The Modified Maryland School Assessment (Mod-MSA) Mathematics:

Grades 3 through 8

Technical Report:
2010 Administration

TABLE OF CONTENTS

Table of Contents **1**

List of Tables **4**

List of Figures **7**

Acknowledgements **8**

Overview of this Manual **9**

1. Introduction **11**

1.1. Background 11
1.2. Rationale for the 2010 Mod-MSA: Mathematics 12
1.3. Eligibility Criteria for the Mod-MSA: Mathematics 12
1.4. Test Administration of the 2010 Mod-MSA: Mathematics 12
Test Materials 12
Test Administration Schedule 13
Student Participation 14
Accommodations for Assessment 14
Test Format 15
Security of Test Materials 16
1.5. Quality Control Procedures 16
1.6. Item Bank Construction 18

2. Test Design and Development of the 2010 Mod-MSA: Mathematics **19**

2.1. Test Design and Structure of the 2010 Mod-MSA: Mathematics 19
2.2. Development and Review of the 2010 Mod-MSA: Mathematics 20
National Psychometric Council 21
Content Review Committee 21
Bias Review Committee 21
Vision Review Committee 21
2.3. Operational Form Structure of the 2010 Mod-MSA: Mathematics 21
2.4. Procedures Undertaken for Item Analyses Prior to the Creation of the Operational Form 23
Classical Item Analyses 24
Differential Item Functioning Analyses 24
Item Response Theory (IRT) Analyses 27
2.5. Items Flagged for Inspection Prior to the Creation of the Operational Forms 27
2.6. Items Selected for the 2010 Operational Tests 27
2.7. Scoring Procedures of the 2010 Mod-MSA: Mathematics 29

3. Central Tendency Measures and Other Classical Item Analyses of the 2010 Mod-MSA: Mathematics **30**

3.1. Measures of Central Tendency 30
Standard Error of the Mean 30
3.2. Item Level Descriptive Statistics 32

4. Scale Creation, Equating and Raw Scores to Scale Scores Conversion via Item Response Theory (IRT) Procedures **51**

The Rasch Model 51
4.1. Calibration and Scaling Procedures for Grades 3 to 5 53
4.2. Specifics for Creating the Base Scale for the Mod-MSA: Mathematics Grades 3-5 53
4.3. Calibration and Equating the 2010 Mod-MSA: Mathematics Grades 6-8 54
4.4. Specifics of Linking and Equating the 2010 Mod-MSA Grades 6-8: Mathematics 54
4.5. Reporting Scale Scores for the 2010 Mod-MSA: Mathematics 59
Raw Score to Scale Score Conversion Tables for the Total Score 61
Raw Score to Scale Score Conversion Tables for the Subscales 73
4.6. Score Interpretation 85
Scale Scores 85
Performance Level Descriptors 85
4.7. Final Performance Level Cut Points for the Mod-MSA: Mathematics 85
5. Comparative Classical and IRT Stats Across Years for the Common Items Used in Grades 6-8. 87
6. IRT Item Level Statistics for the 2010 Mod-MSA: Mathematics 94
6.1 Rationale for the Use of the Rasch Model 94
6.2 Fit Statistics for the Rasch Model 94
6.3 Rasch Item Level Statistics 95
7. Test Reliability 114
7.1. Precision and Reliability (Classical Methods) 114
Standard Error of Measurement (SEM) of the Test 114
Cronbach's Alpha (KR_{20}) 115
7.2. IRT Method in Measuring Precision of the Test 115
7.3. Decision Accuracy and Consistency at the Cut Scores 116
8. Test Validity 118
8.1. Test Validity for the 2010 Mod-MSA: Mathematics 118
Content-Related Evidence 118
Evidence from Item Development Methods 118
Evidence from Internal Structure of the Tests 121
8.2. Unidimensionality Analysis for the 2010 Mod-MSA: Mathematics 123
9. Summary of Operational Test Results for the 2010 Mod-MSA: Mathematics 125
9.1 Classical Descriptive Test Statistics (Raw Scores) 125
9.2 Scale Score Descriptive Statistics 125
9.3 Frequency of Students at Each Proficiency Level by Grade 126
9.4. Test Characteristic Curves, Test Information Functions, and Conditional Standard Errors of Measurement 127
10. References 141
Appendix A: Operational Items Selected After Data Review 143
Appendix B: Item Distractor Analyses 150
Appendix C: Frequency Distribution Histograms of Scale Scores 169
Appendix D: Standard Setting Report 176

Appendix E: Alignment of the 2009 Mod-MSA: Mathematics Items to the State Standards 256

Appendix F: Identifying Students for Participation in the Mod-MSA: Mathematics Program 263

ppendix G: Comparability Study of Paper and Pencil, and Online Administration of the Mod-MSA 265

LIST OF TABLES

Table 2.1.1 Test Design for the Mod-MSA, Mathematics, Grades 3-5 19
Table 2.1.2 Test Design for the Mod-MSA, Mathematics, Grades 6-8 20
Table 2.2.1. Responsibility for the 2009 Mod-MSA Test Development 20
Table 2.3.1. The 2010 Mod-MSA, Mathematics Operational Form with Maximum Points Obtainable Per Strand: Grades 3 to 8 23
Table 2.4.1. DIF Classification 26
Table 2.5.1. Summary Stats Used in the Development of the 2010 Mod-MSA Mathematics Operational Forms 27
Table 2.6.1. Number of Items Included as Operational Items with 0 < PB < .10, and flagged Omit Items by Grades 28
Table 3.1.1. Central Tendency Statistics of the Operational Test by Subscales and Grades 31
Table 3.2.1. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 3 33
Table 3.2.2. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 4 36
Table 3.2.3. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 5 39
Table 3.2.4. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 6 42
Table 3.2.5. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 7 45
Table 3.2.6. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 8 48
Table 4.4.1. Unequated Core Linking Item Difficulties of Year 2009 vs. Year 2010: Grade 6 56
Table 4.4.1. Unequated Core Linking Item Difficulties of Year 2009 vs. Year 2010: Grade 6 57
Table 4.4.2. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 7 58
Table 4.4.3. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 8 59
Table 4.5.1. The 2010 Mod-MSA, Mathematics Slope and Intercept for the Transfer of RS to SS Across Grades 60
Table 4.5.2. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 3 61
Table 4.5.3. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 4 63
Table 4.5.4. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 5 65
Table 4.5.5. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 6 67
Table 4.5.6. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 7 69
Table 4.5.7. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 8 71
Table 4.5.8. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 3 73
Table 4.5.9. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 4 75
Table 4.5.10. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 5 77
Table 4.5.11. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 6 79

Table 4.5.12. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 7 81
Table 4.5.13. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 8 83
Table 4.7.1 Mod-MSA: Mathematics Scale Score Cuts by Grades 86
Table 5.1. P-Value and Unequated Rasch Difficulty Comparisons of Core Linking Items Year 2009 with Year 2010: Grade 6 88
Table 5.2. P-Value and Rasch Difficulty Comparisons of Core Linking Items Year 2009 with Year 2010: Grade 7 90
Table 5.3. P-Value and Rasch Difficulty Comparisons of Core Linking Items Year 2009 with Year 2010: Grade 8 92
Table 6.2.1. Criteria to Evaluate Mean-Square Fit Statistics 95
Table 6.3.1. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 3 96
Table 6.3.2. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 4 99
Table 6.3.3. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 5 102
Table 6.3.4. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 6 105
Table 6.3.5. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 7 108
Table 6.3.6. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 8 111
Table 7.3.1. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 3 116
Table 7.3.2. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 4 116
Table 7.3.3. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 5 117
Table 7.3.4. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 6 117
Table 7.3.5. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 7 117
Table 7.3.6. The 2010 Mod-MSA: Mathematics Decision Accuracy and Consistency Indices: Grade 8 117
Table 8.1.1. Category Classification of Items that Showed Moderate or Extreme DIF by Grades 119
Table 8.1.2. The 2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 3 121
Table 8.1.3. The 2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 4 121
Table 8.1.4. The 2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 5 121
Table 8.1.5. The 2009/2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 6 122
Table 8.1.6. The 2009/2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 7 122
Table 8.1.7. The 2009/2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 8 123
Table 8.2.1. Eigenvalues between the First and Second Components of the Mod-MSA: Mathematics .. 124
Table 9.1.1 Classical Descriptive Statistics for the 2010 Mod-MSA: Mathematics: Grades 3 through 8 125
Table 9.2.1. Scale Score Descriptive Statistics Across Years for Mod-MSA, Mathematics: Grades 3-8 126
Table 9.3.1. Percentage of Students in Each Proficiency Level by Grade 126
Table A1. Grade 3 Operational Items 144
Table A2. Grade 4 Operational Items 145
Table A3. Grade 5 Operational Items 146
Table A4. Grade 6 Operational Items 147
Table A5. Grade 7 Operational Items 148
Table A6. Grade 8 Operational Items 149
Table B.1. The 2010 Mod-MSA: Mathematics Grade 3, Distractor Distribution and Distractor-to- Total Correlation Analysis 151
Table B.2. The 2010 Mod-MSA: Mathematics Grade 4, Distractor Distribution and Distractor-to- Total Correlation Analysis 154

Table B.3. The 2010 Mod-MSA: Mathematics Grade 5, Distractor Distribution and Distractor-to- Total Correlation Analysis 157
Table B.4. The 2010 Mod-MSA: Mathematics Grade 6, Distractor Distribution and Distractor-to- Total Correlation Analysis 160
Table B.5. The 2010 Mod-MSA: Mathematics Grade 7, Distractor Distribution and Distractor-to- Total Correlation Analysis 163
Table B.6. The 2010 Mod-MSA: Mathematics Grade 8, Distractor Distribution and Distractor-to- Total Correlation Analysis 166

LIST OF FIGURES

Figure 4.1 Item Characteristic Curve 51
Figure 4.2 Category Response Curves for a One-Step Item 52
Figure 4.4.1. Anchor Evaluation Steps Chart for Mod-MSA 56
Figure 5.1. P-values of the common items for Grade 6, years 2009 and 2010 89
Figure 5.2. Rasch difficulty parameters for the common items, Grade 6, years 2009 and 2010. 89
Figure 5.3. P-values of the common items for Grade 7, years 2009 and 2010 91
Figure 5.4. Rasch difficulty parameters for the common items, Grade 7, years 2009 and 2010.. 91
Figure 5.5. P-values of the common items for Grade 8, years 2009 and 2010 93
Figure 5.6. Rasch difficulty parameters for the common items, Grade 8, years 2009 and 2010.. 93
Figure 9.4.1a. TCC for Grade 3 Mod-MSA: Mathematics 129
Figure 9.4.1c. CSEM for Grade 3 Mod-MSA: Mathematics 130
Figure 9.4.2a. TCC for Grade 4 Mod-MSA: Mathematics 131
Figure 9.4.2b. TIF for Grade 4 Mod-MSA: Mathematics 131
Figure 9.4.2c. CSEM for Grade 4 Mod-MSA: Mathematics 132
Figure 9.4.3a. TCC for Grade 5 Mod-MSA: Mathematics 133
Figure 9.4.3b. TIF for Grade 5 Mod-MSA: Mathematics 133
Figure 9.4.3c. CSEM for Grade 5 Mod-MSA: Mathematics 134
Figure 9.4.4a. TCC Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010) 135
Figure 9.4.4b. TIF Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010)……. 135
Figure 9.4.4c. CSEM Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010) 136
Figure 9.4.5a. TCC Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010) 137
Figure 9.4.5b. TIF Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010) 137
Figure 9.4.5c. CSEM Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010) 138
Figure 9.4.6a. TCC Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010) 139
Figure 9.4.6b. TIF Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010) 139
Figure 9.4.6c. CSEM Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010) 140

ACKNOWLEDGEMENTS

The 2010 *Modified Maryland School Assessment (Mod-MSA): Mathematics 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, Ray Graleski, Paul Katula, Julia Walters, and Leslie Wilson.

OVERVIEW OF THIS MANUAL

The *Modified Maryland School Assessment Program* (Mod-MSA) *Technical Manual* for the 2010 Mathematics 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 the Mod_MSA Mathematics 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: Mathematics

This section describes the test design, development, and the scoring process of the Mod-MSA: Mathematics 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: Mathematics.

Section 3: Central Tendency Measures and Other Classical Item Analyses of the 2010 Mod-MSA: Mathematics

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: Mathematics (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 stats.

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

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 ends with 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 2009 Mod-MSA: Mathematics, 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: Mathematics

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: Mathematics 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-Math Recalibration Results from 3PL IRT to the Rasch Model Using Equipercentile Method* in the 2007 *MSA-Math 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 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) 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: Mathematics

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 2009 Mod-MSA: Mathematics fulfills two main purposes. First, the Mod-MSA: Mathematics 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: Mathematics 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: Mathematics

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

1.4. Test Administration of the 2010 Mod-MSA: Mathematics

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 they needed for testing.

Test-related examiners manuals (EMs) were developed for the 2010 Mod-MSA: Mathematics for use in all grades 3–8. Developed in partnership with MSDE, the EMs contained instructions for preparation and administration of the test. In addition to the EMs, one test administration and coordination manual (TACM) was developed for use 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- Mathematics

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
- Blank scratch paper
- Calculator (all grades)
- Classroom ruler with both U.S. customary and metric measurements (all grades)
- Classroom protractor for grades 6 through 8

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: Mathematics 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 mathematics 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: Mathematics 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 26–50 minutes per test section. All mathematics items tested on Day 1 permitted students to use calculators while Day 2 items did not allow the use of calculators.

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: Mathematics, 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 accommodation had to be based on individual needs and not on a category of disability area, level of instruction, environment, or other group characteristics. Responsibility for confirming the need and appropriateness of an accommodation rested with the LAC and school-based staff involved with each student's instructional program. A master list of all students and their accommodations had to be maintained by the principal and submitted to the LAC, who provided a copy to MSDE upon request. Please refer to Section 1 of the 2010 TACM for further information regarding testing accommodations.

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

The Mod-MSA: Mathematics 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-scorable 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 2010 TACM in section 4.

Verbatim Reading Accommodation and Kurzweil™ Test Form on CD

Students who had a verbatim reading accommodation documented in their Individual Education Plan (IEP), ELL Plan, or Section 504 Plan—and who received that accommodation in regular instruction—received the accommodation on the 2010 Mod-MSA: Mathematics. The accommodation was provided by a live reader or through technology. Section 1 of the 2009 TACM provided information on verbatim reading instruction. Technology used to provide the verbatim reading accommodation was Kurzweil™ reading software. Official, secure electronic copies of the test were ordered through the LAC. MSDE encouraged (but did not require) the use of the Kurzweil™ software to ensure uniformity in the delivery of the verbatim reading accommodation throughout the state.

Students using Kurzweil™ software had to familiarize themselves with its operation prior to the test administration. When there were technical difficulties with Kurzweil™ 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 the direct supervision of a certified TE, into the student's Test/Answer Book with a pre-printed or generic ID label attached.

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

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 (2008):

> 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, 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 demographic 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 update of files

The Maryland test deck was the first order processed through the Maryland system to verify all aspects of the materials packaging, scanning, editing, scoring, summary, and reporting. Pre-determined conditions were included in the test deck to assure the programs were processing all data to meet the requirements of the program with zero defects. Processing of live orders could not proceed until each phase of the test deck had been approved by 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 specialist and psychometrician. 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 student
- Statewide frequency distribution of demographic variables
- District/Building/N-count
- Statewide raw and scale cut score tables
- The use of Proc Means to verify summary statistics
- Item response listing to verify all constructed responses were scored and within the valid range
- Normative data check for all raw scores
- Reader resolution report to verify all readings and resolution combinations

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

1.6. Item Bank Construction

The Mod-MSA: Mathematics 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
- 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 AND DEVELOPMENT OF THE 2010 MOD-MSA: MATHEMATICS

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

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

The 2010 Mod-MSA: Mathematics grades 3 to 8 was administered as a single form, which included more than the required number of operational items. The administration included a mix of operational items (i.e., items on which the students were scored) and some field-test [FT] items. Three items for grade 7 were not scored because of context and/or content irregularities that did not meet with the expectations of MSDE. The selection of the operational items was determined through the data review process only after the test was administered. In other words, 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 mathematics, grades 3 to 5, were augmented from the MSA items.

Table 2.1.1 Test Design for the Mod-MSA, Mathematics, 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	0	-	-
	Modified from Previous MSA and/or Modified MSA Bank Items	100	51	49
	Total	**100**	**51**	**49**
4	Intact from Previous MSA	0	-	-
	Modified from Previous MSA and/or Modified MSA Bank Items	102	51	51
	Total	**102**	**51**	**51**
5	Intact from Previous MSA	0	-	-
	Modified from Previous MSA and/or Modified MSA Bank Items	100	51	49
	Total	**100**	**51**	**49**

Note: 1. 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 Grades 6 to 8 2010 tests on the established 2009 scale. There were 25 common items in each of these three grades.

Table 2.1.2 Test Design for the Mod-MSA, Mathematics, 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	25	25	-
	Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items	71	26	45
	Total	**96**	**51**	**45**
7	Linking Items from 2009	25	25	-
	Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items	73-3=70 items (3 items were DNU	26	44
	Total	**95**	**51**	**44**
8	Linking Items from 2009	25	25	-
	Modified from Previous MSA and/or Modified MSA Bank Items and/or new 2010 Items	73	26	47
	Total	**98**	**51**	**47**

Note: 1. The total number of items is the sum of the operational and field-test items.
2. DNU is the abbreviation for "Do Not Use."

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

Developing the 2010 Mod-MSA: Mathematics 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: Mathematics.

Table 2.2.1. Responsibility for the 2009 Mod-MSA Test Development

Development of the 2009 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: Mathematics 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: Mathematics was appropriately difficult and fair. Committee members were either specialists in mathematics 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 mathematics 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 mathematics 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: Mathematics

The 2010 Mod-MSA in mathematics only had selected response (SR) items (multiple-choice) with two distractors (aside from the correct choice) for each item. As shown in Table 2.2.1, there were 25 items in grades 6 to 8 that were used as common items from 2009 for linking the 2010 assessment to the established scale. The rest of the mathematics items were either modified from the MSA items (grades 3 -5) or were modified versions of the MSA items plus some new developed items (grades 6 to 8). These items required students to select a correct answer from the various alternatives. Each SR item was scored dichotomously (i.e., 0 or 1).

The Mod-MSA: Mathematics was organized under the following content strands for each of the six grades, 3 through 8:

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

These strands were combined to match the same five strands as those reported by the mathematics MSA. For the Mod-MSA: Mathematics, therefore, the final reporting strands were:

1) Algebra
2) Geometry and Measurement
3) Statistics and Probability
4) Numbers and Computation
5) Process

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

Table 2.3.1. The 2010 Mod-MSA, Mathematics Operational Form with Maximum Points Obtainable Per Strand: Grades 3 to 8

Grade	Strand Title	No. of Items	Item Type	Reporting Strand	Reporting Score
3	Total Test	51	*SR*	Total Test	51
	Algebra	9	*SR*	Algebra	9
	Geometry	6	*SR*	Geometry and Measurement	11
	Measurement	5	*SR*		
	Statistics	8	*SR*	Statistics and Probability	10
	Probability	2	*SR*		
	Number Computation	11	*SR*	Number Computation	11
	Process	10	*SR*	Process	10
4	Total Test	51	*SR*	Total Test	51
	Algebra	10	*SR*	Algebra	10
	Geometry	5	*SR*	Geometry and Measurement	10
	Measurement	5	*SR*		
	Statistics	6	*SR*	Statistics and Probability	11
	Probability	5	*SR*		
	Number Computation	10	*SR*	Number Computation	10
	Process	10	*SR*	Process	10
5	Total Test	51	*SR*	Total Test	51
	Algebra	10	*SR*	Algebra	10
	Geometry	4	*SR*	Geometry and Measurement	10
	Measurement	6	*SR*		
	Statistics	6	*SR*	Statistics and Probability	9
	Probability	3	*SR*		
	Number Computation	10	*SR*	Number Computation	10
	Process	12	*SR*	Process	12
6	Total Test	51	*SR*	Total Test	51
	Algebra	10	*SR*	Algebra	10
	Geometry	6	*SR*	Geometry and Measurement	10
	Measurement	4	*SR*		
	Statistics	7	*SR*	Statistics and Probability	10
	Probability	3	*SR*		
	Number Computation	10	*SR*	Number Computation	10
	Process	11	*SR*	Process	11
7	Total Test	51	*SR*	Total Test	51
	Algebra	10	*SR*	Algebra	10
	Geometry	5	*SR*	Geometry and Measurement	9
	Measurement	4	*SR*		
	Statistics	5	*SR*	Statistics and Probability	10
	Probability	5	*SR*		
	Number Computation	10	*SR*	Number Computation	10
	Process	12	*SR*	Process	12
8	Total Test	51	*SR*	Total Test	51
	Algebra	11	*SR*	Algebra	11
	Geometry	5	*SR*	Geometry and Measurement	9
	Measurement	4	*SR*		
	Statistics	6	*SR*	Statistics and Probability	10
	Probability	4	*SR*		
	Number Computation	8	*SR*	Number Computation	8
	Process	13	*SR*	Process	13

2.4. Procedures Undertaken for Item Analyses Prior to the Creation of the Operational Form

The 2010 Mod-MSA: Mathematics 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 (with acceptable statistics) together with the 2010 operational items could be used as common linking items in 2011 in order to place the 2011 examinations on the established 2009/2010 scales.

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

- Classical item analyses
- 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 flagged at > 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 mathematics 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: Mathematics was a single-format examination, comprising only of 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):

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

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.

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:

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

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).

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

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: $\beta_{M-H} = \ln(\alpha_{M\text{-}H})$. β_{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* χ^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* χ^2 or $\|D\| < 1.0$
B	Weak DIF	Significant *M-H* χ^2 and $\|D\| < 1.5$ or Non-significant *M-H* χ^2 and $\|D\| > 1.0$
C	Strong DIF	Significant *M-H* χ^2 and $\|D\| \geq 1.5$

The groupings for the DIF analysis were based on matching students' scores on the Mod-MSA: Mathematics. Four proficiency groupings of the Mod-MSA students were formed at quarter intervals of the total Mod-MSA: Mathematics score. The performance on the Mod-MSA: Mathematics 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 Mathematics Operational Forms

Grade	Total # of Items	Items DNU[1] by MSDE	DIF Flag B (for check only)	DIF Flag C	PB Flag <= 0.10 but > 0[2]	PB Flag < = 0 (Cannot be used)	Items Rejected (DNU by MSDE + C DIF + PB <=0 Flag)	Items Available for Use for Operational Form Building on Statistical Criterion	Items Needed for Each Operational Form
3	100	0	14	1	5	0	1	99	51
4	102	0	6	1	3	3	4	98	51
5	100	0	4	2	15	2	4	96	51
6	96	0	11	2	4	3	5	91	51
7	98	3	11	0	11	3	6	92	51
8	98	0	7	2	9	2	4	94	51

Note: 1.. DNU is a terms used for "Do not use".
2. 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 earlier, the selection of items that were included in the final operational test forms of the 2010 Mod-MSA: Mathematics 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 had 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 eight items, one each in Grades 3, 4, and 2 items each in Grades 5, 6 and 8 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, 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. There was also one common item in Grade 7 that had a zero point biserial but was used because aside from not having a corresponding item to take its place, it also belonged to a suit of four items. None of the omit flags, however, were for more than 5.73% 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	2	0
4	0	0
5	4	0
6	1	1
7	4 (includes 2 common items)	0
8	2	0

As shown in Table 2.5.1, Grade 7 had two common items that had poor PBs. However, they could not be dropped because they were a part of a suite and also because there were no corresponding field-test items to replace them. These poor PB items in Grade 7 do not include the one common item discussed above that had a zero PB.

2.7. Scoring Procedures of the 2010 Mod-MSA: Mathematics

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 OTHER CLASSICAL ITEM ANALYSES OF THE 2010 MOD-MSA: MATHEMATICS

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	Algebra	9	9	961	4.96	1.92	0.06
	Geometry and Measurement	11	11	961	5.89	2.16	0.07
	Statistics and Probability	10	10	961	4.80	2.09	0.07
	Number and Computation	11	11	961	6.12	2.47	0.08
	Process	10	10	961	4.63	2.05	0.07
	Total Test	51	51	961	26.40	8.39	0.27
4	Algebra	10	10	1294	5.15	2.11	0.06
	Geometry and Measurement	10	10	1294	5.17	2.04	0.06
	Statistics and Probability	11	11	1294	5.83	2.33	0.06
	Number and Computation	10	10	1294	5.56	2.15	0.06
	Process	10	10	1294	5.14	2.06	0.06
	Total Test	51	51	1294	26.85	8.25	0.23
5	Algebra	10	10	1430	5.50	2.07	0.05
	Geometry and Measurement	10	10	1430	4.68	1.90	0.05
	Statistics and Probability	9	9	1430	4.38	1.94	0.05
	Number and Computation	10	10	1430	4.59	1.94	0.05
	Process	12	12	1430	5.82	2.25	0.06
	Total Test	51	51	1430	24.97	7.38	0.20
6	Algebra	10	10	1477	5.24	2.08	0.05
	Geometry and Measurement	10	10	1477	3.86	2.11	0.05
	Statistics and Probability	10	10	1477	4.71	2.11	0.06
	Number and Computation	10	10	1477	4.87	2.09	0.05
	Process	11	11	1477	4.98	2.19	0.06
	Total Test	51	51	1477	23.66	8.04	0.21
7	Algebra	10	10	1755	5.14	2.12	0.05
	Geometry and Measurement	9	9	1755	3.99	1.77	0.04
	Statistics and Probability	10	10	1755	4.32	1.97	0.05
	Number and Computation	10	10	1755	5.64	2.08	0.05
	Process	12	12	1755	4.87	1.90	0.05
	Total Test	51	51	1755	23.96	6.87	0.16
8	Algebra	11	11	1920	5.65	2.24	0.05
	Geometry and Measurement	9	9	1920	3.58	1.65	0.04
	Statistics and Probability	10	10	1920	4.28	1.99	0.05
	Number and Computation	8	8	1920	4.18	1.63	0.04
	Process	13	13	1920	5.72	2.38	0.05
	Total Test	51	51	1920	23.40	7.21	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: Mathematics spring 2010 within the framework of 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 proficiency 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: Mathematics based on the equating sample are presented in Table 3.3.1 to Table 3.3.6. 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 number
- N-count (number of students)
- Response options
- *P*-value for multiple choice items (percentage of examinees that answered the item correctly), and item mean for constructed-response items (average number of points earned out of the maximum number of possible points)
- 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: Mathematics Classical Item Statistics: Grade 3

Item Seq. No.	Item UIN	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
1	100000197601	246	669	3	0.77	0.30
2	100000197604	246	669	3	0.50	0.34
3	100000098438	241	669	3	0.44	0.19
4	100000098445	241	669	3	0.59	0.37
5	100000098439	241	669	3	0.55	0.36
6	100000098516	268	669	3	0.58	0.42
7	100000185401	273	669	3	0.35	0.24
8	100000185403	273	669	3	0.38	0.17
9	100000197723	257	669	3	0.54	0.24
10	100000197745	257	669	3	0.57	0.39
11	100000197780	262	669	3	0.46	0.17
12	100000197781	262	669	3	0.53	0.32
13	100000098529	268	669	3	0.38	0.13
14	100000197752	257	669	3	0.41	0.30
15	100000197753	273	669	3	0.31	0.28
16	100000197754	273	669	3	0.27	0.09
17	100000197620	246	669	3	0.70	0.22
18	100000197647	246	669	3	0.55	0.29
19	100000098440	241	669	3	0.68	0.26
20	100000098448	241	669	3	0.63	0.38
21	100000185375	268	669	3	0.61	0.33
22	100000185389	273	669	3	0.45	0.17
23	100000185391	273	669	3	0.42	0.35
24	100000197617	246	669	3	0.36	0.13
25	100000197603	246	669	3	0.66	0.20
26	100000197677	257	669	3	0.41	0.13
27	100000197751	257	669	3	0.26	0.21
28	100000197722	257	669	3	0.42	0.21
29	100000350878	241	669	3	0.56	0.27
30	100000098450	241	669	3	0.57	0.33
31	100000197651	246	669	3	0.74	0.12
32	100000197655	273	669	3	0.22	0.11
33	100000197656	273	669	3	0.71	0.32
34	100000197741	257	669	3	0.48	0.40
35	100000197755	257	669	3	0.52	0.27
36	100000197779	262	669	3	0.62	0.41
37	100000197670	251	669	3	0.54	0.30

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
38	100000197674	273	669	3	0.52	0.34
39	100000197675	273	669	3	0.46	0.31
40	100000098452	241	669	3	0.70	0.41
41	100000098444	241	669	3	0.36	0.21
42	100000197667	251	669	3	0.37	0.25
43	100000197668	273	669	3	0.46	0.28
44	100000197669	273	669	3	0.33	0.25
45	100000098451	241	669	3	0.62	0.44
46	100000098454	241	669	3	0.69	0.33
47	100000098447	241	669	3	0.29	0.12
48	100000185378	268	669	3	0.69	0.38
49	100000197757	257	669	3	0.57	0.47
50	100000197761	273	669	3	0.59	0.33
51	100000197765	273	669	3	0.52	0.14
52	100000098449	241	669	3	0.63	0.43
53	100000098453	241	669	3	0.39	0.17
54	100000197602	246	669	3	0.52	0.15
55	100000197621	246	669	3	0.61	0.31
56	100000197660	246	669	3	0.71	0.37
57	100000197756	257	669	3	0.64	0.31
58	100000098446	241	669	3	0.33	0.23
59	100000098455	241	669	3	0.36	0.09
60	100000197648	246	669	3	0.28	0.09
61	100000197649	273	669	3	0.66	0.34
62	100000197650	273	669	3	0.39	0.12
63	100000185382	268	669	3	0.42	0.23
64	100000185313	268	669	3	0.75	0.37
65	100000197662	251	669	3	0.49	0.44
66	100000197665	251	669	3	0.60	0.34
67	100000098522	268	669	3	0.70	0.42
68	100000185484	273	669	3	0.40	0.23
69	100000185404	273	669	3	0.43	0.29
70	100000197746	257	669	3	0.31	0.05
71	100000197724	257	669	3	0.50	0.40
72	100000197750	257	669	3	0.38	0.17
73	100000098527	268	669	3	0.31	0.19
74	100000185381	268	669	3	0.62	0.37
75	100000098435	241	669	3	0.52	0.34

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

Item Seq. No.	Item CID	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
76	100000185485	273	669	3	0.49	0.40
77	100000185486	273	669	3	0.63	0.41
78	100000185380	268	669	3	0.48	0.28
79	100000185384	268	669	3	0.63	0.37
80	100000185377	268	669	3	0.53	0.35
81	100000197720	257	669	3	0.44	0.30
82	100000185386	268	669	3	0.39	0.24
83	100000185477	273	669	3	0.67	0.32
84	100000185473	273	669	3	0.33	0.34
85	100000197676	251	669	3	0.36	0.13
86	100000197664	251	669	3	0.39	0.29
87	100000098532	268	669	3	0.30	0.11
88	100000185387	268	669	3	0.42	0.39
89	100000185376	268	669	3	0.46	0.22
90	100000197740	257	669	3	0.62	0.39
91	100000197661	251	669	3	0.57	0.28
92	100000098515	268	669	3	0.53	0.49
93	100000185385	268	669	3	0.26	0.14
94	100000197663	251	669	3	0.43	0.22
95	100000197666	251	669	3	0.56	0.31
96	100000098441	241	669	3	0.52	0.04
97	100000185388	273	669	3	0.57	0.30
98	100000185390	273	669	3	0.61	0.42
99	100000185379	268	669	3	0.61	0.28
100	100000185383	268	669	3	0.68	0.48

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criteria.

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
1	100000198111	251	872	3	0.52	0.37
2	100000198113	273	872	3	0.57	0.32
3	100000198114	273	872	3	0.56	0.17
4	100000198138	257	872	3	0.34	0.30
5	100000198123	257	872	3	0.57	0.29
6	100000098578	241	872	3	0.42	0.37
7	100000186564	268	872	3	0.42	0.19
8	100000186565	273	872	3	0.32	0.17
9	100000186566	273	872	3	0.33	0.02
10	100000098583	241	872	3	0.37	0.19
11	100000098584	241	872	3	0.78	0.39
12	100000198124	257	872	3	0.57	0.29
13	100000198153	262	872	3	0.39	0.13
14	100000198098	246	872	3	0.76	0.31
15	100000198100	246	872	3	0.49	0.19
16	100000198133	257	872	3	0.39	0.25
17	100000198134	273	872	3	0.49	0.30
18	100000198135	273	872	3	0.36	-0.08
19	100000198110	251	872	3	0.76	0.32
20	100000098579	241	872	3	0.71	0.30
21	100000098581	241	872	3	0.43	0.33
22	100000198140	262	872	3	0.64	0.34
23	100000198142	273	872	3	0.53	0.20
24	100000198143	273	872	3	0.51	0.24
25	100000198126	257	872	3	0.31	0.03
26	100000198128	257	872	3	0.46	0.17
27	100000198092	246	872	3	0.48	0.13
28	100000198099	246	872	3	0.58	0.13
29	100000198117	251	872	3	0.46	0.38
30	100000198118	273	872	3	0.53	0.28
31	100000198119	273	872	3	0.34	0.12
32	100000098577	241	872	3	0.61	0.32
33	100000098576	241	872	3	0.36	0.03
34	100000098587	241	872	3	0.42	0.19
35	100000198116	251	872	3	0.20	-0.01
36	100000198107	251	872	3	0.53	0.33
37	100000198101	246	872	3	0.39	0.18

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
38	100000198102	273	872	3	0.48	0.29
39	100000198103	273	872	3	0.51	0.31
40	100000198148	262	872	3	0.47	0.26
41	100000198158	262	872	3	0.49	0.37
42	100000098646	268	872	3	0.78	0.41
43	100000186562	273	872	3	0.67	0.36
44	100000186563	273	872	3	0.44	0.18
45	100000198094	246	872	3	0.78	0.35
46	100000198096	246	872	3	0.58	0.32
47	100000198129	257	872	3	0.53	0.38
48	100000198131	273	872	3	0.34	0.19
49	100000198132	273	872	3	0.40	0.21
50	100000198093	246	872	3	0.90	0.38
51	100000098572	241	872	3	0.40	0.38
52	100000098586	241	872	3	0.48	0.32
53	100000198104	246	872	3	0.55	0.12
54	100000198105	273	872	3	0.54	0.26
55	100000198106	273	872	3	0.43	0.23
56	100000198139	257	872	3	0.42	0.21
57	100000098585	241	872	3	0.79	0.33
58	100000098582	241	872	3	0.52	0.38
59	100000207144	251	872	3	0.41	0.17
60	100000198144	262	872	3	0.69	0.47
61	100000198145	273	872	3	0.39	0.29
62	100000198147	273	872	3	0.49	0.24
63	100000186576	268	872	3	0.72	0.37
64	100000098666	268	872	3	0.53	0.24
65	100000198150	262	872	3	0.70	0.45
66	100000198149	262	872	3	0.56	0.33
67	100000186573	268	872	3	0.36	0.29
68	100000186574	273	872	3	0.49	0.26
69	100000186575	273	872	3	0.49	0.31
70	100000186583	268	872	3	0.50	-0.07
71	100000198108	251	872	3	0.44	0.36
72	100000198122	251	872	3	0.37	0.19
73	100000098645	268	872	3	0.44	0.25
74	100000186578	268	872	3	0.73	0.41
75	100000186580	268	872	3	0.63	0.42

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
76	100000198157	262	872	3	0.32	0.22
77	100000098573	241	872	3	0.42	0.28
78	100000186560	273	872	3	0.53	0.21
79	100000186561	273	872	3	0.35	0.20
80	100000198137	257	872	3	0.44	0.34
81	100000098571	241	872	3	0.60	0.39
82	100000186581	268	872	3	0.61	0.30
83	100000207143	268	872	3	0.60	0.36
84	100000198121	251	872	3	0.41	0.23
85	100000098588	241	872	3	0.55	0.29
86	100000098657	268	872	3	0.35	0.25
87	100000186571	273	872	3	0.47	0.12
88	100000186572	273	872	3	0.47	0.27
89	100000198125	257	872	3	0.55	0.37
90	100000198151	262	872	3	0.71	0.40
91	100000098653	268	872	3	0.33	0.21
92	100000186577	268	872	3	0.72	0.38
93	100000098568	241	872	3	0.64	0.41
94	100000186558	273	872	3	0.57	0.38
95	100000186559	273	872	3	0.49	0.21
96	100000186567	268	872	3	0.54	0.31
97	100000186582	268	872	3	0.38	0.21
98	100000198127	257	872	3	0.55	0.38
99	100000098664	268	872	3	0.52	0.32
100	100000098580	241	872	3	0.32	0.15
101	100000186579	268	872	3	0.50	0.34
102	100000198109	251	872	3	0.40	0.32

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criteria.

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
1	100000187371	268	926	3	0.83	0.36
2	100000187367	268	926	3	0.47	0.22
3	100000187378	268	926	3	0.68	0.31
4	100000099084	241	926	3	0.67	0.39
5	100000099080	241	926	3	0.64	0.38
6	100000099090	241	926	3	0.40	0.16
7	100000099087	241	926	3	0.26	0.06
8	100000196233	251	926	3	0.65	0.38
9	100000196246	251	926	3	0.12	-0.04
10	100000196229	251	926	3	0.46	0.25
11	100000196273	257	926	3	0.31	0.34
12	100000196043	273	926	3	0.43	0.30
13	100000196045	273	926	3	0.52	0.21
14	100000196253	257	926	3	0.64	0.40
15	100000196278	257	926	3	0.29	0.32
16	100000196088	246	926	3	0.33	0.15
17	100000187430	268	926	3	0.55	0.30
18	100000187390	268	926	3	0.73	0.25
19	100000196090	246	926	3	0.29	0.15
20	100000196081	246	926	3	0.35	0.14
21	100000187376	268	926	3	0.55	0.16
22	100000187386	268	926	3	0.70	0.14
23	100000187383	268	926	3	0.41	0.18
24	100000196231	251	926	3	0.40	0.17
25	100000196036	273	926	3	0.68	0.27
26	100000196029	273	926	3	0.54	0.30
27	100000196242	251	926	3	0.60	0.34
28	100000099083	241	926	3	0.53	0.19
29	100000099075	241	926	3	0.53	0.34
30	100000099086	241	926	3	0.62	0.35
31	100000196200	246	926	3	0.63	0.30
32	100000196277	257	926	3	0.50	0.26
33	100000196284	262	926	3	0.63	0.36
34	100000196057	273	926	3	0.48	0.27
35	100000196054	273	926	3	0.47	0.26
36	100000196279	262	926	3	0.28	0.19
37	100000099081	241	926	3	0.39	0.13

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

Item Seq. No.	Item UIN	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
38	100000099079	241	926	3	0.63	0.25
39	100000099088	241	926	3	0.54	0.12
40	100000099085	241	926	3	0.69	0.33
41	100000196237	251	926	3	0.39	0.08
42	100000196234	251	926	3	0.75	0.27
43	100000196053	273	926	3	0.64	0.34
44	100000196244	251	926	3	0.20	0.09
45	100000099089	241	926	3	0.72	0.39
46	100000099091	241	926	3	0.54	0.30
47	100000099092	241	926	3	0.54	0.23
48	100000187365	273	926	3	0.40	0.23
49	100000187364	273	926	3	0.46	0.25
50	100000196381	262	926	3	0.76	0.38
51	100000196281	262	926	3	0.41	0.16
52	100000187387	268	926	3	0.22	0.15
53	100000187388	273	926	3	0.40	0.08
54	100000187389	273	926	3	0.35	0.02
55	100000196270	257	926	3	0.35	0.31
56	100000196256	257	926	3	0.53	0.31
57	100000196267	257	926	3	0.85	0.31
58	100000099178	241	926	3	0.53	0.25
59	100000187369	268	926	3	0.40	0.15
60	100000187380	268	926	3	0.39	0.15
61	100000187429	268	926	3	0.56	0.34
62	100000187428	268	926	3	0.59	0.33
63	100000187372	268	926	3	0.24	0.04
64	100000099180	268	926	3	0.45	0.29
65	100000099072	241	926	3	0.56	0.37
66	100000187360	273	926	3	0.38	0.13
67	100000187361	273	926	3	0.51	0.34
68	100000196260	257	926	3	0.41	0.04
69	100000196258	257	926	3	0.32	0.07
70	100000196269	257	926	3	0.22	0.23
71	100000196223	246	926	3	0.52	0.20
72	100000196100	246	926	3	0.80	0.36
73	100000196094	246	926	3	0.47	0.24
74	100000187373	273	926	3	0.35	0.21
75	100000187363	273	926	3	0.36	0.08

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
76	100000196037	273	926	3	0.47	0.16
77	100000187370	268	926	3	0.46	0.28
78	100000187391	268	926	3	0.38	0.28
79	100000187392	268	926	3	0.40	0.11
80	100000187382	268	926	3	0.41	0.15
81	100000196263	257	926	3	0.67	0.28
82	100000196247	257	926	3	0.41	0.28
83	100000196238	251	926	3	0.38	0.09
84	100000196079	246	926	3	0.40	0.22
85	100000196025	273	926	3	0.60	0.22
86	100000187393	273	926	3	0.28	-0.02
87	100000196225	246	926	3	0.31	0.07
88	100000099177	241	926	3	0.48	0.32
89	100000187366	273	926	3	0.45	0.23
90	100000196049	273	926	3	0.58	0.22
91	100000196235	251	926	3	0.28	0.05
92	100000099082	241	926	3	0.44	0.05
93	100000196042	273	926	3	0.40	0.22
94	100000187385	268	926	3	0.49	0.20
95	100000187374	273	926	3	0.45	0.04
96	100000187424	273	926	3	0.40	0.16
97	100000187379	268	926	3	0.55	0.37
98	100000187377	268	926	3	0.24	0.04
99	100000187381	268	926	3	0.26	0.19
100	100000187375	268	926	3	0.38	0.12

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373 & 374= Process
2. These analyses are based on the equating sample used with the exclusion criteria.

Table 3.2.4. The 2010 Mod-MSA: Mathematics Classical Item Statistics: Grade 6

Item Seq. No.	Item UIN	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
1	100000272103	246	852	3	0.72	0.16
2	100000272105	246	852	3	0.62	0.17
3	100000272100	246	852	3	0.30	0.14
4	100000272095	246	852	3	0.21	0.17
5	100000272117	251	852	3	0.41	0.29
6	100000198185	251	852	3	0.41	0.37
7	100000272111	251	852	3	0.35	0.11
8	100000272115	251	852	3	0.37	0.09
9	100000272113	251	852	3	0.41	0.36
10	100000272076	241	852	3	0.58	0.30
11	100000272073	241	852	3	0.53	0.15
12	100000099232	241	852	3	0.60	0.43
13	100000272096	246	852	3	0.55	0.36
14	100000272145	273	852	3	0.47	0.24
15	100000272146	273	852	3	0.49	0.17
16	100000198162	246	852	3	0.43	0.18
17	100000272101	246	852	3	0.46	0.24
18	100000198232	262	852	3	0.33	0.21
19	100000272084	262	852	3	0.34	-0.01
20	100000198210	257	852	3	0.35	0.32
21	100000198211	273	852	3	0.51	0.28
22	100000198212	273	852	3	0.63	0.41
23	100000272129	257	852	3	0.52	0.31
24	100000187836	268	852	3	0.61	0.32
25	100000272156	268	852	3	0.60	0.29
26	100000272158	268	852	3	0.41	0.13
27	100000272159	268	852	3	0.39	0.21
28	100000272066	241	852	3	0.73	0.37
29	100000272077	241	852	3	0.31	0.17
30	100000272068	241	852	3	0.36	0.19
31	100000272075	241	852	3	0.70	0.29
32	100000272072	241	852	3	0.61	0.23
33	100000272067	241	852	3	0.66	0.43
34	100000272131	257	852	3	0.44	0.27
35	100000272127	257	852	3	0.62	0.39
36	100000272134	257	852	3	0.65	0.32
37	100000273844	241	852	3	0.50	0.18

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
38	100000273626	273	852	3	0.31	0.17
39	100000272143	273	852	3	0.54	0.37
40	100000272144	273	852	3	0.45	0.27
41	100000198178	246	852	3	0.47	0.26
42	100000272099	246	852	3	0.28	0.08
43	100000272148	273	852	3	0.42	0.14
44	100000272149	273	852	3	0.49	0.15
45	100000198235	262	852	3	0.48	0.19
46	100000272080	262	852	3	0.69	0.31
47	100000272082	262	852	3	0.66	0.31
48	100000272114	251	852	3	0.26	0.20
49	100000272116	251	852	3	0.40	0.32
50	100000198186	251	852	3	0.35	0.29
51	100000272112	251	852	3	0.34	0.09
52	100000099325	241	852	3	0.65	0.43
53	100000272128	257	852	3	0.56	0.30
54	100000272125	257	852	3	0.45	0.22
55	100000272069	241	852	3	0.58	0.35
56	100000272155	268	852	3	0.23	-0.02
57	100000272166	268	852	3	0.39	0.27
58	100000272167	268	852	3	0.25	0.28
59	100000187852	268	852	3	0.60	0.37
60	100000272065	241	852	3	0.43	0.38
61	100000198218	257	852	3	0.52	0.29
62	100000198220	273	852	3	0.53	0.41
63	100000198219	273	852	3	0.50	0.35
64	100000187841	268	852	3	0.60	0.32
65	100000272161	268	852	3	0.27	0.18
66	100000272163	268	852	3	0.45	0.35
67	100000198170	246	852	3	0.23	0.24
68	100000272104	246	852	3	0.45	0.17
69	100000272150	273	852	3	0.44	0.17
70	100000272151	273	852	3	0.28	0.05
71	100000272098	246	852	3	0.60	0.28
72	100000272071	241	852	3	0.74	0.25
73	100000099236	241	852	3	0.62	0.34
74	100000273837	241	852	3	0.65	0.34
75	100000273598	273	852	3	0.51	0.20

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

Item Seq. No.	Item CID	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
76	100000272141	273	852	3	0.63	0.28
77	100000198202	257	852	3	0.62	0.38
78	100000272126	257	852	3	0.54	0.29
79	100000272133	257	852	3	0.38	0.35
80	100000272132	257	852	3	0.63	0.33
81	100000272165	268	852	3	0.53	0.10
82	100000272152	273	852	3	0.38	0.22
83	100000273858	273	852	3	0.40	0.14
84	100000272160	268	852	3	0.69	0.34
85	100000187850	268	852	3	0.66	0.32
86	100000272157	268	852	3	0.18	0.22
87	100000272154	273	852	3	0.46	0.12
88	100000272153	273	852	3	0.57	0.26
89	100000272162	268	852	3	0.55	0.31
90	100000099252	241	852	3	0.35	0.11
91	100000198165	273	852	3	0.44	0.30
92	100000272097	246	852	3	0.52	0.26
93	100000187833	273	852	3	0.35	0.21
94	100000272102	246	852	3	0.47	0.00
95	100000198174	273	852	3	0.40	0.19
96	100000272164	268	852	3	0.34	0.19

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373b & 374= Process
2. These analyses are based on the equating sample used with the exclusion criteria.

