

# Missouri Assessment Program Technical Report 2006

November 27, 2006



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

Published by CTB/McGraw-Hill, a Division of the Educational and Professional Publishing Group of The McGraw-Hill Companies, Inc.,  
20 Ryan Ranch Road, Monterey, CA 93940-5703. Copyright© 2006 by CTB/McGraw-Hill. All rights reserved.

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 © 2006 by CTB/McGraw-Hill LLC. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of CTB/McGraw-Hill. *TerraNova* is a registered trademark, and Bookmark Standard Setting Procedure is a trademark, of The McGraw-Hill Companies, Inc.

# Table of Contents

<b>PART 1: OVERVIEW .....</b>	<b>1</b>
<b>PART 2: VALIDITY AND THE USES OF TEST SCORES.....</b>	<b>2</b>
USES OF TEST SCORES .....	2
TEST-LEVEL SCORES .....	2
<i>Use of Test-Level Scores</i> .....	3
CONTENT STANDARD SUBSCORES .....	4
<i>The Use of the Content Standard Subscores</i> .....	4
PROCESS STANDARD SUBSCORES .....	5
<i>The Use of the Process Standard Subscores</i> .....	5
<b>PART 3: TEST CONTENT DEVELOPMENT .....</b>	<b>6</b>
TEST DESIGN .....	6
MATHEMATICS ITEM DEVELOPMENT.....	6
ALIGNMENT TO GRADE-LEVEL EXPECTATIONS.....	9
COMMUNICATION ARTS ITEM DEVELOPMENT.....	9
<b>PART 4: RELIABILITY AND CONSTRUCT-RELATED VALIDITY.....</b>	<b>35</b>
MINIMIZATION OF CONSTRUCT-IRRELEVANT VARIANCE AND UNDER-REPRESENTATION.....	35
RELIABILITY .....	35
<i>Test Reliability</i> .....	36
<i>Standard Error of Measurement</i> .....	36
<i>Decision Accuracy and Consistency</i> .....	37
CONVERGENT VALIDITY.....	38
<i>Principal Components Analysis</i> .....	38
DIVERGENT (DISCRIMINANT) VALIDITY .....	39
ANALYSES BY CONTENT STANDARD .....	40
<i>Reliability of Content Standards</i> .....	40
<i>Correlations among Content Standard Subscores</i> .....	40
<b>PART 5: SCALING AND LINKING.....</b>	<b>64</b>
SCALING .....	64
ITEM RESPONSE THEORY .....	64
MODEL FIT .....	65
<i>Linking Method and Results</i> .....	69
LOWEST AND HIGHEST OBTAINABLE SCALE SCORES .....	70
<b>PART 6: TEST AND ITEM STATISTICS .....</b>	<b>90</b>
TEST-LEVEL STATISTICS .....	90
ITEM-LEVEL STATISTICS .....	90
<b>PART 7: FAIRNESS .....</b>	<b>106</b>
MINIMIZING BIAS THROUGH CAREFUL TEST DEVELOPMENT .....	106
EVALUATING BIAS THROUGH DIFFERENTIAL ITEM FUNCTIONING STATISTICS.....	106
EVALUATING BIAS THROUGH IMPACT ANALYSIS .....	108
<b>PART 8: ACHIEVEMENT-LEVEL STANDARD SETTING .....</b>	<b>129</b>
2005 MAP STANDARD SETTING .....	129
2006 DATA REVIEW .....	129
<b>PART 9: CONSTRUCTED-RESPONSE SCORING.....</b>	<b>132</b>

MAP SCORING PROCESS .....	132
<i>Hand-Scoring Process Used for MAP</i> .....	132
INTER-RATER RELIABILITY .....	135
<b>PART 10: STUDENT DEMOGRAPHICS</b> .....	<b>140</b>
CALIBRATION SAMPLE .....	140
CALIBRATION AND CENSUS DEMOGRAPHICS .....	140
STUDENT PARTICIPATION .....	141
TEST ACCOMMODATIONS .....	141
<b>REFERENCES</b> .....	<b>163</b>
<b>APPENDIX</b> .....	<b>1635</b>

## Table of Tables

TABLE 3. 1: CONTENT COVERAGE: MAP 2006 ITEM MAPS.....	12
TABLE 3. 2: MAP 2006 CONTENT STANDARD ITEM/POINT DISTRIBUTIONS.....	23
TABLE 3. 3: MAP 2006 NUMBER OF ITEMS/POINTS MEASURING PROCESS STANDARDS.....	30
TABLE 4. 1: RELIABILITY IN COMMUNICATION ARTS.....	40
TABLE 4. 2: RELIABILITY IN MATHEMATICS.....	41
TABLE 4. 3: CRONBACH’S ALPHA BY SUBGROUP, COMMUNICATION ARTS 2006.....	41
TABLE 4. 4: CRONBACH’S ALPHA BY SUBGROUP, MATHEMATICS 2006.....	42
TABLE 4. 5: DECISION ACCURACY AND CONSISTENCY CONDITIONED ON LEVEL OF ACHIEVEMENT.....	42
TABLE 4. 6: DECISION ACCURACY AND CONSISTENCY AT ACHIEVEMENT CUT POINTS.....	43
TABLE 4. 7: PRINCIPAL COMPONENT ANALYSIS FOR COMMUNICATION ARTS.....	44
TABLE 4. 8: PRINCIPAL COMPONENT ANALYSIS FOR MATHEMATICS.....	45
TABLE 4. 9 INTER-CORRELATION AMONG CONTENT AREA SCALE SCORES.....	46
TABLE 4. 10: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 3.....	46
TABLE 4. 11: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 4.....	46
TABLE 4. 12: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 5.....	47
TABLE 4. 13: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 6.....	47
TABLE 4. 14: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 7.....	47
TABLE 4. 15: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 8.....	47
TABLE 4. 16: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: COMMUNICATION ARTS GRADE 11.....	48
TABLE 4. 17: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 3.....	48
TABLE 4. 18: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 4.....	48
TABLE 4. 19: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 5.....	48
TABLE 4. 20: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 6.....	49
TABLE 4. 21: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 7.....	49
TABLE 4. 22: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 8.....	49
TABLE 4. 23: RELIABILITY OF EACH CONTENT STANDARD AND CORRELATION AMONG CONTENT STANDARDS: MATHEMATICS GRADE 10.....	49
TABLE 5. 1: ITEM FIT STATISTICS FOR MISFITTING ITEMS.....	67
TABLE 5. 2: LOSS AND HOSS VALUES BY GRADE AND CONTENT AREA.....	70
TABLE 6. 1: MAP MEANS, STANDARD DEVIATIONS FOR SCALE/RAW SCORES, P-VALUES, ITEM-TOTAL CORRELATION: COMMUNICATION ARTS 2006.....	91
TABLE 6. 2: MAP MEANS, STANDARD DEVIATIONS FOR SCALE/RAW SCORES, P-VALUES, ITEM-TOTAL CORRELATION: MATHEMATICS 2006.....	91
TABLE 6. 3: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 3.....	92
TABLE 6. 4: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 4.....	93

TABLE 6. 5: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 5 .....	94
TABLE 6. 6: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 6 .....	95
TABLE 6. 7: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 7 .....	96
TABLE 6. 8: ITEM STATISTICS: COMMUNICATION ARTS, GRADE 8 .....	97
TABLE 6. 9: ITEM STATISTICS: COMMUNICATION ARTS GRADE 11 .....	98
TABLE 6. 10: ITEM STATISTICS: MATHEMATICS, GRADE 3 .....	99
TABLE 6. 11: ITEM STATISTICS: MATHEMATICS, GRADE 4 .....	100
TABLE 6. 12: ITEM STATISTICS: MATHEMATICS, GRADE 5 .....	101
TABLE 6. 13: ITEM STATISTICS: MATHEMATICS, GRADE 6 .....	102
TABLE 6. 14: ITEM STATISTICS: MATHEMATICS, GRADE 7 .....	103
TABLE 6. 15: ITEM STATISTICS: MATHEMATICS, GRADE 8 .....	104
TABLE 6. 16: ITEM STATISTICS: MATHEMATICS, GRADE 10 .....	105
TABLE 7. 1: 2006 MAP SMD STATISTICS: NUMBER OF FLAGGED ITEMS, COMMUNICATION ARTS .....	109
TABLE 7. 2: 2006 MAP SMD STATISTICS: NUMBER OF FLAGGED ITEMS, MATHEMATICS .....	112
TABLE 7. 3: IMPACT ANALYSIS, GRADE 03 COMMUNICATION ARTS .....	115
TABLE 7. 4: IMPACT ANALYSIS, GRADE 04 COMMUNICATION ARTS .....	116
TABLE 7. 5: IMPACT ANALYSIS, GRADE 05 COMMUNICATION ARTS .....	117
TABLE 7. 6: IMPACT ANALYSIS, GRADE 06 COMMUNICATION ARTS .....	118
TABLE 7. 7: IMPACT ANALYSIS, GRADE 07 COMMUNICATION ARTS .....	119
TABLE 7. 8: IMPACT ANALYSIS, GRADE 08 COMMUNICATION ARTS .....	120
TABLE 7. 9: IMPACT ANALYSIS, GRADE 11 COMMUNICATION ARTS .....	121
TABLE 7. 10: IMPACT ANALYSIS, GRADE 03 MATHEMATICS .....	122
TABLE 7. 11: IMPACT ANALYSIS, GRADE 04 MATHEMATICS .....	123
TABLE 7. 12: IMPACT ANALYSIS, GRADE 05 MATHEMATICS .....	124
TABLE 7. 13: IMPACT ANALYSIS, GRADE 06 MATHEMATICS .....	125
TABLE 7. 14: IMPACT ANALYSIS, GRADE 07 MATHEMATICS .....	126
TABLE 7. 15: IMPACT ANALYSIS, GRADE 08 MATHEMATICS .....	127
TABLE 7. 16: IMPACT ANALYSIS, GRADE 10 MATHEMATICS .....	128
TABLE 8. 1: RECOMMENDED CUT SCORES AND IMPACT DATA FROM 2005 MAP STANDARD SETTING .....	130
TABLE 8. 2: CUT SCORES FROM THE PHASE I 2006 DATA REVIEW .....	131
TABLE 8. 3: FINAL CUT SCORES FROM THE PHASE II 2006 DATA REVIEW .....	131
TABLE 9. 1: INTER-RATER RELIABILITY, COMMUNICATION ARTS .....	136
TABLE 9. 2: INTER-RATER RELIABILITY, MATHEMATICS .....	138
TABLE 10. 1: NUMBER OF SCHOOLS INVITED TO PARTICIPATE IN THE CALIBRATION SAMPLE .....	140
TABLE 10. 2: SUMMARY OF CALIBRATION AND CENSUS DATA: COMMUNICATION ARTS .....	143
TABLE 10. 3: SUMMARY OF CALIBRATION AND CENSUS DATA: MATHEMATICS .....	146
TABLE 10. 4: PARTICIPATION RATES: ALL STUDENTS .....	149
TABLE 10. 5: PARTICIPATION RATES: MALES .....	149
TABLE 10. 6: PARTICIPATION RATES: FEMALES .....	149
TABLE 10. 7: PARTICIPATION RATES: WHITE .....	150
TABLE 10. 8: PARTICIPATION RATES: BLACK .....	150
TABLE 10. 9: PARTICIPATION RATES: HISPANIC .....	150
TABLE 10. 10: PARTICIPATION RATES: ASIAN/PACIFIC ISLANDER .....	151
TABLE 10. 11: PARTICIPATION RATES: NATIVE AMERICAN/ALASKAN .....	151
TABLE 10. 12: PARTICIPATION RATES: STUDENTS RECEIVING FREE OR REDUCED-PRICE LUNCH (SES) .....	151
TABLE 10. 13: PARTICIPATION RATES: MIGRANT STUDENTS .....	152
TABLE 10. 14: PARTICIPATION RATES: SPECIAL EDUCATION (IEP) STUDENTS .....	152
TABLE 10. 15: PARTICIPATION RATES: ENGLISH LEARNERS (LEP, ELL) .....	152
TABLE 10. 16: PARTICIPATION RATES: STUDENTS WITH DISABILITIES .....	153
TABLE 10. 17: PARTICIPATION RATES: STUDENTS RECEIVING ACCOMMODATIONS .....	153

TABLE 10. 18: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, MAP 2006 .....	154
TABLE 10. 19: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, BRAILLE-EDITION MAP 2006 .....	158
TABLE 10. 20: NUMBER AND PERCENTAGE OF STUDENTS RECEIVING ACCOMMODATIONS BY ACCOMMODATION TYPE, LARGE-PRINT EDITION MAP 2006 .....	160

## Table of Figures

FIGURE 4. 1: SEM PLOT COMMUNICATION ARTS GRADE 3.....	50
FIGURE 4. 2: SEM PLOT COMMUNICATION ARTS GRADE 4.....	50
FIGURE 4. 3: SEM PLOT COMMUNICATION ARTS GRADE 5.....	51
FIGURE 4. 4: SEM PLOT COMMUNICATION ARTS GRADE 6.....	51
FIGURE 4. 5: SEM PLOT COMMUNICATION ARTS GRADE 7.....	52
FIGURE 4. 6: SEM PLOT COMMUNICATION ARTS GRADE 8.....	52
FIGURE 4. 7: SEM PLOT COMMUNICATION ARTS GRADE 11.....	53
FIGURE 4. 8: SEM PLOT MATHEMATICS GRADE 3 .....	53
FIGURE 4. 9: SEM PLOT MATHEMATICS GRADE 4 .....	54
FIGURE 4. 10: SEM PLOT MATHEMATICS GRADE 5 .....	54
FIGURE 4. 11: SEM PLOT MATHEMATICS GRADE 6 .....	55
FIGURE 4. 12: SEM PLOT MATHEMATICS GRADE 7 .....	55
FIGURE 4. 13: SEM PLOT MATHEMATICS GRADE 8 .....	56
FIGURE 4. 14: SEM PLOT MATHEMATICS GRADE 10 .....	56
FIGURE 4. 15: SCREE PLOT COMMUNICATION ARTS GRADE 3.....	57
FIGURE 4. 16: SCREE PLOT COMMUNICATION ARTS GRADE 4.....	57
FIGURE 4. 17: SCREE PLOT COMMUNICATION ARTS GRADE 5.....	58
FIGURE 4. 18: SCREE PLOT COMMUNICATION ARTS GRADE 6.....	58
FIGURE 4. 19: SCREE PLOT COMMUNICATION ARTS GRADE 7.....	59
FIGURE 4. 20: SCREE PLOT COMMUNICATION ARTS GRADE 8.....	59
FIGURE 4. 21: SCREE PLOT COMMUNICATION ARTS GRADE 11.....	60
FIGURE 4. 22: SCREE PLOT MATHEMATICS GRADE 3 .....	60
FIGURE 4. 23: SCREE PLOT MATHEMATICS GRADE 4 .....	61
FIGURE 4. 24: SCREE PLOT MATHEMATICS GRADE 5 .....	61
FIGURE 4. 25: SCREE PLOT MATHEMATICS GRADE 6 .....	62
FIGURE 4. 26: SCREE PLOT MATHEMATICS GRADE 7 .....	62
FIGURE 4. 27: SCREE PLOT MATHEMATICS GRADE 8 .....	63
FIGURE 4. 28: SCREE PLOT MATHEMATICS GRADE 10 .....	63
FIGURE 5. 1. ITEM CHARACTERISTIC CURVE FOR GRADE 3 COMMUNICATION ARTS, ITEM 3 ( $Z = 11.00$ ) .....	71
FIGURE 5. 2. ITEM CHARACTERISTIC CURVE FOR GRADE 3 COMMUNICATION ARTS, ITEM 26 ( $Z = 20.03$ ) .....	72
FIGURE 5. 3. ITEM CHARACTERISTIC CURVE FOR GRADE 3 COMMUNICATION ARTS, ITEM 43 ( $Z = 12.78$ ).....	72
FIGURE 5. 4. ITEM CHARACTERISTIC CURVE FOR GRADE 4 COMMUNICATION ARTS, ITEM 19 ( $Z = 19.46$ ).....	73
FIGURE 5. 5. ITEM CHARACTERISTIC CURVE FOR GRADE 5 COMMUNICATION ARTS, ITEM 9 ( $Z = 11.29$ ) .....	74
FIGURE 5. 6. ITEM CHARACTERISTIC CURVE FOR GRADE 11 COMMUNICATION ARTS, ITEM 5A ( $Z = 19.03$ ).....	74
FIGURE 5. 7. ITEM CHARACTERISTIC CURVE FOR GRADE 11 COMMUNICATION ARTS, ITEM 42 ( $Z = 11.61$ ) .....	75
FIGURE 5. 8. ITEM CHARACTERISTIC CURVE FOR GRADE 4 MATHEMATICS, ITEM 14 ( $Z = 17.38$ ) .....	76
FIGURE 5. 9. ITEM CHARACTERISTIC CURVE FOR GRADE 5 MATHEMATICS, ITEM 17 ( $Z = 11.40$ ) .....	76
FIGURE 5. 10. ITEM CHARACTERISTIC CURVE FOR GRADE 5 MATHEMATICS, ITEM 35 ( $Z = 14.28$ ) .....	77
FIGURE 5. 11. ITEM CHARACTERISTIC CURVE FOR GRADE 7 MATHEMATICS, ITEM 35 ( $Z = 12.93$ ) .....	78
FIGURE 5. 12. ITEM CHARACTERISTIC CURVE FOR GRADE 10 MATHEMATICS, ITEM 20 ( $Z = 12.83$ ) .....	79
FIGURE 5. 13. ITEM CHARACTERISTIC CURVE FOR GRADE 10 MATHEMATICS, ITEM 26 ( $Z = 10.78$ ) .....	80
FIGURE 5. 14. COMMUNICATION ARTS, GRADE 3 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	81
FIGURE 5. 15. COMMUNICATION ARTS, GRADE 4 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	81
FIGURE 5. 16. COMMUNICATION ARTS, GRADE 5 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	82
FIGURE 5. 17. COMMUNICATION ARTS, GRADE 6 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	82
FIGURE 5. 18. COMMUNICATION ARTS, GRADE 7 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	83

FIGURE 5. 19. COMMUNICATION ARTS, GRADE 8 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	83
FIGURE 5. 20. COMMUNICATION ARTS, GRADE 11 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	84
FIGURE 5. 21. MATHEMATICS, GRADE 3 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	84
FIGURE 5. 22. MATHEMATICS, GRADE 4 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	85
FIGURE 5. 23. MATHEMATICS, GRADE 5 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	85
FIGURE 5. 24. MATHEMATICS, GRADE 6 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	86
FIGURE 5. 25. MATHEMATICS, GRADE 7 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	86
FIGURE 5. 26. MATHEMATICS, GRADE 8 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	87
FIGURE 5. 27. MATHEMATICS, GRADE 10 TEST CHARACTERISTIC CURVES (TCC) FOR THE WHOLE TEST, FOR THE INPUTTED ANCHOR ITEMS, AND FOR THE ESTIMATED ANCHOR ITEMS .....	87
FIGURE 5. 28. SELECTED SCALE SCORES PLOTTED BY GRADE LEVEL, MATHEMATICS MAP 2006 .....	88
FIGURE 5. 29. SELECTED SCALE SCORES PLOTTED BY GRADE LEVEL, COMMUNICATION ARTS MAP .....	88
FIGURE 5. 30. 2006 COMMUNICATION ARTS TEST CHARACTERISTIC CURVES (TCC) BY GRADE .....	89
FIGURE 5. 31. 2006 MATHEMATICS TEST CHARACTERISTIC CURVES (TCC) BY GRADE.....	89

## Part 1: Overview

The 2006 Missouri Assessment Program (MAP) is designed to measure students' knowledge of Communication Arts and Mathematics.<sup>1</sup> This administration of the Communication Arts and Mathematics MAP marks the beginning of a new phase of MAP in which all grade levels are tested.

Historically, 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 operational test forms. In the spring of 2006, Missouri administered assessments in Communication Arts to students in Grades 3–8 and 11, and Mathematics to students in Grades 3–8 and 10.

This report provides a technical overview of the Communication Arts and Mathematics assessments of the 2006 Missouri Assessment Program. As such, it presents evidence for the validity of the 2006 MAP scores. This first section, Part 1, provides a brief historical overview of the Communication Arts and Mathematics assessments of the MAP. Part 2 discusses the concept of validity. Validity hinges on the uses of test scores. In this section, we focus on the uses of 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 discusses reliability and construct-related validity. In this chapter, the assumption is tested that the content-area MAPs measure only one construct. For example, the grade-level Mathematics MAP should measure one primary dimension (Mathematics). Part 5 presents the scaling and linking procedures, as well as the results of these procedures. Part 6 reviews the test- and item-level characteristics. Part 7 contains an overview of the statistical and developmental processes used to ensure fairness of the MAP for all examinees. Part 8 reviews the results of the achievement-level setting. Part 9 discusses the scoring of constructed-response items, as well as the results of the inter-rater reliability studies. Finally, Part 10 reports the demographic characteristics of the calibration (i.e., early-return) sample and census data. Some analyses in this document are based on the calibration sample while other are based on census data. The type of data used for particular analyses is indicated throughout the Technical 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.

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

Validity is one of the most important components of any testing program. The information provided in this Technical Report constitutes *validity evidence* that supports the uses of the test scores. The following text 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. In gathering evidence related to the validity of the test scores, our first step should be to examine the uses of the test scores. In this section, we examine some possible uses of the MAP test scores.

The following sections (Parts 3 – 10) of this Technical Report provide validity evidence for these uses, as well as technical support for some of the interpretations and uses of test scores. The information in Parts 3 – 10 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 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 ability 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 ability each shows in each content area. Classification is based on the level of student ability 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, in varying ways, indicate a student's ability 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 MAP scores are credible individual ability 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 ability 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 on the MAP is determined for each content area. The overall scale score for a content area is a quantification of the ability being measured by either the Communication Arts or Mathematics test. The psychometric approach to the definition of ability is operational; in other words, the test score defines the ability.

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 ability 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 on the Communication Arts or Mathematics MAP is reported in one of four levels of achievement: *Below Basic*, *Basic*, *Proficient*, or *Advanced*. These levels are based on the student's scale score. The cut scores for the levels of achievement were first recommended by Missouri teachers at the Bookmark Standard Setting Workshop in December 2005 and reflect the teachers' 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. Thus, MAP achievement levels reflect the achievement standards and abilities intended by the Missouri legislature, Missouri teachers, and DESE. See Part 8 for an explanation of the achievement-level setting process.

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

MAP scale scores and achievement levels provide summary evidence of student ability in Communication Arts or Mathematics. Classroom teachers may use these scores as evidence of student ability 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 curricula to the Show-Me Standards/GLE strands. The results presented in this Technical Report provide evidence that the scale scores are a valid and reliable indicator of student performance in both Communication Arts and Mathematics.

The achievement levels reflect the expectations of Missouri educators and citizens. They were derived from a successful standard setting workshop in which Missouri educators and citizens specified their expectations for students in each achievement level. 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. Part 8 of this Technical Report and the *Bookmark Standard Setting Technical Report for the Missouri Assessment Program* (2005) provide evidence for the validity of the achievement levels.

### ***Content Standard Subscores***

The Content Standard subscores indicate student performance in terms of the number- and percent-correct scores for each Content Standard in Communication Arts and each GLE strand in Mathematics. Content Standard subscores are reported at four reporting levels: the state, the school district, the school, and the student. When aggregated over all the students in the school, district, or state, the means for Content Standard subscores become dependable. These means are intended as indicators of the performance of the school or district in teaching students the knowledge, skills, and abilities defined for each subject area. On the Content Standards Summary Report for schools and districts, standard errors of the means are reported using a 95% confidence interval.

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

The purpose of reporting Content Standard subscores on MAP is to show that students who exhibit the overall ability being measured have 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 ability 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. District and school administrators may compare their aggregate results with the state mean to better understand their strengths and weaknesses within a content area. State-level administrators may use aggregate results to ascertain the extent to which Missouri schools and districts have aligned their curricula to the Show-Me Standards/GLE strands. 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.

Caution should be exercised when comparing Content Standard 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 ability 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 ability 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 Standard 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 students with 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 to 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 for 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 in the 2006 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 has to do with 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.

### *Mathematics Item Development*

Planning and preparation for the development of item content to be used on the 2006 MAP Operational Test forms commenced in late 2002. The plan specified an item development and selection cycle that included an initial item writing 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) at Grade levels 3–8 and 10. The entire development process, which involved Missouri educators at each step along the way, is described in greater detail below.

In May of 2003, a group composed of Missouri educators, Regional Facilitators, DESE staff, and CTB personnel participated in a Mathematics item writing workshop (IWW) 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 test forms. (This and other key item development workshops were held in conjunction with the Communication Arts item development team and its participants.) The workshop was conducted with participants selected by DESE to represent educational sites throughout Missouri, and the purpose of the workshop was to write 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. Overall, the item writing workshop in May 2003 provided a basis upon which items written for the Mathematics assessment could be selected for use on small-scale pilot studies administered throughout Missouri.

Mathematics items were refined after the initial item writing workshop via collaboration between DESE and CTB. Some items from the workshop were considered to be unusable, so additional SR items were developed by CTB to help supplement the item pool (particularly selected-response items) and reviewed by DESE. This interaction led to the production of Mathematics test forms used for the local pilot studies. The small-scale pilot was administered in October/November of 2003 for students at Grades 3–8 and 10 in a limited number of classrooms throughout Missouri. Six 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 pilot forms for Grades 4, 8, and 10 contained twelve selected-response items, four constructed-response items, and one performance event. Then, 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 Workshop.

The purpose of the Score, Revise, and Rewrite (SRR) Workshop was for the participants to score the mathematics items piloted in classrooms in Missouri in October/November of 2003, and to make adjustments to the items and/or rubrics 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. The participants individually scored the students' pilot forms, tallied the results, and then reviewed the items as a group. CTB and DESE personnel were present to facilitate the Score, Revise, and Rewrite workshop. 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 particularly successful in this regard.

The Content and Bias Review (CBR) workshop was conducted in March 2004 with DESE, Missouri educators, Regional Facilitators, and CTB staff involved and was held at the Inn at Grand Glaize in Lake of the Ozarks, Missouri (this workshop was conducted simultaneously with the Communication Arts review). For the Content Review, DESE invited participants from educational sites throughout Missouri to review the Mathematics items and scoring guides for content accuracy and grade-level appropriateness. Participants also 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. 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 (a separate group of people composed of representatives from various backgrounds whose purpose was to screen the items for any racial, socioeconomic, gender, or other sensitivity issues) completed their review of the Mathematics content without rejecting any items; however, a small percentage of the items were revised by the Bias Review committee during their review. At the conclusion of the Content and Bias Review, the final count showed that only 2 items had been rejected from the entire Mathematics item pool, while 635 items had been accepted (or accepted with revision) to become candidates for MAP field testing.

The outcome of the Mathematics item development process from early 2003 to mid-2004 was successful in providing a quantity of custom-written items needed to meet the requirements of the upcoming field test.

During the remaining months of 2004, the Mathematics item pool was used as the basis for the formation of four stand-alone Field Test forms. 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, Mathematics field-test forms A through D were produced as printed test books and prepared for field testing.

The MAP Spring 2005 Mathematics Field Test was successfully administered to Grades 3–8 and 10 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 2006/2007.

The use of a *TerraNova* portion for the Mathematics test was specified from the earliest stages of development, and the use of the *TerraNova* Survey and its match/alignment to the Missouri Content 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 entire test, which includes both of the norm-referenced (NRT) custom portions. As an NRT product, *TerraNova* items are pre-classified to an existing set of *TerraNova* Mathematics standards. In many cases, the alignment of *TerraNova* items to Missouri Content 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 Content 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 Operational test selection was under way.

Just prior to initiating the item selection submission/approval process in fall of 2005, new perspectives arose with regard to the alignment of *TerraNova* items to the Missouri Content Standards/strands; 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 2006 Operational Test, which was administered in March/April of 2006.

### ***Alignment to Grade-Level Expectations***

During and after the administration of the Spring 2006 Operational Test in Mathematics, the back-and-forth process for reviewing mock-up score reports showed a key difference between the 2005 and 2006 Operational Tests in terms of the desired reporting structure; primarily, the 2005 reporting structure was based on Missouri's Content Standards (and Process Standards), whereas the 2006 reporting structure needed to be altered to support the more current DESE curriculum policy, which emphasized *GLE Strands* in Mathematics over the previously promoted *Content Standards* in Mathematics. ("GLEs," a.k.a. Grade Level Expectations, were published by DESE in 2004, which was in the middle of the development cycle for items that were ultimately used operationally in 2006.) It was determined that, in order to provide for the reporting transition to GLE Strands in Mathematics, item alignments for the entire 2006 Operational assessment would need to be checked and cross-checked to determine the impact of assigning items to GLE Strands versus Content Standards. This analysis was performed by CTB, and it was determined that if the realignment were to take place, the transition from Content Standards to GLE Strands as item attributes would have minimal impact on the target blueprint(s). Therefore, the 2006 Operational test in Mathematics was realigned to GLE Strands with DESE approval, and the 2006 Operational score reports reflect this change.

### ***Communication Arts Item Development***

Planning and preparation for the development of item content to be used on the 2006 MAP Operational Test forms commenced in late 2002. The Communication Arts item development plan specified an item development and selection cycle that ran parallel to the Mathematics item development effort: an initial passage selection/item writing workshop (spring 2003), a local pilot study (fall 2003), a content and bias review (spring 2004), item refinements and form

creation (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 and fall 2005), a formal standard-setting process (winter 2005), and ultimately, operational testing (spring 2006) at Grades 3–8 and 11. During the item development process, key workshops were held in conjunction with the Mathematics item development team and its participants.

In May of 2003, a group composed of Missouri educators, Regional Facilitators, DESE staff, and CTB personnel participated in a Communication Arts passage selection and item writing workshop (IWW) 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 test forms. The workshop was conducted with participants selected by DESE to represent educational sites throughout Missouri. During the first part of the workshop, 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 item writing workshop via collaboration between DESE and CTB, which led to the production of Communication Arts test forms used for the local pilot studies. The small-scale pilot was administered in October/November of 2003 for students at Grades 3–8 and 11 in a limited number of classrooms throughout Missouri. Six 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 pilot forms for Grades 3, 7, and 11 contained two selected-response items, four constructed-response items, and one writing prompt. Then, 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.

The purpose of the SRR Workshop was for the participants to score the Communication Arts items piloted in classrooms in Missouri in October/November of 2003, and to make adjustments to the items and/or 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 MAP, and this workshop was viewed by participants as very successful in this regard.

The Content and Bias Review (CBR) workshop was conducted in March 2004 with DESE, Missouri educators, regional facilitators, and CTB staff involved and was conducted at the Inn at Grand Glaize in Lake of the Ozarks, Missouri. For the Content Review, DESE invited over 30 participants (Missouri educators) from educational sites throughout Missouri to review the Communication Arts passages, items, writing prompts, and scoring guides. Training focused on verification of content and process standard designations and the alignment of content to the

Missouri Grade Level Expectations. The participants helped to realign the items and revise items and item alignments where necessary. The Content Review was followed by a Bias Review. The Bias review committee (a separate group composed of representatives from various regions and backgrounds whose purpose was to screen the items for any racial, socioeconomic, gender, or other sensitivity issues) reviewed the material during the same week. At the conclusion of the Content and Bias Review, four items had been rejected from the entire Communication Arts pool, while 257 items had been accepted (or accepted with revision) as candidates for MAP field testing.

During the remaining months of 2004, the Communication Arts item pool was used as the basis for the formation of four stand-alone Field Test forms. The custom-written material was arranged into test forms using *TerraNova* Survey as the anchor item set (this would also encompass the NRT portion of the test when the forms went to operational testing in 2006). 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. Communication Arts field test forms A through D were then produced as printed test books and prepared for field testing.

The MAP Spring 2005 Communication Arts Field Test was successfully administered to Grades 3–8 and 10 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 2006/2007.

Like the Mathematics section of the MAP, the use of a *TerraNova* portion for the Communication Arts tests was specified from the earliest stages of development, and the use of the *TerraNova* Survey and its match/alignment to the Missouri Content 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 entire test, which includes both the NRT and custom portions. *TerraNova* items are pre-classified to an existing set of *TerraNova* Reading and Language Arts standards, and in many cases the alignment of *TerraNova* items to Missouri Content 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 content standards took a less prominent role leading up to the Field Test, but ultimately all *TerraNova* items were matched via a collaborative process between CTB and DESE, with final matches approved by DESE. One notable aspect of the match included a very small number of *TerraNova* items that were originally mapped to *TerraNova* Language standards for the TN shelf product, but were identified by DESE as more properly aligned to Reading standards when the Show-Me Standards were used as the basis for the match. However, this did not affect the core NRT portion of the test (*TerraNova* Reading), as only those items previously identified as *TerraNova* Reading items were used for NRT scores.

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 2006 Operational Test, which was administered in March/April of 2006.

**Table 3. 1: Content Coverage: MAP 2006 Item Maps**

Communication Arts Grade 3							
Content Standard		CR Item #	SR Item #	CR Item #	CR Item #	SR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)	(Session 2)	(Session 3)	(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	3, 4, 5, 6A	1, 2		22, 23, 43		7, 8, 9, 10, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 40, 41
3	Reading – Nonfiction				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,

MAP 2006 Item Maps (cont'd)

Communication Arts Grade 4							
Content Standard		CR Item #	SR Item #		CR Item #	SR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)		(Session 2)	(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

**MAP 2006 Item Maps (cont'd)**

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

MAP 2006 Item Maps (cont'd)

Communication Arts Grade 6							
Content Standard		CR Item #	SR Item #		CR Item #	SR Item #	TN Reading NRT SR Item #
		(Session 1)	(Session 1)		(Session 2)	(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, 5, 6A	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	11	6, 7, 8, 9, 10, 25, 26, 27, 28, 29,
4	Writing Formally & Informally	6B					
5	Combined Reading from Standards 2 & 3	3, 4, 5, 6A	1, 2		17, 18, 19, 43	11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36

**MAP 2006 Item Maps (cont'd)**

<b>Communication Arts Grade 7</b>							
<b>Content Standard</b>		<b>CR Item #</b>	<b>SR Item #</b>	<b>CR Item #</b>	<b>CR Item #</b>	<b>SR Item #</b>	<b>TN Reading NRT SR Item #</b>
		<b>(Session 1)</b>	<b>(Session 1)</b>	<b>(Session 2)</b>	<b>(Session 3)</b>	<b>(Session 3)</b>	<b>(Session 3)</b>
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	3, 4, 5A, 6	1, 2		19		1, 2, 3, 4, 5, 6, 7, 15, 16, 17, 18
3	Reading - Nonfiction				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	5B		1WP			
5	Combined Reading from Standards 2 & 3	3, 4, 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

**MAP 2006 Item Maps (cont'd)**

<b>Communication Arts Grade 8</b>							
<b>Content Standard</b>		<b>CR Item #</b>	<b>SR Item #</b>		<b>CR Item #</b>	<b>SR Item #</b>	<b>TN Reading NRT SR Item #</b>
		<b>(Session 1)</b>	<b>(Session 1)</b>		<b>(Session 2)</b>	<b>(Session 2)</b>	<b>(Session 2)</b>
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

**MAP 2006 Item Maps (cont'd)**

<b>Communication Arts Grade 11</b>							
<b>Content Standard</b>		<b>CR Item #</b>	<b>SR Item #</b>	<b>CR Item #</b>	<b>CR Item #</b>	<b>SR Item #</b>	<b>TN Reading NRT SR Item #</b>
		<b>(Session 1)</b>	<b>(Session 1)</b>	<b>(Session 2)</b>	<b>(Session 3)</b>	<b>(Session 3)</b>	<b>(Session 3)</b>
1	Speaking/Writing Standard English					19, 20, 28, 29, 30, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52	
2	Reading – Fiction/Poetry/ Drama	3, 4, 5A, 6A	1, 2		42		23, 24, 25, 26, 27, 31, 32, 33, 34
3	Reading - Nonfiction				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
4	Writing Formally & Informally	5B, 6B		1WP			
5	Combined Reading from Standards 2 & 3	3, 4, 5A, 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

**MAP 2006 Item Maps (cont'd)**

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

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

MAP 2006 Item Maps (cont'd)

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	6, 24	3, 9, 15, 20, 25, 26	34	5, 15, 27, 30
3	Geometric and Spatial Relationships	17	1, 5, 7, 10, 16, 18, 22	35	13, 16, 25
4	Measurement	8	2, 4, 11, 19, 23, 27		17, 22, 23, 24, 26
5	Data and Probability		12, 13, 14, 21	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, 6, 9, 11, 12, 15, 16, 20, 21	33	13, 17, 26
3	Geometric and Spatial Relationships	7, 14, 19	5, 10, 22, 27	32	22, 28
4	Measurement	24	3, 4, 8, 13, 18, 23		21, 25, 30, 31
5	Data and Probability		1, 17, 25, 26	34	10, 11, 12, 29

MAP 2006 Item Maps (cont'd)

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		3, 6, 12, 15, 16, 17, 18	33	20, 25, 27, 29
3	Geometric and Spatial Relationships	7, 27	4, 11, 20		16, 21, 26, 30
4	Measurement	21	8, 13, 19, 25, 26	35	22, 23
5	Data and Probability	14	1, 2, 5, 9, 10, 22, 23, 24	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	4, 7	1, 3, 8, 9, 11, 15, 18, 20, 24	32	10, 21, 30, 31
3	Geometric and Spatial Relationships	17, 22,	10, 23, 26, 29	33	11, 14, 22
4	Measurement	30	5, 6, 25	34	17, 27, 28
5	Data and Probability	12, 27	2, 13, 14, 16, 19, 21, 28		8, 9, 13, 23

**MAP 2006 Item Maps (cont'd)**

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

**Table 3. 2: MAP 2006 Content Standard Item/Point Distributions**

<b>Communication Arts Grade 3</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	15	15	0	0	15	15	20%
2	Reading – Fiction/Poetry/Drama	18	7	2	27	18	15	2	35	47%
3	Reading – Nonfiction	12	3	0	15	12	6	0	18	24%
4	Writing Formally & Informally	0	2	0	2	0	6	0	6	8%
5	Combined Reading*	30	10	2	42	30	21	2	53	72%
<i>Totals**</i>		30	12	17	57	30	27	17	74	100%

<b>Communication Arts Grade 4</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	10	10	0	0	10	10	15%
2	Reading – Fiction/Poetry/Drama	29	4	0	33	29	9	0	38	58%
3	Reading Nonfiction	6	4	2	12	6	8	2	16	24%
4	Writing Formally & Informally	0	1	0	1	0	2	0	2	3%
5	Combined Reading*	35	8	2	45	35	17	2	54	82%
<i>Totals**</i>		35	9	12	56	35	19	12	66	100%

\* Combined Reading includes all items measuring Standards 2 and 3, and is not included in the totals.

\*\* These totals accurately reflect the total number of items in the test book(s), but may not equal the sum of content standards 1 through 4 due to the fact that in some cases a single item may report to content standards 2 and 3. Therefore, items reporting to more than one content standard are only counted once in this total.

**MAP 2006 Content Standard Item/Point Distribution (cont'd)**

<b>Communication Arts Grade 5</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	12	12	0	0	12	12	18%
2	Reading – Fiction/Poetry/Drama	16	3	0	19	16	6	0	22	33%
3	Reading Nonfiction	16	6	3	25	16	13	3	32	48%
4	Writing Formally & Informally	0	1	0	1	0	1	0	1	1%
5	Combined Reading*	32	9	3	44	32	19	3	54	81%
<i>Totals**</i>		32	10	15	56	32	20	15	67	100%

<b>Communication Arts Grade 6</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	13	13	0	0	13	13	20%
2	Reading – Fiction/Poetry/Drama	21	7	2	30	21	15	2	38	58%
3	Reading Nonfiction	10	1	1	12	10	2	1	13	20%
4	Writing Formally & Informally	0	1	0	1	0	2	0	2	3%
5	Combined Reading*	31	8	3	42	31	17	3	51	77%
<i>Totals**</i>		31	9	16	56	31	19	16	66	100%

\* Combined reading includes all items measuring Standards 2 and 3, and is not included in the totals.

\*\* These totals accurately reflect the total number of items in the test book(s), but may not equal the sum of content standards 1 through 4 due to the fact that in some cases a single item may report to content standards 2 and 3. Therefore, items reporting to more than one content standard are only counted once in this total.

**MAP 2006 Content Standard Item/Point Distribution (cont'd)**

<b>Communication Arts Grade 7</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	16	16	0	0	16	16	22%
2	Reading – Fiction/Poetry/Drama	11	5	2	18	11	10	2	23	32%
3	Reading Nonfiction	22	2	0	24	22	5	0	27	38%
4	Writing Formally & Informally	0	2	0	2	0	6	0	6	8%
5	Combined Reading*	33	7	2	42	33	15	2	50	69%
<i>Totals**</i>		33	9	18	60	33	21	18	72	100%

<b>Communication Arts Grade 8</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	15	15	0	0	15	15	22%
2	Reading – Fiction/Poetry/Drama	15	2	0	17	15	4	0	19	28%
3	Reading Nonfiction	19	6	2	27	19	13	2	34	49%
4	Writing Formally & Informally	0	1	0	1	0	1	0	1	1%
5	Combined Reading*	34	8	2	44	34	17	2	53	77%
<i>Totals**</i>		34	9	17	60	34	18	17	69	100%

\* Combined reading includes all items measuring Standards 2 and 3, and is not included in the totals.

\*\* These totals accurately reflect the total number of items in the test book(s).

**MAP 2006 Content Standard Item/Point Distribution (cont'd)**

<b>Communication Arts Grade 11</b>										
<b>Content Standard</b>		<b>TN NRT Items</b>	<b>CR/WP 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>
1	Speaking/Writing Standard English	0	0	15	15	0	0	15	15	20%
2	Reading – Fiction/Poetry/Drama	9	5	2	16	9	11	2	22	30%
3	Reading Nonfiction	25	2	0	27	25	5	0	30	41%
4	Writing Formally & Informally	0	3	0	3	0	7	0	7	9%
5	Combined Reading*	34	7	2	43	34	16	2	52	70%
<i>Totals**</i>		34	10	17	61	34	23	17	74	100%

\* Combined reading includes all items measuring Standards 2 and 3, and is not included in the totals.

\*\* These totals accurately reflect the total number of items in the test book(s).

<b>Mathematics Grade 3</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>
1	Number and Operations	15	2	0	17	15	4	0	19	28%
2	Algebraic Relationships	4	1	7	12	4	2	7	13	19%
3	Geometric and Spatial Relationships	3	0	6	9	3	0	6	9	13%
4	Measurement	4	2	5	11	4	4	5	13	19%
5	Data and Probability	4	2	5	11	4	4	5	13	19%
<i>Total</i>		30	7	23	60	30	14	23	67	100%

**MAP 2006 Content Standard Item/Point Distribution (cont'd)**

<b>Mathematics Grade 4</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>
1	Number and Operations	15	2	0	17	15	4	0	19	25%
2	Algebraic Relationships	3	4	4	11	3	8	4	15	19%
3	Geometric and Spatial Relationships	4	0	12	16	4	0	12	16	21%
4	Measurement	6	2	5	13	6	4	5	15	19%
5	Data and Probability	4	2	2	8	4	6	2	12	16%
<i>Total</i>		32	10	23	65	32	22	23	77	100%

<b>Mathematics Grade 5</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>
1	Number and Operations	15	0	0	15	15	0	0	15	21%
2	Algebraic Relationships	4	3	6	13	4	6	6	16	23%
3	Geometric and Spatial Relationships	3	2	7	12	3	5	7	15	21%
4	Measurement	5	1	6	12	5	2	6	13	18%
5	Data and Probability	5	1	4	10	5	3	4	12	17%
<i>Total</i>		32	7	23	62	32	16	23	71	100%

**MAP 2006 Content Standard Item/Point Distribution (cont'd)**

<b>Mathematics Grade 6</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>
1	Number and Operations	18	0	0	18	18	0	0	18	26%
2	Algebraic Relationships	3	1	9	13	3	2	9	14	20%
3	Geometric and Spatial Relationships	2	4	4	10	2	9	4	15	22%
4	Measurement	4	1	6	11	4	2	6	12	17%
5	Data and Probability	4	1	4	9	4	2	4	10	14%
<i>Total</i>		31	7	23	61	31	15	23	69	100%

<b>Mathematics Grade 7</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>
1	Number and Operations	19	0	0	19	19	0	0	19	27%
2	Algebraic Relationships	4	1	7	12	4	2	7	13	18%
3	Geometric and Spatial Relationships	4	2	3	9	4	4	3	11	15%
4	Measurement	2	2	5	9	2	5	5	12	17%
5	Data and Probability	3	2	8	13	3	5	8	16	23%
<i>Total</i>		32	7	23	62	32	16	23	71	100%

MAP 2006 Content Standard Item/Point Distribution (cont'd)

Mathematics Grade 8										
Content Standard		TN NRT Items	CR/PE Items	SR Items	Total Items	TN Points	CR/PE Points	SR Points	Total Points	% of Total Points
1	Number and Operations	17	0	0	17	17	0	0	17	22%
2	Algebraic Relationships	4	3	9	16	4	6	9	19	25%
3	Geometric and Spatial Relationships	3	3	4	10	3	6	4	13	17%
4	Measurement	3	2	3	8	3	6	3	12	16%
5	Data and Probability	4	2	7	13	4	4	7	15	20%
<i>Total</i>		31	10	23	64	31	22	23	76	100%

Mathematics Grade 10										
Content Standard		TN NRT Items	CR/PE Items	SR Items	Total Items	TN Points	CR/PE Points	SR Points	Total Points	% of Total Points
1	Number and Operations	12	0	0	12	12	0	0	12	16%
2	Algebraic Relationships	4	4	11	19	4	8	11	23	31%
3	Geometric and Spatial Relationships	4	2	5	11	4	6	5	15	20%
4	Measurement	1	3	4	8	1	6	4	11	15%
5	Data and Probability	4	2	5	11	4	5	5	14	19%
<i>Total</i>		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 MAP paper-and-pencil items.

