

# Missouri Assessment Program

## Technical Report 2007

November 28, 2007



CTB/McGraw-Hill  
20 Ryan Ranch Road  
Monterey, CA 93940

**Developed and published by CTB/McGraw-Hill LLC, a subsidiary of The McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2007 by Missouri Department of Elementary and Secondary Education. All rights reserved. Only Missouri State educators and citizens may copy and/or download and print the document, located online at <http://dese.mo.gov/divimprove/assess/index.html>.**

# Table of Contents

<b>TABLE OF TABLES .....</b>	<b>III</b>
<b>TABLE OF FIGURES .....</b>	<b>VI</b>
<b>PART 1: OVERVIEW .....</b>	<b>1</b>
BACKGROUND .....	1
ADMINISTRATION .....	1
STUDENT PERFORMANCE.....	2
VALIDITY AND TEST SCORES.....	2
<b>PART 2: VALIDITY AND THE USES OF TEST SCORES.....</b>	<b>4</b>
USES OF TEST SCORES .....	4
TEST-LEVEL SCORES .....	4
<i>Use of Test-Level Scores</i> .....	6
CONTENT STANDARD SUBSCORES.....	6
<i>The Use of the Content Standard Subscores</i> .....	6
PROCESS STANDARD SUBSCORES .....	6
<i>The Use of the Process Standard Subscores</i> .....	7
<b>PART 3: TEST CONTENT DEVELOPMENT .....</b>	<b>8</b>
TEST DESIGN .....	8
ITEM DEVELOPMENT .....	8
<i>Item Writing</i> .....	9
<i>Local Pilot Test</i> .....	9
<i>Score, Revise, Rewrite Workshop</i> .....	10
<i>Content and Bias Review Workshop</i> .....	10
<i>Field-test Selection and Administration</i> .....	10
<i>Operational-test Selection</i> .....	11
<b>PART 4: STUDENT DEMOGRAPHICS AND PERFORMANCE.....</b>	<b>27</b>
CALIBRATION SAMPLE .....	27
CALIBRATION, CENSUS DEMOGRAPHICS AND TEST PERFORMANCE .....	27
LONGITUDINAL COMPARISONS .....	28
STUDENT PARTICIPATION .....	28
TEST ACCOMMODATIONS .....	29
IMPACT ANALYSIS.....	29
<b>PART 5: CONSTRUCT VALIDITY .....</b>	<b>64</b>
MINIMIZATION OF CONSTRUCT-IRRELEVANT VARIANCE AND UNDER-REPRESENTATION.....	64
RELIABILITY .....	65
<i>Test Reliability</i> .....	65
<i>Standard Error of Measurement</i> .....	66
<i>Decision Accuracy and Consistency</i> .....	66
CONVERGENT VALIDITY.....	68
<i>Principal Components Analysis</i> .....	68
<i>Analyses by Content Standard</i> .....	69
DIVERGENT (DISCRIMINANT) VALIDITY .....	70
<b>PART 6: SCALING AND LINKING.....</b>	<b>102</b>
ITEM RESPONSE THEORY .....	102
<i>Model Fit</i> .....	103

SCALING .....	105
<i>Linking Method and Results</i> .....	106
<i>Vertical Scale</i> .....	106
LOWEST AND HIGHEST OBTAINABLE SCALE SCORES .....	107
<b>PART 7: TEST AND ITEM STATISTICS .....</b>	<b>132</b>
TEST-LEVEL STATISTICS .....	132
ITEM-LEVEL STATISTICS .....	132
<b>PART 8. FAIRNESS.....</b>	<b>148</b>
MINIMIZING BIAS THROUGH CAREFUL TEST DEVELOPMENT .....	148
EVALUATING BIAS THROUGH DIFFERENTIAL ITEM FUNCTIONING STATISTICS.....	149
<b>PART 9. CONSTRUCTED-RESPONSE SCORING .....</b>	<b>156</b>
MAP SCORING PROCESS .....	156
<i>Handscoring Process Used for MAP</i> .....	156
INTER-RATER RELIABILITY .....	159
<b>REFERENCES .....</b>	<b>164</b>

## Table of Tables

TABLE 1. 1: PARTICIPATION RATES, ALL STUDENTS .....	3
TABLE 1. 2: PERCENTAGE OF STUDENTS CLASSIFIED AS <i>PROFICIENT</i> OR <i>ADVANCED</i> IN 2006 AND 2007 USING CENSUS DATA, COMMUNICATION ARTS AND MATHEMATICS .....	3
TABLE 3. 1: CONTENT COVERAGE, MAP 2007 ITEM MAPS .....	13
TABLE 3. 2: MAP 2007 CONTENT STANDARD ITEM/POINT DISTRIBUTIONS, COMMUNICATION ARTS .....	21
TABLE 3. 3: MAP 2007 CONTENT STANDARD ITEM/POINT DISTRIBUTIONS, MATHEMATICS .....	22
TABLE 3. 4: MAP 2007 NUMBER OF ITEMS/POINTS MEASURING PROCESS STANDARDS, COMMUNICATION ARTS .....	23
TABLE 3. 5: MAP 2007 NUMBER OF ITEMS/POINTS MEASURING PROCESS STANDARDS, MATHEMATICS .....	25
TABLE 4. 1: NUMBER OF SCHOOLS WHO PARTICIPATED IN THE CALIBRATION SAMPLE .....	31
TABLE 4. 2: SUMMARY OF CALIBRATION AND CENSUS DATA, COMMUNICATION ARTS .....	32
TABLE 4. 3: SUMMARY OF CALIBRATION AND CENSUS DATA, MATHEMATICS .....	35
TABLE 4. 4: COMPARISON OF STATE-LEVEL MEANS, 2006 AND 2007 CENSUS DATA .....	37
TABLE 4. 5: COMPARISON OF PERCENT OF STUDENTS IN EACH ACHIEVEMENT LEVEL USING 2006 AND 2007 CENSUS DATA, COMMUNICATION ARTS.....	38
TABLE 4. 6: COMPARISON OF PERCENT OF STUDENTS IN EACH ACHIEVEMENT LEVEL USING 2006 AND 2007 CENSUS DATA, MATHEMATICS .....	39
TABLE 4. 7: PARTICIPATION RATES, ALL STUDENTS .....	39
TABLE 4. 8: PARTICIPATION RATES, MALES.....	40
TABLE 4. 9: PARTICIPATION RATES, FEMALES .....	40
TABLE 4. 10: PARTICIPATION RATES, WHITE .....	40
TABLE 4. 11: PARTICIPATION RATES, BLACK.....	41
TABLE 4. 12: PARTICIPATION RATES, HISPANIC .....	41
TABLE 4. 13: PARTICIPATION RATES, ASIAN/PACIFIC ISLANDER .....	41
TABLE 4. 14: PARTICIPATION RATES, NATIVE AMERICAN/ALASKAN.....	42
TABLE 4. 15: PARTICIPATION RATES, STUDENTS RECEIVING FREE OR REDUCED-PRICE LUNCH (SES) .....	42
TABLE 4. 16: PARTICIPATION RATES, MIGRANT STUDENTS .....	42
TABLE 4. 17: PARTICIPATION RATES, SPECIAL EDUCATION (IEP) STUDENTS .....	43
TABLE 4. 18: PARTICIPATION RATES, ENGLISH LEARNERS (LEP, ELL).....	43
TABLE 4. 19: PARTICIPATION RATES, STUDENTS WITH DISABILITIES.....	43
TABLE 4. 20: PARTICIPATION RATES, STUDENTS RECEIVING ACCOMMODATIONS .....	44
TABLE 4. 21: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, REGULAR EDITION.....	45
TABLE 4. 22: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, BRAILLE EDITION .....	49
TABLE 4. 23: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, LARGE PRINT EDITION.....	53
TABLE 4. 24: IMPACT ANALYSIS, GRADE 3 COMMUNICATION ARTS .....	57
TABLE 4. 25: IMPACT ANALYSIS, GRADE 4 COMMUNICATION ARTS .....	57
TABLE 4. 26: IMPACT ANALYSIS, GRADE 5 COMMUNICATION ARTS .....	58
TABLE 4. 27: IMPACT ANALYSIS, GRADE 6 COMMUNICATION ARTS .....	58
TABLE 4. 28: IMPACT ANALYSIS, GRADE 7 COMMUNICATION ARTS .....	59
TABLE 4. 29: IMPACT ANALYSIS, GRADE 8 COMMUNICATION ARTS .....	59
TABLE 4. 30: IMPACT ANALYSIS, GRADE 11 COMMUNICATION ARTS .....	60
TABLE 4. 31: IMPACT ANALYSIS, GRADE 3 MATHEMATICS.....	60
TABLE 4. 32: IMPACT ANALYSIS, GRADE 4 MATHEMATICS.....	61
TABLE 4. 33: IMPACT ANALYSIS, GRADE 5 MATHEMATICS.....	61
TABLE 4. 34: IMPACT ANALYSIS, GRADE 6 MATHEMATICS.....	62
TABLE 4. 35: IMPACT ANALYSIS, GRADE 7 MATHEMATICS.....	62
TABLE 4. 36: IMPACT ANALYSIS, GRADE 8 MATHEMATICS.....	63
TABLE 4. 37: IMPACT ANALYSIS, GRADE 10 MATHEMATICS.....	63

TABLE 5. 1: RELIABILITY, COMMUNICATION ARTS .....	70
TABLE 5. 2: RELIABILITY, MATHEMATICS.....	72
TABLE 5. 3: CRONBACH’S ALPHA BY SUBGROUP, COMMUNICATION ARTS.....	72
TABLE 5. 4: CRONBACH’S ALPHA BY SUBGROUP, MATHEMATICS .....	73
TABLE 5. 5: DECISION ACCURACY AND CONSISTENCY CONDITIONED ON LEVEL OF ACHIEVEMENT, COMMUNICATION ARTS AND MATHEMATICS .....	73
TABLE 5. 6: DECISION ACCURACY AND CONSISTENCY AT ACHIEVEMENT CUT POINTS, COMMUNICATION ARTS AND MATHEMATICS .....	74
TABLE 5. 7: PRINCIPAL COMPONENT ANALYSIS, COMMUNICATION ARTS .....	75
TABLE 5. 8: PRINCIPAL COMPONENT ANALYSIS, MATHEMATICS .....	76
TABLE 5. 9: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATIONS AMONG CONTENT STANDARDS, GRADE 3 COMMUNICATION ARTS .....	77
TABLE 5. 10: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 4 COMMUNICATION ARTS .....	77
TABLE 5. 11: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 5 COMMUNICATION ARTS .....	78
TABLE 5. 12: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 6 COMMUNICATION ARTS .....	78
TABLE 5. 13: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 7 COMMUNICATION ARTS .....	78
TABLE 5. 14: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 8 COMMUNICATION ARTS .....	78
TABLE 5. 15: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 11 COMMUNICATION ARTS .....	79
TABLE 5. 16: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 3 MATHEMATICS .....	79
TABLE 5. 17: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 4 MATHEMATICS .....	79
TABLE 5. 18: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 5 MATHEMATICS .....	79
TABLE 5. 19: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 6 MATHEMATICS .....	80
TABLE 5. 20: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 7 MATHEMATICS .....	80
TABLE 5. 21: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 8 MATHEMATICS .....	80
TABLE 5. 22: RELIABILITY OF EACH CONTENT STANDARD AND UNCORRECTED AND CORRECTED CORRELATION AMONG CONTENT STANDARDS, GRADE 10 MATHEMATICS .....	80
TABLE 5. 23: INTER-CORRELATION OF SCALE SCORES, COMMUNICATION ARTS AND MATHEMATICS.....	81
TABLE 6. 1: ITEM FIT STATISTICS FOR MISFITTING ITEMS, COMMUNICATION ARTS (CA) AND MATHEMATICS (MA) .....	108
TABLE 6. 2: LOSS AND HOSS VALUES, COMMUNICATION ARTS AND MATHEMATICS.....	108
TABLE 7. 1: MAP MEANS, STANDARD DEVIATIONS FOR RAW SCORES, P-VALUES, ITEM-TOTAL CORRELATION ( $R_{IT}$ ): COMMUNICATION ARTS.....	133
TABLE 7. 2: MAP MEANS, STANDARD DEVIATIONS FOR RAW SCORES, P-VALUES, ITEM-TOTAL CORRELATION ( $R_{IT}$ ): MATHEMATICS .....	133
TABLE 7. 3: ITEM STATISTICS, GRADE 3 COMMUNICATION ARTS AND MATHEMATICS.....	134
TABLE 7. 4: ITEM STATISTICS, GRADE 4 COMMUNICATION ARTS AND MATHEMATICS.....	136
TABLE 7. 5: ITEM STATISTICS, GRADE 5 COMMUNICATION ARTS AND MATHEMATICS.....	138
TABLE 7. 6: ITEM STATISTICS, GRADE 6 COMMUNICATION ARTS AND MATHEMATICS.....	140
TABLE 7. 7: ITEM STATISTICS, GRADE 7 COMMUNICATION ARTS AND MATHEMATICS.....	142
TABLE 7. 8: ITEM STATISTICS, GRADE 8 COMMUNICATION ARTS AND MATHEMATICS.....	144
TABLE 7. 9: ITEM STATISTICS, GRADE 11 COMMUNICATION ARTS AND GRADE 10 MATHEMATICS .....	146

TABLE 8. 1: NUMBER OF FLAGGED ITEMS FOR DIF USING SMD STATISTICS, COMMUNICATION ARTS.....	152
TABLE 8. 2: NUMBER OF FLAGGED ITEMS FOR DIF USING SMD STATISTICS, MATHEMATICS.....	154
TABLE 9. 1: INTER-RATER RELIABILITY, COMMUNICATION ARTS.....	160
TABLE 9. 2: INTER-RATER RELIABILITY, MATHEMATICS.....	162

## Table of Figures

FIGURE 5. 1: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 3 COMMUNICATION ARTS.....	81
FIGURE 5. 2: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 4 COMMUNICATION ARTS.....	81
FIGURE 5. 3: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 5 COMMUNICATION ARTS.....	82
FIGURE 5. 4: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 6 COMMUNICATION ARTS.....	82
FIGURE 5. 5: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 7 COMMUNICATION ARTS.....	83
FIGURE 5. 6: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 8 COMMUNICATION ARTS.....	83
FIGURE 5. 7: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 11 COMMUNICATION ARTS.....	84
FIGURE 5. 8: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 3 MATHEMATICS .....	84
FIGURE 5. 9: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 4 MATHEMATICS .....	85
FIGURE 5. 10: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 5 MATHEMATICS .....	85
FIGURE 5. 11: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 6 MATHEMATICS .....	86
FIGURE 5. 12: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 7 MATHEMATICS .....	86
FIGURE 5. 13: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 8 MATHEMATICS .....	87
FIGURE 5. 14: STANDARD ERROR OF MEASUREMENT PLOT, GRADE 10 MATHEMATICS .....	87
FIGURE 5. 15: SCREE PLOT, GRADE 3 COMMUNICATION ARTS.....	88
FIGURE 5. 16: SCREE PLOT, GRADE 4 COMMUNICATION ARTS.....	89
FIGURE 5. 17: SCREE PLOT, GRADE 5 COMMUNICATION ARTS.....	90
FIGURE 5. 18: SCREE PLOT, GRADE 6 COMMUNICATION ARTS.....	91
FIGURE 5. 19: SCREE PLOT, GRADE 7 COMMUNICATION ARTS.....	92
FIGURE 5. 20: SCREE PLOT, GRADE 8 COMMUNICATION ARTS.....	93
FIGURE 5. 21: SCREE PLOT, GRADE 11 COMMUNICATION ARTS.....	94
FIGURE 5. 22: SCREE PLOT, GRADE 3 MATHEMATICS .....	95
FIGURE 5. 23: SCREE PLOT, GRADE 4 MATHEMATICS .....	96
FIGURE 5. 24: SCREE PLOT, GRADE 5 MATHEMATICS .....	97
FIGURE 5. 25: SCREE PLOT, GRADE 6 MATHEMATICS .....	98
FIGURE 5. 26: SCREE PLOT, GRADE 7 MATHEMATICS .....	99
FIGURE 5. 27: SCREE PLOT, GRADE 8 MATHEMATICS .....	100
FIGURE 5. 28: SCREE PLOT, GRADE 10 MATHEMATICS .....	101
FIGURE 6. 1: ITEM CHARACTERISTIC CURVE FOR SESSION 3, ITEM 26 ( $Z = 11.68$ ), GRADE 3 COMMUNICATION ARTS.....	109
FIGURE 6. 2: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 19 ( $Z = 17.65$ ), GRADE 4 COMMUNICATION ARTS.....	109
FIGURE 6. 3: ITEM CHARACTERISTIC CURVE FOR SESSION 1, ITEM 5 ( $Z = 12.61$ ), GRADE 11 COMMUNICATION ARTS.....	110
FIGURE 6. 4: ITEM CHARACTERISTIC CURVE FOR SESSION 3, ITEM 3 ( $Z = 16.55$ ), GRADE 11 COMMUNICATION ARTS.....	110
FIGURE 6. 5: ITEM CHARACTERISTIC CURVE FOR SESSION 1, ITEM 8 ( $Z = 20.52$ ), GRADE 3 MATHEMATICS .....	111
FIGURE 6. 6: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 34 ( $Z = 16.41$ ), GRADE 4 MATHEMATICS .....	111
FIGURE 6. 7: ITEM CHARACTERISTIC CURVE FOR SESSION 1, ITEM 24 ( $Z = 18.05$ ), GRADE 5 MATHEMATICS .....	112
FIGURE 6. 8: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 34 ( $Z = 11.81$ ), GRADE 5 MATHEMATICS .....	112
FIGURE 6. 9: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 22 ( $Z = 18.35$ ), GRADE 7 MATHEMATICS .....	113
FIGURE 6. 10: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 33 ( $Z = 12.92$ ), GRADE 7 MATHEMATICS .....	113
FIGURE 6. 11: ITEM CHARACTERISTIC CURVE FOR SESSION 2, ITEM 35 ( $Z = 17.29$ ), GRADE 7 MATHEMATICS .....	114
FIGURE 6. 12: ITEM CHARACTERISTIC CURVE FOR SESSION 1 ITEM 25 ( $Z = 15.00$ ), GRADE 10 MATHEMATICS .....	114
FIGURE 6. 13: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 3 COMMUNICATION ARTS .....	115
FIGURE 6. 14: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 4 COMMUNICATION ARTS .....	116
FIGURE 6. 15: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 5 COMMUNICATION ARTS .....	117

FIGURE 6. 16: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 6 COMMUNICATION ARTS .....	118
FIGURE 6. 17: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 7 COMMUNICATION ARTS .....	119
FIGURE 6. 18: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 8 COMMUNICATION ARTS .....	120
FIGURE 6. 19: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 11 COMMUNICATION ARTS .....	121
FIGURE 6. 20: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 3 MATHEMATICS .....	122
FIGURE 6. 21: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 4 MATHEMATICS .....	123
FIGURE 6. 22: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 5 MATHEMATICS .....	124
FIGURE 6. 23: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 6 MATHEMATICS .....	125
FIGURE 6. 24: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 7 MATHEMATICS .....	126
FIGURE 6. 25: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 8 MATHEMATICS .....	127
FIGURE 6. 26: TEST CHARACTERISTIC CURVES (TCC) FOR THE INPUTTED ANCHOR ITEMS AND FOR THE ESTIMATED ANCHOR ITEMS, GRADE 10 MATHEMATICS .....	128
FIGURE 6. 27: CROSS-GRADE ARTICULATION OF SCALE SCORES AT SELECTED PERCENTILES, COMMUNICATION ARTS .....	129
FIGURE 6. 28: CROSS-GRADE ARTICULATION OF SCALE SCORES AT SELECTED PERCENTILES, MATHEMATICS .....	130
FIGURE 6. 29: TEST CHARACTERISTIC CURVES (TCC) BY GRADE, COMMUNICATION ARTS .....	131
FIGURE 6. 30: TEST CHARACTERISTIC CURVES (TCC) BY GRADE, MATHEMATICS .....	131

## Part 1: Overview

The 2007 Missouri Assessment Program (MAP) marked the second administration of grade-level MAP in Missouri. The MAP is designed to measure students' knowledge of Communication Arts and Mathematics<sup>1</sup>. This report provides a technical overview of the Communication Arts and Mathematics assessments of the 2007 MAP. As such, it presents evidence for the validity of the 2007 MAP scores.

### *Background*

Historically, the MAP was designed to be a grade-span test: Grades 3, 7, and 11 in Communication Arts and Grades 4, 8, and 10 in Mathematics. In 2003, the Department of Elementary and Secondary Education (DESE) contracted with CTB/McGraw-Hill to expand the testing program to grade-level testing for Communication Arts and Mathematics. In the spring of 2005, Missouri administered a field test, which was the basis for the construction of the 2006 and 2007 operational test forms.

The MAP was originally designed to measure Missouri's Show-Me Standards. These standards were adopted by the Missouri State Board of Education in 1996. Since their inception, Missouri's Show-Me Standards have been further refined to better delineate Content Standards, Process Standards, and Content Strands/Grade-Level Expectations as Missouri moved from grade-span to grade-level tests to comply with the requirements of No Child Left Behind. The MAP tests have therefore undergone multiple alignment analyses to ensure that MAP content reflects these refinements. Further details of the development of the 2007 MAP may be found in Part 3 of this report.

### *Administration*

In the spring of 2007, Missouri administered assessments in Communication Arts to students in Grades 3–8 and 11, and Mathematics to students in Grades 3–8 and 10. The MAP was administered from March 26 to May 4, 2007. Schools participating in the early-return sample administered tests from March 26 to April 19, 2007. The early-return and regular test windows were expanded by one week to accommodate schools whose academic schedules had been affected by winter weather.

Approximately 550 districts administered the Communication Arts and/or Mathematics MAP tests in Grades 3 through 8 and/or high school. Table 1.1 shows the number of students tested in each grade and the percentage of students with reportable test scores based on the census data.<sup>2</sup> Further analysis of participation rates is provided in Part 4 of this report.

---

<sup>1</sup> Science and Social Studies are also assessed as part of the MAP; however, both of these content areas are voluntary assessments that schools and districts choose to administer. This document does not report the results of these assessments.

<sup>2</sup> The census data used in this report does not reflect additional cleaning steps that DESE staff implements once CTB releases data to DESE; therefore, the numbers in this report may differ from numbers that DESE reports using their cleaned data.

### ***Student Performance***

This is the second year of the grade-level MAP testing programs in Communication Arts and Mathematics. Table 1.2 presents the percentage of students classified as *Proficient* or *Advanced* in 2006 and 2007 in both Communication Arts and Mathematics. Small gains in performance were seen in Grades 3 through 8 in both content areas. The largest gains in improvement were seen in Grades 5 and 6 Mathematics. On the high school assessments, there was a small decline in performance on both the Communication Arts and Mathematics tests. More information on student performance may be found in Part 4 of this report.

### ***Validity and Test Scores***

Most sections of this technical report are designed to provide validity evidence to support the use of the MAP test scores. Part 2 discusses the concept of validity and the uses of the MAP scores. Part 3 focuses on the test development process used to create the MAP. The discussions in this section are important to the content-related validity of the MAP scores. Part 4 presents information on student performance. Part 5 discusses reliability and construct-related validity. In this section, the assumption that the content area MAPs measure only one construct is evaluated. For example, the grade-level Mathematics MAP should measure one primary dimension (Mathematics). Part 6 presents the scaling and linking procedures and the results. Part 7 reviews the test- and item-level characteristics. Part 8 overviews the statistical and development processes used to assure fairness of the MAP for all examinees. Part 9 discusses the scoring of constructed-response items, as well as the results of the inter-rater reliability studies. Some analyses in this document are based on the calibration sample while others are based on census data. The source of data used for particular analyses is indicated throughout the Technical Report.

**Table 1. 1: Participation Rates, All Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	67259	98.64	67257	99.08
4	65982	98.93	65975	99.07
5	66098	99.04	66075	99.13
6	67045	98.81	67039	98.95
7	68404	98.19	68405	98.76
8	71200	98.58	71190	98.61
10	-	-	69488	98.33
11	62946	97.67	-	-

**Table 1. 2: Percentage of Students Classified as *Proficient* or *Advanced* in 2006 and 2007 using Census Data, Communication Arts and Mathematics**

Grade	Communication Arts			Mathematics		
	2006	2007	2007-2006	2006	2007	2007-2006
3	43.0	43.2	0.2	43.7	45.4	1.7
4	44.3	45.6	1.3	43.8	44.9	1.1
5	45.5	48.3	2.8	43.8	47.0	3.2
6	42.7	44.1	1.4	44.3	48.3	4.0
7	43.5	45.2	1.7	43.5	45.5	2.0
8	42.1	42.2	0.1	40.4	41.2	0.8
HS	42.5	41.5	-1.0	42.3	40.8	-1.5

## **Part 2: Validity and the Uses of Test Scores**

Validity is the overarching component of the MAP testing program. The following excerpt is from the *Standards for Educational and Psychological Testing* (AERA, 1999) [hereafter referred to as the *Standards*]:

Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees.... (p. 17)

As stated by the *Standards*, the validity of a testing program hinges on the use of the test scores. *Validity evidence* that supports the uses of the MAP test scores is provided in this Technical Report. In this section, we examine some possible uses of the MAP test scores.

The following sections (Parts 3–9) of this Technical Report provide additional evidence for these uses, as well as technical support for some of the interpretations and uses of test scores. The information in Parts 3–9 also provide a firm foundation that the MAP tests measure what they are intended to measure. However, this Technical Report cannot anticipate all possible interpretations and uses of the MAP scores. It is recommended that policy and program evaluation studies, in accordance with the *Standards*, be conducted to support some of the uses of the MAP scores.

### ***Uses of Test Scores***

The validity of a test score ultimately rests in how that test score is used. To understand whether a test score is being used properly, we must first understand the purpose of the test. The purpose of the Communication Arts and Mathematics MAPs is to demonstrate student achievement in these content areas in Missouri. As such, the test scores may be used to classify students, schools, districts, and the state with respect to how much achievement each shows in each content area. Classification is based on the level of student achievement demonstrated on the MAP for each content area.

This Technical Report refers to the use of several kinds of scores: the test-level scores (scale scores and achievement levels), the content standard scores, and the process standard scores.

### ***Test-Level Scores***

At the test level, an overall scale score is reported that is based on student performance on the entire test. In addition, an associated level of achievement is reported. These scores indicate, in varying ways, a student's achievement in Communication Arts or Mathematics. Test-level scores are reported at four reporting levels: the state, the school district, the school, and the student.

In part, the rationale for the claim that the MAP scores are credible individual achievement scores is based on the fact that the MAP was developed with items that are very similar to the questions and activities teachers use to teach their students. In fact, custom-written portions of the MAP were directly authored by Missouri educators, edited by both CTB and Missouri educators, and subsequently reviewed and approved for use by Missouri educators. This procedure fosters a close relationship between the items and the Missouri Show-Me Standards, from which the MAP was developed. Portions of the MAP from CTB's item pool were also aligned to Missouri Content Standards, Process Standards, and Grade-Level Expectations (GLEs) to further solidify the Show-Me Standards as the foundation of the MAP. Item development is described in Part 3; however, detailed descriptions of processes used to delineate the knowledge, skills, abilities, including content limits and descriptions for each content area, are beyond the scope of this report.

At the test level, two types of scores are reported to indicate a student's achievement on the MAP: (1) a scale score and (2) its associated level of achievement.

### **1. Scale Score**

A scale score indicating a student's total performance is determined for each content area on the MAP. The overall scale score for a content area quantifies the achievement being measured by either the Communication Arts or Mathematics test. In other words, the scale score represents the students' level of achievement where higher scale scores indicate higher levels of achievement on the test and lower scale scores indicate the opposite. Thus, the scale score defines achievement operationally.

For this reason, validation of the test scores is focused on gathering contextual evidence that supports the test's construct. Psychometric validation of the operational definition of achievement consists of *prima-facie* evidence. Generally, alignment with content standards provides additional support for this *prima-facie* evidence.

### **2. Level of Achievement**

A student's performance on the Communication Arts or Mathematics MAPs is reported in one of four levels of achievement: *Below Basic*, *Basic*, *Proficient*, or *Advanced*. The cut scores for the levels of achievement were recommended by Missouri educators and citizens at the Bookmark Standard Setting Workshop in December 2005 and reflect their expectations of what Missouri students should know and be able to do in each grade/content area. The Missouri Show-Me Standards guided these recommendations, as did Senate Bill 1080<sup>3</sup>. Thus, MAP achievement levels reflect the achievement standards and abilities intended by the Missouri legislature, Missouri teachers, Missouri citizens, and DESE. Descriptions of each level of achievement in terms of what a student should know and be able to do are provided with the MAP score reports.

---

<sup>3</sup> See *Missouri Assessment Program Final Bookmark Standard Setting Technical Report* (CTB/McGraw-Hill, 2005) or Part 8 of the *Missouri Assessment Program Technical Report 2006* (CTB/McGraw-Hill, 2006) for an explanation of the achievement-level setting process.

## **Use of Test-Level Scores**

The MAP scale scores and achievement levels provide summary evidence of student achievement in Communication Arts or Mathematics. Classroom teachers may use these scores as evidence of student achievement in these content areas. At the aggregate level, district and school administrators may use this information for activities such as planning curriculum. At the state level, the aggregate test-level scale scores are appropriate to use for accountability programs associated with No Child Left Behind and the Missouri School Improvement Program. State officials may also use aggregate results to ascertain the extent to which Missouri schools and districts have aligned their curriculum to the Show-Me Standards. The results presented in this Technical Report provide evidence that the scale scores are a valid and reliable indicator of student performance in Communication Arts and Mathematics.

## ***Content Standard Subscores***

The Content Standard subscores indicate student performance in terms of the number- and percent-correct score for each Content Standard in Communication Arts and each GLE strand in Mathematics. Starting in 2007, CTB reports Content Standard subscores only at the student level. These scores may be aggregated by the state, district, or schools to determine the mean Content Standard subscores. These means may be used as indicators of the performance of the school or district in teaching students the knowledge and skills defined for each subject area.

## **The Use of the Content Standard Subscores**

The purpose of reporting Content Standard subscores on the MAP is to show for each student the relationship between the overall achievement being measured and the skills in each of the areas delimited by the Content Standards in Communication Arts and the GLE strands in Mathematics. The Content Standard subscores provide a measure of the extent to which an overall achievement level has the desired breadth of the Missouri Show-Me Standards (and, by association, the GLEs). Teachers may use these subscores for individual students as indicators of strengths and weaknesses, but they are best corroborated by other evidence, such as homework, class participation, diagnostic test scores, or observation. Part 3 of this Technical Report provides content validity evidence that supports the use of the Content Standard subscores. Part 4 of this Technical Report provides evidence of construct validity that further supports the use of the Content Standard subscores.

District and school administrators may compare their aggregate results with the state mean to better understand their strengths and weaknesses within a content area. Caution should be exercised when comparing Content Standards subscores between students or across years. The user should be aware that different items will comprise the Content Standards across years and that these items may vary in difficulty.

## ***Process Standard Subscores***

For each MAP content area, Process Standard and Content Standard subscores are determined from the same pool of items. These items were classified by the particular underlying processes used to teach each item's content, and each item's assigned Process Standard was verified by Missouri teachers in a Content Review workshop specifically designed to fulfill that purpose.

Content Standard and Process Standard subscores generally show a directly proportional relationship, because the same pool of items is used to measure both sets of standards. Process Standard subscores are only reported at the student level.

### **The Use of the Process Standard Subscores**

The purpose of reporting Process Standard subscores on the MAP is to show the achievement of students in each of the areas delimited by the Process Standards in Communication Arts or Mathematics. The Process Standard subscores provide a measure of the extent to which an overall achievement level has the desired breadth of the Missouri Process Standards. When the Process Standard processes are used to teach the subject area content, the Process Standard scores can be said to reflect the strategies Missouri teachers want Missouri students to adopt in the learning and handling of “real world” activities.

Caution should be exercised when making comparisons of Process Standards subscores between students or across years. The user should be aware that different items will comprise the Process Standards across years and that these items may vary in difficulty.

## **Part 3: Test Content Development**

Content-related validity in achievement tests is evidenced by a correspondence between test content and a specification of the content domain. Content-related validity can be demonstrated through consistent adherence to test blueprints, through a high quality test development process that includes review of items for accessibility to English Language Learners and students with disabilities, and through alignment studies performed by independent groups. In this section, CTB will provide a detailed discussion of the test development cycle from aligning items with Missouri's rigorous Show-Me Standards and GLE strands to selecting items for the final operational test form. In particular, this section will show how CTB's Content Development Team followed rigorous procedures to select tests that reflect the full range of content that the MAP is expected to cover.

### ***Test Design***

Evidence of validity based on test content includes information about the test and item specifications. Test development involves creating a design framework from the statement of the construct to be measured. This design evolves from the tension between the constraints of the assessment program and the benefits sought from the examination of students. Many of the benefits sought are not scientific in nature nor are many of the constraints; rather, they are policy considerations. The design emerges from specifications which are originally set forth and modified as a result of these considerations during the test development process. Design elements include such things as number and types of items/tasks for each of the scores reported (tasks are measured by constructed-response items in the MAP). These design elements are documented with *item maps* for the MAP which show the distribution of items/tasks by Content Standards (Communication Arts) and GLE strands (Mathematics) in the 2007 test forms. The item maps also show the design of the test administration by representing the sessions into which the test is divided (*session* assignments determine which items will be taken together).

The other key aspect of the structural framework of the MAP tests is the number of points awarded for each Content and Process Standard. This design element represents a compromise between many constraints, including the target weights for each Content Standard recommended by Missouri teachers, availability of items from field testing, and results of multiple reviews by content specialists.

### ***Item Development***

Planning and preparation for the development of item content to be used on the 2006 and 2007 Communication Arts and Mathematics MAP Operational Test forms commenced in late 2002. The plan specified an item development and selection cycle that included an initial item writing/passage selection workshop (Spring 2003), a local pilot study (Fall 2003), a content and bias review (Spring 2004), item refinements and form construction (Summer, Fall, Winter 2004), a subsequent round of formal field testing (Spring 2005), the selection of operational forms based on statistical data from the field test (Summer, Fall 2005), a formal standard-setting process (Winter 2005), and ultimately, operational testing (Spring 2006 and 2007) at Grade levels 3 through 8 and high school. Each of these steps is described in greater detail below.

## **Item Writing**

In May 2003, a group comprised of Missouri educators, Regional Facilitators, DESE staff, and CTB personnel participated in an Item Writing Workshop (IWW) for Communication Arts and Math at the Resort at Port Arrowhead, located at Lake Ozark, Missouri, which served as the basis for the custom-written portion of the MAP Operational 2006 and 2007 test forms. The workshop was conducted with participants selected by DESE to represent educational sites throughout Missouri.

*Communication Arts:* During the first part of the workshop, Communication Arts participants selected reading passages. Then, participants used selected passages as a basis for writing constructed-response items and writing prompts in preparation for the 2006 and 2007 Operational forms for Grades 3–8 and 11. The items were written over the course of several days with the participation of over 30 Missouri teachers, and the content developed at the workshop was based specifically upon the Missouri Show-Me Standards for Communication Arts. Communication Arts items were refined after the initial IWW, which led to the production of Communication Arts test forms used for the local pilot studies.

*Mathematics:* The Mathematics participants wrote constructed-response and performance-event items along with scoring guides to create a pool of items for the 2006 and 2007 Operational forms for Grades 3–8 and 10. The items were written over the course of several days with the participation of over 30 Missouri teachers, and the content developed at the workshop was based specifically upon the Missouri curriculum/Show-Me Standards. Mathematics items were refined after the initial IWW via collaboration between DESE and CTB. Some items from the workshop were considered to be unusable, so additional selected-response items were developed by CTB to help supplement the item pool (particularly selected-response items) and reviewed by DESE.

Overall, the item writing workshop in May 2003 provided a basis upon which items written for the Communication Arts and Mathematics assessment could be selected for use on small-scale pilot studies administered throughout Missouri.

## **Local Pilot Test**

These items were used to produce Communication Arts and Mathematics test forms for local pilot studies. The small-scale pilot was administered in October/November of 2003 for students at Grades 3–8 and high school in a limited number of classrooms throughout Missouri.

Six Communication Arts forms per grade were piloted, consisting of approximately two selected-response and six constructed-response items for each of Grades 4, 5, 6, and 8. The six Communication Arts pilot forms for Grades 3, 7, and 11 contained two selected-response items, four constructed-response items, and one writing prompt.

Six Mathematics forms per grade were piloted, consisting of approximately twelve selected-response and two constructed-response items for each of Grades 3, 5, 6, and 7. The six

Mathematics pilot forms for Grades 4, 8, and 10 contained twelve selected-response items, four constructed-response items, and one performance event.

In November 2003, the results of the pilot studies underwent further evaluation during the next step of the item development process: the Score, Revise, and Rewrite (SRR) Workshop.

### **Score, Revise, Rewrite Workshop**

The purpose of the SRR Workshop was for the participants to score the Communication Arts or Mathematics items piloted in classrooms in Missouri in October/November of 2003, and to revise the items and/or rubrics/scoring guides based on the scoring process, student results, and subsequent discussion. DESE invited approximately 5 to 7 participants per grade, resulting in the direct participation of over 40 Missouri educators. CTB and DESE personnel were present to facilitate the SRR Workshop. The participants individually scored the students' pilot forms, tallied the results, and then reviewed the items as a group. The Regional Facilitators were also present and participated in the process. Overall, the goal of the workshop was to improve the item quality prior to the Content and Bias Review (CBR) and to ensure that quality items were developed for future use in the Missouri Assessment Program. Most participants commented that this workshop was successful in this regard.

### **Content and Bias Review Workshop**

The CBR workshop was conducted in March 2004, involving DESE and CTB staff, Missouri educators, and Regional Facilitators. The workshop was held at the Inn at Grand Glaize at Lake Ozark, Missouri. For the Content Review, DESE invited participants from educational sites throughout Missouri to review the Communication Arts or Mathematics items and scoring guides for content accuracy and grade level appropriateness. In Communication Arts, participants also reviewed passages. In addition, participants in both content area groups verified each item's alignment to the Missouri curriculum by reviewing the Content Standard, Process Standard, and GLE assignment for each item at the review. In each content area, over 30 Missouri educators participated in the process and helped to realign and revise the items. The Content Review was accomplished over the course of two days and was followed by a Bias Review on the following day.

The Bias review committee was comprised of representatives from various backgrounds whose purpose was to screen the items for any racial, socioeconomic, gender, or other sensitivity issues. This committee could revise or reject items because of issues related to possible bias. Only four Communication Arts items were rejected and no Mathematics items were rejected from their respective pools. The remaining items were either accepted or accepted with revisions.

Following the CBR, 257 items remained in the Communication Arts item pool and 635 items remained in the Mathematics item pool. These items became candidates for the MAP field test.

### **Field-test Selection and Administration**

During the remaining months of 2004, the Communication Arts and Mathematics item pools were used as the basis for the formation of four stand-alone Field Test forms in each content

area. The custom-written material was arranged into test forms using *TerraNova* Survey as the anchor item set (this would also represent the NRT portion of the test). Items were selected and placed into forms so that the combined coverage of the NRT and customized portions of the test met the established blueprint requirements for content coverage; each field test form was constructed using the same design.

As items underwent additional review between DESE and CTB, additional refinements were made to the custom-written items to compensate for issues unforeseen during the previous reviews. Ultimately, four field-test forms in each content area were produced as printed test books and field tested.

The MAP Spring 2005 Mathematics Field Test was successfully administered to Grades 3–8 and high school in May of 2005. The results of the field test generated item statistics that were used to help select two years of parallel operational forms, to be administered in 2006 and 2007.

Statistics measuring differential item function (DIF, see Part 8 of this report) were generated on field-test data. Items flagged for DIF could be included in the operational test selection after they were further reviewed by CTB and DESE staff. During this review, CTB and DESE staff analyzed DIF items for possible content-based causes of the DIF flag. If a content-based reason was not found, then the item could be included in the operational selection.

### **Operational-test Selection**

The use of a *TerraNova* portion for the MAP tests was specified from the earliest stages of development, and the use of the *TerraNova* Survey and its match/alignment to the Missouri standards played an important role in planning for the entire development process leading up to the time of item selection. This is because the test blueprint is applied to the entire test, which includes both the norm-referenced (NRT) and custom portions. As an NRT product, *TerraNova* items are pre-classified to an existing set of *TerraNova* Reading, Language, or Mathematics standards.<sup>4</sup> In many cases, the alignment of *TerraNova* items to Missouri standards could be considered equivalent; nevertheless, for the 2003–2005 development cycle, part of the item development process provided for a DESE review of how the items in the *TerraNova* Survey were matched to the Missouri standards. This match/alignment of *TerraNova* items to Missouri standards took a less prominent role leading up to the Field Test, perhaps in part due to the fact that Field Test scores do not generate the score reports common to Operational Tests, but the matter drew increased attention as planning for the 2006 and 2007 Operational test selections was underway.

When designing the Communication Arts MAP, a very small number of Language *TerraNova* items were identified by DESE as being properly aligned to the Communication Arts Show-Me standards. While items from both the *TerraNova* Reading and Language Arts tests were used for the Communication Arts MAP, only NRT scores from the *TerraNova* Reading test were reported.

---

<sup>4</sup> It's important to note that the Communication Arts MAP is comprised of both Language and Reading items that are scaled together. In the *TerraNova* family of tests, Language and Reading are administered in a single booklet but are scaled separately.

Just prior to initiating the item selection submission/approval process for Mathematics in Fall of 2005, new perspectives arose with regard to the alignment of *TerraNova* items to the Missouri standards; in a number of cases, the existing alignments of items were reconsidered by DESE and determined to be aligned to *TerraNova* Mathematics standards in a different way than similar custom-written Missouri items might be aligned to Missouri standards. As a result, CTB found that if the affected items were realigned, in some instances there were not enough custom-written items in the developed pool to remain in compliance with the established blueprint(s). This necessitated the use of a limited number of items (five or fewer per grade level) from other *TerraNova* editions as a means of filling coverage gaps. Once the modified design was in place and the pool of items available for selection was appended, item selection tasks continued.

Item selections were performed by CTB and provided to DESE for approval. The final selections were done in compliance with strict statistical criteria for the MAP, as required by CTB's Research department, and approved for operational use by DESE based on their adherence to both content requirements and statistical criteria.

Upon receipt of approved item selections, production of the resulting operational test forms commenced. Items were ordered and placed into test books in preparation for operational testing, and the standard process of page reviews between CTB and DESE ensued until final approvals were in place. Then, test books and ancillary materials were printed and distributed in support of the Spring 2007 Operational Test, which was administered in March/April of 2007.

