Missouri

Assessment Program

Grade-Level Assessments

Technical Report 2009

Submitted to
Missouri Department of Elementary and Secondary Education
December 2009

Table of Contents

Table of Tables

Table of Figures

EXECUTIVE SUMMARY

This report is a technical summary of the 2009 operational administration of the Missouri Assessment Program (MAP). The MAP is a grade-level test in Communication Arts and Mathematics administered in Grades 3 through 8. The MAP is a grade-span test in Science administered in Grades 5 and 8. These tests are designed to measure students' knowledge of Communication Arts, Mathematics, and Science. This section provides a summary of the 2009 Technical Report.

E.1 Background

The MAP was originally designed as grade-span tests to measure Missouri's Show-Me Standards. These standards were adopted by the Missouri State Board of Education in 1996. Since their inception, Missouri's Show-Me Standards have been further refined to better delineate Content Standards, Process Standards, and Content Strands/Grade-Level Expectations as Missouri changed their testing program to comply with the requirements of No Child Left Behind. Starting in 2006, grade-level tests were administered in Communication Arts and Mathematics. In 2008, grade-span tests were administered in Science. In 2009, MAP was no longer administered at the high school level. It was replaced by the Missouri End-of-Course Assessments (the technical report for these assessments may be found here: http://dese.mo.gov/divimprove/assess/tech/). The MAP tests have therefore undergone multiple alignment analyses to ensure that MAP content reflects these refinements. Further details of the development of the 2009 MAP may be found in Chapter 3 of this report.

E.2 Administration

In the spring of 2009, Missouri administered grade-level MAPs in Communication Arts and Mathematics to students in Grades 3 through 8 and in Science to students in Grades 5 and 8. The MAP was administered from March 30 to April 24, 2009. A small portion of districts were granted a week-long extension to this testing window because the districts had been adversely affected by winter weather for an extended period of time. For these 29 districts, the test window was March 30 to May 1, 2009. Test administration is discussed in Chapter 4 of this report.

Approximately 550 districts administered Communication Arts and Mathematics MAP tests in Grades 3 through 8. These districts also administered Science MAP tests in Grades 5 and 8. Table E.1 shows participation rates based on the census data.[1] For the purposes of this report, participation rate is defined as the percentage of students who received a valid scale score given the total number of students who received a test book. The accountable column shows the total number of students who received a test book. The percent reportable column shows the percentage of students who received a scale

[1] The census data used in this report does not reflect additional cleaning steps that DESE staff implements once CTB releases data to DESE; therefore, the numbers in this report may differ from those in DESE reports using their cleaned data.

score on MAP. Further analysis of participation rates is provided in Chapter 7 of this report.

E.3 Student Performance

This is the fourth year of the grade-level MAP testing programs in Communication Arts and Mathematics and the second year for the grade-span tests in Science. Tables E.2 and E.3 present the percentage of students classified as *Proficient* or *Advanced* in 2006 through 2009 in Communication Arts and Mathematics, respectively. Table E.4 shows the percentage of students classified as *Proficient* or *Advanced* in 2008 and 2009 on the Science MAP.

For all grades and content areas, except Grade 6 Mathematics, small increases in the percentage of students classified as *Proficient* or *Advanced* were observed. In Grade 6 Mathematics, a small decrease in the percentage of students *Proficient* or *Advanced* was observed. More information on student performance may be found in Chapter 7 of this report.

E.4 Validity and Test Scores

Most sections of this technical report are designed to provide validity evidence to support the use of MAP test scores. Chapter 2 discusses the uses of MAP scores. Chapter 3 discusses the test development process used to create MAP, which is important to the content-related validity of the MAP scores. Chapter 4 presents information on test administration. Chapter 5 discusses the scoring of constructed-response items, as well as the results of the inter-rater reliability studies. Chapter 6 presents the scaling and linking procedures, as well as the results of other operational data analyses. Chapter 7 reviews the results of the 2009 operational administration and overviews the score reports sent to parents, schools, and districts. Chapter 8 highlights the standard-setting procedures used for MAP. Chapter 9 discusses reliability and construct-related validity. In this section, we evaluate the assumption that the content-area MAPs are unidimensional. For example, the grade-level Mathematics MAP should measure one primary dimension (Mathematics). Chapter 10 overviews the statistical and development processes used to assure fairness of the MAP for all examinees. Some analyses in this document are based on the calibration sample while others are based on census data. The sources of data used for particular analyses are indicated throughout the Technical Report.

Table E.1: Participation Rates: All Students

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	67357	99.71	67357	99.81		
4	66709	99.67	66709	99.82		
5	67307	99.67	67307	99.77	67307	99.72
6	65908	99.71	65908	99.77		
7	66531	99.68	66531	99.70		
8	67077	99.50	67077	99.54	67077	99.44

Table E.2: Percentage of Students Classified as *Proficient* or *Advanced* in 2006 through 2009 using Census Data: Communication Arts

	Communication Arts				
Grade	**2006**	**2007**	**2008**	**2009**	**2009-2008**
3	42.4	42.6	40.3	40.3	0.0
4	43.8	45.1	45.1	46.3	1.2
5	45.0	47.8	48.1	48.8	0.7
6	42.2	43.6	47.4	47.7	0.3
7	42.7	44.4	49.0	50.8	1.8
8	41.5	41.6	48.1	49.7	1.6

Table E.3: Percentage of Students Classified as *Proficient* or *Advanced* in 2006 through 2009 using Census Data: Mathematics

	Mathematics				
Grade	**2006**	**2007**	**2008**	**2009**	**2009-2008**
3	43.3	45.0	43.8	44.4	0.6
4	43.4	44.5	44.2	44.4	0.2
5	43.3	46.6	45.8	47.2	1.4
6	43.9	47.8	50.7	50.1	-0.6
7	42.9	44.9	49.5	51.9	2.4
8	39.8	40.6	43.8	46.4	2.6

Table E.4: Percentage of Students Classified as *Proficient* or *Advanced* in 2006 through 2009 using Census Data: Science

	Science		
Grade	**2008**	**2009**	**2009-2008**
5	44.5	45.1	0.6
8	43.2	44.8	1.6

CHAPTER 1: INTRODUCTION

The 2009 Missouri Assessment Program (MAP) marked the fourth administration of grade-level Communication Arts and Mathematics MAP in Missouri. It was the second administration of the grade-span Science MAP at Grades 5 and 8. The MAP is designed to measure students' knowledge of Communication Arts, Mathematics, and Science. This report provides a technical overview of the Communication Arts, Mathematics, and Science assessments of the 2009 MAP. As such, it presents evidence for the validity of the 2009 MAP scores.

This chapter of the Technical Report serves to describe the background, history, purpose, and design of the MAP, followed by an overview of the major sections for the current report.

1.1 Background of the Missouri Assessment Program

The MAP traces its origin to the 1993 Outstanding Schools Act. This act required that Missouri create a statewide assessment system that measured challenging academic standards. From this act, grade-span assessments were created that measured Missouri's Show-Me standards. Originally, MAP was designed to be a grade-span test: Grades 3, 7, and 11 in Communication Arts, Grades 4, 8, and 10 in Mathematics, and Grades 3, 7, and 10 in Science. Table 1.1 provides a brief timeline of the events of the grade-span MAP.

In 2001, the federal No Child Left Behind (NCLB) legislation was enacted, which required states to develop grade-level tests to be administered in Grades 3 through 8 and once in Grades 10 through 12 in both Reading and Mathematics. It also required that states have in place Science assessments to be administered at least once in Grades 3 through 5, Grades 6 through 9, and Grades 10 through 12 by the 2007–2008 school year. Based on the NCLB legislation, student performance, reported in terms of proficiency categories, is used to determine the adequate yearly progress of students at the school, district, and state levels.

In response to NCLB, the Department of Elementary and Secondary Education (DESE) contracted with CTB/McGraw-Hill in 2003 to expand the testing program to grade-level testing for Communication Arts and Mathematics. This contract was renewed in 2007 and extends through 2013. In the spring of 2005, Missouri administered a field test in Communication Arts and Mathematics, which was the basis for the construction of the 2006 and 2007 operational test forms.

The construction of the new Science MAP has been on a different trajectory. In 2005 DESE contracted with CTB/McGraw-Hill to construct a grade-span Science assessment in order to comply with the requirements of NCLB. In the spring of 2006, Missouri administered a field test in Science, which was the basis for the construction of the 2008 and 2009 operational Science forms. The contract to create grade-span Science assessments was renewed in 2007. This contract also extends through 2013.

In 2008, DESE together with Riverside Publishing developed End-of-Course Assessments for use at the high school level. With the development of the new test program, the MAP high school assessments were discontinued. The final administration of the MAP high school assessments was in the spring of 2008.

Table 1.2 shows a timeline of the development history of the NCLB-compliant testing program.

1.2 Purpose of the Missouri Assessment Program

The MAP is designed to measure how well students acquire the skills and knowledge described in Missouri's Grade-Level Expectations (GLEs). The assessments yield information on academic achievement at the student, class, school, district, and state levels. This information is used to diagnose individual student strengths and weaknesses in relation to the instruction of the GLEs and to gauge the overall quality of education throughout Missouri.

1.3 Design of the Missouri Assessment Program

The spring 2009 MAP administration consisted of 14 operational assessments. Within Grades 3 through 8, six versions of the operational form were administered in a grade/content area. These versions were spiraled within classrooms and differed only by a set of embedded field test items. Note that the field test items embedded in the MAP did not contribute to a student's scale score.

Each form contained a norm-referenced test form from which norm-referenced scores were derived. The norm-referenced items served as anchor items to link performance on the 2009 MAP administration to prior administrations. These counted toward the student scale score if they could be mapped to a Missouri GLE. If an item could not be mapped to a Missouri GLE, then it did not count toward the criterion-referenced score, nor was it used as an anchor item. Table 1.3 shows the number of items that could not be mapped to a Missouri GLE. Table 1.4 provides an overview of the 2009 MAP test design.

Braille and large print versions of each operational MAP form were constructed for each grade/content area to enable visually impaired students to participate in MAP testing. At some grade levels/content areas, it was necessary to drop items from the assessment due to difficulties associated with the Braille translation. Table 1.5 lists the number of items that were omitted from the Braille forms. Note that students taking the Braille forms were given full credit for the omitted items.

1.4 Overview of this Report

This Technical Report documents in the subsequent chapters the major activities of the testing cycle. This report provides comprehensive detail that confirms that the processes and procedures applied in the MAP adhered to appropriate professional standards and practices of educational assessment. Ultimately, this report serves to document evidence

that valid inferences about Missouri student performance can be derived from the MAP. An overview of major activities documented within this report is provided below:

Use of Test Scores (Chapter 2)
Chapter 2 of the Technical Report discusses the concept of validity evidence. This Technical Report is comprised of evidence that supports the use of the MAP scores. In Chapter 2, we discuss some of the uses of the MAP scores.

Item and Test Development (Chapter 3)
Chapter 3 of the Technical Report provides a summary of the major test development activities that occurred to create the spring 2009 operational test forms, the embedded field test items, and the materials developed to inform the public about the testing program. As each major event is presented and discussed, the role of the event in contributing to evidence for validity of the use of test results is discussed.

Test Administration (Chapter 4)
Chapter 4 of the Technical Report serves to describe the processes and activities implemented and information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students.

Scoring Constructed-Response Items (Chapter 5)
Chapter 5 of the Technical Report describes the processes and activities for scoring constructed-response items. This chapter discusses how raters are trained and the measures for assuring consistency among scorers. Finally, this chapter presents the results of the inter-rater reliability studies.

Operational Data Analyses (Chapter 6)
Chapter 6 of the Technical Report includes a detailed description of the operational analyses of the 2009 MAP, which are comprised of three major parts: the calibration sample, the classical item analysis and calibration, scaling, and linking using item response theory (IRT) models. This chapter describes the demographics of the calibration sample and compares it to the state census data. It reports the results of the classical item analysis, as well as the results of the calibration, scaling, and linking.

Test Results and Reporting (Chapter 7)
Chapter 7 of the Technical Report contains information on the results of the spring 2009 MAP administration. Detailed summary statistics based on scale scores and achievement level information are also provided. Finally, this chapter presents information on the score reports sent to parents, schools, and districts.

Standard-Setting (Chapter 8)
Chapter 8 of the Technical Report briefly discusses standard setting. It provides an overview of the standard setting activities that occurred for the MAP.

Reliability and Validity Evidence (Chapter 9)
Chapter 9 of the Technical Report provides evidence of reliability and validity of MAP scores. This chapter provides detailed results of the reliability of the tests, as well as information on the decision consistency of the cut scores. It also provides evidence of construct validity for MAP scores.

Fairness (Chapter 10)
Chapter 10 of the Technical Report discusses fairness and how the MAP tests are constructed to be fair to all Missouri students. This chapter summarizes the results of the differential item (DIF) analysis. It also discusses the results of an impact analysis to determine if large differences exist between demographic groups in Missouri.

Table 1. 1: Timeline of Grade-Span MAP

Year	Event
1996	Show-Me Standards Approved
1996	Frameworks for Curriculum Development published
1997	Annotations to the Curriculum Frameworks published
1998	First operational administration of Mathematics MAP (Grades 4, 8, and 10)
1999	First operational administration of Communication Arts MAP (Grades 3, 7, and 11) and Science MAP (Grades 4, 8, and 11)
2000	First operational administration of Social Studies MAP (Grades 4, 8, and 10)
2001	Mathematics Curriculum Supplement published
2005	Last year of grade-span MAP

Table 1. 2: Timeline of Grade-Level MAP

Year	Event
2004	Grade-Level Expectations published
2005	Communication Arts and Mathematics Field Test
2005	Standard Setting for Communication Arts and Mathematics
2006	First Operational Communication Arts and Mathematics MAP
2007	Science Field Test
2008	First Operational Science MAP
2008	Standard Setting for Science
2008	Last Operational Administration of High School MAP
2008	Version 2.0 Grade-Level Expectations (GLEs) published
2009	Last Operational Administration of MAP based on V1.0 GLEs

Table 1. 3: Number of Items that did not Map to a Missouri Grade-Level Expectation

Content	Grade	Number of Items
Science	5	3
	8	2

Table 1. 4: Spring 2009 MAP Test Design

Content	Grade	Number of Forms	Anchor Items	Operational Items	Total Number of OP Items	Total Raw Score Points	Embedded Field Test Items
Communication Arts	3	6	30	27	57	67	7-11
	4	6	35	20	55	63	8-12
	5	6	32	23	55	62	7-12
	6	6	31	24	55	62	8-10
	7	6	33	28	61	72	8-10
	8	6	28	33	61	68	7-10
Mathematics	3	6	30	30	60	67	23-24
	4	6	32	33	65	77	25-26
	5	6	32	30	62	69	23-24
	6	6	31	30	61	68	24
	7	6	32	30	62	69	23-24
	8	6	31	33	64	76	22-23
Science	5	6	22	31	53	79	15-16
	8	6	23	36	59	91	16-17

Table 1. 5: Spring 2009 Items Removed from Braille Forms

Content Area	Grade	Total Number of Items
Communication Arts	4	1
	7	1
Mathematics	3	2
	4	4
	6	2
	7	2
	8	6
Science	5	2
	8	3

CHAPTER 2: THE USES OF TEST SCORES

Validity is the overarching component of the MAP testing program. The following excerpt is from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

> 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 (17).

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

The following sections (Chapters 3 through 10) of this Technical Report provide additional evidence for these uses, as well as technical support for some of the interpretations and uses of test scores. The information in Chapters 3 through 10 also provides 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. To this end, DESE conducted a study on consequential validity that was implemented by the Assessment Resource Center (see *MAP and Missouri Schools: A Consequential Validity Study*, ARC, 2008).

2.1 Uses of Test Scores

The validity of a test score ultimately rests on 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 intended uses of MAP scores include:

- identifying students' strengths and weaknesses on Missouri's Grade-Level Expectations
- communicating expectations for all students
- evaluating school-, district-, and/or state-level programs
- informing stakeholders (teachers, school administrators, district administrators, DESE staff members, parents, and the public) on the status of the progress toward meeting academic achievement standards of the state
- meeting the requirements to measure Adequate Yearly Progress by NCLB
- meeting the requirements of the state's accountability program, Missouri School Improvement Program (MSIP)

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

2.2 Test-Level Scores

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

Custom-written portions of the MAP were directly authored by Missouri educators, edited by DESE and CTB staff, 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 GLEs to further solidify the Show-Me Standards as the foundation of the MAP. As shown in Table 1.3 in the previous chapter, all *TerraNova* items in the Communication Arts and the Mathematics MAP align to Missouri standards. Only three Grade 5 Science items and two Grade 8 Science items did not map to Missouri standards. Item development is described in Chapter 3; however, detailed descriptions of processes used to delineate the knowledge, skills, and abilities, including content limits and descriptions for each content area, are beyond the scope of this report.

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

2.2.1 Scale Scores

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

2.2.2 Levels of Achievement

A student's performance on the Communication Arts, Mathematics, or Science MAP is reported in one of four levels of achievement: *Below Basic*, *Basic*, *Proficient*, or *Advanced*. The cut scores for the levels of achievement were recommended by Missouri educators and citizens at the Bookmark Standard Setting Workshop in December 2005 for Communication Arts and Mathematics and in July 2008 for Science. The cut scores reflect the expectations of Missouri educators and citizens 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 Missouri Senate Bill 1080. (See Chapter 8 of this report for a discussion of MAP standard setting). Thus, MAP achievement levels

reflect the achievement standards and abilities intended by the Missouri legislature, Missouri teachers, Missouri citizens, and DESE. Descriptions of each level of achievement in terms of what a student should know and be able to do are provided with the *Guide to Interpreting Results* (see Chapters 4 and 7).

2.2.3 Use of Test-Level Scores

MAP scale scores and achievement levels provide summary evidence of student achievement in Communication Arts, Mathematics, or Science. Classroom teachers may use these scores as evidence of student achievement in these content areas. At the aggregate level, district and school administrators may use this information for activities such as planning curriculum. At the state level, the aggregate test-level scale scores are used for accountability programs associated with NCLB and the MSIP. The results presented in this Technical Report provide evidence that the scale scores are a valid and reliable indicator of student performance in Communication Arts, Mathematics, and Science.

2.3 Content Standard Subscores

The Content Standard subscores indicate student performance in terms of the number- and percent-correct score for each Content Standard in Communication Arts and each GLE strand in Mathematics and Science. Starting in 2008, Content Standard subscores were reported only through DESE's Crystal Reporting system. These scores may be aggregated by the state, district, or schools to determine the mean Content Standard subscores. These means may be used as indicators of the performance of the school or district in teaching students the knowledge and skills defined for each subject area.

2.3.1 Use of the Content Standard Subscores

The purpose of reporting Content Standard subscores on MAP is to show for each student the relationship between the overall achievement being measured and the skills in each of the areas delimited by the Content Standards in Communication Arts and the GLE strands in Mathematics and Science. 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. Chapter 3 of this Technical Report provides evidence of content validity that supports the use of the Content Standard subscores. Chapter 9 of this Technical Report provides evidence of construct validity that further supports the use of the Content Standard subscores.

District and school administrators may compare their aggregate results with the state mean to better understand their strengths and weaknesses within a content area. Caution should be exercised when comparing Content 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.

2.4 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 through DESE's Crystal Reporting system.

2.4.1 Use of the Process Standard Subscores

The purpose of reporting Process Standard subscores on MAP is to show the achievement of students in each of the areas delimited by the Process Standards in Communication Arts, Mathematics, or Science. When the Process Standard processes are used to teach the subject area content, the Process Standard subscores 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.

CHAPTER 3: TEST CONTENT DEVELOPMENT

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

This chapter is particularly relevant to AERA, APA, & NCME (1999) Standards 3.1, 3.2, and 3.7. It also addresses Standards 3.11, 7.4, and 7.7, which will be discussed in the pertinent section of this chapter. Standards 3.1, 3.2, and 3.7 are from Chapter 3 of the AERA, APA, & NCME (1999) Standards, which is titled *Test Development and Revision.* Each of these Standards will be presented, as will the way the Standard is addressed in this chapter. AERA, APA, & NCME (1999) Standard 3.1 says,

> Tests and testing programs should be developed on a sound scientific basis. Test developers and publishers should compile and document adequate evidence bearing on test development.

The purpose of this chapter is to document the test development process used for MAP. In this chapter, we describe steps taken to create MAP from the development of test specifications to the selection of operational forms.

3.1 Test Specifications

AERA, APA, & NCME (1999) Standard 3.2 says,

> The purpose(s) of the test, definition of the domain, and the test specifications should be stated clearly so that judgments can be made about the appropriateness of the defined domain for the stated purpose(s) of the test and about the relation of items to the dimensions of the domain they are intended to represent.

The purpose of the test is discussed in Chapter 2. MAP domains are generally defined as the knowledge and skills in a subject matter area that are identified within the Missouri GLE and Show-Me Standards documentation. These frameworks are, in turn, based on prior consensus among DESE, Missouri educators, and experienced subject-matter experts that the frameworks represent what is important for teachers to teach and students to learn.

Evidence of validity based on test content includes information about the test specifications, including the test design and test blueprint. Test development involves

creating a design framework from the statement of the construct to be measured. The MAP test specifications evolve from the tension between the constraints of the assessment program and the benefits sought from the examination of students. Many of the benefits sought are not scientific in nature, nor are many of the constraints; rather, they are policy considerations. The 2008–2009 MAP item selection specifications were finalized in August 2007 prior to item selection for the operational forms.

The MAP test specifications consist of a test blueprint and a test design for each grade level/content area. The key structural aspect of the MAP tests is the test blueprint, which specifies the target score points for each Content Standard (Table 3.1). The blueprint 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. Test design elements include such elements as number and types of items/tasks for each of the scores reported (tasks are measured by constructed-response items in MAP). The degree to which the 2009 MAP operational forms matched the test blueprint can be assessed by comparing the targeted score point distributions defined in the test blueprint with the actual point distributions displayed in Tables 3.4–3.7. Actual point distributions on the 2009 MAP operational forms matched blueprint targets within 10%, which was the tolerance for variation approved by DESE.

3.2 Item Development

Item development is discussed in this section in compliance with the AERA, APA, & NCME (1999) Standards. Standard 3.7 states,

> The procedures used to develop, review, and try out items, and to select items from the item pool should be documented. If the items were classified into different categories or subtests according to the test specifications, the procedures used for classification and the appropriateness and accuracy of the classification should be documented.

Planning and preparation for the development of item content for the 2008 and 2009 MAP Operational Test forms began in 2004. The plan specified an item development and selection cycle that included an initial item writing/passage selection workshop; a local pilot study; a content and bias review, item refinements and form construction; a subsequent round of formal field testing; the selection of operational forms based on statistical data from the field test; and ultimately, operational testing at grade levels 3 through 8. Each of these steps is described in greater detail below.

3.2.1 Reading Load

AERA, APA, & NCME (1999) Standard 7.7 is particularly relevant to item development. It says,

> In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct.

MAP item development takes place within well-established content development workflow processes and methodologies. These processes include editing items for both content and style, the latter of which includes multiple reviews of each question to assure proper grammar, punctuation, and compliance to the established style. Clarity and fair access for all examinees also fall within the purview of the style reviews, which occur at scheduled milestones within the overall test development process. A thorough quality assurance review is conducted by a separate entity within the publishing division prior to the actual publication and distribution of the MAP assessments.

During the initial item writing/content development workshops (described later), content developers are provided with specific training about how to write items that require minimal reading loads for assessing content knowledge outside of the reading/communication arts content domain. For example, Mathematics content developers are trained to recognize and eliminate excessive wordiness in question stems; likewise, Science developers are encouraged to use only strictly relevant information in their items, even for those items which require some kind of background explanation of a scenario or scientific experiment.

Once item writing workshops are complete, content development editors review all item content generated at the workshops and perform a post-workshop analysis. During this process, editors reject items which do not meet specific criteria for further development; items which do not directly assess the intended targets or cannot be modified in such a way as to comply with the established style and quality of the existing MAP items (due to excessive wordiness, linguistic complexity, or overall fair access concerns) are summarily filtered out from the pool. Then, only the remaining material is submitted to a thorough style review.

The established MAP content development workflow calls for style reviews to occur at other milestones which include (but are not limited to) pilot testing, formal content and bias reviews, and form selection. Style reviews also occur after the results of the Score, Revise, Rewrite workshops.

3.2.2 Item Writing

Communication Arts and Mathematics: In February 2005, a group comprised of Missouri educators, Regional Instructional Facilitators (RIFs), DESE staff, and CTB personnel participated in an Item Writing Workshop (IWW) for Communication Arts and

Mathematics at the Resort at Port Arrowhead, located at Lake Ozark, Missouri. The workshops were conducted with more than 30 teacher participants per content area. Teacher participants were selected by DESE to represent educational sites throughout Missouri. During the first day of the workshop, Communication Arts participants selected reading passages. During the next three days, Communication Arts participants used selected passages as a basis for writing constructed-response (CR) items and writing prompts for the 2009 Operational forms for grades 3–8. The Mathematics participants wrote CR items and performance-event (PE) items along with scoring guides to create a pool of items for the 2009 Operational forms for grades 3–8. The content developed at the workshop was based specifically upon the Missouri Show-Me Standards and GLEs. Some selected-response (SR) items were developed by CTB after the workshop to help supplement the item pool and reviewed by DESE. Items were refined after the initial item writing workshop which led to the production of local pilot test forms.

Science: In November 2004, a group comprised of Missouri educators, RIFs, DESE staff, and CTB personnel participated in a four-day Science IWW in Columbia, Missouri. The IWW was conducted with 37 teacher participants selected by DESE on the basis of their prior experience and expertise in item development for MAP Science and to represent educational sites throughout Missouri. The purpose of the IWW was to revise existing items and write new items to ensure a well-balanced item pool for the 2009 MAP Science operational tests. The existing items came from the MAP Science item pool previously developed for operational testing at grades 3 and 7. During the first two days of the IWW, the existing items were revised to target the new MAP Science GLEs. These new GLEs were the basis for the 2009 assessment to be administered at grades 5 and 8. During the third and fourth days of the IWW, Science participants wrote new CR items and performance events. A new MAP Science Performance event development template was introduced at the IWW. This template specified the types of tasks and numbers of items that comprise a Performance event.

Overall, the IWWs in November 2004 and February 2005 provided a basis upon which items written for the Communication Arts, Mathematics, and Science assessments could be selected for use on small-scale local pilot tests administered throughout Missouri.

3.2.3 Local Pilot Test

Small-scale pilot tests were administered in March 2005 (Science) and November 2005 (Communication Arts and Mathematics) in a limited number of classrooms throughout Missouri. Teachers who administered the pilot tests were generally selected by DESE from the pool of IWW participants.

Six Communication Arts forms per grade were piloted, consisting of approximately two SR items and six CR items each for grades 4, 5, 6, and 8. The six Communication Arts pilot forms for grades 3 and 7 each contained two selected-response items, four constructed-response items, and one writing prompt. Six Mathematics forms per grade were piloted, consisting of approximately twelve SR items and two CR items each for grades 3, 5, 6, and 7. The six Mathematics pilot forms for grades 4 and 8 each contained twelve SR items, four CR items, and one performance event. Ten Science forms per

grade, consisting of approximately 15 CR items, were piloted for each of grades 5 and 8. In addition to these ten pilot forms, eight performance events were piloted at each grade level.

3.2.4 Score, Revise, Rewrite Workshop

In April 2005 (Science) and February 2006 (Communication Arts and Mathematics), the results of the pilot studies underwent further evaluation during Score, Revise, and Rewrite (SRR) Workshops.

The purpose of the SRR Workshop was for the participants to score the items piloted in Missouri classrooms and to revise the items and rubrics/scoring guides based on the scoring process, student results, and subsequent discussion. DESE invited approximately 5 to 7 participants per grade/content area, resulting in the direct participation of approximately 100 Missouri educators in this step of the development process. 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. Regional Instructional Facilitators were also present and participated in the process. Overall, the goal of the workshop was to improve the item quality prior to the next step in the process, Content and Bias Review, and to ensure that quality items were developed for future use in the MAP. Most participants commented that this workshop was successful in this regard.

3.2.5 Content and Bias Review Workshop

Content and Bias Review (CBR) workshops were conducted in May 2005 (Science) and May 2006 (Communication Arts and Mathematics) with DESE, Missouri educators, RIFs, and CTB staff. Both of the CBR workshops were conducted in Columbia, Missouri. For the Content Review, DESE invited participants from educational sites throughout Missouri to review items, writing prompts, performance events and scoring guides for content accuracy and grade level appropriateness. In Communication Arts, participants also reviewed passages. In addition, participants in all three content areas verified each item's alignment to the Missouri curriculum by reviewing the Content Standard, Process Standard, and GLE assignment. The Content Review was accomplished over the course of one or two days, and was followed by a one- or two-day Bias Review.

The Bias Review committee was comprised of representatives from various backgrounds whose purpose was to screen the items for any racial, socioeconomic, gender, or other sensitivity issues. This follows AERA, APA, & NCME (1999) Standard 7.4, which states,

> Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.

The Bias Review committee could revise or reject items because of issues related to possible bias. Four Communication Arts items, no Mathematics items, and nine Science items were rejected from their respective pools. The remaining items were either accepted outright or accepted with revisions.

For each content area, over 30 Missouri educators participated in the process to help ensure content validity. Greater than 90% of reviewed items were accepted by the CBR committees. The general consensus was that the items as a group were well written and edited, and that the changes made during and after the SRR Workshop had contributed to a smooth CBR workshop. The accepted items became candidates for the next step in the process, the MAP field test.

3.3 Field Test Selection and Administration

The items approved by CBR committees became the basis for the formation of stand-alone Field Test forms administered in 2006 and 2007. The custom-written material was arranged into test forms using *TerraNova* Survey as a common anchor across forms. (The same anchor would become the norm-referenced test (NRT) portion of the 2008 operational test and is described in more detail in the following section). Field 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.

The MAP Spring 2006 Science Field Test consisted of four parallel forms per grade level, which were successfully administered in grades 5 and 8 in May 2006. The MAP Spring 2007 Communication Arts and Mathematics Field Tests consisted of six parallel forms per grade/content area which were successfully administered in grades 3–8 in May 2007. All field test forms were reviewed and approved by DESE prior to administration. The field tests generated item statistics that were used to help select two years of parallel operational forms, to be administered in 2008 and 2009.

3.4 Operational Test Selection

The use of an embedded *TerraNova* Survey provides both an anchor in the MAP tests and an NRT subtest, which is a requirement of the MAP. For most grade/content areas, the intact *TerraNova* Survey Form D was embedded in the 2006 and 2007 Field Tests and again in the 2009 operational tests. For grade 8 Communication Arts, one passage and item set was also selected from an alternate form of *TerraNova* Survey Form C due to an author's denial of permissions.

A small number of items from the Language Arts section of *TerraNova* Survey were identified by DESE as being aligned to Missouri's "Writing Standard English" content standard. To supplement the custom items and fulfill the blueprint, a selection of these *TerraNova* Language Arts items, plus the intact *TerraNova* Reading section of Survey Form D, were embedded in the 2007 Communication Arts Field Test and the 2009 Communication Arts operational test.

The use of the *TerraNova* Survey and its match to the Missouri standards plays an important role in planning for the entire development process leading up to the time of item selection. This is because the test blueprint is applied to the entire test, which includes both the NRT and custom portions. As an NRT product, *TerraNova* items are pre-classified to an existing set of *TerraNova* Reading, Language, Mathematics, or Science standards.[2] In many cases, the match of *TerraNova* items to Missouri standards could be considered equivalent; nevertheless, the item development process provided for a DESE review of how the items in the *TerraNova* Survey were matched to the Missouri standards. The match of *TerraNova* items to Missouri standards was initially assessed by DESE in 2004 and then verified by DESE in October 2007 and August 2008.

Operational item selections for 2009 were performed in September–October 2007 by CTB. The selection process followed strict statistical criteria specified by CTB's Research department and approved by DESE. The selection criteria were based on both content requirements and statistical criteria, including the following:

1. *TerraNova* Survey Form D is the anchor for all grades and content areas, with exceptions, as noted above.
2. Test length and item types match the DESE-approved test design.
3. Content coverage matches DESE-approved test blueprint.
4. The following items were to be avoided, whenever possible:
    a. For CR items: 3+ point items where more than 50% were able to attain the top score points.
    b. *p*-value ≤ 0.20 or ≥ 0.90
    c. Omit rates $\geq 5\%$
    d. Poor Fit statistics (Q1)
    e. Significant DIF statistics:
        i. If an item with DIF had to be included for blueprint coverage, examine the item to determine if any content reason exists for the DIF flag (sometimes items will demonstrate statistical bias but no content reason can be determined for the bias).
        ii. Obtain DESE permission to use the DIF item (meaning someone from DESE should examine the item and agree that no content reason can be determined for the statistical bias).
5. Statistical properties of the test:
    a. ITEMWIN software must be used to select forms.
    b. Standard Error of Measurement (SEM) and Test Characteristic Curve (TCC) of 2009 operational test must match within 5% of 2008 MAP

Production of the 2009 operational test forms and ancillary materials commenced in June 2008. 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

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

final approvals were in place in December 2008. Then, test books and ancillary materials were printed and distributed in support of the Spring 2009 Operational Test.

3.5 Universal Design

Assessments that are universally designed allow participation of the widest possible range of students, resulting in more valid inferences about students' performance. Universally designed assessments may reduce the need for accommodations by reducing or eliminating access barriers associated with the tests themselves. Table 3.2 presents the elements of universal design (Thompson & Thurlow, 2002). The elements of Universal Design are relevant to both item development and form construction. This section addresses how the elements of Universal Design were addressed in the construction of the Spring 2009 test forms.

Universal design requires that assessments need to measure the performance of students with a wide range of abilities and skill repertoires, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content. To accommodate the greatest number of students within MAP, the regular print assessment includes simple, clear, and intuitive instructions and procedures, maximum readability and comprehensibility, and maximum legibility. All of these design components are addressed primarily through the physical layout and formatting of the test books. The page specifications and template for test book pages define how directions and test items are placed on the pages, the location and appearance of headers and footers, spacing between an item stem and answer choices, and other page elements to ensure a consistent, legible appearance of printed test books. Written instructions in the test books at the beginning of each test session are clearly and simply stated, and the wording of such instructions is standardized as much as possible across content areas and grade levels to ensure clarity and consistency.

The MAP test books are designed to minimize distractions and to support navigation through the test book. In Grade 3 Communication Arts, the test items are read aloud to the students. In all grade levels and content areas, a "full-page stop" at the end of each testing session indicates that the students cannot turn the page until instructed by the test examiner. Right-facing pages within a session have a "go on" arrow at the bottom right-hand corner to indicate that the test session continues on the next page. Any pages that are intentionally left blank are labeled "Do Not Mark on this Page" to indicate that there are no test materials on that page.

3.6 Accommodations

Students with disabilities or who are English Language Learners may be provided test administration accommodation based on their Individualized Education Plan (IEP). More information on accommodations can be found in Section 4.4.2 of Chapter 4. Accommodation code definitions can be found on the DESE website at:

http://www.dese.mo.gov/divimprove/assess/special.html.

Braille and large print versions were constructed for each grade/content area to enable visually-impaired students to participate in MAP testing. DESE conducted two meetings with a committee of teachers of to ensure the 2009 MAP assessment would be accessible to visually-challenged students. During the first meeting, in September 2007, the entire pool of items available for the 2009 operational test were reviewed to determine which could not be Brailled and to make recommendations for how to transcribe those that were appropriate for the Braille version. Specific recommendations were provided to the transcribers and an Independent Braille expert, who collaborated to produce the Braille proof and the teacher's notes that accompany the Braille form. During the second review meeting in January 2009, DESE and a teacher committee reviewed the 2009 Braille version of Form A of each grade level and made recommendations, as needed, for how to modify the transcription to best serve the needs of visually-challenged students.

While the goal is to maximize the number of items on the Braille form, it was not possible to transcribe all items into Braille, as some items represent concepts that are simply not appropriate for students who take the Braille form. At some grade levels/content areas, it was necessary to omit items from the Braille version due to bias issues or excessive difficulty associated with the Braille transcription. Table 3.3 lists the items that were omitted from the 2009 Braille versions. The concerns noted by the committee for items that were dropped from the Braille form will be brought to the attention of assessment editors and item writers to guide future item development.

3.7 Content and Process Standards

Test content evidence of validity is provided for the MAP with the specification of each of the Content and Process Standards that are influential in acquiring the skills tested in the items/tasks used in each of the MAP tests. If teachers teach using the Content and Process Standards as intended, then student performance should improve on those items that were identified as implicitly tapping these habits of mind and/or explicitly written and clearly intended to measure specific Content Standards.

AERA, APA, & NCME (1999) Standard 3.11 says,

> Test developers should document the extent to which the content domain of a test represents the defined domain and test specifications.

Table 3.4 provides the distribution of items and points by Content Standard for Communication Arts. Tables 3.5 and 3.6 provide the same distribution by GLE strand for Mathematics and Science, respectively. (GLE strands are the reported categories for these content domains; however, GLEs remain linked directly to the Content Standards.) Lastly, tables 3.7 through 3.9 show the distribution of items and points by Process Strand for Communication Arts, Mathematics, and Science, respectively.

3.8 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the MAP assessments. The efforts by DESE and CTB/McGraw-Hill in

developing the MAP address multiple best practices of the test industry but in particular are related to the following AERA, APA, & NCME (1999) Standards:

- Standard 3.1 — Tests and testing programs should be developed on a sound scientific basis. Test developers and publishers should compile and document adequate evidence bearing on test development.
- Standard 3.2 — The purpose(s) of the test, definition of the domain, and the test specifications should be stated clearly so that judgments can be made about the appropriateness of the defined domain for the stated purpose(s) of the test and about the relation of items to the dimensions of the domain they are intended to represent.
- Standard 3.7 — The procedures used to develop, review, and try out items, and to select items from the item pool should be documented. If the items were classified into different categories or subtests according to the test specifications, the procedures used for classification and the appropriateness and accuracy of the classification should be documented.
- Standard 3.11 — Test developers should document the extent to which the content domain of a test represents the defined domain and test specifications.
- Standard 7.4 — Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.
- Standard 7.7 — In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct.

Table 3. 1: MAP Test Blueprint: Target Score Points by Content Standard (Communication Arts) or GLE Strand (Mathematics and Science)

Content Area	Grade							
Content Standard/ GLE Strand	**3**	**4**	**5**	**6**	**7**	**8**	**10**	**11**
Reading								
Speaking/Writing Standard English	15	10	12	13	16	15		15
Reading—Fiction & Nonfiction	48	54	52	51	50	53		52
Writing Formally & Informally	6	2	2	1	7	1		6
Mathematics								
Number and Operations	25	19	14	14	14	12	11	
Algebraic Relationships	14	16	14	14	14	23	23	
Geometric and Spatial Relationships	14	16	14	14	14	15	15	
Measurement	10	16	14	14	14	12	11	
Data and Probability	7	11	14	14	14	15	15	
Science								
Matter and Energy			11			11		12
Force and Motion			8			7		10
Living Organisms			8			10		11
Ecology			9			8		8
Earth Systems			10			11		8
Universe			9			9		8
Scientific Inquiry			21			24		27
Science, Technology, and Human Activity			7			6		6

Table 3. 2: Elements of Universal Design

Element	Explanation
Inclusive Assessment Population	Tests designed for state, district, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field testing procedures.
Precisely Defined Constructs	The specific constructs tested must be clearly defined so that all construct irrelevant cognitive, sensory, emotional, and physical barriers can be removed.
Accessible, Non-Biased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amenable to Accommodations	The test design facilitates the use of needed accommodations (e.g., all items can be Brailled).
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and presented in understandable language.
Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words are kept to a minimum) to produce readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, to tables, figures, and illustrations, and to response formats.

Table 3. 3: Items Omitted from the MAP Spring 2009 Braille Version

Grade	Content Area	Type	Session	Item
3	Mathematics	SR	2	26
		CR	3	1
4	Communication Arts	SR	2	4
	Mathematics	SR	2	24
		CR	3	1
		CR	3	3
		CR	3	5
5	Science	SR	2	9
		CR	3	2
6	Mathematics	SR	1	4
		SR	1	6
7	Communication Arts	SR	3	16
	Mathematics	SR	1	9
		CR	3	7
8	Mathematics	PE	1	31
		SR	2	17
		CR	3	2
		CR	3	5
		CR	3	6
		CR	3	9
	Science	SR	2	4
		CR	3	5
		CR	3	6

Table 3. 4: MAP 2009 Content Standard Item/Point Distributions, Communication Arts

Grade	Content Standard	TN NRT Items	SR Items	CR/PE Items	Total Items	SR Points	CR/PE Points	Total Points	% of Total Points
3	Speaking/Writing Standard English	1	15		16	16		16	24%
	Reading Fiction/Poetry/Drama	20	2	4	26	22	8	30	45%
	Reading Nonfiction	9		3	12	9	6	15	22%
	Writing Formally & Informally	0		3	3	0	6	6	9%
	Combined Reading from Standards 2 & 3	29	2	7	38	31	14	45	67%
	Total	30	17	10	57	47	20	67	100%
4	Speaking/Writing Standard English	0	10		10	10		10	16%
	Reading Fiction/Poetry/Drama	35		3	38	35	6	41	65%
	Reading Nonfiction	0	2	4	6	2	8	10	16%
	Writing Formally & Informally	0		1	1	0	2	2	3%
	Combined Reading from Standards 2 & 3	35	2	7	44	37	14	51	81%
	Total	35	12	8	55	47	16	63	100%
5	Speaking/Writing Standard English	0	13		13	13		13	21%
	Reading Fiction/Poetry/Drama	19	2	7	28	21	14	35	56%
	Reading Nonfiction	13			13	13		13	21%
	Writing Formally & Informally	0		1	1	0	1	1	2%
	Combined Reading from Standards 2 & 3	32	2	7	41	34	14	48	77%
	Total	32	15	8	55	47	15	62	100%
6	Speaking/Writing Standard English	0	13		13	13		13	21%
	Reading Fiction/Poetry/Drama	17			17	17		17	27%
	Reading Nonfiction	14	3	7	24	17	14	31	50%
	Writing Formally & Informally	0		1	1	0	1	1	2%
	Combined Reading from Standards 2 & 3	31	3	7	41	34	14	48	77%
	Total	31	16	8	55	47	15	62	100%
7	Speaking/Writing Standard English	0	16		16	16		16	22%
	Reading Fiction/Poetry/Drama	27		3	30	27	6	33	46%
	Reading Nonfiction	7	2	4	13	9	8	17	24%
	Writing Formally & Informally	0		3	3	0	7	7	10%
	Combined Reading from Standards 2 & 3	33	2	7	42	35	14	49	68%
	Total	33	18	10	61	51	21	72	100%
8	Speaking/Writing Standard English	0	15		15	15		15	22%
	Reading Fiction/Poetry/Drama	10	2	7	19	12	14	26	38%
	Reading Nonfiction	24			24	24		24	35%
	Writing Formally & Informally	0		3	3	0	3	3	4%
	Combined Reading from Standards 2 & 3	34	2	7	43	36	14	50	74%
	Total	34	17	10	61	51	17	68	100%

Table 3. 5: MAP 2009 GLE Strand Item/Point Distributions, Mathematics

Grade	GLE Strand	TN NRT Items	SR Items	CR/PE Items	Total Items	SR Points	CR/PE Points	Total Points	% of Total Points
3	Number and Operations	14	5	3	22	19	6	25	37%
	Algebraic Relationships	3	5	2	10	8	4	12	18%
	Geometric and Spatial Relationships	4	9		13	13	0	13	19%
	Measurement	4	4		8	8	0	8	12%
	Data and Probability	5		2	7	5	4	9	13%
	Total	30	23	7	60	53	14	67	100%
4	Number and Operations	17		1	18	17	4	21	27%
	Algebraic Relationships	4	6	3	13	10	6	16	21%
	Geometric and Spatial Relationships	4	6	2	12	10	4	14	18%
	Measurement	4	5	3	12	9	6	15	19%
	Data and Probability	3	6	1	10	9	2	11	14%
	Total	32	23	10	65	55	22	77	100%
5	Number and Operations	16			16	16	0	16	23%
	Algebraic Relationships	3	6	2	11	9	4	13	19%
	Geometric and Spatial Relationships	3	8	1	12	11	2	13	19%
	Measurement	6	4	2	12	10	4	14	20%
	Data and Probability	4	5	2	11	9	4	13	19%
	Total	32	23	7	62	55	14	69	100%
6	Number and Operations	17			17	17	0	17	25%
	Algebraic Relationships	4	4	2	10	8	4	12	18%
	Geometric and Spatial Relationships	2	8	1	11	10	2	12	18%
	Measurement	4	6	2	12	10	4	14	21%
	Data and Probability	4	5	2	11	9	4	13	19%
	Total	31	23	7	61	54	14	68	100%
7	Number and Operations	16			16	16	0	16	23%
	Algebraic Relationships	2	7	2	11	9	4	13	19%
	Geometric and Spatial Relationships	4	6	2	12	10	4	14	20%
	Measurement	4	5	2	11	9	4	13	19%
	Data and Probability	6	5	1	12	11	2	13	19%
	Total	32	23	7	62	55	14	69	100%
8	Number and Operations	14			14	14	0	14	18%
	Algebraic Relationships	5	10	5	20	15	10	25	33%
	Geometric and Spatial Relationships	4	3	3	10	7	6	13	17%
	Measurement	3	6		9	9	0	9	12%
	Data and Probability	5	4	2	11	9	6	15	20%
	Total	31	23	10	64	54	22	76	100%

Table 3. 6: MAP 2009 GLE Strand Item/Point Distributions, Science

Grade	GLE Strand	TN NRT Items	CR/PE Items	Total Items	SR Points	CR/PE Points	Total Points	% of Total Points
5	Matter and Energy	2	4	6	2	8	10	13%
	Force and Motion	1	3	4	1	6	7	9%
	Living Organisms	4	2	6	4	4	8	10%
	Ecology	3	3	6	3	6	9	11%
	Earth Systems	2	4	6	2	8	10	13%
	Universe	2	3	5	2	6	8	10%
	Scientific Inquiry	5	10	15	5	15	20	25%
	Science, Technology, & Human Activity	3	2	5	3	4	7	9%
	Total	22	31	53	22	57	79	100%
8	Matter and Energy	2	5	7	2	10	12	13%
	Force and Motion	3	2	5	3	4	7	8%
	Living Organisms	3	4	7	3	8	11	12%
	Ecology	2	3	5	2	6	8	9%
	Earth Systems	5	3	8	5	6	11	12%
	Universe		4	4	0	8	8	9%
	Scientific Inquiry	7	12	19	7	20	27	30%
	Science, Technology, & Human Activity	1	3	4	1	6	7	8%
	Total	23	36	59	23	68	91	100%

Table 3. 7: MAP 2009 Number of Items/Points Measuring Process Standards, Communication Arts

Grade Level	Process Standard	NRT Items	Custom Items	Total Items	SR Points	CR Pts	Total Points
3	1.5	13	0	13	13	0	13
	1.6	8	5	13	9	8	17
	2.1		3	3	0	6	6
	2.2		15	15	15	0	15
	3.1		1	1	0	2	2
	3.5	9	3	12	10	4	14
4	1.1		1	1	0	2	2
	1.5	3	0	3	3	0	3
	1.6	20	4	24	21	6	27
	2.1		1	1	0	2	2
	2.2		10	10	10	0	10
	3.5	12	4	16	13	6	19
5	1.5	1	0	1	1	0	1
	1.6	18	2	20	19	2	21
	2.1		5	5	4	1	5
	2.2		9	9	9	0	9
	3.5	13	7	20	14	12	26
6	1.4		1	1	1	0	1
	1.6	24	3	27	25	4	29
	2.1		1	1	0	1	1
	2.2		13	13	13	0	13
	2.4	1	0	1	1	0	1
	3.5	6	6	12	7	10	17
7	1.5	5	0	5	5	0	5
	1.6	12	4	16	13	6	19
	1.8		1	1	0	2	2
	2.1		2	2	0	5	5
	2.2		16	16	16	0	16
	2.4	3	0	3	3	0	3
	3.5	13	5	18	14	8	22
8	1.5	6	2	8	8	0	8
	1.6	20	7	27	24	5	29
	2.1		3	3	0	3	3
	2.2		15	15	15	0	15
	2.4	2	0	2	2	0	2
	3.5		6	6	1	10	11

Table 3. 8: MAP 2009 Number of Items/Points Measuring Process Standards, Mathematics

Grade Level	Process Standard	NRT Items	Custom Items	Total Items	SR Points	CR Pts	Total Points
3	1.10		6	6	5	2	7
	1.5	6	0	6	6	0	6
	1.6	8	11	19	17	4	21
	2.1		1	1	1	0	1
	3.1	1	0	1	1	0	1
	3.2		1	1		2	2
	3.3	13	7	20	18	4	22
	3.5	2	0	2	2	0	2
	3.6		3	3	3	0	3
	4.1		1	1		2	2
4	1.10	2	4	6	6	0	6
	1.5		1	1	1	0	1
	1.6	5	6	11	10	2	12
	1.8		1	1		2	2
	3.1	11	3	14	14	0	14
	3.2		4	4	1	6	7
	3.3	14	4	18	15	6	21
	3.5		5	5	4	2	6
	3.6		5	5	4	4	8
5	1.1		1	1	1	0	1
	1.10		4	4	1	6	7
	1.5	4	0	4	4	0	4
	1.6		10	10	9	2	11
	3.1	2	6	8	8	0	8
	3.2		1	1		2	2
	3.3	24	4	28	27	2	29
	3.5	1	0	1	1	0	1
	3.6		4	4	3	2	5
	3.7	1	0	1	1	0	1
6	1.10		4	4	3	2	5
	1.5	4	0	4	4	0	4
	1.6	3	4	7	7	0	7
	1.7	3	0	3	3	0	3
	1.8		2	2	2	0	2
	3.1	6	10	16	15	2	17
	3.2		4	4	2	4	6
	3.3	12	2	14	13	2	15
	3.5	2	0	2	2	0	2
	3.6		5	5	3	4	7
	3.7	1	0	1	1	0	1
7	1.10		3	3	2	2	4
	1.4		1	1		2	2
	1.5	3	0	3	3	0	3
	1.6	2	7	9	8	2	10
	1.7		1	1	1	0	1
	3.1	8	7	15	15	0	15

Table 3. 8: MAP 2009 Number of Items/Points Measuring Process Standards, Mathematics (Cont'd)

Grade Level	Process Standard	NRT Items	Custom Items	Total Items	SR Points	CR Pts	Total Points
7	3.2		4	4	3	2	5
	3.3	18	3	21	21	0	21
	3.5		1	1	1	0	1
	3.6		3	3	1	4	5
	3.7	1	0	1	1	0	1
	4.1		1	1		2	2
8	1.4		2	2	2	0	2
	1.5	6	0	6	6	0	6
	1.6	4	11	15	12	6	18
	1.8	1	2	3	1	6	7
	3.1	2	3	5	5	0	5
	3.2		2	2	1	2	3
	3.3	18	2	20	18	4	22
	3.4		3	3	3	0	3
	3.5		1	1	1	0	1
	3.6		5	5	4	2	6
	3.8		2	2	1	2	3

Table 3. 9: MAP 2009 Number of Items/Points Measuring Process Standards, Science

Grade Level	Process Standard	NRT Items (SR)	Custom/CR/, Other Items	Total Items	NRT Points (SR)	Custom/CR/, Other Pts	Total Points
5	1.1		2	2		2	2
	1.10	10	9	19	10	18	28
	1.3	2	5	7	2	9	11
	1.5	5		5	5	0	5
	1.6	3	9	12	3	16	19
	1.7		2	2		2	2
	1.8		1	1		4	4
	3.5	2	3	5	2	6	8
8	1.1		2	2		2	2
	1.10	16	16	32	16	31	47
	1.3	1	4	5	1	8	9
	1.5	3	1	4	3	1	4
	1.6	3	8	11	3	16	19
	1.7		2	2		2	2
	1.8		2	2		6	6
	3.6		1	1		2	2

CHAPTER 4: TEST ADMINISTRATION

Chapter 4 of the Technical Report describes the processes and activities implemented and information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. According to the AERA, APA, & NCME *Standards* (1999), the "usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions" (61). Chapter 4 examines how test administration procedures implemented for the MAP strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

Chapter 4 demonstrates adherence to AERA, APA, & NCME (1999) Standards 3.19, 3.20, 5.1, 5.2, 5.3, 5.4, 5.6, and 5.7,in the MAP program. Each Standard will be explicated within the relevant section of this chapter.

4.1 Training of Districts

To ensure that the MAP tests are administered and scored in accordance with the department's mandates, DESE takes a primary role in communicating with and training district personnel. The development of the MAP tests is a collaborative effort between DESE and CTB/McGraw-Hill. DESE conveys to districts the purpose of the MAP program and that test administration must be consistent with test industry standards, as well as meet the State Board of Education policies and the mandates of both state and federal legislation. To accomplish these goals, DESE provides train-the-trainer opportunities for the RIFs. The RIFs convey test administration training to districts. The RIFs also conduct Quality Assurance visits during testing to ensure district adherence to the standardized administration of the tests.

The RIFs are responsible to districts within their region. The RIFs disseminate information to each district, offer assistance with test administration, and serve as the liaisons between DESE and the districts. DESE departmental staff also communicates directly with districts, answering questions particular to the MAP program as well as general assessment questions. DESE staff also provides assistance with MAP data and interpretation of MAP test results.

The Director of Assessment and the Assistant Director of Assessment trained the RIFs in the following components of MAP test administration: the *Test Coordinator's Manual*; the *Examiner's Manual*; the dates for testing; appropriate protocols for test administration and security; guidance on the timing and administration of tests; and changes made to the test since 2008. Appendix A of this report contains DESE's presentations on the *Test Coordinator's Manual* and the *Examiner's Manual*. The RIFs, in turn, used this information to train district-level staff. Appendix A also contains one of the presentations that was compiled by RIFs in the St. Louis region. It is representative of the information that other RIFs would use in their presentations.

4.2 Ancillary Materials

Test administration ancillary materials for the MAP contribute to the body of evidence of the validity of score interpretation. This section examines how the test materials address the AERA, APA, & NCME (1999) Standards related to test administration procedures.

For the spring 2009 test administration, CTB/McGraw-Hill produced two types of administration manuals: the *Test Coordinator's Manual* and the *Examiner's Manual*. DESE Curriculum and Assessment staff review, provide feedback, and give final approval for each manual.

The *Test Coordinator's Manual* is common to all grades and content areas. It provides an overview of MAP and any changes made to MAP for 2009. It gives guidelines for testing, such as the inclusion of special populations, the use of translators, and the invalidation procedures. It also details the Test Coordinator's role in the testing process by outlining nine steps the Test Coordinator should follow. These steps are:

Step 1: Review Testing Materials
Step 2: Distribute Testing Materials
Step 3: Collect Testing Materials
Step 4: Check the Organization of Materials Collected
Step 5: Check the Student Information Sheet (SIS)
Step 6: Check the Group Information Sheet (GIS)
Step 7: Complete the School/Group List
Step 8: Organize Materials for the District Test Coordinator
Step 9: Package and Ship Testing Materials

The *Examiner's Manual*s are specific to each grade, and for Grade 3, it was specific for each form. The MAP *Examiner's Manual*s also outline steps that should be followed when administering MAP. These steps include:

Step 1: Preparing for Testing
Step 2: Organize Your Classroom
Step 3: Check your Testing Materials
Step 4: Before Testing
Step 5: Administer the Test
Step 6: Invalidations and Make-ups
Step 7: After Testing: Student Status Coding
Step 8: Assemble Materials for Return

These steps provide instructions on pre-test and post-test procedures, such as:

- Test security
- Standardized testing protocols for norm-referenced information
- Using student barcode labels

- Completing the student information sheet, including recording test accommodations

This section presents the AERA, APA, & NCME (1999) Standards relevant to test administration and how information in the MAP *Examiner's Manual*s and *Test Coordinator Manual* address these Standards.

Standard 3.19 *The directions for test administration should be presented with sufficient clarity and emphasis so that it is possible for others to replicate adequately the administration conditions under which the data on reliability and validity, and, where appropriate, norms were obtained.*

The MAP *Examiner's Manual*s provide instructions for before-, during-, and after-testing activities with sufficient detail and clarity to support reliable test administrations by qualified test administrators. To ensure uniform administration conditions throughout the state, instructions in the *Examiner's Manual*s describe the following: the materials that the examiner and students need for testing; how to verify that pre-coded student information on student barcode labels is correct; how to fill out the Student Information Sheet if the student barcode label is incorrect; how to prepare the testing environment; the test schedule, including testing times; and how to administer the tests.

Standard 3.20 *The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample material, practice or sample questions, criteria for scoring, and a representative item identified with each major area in the test's classification or domain should be provided to the test takers prior to the administration of the test or included in the testing material as part of the standard administration instructions.*

To ensure clarity of instructions to students, the manuals include scripts that the examiner is instructed to read verbatim to students. Examiners are instructed to follow the script and to repeat any part of the directions as many times as needed, but to not modify the words used. Examiners may use professional judgment to respond to student questions, but they may not reword test items, suggest answers, or evaluate student work during the testing session. A sample of a script is presented in Figure 4.1.

Sample test items are provided in each content area to familiarize students with how to fill in answers. Sample items are also provided in the *Examiner's Manual*s.

Standard 5.1 *Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer, unless the situation or a test taker's disability dictates that an exception should be made..*

To ensure the usefulness and interpretability of test scores and to minimize sources of construct-irrelevant variance, it is essential that the MAP is administered according to the prescribed test schedule. The *Test Coordinator's Manual* includes instructions for scheduling the test within the state testing window of March 30 through April 24, 2009, with a one-week extension until May 1, 2009 for 29 districts. The *Examiner's Manual*s

contain the schedule for timing each test session and whether timing is to be strictly enforced. The test timing schedule is presented in Table 4.1.

Standard 5.2 *Modifications or disruptions of standardized test administration procedures or scoring should be documented.*

DESE staff administer reports on testing concerns which have a wide range of improper activities that may occur during testing including the following: copying and reviewing MAP test questions with students; cueing students during testing either verbally or with written materials on the classroom walls; cueing students nonverbally, such as tapping or nodding the head; using a calculator on parts of the test where it is not allowed; allowing too much time on *TerraNova* sections of the test; allowing students to correct or complete answers after tests have been returned to the teacher; splitting sessions into two parts; ignoring the standardized directions in the test books; reading the Communication Arts test to students; paraphrasing parts of the test to students; changing or completing (or allowing other school personnel to change or complete) student answers; allowing accommodations that are not written in the IEP; allowing non-IEP students accommodations; allowing students to use dictionaries on parts of MAP other than the writing prompt; or defining terms on the test.

Testing concerns are gathered from school officials, students, parents, and other interested parties who call DESE to state their allegation. A narrative of the conversation is written and read back to them. The superintendent of the district in which the allegation is made is then contacted and read the narrative. A letter is sent to confirm the conversation and to ask the superintendent to investigate the claim. A MAP Quality Assurance District Response Report is sent for the superintendent to use for replying to the allegation. This report is shown in Figure 4.2.

All of these narratives, letters, and reports are given to the Data, Accountability, and Accreditation section in order to make accountability decisions.

Standard 5.4 *The testing environment should furnish reasonable comfort with minimal distractions.*

Step 2 in the Examiner's Manual overviews the steps that teachers should take to prepare their classroom for administering the MAP test. These include:

- Plan for the distribution and collection of materials.
- Plan seating arrangements. Allow enough space between students to prevent the sharing of answers.
- Eliminate distractions such as bells or telephones.
- Use a Do Not Disturb sign on the door of the testing room.
- Make sure classroom maps, charts, and any other materials that relate to the content and processes of the test are out of the students' view.
- When administering the timed portion of the test, write on the board the starting and stopping times for the test.

Standard 5.6 *Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.*

The *Examiner's Manual* and *Test Coordinator's Manual* present instructions for post-test activities to ensure that test materials are handled properly and to ensure the integrity of student information and test scores. Detailed instructions guide test examiners in completing required information on students' scannable test books. For students who were administered a large print or Braille version of the MAP, examiners are instructed to transcribe students' responses from the large print test or Braille test book to a regular-edition test book exactly as they responded in the large print or Braille test book.

Standard 5.7 *Test users have the responsibility of protecting the security of test materials at all times.*

Throughout the manuals, test coordinators and examiners are reminded of test security requirements and procedures to maintain test security. Specific actions that are direct violations of test security are so noted. Detailed information about test security procedures are presented in Section 4.3.

4.2.1 Return Material Forms and Guidelines

The *Test Coordinator's Manual* instructs test coordinators in procedures for organizing and packing materials and returning them to CTB/McGraw-Hill for scanning and scoring. DESE curriculum and assessment staff have opportunities to review, provide feedback, and have final approval. The purpose of the instructions is to ensure that used and unused test materials are properly accounted for and student answer documents are organized properly for return shipment. Proper organization of materials and accurate completion of the school/group list document contributes to accurate score reports and helps in delivery of such reports in a timely manner.

4.2.2 Security Forms

As soon as test books are received by a district, the district test coordinator assures that the first and last security barcode on the tests match the packing list they received. The district test coordinator then packages the tests to be sent to schools. Upon returning tests to CTB/McGraw-Hill, school and district test coordinators are required to complete and submit a *Test Book Accountability Form* that details the number of scorable and nonscorable books returned. This form also requires that districts/schools document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books. The *Test Book Accountability Form* is shown in Figure 4.3.

4.2.3 Interpretive Guides

Essential to making valid interpretations of test scores is an understanding of what the test scores mean and how to interpret score reports. The *Guide to Interpreting Results* is written for Missouri teachers and administrators who receive MAP score reports from the 2009 administration. More detail about the guide can be found in Chapter 7.

4.3 Test Security Measures

Maintaining the security of all test materials is crucial to preventing the possibility of random or systematic errors, such as unauthorized exposure of test items, that would affect the valid interpretation of test scores. Several test security measures are implemented for the MAP. Test security procedures are discussed throughout the *Test Examiner Manuals* and *Test Coordinator's Manual*.

Test coordinators and examiners are instructed to keep all test materials in locked storage, except during actual test administration, and access to secure materials must be restricted to authorized individuals only (e.g., test examiners and the school test coordinator). During the testing sessions, test examiners are directly responsible for the security of the MAP and must account for all test materials at all times. The test examiners must supervise the test administrations at all times.

4.4 Test Administration

The 2009 test was administered to students within the state testing window of March 30 to April 24, 2009, with a one-week extension until May 1, 2009 for 29 districts adversely affected by winter weather. Table 4.2 shows those districts who were given a one-week extension of the testing window. Systems chose when and how to administer the MAP within this window. Each session within each content area of the MAP was required to be administered in one block of time.

4.4.1 Time

Each section of each content area test was timed to provide sufficient time for students to attempt all items. The *Examiner's Manual*s provided examiners with timing guidelines for the custom portions of MAP. Strict timing guidelines were given for the norm-referenced portions of the test. For MAP's custom sessions, examiners were instructed to allow students to complete the assessment if s/he was making adequate progress. For the norm-referenced portion of the test, students received an accommodation for additional time if so needed and documented on their individualized education plan. The timing schedule of the MAP is presented in Table 4.1.

4.4.2 Accommodations

Accommodations are allowed on MAP. Test accommodations may be used with students who qualify under IDEA and have an IEP or Section 504 of the Americans' with Disabilities Act and have a 504 plan, or who are identified as an English Language Learner. Accommodations must be specified in the qualifying student's individual plan and must be consistent with accommodations used during daily classroom instruction and testing. The use of any accommodation must be indicated on the student information sheet at the time of test administration. AERA, APA, and NCME (1999) standard 5.3, states

When formal procedures have been established for requesting and receiving accommodation, test takers should be informed of these procedures in advance of testing.

In compliance with this, the grade-specific MAP *Examiner's Manual* contains the list of accommodations permissible for the MAP assessments. The table of accommodations presented in the *Examiner's Manual* is shown in Tables 4.3 and 4.4. 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, such as an oral administration of the Communication Arts test or paraphrasing any of the tests. Detailed information regarding testing accommodations can be found at the DESE website:

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

Braille and large print forms are provided to students with vision disabilities.

Tables 4.5 through 4.7 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2009 MAP, the Braille edition of the 2009 MAP, and the large print edition of the 2009 MAP. The analyses in Tables 4.5 through 4.7 are based on census data and include only those students who received accommodations and received a scale score on the Communication Arts, Mathematics, or Science MAP.

In 2009, setting and timing accommodations appear to be the most frequently used for the Communication Arts, Mathematics, and Science MAP. For the Science and Mathematics MAP, having the test read aloud was also among the more frequently used accommodations. For the Mathematics MAP, using calculators was also among the more frequently used accommodations.

On the Braille and large print editions of MAP, the setting and timing accommodations are again among the most frequently used accommodations. Common accommodations for both the Braille and large print editions include using a scribe for the Communication Arts, Mathematics, and Science MAPs, having the test read aloud for the Mathematics and Science MAPs, and using a calculator for the Mathematics MAP.

4.5 Summary

In summary, the overall purpose of each of the test administration workshops and the ancillary materials is to keep districts informed about policies and procedures related to testing in general and the MAP program in particular. The information imparted is clearly related to standardizing the administration of the MAP, maintaining the security of the assessment, allowing access to the assessments for special populations by clearly delineating appropriate accommodations, and by providing guidance on appropriate interpretations of the test results. These communication and training efforts by DESE and the ancillary information developed by CTB/McGraw-Hill address multiple best practices

of the testing industry but in particular are related to the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

- Standard 3.19— The directions for test administration should be presented with sufficient clarity and emphasis so that it is possible for others to replicate adequately the administration conditions under which the data on reliability and validity, and, where appropriate, norms were obtained.
- Standard 3.20— The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample material, practice or sample questions, criteria for scoring, and a representative item identified with each major area in the test's classification or domain should be provided to the test takers prior to the administration of the test or included in the testing material as part of the standard administration instructions.
- Standard 5.1—Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer, unless the situation or a test taker's disability dictates that an exception should be made.
- Standard 5.2— Modifications or disruptions of standardized test administration procedures or scoring should be documented.
- Standard 5.3—When formal procedures have been established for requesting and receiving accommodation, test takers should be informed of these procedures in advance of testing.
- Standard 5.4—The testing environment should furnish reasonable comfort with minimal distractions.
- Standard 5.6—Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.
- Standard 5.7—Test users have the responsibility of protecting the security of test materials at all times.

Table 4. 1: MAP Administration Schedule Timing Guidelines by Session (Time in Minutes)

Grade	Session	Communication Arts	Mathematics	Science
3	1	40 - 55	40 - 55	
	2	60 - 90	65*	
	3	58 - 63**	35 - 45	
	4	50 - 65		
4	1	45 - 55	55 - 75	
	2	63*	65*	
	3	50 - 65	50 - 70	
5	1	45 - 55	40 - 55	65 - 85
	2	63*	65*	70 - 85**
	3	50 - 65	35 - 45	90 - 105
6	1	45 - 55	40 - 45	
	2	64*	65*	
	3	50 - 65	35 - 45	
7	1	45 - 55	40 - 55	
	2	60 - 90	65*	
	3	64*	35 - 45	
	4	50 - 65		
8	1	45 - 55	55 - 75	65 - 85
	2	64*	65*	70 - 85**
	3	50 - 65	50 - 70	90 - 105

*Strictly timed
**Strictly timed, times vary by form

Table 4. 2: Districts Granted a One-Week Extension of the MAP Testing Window

District
Arcadia Valley R II
Bismarck R V
Bloomfield R Xiv
Cape Girardeau 63
Crane R III
Delta R V
East Carter Co R II
Fairview R XI
Fox C 6
Glenwood R VIII
Junction Hill C 12
Kennett 39
Malden R I
Oregon Howell R III
Portageville
Potosi
Puxico R Viii
Richland R I
Richwoods
Risco R II
Riverview Gardens
Senath Hornersville C 8
Sikeston R 6
Southland C 9
Ste Genevieve Co R II
Twin Rivers R X
Van Buren R I
Woodland R IV
Zalma R V

Table 4. 3: MAP Accommodations for Students Who Are English Language Learners

Accommodations List for Students Who Are English Language Learners (ELL) The following are the only accommodations allowed for ELL students:			
Code	**Invalidates**	**Administration Accommodations**	**Description**
04	√	Oral reading of assessment (Not permissible for Communication Arts) See Note 1 (below).	The test examiner reads items verbatim to the student in an isolated setting so that other students will not benefit or be disturbed.
11	√	Oral reading in native language (Not permissible for Communication Arts) See Note 1 (below).	
		Timing Accommodations	**Description**
20		Extend time allotted to complete TerraNova Survey. See Note 2 (below).	ELL students may need to complete the assessments over more than one test period.
21		Administer test using more than allotted periods	Dates for taking the MAP must occur within the MAP testing window.
22		Other: Specify	Other timing accommodations.
		Response Accommodations	**Description**
35		Use of scribe to record student response in test booklet	The student conveys verbal responses to a scribe in an isolated, individual setting so that other students cannot benefit or be disturbed. The scribe cannot suggest ideas, words, or concepts. The scribe records the student's answers verbatim. The student should indicate capitalization and punctuation if language mechanics are being assessed.
		Oral response	The student provides an oral response to the examiner.
43	√	Use of bilingual dictionary (Not permissible for Communication Arts) See Note 1 (below).	
		Setting Accommodations	**Description**
50		Testing individually	The room should be free of noises, conversation, and distractions from adjoining rooms. Individual testing is appropriate when, for example, responses are given orally or questions are paraphrased.
51		Testing with small groups	The location should be free of noises, conversation, and distractions from adjoining rooms. Students may not interact with one another about questions or answers. The examiner must be present at all times. Testing in small groups is not appropriate for students who give responses orally or require paraphrasing of questions.
53		Other: Specify	Other setting accommodations.

