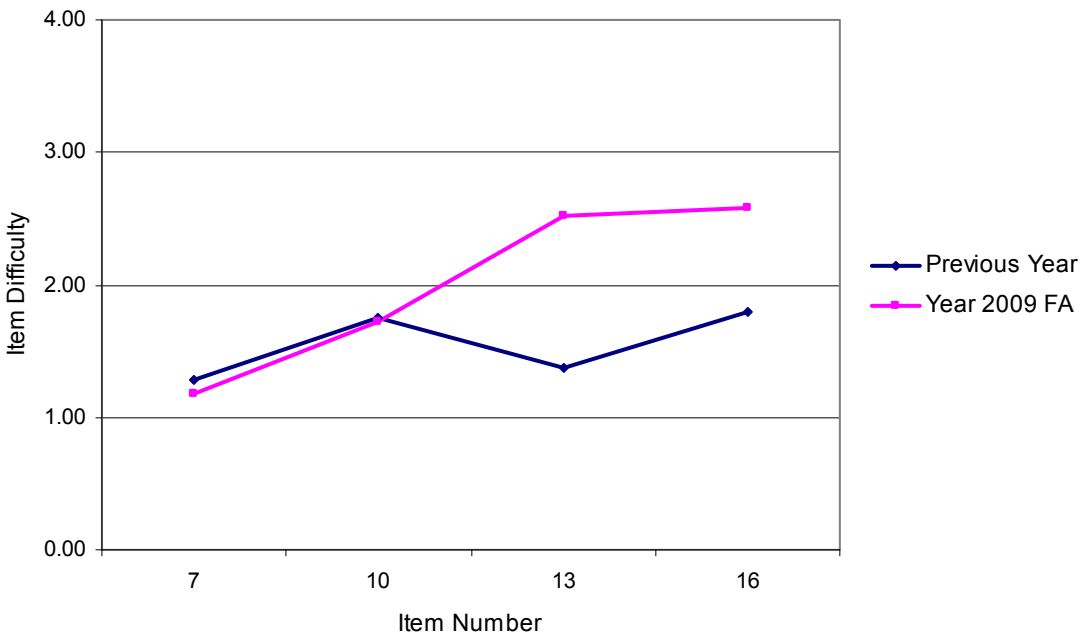
**Table 1.55 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form A**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	7	3514210	BCR	2,245	1.39	0.66	5.21	47.93	42.18	2.27	2.41
2007	10	3514209	BCR	2,245	1.20	0.78	16.84	43.07	34.30	2.76	3.03
2007	13	3327522	BCR	2,204	1.33	0.78	11.57	42.88	37.93	4.67	2.95
2007	16	3327523	BCR	2,204	1.11	0.73	15.43	53.54	24.55	2.77	3.72
2009	7	3514210	BCR	30,717	1.51	0.66	4.07	43.88	47.07	4.33	0.64
2009	10	3514209	BCR	30,717	1.38	0.71	10.59	40.70	45.68	2.11	0.91
2009	13	3327522	BCR	30,717	1.23	0.68	12.48	50.53	35.41	0.61	0.97
2009	16	3327523	BCR	30,717	1.06	0.64	15.30	60.86	21.72	0.64	1.47

**Table 1.56 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form A**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	7	3514210	BCR	1.2866	-3.2991	-0.2057	3.5049
2007	10	3514209	BCR	1.7496	-2.3321	-0.4262	2.7583
2007	13	3327522	BCR	1.3792	-2.2408	-0.1764	2.4172
2007	16	3327523	BCR	1.7986	-2.4663	0.1969	2.2694
2009	7	3514210	BCR	1.1815	-3.1976	0.0552	3.1424
2009	10	3514209	BCR	1.7222	-2.4424	-0.4761	2.9186
2009	13	3327522	BCR	2.5193	-2.9607	-0.5414	3.5021
2009	16	3327523	BCR	2.5818	-3.0308	0.0831	2.9476

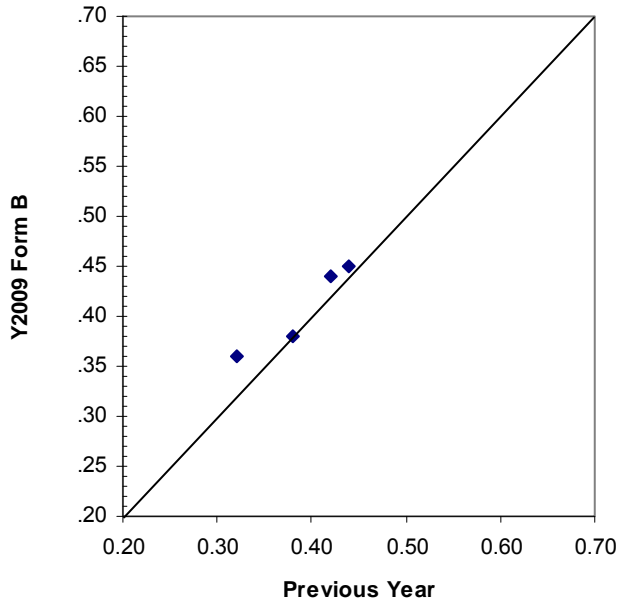
Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.13 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form A**

**Table 1.57 P-Value Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form B**

Item Number	CID	Item Type	Year 07	Year 09
7	3470065	BCR	0.32	0.36
10	3470063	BCR	0.42	0.44
13	3489342	BCR	0.44	0.45
16	3489340	BCR	0.38	0.38



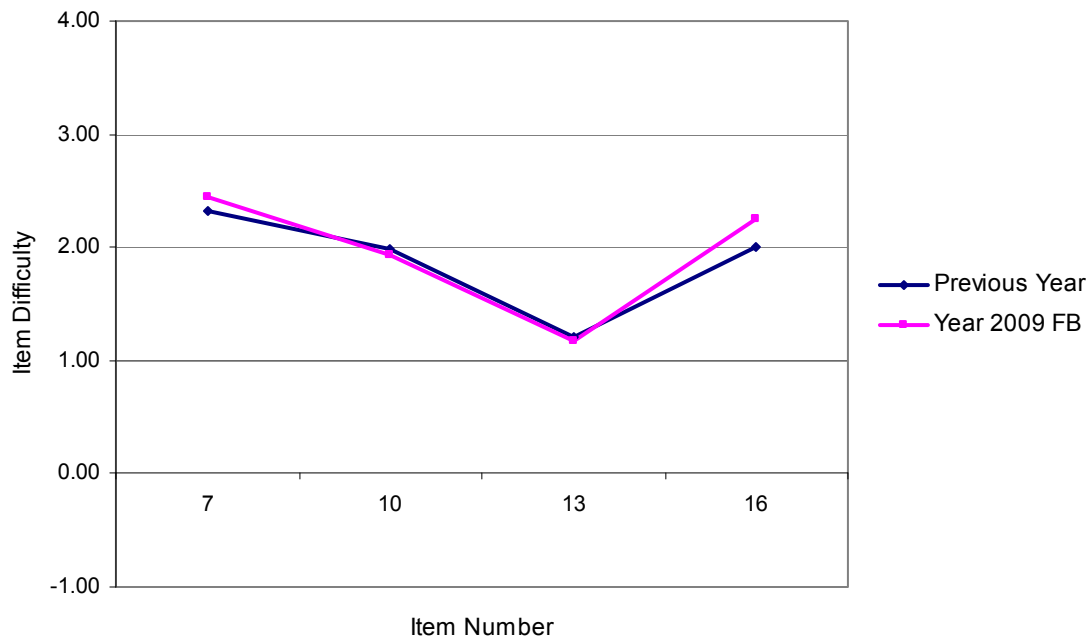
**Table 1.58 Score-Point Distribution Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form B**

Year	Item #	CID	Item Type	N	Mean	SD	Score-Point Distribution (%)				
							0	1	2	3	Omit
2007	7	3470065	BCR	2,165	0.97	0.78	26.37	44.20	24.34	1.43	3.65
2007	10	3470063	BCR	2,165	1.26	0.67	8.64	51.55	35.80	0.79	3.23
2007	13	3489342	BCR	2,190	1.32	0.60	2.97	60.32	32.83	2.01	1.87
2007	16	3489340	BCR	2,190	1.13	0.59	6.76	65.34	23.06	0.59	4.25
2009	7	3470065	BCR	30,235	1.07	0.76	21.89	48.21	26.63	1.92	1.34
2009	10	3470063	BCR	30,235	1.31	0.58	4.65	58.56	35.06	0.75	0.98
2009	13	3489342	BCR	30,235	1.36	0.53	0.98	61.88	35.67	0.82	0.65
2009	16	3489340	BCR	30,235	1.13	0.52	5.60	73.50	19.01	0.62	1.27