**Table 3. 3: MAP 2006 Number of Items/Points Measuring Process Standards**

<b>Communication Arts Grade 3</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</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
3.1	0	2	2	0	3	3
3.5	13	2	15	13	4	17

<b>Communication Arts Grade 4</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	7	0	7	7	0	7
1.6	14	4	18	14	9	23
2.1	0	1	1	0	2	2
2.2	0	10	10	0	10	10
2.4	0	1	1	0	1	1
3.1	2	1	3	2	2	4
3.5	12	4	16	12	7	19

<b>Communication Arts Grade 5</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	2	1	3	2	2	4
1.6	14	2	16	14	5	19
1.7	0	2	2	0	2	2
1.8	0	1	1	0	1	1
2.2	0	12	12	0	12	12
2.4	1	0	1	1	0	1
3.1	1	2	3	1	4	5
3.5	14	3	17	14	5	19
3.7	0	1	1	0	2	2

**Number of Items/Points Measuring Process Standards (cont'd)**

<b>Communication Arts Grade 6</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	4	0	4	4	0	4
1.6	13	1	14	13	2	15
1.7	0	1	1	0	1	1
2.1	0	1	1	0	2	2
2.2	0	13	13	0	13	13
2.4	0	2	2	0	3	3
3.1	2	2	4	2	3	5
3.5	12	5	17	12	11	23

<b>Communication Arts Grade 7</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.1	0	1	1	0	1	1
1.5	1	0	1	1	0	1
1.6	18	3	21	18	7	25
2.1	0	2	2	0	6	6
2.2	0	16	16	0	16	16
2.4	0	1	1	0	1	1
3.1	1	0	1	1	0	1
3.5	13	3	16	13	6	19
3.8	0	1	1	0	2	2

<b>Communication Arts Grade 8</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	6	0	6	6	0	6
1.6	16	4	20	16	7	23
1.7	0	2	2	0	5	5
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	2	2
3.5	10	3	13	10	5	15

**Number of Items/Points Measuring Process Standards (cont'd)**

<b>Communication Arts Grade 11</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	4	0	4	4	0	4
1.6	18	2	20	18	5	23
1.8	0	1	1	0	1	1
2.1	0	2	2	0	6	6
2.2	0	15	15	0	15	15
2.4	0	2	2	0	3	3
3.5	12	4	16	12	8	20
3.6	0	1	1	0	2	2

<b>Mathematics Grade 3</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.2	0	1	1	0	1	1
1.5	5	2	7	5	2	7
1.6	7	15	22	7	16	23
1.8	0	1	1	0	2	2
1.10	1	5	6	1	7	8
2.1	0	1	1	0	2	2
3.1	3	2	5	3	3	6
3.3	12	4	16	12	5	17
3.5	2	1	3	2	2	4

<b>Mathematics Grade 4</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.5	3	8	11	3	8	11
1.6	5	9	14	5	13	18
1.8	0	1	1	0	4	4
1.10	2	0	2	2	0	2
2.1	0	2	2	0	4	4
3.1	13	9	22	13	12	25
3.3	8	3	11	8	3	11
3.5	2	2	4	2	3	5
3.6	0	1	1	0	2	2
4.1	0	1	1	0	4	4

**Number of Items/Points Measuring Process Standards (cont'd)**

<b>Mathematics Grade 5</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.4	0	1	1	0	1	1
1.5	7	7	14	7	10	17
1.6	2	11	13	2	13	15
1.8	0	3	3	0	5	5
1.10	0	5	5	0	5	5
3.1	8	3	11	8	5	13
3.3	15	3	18	15	4	19
3.4	0	1	1	0	1	1
3.5	0	1	1	0	1	1
4.1	0	1	1	0	1	1

<b>Mathematics Grade 6</b>						
<b>Process Standards</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 Points</b>	<b>Total Score Points</b>
1.2	0	1	1	0	1	1
1.4	1	0	1	1	0	1
1.5	3	4	7	3	5	8
1.6	2	6	8	2	7	9
1.7	2	1	3	2	1	3
1.8	0	3	3	0	7	7
1.10	0	1	1	0	1	1
2.1	0	2	2	0	3	3
3.1	7	0	7	7	0	7
3.3	13	7	20	13	7	20
3.4	0	3	3	0	3	3
3.5	1	2	3	1	3	4
3.7	2	0	2	2	0	2
4.1	0	1	1	0	2	2

Number of Items/Points Measuring Process Standards (cont'd)

Mathematics Grade 7						
Process Standards	NRT Items (SR)	Custom/CR/ Other Items	Total Items	NRT Points (SR)	Custom/CR/ Other Points	Total Score Points
1.4	1	1	2	1	1	2
1.5	7	2	9	7	2	9
1.6	0	4	4	0	4	4
1.8	0	3	3	0	6	6
1.10	1	1	2	1	1	2
2.1	0	1	1	0	2	2
3.1	7	4	11	7	6	13
3.3	14	9	23	14	11	25
3.4	0	6	6	0	6	6
3.5	1	1	2	1	2	3
3.7	1	1	2	1	3	4
4.1	0	1	1	0	2	2

Mathematics Grade 8						
Process Standards	NRT Items (SR)	Custom/CR/ Other Items	Total Items	NRT Points (SR)	Custom/CR/ Other Points	Total Score Points
1.2	0	1	1	0	1	1
1.4	0	1	1	0	1	1
1.5	4	5	9	4	6	10
1.6	2	11	13	2	14	16
1.8	1	1	2	1	1	2
1.10	3	0	3	3	0	3
3.1	0	3	3	0	6	6
3.3	18	5	23	18	5	23
3.4	1	1	2	1	4	5
3.5	2	4	6	2	5	7
4.1	0	2	2	0	6	6

Mathematics Grade 10						
Process Standards	NRT Items (SR)	Custom/CR/ Other Items	Total Items	NRT Points (SR)	Custom/CR/ Other Points	Total Score Points
1.5	2	2	4	2	2	4
1.6	6	15	21	6	17	23
1.7	0	1	1	0	2	2
1.10	5	1	6	5	2	7
3.1	3	1	4	3	2	5
3.3	9	8	17	9	8	17
3.4	0	2	2	0	5	5
3.5	0	1	1	0	1	1
3.7	0	1	1	0	2	2
4.1	0	4	4	0	9	9

## **Part 4: Reliability and Construct-Related 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 studies of test reliability, convergent validity, and divergent validity. At the end of this section, we provide analyses by Content Standard.

### ***Minimization of Construct-Irrelevant Variance and Under-Representation***

Minimization of construct-irrelevant variance and construct under-representation is addressed in the following steps of the test development process: 1) specification, 2) item writing, 3) review, 4) field testing, 5) test construction, and 6) calibration.

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. Then, empirical evidence, especially item-level data, is used to infer construct irrelevance.

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.

### ***Reliability***

Reliability refers to the consistency of the students' test scores on parallel forms of a test. A reliable test is one that produces scores that are expected to be relatively stable if the test is administered repeatedly under similar conditions. Often, however, it is 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 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) coefficient 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 0.8 are considered acceptable for tests of moderate lengths.

Cronbach's coefficient 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 coefficient 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 4.1 and 4.2. 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 4.3 and 4.4 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 true score is likely to fall. 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 4.1 through 4.14 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 measurement error increases when there are few observations at a particular ability level. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. Figures 4.1 through 4.14 demonstrate that the tests are designed so that measurement error is minimized in the middle of the scale range where the majority of students are located.

### **Decision Accuracy and Consistency**

The *Standards* also make reference to an additional measurement concern that bears on 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 two independent administrations of the test, or one administration of two parallel test forms. However, it is difficult to obtain data from repeated administrations of the same form because of cost, time, and students' recall of the first administration. Also, it is difficult to construct two parallel forms. 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 to find true scores corresponding to observed scores.

CTB used the Livingston-Lewis (1995) methodology to calculate these statistics on the 2006 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 4.5 reports the decision accuracy and consistency classifications conditioned on each level of achievement for each grade/content area. In Table 4.5, 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, 81% of the Grade 4 Mathematics students who were estimated to have a true status of *Below Basic* were correctly classified on the Grade 4 Mathematics test by their observed score.

In Table 4.5, 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, 57% 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 4.6. To evaluate decisions at specific cut points, the joint distribution of all the 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 4.6, all accuracy and consistency statistics conditioned on cut point are above 85%.

### ***Convergent Validity***

Convergent validity is a subtype of construct validity that can be estimated by the extent to which measures of constructs that theoretically should be related to each other are, in fact, observed as related to each other. 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.

This Technical Report summarizes additional statistics that contribute to construct validity (Cronbach's coefficient alpha reported previously in this section and item fit reported in Part 5). The internal consistency coefficient (Cronbach's alpha) reported above is a measure of item homogeneity. In order for a group of items to be homogeneous, they must measure the same construct (construct validity) or represent the same content domain (content validity). Because IRT models were used to calibrate test items and to report student scores, item fit is also 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 5, very few items (13) were flagged for poor model/data fit.

### **Principal Components Analysis**

As another measure of construct validity, CTB examined the unidimensionality of each grade-level MAP test. One of the underlying assumptions of the IRT models used to scale the MAP is that the tests being calibrated are unidimensional, that is, items comprising MAP in each grade/content area measure a single content domain. For example, mathematics items should measure mathematics ability and not measure reading skills. 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 4.15 to 4.28, 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 factors.

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 4.7 and 4.8). To further investigate the unidimensionality of the Communication Arts and Mathematics tests, the ratio of the first eigenvalue to the second eigenvalue was explored (see Tables 4.7 and 4.8). 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 variance such as factual knowledge irrelevant for doing well in a subject does not appear to create significant nuisance factors.

### ***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 MAP subject area tests in 2006. These correlations are based on the census data and the results are shown in Table 4.9. The correlation coefficients ranged from 0.735 (between Communication Arts and Mathematics in Grade 3) to 0.767 (between Communication Arts and Mathematics in Grade 7). The correlation coefficients suggest that individual student scores for Communication Arts and Mathematics are moderately to highly related. These coefficients are not so low as to call into question whether these tests are tapping into achievement constructs, and not so high as to arouse suspicion that the intended constructs are not distinct.

### ***Analyses by Content Standard***

Two sets of analyses were conducted for the content standard level. 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 4.10 through 4.23 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 between the content standards. Tables 4.10 through 4.23 report the Pearson product-moment correlation coefficients among the Content Standard subscores in the off-diagonal portion of the matrix. The reliability coefficients for the Content Standard subscores are shown on the diagonal in each table. The correlation coefficients in Tables 4.10 to 4.23 should be interpreted in the context of the reliability coefficient. In general, we expect to see lower correlation coefficients between variables that are less reliable.

Overall, the correlation 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 correlation coefficients measuring the relationships between Content Standards to those measuring the relationships between content areas (Table 4.9). We expect to see a more modest relationship (smaller correlation coefficients) reported between the Content Standards as a consequence of the lower number of items measuring each content standard (e.g., Writing Formally and Informally).

**Table 4. 1: Reliability in 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	0.912
<b>4</b>	56	66	0.910
<b>5</b>	56	65	0.912
<b>6</b>	56	66	0.903
<b>7</b>	60	72	0.909
<b>8</b>	60	69	0.912
<b>11</b>	61	74	0.911

**Table 4. 2: Reliability in Mathematics**

Grade	Number of Items	Number of Score Points	Cronbach's Alpha
3	60	67	0.907
4	65	77	0.925
5	62	71	0.919
6	61	69	0.920
7	62	71	0.924
8	64	76	0.923
10	61	75	0.940

**Table 4. 3: Cronbach's Alpha by Subgroup, Communication Arts 2006**

	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)	0.91	0.90	0.90	0.90	0.90	0.91	0.89
	Hispanic	0.91	0.90	0.90	0.91	0.90	0.92	0.90
	Asian/Pacific Islander	0.91	0.89	0.92	0.92	0.92	0.93	0.92
	Native American	0.90	0.90	0.91	0.89	0.90	0.93	0.90
Gender	Female	0.90	0.89	0.90	0.90	0.90	0.91	0.90
	Male	0.91	0.91	0.91	0.91	0.91	0.92	0.91
ELL Status	Non-ELL	0.90	0.90	0.91	0.91	0.91	0.92	0.91
	ELL	0.91	0.90	0.90	0.90	0.90	0.91	0.89
IEP Status	Non-IEP	0.88	0.87	0.89	0.88	0.89	0.90	0.89
	IEP	0.92	0.92	0.91	0.90	0.89	0.90	0.88
SES Status	Non-SES	0.89	0.89	0.90	0.89	0.90	0.91	0.90
	SES	0.90	0.90	0.90	0.90	0.90	0.91	0.90
Disability	No Disability	0.88	0.87	0.89	0.88	0.89	0.90	0.89
	Disability	0.92	0.92	0.91	0.90	0.89	0.90	0.87
Accommodations	No Accommodations	0.88	0.88	0.89	0.89	0.89	0.90	0.89
	Accommodations	0.90	0.90	0.89	0.88	0.88	0.89	0.86
Migrant Status	Non-Migrant	0.90	0.90	0.91	0.91	0.91	0.92	0.91
	Migrant	0.91	0.90	0.91	0.92	0.93	0.93	0.90

**Table 4. 4: Cronbach’s Alpha by Subgroup, Mathematics 2006**

		Grade Level						
Group		3	4	5	6	7	8	10
Ethnicity	White (not Hispanic)	0.89	0.92	0.90	0.92	0.92	0.92	0.93
	Black (not Hispanic)	0.91	0.92	0.91	0.91	0.89	0.90	0.91
	Hispanic	0.90	0.92	0.91	0.91	0.91	0.92	0.93
	Asian/Pacific Islander	0.90	0.93	0.91	0.93	0.93	0.94	0.95
	Native American	0.89	0.92	0.92	0.88	0.92	0.92	0.93
Gender	Female	0.90	0.92	0.91	0.92	0.92	0.92	0.94
	Male	0.91	0.93	0.92	0.93	0.93	0.93	0.94
ELL Status	Non-ELL	0.90	0.92	0.91	0.92	0.92	0.93	0.94
	ELL	0.91	0.93	0.92	0.92	0.91	0.92	0.92
IEP Status	Non-IEP	0.89	0.91	0.90	0.91	0.92	0.92	0.93
	IEP	0.93	0.94	0.93	0.92	0.90	0.89	0.89
SES Status	Non-SES	0.89	0.91	0.90	0.92	0.92	0.92	0.94
	SES	0.90	0.92	0.91	0.92	0.91	0.91	0.93
Disability	No Disability	0.89	0.91	0.90	0.91	0.92	0.92	0.93
	Disability	0.93	0.94	0.93	0.92	0.89	0.89	0.89
Accommodations	No Accommodations	0.89	0.91	0.90	0.92	0.92	0.92	0.93
	Accommodations	0.91	0.93	0.92	0.90	0.88	0.88	0.87
Migrant Status	Non-Migrant	0.90	0.92	0.91	0.92	0.92	0.93	0.94
	Migrant	0.91	0.93	0.91	0.91	0.92	0.91	0.91

**Table 4. 5: Decision Accuracy and Consistency Conditioned on Level of Achievement**

		Accuracy				Consistency			
		Below Basic	Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.
Mathematics	3	0.78	0.84	0.81	0.79	0.68	0.79	0.71	0.74
	4	0.81	0.86	0.82	0.79	0.73	0.82	0.71	0.75
	5	0.83	0.85	0.82	0.8	0.71	0.81	0.71	0.75
	6	0.82	0.84	0.81	0.79	0.75	0.79	0.7	0.74
	7	0.85	0.81	0.8	0.8	0.81	0.74	0.67	0.75
	8	0.85	0.8	0.74	0.81	0.81	0.72	0.59	0.77
	10	0.87	0.78	0.82	0.82	0.85	0.69	0.71	0.78
Communication Arts	3	0.81	0.85	0.7	0.82	0.71	0.81	0.55	0.77
	4	0.8	0.83	0.73	0.8	0.74	0.76	0.59	0.8
	5	0.79	0.83	0.73	0.8	0.73	0.78	0.57	0.77
	6	0.81	0.81	0.77	0.76	0.73	0.77	0.6	0.73
	7	0.83	0.83	0.73	0.8	0.74	0.78	0.57	0.74
	8	0.82	0.84	0.64	0.82	0.74	0.8	0.47	0.77
	11	0.83	0.83	0.79	0.77	0.69	0.82	0.61	0.75

**Table 4. 6: Decision Accuracy and Consistency at Achievement Cut Points**

		Accuracy			Consistency		
		Below Basic/Basic	Basic/ Proficient	Proficient/ Advanced	Below Basic/Basic	Basic/ Proficient	Proficient/ Advanced
Mathematics	3	0.97	0.9	0.96	0.95	0.86	0.94
	4	0.97	0.91	0.92	0.95	0.87	0.94
	5	0.97	0.91	0.96	0.95	0.87	0.94
	6	0.96	0.91	0.96	0.94	0.87	0.94
	7	0.94	0.91	0.96	0.91	0.88	0.94
	8	0.93	0.91	0.95	0.91	0.87	0.93
	10	0.94	0.92	0.96	0.91	0.89	0.95
Communication Arts	3	0.96	0.9	0.94	0.95	0.86	0.91
	4	0.96	0.9	0.94	0.93	0.86	0.91
	5	0.96	0.9	0.94	0.94	0.86	0.91
	6	0.95	0.89	0.95	0.93	0.85	0.92
	7	0.95	0.9	0.95	0.92	0.86	0.92
	8	0.95	0.9	0.94	0.93	0.87	0.91
	11	0.96	0.9	0.95	0.94	0.87	0.93

**Table 4. 7: Principal Component Analysis for 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	10.48	18.39	18.39
Second Component	1.83	3.21	21.59
Ratio (First/Second)	5.73		
<b>Grade 4</b>			
First Component	10.33	18.44	18.44
Second Component	1.94	3.47	21.91
Ratio (First/Second)	5.31		
<b>Grade 5</b>			
First Component	10.23	18.26	18.26
Second Component	1.53	2.73	20.99
Ratio (First/Second)	6.70		
<b>Grade 6</b>			
First Component	9.59	17.13	17.13
Second Component	1.46	2.61	19.74
Ratio (First/Second)	6.56		
<b>Grade 7</b>			
First Component	10.21	17.01	17.01
Second Component	1.91	3.19	20.20
Ratio (First/Second)	5.33		
<b>Grade 8</b>			
First Component	10.30	17.16	17.16
Second Component	1.90	3.17	20.32
Ratio (First/Second)	5.42		
<b>Grade 11</b>			
First Component	9.53	15.62	15.62
Second Component	1.70	2.78	18.40
Ratio (First/Second)	5.62		

**Table 4. 8: Principal Component Analysis for 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	10.08	16.52	16.52
Second Component	3.97	6.51	23.02
Ratio (First/Second)	2.54		
<b>Grade 4</b>			
First Component	11.95	18.38	18.38
Second Component	1.78	2.74	21.12
Ratio (First/Second)	6.71		
<b>Grade 5</b>			
First Component	11.47	18.50	18.50
Second Component	1.83	2.95	21.45
Ratio (First/Second)	6.27		
<b>Grade 6</b>			
First Component	11.20	18.36	18.36
Second Component	1.62	2.66	21.02
Ratio (First/Second)	6.90		
<b>Grade 7</b>			
First Component	11.85	19.12	19.12
Second Component	1.92	3.09	22.21
Ratio (First/Second)	6.18		
<b>Grade 8</b>			
First Component	11.46	17.90	17.90
Second Component	1.98	3.09	20.99
Ratio (First/Second)	5.80		
<b>Grade 10</b>			
First Component	14.34	23.51	23.51
Second Component	1.65	2.71	26.22
Ratio (First/Second)	8.68		

**Table 4.9 Inter-Correlation among Content Area Scale Scores**

Grade	Content Area	Communication Arts	Mathematics
3	Communication Arts	1.000	0.735
	Mathematics	0.735	1.000
4	Communication Arts	1.000	0.736
	Mathematics	0.736	1.000
5	Communication Arts	1.000	0.738
	Mathematics	0.738	1.000
6	Communication Arts	1.000	0.755
	Mathematics	0.755	1.000
7	Communication Arts	1.000	0.767
	Mathematics	0.767	1.000
8	Communication Arts	1.000	0.765
	Mathematics	0.765	1.000

**Table 4.10: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 3**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.69				
02	Reading Fiction/Poetry/Drama	27	0.69	0.84			
03	Reading Nonfiction	15	0.66	0.81	0.74		
04	Writing Formally & Informally	2	0.48	0.54	0.49	0.37	
05	Combined Reading	40	0.71	0.98	0.90	0.55	0.88

**Table 4.11: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 4**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	10	0.55				
02	Reading Fiction/Poetry/Drama	33	0.62	0.87			
03	Reading Nonfiction	12	0.57	0.75	0.73		
04	Writing Formally & Informally	1	0.31	0.40	0.37	---	
05	Combined Reading	45	0.64	0.97	0.89	0.42	0.90

**Table 4. 12: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 5**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	12	0.65				
02	Reading Fiction/Poetry/Drama	19	0.64	0.80			
03	Reading Nonfiction	25	0.67	0.81	0.83		
04	Writing Formally & Informally	1	0.33	0.35	0.39	---	
05	Combined Reading	43	0.69	0.93	0.97	0.39	0.90

**Table 4. 13: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 6**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	13	0.64				
02	Reading Fiction/Poetry/Drama	30	0.67	0.85			
03	Reading Nonfiction	12	0.60	0.76	0.76		
04	Writing Formally & Informally	1	0.39	0.49	0.42	---	
05	Combined Reading	42	0.68	0.98	0.88	0.50	0.90

**Table 4. 14: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 7**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	16	0.67				
02	Reading Fiction/Poetry/Drama	18	0.63	0.76			
03	Reading Nonfiction	24	0.67	0.78	0.85		
04	Writing Formally & Informally	2	0.48	0.54	0.51	0.55	
05	Combined Reading	42	0.69	0.93	0.96	0.56	0.89

**Table 4. 15: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 8**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.69				
02	Reading Fiction/Poetry/Drama	17	0.67	0.81			
03	Reading Nonfiction	27	0.65	0.78	0.86		
04	Writing Formally & Informally	1	0.22	0.28	0.38	---	
05	Combined Reading	44	0.70	0.92	0.96	0.36	0.91

**Table 4. 16: Reliability of Each Content Standard and Correlation among Content Standards: Communication Arts Grade 11**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Speaking/Writing Standard English	15	0.62				
02	Reading Fiction/Poetry/Drama	16	0.61	0.75			
03	Reading Nonfiction	27	0.65	0.77	0.83		
04	Writing Formally & Informally	3	0.49	0.60	0.57	0.48	
05	Combined Reading	43	0.67	0.92	0.96	0.62	0.89

**Table 4. 17: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 3**

No.	Content Standard	01	02	03	04	05	01
01	Number and Operations	17	0.73				
02	Algebraic Relationship	12	0.71	0.68			
03	Geometric and Spatial	9	0.57	0.56	0.52		
04	Measurement	11	0.69	0.64	0.52	0.60	
05	Data and Probability	11	0.73	0.70	0.57	0.65	0.71

**Table 4. 18: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 4**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	17	0.78				
02	Algebraic Relationship	11	0.73	0.69			
03	Geometric and Spatial	16	0.65	0.67	0.73		
04	Measurement	13	0.75	0.73	0.68	0.76	
05	Data and Probability	8	0.69	0.68	0.64	0.68	0.58

**Table 4. 19: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 5**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	15	0.73				
02	Algebraic Relationship	13	0.72	0.76			
03	Geometric and Spatial	12	0.59	0.62	0.62		
04	Measurement	12	0.66	0.71	0.65	0.73	
05	Data and Probability	10	0.65	0.67	0.56	0.61	0.65

**Table 4. 20: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 6**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	18	0.79				
02	Algebraic Relationship	13	0.74	0.74			
03	Geometric and Spatial	10	0.64	0.68	0.72		
04	Measurement	11	0.67	0.70	0.68	0.71	
05	Data and Probability	9	0.68	0.68	0.64	0.61	0.63

**Table 4. 21: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 7**

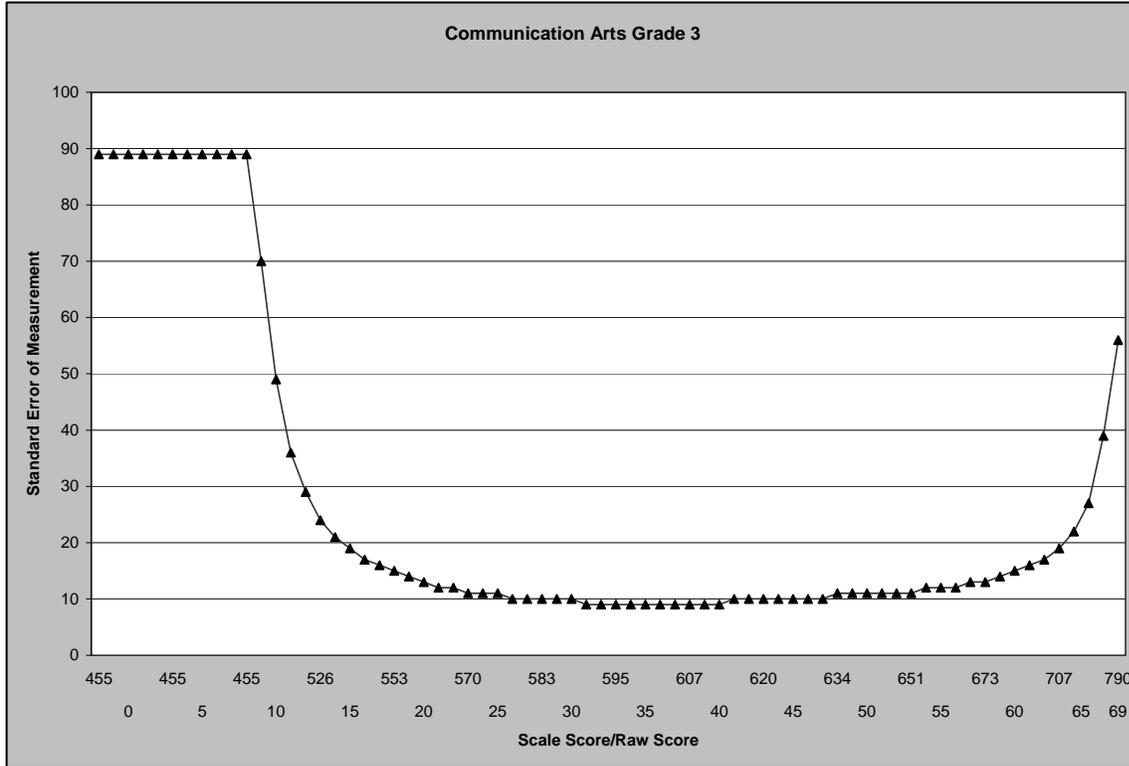
No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	19	0.82				
02	Algebraic Relationship	12	0.74	0.72			
03	Geometric and Spatial	9	0.64	0.64	0.67		
04	Measurement	9	0.67	0.68	0.69	0.68	
05	Data and Probability	13	0.74	0.71	0.64	0.68	0.70

**Table 4. 22: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 8**

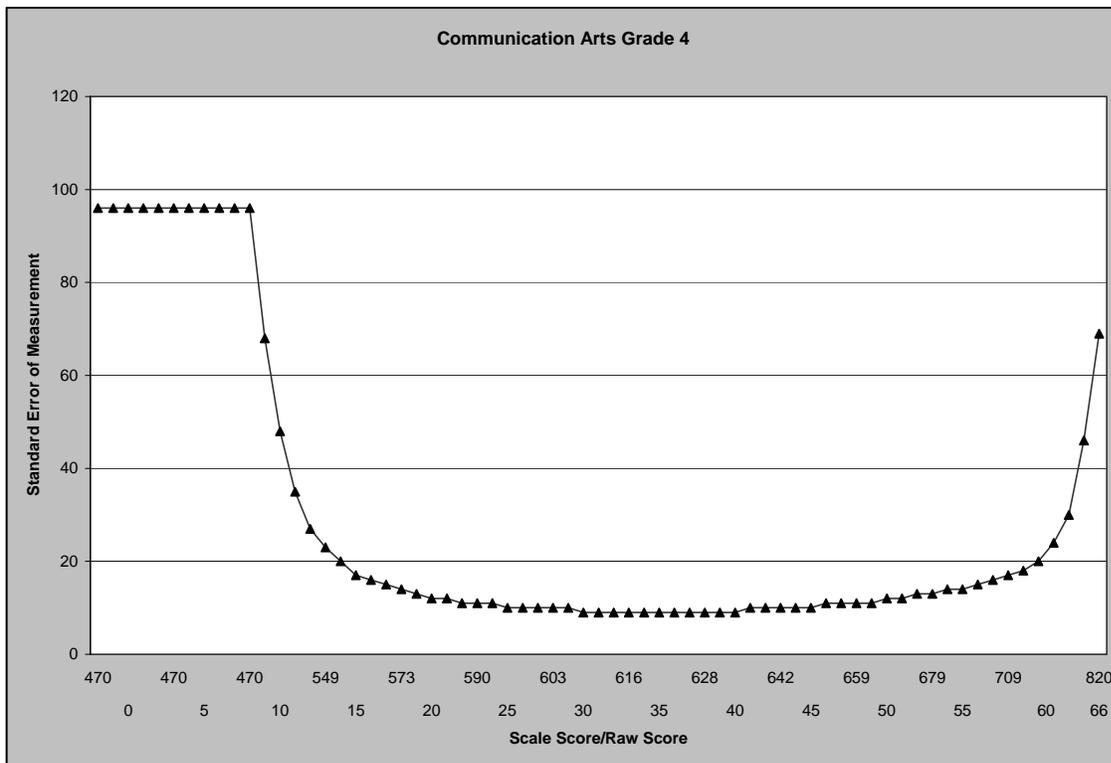
No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	17	0.77				
02	Algebraic Relationship	16	0.75	0.78			
03	Geometric and Spatial	10	0.67	0.72	0.71		
04	Measurement	8	0.68	0.74	0.72	0.69	
05	Data and Probability	13	0.71	0.74	0.69	0.69	0.69

**Table 4. 23: Reliability of Each Content Standard and Correlation among Content Standards: Mathematics Grade 10**

No.	Content Standard	Number of Items	01	02	03	04	05
01	Number and Operations	12	0.71				
02	Algebraic Relationship	19	0.77	0.84			
03	Geometric and Spatial	11	0.71	0.80	0.73		
04	Measurement	8	0.71	0.81	0.78	0.78	
05	Data and Probability	11	0.70	0.76	0.71	0.72	0.68



**Figure 4. 1: SEM Plot Communication Arts Grade 3**



**Figure 4. 2: SEM Plot Communication Arts Grade 4**

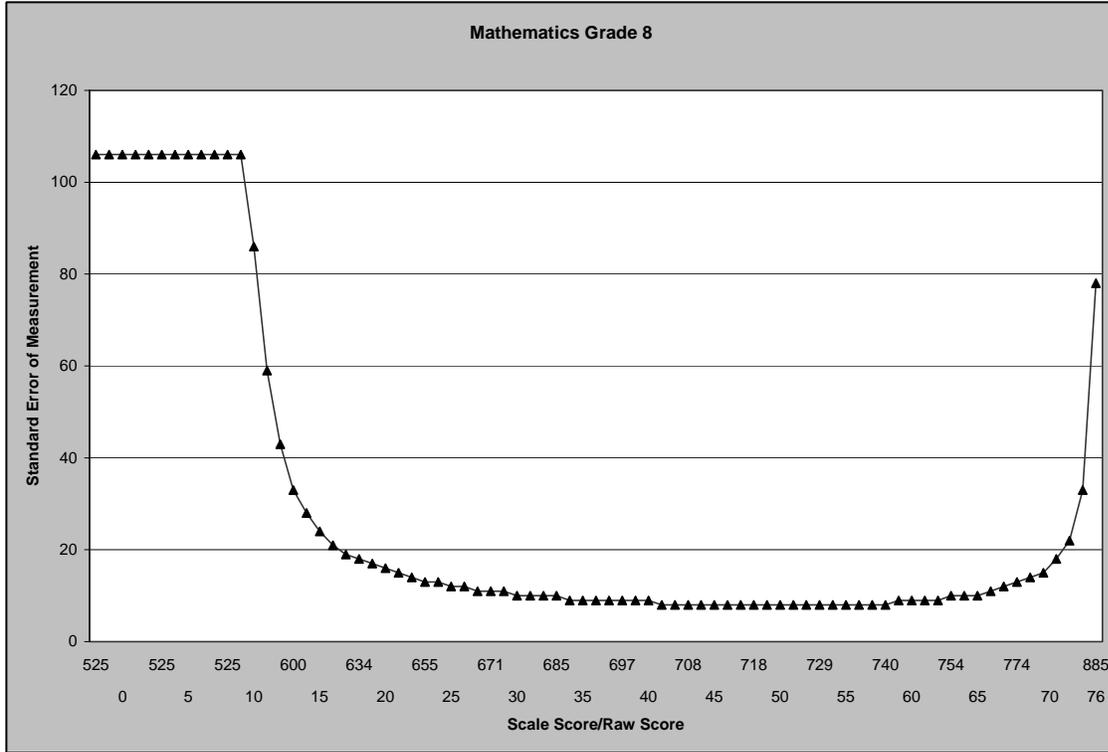




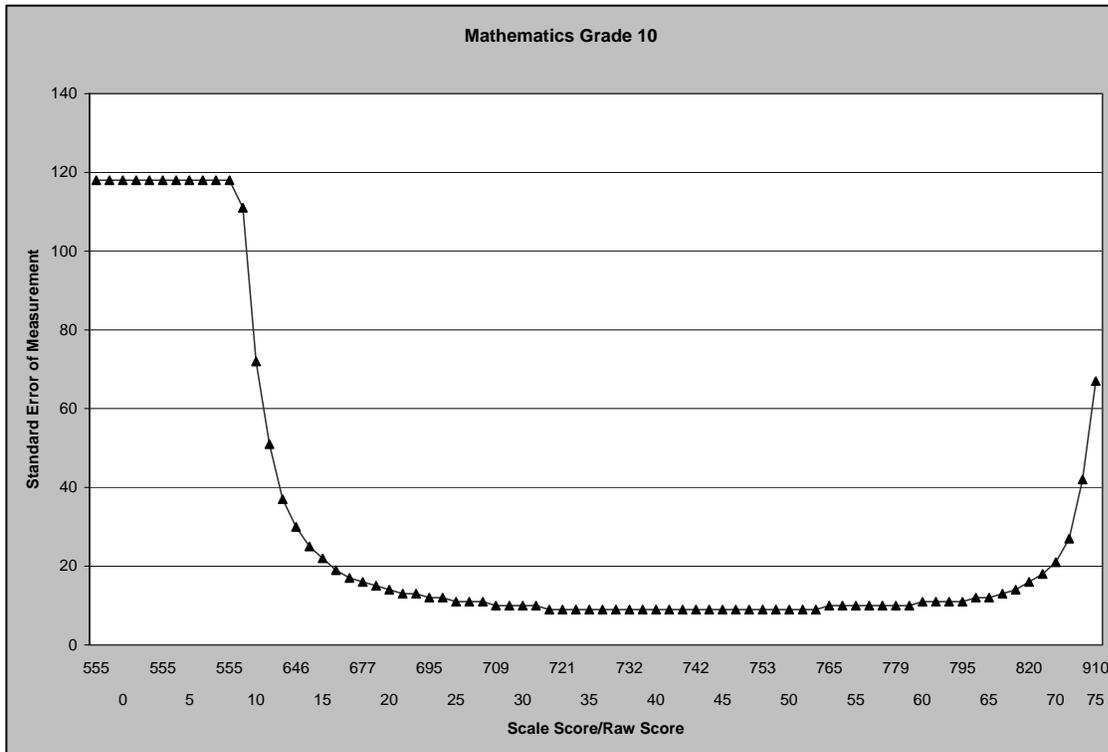






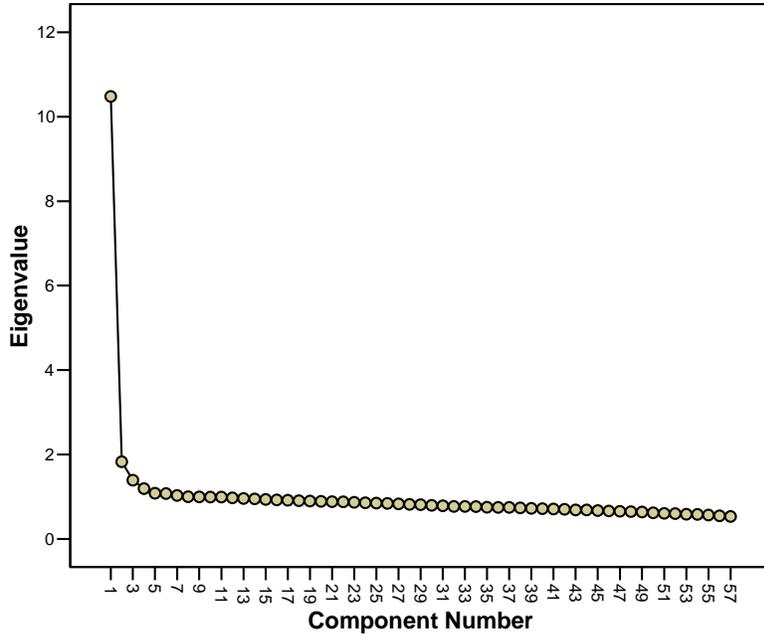


**Figure 4. 13: SEM Plot Mathematics Grade 8**



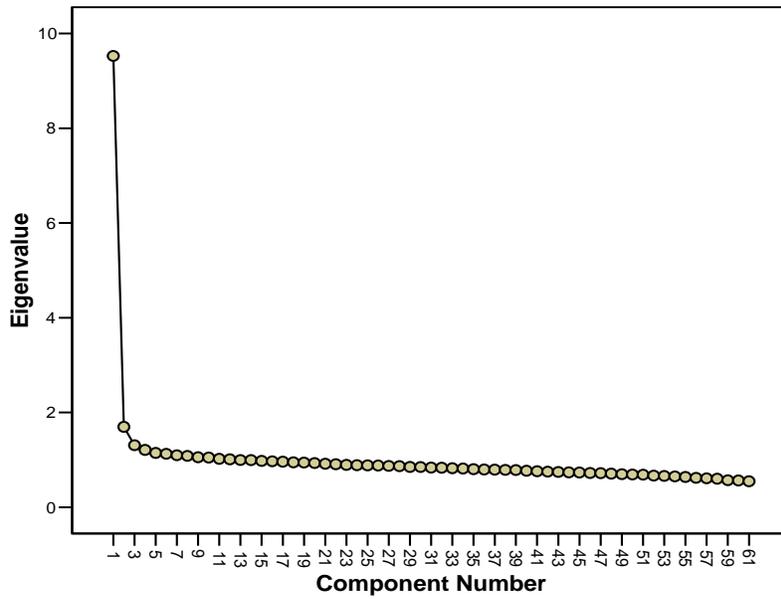
**Figure 4. 14: SEM Plot Mathematics Grade 10**