**Table 3. 1: Content Coverage, MAP 2007 Item Maps**

**Communication Arts Grade 3**

Content Standard		CR Item #	SR Item #	CR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 3)	(Session 3)
1	Speaking/Writing Standard English					11, 12, 13, 20, 21, 31, 32, 33, 42, 44, 45, 46, 47, 48, 49
2	Reading – Fiction/Poetry/ Drama				22, 23, 43	7, 8, 9, 10, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 40, 41,
3	Reading – Nonfiction	3, 4, 5, 6A	1, 2		22, 23, 24	1, 2, 3, 4, 5, 6, 14, 15, 16, 17, 18, 19,
4	Writing Formally & Informally	6B		1WP		
5	Combined Reading from Standards 2 & 3	3, 4, 5, 6A	1, 2		22, 23, 24, 43	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 14, 15, 16, 17, 18, 19, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 40, 41,

**Communication Arts Grade 4**

Content Standard		CR Item #	SR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Speaking/Writing Standard English				15, 16, 35, 36, 44, 45, 46, 47, 48, 49
2	Reading – Fiction/Poetry/ Drama			19, 20, 21, 43	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34
3	Reading – Nonfiction	3, 4, 5, 6A	1, 2		37, 38, 39, 40, 41, 42
4	Writing Formally & Informally	6B			
5	Combined Reading from Standards 2 & 3	3, 4, 5, 6A	1, 2	19, 20, 21, 43	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 37, 38, 39, 40, 41, 42

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Communication Arts Grade 5**

Content Standard		CR Item #	SR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Speaking/Writing Standard English				24, 25, 26, 27, 28, 29, 44, 45, 46, 47, 48, 49
2	Reading – Fiction/Poetry/ Drama	3, 4A, 5, 6	1, 2	17, 18, 43A	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
3	Reading – Nonfiction			42, 43A	19, 20, 21, 22, 23, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41
4	Writing Formally & Informally	4B			
5	Combined Reading from Standards 2 & 3	3, 4A, 5, 6	1, 2	17, 18, 42, 43A	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41

**Communication Arts Grade 6**

Content Standard		CR Item #	SR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Speaking/Writing Standard English				12, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49
2	Reading – Fiction/Poetry/ Drama	3, 4, 5A, 6	1, 2	17, 19, 43	1, 2, 3, 4, 5, 13, 14, 15, 16, 20, 21, 22, 23, 24, 30, 31, 32, 33, 34, 35, 36
3	Reading – Nonfiction			18	6, 7, 8, 9, 10, 11, 25, 26, 27, 28, 29
4	Writing Formally & Informally	5B			
5	Combined Reading from Standards 2 & 3	3, 4, 5A, 6	1, 2	17, 18, 19, 43	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Communication Arts Grade 7**

Content Standard		CR Item #	SR Item #	CR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 3)	(Session 3)
1	Speaking/Writing Standard English					26, 27, 28, 29, 30, 31, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52
2	Reading – Fiction/Poetry/ Drama				19	1, 2, 3, 4, 5, 6, 7, 15, 16, 17, 18
3	Reading – Nonfiction	3, 4A, 5A, 6	1, 2		20, 42	8, 9, 10, 11, 12, 13, 14, 21, 22, 23, 24, 25, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41
4	Writing Formally & Informally	4B, 5B		1WP		
5	Combined Reading from Standards 2 & 3	3, 4A, 5A, 6	1, 2		19, 20, 42	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41

**Communication Arts Grade 8**

Content Standard		CR Item #	SR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Speaking/Writing Standard English				32, 33, 34, 35, 36, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53
2	Reading – Fiction/Poetry/ Drama			42, 43	22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 37, 38, 39, 40, 41
3	Reading – Nonfiction	3, 4, 5, 6A	1, 2	20, 21	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
4	Writing Formally & Informally	6B			
5	Combined Reading from Standards 2 & 3	3, 4, 5, 6A	1, 2	20, 21, 42, 43	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 37, 38, 39, 40, 41

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Communication Arts Grade 11**

Content Standard		CR Item #	SR Item #	CR Item #	TN Reading NRT CR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 3)	(Session 3)
<b>1</b>	Speaking/Writing Standard English					19, 20, 28, 29, 30, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52
<b>2</b>	Reading – Fiction/Poetry/ Drama				42	23, 24, 25, 26, 27, 31, 32, 33, 34
<b>3</b>	Reading – Nonfiction	3, 4, 5, 6A	1, 2		21, 22	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 35, 36, 37, 38, 39, 40, 41
<b>4</b>	Writing Formally & Informally	6B		1WP		
<b>5</b>	Combined Reading from Standards 2 & 3	3, 4, 5, 6A	1, 2		21, 22, 42	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 23, 24, 25, 26, 27, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Mathematics Grade 3**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations		5	33	1, 2, 3, 4, 5, 6, 8, 9, 11, 13, 18, 19, 21, 22, 30
2	Algebraic Relationships	8	9, 13, 20, 22, 23		12, 15, 20, 29
3	Geometric and Spatial Relationships		7, 12, 14, 16, 17, 18, 19, 25, 26		16, 17, 23
4	Measurement	15	1, 2, 6, 10, 11	32	7, 10, 14, 28
5	Data and Probability	24, 27	3, 4, 21	31	24, 25, 26, 27

**Mathematics Grade 4**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations		13, 19	34	1, 2, 3, 4, 5, 6, 9, 10, 13, 15, 17, 18, 26, 27, 32
2	Algebraic Relationships	25, 28, 30	6, 10, 24	35	8, 16, 25
3	Geometric and Spatial Relationships	5, 21	1, 3, 7, 8, 9, 11, 17, 23		19, 20, 24, 28
4	Measurement	15	2, 18, 22, 27	33	7, 21, 22, 23, 29, 31
5	Data and Probability	12	4, 14, 16, 20, 26, 29		11, 12, 14, 30

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Mathematics Grade 5**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations				1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 14, 19, 21, 28, 31
2	Algebraic Relationships	5	1, 10, 12, 18, 23, 27	34	5, 15, 27, 30
3	Geometric and Spatial Relationships	17	2, 6, 13, 16, 21, 25	35	13, 16, 25
4	Measurement	24	4, 7, 9, 14, 15, 19, 22		17, 22, 23, 24, 26
5	Data and Probability	8	3, 11, 20, 26	33	12, 18, 20, 29, 32

**Mathematics Grade 6**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations				1, 2, 3, 4, 5, 6, 7, 8, 9, 14, 15, 16, 18, 19, 20, 23, 24, 27
2	Algebraic Relationships		2, 4, 5, 6, 12, 13, 17, 20, 25	33	13, 17, 26
3	Geometric and Spatial Relationships	14	1, 10, 11, 21, 23, 26	32	22, 28
4	Measurement		3, 16, 22, 27		21, 25, 30, 31
5	Data and Probability	7, 19, 24	8, 9, 15, 18	34	10, 11, 12, 29

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Mathematics Grade 7**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations				1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 18, 19, 24, 28, 32
2	Algebraic Relationships		4, 5, 10, 15, 17	33	20, 25, 27, 29
3	Geometric and Spatial Relationships	7, 21	1, 6, 11, 12, 19		16, 21, 26, 30
4	Measurement		3, 9, 13, 23, 24, 25, 26	35	22, 23
5	Data and Probability	14, 27	2, 8, 16, 18, 20, 22	34	12, 17, 31

**Mathematics Grade 8**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations				1, 2, 3, 4, 5, 6, 7, 12, 15, 16, 18, 19, 20, 24, 25, 26, 29
2	Algebraic Relationships	12, 17	1, 9, 11, 13, 16, 20, 21, 24, 25, 29	32	10, 21, 30, 31
3	Geometric and Spatial Relationships	7, 27	5, 10, 15, 23	33	11, 14, 22
4	Measurement		2, 3, 6, 14, 19, 28	34	17, 27, 28
5	Data and Probability	4, 22, 30	8, 18, 26		8, 9, 13, 23

**Table 3. 1: Content Coverage, MAP 2007 Item Maps (Continued)**

**Mathematics Grade 10**

Content Strand/GLE		CR Item #	SR Item #	CR Item #	TN NRT Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 2)
1	Number and Operations				1, 2, 3, 4, 9, 11, 13, 14, 15, 19, 23, 24
2	Algebraic Relationships	10, 29	1, 7, 9, 11, 14, 16, 19, 22, 23, 24, 26	26, 29	10, 12, 17, 18
3	Geometric and Spatial Relationships	15, 32	3, 4, 8, 12, 28		6, 20, 21, 22
4	Measurement	5, 25	2, 13, 17, 31	27	16
5	Data and Probability	20	6, 18, 21, 27, 30	28	5, 7, 8, 25

**Table 3. 2: MAP 2007 Content Standard Item/Point Distributions, Communication Arts**

<b>Grade</b>	<b>Content Standard</b>	<b>TN NRT Items</b>	<b>CR/PE Items</b>	<b>SR Items</b>	<b>Total Items</b>	<b>TN Points</b>	<b>CR/PE Points</b>	<b>SR Points</b>	<b>Total Points</b>	<b>% of Total Points</b>
<b>3</b>	Speaking/writing standard English	0	0	15	15	0	0	15	15	22
	Reading fiction/poetry/drama	18	3	0	21	18	7	0	25	36
	Reading nonfiction	12	7	2	21	12	14	2	28	41
	Writing formally & informally	0	2	0	2	0	6	0	6	9
	Combined Reading from Standards 2 & 3	30	8	2	40	30	16	2	48	70
	Total	30	10	17	57	30	22	17	69	100
<b>4</b>	Speaking/writing standard English	0	0	10	10	0	0	10	10	15
	Reading fiction/poetry/drama	29	4	0	33	29	9	0	38	58
	Reading nonfiction	6	4	2	12	6	8	2	16	24
	Writing formally & informally	0	1	0	1	0	2	0	2	3
	Combined Reading from Standards 2 & 3	35	8	2	45	35	17	2	54	82
	Total	35	9	12	56	35	19	12	66	100
<b>5</b>	Speaking/writing standard English	0	0	12	12	0	0	12	12	18
	Reading fiction/poetry/drama	16	7	2	25	16	14	2	32	48
	Reading nonfiction	16	2	1	19	16	5	1	22	33
	Writing formally & informally	0	1	0	1	0	2	0	2	3
	Combined Reading from Standards 2 & 3	32	8	3	43	32	17	3	52	79
	Total	32	9	15	56	32	19	15	66	100
<b>6</b>	Speaking/writing standard English	0	0	13	13	0	0	13	13	20
	Reading fiction/poetry/drama	21	7	2	30	21	15	2	38	58
	Reading nonfiction	10	1	1	12	10	2	1	13	20
	Writing formally & informally	0	1	0	1	0	1	0	1	2
	Combined Reading from Standards 2 & 3	31	8	3	42	31	17	3	51	78
	Total	31	9	16	56	31	18	16	65	100
<b>7</b>	Speaking/writing standard English	0	0	16	16	0	0	16	16	22
	Reading fiction/poetry/drama	11	1	0	12	11	2	0	13	18
	Reading nonfiction	22	6	2	30	22	13	2	37	51
	Writing formally & informally	0	3	0	3	0	7	0	7	10
	Combined Reading from Standards 2 & 3	33	7	2	42	33	15	2	50	68
	Total	33	10	18	61	33	22	18	73	100
<b>8</b>	Speaking/writing standard English	0	0	15	15			15	15	22
	Reading fiction/poetry/drama	15	2	0	17	15	4		19	28
	Reading nonfiction	19	6	2	27	19	13	2	34	49
	Writing formally & informally	0	1	0	1		1		1	1
	Combined Reading from Standards 2 & 3	34	8	2	44	34	17	2	53	77
	Total	34	9	17	60	34	18	17	69	100
<b>11</b>	Speaking/writing standard English	0	0	15	15			15	15	21
	Reading fiction/poetry/drama	9	1		10	9	3		12	16
	Reading nonfiction	25	6	2	33	25	13	2	40	55
	Writing formally & informally		2		2		6		6	8
	Combined Reading from Standards 2 & 3	34	7	2	43	34	16	2	52	71
	Total	34	9	17	60	34	22	17	73	100

**Table 3. 3: MAP 2007 Content Standard Item/Point Distributions, Mathematics**

<b>Grade</b>	<b>Content Standard</b>	<b>TN NRT Items</b>	<b>CR/PE Items</b>	<b>SR Items</b>	<b>Total Items</b>	<b>TN Points</b>	<b>CR/PE Points</b>	<b>SR Points</b>	<b>Total Points</b>	<b>% of Total Points</b>
<b>3</b>	Number and Operations	15	1	1	17	15	2	1	18	27
	Algebraic Relationships	4	1	5	10	4	2	5	11	16
	Geometric and Spatial Relationships	3		9	12	3		9	12	18
	Measurement	4	2	5	11	4	4	5	13	19
	Data and Probability	4	3	3	10	4	6	3	13	19
	Total	30	7	23	60	30	14	23	67	100
<b>4</b>	Number and Operations	15	1	2	18	15	2	2	19	25
	Algebraic Relationships	3	4	3	10	3	10	3	16	21
	Geometric and Spatial Relationships	4	2	8	14	4	4	8	16	21
	Measurement	6	2	4	12	6	4	4	14	18
	Data and Probability	4	1	6	11	4	2	6	12	16
	Total	32	10	23	65	32	22	23	77	100
<b>5</b>	Number and Operations	15	0	0	15	15	0	0	15	21
	Algebraic Relationships	4	2	6	12	4	4	6	14	20
	Geometric and Spatial Relationships	3	2	6	11	3	5	6	14	20
	Measurement	5	1	7	13	5	2	7	14	20
	Data and Probability	5	2	4	11	5	5	4	14	20
	Total	32	7	23	62	32	16	23	71	100
<b>6</b>	Number and Operations	18	0	0	18	18	0	0	18	26
	Algebraic Relationships	3	1	9	13	3	2	9	14	20
	Geometric and Spatial Relationships	2	2	6	10	2	5	6	13	19
	Measurement	4	0	4	8	4	0	4	8	12
	Data and Probability	4	4	4	12	4	8	4	16	23
	Total	31	7	23	61	31	15	23	69	100
<b>7</b>	Number and Operations	19	0	0	19	19	0	0	19	27
	Algebraic Relationships	4	1	5	10	4	2	5	11	15
	Geometric and Spatial Relationships	4	2	5	11	4	4	5	13	18
	Measurement	2	1	7	10	2	3	7	12	17
	Data and Probability	3	3	6	12	3	7	6	16	23
	Total	32	7	23	62	32	16	23	71	100
<b>8</b>	Number and Operations	17	0	0	17	17	0	0	17	22
	Algebraic Relationships	4	3	10	17	4	6	10	20	26
	Geometric and Spatial Relationships	3	3	4	10	3	6	4	13	17
	Measurement	3	1	6	10	3	2	6	11	14
	Data and Probability	4	3	3	10	4	8	3	15	20
	Total	31	10	23	64	31	22	23	76	100
<b>10</b>	Number and Operations	12	0	0	12	12	0	0	12	16
	Algebraic Relationships	4	4	11	19	4	8	11	23	31
	Geometric and Spatial Relationships	4	2	5	11	4	6	5	15	20
	Measurement	1	3	4	8	1	6	4	11	15
	Data and Probability	4	2	5	11	4	5	5	14	19
	Total	25	11	25	61	25	25	25	75	100

Test content evidence of validity is provided for the MAP with the specification of each of the Process Standards that are influential in acquiring the skills tested in the items/tasks used in each of the MAP tests. If teachers teach by the Process Standards as intended, then student performance should improve on those items that were identified as implicitly tapping these habits of mind. The following charts provide the distribution of items and points by Process Standards deemed addressable using the MAP paper-and-pencil items.

**Table 3. 4: MAP 2007 Number of Items/Points Measuring Process Standards, Communication Arts**

<b>Grade Level</b>	<b>Process Standard</b>	<b>NRT Items (SR)</b>	<b>Custom/CR/Other Items</b>	<b>Total Items</b>	<b>NRT Points (SR)</b>	<b>Custom/CR/Other Pts</b>	<b>Total Points</b>
<b>3</b>	1.5	4	2	6	4	4	8
	1.6	13	4	17	13	7	20
	2.1	0	2	2	0	6	6
	2.2	0	15	15	0	15	15
	3.5	13	4	17	13	7	20
<b>4</b>	1.5	7	0	7	7	0	7
	1.6	14	3	17	14	7	21
	2.1	0	1	1	0	2	2
	2.2	0	10	10	0	10	10
	2.4	0	2	2	0	3	3
	3.1	2	0	2	2	0	2
	3.5	12	5	17	12	9	21
<b>5</b>	1.5	2	1	3	2	2	4
	1.6	14	1	15	14	3	17
	1.7	0	1	1	0	1	1
	2.1	0	1	1	0	2	2
	2.2	0	12	12	0	12	12
	2.4	1	1	2	1	1	2
	3.1	1	1	2	1	1	2
	3.5	14	5	19	14	10	24
	3.7	0	1	1	0	2	2
<b>6</b>	1.5	4	0	4	4	0	4
	1.6	13	0	13	13	0	13
	1.8	0	1	1	0	1	1
	2.2	0	13	13	0	13	13
	2.4	0	3	3	0	5	5
	3.1	2	2	4	2	4	6
	3.5	12	6	18	12	11	23
<b>7</b>	1.1	0	1	1	0	2	2
	1.5	1	0	1	1	0	1
	1.6	18	5	23	18	10	28
	1.7	0	1	1	0	1	1
	1.8	0	1	1	0	1	1
	2.1	0	2	2	0	6	6
	2.2	0	16	16	0	16	16
	3.1	1	0	1	1	0	1
	3.5	13	2	15	13	4	17

**Table 3. 4: MAP 2007 Number of Items/Points Measuring Process Standards, Communication Arts  
(Continued)**

<b>Grade Level</b>	<b>Process Standard</b>	<b>NRT Items (SR)</b>	<b>Custom/CR/ Other Items</b>	<b>Total Items</b>	<b>NRT Points (SR)</b>	<b>Custom/CR/ Other Pts</b>	<b>Total Points</b>
<b>8</b>	1.5	6	0	6	6	0	6
	1.6	16	4	20	16	7	23
	1.7	0	1	1	0	3	3
	2.1	0	1	1	0	1	1
	2.2	0	15	15	0	15	15
	2.4	2	0	2	2	0	2
	3.1	0	1	1	0	1	1
	3.5	10	4	14	10	8	18
<b>11</b>	1.5	4	0	4	4	0	4
	1.6	18	3	21	18	6	24
	2.1	0	2	2	0	6	6
	2.2	0	15	15	0	15	15
	2.4	0	1	1	0	1	1
	3.5	12	5	17	12	11	23

**Table 3. 5: MAP 2007 Number of Items/Points Measuring Process Standards, Mathematics**

<b>Grade Level</b>	<b>Process Standard</b>	<b>NRT Items (SR)</b>	<b>Custom/CR/ Other Items</b>	<b>Total Items</b>	<b>NRT Points (SR)</b>	<b>Custom/CR/ Other Pts</b>	<b>Total Points</b>
<b>3</b>	1.1	1	8	9	1	11	12
	1.5	5	7	12	5	8	13
	1.6	7	11	18	7	11	18
	2.1	0	1	1	0	2	2
	3.1	3	3	6	3	5	8
	3.3	12	3	15	12	4	16
	3.5	2	1	3	2	2	4
	4.1	0	1	1	0	2	2
<b>4</b>	1.1	2	0	2	2	0	2
	1.5	3	10	13	3	10	13
	1.6	5	10	15	5	14	19
	1.8	0	1	1	0	4	4
	2.1	0	3	3	0	6	6
	3.1	13	4	17	13	6	19
	3.3	8	7	15	8	12	20
	3.5	2	1	3	2	2	4
	3.6	0	1	1	0	1	1
<b>5</b>	1.1	0	7	7	0	8	8
	1.4	0	1	1	0	1	1
	1.5	7	6	13	7	8	15
	1.6	2	8	10	2	9	11
	1.7	0	1	1	0	1	1
	1.8	0	3	3	0	7	7
	3.1	8	3	11	8	4	12
	3.3	15	2	17	15	2	17
	3.5	0	1	1	0	1	1
	4.1	0	3	3	0	4	4
<b>6</b>	1.1	0	8	8	0	11	11
	1.4	1	0	1	1	0	1
	1.5	3	4	7	3	4	7
	1.6	2	5	7	2	6	8
	1.7	2	1	3	2	1	3
	1.8	0	1	1	0	3	3
	2.1	0	1	1	0	1	1
	3.1	7	0	7	7	0	7
	3.3	13	5	18	13	5	18
	3.4	0	1	1	0	1	1
	3.5	1	3	4	1	4	5
	3.7	2	0	2	2	0	2
	4.1	0	1	1	0	2	2

**Table 3. 5: MAP 2007 Number of Items/Points Measuring Process Standards, Mathematics (Continued)**

<b>Grade Level</b>	<b>Process Standard</b>	<b>NRT Items (SR)</b>	<b>Custom/CR/Other Items</b>	<b>Total Items</b>	<b>NRT Points (SR)</b>	<b>Custom/CR/Other Pts</b>	<b>Total Points</b>
<b>7</b>	1.1	1	2	3	1	3	4
	1.3	0	1	1	0	1	1
	1.4	1	2	3	1	2	3
	1.5	7	6	13	7	7	14
	1.6	0	6	6	0	6	6
	1.8	0	2	2	0	5	5
	3.1	7	2	9	7	4	11
	3.3	14	6	20	14	9	23
	3.4	0	4	4	0	4	4
	3.5	1	2	3	1	2	3
	3.7	1	1	2	1	3	4
4.1	0	1	1	0	2	2	
<b>8</b>	1.1	3	0	3	3	0	3
	1.4	0	1	1	0	1	1
	1.5	4	4	8	4	6	10
	1.6	2	14	16	2	18	20
	1.8	1	1	2	1	1	2
	3.2	0	1	1	0	2	2
	3.3	18	6	24	18	6	24
	3.4	1	1	2	1	1	2
	3.5	2	3	5	2	6	8
	3.7	0	1	1	0	2	2
	4.1	0	1	1	0	2	2
<b>10</b>	1.1	5	3	8	5	8	13
	1.5	2	5	7	2	5	7
	1.6	6	13	19	6	17	23
	1.8	0	1	1	0	2	2
	3.1	3	1	4	3	2	5
	3.3	9	5	14	9	5	14
	3.5	0	4	4	0	4	4
	3.6	0	1	1	0	1	1
	3.7	0	1	1	0	1	1
	4.1	0	2	2	0	5	5

## **Part 4: Student Demographics and Performance**

In this section, information is presented on the demographic composition of the Missouri student population and on the performance of those students on the Communication Arts and Mathematics MAP. Demographic information was collected on all students and includes gender, race/ethnicity, free or reduced-price lunch (SES), migrant status, Individualized Education Program (IEP), limited English proficient students (LEP), students with disabilities, and students with accommodations. This section summarizes the purpose of the calibration sample, reports the demographics and performance of the calibration sample and the census data, summarizes performance on the longitudinal test, reviews the participation rates of students in eight subgroups, summarizes the usage of test accommodations in the 2007 MAP administration, and summarizes impact analyses for various subgroups.

### ***Calibration Sample***

The calibration sample (also referred to as the early-return sample) was selected by DESE to be representative of the state in terms of free and reduced-price lunch, race/ethnicity, and school/district size variables. In general, these samples ranged in size from 3800 to 4700 students per grade and were used for classical item analyses and IRT analyses of the operational data. Table 4.1 shows the number of schools by grade level that were invited to be included in the calibration sample. The schools in this sample were required to complete testing during the first part of the testing window and to return materials ahead of other schools in the state.

The use of a calibration sample is fairly common practice when analyzing data for a state testing program. By using a representative calibration sample, states are able to expedite the return of score reports to the districts.

### ***Calibration, Census Demographics and Test Performance***

Missouri administered MAP assessments in Communication Arts and Mathematics to students in Grades 3 through 8 and high school. The Communication Arts assessment was administered to students in Grade 11; Mathematics was administered to students in Grade 10.

Table 4.2 and 4.3 summarize the number of students in the calibration sample and the overall (census) number of Missouri students in each grade who took the Communication Arts and Mathematics MAPs, respectively. Tables 4.2 and 4.3 also summarize the distribution of students by gender and race/ethnicity. In addition, the mean scale scores and standard deviations for all students and by gender and race/ethnicity are presented in these tables. Looking across Tables 4.2 and 4.3, the calibration sample appears to be representative of the state in terms of demographic composition; in other words, the distributions of students by gender and race/ethnicity are similar between the calibration

and census data. Moreover, the calibration sample is similar to the state in terms of ability both at the aggregate level and when disaggregated by gender and race/ethnicity.

### ***Longitudinal Comparisons***

It is often desirable to examine the scores of students across time. The data in this section compares student performance on the MAP using census data from 2006 and 2007.

Table 4.4 compares the state-level aggregate means. Table 4.4 shows that in most grades there was little change in the mean scale score between 2006 and 2007. The largest change in mean scale score occurred on the Grade 5 and 6 Mathematics MAP, where the mean scale score increased by slightly over 3 scale score points from 2006. A decrease in the mean scale score was observed in Grades 3 and 11 Communication Arts and Grade 10 Mathematics; however, this decrease was less than a scale score point.

Table 4.5 shows the percent of students in each achievement level in 2006 and 2007 on the Communication Arts test. In all grades except high school, the percentage of Proficient and Advanced students was about the same or slightly higher in 2007 than in 2006. There were 1% fewer students classified as Proficient or Advanced in 2007 Grade 11 Communication Arts than there were in 2006. In Grade 5 Communication Arts, nearly 3% more students were classified as Proficient or Advanced in 2007 as compared to 2006.

Table 4.6 shows the percent of students in each achievement level in 2006 and 2007 on the Mathematics test. In all grades except high school, the percentage of Proficient and Advanced students was about the same or slightly higher in 2007 than in 2006. The percentage in Proficient and Advanced in 2007 improved by about 3% in Grade 5 Mathematics and by 4% in Grade 6 Mathematics compared to the 2006 data. There were nearly 1.5% fewer students classified as Proficient or Advanced in 2007, Grade 10 Mathematics data than there were in 2006.

### ***Student Participation***

For the MAP, the following are the major reporting subgroups identified by DESE:

- Gender: Female and Male
- Race and Ethnicity: White, Black, Hispanic, Asian/Pacific Islander, and Native American/Alaskan
- SES: Students receiving free or reduced-price lunch
- Migrant: Migrant students
- IEP: Special education students
- LEP, ELL: English language learners
- Disability: Students with disabilities
- Accommodations: Students receiving testing accommodations

The participation rates of all students and the eight subgroups were computed using the census data. These participation rates are summarized in Tables 4.7 through 4.20. The participation rates were defined as the ratio of the numbers of reportable students to the

numbers of accountable students in each grade/content area. The numbers of accountable students include students who are reportable and students whose scale score in a content area cannot be determined. Students who are not reportable may not have attempted all test sessions or their test may have been invalidated by the teacher.

### ***Test Accommodations***

An Individualized Education Program (IEP) team makes decisions about how students with disabilities will most appropriately participate in assessment programs. These decisions included whether a student would participate in the MAP subject-area assessments (with or without accommodations) or the MAP-Alternate assessment. When making the decision about participation in the MAP subject-area assessments, the IEP team must also consider the student's need for accommodations. If the team decides that the MAP subject-area assessments are not appropriate for an individual student, even with the use of accommodations, then it can determine the student's eligibility for the MAP-Alternate.

The grade-specific *Missouri Assessment Program Examiner's Manual* (2007) contains the list of accommodations permissible for the MAP assessments. If a specific accommodation is not on the list of accommodations in the *Examiner's Manual*, the accommodation may still be permitted. However, for accountability purposes, there are some accommodations that will invalidate a student's test results. All assessment accommodation decisions made by the IEP team must be documented in the IEP. Detailed information regarding testing accommodations can be found at the DESE website:

<http://dese.mo.gov/divimprove/assess/ancillaries.html>

Tables 4.21 through 4.23 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2007 MAP, the Braille Edition of the 2007 MAP, and the Large-Print Edition of the 2007 MAP. For the 2007 MAP, setting and timing accommodations appear to be the most frequently used for both Communication Arts and Mathematics. For the Mathematics MAP, using calculators and having the test read aloud are also among the more frequently used accommodations. On the Braille and Large-Print editions of the MAP, the setting and timing accommodations are again among the most frequently used accommodations. Common accommodations on both the Braille and Large-Print Editions include using a scribe on both Communication Arts and Mathematics MAP, and having the test read aloud and using a calculator on the Mathematics MAP. The analyses in Tables 4.21 through 4.23 are based on census data and include only those students receiving accommodations and who received a scale score on the Communication Arts or Mathematics MAP.

### ***Impact Analysis***

The impact of achievement testing on different subgroups of students can be determined and reported in the form of average scores and also in terms of test score reliability. Tables 4.24 through 4.37 present the scale score means and standard deviations, numbers

of students, effect size (Cohen's  $d$ ) and test form reliability statistics (Coefficient Alpha, see Part 5) for various subgroups of interest.

One way to evaluate the magnitude of the differences between subgroups is to calculate the effect size. Cohen's  $d$  was used to calculate the effect size. Cohen's  $d$  is given by the formula:

$$d = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}}$$

where  $\bar{x}_a$  is the mean score of group A,  $\bar{x}_b$  is the mean score of group b,  $s_a^2$  is the variance of group A,  $s_b^2$  is the variance of group B,  $n_a$  is the number of students in group A, and  $n_b$  is the number of students in group B.

Cohen's  $d$ , then, expresses the difference in group means in terms of the standard deviation. For example if  $d = .34$  for two groups, then it may be interpreted as the mean difference between the two groups is .34 of the pooled standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the  $d$  statistic:  $d = .20$  is a small effect size,  $d = .50$  is a medium effect size, and  $d = .80$  is a large effect size. Even with these guidelines, caution should be used when judging the differences between the groups compared, as there is debate in the measurement field regarding the appropriateness of these guidelines for standardized testing.

Using Cohen's (1988) guidelines, certain trends become apparent in Tables 4.24 through 4.37. On the Communication Arts test in all grades, gender has a small effect on mean test scores where girls outperform boys. In all grades of the Communication Arts and Mathematics tests, IEP, disability, and accommodation tend to have a large effect on the mean test scores, as students in those groups underperform students who are not in those groups. On both assessments in all grades, ELL has a medium (or, in a few cases, large) effect on mean test scores where ELL students underperform other students on the MAP. On both assessments across grades, SES has a medium effect on mean test scores where students assigned to the SES category underperform other students on the MAP.

In terms of the race/ethnicity in all grades, there is a moderate difference in mean Communication Arts test scores for black students compared to white students, where black students underperform white students on average. There is a moderate difference in mean Mathematics tests scores for black students compared to white students in grades 3 through 5. In grades 6 through 8 and 10, there is a large difference in mean Mathematics test scores for black students compared to white students.

Tables 4.24 through 4.37 show that there is a small to moderate difference in mean Communication Arts and Mathematics test scores for Hispanic students compared to white students across all grades. There tends to be a small difference in mean Communication Arts and Mathematics test scores for Native American students

compared to white students across all grades. Finally, there is a small difference in mean Communication Arts and Mathematics test scores for Asian/Pacific Islander (API) students compared to white students, where API students tend to outperform white students in almost all grades.

Tables 4.24 through 4.37 also present the Coefficient Alpha reliability statistic for each subgroup. For each subgroup, the reliability statistic is in the high 0.80 range or low 0.90 range. This indicates that the tests are a reliable measure for each subgroup.

**Table 4. 1: Number of Schools who Participated in the Calibration Sample**

<b>Grade Level</b>	<b>Number of Schools</b>
<b>3</b>	71
<b>4</b>	70
<b>5</b>	65
<b>6</b>	41
<b>7</b>	40
<b>8</b>	39
<b>10</b>	36
<b>11</b>	36

**Table 4. 2: Summary of Calibration and Census Data, Communication Arts**

	Communication Arts Grade 3							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	4248		638.97	38.51	66347		639.58	38.04
<b>Gender</b>								
Male	2146	50.52	634.95	37.80	33962	51.19	635.00	38.37
Female	2096	49.34	643.13	38.76	32195	48.53	644.49	37.03
Unknown	6	0.14	621.50	49.67	190	0.29	625.72	43.44
<b>Race/Ethnicity</b>								
White	3237	76.20	643.52	37.14	50131	75.56	644.37	36.32
Black	702	16.53	622.76	38.44	12118	18.26	622.11	38.82
Hispanic	218	5.13	621.34	40.10	2576	3.88	626.04	38.37
Asian/Pacific Islander	64	1.51	647.66	39.52	1154	1.74	648.26	40.70
Native American/Alaskan	22	0.52	642.95	24.74	265	0.40	635.03	30.54
Unknown	3237	76.20	643.52	37.14	103	0.16	616.61	46.04
	<b>Communication Arts Grade 4</b>							
<b>All Students</b>	4135		656.77	37.02	65274		656.11	39.51
<b>Gender</b>								
Male	2071	50.08	651.03	37.57	33104	50.72	650.50	40.31
Female	2048	49.53	662.74	35.54	31974	48.98	661.98	37.75
Unknown	16	0.39	636.50	28.54	196	0.30	644.41	42.35
<b>Race/Ethnicity</b>								
White	3162	76.47	661.45	35.78	49616	76.01	660.85	37.72
Black	675	16.32	638.68	35.73	11782	18.05	637.53	40.73
Hispanic	191	4.62	643.44	36.57	2341	3.59	644.50	38.38
Asian/Pacific Islander	73	1.77	663.27	40.19	1145	1.75	665.84	42.96
Native American/Alaskan	24	0.58	644.92	43.92	260	0.40	655.10	39.64
Unknown	10	0.24	632.50	30.36	130	0.20	655.12	38.83
	<b>Communication Arts Grade 5</b>							
<b>All Students</b>	3889		670.49	35.94	65461		671.01	37.14
<b>Gender</b>								
Male	2006	51.58	667.11	35.42	33252	50.80	666.17	37.97
Female	1876	48.24	674.06	36.18	32039	48.94	676.14	35.49
Unknown	7	0.18	679.86	25.70	170	0.26	651.02	44.85
<b>Race/Ethnicity</b>								
White	2909	74.80	675.18	34.51	50222	76.72	675.10	35.57
Black	687	17.67	652.34	37.19	11530	17.61	654.36	38.02
Hispanic	186	4.78	660.52	30.10	2209	3.37	660.22	37.47
Asian/Pacific Islander	69	1.77	683.16	34.97	1160	1.77	682.53	39.78
Native American/Alaskan	28	0.72	661.43	36.09	266	0.41	666.83	40.27
Unknown	10	0.26	674.90	15.32	74	0.11	642.72	54.28

**Table 4. 2: Summary of Calibration and Census Data, Communication Arts (Continued)**

	Communication Arts Grade 6							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	3992		666.20	33.52	66247		667.99	34.63
<b>Gender</b>								
Male	2023	50.68	662.37	35.78	34076	51.44	663.90	36.32
Female	1961	49.12	670.23	30.54	31999	48.30	672.43	32.04
Unknown	8	0.20	648.00	20.78	172	0.26	650.48	47.79
<b>Race/Ethnicity</b>								
White	2972	74.45	670.89	32.18	50480	76.20	672.68	32.41
Black	730	18.29	650.98	32.10	12005	18.12	649.34	36.13
Hispanic	219	5.49	651.69	36.81	2237	3.38	657.60	36.30
Asian/Pacific Islander	60	1.50	673.83	33.77	1179	1.78	679.15	35.55
Native American/Alaskan	10	0.25	658.40	29.96	273	0.41	662.62	37.58
Unknown	1	0.03	658.00		73	0.11	643.32	51.78
	<b>Communication Arts Grade 7</b>							
<b>All Students</b>	4739		672.20	35.69	67167		672.11	36.26
<b>Gender</b>								
Male	2390	50.43	667.83	38.07	34353	51.15	667.71	38.29
Female	2343	49.44	676.68	32.52	32588	48.52	676.86	33.30
Unknown	6	0.13	661.33	24.83	226	0.34	656.01	40.87
<b>Race/Ethnicity</b>								
White	3763	79.40	676.60	34.19	51125	76.12	677.75	33.85
Black	706	14.90	653.51	34.42	12424	18.50	650.60	36.57
Hispanic	196	4.14	655.79	40.08	2145	3.19	658.85	37.43
Asian/Pacific Islander	51	1.08	676.71	39.27	1027	1.53	683.26	39.00
Native American/Alaskan	18	0.38	657.33	41.05	310	0.46	667.24	35.64
Unknown	5	0.11	652.80	37.88	136	0.20	652.20	40.79
	<b>Communication Arts Grade 8</b>							
<b>All Students</b>	4549		689.25	34.78	70187		686.90	37.54
<b>Gender</b>								
Male	2305	50.67	683.70	35.19	35824	51.04	680.73	37.76
Female	2217	48.74	695.30	33.15	34030	48.48	693.55	36.08
Unknown	27	0.59	665.33	41.82	333	0.47	670.58	41.38
<b>Race/Ethnicity</b>								
White	3651	80.26	693.25	33.69	53308	75.95	692.45	35.91
Black	625	13.74	670.71	32.72	13166	18.76	665.61	35.82
Hispanic	182	4.00	674.66	37.46	2104	3.00	675.73	34.85
Asian/Pacific Islander	52	1.14	693.44	34.25	1035	1.47	699.49	40.70
Native American/Alaskan	18	0.40	678.39	38.18	334	0.48	681.01	38.26
Unknown	21	0.46	670.48	37.13	240	0.34	672.75	41.66

**Table 4. 2: Summary of Calibration and Census Data, Communication Arts (Continued)**

	Communication Arts Grade 11							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	3867		714.52	31.60	61482		715.86	31.32
<b>Gender</b>								
Male	1946	50.32	709.04	33.04	30624	49.81	711.28	33.00
Female	1910	49.39	720.31	28.72	30620	49.80	720.58	28.74
Unknown	11	0.28	677.36	51.24	238	0.39	696.62	32.42
<b>Race/Ethnicity</b>								
White	3283	84.90	717.37	30.35	49363	80.29	719.35	29.92
Black	453	11.71	694.26	32.54	9264	15.07	697.87	31.83
Hispanic	67	1.73	707.03	31.38	1476	2.40	707.54	31.08
Asian/Pacific Islander	42	1.09	722.57	31.21	969	1.58	726.05	32.37
Native American/Alaskan	15	0.39	728.00	22.39	292	0.47	714.80	31.66
Unknown	7	0.18	682.00	56.90	118	0.19	690.58	33.93

**Table 4. 3: Summary of Calibration and Census Data, Mathematics**

	Mathematics Grade 3							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	4198		622.71	39.19	66640		622.40	38.72
<b>Gender</b>								
Male	2119	50.48	621.86	39.16	34169	51.27	621.77	39.37
Female	2071	49.33	623.63	39.26	32280	48.44	623.20	37.93
Unknown	8	0.19	609.50	20.44	191	0.29	598.94	43.85
<b>Race/Ethnicity</b>								
White	3200	76.23	628.04	37.69	50258	75.42	627.87	36.69
Black	693	16.51	602.23	37.96	12202	18.31	601.16	38.96
Hispanic	219	5.22	605.98	36.20	2622	3.93	611.26	37.12
Asian/Pacific Islander	63	1.50	637.02	47.36	1186	1.78	636.14	41.27
Native American/Alaskan	22	0.52	618.18	33.46	267	0.40	618.75	33.92
Unknown	1	0.02	614.00		105	0.16	601.33	43.42
	Mathematics Grade 4							
<b>All Students</b>	4070		645.31	35.47	65363		644.47	36.56
<b>Gender</b>								
Male	2043	50.20	644.90	36.95	33172	50.75	644.11	37.44
Female	2017	49.56	645.74	33.93	32019	48.99	644.93	35.55
Unknown	10	0.25	642.20	30.61	172	0.26	629.70	46.84
<b>Race/Ethnicity</b>								
White	3112	76.46	650.37	33.92	49657	75.97	649.86	34.13
Black	672	16.51	622.22	33.91	11789	18.04	622.29	37.35
Hispanic	185	4.55	640.97	32.16	2377	3.64	635.49	34.72
Asian/Pacific Islander	73	1.79	658.51	34.95	1187	1.82	658.99	41.23
Native American/Alaskan	21	0.52	630.05	36.74	259	0.40	643.16	36.27
Unknown	7	0.17	636.43	20.12	94	0.14	629.77	44.39
	Mathematics Grade 5							
<b>All Students</b>	3971		662.02	42.82	65498		663.21	41.50
<b>Gender</b>								
Male	2038	51.32	663.63	43.26	33271	50.80	663.43	42.37
Female	1907	48.02	660.27	42.34	32049	48.93	663.10	40.54
Unknown	26	0.65	663.62	39.53	178	0.27	641.63	43.90
<b>Race/Ethnicity</b>								
White	2987	75.22	668.24	40.87	50176	76.61	668.99	39.52
Black	682	17.17	635.68	42.06	11527	17.60	638.62	40.25
Hispanic	189	4.76	653.31	37.64	2247	3.43	652.21	38.65
Asian/Pacific Islander	71	1.79	679.45	43.68	1185	1.81	681.28	44.69
Native American/Alaskan	27	0.68	655.85	42.63	267	0.41	661.10	41.69
Unknown	15	0.38	659.33	30.89	96	0.15	632.91	46.04

**Table 4. 3: Summary of Calibration and Census Data, Mathematics (Continued)**

	Mathematics Grade 6							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	4010		674.13	41.14	66332		676.31	41.75
<b>Gender</b>								
Male	2034	50.72	674.42	42.95	34107	51.42	676.25	43.56
Female	1972	49.18	673.85	39.24	32038	48.30	676.52	39.62
Unknown	4	0.10	668.50	22.34	187	0.28	649.78	52.47
<b>Race/Ethnicity</b>								
White	2977	74.24	681.19	39.17	50439	76.04	682.81	38.90
Black	732	18.25	650.32	38.68	12013	18.11	649.86	41.79
Hispanic	220	5.49	654.54	40.98	2285	3.44	663.43	40.79
Asian/Pacific Islander	65	1.62	688.77	33.77	1214	1.83	695.77	44.73
Native American/Alaskan	12	0.30	661.67	38.99	276	0.42	668.74	39.05
Unknown	4	0.10	658.25	36.09	105	0.16	654.59	49.68
	<b>Mathematics Grade 7</b>							
<b>All Students</b>	4753		678.08	41.72	67554		677.41	42.62
<b>Gender</b>								
Male	2395	50.39	677.45	44.04	34577	51.18	676.68	44.02
Female	2340	49.23	678.77	39.25	32666	48.36	678.38	41.02
Unknown	18	0.38	671.44	36.72	311	0.46	656.36	41.93
<b>Race/Ethnicity</b>								
White	3764	79.19	684.70	39.46	51276	75.90	684.95	39.45
Black	708	14.90	649.31	38.68	12519	18.53	647.71	41.30
Hispanic	203	4.27	654.02	42.86	2191	3.24	662.91	40.37
Asian/Pacific Islander	52	1.09	686.56	40.67	1060	1.57	697.17	47.36
Native American/Alaskan	15	0.32	668.40	38.56	310	0.46	673.72	34.35
Unknown	11	0.23	680.45	25.23	198	0.29	661.68	43.01
	<b>Mathematics Grade 8</b>							
<b>All Students</b>	4549		701.41	39.64	70204		698.33	41.98
<b>Gender</b>								
Male	2314	50.87	701.34	41.15	35901	51.14	697.72	43.83
Female	2222	48.85	701.59	38.03	34086	48.55	699.12	39.83
Unknown	13	0.29	681.62	31.87	217	0.31	675.03	49.73
<b>Race/Ethnicity</b>								
White	3660	80.46	706.64	37.85	53379	76.03	705.46	39.15
Black	624	13.72	674.17	38.55	13156	18.74	669.90	40.54
Hispanic	184	4.04	687.88	38.26	2147	3.06	686.72	38.61
Asian/Pacific Islander	54	1.19	714.13	32.54	1051	1.50	720.79	42.63
Native American/Alaskan	20	0.44	696.80	38.93	331	0.47	692.63	43.81
Unknown	7	0.15	663.86	38.89	140	0.20	675.50	45.81

**Table 4. 3: Summary of Calibration and Census Data, Mathematics (Continued)**

	Mathematics Grade 10							
	Calibration Sample				Census Data			
			Scale Score				Scale Score	
	N	%	M	SD	N	%	M	SD
<b>All Students</b>	4484		721.11	47.61	68330		723.69	47.83
<b>Gender</b>								
Male	2215	49.40	720.90	50.10	34132	49.95	723.55	50.17
Female	2260	50.40	721.31	45.10	33902	49.62	724.07	45.27
Unknown	9	0.20	720.22	33.86	296	0.43	695.66	47.34
<b>Race/Ethnicity</b>								
White	3616	80.64	727.87	45.34	53602	78.45	731.14	44.47
Black	684	15.25	688.22	44.28	11409	16.70	689.61	47.22
Hispanic	113	2.52	703.41	45.54	1803	2.64	708.48	46.63
Asian/Pacific Islander	47	1.05	731.26	53.88	1049	1.54	745.78	49.97
Native American/Alaskan	12	0.27	709.58	59.41	323	0.47	718.74	45.08
Unknown	12	0.27	697.33	53.54	144	0.21	690.57	49.86

**Table 4. 4. Comparison of State-Level Means, 2006 and 2007 Census Data**

Grade	Year	Communication Arts			Mathematics		
		N	Mean SS	S.D. SS	N	Mean SS	S.D. SS
3	2006	64486	639.86	36.84	64763	621.59	39.11
	2007	66347	639.58	38.04	66640	622.40	38.72
4	2006	65179	654.55	38.56	65306	643.88	37.07
	2007	65274	656.11	39.51	65363	644.47	36.56
5	2006	66007	668.18	37.09	66123	660.06	39.99
	2007	65461	671.01	37.14	65498	663.21	41.50
6	2006	66948	666.85	33.70	67017	673.30	39.80
	2007	66247	667.99	34.63	66332	676.31	41.75
7	2006	70290	671.63	37.06	70698	675.38	41.27
	2007	67167	672.11	36.26	67554	677.41	42.62
8	2006	72483	686.85	37.87	72542	697.73	40.37
	2007	70187	686.90	37.54	70204	698.33	41.98
High School	2006	60004	716.69	31.42	68083	724.46	51.18
	2007	61482	715.86	31.32	68330	723.69	47.83

**Table 4. 5: Comparison of Percent of Students in each Achievement Level using 2006 and 2007 Census Data, Communication Arts**

<b>Grade</b>	<b>Year</b>	<b>N</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	<b>Prof. &amp; Adv.</b>
<b>3</b>	2006	64486	8.9	48.1	26.0	17.0	43.0
	2007	66347	9.5	47.3	26.1	17.1	43.2
<b>4</b>	2006	65179	10.7	45.0	29.1	15.2	44.3
	2007	65274	10.6	43.8	28.5	17.0	45.6
<b>5</b>	2006	66007	9.2	45.3	29.9	15.6	45.5
	2007	65461	8.4	43.3	30.1	18.2	48.3
<b>6</b>	2006	66948	12.0	45.3	32.0	10.7	42.7
	2007	66247	11.4	44.5	32.2	11.9	44.1
<b>7</b>	2006	70290	14.0	42.5	31.1	12.4	43.5
	2007	67167	13.3	41.5	33.4	11.8	45.2
<b>8</b>	2006	72483	9.3	48.7	27.0	15.1	42.1
	2007	70187	8.8	49.0	27.3	14.8	42.2
<b>11</b>	2006	60004	10.2	47.3	31.5	11.0	42.5
	2007	61482	10.5	48.0	32.3	9.3	41.5

**Table 4. 6: Comparison of Percent of Students in each Achievement Level using 2006 and 2007 Census Data, Mathematics**

Grade	Year	N	Below Basic	Basic	Proficient	Advanced	Prof & Adv
3	2006	64763	7.2	49.1	33.6	10.1	43.7
	2007	66640	7.3	47.3	35.3	10.1	45.4
4	2006	65306	8.3	47.9	34.7	9.1	43.8
	2007	65363	8.2	46.9	35.5	9.4	44.9
5	2006	66123	8.0	48.2	33.0	10.8	43.8
	2007	65498	7.7	45.3	33.4	13.5	47.0
6	2006	67017	11.1	44.5	34.7	9.6	44.3
	2007	66332	11.2	40.5	35.9	12.4	48.3
7	2006	70698	17.6	38.9	33.1	10.4	43.5
	2007	67554	17.0	37.5	33.6	11.9	45.5
8	2006	72542	21.3	38.3	28.0	12.4	40.4
	2007	70204	21.7	37.1	27.0	14.2	41.2
10	2006	68083	24.4	33.3	32.0	10.4	42.3
	2007	68330	24.2	35.1	32.7	8.1	40.8

**Table 4. 7: Participation Rates, All Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	67259	98.64	67257	99.08
4	65982	98.93	65975	99.07
5	66098	99.04	66075	99.13
6	67045	98.81	67039	98.95
7	68404	98.19	68405	98.76
8	71200	98.58	71190	98.61
10	-	-	69488	98.33
11	62946	97.67	-	-

**Table 4. 8: Participation Rates, Males**

<b>Grade</b>	<b>Accountable in Comm. Arts</b>	<b>Percent Reportable in Comm. Arts</b>	<b>Accountable in Mathematics</b>	<b>Percent Reportable in Mathematics</b>
<b>3</b>	34532	98.35	34528	98.96
<b>4</b>	33529	98.73	33551	98.87
<b>5</b>	33648	98.82	33625	98.95
<b>6</b>	34575	98.56	34558	98.69
<b>7</b>	35164	97.69	35116	98.47
<b>8</b>	36473	98.22	36541	98.25
<b>10</b>	-	-	34774	98.15
<b>11</b>	31479	97.28	-	-

**Table 4. 9: Participation Rates, Females**

<b>Grade</b>	<b>Accountable in Comm. Arts</b>	<b>Percent Reportable in Comm. Arts</b>	<b>Accountable in Mathematics</b>	<b>Percent Reportable in Mathematics</b>
<b>3</b>	32501	99.06	32507	99.30
<b>4</b>	32242	99.17	32240	99.31
<b>5</b>	32269	99.29	32258	99.35
<b>6</b>	32289	99.10	32289	99.22
<b>7</b>	32991	98.78	32969	99.08
<b>8</b>	34379	98.98	34416	99.04
<b>10</b>	-	-	34349	98.70
<b>11</b>	31152	98.29	-	-

**Table 4. 10: Participation Rates, White**

<b>Grade</b>	<b>Accountable in Comm. Arts</b>	<b>Percent Reportable in Comm. Arts</b>	<b>Accountable in Mathematics</b>	<b>Percent Reportable in Mathematics</b>
<b>3</b>	50690	98.90	50685	99.16
<b>4</b>	50050	99.13	50079	99.16
<b>5</b>	50651	99.15	50608	99.15
<b>6</b>	50992	99.00	50941	99.01
<b>7</b>	51872	98.56	51830	98.93
<b>8</b>	53941	98.83	54025	98.80
<b>10</b>	-	-	54361	98.60
<b>11</b>	50317	98.10	-	-

**Table 4. 11: Participation Rates, Black**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	12340	98.20	12340	98.88
4	11926	98.79	11931	98.81
5	11653	98.94	11640	99.03
6	12185	98.52	12182	98.61
7	12785	97.18	12766	98.07
8	13445	97.92	13440	97.89
10	-	-	11698	97.53
11	9637	96.13	-	-

**Table 4. 12: Participation Rates, Hispanic**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	2643	97.47	2641	99.28
4	2410	97.14	2402	98.96
5	2257	97.87	2262	99.34
6	2305	97.05	2312	98.83
7	2217	96.75	2216	98.87
8	2160	97.41	2173	98.80
10	-	-	1823	98.90
11	1510	97.75	-	-

**Table 4. 13: Participation Rates, Asian/Pacific Islander**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	1194	96.65	1195	99.25
4	1197	95.66	1201	98.83
5	1192	97.32	1194	99.25
6	1209	97.52	1218	99.67
7	1065	96.43	1071	98.97
8	1062	97.46	1063	98.87
10	-	-	1062	98.78
11	1004	96.51	-	-

**Table 4. 14: Participation Rates, Native American/Alaskan**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	267	99.25	268	99.63
4	262	99.24	261	99.23
5	267	99.63	268	99.63
6	278	98.20	279	98.92
7	319	97.18	317	97.79
8	343	97.38	338	97.93
10	-	-	331	97.58
11	297	98.32	-	-

**Table 4. 15: Participation Rates, Students Receiving Free or Reduced-Price Lunch (SES)**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	30252	98.27	30223	98.86
4	29056	98.70	29031	98.87
5	28504	98.84	28466	98.94
6	27621	98.40	27643	98.60
7	27130	97.53	27171	98.32
8	27154	98.03	27287	98.10
10	-	-	21934	97.86
11	17279	96.72	-	-

**Table 4. 16: Participation Rates, Migrant Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	105	98.10	98	98.98
4	100	98.00	96	98.96
5	108	98.15	111	98.20
6	207	96.62	207	98.55
7	199	97.99	198	98.99
8	66	93.94	60	96.67
10	-	-	68	97.06
11	56	98.21	-	-

**Table 4. 17: Participation Rates, Special Education (IEP) Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	10592	94.25	10583	95.36
4	10341	94.95	10324	95.01
5	10059	95.28	10036	95.23
6	9744	94.46	9687	94.44
7	9478	92.32	9447	94.03
8	10100	93.99	10076	94.08
10	-	-	9015	93.63
11	7557	91.04	-	-