**Table 1.59 Rasch Item and Step Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009:  
Grade 8 Form B**

Year	Item #	CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2007	7	3470065	BCR	2.3157	-2.1770	-0.3799	2.5569
2007	10	3470063	BCR	1.9832	-3.3219	-0.4688	3.7907
2007	13	3489342	BCR	1.2063	-3.7760	0.4545	3.3215
2007	16	3489340	BCR	1.9975	-3.7463	0.2146	3.5318
2009	7	3470065	BCR	2.4400	-2.2582	-0.1749	2.4331
2009	10	3470063	BCR	1.9256	-3.8668	0.1164	3.7504
2009	13	3489342	BCR	1.1781	-5.0342	0.7254	4.3088
2009	16	3489340	BCR	2.2508	-4.0963	0.8369	3.2594

Note. Rasch item and step difficulties were placed on a common scale.



**Figure 1.14 Rasch Item Difficulty Comparisons of BCR Items for Year 2007 vs. Year 2009: Grade 8 Form B**

## 1.9 Linking, Equating, and Scaling Procedures

For the purpose of year-to-year linking and equating, we constructed a 2009 linking pool: we included only operational selected-response (SR) items (i.e., multiple choice items) that appeared in both years (i.e., 2007 and 2009). It should be noted that all the classical and Rasch analyses of the 2007 assessment were conducted with field-test samples. After setting up the linking pool, we then conducted a stability check of linking items and decided which items should be excluded from or which item should remain in the linking pool. During the calibration and equating process, we kept and fixed the original operational Rasch item difficulty parameters (i.e., 2007) 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 2003 (i.e., grades 3, 5, and 8) or 2004 assessment (i.e., 4, 6, and 7) and all the scale scores of different years were comparable within each content and grade.

### Stratified Random Sampling Procedures

To select equating samples, a stratified random sampling method was applied in 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-Reading 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

Robust z values 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 B
- The correlation between operational form A and B item difficulties
- The difference between operational form A and B 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 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-Reading 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 B
- Correlate Rasch item difficulties between form A and form B
- 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 1.60 through 1.65 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 2003 or 2004 assessment).



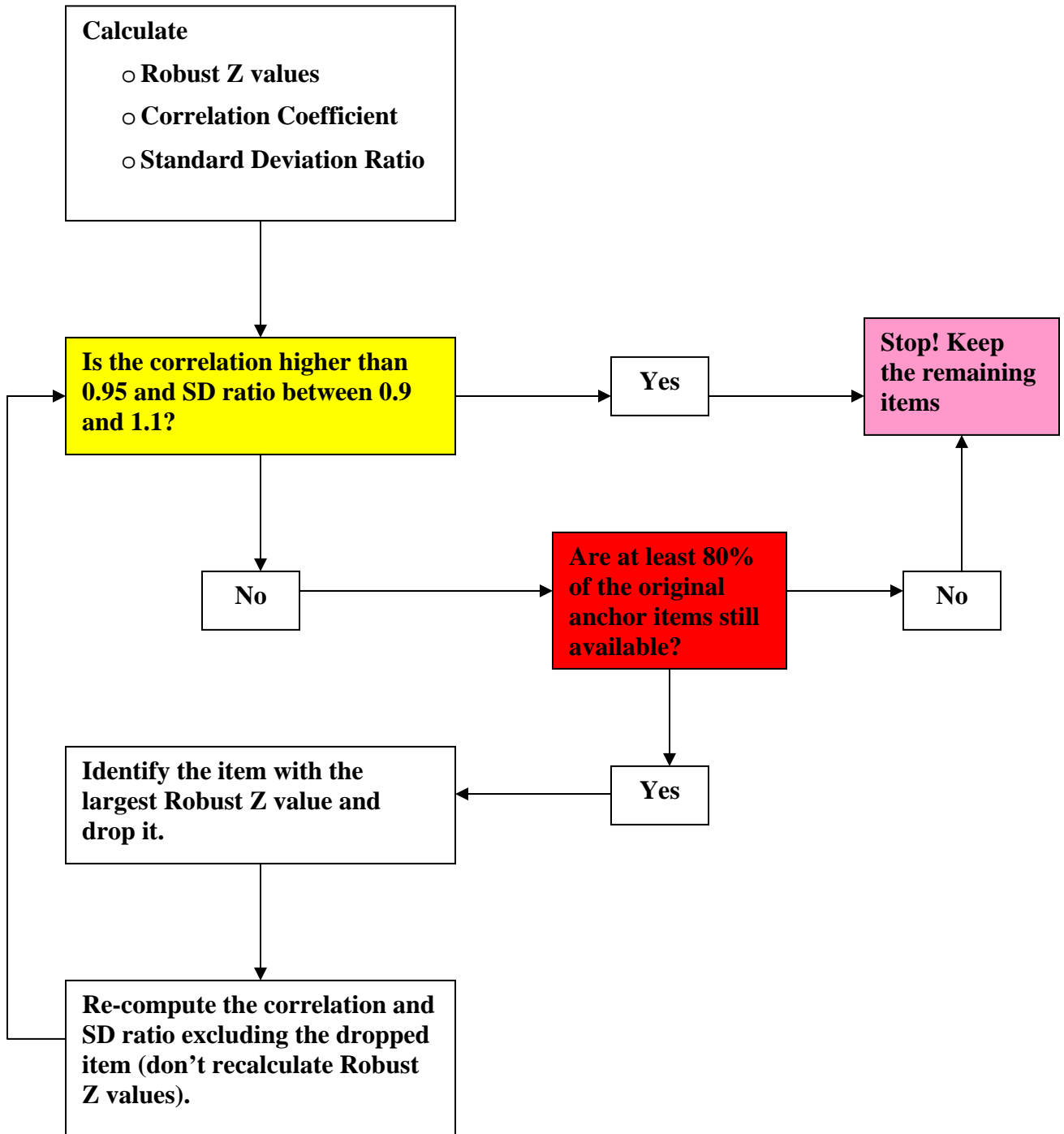


Figure 1.15 Anchor Evaluation Steps Chart for MSA-Reading

**Table 1.60 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 3**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
1	-1.3708	-1.9877	-2.8338	1	-1.3708	-1.9568	-3.6815
2	-3.3089	-3.2452	.3652	2	-3.3089	-3.2466	-.2753
3	-0.1994	0.0594	1.2823	3	-0.1994	0.0266	.5848
4	-1.1969	-1.2702	-.2787	4	-1.1969	-1.1682	-.4519
5	-1.4394	-1.4803	-.1264	5	-1.4394	-1.4897	-.8669
6	-2.1425	-2.1551	.0066	6	-2.1425	-2.2075	-.9442
7	0.1005	0.2326	.6867	7	0.1005	0.2928	.4077
8	0.4026	0.3155	-.3436	8	0.4026	0.4108	-.5596
9	-0.3764	-0.3244	.3102	9	-0.3764	-0.2493	.0652
22	-0.3948	-0.5623	-.7215	22	-0.3948	-0.5862	-1.6083
23	-0.021	-0.3393	-1.4303	23	-0.0210	-0.2944	-2.0391
24	0.2441	0.1219	-.5086	24	0.2441	0.2353	-.6489
25	1.051	1.0790	.1974	25	1.0510	1.1684	.0142
26	-0.0453	-0.0593	.0000	26	-0.0453	-0.0036	-.3835
27	1.1168	1.2998	.9260	27	1.1168	1.4137	.9573
28	-0.7649	-0.6491	.6101	28	-0.7649	-0.5580	.4844
29	0.2949	0.2763	-.0216	29	0.2949	0.4096	.0000
36	0.5978	0.6616	.3657	36	0.5978	0.8932	.9494
37	-1.387	-1.1954	.9664	37	-1.3870	-1.2002	.3788
38	-0.8873	-0.7730	.6031	38	-0.8873	-0.7483	.1277
39	-0.5402	-0.2787	1.2949	39	-0.5402	-0.3090	.6121
40	0.829	1.1165	1.4172	40	0.8290	1.0869	.7524
41	-0.5329	-0.4129	.6298	41	-0.5329	-0.3839	.1802
42	0.3757	0.5821	1.0360	42	0.3757	0.5942	.5454
43	0.5816	0.7467	.8418	43	0.5816	0.8732	.9294
10A	-1.0069	-1.3060	-1.3401	10B	-0.0200	-0.2932	-2.0381
12A	-0.3612	-0.4883	-.5316	12B	-0.6084	-0.2975	1.0309
13A	-0.3214	-0.4743	-.6529	13B	-0.0259	-0.2030	-1.5331
15A	0.3988	0.1948	-.8931	15B	-0.1183	-0.2341	-1.2111
16A	0.8294	0.3166	-2.3445	16B	0.9690	0.9405	-.7524
18A	-0.0978	-0.4528	-1.6028	18B	0.3083	0.3779	-.2370
19A	0.0727	-0.1107	-.7962	19B	-0.5740	-0.4055	.2827
21A	0.2195	-0.2519	-2.1499	21B	0.2044	-0.0506	-1.9424