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

Item Seq. No.[1]	Item CID	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
1	100000322134	246	1000	3	0.63	0.32
2	100000273428	273	1000	3	0.29	0.13
3	100000273431	246	1000	3	0.27	0.04
4	100000272267	273	1000	3	0.80	0.12
5	100000197499	257	1000	3	0.59	0.34
6	100000272217	257	1000	3	0.60	0.19
7	100000322127	246	1000	3	0.20	0.24
8	100000322131	246	1000	3	0.15	0.01
9	100000197494	251	1000	3	0.35	0.23
10	100000322136	251	1000	3	0.27	-0.06
11	100000322138	251	1000	3	0.26	0.21
12	100000197496	251	1000	3	0.21	0.23
13	100000322166	241	1000	3	0.63	0.29
14	100000322165	241	1000	3	0.43	0.18
15	100000322169	241	1000	3	0.33	0.09
16	100000099488	268	1000	3	0.83	0.35
17	100000322151	268	1000	3	0.34	0.18
18	100000322141	268	1000	3	0.54	0.26
19	100000322152	268	1000	3	0.52	0.23
20	100000322140	268	1000	3	0.58	0.19
21	100000287243	273	1000	3	0.76	0.26
22	100000272268	273	1000	3	0.63	0.19
23	100000287244	251	1000	3	0.16	0.17
24	100000272270	273	1000	3	0.32	0.11
25	100000322137	251	1000	3	0.31	0.12
26	100000322139	251	1000	3	0.26	0.21
27	100000272201	251	1000	3	0.38	0.00
28	100000322164	241	1000	3	0.55	0.25
29	100000272171	241	1000	3	0.60	0.15
30	100000099420	241	1000	3	0.63	0.12
31	100000099421	241	1000	3	0.72	0.29
32	100000272271	273	1000	3	0.37	0.18
33	100000272216	257	1000	3	0.43	0.10
34	100000272272	273	1000	3	0.47	0.21
35	100000322175	262	1000	3	0.57	0.26
36	100000322172	262	1000	3	0.31	0.21
37	100000197528	262	1000	3	0.34	0.20

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

Item Seq. No.[1]	Item CID	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
38	100000374151	273	1000	3	0.37	0.02
39	100000374152	246	1000	3	0.41	0.01
40	100000196052	273	1000	3	0.42	0.00
41	100000196055	273	1000	3	0.33	0.15
42	100000322129	246	1000	3	0.62	0.26
43	100000197481	246	1000	3	0.59	0.35
44	100000273423	273	1000	3	0.47	0.19
46	100000273427	241	1000	3	0.37	0.02
48	100000272203	251	1000	3	0.52	0.37
49	100000322153	268	1000	3	0.58	0.40
50	100000272246	268	1000	3	0.29	0.07
51	100000322154	268	1000	3	0.62	0.22
52	100000322128	246	1000	3	0.29	0.14
53	100000322173	262	1000	3	0.34	0.04
54	100000322174	262	1000	3	0.28	0.03
55	100000273415	273	1000	3	0.29	0.03
56	100000273416	262	1000	3	0.17	0.12
57	100000272274	273	1000	3	0.41	0.16
58	100000272186	246	1000	3	0.65	0.23
59	100000322135	246	1000	3	0.65	0.36
60	100000322132	246	1000	3	0.48	0.12
61	100000272187	246	1000	3	0.46	0.25
63	100000374196	268	1000	3	0.34	0.27
64	100000272248	268	1000	3	0.58	0.21
65	100000322146	268	1000	3	0.33	0.13
66	100000197501	257	1000	3	0.46	0.24
67	100000322158	241	1000	3	0.44	0.53
68	100000322160	241	1000	3	0.48	0.38
69	100000322163	241	1000	3	0.46	0.29
70	100000272174	241	1000	3	0.41	0.39
71	100000099504	268	1000	3	0.32	0.22
72	100000322142	268	1000	3	0.80	0.26
73	100000167818	268	1000	3	0.63	0.45
74	100000322143	268	1000	3	0.80	0.31
75	100000322147	268	1000	3	0.49	0.29
76	100000099493	268	1000	3	0.69	0.27
77	100000273417	273	1000	3	0.47	0.12
78	100000273444	241	1000	3	0.47	0.41

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

Item Seq. No.[3]	Item CID	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
79	100000273445	273	1000	3	0.33	0.15
80	100000197524	262	1000	3	0.19	0.23
81	100000272247	268	1000	3	0.68	0.26
82	100000322149	268	1000	3	0.68	0.41
83	100000322150	268	1000	3	0.50	0.17
84	100000322144	268	1000	3	0.74	0.27
85	100000322148	268	1000	3	0.61	0.44
86	100000099410	241	1000	3	0.43	0.37
87	100000322161	241	1000	3	0.42	0.27
88	100000099413	241	1000	3	0.25	0.20
89	100000322157	241	1000	3	0.75	0.23
90	100000099404	241	1000	3	0.49	0.44
91	100000196255	273	1000	3	0.55	0.22
92	100000197522	257	1000	3	0.59	0.23
93	100000196257	273	1000	3	0.37	0.18
94	100000196261	273	1000	3	0.47	0.08
95	100000322171	241	1000	3	0.48	0.43
96	100000322162	241	1000	3	0.47	0.30
97	100000322159	241	1000	3	0.42	0.36
98	100000099490	268	1000	3	0.46	0.29

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process

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

3. Item #s 45, 47, 62 were not scored

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
1	100000322198	241	1017	3	0.49	-0.04
2	100000322187	241	1017	3	0.41	0.11
3	100000197359	251	1017	3	0.60	0.24
4	100000272279	251	1017	3	0.15	0.15
5	100000272278	251	1017	3	0.37	0.07
6	100000272340	241	1017	3	0.47	0.24
7	100000099548	241	1017	3	0.74	0.28
8	100000322121	262	1017	3	0.54	0.25
9	100000322186	241	1017	3	0.73	0.17
10	100000322122	262	1017	3	0.59	0.01
11	100000322120	262	1017	3	0.55	0.31
12	100000099550	241	1017	3	0.43	0.34
13	100000272296	273	1017	3	0.26	0.06
14	100000272337	241	1017	3	0.28	0.22
15	100000272295	273	1017	3	0.29	0.14
16	100000272297	273	1017	3	0.51	0.10
17	100000272323	257	1017	3	0.45	0.08
18	100000322119	257	1017	3	0.58	0.22
19	100000322118	257	1017	3	0.47	0.33
20	100000322084	246	1017	3	0.32	0.11
21	100000197347	246	1017	3	0.38	0.21
22	100000322081	246	1017	3	0.54	0.23
23	100000272309	246	1017	3	0.33	0.04
24	100000322209	268	1017	3	0.53	0.21
25	100000322208	268	1017	3	0.65	0.29
26	100000322205	268	1017	3	0.42	0.10
27	100000322207	268	1017	3	0.77	0.29
28	100000197387	262	1017	3	0.42	0.41
29	100000286862	273	1017	3	0.39	0.02
30	100000272790	273	1017	3	0.58	0.32
31	100000286872	262	1017	3	0.40	0.07
32	100000373799	268	1017	3	0.37	0.21
33	100000272368	268	1017	3	0.34	-0.02
34	100000322200	268	1017	3	0.34	0.32
35	100000099626	268	1017	3	0.72	0.23
36	100000286792	273	1017	3	0.45	0.18
37	100000272298	273	1017	3	0.27	0.12

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

Item Seq. No.	Item UIN	Strand No.[1]	n- Count[2]	Response Options	P-Value	Point-Biserial
38	100000286786	251	1017	3	0.76	0.13
39	100000272300	273	1017	3	0.15	0.12
40	100000322079	251	1017	3	0.44	0.08
41	100000322078	251	1017	3	0.59	0.23
42	100000197367	257	1017	3	0.36	0.27
43	100000286859	273	1017	3	0.33	0.31
44	100000272305	273	1017	3	0.30	0.29
45	100000286860	257	1017	3	0.27	0.23
46	100000322116	257	1017	3	0.49	0.06
47	100000322117	257	1017	3	0.83	0.31
48	100000286845	273	1017	3	0.42	0.24
49	100000272301	273	1017	3	0.15	0.16
50	100000286846	246	1017	3	0.30	0.15
51	100000322080	246	1017	3	0.46	0.23
52	100000272353	262	1017	3	0.48	0.21
53	100000167565	273	1017	3	0.48	0.19
54	100000167566	273	1017	3	0.30	0.17
55	100000174558	262	1017	3	0.26	0.19
56	100000322204	268	1017	3	0.34	0.19
57	100000322206	268	1017	3	0.81	0.26
58	100000322199	268	1017	3	0.41	0.10
59	100000099639	268	1017	3	0.40	0.29
60	100000272308	246	1017	3	0.43	0.22
61	100000197322	246	1017	3	0.45	0.20
62	100000322082	246	1017	3	0.42	0.24
63	100000322115	257	1017	3	0.74	0.27
64	100000197365	257	1017	3	0.38	0.24
65	100000272339	241	1017	3	0.56	0.35
66	100000197377	273	1017	3	0.61	0.20
67	100000197373	273	1017	3	0.44	0.20
68	100000197375	257	1017	3	0.18	0.24
69	100000197376	273	1017	3	0.64	0.23
70	100000322201	268	1017	3	0.89	0.19
71	100000322203	268	1017	3	0.28	0.22
72	100000322193	241	1017	3	0.48	0.14
73	100000099553	241	1017	3	0.68	0.37
74	100000322185	241	1017	3	0.53	0.33
75	100000322195	241	1017	3	0.44	0.28

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

Item Seq. No.	Item CID	Strand No.[1]	n-Count[2]	Response Options	P-Value	Point-Biserial
76	100000373853	273	1017	3	0.48	0.36
77	100000167561	273	1017	3	0.58	0.34
78	100000373854	241	1017	3	0.52	0.41
79	100000322188	241	1017	3	0.58	0.43
80	100000197319	246	1017	3	0.36	0.23
81	100000322196	241	1017	3	0.65	0.20
82	100000322191	241	1017	3	0.52	0.23
83	100000322189	241	1017	3	0.76	0.37
84	100000099647	241	1017	3	0.52	0.28
85	100000286880	273	1017	3	0.37	0.18
86	100000272303	273	1017	3	0.57	0.17
87	100000286879	241	1017	3	0.41	0.17
88	100000286793	273	1017	3	0.40	0.16
89	100000286797	257	1017	3	0.39	0.28
90	100000272293	273	1017	3	0.74	0.24
91	100000272294	273	1017	3	0.64	0.25
92	100000322192	241	1017	3	0.56	0.34
93	100000322190	241	1017	3	0.29	0.21
94	100000322202	268	1017	3	0.77	0.27
95	100000272367	268	1017	3	0.80	0.24
96	100000099631	268	1017	3	0.35	0.19
97	100000322197	241	1017	3	0.41	0.35
98	100000322194	241	1017	3	0.64	0.21

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process

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 (IRT) 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: Mathematics. 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: Mathematics 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).

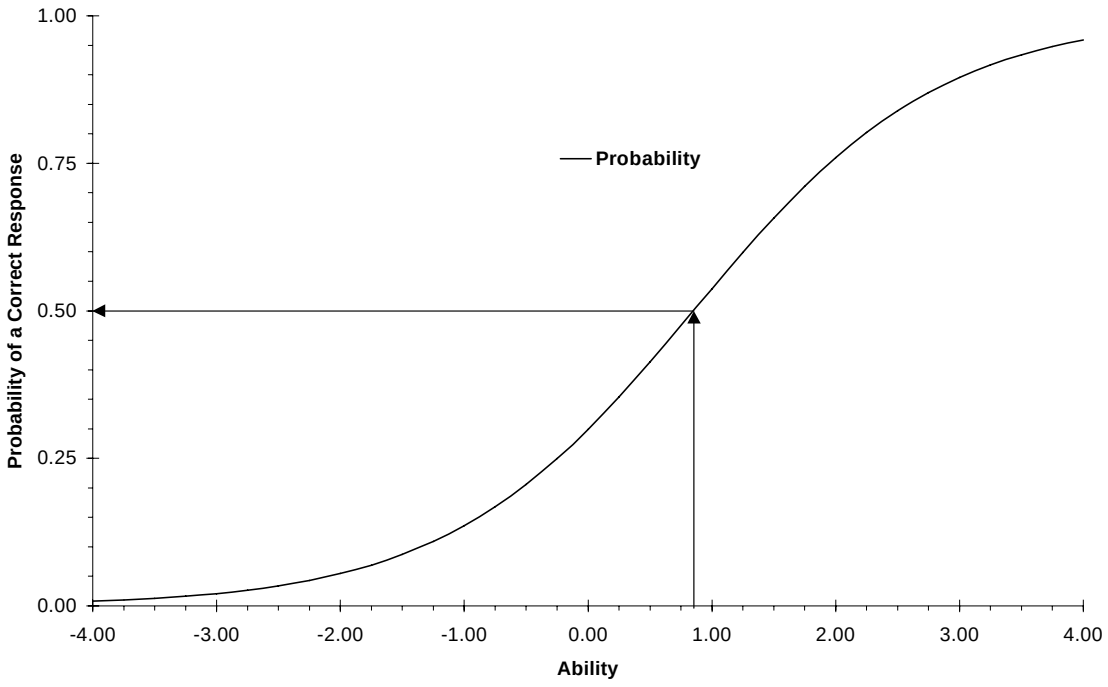


Figure 4.1 Item Characteristic Curve

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).

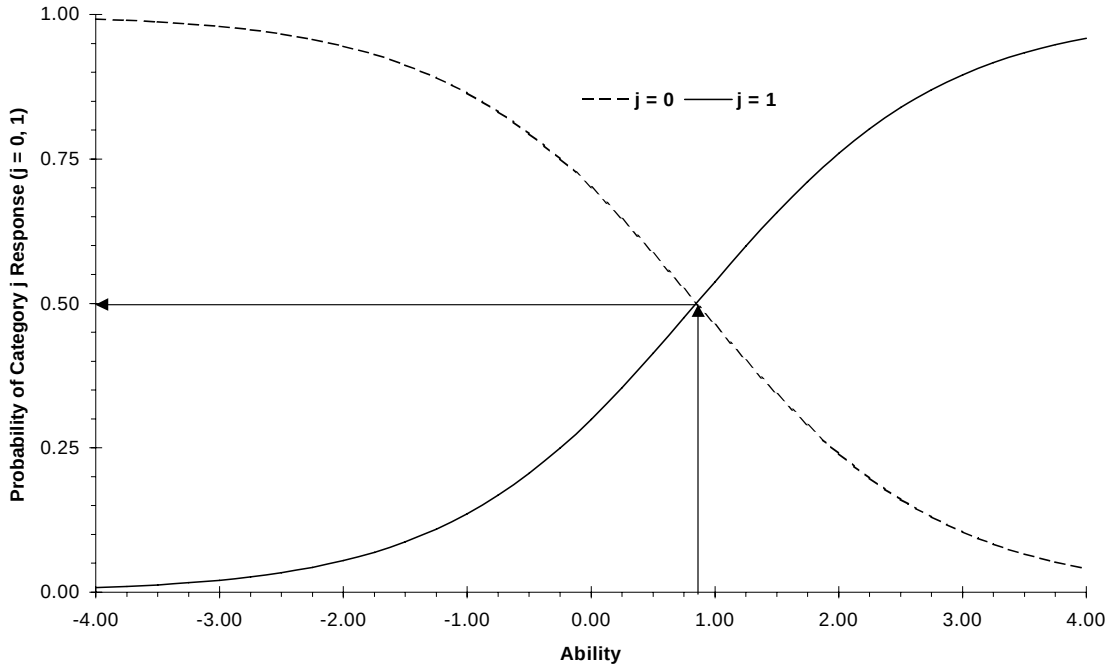


Figure 4.2 Category Response Curves for a One-Step Item

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 person 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: Mathematics was used to establish the base scale for the assessment in the area of mathematics 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: Mathematics 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 and Equating the 2010 Mod-MSA: Mathematics Grades 6-8

The base scale for the Mod-MSA: Mathematics 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.

The 2010 Mod-MSA, Math program included Algebra, Geometry and Measurement, Statistics and Probability, Numbers and Computations, and Process as the five scoring strands. The 2010 common items selected for linking were those items that had been administered in 2009. The pool of common items followed the same proportion of strand representation on the 2010 form as they did on the 2009 form. They also were placed at the same location in 2010 as they were on the 2009 form. Consequently, these items were used to put the 2010 assessments on the same base scale created during the 2009 assessment. In terms of year-to-year linking purpose, item and structure calibration parameters of the 2010 linking items were fixed with those of 2009 linking items which were already put on the 2009 common base scale. The stability of linking common items was evaluated using generalized robust z procedures, correlation coefficients, and standard deviation ratios discussed above.

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

The 2010 Mod-MSA was calibrated and equated by fixing item parameters of common linking items. For Grades 6-8, there were 25 common items used in each of the three grades for use as linking items in the equating process. Items in these grades were first 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, create the RS to SS tables and archive non-operational items in the item bank.

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.

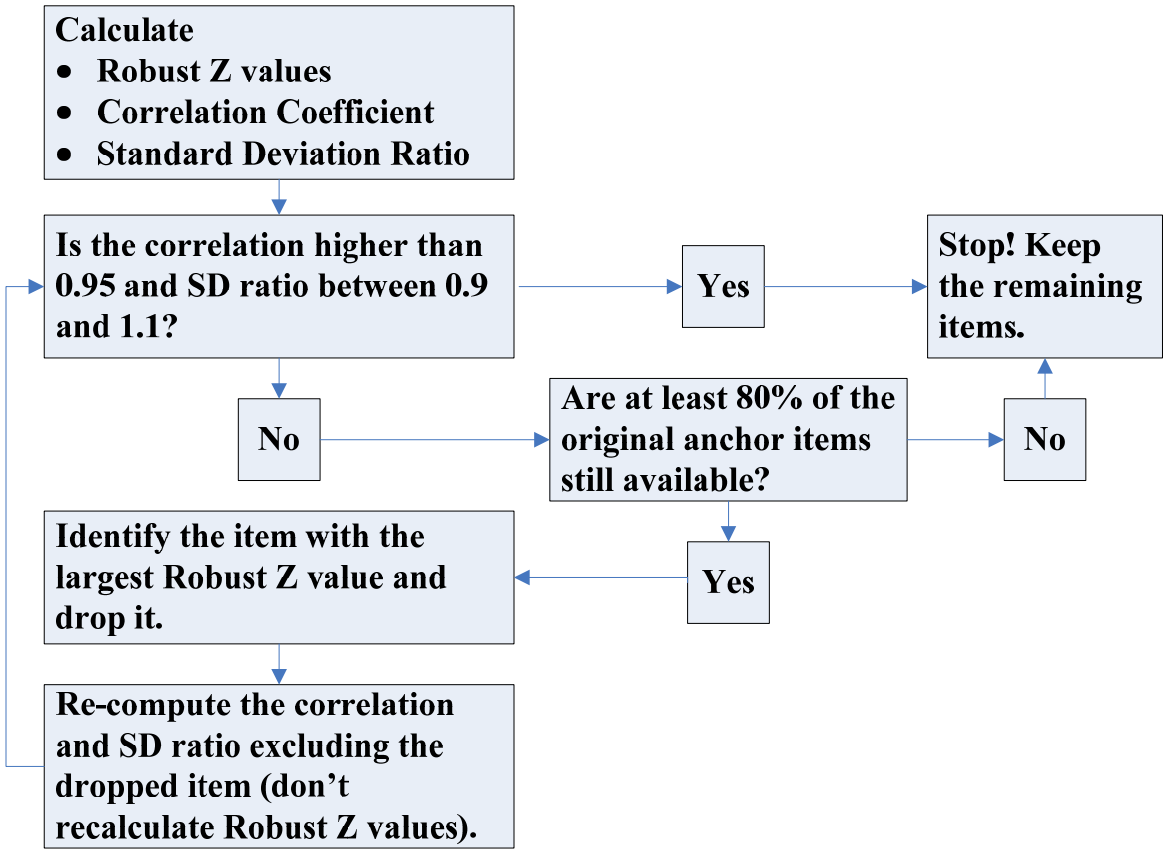


Figure 4.4.1. Anchor Evaluation Steps Chart for Mod-MSA

Table 4.4.1. Unequated Core Linking Item Difficulties of Year 2009 vs. Year 2010: Grade 6

Item No.	Item Seq. No.	Rasch Diff. 2009	Rasch Diff.2010	Robust Z*
1	6	0.3977	0.3028	0.23
2	12	-0.4626	-0.5599	0.24
3	16	0.3156	0.2080	0.29
4	18	0.6694	0.6908	-0.31
5	20	0.3908	0.5815	-1.09
6	21	-0.1220	-0.1674	0.00
7	22	-0.8737	-0.6721	-1.14
8	24	-0.5550	-0.5970	-0.02
9	41	0.0892	0.0166	0.13
10	45	0.1385	-0.0091	0.47
11	50	0.3190	0.5872	-1.44
12	52	-0.5054	-0.7925	1.11
13	59	-0.3809	-0.5283	0.47
14	61	-0.5584	-0.1776	**-1.96**
15	62	-0.4037	-0.2235	-1.04
16	63	-0.1446	-0.1266	-0.29
17	64	-0.0767	-0.5441	**1.94**
18	67	1.4827	1.2262	0.97
19	73	-0.3322	-0.6559	1.28
20	77	-0.4955	-0.6344	0.43
21	85	-0.7997	-0.8372	-0.04
22	90	1.0009	0.5815	**1.72**
23	91	-0.0184	0.1662	-1.06
24	93	0.5310	0.5929	-0.49
25	95	0.1881	0.3347	-0.88

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.	Previous Year-2009	2010	Robust Z
1	5	-0.4387	-0.5594	0.40
2	9	0.8482	0.4942	1.35
3	12	1.4407	1.2821	0.56
4	16	-2.0862	-1.8465	-1.06
5	30	-1.0134	-0.7346	-1.22
6	31	-1.1523	-1.1412	-0.13
7	37	0.4196	0.5750	-0.72
8	38	0.4109	0.4289	-0.16
9	39	0.1404	0.2299	-0.45
10	40	0.1294	0.1945	-0.35
11	41	0.5938	0.5846	-0.05
12	43	-0.3334	-0.5461	0.78
13	66	0.3822	0.0026	1.45
14	71	0.6547	0.6430	-0.04
15	73	-0.4248	-0.7209	1.11
16	76	-0.5911	-1.0196	**1.65**
17	80	1.6427	1.3751	1.00
18	86	0.5339	0.1329	1.54
19	88	1.0713	1.0332	0.07
20	90	0.1073	-0.1050	0.77
21	91	-0.3721	-0.3932	0.00
22	92	-0.8505	-0.5506	-1.30
23	93	0.3308	0.3967	-0.35
24	94	-0.1444	-0.0405	-0.51
25	98	0.2911	0.0069	1.07

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.	Previous Year-2009	2010	Robust Z
1	3	-0.6919	-0.5431	0.23
2	7	-1.2618	-1.1996	0.82
3	12	0.1130	0.2051	0.62
4	21	0.2121	0.4107	-0.11
5	28	0.3845	0.2656	**2.07**
6	32	0.0630	0.4601	-1.47
7	35	-1.2755	-1.1001	0.05
8	42	0.4466	0.5237	0.72
9	53	-0.1516	-0.016	0.32
10	54	0.6133	0.7956	0.00
11	55	0.7327	1.0196	-0.72
12	59	0.0583	0.3487	-0.74
13	61	-0.2045	0.1153	-0.94
14	64	-0.0053	0.4197	**-1.66**
15	66	-0.9339	-0.5869	-1.13
16	67	0.0276	0.1751	0.24
17	68	1.2281	1.5173	-0.73
18	69	-1.1337	-0.7028	**-1.70**
19	73	-0.9660	-0.8738	0.62
20	76	-0.2869	-0.0202	-0.58
21	77	-0.6825	-0.4262	-0.51
22	78	-0.2434	-0.1886	0.87
23	80	0.2761	0.5237	-0.45
24	84	-0.2686	-0.1971	0.76
25	96	0.5341	0.5932	0.84

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

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

The Mod-MSA reports student scores on the total performance of students on the mathematics 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 Mathematics, 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. Please note that no scale score was allowed to fall below 2 (the LOSS) or exceed 98 (the HOSS). 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 mathematics test score.

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

Grade	Slope	Intercept
3	15.5187	49.4835
4	15.0818	48.9387
5	17.8819	51.4991
6	16.7632	52.1350
7	18.6899	51.4473
8	20.7023	56.8048

Raw Score to Scale Score Conversion Tables for the Total Score

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

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.4562	2.0067	2	31	-	-
1	-4.0494	1.0136	2	16	-	-
2	-3.3287	0.7264	2	11	-	-
3	-2.8954	0.6011	5	9	-	-
4	-2.5795	0.5276	9	8	2	17
5	-2.3278	0.4783	13	7	6	20
6	-2.1165	0.4425	17	7	10	24
7	-1.9329	0.4153	19	6	13	25
8	-1.7695	0.3938	22	6	16	28
9	-1.6214	0.3764	24	6	18	30
10	-1.4854	0.3620	26	6	20	32
11	-1.3586	0.3500	28	5	23	33
12	-1.2397	0.3399	30	5	25	35
13	-1.1272	0.3313	32	5	27	37
14	-1.0199	0.3239	34	5	29	39
15	-0.9170	0.3176	35	5	30	40
16	-0.8179	0.3122	37	5	32	42
17	-0.7220	0.3076	38	5	33	43
18	-0.6286	0.3036	40	5	35	45
19	-0.5374	0.3003	41	5	36	46
20	-0.4481	0.2975	43	5	38	48
21	-0.3602	0.2953	44	5	39	49
22	-0.2736	0.2935	45	5	40	50
23	-0.1878	0.2922	47	5	42	52
24	-0.1026	0.2914	48	5	43	53
25	-0.0180	0.2909	49	5	44	54
26	0.0667	0.2909	51	5	46	56
27	0.1514	0.2913	52	5	47	57
28	0.2364	0.2921	53	5	48	58
29	0.3222	0.2934	54	5	49	59
30	0.4088	0.2951	56	5	51	61
31	0.4964	0.2973	57	5	52	62
32	0.5857	0.3000	59	5	54	64
33	0.6767	0.3033	60	5	55	65
34	0.7697	0.3072	61	5	56	66
35	0.8656	0.3118	63	5	58	68
36	0.9645	0.3172	64	5	59	69
37	1.0670	0.3235	66	5	61	71
38	1.1740	0.3308	68	5	63	73
39	1.2862	0.3394	69	5	64	74
40	1.4047	0.3495	71	5	66	76
41	1.5310	0.3614	73	6	67	79
42	1.6667	0.3758	75	6	69	81
43	1.8142	0.3931	78	6	72	84
44	1.9771	0.4146	80	6	74	86

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
45	2.1601	0.4419	83	7	76	90
46	2.3707	0.4776	86	7	79	93
47	2.6218	0.5270	90	8	82	98
48	2.9369	0.6005	95	9	-	-
49	3.3695	0.7259	98	11	-	-
50	4.0894	1.0132	98	16	-	-
51	5.4954	2.0066	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 4.5.3. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 4

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.4887	2.0070	2	30	-	-
1	-4.0813	1.0139	2	15	-	-
2	-3.3599	0.7269	2	11	-	-
3	-2.9260	0.6016	5	9	-	-
4	-2.6097	0.5280	10	8	2	18
5	-2.3576	0.4786	13	7	6	20
6	-2.1460	0.4428	17	7	10	24
7	-1.9623	0.4154	19	6	13	25
8	-1.7989	0.3938	22	6	16	28
9	-1.6508	0.3763	24	6	18	30
10	-1.5147	0.3618	26	5	21	31
11	-1.3882	0.3498	28	5	23	33
12	-1.2695	0.3395	30	5	25	35
13	-1.1574	0.3308	31	5	26	36
14	-1.0504	0.3233	33	5	28	38
15	-0.9480	0.3168	35	5	30	40
16	-0.8494	0.3113	36	5	31	41
17	-0.7540	0.3065	38	5	33	43
18	-0.6614	0.3025	39	5	34	44
19	-0.5709	0.2991	40	5	35	45
20	-0.4823	0.2962	42	4	38	46
21	-0.3953	0.2938	43	4	39	47
22	-0.3096	0.2920	44	4	40	48
23	-0.2248	0.2906	46	4	42	50
24	-0.1406	0.2896	47	4	43	51
25	-0.0570	0.2891	48	4	44	52
26	0.0266	0.2890	49	4	45	53
27	0.1102	0.2893	51	4	47	55
28	0.1940	0.2901	52	4	48	56
29	0.2786	0.2913	53	4	49	57
30	0.3638	0.2929	54	4	50	58
31	0.4503	0.2951	56	4	52	60
32	0.5380	0.2977	57	4	53	61
33	0.6277	0.3010	58	5	53	63
34	0.7193	0.3048	60	5	55	65
35	0.8137	0.3094	61	5	56	66
36	0.9109	0.3147	63	5	58	68
37	1.0120	0.3210	64	5	59	69
38	1.1173	0.3283	66	5	61	71
39	1.2279	0.3369	67	5	62	72
40	1.3448	0.3470	69	5	64	74
41	1.4692	0.3590	71	5	66	76
42	1.6032	0.3734	73	6	67	79
43	1.7489	0.3908	75	6	69	81
44	1.9099	0.4123	78	6	72	84
45	2.0909	0.4397	80	7	73	87

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
46	2.2997	0.4756	84	7	77	91
47	2.5488	0.5251	87	8	79	95
48	2.8619	0.5988	92	9	-	-
49	3.2924	0.7244	98	11	-	-
50	4.0101	1.0121	98	15	-	-
51	5.4145	2.0061	98	30	-	-

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, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 5

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.4277	2.0065	2	36	-	-
1	-4.0219	1.0130	2	18	-	-
2	-3.3024	0.7256	2	13	-	-
3	-2.8703	0.6001	2	11	-	-
4	-2.5557	0.5265	6	9	-	-
5	-2.3051	0.4771	10	9	2	19
6	-2.0951	0.4412	14	8	6	22
7	-1.9126	0.4139	17	7	10	24
8	-1.7504	0.3924	20	7	13	27
9	-1.6034	0.3749	23	7	16	30
10	-1.4683	0.3605	25	6	19	31
11	-1.3428	0.3485	27	6	21	33
12	-1.2249	0.3384	30	6	24	36
13	-1.1134	0.3297	32	6	26	38
14	-1.0071	0.3224	33	6	27	39
15	-0.9053	0.3160	35	6	29	41
16	-0.8072	0.3106	37	6	31	43
17	-0.7121	0.3060	39	5	34	44
18	-0.6197	0.3020	40	5	35	45
19	-0.5296	0.2987	42	5	37	47
20	-0.4413	0.2959	44	5	39	49
21	-0.3544	0.2937	45	5	40	50
22	-0.2686	0.2919	47	5	42	52
23	-0.1838	0.2907	48	5	43	53
24	-0.0996	0.2898	50	5	45	55
25	-0.0158	0.2894	51	5	46	56
26	0.0680	0.2894	53	5	48	58
27	0.1519	0.2898	54	5	49	59
28	0.2360	0.2907	56	5	51	61
29	0.3209	0.2920	57	5	52	62
30	0.4067	0.2937	59	5	54	64
31	0.4935	0.2959	60	5	55	65
32	0.5819	0.2987	62	5	57	67
33	0.6721	0.3020	64	5	59	69
34	0.7645	0.3060	65	5	60	70

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
35	0.8595	0.3106	67	6	61	73
36	0.9577	0.3161	69	6	63	75
37	1.0595	0.3224	70	6	64	76
38	1.1658	0.3298	72	6	66	78
39	1.2774	0.3385	74	6	68	80
40	1.3954	0.3487	76	6	70	82
41	1.5211	0.3607	79	6	73	85
42	1.6563	0.3751	81	7	74	88
43	1.8035	0.3926	84	7	77	91
44	1.9659	0.4142	87	7	80	94
45	2.1486	0.4415	90	8	82	98
46	2.3590	0.4774	94	9	-	-
47	2.6099	0.5269	98	9	-	-
48	2.9249	0.6005	98	11	-	-
49	3.3575	0.7259	98	13	-	-
50	4.0775	1.0133	98	18	-	-
51	5.4837	2.0067	98	36	-	-

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, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 6

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.3923	2.0065	2	34	-	-
1	-3.9866	1.0129	2	17	-	-
2	-3.2672	0.7255	2	12	-	-
3	-2.8352	0.6001	5	10	-	-
4	-2.5205	0.5266	10	9	-	-
5	-2.2699	0.4772	14	8	6	22
6	-2.0596	0.4414	18	7	11	25
7	-1.8770	0.4142	21	7	14	28
8	-1.7145	0.3927	23	7	16	30
9	-1.5672	0.3753	26	6	20	32
10	-1.4318	0.3610	28	6	22	34
11	-1.3059	0.3490	30	6	24	36
12	-1.1877	0.3390	32	6	26	38
13	-1.0757	0.3304	34	6	28	40
14	-0.9690	0.3231	36	5	31	41
15	-0.8667	0.3168	38	5	33	43
16	-0.7681	0.3114	39	5	34	44
17	-0.6725	0.3069	41	5	36	46
18	-0.5796	0.3030	42	5	37	47
19	-0.4888	0.2997	44	5	39	49
20	-0.3997	0.2970	45	5	40	50
21	-0.3123	0.2948	47	5	42	52
22	-0.2259	0.2931	48	5	43	53
23	-0.1404	0.2918	50	5	45	55
24	-0.0554	0.2910	51	5	46	56
25	0.0291	0.2906	53	5	48	58
26	0.1135	0.2906	54	5	49	59
27	0.1981	0.2911	55	5	50	60
28	0.2831	0.2920	57	5	52	62
29	0.3687	0.2933	58	5	53	63
30	0.4551	0.2950	60	5	55	65
31	0.5428	0.2973	61	5	56	66
32	0.6321	0.3001	63	5	58	68
33	0.7231	0.3034	64	5	59	69
34	0.8163	0.3073	66	5	61	71

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
35	0.9122	0.3120	67	5	62	72
36	1.0111	0.3174	69	5	64	74
37	1.1138	0.3237	71	5	66	76
38	1.2209	0.3311	73	6	67	79
39	1.3335	0.3397	74	6	68	80
40	1.4522	0.3499	76	6	70	82
41	1.5787	0.3619	79	6	73	85
42	1.7149	0.3763	81	6	75	87
43	1.8628	0.3936	83	7	76	90
44	2.0261	0.4152	86	7	79	93
45	2.2096	0.4424	89	7	82	96
46	2.4208	0.4782	93	8	-	-
47	2.6725	0.5276	97	9	-	-
48	2.9883	0.6011	98	10	-	-
49	3.4217	0.7264	98	12	-	-
50	4.1423	1.0136	98	17	-	-
51	5.5490	2.0068	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 4.5.6. The 2010 Mod-MSA, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 7

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.5025	2.0096	2	38	-	-
1	-4.0877	1.0186	2	19	-	-
2	-3.3574	0.7327	2	14	-	-
3	-2.9155	0.6078	2	11	-	-
4	-2.5919	0.5345	3	10	-	-
5	-2.3333	0.4851	8	9	-	-
6	-2.1158	0.4493	12	8	4	20
7	-1.9264	0.4220	15	8	7	23
8	-1.7577	0.4003	19	7	12	26
9	-1.6045	0.3828	21	7	14	28
10	-1.4637	0.3683	24	7	17	31
11	-1.3325	0.3562	27	7	20	34
12	-1.2094	0.3459	29	6	23	35
13	-1.0927	0.3372	31	6	25	37
14	-0.9817	0.3297	33	6	27	39
15	-0.8750	0.3233	35	6	29	41
16	-0.7723	0.3178	37	6	31	43
17	-0.6728	0.3131	39	6	33	45
18	-0.5761	0.3091	41	6	35	47
19	-0.4817	0.3057	42	6	36	48
20	-0.3891	0.3028	44	6	38	50
21	-0.2981	0.3005	46	6	40	52
22	-0.2083	0.2987	48	6	42	54
23	-0.1195	0.2974	49	6	43	55
24	-0.0314	0.2965	51	6	45	57
25	0.0563	0.2960	52	6	46	58
26	0.1439	0.2959	54	6	48	60
27	0.2315	0.2963	56	6	50	62
28	0.3195	0.2971	57	6	51	63
29	0.4080	0.2983	59	6	53	65
30	0.4975	0.3000	61	6	55	67
31	0.5881	0.3022	62	6	56	68
32	0.6802	0.3048	64	6	58	70
33	0.7742	0.3081	66	6	60	72
34	0.8702	0.3120	68	6	62	74

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
35	0.9690	0.3166	70	6	64	76
36	1.0708	0.3219	71	6	65	77
37	1.1765	0.3282	73	6	67	79
38	1.2865	0.3355	75	6	69	81
39	1.4019	0.3440	78	6	72	84
40	1.5236	0.3541	80	7	73	87
41	1.6532	0.3660	82	7	75	89
42	1.7923	0.3803	85	7	78	92
43	1.9433	0.3976	88	7	81	95
44	2.1097	0.4190	91	8	83	98
45	2.2964	0.4461	94	8	-	-
46	2.5109	0.4817	98	9	-	-
47	2.7658	0.5308	98	10	-	-
48	3.0851	0.6040	98	11	-	-
49	3.5220	0.7289	98	14	-	-
50	4.2463	1.0155	98	19	-	-
51	5.6557	2.0076	98	38	-	-

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, Mathematics: Total Raw Score to Scale Score Conversion Table: Grade 8

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
0	-5.6620	2.0078	2	42	-	-
1	-4.2524	1.0156	2	21	-	-
2	-3.5278	0.7290	2	15	-	-
3	-3.0907	0.6043	2	13	-	-
4	-2.7710	0.5312	2	11	-	-
5	-2.5155	0.4823	5	10	-	-
6	-2.3004	0.4468	9	9	-	-
7	-2.1131	0.4197	13	9	4	22
8	-1.9461	0.3985	17	8	9	25
9	-1.7943	0.3812	20	8	12	28
10	-1.6545	0.3670	23	8	15	31
11	-1.5242	0.3551	25	7	18	32
12	-1.4017	0.3451	28	7	21	35
13	-1.2857	0.3366	30	7	23	37
14	-1.1749	0.3293	32	7	25	39
15	-1.0686	0.3230	35	7	28	42
16	-0.9660	0.3176	37	7	30	44
17	-0.8666	0.3130	39	6	33	45
18	-0.7699	0.3091	41	6	35	47
19	-0.6754	0.3058	43	6	37	49
20	-0.5828	0.3030	45	6	39	51
21	-0.4917	0.3008	47	6	41	53
22	-0.4018	0.2990	48	6	42	54
23	-0.3128	0.2977	50	6	44	56
24	-0.2245	0.2968	52	6	46	58
25	-0.1365	0.2964	54	6	48	60
26	-0.0486	0.2963	56	6	50	62
27	0.0393	0.2967	58	6	52	64
28	0.1276	0.2976	59	6	53	65
29	0.2165	0.2988	61	6	55	67
30	0.3062	0.3005	63	6	57	69
31	0.3972	0.3027	65	6	59	71
32	0.4895	0.3054	67	6	61	73
33	0.5839	0.3087	69	6	63	75
34	0.6804	0.3126	71	6	65	77
35	0.7794	0.3172	73	7	66	80

Raw Score	Proficiency Estimate	SE	SS[1]	SE(SS)	SS-1SE (SS)[2]	SS+1SE (SS)[2]
36	0.8817	0.3226	75	7	68	82
37	0.9878	0.3289	77	7	70	84
38	1.0984	0.3362	80	7	73	87
39	1.2142	0.3448	82	7	75	89
40	1.3366	0.3549	84	7	77	91
41	1.4667	0.3669	87	8	79	95
42	1.6064	0.3811	90	8	82	98
43	1.7581	0.3985	93	8	-	-
44	1.9254	0.4199	97	9	-	-
45	2.1128	0.4470	98	9	-	-
46	2.3282	0.4827	98	10	-	-
47	2.5841	0.5318	98	11	-	-
48	2.9045	0.6050	98	13	-	-
49	3.3427	0.7297	98	15	-	-
50	4.0683	1.0162	98	21	-	-
51	5.4789	2.0081	98	42	-	-

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, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 3

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	32	-	-
Algebra	1	13	17	-	-
Algebra	2	27	13	14	40
Algebra	3	36	11	25	47
Algebra	4	44	11	33	55
Algebra	5	51	11	40	62
Algebra	6	59	11	48	70
Algebra	7	68	13	55	81
Algebra	8	82	17	-	-
Algebra	9	98	32	-	-
Geometry and Measurement	0	2	32	-	-
Geometry and Measurement	1	11	17	-	-
Geometry and Measurement	2	24	12	12	36
Geometry and Measurement	3	32	11	21	43
Geometry and Measurement	4	39	10	29	49
Geometry and Measurement	5	46	10	36	56
Geometry and Measurement	6	52	10	42	62
Geometry and Measurement	7	58	10	48	68
Geometry and Measurement	8	65	11	54	76
Geometry and Measurement	9	74	12	62	86
Geometry and Measurement	10	86	16	-	-
Geometry and Measurement	11	98	32	-	-
Statistics and Probability	0	2	32	-	-
Statistics and Probability	1	17	16	-	-
Statistics and Probability	2	30	12	18	42
Statistics and Probability	3	39	11	28	50
Statistics and Probability	4	46	10	36	56
Statistics and Probability	5	52	10	42	62
Statistics and Probability	6	59	10	49	69
Statistics and Probability	7	66	11	55	77
Statistics and Probability	8	75	12	63	87
Statistics and Probability	9	88	17	-	-
Statistics and Probability	10	98	32	-	-
Numbers and Computation	0	2	32	-	-
Numbers and Computation	1	9	17	-	-
Numbers and Computation	2	22	12	10	34
Numbers and Computation	3	31	11	20	42
Numbers and Computation	4	38	10	28	48
Numbers and Computation	5	44	10	34	54
Numbers and Computation	6	50	10	40	60
Numbers and Computation	7	57	10	47	67
Numbers and Computation	8	64	11	53	75
Numbers and Computation	9	72	12	60	84

Table 4.5.8. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 3 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Numbers and Computation	10	85	17	-	-
Numbers and Computation	11	98	32	-	-
Process	0	2	32	-	-
Process	1	17	17	-	-
Process	2	31	13	18	44
Process	3	39	11	28	50
Process	4	47	10	37	57
Process	5	54	10	44	64
Process	6	60	10	50	70
Process	7	68	11	57	79
Process	8	77	13	64	90
Process	9	90	17	-	-
Process	10	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 4.5.9. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 4

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	31	-	-
Algebra	1	13	16	-	-
Algebra	2	26	12	14	38
Algebra	3	35	11	24	46
Algebra	4	43	10	33	53
Algebra	5	49	10	39	59
Algebra	6	56	10	46	66
Algebra	7	63	11	52	74
Algebra	8	72	12	60	84
Algebra	9	85	16	-	-
Algebra	10	98	31	-	-
Geometry and Measurement	0	2	31	-	-
Geometry and Measurement	1	14	16	-	-
Geometry and Measurement	2	27	12	15	39
Geometry and Measurement	3	35	11	24	46
Geometry and Measurement	4	43	10	33	53
Geometry and Measurement	5	49	10	39	59
Geometry and Measurement	6	56	10	46	66
Geometry and Measurement	7	63	11	52	74
Geometry and Measurement	8	71	12	59	83
Geometry and Measurement	9	84	16	-	-
Geometry and Measurement	10	98	31	-	-
Statistics and Probability	0	2	31	-	-
Statistics and Probability	1	13	16	-	-
Statistics and Probability	2	25	12	13	37
Statistics and Probability	3	33	10	23	43
Statistics and Probability	4	40	10	30	50
Statistics and Probability	5	46	9	37	55
Statistics and Probability	6	52	9	43	61
Statistics and Probability	7	58	10	48	68
Statistics and Probability	8	64	10	54	74
Statistics and Probability	9	73	12	61	85
Statistics and Probability	10	85	16	-	-
Statistics and Probability	11	98	31	-	-
Numbers and Computation	0	2	31	-	-
Numbers and Computation	1	11	16	-	-
Numbers and Computation	2	24	12	12	36
Numbers and Computation	3	33	11	22	44
Numbers and Computation	4	40	10	30	50
Numbers and Computation	5	46	10	36	56
Numbers and Computation	6	53	10	43	63
Numbers and Computation	7	60	11	49	71
Numbers and Computation	8	69	12	57	81
Numbers and Computation	9	81	16	65	97
Numbers and Computation	10	98	31	-	-

Table 4.5.9. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 4 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Process	0	2	31	-	-
Process	1	16	16	-	-
Process	2	28	12	16	40
Process	3	37	10	27	47
Process	4	43	10	33	53
Process	5	50	10	40	60
Process	6	56	10	46	66
Process	7	63	10	53	73
Process	8	71	12	59	83
Process	9	83	16	-	-
Process	10	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, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 5

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	36	-	-
Algebra	1	7	19	-	-
Algebra	2	22	14	8	36
Algebra	3	32	13	19	45
Algebra	4	40	12	28	52
Algebra	5	47	11	36	58
Algebra	6	55	12	43	67
Algebra	7	63	13	50	76
Algebra	8	73	14	59	87
Algebra	9	88	19	-	-
Algebra	10	98	36	-	-
Geometry and Measurement	0	2	36	-	-
Geometry and Measurement	1	11	19	2	30
Geometry and Measurement	2	26	15	11	41
Geometry and Measurement	3	37	13	24	50
Geometry and Measurement	4	46	12	34	58
Geometry and Measurement	5	54	12	42	66
Geometry and Measurement	6	62	12	50	74
Geometry and Measurement	7	70	13	57	83
Geometry and Measurement	8	81	15	66	96
Geometry and Measurement	9	96	19	-	-
Geometry and Measurement	10	98	36	-	-
Statistics and Probability	0	2	36	-	-
Statistics and Probability	1	12	19	-	-
Statistics and Probability	2	28	15	13	43
Statistics and Probability	3	39	13	26	52
Statistics and Probability	4	48	13	35	61
Statistics and Probability	5	57	13	44	70
Statistics and Probability	6	66	13	53	79
Statistics and Probability	7	77	15	62	92
Statistics and Probability	8	92	19	-	-
Statistics and Probability	9	98	36	-	-
		98	42	-	-
Numbers and Computation	0	2	36	-	-
Numbers and Computation	1	14	19	-	-
Numbers and Computation	2	28	14	14	42
Numbers and Computation	3	38	13	25	51
Numbers and Computation	4	47	12	35	59
Numbers and Computation	5	54	12	42	66
Numbers and Computation	6	62	12	50	74
Numbers and Computation	7	70	13	57	83
Numbers and Computation	8	80	14	66	94
Numbers and Computation	9	95	19	-	-
Numbers and Computation	10	98	36	-	-

Table 4.5.10. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 5 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Process	0	2	36	-	-
Process	1	8	19	-	-
Process	2	23	14	9	37
Process	3	32	12	20	44
Process	4	40	11	29	51
Process	5	46	11	35	57
Process	6	52	11	41	63
Process	7	59	11	48	70
Process	8	65	11	54	76
Process	9	73	12	61	85
Process	10	82	14	68	96
Process	11	96	19	-	-
Process	12	98	36	-	-

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, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 6

