# The Modified Maryland School Assessment (Mod-MSA)

# Reading:

## Grades 3 through 8

**Technical Report:**

**2010 Administration**

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ACKNOWLEDGEMENTS

The 2010 *Modified Maryland School Assessment (Mod-MSA): Reading Technical Report* is the product of the combined resources of the research scientist, the research analyst, the content experts, and the program manager at Pearson.

Members of the National Psychometric Council, Bert Green, Huynh Huynh, Bob Lissitz, Mark Moody, Bill Schafer, and Steve Wise reviewed this report. Special acknowledgement is given to the following Maryland State Department of Education (MSDE) staff members for their input and guidance: Janet Bagsby, Trinell Bowman, Matthew Dammann, Damon Hartzler, Ann Herrmann, Ray Scott, and Leslie Wilson.

## OVERVIEW OF THIS MANUAL

The *Modified Maryland School Assessment Program* (Mod-MSA) *Technical Manual* for the 2010 Reading Administration is organized around ten major sections, plus a section for the appendices. An overview of this manual is provided below.

### Section 1: Introduction

This section presents Mod-MSA Reading test's background, rationale, eligibility criteria, test administration (test materials, test administration schedule, student participation, accommodation used for assessment, test format, and security of test materials), quality control procedures, and item bank construction.

### Section 2: Test Design, Development, and Scoring of the 2010 Mod-MSA: Reading

This section describes the test design, development, and the scoring process of the Mod-MSA: Reading test. It provides the test design and structure, development and review of the test (National Psychometric Council, Content Review Committee, Bias Review Committee, and Vision Review Committee), structure of the operational test, item analyses undertaken prior to the creation of the operational test (classical item analyses, differential item functioning analyses, and item response theory analyses), items flagged for inspection prior to the creation of the operational form, the items selected for the 2010 operational tests, and the scoring procedure for the Mod-MSA: Reading.

### Section 3: Central Tendency Measures and other Classical Item Analyses of the 2010 Mod-MSA: Reading

This section provides central tendency measures and classical item analyses. It includes explanation of the standard error of the mean, the measure of central tendencies, and item level descriptive statistics.

### Section 4: Scale Creation, Equating, and Raw Scores to Scale Scores Conversion via Item Response Theory (IRT) Procedures

This section explains the item response theory (IRT) procedures (the Rasch model and the conditional standard error of measurement); calibration and scaling procedure for Grades 3-5; specifics of creating the base scale for the Mod-MSA: Reading (Grades 3 -5); calibration and equating for Grades 6 - 8; specifics of scaling and equating for Grades 6-8 (generalized Robust-Z procedure; reporting scale scores (raw score to scale score conversion for the total test score, and by subscales, guide for selecting linking items and step-by-step procedure for selecting linking items); tables of unequated linking item difficulties; reporting of the scale scores (tables of raw scores to scale scores by total test and also by sub-scales/strands); score interpretation (scale scores and performance level descriptors); and the final scale score cutoffs at each performance level.

**Section 5: Comparative Classical and IRT Stats Across Years for the Common Items Used in Grades 6-8**

This section compares common items used for linking Grades 6 to 8 across the two years of operation. Both p-values from classical statistical analysis and the Rasch difficulties from IRT analysis are compared for each grade across 2009 and 2010. Graphical displays are also provided for these statistics.

**Section 6: IRT Item Statistics for the 2010 Mod-MSA: Reading**

This section begins with a description for the rationale for the use of the Rasch model. It then provides an explanation of the fit indices and the provision of IRT item level analyses.

**Section 7: Test Reliability**

This section describes the precision and reliability undertaken by classical methods (standard error of measurement of the test and the Cronbach's Alpha). It also includes a brief statement relating to IRT methods in measuring the precision of the test and explains the decision accuracy and consistency at the cut scores. Tables relating to accuracy and consistency are also provided in this section.

**Section 8: Test Validity**

To investigate the validity evidence of the 2010 Mod-MSA: Reading, content-related evidence, evidence from item development methods, differential item functioning (DIF) analysis on gender and ethnicity (white and black students), and evidence from internal structure were collected. Also, a study comparing the mode of administration was undertaken by Pearson to validate the online administration of the test. Evidence with respect to the unidimensionality of the test was undertaken by a principal component analysis.

**Section 9: Summary of Operational Test Results for the 2010 Mod-MSA: Reading**

This section presents the summary of the operational test results. It includes the classical descriptive test statistics of the raw scores, scale score descriptive statistics for the test, frequency of students in each proficiency level by grades, and an explanation and depiction of the test characteristic curve (TCC), the test information curve, and the conditional standard error of measurement (CSEM) of the tests.

**Section 10: References**

This section presents the references used in producing this report.

**Section 11: Appendices**

This section presents Appendix A to Appendix G. Appendix A provides a list of operational items selected after data review, while Appendix B and Appendix C provide the item distractor analyses, and the frequency distribution and histograms of the scale scores, respectively. The standard setting report is provided in Appendix D, and the alignment of the test items with the state standards is provided in Appendix E. Appendix F provides the criterion used in identifying

students with disabilities for participation in the Mod-MSA: Reading program. Finally, Appendix G provides the comparability study of the paper-and-pencil and online modes of Mod-MSA administration for Grades 4-5.

# 1. INTRODUCTION

## 1.1. Background

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

The MSA assesses students' proficiency in mathematics and reading, and it is administered to students in grades 3 through 8. It should be noted that in 2007 the MSA was administered using a new vendor and applying a different IRT method (e.g., the Rasch model); therefore, a transformation of scale scores using equipercentile method was conducted in that year. Detailed information on scale score transformation can be found in Appendix C, *Year 2006 MSA-Reading Recalibration Results from 3PL IRT to the Rasch Model Using Equipercentile Method* in the 2007 *MSA-Reading Technical Report*.

Also in 2007, the U.S. Department of Education issued guidance for the development of Alternative Assessment based on Modified Academic Achievement Standards (also known as AA-MAAS or "Modified Assessments"). These guidelines are based on grade-level academic content standards and modified academic achievement standards. Adhering to these guidelines, Maryland, in 2008, developed the Modified Maryland School Assessment (Mod-MSA) based on grade-level academic content standards and modified academic achievement standards.

The Mod-MSA is an alternate assessment to the Maryland School Assessment Program (MSA) for students with disabilities who meet specific eligibility criteria and who are unable to participate in the MSA, even with accommodations. Students are identified through the Individualized Education Program (IEP) process for participation in the Mod-MSA.

Prior to the first administration of the Mod-MSA tests (Grades 6 to 8 in spring 2009 and Grades 3 to 5 in spring 2010), approximately 95% of the students, regardless of their eligibility, had taken the Maryland School Assessments (MSA) examination. The Mod-MSA assessments in reading and mathematics were designed for students with disabilities who, based on a decision-making process undertaken by their Individual Educational Planning (IEP) team, met specific eligibility criteria. The Mod-MSA tests, as stated above, are alternatives to the tests in the MSA Program. The alternate assessments based on modified achievement standards (AA-MAS) are commonly referred to as 2% assessments. They are specified by the guidelines set by the U.S. Department of Education (DOE) and are based on the U.S. DOE final rule, of April 9, 2007[1]. According to the rule, although states may test more than 2% of the population using the AA-MAS, they may report only 2% as proficient or above proficiency, for Adequate Yearly Progress (AYP) determinations.

The Mod-MSA assesses and reports student attainment on modified indicators and objectives in mathematics and reading content standards. In 2009, the test was administered concurrently with

[1] U.S. DOE rule published Monday, April 9, 2007, in the Federal Register as "Title I-Improving the Academic Achievement of the Disadvantaged; Individual of Disabilities Education Act, Final Rule."

the MSA to students in grades 6 through 8 while in 2010, grades 3 through 5 were also included in the Mod-MSA administration.

## 1.2. Rationale for the 2010 Mod-MSA: Reading

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

The State Curriculum, 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 State Curriculum has specified standards statements, topics, indicators, and objectives. Standards are broad, measurable statements of what students should know and be able to do. Topics, indicators, and objectives provide more specific content knowledge and skills that are unique at each grade level.

While 100% of the standards should be tested, it was not the case that every indicator would necessarily be tested each year for the Mod-MSA. Consequently, the State Curriculum has specified curricular indicators and objectives for the Mod-MSA that have contributed directly to measuring content standards that were aligned to the MSA.

By measuring students' achievement against the established academic standards, the 2010 Mod-MSA: Reading fulfills two main purposes. First, the Mod-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 Mod-MSA: Reading serves as an accountability tool to measure performance levels of individual students, schools, and districts against the new academic standards.

## 1.3. Eligibility Criteria for the Mod-MSA: Reading

Appendix F provides the criteria that were used for identifying students with disabilities for participation in the Mod-MSA.

## 1.4. Test Administration of the 2010 Mod-MSA: Reading

### 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 local education agency (LEA) was used to track the distribution and return of test books.

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

Test-related examiners manuals (EMs) were developed for the 2010 Mod-MSA: Reading 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 collaboratively by Pearson and MSDE. 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.

For the test examiner, Pearson provided the following materials:

- Examiner's Manual—Reading

For each student, Pearson provided the following materials:

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

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

- Two No. 2 pencils with erasers

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

- Sign for the door reading "Testing: Do not disturb"
- Digital clock or a watch, or clock with a second hand

**Test Administration Schedule**

The primary test window for Mod-MSA: Reading was established by MSDE (March 8–17, 2010, with online testing held March 8-23, and make-up testing held March 18–23, 2010). The test materials were delivered to schools (Examiner's Manuals, Test/Answer Books, and Test Coordinator's Kit) on or before February 22, 2010. However, each Local Education Agency (LEA) set a specific schedule for administration of the Mod-MSA: Reading within that window for their district. For a given grade and content area, all testing had to take place on the same schedule within each LEA. Each LEA schedule was submitted to MSDE in advance and approved by the state. For example, all grade 6 reading had to be administered on the same days throughout the LEA. In addition, each content area at each grade was tested on two days during the window.

The Mod-MSA: Reading schedule allowed approximately 2 1/2 hours for testing on each of the two days (including preparation time and breaks). Unless a student's IEP provided for extended time, students were required to submit their test books at the end of testing regardless whether they had answered all items. Unanswered items received a score of zero.

Students were allotted 10–52 minutes per test section. Shorter times were allotted for stand-alone or vocabulary-type items while the longer sections included 20 minutes to read each passage and answer 3 or 4 items relating to the passage.

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 Mod-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 main office of each school and giving proper notification that an MSDE monitor was in the building.

#### Student Participation

All students in grades 3 through 8 had to participate in the 2010 MSA or Mod-MSA. The only exception was that students with severe cognitive disabilities were assessed by the *Alternate Maryland School Assessment* (ALT-MSA) instead of either the regular MSA or Mod-MSA.

#### 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 designated as 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. All accommodations 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 Mod-MSA: Reading was administered to those requiring (1) large-print Student Test/Answer Books, (2) Braille Test Books, or (3) Kurzweil™ Test Forms on CD for a verbatim reading accommodation. For large-print and Braille Test Books, 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 returned with non-scoreable materials. Any Test/Answer Books used as source documents for transcription were invalidated by drawing a large slash across the student demographic page with a black permanent marker.

Once the student responses had been transcribed, the transcribed Test/Answer Book was returned for scoring with the standard-size materials. Specific packing instructions are provided in the 2009 TACM in section 4.

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

Students who had a verbatim reading accommodation documented in their Individual Education Plan (IEP), ELL Plan, or Section 504 Plan—and who received that accommodation in regular instruction—received the accommodation on the 2010 Mod-MSA. The accommodation was provided by a live reader or through technology. Section 1 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™ software, 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 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 certified TE direct supervision into the student's Test/Answer Book with a pre-printed or generic ID label attached.

A student whose IEP, 504 Plan, or ELL plan permitted the use of a word processor had his/her responses transcribed by hand or under the direct supervision of an eligible TE or STC exactly as the student entered his/her responses on the word processor. The student's responses were always transcribed at the school level into the student's Test/Answer Book with the pre-printed or generic ID label attached. After the student's responses had been transcribed, the word processor memory was cleared. The original word-processed printout was returned to Pearson with the non-scorable materials.

#### Test Format

All grade levels of the Mod-MSA: Reading used either a test book format, in which students wrote their answers directly in the test book, or used an online format which presented the test items on a computer screen and allowed students to select their answer choices by clicking on the corresponding answer bubble displayed onscreen. There was one form per grade of the Mod-MSA: Reading. .

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

**Security of Test Materials**

The following code of ethics conforms to the *Standards for Educational and Psychological Testing* developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education:

> It is breach of professional ethics for school personnel to provide verbal or nonverbal clues or answers, teach items on the test, share writing prompts, coach, hint, or in any way influence a student's performance during the testing situation. A breach of ethics may result in invalidation of test results and local education agency or MSDE disciplinary action. (Pearson, 2008, p. 13)

The Test/Answer Books for the 2010 Mod-MSA 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:

> 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. 13)

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

## 1.5. Quality Control Procedures

As a standard quality control procedure, Pearson created a test deck for the Mod-MSA program. The test deck began when Quality Assurance (QA) 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 QA Department. Pearson 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 ensure 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 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 and invalid multiple choice responses

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 file updates

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 Pearson's Quality Assurance Department. An issues log with sign-off approvals was utilized to ensure Pearson was addressing any issues that arose in the review of the test deck data across all functional groups at the company.

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

Pearson 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. The programs included, but were not limited to:

- Statewide duplicate students
- Statewide frequency distribution of demographic variables
- District/Building/N-counts
- Statewide raw to scale cut score tables
- The use of Proc Means to verify summary statistics
- Item response listing to verify all constructed responses were scored and were 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, Pearson posted the statewide student files to a secure FTP site for review by MSDE.

### 1.6. Item Bank Construction

The Mod-MSA: Reading item bank is constructed and maintained by Pearson in the form of computer files and paper copies. This enables test items to be readily available to both Pearson and MSDE staff for reference, test construction, test book design, and printing.

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

- Unique item number (UIN)
- Test administration year and season
- Test form
- Grade level
- Item type
- Item stem and options
- Passage code and title
- Subject code and description
- Process code and description
- Standard code and description
- Topic code and description
- Indicator code and description
- Objective code and description
- Item status
- Item statistics (Classical and Rasch)

# 2. TEST DESIGN, DEVELOPMENT, AND SCORING OF THE 2010 MOD-MSA: READING

## 2.1. Test Design and Structure of the 2010 Mod-MSA: Reading

In 2010, a single form in reading was created for each grade level from 3 through 8.

The 2010 Mod-MSA: Reading was administered as a single form, which included more than the required number of operational items. The 2010 administration included a mix of items (i.e., operational [OP] items on which the students will be scored and some field-test [FT] items). There were also some items that were used intact from the MSA item bank. The operational items were selected through the data review process only after the test was administered. All items were considered FT items until they were specified differently by the data review committee.

As shown in Table 2.1.1, the Mod-MSA items for reading, grades 3 to 5, were used intact from the MSA item bank or were augmented from the MSA items.

**Table 2.1.1 Test Design for the Mod-MSA, Reading, Grades 3-5**

| Grade | Item Type | Total # of Items | No. of Operational Items After Data Review | No. of Field-Test Items After Data Review |
|---|---|---|---|---|
| 3 | Intact from Previous MSA | 9 | 9 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items, and/or intact from Previous MSA | 60 (Includes four intact MSA items) | 36 | 24 |
| | **Total** | **69** | **45** | **24** |
| 4 | Intact from Previous MSA | 9 | 9 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items, and/or intact from Previous MSA | 59 (Includes three intact MSA items) | 36 | 23 |
| | **Total** | **68** | **45** | **23** |
| 5 | Intact from Previous MSA | 9 | 9 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items, and/or intact from Previous MSA | 58 (Includes two intact MSA items) | 36 | 22 |
| | **Total** | **67** | **45** | **22** |

Note: The total number of items is the sum of the operational and field-test items.

Table 2.1.2 shows that grades 6 to 8 included common items from the 2009 administration (together with newly created items for the Mod-MSA) to help place the 2010 Grades 6 to 8 tests on the established 2009 scale. Except for Grade 6, which had 23 common items, Grade 7 and 8 had 25 common items.

**Table 2.1.2 Test Design for the Mod-MSA, Reading (Grades 6-8)**

| Grade | Item Type | Total # of Items | No. of Operational Items After Data Review | No. of Field-Test Items After Data Review |
|---|---|---|---|---|
| 6 | Linking Items from 2009 | 23 | 23 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items | 37 | 22 | 15 |
| | **Total** | **60** | **45** | **15** |
| 7 | Linking Items from 2009 | 25 | 25 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items | 30 | 20 | 10 |
| | **Total** | **55** | **45** | **10** |
| 8 | Linking Items from 2009 | 25 | 25 | - |
| | Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items | 30 | 20 | 10 |
| | **Total** | **55** | **45** | **10** |

Note: The total number of items is the sum of the operational and field-test items.

## 2.2. Development and Review of the 2010 Mod-MSA: Reading

Developing the 2010 Mod-MSA: Reading was a complex process. It required a great deal of involvement from MSDE, Pearson, and local school systems. In addition, teachers, administrators, and content specialists from all over Maryland were recruited for different test development committees. These individuals reviewed test forms and items to ensure that they measured students' knowledge and skills fairly and without bias. Table 2.2.1 identifies which groups were responsible for developing the 2010 Mod-MSA: Reading.

**Table 2.2.1. Responsibility for the 2010 Mod-MSA Test Development**

| Development of the 2010 Mod-MSA | Primary Responsibility |
|---|---|
| Development of Preliminary Blueprints and Item Specifications | Pearson, MSDE, NPC |
| Development of Operational Form Requirements and Blueprint Session | MSDE |
| Item Writing | MSDE; Pearson |
| Item Review | Pearson; MSDE; NPC; Content Review Committee |
| Bias Review | Pearson; MSDE; Bias Review Committee |
| Vision Review | Pearson: MSDE; Vision Review Committee |
| Modification of Special Forms | Pearson; MSDE |
| Review of Special Forms | MSDE |
| Construction of Operational Test Forms | Pearson; MSDE; NPC |
| Construction of Field Test Forms | Pearson; MSDE |
| Review of Operational Test Forms | MSDE |
| Final Construction of Test Forms | Pearson; MSDE |

**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 Mod-MSA: Reading program. For example, they made recommendations to MSDE on issues such as test blueprints, operational form construction, field test design, item analysis, item selection for scoring purposes, linking, equating and scaling issues, and other relevant statistical and psychometric issues.

**Content Review Committee**

Content review committee members ensured that the Mod-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 on the reading tests. They looked for indications of bias that would affect the performance of an identifiable group of students on the test and across the mode of administration (i.e., online, and paper and pencil). Committee members discussed and, if necessary, rejected items based on gender, ethnic, religious, geographical, or mode of administration bias.

**Vision Review Committee**

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

## 2.3. Operational Form Structure of the 2010 Mod-MSA: Reading

The 2010 Mod-MSA: Reading only had selected response (SR) items (multiple-choice) with two or three distractors (aside from the correct choice) for each item. As shown in Table 2.2.1, there were 13, 12, and 11 vocabulary items each for grade 3, 4, and 5, respectively, that were used intact from the MSA. Also, there were 23 items in grade 6 and 25 items each in grade 7 and 8 that were used as common items from 2009 for linking the 2010 assessment to the established scale. The rest of the reading items were modified versions of the MSA items (grades 3-5) or were modified version of the MSA items plus newly developed items (grades 6-8). The intact items had three distractors, while the modified items had only two distractors. These items required students to select a correct answer from the various alternatives. The total number of scored items (operational) was selected so that the total score of the test matched that with the corresponding MSA examination both in terms of the total score and the scores of the different strands. Each SR item was scored dichotomously (i.e., 0 or 1).

The Mod-MSA: Reading test was to have the same number of items as the MSA test, i.e., 45 items for each of grades 3 through 8. These items reflected the same proportion of scores across the different subscales, i.e., the strands.

The Mod-MSA: Reading test was organized under the following content strands for each of the three grades (3–8):

1. Reading vocabulary (multiple meaning)
2. Reading vocabulary (words in context)
3. Literary
4. Informational

These strands were combined to match the same three strands as those reported by the reading MSA. For the Mod-MSA: Reading, therefore, the final reporting strands were:

1. General reading (combination of reading vocabulary, i.e., multiple meaning and words in context)
2. Literary
3. Informational

Table 2.3.1 provides the score for the reading operational tests based on the number of items used for each strand and grade level.

**Table 2.3.1. The 2010 Mod-MSA, Reading Operational Form with Maximum Points Obtainable Per Strand: Grades 6 to 8**

| Grade | Strand Title | No. of Items | Item Type | Reporting Strand | Reporting Score |
|---|---|---|---|---|---|
| 3 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading | 16 | *SR* | General | 16 |
| | Literary | 14 | *SR* | Literary | 14 |
| | Informational | 15 | *SR* | Informational | 15 |
| 4 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading | 15 | *SR* | General | 15 |
| | Literary | 15 | *SR* | Literary | 15 |
| | Informational | 15 | *SR* | Informational | 15 |
| 5 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading | 15 | *SR* | General | 15 |
| | Literary | 15 | *SR* | Literary | 15 |
| | Informational | 15 | *SR* | Informational | 15 |
| 6 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading | 15 | *SR* | General | 15 |
| | Literary | 15 | *SR* | Literary | 15 |
| | Informational | 15 | *SR* | Informational | 15 |
| 7 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading) | 15 | *SR* | General | 15 |
| | Literary | 15 | *SR* | Literary | 15 |
| | Informational | 15 | *SR* | Informational | 15 |
| 8 | Total Test | 45 | *SR* | Total Test | 45 |
| | General Reading | 16 | *SR* | General | 16 |
| | Literary | 14 | *SR* | Literary | 14 |
| | Informational | 15 | *SR* | Informational | 15 |

## 2.4. Item Analyses Undertaken Prior to the Creation of the Operational Forms

The 2010 Mod-MSA: Reading was administered as a single form, which included more than the required number of operational items. After administration of the form, operational items were selected during data review on the basis of their statistics and the number of items required for each strand of the operational test (see Table 2.3.1, above). All items not selected as operational were banked with their respective statistics as field test (FT) items. These items together with the 2009 operational items could be used as common linking items in 2011 in order to place the 2011 examinations on the established 2009/2010 scale.

The statistical considerations for the selection of operational items were based on the following:

- Classical item analysis
- Differential item functioning (DIF) analyses
- IRT analyses

All analyses provided in this report are (as indicated in the different tables) based either on the equating sample or the population. The equating sample was approximately 60 percent of the total population.

#### Classical Item Analyses

Classical item analyses included the calculation of p-values, the point-biserials, distractor-to-total correlations, and distractor frequency analysis.

Items were flagged for further scrutiny if:

- An item distractor was not selected by any students (i.e., nonfunctional distractor), or selected by a large number of high proficiency students, with low selection from other proficiency groupings (i.e., ambiguous distractor).
- An item *p*-value was less than .10 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.
- Omit rate was > 5%.

All items required a careful decision for inclusion in the operational form. For example, an item that was flagged as having a point-biserial < 0.10 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 was made for including it in the operational test form, i.e., learning the content was deemed a necessary factor for an item's inclusion in spite of it having poor statistics that were not related to miskeyed items.

#### 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.

During the item development period, prior to statistical analysis of DIF, all items were subjected to the scrutiny of the Bias Review Committee. As explained in Section 2.2, the Bias Review

Committee examined each reading item, looking for indications of bias that could impact the performance of an identifiable group of students. They discussed or rejected items biased on gender, ethnic, religious, or geographical bias.

After items were scored, statistical item analysis pertaining to DIF was undertaken. In this analysis, the gender reference group was males, and the ethnic reference group was white. The gender focal group was females and the ethnic focal group was black (African Americans). The total score of each operational form was used as the matching variable.

Since the 2010 Mod-MSA: Reading was a single-format examination comprising only SR items, the DIF procedure used consisted of the Mantel-Haenszel contingency procedure (Mantel & Haenszel, 1959) together with an effect-size approach[2] based on the delta scale (Camilli & Shepard, 1994).

### The Mantel Haenszel Chi-Square

The Mantel and Haenszel (1959) chi-square, which approximately follows a chi-square distribution with one degree of freedom, can be formulated as per the following (from Camilli & Shepard, 1994):

$$\text{MH } \chi^2 = \frac{\left\{\left|\sum_{j=1}^{S}\left[A_j - E(A_j)\right]\right| - 1/2\right\}^2}{\sum_{j=1}^{S} VAR(A_j)}, \text{ where}$$

$A_j$ and $E(A_j)$ are the observed number of correct responses and the expected number on the item, respectively for the reference group, while $VAR(A_j)$ is the variance associated with the observed score.

### The Delta Scale

The odds of a correct response are *P/Q* or *P/(1-P)*. The odds ratio, on the other hand, is simply the odds of a correct response of the reference group divided by the odds of a correct response of the focal group.

For a given item, the odds ratio is defined as follows:

$$\alpha_{M-H} = \frac{P_r / Q_r}{P_f / Qf}.$$

The corresponding null hypothesis is that the odds of getting the item correct are equal for the two groups. Thus, the odds ratio is equal to 1:

$$H_0\text{: } \alpha_{M-H} = \frac{P_r / Q_r}{P_f / Qf} = 1.$$

In order to calculate the delta scale, the Mantel and Haenszel (1959) log odds ratio was calculated using the following equation:

[2] For a detailed discussion on Mantel-Haenszel chi-square, the delta scale and ETS categories, please refer to Camilli and Shepard (1994).

$$\alpha_{MH} = \frac{\sum_{j=1}^{S} A_j D_j / T_j}{\sum_{j=1}^{S} B_j C_j / T_j}$$ , where

the various variables in the equation are from the following 2 x 2 contingency table for the *jth* total score on the test (Camilli & Shepard, 1994, p. 106).

Score on studied item with general notation

| | | 1 | 0 | Total |
|---|---|---|---|---|
| Group | R | $A_j$ | $B_j$ | $n_{Rj}$ |
| | F | $C_j$ | $D_j$ | nFj |
| | | $m_{1j}$ | $m_{0j}$ | $T_j$ |

The log odds ratio is a transformation of the odds ratio with its range being in the interval $-\infty$ to $+\infty$. The simple natural logarithm transformation of this odds ratio is symmetrical around zero, in which zero has the interpretation of equal odds. The odds ratio is transformed into a log odds ratio as per the following: $\beta_{M-H} = \ln(\alpha_{M\text{-}H})$. $\beta_{M-H}$, also has the advantage of being transformed linearly to other interval scale metrics (Camilli & Shepard, 1994). This fact is utilized in creating the delta scale (*D*), which is defined as $D = -2.35\beta_{M-H}$.

## DIF Classification

The *M-H* $\chi^2$ is examined in conjunction with the delta scale (*D*) to obtain DIF classifications depicted in Table 2.4.1, below.

**Table 2.4.1. DIF Classification**

| Category | Description | Criterion |
|---|---|---|
| A | No DIF | Non-significant *M-H* $\chi^2$ or $\|D\| < 1.0$ |
| B | Weak DIF | Significant *M-H* $\chi^2$ and $\|D\| < 1.5$ or |
| | | Non-significant *M-H* $\chi^2$ and $\|D\| \geq 1.0$ |
| C | Strong DIF | Significant *M-H* $\chi^2$ and $\|D\| \geq 1.5$ |

The groupings for the DIF analysis were based on matching students' scores on the Mod-MSA: Reading. Four proficiency groupings of the Mod-MSA students were formed at quarter intervals of the total Mod-MSA: Reading score. The performance on the Mod-MSA: Reading for the four proficiency-matched groups (gender, and ethnicity) was then compared for each item to evaluate potential differential performance by groups.

Items that were flagged as showing DIF (Category 'B', i.e., moderate DIF, and category 'C', i.e., extreme DIF) were subjected to further examination. For each of these items, experts judged whether the differential difficulty of the item was unfairly related to group membership based on the following guidelines:

- If the difficulty of the item was unfairly related to group membership, then the item should not be used at all.

- If the difficulty of the item was related to group membership, then the item should only be used if there was no other item matching the test alignment requirements presented in Appendix E.

All DIF results were stored in the Maryland item bank.

**Item Response Theory (IRT) Analyses**

Rasch fit statistics, infit and outfit (see Section 6.2) were used to examine model fit to the data. Items with fit indices < 0.5 or > 2.00 were flagged for misfit because, according to Linacre and Wright (1999), the inclusion of these items could be unproductive to the measurement system (< 0.5) or they could degrade the measurement system (> 2.0).

## 2.5. Items Flagged for Inspection Prior to the Creation of the Operational Forms

The following table provides content by grade summary with respect to the total number of items administered and the number of items that were flagged strictly on the basis of the statistics (classical, DIF and IRT) discussed above.

**Table 2.5.1. Summary Stats Used in the Development of the 2010 Mod-MSA: Reading Operational Form**

| Grade | Total # of Items | DIF Flag B (for check only) | DIF Flag C | PB Flag <= 0.10 but > 0[1] | PB Flag < = 0 (Cannot be used) | Items Rejected (C DIF + PB <=0 Flag | Items Used for Operational Form Building Based on Statistical Criteria | Items Needed for Each Operational Form |
|---|---|---|---|---|---|---|---|---|
| 3 | 69 | 2 | 1 | 5 | 0 | 1 | 68 | 45 |
| 4 | 68 | 5 | 1 | 3 | 1 | 2 | 66 | 45 |
| 5 | 61 | 4 | 1 | 6 | 3 | 4 | 57 | 45 |
| 6 | 60 | 10 | 1 | 6 | 0 | 1 | 59 | 45 |
| 7 | 55 | 8 | 2 | 2 | 1 | 3 | 52 | 45 |
| 8 | 55 | 4 | 0 | 4 | 1 | 1 | 54 | 45 |

Note: 1. Items in this column were generally not used unless a substitute could not be found for it.

As can be seen from the table, other than the point biserial (PB) and the DIF flags, all other statistical indices were well within the acceptable criteria. No items were flagged based on the fit analyses. For the PB we checked every item < 0.15 internally for the items being wrongly keyed. No such items were found across content and grade, even though some of the items had negative PBs.

## 2.6. Items Selected for the 2010 Operational Tests

As discussed above, the selection of items that were included in the final operational test forms of the 2010 Mod-MSA: Reading examination required a careful consideration based on test design, classical item analyses, DIF analyses, and IRT analyses. The general guidelines for the creation of the operational forms were as follows:

- Do not include items that are too easy or too hard.
- Do not include items with DIF classifications "C" for the SR 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 higher than 2.0.

Appendix A provides a list of item UIN numbers used to produce the operational form (the core items) from the total items administered in 2010.

Item level descriptive statistics (i.e., p-values and point biserials) are provided in Section 3.2 (Tables 3.2.1 to 3.2.6). Classical (i.e., p-values and point biserials) as well as IRT item analyses (i.e., Rasch difficulty and fit analysis) are provided in Section 5.3 (Tables 5.3.1 to 5.3.6). Appendix B provides item analysis by distractors while differential item functioning (DIF) analysis is provided in Appendix C.

As shown in Tables 3.2.1 to 3.2.6, there were several items across grades that had negative and zero point biserials. As explained earlier, these items were examined by content specialists for key and content accuracy, but no items were found that were wrongly keyed.

DIF analyses were conducted for gender and between white and African-Americans using the delta scale, D ($D = -2.35\log_e(\alpha_{MH})$, where $\log_e(\alpha_{MH})$ is the Mantel-Haenszel log odds ratio), in combination with the Mantel-Haenszel significant test of DIF detection (see Appendix C). Items with flags for moderate DIF (flag with B) were examined for bias. All items that were flagged as C were not included in the operational form. Only six items, one each in Grades 3, 4, 5, and 6, and two in Grade 7, had a DIF classification of 'C'.

The MSDE and Pearson worked collaboratively to select items for 2010 operational scoring and evaluate the psychometric properties of these operational item sets. In accordance with the NPC's recommendation, no items with negative point biserial correlations were selected for operational scoring. However, in spite of our intention of abiding by the terms of rejection outlined above, some items that had PBs less than 0.10 (but not negative or zero PBs), and items that were omitted by more than 5% of the students were included as operational items because of not having corresponding substitute items to use. None of the omit flags, however, were for more than 5.95% of the students.

**Table 2.6.1. Number of Items Included as Operational Items with 0 < PB < .10, and flagged Omit Items by Grades**

| Grade | 0 < PB < 0.10 | Omit Flag |
|---|---|---|
| 3 | 0 | 1 |
| 4 | 0 | 0 |
| 5 | 1 | 0 |
| 6 | 1 | 4 |
| 7 | 1 Common Item | 0 |
| 8 | 2 (includes 1 common item) | 0 |

As shown in Table 2.6.1 above, there was one common item in Grade 8 and Grade 7 that had 0 < PB < .10. It was used because no substitutes could be found to replace this item.

In Reading there were several items with four distractors (instead of three) that were used intact from the MSA. The number of such items by grades is provided in Table 2.6.2, below.

**Table 2.6.2. No of Intact MSA Items Used by Grades**

| Grade | No. of Intact Items Used | No. of Intact Items Used that were also Common Items |
|---|---|---|
| 3 | 9 | N/A |
| 4 | 9 | N/A |
| 5 | 9 | N/A |
| 6 | 9 | 9 |
| 7 | 5 | 5 |
| 8 | 5 | 5 |

## 2.7. Scoring Procedures of the 2010 Mod-MSA: Reading

Students' responses were machine-scored. Once received by Pearson, Test/Answer Books were scanned into an electronic imaging system so that the information necessary to score responses was captured and converted into an electronic format. Students' identification and demographic information, school information, and answers were converted to alphanumeric format.

After students' responses were scanned, the scoring key was applied to the captured item responses. Correct answers were assigned a score of one point. Incorrect answers, blank responses (omits), and responses with multiple marks were assigned a score of zero.

# 3. CENTRAL TENDENCY MEASURES AND CLASSICAL ITEM ANALYSES OF THE 2010 MOD-MSA: READING

This section provides central tendency statistics and results of classical statistical item analyses for students in grades 3 through 8.

The analyses provided for the central tendency measures of the operational test are based on the statewide population. However, the item level statistics are based on the equating sample mainly because those are the item statistics that were used in selecting the operational items and are stored in the item bank. In the absence of a table note, the analyses in this report will be assumed to be based on the equating sample.

## 3.1. Measures of Central Tendency

The classical measures of central tendency, variability, and score precision of raw scores are presented in Table 3.2.1 by grades for each strand as well as the total operational test. The tables include the following:

- Number of items by strands and the total test
- Maximum score attainable
- N-Count (sample size)
- Mean (average raw score)
- SD (standard deviation)
- SE (standard error of the mean)

**Standard Error of the Mean**

The standard error of the mean (SE) is an estimate of the magnitude of sampling error associated in the estimation of the population mean. It is defined as follows:

$$SE=\frac{\hat{\sigma}}{\sqrt{n}}, \text{ where}$$

SE = standard error of the mean
$\hat{\sigma}$ = standard deviation of the sample
n = number of responses in the sample

**Table 3.1.1. Central Tendency Statistics of the Operational Test by Subscales and Grades**

| Grade | Modality (Strand) | # Items | Max Points | N Count | RS Mean | SD | SE |
|---|---|---|---|---|---|---|---|
| 3 | General Reading | 16 | 16 | 1058 | 9.76 | 2.94 | 0.09 |
| | Literary | 14 | 14 | 1058 | 8.37 | 2.81 | 0.09 |
| | Informational | 15 | 15 | 1058 | 7.74 | 2.90 | 0.09 |
| | Total Test | 45 | 45 | 1058 | 25.87 | 7.36 | 0.23 |
| 4 | General Reading | 15 | 15 | 1335 | 9.42 | 2.82 | 0.08 |
| | Literary | 15 | 15 | 1335 | 7.80 | 2.76 | 0.08 |
| | Informational | 15 | 15 | 1335 | 7.61 | 2.86 | 0.08 |
| | Total Test | 45 | 45 | 1335 | 24.83 | 7.04 | 0.19 |
| 5 | General Reading | 15 | 15 | 1459 | 10.21 | 2.97 | 0.08 |
| | Literary | 15 | 15 | 1459 | 7.83 | 2.66 | 0.07 |
| | Informational | 15 | 15 | 1459 | 6.65 | 2.56 | 0.07 |
| | Total Test | 45 | 45 | 1459 | 26.68 | 6.71 | 0.16 |
| 6 | General Reading | 15 | 15 | 1356 | 9.05 | 2.92 | 0.08 |
| | Literary | 15 | 15 | 1356 | 8.33 | 2.73 | 0.07 |
| | Informational | 15 | 15 | 1356 | 7.84 | 2.73 | 0.07 |
| | Total Test | 45 | 45 | 1356 | 25.22 | 6.92 | 0.19 |
| 7 | General Reading | 15 | 15 | 1702 | 9.83 | 2.47 | 0.06 |
| | Literary | 15 | 15 | 1702 | 8.98 | 2.74 | 0.07 |
| | Informational | 15 | 15 | 1702 | 7.16 | 2.79 | 0.07 |
| | Total Test | 45 | 45 | 1702 | 25.97 | 6.54 | 0.16 |
| 8 | General Reading | 15 | 16 | 1969 | 9.75 | 2.87 | 0.06 |
| | Literary | 15 | 14 | 1969 | 8.77 | 2.77 | 0.06 |
| | Informational | 15 | 15 | 1969 | 8.71 | 2.65 | 0.06 |
| | Total Test | 45 | 45 | 1969 | 27.23 | 6.89 | 0.16 |

Note: Analyses were conducted with the statewide population after applying equating exclusion criteria.

## 3.2. Item-Level Descriptive Statistics

This section presents the raw score summary statistics for all items in the Mod-MSA: Reading spring 2010 within the framework of classical test theory (CTT). The *p*-value for each item is defined as the proportion of students that answer an item correctly for the multiple choice items. A high *p*-value means that an item is easy; a low *p*-value means that an item is difficult.

The point biserial correlation for each item is an index of the association between the item score and the total test score. It shows the ability of the item to discriminate between low proficiency and high proficiency students. An item with a high point biserial correlation discriminates more effectively between the low and the high proficiency students than a low point biserial correlation.

The item-level statistics for the operational and the embedded field test items for the 2010 Mod-MSA: Reading based on the equating sample are presented in Table 3.2.1 to Table 3.2.6. The following item information and statistics are presented for each item:

- Item number based on their sequential appearance in the form.

- Item UIN number
- Item strand
- N-count (number of students)
- Response options
- *P*-value for multiple choice items (percentage of examinees that answered the item correctly)
- Point biserial (index of discrimination between high and low scoring students)

Item distractor analyses are provided in Appendix B. These analyses include

- Item UIN number
- Item strand number
- Item mean
- Item standard deviation
- Percentage of distribution by distractors
- Percent of omits
- Distractor-to-total correlation.

**Table 3.2.1 The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 3**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213631 | 208 | 813 | 4 | 0.73 | 0.32 |
| 2 | 100000360181 | 208 | 813 | 4 | 0.29 | 0.26 |
| 3 | 100000213633 | 208 | 813 | 4 | 0.80 | 0.48 |
| 4 | 100000213634 | 208 | 813 | 4 | 0.38 | 0.34 |
| 5 | 100000360182 | 208 | 813 | 4 | 0.42 | 0.32 |
| 6 | 100000101512 | 208 | 813 | 3 | 0.51 | 0.25 |
| 7 | 100000101513 | 208 | 813 | 3 | 0.56 | 0.25 |
| 8 | 100000101514 | 208 | 813 | 3 | 0.62 | 0.38 |
| 9 | 100000360183 | 208 | 813 | 3 | 0.58 | 0.24 |
| 10 | 100000101516 | 208 | 813 | 3 | 0.42 | 0.21 |
| 11 | 100000101517 | 208 | 813 | 3 | 0.32 | 0.20 |
| 12 | 100000101518 | 208 | 813 | 3 | 0.56 | 0.20 |
| 13 | 100000360184 | 208 | 813 | 3 | 0.48 | 0.23 |
| 14 | 100000101969 | 208 | 813 | 3 | 0.69 | 0.46 |
| 15 | 100000101970 | 208 | 813 | 3 | 0.50 | 0.36 |
| 16 | 100000101971 | 209 | 813 | 3 | 0.40 | 0.22 |
| 17 | 100000101972 | 209 | 813 | 3 | 0.46 | 0.14 |
| 18 | 100000101973 | 209 | 813 | 3 | 0.37 | 0.17 |
| 19 | 100000101974 | 209 | 813 | 3 | 0.43 | 0.32 |
| 20 | 100000101975 | 209 | 813 | 3 | 0.45 | 0.15 |
| 21 | 100000101932 | 208 | 813 | 3 | 0.59 | 0.27 |
| 22 | 100000101934 | 209 | 813 | 3 | 0.52 | 0.31 |
| 23 | 100000101935 | 209 | 813 | 3 | 0.40 | 0.07 |
| 24 | 100000101933 | 208 | 813 | 3 | 0.37 | 0.23 |
| 25 | 100000101936 | 209 | 813 | 3 | 0.52 | 0.21 |
| 26 | 100000101937 | 209 | 813 | 3 | 0.63 | 0.35 |
| 27 | 100000101938 | 209 | 813 | 3 | 0.47 | 0.29 |
| 28 | 100000346452 | 208 | 813 | 3 | 0.60 | 0.29 |
| 29 | 100000346453 | 208 | 813 | 3 | 0.59 | 0.37 |
| 30 | 100000346444 | 209 | 813 | 3 | 0.44 | 0.26 |
| 31 | 100000346446 | 209 | 813 | 3 | 0.39 | 0.08 |
| 32 | 100000346450 | 209 | 813 | 3 | 0.74 | 0.40 |
| 33 | 100000346445 | 209 | 813 | 3 | 0.46 | 0.38 |
| 34 | 100000346448 | 209 | 813 | 3 | 0.49 | 0.27 |
| 35 | 100000260458 | 209 | 813 | 3 | 0.42 | 0.18 |
| 36 | 100000260460 | 209 | 813 | 3 | 0.59 | 0.21 |
| 37 | 100000260461 | 209 | 813 | 3 | 0.56 | 0.30 |
| 38 | 100000312661 | 208 | 813 | 3 | 0.58 | 0.27 |
| 39 | 100000260457 | 209 | 813 | 3 | 0.51 | 0.20 |
| 40 | 100000260459 | 209 | 813 | 3 | 0.37 | 0.19 |
| 41 | 100000260465 | 208 | 813 | 3 | 0.58 | 0.26 |
| 42 | | | 813 | 3 | | |
| | 100000101905 | 208 | | | 0.74 | 0.36 |

Table 3.2.1 The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 3 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 43 | 100000101906 | 208 | 813 | 3 | 0.40 | 0.28 |
| 44 | 100000101907 | 212 | 813 | 3 | 0.48 | 0.32 |
| 45 | 100000101908 | 212 | 813 | 3 | 0.71 | 0.47 |
| 46 | 100000101909 | 212 | 813 | 3 | 0.58 | 0.38 |
| 47 | 100000101910 | 212 | 813 | 3 | 0.56 | 0.05 |
| 48 | 100000101911 | 212 | 813 | 3 | 0.70 | 0.37 |
| 49 | 100000101529 | 212 | 813 | 3 | 0.83 | 0.38 |
| 50 | 100000101530 | 212 | 813 | 3 | 0.65 | 0.43 |
| 51 | 100000101531 | 212 | 813 | 3 | 0.24 | 0.06 |
| 52 | 100000101532 | 212 | 813 | 3 | 0.46 | 0.13 |
| 53 | 100000101527 | 208 | 813 | 3 | 0.60 | 0.33 |
| 54 | 100000101528 | 208 | 813 | 3 | 0.53 | 0.34 |
| 55 | 100000101533 | 212 | 813 | 3 | 0.69 | 0.31 |
| 56 | 100000260338 | 208 | 813 | 3 | 0.65 | 0.22 |
| 57 | 100000260342 | 209 | 813 | 3 | 0.51 | 0.30 |
| 58 | 100000260345 | 212 | 813 | 3 | 0.32 | 0.14 |
| 59 | 100000365154 | 212 | 813 | 3 | 0.44 | 0.26 |
| 60 | 100000260349 | 209 | 813 | 3 | 0.23 | 0.15 |
| 61 | 100000260339 | 212 | 813 | 3 | 0.57 | 0.35 |
| 62 | 100000300707 | 208 | 813 | 3 | 0.40 | 0.20 |
| 63 | 100000260364 | 212 | 813 | 3 | 0.38 | 0.04 |
| 64 | 100000260368 | 212 | 813 | 3 | 0.67 | 0.34 |
| 65 | 100000260371 | 212 | 813 | 3 | 0.51 | 0.16 |
| 66 | 100000300557 | 212 | 813 | 3 | 0.39 | 0.28 |
| 67 | 100000260372 | 208 | 813 | 3 | 0.74 | 0.44 |
| 68 | 100000260365 | 212 | 813 | 3 | 0.73 | 0.35 |
| 69 | 100000300552 | 208 | 813 | 3 | 0.65 | 0.47 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria.

**Table 3.2.2. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 4**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213644 | 208 | 967 | 3 | 0.36 | 0.14 |
| 2 | 100000213645 | 208 | 967 | 3 | 0.50 | 0.14 |
| 3 | 100000213646 | 208 | 967 | 3 | 0.47 | 0.12 |
| 4 | 100000213647 | 208 | 967 | 3 | 0.59 | 0.33 |
| 5 | 100000360190 | 208 | 967 | 3 | 0.23 | 0.12 |
| 6 | 100000213637 | 208 | 967 | 4 | 0.69 | 0.39 |
| 7 | 100000213638 | 208 | 967 | 4 | 0.69 | 0.38 |
| 8 | 100000213639 | 208 | 967 | 4 | 0.36 | 0.12 |
| 9 | 100000360191 | 208 | 967 | 4 | 0.38 | 0.25 |
| 10 | 100000213641 | 208 | 967 | 4 | 0.78 | 0.36 |
| 11 | 100000213642 | 208 | 967 | 4 | 0.66 | 0.40 |
| 12 | 100000360192 | 208 | 967 | 4 | 0.73 | 0.39 |
| 13 | 100000357134 | 209 | 967 | 3 | 0.67 | 0.33 |
| 14 | 100000357132 | 208 | 967 | 3 | 0.45 | 0.18 |
| 15 | 100000357133 | 208 | 967 | 3 | 0.74 | 0.37 |
| 16 | 100000357135 | 209 | 967 | 3 | 0.45 | 0.10 |
| 17 | 100000357136 | 209 | 967 | 3 | 0.57 | 0.36 |
| 18 | 100000357137 | 209 | 967 | 3 | 0.56 | 0.36 |
| 19 | 100000357138 | 209 | 967 | 3 | 0.41 | 0.27 |
| 20 | 100000357106 | 209 | 967 | 3 | 0.52 | 0.16 |
| 21 | 100000357104 | 208 | 967 | 3 | 0.65 | 0.37 |
| 22 | 100000357107 | 209 | 967 | 3 | 0.61 | 0.34 |
| 23 | 100000357105 | 208 | 967 | 3 | 0.63 | 0.31 |
| 24 | 100000357108 | 209 | 967 | 3 | 0.49 | 0.28 |
| 25 | 100000357109 | 209 | 967 | 3 | 0.73 | 0.36 |
| 26 | 100000357110 | 209 | 967 | 3 | 0.59 | 0.17 |
| 27 | 100000260487 | 209 | 967 | 3 | 0.41 | 0.07 |
| 28 | 100000260486 | 209 | 967 | 3 | 0.25 | 0.28 |
| 29 | 100000260483 | 208 | 967 | 3 | 0.57 | 0.27 |
| 30 | 100000260481 | 208 | 967 | 3 | 0.50 | 0.35 |
| 31 | 100000260489 | 209 | 967 | 3 | 0.64 | 0.26 |
| 32 | 100000260488 | 209 | 967 | 3 | 0.26 | 0.25 |
| 33 | 100000260492 | 209 | 967 | 3 | 0.28 | 0.16 |
| 34 | 100000269896 | 209 | 967 | 3 | 0.55 | 0.31 |
| 35 | 100000271197 | 208 | 967 | 3 | 0.38 | 0.35 |
| 36 | 100000271198 | 208 | 967 | 3 | 0.57 | 0.33 |
| 37 | 100000269900 | 209 | 967 | 3 | 0.33 | 0.15 |
| 38 | 100000269897 | 209 | 967 | 3 | 0.46 | 0.19 |
| 39 | 100000269898 | 209 | 967 | 3 | 0.41 | 0.26 |
| 40 | 100000269899 | 209 | 967 | 3 | 0.42 | 0.25 |
| 41 | 100000357097 | 212 | 967 | 3 | 0.27 | 0.08 |
| 42 | 100000357095 | 208 | 967 | 3 | 0.60 | 0.28 |

Table 3.2.2. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 4 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 43 | 100000357098 | 212 | 967 | 3 | 0.33 | 0.23 |
| 44 | 100000462160 | 212 | 967 | 3 | 0.31 | 0.00 |
| 45 | 100000357100 | 212 | 967 | 3 | 0.40 | 0.13 |
| 46 | 100000357101 | 212 | 967 | 3 | 0.59 | 0.26 |
| 47 | 100000462157 | 208 | 967 | 3 | 0.58 | 0.33 |
| 48 | 100000101997 | 208 | 967 | 3 | 0.52 | 0.21 |
| 49 | 100000101996 | 212 | 967 | 3 | 0.43 | 0.33 |
| 50 | 100000101998 | 212 | 967 | 3 | 0.44 | 0.21 |
| 51 | 100000101999 | 212 | 967 | 3 | 0.42 | 0.24 |
| 52 | 100000102000 | 212 | 967 | 3 | 0.52 | 0.22 |
| 53 | 100000102001 | 212 | 967 | 3 | 0.49 | 0.29 |
| 54 | 100000200070 | 208 | 967 | 3 | 0.51 | 0.30 |
| 55 | 100000102025 | 208 | 967 | 3 | 0.30 | 0.14 |
| 56 | 100000102026 | 208 | 967 | 3 | 0.44 | 0.15 |
| 57 | 100000102027 | 212 | 967 | 3 | 0.53 | 0.36 |
| 58 | 100000102023 | 212 | 967 | 3 | 0.47 | 0.27 |
| 59 | 100000102024 | 212 | 967 | 3 | 0.69 | 0.21 |
| 60 | 100000102028 | 212 | 967 | 3 | 0.46 | 0.34 |
| 61 | 100000102029 | 212 | 967 | 3 | 0.58 | 0.31 |
| 62 | 100000301035 | 208 | 967 | 3 | 0.67 | 0.39 |
| 63 | 100000267470 | 208 | 967 | 3 | 0.63 | 0.38 |
| 64 | 100000267467 | 212 | 967 | 3 | 0.85 | 0.34 |
| 65 | 100000301034 | 212 | 967 | 3 | 0.69 | 0.32 |
| 66 | 100000267473 | 212 | 967 | 3 | 0.50 | 0.20 |
| 67 | 100000267474 | 212 | 967 | 3 | 0.53 | 0.06 |
| 68 | 100000267472 | 212 | 967 | 3 | 0.61 | 0.28 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria.

**Table 3.2.3. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 5**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213650 | 208 | 1043 | 4 | 0.67 | 0.30 |
| 2 | 100000213651 | 208 | 1043 | 4 | 0.69 | 0.32 |
| 3 | 100000213652 | 208 | 1043 | 4 | 0.68 | 0.28 |
| 4 | 100000213653 | 208 | 1043 | 4 | 0.73 | 0.37 |
| 5 | 100000360196 | 208 | 1043 | 4 | 0.38 | 0.16 |
| 6 | 100000213655 | 208 | 1043 | 4 | 0.73 | 0.31 |
| 7 | 100000213656 | 208 | 1043 | 4 | 0.78 | 0.41 |
| 8 | 100000213657 | 208 | 1043 | 4 | 0.69 | 0.39 |
| 9 | 100000213658 | 208 | 1043 | 4 | 0.49 | 0.08 |
| 10 | 100000213659 | 208 | 1043 | 4 | 0.65 | 0.29 |
| 11 | 100000360197 | 208 | 1043 | 4 | 0.62 | 0.31 |
| 12 | 100000102095 | 209 | 1043 | 3 | 0.51 | 0.24 |
| 13 | 100000102093 | 208 | 1043 | 3 | 0.73 | 0.33 |
| 14 | 100000102094 | 208 | 1043 | 3 | 0.32 | 0.23 |
| 15 | 100000102096 | 209 | 1043 | 3 | 0.50 | 0.20 |
| 16 | 100000102097 | 209 | 1043 | 3 | 0.37 | 0.05 |
| 17 | 100000102098 | 209 | 1043 | 3 | 0.37 | 0.11 |
| 18 | 100000102099 | 209 | 1043 | 3 | 0.34 | 0.13 |
| 19 | 100000102113 | 209 | 1043 | 3 | 0.51 | 0.35 |
| 20 | 100000102111 | 208 | 1043 | 3 | 0.70 | 0.35 |
| 21 | 100000102114 | 209 | 1043 | 3 | 0.51 | 0.30 |
| 22 | 100000102115 | 209 | 1043 | 3 | 0.34 | 0.25 |
| 23 | 100000102116 | 209 | 1043 | 3 | 0.50 | 0.28 |
| 24 | 100000102112 | 208 | 1043 | 3 | 0.72 | 0.27 |
| 25 | 100000102117 | 209 | 1043 | 3 | 0.35 | 0.05 |
| 26 | 100000102104 | 209 | 1043 | 3 | 0.53 | 0.23 |
| 27 | 100000102102 | 208 | 1043 | 3 | 0.60 | 0.21 |
| 28 | 100000102103 | 208 | 1043 | 3 | 0.61 | 0.23 |
| 29 | 100000102105 | 209 | 1043 | 3 | 0.32 | 0.10 |
| 30 | 100000102106 | 209 | 1043 | 3 | 0.65 | 0.28 |
| 31 | 100000102107 | 209 | 1043 | 3 | 0.33 | 0.22 |
| 32 | 100000102108 | 209 | 1043 | 3 | 0.40 | 0.20 |
| 33 | 100000102086 | 209 | 1043 | 3 | 0.37 | 0.05 |
| 34 | 100000102084 | 208 | 1043 | 3 | 0.62 | 0.23 |
| 35 | 100000102087 | 209 | 1043 | 3 | 0.42 | 0.13 |
| 36 | 100000102088 | 209 | 1043 | 3 | 0.30 | 0.06 |
| 37 | 100000102089 | 209 | 1043 | 3 | 0.23 | -0.03 |
| 38 | 100000102085 | 208 | 1043 | 3 | 0.55 | 0.15 |
| 39 | 100000102090 | 209 | 1043 | 3 | 0.38 | 0.24 |
| 40 | 100000102046 | 208 | 1043 | 3 | 0.78 | 0.40 |
| 41 | 100000102048 | 212 | 1043 | 3 | 0.55 | 0.18 |

Table 3.2.3. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 5 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 42 | 100000102049 | 212 | 1043 | 3 | 0.33 | 0.06 |
| 43 | 100000102050 | 212 | 1043 | 3 | 0.70 | 0.24 |
| 44 | 100000102051 | 212 | 1043 | 3 | 0.47 | 0.16 |
| 45 | 100000102052 | 208 | 1043 | 3 | 0.63 | 0.28 |
| 46 | 100000102047 | 212 | 1043 | 3 | 0.61 | 0.35 |
| 47 | 100000102056 | 208 | 1043 | 3 | 0.47 | 0.17 |
| 48 | 100000102058 | 212 | 1043 | 3 | 0.18 | -0.04 |
| 49 | 100000102059 | 212 | 1043 | 3 | 0.60 | 0.26 |
| 50 | 100000102060 | 212 | 1043 | 3 | 0.36 | 0.20 |
| 51 | 100000102061 | 212 | 1043 | 3 | 0.50 | 0.23 |
| 52 | 100000102057 | 208 | 1043 | 3 | 0.39 | 0.13 |
| 53 | 100000102062 | 212 | 1043 | 3 | 0.27 | 0.04 |
| 54 | 100000102068 | 212 | 1043 | 3 | 0.58 | 0.28 |
| 55 | 100000102069 | 212 | 1043 | 3 | 0.37 | 0.31 |
| 56 | 100000102066 | 208 | 1043 | 3 | 0.54 | 0.18 |
| 57 | 100000102067 | 208 | 1043 | 3 | 0.67 | 0.34 |
| 58 | 100000102070 | 212 | 1043 | 3 | 0.24 | -0.07 |
| 59 | 100000102071 | 212 | 1043 | 3 | 0.47 | 0.13 |
| 60 | 100000102072 | 212 | 1043 | 3 | 0.62 | 0.36 |
| 61 | 100000267477 | 212 | 1043 | 3 | 0.47 | 0.25 |
| 62 | 100000268378 | 212 | 1043 | 3 | 0.30 | 0.25 |
| 63 | 100000267485 | 208 | 1043 | 3 | 0.38 | 0.29 |
| 64 | 100000268380 | 212 | 1043 | 3 | 0.43 | 0.28 |
| 65 | 100000267481 | 212 | 1043 | 3 | 0.29 | 0.16 |
| 66 | 100000301339 | 208 | 1043 | 3 | 0.47 | 0.21 |
| 67 | 100000303033 | 212 | 1043 | 3 | 0.59 | 0.37 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria.

**Table 3.2.4 The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 6**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213662 | 208 | 975 | 4 | 0.69 | 0.41 |
| 2 | 100000213663 | 208 | 975 | 4 | 0.76 | 0.47 |
| 3 | 100000213664 | 208 | 975 | 4 | 0.54 | 0.28 |
| 4 | 100000213665 | 208 | 975 | 4 | 0.65 | 0.40 |
| 5 | 100000213666 | 208 | 975 | 4 | 0.57 | 0.17 |
| 6 | 100000360201 | 208 | 975 | 4 | 0.19 | 0.12 |
| 7 | 100000213668 | 208 | 975 | 4 | 0.68 | 0.37 |
| 8 | 100000213669 | 208 | 975 | 4 | 0.25 | 0.18 |
| 9 | 100000213670 | 208 | 975 | 4 | 0.77 | 0.34 |
| 10 | 100000213671 | 208 | 975 | 4 | 0.52 | 0.28 |
| 11 | 100000360202 | 208 | 975 | 4 | 0.69 | 0.36 |
| 12 | 100000102192 | 209 | 975 | 3 | 0.60 | 0.42 |
| 13 | 100000102194 | 209 | 975 | 3 | 0.54 | 0.21 |
| 14 | 100000102191 | 208 | 975 | 3 | 0.68 | 0.40 |
| 15 | 100000102195 | 209 | 975 | 3 | 0.61 | 0.26 |
| 16 | 100000102196 | 209 | 975 | 3 | 0.44 | 0.23 |
| 17 | 100000102174 | 209 | 975 | 3 | 0.51 | 0.14 |
| 18 | 100000102175 | 209 | 975 | 3 | 0.33 | 0.15 |
| 19 | 100000102173 | 208 | 975 | 3 | 0.64 | 0.33 |
| 20 | 100000102176 | 209 | 975 | 3 | 0.43 | 0.26 |
| 21 | 100000102178 | 209 | 975 | 3 | 0.51 | 0.19 |
| 22 | 100000269999 | 208 | 975 | 3 | 0.86 | 0.37 |
| 23 | 100000269997 | 209 | 975 | 3 | 0.56 | 0.25 |
| 24 | 100000270006 | 209 | 975 | 3 | 0.71 | 0.40 |
| 25 | 100000269998 | 209 | 975 | 3 | 0.44 | 0.17 |
| 26 | 100000270000 | 208 | 975 | 3 | 0.69 | 0.35 |
| 27 | 100000270005 | 209 | 975 | 3 | 0.66 | 0.38 |
| 28 | 100000270007 | 209 | 975 | 3 | 0.45 | 0.25 |
| 29 | 100000270016 | 208 | 975 | 3 | 0.30 | 0.07 |
| 30 | 100000270018 | 208 | 975 | 3 | 0.45 | 0.12 |
| 31 | 100000270021 | 209 | 975 | 3 | 0.39 | 0.17 |
| 32 | 100000270022 | 209 | 975 | 3 | 0.47 | 0.02 |
| 33 | 100000270023 | 209 | 975 | 3 | 0.41 | 0.07 |
| 34 | 100000270013 | 209 | 975 | 3 | 0.41 | 0.23 |
| 35 | 100000270020 | 209 | 975 | 3 | 0.44 | 0.28 |
| 36 | 100000257087 | 208 | 975 | 3 | 0.75 | 0.33 |
| 37 | 100000257088 | 208 | 975 | 3 | 0.58 | 0.40 |
| 38 | 100000257113 | 212 | 975 | 3 | 0.36 | 0.05 |
| 39 | 100000257114 | 212 | 975 | 3 | 0.42 | 0.18 |
| 40 | 100000257118 | 212 | 975 | 3 | 0.67 | 0.31 |
| 41 | 100000257119 | 212 | 975 | 3 | 0.62 | 0.36 |

Table 3.2.4. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 6 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 42 | 100000257120 | 212 | 975 | 3 | 0.42 | 0.17 |
| 43 | 100000257102 | 212 | 975 | 3 | 0.52 | 0.36 |
| 44 | 100000257072 | 208 | 975 | 3 | 0.34 | 0.08 |
| 45 | 100000257103 | 212 | 975 | 3 | 0.44 | 0.15 |
| 46 | 100000257104 | 212 | 975 | 3 | 0.50 | 0.21 |
| 47 | 100000257071 | 208 | 975 | 3 | 0.59 | 0.28 |
| 48 | 100000257100 | 212 | 975 | 3 | 0.55 | 0.34 |
| 49 | 100000257101 | 212 | 975 | 3 | 0.30 | 0.12 |
| 50 | 100000102136 | 212 | 975 | 3 | 0.70 | 0.29 |
| 51 | 100000102137 | 212 | 975 | 3 | 0.54 | 0.18 |
| 52 | 100000102139 | 212 | 975 | 3 | 0.60 | 0.08 |
| 53 | 100000102140 | 212 | 975 | 3 | 0.55 | 0.16 |
| 54 | 100000257081 | 208 | 975 | 3 | 0.74 | 0.35 |
| 55 | 100000257110 | 212 | 975 | 3 | 0.60 | 0.34 |
| 56 | 100000257083 | 208 | 975 | 3 | 0.46 | 0.39 |
| 57 | 100000257106 | 212 | 975 | 3 | 0.59 | 0.41 |
| 58 | 100000257107 | 212 | 975 | 3 | 0.39 | 0.28 |
| 59 | 100000257109 | 212 | 975 | 3 | 0.56 | 0.45 |
| 60 | 100000257112 | 212 | 975 | 3 | 0.83 | 0.49 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria.

**Table 3.2.5. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 7**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213674 | 208 | 1158 | 4 | 0.67 | 0.42 |
| 2 | 100000213675 | 208 | 1158 | 4 | 0.78 | 0.34 |
| 3 | 100000213676 | 208 | 1158 | 4 | 0.35 | 0.23 |
| 4 | 100000360204 | 208 | 1158 | 4 | 0.29 | 0.17 |
| 5 | 100000213678 | 208 | 1158 | 4 | 0.74 | 0.33 |
| 6 | 100000213679 | 208 | 1158 | 4 | 0.68 | 0.40 |
| 7 | 100000360205 | 208 | 1158 | 4 | 0.35 | 0.33 |
| 8 | 100000270578 | 208 | 1158 | 3 | 0.62 | 0.28 |
| 9 | 100000270572 | 209 | 1158 | 3 | 0.52 | 0.13 |
| 10 | 100000270574 | 209 | 1158 | 3 | 0.42 | 0.29 |
| 11 | 100000322011 | 209 | 1158 | 3 | 0.62 | 0.29 |
| 12 | 100000270580 | 208 | 1158 | 3 | 0.81 | 0.41 |
| 13 | 100000270571 | 209 | 1158 | 3 | 0.38 | 0.18 |
| 14 | 100000270575 | 209 | 1158 | 3 | 0.70 | 0.43 |
| 15 | 100000270109 | 209 | 1158 | 3 | 0.42 | 0.25 |
| 16 | 100000270112 | 209 | 1158 | 3 | 0.39 | 0.11 |
| 17 | 100000270113 | 208 | 1158 | 3 | 0.52 | 0.22 |
| 18 | 100000270114 | 208 | 1158 | 3 | 0.56 | 0.25 |
| 19 | 100000270111 | 209 | 1158 | 3 | 0.41 | -0.01 |
| 20 | 100000270107 | 209 | 1158 | 3 | 0.34 | 0.07 |
| 21 | 100000270121 | 209 | 1158 | 3 | 0.56 | 0.30 |
| 22 | 100000102276 | 209 | 1158 | 3 | 0.48 | 0.25 |
| 23 | 100000102277 | 209 | 1158 | 3 | 0.34 | 0.20 |
| 24 | 100000102274 | 208 | 1158 | 3 | 0.41 | 0.13 |
| 25 | 100000102278 | 209 | 1158 | 3 | 0.46 | 0.31 |
| 26 | 100000102279 | 209 | 1158 | 3 | 0.41 | 0.24 |
| 27 | 100000102263 | 208 | 1158 | 3 | 0.58 | 0.25 |
| 28 | 100000102267 | 209 | 1158 | 3 | 0.43 | 0.08 |
| 29 | 100000102268 | 209 | 1158 | 3 | 0.50 | 0.36 |
| 30 | 100000102269 | 209 | 1158 | 3 | 0.42 | 0.22 |
| 31 | 100000102266 | 209 | 1158 | 3 | 0.33 | 0.14 |
| 32 | 100000257181 | 208 | 1158 | 3 | 0.65 | 0.14 |
| 33 | 100000257182 | 208 | 1158 | 3 | 0.79 | 0.34 |
| 34 | 100000257216 | 212 | 1158 | 3 | 0.69 | 0.36 |
| 35 | 100000257218 | 212 | 1158 | 3 | 0.76 | 0.39 |
| 36 | 100000257219 | 212 | 1158 | 3 | 0.47 | 0.13 |
| 37 | 100000257215 | 212 | 1158 | 3 | 0.87 | 0.26 |
| 38 | 100000257221 | 212 | 1158 | 3 | 0.53 | 0.20 |
| 39 | 100000102244 | 208 | 1158 | 3 | 0.84 | 0.41 |
| 40 | 100000102246 | 212 | 1158 | 3 | 0.58 | 0.20 |
| 41 | 100000102247 | 212 | 1158 | 3 | 0.75 | 0.47 |

Table 3.2.5. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 7 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 42 | 100000102248 | 212 | 1158 | 3 | 0.51 | 0.15 |
| 43 | 100000102245 | 208 | 1158 | 3 | 0.80 | 0.34 |
| 44 | 100000102218 | 212 | 1158 | 3 | 0.50 | 0.35 |
| 45 | 100000102220 | 212 | 1158 | 3 | 0.58 | 0.37 |
| 46 | 100000102221 | 212 | 1158 | 3 | 0.47 | 0.35 |
| 47 | 100000102222 | 212 | 1158 | 3 | 0.61 | 0.29 |
| 48 | 100000102217 | 208 | 1158 | 3 | 0.63 | 0.24 |
| 49 | 100000257202 | 208 | 1158 | 3 | 0.84 | 0.29 |
| 50 | 100000257769 | 212 | 1158 | 3 | 0.53 | 0.12 |
| 51 | 100000257770 | 212 | 1158 | 3 | 0.40 | 0.14 |
| 52 | 100000257773 | 212 | 1158 | 3 | 0.73 | 0.38 |
| 53 | 100000322795 | 208 | 1158 | 3 | 0.52 | 0.21 |
| 54 | 100000257767 | 212 | 1158 | 3 | 0.66 | 0.26 |
| 55 | 100000257772 | 212 | 1158 | 3 | 0.51 | 0.36 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria

**Table 3.2.6. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 8**

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 1 | 100000213683 | 208 | 1268 | 4 | 0.83 | 0.23 |
| 2 | 100000213684 | 208 | 1268 | 4 | 0.67 | 0.22 |
| 3 | 100000360208 | 208 | 1268 | 4 | 0.43 | 0.15 |
| 4 | 100000213686 | 208 | 1268 | 4 | 0.50 | 0.26 |
| 5 | 100000213687 | 208 | 1268 | 4 | 0.60 | 0.24 |
| 6 | 100000213688 | 208 | 1268 | 4 | 0.63 | 0.32 |
| 7 | 100000360209 | 208 | 1268 | 4 | 0.53 | 0.26 |
| 8 | 100000270136 | 209 | 1268 | 3 | 0.44 | 0.11 |
| 9 | 100000270132 | 209 | 1268 | 3 | 0.80 | 0.26 |
| 10 | 100000273591 | 208 | 1268 | 3 | 0.55 | 0.28 |
| 11 | 100000270131 | 209 | 1268 | 3 | 0.38 | 0.22 |
| 12 | 100000270135 | 209 | 1268 | 3 | 0.52 | 0.24 |
| 13 | 100000270137 | 209 | 1268 | 3 | 0.57 | 0.18 |
| 14 | 100000304610 | 208 | 1268 | 3 | 0.37 | 0.16 |
| 15 | 100000102331 | 209 | 1268 | 3 | 0.86 | 0.28 |
| 16 | 100000102329 | 208 | 1268 | 3 | 0.56 | 0.12 |
| 17 | 100000102330 | 208 | 1268 | 3 | 0.60 | 0.30 |
| 18 | 100000102334 | 209 | 1268 | 3 | 0.47 | 0.06 |
| 19 | 100000102335 | 209 | 1268 | 3 | 0.85 | 0.30 |
| 20 | 100000270154 | 209 | 1268 | 3 | 0.40 | 0.03 |
| 21 | 100000270155 | 209 | 1268 | 3 | 0.43 | -0.02 |
| 22 | 100000270158 | 209 | 1268 | 3 | 0.50 | 0.23 |
| 23 | 100000273597 | 208 | 1268 | 3 | 0.70 | 0.31 |
| 24 | 100000270163 | 209 | 1268 | 3 | 0.50 | 0.24 |
| 25 | 100000270159 | 209 | 1268 | 3 | 0.77 | 0.36 |
| 26 | 100000273595 | 208 | 1268 | 3 | 0.40 | 0.06 |
| 27 | 100000102341 | 208 | 1268 | 3 | 0.67 | 0.35 |
| 28 | 100000102343 | 209 | 1268 | 3 | 0.57 | 0.23 |
| 29 | 100000102344 | 209 | 1268 | 3 | 0.70 | 0.28 |
| 30 | 100000102346 | 209 | 1268 | 3 | 0.55 | 0.37 |
| 31 | 100000102345 | 209 | 1268 | 3 | 0.63 | 0.37 |
| 32 | 100000257161 | 212 | 1268 | 3 | 0.53 | 0.15 |
| 33 | 100000302728 | 212 | 1268 | 3 | 0.42 | 0.10 |
| 34 | 100000257787 | 208 | 1268 | 3 | 0.37 | 0.16 |
| 35 | 100000257788 | 208 | 1268 | 3 | 0.73 | 0.33 |
| 36 | 100000257158 | 212 | 1268 | 3 | 0.56 | 0.35 |
| 37 | 100000257160 | 212 | 1268 | 3 | 0.60 | 0.32 |
| 38 | 100000257163 | 212 | 1268 | 3 | 0.74 | 0.32 |
| 39 | 100000102320 | 208 | 1268 | 3 | 0.78 | 0.22 |
| 40 | 100000102322 | 212 | 1268 | 3 | 0.65 | 0.13 |
| 41 | 100000102323 | 212 | 1268 | 3 | 0.51 | 0.14 |

Table 3.2.6. The 2010 Mod-MSA: Reading Classical Item Statistics: Grade 8 (Continued)

| Item Seq. No | Item UIN | Strand No.[1] | N-Count[2] | Response Options | P-Value | Point-Biserial |
|---|---|---|---|---|---|---|
| 42 | 100000102324 | 212 | 1268 | 3 | 0.62 | 0.39 |
| 43 | 100000102321 | 208 | 1268 | 3 | 0.76 | 0.31 |
| 44 | 100000257139 | 212 | 1268 | 3 | 0.31 | 0.05 |
| 45 | 100000257140 | 212 | 1268 | 3 | 0.71 | 0.12 |
| 46 | 100000257141 | 212 | 1268 | 3 | 0.48 | 0.15 |
| 47 | 100000257144 | 212 | 1268 | 3 | 0.58 | 0.28 |
| 48 | 100000257777 | 208 | 1268 | 3 | 0.46 | 0.21 |
| 49 | 100000302674 | 208 | 1268 | 3 | 0.56 | 0.22 |
| 50 | 100000302675 | 212 | 1268 | 3 | 0.70 | 0.37 |
| 51 | 100000102304 | 212 | 1268 | 3 | 0.57 | 0.18 |
| 52 | 100000102305 | 212 | 1268 | 3 | 0.66 | 0.30 |
| 53 | 100000102302 | 208 | 1268 | 3 | 0.53 | 0.33 |
| 54 | 100000102307 | 212 | 1268 | 3 | 0.68 | 0.32 |
| 55 | 100000102308 | 212 | 1268 | 3 | 0.66 | 0.39 |

Note: 1. 208=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criteria.

