1. OVERVIEW OF THE 2010 *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 2010 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 2010 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 in 2010 a linking pool which included only operational selected-response items (i.e., multiple-choice items). These items appeared both in 2010 and in 2008. 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., 2008 assessment) that remained through the stability check to put the 2010 assessment on a common scale. Accordingly, all scale scores of the 2010 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) through 2008 scale which were on the same scale with 2003 or 2004.

1.1 Purposes/Uses of the 2010 MSA-Reading

By measuring students' achievement against the new academic standards, the 2010 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 2010 MSA-Reading Blueprint*.

1.3 Development and Review of the 2010 MSA-Reading Items and Test

As seen in Table 1.1, the development of the 2010 MSA-Reading test required the involvement of four groups in addition to MSDE and Pearson. It should be noted that the same procedures used for the 2009 administration were used for the 2010 administration. 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 2010 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 since this form is usually used for large print and braille forms.

Table 1.1 The 2010 MSA-Reading Responsibility for Test Development

Development of the 2010 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 and 5 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, and 6 (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 2010 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.

Table1.2 The 2010 MSA-Reading Test Form Design: Grades 3 through 8

	Operational Item Sets		Field Test Item Sets					
	A	B	1	2	3	4	5	6
Form 1	X		X					
Form 2		X		X				
Form 3	X				X			
Form 4		X				X		
Form 5	X						X	
Form 6		X						X

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

Item Types

The 2010 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 2010 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 2010 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 2010 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	**37**
B	16	0	16	8	2	10	9	2	11	**37**

Table 1.4 The 2010 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	**37**
B	15	0	15	9	2	11	9	2	11	**37**

Table 1.5 The 2010 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	**37**
B	15	0	15	9	2	11	9	2	11	**37**

Table 1.6 The 2010 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 2010 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 2010 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 2010 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 2010 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 2010 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 2010 MSA-Reading. The following table and figure show an example of the 2010 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 2010 Reading Operational Test Construction Using the Rasch Model: Grade 3 Form A

Item Type	P-value	A	D_{i1}	D_{i2}	D_{i3}
BCR	0.49	1.00	-2.9074	1.0900	6.4369
BCR	0.44	1.00	-0.9902	1.2648	6.6182
BCR	0.46	1.00	-1.2429	1.1387	5.7040
BCR	0.34	1.00	-0.2048	2.1489	6.5361
SR	0.93	1.00	-1.3708		
SR	0.98	1.00	-3.3089		
SR	0.74	1.00	-0.1994		
SR	0.87	1.00	-1.1969		
SR	0.89	1.00	-1.4394		
SR	0.94	1.00	-2.1425		
SR	0.68	1.00	0.1005		
SR	0.64	1.00	0.4026		
SR	0.77	1.00	-0.3764		
SR	0.79	1.00	-0.4038		
SR	0.84	1.00	-0.7744		
SR	0.87	1.00	-1.1230		
SR	0.81	1.00	-0.6469		
SR	0.76	1.00	-0.2203		
SR	0.54	1.00	1.0333		
SR	0.66	1.00	-0.4174		
SR	0.51	1.00	1.0945		
SR	0.45	1.00	-1.1558		
SR	0.92	1.00	0.2728		
SR	0.74	1.00	1.7277		
SR	0.78	1.00	-1.2720		
SR	0.66	1.00	0.3685		
SR	0.51	1.00	1.1796		
SR	0.45	1.00	1.4373		
SR	0.92	1.00	-1.8724		
SR	0.74	1.00	-0.1123		
SR	0.78	1.00	-0.3721		
SR	0.88	1.00	-1.1963		
SR	0.95	1.00	-2.2073		
SR	0.41	1.00	1.6804		
SR	0.64	1.00	0.4434		
SR	0.81	1.00	-0.5752		
SR	0.59	1.00	0.7773		
SR	0.57	1.00	0.6881		
SR	0.42	1.00	1.4344		

Table 1.9 (Continued)

Item Type	P-value	A	D_{i1}	D_{i2}	D_{i3}
SR	0.83	1.00	-0.8740		
SR	0.68	1.00	0.0445		
SR	0.59	1.00	0.6246		
SR	0.72	1.00	-0.1433		
SR	0.58	1.00	0.6215		
SR	0.68	1.00	0.0142		

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 ($D_{ij} = D_i + F_{ij}$)

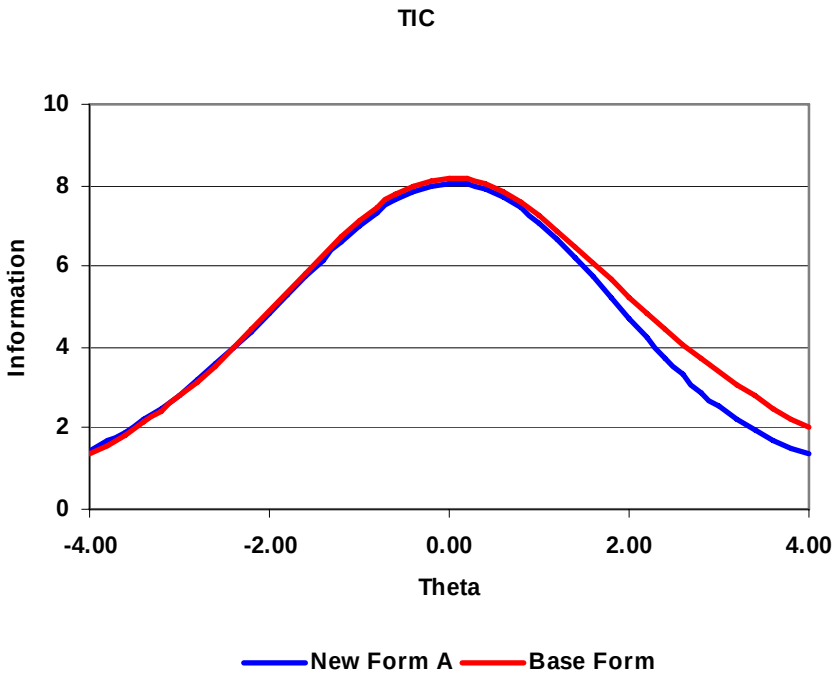

Figure 1.1 Test Information Curves of Base Form vs. Current Year's Reading Operational Test Form

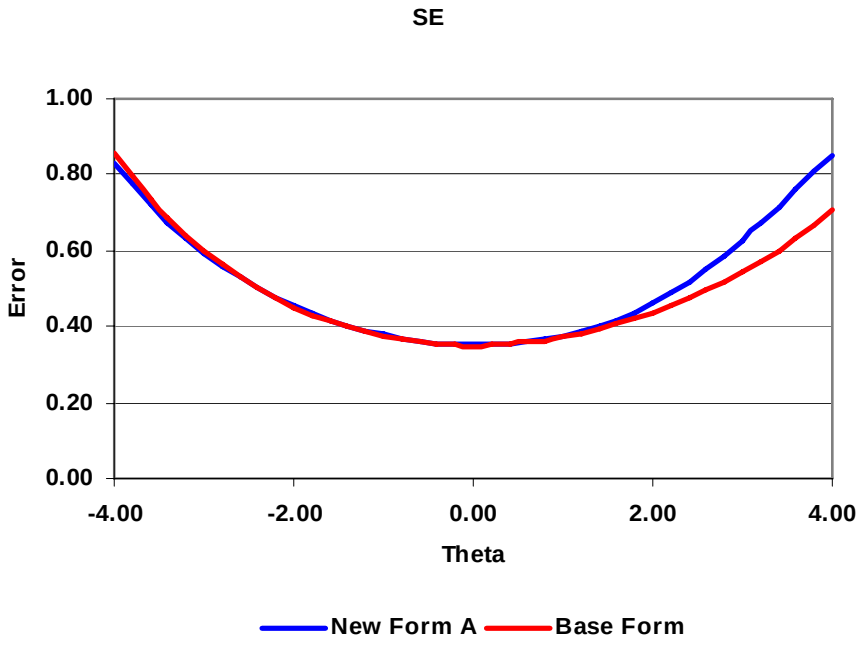

Figure 1.2 Standard Errors of Base Form vs. Current Year's Reading Operational Test Form

1.6 Test Administration of the 2010 MSA-Reading

The 2010 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 2010 test administration. Detailed information about the 2010 test administration can be obtained from the 2010 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 2010 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 8-17, 2010, with make-up testing held March 18-23, 2010). 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 2010 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) February 16-22, 2010
- Mathematics Primary Testing Window March 8 – March 17, 2010
- Make-up Testing Window March 18 – March 23, 2010

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 2010 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 2010 MSA-Reading, and all students in grades 6 through 8 had to participate in either the 2010 MSA-Reading or Mod-MSA-Reading. All students in grade 6 through 8 had to participate in the 2010 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, Administrating, and Evaluating the Use of Accommodations for Instruction and Assessment" (MAM). A copy of the most recent edition of this document is available electronically on the LAC and STC web pages at https://docushare.msde.state.md.us/docushare.

No accommodations could be made for students merely because they were members of an instructional group. Any accommodation had to be based on individual needs and not on a category of disability area, level of instruction, environment, or other group characteristics. Responsibility for confirming the need and appropriateness of an accommodation rested with the LAC and school-based staff involved with each student's instructional program. A master list of all students and their accommodations had to be maintained by the principal and submitted to the LAC, who provided a copy to MSDE upon request. Please refer to section 1 of the 2010 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 2010 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 2010 MSA-Reading. The accommodation was provided by a live reader or through technology. Appendix L of the 2010 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, 2010):

> 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 2010 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, 2010):

> Violation of security can result in prosecution and/or penalties as imposed by the Maryland State Board of Education and/or State Superintendent of Schools in accordance with the COMAR 13A.03.04 and 13A.12.05. (p. 11)

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

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

1.7 Scoring Procedures of the 2010 MSA-Reading

Students' responses to *SR* items were machine-scored, and their responses to *BCR* items were individually read and scored by Pearson. It should be noted that the same procedures used for scoring the 2009 BCR items were used for the 2010 BCR items.

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:

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

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

- For example:

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

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 2010 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 2010 MSA-Reading Form-to-Form Linking Common Items

Grade	Form	No. of Items	N	M	SD
3	A	25	26,273	18.78	4.20
	B	25	26,671	18.92	4.17
4	A	25	29,706	18.51	4.45
	B	25	29,549	18.61	4.40
5	A	25	29,203	18.80	3.78
	B	25	28,993	18.75	3.82
6	A	25	29,968	17.94	3.89
	B	25	29,222	18.13	3.80
7	A	25	29,782	19.32	4.30
	B	25	29,202	19.46	4.26
8	A	25	30,197	18.74	4.30
	B	25	29,668	18.81	4.23

Note. Form A designates the identical operational portion of Forms 1, 3, and 5. Form B designates the identical operational portion of Forms 2, 4, and 6.
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 2010 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 2010 assessment. The statistics of the previous year's assessment (i.e., 2008) were calculated based on a smaller field-test sample while the 2010 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 2010 p-values were almost the same or slightly increased compared to the 2008 p-values across all grades.

Table 1.11 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 3 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3399914	0.93	0.94	26	100000092693	0.78	0.90
3	3588051	0.98	0.98	27	100000092690	0.74	0.78
4	3588052	0.72	0.75	28	100000092694	0.66	0.71
6	3588010	0.89	0.90	29	100000154813	0.45	0.49
7	3588015	0.90	0.92	30	100000092724	0.95	0.95
8	3588020	0.94	0.96	31	100000092725	0.41	0.42
10	3588023	0.70	0.72	32	100000092715	0.64	0.68
11	3588035	0.67	0.70	33	100000092722	0.59	0.58
12	3588039	0.78	0.79	40	100000092739	0.53	0.56
14	**3571618**	**0.59**	**0.62**	41	100000092731	0.68	0.61
16	**3571625**	**0.72**	**0.74**	42	100000092728	0.79	0.74
17	**3571620**	**0.58**	**0.63**	43	100000092735	0.89	0.87
19	**3571623**	**0.68**	**0.74**	44	100000092754	0.79	0.76
20	**3571710**	**0.83**	**0.89**	45	100000092756	0.84	0.82
22	**3571716**	**0.42**	**0.44**	46	100000092753	0.54	0.51
23	**3571717**	**0.57**	**0.55**	47	100000092761	0.76	0.74
25	**3571713**	**0.68**	**0.71**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
3	Previous Year	33	0.72	0.15
	2010	33	0.73	0.16

Table 1.12 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 3 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3399914	0.93	0.95	26	100000092693	0.78	0.89
3	3588051	0.98	0.98	27	100000092690	0.74	0.79
4	3588052	0.72	0.75	28	100000092694	0.66	0.71
6	3588010	0.89	0.91	29	100000154813	0.45	0.51
7	3588015	0.90	0.92	30	100000092724	0.95	0.95
8	3588020	0.94	0.96	31	100000092725	0.41	0.42
10	3588023	0.70	0.74	32	100000092715	0.64	0.68
11	3588035	0.67	0.71	33	100000092722	0.59	0.58
12	3588039	0.78	0.80	40	100000092739	0.53	0.59
14	**3571489**	**0.68**	**0.69**	41	100000092731	0.68	0.61
16	**3571483**	**0.60**	**0.67**	42	100000092728	0.79	0.74
17	**3571485**	**0.52**	**0.59**	43	100000092735	0.89	0.87
19	**3571482**	**0.67**	**0.76**	44	100000092754	0.79	0.77
20	**3375183**	**0.78**	**0.74**	45	100000092756	0.84	0.83
22	**3375179**	**0.70**	**0.74**	46	100000092753	0.54	0.51
23	**3375182**	**0.43**	**0.46**	47	100000092761	0.76	0.75
25	**3375184**	**0.62**	**0.71**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
3	Previous Year	33	0.71	0.15
	2010	33	0.74	0.15

Table 1.13 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 4 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3588095	0.65	0.65	25	100000092940	0.60	0.62
2	3399931	0.80	0.80	26	100000092932	0.75	0.78
3	3588096	0.78	0.79	27	100000092942	0.67	0.70
4	3595149	0.94	0.95	28	100000092934	0.77	0.79
6	3399943	0.96	0.96	29	100000092922	0.58	0.58
7	3399944	0.97	0.97	30	100000092929	0.73	0.79
8	3588105	0.56	0.55	31	100000092927	0.75	0.79
10	3588111	0.94	0.94	32	100000092925	0.63	0.67
11	3588114	0.89	0.89	39	100000093046	0.68	0.65
13	**3374669**	**0.61**	**0.63**	40	100000093045	0.58	0.54
15	**3374670**	**0.57**	**0.65**	41	100000093041	0.65	0.54
16	**3374673**	**0.82**	**0.84**	42	100000093042	0.53	0.54
18	**3374672**	**0.78**	**0.79**	43	100000093008	0.77	0.74
19	**3557823**	**0.57**	**0.58**	44	100000093004	0.65	0.59
21	**3557825**	**0.77**	**0.81**	45	100000093014	0.92	0.89
22	**3557826**	**0.83**	**0.84**	46	100000093009	0.82	0.80
24	**3557822**	**0.53**	**0.54**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
4	Previous Year	33	0.73	0.13
	2010	33	0.73	0.14

Table 1.14 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 4 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3588095	0.65	0.66	25	100000092940	0.60	0.61
2	3399931	0.80	0.80	26	100000092932	0.75	0.78
3	3588096	0.78	0.80	27	100000092942	0.67	0.70
4	3595149	0.94	0.95	28	100000092934	0.77	0.79
6	3399943	0.96	0.96	29	100000092922	0.58	0.59
7	3399944	0.97	0.97	30	100000092929	0.73	0.79
8	3588105	0.56	0.55	31	100000092927	0.75	0.80
10	3588111	0.94	0.94	32	100000092925	0.63	0.66
11	3588114	0.89	0.89	39	100000093046	0.68	0.69
13	**3560628**	**0.67**	**0.70**	40	100000093045	0.58	0.54
15	**3560629**	**0.54**	**0.54**	41	100000093041	0.65	0.56
16	**3560635**	**0.73**	**0.73**	42	100000093042	0.53	0.55
18	**3560630**	**0.64**	**0.65**	43	100000093008	0.77	0.74
19	**3560658**	**0.75**	**0.78**	44	100000093004	0.65	0.59
21	**3560659**	**0.63**	**0.67**	45	100000093014	0.92	0.89
22	**3560656**	**0.66**	**0.70**	46	100000093009	0.82	0.81
24	**3560652**	**0.84**	**0.90**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
4	Previous Year	33	0.73	0.13
	2010	33	0.74	0.13

Table 1.15 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 5 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3400077	0.87	0.87	24	100000096744	0.65	0.67
2	3400080	0.79	0.81	25	100000096741	0.57	0.74
3	3400086	0.92	0.93	26	100000096750	0.84	0.86
4	3400088	0.94	0.95	27	100000096749	0.86	0.90
6	3451551	0.93	0.94	28	100000096809	0.49	0.51
7	3451440	0.96	0.96	29	100000096807	0.75	0.76
8	3451552	0.91	0.92	30	100000096810	0.66	0.67
9	3588453	0.49	0.50	31	100000096806	0.55	0.63
10	3588454	0.93	0.94	38	100000096609	0.53	0.55
12	**3374725**	**0.76**	**0.80**	39	100000096600	0.88	0.87
14	**3374727**	**0.62**	**0.64**	40	100000096610	0.60	0.59
15	**3374729**	**0.73**	**0.76**	41	100000096603	0.54	0.54
17	**3374728**	**0.73**	**0.74**	42	100000096642	0.87	0.85
18	**3557878**	**0.42**	**0.38**	43	100000096635	0.73	0.68
20	**3557876**	**0.73**	**0.69**	44	100000096636	0.75	0.71
21	**3557880**	**0.65**	**0.71**	45	100000096641	0.48	0.45
23	**3557881**	**0.67**	**0.72**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
5	Previous Year	33	0.72	0.16
	2010	33	0.73	0.16

Table 1.16 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 5 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3400077	0.87	0.88	24	100000096744	0.65	0.66
2	3400080	0.79	0.81	25	100000096750	0.84	0.87
3	3400086	0.92	0.93	26	100000096741	0.57	0.68
4	3400088	0.94	0.95	27	100000096749	0.86	0.88
6	3451551	0.93	0.94	28	100000096809	0.49	0.50
7	3451440	0.96	0.96	29	100000096807	0.75	0.76
8	3451552	0.91	0.92	30	100000096810	0.66	0.67
9	3588453	0.49	0.50	31	100000096806	0.55	0.63
10	3588454	0.93	0.94	38	100000096609	0.53	0.56
12	**3571702**	**0.66**	**0.76**	39	100000096600	0.88	0.87
14	**3571696**	**0.69**	**0.75**	40	100000096610	0.60	0.59
15	**3571703**	**0.70**	**0.76**	41	100000096603	0.54	0.55
17	**3571705**	**0.52**	**0.59**	42	100000096642	0.87	0.86
18	**3557865**	**0.63**	**0.72**	43	100000096635	0.73	0.68
20	**3557861**	**0.75**	**0.79**	44	100000096636	0.75	0.71
21	**3557868**	**0.50**	**0.50**	45	100000096641	0.48	0.45
23	**3557864**	**0.58**	**0.58**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
5	Previous Year	33	0.71	0.16
	2010	33	0.73	0.15