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

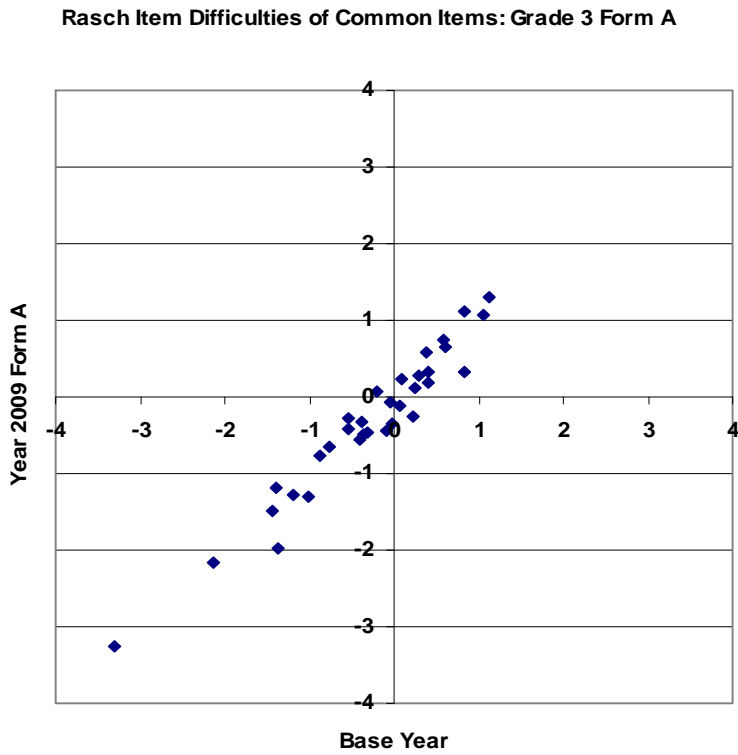
*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.

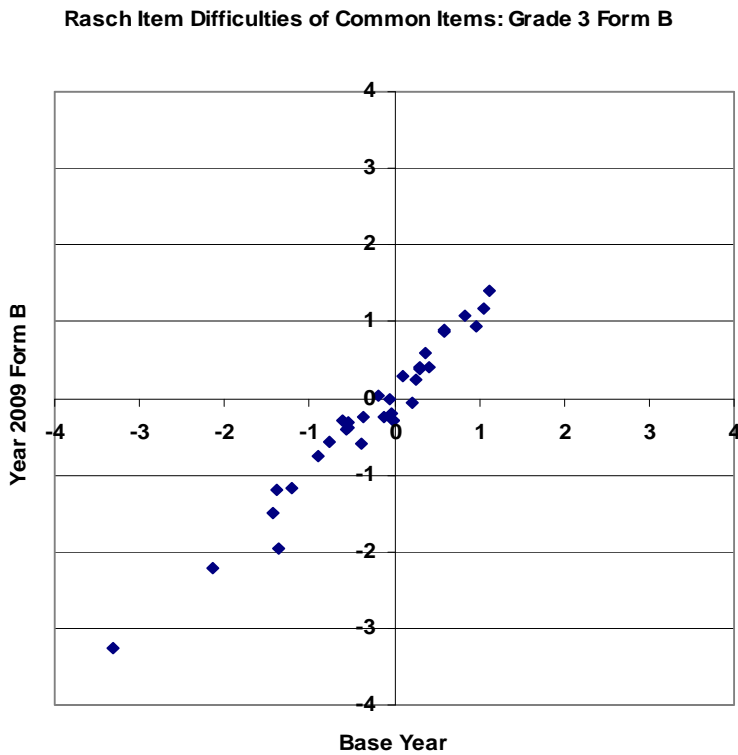
**Form Statistics**

Form Statistics	Previous Base Form	2009 Form A	Previous Base Form	2009 Form B
Mean	-.281	-.328	-.269	-.217
SD	.950	.983	.948	1.011
		2009 Form A		2009 Form B
Correlation		<b>.972</b>		<b>.980</b>
SD Ratio		<b>103%</b>		<b>107%</b>
Mean Diff		-.046		.052
Median Diff		-.014		.115
IQR Diff		.288		.257

Based on correlation coefficients, SD ratios, robust z values, and item difficulty plot, none of the linking common items were dropped from the linking pool.



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



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

**Table 1.61 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 4**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
1	0.8485	0.3571	-.0152	1	0.8485	0.2796	.0146
2	-0.2773	-0.6124	.6817	2	-0.2773	-0.6946	.7289
3	-0.0474	-0.5057	.1324	3	-0.0474	-0.6161	.0155
4	-1.694	-2.2060	-.1070	4	-1.6940	-2.2923	-.1239
<b>5*</b>	<b>-1.3091</b>	<b>-2.5919</b>	<b>-3.5436</b>	<b>5*</b>	<b>-1.3091</b>	<b>-2.7488</b>	<b>-4.0885</b>
<b>6*</b>	<b>-1.4119</b>	<b>-2.8926</b>	<b>-4.4259</b>	<b>6*</b>	<b>-1.4119</b>	<b>-2.8638</b>	<b>-4.1459</b>
7	1.2863	0.8432	.2002	7	1.2863	0.7578	.2050
8	-1.7138	-2.4408	-1.0656	8	-1.7138	-2.5214	-1.1101
9	-0.9493	-1.6549	-.9702	9	-0.9493	-1.7223	-.9471
22	1.0781	0.3584	-1.0330	22	1.0781	0.2745	-1.0913
23	0.5162	-0.4655	-2.2011	23	0.5162	-0.4750	-1.9752
24	0.7249	0.1186	-.5274	24	0.7249	0.0802	-.3425
25	1.3574	1.2125	1.5297	25	1.3574	1.2224	2.0591
26	0.9364	0.5327	.3758	26	0.9364	0.4857	.5715
27	1.4992	0.8864	-.5564	27	1.4992	0.8232	-.4900
28	1.2577	0.8905	.5386	28	1.2577	0.8268	.6648
29	0.1196	-0.3717	-.0147	29	0.1196	-0.4706	-.0858
36	0.6644	0.2234	.2095	36	0.6644	0.2478	.7322
37	0.9481	0.7480	1.2836	37	0.9481	0.7148	1.5959
38	0.7393	0.3082	.2537	38	0.7393	0.3336	.7836
39	-0.0274	-0.5066	.0392	39	-0.0274	-0.6360	-.1725
40	1.0705	0.6680	.3812	40	1.0705	0.7749	1.3023
41	0.0061	-0.3150	.7441	41	0.0061	-0.3027	1.2402
42	0.081	-0.2768	.5805	42	0.0810	-0.3803	.5216
43	1.4931	1.2033	.8837	43	1.4931	1.0755	.7275
10A	0.0996	-0.6344	-1.0968	10B	0.3535	-0.4312	-1.0022
12A	0.2402	-0.6939	-1.9889	12B	0.3765	-0.2578	-.2935
13A	0.3033	-0.5334	-1.5547	13B	-0.2149	-0.9341	-.6936
15A	0.4782	0.4322	1.9706	15B	1.2219	0.6496	-.0014
16A	1.9046	1.3277	-.3964	16B	1.1396	0.4219	-.6865
18A	0.9083	0.4311	.0482	18B	1.4526	0.8806	.0000
19A	0.7115	0.2235	.0000	19B	0.4873	-0.0743	.0490
21A	0.1834	-0.4925	-.8377	21B	-0.5476	-1.0411	.3699

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.