NOTES
Note 1 *Oral reading, oral reading in native language, or the use of a bilingual dictionary during the Communication Arts test will result in the LOSS (Lowest Obtainable Scale Score).*
Note 2 *If used, the student score cannot be compared with scores generated under standard conditions.*

Table 4. 4: MAP Accommodations for Students with Disabilities

Accommodations List for Students with Disabilities			
Code	**Invalidates**	**Administration Accommodations**	**Description**
01		Braille edition of assessment	Braille editions of the assessment require special processing. Consult your Braille edition test materials for specific instructions.
02		Large Print edition of assessment	Large Print editions of the assessment require special processing. Consult your Large Print test materials for specific instructions.
04	√	Oral reading of assessment. See Note 1 (below).	The test examiner reads items verbatim to the student in an isolated setting so that other students will not benefit or be disturbed.
04		Oral reading of assessment to Blind/Partial Sight students. See Note 1.	The test examiner reads items verbatim to the student who cannot read Braille in an isolated setting so that other students will not benefit or be disturbed.
05		Signing	A certified sign language interpreter or deaf education instructor signs the Mathematics and/or Science test (directions and test items are allowed) and/or the directions only of the Communication Arts test to the student.
06	√	Paraphrasing See Note 2.	The test examiner paraphrases questions to help student understanding in an isolated setting. Terms may be defined as long as they: 1) are not the actual concept or content being assessed, 2) would not give clues, or 3) would not disclose the answer.
10		**Other administration accommodations**	
		Use of assistive devices	An assistive device, which permits a student to read and/or respond to the assessment, is used. Examples of assistive devices include computers that assist students with fine-motor problems, text enlargers that enable students to independently read and answer test questions, or augmentative communication devices.
		Use of visual aids: Specify	Visual aids include any type of optical or non-optical devices used to enhance visual capability. Examples of visual aids include bold-line felt-tip markers, lamps, filters, bold-lined paper, writing guides, or other adaptations that alter the visual environment by adjusting the space, illumination, color, contrast, or other physical features of the environment.
		Timing Accommodations	**Description**
20		Extend time allotted to complete TerraNova Survey. See Note 3.	Extended time to complete the TerraNova Survey is allowed for a student whose disability may cause him/her to be unable to meet time constraints.
21		Administer assessment using more than allotted periods	Students with disabilities may need to complete the assessments over more than one test period as a result of fatigue and/or loss of concentration. Some students may require additional breaks. Dates for taking the MAP must occur within the MAP testing window.
22		Other: Specify	Other timing accommodations
		Response Accommodations	**Description**
35		Use of scribe to record student response in test booklet	The student conveys verbally or signs responses to a scribe in an isolated, individual setting so that other students cannot benefit or be disturbed. The scribe cannot suggest ideas, words, or concepts. The scribe records the student's answers verbatim. The student should indicate capitalization and punctuation if language mechanics are being assessed.
		Student taped response	The student speaks responses into a tape recorder in an isolated setting so that other students cannot benefit or be disturbed. The test examiner must be present at all times.
		Signed response	The student uses sign language to convey responses. A certified sign language interpreter or deaf education instructor records responses.
		Pointing to respond	The student points to correct responses and the administrator records responses in the MAP test booklet.
		Oral response	The student provides oral responses to the test examiner.
		Use of a Brailler	A student records responses using a Brailler. Examples of a Brailler include a Braillewriter, a slate and stylus, or an electronic Brailler note taker.

NOTES
Note 1 *Oral reading of the Communication Arts test results in the LOSS (Lowest Obtainable Scale Score). Students identified as blind/ visually impaired (who do not read Braille) may use the oral reading accommodation if it is their primary instructional method.*
Note 2 *Paraphrasing test questions invalidates all MAP assessment student scores for accountability purposes.*
Note 3 *If used, the student score cannot be compared with scores generated under standard conditions.*
Note 4 *Use of magnifying equipment, amplification equipment, graph paper, and testing with the teacher facing the student are not listed as accommodations because these are no longer required to be reported as accommodations for the MAP tests.*

Table 4. 4: MAP Accommodations for Students with Disabilities (cont'd)

Accommodations List for Students with Disabilities			
Code	**Invalidates**	**Administration Accommodations**	**Description**
		Use of a communication device	The student uses a communication device to provide responses to the test examiner.
		Use of a computer/word processor/typewriter to respond	The student uses a computer/word processor to write the responses. (Provide a non-networked computer to avoid inappropriate use of the computer to access answers.) The student uses a typewriter to write the responses.
39		Use of a calculator/math table/ abacus	In sessions of the MAP where calculators are allowed, the accommodation code should not be marked. The use of a calculator represents an accommodation when it is used on a section of the assessment for which calculator use is not allowed. Students may use talking calculators, but only in an isolated setting. Students may use tables to assist in simple addition, subtraction, multiplication, and division facts using whole numbers. Students may use an abacus to perform mathematical computations by sliding beads along rods.
44		Other: Specify. See Note 4.	Other response accommodations
		Setting Accommodations	**Description**
50		Testing individually	The location should be free of noises, conversation, and distractions from adjoining rooms. Individual testing is appropriate when, for example, responses are given orally or questions are paraphrased.
51		Testing in small groups	The location should be free of noises, conversation, and distractions from adjoining rooms. Students may not interact with one another about questions or answers. The test examiner must be present at all times. Testing in small groups is not appropriate for students who give responses orally or require paraphrasing of questions.
53		Other: Specify	Other setting accommodations

Table 4. 5: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Regular Edition

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
3	Regular Edition	67123	100.00%	67195	100.00%		
	Oral reading	56	0.08%	4612	6.86%		
	Oral reading blind	5	0.01%				
	Signing of assessment	4	0.01%	17	0.03%		
	Paraphrasing	3	0.00%	2	0.00%		
	Other administration	128	0.19%	86	0.13%		
	Oral reading in native language	13	0.02%	157	0.23%		
	Extend time—*TerraNova* session	2869	4.27%	2887	4.30%		
	Administer using > allotted periods	2790	4.16%	2752	4.10%		
	Other timing	592	0.88%	567	0.84%		
	Use of scribe	2015	3.00%	1804	2.68%		
	Use of calculator, math table, etc.	63	0.09%	1513	2.25%		
	Use of bilingual dictionary	0	0.00%	43	0.06%		
	Other response	106	0.16%	106	0.16%		
	Testing individually	2085	3.11%	1972	2.93%		
	Testing in small group	4492	6.69%	4749	7.07%		
	Other setting	285	0.42%	281	0.42%		
4	Regular Edition	66448	100.00%	66544	100.00%		
	Oral reading	42	0.06%	4765	7.16%		
	Oral reading blind	6	0.01%				
	Signing of assessment	8	0.01%	20	0.03%		
	Paraphrasing	2	0.00%	5	0.01%		
	Other administration	121	0.18%	81	0.12%		
	Oral reading in native language	14	0.02%	203	0.31%		
	Extend time—*TerraNova* session	3017	4.54%	3134	4.71%		
	Administer using > allotted periods	3078	4.63%	3082	4.63%		
	Other timing	655	0.99%	652	0.98%		
	Use of scribe	1999	3.01%	1866	2.80%		
	Use of calculator, math table, etc.	82	0.12%	1976	2.97%		
	Use of bilingual dictionary	1	0.00%	53	0.08%		
	Other response	85	0.13%	102	0.15%		
	Testing individually	2162	3.25%	2060	3.10%		
	Testing in small group	4703	7.08%	4989	7.50%		
	Other setting	325	0.49%	329	0.49%		
5	Regular Edition	67025	100.00%	67097	100.00%	67060	100.00%
	Oral reading	28	0.04%	5113	7.62%	4948	7.38%
	Oral reading blind	6	0.01%				
	Signing of assessment	5	0.01%	24	0.04%	24	0.04%
	Paraphrasing	0	0.00%	6	0.01%	0	0.00%
	Other administration	159	0.24%	102	0.15%	97	0.14%

Table 4. 5: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Regular Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
5	Oral reading in native language	7	0.01%	176	0.26%	164	0.24%
	Extend time—*TerraNova* session	3104	4.63%	3257	4.85%	3002	4.48%
	Administer using > allotted periods	3308	4.94%	3287	4.90%	3140	4.68%
	Other timing	670	1.00%	675	1.01%	615	0.92%
	Use of scribe	1895	2.83%	1788	2.66%	1825	2.72%
	Use of calculator, math table, etc.	98	0.15%	2554	3.81%	871	1.30%
	Use of bilingual dictionary	1	0.00%	42	0.06%	57	0.08%
	Other response	115	0.17%	124	0.18%	123	0.18%
	Testing individually	2066	3.08%	1922	2.86%	1903	2.84%
	Testing in small group	5460	8.15%	5809	8.66%	5523	8.24%
	Other setting	342	0.51%	345	0.51%	336	0.50%
6	Regular Edition	65661	100.00%	65698	100.00%		
	Oral reading	38	0.06%	4424	6.73%		
	Oral reading blind	6	0.01%				
	Signing of assessment	2	0.00%	12	0.02%		
	Paraphrasing	2	0.00%	4	0.01%		
	Other administration	103	0.16%	65	0.10%		
	Oral reading in native language	8	0.01%	176	0.27%		
	Extend time—*TerraNova* session	2702	4.12%	2737	4.17%		
	Administer using > allotted periods	2753	4.19%	2756	4.19%		
	Other timing	566	0.86%	576	0.88%		
	Use of scribe	1328	2.02%	1121	1.71%		
	Use of calculator, math table, etc.	122	0.19%	3112	4.74%		
	Use of bilingual dictionary	1	0.00%	72	0.11%		
	Other response	90	0.14%	77	0.12%		
	Testing individually	1517	2.31%	1386	2.11%		
	Testing in small group	5631	8.58%	5872	8.94%		
	Other setting	223	0.34%	231	0.35%		
7	Regular Edition	66271	100.00%	66287	100.00%		
	Oral reading	37	0.06%	3777	5.70%		
	Oral reading blind	4	0.01%				
	Signing of assessment	9	0.01%	25	0.04%		
	Paraphrasing	11	0.02%	8	0.01%		
	Other administration	69	0.10%	49	0.07%		
	Oral reading in native language	14	0.02%	113	0.17%		
	Extend time—*TerraNova* session	2206	3.33%	2241	3.38%		
	Administer using > allotted periods	2440	3.68%	2412	3.64%		
	Other timing	532	0.80%	527	0.80%		
	Use of scribe	992	1.50%	716	1.08%		
	Use of calculator, math table, etc.	203	0.31%	3467	5.23%		
	Use of bilingual dictionary	1	0.00%	82	0.12%		

Table 4. 5: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Regular Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
7	Other response	54	0.08%	39	0.06%		
	Testing individually	1158	1.75%	994	1.50%		
	Testing in small group	5517	8.32%	5685	8.58%		
	Other setting	151	0.23%	168	0.25%		
8	Regular Edition	66691	100.00%	66719	100.00%	66654	100.00%
	Oral reading	51	0.08%	3622	5.43%	3640	5.46%
	Oral reading blind	4	0.01%				
	Signing of assessment	1	0.00%	19	0.03%	18	0.03%
	Paraphrasing	7	0.01%	5	0.01%	5	0.01%
	Other administration	73	0.11%	52	0.08%	51	0.08%
	Oral reading in native language	5	0.01%	130	0.19%	121	0.18%
	Extend time—*TerraNova* session	2167	3.25%	2206	3.31%	2110	3.17%
	Administer using > allotted periods	2371	3.56%	2381	3.57%	2314	3.47%
	Other timing	539	0.81%	553	0.83%	543	0.81%
	Use of scribe	718	1.08%	609	0.91%	672	1.01%
	Use of calculator, math table, etc.	156	0.23%	3697	5.54%	2403	3.61%
	Use of bilingual dictionary	2	0.00%	106	0.16%	111	0.17%
	Other response	44	0.07%	43	0.06%	44	0.07%
	Testing individually	906	1.36%	817	1.22%	840	1.26%
	Testing in small group	5582	8.37%	5812	8.71%	5637	8.46%
	Other setting	167	0.25%	165	0.25%	162	0.24%

Table 4. 6: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Braille Edition

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

Table 4. 6: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Braille Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
5	Oral reading in native language	0	0.00%	0	0.00%	0	0.00%
	Extend time—*TerraNova* session	2	33.33%	2	33.33%	1	16.67%
	Administer using > allotted periods	0	0.00%	0	0.00%	0	0.00%
	Other timing	1	16.67%	1	16.67%	1	16.67%
	Use of scribe	2	33.33%	3	50.00%	2	33.33%
	Use of calculator, math table, etc.	0	0.00%	2	33.33%	0	0.00%
	Use of bilingual dictionary	0	0.00%	0	0.00%	0	0.00%
	Other response	0	0.00%	0	0.00%	0	0.00%
	Testing individually	4	66.67%	4	66.67%	4	66.67%
	Testing in small group	0	0.00%	0	0.00%	0	0.00%
	Other setting	0	0.00%	0	0.00%	0	0.00%
6	Braille Edition	NR		NR			
	Oral reading						
	Oral reading blind						
	Signing of assessment						
	Paraphrasing						
	Other administration						
	Oral reading in native language						
	Extend time—*TerraNova* session						
	Administer using > allotted periods						
	Other timing						
	Use of scribe						
	Use of calculator, math table, etc.						
	Use of bilingual dictionary						
	Other response						
	Testing individually						
	Testing in small group						
	Other setting						
7	Braille Edition	8	100.00%	9	100.00%		
	Oral reading	0	0.00%	0	0.00%		
	Oral reading blind	0	0.00%				
	Signing of assessment	0	0.00%	0	0.00%		
	Paraphrasing	0	0.00%	0	0.00%		
	Other administration	1	12.50%	0	0.00%		
	Oral reading in native language	0	0.00%	0	0.00%		
	Extend time—*TerraNova* session	6	75.00%	6	66.67%		
	Administer using > allotted periods	5	62.50%	5	55.56%		
	Other timing	0	0.00%	0	0.00%		
	Use of scribe	3	37.50%	3	33.33%		
	Use of calculator, math table, etc.	0	0.00%	5	55.56%		
	Use of bilingual dictionary	0	0.00%	0	0.00%		

Table 4. 6: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Braille Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
7	Other response	0	0.00%	0	0.00%		
	Testing individually	4	50.00%	3	33.33%		
	Testing in small group	0	0.00%	0	0.00%		
	Other setting	0	0.00%	0	0.00%		
8	Braille Edition	8	100.00%	9	100.00%	7	100.00%
	Oral reading	0	0.00%	2	22.22%	1	14.29%
	Oral reading blind	1	12.50%				
	Signing of assessment	0	0.00%	0	0.00%	0	0.00%
	Paraphrasing	0	0.00%	0	0.00%	0	0.00%
	Other administration	1	12.50%	1	11.11%	1	14.29%
	Oral reading in native language	0	0.00%	0	0.00%	0	0.00%
	Extend time—*TerraNova* session	5	62.50%	6	66.67%	5	71.43%
	Administer using > allotted periods	7	87.50%	6	66.67%	6	85.71%
	Other timing	1	12.50%	1	11.11%	0	0.00%
	Use of scribe	4	50.00%	5	55.56%	3	42.86%
	Use of calculator, math table, etc.	1	12.50%	5	55.56%	1	14.29%
	Use of bilingual dictionary	0	0.00%	0	0.00%	0	0.00%
	Other response	0	0.00%	0	0.00%	0	0.00%
	Testing individually	6	75.00%	7	77.78%	5	71.43%
	Testing in small group	0	0.00%	0	0.00%	0	0.00%
	Other setting	0	0.00%	0	0.00%	0	0.00%

NR=Not reported due to sample size less than 5 students

Table 4. 7: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Large Print Edition

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
3	Large Print Edition	34	100.00%	32	100.00%		
	Oral reading	1	2.94%	16	50.00%		
	Oral reading blind	2	5.88%				
	Signing of assessment	1	2.94%	0	0.00%		
	Paraphrasing	1	2.94%	0	0.00%		
	Other administration	4	11.76%	4	12.50%		
	Oral reading in native language	1	2.94%	0	0.00%		
	Extend time—T*erraNova* session	16	47.06%	11	34.38%		
	Administer using > allotted periods	17	50.00%	15	46.88%		
	Other timing	3	8.82%	2	6.25%		
	Use of scribe	13	38.24%	11	34.38%		
	Use of calculator, math table, etc.	2	5.88%	8	25.00%		
	Use of bilingual dictionary	1	2.94%	0	0.00%		
	Other response	3	8.82%	2	6.25%		
	Testing individually	16	47.06%	14	43.75%		
	Testing in small group	16	47.06%	16	50.00%		
	Other setting	4	11.76%	2	6.25%		
4	Large Print Edition	38	100.00%	37	100.00%		
	Oral reading	0	0.00%	17	45.95%		
	Oral reading blind	3	7.89%				
	Signing of assessment	1	2.63%	0	0.00%		
	Paraphrasing	1	2.63%	0	0.00%		
	Other administration	1	2.63%	1	2.70%		
	Oral reading in native language	0	0.00%	0	0.00%		
	Extend time—*TerraNova* session	12	31.58%	10	27.03%		
	Administer using > allotted periods	12	31.58%	12	32.43%		
	Other timing	1	2.63%	1	2.70%		
	Use of scribe	22	57.89%	21	56.76%		
	Use of calculator, math table, etc.	0	0.00%	7	18.92%		
	Use of bilingual dictionary	0	0.00%	0	0.00%		
	Other response	1	2.63%	2	5.41%		
	Testing individually	17	44.74%	17	45.95%		
	Testing in small group	11	28.95%	11	29.73%		
	Other setting	2	5.26%	2	5.41%		
5	Large Print Edition	52	100.00%	52	100.00%	52	100.00%
	Oral reading	1	1.92%	26	50.00%	24	46.15%
	Oral reading blind	3	5.77%				
	Signing of assessment	0	0.00%	1	1.92%	1	1.92%
	Paraphrasing	0	0.00%	0	0.00%	0	0.00%
	Other administration	3	5.77%	2	3.85%	2	3.85%

Table 4. 7: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Large Print Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
5	Oral reading in native language	0	0.00%	0	0.00%	0	0.00%
	Extend time—*TerraNova* session	18	34.62%	19	36.54%	18	34.62%
	Administer using > allotted periods	16	30.77%	15	28.85%	15	28.85%
	Other timing	1	1.92%	1	1.92%	0	0.00%
	Use of scribe	23	44.23%	23	44.23%	23	44.23%
	Use of calculator, math table, etc.	0	0.00%	11	21.15%	5	9.62%
	Use of bilingual dictionary	0	0.00%	0	0.00%	0	0.00%
	Other response	3	5.77%	3	5.77%	3	5.77%
	Testing individually	27	51.92%	25	48.08%	24	46.15%
	Testing in small group	17	32.69%	21	40.38%	20	38.46%
	Other setting	3	5.77%	3	5.77%	2	3.85%
6	Large Print Edition	51	100.00%	53	100.00%		
	Oral reading	0	0.00%	21	39.62%		
	Oral reading blind	3	5.88%				
	Signing of assessment	0	0.00%	0	0.00%		
	Paraphrasing	0	0.00%	0	0.00%		
	Other administration	4	7.84%	3	5.66%		
	Oral reading in native language	0	0.00%	1	1.89%		
	Extend time—*TerraNova* session	17	33.33%	17	32.08%		
	Administer using > allotted periods	19	37.25%	19	35.85%		
	Other timing	5	9.80%	5	9.43%		
	Use of scribe	25	49.02%	25	47.17%		
	Use of calculator, math table, etc.	1	1.96%	19	35.85%		
	Use of bilingual dictionary	0	0.00%	0	0.00%		
	Other response	2	3.92%	1	1.89%		
	Testing individually	26	50.98%	24	45.28%		
	Testing in small group	17	33.33%	20	37.74%		
	Other setting	1	1.96%	1	1.89%		
7	Large Print Edition	37	100.00%	34	100.00%		
	Oral reading	1	2.70%	8	23.53%		
	Oral reading blind	6	16.22%				
	Signing of assessment	0	0.00%	0	0.00%		
	Paraphrasing	0	0.00%	0	0.00%		
	Other administration	1	2.70%	0	0.00%		
	Oral reading in native language	1	2.70%	0	0.00%		
	Extend time—*TerraNova* session	9	24.32%	9	26.47%		
	Administer using > allotted periods	12	32.43%	10	29.41%		
	Other timing	2	5.41%	2	5.88%		
	Use of scribe	17	45.95%	14	41.18%		
	Use of calculator, math table, etc.	0	0.00%	16	47.06%		
	Use of bilingual dictionary	0	0.00%	0	0.00%		

Table 4. 7: Number and Percent of Students Receiving Accommodations by Accommodation Type, MAP 2009 Large Print Edition (Cont'd)

Grade	Accommodation	Communication Arts		Mathematics		Science	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
7	Other response	2	5.41%	1	2.94%		
	Testing individually	15	40.54%	14	41.18%		
	Testing in small group	11	29.73%	10	29.41%		
	Other setting	2	5.41%	2	5.88%		
8	Large Print Edition	42	100.00%	42	100.00%	41	100.00%
	Oral reading	1	2.38%	18	42.86%	17	41.46%
	Oral reading blind	3	7.14%				
	Signing of assessment	0	0.00%	0	0.00%	0	0.00%
	Paraphrasing	0	0.00%	0	0.00%	0	0.00%
	Other administration	0	0.00%	0	0.00%	0	0.00%
	Oral reading in native language	0	0.00%	0	0.00%	0	0.00%
	Extend time—*TerraNova* session	10	23.81%	10	23.81%	10	24.39%
	Administer using > allotted periods	19	45.24%	18	42.86%	18	43.90%
	Other timing	3	7.14%	4	9.52%	3	7.32%
	Use of scribe	17	40.48%	17	40.48%	16	39.02%
	Use of calculator, math table, etc.	3	7.14%	17	40.48%	12	29.27%
	Use of bilingual dictionary	0	0.00%	0	0.00%	0	0.00%
	Other response	3	7.14%	2	4.76%	2	4.88%
	Testing individually	18	42.86%	17	40.48%	16	39.02%
	Testing in small group	13	30.95%	13	30.95%	13	31.71%
	Other setting	0	0.00%	0	0.00%	0	0.00%

Figure 4. 1: Sample Script of Examiner's Manual

Directions for Administering Mathematics

SESSION 1

Punch out all the rulers and pattern blocks prior to testing.

Before administering the test, be sure that students understand what each picture means.

This picture means that you will use your ruler.

This picture means that you will use your pattern blocks.

Teachers may keep the rulers and pattern blocks after the test is administered.

Distribute the test books, rulers, pattern blocks, and scratch paper. Scratch paper may include graph or grid paper. If this is the first day of testing, check to see that students write their names and district/school on their test books. Ensure that all students use a non-mechanical No. 2 pencil.

Before administering the test, take a moment to have your students look through the test book. Point out the STOP pages. Tell the students that whenever they see one of the STOP pages, they should not continue.

SAY

You may use these rulers and pattern blocks in Session 1. You may not need to use both of them.

For the questions in this session, you will select from a list of given answer choices. Use scratch paper to work the problems. Remember to fill in the circle in the test book that goes with the answer you choose.

You should read each question very carefully and do your best to answer clearly and completely. Your score on these questions will depend on how well you follow directions and show your understanding of mathematics.

Remember, in this session do NOT show any work in your test book UNLESS the question asks you to do so. You should mark ONLY the answer in your test book. Use the scratch paper provided by your teacher to figure out or solve the problem.

Open your test book to Session 1 of Mathematics.

Figure 4. 2: District Report Form

MAP Quality Assurance Concern
District Response Report

FOR DISTRICT/SCHOOL USE ONLY

Please return completed report to: Mr. Bill Gerling, Assistant Director of Assessment, Department of Elementary and Secondary Education, P. O. Box 480, Jefferson City, MO 65102-0480. The report is due within one month from date of receipt.

District Name: _____ **Building Name:** _____
County/District/School Code: _____

Please explain in detail the district response to the concern noted and the manner in which it has been/will be addressed to prevent future recurrence:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Principal Name (type or print): _____

Principal Signature: _____ **Date:** _____

Superintendent Name (type or print): _____

Superintendent Signature: _____ **Date:** _____

MAP District Response Report

Figure 4. 3: Test Book Accountability Form

MAP 2009 TEST BOOK ACCOUNTABILITY FORM

District Name: _____ District #: _____

School Name: _____ School #: _____

This form provides start-to-finish accountability for the MAP test books assigned to your school.

School Coordinator:

1. Complete the table below, providing any additional information on the back as required. Be sure to sign at the bottom of the page.
2. Include Braille, Large Print and accompanying standard editions on Lines 4, 6 and 7.
3. Retain a copy of this form for your own records, along with a photocopy of the security barcode ranges printed on the test book packages.
4. Return the completed form to your District Coordinator.

District Coordinator:

5. Complete a copy of this form for District Overage, providing any additional information on the back as required.
6. Fax the forms for all schools in your district to CTB - Fax # 866-405-4086. CTB may contact you to clarify any discrepancies on your schools' forms.

TEST BOOKS RECEIVED							
		GR 3	GR 4	GR 5	GR 6	GR 7	GR 8
(1)	Number of test books listed on packing list	+	+	+	+	+	+
(2)	Test books missing from shipment (Fill out A on back)	–	–	–	–	–	–
(3)	Extra books received in shipment (Fill out B on back)	+	+	+	+	+	+
(4)	Additional books from district office (Fill out C on back)	+	+	+	+	+	+
(5)	Total test books received (Add lines 1, 3 and 4; then subtract line 2)						
TEST BOOKS RETURNED							
		GR 3	GR 4	GR 5	GR 6	GR 7	GR 8
(6)	Number of tests administered	+	+	+	+	+	+
(7)	Number of unused test books (inc. Braille/LP)	+	+	+	+	+	+
(8)	Total test books returned (Sum of lines 6 and 7)						
TEST BOOKS NOT RETURNED							
		GR 3	GR 4	GR 5	GR 6	GR 7	GR 8
(9)	Test books securely destroyed (Fill out D on back)	+	+	+	+	+	+
(10)	Test books unaccounted for (Fill out D on back)	+	+	+	+	+	+
(11)	Total test books not returned (Sum of lines 9 - 10)						

I confirm that Line 5 = Line 8 + Line 11.

School Test Coordinator: _____ Date Faxed: _____

Print Name: _____

A. Security barcode numbers of test books missing from shipment	B. Security barcode numbers of extra test books received in shipment	C. Security barcode numbers of additional test books from district office, Braille or LP

D.	
Security barcode numbers of lost or securely destroyed test books	Explanation

CHAPTER 5: CONSTRUCTED-RESPONSE SCORING

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

Chapter 5 adheres to AERA, APA, & NCME Standards 3.22, 3.23, and 5.9. Each of these Standards will be presented in the pertinent section of this chapter. Standard 3.22 provides some general guidance for Chapter 5:

> Procedures for scoring and, if relevant, scoring criteria should be presented by the test developer in sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scores or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical if test can be scored locally.

Chapter 5 explains the procedures used for scoring the MAP constructed-response items. The scoring criteria used for each item is not presented in this chapter to preserve the integrity of the items for future use.

5.1 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.

5.1.1 Selection of Scoring Evaluators

AERA, APA, & NCME (1999) Standard 3.23 specifies:

> The process for selecting, training, and qualifying scorers should be documented by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the score scale, and the procedures for training scorers should result in a degree of agreement among scorers that allows for the scores to be interpreted as originally intended by the test developer. Scorer reliability and potential drift over time in raters' scoring standards should be evaluated and reported by the person(s) responsible for conducting the training session.

Sections 5.1.1 and 5.1.2 explain how scorers are selected and trained for the MAP handscoring process. Section 5.1.3 describes how the scorers are monitored throughout the MAP handscoring process.

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 2009 project was staffed with a large number of returning evaluators and team leaders who had previous experience with MAP and other handscoring projects. Kelly Services also recruited new team leaders and evaluators for employment. Recruitment sources included advertisements in newspapers in Indianapolis, Indiana; Mather, California, and nearby areas; and Internet sources.

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

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

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

When Kelly Services determined 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 nondisclosure agreement outlining the CTB/McGraw-Hill business ethics and security procedures.

5.1.2 Handscoring Training Process

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

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

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

Training Material Review Meeting
CTB prepared all anchors, scoring guides, and student response samples for DESE and Missouri participant review. 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 Block (RIB) consisting of one item to be scored, except in Grades 5 and 8 Science where there was one multi-item RIB. Evaluators and team leaders were trained using the following steps:

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

All evaluators were trained and qualified using the same procedures and criteria used for the team leaders. Qualification standards for every item were predetermined by DESE. In order to score an item, readers must have met the specific standards for that item. The qualification standards were:

- 4-point item: 80% qualification
- 3-point item: 80% qualification
- 2-point item: 90% qualification
- 1-point item: 100% qualification

5.1.3 Monitoring the Scoring Process

AERA, APA, & NCME (1999) Standard 5.9 says:

> When test scoring involves human judgment, scoring rubrics should specify criteria for scoring. Adherence to established scoring criteria should be monitored and checked regularly. Monitoring procedures should be documented.

Section 5.1.3 explains the monitoring procedures that CTB uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics are available for all CR items, which specify the criteria for scoring those CR items. These rubrics will not be presented here in order to preserve the integrity of the items for use in future MAP forms.

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

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

Approximately 5% of Communication Arts, Mathematics, and Science papers were scored by a second reader to establish inter-rater reliability statistics for all constructed-response items. This procedure is called a "double-blind read," because the second reader does not know the first reader's score.

Recalibration of Raters
Recalibration in handscoring refers to the process in which scorers/raters who begin to drift away from scoring accuracy are realigned to correct scoring. After a thorough review of the rubric, anchors, and training papers, a recalibration round is administered to a reader who has drifted; accuracy on this round must meet or exceed the qualification rate. A scorer who continues to exhibit drift is released.

5.1.4 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.

5.2 Inter-Rater Reliability

Approximately 5% of papers in Communication Arts, Mathematics, and Science were scored independently by a second reader. The statistics for the inter-rater reliability were calculated by form, and six different forms were administered at each grade. To determine the reliability of scoring, the mean percentage of perfect agreement and adjacent agreement between the two readers was averaged across the six forms. The standard deviation for each of these was also calculated.

For each item on each form, a weighted kappa was calculated to reflect the level of improvement beyond the chance level in the consistency of scoring. These weighted kappa values were averaged across forms and are presented in Tables 5.1 to 5.3. To aid in the interpretation of Kappa, the following cutoffs have been suggested (Landis & Koch, 1977; Altman, 1991):

Kappa Value	Strength of Agreement
0	None
<0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very good

All Communication Arts, Mathematics, and Science items show good inter-rater agreement. As shown in Table 5.1, raters demonstrated at least 92% perfect and adjacent agreement for all Communication Arts items. Except for two items, the strength of the inter-rater agreement may be interpreted as good or very good as indicated by the mean weighted Kappa values. One Grade 5 item (Session 1, Item 6B) and one Grade 7 item (Session 4, Item 3B) had Kappa values that indicate only moderate agreement between the raters.

As shown in Table 5.2, raters demonstrated above 98% perfect and adjacent agreement for all Mathematics items. The mean weighted Kappa values indicate that there was very good inter-rater agreement for all Mathematics items.

As shown in Table 5.3, raters demonstrated above 93% perfect and adjacent agreement for all Science items. The mean weighted Kappa statistic indicate good or very good inter-rater agreement for all Science items, with the exception of one Grade 8 item (Session 3, Item 7) that only had moderate agreement between the raters.

5.3 Summary

The information presented in this chapter summarizes the steps taken by CTB to ensure accuracy in the handscoring process. The inter-rater reliability statistics presented in Section 5.2 demonstrate that the items are scored reliably. These efforts by CTB address

multiple best practices of the testing industry, but are particularly related to AERA, APA, & NCME (1999) Standards 3.22, 3.23, and 5.9.:

- Standard 3.22—Procedures for scoring and, if relevant, scoring criteria should be presented by the test developer in sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scores or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical if test can be scored locally.
- Standard 3.23—The process for selecting, training, and qualifying scorers should be documented by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the score scale, and the procedures for training scorers should result in a degree of agreement among scorers that allows for the scores to be interpreted as originally intended by the test developer. Scorer reliability and potential drift over time in raters' scoring standards should be evaluated and reported by the person(s) responsible for conducting the training session.
- Standard 5.9—When test scoring involves human judgment, scoring rubrics should specify criteria for scoring. Adherence to established scoring criteria should be monitored and checked regularly. Monitoring procedures should be documented.

Table 5. 1: Inter-rater Reliability, Communication Arts

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
3	3	1	2	81	2	15	2	96	0.73	0.02
	4	1	2	83	2	17	2	99	0.79	0.02
	5	1	2	82	2	16	2	98	0.75	0.02
	6A	1	2	82	2	17	2	100	0.80	0.03
	6B	1	1	99	1	1	1	100	0.81	0.07
	6C	1	1	99	1	1	0	100	0.67	0.14
	1	2	4	66	2	32	2	97	0.65	0.04
	1	4	2	74	2	24	2	97	0.76	0.04
	2	4	2	89	1	11	1	99	0.91	0.02
	3	4	2	94	1	5	1	100	0.96	0.01
4	3	1	2	95	1	6	1	100	0.94	0.01
	4	1	2	72	1	26	1	98	0.70	0.01
	5	1	2	88	1	11	1	99	0.91	0.02
	6	1	2	88	1	11	1	99	0.86	0.03
	1	3	2	92	1	8	1	100	0.83	0.03
	2	3	2	84	1	14	1	98	0.84	0.02
	3A	3	2	79	1	19	1	97	0.79	0.02
	3B	3	2	90	1	10	1	99	0.80	0.03
5	3	1	2	83	1	15	2	98	0.77	0.02
	4	1	2	69	3	28	2	97	0.64	0.02
	5	1	2	70	2	27	1	97	0.67	0.05
	6A	1	2	78	2	19	2	97	0.75	0.03
	6B	1	1	89	2	10	2	100	0.48	0.09
	1	3	2	70	3	28	3	98	0.67	0.05
	2	3	2	73	2	26	1	99	0.73	0.03
	3	3	2	70	2	28	2	99	0.67	0.03
6	3	1	2	84	1	16	1	100	0.85	0.02
	4	1	2	77	2	22	2	99	0.77	0.02
	5	1	2	72	2	28	3	99	0.71	0.03
	6A	1	2	83	2	16	2	99	0.82	0.02
	6B	1	1	92	2	8	1	100	0.61	0.06
	1	3	2	82	1	17	1	99	0.79	0.03
	2	3	2	79	3	18	3	96	0.71	0.00
	3	3	2	68	2	29	2	97	0.62	0.03
7	3	1	2	75	1	24	1	99	0.74	0.02
	4	1	2	77	2	21	3	97	0.69	0.03
	5	1	2	73	2	26	3	99	0.72	0.02
	6A	1	2	76	1	23	1	99	0.78	0.03
	6B	1	2	88	1	12	1	100	0.84	0.02
	1	2	4	65	2	35	2	99	0.64	0.05

Table 5. 1: Inter-rater Reliability, Communication Arts (Cont'd)

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
7	1	4	2	75	1	23	2	98	0.73	0.02
	2	4	2	73	2	25	2	98	0.72	0.04
	3A	4	2	82	1	17	2	98	0.82	0.01
	3B	4	1	81	1	19	1	99	0.50	0.04
8	3	1	2	66	3	27	2	92	0.61	0.04
	4	1	2	80	2	19	2	98	0.79	0.02
	5	1	2	64	2	32	2	96	0.63	0.02
	6A	1	2	69	3	26	3	95	0.68	0.02
	6B	1	1	83	2	16	2	99	0.65	0.04
	1	3	2	68	1	30	1	98	0.66	0.03
	2	3	2	76	2	20	1	96	0.74	0.03
	3A	3	2	80	2	19	2	99	0.76	0.02
	3B	3	1	96	2	4	1	100	0.91	0.02
	3C	3	1	95	2	5	1	100	0.76	0.03

* The percent perfect & adjacent may not add up to 100 for 1-point items due to the percent discrepant. The percent discrepant includes the cases where one rater assigned a score and the other rater assigned a condition code. With 2- or more point items, it also refers to the cases where the assigned score varied by more than 1 point.

Table 5. 2: Inter-rater Reliability, Mathematics

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
	1	3	2	98	0	1	0	99	0.96	0.01
	2	3	2	98	1	2	1	100	0.98	0.01
	3	3	2	89	1	10	1	99	0.92	0.01
3	4	3	2	92	1	8	1	100	0.91	0.01
	5	3	2	88	2	12	2	100	0.87	0.02
	6	3	2	99	1	1	1	100	0.99	0.01
	7	3	2	93	2	7	1	100	0.91	0.02
	31	1	4	75	2	23	2	98	0.92	0.01
	1	3	2	93	1	7	1	100	0.86	0.01
	2	3	2	91	1	9	1	100	0.93	0.01
	3	3	2	93	1	6	1	99	0.91	0.02
	4	3	2	94	1	5	1	99	0.96	0.01
4	5	3	2	95	1	3	1	98	0.93	0.01
	6	3	2	95	1	5	1	100	0.94	0.01
	7	3	2	87	1	13	1	100	0.88	0.01
	8	3	2	97	1	2	1	99	0.97	0.01
	9	3	2	98	0	2	1	100	0.98	0.01
	1	3	2	98	1	2	1	100	0.97	0.01
	2	3	2	96	1	3	0	99	0.96	0.01
	3	3	2	75	2	25	2	100	0.81	0.02
5	4	3	2	91	2	9	2	99	0.93	0.02
	5	3	2	89	2	11	2	100	0.92	0.01
	6	3	2	81	2	19	2	99	0.77	0.04
	7	3	2	92	2	8	2	100	0.91	0.02
	1	3	2	87	1	13	1	100	0.86	0.02
	2	3	2	95	1	5	1	100	0.97	0.01
	3	3	2	93	1	7	1	100	0.93	0.01
6	4	3	2	97	1	3	0	100	0.98	0.01
	5	3	2	93	1	7	1	100	0.91	0.01
	6	3	2	98	1	1	1	98	0.96	0.00
	7	3	2	92	1	7	1	100	0.94	0.01
	1	3	2	94	1	6	1	100	0.94	0.00
	2	3	2	87	1	13	1	100	0.91	0.01
	3	3	2	91	1	8	1	100	0.94	0.01
7	4	3	2	95	1	5	1	100	0.96	0.01
	5	3	2	95	1	5	1	100	0.96	0.01
	6	3	2	96	0	4	1	100	0.97	0.00
	7	3	2	97	1	3	0	100	0.97	0.01
8	31	1	4	85	1	15	1	100	0.91	0.01
	1	3	2	95	1	4	1	99	0.95	0.02

Table 5. 2. Inter-rater Reliability, Mathematics (Cont'd)

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
8	2	3	2	95	2	4	2	99	0.95	0.02
	3	3	2	92	2	7	1	99	0.96	0.01
	4	3	2	95	1	5	1	100	0.93	0.02
	5	3	2	84	1	15	2	99	0.81	0.02
	6	3	2	97	1	3	1	100	0.95	0.01
	7	3	2	98	1	2	1	101	0.98	0.01
	8	3	2	94	1	6	1	100	0.92	0.02
	9	3	2	98	1	2	1	100	0.97	0.01

* The percent perfect & adjacent may not add up to 100 for 1-point items due to the percent discrepant. The percent discrepant includes the cases where one rater assigned a score and the other rater assigned a condition code in addition to the cases where the score assigned varied by more than 1 point.

Table 5. 3: Inter-rater Reliability, Science

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
	1	1	2	95	1	5	1	100	0.95	0.01
	2	1	2	83	1	16	1	99	0.84	0.02
	3	1	2	93	1	7	1	100	0.91	0.02
	4	1	2	81	1	18	1	100	0.82	0.01
	5	1	2	90	1	10	1	100	0.91	0.01
	7	1	2	87	2	12	2	100	0.87	0.02
	8	1	2	89	2	10	2	99	0.89	0.02
	9	1	2	81	2	17	2	98	0.82	0.02
	10	1	2	94	1	6	1	99	0.93	0.01
	11	1	2	89	2	9	2	98	0.88	0.02
	26	2	2	84	1	16	1	100	0.82	0.02
	27	2	2	77	2	22	2	99	0.76	0.04
	28	2	2	87	1	13	1	99	0.88	0.02
	29	2	2	97	1	3	1	100	0.97	0.00
	30	2	2	92	1	8	1	100	0.94	0.01
5	31	2	2	95	1	5	1	100	0.94	0.01
	32	2	2	97	1	3	1	100	0.98	0.01
	33	2	2	93	1	7	1	100	0.90	0.01
	34	2	2	96	1	4	1	99	0.97	0.01
	35	2	2	93	1	7	0	100	0.90	0.01
	36	2	2	91	2	9	2	100	0.90	0.03
	1	3	2	99	1	1	1	100	0.85	0.06
	2	3	4	84	2	11	1	95	0.93	0.01
	3	3	1	94	2	5	1	100	0.89	0.02
	4	3	1	99	1	1	1	100	0.97	0.01
	5	3	1	92	1	8	1	100	0.73	0.05
	6	3	2	79	2	20	2	99	0.79	0.02
	7	3	1	91	2	9	2	100	0.78	0.04
	8	3	1	86	1	14	1	100	0.66	0.02
	9	3	1	85	1	15	1	100	0.65	0.02
	10	3	1	98	1	2	1	100	0.96	0.01
	1	1	2	82	2	17	1	99	0.83	0.01
	2	1	2	92	2	8	1	100	0.93	0.02
	3	1	2	89	1	11	1	100	0.91	0.00
	4	1	2	87	1	13	1	100	0.88	0.02
8	5	1	2	72	1	26	2	98	0.71	0.02
	6	1	2	98	1	2	0	100	0.97	0.01
	7	1	2	94	1	6	1	100	0.87	0.03
	8	1	2	89	2	11	2	100	0.91	0.02
	9	1	2	86	1	13	1	100	0.82	0.03

Table 5. 3: Inter-rater Reliability, Science (Cont'd)

Grade	Itm #	Ses-sion	# Pts	Mean % Perfect	SD % Perfect	Mean % Adjacent	SD % Adjacent	% Perfect & Adjacent*	Kappa Mean	Kappa SD
	10	1	2	82	2	16	2	98	0.80	0.03
	11	1	2	88	2	12	1	100	0.84	0.05
	12	1	2	91	1	8	2	99	0.89	0.02
	26	2	2	89	1	11	1	100	0.88	0.01
	27	2	2	85	2	15	1	100	0.86	0.01
	28	2	2	87	1	13	1	100	0.90	0.01
	29	2	2	97	1	3	0	100	0.96	0.01
	30	2	2	97	1	3	1	100	0.96	0.01
	31	2	2	86	1	14	1	100	0.81	0.02
	32	2	2	95	1	4	1	99	0.94	0.01
	33	2	2	91	2	8	2	99	0.78	0.04
	34	2	2	86	1	14	1	100	0.84	0.02
	35	2	2	88	2	12	2	100	0.80	0.03
8	36	2	2	92	1	7	1	100	0.86	0.03
	37	2	2	85	2	14	2	99	0.79	0.03
	1	3	2	95	2	5	1	100	0.96	0.01
	2	3	2	95	1	5	1	100	0.97	0.01
	3	3	1	90	1	10	1	100	0.78	0.02
	4	3	1	99	1	0	1	99	0.99	0.01
	5	3	4	80	1	18	1	98	0.91	0.01
	6	3	1	91	2	8	1	99	0.84	0.03
	7	3	1	83	1	17	1	100	0.53	0.02
	8	3	1	91	1	9	1	100	0.81	0.01
	9	3	1	89	1	11	1	100	0.78	0.03
	10	3	2	87	2	13	1	99	0.86	0.01
	11	3	3	76	2	17	1	93	0.80	0.02
	12	3	1	85	2	15	1	100	0.67	0.03

* The percent perfect & adjacent may not add up to 100 for 1-point items due to the percent discrepant. The percent discrepant includes the cases where one rater assigned a score and the other rater assigned a condition code in addition to the cases where the score assigned varied by more than 1 point.

CHAPTER 6: OPERATIONAL DATA ANALYSES

This chapter of the MAP Technical Report describes the analyses that occurred on the operational data. These analyses include a classical item analysis and examination of the raw scores and an IRT analysis involving calibrating, scaling, and linking. All of these analyses were conducted using the calibration sample.

In this section, we first discuss the calibration sample. Next, we present the classical item statistics, including aggregate raw score statistics and individual item-level statistics. Then, we discuss the IRT models used for calibrating the data and address how well these models fit the Missouri data. If the IRT models fit the empirical item response distributions for the population for which we want to make generalizations (i.e., Missouri students), then the claim is strengthened that the scores are valid indicators of an underlying ability. The lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) for MAP are presented. Finally, we provide a general overview of scaling and discuss the methods used to link the MAP results to the *TerraNova* scale.

Chapter 6 demonstrates adherence in the MAP program to AERA, APA, & NCME (1999) Standards 1.5, 2.8, 3.18, 4.2, 4.11, 4.13, and 6.4. Each Standard will be explicated within the appropriate section of this chapter. Standard 6.4 provides general guidance that is relevant to this chapter. It states:

> The population for whom the test is intended and the test specifications should be documented. If applicable, the item pool and scale development procedures should be described in the relevant test manuals. If normative data are provided, the norming population should be described in terms of relevant demographic variables, and the year(s) in which the data were collected should be reported.

In section 6.1, we will discuss the calibration sample and compare it to the general population. The test specifications and item pool are discussed in Chapter 3. The scale development procedures are discussed in section 6.4 of this chapter. Information regarding reported data are discussed in detail in Chapter 7. Information on the normative data may be found in the *TerraNova, Third Edition: Technical Addendum Forms E and F* (2009).

6.1 Calibration Sample

In this section we describe the calibration sample in adherence to Standard 1.5 of the AERA, APA, & NCME (1999) Standards. Standard 1.5 states:

> The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics.

In 2009, the grade-level calibration samples were comprised of at least 80% of the total student population for that grade. Several large school districts were identified for

inclusion in the 80% sample. These districts are listed in Table 6.1. Data from these districts had to be included in the calibration sample before data analyses procedures could begin. These large districts were identified because past data processing has demonstrated that large districts often return data at the end of the data-return window while small districts often return data early in the data-return window. Since the calibration sample was going to be based on the first 80% of data to be returned, it was important to identify large districts to ensure the calibration data were representative of the state.

Tables 6.2 through 6.4 examine the representativeness of the calibration sample compared to the census data. These tables demonstrate that the calibration sample was representative of the state. It should be noted that data from private schools was inadvertently left in the calibration data. These data comprised no more than 0.08% of the data, and it was determined that their inclusion did not effect the results.

6.2 Classical 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.

6.2.1. Test-Level Statistics

Tables 6.5 through 6.7 present the number of items and score points on each test, as well as the mean and standard deviation of the raw scores, *p*-values and item-total correlations (also known as item discrimination values) for each grade level of Communication Arts, Mathematics, and Science, respectively.

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

6.2.2. Item-Level Statistics

Tables 6.8 through 6.13 present the item statistics for each item by grade/content area. The tables include test session, item booklet number and part (if applicable), *p*-values, item-total correlations (R_{it}), and omit rates for each item by grade/content area. The constructed-response (CR) items appear in the tables first, followed by the multiple-choice (MC) items.

p-value: The *p*-value is a measure of item difficulty. For a multiple-choice item, the *p*-value is calculated from the number of students who correctly responded to an item divided by the total number of students who attempted 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 more precise 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 Table 6.5, the average *p*-values associated with the Communication Arts MAP range from .67 (Grade 8) to .74 (Grade 4). The average *p*-values associated with the Mathematics MAP (Table 6.6) range from .57 (Grade 8) to .77 (Grade 3). The average *p*-values associated with the Science MAP (Table 6.7) range from .53 (Grade 8) to .59 (Grade 5).

It is important that one examines the range of *p*-values and not just the average *p*-value to determine whether a test measures well. It is desirable for the test to measure well throughout the range of skills present at a given grade. That is, it is important that the items measure the performance of both low-scoring and high-scoring students, as well as students in the center of the distribution. Having a range of *p*-values also helps to prevent floor and/or ceiling effects so that the test does not have large numbers of students at the minimum or maximum possible scores. The Communication Arts MAP has items with *p*-values ranging from the low 0.20s to the 0.90s (see Tables 6.8 through 6.13). The *p*-values on the Mathematics MAP tend to range from the 0.10s and 0.20s to the 0.90s (see Tables 6.8 through 6.13). The Science MAP has items with *p*-values ranging from the 0.10s to the 0.90s. (see Tables 6.10 and 6.13). Such a broad range of *p*-values indicates that the items measure well throughout the range of skills at a given grade, and hence 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, where the item score is included in the total 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. Any item with an item-total correlation below the .20 threshold was further analyzed to assure that the item was correctly keyed.

Omit Rates: The omit rate for each item indicate the percentage of students who did not answer the item. Omit rates can be used to examine possible speededness issues on tests. A test may be speeded if students do not have adequate time to answer all questions on the test. As a rule of thumb, an item is said to have a high-omit rate if more than 5% of students failed to respond to the item.

This examination of omit rates complies with Standards 2.8 and 3.18 of the AERA, APA, & NCME (1999) Standards. Both Standards are concerned with speededness of a test. Standard 2.8 states:

> Test users should be informed about the degree to which rate of work may affect examinee performance.

The results in this section will show that, overall, student test scores are not adversely affected by the rate at which they complete the test. In general, students have ample time to complete all sections of the test. Related to this, Standard 3.18 states:

> For tests that have time limits, test development research should examine the degree to which scores include a speed component and evaluate the appropriateness of that component, given the domain the test is designed to measure.

Again, the results presented in Tables 6.8 through 6.13 show that the majority of tests did not have a speed component. These results are particularly relevant to the *TerraNova* component of the test, which is a strictly timed administration. The results of our analyses suggest that the majority of students were able to complete the test in the prescribed amount of time.

In examining Tables 6.8 through 6.13, there were four tests that demonstrate a small amount of speededness at the end of an administration section, which are Grades 3, 4, and 7 Mathematics and Grade 8 Science. Items 22 and 23 from Session 2 of the Grade 3 Mathematics had omit rates above 5%. These items occurred at the end of an administration section within Session 2, indicating that students may not have had enough time to complete the section. Similarly, items 11 and 12 from Session 2 of the Grade 4 Mathematics had omit rates above 5%, and these items occurred at the end of an administration section. Item 9 from Session 2 of the Grade 7 Mathematics had an omit rate above 5% and occurred at the end of an administration section. In Grade 8 Science, items 36 and 37 in Session 2 had omit rates above 5%. In all cases, the items immediately preceding the aforementioned items had omit rates that were slowly creeping upward.

There were a limited number of other items that had high omit rates. Item 18 in Session 2 of the Grade 8 Science test had an omit rate above 5%. This may have been due to the layout of the item on the page. The items were displayed in two columns, and Item 18 was in the left-hand column. Students may not have seen the item.

Item 10 in Session 1 and item 6 in Session 3 of the Grade 8 Science test had omit rates above 10%. These were CR items. In the cases of item 10 in Session 1, this was a difficult item (*p*-value = .30). Item 6 in Session 3 was located under a large graphic and likely went unseen by students.

6.3 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 θ will respond correctly to multiple-choice item *j* is

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

In the equation, 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-ability student. Under the 2PPC model, the probability that a student with trait or scale score θ will respond in category *k* to

partial-credit item *j* is

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

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.

6.3.1 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, $10(m_j\text{-}1)$, 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, and cutoff values for flagging an item based on Z_j have been developed and were

used to identify items for the item review. The cutoff value is (N/1500 x 4) for a given test, where N is the sample size.

Twelve MAP operational items were flagged for poor fit. In Communication Arts, one item was flagged for poor fit in each of Grades 3 and 4, and two items were flagged for poor fit in Grade 8. In Mathematics, one item was flagged for poor fit in Grade 3, two items were flagged for poor fit in Grade 6, and four items were flagged for poor fit in Grade 8. In Science, one item was flagged for poor fit in Grade 8. Table 6.14 shows the chi-square statistic and the Z-statistic for each flagged item. The average percent across ten cells of observed percentage correct and predicted percentage correct is also provided. The difference between the observed and predicted percentages provides an indication of how well the modeled response curves reflect the empirical curves.

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

Figures 6.1 to 6.12 show the item characteristic curves for each of the misfitting MAP items. The smooth line in each of these figures represents the predicted relationship between examinee performance on the item and examinee ability, and the jagged line represents the observed relationship.[3] Large differences between the two lines indicate poor fit. Each figure also shows the distribution of theta scores, so that the fit between observed and predicted performance at different ability levels can be interpreted in light of the overall distribution of examinees.

With large numbers of observations such as there are for the Missouri calibration samples, items may be flagged for statistically significant differences; however, these differences may not be of practical importance. In the case of the twelve MAP items flagged for misfit, the differences do not seem to be of practical importance. Misfitting items that have content validity are often retained for use in one 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. No MAP test had more than four items flagged for misfit.

Figure 6.1 presents the ICC for Session 2, Item 1 (4-point constructed-response item) on the Grade 3 Communication Arts test. As shown, there is poor fit at the lower end of levels 1 through 3 (students who scored 0, 1, or 2 out of 4). Levels 4 and 5 show spikes at

[3] 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.

the higher ends of the ability distribution for students who scored 3 out of 4 and 4 out of 4, respectively.

Figure 6.2 presents the ICC for Session 2, Item 17 (SR item) on the Grade 4 Communication Arts test. This figure shows this is an easy item. There appears to be somewhat poor fit in the lower end of the ability distribution.

Figure 6.3 presents the ICC for Session 2, Item 11 (SR item) on the Grade 8 Communication Arts test. There is poor fit throughout the ability range.

Figure 6.4 presents the ICC for Session 2, Item 38 (SR item) on the Grade 8 Communication Arts test. There is poor fit throughout the ability range, particularly at the low end.

Figure 6.5 presents the ICC for Session 3, Item 7 (2- point CR item) on the Grade 3 Mathematics test. There is poor fit throughout the ability distribution of levels 1, 2, and 3.

Figure 6.6 presents the ICC for Session 2, Item 30 (SR item) on the Grade 6 Mathematics test. There is poor fit throughout the ability distribution.

Figure 6.7 presents the ICC for Session 3, Item 3 (2-point CR item) on the Grade 6 Mathematics test. As shown, there is poor fit at the low end of the ability distribution for level 1, throughout the distribution for level 2, and at the upper end for level 3.

Figure 6.8 presents the ICC for Session 2, Item 16 (SR item) on the Grade 8 Mathematics test. As shown, there is poor fit throughout the ability distribution for this item.

Figure 6.9 presents the ICC for Session 1, Item 20 (2-point CR item) on the Grade 8 Mathematics test. As shown, there is poor fit throughout the lower end of the ability distribution for this item.

Figure 6.10 presents the ICC for Session 3, Item 8 (SR item) on the Grade 8 Mathematics test. As shown, there is poor fit through the upper end of the ability distribution for all three levels of this item.

Figure 6.11 presents the ICC for Session 3, Item 4 (2-point CR item) on the Grade 8 Mathematics test. There is poor fit at the lower end of the ability distribution for level 1. There is poor fit throughout the ability distribution for level 2. Level 3 shows poor fit at the upper end of the ability distribution.

Figure 6.12 presents the ICC for Session 3, Item 5 (4-point CR item) on the Grade 8 Science test. As shown, there is poor fit at the low end of the ability distribution for levels 1 and 2. There is poor fit at the upper end of the ability distribution for levels 3, 4, and 5.

6.4 Scaling

The purpose of scaling a test is to enhance its validity by increasing the comparability of test takers' scores. In this section, we explicate the way in which the MAP scales are produced to comply with Standard 4.2 of the AERA, APA, & NCME (1999) Standards, which states:

> The construction of scales used for reporting scores should be described clearly in the test documents.

The MAP scores are produced using the three-parameter logistic, two-parameter partial credit (3PL/2PPC) IRT model (explained previously) 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 MC items and CR items. In PARDUX, items are calibrated based on IRT, using the 3PL model (Lord & 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 & Lord (1983).

6.4.1 Linking Methods

CTB uses a common-item, non-equivalent groups design to link the current year's assessment to the established MAP scale. The embedded *TerraNova* form serves as the anchor set, and the non-equivalent groups are comprised of at least 80% of the census data in each grade. After the initial IRT item calibration, item parameters were linked to the MAP scale using the Stocking & Lord (1983) equating procedure.

Standard 4.11 of the AERA, APA, & NCME (1999) Standards states:

> When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions or other linkages were established and on the accuracy of equating functions.

The Stocking & Lord (1983) procedure minimizes the mean squared difference between the two TCCs, one based on estimates from the previous calibration and the other on transformed estimates from the current calibration. Let $\hat{\psi}_j$ be the test characteristic curve based on estimates from a previous calibration and $\hat{\psi}_j^*$ be the test characteristic curve based on transformed estimates from the current calibration.

$$\hat{\psi}_j = \hat{\psi}(\theta_j) = \sum_{i=1}^{n} P_i(\theta_j; a_i, b_i, c_i),$$

$$\hat{\psi}_j^* = \hat{\psi}(\theta_j) = \sum_{i=1}^{n} P_i(\theta_j; \frac{a_i}{M_1}, M_1 b_i + M_2, c_i),$$

The TCC method determines the scaling constants (M_1 and M_2) by minimizing the following quadratic loss function (F):

$$F = \frac{1}{N} \sum_{a=1}^{N} (\hat{\psi}_j - \hat{\psi}_j^*)^2.$$

The standard error of the equating (SEE) is difficult and cumbersome to estimate for IRT equating procedures, like Stocking and Lord (Kolen & Brennan, 1995; Michaelides & Haertel, 2004). The estimation of the SEE is beyond the scope of this report. It is anticipated that the SEE would be small because 80% of the census data is used for the purposes of linking each year. The large sample size (55,000 +) should ensure that the equating estimates are fairly stable.

6.4.2 Anchor Items

AERA, APA, & NCME (1999) Standard 4.13 requires information about the anchors, stating:

> In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used, as in some IRT-based and classical equating studies, the representativeness and psychometric characteristics of anchor items should be presented.

The content representation of the anchor items is shown in Tables 3.4, 3.5, and 3.6 of Chapter 3. Appendix B provides further details on psychometric characteristics of the anchor items.

6.4.3 Vertical Scale

The scale on which the MAP scale scores are reported is based in part on the *Terra Nova* 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 the scale score at selected percentiles as grade level increases. Figures 6.13, 6.14, and 6.15 display

the scale scores for several points on the score distributions for each grade of the Communication Arts, Mathematics, and Science MAP, respectively. These scale scores indicate the growth, or change, in score by grade at the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles. Ideally, the scale score associated with each percentile will increase from grade to grade. Figure 6.13 shows the selected percentiles for the Communication Arts MAP. Considering all but the 1st and 99th percentiles, the scale scores progress upward from Grades 3 through 5 and then flatten from Grades 5 to 6 before continuing to progress upward again from Grade 7 to 8. At the 1st percentile, there is a decrease in scale score from Grades 5 to 7.

Figure 6.14 shows the selected percentiles for the Mathematics MAP. Except for the 1st percentile, there is an upward progression of scale scores across all grades. At the 1st percentile, there is a decrease in scale score between Grades 4 and 5 and Grades 6 and 7.

Figure 6.15 shows the selected percentiles for the Science MAP. There is an upward progression of scale scores across the two Science grades.

Figures 6.16 to 6.18 show the TCCs by grade for the MAP Communication Arts, Mathematics, and Science, respectively. 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. Figure 6.16 shows that the TCCs for Communication Arts for Grades 5, 6, and 7 overlap. Grades 5 and 6 TCCs are very close to each other, separating only in the middle of the TCCs. The Grade 7 TCC crosses the Grades 5 and 6 TCCs at the lower end. During the selection of the forms, the pre-equated TCCs were examined and efforts were made to further separate the Grades 5 through 7 TCCs while, at the same time, protecting against scale drift. The available item pool was insufficient to create tests that resulted in the optimal increases in test difficulty. For Grade 7, the mean scale score is higher than Grades 5 and 6. The Grades 5 and 6 mean scale scores were nearly identical. DESE continues to work on differentiating skills in these grades, which may help pull apart the Grades 5 and 6 TCCs.

For both Mathematics (Figure 6.17) and Science (Figure 6.18), the TCCs indicate that test difficulty increases with grade level.

6.4.4 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 LOSS and the HOSS. Table 6.15 shows the LOSS and HOSS values used for each grade of the Communication Arts, Mathematics, and Science MAPs.

6.5 Item-Pattern Scoring

MAP scale scores are derived using item-pattern scoring; thus, these scale scores are based on the student's responses to all items on a given test, and scale scores account for the characteristics of the items that are in the test (such as item difficulty). A scale score can be interpreted as a highly probable estimate of a student's ability in a given content area.

Using item-pattern scoring, a student's scale score is based on the student's responses to each item (his/her item-response vector). Each item uses optimal item weights in terms of item information, meaning that items do not contribute equally to the overall scale score. Students with the same raw score may be assigned to different scale scores, depending on which items they answered correctly.

The procedures applied here are similar to those followed in the development of the *TerraNova* and *TerraNova* 2nd edition tests. For additional information on the technical details of the item-pattern scoring, readers can also refer to Yen & Candell (1991) and to technical report for *TerraNova* 2nd Edition (CTB/McGraw-Hill, 2003).

6.6 Summary

In summary, the overall purpose of the operational data analysis is to ensure that the test items, as well as the overall test, are functioning appropriately. It also helps maintain the test scale across the years so that test results may be appropriately compared across years. The data analyses undertaken by CTB/McGraw-Hill address multiple best practices of the testing industry but in particular are related to the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

- Standard 1.5—The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics.
- Standard 2.8—Test users should be informed about the degree to which rate of work may affect examinee performance.
- Standard 3.18—For tests that have time limits, test development research should examine the degree to which scores include a speed component and evaluate the appropriateness of that component, given the domain the test is designed to measure.
- Standard 4.2—The construction of scales used for reporting scores should be described clearly in the test documents.
- Standard 4.11—When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions or other linkages were established and on the accuracy of equating functions.
- Standard 4.13—In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically

determined relationships among test scores. If anchor items are used, as in some IRT-based and classical equating studies, the representativeness and psychometric characteristics of anchor items should be presented.

- Standard 6.4—The population for who the test is intended and the test specifications should be documented. If applicable, the item pool and scale development procedures should be described in the relevant test manuals. If normative data are provided, the norming population should be described in terms of relevant demographic variables, and the year(s) in which the data were collected should be reported.