Table 1.17 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 6 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3400102	0.93	0.93	24	100000090802	0.76	0.73
2	3400104	0.97	0.97	25	100000090803	0.40	0.40
3	3595144	0.85	0.84	26	100000090796	0.71	0.74
4	3400107	0.89	0.88	27	100000090807	0.79	0.82
5	3588412	0.81	0.81	28	3588289	0.56	0.56
7	3451451	0.93	0.93	29	3588290	0.73	0.70
8	3451452	0.51	0.48	30	3588292	0.77	0.74
9	3451553	0.80	0.79	31	3588288	0.45	0.43
10	3451453	0.93	0.94	38	100000155728	0.81	0.82
12	**3568763**	**0.76**	**0.80**	39	100000155731	0.74	0.75
14	**3568767**	**0.63**	**0.68**	40	100000155727	0.48	0.44
15	**3568762**	**0.57**	**0.63**	41	100000155734	0.90	0.86
17	**3568764**	**0.69**	**0.75**	42	100000090726	0.42	0.41
18	**3562051**	**0.56**	**0.61**	43	100000090724	0.72	0.67
20	**3562050**	**0.63**	**0.72**	44	100000090722	0.80	0.77
21	**3562047**	**0.69**	**0.74**	45	100000090727	0.48	0.50
23	**3562052**	**0.76**	**0.82**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
6	Previous Year	33	0.71	0.16
	2010	33	0.72	0.16

Table 1.18 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 6 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3400102	0.93	0.93	24	100000090802	0.76	0.76
2	3400104	0.97	0.97	25	100000090803	0.40	0.39
3	3595144	0.85	0.85	26	100000090796	0.71	0.74
4	3400107	0.89	0.89	27	100000090807	0.79	0.83
5	3588412	0.81	0.82	28	3588289	0.56	0.57
7	3451451	0.93	0.93	29	3588290	0.73	0.71
8	3451452	0.51	0.50	30	3588292	0.77	0.76
9	3451553	0.80	0.81	31	3588288	0.45	0.44
10	3451453	0.93	0.94	38	100000155728	0.81	0.86
12	**3489699**	**0.62**	**0.68**	39	100000155731	0.74	0.74
14	**3489697**	**0.66**	**0.72**	40	100000155727	0.48	0.45
15	**3489701**	**0.77**	**0.85**	41	100000155734	0.90	0.87
17	**3489700**	**0.62**	**0.63**	42	100000090726	0.42	0.40
18	**3562086**	**0.77**	**0.83**	43	100000090724	0.72	0.71
20	**3562081**	**0.58**	**0.58**	44	100000090722	0.80	0.77
21	**3562084**	**0.67**	**0.75**	45	100000090727	0.48	0.48
23	**3562083**	**0.73**	**0.76**				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
6	Previous Year	33	0.71	0.16
	2010	33	0.73	0.17

Table 1.19 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 7 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3400135	0.96	0.96	24	100000090960	0.85	0.87
2	3400120	0.94	0.93	25	100000090951	0.89	0.90
3	3400132	0.80	0.82	26	100000090956	0.56	0.59
5	3451470	0.91	0.92	27	100000090954	0.70	0.68
6	3451556	0.95	0.95	28	100000092195	0.79	0.79
8	**3560779**	**0.63**	**0.65**	29	100000092193	0.89	0.90
10	**3560780**	**0.56**	**0.60**	30	100000090940	0.62	0.59
11	**3560777**	**0.57**	**0.57**	31	100000092198	0.63	0.64
13	**3560778**	**0.61**	**0.68**	38	100000091010	0.73	0.68
14	**3577903**	**0.70**	**0.76**	39	100000091002	0.79	0.73
16	**3577911**	**0.75**	**0.77**	40	100000091004	0.84	0.77
17	**3577907**	**0.57**	**0.58**	41	100000091007	0.85	0.78
19	**3577913**	**0.83**	**0.82**	42	100000085672	0.77	0.67
20	100000090921	0.82	0.85	43	100000149414	0.73	0.69
21	100000090915	0.65	0.66	44	100000085675	0.84	0.76
22	100000090922	0.76	0.77	45	100000085673	0.70	0.66
23	100000090920	0.74	0.76				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
7	Previous Year	33	0.75	0.12
	2010	33	0.75	0.12

Table 1.20 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 7 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3400135	0.96	0.96	24	100000090960	0.85	0.87
2	3400120	0.94	0.93	25	100000090951	0.89	0.90
3	3400132	0.80	0.83	26	100000090956	0.56	0.61
5	3451470	0.91	0.92	27	100000090954	0.70	0.69
6	3451556	0.95	0.95	28	100000092195	0.79	0.79
8	**3560794**	**0.68**	**0.76**	29	100000092193	0.89	0.91
10	**3560789**	**0.77**	**0.85**	30	100000090940	0.62	0.59
11	**3560787**	**0.63**	**0.70**	31	100000092198	0.63	0.64
13	**3560790**	**0.75**	**0.81**	38	100000091010	0.73	0.69
14	**3562063**	**0.64**	**0.70**	39	100000091002	0.79	0.75
16	**3562059**	**0.60**	**0.65**	40	100000091004	0.84	0.79
17	**3562060**	**0.58**	**0.60**	41	100000091007	0.85	0.78
19	**3562058**	**0.55**	**0.63**	42	100000085672	0.77	0.67
20	100000090921	0.82	0.86	43	100000149414	0.73	0.70
21	100000090915	0.65	0.66	44	100000085675	0.84	0.77
22	100000090922	0.76	0.78	45	100000085673	0.70	0.64
23	100000090920	0.74	0.77				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
7	Previous Year	33	0.75	0.12
	2010	33	0.76	0.11

Table 1.21 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 8 Form A

Item Seq. No.	Item CID	Previous Year	Y10 FA	Item Seq. No.	Item CID	Previous Year	Y10 FA
1	3400154	0.95	0.96	24	100000091099	0.87	0.88
2	3400158	0.93	0.94	25	100000091103	0.74	0.76
4	3451476	0.85	0.85	26	100000091104	0.83	0.84
5	3451557	0.84	0.83	27	100000091106	0.59	0.60
6	3451558	0.91	0.92	28	100000091150	0.65	0.67
8	**3570924**	**0.81**	**0.87**	29	100000157294	0.51	0.52
10	**3570922**	**0.49**	**0.57**	30	100000091148	0.78	0.79
11	**3570918**	**0.65**	**0.64**	31	100000157293	0.60	0.62
13	**3570919**	**0.63**	**0.72**	38	100000091190	0.79	0.75
14	**3571801**	**0.54**	**0.55**	39	100000091197	0.64	0.57
16	**3571807**	**0.57**	**0.67**	40	100000091196	0.79	0.71
17	**3571806**	**0.72**	**0.78**	41	100000091195	0.55	0.51
19	**3571805**	**0.88**	**0.91**	42	100000091227	0.80	0.65
20	100000091136	0.71	0.73	43	100000091229	0.77	0.72
21	100000091138	0.83	0.88	44	100000091232	0.90	0.77
22	100000091139	0.60	0.65	45	100000091234	0.89	0.79
23	100000091140	0.80	0.85				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
8	Previous Year	33	0.74	0.13
	2010	33	0.74	0.13

Table 1.22 P-Value Comparisons of Core Linking Items for Previous Year vs. Year 2010: Grade 8 Form B

Item Seq. No.	Item CID	Previous Year	Y10 FB	Item Seq. No.	Item CID	Previous Year	Y10 FB
1	3400154	0.95	0.96	24	100000091099	0.87	0.89
2	3400158	0.93	0.94	25	100000091103	0.74	0.78
4	3451476	0.85	0.86	26	100000091104	0.83	0.85
5	3451557	0.84	0.84	27	100000091106	0.59	0.62
6	3451558	0.91	0.92	28	100000091150	0.65	0.66
8	**3560850**	**0.57**	**0.68**	29	100000157294	0.51	0.51
10	**3560848**	**0.50**	**0.63**	30	100000091148	0.78	0.81
11	**3560852**	**0.62**	**0.69**	31	100000157293	0.60	0.61
13	**3560851**	**0.74**	**0.84**	38	100000091190	0.79	0.75
14	**3570910**	**0.71**	**0.82**	39	100000091197	0.64	0.56
16	**3570905**	**0.55**	**0.70**	40	100000091196	0.79	0.71
17	**3570907**	**0.82**	**0.86**	41	100000091195	0.55	0.51
19	**3570912**	**0.74**	**0.82**	42	100000091227	0.80	0.65
20	100000091136	0.71	0.74	43	100000091229	0.77	0.70
21	100000091138	0.83	0.89	44	100000091232	0.90	0.77
22	100000091139	0.60	0.64	45	100000091234	0.89	0.79
23	100000091140	0.80	0.85				

Note. Bold-faced items are sessions 2 (Literacy) and 3 (informational) items.

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

Grade	Year	No. of Items	M	SD
8	Previous Year	33	0.74	0.13
	2010	33	0.75	0.12

Validation Check with the 2010 Operational BCR Items

To collect information about how much the same BCR items that appeared in both 2008 and 2010 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 2008 assessment and appeared as operational test items on the 2010 assessment, as shown in Table 1.23. It should be noted that these items were administered in sessions 2 (Literary) and 3 (Informational) during the first testing day of the 2010 assessment but appeared in the last session on the 2008 assessment. The item numbers in Tables 1.24 through 1.59 were assigned based on the 2010 assessment. Detailed information about the specific test design and construction of Year 2010 can be obtained from section 1.4, *Test Structure of the 2010 MSA-Reading*.

While the 2008 p-value was calculated with a field test sample, the 2010 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 2010 p-values were almost the same or somewhat increased compared to those of the 2008 assessment.

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 2010 item difficulties stayed almost the same or became a little lower compared to that of the 2008 assessment across all the grades. It should be noted that all of the Rasch item and step difficulty parameters were on a common scale (i.e., linked to the 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.

Table 1.23 Form Identification for Items Appearing in both 2008 and 2010: Grades 3 through 8

Grade	Year 2008	Year 2010
3	Form 1, 2	Form A
	Form 1, 8	Form B
4	Form 7, 8	Form A
	Form 1, 5	Form B
5	Form 1, 5	Form A
	Form 2, 9	Form B
6	Form 1, 6	Form A
	Form 7, 8	Form B
7	Form 1, 5	Form A
	Form 3, 4	Form B
8	Form 4, 5	Form A
	Form 2, 9	Form B

Table 1.24 P-Value Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 3 Form A

Item Number	CID	Item Type	Previous Year	Year 10
15	3571629	BCR	0.46	0.49
18	3571626	BCR	0.34	0.37
21	3571719	BCR	0.49	0.53
24	3571720	BCR	0.44	0.42

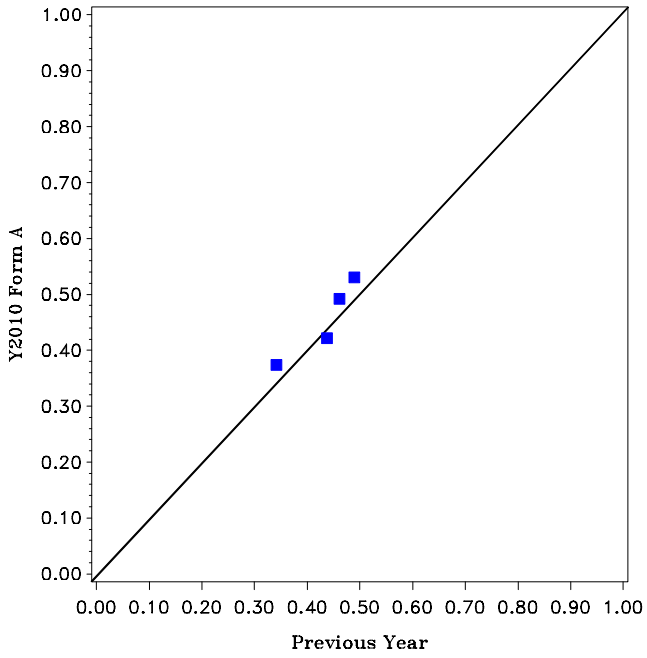

Table 1.25 Score-Point Distribution Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 3 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	15	3571629	BCR	2,598	1.39	0.67	8.12	43.76	45.77	1.12	1.23
2008	18	3571626	BCR	2,598	1.03	0.71	20.36	50.92	25.44	0.35	2.93
2008	21	3571719	BCR	2,558	1.47	0.55	1.95	48.87	48.36	0.55	0.27
2008	24	3571720	BCR	2,558	1.31	0.68	10.75	45.39	42.38	0.43	1.06
2010	15	3571629	BCR	26,273	1.48	0.63	4.29	45.08	47.79	2.45	0.30
2010	18	3571626	BCR	26,273	1.12	0.72	17.64	48.96	30.86	0.54	1.36
2010	21	3571719	BCR	26,273	1.60	0.58	2.26	36.91	58.55	1.85	0.30
2010	24	3571720	BCR	26,273	1.27	0.69	12.08	47.73	38.51	0.71	0.65

Table 1.26 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 3 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	15	3571629	BCR	1.8666	-3.1095	-0.7279	3.8374
2008	18	3571626	BCR	2.8267	-3.0315	-0.6778	3.7094
2008	21	3571719	BCR	1.5398	-4.4472	-0.4498	4.8971
2008	24	3571720	BCR	2.2976	-3.2878	-1.0328	4.3206
2010	15	3571629	BCR	1.3455	-3.2269	-0.5172	3.7441
2010	18	3571626	BCR	2.5666	-2.8070	-0.6925	3.4995
2010	21	3571719	BCR	1.0356	-3.7359	-0.3326	4.0685
2010	24	3571720	BCR	2.4702	-3.2455	-0.5997	3.8452

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

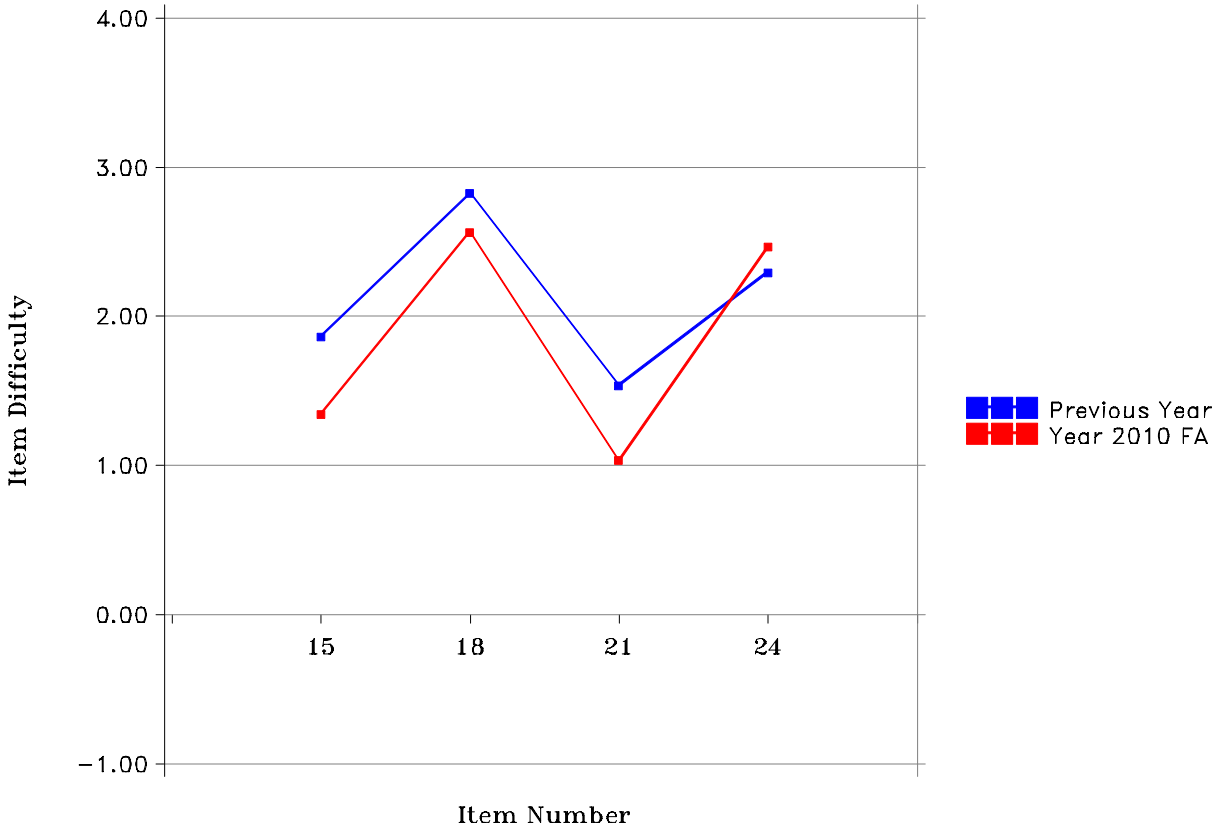

Figure 1.3 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 3 Form A

Table 1.27 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 3 Form B

Item Number	CID	Item Type	Previous Year	Year 10
15	3571492	BCR	0.52	0.55
18	3571493	BCR	0.41	0.51
21	3375185	BCR	0.39	0.40
24	3375187	BCR	0.46	0.48

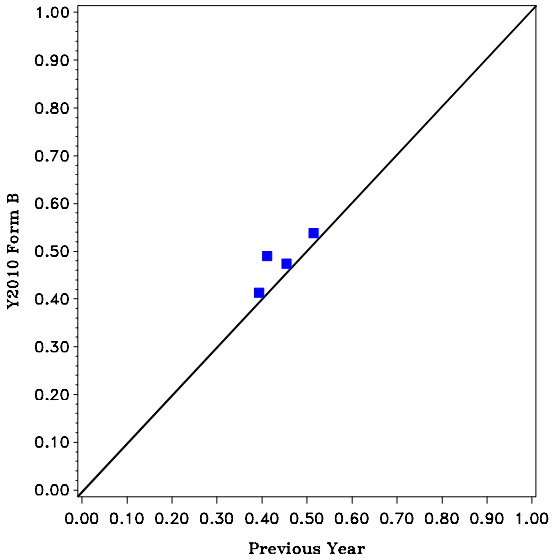

Table 1.28 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 3 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	15	3571492	BCR	2,242	1.55	0.65	6.56	32.47	58.61	1.65	0.71
2008	18	3571493	BCR	2,242	1.24	0.67	10.93	52.23	34.26	0.98	1.61
2008	21	3375185	BCR	2,197	1.18	0.68	12.92	54.71	30.00	1.23	1.14
2008	24	3375187	BCR	2,450	1.37	0.62	4.93	49.47	43.06	0.41	2.12
2010	15	3571492	BCR	14,636	1.64	0.60	2.75	33.08	60.13	3.46	0.44
2010	18	3571493	BCR	14,528	1.53	0.64	5.14	38.05	52.60	2.51	1.18
2010	21	3375185	BCR	14,586	1.20	0.77	19.62	41.17	36.35	1.72	0.78
2010	24	3375187	BCR	14,537	1.43	0.63	5.97	44.32	47.03	1.16	1.12