**Form Statistics**

	Previous	2009	Previous	2009
Form Statistics	Base Form	Form A	Base Form	Form B
Mean	.364	-.195	.347	-.261
SD	.930	1.109	.942	1.117
Correlation Coefficient		.972		.977
SD Ratio		<b>119%</b>		<b>118%</b>
Mean Diff		-.559		-.608
Median Diff		-.488		-.572
IQR Diff		.303		.287

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 5 and 6 appearing on both forms were dropped from the linking pool.

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

Correlation Coefficient	<b>.974</b>	<b>.982</b>
SD Ratio	<b>109%</b>	<b>108%</b>

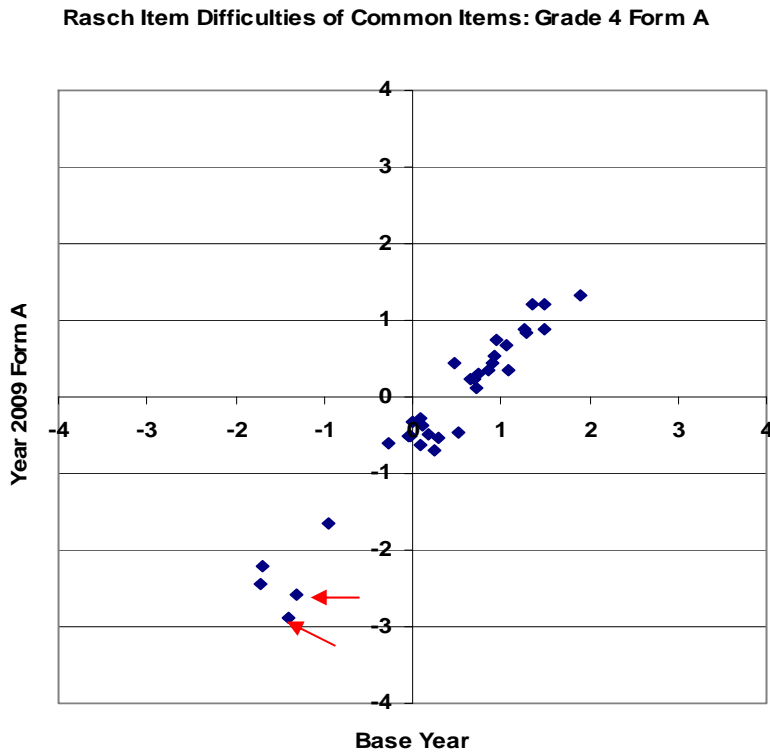


Figure 1.18 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 4 Form A

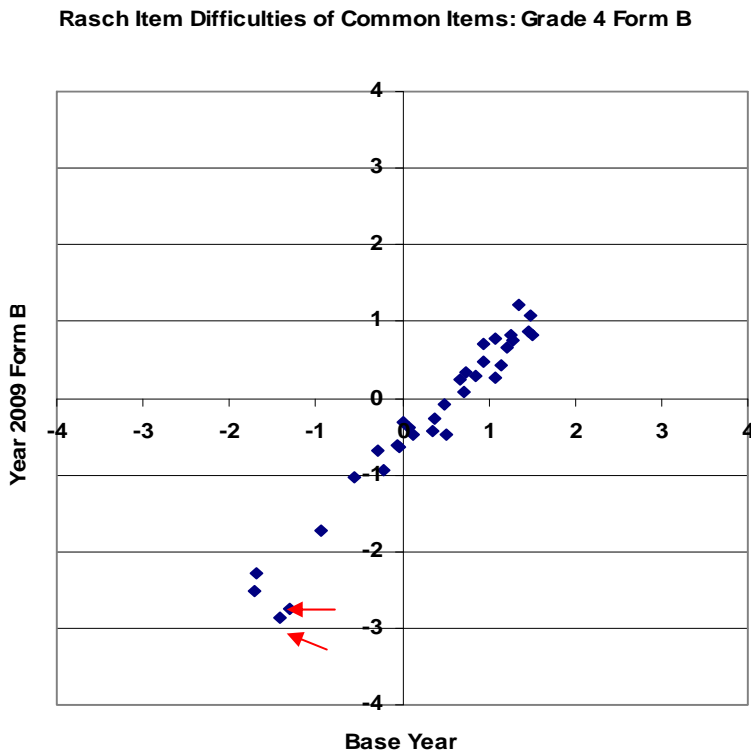


Figure 1.19 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 4 Form B

**Table 1.62 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 5**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
1	-0.6371	-0.8749	-.4379	1	-0.6371	-1.0027	-.4716
2	-0.2093	-0.4176	-.3291	2	-0.2093	-0.4800	-.1503
3	-1.2263	-1.4995	-.5685	3	-1.2263	-1.6520	-.6750
4	-1.4827	-1.8174	-.7954	4	-1.4827	-2.0212	-1.0568
5	-1.3213	-1.6520	-.7806	5	-1.3213	-1.7098	-.5491
6	-1.8707	-2.2796	-1.0691	6	-1.8707	-2.4124	-1.0677
7	-1.0118	-1.4698	-1.2503	7	-1.0118	-1.4995	-.8849
8	1.4561	1.2817	-.2040	8	1.4561	1.2298	.0000
9	-1.7612	-1.6206	.9581	9	-1.7612	-1.8838	.3510
22	0.1781	0.1932	.4951	22	0.1781	-0.1582	-.3724
23	1.4551	1.0524	-1.0463	23	1.4551	1.0626	-.5626
24	0.3030	-0.0716	-.9426	24	0.3030	-0.1074	-.6232
25	-0.5030	-0.7856	-.6032	25	-0.5030	-1.0467	-1.0744
26	0.2006	0.2362	.5707	26	0.2006	0.1385	.5558
27	0.6870	0.6170	.1811	27	0.6870	0.4813	.0697
28	0.5988	0.3280	-.5597	28	0.5988	0.2341	-.4685
29	0.7593	0.2069	-1.5985	29	0.7593	0.0888	-1.5037
36	0.4436	0.4860	.5958	36	0.4436	0.4276	.7119
37	0.8269	0.4776	-.8493	37	0.8269	0.2336	-1.2423
38	0.7861	0.7265	.2195	38	0.7861	0.5779	.0613
39	1.1761	1.1309	.2726	39	1.1761	0.9393	-.0355
40	0.7230	0.7610	.5796	40	0.7230	0.7118	.7281
41	0.3088	0.4329	.8972	41	0.3088	0.3488	.9015
42	-0.0371	-0.0092	.5423	42	-0.0371	0.0206	.9614
43	-0.1849	0.0280	1.2248	43	-0.1849	-0.1947	.7329
10A	-0.8817	-1.3259	-1.1994	<b>10B*</b>	<b>-1.2624</b>	<b>-2.0112</b>	<b>-1.7687</b>
12A	0.1471	0.1031	.2771	12B	-0.2621	-0.0500	1.4840
13A	0.1540	0.5166	1.7771	<b>13B*</b>	<b>0.0092</b>	<b>0.5333</b>	<b>2.5402</b>
15A	0.1906	-0.1333	-.7555	15B	0.1364	0.2353	1.1009
16A	-0.1963	-0.0361	1.0304	16B	-0.2328	-0.0382	1.4248
18A	-1.2429	-1.3620	.0000	18B	0.5029	0.7255	1.5196
19A	0.1071	0.1073	.4401	19B	1.1421	1.0316	.3920
21A	0.1229	0.1893	.6843	21B	-0.4039	-0.4850	.4915

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.