# 4. SCALE CREATION, EQUATING AND RAW SCORES TO SCALE SCORES CONVERSION VIA ITEM RESPONSE THEORY PROCEDURES

For the 2010 administration, there was no equating for Grades 3 to 5 as this was the first year of implementation of the Mod-MSA examinations for these grades. However, grades 6 to 8 forms were linked together by the common items non-equivalent groups (CINEG, Kolen & Brennan, 2004) design.

The Rasch model (Rasch, 1960) was used to develop, calibrate, and scale the Mod-MSA: Reading. The Rasch measurement model is regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All item and test analyses, including item-fit analysis, scaling, diagnosis, and performance prediction were accomplished within this framework. The statistical software used to calibrate and scale the Mod-MSA: Reading was WINSTEPS Version 3.46 (Linacre & Wright, 2000).

**The Rasch Model**

The most basic expression of the Rasch model is in the item characteristic curve (ICC). It shows the probability of a correct response to an item as a function of the ability, i.e., the proficiency level. The probability of a correct response is bounded by 1 (certainty of a correct response) and 0 (certainty of an incorrect response).

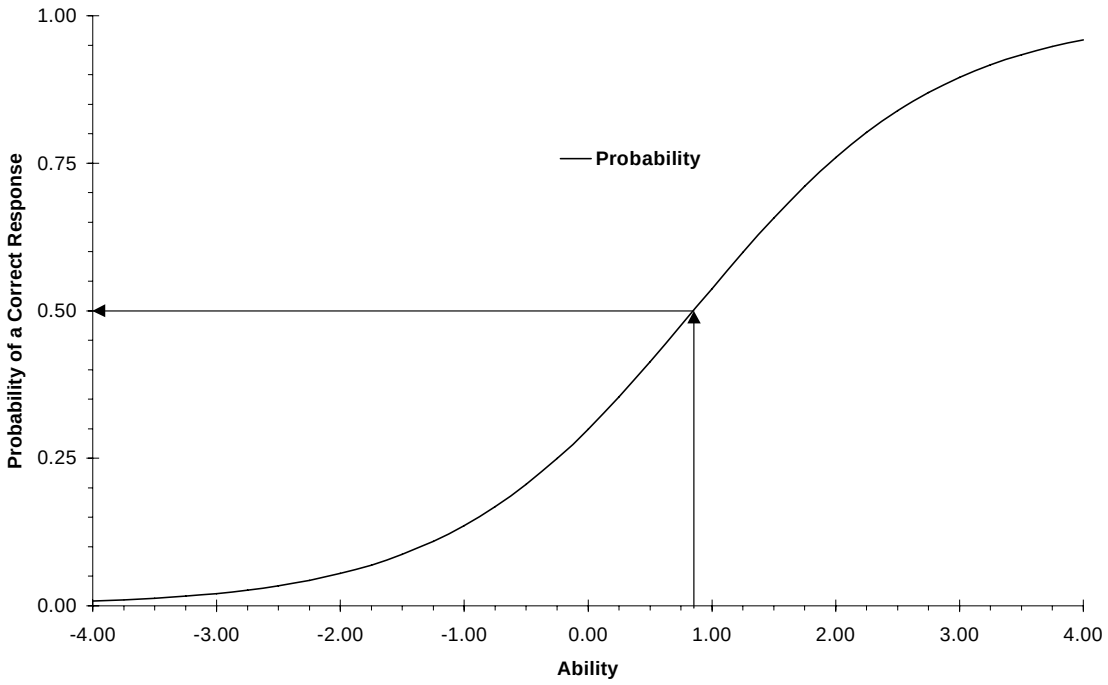



**Figure 4.1 Item Characteristic Curve**

As an example, consider Figure 4.1 which depicts an item that falls at approximately 0.85 on the ability, i.e., the proficiency (horizontal) scale. When a person answers an item at the same level as his or her proficiency, then that person has a probability of roughly 50% of answering the item correctly. Another way of expressing this is that if we have a group of 100 people, all of whom have a proficiency of 0.85, we would expect about 50% of them to answer the item correctly. A person whose proficiency was above 0.85 would a higher probability of getting the item right, while a person whose proficiency is below 0.85 would have a lower probability of getting the item right. This makes intuitive sense and is the basic formulation of Rasch measurement for test items having only two possible categories (i.e., wrong or right).

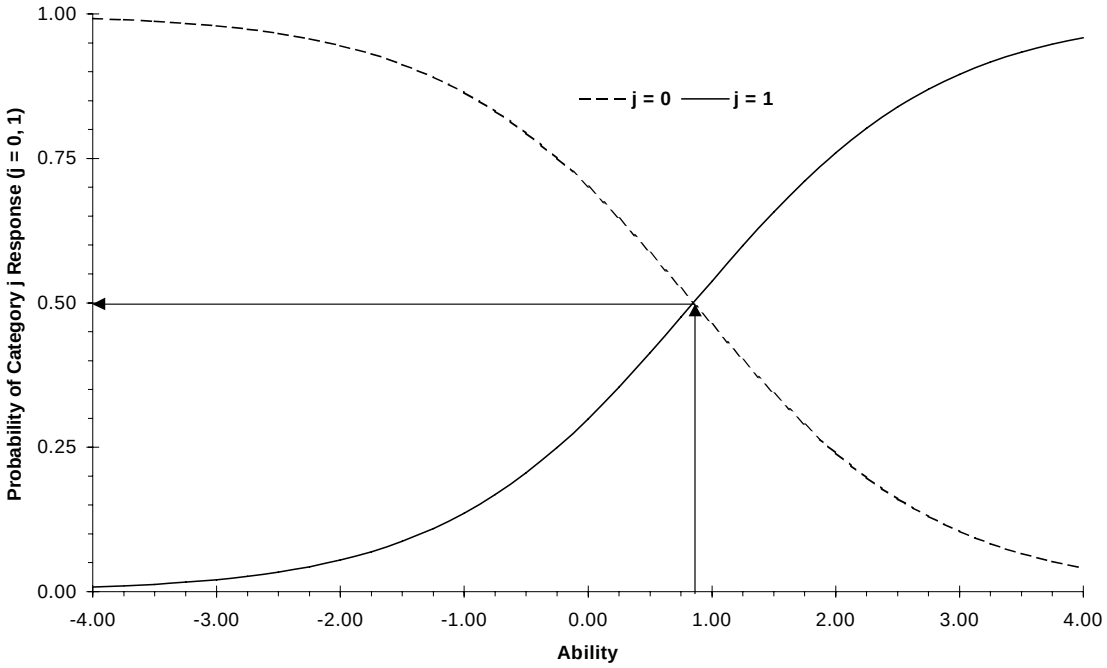



**Figure 4.2 Category Response Curves for a One-Step Item**

Figure 4.2 extends this formulation to show the probabilities of obtaining a wrong answer or a right answer. The curve on the left ($j = 0$) shows the probability of getting a score of "0" while the curve on the right ($j = 1$) shows the probability of getting a score of "1". The point at which the two curves cross indicates the transition point on the proficiency scale where the most likely response changes from a "0" to a "1." Here, the probability of answering the item correctly is 50%.

One important property of the Rasch model is its ability to separate the estimation of item/task parameters from the person parameters. With the Rasch model, the total score given by the sum of the categories in which a person responds is a sufficient statistic for estimating a person's proficiency (i.e., no additional information need be estimated). The total number of responses

across examinees in a particular category is a sufficient statistic for estimating the step difficulty for that category. Thus with the Rasch model, the same total score will yield the same proficiency estimate for different examinees.

The parameters estimated by this model are (1) a proficiency estimate for each person, (2) $m_i$ threshold (difficulty) estimate for each item. From these estimates, the conditional standard error estimates associated with proficiency and the standard error of the difficulty parameter estimates of each item can be calculated (See Section 8.4 for the derivation of the conditional standard error of measurement and the confidence interval set at each proficiency level).

## 4.1. Calibration and Scaling Procedures for Grades 3 to 5

For the 2010 administration, there was no equating for Grades 3 to 5 as this was the first year of implementation of the Mod-MSA examinations for these grades. However, for 2010, a new form of the test was created for Grades 6 to 8 and these forms were linked together by the c*ommon items non-equivalent groups* (CINEG, Kolen & Brennan, 2004) design.

The calibration of the spring 2010 administration of the Mod-MSA: Reading was used to establish the base scale for the assessment in the area of reading at grades 3–5. Item parameters were calibrated using the Rasch measurement model, which placed all items on a common scale. Although the Rasch model is fairly robust, when setting the base scale for an assessment program it is desirable to minimize as many sources of error as practical during the calibration process. This calibration was, therefore, conducted using a two-phase approach. In the first phase only items with acceptable classical item statistics (i.e., non-negative point biserial correlations) and IRT model fit were included. This phase of calibration established the base scale. During the second phase of calibration the items excluded from phase one were placed on the established base scale. This was accomplished by anchoring the parameters obtained for the items included in phase one to their base scale values and only allowing the parameters of the items with less acceptable classical stats (those excluded from phase one) to be freely estimated. This method placed the parameters of the poorly functioning items on the base scale (thereby allowing these items to be selected for operational scoring if necessary) while ensuring that these items did not unduly influence the parameters of those items with acceptable statistics.

Following calibration, all items were sent to data review. Those items not selected as operational items, but not labeled as "do not use" (DNU) during data review, were archived in the item bank for possible future use. RS to SS tables were then created using the established scale parameters of the items selected for operational scoring.

## 4.2. Specifics for Creating the Base Scale for the Mod-MSA: Reading Grades 3-5

The base scale was created for each grade 3 to 5 and content area based on the strength of the items' classical statistics. Items that had poor classical statistics were not included in the creation of the base scale for each grade and content area (for the purposes of this calibration poor item statistics means a negative point biserial correlation).

Items selected from above were calibrated using the Rasch model. From these items, all items showing poor infit and outfit stats (>2.00 and < 0.5) were dropped from the creation of the base scale.

All the items that were excluded from the creation of the base scale were placed on this scale by floating them (keeping their calibration values unanchored) while anchoring the base-scale items to their established calibrated values.

Operational item calibration took place after an identification of these items from data review. The operational form item calibrations remained the same as those established on the above scale for the creation of the RS to SS tables. The non-operational items with their respective calibrations were banked as FT items.

The specific steps in the process were as follows:

1. Conduct classical item analysis of all items on a test.
2. Conduct Rasch calibration of all items on a test that do ***not*** have negative point biserial correlations (based on results of Step 1).
3. Conduct Rasch calibration of all items used in Step 2 that show acceptable infit and outfit(< =2.00 and > =0.5) – this step establishes the base scale for the test.
4. Place the items excluded at Steps 2 and 3 on the base scale by conducting a Rasch calibration with all items used in Step 3 anchored to their base scale values.
5. Submit items for data review with their respective calibrations obtained as outlined above.
6. Create RS to SS scales (for total scores and strand scores), using base scale parameters of the items selected for operational scoring by data review members.

## 4.3. Calibration Equating the 2010 Mod-MSA: Reading Grades 6-8

The base scale for the Mod-MSA: Reading Grades 6 to 8 had been created in 2009. The procedures followed in creating the base scale were the same as those explained above in creating the scale for Grades 3 to 5 in 2010.

For Grades 6 there were 23 common items for use as linking items in the equating process while Grades 7 and 8 had 25 such common items. Items in these grades were placed on the 2009 established scale through the equating process. The calibrations of these items were then sent to data review and the same process was followed as in the Grade 3 to 5 calibration process described above to select operational items, and archiving non-operational items into the item bank.

## 4.4. Specifics of Linking and Equating the 2010 Mod-MSA Grades 6-8: Reading

The 2010 Mod-MSA was calibrated and equated by fixing item parameters of common linking items. To select unstable common items (outliers) from being linking items, the Robust Z procedure was used.

### *4.4.1. Generalized Robust Z Procedure*

Generalized robust z values were calculated by the following procedures:

- Calculate the mean and standard deviation of the linking pool's structure measure parameters ( $D_{ij}$ ) for the 2010 form
- Obtain the ratio of the standard deviations between form 2009 and form 2010
- Obtain the correlation between form 2009 and form 2010 structure measure parameters

- Calculate the difference between form 2009 and 2010 structure measure parameters for each item in the linking pool
- Calculate the mean of the differences calculated above
- Calculate the median of the differences
- Calculate the interquartile range of the differences
- Calculate the robust z for each structure measure parameter in the linking pool where the robust z is defined as (the difference between form 2009 and form 2010 item measure parameters minus the median of the differences) / (interquartile range multiplied by 0.74)
- Calculate the absolute z value of each item measure parameter

### *4.4.2. Guidelines for Selecting Linking Items*

Once the above calculations are made, the following guidelines will determine possible sets of common items to be used for the Rasch equating (SCDE, 2001):

- Try not to include those items with an averaged absolute robust z exceeding 1.645
- Consider that the ratio of the standard deviations of form 2009 and form 2010 item measure parameters should be in the 90 to 110 percent range
- The correlation coefficient of form 2009 and 2010 should be greater than .95
- Do not eliminate more than 20 percent of total score point of the linking pool items

### *4.4.3. Step-by-step Procedure for Selecting Linking Items*

1. Calculate robust Z for all items, the correlation between the fixed Rasch difficulties and the estimated Rasch difficulties, and the ratio of the standard deviations for the fixed and estimated Rasch difficulties. .
2. Check the correlation and ratio of SD of fixed and estimated Rasch parameters. If correlation is greater than 0.95 and ratio is between 0.9 and 1.1 then stop.
3. Choose the item with the largest absolute value of robust Z that is greater than 1.645 and drop from linking set. If no items have a robust Z with an absolute value greater than 1.645 then stop.
4. If the deletion of one more item from the linking set would result in 20% or more of the linking set items being dropped, then stop.
5. Recalculate correlation and SD ratio for remaining items and return to step 1. Do NOT recalculate robust Z values.

The step-by-step procedure is graphically displayed in Figure 4.4.1, below. Tables 4.4.1 to 4.4.3 provide the unequated Rasch item difficulty comparison of the core linking items between 2009 and 2010 for grades 6 to 8 together with their robust z values.

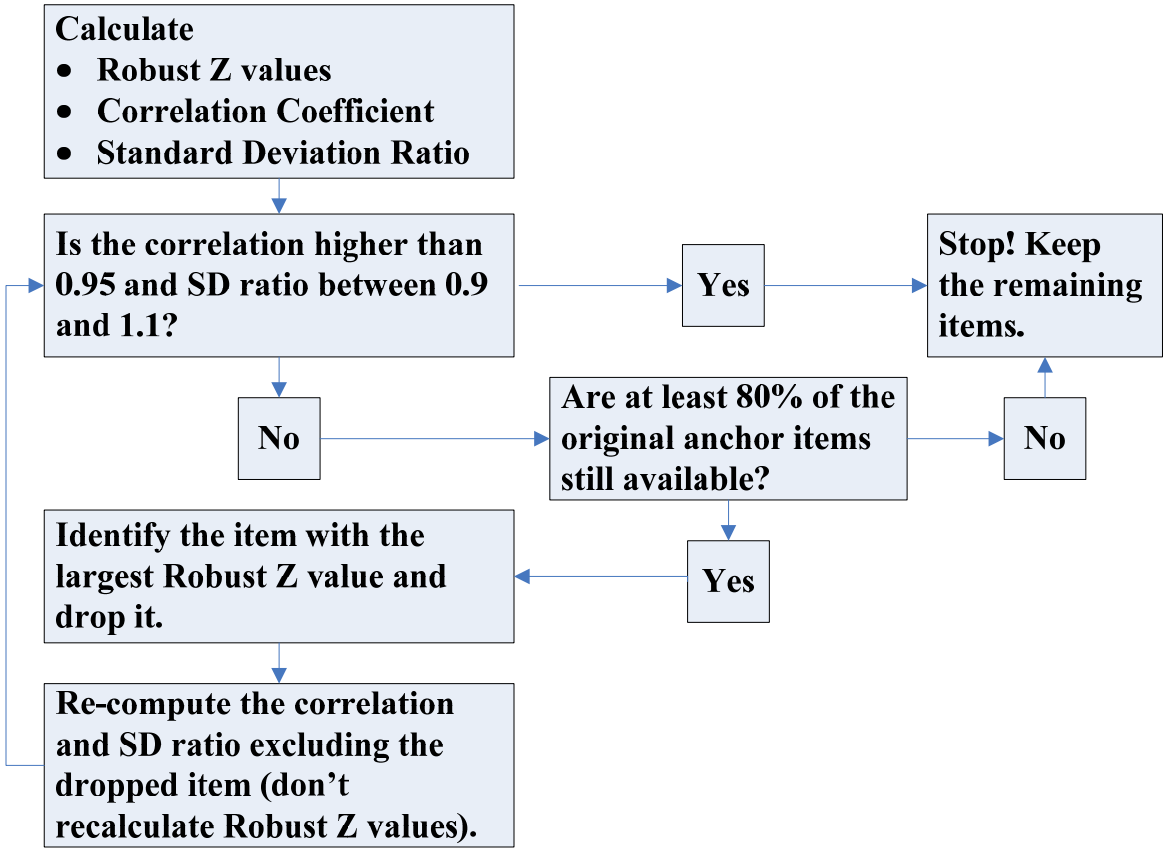



**Figure 4.3.1. Anchor Evaluation Steps Chart for Mod-MSA**

**Table 4.4.1. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 6**

| Item No. | Item Seq. No. | Rasch Diff. 2009 | Rasch Diff.2010 | Robust Z* |
|---|---|---|---|---|
| 1 | 1 | -0.7867 | -0.6640 | -0.67 |
| 2 | 2 | -1.2831 | -1.0811 | -1.32 |
| 3 | 3 | -0.3530 | 0.0364 | **-2.86** |
| 4 | 4 | -0.4888 | -0.4854 | 0.32 |
| 5 | 5 | -0.0595 | -0.1022 | 0.70 |
| 6 | 7 | -0.5508 | -0.6483 | 1.15 |
| 7 | 8 | 1.3285 | 1.4436 | -0.60 |
| 8 | 9 | -0.8831 | -1.1560 | **2.59** |
| 9 | 10 | 0.1881 | 0.1052 | 1.03 |
| 10 | 12 | -0.4133 | -0.2714 | -0.82 |
| 11 | 13 | -0.1792 | 0.0364 | -1.43 |
| 12 | 14 | -0.6591 | -0.6172 | 0.00 |
| 13 | 15 | -0.3109 | -0.3000 | 0.26 |
| 14 | 16 | 0.2119 | 0.4491 | -1.61 |
| 15 | 17 | 0.0695 | 0.1464 | -0.29 |
| 16 | 18 | 0.8270 | 0.9642 | -0.78 |
| 17 | 19 | -0.4815 | -0.4457 | 0.05 |
| 18 | 20 | 0.4870 | 0.4908 | 0.31 |
| 19 | 21 | 0.3073 | 0.1738 | 1.44 |
| 20 | 50 | -0.5472 | -0.7221 | **1.78** |
| 21 | 51 | 0.0153 | 0.0088 | 0.40 |
| 22 | 52 | -0.3179 | -0.2571 | -0.16 |
| 23 | 53 | -0.2517 | -0.0327 | -1.46 |

Note: Bold, underlined values are for Robust Z > 1.645

**Table 4.4.2. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 7**

| Item No. | Item Seq. No. | Rasch Diff. 2009 | Rasch Diff.2010 | Robust Z* |
|---|---|---|---|---|
| 1 | 1 | -0.7780 | -0.5090 | -0.57 |
| 2 | 2 | -1.3540 | -1.1078 | -0.03 |
| 3 | 3 | 0.8276 | 0.9922 | **1.91** |
| 4 | 5 | -0.9143 | -0.8588 | **4.51** |
| 5 | 6 | -0.8469 | -0.5615 | -0.96 |
| 6 | 22 | 0.0908 | 0.3745 | -0.92 |
| 7 | 23 | 0.7546 | 1.0260 | -0.63 |
| 8 | 24 | 0.4769 | 0.7183 | 0.08 |
| 9 | 25 | 0.0775 | 0.4595 | **-3.26** |
| 10 | 26 | 0.4687 | 0.7064 | 0.17 |
| 11 | 27 | -0.1197 | -0.0485 | **4.13** |
| 12 | 28 | 0.4089 | 0.6273 | 0.63 |
| 13 | 29 | 0.0297 | 0.2819 | -0.17 |
| 14 | 30 | 0.4035 | 0.6628 | -0.34 |
| 15 | 31 | 0.6886 | 1.1033 | **-4.04** |
| 16 | 39 | -1.6963 | -1.4900 | 0.92 |
| 17 | 40 | -0.1924 | -0.0761 | **3.06** |
| 18 | 41 | -1.1534 | -0.9085 | 0.00 |
| 19 | 42 | 0.0244 | 0.2742 | -0.12 |
| 20 | 43 | -1.3660 | -1.2209 | **2.37** |
| 21 | 44 | 0.1918 | 0.2973 | **3.32** |
| 22 | 45 | -0.2739 | -0.0642 | 0.84 |
| 23 | 46 | 0.1732 | 0.4208 | -0.06 |
| 24 | 47 | -0.4429 | -0.1798 | -0.43 |
| 25 | 48 | -0.5024 | -0.2937 | 0.86 |

Note: Bold, underlined values are for Robust Z > 1.645

**Table 4.4.3. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 8**

| Item No. | Item Seq. No. | Rasch Diff. 2009 | Rasch Diff.2010 | Robust Z* |
|---|---|---|---|---|
| 1 | 1 | -1.1648 | -1.3041 | 0.58 |
| 2 | 2 | -0.4197 | -0.3775 | -1.13 |
| 3 | 4 | 0.4446 | 0.3926 | -0.24 |
| 4 | 5 | 0.0045 | -0.0457 | -0.26 |
| 5 | 6 | -0.1989 | -0.1547 | -1.15 |
| 6 | 15 | -1.3174 | -1.5258 | 1.22 |
| 7 | 16 | 0.2613 | 0.1423 | 0.38 |
| 8 | 17 | 0.0618 | -0.0170 | 0.01 |
| 9 | 18 | 0.3175 | 0.5589 | **-3.00** |
| 10 | 19 | -1.1686 | -1.4334 | **1.75** |
| 11 | 27 | -0.2773 | -0.3581 | 0.03 |
| 12 | 28 | 0.2393 | 0.1213 | 0.38 |
| 13 | 29 | -0.5397 | -0.5208 | -0.91 |
| 14 | 30 | 0.2686 | 0.1913 | -0.01 |
| 15 | 31 | -0.0760 | -0.1806 | 0.25 |
| 16 | 39 | -0.9453 | -0.9953 | -0.26 |
| 17 | 40 | -0.2879 | -0.2740 | -0.86 |
| 18 | 41 | 0.2760 | 0.3822 | **-1.73** |
| 19 | 42 | 0.0370 | -0.1072 | 0.62 |
| 20 | 43 | -0.8923 | -0.8741 | -0.90 |
| 21 | 51 | 0.2711 | 0.1002 | 0.87 |
| 22 | 52 | -0.2458 | -0.3158 | -0.08 |
| 23 | 53 | 0.4618 | 0.2783 | 0.99 |
| 24 | 54 | -0.3358 | -0.4206 | 0.06 |
| 25 | 55 | -0.1834 | -0.3043 | 0.40 |

Note: Bold, underlined values are for Robust Z > 1.645

## 4.5. Reporting Scale Scores for the 2010 Mod-MSA: Reading

The Mod-MSA reports student scores on the total performance of students on the reading examination (total score) as well as the reporting of their strand scores outlined in Section 2.3.

In order to facilitate the use and interpretation of the results of the 2010 Mod-MSA Reading, a scale score was created for each point on the raw score tables (total scores as well as strand scores) that had a mean = 50; a standard deviation = 12; and the lowest and highest obtainable scale scores (LOSS and HOSS) as 2 and 98, respectively. As is the case with standard MSA, the lowest obtainable raw score (zero) was automatically set to the LOSS and the highest obtainable raw score (51) set to the HOSS in the event that the actual scale score associated with these raw scores fell above or below these values, respectively.

Once RS to Theta tables were produced by the WINSTEPS 3.46 program after data review, theta to scale score constants were calculated using the following formula:

$$SS = Slope \times Theta + Intercept$$

$$SEM_{CSS} = Slope \times SEM_{CT}$$

where

*Slope* = 12 / the standard deviation of the theta values, and

*Intercept* = 50 – slope × mean of the theta values

*Theta* = the *IRT* proficiency estimate at a particular raw score on the scoring continuum

$SEM_{CSS}$ = the standard error of the scale score, and

$SEM_{CT}$ = the standard error conditional on proficiency (theta) estimates

Table 4.5.1 depicts the slope and intercept that were used for each grade. It should be noted that the same slopes and intercepts were used for Grades 6 to 8 as those used in 2009. Similarly, the same slopes and intercept for each of the grades 3 to 8 will be used for future administrations. Total raw score to scale score conversion tables for Grades 3-8 are provided in Tables 4.5.2 to 4.5.7, while strand level RS to SS are provided in Tables 4.5.8 to Tables 4.5.13.

Each student's total raw score for the strands was a summation of the individual item score within a strand level. The strand levels were classified as stated in section 2.3 and the item parameters within each strand was obtained using the Winsteps program in the same manner as those obtained for the total test. Once the item parameters were available, thetas (student proficiency scores) were calculated for each raw score point that could be obtained within each strand. The thetas were transferred to scale scores, using the same slope and intercept as that which were applied for the total reading test score.

**Table 4.5.1. The 2010 Mod-MSA, Reading Slope and Intercept for the Transfer of RS to SS Across Grades**

| Grade | Slope | Intercept |
|---|---|---|
| 3 | 13.8375 | 47.9876 |
| 4 | 15.3069 | 48.7765 |
| 5 | 15.8311 | 49.1418 |
| 6 | 18.4057 | 48.0880 |
| 7 | 17.1743 | 47.9523 |
| 8 | 16.5287 | 43.8652 |

## Raw Score to Scale Score Conversion Tables for the Total Score

### Table 4.5.2. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 3

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.5040 | 2.0073 | 2 | 28 | - | - |
| 1 | -4.0956 | 1.0147 | 2 | 14 | - | - |
| 2 | -3.3724 | 0.7282 | 2 | 10 | - | - |
| 3 | -2.9365 | 0.6034 | 7 | 8 | - | - |
| 4 | -2.6179 | 0.5303 | 12 | 7 | 5 | 19 |
| 5 | -2.3632 | 0.4815 | 15 | 7 | 8 | 22 |
| 6 | -2.1488 | 0.4462 | 18 | 6 | 12 | 24 |
| 7 | -1.9620 | 0.4194 | 21 | 6 | 15 | 27 |
| 8 | -1.7951 | 0.3984 | 23 | 6 | 17 | 29 |
| 9 | -1.6432 | 0.3815 | 25 | 5 | 20 | 30 |
| 10 | -1.5030 | 0.3677 | 27 | 5 | 22 | 32 |
| 11 | -1.3721 | 0.3563 | 29 | 5 | 24 | 34 |
| 12 | -1.2486 | 0.3467 | 31 | 5 | 26 | 36 |
| 13 | -1.1312 | 0.3387 | 32 | 5 | 27 | 37 |
| 14 | -1.0189 | 0.3320 | 34 | 5 | 29 | 39 |
| 15 | -0.9107 | 0.3263 | 35 | 5 | 30 | 40 |
| 16 | -0.8057 | 0.3216 | 37 | 4 | 33 | 41 |
| 17 | -0.7035 | 0.3178 | 38 | 4 | 34 | 42 |
| 18 | -0.6036 | 0.3147 | 40 | 4 | 36 | 44 |
| 19 | -0.5054 | 0.3123 | 41 | 4 | 37 | 45 |
| 20 | -0.4084 | 0.3105 | 42 | 4 | 38 | 46 |
| 21 | -0.3124 | 0.3093 | 44 | 4 | 40 | 48 |
| 22 | -0.2169 | 0.3087 | 45 | 4 | 41 | 49 |
| 23 | -0.1216 | 0.3087 | 46 | 4 | 42 | 50 |
| 24 | -0.0262 | 0.3093 | 48 | 4 | 44 | 52 |
| 25 | 0.0698 | 0.3104 | 49 | 4 | 45 | 53 |
| 26 | 0.1666 | 0.3122 | 50 | 4 | 46 | 54 |
| 27 | 0.2649 | 0.3146 | 52 | 4 | 48 | 56 |
| 28 | 0.3647 | 0.3177 | 53 | 4 | 49 | 57 |
| 29 | 0.4669 | 0.3215 | 54 | 4 | 50 | 58 |
| 30 | 0.5717 | 0.3262 | 56 | 5 | 51 | 61 |
| 31 | 0.6799 | 0.3318 | 57 | 5 | 52 | 62 |
| 32 | 0.7922 | 0.3385 | 59 | 5 | 54 | 64 |
| 33 | 0.9094 | 0.3465 | 61 | 5 | 56 | 66 |
| 34 | 1.0328 | 0.3561 | 62 | 5 | 57 | 67 |
| 35 | 1.1635 | 0.3675 | 64 | 5 | 59 | 69 |
| 36 | 1.3036 | 0.3813 | 66 | 5 | 61 | 71 |
| 37 | 1.4553 | 0.3982 | 68 | 6 | 62 | 74 |
| 38 | 1.6220 | 0.4192 | 70 | 6 | 64 | 76 |
| 39 | 1.8087 | 0.4460 | 73 | 6 | 67 | 79 |
| 40 | 2.0231 | 0.4813 | 76 | 7 | 69 | 83 |
| 41 | 2.2776 | 0.5302 | 80 | 7 | 73 | 87 |
| 42 | 2.5962 | 0.6033 | 84 | 8 | 76 | 92 |
| 43 | 3.0319 | 0.7280 | 90 | 10 | - | - |
| 44 | 3.7549 | 1.0147 | 98 | 14 | - | - |
| 45 | 5.1633 | 2.0072 | 98 | 28 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors may not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.3. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 4**

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.4793 | 2.0073 | 2 | 31 | - | - |
| 1 | -4.0708 | 1.0148 | 2 | 16 | - | - |
| 2 | -3.3476 | 0.7283 | 2 | 11 | - | - |
| 3 | -2.9115 | 0.6035 | 4 | 9 | - | - |
| 4 | -2.5927 | 0.5305 | 9 | 8 | - | - |
| 5 | -2.3379 | 0.4817 | 13 | 7 | 6 | 20 |
| 6 | -2.1233 | 0.4464 | 16 | 7 | 9 | 23 |
| 7 | -1.9361 | 0.4197 | 19 | 6 | 13 | 25 |
| 8 | -1.7689 | 0.3988 | 22 | 6 | 16 | 28 |
| 9 | -1.6168 | 0.3820 | 24 | 6 | 18 | 30 |
| 10 | -1.4762 | 0.3682 | 26 | 6 | 20 | 32 |
| 11 | -1.3449 | 0.3568 | 28 | 5 | 23 | 33 |
| 12 | -1.2210 | 0.3474 | 30 | 5 | 25 | 35 |
| 13 | -1.1032 | 0.3394 | 32 | 5 | 27 | 37 |
| 14 | -0.9903 | 0.3328 | 34 | 5 | 29 | 39 |
| 15 | -0.8815 | 0.3272 | 35 | 5 | 30 | 40 |
| 16 | -0.7759 | 0.3226 | 37 | 5 | 32 | 42 |
| 17 | -0.6731 | 0.3188 | 38 | 5 | 33 | 43 |
| 18 | -0.5725 | 0.3158 | 40 | 5 | 35 | 45 |
| 19 | -0.4735 | 0.3135 | 42 | 5 | 37 | 47 |
| 20 | -0.3758 | 0.3118 | 43 | 5 | 38 | 48 |
| 21 | -0.2790 | 0.3107 | 45 | 5 | 40 | 50 |
| 22 | -0.1827 | 0.3102 | 46 | 5 | 41 | 51 |
| 23 | -0.0865 | 0.3103 | 47 | 5 | 42 | 52 |
| 24 | 0.0100 | 0.3109 | 49 | 5 | 44 | 54 |
| 25 | 0.1069 | 0.3121 | 50 | 5 | 45 | 55 |
| 26 | 0.2049 | 0.3140 | 52 | 5 | 47 | 57 |
| 27 | 0.3042 | 0.3165 | 53 | 5 | 48 | 58 |
| 28 | 0.4054 | 0.3196 | 55 | 5 | 50 | 60 |
| 29 | 0.5088 | 0.3236 | 57 | 5 | 52 | 62 |
| 30 | 0.6150 | 0.3283 | 58 | 5 | 53 | 63 |
| 31 | 0.7245 | 0.3340 | 60 | 5 | 55 | 65 |
| 32 | 0.8384 | 0.3408 | 62 | 5 | 57 | 67 |
| 33 | 0.9572 | 0.3489 | 63 | 5 | 58 | 68 |
| 34 | 1.0822 | 0.3585 | 65 | 5 | 60 | 70 |
| 35 | 1.2148 | 0.3700 | 67 | 6 | 61 | 73 |
| 36 | 1.3567 | 0.3839 | 70 | 6 | 64 | 76 |
| 37 | 1.5105 | 0.4008 | 72 | 6 | 66 | 78 |
| 38 | 1.6793 | 0.4218 | 74 | 6 | 68 | 80 |
| 39 | 1.8683 | 0.4486 | 77 | 7 | 70 | 84 |
| 40 | 2.0850 | 0.4839 | 81 | 7 | 74 | 88 |
| 41 | 2.3420 | 0.5326 | 85 | 8 | 77 | 93 |
| 42 | 2.6631 | 0.6055 | 90 | 9 | - | - |
| 43 | 3.1017 | 0.7300 | 96 | 11 | - | - |
| 44 | 3.8277 | 1.0162 | 98 | 16 | - | - |
| 45 | 5.2383 | 2.0080 | 98 | 31 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.4. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 5**

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.5057 | 2.0075 | 2 | 32 | - | - |
| 1 | -4.0969 | 1.0151 | 2 | 16 | - | - |
| 2 | -3.3732 | 0.7286 | 2 | 12 | - | - |
| 3 | -2.9367 | 0.6039 | 3 | 10 | - | - |
| 4 | -2.6173 | 0.5310 | 8 | 8 | - | - |
| 5 | -2.3619 | 0.4823 | 12 | 8 | 4 | 20 |
| 6 | -2.1466 | 0.4472 | 15 | 7 | 8 | 22 |
| 7 | -1.9587 | 0.4206 | 18 | 7 | 11 | 25 |
| 8 | -1.7908 | 0.3997 | 21 | 6 | 15 | 27 |
| 9 | -1.6379 | 0.3829 | 23 | 6 | 17 | 29 |
| 10 | -1.4966 | 0.3692 | 25 | 6 | 19 | 31 |
| 11 | -1.3645 | 0.3579 | 28 | 6 | 22 | 34 |
| 12 | -1.2399 | 0.3485 | 30 | 6 | 24 | 36 |
| 13 | -1.1212 | 0.3406 | 31 | 5 | 26 | 36 |
| 14 | -1.0076 | 0.3340 | 33 | 5 | 28 | 38 |
| 15 | -0.8978 | 0.3285 | 35 | 5 | 30 | 40 |
| 16 | -0.7915 | 0.3239 | 37 | 5 | 32 | 42 |
| 17 | -0.6878 | 0.3202 | 38 | 5 | 33 | 43 |
| 18 | -0.5862 | 0.3172 | 40 | 5 | 35 | 45 |
| 19 | -0.4864 | 0.3148 | 41 | 5 | 36 | 46 |
| 20 | -0.3880 | 0.3131 | 43 | 5 | 38 | 48 |
| 21 | -0.2902 | 0.3120 | 45 | 5 | 40 | 50 |
| 22 | -0.1931 | 0.3115 | 46 | 5 | 41 | 51 |
| 23 | -0.0960 | 0.3115 | 48 | 5 | 43 | 53 |
| 24 | 0.0011 | 0.3122 | 49 | 5 | 44 | 54 |
| 25 | 0.0990 | 0.3133 | 51 | 5 | 46 | 56 |
| 26 | 0.1976 | 0.3151 | 52 | 5 | 47 | 57 |
| 27 | 0.2976 | 0.3175 | 54 | 5 | 49 | 59 |
| 28 | 0.3995 | 0.3206 | 55 | 5 | 50 | 60 |
| 29 | 0.5033 | 0.3244 | 57 | 5 | 52 | 62 |
| 30 | 0.6102 | 0.3291 | 59 | 5 | 54 | 64 |
| 31 | 0.7202 | 0.3347 | 61 | 5 | 56 | 66 |
| 32 | 0.8343 | 0.3413 | 62 | 5 | 57 | 67 |
| 33 | 0.9535 | 0.3493 | 64 | 6 | 58 | 70 |
| 34 | 1.0787 | 0.3587 | 66 | 6 | 60 | 72 |
| 35 | 1.2114 | 0.3701 | 68 | 6 | 62 | 74 |
| 36 | 1.3533 | 0.3838 | 71 | 6 | 65 | 77 |
| 37 | 1.5069 | 0.4005 | 73 | 6 | 67 | 79 |
| 38 | 1.6755 | 0.4214 | 76 | 7 | 69 | 83 |
| 39 | 1.8640 | 0.4480 | 79 | 7 | 72 | 86 |
| 40 | 2.0801 | 0.4831 | 82 | 8 | 74 | 90 |
| 41 | 2.3364 | 0.5318 | 86 | 8 | 78 | 94 |
| 42 | 2.6567 | 0.6047 | 91 | 10 | - | - |
| 43 | 3.0941 | 0.7292 | 98 | 12 | - | - |
| 44 | 3.8186 | 1.0155 | 98 | 16 | - | - |
| 45 | 5.2282 | 2.0076 | 98 | 32 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.5 The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 6**

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.3858 | 2.0071 | 2 | 37 | - | - |
| 1 | -3.9781 | 1.0143 | 2 | 19 | - | - |
| 2 | -3.2560 | 0.7274 | 2 | 13 | - | - |
| 3 | -2.8211 | 0.6025 | 2 | 11 | - | - |
| 4 | -2.5036 | 0.5294 | 2 | 10 | - | - |
| 5 | -2.2499 | 0.4804 | 7 | 9 | - | - |
| 6 | -2.0366 | 0.4450 | 11 | 8 | 3 | 19 |
| 7 | -1.8506 | 0.4182 | 14 | 8 | 6 | 22 |
| 8 | -1.6848 | 0.3971 | 17 | 7 | 10 | 24 |
| 9 | -1.5340 | 0.3802 | 20 | 7 | 13 | 27 |
| 10 | -1.3947 | 0.3663 | 22 | 7 | 15 | 29 |
| 11 | -1.2649 | 0.3549 | 25 | 7 | 18 | 32 |
| 12 | -1.1423 | 0.3453 | 27 | 6 | 21 | 33 |
| 13 | -1.0259 | 0.3373 | 29 | 6 | 23 | 35 |
| 14 | -0.9144 | 0.3306 | 31 | 6 | 25 | 37 |
| 15 | -0.8070 | 0.3250 | 33 | 6 | 27 | 39 |
| 16 | -0.7030 | 0.3203 | 35 | 6 | 29 | 41 |
| 17 | -0.6017 | 0.3165 | 37 | 6 | 31 | 43 |
| 18 | -0.5025 | 0.3134 | 39 | 6 | 33 | 45 |
| 19 | -0.4050 | 0.3110 | 41 | 6 | 35 | 47 |
| 20 | -0.3089 | 0.3093 | 42 | 6 | 36 | 48 |
| 21 | -0.2137 | 0.3081 | 44 | 6 | 38 | 50 |
| 22 | -0.1189 | 0.3076 | 46 | 6 | 40 | 52 |
| 23 | -0.0244 | 0.3076 | 48 | 6 | 42 | 54 |
| 24 | 0.0704 | 0.3082 | 49 | 6 | 43 | 55 |
| 25 | 0.1659 | 0.3094 | 51 | 6 | 45 | 57 |
| 26 | 0.2621 | 0.3113 | 53 | 6 | 47 | 59 |
| 27 | 0.3597 | 0.3137 | 55 | 6 | 49 | 61 |
| 28 | 0.4591 | 0.3168 | 57 | 6 | 51 | 63 |
| 29 | 0.5607 | 0.3207 | 58 | 6 | 52 | 64 |
| 30 | 0.6650 | 0.3255 | 60 | 6 | 54 | 66 |
| 31 | 0.7728 | 0.3312 | 62 | 6 | 56 | 68 |
| 32 | 0.8847 | 0.3379 | 64 | 6 | 58 | 70 |
| 33 | 1.0015 | 0.3460 | 67 | 6 | 61 | 73 |
| 34 | 1.1245 | 0.3556 | 69 | 7 | 62 | 76 |
| 35 | 1.2550 | 0.3671 | 71 | 7 | 64 | 78 |
| 36 | 1.3948 | 0.3810 | 74 | 7 | 67 | 81 |
| 37 | 1.5463 | 0.3980 | 77 | 7 | 70 | 84 |
| 38 | 1.7129 | 0.4191 | 80 | 8 | 72 | 88 |
| 39 | 1.8997 | 0.4460 | 83 | 8 | 75 | 91 |
| 40 | 2.1139 | 0.4813 | 87 | 9 | 78 | 96 |
| 41 | 2.3685 | 0.5303 | 92 | 10 | - | - |
| 42 | 2.6872 | 0.6034 | 98 | 11 | - | - |
| 43 | 3.1232 | 0.7283 | 98 | 13 | - | - |
| 44 | 3.8465 | 1.0149 | 98 | 19 | - | - |
| 45 | 5.2553 | 2.0074 | 98 | 37 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.6. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 7**

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.6365 | 2.0084 | 2 | 34 | - | - |
| 1 | -4.2246 | 1.0170 | 2 | 17 | - | - |
| 2 | -3.4971 | 0.7312 | 2 | 13 | - | - |
| 3 | -3.0568 | 0.6069 | 2 | 10 | - | - |
| 4 | -2.7339 | 0.5343 | 2 | 9 | - | - |
| 5 | -2.4750 | 0.4857 | 5 | 8 | - | - |
| 6 | -2.2566 | 0.4507 | 9 | 8 | - | - |
| 7 | -2.0657 | 0.4241 | 12 | 7 | 5 | 19 |
| 8 | -1.8949 | 0.4032 | 15 | 7 | 8 | 22 |
| 9 | -1.7394 | 0.3864 | 18 | 7 | 11 | 25 |
| 10 | -1.5955 | 0.3726 | 21 | 6 | 15 | 27 |
| 11 | -1.4610 | 0.3612 | 23 | 6 | 17 | 29 |
| 12 | -1.3340 | 0.3516 | 25 | 6 | 19 | 31 |
| 13 | -1.2132 | 0.3436 | 27 | 6 | 21 | 33 |
| 14 | -1.0976 | 0.3368 | 29 | 6 | 23 | 35 |
| 15 | -0.9861 | 0.3311 | 31 | 6 | 25 | 37 |
| 16 | -0.8781 | 0.3263 | 33 | 6 | 27 | 39 |
| 17 | -0.7729 | 0.3224 | 35 | 6 | 29 | 41 |
| 18 | -0.6700 | 0.3192 | 36 | 5 | 31 | 41 |
| 19 | -0.5690 | 0.3167 | 38 | 5 | 33 | 43 |
| 20 | -0.4694 | 0.3147 | 40 | 5 | 35 | 45 |
| 21 | -0.3707 | 0.3134 | 42 | 5 | 37 | 47 |
| 22 | -0.2728 | 0.3127 | 43 | 5 | 38 | 48 |
| 23 | -0.1750 | 0.3126 | 45 | 5 | 40 | 50 |
| 24 | -0.0772 | 0.3130 | 47 | 5 | 42 | 52 |
| 25 | 0.0209 | 0.3139 | 48 | 5 | 43 | 53 |
| 26 | 0.1200 | 0.3155 | 50 | 5 | 45 | 55 |
| 27 | 0.2202 | 0.3178 | 52 | 5 | 47 | 57 |
| 28 | 0.3221 | 0.3207 | 53 | 6 | 47 | 59 |
| 29 | 0.4260 | 0.3243 | 55 | 6 | 49 | 61 |
| 30 | 0.5326 | 0.3288 | 57 | 6 | 51 | 63 |
| 31 | 0.6425 | 0.3343 | 59 | 6 | 53 | 65 |
| 32 | 0.7563 | 0.3408 | 61 | 6 | 55 | 67 |
| 33 | 0.8751 | 0.3486 | 63 | 6 | 57 | 69 |
| 34 | 0.9998 | 0.3580 | 65 | 6 | 59 | 71 |
| 35 | 1.1319 | 0.3692 | 67 | 6 | 61 | 73 |
| 36 | 1.2731 | 0.3828 | 70 | 7 | 63 | 77 |
| 37 | 1.4259 | 0.3995 | 72 | 7 | 65 | 79 |
| 38 | 1.5937 | 0.4204 | 75 | 7 | 68 | 82 |
| 39 | 1.7812 | 0.4470 | 79 | 8 | 71 | 87 |
| 40 | 1.9963 | 0.4821 | 82 | 8 | 74 | 90 |
| 41 | 2.2515 | 0.5308 | 87 | 9 | 78 | 96 |
| 42 | 2.5707 | 0.6037 | 92 | 10 | - | - |
| 43 | 3.0070 | 0.7284 | 98 | 13 | - | - |
| 44 | 3.7303 | 1.0149 | 98 | 17 | - | - |
| 45 | 5.1389 | 2.0073 | 98 | 34 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.7. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 8**

| Raw Score | Proficiency Estimates | SE | SS[1] | SE(SS) | SS-1SE (SS)[2] | SS+1SE (SS)[2] |
|---|---|---|---|---|---|---|
| 0 | -5.3918 | 2.0078 | 2 | 33 | - | - |
| 1 | -3.9819 | 1.0156 | 2 | 17 | - | - |
| 2 | -3.2570 | 0.7294 | 2 | 12 | - | - |
| 3 | -2.8193 | 0.6048 | 2 | 10 | - | - |
| 4 | -2.4990 | 0.5319 | 3 | 9 | - | - |
| 5 | -2.2427 | 0.4830 | 7 | 8 | - | - |
| 6 | -2.0268 | 0.4478 | 10 | 7 | 3 | 17 |
| 7 | -1.8385 | 0.4210 | 13 | 7 | 6 | 20 |
| 8 | -1.6704 | 0.4000 | 16 | 7 | 9 | 23 |
| 9 | -1.5173 | 0.3831 | 19 | 6 | 13 | 25 |
| 10 | -1.3760 | 0.3692 | 21 | 6 | 15 | 27 |
| 11 | -1.2440 | 0.3577 | 23 | 6 | 17 | 29 |
| 12 | -1.1195 | 0.3481 | 25 | 6 | 19 | 31 |
| 13 | -1.0012 | 0.3400 | 27 | 6 | 21 | 33 |
| 14 | -0.8880 | 0.3332 | 29 | 6 | 23 | 35 |
| 15 | -0.7790 | 0.3275 | 31 | 5 | 26 | 36 |
| 16 | -0.6733 | 0.3227 | 33 | 5 | 28 | 38 |
| 17 | -0.5705 | 0.3187 | 34 | 5 | 29 | 39 |
| 18 | -0.4700 | 0.3155 | 36 | 5 | 31 | 41 |
| 19 | -0.3712 | 0.3130 | 38 | 5 | 33 | 43 |
| 20 | -0.2739 | 0.3112 | 39 | 5 | 34 | 44 |
| 21 | -0.1774 | 0.3099 | 41 | 5 | 36 | 46 |
| 22 | -0.0817 | 0.3092 | 43 | 5 | 38 | 48 |
| 23 | 0.0139 | 0.3091 | 44 | 5 | 39 | 49 |
| 24 | 0.1096 | 0.3096 | 46 | 5 | 41 | 51 |
| 25 | 0.2057 | 0.3107 | 47 | 5 | 42 | 52 |
| 26 | 0.3027 | 0.3123 | 49 | 5 | 44 | 54 |
| 27 | 0.4010 | 0.3147 | 50 | 5 | 45 | 55 |
| 28 | 0.5009 | 0.3177 | 52 | 5 | 47 | 57 |
| 29 | 0.6030 | 0.3214 | 54 | 5 | 49 | 59 |
| 30 | 0.7077 | 0.3260 | 56 | 5 | 51 | 61 |
| 31 | 0.8158 | 0.3316 | 57 | 5 | 52 | 62 |
| 32 | 0.9279 | 0.3382 | 59 | 6 | 53 | 65 |
| 33 | 1.0449 | 0.3462 | 61 | 6 | 55 | 67 |
| 34 | 1.1680 | 0.3557 | 63 | 6 | 57 | 69 |
| 35 | 1.2984 | 0.3671 | 65 | 6 | 59 | 71 |
| 36 | 1.4381 | 0.3808 | 68 | 6 | 62 | 74 |
| 37 | 1.5894 | 0.3977 | 70 | 7 | 63 | 77 |
| 38 | 1.7558 | 0.4187 | 73 | 7 | 66 | 80 |
| 39 | 1.9420 | 0.4454 | 76 | 7 | 69 | 83 |
| 40 | 2.1558 | 0.4808 | 79 | 8 | 71 | 87 |
| 41 | 2.4097 | 0.5296 | 84 | 9 | 75 | 93 |
| 42 | 2.7276 | 0.6027 | 89 | 10 | - | - |
| 43 | 3.1628 | 0.7276 | 96 | 12 | - | - |
| 44 | 3.8851 | 1.0144 | 98 | 17 | - | - |
| 45 | 5.2929 | 2.0071 | 98 | 33 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Raw Score to Scale Score Conversion Tables for the Subscales

### Table 4.5.8. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 3

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 28 | - | - |
| General Reading | 1 | 5 | 14 | - | - |
| General Reading | 2 | 15 | 11 | 4 | 26 |
| General Reading | 3 | 22 | 9 | 13 | 31 |
| General Reading | 4 | 28 | 8 | 20 | 36 |
| General Reading | 5 | 32 | 8 | 24 | 40 |
| General Reading | 6 | 36 | 7 | 29 | 43 |
| General Reading | 7 | 40 | 7 | 33 | 47 |
| General Reading | 8 | 44 | 7 | 37 | 51 |
| General Reading | 9 | 48 | 7 | 41 | 55 |
| General Reading | 10 | 52 | 7 | 45 | 59 |
| General Reading | 11 | 56 | 8 | 48 | 64 |
| General Reading | 12 | 60 | 8 | 52 | 68 |
| General Reading | 13 | 66 | 9 | 57 | 75 |
| General Reading | 14 | 73 | 11 | 62 | 84 |
| General Reading | 15 | 84 | 14 | 70 | 98 |
| General Reading | 16 | 98 | 28 | - | - |
| Literary | 0 | 2 | 28 | - | - |
| Literary | 1 | 7 | 14 | - | - |
| Literary | 2 | 18 | 11 | 7 | 29 |
| Literary | 3 | 25 | 9 | 16 | 34 |
| Literary | 4 | 30 | 8 | 22 | 38 |
| Literary | 5 | 35 | 8 | 27 | 43 |
| Literary | 6 | 40 | 8 | 32 | 48 |
| Literary | 7 | 44 | 8 | 36 | 52 |
| Literary | 8 | 48 | 8 | 40 | 56 |
| Literary | 9 | 53 | 8 | 45 | 61 |
| Literary | 10 | 57 | 8 | 49 | 65 |
| Literary | 11 | 63 | 9 | 54 | 72 |
| Literary | 12 | 70 | 11 | 59 | 81 |
| Literary | 13 | 81 | 15 | 66 | 96 |
| Literary | 14 | 98 | 28 | - | - |
| Informational | 0 | 2 | 28 | - | - |
| Informational | 1 | 11 | 14 | - | - |
| Informational | 2 | 22 | 11 | 11 | 33 |
| Informational | 3 | 29 | 9 | 20 | 38 |
| Informational | 4 | 34 | 8 | 26 | 42 |
| Informational | 5 | 39 | 8 | 31 | 47 |
| Informational | 6 | 43 | 7 | 36 | 50 |
| Informational | 7 | 47 | 7 | 40 | 54 |
| Informational | 8 | 51 | 7 | 44 | 58 |
| Informational | 9 | 55 | 7 | 48 | 62 |
| Informational | 10 | 59 | 8 | 51 | 67 |
| Informational | 11 | 64 | 8 | 56 | 72 |
| Informational | 12 | 69 | 9 | 60 | 78 |
| Informational | 13 | 76 | 11 | 65 | 87 |
| Informational | 14 | 87 | 14 | - | - |
| Informational | 15 | 98 | 28 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Table 4.5.9. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 4

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 31 | - | - |
| General Reading | 1 | 2 | 16 | - | - |
| General Reading | 2 | 11 | 12 | - | - |
| General Reading | 3 | 19 | 10 | 9 | 29 |
| General Reading | 4 | 25 | 9 | 16 | 34 |
| General Reading | 5 | 30 | 9 | 21 | 39 |
| General Reading | 6 | 35 | 8 | 27 | 43 |
| General Reading | 7 | 39 | 8 | 31 | 47 |
| General Reading | 8 | 44 | 8 | 36 | 52 |
| General Reading | 9 | 48 | 8 | 40 | 56 |
| General Reading | 10 | 53 | 9 | 44 | 62 |
| General Reading | 11 | 58 | 9 | 49 | 67 |
| General Reading | 12 | 64 | 10 | 54 | 74 |
| General Reading | 13 | 72 | 12 | 60 | 84 |
| General Reading | 14 | 84 | 16 | - | - |
| General Reading | 15 | 98 | 31 | - | - |
| Literary | 0 | 2 | 31 | - | - |
| Literary | 1 | 7 | 16 | - | - |
| Literary | 2 | 19 | 12 | 7 | 31 |
| Literary | 3 | 27 | 10 | 17 | 37 |
| Literary | 4 | 33 | 9 | 24 | 42 |
| Literary | 5 | 38 | 9 | 29 | 47 |
| Literary | 6 | 42 | 8 | 34 | 50 |
| Literary | 7 | 47 | 8 | 39 | 55 |
| Literary | 8 | 51 | 8 | 43 | 59 |
| Literary | 9 | 55 | 8 | 47 | 63 |
| Literary | 10 | 60 | 9 | 51 | 69 |
| Literary | 11 | 65 | 9 | 56 | 74 |
| Literary | 12 | 71 | 10 | 61 | 81 |
| Literary | 13 | 79 | 12 | 67 | 91 |
| Literary | 14 | 90 | 16 | - | - |
| Literary | 15 | 98 | 31 | - | - |
| Informational | 0 | 2 | 31 | - | - |
| Informational | 1 | 7 | 16 | - | - |
| Informational | 2 | 19 | 12 | 7 | 31 |
| Informational | 3 | 27 | 10 | 17 | 37 |
| Informational | 4 | 33 | 9 | 24 | 42 |
| Informational | 5 | 38 | 9 | 29 | 47 |
| Informational | 6 | 43 | 8 | 35 | 51 |
| Informational | 7 | 48 | 8 | 40 | 56 |
| Informational | 8 | 52 | 8 | 44 | 60 |
| Informational | 9 | 57 | 8 | 49 | 65 |
| Informational | 10 | 62 | 9 | 53 | 71 |
| Informational | 11 | 67 | 9 | 58 | 76 |
| Informational | 12 | 73 | 10 | 63 | 83 |
| Informational | 13 | 81 | 12 | 69 | 93 |
| Informational | 14 | 93 | 16 | - | - |
| Informational | 15 | 98 | 31 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Table 4.5.10. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 5

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 32 | - | - |
| General Reading | 1 | 2 | 16 | - | - |
| General Reading | 2 | 7 | 12 | - | - |
| General Reading | 3 | 15 | 10 | 5 | 25 |
| General Reading | 4 | 21 | 9 | 12 | 30 |
| General Reading | 5 | 26 | 9 | 17 | 35 |
| General Reading | 6 | 31 | 8 | 23 | 39 |
| General Reading | 7 | 35 | 8 | 27 | 43 |
| General Reading | 8 | 40 | 8 | 32 | 48 |
| General Reading | 9 | 44 | 8 | 36 | 52 |
| General Reading | 10 | 49 | 9 | 40 | 58 |
| General Reading | 11 | 54 | 9 | 45 | 63 |
| General Reading | 12 | 60 | 10 | 50 | 70 |
| General Reading | 13 | 68 | 12 | 56 | 80 |
| General Reading | 14 | 80 | 16 | 64 | 96 |
| General Reading | 15 | 98 | 32 | - | - |
| Literary | 0 | 2 | 32 | - | - |
| Literary | 1 | 6 | 16 | - | - |
| Literary | 2 | 18 | 12 | 6 | 30 |
| Literary | 3 | 26 | 10 | 16 | 36 |
| Literary | 4 | 32 | 9 | 23 | 41 |
| Literary | 5 | 38 | 9 | 29 | 47 |
| Literary | 6 | 42 | 9 | 33 | 51 |
| Literary | 7 | 47 | 8 | 39 | 55 |
| Literary | 8 | 51 | 8 | 43 | 59 |
| Literary | 9 | 56 | 9 | 47 | 65 |
| Literary | 10 | 61 | 9 | 52 | 70 |
| Literary | 11 | 66 | 9 | 57 | 75 |
| Literary | 12 | 72 | 10 | 62 | 82 |
| Literary | 13 | 80 | 12 | 68 | 92 |
| Literary | 14 | 93 | 17 | | |
| Literary | 15 | 98 | 32 | | |
| Informational | 0 | 2 | 32 | - | - |
| Informational | 1 | 11 | 16 | - | - |
| Informational | 2 | 23 | 12 | 11 | 35 |
| Informational | 3 | 31 | 10 | 21 | 41 |
| Informational | 4 | 37 | 9 | 28 | 46 |
| Informational | 5 | 43 | 9 | 34 | 52 |
| Informational | 6 | 47 | 9 | 38 | 56 |
| Informational | 7 | 52 | 8 | 44 | 60 |
| Informational | 8 | 56 | 8 | 48 | 64 |
| Informational | 9 | 61 | 9 | 52 | 70 |
| Informational | 10 | 66 | 9 | 57 | 75 |
| Informational | 11 | 71 | 9 | 62 | 80 |
| Informational | 12 | 77 | 10 | 67 | 87 |
| Informational | 13 | 85 | 12 | 73 | 97 |
| Informational | 14 | 97 | 16 | - | - |
| Informational | 15 | 98 | 32 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Table 4.5.11. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 6

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 37 | - | - |
| General Reading | 1 | 2 | 19 | - | - |
| General Reading | 2 | 7 | 14 | - | - |
| General Reading | 3 | 16 | 12 | 4 | 28 |
| General Reading | 4 | 23 | 11 | 12 | 34 |
| General Reading | 5 | 29 | 10 | 19 | 39 |
| General Reading | 6 | 35 | 10 | 25 | 45 |
| General Reading | 7 | 40 | 10 | 30 | 50 |
| General Reading | 8 | 46 | 10 | 36 | 56 |
| General Reading | 9 | 51 | 10 | 41 | 61 |
| General Reading | 10 | 57 | 11 | 46 | 68 |
| General Reading | 11 | 63 | 11 | 52 | 74 |
| General Reading | 12 | 71 | 12 | 59 | 83 |
| General Reading | 13 | 80 | 14 | 66 | 94 |
| General Reading | 14 | 95 | 19 | - | - |
| General Reading | 15 | 98 | 37 | - | - |
| Literary | 0 | 2 | 37 | - | - |
| Literary | 1 | 2 | 19 | - | - |
| Literary | 2 | 12 | 14 | - | - |
| Literary | 3 | 21 | 12 | 9 | 33 |
| Literary | 4 | 28 | 11 | 17 | 39 |
| Literary | 5 | 34 | 10 | 24 | 44 |
| Literary | 6 | 40 | 10 | 30 | 50 |
| Literary | 7 | 45 | 10 | 35 | 55 |
| Literary | 8 | 50 | 10 | 40 | 60 |
| Literary | 9 | 55 | 10 | 45 | 65 |
| Literary | 10 | 60 | 10 | 50 | 70 |
| Literary | 11 | 67 | 11 | 56 | 78 |
| Literary | 12 | 74 | 12 | 62 | 86 |
| Literary | 13 | 83 | 14 | 69 | 97 |
| Literary | 14 | 97 | 19 | - | - |
| Literary | 15 | 98 | 37 | - | - |
| Informational | 0 | 2 | 37 | - | - |
| Informational | 1 | 2 | 19 | - | - |
| Informational | 2 | 14 | 14 | - | - |
| Informational | 3 | 23 | 12 | 11 | 35 |
| Informational | 4 | 30 | 11 | 19 | 41 |
| Informational | 5 | 36 | 10 | 26 | 46 |
| Informational | 6 | 42 | 10 | 32 | 52 |
| Informational | 7 | 47 | 10 | 37 | 57 |
| Informational | 8 | 52 | 10 | 42 | 62 |
| Informational | 9 | 58 | 10 | 48 | 68 |
| Informational | 10 | 63 | 10 | 53 | 73 |
| Informational | 11 | 69 | 11 | 58 | 80 |
| Informational | 12 | 77 | 12 | 65 | 89 |
| Informational | 13 | 86 | 14 | - | - |
| Informational | 14 | 98 | 19 | - | - |
| Informational | 15 | 98 | 37 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Table 4.5.12. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 7

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 35 | - | - |
| General Reading | 1 | 2 | 18 | - | - |
| General Reading | 2 | 3 | 13 | - | - |
| General Reading | 3 | 12 | 11 | - | - |
| General Reading | 4 | 19 | 10 | 9 | 29 |
| General Reading | 5 | 25 | 10 | 15 | 35 |
| General Reading | 6 | 30 | 10 | 20 | 40 |
| General Reading | 7 | 36 | 9 | 27 | 45 |
| General Reading | 8 | 41 | 9 | 32 | 50 |
| General Reading | 9 | 46 | 10 | 36 | 56 |
| General Reading | 10 | 51 | 10 | 41 | 61 |
| General Reading | 11 | 58 | 11 | 47 | 69 |
| General Reading | 12 | 65 | 12 | 53 | 77 |
| General Reading | 13 | 73 | 13 | 60 | 86 |
| General Reading | 14 | 87 | 18 | - | - |
| General Reading | 15 | 98 | 35 | - | - |
| Literary | 0 | 2 | 35 | - | - |
| Literary | 1 | 2 | 18 | - | - |
| Literary | 2 | 9 | 13 | - | - |
| Literary | 3 | 17 | 11 | 6 | 28 |
| Literary | 4 | 24 | 10 | 14 | 34 |
| Literary | 5 | 30 | 10 | 20 | 40 |
| Literary | 6 | 35 | 9 | 26 | 44 |
| Literary | 7 | 40 | 9 | 31 | 49 |
| Literary | 8 | 45 | 9 | 36 | 54 |
| Literary | 9 | 50 | 9 | 41 | 59 |
| Literary | 10 | 55 | 10 | 45 | 65 |
| Literary | 11 | 61 | 10 | 51 | 71 |
| Literary | 12 | 68 | 11 | 57 | 79 |
| Literary | 13 | 76 | 13 | 63 | 89 |
| Literary | 14 | 90 | 18 | - | - |
| Literary | 15 | 98 | 35 | - | - |
| Informational | 0 | 2 | 35 | - | - |
| Informational | 1 | 4 | 18 | - | - |
| Informational | 2 | 18 | 13 | 5 | 31 |
| Informational | 3 | 26 | 11 | 15 | 37 |
| Informational | 4 | 33 | 10 | 23 | 43 |
| Informational | 5 | 39 | 10 | 29 | 49 |
| Informational | 6 | 44 | 9 | 35 | 53 |
| Informational | 7 | 49 | 9 | 40 | 58 |
| Informational | 8 | 54 | 9 | 45 | 63 |
| Informational | 9 | 58 | 9 | 49 | 67 |
| Informational | 10 | 64 | 10 | 54 | 74 |
| Informational | 11 | 69 | 10 | 59 | 79 |
| Informational | 12 | 76 | 11 | 65 | 87 |
| Informational | 13 | 84 | 13 | 71 | 97 |
| Informational | 14 | 98 | 18 | - | - |
| Informational | 15 | 98 | 35 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## Table 4.5.13. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 8

| Subscale Strand | Raw Score | Scale Score[1] (SS) | Standard Error (*SEM*) | (SS – 1*SEM*)[2] | (SS + 1*SEM*)[2] |
|---|---|---|---|---|---|
| General Reading | 0 | 2 | 33 | - | - |
| General Reading | 1 | 2 | 17 | - | - |
| General Reading | 2 | 8 | 13 | - | - |
| General Reading | 3 | 17 | 11 | 6 | 28 |
| General Reading | 4 | 23 | 10 | 13 | 33 |
| General Reading | 5 | 29 | 9 | 20 | 38 |
| General Reading | 6 | 34 | 9 | 25 | 43 |
| General Reading | 7 | 38 | 9 | 29 | 47 |
| General Reading | 8 | 43 | 9 | 34 | 52 |
| General Reading | 9 | 47 | 9 | 38 | 56 |
| General Reading | 10 | 52 | 9 | 43 | 61 |
| General Reading | 11 | 57 | 9 | 48 | 66 |
| General Reading | 12 | 62 | 10 | 52 | 72 |
| General Reading | 13 | 69 | 11 | 58 | 80 |
| General Reading | 14 | 77 | 13 | 64 | 90 |
| General Reading | 15 | 90 | 17 | - | - |
| General Reading | 16 | 98 | 33 | - | - |
| Literary | 0 | 2 | 33 | - | - |
| Literary | 1 | 2 | 17 | - | - |
| Literary | 2 | 12 | 13 | - | - |
| Literary | 3 | 20 | 11 | 9 | 31 |
| Literary | 4 | 27 | 10 | 17 | 37 |
| Literary | 5 | 32 | 9 | 23 | 41 |
| Literary | 6 | 37 | 9 | 28 | 46 |
| Literary | 7 | 42 | 9 | 33 | 51 |
| Literary | 8 | 47 | 9 | 38 | 56 |
| Literary | 9 | 52 | 9 | 43 | 61 |
| Literary | 10 | 58 | 10 | 48 | 68 |
| Literary | 11 | 64 | 11 | 53 | 75 |
| Literary | 12 | 72 | 13 | 59 | 85 |
| Literary | 13 | 85 | 17 | - | - |
| Literary | 14 | 98 | 33 | - | - |
| Informational | 0 | 2 | 34 | - | - |
| Informational | 1 | 2 | 17 | - | - |
| Informational | 2 | 11 | 13 | - | - |
| Informational | 3 | 20 | 11 | 9 | 31 |
| Informational | 4 | 26 | 10 | 16 | 36 |
| Informational | 5 | 32 | 10 | 22 | 42 |
| Informational | 6 | 38 | 9 | 29 | 47 |
| Informational | 7 | 43 | 9 | 34 | 52 |
| Informational | 8 | 48 | 9 | 39 | 57 |
| Informational | 9 | 52 | 9 | 43 | 61 |
| Informational | 10 | 58 | 9 | 49 | 67 |
| Informational | 11 | 63 | 10 | 53 | 73 |
| Informational | 12 | 70 | 11 | 59 | 81 |
| Informational | 13 | 78 | 13 | 65 | 91 |
| Informational | 14 | 91 | 17 | - | - |
| Informational | 15 | 98 | 33 | - | - |

*Note*. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

## 4.6. Score Interpretation

Interpretation of the 2010 Mod-MSA: Reading test scores depends primarily on the understanding of the scale score and the performance level descriptors.

### Scale Scores

As explained in section 4.5, Reporting Scale Scores for the 2010 Mod-MSA: Reading, the tests produced scale scores that ranged between 2 and 98. These scale scores have the same meaning within the same grade, but are not comparable across grade levels.

It should be noted that for scale scores, a higher score simply means a higher performance on the reading tests. Performance levels and descriptions can then be used to give specific interpretation to the scale scores because they are developed to bring meaning to those scale scores.

### Performance Level Descriptors

As explained previously, performance level descriptors provide specific information about students' performance levels and help interpret the 2010 Mod-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 (http://mdk12.org/instruction/curriculum/index.html):

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

The proficient levels described above were translated as classification scale score cuts through a standard setting procedure discussed in Appendix D.

## 4.7. Final Performance Level Cut Points for the Mod-MSA: Reading

For grade 3-5 a standard setting procedure was undertaken (see Appendix D) to obtain the cuts at the performance levels. The final cut points adopted by MSDE for the 2010 administration of the Mod-MSA: Reading test, grades 3-5 in raw score points, scale score, and theta metric were adjusted by the executive committee. There are two cut points that correspond to the three performance levels discussed above. Any score below the proficient cut point is the basic performance level.

Table 4.7.1 contains information about the cutoff scale score of each performance level. It should be noted that the same cutoff scores set by the standard setting procedure in 2009 for grades 6-8 were applied in 2010.

**Table 4.7.1 Mod-MSA: Reading Scale Score Cut Scores: Grades 3 through 8**

| Grade | Cut Scores at Performance Levels | |
|---|---|---|
| | Proficient | Advanced |
| 3 | 54 | 64 |
| 4 | 53 | 65 |
| 5 | 53 | 69 |
| 6 | 54 | 67 |
| 7 | 56 | 72 |
| 8 | 54 | 66 |

# 5. COMPARATIVE CLASSICAL AND IRT STATS ACROSS YEARS FOR THE COMMON ITEMS USED IN GRADES 6-8

The p-values of common items were compared across the two years (2009 and 2010) for Grades 6 to 8 to collect information about how much a possible year-to-year linking common item had changed in consecutive years with regard to item difficulty. Grades 3-5 had no common items as this was the first year of administration.

Tables 5.1 to 5.3 provide both the p-values and the Rasch difficulties of the common items across the two years of administration for grades 6-8. Figures 5.1, 5.3, and 5.5 graphically display the shifts in p-values for the common items across the two years. Figures 5.2, 5.4 and 5.6 display the same for the Rasch item difficulty parameters prior to placing them on the common scale.

.