Table 6. 1: Large Districts that Were Included in the 80% Calibration Sample

District Name
Columbia
St Joseph
North Kansas
Springfield
Blue Springs
Lee's Summit
Kansas City
Fort Zumwalt
Francis Howell
Hazelwood
Ferguson Florrisant
Rockwood
Mehlville
Parkway
St. Louis City

Table 6. 2: Summary of Calibration and Census Data: Communication Arts

	Communication Arts, Grade 3				
	Calibration Sample		Census Data		Diff (Calib % - Census %)
	N	%	N	%	
All Students	59323		67163		
Gender					
Male	30424	51.29	34461	51.31	-0.02
Female	28838	48.61	32633	48.59	0.02
Unknown	61	0.10	69	0.10	0.00
Race/Ethnicity					
White	44495	75.00	50411	75.06	-0.06
Black	10668	17.98	12138	18.07	-0.09
Hispanic	2606	4.39	2924	4.35	0.04
Asian/Pacific Islander	1252	2.11	1341	2.00	0.11
Native American/Alaskan	233	0.39	271	0.40	-0.01
Unknown	69	0.12	78	0.12	0.00
	Communication Arts, Grade 4				
All Students	59975		66490		
Gender					
Male	30696	51.18	34054	51.22	-0.04
Female	29218	48.72	32370	48.68	0.04
Unknown	61	0.10	66	0.10	0.00
Race/Ethnicity					
White	45317	75.56	50182	75.47	0.09
Black	10650	17.76	11965	18.00	-0.24
Hispanic	2499	4.17	2738	4.12	0.05
Asian/Pacific Islander	1180	1.97	1244	1.87	0.10
Native American/Alaskan	270	0.45	297	0.45	0.00
Unknown	59	0.10	64	0.10	0.00
	Communication Arts, Grade 5				
All Students	55944		67083		
Gender					
Male	28747	51.39	34415	51.30	0.09
Female	27096	48.43	32545	48.51	-0.08
Unknown	101	0.18	123	0.18	0.00
Race/Ethnicity					
White	41930	74.95	50657	75.51	-0.56
Black	10344	18.49	12113	18.06	0.43
Hispanic	2245	4.01	2672	3.98	0.03
Asian/Pacific Islander	1085	1.94	1234	1.84	0.10
Native American/Alaskan	238	0.43	283	0.42	0.01
Unknown	102	0.18	124	0.18	0.00

Table 6. 2: Summary of Calibration and Census Data: Communication Arts (Cont'd)

	Communication Arts, Grade 6				
	Calibration Sample		Census Data		Diff (Calib % - Census %)
	N	%	N	%	
All Students	65691		65716		
Gender					
Male	33422	50.88	33450	50.90	-0.02
Female	32211	49.03	32208	49.01	0.02
Unknown	58	0.09	58	0.09	0.00
Race/Ethnicity					
White	50109	76.28	50138	76.29	-0.01
Black	11564	17.60	11567	17.60	0.00
Hispanic	2459	3.74	2460	3.74	0.00
Asian/Pacific Islander	1220	1.86	1213	1.85	0.01
Native American/Alaskan	292	0.44	291	0.44	0.00
Unknown	47	0.07	47	0.07	0.00
	Communication Arts, Grade 7				
All Students	62856		66316		
Gender					
Male	32042	50.98	33806	50.98	0.00
Female	30759	48.94	32457	48.94	0.00
Unknown	55	0.09	53	0.08	0.01
Race/Ethnicity					
White	48109	76.54	50989	76.89	-0.35
Black	11062	17.60	11486	17.32	0.28
Hispanic	2204	3.51	2296	3.46	0.05
Asian/Pacific Islander	1155	1.84	1207	1.82	0.02
Native American/Alaskan	273	0.43	287	0.43	0.00
Unknown	53	0.08	51	0.08	0.00
	Communication Arts, Grade 8				
All Students	64354		66741		
Gender					
Male	33099	51.43	34325	51.43	0.00
Female	31160	48.42	32317	48.42	0.00
Unknown	95	0.15	99	0.15	0.00
Race/Ethnicity					
White	49529	76.96	51012	76.43	0.53
Black	11090	17.23	11778	17.65	-0.42
Hispanic	2158	3.35	2317	3.47	-0.12
Asian/Pacific Islander	1192	1.85	1235	1.85	0.00
Native American/Alaskan	283	0.44	293	0.44	0.00
Unknown	102	0.16	106	0.16	0.00

Table 6. 3: Summary of Calibration and Census Data: Mathematics

	Mathematics, Grade 3				
	Calibration Sample		Census Data		Diff (Calib % - Census %)
	N	%	N	%	
All Students	59469		67232		
Gender					
Male	30516	51.31	34501	51.32	-0.01
Female	28892	48.58	32663	48.58	0.00
Unknown	61	0.10	68	0.10	0.00
Race/Ethnicity					
White	44556	74.92	50410	74.98	-0.06
Black	10678	17.96	12139	18.06	-0.10
Hispanic	2638	4.44	2953	4.39	0.05
Asian/Pacific Islander	1292	2.17	1379	2.05	0.12
Native American/Alaskan	235	0.40	273	0.41	-0.01
Unknown	70	0.12	78	0.12	0.00
	Mathematics, Grade 4				
All Students	60130		66587		
Gender					
Male	30800	51.22	34122	51.24	-0.02
Female	29269	48.68	32399	48.66	0.02
Unknown	61	0.10	66	0.10	0.00
Race/Ethnicity					
White	45366	75.45	50191	75.38	0.07
Black	10679	17.76	11982	17.99	-0.23
Hispanic	2524	4.20	2759	4.14	0.06
Asian/Pacific Islander	1231	2.05	1293	1.94	0.11
Native American/Alaskan	270	0.45	297	0.45	0.00
Unknown	60	0.10	65	0.10	0.00
	Mathematics, Grade 5				
All Students	56030		67155		
Gender					
Male	28788	51.38	34444	51.29	0.09
Female	27140	48.44	32588	48.53	-0.09
Unknown	102	0.18	123	0.18	0.00
Race/Ethnicity					
White	41960	74.89	50671	75.45	-0.56
Black	10349	18.47	12118	18.04	0.43
Hispanic	2267	4.05	2696	4.01	0.04
Asian/Pacific Islander	1114	1.99	1264	1.88	0.11
Native American/Alaskan	238	0.42	283	0.42	0.00
Unknown	102	0.18	123	0.18	0.00

Table 6. 3: Summary of Calibration and Census Data: Mathematics (Cont'd)

	Mathematics, Grade 6				
	Calibration Sample		Census Data		Diff (Calib % - Census %)
	N	%	N	%	
All Students	65774		65755		
Gender					
Male	33480	50.90	33468	50.90	0.00
Female	32234	49.01	32227	49.01	0.00
Unknown	60	0.09	60	0.09	0.00
Race/Ethnicity					
White	50133	76.22	50132	76.24	-0.02
Black	11575	17.60	11566	17.59	0.01
Hispanic	2476	3.76	2475	3.76	0.00
Asian/Pacific Islander	1249	1.90	1242	1.89	0.01
Native American/Alaskan	293	0.45	292	0.44	0.01
Unknown	48	0.07	48	0.07	0.00
	Mathematics, Grade 7				
All Students	62924		66330		
Gender					
Male	32081	50.98	33803	50.96	0.02
Female	30790	48.93	32476	48.96	-0.03
Unknown	53	0.08	51	0.08	0.00
Race/Ethnicity					
White	48135	76.50	50976	76.85	-0.35
Black	11058	17.57	11473	17.30	0.27
Hispanic	2230	3.54	2316	3.49	0.05
Asian/Pacific Islander	1180	1.88	1232	1.86	0.02
Native American/Alaskan	271	0.43	285	0.43	0.00
Unknown	50	0.08	48	0.07	0.01
	Mathematics, Grade 8				
All Students	64432		66770		
Gender					
Male	33152	51.45	34345	51.44	0.01
Female	31186	48.40	32327	48.42	-0.02
Unknown	94	0.15	98	0.15	0.00
Race/Ethnicity					
White	49540	76.89	50998	76.38	0.51
Black	11109	17.24	11779	17.64	-0.40
Hispanic	2184	3.39	2338	3.50	-0.11
Asian/Pacific Islander	1215	1.89	1257	1.88	0.01
Native American/Alaskan	282	0.44	292	0.44	0.00
Unknown	102	0.16	106	0.16	0.00

Table 6. 4: Summary of Calibration and Census Data: Science

	Science, Grade 5				
	Calibration Sample		Census Data		Diff (Calib % - Census %)
	N	%	N	%	
All Students	55996		67118		
Gender					
Male	28766	51.37	34423	51.29	0.08
Female	27129	48.45	32573	48.53	-0.08
Unknown	101	0.18	122	0.18	0.00
Race/Ethnicity					
White	41938	74.89	50651	75.47	-0.58
Black	10340	18.47	12105	18.04	0.43
Hispanic	2265	4.04	2692	4.01	0.03
Asian/Pacific Islander	1114	1.99	1265	1.88	0.11
Native American/Alaskan	237	0.42	282	0.42	0.00
Unknown	102	0.18	123	0.18	0.00
	Science, Grade 8				
All Students	64354		66702		
Gender					
Male	33101	51.44	34297	51.42	0.02
Female	31158	48.42	32306	48.43	-0.01
Unknown	95	0.15	99	0.15	0.00
Race/Ethnicity					
White	49519	76.95	50967	76.41	0.54
Black	11064	17.19	11753	17.62	-0.43
Hispanic	2175	3.38	2329	3.49	-0.11
Asian/Pacific Islander	1212	1.88	1254	1.88	0.00
Native American/Alaskan	281	0.44	292	0.44	0.00
Unknown	103	0.16	107	0.16	0.00

Table 6. 5: MAP Means, Standard Deviations for Raw Scores, *p*-values, Item-Total Correlation (R_{it}): Communication Arts 2009

Grade	Total Items	Total Points	Mean Raw Score (SD)	Mean *p*-value (SD)	Mean R_{it} (SD)
3	57	67	47.58 (10.91)	0.72 (0.16)	0.38 (0.08)
4	55	63	44.97 (10.48)	0.74 (0.17)	0.40 (0.08)
5	55	62	42.63 (11.11)	0.70 (0.15)	0.40 (0.10)
6	55	62	41.50 (10.51)	0.70 (0.17)	0.38 (0.08)
7	61	72	49.43 (12.15)	0.70 (0.15)	0.39 (0.08)
8	61	68	44.69 (11.42)	0.67 (0.19)	0.37 (0.09)

Table 6. 6: MAP Means, Standard Deviations for Raw Scores, *p*-values, Item-Total Correlation (R_{it}): Mathematics 2009

Grade	Total Items	Total Points	Mean Raw Score (SD)	Mean *p*-value (SD)	Mean R_{it} (SD)
3	60	67	50.91 (10.92)	0.77 (0.14)	0.39 (0.09)
4	65	77	53.59 (13.52)	0.72 (0.16)	0.39 (0.10)
5	62	69	45.69 (12.11)	0.69 (0.16)	0.38 (0.09)
6	61	68	45.56 (12.41)	0.69 (0.13)	0.39 (0.08)
7	62	69	41.37 (13.29)	0.62 (0.17)	0.39 (0.10)
8	64	76	40.75 (14.13)	0.57 (0.20)	0.40 (0.11)

Table 6. 7: MAP Means, Standard Deviations for Raw Scores, *p*-values, Item-Total Correlation (R_{it}): Science 2009

Grade	Total Items	Total Points	Mean Raw Score (SD)	Mean *p*-value (SD)	Mean R_{it} (SD)
5	53	79	43.79 (13.65)	0.59 (0.22)	0.39 (0.11)
8	59	91	43.20 (15.63)	0.53 (0.24)	0.42 (0.10)

Table 6. 8: Item Statistics: Grade 3

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
1	1	0.91	0.36	0.11	1	1	0.66	0.44	0.77
1	2	0.90	0.43	0.11	1	2	0.74	0.34	0.18
1	3	0.71	0.23	0.33	1	3	0.89	0.40	0.12
1	4	0.74	0.40	0.41	1	4	0.79	0.39	0.22
1	5	0.78	0.32	0.45	1	5	0.90	0.48	0.22
1	6A	0.59	0.47	0.60	1	6	0.86	0.35	0.21
1	6B	0.97	0.30	0.59	1	7	0.88	0.36	0.31
1	6C	0.98	0.25	0.59	1	8	0.87	0.47	0.21
1	7	0.53	0.30	0.25	1	9	0.65	0.44	0.44
1	8	0.87	0.30	0.28	1	10	0.69	0.09	0.47
1	9	0.70	0.41	0.30	1	11	0.81	0.50	0.39
1	10	0.64	0.28	0.29	1	12	0.79	0.33	0.29
1	11	0.75	0.31	0.37	1	13	0.74	0.47	0.36
1	12	0.45	0.23	0.65	1	14	0.79	0.34	0.49
2	1	0.69	0.48	0.32	1	15	0.64	0.53	0.50
3	1	0.94	0.40	0.16	1	16	0.65	0.42	0.50
3	2	0.96	0.37	0.24	1	17	0.54	0.45	0.44
3	3	0.76	0.40	0.33	1	18	0.59	0.45	0.31
3	4	0.87	0.31	0.37	1	19	0.88	0.44	0.24
3	5	0.96	0.33	0.19	1	20	0.81	0.41	0.62
3	6	0.97	0.33	0.27	1	21	0.83	0.42	0.58
3	7	0.87	0.42	0.30	1	22	0.92	0.34	0.42
3	8	0.79	0.39	1.31	1	23	0.84	0.38	0.43
3	9	0.62	0.27	2.29	2	1	0.93	0.28	0.15
3	10	0.90	0.44	0.39	2	2	0.90	0.29	0.29
3	11	0.71	0.30	0.46	2	3	0.77	0.34	0.36
3	12	0.70	0.28	0.65	2	4	0.60	0.40	2.07
3	13	0.49	0.33	0.89	2	5	0.74	0.48	2.05
3	14	0.39	0.32	1.98	2	6	0.93	0.33	0.70
3	15	0.79	0.50	0.77	2	7	0.73	0.34	1.65
3	16	0.57	0.27	1.17	2	8	0.52	0.35	2.68
3	17	0.77	0.31	1.51	2	9	0.91	0.37	0.39
3	18	0.79	0.47	2.37	2	10	0.85	0.30	0.15
3	19	0.78	0.44	2.69	2	11	0.99	0.21	0.27
3	20	0.82	0.50	3.40	2	12	0.94	0.35	0.69

Table 6. 8: Item Statistics: Grade 3 (Cont'd)

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
3	21	0.88	0.37	4.04	2	13	0.90	0.30	0.36
3	22	0.83	0.47	0.31	2	14	0.93	0.31	0.31
3	23	0.69	0.49	0.76	2	15	0.75	0.51	0.44
3	24	0.81	0.27	0.77	2	16	0.94	0.39	0.81
3	25	0.54	0.33	0.77	2	17	0.95	0.29	0.40
3	26	0.81	0.47	1.06	2	18	0.89	0.38	0.43
3	27	0.37	0.23	1.73	2	19	0.74	0.31	1.32
3	28	0.52	0.39	2.47	2	20	0.96	0.31	0.24
3	29	0.64	0.30	0.22	2	21	0.61	0.42	1.04
3	30	0.53	0.48	0.26	2	22	0.74	0.26	5.92
3	31	0.61	0.38	0.36	2	23	0.86	0.39	5.49
3	32	0.79	0.46	0.65	2	24	0.88	0.42	0.85
3	33	0.66	0.45	0.96	2	25	0.81	0.40	0.69
3	34	0.79	0.51	0.99	2	26	0.74	0.22	0.31
3	35	0.82	0.33	0.88	2	27	0.79	0.34	0.30
3	36	0.83	0.36	0.36	2	28	0.54	0.48	0.55
3	37	0.68	0.39	0.54	2	29	0.69	0.47	0.55
3	38	0.65	0.38	0.68	2	30	0.61	0.46	0.89
3	39	0.48	0.31	1.05	3	1	0.93	0.41	0.27
4	1	0.46	0.23	0.47	3	2	0.71	0.58	0.22
4	2	0.63	0.47	0.55	3	3	0.42	0.54	0.19
4	3	0.69	0.58	0.58	3	4	0.57	0.43	0.41
					3	5	0.50	0.46	0.45
					3	6	0.82	0.40	0.82
					3	7	0.72	0.32	0.27

Table 6. 9: Item Statistics: Grade 4

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
1	1	0.76	0.38	0.09	1	1	0.80	0.45	0.54
1	2	0.67	0.48	0.10	1	2	0.82	0.41	0.18
1	3	0.52	0.35	0.41	1	3	0.91	0.29	0.18
1	4	0.62	0.43	0.52	1	4	0.67	0.34	0.24
1	5	0.60	0.47	0.18	1	5	0.66	0.40	0.21
1	6	0.24	0.39	0.42	1	6	0.90	0.40	0.44
1	7	0.91	0.22	0.13	1	7	0.78	0.42	0.24
1	8	0.92	0.39	0.27	1	8	0.41	0.33	0.33
1	9	0.86	0.33	0.22	1	9	0.95	0.32	0.13
1	10	0.85	0.38	0.38	1	10	0.51	0.31	0.56
1	11	0.74	0.34	0.38	1	11	0.86	0.44	0.30
1	12	0.55	0.29	0.40	1	12	0.53	0.37	0.29
2	1	0.84	0.38	0.13	1	13	0.74	0.36	0.28
2	2	0.98	0.22	0.15	1	14	0.85	0.25	0.32
2	3	0.87	0.39	0.23	1	15	0.72	0.29	0.35
2	4	0.90	0.41	0.46	1	16	0.60	0.24	0.47
2	5	0.86	0.44	0.25	1	17	0.67	0.40	0.31
2	6	0.82	0.34	0.58	1	18	0.82	0.33	0.39
2	7	0.97	0.34	0.49	1	19	0.59	0.42	0.47
2	8	0.74	0.42	0.22	1	20	0.72	0.54	0.45
2	9	0.50	0.44	0.37	1	21	0.54	0.34	0.62
2	10	0.76	0.38	0.26	1	22	0.53	0.38	0.28
2	11	0.52	0.23	0.24	1	23	0.80	0.49	0.27
2	12	0.82	0.43	0.40	1	31	0.34	0.64	0.43
2	13	0.40	0.25	0.31	2	1	0.90	0.20	0.18
2	14	0.83	0.37	0.66	2	2	0.77	0.43	0.38
2	15	0.83	0.45	0.95	2	3	0.73	0.37	1.14
2	16	0.86	0.38	0.64	2	4	0.69	0.36	1.63
2	17	0.93	0.40	0.85	2	5	0.87	0.46	0.56
2	18	0.88	0.49	1.53	2	6	0.62	0.54	1.17
2	19	0.94	0.42	1.04	2	7	0.49	0.42	2.23
2	20	0.62	0.38	1.45	2	8	0.87	0.39	2.74
2	21	0.87	0.37	1.39	2	9	0.92	0.36	3.55
2	22	0.82	0.48	0.19	2	10	0.70	0.51	4.78
2	23	0.87	0.38	0.26	2	11	0.70	0.31	6.27

Table 6. 9: Item Statistics: Grade 4 (Cont'd)

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
2	24	0.88	0.44	0.49	2	12	0.72	0.33	7.74
2	25	0.85	0.42	0.27	2	13	0.93	0.30	0.15
2	26	0.66	0.41	0.61	2	14	0.87	0.22	0.28
2	27	0.70	0.42	0.86	2	15	0.71	0.41	2.39
2	28	0.84	0.35	0.46	2	16	0.92	0.13	0.32
2	29	0.87	0.40	0.43	2	17	0.90	0.35	0.51
2	30	0.89	0.39	0.52	2	18	0.50	0.47	0.53
2	31	0.80	0.57	1.69	2	19	0.66	0.49	0.34
2	32	0.84	0.30	3.56	2	20	0.67	0.52	0.41
2	33	0.56	0.22	0.97	2	21	0.78	0.34	0.61
2	34	0.83	0.50	1.08	2	22	0.84	0.38	0.42
2	35	0.48	0.32	1.41	2	23	0.72	0.48	1.09
2	36	0.82	0.37	1.53	2	24	0.93	0.21	0.55
2	37	0.73	0.53	1.72	2	25	0.83	0.43	0.58
2	38	0.72	0.47	2.02	2	26	0.95	0.23	0.40
2	39	0.72	0.47	2.16	2	27	0.85	0.43	0.41
3	1	0.43	0.44	0.37	2	28	0.81	0.44	0.54
3	2	0.43	0.39	0.40	2	29	0.87	0.50	0.43
3	3A	0.49	0.44	0.47	2	30	0.98	0.23	0.36
3	3B	0.89	0.46	0.46	2	31	0.95	0.31	0.38
					2	32	0.60	0.31	0.25
					3	1	0.88	0.35	0.43
					3	2	0.66	0.60	0.31
					3	3	0.68	0.44	0.41
					3	4	0.42	0.47	0.92
					3	5	0.60	0.51	0.54
					3	6	0.47	0.42	0.30
					3	7	0.61	0.52	0.41
					3	8	0.53	0.33	0.43
					3	9	0.68	0.50	0.49

Table 6. 10: Item Statistics: Grade 5

Communication Arts					Mathematics					Science				
Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate
1	1	0.67	0.37	0.08	1	1	0.81	0.36	0.08	1	1	0.65	0.49	0.53
1	2	0.87	0.39	0.11	1	2	0.66	0.41	0.21	1	2	0.52	0.51	0.59
1	3	0.54	0.51	0.29	1	3	0.73	0.24	0.20	1	3	0.54	0.44	0.35
1	4	0.56	0.45	0.40	1	4	0.57	0.26	0.31	1	4	0.45	0.51	1.21
1	5	0.64	0.52	0.29	1	5	0.65	0.42	0.31	1	5	0.43	0.50	0.47
1	6A	0.78	0.57	0.87	1	6	0.80	0.35	0.21	1	7	0.53	0.47	0.61
1	6B	0.90	0.41	0.86	1	7	0.69	0.32	0.21	1	8	0.29	0.48	1.03
1	7	0.65	0.30	0.11	1	8	0.53	0.30	0.16	1	9	0.60	0.33	0.91
1	8	0.65	0.27	0.18	1	9	0.54	0.24	0.18	1	10	0.22	0.36	1.16
1	9	0.62	0.21	0.25	1	10	0.76	0.39	0.28	1	11	0.23	0.48	2.46
1	10	0.64	0.38	0.27	1	11	0.91	0.37	0.26	2	1	0.92	0.19	0.14
1	11	0.49	0.25	0.35	1	12	0.77	0.51	0.39	2	2	0.93	0.27	0.18
1	12	0.78	0.28	0.31	1	13	0.68	0.31	0.27	2	3	0.92	0.25	0.16
2	1	0.92	0.29	0.20	1	14	0.30	0.17	0.60	2	4	0.91	0.22	0.20
2	2	0.84	0.43	0.29	1	15	0.72	0.38	0.21	2	6	0.88	0.33	0.40
2	3	0.62	0.41	0.30	1	16	0.52	0.18	0.26	2	7	0.91	0.27	0.48
2	4	0.88	0.32	0.21	1	17	0.82	0.30	0.28	2	8	0.95	0.29	0.31
2	5	0.64	0.36	0.40	1	18	0.71	0.32	0.30	2	9	0.96	0.22	0.41
2	6	0.83	0.41	0.95	1	19	0.64	0.42	0.37	2	10	0.89	0.38	1.99
2	7	0.78	0.38	1.50	1	20	0.74	0.41	0.37	2	11	0.70	0.14	4.76
2	8	0.77	0.44	0.43	1	21	0.71	0.43	0.49	2	12	0.66	0.14	0.30
2	9	0.71	0.45	0.55	1	22	0.44	0.36	0.28	2	13	0.58	0.24	0.31
2	10	0.88	0.40	1.82	1	23	0.55	0.32	0.13	2	14	0.84	0.42	0.26
2	11	0.78	0.25	2.81	2	1	0.76	0.32	0.27	2	15	0.69	0.44	0.38
2	12	0.90	0.45	0.55	2	2	0.77	0.28	0.37	2	16	0.56	0.50	0.81
2	13	0.60	0.20	0.82	2	3	0.78	0.37	3.64	2	17	0.69	0.41	1.31
2	14	0.75	0.33	1.71	2	4	0.86	0.48	0.31	2	20	0.49	0.35	0.28
2	15	0.63	0.35	0.98	2	5	0.74	0.43	0.55	2	21	0.65	0.42	0.49
2	16	0.85	0.49	2.60	2	6	0.72	0.50	0.96	2	22	0.51	0.36	1.47
2	17	0.93	0.36	1.15	2	7	0.86	0.36	2.72	2	23	0.53	0.48	0.34
2	18	0.76	0.46	1.48	2	8	0.81	0.35	2.55	2	24	0.35	0.30	0.61
2	19	0.70	0.48	5.10	2	9	0.58	0.33	3.66	2	25	0.40	0.47	0.30
2	20	0.78	0.27	1.94	2	10	0.92	0.34	0.19	2	26	0.72	0.46	0.21
2	21	0.80	0.41	2.25	2	11	0.84	0.54	0.36	2	27	0.62	0.44	0.25
2	22	0.95	0.40	0.13	2	12	0.77	0.53	0.67	2	28	0.62	0.41	0.66

Table 6. 10: Item Statistics: Grade 5 (Cont'd)

Communication Arts					Mathematics					Science				
Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate
2	23	0.81	0.53	0.16	2	13	0.89	0.42	1.39	2	29	0.64	0.56	0.58
2	24	0.69	0.20	0.29	2	14	0.73	0.44	1.71	2	30	0.46	0.56	0.50
2	25	0.69	0.46	0.25	2	15	0.55	0.41	2.31	2	31	0.51	0.43	0.39
2	26	0.51	0.43	0.64	2	16	0.69	0.46	0.17	2	32	0.50	0.55	0.69
2	27	0.45	0.30	0.98	2	17	0.96	0.28	0.16	2	33	0.46	0.38	0.90
2	28	0.62	0.48	0.35	2	18	0.85	0.38	0.24	2	34	0.48	0.58	1.78
2	29	0.67	0.46	0.36	2	19	0.82	0.35	0.33	2	35	0.36	0.37	0.52
2	30	0.84	0.45	0.25	2	20	0.85	0.39	0.54	2	36	0.37	0.41	0.66
2	31	0.76	0.49	0.30	2	21	0.94	0.19	0.67	3	1	0.99	0.12	0.56
2	32	0.78	0.43	0.49	2	22	0.82	0.42	0.38	3	2	0.37	0.48	1.99
2	33	0.93	0.42	0.32	2	23	0.81	0.41	0.39	3	3	0.44	0.45	1.87
2	34	0.77	0.57	0.54	2	24	0.60	0.46	0.42	3	4	0.74	0.49	1.22
2	35	0.63	0.33	0.61	2	25	0.84	0.27	0.53	3	5	0.16	0.22	0.81
2	36	0.42	0.29	0.49	2	26	0.75	0.29	0.36	3	6	0.64	0.38	1.19
2	37	0.80	0.53	0.75	2	27	0.90	0.32	0.45	3	7	0.76	0.32	0.82
2	38	0.74	0.43	0.88	2	28	0.62	0.45	1.44	3	8	0.30	0.37	1.03
2	39	0.37	0.23	0.94	2	29	0.59	0.46	0.47	3	9	0.32	0.40	2.07
3	1	0.31	0.36	0.35	2	30	0.60	0.47	0.70	3	10	0.41	0.45	0.76
3	2	0.41	0.42	0.46	2	31	0.56	0.31	0.74					
3	3	0.73	0.57	0.54	2	32	0.77	0.31	0.43					
					3	1	0.70	0.38	0.81					
					3	2	0.33	0.43	0.36					
					3	3	0.51	0.48	0.32					
					3	4	0.50	0.54	0.43					
					3	5	0.53	0.47	0.49					
					3	6	0.32	0.41	0.40					
					3	7	0.22	0.34	0.41					

Table 6. 11: Item Statistics: Grade 6

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
1	1	0.83	0.33	0.12	1	1	0.49	0.36	0.13
1	2	0.82	0.33	0.11	1	2	0.76	0.48	0.20
1	3	0.62	0.35	0.26	1	3	0.70	0.40	0.17
1	4	0.55	0.49	0.34	1	4	0.83	0.31	1.24
1	5	0.56	0.45	0.44	1	5	0.58	0.29	0.26
1	6A	0.31	0.47	0.70	1	6	0.58	0.32	0.26
1	6B	0.90	0.35	0.69	1	7	0.57	0.23	0.33
1	7	0.58	0.34	0.28	1	8	0.85	0.40	0.33
1	8	0.83	0.41	0.26	1	9	0.59	0.21	0.29
1	9	0.77	0.22	0.53	1	10	0.60	0.34	0.40
1	10	0.71	0.08	0.24	1	11	0.60	0.46	0.30
1	11	0.65	0.38	0.29	1	12	0.71	0.41	0.26
1	12	0.74	0.37	0.31	1	13	0.75	0.20	0.21
2	1	0.76	0.41	0.18	1	14	0.61	0.39	0.20
2	2	0.83	0.33	0.22	1	15	0.64	0.45	0.35
2	3	0.93	0.43	0.65	1	16	0.64	0.44	0.37
2	4	0.92	0.35	1.11	1	17	0.88	0.40	0.18
2	5	0.88	0.37	1.98	1	18	0.85	0.43	0.33
2	6	0.69	0.36	0.23	1	19	0.80	0.39	0.27
2	7	0.79	0.36	0.25	1	20	0.72	0.40	0.29
2	8	0.87	0.30	0.24	1	21	0.68	0.39	0.28
2	9	0.64	0.37	0.40	1	22	0.79	0.33	0.45
2	10	0.42	0.32	0.58	1	23	0.91	0.38	0.29
2	11	0.93	0.42	0.52	2	1	0.78	0.23	0.17
2	12	0.94	0.32	1.05	2	2	0.77	0.31	0.31
2	13	0.70	0.43	0.46	2	3	0.75	0.24	2.13
2	14	0.92	0.43	0.50	2	4	0.77	0.40	0.28
2	15	0.86	0.45	0.61	2	5	0.84	0.38	0.60
2	16	0.68	0.41	0.90	2	6	0.73	0.41	1.09
2	17	0.68	0.26	1.17	2	7	0.58	0.35	1.62
2	18	0.81	0.34	1.44	2	8	0.83	0.40	2.15
2	19	0.64	0.34	1.46	2	9	0.95	0.27	0.13
2	20	0.53	0.32	1.64	2	10	0.85	0.32	0.27
2	21	0.70	0.44	0.40	2	11	0.58	0.31	4.30
2	22	0.63	0.33	0.73	2	12	0.68	0.41	0.21

Table 6. 11: Item Statistics: Grade 6 (Cont'd)

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
2	23	0.72	0.43	0.40	2	13	0.73	0.41	0.36
2	24	0.58	0.28	0.28	2	14	0.82	0.31	0.48
2	25	0.64	0.43	0.37	2	15	0.78	0.49	0.52
2	26	0.67	0.45	0.61	2	16	0.79	0.49	0.58
2	27	0.74	0.28	0.67	2	17	0.73	0.33	0.43
2	28	0.75	0.41	0.54	2	18	0.84	0.45	0.47
2	29	0.81	0.23	2.06	2	19	0.77	0.42	0.69
2	30	0.70	0.38	0.54	2	20	0.79	0.51	0.83
2	31	0.43	0.31	0.43	2	21	0.52	0.39	0.94
2	32	0.92	0.42	0.32	2	22	0.71	0.42	1.23
2	33	0.72	0.49	0.38	2	23	0.62	0.39	1.12
2	34	0.84	0.41	0.69	2	24	0.77	0.55	0.42
2	35	0.83	0.45	0.90	2	25	0.66	0.43	0.50
2	36	0.47	0.25	1.26	2	26	0.67	0.47	0.71
2	37	0.51	0.29	1.38	2	27	0.76	0.46	0.66
2	38	0.78	0.50	1.51	2	28	0.63	0.36	0.74
2	39	0.45	0.28	1.96	2	29	0.57	0.45	1.01
3	1	0.68	0.47	0.36	2	30	0.65	0.22	0.91
3	2	0.25	0.39	1.04	2	31	0.63	0.34	1.00
3	3	0.34	0.44	0.73	3	1	0.81	0.37	0.28
					3	2	0.45	0.31	0.28
					3	3	0.56	0.45	0.59
					3	4	0.49	0.47	0.66
					3	5	0.27	0.47	0.69
					3	6	0.59	0.47	0.48
					3	7	0.41	0.45	0.66

Table 6. 12: Item Statistics: Grade 7

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
1	1	0.49	0.35	0.22	1	1	0.67	0.32	0.18
1	2	0.88	0.47	0.12	1	2	0.48	0.31	0.61
1	3	0.49	0.45	0.46	1	3	0.50	0.45	0.34
1	4	0.32	0.33	0.70	1	4	0.71	0.38	0.16
1	5	0.49	0.35	0.70	1	5	0.58	0.43	0.26
1	6A	0.67	0.56	2.76	1	6	0.54	0.35	0.19
1	6B	0.70	0.36	2.71	1	7	0.49	0.37	0.42
1	7	0.31	0.27	0.33	1	8	0.73	0.24	0.41
1	8	0.65	0.23	0.63	1	9	0.78	0.43	0.17
1	9	0.65	0.26	0.37	1	10	0.81	0.36	0.25
1	10	0.85	0.39	1.24	1	11	0.44	0.41	0.32
1	11	0.63	0.35	2.63	1	12	0.47	0.21	0.30
1	12	0.86	0.34	0.32	1	13	0.63	0.56	0.38
1	13	0.86	0.30	0.99	1	14	0.48	0.14	0.58
1	14	0.91	0.37	0.71	1	15	0.62	0.27	0.31
1	15	0.60	0.26	0.49	1	16	0.60	0.51	0.41
1	16	0.74	0.36	0.42	1	17	0.44	0.36	0.31
2	1	0.72	0.55	0.45	1	18	0.58	0.56	0.39
3	1	0.95	0.33	0.28	1	19	0.68	0.41	0.26
3	2	0.63	0.32	0.32	1	20	0.36	0.39	0.38
3	3	0.88	0.41	0.47	1	21	0.80	0.21	0.33
3	4	0.78	0.48	0.71	1	22	0.78	0.42	0.45
3	5	0.86	0.18	0.28	1	23	0.51	0.32	0.31
3	6	0.89	0.42	0.31	2	1	0.74	0.30	0.23
3	7	0.50	0.27	0.54	2	2	0.74	0.39	0.21
3	8	0.89	0.32	0.35	2	3	0.53	0.34	0.87
3	9	0.92	0.32	0.59	2	4	0.63	0.50	1.01
3	10	0.93	0.36	0.79	2	5	0.58	0.36	0.81
3	11	0.84	0.40	0.47	2	6	0.59	0.33	1.37
3	12	0.77	0.42	1.25	2	7	0.52	0.43	2.22
3	13	0.80	0.39	0.60	2	8	0.64	0.30	4.50
3	14	0.71	0.42	0.80	2	9	0.62	0.40	6.18
3	15	0.83	0.45	1.08	2	10	0.99	0.12	0.14
3	16	0.65	0.29	3.46	2	11	0.87	0.44	0.23
3	17	0.75	0.49	1.34	2	12	0.91	0.32	1.12

Table 6. 12: Item Statistics: Grade 7 (Cont'd)

Communication Arts					Mathematics				
Session	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**
3	18	0.75	0.49	1.56	2	13	0.91	0.34	2.72
3	19	0.76	0.32	1.75	2	14	0.85	0.34	0.23
3	20	0.74	0.34	1.92	2	15	0.70	0.39	0.30
3	21	0.81	0.47	0.26	2	16	0.90	0.40	0.50
3	22	0.60	0.40	0.39	2	17	0.92	0.32	0.22
3	23	0.71	0.43	0.38	2	18	0.62	0.40	0.33
3	24	0.43	0.31	1.02	2	19	0.83	0.39	1.61
3	25	0.75	0.41	0.39	2	20	0.66	0.43	0.68
3	26	0.77	0.39	0.40	2	21	0.49	0.40	0.47
3	27	0.85	0.27	0.60	2	22	0.53	0.30	1.62
3	28	0.54	0.35	0.85	2	23	0.79	0.47	0.38
3	29	0.73	0.47	0.50	2	24	0.60	0.47	1.04
3	30	0.45	0.24	1.06	2	25	0.78	0.37	0.28
3	31	0.57	0.32	0.50	2	26	0.64	0.47	0.76
3	32	0.59	0.35	0.60	2	27	0.66	0.38	0.37
3	33	0.81	0.47	0.98	2	28	0.76	0.50	0.41
3	34	0.80	0.40	1.09	2	29	0.53	0.45	0.68
3	35	0.70	0.46	1.54	2	30	0.68	0.42	0.74
3	36	0.68	0.44	1.12	2	31	0.64	0.59	0.62
3	37	0.82	0.53	1.42	2	32	0.51	0.46	0.97
3	38	0.64	0.38	1.33	3	1	0.24	0.50	0.42
3	39	0.55	0.40	1.47	3	2	0.58	0.39	1.35
4	1	0.68	0.41	0.78	3	3	0.41	0.55	0.35
4	2	0.53	0.40	1.19	3	4	0.49	0.54	1.36
4	3A	0.74	0.54	1.65	3	5	0.25	0.45	0.89
4	3B	0.77	0.36	1.65	3	6	0.62	0.59	0.92
					3	7	0.24	0.41	0.54

Table 6. 13: Item Statistics: Grade 8

Communication Arts					Mathematics					Science				
Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate
1	1	0.87	0.33	0.17	1	1	0.41	0.52	0.17	1	1	0.60	0.36	2.12
1	2	0.67	0.37	0.23	1	2	0.88	0.42	0.28	1	2	0.51	0.37	2.23
1	3	0.40	0.40	1.38	1	3	0.64	0.18	0.28	1	3	0.59	0.58	0.63
1	4	0.44	0.50	1.21	1	4	0.52	0.41	0.20	1	4	0.39	0.48	0.63
1	5	0.42	0.39	1.67	1	5	0.63	0.46	0.39	1	5	0.42	0.45	1.94
1	6A	0.67	0.55	3.93	1	6	0.48	0.40	0.37	1	6	0.26	0.19	0.71
1	6B	0.68	0.41	3.92	1	7	0.73	0.33	0.47	1	7	0.40	0.37	1.47
1	7	0.50	0.26	0.35	1	8	0.67	0.43	0.35	1	8	0.46	0.50	1.59
1	8	0.42	0.23	0.67	1	9	0.34	0.43	0.31	1	9	0.19	0.40	1.06
1	9	0.60	0.19	0.38	1	10	0.74	0.05	0.31	1	10	0.30	0.52	10.44
1	10	0.56	0.08	1.58	1	11	0.49	0.52	0.31	1	11	0.23	0.52	2.83
1	11	0.79	0.37	1.59	1	12	0.49	0.54	0.57	1	12	0.18	0.48	5.62
1	12	0.82	0.34	0.34	1	13	0.60	0.26	0.27	2	1	0.92	0.19	0.32
1	13	0.69	0.36	0.52	1	14	0.30	0.49	0.29	2	2	0.91	0.32	0.33
1	14	0.50	0.29	0.62	1	15	0.43	0.27	0.53	2	3	0.89	0.39	0.34
1	15	0.48	0.32	0.87	1	16	0.38	0.47	0.39	2	4	0.89	0.27	0.36
1	16	0.85	0.36	0.52	1	17	0.67	0.49	0.62	2	5	0.97	0.24	0.36
2	1	0.96	0.30	0.19	1	18	0.53	0.43	0.85	2	6	0.82	0.37	0.48
2	2	0.94	0.30	0.21	1	19	0.30	0.37	0.76	2	7	0.72	0.32	0.48
2	3	0.89	0.39	0.31	1	20	0.26	0.43	0.42	2	8	0.73	0.34	0.46
2	4	0.93	0.32	0.57	1	21	0.46	0.40	0.42	2	9	0.85	0.37	0.57
2	5	0.83	0.46	0.22	1	22	0.51	0.49	0.45	2	10	0.87	0.36	0.63
2	6	0.90	0.25	0.37	1	23	0.52	0.30	0.39	2	11	0.74	0.34	0.59
2	7	0.54	0.23	0.50	1	31	0.45	0.66	0.72	2	12	0.76	0.35	0.59
2	8	0.97	0.35	0.22	2	1	0.80	0.33	0.22	2	13	0.67	0.46	0.62
2	9	0.83	0.52	0.25	2	2	0.80	0.34	0.35	2	14	0.77	0.38	0.66
2	10	0.81	0.43	0.33	2	3	0.78	0.40	0.65	2	15	0.80	0.47	0.70
2	11	0.66	0.21	0.41	2	4	0.95	0.24	0.26	2	16	0.65	0.51	0.61
2	12	0.61	0.35	0.60	2	5	0.40	0.29	0.45	2	17	0.72	0.23	0.53
2	13	0.79	0.26	0.42	2	6	0.42	0.36	0.41	2	18	0.56	0.36	5.60
2	14	0.66	0.36	0.52	2	7	0.76	0.35	0.47	2	20	0.73	0.45	0.78
2	15	0.72	0.36	0.84	2	8	0.77	0.29	0.44	2	21	0.59	0.26	0.69
2	16	0.80	0.34	0.46	2	9	0.37	0.35	0.30	2	22	0.46	0.41	1.23
2	17	0.84	0.40	0.63	2	10	0.84	0.39	0.30	2	23	0.92	0.25	0.74
2	18	0.60	0.35	0.74	2	11	0.70	0.29	0.37	2	25	0.51	0.25	0.60

Table 6. 13: Item Statistics: Grade 8 (Cont'd)

Communication Arts					Mathematics					Science				
Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate	Session	Item	*p*-value	R_{it}	Omit Rate
2	19	0.33	0.24	0.81	2	12	0.85	0.37	0.88	2	26	0.76	0.49	0.99
2	20	0.36	0.24	1.10	2	13	0.76	0.38	0.59	2	27	0.52	0.46	1.95
2	21	0.76	0.38	0.28	2	14	0.69	0.40	0.60	2	28	0.57	0.53	1.89
2	22	0.94	0.39	0.32	2	15	0.91	0.41	0.37	2	29	0.21	0.55	1.97
2	23	0.80	0.47	0.41	2	16	0.80	0.28	0.29	2	30	0.39	0.60	2.21
2	24	0.77	0.35	0.42	2	17	0.87	0.27	0.31	2	31	0.38	0.37	0.90
2	25	0.75	0.47	0.27	2	18	0.61	0.48	0.44	2	32	0.27	0.46	4.60
2	26	0.82	0.42	0.37	2	19	0.66	0.54	0.42	2	33	0.13	0.37	4.62
2	27	0.53	0.37	0.51	2	20	0.57	0.29	0.53	2	34	0.26	0.46	2.46
2	28	0.88	0.40	0.42	2	21	0.88	0.37	0.39	2	35	0.27	0.46	2.28
2	29	0.67	0.39	0.53	2	22	0.81	0.32	0.51	2	36	0.15	0.33	6.30
2	30	0.88	0.47	0.47	2	23	0.73	0.38	0.41	2	37	0.23	0.46	6.51
2	31	0.44	0.30	1.12	2	24	0.80	0.29	0.44	3	1	0.41	0.45	2.43
2	32	0.47	0.24	2.03	2	25	0.36	0.31	0.57	3	2	0.54	0.53	3.04
2	33	0.89	0.45	2.41	2	26	0.68	0.53	0.59	3	3	0.37	0.39	1.59
2	34	0.62	0.42	1.09	2	27	0.45	0.34	1.13	3	4	0.84	0.33	4.85
2	35	0.82	0.44	0.56	2	28	0.62	0.53	1.69	3	5	0.42	0.54	1.89
2	36	0.65	0.44	0.94	2	29	0.38	0.32	1.71	3	6	0.60	0.50	10.40
2	37	0.39	0.35	0.67	2	30	0.39	0.32	0.93	3	7	0.26	0.35	1.79
2	38	0.42	0.24	0.90	2	31	0.36	0.44	1.02	3	8	0.55	0.48	1.29
2	39	0.26	0.23	0.90	3	1	0.35	0.46	0.90	3	9	0.58	0.52	2.43
3	1	0.57	0.47	0.71	3	2	0.25	0.41	4.80	3	10	0.55	0.37	6.89
3	2	0.35	0.51	5.01	3	3	0.45	0.55	3.16	3	11	0.36	0.54	3.60
3	3A	0.78	0.49	1.67	3	4	0.52	0.41	2.17	3	12	0.38	0.43	4.54
3	3B	0.77	0.39	1.66	3	5	0.40	0.55	1.51					
3	3C	0.92	0.37	1.67	3	6	0.16	0.53	0.89					
					3	7	0.50	0.47	1.53					
					3	8	0.33	0.57	1.85					
					3	9	0.29	0.47	1.77					

Table 6. 14: Item Fit Statistics for Misfitting Items

Content	Grade	Session	Item	Chi-Square	DF	Total N	Z	Ob-served	Pre-dicted	Obs-Pred
CA	3	2	1	1644.76	35	59093	192.40	0.68	0.68	0.00
CA	4	2	17	613.02	7	59451	161.97	0.93	0.92	0.00
CA	8	2	11	654.68	7	64081	173.10	0.66	0.66	0.00
CA	8	2	38	1075.89	7	63767	285.67	0.42	0.41	0.01
MA	3	3	7	1079.40	17	59123	182.20	0.72	0.72	0.00
MA	6	2	30	788.81	7	65136	208.95	0.65	0.66	0.00
MA	6	3	3	1762.75	17	65343	299.39	0.56	0.56	0.00
MA	8	2	16	882.24	7	64236	233.92	0.80	0.80	0.00
MA	8	1	20	3138.47	7	64155	836.92	0.26	0.31	-0.05
MA	8	3	8	2019.77	17	63233	343.47	0.33	0.33	0.00
MA	8	3	4	4304.54	17	63023	735.31	0.52	0.52	0.00
SC	8	3	5	1718.08	35	63137	201.17	0.42	0.42	0.00

Table 6. 15: LOSS and HOSS Values by Grade and Content Area

Grade	Communication Arts		Mathematics		Science	
	LOSS	HOSS	LOSS	HOSS	LOSS	HOSS
3	455	790	450	780		
4	470	820	465	805		
5	485	840	480	830	470	855
6	505	855	495	845		
7	515	865	510	860		
8	530	875	525	885	540	895

Figure 6. 1: Item characteristic curve for Grade 3 Communication Arts, Session 2 Item 1

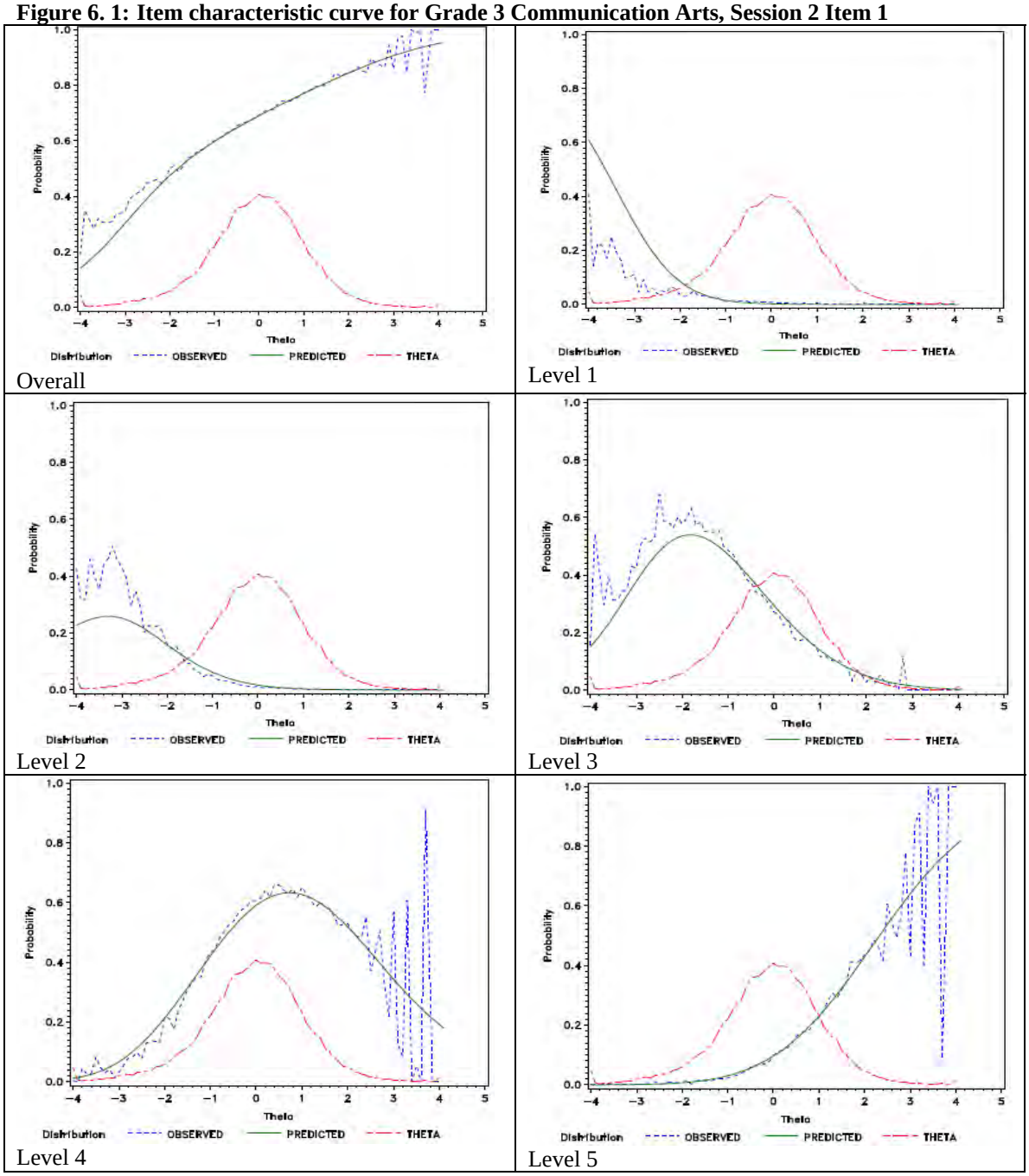


Figure 6. 2: Item characteristic curve for Grade 4 Communication Arts, Session 2 Item 17

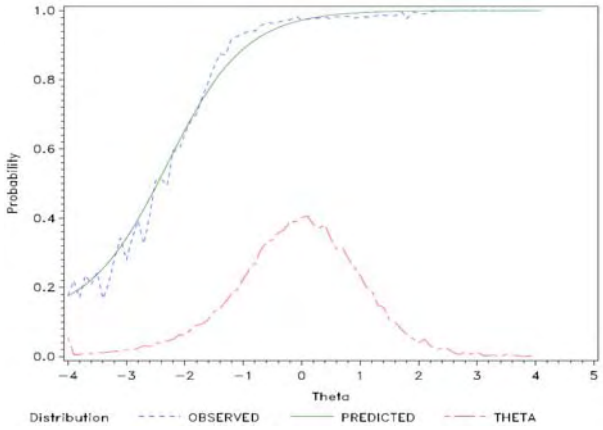


Figure 6. 3: Item characteristic curve for Grade 8 Communication Arts, Session 2 Item 11

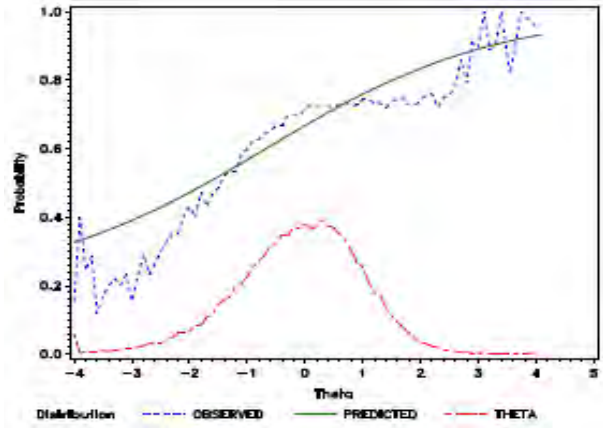


Figure 6. 4: Item characteristic curve for Grade 8 Communication Arts, Session 2 Item 38

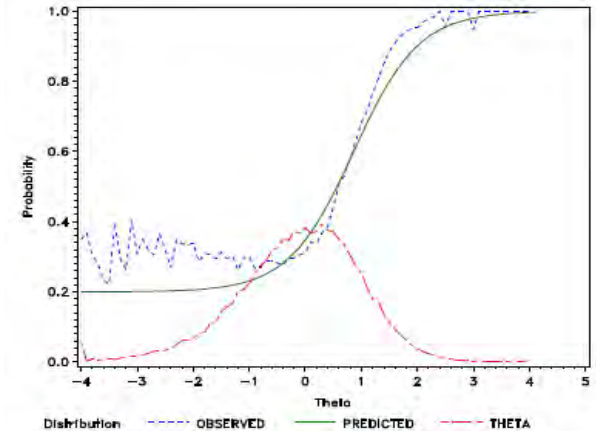


Figure 6. 5: Item characteristic curve for Grade 3 Mathematics, Session 3 Item 7

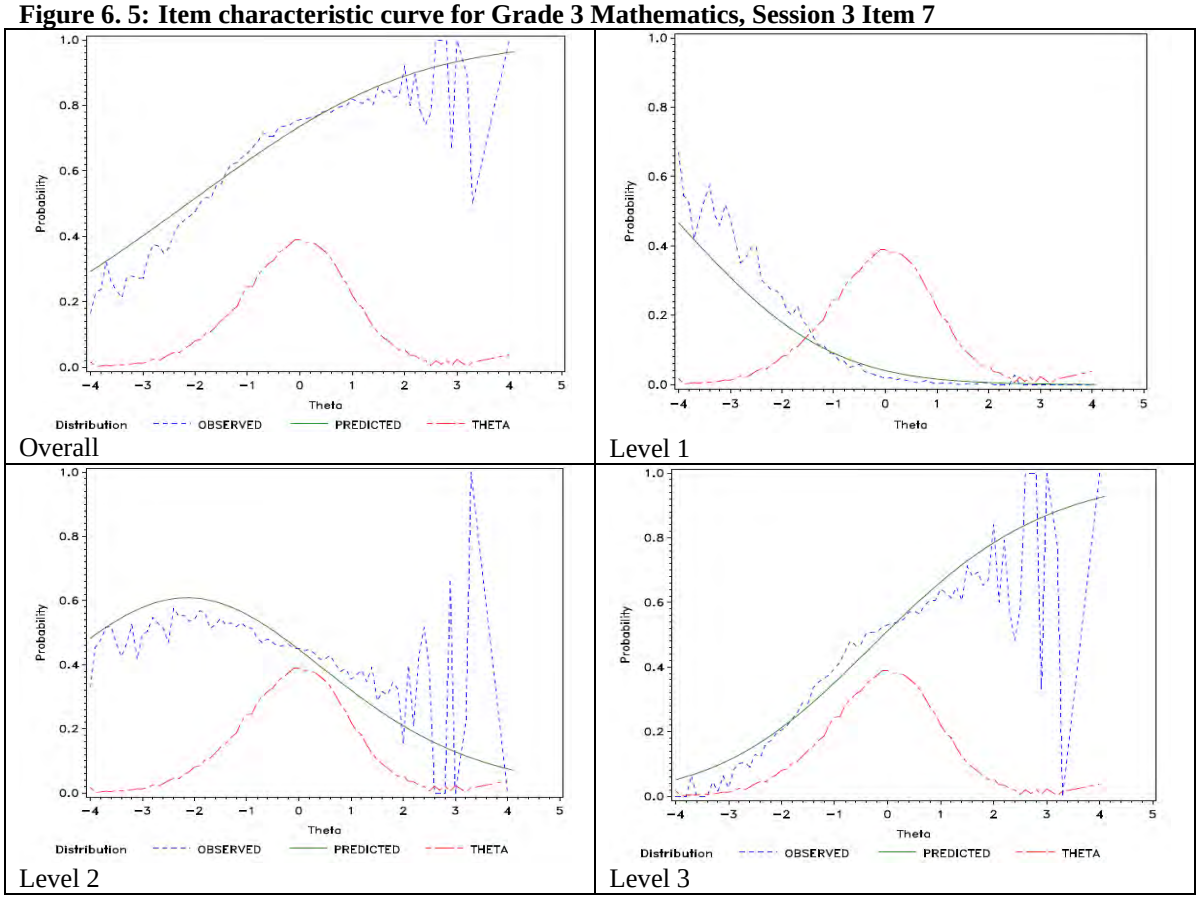


Figure 6. 6: Item characteristic curve for Grade 6 Mathematics, Session 2 Item 30

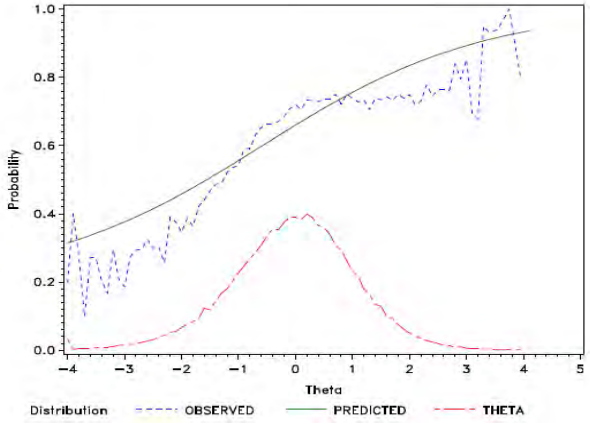


Figure 6. 7: Item characteristic curve for Grade 6 Mathematic, Session 3 Item 3

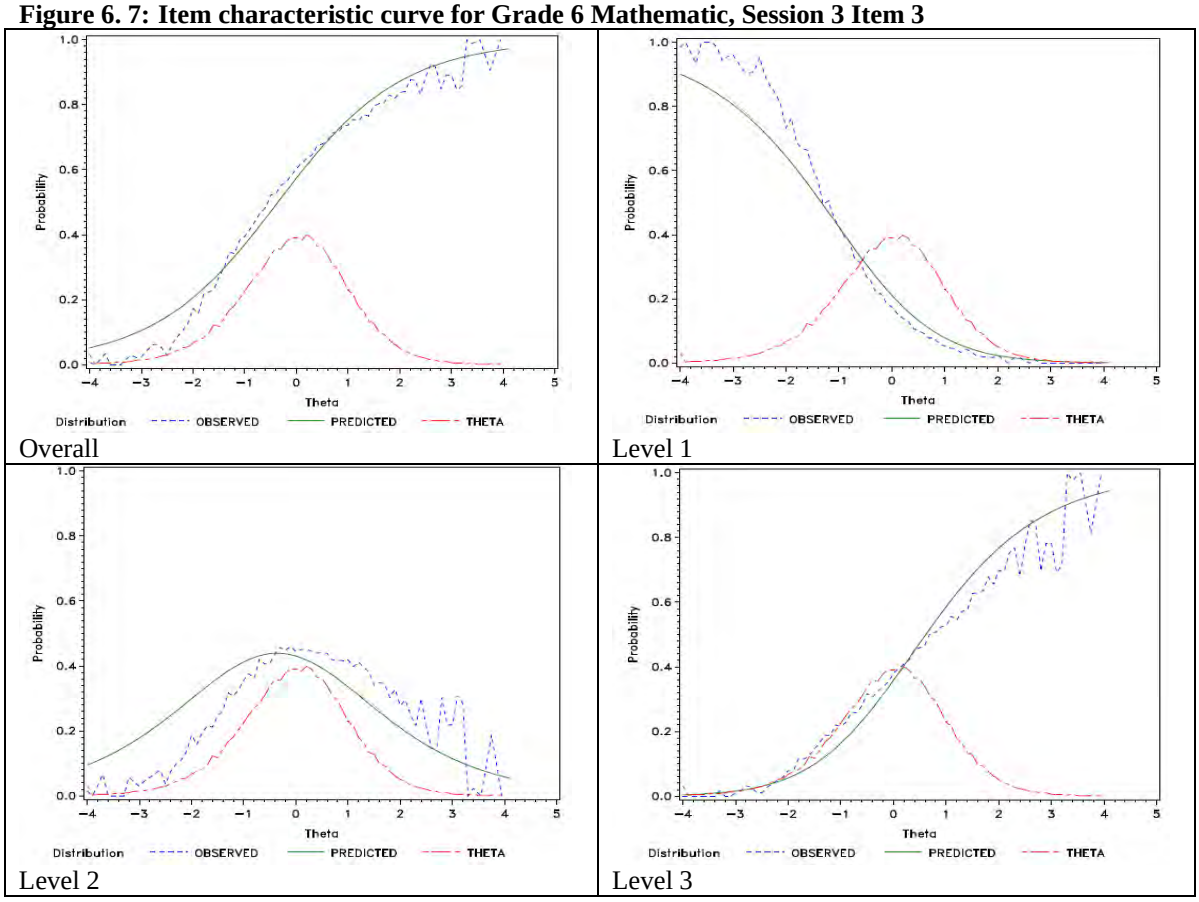


Figure 6. 8: Item characteristic curve for Grade 8 Mathematics, Session 2 Item 16

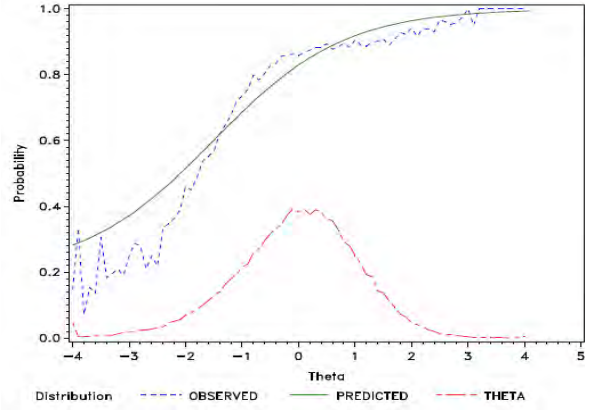


Figure 6. 9: Item characteristic curve for Grade 8 Mathematics, Session 1 Item 20

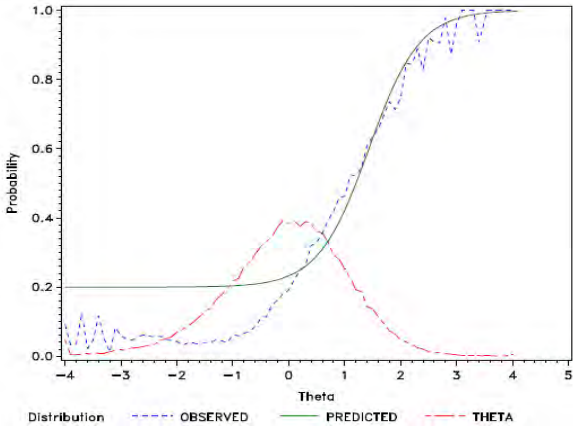


Figure 6. 10: Item characteristic curve for Grade 8 Mathematics, Session 3 Item 8

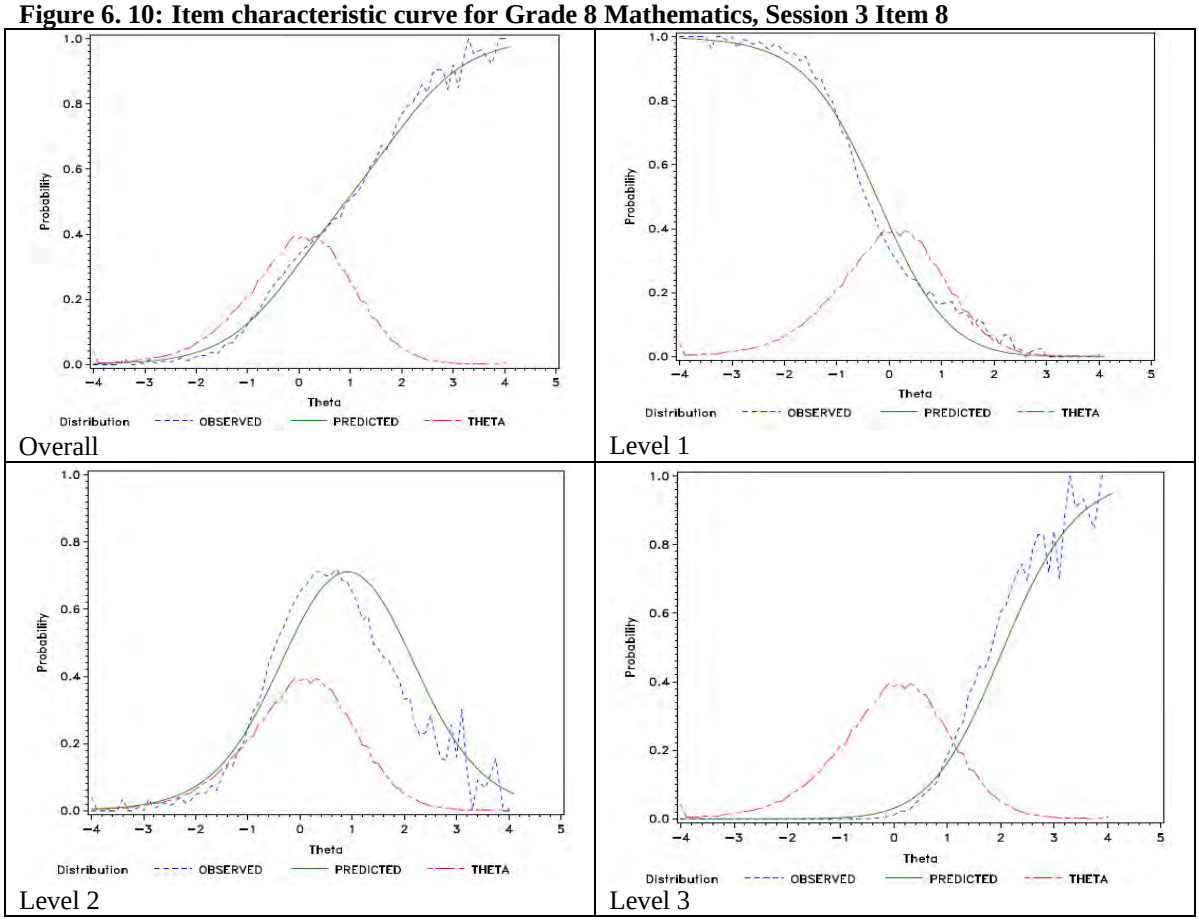


Figure 6. 11: Item characteristic curve for Grade 8 Mathematics, Session 3 Item 4

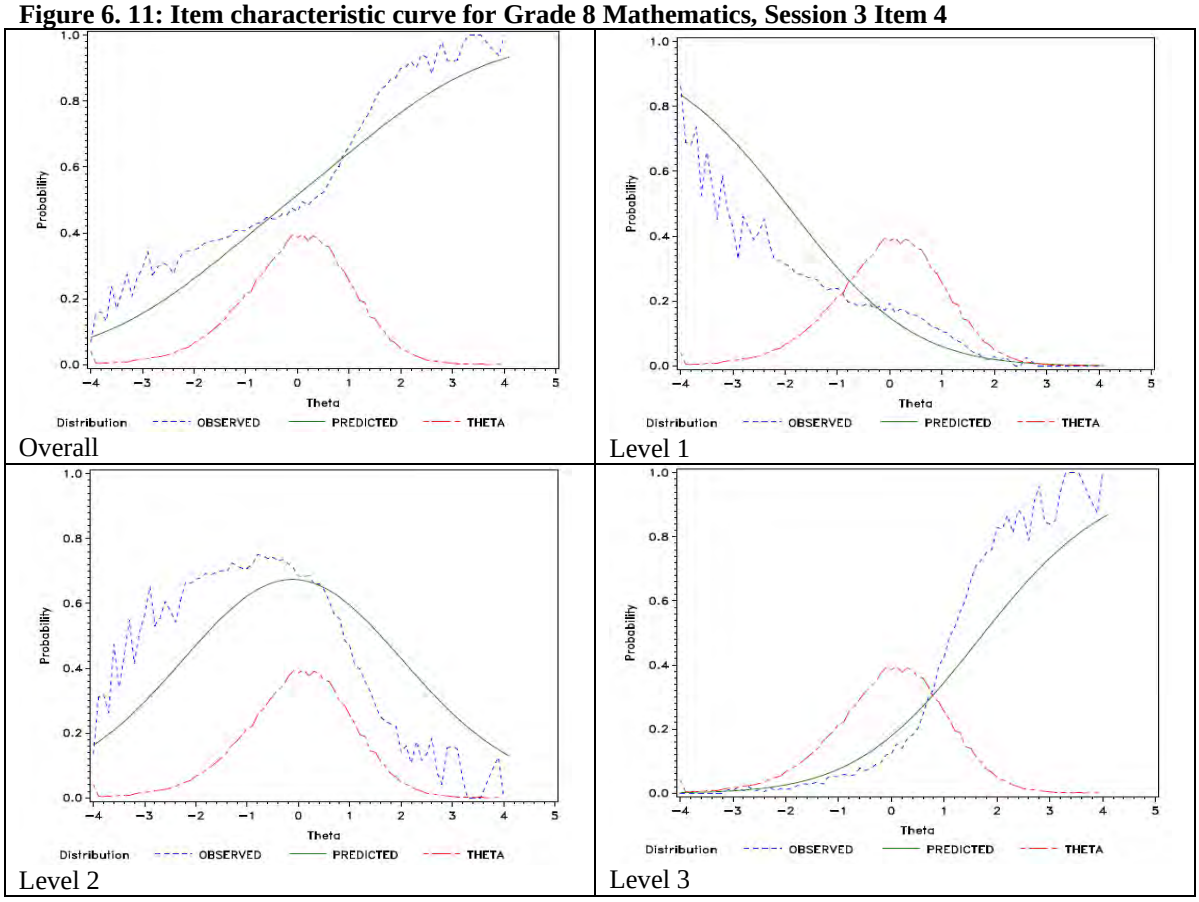


Figure 6. 12: Item characteristic curve for Grade 8 Science, Session 3 Item 5

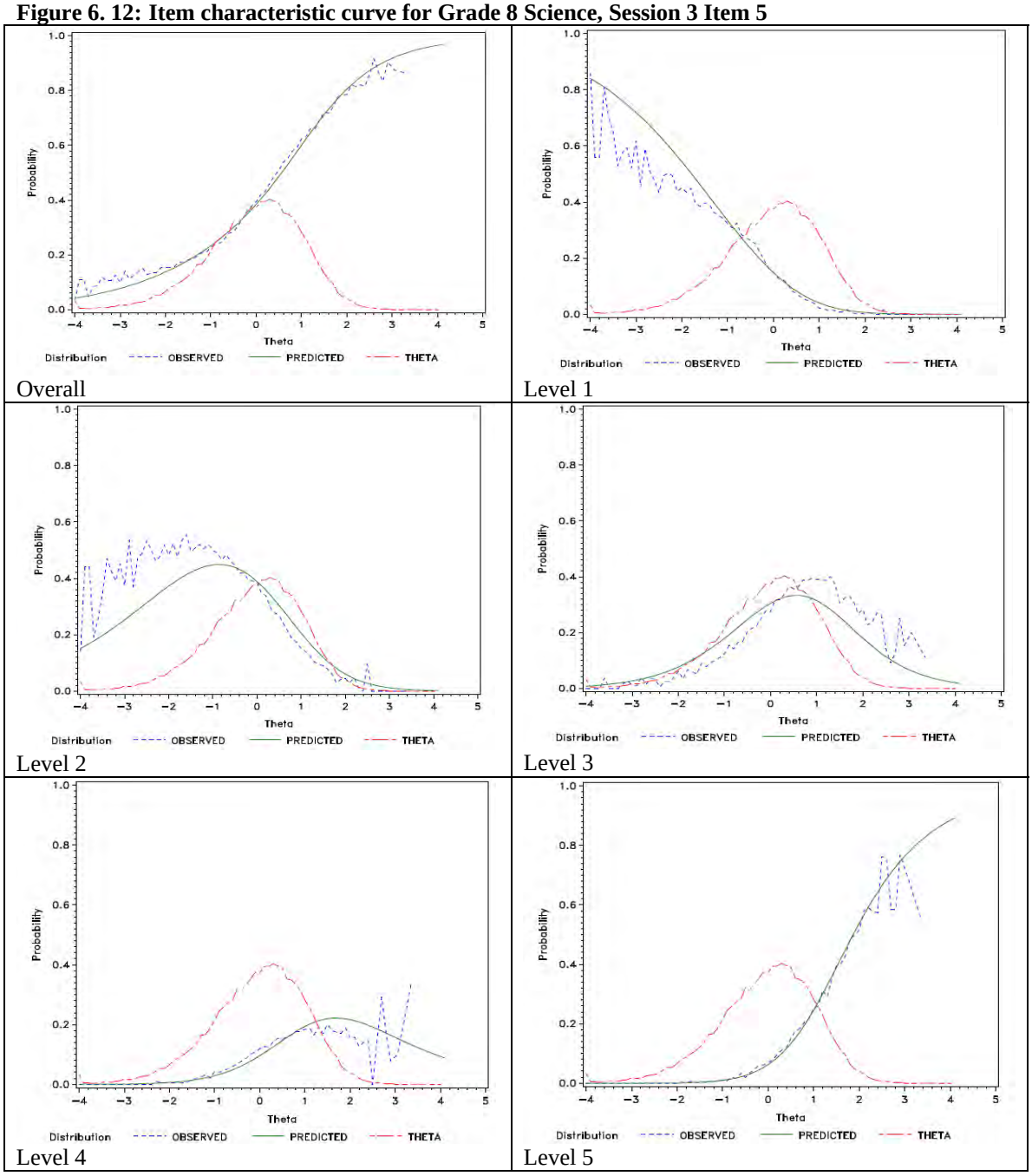


Figure 6. 13: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Communication Arts MAP

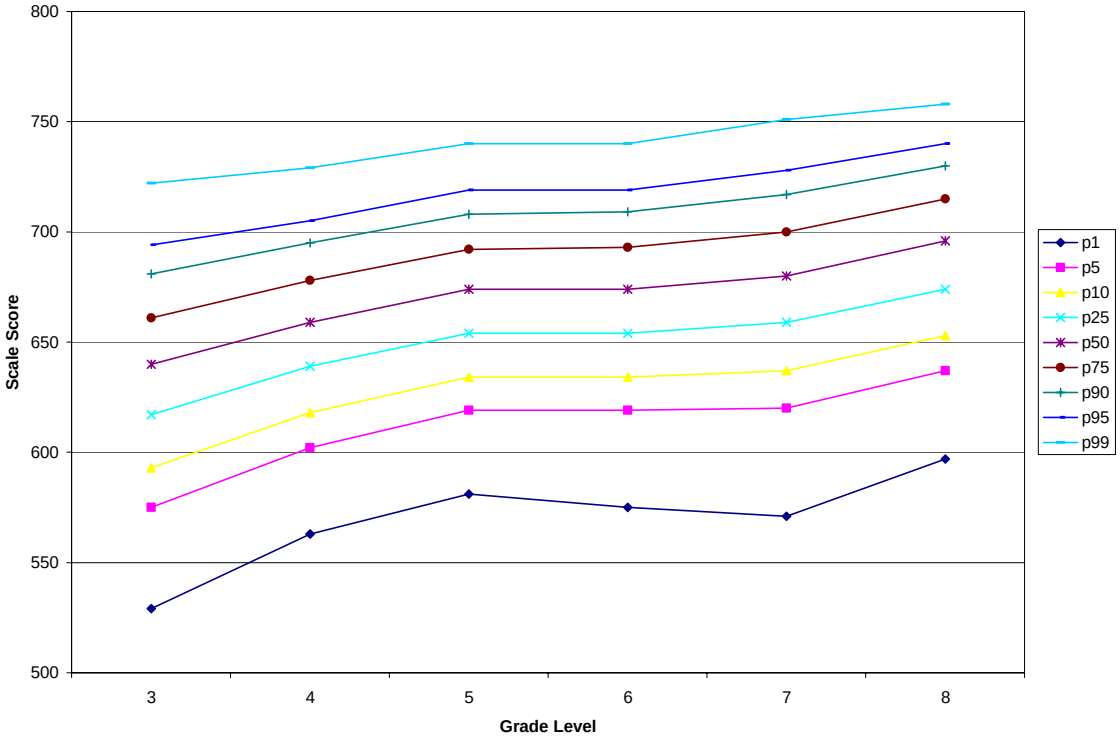


Figure 6. 14: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Mathematics MAP