**Scree Plot**



**Figure 4. 15: Scree Plot Communication Arts Grade 3**

**Scree Plot**



**Figure 4. 16: Scree Plot Communication Arts Grade 4**

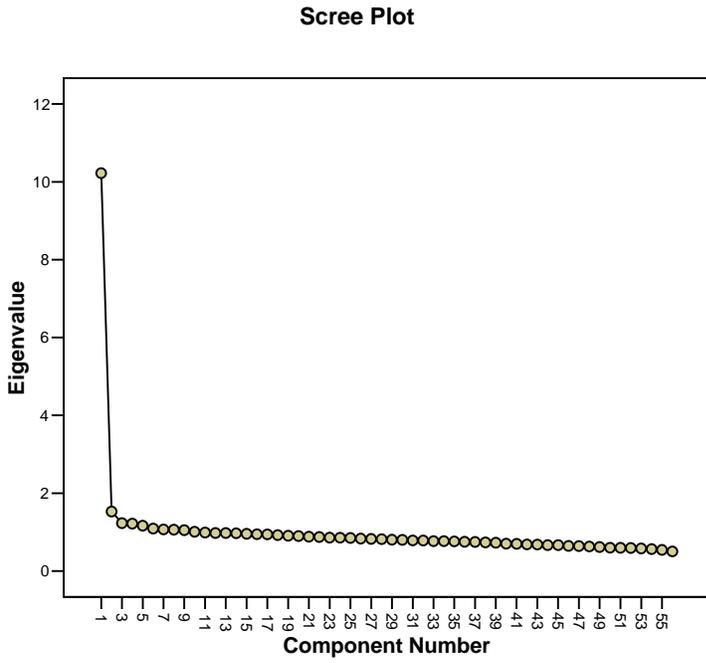


Figure 4. 17: Scree Plot Communication Arts Grade 5

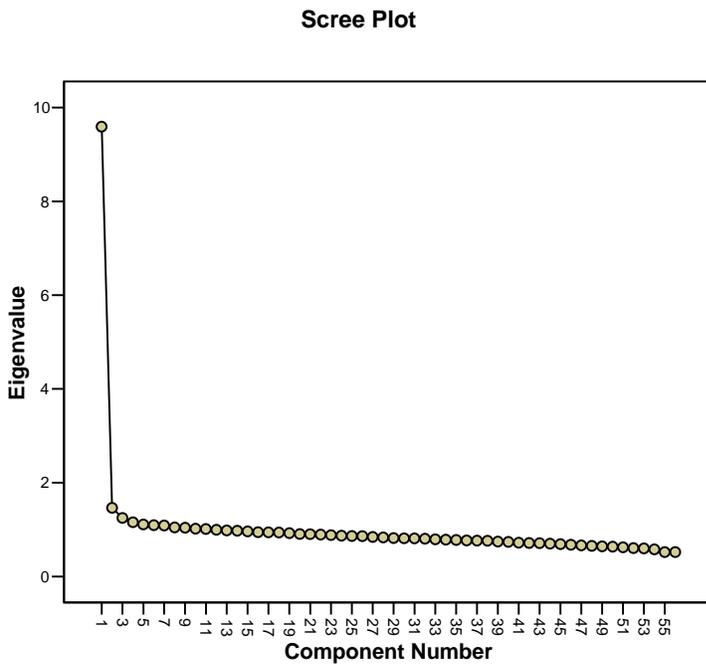


Figure 4. 18: Scree Plot Communication Arts Grade 6

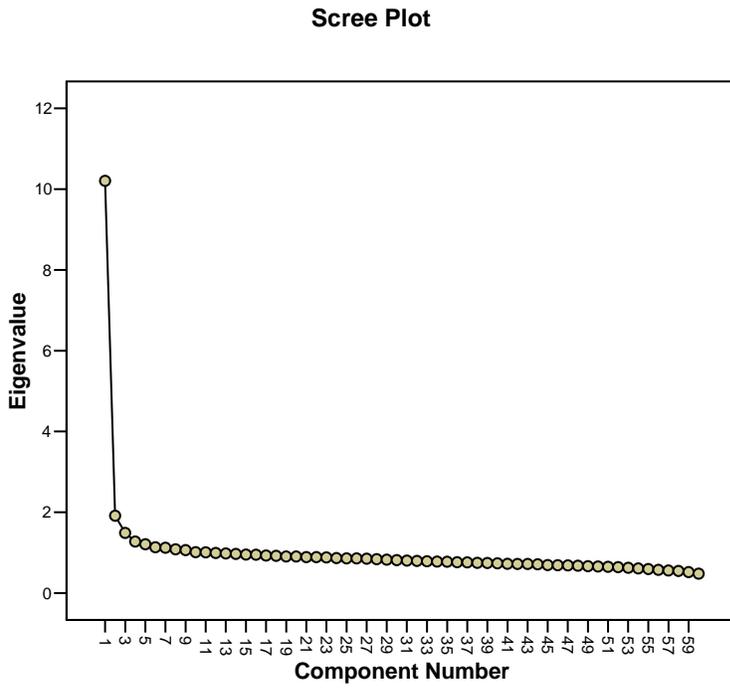


Figure 4. 19: Scree Plot Communication Arts Grade 7

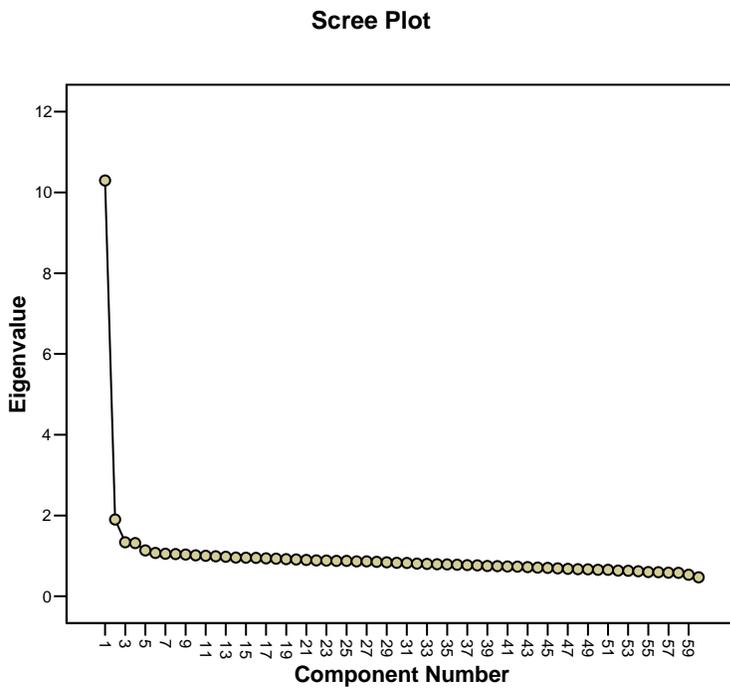


Figure 4. 20: Scree Plot Communication Arts Grade 8

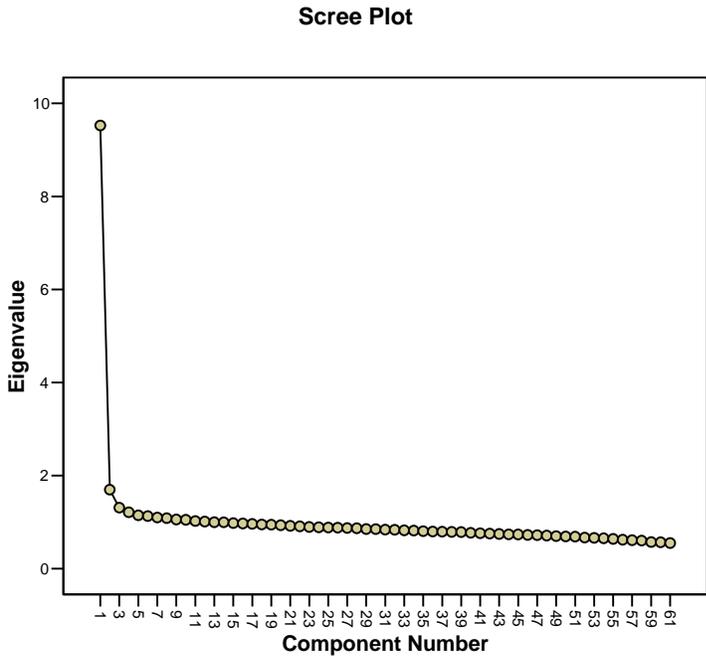


Figure 4. 21: Scree Plot Communication Arts Grade 11

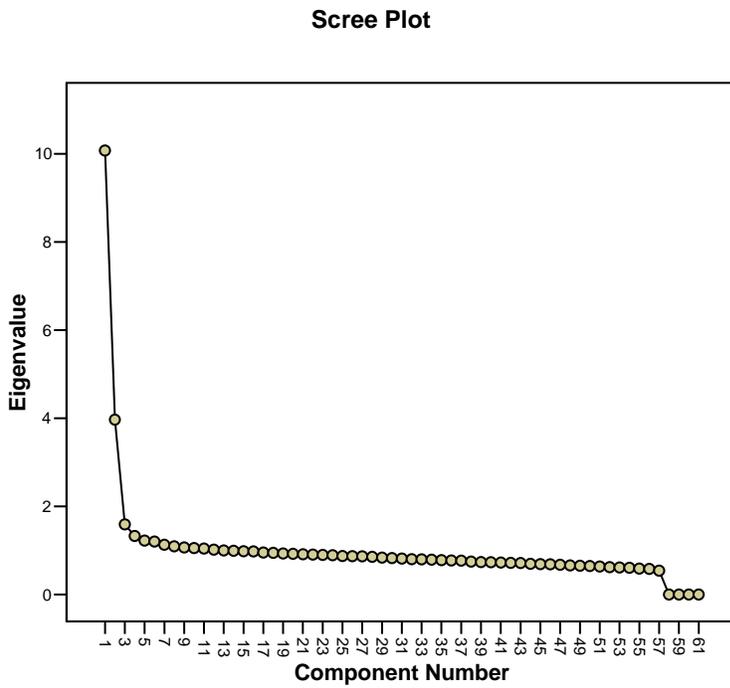


Figure 4. 22: Scree Plot Mathematics Grade 3

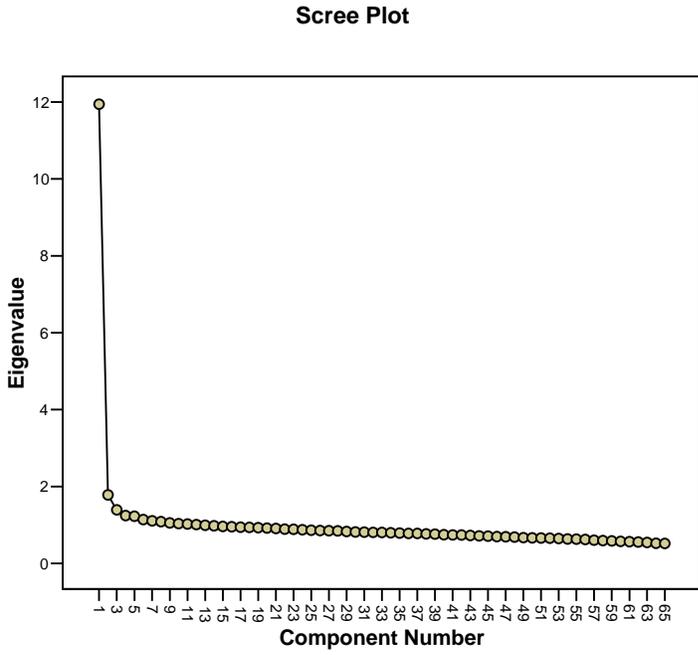


Figure 4. 23: Scree Plot Mathematics Grade 4

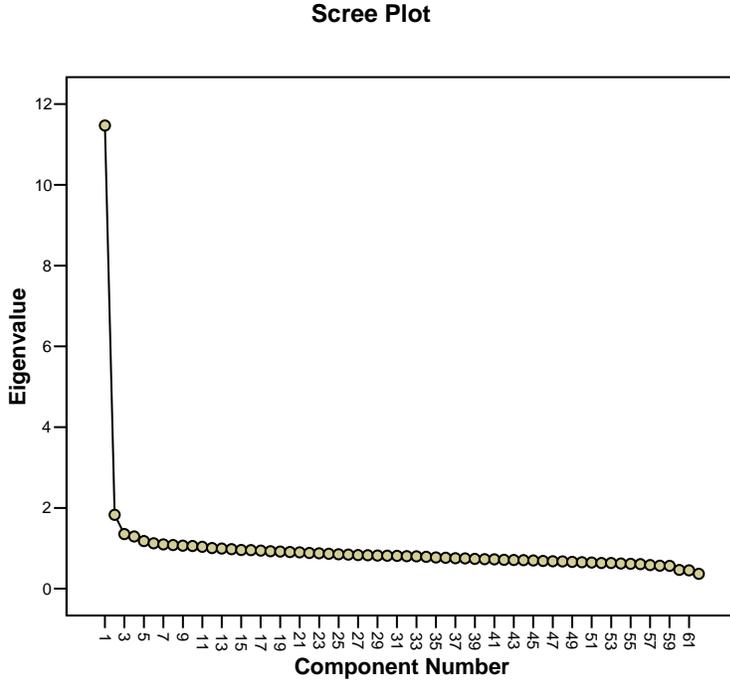


Figure 4. 24: Scree Plot Mathematics Grade 5

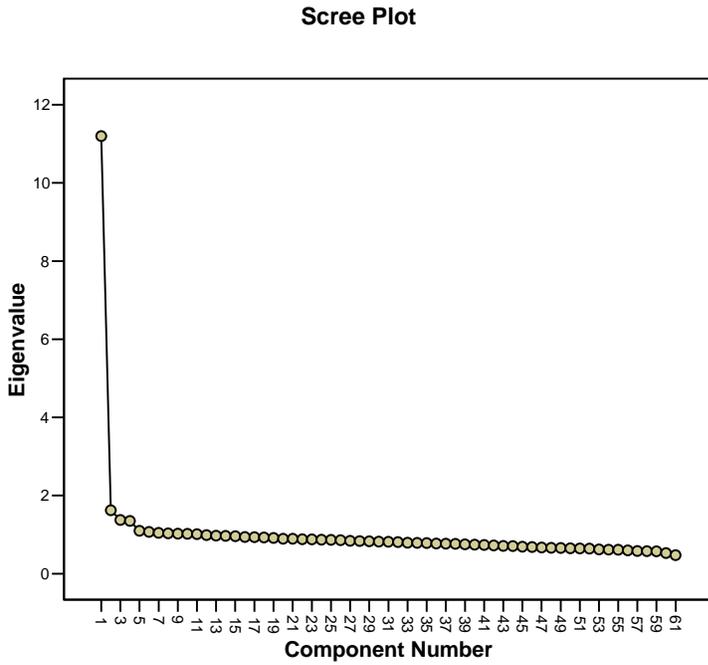


Figure 4. 25: Scree Plot Mathematics Grade 6

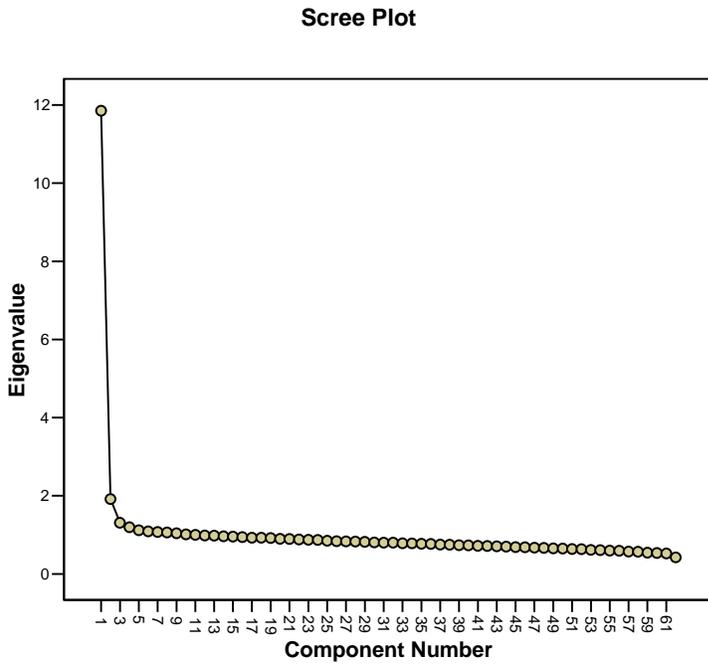


Figure 4. 26: Scree Plot Mathematics Grade 7

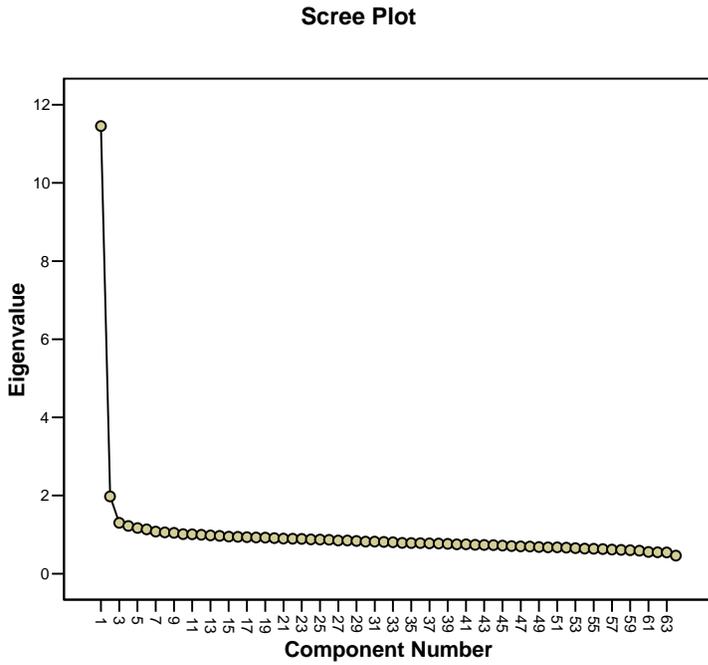


Figure 4. 27: Scree Plot Mathematics Grade 8

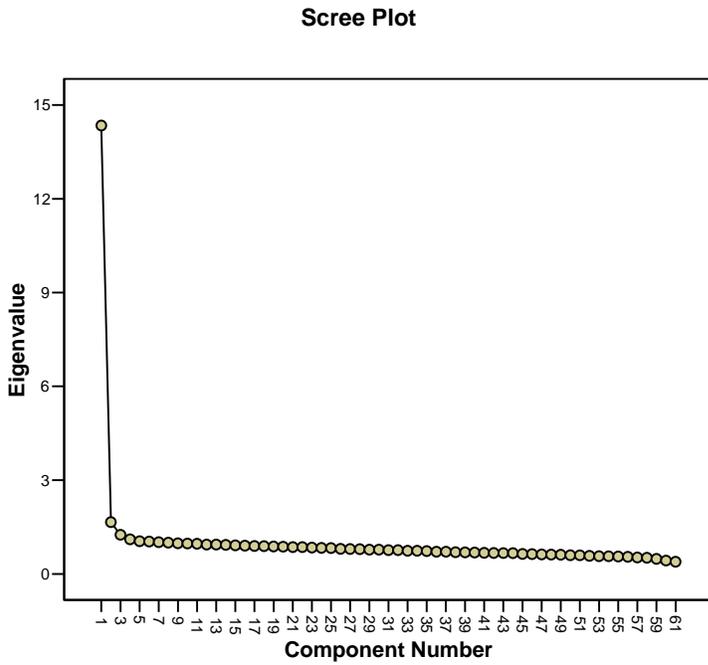


Figure 4. 28: Scree Plot Mathematics Grade 10

## Part 5: Scaling and Linking

Scaling and linking were conducted using the calibration sample described in Part 10. In this section, we first provide a general overview of scaling, then discuss the item response theory (IRT) models used for calibrating the data. We 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 discuss the methods used to link the MAP results to the *TerraNova* scale.

### *Scaling*

The purpose of scaling a test is to enhance its validity by increasing the comparability of test takers' scores. Number-correct raw scores are easily misinterpreted because examinees with the same number of correct items may differ in their ability on the underlying continuum the test measures. This happens because items differ in kind and degree in what they measure. A score that merely counts items correct ignores this important fact.

Fortunately, transformations exist that 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, 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 of complex 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).

### *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.<sup>2</sup>

### 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 useful for flagging items that fit relatively poorly.  $Z_j$  is sensitive to sample size,

<sup>2</sup> Item 8 in Grade 3 Communication Arts did not converge during the calibration process. Item parameters were created for this item by handfitting an item curve to the data.

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 cut-off value is  $(N/1500 \times 4)$  for a given test, where  $N$  is the sample size.

Model fit information is obtained from the  $Z$ -statistic. The  $Z$ -statistic is a transformation of the chi-square ( $Q1$ ) statistic that takes into account differing numbers of score levels as well as sample size:

$$Z_j = \frac{(Q_{1j} - DF)}{\sqrt{2DF}}, \text{ where } j = \text{item } j.$$

The  $Z$ -statistic is an index of the degree to which observed proportions of students with each item score are close to the proportions that would be predicted by the estimated thetas and item parameters. These values are computed for ten intervals corresponding to deciles of the theta distribution. The  $Z$ -statistic is used to characterize item fit. The critical value of  $Z$  is different for each grade because it is dependent on sample size.

Thirteen MAP operational items were flagged for poor fit. In Communication Arts, three items were flagged for poor fit in Grade 3, one item was flagged for poor fit in each of Grades 4 and 5, and two items were flagged for poor fit in Grade 11. In Mathematics, one item was flagged for poor fit in each of Grades 4 and 7, and two items were flagged for poor fit in each of Grades 5 and 10. Table 5.1 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 fewer students occupy these scale score ranges. 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.

**Table 5. 1: Item Fit Statistics for Misfitting Items**

<b>CONTENT</b>	<b>Grade</b>	<b>Session</b>	<b>Item</b>	<b>Chi Square</b>	<b>DF</b>	<b>Total N</b>	<b>Z</b>	<b>Observed</b>	<b>Predicted</b>	<b>Obs-Pred</b>
Comm. Arts	3	S1	3	81.11	17	4099	11.00	0.5965	0.5988	-.0023
Comm. Arts	3	S3	26	81.93	7	4107	20.03	0.6043	0.6099	-.0056
Comm. Arts	3	S3	43	91.52	17	4106	12.78	0.7622	0.7617	0.0005
Comm. Arts	4	S2	19	130.47	17	4037	19.46	0.7312	0.7317	-.0004
Comm. Arts	5	S2	9	49.24	7	3763	11.29	0.7491	0.7476	0.0015
Comm. Arts	11	S1	5A	127.98	17	3988	19.03	0.5226	0.5248	-.0022
Comm. Arts	11	S3	42	109.71	26	3979	11.61	0.4108	0.4150	-.0042
Mathematics	4	S1	14	118.34	17	4067	17.38	0.7690	0.7676	0.0013
Mathematics	5	S2	17	49.65	7	3780	11.40	0.7447	0.7447	0.0000
Mathematics	5	S2	35	128.95	26	3788	14.28	0.4122	0.4151	-.0029
Mathematics	7	S2	35	119.21	26	4339	12.93	0.4281	0.4313	-.0032
Mathematics	10	S1	20	91.83	17	4160	12.83	0.3952	0.3976	-.0024
Mathematics	10	S1	26	79.86	17	4024	10.78	0.3108	0.3136	-.0029

Figures 5.1 through 5.13 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>3</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 light of 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 thirteen MAP items flagged for misfit, the differences do not seem to be of practical importance. Because the understanding of the causes of misfit are currently the subject of continuing investigation, misfitting items that have content validity are often retained for use in an assessment and monitored over a period of usage. A large number of misfitting items in an assessment would indicate that caution should be exercised in the interpretation of the overall score.

Figure 5.1 presents the ICC for Item 3 (Session 1) on the Grade 3 Communication Arts MAP. This is a 2-point, constructed-response item. As shown, there is poor fit at the lower end of level 1 (score point 0). There is poor fit at the lower and upper ends of level 2 (score point 1). At level 3 (score point 2), there is poor fit toward the upper end of the ability distribution.

<sup>3</sup> For constructed response items, there will be one graph for each score level. For example, a 2-point item will have three graphs, for 0, 1, and 2 score points.

Figure 5.2 presents the ICC for Item 26 (Session 3) on the Grade 3 Communication Arts MAP. Item 26 is a selected-response item. As shown, fewer lower ability students answered Item 26 correctly than the IRT model predicted.

Figure 5.3 presents the ICC for Item 43 (Session 3) on the Grade 3 Communication Arts test. This is a 2-point, constructed-response item. All levels appear to have reasonable fit throughout the distribution range; however, the spike in the middle of the distribution of level 2 may account for the item being flagged for poor fit.

Figure 5.4 presents the ICC for Item 19 (Session 2) on the Grade 4 Communication Arts test. This is a 2-point, constructed-response item. There is poor fit at the lower end of ability distribution for levels 1, 2, and 3. There is good fit throughout the rest of the ability distribution for level 1. There is poor fit at the upper end of the ability distribution for levels 2 and 3. At the lower end of levels 2 and 3, there are fewer students than expected obtaining a score of 1 out of 2 or 2 out of 2, respectively. Conversely, at the lower end of level 1, more students than expected obtained a score of 0 out of 2.

Figure 5.5 presents the ICC for Item 9 (Session 2) on the Grade 5 Communication Arts MAP. Item 9 is a selected-response item. As shown, fewer lower ability students answered this item correctly than the IRT model predicted.

Figure 5.6 presents the ICC for Item 5A (Session 1) on the Grade 11 Communication Arts test. This is a 2-point, constructed-response item. There is poor fit at the lower end of ability distribution for levels 1 and 2; however, there is good fit throughout the rest of the ability distribution for both levels. At the lower end of level 2, there are fewer students than expected obtaining a score of 1 out of 2. Conversely, at the lower end of level 1, there are more students than expected obtaining a score of 0 out of 2.

Figure 5.7 presents the ICC for Item 42 (Session 3) on the Grade 11 Communication Arts test. This is a 3-point, constructed-response item. There appears to be poor fit throughout the ability distribution of level 2.

Figure 5.8 presents the ICC for Item 14 (Session 1) on the Grade 4 Mathematics test. This is a 2-point, constructed-response item. There is poor fit at the lower end and middle of ability distribution for level 1, poor fit throughout the ability distribution for level 2, and poor fit at tails of the distribution for level 3.

Figure 5.9 presents the ICC for Item 17 (Session 2) on the Grade 5 Mathematics MAP. Item 17 is a selected-response item. As shown, fewer lower ability and more higher ability students answered Item 17 correctly than the IRT model predicted.

Figure 5.10 presents the ICC for Item 35 (Session 2) on the Grade 5 Mathematics test. This is a 3-point, constructed-response item. There is poor fit throughout the ability distribution of level 1. Levels 3 and 4 show poor fit at the upper end.

Figure 5.11 presents the ICC for Item 35 (Session 2) on the Grade 7 Mathematics test. This is a 3-point, constructed-response item. There is poor fit at the lower end of ability for levels 1 and 2

but good fit throughout the rest of the ability distribution. At the lower end of levels 1 and 2, fewer students than expected obtained those score points.

Figure 5.12 presents the ICC for Item 20 (Session 1) on the Grade 10 Mathematics test. This is a 2-point, constructed-response item. There is poor fit through the middle to upper range of the ability distribution for level 2.

Figure 5.13 presents the ICC for Item 26 (Session 1) on the Grade 10 Mathematics test. This is a 2-point, constructed-response item. There is poor fit at the upper end of ability distribution for levels 1, 2, and 3; however, there is good fit throughout the rest of the ability distribution for all three levels. At the upper end of levels 1 and 2, there are fewer students than expected obtaining a 0 or 1 score point. Conversely, at the upper end of level 3, there are more students than expected obtaining a score of 2 out of 2.

### **Linking Method and Results**

After the initial IRT item calibration, item parameters were linked to the *TerraNova* scale using the Stocking & Lord (1983) equating procedure. This approach takes advantage of the vertical properties of the *TerraNova* scale and allows for comparison of the results from the 2006 administration to results from future administrations.

For the linking, the intact *TerraNova* Survey items served as anchors. Linking was performed using a test characteristic curve (TCC) method proposed by Stocking & Lord (1983), and implemented using PARDUX (Burket, 1995). Figures 5.14 through 5.27 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 lightest TCC lines in the plots (denoted “Anc Input”) are the TCCs for the original *TerraNova* anchor items. The medium-dark lines (denoted “Anc Est”) 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 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 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 mean scale scores as grade level increases. Figures 5.28 and 5.29 display the scale scores for several points on the score distributions for each grade of the MAP. These scale scores indicate the growth, or change, in score by grade of (a) the lowest and highest obtainable scale scores (LOSS and HOSS), (b) the mean scale score, and (c) the scale scores one standard deviation below and above the mean (-1 and +1 SD).

For Mathematics, the mean scale scores range from 618 to 726, as shown in Figure 5.28. Steady mean growth in the scale occurs from Grades 3 to 6, followed by a decline in rate of mean

growth in the middle grades, and an increase in the rate of mean growth in Grades 8 and 10. For Communication Arts, mean scale scores range from 640 to 718, as shown in Figure 5.29. Steady mean growth in the scale occurs from Grades 3 to 5, followed by a decline in the rate of mean growth in the middle grades, and an increase in the rate of mean growth in Grades 8 and 11. However, the average mean growth across grades is approximately the equivalent of the standard error of measurement for these grade level forms so the mean growth is about one standard error per year. This suggests the grade level tests are accurate enough to detect the expected yearly growth for an individual student who is at the median with some confidence. Students at other points in the ability distribution would need scale scores to increase more than this in order to evince evidence of true growth.

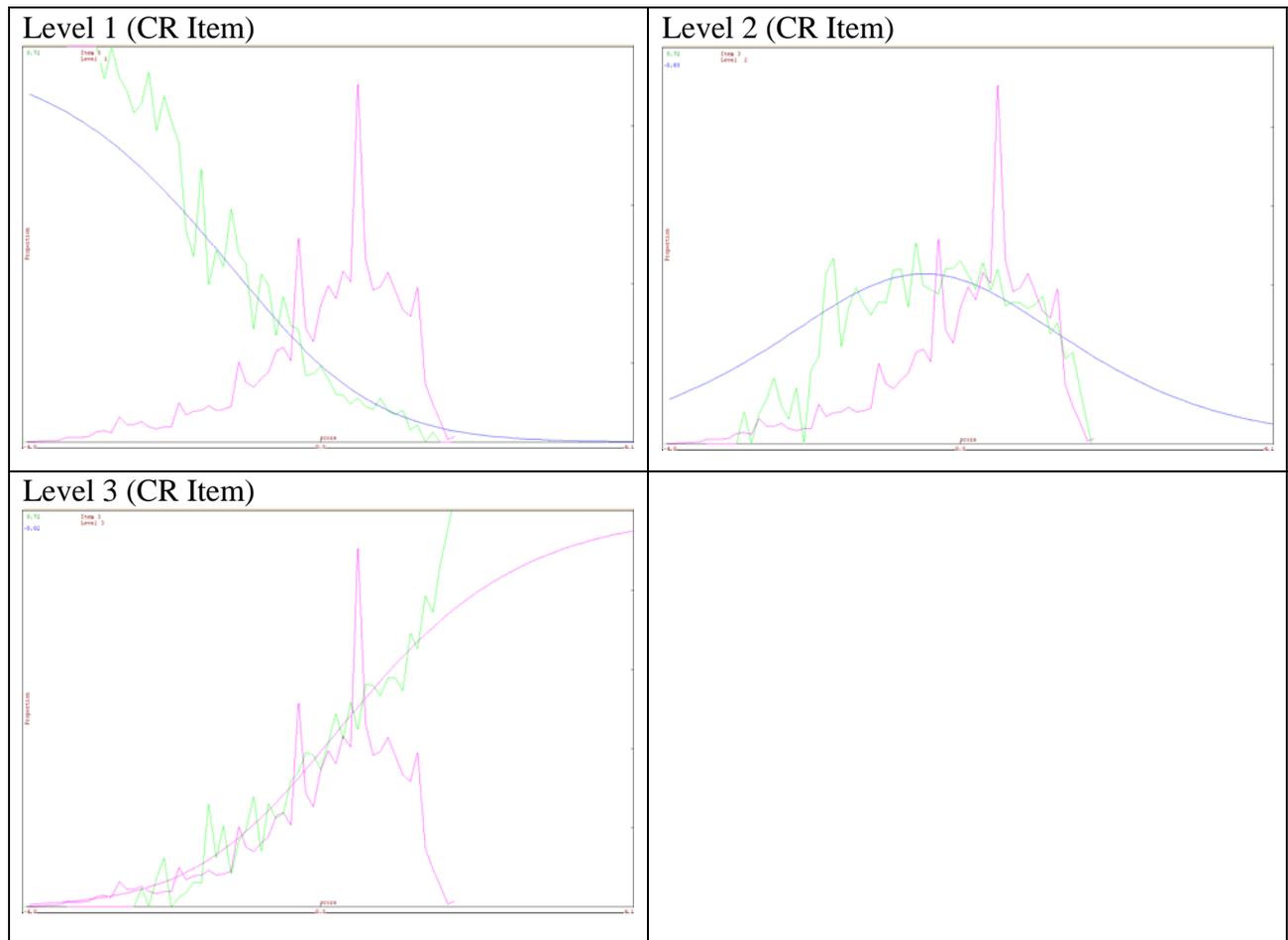
Figure 5.30 shows the TCCs by grade for the Communication Arts MAP, and Figure 5.31 shows the TCCs by grade for the Mathematics MAP. Because these tests were linked to the *TerraNova* scale, they have an underlying vertical scale. By plotting the TCCs together, we can demonstrate that the tests increase 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 through 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, occasionally these estimates have standard errors of measurement that are very large, 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 5.2 shows the LOSS and HOSS values used for each grade of the Communication Arts and Mathematics MAPs.

**Table 5. 2: LOSS and HOSS Values by Grade and Content Area**

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 5. 1. Item characteristic curve for Grade 3 Communication Arts, Item 3 (Z = 11.00)**

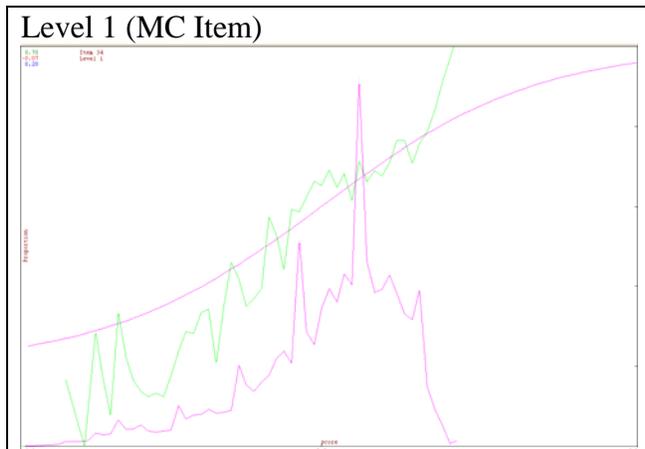


Figure 5. 2. Item characteristic curve for Grade 3 Communication Arts, Item 26 ( $Z = 20.03$ )

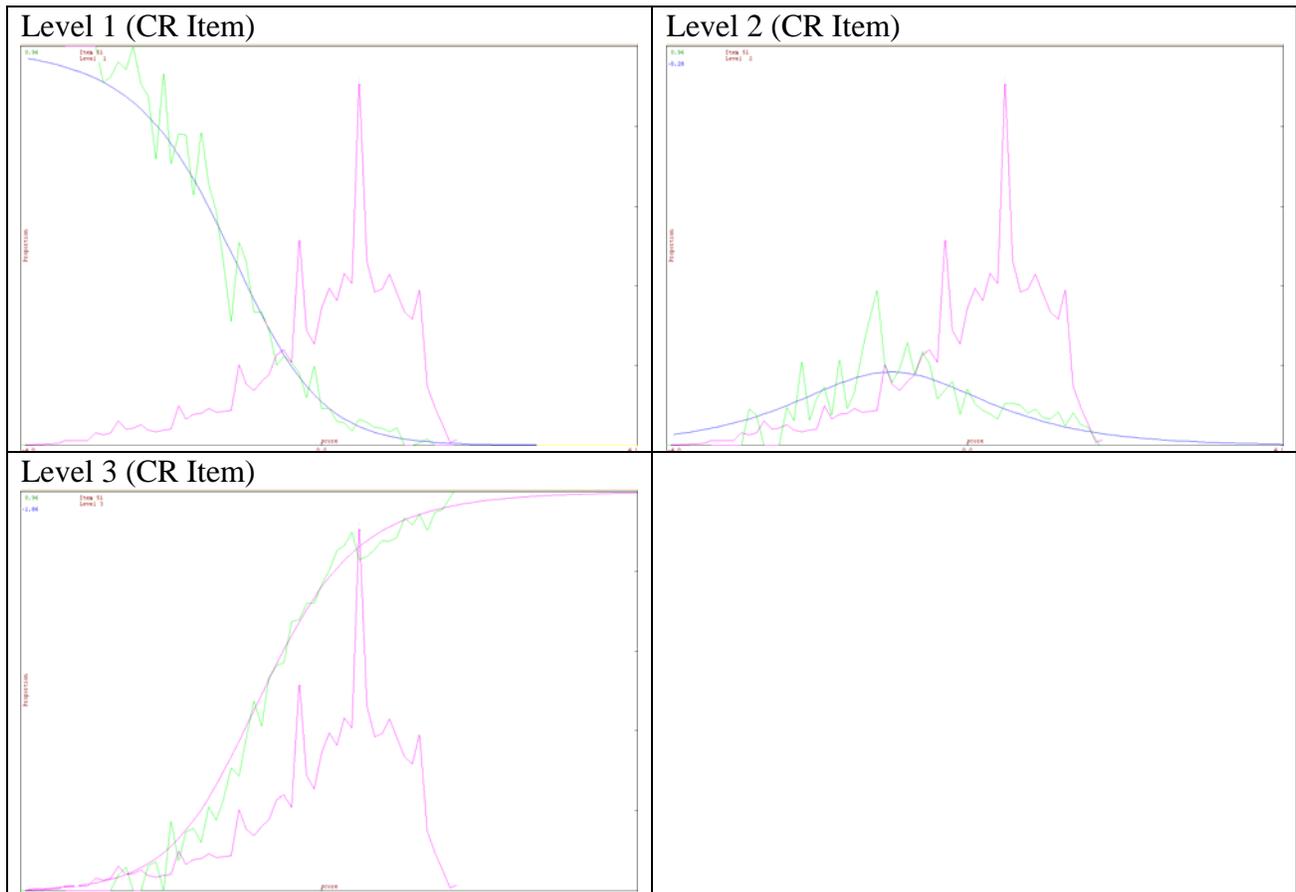


Figure 5. 3. Item characteristic curve for Grade 3 Communication Arts, Item 43 ( $Z = 12.78$ )

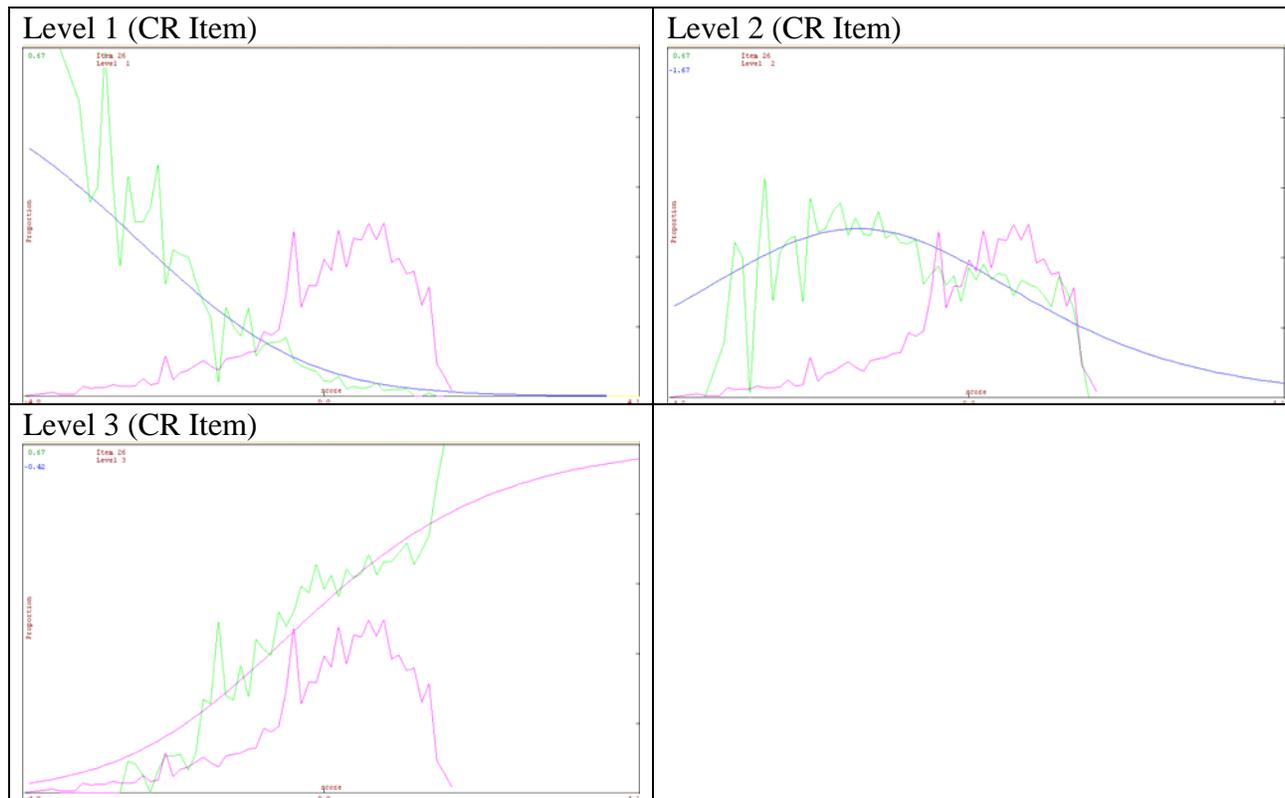


Figure 5. 4. Item characteristic curve for Grade 4 Communication Arts, Item 19 (Z = 19.46)

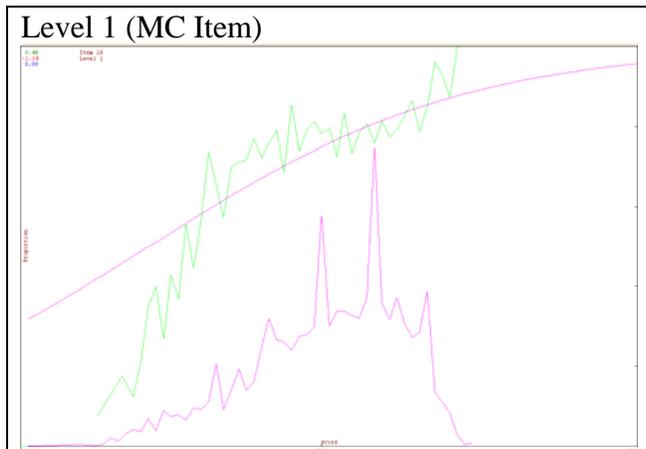


Figure 5. 5. Item characteristic curve for Grade 5 Communication Arts, Item 9 (Z = 11.29)

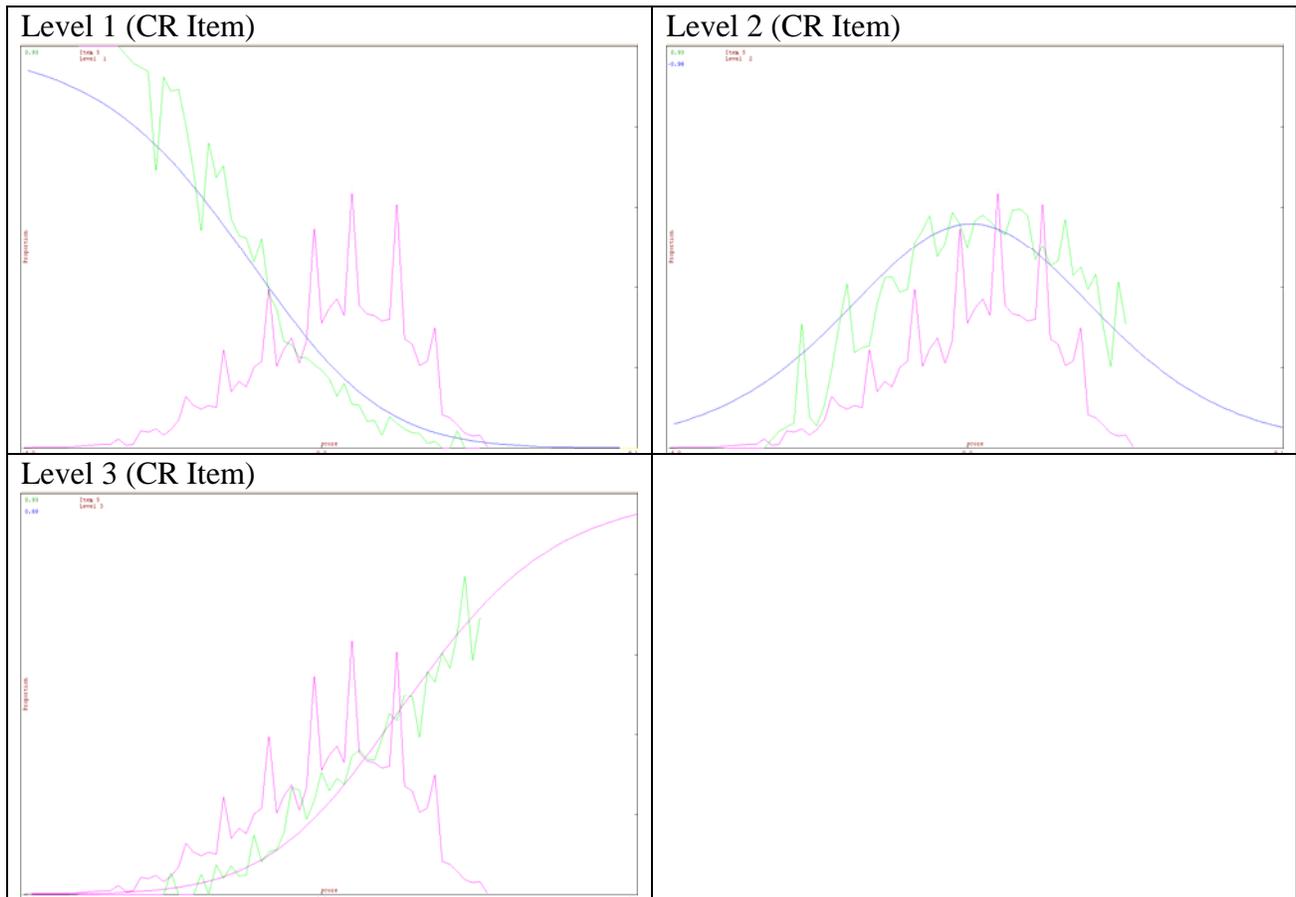


Figure 5. 6. Item characteristic curve for Grade 11 Communication Arts, Item 5A (Z = 19.03)

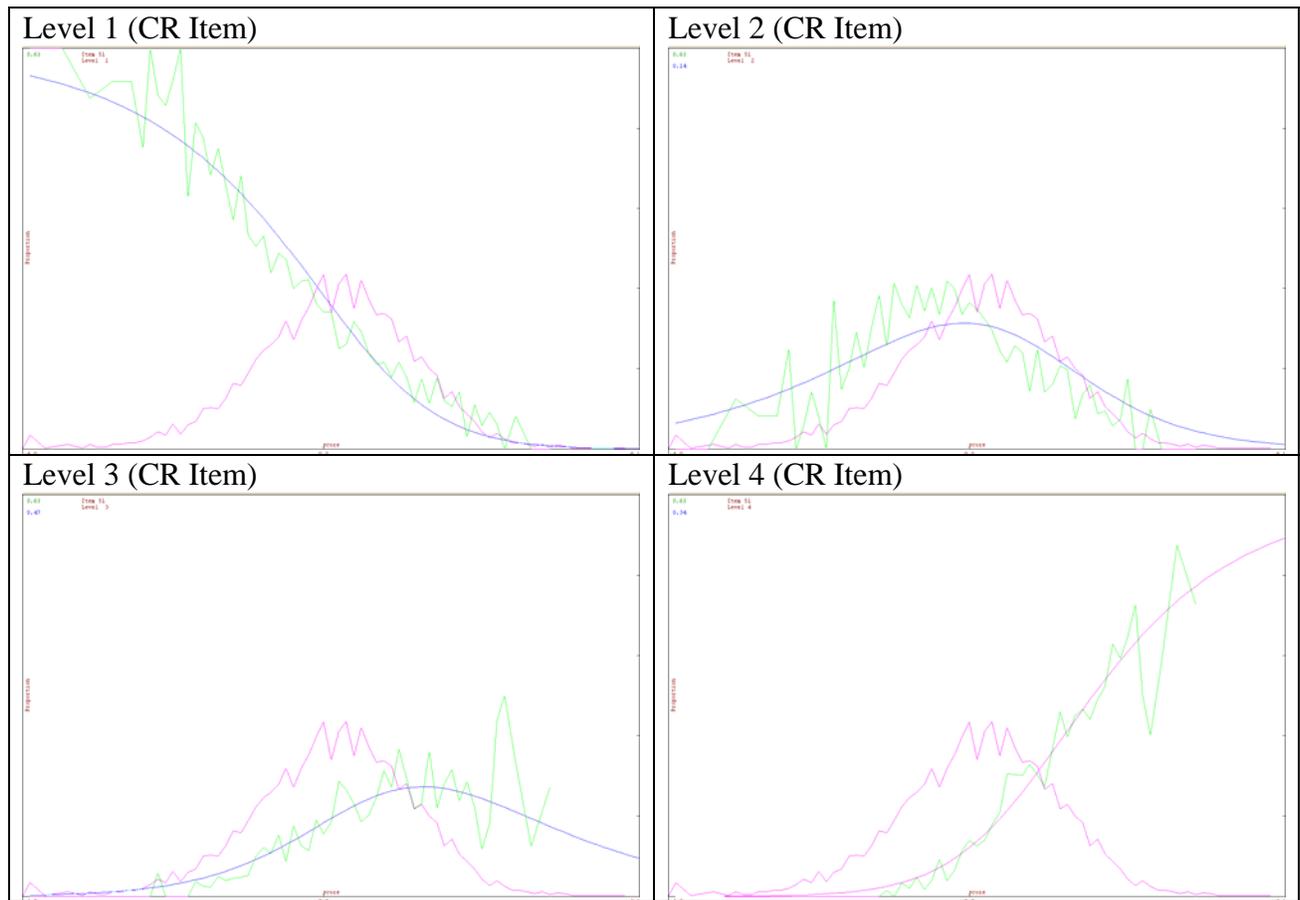


Figure 5. 7. Item characteristic curve for Grade 11 Communication Arts, Item 42 ( $Z = 11.61$ )