**Table 5.1. P-Value and Unequated Rasch Difficulty Comparisons of Core Common Items Year 2009 with Year 2010: Grade 6**

| Item No | Item CID | P-Value 2009 | P-Value 2010 | Difference in P-Values | Rasch Difficulty 2009 | Rasch Difficulty 2010 |
|---|---|---|---|---|---|---|
| 1 | 100000213662 | 0.69 | 0.69 | 0.00 | -0.7867 | -0.6640 |
| 2 | 100000213663 | 0.78 | 0.76 | -0.02 | -1.2831 | -1.0811 |
| 3 | 100000213664 | 0.60 | 0.54 | -0.06 | -0.3530 | 0.0364 |
| 4 | 100000213665 | 0.63 | 0.65 | 0.02 | -0.4888 | -0.4854 |
| 5 | 100000213666 | 0.54 | 0.57 | 0.03 | -0.0595 | -0.1022 |
| 6 | 100000213668 | 0.65 | 0.68 | 0.03 | -0.5508 | -0.6483 |
| 7 | 100000213669 | 0.24 | 0.25 | 0.01 | 1.3285 | 1.4436 |
| 8 | 100000213670 | 0.71 | 0.77 | 0.06 | -0.8831 | -1.1560 |
| 9 | 100000213671 | 0.48 | 0.52 | 0.04 | 0.1881 | 0.1052 |
| 10 | 100000102192 | 0.62 | 0.60 | -0.02 | -0.4133 | -0.2714 |
| 11 | 100000102194 | 0.56 | 0.54 | -0.02 | -0.1792 | 0.0364 |
| 12 | 100000102191 | 0.67 | 0.68 | 0.01 | -0.6591 | -0.6172 |
| 13 | 100000102195 | 0.59 | 0.61 | 0.02 | -0.3109 | -0.3000 |
| 14 | 100000102196 | 0.47 | 0.44 | -0.03 | 0.2119 | 0.4491 |
| 15 | 100000102174 | 0.51 | 0.51 | 0.00 | 0.0695 | 0.1464 |
| 16 | 100000102175 | 0.34 | 0.33 | -0.01 | 0.8270 | 0.9642 |
| 17 | 100000102173 | 0.63 | 0.64 | 0.01 | -0.4815 | -0.4457 |
| 18 | 100000102176 | 0.41 | 0.43 | 0.02 | 0.4870 | 0.4908 |
| 19 | 100000102178 | 0.45 | 0.51 | 0.06 | 0.3073 | 0.1738 |
| 20 | 100000102136 | 0.65 | 0.70 | 0.05 | -0.5472 | -0.7221 |
| 21 | 100000102137 | 0.52 | 0.54 | 0.02 | 0.0153 | 0.0088 |
| 22 | 100000102139 | 0.60 | 0.60 | 0.00 | -0.3179 | -0.2571 |
| 23 | 100000102140 | 0.58 | 0.55 | -0.03 | -0.2517 | -0.0327 |
| Mean Rasch | | | | | -0.0109 | -0.0014 |

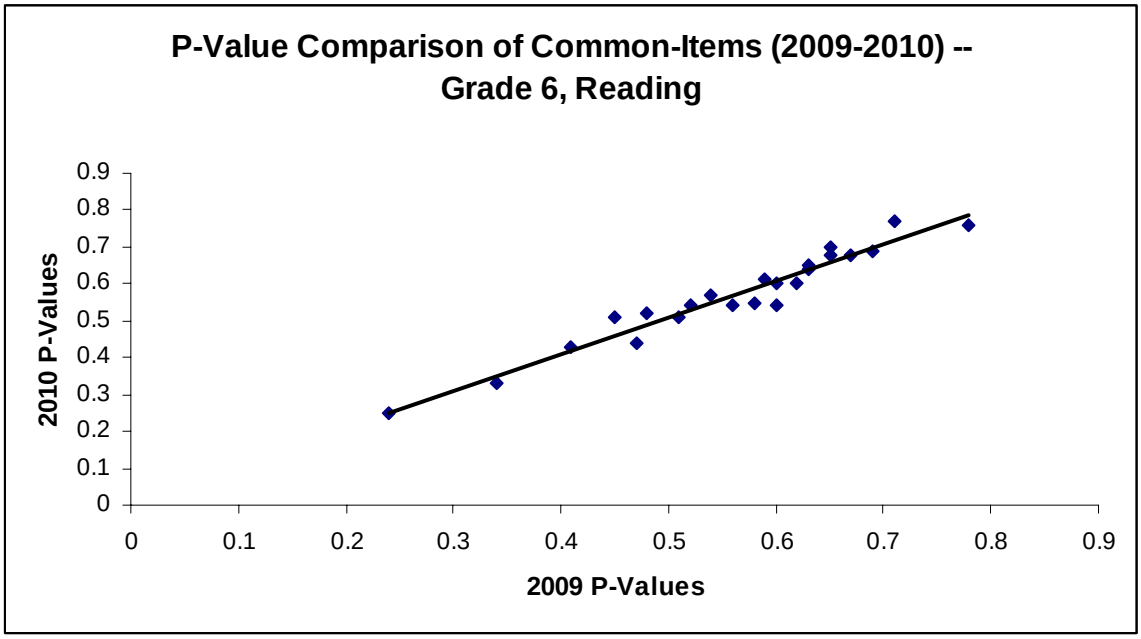



**Figure 5.1. P-values of the common items for Grade 6, years 2009 and 2010**

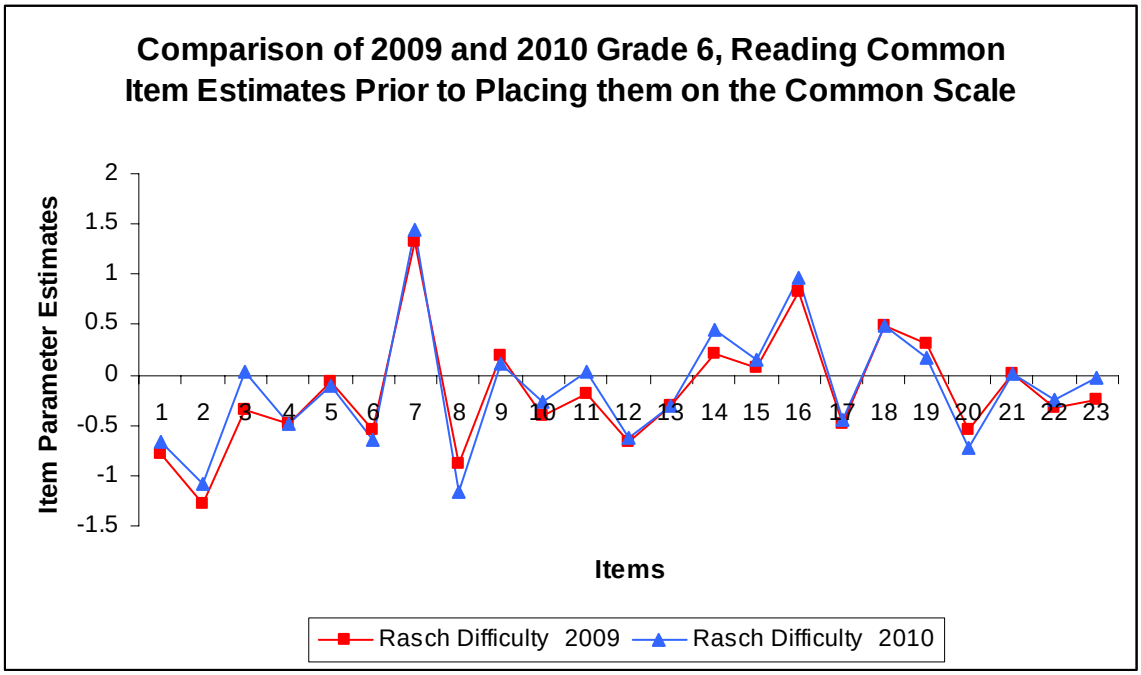



**Figure 5.2. Rasch difficulty parameters for the common items, Grade 6, years 2009 and 2010**

**Table 5.2. P-Value and Rasch Difficulty Comparisons of Core Linking Items Year 2009 with Year 2010: Grade 7**

| Item No. | Item CID | P-Value 2009 | P-Value 2010 | Difference in P-Values | Rasch Difficulty 2009 | Rasch Difficulty 2010 |
|---|---|---|---|---|---|---|
| 1 | 100000213674 | 0.69 | 0.67 | -0.02 | -0.7780 | -0.5090 |
| 2 | 100000213675 | 0.80 | 0.78 | -0.02 | -1.3540 | -1.1078 |
| 3 | 100000213676 | 0.34 | 0.35 | 0.01 | 0.8276 | 0.9922 |
| 4 | 100000213678 | 0.72 | 0.74 | 0.02 | -0.9143 | -0.8588 |
| 5 | 100000213679 | 0.71 | 0.68 | -0.03 | -0.8469 | -0.5615 |
| 6 | 100000102276 | 0.50 | 0.48 | -0.02 | 0.0908 | 0.3745 |
| 7 | 100000102277 | 0.36 | 0.34 | -0.02 | 0.7546 | 1.0260 |
| 8 | 100000102274 | 0.42 | 0.41 | -0.01 | 0.4769 | 0.7183 |
| 9 | 100000102278 | 0.51 | 0.46 | -0.05 | 0.0775 | 0.4595 |
| 10 | 100000102279 | 0.42 | 0.41 | -0.01 | 0.4687 | 0.7064 |
| 11 | 100000102263 | 0.55 | 0.58 | 0.03 | -0.1197 | -0.0485 |
| 12 | 100000102267 | 0.43 | 0.43 | 0.00 | 0.4089 | 0.6273 |
| 13 | 100000102268 | 0.52 | 0.50 | -0.02 | 0.0297 | 0.2819 |
| 14 | 100000102269 | 0.43 | 0.42 | -0.01 | 0.4035 | 0.6628 |
| 15 | 100000102266 | 0.37 | 0.33 | -0.04 | 0.6886 | 1.1033 |
| 16 | 100000102244 | 0.84 | 0.84 | 0.00 | -1.6963 | -1.4900 |
| 17 | 100000102246 | 0.57 | 0.58 | 0.01 | -0.1924 | -0.0761 |
| 18 | 100000102247 | 0.76 | 0.75 | -0.01 | -1.1534 | -0.9085 |
| 19 | 100000102248 | 0.52 | 0.51 | -0.01 | 0.0244 | 0.2742 |
| 20 | 100000102245 | 0.80 | 0.80 | 0.00 | -1.3660 | -1.2209 |
| 21 | 100000102218 | 0.48 | 0.50 | 0.02 | 0.1918 | 0.2973 |
| 22 | 100000102220 | 0.59 | 0.58 | -0.01 | -0.2739 | -0.0642 |
| 23 | 100000102221 | 0.49 | 0.47 | -0.02 | 0.1732 | 0.4208 |
| 24 | 100000102222 | 0.62 | 0.61 | -0.01 | -0.4429 | -0.1798 |
| 25 | 100000102217 | 0.64 | 0.63 | -0.01 | -0.5024 | -0.2937 |
| Mean Rasch | | | | | -0.2010 | 0.0250 |

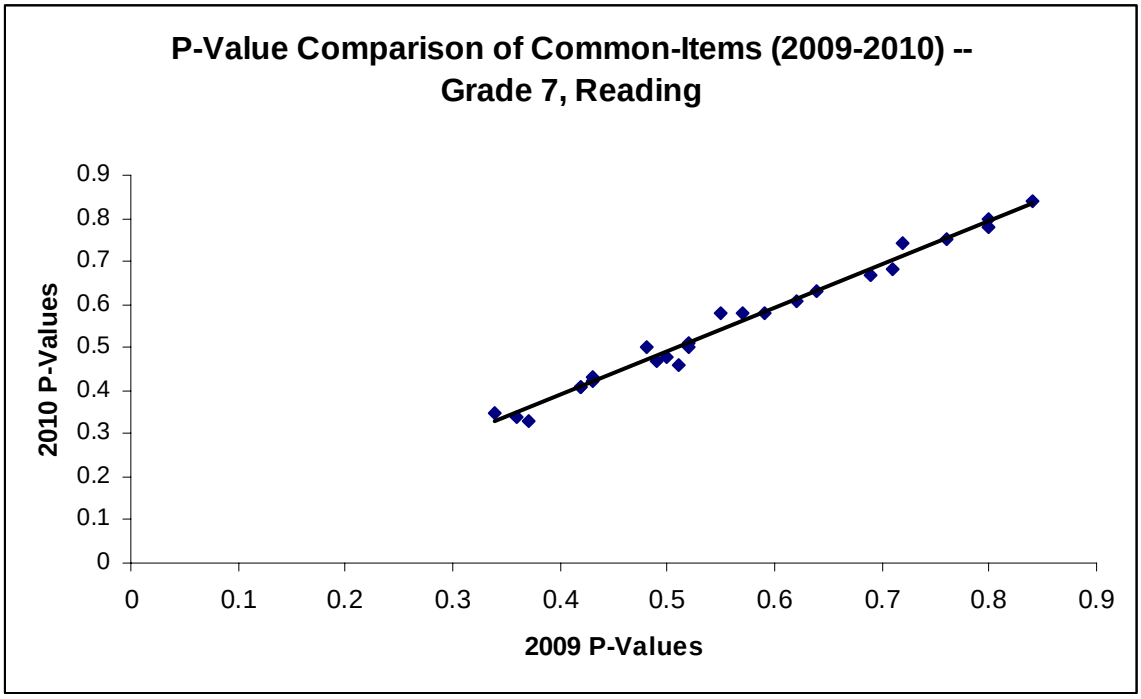



**Figure 5.3. P-values of the common items for Grade 7, years 2009 and 2010**

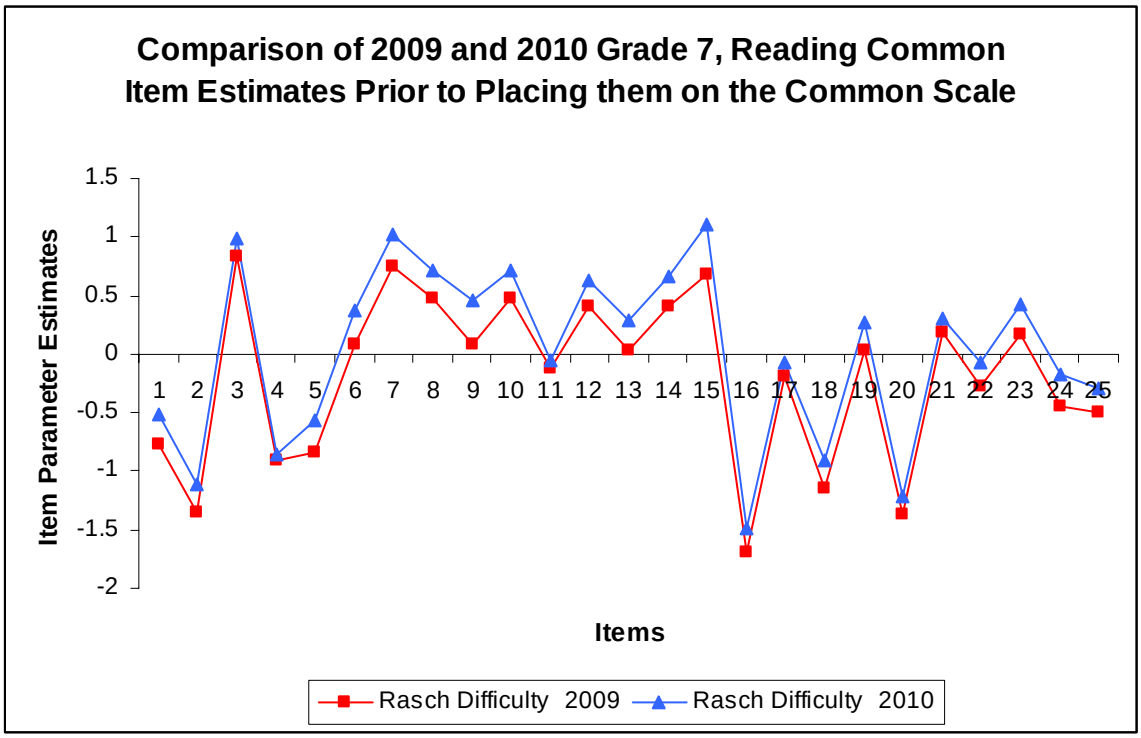



**Figure 5.4. Rasch difficulty parameters for the common items, Grade 7, years 2009 and 2010**

**Table 5.3. P-Value and Rasch Difficulty Comparisons of Core Linking Items Year 2009 with Year 2010: Grade 8**

| Item No. | Item CID | P-Value 2009 | P-Value 2010 | Difference in P-Values | Rasch Difficulty 2009 | Rasch Difficulty 2010 |
|---|---|---|---|---|---|---|
| 1 | 100000213683 | 0.80 | 0.83 | 0.03 | -1.1648 | -1.3041 |
| 2 | 100000213684 | 0.67 | 0.67 | 0.00 | -0.4197 | -0.3775 |
| 3 | 100000213686 | 0.48 | 0.50 | 0.02 | 0.4446 | 0.3926 |
| 4 | 100000213687 | 0.58 | 0.60 | 0.02 | 0.0045 | -0.0457 |
| 5 | 100000213688 | 0.63 | 0.63 | 0.00 | -0.1989 | -0.1547 |
| 6 | 100000102331 | 0.83 | 0.86 | 0.03 | -1.3174 | -1.5258 |
| 7 | 100000102329 | 0.52 | 0.56 | 0.04 | 0.2613 | 0.1423 |
| 8 | 100000102330 | 0.57 | 0.60 | 0.03 | 0.0618 | -0.0170 |
| 9 | 100000102334 | 0.51 | 0.47 | -0.04 | 0.3175 | 0.5589 |
| 10 | 100000102335 | 0.80 | 0.85 | 0.05 | -1.1686 | -1.4334 |
| 11 | 100000102341 | 0.64 | 0.67 | 0.03 | -0.2773 | -0.3581 |
| 12 | 100000102343 | 0.53 | 0.57 | 0.04 | 0.2393 | 0.1213 |
| 13 | 100000102344 | 0.70 | 0.70 | 0.00 | -0.5397 | -0.5208 |
| 14 | 100000102346 | 0.52 | 0.55 | 0.03 | 0.2686 | 0.1913 |
| 15 | 100000102345 | 0.60 | 0.63 | 0.03 | -0.0760 | -0.1806 |
| 16 | 100000102320 | 0.77 | 0.78 | 0.01 | -0.9453 | -0.9953 |
| 17 | 100000102322 | 0.64 | 0.65 | 0.01 | -0.2879 | -0.2740 |
| 18 | 100000102323 | 0.52 | 0.51 | -0.01 | 0.2760 | 0.3822 |
| 19 | 100000102324 | 0.57 | 0.62 | 0.05 | 0.0370 | -0.1072 |
| 20 | 100000102321 | 0.76 | 0.76 | 0.00 | -0.8923 | -0.8741 |
| 21 | 100000102304 | 0.52 | 0.57 | 0.05 | 0.2711 | 0.1002 |
| 22 | 100000102305 | 0.64 | 0.66 | 0.02 | -0.2458 | -0.3158 |
| 23 | 100000102302 | 0.48 | 0.53 | 0.05 | 0.4618 | 0.2783 |
| 24 | 100000102307 | 0.65 | 0.68 | 0.03 | -0.3358 | -0.4206 |
| 25 | 100000102308 | 0.62 | 0.66 | 0.04 | -0.1834 | -0.3043 |
| Mean Rasch | | | | | -0.2164 | -0.2817 |

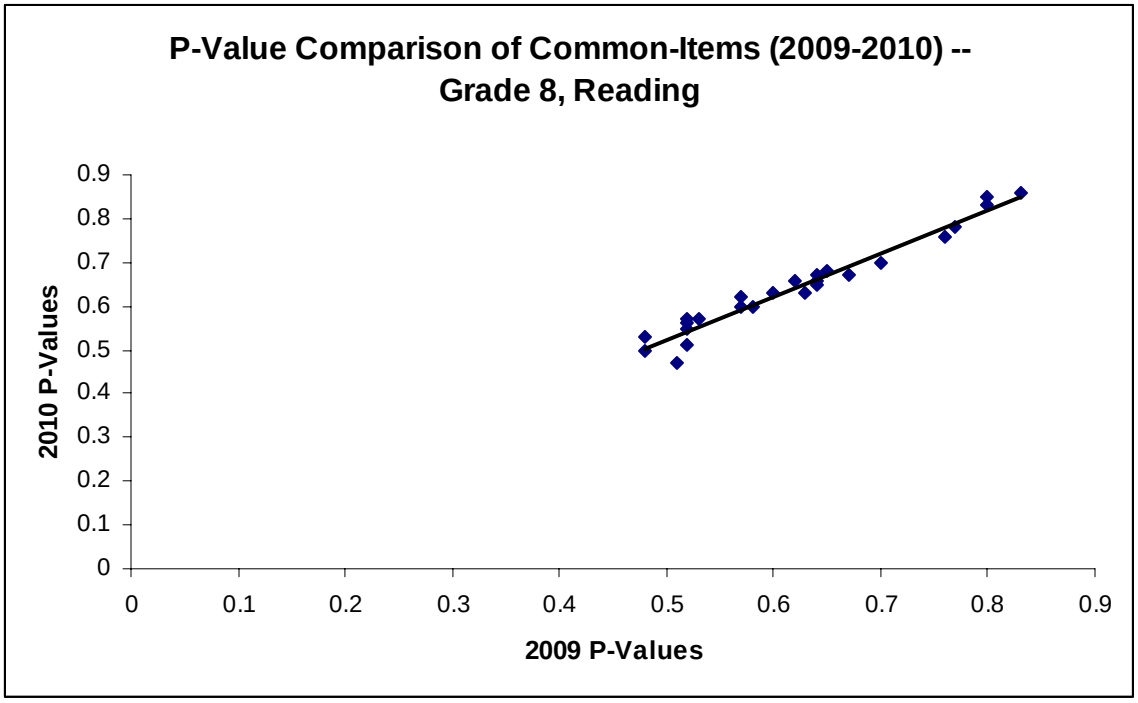



**Figure 5.5. P-values of the common items for Grade 8, years 2009 and 2010**

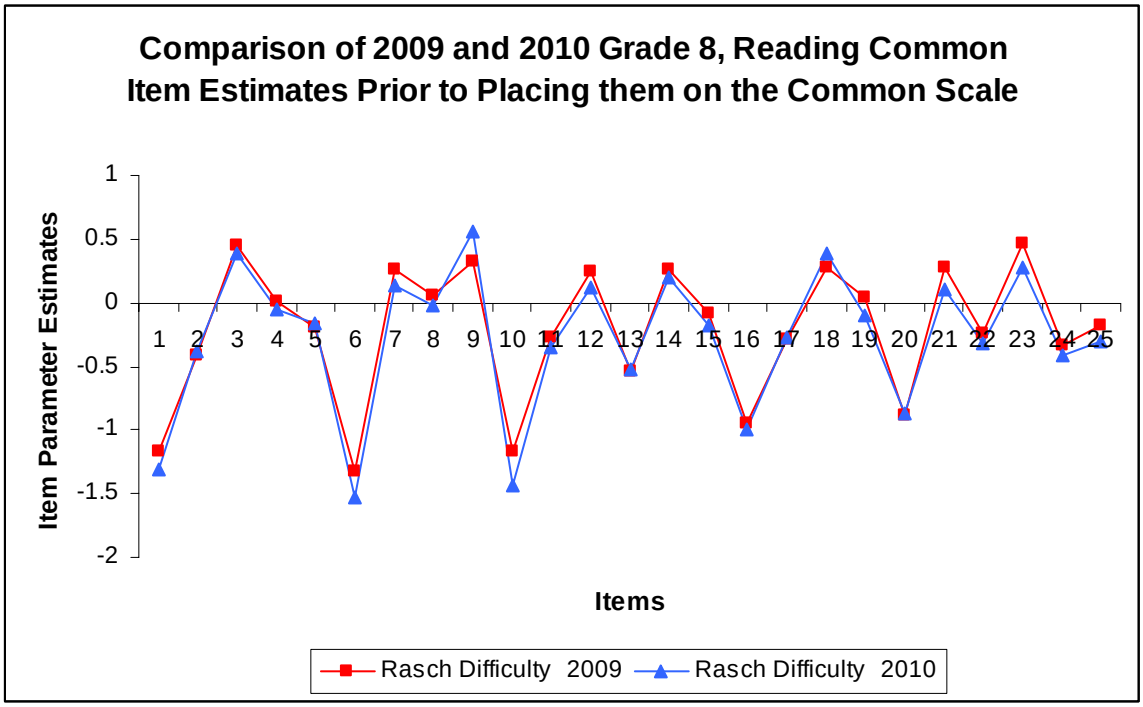



**Figure 5.6. Rasch difficulty parameters for the common items, Grade 8, years 2009 and 2010**

# 6. IRT ITEM LEVEL STATISTICS FOR THE 2010 MOD-MSA: READING

## 6.1 Rationale for the Use of the Rasch Model

In addition to reporting raw score summary statistics and item level statistics using the classical test theory (CTT), the items on the Mod-MSA: Reading test were also analyzed within the framework of item response theory (IRT). The Rasch model (Rasch, 1960) for dichotomous items was used for developing, scoring, and reporting the Mod-MSA: Reading assessment. These models were recommended for several reasons.

First, the sample size requirements for calibration, scaling, and equating under the Rasch model are significantly smaller than for other IRT models. For example, the Rasch model requires about 400 examinees per form for equating versus approximately 1,500 examinees per form under the 3PL IRT model (Kolen and Brennan, 2004, p. 288). This requirement for the Rasch model was particularly useful since the counts for the Mod-MSA were expected to be low.

Second, for the requirements of the Mod-MSA: Reading program, the Rasch model has one characteristic that makes it very useful. There exists a one-to-one relationship between raw scores and scale scores. That is, a student who answers a certain number of items correctly will receive the same scale score as a second student with the same raw score, regardless of which particular items within the test form were answered correctly.

## 6.2 Fit Statistics for the Rasch Model

Fit statistics are used for evaluating the goodness-of-fit of a model to the data. They are calculated by comparing the observed and expected trace lines obtained for an item after parameter estimates are obtained using a particular model. WINSTEPS provides two kinds of fit statistics called mean-squares that show the size of the randomness or amount of distortion of the measurement system.

The outfit and the infit statistics are used in order to ascertain the suitability of the data for constructing variables and making measures with the Rasch model. These fit statistics are mean square standardized residuals for item by person responses averaged over persons and partitioned between proficiency groups (outfit) and within proficiency groups (infit). When the observed item characteristic curve (ICC) departs from the expected ICC from a reference value of 1, there is an expectation of high proficiency students failing on an easy item or low proficiency students succeeding on a difficult one. The outfit mean square evaluates the agreement between the observed ICC and the best fitting Rasch model curve over the proficiency sub-groups. It is a standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's proficiency level. The infit, on the other hand, is a within-group mean square, which summarizes the degree of misfit remaining within proficiency groups after the between-group misfit has been removed from the total. The infit, therefore, is a standardized information-weighted mean square statistic, which is more sensitive to unexpected responses to items near the person's proficiency level.

Outfit mean-squares are influenced by outliers and are usually easy to diagnose and remedy. Infit mean-squares, on the other hand, are influenced by response patterns and are harder to diagnose and remedy. In general, mean-squares near 1.0 indicate little distortion of the measurement system, while values less than 1.0 indicate that observations are too predictable (redundancy,

model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit).

Generally speaking, when item fit indices are lower then 0.5, they do not discriminate well and show greater than expected degree of consistency. Similarly, a fit value higher than 1.5 indicates inconsistency in examinee scores on the item, i.e., some unexpectedly high scores are obtained by low proficiency candidates, and low scores for high proficiency candidates. Linacre and Wright (1999) provide an overall guideline for evaluating mean-square fit statistics (see Table 6.2.1, below).

**Table 6.2.1. Criteria to Evaluate Mean-Square Fit Statistics**

| Mean-Square | Interpretation |
|---|---|
| > 2.0 | Distorts or degrades the measurement system |
| 1.5 – 2.0 | Unproductive for construction of measurement, but not degrading |
| 0.5 – 1.5 | Productive for measurement |
| < 0.5 | Unproductive for measurement, but not degrading. May produce misleadingly good reliabilities and separations |

Note: Adapted from Linacre & Wright, 1999.

In our analysis, items were flagged if they distorted measurements, degraded the measurement system, or were unproductive for measurement, i.e., if the MSQ > 2.0 logits or MSQ < 0.5 logits.

### 6.3. Rasch Item Level Statistics

Table 6.3.1 to Table 6.3.6 provide both classical and IRT-based item parameters and includes:

- UIN
- Item sequence number
- Item strand number
- N-Count
- Rasch difficulty estimate
- Standard error of the Rasch difficulty (SE)
- Mean-square infit
- Mean-square outfit

## Table 6.3.1. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 3

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213631 | 1 | 208 | 813 | -0.9787 | 0.0822 | 0.93 | 0.93 |
| 100000360181 | 2 | 208 | 813 | 1.1187 | 0.0816 | 1.00 | 1.03 |
| 100000213633 | 3 | 208 | 813 | -1.3842 | 0.0902 | 0.84 | 0.69 |
| 100000213634 | 4 | 208 | 813 | 0.6841 | 0.0766 | 0.96 | 0.96 |
| 100000360182 | 5 | 208 | 813 | 0.4547 | 0.0751 | 0.98 | 0.97 |
| 100000101512 | 6 | 208 | 813 | 0.0840 | 0.0741 | 1.02 | 1.03 |
| 100000101513 | 7 | 208 | 813 | -0.1360 | 0.0745 | 1.02 | 1.02 |
| 100000101514 | 8 | 208 | 813 | -0.4404 | 0.0760 | 0.93 | 0.89 |
| 100000360183 | 9 | 208 | 813 | -0.2249 | 0.0748 | 1.02 | 1.02 |
| 100000101516 | 10 | 208 | 813 | 0.4886 | 0.0753 | 1.05 | 1.07 |
| 100000101517 | 11 | 208 | 813 | 0.9633 | 0.0795 | 1.04 | 1.07 |
| 100000101518 | 12 | 208 | 813 | -0.1582 | 0.0745 | 1.05 | 1.06 |
| 100000360184 | 13 | 208 | 813 | 0.2103 | 0.0742 | 1.04 | 1.03 |
| 100000101969 | 14 | 208 | 813 | -0.7516 | 0.0791 | 0.86 | 0.80 |
| 100000101970 | 15 | 208 | 813 | 0.1005 | 0.0741 | 0.94 | 0.94 |
| 100000101971 | 16 | 209 | 813 | 0.5568 | 0.0757 | 1.04 | 1.05 |
| 100000101972 | 17 | 209 | 813 | 0.3097 | 0.0745 | 1.10 | 1.11 |
| 100000101973 | 18 | 209 | 813 | 0.7253 | 0.0770 | 1.06 | 1.11 |
| 100000101974 | 19 | 209 | 813 | 0.4210 | 0.0749 | 0.97 | 0.97 |
| 100000101975 | 20 | 209 | 813 | 0.3541 | 0.0746 | 1.09 | 1.10 |
| 100000101932 | 21 | 208 | 813 | -0.2753 | 0.0750 | 1.00 | 1.00 |
| 100000101934 | 22 | 209 | 813 | 0.0401 | 0.0741 | 0.98 | 0.99 |
| 100000101935 | 23 | 209 | 813 | 0.5682 | 0.0758 | 1.15 | 1.16 |
| 100000101933 | 24 | 208 | 813 | 0.7312 | 0.0770 | 1.03 | 1.05 |
| 100000101936 | 25 | 209 | 813 | 0.0072 | 0.0742 | 1.05 | 1.03 |
| 100000101937 | 26 | 209 | 813 | -0.4984 | 0.0765 | 0.94 | 0.92 |
| 100000101938 | 27 | 209 | 813 | 0.2268 | 0.0743 | 0.99 | 0.99 |
| 100000346452 | 28 | 208 | 813 | -0.3545 | 0.0754 | 0.98 | 0.97 |
| 100000346453 | 29 | 208 | 813 | -0.2978 | 0.0751 | 0.93 | 0.92 |
| 100000346444 | 30 | 209 | 813 | 0.4042 | 0.0748 | 1.01 | 1.02 |
| 100000346446 | 31 | 209 | 813 | 0.5912 | 0.0759 | 1.15 | 1.17 |
| 100000346450 | 32 | 209 | 813 | -1.0333 | 0.0831 | 0.90 | 0.80 |
| 100000346445 | 33 | 209 | 813 | 0.2931 | 0.0744 | 0.93 | 0.93 |
| 100000346448 | 34 | 209 | 813 | 0.1499 | 0.0742 | 1.01 | 1.00 |
| 100000260458 | 35 | 209 | 813 | 0.4942 | 0.0753 | 1.07 | 1.10 |
| 100000260460 | 36 | 209 | 813 | -0.2978 | 0.0751 | 1.04 | 1.04 |

Table A. 6.3.1. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 3 (Continued)

| Item CID | Item Seq. No.[1] | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000260461 | 37 | 209 | 813 | -0.1693 | 0.0746 | 0.98 | 1.02 |
| 100000312661 | 38 | 208 | 813 | -0.2585 | 0.0749 | 1.00 | 0.99 |
| 100000260457 | 39 | 209 | 813 | 0.0511 | 0.0741 | 1.06 | 1.06 |
| 100000260459 | 40 | 209 | 813 | 0.7253 | 0.0770 | 1.07 | 1.09 |
| 100000260465 | 41 | 208 | 813 | -0.2585 | 0.0749 | 1.01 | 1.00 |
| 100000101905 | 42 | 208 | 813 | -1.0403 | 0.0833 | 0.92 | 0.85 |
| 100000101906 | 43 | 208 | 813 | 0.5625 | 0.0757 | 1.00 | 1.02 |
| 100000101907 | 44 | 212 | 813 | 0.1883 | 0.0742 | 0.97 | 0.98 |
| 100000101908 | 45 | 212 | 813 | -0.8661 | 0.0806 | 0.85 | 0.78 |
| 100000101909 | 46 | 212 | 813 | -0.2417 | 0.0749 | 0.93 | 0.91 |
| 100000101910 | 47 | 212 | 813 | -0.1582 | 0.0745 | 1.16 | 1.22 |
| 100000101911 | 48 | 212 | 813 | -0.8211 | 0.0799 | 0.92 | 0.88 |
| 100000101529 | 49 | 212 | 813 | -1.5996 | 0.0957 | 0.89 | 0.78 |
| 100000101530 | 50 | 212 | 813 | -0.5632 | 0.0770 | 0.88 | 0.84 |
| 100000101531 | 51 | 212 | 813 | 1.3922 | 0.0862 | 1.12 | 1.28 |
| 100000101532 | 52 | 212 | 813 | 0.2820 | 0.0744 | 1.11 | 1.13 |
| 100000101527 | 53 | 208 | 813 | -0.3488 | 0.0754 | 0.96 | 0.95 |
| 100000101528 | 54 | 208 | 813 | 0.0017 | 0.0742 | 0.96 | 0.94 |
| 100000101533 | 55 | 212 | 813 | -0.7956 | 0.0796 | 0.96 | 0.95 |
| 100000260338 | 56 | 208 | 813 | -0.5513 | 0.0769 | 1.02 | 1.11 |
| 100000260342 | 57 | 209 | 813 | 0.0621 | 0.0741 | 0.98 | 0.97 |
| 100000260345 | 58 | 212 | 813 | 0.9696 | 0.0796 | 1.07 | 1.16 |
| 100000365154 | 59 | 212 | 813 | 0.3708 | 0.0747 | 1.02 | 1.02 |
| 100000260349 | 60 | 209 | 813 | 1.4755 | 0.0879 | 1.06 | 1.17 |
| 100000260339 | 61 | 212 | 813 | -0.1804 | 0.0746 | 0.95 | 0.92 |
| 100000300707 | 62 | 208 | 813 | 0.5511 | 0.0756 | 1.06 | 1.07 |
| 100000260364 | 63 | 212 | 813 | 0.6665 | 0.0765 | 1.17 | 1.23 |
| 100000260368 | 64 | 212 | 813 | -0.6775 | 0.0782 | 0.93 | 0.94 |
| 100000260371 | 65 | 212 | 813 | 0.0566 | 0.0741 | 1.09 | 1.09 |
| 100000300557 | 66 | 212 | 813 | 0.5912 | 0.0759 | 1.00 | 1.01 |
| 100000260372 | 67 | 208 | 813 | -1.0333 | 0.0831 | 0.87 | 0.80 |
| 100000260365 | 68 | 212 | 813 | -0.9719 | 0.0821 | 0.94 | 0.86 |
| 100000300552 | 69 | 208 | 813 | -0.5572 | 0.0770 | 0.86 | 0.82 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

## Table 6.3.2. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 4

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213644 | 1 | 208 | 967 | 0.7037 | 0.0705 | 1.07 | 1.11 |
| 100000213645 | 2 | 208 | 967 | 0.0750 | 0.0677 | 1.08 | 1.10 |
| 100000213646 | 3 | 208 | 967 | 0.2075 | 0.0679 | 1.10 | 1.12 |
| 100000213647 | 4 | 208 | 967 | -0.3436 | 0.0686 | 0.95 | 0.95 |
| 100000360190 | 5 | 208 | 967 | 1.3628 | 0.0792 | 1.05 | 1.20 |
| 100000213637 | 6 | 208 | 967 | -0.8113 | 0.0725 | 0.90 | 0.85 |
| 100000213638 | 7 | 208 | 967 | -0.8061 | 0.0724 | 0.91 | 0.87 |
| 100000213639 | 8 | 208 | 967 | 0.6938 | 0.0704 | 1.10 | 1.12 |
| 100000360191 | 9 | 208 | 967 | 0.5962 | 0.0697 | 1.00 | 1.01 |
| 100000213641 | 10 | 208 | 967 | -1.2995 | 0.0799 | 0.91 | 0.82 |
| 100000213642 | 11 | 208 | 967 | -0.6477 | 0.0708 | 0.90 | 0.86 |
| 100000360192 | 12 | 208 | 967 | -1.0009 | 0.0749 | 0.89 | 0.84 |
| 100000357134 | 13 | 209 | 967 | -0.7234 | 0.0715 | 0.95 | 0.91 |
| 100000357132 | 14 | 208 | 967 | 0.2672 | 0.0680 | 1.05 | 1.06 |
| 100000357133 | 15 | 208 | 967 | -1.0692 | 0.0759 | 0.91 | 0.84 |
| 100000357135 | 16 | 209 | 967 | 0.3041 | 0.0681 | 1.11 | 1.14 |
| 100000357136 | 17 | 209 | 967 | -0.2365 | 0.0682 | 0.93 | 0.92 |
| 100000357137 | 18 | 209 | 967 | -0.2041 | 0.0681 | 0.93 | 0.91 |
| 100000357138 | 19 | 209 | 967 | 0.4532 | 0.0688 | 0.99 | 1.00 |
| 100000357106 | 20 | 209 | 967 | -0.0254 | 0.0677 | 1.07 | 1.09 |
| 100000357104 | 21 | 208 | 967 | -0.5981 | 0.0703 | 0.92 | 0.89 |
| 100000357107 | 22 | 209 | 967 | -0.4096 | 0.0690 | 0.94 | 0.92 |
| 100000357105 | 23 | 208 | 967 | -0.5150 | 0.0697 | 0.96 | 0.98 |
| 100000357108 | 24 | 209 | 967 | 0.0978 | 0.0677 | 0.99 | 0.98 |
| 100000357109 | 25 | 209 | 967 | -1.0291 | 0.0753 | 0.92 | 0.87 |
| 100000357110 | 26 | 209 | 967 | -0.3249 | 0.0685 | 1.05 | 1.07 |
| 100000260487 | 27 | 209 | 967 | 0.4626 | 0.0688 | 1.13 | 1.15 |
| 100000260486 | 28 | 209 | 967 | 1.2707 | 0.0776 | 0.97 | 0.98 |
| 100000260483 | 29 | 208 | 967 | -0.2411 | 0.0682 | 0.99 | 0.98 |
| 100000260481 | 30 | 208 | 967 | 0.0476 | 0.0677 | 0.94 | 0.93 |
| 100000260489 | 31 | 209 | 967 | -0.5735 | 0.0701 | 0.99 | 1.02 |
| 100000260488 | 32 | 209 | 967 | 1.2055 | 0.0765 | 0.98 | 1.02 |
| 100000260492 | 33 | 209 | 967 | 1.1306 | 0.0754 | 1.05 | 1.10 |
| 100000269896 | 34 | 209 | 967 | -0.1489 | 0.0679 | 0.97 | 0.96 |
| 100000271197 | 35 | 208 | 967 | 0.6155 | 0.0698 | 0.94 | 0.93 |
| 100000271198 | 36 | 208 | 967 | -0.2596 | 0.0682 | 0.95 | 0.93 |

Table 6.3.2. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 4 (Continued)

| Item CID | Item Seq. No.[1] | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000269900 | 37 | 209 | 967 | 0.8244 | 0.0716 | 1.07 | 1.11 |
| 100000269897 | 38 | 209 | 967 | 0.2350 | 0.0679 | 1.05 | 1.06 |
| 100000269898 | 39 | 209 | 967 | 0.4673 | 0.0688 | 1.00 | 1.00 |
| 100000269899 | 40 | 209 | 967 | 0.4156 | 0.0686 | 1.01 | 1.01 |
| 100000357097 | 41 | 212 | 967 | 1.1649 | 0.0759 | 1.11 | 1.17 |
| 100000357095 | 42 | 208 | 967 | -0.3954 | 0.0689 | 0.99 | 0.97 |
| 100000357098 | 43 | 212 | 967 | 0.8553 | 0.0720 | 1.02 | 1.03 |
| 100000462160 | 44 | 212 | 967 | 0.9548 | 0.0731 | 1.16 | 1.25 |
| 100000357100 | 45 | 212 | 967 | 0.5337 | 0.0692 | 1.09 | 1.10 |
| 100000357101 | 46 | 212 | 967 | -0.3483 | 0.0686 | 1.00 | 0.97 |
| 100000462157 | 47 | 208 | 967 | -0.2969 | 0.0684 | 0.95 | 0.94 |
| 100000101997 | 48 | 208 | 967 | -0.0391 | 0.0677 | 1.03 | 1.03 |
| 100000101996 | 49 | 212 | 967 | 0.3690 | 0.0684 | 0.95 | 0.95 |
| 100000101998 | 50 | 212 | 967 | 0.3411 | 0.0683 | 1.04 | 1.04 |
| 100000101999 | 51 | 212 | 967 | 0.4109 | 0.0686 | 1.01 | 1.03 |
| 100000102000 | 52 | 212 | 967 | -0.0299 | 0.0677 | 1.03 | 1.03 |
| 100000102001 | 53 | 212 | 967 | 0.1069 | 0.0677 | 0.98 | 0.98 |
| 100000200070 | 54 | 208 | 967 | 0.0020 | 0.0677 | 0.97 | 0.97 |
| 100000102025 | 55 | 208 | 967 | 0.9870 | 0.0735 | 1.07 | 1.12 |
| 100000102026 | 56 | 208 | 967 | 0.3318 | 0.0682 | 1.07 | 1.10 |
| 100000102027 | 57 | 212 | 967 | -0.0893 | 0.0678 | 0.93 | 0.91 |
| 100000102023 | 58 | 212 | 967 | 0.1846 | 0.0678 | 1.00 | 1.00 |
| 100000102024 | 59 | 212 | 967 | -0.8166 | 0.0725 | 1.01 | 1.07 |
| 100000102028 | 60 | 212 | 967 | 0.2350 | 0.0679 | 0.95 | 0.94 |
| 100000102029 | 61 | 212 | 967 | -0.2782 | 0.0683 | 0.96 | 0.96 |
| 100000301035 | 62 | 208 | 967 | -0.7234 | 0.0715 | 0.90 | 0.86 |
| 100000267470 | 63 | 208 | 967 | -0.5393 | 0.0699 | 0.91 | 0.87 |
| 100000267467 | 64 | 212 | 967 | -1.7952 | 0.0918 | 0.91 | 0.75 |
| 100000301034 | 65 | 212 | 967 | -0.8324 | 0.0727 | 0.94 | 0.94 |
| 100000267473 | 66 | 212 | 967 | 0.0476 | 0.0677 | 1.04 | 1.06 |
| 100000267474 | 67 | 212 | 967 | -0.0802 | 0.0678 | 1.13 | 1.16 |
| 100000267472 | 68 | 212 | 967 | -0.4287 | 0.0691 | 0.98 | 0.98 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

## Table 6.3.3. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 5

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213650 | 1 | 208 | 1043 | -0.7119 | 0.0682 | 0.95 | 0.94 |
| 100000213651 | 2 | 208 | 1043 | -0.8400 | 0.0696 | 0.93 | 0.94 |
| 100000213652 | 3 | 208 | 1043 | -0.7919 | 0.0690 | 0.96 | 0.94 |
| 100000213653 | 4 | 208 | 1043 | -1.0097 | 0.0718 | 0.90 | 0.86 |
| 100000360196 | 5 | 208 | 1043 | 0.5530 | 0.0663 | 1.04 | 1.04 |
| 100000213655 | 6 | 208 | 1043 | -1.0357 | 0.0721 | 0.94 | 0.89 |
| 100000213656 | 7 | 208 | 1043 | -1.3124 | 0.0768 | 0.88 | 0.79 |
| 100000213657 | 8 | 208 | 1043 | -0.8400 | 0.0696 | 0.90 | 0.85 |
| 100000213658 | 9 | 208 | 1043 | 0.0939 | 0.0646 | 1.09 | 1.10 |
| 100000213659 | 10 | 208 | 1043 | -0.6382 | 0.0675 | 0.96 | 0.93 |
| 100000360197 | 11 | 208 | 1043 | -0.4994 | 0.0664 | 0.94 | 0.95 |
| 100000102095 | 12 | 209 | 1043 | -0.0061 | 0.0646 | 0.99 | 0.99 |
| 100000102093 | 13 | 208 | 1043 | -1.0513 | 0.0724 | 0.93 | 0.88 |
| 100000102094 | 14 | 208 | 1043 | 0.8634 | 0.0691 | 0.99 | 1.00 |
| 100000102096 | 15 | 209 | 1043 | 0.0314 | 0.0645 | 1.01 | 1.02 |
| 100000102097 | 16 | 209 | 1043 | 0.6284 | 0.0669 | 1.10 | 1.14 |
| 100000102098 | 17 | 209 | 1043 | 0.6195 | 0.0668 | 1.06 | 1.10 |
| 100000102099 | 18 | 209 | 1043 | 0.7509 | 0.0679 | 1.05 | 1.08 |
| 100000102113 | 19 | 209 | 1043 | -0.0061 | 0.0646 | 0.93 | 0.92 |
| 100000102111 | 20 | 208 | 1043 | -0.8839 | 0.0701 | 0.92 | 0.87 |
| 100000102114 | 21 | 209 | 1043 | -0.0103 | 0.0646 | 0.96 | 0.95 |
| 100000102115 | 22 | 209 | 1043 | 0.7555 | 0.0680 | 0.98 | 0.98 |
| 100000102116 | 23 | 209 | 1043 | 0.0314 | 0.0645 | 0.97 | 0.96 |
| 100000102112 | 24 | 208 | 1043 | -0.9637 | 0.0711 | 0.95 | 0.95 |
| 100000102117 | 25 | 209 | 1043 | 0.7096 | 0.0675 | 1.10 | 1.14 |
| 100000102104 | 26 | 209 | 1043 | -0.0854 | 0.0646 | 1.00 | 0.99 |
| 100000102102 | 27 | 208 | 1043 | -0.4120 | 0.0658 | 1.00 | 1.01 |
| 100000102103 | 28 | 208 | 1043 | -0.4599 | 0.0661 | 0.99 | 1.00 |
| 100000102105 | 29 | 209 | 1043 | 0.8587 | 0.0690 | 1.07 | 1.12 |
| 100000102106 | 30 | 209 | 1043 | -0.6201 | 0.0673 | 0.96 | 0.96 |
| 100000102107 | 31 | 209 | 1043 | 0.8255 | 0.0687 | 1.00 | 1.02 |
| 100000102108 | 32 | 209 | 1043 | 0.4832 | 0.0659 | 1.01 | 1.03 |
| 100000102086 | 33 | 209 | 1043 | 0.6061 | 0.0667 | 1.10 | 1.14 |
| 100000102084 | 34 | 208 | 1043 | -0.4994 | 0.0664 | 1.00 | 0.99 |
| 100000102087 | 35 | 209 | 1043 | 0.3757 | 0.0653 | 1.06 | 1.07 |
| 100000102088 | 36 | 209 | 1043 | 0.9851 | 0.0705 | 1.10 | 1.11 |

Table 6.3.3. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 5 (Continued)

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000102089 | 37 | 209 | 1043 | 1.3313 | 0.0758 | 1.13 | 1.28 |
| 100000102085 | 38 | 208 | 1043 | -0.1734 | 0.0648 | 1.05 | 1.06 |
| 100000102090 | 39 | 209 | 1043 | 0.5618 | 0.0664 | 0.99 | 0.99 |
| 100000102046 | 40 | 208 | 1043 | -1.3482 | 0.0775 | 0.88 | 0.78 |
| 100000102048 | 41 | 212 | 1043 | -0.1734 | 0.0648 | 1.03 | 1.04 |
| 100000102049 | 42 | 212 | 1043 | 0.8208 | 0.0686 | 1.09 | 1.14 |
| 100000102050 | 43 | 212 | 1043 | -0.8497 | 0.0697 | 0.98 | 0.97 |
| 100000102051 | 44 | 212 | 1043 | 0.1732 | 0.0647 | 1.04 | 1.05 |
| 100000102052 | 45 | 208 | 1043 | -0.5437 | 0.0667 | 0.96 | 0.96 |
| 100000102047 | 46 | 212 | 1043 | -0.4511 | 0.0661 | 0.93 | 0.91 |
| 100000102056 | 47 | 208 | 1043 | 0.1816 | 0.0647 | 1.03 | 1.05 |
| 100000102058 | 48 | 212 | 1043 | 1.6502 | 0.0824 | 1.11 | 1.34 |
| 100000102059 | 49 | 212 | 1043 | -0.3903 | 0.0657 | 0.98 | 0.96 |
| 100000102060 | 50 | 212 | 1043 | 0.6553 | 0.0671 | 1.01 | 1.02 |
| 100000102061 | 51 | 212 | 1043 | 0.0397 | 0.0645 | 1.00 | 1.00 |
| 100000102057 | 52 | 208 | 1043 | 0.5399 | 0.0662 | 1.06 | 1.07 |
| 100000102062 | 53 | 212 | 1043 | 1.1125 | 0.0723 | 1.08 | 1.18 |
| 100000102068 | 54 | 212 | 1043 | -0.2875 | 0.0652 | 0.96 | 0.95 |
| 100000102069 | 55 | 212 | 1043 | 0.6284 | 0.0669 | 0.94 | 0.95 |
| 100000102066 | 56 | 208 | 1043 | -0.1356 | 0.0647 | 1.03 | 1.03 |
| 100000102067 | 57 | 208 | 1043 | -0.7446 | 0.0685 | 0.93 | 0.89 |
| 100000102070 | 58 | 212 | 1043 | 1.2867 | 0.0750 | 1.14 | 1.33 |
| 100000102071 | 59 | 212 | 1043 | 0.1816 | 0.0647 | 1.06 | 1.08 |
| 100000102072 | 60 | 212 | 1043 | -0.4950 | 0.0663 | 0.92 | 0.89 |
| 100000267477 | 61 | 212 | 1043 | 0.1648 | 0.0647 | 0.99 | 0.98 |
| 100000268378 | 62 | 212 | 1043 | 0.9505 | 0.0701 | 0.97 | 0.98 |
| 100000267485 | 63 | 208 | 1043 | 0.5883 | 0.0666 | 0.96 | 0.95 |
| 100000268380 | 64 | 212 | 1043 | 0.3289 | 0.0651 | 0.97 | 0.96 |
| 100000267481 | 65 | 212 | 1043 | 1.0353 | 0.0712 | 1.03 | 1.05 |
| 100000301339 | 66 | 208 | 1043 | 0.1648 | 0.0647 | 1.01 | 1.02 |
| 100000303033 | 67 | 212 | 1043 | -0.3602 | 0.0655 | 0.91 | 0.89 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

## Table 6.3.4. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 6

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213662 | 1 | 208 | 975 | -0.7867 | 0.0731 | 0.91 | 0.88 |
| 100000213663 | 2 | 208 | 975 | -1.2831 | 0.0810 | 0.91 | 0.82 |
| 100000213664 | 3 | 208 | 975 | -0.3530 | 0.0692 | 1.06 | 1.08 |
| 100000213665 | 4 | 208 | 975 | -0.4888 | 0.0701 | 0.89 | 0.86 |
| 100000213666 | 5 | 208 | 975 | -0.0595 | 0.0679 | 1.06 | 1.07 |
| 100000360201 | 6 | 208 | 975 | 1.7291 | 0.0846 | 1.05 | 1.15 |
| 100000213668 | 7 | 208 | 975 | -0.5508 | 0.0707 | 0.89 | 0.86 |
| 100000213669 | 8 | 208 | 975 | 1.3285 | 0.0769 | 1.00 | 1.06 |
| 100000213670 | 9 | 208 | 975 | -0.8831 | 0.0744 | 0.82 | 0.77 |
| 100000213671 | 10 | 208 | 975 | 0.1881 | 0.0677 | 1.00 | 1.00 |
| 100000360202 | 11 | 208 | 975 | -0.7472 | 0.0727 | 0.92 | 0.92 |
| 100000102192 | 12 | 209 | 975 | -0.4133 | 0.0696 | 0.91 | 0.88 |
| 100000102194 | 13 | 209 | 975 | -0.1792 | 0.0683 | 1.06 | 1.08 |
| 100000102191 | 14 | 208 | 975 | -0.6591 | 0.0717 | 0.90 | 0.87 |
| 100000102195 | 15 | 209 | 975 | -0.3109 | 0.0689 | 1.00 | 1.02 |
| 100000102196 | 16 | 209 | 975 | 0.2119 | 0.0677 | 1.02 | 1.03 |
| 100000102174 | 17 | 209 | 975 | 0.0695 | 0.0677 | 1.09 | 1.11 |
| 100000102175 | 18 | 209 | 975 | 0.8270 | 0.0707 | 1.03 | 1.12 |
| 100000102173 | 19 | 208 | 975 | -0.4815 | 0.0701 | 0.95 | 0.94 |
| 100000102176 | 20 | 209 | 975 | 0.4870 | 0.0684 | 1.01 | 1.02 |
| 100000102178 | 21 | 209 | 975 | 0.3073 | 0.0679 | 1.07 | 1.09 |
| 100000269999 | 22 | 208 | 975 | -1.8723 | 0.0958 | 0.89 | 0.74 |
| 100000269997 | 23 | 209 | 975 | -0.1412 | 0.0682 | 1.01 | 1.01 |
| 100000270006 | 24 | 209 | 975 | -0.8600 | 0.0741 | 0.90 | 0.85 |
| 100000269998 | 25 | 209 | 975 | 0.4146 | 0.0682 | 1.07 | 1.11 |
| 100000270000 | 26 | 208 | 975 | -0.7262 | 0.0724 | 0.94 | 0.90 |
| 100000270005 | 27 | 209 | 975 | -0.5977 | 0.0711 | 0.91 | 0.88 |
| 100000270007 | 28 | 209 | 975 | 0.3684 | 0.0680 | 1.01 | 1.01 |
| 100000270016 | 29 | 208 | 975 | 1.0636 | 0.0732 | 1.12 | 1.20 |
| 100000270018 | 30 | 208 | 975 | 0.3499 | 0.0680 | 1.10 | 1.14 |
| 100000270021 | 31 | 209 | 975 | 0.6502 | 0.0693 | 1.06 | 1.10 |
| 100000270022 | 32 | 209 | 975 | 0.2809 | 0.0678 | 1.18 | 1.23 |
| 100000270023 | 33 | 209 | 975 | 0.5597 | 0.0688 | 1.14 | 1.16 |
| 100000270013 | 34 | 209 | 975 | 0.5644 | 0.0688 | 1.02 | 1.03 |
| 100000270020 | 35 | 209 | 975 | 0.4193 | 0.0682 | 0.99 | 0.98 |
| 100000257087 | 36 | 208 | 975 | -1.1007 | 0.0776 | 0.95 | 0.88 |

Table 6.3.4. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 6 (Continued)

| Item CID | Item Seq. No.[1] | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000257088 | 37 | 208 | 975 | -0.2438 | 0.0686 | 0.91 | 0.88 |
| 100000257113 | 38 | 212 | 975 | 0.7570 | 0.0701 | 1.14 | 1.22 |
| 100000257114 | 39 | 212 | 975 | 0.4985 | 0.0685 | 1.06 | 1.07 |
| 100000257118 | 40 | 212 | 975 | -0.6383 | 0.0715 | 0.96 | 0.95 |
| 100000257119 | 41 | 212 | 975 | -0.3959 | 0.0695 | 0.94 | 0.92 |
| 100000257120 | 42 | 212 | 975 | 0.4892 | 0.0684 | 1.06 | 1.09 |
| 100000257102 | 43 | 212 | 975 | 0.0475 | 0.0677 | 0.94 | 0.93 |
| 100000257072 | 44 | 208 | 975 | 0.8968 | 0.0713 | 1.12 | 1.16 |
| 100000257103 | 45 | 212 | 975 | 0.4193 | 0.0682 | 1.08 | 1.11 |
| 100000257104 | 46 | 212 | 975 | 0.1390 | 0.0677 | 1.04 | 1.05 |
| 100000257071 | 47 | 208 | 975 | -0.2673 | 0.0687 | 0.99 | 0.97 |
| 100000257100 | 48 | 212 | 975 | -0.1088 | 0.0681 | 0.95 | 0.94 |
| 100000257101 | 49 | 212 | 975 | 1.0636 | 0.0732 | 1.08 | 1.14 |
| 100000102136 | 50 | 212 | 975 | -0.5472 | 0.0706 | 0.92 | 0.89 |
| 100000102137 | 51 | 212 | 975 | 0.0153 | 0.0678 | 1.06 | 1.07 |
| 100000102139 | 52 | 212 | 975 | -0.3179 | 0.0690 | 1.13 | 1.17 |
| 100000102140 | 53 | 212 | 975 | -0.2517 | 0.0686 | 1.10 | 1.11 |
| 100000257081 | 54 | 208 | 975 | -1.0121 | 0.0762 | 0.94 | 0.87 |
| 100000257110 | 55 | 212 | 975 | -0.3336 | 0.0691 | 0.96 | 0.93 |
| 100000257083 | 56 | 208 | 975 | 0.3131 | 0.0679 | 0.92 | 0.89 |
| 100000257106 | 57 | 212 | 975 | -0.2768 | 0.0687 | 0.90 | 0.89 |
| 100000257107 | 58 | 212 | 975 | 0.6310 | 0.0692 | 0.99 | 0.99 |
| 100000257109 | 59 | 212 | 975 | -0.1366 | 0.0681 | 0.88 | 0.85 |
| 100000257112 | 60 | 212 | 975 | -1.5718 | 0.0874 | 0.82 | 0.67 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

## Table 6.3.5. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 7

| Item CID | Item Seq. No.[1] | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213674 | 1 | 208 | 1158 | -0.7780 | 0.0663 | 0.90 | 0.86 |
| 100000213675 | 2 | 208 | 1158 | -1.3540 | 0.0744 | 0.92 | 0.92 |
| 100000213676 | 3 | 208 | 1158 | 0.8276 | 0.0655 | 1.03 | 1.06 |
| 100000360204 | 4 | 208 | 1158 | 1.0398 | 0.0677 | 1.03 | 1.11 |
| 100000213678 | 5 | 208 | 1158 | -0.9143 | 0.0678 | 0.88 | 0.85 |
| 100000213679 | 6 | 208 | 1158 | -0.8469 | 0.0671 | 0.92 | 0.89 |
| 100000360205 | 7 | 208 | 1158 | 0.7509 | 0.0649 | 0.95 | 0.93 |
| 100000270578 | 8 | 208 | 1158 | -0.4898 | 0.0639 | 0.99 | 0.96 |
| 100000270572 | 9 | 209 | 1158 | -0.0167 | 0.0622 | 1.09 | 1.11 |
| 100000270574 | 10 | 209 | 1158 | 0.4180 | 0.0629 | 0.98 | 0.99 |
| 100000322011 | 11 | 209 | 1158 | -0.4939 | 0.0640 | 0.98 | 0.97 |
| 100000270580 | 12 | 208 | 1158 | -1.5496 | 0.0783 | 0.88 | 0.76 |
| 100000270571 | 13 | 209 | 1158 | 0.6143 | 0.0639 | 1.05 | 1.06 |
| 100000270575 | 14 | 209 | 1158 | -0.8979 | 0.0676 | 0.88 | 0.82 |
| 100000270109 | 15 | 209 | 1158 | 0.4417 | 0.0630 | 1.02 | 1.01 |
| 100000270112 | 16 | 209 | 1158 | 0.5778 | 0.0637 | 1.10 | 1.15 |
| 100000270113 | 17 | 208 | 1158 | -0.0321 | 0.0622 | 1.03 | 1.04 |
| 100000270114 | 18 | 208 | 1158 | -0.2186 | 0.0626 | 1.01 | 1.01 |
| 100000270111 | 19 | 209 | 1158 | 0.4894 | 0.0632 | 1.18 | 1.28 |
| 100000270107 | 20 | 209 | 1158 | 0.8187 | 0.0654 | 1.14 | 1.19 |
| 100000270121 | 21 | 209 | 1158 | -0.1912 | 0.0625 | 0.98 | 0.97 |
| 100000102276 | 22 | 209 | 1158 | 0.0908 | 0.0622 | 1.01 | 1.01 |
| 100000102277 | 23 | 209 | 1158 | 0.7546 | 0.0649 | 1.02 | 1.04 |
| 100000102274 | 24 | 208 | 1158 | 0.4769 | 0.0631 | 1.09 | 1.12 |
| 100000102278 | 25 | 209 | 1158 | 0.0775 | 0.0622 | 0.97 | 0.96 |
| 100000102279 | 26 | 209 | 1158 | 0.4687 | 0.0631 | 1.01 | 1.03 |
| 100000102263 | 27 | 208 | 1158 | -0.1197 | 0.0624 | 1.00 | 1.00 |
| 100000102267 | 28 | 209 | 1158 | 0.4089 | 0.0628 | 1.12 | 1.18 |
| 100000102268 | 29 | 209 | 1158 | 0.0297 | 0.0622 | 0.94 | 0.92 |
| 100000102269 | 30 | 209 | 1158 | 0.4035 | 0.0628 | 1.03 | 1.03 |
| 100000102266 | 31 | 209 | 1158 | 0.6886 | 0.0644 | 1.04 | 1.06 |
| 100000257181 | 32 | 208 | 1158 | -0.6181 | 0.0649 | 1.08 | 1.10 |
| 100000257182 | 33 | 208 | 1158 | -1.3911 | 0.0751 | 0.93 | 0.88 |
| 100000257216 | 34 | 212 | 1158 | -0.7990 | 0.0665 | 0.93 | 0.90 |
| 100000257218 | 35 | 212 | 1158 | -1.2130 | 0.0720 | 0.89 | 0.85 |
| 100000257219 | 36 | 212 | 1158 | 0.1995 | 0.0623 | 1.10 | 1.12 |

Table 6.3.5. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 7 (Continued)

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000257215 | 37 | 212 | 1158 | -2.0275 | 0.0906 | 0.96 | 0.86 |
| 100000257221 | 38 | 212 | 1158 | -0.0708 | 0.0623 | 1.05 | 1.06 |
| 100000102244 | 39 | 208 | 1158 | -1.6963 | 0.0816 | 0.85 | 0.74 |
| 100000102246 | 40 | 212 | 1158 | -0.1924 | 0.0625 | 1.04 | 1.04 |
| 100000102247 | 41 | 212 | 1158 | -1.1534 | 0.0711 | 0.85 | 0.75 |
| 100000102248 | 42 | 212 | 1158 | 0.0244 | 0.0622 | 1.09 | 1.10 |
| 100000102245 | 43 | 208 | 1158 | -1.3660 | 0.0746 | 0.87 | 0.86 |
| 100000102218 | 44 | 212 | 1158 | 0.1918 | 0.0622 | 0.94 | 0.94 |
| 100000102220 | 45 | 212 | 1158 | -0.2739 | 0.0628 | 0.92 | 0.91 |
| 100000102221 | 46 | 212 | 1158 | 0.1732 | 0.0622 | 0.94 | 0.93 |
| 100000102222 | 47 | 212 | 1158 | -0.4429 | 0.0637 | 0.99 | 0.96 |
| 100000102217 | 48 | 208 | 1158 | -0.5024 | 0.0640 | 1.01 | 1.02 |
| 100000257202 | 49 | 208 | 1158 | -1.7436 | 0.0828 | 0.95 | 0.86 |
| 100000257769 | 50 | 212 | 1158 | -0.0747 | 0.0623 | 1.10 | 1.14 |
| 100000257770 | 51 | 212 | 1158 | 0.5174 | 0.0633 | 1.07 | 1.15 |
| 100000257773 | 52 | 212 | 1158 | -1.0240 | 0.0692 | 0.90 | 0.86 |
| 100000322795 | 53 | 208 | 1158 | -0.0244 | 0.0622 | 1.04 | 1.04 |
| 100000257767 | 54 | 212 | 1158 | -0.6519 | 0.0652 | 1.01 | 0.98 |
| 100000257772 | 55 | 212 | 1158 | 0.0451 | 0.0622 | 0.94 | 0.92 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

## Table 6.3.6. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 8

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000213683 | 1 | 208 | 1268 | -1.1648 | 0.0749 | 0.92 | 0.93 |
| 100000213684 | 2 | 208 | 1268 | -0.4197 | 0.0634 | 1.03 | 1.05 |
| 100000360208 | 3 | 208 | 1268 | 0.7685 | 0.0595 | 1.06 | 1.07 |
| 100000213686 | 4 | 208 | 1268 | 0.4446 | 0.0589 | 0.99 | 0.98 |
| 100000213687 | 5 | 208 | 1268 | 0.0045 | 0.0601 | 1.00 | 1.01 |
| 100000213688 | 6 | 208 | 1268 | -0.1989 | 0.0614 | 0.97 | 0.96 |
| 100000360209 | 7 | 208 | 1268 | 0.3432 | 0.0590 | 0.99 | 0.98 |
| 100000270136 | 8 | 209 | 1268 | 0.7085 | 0.0593 | 1.08 | 1.10 |
| 100000270132 | 9 | 209 | 1268 | -1.0472 | 0.0726 | 0.96 | 0.93 |
| 100000273591 | 10 | 208 | 1268 | 0.2316 | 0.0592 | 0.97 | 0.96 |
| 100000270131 | 11 | 209 | 1268 | 0.9731 | 0.0605 | 1.00 | 1.01 |
| 100000270135 | 12 | 209 | 1268 | 0.3501 | 0.0590 | 1.00 | 0.99 |
| 100000270137 | 13 | 209 | 1268 | 0.1296 | 0.0595 | 1.04 | 1.03 |
| 100000304610 | 14 | 208 | 1268 | 1.0504 | 0.0609 | 1.04 | 1.06 |
| 100000102331 | 15 | 209 | 1268 | -1.3174 | 0.0784 | 0.85 | 0.78 |
| 100000102329 | 16 | 208 | 1268 | 0.2613 | 0.0591 | 1.07 | 1.08 |
| 100000102330 | 17 | 208 | 1268 | 0.0618 | 0.0598 | 0.95 | 0.94 |
| 100000102334 | 18 | 209 | 1268 | 0.3175 | 0.0590 | 1.14 | 1.17 |
| 100000102335 | 19 | 209 | 1268 | -1.1686 | 0.0750 | 0.82 | 0.74 |
| 100000270154 | 20 | 209 | 1268 | 0.9005 | 0.0601 | 1.14 | 1.17 |
| 100000270155 | 21 | 209 | 1268 | 0.7543 | 0.0594 | 1.17 | 1.20 |
| 100000270158 | 22 | 209 | 1268 | 0.4612 | 0.0589 | 1.01 | 1.01 |
| 100000273597 | 23 | 208 | 1268 | -0.4749 | 0.0639 | 0.95 | 0.92 |
| 100000270163 | 24 | 209 | 1268 | 0.4577 | 0.0589 | 1.00 | 0.99 |
| 100000270159 | 25 | 209 | 1268 | -0.8571 | 0.0692 | 0.91 | 0.84 |
| 100000273595 | 26 | 208 | 1268 | 0.8933 | 0.0600 | 1.12 | 1.15 |
| 100000102341 | 27 | 208 | 1268 | -0.2773 | 0.0620 | 0.92 | 0.88 |
| 100000102343 | 28 | 209 | 1268 | 0.2393 | 0.0592 | 1.00 | 0.99 |
| 100000102344 | 29 | 209 | 1268 | -0.5397 | 0.0647 | 0.99 | 0.98 |
| 100000102346 | 30 | 209 | 1268 | 0.2686 | 0.0591 | 0.92 | 0.90 |
| 100000102345 | 31 | 209 | 1268 | -0.0760 | 0.0605 | 0.91 | 0.88 |
| 100000257161 | 32 | 212 | 1268 | 0.3084 | 0.0590 | 1.06 | 1.06 |
| 100000302728 | 33 | 212 | 1268 | 0.8110 | 0.0596 | 1.09 | 1.10 |
| 100000257787 | 34 | 208 | 1268 | 1.0208 | 0.0608 | 1.04 | 1.06 |
| 100000257788 | 35 | 208 | 1268 | -0.6432 | 0.0660 | 0.94 | 0.87 |
| 100000257158 | 36 | 212 | 1268 | 0.1965 | 0.0593 | 0.93 | 0.92 |

Table 6.3.6. The 2010 Mod-MSA Reading IRT Item Parameters: Grade 8 (Continued)

| Item CID | Item Seq. No. | Item Strand No. | n- Count | Rasch Difficulty | *SE* | MS Infit | MS Outfit |
|---|---|---|---|---|---|---|---|
| 100000257160 | 37 | 212 | 1268 | 0.0262 | 0.0600 | 0.95 | 0.93 |
| 100000257163 | 38 | 212 | 1268 | -0.7049 | 0.0668 | 0.94 | 0.88 |
| 100000102320 | 39 | 208 | 1268 | -0.9453 | 0.0707 | 0.99 | 0.94 |
| 100000102322 | 40 | 212 | 1268 | -0.2879 | 0.0621 | 1.08 | 1.12 |
| 100000102323 | 41 | 212 | 1268 | 0.2760 | 0.0591 | 1.08 | 1.09 |
| 100000102324 | 42 | 212 | 1268 | 0.0370 | 0.0599 | 0.89 | 0.88 |
| 100000102321 | 43 | 208 | 1268 | -0.8923 | 0.0698 | 0.98 | 0.91 |
| 100000257139 | 44 | 212 | 1268 | 1.3316 | 0.0633 | 1.10 | 1.16 |
| 100000257140 | 45 | 212 | 1268 | -0.5410 | 0.0647 | 1.06 | 1.08 |
| 100000257141 | 46 | 212 | 1268 | 0.5584 | 0.0590 | 1.06 | 1.06 |
| 100000257144 | 47 | 212 | 1268 | 0.0977 | 0.0597 | 0.97 | 0.96 |
| 100000257777 | 48 | 208 | 1268 | 0.6245 | 0.0591 | 1.02 | 1.02 |
| 100000302674 | 49 | 208 | 1268 | 0.1895 | 0.0593 | 1.01 | 1.01 |
| 100000302675 | 50 | 212 | 1268 | -0.4505 | 0.0637 | 0.92 | 0.86 |
| 100000102304 | 51 | 212 | 1268 | 0.2711 | 0.0591 | 1.03 | 1.03 |
| 100000102305 | 52 | 212 | 1268 | -0.2458 | 0.0617 | 0.95 | 0.92 |
| 100000102302 | 53 | 208 | 1268 | 0.4618 | 0.0589 | 0.95 | 0.94 |
| 100000102307 | 54 | 212 | 1268 | -0.3358 | 0.0625 | 0.93 | 0.91 |
| 100000102308 | 55 | 212 | 1268 | -0.1834 | 0.0613 | 0.89 | 0.85 |

Note: 1. 203, 204 & 205=General Reading, 209=Literary, 212=Informational
2. These analyses are based on the equating sample used with the exclusion criterion.

# 7. TEST RELIABILITY

## 7.1. Precision and Reliability (Classical Methods)

**Standard Error of Measurement (SEM) of the Test**

Classical test theory is based on the following assumptions (Andrich & Luo, 2004):

- Each person *v* has a true score on the construct, usually denoted by the variable $T_{v.}$
- The best overall indicator of the person's true score is the sum of the scores on the items and is usually denoted by the variable $X_{v.}$
- This observed score will have an error for each person, usually denoted by $E_{v.}$
- These errors are not correlated with the true score.
- Across a population of people, the errors sum to 0 and they are normally distributed.

Based on these assumptions, useful indices are available within the framework of classical test theory (CTT) for estimating the precision of the raw test scores and the reliability of assessments. Within CTT, an observed test score is defined as an imprecise estimate of a student's true (and unobservable) proficiency level and is composed of two components. The first component is referred to as "true score" and is the portion of the observed score that is directly dependent on the student's proficiency level. The second is an error component (error) and is the portion of the score that is attributable to random error, that is, the portion of the score attributable to factors unrelated to the student's proficiency. Error for any student is normally distributed around that student's true score with a mean of zero and an arbitrary standard deviation. Suppose it were possible to give an exam to one student a large number of times without any practice effects. If we were to examine the resulting distribution of scores, we would find a normal distribution with a certain mean and a certain standard deviation about the mean. The mean of the resulting distribution is the student's true score according to the definition of error given above. For each student who responds to the exam, error is normally distributed with a mean of zero. However, the standard deviation of the error distribution is idiosyncratic to each student (though it tends to be larger toward the low and high ends of the exam for most tests). If we wanted to estimate what would likely be the standard deviation of this distribution of errors for any arbitrary examinee, the best estimate would be the mean of the standard deviations of the error distribution across all examinees. This quantity is called the standard error of measurement (SEM).