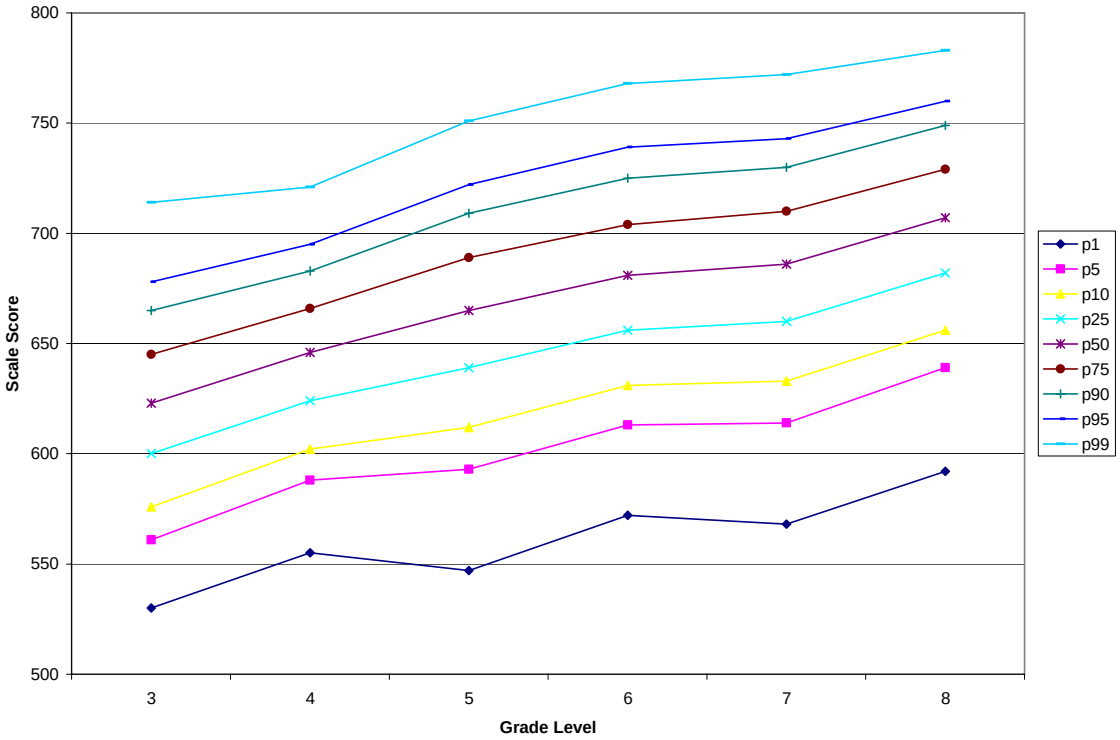


Figure 6. 15: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Science MAP

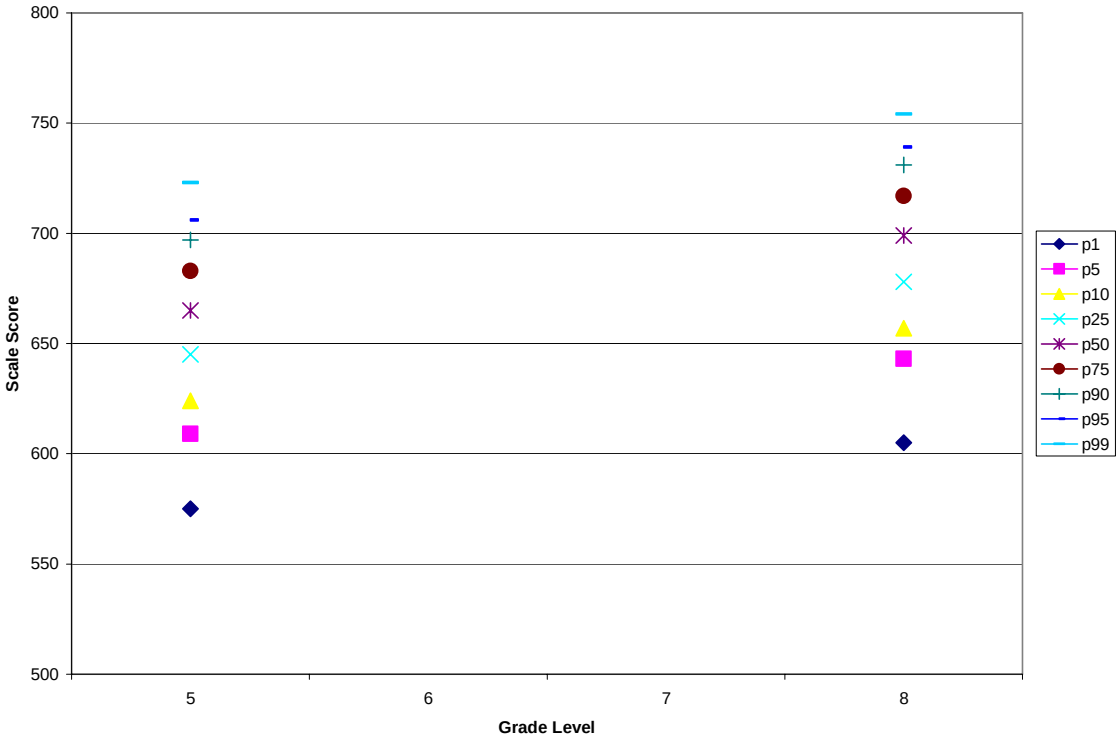


Figure 6. 16: Communication Arts Test Characteristic Curves by grade, 2009

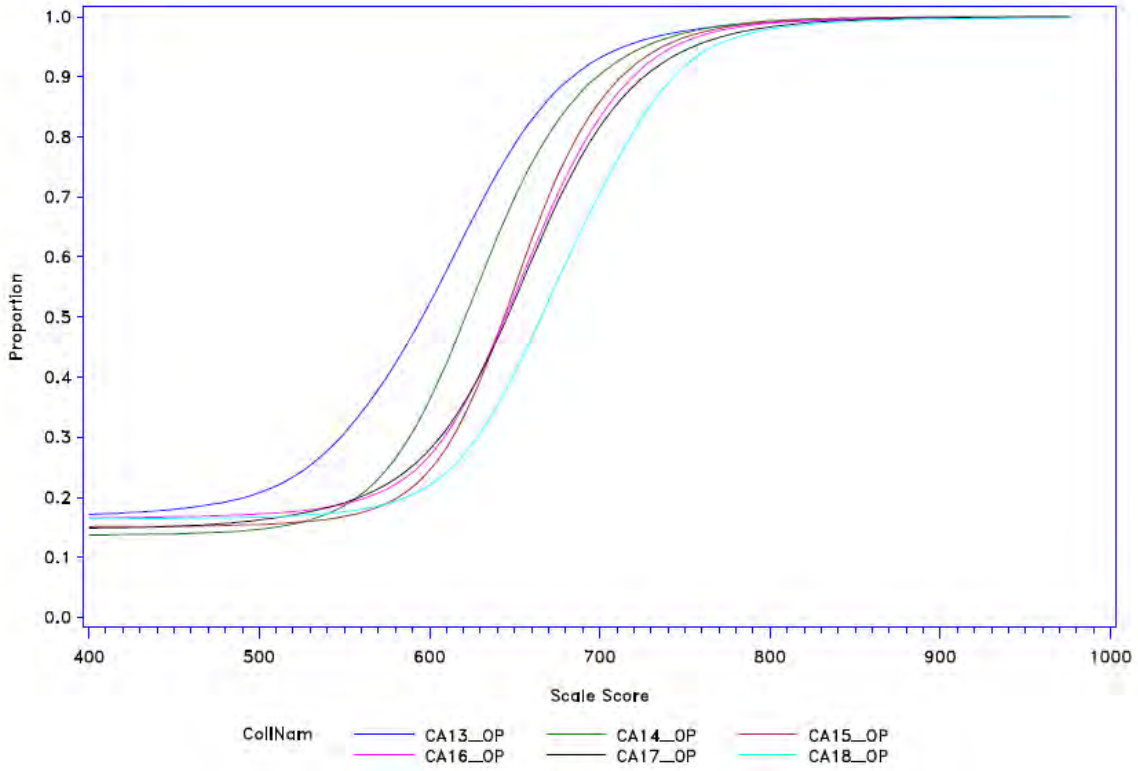


Figure 6. 17: Mathematics Test Characteristic Curves by grade, 2009

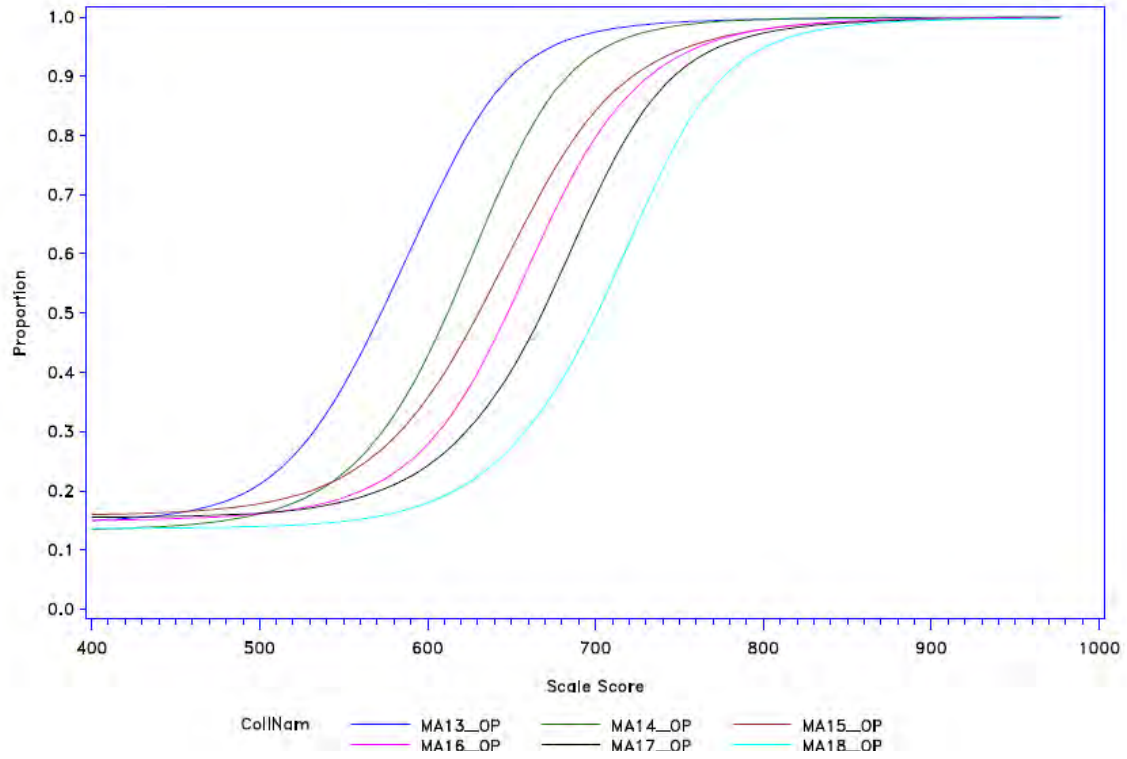
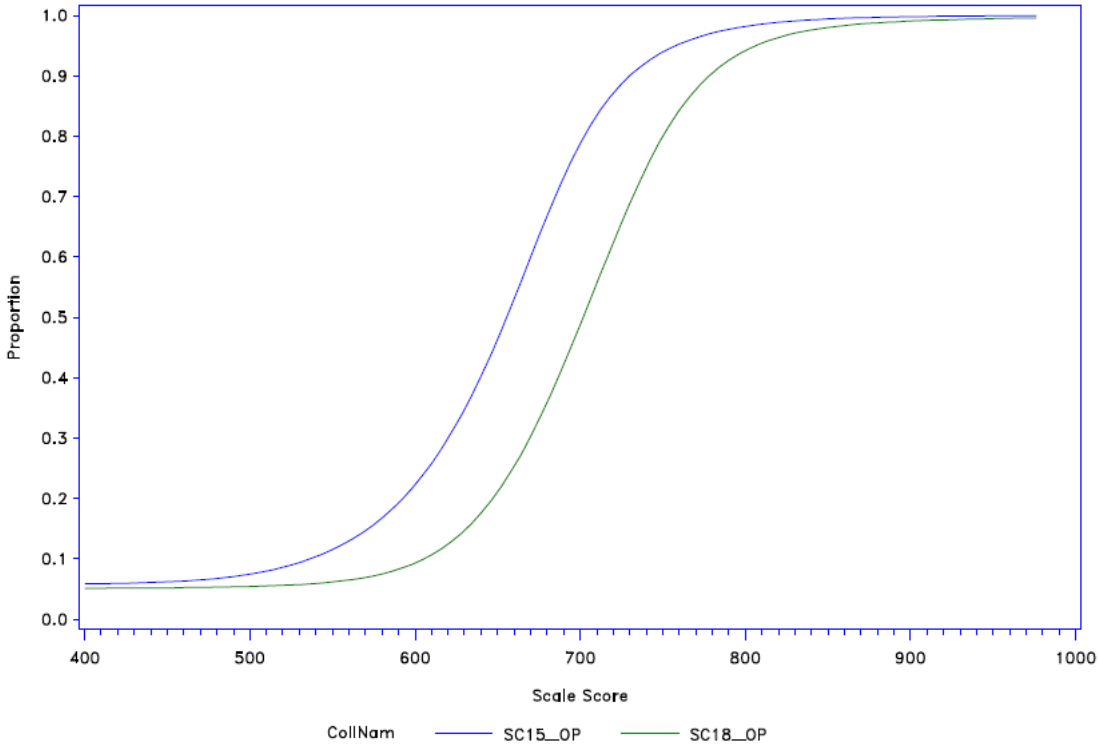


Figure 6. 18: Science Test Characteristic Curves by grade, 2009



CHAPTER 7: TEST RESULTS

This chapter of the Technical Report contains information on the results of the Spring 2009 administration of the MAP. The scale score results are presented here. Performance level information is also provided. Presenting the results by performance level translates the quantitative scale provided through scale scores into a qualitative description of student performance: *Below Basic, Basic, Proficient,* and *Advanced.*

While the scale score provides an essential quantitative reference to student performance, the performance level information speaks directly to requirements of the NCLB Act, as well as plainly outlines the scores to parents, students, and educators. When combined, scale scores, performance levels, and Lexile scores provide a comprehensive set of tools to assess Missouri student performance by content and grade level.

This chapter also provides description of the score reports, data structure, and interpretive guide. The AERA, APA, & NCME (1999) Standards addressed in Chapter 7 are 4.1, 5.10, 6.2, and 13.19. Each Standard will be presented in the pertinent section of this chapter.

Results presented below are based on census data. The results presented here may differ slightly from the official state summary report of all student populations due to ongoing resolution of test materials and student information. The results in the following tables are presented as evidence of reliability and validity of the scores from the MAP assessments, and should not be used for state accountability purposes.

7.1 Student Participation

The following are subgroups reported during the administration of MAP (other demographic information is collected separately and merged into MAP data after CTB sends DESE the General Research File):

- Gender: Female and Male
- Race and Ethnicity: White, Black, Hispanic, Asian/Pacific Islander, and Native American/Alaskan
- Accommodations: Students receiving testing accommodations

For the purposes of this report, participation rate is defined as the percentage of students who received a valid scale score given the total number of students who received a test book. These participation rates are summarized in Tables 7.1 through 7.9. The tables show both the percentage of students classified as reportable and the number of students classified as accountable. Reportable students include all students with a valid scale score. Accountable students include all students for whom a test book was submitted. These include students who should have received a MAP scale score, but did not take the test and could not be assigned a scale score.

7.2 Current Administration Data

The Communication Arts and Mathematics MAP assessments were administered to students in Grades 3 through 8. The Science MAP assessments were administered to students in Grades 5 and 8.

Tables 7.10 through 7.12 provide a summary of the scale scores based on the state population for the 2009 administration of the MAP. In compliance with AERA, APA, & NCME (1999) Standard 13.19, these tables present the number of students, mean and standard deviation of scale scores, and scale scores at specific percentile points. Standard 13.19 states:

> In educational settings, when average or summary scores for groups of students are reported, they should be supplemented with additional information about the sample size and shape or dispersion of score distributions.

7.3 Cross-year, Cross-sectional Comparisons

It is often desirable to examine the scores of students across time. The data in this section compare student performance on the MAP using census data from 2006 through 2009. It should be noted that beginning in 2008, invalidated students were assigned to the LOSS and to the *Below Basic* achievement level. Prior to 2008, invalidated students did not receive a scale score.

Table 7.13 shows the state-level means for all grades from 2006 through 2009 for Communication Arts and Mathematics and from 2008 and 2009 for Science. The Science MAP was administered for the first time in 2008. As shown in Table 7.13, the mean scale scores in all grades and content areas were stable across years.

Table 7.14 shows the percentage of students in each achievement level in 2006 through 2009 on the Communication Arts test. The percentages at or above *Proficient* tended to be stable from 2008 to 2009, with small increases in all grades except Grade 3 which did not change.

Table 7.15 shows the percentage of students in each achievement level in 2006 through 2009 on the Mathematics test. As compared to 2008, increases in the percentage of students at or above *Proficient* were observed in all grades in 2009, except Grade 6 which decreased slightly.

Table 7.16 shows the percentage of students in each achievement level in 2008 and 2009 on the Science test. In Grades 5 and 8, the percentage of students at or above *Proficient* increased from 2008 to 2009.

7.4 Reports

Score reports are the primary means of communicating test scores to relevant district personnel (i.e., testing coordinators or superintendents), teachers, and parents. AERA, APA, and NCME (1999) standard 5.10 states:

> When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate interpretations. The interpretations should describe in simple language what the test covers, what scores mean, the precision of the scores, common misinterpretations of test scores, and how scores will be used.

Standard 4.1 is related in that it says:

> Test documents should provide test users with clear explanations of the meaning and intended interpretations of derived score scales, as well as their limitations.

Interpretations related to the test scores are disseminated in two ways: (1) the individual score report, and (2) the *Guide to Interpreting Results* (CTB/McGraw-Hill, 2009).

In addition to providing interpretation, it is important that the information is understandable by the target audience. Standard 6.2 of the AERA, APA, & NCME (1999) Standards states:

> Test documents should be complete, accurate, and clearly written so that the intended reader can readily understand the content.

The staffs at DESE and CTB strive to create documents that will be accessible to parents, teachers, and laypeople alike.

The individual student report is the primary means for sharing student test results with parents. As such, it should be a stand-alone document from which parents can glean relevant information so they understand their child's test score. In 2008, the individual MAP student reports were redesigned so that they were more parent-friendly. These changes include improved interpretations of the MAP scale score, *TerraNova* scale score, and Lexile score. In addition, the state mean score is now provided, as are activities that parents may engage in to help their children improve their skills within the content area in accordance with the Missouri Curriculum Framework. The new score reports also simplify the way in which the scale score and performance level are presented and interpreted. Starting in 2008, parents no longer receive scores for content/knowledge standards or for process/performance standards.

The *Guide to Interpreting Results* is intended for use by school and district personnel so that they can interpret their score reports. It provides a context for the score reports in that it outlines the history and purpose of MAP. It also overviews the Missouri Show-Me Standards and GLE Strands. It provides greater detail on the types of scores reported on the individual student report, and it provides all of the abbreviated achievement level

descriptors (ALDs), as well as the web location of the detailed ALDs. Finally, it outlines each piece of the individual student report and overviews the student label. The *Guide to Interpreting Results* is located on DESE's website at:

http://dese.mo.gov/divimprove/assess/map/

7.4.1. Description of Each Type of Report

In this section, descriptions for the following reports are provided: Individual Student Report, Student Score Label, online Crystal Reports, District Performance Summary Reports, and School Performance Summary Reports. Table 7.17 shows each report type and for whom the report is intended.

Individual Student Report

One copy of the Individual Student Report (ISR) is provided to schools to be sent home to the parents. On the left side of the page, results for a given content area are shown, including the student's MAP scale score, the state mean MAP scale score, the National Percentile score from the *TerraNova* section of MAP, and a brief definition of the National Percentile. On the Communication Arts ISR, the student's Lexile measure is also reported, along with a brief explanation of the Lexile measure and a website where more information may be obtained.

In the middle of the page, the student's scale score is shown again along with the achievement level associated with that scale score. This is followed by a brief explanation of what the achievement level means. When a student does not receive a scale score, then he or she will receive either "Level Not Determined" (LND) or "Invalidated" in place of the MAP scale score. No achievement level is assigned for the LND students. Invalidated students are assigned to the LOSS and to the *Below Basic* achievement level. A brief explanation accompanies the meaning of LND or invalidated.

On the right side of the page are recommended activities based on the child's achievement level. These are generic activities that are targeted to all students within an achievement level, not specific activities targeted at the individual student.
A sample report is provided in Appendix C, Figure C.1.

Student Score Label

The Student Score Label is designed so that each student's test results can be placed in the student's permanent record. A label is provided for every student who participated in the spring 2009 administration of the MAP. Each label has a self-adhesive backing so that it can be peeled from the sheet and placed in the student's cumulative school record. The label presents a snapshot of the student's results on the MAP. Separate labels are generated for each grade and content area; thus, a student will have multiple labels for each of the content areas administered within a grade. The label lists the student's scale score and National Percentile for each content area. For the Communication Arts MAP, the label also lists the student's Lexile measure.

CTB/McGraw-Hill provided multiple labels per student submitted for scoring. The labels are provided in print only. A sample report is provided in Appendix C, Figure C.2.

Online Crystal Reports

Schools and districts are able to access summary level reports through the online Crystal Reports tool. This tool allows district and school administrators to create on-the-fly reports containing information relevant to their data needs. There are several reporting options available through the Crystal Reports tool, including administrative reports, AYP reports, achievement level reports, content standard reports, and item analysis reports. Table 7.18 lists each of the major report headings and the sub-reports found under each reporting type.

For each sub-report, a user selects various filters such as year, grade/content area, and level of reporting (state, district, school) in order to create the desired report. For the Content Standard Reports, the user may also disaggregate results by various subgroups (e.g. race, disability).

A detailed discussion of all reports is beyond the scope of this document. Only those reports that are first-level analyses of MAP data will be discussed. The Achievement Level-5 reports will not be discussed as these are summaries of the pre-NCLB testing program. The AYP reports also will not be discussed nor will some of the Administrative Reports, including the High School Career Education Student Summary, Level Not Determined, and Map Alternate reports. Examples of all reports discussed are provided in Appendix C.

The Crystal Reports tool is accessed through DESE's website. Each school and/or district is assigned a user name and password so that it can access the site.

Administrative Reports

These reports provide student-level test data. Based on only the MAP test results, four reports are generated: MAP Scale Score Summary, MAP Student Demographic, Student Achievement Level, and Student Report.

MAP Scale Score Summary: This report lists each student in the school or district along with his/her MOSIS ID, testing year, content area, grade level, MAP scale score, achievement level, and *TerraNova* National Percentile. An example is included in Appendix C, Figure C.3.

MAP Student Demographic: This report lists each student in the school or district along with their date of birth (DOB), content area, CTB number, MOSIS ID, district ID, and relevant demographic information, including the student's race; the student's disability diagnosis; if the student has been in the district for less than a year; if the student has been in the building for less than a year; if the student is Limited English Proficiency (LEP); if the student qualifies for free and reduced lunch (SES); if the student has an individualized education plan (IEP); if the student is an English-language learner (ELL)/LEP who has been in the school for less than one year and in the country for less than three years; if the student is an LEP/ELL Title 3, the number of months the

LEP/ELL student has been in the U.S.; if the student took MAP-A; and if the student is Title I. An example is included in Appendix C, Figure C.4.

Student Achievement Level: This report lists all of the students in a school or district along with the year of testing, content area, grade level, achievement level, and MOSIS ID. An example is included in Appendix C, Figure C.5.

Achievement Level–4 Levels

These reports contain summary information on school or district performance in terms of the four MAP achievement levels. There are two types of achievement level reports: Achievement Level 4 Chart and Achievement Level 4 Report.

Achievement Level 4 Chart: This report charts the percentage of students classified as *Proficient* or *Advanced* across all grade levels tested in a particular content area. State-level, district-level, and/or school-level performance may be displayed on the chart. An example is included in Appendix C, Figure C.6.

Achievement Level 4 Report: This report summarizes the number and percentage of students in each achievement level. This report is comprised of 19 columns: Total; content area; grade; year; number of accountable (ACC) students; number of reportable (REP) students; number and percentage of students whose achievement level was not determined (LND); number and percentage of students classified in the *Below Basic* (BB) achievement level; number and percentage of students classified in the *Basic* (B) achievement level; number and percentage of students classified in the *Proficient* (P) achievement level; number and percentage of students classified in the *Advanced* (A) achievement level; MAP index score; mean MAP scale score; and the median *TerraNova* national percentile. The first column, Total, shows if aggregated or disaggregated information is being shown. A key to the abbreviations is found in the bottom left corner, as is the computation details for the MAP Index score. An example is included in Appendix C, Figure C.7.

Content Standard

The content standard reports summarize information about the content standards (CS).

Content Standards Report: This report has 14 columns: content area; grade level; category/type; year; percentage of points earned on content standard 1 (CS-1); points possible (PP) on CS-1; percentage of points earned on CS-2; PP on CS-2; percentage of points earned on CS-3; PP on CS-3; percentage of points earned on CS-4; PP on CS-4; percentage of points earned on CS-5; and PP on CS-5. The category/type column indicates if the data is aggregated or disaggregated data. An example is included in Appendix C, Figure C.8.

Content Standards Detail: This report shows the percentage of points each student achieved on each content standard within a particular content area. An example is included in Appendix C, Figure C.9.

Item Analysis Expanded
This set of reports provides detailed item-level results for the school or district aggregated either by the content standard or process standard.

Content Standard IBD EX: The Content Standard Item Benchmark Descriptions (IBD) Extended (EX) report contains item-level detail aggregated by content standard. The report is comprised of 11 columns: school code (SC), grade level (GR), standard number and description (desc.), code for the grade-level expectation (GLE), description of the GLE, depth of knowledge (DOK) of the item, session/item number where the item was in the operational test, question type (QT), points possible for the item, average points (avg pts) earned by students in the district on that item, and percentage of points earned by the students in the district on that item. An example is included in Appendix C, Figure C.10.

Goal Process Standard IBD EX: The Goal Process Standard Item Benchmark Descriptions (IBD) Extended (EX) report contains item-level detail aggregated by the goal process standard. The report is comprised of 12 columns: school code (SC), grade level (GR), goal, standard description (desc.), code for the grade-level expectation (GLE), description of the GLE, depth of knowledge (DOK) of the item, session/item number where the item was in the operational test, question type (QT), points possible for the item, the average points (avg pts) earned by students in the district on that item, and percentage of points earned by the students in the district on that item. An example is included in Appendix C, Figure C.11.

School/District Summary Reports

CTB provides DESE with school and district summary reports for each school and district in the state. These reports are intended for the sole use of DESE and are not distributed to schools and districts. These reports provide performance information for all students within a school or district who took the MAP.

The school or district is listed in the left-most column along with the purpose of the report. The main section of the Summary Report consists of a table that divides students from the school or district into achievement levels. The *Reportable* column shows the number of students with valid MAP scale scores. The *Accountable* column should equal the grade-level enrollment at the time the MAP was administered.

Within both the *Reportable* and *Accountable* columns, students are categorized as *Advanced, Proficient, Basic,* or *Below Basic.* The number and percentage of students falling into each achievement level is reported. A short description of the knowledge skills and abilities associated with each achievement level is also reported. Students who are not assigned to an achievement level will be classified as *Level Not Determined.* A short descriptor is also associated with this categorization.

Below this table, the norm-referenced summary statistics are reported for each school or district. The norm-referenced information includes the National Percentile (NP)

associated with the Mean Normal Curve Equivalent, the median NP, and the number of students with *TerraNova* scores.

On the back of these reports, the terms *Reportable* and *Accountable* are defined. A sample of the School/District Summary Report is provided in Appendix C, Figure C.12.

7.5 Data Structures

A data file referred to as General Research File (GRF) was provided to DESE by CTB/McGraw-Hill. It contains one record for every test book submitted; each record contains demographic information for each student as well as item responses, raw score, content and process standard raw scores, and scale score data for each content area.

7.5.1. General Research File

The layout for a state level GRF is included in Appendix C.

7.6 Interpreting Test Results

Individual Student Reports and Student Labels
The *Guide to Interpreting Results* was written for Missouri teachers and administrators who receive score reports from the 2009 administration of the MAP. The *Guide to Interpreting Results* was developed collaboratively by CTB/McGraw-Hill and DESE staff. DESE staff have opportunities to review, provide feedback, and give final approval.

This guide has six sections. The first section presents an overview of key terms and test-related concepts. The second section presents the Show-Me Content Standards/GLE Strands. The third section presents the Show-Me Performance Standards. The fourth section discusses assessment terms and the types of scores that will be presented on the score reports. The fifth section presents the achievement-level descriptors for all grade/content areas. Finally, the sixth section presents sample score reports.

The 2009 edition is available on the DESE website at:

http://dese.mo.gov/divimprove/assess/map/mapgenresources.html

Crystal Reports
Training for the Crystal Report tool is provided through DESE's RIFs as well as through online help tools. Appendix C contains an example of a training session provided by the St. Louis RIFs.

7.7 Summary

In summary, the overall purpose of reporting test results is to communicate various aggregations of student performance to stakeholders. These results are presented in the context of score reports that aid the user in understanding the meaning of the test scores. The reports and ancillary information developed by CTB/McGraw-Hill address multiple

best practices of the testing industry but in particular are related to the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

- Standard 4.1—Test documents should provide test users with clear explanations of the meaning and intended interpretations of derived score scales, as well as their limitations.
- Standard 5.10—When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate interpretations. The interpretations should describe in simple language what the test covers, what scores mean, the precision of the scores, common misinterpretations of test scores, and how scores will be used.
- Standard 6.2—Test documents should be complete, accurate, and clearly written so that the intended reader can readily understand the content.
- Standard 13.19—In educational settings, when average or summary scores for groups of students are reported, they should be supplemented with additional information about the sample size and shape or dispersion of score distributions.

Table 7. 1: Participation Rates: All Students

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	67357	99.71	67357	99.81		
4	66709	99.67	66709	99.82		
5	67307	99.67	67307	99.77	67307	99.72
6	65908	99.71	65908	99.77		
7	66531	99.68	66531	99.70		
8	67077	99.50	67077	99.54	67077	99.44

Table 7. 2: Participation Rates: Males

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	34570	99.68	34570	99.80		
4	34194	99.59	34194	99.79		
5	34532	99.66	34532	99.75	34532	99.68
6	33559	99.68	33559	99.73		
7	33923	99.66	33923	99.65		
8	34508	99.47	34508	99.53	34508	99.39

Table 7. 3: Participation Rates: Females

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	32718	99.74	32718	99.83		
4	32447	99.76	32447	99.85		
5	32645	99.69	32645	99.83	32645	99.78
6	32282	99.77	32282	99.83		
7	32547	99.72	32547	99.78		
8	32450	99.59	32450	99.62	32450	99.56

Table 7. 4: Participation Rates: White

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	50497	99.83	50497	99.83		
4	50276	99.81	50276	99.83		
5	50761	99.80	50761	99.82	50761	99.78
6	50229	99.82	50229	99.81		
7	51105	99.77	51105	99.75		
8	51174	99.68	51174	99.66	51174	99.60

Table 7. 5: Participation Rates: Black

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	12170	99.74	12170	99.75		
4	12009	99.63	12009	99.78		
5	12162	99.60	12162	99.64	12162	99.53
6	11609	99.64	11609	99.63		
7	11525	99.66	11525	99.55		
8	11874	99.19	11874	99.20	11874	98.98

Table 7. 6: Participation Rates: Hispanic

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	2958	98.85	2958	99.83		
4	2765	99.02	2765	99.78		
5	2702	98.89	2702	99.78	2702	99.63
6	2476	99.35	2476	99.96		
7	2321	98.92	2321	99.78		
8	2348	98.68	2348	99.57	2348	99.19

Table 7. 7: Participation Rates: Asian/Pacific Islander

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	1380	97.17	1380	99.93		
4	1295	96.06	1295	99.85		
5	1268	97.32	1268	99.68	1268	99.76
6	1245	97.43	1245	99.76		
7	1235	97.73	1235	99.76		
8	1260	98.02	1260	99.76	1260	99.52

Table 7. 8: Participation Rates: Native American/Alaskan

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	273	99.27	273	100.00		
4	297	100.00	297	100.00		
5	283	100.00	283	100.00	283	99.65
6	293	99.32	293	99.66		
7	287	100.00	287	99.30		
8	293	100.00	293	99.66	293	99.66

Table 7. 9: Participation Rates: Students Receiving Accommodations

Grade	Accountable in Comm. Arts	Percent Reportable in Comm. Arts	Accountable in Mathematics	Percent Reportable in Mathematics	Accountable in Science	Percent Reportable in Science
3	6419	99.88	6618	99.88		
4	6701	99.76	6915	99.91		
5	7385	99.73	7667	99.87	7344	99.84
6	7118	99.82	7345	99.86		
7	6870	99.83	7045	99.77		
8	6758	99.73	7003	99.77	6838	99.65

Table 7. 10: Summary Statistics for Communication Arts

Grade	N	Mean	Std. Dev.	Scale Scores by Percentiles				
				10	25	50	75	90
3	67,163	637.43	38.18	593	617	640	661	681
4	66,490	656.77	33.41	618	639	659	678	695
5	67,083	671.58	32.84	634	654	674	692	708
6	65,716	671.67	33.04	634	654	674	692	709
7	66,316	677.68	34.75	637	659	680	700	717
8	66,741	692.56	33.31	653	674	695	715	730

Table 7. 11: Summary Statistics for Mathematics

Grade	N	Mean	Std. Dev.	Scale Scores by Percentiles				
				10	25	50	75	90
3	67,232	621.67	36.76	576	600	623	645	664
4	66,587	644.20	33.89	602	624	646	666	683
5	67,155	662.07	40.52	612	639	665	688	708
6	65,755	678.87	39.56	631	656	681	704	725
7	66,330	683.63	40.72	633	660	686	710	731
8	66,770	703.60	38.63	656	682	707	729	748

Table 7. 12: Summary Statistics for Science

Grade	N	Mean	Std. Dev.	Scale Scores by Percentiles				
				10	25	50	75	90
5	67118	662.22	30.40	624	645	665	683	697
8	66702	695.65	30.94	657	678	699	717	731

Table 7. 13: Comparison of State-Level Means, 2006 through 2009 Census Data

		Communication Arts			Mathematics			Science		
Grade	**Year**	**N**	**Mean SS**	**S.D. SS**	**N**	**Mean SS**	**S.D. SS**	**N**	**Mean SS**	**S.D. SS**
3	2006	64,486	639.86	36.84	64,763	621.59	39.11			
	2007	66,347	639.58	38.04	66,640	622.40	38.72			
	2008	66,179	637.60	37.54	66,258	621.65	36.92			
	2009	67,163	637.43	38.18	67,232	621.67	36.76			
4	2006	65,179	654.55	38.56	65,306	643.88	37.07			
	2007	65,274	656.11	39.51	65,363	644.47	36.56			
	2008	66,873	655.61	33.63	66,944	644.18	34.19			
	2009	66,490	656.77	33.41	66,587	644.20	33.89			
5	2006	66,007	668.18	37.09	66,123	660.06	39.99			
	2007	65,461	671.01	37.14	65,498	663.21	41.50			
	2008	65,544	671.48	33.71	65,636	661.43	40.73	65,586	661.64	31.52
	2009	67,083	671.58	32.84	67,155	662.07	40.52	67,118	662.22	30.40
6	2006	66,948	666.85	33.70	67,017	673.30	39.80			
	2007	66,247	667.99	34.63	66,332	676.31	41.75			
	2008	65,672	671.27	33.50	65,716	678.46	41.13			
	2009	65,716	671.67	33.04	65,755	678.87	39.56			
7	2006	70,290	671.63	37.06	70,698	675.38	41.27			
	2007	67,167	672.11	36.26	67,554	677.41	42.62			
	2008	66,701	675.87	35.08	66,727	681.15	41.38			
	2009	66,316	677.68	34.75	66,330	683.63	40.72			
8	2006	72,483	686.85	37.87	72,542	697.73	40.37			
	2007	70,187	686.90	37.54	70,204	698.33	41.98			
	2008	67,278	691.05	33.57	67,312	701.30	39.40	67,209	694.36	30.67
	2009	66,741	692.56	33.31	66,770	703.60	38.63	66,702	695.65	30.94

Table 7. 14: Comparison of Percent of Students in each Achievement Level, Communication Arts 2006 through 2009 Census Data

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv
3	2006	65,344	1.3	8.8	47.5	25.7	16.7	42.4
	2007	67,259	1.4	9.4	46.6	25.8	16.8	42.6
	2008	66,357	0.3	9.3	50.2	25.2	15.1	40.3
	2009	67,357	0.3	9.6	49.8	25.1	15.2	40.3
4	2006	65,849	1.0	10.6	44.5	28.8	15.0	43.8
	2007	65,982	1.1	10.5	43.4	28.2	16.8	45.1
	2008	67,049	0.3	8.0	46.7	33.4	11.7	45.1
	2009	66,709	0.3	7.6	45.8	33.6	12.7	46.3
5	2006	66,704	1.0	9.1	44.8	29.6	15.4	45.0
	2007	66,098	1.0	8.3	42.9	29.8	18.0	47.8
	2008	65,734	0.3	6.4	45.1	32.2	15.9	48.1
	2009	67,307	0.3	6.3	44.6	33.9	14.9	48.8
6	2006	67,709	1.1	11.9	44.8	31.6	10.6	42.2
	2007	67,045	1.2	11.2	44.0	31.8	11.7	43.6
	2008	65,830	0.2	9.0	43.5	34.0	13.4	47.4
	2009	65,908	0.3	8.6	43.4	33.8	13.9	47.7
7	2006	71,632	1.9	13.7	41.8	30.5	12.2	42.7
	2007	68,404	1.8	13.1	40.7	32.8	11.6	44.4
	2008	66,923	0.3	10.0	40.7	36.1	12.9	49.0
	2009	66,531	0.3	8.7	40.3	37.2	13.6	50.8
8	2006	73,516	1.4	9.1	48.0	26.6	15.0	41.5
	2007	71,200	1.4	8.7	48.3	26.9	14.6	41.6
	2008	67,574	0.4	5.7	45.8	33.1	15.0	48.1
	2009	67,077	0.5	5.3	44.5	33.4	16.3	49.7

Table 7. 15: Comparison of Percent of Students in each Achievement Level, Mathematics 2006 through 2009 Census Data

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv
3	2006	65,325	0.9	7.2	48.7	33.3	10.0	43.3
	2007	67,257	0.9	7.2	46.9	35.0	10.0	45.0
	2008	66,357	0.1	6.5	49.6	35.0	8.8	43.8
	2009	67,357	0.2	6.8	48.5	35.6	8.8	44.4
4	2006	65,845	0.8	8.3	47.5	34.4	9.0	43.4
	2007	65,975	0.9	8.1	46.5	35.2	9.3	44.5
	2008	67,049	0.2	7.6	48.0	36.0	8.2	44.2
	2009	66,709	0.2	7.3	48.2	36.6	7.8	44.4
5	2006	66,703	0.9	8.1	47.8	32.7	10.6	43.3
	2007	66,075	0.9	7.6	44.9	33.1	13.4	46.6
	2008	65,734	0.1	7.5	46.5	34.4	11.4	45.8
	2009	67,307	0.2	7.5	45.1	35.6	11.6	47.2
6	2006	67,706	1.0	11.1	44.1	34.4	9.5	43.9
	2007	67,039	1.1	11.1	40.0	35.5	12.3	47.8
	2008	65,830	0.2	9.5	39.6	37.8	12.9	50.7
	2009	65,908	0.2	8.9	40.7	37.5	12.6	50.1
7	2006	71,575	1.2	17.4	38.5	32.7	10.2	42.9
	2007	68,405	1.2	16.7	37.1	33.2	11.7	44.9
	2008	66,923	0.3	13.9	36.3	36.7	12.8	49.5
	2009	66,531	0.3	12.5	35.2	37.6	14.3	51.9
8	2006	73,523	1.3	21.1	37.8	27.6	12.2	39.8
	2007	71,190	1.4	21.4	36.6	26.6	14.0	40.6
	2008	67,574	0.4	18.0	37.7	29.9	13.9	43.8
	2009	67,077	0.5	16.4	36.8	31.5	14.9	46.4

Table 7. 16: Percent of Students in each Achievement Level, Science 2009 Census Data

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv
5	2008	65,734	0.2	11.2	44.0	29.6	14.9	44.5
	2009	67,307	0.3	10.6	44.1	30.3	14.8	45.1
8	2008	67,574	0.5	19.3	37.0	36.7	6.5	43.2
	2009	67,077	0.6	18.2	36.5	37.2	7.6	44.8

Table 7. 17: Summary of Score Reports for Spring 2009

Score Report		Paper Report		Electronic Report		
		Parent	Teacher	Principal	System	DESE
Student Score Labels			X			
Individual Student Report		X				
Performance Summary Report	School Performance Summary Report					X
	District Performance Summary Report					X
	Crystal Reports			X	X	

Table 7. 18: Types of Reports Available to Districts through Crystal Reports

Crystal Report	Sub Reports
Administrative Report	Level Not Determined MAP Alternate MAP Scale Score Summary MAP Student Demographic Student Achievement Level
AYP	AYP AYP Additional Indicator AYP Growth Report AYP Growth Target Met AYP Growth Trajectory AYP Summary
Achievement Level-4 Levels	Achievement Level 4 Chart Achievement Level 4 Report
Achievement Level-5 Levels	Achievement Level 5 Chart Top2 Bottom2 Achievement Level 5 Chart Top2 Bottom2 State District Achievement Level 5 Report MAP Index Chart 5 Levels MAP Index Chart 5 Levels District State
Content Standard	Content Standards Content Standards Detail
Item Analysis Expanded	Content Standard IBD EX Goal Process Standard IBD EX

CHAPTER 8: ACHIEVEMENT-LEVEL SETTING

A Bookmark standard setting was held in 2005 to establish cut scores for the Communication Arts and Mathematics MAPs. Another Bookmark standard setting was held in 2008 to establish cut scores for the Science MAP. In this chapter, we briefly describe the MAP achievement-level setting, and we present the cut scores established and the achievement-level descriptors derived from the achievement-level setting.

A detailed discussion of the Communication Arts and Mathematics achievement-level setting may be found in the *Missouri Assessment Program Final Bookmark Standard Setting Technical Report* (2005). A detailed discussion of the Science achievement-level setting may be found in the *Missouri Assessment Program Bookmark Standard Setting Technical Report 2008 for Missouri Achievement-Level Setting Grades 5, 8, and 11 Science* (2008). These Technical Reports address AERA, APA, and NCME Standard 4.19:

> When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented.

We briefly overview the rationale and procedures used for MAP standard setting below.

In terms of the validity of the MAP scores, it is essential to understand that descriptors and cut scores are established in a collaborative, participatory process, largely driven by the input of Missouri teachers and educators. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, and cut scores in particular.

8.1. Legislation Affecting MAP Standard Setting

A modified Bookmark Standard Setting Procedure (BSSP) was used to establish cut scores for the Communication Arts and Mathematics MAPs for Grades 3 through 8 and high school and Science for Grades 5, 8, and 11. A modification of the Bookmark was used to meet the requirements of Senate Bill 1080, which requires that cut scores be established for the MAPs that are like the cut scores established for the National Assessment of Educational Progress (NAEP).

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.* For Communication Arts and Mathematics, no fewer than 26% and no more than 44% of

students could be classified as *Proficient* or *Advanced*. For Science, no fewer than 27% of and no more than 48% of students could be classified as *Proficient* or *Advanced*.

The pre-specified range was determined using the results from NAEP and MAP. For all three subject areas, the high end of the range (in terms of scale score points) was based on NAEP results. This was the lowest percentage of students classified as *Proficient* or *Advanced* on the NAEP test for Grades 4 and 8 Reading, Mathematics, and Science using both national and state data.

The low end of the range (in terms of scale score points) was based on the 2005 MAP results for the Communication Arts and Mathematics standard setting and on the 2007 MAP results for Science. This was the highest percentage of students classified as *Proficient* or *Advanced* on the previous years' tests.

8.2. Bookmark Standard Setting Procedure

A modified BSSP was used to establish cut scores on the Communication Arts, Mathematics, and Science MAP. At both workshops, the BSSP involved three rounds of discussion and voting. AERA, APA, & NCME (1999) Standard 4.21 says

> When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of item or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge and experience to bear in a reasonable way.

The Technical Reports associated with each standard setting give detailed reports of the standard setting design and procedure. Here, we discuss the major activities of the three rounds.

Round 1: In this round, panelists discussed target students (the students for whom they were placing cut scores), took the test, studied and discussed the test items in order of difficulty, and made initial recommendations of cut scores.

Round 2: In this round, panelists were shown their Round 1 recommendations and the percentage of students in each achievement level as a result of their Round 1 recommendations. They discussed their Round 1 recommendations for cut scores and made another recommendation based on their Round 2 discussions.

Round 3: In this round, panelists were shown their Round 2 recommendations and the percentage of students in each achievement level as a result of their Round 2 recommendations. They discussed their Round 2 recommendations for cut scores and made another recommendation based on their Round 3 discussions.

Following Round 3, panelists wrote draft achievement-level descriptors which were later edited by CTB and DESE staff.

The Missouri State Board of Education approved the cut scores as recommended by the standard-setting panelists.

8.3. Cut Scores

In this section, we present the cut scores for each grade/content area of MAP. Tables 8.1 through 8.3 show the cut scores for Communication Arts, Mathematics, and Science, respectively. Please note that we only present the cut scores for Grades 3 through 8. The high school MAPs are no longer part of the assessment system.

8.4. Achievement-Level Descriptors

In Appendix D of this report, we present the short achievement-level descriptors that were drafted during the standard setting and finalized between CTB and DESE staff after the standard setting. We only present the short achievement-level descriptors for those grades that are currently part of the MAP.

8.5. Summary

This chapter presented a brief overview of the standard setting process used for the grade-level MAPs, as well as the rationale behind the standard setting. The standard settings are addressed in more detail in the relevant Technical Reports. The standard settings undertaken by CTB/McGraw-Hill address the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

- Standard 4.19—When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented.
- Standard 4.21—When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of item or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge and experience to bear in a reasonable way.

Table 8. 1: Communication Arts Cut Scores

Grade	Cut Scores		
	Basic	Proficient	Advanced
3	592	648	673
4	612	662	691
5	625	675	702
6	631	676	704
7	634	680	712
8	639	696	723

Table 8. 2: Mathematics Cut Scores

Grade	Cut Scores		
	Basic	Proficient	Advanced
3	568	628	667
4	596	651	688
5	605	668	706
6	628	681	721
7	640	685	724
8	670	710	741

Table 8. 3: Science Cut Scores

Grade	Cut Scores		
	Basic	Proficient	Advanced
5	626	669	692
8	671	703	735

CHAPTER 9: EVIDENCE OF CONSTRUCT-RELATED VALIDITY

Evidence for construct-related validity—the meaning of test scores and the inferences they support—is the central concept underlying the MAP validation process. In this section, CTB presents evidence of construct-related validity through studies of test reliability, convergent validity, and divergent validity. All analyses in this section are based on census data.

Chapter 9 of this report demonstrates the adherence to AERA, APA, & NCME (1999) Standards 1.11, 1.18, 2.1, 2.2, 2.4, 2.14, and 2.15. Each standard will be discussed in the pertinent section of this chapter.

9.1 Minimization of Construct-Irrelevant Variance and Construct-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 (see Chapter 3 for more information on 1 through 5 and Chapter 6 for more information on 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. 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 appropriately represented.

9.2 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 AERA, APA, & NCME (1999) Standards indicate:

> . . . 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 (27).

In accordance with the AERA, APA, & NCME (1999) 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.

There are several specific AERA, APA, & NCME (1999) Standards that this chapter addresses. These include Standards 2.1, 2.2, and 2.4, each articulated below.

Standard 2.1 *For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported.*

The total score reliabilities are discussed in 9.2.1. of this chapter. The subscore reliabilities and SEMs are presented in Section 9.4.3. The SEM of the total score is discussed in Section 9.2.2.

Standard 2.2 *The standard error of measurement, both overall and conditional (if relevant), should be reported both in raw score or original scale units and in units of each derived score recommended for use in test interpretation.*

The overall SEM is discussed in Section 9.2.2 and is reported in scale score units. The conditional SEM is discussed in Section 9.2.3.

Standard 2.4 *Each method of quantifying the precision or consistency of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select examinees for reliability analyses and descriptive statistics on these samples should be reported.*

Section 9.2 discusses different ways of measuring test reliability, including reliability of raw scores, overall SEM, IRT-based conditional SEM, and decision consistency of achievement-level classifications. The sample on which these statistics are computed is discussed in Section 6.1 of Chapter 6.

9.2.1. 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 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

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma_i^2}{\sigma_X^2}\right], \qquad (9.1)$$

where n is the number of items on the test, σ_i^2 is the variance of item i and σ_x^2 is the variance of the total test score.

Total test reliability measures such as Cronbach's coefficient alpha and SEM 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 9.1, 9.2, and 9.3 for Communication Arts, Mathematics, and Science, respectively. These reliability coefficients were computed using the census data. All reliability statistics are 0.90 or greater for all tests indicating acceptable reliability. The reliability statistics by subgroup are reported and discussed in Chapter 10.

9.2.2. Standard Error of Measurement

The reliability of reported test scores can be characterized by the standard errors associated with the scores. The 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. The SEM is an index of the random variability in test scores and is defined as follows:

$$\text{SEM} = SD\sqrt{1 - R_{xx'}}, \qquad (9.2)$$

where SD represents standard deviation of the raw score distribution, and $R_{xx'}$ is estimated by $\hat{\alpha}$ as expressed in Equation 9.1.

The overall SEM is expressed in scale score units and is a test level statistic. The SEM is summarized in Table 9.4 with respect to all students and each subgroup. There were some observable trends in SEM by the subgroups in all grades and content areas. The SEM for

females tended to be smaller than for males. The SEM was smaller for non-accommodated students than for accommodated students.

9.2.3. Conditional Standard Error of Measurement

In contrast to SEM, the conditional standard errors of measurement (CSEM) express the degree of measurement error in scale score units and are conditioned on the ability of the student. We report the CSEM in support of AERA, APA, & NCME (1999) Standard 2.14, which states:

> Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of the cut scores.

In further compliance with Standard 2.14, the CSEM of each cut score is reported in Table 9.5.

The CSEMs are defined as the reciprocal of the square root of the test information function and can be estimated across all points of the ability continuum (Hambleton & Swaminathan, 1985):

$$\text{CSEM}(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}}, \tag{9.3}$$

where $I(\theta_i)$ is the test information function, as a sum of item information function 2, obtained as

$$I(\theta_i) = \sum_j \frac{p'_{ij}(\theta_i)^2}{p_{ij}(\theta_i)\, q_{ij}(\theta_i)}, \tag{9.4}$$

where $p'_{ij}(\theta_i)$ is the derivative of $p_{ij}(\theta_i)$ and $q_{ij}(\theta_i) = 1 - p_{ij}(\theta_i)$.

Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scale scores) and are smaller in the middle of the score distribution and higher at the tails. This pattern is seen for all MAP CSEMs and is to be expected when IRT methods are used. The CSEMs at the three cut scores that define the performance levels are presented in Table 9.5 and range from 6 to 15 scale score points.

Figures 9.1 through 9.14 display the CSEM curves and cut scores 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 9.1 through 9.14 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.

9.2.4. Classification Accuracy and Consistency

The *Standards* (AERA, APA, & NCME, 1999) 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.

Classification Consistency: Classification consistency (also known as 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. It is difficult, however, 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. A common practice, therefore, is to estimate decision consistency from one administration of a test. These analyses directly address AERA, APA, & NCME (1999) Standard 2.15, which states:

> When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument.

Classification Accuracy: Classification 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). It is common to estimate classification accuracy by utilizing a psychometric model to find true scores corresponding to observed scores.

In other words, classification *consistency* refers to the agreement between two observed scores, while classification *accuracy* refers to the agreement between the observed score and the true score. A straightforward approach to classification consistency estimation can be expressed in terms of a contingency table representing the probability of a particular classification outcome under specific scenarios. For example, the following table is a contingency table of $(H+1) \times (H+1)$, where H is the number of cut scores, such that two cut scores yield a 3×3 contingency table.

Example of Contingency Table with 2 Cut Scores

	Level 1	Level 2	Level 3	Sum
Level 1	P_{11}	P_{21}	P_{31}	$P_{\cdot 1}$
Level 2	P_{12}	P_{22}	P_{32}	$P_{\cdot 2}$
Level 3	P_{13}	P_{23}	P_{33}	$P_{\cdot 3}$
Sum	$P_{1\cdot}$	$P_{2\cdot}$	$P_{3\cdot}$	1.0

CTB used a method suggested by Kolen and Kim (2005) for estimating consistency and accuracy that involves the generation of item responses using item parameters based on the IRT model (see also Kim, Choi, Um, & Kim, 2006; Kim, Barton, & Kim, 2007). Two sets of item responses are generated using a set of item parameters and an examinee's ability distribution from a single test administration. These two sets of item responses are considered as an examinee's responses on two administrations of the same form. The procedure is described below and is implemented with KKCLASS software (Kim, 2005).

- Step 1: Obtain item parameters (**I**) and ability distribution weight ($\hat{g}(\theta)$) at each quadrature point from a single test.
- Step 2: Compute two raw scores at each quadrature point. At a given quadrature point θ_i, generate two sets of item responses using the item parameters from a test form, assuming that the same test form was administered twice to an examinee with the true ability θ_i.
- Step 3: Construct a classification matrix at each quadrature point. Determine the joint event for the cells in table above using the raw scores obtained from Step 2.
- Step 4: Repeat Steps 2 and 3 *R* times and get average values over *R* replications.
- Step 5: Multiply distribution weight ($\hat{g}(\theta)$) by average values in Step 4 for each quadrature point, and sum across all quadrature points. From this final contingency table, decision consistency indices, such as consistency agreement and kappa, can be computed.
- Step 6. Because examinee ability is estimated at each quadrature point, this quadrature point can be considered the true score. Therefore, decision accuracy is computed using both examinee estimated ability (observed scores) and quadrature point (true score).

Tables 9.6 and 9.7 show the results for the 2009 MAP classification analyses. Classification consistency and classification accuracy condition on performance level (Table 9.6) and on cut score (Table 9.7) are presented. As can be seen in Table 9.6, classification accuracy conditioned on achievement level ranged from 0.66 to 0.89, and classification consistency conditioned on achievement level ranged from 0.56 to 0.84. The magnitude of classification consistency and accuracy measures is influenced by key features of the test design including the number of items, number of cut scores, and the reliability and associated SEM.

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. 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.* Table 9.7 shows the classification accuracy and consistency estimates when conditioned on MAP cut scores. The classification accuracy statistics are above 0.90 while the classification consistency statistics are at or above 0.87. These results suggest that consistent and accurate performance level classifications are being made for students in Missouri based on the MAP.

9.2.5. 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 Chapter 6). 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 Chapter 6, only 12 items total were flagged for poor model/data fit across all 14 grade/content area MAPs.

9.3 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 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. Standard 1.11 of the AERA, APA, & NCME (1999) Standards says,

> If the rationale for a test use or interpretation depends on premises about the relationship among parts of the test, evidence concerning internal structure should be provided.

In this section, we examine the internal structure by evaluating the unidimensionality assumption 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 9.15 to 9.28, scree plots (Cattell, 1966) of eigenvalues are presented to illustrate the relative dominance of the first principal component in each MAP test. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other factors.

For all grade/content area MAPs, the ratio of the variance accounted for by the first factor to the second and third is sufficiently large to support the claim that these tests are unidimensional. All of the MAP subject area tests exhibit first principal components accounting for more than 17% of the test variance (see Tables 9.8 through 9.10). To further investigate the unidimensionality of the Communication Arts, Mathematics, and Science tests, the ratio of the first eigenvalue to the second eigenvalue was explored (see Tables 9.8 through 9.10). These ratios show that the first eigenvalue is at least five 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, Mathematics, and Science tests are essentially unidimensional.

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 to doing well in a subject, does not appear to create significant nuisance factors.

9.4 Analyses by Content Standard

Three sets of analyses were conducted for the content standard level in another attempt to assess the construct validity of MAP. First, the reliability of each Content Standard was computed. Second, correlation coefficients that measure the relationship between the Content Standards were computed. Finally, the SEM was computed for each reportable content standard.

9.4.1. Reliability of Content Standards

Cronbach's (1951) coefficient alpha was computed for each of the Content Standards by grade/content area using the census data. Tables 9.11 through 9.13 report the reliability

statistics along the diagonal of each matrix for each grade/content area. Reliability indices, such as Cronbach's coefficient alpha, are a function of the number of test items. It is expected that coefficient alpha would be low for a Content Standard assessed by a small number of items (e.g., Writing Formally and Informally).

9.4.2. 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 9.11 through 9.13 report the uncorrected Pearson product-moment (PPM) correlation coefficients, the PPM corrected for attenuation (CAPPM), in addition to the reliability coefficients described above. The PPM among the Content Standard subscores is presented below the diagonal portion of the matrix, the CAPPM is presented above the diagonal portion of the matrix, and the reliability coefficients are shown on the diagonal in each table.

The uncorrected PPM in Tables 9.11 through 9.13 should be interpreted in the context of the reliability coefficient. In general, we expect to see lower PPM coefficients between variables that are less reliable. Overall, the PPM coefficients show that performance on one Content Standard is moderately to strongly related to performance on another Content Standard within the same content area. As noted above, the value of the correlation coefficients will be affected by the limited number of items measuring each Content Standard. So, caution should be used when comparing the PPM coefficients measuring the relationships between Content Standards to those measuring the relationships between content areas (Table 9.17). 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).

Indeed, the PPM between two content standard subscores may be artificially low because of measurement error.

AERA, APA, & NCME (1999) Standard 1.18, states:

> When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported.

We can correct for the attenuation of the PPM statistically using Spearman's formula:

$$CAPPM = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}, \tag{9.5}$$

where r_{xy} is the PPM between two content standards, r_{xx} is the reliability of one of those content standards, and r_{yy} is the reliability for the other content standard.

Across all tables, the CAPPM indicate strong relationships between the content standards. In some cases, the CAPPM is greater than 1.0. “Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed” (Schumacker, 1996). The strong relationships suggested by the CAPPM in Tables 9.11 through 9.13 are further evidence of the validity of the test construct. Since the overall content area is comprised of the content standard subscores and the content area is expected to measure a single dimension, then we would expect that these subscores are also highly related.

9.4.3. Standard Error of Measurement of Content Standards

In this chapter, we report the SEM associated with each of the content standards in Tables 9.14 through 9.16 for Communication Arts, Mathematics, and Science, respectively. These SEMs are reported in the percent correct metric as content standards are reported in that metric.

9.5 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 Mathematics and Communication Arts scale scores for students who took both MAP subject area tests in 2009. These correlations are based on the census data and the results are shown in Table 9.17. The correlation coefficients ranged from 0.75 (between Communication Arts and Mathematics in Grades 3 through 6) to 0.84 (between Mathematics and Science in Grade 8). The correlation coefficients suggest that individual student scores for Communication Arts and Mathematics are moderately to highly related. The correlation coefficients between Science and the other two content areas suggest that the Science MAP is highly related to the Communication Arts and Mathematics MAP. The tests are not perfectly related to each other, suggesting that different constructs are being tapped; however, the test scores do appear at least moderately related to one another, suggesting they are tapping into a similar knowledge base. This is especially true of the Science test. The Science MAP is comprised of many constructed-response items, which may help account for its relationship with the Communication Arts test. The Science MAP tests similar thinking skills and item types as are found in the Mathematics MAP, which may help account for the strong correlation between the Science and Mathematics test scores.

9.6 Summary

In summary, the overall purpose of each of the test administration workshops and the ancillary materials is to keep districts informed about policies and procedures related to testing in general and the MAP program in particular. The information imparted is clearly

related to standardizing the administration of the MAP, maintaining the security of the assessment, allowing access to the assessments for special populations by clearly delineating appropriate accommodations, and by providing guidance on appropriate interpretations of the test results. These communication and training efforts by DESE and the ancillary information developed by CTB/McGraw-Hill address multiple best practices of the testing industry but in particular are related to the following *Standards for Educational and Psychological Testing* (1999):

- Standard 1.11—If the rational for a test use or interpretation depends on premises about the relationship among parts of the test, evidence concerning internal structure should be provided.
- Standard 1.18—When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported.
- Standard 2.1—For each total scores, subscore, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported.
- Standard 2.2—The standard error of measurement, both overall and conditional (if relevant), should be reported both in raw score or original scale units and in units of each derived score recommended for use in test interpretation.
- Standard 2.4—Each method of quantifying the precision or consistency of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select examinees for reliability analyses and descriptive statistics on these samples should be reported.
- Standard 2.14—Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of the cut scores.
- Standard 2.15—When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same for or alternate forms of the instrument.

Table 9. 1: Reliability in Communication Arts

Grade	Number of Items	Number of Score Points	Cronbach's Alpha
3	57	67	0.90
4	55	63	0.92
5	55	62	0.92
6	55	62	0.90
7	61	72	0.92
8	61	68	0.91

Table 9. 2: Reliability in Mathematics

Grade	Number of Items	Number of Score Points	Cronbach's Alpha
3	60	67	0.92
4	65	77	0.92
5	62	69	0.91
6	61	68	0.92
7	62	69	0.92
8	64	76	0.93

Table 9. 3: Reliability in Science

Grade	Number of Items	Number of Score Points	Cronbach's Alpha
5	53	79	0.91
8	59	91	0.93

Table 9. 4: SEM by Subgroup

Grade	Category	Group	CA SEM	MA SEM	SC SEM
3	**Overall**		11.77	10.53	
	Ethnicity	White (not Hispanic)	11.97	10.86	
		Black (not Hispanic)	11.89	10.68	
		Hispanic	11.65	10.48	
		Asian/Pacific Islander	12.23	11.63	
		Native American	11.59	9.85	
	Gender	Male	11.68	10.73	
		Female	11.61	10.64	
	Accommo-dations	No	11.38	10.45	
		Yes	13.81	10.71	
4	**Overall**		9.63	9.40	
	Ethnicity	White (not Hispanic)	9.49	9.47	
		Black (not Hispanic)	10.03	9.15	
		Hispanic	9.41	9.27	
		Asian/Pacific Islander	7.32	10.12	
		Native American	9.78	9.72	
	Gender	Male	9.63	9.25	
		Female	9.53	9.25	
	Accommo-dations	No	9.29	9.01	
		Yes	11.79	10.38	
5	**Overall**		9.57	11.88	9.02
	Ethnicity	White (not Hispanic)	9.28	11.89	8.48
		Black (not Hispanic)	10.38	12.64	9.76
		Hispanic	9.57	11.68	9.32
		Asian/Pacific Islander	9.82	12.38	9.12
		Native American	8.44	11.12	9.19
	Gender	Male	9.63	11.76	9.41
		Female	9.32	11.79	8.77
	Accommo-dations	No	8.98	11.70	8.92
		Yes	12.52	13.35	10.58
6	**Overall**		10.24	11.47	
	Ethnicity	White (not Hispanic)	9.92	11.00	
		Black (not Hispanic)	10.84	12.21	
		Hispanic	10.06	11.25	
		Asian/Pacific Islander	10.38	12.19	
		Native American	10.64	11.46	
	Gender	Male	10.40	11.52	
		Female	10.12	11.48	
	Accommo-dations	No	9.81	11.53	
		Yes	13.51	12.35	

Table 9. 4: SEM by Subgroup (Cont'd)

Grade	Category	Group	CA SEM	MA SEM	SC SEM
7	**Overall**		9.95	11.23	
	Ethnicity	White (not Hispanic)	9.93	10.73	
		Black (not Hispanic)	10.06	12.23	
		Hispanic	9.76	11.21	
		Asian/Pacific Islander	10.37	11.50	
		Native American	10.32	11.29	
	Gender	Male	10.11	11.14	
		Female	9.82	11.09	
	Accommo-dations	No	9.53	10.45	
		Yes	12.62	13.70	
8	**Overall**		10.10	10.29	8.19
	Ethnicity	White (not Hispanic)	9.88	10.06	7.73
		Black (not Hispanic)	10.99	11.85	9.40
		Hispanic	9.99	10.58	8.16
		Asian/Pacific Islander	10.23	9.98	8.21
		Native American	10.54	10.29	8.50
	Gender	Male	10.42	10.73	8.54
		Female	9.74	10.30	7.78
	Accommo-dations	No	9.56	9.85	7.91
		Yes	12.95	14.09	10.15

Table 9. 5: Conditional Standard Error of Measurement at the Basic, Proficient, & Advanced Cut Scores

Content Area	Grade	Basic		Proficient		Advanced	
		Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM
Communication Arts	3	592	10	648	9	673	12
	4	612	8	662	9	691	12
	5	625	8	675	8	702	10
	6	631	9	676	8	704	10
	7	634	9	680	8	712	11
	8	639	9	696	8	723	9
Mathematics	3	568	9	628	9	667	15
	4	596	9	651	8	688	12
	5	605	12	668	9	706	13
	6	628	11	681	9	721	13
	7	640	12	685	8	724	10
	8	670	10	710	7	741	7
Science	5	626	9	669	7	692	8
	8	671	7	703	6	735	7

Table 9. 6: Decision Accuracy and Consistency Conditioned on Level of Achievement

Content Area	Grade	Accuracy				Consistency			
		Below Basic	Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.
Communication Arts	3	0.86	0.87	0.66	0.87	0.80	0.84	0.56	0.72
	4	0.88	0.87	0.72	0.82	0.79	0.81	0.67	0.66
	5	0.87	0.88	0.74	0.84	0.80	0.83	0.68	0.73
	6	0.83	0.84	0.74	0.84	0.78	0.80	0.67	0.67
	7	0.87	0.84	0.81	0.81	0.79	0.81	0.71	0.67
	8	0.85	0.88	0.74	0.84	0.77	0.84	0.67	0.71
Mathematics	3	0.85	0.88	0.78	0.83	0.76	0.84	0.71	0.65
	4	0.86	0.89	0.80	0.84	0.76	0.84	0.73	0.65
	5	0.85	0.87	0.77	0.84	0.77	0.83	0.72	0.68
	6	0.84	0.85	0.81	0.81	0.78	0.82	0.73	0.68
	7	0.87	0.82	0.83	0.85	0.77	0.77	0.77	0.72
	8	0.86	0.84	0.83	0.88	0.80	0.79	0.74	0.81
Science	5	0.85	0.84	0.78	0.85	0.78	0.81	0.67	0.74
	8	0.87	0.84	0.83	0.82	0.81	0.78	0.79	0.73

Table 9. 7: Decision Accuracy and Consistency at Achievement Cut Points

Content Area	Grade	Accuracy			Consistency		
		Below Basic/ Basic	Basic/ Prof.	Prof./Adv.	Below Basic/ Basic	Basic/ Prof.	Prof./Adv.
Communication Arts	3	0.97	0.91	0.93	0.96	0.88	0.90
	4	0.98	0.91	0.93	0.97	0.87	0.90
	5	0.98	0.92	0.93	0.97	0.89	0.90
	6	0.97	0.91	0.93	0.96	0.88	0.90
	7	0.97	0.92	0.93	0.96	0.88	0.90
	8	0.98	0.91	0.93	0.98	0.88	0.90
Mathematics	3	0.98	0.92	0.94	0.97	0.88	0.91
	4	0.98	0.92	0.95	0.96	0.88	0.93
	5	0.97	0.92	0.94	0.96	0.89	0.91
	6	0.97	0.92	0.94	0.96	0.89	0.91
	7	0.96	0.93	0.95	0.94	0.90	0.92
	8	0.95	0.93	0.96	0.94	0.90	0.93
Science	5	0.96	0.92	0.94	0.95	0.89	0.92
	8	0.95	0.92	0.96	0.94	0.90	0.95

Table 9. 8: Principal Component Analysis for Communication Arts

Grade	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
Grade 3			
First Component	10.31	18.10	18.10
Second Component	2.13	3.74	21.84
Ratio (First/Second)	4.83		
Grade 4			
First Component	10.76	19.56	19.56
Second Component	1.72	3.13	22.69
Ratio (First/Second)	6.25		
Grade 5			
First Component	10.86	19.74	19.74
Second Component	1.49	2.70	22.44
Ratio (First/Second)	7.30		
Grade 6			
First Component	9.79	17.79	17.79
Second Component	1.66	3.01	20.80
Ratio (First/Second)	5.91		
Grade 7			
First Component	11.18	18.33	18.33
Second Component	1.57	2.57	20.91
Ratio (First/Second)	7.12		
Grade 8			
First Component	10.50	17.21	17.21
Second Component	2.03	3.33	20.54
Ratio (First/Second)	5.18		

Table 9. 9: Principal Component Analysis for Mathematics

Grade	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
Grade 3			
First Component	11.14	18.57	18.57
Second Component	1.80	3.00	21.57
Ratio (First/Second)	6.19		
Grade 4			
First Component	12.23	18.82	18.82
Second Component	1.76	2.71	21.53
Ratio (First/Second)	6.95		
Grade 5			
First Component	10.81	17.44	17.44
Second Component	1.60	2.58	20.02
Ratio (First/Second)	6.76		
Grade 6			
First Component	11.14	18.26	18.26
Second Component	1.41	2.30	20.57
Ratio (First/Second)	7.93		
Grade 7			
First Component	11.80	19.03	19.03
Second Component	1.86	3.00	22.03
Ratio (First/Second)	6.34		
Grade 8			
First Component	12.53	19.58	19.58
Second Component	2.14	3.35	22.93
Ratio (First/Second)	5.85		

Table 9. 10: Principal Component Analysis for Science

Grade	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
Grade 5			
First Component	10.28	19.40	19.40
Second Component	1.74	3.28	22.68
Ratio (First/Second)	5.91		
Grade 8			
First Component	12.50	21.18	21.18
Second Component	1.79	3.03	24.21
Ratio (First/Second)	6.99		

Table 9. 11: Reliability (Diagonal) of Each Content Standard, Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) Among Content Standards: Communication Arts

Grade	No.	Content Standard	Number of Items	1	2	3	4	5
3	1	Speaking/Writing Standard English	16	0.74	0.89	0.89		0.90
	2	Reading Fiction/Poetry/Drama	26	0.70	0.83	0.96		1.14
	3	Reading Nonfiction	12	0.62	0.71	0.66		1.14
	4	Writing Formally/Informally	NR*					
	5	Combined Reading	38	0.72	0.97	0.86		0.87
4	1	Speaking/Writing Standard English	10	0.62	0.89	0.88		0.90
	2	Reading Fiction/Poetry/Drama	38	0.66	0.89	0.93		1.10
	3	Reading Nonfiction	6	0.53	0.68	0.60		1.10
	4	Writing Formally/Informally	NR					
	5	Combined Reading	44	0.67	0.98	0.80		0.90
5	1	Speaking/Writing Standard English	13	0.71	0.89	0.86		0.90
	2	Reading Fiction/Poetry/Drama	28	0.69	0.86	0.91		1.11
	3	Reading Nonfiction	13	0.63	0.73	0.75		1.05
	4	Writing Formally/Informally	NR					
	5	Combined Reading	41	0.72	0.98	0.85		0.90
6	1	Speaking/Writing Standard English	13	0.69	0.91	0.92		0.93
	2	Reading Fiction/Poetry/Drama	17	0.66	0.77	0.94		1.08
	3	Reading Nonfiction	24	0.69	0.74	0.82		1.14
	4	Writing Formally/Informally	NR					
	5	Combined Reading	41	0.72	0.90	0.96		0.88
7	1	Speaking/Writing Standard English	16	0.72	0.89	0.90	0.90	0.90
	2	Reading Fiction/Poetry/Drama	30	0.70	0.86	0.98	0.91	1.11
	3	Reading Nonfiction	13	0.65	0.78	0.74	0.97	1.11
	4	Writing Formally/Informally	3	0.54	0.59	0.58	0.49	0.93
	5	Combined Reading	42	0.72	0.97	0.90	0.62	0.89
8	1	Speaking/Writing Standard English	15	0.66	0.87	0.88		0.89
	2	Reading Fiction/Poetry/Drama	19	0.64	0.81	0.93		1.12
	3	Reading Nonfiction	24	0.64	0.75	0.80		1.09
	4	Writing Formally/Informally	NR					
	5	Combined Reading	43	0.68	0.95	0.92		0.89