**Table 4. 18: Participation Rates, English Learners (LEP, ELL)**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	2578	96.04	2579	99.46
4	2286	95.14	2322	99.22
5	2086	95.69	2096	99.28
6	1972	94.62	1963	99.03
7	1837	93.09	1827	98.63
8	1808	95.69	1815	98.84
10	-	-	1757	99.20
11	1381	96.52	-	-

**Table 4. 19: Participation Rates, Students with Disabilities**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
3	10801	94.32	10792	95.41
4	10543	94.97	10531	95.04
5	10246	95.27	10234	95.25
6	9889	94.35	9847	94.40
7	9608	92.50	9612	94.10
8	10184	93.89	10170	93.94
10	-	-	9071	93.63
11	7601	91.05	-	-

**Table 4. 20: Participation Rates, Students Receiving Accommodations**

<b>Grade</b>	<b>Accountable in Comm. Arts</b>	<b>Percent Reportable in Comm. Arts</b>	<b>Accountable in Mathematics</b>	<b>Percent Reportable in Mathematics</b>
<b>3</b>	6395	98.06	6605	99.58
<b>4</b>	7036	99.35	7353	99.56
<b>5</b>	7400	99.35	7536	99.46
<b>6</b>	7432	98.91	7541	99.34
<b>7</b>	7050	96.68	7241	98.91
<b>8</b>	7411	98.50	7650	98.77
<b>10</b>	-	-	6177	97.99
<b>11</b>	4982	95.89	-	-

**Table 4. 21: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Regular Edition**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Regular Edition	66302	100.00	66599	100.00
	Oral reading	74	0.11	4066	6.11
	Signing of assessment	5	0.01	26	0.04
	Paraphrasing	8	0.01	14	0.02
	Other administration	157	0.24	125	0.19
	Oral reading in native language	8	0.01	123	0.18
	Extend time-TerraNova session	2520	3.80	2543	3.82
	Administer using > allotted periods	2416	3.64	2314	3.47
	Other timing	498	0.75	488	0.73
	Use of scribe	1986	3.00	1716	2.58
	Use of calculator, math table, etc.	163	0.25	1221	1.83
	Use of bilingual dictionary	7	0.01	24	0.04
	Other response	85	0.13	78	0.12
	Testing individually	1916	2.89	1781	2.67
	Testing in small group	4446	6.71	4838	7.26
	Other setting	292	0.44	294	0.44
4	Regular Edition	65226	100.00	65314	100.00
	Oral reading	110	0.17	4634	7.09
	Signing of assessment	2	0.00	12	0.02
	Paraphrasing	11	0.02	8	0.01
	Other administration	191	0.29	158	0.24
	Oral reading in native language	1	0.00	96	0.15
	Extend time-TerraNova session	2805	4.30	2902	4.44
	Administer using > allotted periods	2677	4.10	2646	4.05
	Other timing	469	0.72	486	0.74
	Use of scribe	1914	2.93	1753	2.68
	Use of calculator, math table, etc.	169	0.26	1557	2.38
	Use of bilingual dictionary	5	0.01	27	0.04
	Other response	66	0.10	78	0.12
	Testing individually	1914	2.93	1823	2.79
	Testing in small group	5127	7.86	5434	8.32
	Other setting	272	0.42	259	0.40

**Table 4. 21: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Regular Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
5	Regular Edition	65410	100.00	65448	100.00
	Oral reading	88	0.13	4458	6.81
	Signing of assessment	8	0.01	25	0.04
	Paraphrasing	8	0.01	13	0.02
	Other administration	150	0.23	148	0.23
	Oral reading in native language	2	0.00	95	0.15
	Extend time-TerraNova session	2811	4.30	2783	4.25
	Administer using > allotted periods	2920	4.46	2795	4.27
	Other timing	602	0.92	602	0.92
	Use of scribe	1656	2.53	1438	2.20
	Use of calculator, math table, etc.	210	0.32	2133	3.26
	Use of bilingual dictionary	4	0.01	23	0.04
	Other response	56	0.09	56	0.09
	Testing individually	1786	2.73	1635	2.50
	Testing in small group	5681	8.69	5942	9.08
Other setting	273	0.42	295	0.45	
6	Regular Edition	66200	100.00	66284	100.00
	Oral reading	90	0.14	4111	6.20
	Signing of assessment	1	0.00	17	0.03
	Paraphrasing	14	0.02	17	0.03
	Other administration	109	0.16	102	0.15
	Oral reading in native language	4	0.01	90	0.14
	Extend time-TerraNova session	2834	4.28	2841	4.29
	Administer using > allotted periods	2377	3.59	2227	3.36
	Other timing	555	0.84	534	0.81
	Use of scribe	1130	1.71	924	1.39
	Use of calculator, math table, etc.	376	0.57	2862	4.32
	Use of bilingual dictionary	3	0.00	71	0.11
	Other response	58	0.09	45	0.07
	Testing individually	1230	1.86	1126	1.70
	Testing in small group	5986	9.04	6178	9.32
Other setting	167	0.25	161	0.24	

**Table 4. 21: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Regular Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
7	Regular Edition	67137	100.00	67521	100.00
	Oral reading	74	0.11	3368	4.99
	Signing of assessment	0	0.00	7	0.01
	Paraphrasing	10	0.01	16	0.02
	Other administration	96	0.14	97	0.14
	Oral reading in native language	7	0.01	81	0.12
	Extend time-TerraNova session	2451	3.65	2554	3.78
	Administer using > allotted periods	2030	3.02	2039	3.02
	Other timing	471	0.70	421	0.62
	Use of scribe	759	1.13	570	0.84
	Use of calculator, math table, etc.	411	0.61	3121	4.62
	Use of bilingual dictionary	29	0.04	100	0.15
	Other response	37	0.06	17	0.03
	Testing individually	752	1.12	707	1.05
	Testing in small group	5720	8.52	5945	8.80
Other setting	104	0.15	115	0.17	
8	Regular Edition	70156	100.00	70172	100.00
	Oral reading	115	0.16	3291	4.69
	Signing of assessment	2	0.00	7	0.01
	Paraphrasing	18	0.03	19	0.03
	Other administration	98	0.14	121	0.17
	Oral reading in native language	3	0.00	97	0.14
	Extend time-TerraNova session	2552	3.64	2539	3.62
	Administer using > allotted periods	2268	3.23	2194	3.13
	Other timing	481	0.69	395	0.56
	Use of scribe	623	0.89	423	0.60
	Use of calculator, math table, etc.	501	0.71	3458	4.93
	Use of bilingual dictionary	22	0.03	81	0.12
	Other response	29	0.04	27	0.04
	Testing individually	703	1.00	642	0.91
	Testing in small group	6137	8.75	6326	9.01
Other setting	128	0.18	104	0.15	

**Table 4. 21: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Regular Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
HS	Regular Edition	61459	100.00	68299	100.00
	Oral reading	99	0.16	2238	3.28
	Signing of assessment	0	0.00	19	0.03
	Paraphrasing	10	0.02	20	0.03
	Other administration	48	0.08	92	0.13
	Oral reading in native language	4	0.01	42	0.06
	Extend time-TerraNova session	1658	2.70	2056	3.01
	Administer using > allotted periods	1020	1.66	1260	1.84
	Other timing	226	0.37	312	0.46
	Use of scribe	241	0.39	171	0.25
	Use of calculator, math table, etc.	128	0.21	2832	4.15
	Use of bilingual dictionary	3	0.00	28	0.04
	Other response	17	0.03	13	0.02
	Testing individually	287	0.47	282	0.41
	Testing in small group	4138	6.73	5105	7.47
	Other setting	74	0.12	104	0.15

**Table 4. 22: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille Edition**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Braille	6	100.00	6	100.00
	Oral reading	2	33.33	3	50.00
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	3	50.00	1	16.67
	Administer using > allotted periods	0	0.00	0	0.00
	Other timing	0	0.00	0	0.00
	Use of scribe	4	66.67	4	66.67
	Use of calculator, math table, etc.	0	0.00	2	33.33
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	6	100.00	6	100.00
	Testing in small group	0	0.00	0	0.00
	Other setting	1	16.67	1	16.67
	4	Braille	5	100.00	5
Oral reading		0	0.00	1	20.00
Signing of assessment		0	0.00	0	0.00
Paraphrasing		0	0.00	0	0.00
Other administration		1	20.00	1	20.00
Oral reading in native language		0	0.00	0	0.00
Extend time-TerraNova session		1	20.00	1	20.00
Administer using > allotted periods		1	20.00	1	20.00
Other timing		1	20.00	0	0.00
Use of scribe		2	40.00	2	40.00
Use of calculator, math table, etc.		0	0.00	1	20.00
Use of bilingual dictionary		0	0.00	0	0.00
Other response		0	0.00	0	0.00
Testing individually		4	80.00	3	60.00
Testing in small group		0	0.00	0	0.00
Other setting		0	0.00	0	0.00

**Table 4. 22: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
5	Braille	6	100.00	6	100.00
	Oral reading	0	0.00	0	0.00
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	0	0.00	1	16.67
	Administer using > allotted periods	3	50.00	3	50.00
	Other timing	0	0.00	0	0.00
	Use of scribe	3	50.00	3	50.00
	Use of calculator, math table, etc.	0	0.00	0	0.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	3	50.00	3	50.00
	Testing in small group	2	33.33	1	16.67
Other setting	0	0.00	0	0.00	
6	Braille	7	100.00	7	100.00
	Oral reading	0	0.00	1	14.29
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	4	57.14	4	57.14
	Administer using > allotted periods	3	42.86	3	42.86
	Other timing	0	0.00	0	0.00
	Use of scribe	2	28.57	2	28.57
	Use of calculator, math table, etc.	1	14.29	2	28.57
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	5	71.43	5	71.43
	Testing in small group	1	14.29	1	14.29
Other setting	0	0.00	0	0.00	

**Table 4. 22: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
7	Braille	6	100.00	5	100.00
	Oral reading	1	16.67	2	40.00
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	3	50.00	2	40.00
	Administer using > allotted periods	2	33.33	1	20.00
	Other timing	0	0.00	0	0.00
	Use of scribe	3	50.00	3	60.00
	Use of calculator, math table, etc.	0	0.00	1	20.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	2	33.33	1	20.00
	Testing individually	4	66.67	4	80.00
	Testing in small group	1	16.67	1	20.00
Other setting	0	0.00	0	0.00	
8	Braille	6	100.00	8	100.00
	Oral reading	0	0.00	3	37.50
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	0	0.00	1	12.50
	Administer using > allotted periods	3	50.00	4	50.00
	Other timing	0	0.00	0	0.00
	Use of scribe	3	50.00	3	37.50
	Use of calculator, math table, etc.	0	0.00	2	25.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	2	33.33	3	37.50
	Testing in small group	2	33.33	2	25.00
Other setting	0	0.00	0	0.00	

**Table 4. 22: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
HS	Braille	6	100.00	8	100.00
	Oral reading	0	0.00	3	37.50
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	1	12.50
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	1	16.67	2	25.00
	Administer using > allotted periods	4	66.67	5	62.50
	Other timing	1	16.67	0	0.00
	Use of scribe	2	33.33	5	62.50
	Use of calculator, math table, etc.	0	0.00	5	62.50
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	1	12.50
	Testing individually	4	66.67	6	75.00
	Testing in small group	1	16.67	2	25.00
	Other setting	0	0.00	1	12.50

**Table 4. 23: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large Print Edition**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Large Print	39	100.00	35	100.00
	Oral reading	3	7.69	17	48.57
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	1	2.56	0	0.00
	Other administration	2	5.13	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	11	28.21	9	25.71
	Administer using > allotted periods	15	38.46	15	42.86
	Other timing	3	7.69	2	5.71
	Use of scribe	23	58.97	18	51.43
	Use of calculator, math table, etc.	0	0.00	7	20.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	1	2.56	1	2.86
	Testing individually	21	53.85	18	51.43
	Testing in small group	14	35.90	15	42.86
	Other setting	2	5.13	2	5.71
4	Large Print	43	100.00	44	100.00
	Oral reading	1	2.33	17	38.64
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	2	4.65	3	6.82
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	14	32.56	13	29.55
	Administer using > allotted periods	19	44.19	19	43.18
	Other timing	2	4.65	2	4.55
	Use of scribe	21	48.84	17	38.64
	Use of calculator, math table, etc.	0	0.00	6	13.64
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	1	2.33	1	2.27
	Testing individually	13	30.23	14	31.82
	Testing in small group	24	55.81	21	47.73
	Other setting	1	2.33	1	2.27

**Table 4. 23: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large Print Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
5	Large Print	45	100.00	44	100.00
	Oral reading	3	6.67	17	38.64
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	3	6.67	1	2.27
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	17	37.78	16	36.36
	Administer using > allotted periods	12	26.67	15	34.09
	Other timing	3	6.67	2	4.55
	Use of scribe	16	35.56	15	34.09
	Use of calculator, math table, etc.	5	11.11	11	25.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	1	2.22	2	4.55
	Testing individually	15	33.33	15	34.09
	Testing in small group	18	40.00	18	40.91
Other setting	0	0.00	0	0.00	
6	Large Print	40	100.00	41	100.00
	Oral reading	3	7.50	14	34.15
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	2	5.00	3	7.32
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	12	30.00	13	31.71
	Administer using > allotted periods	13	32.50	10	24.39
	Other timing	5	12.50	4	9.76
	Use of scribe	23	57.50	22	53.66
	Use of calculator, math table, etc.	1	2.50	16	39.02
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	2	5.00	1	2.44
	Testing individually	22	55.00	19	46.34
	Testing in small group	15	37.50	16	39.02
Other setting	2	5.00	1	2.44	

**Table 4. 23: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large Print Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
7	Large Print	24	100.00	28	100.00
	Oral reading	2	8.33	7	25.00
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	0	0.00	1	3.57
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	3	12.50	6	21.43
	Administer using > allotted periods	5	20.83	5	17.86
	Other timing	3	12.50	3	10.71
	Use of scribe	10	41.67	8	28.57
	Use of calculator, math table, etc.	1	4.17	6	21.43
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	10	41.67	10	35.71
	Testing in small group	9	37.50	12	42.86
Other setting	0	0.00	0	0.00	
8	Large Print	25	100.00	24	100.00
	Oral reading	1	4.00	8	33.33
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	1	4.00	1	4.17
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	5	20.00	7	29.17
	Administer using > allotted periods	7	28.00	9	37.50
	Other timing	0	0.00	0	0.00
	Use of scribe	7	28.00	3	12.50
	Use of calculator, math table, etc.	1	4.00	6	25.00
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	5	20.00	3	12.50
	Testing in small group	9	36.00	14	58.33
Other setting	0	0.00	0	0.00	

**Table 4. 23: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large Print Edition (Continued)**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
HS	Large Print	17	100.00	23	100.00
	Oral reading	0	0.00	7	30.43
	Signing of assessment	0	0.00	0	0.00
	Paraphrasing	0	0.00	0	0.00
	Other administration	1	5.88	0	0.00
	Oral reading in native language	0	0.00	0	0.00
	Extend time-TerraNova session	6	35.29	6	26.09
	Administer using > allotted periods	1	5.88	3	13.04
	Other timing	0	0.00	1	4.35
	Use of scribe	6	35.29	4	17.39
	Use of calculator, math table, etc.	0	0.00	6	26.09
	Use of bilingual dictionary	0	0.00	0	0.00
	Other response	0	0.00	0	0.00
	Testing individually	6	35.29	4	17.39
	Testing in small group	8	47.06	11	47.83
	Other setting	0	0.00	1	4.35

**Table 4. 24: Impact Analysis, Grade 3 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50131	644.37	36.32		0.90
	Black (not Hispanic)	12118	622.11	38.82	0.60	0.92
	Hispanic	2576	626.04	38.37	0.50	0.91
	Asian/Pacific Islander	1154	648.26	40.70	-0.11	0.91
	Native American	265	635.03	30.54	0.26	0.88
	Unknown	103	616.61	46.04	0.76	0.94
<b>Gender</b>	Male	33962	635.00	38.37		0.91
	Female	32195	644.49	37.03	-0.25	0.90
<b>ELL Status</b>	Non ELL	63871	640.45	37.68		0.91
	ELL	2476	616.95	40.30	0.62	0.92
<b>IEP Status</b>	Non IEP	56364	644.57	34.22		0.89
	IEP	9983	611.38	45.52	0.92	0.93
<b>SES Status</b>	Non SES	36617	649.06	35.26		0.89
	SES	29730	627.90	38.09	0.58	0.91
<b>Disability</b>	No Disability	56160	644.58	34.26		0.89
	Disability	10187	612.01	45.37	0.90	0.93
<b>Accommodations</b>	No Accommodations	60076	644.16	34.51		0.89
	Accommodations	6271	595.66	42.07	1.39	0.92
<b>Migrant Status</b>	Non Migrant	66244	639.63	38.03		0.91
	Migrant	103	608.40	31.79	0.82	0.89

**Table 4. 25: Impact Analysis, Grade 4 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	49616	660.85	37.72		0.90
	Black (not Hispanic)	11782	637.53	40.73	0.61	0.91
	Hispanic	2341	644.50	38.38	0.43	0.90
	Asian/Pacific Islander	1145	665.84	42.96	-0.13	0.91
	Native American	260	655.10	39.64	0.15	0.91
	Unknown	130	655.12	38.83	0.15	0.90
<b>Gender</b>	Male	33104	650.50	40.31		0.91
	Female	31974	661.98	37.75	-0.29	0.90
<b>ELL Status</b>	Non ELL	63099	656.93	39.19		0.90
	ELL	2175	632.22	41.22	0.63	0.91
<b>IEP Status</b>	Non IEP	55455	661.93	34.50		0.87
	IEP	9819	623.20	48.79	1.05	0.93
<b>SES Status</b>	Non SES	36596	665.70	36.84		0.89
	SES	28678	643.86	39.42	0.57	0.91
<b>Disability</b>	No Disability	55261	661.95	34.47		0.87
	Disability	10013	623.84	48.86	1.03	0.93
<b>Accommodations</b>	No Accommodations	58284	661.56	34.89		0.88
	Accommodations	6990	610.59	45.89	1.43	0.92
<b>Migrant Status</b>	Non Migrant	65176	656.15	39.48		0.91
	Migrant	98	629.39	45.68	0.68	0.92

**Table 4. 26: Impact Analysis, Grade 5 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50222	675.10	35.57		0.90
	Black (not Hispanic)	11530	654.36	38.02	0.58	0.91
	Hispanic	2209	660.22	37.47	0.42	0.90
	Asian/Pacific Islander	1160	682.53	39.78	-0.21	0.91
	Native American	266	666.83	40.27	0.23	0.91
	Unknown	74	642.72	54.28	0.91	0.92
<b>Gender</b>	Male	33252	666.17	37.97		0.91
	Female	32039	676.14	35.49	-0.27	0.90
<b>ELL Status</b>	Non ELL	63465	671.73	36.78		0.91
	ELL	1996	648.16	41.16	0.64	0.91
<b>IEP Status</b>	Non IEP	55876	677.12	31.50		0.88
	IEP	9584	635.41	46.42	1.22	0.92
<b>SES Status</b>	Non SES	37287	680.34	34.43		0.89
	SES	28174	658.66	37.01	0.61	0.91
<b>Disability</b>	No Disability	55700	677.17	31.45		0.88
	Disability	9761	635.87	46.44	1.21	0.92
<b>Accommodations</b>	No Accommodations	58109	676.59	32.06		0.88
	Accommodations	7352	626.92	44.38	1.51	0.90
<b>Migrant Status</b>	Non Migrant	65355	671.06	37.11		0.91
	Migrant	106	642.53	44.17	0.77	0.91

**Table 4. 27: Impact Analysis, Grade 6 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50480	672.68	32.41		0.90
	Black (not Hispanic)	12005	649.34	36.13	0.70	0.91
	Hispanic	2237	657.60	36.30	0.46	0.91
	Asian/Pacific Islander	1179	679.15	35.55	-0.20	0.90
	Native American	273	662.62	37.58	0.31	0.91
	Unknown	73	643.32	51.78	0.91	0.93
<b>Gender</b>	Male	34076	663.90	36.32		0.91
	Female	31999	672.43	32.04	-0.25	0.90
<b>ELL Status</b>	Non ELL	64381	668.82	34.13		0.91
	ELL	1866	639.14	39.00	0.87	0.91
<b>IEP Status</b>	Non IEP	57043	673.85	28.81		0.88
	IEP	9204	631.66	44.21	1.34	0.92
<b>SES Status</b>	Non SES	39066	675.93	31.99		0.89
	SES	27180	656.57	35.08	0.58	0.91
<b>Disability</b>	No Disability	56917	673.87	28.80		0.88
	Disability	9330	632.12	44.24	1.33	0.92
<b>Accommodations</b>	No Accommodations	58896	673.30	29.30		0.88
	Accommodations	7351	625.45	43.52	1.57	0.91
<b>Migrant Status</b>	Non Migrant	66047	668.03	34.59		0.91
	Migrant	200	654.95	44.12	0.38	0.93

**Table 4. 28: Impact Analysis, Grade 7 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	51125	677.75	33.85		0.90
	Black (not Hispanic)	12424	650.60	36.57	0.79	0.91
	Hispanic	2145	658.85	37.43	0.56	0.91
	Asian/Pacific Islander	1027	683.26	39.00	-0.16	0.92
	Native American	310	667.24	35.64	0.31	0.90
	Unknown	136	652.20	40.79	0.75	0.92
<b>Gender</b>	Male	34353	667.71	38.29		0.92
	Female	32588	676.86	33.30	-0.25	0.90
<b>ELL Status</b>	Non ELL	65457	672.94	35.82		0.91
	ELL	1710	640.58	39.03	0.90	0.91
<b>IEP Status</b>	Non IEP	58416	678.22	31.30		0.89
	IEP	8750	631.32	40.53	1.44	0.91
<b>SES Status</b>	Non SES	40702	680.50	33.88		0.90
	SES	26461	659.21	36.02	0.61	0.91
<b>Disability</b>	No Disability	58280	678.25	31.29		0.89
	Disability	8887	631.86	40.63	1.42	0.91
<b>Accommodations</b>	No Accommodations	60351	677.36	31.89		0.89
	Accommodations	6816	625.68	39.45	1.63	0.90
<b>Migrant Status</b>	Non Migrant	66972	672.11	36.26		0.91
	Migrant	195	673.42	36.82	-0.04	0.92

**Table 4. 29: Impact Analysis, Grade 8 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	53308	692.45	35.91		0.91
	Black (not Hispanic)	13166	665.61	35.82	0.75	0.91
	Hispanic	2104	675.73	34.85	0.47	0.91
	Asian/Pacific Islander	1035	699.49	40.70	-0.20	0.92
	Native American	334	681.01	38.26	0.32	0.92
	Unknown	240	672.75	41.66	0.55	0.93
<b>Gender</b>	Male	35824	680.73	37.76		0.92
	Female	34030	693.55	36.08	-0.35	0.91
<b>ELL Status</b>	Non ELL	68457	687.65	37.25		0.92
	ELL	1730	657.03	36.65	0.82	0.91
<b>IEP Status</b>	Non IEP	60694	693.60	32.33		0.90
	IEP	9493	644.05	40.13	1.48	0.91
<b>SES Status</b>	Non SES	43566	694.79	35.84		0.91
	SES	26620	673.97	36.65	0.58	0.92
<b>Disability</b>	No Disability	60625	693.60	32.33		0.90
	Disability	9562	644.39	40.30	1.47	0.91
<b>Accommodations</b>	No Accommodations	62887	692.40	33.16		0.90
	Accommodations	7300	639.46	39.52	1.63	0.90
<b>Migrant Status</b>	Non Migrant	70125	686.92	37.53		0.92
	Migrant	62	658.76	34.30	0.75	0.89

**Table 4. 30: Impact Analysis, Grade 11 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	49363	719.35	29.92		0.90
	Black (not Hispanic)	9264	697.87	31.83	0.71	0.90
	Hispanic	1476	707.54	31.08	0.39	0.90
	Asian/Pacific Islander	969	726.05	32.37	-0.22	0.92
	Native American	292	714.80	31.66	0.15	0.91
	Unknown	118	690.58	33.93	0.96	0.90
<b>Gender</b>	Male	30624	711.28	33.00		0.91
	Female	30620	720.58	28.74	-0.30	0.90
<b>ELL Status</b>	Non ELL	60149	716.37	31.06		0.91
	ELL	1333	692.67	34.30	0.76	0.90
<b>IEP Status</b>	Non IEP	54601	720.85	26.79		0.89
	IEP	6880	676.24	36.10	1.59	0.89
<b>SES Status</b>	Non SES	44768	720.55	29.62		0.90
	SES	16713	703.29	32.29	0.57	0.90
<b>Disability</b>	No Disability	54561	720.86	26.76		0.89
	Disability	6921	676.40	36.24	1.59	0.89
<b>Accommodations</b>	No Accommodations	56705	719.51	27.90		0.89
	Accommodations	4777	672.46	36.70	1.75	0.88
<b>Migrant Status</b>	Non Migrant	61427	715.88	31.31		0.91
	Migrant	55	687.67	29.72	0.90	0.88

**Table 4. 31: Impact Analysis, Grade 3 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50258	627.87	36.69		0.90
	Black (not Hispanic)	12202	601.16	38.96	0.72	0.92
	Hispanic	2622	611.26	37.12	0.45	0.90
	Asian/Pacific Islander	1186	636.14	41.27	-0.22	0.91
	Native American	267	618.75	33.92	0.25	0.89
	Unknown	105	601.33	43.42	0.72	0.94
<b>Gender</b>	Male	34169	621.77	39.37		0.91
	Female	32280	623.20	37.93	-0.04	0.91
<b>ELL Status</b>	Non ELL	64075	623.09	38.50		0.91
	ELL	2565	605.19	40.36	0.46	0.92
<b>IEP Status</b>	Non IEP	56548	626.37	36.33		0.90
	IEP	10092	600.13	43.84	0.70	0.93
<b>SES Status</b>	Non SES	36763	632.02	36.83		0.90
	SES	29877	610.56	37.70	0.58	0.91
<b>Disability</b>	No Disability	56343	626.40	36.30		0.90
	Disability	10297	600.49	43.90	0.69	0.93
<b>Accommodations</b>	No Accommodations	60063	626.25	36.48		0.90
	Accommodations	6577	587.24	40.82	1.06	0.92
<b>Migrant Status</b>	Non Migrant	66543	622.43	38.72		0.91
	Migrant	97	599.16	33.19	0.60	0.90

**Table 4. 32: Impact Analysis, Grade 4 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	49657	649.86	34.13		0.92
	Black (not Hispanic)	11789	622.29	37.35	0.79	0.93
	Hispanic	2377	635.49	34.72	0.42	0.92
	Asian/Pacific Islander	1187	658.99	41.23	-0.27	0.93
	Native American	259	643.16	36.27	0.20	0.92
	Unknown	94	629.77	44.39	0.59	0.94
<b>Gender</b>	Male	33172	644.11	37.44		0.93
	Female	32019	644.93	35.55	-0.02	0.92
<b>ELL Status</b>	Non ELL	63059	645.14	36.30		0.93
	ELL	2304	626.28	38.95	0.52	0.93
<b>IEP Status</b>	Non IEP	55554	648.77	33.41		0.91
	IEP	9809	620.15	43.50	0.82	0.94
<b>SES Status</b>	Non SES	36658	653.81	34.20		0.91
	SES	28703	632.56	36.01	0.61	0.93
<b>Disability</b>	No Disability	55354	648.79	33.38		0.91
	Disability	10009	620.59	43.55	0.80	0.94
<b>Accommodations</b>	No Accommodations	58042	648.70	33.61		0.91
	Accommodations	7321	610.98	41.44	1.10	0.93
<b>Migrant Status</b>	Non Migrant	65268	644.50	36.56		0.93
	Migrant	95	627.47	35.57	0.47	0.92

**Table 4. 33: Impact Analysis, Grade 5 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50176	668.99	39.52		0.91
	Black (not Hispanic)	11527	638.62	40.25	0.77	0.92
	Hispanic	2247	652.21	38.65	0.43	0.91
	Asian/Pacific Islander	1185	681.28	44.69	-0.31	0.92
	Native American	267	661.10	41.69	0.20	0.92
	Unknown	96	632.91	46.04	0.91	0.93
<b>Gender</b>	Male	33271	663.43	42.37		0.92
	Female	32049	663.10	40.54	0.01	0.92
<b>ELL Status</b>	Non ELL	63417	663.92	41.28		0.92
	ELL	2081	641.53	42.56	0.54	0.92
<b>IEP Status</b>	Non IEP	55941	668.79	37.94		0.91
	IEP	9557	630.54	46.14	0.97	0.93
<b>SES Status</b>	Non SES	37331	673.59	39.42		0.91
	SES	28165	649.45	40.18	0.61	0.92
<b>Disability</b>	No Disability	55750	668.83	37.91		0.91
	Disability	9748	631.03	46.24	0.96	0.93
<b>Accommodations</b>	No Accommodations	58003	668.43	38.27		0.91
	Accommodations	7495	622.79	43.32	1.19	0.92
<b>Migrant Status</b>	Non Migrant	65389	663.25	41.49		0.92
	Migrant	109	635.92	40.76	0.66	0.92

**Table 4. 34: Impact Analysis, Grade 6 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	50439	682.81	38.90		0.91
	Black (not Hispanic)	12013	649.86	41.79	0.83	0.91
	Hispanic	2285	663.43	40.79	0.50	0.91
	Asian/Pacific Islander	1214	695.77	44.73	-0.33	0.92
	Native American	276	668.74	39.05	0.36	0.91
	Unknown	105	654.59	49.68	0.72	0.92
<b>Gender</b>	Male	34107	676.25	43.56		0.92
	Female	32038	676.52	39.62	-0.01	0.91
<b>ELL Status</b>	Non ELL	64388	677.27	41.29		0.92
	ELL	1944	644.47	44.31	0.79	0.91
<b>IEP Status</b>	Non IEP	57184	682.45	37.49		0.90
	IEP	9148	637.89	46.31	1.15	0.91
<b>SES Status</b>	Non SES	39074	686.42	39.36		0.91
	SES	27256	661.82	40.80	0.62	0.91
<b>Disability</b>	No Disability	57036	682.49	37.49		0.90
	Disability	9296	638.39	46.33	1.14	0.92
<b>Accommodations</b>	No Accommodations	58841	682.01	37.74		0.90
	Accommodations	7491	631.49	44.58	1.33	0.90
<b>Migrant Status</b>	Non Migrant	66128	676.35	41.75		0.92
	Migrant	204	663.42	42.17	0.31	0.92

**Table 4. 35: Impact Analysis, Grade 7 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	51276	684.95	39.45		0.92
	Black (not Hispanic)	12519	647.71	41.30	0.94	0.91
	Hispanic	2191	662.91	40.37	0.56	0.91
	Asian/Pacific Islander	1060	697.17	47.36	-0.31	0.93
	Native American	310	673.72	34.35	0.29	0.89
	Unknown	198	661.68	43.01	0.59	0.92
<b>Gender</b>	Male	34577	676.68	44.02		0.93
	Female	32666	678.38	41.02	-0.04	0.92
<b>ELL Status</b>	Non ELL	65752	678.26	42.28		0.92
	ELL	1802	646.31	43.25	0.76	0.92
<b>IEP Status</b>	Non IEP	58671	683.75	38.41		0.92
	IEP	8883	635.51	45.27	1.23	0.91
<b>SES Status</b>	Non SES	40839	687.32	40.59		0.92
	SES	26715	662.26	41.18	0.61	0.92
<b>Disability</b>	No Disability	58509	683.78	38.39		0.92
	Disability	9045	636.18	45.52	1.21	0.91
<b>Accommodations</b>	No Accommodations	60392	682.98	38.92		0.92
	Accommodations	7162	630.43	43.53	1.37	0.90
<b>Migrant Status</b>	Non Migrant	67358	677.42	42.63		0.93
	Migrant	196	674.26	39.89	0.07	0.92

**Table 4. 36: Impact Analysis, Grade 8 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	53379	705.46	39.15		0.93
	Black (not Hispanic)	13156	669.90	40.54	0.90	0.90
	Hispanic	2147	686.72	38.61	0.48	0.92
	Asian/Pacific Islander	1051	720.79	42.63	-0.39	0.94
	Native American	331	692.63	43.81	0.33	0.93
	Unknown	140	675.50	45.81	0.76	0.93
<b>Gender</b>	Male	35901	697.72	43.83		0.94
	Female	34086	699.12	39.83	-0.03	0.93
<b>ELL Status</b>	Non ELL	68410	699.16	41.63		0.93
	ELL	1794	666.73	43.10	0.78	0.91
<b>IEP Status</b>	Non IEP	60724	705.04	37.12		0.93
	IEP	9480	655.36	45.74	1.29	0.90
<b>SES Status</b>	Non SES	43434	707.54	40.13		0.93
	SES	26769	683.39	40.60	0.60	0.92
<b>Disability</b>	No Disability	60650	705.04	37.09		0.93
	Disability	9554	655.75	46.02	1.28	0.90
<b>Accommodations</b>	No Accommodations	62648	704.04	37.74		0.93
	Accommodations	7556	651.00	45.35	1.42	0.89
<b>Migrant Status</b>	Non Migrant	70146	698.35	41.98		0.93
	Migrant	58	681.16	37.72	0.41	0.93

**Table 4. 37: Impact Analysis, Grade 10 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Effect Size</b>	<b>Coefficient Alpha</b>
<b>Ethnicity</b>	White (not Hispanic)	53602	731.14	44.47		0.93
	Black (not Hispanic)	11409	689.61	47.22	0.92	0.91
	Hispanic	1803	708.48	46.63	0.51	0.93
	Asian/Pacific Islander	1049	745.78	49.97	-0.33	0.94
	Native American	323	718.74	45.08	0.28	0.93
	Unknown	144	690.57	49.86	0.91	0.92
<b>Gender</b>	Male	34132	723.55	50.17		0.94
	Female	33902	724.07	45.27	-0.01	0.93
<b>ELL Status</b>	Non ELL	66587	724.56	47.49		0.94
	ELL	1743	690.48	48.98	0.72	0.92
<b>IEP Status</b>	Non IEP	59889	731.28	42.15		0.93
	IEP	8441	669.78	50.96	1.42	0.90
<b>SES Status</b>	Non SES	46863	732.53	45.37		0.93
	SES	21465	704.38	47.37	0.61	0.92
<b>Disability</b>	No Disability	59837	731.32	42.12		0.93
	Disability	8493	669.93	51.05	1.42	0.90
<b>Accommodations</b>	No Accommodations	62277	729.37	43.56		0.93
	Accommodations	6053	665.19	50.47	1.54	0.89
<b>Migrant Status</b>	Non Migrant	68264	723.73	47.81		0.94
	Migrant	66	679.73	46.62	0.92	0.91

## **Part 5: Construct Validity**

Construct validity—the meaning of test scores and the inferences they support—is the central concept underlying the MAP validation process. Evidence for construct validity is comprehensive and integrates evidence from both content- and criterion-related validity. In this section, CTB presents evidence of construct validity through the minimization of construct-irrelevant variance and under-representation in the test-development process and through studies of test reliability, convergent validity, and divergent validity

### ***Minimization of Construct-irrelevant Variance and Under-representation***

Construct-irrelevant variance refers to error variance that is caused by factors unrelated to the constructs measured by the test. For example, when tests are not administered under standardized conditions (e.g., one administration may be timed, but another administration may be untimed), differences in student performance related to different administration conditions may result. Careful specification of content and review of the items representing that content are first steps in minimizing construct-irrelevant variance.

Construct under-representation occurs when the content of the assessment does not reflect the full range of content that the assessment is expected to cover. The MAP is designed to represent the Show-Me Standards/GLE strands. Specification and review, in which test blueprints are developed and reviewed, are primary steps in the development process designed to ensure that content is equitably represented.

Part 3 of this Technical Report addresses the content-development process that was followed for all of the MAP tests. In Part 3, we described all the content-development steps taken that minimize construct-irrelevant variance and construct under-representation, including 1) specification, 2) item writing, 3) review, 4) field testing, and 5) test construction.

Further empirical evidence is then gathered to assess construct irrelevance; in other words, evidence of construct validity is gathered. This evidence includes the results of the calibration (see Part 6). Because Item Response Theory models were used to calibrate test items and to report student scores, item fit is relevant to construct validity. The extent to which test items function as the IRT model prescribes is relevant to the validation of test scores. As shown in Part 6, only 12 items across all grade/content areas were flagged for poor model/data fit across all 14 grade/content area MAPs.

Other evidence is gathered through special studies of the data (such as the studies in this chapter). This section summarizes additional statistics that contribute to the evidence for construct validity, including measures of reliability, convergent validity, and divergent validity.

## ***Reliability***

A common way to conceptualize reliability is the consistency of the students' test scores on parallel forms of the same test when they are administered under the same conditions. A reliable test produces scores that are relatively stable if the test is administered repeatedly under similar conditions. It is usually impractical to administer multiple forms of the test, and reliability is estimated on a single administration of the test. This type of reliability, known as internal consistency, provides an estimate of how consistently examinees perform across items within a test during a single test administration (Crocker & Algina, 1986). Reliability is a necessary but not a sufficient condition of validity.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) indicate that:

. . . reliability evidence may be reported in terms of variances or standard deviations of measurement errors, in terms of one or more coefficients, or in terms of IRT-based test information functions. (p. 27)

In accordance with the AERA/APA/NCME *Standards* and developing and maintaining tests of the highest quality, CTB has calculated the reliability of each MAP test in a variety of ways: reliability of raw scores, overall standard error of measurement, IRT-based conditional standard error of measurement, and decision consistency of achievement level classifications.

## **Test Reliability**

The reliability of raw scores on the MAP tests was evaluated using Cronbach's (1951) Alpha, which is a lower-bound estimate of test reliability. The reliability coefficient is a ratio of the variance of true test scores to those of the observed scores, with the values ranging from 0 to 1. The closer the value of the reliability coefficient is to 1, the more consistent the scores, where 1.00 refers to a perfectly consistent test. As a rule of thumb, reliability coefficients that are equal to or greater than 0.9 are considered acceptable for tests of lengths similar to the MAP.

Cronbach's Alpha was computed using the formula

$$C_{\alpha} = \frac{n}{n-1} \left[ 1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_x^2} \right],$$

where  $n$  is the number of items on the test,  $\sigma_i^2$  is the variance of item  $i$  and  $\sigma_x^2$  is the variance of the total test score.

Total test reliability measures, such as Cronbach's Alpha and standard error of measurement, consider the consistency (reliability) of performance over all test questions in a given form—the results of which imply how well the questions measure the content domain and could continue to do so over repeated administrations. The number of items in the test influences these statistics; a longer test can be expected to be more reliable than a shorter test.

The reliability coefficients for the MAP are reported in Tables 5.1 and 5.2. These reliability coefficients were computed using the calibration sample selected by DESE. All reliability statistics are over .90 for all tests indicating acceptable reliability. Tables 5.3 and 5.4 display the reliability coefficients by subgroup at all grade levels in Communication Arts and Mathematics, respectively. The results in these tables show that the tests have acceptable reliability for all subgroups.

### **Standard Error of Measurement**

The reliability of reported test scores can be characterized by the standard errors associated with the scores. The standard error of measurement (SEM) may be used to determine the range within which a student's observed score would likely be for a given true score. An observed score should be regarded not as a student's true score, but as an estimate of a student's true score. It is expected that 68% of the time a student's score obtained from a single test administration would fall within one SEM of the student's true score and that 95% of the time the obtained score would fall within approximately two standard errors of the true score.

Figures 5.1 through 5.14 display the conditional SEM curves for each grade/content area. The estimates of measurement error tend to be higher at the low and high ends of the scale score range. The test is designed so that measurement error is minimized in the middle of the scale range where the majority of students are located; thus, the measurement error increases on the high and low end of the ability range.

### **Decision Accuracy and Consistency**

The *Standards* also make reference to an additional measurement issue that concerns evidence for validity:

Some authorities have proposed that a semantic distinction be made between “reliability of scores” and “degree of agreement in classification.” The former term would be reserved for analysis of score variation under repeated measurement. The term *classification consistency* . . . , rather than reliability, would be used in discussions of consistency of classification. Adoption of such usage would make it clear that the importance of an error of any given size depends on the proximity of the examinee's score to the cut score. (p. 30)

*Decision Consistency:* Classification consistency or decision consistency is defined as the extent to which the classifications of students agree on the basis of multiple administrations of the test, either with the same form or with parallel forms. However, it is usually not feasible to obtain data from repeated administrations because of cost,

testing time, and (in the case of using the same form) students' recall of the first administration. Therefore, a common practice is to estimate decision consistency from one administration of a test.

*Decision Accuracy:* Decision accuracy is defined as the extent to which the actual classifications of test takers agree with classifications that would be made on the basis of their true scores (Livingston & Lewis, 1995). That is, decision consistency refers to the agreement between two observed scores, while decision accuracy refers to the agreement between the observed score and the true score. It is common to estimate decision accuracy by assuming the psychometric model (see below) to find true scores corresponding to observed scores.

CTB used the Livingston-Lewis (1995) methodology to calculate these statistics on the 2007 MAP results. The Livingston-Lewis procedure utilizes a beta-binomial model that requires two steps: 1) fitting proportion-correct true scores to a four-parameter beta distribution; and 2) using the binomial distribution to estimate classification accuracy and consistency. All calculations for decision accuracy and consistency are based on census data.

Table 5.5 reports the decision accuracy and consistency classifications conditioned on each level of achievement for each grade/content area. In Table 5.5, the accuracy conditioned on level of achievement indicates the percentage of students correctly classified into a level of achievement given their true score status. For example, 88% of the Grade 8 Mathematics students who were estimated to have a true status of *Below Basic* were correctly classified on the Grade 8 Mathematics test by their observed score.

In Table 5.5, the consistency conditioned on level of achievement indicates the percentage of students whose classification would be in the same level of achievement based on a hypothetical alternate form of the test. For example, 62% of Grade 7 Communication Arts students whose performance was classified as *Proficient* would be classified in the same level based on the hypothetical alternate form, if they had taken it.

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. These results are reported in Table 5.6. To evaluate decisions at specific cut points, the joint distribution of all performance levels is collapsed into a dichotomized distribution around that specific cut point. As an example, the dichotomization at the cut point between the *Basic* and *Proficient* classifications was formed. The proportion of correct classifications below this particular cut point is equal to the sum of all the cells at the levels *Below Basic* and *Basic*, and the proportion of correct classifications above that particular cut point is equal to the sum of all the cells at the levels *Proficient* and *Advanced*. As shown in Table 5.6, all accuracy statistics conditioned on cut point are above 90%, and all decision statistics conditioned on cut point are above 85%.

## ***Convergent Validity***

Convergent validity is a subtype of construct validity. Convergent validity occurs when theoretically-related constructs are observed to be related. Analyses of the internal structure of a test can indicate the extent to which the relationships among test items conform to the construct the test purports to measure. For example, the MAP Mathematics test is designed to measure a single overall construct—Mathematics achievement; therefore, the items comprising the Mathematics MAP should only measure Mathematics, not Science, Language, or Reading.

## **Principal Components Analysis**

As another measure of construct validity, CTB examined the unidimensionality of each grade-level MAP test. The test is unidimensional if items comprising the test measure a single content domain. For example, mathematics items should measure mathematics ability and not measure reading skills. When a test is shown to measure a single construct (i.e. the test is unidimensional), then this provides further evidence of construct validity. If the test is not reasonably unidimensional, then multiple factors are needed to define and explain student performance on the test.

One of the underlying assumptions of the IRT models used to scale the MAP is that the tests being calibrated are unidimensional. The unidimensionality assumption is in practice a testable hypothesis that is commonly evaluated through Principal Components Analysis (PCA). This analysis seeks evidence that there exists a single primary factor, the first principal component, which accounts for much of the relationship between items. The presence of a single or dominant factor suggests that a test is sufficiently unidimensional (i.e., measures one underlying construct).

A principal components factor analysis was conducted on each grade/content area MAP. A large first principal component is evident in each analysis. In Figures 5.15 to 5.28, scree plots (Cattell 1966) of eigenvalues are presented to illustrate the relative dominance of the first principal component in each of the MAP tests. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other minor factors.

For all grade/content area MAPs, the ratio of the variance accounted for by the first factor to the second and third is sufficiently large to support the claim that these tests are essentially unidimensional. All of the MAP subject area tests exhibit first principal components accounting for more than 10% of the test variance (see Tables 5.7 and 5.8). To further investigate the unidimensionality of the Communication Arts and Mathematics tests, the ratio of first eigenvalue to the second eigenvalue was explored (see Tables 5.7 and 5.8). These ratios show that the first eigenvalue is at least three times as large as the second eigenvalue for most of the grade/content areas. This substantial difference in magnitude indicates that one factor appears to be dominant and that the Communication Arts and Mathematics tests are essentially one-dimensional.

This evidence supports the claim that there is a dominant dimension underlying the items/tasks in each test and that scores from each test represent performance primarily

determined by that ability. Construct-irrelevant factors such as factual knowledge irrelevant for doing well in a subject does not appear to significantly affect performance.

### **Analyses by Content Standard**

Two sets of analyses were conducted for the content standard level to help assess the construct validity of the MAP. First, the reliability of each Content Standard was computed. Second, correlation coefficients that measure the relationship between the Content Standards were computed.

### **Reliability of Content Standards**

Cronbach's (1951) Alpha was computed for each of the Content Standards by grade/content area using the census data. Tables 5.10 through 5.23 report the reliability statistics along the diagonal of each matrix for each grade/content area. Reliability indices, such as Cronbach's Alpha, are a function of the number of test items. It is expected that Cronbach's Alpha would be low for a Content Standard assessed by a small number of items (e.g., Writing Formally and Informally).

### **Correlations among Content Standard Subscores**

In this section, we measure the strength of the interrelationships among the Content Standards by computing correlation among the content standards. Tables 5.9 through 5.22 report the uncorrected Pearson product-moment (PPM) correlation coefficients, the PPM corrected for attenuation (CAPPM), in addition to the reliability coefficients described above. The PPM among the Content Standard subscores is presented below the diagonal portion of the matrix, the CAPPM is presented above the diagonal portion of the matrix, and the reliability coefficients are shown on the diagonal in each table. When the Content Standard is comprised of one item, then no reliability coefficient is reported for that subscore.

The uncorrected correlation coefficients (PPM) in Tables 5.9 to 5.22 should be interpreted in the context of the reliability coefficient which is directly affected by the number of items. In general, we expect to see lower PPM coefficients between variables that are less reliable. Overall, the PPM coefficients show that performance on one Content Standard is moderately-to-strongly related to performance on another Content Standard within the same content area. As noted above, the value of the correlation coefficients will be affected by the limited number of items measuring each Content Standard. So, caution should be used when comparing the PPM coefficients measuring the relationships between Content Standards to those measuring the relationships between content areas which have larger numbers of items (Table 5.23). We expect to see a weaker relationship (smaller correlation coefficients) reported between the Content Standards as a consequence of the fewer number of items measuring each content standard (e.g., Writing Formally and Informally).

Indeed, the PPM between two content standard subscores may be artificially low because of measurement error. We can correct for the attenuation of the PPM statistically using Spearman's (1904, 1910) formula:

$$CAPP\text{M} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$$

Where  $r_{xy}$  is the PPM between two content standards,  $r_{xx}$  is the reliability of one of those content standards, and  $r_{yy}$  is the reliability for the other content standard. For Content Standards comprised of one item, no correction can be applied to the correlation coefficient (because no reliability coefficient is available).

Across all tables, the CAPPM indicate strong relationships between the content standards. In some cases, the CAPPM is greater than 1.0. “Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed” (Schumacker, 1996). The strong relationships suggested by the CAPPM in Tables 5.9 to 5.22 are further evidence of the validity of the test construct. Since the overall content area is comprised of the content standard subscores and the content area is expected to measure a single dimension, then we would expect that these subscores are also highly related.

### ***Divergent (Discriminant) Validity***

Measures of different constructs should not be highly correlated with each other. Divergent validity is a subtype of construct validity that can be assessed by the extent to which measures of constructs that theoretically should not be related to each other are, in fact, observed as not related to each other. Typically, correlation coefficients among measures of unrelated or distantly related constructs are examined in support of divergent validity.

