Missouri

Assessment Program

Grade-Level Assessments

Science Grades 5 and 8

Technical Report 2015

DRAFT

Submitted to
Missouri Department of Elementary and Secondary Education
February 2016

Table of Contents

Table of Tables

Table of Figures

EXECUTIVE SUMMARY

This report is a technical summary of the 2015 administration of the Science component of Missouri Assessment Program (MAP). The Science MAP is a summative assessment administered in Grades 5 and 8. These tests are designed to measure students' knowledge of Science and they are aligned with Missouri Learning Standards. The 2015 Science tests were created by Iowa Testing Programs (ITP), a research, development, and outreach unit in the College of Education at the University of Iowa, and the Missouri Department of Elementary and Secondary Education (DESE). These assessments included Missouri-owned items and items supplied by ITP and licensed to DESE. Except for Braille, large print, and accommodated paper-and-pencil forms, the Science assessments were administered online. This section provides a summary of the Spring 2015 MAP Science Grades 5 and 8 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 its testing program to comply with the requirements of No Child Left Behind. Starting in 2008, Science tests were administered for the first time. Science assessments are aligned with Science Grade-Level Expectations (GLE) (for details refer to http://dese.mo.gov/college-career-readiness/curriculum/proposed-missouri-learning-standards-update/current-missouri). Further details of the development of the 2014–15 Science MAP tests can be found in Chapter 3 of this report.

E.2 Administration

In the spring of 2015, Missouri administered summative assessments in Science to students in Grades 5 and 8. The assessment was administered from March 30 to May 22, 2015. Test administration is discussed in Chapter 4 of this report.

Approximately 560 districts and charter schools administered Science MAP tests in Grades 5 and 8. Table E.1 shows participation rates based on Missouri student 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 attempted to take the online test or received a test book. The Accountable columns show the total number of students who attempted to take the online test or received a test book. The Percent Reportable columns show the percentage of students who received a scale score on the Science MAP. Further analysis of participation rates is provided in Chapter 7 of this report.

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

E.3 Student Performance

This is the eighth year of the Science MAP testing program. Table E.2 presents the percentage of students classified as *Proficient* or *Advanced* in 2008 through 2015 on the Science assessments. Approximately 47% of students in Grade 5 and 49% of students in Grade 8 were classified as *Proficient* or *Advanced* in Science after the Spring 2015 test administration. 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 the MAP test scores. The scores are used to identify students' strengths and weaknesses in Missouri's student performance; to inform stakeholders (teachers, school administrators, district administrators, DESE staff members, parents, and the public) about the status of the progress toward meeting academic achievement standards of the state; and to meet the requirements of the state's accountability program.

Evidence of validity based on test content was supported by the test specifications, including the test design and test blueprint. Missouri Science assessments were developed in alignment with Missouri Learning Standards and Grade Level Expectations. They were built using Missouri Science item pool and Iowa Test of Basic Skills item pool developed by the University of Iowa.

With exceptions of Braille forms, large print forms, and a limited number of paper-and-pencil test forms, Science assessments were administered on-line in a standardized manner further supporting score validity. Accommodations and designated supports were available for students for whom such aids were deemed appropriate and indicated in their Individualized Education Programs.

Scoring of technology-enhanced and constructed-response items followed predefined scoring criteria. The technology-enhanced items were auto-scored. Constructed-response items were scored by human readers. The inter-rater reliability statistics demonstrated that the items were scored reliably.

The test scaling and linking was conducted using the item response theory methodology. The 2015 Science assessments were equated to MAP Science scales and students were scored using item parameters estimated after the 2015 test administration. The item response theory models used for test scaling were appropriate for the test data supporting the operational data analysis and ensuring that the test items, as well as the overall tests, were functioning appropriately. The cut scores used for classification of students into different performance levels and associated achievement level descriptors were established during standard setting in a collaborative and participatory process further supporting the validity and interpretation of the MAP Science scores.

Evidence for Science assessment construct-related validity—the meaning of test scores and the inferences they support—was provided through reliability, convergent validity, and divergent validity studies. The reliability analysis results indicated that the Science tests produce scores that would be relatively stable if the tests were administered repeatedly under similar conditions. The assumption that the Science MAP tests were unidimensional (that is the grade level test measured one primary dimension) was confirmed through principal component analysis. The divergent

validity of the Science tests was evaluated through the correlations computed between the Science scale scores and the English Language Arts/Literacy and Mathematics scores. The student scores were found to be highly but not perfectly related to each other suggesting that while different constructs are being measured, the three assessments may also be tapping into a similar knowledge base or general underlying ability. In addition, test fairness was evaluated through differential item functioning analysis and analysis of differences in test performance among subgroups.