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	34	-	-
Algebra	1	11	18	-	-
Algebra	2	25	14	11	39
Algebra	3	35	12	23	47
Algebra	4	43	11	32	54
Algebra	5	50	11	39	61
Algebra	6	57	11	46	68
Algebra	7	65	12	53	77
Algebra	8	75	14	61	89
Algebra	9	89	18	-	-
Algebra	10	98	34	-	-
Geometry and Measurement	0	2	34	-	-
Geometry and Measurement	1	21	18	3	39
Geometry and Measurement	2	35	13	22	48
Geometry and Measurement	3	44	12	32	56
Geometry and Measurement	4	52	11	41	63
Geometry and Measurement	5	59	11	48	70
Geometry and Measurement	6	66	11	55	77
Geometry and Measurement	7	74	12	62	86
Geometry and Measurement	8	83	14	69	97
Geometry and Measurement	9	97	18	-	-
Geometry and Measurement	10	98	34	-	-
Statistics and Probability	0	2	34	-	-
Statistics and Probability	1	14	18	-	-
Statistics and Probability	2	28	13	15	41
Statistics and Probability	3	38	12	26	50
Statistics and Probability	4	45	11	34	56
Statistics and Probability	5	53	11	42	64
Statistics and Probability	6	60	11	49	71
Statistics and Probability	7	67	12	55	79
Statistics and Probability	8	77	13	64	90
Statistics and Probability	9	91	18	-	-
Statistics and Probability	10	98	34	-	-
Numbers and Computation	0	2	34	-	-
Numbers and Computation	1	12	18	-	-
Numbers and Computation	2	27	14	13	41
Numbers and Computation	3	36	12	24	48
Numbers and Computation	4	45	11	34	56
Numbers and Computation	5	52	11	41	63
Numbers and Computation	6	60	11	49	71
Numbers and Computation	7	68	12	56	80
Numbers and Computation	8	78	14	64	92
Numbers and Computation	9	93	18	-	-
Numbers and Computation	10	98	34	-	-

Table 4.5.11. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 6 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Process	0	-	-	-	-
Process	1	13	18	-	-
Process	2	26	13	13	39
Process	3	36	12	24	48
Process	4	43	11	32	54
Process	5	50	10	40	60
Process	6	56	10	46	66
Process	7	63	11	52	74
Process	8	70	12	58	82
Process	9	79	13	66	92
Process	10	93	18	-	-
Process	11	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 4.5.12. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 7

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	38	-	-
Algebra	1	5	20	-	-
Algebra	2	21	15	6	36
Algebra	3	32	13	19	45
Algebra	4	41	13	28	54
Algebra	5	49	12	37	61
Algebra	6	58	13	45	71
Algebra	7	67	13	54	80
Algebra	8	78	15	63	93
Algebra	9	94	20	-	-
Algebra	10	98	38	-	-
Geometry and Measurement	0	2	38	-	-
Geometry and Measurement	1	14	20	-	-
Geometry and Measurement	2	31	16	15	47
Geometry and Measurement	3	42	14	28	56
Geometry and Measurement	4	52	13	39	65
Geometry and Measurement	5	61	13	48	74
Geometry and Measurement	6	71	14	57	85
Geometry and Measurement	7	83	16	-	-
Geometry and Measurement	8	98	20	-	-
Geometry and Measurement	9	98	38	-	-
Statistics and Probability	0	2	38	-	-
Statistics and Probability	1	11	20	-	-
Statistics and Probability	2	27	15	12	42
Statistics and Probability	3	39	14	25	53
Statistics and Probability	4	48	13	35	61
Statistics and Probability	5	57	13	44	70
Statistics and Probability	6	65	13	52	78
Statistics and Probability	7	75	14	61	89
Statistics and Probability	8	86	16	-	-
Statistics and Probability	9	98	20	-	-
Statistics and Probability	10	98	38	-	-
Numbers and Computation	0	2	38	-	-
Numbers and Computation	1	2	21	-	-
Numbers and Computation	2	16	16	-	-
Numbers and Computation	3	28	14	14	42
Numbers and Computation	4	37	13	24	50
Numbers and Computation	5	46	13	33	59
Numbers and Computation	6	55	13	42	68
Numbers and Computation	7	64	14	50	78
Numbers and Computation	8	75	15	60	90
Numbers and Computation	9	91	20	-	-
Numbers and Computation	10	98	38	-	-

Table 4.5.12. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 7 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Process	0	2	38	-	-
Process	1	11	20	-	-
Process	2	26	15	11	41
Process	3	36	13	23	49
Process	4	44	12	32	56
Process	5	51	11	40	62
Process	6	57	11	46	68
Process	7	64	11	53	75
Process	8	71	12	59	83
Process	9	78	13	65	91
Process	10	88	15	-	-
Process	11	98	20	-	-
Process	12	98	38	-	-

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, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 8

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Algebra	0	2	42	-	-
Algebra	1	2	22	-	-
Algebra	2	16	17	-	-
Algebra	3	28	15	13	43
Algebra	4	37	14	23	51
Algebra	5	46	13	33	59
Algebra	6	54	13	41	67
Algebra	7	62	13	49	75
Algebra	8	72	14	58	86
Algebra	9	83	17	-	-
Algebra	10	98	22	-	-
Algebra	11	98	42	-	-
Geometry and Measurement	0	2	42	-	-
Geometry and Measurement	1	15	22	-	-
Geometry and Measurement	2	33	17	16	50
Geometry and Measurement	3	45	15	30	60
Geometry and Measurement	4	56	15	41	71
Geometry and Measurement	5	66	15	51	81
Geometry and Measurement	6	77	15	62	92
Geometry and Measurement	7	90	17	-	-
Geometry and Measurement	8	98	23	-	-
Geometry and Measurement	9	98	42	-	-
Statistics and Probability	0	2	42	-	-
Statistics and Probability	1	8	23	-	-
Statistics and Probability	2	26	17	9	43
Statistics and Probability	3	39	15	24	54
Statistics and Probability	4	49	14	35	63
Statistics and Probability	5	59	14	45	73
Statistics and Probability	6	68	14	54	82
Statistics and Probability	7	78	15	63	93
Statistics and Probability	8	90	17	-	-
Statistics and Probability	9	98	22	-	-
Statistics and Probability	10	98	42	-	-
Numbers and Computation	0	2	42	-	-
Numbers and Computation	1	3	23	-	-
Numbers and Computation	2	22	18	4	40
Numbers and Computation	3	36	16	20	52
Numbers and Computation	4	48	16	32	64
Numbers and Computation	5	60	16	44	76
Numbers and Computation	6	73	18	55	91
Numbers and Computation	7	91	23	-	-
Numbers and Computation	8	98	42	-	-

Table 4.5.13. The 2010 Mod-MSA, Mathematics: Raw Score to Scale Score Conversion by Sub-Scales/Strands: Grade 8 (Continued)

Subscale Strand	Raw Score	Scale Score (SS)[1]	Standard Error (*SEM*)	SS – 1*SEM*[2]	SS + 1*SEM*[2]
Process	0	2	42	-	-
Process	1	2	22	-	-
Process	2	18	16	2	34
Process	3	29	14	15	43
Process	4	38	13	25	51
Process	5	46	12	34	58
Process	6	53	12	41	65
Process	7	60	12	48	72
Process	8	67	12	55	79
Process	9	75	13	62	88
Process	10	84	14	70	98
Process	11	95	17	-	-
Process	12	98	22	-	-
Process	13	98	42	-	-

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: Mathematics 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: Mathematics, the test 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 mathematics 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: Mathematics 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: Mathematics

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: Mathematics 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: Mathematics Scale Score Cuts by Grades

Grade	Cut Score at Performance Level	
	Proficient	Advanced
3	54	66
4	53	67
5	54	69
6	56	69
7	54	71
8	60	73

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 Linking 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	100000198185	0.38	0.41	0.03	0.3977	0.3028
2	100000099232	0.57	0.60	0.03	-0.4626	-0.5599
3	100000198162	0.40	0.43	0.03	0.3156	0.2080
4	100000198232	0.32	0.33	0.01	0.6694	0.6908
5	100000198210	0.38	0.35	-0.03	0.3908	0.5815
6	100000198211	0.49	0.51	0.02	-0.1220	-0.1674
7	100000198212	0.66	0.63	-0.03	-0.8737	-0.6721
8	100000187836	0.59	0.61	0.02	-0.5550	-0.5970
9	100000198178	0.45	0.47	0.02	0.0892	0.0166
10	100000198235	0.44	0.48	0.04	0.1385	-0.0091
11	100000198186	0.40	0.35	-0.05	0.3190	0.5872
12	100000099325	0.58	0.65	0.07	-0.5054	-0.7925
13	100000187852	0.55	0.60	0.05	-0.3809	-0.5283
14	100000198218	0.59	0.52	-0.07	-0.5584	-0.1776
15	100000198220	0.56	0.53	-0.03	-0.4037	-0.2235
16	100000198219	0.50	0.50	0.00	-0.1446	-0.1266
17	100000187841	0.48	0.60	0.12	-0.0767	-0.5441
18	100000198170	0.18	0.23	0.05	1.4827	1.2262
19	100000099236	0.54	0.62	0.08	-0.3322	-0.6559
20	100000198202	0.58	0.62	0.04	-0.4955	-0.6344
21	100000187850	0.65	0.66	0.01	-0.7997	-0.8372
22	100000099252	0.26	0.35	0.09	1.0009	0.5815
23	100000198165	0.47	0.44	-0.03	-0.0184	0.1662
24	100000187833	0.35	0.35	0.00	0.5310	0.5929
25	100000198174	0.42	0.40	-0.02	0.1881	0.3347
Mean Rasch					-0.0082	-0.0495

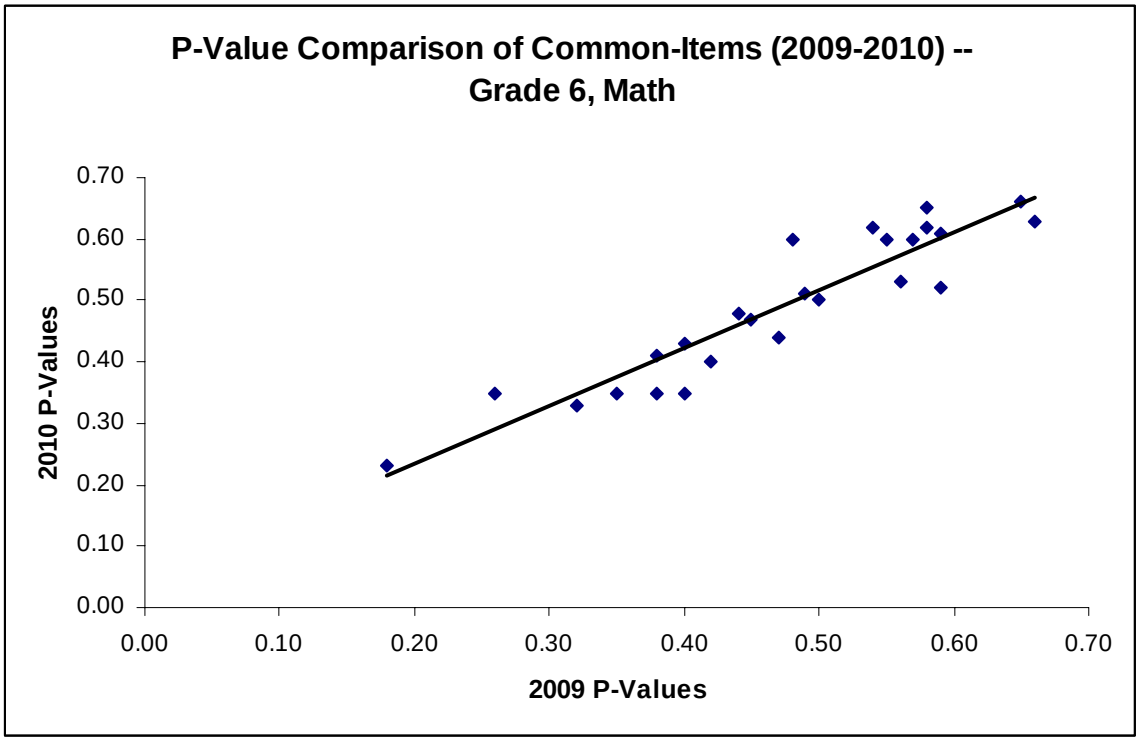


Figure 5.1. P-values of the common items for Grade 6, years 2009 and 2010

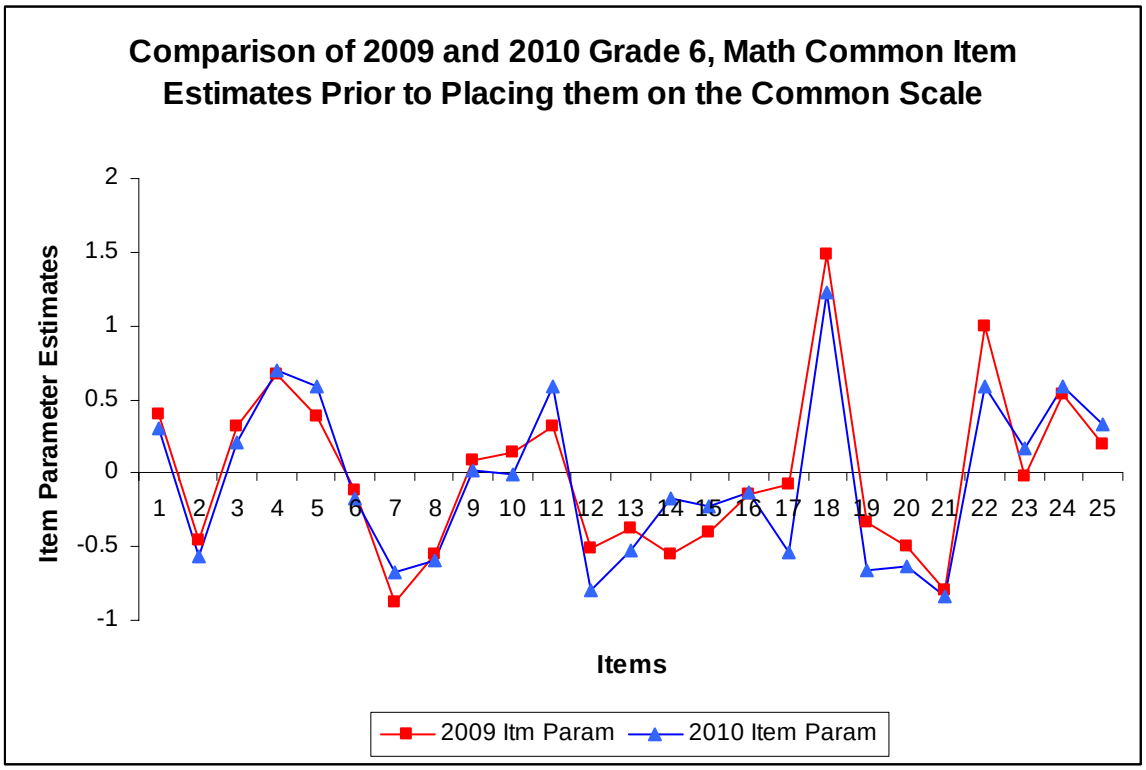


Figure 5.2. Rasch difficulty parameters for the common items, Grade 6, years 2009 and 2010

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	100000197499	0.58	0.59	0.01	-0.4387	-0.5594
2	100000197494	0.30	0.35	0.05	0.8482	0.4942
3	100000197496	0.19	0.21	0.02	1.4407	1.2821
4	100000099488	0.87	0.83	-0.04	-2.0862	-1.8465
5	100000099420	0.70	0.63	-0.07	-1.0134	-0.7346
6	100000099421	0.73	0.72	-0.01	-1.1523	-1.1412
7	100000197528	0.38	0.34	-0.04	0.4196	0.5750
8	100000374151	0.39	0.37	-0.02	0.4109	0.4289
9	100000374152	0.45	0.41	-0.04	0.1404	0.2299
10	100000196052	0.45	0.42	-0.03	0.1294	0.1945
11	100000196055	0.35	0.33	-0.02	0.5938	0.5846
12	100000197481	0.56	0.59	0.03	-0.3334	-0.5461
13	100000197501	0.39	0.46	0.07	0.3822	0.0026
14	100000099504	0.33	0.32	-0.01	0.6547	0.6430
15	100000167818	0.58	0.63	0.05	-0.4248	-0.7209
16	100000099493	0.62	0.69	0.07	-0.5911	-1.0196
17	100000197524	0.17	0.19	0.02	1.6427	1.3751
18	100000099410	0.36	0.43	0.07	0.5339	0.1329
19	100000099413	0.25	0.25	0.00	1.0713	1.0332
20	100000099404	0.45	0.49	0.04	0.1073	-0.1050
21	100000196255	0.57	0.55	-0.02	-0.3721	-0.3932
22	100000197522	0.67	0.59	-0.08	-0.8505	-0.5506
23	100000196257	0.40	0.37	-0.03	0.3308	0.3967
24	100000196261	0.51	0.47	-0.04	-0.1444	-0.0405
25	100000099490	0.41	0.46	0.05	0.2911	0.0069
Mean Rasch					0.0636	-0.0111

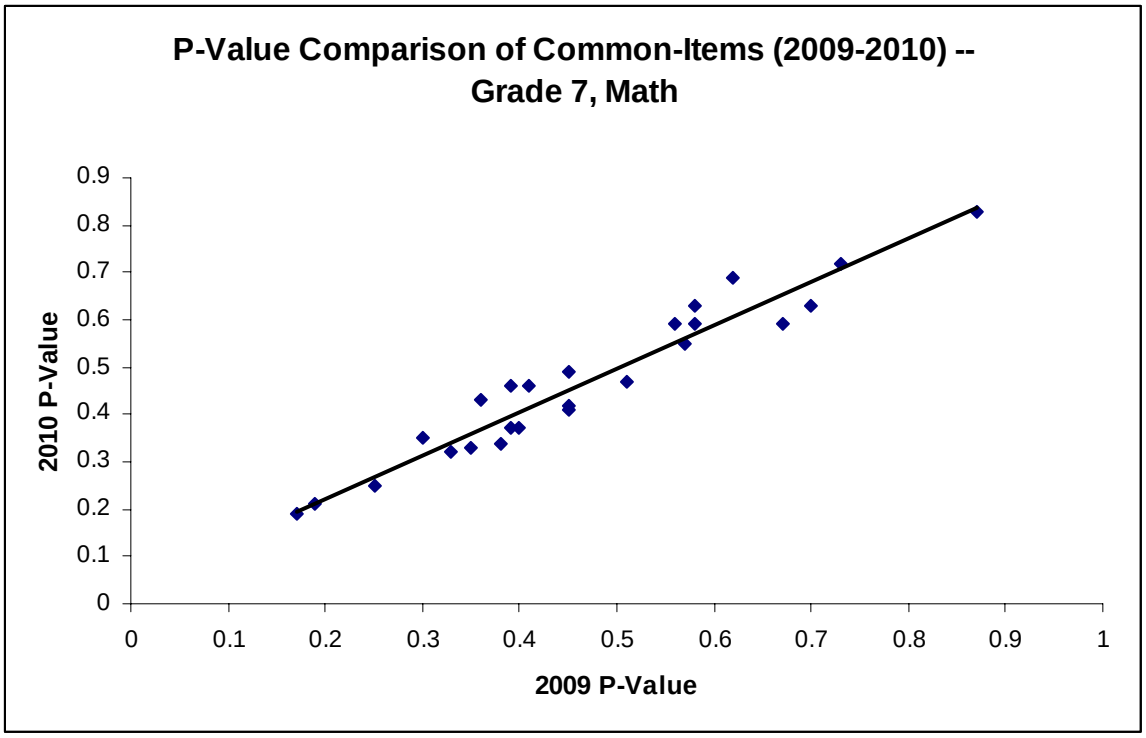


Figure 5.3. P-values of the common items for Grade 7, years 2009 and 2010

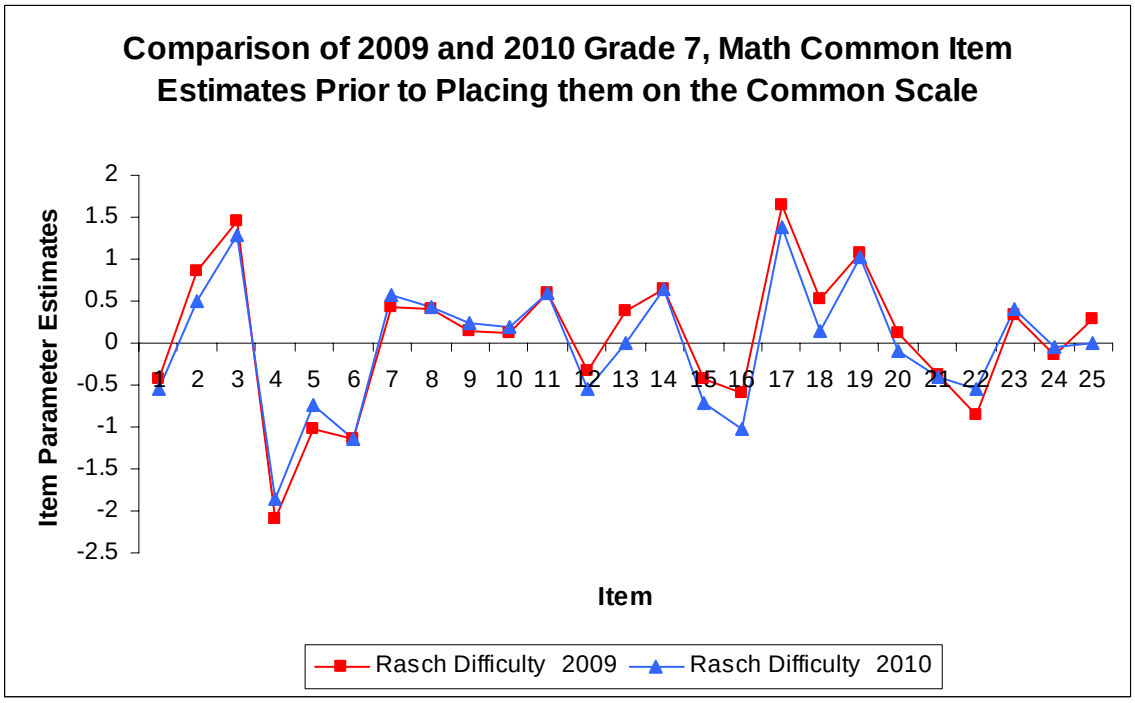


Figure 5.4. Rasch difficulty parameters for the common items, Grade 7, years 2009 and 2010

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	100000197359	0.58	0.60	0.02	-0.6919	-0.5431
2	100000099548	0.71	0.74	0.03	-1.2618	-1.1996
3	100000099550	0.40	0.43	0.03	0.1130	0.2051
4	100000197347	0.37	0.38	0.01	0.2121	0.4107
5	100000197387	0.34	0.42	0.08	0.3845	0.2656
6	100000373799	0.41	0.37	-0.04	0.0630	0.4601
7	100000099626	0.71	0.72	0.01	-1.2755	-1.1001
8	100000197367	0.32	0.36	0.04	0.4466	0.5237
9	100000167565	0.46	0.48	0.02	-0.1516	-0.0160
10	100000167566	0.29	0.30	0.01	0.6133	0.7956
11	100000174558	0.27	0.26	-0.01	0.7327	1.0196
12	100000099639	0.41	0.40	-0.01	0.0583	0.3487
13	100000197322	0.47	0.45	-0.02	-0.2045	0.1153
14	100000197365	0.42	0.38	-0.04	-0.0053	0.4197
15	100000197377	0.64	0.61	-0.03	-0.9339	-0.5869
16	100000197373	0.42	0.44	0.02	0.0276	0.1751
17	100000197375	0.19	0.18	-0.01	1.2281	1.5173
18	100000197376	0.68	0.64	-0.04	-1.1337	-0.7028
19	100000099553	0.65	0.68	0.03	-0.9660	-0.8738
20	100000373853	0.49	0.48	-0.01	-0.2869	-0.0202
21	100000167561	0.58	0.58	0.00	-0.6825	-0.4262
22	100000373854	0.48	0.52	0.04	-0.2434	-0.1886
23	100000197319	0.36	0.36	0.00	0.2761	0.5237
24	100000099647	0.48	0.52	0.04	-0.2686	-0.1971
25	100000099631	0.31	0.35	0.04	0.5341	0.5932
Mean Rasch					-0.1367	0.0601

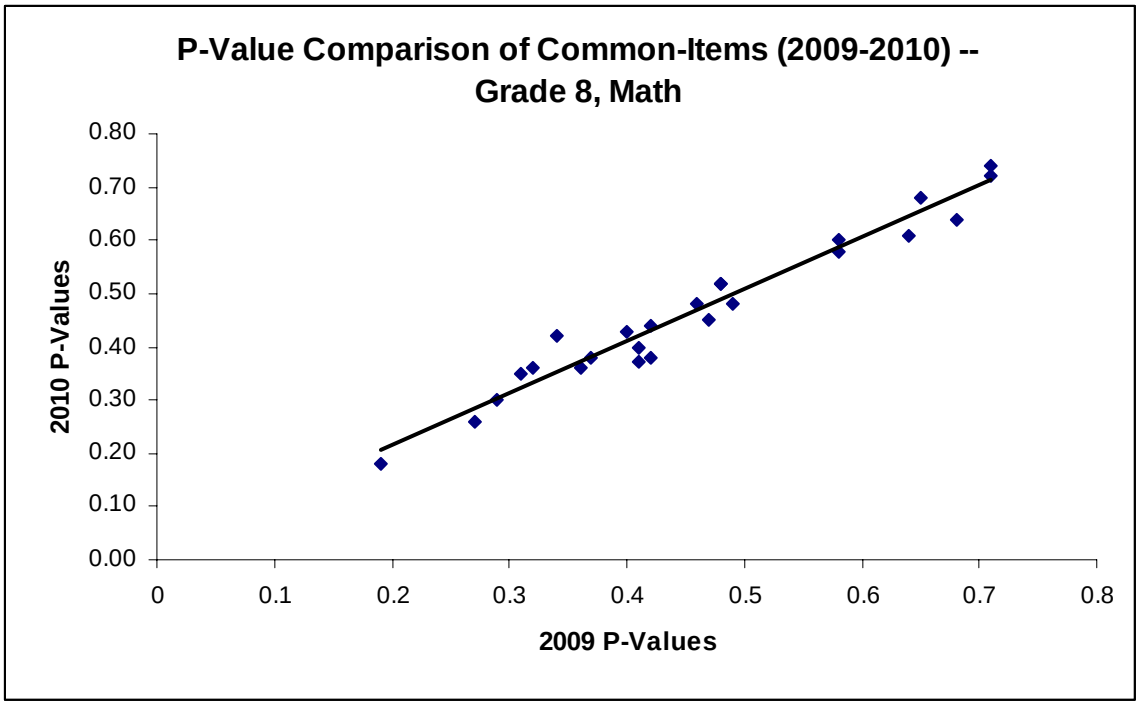


Figure 5.5. P-values of the common items for Grade 8, years 2009 and 2010

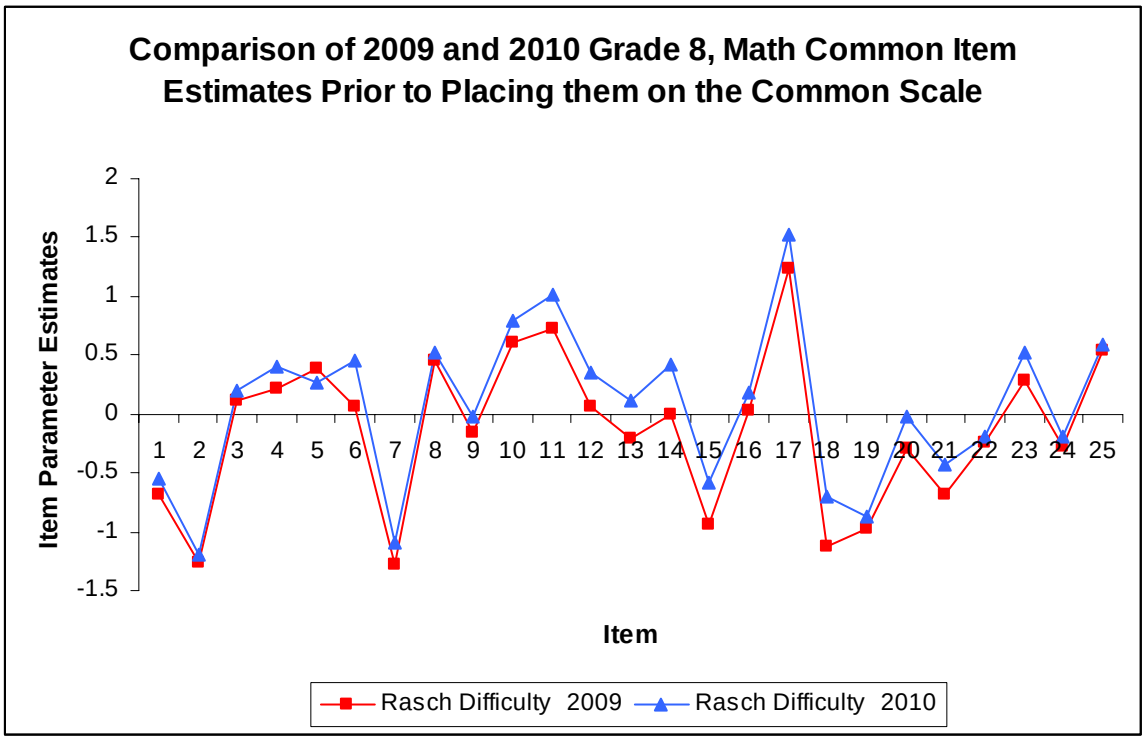


Figure 5.6. Rasch difficulty parameters for the common items, Grade 8, years 2009 and 2010

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

6.1 Rationale for the Use of the Rasch Model

In addition to reporting raw score summary statistics and item level statistics using classical test theory (CTT), the items on the Mod-MSA: Mathematics 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: Mathematics assessment. This model was 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 n-counts for the Mod-MSA: Mathematics were expected to be low.

Second, for the requirements of the Mod-MSA: Mathematics examination, 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 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 (MSQ)	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 provides 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 Mathematics IRT Item Parameters: Grade 3

Item UIN	Item Seq. No.	Item Strand No.	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000197601	1	246	669	-1.3233	0.0954	0.95	0.94
100000197604	2	246	669	0.0165	0.0816	0.96	0.97
100000098438	3	241	669	0.2773	0.0823	1.07	1.09
100000098445	4	241	669	-0.3712	0.0825	0.93	0.96
100000098439	5	241	669	-0.2097	0.0818	0.94	0.92
100000098516	6	268	669	-0.3238	0.0822	0.90	0.88
100000185401	7	273	669	0.6825	0.0855	1.03	1.07
100000185403	8	273	669	0.5604	0.0842	1.09	1.10
100000197723	9	257	669	-0.1630	0.0817	1.03	1.02
100000197745	10	257	669	-0.2901	0.0821	0.92	0.90
100000197780	11	262	669	0.1898	0.0819	1.08	1.12
100000197781	12	262	669	-0.1031	0.0816	0.98	0.95
100000098529	13	268	669	0.5533	0.0842	1.11	1.14
100000197752	14	257	669	0.4000	0.0830	0.99	1.00
100000197753	15	273	669	0.8929	0.0881	1.00	1.03
100000197754	16	273	669	1.1103	0.0916	1.13	1.20
100000197620	17	246	669	-0.9309	0.0880	1.01	1.08
100000197647	18	246	669	-0.1896	0.0818	0.99	0.98
100000098440	19	241	669	-0.8020	0.0863	1.00	0.99
100000098448	20	241	669	-0.5641	0.0838	0.92	0.89
100000185375	21	268	669	-0.4600	0.0830	0.96	0.95
100000185389	22	273	669	0.2167	0.0820	1.08	1.09
100000185391	23	273	669	0.3725	0.0828	0.96	0.95
100000197617	24	246	669	0.6535	0.0852	1.12	1.15
100000197603	25	246	669	-0.7356	0.0855	1.04	1.07
100000197677	26	257	669	0.4069	0.0830	1.12	1.14
100000197751	27	257	669	1.1527	0.0924	1.03	1.12
100000197722	28	257	669	0.3520	0.0827	1.06	1.07
100000350878	29	241	669	-0.2498	0.0819	1.01	1.00
100000098450	30	241	669	-0.3103	0.0822	0.97	0.94
100000197651	31	246	669	-1.1318	0.0914	1.06	1.18
100000197655	32	273	669	1.4240	0.0982	1.11	1.17
100000197656	33	273	669	-0.9621	0.0885	0.95	0.91
100000197741	34	257	669	0.0963	0.0817	0.93	0.91
100000197755	35	257	669	-0.0898	0.0816	1.01	1.00
100000197779	36	262	669	-0.5153	0.0834	0.90	0.86
100000197670	37	251	669	-0.1563	0.0817	0.99	0.98

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

Item UIN	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000197674	38	251	669	-0.0898	0.0816	0.96	0.96
100000197675	39	273	669	0.2100	0.0820	0.98	0.99
100000098452	40	273	669	-0.9309	0.0880	0.89	0.83
100000098444	41	241	669	0.6680	0.0853	1.06	1.07
100000197667	42	241	669	0.5960	0.0846	1.03	1.04
100000197668	43	251	669	0.1965	0.0820	1.01	1.00
100000197669	44	273	669	0.7786	0.0866	1.03	1.04
100000098451	45	273	669	-0.5501	0.0837	0.88	0.85
100000098454	46	241	669	-0.8469	0.0868	0.94	0.92
100000098447	47	241	669	1.0116	0.0899	1.11	1.18
100000185378	48	241	669	-0.8620	0.0871	0.92	0.88
100000197757	49	268	669	-0.3103	0.0822	0.87	0.85
100000197761	50	257	669	-0.3916	0.0826	0.96	0.96
100000197765	51	273	669	-0.0632	0.0816	1.10	1.12
100000098449	52	273	669	-0.5641	0.0838	0.89	0.85
100000098453	53	241	669	0.5251	0.0839	1.08	1.10
100000197602	54	241	669	-0.0632	0.0816	1.09	1.11
100000197621	55	246	669	-0.4876	0.0832	0.97	0.97
100000197660	56	246	669	-0.9621	0.0885	0.92	0.87
100000197756	57	246	669	-0.6348	0.0844	0.96	1.03
100000098446	58	257	669	0.7786	0.0866	1.04	1.06
100000098455	59	241	669	0.6607	0.0852	1.15	1.18
100000197648	60	241	669	1.0523	0.0906	1.14	1.21
100000197649	61	246	669	-0.7284	0.0854	0.94	0.92
100000197650	62	273	669	0.4970	0.0837	1.13	1.14
100000185382	63	273	669	0.3725	0.0828	1.04	1.06
100000185313	64	268	669	-1.1910	0.0925	0.91	0.83
100000197662	65	268	669	0.0630	0.0816	0.90	0.89
100000197665	66	251	669	-0.4463	0.0829	0.95	0.93
100000098522	67	251	669	-0.9001	0.0876	0.89	0.82
100000185484	68	268	669	0.4691	0.0835	1.04	1.05
100000185404	69	273	669	0.3316	0.0826	1.01	1.01
100000197746	70	273	669	0.9084	0.0884	1.17	1.22
100000197724	71	257	669	0.0231	0.0816	0.93	0.90
100000197750	72	257	669	0.5604	0.0842	1.08	1.11
100000098527	73	257	669	0.9162	0.0885	1.06	1.10
100000185381	74	268	669	-0.5153	0.0834	0.93	0.91
100000098435	75	268	669	-0.0964	0.0816	0.97	0.94

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

Item UIN	Item Seq. No.[1]	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000185485	76	241	669	0.0763	0.0817	0.93	0.91
100000185486	77	273	669	-0.5641	0.0838	0.91	0.86
100000185380	78	273	669	0.1030	0.0817	1.01	1.00
100000185384	79	268	669	-0.5711	0.0838	0.93	0.91
100000185377	80	268	669	-0.1230	0.0816	0.96	0.94
100000197720	81	268	669	0.2841	0.0823	0.99	0.99
100000185386	82	257	669	0.5040	0.0838	1.03	1.04
100000185477	83	268	669	-0.7797	0.0860	0.97	0.90
100000185473	84	273	669	0.8011	0.0869	0.96	0.96
100000197676	85	273	669	0.6318	0.0849	1.12	1.15
100000197664	86	251	669	0.5251	0.0839	1.00	1.01
100000098532	87	251	669	0.9398	0.0888	1.11	1.19
100000185387	88	268	669	0.3520	0.0827	0.93	0.92
100000185376	89	268	669	0.1831	0.0819	1.05	1.08
100000197740	90	268	669	-0.5222	0.0834	0.92	0.90
100000197661	91	257	669	-0.3036	0.0821	1.00	0.99
100000098515	92	251	669	-0.1230	0.0816	0.86	0.84
100000185385	93	268	669	1.1612	0.0926	1.10	1.14
100000197663	94	268	669	0.3452	0.0827	1.05	1.05
100000197666	95	251	669	-0.2364	0.0819	0.98	0.97
100000098441	96	251	669	-0.0765	0.0816	1.17	1.21
100000185388	97	241	669	-0.3036	0.0821	0.98	1.02
100000185390	98	273	669	-0.4738	0.0831	0.90	0.87
100000185379	99	273	669	-0.4738	0.0831	0.99	0.98
100000185383	100	268	669	-0.8169	0.0865	0.85	0.78

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criterion.

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

Item UIN	Item Seq. No.	Item Strand No	n-Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000198111	1	251	872	-0.0382	0.0711	0.94	0.94
100000198113	2	273	872	-0.2475	0.0716	0.97	0.96
100000198114	3	273	872	-0.2269	0.0715	1.06	1.10
100000198138	4	257	872	0.7870	0.0752	0.98	0.99
100000198123	5	257	872	-0.2629	0.0716	0.98	0.99
100000098578	6	241	872	0.4073	0.0722	0.93	0.96
100000186564	7	268	872	0.4177	0.0722	1.06	1.09
100000186565	8	273	872	0.8849	0.0764	1.06	1.09
100000186566	9	273	872	0.8327	0.0757	1.17	1.22
100000098583	10	241	872	0.6531	0.0739	1.06	1.07
100000098584	11	241	872	-1.3010	0.0838	0.90	0.81
100000198124	12	257	872	-0.2732	0.0717	0.99	0.98
100000198153	13	262	872	0.5394	0.0730	1.10	1.12
100000198098	14	246	872	-1.1913	0.0817	0.94	0.97
100000198100	15	246	872	0.1141	0.0712	1.06	1.07
100000198133	16	257	872	0.5394	0.0730	1.02	1.03
100000198134	17	273	872	0.1090	0.0712	0.98	0.97
100000198135	18	273	872	0.6812	0.0742	1.23	1.30
100000198110	19	251	872	-1.1847	0.0816	0.94	0.91
100000098579	20	241	872	-0.9080	0.0772	0.96	0.94
100000098581	21	241	872	0.3708	0.0720	0.97	0.96
100000198140	22	262	872	-0.5900	0.0737	0.94	0.93
100000198142	23	273	872	-0.0891	0.0712	1.05	1.05
100000198143	24	273	872	-0.0129	0.0711	1.02	1.01
100000198126	25	257	872	0.9559	0.0773	1.15	1.24
100000198128	26	257	872	0.2315	0.0714	1.07	1.07
100000198092	27	246	872	0.1396	0.0712	1.10	1.12
100000198099	28	246	872	-0.2939	0.0718	1.09	1.11
100000198117	29	251	872	0.2417	0.0715	0.93	0.92
100000198118	30	273	872	-0.0789	0.0712	0.99	0.99
100000198119	31	273	872	0.7643	0.0750	1.10	1.13
100000098577	32	241	872	-0.4665	0.0727	0.96	0.95
100000098576	33	241	872	0.6862	0.0742	1.16	1.20
100000098587	34	241	872	0.3864	0.0721	1.06	1.11
100000198116	35	251	872	1.5552	0.0880	1.16	1.37
100000198107	36	251	872	-0.1043	0.0712	0.96	0.96
100000198101	37	246	872	0.5609	0.0732	1.07	1.09

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

Item UIN	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000198102	38	273	872	0.1345	0.0712	0.99	0.99
100000198103	39	273	872	0.0125	0.0711	0.98	0.97
100000198148	40	262	872	0.1701	0.0713	1.01	1.00
100000198158	41	262	872	0.0887	0.0712	0.94	0.91
100000098646	42	268	872	-1.3436	0.0847	0.89	0.77
100000186562	43	273	872	-0.7340	0.0751	0.93	0.90
100000186563	44	273	872	0.2983	0.0717	1.06	1.07
100000198094	45	246	872	-1.3436	0.0847	0.92	0.86
100000198096	46	246	872	-0.3042	0.0718	0.96	0.98
100000198129	47	257	872	-0.0789	0.0712	0.93	0.92
100000198131	48	273	872	0.7643	0.0750	1.06	1.07
100000198132	49	273	872	0.5074	0.0728	1.04	1.07
100000198093	50	246	872	-2.3400	0.1165	0.89	0.59
100000098572	51	241	872	0.5021	0.0728	0.93	0.92
100000098586	52	241	872	0.1192	0.0712	0.97	0.96
100000198104	53	246	872	-0.1706	0.0714	1.09	1.11
100000198105	54	273	872	-0.1196	0.0712	1.01	0.99
100000198106	55	273	872	0.3396	0.0718	1.03	1.04
100000198139	56	257	872	0.4230	0.0723	1.04	1.05
100000098585	57	241	872	-1.3872	0.0857	0.93	0.85
100000098582	58	241	872	-0.0484	0.0711	0.93	0.91
100000207144	59	251	872	0.4703	0.0725	1.07	1.08
100000198144	60	262	872	-0.8140	0.0760	0.86	0.79
100000198145	61	273	872	0.5394	0.0730	0.99	1.00
100000198147	62	273	872	0.1039	0.0712	1.02	1.02
100000186576	63	268	872	-1.0052	0.0786	0.91	0.88
100000098666	64	268	872	-0.0738	0.0712	1.02	1.02
100000198150	65	262	872	-0.8961	0.0770	0.87	0.80
100000198149	66	262	872	-0.2115	0.0715	0.96	0.94
100000186573	67	268	872	0.6917	0.0743	0.99	1.00
100000186574	68	273	872	0.0887	0.0712	1.01	1.02
100000186575	69	273	872	0.0887	0.0712	0.98	0.96
100000186583	70	268	872	0.0510	0.0711	1.23	1.31
100000198108	71	251	872	0.3189	0.0718	0.95	0.94
100000198122	72	251	872	0.6367	0.0738	1.06	1.06
100000098645	73	268	872	0.3189	0.0718	1.01	1.02
100000186578	74	268	872	-1.0551	0.0794	0.90	0.81
100000186580	75	268	872	-0.5198	0.0731	0.89	0.86

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

Item CID	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000198157	76	262	872	0.8908	0.0764	1.02	1.08
100000098573	77	241	872	0.3864	0.0721	0.99	1.00
100000186560	78	273	872	-0.0789	0.0712	1.04	1.04
100000186561	79	273	872	0.7362	0.0747	1.05	1.07
100000198137	80	257	872	0.2983	0.0717	0.96	0.95
100000098571	81	241	872	-0.4032	0.0723	0.92	0.89
100000186581	82	268	872	-0.4665	0.0727	0.97	0.96
100000207143	83	268	872	-0.4137	0.0724	0.94	0.91
100000198121	84	251	872	0.4651	0.0725	1.03	1.03
100000098588	85	241	872	-0.1859	0.0714	0.98	0.98
100000098657	86	268	872	0.7306	0.0747	1.02	1.05
100000186571	87	273	872	0.2008	0.0714	1.10	1.12
100000186572	88	273	872	0.1854	0.0713	1.01	1.00
100000198125	89	257	872	-0.1808	0.0714	0.93	0.91
100000198151	90	262	872	-0.9440	0.0777	0.90	0.84
100000098653	91	268	872	0.8327	0.0757	1.04	1.06
100000186577	92	268	872	-0.9928	0.0784	0.91	0.86
100000098568	93	241	872	-0.6063	0.0738	0.91	0.85
100000186558	94	273	872	-0.2732	0.0717	0.93	0.90
100000186559	95	273	872	0.1039	0.0712	1.04	1.04
100000186567	96	268	872	-0.1502	0.0713	0.97	0.97
100000186582	97	268	872	0.5716	0.0733	1.05	1.06
100000198127	98	257	872	-0.1808	0.0714	0.93	0.91
100000098664	99	268	872	-0.0230	0.0711	0.97	0.96
100000098580	100	241	872	0.8732	0.0762	1.07	1.15
100000186579	101	268	872	0.0531	0.0711	0.96	0.95
100000198109	102	251	872	0.4915	0.0727	0.97	0.97

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criterion.