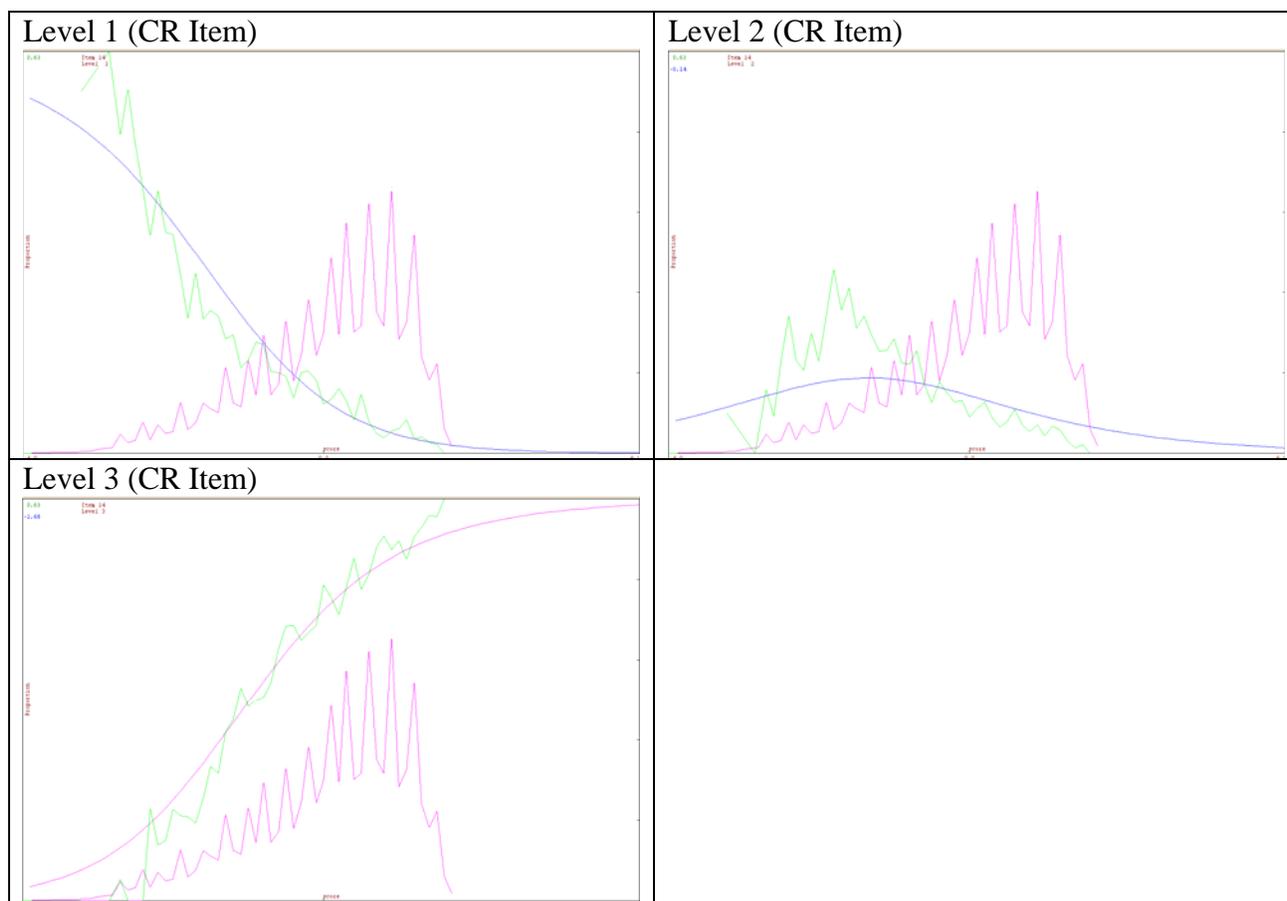


Figure 5. 8. Item characteristic curve for Grade 4 Mathematics, Item 14 (Z = 17.38)

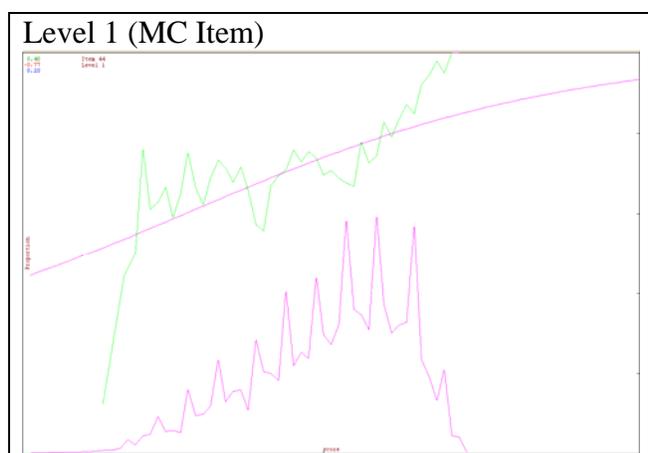
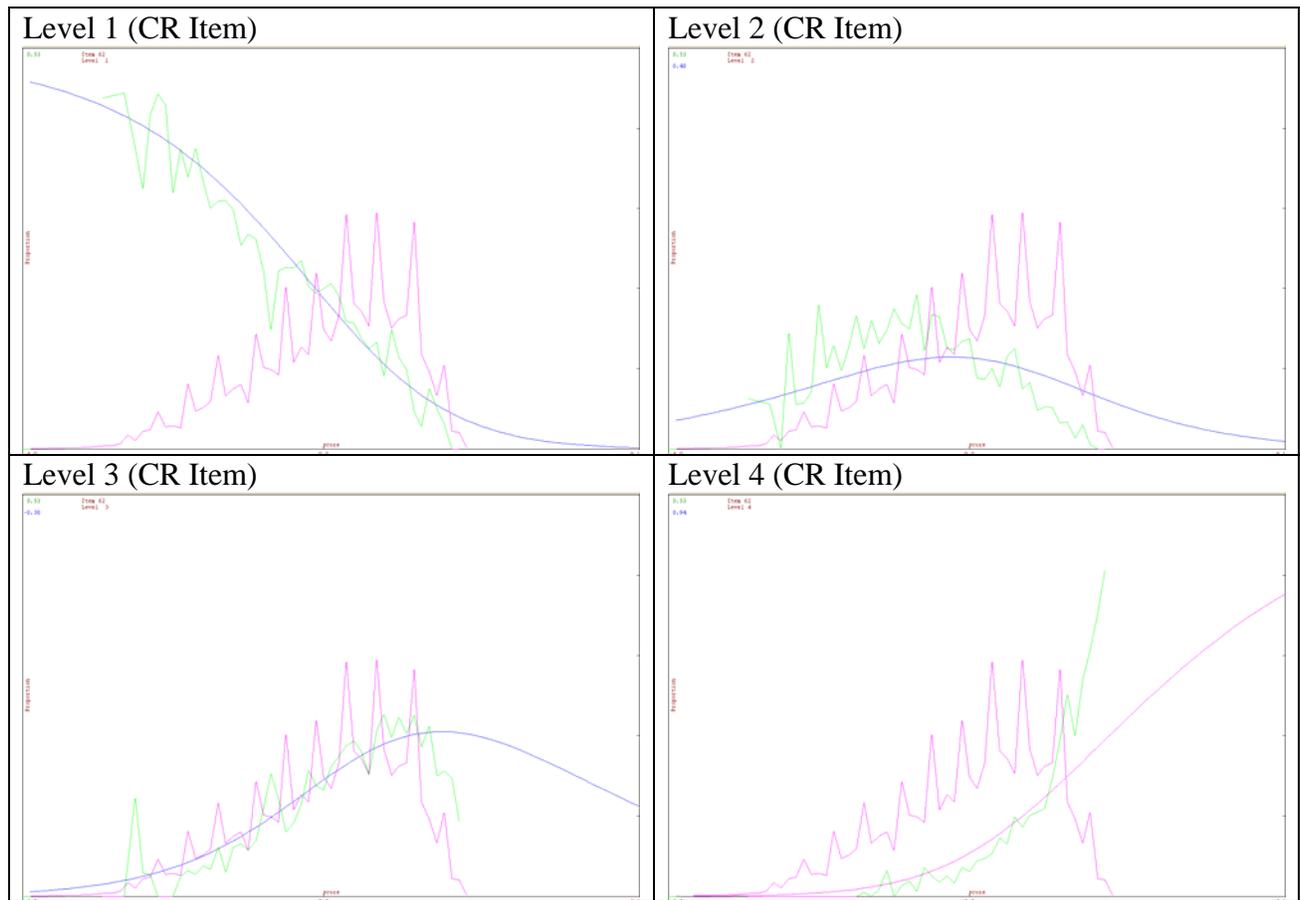


Figure 5. 9. Item characteristic curve for Grade 5 Mathematics, Item 17 (Z = 11.40)



**Figure 5. 10. Item characteristic curve for Grade 5 Mathematics, Item 35 (Z = 14.28)**

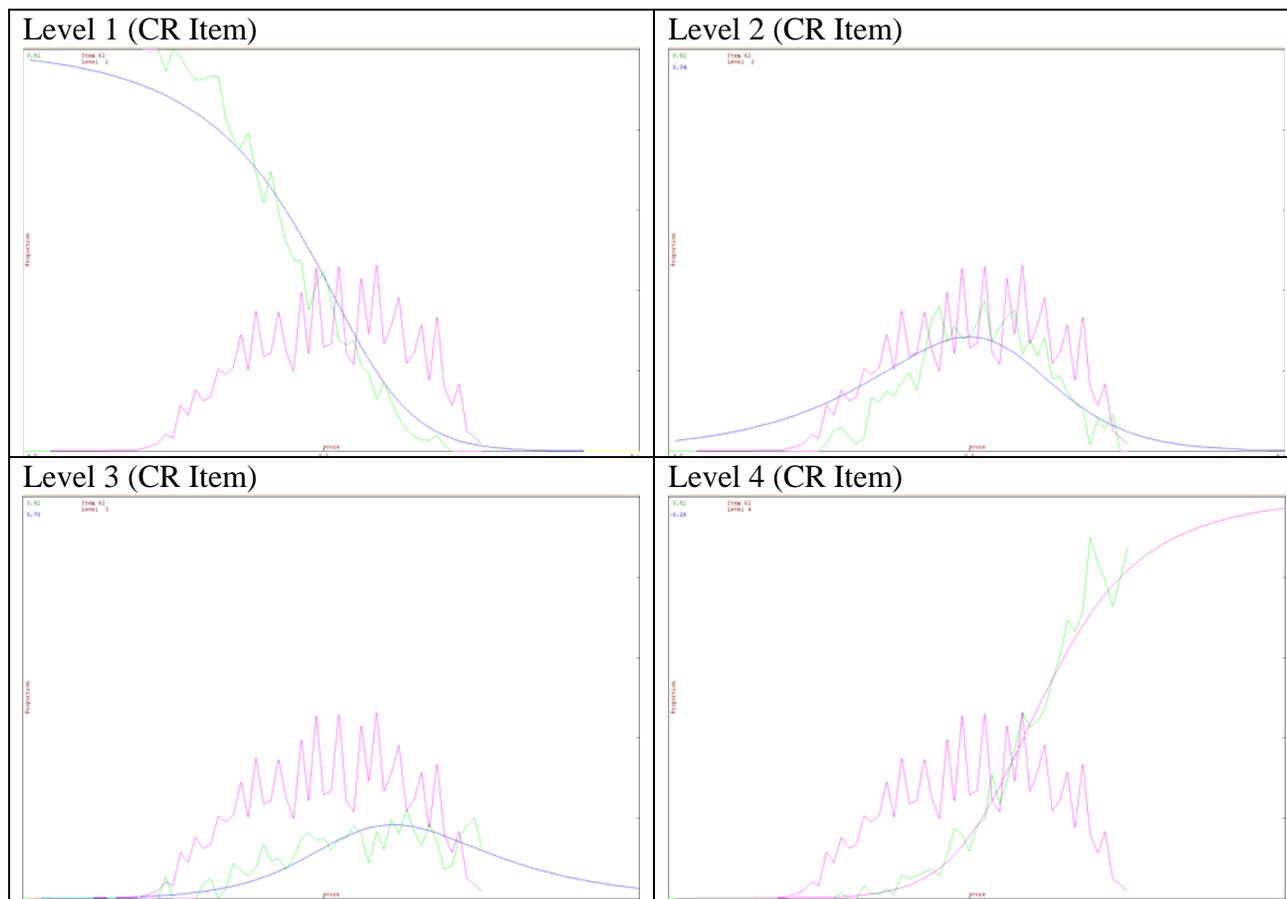


Figure 5. 11. Item characteristic curve for Grade 7 Mathematics, Item 35 ( $Z = 12.93$ )

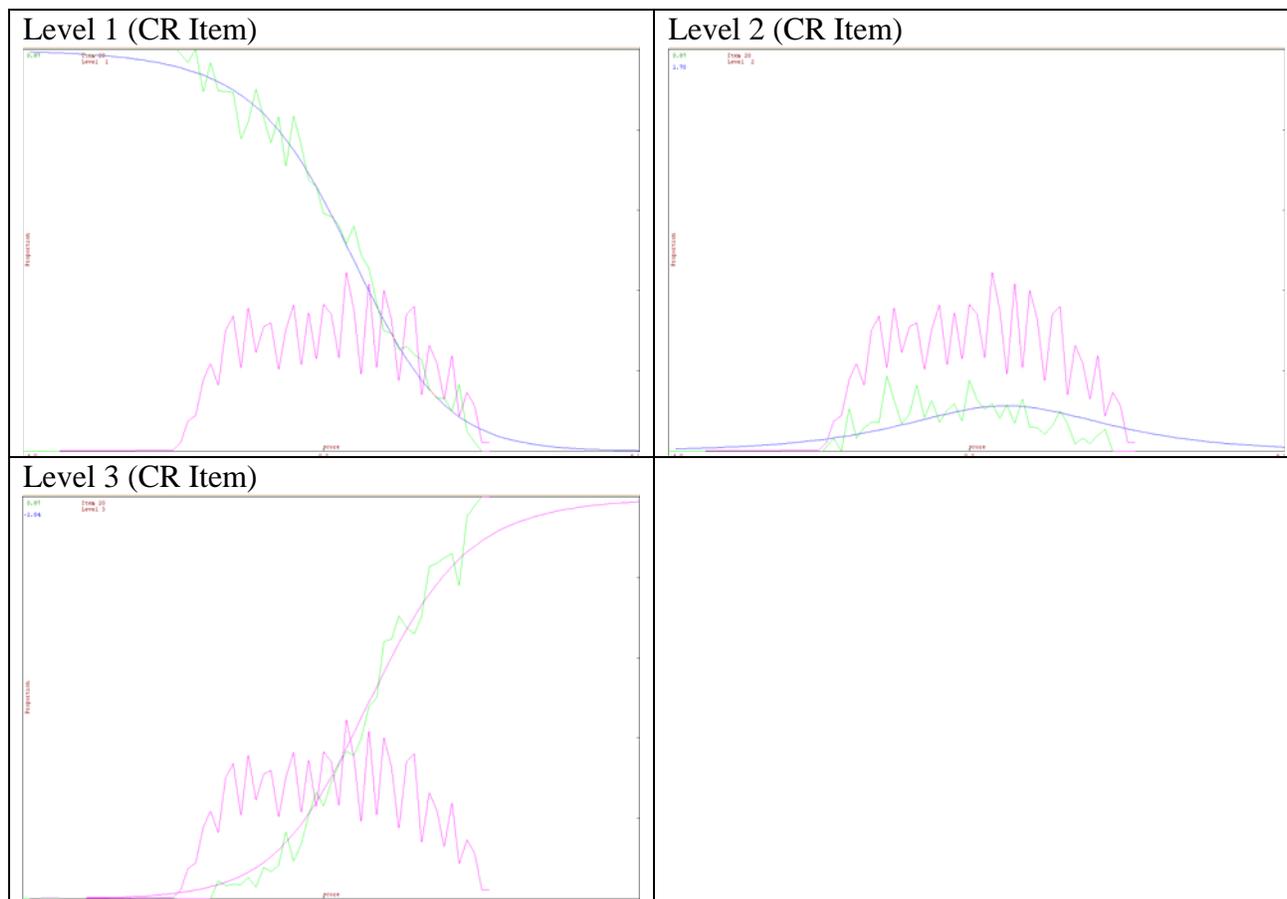


Figure 5. 12. Item characteristic curve for Grade 10 Mathematics, Item 20 ( $Z = 12.83$ )

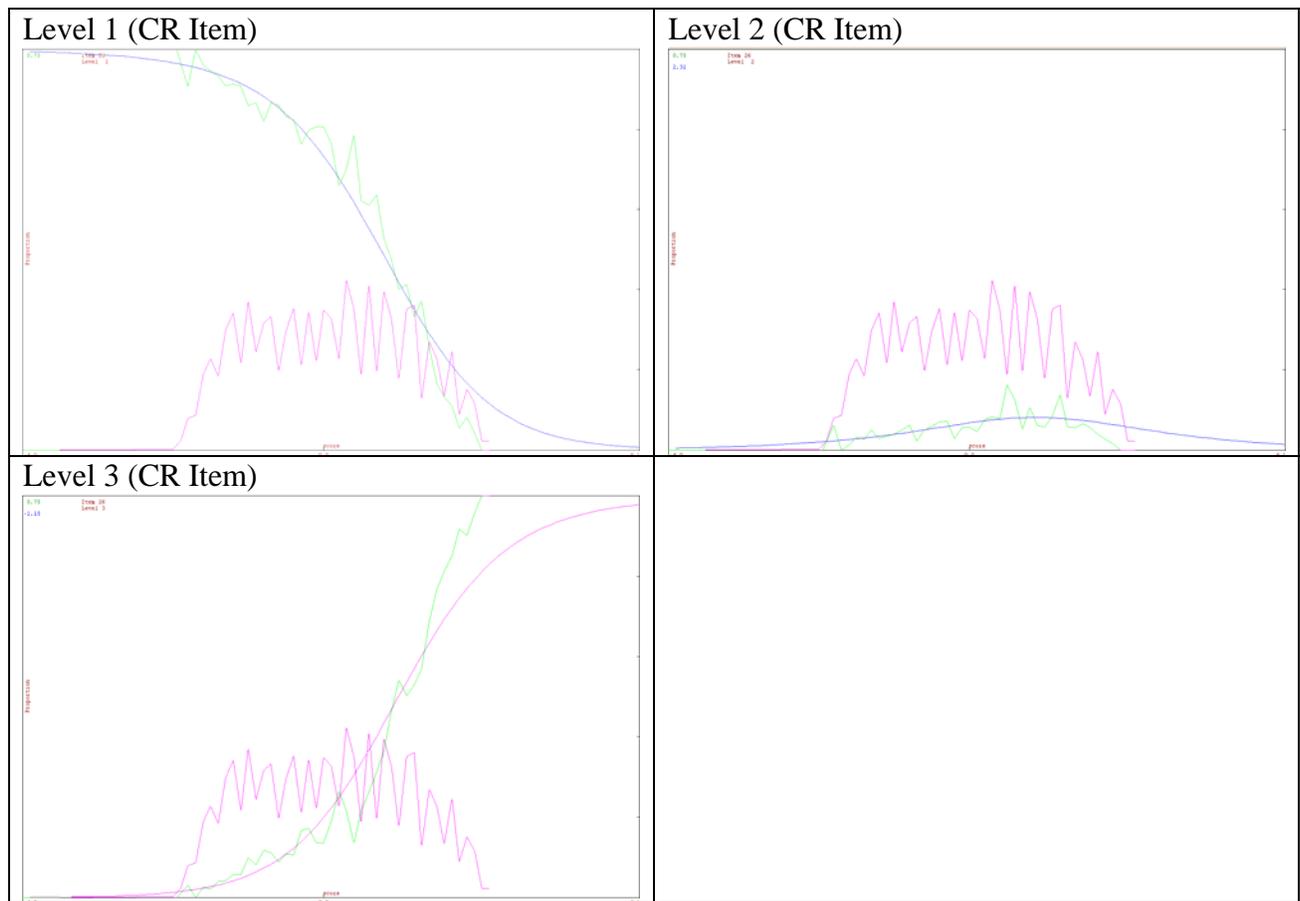
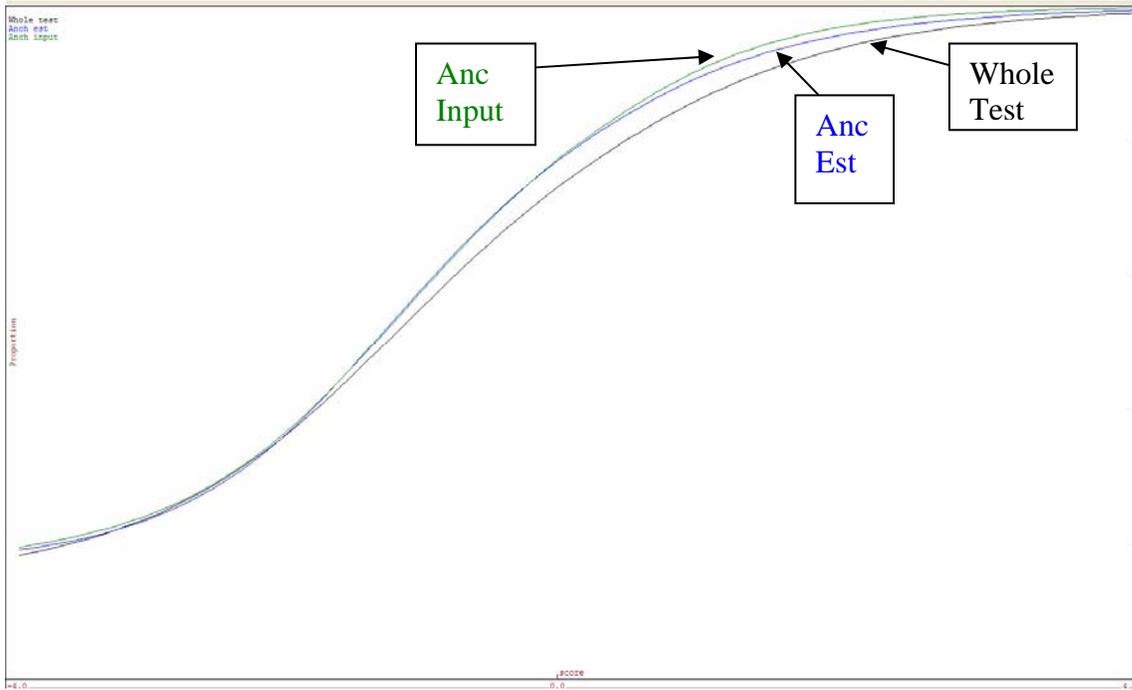
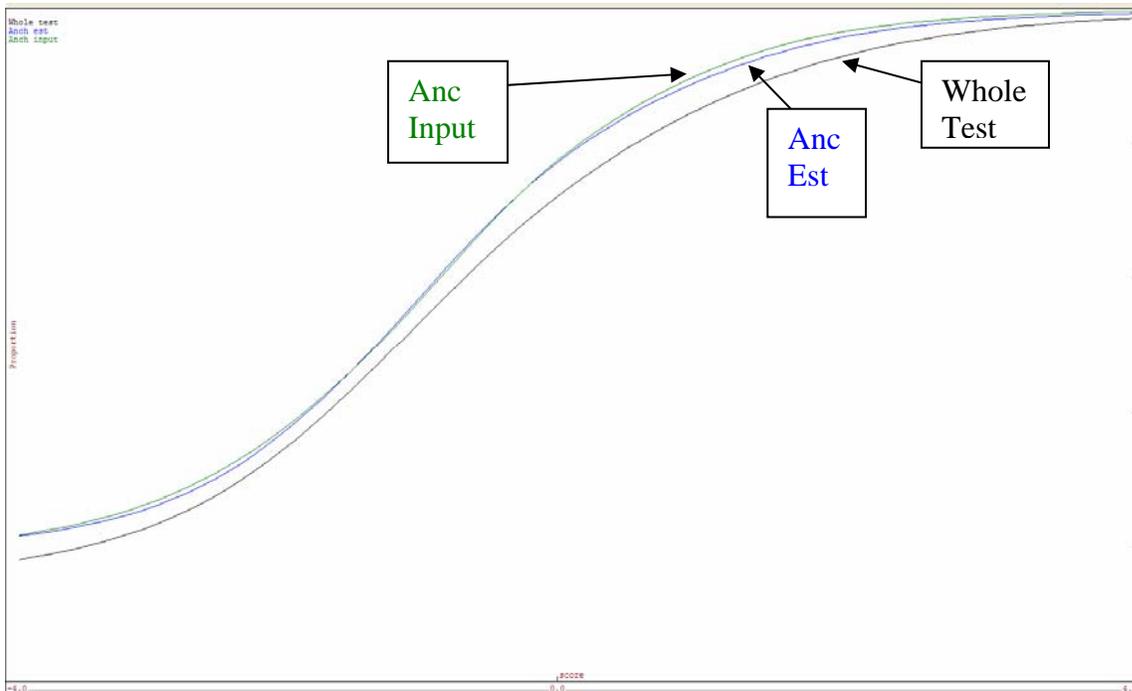


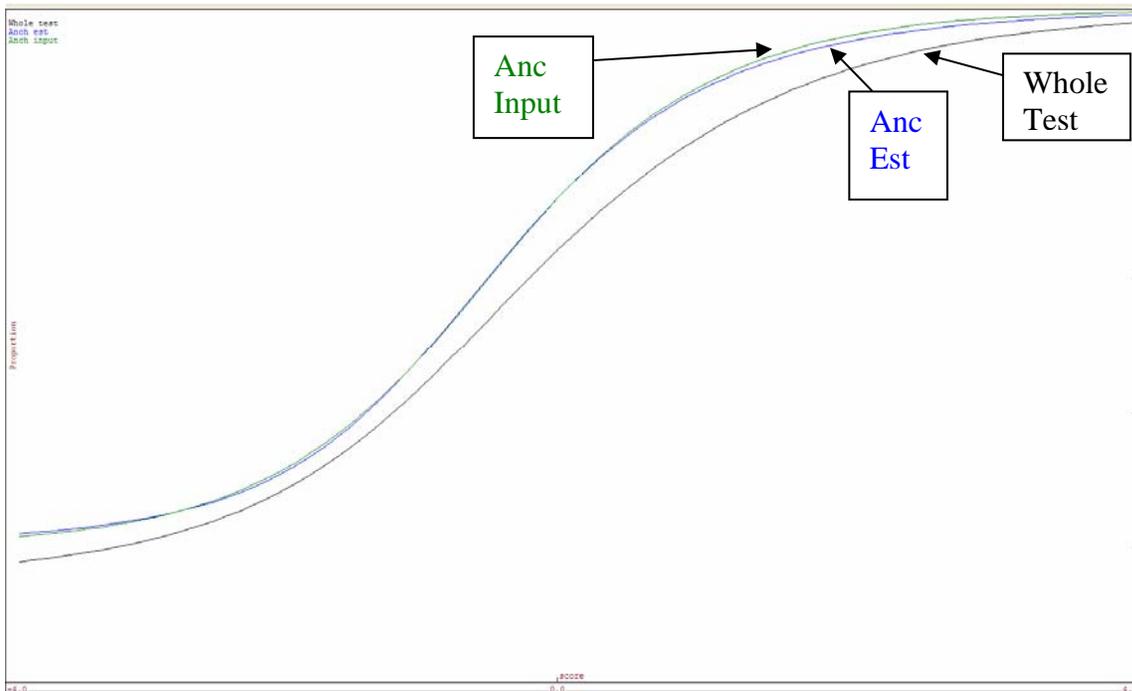
Figure 5.13. Item characteristic curve for Grade 10 Mathematics, Item 26 ( $Z = 10.78$ )



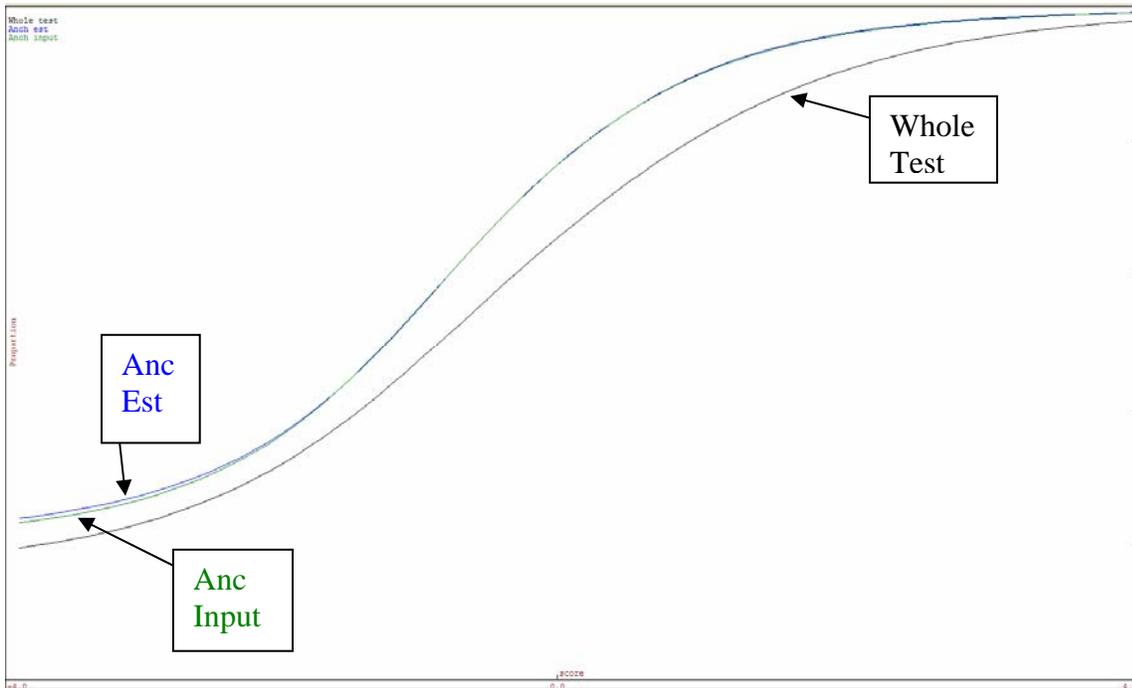
**Figure 5.14. Communication Arts, Grade 3 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



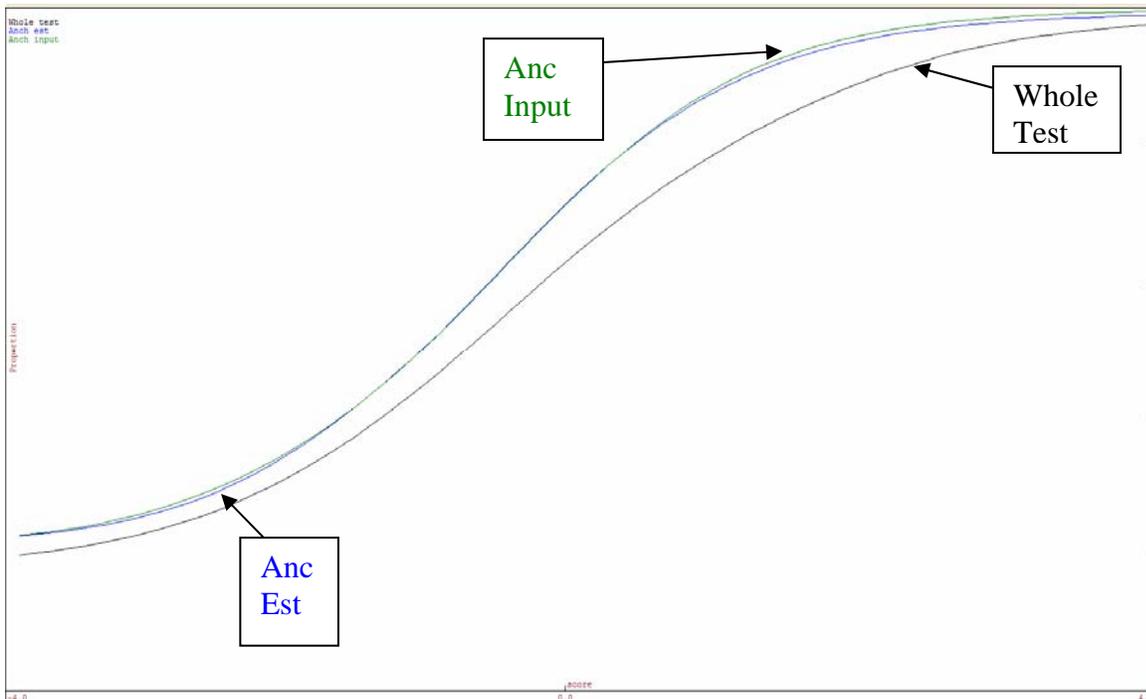
**Figure 5.15. Communication Arts, Grade 4 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



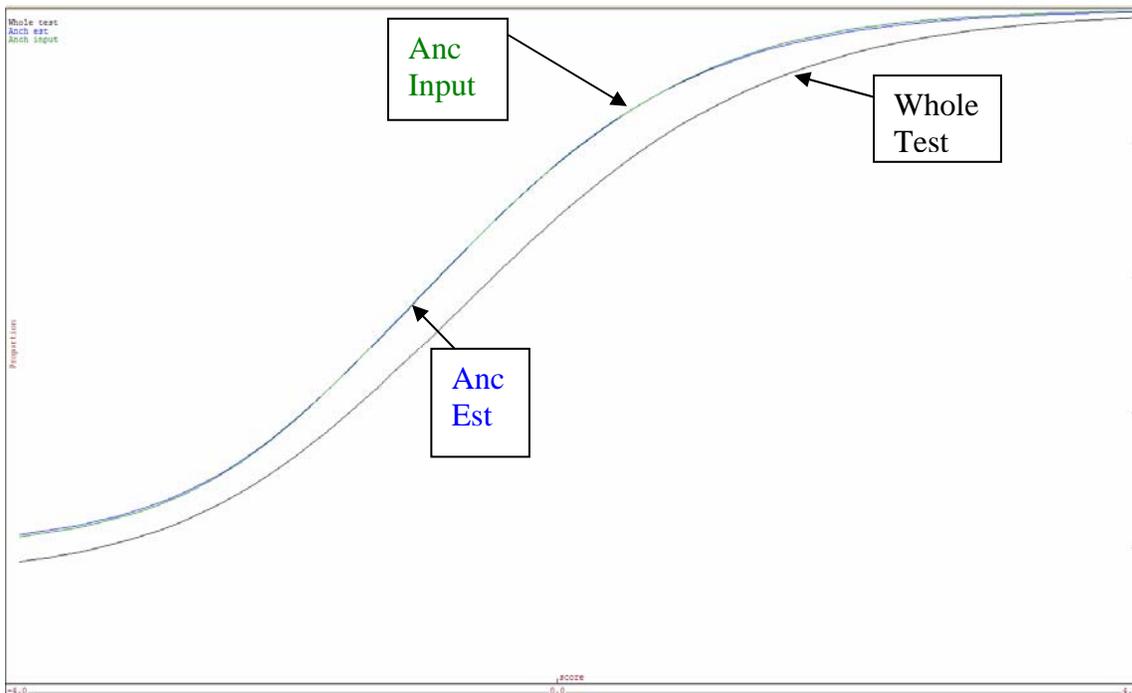
**Figure 5. 16. Communication Arts, Grade 5 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



**Figure 5. 17. Communication Arts, Grade 6 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



**Figure 5. 18. Communication Arts, Grade 7 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



**Figure 5. 19. Communication Arts, Grade 8 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**

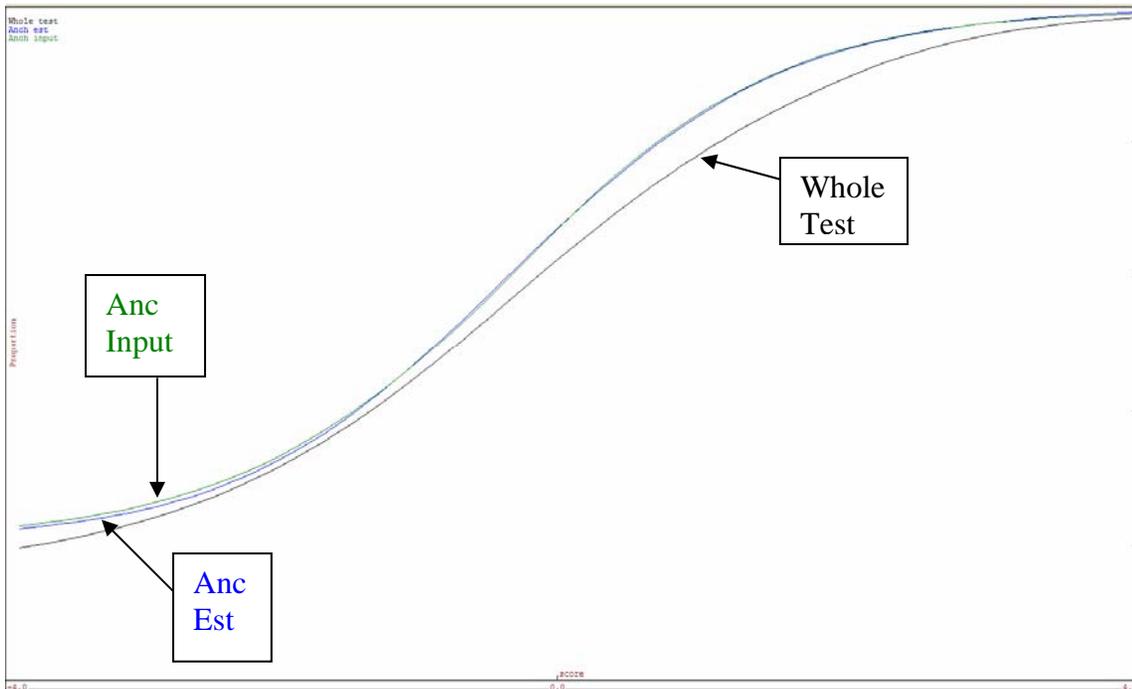


Figure 5. 20. Communication Arts, Grade 11 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items

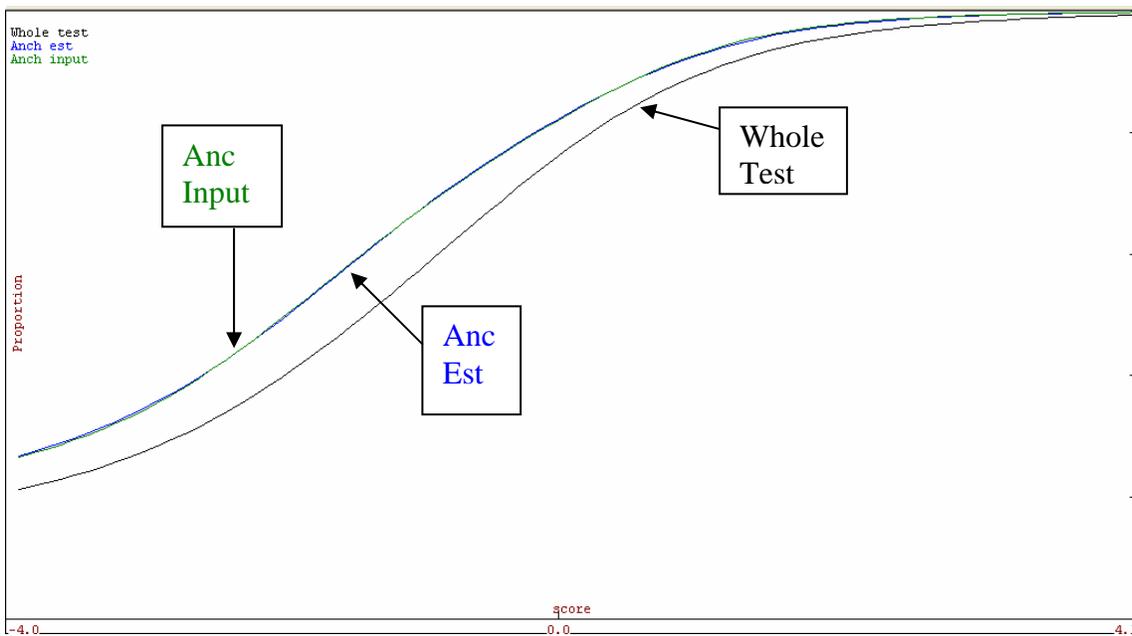
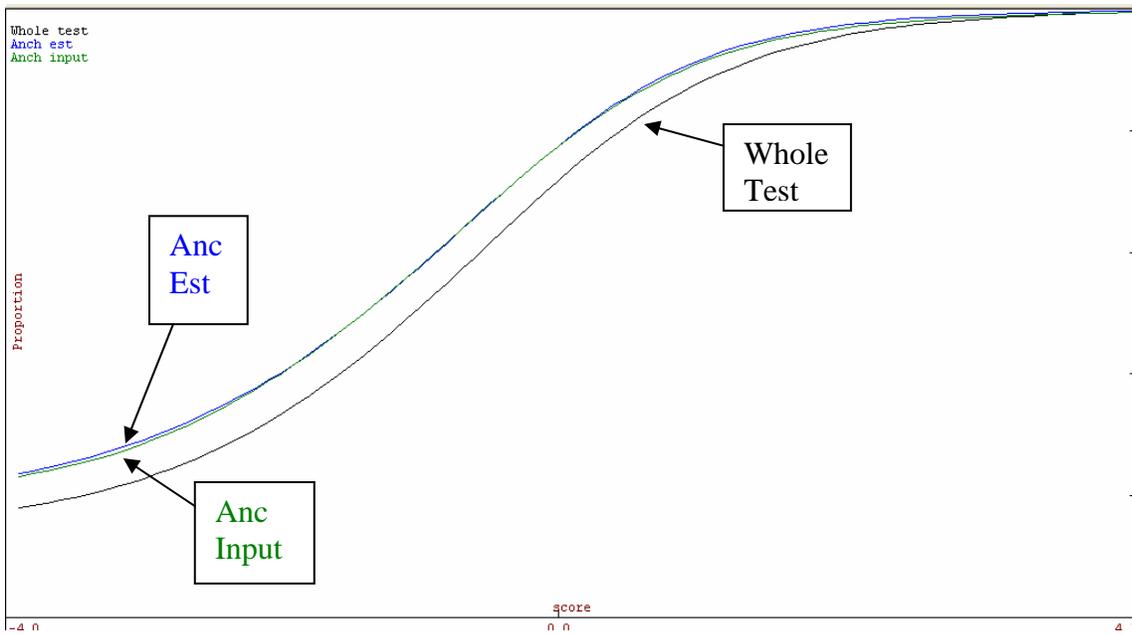
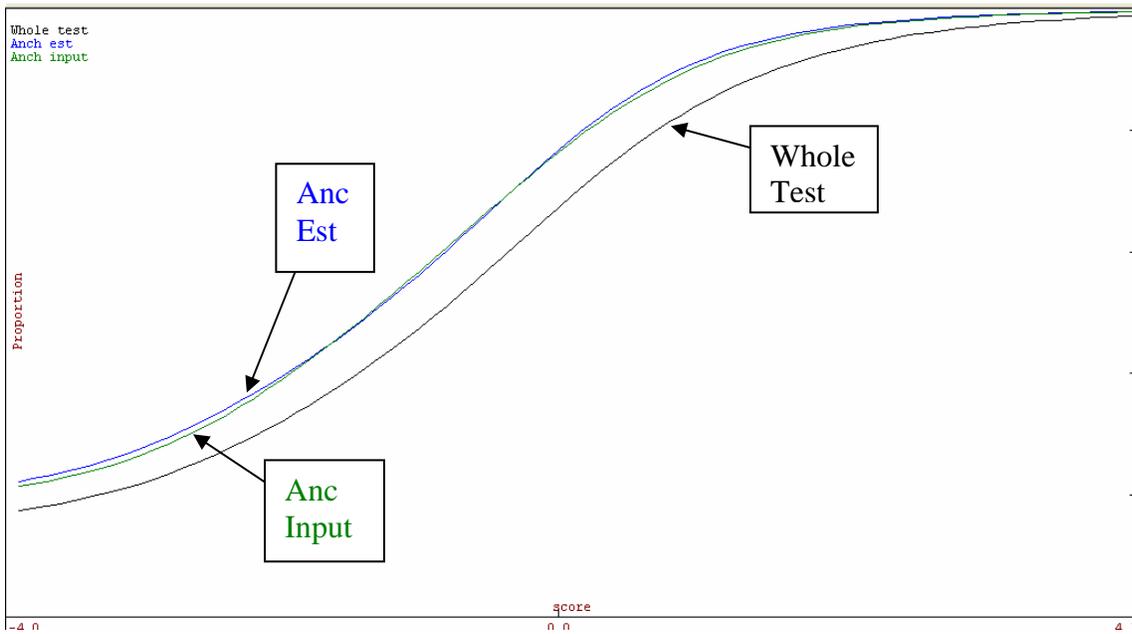