Table 1.29 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 3 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	15	3571492	BCR	1.5856	-2.8992	-1.0981	3.9974
2008	18	3571493	BCR	2.1795	-3.2246	-0.3997	3.6244
2008	21	3375185	BCR	2.1690	-3.0299	-0.2015	3.2314
2008	24	3375187	BCR	2.0716	-3.9839	-0.6980	4.6819
2010	15	3571492	BCR	1.1436	-3.4613	-0.6072	4.0685
2010	18	3571493	BCR	1.5943	-3.2135	-0.5673	3.7808
2010	21	3375185	BCR	2.4265	-2.4750	-0.7587	3.2337
2010	24	3375187	BCR	1.9958	-3.5656	-0.7362	4.3018

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

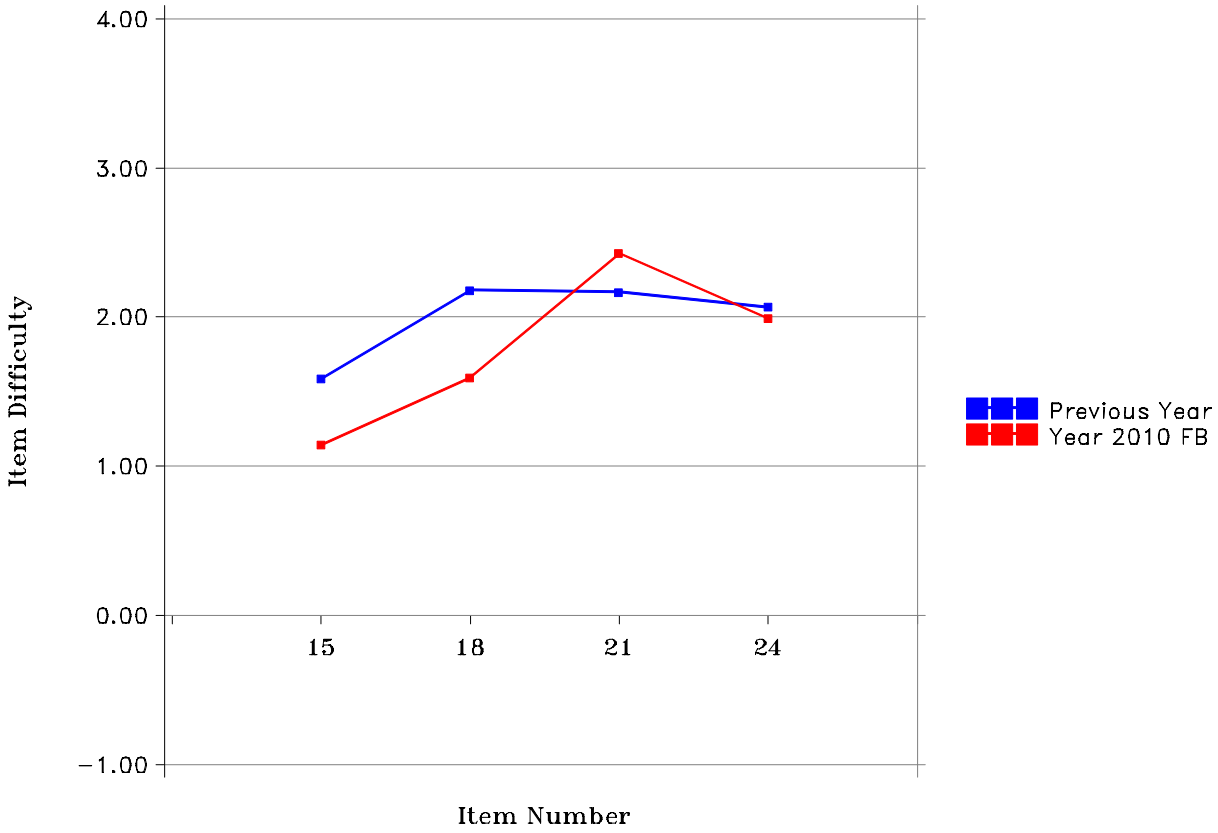

Figure 1.4 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 3 Form B

Table 1.30 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 4 Form A

Item Number	CID	Item Type	Previous Year	Year 10
14	3374845	BCR	0.48	0.53
17	3374844	BCR	0.47	0.51
20	3557831	BCR	0.38	0.39
23	3557830	BCR	0.42	0.42

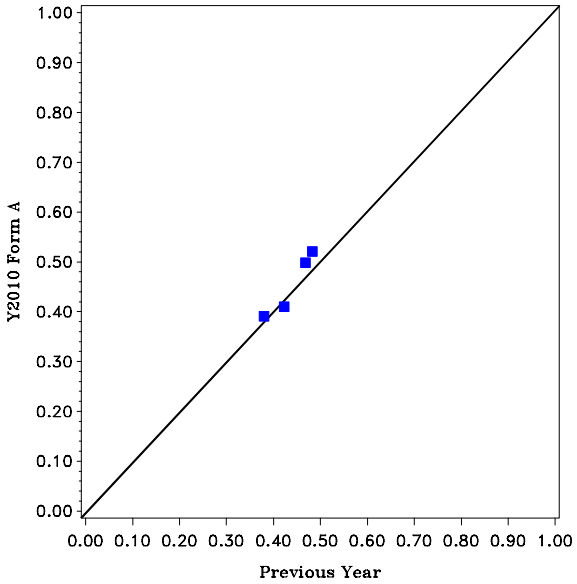

Table 1.31 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 4 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	14	3374845	BCR	2,164	1.45	0.61	5.03	43.39	50.28	0.42	0.88
2008	17	3374844	BCR	2,381	1.41	0.55	2.10	54.01	42.80	0.38	0.71
2008	20	3557831	BCR	2,582	1.14	0.58	9.45	66.27	23.12	0.54	0.62
2008	23	3557830	BCR	2,582	1.27	0.62	6.86	58.64	32.30	1.32	0.89
2010	14	3374845	BCR	14,636	1.58	0.57	2.64	36.42	59.33	0.98	0.23
2010	17	3374844	BCR	14,604	1.54	0.56	1.54	43.94	52.37	1.51	0.45
2010	20	3557831	BCR	14,626	1.17	0.55	7.53	67.01	24.76	0.16	0.30
2010	23	3557830	BCR	14,602	1.26	0.56	4.96	63.61	29.90	0.80	0.46

Table 1.32 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 4 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	14	3374845	BCR	2.3989	-3.8405	-0.9598	4.8003
2008	17	3374844	BCR	2.1120	-4.7233	-0.3093	5.0326
2008	20	3557831	BCR	2.8182	-3.7791	0.1156	3.6634
2008	23	3557830	BCR	2.2921	-3.5614	0.0616	3.4998
2010	14	3374845	BCR	2.1548	-3.8957	-1.0699	4.9656
2010	17	3374844	BCR	1.8215	-4.3666	-0.2854	4.6520
2010	20	3557831	BCR	3.4417	-4.6455	-0.3241	4.9696
2010	23	3557830	BCR	2.6758	-4.4942	0.0717	4.4225

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

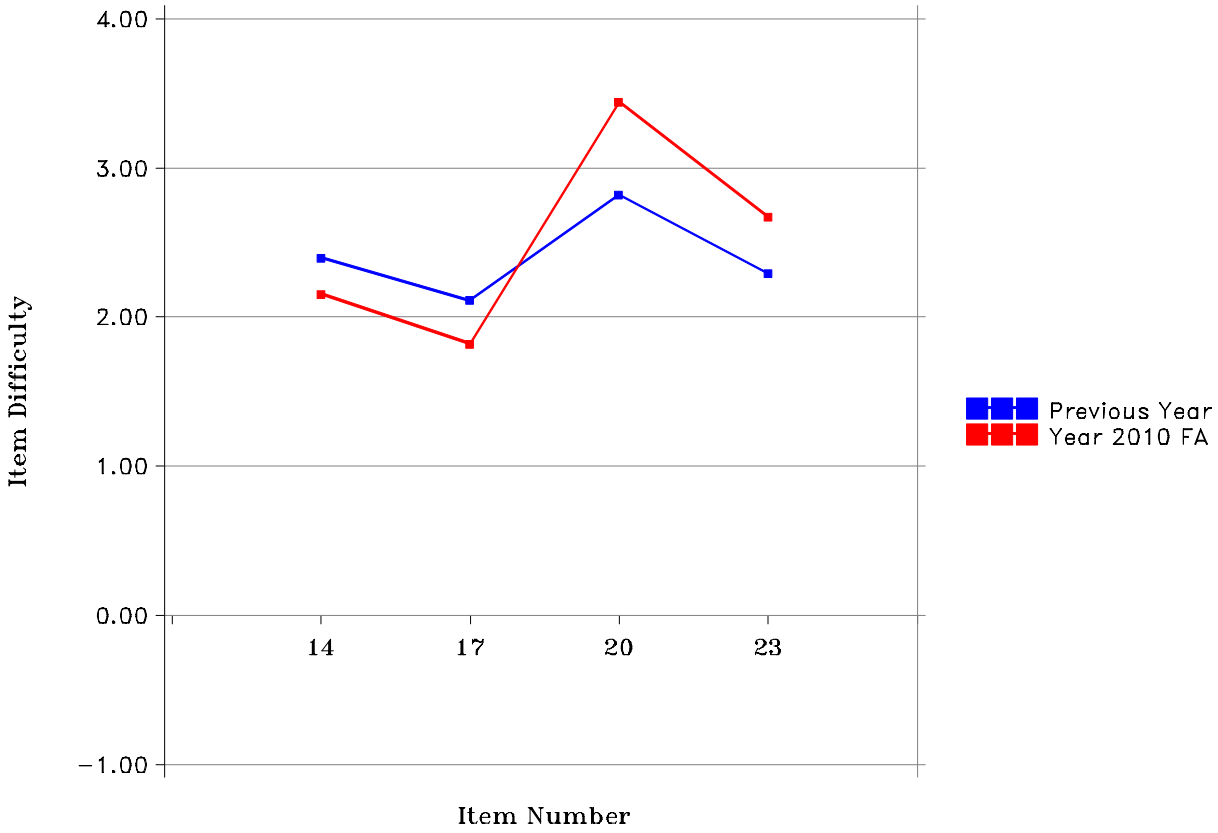

Figure 1.5 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 4 Form A

Table 1.33 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 4 Form B

Item Number	CID	Item Type	Previous Year	Year 10
14	3560636	BCR	0.36	0.40
17	3560637	BCR	0.42	0.42
20	3560662	BCR	0.47	0.49
23	3560660	BCR	0.42	0.40

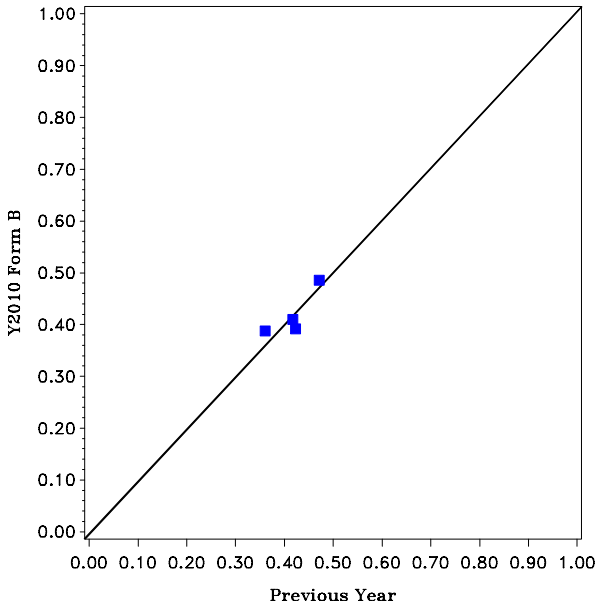

Table 1.34 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 4 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	14	3560636	BCR	2,264	1.08	0.50	6.98	76.33	15.15	0.57	0.97
2008	17	3560637	BCR	2,264	1.26	0.52	2.65	67.27	28.75	0.27	1.06
2008	20	3560662	BCR	2,572	1.42	0.52	0.82	56.57	42.03	0.43	0.16
2008	23	3560660	BCR	2,572	1.27	0.52	2.57	68.00	28.54	0.70	0.19
2010	14	3560636	BCR	14,571	1.20	0.58	7.78	63.75	26.98	0.58	0.55
2010	17	3560637	BCR	14,572	1.25	0.52	3.19	68.09	27.38	0.59	0.55
2010	20	3560662	BCR	14,627	1.48	0.57	1.89	49.60	46.25	2.03	0.17
2010	23	3560660	BCR	14,598	1.19	0.48	2.29	77.18	18.52	1.43	0.37

Table 1.35 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 4 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	14	3560636	BCR	2.8057	-4.1303	0.7762	3.3541
2008	17	3560637	BCR	2.6159	-4.8353	0.0388	4.7965
2008	20	3560662	BCR	1.9064	-5.3535	0.0310	5.3225
2008	23	3560660	BCR	2.2437	-4.6336	0.3992	4.2344
2010	14	3560636	BCR	3.0104	-4.2257	-0.1342	4.3599
2010	17	3560637	BCR	2.5387	-4.7019	0.2433	4.4586
2010	20	3560662	BCR	1.7103	-4.1660	-0.0112	4.1772
2010	23	3560660	BCR	2.1940	-4.8789	1.2263	3.6526

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

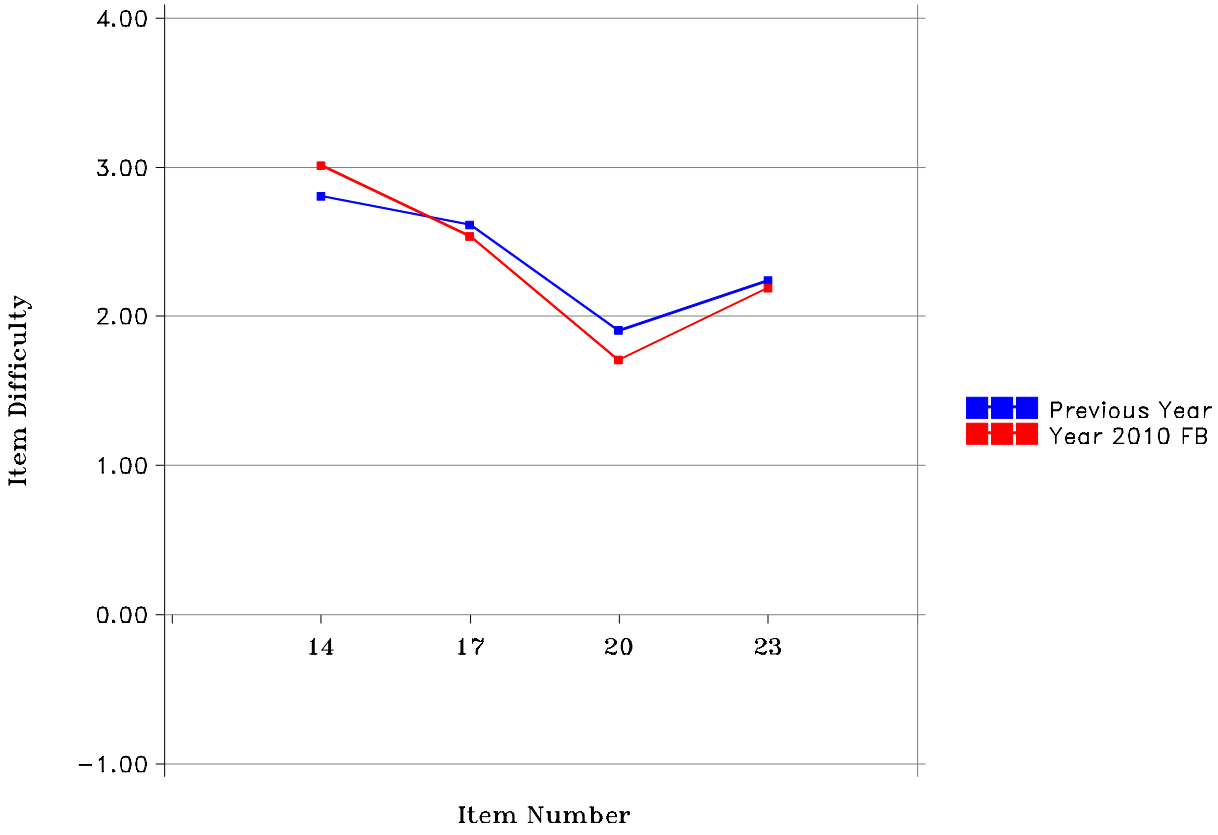

Figure 1.6 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 4 Form B

Table 1.36 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 5 Form A

Item Number	CID	Item Type	Previous Year	Year 10
13	3374874	BCR	0.40	0.44
16	3374873	BCR	0.24	0.28
19	3557885	BCR	0.55	0.54
22	3557883	BCR	0.45	0.43

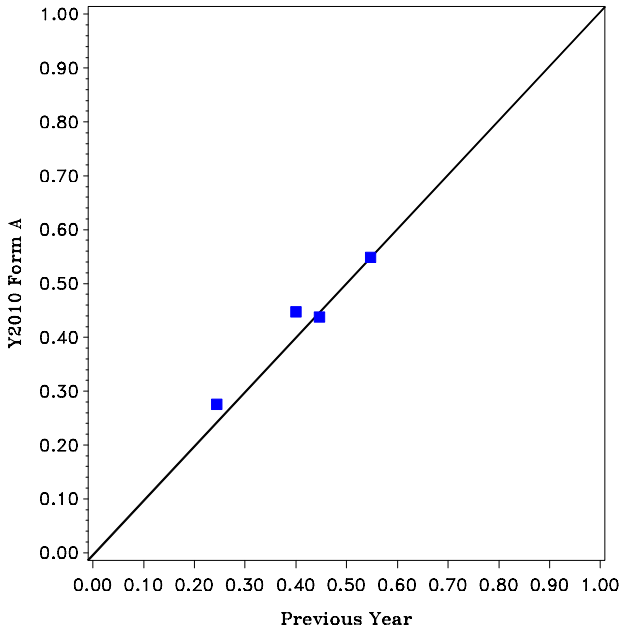

Table 1.37 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 5 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	13	3374874	BCR	2,209	1.20	0.83	21.46	39.11	34.45	4.12	0.86
2008	16	3374873	BCR	2,496	0.73	0.68	2.28	37.46	47.40	12.54	0.32
2008	19	3557885	BCR	2,607	1.65	0.61	1.65	36.63	56.23	5.14	0.35
2008	22	3557883	BCR	2,607	1.34	0.65	7.33	51.94	37.97	2.15	0.61
2010	13	3374874	BCR	14,833	1.31	0.73	13.13	45.51	38.21	2.80	0.26
2010	16	3374873	BCR	14,799	0.85	0.74	34.74	45.21	18.49	0.91	0.49
2010	19	3557885	BCR	14,858	1.62	0.54	0.85	37.95	58.89	2.20	0.09
2010	22	3557883	BCR	14,845	1.30	0.58	3.44	65.36	28.41	2.60	0.18