To assess the divergent validity of the MAP tests, correlations were computed between the Math and Communication Arts scale scores for students who took both of the MAP subject area tests in 2007. These correlations are based on the census data and the results are shown in Table 5.23. The correlation coefficients ranged from 0.75 (between Communication Arts and Mathematics in Grades 4 and 5) to 0.78 (between Communication Arts and Mathematics in Grade 8). The correlation coefficients suggest that individual student scores for Communication Arts and Mathematics are moderately- to highly-related constructs. These correlation coefficients are similar to what was seen in 2006.

**Table 5. 1: Reliability, Communication Arts**

<b>Grade</b>	<b>Number of Items</b>	<b>Number of Score Points</b>	<b>Cronbach’s Alpha</b>
<b>3</b>	57	69	0.91
<b>4</b>	56	66	0.90
<b>5</b>	56	66	0.91
<b>6</b>	56	65	0.90
<b>7</b>	61	73	0.91
<b>8</b>	60	69	0.91
<b>11</b>	60	73	0.91



**Table 5. 2: Reliability, Mathematics**

Grade	Number of Items	Number of Score Points	Cronbach's Alpha
3	60	67	0.91
4	65	77	0.93
5	62	71	0.93
6	61	69	0.92
7	62	71	0.92
8	64	76	0.93
10	61	75	0.94

**Table 5. 3: Cronbach's Alpha by Subgroup, Communication Arts**

	Group	Grade Level						
		3	4	5	6	7	8	11
Ethnicity	White (not Hispanic)	0.90	0.90	0.90	0.90	0.90	0.91	0.90
	Black (not Hispanic)	0.92	0.91	0.91	0.91	0.91	0.91	0.90
	Hispanic	0.91	0.90	0.90	0.91	0.91	0.91	0.90
	Asian/Pacific Islander	0.91	0.91	0.91	0.90	0.92	0.92	0.92
	Native American	0.88	0.91	0.91	0.91	0.90	0.92	0.91
Gender	Male	0.91	0.91	0.91	0.91	0.92	0.92	0.91
	Female	0.90	0.90	0.90	0.90	0.90	0.91	0.90
ELL Status	Non-ELL	0.91	0.90	0.91	0.91	0.91	0.92	0.91
	ELL	0.92	0.91	0.91	0.91	0.91	0.91	0.90
IEP Status	Non-IEP	0.89	0.87	0.88	0.88	0.89	0.90	0.89
	IEP	0.93	0.93	0.92	0.92	0.91	0.91	0.89
SES Status	Non-SES	0.89	0.89	0.89	0.89	0.90	0.91	0.90
	SES	0.91	0.91	0.91	0.91	0.91	0.92	0.90
Disability	No Disability	0.89	0.87	0.88	0.88	0.89	0.90	0.89
	Disability	0.93	0.93	0.92	0.92	0.91	0.91	0.89
Accommodations	No Accommodations	0.89	0.88	0.88	0.88	0.89	0.90	0.89
	Accommodations	0.92	0.92	0.90	0.91	0.90	0.90	0.88
Migrant Status	Non-Migrant	0.91	0.91	0.91	0.91	0.91	0.92	0.91
	Migrant	0.89	0.92	0.91	0.93	0.92	0.89	0.88

**Table 5. 4: Cronbach’s Alpha by Subgroup, Mathematics**

	Group	Grade Level						
		3	4	5	6	7	8	10
Ethnicity	White (not Hispanic)	0.90	0.92	0.91	0.91	0.92	0.93	0.93
	Black (not Hispanic)	0.92	0.93	0.92	0.91	0.91	0.90	0.91
	Hispanic	0.90	0.92	0.91	0.91	0.91	0.92	0.93
	Asian/Pacific Islander	0.91	0.93	0.92	0.92	0.93	0.94	0.94
	Native American	0.89	0.92	0.92	0.91	0.89	0.93	0.93
Gender	Male	0.91	0.93	0.92	0.92	0.93	0.94	0.94
	Female	0.91	0.92	0.92	0.91	0.92	0.93	0.93
ELL Status	Non-ELL	0.91	0.93	0.92	0.92	0.92	0.93	0.94
	ELL	0.92	0.93	0.92	0.91	0.92	0.91	0.92
IEP Status	Non-IEP	0.90	0.91	0.91	0.90	0.92	0.93	0.93
	IEP	0.93	0.94	0.93	0.91	0.91	0.90	0.90
SES Status	Non-SES	0.90	0.91	0.91	0.91	0.92	0.93	0.93
	SES	0.91	0.93	0.92	0.91	0.92	0.92	0.92
Disability	No Disability	0.90	0.91	0.91	0.90	0.92	0.93	0.93
	Disability	0.93	0.94	0.93	0.92	0.91	0.90	0.90
Accommodations	No Accommodations	0.90	0.91	0.91	0.90	0.92	0.93	0.93
	Accommodations	0.92	0.93	0.92	0.90	0.90	0.89	0.89
Migrant Status	Non-Migrant	0.91	0.93	0.92	0.92	0.93	0.93	0.94
	Migrant	0.90	0.92	0.92	0.92	0.92	0.93	0.91

**Table 5. 5: Decision Accuracy and Consistency Conditioned on Level of Achievement, Communication Arts and Mathematics**

Content Area	Grade	Accuracy				Consistency			
		Below Basic	Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.
Communication Arts	3	0.83	0.84	0.64	0.85	0.71	0.79	0.53	0.76
	4	0.83	0.80	0.65	0.84	0.71	0.73	0.54	0.75
	5	0.83	0.82	0.66	0.85	0.70	0.76	0.55	0.76
	6	0.83	0.82	0.69	0.84	0.72	0.76	0.59	0.73
	7	0.83	0.81	0.72	0.84	0.72	0.75	0.62	0.73
	8	0.83	0.85	0.66	0.85	0.70	0.81	0.56	0.76
	11	0.83	0.85	0.73	0.83	0.72	0.80	0.64	0.72
	Mathematics	3	0.82	0.85	0.76	0.83	0.69	0.80	0.68
4		0.84	0.86	0.79	0.85	0.72	0.81	0.71	0.74
5		0.83	0.86	0.76	0.86	0.71	0.81	0.68	0.76
6		0.84	0.82	0.76	0.84	0.73	0.76	0.67	0.74
7		0.86	0.79	0.75	0.85	0.77	0.72	0.66	0.75
8		0.88	0.77	0.70	0.86	0.81	0.70	0.60	0.77
11		0.89	0.77	0.79	0.85	0.83	0.68	0.71	0.75

**Table 5. 6: Decision Accuracy and Consistency at Achievement Cut Points, Communication Arts and Mathematics**

Content Area	Grade	Accuracy			Consistency		
		Below Basic/ Basic	Basic/ Prof.	Prof./ Adv.	Below Basic/ Basic	Basic/ Prof.	Prof./ Adv.
Communication Arts	3	0.96	0.91	0.93	0.94	0.87	0.91
	4	0.95	0.90	0.93	0.92	0.86	0.90
	5	0.95	0.90	0.93	0.94	0.86	0.90
	6	0.94	0.90	0.94	0.92	0.86	0.92
	7	0.94	0.90	0.94	0.92	0.86	0.92
	8	0.96	0.91	0.94	0.94	0.87	0.91
	11	0.95	0.91	0.95	0.93	0.87	0.93
Mathematics	3	0.96	0.90	0.95	0.95	0.87	0.93
	4	0.96	0.91	0.96	0.95	0.88	0.94
	5	0.97	0.91	0.95	0.95	0.88	0.92
	6	0.95	0.91	0.95	0.93	0.87	0.92
	7	0.94	0.91	0.95	0.91	0.88	0.93
	8	0.93	0.92	0.95	0.90	0.88	0.93
	10	0.93	0.92	0.96	0.90	0.89	0.95

**Table 5. 7: Principal Component Analysis, Communication Arts**

<b>Grade</b>	<b>Eigenvalue</b>	<b>Percent of Variance Explained</b>	<b>Cumulative Percent of Variance Explained</b>
<b>Grade 3</b>			
First Component	10.08	17.68	17.68
Second Component	1.84	3.23	20.91
Ratio (First/Second)	5.47		
<b>Grade 4</b>			
First Component	9.60	17.13	17.13
Second Component	1.91	3.42	20.55
Ratio (First/Second)	5.01		
<b>Grade 5</b>			
First Component	10.07	17.98	17.98
Second Component	1.49	2.65	20.63
Ratio (First/Second)	6.78		
<b>Grade 6</b>			
First Component	9.70	17.32	17.32
Second Component	1.42	2.53	19.85
Ratio (First/Second)	6.85		
<b>Grade 7</b>			
First Component	10.08	16.52	16.52
Second Component	1.99	3.25	19.78
Ratio (First/Second)	5.08		
<b>Grade 8</b>			
First Component	10.59	17.64	17.64
Second Component	1.80	3.00	20.64
Ratio (First/Second)	5.88		
<b>Grade 11</b>			
First Component	9.55	15.91	15.91
Second Component	1.73	2.88	18.79
Ratio (First/Second)	5.53		

**Table 5. 8: Principal Component Analysis, Mathematics**

<b>Grade</b>	<b>Eigenvalue</b>	<b>Percent of Variance Explained</b>	<b>Cumulative Percent of Variance Explained</b>
<b>Grade 3</b>			
First Component	9.97	16.61	16.61
Second Component	2.11	3.52	20.14
Ratio (First/Second)	4.71		
<b>Grade 4</b>			
First Component	12.22	18.80	18.80
Second Component	1.87	2.88	21.68
Ratio (First/Second)	6.52		
<b>Grade 5</b>			
First Component	12.04	19.42	19.42
Second Component	1.92	3.10	22.51
Ratio (First/Second)	6.27		
<b>Grade 6</b>			
First Component	10.75	17.62	17.62
Second Component	1.71	2.80	20.42
Ratio (First/Second)	6.29		
<b>Grade 7</b>			
First Component	12.13	19.56	19.56
Second Component	1.96	3.17	22.73
Ratio (First/Second)	6.18		
<b>Grade 8</b>			
First Component	12.34	19.29	19.29
Second Component	1.87	2.92	22.21
Ratio (First/Second)	6.60		
<b>Grade 10</b>			
First Component	13.10	21.47	21.47
Second Component	1.68	2.76	24.23
Ratio (First/Second)	7.78		

**Table 5. 9: Reliability of Each Content Standard and Uncorrected and Corrected Correlations among Content Standards, Grade 3 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.69	0.90	0.92	0.98	0.91
02	Reading Fiction/Poetry/Drama	21	0.67	0.81	1.02	0.96	1.12
03	Reading Nonfiction	21	0.68	0.82	0.80	1.03	1.13
04	Writing Formally & Informally	2	0.50	0.53	0.57	0.38	1.00
05	Combined Reading	40	0.71	0.95	0.95	0.58	0.89

**Table 5. 10: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 4 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	10	0.56	0.89	0.88	--	0.89
02	Reading Fiction/Poetry/Drama	33	0.62	0.87	0.94	--	1.10
03	Reading Nonfiction	12	0.54	0.72	0.68	--	1.12
04	Writing Formally & Informally	1	0.35	0.47	0.41	--	--
05	Combined Reading	45	0.63	0.97	0.87	0.48	0.89

**Table 5. 11: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 5 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	12	0.65	0.89	0.92	--	0.92
02	Reading Fiction/Poetry/Drama	25	0.65	0.82	0.98	--	1.12
03	Reading Nonfiction	19	0.67	0.80	0.81	--	1.10
04	Writing Formally & Informally	1	0.30	0.38	0.33	--	-
05	Combined Reading	43	0.70	0.96	0.93	0.37	0.89

**Table 5. 12: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 6 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	13	0.64	0.89	0.86	--	0.90
02	Reading Fiction/Poetry/Drama	30	0.66	0.85	0.95	--	1.13
03	Reading Nonfiction	12	0.60	0.76	0.76	--	1.07
04	Writing Formally & Informally	1	0.31	0.40	0.34	--	-
05	Combined Reading	42	0.68	0.98	0.88	0.40	0.89

**Table 5. 13: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 7 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	16	0.66	0.84	0.89	0.88	0.89
02	Reading Fiction/Poetry/Drama	12	0.56	0.67	0.96	0.91	1.09
03	Reading Nonfiction	30	0.67	0.73	0.86	0.90	1.12
04	Writing Formally & Informally	3	0.49	0.51	0.57	0.47	0.90
05	Combined Reading	42	0.68	0.84	0.98	0.58	0.89

**Table 5. 14: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 8 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.69	0.89	0.84	--	0.88
02	Reading Fiction/Poetry/Drama	17	0.67	0.82	0.93	--	1.07
03	Reading Nonfiction	27	0.65	0.78	0.86	--	1.09
04	Writing Formally & Informally	1	0.13	0.14	0.15	--	-
05	Combined Reading	44	0.70	0.92	0.96	0.16	0.91

**Table 5. 15: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 11 Communication Arts**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.63	0.91	0.88	0.84	0.89
02	Reading Fiction/Poetry/Drama	10	0.58	0.64	1.01	0.90	1.14
03	Reading Nonfiction	33	0.65	0.75	0.86	0.91	1.12
04	Writing Formally & Informally	2	0.47	0.51	0.6	0.5	0.90
05	Combined Reading	43	0.67	0.86	0.98	0.6	0.89

**Table 5. 16: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 3 Mathematics**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	17	0.74	0.99	0.92	1.00	0.99
02	Algebraic Relationship	10	0.71	0.69	0.93	1.01	0.99
03	Geometric and Spatial	12	0.63	0.61	0.63	0.91	0.95
04	Measurement	11	0.69	0.67	0.58	0.64	0.98
05	Data and Probability	10	0.70	0.67	0.62	0.64	0.67

**Table 5. 17: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 4 Mathematics**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	18	0.79	1.00	0.91	0.97	0.89
02	Algebraic Relationship	10	0.74	0.69	0.98	1.01	0.92
03	Geometric and Spatial	14	0.69	0.69	0.72	0.93	0.93
04	Measurement	12	0.75	0.73	0.69	0.76	0.85
05	Data and Probability	11	0.64	0.62	0.64	0.60	0.66

**Table 5. 18: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 5 Mathematics**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	15	0.72	0.96	0.90	0.91	0.94
02	Algebraic Relationship	12	0.69	0.72	0.94	0.97	0.94
03	Geometric and Spatial	11	0.62	0.65	0.66	0.93	0.94
04	Measurement	13	0.67	0.72	0.66	0.76	0.89
05	Data and Probability	11	0.67	0.67	0.64	0.65	0.70

**Table 5. 19: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 6 Mathematics**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	18	0.80	0.96	0.89	0.90	0.95
02	Algebraic Relationship	13	0.72	0.71	0.94	0.98	1.01
03	Geometric and Spatial	10	0.63	0.63	0.63	0.92	0.97
04	Measurement	8	0.60	0.61	0.54	0.55	0.97
05	Data and Probability	12	0.72	0.72	0.65	0.61	0.72

**Table 5. 20: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 7 Mathematics**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	19	0.82	0.98	0.87	0.98	0.96
02	Algebraic Relationship	10	0.73	0.67	0.92	1.04	1.02
03	Geometric and Spatial	11	0.67	0.64	0.72	1.08	0.96
04	Measurement	10	0.69	0.66	0.71	0.60	1.07
05	Data and Probability	12	0.74	0.71	0.69	0.70	0.72

**Table 5. 21: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 8 Mathematics**

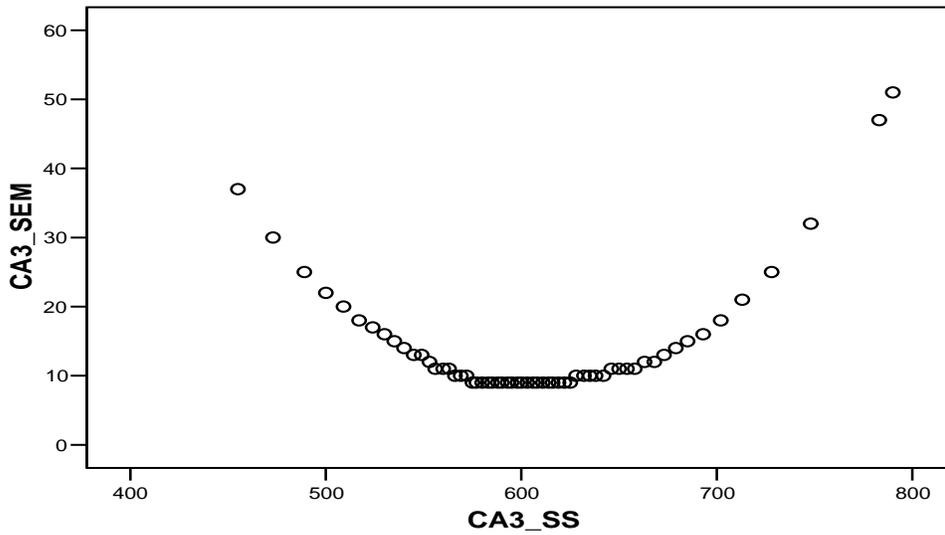
No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	17	0.77	0.96	0.96	0.91	0.99
02	Algebraic Relationship	17	0.75	0.80	1.01	0.96	1.04
03	Geometric and Spatial	10	0.71	0.76	0.71	1.04	1.04
04	Measurement	10	0.66	0.71	0.72	0.68	0.99
05	Data and Probability	10	0.72	0.77	0.73	0.68	0.69

**Table 5. 22: Reliability of Each Content Standard and Uncorrected and Corrected Correlation among Content Standards, Grade 10 Mathematics**

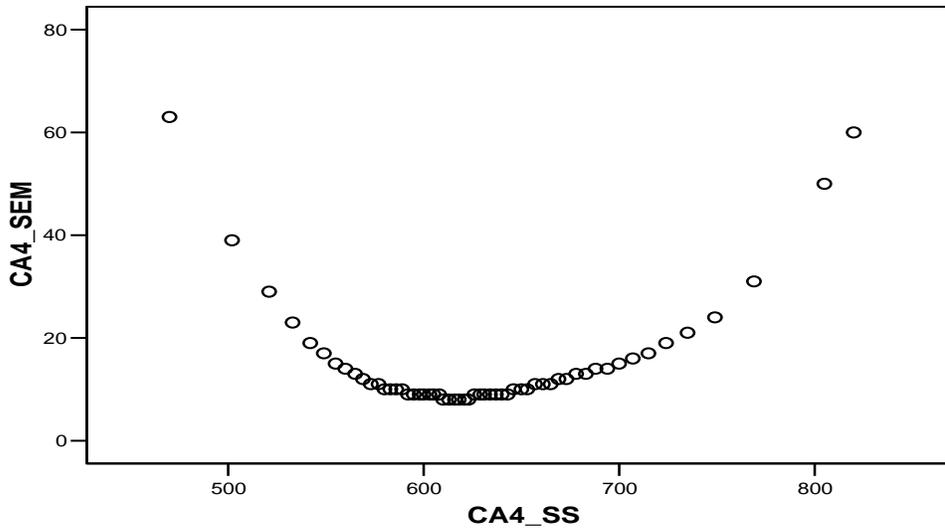
No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	12	0.71	1.00	0.96	0.94	0.99
02	Algebraic Relationship	19	0.76	0.81	1.01	0.99	1.01
03	Geometric and Spatial	11	0.68	0.76	0.70	1.00	1.01
04	Measurement	8	0.69	0.78	0.73	0.76	0.98
05	Data and Probability	11	0.71	0.78	0.72	0.73	0.73