Figure 5. 21. Mathematics, Grade 3 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items



**Figure 5. 22. Mathematics, Grade 4 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



**Figure 5. 23. Mathematics, Grade 5 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**

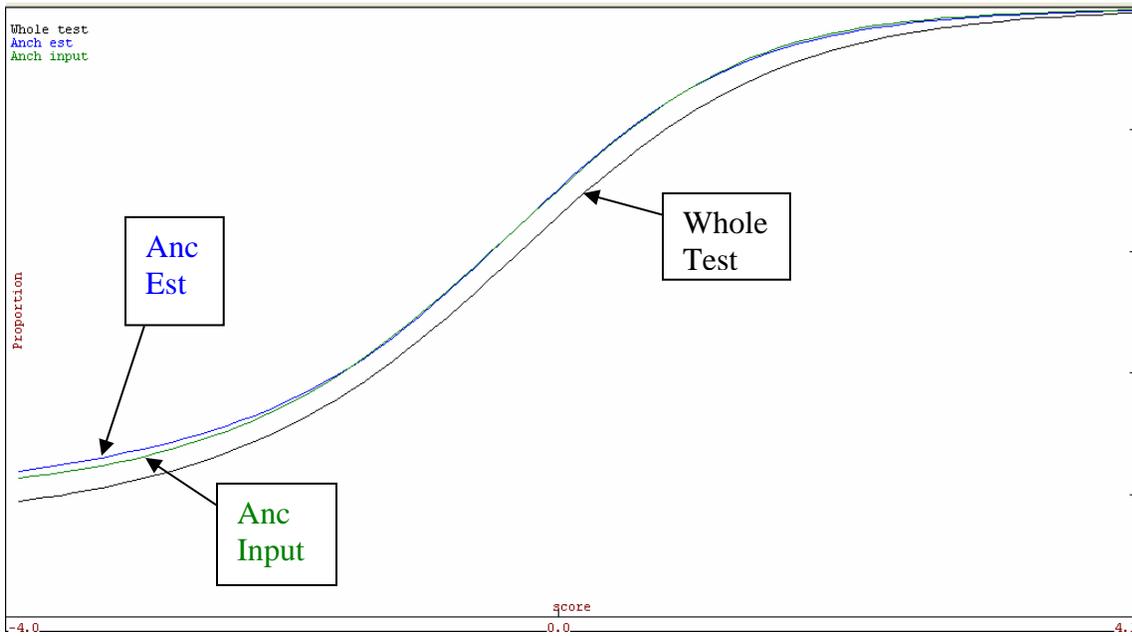


Figure 5. 24. Mathematics, Grade 6 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items

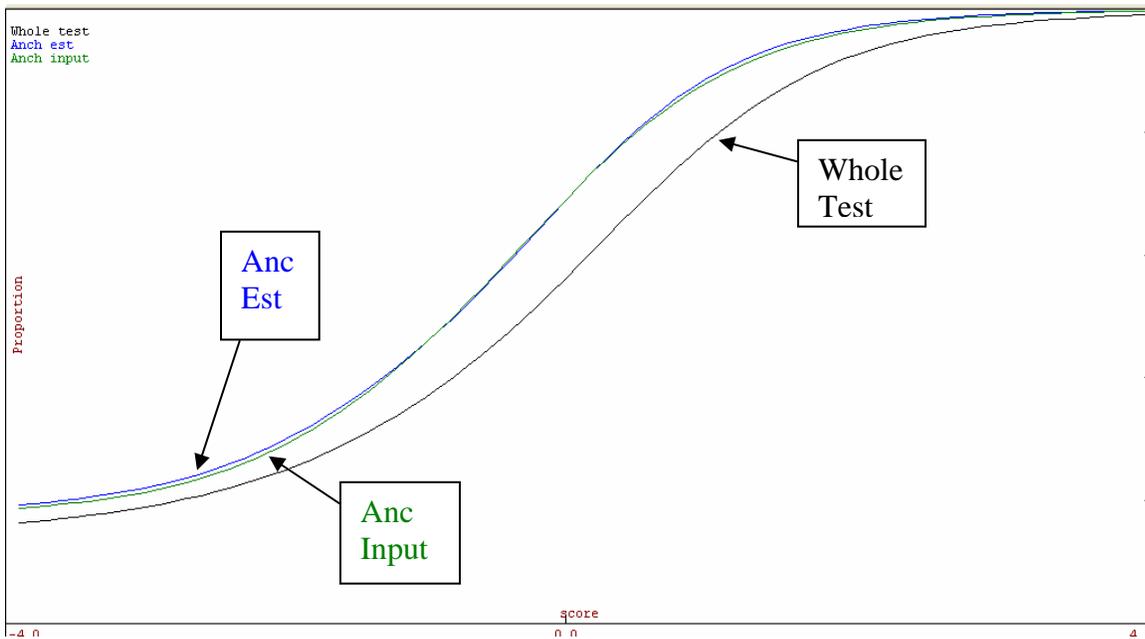
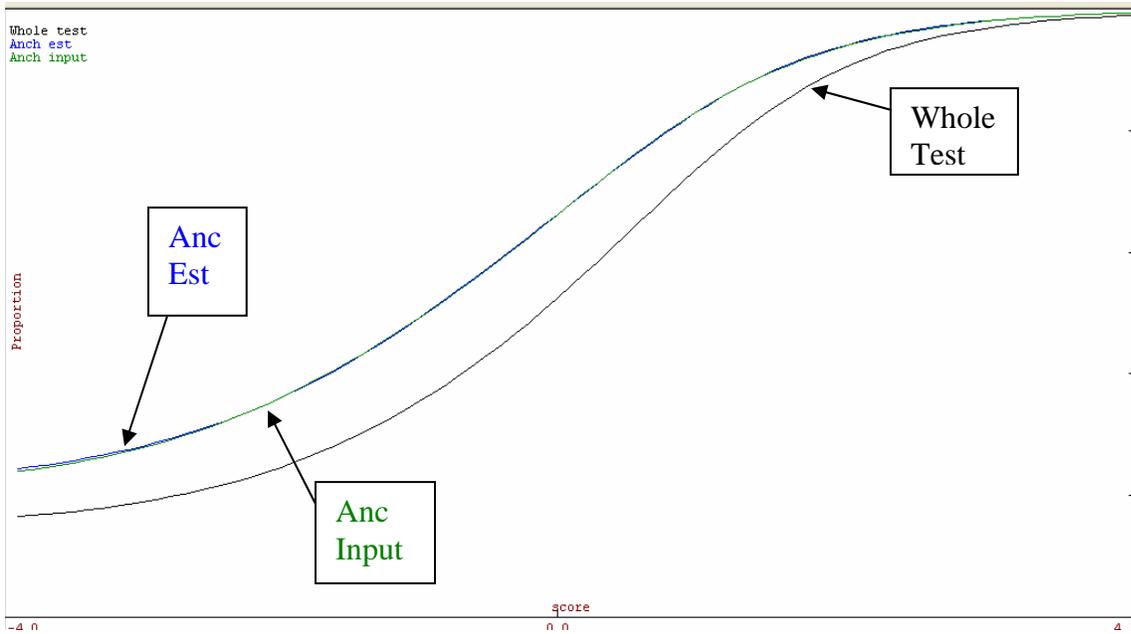
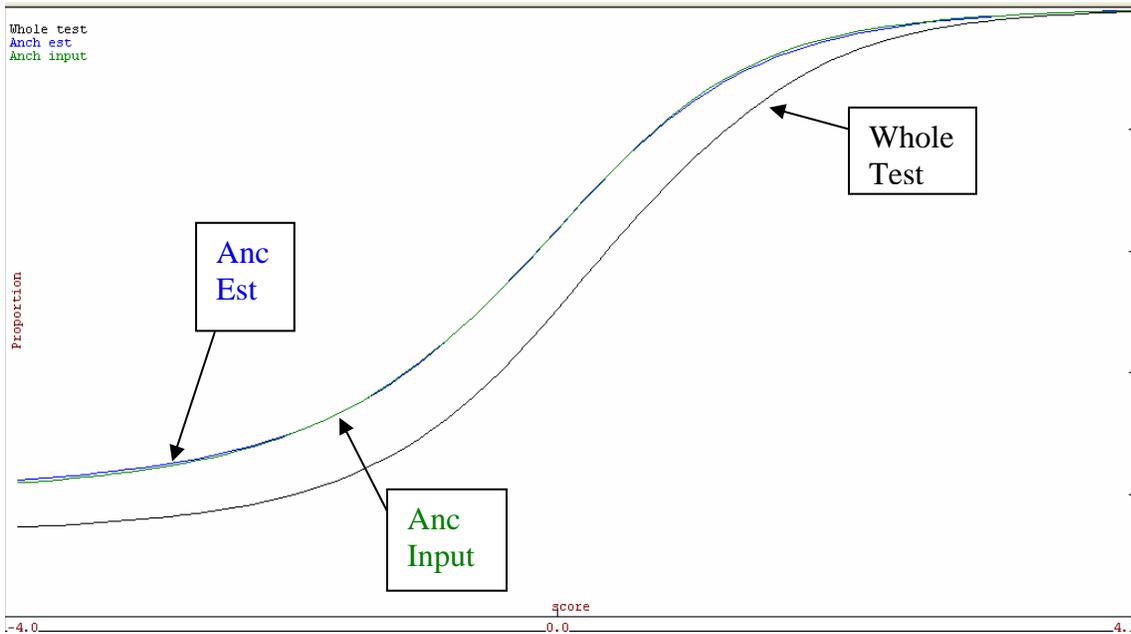


Figure 5. 25. Mathematics, Grade 7 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items



**Figure 5. 26. Mathematics, Grade 8 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**



**Figure 5. 27. Mathematics, Grade 10 Test Characteristic Curves (TCC) for the whole test, for the inputted anchor items, and for the estimated anchor items**

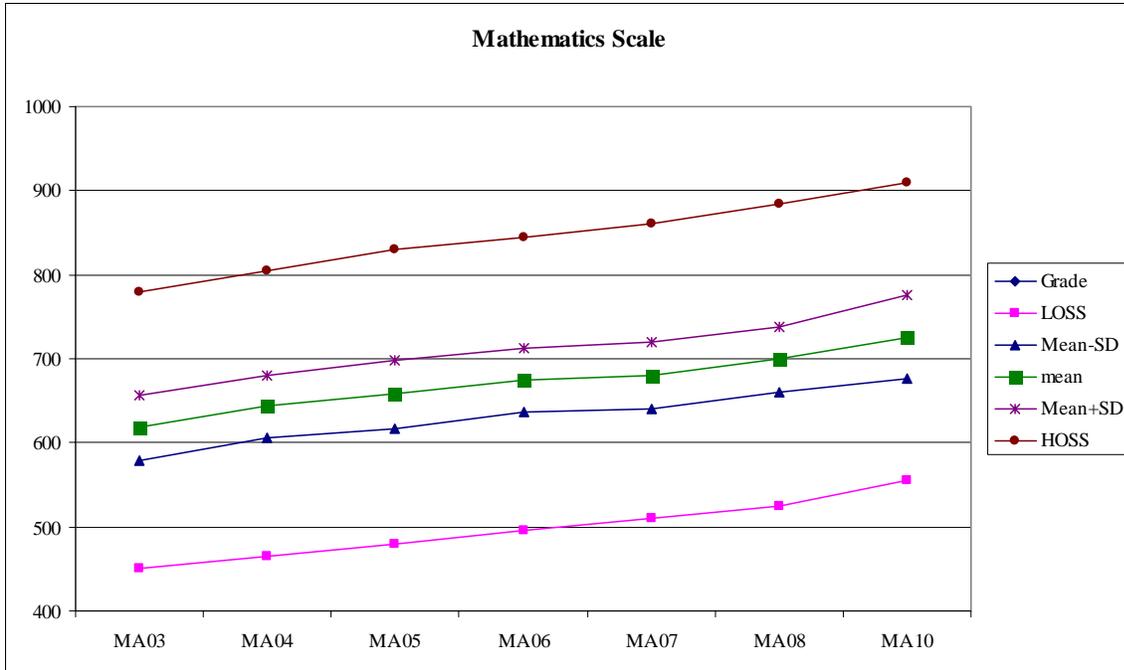


Figure 5.28 Selected Scale Scores Plotted by Grade Level, Mathematics MAP 2006

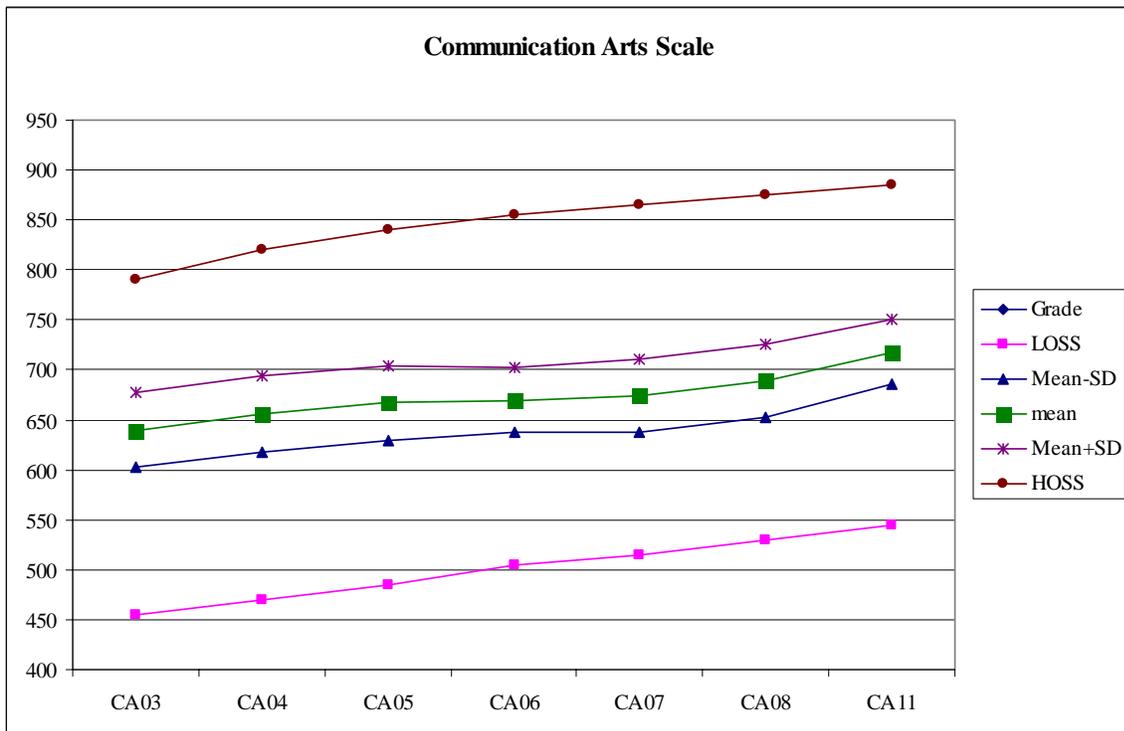


Figure 5.29 Selected Scale Scores Plotted by Grade Level, Communication Arts MAP

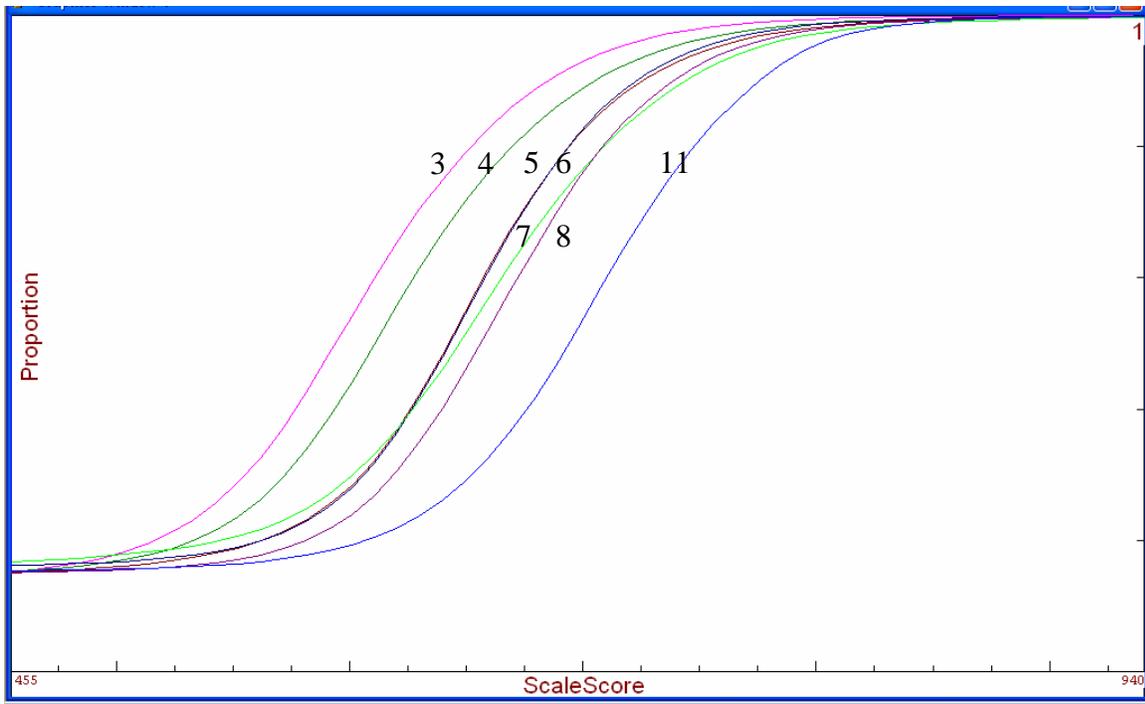


Figure 5.30. 2006 Communication Arts Test Characteristic Curves (TCC) by grade

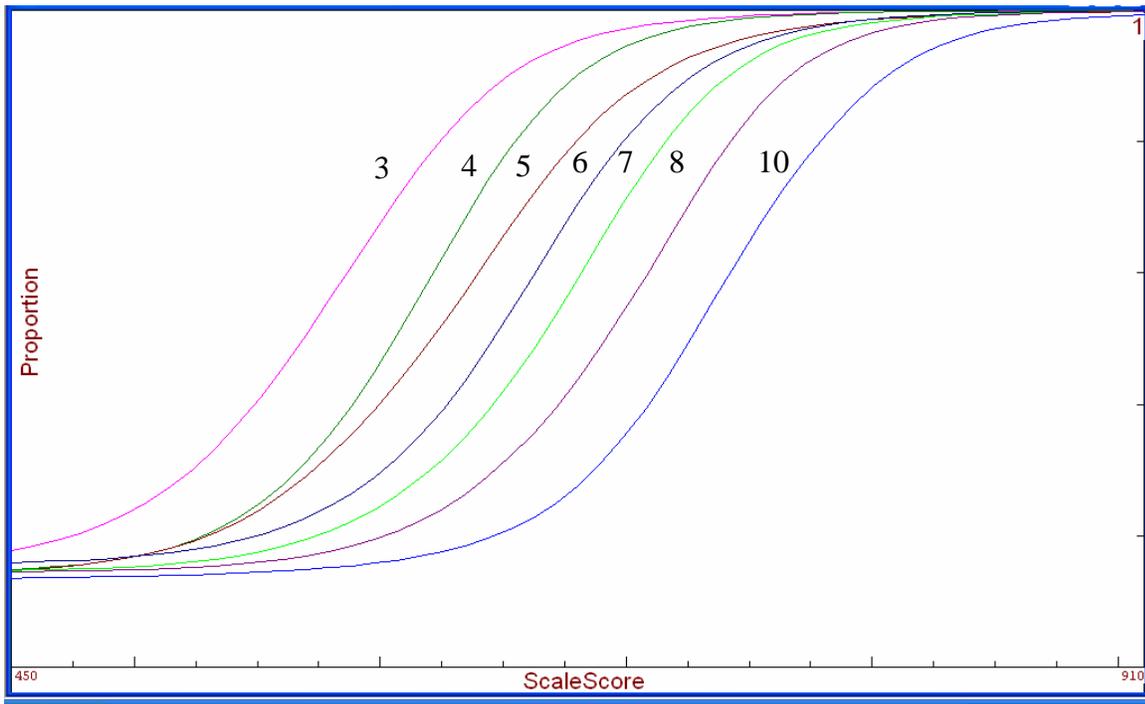


Figure 5.31. 2006 Mathematics Test Characteristic Curves (TCC) by grade

## Part 6: 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 6.1 and 6.2 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 means and standard deviations of the scale scores, 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 10.

The mean *p-value* is the average of all item *p-values* of a specific grade/content area. The mean item-total correlation 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 6.3 through 6.16 present the item statistics for each item by grade/content area. The tables include item booklet number and part (if applicable), test session, number of score points, location (loc), information (info), *p-value*, and item-total correlation ( $R_{it}$ ) for each item by grade/content area. The constructed-response (CR) items appear in the tables first, followed by the multiple-choice (MC) items.

*Location and Information:* IRT location refers to the scale score at which the maximum information is provided by an item. In classical testing theory, ability level and item difficulty are confounded in *p-values*; therefore, the *p-values* for the test items do not reveal the relative difficulty of the test items. IRT analyses allow one to separate more clearly ability level and item difficulty. The IRT locations associated with different items can be compared to evaluate their relative difficulty. Moreover, IRT location can be used to help evaluate whether the test items cover a range of difficulty levels on the scale score metric. The IRT item location indicates the scale score at which the item is measuring the most accurately; the higher the location, the more difficult the item. This scale score value for a 2PPC item is comparable in meaning to the IRT item location for a 3PL item (Lord 1980). To obtain an IRT location for each item, the item function for each item is first defined, and then the maximum amount of information provided by an item for a given ability level is computed. The scale score at which the maximum information is obtained is the item location. The maximum amount of information provided by the item at this particular scale score location is presented as IRT item information.

*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 rather

than on purely statistical criteria. As demonstrated in Tables 6.1 and 6.2, the average *p-values* associated with the MAP range from .56 to .77.

Because it is desirable for a test to measure well throughout the range of skills present for a given grade, the range of *p-values* and not just the average *p-value* must be examined. That is, it is important that the items measure the performance of both low- and high-scoring students, as well as students in the center of the distribution. Having a range of item difficulties also helps to prevent floor and/or ceiling effects, which occur if large numbers of students have scores at the minimum or maximum possible scores. The individual *p-values* (see Tables 6.3–6.16) on the MAP ranged from 0.10s to the 0.90s. The broad range of *p-values* indicates that the items span a range of difficulties. In other words, the items measure well throughout the range of skills for a given grade, which supports the accuracy of the MAP test scores.

*Item-Total Correlations:* An item-total correlation is the correlation between an item 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 majority of the items in the MAP had item-test correlations above this threshold.

**Table 6. 1: MAP Means, Standard Deviations for Scale/Raw Scores, P-Values, Item-Total Correlation: Communication Arts 2006**

	Mean Scale Score (SD)	Mean Raw Score (SD)	Mean P-Value (SD)	Mean Item-Total Correlation (SD)	Total Items	Total Points
<b>Grade 3</b>	639.86 (36.84)	50.15 (10.95)	0.7129 (0.1645)	0.4233 (0.0902)	57	74
<b>Grade 4</b>	654.55 (38.56)	47.15 (10.69)	0.7216 (0.1728)	0.4192 (0.1036)	56	66
<b>Grade 5</b>	668.18 (37.09)	42.48 (11.41)	0.6550 (0.0991)	0.4181 (0.0991)	56	67
<b>Grade 6</b>	666.85 (33.70)	42.61 (11.46)	0.6878 (0.1667)	0.4013 (0.1005)	56	66
<b>Grade 7</b>	671.63 (37.06)	44.79 (12.09)	0.6566 (0.1915)	0.4013 (0.1095)	60	72
<b>Grade 8</b>	686.85 (37.87)	46.11 (12.64)	0.6830 (0.1686)	0.4067 (0.1061)	60	69
<b>Grade 11</b>	716.69 (31.42)	46.47 (11.98)	0.6487 (0.1996)	0.3886 (0.1096)	61	74

**Table 6. 2: MAP Means, Standard Deviations for Scale/Raw Scores, P-Values, Item-Total Correlation: Mathematics 2006**

	Mean Scale Score (SD)	Mean Raw Score (SD)	Mean P-Value (SD)	Mean Item-Total Correlation (SD)	Total Items	Total Points
<b>Grade 3</b>	621.59 (39.11)	50.80 (10.65)	0.7728 (0.1435)	0.3800 (0.1061)	60	67
<b>Grade 4</b>	643.88 (37.07)	55.17 (13.66)	0.7241 (0.1438)	0.4195 (0.0932)	65	77
<b>Grade 5</b>	660.06 (39.99)	49.09 (12.09)	0.7017 (0.1685)	0.4159 (0.0992)	62	71
<b>Grade 6</b>	673.30 (39.80)	45.00 (12.84)	0.6637 (0.1533)	0.4129 (0.0971)	61	69
<b>Grade 7</b>	675.38 (41.27)	40.80 (14.01)	0.6149 (0.1815)	0.4271 (0.1038)	62	71
<b>Grade 8</b>	697.73 (40.37)	42.07 (14.42)	0.5977 (0.1876)	0.4107 (0.0982)	64	76
<b>Grade 10</b>	724.46 (51.18)	38.92 (16.54)	0.5557 (0.1605)	0.4637 (0.1205)	61	75

**Table 6. 3: Item Statistics: Communication Arts, Grade 3**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	620	0.266	0.597	0.502
	4		1	2	613	0.291	0.643	0.506
	5		1	2	618	0.327	0.627	0.519
	6	A	1	2	602	0.335	0.715	0.524
	6	B	1	2	552	0.743	0.939	0.458
	1		2	4	601	0.399	0.699	0.574
	22		3	3	548	0.324	0.754	0.459
	23		3	2	586	0.284	0.767	0.468
	24		3	1	576	0.472	0.883	0.462
	43		3	2	602	0.681	0.764	0.601
	MC	1		1	1	646	0.249	0.636
2			1	1	666	0.178	0.496	0.331
1			3	1	600	0.194	0.819	0.396
2			3	1	592	0.700	0.879	0.531
3			3	1	647	0.320	0.578	0.417
4			3	1	582	0.377	0.909	0.431
5			3	1	615	0.121	0.609	0.355
6			3	1	588	0.422	0.881	0.464
7			3	1	561	0.342	0.957	0.331
8			3	1	572	0.391	0.881	0.438
9			3	1	606	0.718	0.857	0.519
10			3	1	554	0.223	0.947	0.313
11			3	1	605	0.831	0.851	0.559
12			3	1	621	0.199	0.848	0.362
13			3	1	671	0.225	0.560	0.304
14			3	1	608	0.405	0.821	0.467
15			3	1	666	0.318	0.421	0.382
16			3	1	664	0.735	0.484	0.358
17			3	1	636	0.115	0.623	0.335
18			3	1	586	0.525	0.909	0.467
19			3	1	616	0.362	0.776	0.466
20			3	1	659	0.491	0.474	0.415
21			3	1	608	0.390	0.791	0.486
25			3	1	659	0.175	0.542	0.361
26			3	1	649	0.075	0.605	0.322
27			3	1	677	0.122	0.447	0.292
28			3	1	623	0.645	0.731	0.559
29			3	1	574	0.434	0.878	0.463
30			3	1	633	0.495	0.662	0.487
31			3	1	577	0.548	0.934	0.449
32			3	1	611	0.247	0.777	0.446
33			3	1	679	0.219	0.504	0.272
34			3	1	581	1.044	0.936	0.510
35			3	1	599	0.173	0.706	0.395
36			3	1	674	0.201	0.475	0.274
37			3	1	612	0.674	0.829	0.534
38			3	1	609	0.453	0.797	0.484
39			3	1	625	0.215	0.713	0.408
40			3	1	587	0.949	0.919	0.517
41			3	1	596	0.644	0.877	0.511
42			3	1	678	0.198	0.430	0.304
44			3	1	602	0.086	0.777	0.303
45			3	1	598	0.325	0.850	0.448
46			3	1	659	0.263	0.553	0.365
47			3	1	622	0.098	0.699	0.336
48			3	1	687	0.091	0.441	0.258
49			3	1	697	0.215	0.385	0.239

**Table 6. 4: Item Statistics: Communication Arts, Grade 4**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	634	0.286	0.607	0.505
	4		1	2	668	0.273	0.439	0.487
	5		1	2	663	0.473	0.453	0.555
	6	A	1	2	639	0.350	0.589	0.543
	6	B	1	2	574	0.293	0.859	0.428
	19		2	2	600	0.199	0.733	0.439
	20		2	3	602	1.477	0.861	0.666
	21		2	2	676	0.235	0.389	0.439
	43		2	2	587	0.970	0.910	0.539
	MC	1		1	1	671	0.241	0.600
2			1	1	647	0.493	0.734	0.483
1			2	1	574	0.343	0.958	0.336
2			2	1	574	0.489	0.967	0.352
3			2	1	689	0.359	0.444	0.348
4			2	1	620	0.377	0.818	0.473
5			2	1	584	1.114	0.966	0.426
6			2	1	607	0.302	0.878	0.431
7			2	1	612	0.170	0.722	0.394
8			2	1	599	0.343	0.905	0.430
9			2	1	568	0.145	0.934	0.278
10			2	1	609	0.555	0.875	0.509
11			2	1	589	0.621	0.940	0.437
12			2	1	660	0.100	0.628	0.304
13			2	1	611	0.721	0.889	0.507
14			2	1	658	0.570	0.626	0.496
15			2	1	595	0.297	0.910	0.399
16			2	1	665	0.290	0.673	0.382
17			2	1	606	0.556	0.893	0.497
18			2	1	692	0.366	0.409	0.324
22			2	1	622	0.180	0.800	0.390
23			2	1	616	0.335	0.830	0.465
24			2	1	623	0.178	0.797	0.375
25			2	1	598	0.056	0.814	0.261
26			2	1	703	0.441	0.379	0.267
27			2	1	640	0.228	0.757	0.442
28			2	1	626	0.385	0.781	0.505
29			2	1	619	0.787	0.864	0.549
30			2	1	647	0.188	0.698	0.383
31			2	1	628	1.371	0.821	0.605
32			2	1	644	0.796	0.739	0.519
33			2	1	643	0.480	0.717	0.511
34			2	1	673	0.184	0.488	0.376
35			2	1	602	0.232	0.879	0.389
36			2	1	676	0.063	0.574	0.276
37			2	1	630	0.649	0.794	0.550
38			2	1	691	0.031	0.534	0.213
39			2	1	685	0.172	0.481	0.336
40			2	1	614	0.737	0.895	0.520
41			2	1	618	0.636	0.895	0.483
42			2	1	651	0.282	0.664	0.446
44			2	1	599	0.209	0.885	0.382
45			2	1	705	0.027	0.529	0.192
46			2	1	670	0.161	0.559	0.365
47			2	1	641	0.214	0.791	0.397
48			2	1	621	0.138	0.792	0.352
49			2	1	735	0.280	0.289	0.153

**Table 6. 5: Item Statistics: Communication Arts, Grade 5**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	618	0.784	0.836	0.548
	4		1	2	622	0.306	0.736	0.471
	5		1	2	690	0.160	0.423	0.401
	6	A	1	2	632	0.182	0.661	0.412
	6	B	1	1	687	0.326	0.369	0.439
	17		2	2	655	0.484	0.580	0.556
	18		2	2	683	0.275	0.432	0.477
	42		2	3	684	1.000	0.400	0.645
	43		2	2	627	0.329	0.714	0.474
	MC	1		1	1	669	0.191	0.628
2			1	1	724	0.310	0.337	0.252
1			2	1	650	0.487	0.754	0.490
2			2	1	609	0.153	0.786	0.331
3			2	1	626	0.888	0.895	0.485
4			2	1	626	0.855	0.885	0.490
5			2	1	651	0.664	0.750	0.531
6			2	1	666	0.328	0.663	0.429
7			2	1	649	0.908	0.801	0.532
8			2	1	733	0.167	0.390	0.221
9			2	1	592	0.057	0.749	0.232
10			2	1	663	0.170	0.637	0.373
11			2	1	674	0.679	0.643	0.450
12			2	1	637	0.452	0.840	0.458
13			2	1	617	0.552	0.916	0.421
14			2	1	667	0.346	0.620	0.446
15			2	1	646	0.548	0.798	0.504
16			2	1	700	0.250	0.437	0.334
19			2	1	645	0.235	0.767	0.406
20			2	1	623	0.225	0.856	0.361
21			2	1	655	0.208	0.710	0.386
22			2	1	647	0.234	0.754	0.407
23			2	1	652	0.852	0.733	0.557
24			2	1	628	0.137	0.813	0.334
25			2	1	648	0.237	0.752	0.405
26			2	1	666	0.175	0.654	0.376
27			2	1	662	0.996	0.639	0.586
28			2	1	671	0.281	0.641	0.414
29			2	1	656	0.566	0.761	0.488
30			2	1	640	0.705	0.818	0.513
31			2	1	665	0.092	0.628	0.307
32			2	1	670	0.281	0.603	0.435
33			2	1	637	0.588	0.856	0.474
34			2	1	655	0.796	0.752	0.532
35			2	1	658	0.681	0.763	0.499
36			2	1	662	0.457	0.695	0.475
37			2	1	651	0.557	0.778	0.485
38			2	1	684	0.733	0.508	0.491
39			2	1	705	0.101	0.495	0.270
40			2	1	698	0.362	0.490	0.315
41			2	1	702	0.162	0.483	0.306
44			2	1	634	0.154	0.798	0.350
45			2	1	697	0.335	0.485	0.359
46			2	1	704	0.113	0.503	0.280
47			2	1	674	0.161	0.638	0.352
48			2	1	714	0.274	0.411	0.281
49			2	1	734	0.025	0.498	0.187

**Table 6. 6: Item Statistics: Communication Arts, Grade 6**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	681	0.507	0.426	0.512
	4		1	2	697	0.549	0.302	0.473
	5		1	2	689	0.575	0.365	0.521
	6	A	1	2	669	0.849	0.513	0.599
	6	B	1	2	624	0.717	0.815	0.523
	17		2	2	663	0.139	0.525	0.345
	18		2	2	636	0.950	0.752	0.619
	19		2	2	712	0.372	0.246	0.402
	43		2	3	680	0.526	0.415	0.550
	MC	1		1	1	681	0.168	0.663
2			1	1	618	0.142	0.859	0.284
1			2	1	613	0.300	0.915	0.333
2			2	1	661	0.339	0.706	0.419
3			2	1	647	0.754	0.827	0.510
4			2	1	644	0.402	0.775	0.450
5			2	1	604	0.230	0.924	0.290
6			2	1	647	0.937	0.822	0.516
7			2	1	613	0.145	0.876	0.274
8			2	1	644	0.177	0.758	0.351
9			2	1	680	0.760	0.562	0.468
10			2	1	644	1.209	0.847	0.534
11			2	1	670	0.329	0.672	0.400
12			2	1	618	0.419	0.917	0.377
13			2	1	645	0.674	0.777	0.504
14			2	1	644	0.460	0.813	0.439
15			2	1	650	0.494	0.773	0.470
16			2	1	638	0.866	0.879	0.488
20			2	1	691	0.342	0.544	0.351
21			2	1	635	0.346	0.837	0.411
22			2	1	681	0.032	0.621	0.192
23			2	1	648	0.052	0.718	0.225
24			2	1	627	0.784	0.902	0.464
25			2	1	679	0.414	0.626	0.395
26			2	1	670	0.214	0.614	0.371
27			2	1	666	0.378	0.728	0.400
28			2	1	650	0.575	0.798	0.474
29			2	1	687	0.577	0.526	0.427
30			2	1	635	0.598	0.850	0.463
31			2	1	708	0.193	0.484	0.259
32			2	1	654	0.217	0.736	0.371
33			2	1	649	0.759	0.790	0.522
34			2	1	650	0.586	0.798	0.479
35			2	1	662	0.379	0.706	0.454
36			2	1	651	0.253	0.740	0.399
37			2	1	613	0.602	0.942	0.369
38			2	1	609	0.197	0.903	0.300
39			2	1	673	0.319	0.687	0.385
40			2	1	666	0.185	0.671	0.366
41			2	1	669	0.225	0.652	0.378
42			2	1	672	0.140	0.638	0.327
44			2	1	693	0.223	0.555	0.317
45			2	1	707	0.109	0.549	0.239
46			2	1	706	0.390	0.453	0.288
47			2	1	630	0.201	0.839	0.350
48			2	1	733	0.162	0.525	0.158
49			2	1	680	0.335	0.662	0.377

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

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	737	0.170	0.230	0.342
	4		1	2	706	0.279	0.334	0.454
	5	A	1	2	697	0.242	0.377	0.451
	5	B	1	2	613	0.521	0.856	0.487
	6		1	2	686	0.239	0.440	0.467
	1		2	4	609	0.680	0.707	0.583
	19		3	2	653	0.544	0.654	0.581
	20		3	3	680	0.332	0.402	0.522
	42		3	2	663	0.253	0.564	0.466
	MC	1		1	1	714	0.255	0.426
2			1	1	656	0.637	0.801	0.508
1			3	1	616	0.182	0.885	0.331
2			3	1	602	0.355	0.947	0.325
3			3	1	621	0.561	0.913	0.432
4			3	1	596	0.418	0.962	0.307
5			3	1	602	1.203	0.951	0.411
6			3	1	651	0.186	0.751	0.386
7			3	1	703	0.006	0.619	0.126
8			3	1	652	0.413	0.807	0.471
9			3	1	653	0.238	0.734	0.429
10			3	1	659	0.489	0.756	0.504
11			3	1	664	1.032	0.706	0.567
12			3	1	664	0.546	0.724	0.494
13			3	1	623	0.872	0.927	0.455
14			3	1	678	0.476	0.610	0.470
15			3	1	665	0.652	0.752	0.497
16			3	1	643	1.001	0.848	0.556
17			3	1	695	0.147	0.561	0.339
18			3	1	681	0.239	0.610	0.405
21			3	1	682	0.183	0.606	0.374
22			3	1	674	0.354	0.690	0.449
23			3	1	617	0.479	0.922	0.404
24			3	1	706	0.227	0.412	0.373
25			3	1	672	0.552	0.687	0.493
26			3	1	674	0.244	0.723	0.391
27			3	1	673	0.224	0.680	0.399
28			3	1	679	0.337	0.686	0.420
29			3	1	614	0.093	0.852	0.292
30			3	1	657	0.554	0.767	0.518
31			3	1	631	0.457	0.880	0.451
32			3	1	688	0.298	0.607	0.407
33			3	1	660	0.502	0.766	0.500
34			3	1	638	0.322	0.823	0.440
35			3	1	708	0.178	0.473	0.321
36			3	1	703	0.771	0.428	0.430
37			3	1	700	0.412	0.446	0.427
38			3	1	712	0.189	0.482	0.313
39			3	1	676	0.146	0.654	0.360
40			3	1	656	0.887	0.802	0.548
41			3	1	658	0.220	0.749	0.413
43			3	1	632	0.206	0.844	0.375
44			3	1	609	0.023	0.784	0.178
45			3	1	641	0.127	0.787	0.338
46			3	1	734	0.201	0.514	0.199
47			3	1	697	0.298	0.555	0.373
48			3	1	662	0.302	0.739	0.448
49			3	1	710	0.454	0.524	0.299
50			3	1	747	0.646	0.204	0.142
51			3	1	753	1.019	0.384	0.101
52			3	1	730	0.776	0.330	0.217

**Table 6. 8: Item Statistics: Communication Arts, Grade 8**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	646	0.478	0.763	0.531
	4		1	2	652	0.630	0.760	0.562
	5		1	2	660	0.316	0.672	0.491
	6	A	1	2	638	0.788	0.852	0.544
	6	B	1	1	604	0.789	0.966	0.334
	20		2	2	673	0.406	0.624	0.526
	21		2	3	702	0.614	0.575	0.582
	42		2	2	681	0.656	0.571	0.596
	43		2	2	671	0.849	0.654	0.621
	MC	1		1	1	737	0.139	0.441
2			1	1	705	0.351	0.590	0.389
1			2	1	633	0.542	0.927	0.422
2			2	1	655	0.436	0.817	0.474
3			2	1	672	0.163	0.783	0.354
4			2	1	659	0.501	0.821	0.495
5			2	1	663	0.455	0.795	0.484
6			2	1	714	0.125	0.558	0.298
7			2	1	659	0.277	0.810	0.422
8			2	1	652	0.287	0.823	0.422
9			2	1	642	0.515	0.913	0.429
10			2	1	680	0.256	0.662	0.419
11			2	1	649	0.556	0.854	0.496
12			2	1	682	0.059	0.674	0.264
13			2	1	634	0.190	0.788	0.365
14			2	1	638	0.582	0.916	0.450
15			2	1	648	0.825	0.885	0.510
16			2	1	622	0.987	0.962	0.397
17			2	1	638	0.248	0.853	0.385
18			2	1	664	0.404	0.794	0.473
19			2	1	624	0.198	0.896	0.329
22			2	1	710	0.746	0.459	0.458
23			2	1	673	0.283	0.699	0.433
24			2	1	734	0.187	0.422	0.273
25			2	1	691	0.531	0.637	0.481
26			2	1	673	0.949	0.800	0.525
27			2	1	682	0.466	0.700	0.479
28			2	1	675	0.255	0.702	0.434
29			2	1	629	0.091	0.745	0.281
30			2	1	638	0.547	0.919	0.433
31			2	1	654	0.356	0.834	0.434
32			2	1	698	0.660	0.548	0.489
33			2	1	688	0.519	0.662	0.482
34			2	1	747	0.513	0.351	0.201
35			2	1	654	0.285	0.830	0.420
36			2	1	700	0.247	0.614	0.379
37			2	1	692	0.863	0.694	0.455
38			2	1	674	0.312	0.773	0.437
39			2	1	680	0.266	0.713	0.444
40			2	1	721	0.317	0.492	0.322
41			2	1	722	0.295	0.460	0.328
44			2	1	693	0.226	0.586	0.405
45			2	1	734	0.388	0.320	0.320
46			2	1	678	0.212	0.698	0.421
47			2	1	683	0.213	0.653	0.404
48			2	1	720	0.107	0.562	0.279
49			2	1	719	0.113	0.578	0.276
50			2	1	761	0.052	0.476	0.184
51			2	1	745	0.134	0.450	0.231
52			2	1	755	0.339	0.259	0.212
53			2	1	767	0.015	0.511	0.162

**Table 6. 9: Item Statistics: Communication Arts Grade 11**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	3		1	2	749	0.254	0.327	0.400
	4		1	2	741	0.361	0.361	0.458
	5	A	1	2	716	0.478	0.523	0.532
	5	B	1	1	637	0.398	0.940	0.276
	6	A	1	2	677	0.419	0.759	0.464
	6	B	1	2	637	0.572	0.916	0.377
	1		2	4	664	1.167	0.678	0.651
	21		3	3	657	0.266	0.681	0.432
	22		3	2	716	0.808	0.537	0.598
	42		3	3	733	0.595	0.412	0.556
	MC	1		1	1	701	0.336	0.812
2			1	1	714	0.281	0.727	0.388
1			3	1	753	0.622	0.450	0.335
2			3	1	650	0.410	0.954	0.286
3			3	1	731	1.175	0.524	0.529
4			3	1	738	1.190	0.495	0.447
5			3	1	723	0.497	0.677	0.421
6			3	1	664	0.195	0.793	0.327
7			3	1	714	0.216	0.671	0.368
8			3	1	675	0.641	0.885	0.430
9			3	1	688	0.295	0.792	0.395
10			3	1	694	0.313	0.760	0.408
11			3	1	679	0.135	0.818	0.289
12			3	1	770	0.912	0.268	0.239
13			3	1	719	0.285	0.595	0.399
14			3	1	694	0.799	0.833	0.485
15			3	1	694	0.258	0.797	0.374
16			3	1	699	0.582	0.787	0.474
17			3	1	729	0.213	0.620	0.357
18			3	1	801	0.401	0.181	0.150
19			3	1	741	0.159	0.632	0.293
20			3	1	720	0.391	0.700	0.408
23			3	1	672	0.504	0.916	0.381
24			3	1	769	0.195	0.442	0.249
25			3	1	704	0.489	0.759	0.455
26			3	1	719	0.386	0.629	0.434
27			3	1	726	0.525	0.632	0.430
28			3	1	687	0.626	0.845	0.462
29			3	1	684	0.078	0.774	0.246
30			3	1	740	0.142	0.588	0.287
31			3	1	681	0.862	0.905	0.447
32			3	1	665	0.722	0.944	0.371
33			3	1	687	0.281	0.780	0.387
34			3	1	706	0.768	0.732	0.525
35			3	1	707	0.469	0.706	0.457
36			3	1	749	1.211	0.373	0.385
37			3	1	734	0.260	0.597	0.362
38			3	1	707	0.947	0.768	0.497
39			3	1	706	1.514	0.778	0.545
40			3	1	705	0.575	0.720	0.486
41			3	1	717	0.430	0.679	0.416
43			3	1	690	0.251	0.806	0.367
44			3	1	709	0.602	0.703	0.494
45			3	1	671	0.389	0.899	0.365
46			3	1	729	0.237	0.681	0.342
47			3	1	703	0.353	0.753	0.425
48			3	1	751	0.484	0.478	0.321
49			3	1	782	1.748	0.270	0.097
50		3	1	777	0.842	0.276	0.176	
51		3	1	778	0.922	0.191	0.183	
52		3	1	783	0.319	0.362	0.182	