Table E.1: Participation Rates: All Students

Grade	Accountable in Science	Percent Reportable in Science
5	66412	99.95%
8	66526	99.86%

Table E.2: Percentage of Students Classified as *Proficient* or *Advanced* in 2008 through 2015 Using Census Data: Science

	Science								
Grade	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2015 – 2014**
5	44.5	45.1	48.9	50.5	51.4	51.3	47.3	47.0	-0.3
8	43.2	44.8	48.0	50.0	49.6	50.1	51.9	48.9	-3.0

CHAPTER 1: INTRODUCTION

The 2015 Missouri Assessment Program (MAP) marked the eighth administration of the grade-level Science test at Grades 5 and 8. The assessment is designed to measure students' knowledge of Science. This report provides a technical overview of the Science assessments of the 2014–15 MAP. As such, it presents evidence for the validity of the 2014–15 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 of the current report.

1.1 Background of the Missouri Science Assessment Program

In 2001, the federal No Child Left Behind (NCLB) legislation, which required states to develop grade-level tests in both Reading and Mathematics to be administered annually in Grades 3 through 8 and once in Grades 10 through 12, was enacted. 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 2007–08 school year. In accordance with 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 Missouri Department of Elementary and Secondary Education (DESE) contracted with CTB/McGraw-Hill (CTB) in 2005 to construct Science assessments 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 2010 operational Science forms. The contract to create Science assessments was renewed in 2007 and lasted until 2014. DESE contracted with ITP to develop the Science assessments for the 2014-15 test administration. Data Recognition Corporation (DRC)/CTB supported the test 2014-15 administration, data analysis, student scoring, and reporting. Table 1.1 shows a timeline of the development history of the Science testing program.

1.2 Purpose of the Missouri Assessment Program

The MAP Science tests are designed to measure how well students acquire the skills and knowledge described in the Missouri Grade-Level Expectations (GLE). 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 and to gauge the overall quality of education throughout Missouri.

1.3 Design of the Missouri Assessment Program

Two regular Science test forms were administered in each grade level. In addition, Braille and large print test forms were constructed for each grade to enable visually impaired students to participate in Science testing. In Grade 5, the Braille and large print forms consisted of the same items as one of the regular forms. In Grade 8, the Braille and large print forms contained some items that were not common with regular forms.

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 details that confirm that the processes and procedures applied in the Science MAP adhere 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 Science assessments. 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 composed of evidence that supports the use of the Science scores. In Chapter 2, we discuss some of the uses of the Science scores.

Item and Test Development (Chapter 3)
Chapter 3 of the Technical Report provides a summary of the test development activities that occurred to create the Spring 2015 operational test forms 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 of Constructed-Response and Technology-Enhanced Items (Chapter 5)
Chapter 5 of the Technical Report describes the processes and activities for scoring constructed-response and technology-enhanced items. This chapter also discusses the measures for training raters and 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 2015 Science MAP, which are composed of three major parts: the classical item analysis; calibration, scaling, and linking using item response theory (IRT) models; and student scoring. This chapter also describes the demographics of the calibration samples 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 2015 Science 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 districts.

Standard Setting (Chapter 8)
Chapter 8 of the Technical Report briefly discusses standard setting and describes the cut score review activities that occurred for the Science tests in the spring of 2015.

Reliability and Validity Evidence (Chapter 9)
Chapter 9 of the Technical Report provides evidence of reliability and validity of the Science 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 the Science scores.

Fairness (Chapter 10)
Chapter 10 of the Technical Report discusses fairness and how the Science tests are constructed to be fair to all Missouri students. This chapter summarizes the results of the differential item functioning (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 the Grade-Level MAP: Science

Year	Event
2004	Grade-Level Expectations published
2007	Science field test
2008	First operational Science MAP
2008	Standard setting for Science
2008	Version 2.0 Grade-Level Expectations (GLEs) published
2009	Last operational administration of Science MAP based on V1.0 GLEs
2010	First operational administration of Science MAP based on V2.0 GLEs
2011-2015	Subsequent operational administrations of Science MAP based on V2.0 GLEs
2015	Cut score validation for Science

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* (hereafter the *Standards*; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014):

> 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. Different components of validity evidence … include 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 test takers, as appropriate to the test interpretation in question. (22)

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 Science test scores is provided in this Technical Report. In this section, we examine some possible uses of the Science 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 of evidence that the Science MAP tests measure what they are intended to measure. However, this Technical Report cannot anticipate all possible interpretations and uses of the test 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 Science test scores.