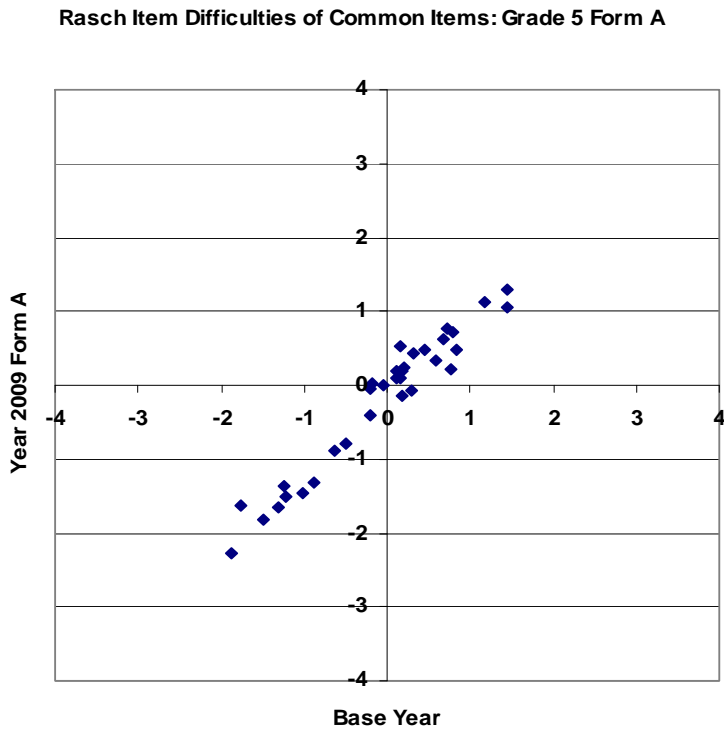
**Form Statistics**

Form Statistics	Previous Year	2009	Previous Year	2009
	Form A	Form A	Form B	Form B
Mean	-.059	-.196	-.022	-.234
SD	.892	.951	.911	1.019
Correlation Coefficient		.972		.960
SD Ratio		107%		<b>112%</b>
Mean Diff		-.138		-.213
Median Diff		-.119		-.226
IQR Diff		.366		.399

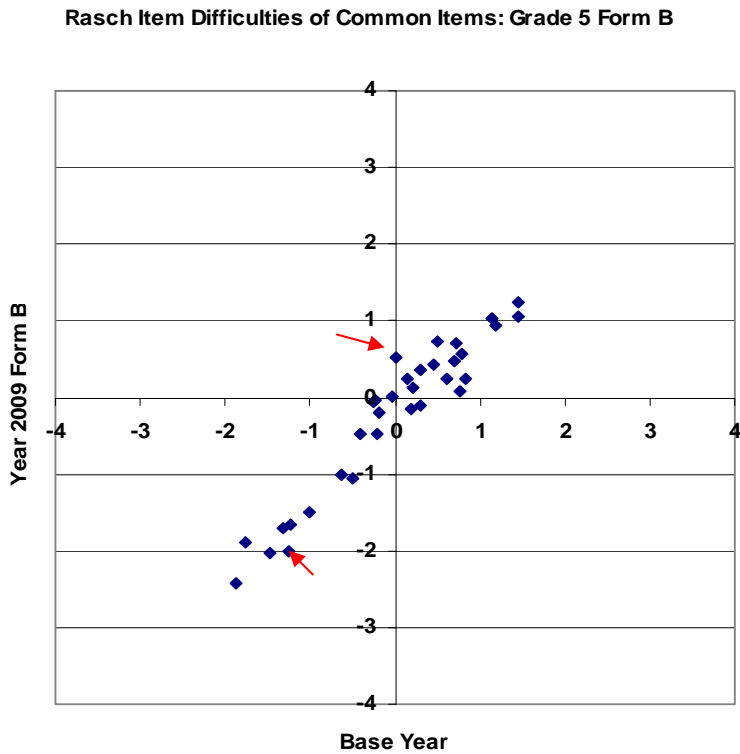
Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 10 and 13 appearing on Form B were dropped from the linking pool.

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

Correlation Coefficient	<b>.972</b>	<b>.968</b>
SD Ratio	<b>107%</b>	<b>109%</b>



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



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

**Table 1.63 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 6**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
1	-1.3336	-1.8853	-2.4265	1	-1.3336	-1.7859	-1.6427
<b>2*</b>	<b>-2.0006</b>	<b>-2.8271</b>	<b>-3.8168</b>	<b>2*</b>	<b>-2.0006</b>	<b>-2.7231</b>	<b>-2.8074</b>
3	-0.9089	-0.7324	1.2578	3	-0.9089	-0.8456	.5798
4	-1.1479	-1.1545	.3314	4	-1.1479	-1.0593	.6888
5	-0.5668	-0.4834	.7867	5	-0.5668	-0.4108	.9794
6	-1.4246	-1.7858	-1.4627	6	-1.4246	-1.7790	-1.2207
7	1.0944	1.3059	1.4348	7	1.0944	1.2956	1.1742
8	-0.485	-0.3826	.8829	8	-0.4850	-0.5464	.0422
9	-1.5147	-1.9450	-1.8123	9	-1.5147	-1.9852	-1.7212
22	1.2935	1.2793	.2929	22	1.2935	1.1158	-.4591
23	-0.9772	-1.0786	-.1482	23	-0.9772	-0.9903	.2504
24	-0.6738	-0.9398	-.9810	24	-0.6738	-1.0217	-1.1927
25	-0.3487	-0.5102	-.4523	25	-0.3487	-0.5592	-.6005
26	1.0501	1.0252	.2388	26	1.0501	0.9625	-.0707
27	0.4286	0.0641	-1.4794	27	0.4286	0.1784	-.7716
28	0.2397	-0.0253	-.9759	28	0.2397	-0.0852	-1.0936
29	1.7013	1.5473	-.4144	29	1.7013	1.4687	-.6957
36	0.5357	0.4986	.1771	36	0.5357	0.4045	-.2586
37	1.2684	1.2882	.4650	37	1.2684	1.3388	.6104
38	0.7073	0.6352	.0000	38	0.7073	0.7039	.2923
39	-0.1523	-0.2091	.0774	39	-0.1523	-0.3529	-.5578
40	-0.8265	-0.8542	.2246	40	-0.8265	-0.8516	.1987
41	0.5323	0.5334	.3703	41	0.5323	0.5757	.4940
42	1.3246	1.2574	.0248	42	1.3246	1.2534	.0000
43	0.8915	0.9937	.8818	43	0.8915	1.0172	.8487
10A	-0.5665	-0.8589	-1.1146	10B	0.7440	0.8439	.7375
12A	0.0720	-0.4183	-2.1158	12B	0.1213	0.3197	1.1621
13A	-0.1510	0.0084	1.1712	13B	-0.0287	-0.4853	-1.6613
15A	0.2778	0.1070	-.4994	15B	0.5913	0.7286	.8987
16A	0.1214	0.3948	1.7480	16B	-0.3767	-0.7530	-1.3151
18A	-0.2321	-0.3501	-.2322	18B	-0.5113	-0.4858	.4168
19A	0.0309	-0.3387	-1.5052	19B	0.4651	0.2473	-.6319
21A	1.0627	0.8135	-.8960	21B	-0.0913	-0.1041	.2517

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.

**Form Statistics**

Form Statistics	Previous Year	2009	Previous Year	2009
	Form A	Form A	Form B	Form B
Mean	-.021	-.152	-.012	-.132
SD	.952	1.076	.951	1.068
Correlation Coefficient		.979		.981
SD Ratio		<b>113%</b>		<b>112%</b>
Mean Diff		-.132		-.121
Median Diff		-.072		-.071
IQR Diff		.267		.314

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 2 appearing on both forms was dropped from the linking pool.