Table 1.38 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 5 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	13	3374874	BCR	1.7248	-1.8343	-0.4197	2.2540
2008	16	3374873	BCR	3.1844	-2.5412	-0.2708	2.8119
2008	19	3557885	BCR	0.9340	-3.5315	-0.0646	3.5961
2008	22	3557883	BCR	1.9597	-3.1704	-0.0994	3.2698
2010	13	3374874	BCR	2.1195	-2.5645	-0.3842	2.9487
2010	16	3374873	BCR	3.2685	-2.3037	-0.4068	2.7106
2010	19	3557885	BCR	1.0431	-4.4367	-0.2013	4.6380
2010	22	3557883	BCR	1.7525	-3.9147	0.6829	3.2318

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

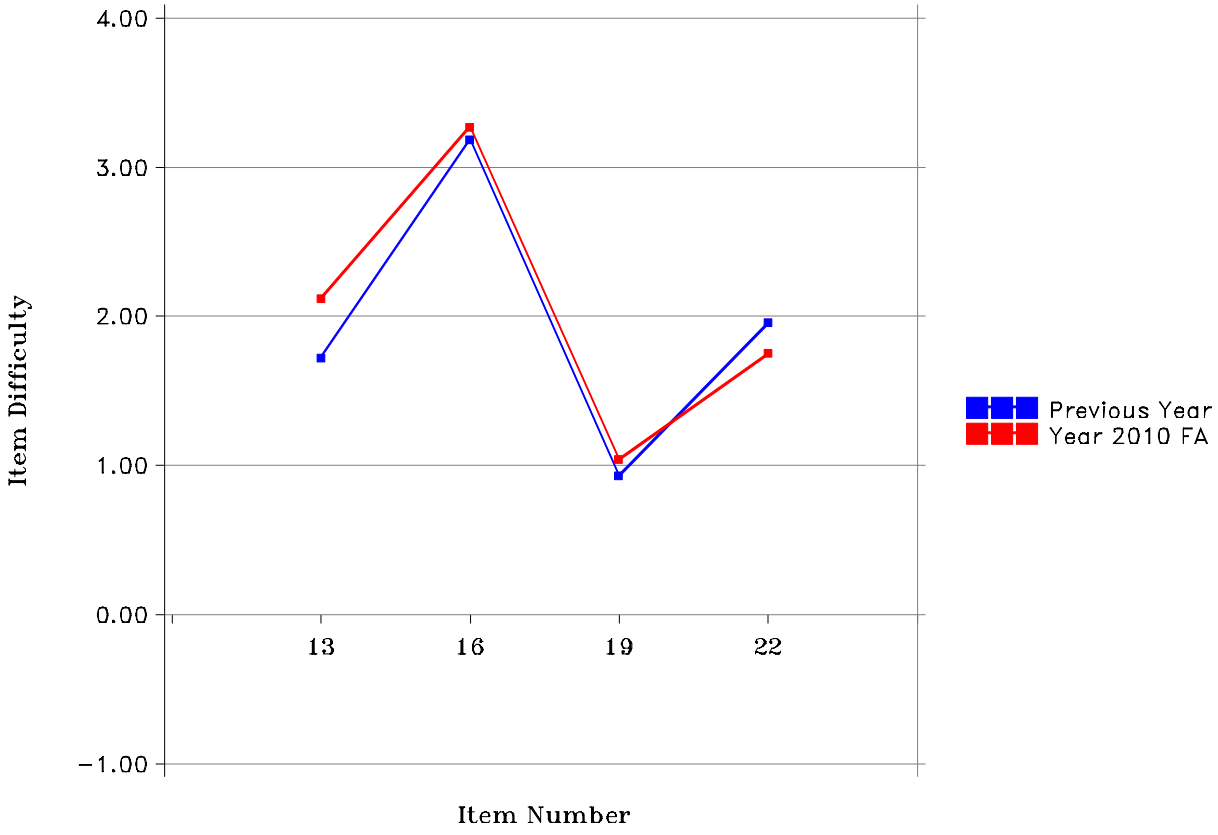

Figure 1.7 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 5 Form A

Table 1.39 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 5 Form B

Item Number	CID	Item Type	Previous Year	Year 10
13	3571706	BCR	0.38	0.34
16	3571707	BCR	0.41	0.45
19	3557869	BCR	0.51	0.56
22	3557870	BCR	0.50	0.53

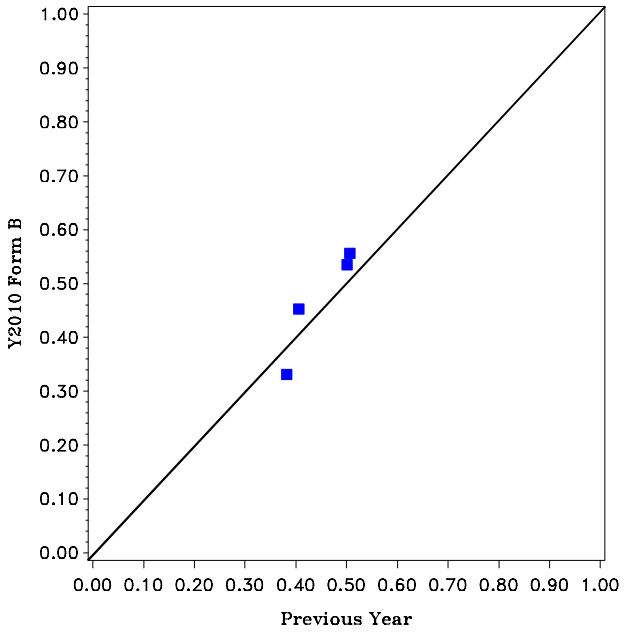

Table 1.40 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 5 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	13	3571706	BCR	2,620	1.15	0.70	16.83	50.34	31.18	0.73	0.92
2008	16	3571707	BCR	2,620	1.22	0.60	5.88	62.02	28.47	1.03	2.60
2008	19	3557869	BCR	2,308	1.52	0.64	5.68	37.39	54.77	1.73	0.43
2008	22	3557870	BCR	2,308	1.51	0.71	6.50	40.38	46.62	5.63	0.87
2010	13	3571706	BCR	14,769	1.02	0.77	28.58	41.13	29.76	0.28	0.11
2010	16	3571707	BCR	14,726	1.35	0.59	3.95	58.27	35.50	1.87	0.40
2010	19	3557869	BCR	14,754	1.69	0.58	3.38	26.72	67.16	2.48	0.21
2010	22	3557870	BCR	14,745	1.60	0.61	2.75	37.86	55.06	4.02	0.27

Table 1.41 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 5 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	13	3571706	BCR	2.7259	-2.9646	-0.6169	3.5815
2008	16	3571707	BCR	2.1992	-3.7998	0.1575	3.6423
2008	19	3557869	BCR	1.8029	-3.1093	-0.8163	3.9256
2008	22	3557870	BCR	1.4530	-2.6741	-0.2141	2.8882
2010	13	3571706	BCR	3.4794	-2.7377	-1.2920	4.0297
2010	16	3571707	BCR	1.9766	-3.8669	0.1829	3.6840
2010	19	3557869	BCR	1.4886	-2.9558	-1.1427	4.0985
2010	22	3557870	BCR	1.3596	-3.3973	-0.3021	3.6994

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

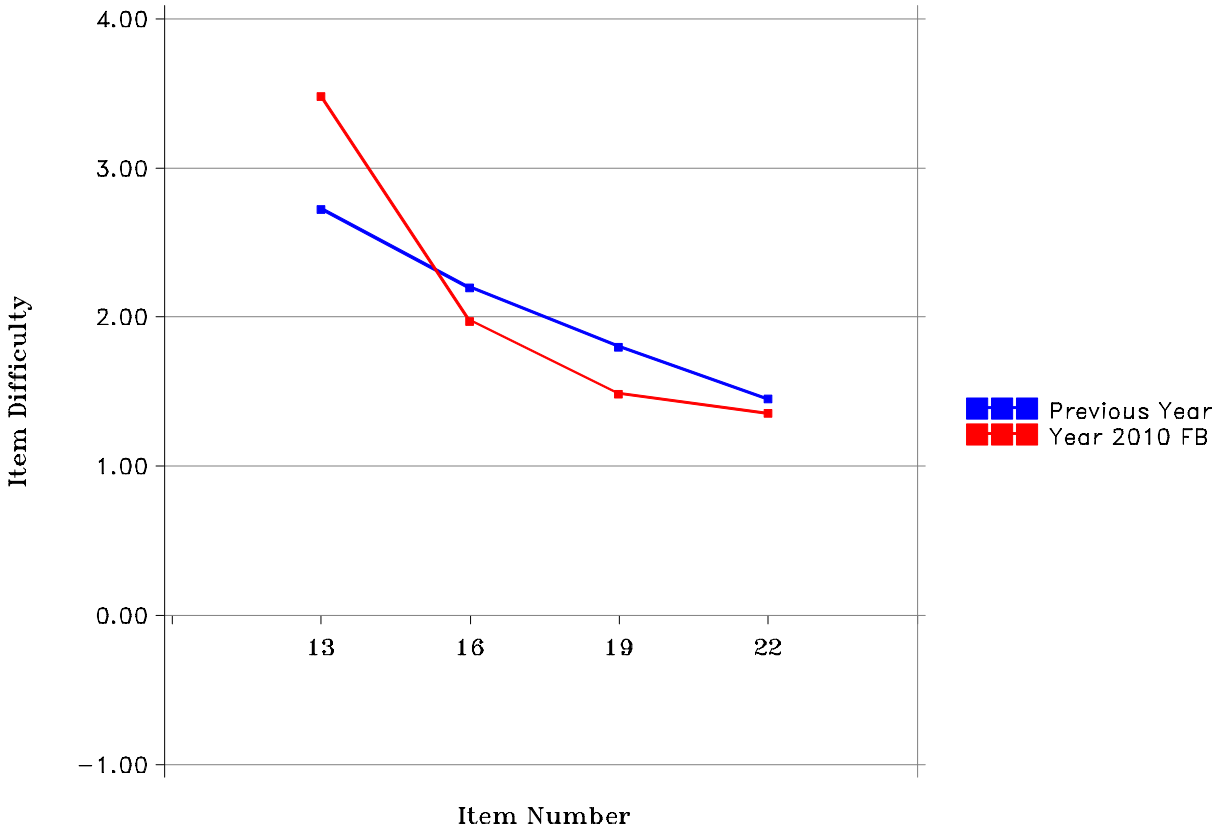

Figure 1.8 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 5 Form B

Table 1.42 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 6 Form A

Item Number	CID	Item Type	Previous Year	Year 10
13	3568771	BCR	0.39	0.43
16	3568770	BCR	0.38	0.43
19	3562055	BCR	0.62	0.66
22	3562053	BCR	0.42	0.49

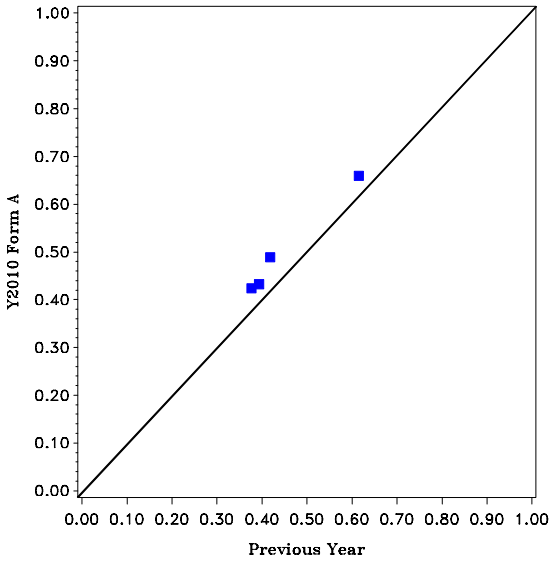

Table 1.43 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 6 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	13	3568771	BCR	2,595	1.18	0.57	6.94	64.70	26.59	0.15	1.62
2008	16	3568770	BCR	2,595	1.13	0.68	14.03	57.34	25.47	1.66	1.50
2008	19	3562055	BCR	2,635	1.85	0.78	2.88	26.64	49.34	19.85	1.29
2008	22	3562053	BCR	2,635	1.26	0.70	12.64	47.06	38.03	0.95	1.33
2010	13	3568771	BCR	29,809	1.30	0.54	3.48	62.60	33.09	0.38	0.31
2010	16	3568770	BCR	29,751	1.28	0.59	4.84	63.70	28.71	2.12	0.50
2010	19	3562055	BCR	29,766	1.99	0.69	1.12	20.38	56.27	21.72	0.45
2010	22	3562053	BCR	29,740	1.48	0.63	5.91	41.01	51.43	1.03	0.54

Table 1.44 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 6 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	13	3568771	BCR	2.8479	-4.3605	-0.3985	4.7590
2008	16	3568770	BCR	2.3006	-2.9291	0.0955	2.8336
2008	19	3562055	BCR	0.2956	-2.3185	-0.0575	2.3760
2008	22	3562053	BCR	2.3452	-3.1262	-0.7478	3.8739
2010	13	3568771	BCR	2.3640	-4.4960	-0.1774	4.6734
2010	16	3568770	BCR	1.8859	-3.6833	0.4747	3.2086
2010	19	3562055	BCR	0.0533	-2.6229	-0.0592	2.6821
2010	22	3562053	BCR	2.0424	-3.3351	-0.8743	4.2094

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

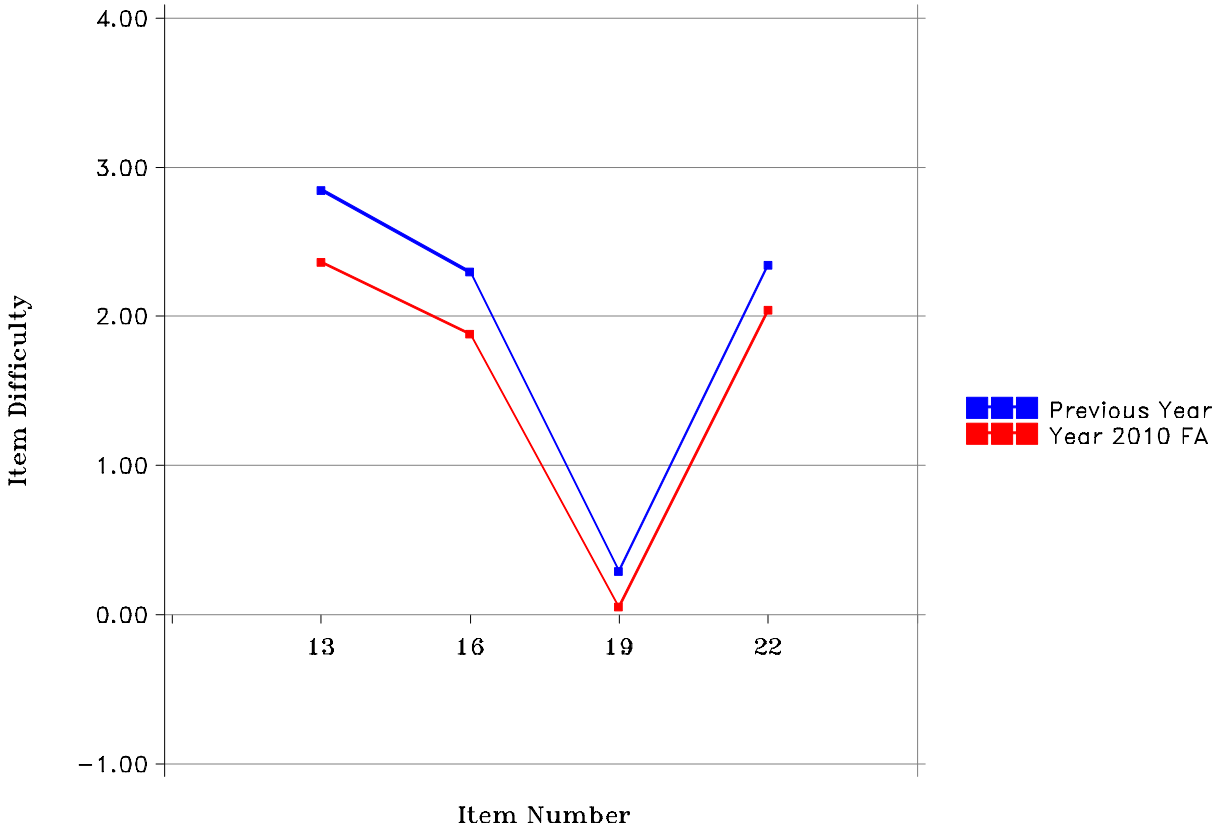

Figure 1.9 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 6 Form A

Table 1.45 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 6 Form B

Item Number	CID	Item Type	Previous Year	Year 10
13	3489708	BCR	0.37	0.44
16	3489707	BCR	0.39	0.43
19	3562091	BCR	0.36	0.37
22	3562092	BCR	0.45	0.47

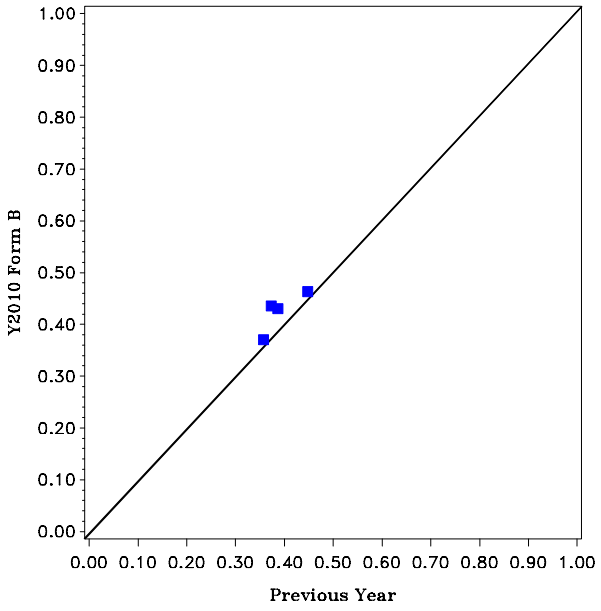

Table 1.46 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 6 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	13	3489708	BCR	2,265	1.12	0.67	15.59	55.06	28.04	0.40	0.93
2008	16	3489707	BCR	2,265	1.16	0.56	7.46	66.09	24.81	0.22	1.41
2008	19	3562091	BCR	2,281	1.07	0.74	21.04	47.74	28.06	1.18	1.97
2008	22	3562092	BCR	2,281	1.35	0.60	2.98	58.40	35.16	1.97	1.49
2010	13	3489708	BCR	29,090	1.31	0.62	7.43	54.41	36.69	1.13	0.21
2010	16	3489707	BCR	29,016	1.30	0.56	4.77	60.37	33.74	0.51	0.47
2010	19	3562091	BCR	28,954	1.12	0.80	24.52	39.48	33.72	1.51	0.68
2010	22	3562092	BCR	29,032	1.40	0.58	3.76	53.62	40.92	1.24	0.41