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

Item CID	Item Seq. No.	Item Strand No	n-Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000187371	1	268	926	-1.7441	0.0884	0.90	0.79
100000187367	2	268	926	0.0464	0.0684	1.01	1.00
100000187378	3	268	926	-0.9103	0.0727	0.94	0.92
100000099084	4	241	926	-0.8268	0.0718	0.90	0.88
100000099080	5	241	926	-0.7302	0.0709	0.91	0.88
100000099090	6	241	926	0.3591	0.0697	1.04	1.05
100000099087	7	241	926	1.0158	0.0772	1.08	1.14
100000196233	8	251	926	-0.7706	0.0712	0.91	0.87
100000196246	9	251	926	2.0213	0.1031	1.07	1.40
100000196229	10	251	926	0.0931	0.0685	0.99	0.99
100000196273	11	257	926	0.7665	0.0736	0.93	0.93
100000196043	12	273	926	0.2106	0.0689	0.96	0.96
100000196045	13	273	926	-0.1950	0.0682	1.01	1.02
100000196253	14	257	926	-0.7102	0.0707	0.89	0.87
100000196278	15	257	926	0.8656	0.0750	0.94	0.94
100000196088	16	246	926	0.6498	0.0723	1.05	1.06
100000187430	17	268	926	-0.3067	0.0684	0.96	0.95
100000187390	18	268	926	-1.1763	0.0764	0.97	0.95
100000196090	19	246	926	0.8544	0.0748	1.04	1.06
100000196081	20	246	926	0.5929	0.0717	1.05	1.06
100000187376	21	268	926	-0.2974	0.0684	1.03	1.06
100000187386	22	268	926	-1.0016	0.0739	1.03	1.04
100000187383	23	268	926	0.3156	0.0695	1.02	1.04
100000196231	24	251	926	0.3639	0.0698	1.03	1.04
100000196036	25	273	926	-0.9156	0.0728	0.96	0.95
100000196029	26	273	926	-0.2834	0.0684	0.96	0.95
100000196242	27	251	926	-0.5340	0.0694	0.93	0.91
100000099083	28	241	926	-0.2043	0.0682	1.02	1.04
100000099075	29	241	926	-0.2322	0.0683	0.93	0.92
100000099086	30	241	926	-0.6164	0.0700	0.92	0.90
100000196200	31	246	926	-0.6507	0.0702	0.95	0.94
100000196277	32	257	926	-0.0883	0.0682	0.98	0.98
100000196284	33	262	926	-0.6507	0.0702	0.92	0.88
100000196057	34	273	926	-0.0048	0.0683	0.98	0.97
100000196054	35	273	926	0.0371	0.0683	0.98	0.97
100000196279	36	262	926	0.9052	0.0755	1.01	1.05
100000099081	37	241	926	0.3931	0.0700	1.06	1.06

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

Item CID	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000099079	38	241	926	-0.6458	0.0702	0.98	0.97
100000099088	39	241	926	-0.2554	0.0683	1.06	1.09
100000099085	40	241	926	-0.9637	0.0734	0.93	0.89
100000196237	41	251	926	0.3834	0.0699	1.09	1.10
100000196234	42	251	926	-1.2535	0.0777	0.95	0.92
100000196053	43	273	926	-0.7152	0.0707	0.93	0.90
100000196244	44	251	926	1.3858	0.0845	1.04	1.14
100000099089	45	241	926	-1.1243	0.0756	0.90	0.84
100000099091	46	241	926	-0.2508	0.0683	0.96	0.95
100000099092	47	241	926	-0.2834	0.0684	1.00	0.99
100000187365	48	273	926	0.3542	0.0697	1.00	1.00
100000187364	49	273	926	0.0931	0.0685	0.99	1.00
100000196381	50	262	926	-1.3147	0.0788	0.89	0.83
100000196281	51	262	926	0.2916	0.0693	1.04	1.04
100000187387	52	268	926	1.2615	0.0818	1.03	1.07
100000187388	53	273	926	0.3494	0.0697	1.08	1.10
100000187389	54	273	926	0.5674	0.0714	1.12	1.15
100000196270	55	257	926	0.5878	0.0716	0.95	0.95
100000196256	56	257	926	-0.2043	0.0682	0.95	0.95
100000196267	57	257	926	-1.9350	0.0938	0.92	0.80
100000099178	58	241	926	-0.2136	0.0682	0.98	0.98
100000187369	59	268	926	0.3397	0.0696	1.04	1.05
100000187380	60	268	926	0.3785	0.0699	1.05	1.05
100000187429	61	268	926	-0.3301	0.0685	0.94	0.93
100000187428	62	268	926	-0.4670	0.0691	0.94	0.92
100000187372	63	268	926	1.1509	0.0796	1.08	1.16
100000099180	64	268	926	0.1352	0.0686	0.96	0.97
100000099072	65	241	926	-0.3395	0.0685	0.92	0.90
100000187360	66	273	926	0.4274	0.0702	1.05	1.07
100000187361	67	273	926	-0.1393	0.0682	0.94	0.92
100000196260	68	257	926	0.2964	0.0693	1.10	1.13
100000196258	69	257	926	0.7182	0.0730	1.09	1.11
100000196269	70	257	926	1.2415	0.0814	0.98	1.02
100000196223	71	246	926	-0.1857	0.0682	1.01	1.03
100000196100	72	246	926	-1.5658	0.0840	0.90	0.82
100000196094	73	246	926	0.0464	0.0684	0.99	0.99
100000187373	74	273	926	0.5725	0.0715	1.01	1.01
100000187363	75	273	926	0.5068	0.0709	1.08	1.11

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

Item CID	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000196037	76	273	926	0.0557	0.0684	1.04	1.05
100000187370	77	268	926	0.0650	0.0684	0.97	0.97
100000187391	78	268	926	0.4323	0.0703	0.97	0.97
100000187392	79	268	926	0.3494	0.0697	1.07	1.09
100000187382	80	268	926	0.3108	0.0694	1.05	1.06
100000196263	81	257	926	-0.8372	0.0719	0.96	0.94
100000196247	82	257	926	0.2964	0.0693	0.97	0.96
100000196238	83	251	926	0.4225	0.0702	1.08	1.10
100000196079	84	246	926	0.3445	0.0696	1.01	1.01
100000196025	85	273	926	-0.5291	0.0694	1.00	0.99
100000187393	86	273	926	0.9109	0.0756	1.14	1.18
100000196225	87	246	926	0.7611	0.0736	1.09	1.13
100000099177	88	241	926	-0.0001	0.0683	0.95	0.94
100000187366	89	273	926	0.1258	0.0686	1.00	1.00
100000196049	90	273	926	-0.4243	0.0689	1.00	0.99
100000196235	91	251	926	0.9280	0.0759	1.09	1.14
100000099082	92	241	926	0.1634	0.0687	1.10	1.11
100000196042	93	273	926	0.3639	0.0698	1.00	1.02
100000187385	94	268	926	-0.0698	0.0682	1.02	1.02
100000187374	95	273	926	0.1165	0.0685	1.11	1.11
100000187424	96	273	926	0.3204	0.0695	1.04	1.06
100000187379	97	268	926	-0.2881	0.0684	0.92	0.93
100000187377	98	268	926	1.1700	0.0800	1.08	1.17
100000187381	99	268	926	1.0398	0.0776	1.02	1.02
100000187375	100	268	926	0.4521	0.0704	1.06	1.07

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criterion.

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

Item CID	Item Seq. No.	Item Strand No	n-Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000272103	1	246	852	-1.1141	0.0789	1.02	1.05
100000272105	2	246	852	-0.6106	0.0732	1.03	1.05
100000272100	3	246	852	0.8584	0.0785	1.06	1.11
100000272095	4	246	852	1.3776	0.0880	1.03	1.09
100000272117	5	251	852	0.3108	0.0729	0.98	0.97
100000198185	6	251	852	0.3977	0.0735	0.94	0.93
100000272111	7	251	852	0.5886	0.0752	1.10	1.12
100000272115	8	251	852	0.4885	0.0742	1.11	1.11
100000272113	9	251	852	0.3055	0.0729	0.94	0.94
100000272076	10	241	852	-0.4521	0.0723	0.96	0.97
100000272073	11	241	852	-0.2201	0.0716	1.06	1.08
100000099232	12	241	852	-0.4626	0.0724	0.88	0.84
100000272096	13	246	852	-0.3021	0.0718	0.93	0.90
100000272145	14	273	852	0.0661	0.0718	1.01	1.01
100000272146	15	273	852	-0.0210	0.0716	1.05	1.06
100000198162	16	246	852	0.3156	0.0729	1.06	1.08
100000272101	17	246	852	0.1022	0.0719	1.01	1.01
100000198232	18	262	852	0.6694	0.0760	1.01	1.04
100000272084	19	262	852	0.6396	0.0757	1.18	1.22
100000198210	20	257	852	0.3908	0.0734	0.92	0.91
100000198211	21	273	852	-0.1220	0.0715	0.98	0.97
100000198212	22	273	852	-0.8737	0.0757	0.97	0.92
100000272129	23	257	852	-0.1792	0.0716	0.96	0.94
100000187836	24	268	852	-0.5550	0.0729	0.95	0.91
100000272156	25	268	852	-0.5098	0.0726	0.97	0.96
100000272158	26	268	852	0.3002	0.0728	1.09	1.09
100000272159	27	268	852	0.4177	0.0736	1.04	1.04
100000272066	28	241	852	-1.1644	0.0797	0.91	0.83
100000272077	29	241	852	0.7797	0.0774	1.05	1.08
100000272068	30	241	852	0.5438	0.0747	1.04	1.06
100000272075	31	241	852	-0.9925	0.0771	0.95	0.92
100000272072	32	241	852	-0.5679	0.0730	1.00	1.02
100000272067	33	241	852	-0.7854	0.0747	0.88	0.81
100000272131	34	257	852	0.1954	0.0723	0.99	0.99
100000272127	35	257	852	-0.6267	0.0734	0.90	0.87
100000272134	36	257	852	-0.7465	0.0744	0.94	0.93
100000273844	37	241	852	-0.0670	0.0716	1.05	1.06

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

Item CID	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000273626	38	273	852	0.7857	0.0775	1.06	1.07
100000272143	39	273	852	-0.2713	0.0717	0.92	0.91
100000272144	40	273	852	0.1331	0.0720	0.99	0.99
100000198178	41	246	852	0.0892	0.0719	1.00	0.99
100000272099	42	246	852	0.9521	0.0799	1.10	1.18
100000272148	43	273	852	0.2581	0.0726	1.08	1.10
100000272149	44	273	852	-0.0261	0.0716	1.07	1.09
100000198235	45	262	852	0.1385	0.0720	1.05	1.05
100000272080	46	262	852	-0.9511	0.0766	0.94	0.90
100000272082	47	262	852	-0.7854	0.0747	0.94	0.94
100000272114	48	251	852	1.0429	0.0814	1.02	1.09
100000272116	49	251	852	0.3533	0.0732	0.96	0.96
100000198186	50	251	852	0.3190	0.0729	0.93	0.93
100000272112	51	251	852	0.6396	0.0757	1.11	1.14
100000099325	52	241	852	-0.5054	0.0726	0.84	0.80
100000272128	53	257	852	-0.3227	0.0718	0.97	0.97
100000272125	54	257	852	0.1486	0.0721	1.02	1.02
100000272069	55	241	852	-0.4469	0.0723	0.93	0.92
100000272155	56	268	852	1.2143	0.0845	1.17	1.27
100000272166	57	268	852	0.4070	0.0736	0.99	0.99
100000272167	58	268	852	1.1235	0.0828	0.98	0.98
100000187852	59	268	852	-0.3809	0.0720	0.91	0.88
100000272065	60	241	852	0.2110	0.0723	0.92	0.91
100000198218	61	257	852	-0.1537	0.0715	0.98	0.97
100000198220	62	273	852	-0.4037	0.0721	0.93	0.91
100000198219	63	273	852	-0.1446	0.0715	0.94	0.92
100000187841	64	268	852	-0.5203	0.0727	0.95	0.93
100000272161	65	268	852	1.0297	0.0811	1.04	1.09
100000272163	66	268	852	0.1435	0.0721	0.94	0.93
100000198170	67	246	852	1.4827	0.0905	1.11	1.25
100000272104	68	246	852	0.1538	0.0721	1.06	1.07
100000272150	69	273	852	0.1902	0.0723	1.05	1.06
100000272151	70	273	852	0.9585	0.0800	1.14	1.17
100000272098	71	246	852	-0.5203	0.0727	0.98	0.97
100000272071	72	241	852	-1.2222	0.0806	0.96	0.96
100000099236	73	241	852	-0.3322	0.0718	0.91	0.89
100000273837	74	241	852	-0.7631	0.0745	0.93	0.90
100000273598	75	273	852	-0.1333	0.0715	1.03	1.04

Table 6.3. 4. The 2010 Mod-MSA Mathematics IRT Item Parameters: Grade 6 (Continued)

Item CID	Item Seq. No.	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000272141	76	273	852	-0.6375	0.0734	0.97	0.96
100000198202	77	257	852	-0.4955	0.0725	0.90	0.87
100000272126	78	257	852	-0.2662	0.0717	0.97	0.96
100000272133	79	257	852	0.4394	0.0738	0.94	0.93
100000272132	80	257	852	-0.6699	0.0737	0.93	0.94
100000272165	81	268	852	-0.2150	0.0716	1.09	1.11
100000272152	82	273	852	0.4340	0.0738	1.03	1.03
100000273858	83	273	852	0.3373	0.0731	1.07	1.08
100000272160	84	268	852	-0.9278	0.0763	0.93	0.89
100000187850	85	268	852	-0.7997	0.0749	0.93	0.92
100000272157	86	268	852	1.5820	0.0930	0.98	1.09
100000272154	87	273	852	0.1022	0.0719	1.09	1.12
100000272153	88	273	852	-0.4053	0.0721	0.99	0.99
100000272162	89	268	852	-0.2867	0.0717	0.96	0.96
100000099252	90	241	852	0.6055	0.0753	1.10	1.12
100000198165	91	273	852	-0.0184	0.0716	0.97	0.96
100000272097	92	246	852	-0.1486	0.0715	1.00	0.99
100000187833	93	273	852	0.5310	0.0746	1.01	1.01
100000272102	94	246	852	0.0507	0.0718	1.16	1.20
100000198174	95	273	852	0.1881	0.0722	1.02	1.03
100000272164	96	268	852	0.6282	0.0756	1.04	1.06

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
2. These analyses are based on the equating sample used with the exclusion criterion.

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

Item CID	Item Seq. No.[1]	Item Strand No	n-Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000322134	1	246	1000	-0.6553	0.0677	0.94	0.93
100000273428	2	273	1000	0.8808	0.0722	1.05	1.07
100000273431	3	246	1000	0.9870	0.0737	1.10	1.14
100000272267	4	273	1000	-1.5249	0.0801	1.01	1.10
100000197499	5	257	1000	-0.4387	0.0664	0.93	0.91
100000272217	6	257	1000	-0.4978	0.0667	1.01	1.03
100000322127	7	246	1000	1.4048	0.0815	0.97	0.97
100000322131	8	246	1000	1.7655	0.0907	1.08	1.23
100000197494	9	251	1000	0.8482	0.0718	1.10	1.14
100000322136	10	251	1000	1.0089	0.0741	1.15	1.22
100000322138	11	251	1000	1.0143	0.0742	0.99	1.03
100000197496	12	251	1000	1.4407	0.0823	1.02	1.07
100000322166	13	241	1000	-0.6370	0.0676	0.96	0.95
100000322165	14	241	1000	0.2217	0.0664	1.03	1.03
100000322169	15	241	1000	0.6849	0.0698	1.08	1.09
100000099488	16	268	1000	-2.0862	0.0953	1.15	0.96
100000322151	17	268	1000	0.6462	0.0694	1.02	1.04
100000322141	18	268	1000	-0.2533	0.0658	0.98	0.97
100000322152	19	268	1000	-0.1758	0.0657	1.00	0.99
100000322140	20	268	1000	-0.4140	0.0663	1.01	1.02
100000287243	21	273	1000	-1.3008	0.0757	0.96	0.91
100000272268	22	273	1000	-0.6325	0.0675	1.01	1.02
100000287244	23	251	1000	1.6696	0.0880	0.99	1.05
100000272270	24	273	1000	0.7142	0.0702	1.06	1.09
100000322137	25	251	1000	0.7787	0.0709	1.05	1.08
100000322139	26	251	1000	1.0198	0.0742	1.00	1.01
100000272201	27	251	1000	0.4451	0.0677	1.13	1.15
100000322164	28	241	1000	-0.2792	0.0659	0.99	0.98
100000272171	29	241	1000	-0.5378	0.0669	1.04	1.05
100000099420	30	241	1000	-1.0134	0.0713	1.17	1.30
100000099421	31	241	1000	-1.1523	0.0733	0.98	0.96
100000272271	32	273	1000	0.4679	0.0678	1.03	1.03
100000272216	33	257	1000	0.2305	0.0664	1.07	1.08
100000272272	34	273	1000	0.0392	0.0658	1.01	1.01
100000322175	35	262	1000	-0.4009	0.0662	0.98	0.97
100000322172	36	262	1000	0.7636	0.0707	1.00	1.01
100000197528	37	262	1000	0.4196	0.0675	0.97	0.97

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

Item CID	Item Seq. No.[1]	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000374151	38	273	1000	0.4109	0.0674	1.10	1.12
100000374152	39	246	1000	0.1404	0.0661	1.11	1.12
100000196052	40	273	1000	0.1294	0.0660	1.12	1.13
100000196055	41	273	1000	0.5938	0.0689	1.02	1.04
100000322129	42	246	1000	-0.6098	0.0674	0.97	0.96
100000197481	43	246	1000	-0.3334	0.0660	0.91	0.89
100000273423	44	273	1000	0.0565	0.0658	1.02	1.02
100000273427	46		1000	0.4725	0.0679	1.12	1.13
100000272203	48	241	1000	-0.1586	0.0657	0.92	0.90
100000322153	49		1000	-0.4140	0.0663	0.90	0.88
100000272246	50	251	1000	0.8652	0.0720	1.09	1.11
100000322154	51	268	1000	-0.5872	0.0672	1.00	1.00
100000322128	52	268	1000	0.8860	0.0723	1.03	1.09
100000322173	53	268	1000	0.6127	0.0691	1.11	1.12
100000322174	54	246	1000	0.9122	0.0726	1.10	1.16
100000273415	55	262	1000	0.8860	0.0723	1.11	1.14
100000273416	56	262	1000	1.6088	0.0864	1.02	1.09
100000272274	57	273	1000	0.2878	0.0667	1.04	1.05
100000272186	58	262	1000	-0.7524	0.0685	0.99	1.00
100000322135	59	273	1000	-0.7618	0.0686	0.92	0.88
100000322132	60	246	1000	0.0047	0.0657	1.06	1.06
100000272187	61	246	1000	0.0694	0.0659	0.99	0.98
100000374196	63	246	1000	0.6414	0.0694	0.97	0.98
100000272248	64	246	1000	-0.4228	0.0663	1.01	1.01
100000322146	65		1000	0.6995	0.0700	1.05	1.09
100000197501	66	268	1000	0.3822	0.0672	1.06	1.06
100000322158	67	268	1000	0.1692	0.0662	0.83	0.81
100000322160	68	268	1000	0.0047	0.0657	0.91	0.90
100000322163	69	257	1000	0.0910	0.0659	0.97	0.97
100000272174	70	241	1000	0.2878	0.0667	0.91	0.90
100000099504	71	241	1000	0.6547	0.0695	0.98	1.01
100000322142	72	241	1000	-1.5249	0.0801	0.96	0.91
100000167818	73	241	1000	-0.4248	0.0663	0.84	0.82
100000322143	74	268	1000	-1.5638	0.0809	0.93	0.87
100000322147	75	268	1000	-0.0598	0.0657	0.96	0.96
100000099493	76	268	1000	-0.5911	0.0672	0.90	0.88
100000273417	77	268	1000	0.0435	0.0658	1.06	1.07
100000273444	78	268	1000	0.0392	0.0658	0.90	0.88

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

Item CID	Item Seq. No.[1]	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000273445	79	273	1000	0.6558	0.0695	1.04	1.06
100000197524	80	262	1000	1.6427	0.0873	1.10	1.13
100000272247	81	268	1000	-0.9105	0.0701	0.97	0.94
100000322149	82	268	1000	-0.8909	0.0699	0.89	0.84
100000322150	83	268	1000	-0.0984	0.0657	1.03	1.03
100000322144	84	268	1000	-1.1891	0.0738	0.96	0.93
100000322148	85	268	1000	-0.5737	0.0671	0.87	0.85
100000099410	86	241	1000	0.5339	0.0684	1.00	1.03
100000322161	87	241	1000	0.2657	0.0666	0.98	0.97
100000099413	88	241	1000	1.0713	0.0751	0.99	1.00
100000322157	89	241	1000	-1.2275	0.0745	0.98	0.96
100000099404	90	241	1000	0.1073	0.0660	0.89	0.88
100000196255	91	273	1000	-0.3721	0.0661	1.01	1.01
100000197522	92	257	1000	-0.8505	0.0694	1.11	1.13
100000196257	93	273	1000	0.3308	0.0669	1.01	1.01
100000196261	94	273	1000	-0.1444	0.0657	1.09	1.10
100000322171	95	241	1000	-0.0039	0.0657	0.89	0.88
100000322162	96	241	1000	0.0608	0.0658	0.96	0.97
100000322159	97	241	1000	0.2569	0.0665	0.92	0.92
100000099490	98	268	1000	0.2911	0.0667	1.00	0.99

Note: 1. Item #s 46, 49, and 65 were not scored
2. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
3. These analyses are based on the equating sample used with the exclusion criterion.

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

Item CID	Item Seq. No.	Item Strand No	n-Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000322198	1	241	1017	-0.2442	0.0650	1.14	1.23
100000322187	2	241	1017	0.1063	0.0662	1.07	1.08
100000197359	3	251	1017	-0.6919	0.0659	0.98	0.97
100000272279	4	251	1017	1.6028	0.0915	1.04	1.04
100000272278	5	251	1017	0.2661	0.0673	1.08	1.10
100000272340	6	241	1017	-0.1682	0.0651	0.99	0.99
100000099548	7	241	1017	-1.2618	0.0711	0.90	0.86
100000322121	8	262	1017	-0.4506	0.0651	0.98	0.97
100000322186	9	241	1017	-1.3614	0.0726	1.00	1.02
100000322122	10	262	1017	-0.6724	0.0658	1.11	1.16
100000322120	11	262	1017	-0.4971	0.0652	0.95	0.93
100000099550	12	241	1017	0.1130	0.0663	0.96	0.95
100000272296	13	273	1017	0.8104	0.0738	1.09	1.15
100000272337	14	241	1017	0.7461	0.0728	1.00	1.00
100000272295	15	273	1017	0.6886	0.0720	1.04	1.06
100000272297	16	273	1017	-0.3452	0.0650	1.06	1.07
100000272323	17	257	1017	-0.0706	0.0654	1.08	1.12
100000322119	18	257	1017	-0.6207	0.0656	1.00	0.98
100000322118	19	257	1017	-0.1682	0.0651	0.94	0.93
100000322084	20	246	1017	0.5331	0.0700	1.08	1.09
100000197347	21	246	1017	0.2121	0.0669	1.01	1.02
100000322081	22	246	1017	-0.4802	0.0651	0.99	0.98
100000272309	23	246	1017	0.4847	0.0694	1.11	1.13
100000322209	24	268	1017	-0.4379	0.0651	1.01	0.99
100000322208	25	268	1017	-0.9425	0.0676	0.95	0.93
100000322205	26	268	1017	0.0367	0.0658	1.07	1.07
100000322207	27	268	1017	-1.6002	0.0766	0.94	0.89
100000197387	28	262	1017	0.3845	0.0684	0.99	0.98
100000286862	29	273	1017	0.2078	0.0669	1.12	1.17
100000272790	30	273	1017	-0.6465	0.0657	0.94	0.92
100000286872	31	262	1017	0.1457	0.0665	1.09	1.11
100000373799	32	268	1017	0.0630	0.0660	0.98	0.98
100000272368	33	268	1017	0.4370	0.0689	1.14	1.18
100000322200	34	268	1017	0.4181	0.0687	0.94	0.94
100000099626	35	268	1017	-1.2755	0.0713	0.96	0.97
100000286792	36	273	1017	-0.0663	0.0654	1.02	1.02
100000272298	37	273	1017	0.7941	0.0735	1.06	1.08

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

Item CID	Item Seq. No.[1]	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000286786	38	251	1017	-1.5026	0.0749	1.01	1.05
100000272300	39	273	1017	1.5778	0.0908	1.04	1.10
100000322079	40	251	1017	-0.0321	0.0655	1.08	1.08
100000322078	41	251	1017	-0.6680	0.0658	0.99	1.01
100000197367	42	257	1017	0.4466	0.0690	1.01	1.02
100000286859	43	273	1017	0.4703	0.0692	0.95	0.95
100000272305	44	273	1017	0.6022	0.0708	0.96	0.96
100000286860	45	257	1017	0.7834	0.0733	0.99	1.01
100000322116	46	257	1017	-0.2568	0.0650	1.09	1.09
100000322117	47	257	1017	-1.9559	0.0846	0.93	0.82
100000286845	48	273	1017	0.0540	0.0659	0.99	0.99
100000272301	49	273	1017	1.5290	0.0894	0.99	1.08
100000286846	50	246	1017	0.6274	0.0711	1.03	1.08
100000322080	51	246	1017	-0.1259	0.0652	1.00	1.00
100000272353	52	262	1017	-0.2062	0.0651	1.01	1.01
100000167565	53	273	1017	-0.1516	0.0652	1.02	1.02
100000167566	54	273	1017	0.6133	0.0709	1.04	1.05
100000174558	55	262	1017	0.7327	0.0726	0.98	0.99
100000322204	56	268	1017	0.4134	0.0686	1.02	1.03
100000322206	57	268	1017	-1.8315	0.0816	0.95	0.89
100000322199	58	268	1017	0.1106	0.0662	1.07	1.08
100000099639	59	268	1017	0.0583	0.0660	0.95	0.95
100000272308	60	246	1017	0.0108	0.0657	1.00	1.04
100000197322	61	246	1017	-0.2045	0.0651	1.01	1.00
100000322082	62	246	1017	0.0627	0.0660	0.99	1.01
100000322115	63	257	1017	-1.4201	0.0735	0.95	0.92
100000197365	64	257	1017	-0.0053	0.0657	0.97	0.96
100000272339	65	241	1017	-0.5396	0.0653	0.92	0.91
100000197377	66	273	1017	-0.9339	0.0675	1.04	1.05
100000197373	67	273	1017	0.0276	0.0658	1.02	1.02
100000197375	68	257	1017	1.2281	0.0818	0.92	0.96
100000197376	69	273	1017	-1.1337	0.0695	1.06	1.04
100000322201	70	268	1017	-2.5038	0.1021	0.97	0.89
100000322203	71	268	1017	0.7198	0.0724	1.00	1.01
100000322193	72	241	1017	-0.2147	0.0651	1.04	1.04
100000099553	73	241	1017	-0.9660	0.0678	0.88	0.84
100000322185	74	241	1017	-0.4084	0.0650	0.94	0.93
100000322195	75	241	1017	-0.0150	0.0656	0.97	0.96

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

Item CID	Item Seq. No.[1]	Item Strand No	n- Count	Rasch Difficulty	*SE*	MS Infit	MS Outfit
100000373853	76	273	1017	-0.2869	0.0650	0.92	0.92
100000167561	77	273	1017	-0.6825	0.0658	0.94	0.92
100000373854	78	241	1017	-0.2434	0.0650	0.90	0.88
100000322188	79	241	1017	-0.6293	0.0656	0.88	0.86
100000197319	80	246	1017	0.2761	0.0674	0.99	0.99
100000322196	81	241	1017	-0.9562	0.0677	1.00	0.99
100000322191	82	241	1017	-0.3621	0.0650	0.99	1.00
100000322189	83	241	1017	-1.5423	0.0756	0.90	0.81
100000099647	84	241	1017	-0.2686	0.0650	0.97	0.96
100000286880	85	273	1017	0.2571	0.0672	1.03	1.03
100000272303	86	273	1017	-0.5737	0.0654	1.02	1.02
100000286879	87	241	1017	0.1150	0.0663	1.03	1.04
100000286793	88	273	1017	0.1281	0.0663	1.04	1.03
100000286797	89	257	1017	0.1722	0.0666	0.97	0.96
100000272293	90	273	1017	-1.4147	0.0734	0.97	0.94
100000272294	91	273	1017	-0.9062	0.0673	0.97	0.95
100000322192	92	241	1017	-0.5694	0.0654	0.93	0.91
100000322190	93	241	1017	0.6732	0.0717	1.00	1.02
100000322202	94	268	1017	-1.6002	0.0766	0.95	0.89
100000272367	95	268	1017	-1.7854	0.0805	0.96	0.91
100000099631	96	268	1017	0.5341	0.0700	1.07	1.09
100000322197	97	241	1017	0.0757	0.0660	0.93	0.92
100000322194	98	241	1017	-0.8881	0.0671	0.99	1.00

Note: 1. 241= Algebra, 246=Geometry, 251=Measurement, 257=Statistics, 262= Probability, 268=Numbers and Computation, and 373&374= Process
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

σ_x^2 = the variance of the observed score in a population of persons,

σ_t^2 = the variance of their true score variance, and

σ_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: Mathematics: Grades 3 through 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: Mathematics 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 error 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, 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 was less than decision accuracy.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.90	0.05	0.05	0.86
BP : A	0.94	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.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.89	0.07	0.04	0.84
BP : A	0.95	0.04	0.01	0.93

Note. B: PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.88	0.06	0.06	0.83
BP : A	0.96	0.03	0.01	0.94

Note. B: PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.91	0.05	0.04	0.88
BP : A	0.96	0.02	0.01	0.95

Note. B: PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.86	0.09	0.05	0.81
BP : A	0.96	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.

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

Performance Cut	Accuracy	False Positive	False Negative	Consistency
B : PA	0.92	0.06	0.03	0.88
BP : A	0.96	0.03	0.01	0.94

Note. B: PA denotes the cut between Basic and Proficient, while BP:A denotes the cut between Proficient and Advanced.

8. TEST VALIDITY

8.1. Test Validity for the 2010 Mod-MSA: Mathematics

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: Mathematics, 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: Mathematics 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: Mathematics is ongoing and continuous. Content specialists, teachers from across Maryland, Pearson, and MSDE were greatly involved in developing and reviewing items. Committees such as content review, bias review, and vision review reviewed all of the items, which were finally stored in a Maryland item bank. Specifically, an internal review by MSDE and Pearson staff for content alignment and quality 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: Mathematics 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 (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; Mathematics 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	5	100000098439	+B	A
	11	100000197780	A	+B
	15	100000197753	+B	A
	24	100000197617	A	+B
	25	100000197603	A	+B
	31	100000197651	A	-B
	36	100000197779	A	+B
	56	100000197660	-B	A
	61	100000197649	-B	A
	62	100000197650	A	+B
	66	100000197665	A	-B
	70	100000197746	A	-B
	72	100000197750	A	-C
	90	100000197740	A	-B
	93	100000185385	+B	A
	97	100000185388	A	+B
4	8	100000186565	A	-B
	14	100000198098	A	-B
	43	100000186562	-B	A
	47	100000198129	A	+C
	82	100000186581	A	-B
	85	100000098588	A	+B
	91	100000098653	+B	-B

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

Grade	Item Sequence No.[1]	Item CID No.	DIF Classification[2]	
			Gender	White/African-American
5	27	100000196242	+C	A
	43	100000196053	-B	+C
	58	100000099178	A	-B
	79	100000187392	A	-B
	100	100000187375	A	-B
6	11	100000272073	A	-B
	23	100000272129	A	+B
	35	100000272127	A	-B
	38	100000273626	A	+B
	42	100000272099	+B	A
	45	100000198235	+B	A
	49	100000272116	+C	A
	62	100000198220	A	-C
	66	100000272163	+B	A
	72	100000272071	A	-B
	74	100000273837	A	-B
	84	100000272160	A	-B
	86	100000272157	A	-B
	89	100000272162	A	-B
7	6	100000272217	-B	A
	10	100000322136	+B	A
	11	100000322138	A	+B
	12	100000197496	A	+B
	16	100000099488	A	-B
	19	100000322152	A	-B
	39	100000374152	A	-B
	61	100000272187	A	+B
	71	100000099504	+B	A
	73	100000167818	A	-B
	89	100000322157	A	-B
	96	100000322162	-B	A
8	25	100000322208	A	-B
	33	100000272368	A	-B
	34	100000322200	+B	A
	35	100000099626	A	-B
	38	100000286786	-C	A
	47	100000322117	-C	A
	52	100000272353	+B	A
	57	100000322206	A	-B
	78	100000373854	-B	-B
	86	100000272303	-B	A

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 criterion

Evidence from Internal Structure of the Tests

As explained in Section 2.3, the 2010 Mod-MSA: Mathematics contains five mathematics reporting strands: Algebra, Geometry and Measurement, Statistics and Probability, Numbers and Computation, and Process. Even though these are individual strands are "locally independent," they measure the same underlying mathematics 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 mathematics strands for each of the three grades 3 through 8, respectively.

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

Strand (Subscale)	N	Mean	SD	A	GM	SP	NC	P
Algebra (A)	669	4.79	1.91	1.00				
Geometry and Measurement (GM)	669	5.64	2.20	0.50	1.00			
Statistics and Probability (SP)	669	4.62	2.05	0.42	0.42	1.00		
Numbers and Computation (NC)	669	5.89	2.45	0.52	0.57	0.49	1.00	
Process (P)	669	4.47	2.01	0.46	0.50	0.42	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 8.1.3. The 2010 Mod-MSA, Math Strand (Cluster) Correlations: Grade 4

Strand (Subscale)	N	Mean	SD	A	GM	SP	NC	P
Algebra (A)	872	5.06	2.05	1.00				
Geometry and Measurement (GM)	872	5.07	2.04	0.48	1.00			
Statistics and Probability (SP)	872	5.62	2.34	0.51	0.47	1.00		
Numbers and Computation (NC)	872	5.45	2.19	0.55	0.47	0.54	1.00	
Process (P)	872	5.02	2.04	0.46	0.46	0.47	0.47	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, Math Strand (Cluster) Correlations: Grade 5

Strand (Subscale)	N	Mean	SD	A	GM	SP	NC	P
Algebra (A)	926	5.31	2.10	1.00				
Geometry and Measurement (GM)	926	4.56	1.87	0.38	1.00			
Statistics and Probability (SP)	926	4.22	1.91	0.51	0.40	1.00		
Numbers and Computation (NC)	926	4.47	1.95	0.46	0.36	0.45	1.00	
Process (P)	926	5.62	2.21	0.41	0.32	0.47	0.36	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, Math Strand (Cluster) Correlations: Grade 6

	2009								2010							
Strand (Subscale)	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**
Algebra (A)	1344	5.19	2.08	1.00					852	5.23	2.11	1.00				
Geometry and Measurement (GM)	1344	4.03	1.97	0.44	1.00				852	3.97	2.07	0.41	1.00			
Statistics and Probability (SP)	1344	5.23	1.88	0.43	0.38	1.00			852	4.77	2.09	0.49	0.43	1.00		
Numbers and Computation (NC)	1344	4.29	1.90	0.45	0.41	0.41	1.00		852	4.85	2.12	0.48	0.40	0.48	1.00	
Process (P)	1344	5.06	2.17	0.45	0.45	0.47	0.48	1.00	852	5.00	2.22	0.48	0.40	0.50	0.46	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, Math Strand (Cluster) Correlations: Grade 7

	2009								2010							
Strand (Subscale)	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**
Algebra (A)	1575	4.64	1.79	1.00					1000	5.09	2.13	1.00				
Geometry and Measurement (GM)	1575	3.43	1.68	0.31	1.00				1000	3.94	1.74	0.39	1.00			
Statistics and Probability (SP)	1575	5.00	2.03	0.36	0.33	1.00			1000	4.25	1.90	0.35	0.37	1.00		
Numbers and Computation (NC)	1575	5.30	1.88	0.37	0.28	0.41	1.00		1000	5.58	2.09	0.50	0.40	0.34	1.00	
Process (P)	1575	5.54	2.18	0.33	0.28	0.49	0.34	1.00	1000	4.87	1.85	0.28	0.25	0.34	0.31	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, Math Strand (Cluster) Correlations: Grade 8

	2009								2010							
Strand (Subscale)	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**	**N**	**Mean**	**SD**	**A**	**GM**	**SP**	**NC**	**P**
Algebra (A)	1846	5.40	2.01	1.00					1017	5.74	2.25	1.00				
Geometry and Measurement (GM)	1846	3.56	1.62	0.33	1.00				1017	3.52	1.60	0.33	1.00			
Statistics and Probability (SP)	1846	3.60	1.80	0.37	0.31	1.00			1017	4.22	1.99	0.43	0.30	1.00		
Numbers and Computation (NC)	1846	3.65	1.49	0.31	0.28	0.32	1.00		1017	4.22	1.57	0.39	0.31	0.35	1.00	
Process (P)	1846	6.26	2.18	0.39	0.25	0.32	0.29	1.00	1017	5.68	2.33	0.49	0.31	0.49	0.38	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: Mathematics

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

To check the unidimensionality of the 2010 Mod-MSA: Mathematics, 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: Mathematics. As shown in the table, the first factor extracted a much large amount of eigenvalues across all grades.

Table 8.2.1. Eigenvalues between the First and Second Components of the Mod-MSA: Mathematics

Grade	First Eigenvalue	Second Eigenvalue
3	9.77	2.25
4	9.19	2..09
5	7.40	1.82
6	8.77	2.46
7	6.98	2.77
8	7.41	2.16

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

9. SUMMARY OF OPERATIONAL TEST RESULTS FOR THE 2010 MOD-MSA: MATHEMATICS

This section presents both the raw score and scaled score summaries for the Mod-MSA: Mathematics 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: Mathematics. For grades 6-8, the 2009 results are also depicted in the above mentioned tables.

9.1 Classical Descriptive Test Statistics (Raw Scores)

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 (KR20)
- Standard error of measurement (SEM)

Table 9.1.1 Classical Descriptive Statistics for the 2010 Mod-MSA: Mathematics: 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	51	0	51	N/A	N/A	N/A	N/A	N/A	961	26.40	8.39	0.85	3.27
4	51	0	51	N/A	N/A	N/A	N/A	N/A	1294	26.85	8.25	0.84	3.30
5	51	0	51	N/A	N/A	N/A	N/A	N/A	1430	24.97	7.38	0.80	3.38
6	51	0	51	1345	24.09	7.51	0.81	3.25	1477	23.66	8.04	0.83	3.28
7	51	0	51	1564	23.90	6.63	0.76	3.26	1755	23.96	6.87	0.77	3.30
8	51	0	51	1844	22.46	6.28	0.73	3.29	1920	23.40	7.21	0.79	3.27

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

9.2 Scale Score Descriptive Statistics

Table 9.2.1 provides information about scale score descriptive statistics of the test by grade 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, Mathematics: Grades 3-8

Grade	Year	*N*	*M*	*SD*	*P10*	*Q1*	*Mdn*	*Q3*	*P90*	*IQR*
3	2010	961	51.45	12.40	37	43	49	59	69	16
4	2010	1294	50.83	11.83	38	42	49	57	67	15
5	2010	1430	51.30	12.08	37	44	50	59	69	15
6	2009	1345	50.02	12.00	36	41	48	57	66	16
	2010	1477	50.70	12.81	38	42	48	57	67	15
7	2009	1564	49.85	11.84	36	42	48	57	65	15
	2010	1755	50.66	12.10	37	42	49	57	68	15
8	2009	1844	50.05	12.00	36	42	49	57	66	15
	2010	1920	50.00	14.10	35	41	48	59	69	18

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 cutoff 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	961	62.33	23.62	14.05
4	1294	60.43	29.44	10.12
5	1430	62.31	27.62	10.07
6	1477	72.24	18.01	9.75
7	1755	62.34	31.74	5.93
8	1920	77.76	14.38	7.86

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 θ scores is known as the test characteristic curve (TCC). The TCC for each grade 3 through 8 are provided in Figure 9.4.1a, 9.4.2a, 9.4.3a, 9.4.4a, 9.4.5a, and 9.4.6a 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 9.4.1b, 9.4.2b, 9.4.3b, 9.4.4b, 9.4.5b, and 9.4.6b. 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 β_v is person's proficiency and δ_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 statistics are provided for each grade in Figures 9.4.1c, 9.4.2c, 9.4.3c, 9.4.4c, 9.4.5c, and 9.4.6c, respectively. For grades 6 to 8, a comparison is made across years in these figures.