**Table 6. 10: Item Statistics: Mathematics, Grade 3**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	9		1	2	586	0.481	0.722	0.549
	16		1	2	604	0.475	0.637	0.557
	23		1	2	620	0.511	0.514	0.573
	27		1	2	599	0.809	0.686	0.628
	31		2	2	570	0.329	0.772	0.477
	32		2	2	607	0.150	0.554	0.423
	33		2	2	604	0.894	0.642	0.652
MC	1		1	1	624	0.185	0.643	0.392
	2		1	1	572	0.112	0.837	0.323
	3		1	1	606	0.264	0.676	0.452
	4		1	1	610	0.107	0.697	0.320
	5		1	1	581	0.483	0.857	0.483
	6		1	1	526	0.285	0.967	0.258
	7		1	1	629	0.245	0.562	0.414
	8		1	1	580	0.252	0.845	0.419
	10		1	1	590	0.103	0.773	0.296
	11		1	1	609	0.470	0.726	0.499
	12		1	1	619	0.519	0.576	0.509
	13		1	1	615	0.200	0.667	0.388
	14		1	1	565	0.258	0.891	0.382
	15		1	1	617	0.228	0.660	0.389
	17		1	1	603	0.144	0.729	0.365
	18		1	1	605	0.034	0.693	0.236
	19		1	1	493	0.208	0.987	0.154
	20		1	1	573	0.323	0.882	0.417
	21		1	1	606	0.113	0.680	0.313
	22		1	1	558	0.344	0.924	0.380
	24		1	1	662	0.092	0.606	0.241
	25		1	1	570	0.392	0.903	0.426
	26		1	1	613	0.104	0.624	0.332
	1		2	1	557	0.135	0.870	0.305
	2		2	1	542	0.170	0.921	0.281
	3		2	1	532	0.118	0.918	0.213
	4		2	1	582	0.270	0.900	0.348
	5		2	1	606	0.459	0.722	0.475
	6		2	1	536	0.274	0.964	0.268
	7		2	1	600	0.251	0.699	0.445
8		2	1	599	0.166	0.714	0.407	
9		2	1	568	0.214	0.866	0.364	
10		2	1	562	0.341	0.920	0.388	
11		2	1	560	0.493	0.935	0.375	
12		2	1	554	0.317	0.936	0.356	
13		2	1	563	0.464	0.912	0.414	
14		2	1	568	0.129	0.862	0.313	
15		2	1	562	0.096	0.842	0.310	
16		2	1	537	0.217	0.948	0.292	
17		2	1	526	0.137	0.943	0.261	
18		2	1	504	0.167	0.977	0.179	
19		2	1	617	0.222	0.691	0.386	
20		2	1	587	0.223	0.707	0.418	
21		2	1	619	0.123	0.638	0.369	
22		2	1	583	0.241	0.804	0.397	
23		2	1	406	0.018	0.934	0.148	
24		2	1	519	0.340	0.981	0.259	
25		2	1	557	0.836	0.947	0.413	
26		2	1	586	0.590	0.809	0.508	
27		2	1	601	0.273	0.763	0.450	
28		2	1	642	0.335	0.565	0.372	
29		2	1	648	0.448	0.429	0.405	
30		2	1	640	0.760	0.531	0.438	

**Table 6. 11: Item Statistics: Mathematics, Grade 4**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	5		1	2	633	0.658	0.585	0.602
	11		1	2	627	0.483	0.618	0.564
	14		1	2	596	0.308	0.769	0.470
	19		1	2	660	0.305	0.392	0.463
	23		1	2	635	0.887	0.576	0.634
	25		1	2	670	0.531	0.321	0.500
	30		1	4	592	0.861	0.787	0.645
	33		2	2	626	0.700	0.628	0.623
	34		2	2	609	0.453	0.725	0.536
	35		2	2	616	0.436	0.659	0.554
MC	1		1	1	587	0.316	0.905	0.386
	2		1	1	621	0.236	0.731	0.426
	3		1	1	643	0.212	0.664	0.390
	4		1	1	655	0.198	0.585	0.365
	6		1	1	623	0.316	0.738	0.459
	7		1	1	659	0.112	0.638	0.299
	8		1	1	633	0.201	0.660	0.412
	9		1	1	593	0.447	0.903	0.428
	10		1	1	588	0.277	0.898	0.378
	12		1	1	640	0.329	0.751	0.409
	13		1	1	596	0.079	0.809	0.273
	15		1	1	610	0.192	0.804	0.383
	16		1	1	583	0.356	0.911	0.391
	17		1	1	609	0.176	0.804	0.380
	18		1	1	619	0.120	0.747	0.346
	20		1	1	612	0.520	0.854	0.479
	21		1	1	632	0.087	0.693	0.303
	22		1	1	592	0.187	0.867	0.356
	24		1	1	672	0.405	0.499	0.351
	26		1	1	659	0.368	0.499	0.433
	27		1	1	642	0.332	0.693	0.431
	28		1	1	669	0.251	0.534	0.344
	29		1	1	648	0.233	0.623	0.399
	1		2	1	597	0.153	0.846	0.341
	2		2	1	628	0.176	0.694	0.391
	3		2	1	612	0.337	0.862	0.430
	4		2	1	652	0.265	0.621	0.393
	5		2	1	647	0.075	0.632	0.284
	6		2	1	554	0.136	0.865	0.283
	7		2	1	617	0.626	0.803	0.530
	8		2	1	638	0.181	0.662	0.373
	9		2	1	629	0.880	0.722	0.574
10		2	1	642	0.323	0.689	0.417	
11		2	1	540	0.458	0.987	0.217	
12		2	1	563	0.430	0.964	0.313	
13		2	1	663	0.763	0.459	0.459	
14		2	1	587	0.119	0.751	0.312	
15		2	1	599	0.257	0.845	0.401	
16		2	1	598	0.260	0.864	0.394	
17		2	1	594	0.933	0.932	0.453	
18		2	1	646	0.273	0.599	0.415	
19		2	1	581	0.168	0.892	0.317	
20		2	1	616	0.296	0.815	0.441	
21		2	1	644	0.209	0.680	0.377	
22		2	1	594	0.400	0.893	0.426	
23		2	1	631	0.705	0.737	0.539	
24		2	1	585	0.311	0.902	0.381	
25		2	1	589	0.216	0.878	0.362	
26		2	1	625	0.725	0.751	0.558	
27		2	1	641	0.268	0.653	0.416	
28		2	1	572	0.134	0.887	0.290	
29		2	1	635	0.736	0.692	0.543	
30		2	1	600	0.224	0.845	0.382	
31		2	1	659	0.348	0.525	0.420	
32		2	1	658	0.567	0.562	0.427	

**Table 6. 12: Item Statistics: Mathematics, Grade 5**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	6		1	2	669	0.345	0.430	0.532
	8		1	2	651	0.356	0.547	0.561
	17		1	2	702	0.233	0.287	0.430
	24		1	2	633	0.342	0.646	0.556
	33		2	3	582	0.727	0.851	0.579
	34		2	2	710	0.545	0.193	0.471
	35		2	3	669	0.228	0.412	0.504
MC	1		1	1	649	0.117	0.657	0.372
	2		1	1	674	0.533	0.498	0.489
	3		1	1	583	0.279	0.935	0.351
	4		1	1	653	1.297	0.606	0.652
	5		1	1	697	0.176	0.432	0.343
	7		1	1	663	0.119	0.578	0.365
	9		1	1	617	0.262	0.825	0.444
	10		1	1	697	0.195	0.467	0.329
	11		1	1	680	0.458	0.474	0.454
	12		1	1	573	0.223	0.944	0.304
	13		1	1	616	0.226	0.836	0.418
	14		1	1	608	0.171	0.822	0.392
	15		1	1	635	0.488	0.792	0.519
	16		1	1	664	0.219	0.583	0.436
	18		1	1	608	0.160	0.845	0.382
	19		1	1	638	0.176	0.755	0.405
	20		1	1	612	0.310	0.833	0.455
	21		1	1	563	0.113	0.930	0.253
	22		1	1	674	0.326	0.583	0.424
	23		1	1	669	0.434	0.539	0.496
	25		1	1	634	0.375	0.742	0.513
	26		1	1	660	0.118	0.696	0.342
	27		1	1	670	0.037	0.617	0.240
	1		2	1	602	0.110	0.825	0.324
	2		2	1	637	0.290	0.717	0.481
	3		2	1	643	0.087	0.675	0.325
	4		2	1	641	0.308	0.684	0.493
	5		2	1	628	0.233	0.769	0.444
	6		2	1	617	0.216	0.810	0.428
	7		2	1	626	0.252	0.789	0.453
	8		2	1	663	0.237	0.621	0.437
	9		2	1	575	0.339	0.911	0.377
10		2	1	554	0.038	0.878	0.199	
11		2	1	564	0.436	0.947	0.335	
12		2	1	635	0.300	0.780	0.478	
13		2	1	613	0.060	0.784	0.274	
14		2	1	636	0.131	0.708	0.377	
15		2	1	620	0.683	0.828	0.550	
16		2	1	571	0.125	0.921	0.273	
17		2	1	610	0.020	0.745	0.173	
18		2	1	559	0.664	0.984	0.249	
19		2	1	650	0.270	0.663	0.467	
20		2	1	572	0.408	0.956	0.325	
21		2	1	575	0.102	0.818	0.304	
22		2	1	620	0.093	0.768	0.323	
23		2	1	652	0.565	0.695	0.521	
24		2	1	681	0.358	0.479	0.434	
25		2	1	623	0.190	0.803	0.417	
26		2	1	653	0.426	0.659	0.538	
27		2	1	663	0.599	0.575	0.541	
28		2	1	668	0.283	0.658	0.414	
29		2	1	619	0.286	0.852	0.429	
30		2	1	641	0.219	0.746	0.438	
31		2	1	667	0.339	0.604	0.464	
32		2	1	639	0.348	0.766	0.492	

**Table 6. 13: Item Statistics: Mathematics, Grade 6**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	7		1	2	691	0.484	0.397	0.560
	14		1	2	650	0.188	0.590	0.460
	19		1	2	637	0.422	0.714	0.548
	24		1	2	705	0.448	0.326	0.539
	32		2	3	635	0.514	0.629	0.588
	33		2	2	634	0.499	0.757	0.535
	34		2	2	627	0.482	0.806	0.502
	MC	1		1	1	601	0.310	0.944
2			1	1	617	0.154	0.874	0.319
3			1	1	668	0.385	0.755	0.451
4			1	1	656	0.175	0.735	0.389
5			1	1	685	0.097	0.579	0.327
6			1	1	705	0.137	0.567	0.308
8			1	1	690	0.309	0.538	0.428
9			1	1	697	0.428	0.406	0.475
10			1	1	709	0.134	0.535	0.304
11			1	1	715	0.264	0.427	0.343
12			1	1	660	0.981	0.693	0.597
13			1	1	667	0.505	0.685	0.510
15			1	1	678	0.541	0.553	0.530
16			1	1	657	0.207	0.749	0.405
17			1	1	694	0.078	0.574	0.298
18			1	1	693	0.460	0.499	0.470
20			1	1	723	0.203	0.424	0.300
21			1	1	677	0.281	0.628	0.438
22			1	1	579	0.214	0.964	0.230
23			1	1	701	0.410	0.487	0.407
25			1	1	640	0.218	0.829	0.361
26			1	1	648	0.240	0.768	0.409
27			1	1	678	0.153	0.596	0.382
1			2	1	591	0.074	0.871	0.237
2			2	1	620	0.080	0.808	0.281
3			2	1	651	0.179	0.735	0.401
4			2	1	644	0.559	0.804	0.505
5			2	1	661	0.203	0.696	0.408
6			2	1	627	0.260	0.778	0.423
7			2	1	682	0.068	0.638	0.279
8			2	1	631	0.191	0.821	0.371
9		2	1	644	0.764	0.819	0.522	
10		2	1	571	0.316	0.981	0.203	
11		2	1	629	0.249	0.861	0.385	
12		2	1	661	0.670	0.724	0.548	
13		2	1	675	0.464	0.655	0.502	
14		2	1	688	0.348	0.522	0.461	
15		2	1	629	0.054	0.785	0.243	
16		2	1	660	0.392	0.733	0.470	
17		2	1	682	0.089	0.673	0.307	
18		2	1	642	0.706	0.843	0.496	
19		2	1	678	0.582	0.644	0.500	
20		2	1	638	0.227	0.824	0.396	
21		2	1	675	0.339	0.664	0.448	
22		2	1	678	0.349	0.642	0.464	
23		2	1	658	0.113	0.720	0.338	
24		2	1	700	0.212	0.475	0.395	
25		2	1	686	0.488	0.645	0.457	
26		2	1	667	0.397	0.691	0.495	
27		2	1	678	0.382	0.660	0.453	
28		2	1	664	0.155	0.754	0.367	
29		2	1	695	0.847	0.495	0.490	
30		2	1	723	0.256	0.388	0.314	
31		2	1	718	0.483	0.391	0.318	

**Table 6. 14: Item Statistics: Mathematics, Grade 7**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	7		1	2	725	0.535	0.235	0.518
	14		1	2	722	0.102	0.324	0.353
	21		1	2	712	0.786	0.267	0.568
	27		1	2	680	0.369	0.509	0.562
	33		2	2	644	0.505	0.727	0.566
	34		2	3	668	0.489	0.555	0.621
	35		2	3	694	0.698	0.428	0.652
MC	1		1	1	690	0.308	0.663	0.431
	2		1	1	618	0.055	0.825	0.247
	3		1	1	641	0.334	0.802	0.452
	4		1	1	711	0.094	0.560	0.301
	5		1	1	719	0.217	0.457	0.355
	6		1	1	665	0.189	0.786	0.386
	8		1	1	724	0.471	0.332	0.390
	9		1	1	683	0.301	0.560	0.490
	10		1	1	690	0.347	0.590	0.469
	11		1	1	657	0.036	0.710	0.237
	12		1	1	730	0.679	0.335	0.309
	13		1	1	697	0.408	0.536	0.472
	15		1	1	689	0.468	0.593	0.500
	16		1	1	738	0.875	0.217	0.288
	17		1	1	702	0.266	0.559	0.406
	18		1	1	719	0.346	0.485	0.345
	19		1	1	753	0.268	0.301	0.221
	20		1	1	695	0.306	0.499	0.470
	22		1	1	560	0.429	0.992	0.142
	23		1	1	614	0.228	0.910	0.328
	24		1	1	701	0.841	0.487	0.496
	25		1	1	699	0.964	0.434	0.565
	26		1	1	729	0.439	0.369	0.322
	1		2	1	669	0.111	0.736	0.345
	2		2	1	668	0.299	0.682	0.474
	3		2	1	703	0.312	0.564	0.404
	4		2	1	619	0.158	0.789	0.357
	5		2	1	625	0.163	0.768	0.370
	6		2	1	625	0.251	0.872	0.379
	7		2	1	674	0.296	0.620	0.491
	8		2	1	655	0.157	0.760	0.386
	9		2	1	635	0.140	0.828	0.349
10		2	1	609	0.581	0.951	0.330	
11		2	1	682	0.160	0.673	0.392	
12		2	1	638	0.175	0.804	0.375	
13		2	1	676	0.812	0.657	0.595	
14		2	1	675	0.131	0.683	0.379	
15		2	1	624	0.338	0.898	0.382	
16		2	1	699	0.250	0.546	0.421	
17		2	1	665	0.439	0.712	0.526	
18		2	1	658	0.260	0.716	0.457	
19		2	1	652	0.175	0.750	0.397	
20		2	1	674	0.498	0.669	0.533	
21		2	1	649	0.419	0.828	0.466	
22		2	1	649	0.319	0.769	0.466	
23		2	1	657	0.393	0.790	0.475	
24		2	1	674	0.687	0.660	0.586	
25		2	1	673	0.241	0.682	0.443	
26		2	1	705	0.532	0.502	0.452	
27		2	1	631	0.112	0.719	0.332	
28		2	1	675	0.493	0.670	0.541	
29		2	1	695	0.356	0.519	0.481	
30		2	1	710	0.240	0.452	0.401	
31		2	1	689	0.912	0.554	0.577	
32		2	1	699	0.426	0.560	0.455	

**Table 6. 15: Item Statistics: Mathematics, Grade 8**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	4		1	2	715	0.221	0.446	0.478
	7		1	2	734	1.323	0.233	0.606
	12		1	2	657	0.451	0.774	0.492
	17		1	2	732	0.754	0.290	0.589
	22		1	2	710	0.410	0.447	0.552
	27		1	2	718	0.699	0.381	0.601
	30		1	4	749	1.864	0.186	0.653
	32		2	2	692	0.243	0.553	0.475
	33		2	2	717	0.495	0.399	0.563
	34		2	2	720	0.481	0.381	0.564
	MC	1		1	1	704	0.656	0.706
2			1	1	712	0.111	0.598	0.332
3			1	1	729	0.256	0.557	0.355
5			1	1	747	0.343	0.253	0.393
6			1	1	708	0.294	0.622	0.430
8			1	1	705	0.707	0.547	0.559
9			1	1	704	0.700	0.603	0.531
10			1	1	673	0.282	0.771	0.430
11			1	1	672	0.295	0.796	0.424
13			1	1	696	0.165	0.665	0.385
14			1	1	731	1.055	0.411	0.431
15			1	1	770	0.139	0.469	0.201
16			1	1	672	0.150	0.773	0.354
18			1	1	716	0.577	0.471	0.513
19			1	1	740	0.366	0.386	0.371
20			1	1	738	0.190	0.434	0.341
21			1	1	668	0.197	0.801	0.375
23			1	1	657	0.181	0.837	0.352
24			1	1	694	0.422	0.643	0.504
25			1	1	713	0.447	0.567	0.463
26			1	1	708	0.192	0.600	0.396
28			1	1	746	0.208	0.412	0.318
29			1	1	737	0.481	0.372	0.406
1			2	1	642	0.066	0.806	0.256
2			2	1	732	0.558	0.477	0.377
3			2	1	741	0.283	0.372	0.353
4			2	1	651	0.128	0.817	0.317
5			2	1	669	0.298	0.793	0.428
6			2	1	724	0.484	0.592	0.397
7			2	1	705	0.160	0.633	0.370
8			2	1	620	0.037	0.840	0.188
9		2	1	696	0.091	0.667	0.311	
10		2	1	632	0.138	0.895	0.275	
11		2	1	622	0.208	0.935	0.268	
12		2	1	659	0.462	0.849	0.443	
13		2	1	618	0.289	0.908	0.308	
14		2	1	650	0.103	0.831	0.295	
15		2	1	707	0.183	0.683	0.361	
16		2	1	644	0.405	0.911	0.369	
17		2	1	679	0.291	0.763	0.440	
18		2	1	694	0.337	0.678	0.468	
19		2	1	698	0.314	0.656	0.446	
20		2	1	731	0.475	0.444	0.434	
21		2	1	670	0.215	0.765	0.395	
22		2	1	704	0.269	0.689	0.419	
23		2	1	675	0.507	0.803	0.488	
24		2	1	713	0.383	0.558	0.455	
25		2	1	686	0.366	0.759	0.455	
26		2	1	738	0.162	0.491	0.326	
27		2	1	733	0.313	0.439	0.394	
28		2	1	707	0.179	0.601	0.382	
29		2	1	726	0.163	0.577	0.340	
30		2	1	736	0.202	0.435	0.361	
31		2	1	754	0.208	0.433	0.264	

**Table 6. 16: Item Statistics: Mathematics, Grade 10**

Item Type	No.	Part	Sess.	Points	Loc	Info	P-Val.	R <sub>it</sub>
CR	5		1	2	743	0.556	0.411	0.642
	10		1	2	750	0.317	0.393	0.577
	15		1	2	710	0.075	0.560	0.370
	20		1	2	747	0.379	0.396	0.596
	26		1	2	764	0.288	0.312	0.536
	30		1	2	709	0.432	0.633	0.612
	32		1	4	787	0.938	0.240	0.665
	26		2	2	761	0.570	0.303	0.621
	27		2	2	729	1.035	0.505	0.741
	28		2	3	762	0.588	0.374	0.699
	29		2	2	731	0.615	0.488	0.671
MC	1		1	1	731	0.309	0.626	0.512
	2		1	1	806	0.811	0.225	0.228
	3		1	1	719	0.144	0.634	0.438
	4		1	1	789	0.226	0.262	0.380
	6		1	1	748	0.358	0.516	0.504
	7		1	1	726	0.361	0.700	0.497
	8		1	1	719	0.535	0.633	0.613
	9		1	1	742	0.740	0.540	0.575
	11		1	1	778	0.062	0.544	0.277
	12		1	1	771	0.189	0.453	0.381
	13		1	1	688	0.182	0.805	0.415
	14		1	1	711	0.150	0.667	0.438
	16		1	1	786	0.199	0.337	0.362
	17		1	1	731	0.287	0.592	0.515
	18		1	1	687	0.199	0.765	0.442
	19		1	1	746	0.132	0.602	0.391
	21		1	1	708	0.554	0.733	0.562
	22		1	1	707	0.308	0.733	0.506
	23		1	1	730	0.239	0.765	0.414
	24		1	1	788	0.910	0.185	0.417
	25		1	1	746	0.471	0.501	0.547
	27		1	1	744	0.732	0.484	0.598
	28		1	1	779	0.269	0.477	0.334
	29		1	1	800	0.629	0.325	0.247
	31		1	1	780	0.497	0.525	0.289
	1		2	1	678	0.179	0.819	0.406
	2		2	1	766	0.125	0.450	0.368
	3		2	1	754	0.219	0.451	0.462
	4		2	1	744	0.158	0.596	0.416
	5		2	1	738	0.109	0.596	0.391
	6		2	1	600	0.031	0.893	0.181
7		2	1	727	0.109	0.623	0.395	
8		2	1	736	0.581	0.606	0.556	
9		2	1	739	0.133	0.609	0.404	
10		2	1	716	0.065	0.714	0.317	
11		2	1	736	0.253	0.581	0.510	
12		2	1	705	0.374	0.730	0.529	
13		2	1	699	0.484	0.781	0.536	
14		2	1	768	0.281	0.506	0.386	
15		2	1	729	0.147	0.661	0.417	
16		2	1	743	0.164	0.608	0.414	
17		2	1	742	0.311	0.555	0.509	
18		2	1	712	0.317	0.725	0.511	
19		2	1	685	0.047	0.766	0.273	
20		2	1	735	0.079	0.604	0.356	
21		2	1	743	0.649	0.507	0.576	
22		2	1	727	0.223	0.636	0.475	
23		2	1	724	0.205	0.621	0.472	
24		2	1	707	0.127	0.717	0.407	
25		2	1	751	0.158	0.580	0.411	

## **Part 7: Fairness**

### ***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. The item and test development processes that are used to minimize bias are summarized below.

First, careful attention was paid to content validity during the item development and item-selection process. Bias can occur only 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 several 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 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.

Additional steps to reduce bias, including the use of Bias Review committees composed of Missouri participants, are described in more detail in Part 3 of this report.

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

After administering the test, an empirical approach known as differential item functioning (DIF) was used to examine 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.

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, no one test will be equally appropriate for all. 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 cultural-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. Such tests are measuring different things for different groups and can be called biased (Camilli & Shepard, 1994; Green, 1975). 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. As discussed above and in Part 3 of this report, careful attention is given during the test development and test construction processes to lessen the influence of these elements for large numbers of students (including the use of Bias Review committees). Unfortunately, in some cases these elements may continue to play a substantial role. To assess the extent to which items may be performing differently for various subgroups of interest, DIF analyses are conducted after each operational test administration.

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). The SMD statistic applies weights to the focal group and reference group data, adjusting for differences in the distribution of the reference and focal group members across the values of the matching variable. The SMD takes into account the natural ordering of the response levels of the items and has the desirable property of being based on those ability levels where members of the focal group are present. The SMD method has the additional advantage of generally requiring fewer items than the Mantel-Haenszel procedure, which generally requires a moderately long matching test (see, for example, Donoghue, Holland, and Thayer, 1993).

For the analyses reported here, the matching test for each item was all of the items in the test including the item under consideration. For each item, this SMD statistic is computed by finding the weighted differences between item scores for members of the focal group and reference group at each level on the matching test, then averaging the differences across levels. The SMD was then divided by the total item standard deviation to obtain the SMD effect-size. Items were flagged for moderate DIF if the absolute value of the result was between .10 and .19, and for large DIF if the absolute value was greater than .19. 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-price lunches; Reference group is all others.
- ***Disability***: Focal group is students who indicated one or more disabilities; 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 implies 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.

When interpreting these results, it is important to note that test items that function differently for members of different groups are not necessarily biased (Camilli & Shepard, 1994). If DIF statistics merely signal some type of multidimensionality that causes an item to function differently for two groups, then additional analysis is needed to determine whether the source of differential difficulty is relevant or irrelevant to the construct being measured. Only when an item is tapping knowledge or specific ability irrelevant to the construct the test intends to measure, can we say that the item is biased against a particular group. Substantive review and interpretation by test developers of items flagged by DIF procedures is essential for helping developers/test users understand the reasons underlying DIF for members of different groups.

### ***Evaluating Bias through Impact Analysis***

The impact of achievement testing on minorities can be determined and reported in the form of average scores and also in terms of test score reliability. Tables 7.3 through 7.16 present the scale score means and standard deviations, numbers of students, and test form reliability statistics (Coefficient Alpha, see Part 4) for various subgroups of interest.

Some believe that fairness is an issue whenever the measured ability differences between subgroups are overly large; however, a criterion for large difference is lacking. One way to evaluate the magnitude of the differences between subgroups is to see whether the lowest performing subgroup mean differs from the majority subgroup mean by more than one standard deviation on the scale. Tables 7.3 through 7.16 presents the scale score that is one standard deviation below the majority mean for the various subgroups of interest. Across these tables, the only groups that consistently have mean scale scores more than one standard deviation below the majority mean include students with disabilities, students receiving an individual education plan (IEP), or students receiving accommodations.

**Table 7. 1: 2006 MAP SMD Statistics: Number of Flagged Items, Communication Arts**

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>Female</b>	<b>03</b>	31568	1	0	0	0
	<b>04</b>	31905	0	0	0	0
	<b>05</b>	31835	0	2	0	0
	<b>06</b>	32463	2	0	0	0
	<b>07</b>	33913	6	2	0	1
	<b>08</b>	35063	1	1	0	0
	<b>11</b>	29970	3	2	0	0
<b>Asian/Pacific Islander</b>	<b>03</b>	1103	2	2	0	0
	<b>04</b>	1097	1	1	0	0
	<b>05</b>	1134	0	2	0	0
	<b>06</b>	1014	0	0	0	0
	<b>07</b>	968	3	1	0	0
	<b>08</b>	994	2	0	0	0
	<b>11</b>	959	1	1	0	0
<b>Black</b>	<b>03</b>	11679	3	1	0	0
	<b>04</b>	11588	0	0	0	0
	<b>05</b>	12053	1	1	0	0
	<b>06</b>	12610	0	1	0	0
	<b>07</b>	13419	0	2	0	0
	<b>08</b>	13187	1	1	0	0
	<b>11</b>	8842	0	1	0	0
<b>Hispanic</b>	<b>03</b>	2194	0	0	0	0
	<b>04</b>	2162	0	0	0	0
	<b>05</b>	2185	1	1	0	0
	<b>06</b>	2051	0	0	0	0
	<b>07</b>	2052	0	1	0	0
	<b>08</b>	2050	2	1	0	0
	<b>11</b>	1336	1	1	0	0
<b>Native American/Alaskan</b>	<b>03</b>	245	0	0	0	0
	<b>04</b>	275	0	0	0	0
	<b>05</b>	279	0	0	0	0
	<b>06</b>	270	0	0	0	0
	<b>07</b>	326	0	0	0	0
	<b>08</b>	371	1	0	0	0
	<b>11</b>	244	0	1	0	0

**2006 MAP SMD Statistics: Number of Flagged Items, Communication Arts (continued)**

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>ELL</b>	<b>03</b>	1797	1	1	0	0
	<b>04</b>	1612	0	0	0	0
	<b>05</b>	1519	0	1	0	0
	<b>06</b>	1271	1	0	0	0
	<b>07</b>	1162	2	1	0	0
	<b>08</b>	1141	3	2	0	0
	<b>11</b>	666	2	5	0	0

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>IEP</b>	<b>03</b>	9961	1	0	0	0
	<b>04</b>	10129	0	0	0	0
	<b>05</b>	9905	0	0	0	0
	<b>06</b>	9444	0	1	0	0
	<b>07</b>	9865	0	0	0	1
	<b>08</b>	9975	0	0	0	0
	<b>11</b>	6761	0	0	0	1

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>SES</b>	<b>03</b>	28452	0	0	0	0
	<b>04</b>	28308	0	0	0	0
	<b>05</b>	28203	0	0	0	0
	<b>06</b>	27910	0	0	0	0
	<b>07</b>	28713	0	0	0	0
	<b>08</b>	28327	0	0	0	0
	<b>11</b>	15691	0	0	0	0

2006 MAP SMD Statistics: Number of Flagged Items, Communication Arts (continued)

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
Accommodations	03	6287	1	1	0	0
	04	7085	0	1	0	0
	05	7512	0	0	0	0
	06	7059	0	1	0	0
	07	7602	0	1	0	1
	08	7724	0	0	0	0
	11	4601	0	0	0	1

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
Disability	03	9978	1	0	0	0
	04	10083	0	0	0	0
	05	9863	0	0	0	0
	06	9378	0	1	0	0
	07	9819	0	0	0	1
	08	9949	0	0	0	0
	11	6702	0	0	0	1

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
Migrant	03	1886	2	0	0	0
	04	1398	0	0	0	0
	05	1192	0	0	0	0
	06	698	2	0	0	0
	07	569	1	0	0	0
	08	533	2	1	0	0
	11	189	2	2	0	0

**Table 7. 2: 2006 MAP SMD Statistics: Number of Flagged Items, Mathematics**

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>Female</b>	<b>03</b>	31636	1	0	0	0
	<b>04</b>	31941	2	0	0	0
	<b>05</b>	31884	0	1	0	0
	<b>06</b>	32471	1	0	0	0
	<b>07</b>	34002	1	0	0	0
	<b>08</b>	35089	1	2	0	0
	<b>10</b>	33483	1	1	0	0
<b>Asian/Pacific Islander</b>	<b>03</b>	1138	0	1	0	0
	<b>04</b>	1139	0	0	0	0
	<b>05</b>	1165	1	0	0	0
	<b>06</b>	1034	0	0	0	0
	<b>07</b>	995	1	0	0	0
	<b>08</b>	1011	0	0	0	0
	<b>10</b>	1038	1	4	0	1
<b>Black</b>	<b>03</b>	11735	0	0	0	0
	<b>04</b>	11593	0	0	0	0
	<b>05</b>	12058	1	0	0	0
	<b>06</b>	12616	1	2	0	0
	<b>07</b>	13531	0	1	0	0
	<b>08</b>	13164	0	0	0	0
	<b>10</b>	11193	0	2	0	0
<b>Hispanic</b>	<b>03</b>	2258	0	0	0	0
	<b>04</b>	2216	0	0	0	0
	<b>05</b>	2260	2	0	0	0
	<b>06</b>	2098	0	1	0	0
	<b>07</b>	2112	1	0	0	0
	<b>08</b>	2081	0	0	0	0
	<b>10</b>	1734	0	0	0	0
<b>Native American/Alaskan</b>	<b>03</b>	250	1	0	0	0
	<b>04</b>	270	0	0	0	0
	<b>05</b>	283	0	0	0	0
	<b>06</b>	273	0	0	0	0
	<b>07</b>	327	0	0	0	0
	<b>08</b>	368	0	0	0	0
	<b>10</b>	307	0	1	0	0

2006 MAP SMD Statistics: Number of Flagged Items, Mathematics (continued)

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
ELL	03	1913	2	1	0	0
	04	1732	0	1	0	0
	05	1659	1	0	0	0
	06	1364	1	2	0	0
	07	1272	1	0	0	0
	08	1191	0	2	0	0
	10	963	0	4	0	2

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
IEP	03	9961	1	0	0	0
	04	10129	0	0	0	0
	05	9905	0	0	0	0
	06	9444	0	1	0	0
	07	9865	0	0	0	1
	08	9975	0	0	0	0
	11	6761	0	0	0	1

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
SES	03	28613	0	0	0	0
	04	28317	0	0	0	0
	05	28208	0	0	0	0
	06	28022	0	0	0	0
	07	28973	0	0	0	0
	08	28298	0	0	0	0
	10	21003	0	0	0	0

Group	Grade	Sample Size	Moderate Positive	Moderate Negative	Large Positive	Large Negative
Disability	03	10037	0	0	0	0
	04	10112	0	1	0	0
	05	9881	0	0	0	0
	06	9384	0	0	0	0
	07	9942	1	0	0	0
	08	9959	0	0	0	0
	10	8529	0	0	0	0

**2006 MAP SMD Statistics: Number of Flagged Items, Mathematics (continued)**

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>Accommodations</b>	<b>03</b>	6429	2	0	0	0
	<b>04</b>	7346	1	1	0	0
	<b>05</b>	7802	0	0	0	0
	<b>06</b>	7304	0	0	0	0
	<b>07</b>	7970	2	0	0	0
	<b>08</b>	8021	2	0	0	0
	<b>10</b>	6252	1	0	0	0

<b>Group</b>	<b>Grade</b>	<b>Sample Size</b>	<b>Moderate Positive</b>	<b>Moderate Negative</b>	<b>Large Positive</b>	<b>Large Negative</b>
<b>Migrant</b>	<b>03</b>	1904	1	0	0	0
	<b>04</b>	1404	1	1	0	0
	<b>05</b>	1205	2	1	0	0
	<b>06</b>	710	0	0	0	0
	<b>07</b>	577	0	1	0	0
	<b>08</b>	541	0	0	0	0
	<b>10</b>	332	0	1	0	0

**Table 7. 3: Impact Analysis, Grade 03 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	146	633.36		40.12	609.3
	White (not Hispanic)	49119	644.50	0.90	35.23	
	Black (not Hispanic)	11679	622.31	0.91	37.32	
	Hispanic	2194	625.42	0.91	36.35	
	Asian/Pacific Islander	1103	648.08	0.91	38.26	
	Native American	245	641.48	0.90	36.84	
Gender	Female	31568	644.79	0.90	35.67	597.9
	Male	32602	635.19	0.91	37.30	
ELL Status	Non ELL	62689	640.47	0.90	36.61	603.9
	ELL	1797	618.37	0.91	38.26	
IEP Status	Non IEP	54525	644.81	0.88	33.08	611.7
	IEP	9961	612.73	0.92	43.88	
SES Status	Non SES	36034	648.98	0.89	34.26	614.7
	SES	28452	628.31	0.90	36.74	
Disability	No Disability	54508	644.84	0.88	33.05	611.8
	Disability	9978	612.65	0.92	43.90	
Accommodations	No Accommodations	58199	644.46	0.88	33.35	611.1
	Accommodations	6287	597.25	0.90	40.16	
Migrant Status	Non Migrant	62600	640.44	0.90	36.56	603.9
	Migrant	1886	620.57	0.91	40.62	

**Table 7. 4: Impact Analysis, Grade 04 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	139	638.81		40.94	621.9
	White (not Hispanic)	49918	658.92	0.90	37.03	
	Black (not Hispanic)	11588	637.25	0.90	39.39	
	Hispanic	2162	642.87	0.90	38.67	
	Asian/Pacific Islander	1097	664.22	0.89	39.46	
	Native American	275	650.67	0.90	40.58	
Gender	Female	31905	660.59	0.89	36.54	609.4
	Male	32984	648.88	0.91	39.45	
ELL Status	Non ELL	63567	655.14	0.90	38.31	616.8
	ELL	1612	631.29	0.90	40.86	
IEP Status	Non IEP	55050	660.58	0.87	32.89	627.7
	IEP	10129	621.80	0.92	49.16	
SES Status	Non SES	36871	663.55	0.89	35.86	627.7
	SES	28308	642.83	0.90	38.80	
Disability	No Disability	55096	660.56	0.87	32.87	627.7
	Disability	10083	621.69	0.92	49.28	
Accommodations	No Accommodations	58094	660.09	0.88	33.50	626.6
	Accommodations	7085	609.16	0.90	46.51	
Migrant Status	Non Migrant	63781	655.05	0.90	38.33	616.7
	Migrant	1398	631.84	0.90	42.00	

**Table 7. 5: Impact Analysis, Grade 05 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	148	655.24		39.67	637.8
	White (not Hispanic)	50208	672.71	0.90	34.95	
	Black (not Hispanic)	12053	650.58	0.90	39.08	
	Hispanic	2185	656.67	0.90	39.04	
	Asian/Pacific Islander	1134	679.89	0.92	39.96	
	Native American	279	661.65	0.91	40.01	
Gender	Female	31835	672.95	0.90	35.06	625.6
	Male	33925	663.84	0.91	38.26	
ELL Status	Non ELL	64488	668.74	0.91	36.78	632.0
	ELL	1519	644.51	0.90	42.10	
IEP Status	Non IEP	56102	674.44	0.89	30.89	643.5
	IEP	9905	632.73	0.91	47.81	
SES Status	Non SES	37804	677.42	0.90	33.97	643.5
	SES	28203	655.79	0.90	37.48	
Disability	No Disability	56144	674.42	0.89	30.95	643.5
	Disability	9863	632.67	0.91	47.69	
Accommodations	No Accommodations	58495	673.77	0.89	31.63	642.1
	Accommodations	7512	624.67	0.89	46.52	
Migrant Status	Non Migrant	64815	668.52	0.91	36.91	631.6
	Migrant	1192	649.54	0.91	42.09	

**Table 7. 6: Impact Analysis, Grade 06 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	118	647.80		45.01	640.8
	White (not Hispanic)	50885	672.02	0.90	31.23	
	Black (not Hispanic)	12610	647.37	0.90	35.24	
	Hispanic	2051	654.97	0.91	34.61	
	Asian/Pacific Islander	1014	675.97	0.92	35.39	
	Native American	270	664.92	0.89	28.78	
Gender	Female	32463	671.75	0.90	31.44	627.3
	Male	34216	662.34	0.91	35.02	
ELL Status	Non ELL	65677	667.33	0.91	33.43	633.9
	ELL	1271	641.75	0.90	37.85	
IEP Status	Non IEP	57504	672.67	0.88	28.32	644.3
	IEP	9444	631.41	0.90	41.30	
SES Status	Non SES	39038	675.33	0.89	30.59	644.7
	SES	27910	654.98	0.90	34.26	
Disability	No Disability	57570	672.66	0.88	28.32	644.3
	Disability	9378	631.16	0.90	41.27	
Accommodations	No Accommodations	59889	671.76	0.89	29.16	642.6
	Accommodations	7059	625.12	0.88	40.16	
Migrant Status	Non Migrant	66250	667.03	0.91	33.61	633.4
	Migrant	698	648.99	0.92	37.15	

**Table 7. 7: Impact Analysis, Grade 07 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	160	650.84		35.93	642.1
	White (not Hispanic)	53365	677.42	0.90	35.29	
	Black (not Hispanic)	13419	650.19	0.90	35.56	
	Hispanic	2052	658.76	0.90	36.12	
	Asian/Pacific Islander	968	680.94	0.92	38.69	
	Native American	326	669.67	0.90	33.52	
Gender	Female	33913	677.09	0.90	34.25	627.9
	Male	36067	666.67	0.91	38.78	
ELL Status	Non ELL	69128	672.13	0.91	36.86	635.3
	ELL	1162	642.09	0.90	37.08	
IEP Status	Non IEP	60425	678.34	0.89	31.99	646.3
	IEP	9865	630.54	0.89	39.43	
SES Status	Non SES	41577	680.93	0.90	34.66	646.3
	SES	28713	658.16	0.90	36.28	
Disability	No Disability	60471	678.33	0.89	32.01	646.3
	Disability	9819	630.37	0.89	39.32	
Accommodations	No Accommodations	62688	677.20	0.89	32.85	644.4
	Accommodations	7602	625.67	0.88	37.87	
Migrant Status	Non Migrant	69721	671.78	0.91	36.95	634.8
	Migrant	569	653.21	0.93	45.42	

**Table 7. 8: Impact Analysis, Grade 08 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	213	650.00		47.70	656.6
	White (not Hispanic)	55668	692.52	0.91	35.92	
	Black (not Hispanic)	13187	664.65	0.91	36.06	
	Hispanic	2050	674.14	0.92	37.95	
	Asian/Pacific Islander	994	700.03	0.93	43.31	
	Native American	371	681.25	0.93	39.73	
Gender	Female	35063	693.45	0.91	36.42	642.8
	Male	37078	680.85	0.92	38.04	
ELL Status	Non ELL	71342	687.34	0.92	37.64	649.7
	ELL	1141	656.10	0.91	39.79	
IEP Status	Non IEP	62508	693.58	0.90	32.89	660.7
	IEP	9975	644.65	0.90	39.72	
SES Status	Non SES	44156	695.84	0.91	35.83	660.0
	SES	28327	672.85	0.91	36.69	
Disability	No Disability	62534	693.61	0.90	32.85	660.8
	Disability	9949	644.38	0.90	39.67	
Accommodations	No Accommodations	64759	692.50	0.90	33.51	659.0
	Accommodations	7724	639.46	0.89	39.15	
Migrant Status	Non Migrant	71950	686.98	0.92	37.82	649.2
	Migrant	533	669.03	0.93	40.18	

**Table 7. 9: Impact Analysis, Grade 11 Communication Arts**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	137	698.76		36.56	690.6
	White (not Hispanic)	48486	720.43	0.90	29.83	
	Black (not Hispanic)	8842	697.28	0.89	31.79	
	Hispanic	1336	706.01	0.90	30.98	
	Asian/Pacific Islander	959	725.07	0.92	34.12	
	Native American	244	713.91	0.90	33.66	
Gender	Female	29970	721.62	0.90	29.05	692.6
	Male	29719	711.91	0.91	32.83	
ELL Status	Non ELL	59338	717.01	0.91	31.26	685.8
	ELL	666	688.78	0.89	32.99	
IEP Status	Non IEP	53243	721.64	0.89	27.30	694.3
	IEP	6761	677.74	0.88	34.32	
SES Status	Non SES	44313	721.34	0.90	29.82	691.5
	SES	15691	703.57	0.90	32.09	
Disability	No Disability	53302	721.61	0.89	27.34	694.3
	Disability	6702	677.58	0.87	34.20	
Accommodations	No Accommodations	55403	720.33	0.89	28.25	692.1
	Accommodations	4601	672.87	0.86	34.35	
Migrant Status	Non Migrant	59815	716.78	0.91	31.37	685.4
	Migrant	189	689.49	0.90	33.69	

**Table 7. 10: Impact Analysis, Grade 03 Mathematics**

Group		N	Mean	Coefficient Alpha	Std. Dev.	1 SD Below Majority Mean
Ethnicity	Unknown	179	596.79		43.73	589.7
	White (not Hispanic)	49203	627.07	0.89	37.37	
	Black (not Hispanic)	11735	599.87	0.91	38.23	
	Hispanic	2258	610.58	0.90	36.70	
	Asian/Pacific Islander	1138	633.47	0.90	40.84	
	Native American	250	624.77	0.89	37.29	
Gender	Female	31636	621.58	0.90	37.98	581.7
	Male	32798	621.76	0.91	40.11	
ELL Status	Non ELL	62850	622.05	0.90	39.02	583.0
	ELL	1913	606.40	0.91	38.79	
IEP Status	Non IEP	54716	625.52	0.89	36.71	588.8
	IEP	10047	600.15	0.93	44.42	
SES Status	Non SES	36150	631.33	0.89	37.79	593.5
	SES	28613	609.28	0.90	37.22	
Disability	No Disability	54726	625.51	0.89	36.74	588.8
	Disability	10037	600.21	0.93	44.36	
Accommodations	No Accommodations	58334	625.35	0.89	36.99	588.4
	Accommodations	6429	587.43	0.91	41.16	
Migrant Status	Non Migrant	62859	622.17	0.90	38.89	583.3
	Migrant	1904	602.44	0.91	41.45	