Table 1.47 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 6 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	13	3489708	BCR	2.9319	-3.3755	-0.6175	3.9930
2008	16	3489707	BCR	2.8810	-4.2912	-0.2743	4.5655
2008	19	3562091	BCR	2.6662	-2.6346	-0.4469	3.0815
2008	22	3562092	BCR	1.6381	-4.0208	0.3466	3.6742
2010	13	3489708	BCR	2.2188	-3.3459	-0.2957	3.6416
2010	16	3489707	BCR	2.3693	-4.0390	-0.2393	4.2783
2010	19	3562091	BCR	2.6226	-2.0771	-0.8279	2.9050
2010	22	3562092	BCR	1.9025	-3.7652	-0.1568	3.9220

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

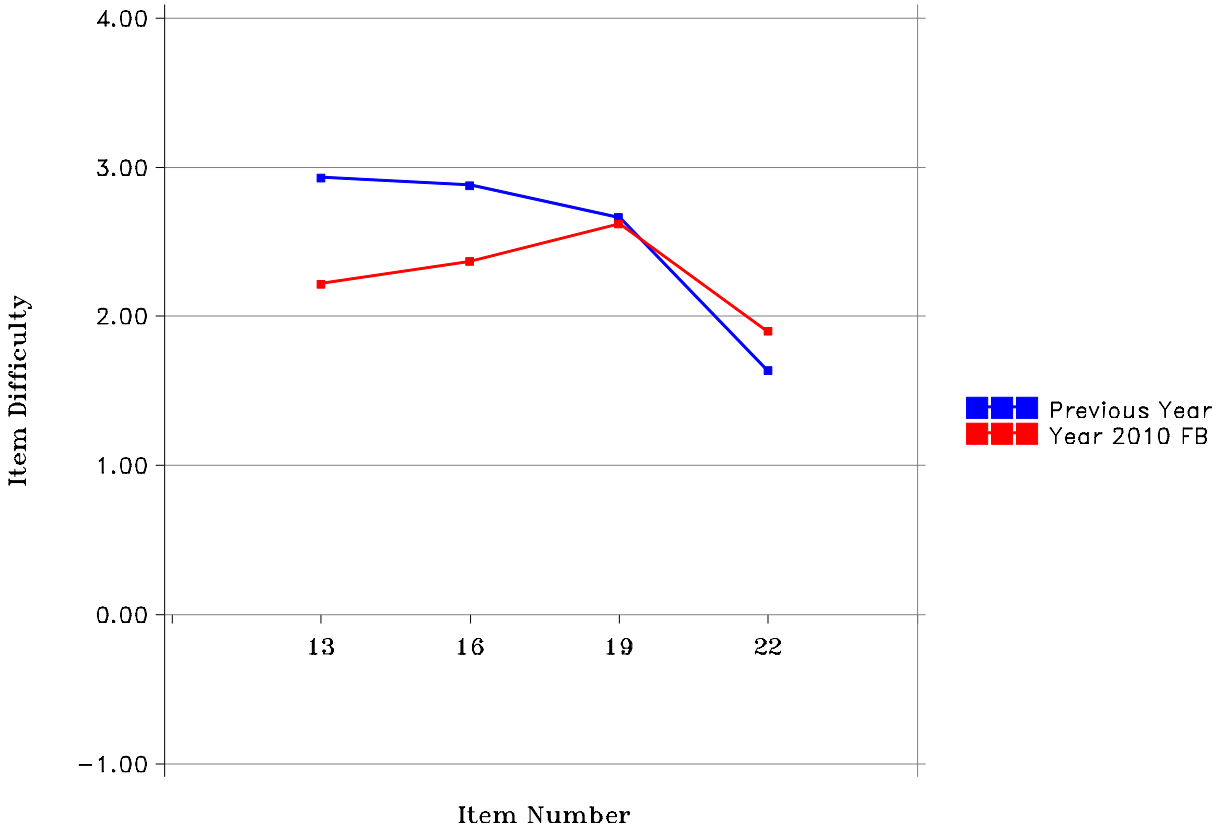

Figure 1.10 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 6 Form B

Table 1.48 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 7 Form A

Item Number	CID	Item Type	Previous Year	Year 10
9	3560786	BCR	0.48	0.47
12	3560784	BCR	0.49	0.52
15	3577916	BCR	0.35	0.38
18	3577919	BCR	0.49	0.48

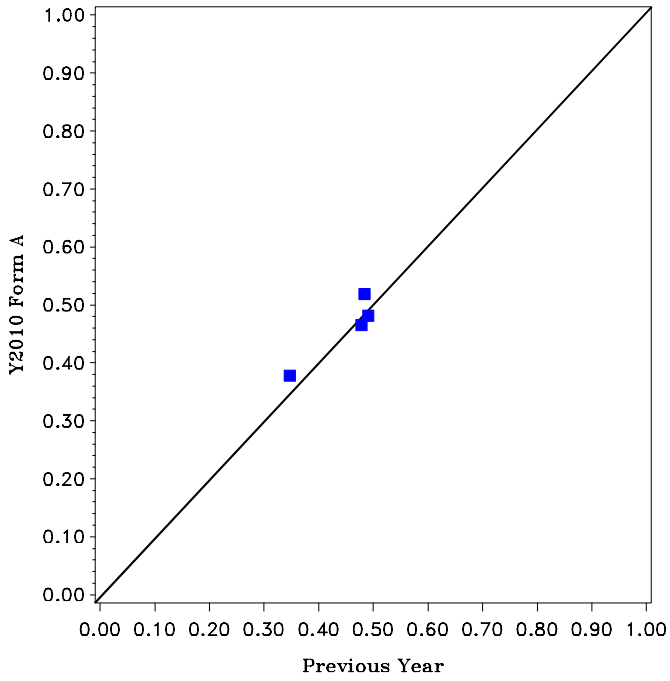

Table 1.49 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 7 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	9	3560786	BCR	2,645	1.44	0.80	11.87	36.67	43.48	6.69	1.29
2008	12	3560784	BCR	2,645	1.46	0.66	3.93	46.92	43.52	3.86	1.78
2008	15	3577916	BCR	2,649	1.04	0.75	22.31	52.51	21.18	3.13	0.87
2008	18	3577919	BCR	2,649	1.47	0.71	6.83	41.45	45.75	4.83	1.13
2010	9	3560786	BCR	29,335	1.41	0.76	12.69	36.81	45.75	3.88	0.70
2010	12	3560784	BCR	29,321	1.57	0.58	2.98	37.73	56.92	1.42	0.75
2010	15	3577916	BCR	29,395	1.14	0.68	15.49	55.41	27.27	1.28	0.50
2010	18	3577919	BCR	29,371	1.45	0.63	5.49	45.15	46.75	1.99	0.58

Table 1.50 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 7 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	9	3560786	BCR	1.4484	-2.0145	-0.3601	2.3746
2008	12	3560784	BCR	1.2433	-3.2515	0.0742	3.1773
2008	15	3577916	BCR	2.2415	-2.2772	0.2504	2.0268
2008	18	3577919	BCR	1.3645	-2.7231	-0.2180	2.9411
2010	9	3560786	BCR	1.9001	-2.3606	-0.6983	3.0589
2010	12	3560784	BCR	1.6647	-3.7561	-0.7960	4.5520
2010	15	3577916	BCR	2.6277	-3.1280	-0.2245	3.3524
2010	18	3577919	BCR	1.8342	-3.4336	-0.4759	3.9095

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

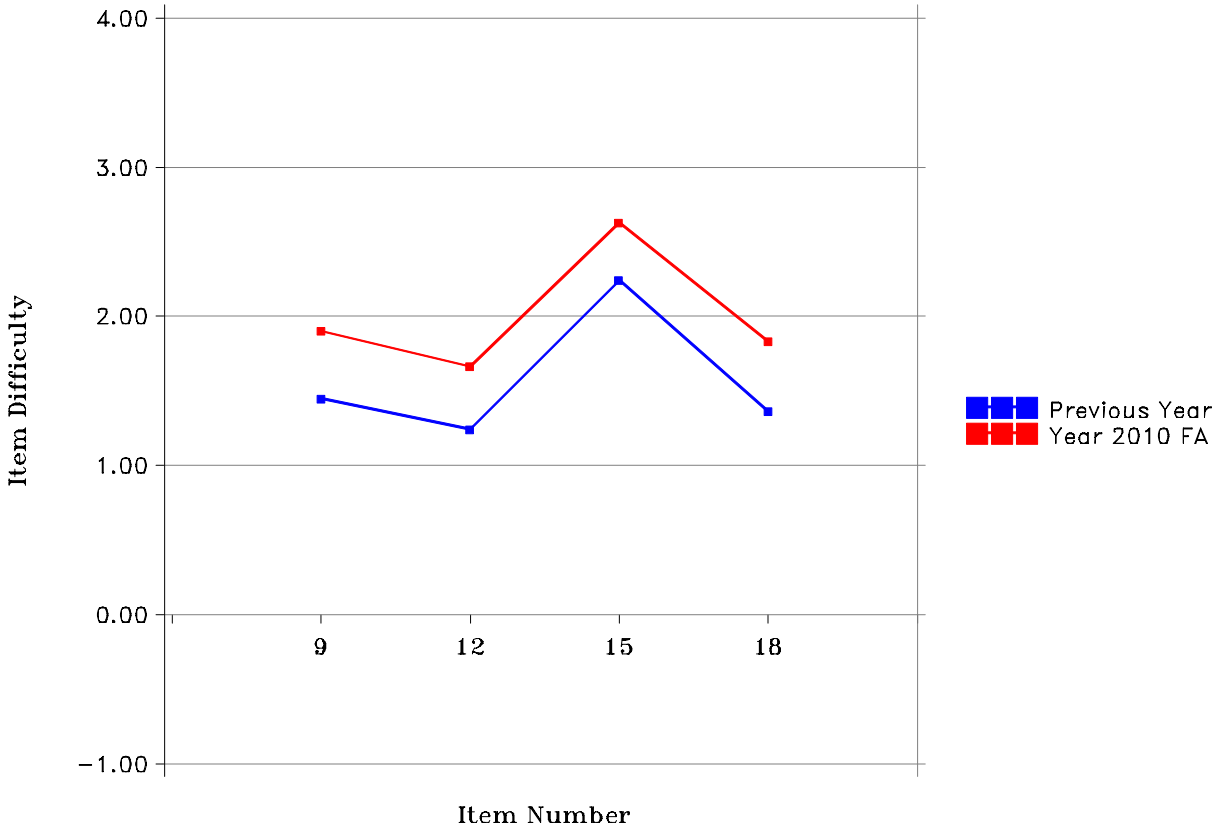

Figure 1.11 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 7 Form A

Table 1.51 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 7 Form B

Item Number	CID	Item Type	Previous Year	Year 10
9	3560795	BCR	0.44	0.56
12	3560796	BCR	0.42	0.48
15	3562065	BCR	0.46	0.49
18	3562066	BCR	0.44	0.44

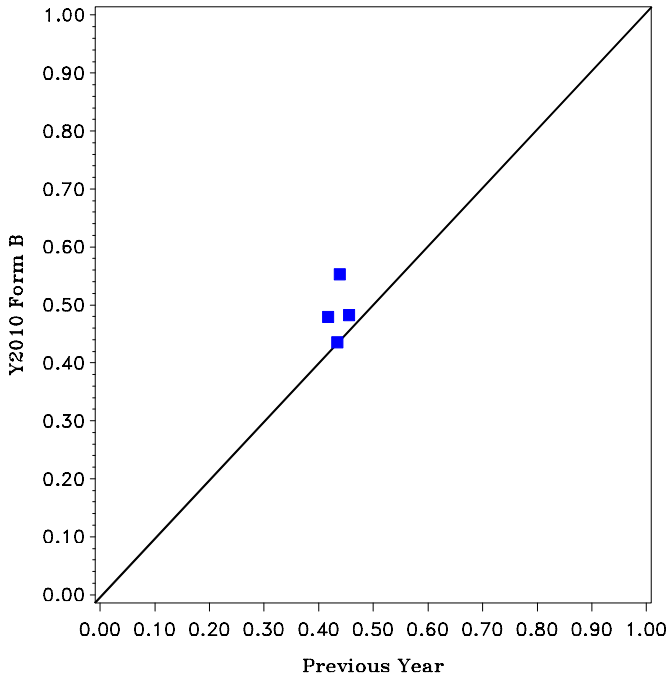

Table 1.52 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 7 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	9	3560795	BCR	2,618	1.32	0.69	8.67	52.44	34.53	3.48	0.88
2008	12	3560796	BCR	2,618	1.25	0.78	16.65	40.37	38.31	2.79	1.87
2008	15	3562065	BCR	2,609	1.37	0.66	5.40	50.40	39.29	2.64	2.26
2008	18	3562066	BCR	2,609	1.31	0.79	14.14	38.79	40.40	3.64	3.03
2010	9	3560795	BCR	28,875	1.67	0.62	2.19	34.30	57.55	5.60	0.28
2010	12	3560796	BCR	28,751	1.45	0.76	11.25	36.28	46.98	4.64	0.70
2010	15	3562065	BCR	28,636	1.47	0.64	3.81	48.95	42.24	3.83	1.10
2010	18	3562066	BCR	28,606	1.33	0.71	12.02	44.53	39.90	2.22	1.21

Table 1.53 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 7 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	9	3560795	BCR	1.7106	-3.0068	0.1364	2.8703
2008	12	3560796	BCR	2.0536	-2.3264	-0.5330	2.8594
2008	15	3562065	BCR	1.5644	-3.2655	-0.0354	3.3009
2008	18	3562066	BCR	1.7899	-2.2379	-0.5161	2.7541
2010	9	3560795	BCR	1.0549	-3.3720	-0.3042	3.6762
2010	12	3560796	BCR	1.8509	-2.3735	-0.6436	3.0172
2010	15	3562065	BCR	1.5632	-3.5277	0.0620	3.4657
2010	18	3562066	BCR	2.2642	-2.8422	-0.5559	3.3981

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

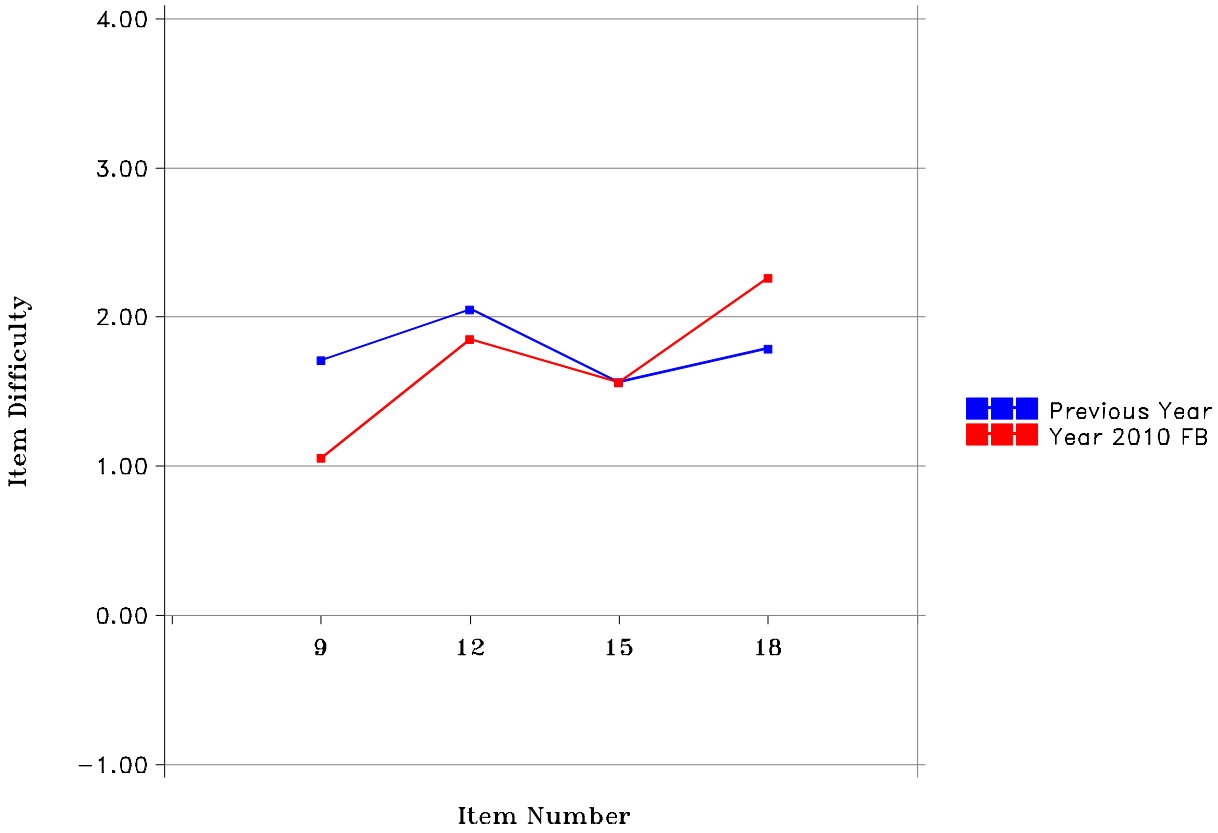

Figure 1.12 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 7 Form B

Table 1.54 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 8 Form A

Item Number	CID	Item Type	Previous Year	Year 10
9	3570926	BCR	0.47	0.49
12	3570928	BCR	0.46	0.47
15	3571809	BCR	0.50	0.52
18	3571811	BCR	0.52	0.54

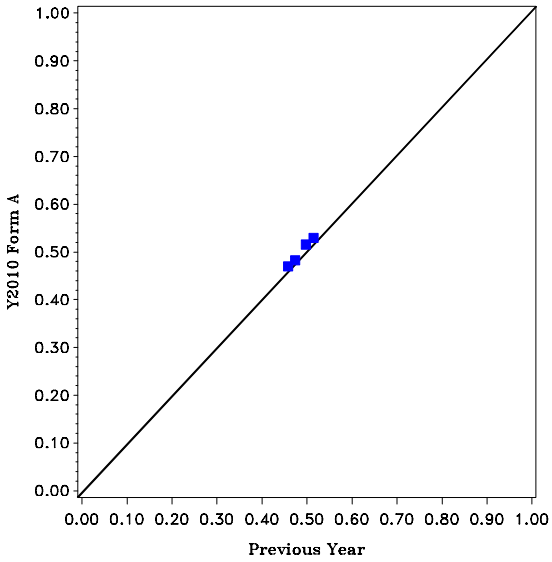

Table 1.55 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 8 Form A

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	9	3570926	BCR	2,620	1.42	0.74	4.20	52.63	32.44	8.28	2.44
2008	12	3570928	BCR	2,620	1.38	0.85	9.47	41.37	34.58	9.01	5.57
2008	15	3571809	BCR	2,389	1.50	0.68	6.28	35.62	54.00	2.01	2.09
2008	18	3571811	BCR	2,389	1.55	0.62	1.26	38.51	54.63	2.26	3.35
2010	9	3570926	BCR	13,959	1.47	0.58	2.39	49.41	45.55	1.93	0.64
2010	12	3570928	BCR	13,902	1.42	0.68	6.53	47.59	40.86	3.82	1.05
2010	15	3571809	BCR	13,946	1.57	0.63	4.25	36.61	55.41	2.79	0.73
2010	18	3571811	BCR	13,900	1.61	0.58	1.61	37.80	56.20	3.09	1.06