From the assumptions outlined and discussed above, the following mathematical formula can be derived:

$$X_v = T_v + E_v .$$

Therefore,

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

where

$\sigma_x^2$ = the variance of the observed score in a population of persons,

$\sigma_t^2$ = the variance of their true score variance, and

$\sigma_e^2$ = the error variance.

The reliability coefficient of the test can be calculated by the following formula:

$$\rho_x = \frac{\sigma_t^2}{\sigma_x^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2}.$$

Thus, the *SEM* is calculated by the following formula:

$$\sigma_e = \sigma_x \sqrt{1 - \rho_x}.$$

The SEM is commonly used in interpreting and reporting individual test scores and score differences on tests (Harvill, 1991). This equation, however, is only useful to estimate true score when the test reliability is reasonably high and the obtained score for the examinee is not an extreme deviation from the mean of the appropriate reference group. Consequently, when we use this equation, we should be careful with statements so that they do not imply greater precision than is actually involved (Harvill, 1991).

The SEM for each grade level of the test is provided in Chapter 9 in Table 9.1.1: Classical Descriptive Statistics for the 2010 Mod-MSA: Reading: Grades 3 though 8.

**Cronbach's Alpha ($KR_{20}$)**

Cronbach Alpha can be calculated by several methods. For dichotomously scored items, one of the best methods is the Kuder Richardson 20 (Crocker & Algina (1986), p.139) to estimate the internal consistency of items in the tests. Since the Mod-MSA: Reading tests include only SR items, the following formula was used to obtain the $KR_{20}$:

$$KR_{20} = \frac{k}{k-1}\left(1 - \frac{\sum pq}{\hat{\sigma}_x^2}\right)$$

$KR_{20}$ = Kuder Richardson 20

$k$ = number of items on the test

$pq$ = variance of item i, and

$\hat{\sigma}_x^2$ = total test variance

$KR_{20}$ is provided as reliability of the test in Table 9.1.1.

## 7.2. IRT Method in Measuring Precision of the Test

The information function (as discussed and provided in Section 9.4) is a function of proficiency and can be used to measure the precision of the test under IRT methods at a specified proficiency level. Conversely, the greater the information, the more precise will be the measurement of proficiency.

The inverse of the information function is the same as the conditional standard error of measurement (CSEM) discussed and provided in Section 9.4. The figures depicting CSEM

provided in Section 9.4 show the standard errors of measurement at different proficiency levels of the examinees.

### 7.3. Decision Accuracy and Consistency at the Cut Scores

The accuracy and consistency analyses make use of the methods outlined and implemented in Livingston and Lewis (1995), Haertel (1996), and Young and Yoon (1998).

The *accuracy* of a decision is the extent to which it would agree with the decisions that would be made if each student could somehow be tested with all possible parallel forms of the assessments. The *consistency* of a decision is the extent to which it would agree with the decisions that would be made if the students had taken a different form of the examination, equal in difficulty and covering the same content as the form they actually took.

Students can be misclassified in one of two ways. Students who were below the proficiency cut score, but were classified (on the basis of the assessment) as being above a cut score, are considered to be *false positives*. Students who were above the proficiency cut score, but were classified as being below a cut score, are considered to be *false negatives*.

For the 2010 Mod-MSA: Reading, Tables 7.3.1 through 7.3.6 include:

- Performance level
- Accuracy classifications
- False positives
- False negatives
- Consistency classifications

The tables illustrate the general rule that decision consistency is less than decision accuracy.

**Table 7.3.1. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 3**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.88 | 0.07 | 0.05 | 0.83 |
| BP : A | 0.93 | 0.05 | 0.02 | 0.90 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. 2. These analyses are based on the statewide population after applying equating exclusion criteria

**Table 7.3.2. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 4**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.88 | 0.06 | 0.06 | 0.83 |
| BP : A | 0.93 | 0.05 | 0.02 | 0.91 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. These analyses are based on the statewide population after applying equating exclusion criteria

**Table 7.3.3. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 5**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.87 | 0.07 | 0.07 | 0.82 |
| BP : A | 0.96 | 0.04 | 0.01 | 0.94 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. These analyses are based on the statewide population after applying equating exclusion criteria

**Table 7.3.4. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 6**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.87 | 0.07 | 0.07 | 0.81 |
| BP : A | 0.92 | 0.06 | 0.02 | 0.88 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. These analyses are based on the statewide population after applying equating exclusion criteria

**Table 7.3.5. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 7**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.88 | 0.08 | 0.05 | 0.83 |
| BP : A | 0.97 | 0.03 | 0.01 | 0.95 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. These analyses are based on the statewide population after applying equating exclusion criteria

**Table 7.3.6. The 2010 Mod-MSA: Reading Decision Accuracy and Consistency Indices: Grade 8**

| Performance Cut | Accuracy | False Positive | False Negative | Consistency |
|---|---|---|---|---|
| B : PA | 0.86 | 0.07 | 0.06 | 0.81 |
| BP : A | 0.95 | 0.04 | 0.02 | 0.92 |

*Note.* B:PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced. These analyses are based on the statewide population after applying equating exclusion criteria

# 8. TEST VALIDITY

## 8.1. Test Validity for the 2010 Mod-MSA: Reading

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 Mod-MSA: Reading, content-related evidence, evidence from item development methods, bias review evidence during test development and for items that showed differential item functioning (DIF), and evidence from internal structure were collected. Also, a study comparing the mode of administration was undertaken by Pearson to validate the online administration of the test.

### 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).

Evidence regarding the alignment between the content in the 2010 Mod-MSA: Reading and the standards of achievement set by MSDE are provided in Appendix E that links each item to the specific standard(s) it measures. Information on the item composition of the operational test forms can be obtained from Section 2.6 *Items Selected for the 2010 Operational Tests.* The selected items are displayed in Appendix A with their UIN numbers.

### Evidence from Item Development Methods

Test development for Mod-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 necessitated a great deal of time and energy. More specific information on item (test) development and review can be obtained in Section 2, *Test Design and Development of the 2010* Mod-MSA: Reading while the standards to which the items were aligned can be obtained from the MSDE website at: http://mdk12.org/instruction/curriculum/index.html.

As explained in Section 2.4 to 2.6, once these items were scored, MSDE and Pearson conducted additional item analysis and content review to select items for the operational form, i.e., the form on which the student scores would be reported. Any item 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.

**Evidence Based on Excluding Bias Items Before and After DIF Analysis**

One important consideration in evaluating the validity of a test is to examine the equity of each item performance between groups of interest. As explained in Section 2.2, all items went through a bias review committee to ascertain that items were not biased with respect to gender, ethnicity, geographical location, etc. Also, as explained in Section 2.4, after items were scored, DIF analysis was undertaken and those items that showed moderate or significant DIF were reviewed for bias with respect to gender, and ethnicity, which included white and black students. More information on DIF analyses can be obtained in Section 2.4, *Differential Item Functioning*.

Items that had moderate or extreme DIF are depicted in Table 8.1.1, below. These items for the Mod-MSA; Reading were checked for content bias, but did not show favoritism on the basis of gender or ethnicity (black vs. white students).

**Table 8.1.1. Category Classification of Items that Showed Moderate or Extreme DIF by Grades**

| | | DIF Classification[2] | | |
|---|---|---|---|---|
| Grade | Item Sequence No.[1] | Item CID No. | Gender | White/African-American |
| 3 | 13 | 100000360184 | A | -B |
| | 27 | 100000101938 | A | +B |
| | 67 | 100000260372 | -C | A |
| 4 | 12 | 100000360192 | A | -B |
| | 16 | 100000357135 | A | -B |
| | 47 | 100000462157 | -B | A |
| | 50 | 100000101998 | A | -B |
| | 56 | 100000102026 | A | +B |
| | 64 | 100000267467 | -C | A |
| 5 | 2 | 100000213651 | A | -B |
| | 4 | 100000213653 | A | -B |
| | 7 | 100000213656 | A | -B |
| | 8 | 100000213657 | -B | A |
| | 9 | 100000213658 | A | -C |
| | 10 | 100000213659 | A | -B |
| | 21 | 100000102114 | +B | A |
| | 48 | 100000102058 | A | A |
| 6 | 9 | 100000213670 | A | -C |
| | 10 | 100000213671 | +B | A |
| | 18 | 100000102175 | A | +C |
| | 24 | 100000270006 | A | -B |
| | 29 | 100000270016 | A | +B |
| | 39 | 100000257114 | A | +B |
| | 40 | 100000257118 | A | -B |

Table 8.1.1. Category Classification of Items that Showed Moderate or Extreme DIF by Grades (Continued)

| | | DIF Classification[2] | | |
|---|---|---|---|---|
| Grade | Item Sequence No.[1] | Item CID No. | Gender | White/African-American |
| 6 | 44 | 100000257072 | +B | A |
| | 48 | 100000257100 | -B | A |
| | 53 | 100000102140 | A | -B |
| | 60 | 100000257112 | A | -B |
| 7 | 7 | 100000360205 | A | -C |
| | 8 | 100000270578 | +B | A |
| | 14 | 100000270575 | +B | A |
| | 35 | 100000257218 | A | -B |
| | 37 | 100000257215 | -B | A |
| | 39 | 100000102244 | A | -C |
| | 41 | 100000102247 | -B | A |
| | 45 | 100000102220 | A | -B |
| | 55 | 100000257772 | A | -B |
| 8 | 19 | 100000102335 | A | -B |
| | 27 | 100000102341 | +B | A |
| | 29 | 100000102344 | A | +B |
| | 50 | 100000302675 | A | -B |

Note: 1. '+' = in favor of the reference group, i.e. males and White Americans while '-' = in favor of the focal group. Extreme DIF = "C", Moderate DIF = "B", and No DIF is classified as an "A".
2. These analyses are based on the equating sample used with the exclusion criteria.

**Evidence from Internal Structure of the Tests**

As explained in Section 2.3, the 2010 Mod-MSA: Reading contains three reading strands: General Reading, Literary Reading, and Informational Reading. Even though these are individual strands are "locally independent," they measure the same underlying reading trait. Therefore, the positive correlation among these strands is an indication of their relationship with each other in measuring the same underlying construct. To ascertain the homogeneity of the test, correlations were calculated to depict the relationship between each strand within a grade. Tables 8.1.2 through 8.1.7 show the correlations among the reading strands for each of the three grades 3 through 8, respectively.

**Table 8.1.2. The 2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 3**

| Strand (Subscale) | N | Mean | SD | GR | L | I |
|---|---|---|---|---|---|---|
| General Reading (GR) | 813 | 10.02 | 3.34 | 1.00 | | |
| Literary (L) | 813 | 7.07 | 2.50 | 0.60 | 1.00 | |
| Informational (I) | 813 | 8.13 | 3.01 | 0.58 | 0.57 | 1.00 |

*Note*. The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

### Table 8.1.3. The 2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 4

| Strand (Subscale) | N | Mean | SD | GR | L | I |
|---|---|---|---|---|---|---|
| General Reading (GR) | 967 | 10.10 | 3.22 | 1.00 | | |
| Literary (L) | 967 | 6.82 | 2.51 | 0.57 | 1.00 | |
| Informational (I) | 967 | 7.48 | 2.85 | 0.60 | 0.51 | 1.00 |

*Note*. The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

### Table 8.1.4. The 2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 5

| Strand (Subscale) | N | Mean | SD | GR | L | I |
|---|---|---|---|---|---|---|
| General Reading (GR) | 1043 | 10.08 | 3.06 | 1.00 | | |
| Literary (L) | 1043 | 7.62 | 2.80 | 0.57 | 1.00 | |
| Informational (I) | 1043 | 6.60 | 2.58 | 0.50 | 0.51 | 1.00 |

*Note*. The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

### Table 8.1.5. The 2009/2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 6

| | 2009 | | | | | | 2010 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Strand (Subscale)** | N | Mean | SD | GR | L | I | N | Mean | SD | GR | L | I |
| General Reading (GR) | 1274 | 8.62 | 2.87 | 1.00 | | | 975 | 8.82 | 3.07 | 1.00 | | |
| Literary (L) | 1274 | 8.35 | 2.71 | 0.47 | 1.00 | | 975 | 8.12 | 2.80 | 0.59 | 1.00 | |
| Informational (I) | 1274 | 7.16 | 2.42 | 0.46 | 0.40 | 1.00 | 975 | 7.59 | 2.78 | 0.57 | 0.54 | 1.00 |

*Note*. The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

### Table 8.1.6. The 2009/2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 7

| | 2009 | | | | | | 2010 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Strand (Subscale)** | N | Mean | SD | GR | L | I | N | Mean | SD | GR | L | I |
| General Reading (GR) | 1630 | 8.62 | 2.49 | 1.00 | | | 1158 | 9.50 | 2.79 | 1.00 | | |
| Literary (L) | 1630 | 8.49 | 2.80 | 0.55 | 1.00 | | 1158 | 8.77 | 2.83 | 0.62 | 1.00 | |
| Informational (I) | 1630 | 7.32 | 2.81 | 0.52 | 0.48 | 1.00 | 1158 | 7.01 | 2.80 | 0.50 | 0.49 | 1.00 |

*Note:* The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

### Table 8.1.7. The 2009/2010 Mod-MSA, Reading Strand (Cluster) Correlations: Grade 8

| | 2009 | | | | | | 2010 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Strand (Subscale)** | N | Mean | SD | GR | L | I | N | Mean | SD | GR | L | I |
| General Reading (GR) | 1792 | 9.51 | 2.80 | 1.00 | | | 1268 | 9.79 | 2.84 | 1.00 | | |
| Literary (L) | 1792 | 7.79 | 2.58 | 0.57 | 1.00 | | 1268 | 8.77 | 2.69 | 0.55 | 1.00 | |
| Informational (I) | 1792 | 9.37 | 2.70 | 0.54 | 0.50 | 1.00 | 1268 | 8.71 | 2.59 | 0.52 | 0.48 | 1.00 |

*Note.* The restriction of the range of scores on the strands could have resulted in the attenuation of the correlation coefficients between any two modalities.

**Evidence of Equity in the Modes of Administration**

See Appendix G: *Comparison of Paper-Pencil Version with the On-Line Version of the Maryland Modified School Assessment (Mod-MSA) in Reading and Mathematics (Grades 4 and 5)* that justifies the use of online testing both as a total test across the two modes of administration and also in the bias-examination of each item's performance across the modes of administration**.**

## 8.2. Unidimensionality Analysis for the 2010 Mod-MSA: Reading

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

To check the unidimensionality of the 2010 Mod-MSA: Reading, correlation coefficients were computed with LISREL 8.5 (Jöreskog & Sörbom, 1993). Principal component analysis was then applied to produce eigenvalues. The first and the second principal component eigenvalues were compared without rotation. Table 8.2.1 summarizes the results of the first and second principal component eigenvalues of the 2010 Mod-MSA: Reading. As shown in the table, the first factor extracted a much large amount of eigenvalues across all grades.

**Table 8.2.1. The 2010 Mod-MSA: Reading Eigenvalues in the First and Second Components**

| Grade | First Eigenvalue | Second Eigenvalue |
|---|---|---|
| 3 | 8.77 | 2.17 |
| 4 | 7.84 | 2.01 |
| 5 | 7.08 | 1.70 |
| 6 | 7.51 | 1.72 |
| 7 | 7.27 | 1.75 |
| 8 | 7.97 | 1.60 |

*Note.* Analyses were conducted with the statewide population after applying equating exclusion criteria.

# 9. SUMMARY OF OPERATIONAL TEST RESULTS OF THE 2010 MOD-MSA: READING

This section presents both the raw score and scaled score summaries for the Mod-MSA: Reading by grade. Table 9.1.1 presents the raw score summary by grade. Table 9.2.1 presents the scale score summary by grade. Table 9.3.1 presents the percentage of students in each of the proficiency levels by grade. In addition, Appendix C provides frequency distributions and histograms of the scale scores of the 2010 Mod-MSA: Reading. For grades 6-8, the 2009 results are also depicted in the above mentioned tables.

## 9.1 Classical Descriptive Test Statistics

Table 9.1.1 contains the classical descriptive statistics of each form for each grade and includes:

- Numbers of students (based on a whole population)
- Numbers of items
- Minimum and maximum points achievable on the test
- Means and standard deviations of raw scores
- Test reliability ($KR_{20}$)
- Standard error of measurement (SEM)

**Table 9.1.1 Classical Descriptive Statistics for the 2010 Mod-MSA: Reading: Grades 3 through 8**

| | | | | 2009 | | | | | 2010 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade | Total # of Items | Min. Point | Max. Point | N | Mean | SD | Reliability | SEM | N | Mean | SD | Reliability | SEM |
| 3 | 45 | 0 | 45 | N/A | N/A | N/A | N/A | N/A | 1058 | 17.53 | 6.45 | 0.82 | 3.08 |
| 4 | 45 | 0 | 45 | N/A | N/A | N/A | N/A | N/A | 1335 | 24.83 | 7.04 | 0.81 | 3.06 |
| 5 | 45 | 0 | 45 | N/A | N/A | N/A | N/A | N/A | 1459 | 24.68 | 6.65 | 0.79 | 3.04 |
| 6 | 45 | 0 | 45 | 1274 | 24.13 | 6.34 | 0.75 | 3.16 | 1356 | 25.22 | 6.92 | 0.81 | 3.03 |
| 7 | 45 | 0 | 45 | 1630 | 24.43 | 6.66 | 0.78 | 3.09 | 1702 | 25.97 | 6.54 | 0.79 | 3.02 |
| 8 | 45 | 0 | 45 | 1792 | 26.68 | 6.71 | 0.79 | 3.06 | 1969 | 27.23 | 6.89 | 0.81 | 3.04 |

*Note*. Analyses were conducted with the statewide population after applying equating exclusion criteria.

## 9.2 Scale Score Descriptive Statistics

Table 8.2.1 provides information about scale score descriptive statistics of the test by grade and years and includes:

- Numbers of students
- Mean and standard deviation of scale scores
- 10% quantile (P10), 25% quantile (Q1), median (P50), 75% quantile (Q3), 90% quantile, and IQR (Interquantile Range=Q3-Q1)

**Table 9.2.1. Scale Score Descriptive Statistics Across Years for Mod-MSA, Reading: Grades 3-8**

| Grade | Year | *N* | *M* | *SD* | *P10* | *Q1* | *Mdn* | *Q3* | *P90* | *IQR* |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 2010 | 1058 | 50.69 | 12.16 | 37 | 42 | 50 | 58 | 66 | 16 |
| 4 | 2010 | 1335 | 50.73 | 11.84 | 37 | 43 | 49 | 58 | 67 | 15 |
| 5 | 2010 | 1459 | 50.65 | 11.44 | 37 | 43 | 51 | 57 | 66 | 14 |
| 6 | 2009 | 1274 | 49.81 | 12.02 | 35 | 41 | 49 | 58 | 66 | 17 |
| | 2010 | 1356 | 52.14 | 13.69 | 35 | 42 | 51 | 60 | 69 | 18 |
| 7 | 2009 | 1630 | 49.75 | 12.38 | 35 | 42 | 49 | 57 | 65 | 15 |
| | 2010 | 1702 | 50.48 | 12.27 | 35 | 42 | 50 | 59 | 67 | 17 |
| 8 | 2009 | 1792 | 50.04 | 12.17 | 35 | 42 | 50 | 59 | 65 | 17 |
| | 2010 | 1969 | 51.68 | 12.57 | 36 | 43 | 52 | 59 | 68 | 16 |

Note: 1. Grades 3-5 have no history since 2010 was the first year of their administration.
2. Analyses were conducted with the statewide population after applying equating exclusion criteria.

## 9.3 Frequency of Students at Each Proficiency Level by Grade

Table 9.3.1 contains the pass rate at each performance level based on the cut scores shown in Table 4.7.1.

**Table 9.3.1. Percentage of Students in Each Proficiency Level by Grade**

| Grade | *N* | Percentage of Performance Level | | |
|---|---|---|---|---|
| | | Basic | Proficient | Advanced |
| 3 | 1058 | 62.19 | 25.43 | 12.38 |
| 4 | 1335 | 60.82 | 26.67 | 12.51 |
| 5 | 1459 | 61.34 | 32.69 | 5.96 |
| 6 | 1356 | 56.27 | 28.32 | 15.41 |
| 7 | 1702 | 68.80 | 26.26 | 4.94 |
| 8 | 1969 | 54.75 | 33.21 | 12.04 |

*Note.* 1. Percentages may not add up to 100% due to rounding.
2. Analyses were conducted with the statewide population after applying equating exclusion criteria.

## 9.4. Test Characteristic Curves, Test Information Functions, and Conditional Standard Errors of Measurement

### Test Characteristic Curves

In IRT models, an item characteristic curve (ICC) permits us to see how the probability of answering correctly depends on the latent trait, i.e., the proficiency of the students. The most common shape of the ICC in practice is the S-shaped curve, which increases monotonically from left to right with the lower asymptote approaching 0 and the upper asymptote approaching 1. Since the logistic model ICC is the probabilistic curve for the item, the score on the test can be presented by the summation of the probabilistic scores of each item plus the error of measurement, i.e.:

$$X = \sum_{g} P_g(\theta) + E \,,$$

where *g* is the item number and *E* is the standard error of measurement. The regression formula shown above for predicting *X* from $\theta$ scores is known as the test characteristic curve (TCC). The TCC for each grade 3 through 8 are provided in Figure 8.4.1a, 8.4.2a, 8.4.3a, 8.4.4a, 8.4.5a, and 8.4.6a, respectively. For grades 6 to 8, a comparison is made across years in these figures.

### Test Information Functions

On a standardized achievement test, items could be too hard for the low- proficiency examinee. Similarly some items may be too easy for everyone and may not help in providing any discrimination for these students. These types of items provide little or no information at the cut scores where it really counts. In most testing situations it becomes necessary for us to understand the information provided by each item across the spectrum of different proficiency levels. Mathematically,

$$I_g(\theta) = \frac{\left[P'_g(\theta)\right]^2}{\left[P_g(\theta)\right]\left[Q_g(\theta)\right]}, \text{ where}$$

*g* is the number of the item, $Q_g(\theta)$ = 1 - $P_g(\theta)$, and the numerator is the first derivative of $P_g(\theta)$, which for the Rasch model = $P_g(\theta)\ Q_g(\theta)$.

The test information function is the sum of the item information functions for all items on the test and is useful in examining the total information provided by the test across the proficiency levels. Symbolically, the test information function at a particular proficiency level can be depicted as:

$$I(\theta) = \sum_{g} I_g(\theta)$$

The test information functions for each grade are provided in Figures 8.4.1b, 8.4.2b, 4.3b, 8.4.4b, 8.4.5b, and 8.4.6b. For grades 6 to 8, a comparison is made across years in these figures.

**Conditional Standard Errors of Measurement (CSEM) of the Tests**

The conditional standard error of measurement is the inverse of the information function. Under the Rasch (i.e., 1-PL IRT) model, the CSEM for each person is as follows (Andrich & Luo, 2004):

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^{L} p_{vi}(1-p_{vi})}}$$

where

$v$ = subscript for a person,

$i$ = subscript for an item,

$L$ = length of the test,

$\hat{\beta}$ = proficiency estimate, and

$p_{vi}$ = the probability that a person answers an item correctly and defined as follows:

$p_{vi} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$ where $\beta_v$ is person's proficiency and $\delta_i$ is item's difficulty.

A confidence band can be found for use in interpreting the proficiency estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by

$$\hat{\beta} \pm SEM$$

For the item standard error, the above equation would be modified so that each item's difficulty estimate will be the summation over the different abilities in the test. Note that the standard error for item difficulty is smallest when the probability of passing is close to the probability of failing. That is, the standard error is small when an item is near the threshold level for many persons in the sample (Embretson & Reise, 2000). These stats are provided for each grade in Figures 8.4.1c, 8.4.2c, 8.4.3c, 8.4.4c, 8.4.5c, and 8.4.6c respectively. For grades 6 to 8, a comparison is made across years in these figures.

## Grade 3

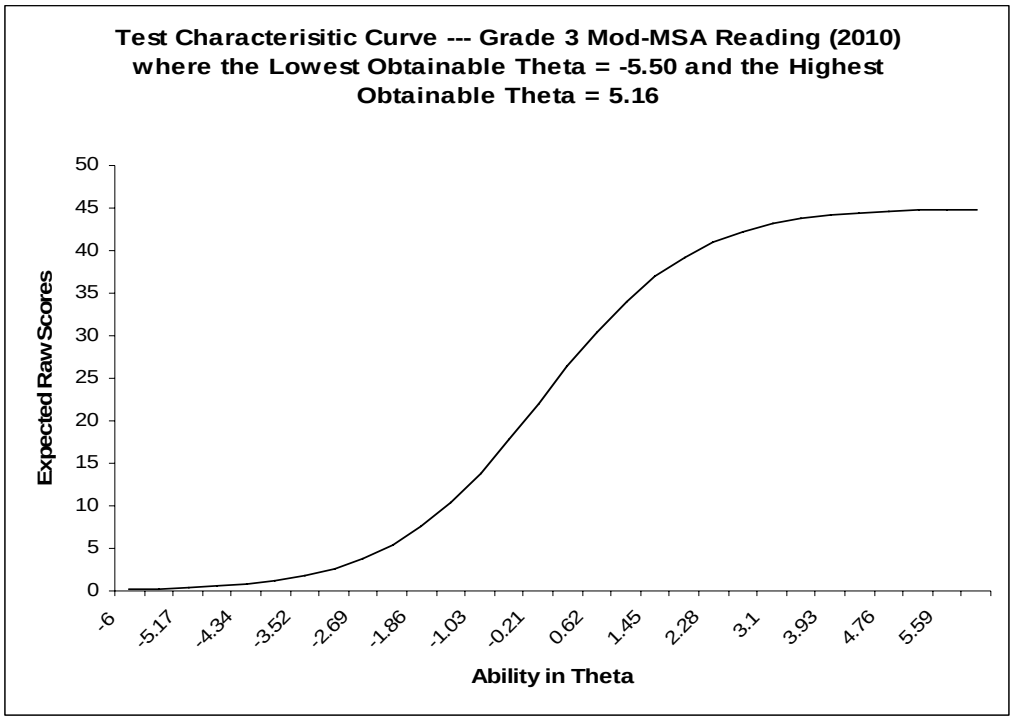



**Figure 9.4.1a. TCC for Grade 3 Mod-MSA: Reading**

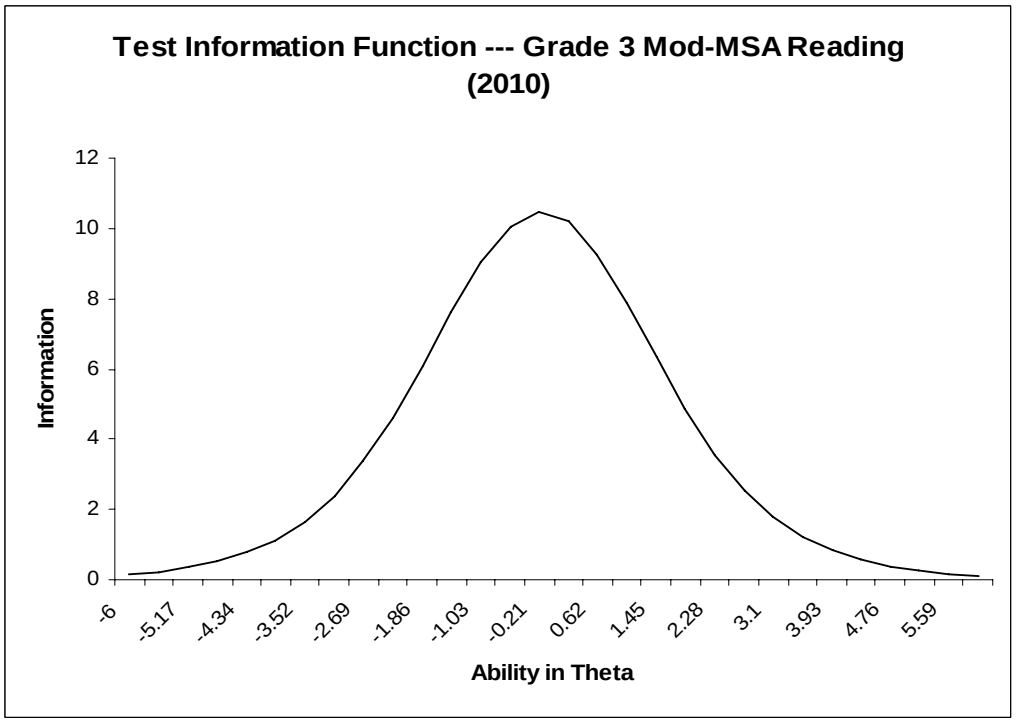



**Figure 9.4.1b. TIF for Grade 3 Mod-MSA: Reading**

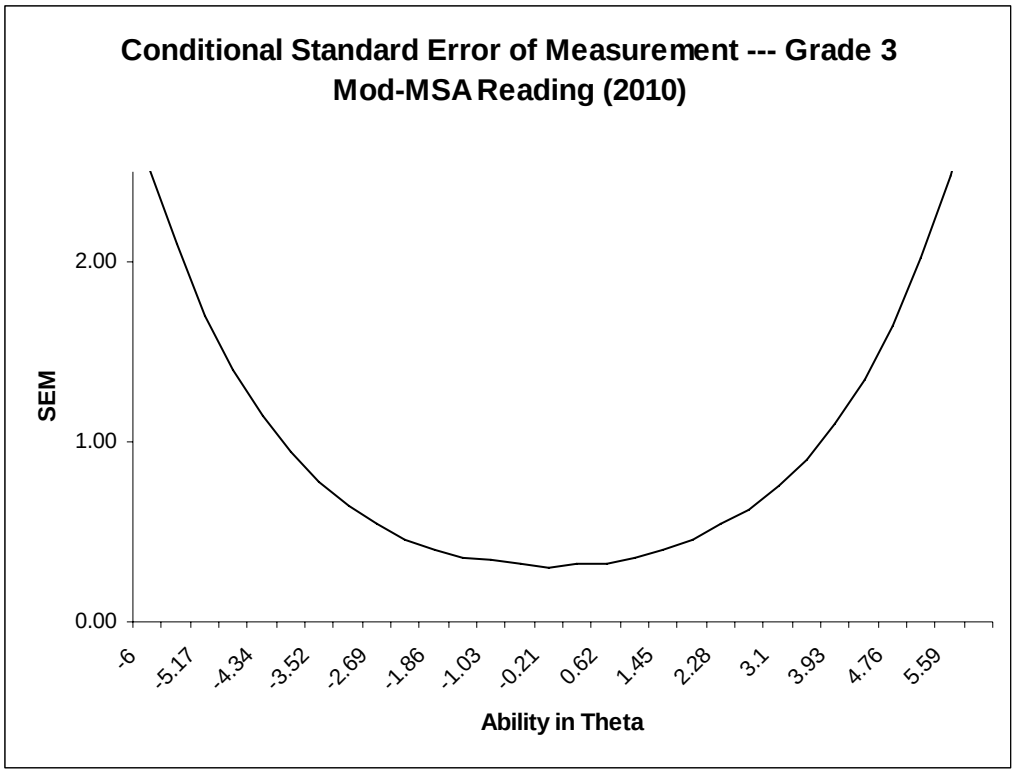



**Figure 9.4.1c. CSEM for Grade 3 Mod-MSA: Reading**

## Grade 4

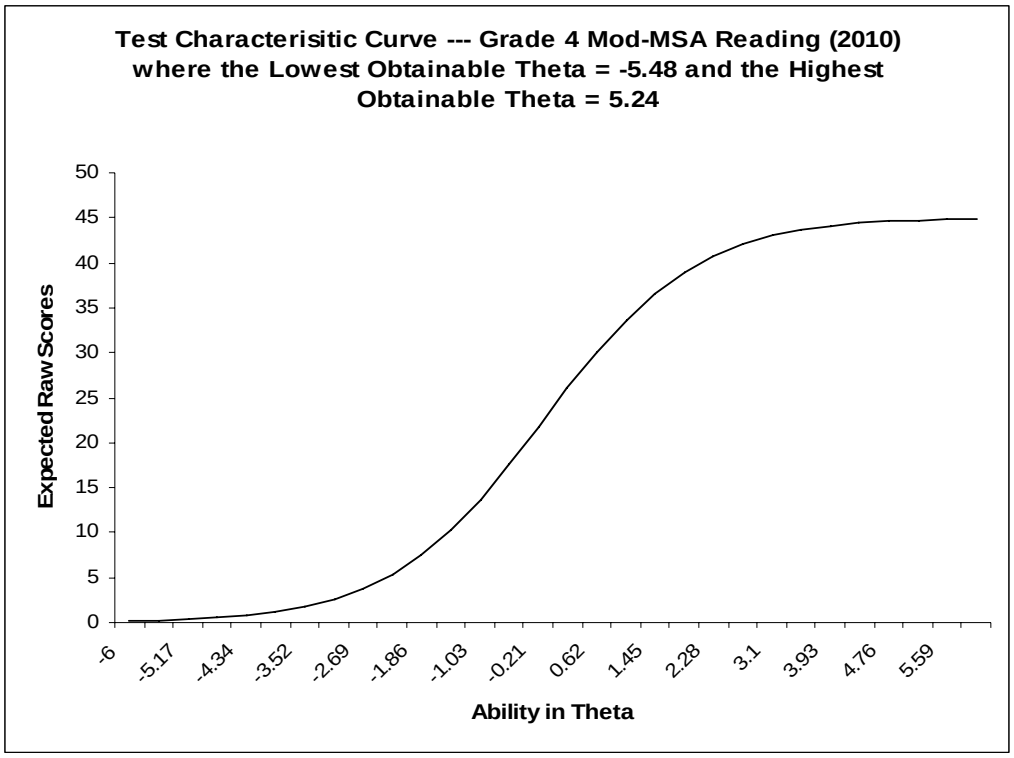



**Figure 9.4.2a. TCC for Grade 4 Mod-MSA: Reading**

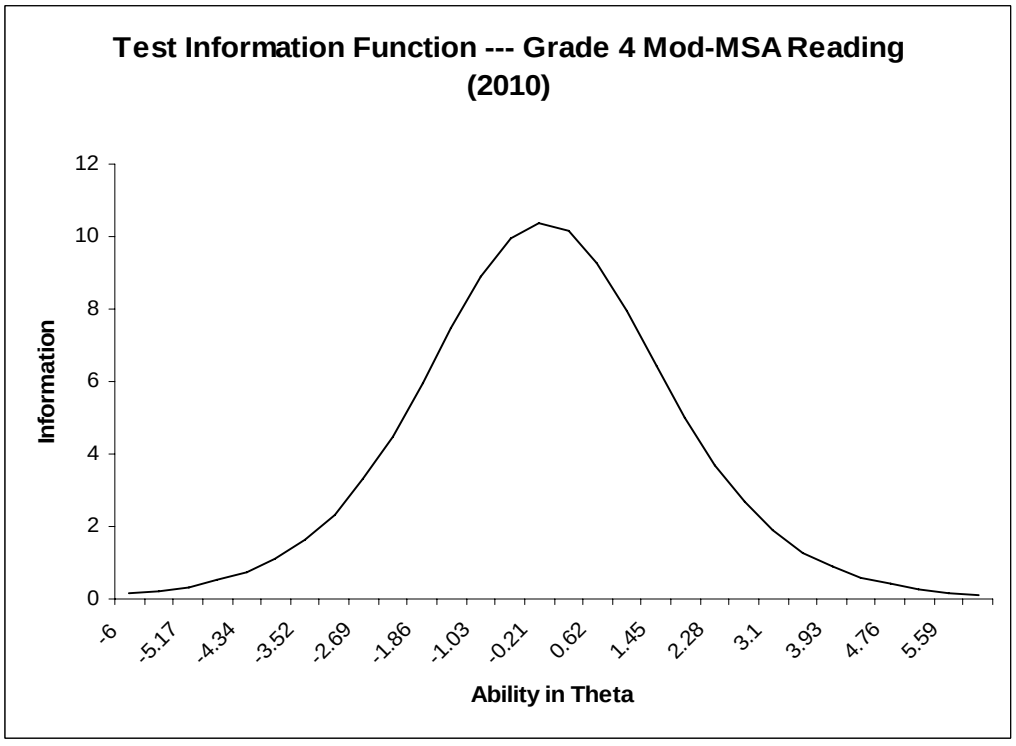



**Figure 9.4.2b. TIF for Grade 4 Mod-MSA: Reading**

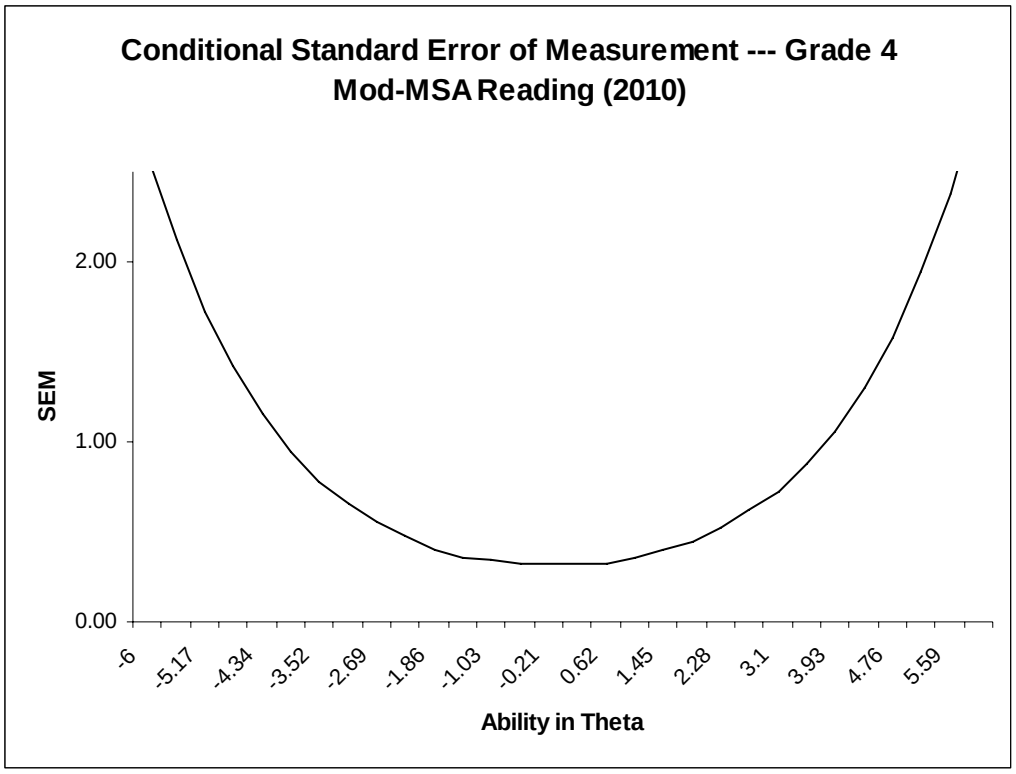



**Figure 9.4.2c. CSEM for Grade 4 Mod-MSA: Reading**

## Grade 5

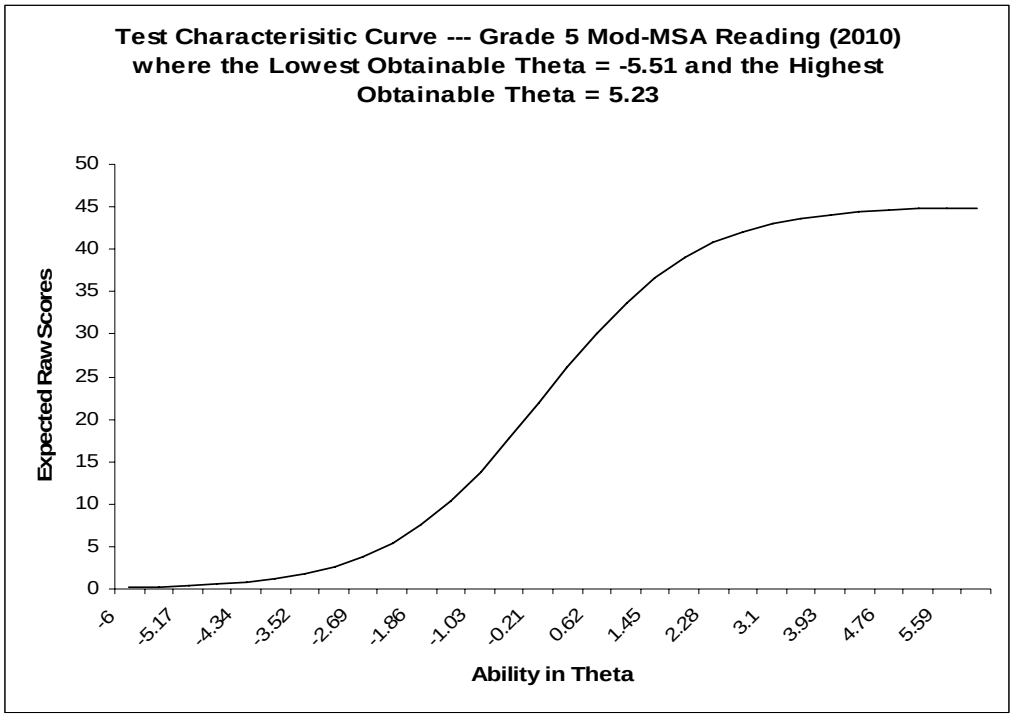



**Figure 9.4.3a. TCC for Grade 5 Mod-MSA: Reading**

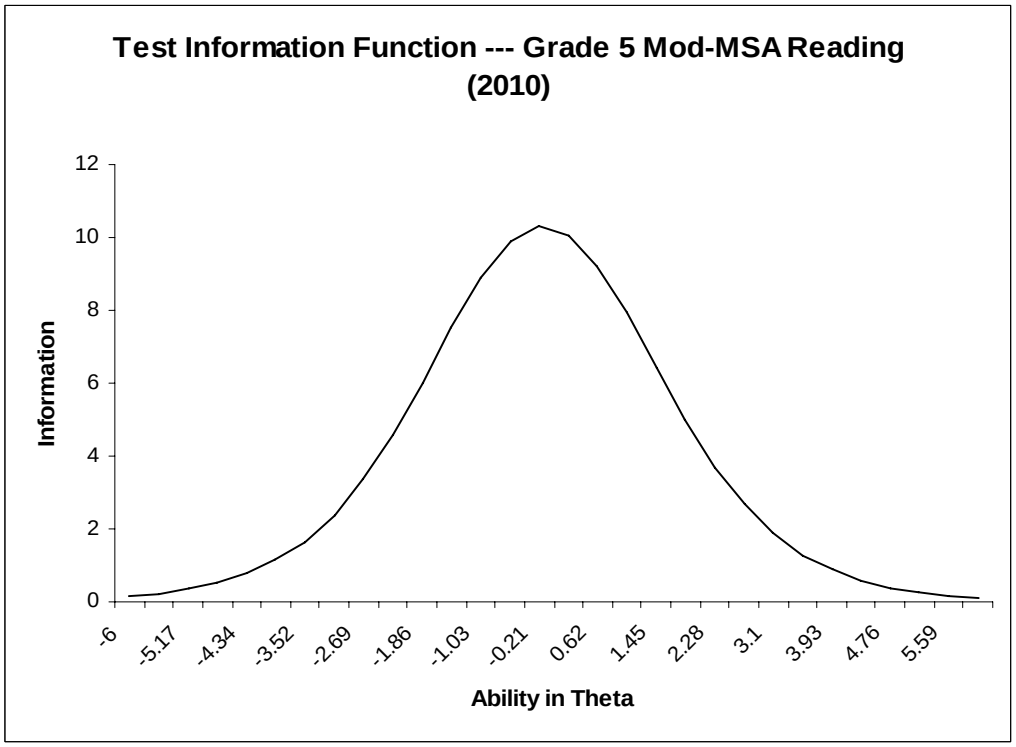



**Figure 9.4.3b. TIF for Grade 5 Mod-MSA: Reading**

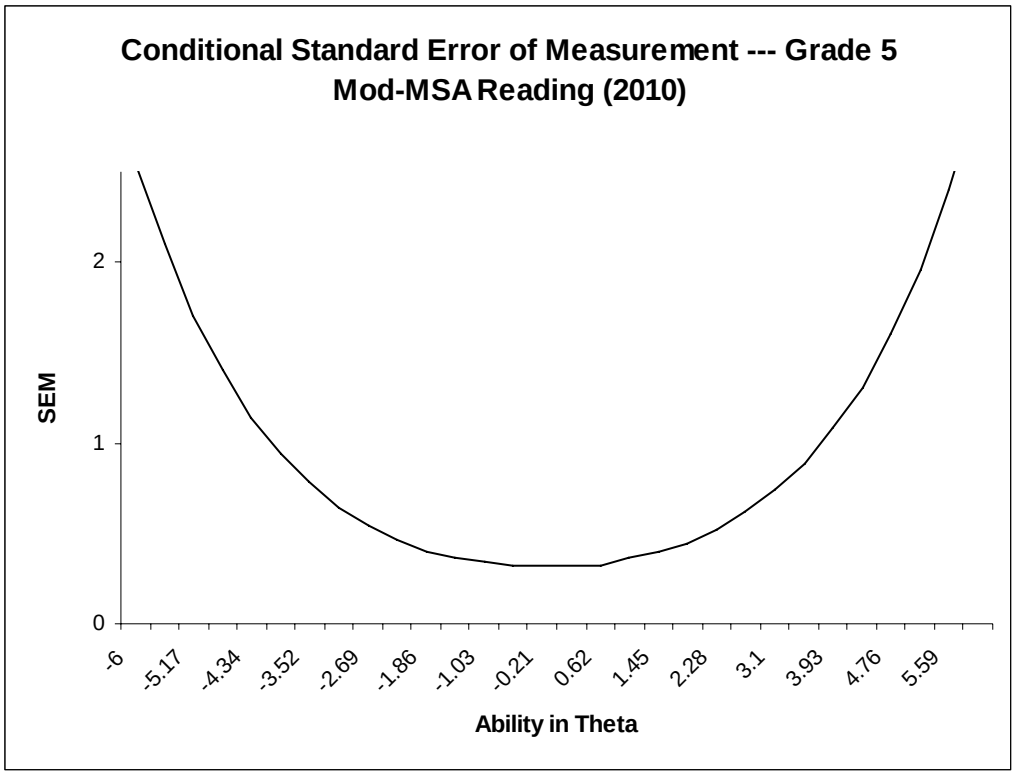



**Figure 9.4.3c. CSEM for Grade 5 Mod-MSA: Reading**

## Grade 6

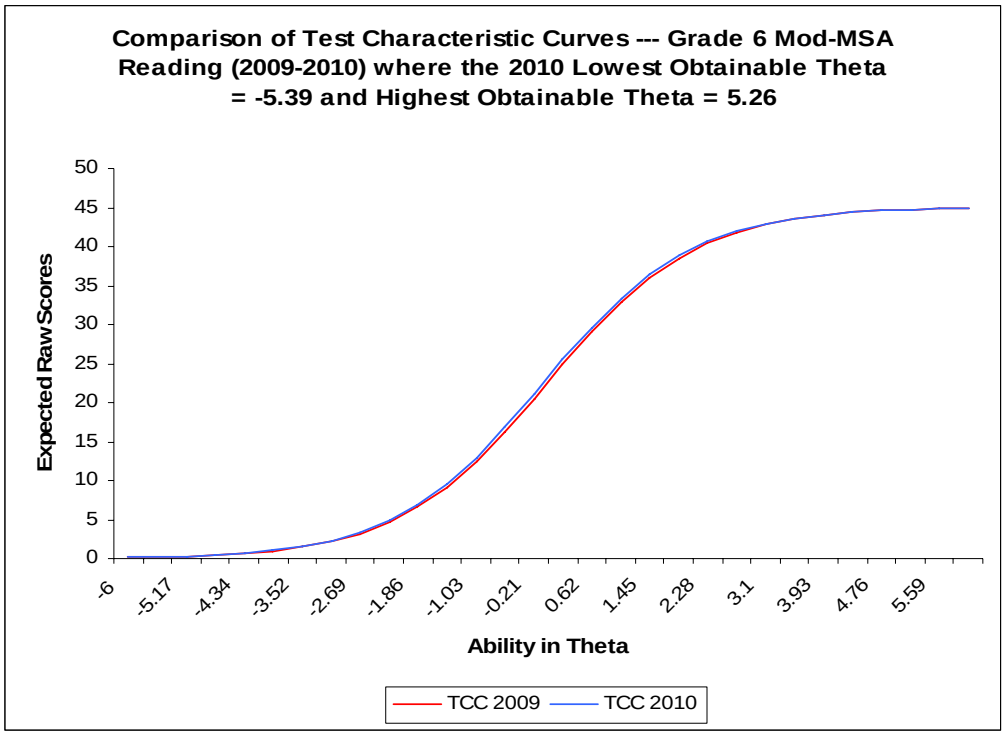



**Figure 9.4.4a. TCC Comparison for Grade 6 Mod-MSA: Reading**

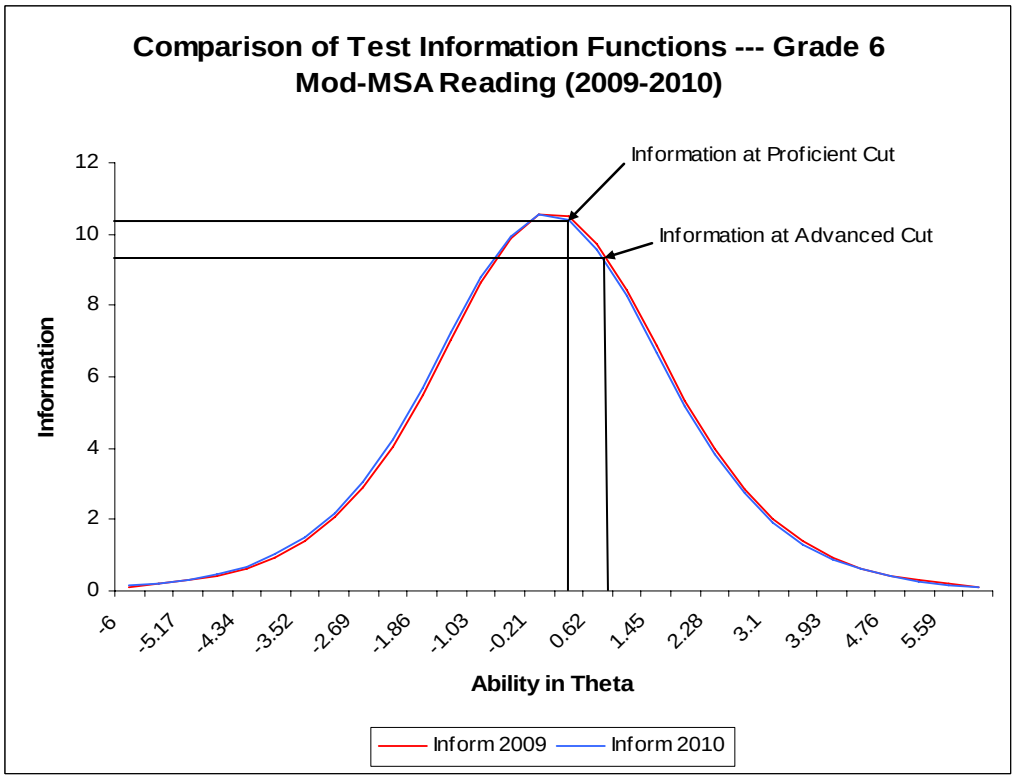



**Figure 9.4.4b. TIF Comparison for Grade 6 Mod-MSA: Reading (2009 with 2010)**

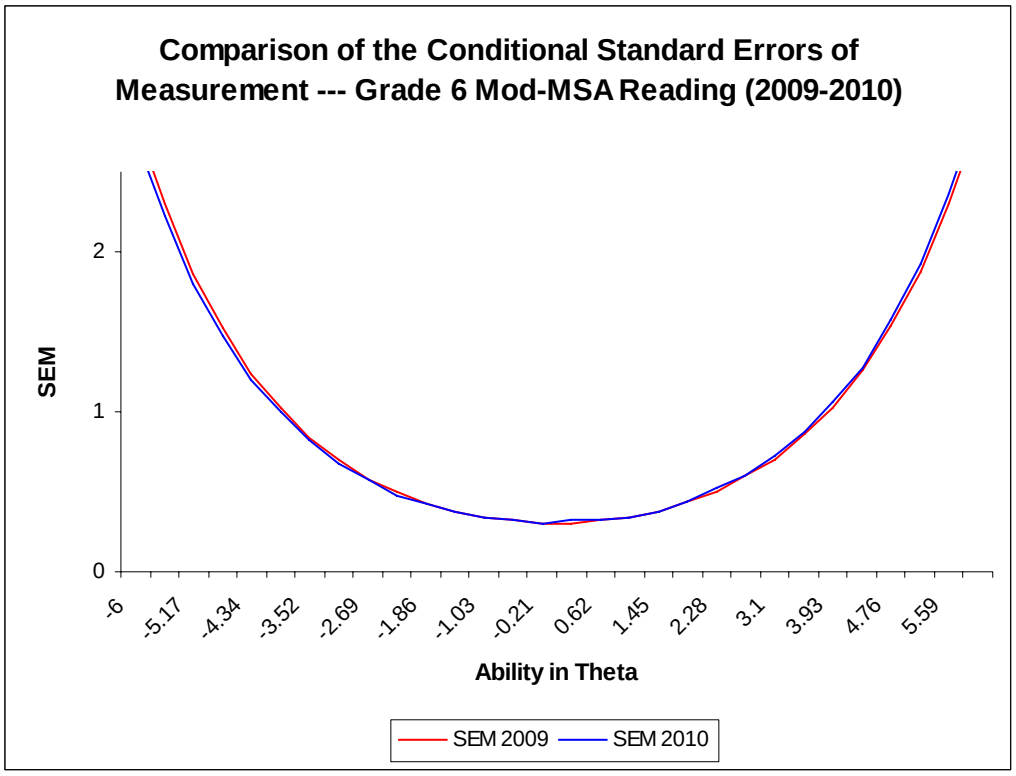



**Figure 9.4.4c. CSEM Comparison for Grade 6 Mod-MSA: Reading**

## Grade 7

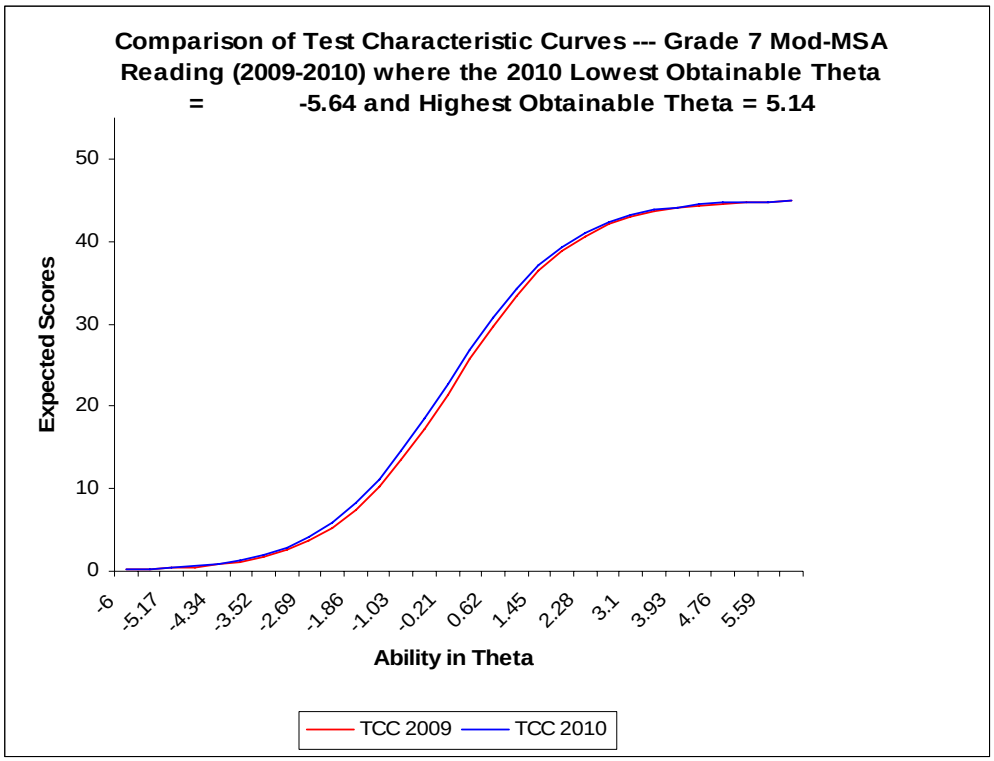



**Figure 9.4.5a. TCC Comparison for Grade 7 Mod-MSA: Reading**

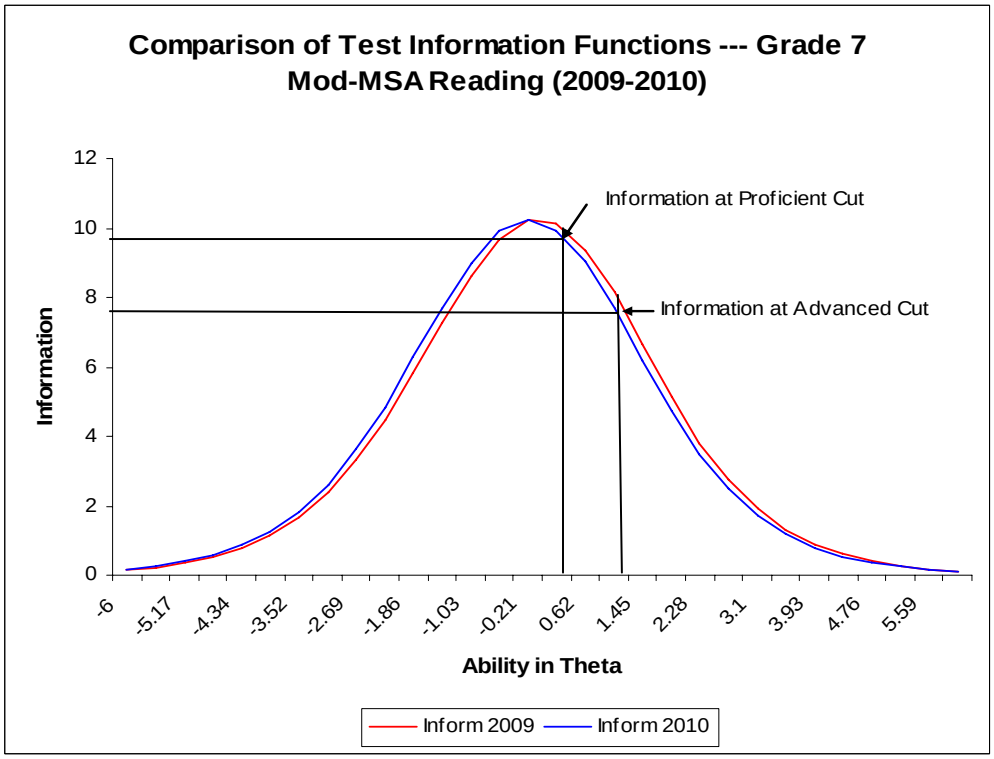



**Figure 9.5.5b. TIF Comparison for Grade 7 Mod-MSA: Reading**

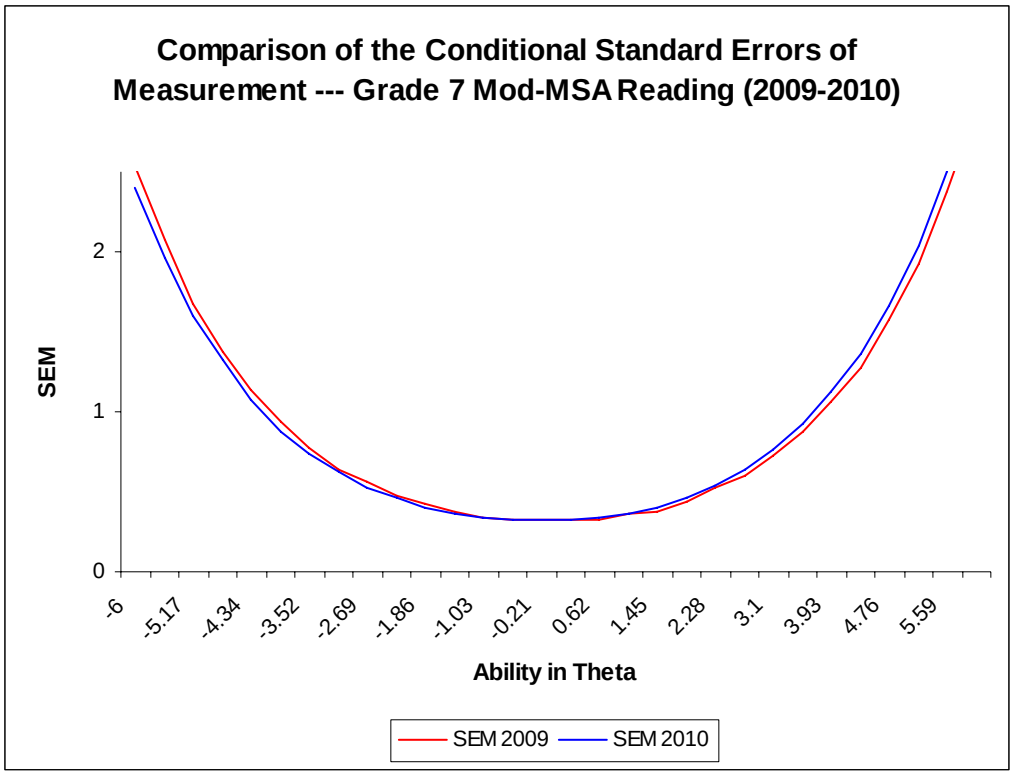



**Figure 9.5.5c. CSEM Comparison for Grade 7 Mod-MSA: Reading**

## Grade 8

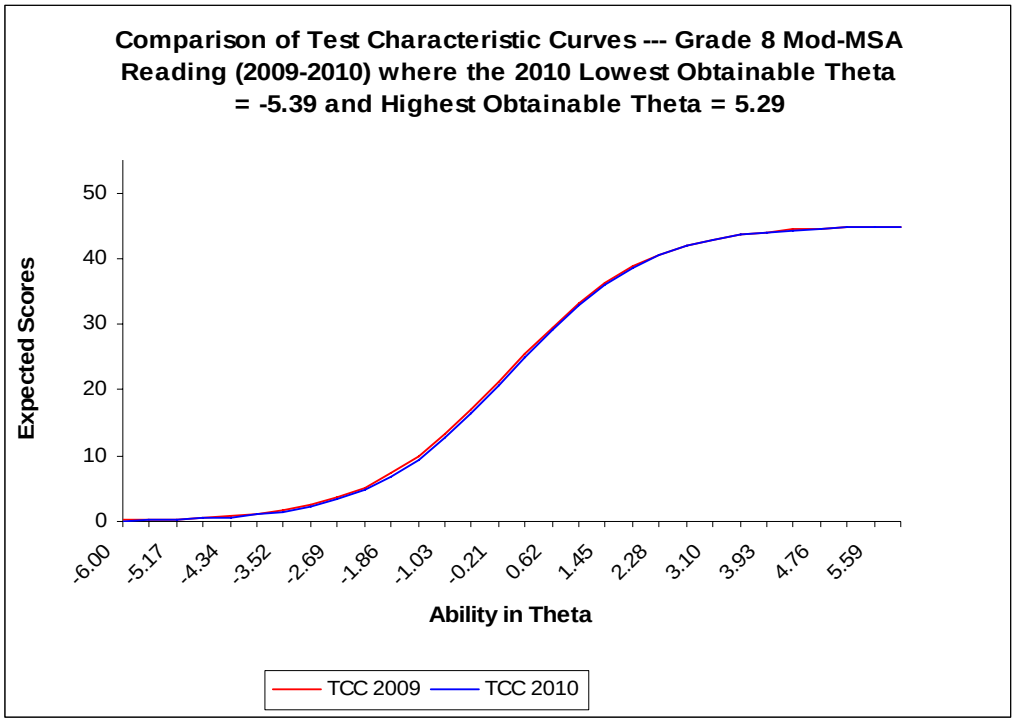



**Figure 9.4.6a. TCC Comparison for Grade 8 Mod-MSA: Reading**

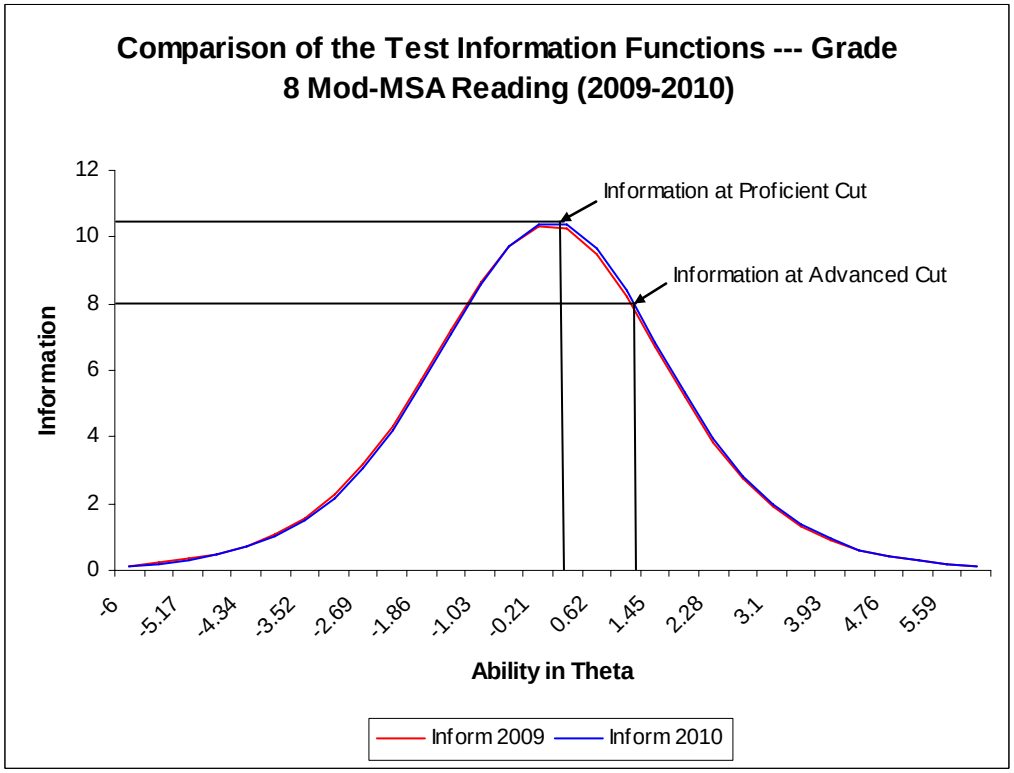



**Figure 9.4.6b. TIF Comparison for Grade 8 Mod-MSA: Reading**

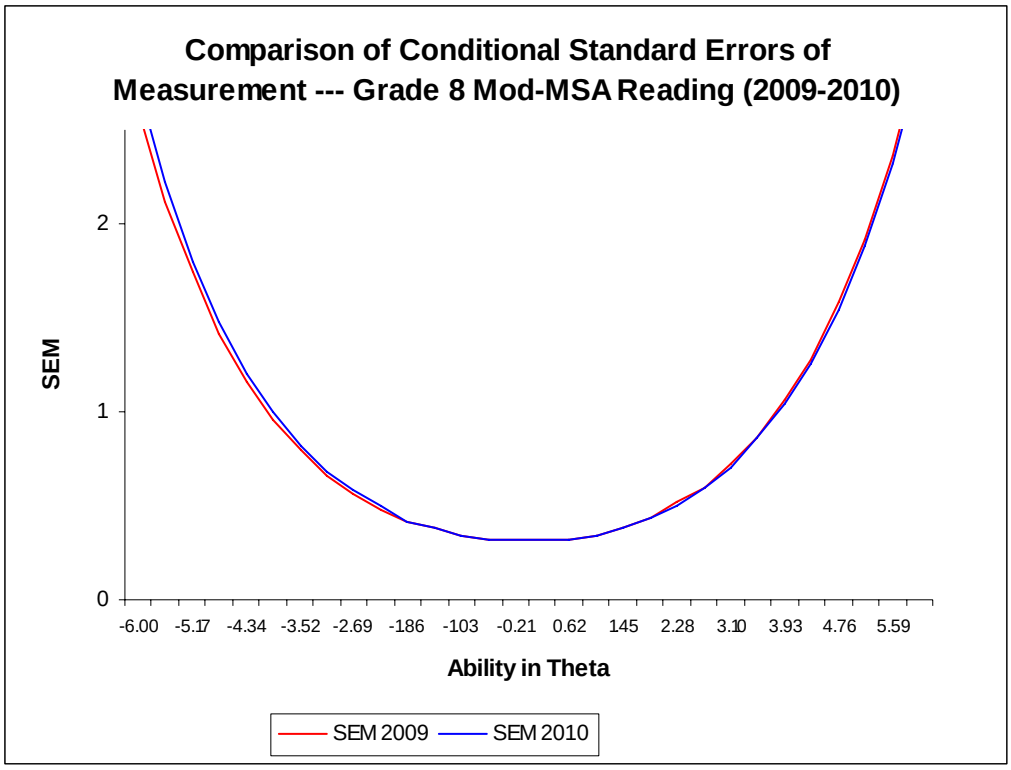



**Figure 9.4.6c. CSEM Comparison for Grade 8 Mod-MSA: Reading**

## 10. REFERENCES


AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, D.C.: Author.

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report* (Technical Report). Washington, DC: National Center for Educational Statistics.

Andrich, A. (1988). *Rasch models for measurement*. Newbury Park, CA: SAGE Publications, Inc.

Andrich, A. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, & H. H. Lovibond (Eds.) *Mathematical and theoretical systems*. North-Holland: Elsevier Science Publisher B.V.

Andrich, A., & Luo, G. (2004). *Modern measurement and analysis in social science*. Murdoch University, Perth, Western Australia.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: SAGE Publications.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt Rinehart Wilson.

CTB/McGraw-Hill (2004, August). *The Maryland standard setting technical report*. (Technical Report). Monterey, CA: CTB/McGraw-Hill.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed-response and differential item functioning: A pragmatic approach* (ETS Research Report No. 91-49). Princeton, NJ: Educational Testing Service.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists.* New Jersey: Lawrence Erlbaum Associates, Publishers.

Haertel, E. H. (1996). *Estimating the decision consistency from a single administration of a performance assessment battery. A report on the National Board of Professional Teaching Standards McGEN Assessment*. Palo Alto, CA: Stanford University.

Harcourt, Inc. (2008, January). *Maryland School Assessment-Reading and Mathematics: Test administration and coordination manual*. San Antonio, TX: Harcourt Inc.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practice,* 10, 181-189.

Huynh, H., Meyer III, J. P., & Barton, K. (2000). *Technical documentation for the 1999 Palmetto achievement challenge tests of English language arts and mathematics, grades three through eight* (Technical Report). Columbia: South Carolina Department of Education.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 & PRELIS 2: User's reference guide*. Chicago: Scientific Software International.

Linacre, J. M. (2003). *A user's guide to WINSTEPS: Rasch-model computer program.* ISBN 0-941938-03-4. Chicago, IL: MESA Press.

Linacre, J. M., & Wright, B. D. (1999). *A user's guide to WINSTEPS, BIGSTEPS, MINISTEP: Rasch-model computer program.* Chicago, IL: MESA Press.

Linacre, J. M., & Wright, B. D. (2000). *A user's guide to WINSTEPS: Rasch-model computer program.* Chicago, IL: MESA Press.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32,* 179-197.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8,* 452-461.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrica, 47,* 149-174.