Grade 3

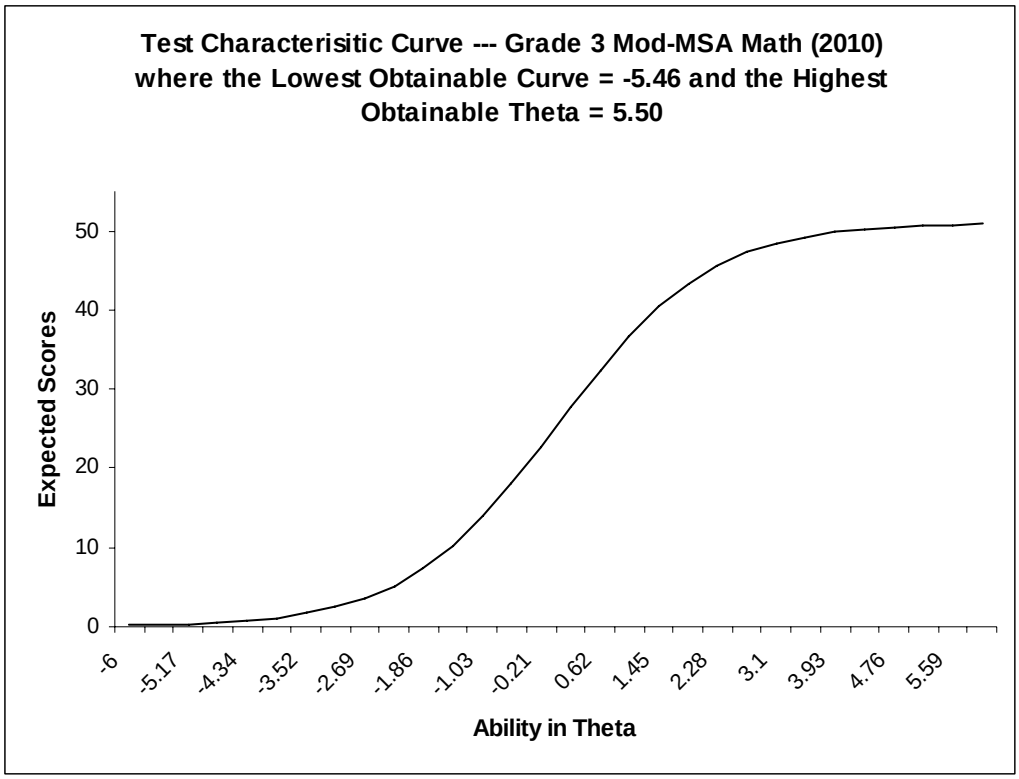


Figure 9.4.1a. TCC for Grade 3 Mod-MSA: Mathematics

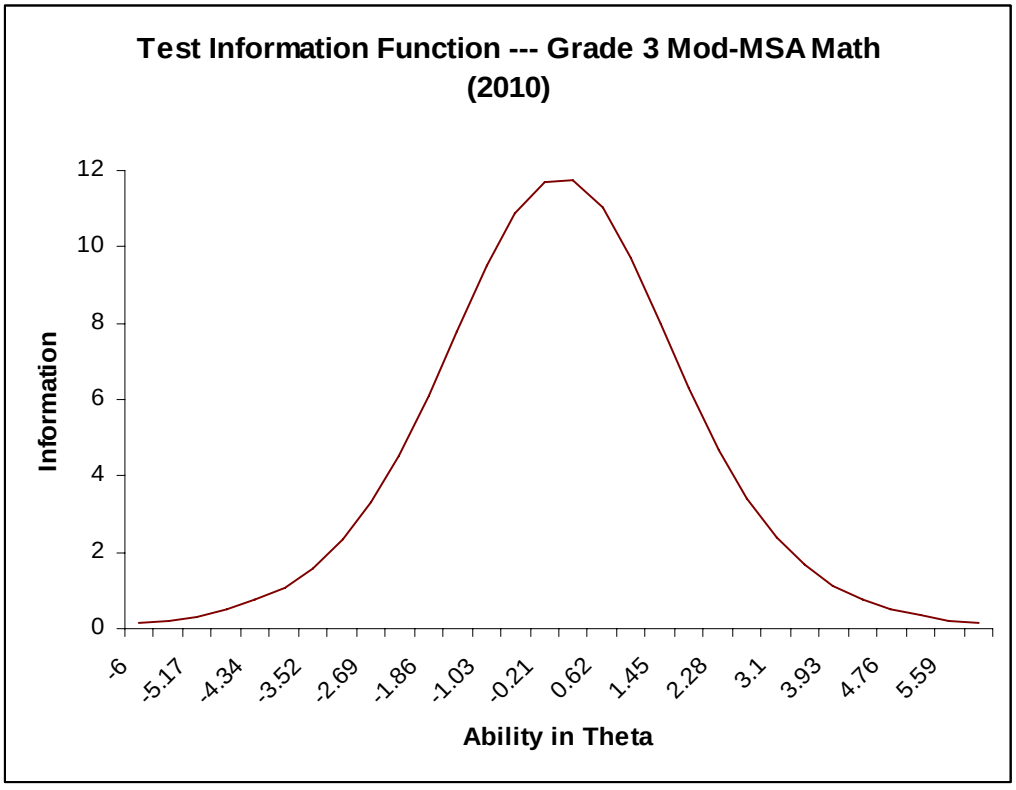


Figure 9.4.1b. TIF for Grade 3 Mod-MSA: Mathematics

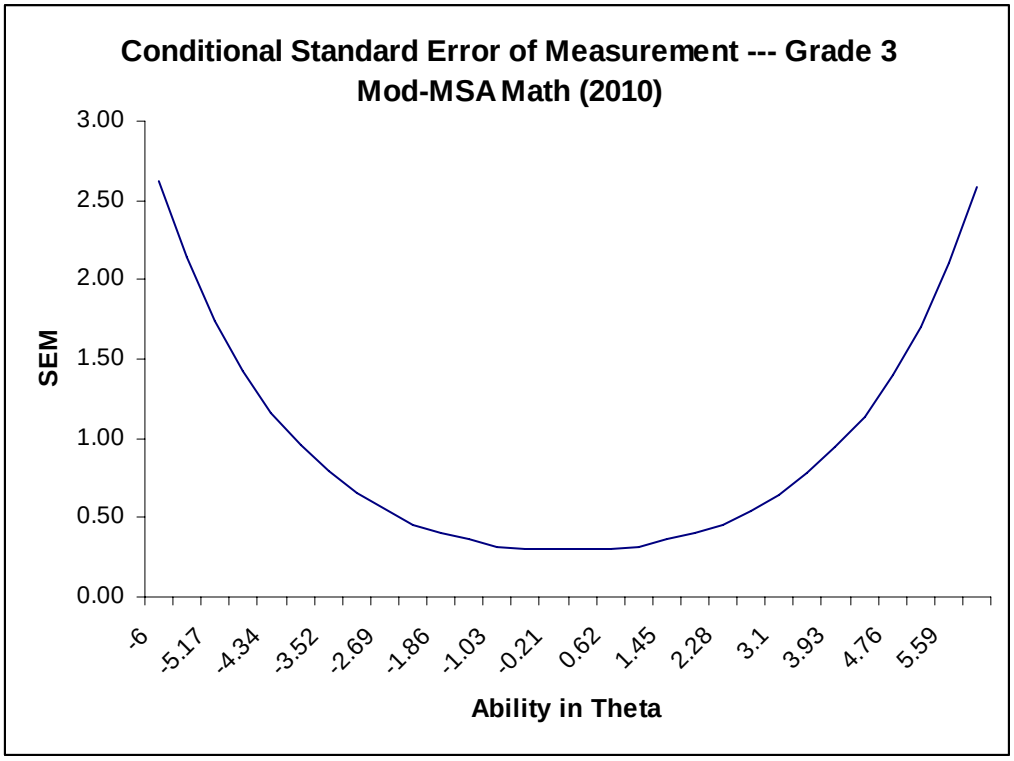


Figure 9.4.1c. CSEM for Grade 3 Mod-MSA: Mathematics

Grade 4

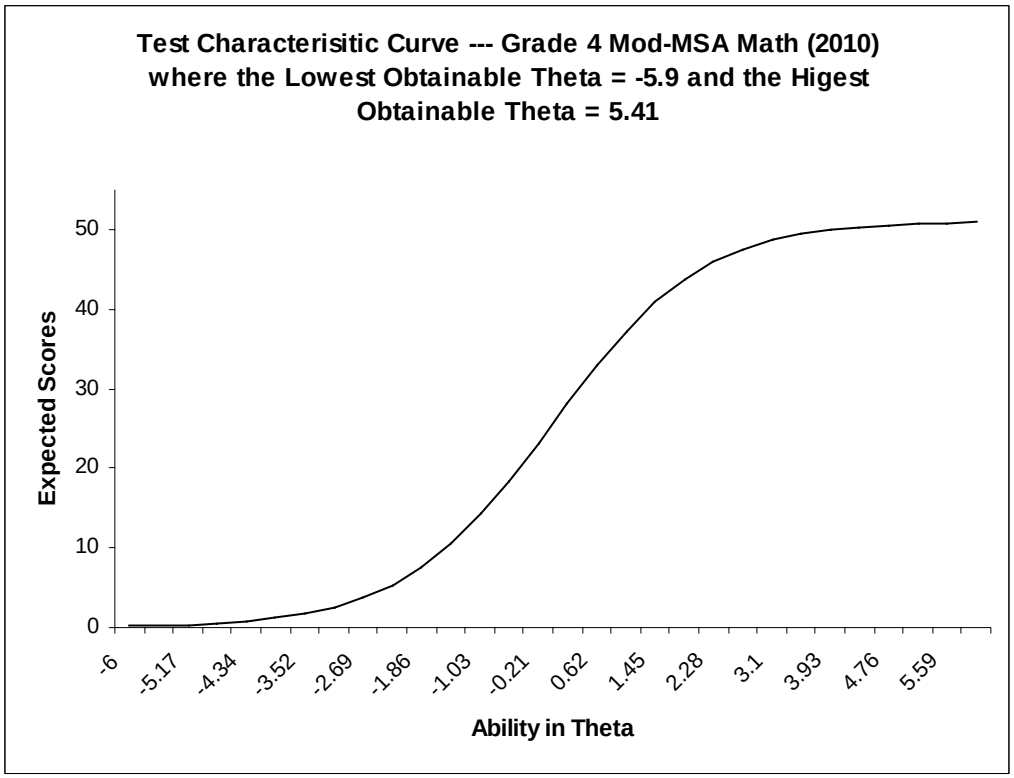


Figure 9.4.2a. TCC for Grade 4 Mod-MSA: Mathematics

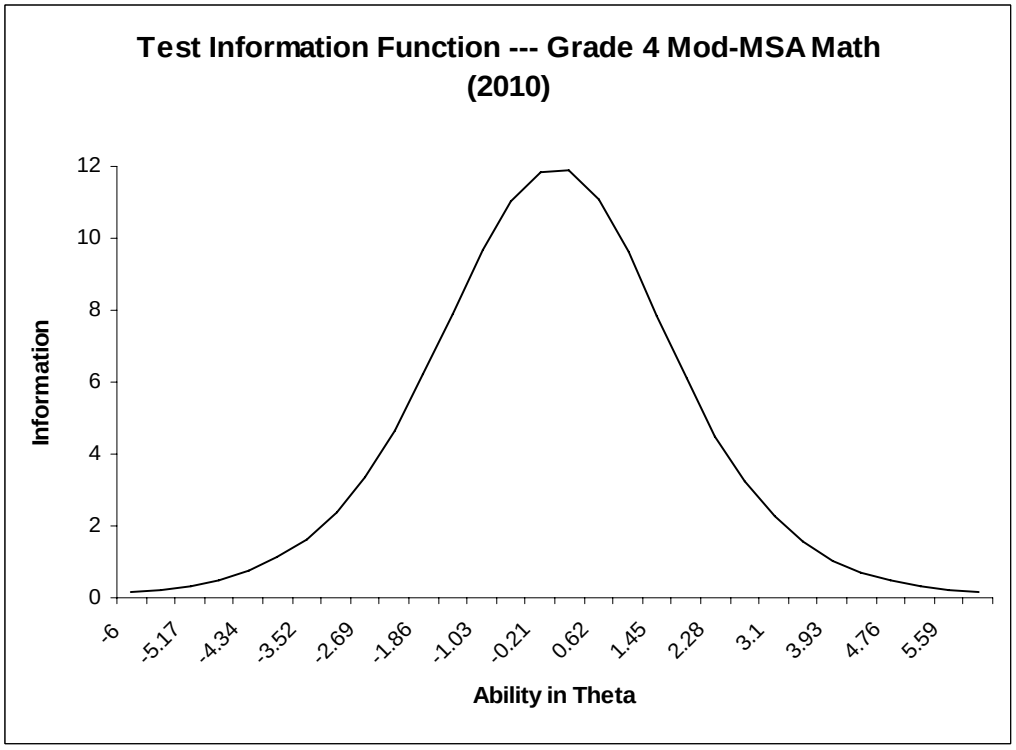


Figure 9.4.2b. TIF for Grade 4 Mod-MSA: Mathematics

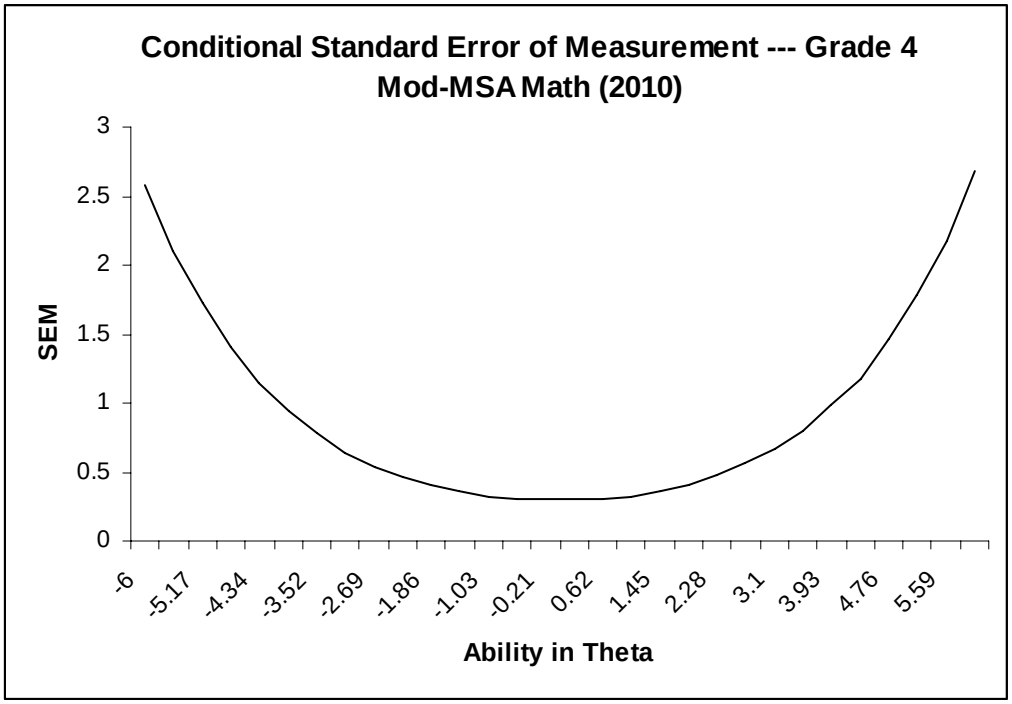


Figure 9.4.2c. CSEM for Grade 4 Mod-MSA: Mathematics

Grade 5

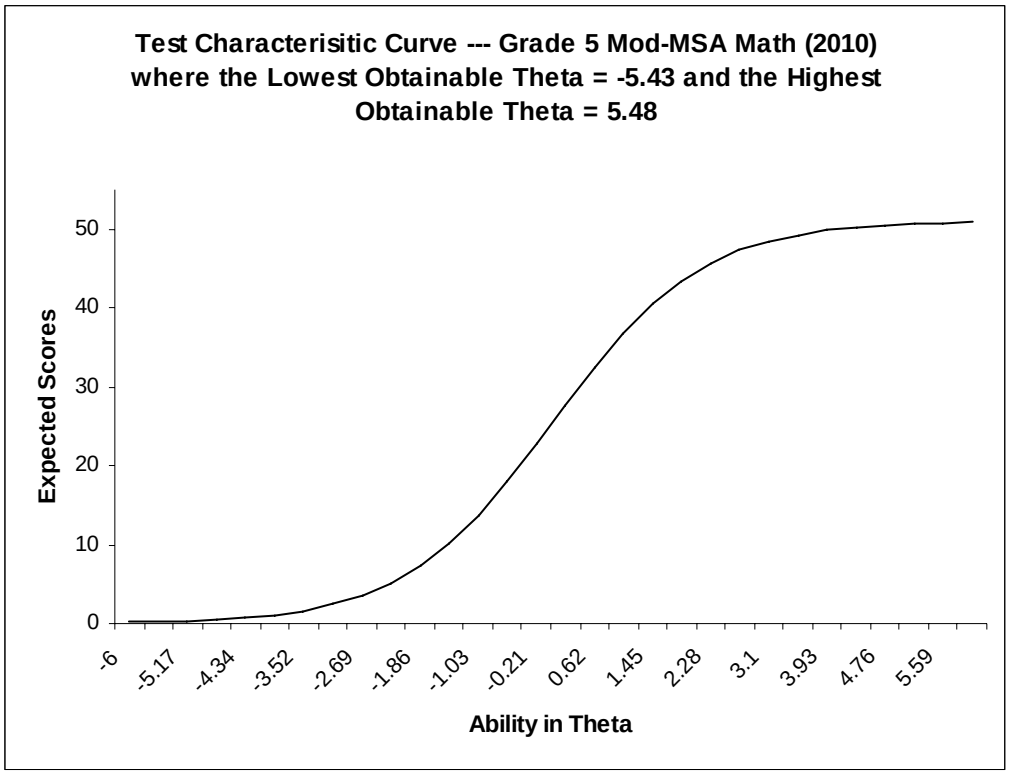


Figure 9.4.3a. TCC for Grade 5 Mod-MSA: Mathematics

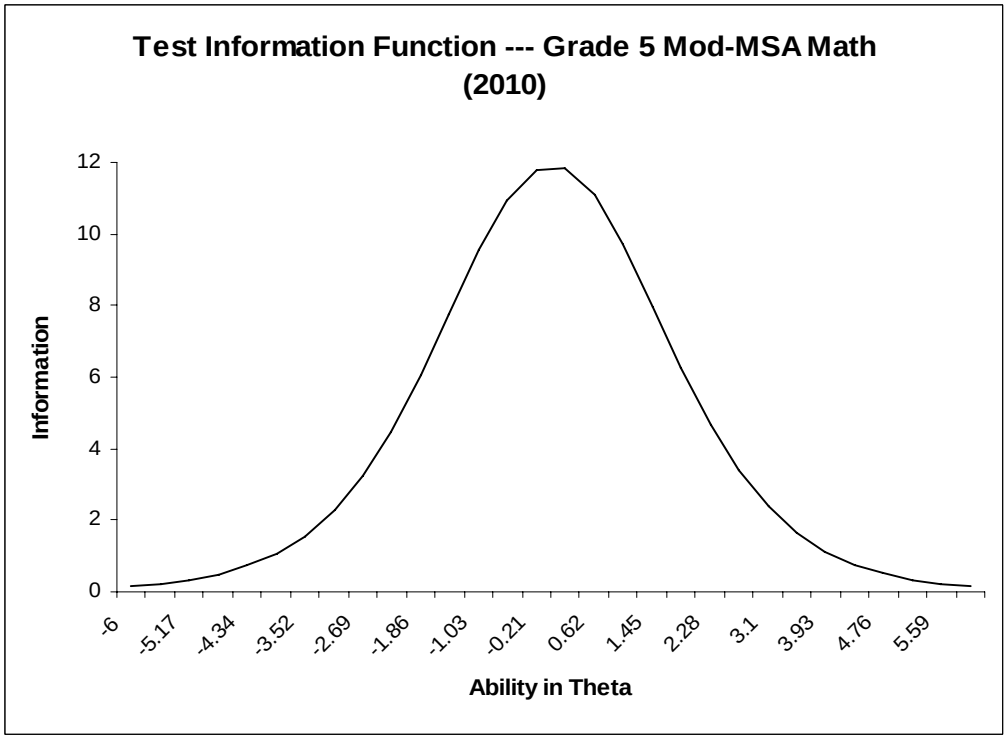


Figure 9.4.3b. TIF for Grade 5 Mod-MSA: Mathematics

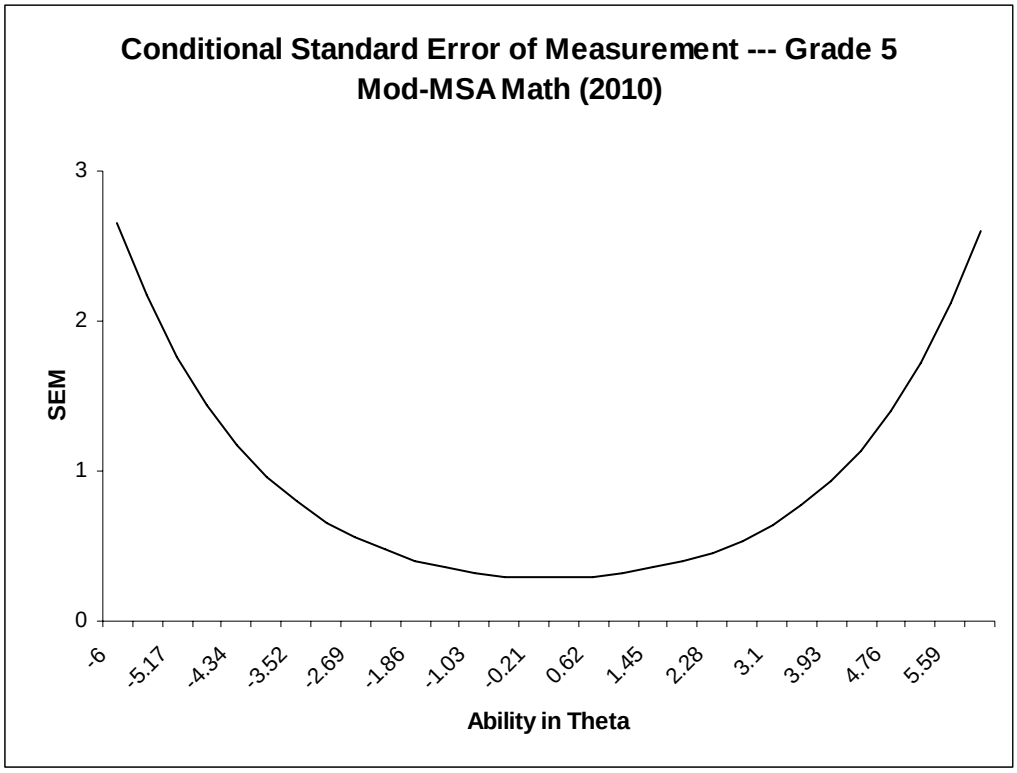


Figure 9.4.3c. CSEM for Grade 5 Mod-MSA: Mathematics

Grade 6

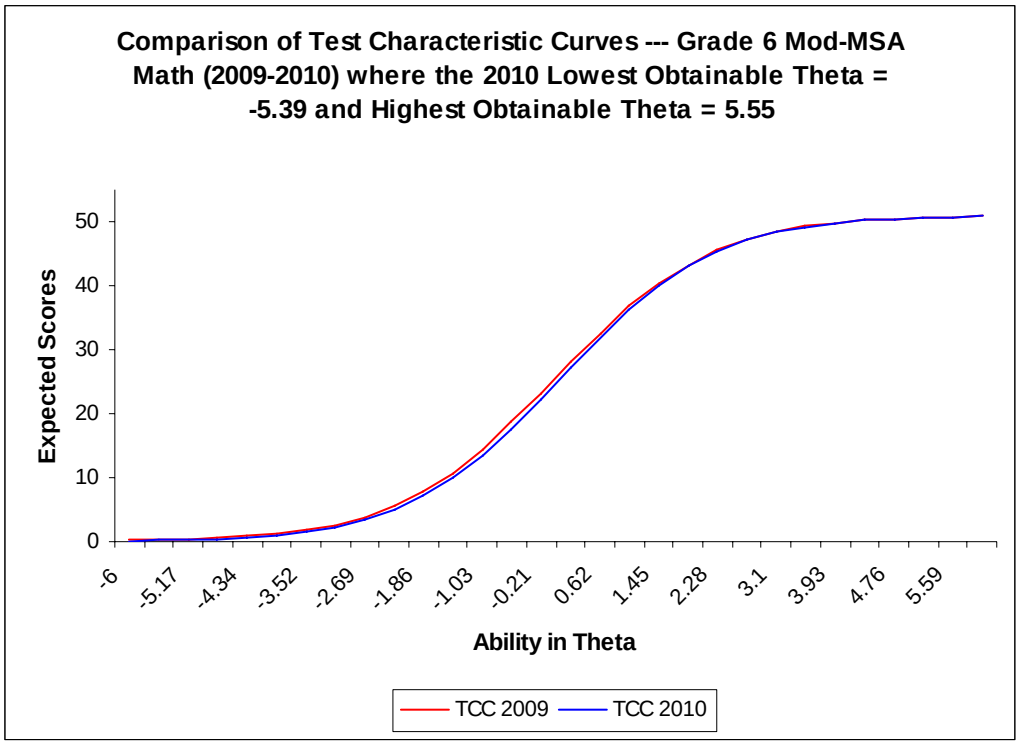


Figure 9.4.4a. TCC Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010)

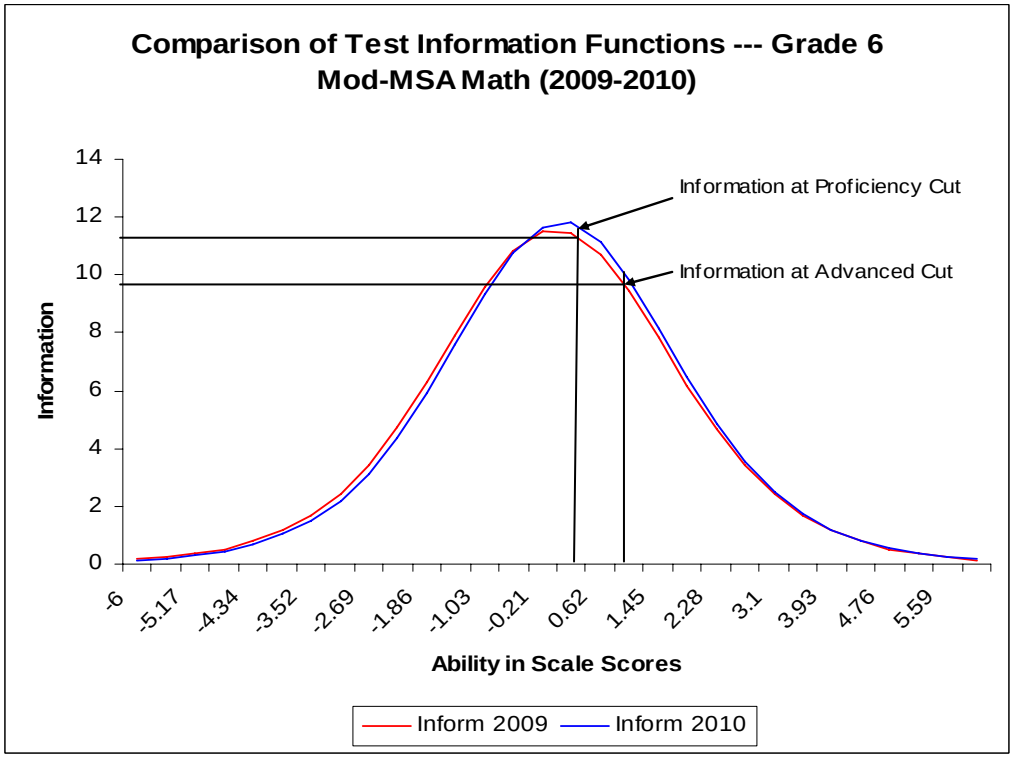


Figure 9.4.4b. TIF Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010)

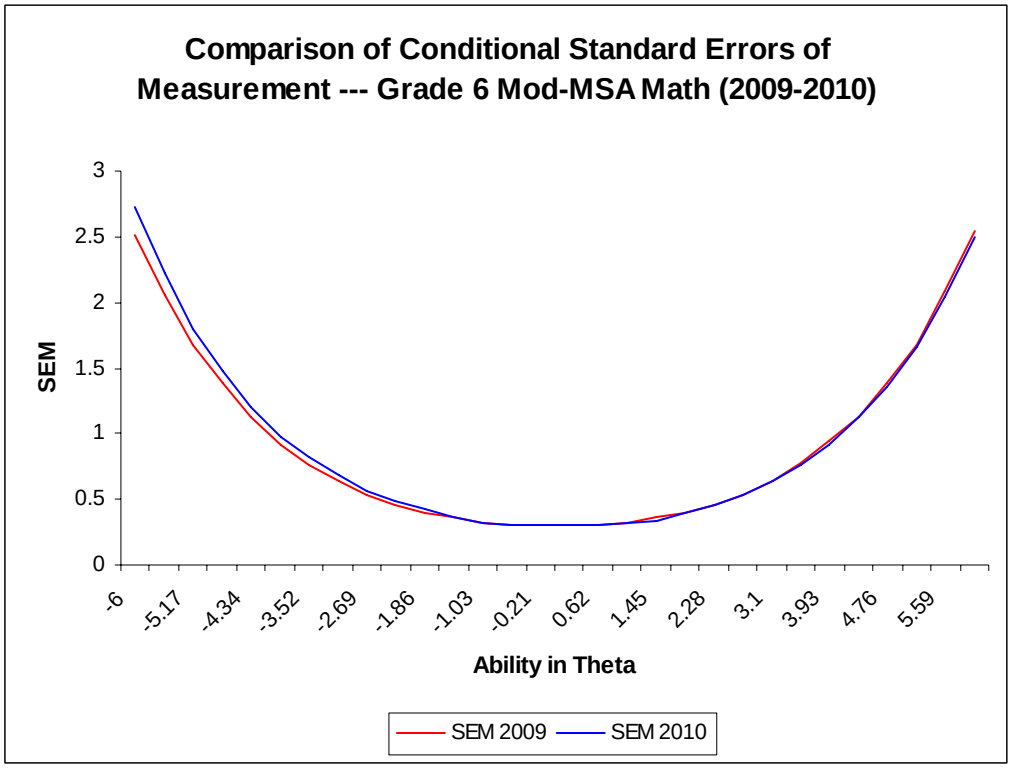


Figure 9.4.4c. CSEM Comparison for Grade 6 Mod-MSA: Mathematics (2009 with 2010)

Grade 7

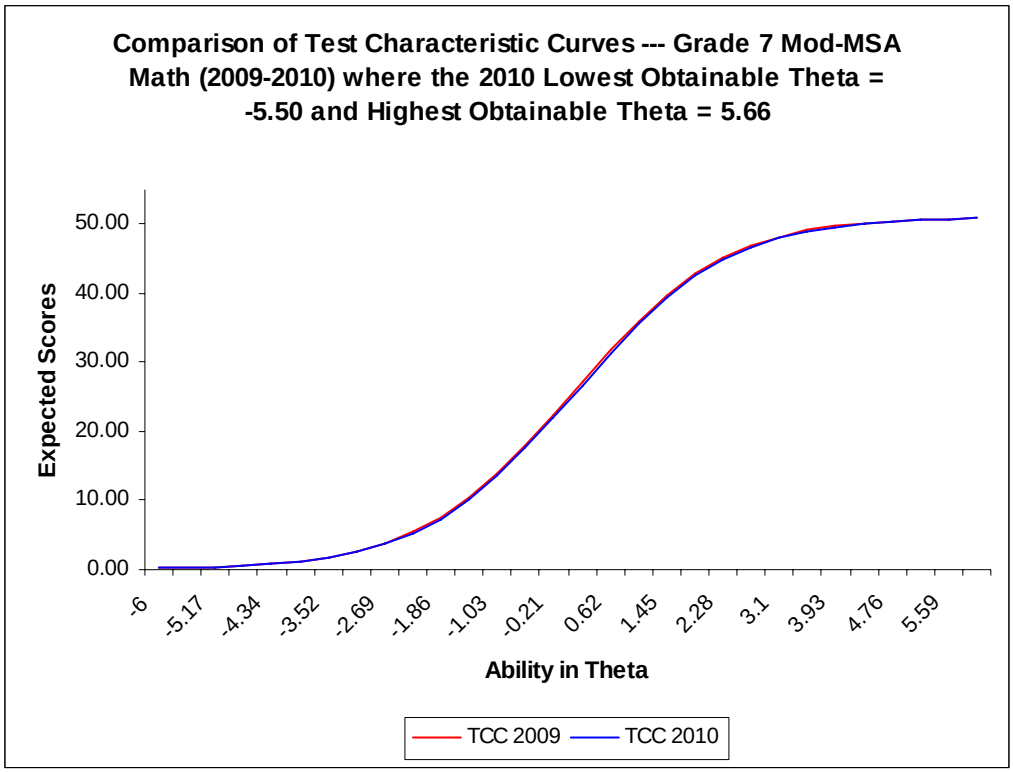


Figure 9.4.5a. TCC Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010)

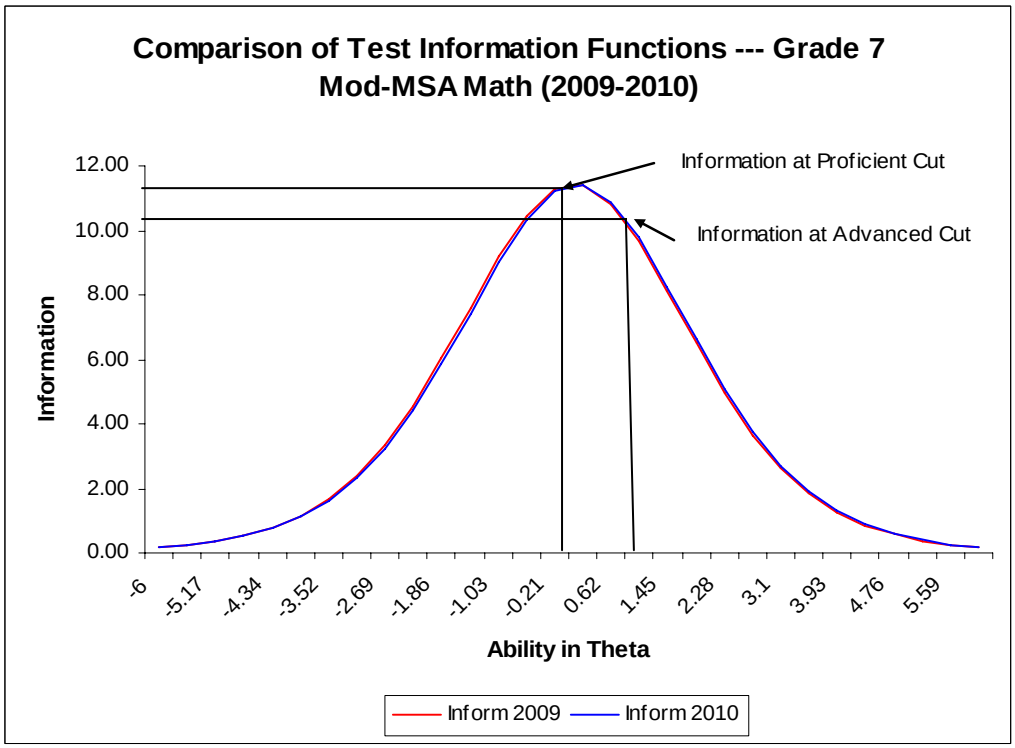


Figure 9.4.5b. TIF Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010)

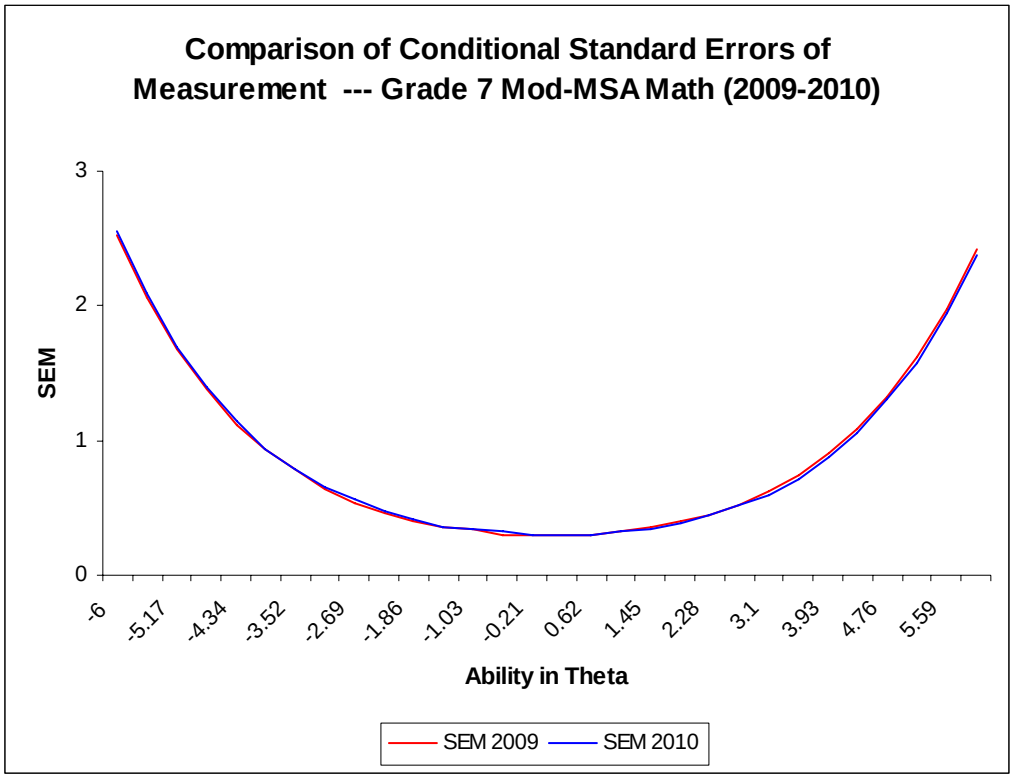


Figure 9.4.5c. CSEM Comparison for Grade 7 Mod-MSA: Mathematics (2009 with 2010)

Grade 8

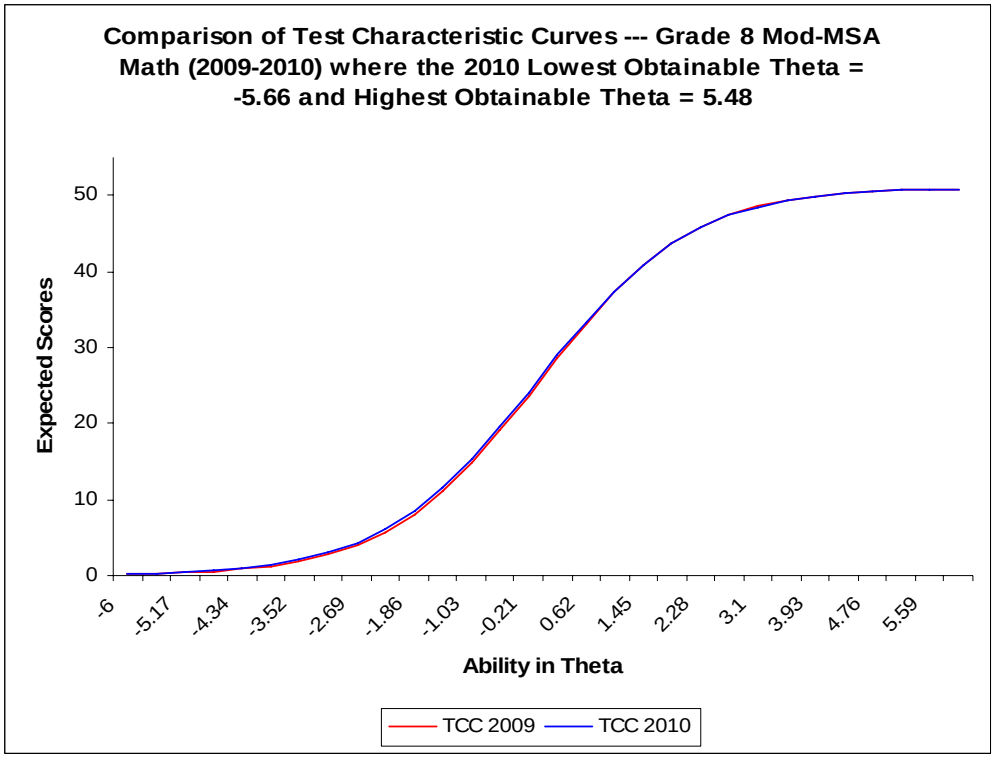


Figure 9.4.6a. TCC Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010)

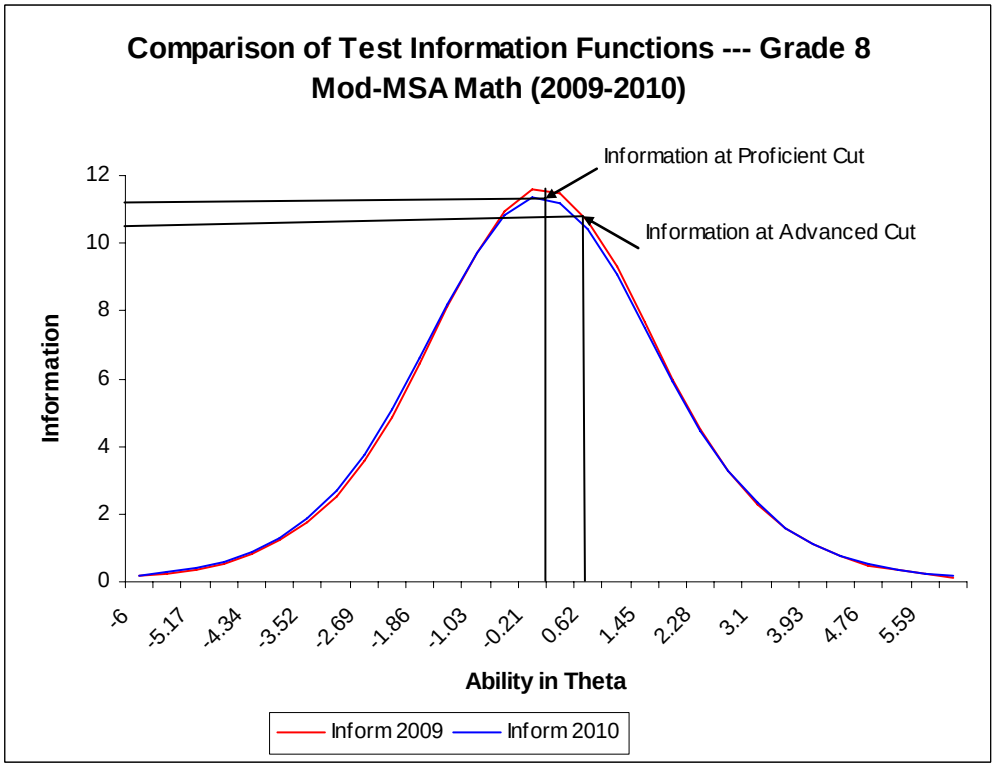


Figure 9.4.6b. TIF Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010)

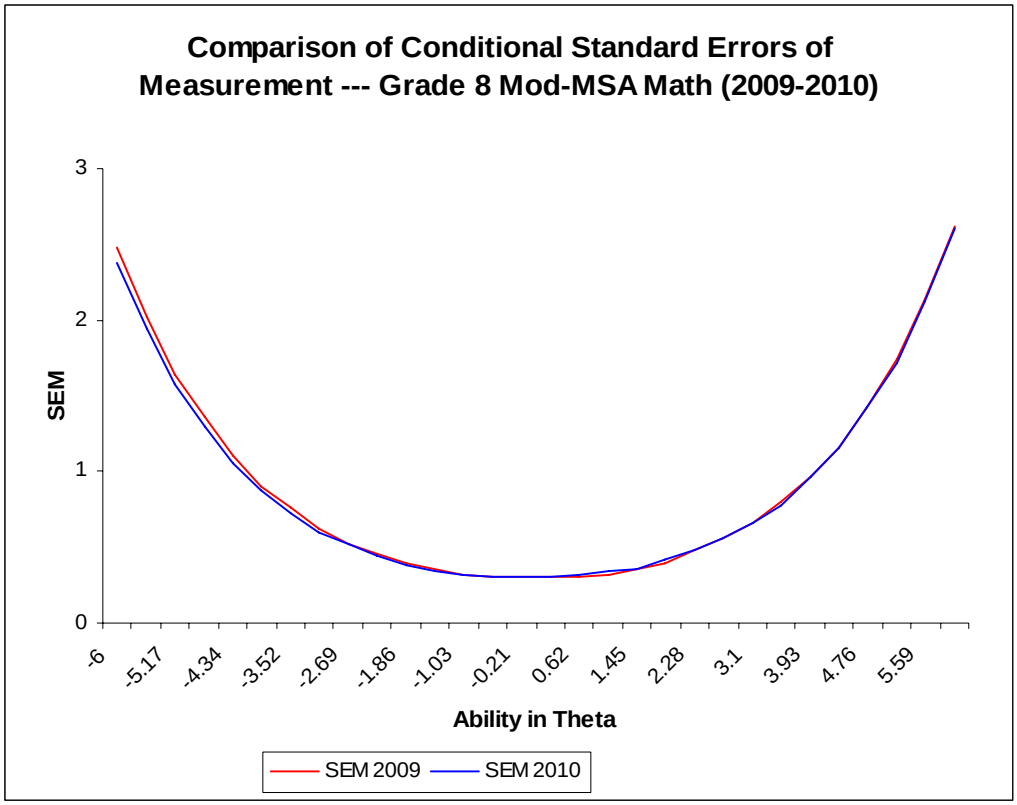


Figure 9.4.6c. CSEM Comparison for Grade 8 Mod-MSA: Mathematics (2009 with 2010)

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/SRGraw-Hill (2004, August). *The Maryland standard setting technical report*. (Technical Report). Monterey, CA: CTB/SRGraw-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 SRGEN Assessment*. Palo Alto, CA: Stanford University.

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.

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

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.

Kolen, M. J., and Brennan, R. L. (1995). Test equating methods and practices. New York: Springer-Verlag.

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.

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	100000107601	54	100000197602
3	2	100000197604	55	100000107621
3	3	100000098438	57	100000197756
3	4	100000098445	58	100000098446
3	6	100000098516	60	100000197648
3	7	100000185401	61	100000197649
3	8	100000185403	62	100000197650
3	9	100000197723	63	100000185382
3	11	100000197780	64	100000185313
3	12	100000197781	65	100000197662
3	14	100000197752	71	100000197724
3	15	100000197753	73	100000098527
3	16	100000197654	74	100000185381
3	18	100000197647	75	100000098435
3	19	100000098440	76	100000185485
3	26	100000197677	77	100000185486
3	27	100000197751	78	100000185380
3	28	100000197722	79	100000185384
3	29	100000350878	81	100000197720
3	37	100000197670	86	100000197664
3	38	100000197674	88	100000185387
3	39	100000197675	89	100000185376
3	41	100000098444	91	100000197661
3	46	100000098454	92	100000098515
3	48	100000185378	94	100000197663
3	52	100000098449		

Table A2. Grade 4 Operational Items

Grade	Sequence No.	UIN No.	Sequence No.	UIN No.
4	1	100000198111	64	100000098666
4	2	100000198113	65	100000198150
4	3	100000198114	67	100000186573
4	5	100000198123	68	100000186574
4	6	100000098578	69	100000186575
4	10	100000098583	71	100000198108
4	20	100000098579	72	100000198122
4	22	100000198140	73	100000098645
4	23	100000198142	74	100000186578
4	24	100000198143	75	100000186580
4	26	100000198128	76	100000198157
4	27	100000198092	77	100000098573
4	28	100000198099	78	100000186560
4	34	100000098587	79	100000196561
4	36	100000198107	80	100000198137
4	37	100000198101	81	100000098571
4	38	100000198102	83	100000207143
4	39	100000198103	84	100000198121
4	40	100000198148	89	100000198125
4	41	100000198158	92	100000186577
4	45	100000198094	96	100000186567
4	46	100000198096	97	100000186582
4	52	100000098586	98	100000198127
4	56	100000198139	99	100000098664
4	57	100000098585	100	100000098580
4	58	100000098582		

Table A3. Grade 5 Operational Items

Grade	Sequence No.	UIN No.	Sequence No.	UIN No.
5	2	100000187367	51	100000196281
5	5	100000099080	52	100000187387
5	6	100000099090	53	100000187389
5	8	100000196233	54	100000187388
5	10	100000196229	55	100000196270
5	11	100000196273	56	100000196256
5	12	100000196043	60	100000187380
5	13	100000196045	61	100000187429
5	14	100000196253	62	100000187428
5	16	100000196088	64	100000099180
5	21	100000187376	65	100000099072
5	24	100000196231	66	100000187360
5	25	100000196036	67	100000187361
5	26	100000196029	71	100000196223
5	28	100000099083	73	100000196094
5	29	100000099075	77	100000187370
5	30	100000099086	78	100000187391
5	33	100000196284	80	100000187382
5	34	100000196057	81	100000196263
5	35	100000196054	82	100000196247
5	36	100000196279	83	100000196238
5	37	100000099081	84	100000196079
5	38	100000099079	88	100000099177
5	42	100000196234	89	100000187366
5	44	100000196244	93	100000196042
5	46	100000099091		

Table A4. Grade 6 Operational Items

Grade	Sequence No.	UIN No.	Sequence No.	UIN No.
6	5	100000272117	58	100000272167
6	6	100000198185	59	100000187852
6	10	100000272076	60	100000272065
6	12	100000099232	61	100000198218
6	16	100000198162	62	100000198220
6	17	100000272101	63	100000198219
6	18	100000198232	64	100000187841
6	20	100000198210	65	100000272161
6	21	100000198211	66	100000272163
6	22	100000198212	67	100000198170
6	24	100000187836	68	100000272104
6	27	100000272159	69	100000272150
6	29	100000272077	70	100000272151
6	30	100000272068	73	100000099236
6	31	100000272075	77	100000198202
6	32	100000272072	79	100000272133
6	34	100000272131	82	100000272152
6	39	100000272143	84	100000272160
6	41	100000198178	85	100000187850
6	45	100000198235	90	100000099252
6	47	100000272082	91	100000198165
6	48	100000272114	92	100000272097
6	50	100000198186	93	100000187833
6	52	100000099325	95	100000198174
6	53	100000272128	96	100000272164
6	54	100000272125		

Table A5. Grade 7 Operational Items

Grade	Sequence No.	UIN No.	Sequence No.	UIN No.
7	5	100000197499	56	100000273416
7	6	100000272217	57	100000272274
7	9	100000197494	58	100000272186
7	12	100000197496	63	100000374196
7	13	100000322166	66	100000197501
7	16	100000099488	69	100000322163
7	25	100000322137	71	100000099504
7	29	100000272171	73	100000167818
7	30	100000099420	75	100000322147
7	31	100000099421	76	100000099493
7	32	100000272271	77	100000273417
7	33	100000272216	78	100000273444
7	34	100000272272	79	100000273445
7	35	100000322175	80	100000197524
7	36	100000322172	83	100000322150
7	37	100000197528	84	100000322144
7	38	100000374151	86	100000099410
7	39	100000374152	87	100000322161
7	40	100000196052	88	100000099413
7	41	100000196055	90	100000099404
7	42	100000322129	91	100000196255
7	43	100000197481	92	100000197522
7	48	100000272203	93	100000196257
7	49	100000322153	94	100000196261
7	52	100000322128	98	100000099490
7	55	100000273415		

Table A6. Grade 8 Operational Items

Grade	Sequence No.	UIN No.	Sequence No.	UIN No.
8	3	100000197359	53	100000167565
8	4	100000272279	54	100000167566
8	5	100000272278	55	100000174558
8	6	100000272340	59	100000099639
8	7	100000099548	61	100000197322
8	9	100000322186	63	100000322115
8	10	100000322122	64	100000197365
8	11	100000322120	66	100000197377
8	12	100000099550	67	100000197373
8	19	100000322118	68	100000197375
8	20	100000322084	69	100000197376
8	21	100000197347	73	100000099553
8	24	100000322209	75	100000322195
8	25	100000322208	76	100000373853
8	26	100000322205	77	100000167561
8	28	100000197387	78	100000373854
8	32	100000373799	80	100000197319
8	35	100000099626	82	100000322191
8	41	100000322078	84	100000099647
8	42	100000197367	85	100000286880
8	43	100000286859	86	100000272303
8	44	100000272305	87	100000286879
8	45	100000286860	93	100000322190
8	48	100000286845	94	100000322202
8	49	100000272301	96	100000099631
8	50	100000286846		

APPENDIX B: ITEM DISTRACTOR ANALYSES

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000197601	1	MC	669	0.77	0.42	12.26	10.16	77.28	0.30	-0.16	-0.24	
100000197604	2	MC	669	0.50	0.50	37.97	11.66	49.93	0.45	-0.37	0.02	
100000098438	3	MC	669	0.44	0.50	35.58	20.03	44.10	0.30	-0.03	-0.20	
100000098445	4	MC	669	0.59	0.49	16.44	58.59	24.96		-0.18		-0.26
100000098439	5	MC	669	0.55	0.50	21.52	55.01	23.47		-0.24		-0.19
100000098516	6	MC	669	0.58	0.49	57.55	17.04	25.41			-0.24	-0.27
100000185401	7	MC	669	0.35	0.48	13.45	35.43	50.82	0.30	-0.14		-0.13
100000185403	8	MC	669	0.38	0.49	20.93	37.97	40.96	0.15	-0.18		-0.01
100000197723	9	MC	669	0.54	0.50	11.06	53.96	34.83	0.15	-0.23		-0.10
100000197745	10	MC	669	0.57	0.50	29.30	13.60	56.80	0.30	-0.23	-0.25	
100000197780	11	MC	669	0.46	0.50	29.60	24.07	46.04	0.30	-0.02	-0.17	
100000197781	12	MC	669	0.53	0.50	23.32	52.62	23.77	0.30	-0.16		-0.21
100000098529	13	MC	669	0.38	0.49	34.53	38.12	26.61	0.75	-0.02		-0.12
100000197752	14	MC	669	0.41	0.49	44.99	13.15	41.41	0.45	-0.16	-0.20	
100000197753	15	MC	669	0.31	0.46	54.26	14.05	31.24	0.45	-0.14	-0.16	
100000197754	16	MC	669	0.27	0.45	27.20	30.34	41.85	0.60		-0.06	-0.02
100000197620	17	MC	669	0.70	0.46	17.49	11.96	70.25	0.30	-0.22	-0.05	
100000197647	18	MC	669	0.55	0.50	21.08	24.07	54.56	0.30	-0.15	-0.19	
100000098440	19	MC	669	0.68	0.47	17.34	14.65	67.71	0.30	-0.17	-0.15	
100000098448	20	MC	669	0.63	0.48	21.08	62.78	15.55	0.60	-0.23		-0.23
100000185375	21	MC	669	0.61	0.49	19.13	19.88	60.54	0.45	-0.25	-0.16	
100000185389	22	MC	669	0.45	0.50	23.17	31.09	45.44	0.30	-0.10	-0.09	
100000185391	23	MC	669	0.42	0.49	32.14	42.00	25.71	0.15	-0.14		-0.24
100000197617	24	MC	669	0.36	0.48	18.98	44.69	36.02	0.30	-0.16	0.01	
100000197603	25	MC	669	0.66	0.47	8.82	24.51	66.37	0.30	-0.23	-0.06	
100000197677	26	MC	669	0.41	0.49	36.62	21.97	41.26	0.15	0.02	-0.17	
100000197751	27	MC	669	0.26	0.44	26.46	40.81	32.44	0.30		-0.04	-0.15
100000197722	28	MC	669	0.42	0.49	34.53	22.72	42.45	0.30	-0.05	-0.18	
100000350878	29	MC	669	0.56	0.50	17.64	26.16	55.90	0.30	-0.21	-0.12	
100000098450	30	MC	669	0.57	0.50	57.25	16.59	26.01	0.15		-0.19	-0.20
100000197651	31	MC	669	0.74	0.44	17.04	8.67	73.99	0.30	-0.04	-0.13	
100000197655	32	MC	669	0.22	0.41	31.69	21.97	46.19	0.15	-0.13		0.04
100000197656	33	MC	669	0.71	0.45	13.30	15.70	70.85	0.15	-0.21	-0.19	
100000197741	34	MC	669	0.48	0.50	19.13	48.13	32.59	0.15	-0.24		-0.22
100000197755	35	MC	669	0.52	0.50	15.25	32.14	52.32	0.30	-0.05	-0.25	