**Table 5. 23: Inter-Correlation of Scale Scores, Communication Arts and Mathematics**

Grade	Correlation
3	0.76
4	0.75
5	0.75
6	0.76
7	0.77
8	0.78



**Figure 5. 1: Standard Error of Measurement Plot, Grade 3 Communication Arts**



**Figure 5. 2: Standard Error of Measurement Plot, Grade 4 Communication Arts**

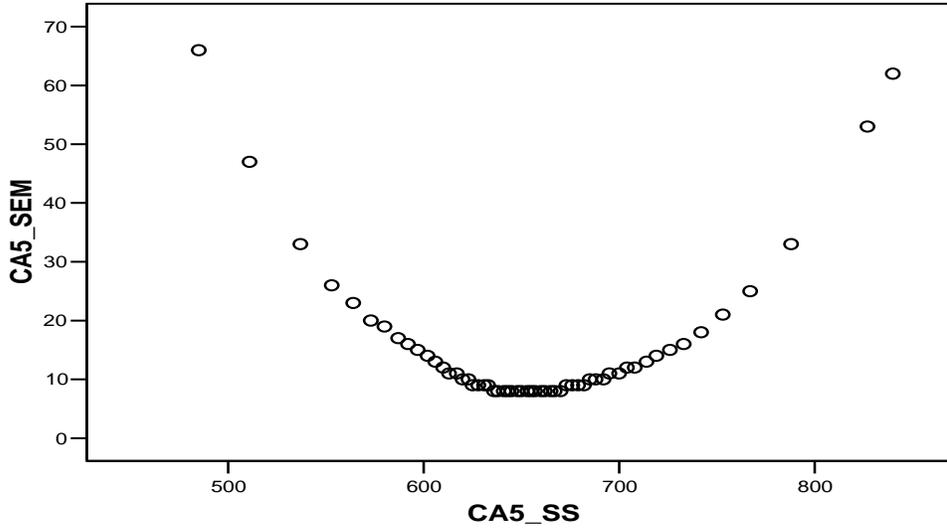


Figure 5. 3: Standard Error of Measurement Plot, Grade 5 Communication Arts

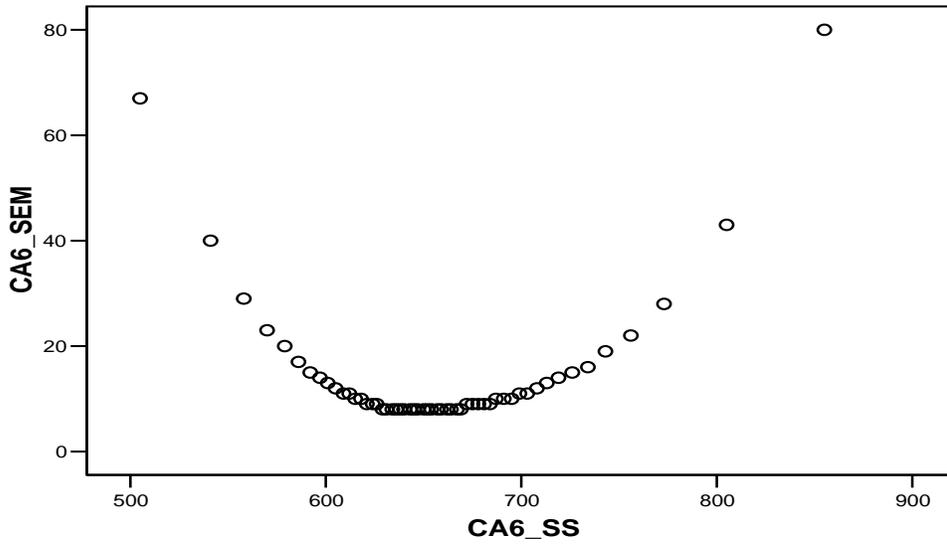


Figure 5. 4: Standard Error of Measurement Plot, Grade 6 Communication Arts

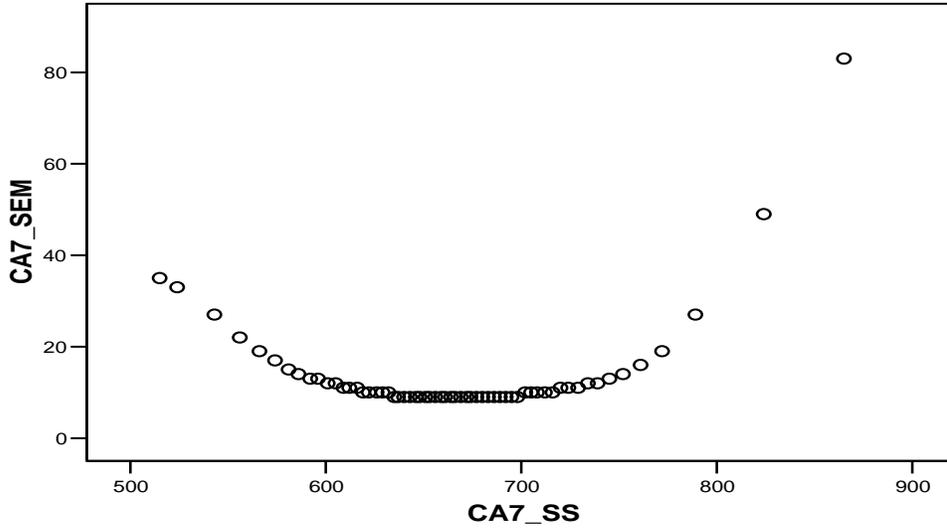


Figure 5. 5: Standard Error of Measurement Plot, Grade 7 Communication Arts

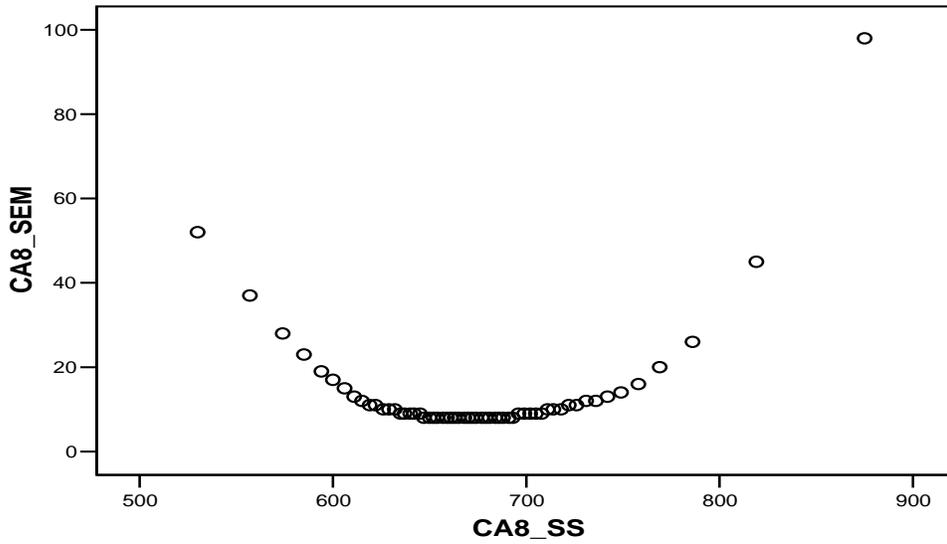


Figure 5. 6: Standard Error of Measurement Plot, Grade 8 Communication Arts

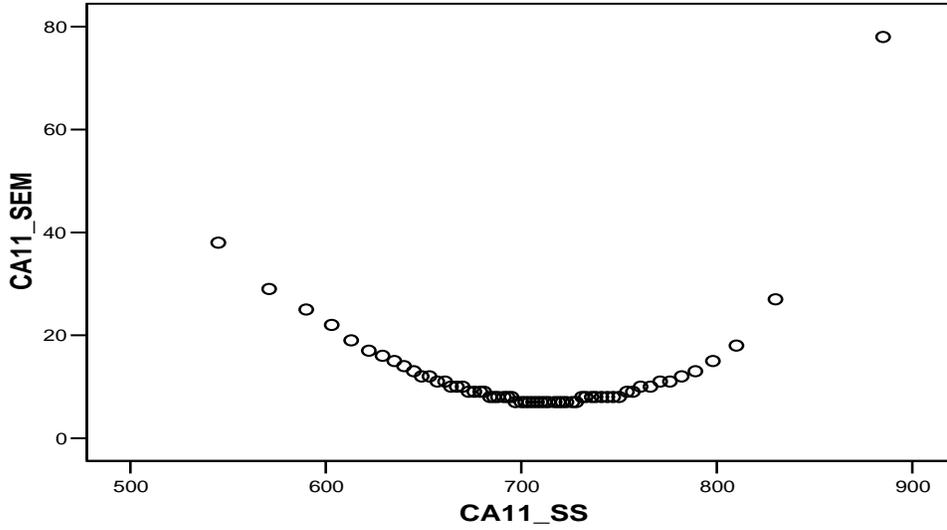


Figure 5. 7: Standard Error of Measurement Plot, Grade 11 Communication Arts

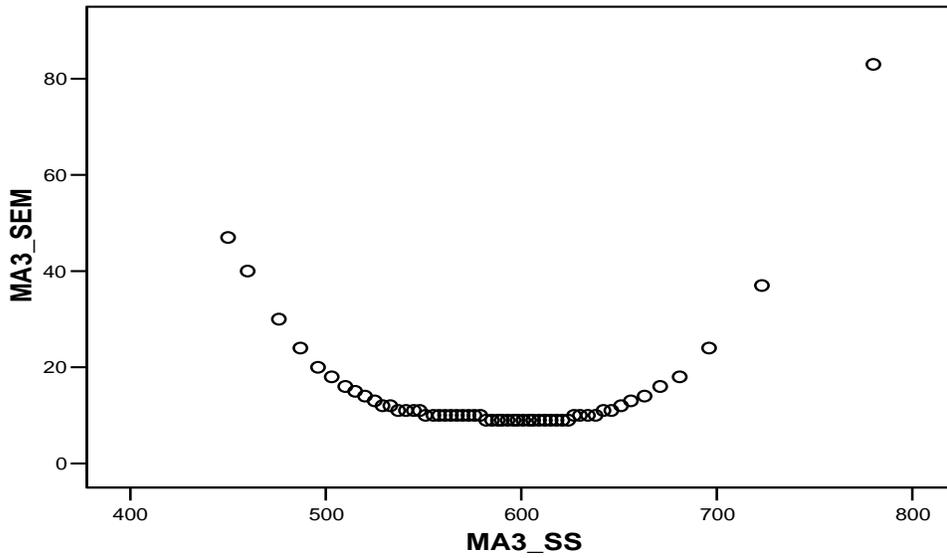


Figure 5. 8: Standard Error of Measurement Plot, Grade 3 Mathematics

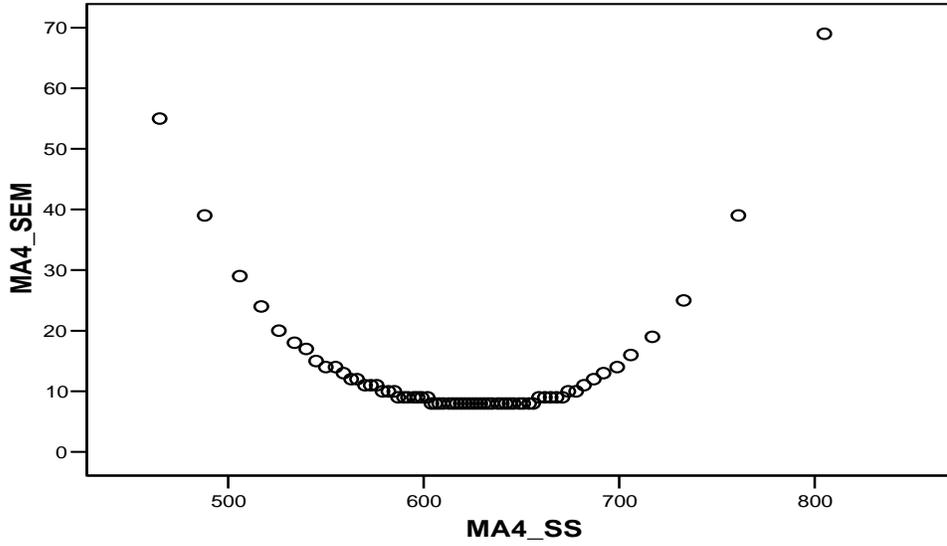


Figure 5. 9: Standard Error of Measurement Plot, Grade 4 Mathematics

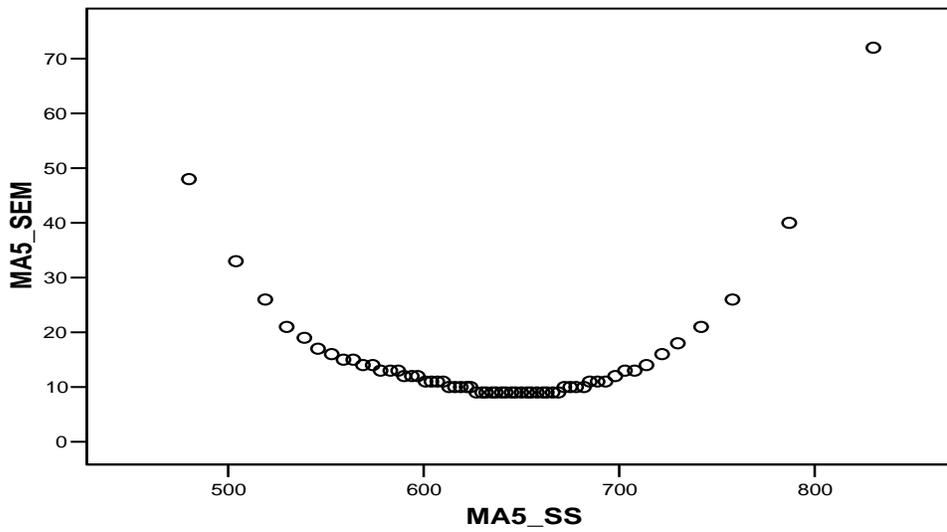


Figure 5. 10: Standard Error of Measurement Plot, Grade 5 Mathematics

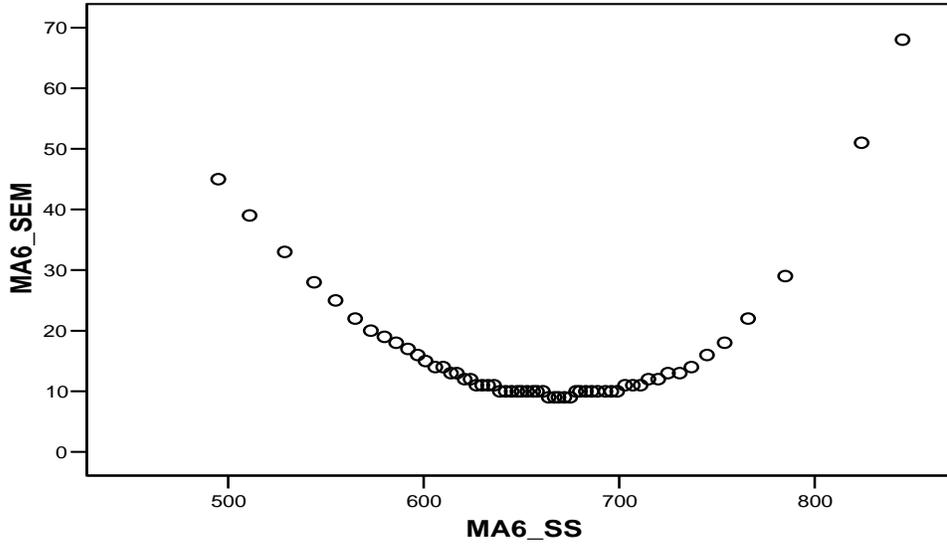


Figure 5. 11: Standard Error of Measurement Plot, Grade 6 Mathematics

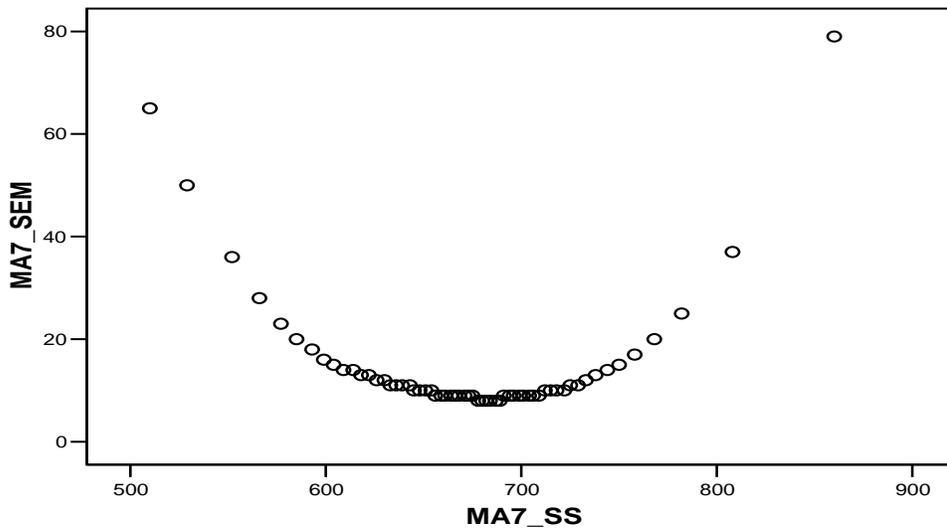


Figure 5. 12: Standard Error of Measurement Plot, Grade 7 Mathematics

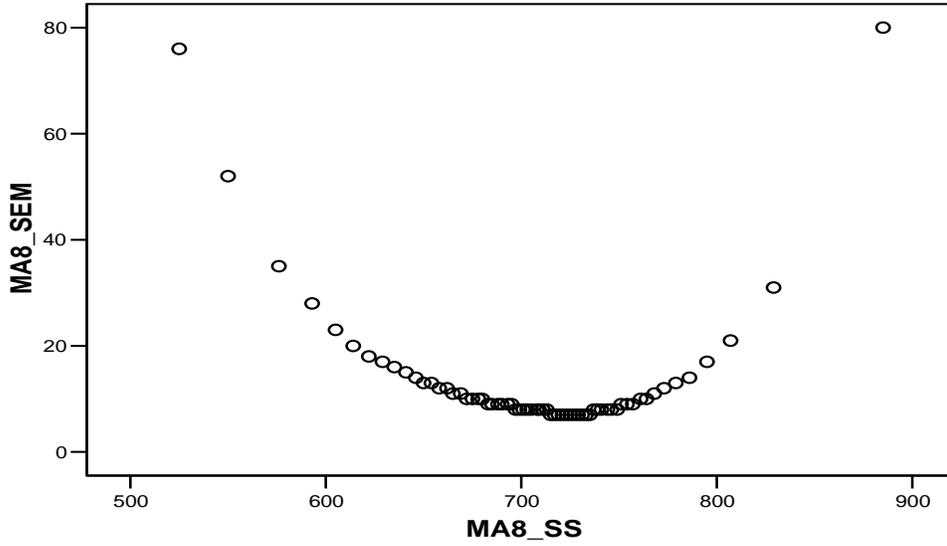


Figure 5. 13: Standard Error of Measurement Plot, Grade 8 Mathematics

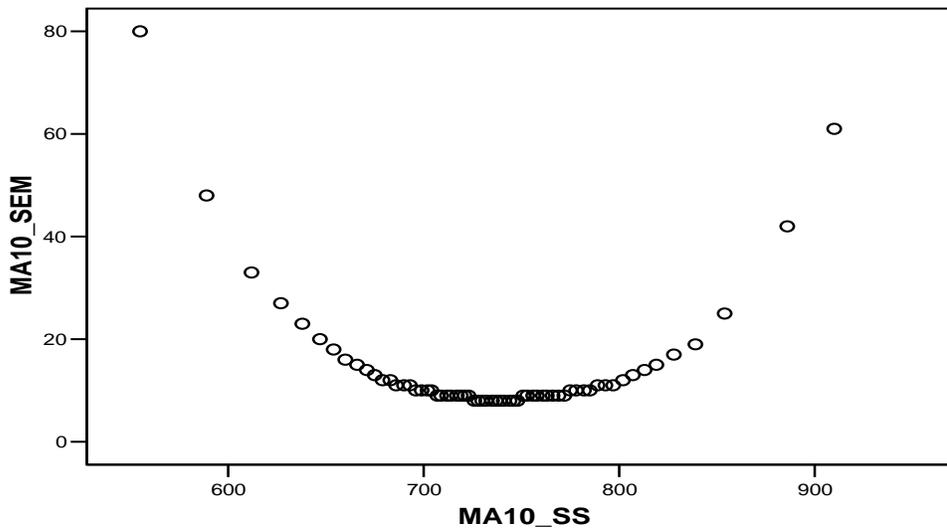


Figure 5. 14: Standard Error of Measurement Plot, Grade 10 Mathematics

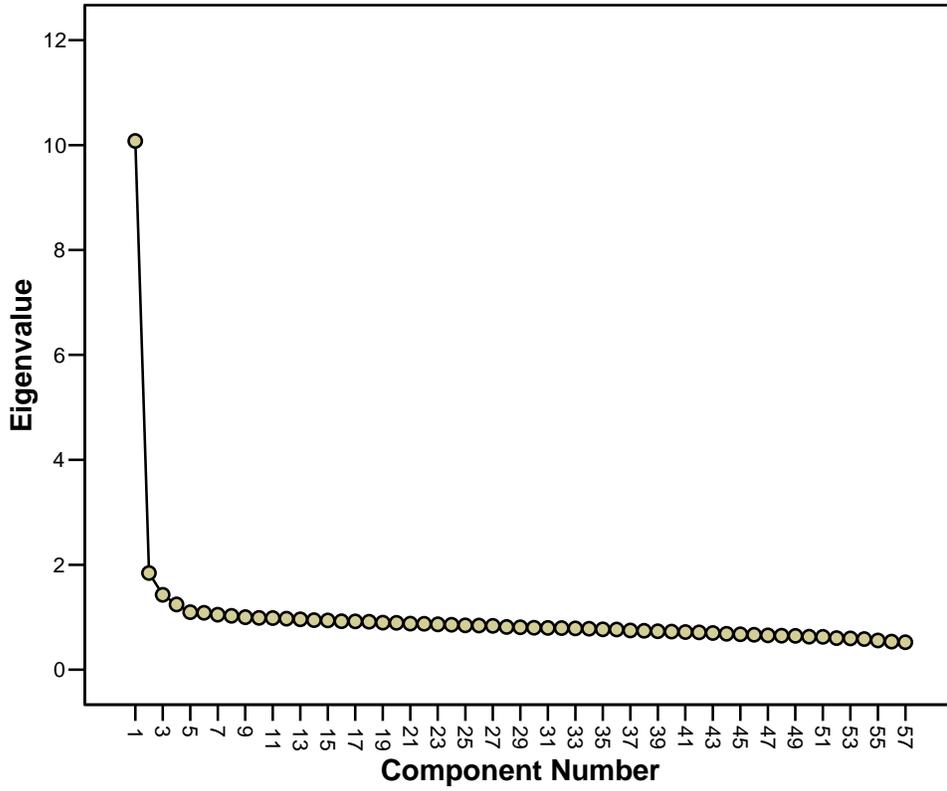


Figure 5. 15: Scree Plot, Grade 3 Communication Arts

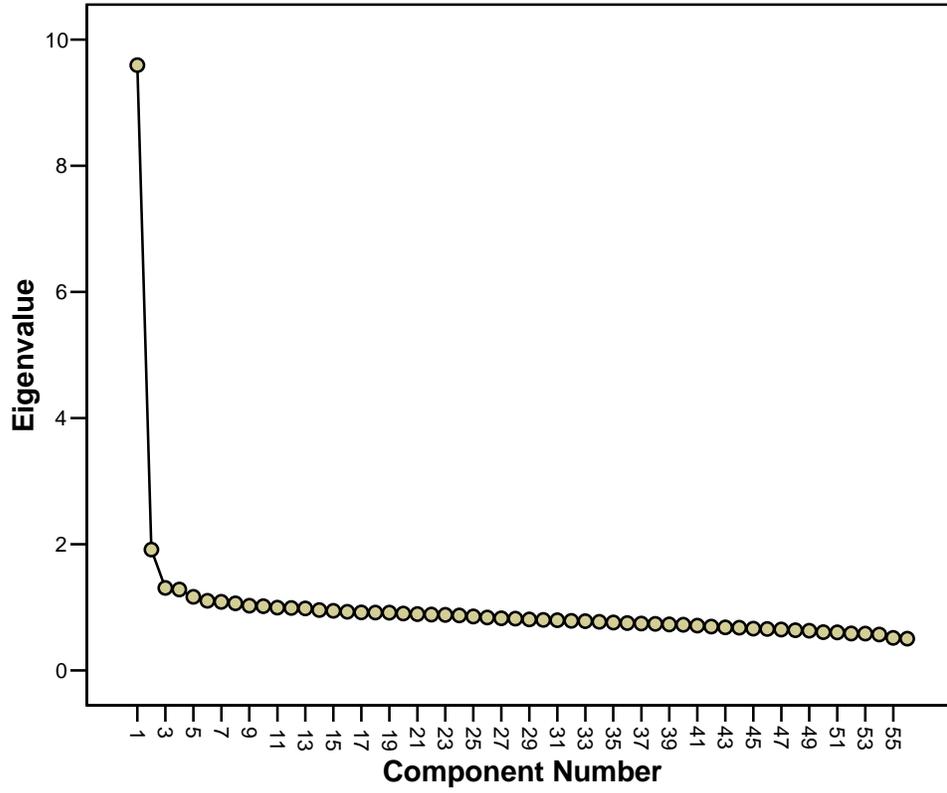


Figure 5. 16: Scree Plot, Grade 4 Communication Arts

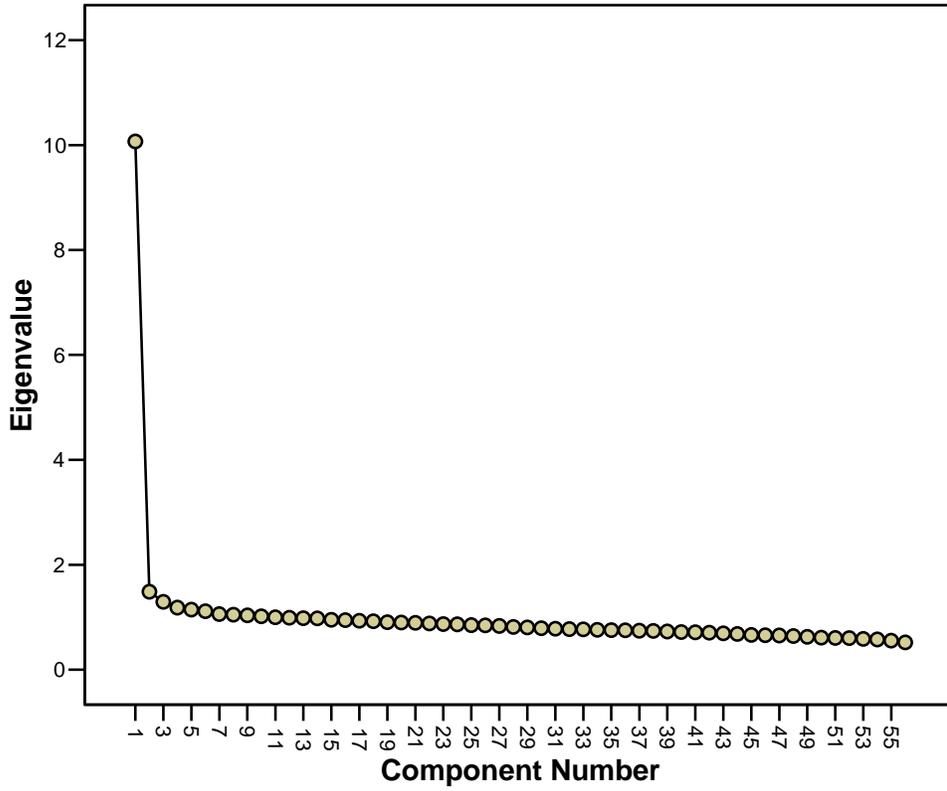


Figure 5. 17: Scree Plot, Grade 5 Communication Arts

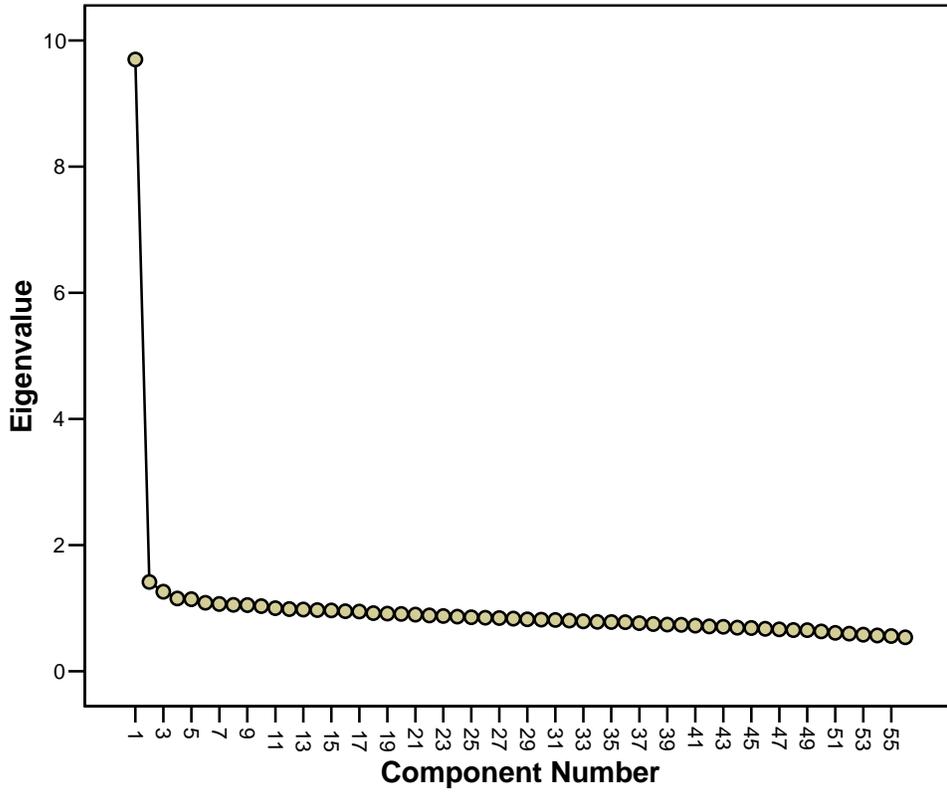


Figure 5. 18: Scree Plot, Grade 6 Communication Arts

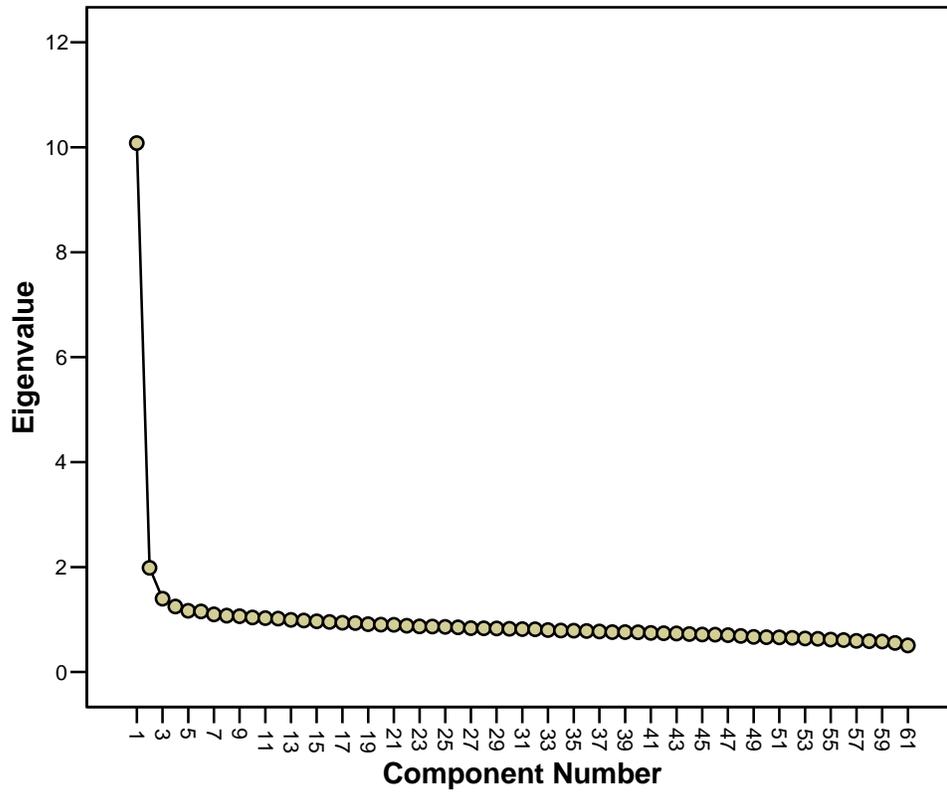


Figure 5. 19: Scree Plot, Grade 7 Communication Arts

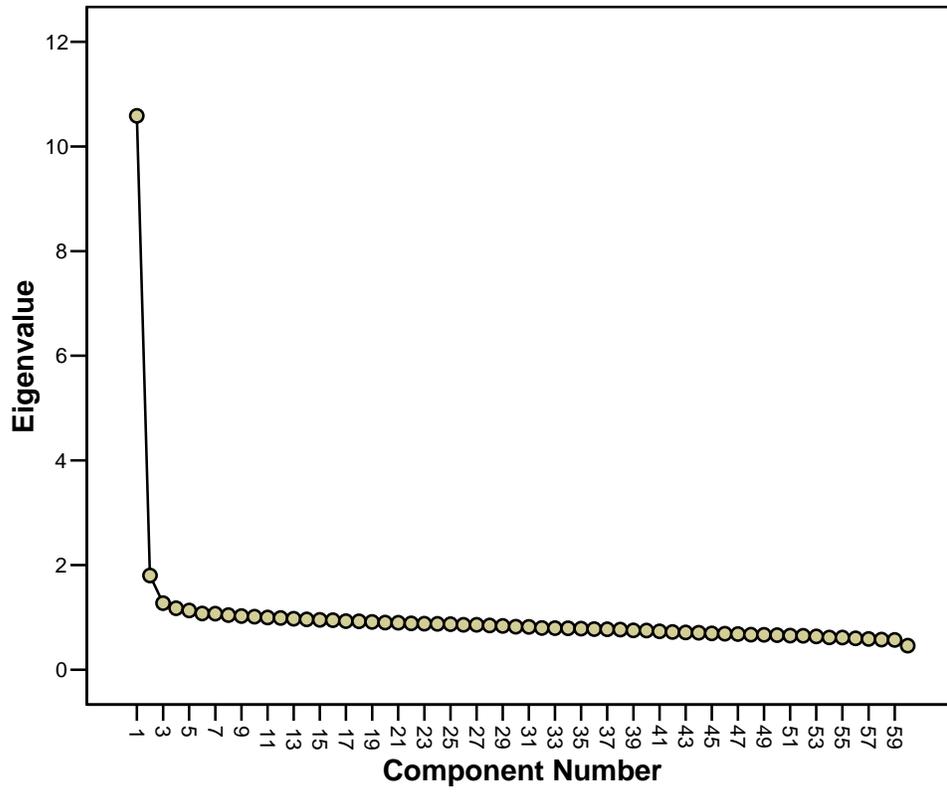


Figure 5. 20: Scree Plot, Grade 8 Communication Arts

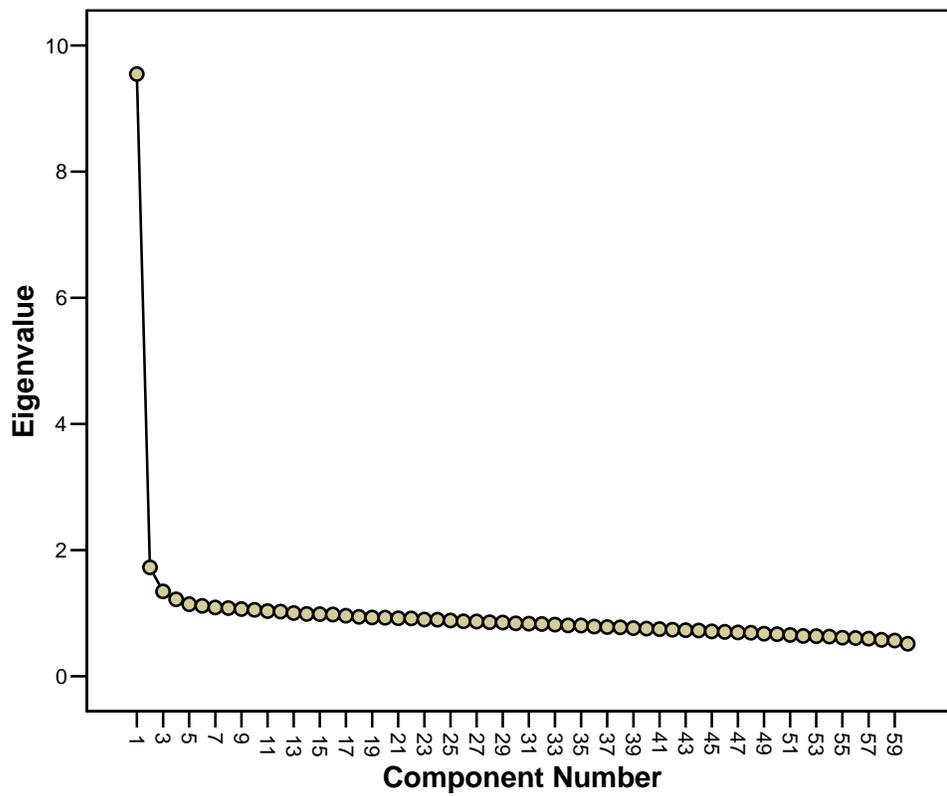


Figure 5. 21: Scree Plot, Grade 11 Communication Arts

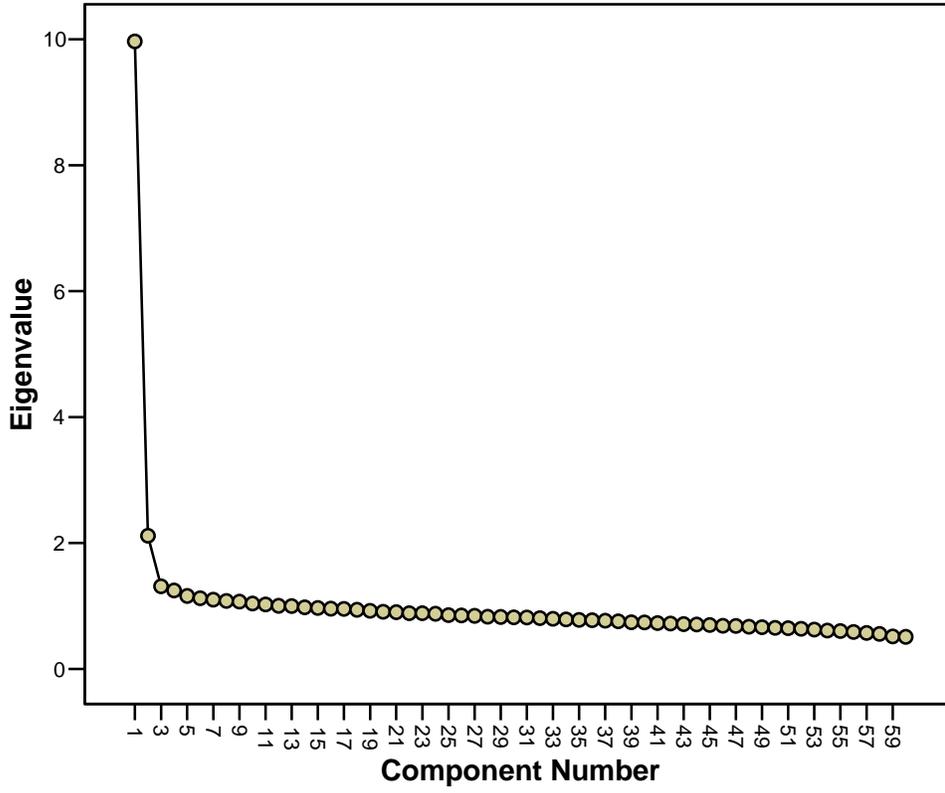


Figure 5. 22: Scree Plot, Grade 3 Mathematics

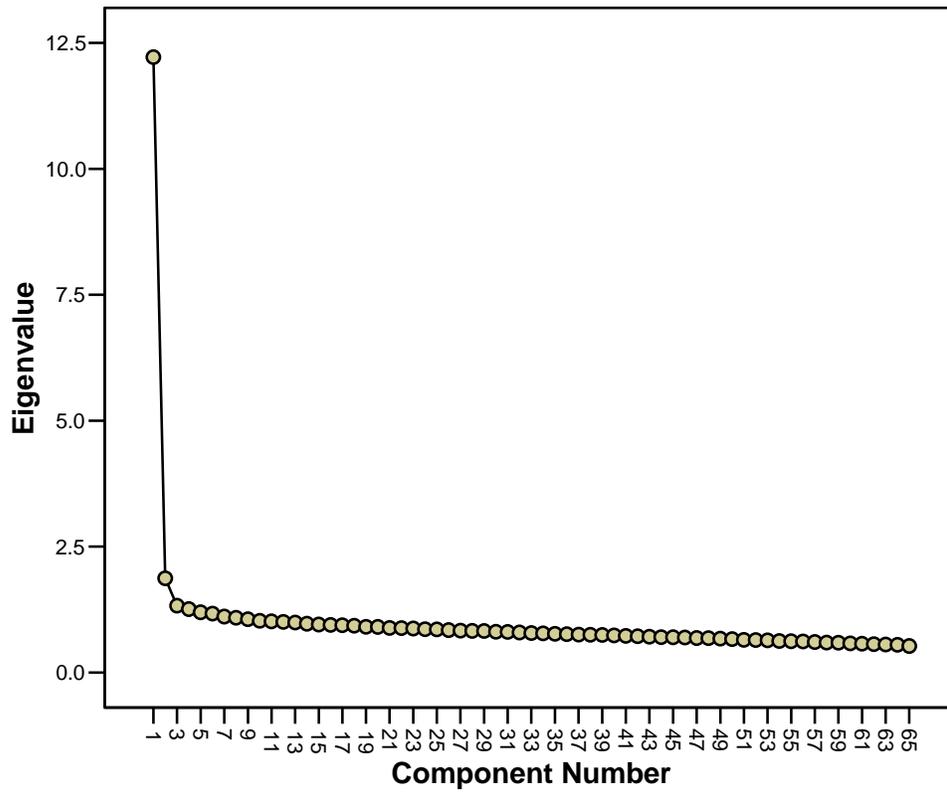


Figure 5. 23: Scree Plot, Grade 4 Mathematics

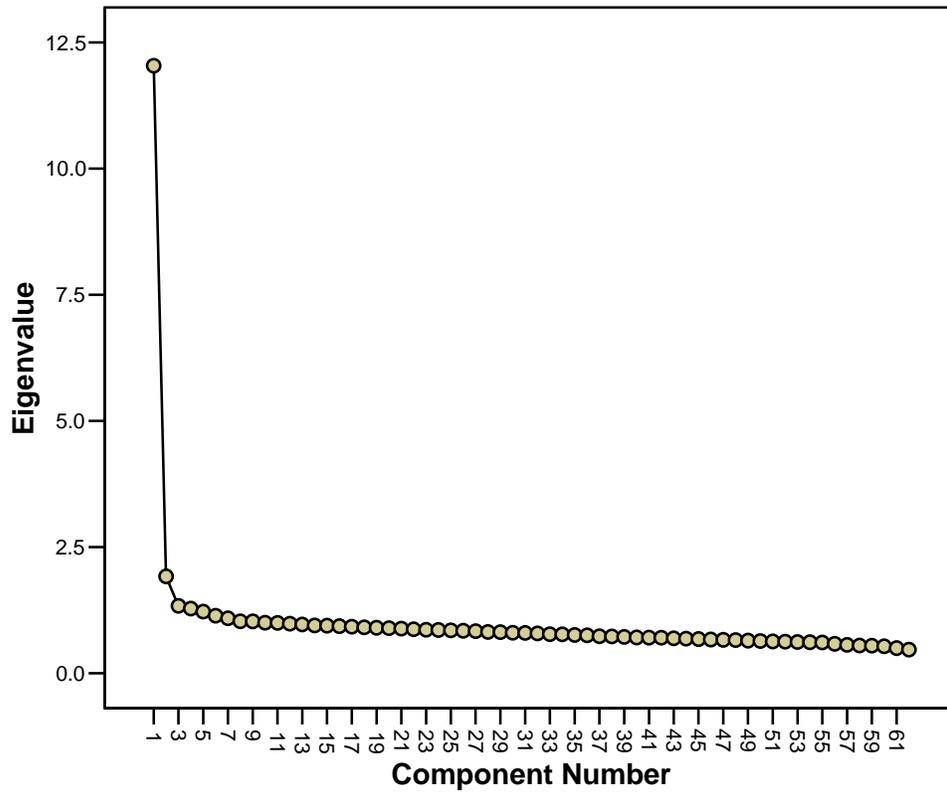


Figure 5. 24: Scree Plot, Grade 5 Mathematics

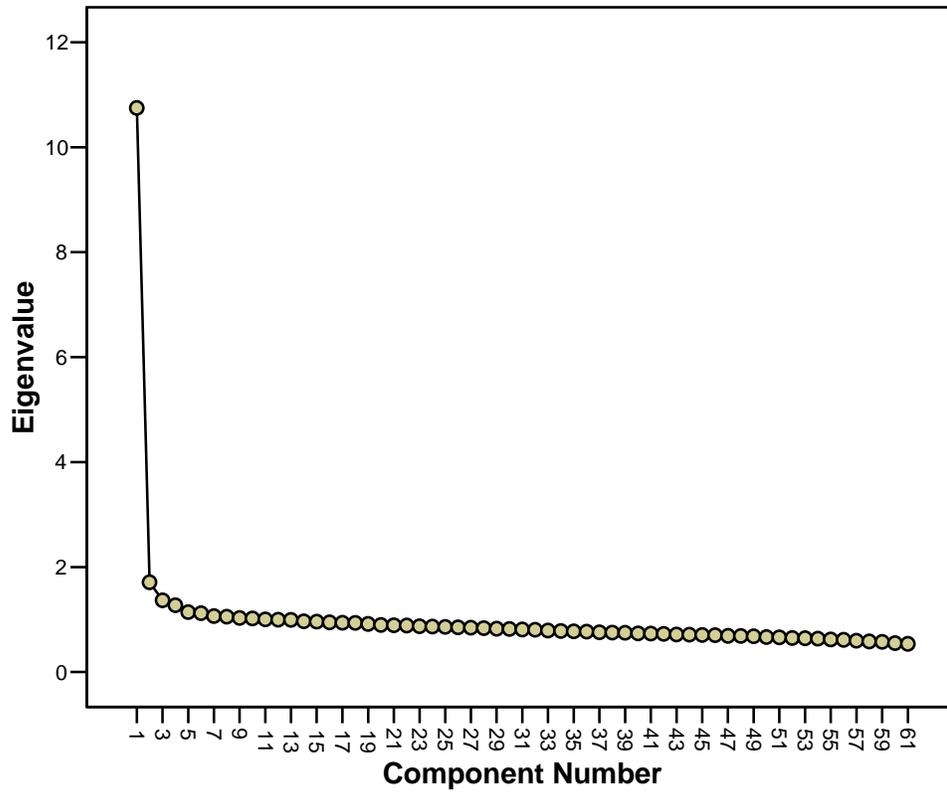


Figure 5. 25: Scree Plot, Grade 6 Mathematics

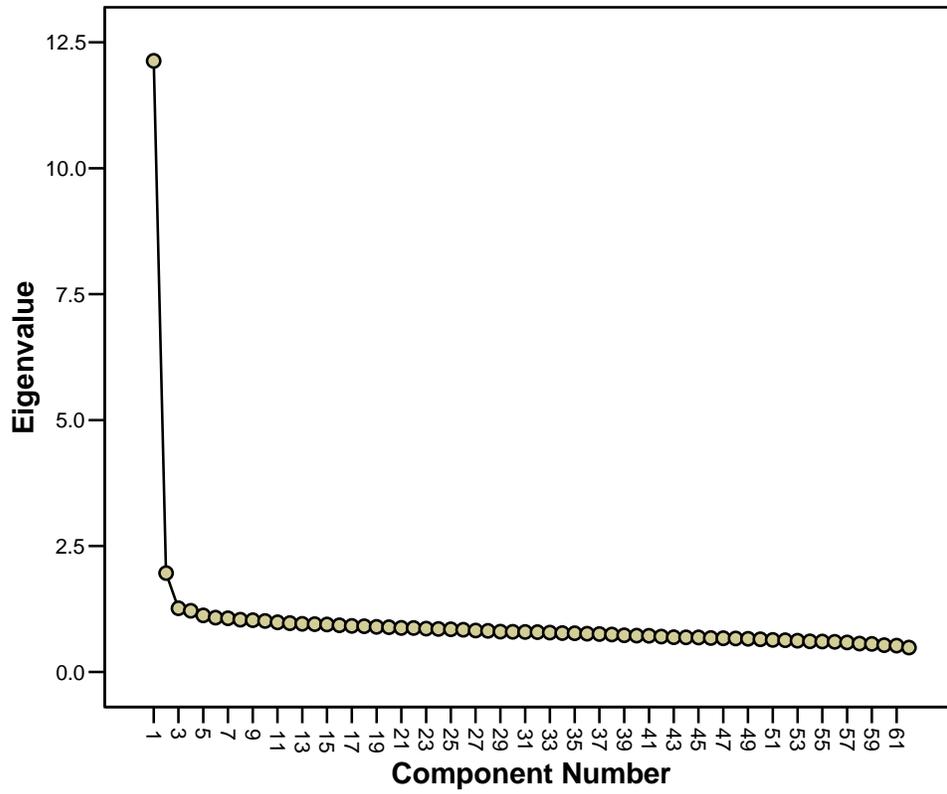


Figure 5. 26: Scree Plot, Grade 7 Mathematics

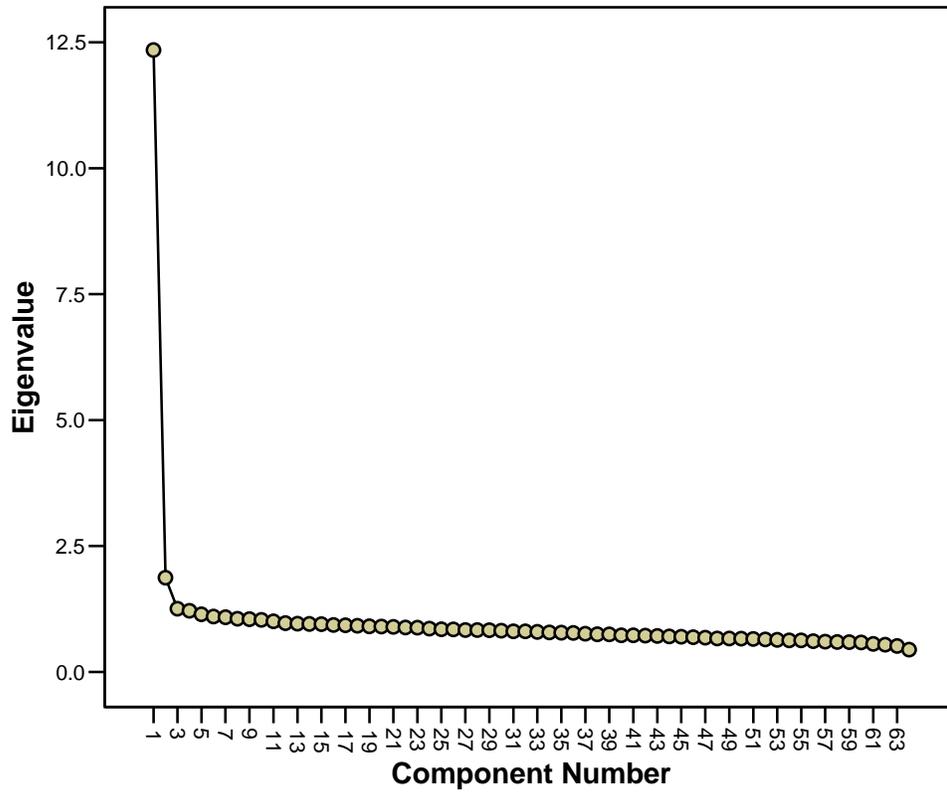


Figure 5. 27: Scree Plot, Grade 8 Mathematics

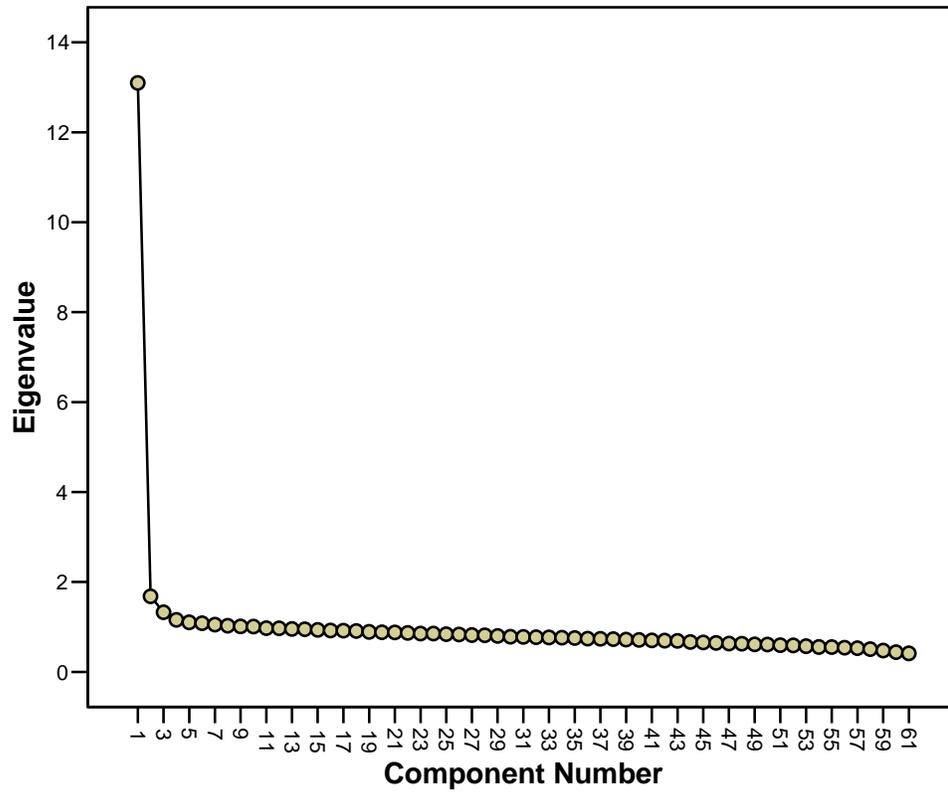


Figure 5. 28: Scree Plot, Grade 10 Mathematics

## Part 6: Scaling and Linking

Scaling and linking were conducted using the calibration sample described in Part 4. In this section, we first discuss the item response theory (IRT) models used for calibrating the data and then address how well these models fit the Missouri data. If the IRT model fits the empirical item response distributions for the population for which we want to make generalizations, that is, Missouri students, then the claim that the scores are valid indicators of an underlying ability is strengthened. The lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) for the MAP are presented. Finally, we provide a general overview of scaling and discuss the methods used to link the MAP results to the *TerraNova* scale.

### *Item Response Theory*

A marginal maximum-likelihood procedure was used to simultaneously estimate the item parameters using the 3PL/2PPC IRT models (Bock & Aitkin, 1981; Thissen, 1982). Under the 3PL model, the probability that a student with trait or scale score  $\theta$  will respond correctly to multiple-choice item  $j$  is

$$P_j(\theta) = c_j + (1 - c_j) / [1 + \exp(-1.7a_j(\theta - b_j))]. \quad (1)$$

In equation (1),  $a_j$  is the item discrimination,  $b_j$  is the item difficulty, and  $c_j$  is the probability of a correct response by a very low-scoring student. Under the 2PPC model, the probability that a student with trait or scale score  $\theta$  will respond in category  $k$  to partial-credit item  $j$  is

$$P_{jk}(\theta) = \exp(z_{jk}) / \sum_{i=1}^{m_j} \exp(z_{ji}), \quad (2)$$

where  $z_{jk} = (k-1)f_j - \sum_{i=0}^{k-1} g_{ji}$ , and  $g_{j0} = 0$  for all  $j$ .

The summary output of the 3PL and 2PPC models is in two different metrics. The location and discrimination parameters for the MC items are in the traditional 3PL metric, and are labeled  $b$  and  $a$ , respectively. In the 2PPC model,  $f$  (alpha) and  $g$  (gamma) are analogous to  $b$  and  $a$ , where alpha is the discrimination parameter and gamma over alpha ( $g/f$ ) is the location where adjacent trace lines cross on the ability scale. Because of the different metrics used, the 3PL parameters  $b$  and  $a$  are not directly comparable to the 2PPC parameters  $f$  and  $g$ ; however, they can be converted to a common metric. The two metrics are related by  $b = g/f$  and  $a = f / 1.7$  (Burket, 1995). As a result of this procedure, the MC and CR items are placed on the same scale. Note that for the 2PPC model, there are  $m_j - 1$  (where  $m_j$  is a score level  $j$ ) independent  $g$ 's and one  $f$ , for a total of  $m_j$  independent parameters estimated for each item, while there is one  $a$  and one  $b$  per item in the 3PL model.

## **Model Fit**

A procedure developed by Yen (1981) was used to assess model-to-data fit for all test items. In this procedure, students are rank ordered on the basis of their  $\hat{\theta}$  values and sorted into ten cells with ten percent of the sample in each cell. Each item  $j$  in each decile  $i$  has a response from  $N_{ij}$  examinees. The fitted IRT models are used to calculate an expected proportion  $E_{ijk}$  of examinees who respond to item  $j$  in category  $k$ . The observed proportion  $O_{ijk}$  is also tabulated for each decile, and the approximate chi-square statistic

$$Q_{1j} = \sum_{i=1}^{10} \sum_{k=1}^{m_j} \frac{N_{ij} (O_{ijk} - E_{ijk})^2}{E_{ijk}},$$

$Q_{1j}$  should be approximately chi-square distributed with degrees of freedom ( $DF$ ) equal to the number of “independent” cells,  $10(m_j-1)$ , minus the number of estimated parameters. For the 3PL model  $m_j = 2$ , so  $DF = 10(2-1) - 3 = 7$ . For the 2PPC model,  $DF = 10(m_j - 1) - m_j = 9m_j - 10$ . Since  $DF$  differs between MC and CR items and between CR items with different score levels  $m_j$ ,  $Q_{1j}$  is transformed, yielding the test statistic

$$Z_j = \frac{Q_{1j} - DF}{\sqrt{2DF}}.$$

This statistic is used for flagging items that fit relatively poorly.  $Z_j$  is sensitive to sample size, and cutoff values for flagging an item based on  $Z_j$  have been developed and were used to identify items for the item review. The cutoff value is  $(N/1500 \times 4)$  for a given test, where  $N$  is the sample size.

Twelve MAP operational items were flagged for poor fit. In Communication Arts, one item was flagged for poor fit in Grade 3 and Grade 4, and two items were flagged for poor fit in Grade 11. In Mathematics, one item was flagged for poor fit in each Grades 3, 4 and 10, two items were flagged for poor fit in Grade 5, and three items were flagged for poor fit in Grade 7. Table 6.1 shows the chi-square statistic and the Z-statistic for each flagged item. The average percent across ten cells of observed percentage correct and predicted percentage correct is also provided. The difference between the observed and predicted percentages provides an indication of how well the modeled response curves reflect the empirical curves.

Each of the flagged items was examined more closely by studying its item characteristic curve (ICC) at each non-zero score point. The ICC models the relationship between the examinees’ performance on an item and the examinees’ underlying ability. In almost all cases for which model misfit occurs, relatively few students occupy these scale score ranges which are at the lower and upper tails of the distribution. Poor fit may occur in one region of the underlying ability distribution when there are relatively few students at that particular point in the distribution. The model functions well for the flagged items in the middle of the theta distribution where the majority of students perform.

Figures 6.1 through 6.12 show the item characteristic curves for each of the misfitting MAP items. The smooth line in each of these figures represents predicted relationship between examinee performance on the item and examinee ability, and the jagged line represents the observed relationship<sup>5</sup>. Large differences between the two lines indicate poor fit. Each figure also shows the distribution of scale scores so that the fit between observed and predicted performance at different ability levels can be interpreted in relationship to the number of examinees at each level.

With large numbers of observations such as there are for the Missouri calibration samples, items may be flagged for statistically significant differences; however, these differences may not be of practical importance. In the case of the twelve MAP items flagged for misfit, the differences do not seem to be of practical importance.

Figure 6.1 presents the ICC for Session 3, Item 26 (selected-response item) on the Grade 3 Communication Arts test. As shown, there is somewhat poor fit throughout the ability distribution.

Figure 6.2 presents the ICC for Session 2, Item 19 (two-point constructed-response item) on the Grade 4 Communication Arts test. As shown, there is poor fit at the lower end of level 1 (students who scored 0 out of 2). Levels 2 and 3 show spikes at the lower and higher ends of the ability distribution for students who scored 1 out of 2 and 2 out of 2, respectively.

Figure 6.3 presents the ICC for Session 1, Item 5 (two-point constructed-response item) on the Grade 11 Communication Arts test. There is poor fit at the lower end of the ability distribution of level 1, at the upper and lower end of level 2, and at the upper end of level 3.

Figure 6.4 presents the ICC for Session 3, Item 3 (selected-response item) on the Grade 11 Communication Arts test. As shown, there is poor fit at the lower end of the ability distribution but good fit throughout the rest of the ability distribution.

Figure 6.5 presents the ICC for Session 1, Item 8 (two-point constructed-response item) on the Grade 3 Mathematics test. There is good fit throughout the ability distribution for level 1. There is poor fit through the middle and upper end of the ability distribution of level 2. Level 3 shows good fit throughout the ability distribution except for a spike at the upper end of both levels.

Figure 6.6 presents the ICC for Session 2, Item 34 (two-point constructed-response item) on the Grade 4 Mathematics test. As shown, there is reasonably good fit throughout the ability distribution for levels 1 and 3. There is somewhat poor fit through the lower and middle portion of the ability distribution of level 2.

---

<sup>5</sup> For constructed response items, there will be one graph for each score level. For example, a two-point item will have three graphs, for 0, 1, and 2 score points.

Figure 6.7 presents the ICC for Session 1, Item 24 (two-point constructed-response item) on the Grade 5 Mathematics test. There is good fit throughout the ability distribution for levels 1 and 3. There is poor fit throughout of the ability distribution of level 2.

Figure 6.8 presents the ICC for Session 2, Item 34 (two-point constructed-response item) on the Grade 5 Mathematics test. As shown, there is good fit throughout the ability distribution for Level 1. There is poor fit at the upper end of the ability distribution for levels 2 and 3.

Figure 6.9 presents the ICC for Session 2, Item 22 (selected-response item) on the Grade 7 Mathematics test. There is somewhat poor fit at the lower end of the ability distribution, while the rest of the distribution shows good fit.

Figure 6.10 presents the ICC for Session 2, Item 33 (two-point constructed-response item) on the Grade 7 Mathematics test. As shown, there is poor fit at the lower end of the ability distribution of level 1 through level 2, and good fit throughout the distribution of level 3.

Figure 6.11 presents the ICC for Session 2, Item 35 (three-point constructed-response item) on the Grade 7 Mathematics test. There is poor fit at the lower end of the ability distribution for level 1 and somewhat poor fit at the lower half of the distribution for level 2. Levels 3 and 4 both show good fit other than the spike at the upper end of the distribution.

Figure 6.12 presents the ICC for Session 1, Item 25 (two-point constructed-response item) on the Grade 10 Mathematics test. As shown, there is reasonably good fit throughout the ability distribution for levels 1 and 2. Level 3 shows a poor fit at the upper end of the ability distribution.

In conclusion, the flagged items tended to show misfit for the parts of the distribution where there were few students. These items should not adversely affect the overall validity of the scores, and it is reasonable to leave the items in the test.

### ***Scaling***

The purpose of scaling a test is to enhance the validity of score interpretations by making test takers' scores comparable. The number-correct raw scores from different forms are not directly comparable because the alternate forms may differ in difficulty. Thus, the same number-correct score may not represent the same ability on two different forms of a test. By using IRT scaling for the MAP test, we are making the test scores from two forms comparable to one another.

Transformations improve this situation by incorporating the variation in item difficulty as information about student ability. In the case of the MAP, these transformations are effected by IRT. Specifically, the MAP scores are produced using a three-parameter logistic, two-parameter partial credit (3PL/2PPC) IRT model that assumes that each of the items and tasks is an independent indicator of the underlying ability governing the

propensity for students to answer an item correctly (or with greater correctness in the case of the multilevel constructed-response items).

Scaling and linking (see below) of the assessment data were performed using PARDUX (Burket, 1995), which is proprietary software developed by CTB/McGraw-Hill. PARDUX is designed to produce a single scale by jointly analyzing data resulting from students' responses to both multiple-choice (MC) items and constructed-response (CR) items. In PARDUX, items are calibrated based on IRT, using the 3PL model (Lord and Novick, 1968) for MC items and the 2PPC model (Yen, 1993) for CR items. PARDUX is also used to link the scales developed by two calibrations through the common-item procedure developed by Stocking and Lord (1983).

### **Linking Method and Results**

After the initial IRT item calibration, items parameters were linked to the *TerraNova* scale using the Stocking and Lord (1983) procedure. This approach takes advantage of the vertical properties of the *TerraNova* scale and allows for comparison of the results from the 2007 administration to results from prior and future administrations.

Linking was performed using a test characteristic curve (TCC) method proposed by Stocking and Lord (1983), and implemented using PARDUX (Burket, 1995). For the linking, the intact *TerraNova* Survey items served as anchors. Figures 6.13 to 6.26 provide plots of the input TCCs and the transformed estimated TCCs for the Communication Arts and Mathematics anchor items by grade. These plots are used to assess the quality of the linking results. The solid TCC lines in the plots (denoted "Anchor") are the TCCs for the original *TerraNova* anchor items. The dashed lines (denoted "Equated") are the TCCs from the new MAP parameter estimates transformed to the original *TerraNova* scale. The closer the two TCCs are to each other at all ability levels, the more confidence we have in the equating result.

### **Vertical Scale**

The scale on which the MAP scale scores are reported is based in part on a standardized achievement test, which makes it possible to report national percentile scores in addition to the criterion-referenced scale scores of the MAP. Although the MAP scale is unique to Missouri, the characteristic growth seen on the scale from grade to grade for the standardized test has been utilized and built upon to give the MAP its vertical scale characteristics. The vertical scale is sometimes referred to as a growth scale.

Evidence of the validity of the MAP growth scale is provided by the increase of the scale score at selected percentiles as grade level increases. Figure 6.27 and Figure 6.28 display the scale scores for several points on the score distributions for each grade of the Communication Arts and Mathematics MAP, respectively. These scale scores indicate the growth, or change, in score by grade at the 1<sup>st</sup>, 10<sup>th</sup>, 20<sup>th</sup>, . . . , 90<sup>th</sup>, and 99<sup>th</sup> percentiles. Ideally, the scale score associated with each percentile will increase from grade to grade. Figure 6.27 shows the selected percentiles for the Communication Arts MAP. Considering all but the 1<sup>st</sup> and 99<sup>th</sup> percentile, the scale scores progress upward from Grades 3 to 5 and then flatten from Grade 5 to 7 before continuing to progress

upward again from Grade 7 to 11. Across all percentiles, there is a very slight dip in performance in Grade 6 from Grade 5.

Figure 6.28 shows the selected percentiles for the Mathematics MAP. Except for the 1<sup>st</sup> percentile, there is an upward progression of scale scores across all grades with only a slight flattening out between Grades 6 and 7. At the 1<sup>st</sup> percentile, there is a decrease in scale score between Grade 8 and Grade 10.

Figure 6.29 shows the TCCs by grade for the MAP Communication Arts, and Figure 6.30 shows the TCCs by grade for the MAP Mathematics. Because these tests were linked to the *TerraNova* scale, they have an underlying vertical scale. By plotting the TCCs together, we can demonstrate that the tests increase in difficulty as the grade levels increase. In almost all cases, the TCCs indicate that test difficulty increases with grade level. For some grade levels (Grades 5 to 8) of Communication Arts, the available item pool was insufficient to create tests that resulted in the optimal increases in test difficulty.

### ***Lowest and Highest Obtainable Scale Scores***

A maximum likelihood procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected by guessing. Also, although maximum likelihood estimates are available for students with extreme scores other than zero or perfect, these estimates generally have large standard errors of measurement, and differences between these extreme values have little meaning. Therefore, scores are established for these students based on a rational but necessarily non-maximum likelihood procedure. These values, which are set separately by grade, are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). Table 6.2 shows the LOSS and HOSS values used for each grade of the Communication Arts and Mathematics MAPs.

**Table 6. 1: Item Fit Statistics for Misfitting Items, Communication Arts (CA) and Mathematics (MA)**

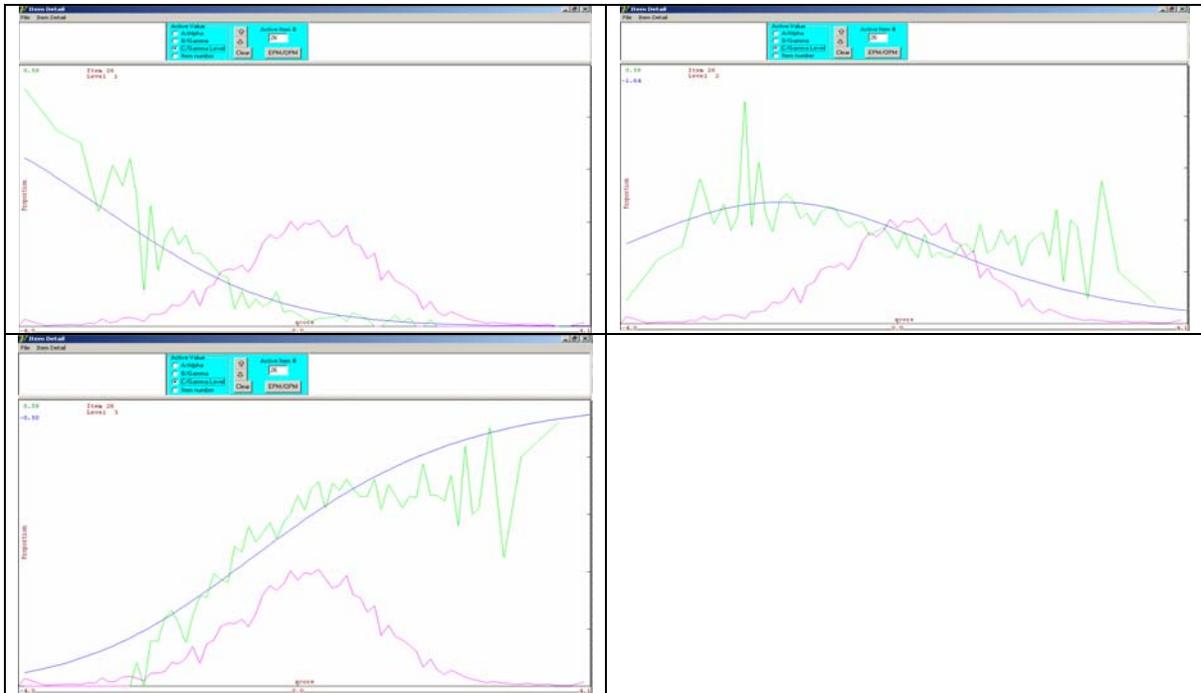
Content	Grade	Ses- sion	Item	Chi Square	DF	Total N	Z	Ob- served	Pre- dicted	Obs- Pred
CA	3	3	26	50.68	7	4234	11.68	0.59	0.60	-0.01
CA	4	2	19	119.94	17	4123	17.65	0.74	0.74	0.00
CA	11	1	5	90.52	17	3847	12.61	0.64	0.64	0.00
CA	11	3	3	68.94	7	3853	16.55	0.49	0.52	-0.03
MA	3	1	8	136.64	17	4162	20.52	0.54	0.55	0.00
MA	4	2	34	112.66	17	4054	16.41	0.75	0.75	0.00
MA	5	1	24	122.24	17	3936	18.05	0.41	0.41	-0.01
MA	5	2	34	85.84	17	3940	11.81	0.21	0.22	-0.01
MA	7	2	22	75.66	7	4734	18.35	0.76	0.76	0.00
MA	7	2	33	92.34	17	4713	12.92	0.73	0.73	0.00
MA	7	2	35	150.64	26	4724	17.29	0.50	0.50	0.00
MA	10	1	25	104.49	17	4282	15.00	0.33	0.33	0.00

**Table 6. 2: LOSS and HOSS Values, Communication Arts and Mathematics**

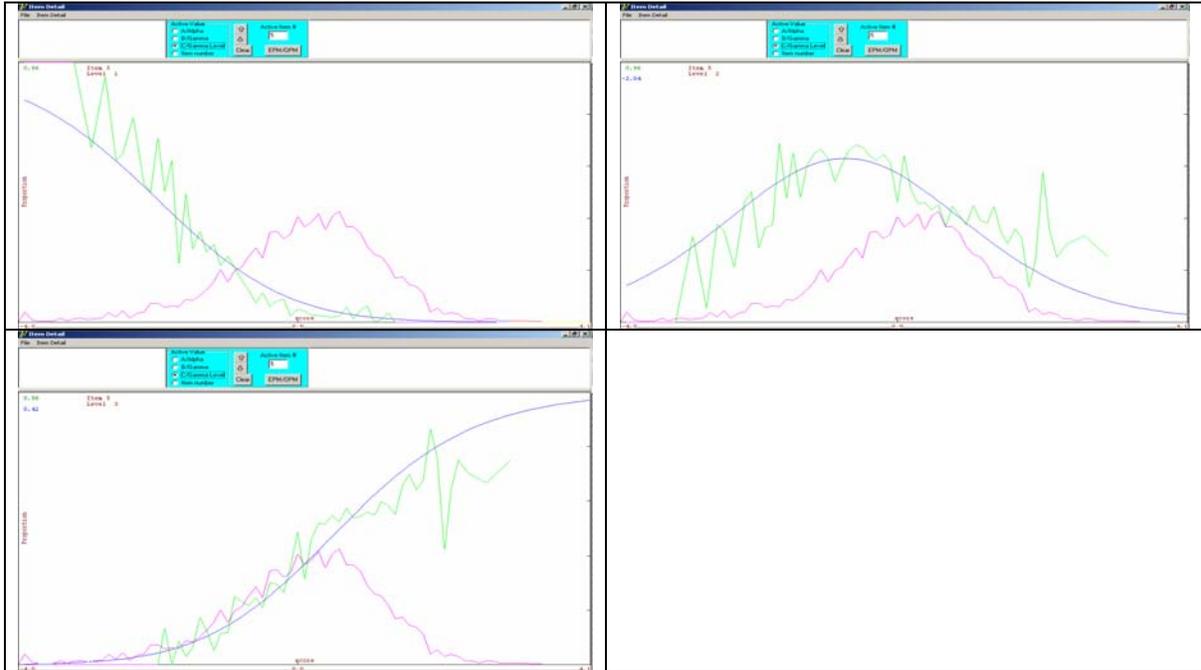
Grade	Communication Arts		Mathematics	
	LOSS	HOSS	LOSS	HOSS
3	455	790	450	780
4	470	820	465	805
5	485	840	480	830
6	505	855	495	845
7	515	865	510	860
8	530	875	525	885
10			555	910
11	545	885		



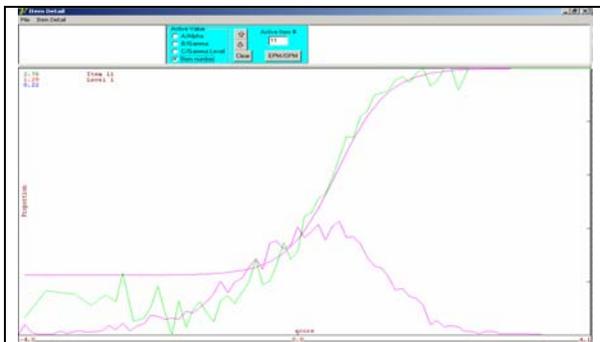
**Figure 6. 1: Item Characteristic Curve for Session 3, Item 26 ( $Z = 11.68$ ), Grade 3 Communication Arts**