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

Correlation Coefficient	<b>.979</b>	<b>.980</b>
SD Ratio	<b>109%</b>	<b>109%</b>

Rasch Item Difficulties of Common Items: Grade 6 Form A

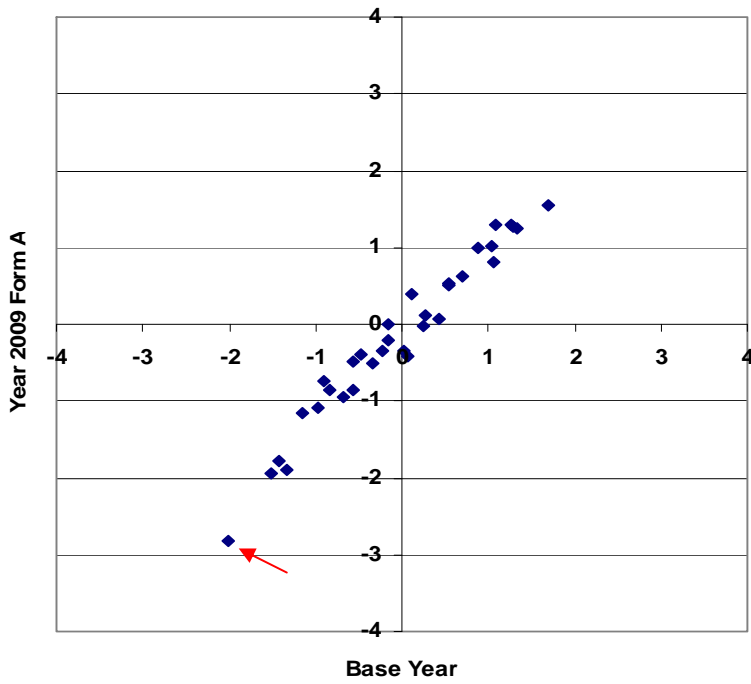


Figure 1.22 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 6 Form A

Rasch Item Difficulties of Common Items: Grade 6 Form B

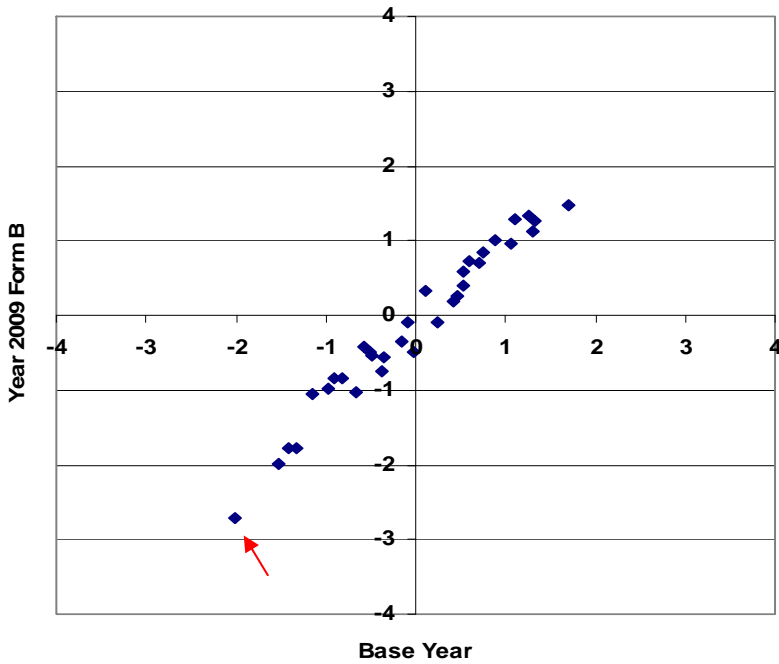


Figure 1.23 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 6 Form B

**Table 1.64 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 7**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
<b>1*</b>	<b>-1.9151</b>	<b>-2.3631</b>	<b>-2.1621</b>	<b>1*</b>	<b>-1.9151</b>	<b>-2.2999</b>	<b>-1.5900</b>
2	-1.5468	-1.8213	-1.4252	2	-1.5468	-1.7920	-1.1106
3	-0.5743	-0.3375	.7462	3	-0.5743	-0.3758	.4131
4	-1.1399	-1.2685	-.8056	4	-1.1399	-1.3454	-.9743
5	-1.8025	-2.1041	-1.5403	5	-1.8025	-2.1307	-1.3957
18	0.5791	0.2425	-1.6890	18	0.5791	0.4912	-.5704
19	-0.5545	-0.6753	-.7725	19	-0.5545	-0.8160	-1.1666
20	0.4841	0.6820	.5810	20	0.4841	0.6266	.2208
21	0.6747	0.3294	-1.7259	21	0.6747	0.3462	-1.3967
22	-0.6174	-0.7140	-.6697	22	-0.6174	-0.7571	-.7483
23	0.6158	0.6713	-.0238	23	0.6158	0.6546	-.1353
24	0.5929	0.4357	-.9271	24	0.5929	0.2707	-1.3750
<b>25*</b>	<b>-1.3746</b>	<b>-2.1937</b>	<b>-3.7381</b>	<b>25*</b>	<b>-1.3746</b>	<b>-2.1963</b>	<b>-3.0904</b>
32	-1.0643	-1.1025	-.4217	32	-1.0643	-1.2075	-.7603
33	0.5496	0.8815	1.1501	33	0.5496	0.8888	.8963
34	-0.0205	0.0574	.0713	34	-0.0205	0.0174	-.1384
35	0.0420	0.0721	-.1317	35	0.0420	0.0580	-.2136
36	0.9240	1.3309	1.4686	36	0.9240	1.2494	.8489
37	0.1653	0.3737	.6256	37	0.1653	0.3241	.2768
38	-0.8093	-0.5692	.7602	38	-0.8093	-0.6336	.3348
39	-0.4818	-0.4051	.0663	39	-0.4818	-0.4335	-.1027
40	-0.1892	0.1819	1.3165	40	-0.1892	0.1789	.9956
41	-0.5607	-0.4307	.2926	41	-0.5607	-0.3491	.4581
42	-0.3731	0.0930	1.7200	42	-0.3731	0.0002	1.0134
43	1.3477	1.4088	.0000	43	1.3477	1.5191	.3201
6A	-0.5886	-0.6997	-.7313	6B	0.4223	0.2404	-.8932
8A	-0.0097	0.0046	-.1988	8B	-0.5513	-0.4731	.0000
9A	-0.9725	-0.7789	.5627	9B	-0.4917	-0.2706	.4907
11A	0.4505	0.5366	.1062	11B	-0.2130	0.2323	1.2607
12A	0.5064	0.5693	.0076	12B	0.4182	0.5961	.3424
14A	0.7000	0.7982	.1576	14B	-0.4213	-0.2175	.4313
15A	0.1619	0.2163	-.0285	15B	0.5905	1.0017	1.1436
17A	1.0169	1.2143	.5788	17B	0.3397	0.7397	1.1051

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.

**Form Statistics**

Form Statistics	Previous Year	2009	Previous Year	2009
	Form A	Form A	Form B	Form B
Mean	-.175	-.163	-.211	-.178
SD	.849	.995	.808	.978
Correlation Coefficient		.970		.965
SD Ratio		<b>117%</b>		<b>121%</b>
Mean Diff		.013		.033
Median Diff		.061		.078
IQR Diff		.318		.394

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 1 and 25 appearing on both forms were dropped from the linking pool.