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000197779	36	MC	669	0.62	0.49	20.93	61.73	17.19	0.15	-0.34		-0.16
100000197670	37	MC	669	0.54	0.50	53.81	19.73	26.31	0.15		-0.12	-0.23
100000197674	38	MC	669	0.52	0.50	52.32	22.72	24.66	0.30		-0.17	-0.22
100000197675	39	MC	669	0.46	0.50	24.07	45.59	30.19	0.15	-0.16		-0.18
100000098452	40	MC	669	0.70	0.46	13.15	70.25	16.44	0.15	-0.24		-0.28
100000098444	41	MC	669	0.36	0.48	35.72	39.16	24.81	0.30		-0.11	-0.10
100000197667	42	MC	669	0.37	0.48	37.22	23.47	39.16	0.15		-0.17	-0.09
100000197668	43	MC	669	0.46	0.50	29.15	45.89	24.81	0.15	-0.08		-0.23
100000197669	44	MC	669	0.33	0.47	33.48	27.50	38.71	0.30		-0.07	-0.16
100000098451	45	MC	669	0.62	0.48	62.48	12.26	24.96	0.30		-0.28	-0.27
100000098454	46	MC	669	0.69	0.46	17.64	68.61	13.45	0.30	-0.15		-0.26
100000098447	47	MC	669	0.29	0.45	46.19	29.00	24.36	0.45	0.03		-0.15
100000185378	48	MC	669	0.69	0.46	68.91	21.52	9.27	0.30		-0.23	-0.25
100000197757	49	MC	669	0.57	0.50	57.25	16.44	26.01	0.30		-0.20	-0.34
100000197761	50	MC	669	0.59	0.49	15.55	59.04	25.11	0.30	-0.19		-0.20
100000197765	51	MC	669	0.52	0.50	27.06	20.78	51.72	0.45	-0.05	-0.10	
100000098449	52	MC	669	0.63	0.48	62.78	9.42	27.35	0.45		-0.21	-0.30
100000098453	53	MC	669	0.39	0.49	25.56	38.71	35.28	0.45	-0.02		-0.14
100000197602	54	MC	669	0.52	0.50	51.72	20.33	27.35	0.60		0.01	-0.16
100000197621	55	MC	669	0.61	0.49	16.89	61.14	21.38	0.60	-0.12		-0.24
100000197660	56	MC	669	0.71	0.45	70.85	11.06	17.64	0.45		-0.23	-0.22
100000197756	57	MC	669	0.64	0.48	16.59	64.28	18.68	0.45	-0.20		-0.17
100000098446	58	MC	669	0.33	0.47	47.09	33.48	18.98	0.45	-0.04		-0.20
100000098455	59	MC	669	0.36	0.48	35.28	35.87	28.40	0.45	0.01		-0.09
100000197648	60	MC	669	0.28	0.45	23.32	47.98	28.25	0.45	-0.17	0.08	
100000197649	61	MC	669	0.66	0.47	66.22	18.68	14.65	0.45		-0.22	-0.19
100000197650	62	MC	669	0.39	0.49	44.39	39.31	15.70	0.60	0.05		-0.19
100000185382	63	MC	669	0.42	0.49	28.70	29.00	42.00	0.30	-0.09	-0.14	
100000185313	64	MC	669	0.75	0.43	12.86	11.81	75.04	0.30	-0.28	-0.19	
100000197662	65	MC	669	0.49	0.50	30.34	48.88	20.48	0.30	-0.28		-0.20
100000197665	66	MC	669	0.60	0.49	23.32	60.24	16.14	0.30	-0.23		-0.17
100000098522	67	MC	669	0.70	0.46	13.00	17.04	69.66	0.30	-0.30	-0.22	
100000185484	68	MC	669	0.40	0.49	39.91	31.99	27.80	0.30		-0.10	-0.13
100000185404	69	MC	669	0.43	0.50	30.49	42.90	26.31	0.30	-0.03		-0.28
100000197746	70	MC	669	0.31	0.46	36.32	30.94	32.44	0.30	0.07		-0.11

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000197724	71	MC	669	0.50	0.50	49.78	19.13	30.64	0.45		-0.21	-0.24
100000197750	72	MC	669	0.38	0.49	49.93	37.97	11.66	0.45	-0.17		0.03
100000098527	73	MC	669	0.31	0.46	48.43	20.33	30.79	0.45	-0.07	-0.11	
100000185381	74	MC	669	0.62	0.49	61.73	20.18	17.49	0.60		-0.20	-0.25
100000098435	75	MC	669	0.52	0.50	17.79	29.30	52.47	0.45	-0.28	-0.12	
100000185485	76	MC	669	0.49	0.50	48.58	28.40	22.57	0.45		-0.13	-0.32
100000185486	77	MC	669	0.63	0.48	25.11	62.78	11.66	0.45	-0.31		-0.18
100000185380	78	MC	669	0.48	0.50	17.34	47.98	34.23	0.45	-0.16		-0.15
100000185384	79	MC	669	0.63	0.48	24.22	62.93	12.41	0.45	-0.24		-0.22
100000185377	80	MC	669	0.53	0.50	53.06	23.47	22.87	0.60		-0.11	-0.29
100000197720	81	MC	669	0.44	0.50	39.01	43.95	16.44	0.60	-0.23		-0.09
100000185386	82	MC	669	0.39	0.49	47.38	39.16	13.00	0.45	-0.09		-0.20
100000185477	83	MC	669	0.67	0.47	18.83	13.45	67.26	0.45	-0.18	-0.21	
100000185473	84	MC	669	0.33	0.47	24.66	33.03	41.70	0.60	-0.06		-0.25
100000197676	85	MC	669	0.36	0.48	36.47	35.28	27.65	0.60		-0.03	-0.09
100000197664	86	MC	669	0.39	0.49	13.45	38.71	47.23	0.60	-0.17		-0.15
100000098532	87	MC	669	0.30	0.46	30.34	17.94	51.27	0.45		0.02	-0.10
100000185387	88	MC	669	0.42	0.49	42.45	16.89	40.21	0.45		-0.17	-0.25
100000185376	89	MC	669	0.46	0.50	46.19	32.88	20.48	0.45		0.01	-0.27
100000197740	90	MC	669	0.62	0.49	11.66	25.86	61.88	0.60	-0.24	-0.24	
100000197661	91	MC	669	0.57	0.50	57.10	31.69	10.76	0.45		-0.17	-0.18
100000098515	92	MC	669	0.53	0.50	16.89	53.06	29.60	0.45	-0.33		-0.24
100000185385	93	MC	669	0.26	0.44	44.99	26.31	28.25	0.45	-0.03		-0.09
100000197663	94	MC	669	0.43	0.49	42.60	24.96	31.99	0.45		-0.12	-0.11
100000197666	95	MC	669	0.56	0.50	24.51	55.61	19.28	0.60	-0.09		-0.28
100000098441	96	MC	669	0.52	0.50	15.99	31.54	52.02	0.45	-0.11	0.06	
100000185388	97	MC	669	0.57	0.50	18.83	57.10	23.47	0.60	-0.16		-0.19
100000185390	98	MC	669	0.61	0.49	60.84	16.44	22.27	0.45		-0.18	-0.31
100000185379	99	MC	669	0.61	0.49	60.84	26.16	12.41	0.60		-0.23	-0.10
100000185383	100	MC	669	0.68	0.47	19.28	68.01	12.26	0.45	-0.32		-0.29

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 for 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: Mathematics Grade 4, Distractor Distribution and Distractor-to- Total Correlation Analysis

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000198111	1	MC	872	0.52	0.50	31.65	51.95	16.06	0.34	-0.29		-0.12
100000198113	2	MC	872	0.57	0.50	56.65	16.17	27.06	0.11		-0.18	-0.19
100000198114	3	MC	872	0.56	0.50	34.98	56.19	8.72	0.11	-0.02		-0.26
100000198138	4	MC	872	0.34	0.47	13.99	33.94	51.72	0.34	-0.09		-0.21
100000198123	5	MC	872	0.57	0.50	33.94	8.83	57.00	0.23	-0.25	-0.08	
100000098578	6	MC	872	0.42	0.49	17.20	41.97	40.71	0.11	-0.19		-0.23
100000186564	7	MC	872	0.42	0.49	41.74	24.54	33.37	0.34		-0.13	-0.06
100000186565	8	MC	872	0.32	0.47	32.00	35.21	32.34	0.46		-0.10	-0.06
100000186566	9	MC	872	0.33	0.47	40.60	26.15	33.03	0.23	0.06	-0.07	
100000098583	10	MC	872	0.37	0.48	30.50	32.34	36.70	0.46	-0.09	-0.09	
100000098584	11	MC	872	0.78	0.42	6.77	77.52	15.14	0.57	-0.20		-0.29
100000198124	12	MC	872	0.57	0.50	57.22	23.39	18.81	0.57		-0.18	-0.15
100000198153	13	MC	872	0.39	0.49	43.35	39.11	17.09	0.46	0.03		-0.18
100000198098	14	MC	872	0.76	0.43	9.98	13.99	75.69	0.34	-0.23	-0.16	
100000198100	15	MC	872	0.49	0.50	48.51	25.57	25.57	0.34		-0.12	-0.08
100000198133	16	MC	872	0.39	0.49	12.96	39.11	47.59	0.34	-0.10		-0.16
100000198134	17	MC	872	0.49	0.50	32.57	48.62	18.58	0.23	-0.19		-0.13
100000198135	18	MC	872	0.36	0.48	35.09	28.56	36.01	0.34	0.11	-0.01	
100000198110	19	MC	872	0.76	0.43	10.89	13.30	75.57	0.23	-0.20	-0.20	
100000098579	20	MC	872	0.71	0.46	18.69	70.53	10.44	0.34	-0.18		-0.19
100000098581	21	MC	872	0.43	0.50	40.71	16.17	42.78	0.34	-0.21	-0.14	
100000198140	22	MC	872	0.64	0.48	19.15	16.51	64.11	0.23	-0.27	-0.14	
100000198142	23	MC	872	0.53	0.50	15.25	31.31	53.10	0.34	-0.13	-0.10	
100000198143	24	MC	872	0.51	0.50	22.82	51.38	25.57	0.23	-0.17		-0.09
100000198126	25	MC	872	0.31	0.46	38.99	30.05	30.62	0.34	0.07	-0.09	
100000198128	26	MC	872	0.46	0.50	21.56	32.34	45.87	0.23	-0.13	-0.06	
100000198092	27	MC	872	0.48	0.50	47.94	30.05	21.79	0.23		-0.13	0.00
100000198099	28	MC	872	0.58	0.49	19.61	22.36	57.68	0.34	-0.12	-0.03	
100000198117	29	MC	872	0.46	0.50	45.64	20.53	33.49	0.34		-0.21	-0.21
100000198118	30	MC	872	0.53	0.50	52.87	23.51	23.39	0.23		-0.20	-0.12
100000198119	31	MC	872	0.34	0.48	45.18	34.40	20.07	0.34	0.07		-0.21
100000098577	32	MC	872	0.61	0.49	61.47	19.95	18.35	0.23		-0.15	-0.23
100000098576	33	MC	872	0.36	0.48	41.51	36.01	22.13	0.34	0.08		-0.13
100000098587	34	MC	872	0.42	0.49	23.39	33.94	42.43	0.23	-0.09	-0.10	
100000198116	35	MC	872	0.20	0.40	14.68	64.68	20.41	0.23	-0.22	0.19	
100000198107	36	MC	872	0.53	0.50	34.40	53.44	11.93	0.23	-0.22		-0.15

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000198101	37	MC	872	0.39	0.49	38.65	19.61	41.51	0.23		-0.20	0.00
100000198102	38	MC	872	0.48	0.50	22.13	48.05	29.47	0.34	-0.10		-0.20
100000198103	39	MC	872	0.51	0.50	50.80	22.82	26.03	0.34		-0.17	-0.17
100000198148	40	MC	872	0.47	0.50	37.50	47.25	14.79	0.46	-0.13		-0.18
100000198158	41	MC	872	0.49	0.50	49.08	28.78	21.90	0.23		-0.15	-0.27
100000098646	42	MC	872	0.78	0.41	14.91	78.21	6.31	0.57	-0.27		-0.27
100000186562	43	MC	872	0.67	0.47	6.65	67.09	25.92	0.34	-0.21		-0.25
100000186563	44	MC	872	0.44	0.50	29.59	25.46	44.38	0.57	-0.08	-0.10	
100000198094	45	MC	872	0.78	0.41	7.22	78.21	13.99	0.57	-0.16		-0.27
100000198096	46	MC	872	0.58	0.49	30.85	57.91	10.78	0.46	-0.19		-0.20
100000198129	47	MC	872	0.53	0.50	52.87	31.19	15.48	0.46		-0.27	-0.16
100000198131	48	MC	872	0.34	0.48	34.40	37.96	27.18	0.46		-0.03	-0.16
100000198132	49	MC	872	0.40	0.49	23.97	39.79	35.78	0.46	-0.01		-0.19
100000198093	50	MC	872	0.90	0.30	3.67	90.25	5.73	0.34	-0.22		-0.27
100000098572	51	MC	872	0.40	0.49	39.91	17.32	42.43	0.34		-0.18	-0.22
100000098586	52	MC	872	0.48	0.50	24.31	48.39	26.95	0.34	-0.05		-0.29
100000198104	53	MC	872	0.55	0.50	15.71	54.93	29.01	0.34	-0.13		-0.01
100000198105	54	MC	872	0.54	0.50	23.62	22.25	53.78	0.34	-0.16	-0.13	
100000198106	55	MC	872	0.43	0.50	28.44	43.46	27.75	0.34	-0.12		-0.11
100000198139	56	MC	872	0.42	0.49	16.97	40.94	41.63	0.46	-0.19	-0.05	
100000098585	57	MC	872	0.79	0.41	78.90	11.58	8.94	0.57		-0.18	-0.25
100000098582	58	MC	872	0.52	0.50	52.18	22.94	24.43	0.46		-0.26	-0.17
100000207144	59	MC	872	0.41	0.49	31.77	27.18	40.60	0.46	-0.02	-0.15	
100000198144	60	MC	872	0.69	0.46	13.65	68.69	17.32	0.34	-0.29		-0.29
100000198145	61	MC	872	0.39	0.49	39.11	36.70	23.85	0.34		-0.10	-0.21
100000198147	62	MC	872	0.49	0.50	48.74	19.50	31.31	0.46		-0.10	-0.15
100000186576	63	MC	872	0.72	0.45	10.55	16.74	72.36	0.34	-0.26	-0.21	
100000098666	64	MC	872	0.53	0.50	52.75	18.12	28.78	0.34		-0.16	-0.12
100000198150	65	MC	872	0.70	0.46	70.30	17.32	12.04	0.34		-0.31	-0.25
100000198149	66	MC	872	0.56	0.50	55.85	20.76	22.82	0.57		-0.22	-0.17
100000186573	67	MC	872	0.36	0.48	31.65	32.11	35.89	0.34	-0.19	-0.10	
100000186574	68	MC	872	0.49	0.50	26.26	49.08	24.31	0.34	-0.13		-0.15
100000186575	69	MC	872	0.49	0.50	32.91	49.08	17.55	0.46	-0.12		-0.23
100000186583	70	MC	872	0.50	0.50	18.12	31.31	49.89	0.69	0.05	0.05	
100000198108	71	MC	872	0.44	0.50	14.91	43.92	40.83	0.34	-0.09		-0.29

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000198122	72	MC	872	0.37	0.48	37.04	48.85	13.76	0.34		-0.17	-0.01
100000098645	73	MC	872	0.44	0.50	37.84	17.43	43.92	0.80	-0.15	-0.14	
100000186578	74	MC	872	0.73	0.44	73.28	10.78	15.25	0.69		-0.23	-0.29
100000186580	75	MC	872	0.63	0.48	16.86	20.18	62.61	0.34	-0.28	-0.23	
100000198157	76	MC	872	0.32	0.47	15.48	52.18	31.88	0.46	-0.04	-0.17	
100000098573	77	MC	872	0.42	0.49	30.05	42.43	27.06	0.46	-0.22		-0.08
100000186560	78	MC	872	0.53	0.50	33.03	52.87	13.65	0.46	-0.05		-0.22
100000186561	79	MC	872	0.35	0.48	33.60	34.98	30.96	0.46	-0.12		-0.06
100000198137	80	MC	872	0.44	0.50	44.38	27.64	27.41	0.57		-0.07	-0.30
100000098571	81	MC	872	0.60	0.49	23.05	60.09	16.40	0.46	-0.31		-0.16
100000186581	82	MC	872	0.61	0.49	17.32	20.87	61.47	0.34	-0.19	-0.16	
100000207143	83	MC	872	0.60	0.49	20.07	19.15	60.32	0.46	-0.17	-0.26	
100000198121	84	MC	872	0.41	0.49	40.25	40.71	18.35	0.69	-0.08		-0.17
100000098588	85	MC	872	0.55	0.50	28.21	16.06	55.28	0.46	-0.09	-0.27	
100000098657	86	MC	872	0.35	0.48	35.09	43.46	20.99	0.46		-0.11	-0.15
100000186571	87	MC	872	0.47	0.50	40.14	12.96	46.56	0.34	-0.01	-0.15	
100000186572	88	MC	872	0.47	0.50	46.90	19.84	32.91	0.34		-0.13	-0.16
100000198125	89	MC	872	0.55	0.50	55.16	21.67	22.59	0.57		-0.19	-0.24
100000198151	90	MC	872	0.71	0.45	71.22	16.74	11.58	0.46		-0.25	-0.25
100000098653	91	MC	872	0.33	0.47	46.90	33.03	19.50	0.57	-0.10		-0.10
100000186577	92	MC	872	0.72	0.45	72.13	13.65	13.65	0.57		-0.20	-0.28
100000098568	93	MC	872	0.64	0.48	17.78	17.20	64.45	0.57	-0.32	-0.17	
100000186558	94	MC	872	0.57	0.50	19.95	57.22	22.36	0.46	-0.17		-0.27
100000186559	95	MC	872	0.49	0.50	24.66	48.74	26.15	0.46	-0.14		-0.09
100000186567	96	MC	872	0.54	0.50	54.47	26.49	18.58	0.46		-0.20	-0.16
100000186582	97	MC	872	0.38	0.49	17.20	38.42	44.04	0.34	-0.07		-0.13
100000198127	98	MC	872	0.55	0.50	12.39	55.16	32.00	0.46	-0.20		-0.25
100000098664	99	MC	872	0.52	0.50	32.34	51.61	15.48	0.57	-0.20		-0.17
100000098580	100	MC	872	0.32	0.47	31.31	35.89	32.22	0.57	-0.12	-0.01	
100000186579	101	MC	872	0.50	0.50	49.89	26.72	17.66	5.73		-0.08	-0.17
100000198109	102	MC	872	0.40	0.49	37.39	40.14	16.86	5.62	-0.13		-0.06

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 for 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: Mathematics Grade 5, Distractor Distribution and Distractor-to- Total Correlation Analysis

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000187371	1	MC	926	0.83	0.38	82.61	10.26	7.02	0.11		-0.26	-0.22
100000187367	2	MC	926	0.47	0.50	24.84	28.29	46.76	0.11	-0.14	-0.10	
100000187378	3	MC	926	0.68	0.47	12.85	18.68	68.25	0.22	-0.18	-0.21	
100000099084	4	MC	926	0.67	0.47	20.09	13.28	66.52	0.11	-0.24	-0.24	
100000099080	5	MC	926	0.64	0.48	64.47	19.55	15.87	0.11		-0.18	-0.29
100000099090	6	MC	926	0.40	0.49	50.97	9.29	39.63	0.11	0.00	-0.25	
100000099087	7	MC	926	0.26	0.44	42.44	31.21	26.24	0.11	0.13	-0.19	
100000196233	8	MC	926	0.65	0.48	21.27	65.33	13.28	0.11	-0.28		-0.19
100000196246	9	MC	926	0.12	0.33	12.10	61.34	26.46	0.11		0.21	-0.20
100000196229	10	MC	926	0.46	0.50	43.41	45.68	10.80	0.11	-0.14		-0.17
100000196273	11	MC	926	0.31	0.46	30.99	13.71	55.08	0.22		-0.22	-0.16
100000196043	12	MC	926	0.43	0.50	43.09	42.98	13.50	0.43	-0.12		-0.26
100000196045	13	MC	926	0.52	0.50	27.32	52.38	19.87	0.43	-0.03		-0.21
100000196253	14	MC	926	0.64	0.48	19.55	64.04	16.09	0.32	-0.28		-0.21
100000196278	15	MC	926	0.29	0.45	29.05	19.65	51.08	0.22		-0.27	-0.07
100000196088	16	MC	926	0.33	0.47	33.37	33.48	32.72	0.43		-0.05	-0.09
100000187430	17	MC	926	0.55	0.50	25.81	54.97	18.47	0.76	-0.16		-0.20
100000187390	18	MC	926	0.73	0.44	12.20	13.50	73.43	0.86	-0.20	-0.13	
100000196090	19	MC	926	0.29	0.46	8.64	29.27	61.45	0.65	-0.14		-0.06
100000196081	20	MC	926	0.35	0.48	45.36	34.56	19.98	0.11	-0.08		-0.05
100000187376	21	MC	926	0.55	0.50	26.57	18.57	54.75	0.11	-0.05	-0.14	
100000187386	22	MC	926	0.70	0.46	70.09	15.33	14.47	0.11		-0.10	-0.07
100000187383	23	MC	926	0.41	0.49	32.40	26.89	40.60	0.11	-0.11	-0.08	
100000196231	24	MC	926	0.40	0.49	48.27	39.52	11.99	0.22	-0.04		-0.18
100000196036	25	MC	926	0.68	0.47	14.90	68.36	16.41	0.32	-0.11		-0.21
100000196029	26	MC	926	0.54	0.50	54.43	24.41	20.84	0.32		-0.19	-0.15
100000196242	27	MC	926	0.60	0.49	27.32	60.15	12.42	0.11	-0.20		-0.22
100000099083	28	MC	926	0.53	0.50	16.41	30.78	52.59	0.22	-0.14	-0.08	
100000099075	29	MC	926	0.53	0.50	19.87	53.24	26.67	0.22	-0.14		-0.25
100000099086	30	MC	926	0.62	0.49	61.99	23.54	14.04	0.43		-0.25	-0.17
100000196200	31	MC	926	0.63	0.48	62.74	21.49	15.44	0.32		-0.23	-0.13
100000196277	32	MC	926	0.50	0.50	18.47	31.32	49.89	0.32	-0.12	-0.16	
100000196284	33	MC	926	0.63	0.48	28.19	62.74	8.96	0.11	-0.21		-0.28
100000196057	34	MC	926	0.48	0.50	20.63	31.21	47.95	0.22	-0.20	-0.11	
100000196054	35	MC	926	0.47	0.50	28.62	46.98	24.08	0.32	-0.08		-0.20

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000196279	36	MC	926	0.28	0.45	26.78	44.82	28.29	0.11	-0.06	-0.11	
100000099081	37	MC	926	0.39	0.49	22.03	38.88	38.98	0.11	-0.12		-0.02
100000099079	38	MC	926	0.63	0.48	14.79	22.46	62.63	0.11	-0.15	-0.16	
100000099088	39	MC	926	0.54	0.50	20.84	53.78	25.16	0.22	-0.07		-0.06
100000099085	40	MC	926	0.69	0.46	69.33	20.19	10.15	0.32		-0.25	-0.18
100000196237	41	MC	926	0.39	0.49	39.09	24.73	35.96	0.22		-0.03	-0.04
100000196234	42	MC	926	0.75	0.43	11.12	74.84	13.93	0.11	-0.20		-0.15
100000196053	43	MC	926	0.64	0.48	13.82	21.81	64.15	0.22	-0.24	-0.19	
100000196244	44	MC	926	0.20	0.40	51.08	20.09	28.62	0.22	0.02		-0.10
100000099089	45	MC	926	0.72	0.45	8.64	72.46	18.79	0.11	-0.19		-0.30
100000099091	46	MC	926	0.54	0.50	21.60	24.62	53.67	0.11	-0.13	-0.21	
100000099092	47	MC	926	0.54	0.50	54.43	32.40	13.07	0.11		-0.10	-0.19
100000187365	48	MC	926	0.40	0.49	28.40	39.74	31.75	0.11	-0.15		-0.09
100000187364	49	MC	926	0.46	0.50	45.68	29.16	24.95	0.22		-0.06	-0.21
100000196381	50	MC	926	0.76	0.43	12.10	11.77	75.92	0.22	-0.30	-0.19	
100000196281	51	MC	926	0.41	0.49	37.47	41.14	21.27	0.11	-0.07		-0.10
100000187387	52	MC	926	0.22	0.41	22.03	47.73	30.13	0.11		-0.07	-0.05
100000187388	53	MC	926	0.40	0.49	39.85	21.92	38.01	0.22		-0.14	0.04
100000187389	54	MC	926	0.35	0.48	23.33	35.10	41.25	0.32	-0.10		0.08
100000196270	55	MC	926	0.35	0.48	33.15	34.67	32.07	0.11	-0.17		-0.14
100000196256	56	MC	926	0.53	0.50	22.25	24.84	52.59	0.32	-0.17	-0.18	
100000196267	57	MC	926	0.85	0.36	85.10	9.61	5.18	0.11		-0.21	-0.22
100000099178	58	MC	926	0.53	0.50	25.49	52.81	21.60	0.11	-0.08		-0.22
100000187369	59	MC	926	0.40	0.49	34.99	40.06	24.84	0.11	-0.02		-0.15
100000187380	60	MC	926	0.39	0.49	39.20	33.48	27.00	0.32		-0.05	-0.09
100000187429	61	MC	926	0.56	0.50	23.65	20.63	55.51	0.22	-0.16	-0.23	
100000187428	62	MC	926	0.59	0.49	58.64	19.98	20.95	0.43		-0.20	-0.20
100000187372	63	MC	926	0.24	0.43	38.55	37.37	23.87	0.22	-0.05	0.02	
100000099180	64	MC	926	0.45	0.50	25.27	29.59	44.71	0.43	-0.17	-0.15	
100000099072	65	MC	926	0.56	0.50	23.43	55.72	20.63	0.22	-0.23		-0.21
100000187360	66	MC	926	0.38	0.49	38.12	37.80	23.76	0.32		-0.05	-0.09
100000187361	67	MC	926	0.51	0.50	24.95	51.08	23.65	0.32	-0.22		-0.17
100000196260	68	MC	926	0.41	0.49	45.25	13.39	41.04	0.32	0.04	-0.12	
100000196258	69	MC	926	0.32	0.47	22.68	45.03	31.97	0.32	-0.12	0.04	

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000196269	70	MC	926	0.22	0.42	22.35	38.98	38.23	0.43		-0.05	-0.14
100000196223	71	MC	926	0.52	0.50	30.67	16.74	52.16	0.43	-0.09	-0.14	
100000196100	72	MC	926	0.80	0.40	11.56	80.02	7.99	0.43	-0.31		-0.15
100000196094	73	MC	926	0.47	0.50	40.28	46.76	12.53	0.43	-0.08		-0.24
100000187373	74	MC	926	0.35	0.48	35.85	34.99	28.51	0.65	-0.02		-0.20
100000187363	75	MC	926	0.36	0.48	36.39	35.64	27.32	0.65		-0.04	-0.04
100000196037	76	MC	926	0.47	0.50	21.81	46.54	31.10	0.54	-0.12		-0.05
100000187370	77	MC	926	0.46	0.50	38.12	46.33	15.12	0.43	-0.16		-0.16
100000187391	78	MC	926	0.38	0.49	9.50	51.94	38.01	0.54	-0.21	-0.15	
100000187392	79	MC	926	0.40	0.49	39.85	41.90	17.71	0.54		-0.07	-0.04
100000187382	80	MC	926	0.41	0.49	24.51	34.56	40.71	0.22	-0.07	-0.08	
100000196263	81	MC	926	0.67	0.47	19.33	66.74	13.82	0.11	-0.21		-0.14
100000196247	82	MC	926	0.41	0.49	20.63	38.12	41.04	0.22	-0.15	-0.15	
100000196238	83	MC	926	0.38	0.49	38.23	31.53	29.81	0.43		-0.04	-0.04
100000196079	84	MC	926	0.40	0.49	39.96	33.26	26.67	0.11		0.05	-0.30
100000196025	85	MC	926	0.60	0.49	25.81	13.93	60.04	0.22	-0.07	-0.21	
100000187393	86	MC	926	0.28	0.45	20.19	28.19	51.51	0.11	-0.16		0.15
100000196225	87	MC	926	0.31	0.46	38.66	31.10	30.13	0.11	-0.03		-0.03
100000099177	88	MC	926	0.48	0.50	34.02	18.03	47.84	0.11	-0.16	-0.22	
100000187366	89	MC	926	0.45	0.50	44.92	36.50	18.47	0.11		-0.07	-0.21
100000196049	90	MC	926	0.58	0.49	23.22	57.67	18.90	0.22	-0.16		-0.10
100000196235	91	MC	926	0.28	0.45	22.14	27.86	49.89	0.11	-0.09		0.03
100000099082	92	MC	926	0.44	0.50	24.19	44.06	31.64	0.11	0.04		-0.09
100000196042	93	MC	926	0.40	0.49	39.52	15.33	45.03	0.11		-0.26	-0.03
100000187385	94	MC	926	0.49	0.50	27.54	22.79	49.46	0.22	-0.04	-0.19	
100000187374	95	MC	926	0.45	0.50	38.66	45.14	15.87	0.32	0.07		-0.14
100000187424	96	MC	926	0.40	0.49	15.33	43.74	40.50	0.43	-0.12	-0.06	
100000187379	97	MC	926	0.55	0.50	54.54	22.79	17.49	5.18		-0.14	-0.15
100000187377	98	MC	926	0.24	0.42	23.54	12.63	58.64	5.18		0.00	0.11
100000187381	99	MC	926	0.26	0.44	25.81	52.16	16.74	5.29		0.05	-0.10
100000187375	100	MC	926	0.38	0.48	30.78	26.46	37.58	5.18	0.05	-0.03	

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 for 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: Mathematics Grade 6, Distractor Distribution and Distractor-to- Total Correlation Analysis

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000272103	1	MC	852	0.72	0.45	13.62	72.30	13.97	0.12	-0.15		-0.05
100000272105	2	MC	852	0.62	0.49	61.97	21.01	16.90	0.12		-0.13	-0.08
100000272100	3	MC	852	0.30	0.46	16.08	29.69	53.87	0.35	-0.06		-0.09
100000272095	4	MC	852	0.21	0.41	64.67	20.77	14.20	0.35	-0.05		-0.11
100000272117	5	MC	852	0.41	0.49	21.01	41.08	37.79	0.12	-0.16		-0.16
100000198185	6	MC	852	0.41	0.49	44.13	14.67	40.73	0.47	-0.22	-0.20	
100000272111	7	MC	852	0.35	0.48	35.09	38.15	26.53	0.23		-0.11	0.01
100000272115	8	MC	852	0.37	0.48	43.43	37.21	19.01	0.35	-0.07		-0.01
100000272113	9	MC	852	0.41	0.49	46.71	11.97	41.20	0.12	-0.25	-0.15	
100000272076	10	MC	852	0.58	0.49	21.60	58.45	19.84	0.12	-0.16		-0.21
100000272073	11	MC	852	0.53	0.50	27.70	53.17	18.90	0.23	-0.02		-0.17
100000099232	12	MC	852	0.60	0.49	16.31	23.12	60.33	0.23	-0.27	-0.26	
100000272096	13	MC	852	0.55	0.50	55.05	28.76	15.73	0.47		-0.21	-0.23
100000272145	14	MC	852	0.47	0.50	27.23	25.70	46.60	0.47	-0.12	-0.15	
100000272146	15	MC	852	0.49	0.50	23.71	48.59	27.58	0.12	-0.10		-0.09
100000198162	16	MC	852	0.43	0.50	42.84	17.84	39.20	0.12		-0.11	-0.09
100000272101	17	MC	852	0.46	0.50	34.04	19.95	45.77	0.23	-0.12	-0.14	
100000198232	18	MC	852	0.33	0.47	32.51	19.25	48.12	0.12		-0.14	-0.08
100000272084	19	MC	852	0.34	0.47	38.03	27.70	34.04	0.23	0.09	-0.07	
100000198210	20	MC	852	0.35	0.48	40.96	24.06	34.74	0.23	-0.15	-0.19	
100000198211	21	MC	852	0.51	0.50	51.41	31.57	16.43	0.59		-0.11	-0.23
100000198212	22	MC	852	0.63	0.48	20.89	15.85	62.79	0.47	-0.27	-0.23	
100000272129	23	MC	852	0.52	0.50	26.88	20.66	52.23	0.23	-0.15	-0.22	
100000187836	24	MC	852	0.61	0.49	61.15	19.01	19.72	0.12		-0.17	-0.22
100000272156	25	MC	852	0.60	0.49	18.43	21.60	59.74	0.23	-0.14	-0.20	
100000272158	26	MC	852	0.41	0.49	41.31	37.56	20.66	0.47		0.02	-0.17
100000272159	27	MC	852	0.39	0.49	28.40	32.63	38.73	0.23	-0.09	-0.12	
100000272066	28	MC	852	0.73	0.44	12.09	73.24	14.44	0.23	-0.27		-0.21
100000272077	29	MC	852	0.31	0.46	29.69	38.62	31.22	0.47	-0.03	-0.12	
100000272068	30	MC	852	0.36	0.48	36.03	36.03	27.58	0.35		-0.10	-0.08
100000272075	31	MC	852	0.70	0.46	19.37	69.95	10.56	0.12	-0.19		-0.19
100000272072	32	MC	852	0.61	0.49	18.54	61.03	20.07	0.35	-0.18		-0.09
100000272067	33	MC	852	0.66	0.47	15.26	18.43	65.73	0.59	-0.26	-0.27	
100000272131	34	MC	852	0.44	0.50	21.60	34.51	43.66	0.23	-0.14	-0.16	
100000272127	35	MC	852	0.62	0.48	62.32	18.43	18.90	0.35		-0.22	-0.26

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000272134	36	MC	852	0.65	0.48	13.38	21.36	64.91	0.35	-0.17	-0.22	
100000273844	37	MC	852	0.50	0.50	49.65	32.98	17.02	0.35		-0.05	-0.16
100000273626	38	MC	852	0.31	0.46	24.30	44.37	31.10	0.23	-0.12	-0.05	
100000272143	39	MC	852	0.54	0.50	25.94	54.34	19.13	0.59	-0.17		-0.26
100000272144	40	MC	852	0.45	0.50	45.07	29.81	24.77	0.35		-0.07	-0.22
100000198178	41	MC	852	0.47	0.50	26.53	47.18	25.82	0.47	-0.15		-0.13
100000272099	42	MC	852	0.28	0.45	42.49	27.93	29.34	0.23	-0.03		-0.04
100000272148	43	MC	852	0.42	0.49	21.95	35.33	42.25	0.47	-0.17	0.01	
100000272149	44	MC	852	0.49	0.50	48.71	16.90	34.15	0.23		-0.14	-0.04
100000198235	45	MC	852	0.48	0.50	32.28	47.77	19.60	0.35	-0.05		-0.17
100000272080	46	MC	852	0.69	0.46	15.61	69.13	15.02	0.23	-0.19		-0.20
100000272082	47	MC	852	0.66	0.47	65.73	19.48	14.44	0.35		-0.25	-0.14
100000272114	48	MC	852	0.26	0.44	57.16	15.85	26.29	0.70	-0.09	-0.11	
100000272116	49	MC	852	0.40	0.49	40.14	13.97	45.54	0.35		-0.13	-0.22
100000198186	50	MC	852	0.35	0.48	56.81	8.10	34.62	0.47	-0.19	-0.14	
100000272112	51	MC	852	0.34	0.47	29.81	34.04	35.68	0.47	0.02		-0.09
100000099325	52	MC	852	0.65	0.48	22.30	65.38	12.09	0.23	-0.26		-0.28
100000272128	53	MC	852	0.56	0.50	55.52	27.46	16.55	0.47		-0.12	-0.23
100000272125	54	MC	852	0.45	0.50	21.36	33.22	44.72	0.70	-0.10	-0.14	
100000272069	55	MC	852	0.58	0.49	58.33	20.31	21.01	0.35		-0.23	-0.18
100000272155	56	MC	852	0.23	0.42	23.36	56.81	19.37	0.47		0.17	-0.18
100000272166	57	MC	852	0.39	0.49	31.81	28.40	38.97	0.82	-0.16	-0.10	
100000272167	58	MC	852	0.25	0.43	24.88	32.63	42.14	0.35		-0.21	-0.04
100000187852	59	MC	852	0.60	0.49	12.56	27.46	59.62	0.35	-0.22	-0.24	
100000272065	60	MC	852	0.43	0.50	16.20	40.38	43.31	0.12	-0.17	-0.26	
100000198218	61	MC	852	0.52	0.50	29.69	51.64	18.66		-0.11		-0.25
100000198220	62	MC	852	0.53	0.50	20.89	20.89	52.70	5.52	-0.19	-0.16	
100000198219	63	MC	852	0.50	0.50	50.47	21.60	27.82	0.12		-0.14	-0.26
100000187841	64	MC	852	0.60	0.49	27.70	12.21	59.98	0.12	-0.24	-0.14	
100000272161	65	MC	852	0.27	0.44	26.53	38.73	34.62	0.12		-0.13	-0.03
100000272163	66	MC	852	0.45	0.50	11.03	44.84	44.01	0.12	-0.18		-0.23
100000198170	67	MC	852	0.23	0.42	69.48	22.77	7.63	0.12	-0.15		-0.11
100000272104	68	MC	852	0.45	0.50	15.96	44.60	39.32	0.12	-0.13		-0.07
100000272150	69	MC	852	0.44	0.50	43.78	37.32	18.66	0.23		-0.04	-0.17

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000272151	70	MC	852	0.28	0.45	27.82	37.09	34.74	0.35		-0.04	0.00
100000272098	71	MC	852	0.60	0.49	21.36	18.54	59.98	0.12	-0.20	-0.13	
100000272071	72	MC	852	0.74	0.44	74.30	13.62	11.85	0.23		-0.17	-0.15
100000099236	73	MC	852	0.62	0.48	10.09	27.23	62.44	0.23	-0.16	-0.25	
100000273837	74	MC	852	0.65	0.48	25.12	9.51	65.26	0.12	-0.29	-0.11	
100000273598	75	MC	852	0.51	0.50	51.17	42.02	6.69	0.12		-0.11	-0.18
100000272141	76	MC	852	0.63	0.48	25.35	62.56	11.85	0.23	-0.18		-0.15
100000198202	77	MC	852	0.62	0.49	61.97	14.79	23.00	0.23		-0.24	-0.23
100000272126	78	MC	852	0.54	0.50	54.23	29.34	16.20	0.23		-0.17	-0.16
100000272133	79	MC	852	0.38	0.49	38.26	18.90	42.49	0.35		-0.25	-0.13
100000272132	80	MC	852	0.63	0.48	16.43	20.07	63.26	0.23	-0.19	-0.21	
100000272165	81	MC	852	0.53	0.50	30.63	53.05	15.96	0.35	-0.10		0.00
100000272152	82	MC	852	0.38	0.49	38.38	38.50	23.12			-0.16	-0.07
100000273858	83	MC	852	0.40	0.49	31.92	40.49	27.35	0.23	-0.01		-0.14
100000272160	84	MC	852	0.69	0.46	19.48	11.74	68.66	0.12	-0.24	-0.18	
100000187850	85	MC	852	0.66	0.47	14.55	19.01	66.31	0.12	-0.16	-0.24	
100000272157	86	MC	852	0.18	0.38	52.70	29.23	17.84	0.23	0.01	-0.19	
100000272154	87	MC	852	0.46	0.50	23.83	45.77	29.81	0.59	-0.02		-0.10
100000272153	88	MC	852	0.57	0.49	19.95	22.54	57.39	0.12	-0.16	-0.14	
100000272162	89	MC	852	0.55	0.50	54.69	23.00	22.18	0.12		-0.28	-0.08
100000099252	90	MC	852	0.35	0.48	34.74	24.06	41.08	0.12		-0.22	0.09
100000198165	91	MC	852	0.44	0.50	29.11	26.88	43.78	0.23	-0.16	-0.16	
100000272097	92	MC	852	0.52	0.50	20.89	51.53	27.46	0.12	-0.23		-0.07
100000187833	93	MC	852	0.35	0.48	34.51	35.33	29.81	0.35		-0.12	-0.09
100000272102	94	MC	852	0.47	0.50	46.95	31.10	21.83	0.12		0.11	-0.12
100000198174	95	MC	852	0.40	0.49	40.02	40.02	19.72	0.23	-0.07		-0.15
100000272164	96	MC	852	0.34	0.47	34.27	25.12	40.49	0.12		-0.01	-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 for 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: Mathematics Grade 7, Distractor Distribution and Distractor-to- Total Correlation Analysis

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000322134	1	MC	1000	0.63	0.48	63.00	24.60	12.30	0.10		-0.20	-0.20
100000273428	2	MC	1000	0.29	0.45	45.00	25.90	28.90	0.20	-0.05	-0.08	
100000273431	3	MC	1000	0.27	0.44	26.90	13.40	59.60	0.10		-0.17	0.09
100000272267	4	MC	1000	0.80	0.40	79.50	16.90	3.50	0.10		-0.06	-0.13
100000197499	5	MC	1000	0.59	0.49	16.10	24.50	59.20	0.20	-0.21	-0.20	
100000272217	6	MC	1000	0.60	0.49	19.40	21.00	59.50	0.10	-0.14	-0.08	
100000322127	7	MC	1000	0.20	0.40	19.90	16.70	63.00	0.40		-0.15	-0.08
100000322131	8	MC	1000	0.15	0.36	63.60	21.30	15.00	0.10	0.08	-0.09	
100000197494	9	MC	1000	0.35	0.48	34.70	29.80	35.20	0.30	-0.11	-0.11	
100000322136	10	MC	1000	0.27	0.44	52.00	26.50	21.10	0.40	0.01		0.06
100000322138	11	MC	1000	0.26	0.44	38.40	34.70	26.40	0.50	-0.06	-0.13	
100000197496	12	MC	1000	0.21	0.40	39.60	39.70	20.60	0.10	-0.17	-0.01	
100000322166	13	MC	1000	0.63	0.48	62.60	23.50	13.80	0.10		-0.16	-0.20
100000322165	14	MC	1000	0.43	0.50	29.10	27.90	42.90	0.10	-0.03	-0.15	
100000322169	15	MC	1000	0.33	0.47	15.20	51.50	32.80	0.50	-0.02	-0.06	
100000099488	16	MC	1000	0.83	0.37	83.10	10.10	6.40	0.40		-0.26	-0.19
100000322151	17	MC	1000	0.34	0.47	47.70	18.40	33.60	0.30	-0.01	-0.21	
100000322141	18	MC	1000	0.54	0.50	21.90	23.80	53.90	0.40	-0.16	-0.14	
100000322152	19	MC	1000	0.52	0.50	37.00	52.10	10.50	0.40	-0.19		-0.07
100000322140	20	MC	1000	0.58	0.49	19.50	57.60	22.60	0.30	-0.05		-0.17
100000287243	21	MC	1000	0.76	0.43	10.60	13.40	75.80	0.20	-0.16	-0.18	
100000272268	22	MC	1000	0.63	0.48	62.50	19.90	17.40	0.20		-0.21	-0.01
100000287244	23	MC	1000	0.16	0.37	25.80	57.60	16.20	0.40	-0.17	0.03	
100000272270	24	MC	1000	0.32	0.47	39.20	32.20	28.20	0.40	0.04		-0.15
100000322137	25	MC	1000	0.31	0.46	31.80	36.90	30.90	0.40	-0.09	-0.02	
100000322139	26	MC	1000	0.26	0.44	26.30	42.10	31.20	0.40		-0.07	-0.12
100000272201	27	MC	1000	0.38	0.49	23.80	37.90	38.00	0.30	-0.17		0.17
100000322164	28	MC	1000	0.55	0.50	54.50	35.10	10.10	0.30		-0.13	-0.20
100000272171	29	MC	1000	0.60	0.49	15.60	60.40	23.60	0.40	-0.06		-0.11
100000099420	30	MC	1000	0.63	0.48	63.10	25.30	11.30	0.30		0.00	-0.18
100000099421	31	MC	1000	0.72	0.45	14.10	71.50	13.90	0.50	-0.16		-0.22
100000272271	32	MC	1000	0.37	0.48	26.90	35.30	37.40	0.40	-0.06	-0.13	
100000272216	33	MC	1000	0.43	0.49	42.70	31.40	25.20	0.70		-0.06	-0.05
100000272272	34	MC	1000	0.47	0.50	47.10	27.00	25.40	0.50		-0.10	-0.12
100000322175	35	MC	1000	0.57	0.49	18.40	24.10	57.30	0.20	-0.04	-0.26	

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000322172	36	MC	1000	0.31	0.46	28.10	40.30	31.20	0.40	-0.03	-0.16	
100000197528	37	MC	1000	0.34	0.47	36.90	29.10	33.50	0.50	0.07	-0.27	
100000374151	38	MC	1000	0.37	0.48	36.60	34.50	28.20	0.70		-0.08	0.07
100000374152	39	MC	1000	0.41	0.49	28.30	29.90	41.00	0.80	-0.04	0.03	
100000196052	40	MC	1000	0.42	0.49	25.80	41.80	31.90	0.50	-0.07		0.07
100000196055	41	MC	1000	0.33	0.47	32.40	33.90	33.30	0.40	-0.05	-0.09	
100000322129	42	MC	1000	0.62	0.49	16.20	62.00	21.40	0.40	-0.19		-0.12
100000197481	43	MC	1000	0.59	0.49	58.90	19.40	21.50	0.20		-0.17	-0.24
100000273423	44	MC	1000	0.47	0.50	46.70	28.70	24.10	0.50		-0.10	-0.11
100000273427	46	MC	1000	0.37	0.48	37.30	41.20	21.00	0.50		0.07	-0.09
100000272203	48	MC	1000	0.52	0.50	22.50	25.60	51.70	0.20	-0.17	-0.25	
100000322153	49	MC	1000	0.58	0.49	28.70	13.40	57.60	0.30	-0.28	-0.19	
100000272246	50	MC	1000	0.29	0.45	27.40	29.20	43.20	0.20	0.11		-0.15
100000322154	51	MC	1000	0.62	0.49	10.20	61.50	28.00	0.30	-0.11		-0.16
100000322128	52	MC	1000	0.29	0.45	22.60	28.80	47.30	1.30	-0.07		-0.06
100000322173	53	MC	1000	0.34	0.47	32.80	32.40	34.30	0.50	0.03	-0.06	
100000322174	54	MC	1000	0.28	0.45	19.90	28.30	51.20	0.60	-0.06		0.02
100000273415	55	MC	1000	0.29	0.45	28.80	37.80	32.70	0.70		-0.02	0.01
100000273416	56	MC	1000	0.17	0.38	57.50	25.20	17.00	0.30	-0.06	-0.02	
100000272274	57	MC	1000	0.41	0.49	41.40	25.10	33.30	0.20		-0.11	-0.06
100000272186	58	MC	1000	0.65	0.48	19.30	65.10	15.10	0.50	-0.15		-0.12
100000322135	59	MC	1000	0.65	0.48	65.30	18.90	15.20	0.60		-0.23	-0.20
100000322132	60	MC	1000	0.48	0.50	47.90	35.10	16.50	0.50		-0.09	-0.05
100000272187	61	MC	1000	0.46	0.50	10.90	42.10	46.40	0.60	-0.11	-0.17	
100000374196	63	MC	1000	0.34	0.47	33.70	23.50	42.40	0.40		-0.07	-0.19
100000272248	64	MC	1000	0.58	0.49	57.80	25.60	16.60			-0.11	-0.15
100000322146	65	MC	1000	0.33	0.47	29.40	38.10	32.50		-0.07	-0.06	
100000197501	66	MC	1000	0.46	0.50	46.20	27.70	26.10			-0.05	-0.22
100000322158	67	MC	1000	0.44	0.50	24.00	31.70	44.10	0.20	-0.27	-0.32	
100000322160	68	MC	1000	0.48	0.50	47.90	31.70	20.20	0.20		-0.25	-0.19
100000322163	69	MC	1000	0.46	0.50	29.60	24.10	45.90	0.40	-0.16	-0.16	
100000272174	70	MC	1000	0.41	0.49	41.40	38.30	20.30			-0.21	-0.22
100000099504	71	MC	1000	0.32	0.47	26.90	32.10	41.00		-0.27		0.03
100000322142	72	MC	1000	0.80	0.40	5.00	15.40	79.50	0.10	-0.14	-0.21	