*NR=Not Reported

Table 9. 12: Reliability (Diagonal) of Each Content Standard, and Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) Among Content Standards: Mathematics

Grade	No.	Content Standard	Number of Items	1	2	3	4	5
3	1	Number and Operations	22	0.83	0.99	0.92	0.95	1.04
	2	Algebraic Relationship	10	0.75	0.70	0.94	0.93	1.02
	3	Geometric and Spatial	13	0.68	0.64	0.66	0.89	0.98
	4	Measurement	8	0.68	0.61	0.57	0.62	0.96
	5	Data and Probability	7	0.68	0.62	0.58	0.54	0.52
4	1	Number and Operations	18	0.79	1.02	0.93	0.98	0.93
	2	Algebraic Relationship	13	0.79	0.76	0.96	0.97	0.95
	3	Geometric and Spatial	12	0.66	0.67	0.64	0.95	0.96
	4	Measurement	12	0.73	0.71	0.63	0.70	0.90
	5	Data and Probability	10	0.65	0.64	0.60	0.58	0.61
5	1	Number and Operations	16	0.77	0.97	0.82	0.98	0.95
	2	Algebraic Relationship	11	0.71	0.69	0.89	1.02	0.99
	3	Geometric and Spatial	12	0.58	0.60	0.65	0.91	0.89
	4	Measurement	12	0.72	0.71	0.61	0.70	1.00
	5	Data and Probability	11	0.65	0.64	0.56	0.65	0.61
6	1	Number and Operations	17	0.78	0.96	0.88	0.94	0.93
	2	Algebraic Relationship	10	0.67	0.63	0.96	0.99	0.99
	3	Geometric and Spatial	11	0.62	0.61	0.64	0.97	0.94
	4	Measurement	12	0.70	0.66	0.65	0.71	0.94
	5	Data and Probability	11	0.68	0.65	0.63	0.66	0.69
7	1	Number and Operations	16	0.76	0.97	0.92	0.97	0.97
	2	Algebraic Relationship	11	0.66	0.61	1.01	1.00	1.02
	3	Geometric and Spatial	12	0.68	0.67	0.73	0.98	0.97
	4	Measurement	11	0.71	0.65	0.70	0.70	0.97
	5	Data and Probability	12	0.73	0.69	0.72	0.71	0.76
8	1	Number and Operations	14	0.71	0.94	0.89	0.86	0.97
	2	Algebraic Relationship	20	0.72	0.83	0.92	0.88	1.00
	3	Geometric and Spatial	10	0.58	0.65	0.62	0.91	0.96
	4	Measurement	9	0.63	0.69	0.61	0.74	0.93
	5	Data and Probability	11	0.70	0.78	0.64	0.69	0.73

Table 9. 13: Reliability (Diagonal) of Each Content Standard, and Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) Among Content Standards: Science

Grade	No.	Content Standard	Number of Items	01	02	03	04	05	06	07	08
5	01	Matter and Energy	6	0.61	1.07	0.96	0.96	1.03	0.99	0.93	0.99
	02	Force and Motion	4	0.56	0.46	1.04	1.00	1.05	1.04	1.00	1.08
	03	Characteristics of Living Organisms	6	0.55	0.51	0.54	1.01	0.99	1.00	0.96	1.05
	04	Interactions of Organisms	6	0.56	0.51	0.56	0.57	0.98	0.97	0.93	1.04
	05	Earth's Processes	6	0.62	0.55	0.56	0.57	0.60	1.00	0.94	1.00
	06	The Universe	5	0.55	0.50	0.52	0.52	0.55	0.51	0.96	1.01
	07	Scientific Inquiry	15	0.61	0.57	0.59	0.59	0.61	0.58	0.71	0.99
	08	Technology and the Environment	5	0.52	0.49	0.52	0.53	0.52	0.49	0.56	0.46
8	01	Matter and Energy	7	0.65	0.98	0.98	0.97	1.00	0.98	0.92	0.96
	02	Force and Motion	5	0.54	0.47	0.91	0.96	0.97	0.99	0.92	0.95
	03	Characteristics of Living Organisms	7	0.63	0.49	0.63	0.98	0.98	0.93	0.89	0.96
	04	Interactions of Organisms	5	0.64	0.54	0.63	0.66	1.00	0.96	0.92	1.01
	05	Earth's Processes	8	0.60	0.50	0.59	0.61	0.56	0.99	0.90	0.98
	06	The Universe	4	0.55	0.47	0.51	0.54	0.52	0.48	0.90	0.93
	07	Scientific Inquiry	19	0.67	0.57	0.65	0.68	0.61	0.57	0.83	0.92
	08	Technology and the Environment	4	0.57	0.48	0.56	0.60	0.54	0.47	0.62	0.54

Table 9. 14: Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Communication Arts Content Standards

Grade	Content Standard	Mean	Std. Deviation	SEM
3	1	66.73	20.23	10.26
	2	70.01	18.52	7.61
	3	74.89	18.66	10.93
	5	71.64	17.31	6.27
4	1	77.85	18.12	11.20
	2	73.32	17.48	5.90
	3	53.44	23.25	14.70
	5	69.41	17.37	5.58
5	1	69.12	21.23	11.45
	2	64.97	19.06	7.08
	3	77.60	20.11	10.15
	5	68.42	18.27	5.89
6	1	70.00	20.00	11.19
	2	74.75	19.60	9.32
	3	60.74	17.90	7.70
	5	65.70	17.39	6.00
7	1	70.61	18.90	10.02
	2	70.71	18.92	7.20
	3	63.35	19.75	10.17
	4	71.12	18.51	13.19
	5	67.91	18.22	5.99
8	1	60.03	19.13	11.19
	2	62.74	20.37	8.78
	3	71.21	16.87	7.64
	5	66.72	17.49	5.91

Table 9. 15: Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Mathematics Content Standards

Grade	Content Standard	Mean	Std. Deviation	SEM
3	1	75.22	18.56	7.63
	2	73.92	20.37	11.23
	3	76.83	18.24	10.64
	4	86.65	17.42	10.74
	5	69.77	20.14	13.97
4	1	67.27	20.75	9.51
	2	68.93	22.01	10.76
	3	76.97	18.53	11.09
	4	59.03	21.83	11.96
	5	80.21	16.72	10.48
5	1	74.37	20.16	9.58
	2	61.81	23.18	12.84
	3	69.71	20.45	12.10
	4	59.94	21.13	11.57
	5	64.24	18.21	11.40
6	1	72.28	20.70	9.64
	2	63.55	21.58	13.13
	3	66.82	22.46	13.49
	4	58.50	22.03	11.89
	5	72.74	20.49	11.39
7	1	63.69	21.73	10.62
	2	59.97	20.68	12.98
	3	57.98	23.42	12.19
	4	50.59	22.22	12.13
	5	67.73	22.39	11.08
8	1	70.74	19.22	10.42
	2	49.15	22.26	9.18
	3	45.34	19.34	12.00
	4	51.67	26.69	13.53
	5	53.33	20.53	10.69

Table 9. 16: Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Science Content Standards

Grade	Content Standard	Mean	Std. Deviation	SEM
5	1	42.24	24.02	15.06
	2	43.66	23.09	17.03
	3	62.55	24.03	16.37
	4	63.49	22.57	14.78
	5	48.15	22.56	14.32
	6	58.80	22.71	15.91
	7	58.35	19.04	10.25
	8	66.29	19.51	14.41
8	1	37.40	18.67	11.03
	2	54.69	22.48	16.43
	3	40.71	20.36	12.34
	4	58.19	25.42	14.73
	5	42.77	17.19	11.36
	6	39.19	22.26	15.99
	7	54.25	21.36	8.91
	8	46.30	25.17	17.04

Table 9. 17: Inter-Correlation of Communication Arts, Mathematics, and Science Scale Scores

Grade	CA/MA	CA/SC	MA/SC
3	0.75	-	-
4	0.75	-	-
5	0.75	0.77	0.80
6	0.75	-	-
7	0.77	-	-
8	0.77	0.81	0.84

Figure 9. 1: SEM Plot Communication Arts, Grade 3

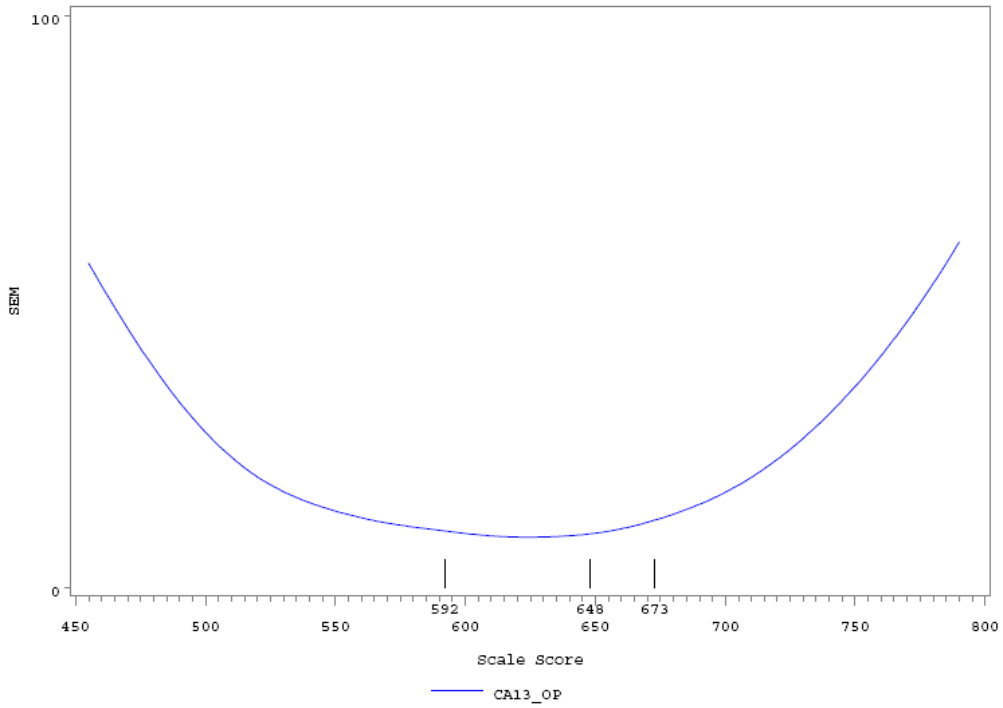


Figure 9. 2: SEM Plot Communication Arts, Grade 4

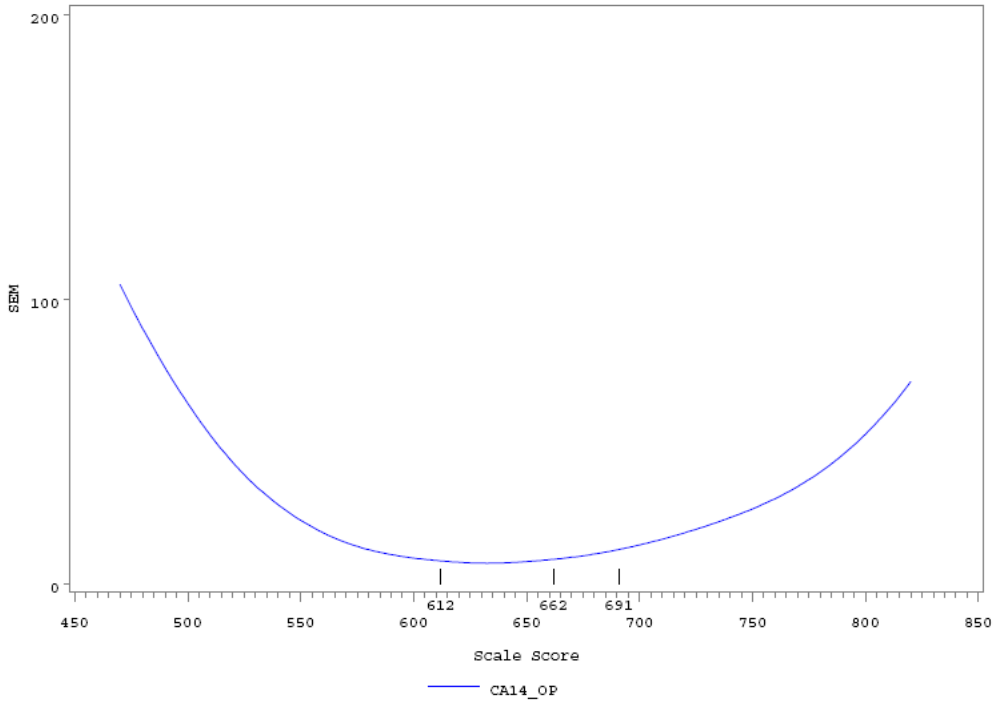


Figure 9. 3: SEM Plot Communication Arts, Grade 5

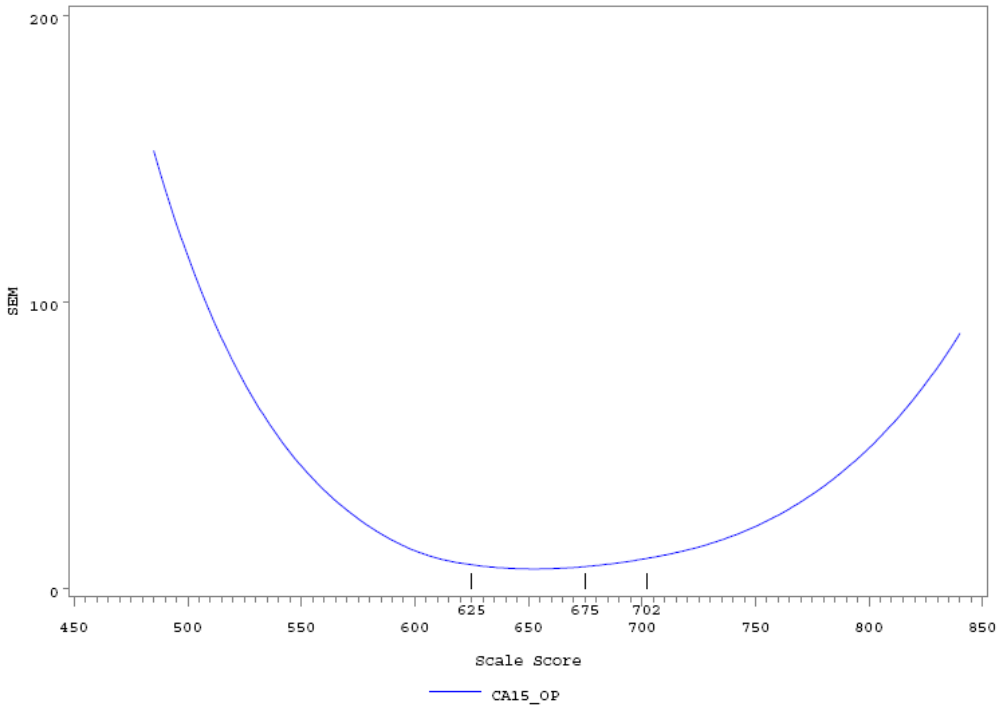


Figure 9. 4: SEM Plot Communication Arts, Grade 6

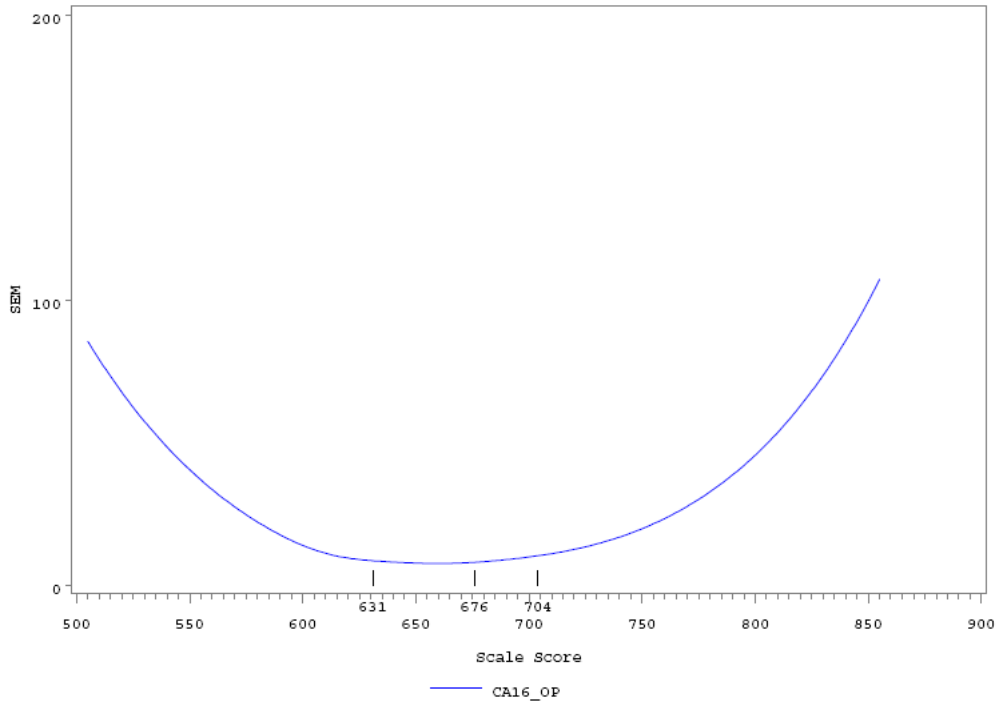


Figure 9. 5: SEM Plot Communication Arts, Grade 7

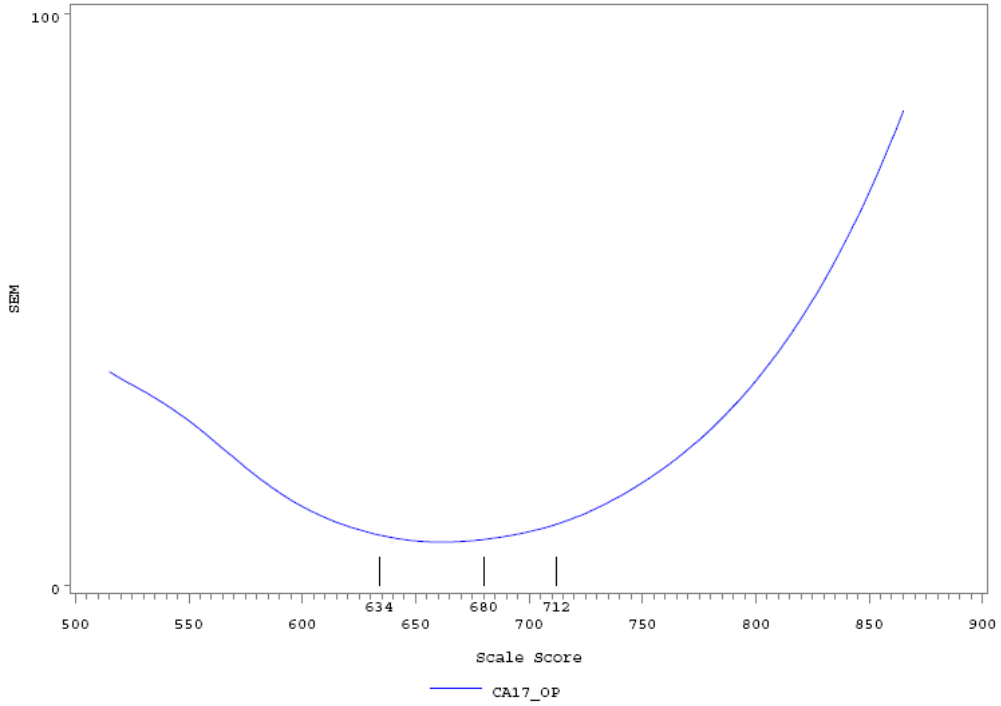


Figure 9. 6: SEM Plot Communication Arts, Grade 8

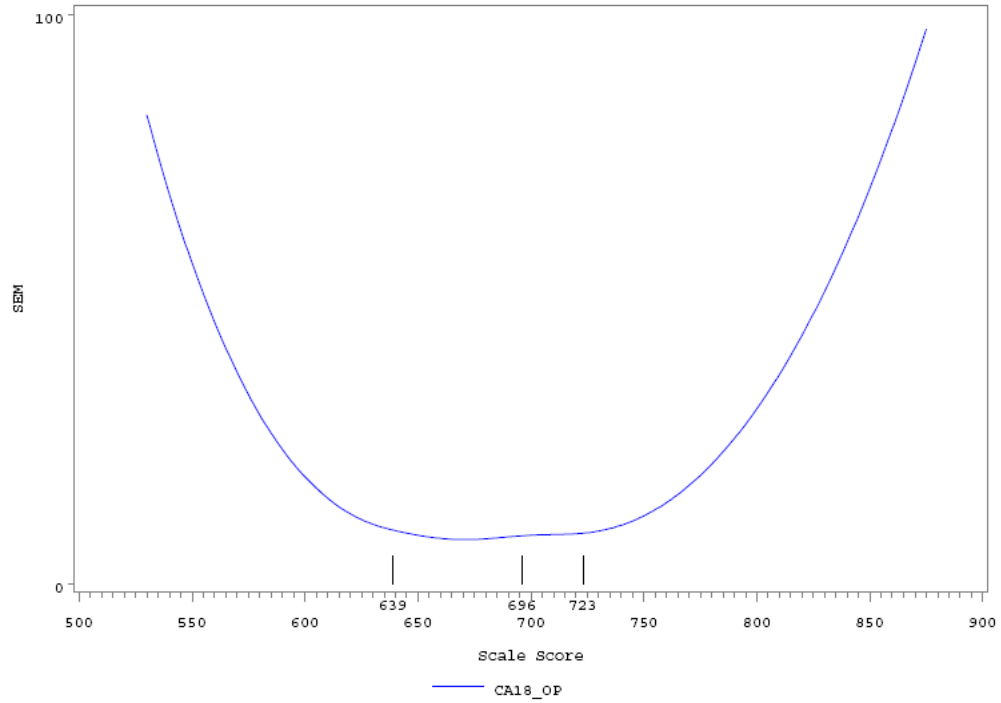


Figure 9. 7: SEM Plot Mathematics, Grade 3

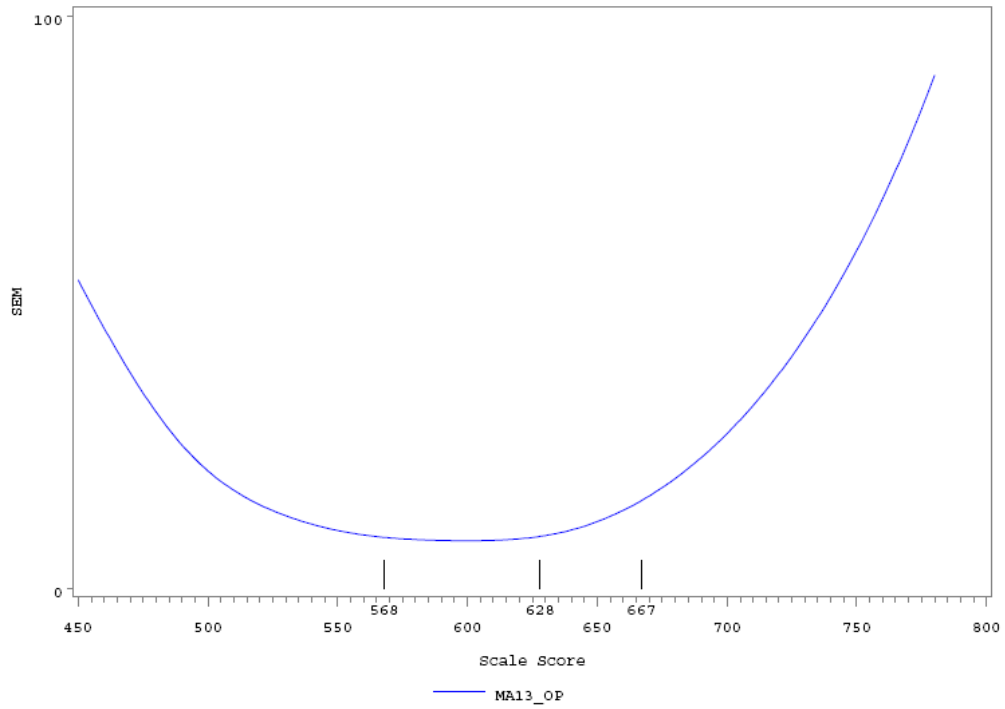


Figure 9. 8: SEM Plot Mathematics, Grade 4

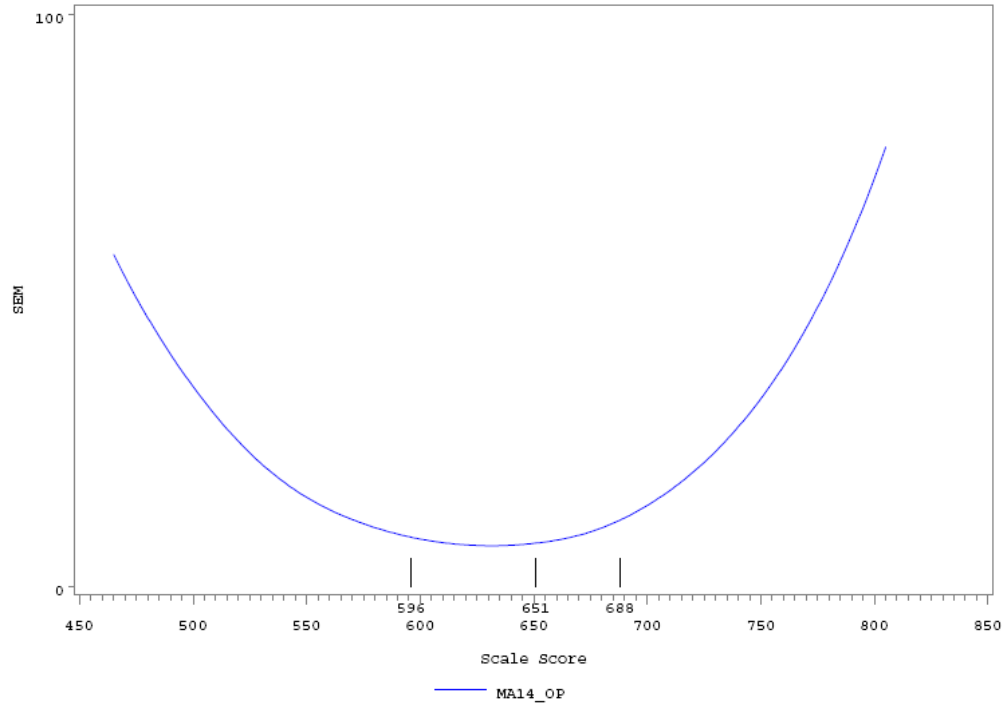


Figure 9. 9: SEM Plot Mathematics, Grade 5

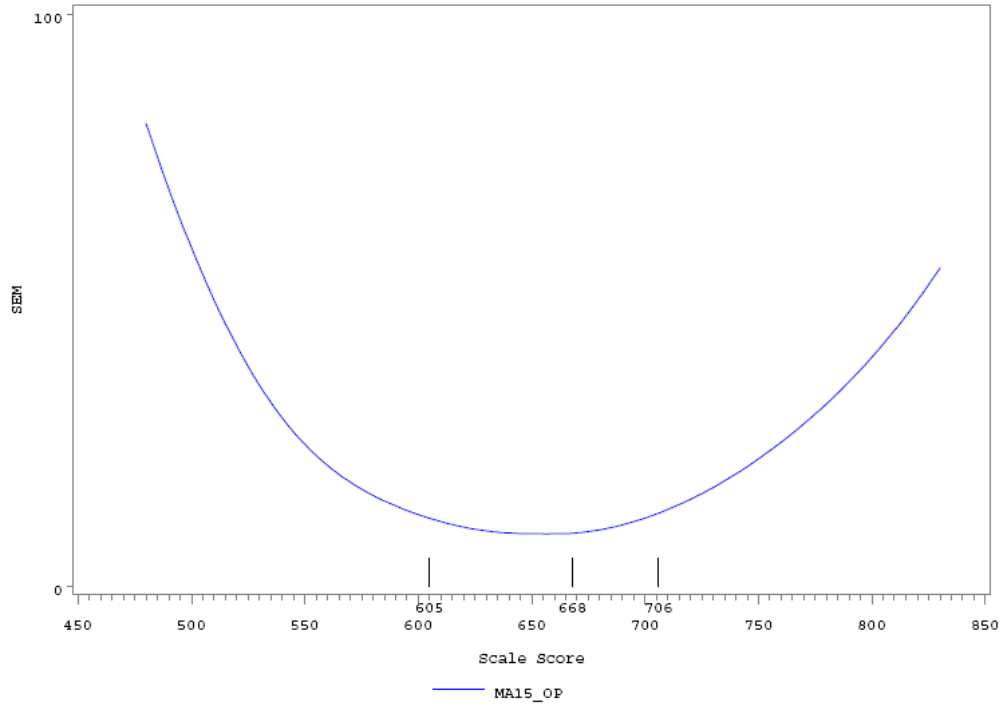


Figure 9. 10: SEM Plot Mathematics, Grade 6

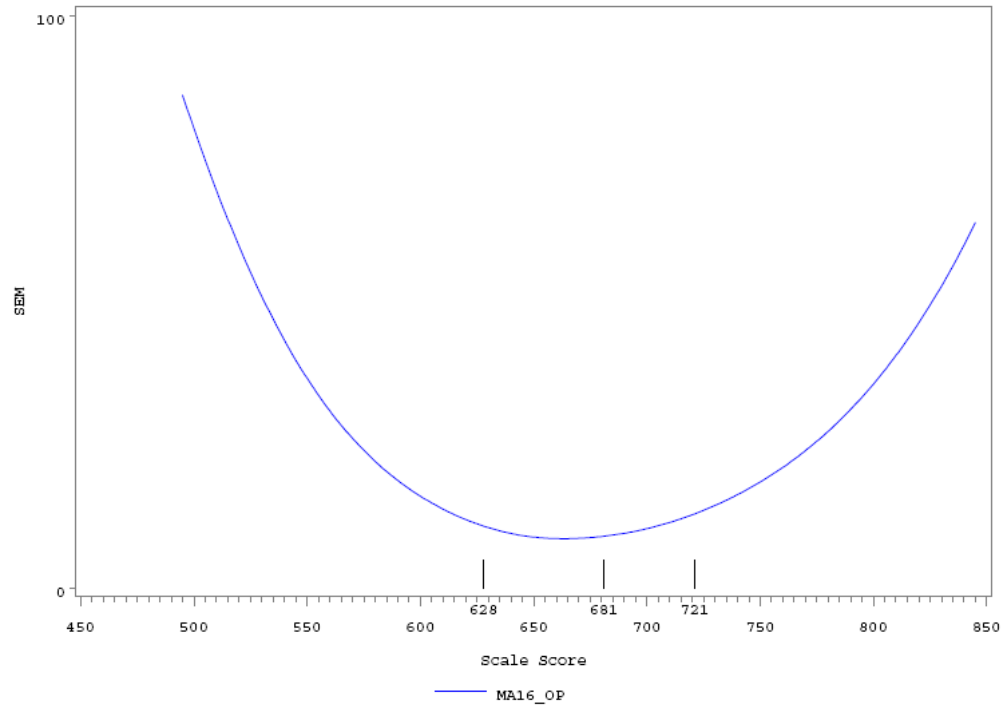


Figure 9. 11: SEM Plot Mathematics, Grade 7

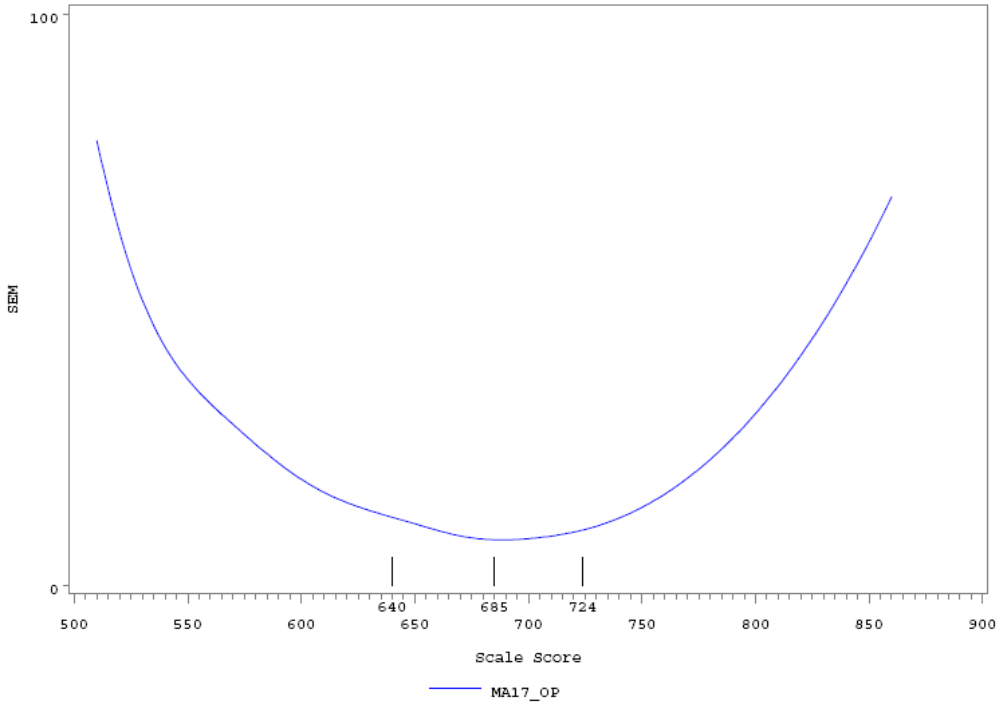


Figure 9. 12: SEM Plot Mathematics, Grade 8

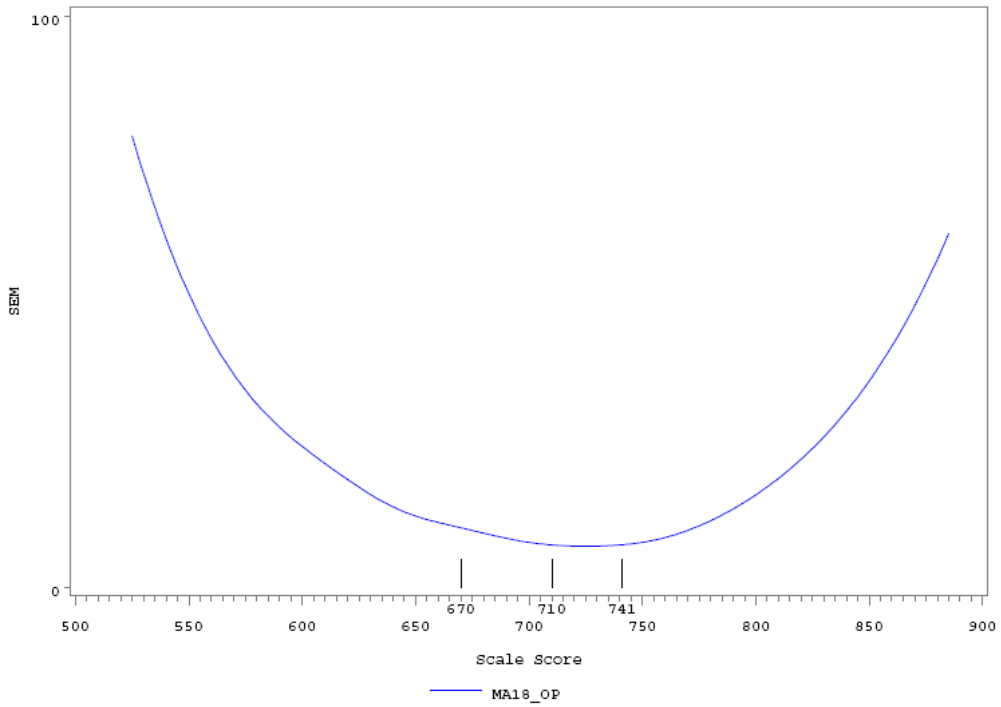


Figure 9. 13: SEM Plot Science, Grade 5

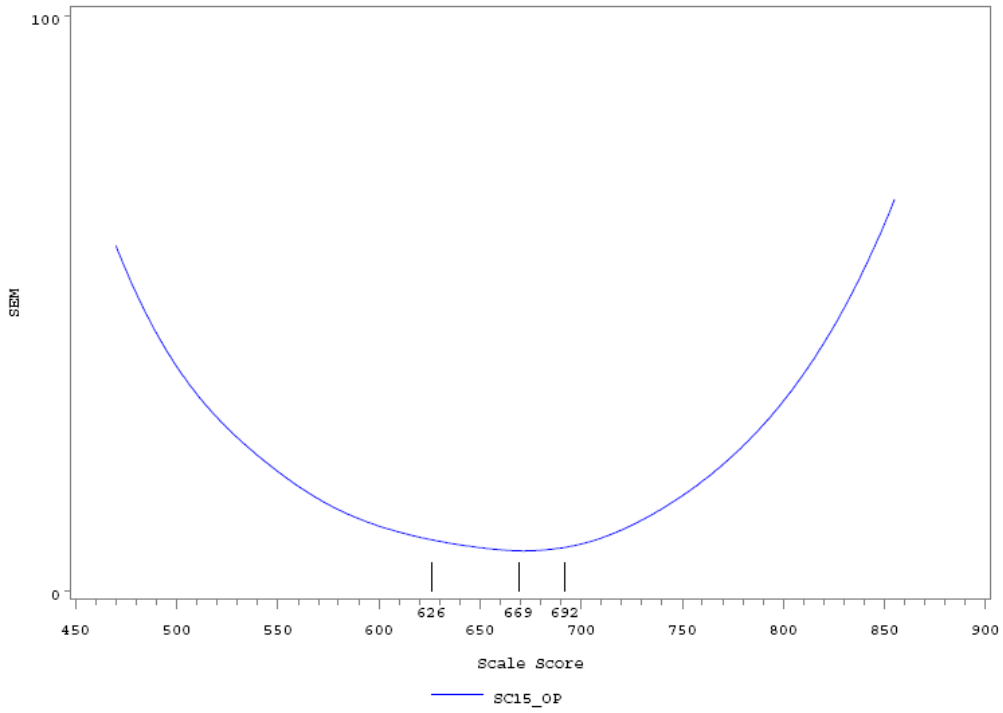


Figure 9. 14: SEM Plot Science, Grade 8

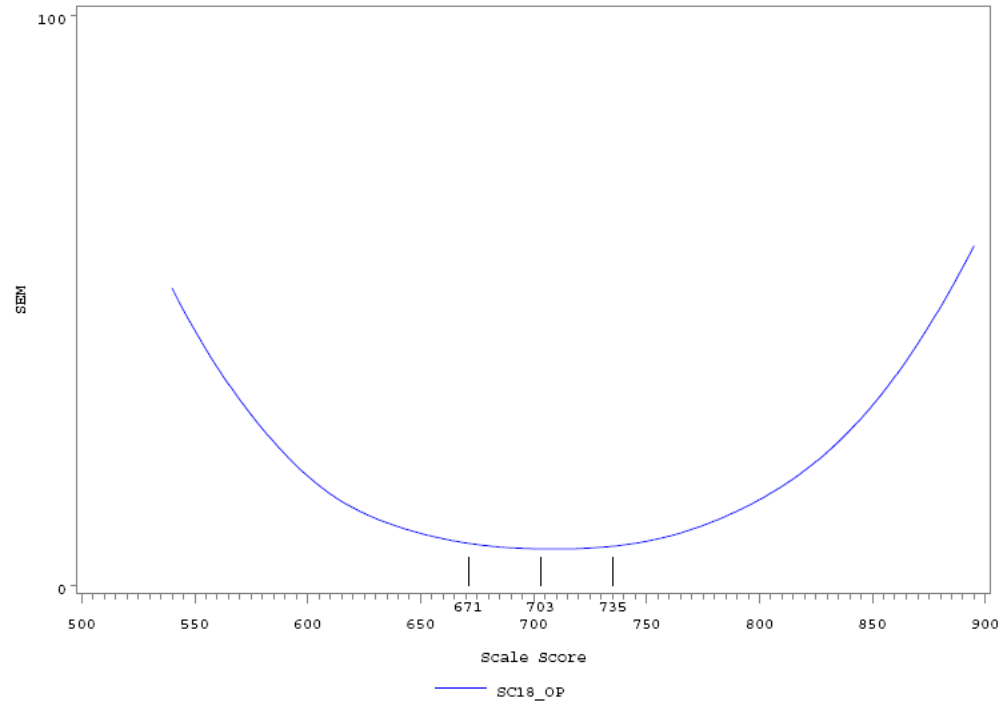


Figure 9. 15: Scree Plot Communication Arts, Grade 3

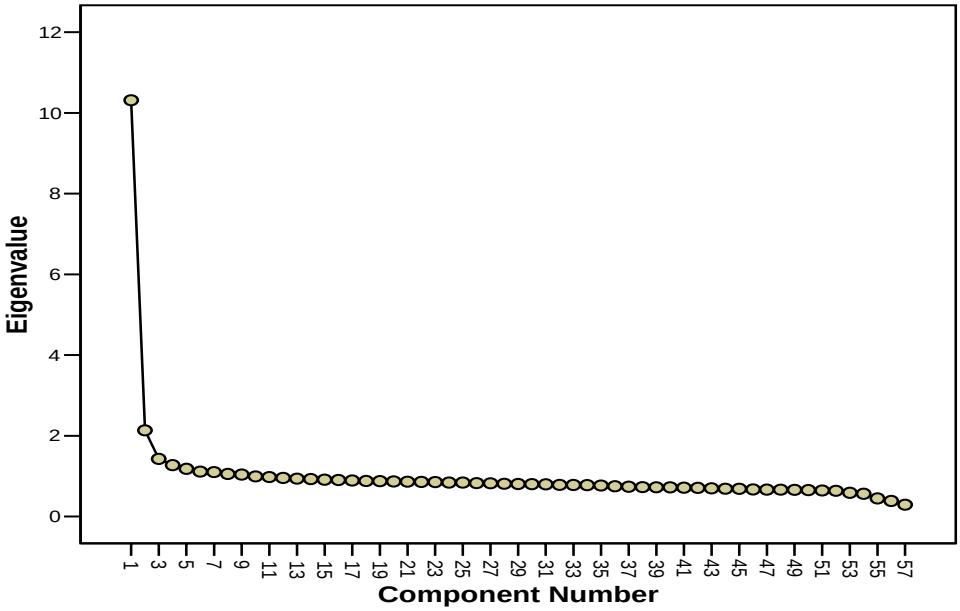


Figure 9. 16: Scree Plot Communication Arts, Grade 4

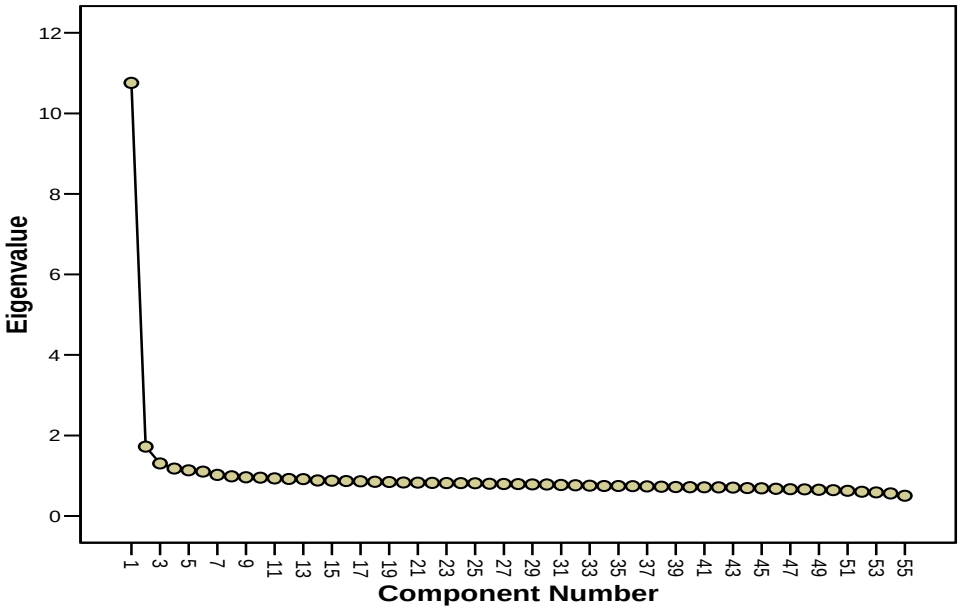


Figure 9. 17: Scree Plot Communication Arts, Grade 5

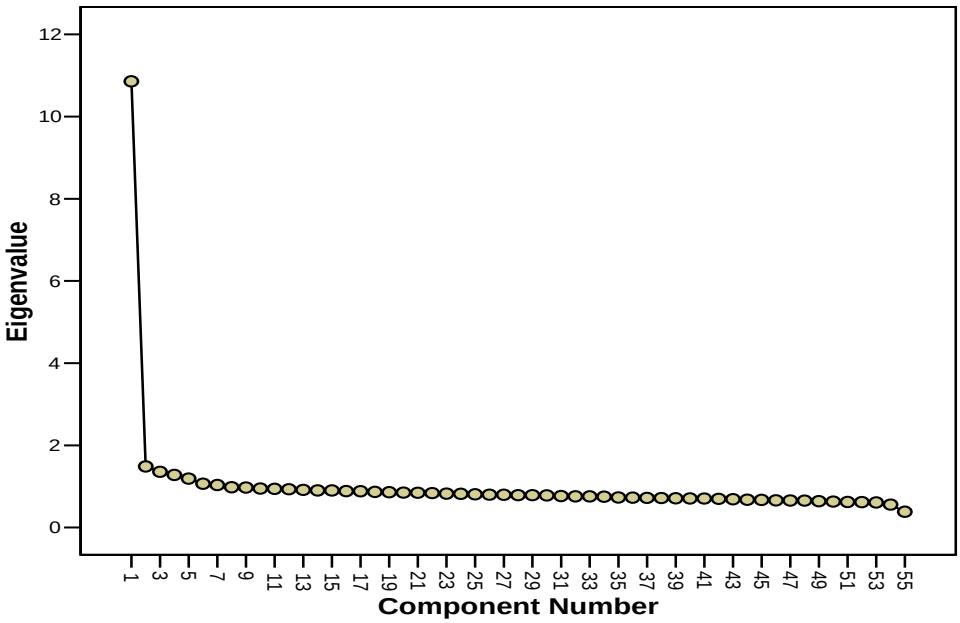


Figure 9. 18: Scree Plot Communication Arts, Grade 6

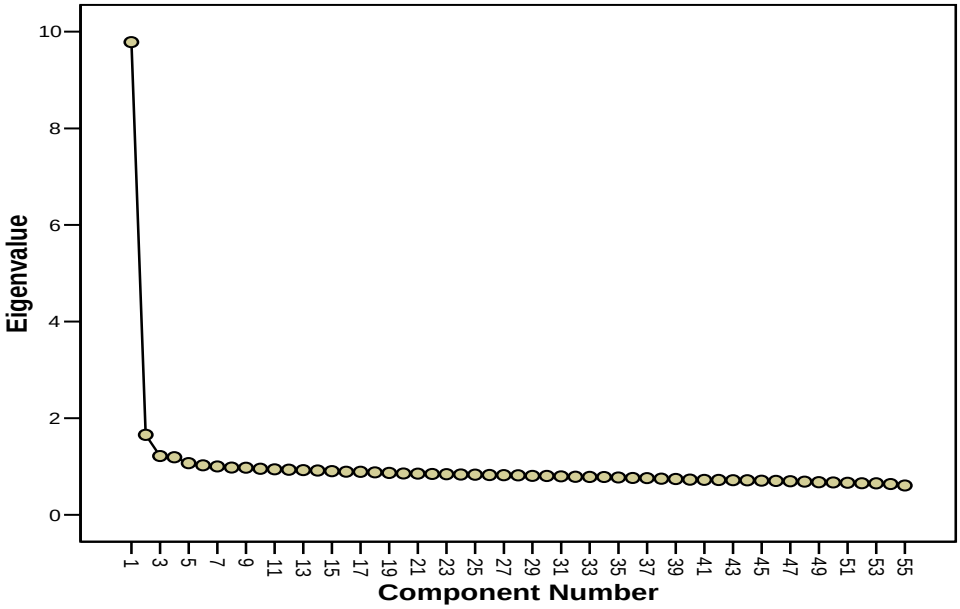


Figure 9. 19: Scree Plot Communication Arts, Grade 7

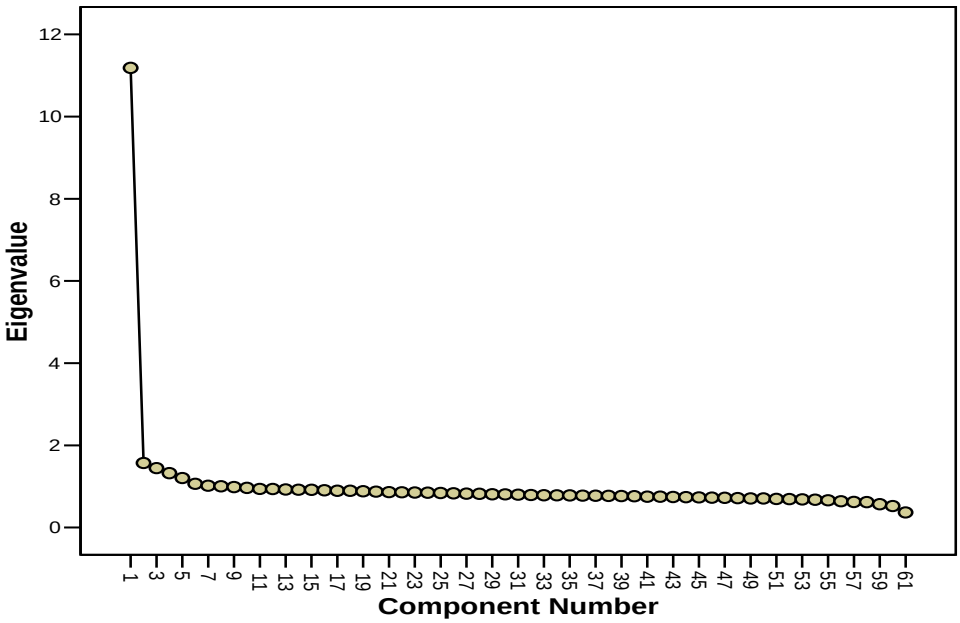


Figure 9. 20: Scree Plot Communication Arts, Grade 8

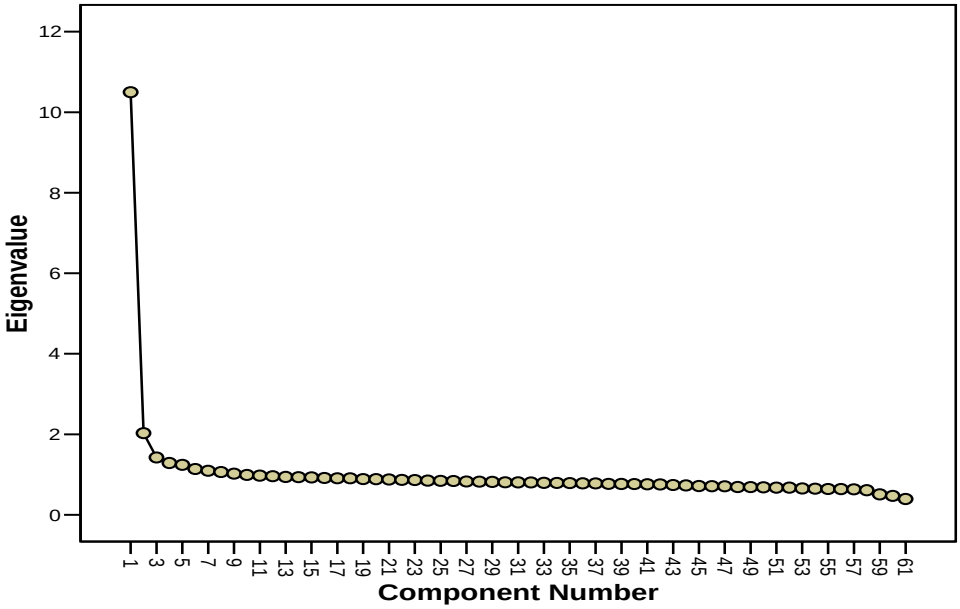


Figure 9. 21: Scree Plot Mathematics, Grade 3

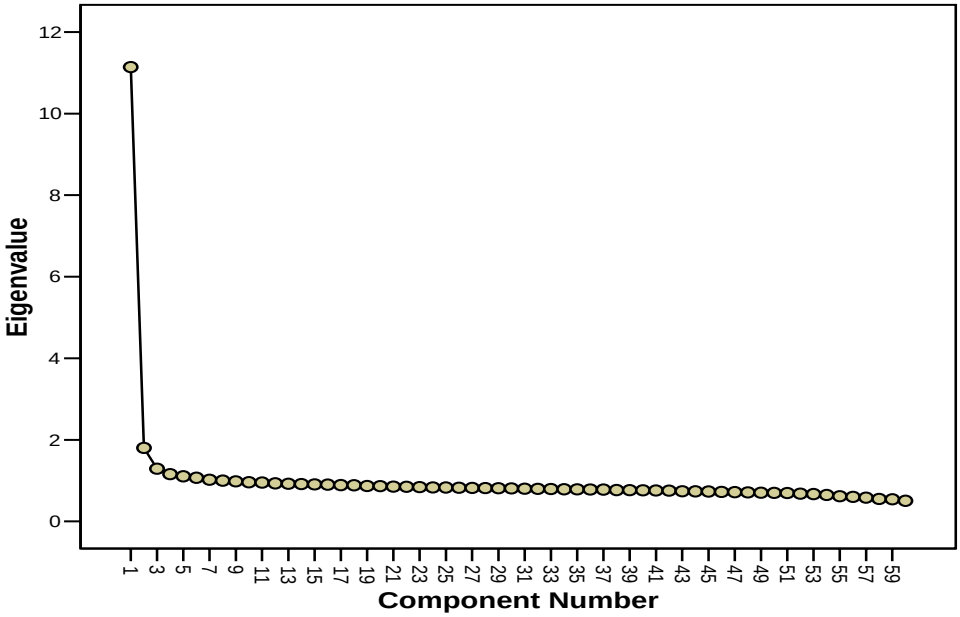


Figure 9. 22: Scree Plot Mathematics, Grade 4

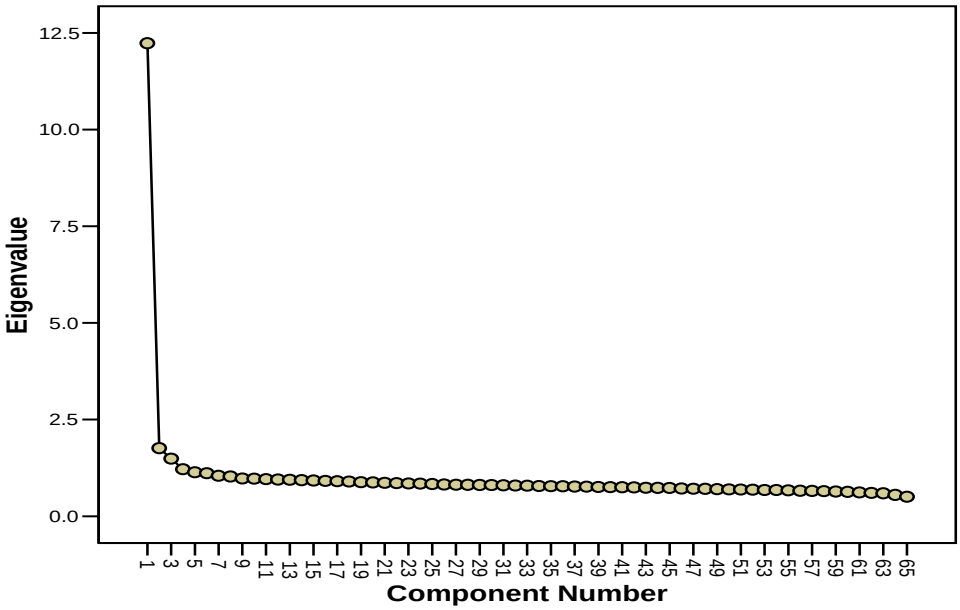


Figure 9. 23: Scree Plot Mathematics, Grade 5

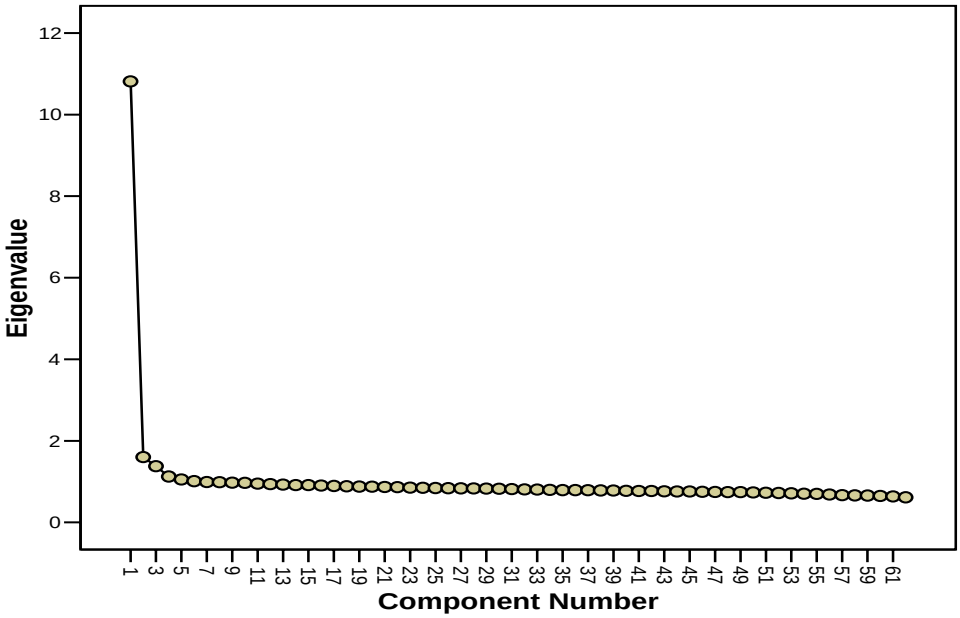


Figure 9. 24: Scree Plot Mathematics, Grade 6

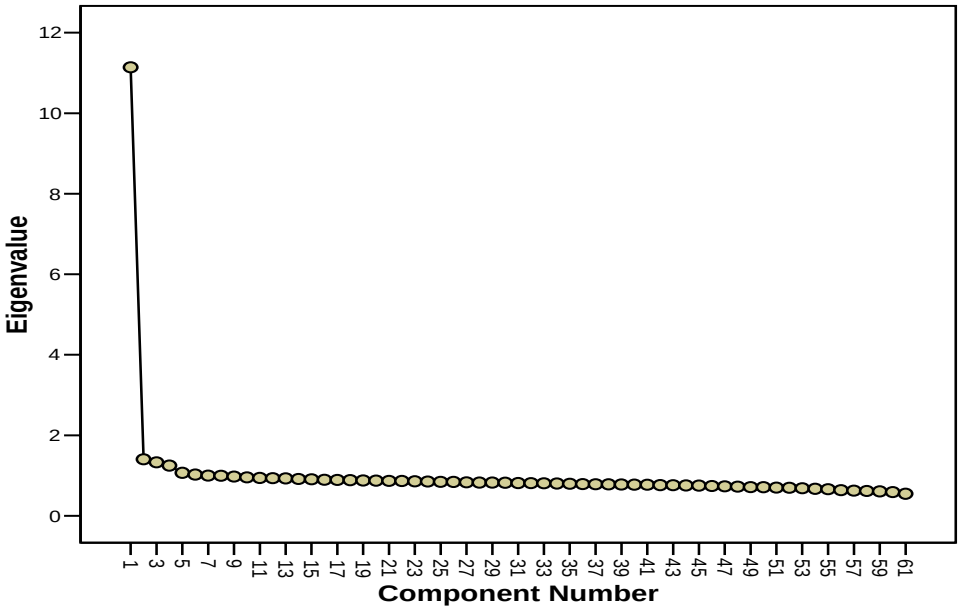


Figure 9. 25: Scree Plot Mathematics, Grade 7

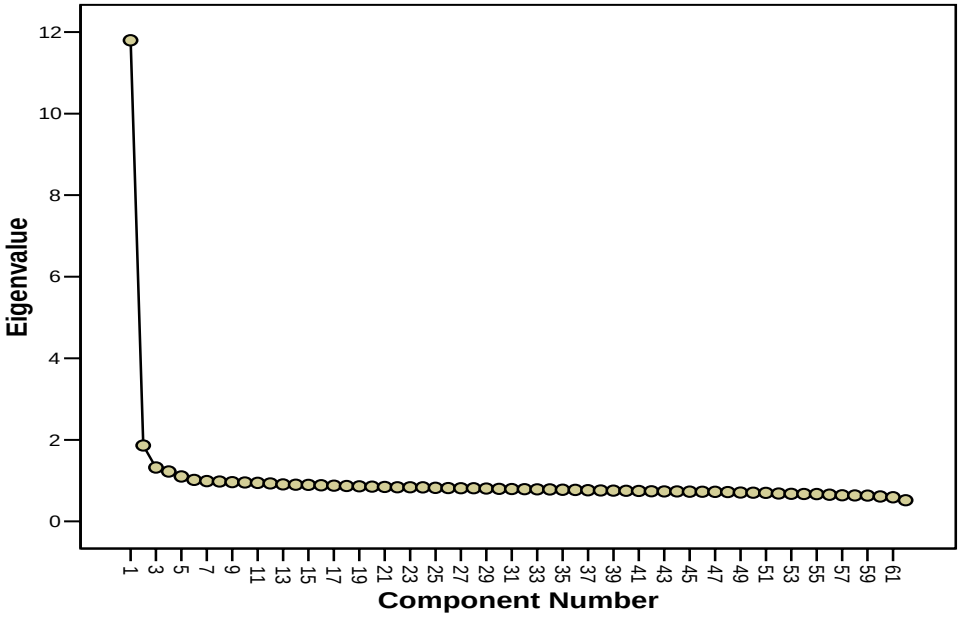


Figure 9. 26: Scree Plot Mathematics, Grade 8

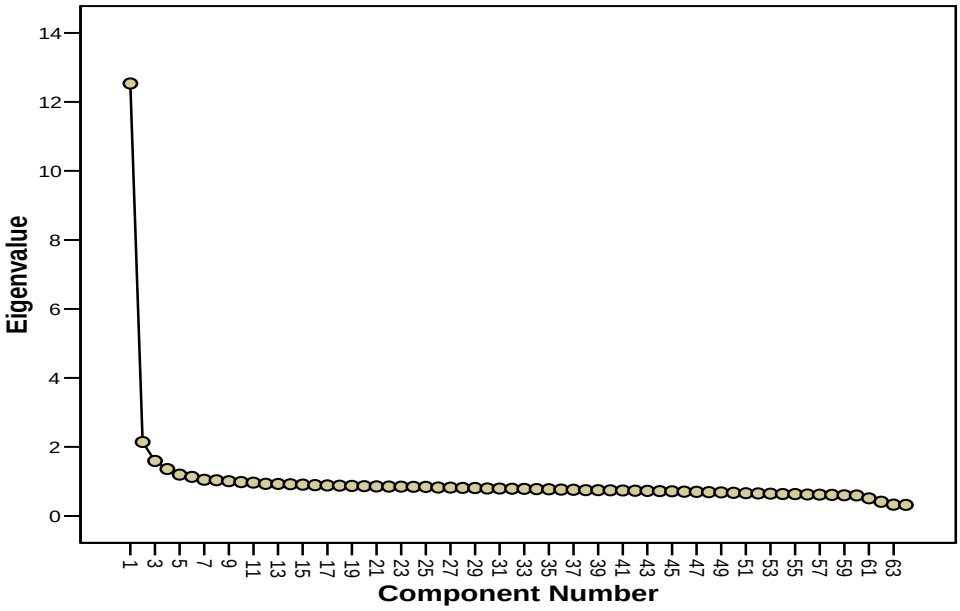


Figure 9. 27: Scree Plot Science, Grade 5

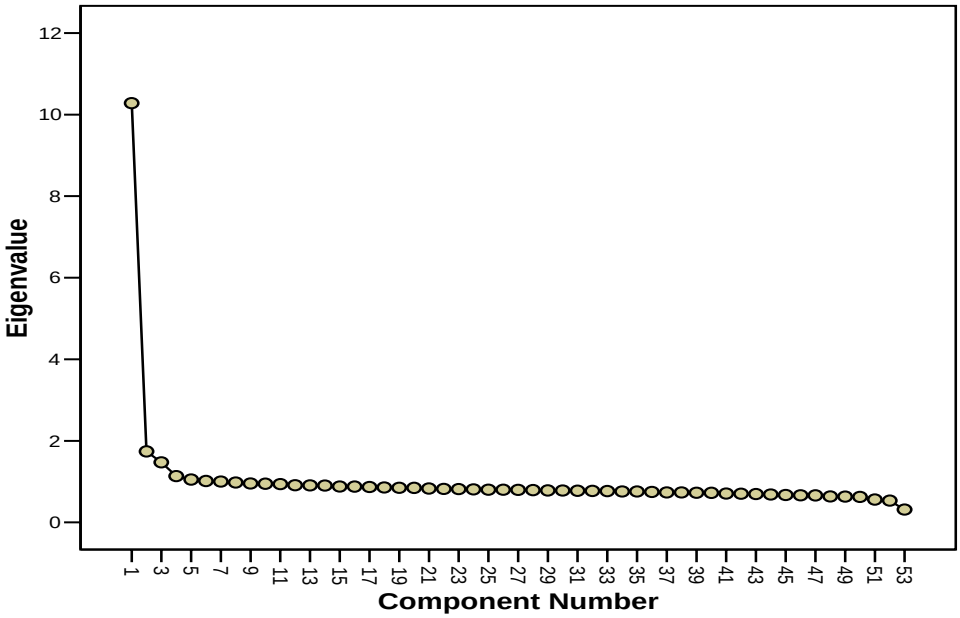
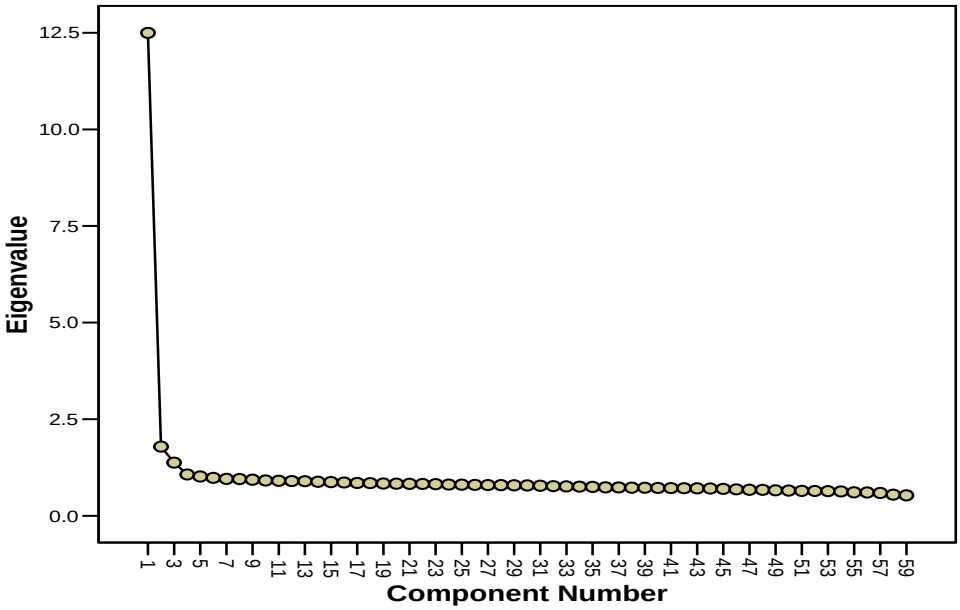


Figure 9. 28: Scree Plot Science, Grade 8



CHAPTER 10: FAIRNESS

As noted in the Standards (AERA, APA, & NCME, 1999), there are varying definitions of fairness. In this chapter, we examine fairness as it relates to minimizing bias on a test. We then look at test performance among varying subgroups assessed by MAP. It should be noted that differences in test performance among subgroups does not mean that a test is unfair—it simply means that groups perform differentially on the test. Even when a test is carefully and properly constructed, differences may exist among subgroups as a result of differences in curriculum or learning by students in the subgroup.

This chapter is particularly relevant to AERA, APA, & NCME (1999) Standards 7.1, 7.2, 7.3, and 7.4. Standards 7.1 through 7.4 are from Chapter 7 of the AERA, APA, & NCME (1999) Standards, which is titled "Fairness in Testing and Test Use." Each of these Standards will be presented as will the way the Standard is addressed in this chapter.

Standard 7.1 *When credible research reports that test scores differ in meaning across examinee subgroups for the type of test in question, then to the extent feasible, the same forms of validity evidence collected for the examinee population as a whole should also be collected for each relevant subgroup. Subgroups may be found to differ with respect to appropriateness of test content, internal structure of test responses, the relation of test scores to other variables, or the response processes employed by individual examinees. Any such findings should receive due consideration in the interpretation and use of scores as well as in the subsequent test revisions.*

There is no particular research on MAP showing that the test scores of examinee subgroups differ in meaning; however, this is an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, CTB has several steps that are followed in item development and selections as is explicated in Section 10.1 of this chapter. Also, DESE conducts content and bias reviews on items as is explained in Chapter 3.

Standard 7.2 *When credible research reports differences in the effects of construct-irrelevant variance across subgroups of test takers on performance on some part of the test, the test should be used if at all only for those subgroups for which evidence indicates that valid inferences can be drawn from test scores.*

Again, there is no research on MAP showing differences in the effects of construct-irrelevant variance across subgroups; however, DESE and CTB undertake steps to minimize construct-irrelevant variance through the test development process outlined in Section 10.1 of this chapter and explained in detail in Chapter 3.