Maryland State Department of Education. (2008). *2008 Maryland School Assessment-Grades 3 through 8 Reading and Mathematics: Test Administration and Coordination Manual*. Baltimore: Maryland State Department of Education.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18,* 5-11.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards* (pp. 249-282). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats, *Applied Measurement in Education,* 8, 111-120.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

Ryan, J. P. (1983). Introduction to latent trait analysis and item response theory. In W. E. Hathaway (Ed.), *Testing in the schools. New directions for testing and measurement, 19,* San Francisco: Jossey-Bass.

South Carolina Department of Education. (2001). *Technical documentation for the 2000 Palmetto achievement challenge tests of English language arts and mathematics* (Technical Report). Columbia: South Carolina Department of Education.

Suen, H. K. (1990). *Principles of test theories*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrica, 51,* 567-577.

Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment.* (CSE Technical Report 475). Center

for the Study of Evaluation, Standards, and Student Testing. Los Angeles, CA: University of California, Los Angeles.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233-251.

# APPENDIX A: OPERATIONAL ITEMS SELECTED AFTER DATA REVIEW

## Table A1. Grade 3 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 3 | 1 | 100000213631 | 35 | 100000260458 |
| 3 | 3 | 100000213633 | 36 | 100000260460 |
| 3 | 5 | 100000360182 | 37 | 100000260461 |
| 3 | 7 | 100000101513 | 39 | 100000260457 |
| 3 | 8 | 100000101514 | 41 | 100000260465 |
| 3 | 9 | 100000360183 | 42 | 100000101905 |
| 3 | 10 | 100000101516 | 44 | 100000101907 |
| 3 | 11 | 100000101517 | 45 | 100000101908 |
| 3 | 12 | 100000101518 | 46 | 100000101909 |
| 3 | 14 | 100000101969 | 48 | 100000101911 |
| 3 | 17 | 100000101972 | 50 | 100000101530 |
| 3 | 18 | 100000101973 | 52 | 100000101532 |
| 3 | 19 | 100000101974 | 54 | 100000101528 |
| 3 | 21 | 100000101932 | 55 | 100000101533 |
| 3 | 22 | 100000101934 | 56 | 100000260338 |
| 3 | 25 | 100000101936 | 57 | 100000260342 |
| 3 | 26 | 100000101937 | 58 | 100000260345 |
| 3 | 27 | 100000101938 | 61 | 100000260339 |
| 3 | 28 | 100000346452 | 64 | 100000260368 |
| 3 | 30 | 100000346444 | 65 | 100000260371 |
| 3 | 32 | 100000346450 | 68 | 100000260365 |
| 3 | 33 | 100000346445 | 69 | 100000300552 |
| 3 | 34 | 100000346448 | | |

## Table A2. Grade 4 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 4 | 1 | 100000213644 | 34 | 100000269896 |
| 4 | 2 | 100000213645 | 37 | 100000269900 |
| 4 | 3 | 100000213646 | 38 | 100000269897 |
| 4 | 4 | 100000213647 | 40 | 100000269899 |
| 4 | 6 | 100000213637 | 43 | 100000357098 |
| 4 | 7 | 100000213638 | 45 | 100000357100 |
| 4 | 10 | 100000213641 | 46 | 100000357101 |
| 4 | 11 | 100000213642 | 47 | 100000462157 |
| 4 | 12 | 100000360192 | 48 | 100000101997 |
| 4 | 13 | 100000357134 | 51 | 100000101999 |
| 4 | 15 | 100000357133 | 52 | 100000102000 |
| 4 | 17 | 100000357136 | 53 | 100000102001 |
| 4 | 18 | 100000357137 | 54 | 100000200070 |
| 4 | 19 | 100000357138 | 56 | 100000102026 |
| 4 | 20 | 100000357106 | 57 | 100000102027 |
| 4 | 21 | 100000357104 | 59 | 100000102024 |
| 4 | 22 | 100000357107 | 60 | 100000102028 |
| 4 | 24 | 100000357108 | 61 | 100000102029 |
| 4 | 25 | 100000357109 | 63 | 100000267470 |
| 4 | 28 | 100000260486 | 65 | 100000301034 |
| 4 | 29 | 100000260483 | 66 | 100000267473 |
| 4 | 31 | 100000260489 | 68 | 100000267472 |
| 4 | 32 | 100000260488 | | |

## Table A3. Grade 5 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 5 | 1 | 100000213650 | 34 | 100000102084 |
| 5 | 2 | 100000213651 | 35 | 100000102087 |
| 5 | 3 | 100000213652 | 36 | 100000102088 |
| 5 | 4 | 100000213653 | 39 | 100000102090 |
| 5 | 6 | 100000213655 | 41 | 100000102048 |
| 5 | 7 | 100000213656 | 43 | 100000102050 |
| 5 | 8 | 100000213657 | 44 | 100000102051 |
| 5 | 10 | 100000213659 | 45 | 100000102052 |
| 5 | 11 | 100000360197 | 46 | 100000102047 |
| 5 | 12 | 100000102095 | 47 | 100000102056 |
| 5 | 13 | 100000102093 | 49 | 100000102059 |
| 5 | 15 | 100000102096 | 50 | 100000102060 |
| 5 | 17 | 100000102098 | 51 | 100000102061 |
| 5 | 18 | 100000102099 | 54 | 100000102068 |
| 5 | 19 | 100000102113 | 55 | 100000102069 |
| 5 | 20 | 100000102111 | 57 | 100000102067 |
| 5 | 21 | 100000102114 | 59 | 100000102071 |
| 5 | 22 | 100000102115 | 60 | 100000102072 |
| 5 | 23 | 100000102116 | 61 | 100000267477 |
| 5 | 26 | 100000102104 | 62 | 100000268378 |
| 5 | 30 | 100000102106 | 64 | 100000268380 |
| 5 | 31 | 100000102107 | 67 | 100000303033 |
| 5 | 32 | 100000102108 | | |

## Table A4. Grade 6 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 6 | 1 | 100000213662 | 30 | 100000270018 |
| 6 | 2 | 100000213663 | 31 | 100000270021 |
| 6 | 3 | 100000213664 | 34 | 100000270013 |
| 6 | 4 | 100000213665 | 35 | 100000270020 |
| 6 | 5 | 100000213666 | 37 | 100000257088 |
| 6 | 7 | 100000213668 | 39 | 100000257114 |
| 6 | 8 | 100000213669 | 40 | 100000257118 |
| 6 | 9 | 100000213670 | 41 | 100000257119 |
| 6 | 10 | 100000213671 | 42 | 100000257120 |
| 6 | 12 | 100000102192 | 43 | 100000257102 |
| 6 | 13 | 100000102194 | 45 | 100000257103 |
| 6 | 14 | 100000102191 | 46 | 100000257104 |
| 6 | 15 | 100000102195 | 47 | 100000257071 |
| 6 | 16 | 100000102196 | 50 | 100000102136 |
| 6 | 17 | 100000102174 | 51 | 100000102137 |
| 6 | 18 | 100000102175 | 52 | 100000102139 |
| 6 | 19 | 100000102173 | 53 | 100000102140 |
| 6 | 20 | 100000102176 | 55 | 100000257110 |
| 6 | 21 | 100000102178 | 56 | 100000257083 |
| 6 | 23 | 100000269997 | 57 | 100000257106 |
| 6 | 24 | 100000270006 | 58 | 100000257107 |
| 6 | 27 | 100000270005 | 59 | 100000257109 |
| 6 | 28 | 100000270007 | | |

## Table A5. Grade 7 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 7 | 1 | 100000213674 | 30 | 100000102269 |
| 7 | 2 | 100000213675 | 31 | 100000102266 |
| 7 | 3 | 100000213676 | 33 | 100000257182 |
| 7 | 5 | 100000213678 | 34 | 100000257216 |
| 7 | 6 | 100000213679 | 35 | 100000257218 |
| 7 | 8 | 100000270578 | 36 | 100000257219 |
| 7 | 9 | 100000270572 | 38 | 100000257221 |
| 7 | 10 | 100000270574 | 39 | 100000102244 |
| 7 | 11 | 100000322011 | 40 | 100000102246 |
| 7 | 14 | 100000270575 | 41 | 100000102247 |
| 7 | 15 | 100000270109 | 42 | 100000102248 |
| 7 | 16 | 100000270112 | 43 | 100000102245 |
| 7 | 17 | 100000270113 | 44 | 100000102218 |
| 7 | 18 | 100000270114 | 45 | 100000102220 |
| 7 | 21 | 100000270121 | 46 | 100000102221 |
| 7 | 22 | 100000102276 | 47 | 100000102222 |
| 7 | 23 | 100000102277 | 48 | 100000102217 |
| 7 | 24 | 100000102274 | 50 | 100000257769 |
| 7 | 25 | 100000102278 | 51 | 100000257770 |
| 7 | 26 | 100000102279 | 52 | 100000257773 |
| 7 | 27 | 100000102263 | 53 | 100000322795 |
| 7 | 28 | 100000102267 | 54 | 100000257767 |
| 7 | 29 | 100000102268 | | |

## Table A6. Grade 8 Operational Items

| Grade | Sequence No. | UIN No. | Sequence No. | UIN No. |
|---|---|---|---|---|
| 8 | 1 | 100000213683 | 30 | 100000102346 |
| 8 | 2 | 100000213684 | 31 | 100000102345 |
| 8 | 4 | 100000213686 | 32 | 100000257161 |
| 8 | 5 | 100000213687 | 34 | 100000257787 |
| 8 | 6 | 100000213688 | 36 | 100000257158 |
| 8 | 8 | 100000270136 | 37 | 100000257160 |
| 8 | 10 | 100000273591 | 38 | 100000257163 |
| 8 | 11 | 100000270131 | 39 | 100000102320 |
| 8 | 12 | 100000270135 | 40 | 100000102322 |
| 8 | 13 | 100000270137 | 41 | 100000102323 |
| 8 | 15 | 100000102331 | 42 | 100000102324 |
| 8 | 16 | 100000102329 | 43 | 100000102321 |
| 8 | 17 | 100000102330 | 45 | 100000257140 |
| 8 | 18 | 100000102334 | 47 | 100000257144 |
| 8 | 19 | 100000102335 | 48 | 100000257777 |
| 8 | 20 | 100000270154 | 49 | 100000302674 |
| 8 | 22 | 100000270158 | 50 | 100000302675 |
| 8 | 23 | 100000273597 | 51 | 100000102304 |
| 8 | 24 | 100000270163 | 52 | 100000102305 |
| 8 | 25 | 100000270159 | 53 | 100000102302 |
| 8 | 27 | 100000102341 | 54 | 100000102307 |
| 8 | 28 | 100000102343 | 55 | 100000102308 |
| 8 | 29 | 100000102344 | | |

# APPENDIX B: ITEM DISTRACTOR ANALYSES

## Table B.1. The 2010 Mod-MSA: Reading Grade 3, Distractor Distribution and Distractor-to- Total Correlation Analysis

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213631 | 1 | MC | 813 | 0.73 | 0.44 | 13.53 | 8.36 | 72.94 | 5.04 | 0.12 | -0.16 | -0.21 | | -0.12 |
| 100000360181 | 2 | MC | 813 | 0.29 | 0.45 | 26.94 | 16.85 | 28.91 | 27.18 | 0.12 | -0.11 | -0.06 | | -0.11 |
| 100000213633 | 3 | MC | 813 | 0.80 | 0.40 | 4.80 | 5.29 | 10.21 | 79.70 | | -0.21 | -0.22 | -0.32 | |
| 100000213634 | 4 | MC | 813 | 0.38 | 0.48 | 25.46 | 24.23 | 37.52 | 12.67 | 0.12 | -0.12 | -0.12 | | -0.18 |
| 100000360182 | 5 | MC | 813 | 0.42 | 0.49 | 13.41 | 42.44 | 13.41 | 30.75 | | -0.23 | | -0.21 | -0.01 |
| 100000101512 | 6 | MC | 813 | 0.51 | 0.50 | 50.68 | 25.46 | 23.25 | 0.62 | | | -0.14 | -0.13 | -0.13 |
| 100000101513 | 7 | MC | 813 | 0.56 | 0.50 | 26.57 | 55.60 | 17.10 | 0.74 | | 0.00 | | -0.31 | -0.12 |
| 100000101514 | 8 | MC | 813 | 0.62 | 0.49 | 27.55 | 8.73 | 62.24 | 1.48 | | -0.28 | -0.13 | | -0.16 |
| 100000360183 | 9 | MC | 813 | 0.58 | 0.49 | 57.56 | 21.16 | 20.17 | 0.98 | 0.12 | | -0.08 | -0.17 | -0.14 |
| 100000101516 | 10 | MC | 813 | 0.42 | 0.49 | 38.38 | 19.31 | 41.70 | 0.62 | | -0.12 | -0.09 | | -0.13 |
| 100000101517 | 11 | MC | 813 | 0.32 | 0.47 | 21.03 | 31.86 | 43.30 | 3.81 | | -0.08 | | -0.02 | -0.28 |
| 100000101518 | 12 | MC | 813 | 0.56 | 0.50 | 22.02 | 20.42 | 56.09 | 1.35 | 0.12 | -0.18 | -0.02 | | -0.15 |
| 100000360184 | 13 | MC | 813 | 0.48 | 0.50 | 31.61 | 47.85 | 20.17 | | 0.37 | -0.21 | | -0.05 | |
| 100000101969 | 14 | MC | 813 | 0.69 | 0.46 | 13.28 | 17.71 | 68.63 | | 0.37 | -0.32 | -0.26 | | |
| 100000101970 | 15 | MC | 813 | 0.50 | 0.50 | 50.31 | 15.13 | 34.32 | | 0.25 | | -0.17 | -0.24 | |
| 100000101971 | 16 | MC | 813 | 0.40 | 0.49 | 29.27 | 40.22 | 30.26 | | 0.25 | -0.02 | | -0.21 | |
| 100000101972 | 17 | MC | 813 | 0.46 | 0.50 | 22.39 | 31.61 | 45.63 | | 0.37 | -0.08 | -0.07 | | |
| 100000101973 | 18 | MC | 813 | 0.37 | 0.48 | 36.65 | 20.17 | 42.80 | | 0.37 | | -0.05 | -0.12 | |
| 100000101974 | 19 | MC | 813 | 0.43 | 0.50 | 33.70 | 43.17 | 22.88 | | 0.25 | -0.16 | | -0.19 | |
| 100000101975 | 20 | MC | 813 | 0.45 | 0.50 | 13.41 | 41.33 | 44.65 | | 0.62 | -0.18 | -0.03 | | |
| 100000101932 | 21 | MC | 813 | 0.59 | 0.49 | 58.67 | 26.32 | 14.76 | | 0.25 | | -0.05 | -0.29 | |
| 100000101934 | 22 | MC | 813 | 0.52 | 0.50 | 17.34 | 30.75 | 51.66 | | 0.25 | -0.19 | -0.17 | | |
| 100000101935 | 23 | MC | 813 | 0.40 | 0.49 | 29.77 | 29.77 | 39.98 | | 0.49 | -0.11 | 0.05 | | |
| 100000101933 | 24 | MC | 813 | 0.37 | 0.48 | 26.45 | 36.53 | 36.41 | | 0.62 | -0.22 | | -0.02 | |
| 100000101936 | 25 | MC | 813 | 0.52 | 0.50 | 29.27 | 17.96 | 52.40 | | 0.37 | -0.10 | -0.14 | | |
| 100000101937 | 26 | MC | 813 | 0.63 | 0.48 | 63.47 | 16.36 | 19.43 | | 0.74 | | -0.15 | -0.29 | |
| 100000101938 | 27 | MC | 813 | 0.47 | 0.50 | 18.94 | 47.48 | 32.84 | | 0.74 | -0.21 | | -0.14 | |
| 100000346452 | 28 | MC | 813 | 0.60 | 0.49 | 19.43 | 19.93 | 60.39 | | 0.25 | -0.20 | -0.15 | | |
| 100000346453 | 29 | MC | 813 | 0.59 | 0.49 | 59.16 | 20.42 | 20.17 | | 0.25 | | -0.19 | -0.25 | |
| 100000346444 | 30 | MC | 813 | 0.44 | 0.50 | 43.54 | 25.22 | 30.87 | | 0.37 | | -0.09 | -0.18 | |
| 100000346446 | 31 | MC | 813 | 0.39 | 0.49 | 39.48 | 24.11 | 36.16 | | 0.25 | | 0.03 | -0.10 | |
| 100000346450 | 32 | MC | 813 | 0.74 | 0.44 | 11.69 | 14.15 | 73.92 | | 0.25 | -0.23 | -0.28 | | |
| 100000346445 | 33 | MC | 813 | 0.46 | 0.50 | 46.00 | 16.36 | 37.39 | | 0.25 | | -0.17 | -0.24 | |
| 100000346448 | 34 | MC | 813 | 0.49 | 0.50 | 12.67 | 49.20 | 37.88 | | 0.25 | -0.21 | | -0.12 | |
| 100000260458 | 35 | MC | 813 | 0.42 | 0.49 | 31.12 | 26.57 | 41.57 | | 0.74 | -0.06 | -0.12 | | |
| 100000260460 | 36 | MC | 813 | 0.59 | 0.49 | 17.59 | 22.51 | 59.16 | | 0.74 | -0.06 | -0.17 | | |
| 100000260461 | 37 | MC | 813 | 0.56 | 0.50 | 14.39 | 56.33 | 28.54 | | 0.74 | -0.09 | | -0.24 | |
| 100000312661 | 38 | MC | 813 | 0.58 | 0.49 | 16.61 | 24.11 | 58.30 | | 0.98 | -0.22 | -0.10 | | |
| 100000260457 | 39 | MC | 813 | 0.51 | 0.50 | 26.08 | 21.77 | 51.41 | | 0.74 | -0.06 | -0.15 | | |

Table B.1. The 2010 Mod-MSA: Reading Grade 3, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000260459 | 40 | MC | 813 | 0.37 | 0.48 | 34.07 | 36.65 | 28.41 | | 0.86 | 0.01 | | -0.19 | |
| 100000260465 | 41 | MC | 813 | 0.58 | 0.49 | 58.30 | 27.92 | 12.79 | | 0.98 | | -0.07 | -0.26 | |
| 100000101905 | 42 | MC | 813 | 0.74 | 0.44 | 6.89 | 18.70 | 74.05 | | 0.37 | -0.29 | -0.19 | | |
| 100000101906 | 43 | MC | 813 | 0.40 | 0.49 | 14.76 | 40.10 | 44.65 | | 0.49 | -0.13 | | -0.17 | |
| 100000101907 | 44 | MC | 813 | 0.48 | 0.50 | 48.34 | 33.09 | 18.20 | | 0.37 | | -0.10 | -0.27 | |
| 100000101908 | 45 | MC | 813 | 0.71 | 0.45 | 9.96 | 70.85 | 18.82 | | 0.37 | -0.27 | | -0.32 | |
| 100000101909 | 46 | MC | 813 | 0.58 | 0.49 | 57.93 | 14.27 | 27.43 | | 0.37 | | -0.25 | -0.20 | |
| 100000101910 | 47 | MC | 813 | 0.56 | 0.50 | 20.79 | 22.76 | 56.09 | | 0.37 | 0.00 | -0.05 | | |
| 100000101911 | 48 | MC | 813 | 0.70 | 0.46 | 9.72 | 69.99 | 19.93 | | 0.37 | -0.22 | | -0.25 | |
| 100000101529 | 49 | MC | 813 | 0.83 | 0.38 | 10.46 | 82.78 | 6.40 | | 0.37 | -0.25 | | -0.24 | |
| 100000101530 | 50 | MC | 813 | 0.65 | 0.48 | 64.82 | 19.80 | 14.88 | | 0.49 | | -0.26 | -0.27 | |
| 100000101531 | 51 | MC | 813 | 0.24 | 0.43 | 24.11 | 31.49 | 44.03 | | 0.37 | | -0.10 | 0.06 | |
| 100000101532 | 52 | MC | 813 | 0.46 | 0.50 | 32.72 | 20.54 | 46.25 | | 0.49 | 0.12 | -0.27 | | |
| 100000101527 | 53 | MC | 813 | 0.60 | 0.49 | 19.56 | 60.27 | 19.68 | | 0.49 | -0.21 | | -0.18 | |
| 100000101528 | 54 | MC | 813 | 0.53 | 0.50 | 52.52 | 24.60 | 22.39 | | 0.49 | | -0.08 | -0.31 | |
| 100000101533 | 55 | MC | 813 | 0.69 | 0.46 | 11.44 | 18.57 | 69.50 | | 0.49 | -0.18 | -0.20 | | |
| 100000260338 | 56 | MC | 813 | 0.65 | 0.48 | 16.85 | 64.58 | 18.08 | | 0.49 | -0.10 | | -0.15 | |
| 100000260342 | 57 | MC | 813 | 0.51 | 0.50 | 14.27 | 34.07 | 51.17 | | 0.49 | -0.11 | -0.22 | | |
| 100000260345 | 58 | MC | 813 | 0.32 | 0.47 | 31.73 | 31.73 | 36.04 | | 0.49 | -0.08 | | -0.04 | |
| 100000365154 | 59 | MC | 813 | 0.44 | 0.50 | 21.28 | 44.28 | 33.95 | | 0.49 | -0.02 | | -0.23 | |
| 100000260349 | 60 | MC | 813 | 0.23 | 0.42 | 22.76 | 46.25 | 30.63 | | 0.37 | | 0.02 | -0.14 | |
| 100000260339 | 61 | MC | 813 | 0.57 | 0.50 | 56.58 | 13.53 | 29.27 | | 0.62 | | -0.30 | -0.14 | |
| 100000300707 | 62 | MC | 813 | 0.40 | 0.49 | 29.89 | 40.34 | 29.27 | | 0.49 | -0.11 | | -0.08 | |
| 100000260364 | 63 | MC | 813 | 0.38 | 0.49 | 37.88 | 25.34 | 36.41 | | 0.37 | | -0.03 | 0.00 | |
| 100000260368 | 64 | MC | 813 | 0.67 | 0.47 | 10.82 | 67.16 | 21.65 | | 0.37 | -0.16 | | -0.25 | |
| 100000260371 | 65 | MC | 813 | 0.51 | 0.50 | 16.36 | 31.86 | 51.29 | | 0.49 | -0.13 | -0.05 | | |
| 100000300557 | 66 | MC | 813 | 0.39 | 0.49 | 39.48 | 27.06 | 32.84 | | 0.62 | | -0.16 | -0.13 | |
| 100000260372 | 67 | MC | 813 | 0.74 | 0.44 | 73.92 | 8.86 | 16.85 | | 0.37 | | -0.29 | -0.27 | |
| 100000260365 | 68 | MC | 813 | 0.73 | 0.45 | 15.01 | 11.81 | 72.82 | | 0.37 | -0.24 | -0.19 | | |
| 100000300552 | 69 | MC | 813 | 0.65 | 0.48 | 11.32 | 64.70 | 18.20 | | 5.78 | -0.10 | | -0.30 | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

## Table B.2. The 2010 Mod-MSA: Reading Grade 4, Distractor Distribution and Distractor-to- Total Correlation Analysis

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213644 | 1 | MC | 967 | 0.36 | 0.48 | 35.88 | 26.89 | 36.19 | 1.03 | | | -0.06 | -0.05 | -0.17 |
| 100000213645 | 2 | MC | 967 | 0.50 | 0.50 | 49.74 | 27.40 | 20.68 | 2.17 | | | -0.02 | -0.08 | -0.21 |
| 100000213646 | 3 | MC | 967 | 0.47 | 0.50 | 20.17 | 32.47 | 46.74 | 0.62 | | -0.07 | -0.05 | | -0.13 |
| 100000213647 | 4 | MC | 967 | 0.59 | 0.49 | 59.15 | 26.78 | 11.79 | 2.28 | | | -0.15 | -0.19 | -0.21 |
| 100000360190 | 5 | MC | 967 | 0.23 | 0.42 | 23.47 | 22.13 | 52.64 | 1.76 | | | 0.01 | -0.06 | -0.20 |
| 100000213637 | 6 | MC | 967 | 0.69 | 0.46 | 10.24 | 11.38 | 68.98 | 9.41 | | -0.16 | -0.27 | | -0.16 |
| 100000213638 | 7 | MC | 967 | 0.69 | 0.46 | 11.58 | 8.17 | 68.87 | 11.38 | | -0.27 | -0.16 | | -0.14 |
| 100000213639 | 8 | MC | 967 | 0.36 | 0.48 | 29.47 | 16.96 | 17.48 | 36.09 | | 0.14 | -0.16 | -0.16 | |
| 100000360191 | 9 | MC | 967 | 0.38 | 0.49 | 24.10 | 16.65 | 38.16 | 20.79 | 0.31 | -0.03 | -0.21 | | -0.07 |
| 100000213641 | 10 | MC | 967 | 0.78 | 0.42 | 77.77 | 6.00 | 7.45 | 8.58 | 0.21 | | -0.14 | -0.24 | -0.19 |
| 100000213642 | 11 | MC | 967 | 0.66 | 0.48 | 8.27 | 14.06 | 11.79 | 65.67 | 0.21 | -0.19 | -0.22 | -0.19 | |
| 100000360192 | 12 | MC | 967 | 0.73 | 0.45 | 72.60 | 4.14 | 11.48 | 10.96 | 0.83 | | -0.17 | -0.27 | -0.18 |
| 100000357134 | 13 | MC | 967 | 0.67 | 0.47 | 67.22 | 17.06 | 15.20 | | 0.52 | | -0.21 | -0.20 | |
| 100000357132 | 14 | MC | 967 | 0.45 | 0.50 | 34.23 | 45.40 | 19.96 | | 0.41 | -0.02 | | -0.20 | |
| 100000357133 | 15 | MC | 967 | 0.74 | 0.44 | 10.86 | 14.89 | 73.84 | | 0.41 | -0.24 | -0.25 | | |
| 100000357135 | 16 | MC | 967 | 0.45 | 0.50 | 33.61 | 44.57 | 21.41 | | 0.41 | 0.08 | | -0.22 | |
| 100000357136 | 17 | MC | 967 | 0.57 | 0.50 | 19.34 | 23.78 | 56.77 | | 0.10 | -0.19 | -0.24 | | |
| 100000357137 | 18 | MC | 967 | 0.56 | 0.50 | 56.05 | 23.58 | 20.27 | | 0.10 | | -0.22 | -0.21 | |
| 100000357138 | 19 | MC | 967 | 0.41 | 0.49 | 26.89 | 41.26 | 31.54 | | 0.31 | -0.15 | | -0.15 | |
| 100000357106 | 20 | MC | 967 | 0.52 | 0.50 | 28.85 | 18.51 | 52.02 | | 0.62 | 0.00 | -0.19 | | |
| 100000357104 | 21 | MC | 967 | 0.65 | 0.48 | 13.86 | 64.63 | 20.99 | | 0.52 | -0.23 | | -0.23 | |
| 100000357107 | 22 | MC | 967 | 0.61 | 0.49 | 20.89 | 60.60 | 17.68 | | 0.83 | -0.15 | | -0.26 | |
| 100000357105 | 23 | MC | 967 | 0.63 | 0.48 | 62.87 | 19.03 | 17.58 | | 0.52 | | -0.20 | -0.18 | |
| 100000357108 | 24 | MC | 967 | 0.49 | 0.50 | 49.22 | 31.54 | 18.61 | | 0.62 | | -0.07 | -0.26 | |
| 100000357109 | 25 | MC | 967 | 0.73 | 0.44 | 13.75 | 12.51 | 73.11 | | 0.62 | -0.25 | -0.21 | | |
| 100000357110 | 26 | MC | 967 | 0.59 | 0.49 | 21.82 | 18.92 | 58.74 | | 0.52 | -0.03 | -0.18 | | |
| 100000260487 | 27 | MC | 967 | 0.41 | 0.49 | 41.05 | 33.71 | 24.61 | | 0.62 | | 0.16 | -0.25 | |
| 100000260486 | 28 | MC | 967 | 0.25 | 0.43 | 25.03 | 43.23 | 31.13 | | 0.62 | | -0.15 | -0.08 | |
| 100000260483 | 29 | MC | 967 | 0.57 | 0.50 | 21.92 | 56.88 | 20.48 | | 0.72 | -0.09 | | -0.22 | |
| 100000260481 | 30 | MC | 967 | 0.50 | 0.50 | 50.36 | 13.24 | 35.68 | | 0.72 | | -0.20 | -0.20 | |
| 100000260489 | 31 | MC | 967 | 0.64 | 0.48 | 13.34 | 22.03 | 64.12 | | 0.52 | -0.23 | -0.09 | | |
| 100000260488 | 32 | MC | 967 | 0.26 | 0.44 | 26.16 | 23.89 | 49.43 | | 0.52 | | -0.08 | -0.14 | |
| 100000260492 | 33 | MC | 967 | 0.28 | 0.45 | 27.51 | 25.23 | 46.74 | | 0.52 | | -0.04 | -0.10 | |
| 100000269896 | 34 | MC | 967 | 0.55 | 0.50 | 16.75 | 27.71 | 54.81 | | 0.72 | -0.18 | -0.17 | | |
| 100000271197 | 35 | MC | 967 | 0.38 | 0.49 | 37.75 | 27.20 | 34.33 | | 0.72 | | -0.10 | -0.25 | |
| 100000271198 | 36 | MC | 967 | 0.57 | 0.49 | 13.86 | 57.29 | 28.02 | | 0.83 | -0.22 | | -0.18 | |
| 100000269900 | 37 | MC | 967 | 0.33 | 0.47 | 33.40 | 30.71 | 34.95 | | 0.93 | | -0.12 | -0.02 | |
| 100000269897 | 38 | MC | 967 | 0.46 | 0.50 | 24.61 | 28.54 | 46.12 | | 0.72 | -0.08 | -0.12 | | |
| 100000269898 | 39 | MC | 967 | 0.41 | 0.49 | 20.68 | 40.95 | 37.64 | | 0.72 | -0.21 | | -0.08 | |
| 100000269899 | 40 | MC | 967 | 0.42 | 0.49 | 42.09 | 25.85 | 31.33 | | 0.72 | | -0.22 | -0.05 | |

Table B.2. The 2010 Mod-MSA: Reading Grade 4, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000357097 | 41 | MC | 967 | 0.27 | 0.44 | 56.15 | 26.89 | 16.86 | | 0.10 | 0.06 | | -0.16 | |
| 100000357095 | 42 | MC | 967 | 0.60 | 0.49 | 60.29 | 23.78 | 15.72 | | 0.21 | | -0.15 | -0.19 | |
| 100000357098 | 43 | MC | 967 | 0.33 | 0.47 | 23.58 | 32.78 | 43.43 | | 0.21 | 0.06 | | -0.26 | |
| 100000357099 | 44 | MC | 967 | 0.31 | 0.46 | 30.82 | 12.93 | 55.95 | | 0.31 | | -0.18 | 0.13 | |
| 100000357100 | 45 | MC | 967 | 0.40 | 0.49 | 34.85 | 25.44 | 39.50 | | 0.21 | 0.02 | -0.15 | | |
| 100000357101 | 46 | MC | 967 | 0.59 | 0.49 | 59.26 | 22.96 | 17.68 | | 0.10 | | -0.17 | -0.14 | |
| 100000357096 | 47 | MC | 967 | 0.58 | 0.49 | 14.37 | 27.09 | 58.12 | | 0.41 | -0.19 | -0.21 | | |
| 100000101997 | 48 | MC | 967 | 0.52 | 0.50 | 13.44 | 34.23 | 52.33 | | | -0.19 | -0.09 | | |
| 100000101996 | 49 | MC | 967 | 0.43 | 0.50 | 30.20 | 26.68 | 43.12 | | | -0.18 | -0.19 | | |
| 100000101998 | 50 | MC | 967 | 0.44 | 0.50 | 26.27 | 43.74 | 29.89 | | 0.10 | -0.04 | | -0.18 | |
| 100000101999 | 51 | MC | 967 | 0.42 | 0.49 | 42.19 | 40.02 | 17.79 | | | | -0.17 | -0.09 | |
| 100000102000 | 52 | MC | 967 | 0.52 | 0.50 | 22.75 | 52.12 | 25.13 | | | -0.07 | | -0.19 | |
| 100000102001 | 53 | MC | 967 | 0.49 | 0.50 | 49.02 | 36.92 | 14.06 | | | | -0.13 | -0.24 | |
| 100000200070 | 54 | MC | 967 | 0.51 | 0.50 | 51.40 | 22.44 | 26.16 | | | | -0.18 | -0.17 | |
| 100000102025 | 55 | MC | 967 | 0.30 | 0.46 | 30.20 | 23.58 | 46.12 | | 0.10 | | -0.17 | 0.02 | |
| 100000102026 | 56 | MC | 967 | 0.44 | 0.50 | 18.92 | 37.02 | 43.95 | | 0.10 | -0.23 | 0.03 | | |
| 100000102027 | 57 | MC | 967 | 0.53 | 0.50 | 15.93 | 53.46 | 30.51 | | 0.10 | -0.21 | | -0.22 | |
| 100000102023 | 58 | MC | 967 | 0.47 | 0.50 | 9.93 | 47.26 | 42.61 | | 0.21 | -0.21 | | -0.14 | |
| 100000102024 | 59 | MC | 967 | 0.69 | 0.46 | 16.03 | 14.79 | 69.08 | | 0.10 | -0.20 | -0.06 | | |
| 100000102028 | 60 | MC | 967 | 0.46 | 0.50 | 46.12 | 16.75 | 36.92 | | 0.21 | | -0.28 | -0.13 | |
| 100000102029 | 61 | MC | 967 | 0.58 | 0.49 | 18.61 | 57.70 | 23.58 | | 0.10 | -0.11 | | -0.25 | |
| 100000301035 | 62 | MC | 967 | 0.67 | 0.47 | 67.22 | 20.37 | 12.31 | | 0.10 | | -0.20 | -0.30 | |
| 100000267470 | 63 | MC | 967 | 0.63 | 0.48 | 63.39 | 12.31 | 24.20 | | 0.10 | | -0.22 | -0.26 | |
| 100000267467 | 64 | MC | 967 | 0.85 | 0.36 | 7.14 | 7.96 | 84.80 | | 0.10 | -0.21 | -0.24 | | |
| 100000301034 | 65 | MC | 967 | 0.69 | 0.46 | 17.17 | 13.34 | 69.39 | | 0.10 | -0.19 | -0.21 | | |
| 100000267473 | 66 | MC | 967 | 0.50 | 0.50 | 23.99 | 50.36 | 25.44 | | 0.21 | -0.12 | | -0.11 | |
| 100000267474 | 67 | MC | 967 | 0.53 | 0.50 | 30.82 | 15.62 | 53.26 | | 0.31 | 0.04 | -0.13 | | |
| 100000267472 | 68 | MC | 967 | 0.61 | 0.49 | 23.06 | 12.41 | 61.01 | | 3.52 | -0.09 | -0.15 | | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

## Table B.3. The 2010 Mod-MSA: Reading Grade 5, Distractor Distribution and Distractor-to- Total Correlation Analysis

| tem CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213650 | 1 | MC | 1043 | 0.67 | 0.47 | 66.73 | 3.55 | 11.51 | 18.22 | | | -0.16 | -0.09 | -0.21 |
| 100000213651 | 2 | MC | 1043 | 0.69 | 0.46 | 2.59 | 11.60 | 16.49 | 69.32 | | -0.09 | -0.24 | -0.14 | |
| 100000213652 | 3 | MC | 1043 | 0.68 | 0.47 | 15.34 | 68.36 | 9.30 | 7.00 | | -0.15 | | -0.19 | -0.09 |
| 100000213653 | 4 | MC | 1043 | 0.73 | 0.45 | 7.96 | 12.18 | 7.29 | 72.58 | | -0.10 | -0.21 | -0.27 | |
| 100000360196 | 5 | MC | 1043 | 0.38 | 0.49 | 38.45 | 7.09 | 3.36 | 51.10 | | | -0.20 | -0.17 | 0.00 |
| 100000213655 | 6 | MC | 1043 | 0.73 | 0.44 | 5.85 | 73.06 | 14.38 | 6.71 | | -0.14 | | -0.16 | -0.20 |
| 100000213656 | 7 | MC | 1043 | 0.78 | 0.42 | 77.85 | 2.97 | 9.88 | 9.30 | | | -0.14 | -0.32 | -0.17 |
| 100000213657 | 8 | MC | 1043 | 0.69 | 0.46 | 13.81 | 10.83 | 69.32 | 5.94 | 0.10 | -0.21 | -0.16 | | -0.25 |
| 100000213658 | 9 | MC | 1043 | 0.49 | 0.50 | 25.98 | 48.80 | 14.19 | 10.93 | 0.10 | 0.14 | | -0.18 | -0.11 |
| 100000213659 | 10 | MC | 1043 | 0.65 | 0.48 | 65.20 | 10.35 | 9.01 | 15.34 | 0.10 | | -0.20 | -0.12 | -0.11 |
| 100000360197 | 11 | MC | 1043 | 0.62 | 0.49 | 14.38 | 62.22 | 8.34 | 14.86 | 0.19 | -0.07 | | -0.17 | -0.23 |
| 100000102095 | 12 | MC | 1043 | 0.51 | 0.50 | 22.53 | 26.37 | 51.10 | | | -0.22 | -0.06 | | |
| 100000102093 | 13 | MC | 1043 | 0.73 | 0.44 | 73.35 | 12.37 | 14.29 | | | | -0.24 | -0.19 | |
| 100000102094 | 14 | MC | 1043 | 0.32 | 0.47 | 31.93 | 23.20 | 44.87 | | | | -0.22 | -0.03 | |
| 100000102096 | 15 | MC | 1043 | 0.50 | 0.50 | 10.74 | 50.24 | 39.02 | | | -0.19 | | -0.09 | |
| 100000102097 | 16 | MC | 1043 | 0.37 | 0.48 | 24.93 | 38.26 | 36.82 | | | -0.07 | 0.01 | | |
| 100000102098 | 17 | MC | 1043 | 0.37 | 0.48 | 32.69 | 37.01 | 30.30 | | | -0.04 | | -0.08 | |
| 100000102099 | 18 | MC | 1043 | 0.34 | 0.47 | 34.23 | 21.76 | 43.82 | | 0.19 | | -0.12 | -0.02 | |
| 100000102113 | 19 | MC | 1043 | 0.51 | 0.50 | 51.10 | 32.79 | 16.11 | | | | -0.17 | -0.26 | |
| 100000102111 | 20 | MC | 1043 | 0.70 | 0.46 | 11.70 | 17.93 | 70.18 | | 0.19 | -0.21 | -0.24 | | |
| 100000102114 | 21 | MC | 1043 | 0.51 | 0.50 | 51.20 | 28.95 | 19.85 | | | | -0.09 | -0.27 | |
| 100000102115 | 22 | MC | 1043 | 0.34 | 0.47 | 30.97 | 34.90 | 34.13 | | | -0.11 | -0.14 | | |
| 100000102116 | 23 | MC | 1043 | 0.50 | 0.50 | 29.15 | 50.24 | 20.61 | | | -0.14 | | -0.20 | |
| 100000102112 | 24 | MC | 1043 | 0.72 | 0.45 | 15.63 | 12.66 | 71.72 | | | -0.14 | -0.22 | | |
| 100000102117 | 25 | MC | 1043 | 0.35 | 0.48 | 36.43 | 35.09 | 28.28 | | 0.19 | 0.02 | | -0.07 | |
| 100000102104 | 26 | MC | 1043 | 0.53 | 0.50 | 52.92 | 16.11 | 30.78 | | 0.19 | | -0.20 | -0.08 | |
| 100000102102 | 27 | MC | 1043 | 0.60 | 0.49 | 23.78 | 15.72 | 60.31 | | 0.19 | -0.21 | -0.03 | | |
| 100000102103 | 28 | MC | 1043 | 0.61 | 0.49 | 22.05 | 61.36 | 16.40 | | 0.19 | -0.06 | | -0.22 | |
| 100000102105 | 29 | MC | 1043 | 0.32 | 0.47 | 32.02 | 41.13 | 26.65 | | 0.19 | | 0.01 | -0.10 | |
| 100000102106 | 30 | MC | 1043 | 0.65 | 0.48 | 17.07 | 17.93 | 64.81 | | 0.19 | -0.13 | -0.21 | | |
| 100000102107 | 31 | MC | 1043 | 0.33 | 0.47 | 36.24 | 30.87 | 32.69 | | 0.19 | -0.11 | -0.10 | | |
| 100000102108 | 32 | MC | 1043 | 0.40 | 0.49 | 29.15 | 39.98 | 30.49 | | 0.38 | 0.00 | | -0.21 | |
| 100000102086 | 33 | MC | 1043 | 0.37 | 0.48 | 24.45 | 37.30 | 37.49 | | 0.77 | -0.08 | | 0.02 | |
| 100000102084 | 34 | MC | 1043 | 0.62 | 0.49 | 13.61 | 23.59 | 62.22 | | 0.58 | -0.10 | -0.17 | | |
| 100000102087 | 35 | MC | 1043 | 0.42 | 0.49 | 22.63 | 42.38 | 34.32 | | 0.67 | -0.13 | | -0.01 | |
| 100000102088 | 36 | MC | 1043 | 0.30 | 0.46 | 29.53 | 46.21 | 23.59 | | 0.67 | | 0.04 | -0.11 | |
| 100000102089 | 37 | MC | 1043 | 0.23 | 0.42 | 48.71 | 27.33 | 23.30 | | 0.67 | 0.15 | -0.13 | | |
| 100000102085 | 38 | MC | 1043 | 0.55 | 0.50 | 29.82 | 54.94 | 14.48 | | 0.77 | -0.01 | | -0.19 | |
| 100000102090 | 39 | MC | 1043 | 0.38 | 0.49 | 38.26 | 31.26 | 29.63 | | 0.86 | | -0.11 | -0.13 | |
| 100000102046 | 40 | MC | 1043 | 0.78 | 0.41 | 7.48 | 78.43 | 13.90 | | 0.19 | -0.27 | | -0.27 | |

Table B.3. The 2010 Mod-MSA: Reading Grade 5, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000102048 | 41 | MC | 1043 | 0.55 | 0.50 | 54.94 | 27.80 | 17.07 | | 0.19 | | -0.07 | -0.15 | |
| 100000102049 | 42 | MC | 1043 | 0.33 | 0.47 | 39.21 | 27.71 | 32.79 | | 0.29 | -0.03 | -0.04 | | |
| 100000102050 | 43 | MC | 1043 | 0.70 | 0.46 | 18.22 | 11.89 | 69.51 | | 0.38 | -0.18 | -0.14 | | |
| 100000102051 | 44 | MC | 1043 | 0.47 | 0.50 | 46.98 | 26.08 | 26.65 | | 0.29 | | -0.11 | -0.08 | |
| 100000102052 | 45 | MC | 1043 | 0.63 | 0.48 | 18.02 | 63.18 | 18.70 | | 0.10 | -0.14 | | -0.21 | |
| 100000102047 | 46 | MC | 1043 | 0.61 | 0.49 | 23.68 | 61.17 | 15.05 | | 0.10 | -0.21 | | -0.23 | |
| 100000102056 | 47 | MC | 1043 | 0.47 | 0.50 | 41.99 | 46.79 | 11.12 | | 0.10 | -0.08 | | -0.14 | |
| 100000102058 | 48 | MC | 1043 | 0.18 | 0.39 | 16.78 | 64.72 | 18.31 | | 0.19 | -0.20 | 0.19 | | |
| 100000102059 | 49 | MC | 1043 | 0.60 | 0.49 | 59.83 | 23.11 | 16.87 | | 0.19 | | -0.21 | -0.10 | |
| 100000102060 | 50 | MC | 1043 | 0.36 | 0.48 | 38.73 | 24.64 | 36.24 | | 0.38 | 0.01 | -0.24 | | |
| 100000102061 | 51 | MC | 1043 | 0.50 | 0.50 | 50.05 | 34.90 | 14.77 | | 0.29 | | -0.09 | -0.21 | |
| 100000102057 | 52 | MC | 1043 | 0.39 | 0.49 | 21.96 | 38.73 | 39.12 | | 0.19 | -0.14 | | -0.01 | |
| 100000102062 | 53 | MC | 1043 | 0.27 | 0.44 | 30.11 | 27.13 | 42.28 | | 0.48 | 0.00 | | -0.04 | |
| 100000102068 | 54 | MC | 1043 | 0.58 | 0.49 | 23.01 | 57.53 | 19.18 | | 0.29 | -0.23 | | -0.11 | |
| 100000102069 | 55 | MC | 1043 | 0.37 | 0.48 | 36.82 | 26.08 | 36.72 | | 0.38 | | -0.24 | -0.10 | |
| 100000102066 | 56 | MC | 1043 | 0.54 | 0.50 | 22.63 | 22.91 | 54.07 | | 0.38 | -0.05 | -0.16 | | |
| 100000102067 | 57 | MC | 1043 | 0.67 | 0.47 | 13.04 | 67.40 | 19.37 | | 0.19 | -0.27 | | -0.16 | |
| 100000102070 | 58 | MC | 1043 | 0.24 | 0.43 | 23.97 | 33.84 | 41.90 | | 0.29 | | 0.06 | 0.01 | |
| 100000102071 | 59 | MC | 1043 | 0.47 | 0.50 | 21.38 | 31.45 | 46.79 | | 0.38 | -0.12 | -0.03 | | |
| 100000102072 | 60 | MC | 1043 | 0.62 | 0.49 | 14.38 | 23.30 | 62.13 | | 0.19 | -0.15 | -0.29 | | |
| 100000267477 | 61 | MC | 1043 | 0.47 | 0.50 | 47.17 | 31.26 | 17.74 | | 3.84 | | -0.04 | -0.10 | |
| 100000268378 | 62 | MC | 1043 | 0.30 | 0.46 | 30.20 | 13.81 | 52.06 | | 3.93 | | -0.12 | -0.01 | |
| 100000267485 | 63 | MC | 1043 | 0.38 | 0.48 | 10.07 | 37.68 | 48.32 | | 3.93 | -0.12 | | -0.07 | |
| 100000268380 | 64 | MC | 1043 | 0.43 | 0.50 | 27.52 | 43.43 | 25.12 | | 3.93 | 0.03 | | -0.19 | |
| 100000267481 | 65 | MC | 1043 | 0.29 | 0.45 | 28.57 | 29.15 | 38.45 | | 3.84 | | 0.03 | -0.03 | |
| 100000301339 | 66 | MC | 1043 | 0.47 | 0.50 | 18.22 | 47.17 | 30.68 | | 3.93 | -0.01 | | -0.07 | |
| 100000303033 | 67 | MC | 1043 | 0.59 | 0.49 | 20.13 | 16.87 | 59.16 | | 3.84 | -0.13 | -0.17 | | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2.

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

## Table B.4. The 2010 Mod-MSA: Reading Grade 6, Distractor Distribution and Distractor-to- Total Correlation Analysis

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213662 | 1 | MC | 975 | 0.69 | 0.46 | 10.36 | 14.26 | 6.77 | 68.51 | 0.10 | -0.30 | -0.17 | -0.17 | |
| 100000213663 | 2 | MC | 975 | 0.76 | 0.43 | 7.69 | 6.26 | 76.10 | 9.85 | 0.10 | -0.22 | -0.13 | | -0.36 |
| 100000213664 | 3 | MC | 975 | 0.54 | 0.50 | 19.49 | 53.64 | 18.56 | 8.31 | | -0.22 | | 0.00 | -0.18 |
| 100000213665 | 4 | MC | 975 | 0.65 | 0.48 | 11.79 | 10.15 | 64.92 | 13.03 | 0.10 | -0.17 | -0.25 | | -0.18 |
| 100000213666 | 5 | MC | 975 | 0.57 | 0.50 | 10.15 | 56.72 | 13.33 | 19.79 | | -0.07 | | -0.05 | -0.12 |
| 100000360201 | 6 | MC | 975 | 0.19 | 0.39 | 19.08 | 43.28 | 18.26 | 19.38 | | | 0.08 | -0.27 | 0.04 |
| 100000213668 | 7 | MC | 975 | 0.68 | 0.47 | 18.46 | 6.87 | 68.21 | 6.36 | 0.10 | -0.22 | -0.09 | | -0.25 |
| 100000213669 | 8 | MC | 975 | 0.25 | 0.43 | 23.90 | 19.38 | 24.51 | 32.10 | 0.10 | -0.01 | -0.21 | | 0.02 |
| 100000213670 | 9 | MC | 975 | 0.77 | 0.42 | 8.51 | 6.05 | 77.33 | 8.10 | | -0.18 | -0.25 | | -0.13 |
| 100000213671 | 10 | MC | 975 | 0.52 | 0.50 | 23.90 | 52.10 | 12.21 | 11.79 | | -0.08 | | -0.20 | -0.13 |
| 100000360202 | 11 | MC | 975 | 0.69 | 0.46 | 4.82 | 16.92 | 9.33 | 68.92 | | -0.12 | -0.19 | -0.24 | |
| 100000102192 | 12 | MC | 975 | 0.60 | 0.49 | 21.33 | 60.41 | 18.05 | | 0.21 | -0.36 | | -0.15 | |
| 100000102194 | 13 | MC | 975 | 0.54 | 0.50 | 53.64 | 26.15 | 20.21 | | | | -0.04 | -0.22 | |
| 100000102191 | 14 | MC | 975 | 0.68 | 0.47 | 18.15 | 14.26 | 67.59 | | | -0.27 | -0.23 | | |
| 100000102195 | 15 | MC | 975 | 0.61 | 0.49 | 7.18 | 61.03 | 31.69 | | 0.10 | -0.25 | | -0.13 | |
| 100000102196 | 16 | MC | 975 | 0.44 | 0.50 | 44.41 | 22.26 | 33.03 | | 0.31 | | -0.13 | -0.13 | |
| 100000102174 | 17 | MC | 975 | 0.51 | 0.50 | 15.59 | 33.23 | 51.18 | | | -0.23 | 0.03 | | |
| 100000102175 | 18 | MC | 975 | 0.33 | 0.47 | 53.54 | 33.44 | 13.03 | | | -0.05 | | -0.14 | |
| 100000102173 | 19 | MC | 975 | 0.64 | 0.48 | 17.03 | 64.10 | 18.87 | | | -0.19 | | -0.22 | |
| 100000102176 | 20 | MC | 975 | 0.43 | 0.50 | 43.49 | 27.18 | 29.23 | | 0.10 | | -0.10 | -0.19 | |
| 100000102178 | 21 | MC | 975 | 0.51 | 0.50 | 19.79 | 29.54 | 50.56 | | 0.10 | -0.02 | -0.19 | | |
| 100000269999 | 22 | MC | 975 | 0.86 | 0.34 | 8.51 | 86.26 | 4.92 | | 0.31 | -0.21 | | -0.32 | |
| 100000269997 | 23 | MC | 975 | 0.56 | 0.50 | 26.67 | 56.21 | 16.82 | | 0.31 | -0.12 | | -0.19 | |
| 100000270006 | 24 | MC | 975 | 0.71 | 0.45 | 71.08 | 10.15 | 18.36 | | 0.41 | | -0.25 | -0.27 | |
| 100000269998 | 25 | MC | 975 | 0.44 | 0.50 | 12.41 | 43.79 | 43.59 | | 0.21 | -0.17 | | -0.05 | |
| 100000270000 | 26 | MC | 975 | 0.69 | 0.46 | 17.03 | 68.51 | 14.15 | | 0.31 | -0.30 | | -0.13 | |
| 100000270005 | 27 | MC | 975 | 0.66 | 0.47 | 65.95 | 18.36 | 15.49 | | 0.21 | | -0.21 | -0.27 | |
| 100000270007 | 28 | MC | 975 | 0.45 | 0.50 | 44.82 | 20.10 | 34.67 | | 0.41 | | -0.21 | -0.08 | |
| 100000270016 | 29 | MC | 975 | 0.30 | 0.46 | 36.72 | 30.26 | 32.72 | | 0.31 | -0.02 | | -0.04 | |
| 100000270018 | 30 | MC | 975 | 0.45 | 0.50 | 25.95 | 28.41 | 45.23 | | 0.41 | -0.10 | -0.03 | | |
| 100000270021 | 31 | MC | 975 | 0.39 | 0.49 | 38.67 | 37.95 | 23.08 | | 0.31 | | -0.01 | -0.17 | |
| 100000270022 | 32 | MC | 975 | 0.47 | 0.50 | 23.69 | 29.13 | 46.77 | | 0.41 | -0.02 | 0.00 | | |
| 100000270023 | 33 | MC | 975 | 0.41 | 0.49 | 24.21 | 40.62 | 34.87 | | 0.31 | -0.22 | | 0.13 | |
| 100000270013 | 34 | MC | 975 | 0.41 | 0.49 | 20.31 | 40.51 | 38.87 | | 0.31 | -0.25 | | -0.02 | |
| 100000270020 | 35 | MC | 975 | 0.44 | 0.50 | 43.69 | 32.41 | 23.59 | | 0.31 | | -0.19 | -0.12 | |
| 100000257087 | 36 | MC | 975 | 0.75 | 0.43 | 8.72 | 75.38 | 15.69 | | 0.21 | -0.21 | | -0.22 | |
| 100000257088 | 37 | MC | 975 | 0.58 | 0.49 | 13.74 | 58.46 | 27.49 | | 0.31 | -0.42 | | -0.11 | |
| 100000257113 | 38 | MC | 975 | 0.36 | 0.48 | 18.05 | 45.33 | 36.41 | | 0.21 | -0.04 | -0.02 | | |
| 100000257114 | 39 | MC | 975 | 0.42 | 0.49 | 20.72 | 36.82 | 41.95 | | 0.51 | -0.24 | 0.02 | | |
| 100000257118 | 40 | MC | 975 | 0.67 | 0.47 | 66.77 | 19.49 | 13.44 | | 0.31 | | -0.04 | -0.37 | |

Table B.4. The 2010 Mod-MSA: Reading Grade 6, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000257119 | 41 | MC | 975 | 0.62 | 0.49 | 61.74 | 13.85 | 24.10 | | 0.31 | | -0.35 | -0.12 | |
| 100000257120 | 42 | MC | 975 | 0.42 | 0.49 | 42.15 | 26.36 | 31.28 | | 0.21 | | -0.16 | -0.03 | |
| 100000257102 | 43 | MC | 975 | 0.52 | 0.50 | 20.62 | 27.28 | 52.00 | | 0.10 | -0.33 | -0.10 | | |
| 100000257072 | 44 | MC | 975 | 0.34 | 0.47 | 33.54 | 29.64 | 36.72 | | 0.10 | | -0.05 | -0.03 | |
| 100000257103 | 45 | MC | 975 | 0.44 | 0.50 | 14.05 | 42.15 | 43.69 | | 0.10 | -0.21 | 0.00 | | |
| 100000257104 | 46 | MC | 975 | 0.50 | 0.50 | 49.95 | 29.85 | 19.69 | | 0.51 | | -0.08 | -0.16 | |
| 100000257071 | 47 | MC | 975 | 0.59 | 0.49 | 21.03 | 58.97 | 19.90 | | 0.10 | -0.06 | | -0.29 | |
| 100000257100 | 48 | MC | 975 | 0.55 | 0.50 | 55.49 | 11.28 | 33.13 | | 0.10 | | -0.22 | -0.21 | |
| 100000257101 | 49 | MC | 975 | 0.30 | 0.46 | 30.26 | 34.77 | 34.87 | | 0.10 | | -0.13 | 0.01 | |
| 100000102136 | 50 | MC | 975 | 0.70 | 0.46 | 19.38 | 69.64 | 10.87 | | 0.10 | -0.17 | | -0.21 | |
| 100000102137 | 51 | MC | 975 | 0.54 | 0.50 | 26.87 | 18.56 | 54.26 | | 0.31 | -0.05 | -0.16 | | |
| 100000102139 | 52 | MC | 975 | 0.60 | 0.49 | 60.10 | 18.15 | 21.54 | | 0.21 | | -0.02 | -0.07 | |
| 100000102140 | 53 | MC | 975 | 0.55 | 0.50 | 22.46 | 22.26 | 55.18 | | 0.10 | -0.02 | -0.17 | | |
| 100000257081 | 54 | MC | 975 | 0.74 | 0.44 | 73.85 | 12.72 | 13.33 | | 0.10 | | -0.27 | -0.18 | |
| 100000257110 | 55 | MC | 975 | 0.60 | 0.49 | 18.67 | 20.82 | 60.41 | | 0.10 | -0.22 | -0.20 | | |
| 100000257083 | 56 | MC | 975 | 0.46 | 0.50 | 8.41 | 46.05 | 39.59 | | 5.95 | -0.15 | | -0.10 | |
| 100000257106 | 57 | MC | 975 | 0.59 | 0.49 | 59.18 | 21.64 | 13.33 | | 5.85 | | -0.16 | -0.10 | |
| 100000257107 | 58 | MC | 975 | 0.39 | 0.49 | 39.08 | 23.90 | 31.18 | | 5.85 | | -0.05 | -0.03 | |
| 100000257109 | 59 | MC | 975 | 0.56 | 0.50 | 56.10 | 15.79 | 22.26 | | 5.85 | | -0.20 | -0.11 | |
| 100000257112 | 60 | MC | 975 | 0.83 | 0.38 | 5.74 | 82.56 | 5.85 | | 5.85 | -0.15 | | -0.21 | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

## Table B.5. The 2010 Mod-MSA: Reading Grade 7, Distractor Distribution and Distractor-to- Total Correlation Analysis

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213674 | 1 | MC | 1158 | 0.67 | 0.47 | 67.44 | 7.17 | 15.72 | 9.50 | 0.17 | | -0.13 | -0.15 | -0.36 |
| 100000213675 | 2 | MC | 1158 | 0.78 | 0.41 | 7.51 | 6.82 | 7.17 | 78.15 | 0.35 | -0.36 | -0.06 | -0.10 | |
| 100000213676 | 3 | MC | 1158 | 0.35 | 0.48 | 34.97 | 10.62 | 28.58 | 25.22 | 0.60 | | -0.25 | 0.01 | -0.08 |
| 100000360204 | 4 | MC | 1158 | 0.29 | 0.46 | 21.50 | 35.41 | 29.45 | 13.21 | 0.43 | -0.14 | 0.08 | | -0.16 |
| 100000213678 | 5 | MC | 1158 | 0.74 | 0.44 | 11.57 | 74.01 | 5.70 | 8.20 | 0.52 | -0.16 | | -0.13 | -0.23 |
| 100000213679 | 6 | MC | 1158 | 0.68 | 0.46 | 17.36 | 68.48 | 4.66 | 8.81 | 0.69 | -0.25 | | -0.18 | -0.18 |
| 100000360205 | 7 | MC | 1158 | 0.35 | 0.48 | 35.15 | 11.31 | 33.42 | 19.17 | 0.95 | | -0.12 | -0.09 | -0.19 |
| 100000270578 | 8 | MC | 1158 | 0.62 | 0.48 | 30.31 | 62.26 | 7.08 | | 0.35 | -0.20 | | -0.14 | |
| 100000270572 | 9 | MC | 1158 | 0.52 | 0.50 | 25.65 | 22.02 | 51.90 | | 0.43 | 0.03 | -0.17 | | |
| 100000270574 | 10 | MC | 1158 | 0.42 | 0.49 | 42.23 | 27.20 | 30.22 | | 0.35 | | -0.23 | -0.07 | |
| 100000322011 | 11 | MC | 1158 | 0.62 | 0.48 | 14.08 | 62.35 | 23.14 | | 0.43 | -0.20 | | -0.16 | |
| 100000270580 | 12 | MC | 1158 | 0.81 | 0.39 | 81.26 | 8.89 | 9.50 | | 0.35 | | -0.32 | -0.21 | |
| 100000270571 | 13 | MC | 1158 | 0.38 | 0.49 | 22.02 | 38.00 | 39.55 | | 0.43 | -0.02 | | -0.14 | |
| 100000270575 | 14 | MC | 1158 | 0.70 | 0.46 | 70.47 | 11.92 | 17.10 | | 0.52 | | -0.26 | -0.28 | |
| 100000270109 | 15 | MC | 1158 | 0.42 | 0.49 | 41.71 | 14.85 | 42.75 | | 0.69 | | -0.13 | -0.14 | |
| 100000270112 | 16 | MC | 1158 | 0.39 | 0.49 | 38.77 | 22.02 | 38.34 | | 0.86 | | -0.21 | 0.08 | |
| 100000270113 | 17 | MC | 1158 | 0.52 | 0.50 | 26.86 | 20.29 | 52.25 | | 0.60 | 0.00 | -0.26 | | |
| 100000270114 | 18 | MC | 1158 | 0.56 | 0.50 | 22.63 | 56.39 | 20.38 | | 0.60 | -0.11 | | -0.17 | |
| 100000270111 | 19 | MC | 1158 | 0.41 | 0.49 | 35.92 | 22.63 | 40.67 | | 0.78 | 0.08 | -0.06 | | |
| 100000270107 | 20 | MC | 1158 | 0.34 | 0.47 | 33.77 | 31.26 | 34.28 | | 0.69 | | -0.21 | 0.16 | |
| 100000270121 | 21 | MC | 1158 | 0.56 | 0.50 | 13.90 | 29.53 | 55.79 | | 0.78 | -0.11 | -0.23 | | |
| 100000102276 | 22 | MC | 1158 | 0.48 | 0.50 | 21.42 | 29.45 | 48.36 | | 0.78 | -0.06 | -0.19 | | |
| 100000102277 | 23 | MC | 1158 | 0.34 | 0.47 | 34.28 | 39.72 | 25.22 | | 0.78 | | -0.09 | -0.09 | |
| 100000102274 | 24 | MC | 1158 | 0.41 | 0.49 | 23.06 | 35.49 | 40.76 | | 0.69 | 0.03 | -0.13 | | |
| 100000102278 | 25 | MC | 1158 | 0.46 | 0.50 | 28.07 | 46.46 | 24.78 | | 0.69 | 0.00 | | -0.33 | |
| 100000102279 | 26 | MC | 1158 | 0.41 | 0.49 | 41.02 | 39.29 | 18.91 | | 0.78 | | -0.07 | -0.19 | |
| 100000102263 | 27 | MC | 1158 | 0.58 | 0.49 | 17.18 | 57.77 | 24.27 | | 0.78 | -0.25 | | -0.05 | |
| 100000102267 | 28 | MC | 1158 | 0.43 | 0.49 | 24.87 | 42.75 | 31.69 | | 0.69 | -0.10 | | 0.03 | |
| 100000102268 | 29 | MC | 1158 | 0.50 | 0.50 | 50.43 | 13.73 | 35.15 | | 0.69 | | -0.21 | -0.20 | |
| 100000102269 | 30 | MC | 1158 | 0.42 | 0.49 | 41.97 | 20.64 | 36.53 | | 0.86 | | -0.21 | -0.02 | |
| 100000102266 | 31 | MC | 1158 | 0.33 | 0.47 | 32.73 | 44.30 | 22.02 | | 0.95 | | -0.01 | -0.13 | |
| 100000257181 | 32 | MC | 1158 | 0.65 | 0.48 | 12.87 | 21.85 | 64.94 | | 0.35 | -0.05 | -0.11 | | |
| 100000257182 | 33 | MC | 1158 | 0.79 | 0.41 | 78.93 | 9.76 | 11.05 | | 0.26 | | -0.29 | -0.14 | |
| 100000257216 | 34 | MC | 1158 | 0.69 | 0.46 | 68.57 | 16.41 | 14.68 | | 0.35 | | -0.20 | -0.23 | |
| 100000257218 | 35 | MC | 1158 | 0.76 | 0.43 | 14.68 | 76.08 | 8.81 | | 0.43 | -0.29 | | -0.20 | |
| 100000257219 | 36 | MC | 1158 | 0.47 | 0.50 | 8.55 | 43.96 | 47.06 | | 0.43 | -0.35 | 0.08 | | |
| 100000257215 | 37 | MC | 1158 | 0.87 | 0.33 | 87.13 | 8.46 | 4.06 | | 0.35 | | -0.16 | -0.19 | |
| 100000257221 | 38 | MC | 1158 | 0.53 | 0.50 | 12.18 | 53.11 | 34.37 | | 0.35 | -0.12 | | -0.11 | |
| 100000102244 | 39 | MC | 1158 | 0.84 | 0.37 | 83.59 | 9.07 | 6.99 | | 0.35 | | -0.33 | -0.19 | |
| 100000102246 | 40 | MC | 1158 | 0.58 | 0.49 | 27.12 | 14.16 | 58.38 | | 0.35 | -0.07 | -0.16 | | |

Table B.5. The 2010 Mod-MSA: Reading Grade 7, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000102247 | 41 | MC | 1158 | 0.75 | 0.43 | 13.64 | 74.87 | 10.97 | | 0.52 | -0.34 | | -0.26 | |
| 100000102248 | 42 | MC | 1158 | 0.51 | 0.50 | 25.73 | 23.06 | 50.60 | | 0.60 | -0.15 | 0.00 | | |
| 100000102245 | 43 | MC | 1158 | 0.80 | 0.40 | 79.88 | 10.02 | 9.33 | | 0.78 | | -0.33 | -0.09 | |
| 100000102218 | 44 | MC | 1158 | 0.50 | 0.50 | 20.98 | 28.07 | 50.09 | | 0.86 | -0.23 | -0.16 | | |
| 100000102220 | 45 | MC | 1158 | 0.58 | 0.49 | 58.12 | 22.54 | 18.57 | | 0.78 | | -0.20 | -0.24 | |
| 100000102221 | 46 | MC | 1158 | 0.47 | 0.50 | 47.32 | 27.81 | 24.09 | | 0.78 | | -0.13 | -0.26 | |
| 100000102222 | 47 | MC | 1158 | 0.61 | 0.49 | 13.21 | 60.62 | 25.65 | | 0.52 | -0.24 | | -0.12 | |
| 100000102217 | 48 | MC | 1158 | 0.63 | 0.48 | 10.54 | 25.91 | 63.04 | | 0.52 | -0.30 | -0.03 | | |
| 100000257202 | 49 | MC | 1158 | 0.84 | 0.37 | 6.74 | 83.85 | 8.98 | | 0.43 | -0.21 | | -0.17 | |
| 100000257769 | 50 | MC | 1158 | 0.53 | 0.50 | 29.02 | 53.20 | 17.27 | | 0.52 | 0.02 | | -0.17 | |
| 100000257770 | 51 | MC | 1158 | 0.40 | 0.49 | 27.12 | 40.07 | 32.47 | | 0.35 | 0.07 | | -0.20 | |
| 100000257773 | 52 | MC | 1158 | 0.73 | 0.45 | 72.80 | 13.56 | 13.30 | | 0.35 | | -0.18 | -0.29 | |
| 100000322795 | 53 | MC | 1158 | 0.52 | 0.50 | 24.53 | 22.80 | 52.07 | | 0.60 | -0.13 | -0.10 | | |
| 100000257767 | 54 | MC | 1158 | 0.66 | 0.48 | 65.63 | 18.05 | 15.98 | | 0.35 | | -0.14 | -0.17 | |
| 100000257772 | 55 | MC | 1158 | 0.51 | 0.50 | 20.21 | 28.76 | 50.52 | | 0.52 | -0.27 | -0.13 | | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2.