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 the MAP Science scores include the following:

- evaluating student strengths and weaknesses
- communicating expectations for all students
- evaluating school-, district-, and state-level programs
- informing stakeholders (teachers, school administrators, district administrators, DESE staff members, parents, and the public) about the status of the progress toward meeting academic achievement standards of the state
- meeting the requirements of the state's accountability program, the Missouri School Improvement Program

This Technical Report refers to the use of the test-level scores: scale scores and achievement levels.

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 Science. Test-level scores are reported at four levels: the state, the school district, the school, and the student.

The following sections discuss two types of test-level scores that are reported to indicate a student's achievement on the MAP Science: (1) the 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 Science MAP. The overall scale score for a content area quantifies the achievement being measured by the Science test. In other words, the scale score represents the student's level of achievement, where higher scale scores indicate higher levels of achievement on the test and lower scale scores indicate lower levels of achievement.

2.2.2 Levels of Achievement

A student's performance on the 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 July 2008 and upheld after the Cut Point Validation workshop in June 2015. The cut scores reflect the expectations of Missouri educators and citizens of what Missouri students should know and be able to do in Science. (See Chapter 8 of this report for a discussion of the MAP standard setting).

Therefore, the 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 in the MAP Grade-Level Assessment *Guide to Interpreting Results* (see also Chapters 4 and 7).

2.2.3 Use of Test-Level Scores

The Science MAP scale scores and achievement levels provide summary evidence of student achievement in 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 curriculum planning. The results presented in this Technical Report provide evidence that the scale scores are a valid and reliable indicator of student performance in Science.

2.3 GLE Strand-Level Subscores

The GLE Strand-level subscores indicate student performance in terms of percent-correct score for each GLE in Science.

2.3.1 Use of the GLE Strand-Level Subscores

The purpose of reporting GLE Strand-level subscores on Science MAP tests 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 GLE Strands. 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 GLE Strand-level subscores. Chapter 9 of this Technical Report provides evidence of construct validity that further supports the use of these subscores.

District and school administrators may compare their aggregate results with the state means to better understand their strengths and weaknesses within a content area. Caution should be exercised when comparing GLE Strand-level subscores between students or across years because different items will comprise these subscores and these items may vary in difficulty between test forms or test administrations.

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. In particular, this section will show how the Science MAP follows rigorous procedures to construct tests that reflect the full range of content that the Science MAP is expected to cover.

This chapter is particularly relevant to AERA, APA, & NCME (2014) Standards 4.0, 4.1, and 4.7. It also addresses Standards 3.1, 3.2, 3.9., 4.12, and 7.4, which will be discussed in pertinent sections of this chapter. Standards 4.0, 4.1, and 4.7 are from Chapter 4 of the AERA, APA, & NCME (2014) *Standards*, "Test Design and Development." Each of these Standards and the way each Standard is addressed will be presented in this chapter. AERA, APA, & NCME (2014) Standard 4.0 states the following:

> Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population. (85)

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

3.1 Test Specifications

AERA, APA, & NCME (2014) Standard 4.1 states the following:

> Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s). (85)

The purpose of the test is discussed in Chapter 2. The Science MAP domains are generally defined as the knowledge and skills that are identified within the Missouri Learning Standards and Grade Level Expectations for Science. This framework, in turn, is based on prior consensus among DESE, Missouri educators, and experienced subject-matter experts that the framework represents 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 Science 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 Science MAP test specifications consist of a test blueprint and a test design for Grades 5 and 8. The 2015 The Missouri DESE provided the test specifications for the 2015 MAP Grade-Level Science Assessment for grades 5 and 8 to ITP. ITP used the test specifications and these items to develop two forms at each grade level for administration in the Spring of 2015. All assessments were fixed forms

The test blueprint specifies the target score points for each GLE Strand, as shown in Table 3.1. The blueprint represents a compromise among many constraints, including the target weights for each GLE recommended by the ITP, 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. The degree to which the 2015 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 Table 3.2. Actual point distributions on the 2015 MAP operational forms matched blueprint targets within 10%, which was the tolerance for variation approved by DESE.

3.2 Item Development

There was no new item development for the 2014–15 Science and previously developed items were included in the 2014–15 test forms. Historically, the MAP item development took 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 ensure proper grammar, punctuation, and compliance with the established style. Clarity and fair access for all examinees also fall within the purview of the style reviews, which occurred at scheduled milestones within the overall test development process. The last field test of MAP Science items occurred during the 2008-09 test administration.

Past item development, reviews, and field-testing processes were in compliance with the following AERA, APA, & NCME (2014) standards.