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000167818	73	MC	1000	0.63	0.48	23.20	13.90	62.80	0.10	-0.22	-0.37	
100000322143	74	MC	1000	0.80	0.40	80.10	5.30	14.50	0.10		-0.19	-0.23
100000322147	75	MC	1000	0.49	0.50	49.40	20.80	29.70	0.10		-0.11	-0.22
100000099493	76	MC	1000	0.69	0.46	69.10	22.00	8.60	0.30		-0.18	-0.18
100000273417	77	MC	1000	0.47	0.50	15.90	47.00	36.70	0.40	-0.10		-0.05
100000273444	78	MC	1000	0.47	0.50	47.10	32.30	20.60			-0.26	-0.21
100000273445	79	MC	1000	0.33	0.47	33.40	37.50	29.10			-0.07	-0.08
100000197524	80	MC	1000	0.19	0.39	31.80	49.00	19.20		-0.04	-0.14	
100000272247	81	MC	1000	0.68	0.47	24.70	68.40	6.70	0.20	-0.19		-0.16
100000322149	82	MC	1000	0.68	0.47	20.80	68.00	11.00	0.20	-0.31		-0.20
100000322150	83	MC	1000	0.50	0.50	9.50	50.30	39.60	0.60	0.03		-0.18
100000322144	84	MC	1000	0.74	0.44	73.80	20.00	5.80	0.40		-0.22	-0.09
100000322148	85	MC	1000	0.61	0.49	11.20	27.40	61.20	0.20	-0.16	-0.37	
100000099410	86	MC	1000	0.43	0.50	11.20	43.20	45.40	0.20	-0.11		-0.29
100000322161	87	MC	1000	0.42	0.49	41.90	32.80	25.00	0.30		-0.23	-0.05
100000099413	88	MC	1000	0.25	0.43	24.70	47.00	27.80	0.50		-0.23	0.06
100000322157	89	MC	1000	0.75	0.44	11.70	74.50	13.80		-0.16		-0.14
100000099404	90	MC	1000	0.49	0.50	48.70	37.10	14.20			-0.35	-0.15
100000196255	91	MC	1000	0.55	0.50	25.10	55.40	18.90	0.60	-0.13		-0.13
100000197522	92	MC	1000	0.59	0.49	30.20	10.50	59.00	0.30	-0.10	-0.22	
100000196257	93	MC	1000	0.37	0.48	37.30	26.90	35.50	0.30		-0.24	0.05
100000196261	94	MC	1000	0.47	0.50	23.80	47.20	28.70	0.30	-0.01		-0.07
100000322171	95	MC	1000	0.48	0.50	27.10	48.10	24.50	0.30	-0.20		-0.29
100000322162	96	MC	1000	0.47	0.50	29.30	23.90	46.60	0.20	-0.09	-0.25	
100000322159	97	MC	1000	0.42	0.49	42.10	22.90	35.00			-0.16	-0.23
100000099490	98	MC	1000	0.46	0.50	46.10	26.90	27.00			-0.19	-0.14

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 for 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: Mathematics Grade 8, Distractor Distribution and Distractor-to- Total Correlation Analysis

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000322198	1	MC	1017	0.49	0.50	41.59	48.87	9.14	0.39	0.06		-0.03
100000322187	2	MC	1017	0.41	0.49	42.97	40.81	16.03	0.20	-0.02		-0.12
100000197359	3	MC	1017	0.60	0.49	25.17	60.37	14.06	0.39	-0.14		-0.16
100000272279	4	MC	1017	0.15	0.35	61.06	24.19	14.55	0.20	-0.12	0.01	
100000272278	5	MC	1017	0.37	0.48	37.27	27.83	34.91			-0.03	-0.05
100000272340	6	MC	1017	0.47	0.50	19.57	33.33	47.10		-0.17	-0.11	
100000099548	7	MC	1017	0.74	0.44	10.91	73.94	15.04	0.10	-0.24		-0.14
100000322121	8	MC	1017	0.54	0.50	21.53	24.78	53.69		-0.05	-0.25	
100000322186	9	MC	1017	0.73	0.44	73.25	14.65	12.00	0.10		-0.07	-0.15
100000322122	10	MC	1017	0.59	0.49	58.80	28.61	12.59			0.02	-0.05
100000322120	11	MC	1017	0.55	0.50	10.32	54.77	34.81	0.10	-0.14		-0.24
100000099550	12	MC	1017	0.43	0.50	28.52	28.42	43.07		-0.22	-0.15	
100000272296	13	MC	1017	0.26	0.44	50.74	22.91	26.35		-0.01	-0.04	
100000272337	14	MC	1017	0.28	0.45	24.88	47.39	27.53	0.20	-0.14	-0.07	
100000272295	15	MC	1017	0.29	0.45	28.61	41.10	30.29			-0.10	-0.03
100000272297	16	MC	1017	0.51	0.50	31.27	51.23	17.31	0.20	-0.07		-0.04
100000272323	17	MC	1017	0.45	0.50	13.37	44.84	41.40	0.39	-0.04		-0.05
100000322119	18	MC	1017	0.58	0.49	57.62	29.40	12.78	0.20		-0.14	-0.14
100000322118	19	MC	1017	0.47	0.50	47.10	21.44	30.97	0.49		-0.21	-0.16
100000322084	20	MC	1017	0.32	0.47	20.75	47.30	31.66	0.29	0.01	-0.10	
100000197347	21	MC	1017	0.38	0.49	42.97	38.45	18.49	0.10	-0.12		-0.11
100000322081	22	MC	1017	0.54	0.50	21.63	54.38	23.70	0.29	-0.15		-0.12
100000272309	23	MC	1017	0.33	0.47	23.40	43.66	32.65	0.29	-0.05	0.01	
100000322209	24	MC	1017	0.53	0.50	21.24	53.39	25.37		-0.11		-0.14
100000322208	25	MC	1017	0.65	0.48	64.80	23.30	11.70	0.20		-0.20	-0.17
100000322205	26	MC	1017	0.42	0.49	34.22	42.38	22.62	0.79	-0.06		-0.04
100000322207	27	MC	1017	0.77	0.42	77.48	15.34	6.59	0.59		-0.22	-0.14
100000197387	28	MC	1017	0.42	0.49	41.69	27.83	29.99	0.49		-0.15	-0.28
100000286862	29	MC	1017	0.39	0.49	20.26	40.81	38.54	0.39	-0.03	0.02	
100000272790	30	MC	1017	0.58	0.49	12.19	29.11	58.21	0.49	-0.22	-0.19	
100000286872	31	MC	1017	0.40	0.49	39.92	20.75	38.94	0.39		-0.16	0.06
100000373799	32	MC	1017	0.37	0.48	37.36	57.82	4.62	0.20		-0.15	-0.12
100000272368	33	MC	1017	0.34	0.47	44.44	21.53	33.63	0.39	0.04	-0.01	
100000322200	34	MC	1017	0.34	0.47	23.89	41.89	34.02	0.20	-0.19	-0.14	
100000099626	35	MC	1017	0.72	0.45	11.41	16.42	72.07	0.10	-0.17	-0.13	

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000286792	36	MC	1017	0.45	0.50	44.74	32.06	23.21			-0.13	-0.07
100000272298	37	MC	1017	0.27	0.44	26.65	54.67	18.58	0.10		-0.07	-0.04
100000286786	38	MC	1017	0.76	0.43	14.26	75.81	9.54	0.39	-0.14		-0.02
100000272300	39	MC	1017	0.15	0.36	60.28	24.58	14.85	0.29	0.02	-0.11	
100000322079	40	MC	1017	0.44	0.50	32.15	43.95	23.50	0.39	-0.02		-0.08
100000322078	41	MC	1017	0.59	0.49	34.41	58.70	6.59	0.29	-0.20		-0.08
100000197367	42	MC	1017	0.36	0.48	37.27	26.65	35.99	0.10	-0.12	-0.16	
100000286859	43	MC	1017	0.33	0.47	33.53	33.43	32.94	0.10	-0.20	-0.11	
100000272305	44	MC	1017	0.30	0.46	29.60	39.92	30.29	0.20	-0.13	-0.15	
100000286860	45	MC	1017	0.27	0.44	26.84	40.31	32.65	0.20		-0.08	-0.13
100000322116	46	MC	1017	0.49	0.50	29.99	49.16	20.75	0.10	0.10		-0.19
100000322117	47	MC	1017	0.83	0.38	5.41	82.89	11.31	0.39	-0.16		-0.24
100000286845	48	MC	1017	0.42	0.49	21.44	36.18	41.99	0.39	-0.11	-0.14	
100000272301	49	MC	1017	0.15	0.36	53.10	31.07	15.44	0.39	-0.02	-0.10	
100000286846	50	MC	1017	0.30	0.46	10.23	29.79	59.98		-0.13		-0.06
100000322080	51	MC	1017	0.46	0.50	30.88	22.81	46.12	0.20	-0.07	-0.19	
100000272353	52	MC	1017	0.48	0.50	29.30	47.98	22.71		-0.21		-0.02
100000167565	53	MC	1017	0.48	0.50	27.53	23.89	48.18	0.39	-0.07	-0.14	
100000167566	54	MC	1017	0.30	0.46	31.37	38.15	30.38	0.10	-0.10	-0.06	
100000174558	55	MC	1017	0.26	0.44	26.16	30.48	43.17	0.20		-0.05	-0.13
100000322204	56	MC	1017	0.34	0.47	33.43	34.12	32.25	0.20	-0.05		-0.14
100000322206	57	MC	1017	0.81	0.39	81.12	11.21	7.67			-0.17	-0.18
100000322199	58	MC	1017	0.41	0.49	24.09	40.71	35.00	0.20	-0.09		-0.02
100000099639	59	MC	1017	0.40	0.49	27.04	32.84	39.82	0.29	-0.15	-0.15	
100000272308	60	MC	1017	0.43	0.50	14.45	42.48	42.97	0.10	-0.17	-0.10	
100000197322	61	MC	1017	0.45	0.50	45.13	39.43	15.44			-0.15	-0.07
100000322082	62	MC	1017	0.42	0.49	30.97	27.14	41.79	0.10	-0.06	-0.21	
100000322115	63	MC	1017	0.74	0.44	74.34	14.65	11.01			-0.22	-0.13
100000197365	64	MC	1017	0.38	0.49	38.25	26.75	34.91	0.10		-0.13	-0.13
100000272339	65	MC	1017	0.56	0.50	30.58	55.75	13.47	0.20	-0.25		-0.17
100000197377	66	MC	1017	0.61	0.49	8.95	29.50	61.36	0.20	0.01	-0.21	
100000197373	67	MC	1017	0.44	0.50	43.76	31.17	24.78	0.29		-0.02	-0.21
100000197375	68	MC	1017	0.18	0.39	46.90	34.61	18.19	0.29	-0.17	-0.02	
100000197376	69	MC	1017	0.64	0.48	63.91	20.55	15.44	0.10		-0.16	-0.13

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

Item CID	Item Seq. No.[1]	Item Type	n-Count	Mean	SD	Distractor Distribution				Distractor/Total Correlations[1]		
						1	2	3	Omit	1	2	3
100000322201	70	MC	1017	0.89	0.31	89.18	2.95	7.87			-0.11	-0.15
100000322203	71	MC	1017	0.28	0.45	25.27	46.61	28.02	0.10	-0.11	-0.10	
100000322193	72	MC	1017	0.48	0.50	16.81	48.18	34.91	0.10	0.06		-0.20
100000099553	73	MC	1017	0.68	0.47	22.81	67.55	9.34	0.29	-0.28		-0.19
100000322185	74	MC	1017	0.53	0.50	13.18	52.70	33.82	0.29	-0.14		-0.25
100000322195	75	MC	1017	0.44	0.50	43.56	34.02	22.12	0.29		-0.16	-0.16
100000373853	76	MC	1017	0.48	0.50	17.40	34.12	48.28	0.20	-0.09	-0.30	
100000167561	77	MC	1017	0.58	0.49	30.29	57.72	11.90	0.10	-0.23		-0.19
100000373854	78	MC	1017	0.52	0.50	15.54	32.06	52.21	0.20	-0.25	-0.25	
100000322188	79	MC	1017	0.58	0.49	23.21	18.58	57.82	0.39	-0.30	-0.21	
100000197319	80	MC	1017	0.36	0.48	31.07	32.45	35.99	0.49	-0.04	-0.19	
100000322196	81	MC	1017	0.65	0.48	21.34	65.09	13.57		-0.14		-0.11
100000322191	82	MC	1017	0.52	0.50	19.67	51.62	28.42	0.29	-0.18		-0.10
100000322189	83	MC	1017	0.76	0.42	8.46	76.50	14.95	0.10	-0.20		-0.29
100000099647	84	MC	1017	0.52	0.50	52.41	23.60	23.89	0.10		-0.22	-0.11
100000286880	85	MC	1017	0.37	0.48	37.46	23.70	38.84			-0.13	-0.06
100000272303	86	MC	1017	0.57	0.50	28.81	56.54	14.65		-0.09		-0.13
100000286879	87	MC	1017	0.41	0.49	40.61	30.09	29.11	0.20		-0.08	-0.10
100000286793	88	MC	1017	0.40	0.49	29.11	30.38	40.31	0.20	-0.03	-0.14	
100000286797	89	MC	1017	0.39	0.49	39.33	43.66	16.72	0.29		-0.17	-0.15
100000272293	90	MC	1017	0.74	0.44	12.68	74.24	12.88	0.20	-0.10		-0.21
100000272294	91	MC	1017	0.64	0.48	17.70	64.01	17.99	0.29	-0.19		-0.13
100000322192	92	MC	1017	0.56	0.50	29.01	56.44	14.45	0.10	-0.29		-0.10
100000322190	93	MC	1017	0.29	0.45	21.44	49.36	28.91	0.29	-0.16	-0.06	
100000322202	94	MC	1017	0.77	0.42	5.60	77.48	16.81	0.10	-0.10		-0.24
100000272367	95	MC	1017	0.80	0.40	5.51	80.43	14.06		-0.15		-0.17
100000099631	96	MC	1017	0.35	0.48	17.31	34.51	48.18		-0.08		-0.11
100000322197	97	MC	1017	0.41	0.49	41.49	47.89	10.52	0.10		-0.24	-0.18
100000322194	98	MC	1017	0.64	0.48	17.80	63.62	18.49	0.10	-0.08		-0.18

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 for 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 Mod-MSA Math Assessment

Grade 3

Histogram Bars

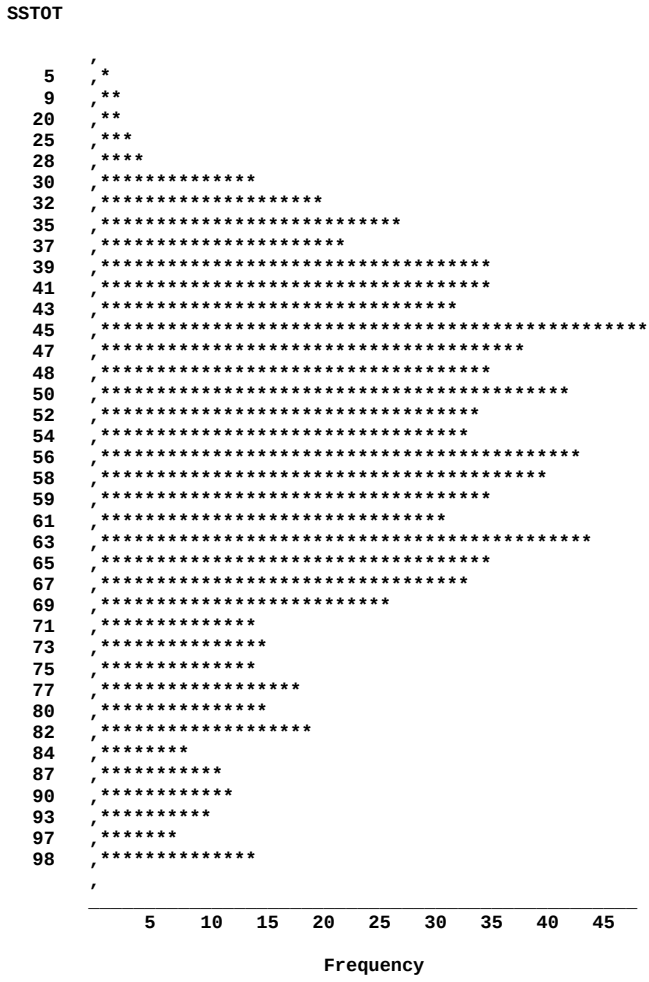
SSTOT		Freq	Cum. Freq	Percent	Cum. Percent
13	*	1	1	0.15	0.15
17	*	1	2	0.15	0.30
20	**	2	4	0.30	0.60
25	*****	5	9	0.75	1.35
28	**	2	11	0.30	1.64
30	*******	7	18	1.05	2.69
32	**********	22	40	3.29	5.98
35	**********	21	61	3.14	9.12
37	**********	33	94	4.93	14.05
39	**********	32	126	4.78	18.83
41	**********	25	151	3.74	22.57
43	**********	25	176	3.74	26.31
45	**********	34	210	5.08	31.39
47	**********	36	246	5.38	36.77
48	**********	35	281	5.23	42.00
50	**********	37	318	5.53	47.53
52	**********	26	344	3.89	51.42
54	**********	25	369	3.74	55.16
56	**********	21	390	3.14	58.30
58	**********	34	424	5.08	63.38
59	**********	23	447	3.44	66.82
61	**********	26	473	3.89	70.70
63	**********	23	496	3.44	74.14
65	**********	16	512	2.39	76.53
67	**********	18	530	2.69	79.22
69	**********	19	549	2.84	82.06
71	**********	16	565	2.39	84.45
73	**********	16	581	2.39	86.85
75	**********	18	599	2.69	89.54
77	**********	13	612	1.94	91.48
80	****	4	616	0.60	92.08
82	**********	12	628	1.79	93.87
84	*******	7	635	1.05	94.92
87	*****	5	640	0.75	95.67
90	*****	5	645	0.75	96.41
93	******	6	651	0.90	97.31
97	*********	9	660	1.35	98.65
98	*********	9	669	1.35	100.00

5 10 15 20 25 30 35

Scale Score Histogram for Mod-MSA Math Assessment

Grade 4

Histogram Bars



Scale Score Histogram for MSA-MOD Math Assessment

Grade 5

Histogram Bars

SSTOT		Freq	Cum. Freq	Percent	Cum. Percent
13	*	1	1	0.11	0.11
20	**	2	3	0.22	0.32
23	***	3	6	0.32	0.65
25	******	6	12	0.65	1.30
28	**********	11	23	1.19	2.48
30	**************	14	37	1.51	4.00
32	**************************	28	65	3.02	7.02
35	**	45	110	4.86	11.88
37	*************************************	39	149	4.21	16.09
39	**************************************	40	189	4.32	20.41
41	*********************************	34	223	3.67	24.08
43	*************************************	39	262	4.21	28.29
45	**	48	310	5.18	33.48
47	**	51	361	5.51	38.98
48	***	54	415	5.83	44.82
50	***	54	469	5.83	50.65
52	**	57	526	6.16	56.80
54	**	47	573	5.08	61.88
56	*******************************	31	604	3.35	65.23
58	**********************************	36	640	3.89	69.11
59	**********************************	37	677	4.00	73.11
61	**	41	718	4.43	77.54
63	**************************	28	746	3.02	80.56
65	**********************	24	770	2.59	83.15
67	**********************	23	793	2.48	85.64
69	************************	25	818	2.70	88.34
71	***************	15	833	1.62	89.96
73	******************	19	852	2.05	92.01
75	******************	19	871	2.05	94.06
77	******	6	877	0.65	94.71
80	************	13	890	1.40	96.11
82	********	8	898	0.86	96.98
84	******	6	904	0.65	97.62
87	*******	7	911	0.76	98.38
90	****	4	915	0.43	98.81
93	*****	5	920	0.54	99.35
97	***	3	923	0.32	99.68
98	***	3	926	0.32	100.00

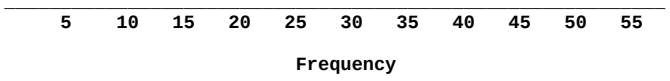
5 10 15 20 25 30 35 40 45 50 55

Frequency

Scale Score Histogram for MSA-MOD Math Assessment
Grade 6

Histogram Bars

SSTOT		Freq	Cum. Freq	Percent	Cum. Percent
	,				
13	,**	2	2	0.23	0.23
20	,*	1	3	0.12	0.35
23	,***	3	6	0.35	0.70
25	,********	8	14	0.94	1.64
28	,******************	18	32	2.11	3.76
30	,************************	24	56	2.82	6.57
32	,**********************	23	79	2.70	9.27
35	,**********************************	35	114	4.11	13.38
37	,*************************************	39	153	4.58	17.96
39	,**********************************	34	187	3.99	21.95
41	,**	47	234	5.52	27.46
43	,**	59	293	6.92	34.39
45	,***	51	344	5.99	40.38
47	,**	47	391	5.52	45.89
48	,**************************************	39	430	4.58	50.47
50	,*************************************	40	470	4.69	55.16
52	,********************************	32	502	3.76	58.92
54	,*****************************	29	531	3.40	62.32
56	,************************************	37	568	4.34	66.67
58	,***	45	613	5.28	71.95
59	,******************************	30	643	3.52	75.47
61	,**********************	22	665	2.58	78.05
63	,**********************	23	688	2.70	80.75
65	,**************************	27	715	3.17	83.92
67	,******************	19	734	2.23	86.15
69	,****************	16	750	1.88	88.03
71	,**********	11	761	1.29	89.32
73	,*********	9	770	1.06	90.38
75	,************	12	782	1.41	91.78
77	,**********	11	793	1.29	93.08
80	,*****	5	798	0.59	93.66
82	,**********	10	808	1.17	94.84
84	,*****	5	813	0.59	95.42
87	,******	6	819	0.70	96.13
90	,*******	7	826	0.82	96.95
93	,*********	9	835	1.06	98.00
97	,******	6	841	0.70	98.71
98	,**********	11	852	1.29	100.00
	,				



Scale Score Histogram for MSA-MOD Math Assessment

Grade 7

Histogram Bars

SSTOT		Freq	Cum. Freq	Percent	Cum. Percent
2	*	1	1	0.10	0.10
17	*	1	2	0.10	0.20
20	*	2	4	0.20	0.40
23	***	5	9	0.50	0.90
25	*****	9	18	0.90	1.80
28	****	8	26	0.80	2.60
30	******	12	38	1.20	3.80
32	********	16	54	1.60	5.40
35	**********	33	87	3.30	8.70
37	**********	45	132	4.50	13.20
39	**********	45	177	4.50	17.70
41	**********	51	228	5.10	22.80
43	**********	65	293	6.50	29.30
45	**********	63	356	6.30	35.60
47	**********	62	418	6.20	41.80
48	**********	61	479	6.10	47.90
50	**********	68	547	6.80	54.70
52	**********	49	596	4.90	59.60
54	**********	56	652	5.60	65.20
56	**********	40	692	4.00	69.20
58	**********	30	722	3.00	72.20
59	**********	35	757	3.50	75.70
61	**********	34	791	3.40	79.10
63	**********	31	822	3.10	82.20
65	**********	29	851	2.90	85.10
67	**********	28	879	2.80	87.90
69	**********	24	903	2.40	90.30
71	**********	29	932	2.90	93.20
73	*******	13	945	1.30	94.50
75	*******	13	958	1.30	95.80
77	***	5	963	0.50	96.30
80	********	15	978	1.50	97.80
82	*****	9	987	0.90	98.70
84	**	3	990	0.30	99.00
87	**	3	993	0.30	99.30
90	**	3	996	0.30	99.60
93	*	1	997	0.10	99.70
98	**	3	1000	0.30	100.00

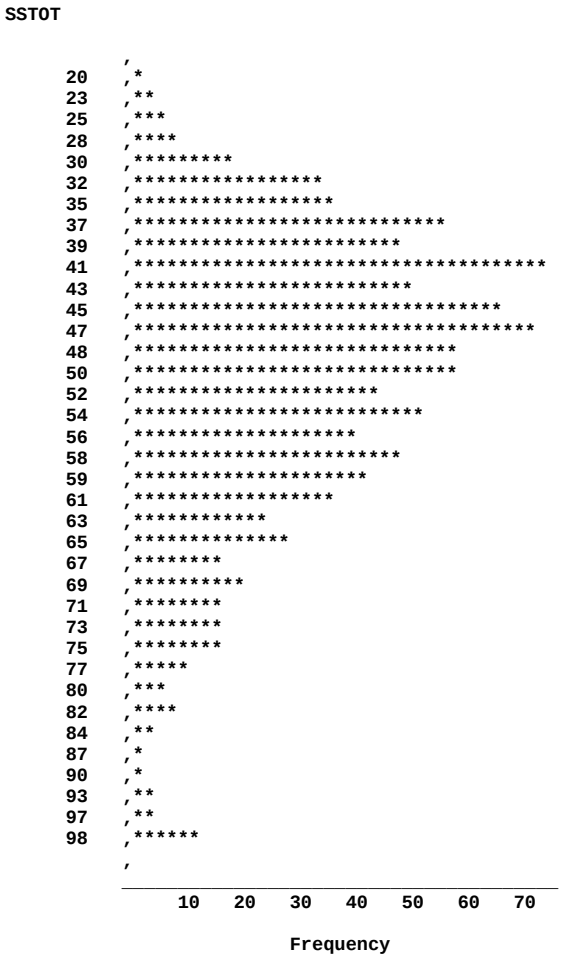
10 20 30 40 50 60

Frequency

Scale Score Histogram for MSA-MOD Math Assessment

Grade 8

Histogram Bars



SSTOT		Freq	Cum. Freq	Percent	Cum. Percent
20	*	1	1	0.10	0.10
23	**	3	4	0.29	0.39
25	***	5	9	0.49	0.88
28	****	8	17	0.79	1.67
30	*********	17	34	1.67	3.34
32	**********	34	68	3.34	6.69
35	**********	36	104	3.54	10.23
37	**********	55	159	5.41	15.63
39	**********	47	206	4.62	20.26
41	**********	73	279	7.18	27.43
43	**********	50	329	4.92	32.35
45	**********	65	394	6.39	38.74
47	**********	71	465	6.98	45.72
48	**********	58	523	5.70	51.43
50	**********	58	581	5.70	57.13
52	**********	44	625	4.33	61.46
54	**********	51	676	5.01	66.47
56	**********	39	715	3.83	70.30
58	**********	48	763	4.72	75.02
59	**********	41	804	4.03	79.06
61	**********	36	840	3.54	82.60
63	**********	24	864	2.36	84.96
65	**********	27	891	2.65	87.61
67	********	15	906	1.47	89.09
69	**********	19	925	1.87	90.95
71	********	15	940	1.47	92.43
73	********	15	955	1.47	93.90
75	********	15	970	1.47	95.38
77	*****	10	980	0.98	96.36
80	***	6	986	0.59	96.95
82	****	7	993	0.69	97.64
84	**	3	996	0.29	97.94
87	*	2	998	0.20	98.13
90	*	2	1000	0.20	98.33
93	**	3	1003	0.29	98.62
97	**	3	1006	0.29	98.92
98	******	11	1017	1.08	100.00

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*[1]

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

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

the round 1 of ratings in addition to the mean and median cut 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.

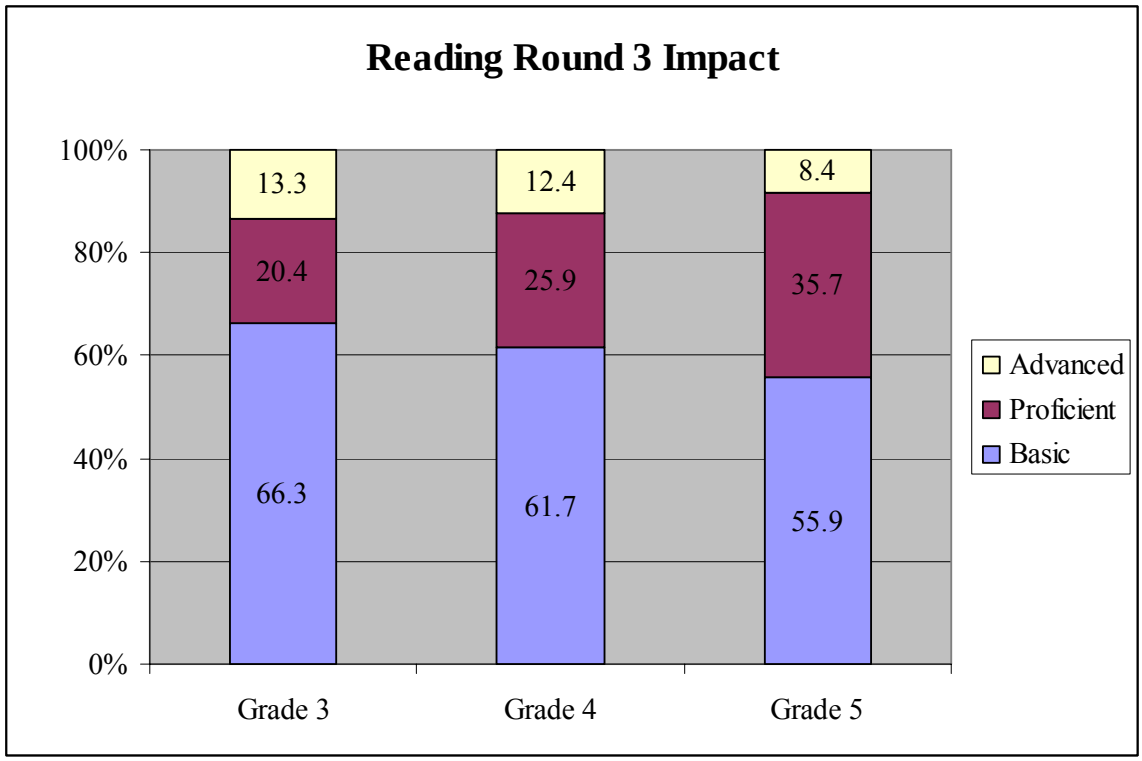


Figure 1. The percentages of students in each performance level, using the final cut scores for Reading by grade.

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.

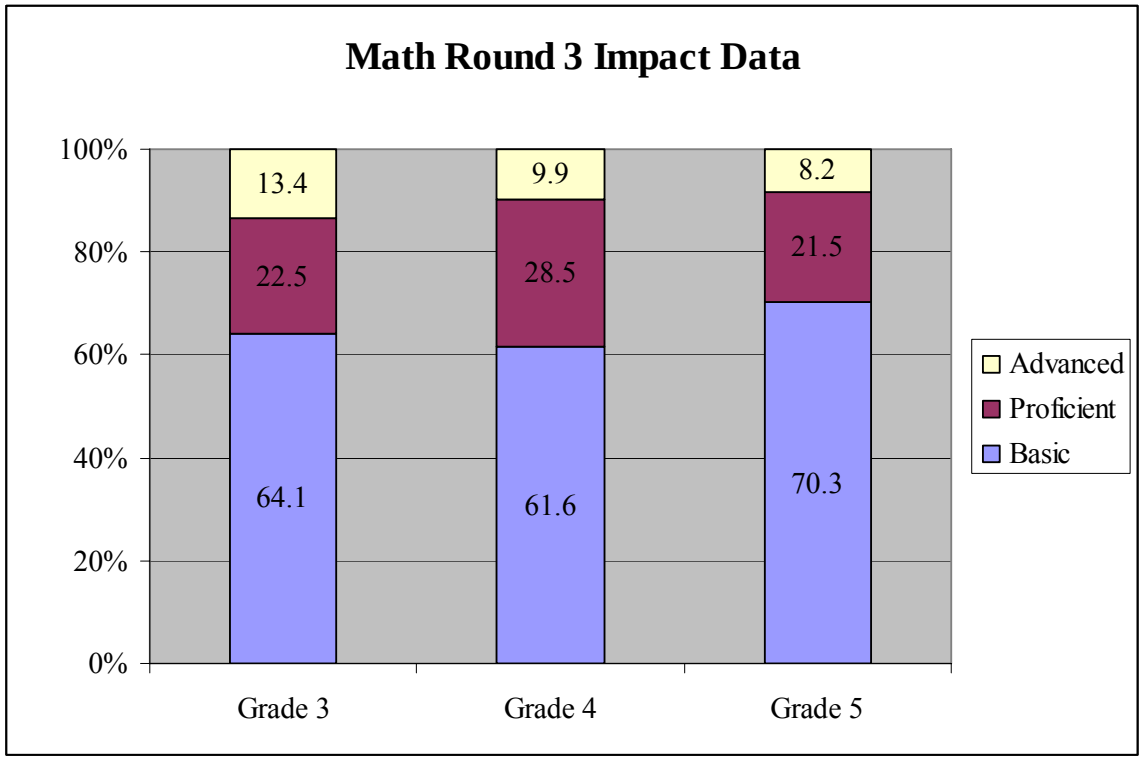


Figure 2. The percentage of students in each performance level, using the final cut scores for Math by grade.

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*[1]

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.

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

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 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.

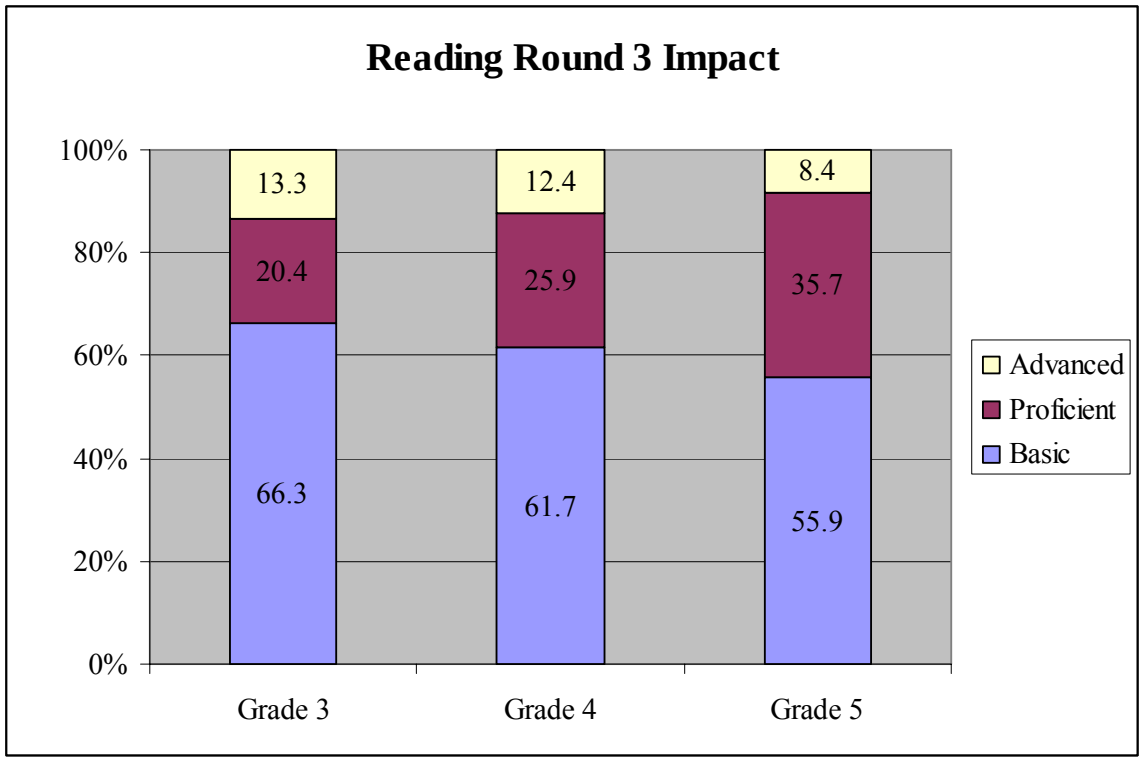


Figure 1. The percentages of students in each performance level, using the final cut scores for Reading by grade.

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.

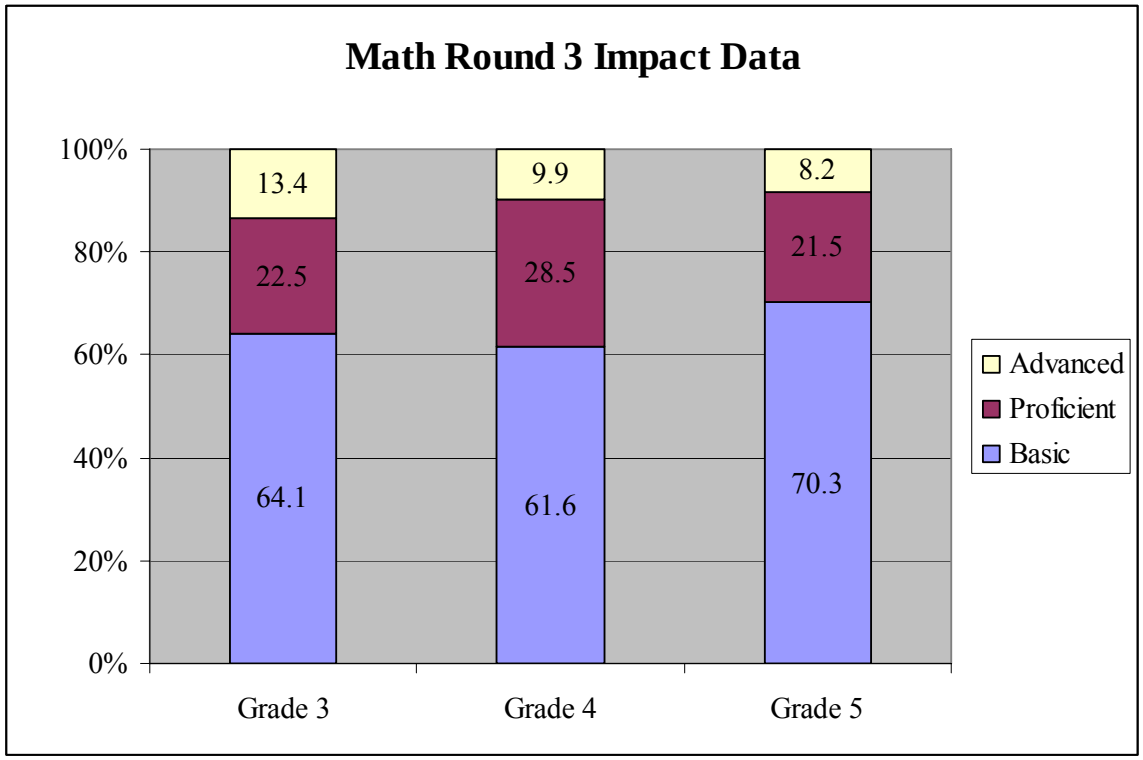


Figure 2. The percentage of students in each performance level, using the final cut scores for Math by grade.

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 (σ^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. σ^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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
27	100000101513	7	208	1B1a	2	56	0.56
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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
26	100000102114	21	212	2A4b	1	60	0.51
27	100000102095	12	212	2A4a	3	60	0.51
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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
27	100000098435	75	241	1A1a	3	59	0.52
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
25	100000186567	96	268	6A3a	1	57	0.54

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
54	100000186582	97	268	6C2a	2	68	0.38
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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
26	100000196223	71	246	2E1a	3	61	0.52
27	100000187361	67	273	7	2	62	0.51
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

Page	Item CID	Item Seq. No	Reporting Strand	Content Standard	Answer Key	Location	P-Value
54	100000187389	54	273	7	2	74	0.35
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.