Table 1.56 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 8 Form A

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	9	3570926	BCR	1.1094	-3.0651	0.6747	2.3904
2008	12	3570928	BCR	1.3617	-2.2167	0.1265	2.0902
2008	15	3571809	BCR	1.6010	-2.9789	-0.8588	3.8377
2008	18	3571811	BCR	1.0534	-3.9557	-0.2935	4.2492
2010	9	3570926	BCR	1.7109	-4.1238	-0.2525	4.3762
2010	12	3570928	BCR	1.7133	-3.0093	-0.1130	3.1224
2010	15	3571809	BCR	1.6913	-3.1884	-0.7752	3.9636
2010	18	3571811	BCR	1.2165	-3.7399	-0.3116	4.0514

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

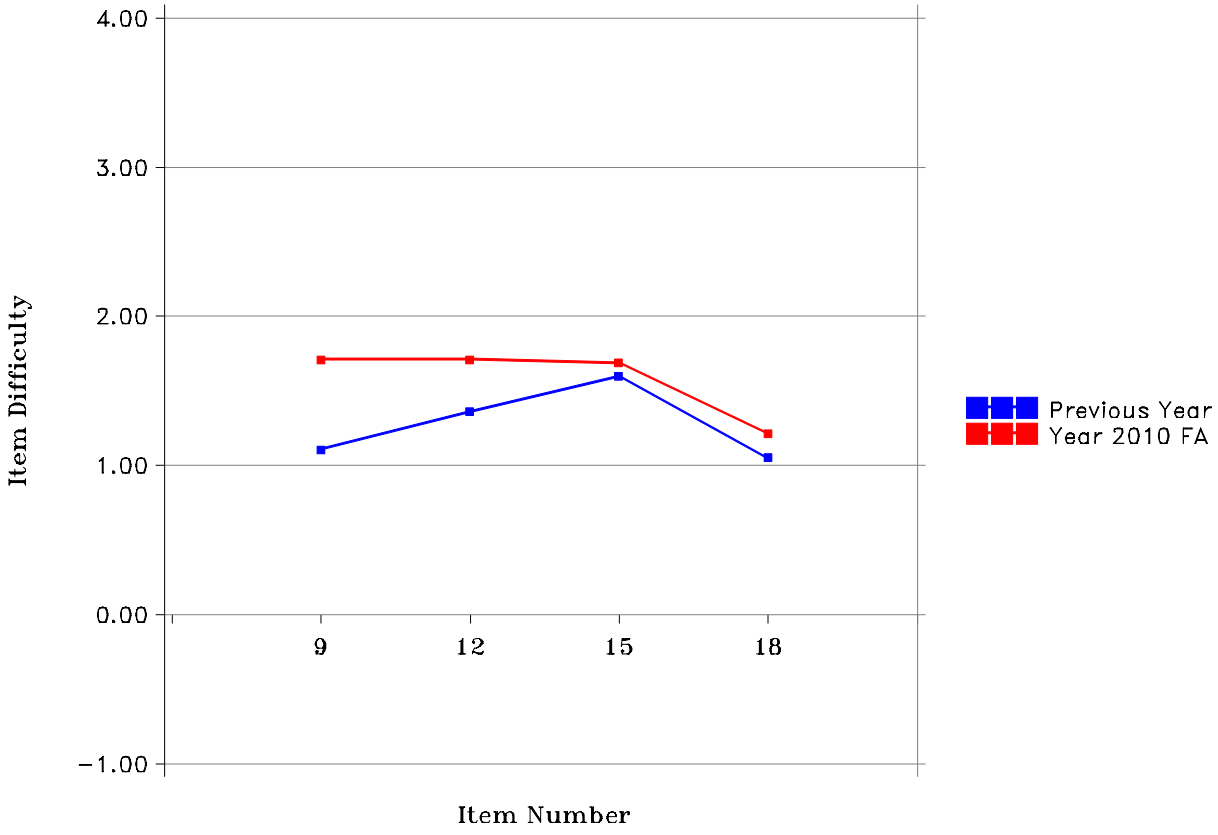

Figure 1.13 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 8 Form A

Table 1.57 P-Value Comparison of Core Items for Previous Year vs. Year 2010: Grade 8 Form B

Item Number	CID	Item Type	Previous Year	Year 10
9	3560856	BCR	0.55	0.62
12	3560855	BCR	0.47	0.51
15	3570913	BCR	0.47	0.51
18	3570916	BCR	0.38	0.45

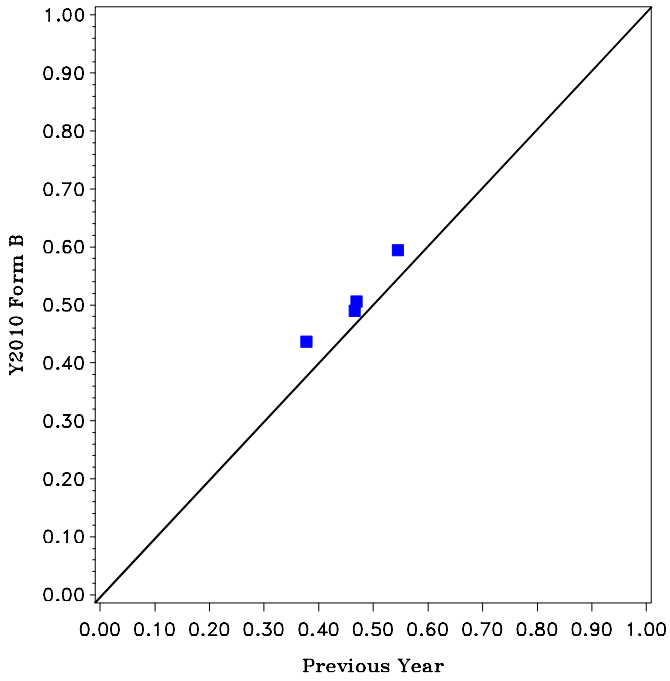

Table 1.58 Score-Point Distribution Comparisons for Previous Year vs. Year 2010: Grade 8 Form B

							Score-Point Distribution (%)				
Year	Item #	Item CID	Item Type	N	Mean	SD	0	1	2	3	Omit
2008	9	3560856	BCR	2,615	1.64	0.68	2.37	28.99	59.58	5.16	3.90
2008	12	3560855	BCR	2,615	1.40	0.65	1.53	48.34	42.33	2.33	5.47
2008	15	3570913	BCR	2,321	1.41	0.61	3.15	48.43	45.24	0.78	2.41
2008	18	3570916	BCR	2,321	1.14	0.76	17.84	42.40	34.77	0.56	4.44
2010	9	3560856	BCR	13,731	1.86	0.57	1.80	18.06	70.87	8.34	0.72
2010	12	3560855	BCR	13,667	1.52	0.58	0.74	48.72	45.73	3.20	1.19
2010	15	3570913	BCR	13,706	1.54	0.62	3.10	42.58	50.09	3.15	0.90
2010	18	3570916	BCR	13,655	1.35	0.72	12.59	40.91	43.41	1.69	1.27

Table 1.59 Rasch Item Difficulty Comparisons for Previous Year vs. Year 2010: Grade 8 Form B

Year	Item #	Item CID	Item Type	Rasch Difficulty	Step 0-1	Step 1-2	Step 2-3
2008	9	3560856	BCR	0.8605	-2.9951	-0.5086	3.5037
2008	12	3560855	BCR	1.2042	-3.9988	0.1901	3.8087
2008	15	3570913	BCR	1.7573	-4.0674	-0.4759	4.5432
2008	18	3570916	BCR	2.6581	-2.8425	-1.0334	3.8759
2010	9	3560856	BCR	0.7717	-2.6448	-0.9787	3.6235
2010	12	3560855	BCR	1.2207	-4.3808	0.3190	4.0618
2010	15	3570913	BCR	1.4923	-3.7058	-0.2043	3.9101
2010	18	3570916	BCR	2.3441	-2.8395	-0.7712	3.6107

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

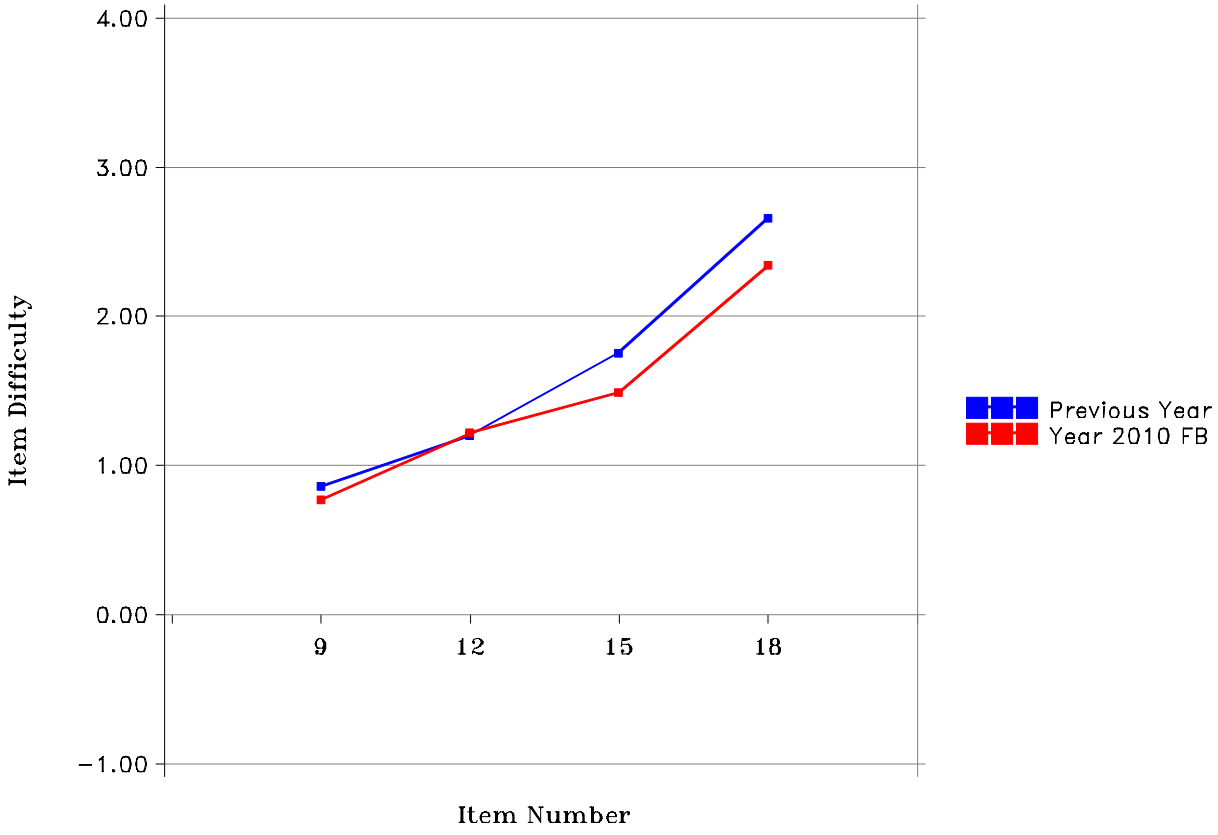

Figure 1.14 Rasch Item Difficulty Comparisons of BCR Items for Previous Year vs. Year 2010: Grade 8 Form B

1.9 Linking, Equating, and Scaling Procedures

For the purpose of year-to-year linking and equating, we constructed a 2010 linking pool: we included only operational selected-response (SR) items (i.e., multiple choice items) that appeared in both years (i.e., 2008 and 2010). It should be noted that all the classical and Rasch analyses of the 2008 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., 2008) of any linking items that remained through the stability check to put the 2010 assessment on a common scale. Accordingly, scale scores of the 2010 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 of grades 3, 4, 5, and 8, a stratified random sampling method was applied in the 2010 state examinee population. To verify that the sample was representative of the statewide examinee population, the distributions of LEA, gender, and ethnicity of the 2010 sample were compared with those of the 2010 population. Appendix A, *The 2010 MSA-Reading Stratified Random Sampling*, provides the results of 2010 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 2010 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 2010 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 2010 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).

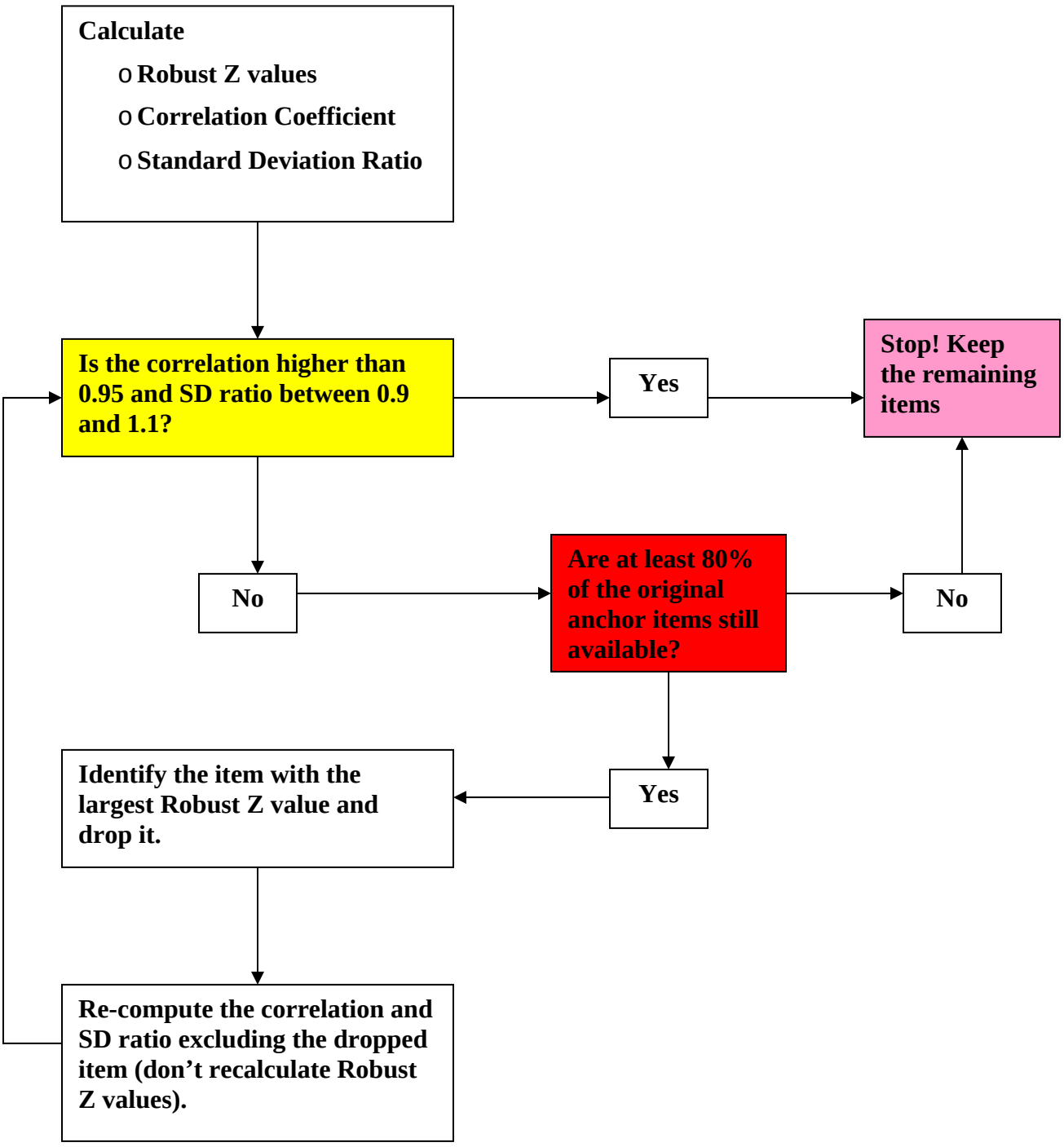

Figure 1.15 Anchor Evaluation Steps Chart for MSA-Reading

Table 1.60 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 3

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	-1.3708	-1.9904	-3.9261	1	-1.3708	-2.1133	-3.4316
3	-3.3089	-3.4759	-.9571	3	-3.3089	-3.3345	.2930
4	-0.1994	-0.0006	1.4425	4	-0.1994	-0.0348	1.2812
6	-1.1969	-1.3126	-.6206	6	-1.1969	-1.3913	-.5840
7	-1.4394	-1.4939	-.2191	7	-1.4394	-1.5115	.0514
8	-2.1425	-2.2694	-.6940	8	-2.1425	-2.1672	.2977
10	0.1005	0.0903	.0715	10	0.1005	0.0185	.0000
11	0.4026	0.3162	-.4284	11	0.4026	0.3	-.1070
12	-0.3764	-0.3597	.2480	12	-0.3764	-0.3848	.3824
26	-0.3721	-1.2306	-5.4933	26	-0.3721	-1.2644	-4.2099
27	-0.1123	-0.3675	-1.5357	27	-0.1123	-0.3767	-.9477
28	0.3685	0.1139	-1.5318	28	0.3685	0.0466	-1.2464
29	1.4373	1.2194	-1.2910	29	1.4373	1.2621	-.4842
30	-2.2073	-2.2804	-.3411	30	-2.2073	-2.2934	-.0213
31	1.6804	1.6222	-.2434	31	1.6804	1.7589	.8339
32	0.4434	0.3603	-.4067	32	0.4434	0.331	-.1579
33	0.7773	0.7826	.1732	33	0.7773	0.8622	.8671
40	1.0945	1.0264	-.3083	40	1.0945	0.8557	-.8147
41	0.2728	0.7024	2.9566	41	0.2728	0.6346	2.3058
42	-0.4174	-0.137	1.9778	42	-0.4174	-0.0909	2.1224
43	-1.272	-1.0177	1.8066	43	-1.272	-1.0774	1.4371
44	-0.4038	-0.3004	.8167	44	-0.4038	-0.3388	.7637
45	-0.7744	-0.6724	.8075	45	-0.7744	-0.7729	.4338
46	1.0333	1.0973	.5583	46	1.0333	1.1518	1.0417
47	-0.2203	-0.0848	1.0273	47	-0.2203	-0.0706	1.2038
14A	0.6246	0.6106	.0466	14B	0.1723	0.3333	1.2625
16A	-0.1433	-0.053	.7308	16B	0.6158	0.4587	-.3902
17A	0.6215	0.6004	.0000	17B	1.0245	0.9057	-.1912
19A	0.0142	-0.0754	-.4494	19B	0.2227	-0.1148	-1.3275
20A	-0.874	-1.2897	-2.5886	20B	0.1027	-0.0387	-.3086
22A	1.4344	1.4607	.3109	22B	0.1116	-0.0337	-.3289
23A	0.6881	0.9664	1.9641	23B	1.1695	1.4067	1.6584
25A	0.0445	0.0853	.4061	25B	0.4932	0.1479	-1.3680

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	2010 Form A	Previous Base Form	2010 Form B
Mean	-.176	-.223	-.130	-.210
SD	1.120	1.190	1.113	1.164

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

	2010 Form A	2010 Form B
Correlation	**.978**	**.974**
SD Ratio	**106%**	**105%**

Values Used for Robust Z Statistics

Mean Diff	-.047	-.080
Median Diff	-.021	-.082
IQR Diff	.206	.260

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.