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

## Table B.6. The 2010 Mod-MSA: Reading Grade 8, Distractor Distribution and Distractor-to- Total Correlation Analysis

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000213683 | 1 | MC | 1268 | 0.83 | 0.38 | 5.76 | 2.05 | 9.07 | 82.97 | 0.16 | -0.06 | -0.14 | -0.18 | |
| 100000213684 | 2 | MC | 1268 | 0.67 | 0.47 | 67.27 | 12.78 | 8.99 | 10.88 | 0.08 | | -0.12 | -0.12 | -0.09 |
| 100000360208 | 3 | MC | 1268 | 0.43 | 0.50 | 5.60 | 42.98 | 11.99 | 39.35 | 0.08 | -0.05 | | -0.18 | 0.00 |
| 100000213686 | 4 | MC | 1268 | 0.50 | 0.50 | 50.39 | 8.75 | 20.98 | 19.72 | 0.16 | | -0.21 | -0.09 | -0.08 |
| 100000213687 | 5 | MC | 1268 | 0.60 | 0.49 | 11.67 | 60.25 | 11.20 | 16.72 | 0.16 | -0.17 | | -0.19 | 0.00 |
| 100000213688 | 6 | MC | 1268 | 0.63 | 0.48 | 62.62 | 11.28 | 19.16 | 6.86 | 0.08 | | -0.16 | -0.16 | -0.14 |
| 100000360209 | 7 | MC | 1268 | 0.53 | 0.50 | 10.17 | 23.82 | 12.93 | 52.60 | 0.47 | -0.04 | -0.21 | -0.08 | |
| 100000270136 | 8 | MC | 1268 | 0.44 | 0.50 | 36.44 | 44.32 | 18.77 | | 0.47 | -0.05 | | -0.07 | |
| 100000270132 | 9 | MC | 1268 | 0.80 | 0.40 | 5.99 | 13.72 | 79.97 | | 0.32 | -0.16 | -0.18 | | |
| 100000273591 | 10 | MC | 1268 | 0.55 | 0.50 | 23.26 | 55.13 | 21.29 | | 0.32 | -0.12 | | -0.21 | |
| 100000270131 | 11 | MC | 1268 | 0.38 | 0.49 | 35.09 | 38.49 | 26.10 | | 0.32 | -0.11 | | -0.11 | |
| 100000270135 | 12 | MC | 1268 | 0.52 | 0.50 | 30.91 | 52.44 | 16.32 | | 0.32 | -0.12 | | -0.15 | |
| 100000270137 | 13 | MC | 1268 | 0.57 | 0.49 | 20.58 | 21.77 | 57.41 | | 0.24 | -0.10 | -0.10 | | |
| 100000304610 | 14 | MC | 1268 | 0.37 | 0.48 | 30.05 | 32.81 | 36.83 | | 0.32 | -0.05 | -0.11 | | |
| 100000102331 | 15 | MC | 1268 | 0.86 | 0.35 | 5.76 | 8.04 | 85.73 | | 0.47 | -0.16 | -0.19 | | |
| 100000102329 | 16 | MC | 1268 | 0.56 | 0.50 | 29.10 | 56.07 | 14.51 | | 0.32 | 0.05 | | -0.21 | |
| 100000102330 | 17 | MC | 1268 | 0.60 | 0.49 | 21.45 | 59.62 | 18.69 | | 0.24 | -0.14 | | -0.22 | |
| 100000102334 | 18 | MC | 1268 | 0.47 | 0.50 | 31.70 | 21.45 | 46.61 | | 0.24 | 0.10 | -0.17 | | |
| 100000102335 | 19 | MC | 1268 | 0.85 | 0.36 | 3.94 | 10.80 | 84.62 | | 0.63 | -0.22 | -0.18 | | |
| 100000270154 | 20 | MC | 1268 | 0.40 | 0.49 | 28.47 | 40.06 | 31.31 | | 0.16 | 0.13 | | -0.14 | |
| 100000270155 | 21 | MC | 1268 | 0.43 | 0.50 | 12.15 | 43.30 | 44.24 | | 0.32 | -0.20 | | 0.16 | |
| 100000270158 | 22 | MC | 1268 | 0.50 | 0.50 | 10.25 | 39.43 | 49.92 | | 0.39 | -0.21 | -0.09 | | |
| 100000273597 | 23 | MC | 1268 | 0.70 | 0.46 | 10.17 | 70.11 | 19.40 | | 0.32 | -0.27 | | -0.14 | |
| 100000270163 | 24 | MC | 1268 | 0.50 | 0.50 | 29.73 | 50.00 | 19.79 | | 0.47 | -0.15 | | -0.11 | |
| 100000270159 | 25 | MC | 1268 | 0.77 | 0.42 | 7.10 | 15.69 | 76.97 | | 0.24 | -0.13 | -0.31 | | |
| 100000273595 | 26 | MC | 1268 | 0.40 | 0.49 | 40.22 | 41.09 | 18.14 | | 0.55 | | 0.03 | -0.10 | |
| 100000102341 | 27 | MC | 1268 | 0.67 | 0.47 | 66.88 | 18.45 | 14.35 | | 0.32 | | -0.22 | -0.20 | |
| 100000102343 | 28 | MC | 1268 | 0.57 | 0.50 | 25.47 | 56.55 | 17.67 | | 0.32 | -0.07 | | -0.20 | |
| 100000102344 | 29 | MC | 1268 | 0.70 | 0.46 | 14.12 | 15.46 | 70.11 | | 0.32 | -0.21 | -0.13 | | |
| 100000102346 | 30 | MC | 1268 | 0.55 | 0.50 | 19.95 | 24.84 | 54.97 | | 0.24 | -0.17 | -0.25 | | |
| 100000102345 | 31 | MC | 1268 | 0.63 | 0.48 | 11.04 | 63.17 | 25.55 | | 0.24 | -0.25 | | -0.21 | |
| 100000257161 | 32 | MC | 1268 | 0.53 | 0.50 | 24.05 | 53.39 | 22.40 | | 0.16 | -0.03 | | -0.15 | |
| 100000302728 | 33 | MC | 1268 | 0.42 | 0.49 | 42.03 | 41.88 | 15.93 | | 0.16 | | 0.00 | -0.13 | |
| 100000257787 | 34 | MC | 1268 | 0.37 | 0.48 | 29.42 | 37.46 | 32.97 | | 0.16 | 0.07 | | -0.23 | |
| 100000257788 | 35 | MC | 1268 | 0.73 | 0.44 | 73.26 | 8.52 | 18.14 | | 0.08 | | -0.22 | -0.21 | |
| 100000257158 | 36 | MC | 1268 | 0.56 | 0.50 | 55.91 | 26.81 | 17.11 | | 0.16 | | -0.21 | -0.20 | |
| 100000257160 | 37 | MC | 1268 | 0.60 | 0.49 | 59.70 | 18.38 | 21.77 | | 0.16 | | -0.19 | -0.19 | |
| 100000257163 | 38 | MC | 1268 | 0.74 | 0.44 | 13.96 | 11.51 | 74.37 | | 0.16 | -0.16 | -0.26 | | |
| 100000102320 | 39 | MC | 1268 | 0.78 | 0.41 | 5.21 | 78.47 | 16.17 | | 0.16 | -0.18 | | -0.12 | |
| 100000102322 | 40 | MC | 1268 | 0.65 | 0.48 | 25.63 | 9.07 | 65.14 | | 0.16 | 0.01 | -0.23 | | |

Table B.6. The 2010 Mod-MSA: Reading Grade 8, Distractor Distribution and Distractor-to-Total Correlation Analysis (Continued)

| Item CID | Item Seq No. | Item Type | n-Count | Mean | SD | Distractor Distribution | | | | | Distractor/Total Correlation[1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 | 4 | Omit | 1 | 2 | 3 | 4 |
| 100000102323 | 41 | MC | 1268 | 0.51 | 0.50 | 50.63 | 20.58 | 28.71 | | 0.08 | | -0.12 | -0.04 | |
| 100000102324 | 42 | MC | 1268 | 0.62 | 0.49 | 61.59 | 27.92 | 10.25 | | 0.24 | | -0.24 | -0.26 | |
| 100000102321 | 43 | MC | 1268 | 0.76 | 0.42 | 8.83 | 76.50 | 14.51 | | 0.16 | -0.24 | | -0.17 | |
| 100000257139 | 44 | MC | 1268 | 0.31 | 0.46 | 42.98 | 31.07 | 25.79 | | 0.16 | -0.07 | | 0.03 | |
| 100000257140 | 45 | MC | 1268 | 0.71 | 0.45 | 8.83 | 19.64 | 71.37 | | 0.16 | -0.09 | -0.06 | | |
| 100000257141 | 46 | MC | 1268 | 0.48 | 0.50 | 41.88 | 10.25 | 47.71 | | 0.16 | -0.10 | -0.08 | | |
| 100000257144 | 47 | MC | 1268 | 0.58 | 0.49 | 32.02 | 58.12 | 9.62 | | 0.24 | -0.17 | | -0.19 | |
| 100000257777 | 48 | MC | 1268 | 0.46 | 0.50 | 32.49 | 21.14 | 46.21 | | 0.16 | -0.03 | -0.21 | | |
| 100000302674 | 49 | MC | 1268 | 0.56 | 0.50 | 21.69 | 56.07 | 22.16 | | 0.08 | -0.10 | | -0.15 | |
| 100000302675 | 50 | MC | 1268 | 0.70 | 0.46 | 15.54 | 69.64 | 14.67 | | 0.16 | -0.22 | | -0.23 | |
| 100000102304 | 51 | MC | 1268 | 0.57 | 0.50 | 7.89 | 34.94 | 57.02 | | 0.16 | -0.18 | -0.08 | | |
| 100000102305 | 52 | MC | 1268 | 0.66 | 0.47 | 9.94 | 23.90 | 66.01 | | 0.16 | -0.16 | -0.22 | | |
| 100000102302 | 53 | MC | 1268 | 0.53 | 0.50 | 53.00 | 16.56 | 30.36 | | 0.08 | | -0.12 | -0.25 | |
| 100000102307 | 54 | MC | 1268 | 0.68 | 0.47 | 68.14 | 16.40 | 15.30 | | 0.16 | | -0.25 | -0.15 | |
| 100000102308 | 55 | MC | 1268 | 0.66 | 0.47 | 65.77 | 15.54 | 18.30 | | 0.39 | | -0.21 | -0.27 | |

Note: 1. Empty cell indicates the correct answer for the particular item. Point biserial (Item-to-Total Correlation) for the correct answer will be the same as the one shown in item analysis in Section 3.2.

2. Percent distribution by distractors may not add to 100 because of rounding

3. These analyses are based on the equating sample used with the exclusion criteria.

# APPENDIX C: FREQUENCY DISTRIBUTION HISTOGRAMS OF SCALE SCORES

## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 3

## Histogram Bars

| SSTOT | | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|---|
| 13 | * | 1 | 1 | 0.12 | 0.12 |
| 16 | * | 1 | 2 | 0.12 | 0.25 |
| 19 | * | 1 | 3 | 0.12 | 0.37 |
| 21 | **** | 4 | 7 | 0.49 | 0.86 |
| 23 | ********** | 12 | 19 | 1.48 | 2.34 |
| 25 | ********** | 13 | 32 | 1.60 | 3.94 |
| 27 | ********** | 17 | 49 | 2.09 | 6.03 |
| 29 | ********** | 13 | 62 | 1.60 | 7.63 |
| 31 | ********** | 25 | 87 | 3.08 | 10.70 |
| 33 | ********** | 27 | 114 | 3.32 | 14.02 |
| 34 | ********** | 20 | 134 | 2.46 | 16.48 |
| 36 | ********** | 31 | 165 | 3.81 | 20.30 |
| 38 | ********** | 33 | 198 | 4.06 | 24.35 |
| 39 | ********** | 32 | 230 | 3.94 | 28.29 |
| 41 | ********** | 38 | 268 | 4.67 | 32.96 |
| 43 | ********** | 43 | 311 | 5.29 | 38.25 |
| 44 | ********** | 39 | 350 | 4.80 | 43.05 |
| 46 | ********** | 37 | 387 | 4.55 | 47.60 |
| 47 | ********** | 35 | 422 | 4.31 | 51.91 |
| 49 | ********** | 47 | 469 | 5.78 | 57.69 |
| 50 | ********** | 30 | 499 | 3.69 | 61.38 |
| 52 | ********** | 44 | 543 | 5.41 | 66.79 |
| 54 | ********** | 33 | 576 | 4.06 | 70.85 |
| 56 | ********** | 35 | 611 | 4.31 | 75.15 |
| 57 | ********** | 17 | 628 | 2.09 | 77.24 |
| 59 | ********** | 31 | 659 | 3.81 | 81.06 |
| 61 | ********** | 28 | 687 | 3.44 | 84.50 |
| 63 | ********** | 26 | 713 | 3.20 | 87.70 |
| 65 | ********** | 25 | 738 | 3.08 | 90.77 |
| 68 | ********** | 15 | 753 | 1.85 | 92.62 |
| 70 | ********** | 15 | 768 | 1.85 | 94.46 |
| 73 | ********** | 13 | 781 | 1.60 | 96.06 |
| 76 | ******** | 8 | 789 | 0.98 | 97.05 |
| 79 | ********* | 9 | 798 | 1.11 | 98.15 |
| 84 | **** | 4 | 802 | 0.49 | 98.65 |
| 89 | **** | 4 | 806 | 0.49 | 99.14 |
| 96 | *** | 3 | 809 | 0.37 | 99.51 |
| 98 | **** | 4 | 813 | 0.49 | 100.00 |

5 10 15 20 25 30 35 40 45

Frequency

## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 4

## Histogram Bars

| SSTOT | | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|---|
| 7 | * | 1 | 1 | 0.10 | 0.10 |
| 13 | * | 1 | 2 | 0.10 | 0.21 |
| 16 | * | 1 | 3 | 0.10 | 0.31 |
| 19 | ** | 3 | 6 | 0.31 | 0.62 |
| 21 | *** | 5 | 11 | 0.52 | 1.14 |
| 23 | ** | 4 | 15 | 0.41 | 1.55 |
| 25 | ******* | 14 | 29 | 1.45 | 3.00 |
| 27 | ********** | 20 | 49 | 2.07 | 5.07 |
| 29 | ********** | 32 | 81 | 3.31 | 8.38 |
| 31 | ********** | 27 | 108 | 2.79 | 11.17 |
| 33 | ********** | 30 | 138 | 3.10 | 14.27 |
| 34 | ********** | 34 | 172 | 3.52 | 17.79 |
| 36 | ********** | 47 | 219 | 4.86 | 22.65 |
| 38 | ********** | 38 | 257 | 3.93 | 26.58 |
| 39 | ********** | 39 | 296 | 4.03 | 30.61 |
| 41 | ********** | 44 | 340 | 4.55 | 35.16 |
| 43 | ********** | 50 | 390 | 5.17 | 40.33 |
| 44 | ********** | 65 | 455 | 6.72 | 47.05 |
| 46 | ********** | 49 | 504 | 5.07 | 52.12 |
| 47 | ********** | 54 | 558 | 5.58 | 57.70 |
| 49 | ********** | 38 | 596 | 3.93 | 61.63 |
| 50 | ********** | 38 | 634 | 3.93 | 65.56 |
| 52 | ********** | 30 | 664 | 3.10 | 68.67 |
| 54 | ********** | 43 | 707 | 4.45 | 73.11 |
| 56 | ********** | 38 | 745 | 3.93 | 77.04 |
| 57 | ********** | 29 | 774 | 3.00 | 80.04 |
| 59 | ********** | 30 | 804 | 3.10 | 83.14 |
| 61 | ********** | 36 | 840 | 3.72 | 86.87 |
| 63 | ********** | 28 | 868 | 2.90 | 89.76 |
| 65 | ********** | 24 | 892 | 2.48 | 92.24 |
| 68 | ********* | 18 | 910 | 1.86 | 94.11 |
| 70 | ********** | 19 | 929 | 1.96 | 96.07 |
| 73 | ****** | 11 | 940 | 1.14 | 97.21 |
| 76 | *** | 5 | 945 | 0.52 | 97.72 |
| 79 | **** | 7 | 952 | 0.72 | 98.45 |
| 84 | **** | 7 | 959 | 0.72 | 99.17 |
| 89 | **** | 7 | 966 | 0.72 | 99.90 |
| 96 | * | 1 | 967 | 0.10 | 100.00 |

10 20 30 40 50 60

Frequency

## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 5

## Histogram Bars

| SSTOT | | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|---|
| 13 | ** | 3 | 3 | 0.29 | 0.29 |
| 16 | ** | 4 | 7 | 0.38 | 0.67 |
| 19 | ** | 4 | 11 | 0.38 | 1.05 |
| 21 | ****** | 11 | 22 | 1.05 | 2.11 |
| 23 | ****** | 12 | 34 | 1.15 | 3.26 |
| 25 | ******* | 14 | 48 | 1.34 | 4.60 |
| 27 | ********* | 17 | 65 | 1.63 | 6.23 |
| 29 | ********** | 20 | 85 | 1.92 | 8.15 |
| 31 | ************* | 26 | 111 | 2.49 | 10.64 |
| 33 | ****************** | 37 | 148 | 3.55 | 14.19 |
| 34 | ******************* | 40 | 188 | 3.84 | 18.02 |
| 36 | ********************** | 43 | 231 | 4.12 | 22.15 |
| 38 | ***************** | 35 | 266 | 3.36 | 25.50 |
| 39 | ************************* | 51 | 317 | 4.89 | 30.39 |
| 41 | ************************ | 49 | 366 | 4.70 | 35.09 |
| 43 | ************************* | 53 | 419 | 5.08 | 40.17 |
| 44 | ************************** | 56 | 475 | 5.37 | 45.54 |
| 46 | ************************ | 52 | 527 | 4.99 | 50.53 |
| 47 | ***************************** | 60 | 587 | 5.75 | 56.28 |
| 49 | ********************************** | 69 | 656 | 6.62 | 62.90 |
| 50 | ********************** | 46 | 702 | 4.41 | 67.31 |
| 52 | ************************ | 52 | 754 | 4.99 | 72.29 |
| 54 | ************************ | 52 | 806 | 4.99 | 77.28 |
| 56 | ********************* | 41 | 847 | 3.93 | 81.21 |
| 57 | ***************** | 33 | 880 | 3.16 | 84.37 |
| 59 | ************ | 24 | 904 | 2.30 | 86.67 |
| 61 | ************ | 23 | 927 | 2.21 | 88.88 |
| 63 | ************** | 28 | 955 | 2.68 | 91.56 |
| 65 | ************** | 27 | 982 | 2.59 | 94.15 |
| 68 | ********* | 17 | 999 | 1.63 | 95.78 |
| 70 | ******** | 16 | 1015 | 1.53 | 97.32 |
| 73 | ***** | 9 | 1024 | 0.86 | 98.18 |
| 76 | *** | 6 | 1030 | 0.58 | 98.75 |
| 79 | *** | 6 | 1036 | 0.58 | 99.33 |
| 84 | ** | 4 | 1040 | 0.38 | 99.71 |
| 89 | * | 2 | 1042 | 0.19 | 99.90 |
| 96 | * | 1 | 1043 | 0.10 | 100.00 |

10 20 30 40 50 60 70

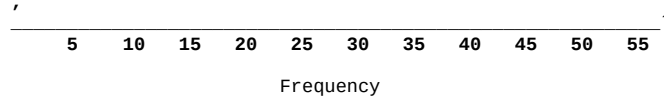
Frequency

## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 6

## Histogram Bars

| SSTOT | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|
| 13 | 1 | 1 | 0.10 | 0.10 |
| 16 | 4 | 5 | 0.41 | 0.51 |
| 19 | 5 | 10 | 0.51 | 1.03 |
| 21 | 10 | 20 | 1.03 | 2.05 |
| 23 | 15 | 35 | 1.54 | 3.59 |
| 25 | 17 | 52 | 1.74 | 5.33 |
| 27 | 25 | 77 | 2.56 | 7.90 |
| 29 | 20 | 97 | 2.05 | 9.95 |
| 31 | 25 | 122 | 2.56 | 12.51 |
| 33 | 21 | 143 | 2.15 | 14.67 |
| 34 | 46 | 189 | 4.72 | 19.38 |
| 36 | 28 | 217 | 2.87 | 22.26 |
| 38 | 28 | 245 | 2.87 | 25.13 |
| 39 | 52 | 297 | 5.33 | 30.46 |
| 41 | 50 | 347 | 5.13 | 35.59 |
| 43 | 36 | 383 | 3.69 | 39.28 |
| 44 | 46 | 429 | 4.72 | 44.00 |
| 46 | 44 | 473 | 4.51 | 48.51 |
| 47 | 53 | 526 | 5.44 | 53.95 |
| 49 | 48 | 574 | 4.92 | 58.87 |
| 50 | 41 | 615 | 4.21 | 63.08 |
| 52 | 55 | 670 | 5.64 | 68.72 |
| 54 | 51 | 721 | 5.23 | 73.95 |
| 56 | 45 | 766 | 4.62 | 78.56 |
| 57 | 36 | 802 | 3.69 | 82.26 |
| 59 | 36 | 838 | 3.69 | 85.95 |
| 61 | 22 | 860 | 2.26 | 88.21 |
| 63 | 31 | 891 | 3.18 | 91.38 |
| 65 | 20 | 911 | 2.05 | 93.44 |
| 68 | 12 | 923 | 1.23 | 94.67 |
| 70 | 16 | 939 | 1.64 | 96.31 |
| 73 | 14 | 953 | 1.44 | 97.74 |
| 76 | 8 | 961 | 0.82 | 98.56 |
| 79 | 4 | 965 | 0.41 | 98.97 |
| 84 | 6 | 971 | 0.62 | 99.59 |
| 89 | 3 | 974 | 0.31 | 99.90 |
| 98 | 1 | 975 | 0.10 | 100.00 |



## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 7

## Histogram Bars

| SSTOT | | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|---|
| 10 | * | 1 | 1 | 0.09 | 0.09 |
| 13 | * | 2 | 3 | 0.17 | 0.26 |
| 16 | ** | 4 | 7 | 0.35 | 0.61 |
| 19 | * | 2 | 9 | 0.17 | 0.78 |
| 21 | **** | 7 | 16 | 0.61 | 1.38 |
| 23 | **** | 7 | 23 | 0.61 | 1.99 |
| 25 | ********* | 17 | 40 | 1.47 | 3.46 |
| 27 | ****** | 11 | 51 | 0.95 | 4.41 |
| 29 | ********** | 22 | 73 | 1.90 | 6.31 |
| 31 | *************** | 29 | 102 | 2.51 | 8.82 |
| 33 | ***************** | 33 | 135 | 2.85 | 11.67 |
| 34 | ******************** | 40 | 175 | 3.46 | 15.13 |
| 36 | ******************** | 40 | 215 | 3.46 | 18.58 |
| 38 | ******************** | 39 | 254 | 3.37 | 21.95 |
| 39 | ************************ | 49 | 303 | 4.24 | 26.19 |
| 41 | ***************************** | 55 | 358 | 4.75 | 30.94 |
| 43 | ************************ | 48 | 406 | 4.15 | 35.09 |
| 44 | ************************ | 50 | 456 | 4.32 | 39.41 |
| 46 | ************************** | 54 | 510 | 4.67 | 44.08 |
| 47 | ******************************* | 64 | 574 | 5.53 | 49.61 |
| 49 | ********************************** | 69 | 643 | 5.96 | 55.57 |
| 50 | ************************ | 47 | 690 | 4.06 | 59.64 |
| 52 | ******************************** | 66 | 756 | 5.70 | 65.34 |
| 54 | ******************************* | 63 | 819 | 5.45 | 70.79 |
| 56 | ************************ | 50 | 869 | 4.32 | 75.11 |
| 57 | ********************** | 45 | 914 | 3.89 | 79.00 |
| 59 | ************************** | 56 | 970 | 4.84 | 83.84 |
| 61 | ********************** | 47 | 1017 | 4.06 | 87.90 |
| 63 | ************* | 25 | 1042 | 2.16 | 90.06 |
| 65 | ******************** | 38 | 1080 | 3.28 | 93.34 |
| 68 | ************* | 25 | 1105 | 2.16 | 95.51 |
| 70 | ********** | 20 | 1125 | 1.73 | 97.23 |
| 73 | ****** | 11 | 1136 | 0.95 | 98.18 |
| 76 | **** | 7 | 1143 | 0.61 | 98.79 |
| 79 | **** | 7 | 1150 | 0.61 | 99.39 |
| 84 | *** | 5 | 1155 | 0.43 | 99.83 |
| 96 | * | 2 | 1157 | 0.17 | 100.00 |

10 20 30 40 50 60 70

Frequency

## Scale Score Histogram for MSA-MOD Reading Assessment

## Grade 8

## Histogram Bars

| SSTOT | | Freq | Cum. Freq | Percent | Cum. Percent |
|---|---|---|---|---|---|
| 2 | ,* | 1 | 1 | 0.08 | 0.08 |
| 16 | ,* | 1 | 2 | 0.08 | 0.16 |
| 19 | ,** | 3 | 5 | 0.24 | 0.39 |
| 23 | ,* | 2 | 7 | 0.16 | 0.55 |
| 25 | ,* | 2 | 9 | 0.16 | 0.71 |
| 27 | ,****** | 11 | 20 | 0.87 | 1.58 |
| 29 | ,********** | 22 | 42 | 1.74 | 3.31 |
| 31 | ,******** | 16 | 58 | 1.26 | 4.57 |
| 33 | ,******** | 16 | 74 | 1.26 | 5.84 |
| 34 | ,********** | 24 | 98 | 1.89 | 7.73 |
| 36 | ,***************** | 34 | 132 | 2.68 | 10.41 |
| 38 | ,***************** | 36 | 168 | 2.84 | 13.25 |
| 39 | ,************************** | 53 | 221 | 4.18 | 17.43 |
| 41 | ,********************** | 47 | 268 | 3.71 | 21.14 |
| 43 | ,***************************** | 57 | 325 | 4.50 | 25.63 |
| 44 | ,********************** | 45 | 370 | 3.55 | 29.18 |
| 46 | ,****************************** | 61 | 431 | 4.81 | 33.99 |
| 47 | ,****************************** | 66 | 497 | 5.21 | 39.20 |
| 49 | ,********************************* | 68 | 565 | 5.36 | 44.56 |
| 50 | ,********************************** | 71 | 636 | 5.60 | 50.16 |
| 52 | ,********************************** | 71 | 707 | 5.60 | 55.76 |
| 54 | ,****************************************** | 81 | 788 | 6.39 | 62.15 |
| 56 | ,****************************** | 61 | 849 | 4.81 | 66.96 |
| 57 | ,***************************** | 59 | 908 | 4.65 | 71.61 |
| 59 | ,************************** | 52 | 960 | 4.10 | 75.71 |
| 61 | ,****************************** | 62 | 1022 | 4.89 | 80.60 |
| 63 | ,********************** | 44 | 1066 | 3.47 | 84.07 |
| 65 | ,************************* | 51 | 1117 | 4.02 | 88.09 |
| 68 | ,******************** | 41 | 1158 | 3.23 | 91.32 |
| 70 | ,**************** | 31 | 1189 | 2.44 | 93.77 |
| 73 | ,**************** | 32 | 1221 | 2.52 | 96.29 |
| 76 | ,********** | 20 | 1241 | 1.58 | 97.87 |
| 79 | ,***** | 10 | 1251 | 0.79 | 98.66 |
| 84 | ,***** | 10 | 1261 | 0.79 | 99.45 |
| 89 | ,*** | 6 | 1267 | 0.47 | 99.92 |
| 96 | ,* | 1 | 1268 | 0.08 | 100.00 |

10 20 30 40 50 60 70 80

Frequency

# APPENDIX D: STANDARD SETTING REPORT

# Maryland Standard Setting for
# The Modified Maryland School Assessment

**May 10–13, 2010**

**FINAL REPORT**

# Maryland Standard Setting for
# The Modified Maryland School Assessment
# May 10–13, 2010

## EXECUTIVE SUMMARY

Committees of Maryland educators were convened from May 10 to 13, 2010, in Towson, Maryland, to set standards for the Modified Maryland School Assessment (Mod-MSA) tests for Reading and Math, Grades 3–5. A total of 134 educators participated for two days per subject to recommend cut scores for these tests. The outcomes of the conference are described in this summary and more detailed information will be provided in a subsequent Standard Setting Technical Report.

The main purpose of the standard setting meetings was to obtain cut score recommendations for each grade within the two content areas for each of the three performance levels: Basic, Proficient, and Advanced. The item mapping procedure was applied to set the recommended standards. Under the item mapping procedure, the panelists are presented with test items and score points in an ordered item book in which each item will appear on a separate page in the book. The panelists are asked to place bookmarks between those items the borderline student for a particular performance level should answer correctly and those item such a student should answer incorrectly.

### Panelists

The panelists met in three committees: a committee representing Grade 3, a committee representing Grade 4, and a committee representing Grade 5. The number of panelists on each committee is shown in Table 1.

**Table 1.** The number of panelists on each committee

| Subject | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| Reading | 23 | 22 | 23 |
| Math | 23 | 22 | 21 |

All the panelists provided voluntary demographic information. Complete demographic information from the panelists will be summarized in the Standard Setting Technical Report. A summary of a subset of panelist demographic information is provided in Table 2, and a summary of the current positions of the panelists appears in Table 3.

**Table 2.** A summary of experience, gender, and ethnicity data for the committees

| Subject | Grade | Years Experience (Average) | Gender | | Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Male | Female | Caucasian | African American | Other | Missing |
| Reading | 3 | 16.35 | 1 | 22 | 19 | 3 | 0 | 1 |
| | 4 | 17.32 | 2 | 20 | 18 | 4 | 0 | 0 |
| | 5 | 18.87 | 2 | 21 | 16 | 6 | 1 | 0 |
| Math | 3 | 14.37 | 1 | 22 | 19 | 2 | 1 | 1 |
| | 4 | 12.48 | 1 | 21 | 15 | 5 | 1 | 1 |
| | 5 | 14.90 | 2 | 19 | 15 | 5 | 1 | 0 |

**Table 3.** Panelists' current positions in Maryland

| Subject | N | Positions | | | | | |
|---|---|---|---|---|---|---|---|
| | | CSO | GET | SET | SES | AP | SLP |
| Reading | 68 | 16 | 17 | 24 | 8 | 2 | 1 |
| Math | 66 | 14 | 22 | 25 | 4 | 1 | 0 |

**Note:** CSO: Content Specialist/Content Supervisor (Central Office); GET: General Education Teacher; SET: Special Education Teacher; SES: Special Education Supervisor (Central Office); AP: Assistant Principal; SLP: Speech and Language Pathologist.

## Method and Procedure

The standard setting conference began on Monday, May 10. The Reading committees met first (on May 10 and 11), followed by the Math committees. The Reading and the Math committees followed identical agendas and processes. Therefore, the process presented in this document applies to both content areas.

Monday morning was devoted to introductions of the staff, to a description of standard setting, and to a description of the Mod-MSA tests and student population. For this stage of the conference, all the panelists met together in one large room. The agendas for the standard setting are shown in Appendix A. All committees within a subject followed the same agenda.

After the midmorning break, the committee members broke into their grade-specific groups. The three committees (Grade 3, Grade 4, and Grade 5) met separately in individual conference rooms. The committee members spent the remainder of the morning working individually to familiarize themselves with the Mod-MSA test items for their grade by completing the items in their ordered item booklets (OIBs). OIBs were constructed for the three grades by using items from the spring 2010 test administration. These booklets were created by augmenting items from the scored form (45 items for Reading and 51 items for Math) with unscored items that had acceptable item statistics to provide the most complete coverage possible of the scale score range

while maintaining content representation. The scale score associated with a response probability of 0.67 was calculated for each item in the OIB on the basis of spring 2010 data. Items were ordered on the basis of these scale scores and sorted from least to most difficult.

Each ordered item book was accompanied by an item map. The item maps for Reading Grade 3, Grade 4, and Grade 5 are shown in Appendices C, D, and E, respectively. Math Grade 3, Grade 4, and Grade 5 can be found in Appendices F, G, and H, respectively. Each item map contains seven pieces of information:

1. Page number
2. A unique item identifier
3. Item position on the administration form
4. Reporting strand
5. Content category or standard
6. Correct option
7. Location (scale score)
8. *p value*[3]

Before and after lunch, the panelists reviewed the Mod-MSA performance level descriptors (PLDs) and created behavioral anchors to clearly and concretely describe "threshold" or minimally qualified students at the Proficient and Advanced performance levels. Throughout this process, the panelists were led through table-level and committee-level discussions by a Pearson facilitator. This process required Monday afternoon. The result from creating performance level descriptors was a set of descriptors for threshold students at the Proficient and Advanced levels.

After this process of PLD review, the panelists received additional training in the item mapping process. This training was provided by a Pearson psychometrician within each committee. After the training, each committee practiced the item mapping process as a group, using a practice OIB constructed from unused Mod-MSA items. This allowed the panelists to gain familiarity with the method and ask questions before beginning the process. For Grade 4, Math and Reading training did not include application to the practice ordered item booklet. Instead, the focus was on the process steps. In the item mapping procedure, the panelists are asked to identify the item in an OIB that is the last item that a threshold student at a given level would be able to correctly answer. The panelists were instructed to identify the last item in an ordered item book that a threshold student at a given level would have a response probability of at least 0.67 of answering correctly (Huynh, 2006).

After this training, all three committees began the standard setting process late Monday afternoon. The standard setting process consisted of three rounds of judgments. The panelists were provided with feedback after each round. The feedback was intended to inform the panelists' decisions but not to dictate their ratings. After round 1, the panelists met in small groups of four or five panelists each. The panelists were provided the cut scores for each panelist in the group based on the round 1 of ratings in addition to the mean and median cut

[3] The P value information was shared during the second round of item mapping.

score at each level for that table. In reviewing the cut score report, the panelists were asked to think about the following:

- How similar are their cut scores to the cut score of the group (i.e., is a given panelist more lenient or stringent than the other panelists)?
- If so, why is this the case?
- Do the panelists have different conceptualizations of these threshold students?

The panelists were informed that there was no intention for them to come to consensus on their cut score judgments but that they should discuss differences to get a feel for why differences existed. After round 1, the panelists were provided with an item map containing *p values*, where a *p values* is an index of student performance on each test item. The panelists were informed that this information was to help them better understand the ordering of items, and that it would not provide any specific insights about the performance of students at a given level.

After round 2, the panelists received the same feedback for each table that was provided after round 1. Next, the panelists were given the mean and median cut scores for the committee, across tables. The Pearson facilitator led the discussion with the panelists from all five tables combined. The facilitator noted the differences and similarities across tables but reminded the panelists that consensus was not required.

Finally, the panelists were provided a graphic display of the impact of using the median cut score for all students. The impact data graphic representation provided the panelists with information on what percentages of students are at each performance level for the populations of interest (all students, African American/Caucasian, and female/male). The panelists were given time to discuss, within the big group, the appropriateness of the committee level cut scores given the proportion of students in each level.

After round 3, the panelists were shown the cut scores they were recommending on the basis of this final round of ratings, the panelists were given the mean and median cut scores for the committee, across tables, and were provided a graphic display of the impact of using the median cut score for all students.

## Results

### Round 3 Cut Scores

The Reading Grade 3, Grade 4, and Grade 5 ordered item books contained 53, 53, and 52 ordered items, respectively. The Math Grade 3, Grade 4, and Grade 5 ordered item books contained 62, 61, and 64 ordered items, respectively. Table 4 summarizes the cut scores after the round 3 final rating for these tests. These are the committees' recommendations based on item location in the ordered item book. The scale score cuts associated with these recommendations and the percentages of students in the Advanced and Proficient performance levels based on these cuts are presented in Table 5. Please note that separate committees made recommendations for each of these tests. Mean, median, minimum, and maximum ratings by

round are presented in Appendix J. Graphs presenting individual ratings across the three rounds by performance level are presented for Reading Grade 3 in Appendix K, Reading Grade 4 in Appendix L, Reading Grade 5 in Appendix M, Math Grade 3 in Appendix N, Math Grade 4 in Appendix O, and Math Grade 5 in Appendix P.

**Table 4.** OIB Cut scores after round 3 by subject and grade

| Subject | Grade | Score | Proficient | Advanced |
|---|---|---|---|---|
| Reading | 3 | Mean | 24.87 | 45.09 |
| | | Median | 24.00 | 45.00 |
| | 4 | Mean | 16.91 | 39.36 |
| | | Median | 18.00 | 42.00 |
| | 5 | Mean | 18.35 | 40.09 |
| | | Median | 18.00 | 41.00 |
| Math | 3 | Mean | 20.17 | 44.30 |
| | | Median | 19.00 | 44.00 |
| | 4 | Mean | 18.86 | 52.68 |
| | | Median | 17.00 | 53.00 |
| | 5 | Mean | 18.52 | 48.76 |
| | | Median | 18.00 | 49.00 |

**Table 5.** Scale score cut scores after round 3 with associated impact by subject

| Subject | Grade | Proficient SS Cut | Percentage Proficient* | Advanced SS Cut | Percentage Advanced |
|---|---|---|---|---|---|
| Reading | 3 | 55 | 20.4 | 65 | 13.3 |
| | 4 | 54 | 25.9 | 66 | 12.4 |
| | 5 | 53 | 35.7 | 69 | 8.4 |
| Math | 3 | 55 | 22.5 | 67 | 13.4 |
| | 4 | 54 | 28.5 | 68 | 9.9 |
| | 5 | 58 | 21.5 | 71 | 8.2 |

*The percentage indicates students who were Proficient but not Advanced.

Figure 1 shows the percentage of students in each performance level, using the cut scores after the round 3 final rating for Reading Grade 3, Grade 4, and Grade 5.

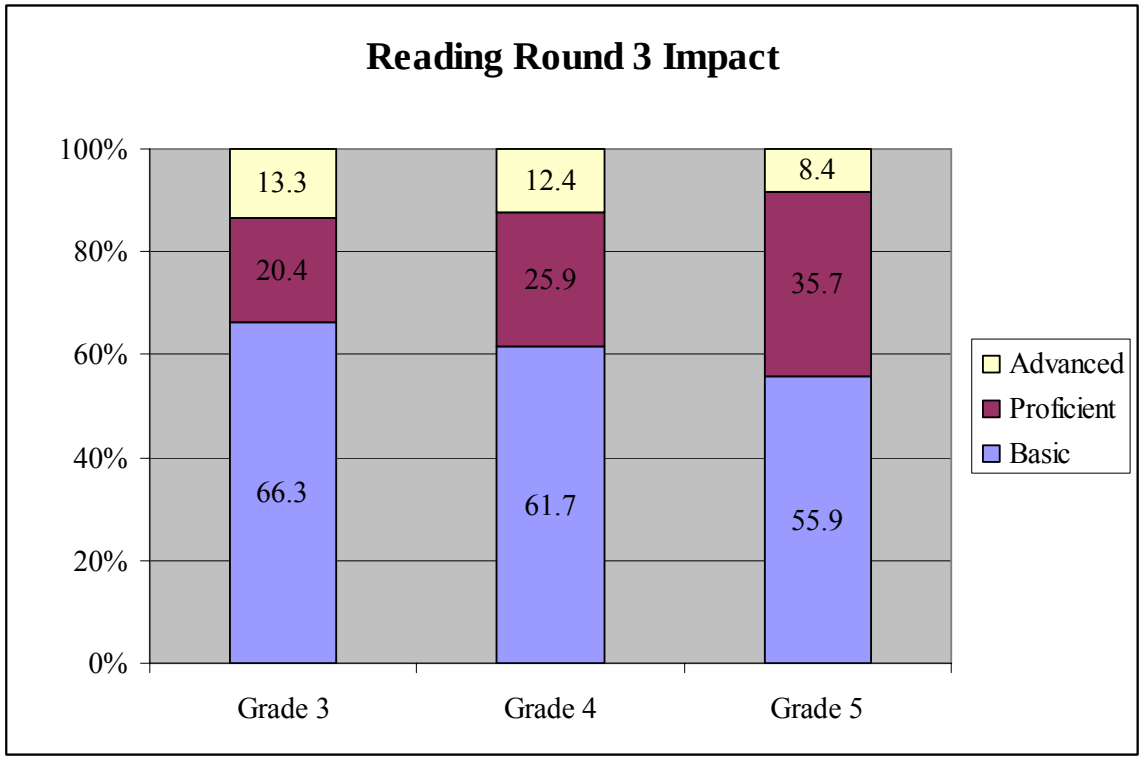



**Figure 1.** The percentages of students in each performance level, using the final cut scores for Reading by grade.

Figure 2 shows the percentage of students in each performance level, using the cut scores after the round 3 final rating for Math Grade 3, Grade 4, and Grade 5.

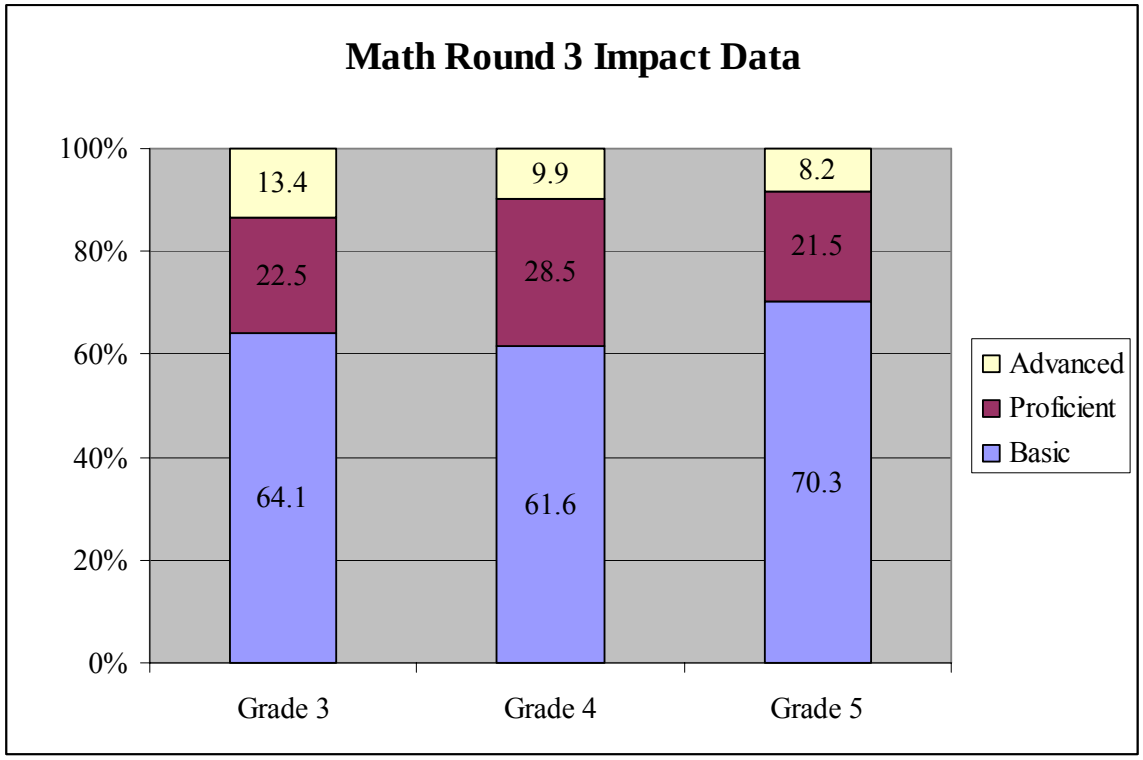



**Figure 2.** The percentage of students in each performance level, using the final cut scores for Math by grade.

# FINAL REPORT

This document provides a detailed description of the standard-setting procedures used with the Maryland Mod-MSA Mathematics and Reading tests. The main purpose of the standard-setting meetings was to obtain cut score recommendations for each grade within the two content areas for each of the three performance levels: Basic, Proficient, and Advanced.

Committees of Maryland educators were convened May 10 through May 13, 2010, in Towson, Maryland, to set standards for the Modified Maryland School Assessment (Mod-MSA) tests for Math and Reading, grades 3 through 5 (see Appendix A for agenda). A total of 134 educators participated for two days per subject to recommend cut scores for these tests. The item mapping procedure was applied to set the recommended standards.

A total of six vendor staff members were involved in conducting the standard setting activity. For each standard setting meeting, a facilitator from Pearson provided training in the implementation of the standard setting procedure and the interpretation and use of feedback data. The Pearson facilitators were: Dr. Daniel Murphy, Dr. Stephen Murphy, and Dr. Kimberly O'Malley. In addition, one staff member from Pearson served the role of a data analyst, supporting the facilitator by taking notes, collecting judge's ratings and performing all analyses required to generate feedback reports. The Pearson data analyst was Morgen Hickey. Two additional Pearson staff members, Scott Hanlin and Andrea Tompkins, were present to oversee the standard setting meeting, coordinate meals, assist the psychometricians, and accommodate any unforeseen requests.

## Panelists

The panelists met in three committees: a committee representing Grade 3, a committee representing Grade 4, and a committee representing Grade 5. The number of panelists on each committee is shown in Table 1.

**Table 1.** The number of panelists on each committee

| Subject | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| Reading | 23 | 22 | 23 |
| Math | 23 | 22 | 21 |

All panelists provided voluntary demographic information, using the form shown in Appendix B. A summary of panelist gender and ethnicity information is provided in Table 2.

**Table 2.** A summary of experience, gender, and ethnicity data for the committees

| Subject | Grade | Years Experience (Average) | Gender | | Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Male | Female | Caucasian | African American | Other | Missing |
| Reading | 3 | 16.35 | 1 | 22 | 19 | 3 | 0 | 1 |
| | 4 | 17.32 | 2 | 20 | 18 | 4 | 0 | 0 |
| | 5 | 18.87 | 2 | 21 | 16 | 6 | 1 | 0 |
| Math | 3 | 14.37 | 1 | 22 | 19 | 2 | 1 | 1 |
| | 4 | 12.48 | 1 | 21 | 15 | 5 | 1 | 1 |
| | 5 | 14.90 | 2 | 19 | 15 | 5 | 1 | 0 |

Table 3 provides a summary of panelists' responses to the question, "Compared to other school districts in Maryland, how would you describe the size of your district?"

**Table 3.** Summary of panelists' responses to district size.

| | Reading | | | Math | | |
|---|---|---|---|---|---|---|
| District Size | Grade 3 | Grade 4 | Grade 5 | Grade 3 | Grade 4 | Grade 5 |
| Large | 11 | 10 | 9 | 9 | 6 | 10 |
| Medium | 8 | 7 | 9 | 9 | 10 | 8 |
| Small | 4 | 5 | 5 | 5 | 6 | 3 |

Table 4 provides a summary of panelists' responses to the question, "Compared to other school districts in Maryland, how would you describe the location of your district?"

**Table 4.** Summary of panelists' responses to district location.

| | Reading | | | Math | | |
|---|---|---|---|---|---|---|
| District Location | Grade 3 | Grade 4 | Grade 5 | Grade 3 | Grade 4 | Grade 5 |
| Rural | 7 | 6 | 9 | 7 | 8 | 4 |
| Suburban | 10 | 8 | 9 | 12 | 10 | 9 |
| Urban | 3 | 5 | 5 | 4 | 3 | 7 |
| Multiple Response | 2 | 3 | | | 1 | |

Table 5 provides a summary of panelists' responses to the question, "Compared to other school districts in Maryland, how would you describe the geographic location of your district?"

**Table 5.** Summary of panelists' responses to district geographic location.

| | Reading | | | Math | | |
|---|---|---|---|---|---|---|
| Geographic Location | Grade 3 | Grade 4 | Grade 5 | Grade 3 | Grade 4 | Grade 5 |
| Central | 10 | 8 | 10 | 7 | 5 | 8 |
| East | 5 | 4 | 2 | 3 | 6 | 2 |
| North | 2 | 2 | 1 | 3 | 1 | 2 |
| South | 3 | 3 | 4 | 5 | 8 | 6 |
| West | 2 | 3 | 5 | | 2 | 2 |
| Multiple Response | 1 | 1 | 1 | 5 | | 1 |

## Method and Procedure

The standard-setting conference began on Monday, May 10. The Reading committees met first (on May 10 and 11), followed by the Math committees. The Reading and the Math committees followed identical agendas and processes. For simplicity the process is presented only once in this document.

The morning of Monday, May 10, was devoted to introductions of the staff, to a description of standard setting, and to a description of the Mod-MSA tests and student population. For this stage of the conference, all panelists met together in one large room. The agendas for the standard setting are shown in Appendix A. All committees within a subject followed the same agenda.

Following the midmorning break, the committee members broke into their grade specific groups. The three committees (Grade 3, Grade 4, and Grade 5) met separately in individual conference rooms. The committees spent the remainder of the morning working individually to familiarize themselves with the Mod-MSA test items for their grade by completing the items in their ordered item booklet (OIB). OIBs were constructed for the three grades using items from the spring 2010 test administration. These booklets were created by augmenting items from the scored form (45 items for Reading and 51 items for Math) with unscored items with acceptable item statistics in order to provide the most complete coverage possible of the scale score range while maintaining content representation. The scale score associated with a response probability of 0.67 was calculated for each item in the OIB based on spring 2010 data. Items were ordered based on these scale scores and sorted from least to most difficult.

Each ordered item book was accompanied by an item map. The item maps for Reading Grade 3, Grade 4, and Grade 5 are shown in Appendices C, D and E, respectively, Math grades 3 through 5 can be found in Appendices F, G, and H, respectively. Each item map contains eight pieces of information:

1. Page number

2. A unique item identifier
3. Item position on the administration form
4. Reporting strand
5. Content category or standard
6. Correct option
7. Location (scale score)
8. *p value*[4]

Before and after lunch, the panelists reviewed the Mod-MSA performance level descriptors (PLDs) and created behavioral anchors to clearly and concretely describe "threshold" or minimally qualified students at the Proficient and Advanced performance levels. Throughout this process the panelists were led thorough table-level and committee-level discussions by a Pearson facilitator. This process required the afternoon of Monday, May 10. The result from creating performance level descriptors was a set of descriptors for threshold students at the Proficient and Advanced levels.

After this process of PLD review, the panelists received additional training in the item mapping process. This training was provided by a Pearson psychometrician within each committee. Following the training, each committee practiced the item mapping process as a group, using a practice OIB constructed from unused Mod-MSA items. This allowed the panelists to gain familiarity with the method and ask questions before beginning the process. For Grade 4, Math and Reading training did not include application to the practice ordered item booklet. Instead, the focus was on the process steps. In the item mapping procedure, the panelists are asked to identify the item in an OIB that is the last item that a threshold student at a given level would be able to correctly answer. The panelists were instructed to identify the last item in an ordered item book that a threshold student at a given level would have a response probability of at least 0.67 of answering correctly (Huynh, 2006).

After this training, all three committees began the standard-setting process late Monday afternoon. The standard-setting process consisted of three rounds of judgments. During each round, panelists were asked to assign cut scores for each performance level. The panelists reviewed the items and placed bookmarks in the item book where they believed the cut scores should be. This was determined as the point at which *threshold* students of that proficiency level have a probability of at least 0.67 of responding correctly to that item and the items before it, and less than that probability of responding correctly to items following it.

"Threshold" examinees are students with the minimum level of proficiency needed to make it into a particular proficiency level. It is this hypothetical population of students that panelists must reference when making judgments about items. Therefore, it is extremely important that each judge have an understanding of what defines this group. was no easy task. The behavioral anchors generated earlier were used to define the knowledge and skills that characterize a typical "threshold" student in each level and provide a frame of reference for conceptualizing this population.

To evaluate whether the training activities successfully helped panelists understand the task, a readiness survey was completed by each panelist prior to each round of judgments (Appendix I). The readiness survey asked panelists to report if they understood the task Pearson facilitators

[4] The P value information was shared during the second round of item mapping.

asked of them as well as any feedback data provided. Results of the readiness survey indicated if panelists unanimously understood their tasks for the rounds, were ready to begin the rounds, and understood the data presented. Table leaders were instructed to check the panelists' answers. If any panelist appeared to have questions about the next task, the table leader was instructed to answer the questions. If additional assistance was needed, the table leader alerted the facilitator to address the remaining questions.

In round one, panelists were divided into small groups. The panelists then worked independently to place the bookmarks.

In round two, still in small groups, panelists compared bookmarks and discussed the differences between them. Panelists were encouraged to describe the reasons they set bookmarks where they did. The discussion addressed all items in the range between the highest and lowest bookmark for a proficiency level. Once the discussion was over, the panelists independently reconsider their bookmark locations.

Following round two, still in small groups, panelists again compared bookmarks and discussed the differences between them. Next, panelists reconvened as a large group where cut score differences across small groups were discussed. A panelist from each small group presented the conclusions of their group.

In a final, third round, panelists independently made final bookmark placements. Panelists were then briefed on the results of their Round 3 ratings.

The cut score at each performance level was determined by computing the median page number recommended across panelists at a given grade level and identifying the scale score associated with this page in the OIB. This represents the minimum scale score that an examinee must attain to be classified at the particular level. Computed cuts could fall between page numbers. In the final report, all median page numbers were rounded to the next higher point if the decimal value is larger than 4 (e.g., 15.5 would become 16) prior to identifying the scale score for the recommended cut.

After the Round 3 rating sheets were collected, Pearson staff members analyzed data and produced the final cut score recommendations. The panelists reconvened and were presented the final cut score recommendations. The panelists were then asked to complete a short questionnaire, evaluating the standard-setting process. The questionnaire asked about panelists' level of comfort with the standard-setting procedure, their understanding of the performance levels and their satisfaction with final cut scores. More information about this is provided in the Evaluation section of this report.

Panelists were provided with feedback between each round. The feedback was intended to inform the panelist's decisions, but not to dictate their ratings. Following Round 1, panelists met in small groups of 5 to 7 panelists. They were provided the cut scores for each panelist based on the Round 1 ratings in addition to the mean and median cut score at each level for that table. In reviewing the cut score report panelists were asked to think about the following:

- How similar are their cut scores are to that of the group (i.e., is a given panelist more lenient or stringent than the other panelists)?
- If so, why is this the case?

- Do panelists have different conceptualization of these threshold students?

Panelists were informed that there was no intention for them to come to consensus on their cut score judgments, but they should discuss differences to get a feel for why differences exist. Following Round 1 panelists were also provided with an item map containing P values, an index of student performance on each test item. Panelists were given this information to help them better understand the ordering of items, but were cautioned that it would not provide any specific insights about the performance of students at a given level.

Following Round 2, panelists received the same table level feedback that was provided following Round 1. Next, panelists were given the mean and median cut scores for the committee (across tables). The Pearson facilitator lead the discussion with all five tables combined. The facilitator noted the differences and similarities across tables but reminded the panelists that consensus was not required.

Panelists were then shown a graphical display of the impact of using the round 2 median cut score. The impact data provided information on what percentage of students fall into each performance level for all students and for sub-populations of interest (African-American/white, female/male). Panelists were given time to discuss, within the big group, the appropriateness of the committee level cut scores given the proportion of students in each level.

Following Round 3, panelists were shown the cut scores they were recommending based on this round of ratings, given the mean and median cut scores for the committee (across tables), and provided a graphical display of the impact of using the median cut score for all students.

## Results

### Round 3 Cut Scores

The Reading Grade 3, Grade 4, and Grade 5 ordered item books contained 53, 53, and 52 ordered items, respectively. The Math Grade 3, Grade 4, and Grade 5 ordered item books contained 62, 61, and 64 ordered items, respectively. Table 6 summarizes the cut scores after the Round 3 final ratings. These are the recommendations from the committees based on item location in the ordered item book. The scale score cuts associated with these recommendations and the percentage of students in the advanced and proficient performance levels based upon these cuts are presented in Table 7. Please note that separate committees made recommendations for each of these tests. Mean, median, minimum, and maximum ratings by round are presented in Appendix J. Graphs presenting individual ratings across the three rounds by performance level are presented for Reading Grade 3 in Appendix K, Reading Grade 4 in Appendix L, Reading Grade 5 in Appendix M, Math Grade 3 in Appendix N, Math Grade 4 in Appendix O, and Math Grade 5 in Appendix P.

**Table 6.** OIB Cut scores after Round 3 by subject and grade.

| Subject | Grade | Score | Proficient | Advanced |
|---|---|---|---|---|
| Reading | 3 | Mean | 24.87 | 45.09 |
| | | Median | 24.00 | 45.00 |

| | | | | |
|---|---|---|---|---|
| | 4 | Mean | 16.91 | 39.36 |
| | | Median | 18.00 | 42.00 |
| | 5 | Mean | 18.35 | 40.09 |
| | | Median | 18.00 | 41.00 |
| Math | 3 | Mean | 20.17 | 44.30 |
| | | Median | 19.00 | 44.00 |
| | 4 | Mean | 18.86 | 52.68 |
| | | Median | 17.00 | 53.00 |
| | 5 | Mean | 18.52 | 48.76 |
| | | Median | 18.00 | 49.00 |

**Table 7.** Scale score cut scores after the Round 3 with associated impact by subject.

| Subject | Grade | Proficient SS Cut | Percentage Proficient* | Advanced SS Cut | Percentage Advanced |
|---|---|---|---|---|---|
| Reading | 3 | 55 | 20.4 | 65 | 13.3 |
| | 4 | 54 | 25.9 | 66 | 12.4 |
| | 5 | 53 | 35.7 | 69 | 8.4 |
| Math | 3 | 55 | 22.5 | 67 | 13.4 |
| | 4 | 54 | 28.5 | 68 | 9.9 |
| | 5 | 58 | 21.5 | 71 | 8.2 |

*The percentage indicates students who were Proficient but not Advanced.

Figure 1 shows the percentage of students in each performance level, using the cut scores after the round 3 final rating for Reading Grade 3, Grade 4, and Grade 5.

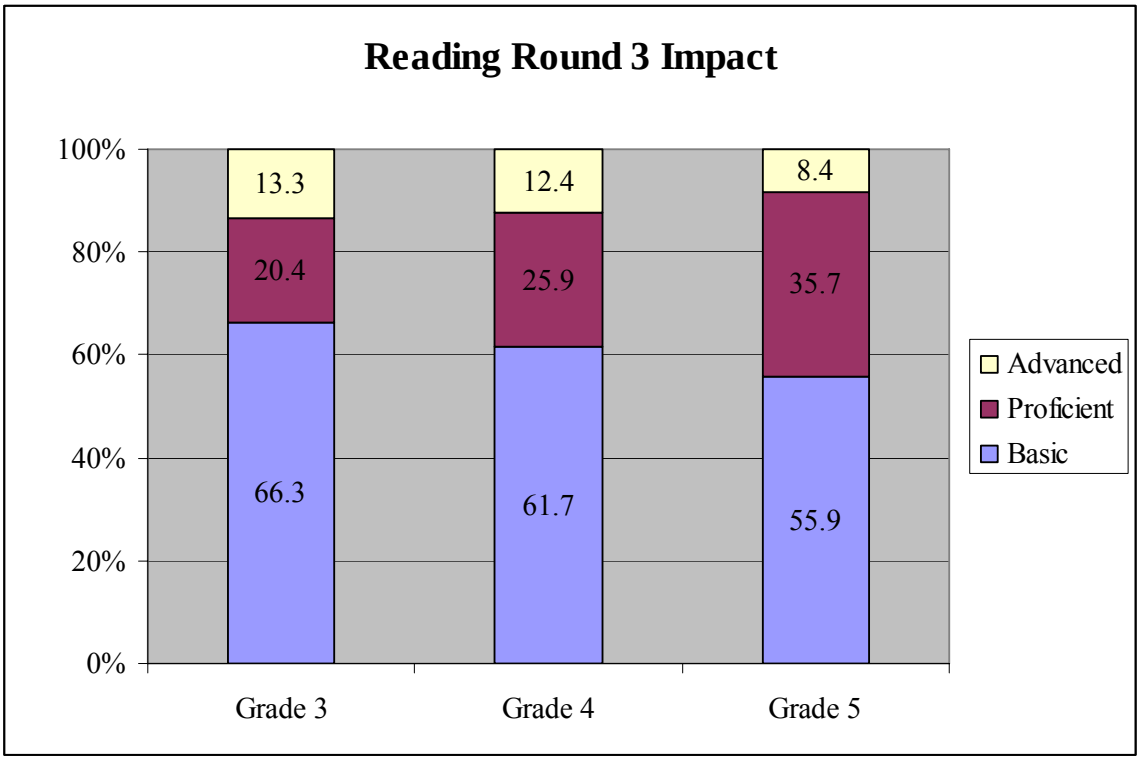



**Figure 1.** The percentages of students in each performance level, using the final cut scores for Reading by grade.

Figure 2 shows the percentage of students in each performance level, using the cut scores after the round 3 final rating for Math Grade 3, Grade 4, and Grade 5.

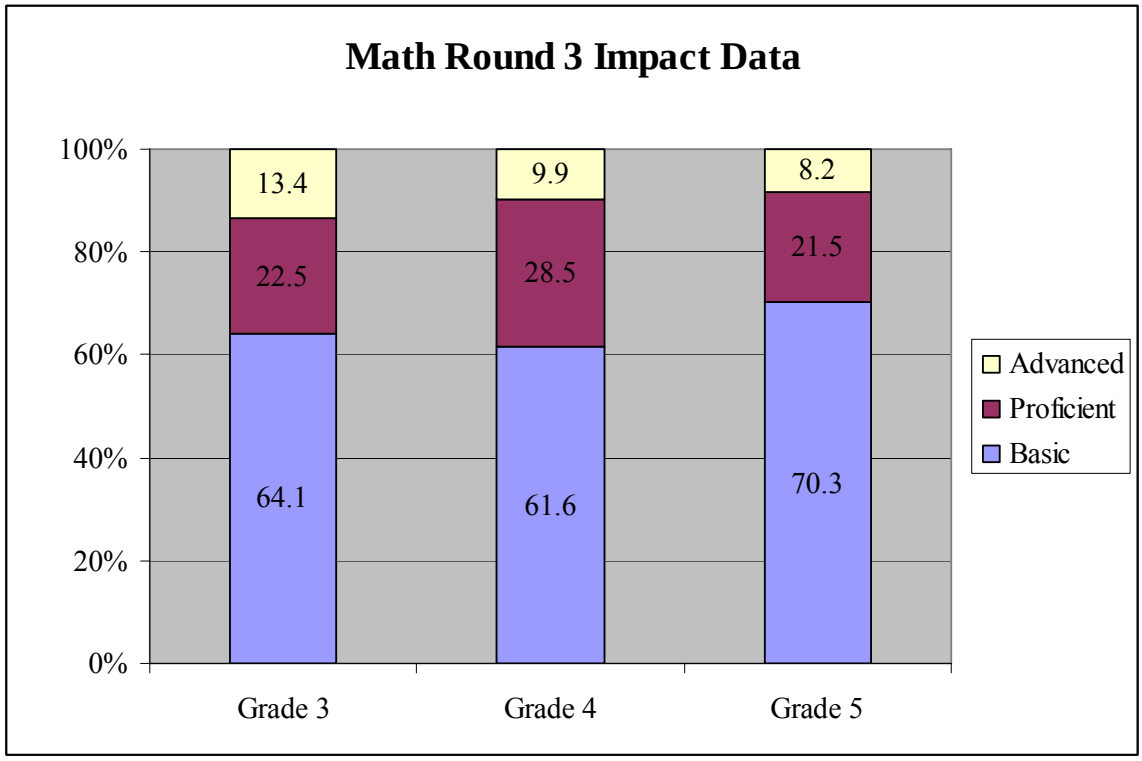



**Figure 2.** The percentage of students in each performance level, using the final cut scores for Math by grade.

**Panelist Variability**

In order to describe the variability in panelists' judgments, a Generalizability Theory (G-Theory) study was performed. This information could be used to determine how similar the cut scores might be if a different set of panelists or different composition of small groups were used to set cut scores. For this investigation, the sources of variability of interest were panelists, small groups, and rounds. For each cut score, the variance associated with each of these sources was estimated using the maximum likelihood SAS VARCOMP procedure. For this study, the number of rounds was treated as a fixed factor (3 rounds in total, a typical practice in standard setting meetings), meaning that if the standard setting meeting was held again, the same number of rounds would be used. In addition, because judges discussed all activities in small groups, their judgments were considered dependent on group membership. Therefore, judges were considered "nested" within tables. Variances components for tables ($\sigma^2_{Tables}$) and judges within tables ($\sigma^2_{Judges:Tables}$) were computed. Computation of the standard errors was made using the following formula (Lee & Lewis, 2008):

$$SE_{cut} = \sqrt{\frac{\sigma^2_{Tables}}{N_{Tables}} + \frac{\sigma^2_{Judge:Table}}{N_{Judges} \bullet N_{Tables}} + \frac{\sigma^2_{Error}}{3N_{Tables} \bullet N_{Judges}}}.$$

Because round was treated as a fixed facet, its variance component was not included in the error term. $\sigma^2_{error}$ was a confounding term and included the variance from the interaction between tables and judges within tables as well as variances unexplained by the defined facets. The sample size in the equation referred to the sample size likely to occur in the Decision Study (D study). Without loss of generality, the sample sizes for the D study were assumed the same as the sample size in the G study. Standard errors were computed for each of the two recommended cut scores associated with each Mod-MSA test. For the purposes of this analysis the recommended cut scores were the scale scores associated with the pages bookmarked during standard setting. Different patterns of variance component estimates and hence standard errors for cut scores were anticipated for different cut scores (Lee & Lewis, 2008).

The conditional standard error of measurement (CSEM) for each recommended scale score cut for each Mod-MSA test was calculated using the following formula:

$$CSEM = 1 \Big/ \sqrt{I(SS)}$$

In this formula *I(SS)* is the amount of psychometric information at a given scale score point; in this case this was the amount of information at each of the two recommended scale score cuts.

The standard error of the cut score ($SE_{cut}$) and conditional standard error of measurement (CSEM) were used to compute a composite standard error ($SEM_{combined}$) calculated using the following formula:

$$SEM_{combined} = \sqrt{(SE_{cut})^2 + (CSEM)^2}$$

These different standard error indices are presented for each test by grade and committee in Table 8.

**Table 8.** Standard Error Indices by Test, Grade and Committee.

| Committee | Grade | Cut | $SE_{cut}$ | *CSEM* | $SEM_{combined}$ |
|---|---|---|---|---|---|
| Reading | 3 | Proficient | 0.46 | 4.00 | 4.03 |
| | | Advanced | 1.26 | 5.00 | 5.16 |
| | 4 | Proficient | 0.93 | 5.00 | 5.09 |
| | | Advanced | 1.67 | 5.00 | 5.27 |
| | 5 | Proficient | 0.97 | 5.00 | 5.09 |
| | | Advanced | 1.74 | 6.00 | 6.25 |
| Math | 3 | Proficient | 0.99 | 5.00 | 5.10 |
| | | Advanced | 0.81 | 5.00 | 5.06 |
| | 4 | Proficient | 1.11 | 4.00 | 4.19 |
| | | Advanced | 0.97 | 5.00 | 5.09 |
| | 5 | Proficient | 0.96 | 5.00 | 5.09 |
| | | Advanced | 1.09 | 6.00 | 6.10 |

Each of these indices was applied to the panel recommended cut scores to produce 1, 2, and 3 standard error bands around the cut score. These results are reported in Appendix Q for Reading Grade 3, Appendix R for Reading Grade 4, Appendix S for Reading Grade 5, Appendix T for Math Grade 3, Appendix U for Math Grade 4, and Appendix V for Math Grade 5.

## Evaluations

Exit surveys were administered following the completion of standard setting for each committee. An exit survey was completed by each panelist. For the Reading Grades 3, 4, and 5 and the Math Grades 3, 4 and 5 committees, these questions and the results are shown in Tables 9, 10,

11, 12, 13, and 14 respectively. Responses to each question were on a five-point scale (1 = Totally Disagree, 5 = Totally Agree).

**Table 9.** The questionnaire results for the Reading Grade 3 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.55 | 5.00 | 2 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.61 | 5.00 | 3 | 5 |
| I could clearly distinguish between student performance levels. | 4.00 | 4.00 | 3 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 4.00 | 4.00 | 2 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.39 | 4.00 | 3 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.43 | 5.00 | 3 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.30 | 4.00 | 3 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Grade 3 Reading Test. | 4.27 | 4.00 | 2 | 5 |

**Table 10.** The questionnaire results for the Reading Grade 4 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.55 | 5.00 | 3 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.59 | 5.00 | 3 | 5 |
| I could clearly distinguish between student performance levels. | 4.05 | 4.00 | 3 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 4.27 | 4.00 | 3 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.41 | 4.50 | 3 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.68 | 5.00 | 3 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.36 | 5.00 | 3 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Reading Grade 4 Test. | 4.23 | 4.00 | 4 | 5 |

**Table 11.** The questionnaire results for the Reading Grade 5 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.50 | 5.00 | 2 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.82 | 5.00 | 4 | 5 |
| I could clearly distinguish between student performance levels. | 4.14 | 4.00 | 3 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 3.73 | 4.00 | 2 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.64 | 5.00 | 3 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.77 | 5.00 | 3 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.41 | 4.00 | 4 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Reading Grade 5 Test. | 4.64 | 5.00 | 4 | 5 |

**Table 12.** The questionnaire results for the Math Grade 3 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.61 | 5.00 | 1 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.73 | 5.00 | 2 | 5 |
| I could clearly distinguish between student performance levels. | 4.43 | 5.00 | 2 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 4.30 | 5.00 | 1 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.48 | 5.00 | 1 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.57 | 5.00 | 1 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.55 | 5.00 | 2 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Math Grade 3 Test. | 4.50 | 5.00 | 1 | 5 |

**Table 13.** The questionnaire results for the Math Grade 4 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.86 | 5.00 | 4 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.91 | 5.00 | 4 | 5 |
| I could clearly distinguish between student performance levels. | 4.18 | 4.00 | 3 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 4.32 | 4.50 | 3 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.52 | 5.00 | 4 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.73 | 5.00 | 4 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.45 | 5.00 | 1 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Math Grade 4 Test. | 4.64 | 5.00 | 4 | 5 |

**Table 14.** The questionnaire results for the Math Grade 5 standard setting committee

| Question | Mean | Median | Min | Max |
|---|---|---|---|---|
| The method for setting standards, item mapping, was conceptually clear. | 4.33 | 4.00 | 2 | 5 |
| I had a good understanding of what the test was intended to measure. | 4.57 | 5.00 | 4 | 5 |
| I could clearly distinguish between student performance levels. | 4.29 | 4.00 | 3 | 5 |
| After the first round of ratings, I felt comfortable with the standard setting procedure. | 4.14 | 4.00 | 2 | 5 |
| I found the feedback on item difficulty useful in setting standards. | 4.33 | 4.00 | 3 | 5 |
| I found the feedback on the ratings of judges compared to other judges useful in setting standards. | 4.76 | 5.00 | 4 | 5 |
| I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards. | 4.29 | 4.00 | 3 | 5 |
| I feel confident that the final cut score recommendations reflect the performance levels associated with the Mod-MSA Math Grade 5 Test. | 4.52 | 5.00 | 3 | 5 |

# References

Huynh, H. (2006). A clarification on the response probability criterion RP67 for standard settings based on bookmark and item mapping. *Educational Measurement: Issues and Practice, 25,* 19–20.

Lee, G, & Lewis, D. (2008). A Generalizability Theory approach to standard error estimates for Bookmark standard settings. *Educational and Psychological Measurement, 68,* 603–620.

## Appendix A
## Agendas for the Mod-MSA Standard Setting Meetings

Modified Maryland School Assessment – Reading
Standard Setting Agenda

DAY 1 – May 10, 2010

| | | |
|---|---|---|
| **Registration** | 8:00-8:30 | Large Group |
| **Opening Remarks** | 8:30-9:15 | Large Group |
| Welcome and Why You Are Here | | |
| Introductions | | |
| Review of Agenda | | |
| Administrative Tasks | | |
| Panelist Information | | |
| **Overview of Standard Setting** | 9:15-9:45 | Large Group |
| Purpose | | |
| Item Mapping Methodology | | |
| **Overview of the Mod-MSA Tests** | 9:45-10:15 | Large Group |
| History | | |
| Purposes | | |
| Test Specifications | | |
| **BREAK** | 10:15-10:30 | |
| **Complete Mod-MSA Test** | 10:30-11:30 | Grade Group |
| **Review Performance Level Descriptors** | 11:30-12:00 | Grade Group |
| Create Behavioral Anchors | | |
| **LUNCH** | 12:00-1:00 | |
| Table Leader Training | | |
| **Review Performance Level Descriptors** | 1:00-2:00 | Grade Group |
| Create Behavioral Anchors | | |

| | | |
|---|---|---|
| **Overview of Standard Setting** | 2:00-2:30 | Grade Group |
| Item Mapping | | |
| Ordered Item Booklet | | |
| Item Map | | |
| Ratings Forms | | |
| **Practice Round** | 2:30-3:00 | Grade Group |
| **BREAK** | 3:00-3:15 | |
| **Round 1 Standard Setting** | 3:15-4:15 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |
| **End of Day Activities** | | |
| Review Day 2 Schedule | | |
| Check in materials | | |
| **END OF DAY 1** | | |
| DAY 2 – May 11, 2010 | | |
| **Breakfast** | 8:00-8:30 | Large Group |
| **Review schedule, answer questions** | 8:30-8:45 | Grade Group |
| **Round 1 Feedback** | 8:45-9:15 | Grade Group |
| Small group discussion of table agreement data | | |
| **Round 2 Standard Setting** | 9:15-10:15 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |

| | | |
|---|---|---|
| **BREAK** | 10:15-10:45 | |
| **Round 2 Feedback** | 10:45-11:15 | Grade Group |
| Small group discussion of table agreement data | | |
| Large-group discussion of group agreement data | | |
| Large-group discussion of impact data | | |
| **Round 3 Standard Setting** | 11:15-12:00 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |
| **LUNCH** | 12:00-1:00 | |
| **Round 3 Feedback** | 1:00-1:15 | Grade Group |
| **End of Day Activities** | 1:15-2:00 | Grade Group |
| Complete Evaluations | | |
| Check in materials | | |

**END OF DAY 2**

Modified Maryland School Assessment – Math

Standard Setting Agenda

DAY 1 – May 12, 2010

| | | |
|---|---|---|
| **Registration** | 8:00-8:30 | Large Group |
| **Opening Remarks** | 8:30-9:15 | Large Group |
| Welcome and Why You Are Here | | |
| Introductions | | |
| Review of Agenda | | |
| Administrative Tasks | | |
| Panelist Information | | |
| **Overview of Standard Setting** | 9:15-9:45 | Large Group |
| Purpose | | |
| Item Mapping Methodology | | |
| **Overview of the Mod-MSA Tests** | 9:45-10:15 | Large Group |
| History | | |
| Purposes | | |
| Test Specifications | | |
| **BREAK** | 10:15-10:30 | |
| **Complete Mod-MSA Test** | 10:30-11:30 | Grade Group |
| **Review Performance Level Descriptors** | 11:30-12:00 | Grade Group |
| Create Behavioral Anchors | | |
| **LUNCH** | 12:00-1:00 | |
| Table Leader Training | | |
| **Review Performance Level Descriptors** | 1:00-2:00 | Grade Group |
| Create Behavioral Anchors | | |

| | | |
|---|---|---|
| **Overview of Standard Setting** | 2:00-2:30 | Grade Group |
| Item Mapping | | |
| Ordered Item Booklet | | |
| Item Map | | |
| Ratings Forms | | |
| **Practice Round** | 2:30-3:00 | Grade Group |
| **BREAK** | 3:00-3:15 | |
| **Round 1 Standard Setting** | 3:15-4:15 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |
| **End of Day Activities** | | |
| Review Day 2 Schedule | | |
| Check in materials | | |
| **END OF DAY 1** | | |
| DAY 2 – May 13, 2010 | | |
| **Breakfast** | 8:00-8:30 | Large Group |
| **Review schedule, answer questions** | 8:30-8:45 | Grade Group |
| **Round 1 Feedback** | 8:45-9:15 | Grade Group |
| Small group discussion of table agreement data | | |
| **Round 2 Standard Setting** | 9:15-10:15 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |
| **BREAK** | 10:15-10:45 | |

| | | |
|---|---|---|
| **Round 2 Feedback** | 10:45-11:15 | Grade Group |
| Small group discussion of table agreement data | | |
| Large-group discussion of group agreement data | | |
| Large-group discussion of impact data | | |
| **Round 3 Standard Setting** | 11:15-12:00 | Grade Group |
| Readiness Form | | |
| Review Method | | |
| Collect page number/item numbers | | |
| **LUNCH** | 12:00-1:00 | |
| **Round 3 Feedback** | 1:00-1:15 | Grade Group |
| **End of Day Activities** | 1:15-2:00 | Grade Group |
| Complete Evaluations | | |
| Check in materials | | |

**END OF DAY 2**

## Appendix B

## Panelist Information Sheet

### Mod-MSA Grade 3 Reading Standard Setting

### Panelist Information Sheet

Judge ID: _____

Please provide the following demographic information that will be used to describe the general characteristics of the panelists who are recommending standards for the Mod-MSA Test.

Your Current Position:

---

---

Courses / Grades Taught / Educational Experience (e.g., teaching experience):

---

---

Gender (circle one): Male Female

Ethnicity:

---

Years of Educational Experience (e.g., years teaching):

---

Compared to other school districts in Maryland, how would you describe the size of your district (circle one)?