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

Correlation Coefficient	<b>.971</b>	<b>.963</b>
SD Ratio	<b>110%</b>	<b>114%</b>

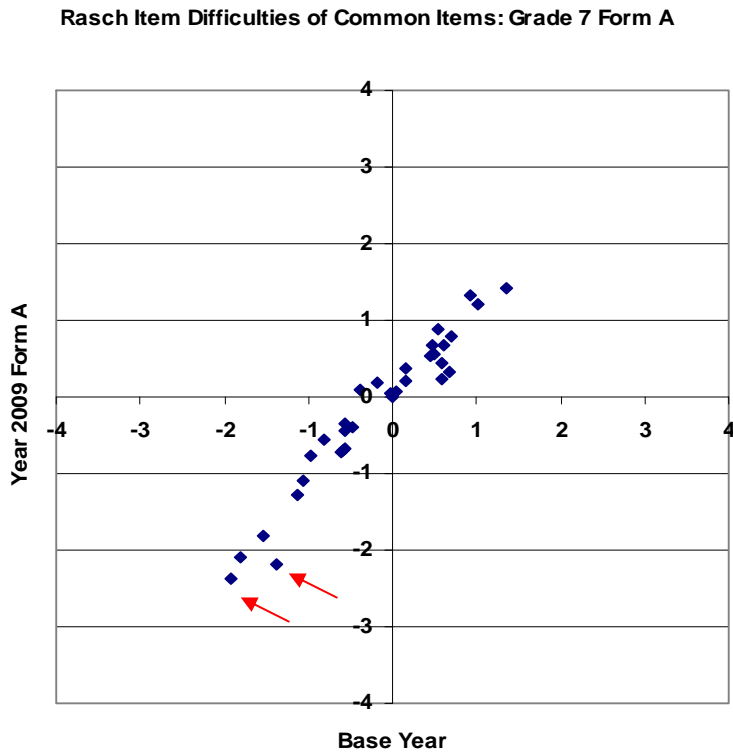


Figure 1.24 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 7 Form A

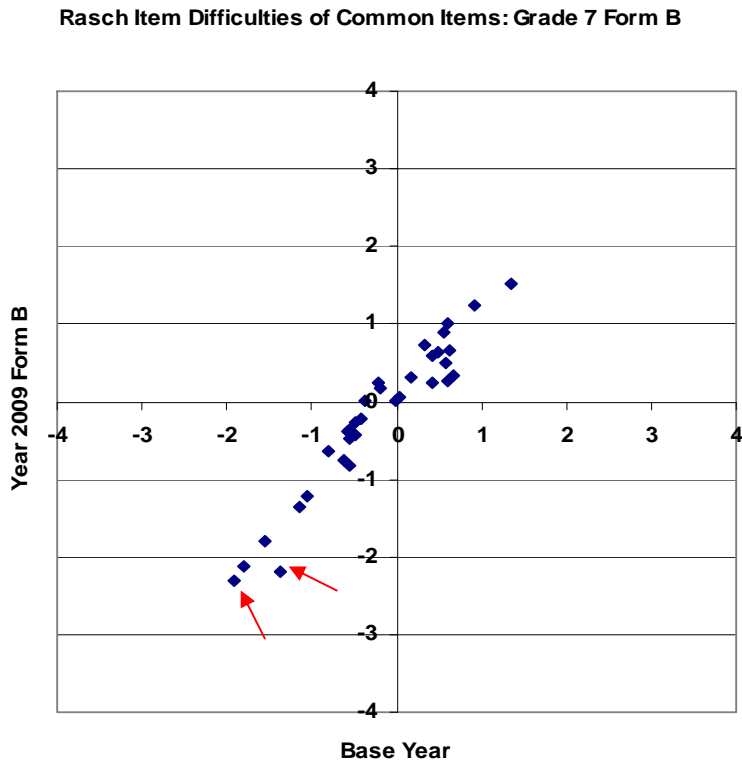


Figure 1.25 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 7 Form B



**Table 1.65 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2009: Grade 8**

Item Seq No.	Previous Year	Y2009 Form A	Robust Z	Item Seq No.	Previous Year	Y2009 Form B	Robust Z
<b>1*</b>	<b>-1.7533</b>	<b>-2.4341</b>	<b>-3.5131</b>	<b>1*</b>	<b>-1.7533</b>	<b>-2.6555</b>	<b>-2.7904</b>
2	-1.6274	-1.8488	-.1592	2	-1.6274	-2.0488	-.9164
3	-0.6076	-0.7950	.0891	3	-0.6076	-0.8102	-.0635
4	-0.6192	-0.7059	.8242	4	-0.6192	-0.7487	.2214
5	-1.3966	-1.6379	-.3044	5	-1.3966	-1.5313	.2011
18	0.4748	-0.1668	-3.2269	18	0.4748	-0.0326	-1.2516
19	0.5684	-0.0240	-2.8677	19	0.5684	-0.0100	-1.5283
20	-0.244	-0.7556	-2.2778	20	-0.2440	-0.6813	-.9783
21	0.5198	0.1242	-1.4309	21	0.5198	0.2367	-.3773
22	0.2093	0.0197	.0730	22	0.2093	0.0150	-.0312
<b>23*</b>	<b>-0.7152</b>	<b>-1.4016</b>	<b>-3.5540</b>	<b>23*</b>	<b>-0.7152</b>	<b>-1.3380</b>	<b>-1.7014</b>
24	0.7855	0.8456	1.8960	24	0.7855	1.0053	1.5829
25	0.3464	0.4471	2.1924	25	0.3464	0.5446	1.4987
32	0.2219	0.0272	.0358	32	0.2219	0.0706	.1364
33	1.2743	1.0233	-.3753	33	1.2743	1.1471	.2304
34	0.0242	-0.1124	.4599	34	0.0242	0.0629	.8770
35	-1.0659	-1.3733	-.7870	35	-1.0659	-1.2522	.0000
36	0.6577	0.4845	.1927	36	0.6577	0.5362	.2526
37	-0.6796	-0.7519	.9294	37	-0.6796	-0.7516	.4455
38	1.0443	1.2212	2.7487	38	1.0443	1.1950	1.3135
39	-1.1736	-1.2959	.5643	39	-1.1736	-1.2900	.2725
40	-0.3586	-0.5700	-.0861	40	-0.3586	-0.5578	-.0503
<b>41*</b>	<b>2.0260</b>	<b>2.4011</b>	<b>4.1957</b>	<b>41*</b>	<b>2.0260</b>	<b>2.3973</b>	<b>2.1734</b>
42	0.4998	0.7211	3.0728	42	0.4998	0.6987	1.5014
43	1.2325	0.9754	-.4198	43	1.2325	0.9835	-.2444
6A	0.1001	-0.2125	-.8250	6B	0.2076	-0.2111	-.9058
8A	0.2165	0.7875	5.6259	8B	0.5854	0.7023	1.1818
9A	-0.3128	-0.5071	.0387	9B	-0.0110	-0.3169	-.4670
11A	0.1675	-0.0250	.0518	<b>11B*</b>	<b>-0.6360</b>	<b>-1.2675</b>	<b>-1.7341</b>
12A	-1.2042	-1.4857	-.5979	<b>12B*</b>	<b>-1.3280</b>	<b>-1.9893</b>	<b>-1.8530</b>
14A	0.2223	-0.1120	-.9834	14B	-0.0090	-0.1986	-.0136
15A	-0.0622	-0.2810	-.1402	15B	0.4904	0.3680	.2491
17A	0.2274	0.0278	.0000	17B	0.3771	0.4277	.9234

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

*Note.* Item parameters of the current year were independently calibrated with a live, stratified random sample.

*Note.* Item parameters of the previous year were estimated with the 2007 field test equating sample and were on a common scale.

*Note.* Characters A and B were used to indicate that they were tested in sessions 2 (Literary Reading) and 3 (Informational Reading). Although these linking items appeared in the same position on each operational form they are unique items.

**Form Statistics**

Form Statistics	Previous Year	2009	Previous Year	2009
	Form A	Form A	Form B	Form B
Mean	-.030	-.224	-.021	-.221
SD	.875	1.011	.896	1.086
Correlation Coefficient		.967		.975
SD Ratio		<b>116%</b>		<b>121%</b>
Mean Diff		-.194		-.201
Median Diff		-.200		-.186
IQR Diff		.185		.347

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 1, 23, and 41 appearing on both forms, item number 11 and 12 appearing on Form B were dropped from the linking pool.