Standard 7.3 *When credible research reports that differential item functioning exists across age, gender, racial/ethnic, cultural, disability, and/or linguistic groups in the population of test takers in the content domain measured by the test, test developers should conduct appropriate studies when feasible. Such research should seek to detect*

and eliminate aspects of test design, content, and format that might bias test scores for particular groups.

CTB conducts DIF studies following the field test and the operational administration of MAP. During the field-test phase of the project, items flagged for DIF will be further examined for possible bias. Items flagged for bias will be removed from the item pool. Section 10.2 of this chapter explains the steps taken to evaluate MAP items through the use of DIF.

Standard 7.4 *Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.*

Section 10.1 of this chapter is directly relevant to Standard 7.4. In this section, we explain the steps taken by CTB to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Section 3.2.5 of Chapter 3 discusses the Content and Bias Review conducted for MAP. This review is also critical in fulfilling Standard 7.4.

10.1 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 were used to minimize bias are summarized below.

First, careful attention was paid to content validity during the item development and item selection processes. 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. 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 comprised of Missouri participants, are described in more detail in Chapter 3 of this report.

10.2 Evaluating Bias through Differential Item Functioning Statistics

After administering the test, an empirical approach known as DIF was used to examine the items. The DIF statistics indicate the degree to which members of a particular subgroup performs 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 into the performances on which the measurement is based. If these tests require that students have culturally-specific knowledge and skills not taught in school, differences in performance among students can occur because of differences in student background and out-of-school learning. 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 Chapter 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 (MH) statistic (1959) and the Standardized Mean Difference (SMD) between the reference and focal groups, proposed by Dorans and Schmitt (1991).

The MH statistic is computed as (Zwick, Donoghue, & Grima, 1993):

$$\text{Mantel } \chi^2 = \frac{\left(\sum_k F_k - \sum_k E(F_k)\right)^2}{\sum_k \text{Var}(F_k)},$$

where F_k is the sum of scores for the focal group at the kth level of the matching variable. Note that the MH statistic is sensitive to N such that larger sample sizes increase the value of chi square.

In addition to the MH chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. Educational Testing Service (ETS) first developed the MH-D DIF statistic. To compute delta, alpha (the odds ratio) is first computed as:

$$\alpha_{MH} = \frac{\sum_{k=1}^{K} N_{r1k} N_{f0k} / N_k}{\sum_{k=1}^{K} N_{f1k} N_{r0k} / N_k},$$

where N_{r1k} is the number of correct responses in the reference group at ability level *k*, N_{f0k} is the number of incorrect responses in the focal group at ability level *k*, N_k is the total number of responses, N_{f1k} is the number of correct responses in the focal group at ability level *k*, and N_{r0k} is the number of incorrect responses in the reference group at ability level *k*. MH-D DIF is then computed as:

$$\text{MH-D DIF} = -2.35 \ln(\alpha_{MH}).$$

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

- Moderate DIF: Significant MH chi-square statistic ($p < 0.05$) and |MH D-DIF| ≤ 1.5
- Large DIF: Significant Mantel-Haenszel chi-square statistic ($p<0.05$) and |MH D-DIF|>1.5

For constructed-response items, an effect size (ES) statistic based on the MH chi-square will be used. The ES is obtained by dividing the SMD statistics by the standard deviation of the item. The SMD is an effect size index of DIF, which is relatively easy to interpret (Zwick et al., 1993). The SMD compares the mean of the reference and focal group, adjusting for the distribution of reference and focal group members on the conditioning variable (Zwick et al., 1993), which for these analyses is the MAP raw score. SMD is computed as (Zwick et al., 1993):

$$SMD = p_{Fk}(\sum_{k} m_{Fk} - \sum_{k} m_{Rk}) ,$$

where p_{fk} = proportion of the focal group members at the kth level of the matching variable, $m_{Fk} = 1/N_{F1k}$ and $m_{Rk} = 1/N_{R1k}$. Items are flagged using the same rules that are used in NAEP:

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

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group. Tables 10.1, 10.2 and 10.3 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.

- ***Accommodations***: Focal group is students who received one or more testing accommodations; 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.

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

Tables 10.1, 10.2, and 10.3 summarize the number of DIF flags by grade for each focal group. They also show the number of items on each test as well as the sample size of each subgroup. For example, in Grade 5 Mathematics, there was one item flagged for DIF for the accommodated subgroup. In this case, the flagged item exhibited moderate negative DIF. Three items were flagged for DIF for the female subgroup: one item exhibited moderate negative DIF while the two exhibited moderate positive DIF. Two items were flagged for moderate negative DIF against the Black subgroup. Finally, four

items were flagged for DIF for Asian/Pacific Islanders, one exhibiting moderate negative DIF and the others exhibiting moderate positive DIF.

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

10.3 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 10.4 through 10.9 present the scale score means and standard deviations, numbers of students, effect size (Cohen's D), and test form reliability statistics (Coefficient Alpha, see Chapter 9) for various subgroups of interest.

10.3.1 Reliability

Tables 10.4 through 10.9 show the test reliability for the various subgroups of interest. This analysis shows that the test reliability is of acceptable magnitude for all of the subgroups.

10.3.2 Effect Size

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

$$d = \frac{\overline{x_a} - \overline{x_b}}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}},$$

where $\overline{x_a}$ is the mean score of group A, $\overline{x_b}$ is the mean score of group B, s_a^2 is the variance of group A, s_b^2 is the variance of group B, n_a is the number of students in group A, and n_b is the number of students in group B.

Cohen's *d,* then, expresses the difference in group means in terms of the standard deviation. For example if *d*=.34 for two groups, then it may be interpreted that the mean difference between the two groups is .34 of the pooled standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the *d* statistic: *d*=.20 is a small effect size, *d*=.50 is a medium effect size, and *d*=.80 is a large effect size.

Using Cohen's (1988) guidelines, certain trends become apparent in Tables 10.4 through 10.9. On the Communication Arts test in all grades, gender has a slight relationship with mean test scores where girls outperform boys. On the Communication Arts, Mathematics, and Science tests in all grades, accommodations tend to have a large relationship with the mean test scores where students in the accommodated groups underperform students who are not in those groups.

In terms of the race/ethnicity in all grades, there is a moderate difference in mean Communication Arts test scores of Black students compared to White students, where Black students underperform white students on average. There is a slight relationship between mean test scores and race/ethnicity where Hispanics underperform White students on the Communication Arts tests. There is a small effect size for the mean test scores where Native Americans underperform White students on Communication Arts in Grades 3, 4, and 8. There is a small difference where Asian/Pacific Islander students outperform White students in Grades 5 and 6 Communication Arts.

There is a strong relationship between the mean Mathematics tests scores and race/ethnicity, where Black students underperformed White students in all grades, except Grade 3 where there is a moderate effect size. There is a small difference in mean Mathematics test scores of Hispanic students compared to White students in Grades 3 through 8, where Hispanic students underperform White students. There is a small effect size for mean test scores of Native American students compared to White students, where Native American students underperform White students in all grades except Grade 6. Finally, there is a small effect size for mean test scores of Asian/Pacific Islander students, where Asian/Pacific Islander students outperform White students in Grades 3 through 8 in Mathematics.

There is a large effect size for mean Science test scores of Black students compared to White students in Grades 5 and 8 where Black students underperform White students. There is a moderate effect size for mean Science test scores of Hispanic students compared to White students in Grades 5 and 8 where Hispanic students underperform White students. There is a small effect size for mean Science test scores of Native American students compared to White students in Grades 5 and 8 where Native American students underperform White students.

10.4 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of MAP. The information in this chapter addresses multiple best practices of the testing industry, and in particular are related to the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME,1999):

- Standard 7.1—When credible research reports that test scores differ in meaning across examinee subgroups for the type of test in question, then to the extent feasible, the same forms of validity evidence collected for the examinee population as a whole should also be collected for each relevant subgroup.

Subgroups may be found to differ with respect to appropriateness of test content, internal structure of test responses, the relation of test scores to other variables, or the response processes employed by individual examinees. Any such findings should receive due consideration in the interpretation and use of scores, as well as in the subsequent test revisions.

- Standard 7.2—When credible research reports differences in the effects of construct-irrelevant variance across subgroups of test takers on performance on some part of the test, the test should be used if at all only for those subgroups for which evidence indicates that valid inferences can be drawn from test scores.
- Standard 7.3—When credible research reports that differential item functioning exists across age, gender, racial/ethnic, cultural, disability, and/or linguistic groups in the population of test takers in the content domain measured by the test, test developers should conduct appropriate studies when feasible. Such research should seek to detect and eliminate aspects of test design, content, and format that might bias test scores for particular groups.
- Standard 7.4—Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.

Table 10. 1: 2009 MAP DIF Statistics: Number of Flagged Items, Communication Arts

Grade	Group	Sample Size	Large Negative	Moderate Negative	Moderate Positive	Large Positive	Number of Items
3	Accommodated	5580	1	3	1		57
	Asian/Pacific Islander	1252		3	2	2	57
	Native American/Alaskan	233				1	57
	Black (not Hispanic)	10668		1		1	57
	Hispanic	2606					57
	Female	28838			1		57
4	Accommodated	5884					55
	Asian/Pacific Islander	1180		5	1	1	55
	Native American/Alaskan	270		1			55
	Black (not Hispanic)	10650					55
	Hispanic	2499					55
	Female	29218					55
5	Accommodated	6099		1			55
	Asian/Pacific Islander	1085		2	2		55
	Native American/Alaskan	238				1	55
	Black (not Hispanic)	10344	1	1	1		55
	Hispanic	2245	1				55
	Female	27096			2		55
6	Accommodated	7048	1	1			55
	Asian/Pacific Islander	1220		2		1	55
	Native American/Alaskan	292					55
	Black (not Hispanic)	11564		1			55
	Hispanic	2459					55
	Female	32211	1	2	2		55
7	Accommodated	6462	2	2			61
	Asian/Pacific Islander	1155		6	1	1	61
	Native American/Alaskan	273				1	61
	Black (not Hispanic)	11062			2	1	61
	Hispanic	2204		2			61
	Female	30759		4		1	61
8	Accommodated	6406					61
	Asian/Pacific Islander	1192	2	8	3		61
	Native American/Alaskan	283					61
	Black (not Hispanic)	11090	1	1	2		61
	Hispanic	2158	1	1			61
	Female	31160	2	3	1	1	61

Table 10. 2: 2009 MAP DIF Statistics: Number of Flagged Items, Mathematics

Grade	Group	Sample Size	Large Negative	Moderate Negative	Moderate Positive	Large Positive	Number of Items
3	Accommodated	5842	1	2	1		60
	Asian/Pacific Islander	1292	3	1	4		60
	Native American/Alaskan	235		2		1	60
	Black (not Hispanic)	10678		3			60
	Hispanic	2638		1			60
	Female	28892		1			60
4	Accommodated	6149		2			65
	Asian/Pacific Islander	1231	3	4	2	1	65
	Native American/Alaskan	270					65
	Black (not Hispanic)	10679		5	2		65
	Hispanic	2524		1			65
	Female	29269		2			65
5	Accommodated	6392		1			62
	Asian/Pacific Islander	1114		1	3		62
	Native American/Alaskan	238					62
	Black (not Hispanic)	10349		2			62
	Hispanic	2267					62
	Female	27140		1	2		62
6	Accommodated	7328			2	1	61
	Asian/Pacific Islander	1249					61
	Native American/Alaskan	293			1		61
	Black (not Hispanic)	11575			1		61
	Hispanic	2476					61
	Female	32234	1	1	1		61
7	Accommodated	6683		2	1	1	62
	Asian/Pacific Islander	1180	1	2			62
	Native American/Alaskan	271					62
	Black (not Hispanic)	11058	2	1	1		62
	Hispanic	2230		1			62
	Female	30790	2	4	1	1	62
8	Accommodated	6705			2	1	64
	Asian/Pacific Islander	1215	2		2		64
	Native American/Alaskan	282					64
	Black (not Hispanic)	11109	2				64
	Hispanic	2184		1			64
	Female	31186		2	1		64

Table 10. 3: 2009 MAP DIF Statistics: Number of Flagged Items, Science

Grade	Group	Sample Size	Large Negative	Moderate Negative	Moderate Positive	Large Positive	Number of Items
5	Accommodated	6111		1			53
	Asian/Pacific Islander	1114	1		1		53
	Native American/Alaskan	237					53
	Black (not Hispanic)	10340		1	1		53
	Hispanic	2265		3			53
	Female	27129	1	2	4		53
8	Accommodated	6538		1			59
	Asian/Pacific Islander	1212		3	1	2	59
	Native American/Alaskan	281		2			59
	Black (not Hispanic)	11064	1		2		59
	Hispanic	2175		1	3		59
	Female	31158		4	3		59

Table 10. 4: Impact Analysis, Grade 3

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	50411	642.58	36.08		0.89
		Black (not Hispanic)	12138	617.94	39.64	0.67	0.91
		Hispanic	2924	625.51	36.83	0.47	0.90
		Asian/Pacific Islander	1341	648.05	38.68	-0.15	0.90
		Native American	271	634.49	40.96	0.22	0.92
	Gender	Male	34461	632.91	38.94		0.91
		Female	32633	642.25	36.71	-0.25	0.90
	Accommo-dations	No	60752	641.92	34.30		0.89
		Yes	6411	594.89	46.02	1.32	0.91
Mathematics	Ethnicity	White (not Hispanic)	50410	627.03	34.34		0.90
		Black (not Hispanic)	12139	600.21	37.76	0.77	0.92
		Hispanic	2953	612.01	34.92	0.44	0.91
		Asian/Pacific Islander	1379	637.00	41.12	-0.29	0.92
		Native American	273	619.13	34.81	0.23	0.92
	Gender	Male	34501	621.76	37.92		0.92
		Female	32663	621.62	35.46	0.00	0.91
	Accommo-dations	No	60622	625.24	34.82		0.91
		Yes	6610	588.93	37.87	1.03	0.92

Table 10. 5: Impact Analysis, Grade 4

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	50182	660.88	31.63		0.91
		Black (not Hispanic)	11965	641.09	35.46	0.61	0.92
		Hispanic	2738	646.64	33.26	0.45	0.92
		Asian/Pacific Islander	1244	666.10	32.75	-0.16	0.95
		Native American	297	652.57	32.59	0.26	0.91
	Gender	Male	34054	651.51	34.05		0.92
		Female	32370	662.33	31.77	-0.33	0.91
	Accommo-dations	No	59805	661.07	29.38		0.90
		Yes	6685	618.32	41.69	1.39	0.92
Mathematics	Ethnicity	White (not Hispanic)	50191	649.39	31.56		0.91
		Black (not Hispanic)	11982	622.65	34.60	0.83	0.93
		Hispanic	2759	636.87	30.91	0.40	0.91
		Asian/Pacific Islander	1293	659.77	35.77	-0.33	0.92
		Native American	297	640.12	34.35	0.29	0.92
	Gender	Male	34122	644.48	34.98		0.93
		Female	32399	643.94	32.69	0.02	0.92
	Accommo-dations	No	59678	647.57	31.87		0.92
		Yes	6909	615.02	36.70	1.00	0.92

Table 10. 6: Impact Analysis, Grade 5

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	50657	675.85	30.93		0.91
		Black (not Hispanic)	12113	654.77	34.60	0.67	0.91
		Hispanic	2672	663.08	31.89	0.41	0.91
		Asian/Pacific Islander	1234	682.21	34.71	-0.21	0.92
		Native American	283	670.35	26.68	0.18	0.90
	Gender	Male	34415	668.40	34.06		0.92
		Female	32545	675.02	31.07	-0.20	0.91
	Accommo-dations	No	59718	676.15	28.40		0.90
		Yes	7365	634.57	41.72	1.38	0.91
Mathematics	Ethnicity	White (not Hispanic)	50671	668.23	37.61		0.90
		Black (not Hispanic)	12118	636.92	42.13	0.81	0.91
		Hispanic	2696	651.92	36.92	0.43	0.90
		Asian/Pacific Islander	1264	680.97	43.78	-0.34	0.92
		Native American	283	658.43	33.53	0.26	0.89
	Gender	Male	34444	661.09	41.57		0.92
		Female	32588	663.20	39.29	-0.05	0.91
	Accommo-dations	No	59498	667.18	37.00		0.90
		Yes	7657	622.37	44.50	1.18	0.91
Science	Ethnicity	White (not Hispanic)	50651	668.14	26.83		0.90
		Black (not Hispanic)	12105	639.24	32.52	1.03	0.91
		Hispanic	2692	652.11	29.47	0.59	0.90
		Asian/Pacific Islander	1265	669.21	34.47	-0.04	0.93
		Native American	282	660.56	27.70	0.28	0.89
	Gender	Male	34423	663.45	31.37		0.91
		Female	32573	661.00	29.23	0.08	0.91
	Accommo-dations	No	59786	665.35	28.20		0.90
		Yes	7332	636.71	35.27	0.99	0.91

Table 10. 7: Impact Analysis, Grade 6

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	50138	675.69	31.38		0.90
		Black (not Hispanic)	11567	654.89	34.28	0.65	0.90
		Hispanic	2460	663.83	31.80	0.38	0.90
		Asian/Pacific Islander	1213	682.67	34.61	-0.22	0.91
		Native American	291	671.68	35.47	0.13	0.91
	Gender	Male	33450	667.19	34.68		0.91
		Female	32208	676.38	30.52	-0.28	0.89
	Accommo-dations	No	58611	676.56	28.31		0.88
		Yes	7105	631.36	40.74	1.51	0.89
Mathematics	Ethnicity	White (not Hispanic)	50132	684.75	36.65		0.91
		Black (not Hispanic)	11566	653.74	40.70	0.83	0.91
		Hispanic	2475	669.14	37.49	0.43	0.91
		Asian/Pacific Islander	1242	696.64	46.09	-0.32	0.93
		Native American	292	678.37	40.53	0.17	0.92
	Gender	Male	33468	678.05	40.74		0.92
		Female	32227	679.79	38.25	-0.04	0.91
	Accommo-dations	No	58420	683.82	36.46		0.90
		Yes	7335	639.48	41.17	1.20	0.91

Table 10. 8: Impact Analysis, Grade 7

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	50989	681.92	33.10		0.91
		Black (not Hispanic)	11486	659.83	35.57	0.66	0.92
		Hispanic	2296	668.96	32.53	0.39	0.91
		Asian/Pacific Islander	1207	686.37	39.18	-0.13	0.93
		Native American	287	676.67	36.49	0.16	0.92
	Gender	Male	33806	672.47	35.76		0.92
		Female	32457	683.15	32.73	-0.31	0.91
	Accommo-dations	No	59458	682.81	30.14		0.90
		Yes	6858	633.17	39.91	1.59	0.90
Mathematics	Ethnicity	White (not Hispanic)	50976	689.83	37.93		0.92
		Black (not Hispanic)	11473	656.62	40.77	0.86	0.91
		Hispanic	2316	672.91	37.36	0.45	0.91
		Asian/Pacific Islander	1232	701.22	46.95	-0.30	0.94
		Native American	285	680.34	37.62	0.25	0.91
	Gender	Male	33803	683.56	42.09		0.93
		Female	32476	683.76	39.20	0.00	0.92
	Accommo-dations	No	59301	688.96	36.94		0.92
		Yes	7029	638.69	43.33	1.33	0.90

Table 10. 9: Impact Analysis, Grade 8

Content Area	Category	Group	N	Mean	Std. Dev.	Effect Size	Coefficient Alpha
Communication Arts	Ethnicity	White (not Hispanic)	51012	697.07	31.24		0.90
		Black (not Hispanic)	11778	674.23	34.74	0.72	0.90
		Hispanic	2317	682.72	33.29	0.46	0.91
		Asian/Pacific Islander	1235	701.56	36.17	-0.14	0.92
		Native American	293	687.97	37.27	0.29	0.92
	Gender	Male	34325	687.51	34.74		0.91
		Female	32317	697.96	30.80	-0.32	0.90
	Accommo-dations	No	60001	697.40	28.82		0.89
		Yes	6740	649.42	39.06	1.60	0.89
Mathematics	Ethnicity	White (not Hispanic)	50998	709.70	35.57		0.92
		Black (not Hispanic)	11779	677.59	39.51	0.88	0.91
		Hispanic	2338	693.32	37.40	0.46	0.92
		Asian/Pacific Islander	1257	722.91	40.75	-0.37	0.94
		Native American	292	696.00	38.90	0.38	0.93
	Gender	Male	34345	702.97	40.55		0.93
		Female	32327	704.34	36.41	-0.04	0.92
	Accommo-dations	No	59783	708.63	34.84		0.92
		Yes	6987	660.57	42.48	1.35	0.89
Science	Ethnicity	White (not Hispanic)	50967	701.43	27.33		0.92
		Black (not Hispanic)	11753	671.78	33.22	1.04	0.92
		Hispanic	2329	686.20	30.83	0.55	0.93
		Asian/Pacific Islander	1254	704.54	33.50	-0.11	0.94
		Native American	292	691.40	32.14	0.37	0.93
	Gender	Male	34297	695.77	32.28		0.93
		Female	32306	695.57	29.42	0.01	0.93
	Accommo-dations	No	59888	699.41	27.96		0.92
		Yes	6814	662.58	35.87	1.28	0.92

References

Altman, D.G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

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.

Assessment Resource Center (2008). *MAP and Missouri Schools: A Consequential Validity Study*. Columbia, MO: Author.

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. Thousand Oaks, CA:Sage Publication.

Candell, G.L. & Drasgow, F. (1988). An iterative procedure for linking metrics bias in item response theory. *Applied Psychological Measurement*, 12(3), 253–260.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

CTB/McGraw-Hill (2003). *TerraNova The 2nd Edition: CAT Technical Report.* Monterey, CA: Author.

CTB/McGraw-Hill (2005). *Missouri Assessment Program Final Bookmark Standard Setting Technical Report.* Monterey, CA: Author.

CTB/McGraw-Hill (2008). *Missouri Assessment Program Bookmark Standard Setting Technical Report 2008 for Missouri Achievement-Level Setting Grades 5, 8, and 11 Science.* Monterey, CA: Author.

CTB/McGraw-Hill (2009). *Guide to Interpreting Results.* Monterey, CA: Author.

Dorans, N.J., & Schmitt, M.P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. Princeton: Educational Testing Service.

CTB/McGraw-Hill (2009). *TerraNova 3rd Edition Technical Addendum: Forms E and F*. Monterey, CA: Author.

Green, D.R. (1975). Procedures for assessing bias in achievement tests. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.

Karkee, T. & Choi, S. (2005). Impact of eliminating anchor items flagged from statistical criteria on test score classifications in common item equating. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 15, 2005.

Kim, D. (2005). KKCLASS [Computer program]. Unpublished.

Kim, D., Barton, K., & Kim, J. (2007). Estimating classification consistency and classification accuracy with pattern scoring. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kim, D., Choi, S., Um, K., & Kim, J. (2006). A comparison of methods for estimating classification consistency. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, CA.

Kolen, M. J. & Brennan, R.L. (1995). *Test Equating: Methods and Practices.* New York: Springer-Verlag.

Kolen, M. J. & Kim, D. (2005). Personal correspondence.

Landis, J.R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.

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

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Macmillan/McGraw-Hill. (1993a). *Guidelines for bias-free publishing*. New York, NY: Author.

Macmillan/McGraw-Hill. (1993b). *Reflecting diversity: Multicultural guidelines for educational publishing professionals*. New York, NY: Author.

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.

Michaelides, M.P., & Haertel, E.H. (2004). Sampling of common items: An unrecognized source of error in test equating. Los Angeles, CA: Center for the Study of Evaluation.

Schumacker, R.E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions,* 10, 479.

Swaminathan, H., Hambleton, R.K., & Algina, J. (1974). Reliability of criterion-reference tests: A decision-theoretic formulation. *Journal of Educational Measurement,* 11(4), 263-267.

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.

Thompson, S. & Thurlow, M. (2002). Universally designed assessments: Better tests for everyone! (Policy Directions. No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [November 9, 2009], from: http://www.cehd.umn.edu/NCEO/OnlinePUbs/Policy14.htm

Voelkle, M., Schwarz, R., Arenson, E., & Ito, K. (2002). An investigation of factors affecting Stocking & Lord equating. (Paper in progress.)

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement,* 5, 245–262.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement,* 30, 187–213.

Yen, W.M. & Candell, G.L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education,* 4, 209-228.

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

Appendix A

DESE Presentation on Test Coordinator's Manual

2009
Test Coordinator Manual Training

1

Missouri Assessment Program
Grade-Level Assessments 2009

Communication Arts.....Grades 3-8
Math.....Grades 3-8
Science.....Grades 5 and 8

2

2009 MAP Grade-Level Assessments' Test Coordinator's Manual (TCM)

The TCM is primarily focused on the responsibilities of:

- District Test Coordinators (DTC's)
- School Test Coordinators (STC's)

The TCM also contains:

- An Overview
- Timelines
- A Glossary of Terms

3

Dates to Remember:

Arrival of Materials

Feb. 20-27, 2009.....Test Coordinator packages to districts

March 13th, 2009.....Test Materials arrive in districts

Testing Window

March 30-April 24, 2009.....Administer Assessments

Testing Materials

April 10, 2009.....Deadline for ordering additional testing materials without incurring additional shipping costs

NOTE: If testing materials are ordered after April 10, the district will be responsible for paying the shipping costs.

April 20, 2009.....Final deadline for ordering additional testing materials

Return of Materials

April 27, 2009.....Deadline to contact CTB/McGraw-Hill for pickup of testing materials

April 30, 2009.....Deadline for testing materials to be picked up by CTB's transportation vendor

Test Results

August-September 2009.....Reports shipped to districts

4

DTC's Responsibilities:

- Checks, reviews and distributes testing materials to the STC's
- Trains STCs on testing processes
- Assumes STC role when necessary
- Guarantees security for all testing materials
- Acts as sole channel for all communication between districts and CTB Service Line
- Collects all testing materials after the test and returns to CTB/McGraw-Hill

5

STC's Responsibilities

- Receives and checks all testing materials from the DTC
- Assumes DTC role when necessary
- Guarantees security of all testing materials
- Disseminates Examiner's Manuals
- Trains all Examiners
- Checks Group Information Sheets (GISs)
- Completes School/Group Lists
- Collects all testing materials after testing, checks and organizes materials for return to the DTC

6

Examiner's Responsibility

- Examiner primary responsibilities are addressed in the 2009 Test Examiner's Manual provided for each grade

7

Summary of 2009 Changes

The MAP tests will begin being conversationally referred to as MAP Grade-Level Assessments. However, no written name change will appear until the 2010 manuals.

MAP Grade-Level Assessments will no longer be administered at Grades 10 and 11. Those assessments are replaced by the MAP End-of-Course Assessments.

A new bubble has been added to the Student Information Sheet (SIS) called, "Not Enrolled in Building".

Guidelines have been added for how to handle test books of students who move during test administration.

8

Summary of 2009 Changes (contd.)

Guidelines for electronic equipment have been added.

Foreign exchange student guidelines have been added.

Guidelines for handling students who recognize literacy-based passages have been added.

The white shipping label for return of test books now reads: "UNUSED/DO NOT SCORE."

9

Summary of 2009 Changes (contd.)

The process for transcribing student responses from a Braille test book to the regular print version has changed. Instructions for this process are included in the test materials sent directly to Examiners.

10

NCLB

requires all Missouri students in grades 3 through 8 to take the MAP Grade-Level Assessment

Only two groups are exempt:

- Group 1: Alternate (MAP-A)
- Group 2: ELL students in the United States, 12 months or fewer, may be exempt from taking the Communication Arts test

11

Guidelines for Special Student Groups

- Accommodations for special populations can be found in the Test Examiners' Manuals
- The following are examples of special populations:
  - IEP students
  - IAP (504)
  - ELL
- Guidelines for testing other students can be found in the TCM
  - Out-of-District students
  - Homebound students
  - Home-schooled students
  - MoVIP students

12

Students Testing Out-of-District

- The home DTC (where the student is enrolled) delivers the testing materials to the serving districts/agencies
- After testing, the completed materials are returned to the home DTC
- The GIS determines where students' results will be reported

13

Homebound Students:

- Must take the test if they are receiving homebound services
- Must have test administered by a trained Examiner who guarantees the security of testing materials
- May be tested either at home or school

14

Home-School Students

- May take part with the local district at the district's discretion
- If receiving services in a subject that is assessed, the student must be tested at the local school. The student will only be tested in the subject area in which they are receiving instruction

15

Student Make-up Sessions

IF	THEN
Absent during one or more sessions --	Schedule make-up session
Student is unable to test during make-up session --	Follow Student Absences Procedures

16

Student Absences

- Write student's name on front of the unused/or incomplete test book
- Affix student barcode label if accurate
- Complete SIS if student barcode label is inaccurate
- Code SIS for absences
- Return test book with scorable books

17

Large Print/Braille Procedures

- Must be transcribed to a regular edition test book to receive score
- Large Print/Braille editions must be labeled, "**Contents transcribed to a regular test book** -- **DO NOT SCORE**"
- Must use special handling and packaging instructions that come with the Large Print/Braille editions
- Must return Large Print/Braille editions with unused testing materials

18

ELL Students

For Mathematics and Science Assessments only:

- ELL students can use their native language to give oral or written responses to assessments
- ELL students' responses must be translated into English and scribed verbatim into a regular test book
- Translators must be trained in administering the MAP Grade-Level tests
- Translators can review tests before administration in a secure environment
- Translators must guarantee security of testing materials
- ELL accommodation codes apply (see Examiner's Manual codes)

19

Invalidations

Six categories for test invalidation:

1. Student discovered cheating
2. Examiner reads any part of the C.A. test to student(s) other than students who are Blind/Visually impaired who do not read Braille
3. Examiner signs any part of the C.A. test
4. Examiner paraphrases test questions in any content area
5. Examiner reads any part of the C.A. test to a student in the student's native language
6. A student uses a bilingual dictionary for any part of the C.A. test

20

Invalidation Procedures

- The STC agrees with the Examiner regarding the test invalidation due to cheating
- STC completes the Teacher Invalidation grid on the SIS
- STC provides demographic information to the DTC
- STC includes invalidated test materials with the scorable testing materials
- DTC sends written communication to DESE

21

Check lists for DTC and STC

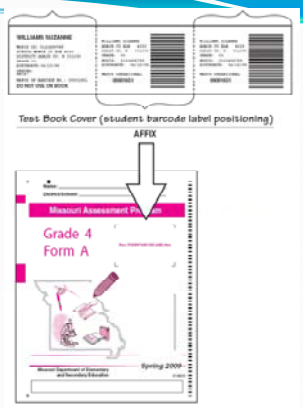
- Convenient check lists are provided to the DTC and STC on pages 8 and 9 of the TCM
- DTC and STC must ensure all functions are completed
- DTC and STC are responsible for both check lists if either of the responsible parties/roles is not available

22

Student Barcode Labels

- Student barcode labels are in the Test Coordinator's Kit
- Student barcode labels include:
  - Information from MOSIS pre-code file
  - Demographic information (but not all biographical data) from the pre-code file
- CTB/McGraw-Hill prints and provides:
  - One biographical master label
  - Two student barcode labels
- DTC and STC do the following:
  - Check student barcode label against the student pre-code roster for accuracy
  - Determine viability of labels
  - Handle barcode exceptions appropriately

23



24

Handling Student Barcode Labels

Error in:
Student Name, Birth Date, Gender, Race/Ethnicity and MOSIS ID
→ *Don't use barcode* ***bubble in all info on SIS***

If label is affixed and then found to be wrong
→ *Place two blank labels over the incorrect label and then bubble all information the SIS*

Wrong student label is affixed
→ *Place blank label over incorrect label* ***Then affix correct label***

25

Using Student Barcode Labels

No student barcode label
→ *Notify local student data management person to enroll/submit the student in core data.* ***Bubble in all info on SIS*** ***Leave barcode area blank!***

26

Step 1: Review Testing Materials

The District Test Coordinator's Kit includes two folders for **EACH** school

- One for the DTC
- One for the STC

27

DTC Folder

- District packing list
- District cover letter
- Materials Ordering and Inventory Information Flyer
- Test Book Accountability Form
- Add/Short Form – District
- Missing Test Materials Form
- Student Barcode Label Instruction Sheet
- Test Coordinator's Manuals
- Blank District Return Shipping Labels:
  - Blue for scorable materials
  - White for unused Inventory

28

STC Folder

- School packing list
- School cover letter
- Add/Short Form – School
- School Group Lists (SGL's)
- Group Information Sheet (GIS) for each teacher and grade
- Test Coordinator's Manuals
- Blank green stock labels
- Return Shipping Labels:
  - Blue for scorable materials
  - White for unused Inventory
- Student Barcode Label Instruction Sheet
- Student Pre-code Roster
- Blank Barcode Label Stock
- Student Barcode Labels

29

Testing Materials

- Packaged by school
- Shipped to the address designated in the online order
- Includes:
  - Examiner's Manuals
  - Test books
  - Large white envelops
  - Ancillary testing materials
  - Braille Cover Letter and Omit Return Instruction Sheet

30

Verify Shipment

- Compare the packing list materials against shipment
- Follow procedures for ordering more materials, if needed, using Add/Short Form

31

MAP Grade-Level Spring 2009
Ordering Additional Testing Materials

Start Date	End Date	Shipping Mode	Shipping Costs
March 13th	April 10th	UPS ground Service	CTB
April 13th	April 20th	Next-day or second-day service	District

32

Step 2: Distribute Testing Materials

DTC's responsibilities:

- Maintains security
- Tracks security barcodes to confirm start and end barcode numbers for each shrink-wrapped bundle
- Matches numbers with packing list
- Reports discrepancies to CTB

33

- **STC's Responsibilities**
  - Maintains testing security
  - Tracks security barcodes to confirm start and end barcode numbers for each shrink-wrapped bundle
  - Matches numbers with packing list
  - Maintains Test Book Accountability Form and documents discrepancies
  - Reports discrepancies to DTC

34

- **Examiner's Responsibilities**
  - Maintains testing security
  - Counts number of books received and keeps documentation in preparation for returning the test books to the STC
  - Reports any problems to the STC

35

Test Book Accountability Form

- Used to ensure 100% accountability of test books
- Deals with the security barcode numbering system
- **Each** STC in the district needs a copy
- Instructions for use outlined on page 14 - 15 in TCM
- Both DTC and STC have pretest and post-test responsibilities regarding the form
- Examiners **do not** get a copy of the form

36

Security Barcode

- Used for test book security
- Used for inventory – each book consecutively numbered
- Used to ensure 100% accountability of test books
- Used for missing inventory reports generated by CTB/McGraw-Hill
- Used by DESE to track barcode numbers, district and school name
- Located on lower right-hand corner directly above "Spring 2009"

37

Security SHALL NOT'S

- Testing materials **shall not** be photocopied, duplicated or made accessible to non-testing personnel
- Testing materials **shall not** be viewed by Examiners before testing
- Testing materials **shall not** be left in an unsecured area at any time, for any reason -- must be locked in a secure cabinet or in a secure room before, between and after testing sessions
- Testing materials **shall not** be discussed between Examiners
- Test books **shall not** be shared between schools

38

Test Administration Musts

1. **DTC** guarantees the security of the testing materials -- every school must have sufficient, satisfactory and locked security
2. DTC houses materials at the district office if the school's security is insufficient, unsatisfactory or unlocked
3. **DTC** distributes the materials to all **STC**'s in the district
4. **DTC** delivers appropriate testing materials for out-of-district students, prior to the first day of testing. Also, the DTC makes arrangements for returning materials after testing is complete
5. **STC** distributes the 2009 *Examiner's Manuals* to all who will administer the test as soon as possible
6. **STC** collects all student draft work and scratch, graph or grid paper and securely destroys after testing
7. **DTC** saves their folder and boxes for use after test administration

39

Materials Needed for Each Examiner

- *Examiner's Manual* for appropriate grade level
- Appropriate quantities of books, manipulatives and reference sheets
- Student barcode labels for each student
- Pre-coded Group Information Sheet (GIS):
  - One for each grade, 3-8
  - No separate GIS for Special Education students.
- Appropriate quantity of large white envelopes (each holds approximately 5-10 books)
- School/Group List for use after test administration

40

Step 3: Collect Testing Materials

- All used and unused books must be accounted for and returned
- One test book is returned for each student
- All manipulatives, reference sheets, scratch, draft, graph or grid paper, extra envelopes and contaminated test books are securely destroyed
- Examiner's Manuals are collected and destroyed
- Bulleted items on pages 20 and 21 are used to ensure that all tasks are completed appropriately

41

Markings in Test Books

In order to be scored properly, the following must take place:

1. Test books completed in ink are transcribed into another test book with a non-mechanical #2 pencil
2. Student responses written on coding tracks/margins are erased and transcribed onto the response line or box
3. All stray marks on the coding track are erased
4. All underlining of text is erased around the answer choices
5. Test books marked with highlighters are transcribed into other test books for scoring

42

Step 4: Check the Organization of Materials Collected

Collect the following after testing:

- Large white envelopes with all items in the exact order prescribed on page 22-23 and the illustration on page 35
- All test books, including student barcodes and/or completed SIS's for each students. Used and unused books are boxed separately

  Unused books include:
  - Test books which are damaged or have manufacturing errors
  - Test books written in a language other than English
  - Test books that are partially used because two test books were used by mistake

    (**NOTE**: Test books that are incomplete because the student was absent should be included with the scorable test books)
  - Test books that are Braille/Large Print editions with contents transcribed to a regular test book
- Completed GIS

43

Step 5: Check Student Information Sheet (SIS)

- Each Examiner's materials MUST be grouped together
- Physical Conditions of SIS can affect/interfere with scoring. Those conditions are listed on pages 24-25
- Biographical data must be checked on the barcode label & pre-code roster
- SIS of the test book must be completed only when:
  - Pre-coded student barcode label cannot be used
  - Pre-code roster information is inaccurate for a student
- A completed SIS and/or a test book with a student barcode label must be received for every eligible student

44

Sample SIS

- A picture of the SIS and explanations of the fields can be found on pages 26-27
- Reminder: Identify Special Populations and Invalidations

45

Step 6: Check the Teacher/Group Information Sheet (GIS)

- GIS provides data that is used on reports – notify the DTC if any errors exist on the GIS
- GIS is submitted for each grade/group/teacher
- GIS has both hand-entered and pre-coded information – **both must be accurate**
- GIS is scannable and **cannot** be photocopied
- GIS's are placed on top of test books whose scores are to be reported together

46

Step 7: Complete School/Group List

- Used by CTB/McGraw-Hill to inventory test books
- Can be photocopied
- Should have an entry for every GIS that was completed
- Includes:
  - Pre-coded:
    - District Name and County/District Code
    - School Name and Code
  - Hand-written:
    - Contact person's phone number
    - Grade/Group/Teacher Name is identical to "Teacher Name" on GIS
    - Grade Number
    - Number of books being returned
- Sample School/Group List on page 33
- Directions for completing fields on page 34

47

Step 8: Organize Materials for the DTC

The STC will:

1. Reuse the CTB/McGraw-Hill green-shaded boxes in which testing materials arrived
2. Package the following scorable materials:
   - School/Group Lists
   - Large white envelopes, organized by grade in ascending order, accompanied by GIS forms
3. Package unused testing materials
4. Affix shipping labels and number each set of boxes separately:
   - Blue labels on scorable books, numbered 1 of X, 2 of X, etc.
   - White labels on unused books, numbered 1 of X, 2 of X, etc.
5. Send materials to the DTC in **unsealed** boxes

48

Step 9: Package and Ship Testing Materials

The DTC will:

- Ensure all testing materials are received from each school in the district
- Verify TCM instructions for the STC were followed exactly
- Contact any STC who delays returning materials
- Add packing material
- Schedule testing material pickup online
- Schedule pickup no later than **April 27, 2009**
- Follow instructions outlined in TCM on page 38-39
- Fax Test Book Accountability Forms to CTB/McGraw-Hill and retain a copy for record-keeping purposes

49

Glossary of Terms

A glossary is listed at the end of the manual.
Most terms are the same as previous years.

50

Level Not Determined (LND)

This designation is for students who did not receive a score for any one of the following three reasons:

1. An SIS is returned to CTB/McGraw-Hill with a blank test book
2. A student does not attempt any items in one or more sessions or
3. A student is absent all 3 sessions

51

Questions????
Questions????

Missouri Department of Elementary and Secondary Education
800-845-3545 (Assessment Section)

52

DESE Presentation on Test Examiner's Manual

Missouri Assessment Program
Examiner's Manual Training
Spring 2009

Produced by the Missouri Department of Elementary and Secondary Education in collaboration with CTB/McGraw-Hill

1

Overview

- Missouri Assessment Program is one of several educational initiatives mandated by the Outstanding School Act of 1993.
- The State Board of Education mandated the Department of Elementary and Secondary Education (DESE) to identify the knowledge, skills, and competencies that Missouri students should acquire.

2

DESE uses testing information to:

- Monitor the progress of Missouri students in meeting the Show-Me standards
- Inform the public and state legislature about students' performance
- Inform state and federal accountability decisions
- Make informed decisions about educational issues
- Improve the services provided to Missouri students

3

Missouri Assessment Program
Grade-level Assessments for spring 2009

Communication Arts:	Grades 3-8
Mathematics:	Grades 3-8
Science:	Grades 5 and 8

4

Testing Time Guidelines:

Testing times vary per grade level – consult the "Testing Schedules" within the examiner's manual for timing guidelines. Testing time has increased given the fact that additional Field Test items are embedded.

5

Types of Questions

Students' content knowledge is measured using three types of test items:

1. **Selected Response Items (multiple-choice questions)**
2. **Constructed-response items (short answers)**
3. **Performance Events or Writing Prompt**

6

Field Test Items

- Field Test items are embedded throughout the test (EFTs)
- The testing period includes additional time for the EFTs

7

The Test Examiner's Manual is constructed so that an Examiner can administer more than one content area in a grade level using one manual.

Steps 1-4:	Contain the same information in all content areas
Step 5:	Contains test administration directions specific to Communication Arts, Mathematics, and Science for relevant grade level
Steps 6-8:	Contain the same information in all content areas

8

Table of Contents

9

Glossary

- Located on the last 2 pages in each Examiner's Manual
- An extensive list of terms and definitions used for testing materials and administering the assessment
- Helpful resource while reviewing the procedures/guidelines for administering the grade-level test

10

Step 1: Preparing for Testing

(Pages 1-4)

- **Testing Schedules** – Page 1 for all content areas
- Review the test directions in the Examiner's Manual in advance
- Testing materials shall not be viewed by Examiners before testing with few exceptions, e.g. translators
- Some sessions are strictly timed. Other sessions allow students who are making adequate progress to complete the test
- Each test session must be completed in one sitting
- Proctors – One proctor for every 20 students in excess of 30
- Accommodations – Discussed in Step 7 (ELL and IEP)

11

Use Standardized Testing Procedures

- Follow the instructions to ensure similar testing conditions are used in all classrooms:
  - Read test directions exactly as written
  - Observe time allowances for strictly-timed sections
- Verify that students understand the directions and how to mark answers:
  - Test directions in the manual can be clarified if necessary
  - Test items and their directions **CANNOT** be clarified or paraphrased

12

Large Print and Braille

- Accommodations must be coded on SIS on inside of front cover of the test book
- Manipulatives are included with Braille and Large Print editions
- Brailler can be used for the Braille edition
- Braille and Large Print test books must be transcribed to regular test book
- Braille and Large Print test books have special instructions that accompany the test books rather than the TC Kit
- Mark the Braille and Large Print book with the words, **"Contents transcribed to a regular test book. DO NOT SCORE"**, and return those test book editions to CTB/McGraw-Hill with the **unused** test books

13

Step 2: Organize Your Classroom

(Page 4)

Plan and organize for:

- Distributing and collecting materials
- Making seating arrangements
- Using a "Do Not Disturb Sign" (not provided)
- Note **start** and **stop** times on the board for the timed sections of the test

14

FOUR DO'S FOR EXAMINERS

Do eliminate noise distractions

Do remove all information from the room that might cue students about test content or process

Do prepare students the day before the test to get a good night's sleep

Do remind students to use **ONLY** No. 2 non-mechanical pencils to bubble in their answers

15

THREE DON'TS FOR EXAMINERS

Don't correct answers that are wrong

Don't limit students -- encourage students as a group to attempt all items

Don't return books to students to have them revise or complete answers

16

Step 3: Check Your Testing Materials

(Pages 5-7)

Examiner materials:

1. Examiner's Manual
2. Student barcode labels
3. Test Books
4. Manipulatives - ***Prepare by having all manipulatives "punched-out" prior to testing***
5. One Group Information Sheet (GIS) per Examiner/Grade
6. Large white envelopes
7. A box or envelope for unused test books
8. Do Not Disturb sign (not provided)

17

Student materials

Items not provided but specified for use in various testing sessions. Read instructions in Examiner's Manual:

- Writing instruments are not provided – students must use non-mechanical, No. 2 pencils
- Scratch, graph or grid paper for all Mathematics sessions
- Book markers (such as **blank** paper strips or index cards)

18

Ancillary Materials for Mathematics:

Manipulatives:

Students **will be** provided with a standard set of manipulatives for use during the test. No other classroom sets of manipulatives will be allowed.

Reference Sheets:

Grades 3 through 7 are provided all necessary formulas and conversions with the items. No separate reference sheet will be provided.

Grades 8 **will be** provided with a reference sheet that contains formulas and conversions.

Calculators:

Grades 3 through 5 are **not** allowed to use calculators during **any** session of the test.

- Grades 6 through 8 may use calculators **except** during **Session 2, Parts 1 and 3**. If schools choose to use calculators on the other parts of the test, students should be given advanced notice and given access to a calculator

19

Ancillary Materials for Communication Arts:

Grade 3: A standard dictionary and extra paper for writing first drafts may be used only during Session 2.

Grades 7: A standard dictionary, thesaurus, grammar handbook, and extra paper for first drafts may be used during Session 2.

Grades 4, 5, 6, 8: No tools may be used for any sessions of the test.

20

Ancillary Materials for Science:

REFERENCE SHEETS:

Grade 5: will not be provided a reference sheet. Necessary formulas and conversions will be included with the item

Grades 8: will be provided with a reference sheet that contains formulas and conversions when required by particular items

CALCULATORS:

Grade 5 and 8: will not be allowed to use calculators during **any** session of the test

21

Security "Shall Nots"

- Testing materials **shall not** be photocopied, duplicated or made accessible to non-testing personnel
- Testing materials **shall not** be viewed by Examiners before testing
- Testing materials **shall not** be left in an unsecured area at any time, for any reason -- must be locked in a secure cabinet or in a secure room before, between and after testing sessions
- Testing materials **shall not** be discussed between Examiners
- Test books **shall not** be shared between schools

22

Security Barcode

- Each test book has a unique barcode, numbered in consecutive order for security and inventory.
- Security barcodes are located on the lower right-hand corner of the students' test books.
- Barcodes should be checked by DTC and STC, and books counted by the Examiner.

23

Accommodations for ELL

- All tests, except Communication Arts, may be read to ELL students in their native language
- ELL students may give responses orally or in writing in their native language
- All native language written responses must be transcribed verbatim to another test book
- Translators must be trained in giving the assessments

24

Step 4: Before Testing

(Pages 7-9)

When all information on the student pre-code roster is correct	→ *Affix the barcode label Do not complete the SIS*
Error in *Student Name, Birth Date, Race/Ethnicity, Gender, MOSIS State ID*	→ *Do not use the barcode label Bubble in ALL info on SIS*
Wrong student label is affixed	→ *Place one blank label over the incorrect label Then affix correct label*

25

Using Student Barcode Labels

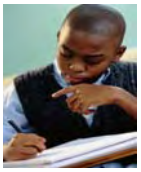
Label attached with error in *Student Name, Birth Date, Race/Ethnicity, Gender, MOSIS State ID*	→ *Place 2 blank labels over incorrect label Bubble in all info on SIS*
No student barcode label	→ *Enroll/submit to core data Leave barcode area blank Bubble in all info on SIS*

26

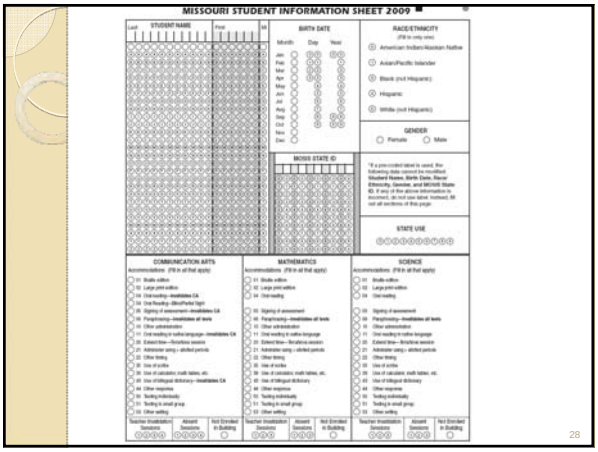
How to Fill Out the SIS

(Page 8)

1. Student Name
2. Birth Date
3. Race/Ethnicity
4. Gender
5. MOSIS State ID



27



28

Step 5: Administer the Test

(Begins on Page 11. Section 5 will vary in length depending on grade level or content being tested).

- Help students relax and be positive
- Explain the purpose of the test
- Point out that some items may be more difficult or may be new to students – *they are not expected to know all the answers. They are only expected to do their best.*

29

Specific Instructions for Test

- Examiner's instructions are arranged by grade level across content areas
- Content area is in alphabetical order– Communication Arts, Mathematics, and Science
- Sample questions are illustrated and printed in the Examiner's Manual for each content
- A starting/stopping time graphic is printed for Examiner's to follow
- Sessions cannot be split over 2 days or over lunch periods
- Break times are printed in the Examiner's Manual
- Notice the wording about not paraphrasing test questions and pronouncing only **one** word per sentence

30

Step 6: Invalidations and Make-ups

Page 37 – Page numbers will vary due to variation in Step 5)

Six categories for test invalidation:

1. Student discovered cheating
2. Examiner reads any part of the C.A. test to student(s) other than students who are Blind/Visually impaired who do not read Braille
3. Examiner signs any part of the C.A. test
4. Examiner paraphrases test questions in any content area
5. Examiner reads any part of the C.A. test to a student in the student's native language
6. A student uses a bilingual dictionary for any part of the C.A. test

31

Examiner's and STC's Responsibilities regarding Invalidations

- The Examiner must provide STC with student information and the reason for the invalidation
- The Examiner and STC must be in agreement about the invalidation due to cheating
- The STC is responsible for filling in appropriate bubble in Accommodations Section of SIS

32

Make-ups

•Students who are absent during one or more sessions of testing should make-up the sessions

If a student cannot participate in make-up testing sessions, follow the procedures in the Examiner's Manual listed in Step 6

33

Accommodations Footnotes

Please observe the footnotes at the bottom of Accommodations pages:

1. Oral reading of CA except for blind students who do not read Braille
2. Paraphrasing test questions invalidates that student's scores for accountability purposes.
3. If paraphrasing is used, student scores cannot be compared with scores generated under standard conditions.
4. Use of magnifying equipment, amplification equipment, graph paper, and testing with the teacher facing student are not listed as accommodations because these are no longer required to be reported as accommodations for the MAP tests.

35

Students Not Tested in content area assessments

Only two groups of students are not required to take the MAP:

1. MAP Alternate (MAP-A) students
2. ELL (English Language Learner) in the United States 12 months or fewer at the time of administration of MAP assessments may be exempted from the Communication Arts test. All other content areas must be assessed.

36

Step 8: Assemble Materials for Return

Pages 43-48 (Page numbers will vary due to Step 5)

Checklist after testing:

√ All student barcode labels are accurate and attached correctly

√ All student identifying information is correct and complete according to Step 7

√ The SIS should not be damaged or torn

√ All applicable accommodations are completed accurately

√ Non-mechanical, No. 2 pencil was used

√ Test books with issues/problems were transcribed and recorded

√ (Continued on next slide)

37

Assemble Materials (cont.)

- √ All loose papers, scratch paper, etc., are removed and securely destroyed
- √ The STC completes and checks Group Information Sheets (GIS)
- √ Examiners must count the test books against the pretest total – both used and unused
- √ **Organize test books whose scores are to be reported together**

38

Assemble Materials (cont.)

Organize materials as shown on picture of envelope

1. MAP test books – each with a student barcode label or with a completed SIS
2. GIS for class/group

NOTE: Do **not** seal the envelopes. The DTC will verify the contents before sealing them.

39

Large White Envelopes

- Each envelope will hold approximately 5–10 test books, depending on grade level
- If an Examiner has more than one envelope, put the GIS in the Examiner's first envelope with as many test books as will fit.
- If multiple envelopes are needed to hold large groups of books, the envelopes should be bundled together and marked on the front upper-left corner "1 of X," "2 of X," "3 of X," etc., with "X" being the total number of envelopes
- The information requested on the front of each envelope must be completed

40

Assemble Materials (cont.)

- Unused test books, Large Print and Braille test books can be placed in a box together labeled "Unused Test Books"
- Draft copies of writing prompt and scratch, graph or grid paper should be given to STC for secure destruction.

41

What should an Examiner do if a problem occurs?

- For any problems or questions, contact the STC who will, in turn, contact the DTC
- The DTC is the sole channel for all communication between districts and CTB/McGraw-Hill customer service line.
- The DTC should also be the channel of communication regarding testing issues with DESE

42

Presentation on Test Administration

Know your P's & Q's For Successful MAP Test Administration

Regional Instructional Facilitators:

Northeast RPDC

Pam Carte
Communication Arts

Myra Collins
Mathematics

Norma Neely
Science

1

Standardized Test

A *standardized test* is a test administered and scored in a consistent manner. The tests are designed in such a way that the "questions, conditions for administering, scoring procedures, and interpretations are consistent" and are "administered and scored in a predetermined, standard manner."

4

Test Validity

Test validity refers to the degree to which the test actually measures what it claims to measure. Test validity is also the extent to which inferences, conclusions, and decisions made on the basis of test scores are appropriate and meaningful.

5

Test Reliability

Test reliability refers to the degree to which a test is consistent and stable in measuring what it is intended to measure. Most simply put, a test is reliable if it is consistent within itself and across time.

6

Test Accountability

For more information contact the Department of Accountability, Data and Accreditation

(573) 751-1362

7

Reminder of 2009 Changes

- Grades 3-8 will have only **one test booklet** that contains all mandated tests for each grade level.
- Two Examples:
  - Grade 5 test booklet will contain the MAP mathematics, communication arts and science tests.
  - Grade 6 test booklet will contain MAP communication arts and mathematics tests.
- High school will still use EOC exam booklets

8

Testing for Spring 2009

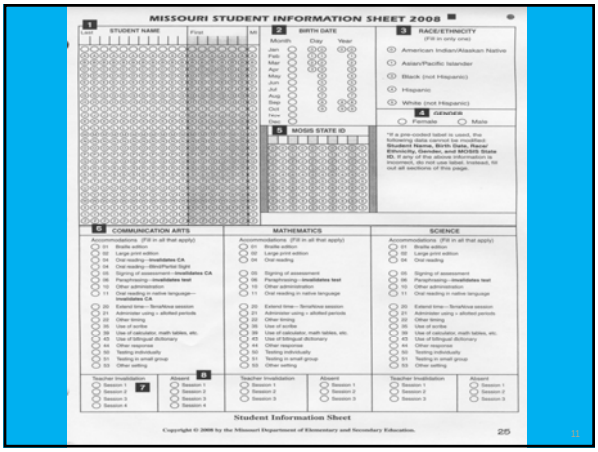
- Four sessions in CA test grades 3 and 7
- Embedded field test items in all tests
- **NO** released items in math and Comm. arts
- **NO** early returns or voluntary tests
- Mandated science in grades 5, 8, and Biology
- Fewer LNDs – participation will get an achievement level

9

Testing for Spring 2009

- One Group Information Sheet (GIS) per grade-level per building.
- Student Information Sheet (SIS) contain information related to the actual testing for each content area separate, example:
  - *Accommodations*
  - *Absent*
  - *Teacher Invalidation*

10



11

Major Things To Think About

Related to-----ONE TEST BOOK

- How does your district/building want the student test book labels sorted and returned to the district?
- How will your school set up testing sessions?
- How will tests be administered since there is only one test booklet?

12

Major Things To Think About Related to-----ONE TEST BOOK

- Ordering Test Book Labels—Decision to sort
- Online enrollment December 1st -15th
- Scheduling?
  - What content/sessions when?
  - Who will administer the test?
  - What training needs to occur?

13

Format for the Spring 2009 Tests

See handout on emphasis of strands

14

Communication Arts
Grades 3 and 7

Session 1:	45 – 55 minutes	
Session 2:	60 – 90 minutes	
Session 3: Part I **Session 3: Part 2**	Approx. 26 minutes (timed) Approx. 26 minutes (timed)	Five minute break between Part 1 and Part 2
Session 4:	50 – 60 minutes	

Timings and item counts are approximate and may vary by grade level.

15

Communication Arts
Grades 4, 5, 6, and 8

Session 1:	45 – 55 minutes	
Session 2: Part I **Session 2: Part 2**	Approx. 26 minutes (timed) Approx. 26 minutes (timed)	Five-minute break between Part I and Part 2
Session 3:	50 – 65 minutes	

Timings and item counts are approximate and may vary by grade level.

16

Mathematics
Grades 3, 5, 6 and 7

Session 1:	40 – 45 minutes
Session 2:	40 minutes (timed)
Session 3:	50 – 70 minutes

Timings and item counts are approximate and may vary by grade level.

17

Mathematics
Grades 4 and 8

Session 1:Part I **Session 1:Part 2**	45 – 50 minutes 15 – 20 minutes	Five-minute break between Part I and Part 2
Session 2:	40 minutes (timed)	
Session 3:	50 – 70 minutes	

Timings and item counts are approximate and may vary by grade level.

18

Science
Grades 5 and 8

Session 1:Part I **Session 1: Part 2**	30 – 40 minutes 30 – 40 minutes	Five-minute break between Part I and Part 2
Session 2: Part 1 **Session 2: Part 2**	25 minutes (timed) 40 – 55 minutes	Five-minute break between Part I and Part 2
Session 3: Part 1 **Session 3: Part 2**	55 – 70 minutes 30 – 40 minutes	Five-minute break between Part I and Part 2

Timings and item counts are approximate and may vary by grade level.

19

MAP SCHEDULING

20

Schedule

- Start several months ahead to plan the schedule your test administration.
- Testing window is 3/30 through 4/24
- RETURN TEST IMMEDIATELY AFTER YOUR DISTRICT TESTING WINDOW
- April 29 is the last day to contact CTB for pick-up to avoid cost to district

21

Plan the Schedule

- **Day of week**
- **Time of day**
- **School schedule**
- **Number of sessions per day**

22

Scheduling

- **Testing all students**
- **One test booklet per student grades 3-8 including all content areas**
- **Be aware of time needed to presort for possible new groupings of content area**
- **Continuity of session**
- **Make-up sessions**

23

Scheduling

- **Need for extra time**
- **Setting up testing groups**
- **Detailed report schedule to MAP office—to include update on changes**

24

TEST SCHEDULING Questions and Answers

25

Can I give more than one testing session per day?

- Yes, but it is important to consider the rigor and length of the sessions.
- For shorter and less demanding sessions, it may be acceptable to schedule two sessions in one day.
- See suggested time table in the examiner's manual.

26

May I break a testing session up so that my students can complete the session throughout the day?

No, each **session** of the test must be completed in one sitting. Breaks can only be taken if specified in the examiner's manual.

27

Can I start the writing prompt in the afternoon and have my students finish it the next morning?

No, each session of the test must be completed in one sitting.

28

How long are breaks during a testing session? Does that include lunch?

Breaks may be taken before or after a testing session, but not during a session unless specified in the examiner's manual. Lunch should be eaten before or after a session.

29

My students need beverages and snacks during testing. Can they have these items at their desks?

Yes, but it is best to have food or drink before the test or during breaks so nothing is spilled on a test booklet. If this happens, the test may have to be retaken or answers transcribed to another test booklet.

30

My students will test better if they are divided into small groups. Is it okay if I divide them into small groups (for example, 5 or less) for the MAP test?

- Small group testing is an accommodation.
- Students may not be tested in smaller groups for the purpose of MAP testing.

31

I want to test all students that I instruct. Can I test different groups of students over the same sessions of the test on different days?

It is not recommended that the test is administered over an extended period of time, because it can lead to test security issues. Testing all grade level/subject area students at the same time during the day is recommended.

32

TEST SECURITY

33

Test Security

Plan *now* to secure storage availability for test books and materials in school buildings and at the district level.

This may be more of an issue if trying to transport or exchange student books between/among different content area teachers for test administration.

34

- **Access to test books before testing**
- **Storage of books**
- **Teacher access to books**
- **No Grading test responses (MAP only)**
- **Beware of "Practice Tests"** not to include

35

TEST SECURITY Questions and Answers

36

I want to view the test booklets before I hand them out to students. How soon can I get the books to preview them?

- Only the school test coordinator and the district test coordinator should have access to tests before testing sessions begin.
- The exception may be a special education teacher pre-selecting items for a student whose IEP requires this accommodation OR an ELL translator who needs to prepare for test administration.
- In both instances, these processes should be done under the supervision of the school test coordinator and in a secure area.

37

Can special education teachers view the test booklets in advance of testing?

Special education teachers can view the test booklets **only** if they have students with IEPs that specify the accommodation of pre-selected items. The previewing process should be done under the supervision of the school test coordinator and in a secure area.

38

All of the teachers in my department want to see what the MAP test looks like. Can they view a test booklet at any time?

No, teachers are not permitted to view the test booklets at any time.

39

I want to grade my students' responses so I can assign classroom points. Is that okay?

No, grading test booklets is not permitted.

40

Is it acceptable if I take notes on the current test? I know those items won't be on the test next year, so does it matter?

No, **all** test questions and their contents are secure.

41

TEST ADMINISTRATION

42

Test Administration

Plan early in the school year for teacher in-services addressing:

- **MAP testing environment/procedures**
- **Balanced assessments—Question Types**

43

What can a proctor do during the test?

- See the handout/next slide: What a Proctor Looks Like and Sounds Like
- No giving cues, clues or hints as to the correct answer or how to go about finding the answer.

44

What does a testing proctor look like and sound like?

LOOKS LIKE	SOUNDS LIKE
•The person is writing the starting/stopping times on the board for timed sessions. •The person is moving around the room and through the aisles to ensure students: •Understand directions •Are on the correct page •Are marking bubbles properly •Are starting/stopping in correct location •Stop when time is over for timed testing sessions. •Do not look ahead/back into previous/past sessions •Stay on task and are not goofing off/wasting time/disturbing others •Have basic needs provided for such as sharpened pencils, tissues for runny noses, white erasers, etc. •Are not getting food or drinks near test books •Have books, etc to read upon completion of the testing session	•The person might be heard saying: •"Sorry. I cannot assist with answers but can only clarify directions." •"You need to do your own thinking." •"I may only pronounce one word per sentence." •"I may not explain or define a word." •"Please do not look back into other sessions." •"Please do not eat snacks/drink beverages during a testing session. You might spill something on the test booklet." •"It is alright to write slightly below the lines/outside the box but do not write on/past the coding lines." •"Find the STOP sign---so you will know where to stop."

Created by Jana Scott, MAP RTF/ University of MO-Columbia, 2007.

Test Administration

- **Wall, Wall, what's on the Wall?**
- **Acceptable student materials during testing**
- **Test stamina**

46

TEST ADMINISTRATION Questions and Answers

47

Can I leave anything on my classroom walls during MAP testing that has been up all year?

No, materials that give clues to correct content and/or processes should be covered or removed from the walls/desks or within view of the students.

If in doubt, cover it up or take it down.

48

If I leave up a "word wall" am I violating test procedures?

It would depend on the information contained in the word wall. If information on the word wall gives clues to content or processes, then it **must** be covered or removed.

49

My students have their own dictionaries that they have made. Can they use these on the MAP test?

No, only a standard dictionary may be used by students in grades 3 and 7 during Session II of the Communication Arts test. (This content session is the only session which allows the use of a dictionary.)

50

Can students make notes or write rough drafts of the constructed response questions and answers for the MAP test?

No, constructed responses must be answered in the test booklet in the space provided.

51

I've been told to look over my students' test booklets and return them if there are unanswered items. Is that okay to do?

No, teachers should not give feedback related to specific test items. They cannot have students change, add to, or supply missing answers.

52

Can I hand out graphic organizers for my students to use during testing and or the prewriting on the writing prompt?

No, all students should create their own graphic organizers during the testing process. The same rule applies for a prewriting strategy for the writing prompt.

53

MOTIVATING STUDENTS TO TRY

54

Share Ideas Your School Uses to Motivate Students to Do Their Best

55

CONCERNS AND ALERTS DURING TESTING AND SCORING

56

Common Concerns Noted During Testing

1. Inappropriate or Improper Use of Accommodations
2. Inappropriate Assistance-Verbal or Non-verbal (Cueing, pointing, pronouncing more than one word per sentence, reviewing test questions, etc.)
3. Returning books for correction or completion
4. Inappropriate Administration (Not using manual, splitting sessions, not following timing guidelines, etc.)
5. Inappropriate materials exposed
6. Photocopied Tests or Created Practice Tests Using MAP Tested Content
7. Test Security Violations
8. Inappropriate use of calculator or dictionary

57

ALERTS Found During the Scoring Process

- Editing Student Responses:
- Teachers should not edit student responses for spelling, grammar, and/or handwriting.
- Readers do a good job at figuring out what students are trying to write and say!

58

ALERTS Found During the Scoring Process

•Students should not look back into other testing sessions of the test while completing the essay portion of the Comm. Arts MAP (Grades 3 and 7 Session II).

•The looking back into other sessions is apparent by the fact that some students' essays contain verbatim quotes taken from a story in a different testing session

59

MAP Manuals Training in the Spring

- It is important to note----today's training does not replace MAP Manuals Training in the Spring.
- More specific and detailed information in regard to form completion, test administration, and test security.
- School and/or District Testing Coordinators should plan on attending the spring manuals training.

60

Please remember to share this information with your staff!

61

Questions

Closure

62

Appendix B

Missouri Assessment Program:
Anchor Evaluation for Communication Arts, Mathematics, and Science

The anchor items were evaluated immediately following the calibration and equating of the Missouri Assessment Program (MAP). This report outlines the methods used to evaluate anchor items for the MAP and the results of the analyses.

Methods Used to Evaluate Anchor Items

For the MAP, two statistical methods are used to evaluate anchor items: (1) iterative linking (Candell & Drasgow, 1988) using Stocking and Lord's (1983) test characteristic curve method, and (2) differences between the item-ability regression curves.

Test Characteristic Curve Method[1]

The Stocking and Lord (1983) procedure, also called the test characteristic curve (TCC) method, minimizes the mean squared difference between the two TCCs, one based on estimates from the previous calibration and the other on transformed estimates from the current calibration. Let $\hat{\psi}_j$ be the test characteristic curve based on estimates from a previous calibration and $\hat{\psi}_j^*$ be the test characteristic curve based on transformed estimates from the current calibration:

$$\hat{\psi}_j = \hat{\psi}(\theta_j) = \sum_{i=1}^{n} P_i(\theta_j; a_i, b_i, c_i),$$

$$\hat{\psi}_j^* = \hat{\psi}(\theta_j) = \sum_{i=1}^{n} P_i(\theta_j; \frac{a_i}{M_1}, M_1 b_i + M_2, c_i),$$

The TCC method determines the scaling constants (M_1 and M_2) by minimizing the following quadratic loss function (F):

$$F = \frac{1}{N} \sum_{a=1}^{N} (\hat{\psi}_j - \hat{\psi}_j^*)^2.$$

Differential item functioning was evaluated by examining previous (input) and transformed (estimated) item parameters. The item with an absolute difference of parameters greater than two times the root mean square deviation was flagged. The difference was also monitored by plotting input and estimated item parameters.

Item Response Theory (IRT) Item-Ability Regression Curves

We will also compute differences between the item-ability regression curves of the anchor items for the 2008 and 2009 calibrations. The differences between the curves will be evaluated using the following statistics:

- UnWtd Mean = Average signed difference in estimated probability.
- UnWtd Mean Abs Dif = Average Absolute (unsigned) difference in estimated probability.

[1] Text explaining the Test Characteristic Curve Method, Delta-Plot Method, and Lord's Chi Square is taken from Karkee and Choi (2005). *Impact of Eliminating Anchor Items Flagged from Statistical Criteria on Test Score Classifications in Common Item Equating.* Paper presented at the American Educational Research Association, Montreal, Canada.

- UnWtd RMSD = Root mean squared difference.

Both unweighted and weighted versions of these statistics will be calculated. Unweighted differences give equal weight to differences across the ability spectrum. Weighted differences assign weights according to the number of test-takers that are impacted.

The weighted versions of these differences are:

- Wtd Mean = Weighted average signed difference in estimated probability.
- Wtd Mean Abs = Weighted average Absolute (unsigned) difference in estimated probability.
- WtdRMSD = Weighted Root mean squared difference.

For the six statistics listed above, differences greater than +.10 are considered large, and differences between +.07 and .10 are considered moderate.

Additionally, the Maximum Absolute difference (MaxAbsDifPC) will be identified. For MaxAbsDIFPC, large differences are those greater than +.15, and moderate differences are all differences between +.125 and .15.

Removal of Anchor Items

While dropping an anchor item flagged based solely on statistical criteria has its simplicity, this option may change the content coverage and equating constants, shift scale score distributions, and affect the classification of students by moving them into different proficiency levels. Before an anchor item may be dropped from an anchor set, the adequacy of the content coverage must be evaluated.

As stated above, an item is removed from the anchor set only if it adversely affects quality of scaling, not desirability of results. As such, CTB will not consider how the removal of an item affects the overall mean scale score or the impact data (percent of students in each achievement level) when recommending items for removal.