Large Medium Small

Compared to other school districts in Maryland, how would you describe the location of your district (circle one)?

Urban Suburban Rural

Compared to other school districts in Maryland, how would you describe the geographic location of your district (circle one)?

North South East West Central

## Appendix C

## Item Map for the Reading Grade 3 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000101529 | 49 | 209 | 3A3c | 2 | 36 | 0.83 |
| **2** | 100000213633 | 3 | 208 | 1D3a | 4 | 39 | 0.8 |
| **3** | 100000101905 | 42 | 208 | 1E4b | 3 | 43 | 0.74 |
| **4** | 100000346450 | 32 | 212 | 2A4c | 3 | 43 | 0.74 |
| **5** | 100000213631 | 1 | 208 | 1D3a | 3 | 44 | 0.73 |
| **6** | 100000260365 | 68 | 209 | 3A7b | 3 | 44 | 0.73 |
| **7** | 100000101908 | 45 | 209 | 3A8b | 2 | 46 | 0.71 |
| **8** | 100000101911 | 48 | 209 | 3A6a | 2 | 46 | 0.7 |
| **9** | 100000101533 | 55 | 209 | 3A7b | 3 | 47 | 0.69 |
| **10** | 100000101969 | 14 | 208 | 1E4c | 3 | 47 | 0.69 |
| **11** | 100000260368 | 64 | 209 | 3A6a | 2 | 48 | 0.67 |
| **12** | 100000101530 | 50 | 209 | 3A3b | 1 | 50 | 0.65 |
| **13** | 100000300552 | 69 | 208 | 1E4c | 2 | 50 | 0.65 |
| **14** | 100000260338 | 56 | 209 | 3A3d | 2 | 50 | 0.65 |
| **15** | 100000101937 | 26 | 212 | 2A5a | 1 | 51 | 0.63 |
| **16** | 100000101514 | 8 | 208 | 1B1a | 3 | 52 | 0.62 |
| **17** | 100000346452 | 28 | 208 | 1E4c | 3 | 53 | 0.6 |
| **18** | 100000101527 | 53 | 208 | 1E4c | 2 | 53 | 0.6 |
| **19** | 100000260460 | 36 | 212 | 2A4g | 3 | 54 | 0.59 |
| **20** | 100000101932 | 21 | 208 | 1E4d | 1 | 54 | 0.59 |
| **21** | 100000260465 | 41 | 208 | 1E4c | 1 | 54 | 0.58 |
| **22** | 100000101909 | 46 | 209 | 3A7b | 1 | 54 | 0.58 |
| **23** | 100000360183 | 9 | 208 | 1B1a | 1 | 55 | 0.58 |
| **24** | 100000260339 | 61 | 209 | 3A7c | 1 | 55 | 0.57 |
| **25** | 100000260461 | 37 | 212 | 2A4b | 2 | 55 | 0.56 |
| **26** | 100000101518 | 12 | 208 | 1D3b | 3 | 56 | 0.56 |
| **27** | 100000101513 | 7 | 208 | 1B1a | 2 | 56 | 0.56 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **28** | 100000101528 | 54 | 208 | 1E4c | 1 | 58 | 0.53 |
| **29** | 100000101936 | 25 | 212 | 2A4a | 3 | 58 | 0.52 |
| **30** | 100000101934 | 22 | 212 | 2A4a | 3 | 58 | 0.52 |
| **31** | 100000260457 | 39 | 212 | 2A3a | 3 | 58 | 0.51 |
| **32** | 100000260371 | 65 | 209 | 3A2b | 3 | 59 | 0.51 |
| **33** | 100000260342 | 57 | 209 | 3A3e | 3 | 59 | 0.51 |
| **34** | 100000101970 | 15 | 208 | 1E4c | 1 | 59 | 0.5 |
| **35** | 100000346448 | 34 | 212 | 2A2b | 2 | 60 | 0.49 |
| **36** | 100000101907 | 44 | 209 | 3A3d | 1 | 60 | 0.48 |
| **37** | 100000101938 | 27 | 212 | 2A2b | 2 | 61 | 0.47 |
| **38** | 100000101532 | 52 | 209 | 3A6a | 3 | 62 | 0.46 |
| **39** | 100000346445 | 33 | 212 | 2A5a | 1 | 62 | 0.46 |
| **40** | 100000101972 | 17 | 212 | 2A4h | 3 | 62 | 0.46 |
| **41** | 100000101975 | 20 | 212 | 2A3b | 3 | 63 | 0.45 |
| **42** | 100000346444 | 30 | 212 | 2A2d | 1 | 63 | 0.44 |
| **43** | 100000101974 | 19 | 212 | 2A5a | 2 | 64 | 0.43 |
| **44** | 100000360182 | 5 | 208 | 1D3a | 2 | 64 | 0.42 |
| **45** | 100000101516 | 10 | 208 | 1D3b | 3 | 65 | 0.42 |
| **46** | 100000260458 | 35 | 212 | 2A4c | 3 | 65 | 0.42 |
| **47** | 100000300707 | 62 | 208 | 1E4a | 2 | 65 | 0.4 |
| **48** | 100000300557 | 66 | 209 | 3A3e | 1 | 66 | 0.39 |
| **49** | 100000260364 | 63 | 209 | 3A8b | 1 | 67 | 0.38 |
| **50** | 100000213634 | 4 | 208 | 1D3a | 3 | 67 | 0.38 |
| **51** | 100000101973 | 18 | 212 | 2A6e | 1 | 68 | 0.37 |
| **52** | 100000101517 | 11 | 208 | 1D3b | 2 | 71 | 0.32 |
| **53** | 100000260345 | 58 | 209 | 3A2b | 2 | 71 | 0.32 |

## Appendix D

## Item Map for the Reading Grade 4 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000213641 | 10 | 208 | 1D3a | 1 | 40 | 0.78 |
| **2** | 100000357133 | 15 | 208 | 1E4c | 3 | 43 | 0.74 |
| **3** | 100000357109 | 25 | 212 | 2A5a | 3 | 44 | 0.73 |
| **4** | 100000360192 | 12 | 208 | 1D3a | 1 | 44 | 0.73 |
| **5** | 100000301034 | 65 | 209 | 3A3d | 3 | 47 | 0.69 |
| **6** | 100000102024 | 59 | 208 | 1E4c | 3 | 47 | 0.69 |
| **7** | 100000213637 | 6 | 208 | 1D3a | 3 | 47 | 0.69 |
| **8** | 100000213638 | 7 | 208 | 1D3a | 3 | 47 | 0.69 |
| **9** | 100000357134 | 13 | 212 | 2A4c | 1 | 49 | 0.67 |
| **10** | 100000213642 | 11 | 208 | 1D3a | 4 | 50 | 0.66 |
| **11** | 100000357104 | 21 | 208 | 1E4b | 2 | 50 | 0.65 |
| **12** | 100000260489 | 31 | 212 | 2A4i | 3 | 51 | 0.64 |
| **13** | 100000267470 | 63 | 208 | 1E4b | 1 | 51 | 0.63 |
| **14** | 100000357105 | 23 | 208 | 1E4c | 1 | 52 | 0.63 |
| **15** | 100000267472 | 68 | 209 | 3A2b | 3 | 53 | 0.61 |
| **16** | 100000357107 | 22 | 212 | 2A4e | 2 | 53 | 0.61 |
| **17** | 100000357101 | 46 | 209 | 3A4b | 1 | 54 | 0.59 |
| **18** | 100000213647 | 4 | 208 | 1D3b | 1 | 54 | 0.59 |
| **19** | 100000462157 | 47 | 209 | 3A4b | 3 | 55 | 0.58 |
| **20** | 100000102029 | 61 | 209 | 3A2b | 2 | 55 | 0.58 |
| **21** | 100000260483 | 29 | 208 | 1E4e | 2 | 56 | 0.57 |
| **22** | 100000357136 | 17 | 212 | 2A5a | 3 | 56 | 0.57 |
| **23** | 100000357137 | 18 | 212 | 2A3a | 1 | 56 | 0.56 |
| **24** | 100000269896 | 34 | 212 | 2A4c | 3 | 57 | 0.55 |
| **25** | 100000102027 | 57 | 209 | 3A3e | 2 | 58 | 0.53 |
| **26** | 100000101997 | 48 | 209 | 3A3a | 3 | 59 | 0.52 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **27** | 100000102000 | 52 | 209 | 3A3b | 2 | 59 | 0.52 |
| **28** | 100000357106 | 20 | 212 | 2A4i | 3 | 59 | 0.52 |
| **29** | 100000200070 | 54 | 208 | 1E4d | 1 | 60 | 0.51 |
| **30** | 100000267473 | 66 | 209 | 3A3c | 2 | 60 | 0.50 |
| **31** | 100000213645 | 2 | 208 | 1D3b | 1 | 61 | 0.50 |
| **32** | 100000357108 | 24 | 212 | 2A4d | 1 | 61 | 0.49 |
| **33** | 100000102001 | 53 | 209 | 3A7b | 1 | 61 | 0.49 |
| **34** | 100000102023 | 58 | 208 | 1E4b | 2 | 62 | 0.47 |
| **35** | 100000213646 | 3 | 208 | 1D3b | 3 | 63 | 0.47 |
| **36** | 100000269897 | 38 | 212 | 2A6e | 3 | 63 | 0.46 |
| **37** | 100000102028 | 60 | 209 | 3A8b | 1 | 63 | 0.46 |
| **38** | 100000357132 | 14 | 208 | 1E4c | 2 | 64 | 0.45 |
| **39** | 100000357135 | 16 | 212 | 2A4g | 2 | 64 | 0.45 |
| **40** | 100000102026 | 56 | 209 | 3A6a | 3 | 65 | 0.44 |
| **41** | 100000101996 | 49 | 209 | 1E4b | 3 | 65 | 0.43 |
| **42** | 100000101999 | 51 | 209 | 3A7b | 1 | 66 | 0.42 |
| **43** | 100000269899 | 40 | 212 | 2A3a | 1 | 66 | 0.42 |
| **44** | 100000357138 | 19 | 212 | 2A2f | 2 | 67 | 0.41 |
| **45** | 100000357100 | 45 | 209 | 3A6a | 3 | 68 | 0.40 |
| **46** | 100000271197 | 35 | 208 | 1E4b | 1 | 69 | 0.38 |
| **47** | 100000213644 | 1 | 208 | 1D3b | 1 | 70 | 0.36 |
| **48** | 100000269900 | 37 | 212 | 2A4g | 1 | 72 | 0.33 |
| **49** | 100000357098 | 43 | 209 | 3A3f | 2 | 73 | 0.33 |
| **50** | 100000102025 | 55 | 209 | 3A3d | 1 | 75 | 0.30 |
| **51** | 100000260492 | 33 | 212 | 2A6e | 1 | 77 | 0.28 |
| **52** | 100000260488 | 32 | 212 | 2A4h | 1 | 78 | 0.26 |
| **53** | 100000260486 | 28 | 212 | 2A4g | 1 | 79 | 0.25 |

## Appendix E

## Item Map for the Reading Grade 5 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000213656 | 7 | 208 | 1D3a | 1 | 40 | 0.78 |
| **2** | 100000102093 | 13 | 208 | 1E4e | 1 | 44 | 0.73 |
| **3** | 100000213655 | 6 | 208 | 1D3a | 2 | 44 | 0.73 |
| **4** | 100000213653 | 4 | 208 | 1D3a | 4 | 44 | 0.73 |
| **5** | 100000102112 | 24 | 208 | 1E4a | 3 | 45 | 0.72 |
| **6** | 100000102111 | 20 | 208 | 1E4b | 3 | 46 | 0.70 |
| **7** | 100000102050 | 43 | 209 | 3A6a | 3 | 47 | 0.70 |
| **8** | 100000213651 | 2 | 208 | 1D3a | 4 | 47 | 0.69 |
| **9** | 100000213657 | 8 | 208 | 1D3a | 3 | 47 | 0.69 |
| **10** | 100000213652 | 3 | 208 | 1D3a | 2 | 48 | 0.68 |
| **11** | 100000102067 | 57 | 208 | 1E4c | 2 | 49 | 0.67 |
| **12** | 100000213650 | 1 | 208 | 1D3a | 1 | 49 | 0.67 |
| **13** | 100000213659 | 10 | 208 | 1D3a | 1 | 50 | 0.65 |
| **14** | 100000102106 | 30 | 212 | 2A4b | 3 | 51 | 0.65 |
| **15** | 100000102052 | 45 | 209 | 3A7b | 2 | 52 | 0.63 |
| **16** | 100000360197 | 11 | 208 | 1D3a | 2 | 52 | 0.62 |
| **17** | 100000102084 | 34 | 208 | 1E4b | 3 | 52 | 0.62 |
| **18** | 100000102072 | 60 | 209 | 3A7c | 3 | 53 | 0.62 |
| **19** | 100000102047 | 46 | 208 | 1E4c | 2 | 53 | 0.61 |
| **20** | 100000102059 | 49 | 209 | 3A3f | 1 | 54 | 0.60 |
| **21** | 100000303033 | 67 | 209 | 3A7a | 3 | 55 | 0.59 |
| **22** | 100000102068 | 54 | 209 | 3A3a | 2 | 56 | 0.58 |
| **23** | 100000102048 | 41 | 209 | 3A3b | 1 | 58 | 0.55 |
| **24** | 100000102066 | 56 | 208 | 1E4b | 3 | 58 | 0.54 |
| **25** | 100000102104 | 26 | 212 | 2A4h | 1 | 59 | 0.53 |
| **26** | 100000102114 | 21 | 212 | 2A4b | 1 | 60 | 0.51 |
| **27** | 100000102095 | 12 | 212 | 2A4a | 3 | 60 | 0.51 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **28** | 100000102113 | 19 | 212 | 2A2d | 1 | 60 | 0.51 |
| **29** | 100000102096 | 15 | 212 | 2A4g | 2 | 61 | 0.50 |
| **30** | 100000102116 | 23 | 212 | 2A5a | 2 | 61 | 0.50 |
| **31** | 100000102061 | 51 | 209 | 3A8b | 1 | 61 | 0.50 |
| **32** | 100000213658 | 9 | 208 | 1D3a | 2 | 62 | 0.49 |
| **33** | 100000267477 | 61 | 209 | 3A3e | 1 | 63 | 0.47 |
| **34** | 100000102051 | 44 | 209 | 3A8b | 1 | 63 | 0.47 |
| **35** | 100000102056 | 47 | 208 | 1E4c | 2 | 63 | 0.47 |
| **36** | 100000102071 | 59 | 209 | 3A3b | 3 | 63 | 0.47 |
| **37** | 100000268380 | 64 | 209 | 3A7b | 2 | 66 | 0.43 |
| **38** | 100000102087 | 35 | 212 | 2A6b | 2 | 66 | 0.42 |
| **39** | 100000102108 | 32 | 212 | 2A3a | 2 | 68 | 0.40 |
| **40** | 100000102057 | 52 | 208 | 1E4b | 2 | 69 | 0.39 |
| **41** | 100000102090 | 39 | 212 | 2A3a | 1 | 69 | 0.38 |
| **42** | 100000102098 | 17 | 212 | 2A5a | 2 | 70 | 0.37 |
| **43** | 100000102069 | 55 | 209 | 3A6a | 1 | 70 | 0.37 |
| **44** | 100000102060 | 50 | 209 | 3A6a | 3 | 71 | 0.36 |
| **45** | 100000102117 | 25 | 212 | 2A3a | 2 | 72 | 0.35 |
| **46** | 100000102099 | 18 | 212 | 2A3a | 1 | 72 | 0.34 |
| **47** | 100000102115 | 22 | 212 | 2A4i | 3 | 72 | 0.34 |
| **48** | 100000102107 | 31 | 212 | 2A6e | 3 | 73 | 0.33 |
| **49** | 100000102094 | 14 | 208 | 1E4b | 1 | 74 | 0.32 |
| **50** | 100000268378 | 62 | 209 | 3A6a | 1 | 75 | 0.30 |
| **51** | 100000102088 | 36 | 212 | 2A4g | 1 | 76 | 0.30 |
| **52** | 100000267481 | 65 | 209 | 3A6c | 1 | 77 | 0.29 |

## Appendix F

## Item Map for the Math Grade 3 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000197601 | 1 | 246 | 2A1a | 3 | 40 | 0.77 |
| **2** | 100000185313 | 64 | 268 | 6A1a | 3 | 42 | 0.75 |
| **3** | 100000197651 | 31 | 246 | 2E2a | 3 | 43 | 0.74 |
| **4** | 100000197660 | 56 | 246 | 2E2a | 1 | 46 | 0.71 |
| **5** | 100000098452 | 40 | 241 | 1B2b | 2 | 46 | 0.70 |
| **6** | 100000185378 | 48 | 268 | 6A2a | 1 | 47 | 0.69 |
| **7** | 100000098454 | 46 | 241 | 1C1a | 2 | 47 | 0.69 |
| **8** | 100000098440 | 19 | 241 | 1A1c | 3 | 48 | 0.68 |
| **9** | 100000197649 | 61 | 273 | 7 | 1 | 49 | 0.66 |
| **10** | 100000197756 | 57 | 257 | 4B1c | 2 | 51 | 0.64 |
| **11** | 100000185384 | 79 | 268 | 6C1c | 2 | 52 | 0.63 |
| **12** | 100000098449 | 52 | 241 | 1B2b | 1 | 52 | 0.63 |
| **13** | 100000185486 | 77 | 273 | 7 | 2 | 52 | 0.63 |
| **14** | 100000185381 | 74 | 268 | 6A3b | 1 | 52 | 0.62 |
| **15** | 100000197621 | 55 | 246 | 2D1a | 2 | 53 | 0.61 |
| **16** | 100000197665 | 66 | 251 | 3C1b | 2 | 54 | 0.60 |
| **17** | 100000197761 | 50 | 273 | 7 | 2 | 54 | 0.59 |
| **18** | 100000098445 | 4 | 241 | 1A2b | 2 | 55 | 0.59 |
| **19** | 100000098516 | 6 | 268 | 6A1d | 1 | 55 | 0.58 |
| **20** | 100000197661 | 91 | 251 | 3A1b | 1 | 56 | 0.57 |
| **21** | 100000350878 | 29 | 241 | 1C1b | 3 | 57 | 0.56 |
| **22** | 100000197647 | 18 | 246 | 2E1a | 3 | 58 | 0.55 |
| **23** | 100000197723 | 9 | 257 | 4A1c | 2 | 58 | 0.54 |
| **24** | 100000197670 | 37 | 251 | 3C1b | 1 | 58 | 0.54 |
| **25** | 100000098515 | 92 | 268 | 6A1c | 2 | 59 | 0.53 |
| **26** | 100000197781 | 12 | 262 | 5B1a | 2 | 59 | 0.53 |
| **27** | 100000098435 | 75 | 241 | 1A1a | 3 | 59 | 0.52 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **28** | 100000197674 | 38 | 273 | 7 | 1 | 59 | 0.52 |
| **29** | 100000197602 | 54 | 246 | 2A1b | 1 | 59 | 0.52 |
| **30** | 100000197604 | 2 | 246 | 2A1c | 3 | 61 | 0.50 |
| **31** | 100000197724 | 71 | 257 | 4B1a | 1 | 61 | 0.50 |
| **32** | 100000197662 | 65 | 251 | 3A1c | 2 | 61 | 0.49 |
| **33** | 100000185485 | 76 | 273 | 7 | 1 | 62 | 0.49 |
| **34** | 100000185380 | 78 | 268 | 6A2b | 2 | 62 | 0.48 |
| **35** | 100000185376 | 89 | 268 | 6A1b | 1 | 63 | 0.46 |
| **36** | 100000197780 | 11 | 262 | 5B1a | 3 | 63 | 0.46 |
| **37** | 100000197675 | 39 | 273 | 7 | 2 | 64 | 0.46 |
| **38** | 100000098438 | 3 | 241 | 1A1a | 3 | 65 | 0.44 |
| **39** | 100000197720 | 81 | 257 | 4A1b | 2 | 65 | 0.44 |
| **40** | 100000197663 | 94 | 251 | 3B1a | 1 | 66 | 0.43 |
| **41** | 100000197722 | 28 | 257 | 4A1c | 3 | 66 | 0.42 |
| **42** | 100000185387 | 88 | 268 | 6C1d | 1 | 66 | 0.42 |
| **43** | 100000185382 | 63 | 268 | 6B1a | 3 | 66 | 0.42 |
| **44** | 100000197752 | 14 | 257 | 4B1b | 3 | 67 | 0.41 |
| **45** | 100000197677 | 26 | 257 | 4A1a | 3 | 67 | 0.41 |
| **46** | 100000185484 | 68 | 273 | 7 | 1 | 68 | 0.40 |
| **47** | 100000197650 | 62 | 273 | 7 | 2 | 68 | 0.39 |
| **48** | 100000197664 | 86 | 251 | 3C1a | 2 | 69 | 0.39 |
| **49** | 100000185403 | 8 | 273 | 7 | 2 | 69 | 0.38 |
| **50** | 100000197667 | 42 | 251 | 3C1a | 1 | 70 | 0.37 |
| **51** | 100000197676 | 85 | 251 | 3C2a | 1 | 70 | 0.36 |
| **52** | 100000098444 | 41 | 241 | 1A2a | 1 | 71 | 0.36 |
| **53** | 100000185401 | 7 | 273 | 7 | 2 | 71 | 0.35 |
| **54** | 100000098446 | 58 | 241 | 1B1a | 2 | 73 | 0.33 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **55** | 100000185473 | 84 | 273 | 7 | 2 | 73 | 0.33 |
| **56** | 100000197753 | 15 | 273 | 7 | 3 | 74 | 0.31 |
| **57** | 100000098527 | 73 | 268 | 6A3a | 3 | 75 | 0.31 |
| **58** | 100000098532 | 87 | 268 | 6C1b | 1 | 75 | 0.30 |
| **59** | 100000098447 | 47 | 241 | 1B1a | 2 | 76 | 0.29 |
| **60** | 100000197648 | 60 | 246 | 2E2a | 3 | 77 | 0.28 |
| **61** | 100000197754 | 16 | 273 | 7 | 1 | 78 | 0.27 |
| **62** | 100000197751 | 27 | 257 | 4A1b | 1 | 78 | 0.26 |

## Appendix G

## Item Map for the Math Grade 4 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000098585 | 57 | 241 | 1B2b | 1 | 39 | 0.79 |
| **2** | 100000198094 | 45 | 246 | 2B1b | 2 | 39 | 0.78 |
| **3** | 100000098584 | 11 | 241 | 1B2b | 2 | 40 | 0.78 |
| **4** | 100000198098 | 14 | 246 | 2B2b | 3 | 42 | 0.76 |
| **5** | 100000186578 | 74 | 268 | 6A2b | 1 | 44 | 0.73 |
| **6** | 100000186576 | 63 | 268 | 6C1f | 3 | 44 | 0.72 |
| **7** | 100000186577 | 92 | 268 | 6A1c | 1 | 45 | 0.72 |
| **8** | 100000098579 | 20 | 241 | 1A2b | 2 | 46 | 0.71 |
| **9** | 100000198150 | 65 | 262 | 5B1a | 1 | 46 | 0.70 |
| **10** | 100000198144 | 60 | 262 | 5B1a | 2 | 47 | 0.69 |
| **11** | 100000186562 | 43 | 273 | 7 | 2 | 49 | 0.67 |
| **12** | 100000098568 | 93 | 241 | 1A1a | 3 | 50 | 0.64 |
| **13** | 100000198140 | 22 | 262 | 5B1a | 3 | 51 | 0.64 |
| **14** | 100000186580 | 75 | 268 | 6A2f | 3 | 52 | 0.63 |
| **15** | 100000186581 | 82 | 268 | 6B1b | 3 | 53 | 0.61 |
| **16** | 100000207143 | 83 | 268 | 6C1c | 3 | 53 | 0.60 |
| **17** | 100000098571 | 81 | 241 | 1A1a | 2 | 54 | 0.60 |
| **18** | 100000198096 | 46 | 246 | 2B2a | 2 | 55 | 0.58 |
| **19** | 100000198099 | 28 | 246 | 2D1a | 3 | 55 | 0.58 |
| **20** | 100000198123 | 5 | 257 | 4B1a | 3 | 56 | 0.57 |
| **21** | 100000198113 | 2 | 273 | 7 | 1 | 56 | 0.57 |
| **22** | 100000198114 | 3 | 273 | 7 | 2 | 56 | 0.56 |
| **23** | 100000198125 | 89 | 257 | 4A1a | 1 | 57 | 0.55 |
| **24** | 100000198127 | 98 | 257 | 4B1a | 2 | 57 | 0.55 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **25** | 100000186567 | 96 | 268 | 6A3a | 1 | 57 | 0.54 |
| **26** | 100000198107 | 36 | 251 | 3A1a | 2 | 58 | 0.53 |
| **27** | 100000198142 | 23 | 273 | 7 | 3 | 58 | 0.53 |
| **28** | 100000186560 | 78 | 273 | 7 | 2 | 58 | 0.53 |
| **29** | 100000098666 | 64 | 268 | 6C1g | 1 | 59 | 0.53 |
| **30** | 100000098582 | 58 | 241 | 1B1b | 1 | 59 | 0.52 |
| **31** | 100000198111 | 1 | 251 | 3C1a | 2 | 59 | 0.52 |
| **32** | 100000098664 | 99 | 268 | 6B1c | 2 | 59 | 0.52 |
| **33** | 100000198143 | 24 | 273 | 7 | 2 | 59 | 0.51 |
| **34** | 100000198103 | 39 | 273 | 7 | 1 | 60 | 0.51 |
| **35** | 100000186579 | 101 | 268 | 6A2e | 1 | 60 | 0.50 |
| **36** | 100000198158 | 41 | 262 | 5B1a | 1 | 61 | 0.49 |
| **37** | 100000186574 | 68 | 273 | 7 | 2 | 61 | 0.49 |
| **38** | 100000186575 | 69 | 273 | 7 | 2 | 61 | 0.49 |
| **39** | 100000098586 | 52 | 241 | 1C1a | 2 | 61 | 0.48 |
| **40** | 100000198102 | 38 | 273 | 7 | 2 | 62 | 0.48 |
| **41** | 100000198092 | 27 | 246 | 2A1a | 1 | 62 | 0.48 |
| **42** | 100000198148 | 40 | 262 | 5B1a | 2 | 62 | 0.47 |
| **43** | 100000198128 | 26 | 257 | 4B1b | 3 | 63 | 0.46 |
| **44** | 100000198137 | 80 | 257 | 4B2a | 1 | 64 | 0.44 |
| **45** | 100000198108 | 71 | 251 | 3B1a | 2 | 64 | 0.44 |
| **46** | 100000098645 | 73 | 268 | 6A1d | 3 | 64 | 0.44 |
| **47** | 100000098587 | 34 | 241 | 1C1a | 3 | 65 | 0.42 |
| **48** | 100000098573 | 77 | 241 | 1A1b | 2 | 65 | 0.42 |
| **49** | 100000098578 | 6 | 241 | 1A2a | 2 | 66 | 0.42 |
| **50** | 100000198139 | 56 | 257 | 4B2a | 3 | 66 | 0.42 |
| **51** | 100000198121 | 84 | 251 | 3C1c | 2 | 67 | 0.41 |
| **52** | 100000098572 | 51 | 241 | 1A1a | 1 | 67 | 0.40 |
| **53** | 100000198101 | 37 | 246 | 2E1a | 1 | 68 | 0.39 |
| **54** | 100000186582 | 97 | 268 | 6C2a | 2 | 68 | 0.38 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **55** | 100000198122 | 72 | 251 | 3C2a | 1 | 69 | 0.37 |
| **56** | 100000098583 | 10 | 241 | 1B2a | 3 | 69 | 0.37 |
| **57** | 100000186573 | 67 | 268 | 6B1b | 3 | 70 | 0.36 |
| **58** | 100000186561 | 79 | 273 | 7 | 2 | 71 | 0.35 |
| **59** | 100000186566 | 9 | 273 | 7 | 3 | 72 | 0.33 |
| **60** | 100000098580 | 100 | 241 | 1B1a | 3 | 73 | 0.32 |
| **61** | 100000198157 | 76 | 262 | 5B1a | 3 | 73 | 0.32 |

## Appendix H

## Item Map for the Math Grade 5 Ordered Item Book

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **1** | 100000196100 | 72 | 246 | 2D1a | 2 | 36 | 0.80 |
| **2** | 100000196234 | 42 | 251 | 3C1b | 2 | 42 | 0.75 |
| **3** | 100000187390 | 18 | 268 | 6B1d | 3 | 43 | 0.73 |
| **4** | 100000099089 | 45 | 241 | 1B2b | 2 | 44 | 0.72 |
| **5** | 100000187386 | 22 | 268 | 6B1c | 1 | 46 | 0.70 |
| **6** | 100000099085 | 40 | 241 | 1B1c | 1 | 47 | 0.69 |
| **7** | 100000196036 | 25 | 273 | 7 | 2 | 48 | 0.68 |
| **8** | 100000196263 | 81 | 257 | 4B1b | 2 | 49 | 0.67 |
| **9** | 100000196233 | 8 | 251 | 3C1a | 2 | 50 | 0.65 |
| **10** | 100000099080 | 5 | 241 | 1A1c | 1 | 51 | 0.64 |
| **11** | 100000196253 | 14 | 257 | 4A1c | 2 | 51 | 0.64 |
| **12** | 100000196284 | 33 | 262 | 5B1a | 2 | 53 | 0.63 |
| **13** | 100000099079 | 38 | 241 | 1A1b | 3 | 53 | 0.63 |
| **14** | 100000099086 | 30 | 241 | 1B1c | 1 | 53 | 0.62 |
| **15** | 100000196025 | 85 | 273 | 7 | 3 | 55 | 0.60 |
| **16** | 100000187428 | 62 | 268 | 6C1f | 1 | 56 | 0.59 |
| **17** | 100000099072 | 65 | 241 | 1A1a | 2 | 58 | 0.56 |
| **18** | 100000187429 | 61 | 268 | 6C1g | 3 | 58 | 0.56 |
| **19** | 100000187376 | 21 | 268 | 6A1d | 3 | 59 | 0.55 |
| **20** | 100000196029 | 26 | 273 | 7 | 1 | 59 | 0.54 |
| **21** | 100000099091 | 46 | 241 | 1C1a | 3 | 60 | 0.54 |
| **22** | 100000099075 | 29 | 241 | 1A1a | 2 | 60 | 0.53 |
| **23** | 100000099083 | 28 | 241 | 1B1b | 3 | 61 | 0.53 |
| **24** | 100000196256 | 56 | 257 | 4A1d | 3 | 61 | 0.53 |
| **25** | 100000196045 | 13 | 273 | 7 | 2 | 61 | 0.52 |
| **26** | 100000196223 | 71 | 246 | 2E1a | 3 | 61 | 0.52 |
| **27** | 100000187361 | 67 | 273 | 7 | 2 | 62 | 0.51 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **28** | 100000196277 | 32 | 257 | 4B2a | 3 | 63 | 0.50 |
| **29** | 100000196057 | 34 | 273 | 7 | 3 | 64 | 0.48 |
| **30** | 100000099177 | 88 | 241 | 1C1b | 3 | 64 | 0.48 |
| **31** | 100000196054 | 35 | 273 | 7 | 2 | 65 | 0.47 |
| **32** | 100000187367 | 2 | 268 | 6A1a | 3 | 65 | 0.47 |
| **33** | 100000196094 | 73 | 246 | 2C1a | 2 | 65 | 0.47 |
| **34** | 100000187370 | 77 | 268 | 6A1b | 2 | 65 | 0.46 |
| **35** | 100000196229 | 10 | 251 | 3A1b | 2 | 66 | 0.46 |
| **36** | 100000187366 | 89 | 273 | 7 | 1 | 66 | 0.45 |
| **37** | 100000099180 | 64 | 268 | 6C1e | 3 | 67 | 0.45 |
| **38** | 100000099082 | 92 | 241 | 1B1a | 2 | 67 | 0.44 |
| **39** | 100000196043 | 12 | 273 | 7 | 2 | 68 | 0.43 |
| **40** | 100000196281 | 51 | 262 | 5A1a | 2 | 69 | 0.41 |
| **41** | 100000196247 | 82 | 257 | 4A1a | 3 | 69 | 0.41 |
| **42** | 100000187382 | 80 | 268 | 6B1b | 3 | 70 | 0.41 |
| **43** | 100000196079 | 84 | 246 | 2A1b | 1 | 70 | 0.40 |
| **44** | 100000187388 | 53 | 273 | 7 | 1 | 70 | 0.40 |
| **45** | 100000099090 | 6 | 241 | 1B2b | 3 | 71 | 0.40 |
| **46** | 100000196231 | 24 | 251 | 3B2a | 2 | 71 | 0.40 |
| **47** | 100000196042 | 93 | 273 | 7 | 1 | 71 | 0.40 |
| **48** | 100000187380 | 60 | 268 | 6B1a | 1 | 71 | 0.39 |
| **49** | 100000099081 | 37 | 241 | 1B1a | 2 | 71 | 0.39 |
| **50** | 100000196238 | 83 | 251 | 3C2a | 1 | 72 | 0.38 |
| **51** | 100000187360 | 66 | 273 | 7 | 1 | 72 | 0.38 |
| **52** | 100000187391 | 78 | 268 | 6C1a | 3 | 72 | 0.38 |
| **53** | 100000187363 | 75 | 273 | 7 | 1 | 73 | 0.36 |
| **54** | 100000187389 | 54 | 273 | 7 | 2 | 74 | 0.35 |

| Page | Item CID | Item Seq. No | Reporting Strand | Content Standard | Answer Key | Location | P-Value |
|---|---|---|---|---|---|---|---|
| **55** | 100000196270 | 55 | 257 | 4B1e | 2 | 75 | 0.35 |
| **56** | 100000196088 | 16 | 246 | 2B1a | 1 | 76 | 0.33 |
| **57** | 100000196258 | 69 | 257 | 4B1a | 3 | 77 | 0.32 |
| **58** | 100000196273 | 11 | 257 | 4B2a | 1 | 78 | 0.31 |
| **59** | 100000196090 | 19 | 246 | 2B2b | 2 | 79 | 0.29 |
| **60** | 100000196279 | 36 | 262 | 5A1a | 3 | 80 | 0.28 |
| **61** | 100000187381 | 99 | 268 | 6B1b | 1 | 83 | 0.26 |
| **62** | 100000187372 | 63 | 268 | 6A1c | 3 | 85 | 0.24 |
| **63** | 100000187387 | 52 | 268 | 6B1d | 1 | 87 | 0.22 |
| **64** | 100000196244 | 44 | 251 | 3C2b | 2 | 89 | 0.20 |

## Appendix I

## Panelist Readiness Survey

**Maryland Mod-MSA Grade 3 Reading**
**Standard Setting Readiness Survey**

Panelist ID: _____

Instructions: Please circle your response to the following questions.

| **Round 1** | | |
|---|---|---|
| I understand my task for Round 1. | **No** | **Yes** |
| I am ready to begin Round 1. | **No** | **Yes** |

| **Round 2** | | |
|---|---|---|
| I understand my task for Round 2. | **No** | **Yes** |
| I understand the panelist agreement data that was presented from Round 1. | **No** | **Yes** |
| I understand the item difficulty data that was provided. | **No** | **Yes** |
| I am ready to begin Round 2. | **No** | **Yes** |

| **Round 3** | | |
|---|---|---|
| I understand my task for Round 3. | **No** | **Yes** |
| I understand the impact data that was presented from Round 2. | **No** | **Yes** |
| I am ready to begin Round 3. | **No** | **Yes** |

## Appendix J

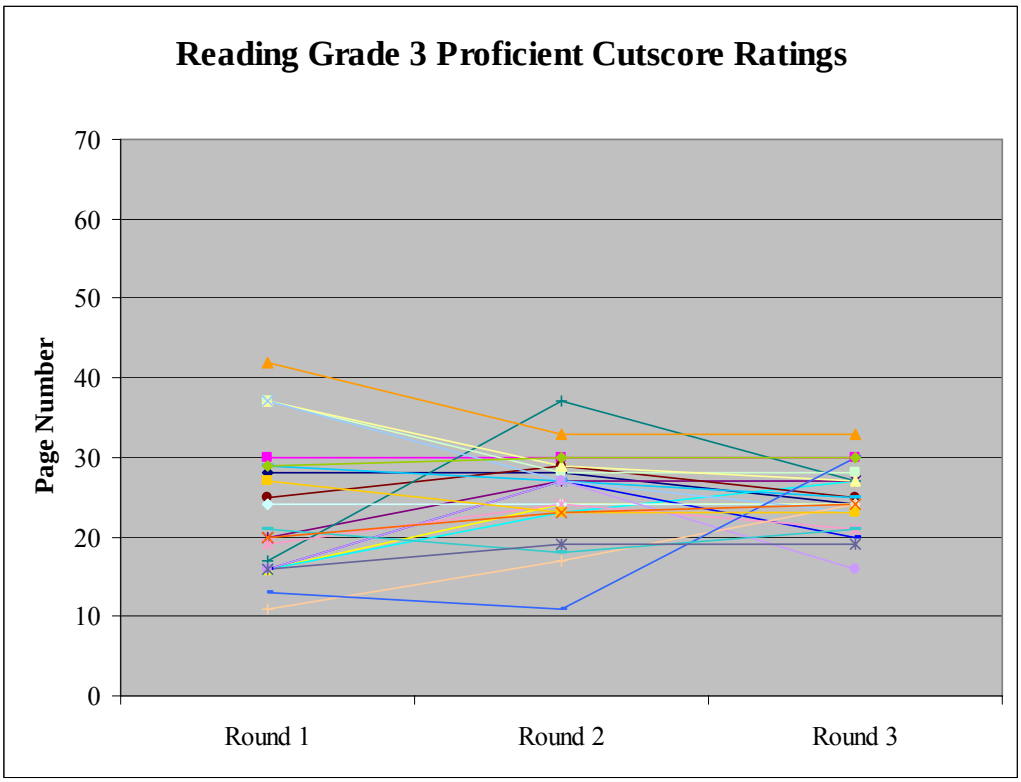
## Mean, Median, Minimum, and Maximum Ratings by Round

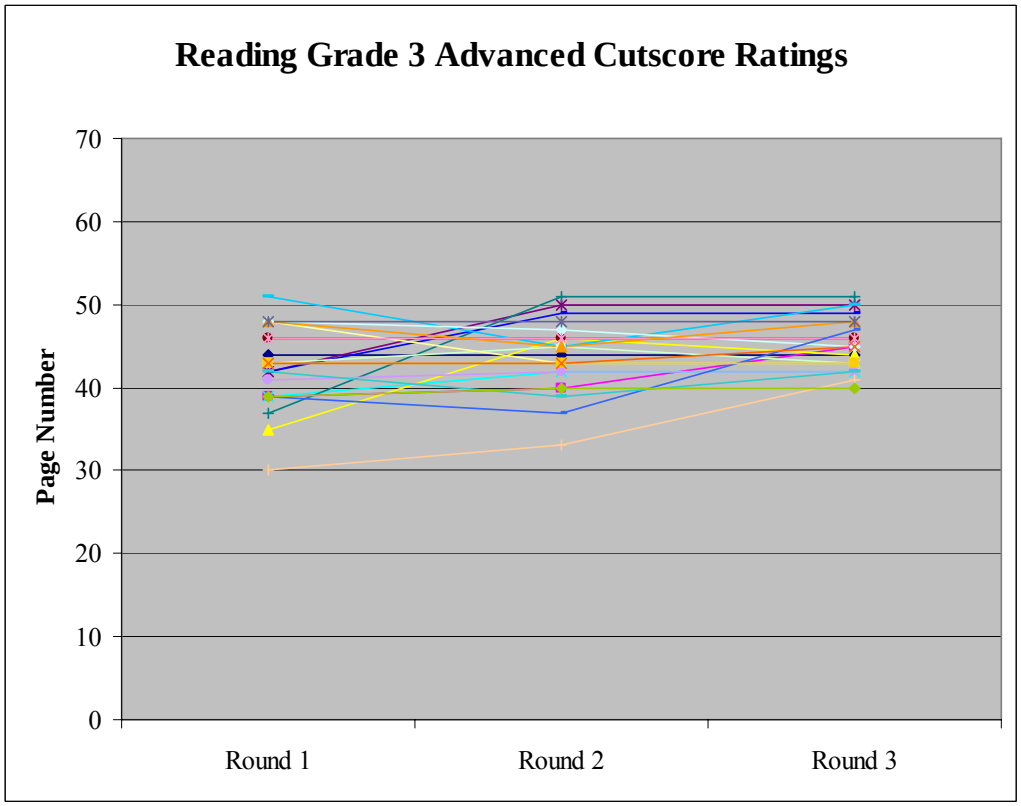
| | Reading Grade 3 | | | | | |
|---|---|---|---|---|---|---|
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 23.74 | 42.43 | 25.43 | 43.78 | 24.87 | 45.09 |
| Median | 21.00 | 43.00 | 27.00 | 44.00 | 24.00 | 45.00 |
| Minimum | 11.00 | 30.00 | 11.00 | 33.00 | 16.00 | 40.00 |
| Maximum | 42.00 | 51.00 | 37.00 | 51.00 | 33.00 | 51.00 |
| | **Reading Grade 4** | | | | | |
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 19.05 | 39.55 | 17.82 | 40.95 | 16.91 | 39.36 |
| Median | 20.00 | 43.00 | 18.00 | 43.00 | 18.00 | 42.00 |
| Minimum | 5.00 | 14.00 | 14.00 | 34.00 | 13.00 | 23.00 |
| Maximum | 29.00 | 48.00 | 24.00 | 48.00 | 20.00 | 48.00 |
| | **Reading Grade 5** | | | | | |
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 15.26 | 35.87 | 18.87 | 39.13 | 18.35 | 40.09 |
| Median | 17.00 | 38.00 | 18.00 | 41.00 | 18.00 | 41.00 |
| Minimum | 2.00 | 5.00 | 8.00 | 27.00 | 17.00 | 34.00 |
| Maximum | 25.00 | 49.00 | 32.00 | 49.00 | 35.00 | 44.00 |

| | Math Grade 3 | | | | | |
|---|---|---|---|---|---|---|
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 20.96 | 43.48 | 19.74 | 43.78 | 20.17 | 44.30 |
| Median | 20.00 | 44.00 | 19.00 | 45.00 | 19.00 | 44.00 |
| Minimum | 7.00 | 28.00 | 11.00 | 19.00 | 12.00 | 34.00 |
| Maximum | 38.00 | 60.00 | 28.00 | 58.00 | 28.00 | 50.00 |
| | **Math Grade 4** | | | | | |
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 24.91 | 48.41 | 22.09 | 50.41 | 18.86 | 52.68 |
| Median | 23.00 | 50.00 | 22.00 | 53.00 | 17.00 | 53.00 |
| Minimum | 16.00 | 31.00 | 17.00 | 35.00 | 16.00 | 48.00 |
| Maximum | 39.00 | 57.00 | 27.00 | 54.00 | 25.00 | 54.00 |
| | **Math Grade 5** | | | | | |
| Round | 1 | | 2 | | 3 | |
| Achievement Level | P | A | P | A | P | A |
| Mean | 20.95 | 50.57 | 20.29 | 50.43 | 18.52 | 48.76 |
| Median | 18.00 | 50.00 | 20.00 | 50.00 | 18.00 | 49.00 |
| Minimum | 9.00 | 41.00 | 15.00 | 46.00 | 13.00 | 43.00 |
| Maximum | 44.00 | 62.00 | 28.00 | 59.00 | 24.00 | 52.00 |

## Appendix K

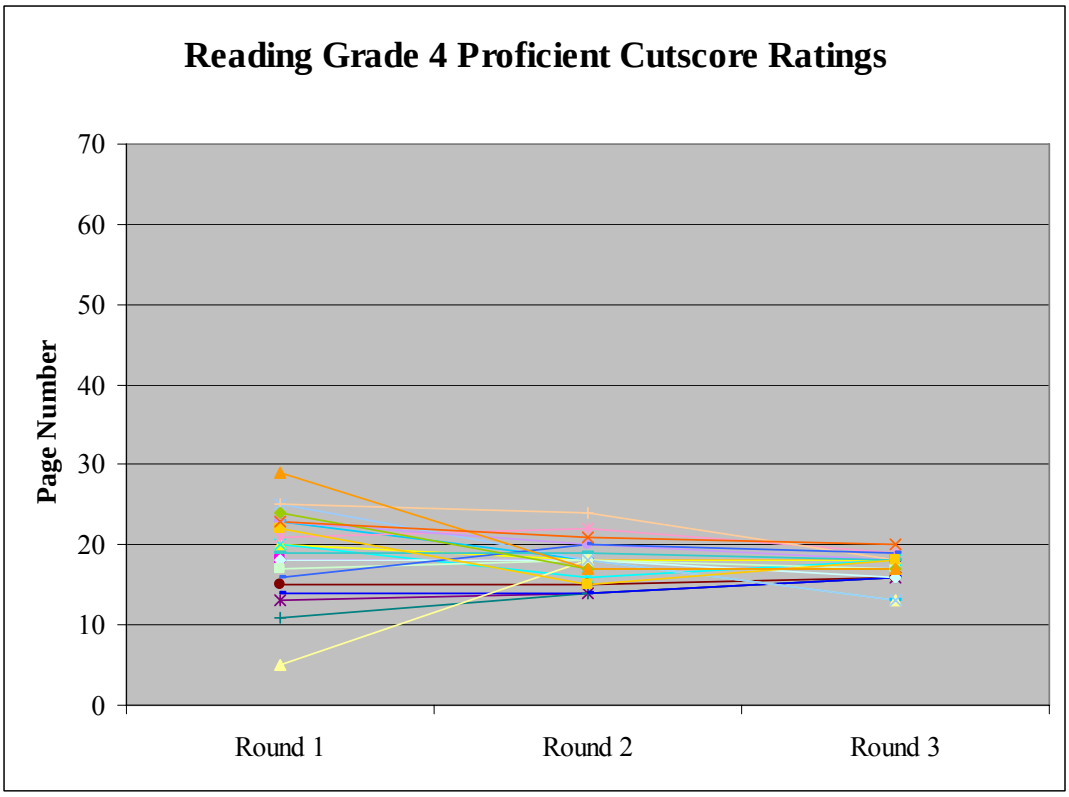
## Individual Ratings across Rounds for Reading Grade 3

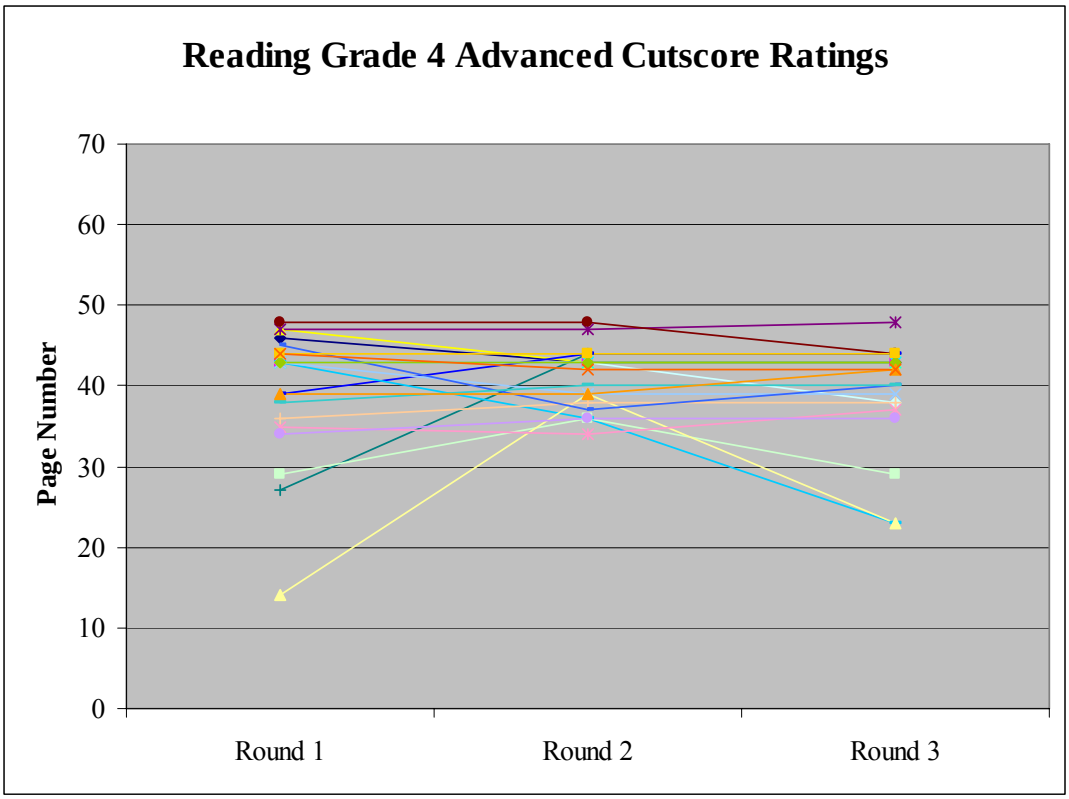


## Appendix L

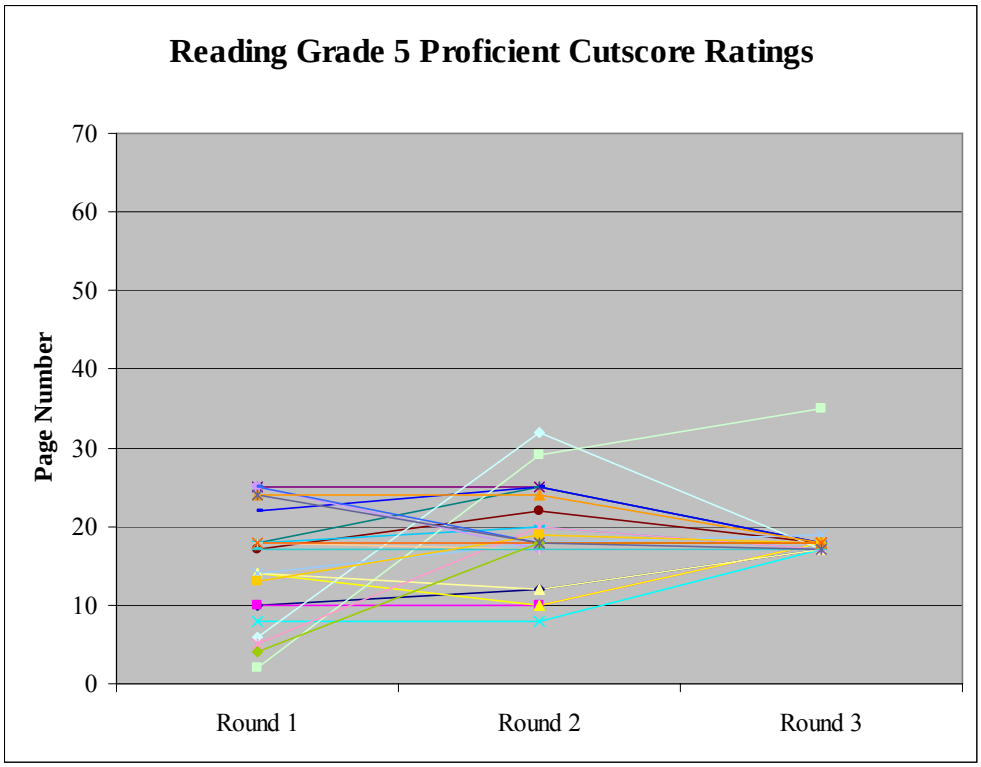
## Individual Ratings across Rounds for Reading Grade 4

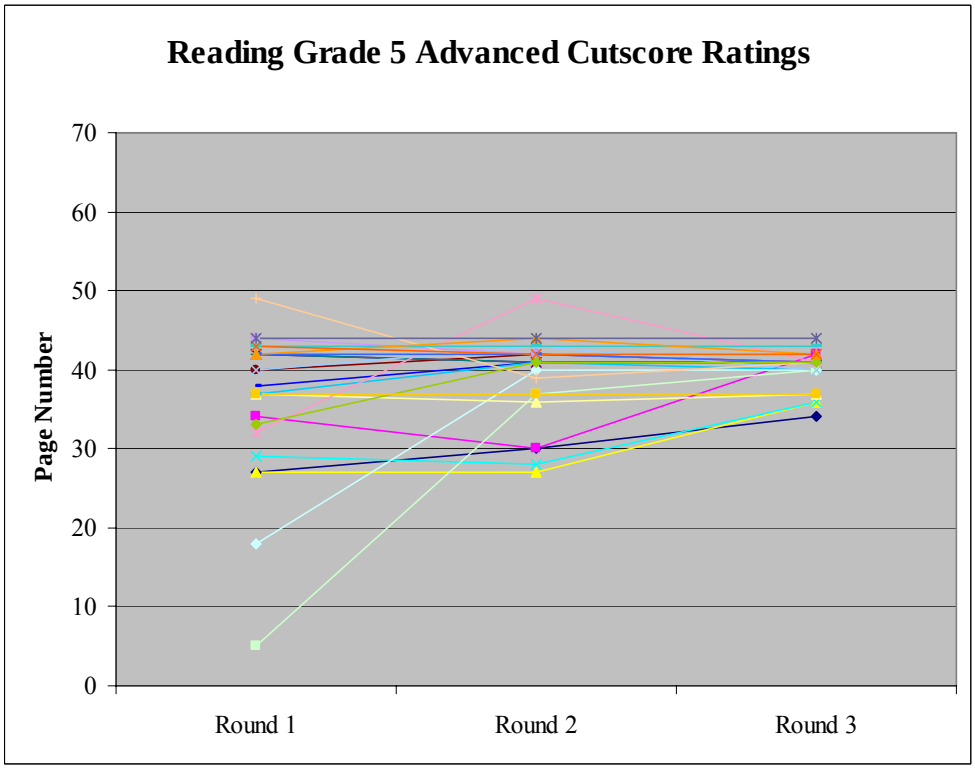



Reading Grade 4 Advanced Cutscore Ratings
Page Number
70
60
50
40
30
20
10
0
Round 1
Round 2
Round 3

## Appendix M

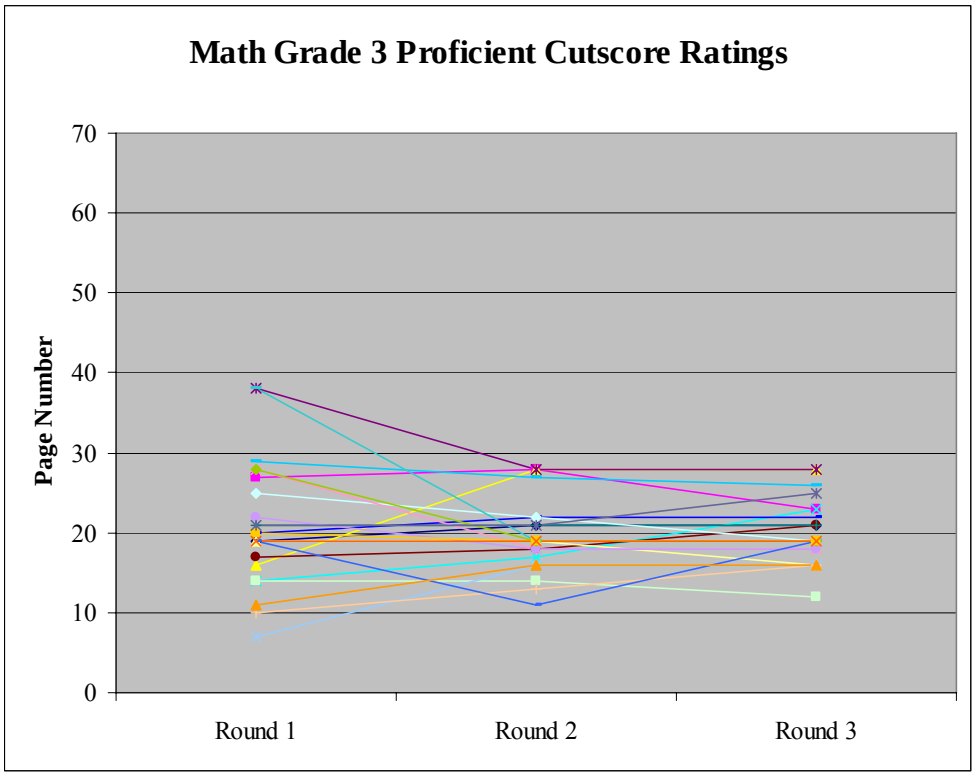
## Individual Ratings across Rounds for Reading Grade 5

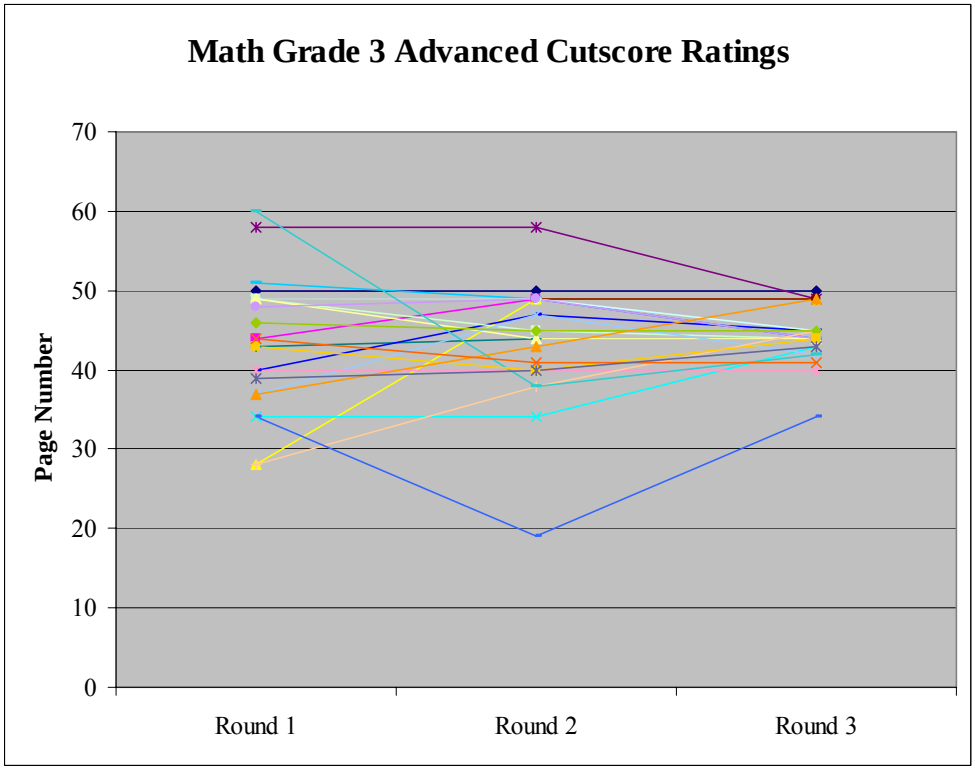


## Appendix N

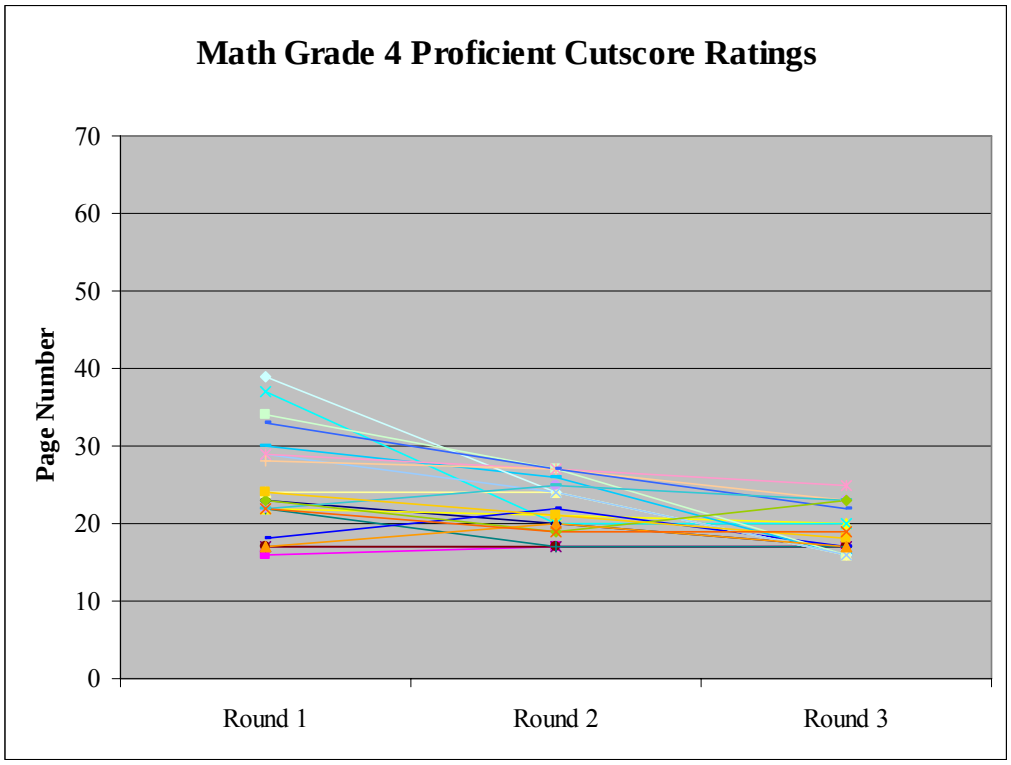
## Individual Ratings across Rounds for Math Grade 3

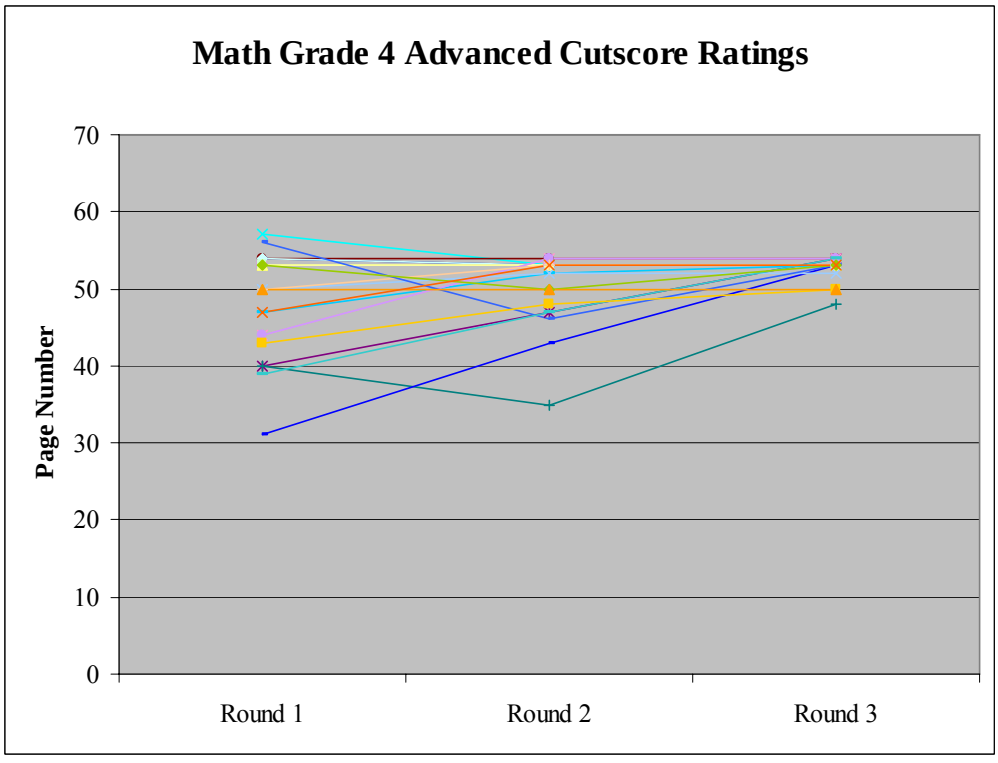


## Appendix O

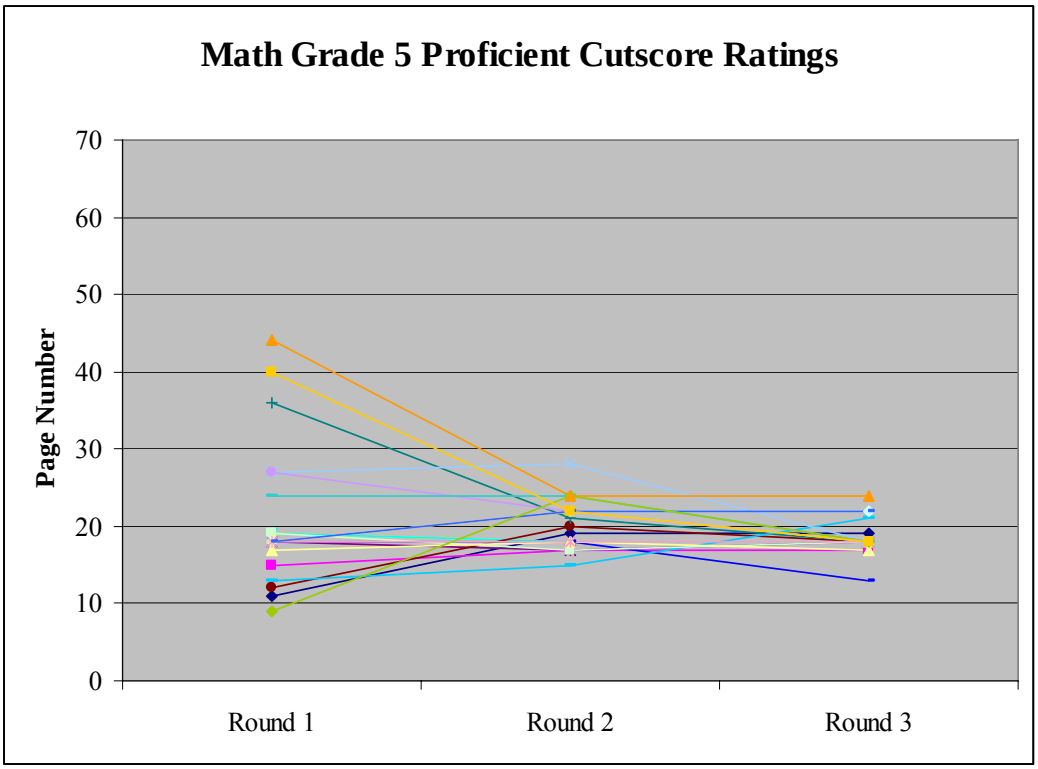
## Individual Ratings across Rounds for Math Grade 4

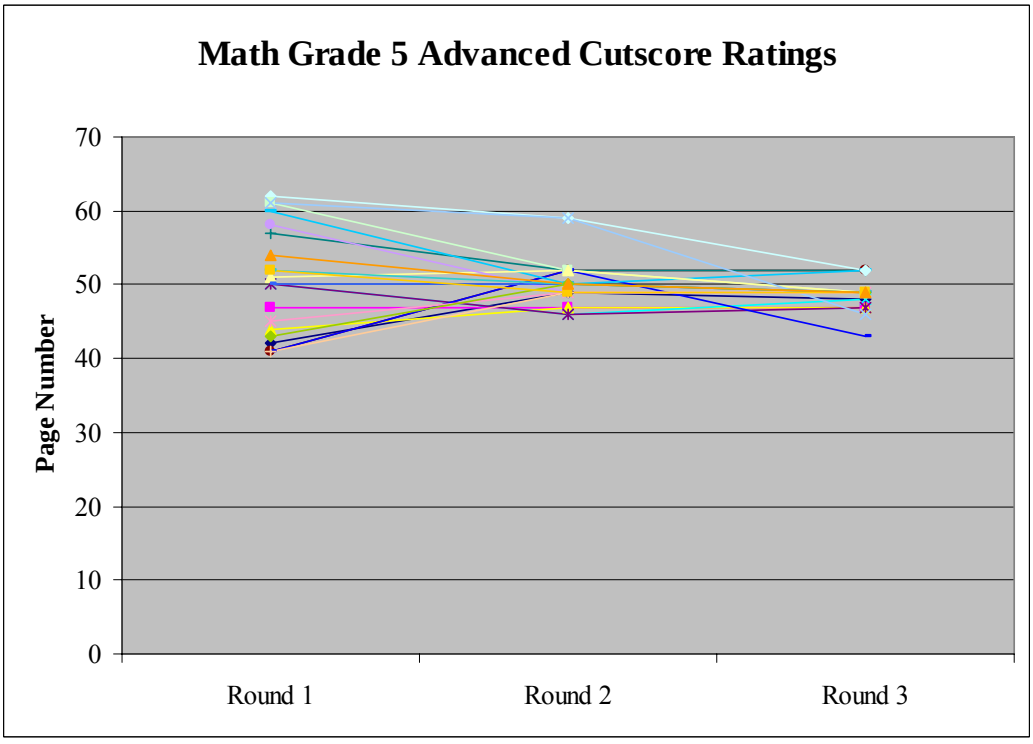


## Appendix P

## Individual Ratings across Rounds for Math Grade 5



# Appendix Q

## Standard Error Bands for Reading Grade 3

| Reading Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 0.46 | 1.26 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 56 | 69 | + 3 SEs |
| Percent of students in each Performance Level | 66.3 | 23.4 | 10.3 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 56 | 68 | + 2 SEs |
| Percent of students in each Performance Level | 66.3 | 23.4 | 10.3 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 55 | 66 | + 1 SEs |
| Percent of students in each Performance Level | 66.3 | 20.4 | 13.3 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 65 | |
| Percent of students in each Performance Level | 66.3 | 20.4 | 13.3 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 55 | 64 | - 1 SEs |
| Percent of students in each Performance Level | 66.3 | 17.0 | 16.7 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 54 | 62 | - 2 SEs |
| Percent of students in each Performance Level | 61.0 | 19.1 | 19.9 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 54 | 61 | - 3 SEs |
| Percent of students in each Performance Level | 61.0 | 19.1 | 19.9 | |
| *Large Group Medians | | | | |

| Reading Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 4.0 | 5.0 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 67 | 80 | + 3 SEs |
| Percent of students in each Performance Level | 89.7 | 4.1 | 6.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 63 | 75 | + 2 SEs |
| Percent of students in each Performance Level | 83.3 | 10.5 | 6.2 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 59 | 70 | + 1 SEs |
| Percent of students in each Performance Level | 74.8 | 14.9 | 10.3 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 65 | |
| Percent of students in each Performance Level | 66.3 | 20.4 | 13.3 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 51 | 60 | - 1 SEs |
| Percent of students in each Performance Level | 51.2 | 25.7 | 23.1 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 47 | 55 | - 2 SEs |
| Percent of students in each Performance Level | 37.2 | 29.1 | 33.7 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 43 | 50 | - 3 SEs |
| Percent of students in each Performance Level | 23.3 | 23.2 | 53.5 | |
| *Large Group Medians | | | | |

| Reading Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 4.03 | 5.16 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 67 | 80 | + 3 SEs |
| Percent of students in each Performance Level | 89.7 | 4.1 | 6.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 63 | 75 | + 2 SEs |
| Percent of students in each Performance Level | 83.3 | 10.5 | 6.2 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 59 | 70 | + 1 SEs |
| Percent of students in each Performance Level | 74.8 | 14.9 | 10.3 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 65 | |
| Percent of students in each Performance Level | 66.3 | 20.4 | 13.3 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 51 | 60 | - 1 SEs |
| Percent of students in each Performance Level | 51.2 | 25.7 | 23.1 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 47 | 55 | - 2 SEs |
| Percent of students in each Performance Level | 37.2 | 29.1 | 33.7 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 43 | 50 | - 3 SEs |
| Percent of students in each Performance Level | 23.3 | 23.2 | 53.5 | |
| *Large Group Medians | | | | |