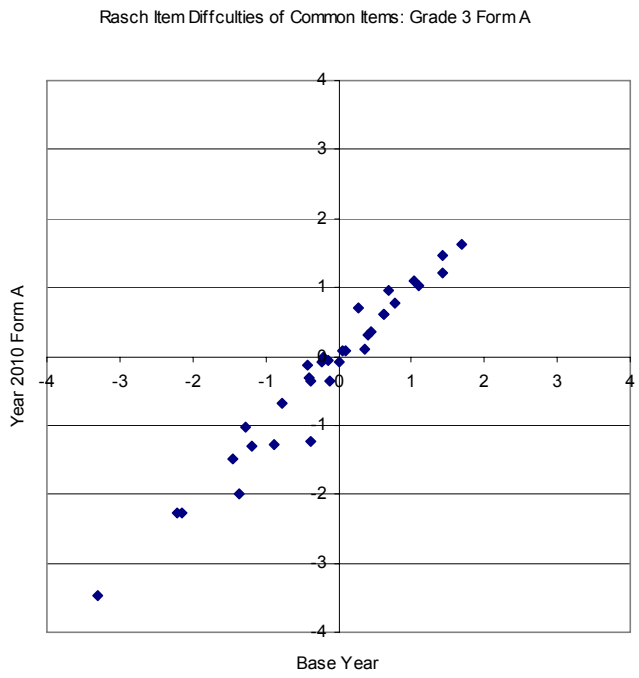

Figure 1.16 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 3 Form A

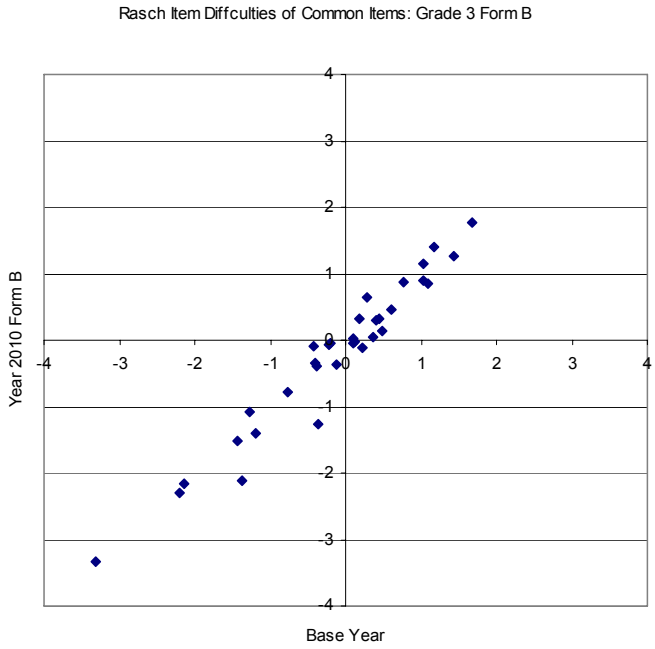

Figure 1.17 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 3 Form B

Table 1.61 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 4

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	0.8485	0.4765	.4132	1	0.8485	0.409	-.0362
2	-0.2773	-0.5362	.9331	2	-0.2773	-0.4766	1.5449
3	-0.0474	-0.5093	.0000	3	-0.0474	-0.4334	.3160
4	-1.694	-2.235	-.3636	4	-1.694	-2.1683	-.2653
6	**-1.3091**	**-2.6286**	**-3.9419**	**6**	**-1.3091**	**-2.3963**	**-4.2996**
7	**-1.4119**	**-3.0823**	**-5.5548**	**7**	**-1.4119**	**-3.0182**	**-7.7165**
8	1.2863	1.0313	.9510	8	1.2863	1.0649	1.3994
10	-1.7138	-2.3059	-.5985	10	-1.7138	-2.2653	-.7734
11	-0.9493	-1.4674	-.2583	11	-0.9493	-1.3466	.2416
25	1.0489	0.5666	-.0938	25	1.0489	0.7039	.5858
26	0.1614	-0.3406	-.1843	26	0.1614	-0.3533	-.5312
27	0.6514	0.179	-.0483	27	0.6514	0.1214	-.6319
28	0.0161	-0.3515	.4334	28	0.0161	-0.4187	-.0053
29	1.1377	0.8114	.6233	29	1.1377	0.7771	.4831
30	0.2403	-0.4683	-1.1339	30	0.2403	-0.4631	-1.7733
31	0.2133	-0.4372	-.8669	31	0.2133	-0.437	-1.4238
32	0.908	0.318	-.5888	32	0.908	0.4326	-.2725
39	0.5973	0.435	1.3771	39	0.5973	0.2158	.3456
40	1.1082	1.0377	1.7990	40	1.1082	1.0076	2.1945
41	0.7541	1.1136	3.7755	41	0.7541	0.9135	3.9060
42	1.3751	1.0251	.5143	42	1.3751	1.0657	.8202
43	0.0633	-0.1136	1.3100	43	0.0633	-0.0889	1.8549
44	0.7616	0.778	2.1985	44	0.7616	0.7287	2.6402
45	-1.4346	-1.4047	2.2605	45	-1.4346	-1.4505	2.7521
46	-0.3025	-0.6243	.6440	46	-0.3025	-0.5781	1.0426
13A	1.1764	0.6127	-.4679	13B	0.7079	0.2258	-.3166
15A	1.0926	0.4219	-.9597	15B	1.3704	0.9738	.2462
16A	-0.6023	-0.8983	.7625	16B	0.375	-0.059	.0000
18A	0.5815	-0.5251	-2.9633	18B	0.8695	0.4454	.0652
19A	1.2015	0.8818	.6536	19B	0.2188	-0.2539	-.2547
21A	0.0718	-0.6007	-.9680	21B	0.9074	0.4625	-.0717
22A	-0.3108	-0.8235	-.2335	22B	0.7248	0.1265	-1.0815
24A	1.4135	1.0604	.5001	24B	-0.4776	-1.4624	-3.6255

Note. Bold-faced items were dropped from the 2010 year-to-year linking pool.

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	2010 Form A	Previous Base Form	2010 Form B
Mean	.202	-.261	.204	-.242
SD	.950	1.137	.918	1.089

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

Correlation Coefficient	**.950**	**.963**
SD Ratio	**120%**	**119%**

Values Used for Robust Z Statistics

Mean Diff	-.462	-.446
Median Diff	-.462	-.434
IQR Diff	.294	.205

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 6 and 7 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	**.958**	**.971**
SD Ratio	**107%**	**107%**

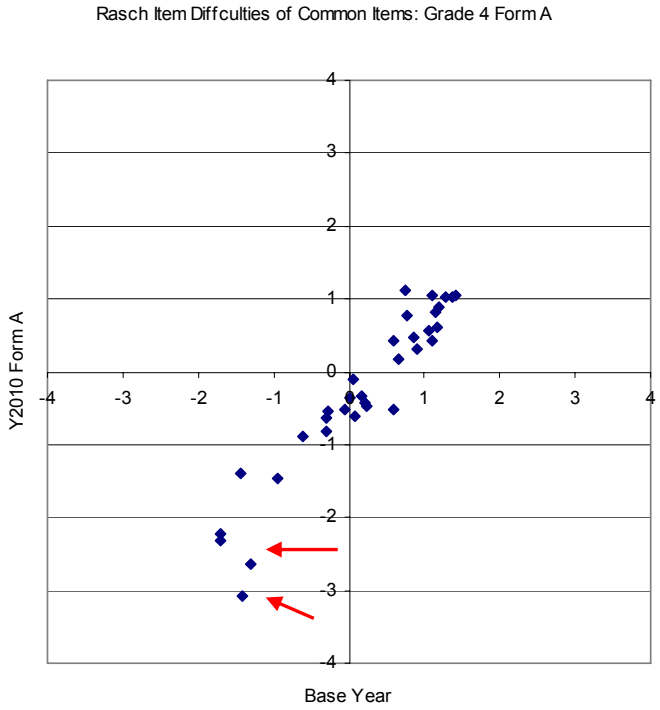

Figure 1.18 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 4 Form A

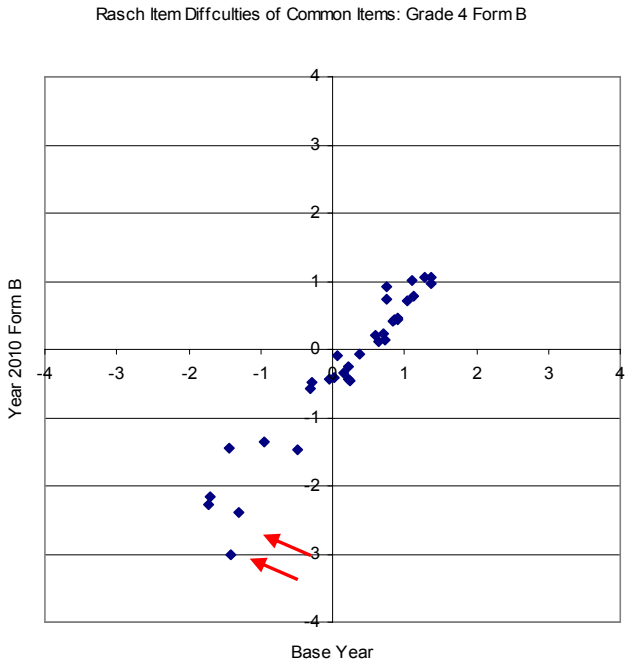

Figure 1.19 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 4 Form B

Table 1.62 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 5

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	-0.6371	-1.0526	-.6658	1	-0.6371	-1.1822	-1.0567
2	-0.2093	-0.4014	.0912	2	-0.2093	-0.5286	-.2432
3	**-1.2263**	**-1.7677**	**-1.0925**	**3**	**-1.2263**	**-2.0338**	**-2.0020**
4	**-1.4827**	**-2.2796**	**-1.9582**	**4**	**-1.4827**	**-2.0844**	**-1.2606**
6	-1.3213	-1.8364	-1.0033	6	-1.3213	-1.9752	-1.4486
7	**-1.8707**	**-2.6185**	**-1.7919**	**7**	**-1.8707**	**-2.4486**	**-1.1748**
8	-1.0118	-1.7492	-1.7566	8	-1.0118	-1.6591	-1.4248
9	1.4561	1.2398	.0091	9	1.4561	1.2369	.1174
10	-1.7612	-2.0257	-.1542	10	-1.7612	-1.9351	.2806
24	0.6018	0.4012	.0623	24	0.6018	0.5097	.5753
25	**1.0194**	**-0.0852**	**-3.0009**	**25**	**1.0194**	**0.3172**	**-1.6226**
26	-0.6907	-0.9608	-.1732	26	-0.6907	-0.943	.0000
27	-0.8114	-1.4009	-1.2555	27	-0.8114	-1.1978	-.4849
28	1.4256	1.2871	.2728	28	1.4256	1.1882	.0519
29	0.0778	-0.06	.2751	29	0.0778	-0.0931	.2915
30	0.5353	0.4359	.4053	30	0.5353	0.3653	.2947
31	1.1113	0.5721	-1.0850	31	1.1113	0.6158	-.8780
38	1.2237	1.0047	.0000	38	1.2237	0.979	.0256
39	-0.9702	-1.0463	.4842	39	-0.9702	-1.0189	.7317
40	0.8757	0.7923	.4595	40	0.8757	0.7684	.5206
41	1.1801	1.0638	.3480	41	1.1801	0.9891	.2190
42	-0.8012	-0.9256	.3206	42	-0.8012	-0.9658	.3142
43	0.218	0.276	.9386	43	0.218	0.283	1.1413
44	0.0925	0.1565	.9590	44	0.0925	0.0828	.8722
45	1.5135	1.4731	.6052	45	1.5135	1.4239	.5843
12A	-0.0933	-0.4982	-.6299	**12B**	**0.5582**	**-0.2109**	**-1.8636**
14A	0.8069	0.5583	-.1003	14B	0.3834	-0.1227	-.9162
15A	0.0189	-0.1692	.1047	15B	0.3344	-0.198	-1.0109
17A	0.2838	-0.0727	-.4659	17B	1.2204	0.7099	-.9320
18A	1.7917	1.8983	1.1033	18B	0.7579	0.0945	-1.4828
20A	0.1876	0.2685	1.0162	20B	0.0859	-0.3422	-.6351
21A	0.6321	0.0909	-1.0918	21B	1.3872	1.2187	.3001
23A	0.5376	0.0416	-.9386	23B	1.0174	0.8449	.2857

Note. Bold-faced items were dropped from the 2010 year-to-year linking pool.

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 Form A	2010 Form A	Previous Year Form B	2010 Form B
Mean	.082	-.224	.130	-.222
SD	1.027	1.164	1.036	1.127

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

Correlation Coefficient	**.974**	**.978**
SD Ratio	**113%**	**109%**

Values Used for Robust Z Statistics

Mean Diff	-.306	-.351
Median Diff	-.219	-.252
IQR Diff	.399	.375

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 3, 4, 7, and 25 appearing on both forms and item number 12 appearing only Form B were dropped from the linking pool.

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

Correlation Coefficient	**.966**	**.976**
SD Ratio	**109%**	**105%**

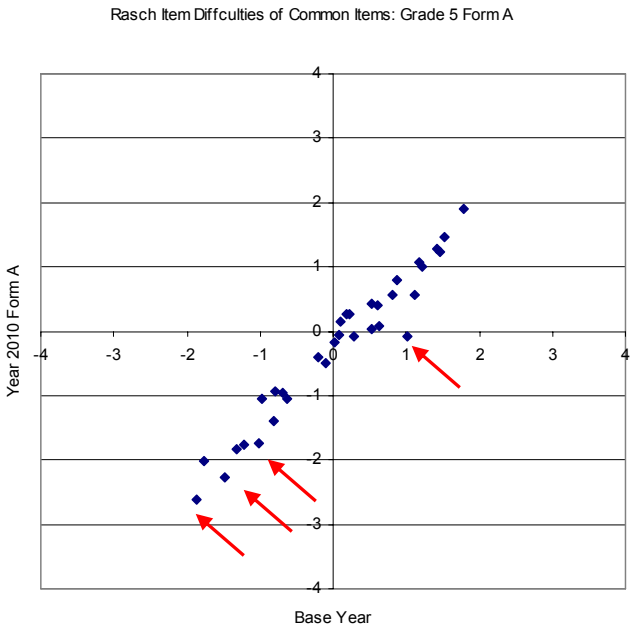

Figure 1.20 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 5 Form A

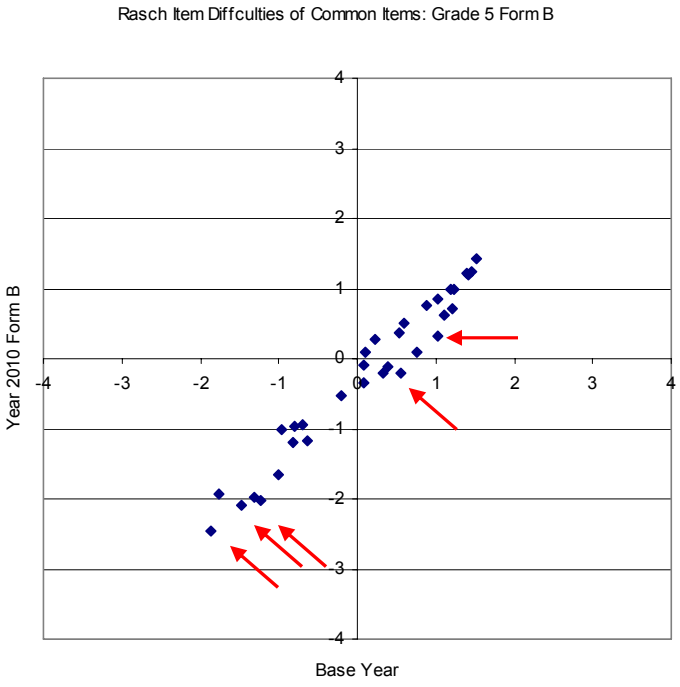

Figure 1.21 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 5 Form B

Table 1.63 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 6

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	-1.3336	-1.8574	-.9370	1	-1.3336	-1.8953	-1.0799
2	-2.0006	-2.7467	-1.7084	2	-2.0006	-2.8311	-2.0795
3	-0.9089	-0.8238	1.1761	3	-0.9089	-0.8905	1.0773
4	-1.1479	-1.2386	.5660	4	-1.1479	-1.3008	.4403
5	-0.5668	-0.612	.7239	5	-0.5668	-0.6844	.5716
7	-1.4246	-1.8851	-.7173	7	-1.4246	-1.8977	-.7504
8	1.0944	1.2382	1.3798	8	1.0944	1.142	1.1859
9	-0.485	-0.4648	.9509	9	-0.485	-0.5819	.6485
10	-1.5147	-2.0853	-1.0994	10	-1.5147	-2.1193	-1.2394
24	-0.0765	-0.0679	.9106	24	-0.0765	-0.2195	.4771
25	1.7746	1.6605	.4848	25	1.7746	1.6438	.5225
26	0.2067	-0.1359	-.3082	26	0.2067	-0.1047	-.1491
27	-0.3212	-0.6666	-.3179	27	-0.3212	-0.7659	-.6448
28	1.0501	0.8448	.1683	28	1.0501	0.7788	.0000
29	0.4286	0.1218	-.1839	29	0.4286	0.0375	-.4455
30	0.2397	-0.1143	-.3477	30	0.2397	-0.2472	-.8017
31	1.7013	1.484	.1267	31	1.7013	1.4041	-.0963
38	-0.3824	-0.706	-.2422	38	-0.3824	-1.0361	-1.4220
39	0.0668	-0.187	.0000	39	0.0668	-0.1616	.1595
40	1.4178	1.4338	.9363	40	1.4178	1.3287	.6775
41	-1.2424	-1.2065	1.0054	41	-1.2424	-1.3051	.7757
42	1.7352	1.5911	.3807	42	1.7352	1.6138	.5574
43	0.1718	0.2795	1.2545	43	0.1718	0.009	.4035
44	-0.3472	-0.3397	.9068	44	-0.3472	-0.3685	.9297
45	1.433	1.1508	-.0986	45	1.433	1.172	.0383
12A	-0.0772	-0.5347	-.7069	12B	0.7297	0.2214	-.8813
14A	0.6277	0.2649	-.3783	14B	0.4921	-0.0021	-.8289
15A	0.9539	0.5193	-.6274	15B	-0.172	-0.9529	-1.8950
17A	0.3013	-0.2049	-.8759	17B	0.7134	0.4953	.1978
18A	0.8371	0.6231	.1381	18B	-0.2169	-0.7598	-1.0100
20A	0.4772	0.0323	-.6632	20B	0.878	0.7154	.4042
21A	0.1191	-0.1182	.0573	21B	0.4139	-0.1777	-1.1911
23A	-0.2807	-0.7392	-.7104	23B	0.0328	-0.2768	-.1424

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 Form A	2010 Form A	Previous Year Form B	2010 Form B
Mean	.077	-.166	.074	-.243
SD	1.011	1.093	1.009	1.110

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

Correlation Coefficient	**.980**	**.980**
SD Ratio	**108%**	**110%**

Values Used for Robust Z Statistics

Mean Diff	-.243	-.317
Median Diff	-.254	-.271
IQR Diff	.389	.363

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, none of items were dropped.