Items removed from the anchor set are still scored as part of the whole test. Anchor items are considered for exclusion from the MAP under the following conditions:

1. Items flagged using the TCC method are considered for exclusion when the correlation between the input and equated item parameters is below .90 for the *b*-parameter or below .80 for the *a*-parameter. If the exclusion of an outlying anchor item increases the correlation to above .90 for the *b*-parameter or above .80 for the *a*-parameter, then the anchor is a candidate for removal.
2. An item is a candidate for removal when it is flagged for large differences on four of the seven statistics considered when examining the differences between the IRT regression curves.
3. Removal of the item will only be considered after alternative explanations have been considered that may explain shifts in performance. For example, performance on the

anchor item may improve because of a statewide initiative emphasizing instruction on a particular set of skills. In this case, improved performance on the item represents true growth in that area. Removing the anchor item may artificially lower test scores.

4. Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of the anchor items across the content standards must remain within 10% of the MAP test blueprint.
5. The number of remaining items will remain at an acceptable level of anchor set reliability. Operationally, this means the anchor set will still be representative of the total test blueprint and that the anchor may not be less than 20% of the total test length.

Results of Analyses

Neither of the analyses revealed any items that were performing in a statistically different manner from 2008.

Detailed Results of the Test Characteristic Curve Method

Tables 1 through 3 provide results for the TCC method. These tables summarize the following information for each grade content area: grade level, number of iterations, scaling constants (M1 and M2), and quadratic loss function (F). Within each grade level, the following information is summarized for each item parameter estimate: difference (Diff), root mean square difference (RMSD), ratio of the standard deviation (SD Ratio), and correlation (r) between input (2008) and estimated (2009) anchor parameters. All correlations of the *a*- and *b*-parameters were greater than .98. No items were flagged using this method.

Please note that the actual TCCs are shown in Figures 1–14. These plots are used to assess the quality of the linking results. The light blue TCC lines in the plots are the TCCs for the input anchor items. The dark blue lines are the TCCs from the 2009 MAP parameter estimates transformed to the MAP scale.[2] The closer the two TCCs are to each other at all ability levels, the more confidence we have in the equating result. In all cases, the input and estimate TCCs overlay each other, making the two curves indistinguishable.

Detailed Results Comparing the IRT Anchor Regression Curves

Tables 4 through 17 present the detailed results for both the original and alternate linking when the IRT Anchor Regression method is used. These tables summarize the seven statistics examined using this method. The headers in the tables are abbreviated as follows:

- UnWtd RMSD = unweighted root mean squared difference
- UnWtd Mean Abs Difference = unweighted average absolute difference in estimated probability.
- UnWtd Max = unweighted maximum absolute difference.
- UnWtd Mean = unweighted average signed difference in estimated probability.
- Wtd RMSD = weighted root mean squared difference.

[2] The *c*-parameters for the MAP test data were fixed to the original *TerraNova c*-parameters in order to provide more accurate equating results (Voelkle, Schwarz, Arenson, & Ito 2002).

- Wtd Mean Abs Difference = weighted average absolute difference in estimated probability.
- Wtd Mean = weighted average signed difference in estimated probability.

Again, for six of the statistics listed above (except the unweighted maximum absolute difference), differences greater than $\pm$.10 were considered large, and differences greater than $\pm$.07 were considered moderate. For maximum absolute difference, large differences were those greater than $\pm$.15, and moderate differences were all differences greater than $\pm$.125. No items were flagged using this approach.

Table 1. Detailed Results from the Test Characteristic Curve Method, Communication Arts

Grade	Iterations	M1	M2	F	Par	Diff	RMSD	SD Ratio	r
3	14	32.634	639.55	0.034	a	0.000	0.001	1.061	0.996
					b	0.172	1.382	1.005	0.999
4	29	27.728	659.84	0.046	a	-0.001	0.001	0.960	0.993
					b	0.025	1.328	0.986	0.999
5	23	27.421	673.04	0.019	a	0.000	0.001	0.997	0.997
					b	-0.051	1.513	1.019	0.998
6	12	27.794	672.59	0.020	a	0.000	0.001	0.966	0.997
					b	0.045	1.019	1.000	0.999
7	19	29.668	679.98	0.020	a	0.000	0.001	1.008	0.996
					b	0.148	1.498	0.979	0.999
8	19	28.173	696.26	0.086	a	0.000	0.002	0.981	0.990
					b	0.231	1.929	0.996	0.998

Table 2. Detailed Results from the Test Characteristic Curve Method, Mathematics

Grade	Iterations	M1	M2	F	Par	Diff	RMSD	SD Ratio	r
3	23	32.168	624.44	0.025	a	0.000	0.001	1.072	0.987
					b	0.047	2.306	0.997	0.997
4	29	30.060	645.59	0.035	a	0.000	0.001	1.030	0.996
					b	-0.249	2.904	1.028	0.997
5	29	35.108	666.72	0.034	a	0.000	0.001	1.018	0.992
					b	0.161	2.178	0.977	0.998
6	23	34.734	680.05	0.007	a	0.000	0.001	1.019	0.991
					b	0.151	2.178	1.002	0.996
7	10	35.821	686.67	0.095	a	0.000	0.001	1.002	0.993
					b	0.175	2.519	1.002	0.997
8	9	33.239	704.85	0.146	a	0.000	0.001	1.005	0.982
					b	0.407	3.739	1.013	0.994

Table 3. Detailed Results from the Test Characteristic Curve Method, Science

Grade	Iterations	M1	M2	F	Par	Diff	RMSD	SD Ratio	r
5	10	27.324	663.28	0.066	a	0.000	0.001	1.044	0.994
					b	-0.223	2.165	0.992	0.998
8	15	27.025	697.71	0.055	a	-0.001	0.002	0.918	0.987
					b	-0.058	1.788	1.001	0.998

Table 4. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 3

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2	0.00	0.00	0.01	0.00	0.00	0.00	0.00
3	0.00	0.00	0.01	0.00	0.00	0.00	0.00
4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
5	0.00	0.00	0.01	0.00	0.00	0.00	0.00
6	0.01	0.00	0.02	0.00	0.00	0.00	0.00
7	0.00	0.00	0.01	0.00	0.01	0.01	0.01
8	0.01	0.01	0.02	0.00	0.01	0.01	0.00
9	0.00	0.00	0.01	0.00	0.00	0.00	0.00
19	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
20	0.00	0.00	0.01	0.00	0.01	0.01	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.01	0.00	0.01	0.00	0.01	0.01	0.01
23	0.00	0.00	0.01	0.00	0.00	0.00	0.00
26	0.00	0.00	0.01	0.00	0.00	0.00	0.00
27	0.00	0.00	0.01	0.00	0.01	0.00	0.00
28	0.00	0.00	0.01	0.00	0.00	0.00	0.00
29	0.00	0.00	0.01	0.00	0.00	0.00	0.00
30	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
31	0.01	0.01	0.02	0.01	0.01	0.01	0.01
32	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
37	0.00	0.00	0.00	0.00	0.00	0.00	0.00
38	0.00	0.00	0.01	0.00	0.00	0.00	0.00
39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.01	0.00	0.01	0.01	0.01
41	0.00	0.00	0.01	0.00	0.01	0.01	0.01
42	0.00	0.00	0.01	0.00	0.00	0.00	0.00
43	0.01	0.01	0.01	0.00	0.01	0.01	0.01
44	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02

Table 5. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 4

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.00	0.00	0.01	0.00	0.01	0.01	0.01
2	0.01	0.00	0.02	0.00	0.00	0.00	0.00
3	0.00	0.00	0.01	0.00	0.00	0.00	0.00
4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.01	0.01	0.03	0.01	0.01	0.01	0.01
7	0.00	0.00	0.01	0.00	0.00	0.00	0.00
14	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
15	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
16	0.02	0.01	0.04	-0.01	0.01	0.00	0.00
17	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
18	0.01	0.00	0.02	0.00	0.01	0.00	0.00
19	0.01	0.01	0.02	-0.01	0.02	0.01	-0.01
20	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
23	0.00	0.00	0.01	0.00	0.00	0.00	0.00
24	0.00	0.00	0.01	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	0.00	0.00	0.01	0.00	0.00	0.00	0.00
27	0.01	0.00	0.01	0.00	0.01	0.01	0.01
28	0.00	0.00	0.01	0.00	0.00	0.00	0.00
29	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
30	0.00	0.00	0.01	0.00	0.01	0.00	0.00
31	0.01	0.01	0.02	-0.01	0.01	0.00	0.00
32	0.01	0.01	0.02	0.01	0.01	0.01	0.01
33	0.01	0.00	0.01	0.00	0.01	0.00	0.00
34	0.01	0.01	0.02	0.00	0.01	0.01	0.01
35	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
37	0.01	0.00	0.01	0.00	0.01	0.01	0.00
38	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
39	0.01	0.01	0.02	0.00	0.01	0.01	0.00
40	0.00	0.00	0.01	0.00	0.00	0.00	0.00
41	0.00	0.00	0.01	0.00	0.00	0.00	0.00
42	0.01	0.00	0.01	0.00	0.01	0.01	0.01
43	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 5

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
7	0.01	0.01	0.03	0.01	0.01	0.01	0.01
8	0.00	0.00	0.01	0.00	0.01	0.00	0.00
9	0.00	0.00	0.01	0.00	0.01	0.00	0.00
10	0.01	0.01	0.02	0.01	0.01	0.01	0.01
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.01	0.01	0.02	0.01	0.01	0.01	0.01
13	0.00	0.00	0.01	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.01	0.00	0.00	0.00	0.00
16	0.01	0.00	0.01	0.00	0.00	0.00	0.00
17	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
18	0.01	0.00	0.01	0.00	0.01	0.01	0.01
19	0.01	0.01	0.01	0.00	0.00	0.00	0.00
20	0.01	0.01	0.02	0.01	0.02	0.02	0.01
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.00	0.00	0.01	0.00	0.00	0.00	0.00
28	0.00	0.00	0.01	0.00	0.00	0.00	0.00
29	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
30	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
31	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
32	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.00	0.00	0.01	0.00	0.00	0.00	0.00
37	0.01	0.01	0.02	0.01	0.02	0.01	0.01
38	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
41	0.01	0.01	0.02	0.01	0.01	0.01	0.01
42	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
43	0.01	0.01	0.02	-0.01	0.01	0.01	0.00
44	0.01	0.01	0.02	0.01	0.01	0.01	0.01
45	0.01	0.01	0.02	0.01	0.01	0.00	0.00

Table 7. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 6

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
7	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
8	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
9	0.01	0.01	0.02	0.01	0.01	0.01	0.01
10	0.01	0.00	0.01	0.00	0.01	0.01	0.01
11	0.01	0.00	0.01	0.00	0.01	0.01	0.01
12	0.01	0.01	0.01	0.00	0.01	0.01	0.01
13	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
14	0.01	0.00	0.01	0.00	0.01	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.01	0.01	0.01	0.00	0.01	0.01	-0.01
20	0.01	0.00	0.02	0.00	0.01	0.01	-0.01
21	0.00	0.00	0.01	0.00	0.00	0.00	0.00
22	0.00	0.00	0.01	0.00	0.01	0.01	0.00
23	0.01	0.01	0.01	0.00	0.01	0.00	0.00
24	0.00	0.00	0.01	0.00	0.01	0.00	0.00
25	0.00	0.00	0.01	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.00	0.00	0.01	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
31	0.01	0.01	0.01	0.00	0.01	0.01	-0.01
32	0.01	0.01	0.02	0.01	0.01	0.01	0.01
33	0.01	0.00	0.01	0.00	0.00	0.00	0.00
34	0.00	0.00	0.00	0.00	0.00	0.00	0.00
35	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
36	0.00	0.00	0.01	0.00	0.00	0.00	0.00
40	0.00	0.00	0.01	0.00	0.01	0.00	0.00
41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
42	0.01	0.01	0.01	0.00	0.01	0.01	0.01
43	0.00	0.00	0.01	0.00	0.01	0.01	0.00

Table 8. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 7

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
11	0.01	0.01	0.02	0.00	0.00	0.00	0.00
12	0.01	0.01	0.01	0.00	0.01	0.01	-0.01
13	0.00	0.00	0.01	0.00	0.00	0.00	0.00
14	0.00	0.00	0.01	0.00	0.01	0.00	0.00
19	0.00	0.00	0.01	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.01	0.00	0.02	0.00	0.01	0.01	0.01
22	0.01	0.00	0.01	0.00	0.01	0.01	0.01
23	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
24	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
25	0.01	0.01	0.02	-0.01	0.02	0.01	-0.01
26	0.00	0.00	0.01	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.01	0.01	0.02	-0.01	0.01	0.00	0.00
29	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.01	0.00	0.01	0.00	0.00	0.00	0.00
31	0.01	0.00	0.01	0.00	0.01	0.01	0.01
32	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
33	0.00	0.00	0.01	0.00	0.00	0.00	0.00
34	0.00	0.00	0.01	0.00	0.01	0.00	0.00
35	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.01	0.01	0.02	0.01	0.01	0.01	0.01
38	0.01	0.00	0.01	0.00	0.00	0.00	0.00
39	0.01	0.01	0.02	0.01	0.01	0.01	0.01
40	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
41	0.01	0.00	0.01	0.00	0.01	0.01	0.01
43	0.00	0.00	0.01	0.00	0.00	0.00	0.00
44	0.00	0.00	0.00	0.00	0.00	0.00	0.00
45	0.00	0.00	0.00	0.00	0.00	0.00	0.00
46	0.01	0.01	0.02	0.01	0.02	0.01	0.01
47	0.00	0.00	0.00	0.00	0.00	0.00	0.00
48	0.00	0.00	0.00	0.00	0.00	0.00	0.00
49	0.01	0.01	0.02	0.01	0.02	0.01	0.01

Table 9. Statistics Comparing IRT Item-Ability Regression Curves, Communication Arts, Grade 8

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.01	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.01	0.00	0.00	0.00	0.00
16	0.01	0.00	0.02	0.00	0.00	0.00	0.00
17	0.02	0.02	0.04	0.02	0.03	0.03	0.03
21	0.00	0.00	0.01	0.00	0.00	0.00	0.00
22	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
23	0.01	0.01	0.02	0.01	0.01	0.01	0.01
24	0.01	0.00	0.01	0.00	0.00	0.00	0.00
25	0.01	0.00	0.01	0.00	0.01	0.01	0.00
26	0.00	0.00	0.01	0.00	0.00	0.00	0.00
27	0.00	0.00	0.01	0.00	0.00	0.00	0.00
28	0.02	0.01	0.03	0.01	0.01	0.01	0.00
31	0.02	0.01	0.05	-0.01	0.03	0.03	-0.02
32	0.01	0.01	0.02	-0.01	0.01	0.00	0.00
33	0.01	0.01	0.01	0.00	0.01	0.01	0.00
34	0.00	0.00	0.00	0.00	0.00	0.00	0.00
35	0.01	0.00	0.02	0.00	0.01	0.01	0.01
36	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
37	0.00	0.00	0.01	0.00	0.00	0.00	0.00
38	0.00	0.00	0.00	0.00	0.00	0.00	0.00
39	0.01	0.00	0.02	0.00	0.01	0.01	0.01
40	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
41	0.00	0.00	0.01	0.00	0.01	0.01	0.00
42	0.01	0.01	0.03	0.00	0.02	0.01	0.00
43	0.02	0.01	0.04	-0.01	0.01	0.01	-0.01

Table 10. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 3

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.01	0.01	0.01	0.00	0.01	0.01	0.00
2	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
5	0.01	0.01	0.03	0.00	0.01	0.01	0.00
6	0.00	0.00	0.01	0.00	0.00	0.00	0.00
7	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
8	0.01	0.00	0.02	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.03	0.02	0.07	0.02	0.04	0.04	0.04
11	0.01	0.01	0.01	0.00	0.01	0.01	0.01
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.01	0.01	0.02	0.01	0.01	0.01	0.01
14	0.02	0.02	0.03	-0.02	0.03	0.03	-0.03
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.01	0.00	0.02	0.00	0.00	0.00	0.00
17	0.02	0.01	0.04	-0.01	0.02	0.02	-0.02
18	0.01	0.00	0.01	0.00	0.01	0.00	0.00
19	0.01	0.01	0.02	-0.01	0.01	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.01	0.01	0.02	0.01	0.01	0.01	0.01
22	0.01	0.01	0.01	0.00	0.01	0.00	0.00
23	0.01	0.01	0.02	0.01	0.01	0.01	0.01
24	0.01	0.01	0.03	0.00	0.02	0.01	0.01
25	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
26	0.01	0.00	0.02	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.02	0.02	0.03	-0.01	0.02	0.02	-0.02
30	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Table 11. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 4

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
5	0.00	0.00	0.01	0.00	0.01	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.01	0.00	0.02	0.00	0.01	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.01	0.01	0.02	0.01	0.01	0.01	0.01
10	0.01	0.01	0.03	0.01	0.01	0.01	0.01
11	0.02	0.01	0.06	-0.01	0.01	0.00	0.00
12	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
13	0.01	0.01	0.03	0.00	0.00	0.00	0.00
14	0.02	0.02	0.04	-0.02	0.03	0.03	-0.03
15	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
16	0.01	0.01	0.02	0.01	0.01	0.01	0.01
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.01	0.01	0.02	0.01	0.01	0.01	0.01
20	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
21	0.00	0.00	0.01	0.00	0.00	0.00	0.00
22	0.00	0.00	0.01	0.00	0.00	0.00	0.00
23	0.01	0.01	0.02	0.01	0.01	0.01	0.01
24	0.01	0.01	0.02	0.01	0.01	0.00	0.00
25	0.01	0.01	0.02	0.01	0.01	0.01	0.01
26	0.00	0.00	0.01	0.00	0.01	0.00	0.00
27	0.01	0.00	0.02	0.00	0.01	0.01	0.01
28	0.01	0.00	0.01	0.00	0.01	0.01	0.01
29	0.01	0.01	0.01	0.00	0.01	0.01	0.00
30	0.00	0.00	0.01	0.00	0.00	0.00	0.00
31	0.01	0.01	0.03	0.01	0.01	0.01	0.01
32	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01

Table 12. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 5

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
3	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
4	0.01	0.01	0.04	-0.01	0.02	0.01	-0.01
5	0.01	0.00	0.01	0.00	0.01	0.01	0.01
6	0.00	0.00	0.01	0.00	0.00	0.00	0.00
7	0.01	0.00	0.01	0.00	0.00	0.00	0.00
8	0.01	0.01	0.02	0.00	0.01	0.01	0.00
9	0.00	0.00	0.01	0.00	0.00	0.00	0.00
10	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.01	0.00	0.00	0.00	0.00
13	0.00	0.00	0.01	0.00	0.01	0.01	0.00
14	0.01	0.01	0.03	-0.01	0.02	0.02	-0.01
15	0.01	0.01	0.03	0.00	0.02	0.02	0.01
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.01	0.01	0.03	0.01	0.02	0.01	0.01
18	0.01	0.00	0.01	0.00	0.01	0.01	0.01
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.01	0.01	0.04	0.00	0.01	0.00	0.00
21	0.01	0.00	0.01	0.00	0.00	0.00	0.00
22	0.01	0.01	0.02	0.00	0.00	0.00	0.00
23	0.00	0.00	0.01	0.00	0.00	0.00	0.00
24	0.02	0.02	0.03	0.00	0.02	0.02	-0.02
25	0.00	0.00	0.01	0.00	0.01	0.01	0.01
26	0.00	0.00	0.01	0.00	0.00	0.00	0.00
27	0.01	0.00	0.01	0.00	0.01	0.01	0.01
28	0.01	0.00	0.01	0.00	0.01	0.01	0.01
29	0.01	0.00	0.01	0.00	0.00	0.00	0.00
30	0.01	0.01	0.01	0.01	0.01	0.01	0.01
31	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
32	0.01	0.01	0.02	0.01	0.01	0.00	0.00

Table 13. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 6

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
3	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
5	0.01	0.01	0.01	0.00	0.01	0.01	0.01
6	0.01	0.01	0.02	0.01	0.02	0.01	0.01
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.01	0.01	0.03	0.01	0.02	0.01	0.01
9	0.01	0.01	0.01	0.01	0.01	0.01	0.01
10	0.01	0.01	0.01	0.00	0.01	0.01	-0.01
11	0.01	0.01	0.02	0.00	0.01	0.01	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.01	0.01	0.02	0.01	0.01	0.01	0.01
14	0.01	0.00	0.01	0.00	0.01	0.00	0.00
15	0.00	0.00	0.01	0.00	0.00	0.00	0.00
16	0.04	0.03	0.06	0.01	0.03	0.02	0.00
17	0.00	0.00	0.01	0.00	0.01	0.00	0.00
18	0.00	0.00	0.01	0.00	0.01	0.01	0.01
19	0.00	0.00	0.01	0.00	0.01	0.01	0.01
20	0.00	0.00	0.01	0.00	0.01	0.01	0.01
21	0.01	0.00	0.01	0.00	0.00	0.00	0.00
22	0.01	0.01	0.03	-0.01	0.02	0.02	-0.01
23	0.01	0.01	0.01	0.00	0.01	0.00	0.00
24	0.01	0.01	0.02	0.00	0.01	0.01	0.01
25	0.02	0.01	0.04	-0.01	0.03	0.03	-0.03
26	0.01	0.00	0.02	0.00	0.00	0.00	0.00
27	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.01	0.01	0.01	0.01	0.01	0.01	0.01
30	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
31	0.00	0.00	0.01	0.00	0.01	0.00	0.00

Table 14. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 7

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.01	0.01	0.02	0.00	0.01	0.01	0.01
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
4	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
5	0.02	0.01	0.04	0.01	0.03	0.03	0.03
6	0.01	0.00	0.01	0.00	0.01	0.00	0.00
7	0.00	0.00	0.01	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.01	0.01	0.03	-0.01	0.01	0.00	0.00
10	0.01	0.01	0.02	0.01	0.02	0.02	0.02
11	0.01	0.01	0.03	0.01	0.01	0.01	0.01
12	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
13	0.01	0.01	0.01	0.00	0.01	0.01	0.01
14	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
15	0.01	0.01	0.04	0.01	0.02	0.02	0.02
16	0.00	0.00	0.01	0.00	0.00	0.00	0.00
17	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
18	0.01	0.01	0.02	0.00	0.01	0.00	0.00
19	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
20	0.02	0.01	0.04	-0.01	0.03	0.03	-0.03
21	0.02	0.01	0.04	-0.01	0.03	0.02	-0.02
22	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
23	0.02	0.02	0.05	0.02	0.03	0.03	0.03
24	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.01	0.00	0.02	0.00	0.01	0.01	0.01
29	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
30	0.01	0.01	0.03	0.01	0.02	0.01	0.01
31	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
32	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01

Table 15. Statistics Comparing IRT Item-Ability Regression Curves, Mathematics, Grade 8

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.01	0.01	0.02	0.01	0.01	0.01	0.01
2	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.01	0.00	0.01	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.01	0.01	0.02	0.01	0.02	0.02	0.02
7	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
8	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
9	0.02	0.02	0.05	-0.02	0.04	0.03	-0.03
10	0.01	0.00	0.01	0.00	0.01	0.01	0.01
11	0.01	0.01	0.02	0.00	0.01	0.01	0.01
12	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
13	0.00	0.00	0.01	0.00	0.01	0.01	0.00
14	0.05	0.04	0.10	0.04	0.08	0.08	0.08
15	0.03	0.02	0.05	0.02	0.02	0.01	0.01
16	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
17	0.03	0.02	0.05	-0.01	0.02	0.01	0.00
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.02	0.01	0.05	-0.01	0.01	0.01	0.00
20	0.01	0.00	0.01	0.00	0.01	0.00	0.00
21	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
22	0.01	0.01	0.03	0.00	0.01	0.01	0.00
23	0.01	0.01	0.01	0.01	0.01	0.01	0.01
24	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
25	0.00	0.00	0.01	0.00	0.00	0.00	0.00
26	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
27	0.01	0.00	0.02	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.00	0.00	0.01	0.00	0.01	0.00	0.00
30	0.02	0.01	0.04	-0.01	0.02	0.01	-0.01
31	0.01	0.01	0.02	0.00	0.01	0.01	-0.01

Table 16. Statistics Comparing IRT Item-Ability Regression Curves, Science, Grade 5

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.02	0.01	0.03	0.01	0.02	0.02	0.02
2	0.00	0.00	0.01	0.00	0.01	0.00	0.00
3	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
5	0.01	0.00	0.01	0.00	0.01	0.00	0.00
6	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
7	0.03	0.02	0.04	0.00	0.02	0.01	0.00
8	0.02	0.01	0.05	-0.01	0.03	0.03	-0.03
9	0.00	0.00	0.01	0.00	0.00	0.00	0.00
10	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
11	0.02	0.01	0.04	0.01	0.02	0.02	0.02
12	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
13	0.01	0.01	0.03	0.01	0.02	0.02	0.01
14	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
15	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
16	0.01	0.01	0.04	0.01	0.02	0.01	0.01
17	0.01	0.01	0.02	0.01	0.02	0.02	0.02
18	0.01	0.01	0.01	0.01	0.01	0.01	0.01
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.01	0.00	0.00	0.00	0.00
21	0.01	0.00	0.01	0.00	0.00	0.00	0.00
22	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Table 17. Statistics Comparing IRT Item-Ability Regression Curves, Science, Grade 8

Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
3	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
4	0.01	0.01	0.02	0.01	0.01	0.01	0.01
5	0.00	0.00	0.01	0.00	0.01	0.01	0.00
6	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.01	0.01	0.02	-0.01	0.01	0.00	0.00
9	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
10	0.01	0.01	0.02	0.00	0.01	0.01	0.01
11	0.01	0.01	0.03	0.01	0.01	0.01	0.01
12	0.01	0.01	0.03	0.00	0.02	0.02	-0.01
13	0.01	0.00	0.02	0.00	0.00	0.00	0.00
14	0.01	0.00	0.02	0.00	0.01	0.01	0.01
15	0.01	0.01	0.01	0.00	0.01	0.01	0.00
16	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.02	0.01	0.04	0.01	0.01	0.00	0.00
19	0.01	0.01	0.03	0.01	0.02	0.02	0.02
20	0.01	0.01	0.02	0.01	0.02	0.02	0.02
21	0.01	0.01	0.02	0.00	0.01	0.01	-0.01
22	0.01	0.01	0.01	0.00	0.01	0.01	0.00
23	0.01	0.01	0.01	0.00	0.01	0.01	0.01

Figure 1. Communication Arts, Grade 3 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

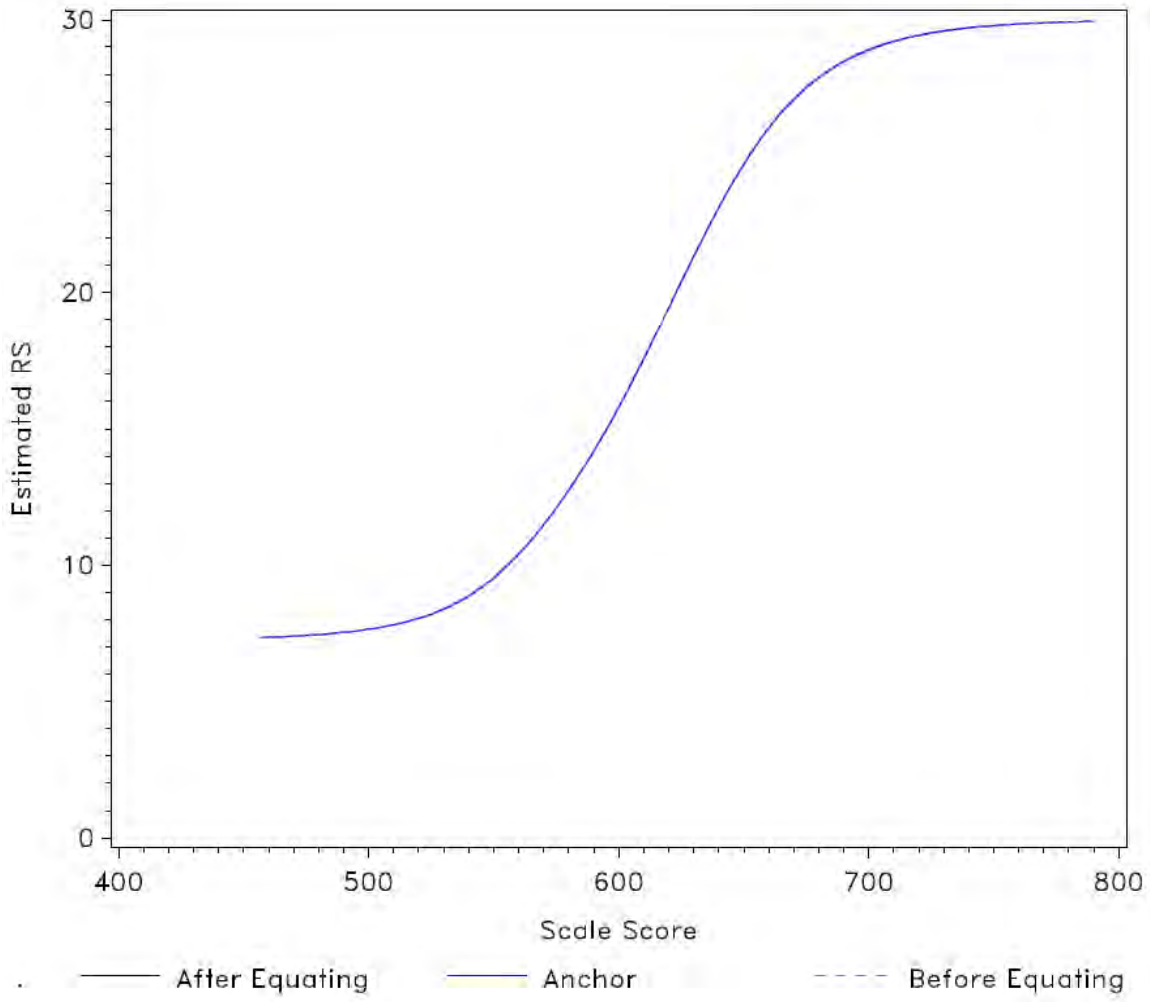


Figure 2. Communication Arts, Grade 4 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

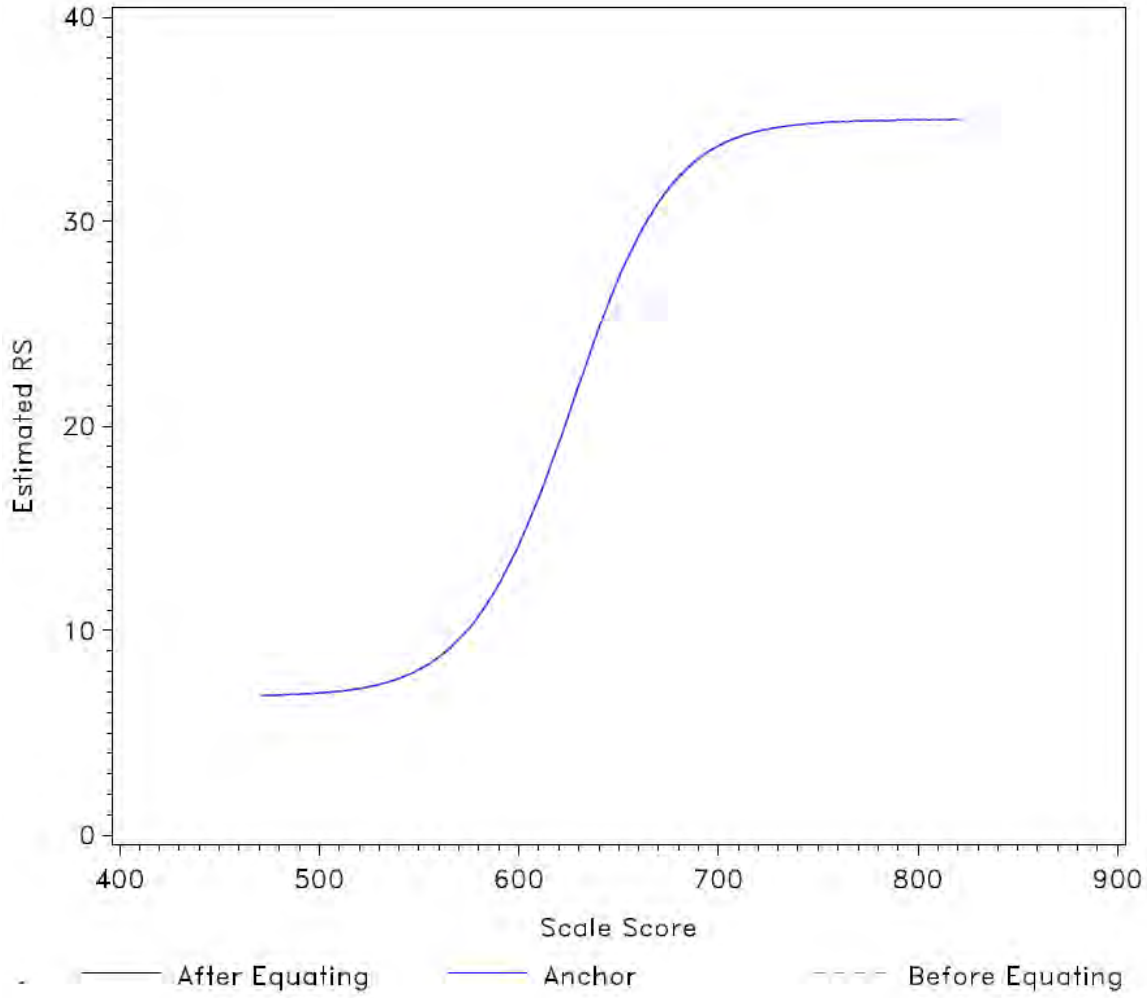


Figure 3. Communication Arts, Grade 5 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

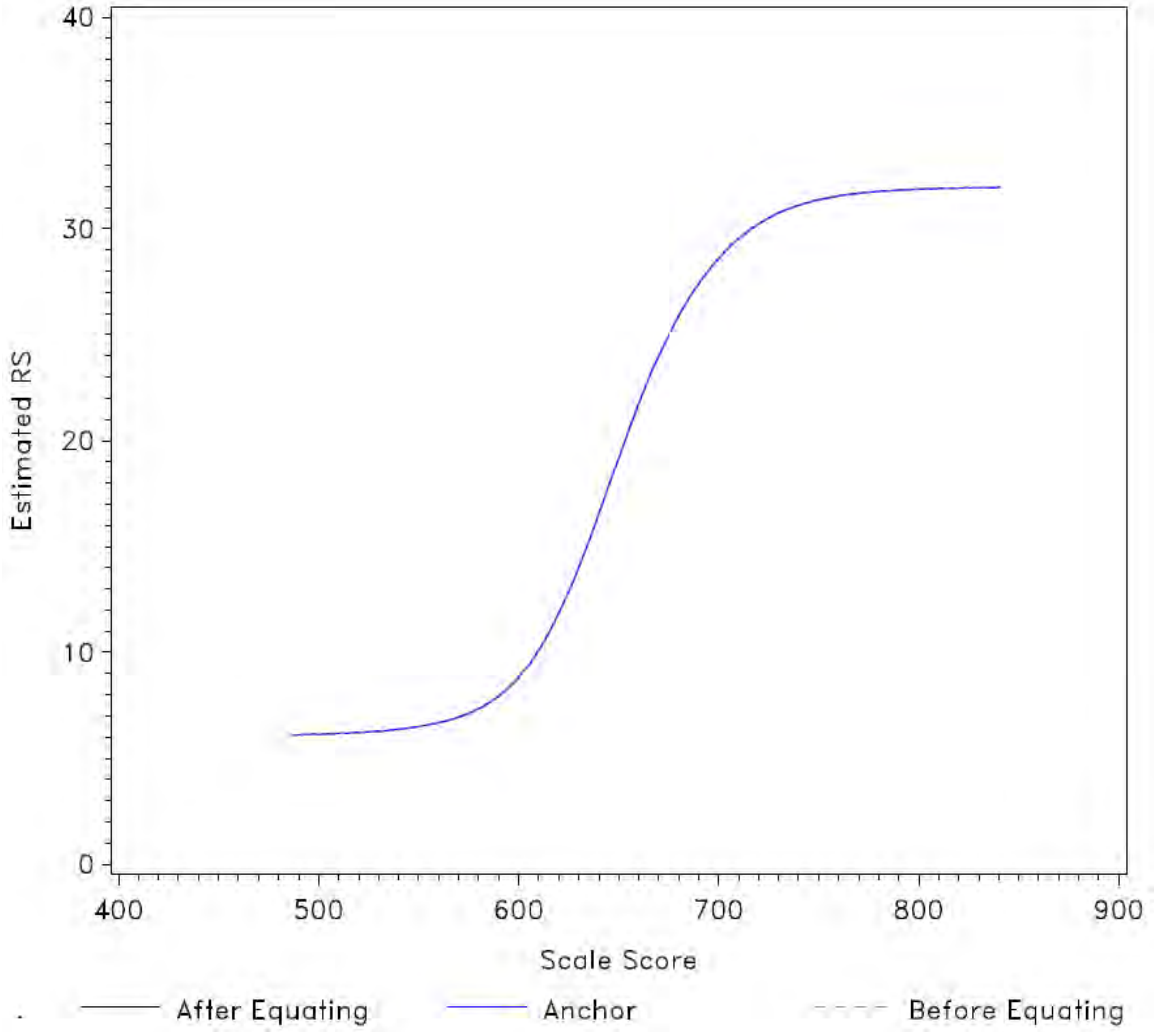


Figure 4. Communication Arts, Grade 6 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

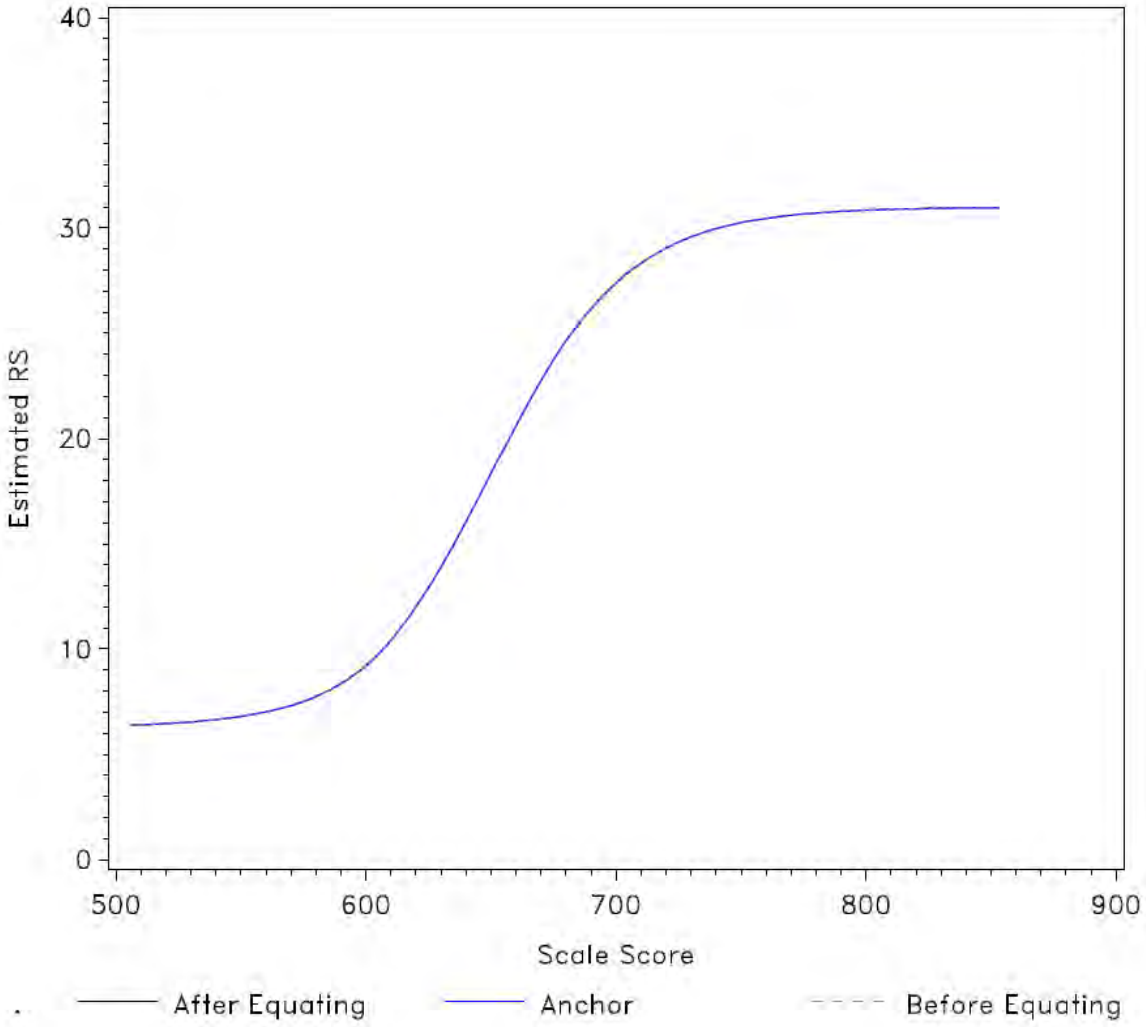


Figure 5. Communication Arts, Grade 7 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

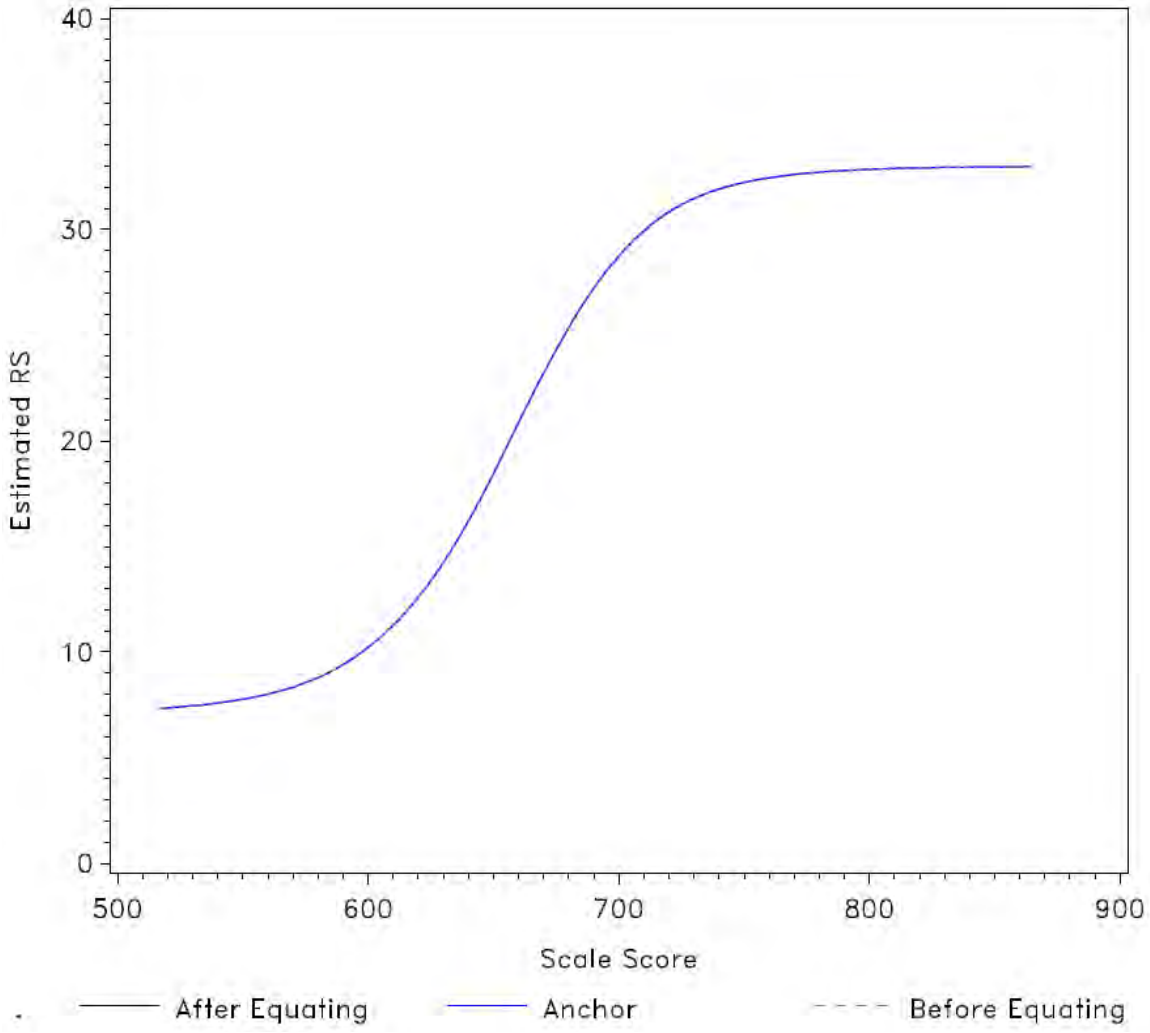


Figure 6. Communication Arts, Grade 8 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

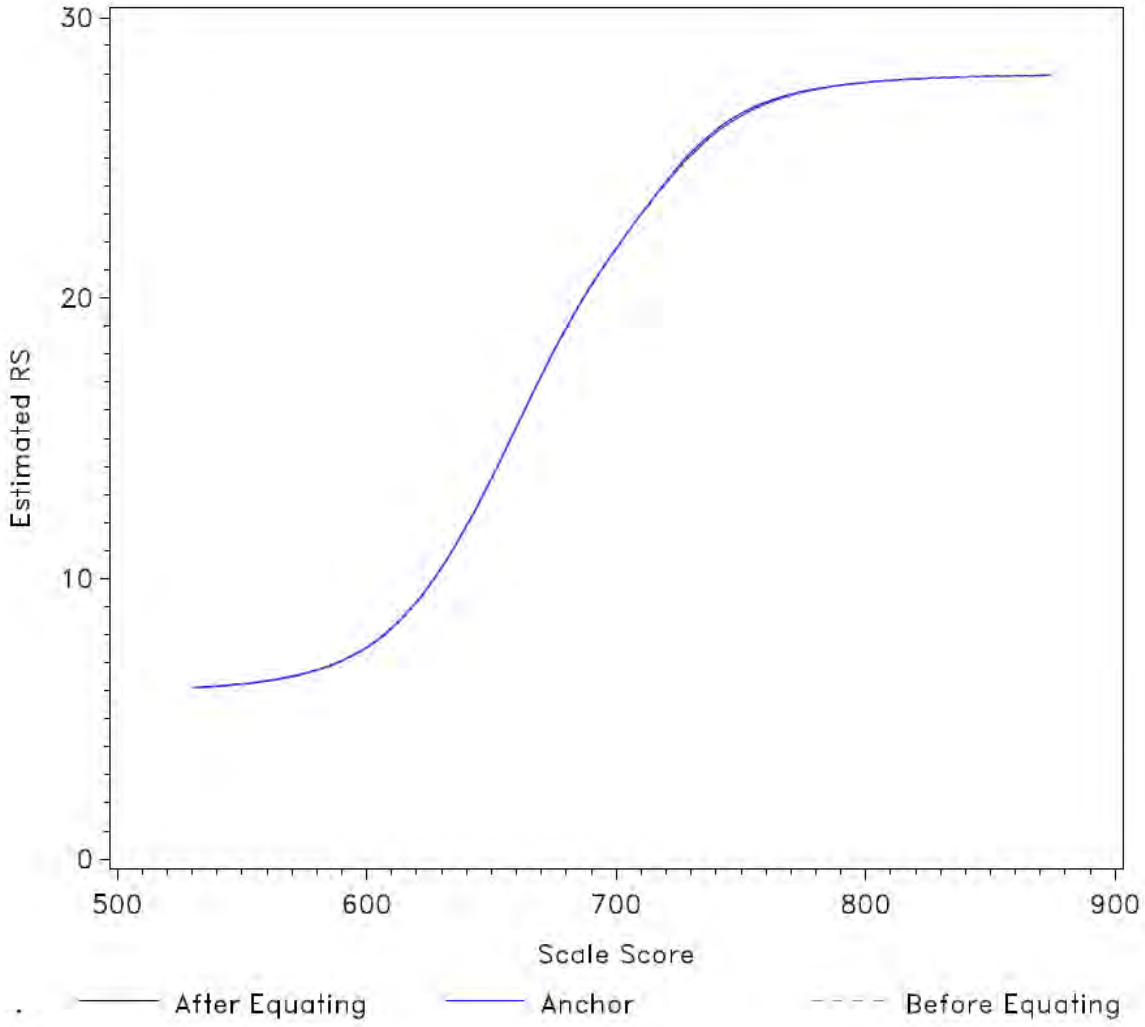


Figure 7. Mathematics, Grade 3 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

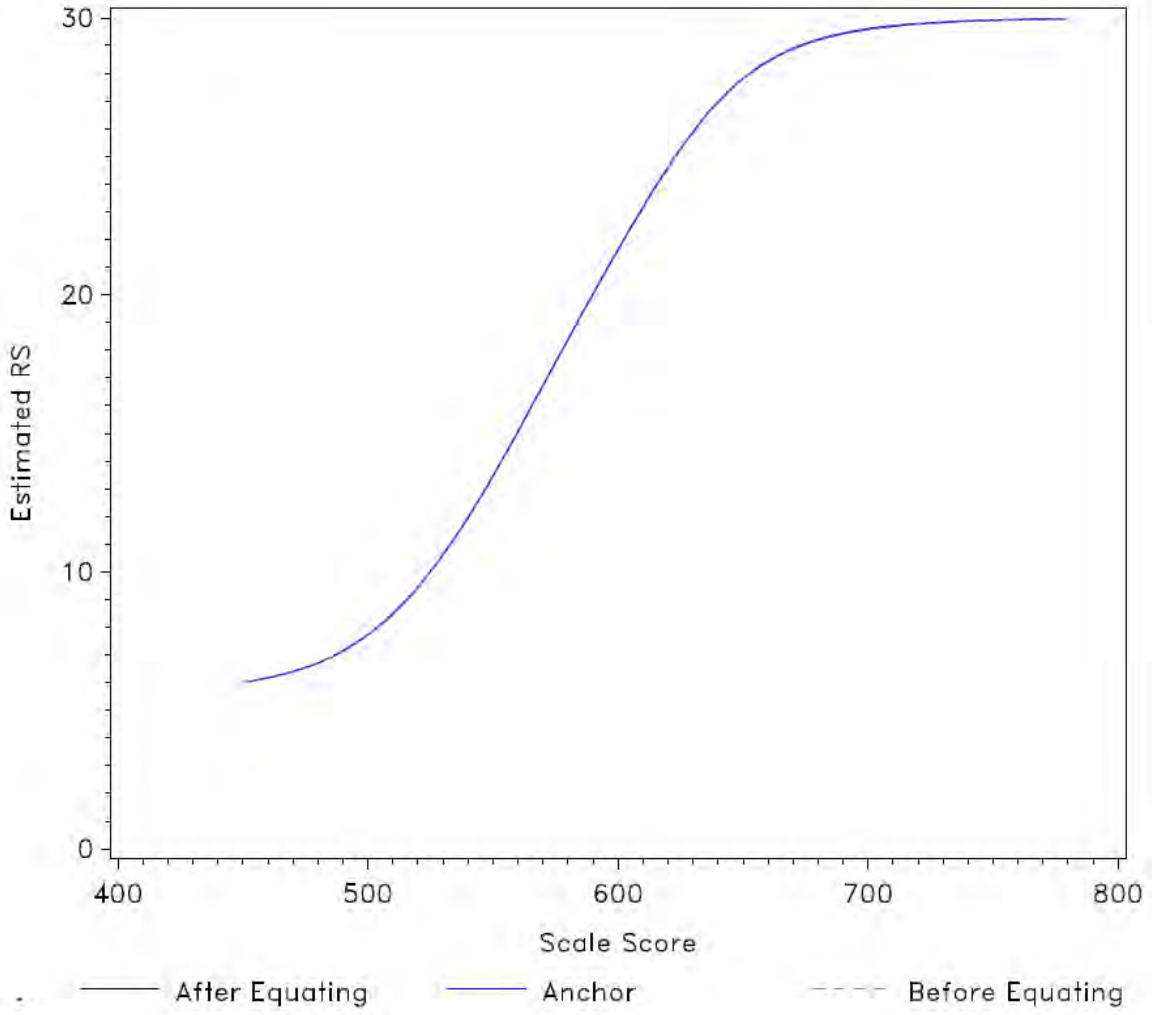


Figure 8. Mathematics, Grade 4 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

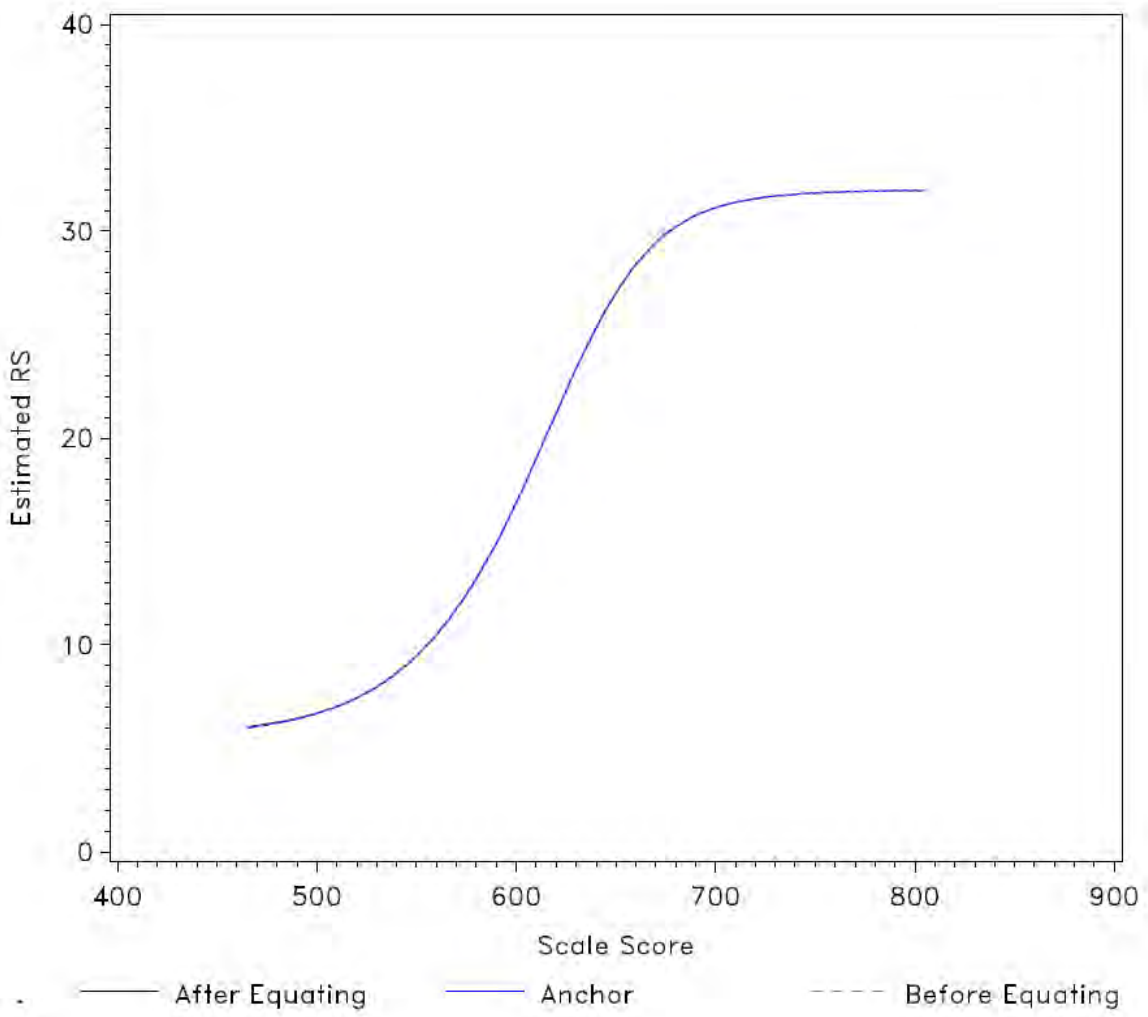


Figure 9. Mathematics, Grade 5 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

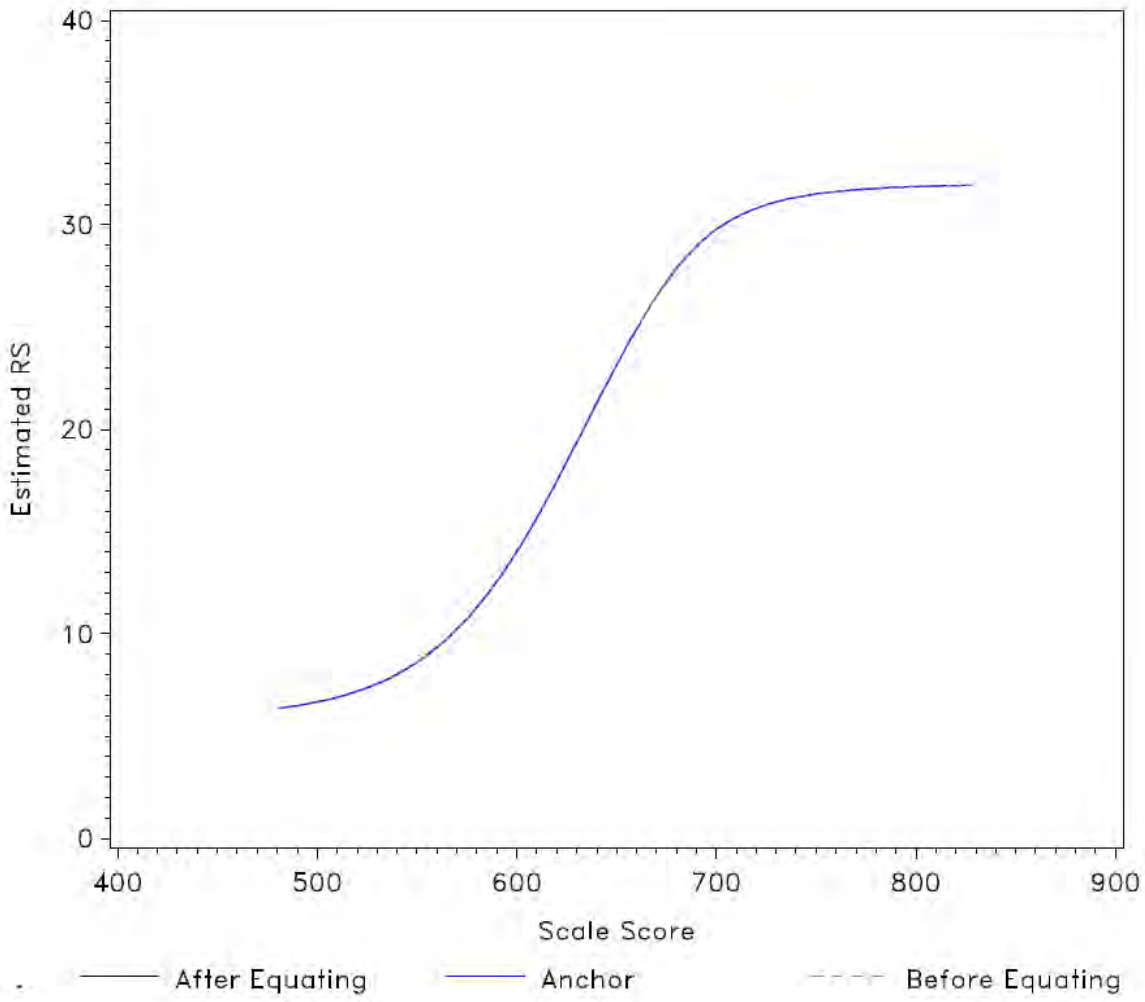


Figure 10. Mathematics, Grade 6 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

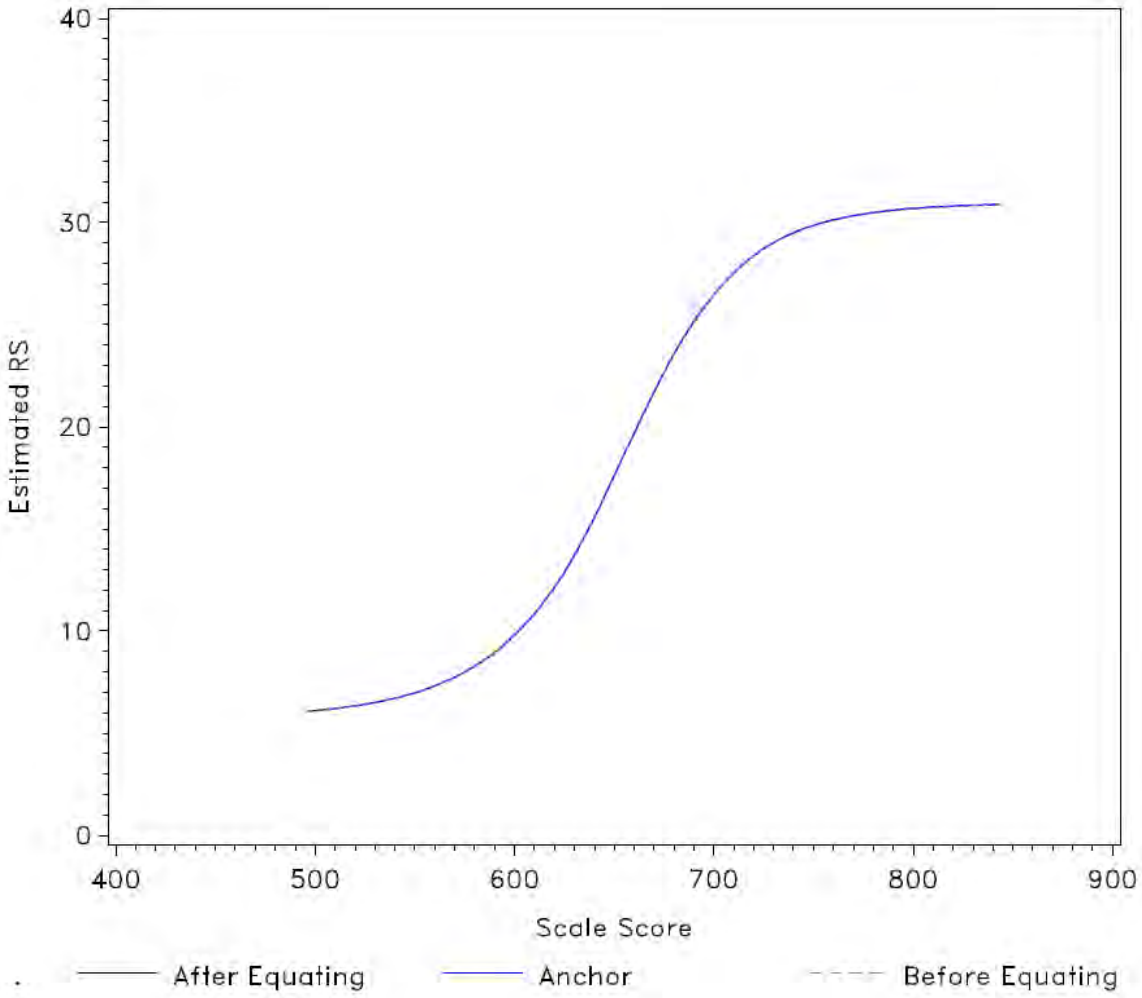


Figure 11. Mathematics, Grade 7 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

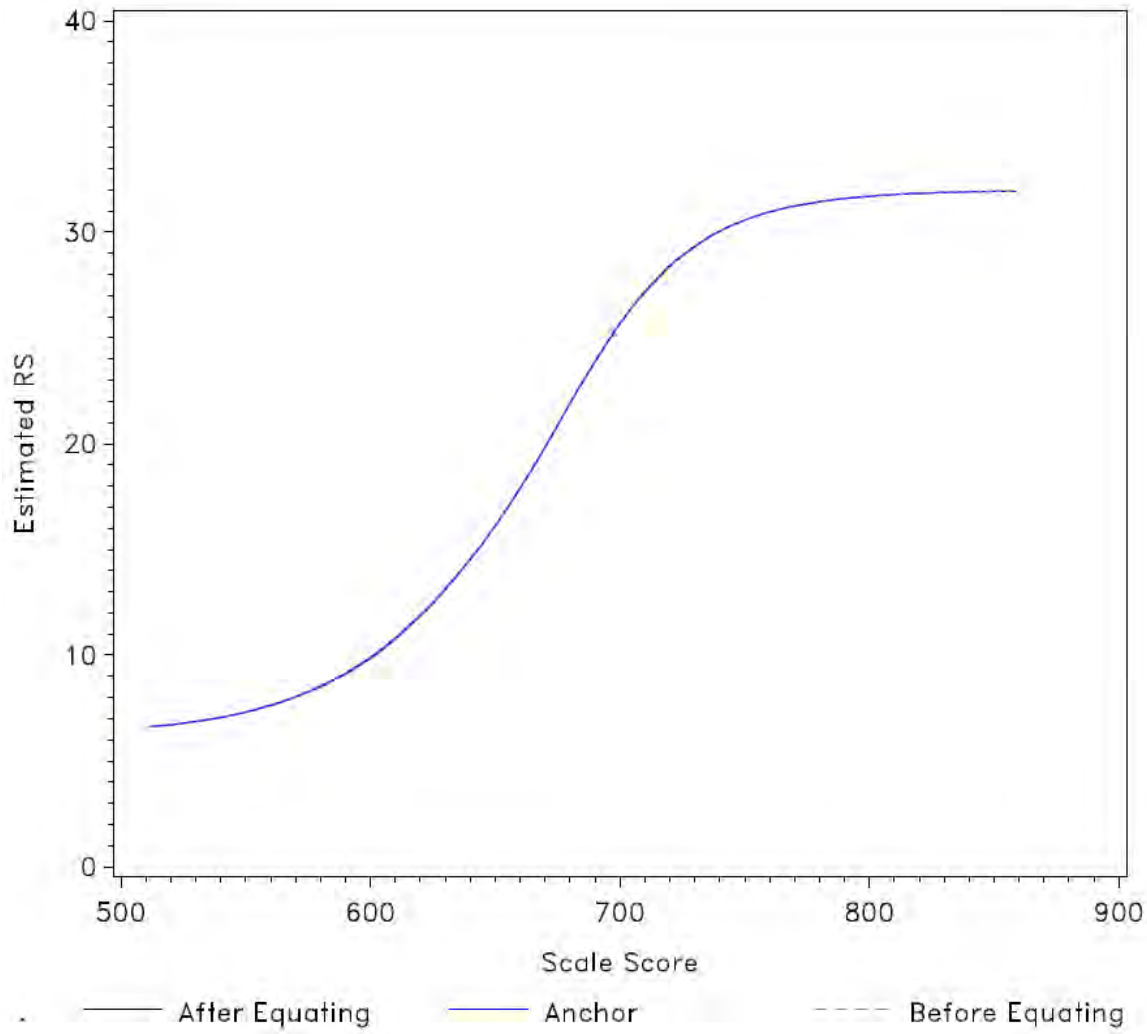


Figure 12. Mathematics, Grade 8 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

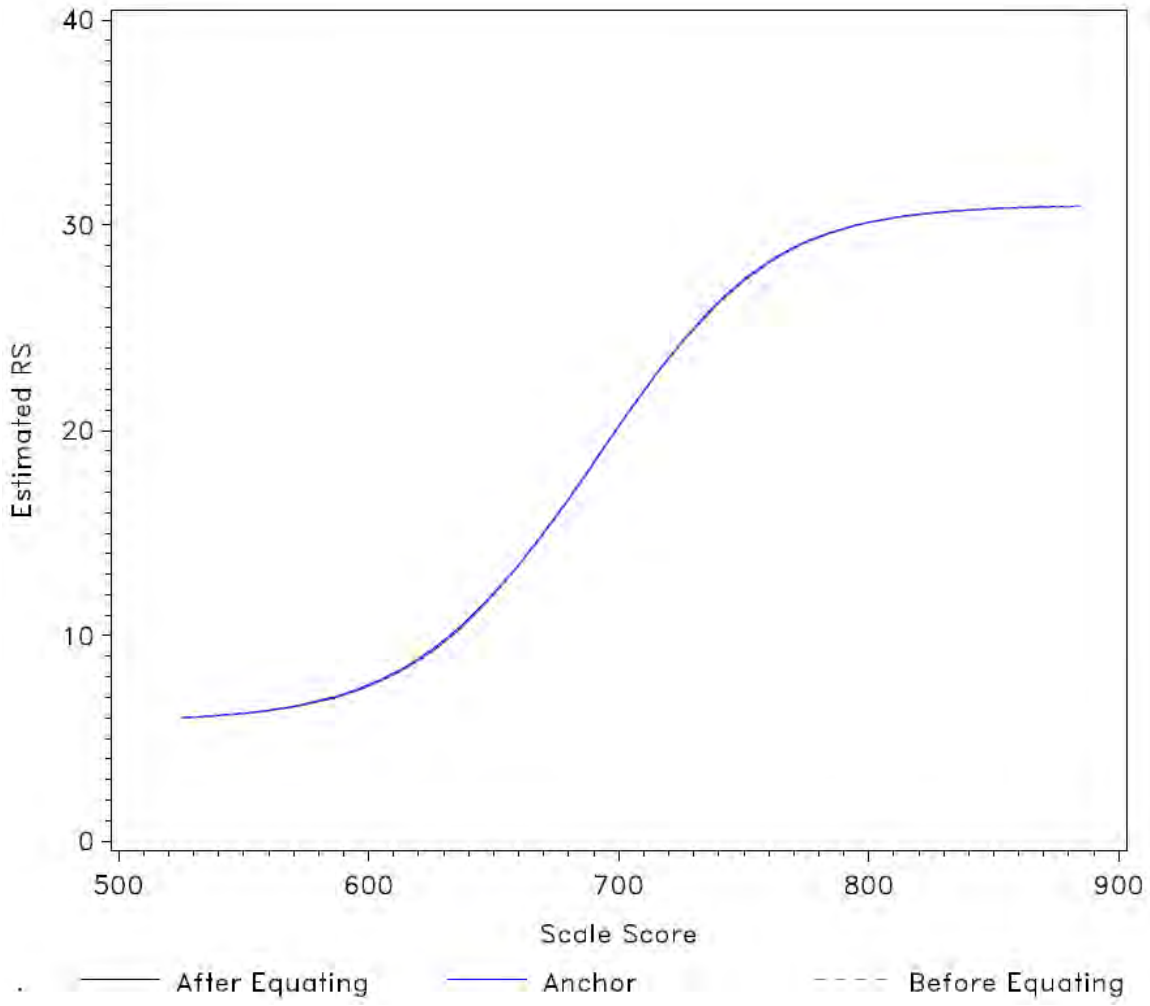


Figure 13. Science, Grade 5 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items