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

nd 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**

nd 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

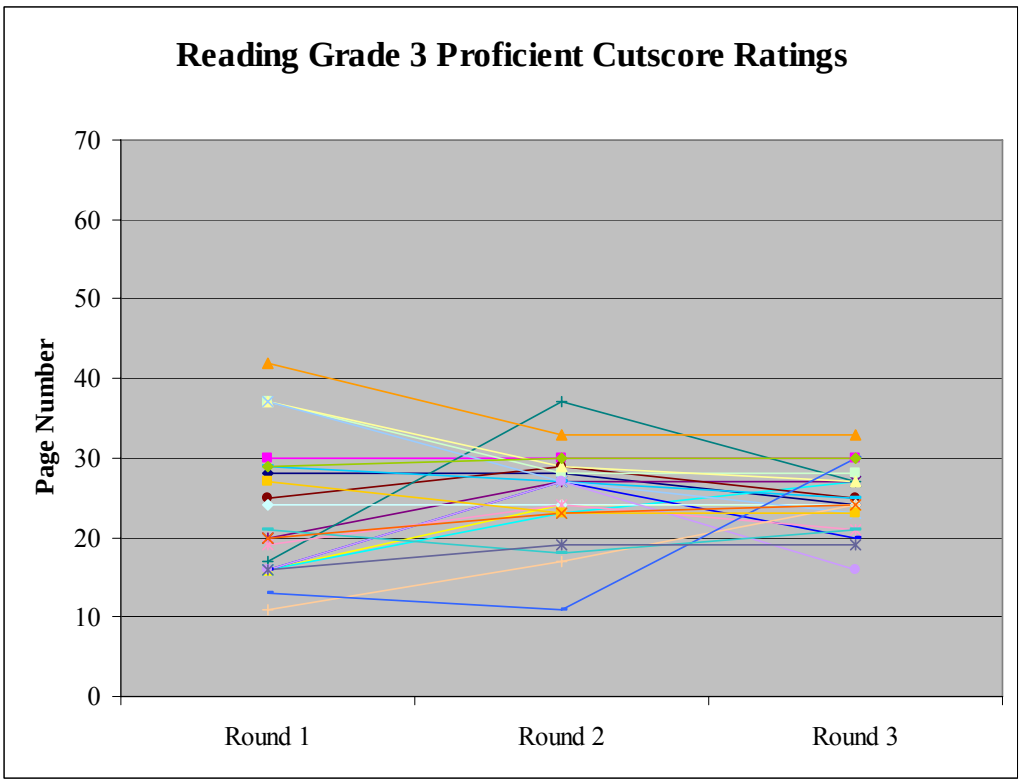
	Reading Grade 3					
Round	1		2		3	
Achievement Level	P	A	P	A	P	A
Mean	23.74	42.43	25.43	43.78	24.87	45.09
Median	21.00	43.00	27.00	44.00	24.00	45.00
Minimum	11.00	30.00	11.00	33.00	16.00	40.00
Maximum	42.00	51.00	37.00	51.00	33.00	51.00
	Reading Grade 4					
Round	1		2		3	
Achievement Level	P	A	P	A	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

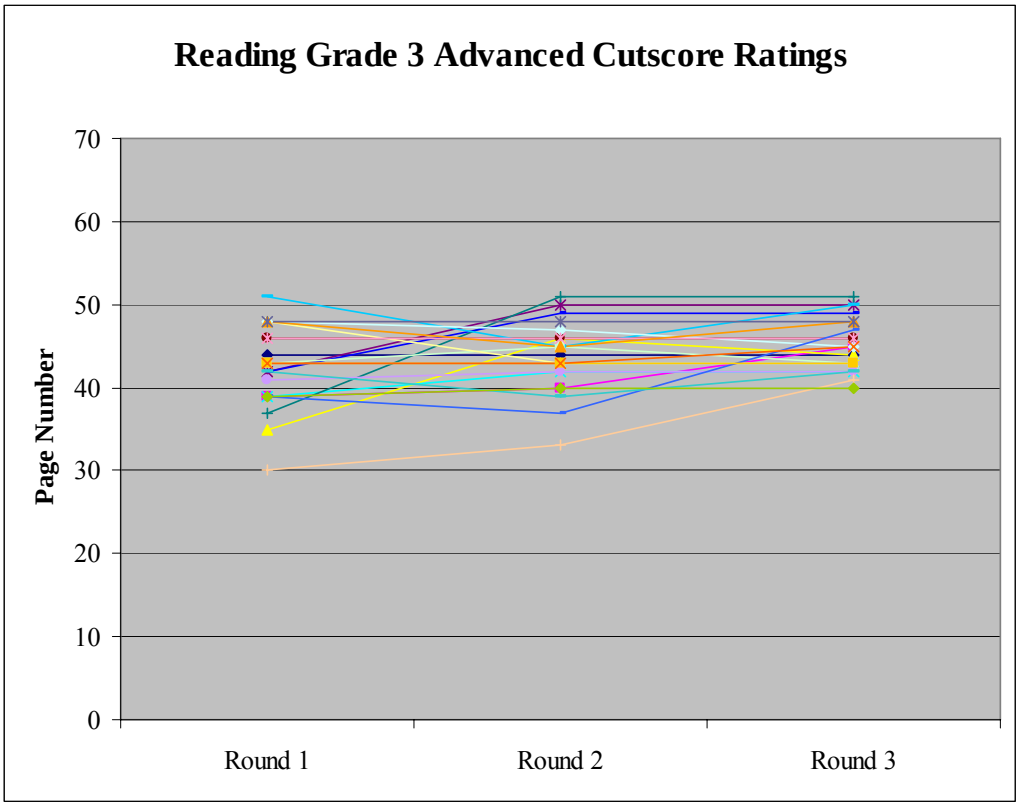
	Math Grade 4					
Round	1		2		3	
Achievement Level	P	A	P	A	P	A
Mean	24.91	48.41	22.09	50.41	18.86	52.68
Median	23.00	50.00	22.00	53.00	17.00	53.00
Minimum	16.00	31.00	17.00	35.00	16.00	48.00
Maximum	39.00	57.00	27.00	54.00	25.00	54.00

	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

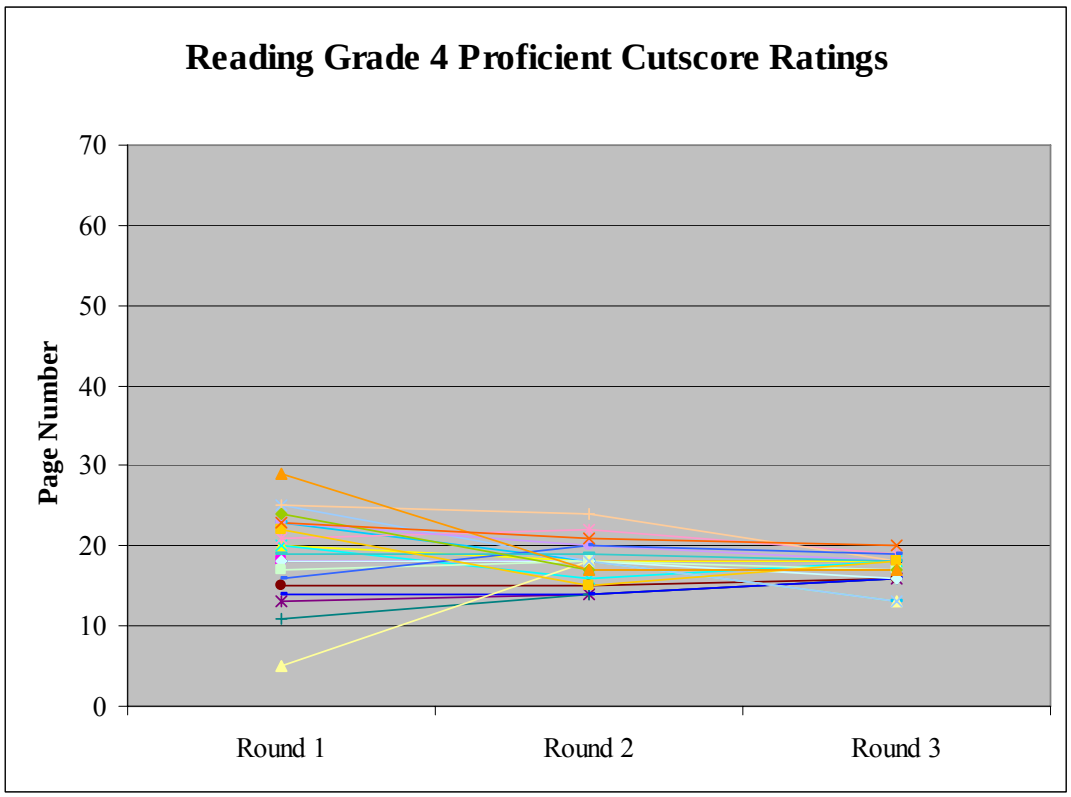
Individual Ratings across Rounds for Reading Grade 3

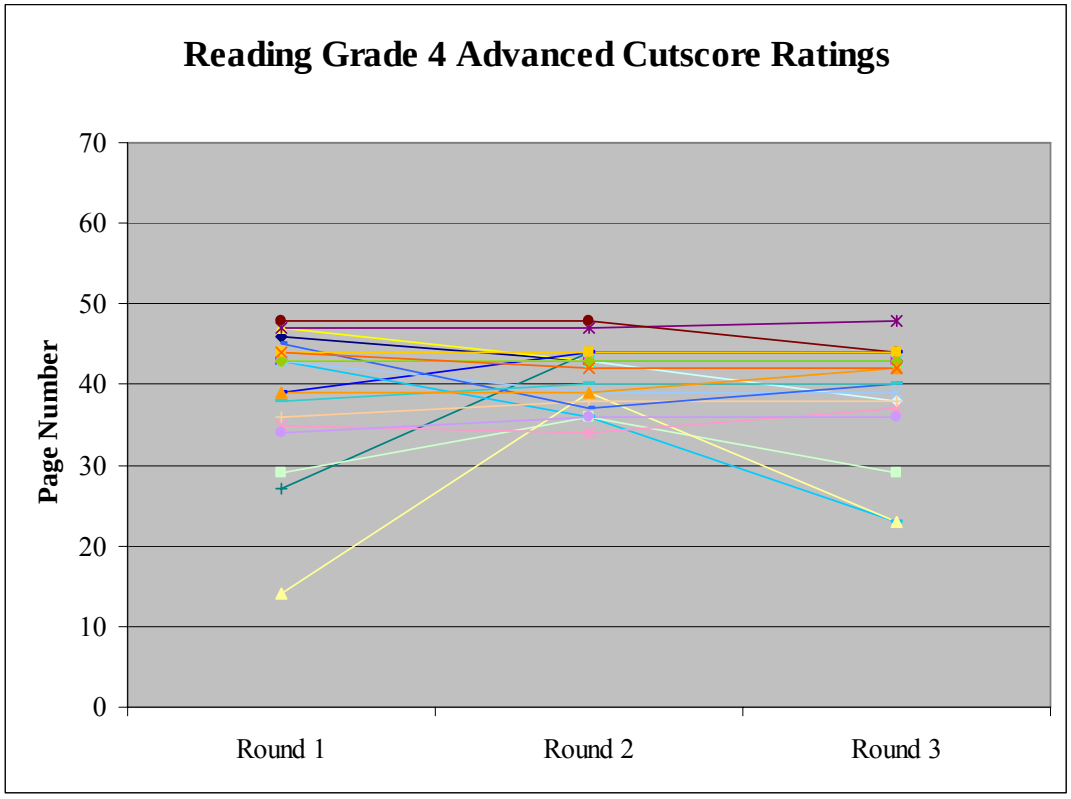



Reading Grade 3 Advanced Cutscore Ratings
Page Number
70
60
50
40
30
20
10
0
Round 1
Round 2
Round 3

Appendix L

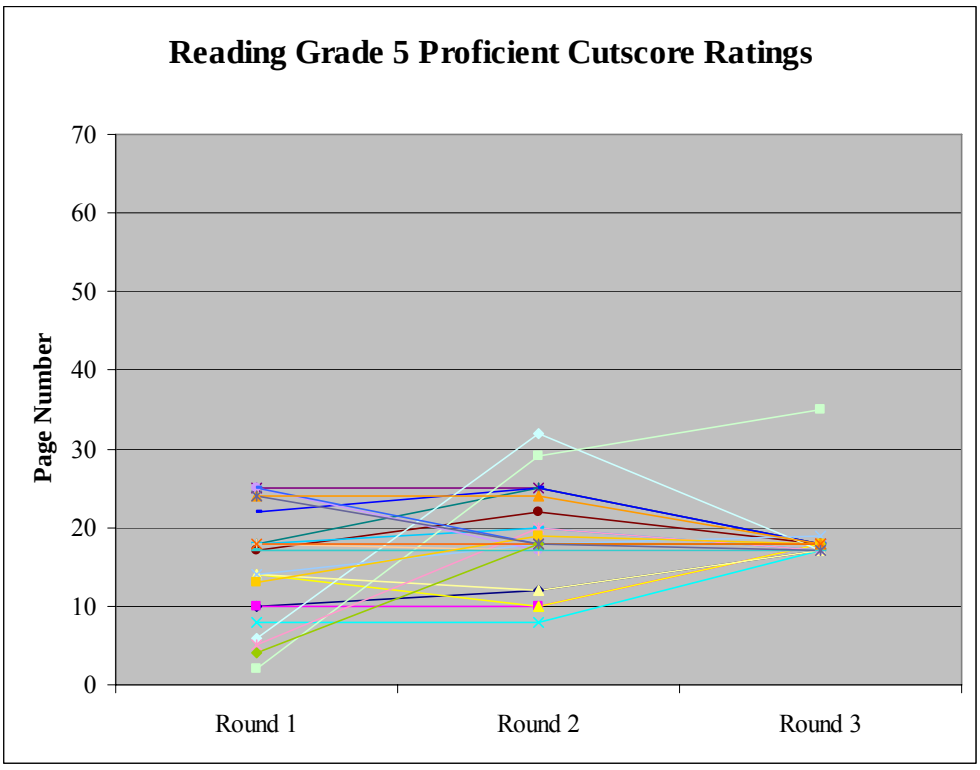
Individual Ratings across Rounds for Reading Grade 4

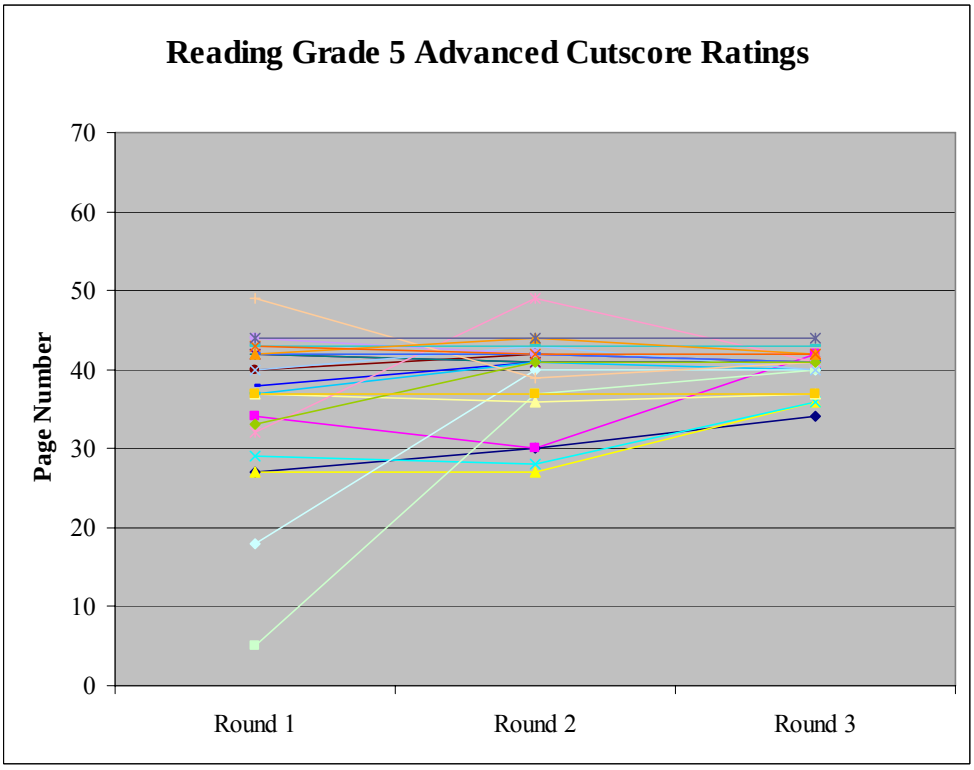



Reading Grade 4 Advanced Cutscore Ratings
Page Number
70
60
50
40
30
20
10
0
Round 1
Round 2
Round 3

Appendix M

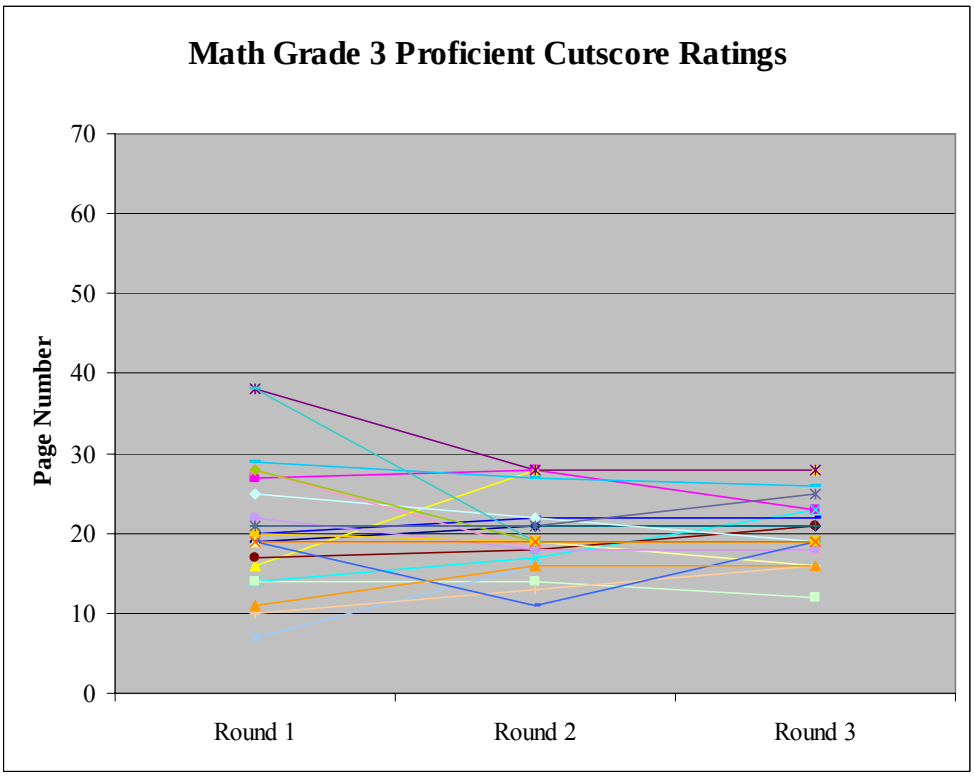
Individual Ratings across Rounds for Reading Grade 5

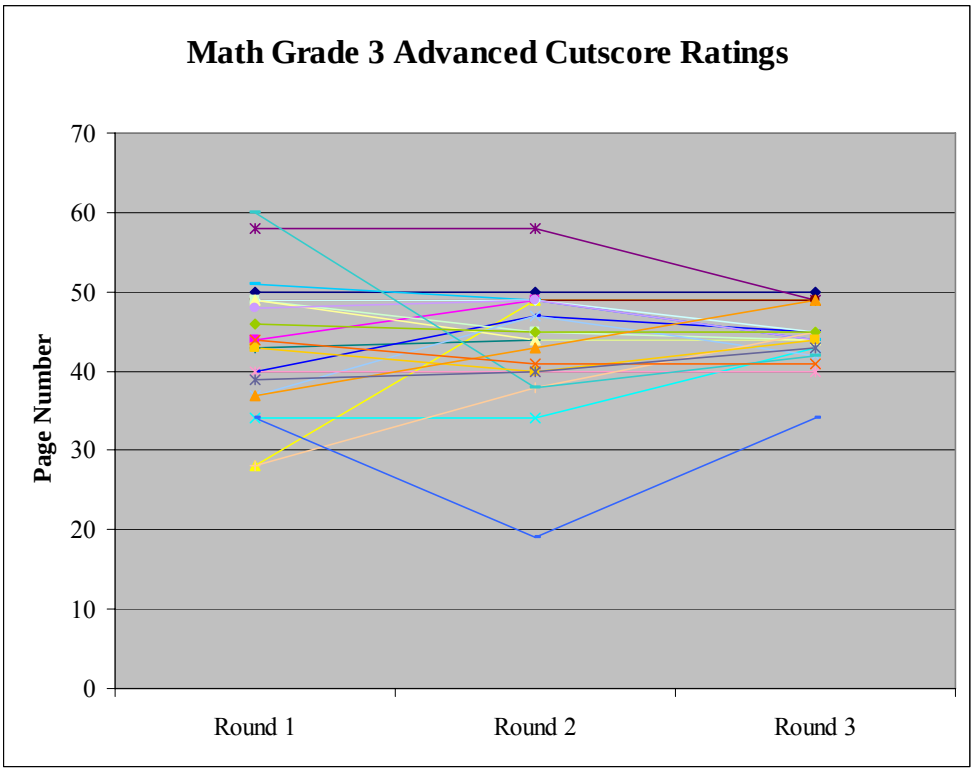




Appendix N

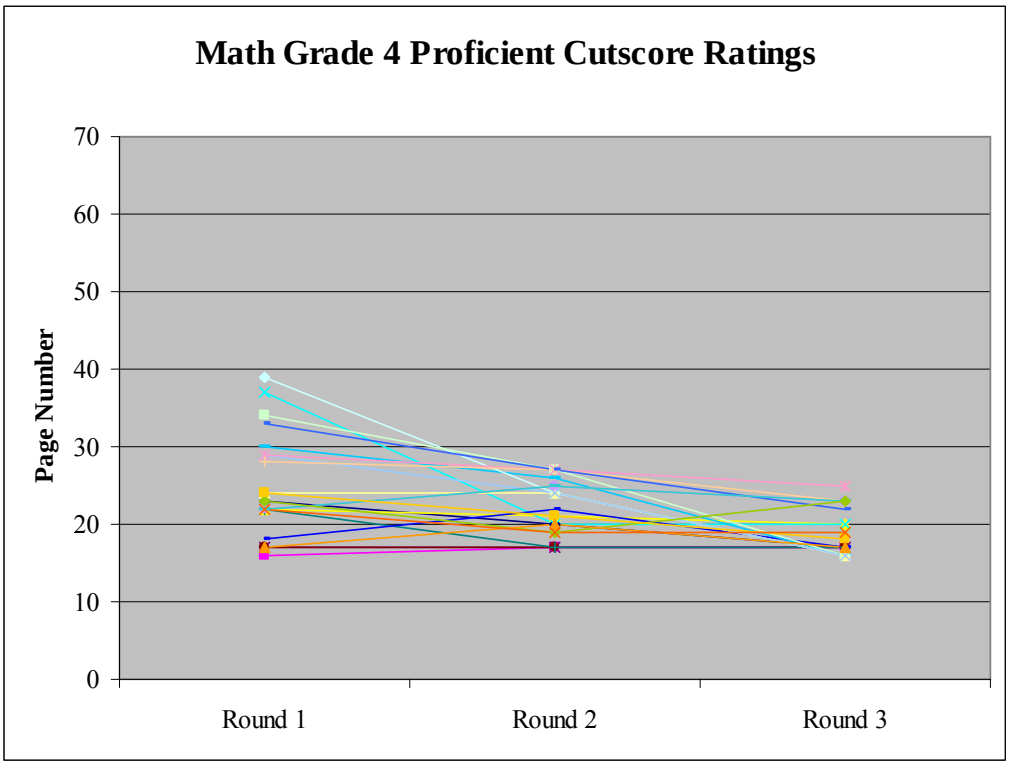
Individual Ratings across Rounds for Math Grade 3

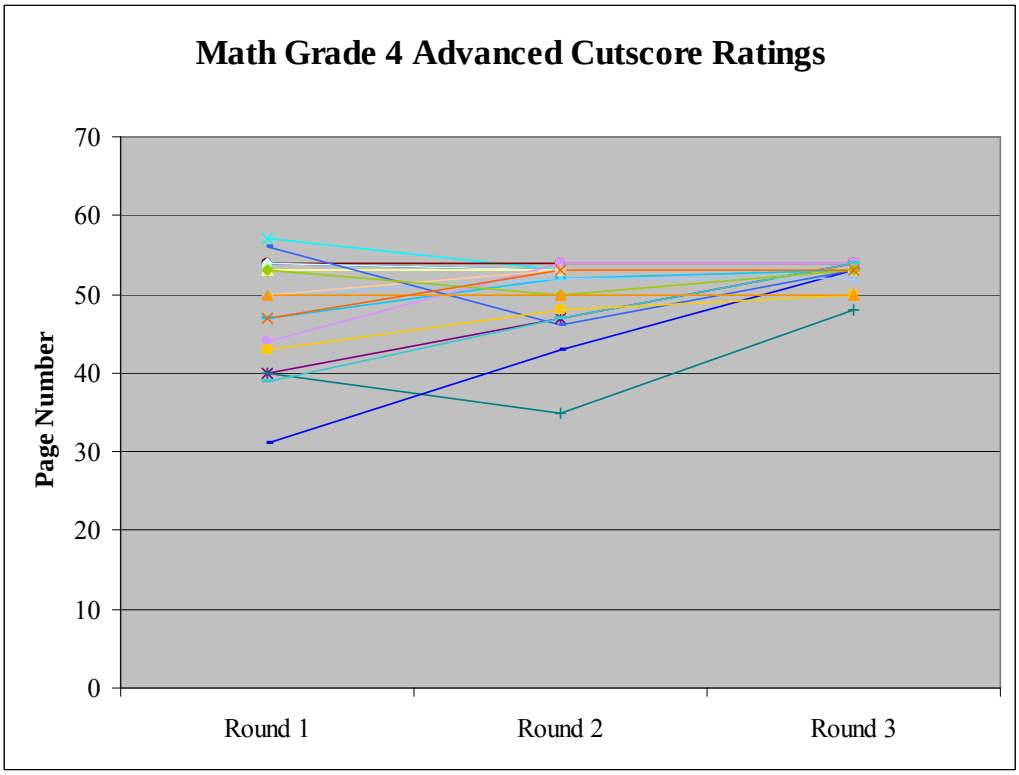



Math Grade 3 Advanced Cutscore Ratings
Page Number
70
60
50
40
30
20
10
0
Round 1
Round 2
Round 3

Appendix O

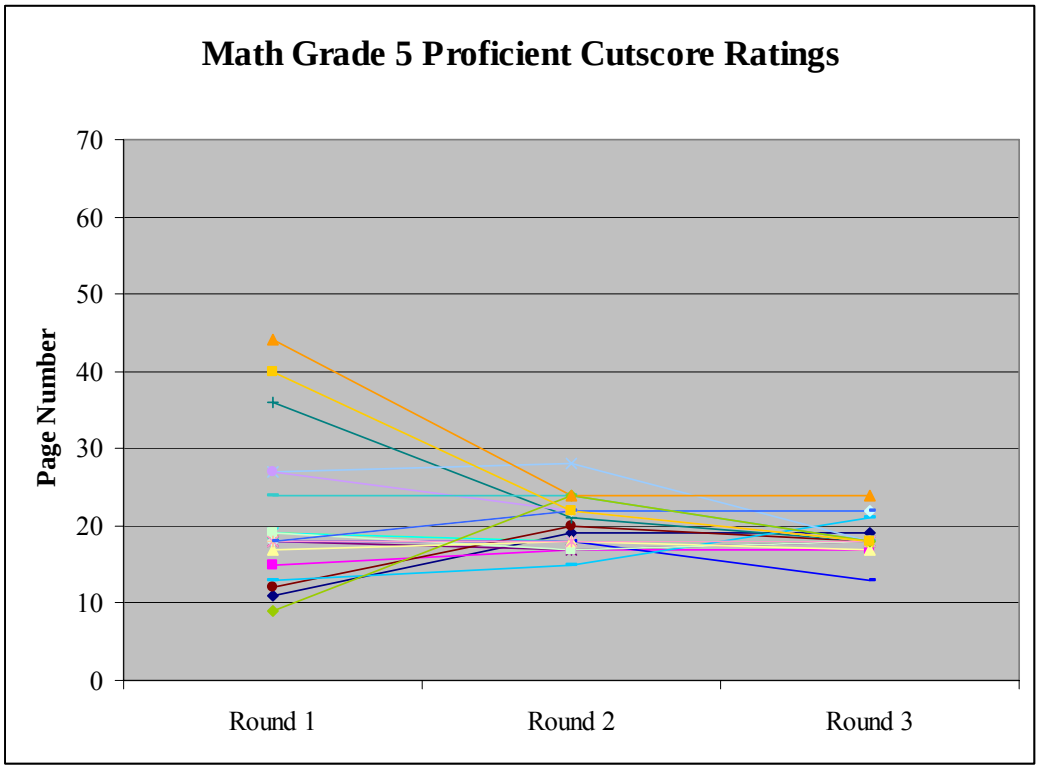
Individual Ratings across Rounds for Math Grade 4

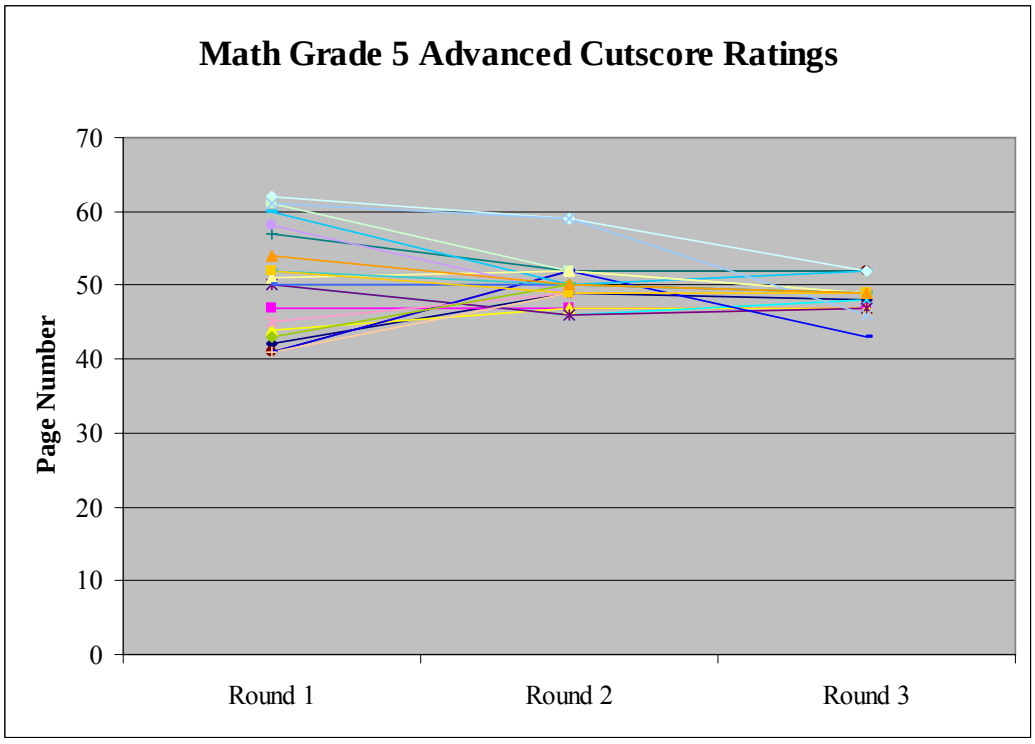



Math Grade 4 Advanced Cutscore Ratings
Page Number
70
60
50
40
30
20
10
0
Round 1
Round 2
Round 3

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 2009 MOD-MSA: MATHEMATICS ITEMS TO THE STATE STANDARDS

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade3.html.

Grade 3

Item #	UIN Number	Strand	Standard Code
1	100000107601	Geometry	2A
2	100000197604	Geometry	2A
3	100000098438	Algebra	1A
4	100000098445	Algebra	1A
6	100000098516	Nos & Comp	6A
7	100000185401	Process	0007
8	100000185403	Process	0007
9	100000197723	Statistics	4A
11	100000197780	Probability	5B
12	100000197781	Probability	5B
14	100000197752	Statistics	4B
15	100000197753	Process	0007
16	100000197654	Process	0007
18	100000197647	Geometry	2E
19	100000098440	Algebra	1A
26	100000197677	Statistics	4A
27	100000197751	Statistics	4A
28	100000197722	Statistics	4A
29	100000350878	Algebra	1C
37	100000197670	Measurement	3C
38	100000197674	Process	0007
39	100000197675	Process	0007
41	100000098444	Algebra	1A
46	100000098454	Algebra	1C
48	100000185378	Nos & Comp	6A
52	100000098449	Algebra	1B
54	100000197602	Geometry	2A
55	100000107621	Geometry	2D
57	100000197756	Statistics	4B
58	100000098446	Algebra	1B
60	100000197648	Geometry	2E
61	100000197649	Process	0007
62	100000197650	Process	0007
63	100000185382	Nos & Comp	6B
64	100000185313	Nos & Comp	6A
65	100000197662	Measurement	3A
71	100000197724	Statistics	4B
73	100000098527	Nos & Comp	6A
74	100000185381	Nos & Comp	6A
75	100000098435	Algebra	1A
76	100000185485	Process	0007
77	100000185486	Process	0007
78	100000185380	Nos & Comp	6A
79	100000185384	Nos & Comp	6C
81	100000197720	Statistics	4A
86	100000197664	Measurement	3C
88	100000185387	Nos & Comp	6C
89	100000185376	Nos & Comp	6A
91	100000197661	Measurement	3A
92	100000098515	Nos & Comp	6A
94	100000197663	Measurement	3B

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade4.html.

Grade 4

Item #	UIN Number	Strand	Standard Code
1	100000198111	Measurement	3C
2	100000198113	Process	0007
3	100000198114	Process	0007
5	100000198123	Statistics	4B
6	100000098578	Algebra	1A
10	100000098583	Algebra	1B
20	100000098579	Algebra	1A
22	100000198140	Probability	5B
23	100000198142	Process	0007
24	100000198143	Process	0007
26	100000198128	Statistics	4B
27	100000198092	Geometry	2A
28	100000198099	Geometry	2D
34	100000098587	Algebra	1C
36	100000198107	Measurement	3A
37	100000198101	Geometry	2E
38	100000198102	Process	0007
39	100000198103	Process	0007
40	100000198148	Probability	5B
41	100000198158	Probability	5B
45	100000198094	Geometry	2B
46	100000198096	Geometry	2B
52	100000098586	Algebra	1C
56	100000198139	Statistics	4B
57	100000098585	Algebra	1B
58	100000098582	Algebra	1B
64	100000098666	Nos & Comp	6C
65	100000198150	Probability	5B
67	100000186573	Nos & Comp	6B
68	100000186574	Process	0007
69	100000186575	Process	0007
71	100000198108	Measurement	3B
72	100000198122	Measurement	3C
73	100000098645	Nos & Comp	6A
74	100000186578	Nos & Comp	6A
75	100000186580	Nos & Comp	6A
76	100000198157	Probability	5B
77	100000098573	Algebra	1A
78	100000186560	Process	0007
79	100000196561	Process	0007
80	100000198137	Statistics	4B
81	100000098571	Algebra	1A
83	100000207143	Nos & Comp	6C
84	100000198121	Measurement	3C
89	100000198125	Statistics	4A
92	100000186577	Nos & Comp	6A
96	100000186567	Nos & Comp	6A
97	100000186582	Nos & Comp	6C
98	100000198127	Statistics	4B
99	100000098664	Nos & Comp	6B
100	100000098580	Algebra	1B

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade5.html.

Grade 5

Item #	UIN Number	Strand	Standard Code
2	100000187367	Nos & Comp	6A
5	100000099080	Algebra	1A
6	100000099090	Algebra	1B
8	100000196233	Measurement	3C
10	100000196229	Measurement	3A
11	100000196273	Statistics	4B
12	100000196043	Process	0007
13	100000196045	Process	0007
14	100000196253	Statistics	4A
16	100000196088	Geometry	2B
21	100000187376	Nos & Comp	6A
24	100000196231	Measurement	3B
25	100000196036	Process	0007
26	100000196029	Process	0007
28	100000099083	Algebra	1B
29	100000099075	Algebra	1A
30	100000099086	Algebra	1B
33	100000196284	Probability	5B
34	100000196057	Process	0007
35	100000196054	Process	0007
36	100000196279	Probability	5A
37	100000099081	Algebra	1B
38	100000099079	Algebra	1A
42	100000196234	Measurement	3C
44	100000196244	Measurement	3C
46	100000099091	Algebra	1C
51	100000196281	Probability	5A
52	100000187387	Nos & Comp	6B
53	100000187389	Process	0007
54	100000187388	Process	0007
55	100000196270	Statistics	4B
56	100000196256	Statistics	4A
60	100000187380	Nos & Comp	6B
61	100000187429	Nos & Comp	6C
62	100000187428	Nos & Comp	6C
64	100000099180	Nos & Comp	6C
65	100000099072	Algebra	1A
66	100000187360	Process	0007
67	100000187361	Process	0007
71	100000196223	Geometry	2E
73	100000196094	Geometry	2C
77	100000187370	Nos & Comp	6A
78	100000187391	Nos & Comp	6C
80	100000187382	Nos & Comp	6B
81	100000196263	Statistics	4B
82	100000196247	Statistics	4A
83	100000196238	Measurement	3C
84	100000196079	Geometry	2A
88	100000099177	Algebra	1C
89	100000187366	Process	0007
93	100000196042	Process	0007

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade6.html.

Grade 6

Item #	UIN Number	Strand	Standard Code
5	100000272117	Measurement	3C
6	100000198185	Measurement	3C
10	100000272076	Algebra	1C
12	100000099232	Algebra	1A
16	100000198162	Geometry	2A
17	100000272101	Geometry	2A
18	100000198232	Probability	5B
20	100000198210	Statistics	4B
21	100000198211	Process	0007
22	100000198212	Process	0007
24	100000187836	Nos & Comp	6A
27	100000272159	Nos & Comp	6A
29	100000272077	Algebra	1C
30	100000272068	Algebra	1B
31	100000272075	Algebra	1C
32	100000272072	Algebra	1B
34	100000272131	Statistics	4B
39	100000272143	Process	0007
41	100000198178	Geometry	2C
45	100000198235	Probability	5C
47	100000272082	Probability	5B
48	100000272114	Measurement	3C
50	100000198186	Measurement	3C
52	100000099325	Algebra	1A
53	100000272128	Statistics	4A
54	100000272125	Statistics	4A
58	100000272167	Nos & Comp	6C
59	100000187852	Nos & Comp	6C
60	100000272065	Algebra	1A
61	100000198218	Statistics	4A
62	100000198220	Process	0007
63	100000198219	Process	0007
64	100000187841	Nos & Comp	6C
65	100000272161	Nos & Comp	6C
66	100000272163	Nos & Comp	6C
67	100000198170	Geometry	2A
68	100000272104	Geometry	2C
69	100000272150	Process	0007
70	100000272151	Process	0007
73	100000099236	Algebra	1B
77	100000198202	Statistics	4A
79	100000272133	Statistics	4B
82	100000272152	Process	0007
84	100000272160	Nos & Comp	6C
85	100000187850	Nos & Comp	6C
90	100000099252	Algebra	1C
91	100000198165	Process	0007
92	100000272097	Geometry	2A
93	100000187833	Process	0007
95	100000198174	Process	0007
96	100000272164	Nos & Comp	6C

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade7.html.

Grade 7

Item #	UIN Number	Strand	Standard Code
05	100000197499	Statistics	4A
06	100000272217	Statistics	4B
09	100000197494	Measurement	3C
12	100000197496	Measurement	3C
13	100000322166	Algebra	1C
16	100000099488	Nos & Comp	6A
25	100000322137	Measurement	3C
29	100000272171	Algebra	1B
30	100000099420	Algebra	1C
31	100000099421	Algebra	1C
32	100000272271	Process	0007
33	100000272216	Statistics	4B
34	100000272272	Process	0007
35	100000322175	Probability	5C
36	100000322172	Probability	5A
37	100000197528	Probability	5B
38	100000374151	Process	0007
39	100000374152	Geometry	2E
40	100000196052	Process	0007
41	100000196055	Process	0007
42	100000322129	Geometry	2A
43	100000197481	Geometry	2D
48	100000272203	Measurement	3C
49	100000322153	Nos & Comp	6C
52	100000322128	Geometry	2A
55	100000273415	Process	0007
56	100000273416	Probability	5A
57	100000272274	Process	0007
58	100000272186	Geometry	2A
63	100000374196	Nos & Comp	6C
66	100000197501	Statistics	4B
69	100000322163	Algebra	1B
71	100000099504	Nos & Comp	6A
73	100000167818	Nos & Comp	6C
75	100000322147	Nos & Comp	6C
76	100000099493	Nos & Comp	6C
77	100000273417	Process	0007
78	100000273444	Algebra	1B
79	100000273445	Process	0007
80	100000197524	Probability	5A
83	100000322150	Nos & Comp	6C
84	100000322144	Nos & Comp	6C
86	100000099410	Algebra	1B
87	100000322161	Algebra	1B
88	100000099413	Algebra	1B
90	100000099404	Algebra	1B
91	100000196255	Process	0007
92	100000197522	Statistics	4B
93	100000196257	Process	0007
94	100000196261	Process	0007
98	100000099490	Nos & Comp	6A

The following tables by grade depict the alignment of the Mod-MSA: Mathematics items to the state standards. For more information regarding the standard codes, visit: http://www.mdk12.org/assessments/vsc/mathematics/bygrade/grade8.html.

Grade 8

Item #	UIN Number	Strand	Standard Code
3	100000197359	Measurement	3C
4	100000272279	Measurement	3C
5	100000272278	Measurement	3C
6	100000272340	Algebra	1B
7	100000099548	Algebra	1A
9	100000322186	Algebra	1A
10	100000322122	Probability	5C
11	100000322120	Probability	5B
12	100000099550	Algebra	1A
19	100000322118	Statistics	4B
20	100000322084	Geometry	2C
21	100000197347	Geometry	2E
24	100000322209	Nos & Comp	6C
25	100000322208	Nos & Comp	6C
26	100000322205	Nos & Comp	6C
28	100000197387	Probability	5B
32	100000373799	Nos & Comp	6C
35	100000099626	Nos & Comp	6A
41	100000322078	Measurement	3C
42	100000197367	Statistics	4B
43	100000286859	Process	0007
44	100000272305	Process	0007
45	100000286860	Statistics	4B
48	100000286845	Process	0007
49	100000272301	Process	0007
50	100000286846	Geometry	2A
53	100000167565	Process	0007
54	100000167566	Process	0007
55	100000174558	Probability	5A
59	100000099639	Nos & Comp	6C
61	100000197322	Geometry	2A
63	100000322115	Statistics	4A
64	100000197365	Statistics	4A
66	100000197377	Process	0007
67	100000197373	Process	0007
68	100000197375	Statistics	4B
69	100000197376	Process	0007
73	100000099553	Algebra	1B
75	100000322195	Algebra	1B
76	100000373853	Process	0007
77	100000167561	Process	0007
78	100000373854	Algebra	1B
80	100000197319	Geometry	2A
82	100000322191	Algebra	1B
84	100000099647	Algebra	1B
85	100000286880	Process	0007
86	100000272303	Process	0007
87	100000286879	Algebra	1B
93	100000322190	Algebra	1B
94	100000322202	Nos & Comp	6C
96	100000099631	Nos & Comp	6C

APPENDIX F: IDENTIFYING STUDENTS FOR PARTICIPATION IN THE MOD-MSA: MATHEMATICS 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).

PPENDIX 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[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 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

[1] 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."

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 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[1] 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).

[1] For a detailed discussion on Mantel-Haenszel Chi-square, the Delta Scale and ETS Categories, please refer to Camilli and Shepard (1994).

Score on studied item with general notation

Group		1	0	Total
	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: $\beta_{M-H} = \ln(\alpha_{M\text{-}H})$. β_{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* χ^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* χ^2 or $\|D\| < 1.0$
B	Weak DIF	Significant *M-H* χ^2 and $\|D\| < 1.5$ or Non-significant *M-H* χ^2 and $\|D\| > 1.0$
C	Strong DIF	Significant *M-H* χ^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.

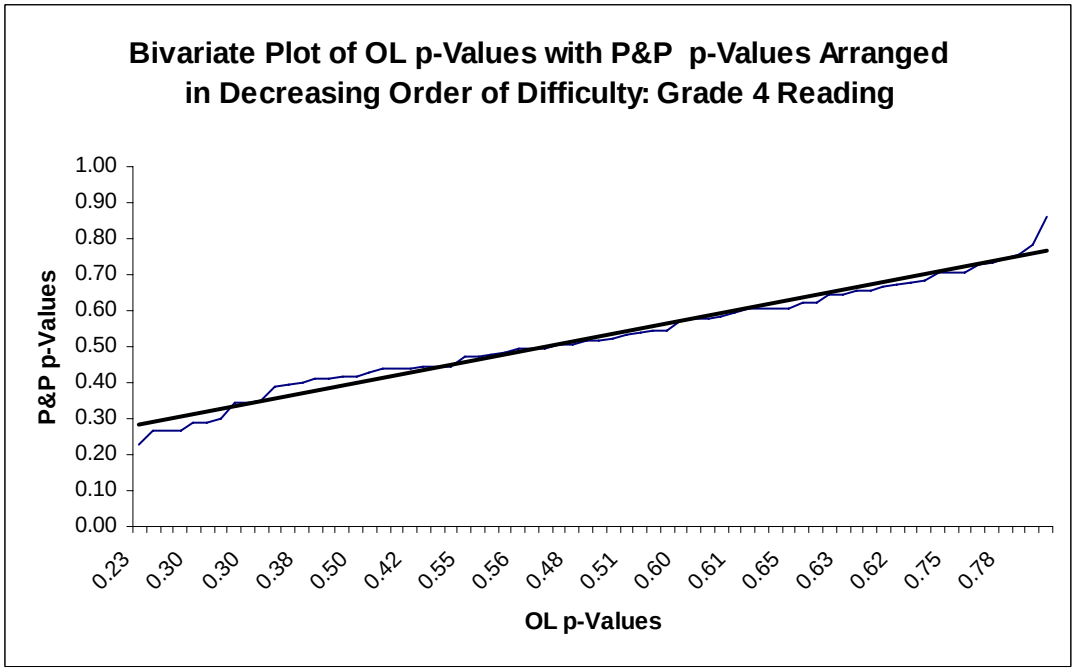


Figure 1: Grade 4 reading item p-values by mode of administration

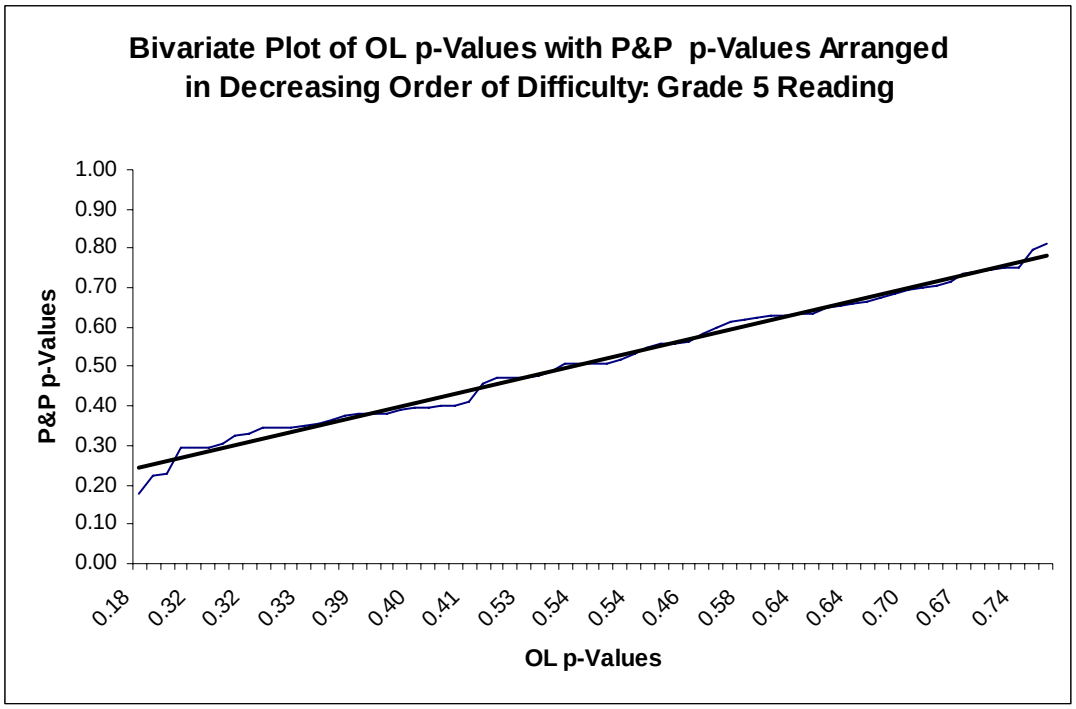


Figure 2: Grade 5 reading item p-values by mode of administration

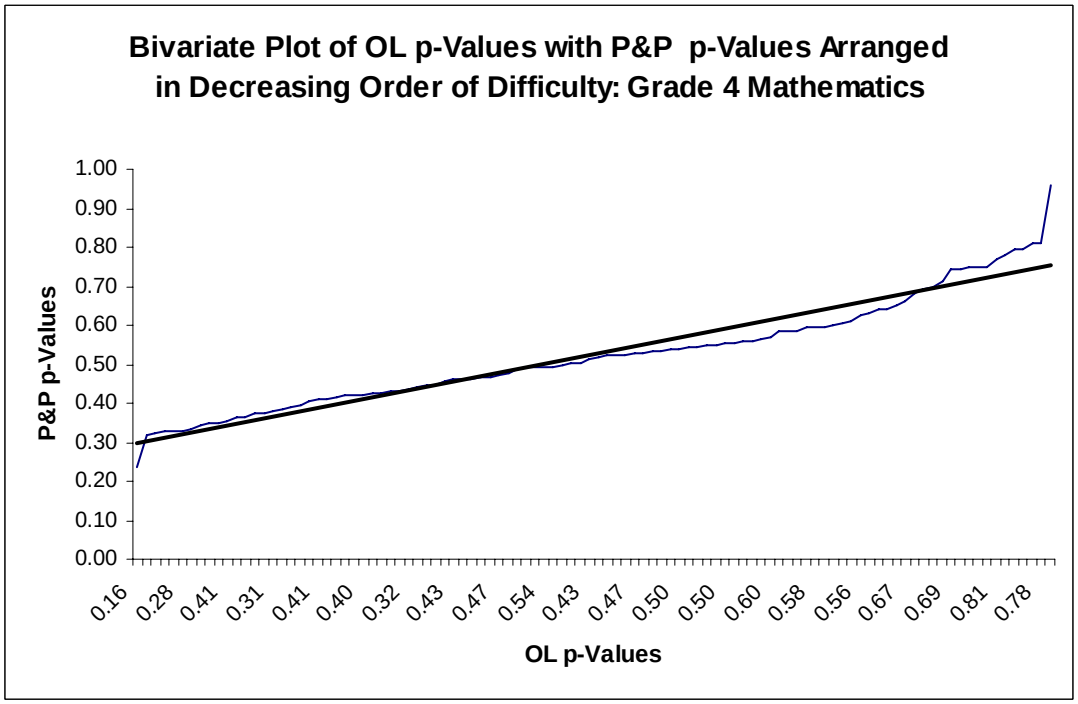


Figure 3: Grade 4 mathematics item p-values by mode of administration

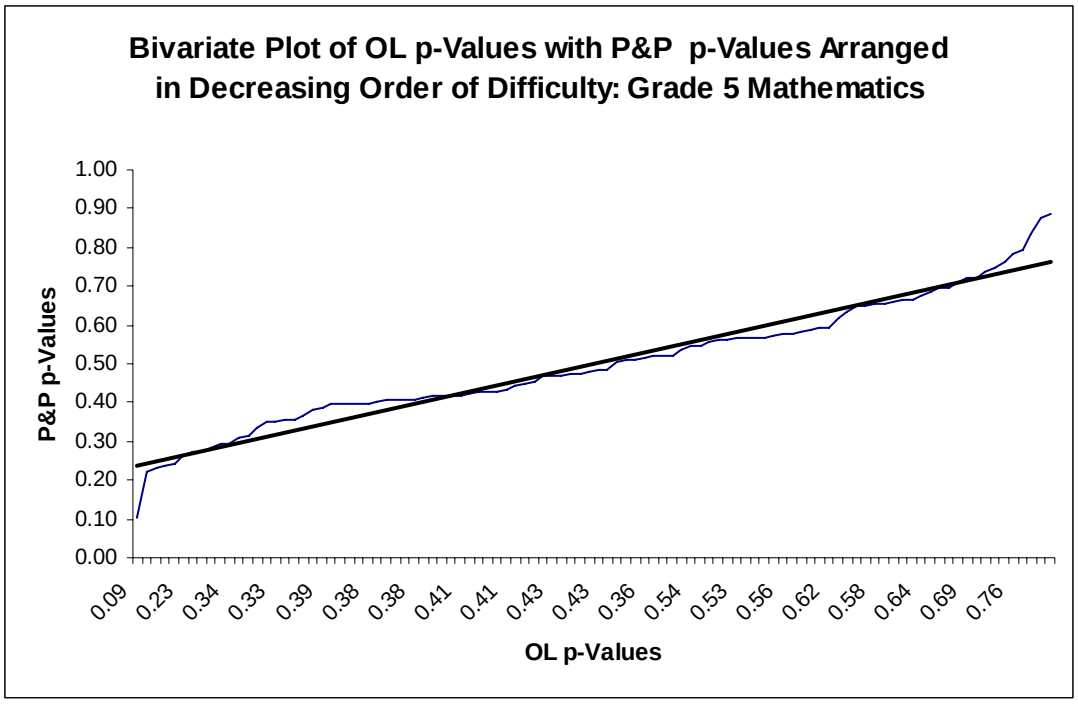


Figure 4: Grade 5 mathematics item p-values by mode of administration

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.