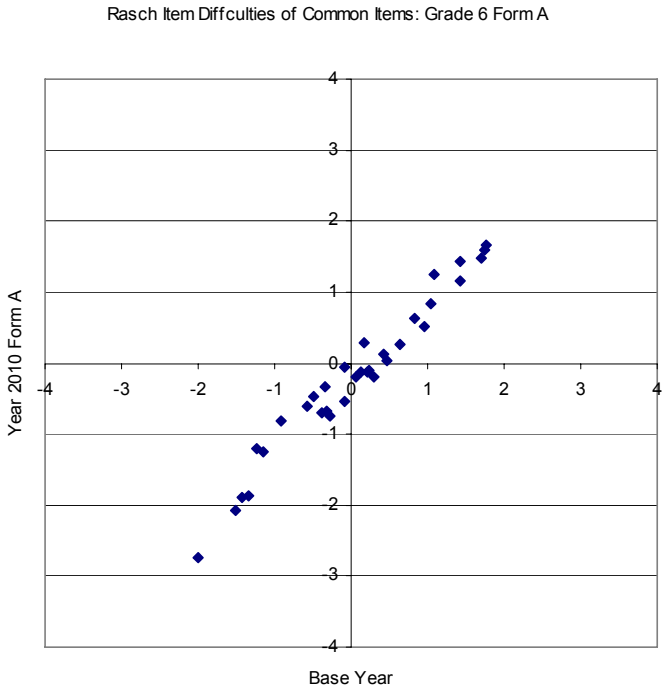

Figure 1.22 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 6 Form A

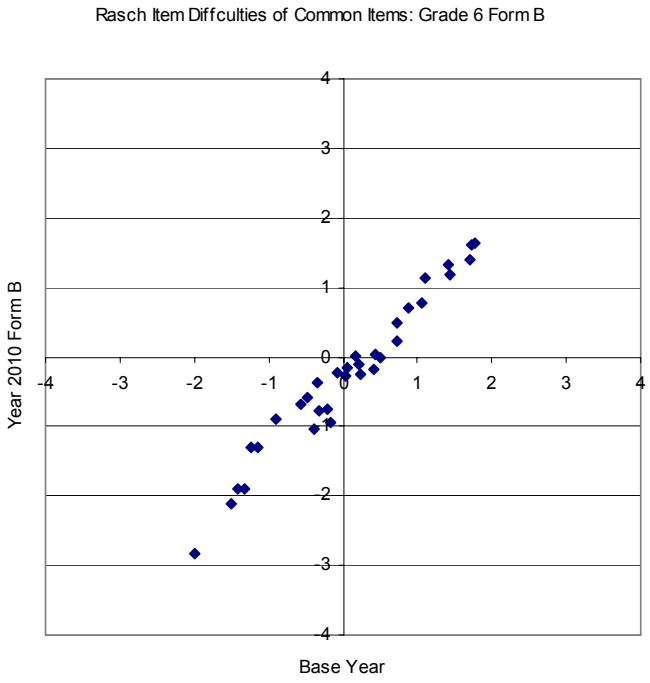

Figure 1.23 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 6 Form B

Table 1.64 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 7

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	-1.9151	-2.431	-1.7908	1	-1.9151	-2.46	-1.7341
2	-1.5468	-1.9246	-1.1977	2	-1.5468	-1.7729	-.5429
3	-0.5743	-0.5931	.3473	3	-0.5743	-0.5448	.4115
5	-1.1399	-1.5956	-1.5330	5	-1.1399	-1.4934	-1.0186
6	-1.8025	-2.1733	-1.1676	6	-1.8025	-2.1363	-.9451
20	-0.4933	-0.803	-.9046	20	-0.4933	-0.8037	-.8577
21	0.5605	0.4478	-.0568	21	0.5605	0.5806	.3764
22	-0.0535	-0.2244	-.3073	22	-0.0535	-0.1502	-.0597
23	0.0344	-0.1467	-.3512	23	0.0344	-0.0871	-.1523
24	-0.6959	-1.0184	-.9597	24	-0.6959	-0.8549	-.2924
25	-1.0941	-1.3415	-.6365	25	-1.0941	-1.3045	-.4843
26	1.1265	0.8743	-.6572	26	1.1265	0.8883	-.5881
27	0.3428	0.3502	.4601	27	0.3428	0.4381	.6572
28	-0.2631	-0.3454	.0740	28	-0.2631	-0.2495	.3521
29	-1.3411	-1.4406	.0000	29	-1.3411	-1.4218	.0000
30	0.7624	0.8252	.6985	30	0.7624	0.9434	.9772
31	0.6796	0.5486	-.1356	31	0.6796	0.6654	.2483
38	0.0964	0.363	1.5756	38	0.0964	0.424	1.5246
39	-0.3345	0.0305	1.9991	39	-0.3345	0.0075	1.5784
40	-0.6932	-0.2566	2.3072	40	-0.6932	-0.2567	1.9312
41	-0.7991	-0.3267	2.4613	41	-0.7991	-0.2295	2.4283
42	**-0.1112**	**0.3549**	**2.4342**	**42**	**-0.1112**	**0.4746**	**2.4887**
43	0.1544	0.2431	.8100	43	0.1544	0.2768	.7584
44	-0.7112	-0.2516	2.4062	44	-0.7112	-0.2209	2.1321
45	0.3510	0.3876	.5857	45	0.3510	0.6073	1.2584
8A	0.5804	0.5059	.1076	8B	0.3175	0.0174	-.8192
10A	0.9417	0.8005	-.1795	10B	-0.2717	-0.6923	-1.2692
11A	0.8806	0.9353	.6636	11B	0.6077	0.364	-.6086
13A	0.6795	0.3524	-.9795	13B	-0.1074	-0.3755	-.6998
14A	0.2019	-0.0988	-.8659	14B	0.4863	0.3743	-.1169
16A	-0.0688	-0.1881	-.0852	16B	0.6729	0.6596	.2517
17A	0.8732	0.9004	.5453	17B	0.7843	0.908	.7632
19A	-0.6846	-0.6228	.6942	19B	0.9313	0.7074	-.5347

Note. Bold-faced items were dropped from the 2010 year-to-year linking pool.

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 Form A	2010 Form A	Previous Year Form B	2010 Form B
Mean	-.184	-.238	-.183	-.204
SD	.829	.919	.810	.909

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

Correlation Coefficient	**.956**	**.948**
SD Ratio	**111%**	**112%**

Values Used for Robust Z Statistics

Mean Diff	-.055	-.021
Median Diff	-.100	-.081
IQR Diff	.314	.362

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 42 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	**.961**	**.955**
SD Ratio	**110%**	**111%**

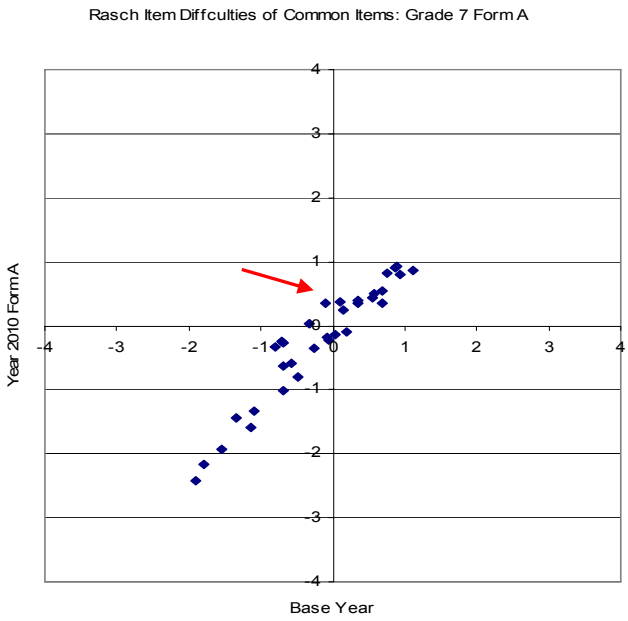

Figure 1.24 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 7 Form A

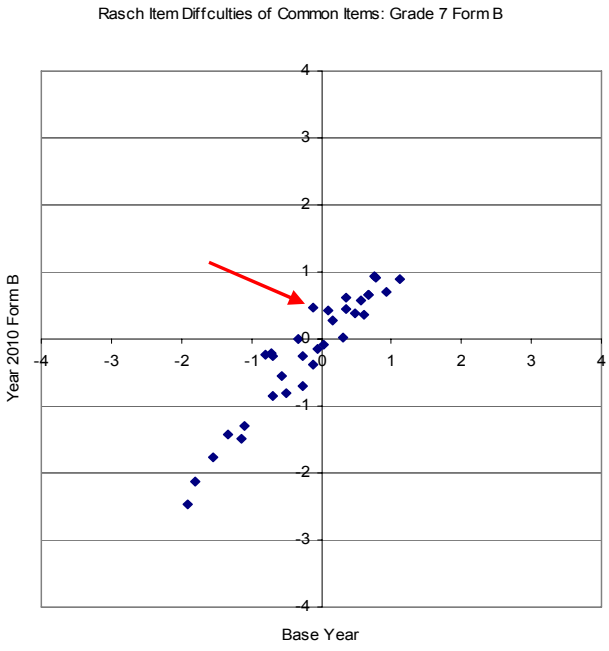

Figure 1.25 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 7 Form B

Table 1.65 Rasch Item Difficulties and Robust Z Values for Previous Year vs. Year 2010: Grade 8

Item Seq No.	Previous Year	Y2010 Form A	Robust Z	Item Seq No.	Previous Year	Y2010 Form B	Robust Z
1	**-1.7533**	**-2.5372**	**-2.0264**	**1**	**-1.7533**	**-2.4954**	**-1.5091**
2	-1.6274	-2.0616	-.9524	2	-1.6274	-1.976	-.5111
4	-0.6076	-0.7226	.0279	4	-0.6076	-0.7468	.0200
5	-0.6192	-0.6277	.3550	5	-0.6192	-0.6893	.1953
6	-1.3966	-1.6787	-.4853	6	-1.3966	-1.6017	-.1471
20	0.2439	0.0492	-.2168	20	0.2439	0.0282	-.1740
21	-0.5739	-1.0418	-1.0559	21	-0.5739	-1.0137	-.7424
22	0.776	0.57	-.2515	22	0.776	0.6289	.0000
23	-0.3472	-0.8297	-1.1007	23	-0.3472	-0.6279	-.3389
24	-0.9291	-1.101	-.1468	24	-0.9291	-1.1351	-.1494
25	0.0425	-0.0351	.1428	25	0.0425	-0.0941	.0266
26	-0.5775	-0.708	-.0197	26	-0.5775	-0.6327	.2331
27	0.9021	0.778	.0000	27	0.9021	0.8963	.3584
28	0.564	0.4606	.0636	28	0.564	0.5284	.2828
29	1.1974	1.2364	.5009	29	1.1974	1.3816	.8403
30	-0.2137	-0.3497	-.0365	30	-0.2137	-0.3963	-.0900
31	0.8184	0.7783	.2580	31	0.8184	0.8342	.4132
38	-0.3439	-0.1426	.9994	38	-0.3439	-0.0781	1.0472
39	0.5919	0.9303	1.4205	39	0.5919	1.0626	1.5669
40	**-0.2946**	**0.1233**	**1.6646**	**40**	**-0.2946**	**0.2278**	**1.6981**
41	1.0198	1.1841	.8857	41	1.0198	1.3077	1.1033
42	**-0.3137**	**0.4291**	**2.6625**	**42**	**-0.3137**	**0.4946**	**2.4232**
43	-0.1356	0.0472	.9426	43	-0.1356	0.2038	1.2339
44	**-1.2722**	**-0.4827**	**2.8059**	**44**	**-1.2722**	**-0.2725**	**2.9087**
45	**-1.1352**	**-0.559**	**2.1508**	**45**	**-1.1352**	**-0.5024**	**1.9781**
8A	-0.525	-0.9565	-.9441	8B	0.8518	0.4049	-.7604
10A	1.2858	0.9424	-.6735	10B	1.1997	0.6005	-1.1467
11A	0.4757	0.5983	.7577	11B	0.5533	0.3506	-.1410
13A	0.5756	0.0699	-1.1720	13B	-0.1105	-0.6094	-.8923
14A	0.9798	1.1209	.8145	14B	-0.0788	-0.5145	-.7320
16A	0.8337	0.3898	-.9822	16B	0.9417	0.3349	-1.1659
17A	0.0322	-0.2667	-.5369	17B	-0.6485	-0.8095	-.0353
19A	-1.2446	-1.5	-.4033	19B	-0.1232	-0.4774	-.5253

Note. Bold-faced items were dropped from the 2010 year-to-year linking pool.

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 Form A	2010 Form A	Previous Year Form B	2010 Form B
Mean	-.108	-.179	-.103	-.163
SD	.865	.955	.831	.901

*Note: mean and standard deviation of Year 10 is calculated with freely calibrated estimates.

Correlation and Standard Deviation Ratio

Correlation Coefficient	**.925**	**.889**
SD Ratio	**110%**	**108%**

Values Used for Robust Z Statistics

Mean Diff	-.070	-.060
Median Diff	-.124	-.147
IQR Diff	.440	.533

Based on correlation coefficients, SD ratios, robust z, and item difficulty plot, item number 1, 40, 42, 44, and 45 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	**.970**	**.947**
SD Ratio	**112%**	**110%**

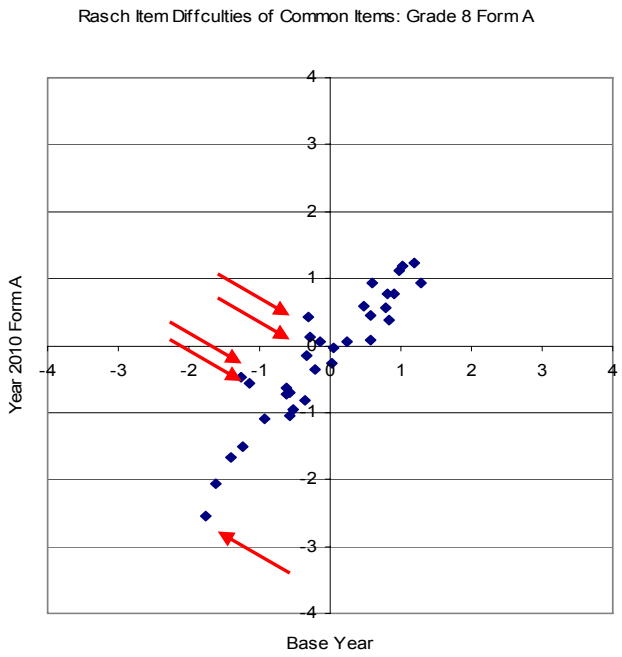

Figure 1.26 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 8 Form A

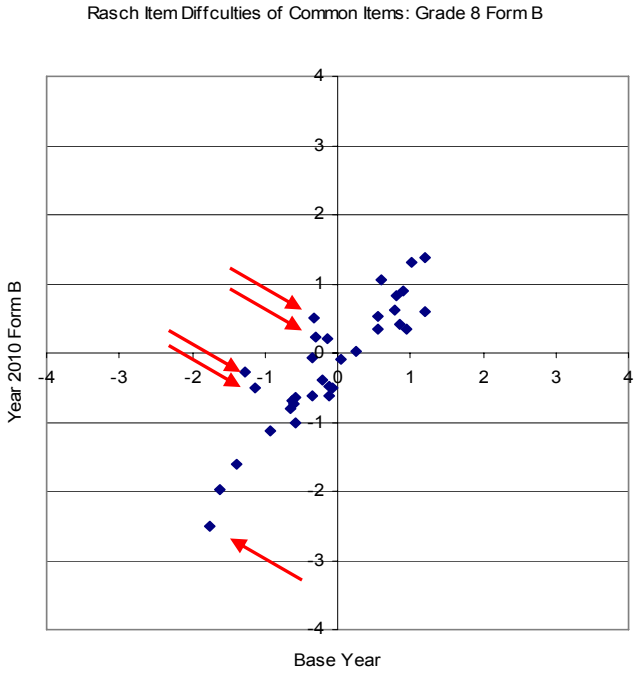

Figure 1.27 Item Difficulty Plot of Previous Year Form vs. Current Year (2010) Form: Grade 8 Form B

Reporting Scale Scores

In order to facilitate the use and interpretation of the results of the 2010 MSA-Reading, the following formula was used to convert each student's ability or theta to the reporting scale score:

$$ReportingAbilityScaleScore = 32.8271 \cdot theta + 362.7449$$

$$ReportingSE = 32.8271 \cdot SE$$

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 2010 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 2010 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 2010 MSA-Reading test scores, two types of scores were created: 240-650 scale scores, and performance levels and descriptions. The scores can be interpreted the same way across different administration years since the tests were on the same scale either on the 2003 administration (i.e., grades 3, 5, and 8) or 2004 administration (i.e., grades 4, 6, and 7) using IRT equating and scaling.

240-650 Scale Scores

As explained in section 1.9, *Linking, Equating, and Scaling Procedures*, the 2010 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 2010 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.marylandpublicshools.org*):

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

As Table 2.1 shows a range of scale scores at each performance level; for example, grade 4 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 2010 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 2010 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 2010 MSA-Reading blueprints provide extensive evidence regarding the alignment between the content in the 2010 MSA-Reading and the *VSC*. It should be noted that the 2010 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 2010 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 2010 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 2010 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 2010 Maryland item bank. More information on *DIF* analyses can be obtained in section 3.7, *Differential Item Functioning*.

Evidence from Internal Structure

The 2010 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 2010 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 2010 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 2010 MSA-Reading was met.

Table 1.67 The 2010 MSA-Reading Eigenvalues between the First and Second Components

Grade	Form	Number of Items	First Eigenvalue	Second Eigenvalue
3	A	37	11.16	1.54
	B	37	12.53	1.45
4	A	37	12.42	1.35
	B	37	11.52	1.47
5	A	37	10.05	1.45
	B	37	10.67	1.51
6	A	37	9.77	1.32
	B	37	8.98	1.50
7	A	37	11.01	1.57
	B	37	11.01	1.47
8	A	37	11.00	1.72
	B	37	11.09	1.64

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 2010 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 2010 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 2010 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 2010 operational items that had been already placed on the base scale during the 2010 operational calibration and equating. For example, each unique field test item appearing on one of five reading test forms (i.e., 1, 3, and 5) 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 2010 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 2010 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.