**Figure 6. 2: Item Characteristic Curve for Session 2, Item 19 ( $Z = 17.65$ ), Grade 4 Communication Arts**



**Figure 6. 3: Item Characteristic Curve for Session 1, Item 5 ( $Z = 12.61$ ), Grade 11 Communication Arts**



**Figure 6. 4: Item Characteristic Curve for Session 3, Item 3 ( $Z = 16.55$ ), Grade 11 Communication Arts**

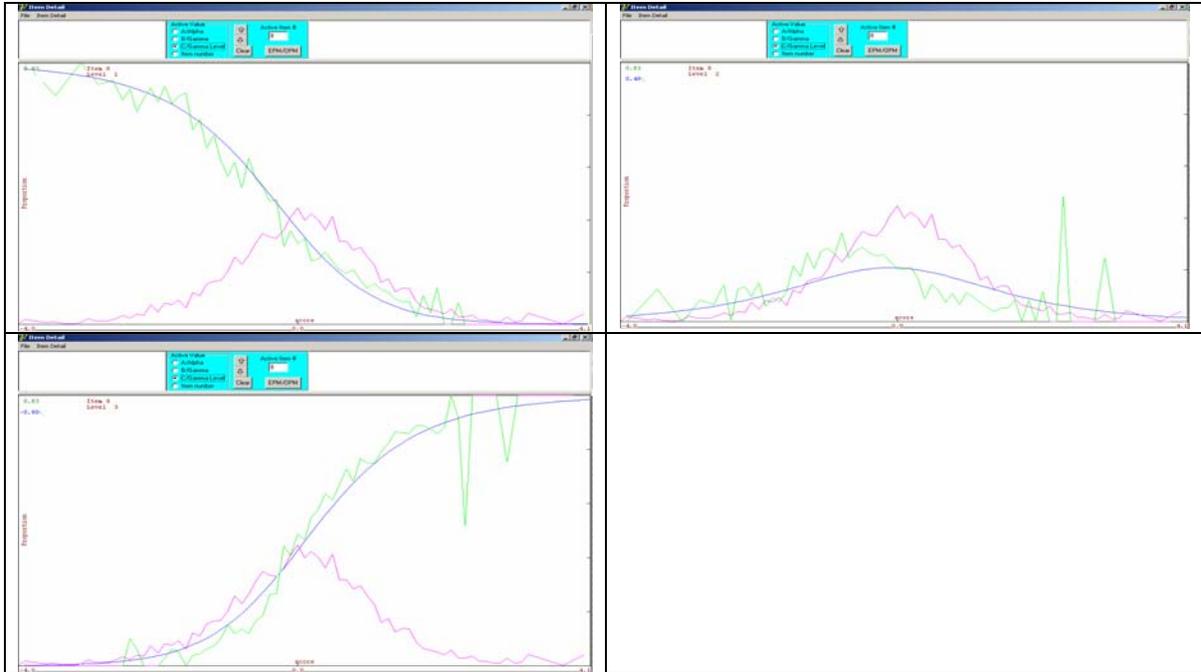


Figure 6. 5: Item Characteristic Curve for Session 1, Item 8 ( $Z = 20.52$ ), Grade 3 Mathematics

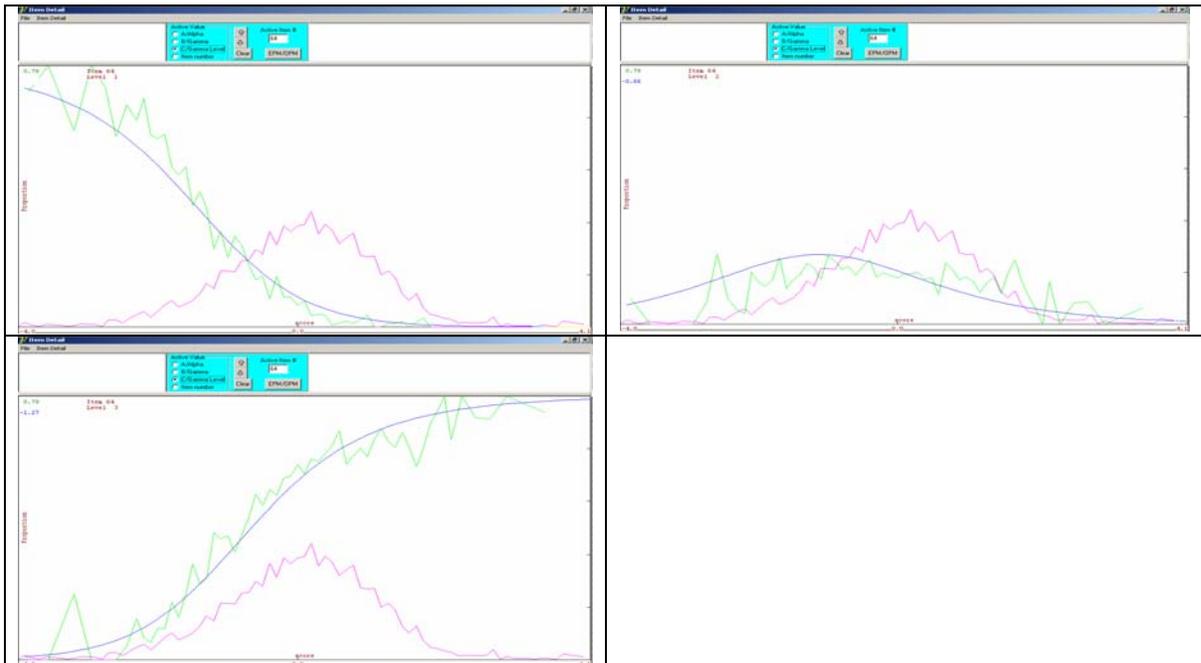


Figure 6. 6: Item Characteristic Curve for Session 2, Item 34 ( $Z = 16.41$ ), Grade 4 Mathematics

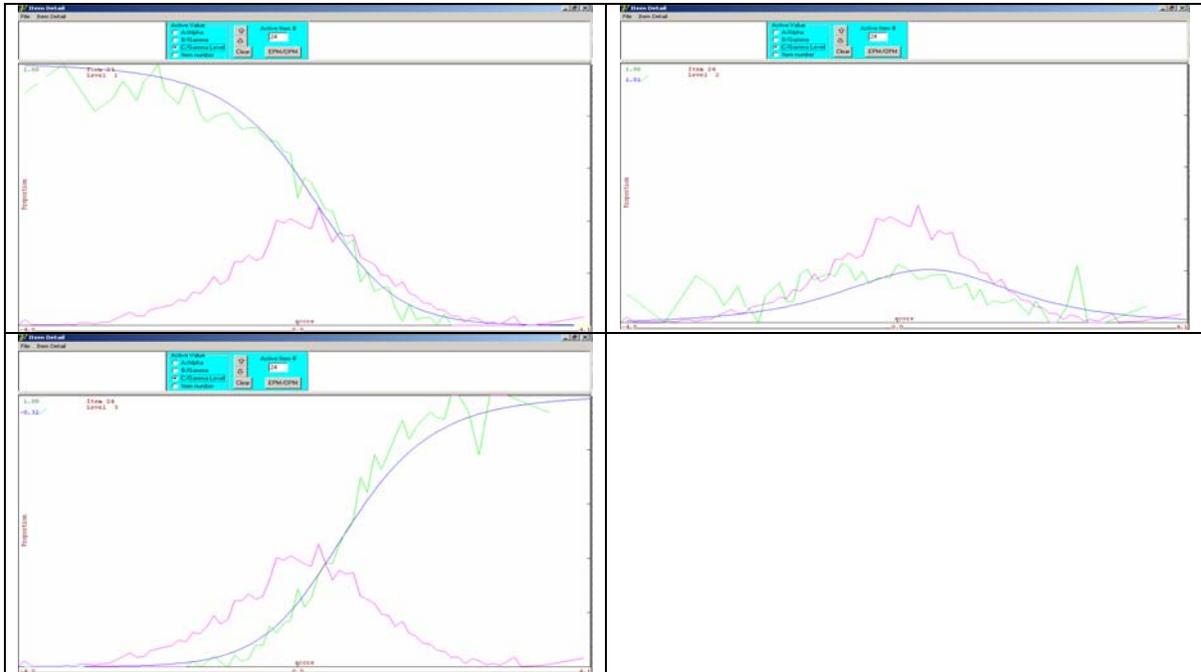


Figure 6. 7: Item Characteristic Curve for Session 1, Item 24 ( $Z = 18.05$ ), Grade 5 Mathematics

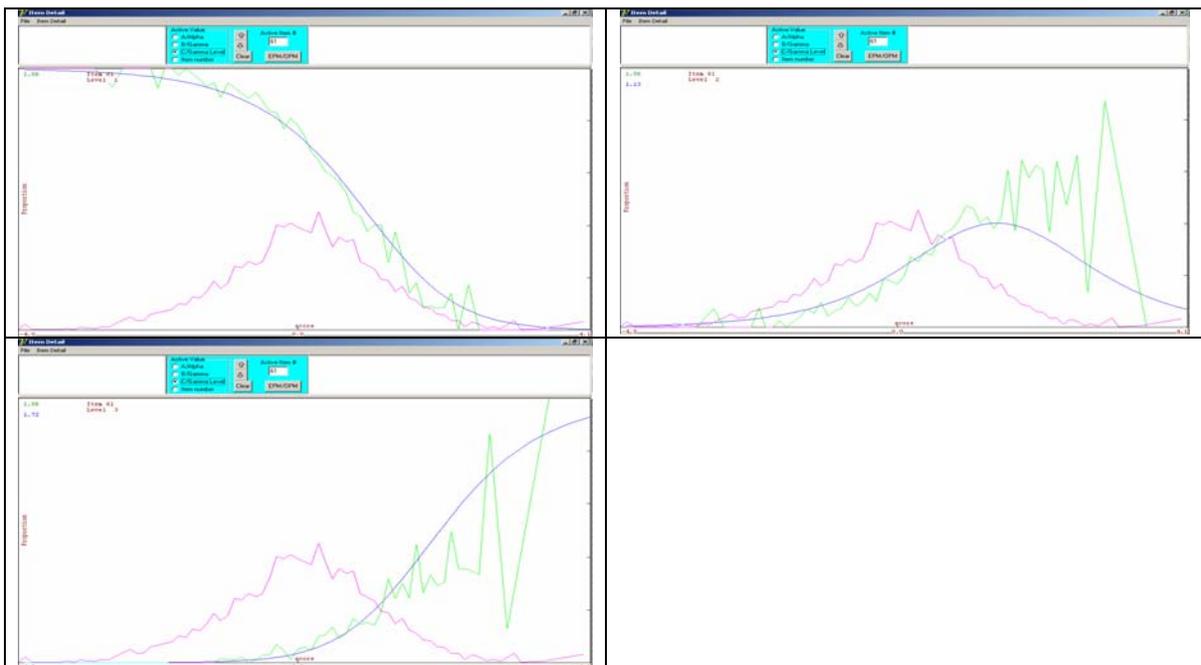


Figure 6. 8: Item Characteristic Curve for Session 2, Item 34 ( $Z = 11.81$ ), Grade 5 Mathematics

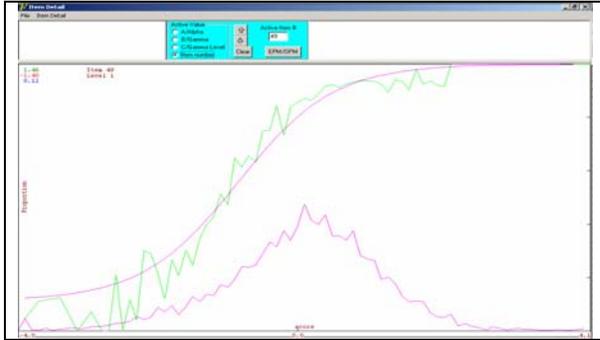


Figure 6. 9: Item Characteristic Curve for Session 2, Item 22 ( $Z = 18.35$ ), Grade 7 Mathematics

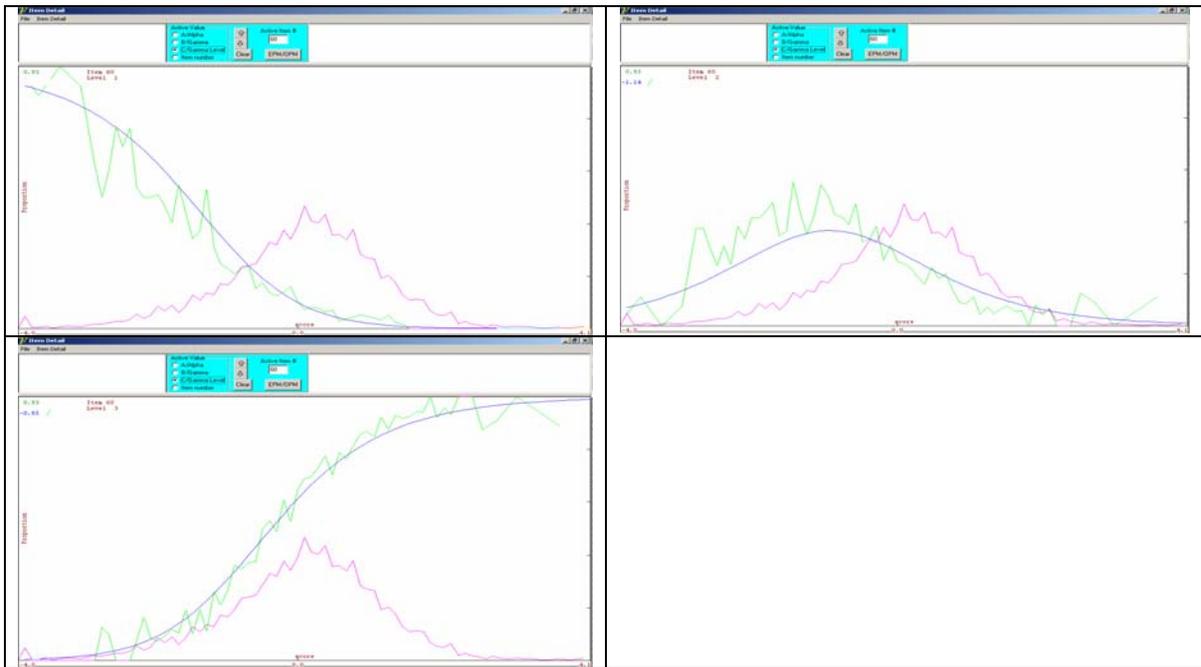


Figure 6. 10: Item Characteristic Curve for Session 2, Item 33 ( $Z = 12.92$ ), Grade 7 Mathematics

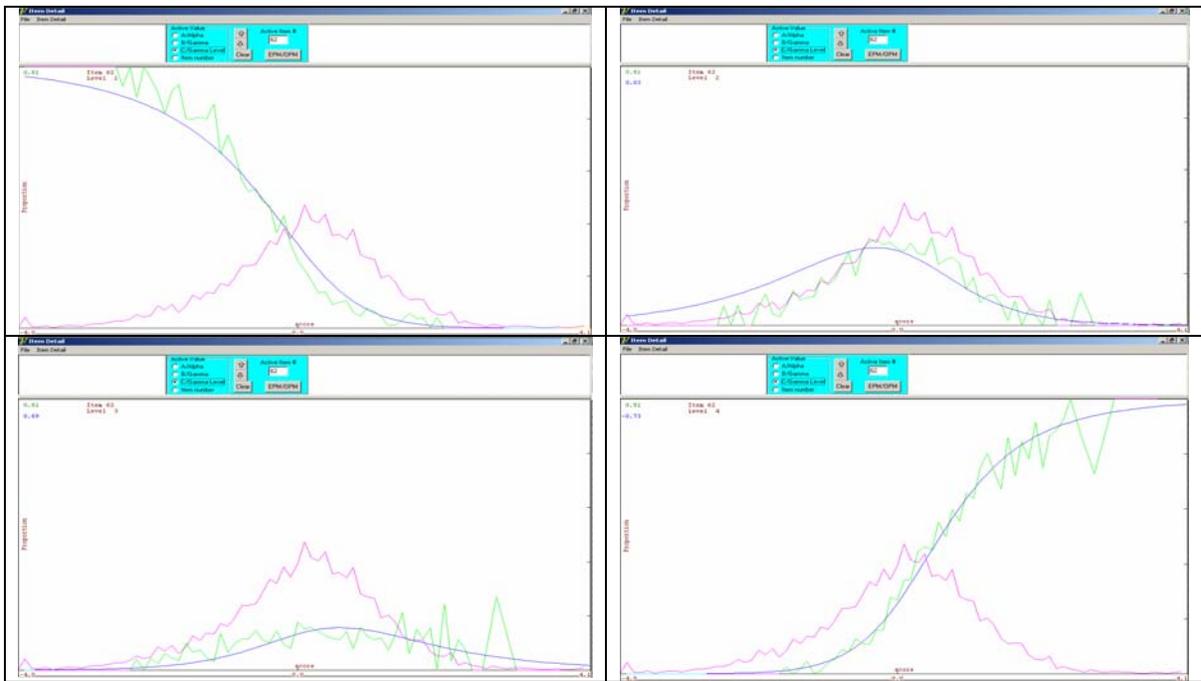


Figure 6. 11: Item Characteristic Curve for Session 2, Item 35 ( $Z = 17.29$ ), Grade 7 Mathematics

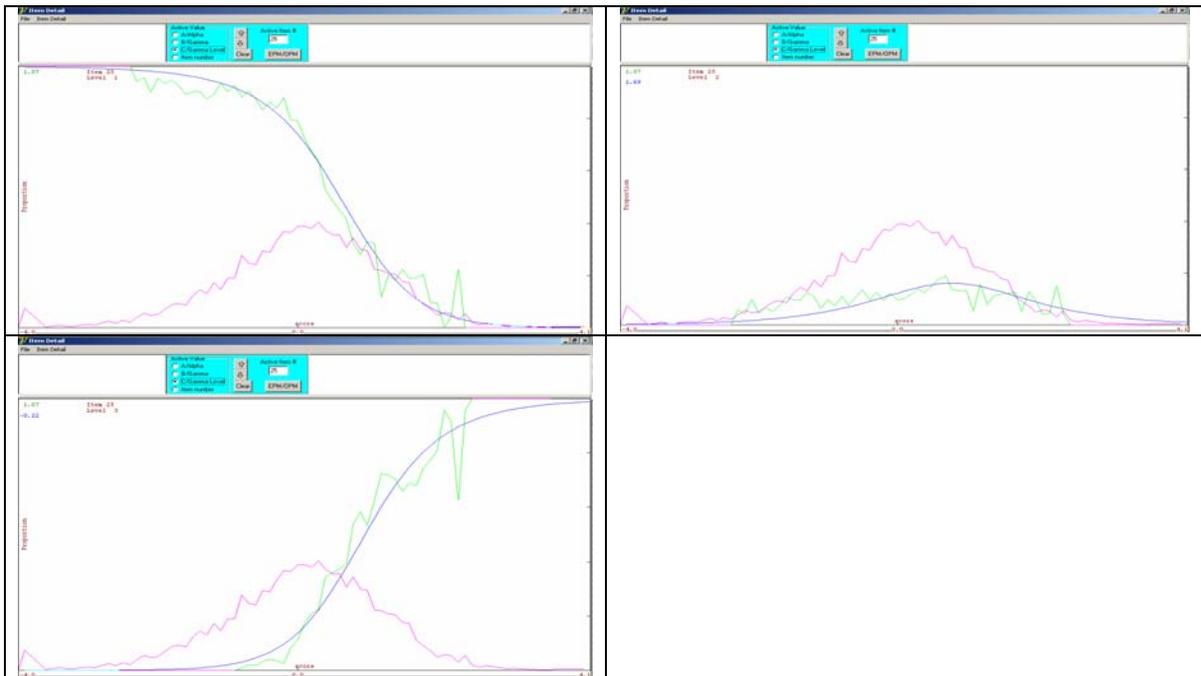
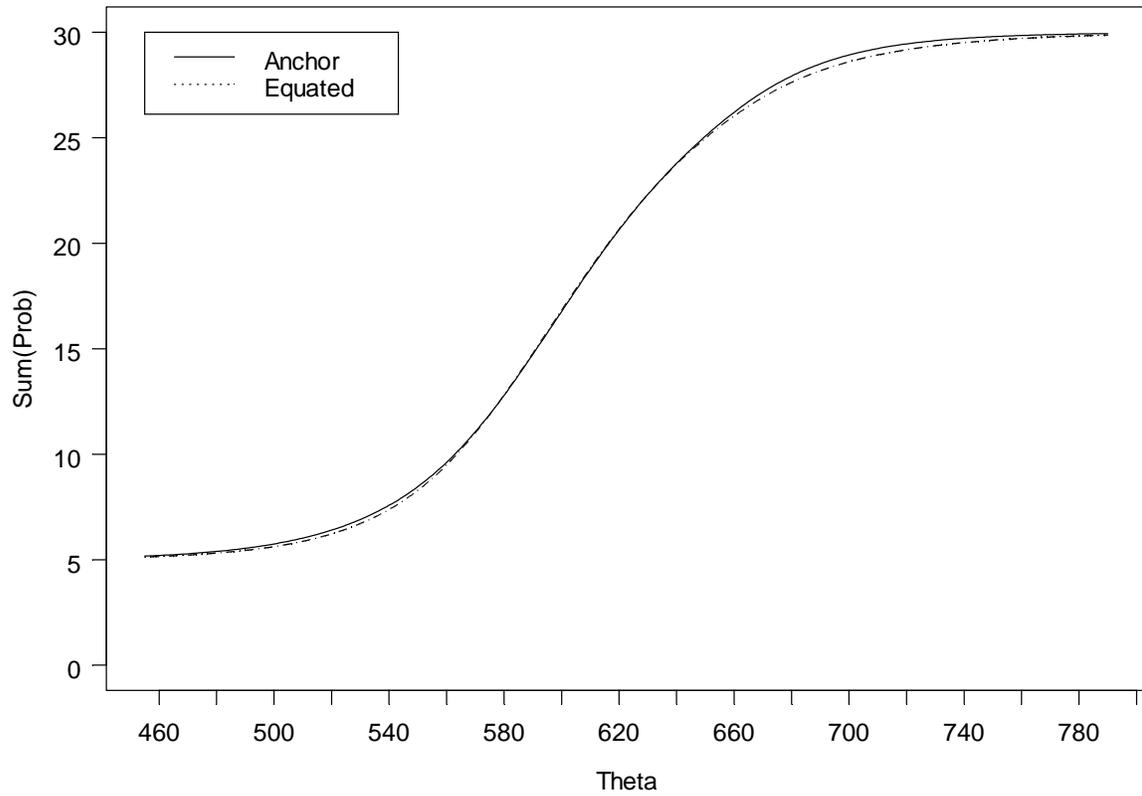
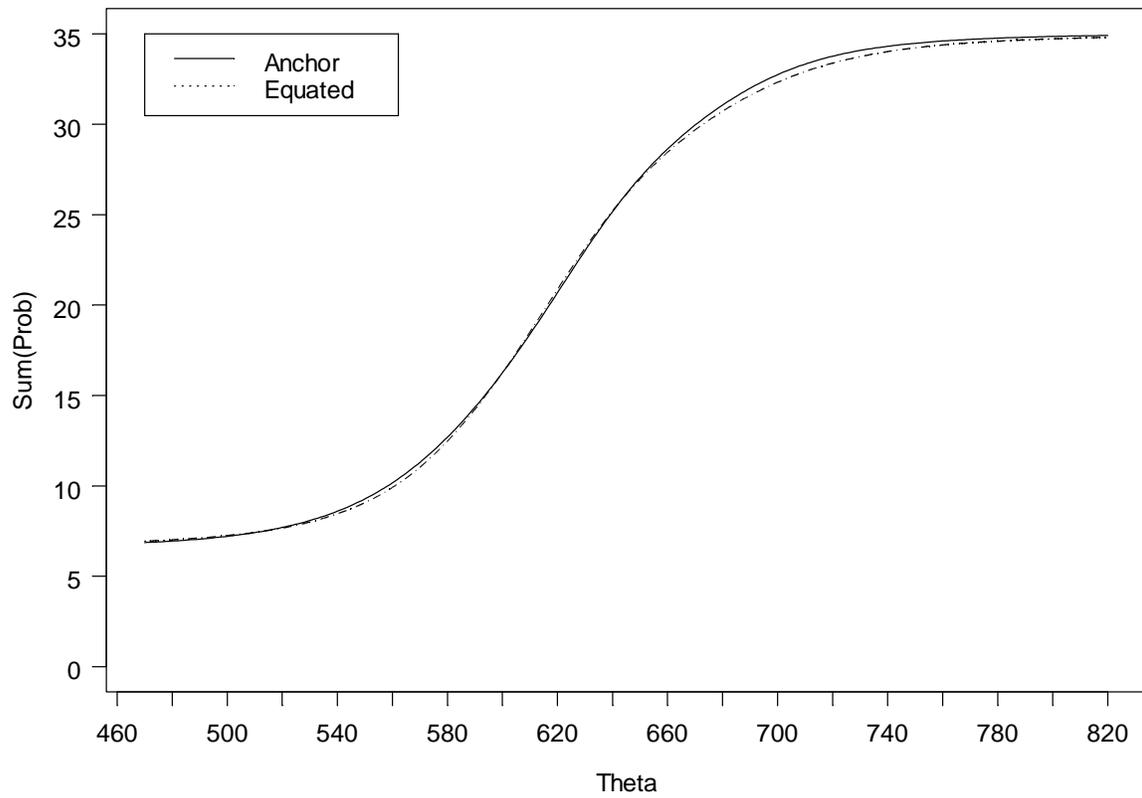


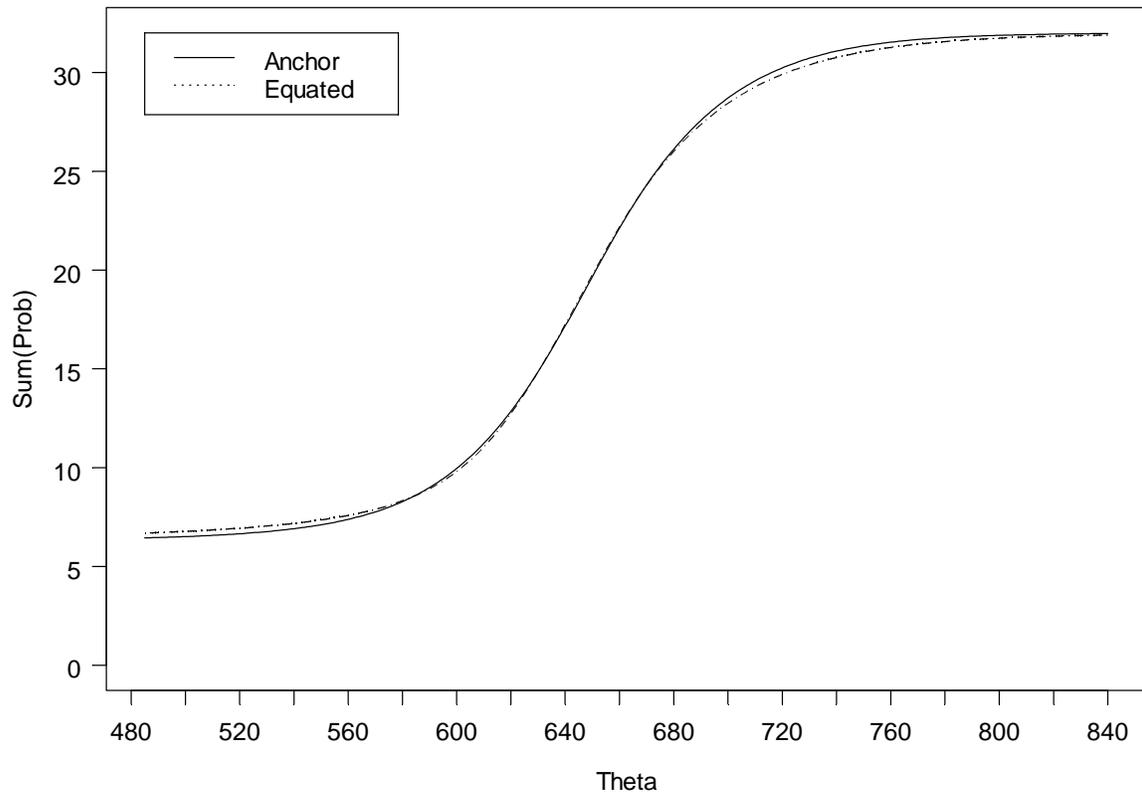
Figure 6. 12: Item Characteristic Curve for Session 1 Item 25 ( $Z = 15.00$ ), Grade 10 Mathematics  
 Copyright © 2007 by Missouri Department of Elementary and Secondary Education



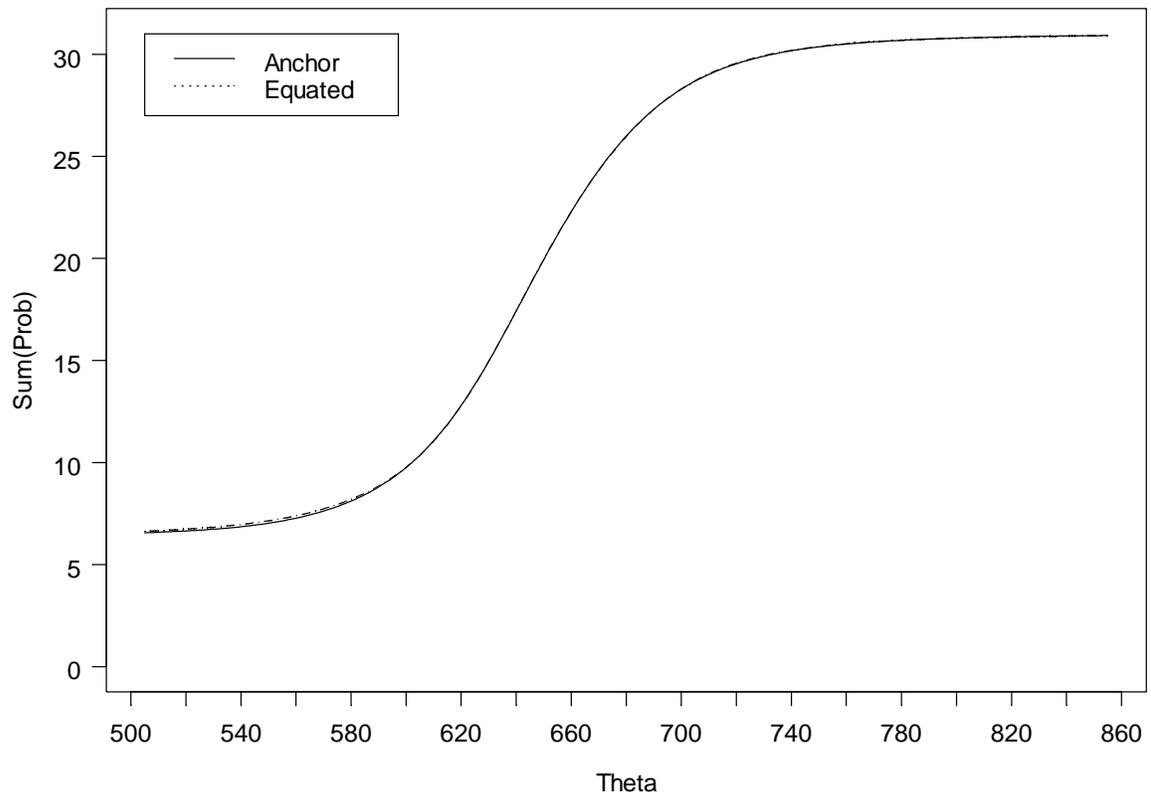
**Figure 6. 13: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 3 Communication Arts**



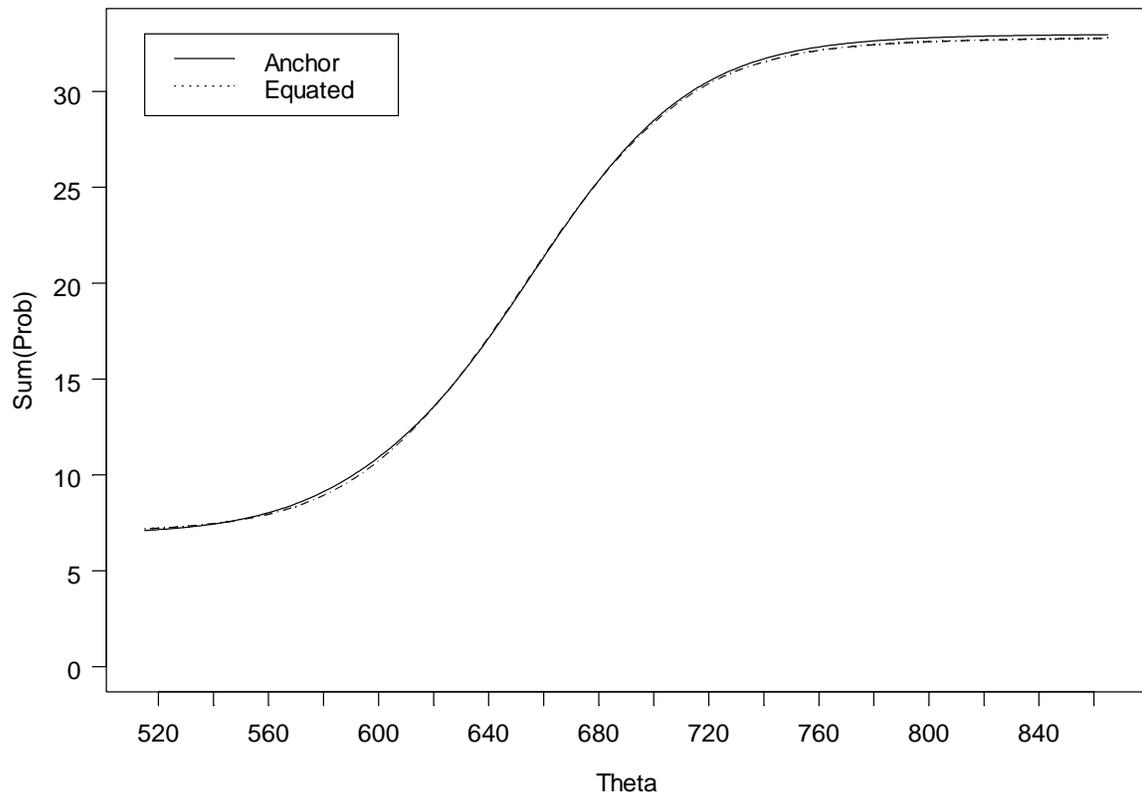
**Figure 6. 14: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 4 Communication Arts**



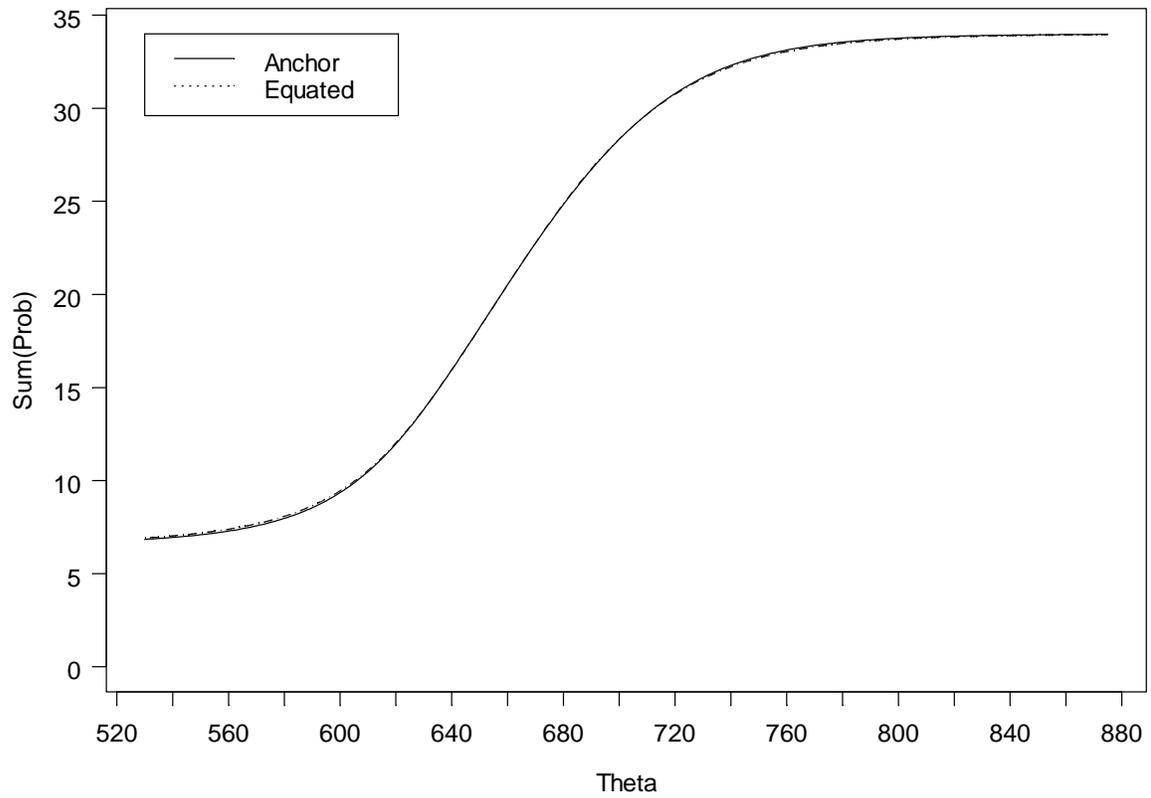
**Figure 6. 15: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 5 Communication Arts**



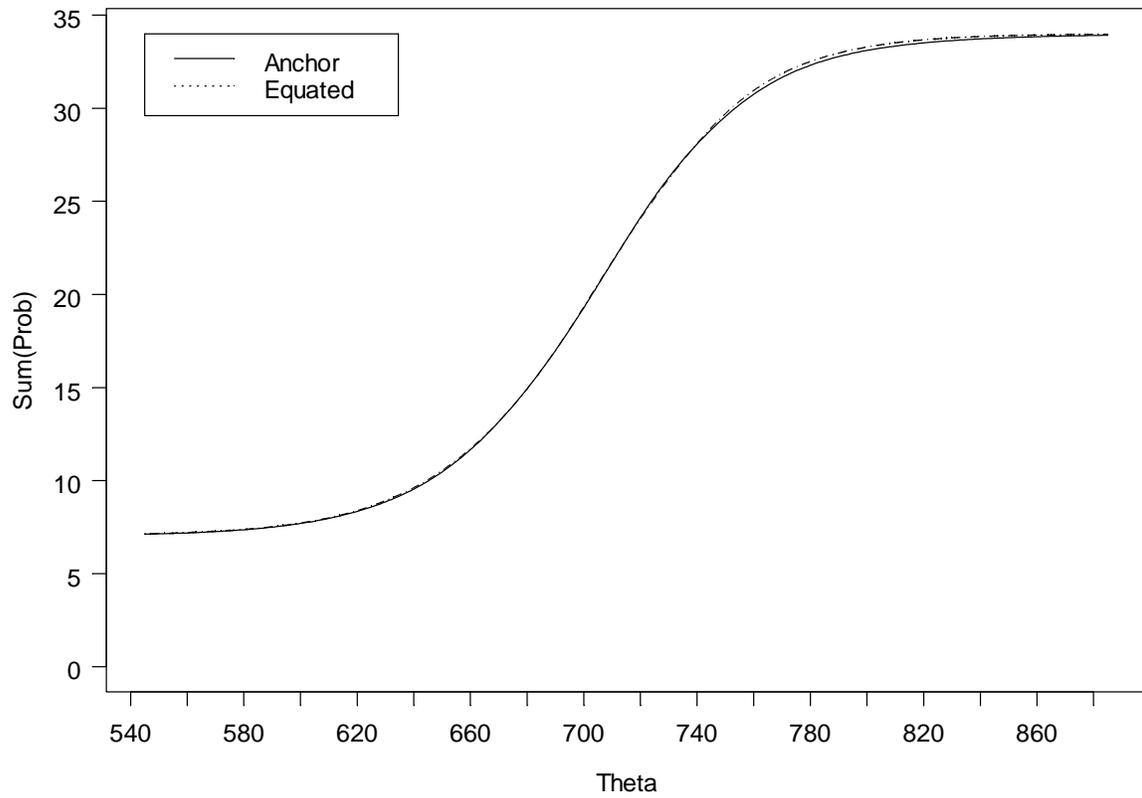
**Figure 6. 16: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 6 Communication Arts**



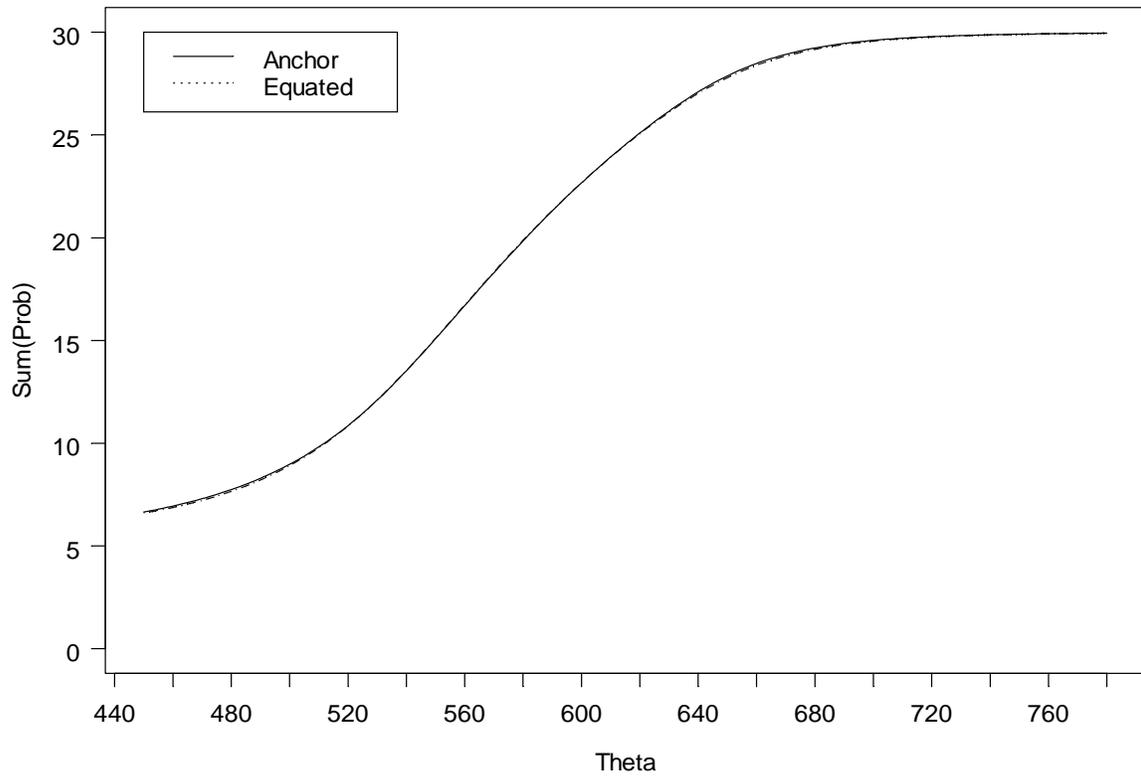
**Figure 6. 17: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 7 Communication Arts**



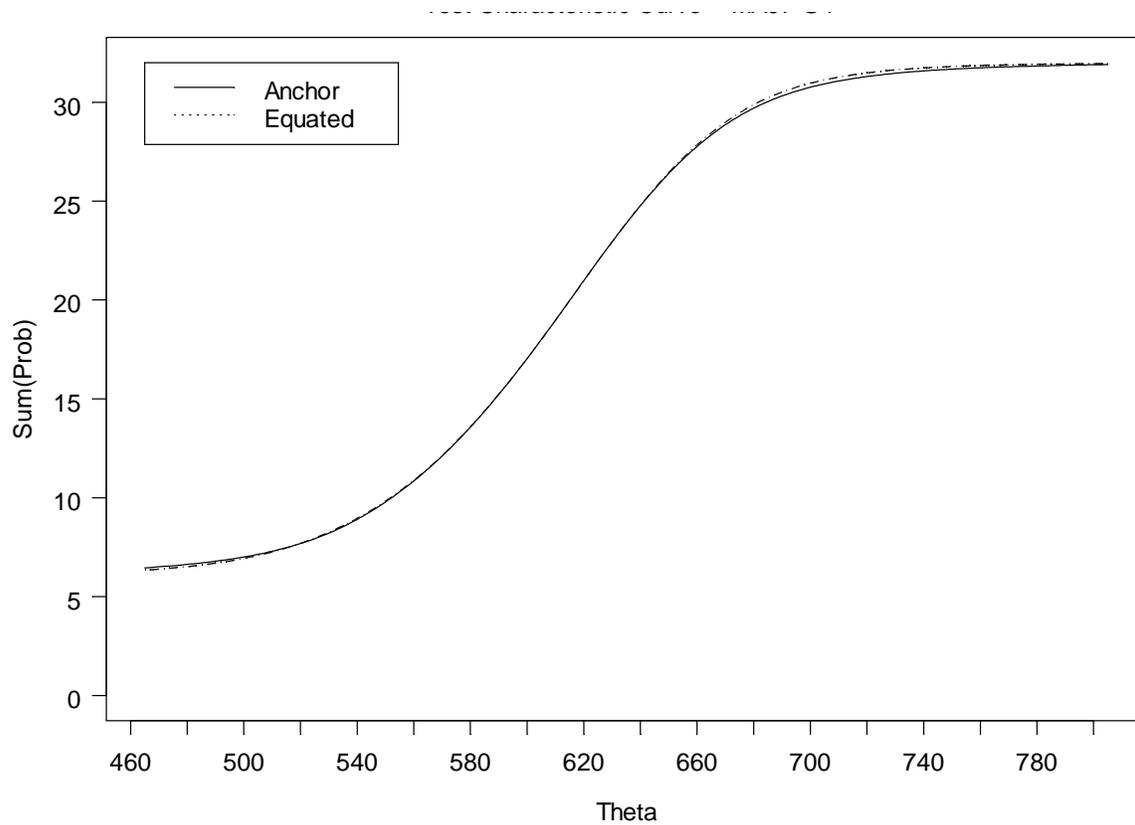
**Figure 6. 18: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 8 Communication Arts**



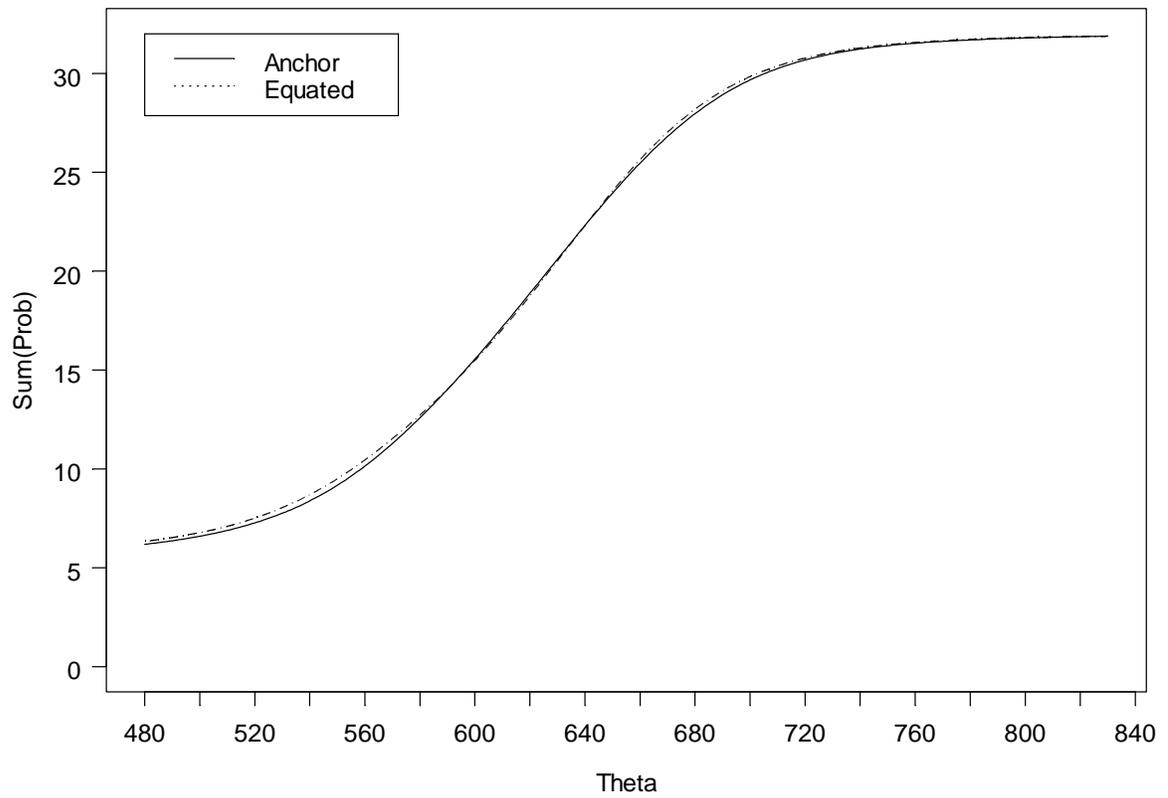
**Figure 6.19: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 11 Communication Arts**