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

Correlation Coefficient	<b>.957</b>	<b>.969</b>
SD Ratio	<b>108%</b>	<b>110%</b>

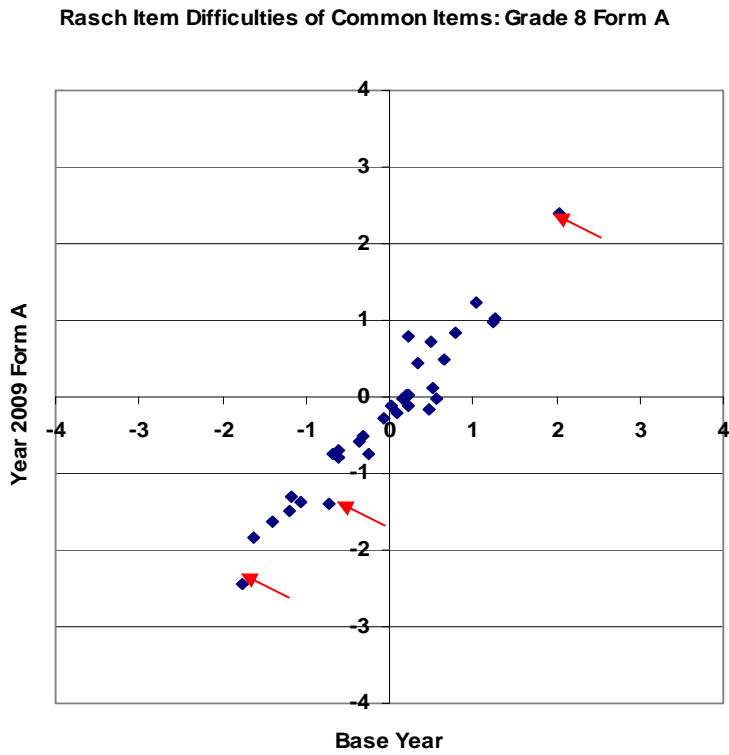


Figure 1.26 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 8 Form A

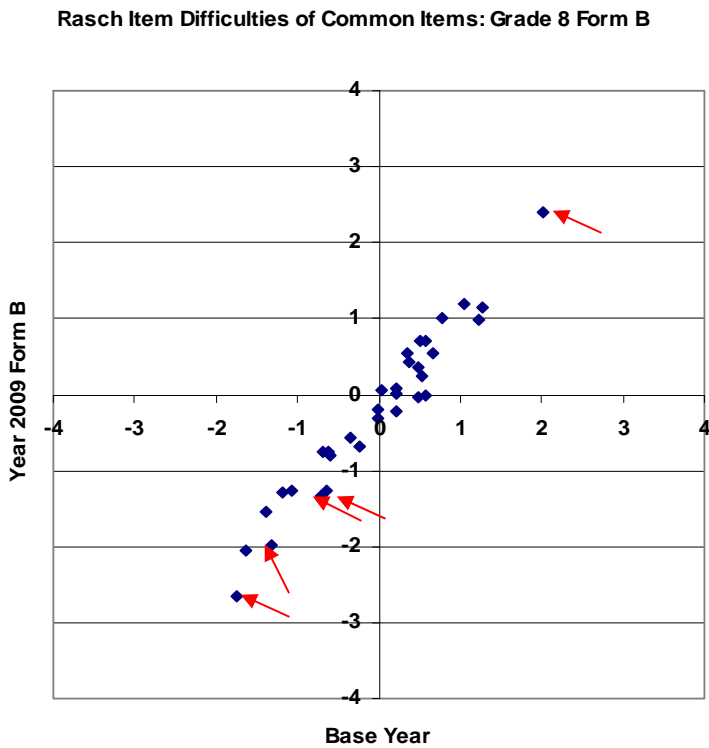


Figure 1.27 Item Difficulty Plot of Previous Year Form vs. Current Year (2009) Form: Grade 8 Form B

### Reporting Scale Scores

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

$$\text{ReportingAbilityScaleScore} = 32.8271 \cdot \text{theta} + 362.7449$$

$$\text{ReportingSE} = 32.8271 \cdot \text{SE}$$

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. It should be noted that these same slopes and intercepts have been used since the 2003 assessment (for grades 3, 5, and 8) or the 2004 assessment (for grades 4, 6, and 7).

**Table 1.66 The 2009 MSA-Reading Slope and Intercept: Grades 3 through 8**

Grade	Slope	Intercept
3	32.4123	384.8579
4	32.8271	362.7449
5	33.0171	380.0082
6	30.4732	373.0575
7	31.9262	377.0054
8	30.3891	376.8316

## 1.10 Score Interpretation

To help provide appropriate interpretation of the 2009 MSA-Reading 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 MSA-Reading produced scale scores that ranged between 240 and 650. These scale scores have the same meaning within the same grade, but those scores are not comparable across grade levels.

It should be noted that for scale scores, a higher score simply means a higher performance on reading 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-Reading 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.

As Table 2.1 shows a range of scale scores at each performance level; for example, grade 4 reading scale scores from 371 to 436 indicate the level of *Proficient*. Students in this level can read grade-appropriate text and demonstrate the ability to comprehend literature and informational passages. Further information about the 2009 MSA-Reading score interpretation can be obtained from the MSDE.

## 1.11 Test Validity

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

Messick (1989) defined validity as follows:

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

This definition implies that test validation is the process of accumulating evidence to support intended use of test scores. Consequently, test validation is a series of 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-Reading, content-related evidence, item development procedures, 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-Reading blueprints provide extensive evidence regarding the alignment between the content in the 2009 MSA-Reading and the VSC. It should be noted that the 2009 MSA-Reading 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. Detailed information about the item composition of the operational test forms can be obtained from section 1.4, *Test Form Design, Specifications, Item Type, and Item Roles* and section 1.5, *Operational Test Form Construction Using the Rasch Model*. In addition, the 2009 MSA-Reading blueprints are presented in Appendix D

### Item Development

Test development for MSA-Reading is ongoing and continuous. Content specialists, teachers from across Maryland, Pearson, and MSDE were greatly involved in developing and reviewing test items. Committees such as content review, bias review, and vision review reviewed all of the items, which were finally stored in the item bank. Specifically, an internal review by MSDE and Pearson staff for alignment and quality required a great deal of time and energy. More specific information on item (test) development and review can be obtained in section 1.3, *Development and Review of the 2009 MSA-Reading*.

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 reading item, looking for indications of bias that would impact the performance of an identifiable group of students. They discussed or rejected items on a basis of gender, ethnic, religious, or geographical bias.

#### **2) *DIF* Statistics**

For DIF analyses, subgroups were first categorized according to either reference or focal groups. For the 2009 MSA-Reading, 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 items, the standardized mean difference (SMD) and the standard deviation (SD), along with the Mantel statistic, were calculated for BCR items. All of the items were classified based on Educational Testing Service (ETS) guidelines. It should be noted that DIF analyses on the operational items indicated that all the items were satisfactory. All the DIF results were archived 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-Reading contains three reading processes: *General Reading*, *Literary Reading*, and *Informational Reading*. Tables 4.3 through 4.14 show correlations among the reading processes.

## 1.12 Unidimensionality Analyses

Measurement implies order and magnitude along a single dimension (Andrich, 1989). Consequently, in the case of scholastic achievement, a 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-Reading, 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 of the polytomously scored reading items. Principal component analysis was then applied to produce eigenvalues. The first and the second principal component eigenvalues were compared *without rotation*. Table 1.67 summarizes the results of the first and second principal component eigenvalues of the 2009 MSA-Reading.

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 a substantially larger eigenvalues across all grades: the size of the eigenvalue of the first component was over ten times that of the second eigenvalue for each form at each grade. As a result, we could conclude that the assumption of unidimensionality for the 2009 MSA-Reading was met.



**Table 1.67 The 2009 MSA-Reading Eigenvalues between the First and Second Components**

Grade	Form	Number of Items	First Eigenvalue	Second Eigenvalue
3	A	37	12.18	1.42
	B	37	12.17	1.43
4	A	37	12.01	1.36
	B	37	11.86	1.45
5	A	37	11.10	1.44
	B	37	10.14	1.52
6	A	37	10.19	1.49
	B	37	10.04	1.57
7	A	37	11.82	1.54
	B	37	11.54	1.43
8	A	37	10.25	1.24
	B	37	9.90	1.28

### 1.13 Field Test Analyses and Item Bank Construction

All field test items embedded in operational forms were subjected to rigorous 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 item bank. The following field test analyses were conducted:

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

#### Classical Item Analyses for *SR* and *BCR* items

Classical item analyses for *SR* and *BCR* 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.

*BCR* 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 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 groups 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* and *BCR* items that were flagged as showing *DIF* were subjected to further examination. For each of these items, for example, reading 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 on all the field test items were archived in the 2009 Maryland item bank. In addition, 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 2003 scale for grades 3, 5, and 8 and the 2004 scale for grades 4, 6, and 7), each field test item was freely calibrated after fixing the 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 reading test forms (i.e., 1, 3, 5, 7, and 9) 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). 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-Reading 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 Assessment, Inc. was to create a test deck for MSA programs. The test deck began when Quality Assurance entered mock data into the enrollment system, which was transferred to the materials requisition system; the order was packaged by our Distribution Center, and shipped to the Quality Assurance Department. We then reviewed the packing list against the data entered, the materials algorithms applied, the materials packaged against the packing list, and the actual packaging of the documents. These documents were then used to create a test deck of mock data, along with advance copies of documents that were received from the printer. Advance printer copies were inclusive of documents throughout the print run to assure we were randomly testing printed documents. The Maryland test deck was a comprehensive set of all documents that:

- 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
- Tested 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.