# Appendix R

## Standard Error Bands for Reading Grade 4

| Reading Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 0.93 | 1.67 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 57 | 71 | + 3 SEs |
| Percent of students in each Performance Level | 69.8 | 22.0 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 56 | 69 | + 2 SEs |
| Percent of students in each Performance Level | 65.5 | 24.6 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 55 | 68 | + 1 SEs |
| Percent of students in each Performance Level | 65.5 | 24.6 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 66 | |
| Percent of students in each Performance Level | 61.7 | 25.9 | 12.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 53 | 64 | - 1 SEs |
| Percent of students in each Performance Level | 57.6 | 26.9 | 15.5 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 52 | 63 | - 2 SEs |
| Percent of students in each Performance Level | 52.7 | 28.5 | 18.8 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 51 | 61 | - 3 SEs |
| Percent of students in each Performance Level | 52.7 | 25.1 | 22.2 | |
| *Large Group Medians | | | | |

| Reading Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 5.0 | 5.0 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 69 | 81 | + 3 SEs |
| Percent of students in each Performance Level | 90.1 | 6.7 | 3.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 64 | 76 | + 2 SEs |
| Percent of students in each Performance Level | 84.5 | 11.6 | 3.9 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 59 | 71 | + 1 SEs |
| Percent of students in each Performance Level | 73.8 | 18.0 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 66 | |
| Percent of students in each Performance Level | 61.7 | 25.9 | 12.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 49 | 61 | - 1 SEs |
| Percent of students in each Performance Level | 41.0 | 36.8 | 22.2 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 44 | 56 | - 2 SEs |
| Percent of students in each Performance Level | 26.6 | 38.9 | 34.5 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 39 | 51 | - 3 SEs |
| Percent of students in each Performance Level | 14.1 | 38.6 | 47.3 | |
| *Large Group Medians | | | | |

| Reading Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 5.09 | 5.27 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 69 | 82 | + 3 SEs |
| Percent of students in each Performance Level | 90.1 | 6.7 | 3.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 64 | 77 | + 2 SEs |
| Percent of students in each Performance Level | 84.5 | 11.6 | 3.9 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 59 | 71 | + 1 SEs |
| Percent of students in each Performance Level | 73.8 | 18.0 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 66 | |
| Percent of students in each Performance Level | 61.7 | 25.9 | 12.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 49 | 61 | - 1 SEs |
| Percent of students in each Performance Level | 41.0 | 36.8 | 22.2 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 44 | 55 | - 2 SEs |
| Percent of students in each Performance Level | 26.6 | 38.9 | 34.5 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 39 | 50 | - 3 SEs |
| Percent of students in each Performance Level | 14.1 | 32.4 | 53.5 | |
| *Large Group Medians | | | | |

# Appendix S

## Standard Error Bands for Reading Grade 5

| Reading Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 0.97 | 1.74 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 56 | 74 | + 3 SEs |
| Percent of students in each Performance Level | 67.2 | 28.6 | 4.2 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 55 | 72 | + 2 SEs |
| Percent of students in each Performance Level | 62.0 | 32.1 | 5.9 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 54 | 71 | + 1 SEs |
| Percent of students in each Performance Level | 62.0 | 32.1 | 5.9 | |
| | | | | |
| Participants Recommended Cut Point* | | 53 | 69 | |
| Percent of students in each Performance Level | 55.9 | 35.7 | 8.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 52 | 67 | - 1 SEs |
| Percent of students in each Performance Level | 50.5 | 35.7 | 13.8 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 51 | 66 | - 2 SEs |
| Percent of students in each Performance Level | 45.4 | 40.8 | 13.8 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 50 | 64 | - 3 SEs |
| Percent of students in each Performance Level | 45.4 | 38.4 | 16.2 | |
| *Large Group Medians | | | | |

| Reading Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 5.0 | 6.0 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 68 | 87 | + 3 SEs |
| Percent of students in each Performance Level | 89.0 | 8.3 | 2.7 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 63 | 81 | + 2 SEs |
| Percent of students in each Performance Level | 83.8 | 13.5 | 2.7 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 58 | 75 | + 1 SEs |
| Percent of students in each Performance Level | 72.0 | 23.8 | 4.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 53 | 69 | |
| Percent of students in each Performance Level | 55.9 | 35.7 | 8.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 48 | 63 | - 1 SEs |
| Percent of students in each Performance Level | 40.1 | 43.7 | 16.2 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 43 | 57 | - 2 SEs |
| Percent of students in each Performance Level | 13.4 | 53.8 | 32.8 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 38 | 51 | - 3 SEs |
| Percent of students in each Performance Level | 13.4 | 32.0 | 54.6 | |
| *Large Group Medians | | | | |

| Reading Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 5.09 | 6.25 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 68 | 88 | + 3 SEs |
| Percent of students in each Performance Level | 89.0 | 8.3 | 2.7 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 63 | 81 | + 2 SEs |
| Percent of students in each Performance Level | 83.8 | 13.5 | 2.7 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 58 | 75 | + 1 SEs |
| Percent of students in each Performance Level | 72.0 | 23.8 | 4.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 53 | 69 | |
| Percent of students in each Performance Level | 55.9 | 35.7 | 8.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 48 | 63 | - 1 SEs |
| Percent of students in each Performance Level | 40.1 | 43.7 | 16.2 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 43 | 57 | - 2 SEs |
| Percent of students in each Performance Level | 13.4 | 53.8 | 32.8 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 38 | 50 | - 3 SEs |
| Percent of students in each Performance Level | 13.4 | 32.0 | 54.6 | |
| *Large Group Medians | | | | |

# Appendix T

## Standard Error Bands for Math Grade 3

| Mathematics Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 0.99 | 0.81 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 58 | 69 | + 3 SEs |
| Percent of students in each Performance Level | 71.3 | 17.7 | 11.0 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 57 | 69 | + 2 SEs |
| Percent of students in each Performance Level | 68.0 | 21.0 | 11.0 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 56 | 68 | + 1 SEs |
| Percent of students in each Performance Level | 64.1 | 24.9 | 11.0 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 67 | |
| Percent of students in each Performance Level | 64.1 | 22.5 | 13.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 54 | 66 | - 1 SEs |
| Percent of students in each Performance Level | 61.0 | 23.4 | 15.6 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 53.0 | 65 | - 2 SEs |
| Percent of students in each Performance Level | 55.8 | 28.6 | 15.6 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 52 | 65 | - 3 SEs |
| Percent of students in each Performance Level | 52.4 | 32.0 | 15.6 | |
| *Large Group Medians | | | | |

| Mathematics Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 5 | 5 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 70 | 82 | + 3 SEs |
| Percent of students in each Performance Level | 89.8 | 6.1 | 4.1 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 65 | 77 | + 2 SEs |
| Percent of students in each Performance Level | 84.4 | 10.5 | 5.1 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 60 | 72 | + 1 SEs |
| Percent of students in each Performance Level | 71.3 | 18.5 | 10.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 67 | |
| Percent of students in each Performance Level | 64.1 | 22.5 | 13.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 50 | 62 | - 1 SEs |
| Percent of students in each Performance Level | 44.7 | 35.0 | 20.3 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 45.0 | 57.0 | - 2 SEs |
| Percent of students in each Performance Level | 25.2 | 42.8 | 32.0 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 40 | 52 | - 3 SEs |
| Percent of students in each Performance Level | 17.3 | 35.1 | 47.6 | |
| *Large Group Medians | | | | |

| Mathematics Grade 3 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 5.10 | 5.06 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 70 | 82 | + 3 SEs |
| Percent of students in each Performance Level | 89.8 | 6.1 | 4.1 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 65 | 77 | + 2 SEs |
| Percent of students in each Performance Level | 84.4 | 10.5 | 5.1 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 60 | 72 | + 1 SEs |
| Percent of students in each Performance Level | 71.3 | 18.5 | 10.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 55 | 67 | |
| Percent of students in each Performance Level | 64.1 | 22.5 | 13.4 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 50 | 62 | - 1 SEs |
| Percent of students in each Performance Level | 44.7 | 35.0 | 20.3 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 45.0 | 57.0 | - 2 SEs |
| Percent of students in each Performance Level | 25.2 | 42.8 | 32.0 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 40 | 52 | - 3 SEs |
| Percent of students in each Performance Level | 17.3 | 35.1 | 47.6 | |
| *Large Group Medians | | | | |

# Appendix U

## Standard Error Bands for Math Grade 4

| Mathematics Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 1.11 | 0.97 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 57 | 71 | + 3 SEs |
| Percent of students in each Performance Level | 68.8 | 23.2 | 8.0 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 56 | 70 | + 2 SEs |
| Percent of students in each Performance Level | 64.7 | 27.3 | 8.0 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 55 | 69 | + 1 SEs |
| Percent of students in each Performance Level | 64.7 | 25.4 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 68 | |
| Percent of students in each Performance Level | 61.6 | 28.5 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 53 | 67 | - 1 SEs |
| Percent of students in each Performance Level | 57.5 | 30.7 | 11.8 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 52 | 66 | - 2 SEs |
| Percent of students in each Performance Level | 53.4 | 32.1 | 14.5 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 51 | 65 | - 3 SEs |
| Percent of students in each Performance Level | 53.4 | 32.1 | 14.5 | |
| *Large Group Medians | | | | |

| Mathematics Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 4.0 | 5.0 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 66 | 78 | + 3 SEs |
| Percent of students in each Performance Level | 85.5 | 7.7 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 62 | 78 | + 2 SEs |
| Percent of students in each Performance Level | 80.9 | 12.3 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 58 | 73 | + 1 SEs |
| Percent of students in each Performance Level | 72.6 | 20.6 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 68 | |
| Percent of students in each Performance Level | 61.6 | 28.5 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 50 | 63 | - 1 SEs |
| Percent of students in each Performance Level | 49.0 | 34.0 | 17.0 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 46 | 58 | - 2 SEs |
| Percent of students in each Performance Level | 36.8 | 35.8 | 27.4 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 42 | 53 | - 3 SEs |
| Percent of students in each Performance Level | 18.9 | 38.6 | 42.5 | |
| *Large Group Medians | | | | |

| Mathematics Grade 4 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 4.15 | 5.09 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 66 | 78 | + 3 SEs |
| Percent of students in each Performance Level | 85.5 | 7.7 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 62 | 78 | + 2 SEs |
| Percent of students in each Performance Level | 80.9 | 12.3 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 58 | 73 | + 1 SEs |
| Percent of students in each Performance Level | 72.6 | 20.6 | 6.8 | |
| | | | | |
| Participants Recommended Cut Point* | | 54 | 68 | |
| Percent of students in each Performance Level | 61.6 | 28.5 | 9.9 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 50 | 63 | - 1 SEs |
| Percent of students in each Performance Level | 49.0 | 34.0 | 17.0 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 46 | 58 | - 2 SEs |
| Percent of students in each Performance Level | 36.8 | 35.8 | 27.4 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 42 | 53 | - 3 SEs |
| Percent of students in each Performance Level | 18.9 | 38.6 | 42.5 | |
| *Large Group Medians | | | | |

# Appendix V

## Standard Error Bands for Math Grade 5

| Mathematics Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error (SE) cut score | | 0.96 | 1.09 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 61 | 74 | + 3 SEs |
| Percent of students in each Performance Level | 78.0 | 15.0 | 7.0 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 60 | 73 | + 2 SEs |
| Percent of students in each Performance Level | 74.6 | 18.4 | 7.0 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 59 | 72 | + 1 SEs |
| Percent of students in each Performance Level | 74.6 | 17.2 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* | | 58 | 71 | |
| Percent of students in each Performance Level | 70.3 | 21.5 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 57 | 70 | - 1 SEs |
| Percent of students in each Performance Level | 66.9 | 23.0 | 10.1 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 56 | 69 | - 2 SEs |
| Percent of students in each Performance Level | 66.9 | 23.0 | 10.1 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 55 | 68 | - 3 SEs |
| Percent of students in each Performance Level | 63.0 | 24.2 | 12.8 | |
| *Large Group Medians | | | | |

| Mathematics Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| Standard Error of Measurement (CSEM) | | 5 | 6 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 73 | 89 | + 3 SEs |
| Percent of students in each Performance Level | 93.0 | 6.2 | 0.8 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 68 | 83 | + 2 SEs |
| Percent of students in each Performance Level | 87.2 | 10.8 | 2.0 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 63 | 77 | + 1 SEs |
| Percent of students in each Performance Level | 80.5 | 15.5 | 4.0 | |
| | | | | |
| Participants Recommended Cut Point* | | 58 | 71 | |
| Percent of students in each Performance Level | 70.3 | 21.5 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 53 | 65 | - 1 SEs |
| Percent of students in each Performance Level | 54.5 | 28.2 | 17.3 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 48 | 59 | - 2 SEs |
| Percent of students in each Performance Level | 37.1 | 37.5 | 25.4 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 43 | 53 | - 3 SEs |
| Percent of students in each Performance Level | 22.1 | 32.4 | 45.5 | |
| *Large Group Medians | | | | |

| Mathematics Grade 5 | | | | |
|---|---|---|---|---|
| Recommended Cut Points*Plus/Minus Selected Standards Errors (SEs) of the Cut Scores | | | | |
| | Basic | Proficient Scale Score | Advanced Scale Score | SE Calculations |
| SEM Combined (SEMcomb) | | 5.09 | 6.10 | |
| | | | | |
| Participants Recommended Cut Point* + 3 SEs | | 73 | 89 | + 3 SEs |
| Percent of students in each Performance Level | 93.0 | 6.2 | 0.8 | |
| | | | | |
| Participants Recommended Cut Point* + 2 SEs | | 68 | 83 | + 2 SEs |
| Percent of students in each Performance Level | 87.2 | 10.8 | 2.0 | |
| | | | | |
| Participants Recommended Cut Point* + 1 SEs | | 63 | 77 | + 1 SEs |
| Percent of students in each Performance Level | 80.5 | 15.5 | 4.0 | |
| | | | | |
| Participants Recommended Cut Point* | | 58 | 71 | |
| Percent of students in each Performance Level | 70.3 | 21.5 | 8.2 | |
| | | | | |
| Participants Recommended Cut Point* - 1 SEs | | 53 | 65 | - 1 SEs |
| Percent of students in each Performance Level | 54.5 | 28.2 | 17.3 | |
| | | | | |
| Participants Recommended Cut Point* - 2 SEs | | 48 | 59 | - 2 SEs |
| Percent of students in each Performance Level | 37.1 | 37.5 | 25.4 | |
| | | | | |
| Participants Recommended Cut Point* - 3 SEs | | 43 | 53 | - 3 SEs |
| Percent of students in each Performance Level | 22.1 | 32.4 | 45.5 | |
| *Large Group Medians | | | | |

# APPENDIX E: ALIGNMENT OF THE 2010 MOD-MSA: READING ITEMS TO THE STATE STANDARDS

The following tables by grade depict the alignment of the Mod-MSA: Reading items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/reading/bygrade/grade3.html and substitute the different grades in the net address for the different grades information.

## Grade 3

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213631 | General | 1D |
| 3 | 100000213633 | General | 1D |
| 5 | 100000360182 | General | 1D |
| 7 | 100000101513 | General | 1B |
| 8 | 100000101514 | General | 1B |
| 9 | 100000360183 | General | 1B |
| 10 | 100000101516 | General | 1D |
| 11 | 100000101517 | General | 1D |
| 12 | 100000101518 | General | 1D |
| 14 | 100000101969 | General | 1E |
| 17 | 100000101972 | Literary | 2A |
| 18 | 100000101973 | Literary | 2A |
| 19 | 100000101974 | Literary | 2A |
| 21 | 100000101932 | General | 1E |
| 22 | 100000101934 | Literary | 2A |
| 25 | 100000101936 | Literary | 2A |
| 26 | 100000101937 | Literary | 2A |
| 27 | 100000101938 | Literary | 2A |
| 28 | 100000346452 | General | 1E |
| 30 | 100000346444 | Literary | 2A |
| 32 | 100000346450 | Literary | 2A |
| 33 | 100000346445 | Literary | 2A |
| 34 | 100000346448 | Literary | 2A |
| 35 | 100000260458 | Literary | 2A |
| 36 | 100000260460 | Literary | 2A |
| 37 | 100000260461 | Literary | 2A |
| 39 | 100000260457 | Literary | 2A |
| 41 | 100000260465 | General | 1E |
| 42 | 100000101905 | General | 1E |
| 44 | 100000101907 | Informational | 3A |
| 45 | 100000101908 | Informational | 3A |
| 46 | 100000101909 | Informational | 3A |
| 48 | 100000101911 | Informational | 3A |
| 50 | 100000101530 | Informational | 3A |
| 52 | 100000101532 | Informational | 3A |
| 54 | 100000101528 | General | 1E |
| 55 | 100000101533 | Informational | 3A |
| 56 | 100000260338 | General | 1E |
| 57 | 100000260342 | Literary | 2A |
| 58 | 100000260345 | Informational | 3A |
| 61 | 100000260339 | Informational | 3A |
| 64 | 100000260368 | Informational | 3A |
| 65 | 100000260371 | Informational | 3A |
| 68 | 100000260365 | Informational | 3A |
| 69 | 100000300552 | General | 1E |

## Grade 4

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213644 | General | 1D |
| 2 | 100000213645 | General | 1D |
| 3 | 100000213646 | General | 1D |
| 4 | 100000213647 | General | 1D |
| 6 | 100000213637 | General | 1D |
| 7 | 100000213638 | General | 1D |
| 10 | 100000213641 | General | 1D |
| 11 | 100000213642 | General | 1D |
| 12 | 100000360192 | General | 1D |
| 13 | 100000357134 | Literary | 2A |
| 15 | 100000357133 | General | 1E |
| 17 | 100000357136 | Literary | 2A |
| 18 | 100000357137 | Literary | 2A |
| 19 | 100000357138 | Literary | 2A |
| 20 | 100000357106 | Literary | 2A |
| 21 | 100000357104 | General | 1E |
| 22 | 100000357107 | Literary | 2A |
| 24 | 100000357108 | Literary | 2A |
| 25 | 100000357109 | Literary | 2A |
| 28 | 100000260486 | Literary | 2A |
| 29 | 100000260483 | General | 1E |
| 31 | 100000260489 | Literary | 2A |
| 32 | 100000260488 | Literary | 2A |
| 34 | 100000269896 | Literary | 2A |
| 37 | 100000269900 | Literary | 2A |
| 38 | 100000269897 | Literary | 2A |
| 40 | 100000269899 | Literary | 2A |
| 43 | 100000357098 | Informational | 3A |
| 45 | 100000357100 | Informational | 3A |
| 46 | 100000357101 | Informational | 3A |
| 47 | 100000462157 | General | 1E |
| 48 | 100000101997 | General | 1E |
| 51 | 100000101999 | Informational | 3A |
| 52 | 100000102000 | Informational | 3A |
| 53 | 100000102001 | Informational | 3A |
| 54 | 100000200070 | General | 1E |
| 56 | 100000102026 | General | 1E |
| 57 | 100000102027 | Informational | 3A |
| 59 | 100000102024 | Informational | 3A |
| 60 | 100000102028 | Informational | 3A |
| 61 | 100000102029 | Informational | 3A |
| 63 | 100000267470 | General | 1E |
| 65 | 100000301034 | Informational | 3A |
| 66 | 100000267473 | Informational | 3A |
| 68 | 100000267472 | Informational | 3A |

## Grade 5

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213650 | General | 1D |
| 2 | 100000213651 | General | 1D |
| 3 | 100000213652 | General | 1D |
| 4 | 100000213653 | General | 1D |
| 6 | 100000213655 | General | 1D |
| 7 | 100000213656 | General | 1D |
| 8 | 100000213657 | General | 1D |
| 10 | 100000213659 | General | 1D |
| 11 | 100000360197 | General | 1D |
| 12 | 100000102095 | Literary | 2A |
| 13 | 100000102093 | General | 1E |
| 15 | 100000102096 | Literary | 2A |
| 17 | 100000102098 | Literary | 2A |
| 18 | 100000102099 | Literary | 2A |
| 19 | 100000102113 | Literary | 2A |
| 20 | 100000102111 | General | 1E |
| 21 | 100000102114 | Literary | 2A |
| 22 | 100000102115 | Literary | 2A |
| 23 | 100000102116 | Literary | 2A |
| 26 | 100000102104 | Literary | 2A |
| 30 | 100000102106 | Literary | 2A |
| 31 | 100000102107 | Literary | 2A |
| 32 | 100000102108 | Literary | 2A |
| 34 | 100000102084 | General | 1E |
| 35 | 100000102087 | Literary | 2A |
| 36 | 100000102088 | Literary | 2A |
| 39 | 100000102090 | Literary | 2A |
| 41 | 100000102048 | Informational | 3A |
| 43 | 100000102050 | Informational | 3A |
| 44 | 100000102051 | Informational | 3A |
| 45 | 100000102052 | General | 1E |
| 46 | 100000102047 | Informational | 3A |
| 47 | 100000102056 | General | 1E |
| 49 | 100000102059 | Informational | 3A |
| 50 | 100000102060 | Informational | 3A |
| 51 | 100000102061 | Informational | 3A |
| 54 | 100000102068 | Informational | 3A |
| 55 | 100000102069 | Informational | 3A |
| 57 | 100000102067 | General | 1E |
| 59 | 100000102071 | Informational | 3A |
| 60 | 100000102072 | Informational | 3A |
| 61 | 100000267477 | Informational | 3A |
| 62 | 100000268378 | Informational | 3A |
| 64 | 100000268380 | Informational | 3A |
| 67 | 100000303033 | Informational | 3A |

## Grade 6

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213662 | General | 1D |
| 2 | 100000213663 | General | 1D |
| 3 | 100000213664 | General | 1D |
| 4 | 100000213665 | General | 1D |
| 5 | 100000213666 | General | 1D |
| 7 | 100000213668 | General | 1D |
| 8 | 100000213669 | General | 1D |
| 9 | 100000213670 | General | 1D |
| 10 | 100000213671 | General | 1D |
| 12 | 100000102192 | Literary | 2A |
| 13 | 100000102194 | Literary | 2A |
| 14 | 100000102191 | General | 1E |
| 15 | 100000102195 | Literary | 2A |
| 16 | 100000102196 | Literary | 2A |
| 17 | 100000102174 | Literary | 2A |
| 18 | 100000102175 | Literary | 2A |
| 19 | 100000102173 | General | 1E |
| 20 | 100000102176 | Literary | 2A |
| 21 | 100000102178 | Literary | 2A |
| 23 | 100000269997 | Literary | 2A |
| 24 | 100000270006 | Literary | 2A |
| 27 | 100000270005 | Literary | 2A |
| 28 | 100000270007 | Literary | 2A |
| 30 | 100000270018 | General | 1E |
| 31 | 100000270021 | Literary | 2A |
| 34 | 100000270013 | Literary | 2A |
| 35 | 100000270020 | Literary | 2A |
| 37 | 100000257088 | General | 1E |
| 39 | 100000257114 | Informational | 3A |
| 40 | 100000257118 | Informational | 3A |
| 41 | 100000257119 | Informational | 3A |
| 42 | 100000257120 | Informational | 3A |
| 43 | 100000257102 | Informational | 3A |
| 45 | 100000257103 | Informational | 3A |
| 46 | 100000257104 | Informational | 3A |
| 47 | 100000257071 | General | 1E |
| 50 | 100000102136 | Informational | 3A |
| 51 | 100000102137 | Informational | 3A |
| 52 | 100000102139 | Informational | 3A |
| 53 | 100000102140 | Informational | 3A |
| 55 | 100000257110 | Informational | 3A |
| 56 | 100000257083 | General | 1E |
| 57 | 100000257106 | Informational | 3A |
| 58 | 100000257107 | Informational | 3A |
| 59 | 100000257109 | Informational | 3A |

## Grade 7

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213674 | General | 1D |
| 2 | 100000213675 | General | 1D |
| 3 | 100000213676 | General | 1D |
| 5 | 100000213678 | General | 1D |
| 6 | 100000213679 | General | 1D |
| 8 | 100000270578 | General | 1E |
| 9 | 100000270572 | Literary | 2A |
| 10 | 100000270574 | Literary | 2A |
| 11 | 100000322011 | Literary | 2A |
| 14 | 100000270575 | Literary | 2A |
| 15 | 100000270109 | Literary | 2A |
| 16 | 100000270112 | Literary | 2A |
| 17 | 100000270113 | General | 1E |
| 18 | 100000270114 | General | 1E |
| 21 | 100000270121 | Literary | 2A |
| 22 | 100000102276 | Literary | 2A |
| 23 | 100000102277 | Literary | 2A |
| 24 | 100000102274 | General | 1E |
| 25 | 100000102278 | Literary | 2A |
| 26 | 100000102279 | Literary | 2A |
| 27 | 100000102263 | General | 1E |
| 28 | 100000102267 | Literary | 2A |
| 29 | 100000102268 | Literary | 2A |
| 30 | 100000102269 | Literary | 2A |
| 31 | 100000102266 | Literary | 2A |
| 33 | 100000257182 | General | 1E |
| 34 | 100000257216 | Informational | 3A |
| 35 | 100000257218 | Informational | 3A |
| 36 | 100000257219 | Informational | 3A |
| 38 | 100000257221 | Informational | 3A |
| 39 | 100000102244 | General | 1E |
| 40 | 100000102246 | Informational | 3A |
| 41 | 100000102247 | Informational | 3A |
| 42 | 100000102248 | Informational | 3A |
| 43 | 100000102245 | General | 1E |
| 44 | 100000102218 | Informational | 3A |
| 45 | 100000102220 | Informational | 3A |
| 46 | 100000102221 | Informational | 3A |
| 47 | 100000102222 | Informational | 3A |
| 48 | 100000102217 | General | 1E |
| 50 | 100000257769 | Informational | 3A |
| 51 | 100000257770 | Informational | 3A |
| 52 | 100000257773 | Informational | 3A |
| 53 | 100000322795 | General | 1E |
| 54 | 100000257767 | Informational | 3A |

## Grade 8

| Item # | UIN Number | Strand | Standard Code |
|---|---|---|---|
| 1 | 100000213683 | General | 1D |
| 2 | 100000213684 | General | 1D |
| 4 | 100000213686 | General | 1D |
| 5 | 100000213687 | General | 1D |
| 6 | 100000213688 | General | 1D |
| 8 | 100000270136 | Literary | 2A |
| 10 | 100000273591 | General | 1E |
| 11 | 100000270131 | Literary | 2A |
| 12 | 100000270135 | Literary | 2A |
| 13 | 100000270137 | Literary | 2A |
| 15 | 100000102331 | Literary | 2A |
| 16 | 100000102329 | General | 1E |
| 17 | 100000102330 | General | 1E |
| 18 | 100000102334 | Literary | 2A |
| 19 | 100000102335 | Literary | 2A |
| 20 | 100000270154 | Literary | 2A |
| 22 | 100000270158 | Literary | 2A |
| 23 | 100000273597 | General | 1E |
| 24 | 100000270163 | Literary | 2A |
| 25 | 100000270159 | Literary | 2A |
| 27 | 100000102341 | General | 1E |
| 28 | 100000102343 | Literary | 2A |
| 29 | 100000102344 | Literary | 2A |
| 30 | 100000102346 | Literary | 2A |
| 31 | 100000102345 | Literary | 2A |
| 32 | 100000257161 | Informational | 3A |
| 34 | 100000257787 | General | 1E |
| 36 | 100000257158 | Informational | 3A |
| 37 | 100000257160 | Informational | 3A |
| 38 | 100000257163 | Informational | 3A |
| 39 | 100000102320 | General | 1E |
| 40 | 100000102322 | Informational | 3A |
| 41 | 100000102323 | Informational | 3A |
| 42 | 100000102324 | Informational | 3A |
| 43 | 100000102321 | General | 1E |
| 45 | 100000257140 | Informational | 3A |
| 47 | 100000257144 | Informational | 3A |
| 48 | 100000257777 | General | 1E |
| 49 | 100000302674 | General | 1E |
| 50 | 100000302675 | Informational | 3A |
| 51 | 100000102304 | Informational | 3A |
| 52 | 100000102305 | Informational | 3A |
| 53 | 100000102302 | General | 1E |
| 54 | 100000102307 | Informational | 3A |
| 55 | 100000102308 | Informational | 3A |

# APPENDIX F: IDENTIFYING STUDENTS FOR PARTICIPATION IN THE MOD-MSA: READING PROGRAM

## Identifying Students for Participation in the Mod-MSA Program

Maryland developed the Alternate Maryland School Assessment in Reading and Mathematics based on Modified Academic Achievement Standards (Mod-MSA) for administration to students in grades 3 through 8. Student participation in the Mod-MSA, either in reading, in mathematics, or in both tested content areas, was qualified through the Individualized Education Program (IEP) process using the participation eligibility criteria, below:

A student eligible for the Mod-MSA is identified based on his/her individual evaluation information and the instructional and service information on his/her IEP. The student is identified as appropriate for instruction and assessment using modified academic achievement standards aligned with the student's grade-level academic content standards. Students pursuing the Mod-MSA are not precluded from completing the requirements for the regular high school diploma. To participate in Mod-MSA, the student must meet each of the following criteria:

- The student is learning based on the State's approved grade-level academic content standards for the grade in which the student is enrolled. There must be sufficient objective evidence demonstrating that the student is not likely to achieve grade-level proficiency within the school year covered by his/her IEP.

**AND**

- The student requires and receives modified academic achievement standards aligned with the Maryland Academic Content Standards (the Voluntary State Curriculum) for the student's grade level during assessments and instruction. In addition, specific accommodations implemented in the testing/assessment and instructional settings may include: test items which are less complex, fewer and shorter reading passages, shorter or less difficult questions, and test items with fewer distractors.

**AND**

- The student must have had consecutive years of individualized intensive instruction in reading **and/or** mathematics consistent with his/her IEP (beginning with the most recent year), and although progress toward grade level standards was made, he/she is not yet making progress at grade level.

**AND**

- The student must demonstrate that he/she cannot attain proficiency in the actual grade level MSA, even with the provision of accommodations based on documented multiple valid and objective measures of the student's progress (or lack of progress). Examples include the end-of course assessments, district-wide assessments, data gathered from classroom assessments, and other formative assessments that can validate documented academic achievement in response to appropriate instruction. There must be enough time to document the progress (or lack of progress) in response to appropriate instruction.

The IEP Team decision-making model to be used in identifying students to participate in Mod-MSA is found in the Maryland Accommodation Manual (MAM).

# APPENDIX G: COMPARABILITY STUDY OF PAPER AND PENCIL, AND ONLINE ADMINISTRATION OF THE MOD-MSA

# Comparison of Paper-Pencil Version with the On-Line Version of the Maryland Modified School Assessment (Mod-MSA) in Reading and Mathematics (Grades 4 and 5)

In recent years, computer based testing in K-12 settings has become popular in consideration of its many advantages. As Way, Davis, and Fitzpatrick (2006) point out, these include savings in cost (no printing and shipment of the paper and pencil format test); improvement in test security; flexibility in test administration; and a base for the utilization of technology in presenting innovative item formats and test delivery algorithms. Above all, on-line (OL) administration provides a quick turnaround of results that could be especially helpful to provide timely feedback to students, teachers, and schools. Furthermore, concern about students' limited familiarity with computers now seems to be displaced by students' preference for computer testing vis-à-vis the paper and pencil (P&P) version (see Glassnapp, Poggio, Poggio, & Yang, 2005).

In comparing the results of the P&P and OL version tests, the main consideration has traditionally been the establishment of a common scale so that scores from the two versions are equivalent. This is often done by matching two groups of students on an external criterion and then comparing their performance. Although an external variable for matching test takers may be difficult to obtain, matching students on a viable external criterion has distinct advantages. The method seems preferable to the costly and rigorous efforts necessary to control for fatigue, student motivation, etc., as would be necessary if a single group of students were to take two versions of a test. Furthermore, assigning (or selecting) students to form randomly equivalent groups may not be a plausible solution because of say, limited technology resources (e.g., a lack of computers at certain schools), or small sample sizes across groups of interest. To date, few studies (e.g., Kim, D. H. & Huynh, H., 2008, 2009; Way, et. al., 2006) have utilized an external matching variable in the comparison of P&P and OL tests.

Once an approach for creating equivalent groups is selected and data in comparable format is obtained or created, analysis of the data is completed using statistical methods, such as, Item Response Theory (IRT), Hierarchical Linear Modeling (HLM), Differential Item Functioning (DIF), Multiple Regression (MR), and Analysis of Covariance (ANCOVA). Selection of an appropriate statistical method not only depends on the design of the study, and availability of data (e.g., large n-counts for IRT), but also on the researcher's goals for the research. For example, the intent may be to compare the performances of the *total* tests across the two groups of examination modes, or the interest may lie in the comparing of *each item's* performance on a test across the two examination modes.

It should be noted that there have been some studies that have compared the P&P and OL versions of tests through the use of item-level analysis. For example, a study by Poggio, Glasnapp, Yang, and Poggio (2005) included both HLM and DIF methods in the analysis, but neither method included an external matching variable. The DIF method was based on random assignment of four forms for the two testing-mode comparison. The HLM method, on the other hand, included three level of analysis where a single group with a counterbalance repeated measure (common persons) design was used for within-student effects at Level 1, and between student effects at Level 2. The students were nested within schools (i.e., each school was assigned to a testing mode), and this variable was used at Level 3. Besides Poggio et al.'s (2005)

study, methods using item-level analysis have been used by some other investigators to study mode differences in test administration, e.g., Keng, McClarty, & Davis, 2006, although Keng, et al.'s study did not use DIF or the external variable design in their research.

Way, et al. (2006), on the other hand, used an external criterion as a matching variable in the comparison of P&P and OL versions of a test. The authors compared test performance of Grade 8 students that tested online with groups from the P&P administration after matching them on their previous spring test performance. In this study, the main purpose of the researchers was to adjust student scores to obtain equivalence across mode of test administration.

This study, as in the previous year (2009) where Grades 7 and 8 were analyzed, uses two methods of comparing test-mode effects. It uses an *external* variable both as a matching variable in forming groups for DIF analysis and as a covariate for the ANCOVA.

## Purpose of the Study

The basic requirements for Mod-MSA reading and mathematics assessments do not call for an adjustment to student scores based on the testing modes. The desired goal is only to note the extent to which modes of assessment influences student performance at both the total test level and at the item level.

The purpose of this study, therefore, is:

1. to analyze whether the total Mod-MSA P&P version differs substantially from the OL version with respect to student achievement, and
2. to identify those items that favor one testing mode, and provide this information to MSDE so that steps may be taken to eliminate or modify these items in order to eliminate bias (if bias exists) in Mod-MSA operational forms.

## Mod-MSA Reading and Mathematics Assessments

In years prior to the first administration of the Mod-MSA Grades 3-5 tests in spring 2010, approximately 95% of the students, except for Grade 3, regardless of their classification, had taken the MSA examination. Grade 3 students had not taken the MSA in 2009 when they would have been in Grade 2 because the MSA examinations are administered to students starting in Grade 3. Therefore, Grade 3 could not be included in this study.

The Mod-MSA assessments in reading and mathematics were designed for students with disabilities who, based on a decision making process undertaken by their Individual Educational Planning (IEP) team, met specific eligibility criteria. The Mod-MSA tests are alternates to the tests in the MSA Program. The Alternate assessments based on modified achievement standards (AA-MAS) are commonly referred to as 2% assessments. They are specified by the guidelines set by the U.S. Department of Education (DOE) on the basis of the U.S. DOE's Final Rule, of April 9, 2007[5]. According to the rule, although states may test more than 2% of the population using the AA-MAS, they may report only 2% as proficient or above proficiency, for Adequate Yearly Progress (AYP) determinations.

The 2010 Mod-MSA reading and mathematics assessments for Grades 4 and 5 are composed of a mixture of items (unaltered MSA items, modified MSA items, and items created specifically for the Mod-MSA assessments). The different Mod-MSA item types are intended to provide students access to the grade level content standards that incorporates variation in test delivery

[5] U.S. DOE's rule published Monday, April 9, 2007, in the Federal Register as "Title I-Improving the Academic Achievement of the Disadvantaged; Individual of Disabilities Education Act, Final Rule."

through a test that is designed to meet the specific learning characteristics of the students in this population. The format includes standard MSA items from the 2009 administration which were modified to allow students in this population greater access to the material. They also include intact MSA items (for reading), and some new items that were created specifically for the 2010 Mod-MSA administration. Other item modifications include, but are not limited to fewer and shorter reading passages, shorter and less complex questions, and test items with fewer item choices. Both the reading and the mathematics tests had more items administered than were required for the final operational test form. Since the newly created and modified items were administered for the first time during the 2010 administration, some of the items produced statistics that were unacceptable to the Data Review Committee (e.g., negative point biserials). The Committee, therefore, eliminated these items with poor statistics from the selection process. Items were then selected from the remaining pool for the final, scored (operational) form.

Both the Mod-MSA reading and mathematics tests contained only dichotomously scored items (i.e., 45 items for reading and 51 items for mathematics for the operational/scored forms). The items used in the administration of the Mod-MSA assessments were based on Maryland's Voluntary State Curriculum (VSCMSC). The test items for the Mod-MSA were aligned to the VSCMSC for the grade being assessed. Despite the similarity between tests (MSA and Mod-MSA), the Mod-MSA tests are considered separate assessments with a unique set of achievement standards (i.e., cut scores). Furthermore, the Mod-MSA was administered as both a P&P and OL version while the MSA was administered only in the P&P format. Local school systems determined which schools would test online based on the availability of computers. In some cases, special education staff worked with individual students to determine the most appropriate assessment mode for a specific student, after he or she was given the opportunity to take the P&P and OL sample test items.

## Research Methodology

### ***The Data Set***

Students from Grade 4 and Grade 5, who completed the Mod-MSA, were included in this study. The students' participation for the Mod-MSA was determined by their Individual Educational Planning (IEP) teams. The number of students in this population was expected to be moderate.

Since most of the students who completed the 2010 Mod-MSA in Grades 4 and 5 also took the 2009 MSA in the same subject area, their scores on the 2009 MSA could be used as a covariate and also as an external matching variable for the DIF analyses. The 2009 MSA was administered only as a P&P test, which further enhanced the use of these scores as a covariate or a matching variable because the administration mode variables in the Mod-MSA were not affected by test mode (i.e., the MSA scores) in determining future performance or group classification.

Although almost all 2010 Mod-MSA students in the two grades of interest had a corresponding score on the MSA in 2009, there was no guarantee that it would be possible to match the two sets of scores for each student. The best identifier for matching students on the two tests was the unique State ID. This was the primary matching method used for identifying Mod-MSA students' 2009 MSA test scores. This matching criterion was not perfect, and the fact that a few

students may not have taken the MSA in 2009, it was expected that the matched sample thus produced would be smaller than the original Mod-MSA student population.

***Methods of Analysis***

This study used two methods of comparing test-mode effects by using an external variable that served as a covariate for the analysis of covariance (ANCOVA) and as a matching variable in forming groups for DIF analysis.

**a. Test Level Analysis:**

In order to examine the differences between groups based on mode of administration, a simple straightforward method would be to use the t-test to determine the significance of the mean differences between groups. However, because the use of a covariate would reduce the within group error and thus produced a more sensitive and powerful test (Stevens, 1990), the ANCOVA was used to compare the P&P and OL version of a test. The covariate in this study was the students' performance on the 2009 MSA because students' Mod-MSA test scores were expected to be positively correlated with their scores on the P&P version of the MSA. As such, students' 2009 MSA scores were considered to be one of the predictors of Mod-MSA test scores, provided the hypothesis of no difference between testing modes for the Mod-MSA examinations was tenable.

A primary benefit of using the ANCOVA method was the *partial* equating of the groups that tested across the two different modes of the Mod-MSA by controlling for students' initial differences (i.e., their differences in achievement on an external variable – the MSA that was expected to correlate with the dependent variable, i.e., the Mod-MSA). Using students' 2009 MSA scores as a covariate equalized the groups on *one* factor, the effects of students' prior knowledge in reading and mathematics that could confound the effects of the testing mode. Instead of testing for significance of the difference in means between the two modes of testing, we tested the difference between the adjusted means of the two modes of testing (i.e. the means that were equalized on the covariate).

It should be noted that the correlation between the Mod-MSA and the MSA scores would be underestimated because of the attenuation effects of the Mod-MSA student population. The population of students taking the Mod-MSA is very different from that of the MSA population. It is likely that these students' MSA scores were in a limited range of the MSA scale vis-à-vis these students' Mod-MSA scores. Nonetheless, we expected the correlation between the dependent and the independent variable to exceed 0.30. This is important, as below this correlation threshold, it is unlikely that the addition of the covariate will lead to an appreciable increase in precision (Cohran, 1957; and Feldt, 1958).

Furthermore, it should also be noted that the assignment of students to the mode of administration was not random. In such cases, the ANCOVA (as in most other statistical analyses) has an important limitation (Anderson, 1963; Lord, 1969). There could be various other variables pertaining to non-randomly assigned schools that may be the cause of differences between the two groups. However, within the limitation of such a possibility, the study gives us a picture of the situation as it exists. This, in our opinion, is justifiable since there is no requirement of adjustments to student scores at this particular time.

In using the ANCOVA, three assumptions regarding the regression part of the covariance analysis, besides those associated with the analysis of variance (ANOVA), have to be met. The ANOVA is fairly robust to violation of its assumptions of normality of the distribution of the dependent variable and the equality of population variance in the two groups. As would be

expected, the ANCOVA is robust to the assumptions associated with the ANOVA, but it is also robust to the second of the three additional ANCOVA assumptions listed below, i.e.:

1. a linear relationship between the dependent variable (i.e., the scores on the Mod-MSA) and the covariate exists;
2. the covariate (i.e., the scores on the MSA) is measured without error; and
3. the homogeneity in the population of the regression slopes for the two groups classified on the basis of the testing mode administered (i.e., there is no covariate - MSA scores - by testing-mode interaction).

The first and the last of the three assumptions listed above were checked for tenability prior to the ANCOVA analysis to ensure that these were not violated. In the event that the homogeneity of the regression slopes was not met (i.e., an interaction effect between the covariate and the mode of administration existed) then limits on the regions of non-significance on the covariate were to be established by the use of the Johnson-Neyman technique (Pedhazur, 1973).

**b. Item Level Analysis:**

For the *item-level* analysis, methods relating to DIF were used to assess the performance of mode effects by items. Groups, based on the mode of test administration, were matched on an external variable (i.e., the students' MSA scores) in this analysis.

Determining that an item is biased requires an inference be made, for which DIF is a necessary, but not sufficient condition (Hambleton, Swaminathan, & Rogers, 1991). Thus, DIF is an important piece of evidence to gather when examining the equivalence of a test across administration modes, but this evidence alone is not sufficient to conclude that an item is biased. This analysis will, however, gives test developers a chance to examine administration method (in this study) with respect to items that may exhibit testing-mode bias, and take this opportunity to eliminate or mitigate the effects of items judged to show bias.

DIF analysis (e.g., the contingency table approaches in Camilli and Shepard, 1994) identifies items that do not function equally between matched groups of individuals. The matching is generally based on equivalency of overall performance, and items that do not perform equally among groups of matched individuals are considered to perform differentially. However, the problem with DIF analysis, in the context of the test-mode comparison, is that student groupings created on the basis of their performance on the administered test may confound the equality criterion of these proficiency groups if the testing mode has systematic differences across some or all items. Specifically, the systematic differences across items will contribute to the score on which students are matched to examine mode effects. It is, therefore, useful to match students on a common non-biased platform (i.e., on a "non-biased" external variable/s prior to DIF analysis).

However, the *unbiased* external criterion must be *an a priori predictor of test performance* for matching students. Because of this, it is important to select an external variable that is not only unbiased with respect to the testing mode, but which would also be a significant predictor of students' performance on the test if, indeed, no test-mode performance-differences exist for the test.

For this study, the Mod-MSA proficiency -groups for the DIF analysis were based on the Mod-MSA students' performance on the corresponding 2009 MSA test (external variable). Groups,

based on the mode of test administration, were then matched on the external variable (i.e., the scores on the MSA) in the analysis of mode effects on each item.

As stated earlier in this paper, most of the same students who took the Mod-MSA assessments in 2010 will have taken the P&P administered MSA in 2009. Since the MSA is a P&P only administration, scores of students on the MSA could be used as a strong unbiased variable for the creation of equivalent groups on the Mod-MSA for the DIF analysis. Because the Mod-MSA is similar to the MSA, student scores on the MSA can be seen as predictive of performance on the Mod-MSA, provided the items on the Mod-MSA indeed has no test-mode effects.

Since the Mod-MSA examinations do not have any polytomously scored items, the Mantel-Haenszel Chi-Square ($MH\chi^2$) together with ETS's Delta Scale were used for the contingency and the effect-size approach[6] to DIF.

The Mantel and Haenszel (1959) chi-square, which approximately follows a chi-square distribution with one degree of freedom, can be formulated as per the following (from Camilli & Shepard, 1994):

$$\text{MH } \chi^2 = \frac{\left\{\left|\sum_{j=1}^{S}\left[A_j - E\left(A_j\right)\right]\right| - 1/2\right\}^2}{\sum_{j=1}^{S} VAR\left(A_j\right)}, \text{ where}$$

$A_j$ and $E(A_j)$ are the observed number of correct responses and the expected number on the item, respectively for the Reference group, while $VAR(A_j)$ is the variance associated with the observed score.

In order to calculate the Delta scale, the Mantel and Haenszel (1959) log odds ratio was calculated using the following equation:

$$\alpha_{MH} = \frac{\sum_{j=1}^{S} A_j D_j / T_j}{\sum_{j=1}^{S} B_j C_j / T_j}, \text{ where}$$

the various variables in the equation are from the following 2 x 2 contingency table for the *jth* total score on the test (Camilli & Shepard, 1994, p. 106).

Score on studied item with general notation

| | | 1 | 0 | Total |
|---|---|---|---|---|
| Group | R | $A_j$ | $B_j$ | $n_{Rj}$ |
| | F | $C_j$ | $D_j$ | nFj |
| | | $m_{1j}$ | $m_{0j}$ | $T_j$ |

The log odds ratio is a transformation of the odds ratio with its range being in the interval $-\infty$ to $+\infty$. The simple natural logarithm transformation of this odds ratio is symmetrical around zero, in which zero has the interpretation of equal odds. The odds ratio is transformed into a log

[6] For a detailed discussion on Mantel-Haenszel Chi-square, the Delta Scale and ETS Categories, please refer to Camilli and Shepard (1994).

odds ratio as per the following: $\beta_{M-H} = \ln(\alpha_{M\text{-}H})$. $\beta_{M-H}$, also has the advantage of being transformed linearly to other interval scale metrics (Camilli & Shepard, 1994). This fact is utilized in creating the Delta scale (*D*), which is defined as $D = -2.35\beta_{M-H}$.

The *M-H* $\chi^2$ is examined in conjunction with the Delta scale (*D*) to obtain DIF classifications depicted in Table 1, below.

**Table 1: DIF Classification**

| Category | Description | Criterion |
|---|---|---|
| A | No DIF | Non-significant *M-H* $\chi^2$ or $\|D\| < 1.0$ |
| B | Weak DIF | Significant *M-H* $\chi^2$ and $\|D\| < 1.5$ or Non-significant *M-H* $\chi^2$ and $\|D\| > 1.0$ |
| C | Strong DIF | Significant *M-H* $\chi^2$ and $\|D\| \geq 1.5$ |

As stated previously, the groupings for the DIF analysis were based on matching students' scores on the MSA. Four proficiency -groupings of the Mod-MSA students were formed at quarter intervals of the total MSA score. All the students who had taken the Mod-MSA were used in this analysis. The Performance on the Mod-MSA for the four external proficiency -matched groups was then compared for each item to evaluate potential differential performance by mode.

The matching method described above for forming equal proficiency groups are the same as those used in conventional DIF analysis with one exception: instead of classifying proficiency groupings based on student performance on the test they have taken, i.e., the Mod-MSA, the four proficiency groupings were classified on the basis of their performance on the MSA. As explained earlier in this paper, this procedure allowed us to bypass the possible confounding effects on student abilities based on systemic differences between the modes of administration on the Mod-MSA tests, keeping in mind that the MSA on which the proficiency groups were classified is a P&P administered test only.

The DIF items identified by this procedure could then to be used to identify *biased* items (with respect to the testing modes) for *future* test forms development.

## Results

As stated earlier, Grade 3 mathematics and reading were not included in the analyses because these students did not have the corresponding MSA scores from 2009. The main matching criterion was the student's State ID. Based on this criterion the samples for Grades 4 and 5 were adequate for our analysis and are depicted in Table 2, below. The respective mean and standard deviation on the Mod-MSA and the MSA for the students' performance are also displayed in the table. We used the SAS statistical program with the Proc GLM option to obtain the adjusted means and the corresponding F-values for the test of homogeneity of slopes, and significance of the main effect, i.e., the equality of the adjusted means between mode groups

**Table 2: Descriptive Statistics by Grade and Content for the Mod-MSA Students who were Identified as Having a Corresponding Score on the MSA.**

| Subject | Grade | Type | N-Count | Mean Mod-MSA | Std. Dev. Mod-MSA | Mean MSA | Std. Dev. MSA |
|---|---|---|---|---|---|---|---|
| Mathematics | 4 | All | 1184 | 26.97 | 8.31 | 31.77 | 11.56 |
| | | OL | 268 | 26.12 | 7.27 | 32.56 | 10.53 |
| | | P&P | 916 | 27.21 | 8.58 | 31.54 | 11.84 |
| Mathematics | 5 | All | 1290 | 25.12 | 7.36 | 25.98 | 9.66 |
| | | OL | 325 | 23.71 | 6.16 | 25.10 | 8.82 |
| | | P&P | 965 | 25.60 | 7.66 | 26.28 | 9.91 |
| Reading | 4 | All | 1225 | 24.83 | 7.04 | 11.69 | 5.09 |
| | | OL | 276 | 24.61 | 6.42 | 12.90 | 5.43 |
| | | P&P | 949 | 24.90 | 7.21 | 11.34 | 4.94 |
| Reading | 5 | All | 1337 | 24.74 | 6.64 | 16.89 | 5.85 |
| | | OL | 335 | 24.27 | 6.14 | 16.87 | 5.72 |
| | | P&P | 1002 | 24.91 | 6.79 | 16.90 | 5.89 |

***Test-Level Analysis***

In order to ascertain the viability of using ANCOVA as an analytical method we first tested the linear correlation between the covariate (the students' 2009 MSA scores) with the Mod-MSA scores. The results are presented in Tables 3.

**Table 3: Correlation Between the 2009 MSA and the 2010 Mod-MSA scores**

| Subject | Grade | Examination Type | N-Count | Correlation Coefficient Between the 2010 Mod-MSA and the 2009 MSA |
|---|---|---|---|---|
| | | | | |
| Mathematics | 4 | Mod-MSA | 1184 | 0.47 |
| | | MSA | 1184 | - |
| Mathematics | 5 | Mod-MSA | 1290 | 0.46 |
| | | MSA | 1290 | - |
| Reading | 4 | Mod-MSA | 1225 | 0.36 |
| | | MSA | 1225 | - |
| Reading | 5 | Mod-MSA | 1337 | 0.47 |
| | | MSA | 1337 | - |

As can be seen from the above table, the correlations range from a low of 36 to a high of 47. Because of the restriction of range of the Mod-MSA student scores, the correlations may be lower than what would be expected if no attenuation had taken place.

The second consideration in the use of the ANCOVA, as discussed above, was the verification of the assumption of equality of the regression slopes (i.e. to test the testing-mode groups' interaction with the MSA scores). These results are presented in Table 4. As can be seen from the table, the homogeneity of the regression slopes is tenable across all grades and content at the 0.05 significance level.

**Table 4: Assessing the Equality of the Regression Slopes**

| Subject | Grade | Source | DF | F-Value | Pr > F |
|---|---|---|---|---|---|
| Mathematics | 4 | MSA × Mode | 1 | 0.85 | 0.3555 |
| | | | | | |
| Mathematics | 5 | MSA × Mode | 1 | 3.34 | 0.0679 |
| | | | | | |
| Reading | 4 | MSA × Mode | 1 | 0.57 | 0.4513 |
| | | | | | |
| Reading | 5 | MSA × Mode | 1 | 0.07 | 0.7985 |

Based on the homogeneity of the regression slopes, as indicated in the table above, we used the ANCOVA to test the difference between the adjusted means of the two mode-administered groups without having to resort to such techniques as the Johnson-Neyman method to establish the limits of the regions of non-significance on the covariate.

The adjusted means and the main effect significance table are provided in Tables 5 and 6, respectively. Table 6 also provides the magnitude of the difference between the adjusted means (i.e., the effect size (ES) measures).

**Table 5: Adjusted Means of OL and P & P Groups**

| Subject | Grade | Adjusted Mean OL | Adjusted Mean P&P |
|---|---|---|---|
| Mathematics | 4 | 25.85 | 27.30 |
| | | | |
| Mathematics | 5 | 24.02 | 25.50 |
| | | | |
| Reading | 4 | 23.99 | 25.08 |
| | | | |
| Reading | 5 | 24.28 | 24.90 |

**Table 6: The *F*-Test for the Main Effects of ANCOVA: Testing for Equality of the Adjusted Means Between Mode Groups**

| Subject | Grade | N-Count | Source | DF | F-Value | Pr > F | Effect Size (ES) Measure |
|---|---|---|---|---|---|---|---|
| Mathematics | 4 | 1184 | Mode Groups | 1 | 8.12 | 0.0045 | 0.08 |
| | | | | | | | |
| Mathematics | 5 | 1290 | Mode Groups | 1 | 12.57 | 0.0004 | 0.10 |
| | | | | | | | |
| Reading | 4 | 1225 | Mode Groups | 1 | 5.83 | 0.0159 | 0.07 |
| | | | | | | | |
| Reading | 5 | 1337 | Mode Groups | 1 | 2.84 | 0.0922 | 0.05 |

As can be seen from Table 5, the adjusted means for the P&P are higher than the OL for each of the content areas across grades (almost negligible for Grade 5 reading), indicating that on an average, groups that took the P&P performed better than those students who took the OL. Since the main effects are significant (Table 6), we rejected the null of no difference between mode-groups at the predetermined 0.05 level for all grades and content areas except for Grade 5 reading where there was no statistically significant difference between modes of administration.

However, in practical terms, the difference in the adjusted means is small as displayed by the effect-size (ES) measures shown in Table 6. The ES for the main effect was calculated by the following formula (Stevens, 1990, p. 143):

$ES = \sqrt{(k-1)F/N}$ , where

k= level of the groups (which in our case = 2), and the F and N values are those that are shown in Table 6 above. The ES values depicted in the table are small as characterized by Cohen (1977) where and ES of around 0.10 is considered small, around 0.25 as medium, and .0.40 as large (Stevens, 1990, p. 89).

### ***Item-Level Analysis***

Prior to analyzing items for DIF, simple p-value charts (Figure 1 to Figure 4) that reflect each item's performance between the modes of administration by grades and content is provided below. These charts give a general idea on item behavior across modes, keeping in mind that no adjustment was made with respect to the proficiency groupings of students between the two modes, and students were not assigned randomly to the modes of administration.

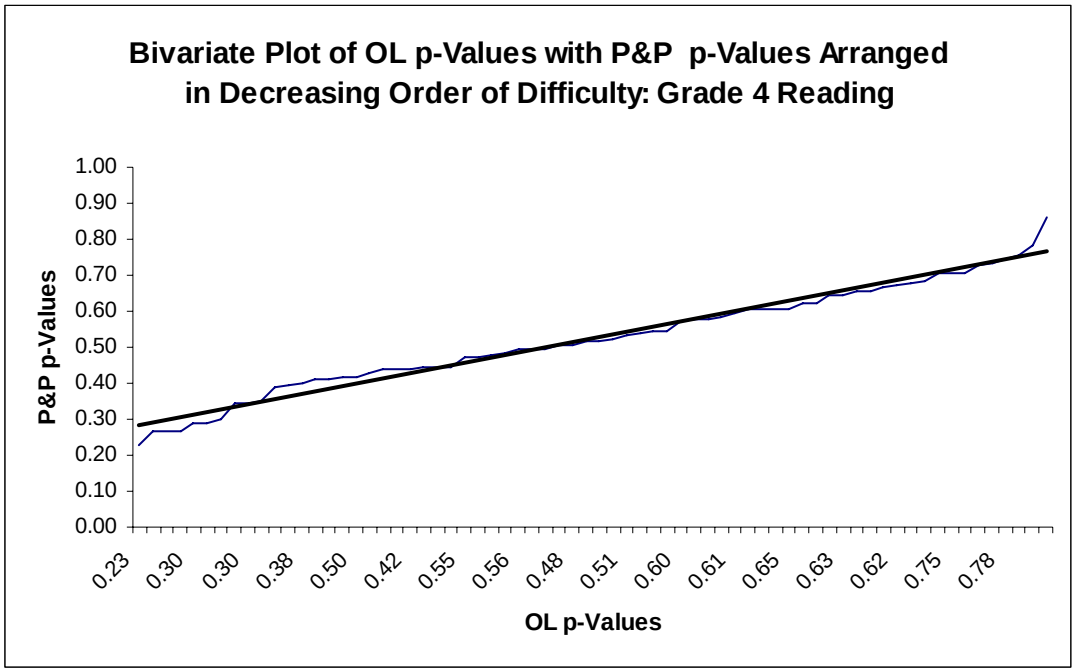



**Figure 1**: Grade 4 reading item p-values by mode of administration

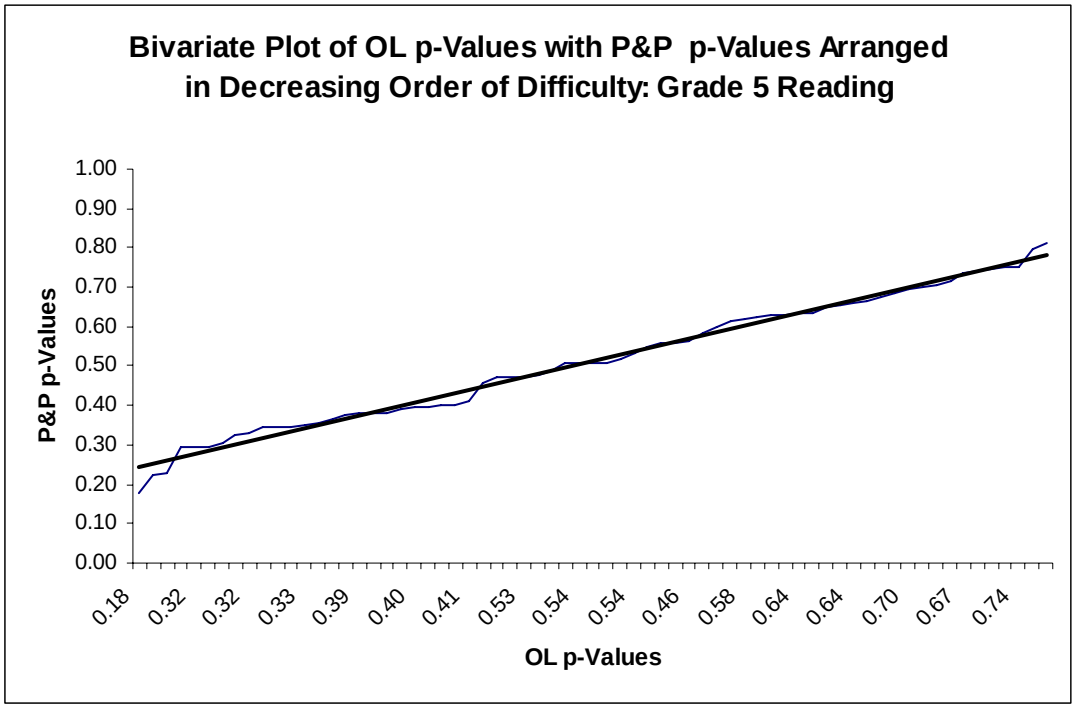



**Figure 2**: Grade 5 reading item p-values by mode of administration

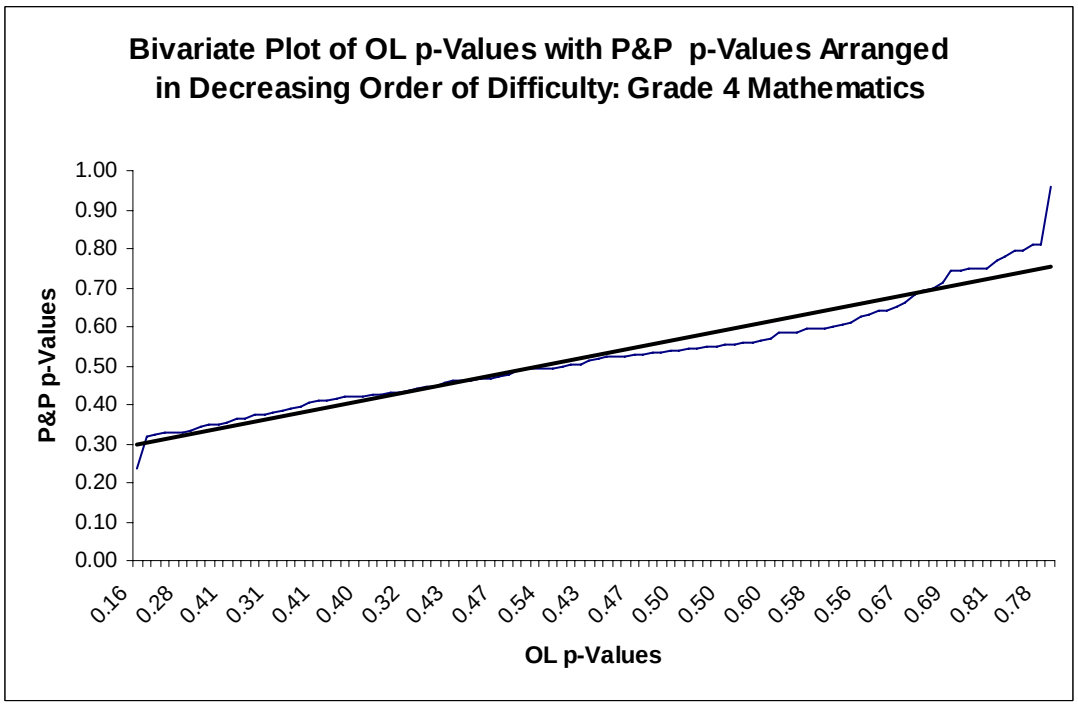



**Figure 3**: Grade 4 mathematics item p-values by mode of administration

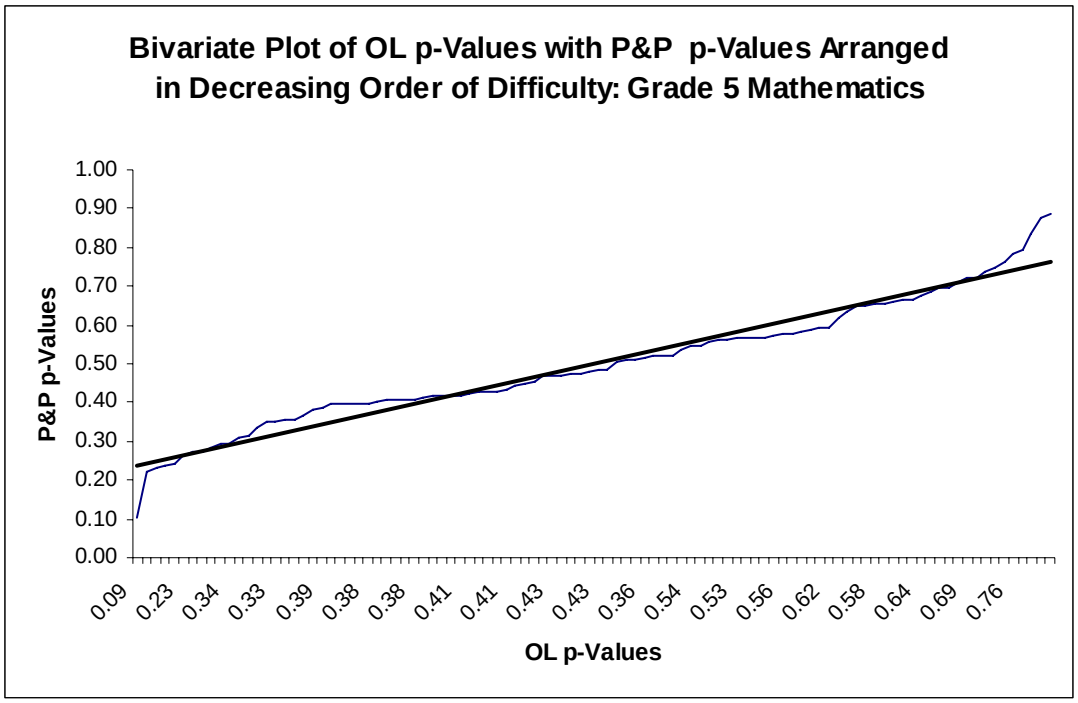



**Figure 4**: Grade 5 mathematics item p-values by mode of administration

By using the 2009 MSA scores as an external variable for matching the mode-administered groups, we found no extreme category, "C" DIF for any of the items (see Table 7, below), which included both operational items and those items that were not used as core items. For this analysis, the SAS program was once again used to calculate the M-H chi-square significance and the effect type measure of the delta scale.

**Table 7: DIF Classification of Flagged Items by Content and Grade**

| Subject | Grade | Item Sequence No. | Item CID No. | M-H Chi-Square | Chi-Square Probability | Delta Scale | DIF Category |
|---|---|---|---|---|---|---|---|
| Mathematics | 4 | 67 | 100000186573 | 9.85 | 0.002 | 1.10 | -B |
| Mathematics | 4 | 84 | 100000198121 | 11.68 | 0.001 | 1.17 | -B |
| | | | | | | | |
| Mathematics | 5 | 73 | 100000196094 | 21.13 | 0.00 | 1.42 | -B |
| | | | | | | | |
| Reading | 4 | 17 | 100000357136 | 10.76 | 0.001 | 1.07 | -B |
| | | | | | | | |
| Reading | 5 | 12 | 100000102095 | 12.38 | 0.000 | 1.05 | -B |

Note: + = in favor of P&P and - = in favor of OL.

All the items that were administered and scored for Grades 4 and 5 reading and mathematics were used in the DIF analysis (i.e., a total of 337 items broken down by 102 items in Grade 4 and 100 items in Grade 5 mathematics, and 68 and 67 items for reading in Grades 4 and 5 respectively). As shown in Table 7, there were two category "B" DIF classifications for mathematics for Grade 4. The remaining grades in reading had one each with "B" classification. All the items with DIF were in favor of OL.

## Comments and Conclusion

The methods described in this study can be seen as two approaches to test the same null hypothesis of no examinee differences in student performance between test modes. However, the results of the two methods have different implications in assessing the impact of testing modes on students who are administered the Mod-MSA. The ANCOVA provides an overall view of test-mode effects by considering the differences between the test-mode groups in terms of the *total* test performance. As such, the results from the analysis can be seen as the total of item effects.

The DIF analysis, on the other hand, tests the hypothesis of no difference between testing-mode groups at the item level. In a sense, the two approaches complement each other by the analysis of individual item behavior as in DIF and the total item behavior as in the ANCOVA.

In our analysis, statistically significant differences were found at three out of four grade levels within a content area for the ANCOVA main effect. It was also found that the differences on average were in favor of those who took the P&P (i.e. the test as a whole with the exception of Grade 5 reading, was slightly harder for OL test takers in comparison to those who took the P&P testing mode). The significant differences, however, could be attributed partly to the attenuation

effects for the Mod-MSA students. Greater precision of estimate would have been possible if the correlation between the independent and the dependent variable was not underestimated.

As we had discussed earlier, the assignment of students to the mode of administration was not random. In such cases, the ANCOVA (as in most other statistical analyses) has an important limitation (Anderson, 1963; Lord, 1969) that needs to be addressed. As Stevens (1990, p.168) points out: "even the use of several covariates will not equate intake groups, and one should not be deluded into thinking that it can. The groups may still differ on some unknown important variable(s)."

In this study, it is quite possible that a non-modeled variable(s) could have had an impact on the groups in question. For example, it is likely that school and student variables (e.g., the degree of schools' encouragement in the use of technology, student non-familiarity with computer testing, etc.) *may* have had an effect on student achievement. Future studies *modeling* these variables may provide some explanation of these hypothetical concerns.

However, it is important to avoid placing too much emphasis on these statistically significant results, as the actual differences between the adjusted means were small as measured by their effect sizes (ES). The variability of the adjusted group means about the grand mean as shown in Table 6 is small. Because of the large sample sizes with respect to the ANCOVA, the least amount of practically insignificant difference (e.g., differences so small as to have a negligible affect on student scores) between groups can show up as being statistically significant. It, therefore, makes sense to examine ES measures as a pragmatic approach in the comparison of mode effects for the Mod-MSA.

It is encouraging to note that the results of this study indicate that all the items used for the 2010 administration for Grades 4 and 5 across the two content areas did not show extreme DIF between modes of administration. The *moderate* DIF shown for a total of a mere five items from 337 items across grades and content areas (far fewer than would be expected by chance at alpha = 0.05) can be scrutinized for mode bias by content specialists.

The very small effect sizes and the relative absence of DIF suggest the viability of using P&P as a replacement for on-line administrations when needed. However, the MSDE is encouraged to continue mode DIF analysis for new items in the future to the extent that the availability of data makes such analysis possible.

## References


Anderson, N. H. (1963). Comparison of different populations: Resistance to extinction and transfer. *Psychological Bulletin*, *70*, 162-179.

Braun, H., Jenkins, F., & Grigg, W. (2006). *Comparing private schools and public schools using hierarchical linear modeling*. (NCES 2006-461). Washington D.C: U. S. Department of Education.

Camilli, G., & Shepard, L. (1994). *MMS: methods for identifying biased test items.* Sage.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Eignor, D. R. (2007). Linking scores derived under different modes of test administration. In Dorans, N. J. , Pommerich, M., & Holland, P. W. (Eds.) *Linking and aligning scores and scales* (pp. 135-137). Springer, N.Y.

Garson, G. David (n.d.). Linear Mixed Models: Random effects, hierarchical linear, multilevel random coefficients, and repeated measure models, from *Statnotes: Topics in multivariate analysis.* Retrieved 06/26/2008 from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

Glasnapp, D. R., Poggio, J., Poggio, A., & Yang, X. (2005). *Student attitudes and perceptions regarding computerized testing and the relationship to performance in large scale assessment programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Thousand Oaks, CA: Sage.

Hoffmann, D. A. (1991). An overview of the logic and rationale of hierarchical linear Models. *Journal of Management, 23*(6), 723-744.

Huang, M., & Lu, E. (2007). *The two-level sample size problem of hierarchical linear modeling: Evidence from simulation experiments*. Paper presented at the annual meeting of the American Political Science Association, Chicago, IL.

Kim, D. H., & Huynh, H. (2008). Computer-based and paper-and-pencil administration mode effects on a statewide end-of-course English test. *Educational and Psychological Measurement, 68(4)*, 554-570.

Kim, D. H., & Huynh, H. (2009). Transition from paper-and-pencil to computer-based testings: examining stability of Rasch latent trait across gender and ethnicity. To appear as a book chapter in Everett V. Smith Jr. & Greg E. Stone (Eds.) *Applications of Rasch measurement in criterion-referenced testing: Practice analysis to score reporting*. Maple Grove, MN: JAM Press.

Keng, L., McClarity, K. L., & Davis, L. L. (2006). *Item-level comparability analysis of on-line and paper administrations of the Texas Assessment of Knowledge and Skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating: methods and practices.* New York, NY: Springer-Verlag.

Linacre, J. M. (2001). *WINSTEPS Rasch measurement program, Version 3.32*. Chicago: John M. Linacre.

Lord, F. M. (1969). Statistical adjustments when comparing pre-existing groups. *Psychological Bulletin*, *70*, 162-179.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute, 22*, 719-748.

Pedhazur, E. J., (1973). *Multiple regression in behavioral research*. Forth Worth, TX: Harcourt Brace Jovanovich.

Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. J. (2005). A comparative evaluation of scored results from computerized and Paper and pencil mathematics testing in a large scale state assessment program. *Journal of Technology, Learning and Assessment*, *3*(6).

Raudenbush, S. W. (1988) Educational applications of hierarchical Models: a review. *Journal of Educational Statistics, 13*, 85-116.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (pp. 238-345). Hillsdale, NJ: Lawrence Erlbaum.

U.S. Department of Education. Fed. Reg. 34 CFR Parts 200 and 300 (April 9, 2007). *Title I-Improving the academic achievement of the disadvantaged; Individual of Disabilities Education Act, Final rule*.

Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006). *Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*(3), 233-251.