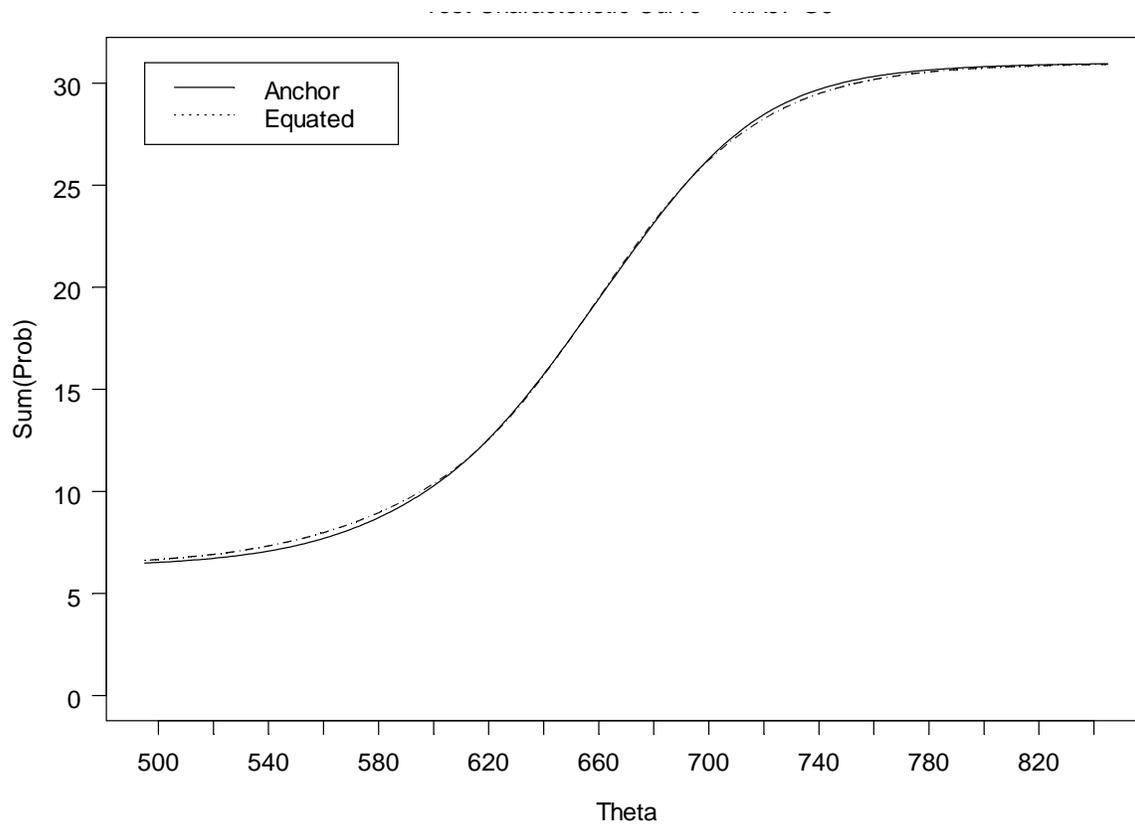
**Figure 6. 20: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 3 Mathematics**



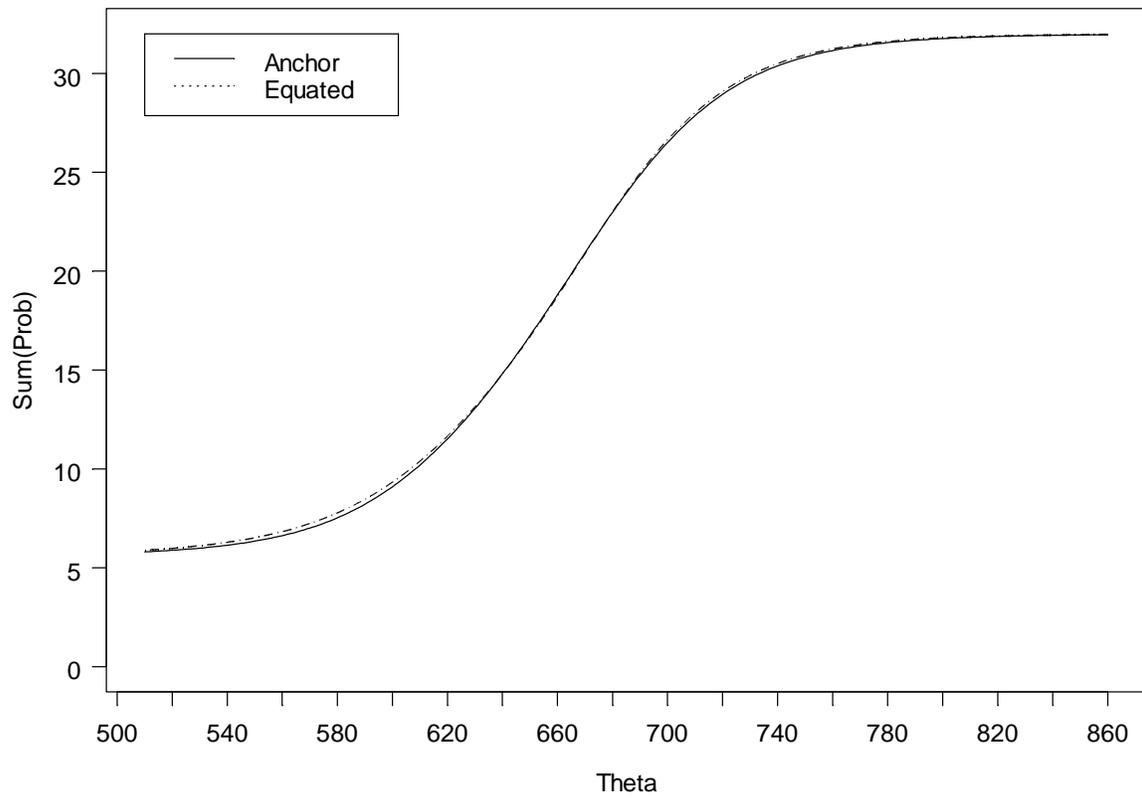
**Figure 6. 21: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 4 Mathematics**



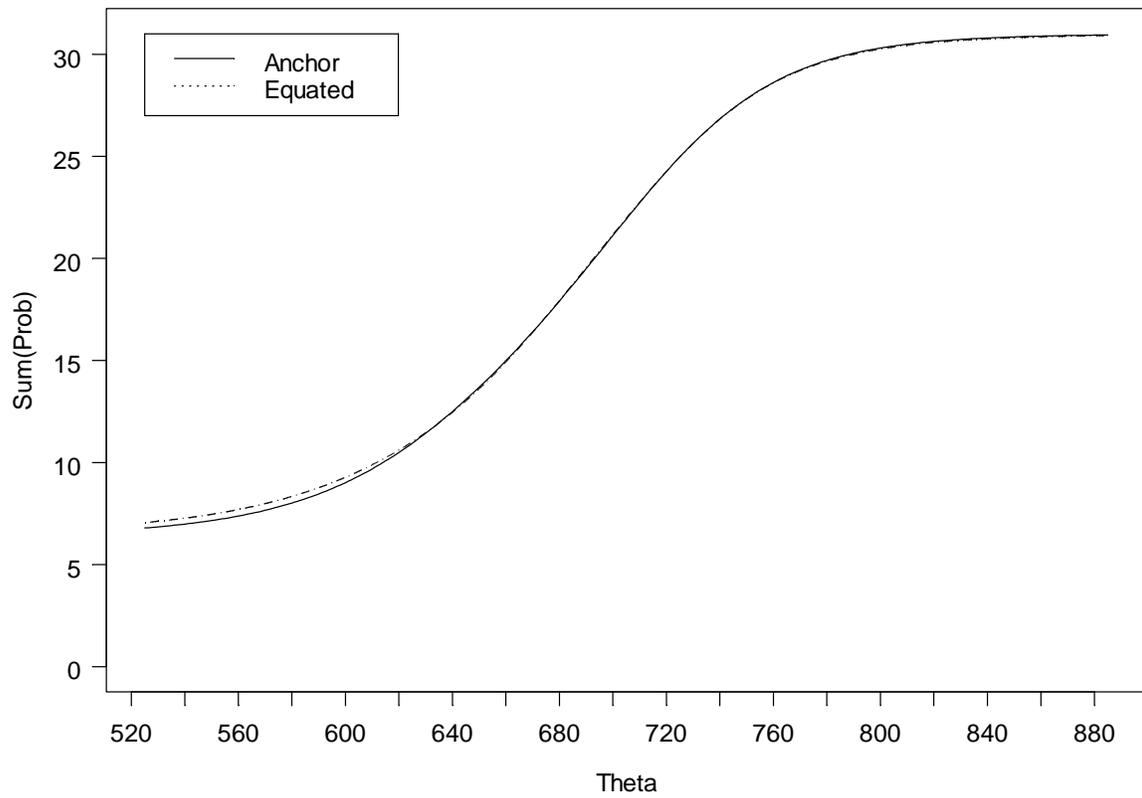
**Figure 6. 22: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 5 Mathematics**



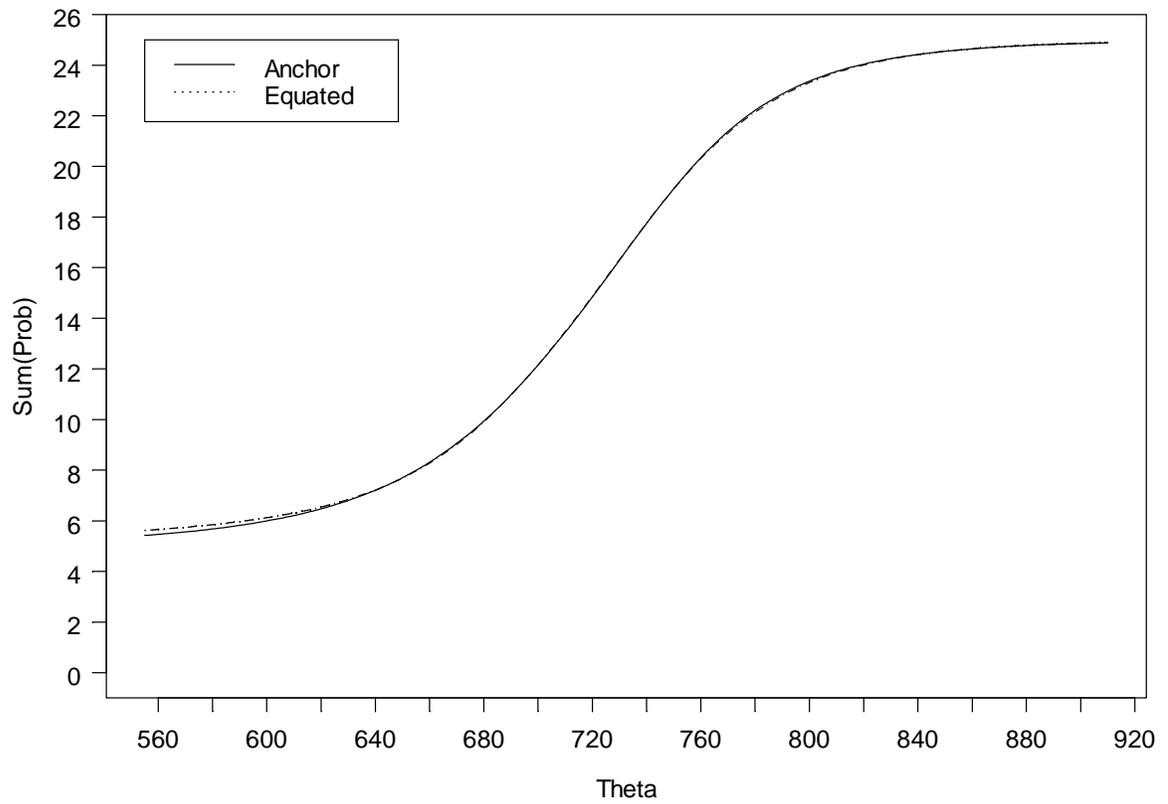
**Figure 6. 23: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 6 Mathematics**



**Figure 6. 24: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 7 Mathematics**



**Figure 6. 25: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 8 Mathematics**



**Figure 6. 26: Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items, Grade 10 Mathematics**

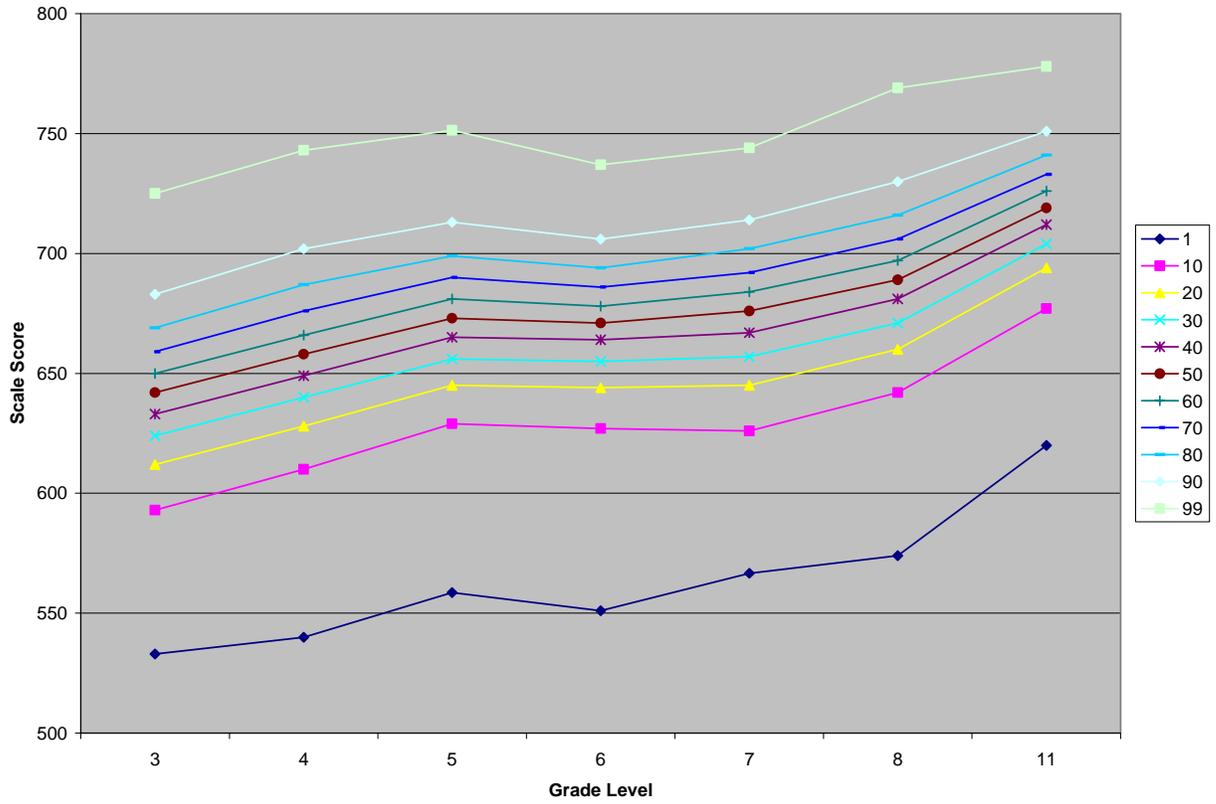


Figure 6. 27: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Communication Arts

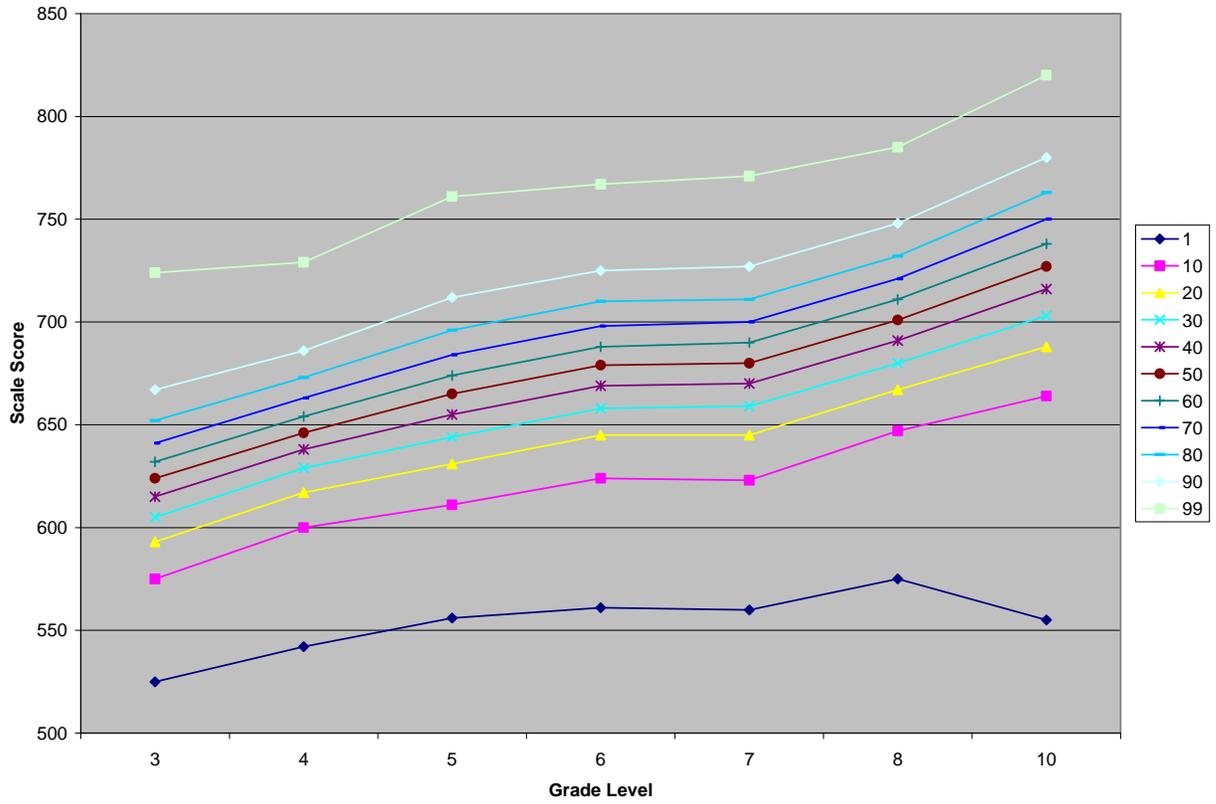


Figure 6. 28: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Mathematics

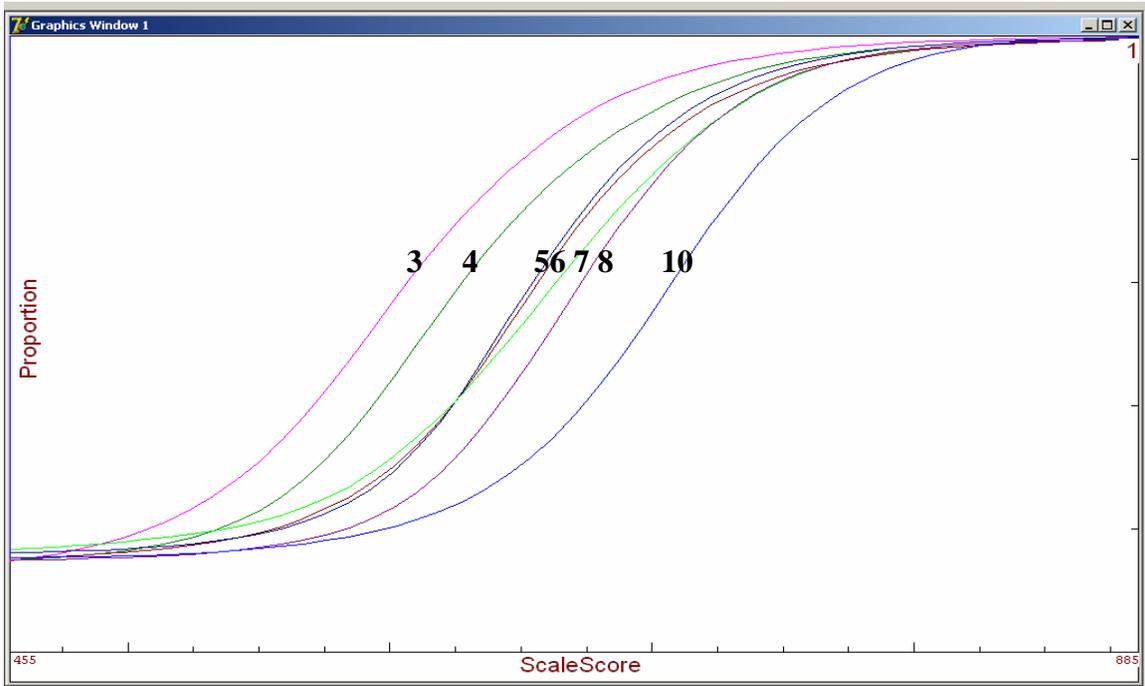


Figure 6.29: Test Characteristic Curves (TCC) by Grade, Communication Arts

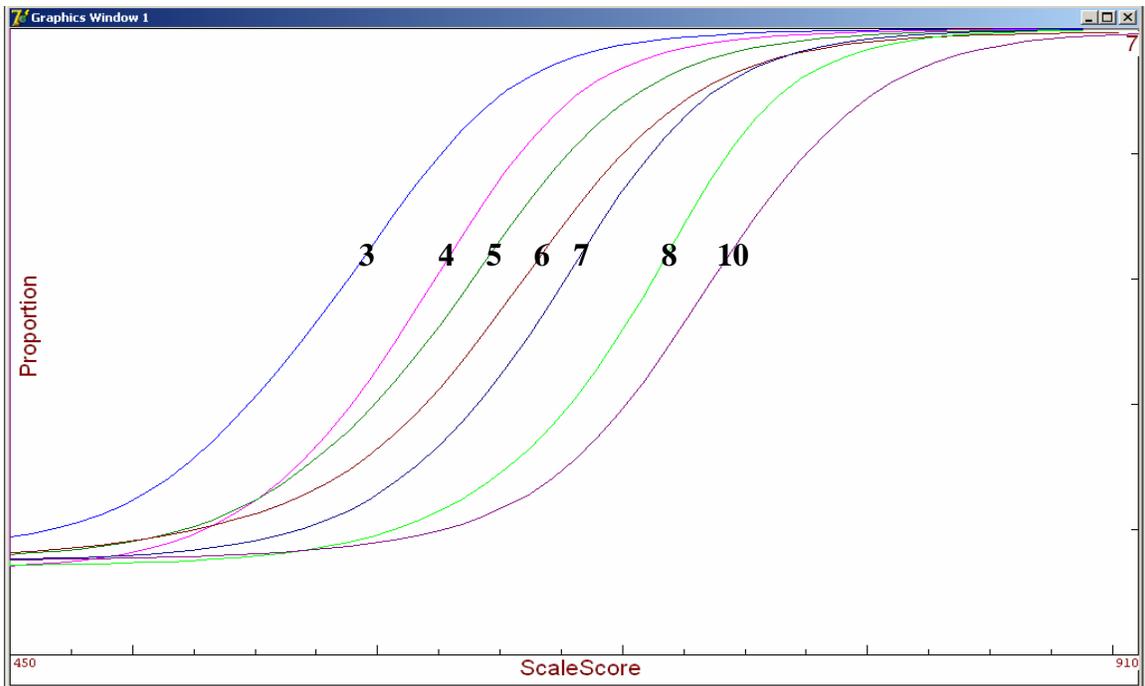


Figure 6.30: Test Characteristic Curves (TCC) by Grade, Mathematics

## Part 7: Test and Item Statistics

In this section, we present summary test statistics for each grade/content area MAP. This is followed by item-level statistics for each grade/content area MAP.

### *Test-Level Statistics*

Tables 7.1 and 7.2 present for each grade level of Communication Arts and Mathematics, respectively, the number of items and score points on each test, as well as the mean and standard deviation of the raw scores, *p-values* and item-total correlations (also known as item discrimination values). The statistics reported in these tables are based on the calibration sample that is described in Part 4.

The mean *p-value* is the average of all item *p-values* of a specific grade/content area. The mean item-total correlation ( $R_{it}$ ) is the average of all item point biserials of a specific grade/content area. The *p-value* and point biserial are explained in the next section, *Item-Level Statistics*.

### *Item-Level Statistics*

Tables 7.3 to 7.16 present the item statistics for each item by grade/content area. The tables include test session, item booklet number and part (if applicable), number of score points, *p-values*, item-total correlations ( $R_{it}$ ), and omit rates for each item by grade/content area. The constructed-response (CR) items appear in the tables first, followed by the multiple-choice (MC) items.

*P-Value:* The *p-value* is a measure of item difficulty. For a multiple-choice item, the *p-value* is calculated from the number of students who correctly respond to an item divided by the total number of students attempting the item. The value is reported as a proportion. For a constructed-response item, the *p-value* is calculated from the average score for the item divided by the maximum points possible and is also reported as a proportion. In terms of *p-values*, test scores tend to be most accurate when their average *p-values* are in the mid 0.50s to low 0.70s. However, in building a criterion referenced test, it is important to select items on the basis of content in addition to statistical criteria. As demonstrated in Tables 7.1 and 7.2, the average *p-values* associated with the MAP range from .57 to .77.

The range of *p-values* as well as the average *p-value* helps to determine whether a test measures the student population well. It is desirable for the test to measure well throughout the range of skills present at a given grade. That is, the items should measure the performance of both low-scoring and high-scoring students, as well as students in the center of the distribution. Having a range of *p-values* also helps to prevent floor and/or ceiling effects, that is, the test does not have large numbers of students at the minimum or maximum possible scores. The Communication Arts MAP has items with *p-values* ranging from the mid 0.10 to the 0.90s (see Tables 7.3 through 7.9). The *p-values* on the Mathematics MAP tend to range from the 0.20s and 0.30s to the 0.90s (see Tables 7.3 through 7.9).

*Item-Total Correlations:* An item-total correlation is the correlation between an item score and the total test score. It indicates how well an item differentiates between low- and high-achieving students. In general, items with correlations below .20 are said to be poorly discriminating. The overwhelming majority of the items in the MAP had item-test correlations above this threshold. Any item with a low item-total correlation was further analyzed to assure that the item was correctly keyed and that it is measuring content appropriate for the test.

*Omit Rate:* The omit rate for each item indicates the percentage of students who did not answer the item. Omit rates can be used to examine whether a test may be speeded; that is, if students do not have adequate time to answer all questions on the test. As a rule of thumb, an item is said to have a high-omit rate if more than 5% of students fail to respond to the item.

**Table 7. 1: MAP Means, Standard Deviations for Raw Scores, P-Values, Item-Total Correlation ( $R_{it}$ ): Communication Arts**

Grade	Total Items	Total Points	Mean Raw Score (SD)	Mean P-Value (SD)	Mean $R_{it}$ (SD)
3	57	69	50.41 (10.62)	.73 (.16)	.42 (.09)
4	56	66	47.20 (10.08)	.72 (.18)	.41 (.10)
5	56	66	43.80 (11.18)	.68 (.16)	.41 (.10)
6	56	65	42.85 (10.90)	.68 (.15)	.41 (.09)
7	61	73	46.63 (11.57)	.66 (.20)	.40 (.11)
8	60	69	46.88 (12.14)	.68 (.17)	.41 (.10)
11	60	73	46.67 (12.03)	.64 (.19)	.40 (.12)

**Table 7. 2: MAP Means, Standard Deviations for Raw Scores, P-Values, Item-Total Correlation ( $R_{it}$ ): Mathematics**

Grade	Total Items	Total Points	Mean Raw Score (SD)	Mean P-Value (SD)	Mean $R_{it}$ (SD)
3	60	67	50.74 (10.61)	.77 (.15)	.40 (.10)
4	65	77	54.77 (13.22)	.74 (.16)	.42 (.09)
5	62	71	49.85 (12.68)	.72 (.15)	.43 (.10)
6	61	69	45.94 (12.25)	.66 (.16)	.41 (.11)
7	62	71	43.47 (13.93)	.63 (.17)	.42 (.11)
8	64	76	41.92 (14.87)	.59 (.19)	.42 (.11)
10	61	75	39.70 (15.33)	.57 (.16)	.44 (.14)

**Table 7. 3: Item Statistics, Grade 3 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.86	0.38	0.19%	1	1	0.53	0.43	0.05%
1	2	0.62	0.42	0.21%	1	2	0.46	0.38	0.10%
1	3	0.61	0.35	0.47%	1	3	0.98	0.26	0.05%
1	4	0.70	0.44	0.33%	1	4	0.96	0.27	0.07%
1	5	0.52	0.47	0.54%	1	5	0.87	0.48	0.14%
1	6A	0.84	0.49	0.42%	1	6	0.78	0.34	0.21%
1	6B	0.94	0.48	0.42%	1	7	0.71	0.51	0.07%
2	1	0.73	0.58	0.00%	1	8	0.54	0.57	0.45%
3	1	0.83	0.40	0.12%	1	9	0.71	0.48	0.24%
3	2	0.88	0.51	0.09%	1	10	0.62	0.24	0.05%
3	3	0.60	0.46	0.35%	1	11	0.96	0.40	0.19%
3	4	0.90	0.41	0.33%	1	12	0.98	0.27	0.05%
3	5	0.62	0.37	0.42%	1	13	0.68	0.52	0.14%
3	6	0.87	0.46	0.57%	1	14	0.85	0.41	0.33%
3	7	0.94	0.37	0.14%	1	15	0.46	0.64	0.55%
3	8	0.89	0.44	0.21%	1	16	0.90	0.44	0.19%
3	9	0.85	0.53	0.68%	1	17	0.55	0.31	0.19%
3	10	0.94	0.32	0.47%	1	18	0.85	0.38	0.05%
3	11	0.86	0.56	0.54%	1	19	0.77	0.35	0.07%
3	12	0.85	0.35	0.47%	1	20	0.69	0.43	0.21%
3	13	0.58	0.32	0.57%	1	21	0.72	0.49	1.76%
3	14	0.82	0.47	0.80%	1	22	0.73	0.41	0.55%
3	15	0.45	0.40	0.28%	1	23	0.77	0.52	1.00%
3	16	0.49	0.36	0.68%	1	24	0.82	0.42	0.24%
3	17	0.61	0.35	5.01%	1	25	0.72	0.48	0.10%
3	18	0.91	0.49	0.54%	1	26	0.61	0.28	0.07%
3	19	0.78	0.47	1.03%	1	27	0.63	0.64	0.19%
3	20	0.47	0.43	1.01%	2	1	0.87	0.33	0.14%
3	21	0.78	0.49	1.31%	2	2	0.91	0.33	0.26%
3	22	0.76	0.44	0.24%	2	3	0.91	0.27	0.33%
3	23	0.77	0.49	1.05%	2	4	0.88	0.41	0.43%
3	24	0.86	0.44	1.43%	2	5	0.72	0.51	0.50%
3	25	0.57	0.34	0.12%	2	6	0.96	0.32	0.21%

**Table 7. 3: Item Statistics, Grade 3 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
3	26	0.59	0.30	0.28%	2	7	0.66	0.45	0.43%
3	27	0.44	0.31	0.87%	2	8	0.71	0.40	0.76%
3	28	0.73	0.55	0.42%	2	9	0.87	0.38	0.12%
3	29	0.89	0.48	0.78%	2	10	0.91	0.37	0.10%
3	30	0.66	0.50	0.64%	2	11	0.92	0.41	0.50%
3	31	0.94	0.46	0.40%	2	12	0.94	0.36	0.57%
3	32	0.79	0.43	0.59%	2	13	0.90	0.44	0.62%
3	33	0.52	0.29	1.38%	2	14	0.86	0.31	0.50%
3	34	0.94	0.49	0.24%	2	15	0.85	0.31	0.67%
3	35	0.72	0.41	0.42%	2	16	0.95	0.27	0.24%
3	36	0.46	0.26	0.42%	2	17	0.94	0.29	0.21%
3	37	0.84	0.53	0.47%	2	18	0.98	0.20	0.17%
3	38	0.80	0.50	1.95%	2	19	0.70	0.38	0.93%
3	39	0.73	0.42	0.47%	2	20	0.69	0.44	0.76%
3	40	0.91	0.50	1.15%	2	21	0.63	0.34	1.14%
3	41	0.87	0.50	1.74%	2	22	0.78	0.38	0.48%
3	42	0.45	0.36	0.68%	2	23	0.93	0.14	0.38%
3	43	0.78	0.61	0.38%	2	24	0.98	0.30	0.19%
3	44	0.80	0.36	0.45%	2	25	0.94	0.44	0.76%
3	45	0.87	0.42	0.47%	2	26	0.79	0.52	1.26%
3	46	0.54	0.38	0.54%	2	27	0.76	0.44	0.69%
3	47	0.69	0.34	0.52%	2	28	0.58	0.37	0.33%
3	48	0.45	0.24	0.52%	2	29	0.45	0.43	0.50%
3	49	0.40	0.21	0.78%	2	30	0.54	0.46	0.41%
					2	31	0.76	0.51	0.17%
					2	32	0.55	0.40	0.17%
					2	33	0.66	0.65	0.31%

**Table 7. 4: Item Statistics, Grade 4 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.67	0.39	0.05%	1	1	0.85	0.35	0.07%
1	2	0.67	0.35	0.05%	1	2	0.41	0.42	0.25%
1	3	0.46	0.48	0.34%	1	3	0.91	0.30	0.07%
1	4	0.44	0.44	0.22%	1	4	0.94	0.31	0.07%
1	5	0.49	0.40	0.12%	1	5	0.78	0.49	0.05%
1	6A	0.46	0.49	0.17%	1	6	0.82	0.43	0.00%
1	6B	0.93	0.42	0.17%	1	7	0.68	0.46	0.02%
2	1	0.97	0.29	0.10%	1	8	0.71	0.35	0.07%
2	2	0.98	0.33	0.10%	1	9	0.61	0.50	0.27%
2	3	0.45	0.35	0.10%	1	10	0.49	0.35	0.25%
2	4	0.83	0.48	0.24%	1	11	0.84	0.31	0.44%
2	5	0.97	0.39	0.39%	1	12	0.86	0.35	0.10%
2	6	0.88	0.42	0.15%	1	13	0.72	0.47	0.07%
2	7	0.74	0.42	0.19%	1	14	0.77	0.53	0.12%
2	8	0.91	0.42	0.17%	1	15	0.45	0.51	0.27%
2	9	0.93	0.30	0.22%	1	16	0.92	0.43	0.17%
2	10	0.88	0.49	0.05%	1	17	0.72	0.42	0.96%
2	11	0.94	0.40	0.19%	1	18	0.75	0.48	0.12%
2	12	0.61	0.25	0.17%	1	19	0.72	0.50	0.22%
2	13	0.90	0.50	0.15%	1	20	0.94	0.35	0.15%
2	14	0.64	0.48	0.39%	1	21	0.39	0.52	0.29%
2	15	0.91	0.40	0.12%	1	22	0.53	0.42	0.25%
2	16	0.67	0.39	0.27%	1	23	0.79	0.38	0.07%
2	17	0.89	0.52	0.07%	1	24	0.84	0.37	0.25%
2	18	0.41	0.34	0.10%	1	25	0.63	0.61	0.74%
2	19	0.74	0.42	0.24%	1	26	0.90	0.37	0.52%
2	20	0.87	0.65	0.19%	1	27	0.52	0.50	0.20%
2	21	0.40	0.46	0.94%	1	28	0.43	0.57	0.42%
2	22	0.83	0.35	0.10%	1	29	0.89	0.33	0.39%
2	23	0.83	0.48	0.07%	1	30	0.43	0.61	0.10%
2	24	0.82	0.35	0.34%	2	1	0.83	0.36	0.15%
2	25	0.82	0.23	0.19%	2	2	0.69	0.40	0.39%
2	26	0.39	0.29	0.44%	2	3	0.86	0.39	1.08%

**Table 7.4: Item Statistics, Grade 4 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
2	27	0.71	0.44	0.87%	2	4	0.60	0.42	1.96%
2	28	0.79	0.49	0.75%	2	5	0.62	0.27	1.20%
2	29	0.86	0.51	0.75%	2	6	0.86	0.31	1.62%
2	30	0.70	0.39	0.89%	2	7	0.79	0.49	2.82%
2	31	0.83	0.55	0.53%	2	8	0.61	0.40	2.80%
2	32	0.74	0.52	0.92%	2	9	0.70	0.51	3.78%
2	33	0.72	0.52	0.73%	2	10	0.67	0.39	4.49%
2	34	0.52	0.36	1.47%	2	11	0.99	0.19	0.07%
2	35	0.87	0.42	0.27%	2	12	0.97	0.27	0.20%
2	36	0.56	0.30	0.73%	2	13	0.46	0.46	0.39%
2	37	0.80	0.54	0.34%	2	14	0.81	0.31	0.74%
2	38	0.53	0.22	0.36%	2	15	0.85	0.37	0.25%
2	39	0.48	0.32	0.48%	2	16	0.87	0.37	0.39%
2	40	0.88	0.48	0.53%	2	17	0.93	0.41	0.47%
2	41	0.89	0.52	0.58%	2	18	0.60	0.42	1.47%
2	42	0.65	0.41	0.63%	2	19	0.89	0.31	1.64%
2	43	0.92	0.54	0.31%	2	20	0.83	0.40	0.42%
2	44	0.89	0.39	0.27%	2	21	0.76	0.38	0.15%
2	45	0.54	0.21	0.24%	2	22	0.90	0.43	0.27%
2	46	0.57	0.37	0.41%	2	23	0.75	0.52	0.22%
2	47	0.78	0.33	0.22%	2	24	0.90	0.36	0.20%
2	48	0.78	0.37	0.19%	2	25	0.88	0.35	0.34%
2	49	0.29	0.14	0.29%	2	26	0.75	0.53	1.20%
					2	27	0.67	0.43	0.86%
					2	28	0.89	0.32	1.20%
					2	29	0.74	0.56	0.37%
					2	30	0.86	0.40	0.44%
					2	31	0.56	0.45	0.57%
					2	32	0.55	0.44	0.86%
					2	33	0.65	0.60	0.17%
					2	34	0.74	0.53	0.29%
					2	35	0.67	0.55	0.10%

**Table 7. 5: Item Statistics, Grade 5 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.91	0.30	0.10%	1	1	0.80	0.49	0.68%
1	2	0.85	0.40	0.13%	1	2	0.50	0.30	0.13%
1	3	0.64	0.44	0.31%	1	3	0.80	0.37	0.20%
1	4A	0.52	0.46	0.23%	1	4	0.66	0.62	0.38%
1	4B	0.96	0.37	0.23%	1	5	0.52	0.54	0.33%
1	5	0.41	0.38	0.33%	1	6	0.79	0.43	0.15%
1	6	0.52	0.32	0.36%	1	7	0.64	0.43	0.25%
2	1	0.78	0.45	0.15%	1	8	0.83	0.46	0.20%
2	2	0.77	0.31	0.08%	1	9	0.64	0.52	0.18%
2	3	0.91	0.47	0.05%	1	10	0.39	0.43	0.30%
2	4	0.90	0.48	0.10%	1	11	0.74	0.44	0.33%
2	5	0.77	0.51	0.08%	1	12	0.73	0.56	0.25%
2	6	0.68	0.41	0.57%	1	13	0.76	0.47	0.18%
2	7	0.81	0.54	1.20%	1	14	0.66	0.54	0.68%
2	8	0.38	0.21	0.54%	1	15	0.49	0.32	0.20%
2	9	0.75	0.21	0.77%	1	16	0.94	0.35	0.15%
2	10	0.66	0.37	1.20%	1	17	0.55	0.60	0.28%
2	11	0.68	0.48	0.95%	1	18	0.85	0.39	0.23%
2	12	0.84	0.50	1.26%	1	19	0.70	0.50	0.38%
2	13	0.90	0.45	0.87%	1	20	0.88	0.44	0.23%
2	14	0.60	0.47	0.98%	1	21	0.58	0.36	0.18%
2	15	0.81	0.50	1.00%	1	22	0.56	0.36	0.71%
2	16	0.46	0.34	1.18%	1	23	0.58	0.32	0.53%
2	17	0.60	0.56	0.36%	1	24	0.41	0.59	0.55%
2	18	0.38	0.50	0.46%	1	25	0.89	0.37	0.25%
2	19	0.78	0.39	0.46%	1	26	0.73	0.35	0.30%
2	20	0.84	0.32	2.93%	1	27	0.63	0.33	0.20%
2	21	0.73	0.42	0.39%	2	1	0.82	0.34	0.18%
2	22	0.78	0.39	0.36%	2	2	0.73	0.49	0.81%
2	23	0.75	0.52	0.51%	2	3	0.73	0.34	0.43%
2	24	0.83	0.35	0.31%	2	4	0.68	0.49	0.20%
2	25	0.76	0.39	0.67%	2	5	0.78	0.45	0.33%
2	26	0.65	0.34	2.18%	2	6	0.80	0.42	0.76%

**Table 7. 5: Item Statistics, Grade 5 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
2	27	0.67	0.57	0.44%	2	7	0.80	0.43	0.98%
2	28	0.66	0.39	0.57%	2	8	0.60	0.44	1.83%
2	29	0.78	0.51	0.54%	2	9	0.93	0.38	0.13%
2	30	0.83	0.51	0.36%	2	10	0.89	0.20	0.33%
2	31	0.64	0.31	0.51%	2	11	0.95	0.36	0.25%
2	32	0.64	0.46	0.39%	2	12	0.80	0.49	0.55%
2	33	0.87	0.49	0.36%	2	13	0.77	0.31	0.33%
2	34	0.76	0.53	0.59%	2	14	0.68	0.42	0.20%
2	35	0.77	0.50	1.18%	2	15	0.85	0.55	0.53%
2	36	0.72	0.50	1.20%	2	16	0.93	0.30	0.60%
2	37	0.78	0.51	0.98%	2	17	0.73	0.18	0.38%
2	38	0.53	0.49	1.46%	2	18	0.98	0.23	0.33%
2	39	0.50	0.30	1.62%	2	19	0.67	0.48	0.73%
2	40	0.49	0.31	2.31%	2	20	0.93	0.26	3.45%
2	41	0.47	0.34	2.72%	2	21	0.82	0.35	0.50%
2	42	0.37	0.62	0.69%	2	22	0.81	0.36	0.93%
2	43	0.70	0.52	0.77%	2	23	0.71	0.55	0.73%
2	44	0.83	0.34	0.31%	2	24	0.54	0.45	1.91%
2	45	0.54	0.37	0.44%	2	25	0.83	0.43	0.28%
2	46	0.53	0.27	0.69%	2	26	0.66	0.52	0.48%
2	47	0.66	0.31	0.44%	2	27	0.63	0.55	0.60%
2	48	0.44	0.29	0.46%	2	28	0.66	0.40	0.68%
2	49	0.52	0.17	0.49%	2	29	0.87	0.43	0.58%
					2	30	0.75	0.45	0.45%
					2	31	0.61	0.48	0.43%
					2	32	0.78	0.51	0.38%
					2	33	0.86	0.58	0.18%
					2	34	0.21	0.49	0.45%
					2	35	0.51	0.60	0.83%

**Table 7. 6: Item Statistics, Grade 6 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.82	0.35	0.00%	1	1	0.90	0.36	0.00%
1	2	0.75	0.38	0.08%	1	2	0.43	0.32	0.07%
1	3	0.71	0.41	0.35%	1	3	0.73	0.38	0.27%
1	4	0.45	0.43	0.30%	1	4	0.52	0.49	0.15%
1	5A	0.53	0.54	0.75%	1	5	0.39	0.24	0.40%
1	5B	0.71	0.45	0.73%	1	6	0.90	0.34	0.52%
1	6	0.44	0.47	1.05%	1	7	0.60	0.56	0.27%
2	1	0.90	0.33	0.05%	1	8	0.79	0.17	0.35%
2	2	0.67	0.44	0.13%	1	9	0.56	0.41	0.12%
2	3	0.83	0.51	0.05%	1	10	0.56	0.32	0.15%
2	4	0.78	0.47	0.13%	1	11	0.85	0.46	0.17%
2	5	0.91	0.33	0.25%	1	12	0.56	0.47	0.17%
2	6	0.83	0.48	0.10%	1	13	0.55	0.45	0.15%
2	7	0.86	0.34	0.15%	1	14	0.98	0.20	0.45%
2	8	0.73	0.39	0.20%	1	15	0.78	0.42	0.05%
2	9	0.53	0.47	0.48%	1	16	0.39	0.42	0.12%
2	10	0.85	0.55	0.40%	1	17	0.90	0.27	0.15%
2	11	0.68	0.35	0.90%	1	18	0.46	0.37	0.27%
2	12	0.91	0.42	0.28%	1	19	0.72	0.51	0.80%
2	13	0.75	0.52	0.75%	1	20	0.34	0.31	0.10%
2	14	0.79	0.44	0.83%	1	21	0.63	0.30	0.45%
2	15	0.74	0.50	1.05%	1	22	0.46	0.11	0.32%
2	16	0.85	0.49	1.30%	1	23	0.65	0.40	0.27%
2	17	0.54	0.34	0.30%	1	24	0.50	0.54	1.47%
2	18	0.75	0.64	0.30%	1	25	0.54	0.40	0.12%
2	19	0.26	0.43	0.75%	1	26	0.32	0.25	0.25%
2	20	0.56	0.35	0.28%	1	27	0.40	0.45	0.60%
2	21	0.83	0.44	0.15%	2	1	0.83	0.27	0.10%
2	22	0.61	0.19	0.20%	2	2	0.78	0.31	0.32%
2	23	0.71	0.25	0.63%	2	3	0.69	0.43	1.47%
2	24	0.87	0.51	1.15%	2	4	0.79	0.56	0.12%
2	25	0.60	0.40	0.28%	2	5	0.67	0.43	0.55%
2	26	0.60	0.38	0.35%	2	6	0.72	0.45	1.34%

**Table 7. 6: Item Statistics, Grade 6 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
2	27	0.69	0.39	3.10%	2	7	0.63	0.33	1.12%
2	28	0.79	0.49	0.28%	2	8	0.78	0.41	1.99%
2	29	0.46	0.42	0.55%	2	9	0.79	0.57	2.51%
2	30	0.83	0.47	0.18%	2	10	0.98	0.24	0.15%
2	31	0.45	0.25	0.15%	2	11	0.87	0.39	0.30%
2	32	0.71	0.39	0.70%	2	12	0.73	0.55	0.42%
2	33	0.76	0.54	1.20%	2	13	0.66	0.49	0.50%
2	34	0.76	0.52	1.55%	2	14	0.49	0.45	0.50%
2	35	0.69	0.46	0.28%	2	15	0.77	0.31	0.72%
2	36	0.71	0.41	0.30%	2	16	0.71	0.47	1.44%
2	37	0.92	0.40	0.33%	2	17	0.65	0.35	1.59%
2	38	0.88	0.36	0.73%	2	18	0.83	0.51	0.47%
2	39	0.65	0.41	0.35%	2	19	0.63	0.54	0.80%
2	40	0.67	0.40	0.70%	2	20	0.82	0.43	0.15%
2	41	0.63	0.38	1.27%	2	21	0.64	0.45	0.57%
2	42	0.61	0.42	0.35%	2	22	0.72	0.48	0.22%
2	43	0.36	0.53	0.63%	2	23	0.67	0.38	0.42%
2	44	0.57	0.31	0.25%	2	24	0.50	0.40	0.42%
2	45	0.53	0.22	0.30%	2	25	0.70	0.45	0.92%
2	46	0.44	0.26	0.60%	2	26	0.68	0.51	0.40%
2	47	0.83	0.38	0.35%	2	27	0.64	0.44	0.50%
2	48	0.50	0.16	0.35%	2	28	0.75	0.40	0.25%
2	49	0.65	0.35	0.30%	2	29	0.54	0.51	0.37%
					2	30	0.46	0.37	0.40%
					2	31	0.38	0.34	0.40%
					2	32	0.66	0.59	0.10%
					2	33	0.76	0.58	0.32%
					2	34	0.85	0.51	0.32%