**Table 7. 11: Impact Analysis, Grade 04 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	149	626.99		45.83	613.7
	White (not Hispanic)	49939	648.79	0.92	35.12	
	Black (not Hispanic)	11593	623.26	0.92	37.59	
	Hispanic	2216	634.94	0.92	34.35	
	Asian/Pacific Islander	1139	658.39	0.93	41.10	
	Native American	270	641.05	0.92	34.91	
Gender	Female	31941	644.05	0.92	36.17	606.0
	Male	33038	643.87	0.93	37.84	
ELL Status	Non ELL	63574	644.35	0.92	36.91	607.4
	ELL	1732	626.61	0.93	38.95	
IEP Status	Non IEP	55156	648.52	0.91	33.58	614.9
	IEP	10150	618.66	0.94	44.30	
SES Status	Non SES	36989	652.64	0.91	35.04	617.6
	SES	28317	632.43	0.92	36.54	
Disability	No Disability	55194	648.54	0.91	33.55	615.0
	Disability	10112	618.40	0.94	44.33	
Accommodations	No Accommodations	57960	648.28	0.91	33.89	614.4
	Accommodations	7346	609.15	0.93	42.41	
Migrant Status	Non Migrant	63902	644.29	0.92	36.90	607.4
	Migrant	1404	625.04	0.93	39.82	

**Table 7. 12: Impact Analysis, Grade 05 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	144	641.46		44.19	627.8
	White (not Hispanic)	50213	665.71	0.90	37.88	
	Black (not Hispanic)	12058	636.92	0.91	39.49	
	Hispanic	2260	650.11	0.91	38.01	
	Asian/Pacific Islander	1165	679.48	0.91	42.23	
	Native American	283	653.75	0.92	40.87	
Gender	Female	31884	659.67	0.91	38.52	619.3
	Male	33951	660.57	0.92	41.25	
ELL Status	Non ELL	64464	660.50	0.91	39.86	620.6
	ELL	1659	642.88	0.92	41.22	
IEP Status	Non IEP	56192	665.72	0.90	36.15	629.6
	IEP	9931	628.03	0.93	45.21	
SES Status	Non SES	37915	669.90	0.90	37.95	631.9
	SES	28208	646.85	0.91	38.83	
Disability	No Disability	56242	665.72	0.90	36.17	629.6
	Disability	9881	627.84	0.93	45.09	
Accommodations	No Accommodations	58321	665.30	0.90	36.51	628.8
	Accommodations	7802	620.88	0.92	42.99	
Migrant Status	Non Migrant	64918	660.41	0.91	39.92	620.5
	Migrant	1205	641.61	0.91	39.28	

**Table 7. 13: Impact Analysis, Grade 06 Mathematics**

<b>Group</b>		<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	113	650.22		45.87	643.6
	White (not Hispanic)	50883	680.29	0.92	36.65	
	Black (not Hispanic)	12616	646.09	0.91	39.70	
	Hispanic	2098	660.42	0.91	36.40	
	Asian/Pacific Islander	1034	691.10	0.93	45.14	
	Native American	273	669.75	0.88	29.34	
Gender	Female	32471	673.95	0.92	38.09	631.5
	Male	34274	672.81	0.93	41.30	
ELL Status	Non ELL	65653	673.78	0.92	39.57	634.2
	ELL	1364	650.34	0.92	43.68	
IEP Status	Non IEP	57566	679.23	0.91	35.60	643.6
	IEP	9451	637.17	0.92	44.61	
SES Status	Non SES	38995	683.75	0.92	37.14	646.6
	SES	28022	658.77	0.92	38.79	
Disability	No Disability	57633	679.20	0.91	35.66	643.5
	Disability	9384	637.07	0.92	44.45	
Accommodations	No Accommodations	59713	678.46	0.92	36.20	642.3
	Accommodations	7304	631.12	0.90	42.69	
Migrant Status	Non Migrant	66307	673.54	0.92	39.74	633.8
	Migrant	710	651.25	0.91	39.22	

**Table 7. 14: Impact Analysis, Grade 07 Mathematics**

Group		N	Mean	Coefficient Alpha	Std. Dev.	1 SD Below Majority Mean
Ethnicity	Unknown	208	645.58		49.74	644.3
	White (not Hispanic)	53525	682.83	0.92	38.51	
	Black (not Hispanic)	13531	647.10	0.89	38.49	
	Hispanic	2112	661.94	0.91	38.65	
	Asian/Pacific Islander	995	695.49	0.93	45.42	
	Native American	327	669.19	0.92	40.24	
Gender	Female	34002	676.57	0.92	39.36	631.7
	Male	36326	674.51	0.93	42.80	
ELL Status	Non ELL	69426	675.89	0.92	41.06	634.8
	ELL	1272	647.65	0.91	43.47	
IEP Status	Non IEP	60685	681.99	0.92	37.11	644.9
	IEP	10013	635.27	0.90	42.50	
SES Status	Non SES	41725	686.32	0.92	38.58	647.7
	SES	28973	659.62	0.91	39.91	
Disability	No Disability	60756	681.96	0.92	37.16	644.8
	Disability	9942	635.18	0.89	42.35	
Accommodations	No Accommodations	62728	680.97	0.92	37.79	643.2
	Accommodations	7970	631.35	0.88	41.09	
Migrant Status	Non Migrant	70121	675.56	0.92	41.20	634.4
	Migrant	577	653.08	0.92	43.64	

**Table 7. 15: Impact Analysis, Grade 08 Mathematics**

Group		N	Mean	Coefficient Alpha	Std. Dev.	1 SD Below Majority Mean
Ethnicity	Unknown	206	668.07		48.98	667.6
	White (not Hispanic)	55712	704.79	0.92	37.18	
	Black (not Hispanic)	13164	668.66	0.90	39.53	
	Hispanic	2081	686.60	0.92	37.12	
	Asian/Pacific Islander	1011	716.80	0.94	44.00	
	Native American	368	694.60	0.92	40.57	
Gender	Female	35089	698.26	0.92	38.54	655.6
	Male	37087	697.48	0.93	41.88	
ELL Status	Non ELL	71351	698.13	0.93	40.16	658.0
	ELL	1191	673.60	0.92	45.22	
IEP Status	Non IEP	62546	704.29	0.92	35.59	668.7
	IEP	9996	656.68	0.89	44.14	
SES Status	Non SES	44244	707.78	0.92	37.58	670.2
	SES	28298	682.00	0.91	39.56	
Disability	No Disability	62583	704.29	0.92	35.59	668.7
	Disability	9959	656.47	0.89	44.05	
Accommodations	No Accommodations	64521	703.37	0.92	36.20	667.2
	Accommodations	8021	652.32	0.88	43.39	
Migrant Status	Non Migrant	72001	697.87	0.93	40.33	657.5
	Migrant	541	678.14	0.91	40.50	

**Table 7. 16: Impact Analysis, Grade 10 Mathematics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Coefficient Alpha</b>	<b>Std. Dev.</b>	<b>1 SD Below Majority Mean</b>
Ethnicity	Unknown	146	690.00		62.87	685.4
	White (not Hispanic)	53665	732.67	0.93	47.28	
	Black (not Hispanic)	11193	686.23	0.91	50.38	
	Hispanic	1734	707.22	0.93	50.58	
	Asian/Pacific Islander	1038	747.79	0.95	53.58	
	Native American	307	718.48	0.93	49.34	
Gender	Female	33483	724.37	0.94	48.55	671.3
	Male	34240	724.83	0.94	53.53	
ELL Status	Non ELL	67120	725.00	0.94	50.93	674.1
	ELL	963	687.44	0.92	54.60	
IEP Status	Non IEP	59476	732.80	0.93	45.32	687.5
	IEP	8607	666.83	0.89	52.21	
SES Status	Non SES	47080	734.14	0.94	48.31	685.8
	SES	21003	702.78	0.93	50.79	
Disability	No Disability	59554	732.71	0.93	45.38	687.3
	Disability	8529	666.89	0.89	52.35	
Accommodations	No Accommodations	61831	730.77	0.93	46.75	684.0
	Accommodations	6252	662.12	0.87	51.34	
Migrant Status	Non Migrant	67751	724.69	0.94	51.07	673.6
	Migrant	332	678.93	0.91	52.65	

## Part 8: Achievement-Level Standard Setting

A Bookmark standard setting was held in 2005 to establish cut scores for the Missouri Assessment Program (MAP). The materials used for this standard setting were based on data from the 2005 field test. Because the standard setting was based on field-test data, DESE also reviewed cut scores using operational data. In this section of the Technical Report, we summarize the results of the standard setting and the data review conducted on operational data.

### *2005 MAP Standard Setting*

In December 2005, a modified Bookmark Standard Setting Procedure was used to establish cut scores for the Communication Arts and Mathematics MAPs for Grades 3–8 and high school. A modification of Bookmark was used to meet the requirements of Senate Bill 1080, which requires that NAEP-like cut scores be established for the MAPs.

Senate Bill 1080 was interpreted such that the *Proficient* achievement level met, but did not exceed, the NAEP performance standards. In other words, the percentage of students who attain *Proficient* on the MAP should be similar to or slightly higher than the percentage attaining *Proficient* on NAEP. The percentage of students in the other three achievement levels would be allowed to vary between NAEP and the MAP.

For the purposes of the MAP standard setting, participants were allowed to recommend *Proficient* cut scores within a pre-specified range. This range was based on the percentage of students who could be classified as either *Proficient* or *Advanced*. At the low end, no fewer than 26% of students could be classified as either *Proficient* or *Advanced*, and, at the high end, approximately 44% of students could be classified as *Proficient* or *Advanced*.

The 2005 standard setting is discussed in more detail in the *Missouri Assessment Program Final Bookmark Standard Setting Technical Report (2005)*.

Table 8.1 summarizes the recommended cut scores and impact data from the 2005 MAP Standard Setting.

### *2006 Data Review*

The cut scores were studied in two phases. In Phase I, an in-depth analysis of the cut scores was conducted using the data from the calibration sample. In Phase II, the cut scores were examined and adjusted using the census data. This final phase was deemed necessary in case the results of the calibration sample differed from those of the state.

*Phase I Review:* First, the cut scores were adjusted using the calibration sample for the operational data. A complete explanation of the review is given in the Standard Setting Report. In short, adjustments were made to the cut scores that took into account the impact data (percentage of students in each achievement level), the content recommendations, and the political situation in Missouri given Senate Bill 1080. See Appendix for *Reviewing 2005 Bookmark Cut Scores with 2006 Operational Data (Phase I)*.

*Phase II Review:* Once census data were available, DESE and CTB again checked the data for a logical progression of cut scores and impact data across the grades. This final check was done to confirm that the impact data did not vary significantly from what was expected as a result of the Phase I data review. DESE established a rule that the impact data could vary by no more than 2 percentage points between the Phase I and Phase II data review.

Cut scores and with the impact data from the Phase I data review are presented in Table 8.2 and those from the Phase II final review are presented in Table 8.3.

**Table 8. 1: Recommended Cut Scores and Impact Data from 2005 MAP Standard Setting**

Content Area	Grade	Cut Score			Impact Data				
		<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Prof &amp; Adv</i>
Comm. Arts	3	587	642	668	10.1	46	26.5	17.4	43.9
	4	605	654	687	12.5	42.7	29.4	15.4	44.8
	5	620	669	699	10.1	45.2	29	15.7	44.7
	6	621	671	703	11	45.3	31.7	12	43.7
	7	622	674	712	14.9	41	31.7	12.4	44.1
	8	627	691	720	9.6	47.7	27.3	15.4	42.7
	11	650	711	745	12.3	43.9	30.5	13.3	43.8
Math	3	559	620	667	11.1	45.1	35.0	8.8	43.8
	4	589	643	677	11.6	44.3	32.1	12.0	44.1
	5	599	658	699	12.7	42.9	33.4	11.0	44.4
	6	613	671	710	13.4	42.3	32.6	11.7	44.3
	7	631	676	714	21.4	34.6	32.3	11.7	44
	8	658	702	741	24.6	35.7	28.2	11.5	39.7
	10	683	723	777	27.2	28.4	34.2	10.2	44.4

**Table 8. 2: Cut Scores from the Phase I 2006 Data Review**

Content Area	Grade	Cut Score			Impact Data				
		<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Prof &amp; Adv</i>
Comm. Arts	3	592	648	673	10.3	46.9	26.0	16.8	42.8
	4	612	662	691	11.7	43.4	30.8	14.1	44.9
	5	625	675	702	10.5	46.2	27.5	15.9	43.4
	6	631	676	704	10.4	44.4	33.7	11.5	45.2
	7	634	684	712	12.2	44.7	30.2	12.9	43.1
	8	645	699	720	9.3	49.0	25.7	15.9	41.6
	11	679	725	753	10.1	45.9	31.4	12.6	44.0
Math	3	568	625	667	9.4	47.2	34.9	8.5	43.4
	4	596	651	688	8.9	47.8	33.5	9.8	43.3
	5	605	665	706	9.6	44.5	35.0	11.0	46.0
	6	628	681	721	9.7	46.2	33.0	11.0	44.0
	7	645	689	726	17.8	38.7	31.8	11.7	43.5
	8	671	712	741	19.3	39.9	29.4	11.4	40.8
	10	695	738	785	24.0	32.2	33.2	10.6	43.8

**Table 8. 3: Final Cut Scores from the Phase II 2006 Data Review**

Content Area	Grade	Cut Score			Impact Data				
		<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Prof &amp; Adv</i>
Comm. Arts	3	592	648	673	8.9	48.1	26.0	17.0	43.0
	4	612	662	691	10.7	45.0	29.1	15.2	44.3
	5	625	675	702	9.2	45.3	29.9	15.6	45.5
	6	631	676	704	12.0	45.3	32.0	10.7	42.7
	7	634	680	712	14.0	42.5	31.1	12.4	43.5
	8	639	696	723	9.3	48.7	27.0	15.1	42.1
	11	679	725	753	10.2	47.3	31.5	11.0	42.5
Math	3	568	628	667	7.2	49.1	33.6	10.1	43.7
	4	596	651	688	8.3	47.9	34.7	9.1	43.8
	5	605	668	706	8.0	48.2	33.0	10.8	43.8
	6	628	681	721	11.1	44.5	34.7	9.6	44.3
	7	640	685	724	17.6	38.9	33.1	10.4	43.5
	8	670	710	741	21.3	38.3	28.0	12.4	40.4
	10	695	738	785	24.4	33.3	32.0	10.4	42.3

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

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

### **Hand-Scoring Process Used for MAP**

Evidence of validity is provided by the procedures described below for hand-scoring.

### **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 2006 project was staffed with a large number of returning evaluators and team leaders who had previous experience with MAP and other hand-scoring projects. Kelly Services also recruited new team leaders and evaluators for employment. Recruitment sources included advertisements in newspapers in Indianapolis, Indiana, and nearby areas and through 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.

## **Hand-Scoring 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 hand-scoring of constructed-response and extended-response items (written essays and performance events).

Missouri operational items have been previously field tested. Prior to actual scoring, hand-scoring 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 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 and 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), which consisted of a group of items to be scored. Evaluators and team leaders were trained using the following steps:

- Reviewing the student response booklet
- Reviewing rubrics
- Reviewing anchor papers
- Explanation of 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 hand-scoring, 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 being administered. All checkset scores had been approved by DESE participants during and immediately following the TMRM.

In addition to the checkset process, CTB's hand-scoring 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 hand-scoring refers to the process in which scorers/raters who begin to drift away from scoring accuracy are realigned to corrected 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 Communication Arts items. Only 10 items had less than 90% perfect and adjacent agreement, and the percent of adjacent and perfect agreement for these items ranged from 85 to 89%. The Kappa statistic for the Communication Arts items ranged from .24 (Grade 8, Item 6B) to .97 (Grade 4, Item 43). As shown in Table 9.2, raters demonstrated above 90% adjacent agreement for all Mathematics items. The Kappa statistic for the Mathematics items ranged from .68 (Grade 10, Item 15) to .99 (Grade 3, Item 31; Grade 10, Item 30).

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

<b>Grade</b>	<b>Item</b>	<b>Percentage of Perfect Agreement</b>	<b>Percentage of Adjacent Agreement</b>	<b>Percentage of Perfect and Adjacent Agreement</b>	<b>Weighted Kappa</b>
3	3	83.73	15.31	99.04	0.88
3	4	77.27	21.29	98.56	0.82
3	5	82.06	14.59	96.65	0.83
3	6A	86.36	12.20	98.56	0.89
3	6B	86.12	11.96	98.08	0.53
3	1	66.51	32.54	99.05	0.69
3	22	81.82	17.22	99.04	0.84
3	23	85.41	13.40	98.81	0.84
3	24	94.74	3.59	98.33	0.82
3	43	85.41	12.20	97.61	0.84
4	3	80.15	18.11	98.26	0.77
4	4	73.20	25.06	98.26	0.66
4	5	79.40	17.37	96.77	0.78
4	6A	67.74	30.27	98.01	0.66
4	6B	76.92	21.59	98.51	0.58
4	19	94.54	3.47	98.01	0.91
4	20	88.34	10.67	99.01	0.92
4	21	83.62	12.90	96.52	0.88
4	43	97.52	1.24	98.76	0.97
5	3	77.86	11.89	89.75	0.79
5	4	69.93	19.81	89.74	0.71
5	5	63.40	25.87	89.27	0.53
5	6A	61.77	26.57	88.34	0.64
5	6B	79.72	9.32	89.04	0.77
5	17	86.48	3.73	90.21	0.96
5	18	58.51	30.77	89.28	0.54
5	42	69.70	18.88	88.58	0.89
5	43	63.17	25.87	89.04	0.62
6	3	66.27	28.67	94.94	0.68
6	4	70.36	21.45	91.81	0.65
6	5	70.84	25.54	96.38	0.74
6	6A	78.07	16.87	94.94	0.82
6	6B	84.82	11.08	95.90	0.80

**Table 9. 1: Inter-Rater Reliability, Communication Arts (continued)**

<b>Grade</b>	<b>Item</b>	<b>Percentage of Perfect Agreement</b>	<b>Percentage of Adjacent Agreement</b>	<b>Percentage of Perfect and Adjacent Agreement</b>	<b>Weighted Kappa</b>
6	17	74.22	22.17	96.39	0.67
6	18	81.45	15.42	96.87	0.79
6	19	65.30	26.99	92.29	0.51
6	43	60.00	31.33	91.33	0.72
7	3	74.72	12.86	87.58	0.68
7	4	82.71	12.86	95.57	0.87
7	5A	75.83	18.85	94.68	0.84
7	5B	86.92	8.65	95.57	0.86
7	6	66.96	24.39	91.35	0.70
7	1	70.73	25.72	96.45	0.70
7	19	70.29	24.61	94.90	0.75
7	20	55.43	31.04	86.47	0.61
7	42	78.05	17.29	95.34	0.84
8	3	77.83	18.63	96.46	0.71
8	4	85.14	11.31	96.45	0.81
8	5	72.28	22.84	95.12	0.71
8	6A	84.92	11.75	96.67	0.75
8	6B	94.24	3.77	98.01	0.24
8	20	74.06	15.30	89.36	0.71
8	21	71.62	21.95	93.57	0.76
8	42	80.71	17.29	98.00	0.85
8	43	81.60	14.63	96.23	0.85
11	3	69.21	22.43	91.64	0.62
11	4	61.58	29.36	90.94	0.55
11	5A	68.50	26.73	95.23	0.68
11	5B	92.36	4.06	96.42	0.66
11	6A	73.03	23.87	96.90	0.70
11	6B	91.65	6.44	98.09	0.83
11	1	63.25	33.65	96.90	0.68
11	21	78.28	19.33	97.61	0.80
11	22	70.41	21.00	91.41	0.72
11	42	56.09	28.88	84.97	0.62

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

Grade	Item	Percentage of Perfect Agreement	Percentage of Adjacent Agreement	Percentage of Perfect and Adjacent Agreement	Weighted Kappa
3	9	91.28	6.05	97.33	0.94
3	16	89.59	5.33	94.92	0.95
3	23	94.19	2.91	97.10	0.98
3	27	94.67	3.15	97.82	0.98
3	31	96.13	1.21	97.34	0.99
3	32	83.29	14.53	97.82	0.81
3	33	93.95	3.87	97.82	0.98
4	5	92.29	3.37	95.66	0.94
4	11	86.99	10.60	97.59	0.91
4	14	90.36	6.75	97.11	0.92
4	19	84.10	10.84	94.94	0.91
4	23	88.92	8.19	97.11	0.95
4	25	79.04	15.90	94.94	0.82
4	30	67.23	27.71	94.94	0.81
4	33	95.42	1.93	97.35	0.98
4	34	86.02	10.84	96.86	0.89
4	35	82.89	14.46	97.35	0.85
5	6	87.93	4.68	92.61	0.97
5	8	86.21	6.16	92.37	0.96
5	17	82.51	9.61	92.12	0.91
5	24	81.53	8.87	90.40	0.90
5	33	86.21	5.67	91.88	0.92
5	34	89.41	2.71	92.12	0.95
5	35	86.70	4.93	91.63	0.97
6	7	93.33	4.14	97.47	0.95
6	14	96.32	1.15	97.47	0.98
6	19	94.71	2.99	97.70	0.97
6	24	75.40	19.08	94.48	0.76
6	32	94.48	3.22	97.70	0.98
6	33	94.02	3.91	97.93	0.96
6	34	90.57	6.44	97.01	0.91
7	7	87.47	5.69	93.16	0.94
7	14	91.80	1.82	93.62	0.87
7	21	90.43	3.87	94.30	0.96
7	27	93.39	4.33	97.72	0.96
7	33	93.17	4.56	97.73	0.96
7	34	85.88	9.79	95.67	0.92
7	35	90.21	6.15	96.36	0.96
8	4	84.42	14.29	98.71	0.79
8	7	93.07	3.90	96.97	0.96
8	12	83.33	13.64	96.97	0.85
8	17	93.72	4.55	98.27	0.96

**Table 9. 2: Inter-Rater Reliability, Mathematics (continued)**

<b>Grade</b>	<b>Item</b>	<b>Percentage of Perfect Agreement</b>	<b>Percentage of Adjacent Agreement</b>	<b>Percentage of Perfect and Adjacent Agreement</b>	<b>Weighted Kappa</b>
8	22	94.59	2.81	97.40	0.97
8	27	87.01	10.82	97.83	0.92
8	30	87.66	7.58	95.24	0.96
8	32	92.42	5.84	98.26	0.97
8	33	81.17	16.02	97.19	0.87
8	34	85.93	11.69	97.62	0.89
10	5	89.19	4.73	93.92	0.96
10	10	88.96	3.60	92.56	0.97
10	15	64.19	28.60	92.79	0.68
10	20	88.51	5.41	93.92	0.96
10	26	86.71	3.83	90.54	0.96
10	30	93.24	0.90	94.14	0.99
10	32	70.27	22.07	92.34	0.87
10	26	91.89	3.83	95.72	0.97
10	27	91.44	3.38	94.82	0.98
10	28	84.23	12.84	97.07	0.91
10	29	92.57	3.60	96.17	0.98

## Part 10: Student Demographics

Demographic information was collected on all students and includes gender, race/ethnicity, free or reduced-price lunch, 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 of the calibration sample and the census data, reviews the participation rates of students in eight subgroups, and summarizes the usage of test accommodations in the 2006 MAP administration.

### *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, racial/ethnic, and school/district size variables. In general, these samples ranged in size from 3700 to 4500 students and were used for classical item analyses and IRT analyses of the operational data. Table 10.1 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 from a state testing program. By using a representative calibration sample, states are able to expedite the return of score reports to the districts.

**Table 10.1: Number of Schools Invited to Participate in the Calibration Sample**

Grade Level	Number of Schools
3	73
4	73
5	65
6	45
7	41
8	41
10	34
11	34

### *Calibration and Census Demographics*

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

Table 10.2 and 10.3 summarize the number of students in the calibration sample and the overall (census) number of Missouri students for each grade of Communication Arts and Mathematics, respectively. Tables 10.2 and 10.3 also summarize the distribution of students by gender and race/ethnicity. These tables also summarize the mean scale scores and standard deviations for all students and by gender and race/ethnicity. Looking across Tables 10.2 and 10.3, 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 for 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.

### ***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 10.4 to 10.17. 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 achievement levels in a content area cannot be determined. Accountable students who did not attempt all the testing sessions in the content area being assessed or were invalidated by the state do not have an achievement level assigned to them.

### ***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 MAP *Examiner’s Manual* 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/divspced/>
- <http://www.dese.mo.gov/divimprove/>

Tables 10.18 through 10.20 summarize the numbers of students receiving accommodations by accommodation type for the 2006 MAP, the Braille Edition of the 2006 MAP, and the Large-Print Edition of the 2006 MAP. For the 2006 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 common accommodations. On the Braille and Large-Print editions of MAP, the setting and timing accommodations are again among the most frequently used accommodations. For these two editions, using a scribe and having the test read aloud are common accommodations. The analyses in Tables 10.18 through 10.20 reflect the complete census data.

**Table 10. 2: Summary of Calibration and Census Data: Communication Arts**

<b>Communication Arts Grade 3</b>								
	<b>Calibration Sample</b>				<b>Census Data</b>			
	<b>Scale Score</b>				<b>Scale Score</b>			
	<b>N</b>	<b>%</b>	<b>M</b>	<b>SD</b>	<b>N</b>	<b>%</b>	<b>M</b>	<b>SD</b>
<b>All Students</b>	4128		637.07	38.20	64486		639.86	36.84
<b>Gender</b>								
Unknown	11	0.27	624.00	33.94	316	0.49	628.65	38.15
Male	2129	51.57	631.73	38.35	32602	50.56	635.19	37.30
Female	1988	48.16	642.85	37.23	31568	48.95	644.79	35.67
<b>Race/Ethnicity</b>								
Unknown	6	0.15	641.83	27.89	146	0.23	633.36	40.12
White	3053	73.96	640.76	37.10	49119	76.17	644.50	35.23
Black	865	20.95	623.17	39.43	11679	18.11	622.31	37.32
Hispanic	136	3.29	635.74	36.69	2194	3.4	625.42	36.35
Asian/Pacific Islander	55	1.33	656.29	29.55	1103	1.71	648.08	38.26
Native American/Alaskan	13	0.31	624.62	34.29	245	0.38	641.48	36.84
<b>Communication Arts Grade 4</b>								
<b>All Students</b>	4047		654.38	38.62	65179		654.55	38.56
<b>Gender</b>								
Unknown	14	0.35	627.36	36.50	290	0.44	634.35	46.24
Male	2063	50.98	648.50	39.44	32984	50.61	648.88	39.45
Female	1970	48.68	660.72	36.67	31905	48.95	660.59	36.54
<b>Race/Ethnicity</b>								
Unknown	6	0.15	644.67	30.82	139	0.21	638.81	40.94
White	3010	74.38	659.89	37.74	49918	76.59	658.92	37.03
Black	850	21.00	635.31	35.34	11588	17.78	637.25	39.39
Hispanic	113	2.79	646.89	38.12	2162	3.32	642.87	38.67
Asian/Pacific Islander	49	1.21	665.37	41.09	1097	1.68	664.22	39.46
Native American/Alaskan	19	0.47	654.11	37.87	275	0.42	650.67	40.58
<b>Communication Arts Grade 5</b>								
<b>All Students</b>	3783		666.43	36.687	66007		668.18	37.09
<b>Gender</b>								
Unknown	9	0.24	624.22	56.19	247	0.37	648.52	47.85
Male	1951	51.57	661.61	37.58	33925	51.4	663.84	38.26
Female	1823	48.19	671.79	34.73	31835	48.23	672.95	35.06
<b>Race/Ethnicity</b>								
Unknown	10	0.26	626.50	45.01	148	0.22	655.24	39.67
White	2676	70.74	671.81	34.56	50208	76.06	672.71	34.95
Black	897	23.71	650.40	36.69	12053	18.26	650.58	39.08
Hispanic	123	3.25	662.70	36.41	2185	3.31	656.67	39.04
Asian/Pacific Islander	61	1.61	682.49	40.32	1134	1.72	679.89	39.96
Native American/Alaskan	16	0.42	657.19	67.89	279	0.42	661.65	40.01

**Table 10. 2: Summary of Calibration and Census Data: Communication Arts (Continued)**

<b>Communication Arts Grade 6</b>								
	<b>Calibration Sample</b>				<b>Census Data</b>			
			<b>Scale Score</b>				<b>Scale Score</b>	
	<b>N</b>	<b>%</b>	<b>M</b>	<b>SD</b>	<b>N</b>	<b>%</b>	<b>M</b>	<b>SD</b>
<b>All Students</b>	4158		668.63	32.61	66948		666.85	33.70
<b>Gender</b>								
Unknown	22	0.53	652.09	39.31	269	0.4	647.98	39.24
Male	2098	50.46	664.25	33.72	34216	51.11	662.34	35.02
Female	2038	49.01	673.32	30.64	32463	48.49	671.75	31.44
<b>Race/Ethnicity</b>								
Unknown	8	0.19	674.13	43.50	118	0.18	647.80	45.01
White	3073	73.91	673.57	30.43	50885	76.01	672.02	31.23
Black	898	21.6	652.16	34.00	12610	18.84	647.37	35.24
Hispanic	120	2.89	664.33	33.13	2051	3.06	654.97	34.61
Asian/Pacific Islander	45	1.08	674.64	34.97	1014	1.51	675.97	35.39
Native American/Alaskan	14	0.34	655.29	35.09	270	0.4	664.92	28.78
<b>Communication Arts Grade 7</b>								
<b>All Students</b>	4376		673.45	37.78	70290		671.63	37.06
<b>Gender</b>								
Unknown	21	0.48	631.19	41.05	310	0.44	650.81	40.92
Male	2217	50.66	668.01	38.97	36067	51.31	666.67	38.78
Female	2138	48.86	679.51	35.31	33913	48.25	677.09	34.25
<b>Race/Ethnicity</b>								
Unknown	10	0.23	622.00	41.85	160	0.23	650.84	35.93
White	3563	81.42	677.24	36.56	53365	75.92	677.42	35.29
Black	606	13.85	652.31	38.16	13419	19.09	650.19	35.56
Hispanic	134	3.06	670.24	32.58	2052	2.92	658.76	36.12
Asian/Pacific Islander	38	0.87	679.34	40.79	968	1.38	680.94	38.69
Native American/Alaskan	25	0.57	675.56	37.21	326	0.46	669.67	33.52
<b>Communication Arts Grade 8</b>								
<b>All Students</b>	4476		688.641	36.506	72483		686.85	37.87
<b>Gender</b>								
Unknown	11	0.25	656.82	46.10	342	0.47	660.35	45.36
Male	2310	51.61	683.22	36.54	37078	51.15	680.85	38.04
Female	2155	48.15	694.61	35.42	35063	48.37	693.45	36.42
<b>Race/Ethnicity</b>								
Unknown	3	0.07	645.67	102.89	213	0.29	650.00	47.70
White	3688	82.39	692.27	35.35	55668	76.8	692.52	35.92
Black	542	12.11	666.00	36.77	13187	18.19	664.65	36.06
Hispanic	152	3.40	684.57	30.50	2050	2.83	674.14	37.95
Asian/Pacific Islander	56	1.25	693.54	28.91	994	1.37	700.03	43.31
Native American/Alaskan	35	0.78	670.46	45.14	371	0.51	681.25	39.73

**Table 10. 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>	4062		717.19	33.96	60004		716.69	31.42
<b>Gender</b>								
Unknown	11	0.27	678.45	30.92	315	0.52	699.67	35.88
Male	1960	48.25	712.59	35.34	29719	49.53	711.91	32.83
Female	2091	51.48	721.71	31.90	29970	49.95	721.62	29.05
<b>Race/Ethnicity</b>								
Unknown	7	0.17	697.71	25.84	137	0.23	698.76	36.56
White	3153	77.62	721.23	31.81	48486	80.8	720.43	29.83
Black	643	15.83	697.28	35.20	8842	14.74	697.28	31.79
Hispanic	133	3.27	708.96	34.13	1336	2.23	706.01	30.98
Asian/Pacific Islander Native	102	2.51	730.86	42.91	959	1.6	725.07	34.12
American/Alaskan	24	0.59	714.38	27.24	244	0.41	713.91	33.66

**Table 10.3: Summary of Calibration and Census Data: Mathematics**

	Mathematics Grade 3							
	Calibration Sample				Census Data			
	N	%	Scale Score		N	%	Scale Score	
M			SD	M			SD	
<b>All Students</b>	4125		617.67	38.56	64763		621.59	39.11
<b>Gender</b>								
Unknown	33	0.80	600.33	27.97	329	0.51	605.88	41.50
Male	2122	51.44	618.20	38.98	32798	50.64	621.76	40.11
Female	1970	47.76	617.40	38.20	31636	48.85	621.58	37.98
<b>Race/Ethnicity</b>								
Unknown	26	0.63	594.12	30.25	179	0.28	596.79	43.73
White	3009	72.95	622.85	37.40	49203	75.97	627.07	37.37
Black	879	21.31	599.80	37.34	11735	18.12	599.87	38.23
Hispanic	141	3.42	615.31	38.16	2258	3.49	610.58	36.70
Asian/Pacific Islander Native	58	1.41	637.09	33.01	1138	1.76	633.47	40.84
American/Alaskan	12	0.29	612.92	35.85	250	0.39	624.77	37.29
	Mathematics Grade 4							
<b>All Students</b>	4087		642.53	36.32	65306		643.88	37.07
<b>Gender</b>								
Unknown	26	0.64	626.88	28.17	327	0.50	627.33	42.70
Male	2071	50.67	641.99	36.21	33038	50.59	643.87	37.84
Female	1990	48.69	643.29	36.49	31941	48.91	644.05	36.17
<b>Race/Ethnicity</b>								
Unknown	7	0.17	625.14	14.45	149	0.23	626.99	45.83
White	3045	74.50	648.43	33.93	49939	76.47	648.79	35.12
Black	851	20.82	620.91	36.05	11593	17.75	623.26	37.59
Hispanic	116	2.84	641.16	35.97	2216	3.39	634.94	34.35
Asian/Pacific Islander Native	49	1.20	655.63	44.94	1139	1.74	658.39	41.10
American/Alaskan	19	0.46	646.11	33.25	270	0.41	641.05	34.91
	Mathematics Grade 5							
<b>All Students</b>	3810		657.25	41.2255	66123		660.06	39.99
<b>Gender</b>								
Unknown	23	0.60	636.30	43.36	288	0.44	642.97	44.09
Male	1952	51.23	656.65	42.32	33951	51.35	660.57	41.25
Female	1835	48.16	658.15	39.94	31884	48.22	659.67	38.52
<b>Race/Ethnicity</b>								
Unknown	16	0.42	621.25	44.87	144	0.22	641.46	44.19
White	2695	70.73	664.23	37.89	50213	75.94	665.71	37.88
Black	895	23.49	634.89	41.24	12058	18.24	636.92	39.49
Hispanic	126	3.31	656.50	44.98	2260	3.42	650.11	38.01
Asian/Pacific Islander Native	62	1.63	686.11	41.73	1165	1.76	679.48	42.23
American/Alaskan	16	0.42	661.56	54.96	283	0.43	653.75	40.87

**Table 10. 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>	4221		674.028	38.89	67017		673.30	39.80
<b>Gender</b>								
Unknown	15	0.36	654.73	48.37	272	0.41	657.79	42.35
Male	2127	50.39	672.86	40.51	34274	51.14	672.81	41.30
Female	2079	49.25	675.36	37.04	32471	48.45	673.95	38.09
<b>Race/Ethnicity</b>								
Unknown	4	0.09	650.50	25.79	113	0.17	650.22	45.87
White	3052	72.31	681.15	36.57	50883	75.93	680.29	36.65
Black	979	23.19	651.67	37.49	12616	18.83	646.09	39.70
Hispanic	123	2.91	667.98	32.00	2098	3.13	660.42	36.40
Asian/Pacific Islander	47	1.11	698.77	46.02	1034	1.54	691.10	45.14
Native American/Alaskan	16	0.38	662.69	38.77	273	0.41	669.75	29.34
	Mathematics Grade 7							
<b>All Students</b>	4390		678.83	40.81	70698		675.38	41.27
<b>Gender</b>								
Unknown	15	0.34	656.13	43.12	370	0.52	651.46	50.71
Male	2228	50.75	677.42	41.91	36326	51.38	674.51	42.80
Female	2147	48.91	680.46	39.53	34002	48.09	676.57	39.36
<b>Race/Ethnicity</b>								
Unknown	8	0.18	632.00	49.53	208	0.29	645.58	49.74
White	3502	79.77	684.16	39.29	53525	75.71	682.83	38.51
Black	688	15.67	652.51	38.58	13531	19.14	647.10	38.49
Hispanic	132	3.01	672.52	36.98	2112	2.99	661.94	38.65
Asian/Pacific Islander	38	0.87	697.00	33.60	995	1.41	695.49	45.42
Native American/Alaskan	22	0.50	676.82	39.03	327	0.46	669.19	40.24
	Mathematics Grade 8							
<b>All Students</b>	4443		699.27	39.02	72542		697.73	40.37
<b>Gender</b>								
Unknown	26	0.59	663.42	48.73	366	0.50	671.25	46.55
Male	2285	51.43	699.46	40.24	37087	51.12	697.48	41.88
Female	2132	47.99	699.51	37.35	35089	48.37	698.26	38.54
<b>Race/Ethnicity</b>								
Unknown	8	0.18	669.13	63.13	206	0.28	668.07	48.98
White	3608	81.21	704.10	36.81	55712	76.80	704.79	37.18
Black	590	13.28	672.14	40.80	13164	18.15	668.66	39.53
Hispanic	143	3.22	693.82	33.33	2081	2.87	686.60	37.12
Asian/Pacific Islander	59	1.33	703.58	34.51	1011	1.39	716.80	44.00
Native American/Alaskan	35	0.79	680.89	53.84	368	0.51	694.60	40.57

**Table 10. 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>	4330		724.92	51.74	68083	724.46	51.18	
<b>Gender</b>								
Unknown	14	0.32	678.00	62.84	360	0.53	698.88	55.07
Male	2173	50.18	726.79	52.99	34240	50.29	724.83	53.53
Female	2143	49.49	723.33	50.19	33483	49.18	724.37	48.55
<b>Race/Ethnicity</b>								
Unknown	12	0.28	703.50	59.86	146	0.21	690.00	62.87
White	3337	77.07	734.07	47.26	53665	78.82	732.67	47.28
Black	782	18.06	688.51	50.86	11193	16.44	686.23	50.38
Hispanic	116	2.68	701.84	53.36	1734	2.55	707.22	50.58
Asian/Pacific Islander Native	66	1.52	739.58	66.91	1038	1.52	747.79	53.58
American/Alaskan	17	0.39	719.29	63.28	307	0.45	718.48	49.34

**Table 10. 4: Participation Rates: All Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	65344	0.99	65325	0.99
04	65849	0.99	65845	0.99
05	66704	0.99	66703	0.99
06	67709	0.99	67706	0.99
07	71632	0.98	71575	0.99
08	73516	0.99	73523	0.99
10	-	-	69229	0.98
11	61593	0.97	-	-

**Table 10. 5: Participation Rates: Males**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	33147	0.98	33145	0.99
04	33381	0.99	33368	0.99
05	34355	0.99	34328	0.99
06	34701	0.99	34705	0.99
07	36907	0.98	36873	0.99
08	37686	0.98	37672	0.98
10	-	-	34880	0.98
11	30670	0.97	-	-

**Table 10. 6: Participation Rates: Females**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	31873	0.99	31840	0.99
04	32162	0.99	32138	0.99
05	32092	0.99	32074	0.99
06	32718	0.99	32719	0.99
07	34362	0.99	34308	0.99
08	35471	0.99	35462	0.99
10	-	-	33945	0.99
11	30555	0.98	-	-

**Table 10. 7: Participation Rates: White**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	49624	0.99	49608	0.99
04	50346	0.99	50340	0.99
05	50645	0.99	50634	0.99
06	51384	0.99	51393	0.99
07	54134	0.99	54106	0.99
08	56322	0.99	56358	0.99
10	-	-	54390	0.99
11	49566	0.98	-	-

**Table 10. 8: Participation Rates: Black**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	11892	0.98	11856	0.99
04	11704	0.99	11698	0.99
05	12180	0.99	12174	0.99
06	12758	0.99	12755	0.99
07	13822	0.97	13782	0.98
08	13464	0.98	13446	0.98
10	-	-	11541	0.97
11	9232	0.96	-	-

**Table 10. 9: Participation Rates: Hispanic**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	2274	0.96	2275	0.99
04	2227	0.97	2233	0.99
05	2273	0.96	2283	0.99
06	2117	0.97	2117	0.99
07	2137	0.96	2132	0.99
08	2111	0.97	2102	0.99
10	-	-	1757	0.99
11	1380	0.97	-	-

**Table 10.10: Participation Rates: Asian/Pacific Islander**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	1156	0.95	1150	0.99
04	1145	0.96	1144	1.00
05	1174	0.97	1173	0.99
06	1046	0.97	1046	0.99
07	1009	0.96	1008	0.99
08	1021	0.97	1023	0.99
10	-	-	1053	0.99
11	994	0.96	-	-

**Table 10.11: Participation Rates: Native American/Alaskan**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	250	0.98	252	0.99
04	276	1.00	271	1.00
05	282	0.99	286	0.99
06	274	0.99	277	0.99
07	331	0.98	327	1.00
08	374	0.99	372	0.99
10	-	-	313	0.98
11	253	0.96	-	-

**Table 10.12: 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
03	28940	0.98	28918	0.99
04	28651	0.99	28590	0.99
05	28588	0.99	28520	0.99
06	28314	0.99	28389	0.99
07	29484	0.97	29481	0.98
08	28882	0.98	28834	0.98
10	-	-	21544	0.97
11	16312	0.96	-	-

**Table 10. 13: Participation Rates: Migrant Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	1941	0.97	1922	0.99
04	1415	0.99	1411	1.00
05	1217	0.98	212	0.99
06	720	0.97	719	0.99
07	595	0.96	580	0.99
08	558	0.96	554	0.98
10	-	-	338	0.98
11	197	0.96	-	-

**Table 10. 14: Participation Rates: Special Education (IEP) Students**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	10451	0.95	10459	0.96
04	10556	0.96	10572	0.96
05	10340	0.96	10353	0.96
06	9930	0.95	9934	0.95
07	10558	0.93	10532	0.95
08	10500	0.95	10518	0.95
10	-	-	9148	0.94
11	7403	0.91	-	-

**Table 10. 15: Participation Rates: English Learners (LEP, ELL)**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	1932	0.93	1924	0.99
04	1744	0.92	1747	0.99
05	1651	0.92	1672	0.99
06	1370	0.93	1376	0.99
07	1288	0.90	1286	0.99
08	1227	0.93	1207	0.99
10	-	-	977	0.99
11	730	0.91	-	-

**Table 10. 16: Participation Rates: Students with Disabilities**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	10493	0.95	10472	0.96
04	10531	0.96	10552	0.96
05	10327	0.96	10336	0.96
06	9878	0.95	9883	0.95
07	10515	0.93	10465	0.95
08	10486	0.95	10497	0.95
10	-	-	9068	0.94
11	7336	0.91	-	-

**Table 10. 17: Participation Rates: Students Receiving Accommodations**

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics
03	6397	0.98	6466	0.99
04	7131	0.99	7374	1.00
05	7573	0.99	7839	1.00
06	7123	0.99	7358	0.99
07	7848	0.97	8049	0.99
08	7842	0.98	8118	0.99
10	-	-	6355	0.98
11	4797	0.96	-	-

**Table 10. 18: Number and Percentage of Students Receiving Accommodations by Accommodation Type, MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Oral reading	198	0.3	3953	6.05
	Signing of assessment	8	0.01	17	0.03
	Paraphrasing	18	0.03	15	0.02
	Other administration	134	0.21	91	0.14
	Oral reading in native language	5	0.01	150	0.23
	Extend time-TerraNova session	2659	4.07	2506	3.84
	Administer using > allotted periods	1925	2.95	1826	2.8
	Other timing	473	0.72	479	0.73
	Use of scribe	1941	2.97	1736	2.66
	Use of calculator, math table, etc.	144	0.22	840	1.29
	Use of bilingual dictionary	8	0.01	42	0.06
	Other response	77	0.12	77	0.12
	Testing individually	1752	2.68	1654	2.53
	Testing in small group	4611	7.06	4747	7.27
	Other setting	263	0.4	248	0.38
4	Oral reading	166	0.25	4531	6.89
	Signing of assessment	9	0.01	24	0.04
	Paraphrasing	27	0.04	25	0.04
	Other administration	134	0.2	116	0.18
	Oral reading in native language	4	0.01	118	0.18
	Extend time-TerraNova session	2948	4.48	2928	4.45
	Administer using > allotted periods	2076	3.15	2070	3.15
	Other timing	492	0.75	495	0.75
	Use of scribe	1946	2.96	1796	2.73
	Use of calculator, math table, etc.	176	0.27	1403	2.13
	Use of bilingual dictionary	4	0.01	47	0.07
	Other response	86	0.13	106	0.16
	Testing individually	1801	2.74	1749	2.66
	Testing in small group	5301	8.06	5558	8.45
	Other setting	271	0.41	276	0.42