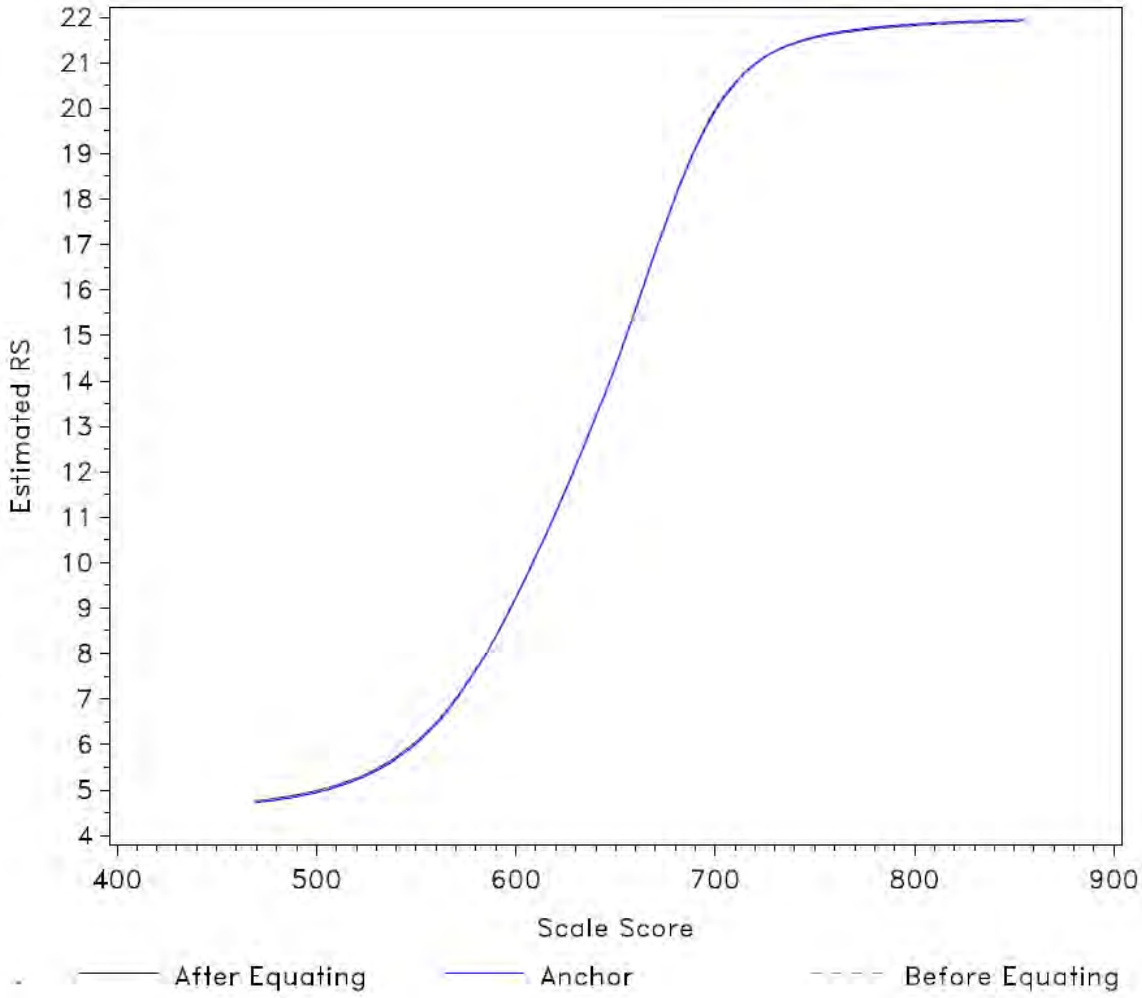
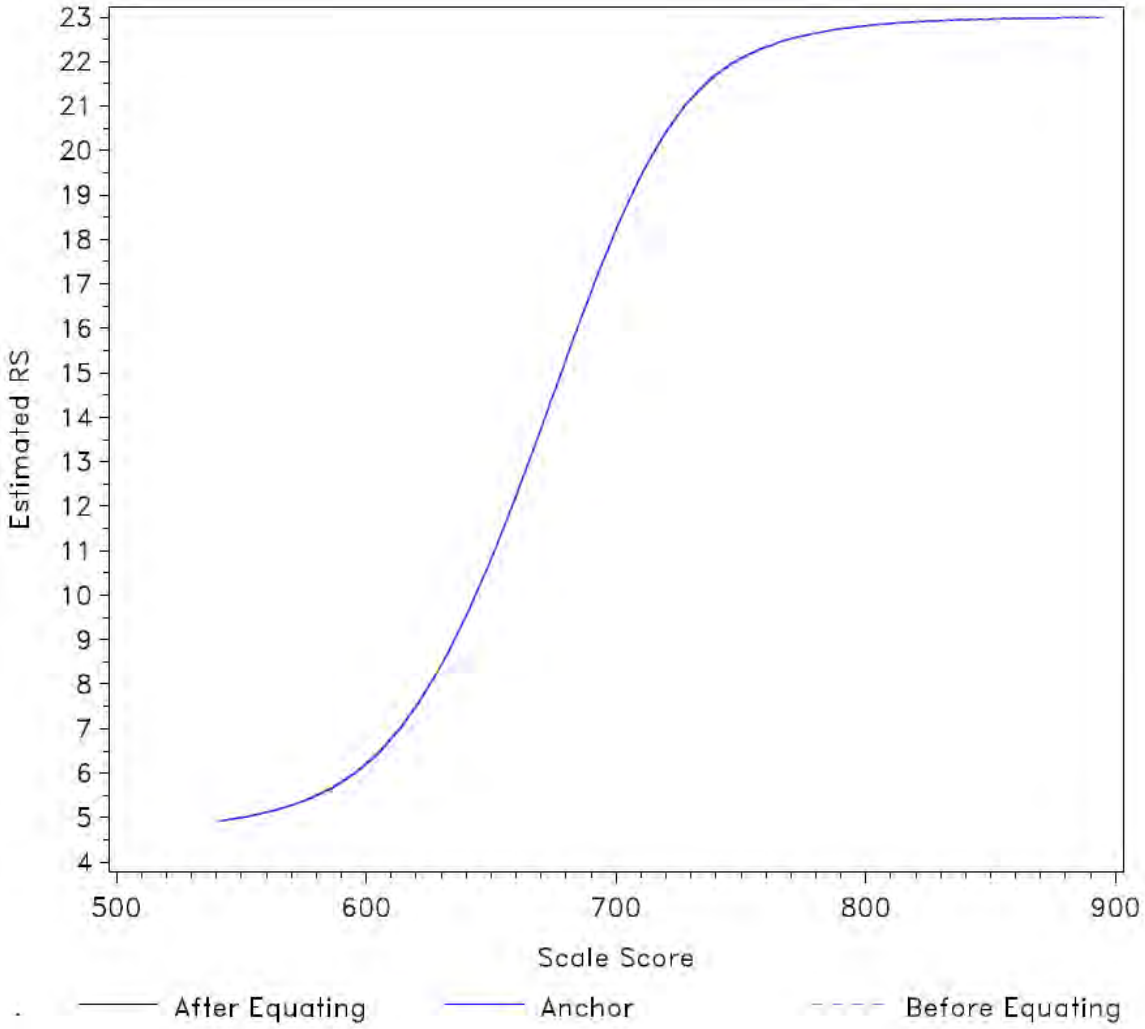


Figure 14. Science, Grade 8 Test Characteristic Curves (TCC) for the Inputted Anchor Items and for the Estimated Anchor Items



Appendix C

Examples of Score Reports

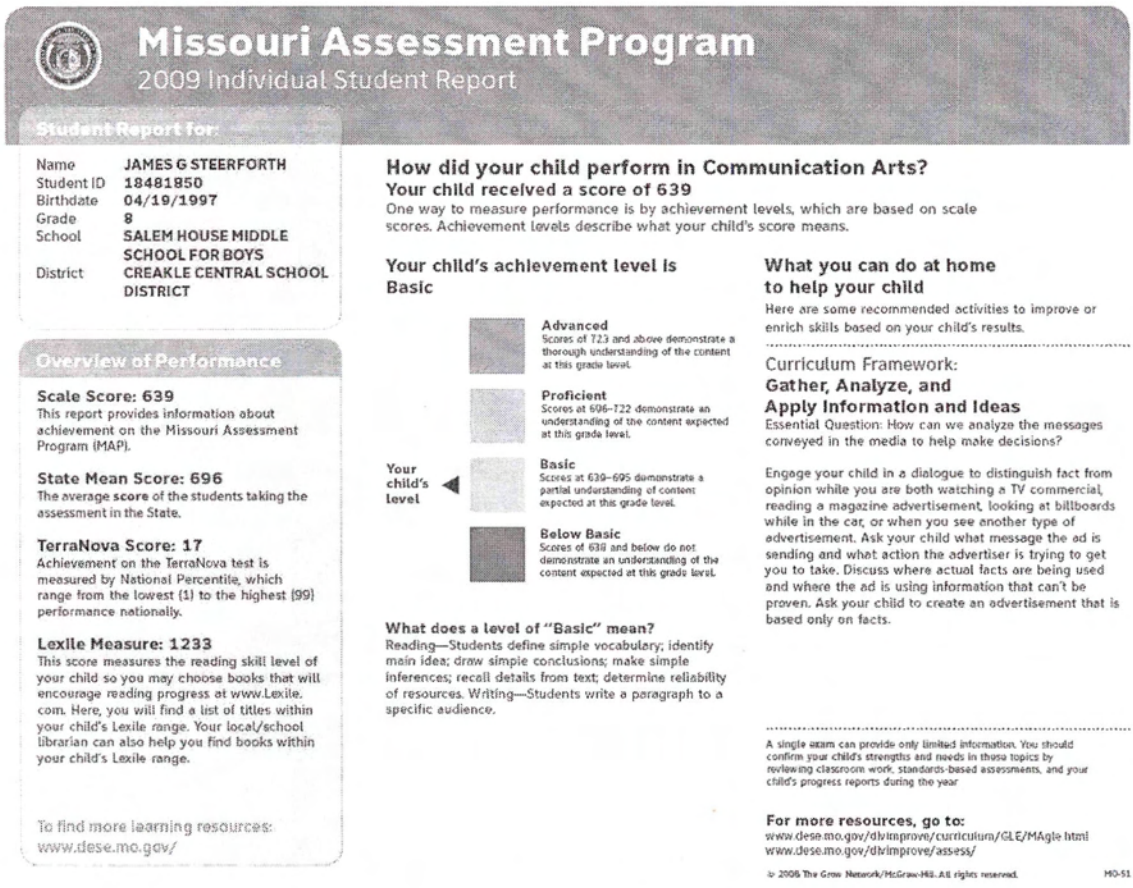


Figure C. 1 Example of Missouri Assessment Program Individual Score Report

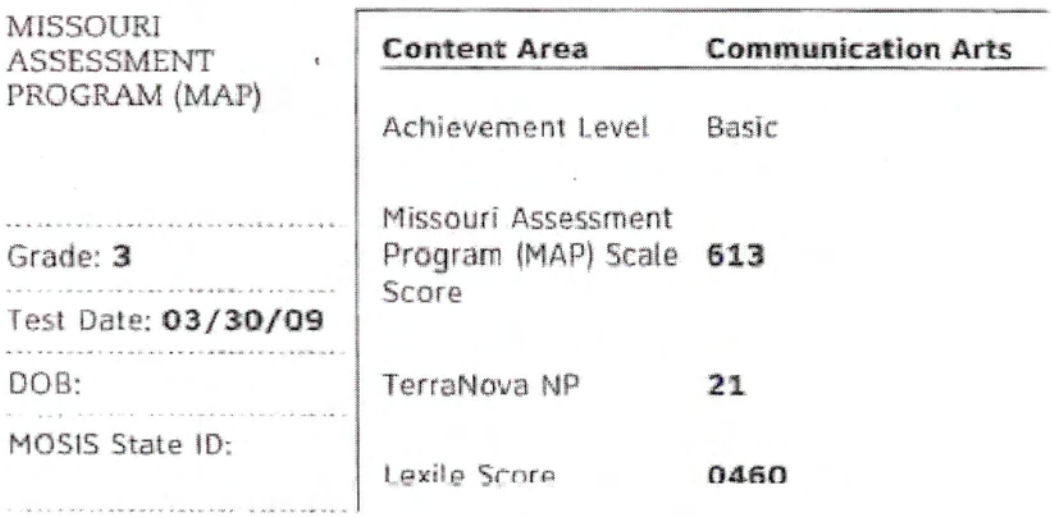


Figure C. 2 Example of Missouri Assessment Program Individual Student Label

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
MAP Student Summary Report

Communication Arts
03
2009

	MOSIS ID	DoB	Year	County District	Content Area	Assessed Grade	MAP Scale Score	EOC Admin Window	EOC Score	MAP A Score	Achievement Level	Terranova Ntnl. Percentile
			2009		CA	03	601				Basic	25
			2009		CA	03	631				Basic	75
			2009		CA	03	653				Proficient	66
			2009		CA	03	644				Basic	69
			2009		CA	03	669				Proficient	84
			2009		CA	03	618				Basic	39
			2009		CA	03	564				Below Basic	4
			2009		CA	03	648				Proficient	84
			2009		CA	03	651				Proficient	59
			2009		CA	03	643				Basic	76
			2009		CA	03	674				Advanced	92
			2009		CA	03	619				Basic	25
			2009		CA	03	602				Basic	28

Figure C. 3 Example of Missouri Assessment Program Crystal Report, MAP Scale Score Summary Report

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
MAP Student Demographics Report

Year	GR	Grade of Student	Name	DOB	Content	MOSIS ID	DIST ID	DIST < 1YR	BUILD < 1YR	LEP	RACE	SES	IEP	ELL/LEP <1 Yr & <3 Yr	LEP ELL Title 3	LEP ELL Mth US	DIS Diag	MAP-A	TITLE I
2009	03				CA		991				5	Y				000			
2009	03				CA		992				5					000			
2009	03				CA		993				5					000			
2009	03				CA		995				5					000			
2009	03				CA		996				5		Y			000	Speech Impair		
2009	03				CA		999				5	Y	Y			000	Speech Impair		
2009	03				CA		841				5	Y	Y			000	Mental Retard		
2009	03				CA		900				5	Y				000			
2009	03				CA		903				5	Y				000			
2009	03				CA		902				5					000			
2009	03				CA		948				5					000			
2009	03				CA		905				5	Y				000			
2009	03				CA		1046	Y	Y		5	Y				000			
2009	03				CA		886				5	Y				000			
2009	03				CA		907				5					000			
2009	03				CA		963				5	Y				000			

Figure C. 4 Example of Missouri Assessment Program Crystal Report, Missouri Student Demographic Report

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
Student Achievement Level

School Name

2009	YEAR	Content Area	Assessed Grade	Grade of Student	Full Name	Achievement Level	MOSIS ID
	2009	Communication Arts	07			Advanced	
	2009	Communication Arts	07			Advanced	
	2009	Communication Arts	08			Advanced	
	2009	Science	08			Advanced	
	2009	Communication Arts	07			Advanced	
	2009	Mathematics	07			Advanced	
	2009	Mathematics	07			Advanced	
	2009	Communication Arts	08			Advanced	
	2009	Mathematics	08			Advanced	
	2009	Science	08			Advanced	
	2009	Communication Arts	08			Advanced	
	2009	Mathematics	07			Advanced	
	2009	Communication Arts	08			Advanced	
	2009	Mathematics	08			Advanced	
	2009	Communication Arts	08			Advanced	

Figure C. 5 Example of Missouri Assessment Program Crystal Report, Student Achievement Level Report

Missouri Assessment Program

MAP Achievement 4 Levels
Percent of of Students Scoring Proficient or Advanced
Year: 2006

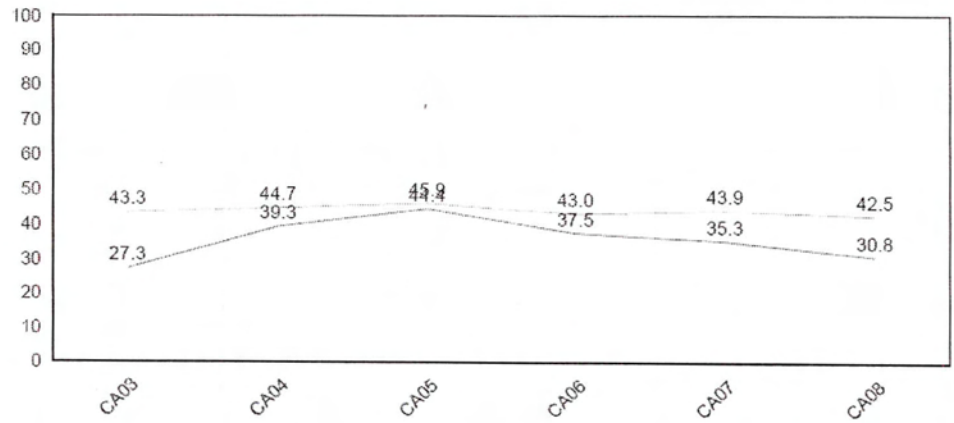


Figure C. 6 Example of Missouri Assessment Program Crystal Report, Achievement Level 4 Chart

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
MAP Achievement Level 4 Report

District Totals

Total	Content Area	Grade	Year	ACC	REP	LND	LND%	BB	% BB	Basic	% Basic	Prof	%Prof	Adv	%Adv	MAP Index*	Mean Scale	Median Terranova
Total	Com. Arts	03	2009	19	19	0	0.0	1	5.3	9	47.4	8	42.1	1	5.3	747.40	638.80	70.00
Total	Com. Arts	04	2009	18	18	0	0.0	2	11.1	7	38.9	7	38.9	2	11.1	750.00	653.20	59.00
Total	Com. Arts	05	2009	20	20	0	0.0	1	5.0	5	25.0	6	30.0	8	40.0	805.00	682.90	75.00
Total	Com. Arts	06	2009	22	22	0	0.0	3	13.6	11	50.0	7	31.8	1	4.5	727.30	663.10	60.00
Total	Com. Arts	07	2009	32	32	0	0.0	3	9.4	17	53.1	9	28.1	3	9.4	737.50	662.60	54.00
Total	Com. Arts	08	2009	14	14	0	0.0	1	7.1	6	42.9	2	14.3	5	35.7	778.60	703.40	81.00

ACC=Accountable; REP=Reportable; LND=Level Not Determined; Prof=Proficient; Adv=Advanced;
Top2 = Percent of Students Scoring Advanced or Proficient
MAP Index* = (pct BB*0)+(pct Basic*7)+(pct Proficient*8)+(pct Advanced*9)

Report as of: 9/30/2009

1

Figure C. 7 Example of Missouri Assessment Program Crystal Report, Achievement Level 4 Report

Missouri Department of Elementary and Secondary Education

Missouri Assessment Program

Content Standards Report

District				CS-1 Number and Operations	PP	CS- 2 Algebraic Relationshi pa	PP	CS-3 Geometric and Spatial Relationsh	PP	CS-4 Measurement	PP	CS-5 Data and Probability	PP
Mathematics 0000	Category/Type												
Mathematics	03 Total/Total	2009	80%	28	68%	12	75%	16	84%	8	70%	9	
Mathematics	04 Total/Total	2009	79%	21	75%	16	78%	14	69%	16	85%	11	
Mathematics	05 Total/Total	2009	79%	18	70%	13	68%	13	71%	14	65%	13	
Mathematics	06 Total/Total	2009	76%	17	59%	12	66%	12	63%	14	69%	13	
Mathematics	07 Total/Total	2009	66%	16	56%	13	56%	14	53%	13	64%	13	
Mathematics	08 Total/Total	2009	70%	14	41%	25	42%	13	48%	9	47%	14	

Average percentage of raw-score points earned by students on each standard.
PP = Points Possible.

Report as of: 10/14/2009

1

Figure C. 8 Example of Missouri Assessment Program Crystal Report, Content Standards Report

Missouri Assessment Program
Content Standards By Student
Missouri Department of Elementary and Secondary Education

Mathematics

Examiner Name:	Grade Level	Year	05-1 Number and Operations Points Possible-25	05-2 Algebraic Relationships Points Possible-12	05-3 Geometric and Spatial Relationships Points Possible-13	05-4 Measurement Points Possible-8	05-5 Data and Probability Points Possible-9
	03	2009	56%	33%	23%	75%	22%
	03	2009	100%	83%	77%	100%	78%
	03	2009	96%	83%	100%	100%	89%
	03	2009	80%	75%	69%	75%	89%
	03	2009	84%	75%	92%	88%	67%
	03	2009	76%	50%	62%	88%	67%
	03	2009	68%	75%	38%	50%	67%
	03	2009	92%	83%	92%	88%	56%
	03	2009	76%	58%	54%	88%	78%
	03	2009	88%	83%	92%	88%	78%
	03	2009	40%	42%	38%	63%	44%
	03	2009	88%	75%	92%	88%	67%
	03	2009	76%	50%	85%	75%	56%
	03	2009	84%	50%	85%	100%	78%
	03	2009	76%	67%	85%	88%	56%
	03	2009	92%	67%	77%	75%	89%
	03	2009	96%	92%	85%	100%	89%
	03	2009	64%	67%	100%	75%	56%
	03	2009	80%	83%	85%	88%	100%

Average percentage of raw-score points earned by students on each standard.

Report as of: 10/14/200
1

Figure C. 9 Example of Missouri Assessment Program Crystal Report, Content Standards Detail Report

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
Average Points Earned with Item Benchmark Descriptions
Content Standard IBD Extended

District School

2009 Communication Arts

SC	GR	Standard Desc.	GLE Code	GLE Description	D.O.K		Session/ Item	QT	Pts Poss	Avg Pts	% Earn
Administration:											
0000	03 1	speaking/writing standard English	W 2 E	In writing, use • correct spelling of simple compounds, homophones, contractions and words with affixes • standard spelling • classroom resources and dictionary to verify correct spelling	1	Recall	1 / 10	MC	1	0.47	47.00
0000	03 1	speaking/writing standard English	W 2 E	In writing, use • correct spelling of simple compounds, homophones, contractions and words with affixes • standard spelling • classroom resources and dictionary to verify correct spelling	1	Recall	1 / 11	MC	1	0.63	63.00
0000	03 1	speaking/writing standard English	W 2 E	In writing, use • correct spelling of simple compounds, homophones, contractions and words with affixes • standard spelling • classroom resources and dictionary to verify correct spelling	1	Recall	1 / 12	MC	1	0.26	26.00
0000	03 1	speaking/writing standard English	W 2 C	In composing text, use • correct ending punctuation in imperative and exclamatory sentences • comma in the greeting and closing of a letter	1	Recall	1 / 7	MC	1	0.53	53.00
0000	03 1	speaking/writing standard English	W 2 C	In composing text, use • correct ending punctuation in imperative and exclamatory sentences • comma in the greeting and closing of a letter	1	Recall	1 / 8	MC	1	0.95	95.00
0000	03 1	speaking/writing standard English	W 2 C	In composing text, use • correct ending punctuation in imperative and exclamatory sentences • comma in the greeting and closing of a letter	1	Recall	1 / 9	MC	1	0.74	74.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	1	Recall	3 / 10	MC	1	0.89	89.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	1	Recall	3 / 11	MC	1	0.74	74.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	2	Skill/Concept	3 / 18	MC	1	0.79	79.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	1	Recall	3 / 19	MC	1	0.63	63.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	1	Recall	3 / 20	MC	1	0.84	84.00
0000	03 1	speaking/writing standard English	W 1 A	Follow a writing process to • independently use a simple graphic organizer in pre-writing • generate a draft • routinely reread and revise work • routinely edit and proofread for capitalization and ending punctuation • independently publish writing	2	Skill/Concept	3 / 21	MC	1	0.89	89.00
0000	03 1	speaking/writing standard English	W 1 A	Follow a writing process to • independently use a simple graphic organizer in pre-writing • generate a draft • routinely reread and revise work • routinely edit and proofread for capitalization and ending punctuation • independently publish writing	2	Skill/Concept	3 / 27	MC	1	0.58	58.00
0000	03 1	speaking/writing standard English	W 1 A	Follow a writing process to • independently use a simple graphic organizer in pre-writing • generate a draft • routinely reread and revise work • routinely edit and proofread for capitalization and ending punctuation • independently publish writing	2	Skill/Concept	3 / 28	MC	1	0.47	47.00
0000	03 1	speaking/writing standard English	W 2 F	In composing text, identify and write sentences: • declarative • interrogative • imperative • exclamatory	1	Recall	3 / 38	MC	1	0.79	79.00
0000	03 1	speaking/writing standard English	W 2 D	Use parts of speech correctly in written text • verbs that agree with the subject • words that answer when, where, why and how questions (adverbs) • words to compare (adverbs)	1	Recall	3 / 39	MC	1	0.58	58.00

Figure C. 10 Example of Missouri Assessment Program Crystal Report, Content Standards IBD EX Report

Missouri Department of Elementary and Secondary Education
Missouri Assessment Program
Average Points Earned with Item Benchmark Descriptions
Goal Process Standard IBD Expanded

District **School**

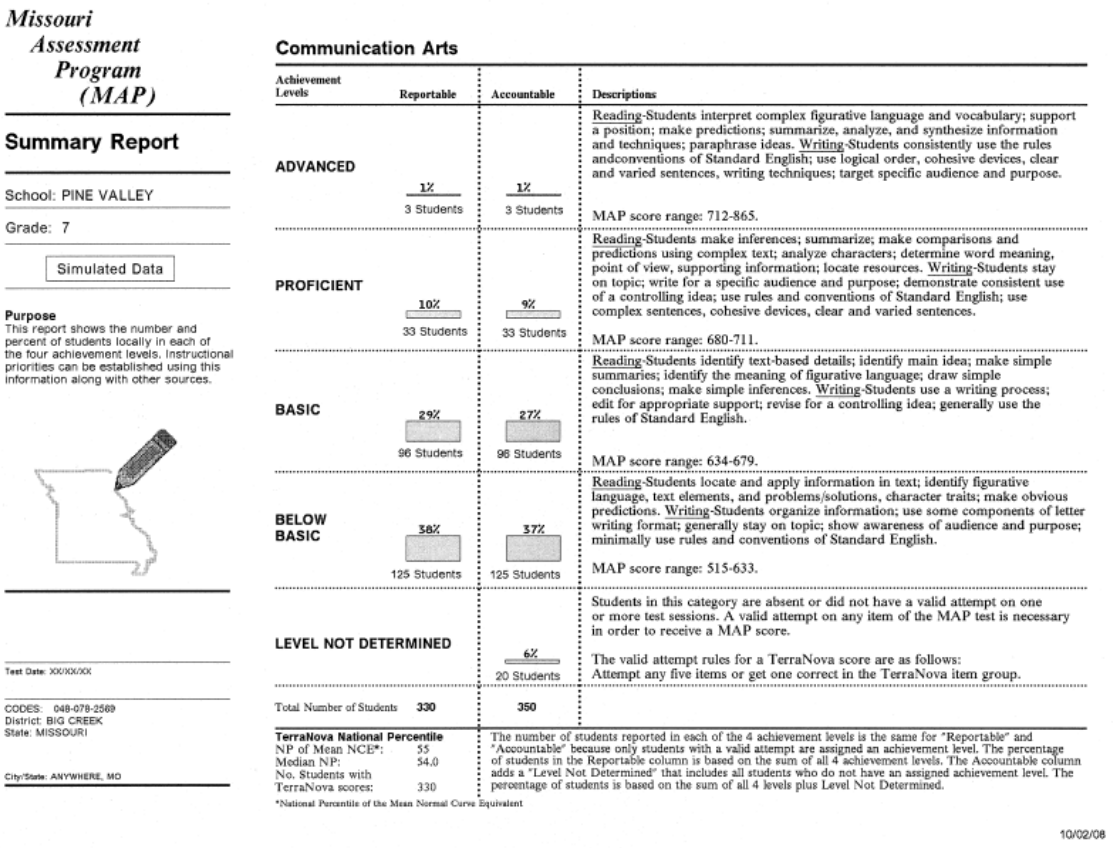
2009 **Communication Arts**

SC	GR	Goal	Standard Desc.	GLE	Code		D.O.K	Session/ Item	QT	Pts. Poss.	Avg. Pts.	% Earn
Administration:												
0000	03	1.5	comprehend/evaluate resources	R 3	C	Use details from text to • answer questions • retell main idea and important details • organize a sequence of events • identify simple cause and effect • draw conclusions • compare and contrast texts • identify author's purpose for writing text • make inferences about problems and solutions	1 Recall	3 / 1	MC	1	0.95	95.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 17	MC	1	0.53	53.00
0000	03	1.5	comprehend/evaluate resources	R 3	C	Use details from text to • answer questions • retell main idea and important details • organize a sequence of events • identify simple cause and effect • draw conclusions • compare and contrast texts • identify author's purpose for writing text • make inferences about problems and solutions	1 Recall	3 / 2	MC	1	0.95	95.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 22	MC	1	1.00	100.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 23	MC	1	0.79	79.00
0000	03	1.5	comprehend/evaluate resources	R 1	E	Develop vocabulary through text, using • base words • synonyms and antonyms • context clues • glossary • dictionary, with assistance	1 Recall	3 / 24	MC	1	0.84	84.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 29	MC	1	0.79	79.00
0000	03	1.5	comprehend/evaluate resources	R 3	C	Use details from text to • answer questions • retell main idea and important details • organize a sequence of events • identify simple cause and effect • draw conclusions • compare and contrast texts • identify author's purpose for writing text • make inferences about problems and solutions	1 Recall	3 / 3	MC	1	0.84	84.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 30	MC	1	0.37	37.00
0000	03	1.5	comprehend/evaluate resources	R 1	H	Apply post-reading skills to identify and explain the relationship between the main idea and supporting details • question to clarify • reflect • analyze • draw conclusions • summarize • paraphrase	1 Recall	3 / 33	MC	1	0.74	74.00
0000	03	1.5	comprehend/evaluate resources	R 1	E	Develop vocabulary through text, using • base words • synonyms and antonyms • context clues • glossary • dictionary, with assistance	1 Recall	3 / 35	MC	1	0.95	95.00
0000	03	1.5	comprehend/evaluate resources	R 2	C	Use details from text to • make inferences about setting, character traits and problem and solution • make predictions • draw conclusions • compare and contrast characters and changes in problems and settings • identify the narrator • identify cause and effect • identify events from the beginning, middle and end • identify author's purpose	1 Recall	3 / 36	MC	1	0.84	84.00

Report as of: 10/14/200

1

Figure C. 11 Example of Missouri Assessment Program Crystal Report, Goal Process Standards IBD EX Report



Missouri Assessment Program (MAP)

Summary Report

School: PINE VALLEY

Grade: 7

Simulated Data

Purpose
This report shows the number and percent of students locally in each of the four achievement levels. Instructional priorities can be established using this information along with other sources.

Test Date: XX/XX/XX

CODES: 048-078-2569
District: BIG CREEK
State: MISSOURI

City/State: ANYWHERE, MO

Communication Arts

Achievement Levels	Reportable	Accountable	Descriptions
ADVANCED	1% 3 Students	1% 3 Students	Reading-Students interpret complex figurative language and vocabulary; support a position; make predictions; summarize, analyze, and synthesize information and techniques; paraphrase ideas. Writing-Students consistently use the rules andconventions of Standard English; use logical order, cohesive devices, clear and varied sentences, writing techniques; target specific audience and purpose. MAP score range: 712-865.
PROFICIENT	10% 33 Students	9% 33 Students	Reading-Students make inferences; summarize; make comparisons and predictions using complex text; analyze characters; determine word meaning, point of view, supporting information; locate resources. Writing-Students stay on topic; write for a specific audience and purpose; demonstrate consistent use of a controlling idea; use rules and conventions of Standard English; use complex sentences, cohesive devices, clear and varied sentences. MAP score range: 680-711.
BASIC	29% 96 Students	27% 96 Students	Reading-Students identify text-based details; identify main idea; make simple summaries; identify the meaning of figurative language; draw simple conclusions; make simple inferences. Writing-Students use a writing process; edit for appropriate support; revise for a controlling idea; generally use the rules of Standard English. MAP score range: 634-679.
BELOW BASIC	38% 125 Students	37% 125 Students	Reading-Students locate and apply information in text; identify figurative language, text elements, and problems/solutions, character traits; make obvious predictions. Writing-Students organize information; use some components of letter writing format; generally stay on topic; show awareness of audience and purpose; minimally use rules and conventions of Standard English. MAP score range: 515-633.
LEVEL NOT DETERMINED		6% 20 Students	Students in this category are absent or did not have a valid attempt on one or more test sessions. A valid attempt on any item of the MAP test is necessary in order to receive a MAP score. The valid attempt rules for a TerraNova score are as follows: Attempt any five items or get one correct in the TerraNova item group.
Total Number of Students	330	350	

TerraNova National Percentile

NP of Mean NCE*:	55
Median NP:	54.0
No. Students with TerraNova scores:	330

The number of students reported in each of the 4 achievement levels is the same for "Reportable" and "Accountable" because only students with a valid attempt are assigned an achievement level. The percentage of students in the Reportable column is based on the sum of all 4 achievement levels. The Accountable column adds a "Level Not Determined" that includes all students who do not have an assigned achievement level. The percentage of students is based on the sum of all 4 levels plus Level Not Determined.

*National Percentile of the Mean Normal Curve Equivalent

10/02/08

Figure C. 12 Example of Missouri Assessment Program Summary Report

Layout of General Research File

MAP'09 GRT

Start	End	Length	Field	Values
			Hierarchical Data	
1	1	1	Mode level	1 = State 2 = District 3 = School 4 = Class
2	11	10	Organization ID	A-Z, 0-9
12	41	30	Element Name - District	Any character, blank
42	43	2	Element Structure Level Number - District	02
44	50	7	Element Number - District	0-9
51	53	3	District/Element Special Codes A-C (Region Code)	Any character, '.'
54	56	3	District/Element Special Codes D-F (District Code)	Any character, '.'
57	76	20	District Special Codes G-Z	Any character, '.'
77	78	2	Grade	03-08
79	108	30	City	Any character, blank
109	110	2	State	MO
111	140	30	Element Name- School	Any character, blank
141	142	2	Element Structure Level Number - School	03
143	149	7	Element Number - School	0-9
150	152	3	School/Element Special Codes A-C (Region Code)	Any character, '.'
153	155	3	School/Element Special Codes D-F (District Code)	Any character, '.'
156	159	4	School/Element Special Codes G-J (School Code)	Any character, '.'
160	175	16	School/Element Special Codes K-Z	Any character, '.'
176	205	30	Element Name- Class	Any character, blank
206	207	2	Element Structure Level Number- Class	04
208	214	7	Element Number-Class	0-9
215	217	3	Class/Element Special Codes A-C (Region Code)	Any character, '.'
218	220	3	Class/Element Special Codes D-F (District Code)	Any character, '.'
221	224	4	Class/Element Special Codes G-J (School Code)	Any character, '.'
225	240	16	Class/Element Special Codes K-Z	Any character, '.'
241	247	7	Student Element Number	0-9
248	275	28	Test Name	"Missouri Assessment Program"
276	278	3	*TerraNova Form/Level*: Communication Arts	D13 = Gr. 3, D14 = Gr. 4, D15 = Gr. 5, D16 = Gr. 6, D17 = Gr. 7 H18 = Gr. 8 blank
279	281	3	*TerraNova Form/Level*: Mathematics	D13 = Gr. 3, D14 = Gr. 4, D15 = Gr. 5, D16 = Gr. 6, D17 = Gr. 7 D18 = Gr. 8 blank
282	284	3	*TerraNova Form/Level*: Science	D15 = Gr. 5, D18 = Gr. 8 blank
285	290	6	Test Date (MMDDYY)	
			Special codes (Length 26)	See Definitions tab
291	300	10	MOSIS State ID	0-9
301	310	10	CTB Use	'.'
311	311	1	Race/ Ethnicity	*0 = Native American or Alaska Native* *1 = Asian/Pacific Islander* *3 = Black (not Hispanic)* *4 = Hispanic* *5 = White (not Hispanic)* *'-' = multi-mark* *'.' = blank*
312	314	3	CTB Use	'.'
315	315	1	Filler (formerly, Flag for Grade 11 Science book)	*Blank*
316	316	1	CTB Use	'.'

Start	End	Length	Field	Values
			User Defined Data	
			Accommodation - CA	
317	317	1	01 Braille edition	*Blank=Not Marked 0=Marked (Communication Arts)*
318	318	1	02 Large Print edition	*Blank=Not Marked 0=Marked (Communication Arts)*
319	319	1	04 Oral reading **– invalidates CA**	*Blank=Not Marked 0=Marked (Communication Arts)*
320	320	1	04 Oral reading – (Blind/Partial Sight)	*Blank=Not Marked 0=Marked (Communication Arts)*
321	321	1	05 Signing of assessment **– invalidates CA**	*Blank=Not Marked 0=Marked (Communication Arts)*
322	322	1	06 Paraphrasing **– invalidates all tests**	*Blank=Not Marked 0=Marked (Communication Arts)*
323	323	1	10 Other Administration	*Blank=Not Marked 0=Marked (Communication Arts)*
324	324	1	11 Oral reading in native language **– invalidates CA**	*Blank=Not Marked 0=Marked (Communication Arts)*
325	325	1	20 Extend time–*TerraNova* session	*Blank=Not Marked 0=Marked (Communication Arts)*
326	326	1	21 Administer using > allotted periods	*Blank=Not Marked 0=Marked (Communication Arts)*
327	327	1	22 Other timing	*Blank=Not Marked 0=Marked (Communication Arts)*
328	328	1	35 Use of scribe	*Blank=Not Marked 0=Marked (Communication Arts)*
329	329	1	39 Use of calculator, math table, etc.	*Blank=Not Marked 0=Marked (Communication Arts)*
330	330	1	43 Use of bilingual dictionary **- invalidates CA**	*Blank=Not Marked 0=Marked (Communication Arts)*
331	331	1	44 Other response	*Blank=Not Marked 0=Marked (Communication Arts)*
332	332	1	50 Testing individually	*Blank=Not Marked 0=Marked (Communication Arts)*
333	333	1	51 Testing in small group	*Blank=Not Marked 0=Marked (Communication Arts)*
334	334	1	53 Other setting	*Blank=Not Marked 0=Marked (Communication Arts)*
335	338	4	Blank for Future Use(4)	
			Accommodation - MA	

Start	End	Length	Field	Values
339	339	1	01 Braille edition	*Blank=Not Marked* *0=Marked* *Mathematics*
340	340	1	02 Large Print edition	*Blank=Not Marked* *0=Marked* *Mathematics*
341	341	1	04 Oral reading	*Blank=Not Marked* *0=Marked* *Mathematics*
342	342	1	05 Signing of assessment	*Blank=Not Marked* *0=Marked* *Mathematics*
343	343	1	06 Paraphrasing – **invalidates all tests**	*Blank=Not Marked* *0=Marked* *Mathematics*
344	344	1	10 Other Administration	*Blank=Not Marked* *0=Marked* *Mathematics*
345	345	1	11 Oral reading in native language	*Blank=Not Marked* *0=Marked* *Mathematics*
346	346	1	20 Extend time–*TerraNova* session	*Blank=Not Marked* *0=Marked* *Mathematics*
347	347	1	21 Administer using > allotted periods	*Blank=Not Marked* *0=Marked* *Mathematics*
348	348	1	22 Other timing	*Blank=Not Marked* *0=Marked* *Mathematics*
349	349	1	35 Use of scribe	*Blank=Not Marked* *0=Marked* *Mathematics*
350	350	1	39 Use of calculator, math table, etc.	*Blank=Not Marked* *0=Marked* *Mathematics*
351	351	1	43 Use of bilingual dictionary	*Blank=Not Marked* *0=Marked* *Mathematics*
352	352	1	44 Other response	*Blank=Not Marked* *0=Marked* *Mathematics*
353	353	1	50 Testing individually	*Blank=Not Marked* *0=Marked* *Mathematics*
354	354	1	51 Testing in small group	*Blank=Not Marked* *0=Marked* *Mathematics*
355	355	1	53 Other setting	*Blank=Not Marked* *0=Marked* *Mathematics*
356	360	5	Blank for Future Use(5)	
			Accommodation - SC	
361	361	1	01 Braille edition	*Blank=Not Marked* *0=Marked* *Science*
362	362	1	02 Large Print edition	*Blank=Not Marked* *0=Marked* *Science*

Start	End	Length	Field	Values
363	363	1	04 Oral reading	*Blank=Not Marked* *0=Marked* *Science*
364	364	1	05 Signing of assessment	*Blank=Not Marked* *0=Marked* *Science*
365	365	1	06 Paraphrasing – **invalidates all tests**	*Blank=Not Marked* *0=Marked* *Science*
366	366	1	10 Other Administration	*Blank=Not Marked* *0=Marked* *Science*
367	367	1	11 Oral reading in native language	*Blank=Not Marked* *0=Marked* *Science*
368	368	1	20 Extend time–*TerraNova* session	*Blank=Not Marked* *0=Marked* *Science*
369	369	1	21 Administer using > allotted periods	*Blank=Not Marked* *0=Marked* *Science*
370	370	1	22 Other timing	*Blank=Not Marked* *0=Marked* *Science*
371	371	1	35 Use of scribe	*Blank=Not Marked* *0=Marked* *Science*
372	372	1	39 Use of calculator, math table, etc.	*Blank=Not Marked* *0=Marked* *Science*
373	373	1	43 Use of bilingual dictionary	*Blank=Not Marked* *0=Marked* *Science*
374	374	1	44 Other response	*Blank=Not Marked* *0=Marked* *Science*
375	375	1	50 Testing individually	*Blank=Not Marked* *0=Marked* *Science*
376	376	1	51 Testing in small group	*Blank=Not Marked* *0=Marked* *Science*
377	377	1	53 Other setting	*Blank=Not Marked* *0=Marked* *Science*
378	392	15	Blank for Future Use	
			Teacher Invalidations	
393	393	1	Teacher Invalidation_CommArts_Session 1	*Blank= No Invalidation marked* *0 = Invalidated this session*
394	394	1	Teacher Invalidation_CommArts_Session 2	*Blank= No Invalidation marked* *0 = Invalidated this session*
395	395	1	Teacher Invalidation_CommArts_Session 3	*Blank= No Invalidation marked* *0 = Invalidated this session*
396	396	1	Teacher Invalidation_CommArts_Session 4	*Blank= No Invalidation marked* *0 = Invalidated this session*
397	397	1	Teacher Invalidation_Mathematics_Session 1	*Blank= No Invalidation marked* *0 = Invalidated this session*
398	398	1	Teacher Invalidation_Mathematics_Session 2	*Blank= No Invalidation marked* *0 = Invalidated this session*

Start	End	Length	Field	Values
399	399	1	Teacher Invalidation_Mathematics_Session 3	*Blank= No Invalidation marked* *0 = Invalidated this session*
400	400	1	Teacher Invalidation_Science_Session 1	*Blank= No Invalidation marked* *0 = Invalidated this session*
401	401	1	Teacher Invalidation_Science_Session 2	*Blank= No Invalidation marked* *0 = Invalidated this session*
402	402	1	Teacher Invalidation_Science_Session 3	*Blank= No Invalidation marked* *0 = Invalidated this session*
403	407	5	Blank for Future Use(5)	
			Absent in Session	
408	408	1	CA Absent Session 1	*Blank= No Absent marked* *0 = Absent this session*
409	409	1	CA Absent Session 2	*Blank= No Absent marked* *0 = Absent this session*
410	410	1	CA Absent Session 3	*Blank= No Absent marked* *0 = Absent this session*
411	411	1	CA Absent Session 4	*Blank= No Absent marked* *0 = Absent this session*
412	412	1	MA Absent Session 1	*Blank= No Absent marked* *0 = Absent this session*
413	413	1	MA Absent Session 2	*Blank= No Absent marked* *0 = Absent this session*
414	414	1	MA Absent Session 3	*Blank= No Absent marked* *0 = Absent this session*
415	415	1	SC Absent Session 1	*Blank= No Absent marked* *0 = Absent this session*
416	416	1	SC Absent Session 2	*Blank= No Absent marked* *0 = Absent this session*
417	417	1	SC Absent Session 3	*Blank= No Absent marked* *0 = Absent this session*
			Not Enrolled in Building for Content Area	
418	418	1	CA Not Enrolled in Building	*Blank=Not Marked* *0=Marked* *(Communication Arts)*
419	419	1	MA Not Enrolled in Building	*Blank=Not Marked* *0=Marked* *Mathematics*
420	420	1	SC Not Enrolled in Building	*Blank=Not Marked* *0=Marked* *Science*
421	423	3	Blank for Future Use(3)	*Blank*
			Student BIO Information	
424	431	8	Student Barcode	0-9, blank
432	439	8	Book Lithocode	0-9, blank
440	447	8	Book Security Barcode	A-Z, 0-9, blank
448	453	6	Birth Date (MMDDYY)	MM=01-12, DD=0-3,&0-9, YY=0,8,9, & 0-9 blank, '-' = multi-mark
454	454	1	Scoring: 'I' = IRT (Communication arts, Math, Science)	'I', blank
455	456	2	Quarter-month - '30' (for TerraNova - Sessions)	30, blank
457	471	15	Last name	A - Z, a - z, blank
472	491	20	First Name	A - Z, a - z, blank
492	492	1	Middle Initial	A - Z, a - z, blank
493	496	4	Chronological Age (in months) - right aligned	0-9, blank
497	497	1	Gender	blank=None marked, 'F'=Female 'M'=Male , '-' = both marked
498	507	10	State Use	0 through 9 for the Marked bubbles. Blank for the bubbles not marked.

Start	End	Length	Field	Values
508	527	20	Blank for Future Use(20)	blank
528	545	18	**Content Area Title - Communication Arts**	"Communication Arts", blank
546	548	3	1st Content Standard	Content Standard Scores - % of points earned - 000 through 100. If a student does not have a MAP score (a session not taken or absent or invalidated), all content standards will be reported as blanks. Content Standard scores will carry leading zeros.
549	551	3	2nd Content Standard	
552	554	3	3rd Content Standard	
555	557	3	4th Content Standard	
558	560	3	5th Content Standard	
561	563	3	6th Content Standard	
564	566	3	7th Content Standard	
567	569	3	8th Content Standard	
570	572	3	9th Content Standard	
573	575	3	10th Content Standard	
576	590	15	Reserved - Filler	blank
591	596	6	1st Process Standard reported	Process Standards Scores - % of pts. earned - 000 through 100; If a student does not have a MAP score (a session not taken or absent or invalidated), all process standards will be reported as blanks including Goal and Standard Number Process Standard scores will carry leading zeros. Position 1 = Goal number Positions 2 to 3 = Standard number Positions 4 to 6 = % of pts. earned
597	602	6	2nd Process Standard reported	
603	608	6	3rd Process Standard reported	
609	614	6	4th Process Standard reported	
615	620	6	5th Process Standard reported	
621	626	6	6th Process Standard reported	
627	632	6	7th Process Standard reported	
633	638	6	8th Process Standard reported	
639	644	6	9th Process Standard reported	
645	650	6	10th Process Standard reported	
651	656	6	11th Process Standard reported	
657	662	6	12th Process Standard reported	
663	668	6	13th Process Standard reported	
669	674	6	14th Process Standard reported	
675	680	6	15th Process Standard reported	
681	681	1	Achievement Level for Comm. Arts	0, 2, 3, 4, 5, blank. See definitions
682	684	3	MAP Scale Score	000-999, blank
685	687	3	CTB use - TerraNova Scale score	000-999, blank
688	691	4	CTB use - TerraNova Norm Year	"2005", blank
692	693	2	TerraNova NP score	00-99, blank
694	697	4	TerraNova Lexile Scores (Comm. Arts only)	0000-9999, blank
698	698	1	MAP Test Status - Communication Arts	blank=valid test. See definitions
699	728	30	Blank for Future Use(30)	
729	746	18	**Content Area Title - Mathematics**	"Mathematics", blank
747	749	3	1st Content Standard	Content Standard Scores - % of points earned - 000 through 100. If a student does not have a MAP score (a session not taken or absent or invalidated), all content standards will be reported as blanks. Content Standard scores will carry leading zeros
750	752	3	2nd Content Standard	
753	755	3	3rd Content Standard	
756	758	3	4th Content Standard	
759	761	3	5th Content Standard	
762	764	3	6th Content Standard	
765	767	3	7th Content Standard	
768	770	3	8th Content Standard	
771	773	3	9th Content Standard	
774	776	3	10th Content Standard	
777	791	15	11th Content Standard	
792	797	6	1st Process Standard reported	Process Standards Scores - % of pts. earned - 000 through 100; If a student does not have a MAP score (a session not taken or absent or invalidated), all process standards will be reported as blanks including Goal and Standard Number Process Standard scores will carry leading zeros.
798	803	6	2nd Process Standard reported	
804	809	6	3rd Process Standard reported	
810	815	6	4th Process Standard reported	
816	821	6	5th Process Standard reported	
822	827	6	6th Process Standard reported	
828	833	6	7th Process Standard reported	
834	839	6	8th Process Standard reported	
840	845	6	9th Process Standard reported	
846	851	6	10th Process Standard reported	
852	857	6	11th Process Standard reported	

Start	End	Length	Field	Values
858	863	6	12th Process Standard reported	
864	869	6	13th Process Standard reported	Position 1 = Goal number
870	875	6	14th Process Standard reported	Positions 2 to 3 = Standard number
876	881	6	15th Process Standard reported	Positions 4 to 6 = % of pts. earned
882	882	1	Achievement Level for Math	0, 2, 3, 4, 5, blank. See definitions
883	885	3	MAP Scale Score	000-999, blank
886	888	3	CTB use - TerraNova Scale score	000-999, blank
889	892	4	CTB use - TerraNova Norm Year	"2005", blank
893	894	2	TerraNova NP score	00-99, blank
895	895	1	MAP Test Status - Mathematics	blank=valid test. See definitions
896	925	30	Blank for Future Use(30)	
926	943	18	**Content Area Title - Science**	"Science", blank
944	946	3	1st Content Standard	Content Standard Scores - % of points earned - 000 through 100. If a student does not have a MAP score (a session not taken or absent or invalidated), all content standards will be reported as blanks. Content Standard scores will carry leading zeros.
947	949	3	2nd Content Standard	
950	952	3	3rd Content Standard	
953	955	3	4th Content Standard	
956	958	3	5th Content Standard	
959	961	3	6th Content Standard	
962	964	3	7th Content Standard	
965	967	3	8th Content Standard	
968	970	3	9th Content Standard	
971	973	3	10th Content Standard	
974	988	15	Reserved - Filler	blank
989	994	6	1st Process Standard reported	Process Standards Scores - % of pts. earned - 000 through 100; If a student does not have a MAP score (a session not taken or absent or invalidated), all process standards will be reported as blanks including Goal and Standard Number. Process Standard scores will carry leading zeros.
995	1000	6	2nd Process Standard reported	
1001	1006	6	3rd Process Standard reported	
1007	1012	6	4th Process Standard reported	
1013	1018	6	5th Process Standard reported	
1019	1024	6	6th Process Standard reported	
1025	1030	6	7th Process Standard reported	
1031	1036	6	8th Process Standard reported	
1037	1042	6	9th Process Standard reported	
1043	1048	6	10th Process Standard reported	
1049	1054	6	11th Process Standard reported	
1055	1060	6	12th Process Standard reported	
1061	1066	6	13th Process Standard reported	Position 1 = Goal number
1067	1072	6	14th Process Standard reported	Positions 2 to 3 = Standard number
1073	1078	6	15th Process Standard reported	Positions 4 to 6 = % of pts. earned
1079	1079	1	Achievement Level for Science	0, 2, 3, 4, 5, blank. See Definitions tab
1080	1082	3	MAP Scale Score	000-999, blank
1083	1085	3	CTB use - TerraNova Scale score	000-999, blank
1086	1089	4	CTB use - TerraNova Norm Year	"2005", blank
1090	1091	2	TerraNova NP score	00-99, blank
1092	1092	1	MAP Test Status - Science	blank=valid test. See Definitions tab
1093	1122	30	Blank for Future Use(30)	blank
			Item Response	See Definitions tab for item suppression information
1123	1152	30	Communication Arts Session 1	
1153	1212	60	Communication Arts Session 2	
1213	1272	60	Communication Arts Session 3	
1273	1292	20	Communication Arts Session 4	
1293	1352	60	Mathematics Session 1	
1353	1412	60	Mathematics Session 2	
1413	1442	30	Mathematics Session 3	
1443	1482	40	Science Session 1	
1483	1542	60	Science Session 2	
1543	1602	60	Science Session 3	

Appendix D

Missouri Assessment Program Achievement Level Descriptors

GRADE 3−SHORT DESCRIPTORS

Below Basic
Reading—Students locate information in text; identify an obvious main idea; define simple words and phrases. Writing—Students show minimal awareness of beginning, middle, end, audience, purpose and controlling idea; attempt to create friendly letters; use graphic organizers.

Basic
Reading—Students make simple comparisons; recall simple sequence of events; make obvious inferences and predictions; use context clues to determine word meaning. Writing—Students use basic parts of speech correctly in simple sentences; show minimal awareness of beginning, middle, end, audience, purpose and controlling idea.

Proficient
Reading—Students locate/identify supporting details, obvious cause and effect; make inferences; use context clues to determine word meaning; make comparisons; recall detailed sequence of events; identify solutions and fact vs. fiction; recognize figurative language; draw obvious conclusions. Writing—Students generally use rules of Standard English; show awareness of audience, purpose, controlling idea, relevant details, beginning, middle and end.

Advanced
Reading—Students identify relevant/supporting information to make predictions and draw conclusions; infer word meaning; infer main idea; make complex comparisons; make complex inferences; categorize information; identify correct sequence of events. Writing—Students consistently apply rules of Standard English; construct complex sentences; use details effectively; have a clear controlling idea, awareness of audience and purpose, beginning, middle and end.

GRADE 4—SHORT DESCRIPTORS

Below Basic
Reading—Students locate information in text; recall stated information; draw obvious conclusions; make simple comparisons and descriptions. Writing—Students write simple letters, minimally use the rules of Standard English; attempt to organize information.

Basic
Reading—Students identify appropriate details; use context clues; make obvious inferences; select vocabulary using context clues. Writing—Students write simple letters with an awareness of an intended audience and purpose; generally use the rules of Standard English.

Proficient
Reading—Students make simple inferences; recall, identify, and use relevant information; draw conclusions; explain figurative language and main idea; use context clues to select vocabulary; identify character traits, sensory details, and simple cause and effect. Writing—Students show organization and awareness of an intended audience and purpose; use the rules of Standard English; use a writing process to revise, edit, and proofread.

Advanced
Reading—Students make complex inferences and comparisons; evaluate simple information; infer cause/effect and word meaning; interpret figurative language; identify author's purpose; identify complex problems/solutions; explain complex main ideas. Writing—Students consistently use the rules of Standard English.

GRADE 5—SHORT DESCRIPTORS

Below Basic
Reading—Students locate/identify information in text; draw simple conclusions; make obvious inferences and predictions; identify character traits. Writing—Students use correct letter writing format; partially organize information.

Basic
Reading—Students identify supporting details, problems/solutions; use context clues; make obvious inferences; give partial summary of action. Writing—Students edit for Standard English.

Proficient
Reading—Students interpret figurative language; infer main idea; identify author's purpose, point of view, the sequence of information, cause/effect, the meaning of vocabulary; summarize; distinguish between fact and opinion; draw conclusions; make inferences and comparisons; support a position. Writing—Students use the rules of Standard English; construct complex sentences; edit for appropriate support, organize information.

Advanced
Reading—Students interpret and draw conclusions from complex information; analyze complex characters; infer author's purpose and word meaning; categorize information; make simple evaluations and judgments; determine the appropriateness of a source and the accuracy of information. Writing—Students consistently use the rules of Standard English; use a writing process to organize information.

GRADE 6—SHORT DESCRIPTORS

Below Basic
Reading—Students locate/identify information in text; make simple inferences; identify main idea, sensory information, figurative language, simple problems or solutions. Writing—Students show awareness of audience and letter format; use simple organizational techniques and graphic organizers; use simple rules of Standard English.

Basic
Reading—Students identify supporting information, simple cause/effect relationships, conflicts, point of view and problem-solving processes. Writing—Students use correct letter writing format; generally use the rules of Standard English including spelling; revise; have a controlling idea.

Proficient
Reading—Students identify author's purpose, supporting details, point of view; describe character traits, plot; identify problems/solutions; support a position with text-based details; draw conclusions; interpret figurative language; make inferences and predictions; locate resources. Writing—Students use the rules of Standard English; construct complex sentences; write for an audience and purpose; organize information.

Advanced
Reading—Students make complex connections; analyze complex characters; evaluate the accuracy and importance of information; draw conclusions and make inferences from complex information, analyze complex characters; determine cause and effect; paraphrase. Writing—Students demonstrate consistent use of a controlling idea and Standard English.

GRADE 7—SHORT DESCRIPTORS

Below Basic
Reading—Students locate and apply information in text; identify figurative language, text elements, and problems/solutions, character traits; make obvious predictions. Writing—Students organize information; use some components of letter writing format; generally stay on topic; show awareness of audience and purpose; minimally use rules and conventions of Standard English.

Basic
Reading—Students identify text-based details; identify main idea; make simple summaries; identify the meaning of figurative language; draw simple conclusions; make simple inferences. Writing—Students use a writing process; edit for appropriate support; revise for a controlling idea; generally use the rules of Standard English.

Proficient
Reading—Students make inferences; summarize; make comparisons and predictions using complex text; analyze characters; determine word meaning, point of view, supporting information; locate resources. Writing—Students stay on topic; write for a specific audience and purpose; demonstrate consistent use of a controlling idea; use rules and conventions of Standard English; use complex sentences, cohesive devices, clear and varied sentences.

Advanced
Reading—Students interpret complex figurative language and vocabulary; support a position; make predictions; summarize, analyze, and synthesize information and techniques; paraphrase ideas. Writing—Students consistently use the rules and conventions of Standard English; use logical order, cohesive devices, clear and varied sentences, writing techniques; target specific audience and purpose.

GRADE 8—SHORT DESCRIPTORS

Below Basic
Reading—Students identify author's purpose, figurative language, plot, and setting; use context clues to choose vocabulary. Writing—Students create a graphic organizer; write a basic paragraph; show some awareness of audience.

Basic
Reading—Students define simple vocabulary; identify main idea; draw simple conclusions; make simple inferences; recall details from text; determine reliability of resources. Writing—Students write a paragraph to a specific audience.

Proficient
Reading—Students summarize; infer vocabulary meaning and cause/effect; interpret figurative language; analyze text features; follow multi-step directions; identify author's technique; analyze text; make inferences, interpretations, predictions, comparisons, using complex material; evaluate evidence, reliability of resources. Writing—Students edit for relevant details and purpose; organize and edit text; consistently use rules/conventions of Standard English.

Advanced
Reading—Students analyze complex information, author's purpose, characters; synthesize information; summarize complex ideas; make complex inferences. Writing—Students edit text correctly applying the rules/conventions of Standard English.

Mathematics

GRADE 3—SHORT DESCRIPTORS

Below Basic
Students use multiplication to model situations; recognize that addition and subtraction are inverse operations; add 2-digit numbers; apply subtraction skills; extend shapes or numbers in a pattern; use number sentences to model situations; use transformations to check congruency of shapes; recognize a line of symmetry; use an appropriate unit on a ruler to measure length; estimate length; interpret information from graphs.

Basic
Students estimate with less-than and greater-than; sort items by size; apply regrouping for adding and subtracting 3-digit numbers; order 3-digit whole numbers; count using numbers and pictures; identify and explain a pattern; use an appropriate unit of measurement; read thermometers; read analog clocks to nearest 5 minutes; use a ruler to measure to the nearest centimeter; compare data; transfer data to graphs.

Proficient
Students identify odd/even numbers; locate landmark numbers; describe change using increase/decrease; perform basic division of 2-digit whole numbers; identify and locate fractional parts; set up/solve simple word problems; recognize 2-D and 3-D shapes; combine 3-D solids; identify 2-D faces of 3-D objects; determine perimeter of polygons; identify appropriate units of measure; add monetary values up to $5.00; use calendars to determine dates; estimate length with fractions.

Advanced
Students estimate and justify results of addition/subtraction of numbers; represent a mathematical situation as a number sentence or an expression; identify multiple lines of symmetry; determine change from $5.00 including different combinations of coins; predict events as likely or unlikely.

GRADE 4—SHORT DESCRIPTORS

Below Basic
Students write and compare decimals to the hundredths place; identify fraction as a part of a whole; describe the results of combining shapes; identify parallel lines; estimate linear measurements; read and compare data on a bar graph; complete tables; create tables or graphs to represent data.

Basic
Students use multiplication to solve problems; analyze patterns using words, tables, and graphs; identify the missing value in a number sentence; identify 2-D and 3-D shapes and attributes; identify the results of transformations; tell time to the nearest minute; use benchmarks to estimate linear measurements; transfer numerical data to a graph; propose and justify conclusions that are based on data.

Proficient
Students compare parts of a whole as fractions; identify place value up to 6-digit whole numbers; decompose/compose whole numbers; represent multiplication using sets/arrays; divide 3-digit by 1-digit numbers; write a number sentence; describe movement on grid using geometric vocabulary; identify lines of symmetry; use standard/metric units to measure; add/subtract money values to $10.00; determine area on grid; read/interpret data on a line plot; analyze and explain data.

Advanced
Students describe constant rates of change; identify strategies to solve problems; describe numeric and geometric patterns; solve problems using graphs, tables, or number sentences; construct a figure with one line of symmetry; determine differences in measures; estimate measurement of angles; determine change from $10.00; identify equivalent linear measures within a system; count combinations of items.

GRADE 5—SHORT DESCRIPTORS

Below Basic
Students recognize equivalent representations of numbers by composing and decomposing numbers up to 5 digits; order decimals to thousandths place; interpret place value to hundred-thousands; determine operations used in numeric patterns; use symmetry to complete figures; make generalizations about geometric patterns; describe attributes of 2-D shapes; identify data on a line graph; make and justify predictions using data; describe, compare, and organize data in a bar graph.

Basic
Students identify place value to the millions place; read, write, and compare unit fractions and decimals to the thousandths place; identify lines of symmetry; identify appropriate units of area; identify appropriate units of measure; use data to create a bar graph and perform calculations using numbers between given intervals.

Proficient
Students multiply decimals to the hundredths place; use estimation in computations; divide 3-digit by 2-digit numbers; add fractions with like denominators; solve problems involving rates of change; extend numeric patterns; complete number sentences; identify faces of 3-D and similar figures; interpret direction on a coordinate grid; calculate area using a grid; compute elapsed time in hours; analyze data in line graphs and tables; explain the probability of a simple event.

Advanced
Students use addition/subtraction of money in a real-world situation; explain and justify the results of calculations; justify and model the results of calculations involving constant rates; use number sentences to model a mathematical situation; analyze characteristics of and identify 3-D figures, quadrilaterals, and angle measures; use a coordinate grid to describe paths and determine distances between points; convert between standard units of measurement.

GRADE 6—SHORT DESCRIPTORS

Below Basic
Students compare and order integers, positive rational numbers, and percents; describe patterns in tables and pictures; identify properties of 2-D and 3-D shapes; identify acute, obtuse, or right angles; identify transformations of 2-D shapes; identify equivalent algebraic expressions using the associative property; read and interpret line and circle graphs.

Basic
Students generate equivalent forms of percents, fractions and decimals; determine a rule for a geometric or numeric pattern; use coordinate geometry to construct and identify 2-D shapes using ordered pairs; use models to compare and explain probabilities; estimate and interpret data in graphs.

Proficient
Students add/subtract positive rational numbers; identify least common multiple and greatest common factor; estimate quotients; determine rate of increase; analyze rates of change; use variables; compare spatial views of 3-D objects; construct polygons; describe transformations; determine area of rectangles; measure angles; convert within a system of measure; interpret and complete a table based on probability; compare/explain data; calculate measures of center.

Advanced
Students estimate and convert measurements; describe solutions to algebraic equations; recognize similarities between 2-D shapes; use properties of basic figures to draw conclusions about angle size; determine area of triangles; solve elapsed time problems; apply formula for perimeter; estimate area of a figure using a coordinate grid; interpret stem-and-leaf plots; determine appropriate data collection methods and questions; interpret data to solve problems.

GRADE 7—SHORT DESCRIPTORS

Below Basic
Students place integers on a number line; identify shapes from a group of 2-D shapes based on a common property; transform 2-D shapes; analyze precision and accuracy using measurement tools; identify unit of measure for volume; interpret bar graphs; use representations of data from bar graphs, circle graphs, stem-and-leaf plots, and box-and-whisker plots; predict outcomes using probability.

Basic
Students multiply and divide positive rational numbers; identify bases and exponents of numbers in exponential form; recognize equivalent numerical representations; solve 2-step problems; use variables to solve inequalities and equations; analyze patterns represented numerically or graphically; read and interpret graphs.

Proficient
Students read/write numbers up to hundred-millions place; compare integers, rational numbers, percents; perform operations with mixed numbers; use circle graphs to recognize relationship of parts to whole; solve fraction/decimal/percent problems; solve proportion/scale problems; use models to solve problems; model with equations; describe and classify 2-D/3-D shapes; apply spatial reasoning to estimate area; solve time problems; solve area problems; calculate measures of center.

Advanced
Students calculate totals involving percents in multi-step problems; extend non-linear patterns; model with inequalities; apply the relationship of corresponding and similar angles; use scale factors on a grid to dilate shapes; describe corresponding angles and sides of similar polygons; solve problems using time conversions; find circumference and area of circles; make conversions using proportions.

GRADE 8—SHORT DESCRIPTORS

Below Basic
Students generalize numeric patterns; generalize relationships between attributes of 2-D shapes; identify the results of subdividing 3-D shapes; identify 3-D figures using a 2-D representation; solve problems involving area; use scales to estimate distance; interpret graphs; find the mean value of a data set; select graphical representations of data; interpret data; make conjectures based on theoretical probability.

Basic
Students perform operations with rational numbers; solve and interpret one-step linear equations; extend geometric patterns; generalize patterns to find a specific term; identify relationships in 3-D objects; calculate the theoretical probability of an event; interpret a scatter plot to determine the relationship between two variables.

Proficient
Students identify equivalent representations of a number; identify mental strategies to solve problems; solve multi-step equations; use symbolic algebra; identify transformations; classify angles; create similar polygons; use coordinate geometry; solve problems involving area; identify appropriate units of measure; convert standard units within a system of measurement; interpret graphic organizers; calculate measures of center.

Advanced
Students estimate the value of square roots; write numbers using scientific notation; solve two-step inequalities; analyze slope and intercept in linear equations; apply the Pythagorean Theorem using coordinate geometry; identify polygons based on their attributes; identify coordinates of vertices of a transformed polygon; use a protractor to measure angles; solve problems involving surface area; select, create, and use appropriate graphical representation of data.

Science

GRADE 5—SHORT DESCRIPTORS

Below Basic
Students identify the relationship between mass and force; classify bodies of water; identify weather instruments and their uses; identify characteristics of the solar system; compare amounts/measurements given in a simple format; identify appropriate tools for simple scientific measurements; identify how technological advances may be helpful to humans.

Basic
Students explain the relationship between mass and force; describe how specialized body structures help animals survive; match environments to the plants and animals they support; identify environmental problems and find solutions; construct part of a graph; determine the appropriate scientific tool and its function in an investigation; determine how technological advances address problems and enhance life.

Proficient
Students describe changes in properties of matter; identify uses of simple machines; explain how work is done; identify forces of magnetism; describe the motion of objects; identify plant parts and their functions; classify vertebrates and invertebrates; classify producers, consumers, or decomposers; predict changes in food chains; identify the effects of human activities on other organisms; describe the Sun as a source of light and heat, or the moon as a reflector of light; explain the day/night cycle; identify characteristics and variables of a fair test; interpret data and make predictions; draw conclusions based on evidence; distinguish between man-made and natural objects; apply problem solving skills to a situation.

Advanced
Students identify energy transformations; predict the effect of heat energy on water; diagram a complete electrical circuit; predict how simple machines affect the force needed to do work; describe the effects of weathering and erosion on Earth's surface; describe relationships in weather data; explain how the Sun's position and the length and position of shadows relate to the time of day; interpret and apply knowledge from a data table; identify appropriate steps, tools and metric units in an investigation; construct a graph and plot data; formulate a question for an investigation.

GRADE 8—SHORT DESCRIPTORS

Below Basic
Students identify simple terms related to matter and energy; demonstrate beginning understanding of properties of light and how it travels; identify structures of plants and animals needed for survival; identify levels of organization in multicellular organisms; read simple graphs and make simple data comparisons.

Basic
Students identify an example of a force; demonstrate simple understanding of how traits are passed from one generation to the next; have a basic understanding of climate; identify a simple hypothesis; recognize a trend in a data table; demonstrate some awareness of how various factors influence and are influenced by science and technology.

Proficient
Students classify types of motion; calculate the speed of an object; demonstrate simple understanding of life processes; classify and/or show relationships between organisms; explain how adaptations help organisms survive; explain how species are affected by environmental change; understand and describe a food web; explain rock and fossil evidence of changes in the Earth; explain how Earth's systems interact; draw conclusions from tables or graphs; demonstrate basic understanding of the solar system; recognize the need for, and calculate, averages; understand the importance of constants in investigations; use appropriate tools and methods to collect data; describe tools and discoveries that advance scientific knowledge.

Advanced
Students explain the physical and chemical properties of matter; apply knowledge of energy and energy transfer; demonstrate understanding of physical and chemical processes of organisms; evaluate the effects of balanced and unbalanced forces; predict the impact of environmental change in ecosystems; justify how adaptations help organisms survive; demonstrate understanding of the water cycle; compare and contrast weather and climate; explain the cause of seasons on Earth; demonstrate understanding of the solar system; apply the concept of light years; construct a complete graph; evaluate experimental design; create testable questions and hypotheses; apply awareness of the influence of science and technology in society.