**Table 7.7: Item Statistics, Grade 7 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.53	0.38	0.06%	1	1	0.65	0.18	0.06%
1	2	0.27	0.38	0.13%	1	2	0.83	0.35	0.06%
1	3	0.46	0.44	0.32%	1	3	0.92	0.25	0.21%
1	4A	0.50	0.44	0.53%	1	4	0.84	0.30	0.04%
1	4B	0.87	0.23	0.74%	1	5	0.54	0.48	0.29%
1	5A	0.82	0.53	0.46%	1	6	0.62	0.40	0.15%
1	5B	0.93	0.45	0.46%	1	7	0.50	0.66	2.27%
1	6	0.26	0.43	1.49%	1	8	0.92	0.41	0.06%
2	1	0.69	0.62	0.00%	1	9	0.41	0.30	0.17%
3	1	0.88	0.35	0.06%	1	10	0.83	0.52	0.17%
3	2	0.94	0.32	0.08%	1	11	0.48	0.47	0.19%
3	3	0.90	0.44	0.15%	1	12	0.70	0.40	0.11%
3	4	0.96	0.31	0.08%	1	13	0.82	0.29	0.19%
3	5	0.94	0.44	0.08%	1	14	0.34	0.56	1.15%
3	6	0.74	0.42	0.30%	1	15	0.63	0.21	0.23%
3	7	0.59	0.12	1.43%	1	16	0.77	0.40	0.13%
3	8	0.81	0.50	0.42%	1	17	0.43	0.28	0.34%
3	9	0.73	0.41	0.51%	1	18	0.31	0.34	0.21%
3	10	0.76	0.53	0.40%	1	19	0.49	0.33	0.13%
3	11	0.72	0.54	0.08%	1	20	0.56	0.33	0.13%
3	12	0.73	0.49	0.32%	1	21	0.22	0.55	4.77%
3	13	0.91	0.47	0.25%	1	22	0.53	0.31	0.25%
3	14	0.58	0.46	0.59%	1	23	0.45	0.30	0.27%
3	15	0.73	0.49	0.55%	1	24	0.30	0.38	0.15%
3	16	0.85	0.55	0.46%	1	25	0.50	0.28	0.63%
3	17	0.58	0.35	0.53%	1	26	0.67	0.44	0.06%
3	18	0.58	0.41	0.59%	1	27	0.33	0.55	1.99%
3	19	0.57	0.54	0.59%	2	1	0.73	0.35	0.27%
3	20	0.24	0.47	0.89%	2	2	0.66	0.47	0.48%
3	21	0.63	0.36	0.30%	2	3	0.56	0.40	0.90%
3	22	0.68	0.44	0.25%	2	4	0.77	0.39	0.78%
3	23	0.91	0.42	0.17%	2	5	0.76	0.38	0.67%
3	24	0.40	0.39	0.57%	2	6	0.86	0.40	1.34%

**Table 7. 7: Item Statistics, Grade 7 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
3	25	0.67	0.48	0.70%	2	7	0.54	0.46	4.79%
3	26	0.73	0.43	0.11%	2	8	0.71	0.38	4.12%
3	27	0.67	0.37	1.41%	2	9	0.75	0.31	4.94%
3	28	0.69	0.43	0.04%	2	10	0.94	0.36	0.25%
3	29	0.87	0.31	0.34%	2	11	0.68	0.41	0.50%
3	30	0.76	0.51	0.13%	2	12	0.81	0.41	0.06%
3	31	0.87	0.45	0.30%	2	13	0.66	0.58	0.46%
3	32	0.57	0.42	0.87%	2	14	0.65	0.38	0.80%
3	33	0.76	0.47	0.46%	2	15	0.87	0.41	0.90%
3	34	0.82	0.43	0.53%	2	16	0.49	0.42	0.69%
3	35	0.45	0.32	0.42%	2	17	0.70	0.54	1.36%
3	36	0.42	0.44	0.51%	2	18	0.72	0.46	1.34%
3	37	0.42	0.38	0.44%	2	19	0.73	0.41	0.74%
3	38	0.45	0.30	0.55%	2	20	0.64	0.52	0.88%
3	39	0.65	0.34	0.68%	2	21	0.82	0.47	1.07%
3	40	0.78	0.56	0.63%	2	22	0.76	0.48	0.34%
3	41	0.73	0.42	0.61%	2	23	0.80	0.46	0.46%
3	42	0.49	0.47	0.61%	2	24	0.64	0.59	1.89%
3	43	0.83	0.36	2.00%	2	25	0.67	0.44	0.34%
3	44	0.75	0.19	2.53%	2	26	0.51	0.41	0.21%
3	45	0.76	0.33	2.02%	2	27	0.71	0.38	0.15%
3	46	0.52	0.21	2.30%	2	28	0.68	0.56	0.32%
3	47	0.56	0.38	2.23%	2	29	0.52	0.47	0.67%
3	48	0.71	0.42	2.17%	2	30	0.51	0.43	0.57%
3	49	0.49	0.28	2.27%	2	31	0.55	0.59	0.88%
3	50	0.21	0.17	2.23%	2	32	0.56	0.41	1.11%
3	51	0.36	0.07	2.17%	2	33	0.72	0.55	0.78%
3	52	0.31	0.23	2.13%	2	34	0.53	0.61	0.67%
					2	35	0.50	0.66	0.55%

**Table 7. 8: Item Statistics, Grade 8 Communication Arts and Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.88	0.45	0.00%	1	1	0.94	0.32	0.09%
1	2	0.90	0.44	0.02%	1	2	0.50	0.48	0.22%
1	3	0.67	0.44	0.33%	1	3	0.76	0.30	0.29%
1	4	0.58	0.51	0.42%	1	4	0.33	0.62	1.42%
1	5	0.62	0.58	0.46%	1	5	0.58	0.43	0.09%
1	6A	0.78	0.53	0.44%	1	6	0.21	0.17	0.11%
1	6B	0.61	0.18	0.44%	1	7	0.34	0.60	1.36%
2	1	0.93	0.33	0.11%	1	8	0.75	0.47	0.15%
2	2	0.81	0.45	0.20%	1	9	0.65	0.41	0.35%
2	3	0.80	0.36	0.04%	1	10	0.41	0.34	0.11%
2	4	0.83	0.48	0.13%	1	11	0.66	0.45	0.24%
2	5	0.78	0.49	0.09%	1	12	0.32	0.64	3.93%
2	6	0.56	0.33	0.15%	1	13	0.28	0.42	0.42%
2	7	0.82	0.43	0.13%	1	14	0.50	0.44	0.15%
2	8	0.81	0.42	0.02%	1	15	0.70	0.40	0.09%
2	9	0.92	0.39	0.20%	1	16	0.47	0.48	0.62%
2	10	0.66	0.46	0.53%	1	17	0.38	0.62	2.39%
2	11	0.86	0.49	0.66%	1	18	0.65	0.45	0.64%
2	12	0.67	0.27	0.02%	1	19	0.27	0.41	0.18%
2	13	0.81	0.36	0.07%	1	20	0.59	0.48	0.20%
2	14	0.91	0.45	0.04%	1	21	0.32	0.46	0.31%
2	15	0.88	0.49	0.13%	1	22	0.40	0.52	0.26%
2	16	0.96	0.37	0.11%	1	23	0.44	0.58	0.11%
2	17	0.85	0.38	0.33%	1	24	0.64	0.44	0.24%
2	18	0.80	0.49	0.51%	1	25	0.43	0.51	0.40%
2	19	0.90	0.32	0.62%	1	26	0.68	0.34	0.55%
2	20	0.65	0.52	0.97%	1	27	0.48	0.59	0.73%
2	21	0.57	0.60	0.90%	1	28	0.31	0.43	0.29%
2	22	0.46	0.45	0.22%	1	29	0.57	0.45	0.26%
2	23	0.70	0.43	0.26%	1	30	0.28	0.71	8.02%
2	24	0.40	0.28	0.35%	2	1	0.81	0.26	0.09%
2	25	0.63	0.46	0.35%	2	2	0.49	0.37	0.22%

2	26	0.79	0.53	0.53%	2	3	0.38	0.38	0.59%
---	----	------	------	-------	---	---	------	------	-------

**Table 7. 8: Item Statistics, Grade 8 Communication Arts and Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
2	27	0.69	0.47	1.05%	2	4	0.82	0.30	0.15%
2	28	0.69	0.48	1.31%	2	5	0.77	0.42	0.44%
2	29	0.75	0.28	0.20%	2	6	0.61	0.42	0.62%
2	30	0.93	0.43	0.24%	2	7	0.63	0.37	0.86%
2	31	0.83	0.44	0.42%	2	8	0.85	0.15	0.51%
2	32	0.56	0.52	0.53%	2	9	0.68	0.30	0.31%
2	33	0.68	0.49	0.53%	2	10	0.90	0.25	0.15%
2	34	0.37	0.25	1.25%	2	11	0.94	0.24	0.13%
2	35	0.84	0.41	0.29%	2	12	0.84	0.44	0.18%
2	36	0.62	0.41	0.29%	2	13	0.91	0.30	0.22%
2	37	0.69	0.47	0.37%	2	14	0.82	0.28	0.13%
2	38	0.77	0.43	0.37%	2	15	0.67	0.39	0.70%
2	39	0.73	0.45	0.40%	2	16	0.92	0.35	0.11%
2	40	0.48	0.32	0.46%	2	17	0.80	0.42	0.15%
2	41	0.45	0.32	0.86%	2	18	0.70	0.49	0.04%
2	42	0.57	0.62	0.90%	2	19	0.64	0.47	0.53%
2	43	0.66	0.65	1.45%	2	20	0.43	0.46	0.75%
2	44	0.61	0.42	0.57%	2	21	0.75	0.38	0.66%
2	45	0.35	0.37	0.57%	2	22	0.74	0.45	0.79%
2	46	0.69	0.48	0.55%	2	23	0.79	0.52	0.26%
2	47	0.65	0.42	0.70%	2	24	0.55	0.45	0.29%
2	48	0.56	0.32	0.84%	2	25	0.73	0.45	2.85%
2	49	0.56	0.31	0.75%	2	26	0.48	0.33	3.53%
2	50	0.47	0.25	0.77%	2	27	0.53	0.43	0.46%
2	51	0.47	0.23	0.90%	2	28	0.55	0.36	0.57%
2	52	0.23	0.23	0.75%	2	29	0.59	0.39	1.07%
2	53	0.49	0.15	0.81%	2	30	0.44	0.38	1.47%
					2	31	0.46	0.25	0.90%
					2	32	0.56	0.48	0.33%
					2	33	0.44	0.58	1.03%
					2	34	0.38	0.57	0.81%

**Table 7.9: Item Statistics, Grade 11 Communication Arts and Grade 10 Mathematics**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
1	1	0.74	0.34	0.10%	1	1	0.60	0.32	0.60%
1	2	0.85	0.41	0.10%	1	2	0.63	0.56	0.36%
1	3	0.66	0.57	0.44%	1	3	0.56	0.26	0.76%
1	4	0.79	0.50	0.65%	1	4	0.96	0.27	0.11%
1	5	0.64	0.53	0.52%	1	5	0.63	0.64	2.23%
1	6A	0.52	0.49	1.19%	1	6	0.35	0.38	0.36%
1	6B	0.87	0.45	1.19%	1	7	0.79	0.46	0.40%
2	1	0.68	0.61	0.00%	1	8	0.45	0.41	0.11%
3	1	0.41	0.30	0.31%	1	9	0.49	0.54	0.20%
3	2	0.95	0.30	0.23%	1	10	0.61	0.57	1.45%
3	3	0.49	0.55	0.36%	1	11	0.51	0.36	0.42%
3	4	0.45	0.42	0.41%	1	12	0.74	0.35	0.11%
3	5	0.65	0.44	0.36%	1	13	0.65	0.40	0.18%
3	6	0.78	0.31	0.41%	1	14	0.65	0.09	0.29%
3	7	0.68	0.39	0.31%	1	15	0.48	0.67	3.30%
3	8	0.88	0.43	0.26%	1	16	0.80	0.48	0.25%
3	9	0.77	0.40	0.36%	1	17	0.37	0.34	0.36%
3	10	0.75	0.38	0.36%	1	18	0.60	0.50	0.33%
3	11	0.82	0.29	0.36%	1	19	0.31	0.12	0.51%
3	12	0.27	0.19	0.44%	1	20	0.50	0.42	1.24%
3	13	0.57	0.44	0.78%	1	21	0.58	0.41	0.31%
3	14	0.80	0.50	0.93%	1	22	0.64	0.55	0.33%
3	15	0.78	0.38	1.13%	1	23	0.72	0.44	0.29%
3	16	0.75	0.50	1.21%	1	24	0.39	0.27	0.60%
3	17	0.58	0.35	1.47%	1	25	0.31	0.62	4.46%
3	18	0.18	0.17	2.30%	1	26	0.78	0.54	0.36%
3	19	0.59	0.31	1.47%	1	27	0.53	0.49	0.31%
3	20	0.66	0.44	1.55%	1	28	0.31	0.31	0.33%
3	21	0.64	0.45	0.80%	1	29	0.31	0.60	5.82%
3	22	0.51	0.60	1.91%	1	30	0.68	0.43	0.22%
3	23	0.90	0.40	0.49%	1	31	0.66	0.48	0.42%

3	24	0.41	0.26	0.54%	1	32	0.16	0.67	5.53%
3	25	0.74	0.48	0.83%	2	1	0.80	0.43	0.20%

**Table 7.9: Item Statistics, Grade 11 Communication Arts and Grade 10 Mathematics (Continued)**

Communication Arts					Mathematics				
Session	Item	P-Value	R <sub>it</sub>	Omit Rate	Session	Item	P-Value	R <sub>it</sub>	Omit Rate
3	26	0.62	0.45	0.75%	2	2	0.41	0.40	0.65%
3	27	0.60	0.43	0.75%	2	3	0.47	0.43	0.25%
3	28	0.81	0.50	0.49%	2	4	0.57	0.41	0.85%
3	29	0.75	0.22	0.52%	2	5	0.60	0.38	0.27%
3	30	0.56	0.34	0.57%	2	6	0.88	0.16	0.02%
3	31	0.88	0.47	0.85%	2	7	0.61	0.41	0.58%
3	32	0.94	0.36	0.65%	2	8	0.57	0.54	0.36%
3	33	0.77	0.42	0.70%	2	9	0.61	0.40	0.56%
3	34	0.70	0.55	0.67%	2	10	0.68	0.31	0.60%
3	35	0.68	0.48	0.54%	2	11	0.55	0.50	0.20%
3	36	0.34	0.33	0.75%	2	12	0.73	0.54	0.38%
3	37	0.59	0.34	0.75%	2	13	0.77	0.52	1.13%
3	38	0.74	0.54	0.65%	2	14	0.48	0.31	1.20%
3	39	0.75	0.54	0.75%	2	15	0.62	0.42	0.25%
3	40	0.69	0.50	0.85%	2	16	0.60	0.42	0.45%
3	41	0.67	0.42	0.70%	2	17	0.54	0.48	0.40%
3	42	0.43	0.62	2.06%	2	18	0.71	0.49	0.42%
3	43	0.79	0.42	0.78%	2	19	0.74	0.26	0.47%
3	44	0.68	0.51	1.06%	2	20	0.55	0.36	0.31%
3	45	0.88	0.36	0.78%	2	21	0.44	0.57	0.60%
3	46	0.64	0.36	0.78%	2	22	0.61	0.44	0.42%
3	47	0.73	0.42	0.72%	2	23	0.61	0.45	0.49%
3	48	0.45	0.29	0.88%	2	24	0.70	0.40	1.11%
3	49	0.28	0.10	1.34%	2	25	0.52	0.42	1.18%
3	50	0.28	0.20	1.00%	2	26	0.26	0.61	2.56%
3	51	0.19	0.17	0.93%	2	27	0.46	0.75	3.65%
3	52	0.32	0.17	0.80%	2	28	0.33	0.66	0.96%
					2	29	0.42	0.65	1.42%

## Part 8. Fairness

A common concern held by educators, psychometricians, and test developers is the fairness of a test. When tests measure different things for different groups, then they can be called biased (Camilli & Shepard, 1994). One goal of the test development cycle is to minimize test bias.

The position of CTB/McGraw-Hill concerning test bias is based on two general propositions. First, students may differ in their background knowledge, cognitive and academic skills, language, attitudes, and values. To the degree that these differences are large, no one curriculum and no one set of instructional materials will be equally suitable for all; therefore, the appropriateness of a test for different subgroups may differ. Furthermore, it is difficult to specify what amount of difference can be called large and to determine how these differences will affect the outcome of a particular test.

Second, schools have been assigned the tasks of developing certain basic cognitive skills and supporting development of these skills equitably among all students. Therefore, there is a need for tests that measure the common skills and bodies of knowledge that are common to all learners. The test publisher's task is to develop assessments that measure these key cognitive skills without introducing extraneous or construct-irrelevant elements in the performances on which the measurement is based. If these tests require that students have culturally-specific knowledge and skills not taught in school, differences in performance among students can occur because of differences in student background and out-of-school learning. In order to lessen this bias, CTB/McGraw-Hill strives to minimize the role of the extraneous elements, thereby, increasing the number of students for whom the test is appropriate.

In this chapter, we discuss the steps taken to minimize bias both through the test development process and through statistical procedures.

### *Minimizing Bias through Careful Test Development*

The development of a test that is fair for all examinees begins in the early stages of planning and development. Part 3 discusses the steps, such as Content and Bias Review Workshops, taken by CTB to minimize bias on the MAP. Below, more detail is given regarding the item and test development processes that were used to minimize bias.

First, careful attention was paid to content validity during the item development and item-selection process. Bias can occur if the test is measuring different things for different groups. By eliminating irrelevant skills or knowledge from the items, the possibility of bias is reduced.

Second, item writers and test developers followed published guidelines for reducing or eliminating bias. These included Guidelines for Bias-Free Publishing (MacMillan/McGraw-Hill, 1993a) and Reflecting Diversity: Multicultural Guidelines for Educational Publishing Professionals (Macmillan/McGraw-Hill, 1993b). Test developers

**Copyright © 2007 by Missouri Department of Elementary and Secondary Education**

reviewed the items and other testing materials with these guidelines in mind. Internal editorial reviews were conducted by at least three different people: a content editor who directly supervised the item writers; a style editor; and a content supervisor. The final test was again reviewed by at least these same people, and was also subjected to an independent review by a quality assurance editor.

Third, careful attention is given to item statistics throughout the test development process. As part of the test assembly process, attempts are made to avoid using or reusing items with poor statistical fit or distractors with positive point biserial correlations, since this may indicate that an item is tapping an ability that is irrelevant to the construct being measured. Differential item functioning (DIF) statistics are also examined during test construction. Items that have exhibited significant DIF against one or more subgroups are removed from further consideration unless it is essential to include them in order to meet content specifications.

### ***Evaluating Bias through Differential Item Functioning Statistics***

After administering both the field and operational tests, an empirical approach known as differential item functioning (DIF) was used to examine student performance on the items. The DIF statistics indicate the degree to which members of a particular subgroup perform better or worse than expected on each item as compared to the reference group. The DIF procedures used and the results of these analyses are detailed in this section.

DIF statistics are used to quantify differences in item performance between two groups after controlling for examinees' overall achievement level. Two DIF statistics that are commonly used for this purpose are the Mantel-Haenszel statistics (1959) and the Standardized Mean Difference (SMD) between the reference and focal groups proposed by Dorans and Schmitt (1991).

For selected-response items, the Mantel-Haenszel ( $\chi^2_{MH}$ ) statistic was used to evaluate potential DIF items. In the MH procedure, subgroups are matched by their raw total test score using a contingency table with  $K$  levels. When applying the MH procedure, the log-odds ratio  $\alpha$  is assumed to be constant across the  $K$  matched levels. The  $\chi^2_{MH}$ , then, estimates a pooled common odds ratio. Taking the natural logarithm of the common odds ratio and its confidence limits and multiplying these with the constant  $-2.35$ , the resulting values may then be placed on the MH delta metric ( $\Delta_{MH}$ ) for interpretive purposes. Items were flagged for DIF using the following criteria:

- Moderate DIF: absolute value of the Mantel-Haenszel ( $\Delta_{MH}$ ) is significantly greater than zero (at the .05 level) and  $-1.5 \leq \Delta_{MH} \leq -1$  or  $1 \leq \Delta_{MH} \leq 1.5$ .
- Large DIF: absolute value of the Mantel-Haenszel ( $\Delta_{MH}$ ) that is significantly greater than zero (at the .05 level) and  $|\Delta_{MH}|$  exceeds 1.5.

For constructed-response items, an effect size (ES) statistic based on the Mantel (1963)  $\chi^2$  will be used. ES is obtained by dividing the standardized mean difference

(SMD) statistics by the standard deviation of the item. (A detailed description of these procedures can be found in Zwick, et al., 1993). Items are flagged using the same rules that are used in NAEP:

- Moderate DIF: If the Mantel statistic is significant ( $p < .05$ ) and  $|ES|$  is between 0.17 and 0.25
- Large DIF: If the Mantel statistic is significant ( $p < .05$ ) and  $|ES| \geq 0.25$

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group. Tables 7.1 and 7.2 show the DIF results for the following subgroups:

- **Gender:** Focal group is Females; reference group is Males.
- **Ethnicity:** Focal groups are Black, Hispanic, Asian/Pacific Islander, Native American/Alaskan; reference group is White.
- **English Language Learners (ELL):** Focal group is students in ELL programs; reference group is all others.
- **Special Education Students:** Focal group is students with Individualized Education Programs (IEPs); reference group is all others.
- **Low Socioeconomic Status (SES):** Focal group is students who received free or reduced lunch; reference group is all others.
- **Disability:** Focal group is students who indicated a disability; reference group is all others.
- **Accommodations:** Focal group is students who received one or more testing accommodations; reference group is all others.
- **Migrant:** Focal group is students who indicated migrant status; reference group is all others.

A negative SMD value indicates that the focal group has a lower mean item score than the reference group, whereas a positive value implies that the focal group has a higher mean item score than the reference group, conditioned on the matching test score.

The DIF analyses are not performed for subgroups of less than 100 students. In these cases, the statistical procedures do not have sufficient power to detect differences should they exist.

Tables 8.1 and 8.2 summarize the number of DIF flags by grade for each focal group. For example in Grade 4 Communication Arts, there are 7 moderate DIF flags for the Asian/Pacific Islander subgroup. In this case, three of the items were flagged in favor of

Asian/Pacific Islander students while four of the items were flagged against Asian/Pacific Islander students. One item was flagged for moderate DIF against Hispanic students, one item was flagged for moderate DIF against Native American/Alaskan students, and one item was flagged for moderate DIF against students receiving accommodations. In Grade 4 Communication Arts, there were not enough Migrant students to analyze the data. In Appendix A, we report the item number along with its DIF flag (moderate or large).

Again, any items included in the MAP (including those items flagged for DIF) have been thoroughly reviewed for content and bias by Missouri teachers, DESE staff, and CTB Content Development staff. Further, these items were reviewed for possible DIF flags during the field-test stage of test development. The DIF flags found on the operational assessment do not necessarily indicate that an item is biased; rather, DIF flags indicate that the item functions differently for members of different groups (Camilli & Shepard, 1994). All items flagged for DIF in the tables above had been thoroughly reviewed before inclusion on the operational MAP to insure that they do not tap knowledge or specific ability irrelevant to the construct the test intends to measure.

During the operational administration, DIF is again measured. At this point, items are not dropped from the scoring process unless they are found to be blatantly biased (both statistically and content wise). It is highly unlikely that an item with blatant bias would be found during the operational administration because of the steps taken during the test development cycle to minimize bias (see above and Part 3). Items flagged for DIF during the operational administration were retained for scoring purposes.

**Table 8. 1: Number of Flagged Items for DIF using SMD Statistics, Communication Arts**

<b>Grade</b>	<b>Group</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>3</b>	Female	32195	1	0	0	0
	Asian/Pacific Islander	1154	2	3	0	2
	Black	12118	0	2	0	0
	Hispanic	2576	0	4	0	0
	Native American/Alaskan	265	2	0	0	0
	ELL	2476	0	2	0	0
	IEP	9983	0	0	0	0
	Migrant	103	4	7	0	0
	SES	29730	0	0	0	0
	Accommodations	6271	0	2	0	0
	Disability	10187	0	0	0	0
<b>4</b>	Female	31974	0	0	0	0
	Asian/Pacific Islander	1145	3	4	0	0
	Black	11782	0	0	0	0
	Hispanic	2341	0	1	0	0
	Native American/Alaskan	260	0	1	0	0
	ELL	2175	0	0	0	0
	IEP	9819	0	0	0	0
	Migrant	98	--	--	--	--
	SES	28678	0	0	0	0
	Accommodations	6990	0	1	0	0
	Disability	10013	0	0	0	0
<b>5</b>	Female	32039	2	2	0	1
	Asian/Pacific Islander	1160	1	2	1	1
	Black	11530	0	1	0	0
	Hispanic	2209	1	4	0	0
	Native American/Alaskan	266	1	1	0	0
	ELL	1996	0	4	0	0
	IEP	9584	0	0	0	0
	Migrant	106	5	3	1	4
	SES	28174	0	0	0	0
	Accommodations	7352	0	0	0	0
	Disability	9761	0	0	0	0
<b>6</b>	Female	31999	3	2	0	0
	Asian/Pacific Islander	1179	1	1	0	0
	Black	12005	0	4	0	0
	Hispanic	2237	0	0	0	0
	Native American/Alaskan	273	0	1	0	0
	ELL	1866	0	0	0	0
	IEP	9204	0	0	0	1
	Migrant	200	0	0	0	0
	SES	27180	0	0	0	0
	Accommodations	7351	0	0	0	1
	Disability	9330	0	0	0	1

**Table 8. 1: Number of Flagged Items for DIF using SMD Statistics, Communication Arts  
(Continued)**

<b>Grade</b>	<b>Group</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>7</b>	Female	32588	3	5	1	2
	Asian/Pacific Islander	1027	5	2	1	0
	Black	12424	1	2	0	0
	Hispanic	2145	0	1	0	1
	Native American/Alaskan	310	0	0	0	0
	ELL	1710	0	0	0	1
	IEP	8750	0	0	0	1
	Migrant	195	2	5	2	0
	SES	26461	0	0	0	0
	Accommodations	6816	0	0	0	1
	Disability	8887	0	0	0	1
<b>8</b>	Female	34030	1	3	0	2
	Asian/Pacific Islander	1035	5	4	0	3
	Black	13166	0	1	0	0
	Hispanic	2104	0	2	0	0
	Native American/Alaskan	334	0	0	0	0
	ELL	1730	1	0	0	2
	IEP	9493	0	0	0	0
	Migrant	62	--	--	--	--
	SES	26620	0	0	0	0
	Accommodations	7300	0	0	0	0
	Disability	9562	0	0	0	0
<b>11</b>	Female	30620	2	3	0	1
	Asian/Pacific Islander	969	2	4	0	1
	Black	9264	0	2	0	0
	Hispanic	1476	0	1	1	0
	Native American/Alaskan	292	2	0	0	0
	ELL	1333	0	1	0	1
	IEP	6880	0	1	0	1
	Migrant	55	--	--	--	--
	SES	16713	0	0	0	0
	Accommodations	4777	0	1	0	1
	Disability	6921	0	1	0	1

**Table 8. 2: Number of Flagged Items for DIF using SMD Statistics, Mathematics**

<b>Grade</b>	<b>Group</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>3</b>	Female	32280	0	0	0	0
	Asian/Pacific Islander	1186	3	2	2	3
	Black	12202	1	1	0	0
	Hispanic	2622	1	3	0	0
	Native American/Alaskan	267	2	0	1	0
	ELL	2565	0	2	0	1
	IEP	10092	0	0	0	0
	Migrant	97	--	--	--	--
	SES	29877	0	0	0	0
	Accommodations	6577	2	1	0	0
	Disability	10297	0	0	0	0
<b>4</b>	Female	32019	0	1	0	0
	Asian/Pacific Islander	1187	3	4	0	1
	Black	11789	2	0	0	0
	Hispanic	2377	2	0	0	0
	Native American/Alaskan	259	2	0	1	0
	ELL	2304	1	3	0	0
	IEP	9809	0	1	0	0
	Migrant	95	--	--	--	--
	SES	28703	0	0	0	0
	Accommodations	7321	0	1	0	0
	Disability	10009	0	1	0	0
<b>5</b>	Female	32049	2	4	0	0
	Asian/Pacific Islander	1185	1	2	0	1
	Black	11527	1	3	0	0
	Hispanic	2247	0	1	0	0
	Native American/Alaskan	267	0	1	0	0
	ELL	2081	0	1	0	0
	IEP	9557	0	0	0	0
	Migrant	109	1	2	1	2
	SES	28165	0	0	0	0
	Accommodations	7495	0	0	0	0
	Disability	9748	0	0	0	0
<b>6</b>	Female	32038	0	3	1	0
	Asian/Pacific Islander	1214	1	2	0	0
	Black	12013	0	1	0	1
	Hispanic	2285	0	1	0	0
	Native American/Alaskan	276	0	1	1	0
	ELL	1944	0	1	0	1
	IEP	9148	0	0	0	0
	Migrant	204	4	3	0	0
	SES	27256	0	0	0	0
	Accommodations	7491	1	0	0	0
	Disability	9296	0	0	0	0

**Table 8. 2: Number of Flagged Items for DIF using SMD Statistics, Mathematics (Continued)**

<b>Grade</b>	<b>Group</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>7</b>	Female	32666	1	1	0	0
	Asian/Pacific Islander	1060	2	2	1	0
	Black	12519	1	3	0	0
	Hispanic	2191	0	1	0	0
	Native American/Alaskan	310	0	0	0	0
	ELL	1802	0	3	0	0
	IEP	8883	1	2	0	0
	Migrant	196	0	0	1	1
	SES	26715	0	0	0	0
	Accommodations	7162	1	2	1	0
	Disability	9045	1	2	0	0
<b>8</b>	Female	34086	1	4	0	1
	Asian/Pacific Islander	1051	1	1	0	0
	Black	13156	1	2	0	0
	Hispanic	2147	0	0	0	0
	Native American/Alaskan	331	0	0	0	0
	ELL	1794	0	0	0	0
	IEP	9480	1	0	0	1
	Migrant	58	--	--	--	--
	SES	26769	0	0	0	0
	Accommodations	7556	1	1	1	1
	Disability	9554	1	0	0	1
<b>10</b>	Female	33902	3	2	0	2
	Asian/Pacific Islander	1049	4	2	0	1
	Black	11409	0	1	0	0
	Hispanic	1803	0	1	0	0
	Native American/Alaskan	323	0	0	0	0
	ELL	1743	0	0	0	1
	IEP	8441	0	0	0	0
	Migrant	66	--	--	--	--
	SES	21465	0	0	0	0
	Accommodations	6053	1	0	0	0
	Disability	8493	0	0	0	0

## **Part 9. Constructed-Response Scoring**

In this section, we first describe the scoring process used for MAP. In particular, we focus on the MAP handscoring process. At the end of this section, we describe and report the results of the inter-rater reliability study conducted on the handscoring of MAP constructed-response items.

### ***MAP Scoring Process***

Multiple-choice items were scored by CTB using electronic scanning equipment. Constructed-response items were scored by human raters who were trained by CTB.

### **Handscoring Process Used for MAP**

Evidence of validity is provided by the procedures described below for handscoring.

### **Selection of Scoring Evaluators**

CTB/McGraw-Hill and Kelly Services strive to develop a highly qualified, experienced core of evaluators so that the integrity of all projects is appropriately maintained.

#### *Recruitment*

The MAP 2007 project was staffed with a large number of returning evaluators and team leaders who had previous experience with MAP and other handscoring projects. Kelly Services also recruited new team leaders and evaluators for employment. Recruitment sources included advertisements in newspapers in Indianapolis, Indiana, Mather, California, and nearby areas and Internet sources.

CTB requires that all evaluators and team leaders possess a bachelor's degree or higher. Kelly Services carefully screened all new applicants and required them to produce either a transcript or a copy of the degree. Kelly Services also required a one- to two-hour interview/screening process. Individuals who did not present proper documentation or had less than desirable work records were eliminated during this process. Kelly Services verified that 100% of all potential evaluators met the degree requirement. All experienced evaluators and team leaders had already successfully completed the screening process.

#### *The Interview Process*

All potential evaluators completed a pre-interview activity. For some parts of the pre-interview activity, applicants were shown examples of test responses and were supplied with a scoring guide. In a brief introduction, they became acquainted with the application of a rubric. After the introduction, applicants applied the scoring guide to score the sample responses. The applicant's scores were used for discussion during the interview process to determine the applicant's trainability as well as his/her ability to understand and implement the standards set forth in the sample scoring guide.

Kelly Services interviewed each applicant and determined the applicant's suitability for a specific content area and grade level. Applicants with strong leadership skills were questioned further to determine whether they were qualified to be team leaders.

When Kelly Services felt applicants were qualified, the applicants were recommended for employment. All assignments were made according to availability and suitability. Before being hired, all employees were required to read, agree to and sign a non-disclosure agreement outlining the CTB/McGraw-Hill business ethics and security procedures.

## **Handscoring Process**

### *Training Material Development*

All materials necessary for scoring were developed by CTB. These materials include the scoring guides and training papers used to complete the handscoring of constructed-response and extended-response items (writing essays and performance events).

Missouri operational items have been previously field tested. Prior to actual scoring, handscoring supervisors assembled materials based on the rubrics. Student answer documents were randomly sampled to ensure that a representative sample of possible responses was used. Supervisors selected anchor papers and training papers and recommended clarifications to rubrics. All materials were presented during the Training Material Review Meeting (TMRM) and scores and annotations were approved by DESE-invited participants.

From this point, training and qualifying materials were developed based on the rubric and scoring philosophies discussed during the TMRM.

### *Training Material Review Meeting*

CTB met with DESE and participants to present all anchors, scoring guides, and a sample of student responses. These materials were scored and annotated based on CTB's interpretation of the rubric and philosophies discussed during previous TMRMs. Each response, score, and annotation was reviewed and updated as needed within the outlined limitations.

### *Training and Qualifying Procedures*

Handscoring involves training and qualifying team leaders and evaluators, monitoring scoring accuracy and production, and ensuring security of both the test materials and the scoring facilities. An explanation of the training and qualification procedures follows.

All readers were trained and qualified in specific Rater Item Blocks (RIBs) consisting of one item to be scored. Evaluators and team leaders were trained using the following steps:

- Reviewing the student response booklet
- Reviewing rubrics

- Reviewing anchor papers
- Explaining scoring strategies, followed by a question-and-answer period
- Scoring a training set, followed by sharing established scores, discussing responses, and answering questions arising from scores
- Scoring and discussing additional training sets
- Administration of Qualifying Round 1
- Administration of Qualifying Round 2 (if necessary)
- Explaining condition codes and sensitive paper procedures
- Explaining non-standard response or computer-generated response (nsr/cgr) procedures
- Explaining un-scannable image procedures

All evaluators were trained and qualified using the same procedures and criteria used for the team leaders.

#### *Team Leader Training (TLT)*

DESE and participants joined CTB team leaders during training. During this time, regional facilitators and team leaders were trained and qualified. These participants served as trainers and team leaders at each of the DHS sites. Once qualification was completed, checksets were presented for approval.

### **Monitoring the Scoring Process**

#### *Daily Accuracy Checks*

Throughout the course of handscoring, calibration sets of pre-scored papers (checksets/validity sets) were administered daily to each scorer to monitor scoring accuracy and to maintain a consistent focus on the established rubrics and guidelines. Checksets were executed via imaging software that provided images in such a way that the reader did not know when a checkset was administered. All checkset scores had been approved by DESE participants during and immediately following the TMRM.

In addition to the checkset process, CTB's handscoring protocol included the use of read-behinds. The read-behind was another valuable rater-reliability monitoring technique that allowed a team leader to review a reader's scored documents, providing feedback and counseling as appropriate.

Approximately 10% of Communication Arts and Mathematics papers from the early-return districts were scored by a second reader to establish inter-rater reliability statistics for all constructed-response items. This procedure is called a "double-blind read" because the second reader does not know the first reader's score.

### *Recalibration of Raters*

Recalibration in handscoring refers to the process in which scorers/raters who begin to drift away from scoring accuracy are realigned to correct scoring.

### **Security**

Security guards were on site whenever employees were present in the building. All employees were issued photo identification badges and were required to wear them in plain view at all times. Visitors and employees who forgot their badges were issued visitors' badges and were required to wear them in plain view. All employees and visitors were subject to inspection of their personal effects.

### ***Inter-rater Reliability***

In the early-return districts, approximately 10% of papers in Communication Arts and Mathematics were scored independently by a second reader. To determine the reliability of scoring, the percent of perfect agreement and adjacent agreement between the two readers was calculated. A weighted kappa was calculated to reflect the level of improvement beyond the chance level in the consistency of scoring. Item-level rater agreement rates and weighted Kappas are presented in Tables 9.1 and 9.2. All Communication Arts and Mathematics items show good inter-rater agreement. As shown in Table 9.1, raters demonstrated at least 90% perfect and adjacent agreement for all but one Communication Arts item. Grade 7 Item 20 had less than 90% perfect and adjacent agreement, and the percent adjacent and perfect agreement for this item was 89%. The Kappa statistic for the Communication Arts items ranged from .56 (Grade 5, Item 18) to .98 (Grade 4, Item 43). As shown in Table 9.2, raters demonstrated above 98% adjacent agreement for all Mathematics items. The Kappa statistic for the Mathematics items ranged from .84 (Grade 3, Item 32; Grade 8, Item 22) to .99 (Grade 3, Item 31; Grade 6, Item 14).

**Table 9. 1: Inter-rater Reliability, Communication Arts**

<b>Grade</b>	<b>Item #</b>	<b>Session</b>	<b># Points</b>	<b>% Perfect</b>	<b>% Adjacent</b>	<b>% Perfect &amp; Adjacent</b>	<b>Kappa</b>
<b>3</b>	3	1	2	79.83	19.90	99.73	0.73
	4	1	2	85.58	14.29	99.87	0.85
	5	1	2	85.82	14.01	99.83	0.84
	6A	1	2	87.21	12.58	99.79	0.88
	6B	1	2	89.99	9.81	99.80	0.85
	1	2	4	64.82	33.77	98.59	0.81
	22	3	3	85.82	14.06	99.88	0.90
	23	3	2	87.37	11.91	99.28	0.88
	24	3	1	95.91	4.09	100.00	0.86
	43	3	2	92.19	6.97	99.16	0.93
<b>4</b>	3	1	2	84.15	14.90	99.05	0.87
	4	1	2	87.12	12.02	99.14	0.88
	5	1	2	83.94	15.23	99.17	0.85
	6A	1	2	83.93	15.38	99.31	0.86
	6B	1	2	91.19	8.42	99.61	0.87
	19	2	2	96.09	2.98	99.07	0.94
	20	2	3	90.52	8.69	99.21	0.93
	21	2	2	85.13	14.13	99.26	0.88
	43	2	2	98.41	1.46	99.87	0.98
<b>5</b>	3	1	2	70.70	27.68	98.38	0.71
	4A	1	2	75.90	18.78	94.68	0.78
	4B	1	2	95.55	4.16	99.71	0.93
	5	1	2	76.54	20.08	96.62	0.76
	6	1	2	72.01	25.79	97.80	0.66
	17	2	2	95.30	4.55	99.85	0.95
	18	2	2	64.07	33.53	97.60	0.56
	42	2	3	81.40	16.80	98.20	0.90
	43	2	2	70.53	28.42	98.95	0.68
<b>6</b>	3	1	2	87.65	11.96	99.61	0.88
	4	1	2	82.51	17.26	99.77	0.86
	5A	1	2	78.83	20.13	98.96	0.83
	5B	1	1	98.10	1.90	100.00	0.96
	6	1	2	82.65	16.73	99.38	0.84
	17	2	2	81.26	17.92	99.18	0.79
	18	2	2	85.53	14.39	99.92	0.87
	19	2	2	72.50	25.94	98.44	0.64
	43	2	3	63.52	32.21	95.73	0.75
<b>7</b>	3	1	2	94.00	5.76	99.76	0.95
	4A	1	2	78.65	20.74	99.39	0.81
	4B	1	1	98.63	1.37	100.00	0.96

**Table 9. 1: Inter-rater Reliability, Communication Arts (Continued)**

<b>Grade</b>	<b>Item #</b>	<b>Session</b>	<b># Points</b>	<b>% Perfect</b>	<b>% Adjacent</b>	<b>% Perfect &amp; Adjacent</b>	<b>Kappa</b>
<b>7</b>	5A	1	2	87.92	11.23	99.15	0.86
	5B	1	2	94.22	5.56	99.78	0.91
	6	1	2	84.88	14.09	98.97	0.81
	1	2	4	68.82	30.74	99.56	0.78
	19	3	2	83.41	15.77	99.18	0.86
	20	3	3	69.34	19.90	89.24	0.64
	42	3	2	83.48	15.94	99.42	0.85
<b>8</b>	3	1	2	95.02	0.51	95.53	0.90
	4	1	2	68.49	28.48	96.97	0.72
	5	1	2	72.04	23.27	95.31	0.72
	6A	1	2	79.45	18.62	98.07	0.83
	6B	1	1	95.50	4.50	100.00	0.92
	20	2	2	75.82	16.63	92.45	0.74
	21	2	3	79.32	18.79	98.11	0.87
	42	2	2	85.69	13.85	99.54	0.90
	43	2	2	86.42	13.26	99.68	0.91
<b>11</b>	3	1	2	74.58	23.86	98.44	0.77
	4	1	2	76.58	22.45	99.03	0.76
	5	1	2	87.91	11.92	99.83	0.89
	6A	1	2	65.78	32.12	97.90	0.70
	6B	1	2	90.70	8.86	99.56	0.92
	1	2	4	64.37	34.52	98.89	0.81
	21	3	3	82.96	16.87	99.83	0.90
	22	3	2	70.11	22.38	92.49	0.71
	42	3	3	62.03	29.49	91.52	0.75

**Table 9. 2: Inter-rater Reliability, Mathematics**

Grade	Item #	Session	# Points	% Perfect	% Adjacent	% Perfect & Adjacent	Kappa
3	8	1	2	96.46	2.88	99.34	0.97
	15	1	2	95.07	4.83	99.90	0.97
	24	1	2	95.26	4.56	99.82	0.94
	27	1	2	92.95	6.61	99.56	0.94
	31	2	2	98.26	1.46	99.72	0.99
	32	2	2	87.07	12.80	99.87	0.84
	33	2	2	95.19	4.60	99.79	0.97
4	5	1	2	96.27	3.65	99.92	0.96
	12	1	2	92.49	7.34	99.83	0.88
	15	1	2	88.89	10.93	99.82	0.92
	21	1	2	93.64	6.29	99.93	0.95
	25	1	2	93.89	6.06	99.95	0.96
	28	1	2	94.79	5.00	99.79	0.96
	30	1	4	85.99	12.48	98.47	0.94
	33	2	2	97.07	2.84	99.91	0.97
	34	2	2	93.31	6.40	99.71	0.94
	35	2	2	90.37	9.63	100.00	0.91
5	5	1	2	95.27	3.88	99.15	0.96
	8	1	2	96.62	3.33	99.95	0.95
	17	1	2	96.47	3.53	100.00	0.98
	24	1	2	96.45	3.45	99.90	0.98
	33	2	3	94.00	5.69	99.69	0.94
	34	2	2	95.53	4.41	99.94	0.95
	35	2	3	94.84	4.92	99.76	0.98
6	7	1	2	85.26	14.66	99.92	0.90
	14	1	2	99.82	0.18	100.00	0.99
	19	1	2	92.26	6.49	98.75	0.92
	24	1	2	88.23	11.47	99.70	0.93
	32	2	3	96.64	3.22	99.86	0.98
	33	2	2	97.05	2.91	99.96	0.97
	34	2	2	94.77	5.12	99.89	0.95
7	7	1	2	96.90	2.67	99.57	0.98
	14	1	2	96.94	2.83	99.77	0.98
	21	1	2	97.95	1.44	99.39	0.98
	27	1	2	94.61	5.24	99.85	0.97
	33	2	2	93.73	6.14	99.87	0.95
	34	2	3	89.17	8.93	98.10	0.95
	35	2	3	94.49	4.04	98.53	0.97
8	4	1	2	94.04	5.87	99.91	0.94

	7	1	2	96.95	2.66	99.61	0.98
--	---	---	---	-------	------	-------	------

**Table 9. 2: Inter-rater Reliability, Mathematics (Continued)**

<b>Grade</b>	<b>Item #</b>	<b>Session</b>	<b># Points</b>	<b>% Perfect</b>	<b>% Adjacent</b>	<b>% Perfect &amp; Adjacent</b>	<b>Kappa</b>
<b>8</b>	12	1	2	96.30	3.50	99.80	0.97
	17	1	2	93.20	6.30	99.50	0.96
	22	1	2	81.83	18.11	99.94	0.84
	27	1	2	97.51	2.18	99.69	0.98
	30	1	4	84.31	14.42	98.73	0.96
	32	2	2	93.04	6.58	99.62	0.96
	33	2	2	84.93	14.31	99.24	0.88
	34	2	2	89.20	10.42	99.62	0.91
<b>10</b>	5	1	2	96.00	3.63	99.63	0.97
	10	1	2	91.29	8.66	99.95	0.93
	15	1	2	96.30	3.48	99.78	0.98
	20	1	2	91.90	7.96	99.86	0.92
	25	1	2	93.99	5.71	99.70	0.96
	29	1	2	92.63	7.13	99.76	0.94
	32	1	4	79.89	18.88	98.77	0.89
	26	2	2	95.55	4.16	99.71	0.96
	27	2	2	92.69	6.90	99.59	0.96
	28	2	3	88.27	11.50	99.77	0.93
	29	2	2	96.80	2.91	99.71	0.98

## References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association, Inc.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, *46*, 443–459.
- Burket, G. R. (1995). PARDUX [Computer program]. Monterey, CA: CTB/McGraw-Hill.
- Camilli, G, & Shepard, A. L. (1994). *Methods for Identifying Biased Test Items*. Sage Publication.
- Cattell, R.B. (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, *1*, 245 -276.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. *16*, 297-334.
- CTB/McGraw-Hill (2005). *Missouri Assessment Program Final Bookmark Standard Setting Technical Report*. Monterey, CA: Author.
- CTB/McGraw-Hill (2006). *Missouri Assessment Program Technical Report 2006*. Monterey, CA: Author.
- CTB/McGraw-Hill (2007). *Missouri Assessment Program Examiner's Manual*. Monterey, CA: Author.
- Dorans, N.J., & Schmitt, M.P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. Princeton: Educational Testing Service.
- 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. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Macmillan/McGraw-Hill. (1993a). *Guidelines for Bias-Free Publish*. New York: Author.
- Macmillan/McGraw-Hill. (1993b). *Reflecting diversity: Multicultural guidelines for educational publishing professionals*. New York: Author.
- 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.
- Spearman C. (1904) The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman C. (1910) Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201–210.
- Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175–186.
- Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5, 245–262.
- Yen, W. M. (1993) . Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233-251.