**Table 10. 18: Number and Percentage of Students Receiving Accommodations by Accommodation Type, MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
5	Oral reading	192	0.29	4475	6.71
	Signing of assessment	3	0	19	0.03
	Paraphrasing	17	0.03	21	0.03
	Other administration	131	0.2	122	0.18
	Oral reading in native language	3	0	173	0.26
	Extend time-TerraNova session	3105	4.66	3065	4.6
	Administer using > allotted periods	2400	3.6	2372	3.56
	Other timing	507	0.76	535	0.8
	Use of scribe	1582	2.37	1495	2.24
	Use of calculator, math table, etc.	251	0.38	1897	2.85
	Use of bilingual dictionary	16	0.02	68	0.1
	Other response	93	0.14	85	0.13
	Testing individually	1557	2.34	1547	2.32
	Testing in small group	6011	9.02	6176	9.27
	Other setting	309	0.46	318	0.48
6	Oral reading	251	0.37	3816	5.64
	Signing of assessment	6	0.01	30	0.04
	Paraphrasing	47	0.07	51	0.08
	Other administration	67	0.1	57	0.08
	Oral reading in native language	11	0.02	114	0.17
	Extend time-TerraNova session	2696	3.98	2661	3.93
	Administer using > allotted periods	1958	2.89	1920	2.84
	Other timing	539	0.8	567	0.84
	Use of scribe	1071	1.58	870	1.29
	Use of calculator, math table, etc.	337	0.5	2674	3.95
	Use of bilingual dictionary	24	0.04	71	0.1
	Other response	65	0.1	51	0.08
	Testing individually	1116	1.65	986	1.46
	Testing in small group	5917	8.74	6195	9.15
	Other setting	165	0.24	144	0.21

**Table 10. 18: Number and Percentage of Students Receiving Accommodations by Accommodation Type, MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
7	Oral reading	269	0.38	3805	5.32
	Signing of assessment	6	0.01	22	0.03
	Paraphrasing	72	0.1	71	0.1
	Other administration	108	0.15	106	0.15
	Oral reading in native language	11	0.02	112	0.16
	Extend time-TerraNova session	2716	3.79	2777	3.88
	Administer using > allotted periods	1986	2.77	1921	2.69
	Other timing	414	0.58	516	0.72
	Use of scribe	851	1.19	590	0.82
	Use of calculator, math table, etc.	510	0.71	3403	4.76
	Use of bilingual dictionary	13	0.02	86	0.12
	Other response	48	0.07	44	0.06
	Testing individually	860	1.2	758	1.06
	Testing in small group	6654	9.29	6822	9.54
	Other setting	135	0.19	132	0.18
8	Oral reading	240	0.33	3695	5.03
	Signing of assessment	3	0	24	0.03
	Paraphrasing	70	0.1	67	0.09
	Other administration	114	0.16	123	0.17
	Oral reading in native language	11	0.01	152	0.21
	Extend time-TerraNova session	2683	3.65	2610	3.55
	Administer using > allotted periods	1973	2.69	1891	2.57
	Other timing	449	0.61	497	0.68
	Use of scribe	664	0.9	568	0.77
	Use of calculator, math table, etc.	581	0.79	3766	5.12
	Use of bilingual dictionary	14	0.02	61	0.08
	Other response	52	0.07	39	0.05
	Testing individually	774	1.05	769	1.05
	Testing in small group	6620	9.01	6822	9.28
	Other setting	158	0.22	130	0.18

**Table 10. 18: Number and Percentage of Students Receiving Accommodations by Accommodation Type, MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
HS	Oral reading	172	0.28	2623	3.79
	Signing of assessment	9	0.01	34	0.05
	Paraphrasing	29	0.05	44	0.06
	Other administration	64	0.1	85	0.12
	Oral reading in native language	5	0.01	93	0.13
	Extend time-TerraNova session	1817	2.95	2257	3.26
	Administer using > allotted periods	786	1.28	1110	1.6
	Other timing	324	0.53	355	0.51
	Use of scribe	272	0.44	243	0.35
	Use of calculator, math table, etc.	165	0.27	3037	4.39
	Use of bilingual dictionary	6	0.01	32	0.05
	Other response	49	0.08	34	0.05
	Testing individually	337	0.55	385	0.56
	Testing in small group	4006	6.51	5343	7.72
	Other setting	131	0.21	124	0.18

**Table 10. 19: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille-Edition MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Oral reading	2	25		
	Signing of assessment	1	12.5		
	Paraphrasing	1	12.5		
	Other administration	2	25		
	Oral reading in native language	1	12.5		
	Extend time-TerraNova session	2	25	1	16.67
	Administer using > allotted periods	1	12.5		
	Use of scribe	2	25	2	33.33
	Testing individually	2	25	3	50
	Testing in small group	2	25	2	33.33
4	Oral reading	1	20	2	33.33
	Extend time-TerraNova session	2	40	2	33.33
	Administer using > allotted periods	1	20	1	16.67
	Use of scribe	2	40	2	33.33
	Use of calculator, math table, etc.			1	16.67
	Other response	2	40	2	33.33
	Testing individually	2	40	2	33.33
5	Large Print edition	1	10		
	Oral reading	2	20	2	25
	Signing of assessment	1	10		
	Paraphrasing	1	10		
	Other administration	1	10		
	Oral reading in native language	1	10		
	Extend time-TerraNova session	4	40	3	37.5
	Administer using > allotted periods	1	10		
	Other timing	2	20	1	12.5
	Use of scribe	6	60	4	50
	Use of calculator, math table, etc.	1	10	1	12.5
	Use of bilingual dictionary	1	10		
	Other response	1	10		
	Testing individually	7	70	5	62.5
	Testing in small group	2	20	1	12.5
6	Oral reading	2	50	2	40
	Extend time-TerraNova session	1	25	1	20
	Administer using > allotted periods	1	25	1	20
	Use of scribe	2	50	3	60
	Testing individually	4	100	4	80
	Testing in small group			1	20

**Table 10. 19: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Braille-Edition MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
7	Oral reading			1	20
	Other administration	1	25	1	20
	Extend time-TerraNova session	1	25		
	Administer using > allotted periods	2	50	1	20
	Use of scribe	2	50	3	60
	Use of calculator, math table, etc.	1	25	3	60
	Testing individually	1	25	3	60
	Testing in small group	1	25	1	20
8	Oral reading			2	25
	Extend time-TerraNova session	3	42.86	5	62.5
	Administer using > allotted periods	4	57.14	4	50
	Use of scribe	2	28.57	3	37.5
	Use of calculator, math table, etc.	1	14.29	6	75
	Testing individually	4	57.14	4	50
	Testing in small group	3	42.86	4	50
HS	Administer using > allotted periods			1	14.29
	Extend time-TerraNova session	3	100	1	14.29
	Administer using > allotted periods			2	28.57
	Other timing			2	28.57
	Use of scribe	2	66.67	2	28.57
	Use of calculator, math table, etc.			1	14.29
	Testing individually	2	66.67	3	42.86
	Testing in small group	1	33.33	1	14.29

**Table 10. 20: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large-Print Edition MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
3	Oral reading	1	2.94	11	33.33
	Other administration	1	2.94	2	6.06
	Oral reading in native language			1	3.03
	Extend time-TerraNova session	5	14.71	4	12.12
	Administer using > allotted periods	11	32.35	10	30.3
	Other timing	2	5.88	1	3.03
	Use of scribe	14	41.18	15	45.45
	Use of calculator, math table, etc.	1	2.94	1	3.03
	Other response	1	2.94	1	3.03
	Testing individually	14	41.18	15	45.45
Testing in small group	11	32.35	10	30.3	
4	Oral reading	5	12.5	17	43.59
	Other administration	2	5	3	7.69
	Extend time-TerraNova session	8	20	8	20.51
	Administer using > allotted periods	8	20	9	23.08
	Other timing	1	2.5	1	2.56
	Use of scribe	13	32.5	15	38.46
	Use of calculator, math table, etc.	2	5	8	20.51
	Other response	2	5	1	2.56
	Testing individually	18	45	20	51.28
	Testing in small group	12	30	9	23.08
Other setting	3	7.5	1	2.56	
5	Braille edition	1	2.63		
	Oral reading	7	18.42	15	37.5
	Signing of assessment	1	2.63		
	Paraphrasing	1	2.63		
	Other administration	5	13.16	4	10
	Oral reading in native language	1	2.63		
	Extend time-TerraNova session	10	26.32	9	22.5
	Administer using > allotted periods	13	34.21	14	35
	Other timing	3	7.89	2	5
	Use of scribe	17	44.74	18	45
	Use of calculator, math table, etc.	2	5.26	8	20
	Use of bilingual dictionary	1	2.63		
	Other response	1	2.63		
Testing individually	16	42.11	19	47.5	
Testing in small group	17	44.74	16	40	
Other setting	1	2.63	1	2.5	

**Table 10. 20: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large-Print Edition MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
6	Oral reading			14	53.85
	Other administration	2	9.52	3	11.54
	Extend time-TerraNova session	6	28.57	11	42.31
	Administer using > allotted periods	6	28.57	5	19.23
	Other timing	3	14.29	2	7.69
	Use of scribe	9	42.86	11	42.31
	Use of calculator, math table, etc.	3	14.29	7	26.92
	Other response	2	9.52	2	7.69
	Testing individually	9	42.86	15	57.69
	Testing in small group	9	42.86	9	34.62
	Other setting	1	4.76	1	3.85
7	Oral reading	1	3.45	12	40
	Signing of assessment			1	3.33
	Oral reading in native language			1	3.33
	Extend time-TerraNova session	7	24.14	6	20
	Administer using > allotted periods	8	27.59	7	23.33
	Other timing			1	3.33
	Use of scribe	10	34.48	10	33.33
	Use of calculator, math table, etc.			5	16.67
	Testing individually	10	34.48	7	23.33
	Testing in small group	10	34.48	12	40
8	Oral reading	5	17.86	13	43.33
	Paraphrasing			1	3.33
	Extend time-TerraNova session	6	21.43	6	20
	Administer using > allotted periods	5	17.86	8	26.67
	Other timing	3	10.71	3	10
	Use of scribe	13	46.43	12	40
	Use of calculator, math table, etc.			8	26.67
	Testing individually	11	39.29	12	40
Testing in small group	8	28.57	10	33.33	
Other setting	1	3.57	1	3.33	

**Table 10. 20: Number and Percentage of Students Receiving Accommodations by Accommodation Type, Large-Print Edition MAP 2006**

Grade	Accommodation	Communication Arts		Mathematics	
		Frequency	Percent	Frequency	Percent
HS	Oral reading	2	11.76	4	21.05
	Signing of assessment			1	5.26
	Other administration			2	10.53
	Extend time-TerraNova session	2	11.76	7	36.84
	Administer using > allotted periods	2	11.76		
	Other timing	1	5.88	1	5.26
	Use of scribe	2	11.76	6	31.58
	Use of calculator, math table, etc.			4	21.05
	Other response			1	5.26
	Testing individually	2	11.76	4	21.05
	Testing in small group	3	17.65	8	42.11
	Other setting			1	5.26

## 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.
- CTB/McGraw-Hill. (2002). *TerraNova, The Second Edition*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2006). *2006 MAP Examiner's Manual*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2006). *2006 MAP Test Guide to Interpretation*. [Guide to Interpreting Results?] Monterey, CA: Author.
- Donoghue, John R., Paul W. Holland, & Dorothy T. Thayer. (1993). A Monte Carlo Study of Factors That Affect the Mantel-Haenszel and Standardization Measures of Differential Item Functioning. pp.137–66 in *Differential Item Functioning*, edited by Paul W. Holland and Howard Wainer. Hillsdale, NJ: Lawrence Erlbaum.
- Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 2, 217–233.
- Dorans, N.J., & Schmitt, M.P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. Princeton: Educational Testing Service.
- Green, D. R. (1975, December). Procedures for assessing bias in achievement tests. Presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.
- 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.
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39, 247–264.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems* (71, 179–181). Hillsdale, NJ: Lawrence Erlbaum.
- Macmillan/McGraw-Hill. (1993). *Reflecting diversity: Multicultural guidelines for educational publishing professionals*. New York: Author.
- Macmillan/McGraw-Hill. (1993). *Guidelines for Bias-Free Publishing*. 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.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230.
- 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., Burket, G. R., & Sykes, R. C. (1988). *Non-unique solutions[should be “solutions”?] to the likelihood equation for the three-parameter logistic model*. Paper presented at the meeting of the Psychometric Society, Los Angeles.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–213.

## APPENDIX

### Reviewing 2005 Bookmark Cut Scores with 2006 Operational Data (Phase I)

Karla L. Egan  
CTB/McGraw-Hill

A Bookmark standard setting was held in 2005 to establish cut scores for the Missouri Assessment Program (MAP). The materials used for this standard setting were based on data from the 2005 field test. When a standard setting is based on field-test data, it is recommended that the state department review the cut scores once operational data are available. To this end, CTB has compiled materials and analyses that the Missouri Department of Elementary and Secondary Education (DESE) may use to review and, if necessary, adjust the cut scores established in 2005.

This document examines the cut scores that were set in 2005 and adjusts those cut scores using 2006 data based on the calibration sample. First, we conducted an Item Order Analysis that compared how much items shifted position between 2005 and 2006. Next, we looked at the difference in Bookmark location values between 2005 and 2006. Then, we conducted a Cut Score Analysis that examined various ways that the cut scores could be adjusted. Finally, we set forth a system of cut scores that may be used by DESE on the MAP tests.

Because this study is being conducted using data from the calibration sample, the cut scores will be revisited once census data is available.

#### ***2005 MAP Standard Setting***

In December 2005, a modified Bookmark Standard Setting Procedure was used to establish cut scores for the Grades 3–8 and high school Communication Arts and Mathematics MAPs. A modification of Bookmark was used to meet the requirements of Senate Bill 1080, which required that NAEP-like cut scores be established for the MAPs.

Senate Bill 1080 was interpreted such that the *Proficient* achievement level met, but did not exceed, the NAEP performance standards. In other words, the percentage of students who attain *Proficient* on MAP should be similar to or slightly higher than the percentage attaining *Proficient* on NAEP. The percentage of students in the other three achievement levels would be allowed to vary between NAEP and MAP.

For the purposes of the MAP standard setting, participants were allowed to recommend *Proficient* cut scores within a pre-specified range. This range was based on the percentage of students who could be classified as either *Proficient* or *Advanced*. At the low end, no fewer than 26% of students could be classified as either *Proficient* or *Advanced*, and, at the high end, approximately 44% of students could be classified as *Proficient* or *Advanced*.

Table 1 summarizes the recommended cut scores and impact data from the 2005 MAP Standard Setting.

**Table 1. Recommended Cut Scores and Impact Data from 2005 MAP Standard Setting**

Content Area	Grade	Cut Score			Impact Data				
		Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.	Prof & Adv
Comm. Arts	3	587	642	668	10.1	46	26.5	17.4	43.9
	4	605	654	687	12.5	42.7	29.4	15.4	44.8
	5	620	669	699	10.1	45.2	29	15.7	44.7
	6	621	671	703	11	45.3	31.7	12	43.7
	7	622	674	712	14.9	41	31.7	12.4	44.1
	8	627	691	720	9.6	47.7	27.3	15.4	42.7
	11	650	711	745	12.3	43.9	30.5	13.3	43.8
Math	3	559	620	667	11.1	45.1	35.0	8.8	43.8
	4	589	643	677	11.6	44.3	32.1	12.0	44.1
	5	599	658	699	12.7	42.9	33.4	11.0	44.4
	6	613	671	710	13.4	42.3	32.6	11.7	44.3
	7	631	676	714	21.4	34.6	32.3	11.7	44
	8	658	702	741	24.6	35.7	28.2	11.5	39.7
	10	683	723	777	27.2	28.4	34.2	10.2	44.4

## Materials

For the purposes of this review, Bookmark locations were calculated for all 2006 operational items. (To see how to calculate Bookmark locations, please see Lewis et al 1996.) Briefly, the Bookmark locations are scale score values that show the necessary ability for a students to have a .55 probability of answering an item correctly<sup>1</sup>. Using these Bookmark locations, item maps were constructed for each grade/content area. These item maps ordered items according to their Bookmark locations. Table 2 shows the number of score points on each 2006 item map.

The 2006 item maps were compared with the 2005 item maps used at the December standard setting. The 2005 item maps contained 2006 as well as some 2007 items. Since the 2005 item maps contained more items than the 2006 maps, the analysis was limited to only those items that appeared on the 2006 operational form. This limitation was necessary since operational IRT parameters were only available for those items that appeared on the 2006 form.

<sup>1</sup> The use of the response probability (RP) of .55 was unique to the MAP standard setting. For an explanation of the reasons for using RP55, please see the *Missouri Assessment Program Final Bookmark Standard Setting Technical Report 2005*.

The analyses in this report were conducted on a calibration sample. The schools in this sample were pre-selected by DESE to be representative of the demographic composition of Missouri schools.

**Table 2. Number of Score Points on 2006 Item Maps**

<b>Content Area</b>	<b>Grade</b>	<b>Number of Score Points</b>
<b>Communication Arts</b>	3	69
	4	66
	5	65
	6	66
	7	72
	8	69*
	11	74
<b>Mathematics</b>	3	67
	4	77
	5	71
	6	69
	7	71
	8	76
	10	75

\*In Grade 8 Communication Arts, only 68 points were used in the analyses. One item on the Grade 8 2006 test did not appear in the Grade 8 2005 Ordered Item Booklet.

### ***Item Order Analyses***

The purpose of the *Item Order Analyses* was to assess how much the order of items changed between 2005 and 2006. The item order refers to the order of items and score points (in the case of CR items) in the Ordered Item Booklet (OIB). Only items that were on the 2006 operational test and that were in the 2005 OIB were included in the analyses. Two analyses were completed to compare the 2005 order of items to the 2006 order of items: (1) Spearman correlations of the 2005 and 2006 rank order; and (2) differences in the 2005 and 2006 rank order.

Table 3 shows the correlations between the 2005 and 2006 rank order of items. As demonstrated in the table, the correlations between the 2005 and 2006 rank orders are quite high. All correlation coefficients are .95 or greater. This suggests that the items stayed in the same relative order between the administration of the field test and the operational test.

**Table 3. Rank-Order Correlations**

Grade	Communication Arts	Mathematics
3	.98	.97
4	.98	.96
5	.98	.97
6	.96	.98
7	.98	.98
8	.98	.98
High School	.95	.96

The second set of analyses examined the amount of item drift in terms of rank order between 2005 and 2006. The absolute difference between the 2005 and 2006 rank order was computed. Tables 4 and 5 show the percentage of items in each grade that changed rank order by 0, 1 or fewer, 2 or fewer, etc., for up to 10 or fewer places. Within a particular row, a higher percentage reflects less change in item order than a lower percentage. For example, in Grade 5 Communication Arts (see Table 4), 47.7% of the items moved less than 1 place in rank order while only 18.3% of the Grade 5 Mathematics (see Table 5) items moved less than 1 place. This means that the items on the Grade 5 Mathematics test changed order more between the field test and the operational test than did items on the Grade 5 Communication Arts test.

On almost all Communication Arts tests (see Table 4), 94% of the items moved 8 or fewer places in terms of rank order. Grade 11 Communication Arts saw the greatest change in item order; even so, 97% of the items moved no more than 14 places in terms of rank order. On all Mathematics tests (see Table 5), 90% of the items moved 10 or fewer places in terms of rank order.

The movement of items merely indicates that there was a difference in performance between the field test and the operational test. The greater movement at a particular grade may indicate that students (and perhaps teachers) took the operational test more seriously than they did the field test. The amount of movement (or lack thereof) does not indicate a problem with either test.

**Table 4. Cumulative Percentage of Items Changing Rank Order, Communication Arts**

Change in Rank Order	Communication Art						
	3	4	5	6	7	8	11
=0	14.5	7.6	23.1	12.1	9.7	10.3	8.1
≤1	34.8	31.8	47.7	22.7	38.9	29.4	23.0
≤2	49.3	62.1	70.8	47.0	55.6	51.5	39.2
≤3	63.8	71.2	78.5	66.7	70.8	63.2	48.6
≤4	79.7	83.3	87.7	77.3	80.6	70.6	54.1
≤5	84.1	89.4	89.2	80.3	84.7	77.9	59.5
≤6	89.9	95.5	93.8	84.8	86.1	82.4	67.6
≤7	91.3	97.0	95.4	87.9	90.3	91.2	75.7
≤8	94.2	98.5		97.0	94.4	94.1	81.1
≤9	95.7		96.9		97.2	97.1	82.4
≤10	97.1		98.5				87.8

**Table 5. Cumulative Percentage of Items Changing Rank Order, Mathematics**

Change in Rank Order	Mathematics						
	3	4	5	6	7	8	10
=0	16.4	11.7	5.6	5.8	7.0	18.4	13.3
≤1	35.8	20.8	18.3	29.0	28.2	35.5	25.3
≤2	53.7	50.6	36.6	49.3	49.3	48.7	32.0
≤3	64.2	54.5	60.6	68.1	60.6	60.5	44.0
≤4	71.6	63.6	67.6	79.7	69.0	65.8	53.3
≤5	80.6	72.7	78.9	88.4	78.9	81.6	58.7
≤6	86.6	76.6	81.7	92.8	83.1	88.2	70.7
≤7	91.0	80.5	91.5	95.7	91.5	92.1	82.7
≤8	94.0	85.7	94.4	98.6	93.0	93.4	85.3
≤9	95.5	88.3	97.2		95.8	96.1	90.7
≤10		90.9			97.2		

### *Difference in Bookmark Location Values*

To better understand how performance changed between the 2005 field test and the 2006 operational test, we examined the difference between the Bookmark location values. If the 2006 Bookmark location values were routinely higher than the 2005 values, this would mean that the items had become more difficult for students. On the other hand, lower Bookmark values in 2006 would indicate that the students improved their performance on the items.

The 2005 Bookmark location values were subtracted from the 2006 values: negative values indicate that the items were easier in 2006, and positive values indicate that the items were more difficult in 2006. Table 6 reports the results of the difference between the 2006 and 2005 Bookmark location values. Table 6 shows the minimum and maximum difference as well as the average difference and the standard deviation of the difference. Across both content areas and in all grades, the Bookmark location values, on average, decreased. The size of the decrease was, however, in some cases quite small (see Grade 3 Communication Arts).

**Table 6. Average Difference Between 2006 and 2005 Bookmark Location Values**

Content Area	Grade	Number of Items	Minimum	Maximum	Mean	Std. Deviation
Communication Arts	3	69	-63.5	28.0	-0.6	13.5
	4	66	-45.5	15.9	-3.8	9.3
	5	65	-28.1	11.6	-1.1	6.7
	6	66	-59.2	12.4	-3.9	11.7
	7	72	-68.9	16.9	-6.4	14.0
	8	68	-31.8	20.7	-2.7	9.3
	11	74	-67.2	26.8	-7.2	18.9
Mathematics	3	67	-76.9	30.4	-4.5	14.0
	4	77	-32.8	18.9	-2.7	9.3
	5	71	-37.2	25.9	-2.3	10.7
	6	69	-26.1	8.1	-4.5	7.6
	7	71	-37.2	21.6	-5.4	9.6
	8	76	-58.4	27.8	-9.9	13.4
	10	75	-53.6	32.1	-3.3	13.2

### *Cut Score Analyses*

The purpose of the *Cut Score Analyses* was to find a set of cut scores that minimized the number of items switching achievement levels and that minimized the change to the 2005 impact data, in order to find cut scores that most closely align with the 2005 participant recommendations.

To accomplish this, three sets of cut scores were applied to the 2006 item maps: (1) the original 2005 cut scores (CS method), (2) the cut scores resulting from the 2005 impact data (ID method), and (3) the cut scores that minimize changes to content (Content method). For the ID method, cut scores were found that most closely aligned with the 2005 impact data. For the Content method, items were assigned to achievement level based on their 2005 Bookmark locations. Then, the items were ordered based on their 2006 Bookmark locations. Finally, cut scores were identified so that the number of items switching achievement levels was minimized.

Once cut scores were identified and applied, impact data resulting from those cut scores was calculated. Tables 7 and 8 summarize the cut scores and impact data that resulted after applying the CS, ID, and Content methods to the 2006 data.

**Table 7. Cut Scores and Impact Data, Communication Arts**

Method	Grade	Cut Score			Impact Data				
		<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Prof. &amp; Adv.</i>
<b>CS Method</b>	<b>3</b>	587	642	668	8.6	44.5	25.6	21.2	46.9
	<b>4</b>	605	654	687	9.2	37.6	35.1	18.1	53.2
	<b>5</b>	620	669	699	8.1	42.1	34.0	15.9	49.9
	<b>6</b>	621	671	703	7.0	41.1	40.4	11.5	51.9
	<b>7</b>	622	674	712	8.0	38.6	40.4	12.9	53.3
	<b>8</b>	627	691	720	4.8	46.5	32.7	15.9	48.7
	<b>11</b>	650	711	745	2.9	34.5	44.3	18.3	62.7
<b>ID Method</b>	<b>3</b>	593	648	673	10.3	46.9	26.0	16.8	42.8
	<b>4</b>	616	662	696	12.8	42.3	30.8	14.1	44.9
	<b>5</b>	626	675	702	10.5	46.2	27.5	15.9	43.4
	<b>6</b>	634	676	706	11.5	43.3	33.7	11.5	45.2
	<b>7</b>	639	684	715	14.9	42.0	30.2	12.9	43.1
	<b>8</b>	645	699	723	9.3	49.0	25.7	15.9	41.7
	<b>11</b>	685	725	753	12.7	43.3	31.4	12.6	44.0
<b>Content Method</b>	<b>3</b>	592	640	669	10.3	39.0	33.9	16.8	50.7
	<b>4</b>	606	650	680	9.2	34.1	34.8	22.0	56.8
	<b>5</b>	617	669	693	7.2	42.9	27.4	22.5	49.9
	<b>6</b>	626	676	704	7.9	46.9	33.7	11.5	45.2
	<b>7</b>	622	671	712	8.0	35.4	43.7	12.9	56.6
	<b>8</b>	624	689	718	4.8	43.6	31.9	19.7	51.6
	<b>11</b>	649	724	737	2.9	53.2	16.3	27.6	44.0

**Table 8. Cut Scores and Impact Data, Mathematics**

Method	Grade	Cut Score			Impact Data				
		<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Prof. &amp; Adv.</i>
CS Method	3	559	620	667	5.6	46.9	38.9	8.5	47.5
	4	589	643	677	6.6	41.2	36.7	15.6	52.2
	5	599	658	699	8.2	39.9	37.9	14.0	51.9
	6	613	671	710	5.4	39.6	38.9	16.0	54.9
	7	631	676	714	10.5	31.7	38.4	19.4	57.7
	8	658	702	741	12.0	37.7	38.9	11.4	50.3
	10	683	723	777	17.9	25.7	43.5	12.9	56.4
ID Method	3	576	625	671	11.7	44.9	34.9	8.5	43.4
	4	603	651	683	12.0	44.7	30.8	12.6	43.3
	5	611	665	706	12.0	42.0	35.0	11.0	45.9
	6	637	681	721	13.9	42.1	33.0	11.0	44.0
	7	651	689	726	21.4	35.2	31.8	11.7	43.5
	8	677	712	743	23.6	35.6	29.4	11.4	40.8
	10	701	738	785	27.5	28.7	33.2	10.6	43.8
Content Method	3	563	616	667	7.1	38.4	46.0	8.5	54.5
	4	586	644	675	5.9	41.8	36.7	15.6	52.2
	5	601	661	735	8.2	42.6	46.2	3.0	49.2
	6	611	672	713	4.7	43.1	39.1	13.1	52.2
	7	628	674	712	9.2	33.0	38.4	19.4	57.7
	8	661	701	729	13.5	33.6	30.7	22.2	52.9
	10	681	724	774	16.0	29.4	40.0	14.6	54.6

**Evaluation of the Cut Score Analyses**

Each set of cut scores was evaluated by (1) calculating the number of items that switched achievement levels (Item Switch) and (2) calculating the difference between the 2005 and 2006 percentage of students *Basic* and above, *Proficient* and above, and *Advanced* (Impact).

Table 9 summarizes the number of items that switched achievement levels and the differences in impact data for each set of cut scores applied from the different methods for Communication Arts. The columns labeled “Item Switch” show the number of items that switched achievement levels when a set of cut scores was applied from the different methods. For example, in Grade 7 Communication Arts, four items switched from the *Basic* achievement level into another achievement level when the cut scores from the CS and Content methods were applied. In the same grade, using the ID method, 10 items switched from the *Basic* achievement level into another achievement level. When low numbers of items switch achievement levels, this indicates a better alignment of the content expectations to the original expectations recommended by the 2005 participants.

The columns labeled “Impact” in Table 9 show the difference between the 2005 and 2006 percentages of students classified as *Basic* and above, *Proficient* and above, and *Advanced*. A positive number represents an increase between 2005 and 2006, and a negative number indicates a decrease from 2005 to 2006. In Grade 7 Communication Arts, 6.9% more students were classified *Basic* or above in 2006 than in 2005 when the CS and Content methods were applied. Smaller numbers in the “Impact” column indicate that the impact data is close to that recommended by participants.

When the CS method was applied, relatively few items switched achievement levels, and there was an increase in the percentage of students classified as *Proficient* or *Advanced*. As Table 9 shows, the percentage of students classified as *Proficient* or above increased anywhere from 3.0% (Grade 3) to 18.9% (Grade 11).

Although the ID method minimized changes to the impact data, there were some fairly large numbers of items switching achievement levels. For example, in Grade 7, 21 (10 + 7 + 4) items switched achievement levels between 2005 and 2006. This represents almost 30% of the items switching levels.

The Content method reversed the situation from the ID method. With the Content method, the number of items switching achievement levels was minimized; however, there were some fairly sizable changes to the impact data. In Grade 7, 12.5% more students would be classified as *Proficient* and above using the Content method.

**Table 9. Summary of Item Switches and Differences Between 2005/2006 Impact Data by Method, Communication Arts**

Method	Grade	Item Switch				Impact		
		<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Basic &amp; above</i>	<i>Prof. &amp; above</i>	<i>Adv.</i>
CS Method	3	4	3	3	2	1.5	3.0	3.8
	4	2	4	3	3	3.3	8.4	2.7
	5	0	3	5	3	2.0	5.2	0.2
	6	0	0	4	3	4.0	8.2	-0.5
	7	0	4	3	3	6.9	9.2	0.5
	8	0	2	4	3	4.8	6.0	0.5
	11	0	8	4	3	9.4	18.9	5.0
ID Method	3	2	3	5	4	-0.2	-1.1	-0.6
	4	0	9	4	6	-0.3	0.1	-1.3
	5	0	3	6	4	-0.4	-1.3	0.2
	6	0	2	3	4	-0.5	1.5	-0.5
	7	0	10	7	4	0.0	-1.0	0.5
	8	0	12	6	4	0.3	-1.0	0.5
	11	0	14	6	6	-0.4	0.2	-0.7
Content Method	3	2	3	1	2	-0.2	6.8	-0.6
	4	0	5	2	1	3.3	12.0	6.6
	5	0	2	6	0	2.9	5.2	6.8
	6	0	0	4	3	3.1	1.5	-0.5
	7	0	4	0	3	6.9	12.5	0.5
	8	0	1	1	2	4.8	8.9	4.3
	11	0	2	5	1	9.4	0.2	14.3

Table 10 summarizes the number of items that switched achievement levels and the differences in impact data for each set of cut scores applied from the different methods for Mathematics. When the CS method was applied, relatively few items switched achievement levels (except in Grade 8 *Advanced*, where 11 items switched achievement levels) and there was increase in the percentage of students classified as *Proficient* or *Advanced*. As Table 7 shows, the percentage of students classified as *Proficient* or above increased anywhere from 3.7% (Grade 3) to 13.7% (Grade 7).

Again the ID cuts minimized changes to the impact; however, there were some fairly large numbers of items switching achievement levels. For example, in Grade 8, 32 (8 + 12 + 12) items switched achievement levels between 2005 and 2006. This represents almost 42% of the items switching levels.

The Content method again reversed the situation from the ID method. With the Content method, the number of items switching achievement levels was minimized; however, there were some fairly sizable changes to the impact data. In Grade 7, 13.7% more students would be classified as *Proficient* and above using the Content method.

**Table 10. Summary of Item Switches and Differences Between 2005/2006 Impact Data by Method, Mathematics**

Method	Grade	Item Switch				Impact		
		<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Basic &amp; above</i>	<i>Prof. &amp; above</i>	<i>Adv.</i>
CS Method	3	5	0	8	1	5.5	3.7	-0.3
	4	1	5	1	2	5.0	8.1	3.6
	5	1	0	2	3	4.5	7.5	3.0
	6	0	1	1	0	8.0	10.6	4.3
	7	0	2	6	4	10.9	13.7	7.7
	8	1	0	3	11	12.6	10.6	-0.1
	10	1	5	5	4	9.3	12.0	2.7
ID method	3	0	7	10	1	-0.6	-0.4	-0.3
	4	0	17	8	2	-0.4	-0.8	0.6
	5	0	4	6	3	0.7	1.5	0.0
	6	0	11	14	4	-0.5	-0.3	-0.7
	7	0	8	10	11	0.0	-0.5	0.0
	8	0	8	12	12	1.0	1.1	-0.1
	10	0	4	18	5	-0.3	-0.6	0.4
Content Method	3	2	2	3	1	4.0	10.7	-0.3
	4	2	3	1	1	5.7	8.1	3.6
	5	1	0	2	3	4.5	4.8	-8.0
	6	0	0	1	0	8.7	7.9	1.4
	7	0	2	3	2	12.2	13.7	7.7
	8	0	1	2	1	11.1	13.2	10.7
	10	1	3	6	1	11.2	10.2	4.4

### *Finding the Final Cut Scores*

To find the final cut scores given the operational data, DESE must decide how to balance the gains in student performance observed between the field test and operational test, the constraints of Senate Bill 1080, the pattern of vertical articulation suggested by the participants, and the vertical scale.

To this end, decisions must be made as to whether it is more important to minimize changes to the content required to reach an achievement level, to minimize changes to the impact data, or to find a compromise between the two positions.

### **Gains in Student Performance**

Student performance improved between the 2005 field test and 2006 operational test. This is indicated by both the higher percentages of students who attained *Proficient* or above when the field-test cut scores were applied to operational data, and the decrease, on average, in Bookmark locations between 2005 and 2006. When 2005 cut scores were applied to the 2006 operational data, an increase in the percentage of students classified as *Proficient* or *Advanced* was observed in each grade. Grade 11 Comm. Arts enjoyed the most dramatic improvement, with almost 19% of students classified as *Proficient* or *Advanced*. On average, Comm. Arts saw an 8.4% increase in the percentage of students classified as *Proficient* or *Advanced*, and Math a 9.5% increase.

The improved student performance may have been due to better motivation on the part of the students and/or teachers on the operational test or, perhaps, teachers and schools better aligned their curricula to meet state standards between the administration of the field test and operational test. Better motivation on an operational test probably does not reflect true improvement in content knowledge; however, improved curriculum alignment most likely would result in improved content knowledge. Determining how much of the gain was due to true growth versus how much was due to better motivation is, if not an impossible task, a very difficult one.

### **Senate Bill 1080**

Given Senate Bill 1080 and its interpretation for standard setting, it may be argued that it is necessary to preserve the impact data for, at least, the *Proficient* cut score. Standard setting participants were not allowed to have more than 44% of students classified as *Proficient* or above. Almost all standard setting groups placed their cut score at the lowest end of the range (to reach the 44% mark). Only Grade 8 Mathematics went into the range to set a cut score; even then, they recommended a cut score for which nearly 40% of students would be classified as *Proficient* or above. For the *Basic* and *Advanced* cut scores, however, standard setting participants were not restricted in setting their cut scores. This may support the argument for minimizing changes to content in the areas of the *Basic* and *Advanced* achievement levels.

### **Vertical Articulation**

During the standard setting process, the participants worked to smooth the data; in other words, the participants recommended a pattern of data that resulted in vertically-articulated data. DESE may want to consider the pattern of impact data that resulted from the standard setting when smoothing the data on the operational test. DESE may want to replicate this pattern during the smoothing process. Communication Arts participants recommended an equipercentile model in which similar numbers of students were classified as *Below Basic*, *Basic*, and *Proficient* or above. For the *Advanced* achievement level, participants recommended a declining model in which more students were classified as *Advanced* in Grade 3 than in Grade 11. In Mathematics, participants recommended an equipercentile model for the *Proficient* and *Advanced* levels (although they did note that they expected fewer students to be classified as *Advanced* in Grade 3). For the *Below Basic* level, they recommended an increasing model in which the percentage of students classified as *Below Basic* would increase between Grade 3 and Grade 10.

### Vertical Scale

The MAP tests are on a vertical scale; therefore, DESE must be cognizant of the cut scores themselves during the smoothing process. The cut scores must increase from grade to grade. DESE must make sure that the cut scores increase across the grades during the smoothing process.

Given these considerations, the following assumptions were used in selecting a set of cut scores for the Communication Arts and Mathematics MAPs:

- Senate Bill 1080 requires NAEP-like cut scores; therefore, the *Proficient* cut score should be constrained as it was for the standard setting. In other words, the *Proficient* cut score will be based on the impact data obtained from the 2005 standard setting.
- Participants were allowed to place the *Basic* cut score; therefore, a compromise position will be found that seeks to minimize changes to content and impact data for the *Basic* cut score.
- Participants were allowed to place the *Advanced* cut score; therefore, a compromise position will be found that seeks to minimize changes to content and impact data for the *Advanced* cut score.
- The cut scores will increase across the grades to meet the constraints of the vertical scale.
- The pattern of vertical articulation suggested by the participants will be retained.
- The same logic should be applied across all grades within both content areas.

Table 11 shows the resulting set of cut scores. It was not possible to meet all assumptions in all cases. For both content areas, the downward shift of Bookmark item locations in some grades made it very difficult to find a middle ground between content and impact recommendations. This problem was exacerbated in some cases because of gaps in the item locations at the high or low end of the scale for which there was sparse or no coverage. In such cases where it was difficult to find a middle ground between content and impact recommendations, we sought to maintain the pattern of impact data suggested by the standard setting participants.

**Table 11. Suggested System of Cut Scores**

Content Area	Grade	Cut Score			Impact Data				
		Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.	Prof & Adv
Comm. Arts	3	592	648	673	10.3	46.9	26.0	16.8	42.8
	4	612	662	691	11.7	43.4	30.8	14.1	44.9
	5	625	675	702	10.5	46.2	27.5	15.9	43.4
	6	631	676	704	10.4	44.4	33.7	11.5	45.2
	7	634	684	712	12.2	44.7	30.2	12.9	43.1
	8	645	699	720	9.3	49.0	25.7	15.9	41.7
	11	679	725	753	10.1	45.9	31.4	12.6	44.0
Math	3	568	625	667	9.4	47.2	34.9	8.5	43.4
	4	596	651	688	8.9	47.8	33.5	9.8	43.3
	5	605	665	706	9.6	44.5	35.0	11.0	45.9
	6	628	681	721	9.7	46.2	33.0	11.0	44.0
	7	645	689	726	17.8	38.7	31.8	11.7	43.5
	8	671	712	741	19.3	39.9	29.4	11.4	40.8
	10	695	738	785	24.0	32.2	33.2	10.6	43.8

**Finding the Suggested System of Cut Scores**

This section explains the logic used to determine the cut scores for a particular grade/content area. In Grade 3 and 5 Communication Arts, the *Basic* cut score was set to balance content and impact recommendations. The *Advanced* cut score was set using impact data due to sparse coverage of items in the region of the cut.

In Grade 4, 6, and 7 Communication Arts, the *Basic* and *Advanced* cut scores were set to balance content and impact recommendations.

In Grade 8 Communication Arts, the downward shift of item locations prohibited balancing content and impact concerns; thus, the *Basic* cut score was set to maintain a similar pattern of impact data as that from the original standard setting. The *Advanced* cut is the original recommended cut score.

In Grade 11 Communication Arts, the downward shift of item locations prohibited balancing content and impact concerns; thus, the *Basic* and *Advanced* cut scores were set to maintain a similar pattern of impact data as that from the original standard setting.

In Grade 3 Mathematics, the *Basic* cut score was set as a trade-off between content and impact. The *Advanced* cut score is the original recommended cut score.

In Grade 4 Mathematics, the downward shift of item locations prohibited balancing content and impact concerns; thus, the *Basic* cut score was set to maintain a pattern of impact data similar to that from the original standard setting. The *Advanced* cut score was set to maintain a pattern of impact data similar to that from the standard setting. This was necessitated by the sparseness of item coverage at the upper end of the scale.

In Grade 5 Mathematics, the *Basic* cut score was set to balance content and impact recommendations. We also maintained the pattern of impact data. The lack of item coverage at the upper end of the scale necessitated that the *Advanced* cut score be set using the ID method.

In Grade 6, 7, and 8 Mathematics, the downward shift of item locations required that the *Basic* and *Advanced* cut scores be set so that the pattern of impact data was maintained. In other words, Table 1 shows that similar numbers of students were classified as *Below Basic* between Grades 5 and 6; furthermore, Table 1 also shows that the percentage of students classified as *Below Basic* increased 7% between Grades 6 and 7. Finally, Table 1 shows that the percentage of *Below Basic* students leveled off again between Grades 7 and 8. During the smoothing process, we followed these patterns..

In Grade 10, the sparseness of item coverage at the low end of the scale necessitated that the *Basic* cut score be established so that the pattern of impact data was maintained. Finally, the *Advanced* cut score was set so that the pattern of impact data was maintained.

#### **Evaluation of the Suggested System of Cut Scores**

Table 12 summarizes the number of items that switched achievement levels and the differences in impact data for the cut scores found in Table 11. The columns labeled “Item Switch” show the number of items that switched achievement levels when a set of cut scores was applied from the different methods. The columns labeled “Impact” show the difference between the percentage of students classified as *Basic* and above, *Proficient* and above, and *Advanced* in 2005 and in 2006.

Table 12 shows that in Communication Arts, the compromise cut scores resulted in fairly small changes to the impact data. In Grades 3–7, the numbers of items switching achievement levels were fairly low. In Grades 8 and 11, over 30% of the items switched achievement levels.

Table 12 shows that in Mathematics, the compromise cut scores resulted in, at least, a 3% increase in the percentage of students classified as *Basic* and above in most grades. The percentages classified as *Proficient* and above, and *Advanced* stayed flat using the compromise method. In Mathematics, it was often harder to reach a compromise between the content and impact data recommendations. This is reflected in the larger numbers of items switching achievement levels in Mathematics.

**Table 12. Summary of Item Switches and Differences Between 2005/2006 Impact Data for Suggested System of Cut Scores in Table 11**

Content Area	Grade	Item Switch				Impact		
		<i>Below Basic</i>	<i>Basic</i>	<i>Prof.</i>	<i>Adv.</i>	<i>Basic &amp; above</i>	<i>Prof. &amp; above</i>	<i>Adv.</i>
Comm. Arts	3	2	2	5	4	-0.2	-1.1	-0.6
	4	0	7	4	6	0.8	0.1	-1.3
	5	0	2	6	4	-0.4	-1.3	0.2
	6	0	2	4	3	0.6	1.5	-0.5
	7	0	6	8	3	2.7	-1.0	0.5
	8	0	12	6	4	0.3	-1.0	0.5
	11	0	13	6	6	2.2	0.2	-0.7
Math	3	0	4	10	1	1.7	-0.4	-0.3
	4	0	11	8	2	2.7	-0.8	-2.2
	5	1	2	6	3	3.1	1.5	0.0
	6	0	8	14	4	3.7	-0.3	-0.7
	7	0	6	10	11	3.6	-0.5	0.0
	8	0	6	12	12	5.3	1.1	-0.1
	10	0	3	18	5	3.2	-0.6	0.4

## Conclusions

This document examined the relationship between the 2005 and 2006 MAP standard setting data. Although the order of the items did not change a great deal between the 2005 field test and the 2006 operational test, the students performed better on the 2006 operational test as reflected in the decreased Bookmark location values. This type of improvement is not completely unexpected given that 2005 was a field test and 2006 was an operational test.

The main purpose of this document was to suggest a system of cut scores for the 2006 MAP, based as closely as possible on the recommendations from the 2005 standard setting. A set of recommendations (see Table 11) was set forth and the assumptions behind those recommendations were explicated. These recommendations should not be considered final; rather, they are meant to give DESE staff something to reflect on and react to.

DESE staff should consider the stated assumptions and determine whether they agree with their own assumptions regarding the cut scores and impact data. If DESE's assumptions differ, then a different set of compromise cut scores may be found based on DESE's assumptions.

DESE may, on the other hand, choose to accept one of the sets of cut scores from the Cut Score Analyses. DESE may choose to accept the 2005 cut scores as they were recommended. DESE

may, instead, choose to minimize changes to the impact data and accept the cut scores associated with the ID method. Finally, DESE may choose to minimize changes to content and accept the cut scores associated with the Content method. Each set of cut scores reflects a different policy decision that DESE may make.

In addition, we must keep in mind that final cut scores cannot be adopted until census data are available. In other words, DESE should plan to apply the cut scores adopted from this study to the census data to confirm that the results are logical and consistent with expectations.

In the end, the decision to accept or adjust the 2006 cut scores to reflect the 2005 standard setting is a policy decision. CTB will assist DESE in adjusting the results of this paper further so that DESE's own assumptions are met.

## **Reference**

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring*. Symposium conducted at the Council of Chief State School Officers National Conference on Large-scale Assessment, Phoenix, AZ.