Standard 3.1 Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

Standard 3.2 Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

Standard 4.7 The procedures used to develop, review, and try out items and to select items from the item pool should be documented. (87)

Detailed information on Science MAP item development, reviews, field-testing can be found in the *Missouri Assessment Program: Grade Level Assessment Technical Report 2009.*

Selected-response (SR) items that were included in the 2015 Science tests were developed by the ITP in years 2000-2005. The standard item development process employed by ITP involves both ITP content area specialists and school teachers. ITP test development staff convenes item writing workshops across the state and trains educators on sound item writing practices. Using the content standards and test blueprint as a guide, educators draw upon their experiences in the classroom to write items that would be engaging to today's students and representative of current curriculum. ITP content experts, copyeditors, and test development experts then review these items for content accuracy, adherence to item development guidelines, fairness, and universal design. The items are also copyedited for clarity of expression as well as grammar and spelling. Once the items are reviewed internally, ITP convenes a panel of educators to evaluate the items. The educators review the items for grade level appropriateness, content relevance, and accuracy. The goal of the educator panels is to confirm that the items are appropriate for the intended students and content area. Items are then edited and refined by ITP test developers as necessary.

3.3 Item Review

The most recent Content and Bias Review (CBR) of MAP Science items was conducted by CTB and American Institute for Research (AIR) in December 2013, during which items were evaluated for content, bias, and suitability for online administration. The participants verified each item's alignment to the Missouri Learning Standards by reviewing the GLE assignment. The accepted items became candidates for the next step in the process, the Science MAP on-line field test. The process of item review adhered to the following AERA, APA, & NCME (2014) standard:

> **Standard 3.2** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

Items developed by ITP were reviewed by a panel of educators after item development and before field testing. The standard process of item review includes review for grade level appropriateness, content relevance, and accuracy. The goal of the educator panels is to confirm that the items are appropriate for the intended students and content area. Items are then edited and refined by ITP test developers as necessary before field test administration.

3.4 Field Test Administration

This section of the report describes field testing and data analysis of items included in the 2014-15 Science assessments.

3.4.1. On-line Field Test Administration and Data Analysis of MAP Items

The items approved by Missouri CBR committees became the basis for the formation of the first on-line field test forms administered Spring 2014. Approximately 150 Science items, previously

administered to Missouri students in paper-and-pencil administrations, were administered online to samples of Grade 5 and 8 students.

The data analysis for the administered on-line Science items was conducted by McGraw-Hill Education CTB in the Summer of 2014. Classical item analysis including computation of *p*-values, computation of item-total test correlations, a distractor analysis for multiple-choice items, and evaluation of item omit rates. Items were calibrated using the Item Response Theory (IRT) methodology. The three-parameter logistic model for multiple choice items and the two-parameter partial credit model for constructed response items were used to estimate item parameters (Bock & Aitkin, 1981; Thissen, 1982). The Stocking and Lord (1983) procedure was used to equate the administered on-line items to the operational test scales using a set of common items administered in the on-line field test and the paper-and-pencil operational assessments. In addition, to evaluate whether the items may function differently for equally able members of different groups, differential item functioning analysis was conducted for all items (Camilli & Shepard, 1994). The results of these data analyses were delivered to DESE for subsequent operational test form selection.

3.4.2. Field Test Administration of ITP Items

Items that were selected from the ITP item pool for inclusion in MAP Science assessments were not field tested in Missouri prior to their operational use due to insufficient time between item selection and form development. These items were field tested nationally between 2001 and 2005. A sample of Missouri students participated in the ITP national research studies conducted by ITP at that time.

3.5 Operational Test Selection

The MAP Grade-Level Science assessments for grades 5 and 8 consist of three sessions. The first session contains constructed response items, the second session contains selected-response items, and the third session contains a performance event. ITP assembled the operational forms using the constructed response items and performance events provided by the DESE. These items were field-tested online in Spring 2014. ITP supplied the selected response items that were field tested on a national sample to ensure that they were grade level appropriate.

Test forms were built from the pool of eligible items according to the test specifications. ITP development staff reviews each test form to make certain that the individual items on the forms were appropriate and error free. Educators and other content experts were convened to conduct another review of the test forms for content accuracy, fairness, and universal design. At this stage in the process, reviews focused not only on individual items, but also on the content representativeness and statistical characteristics of the test as a whole. Note that selected response items were worth one point each, constructed response items were two points each, and the value of items related to the performance event ranged from one to four points each.

The selection criteria were based on both content requirements and statistical criteria, including the following:

1. Test length and item types match the DESE-approved test design.
2. Content coverage matches DESE-approved test blueprint.
3. The following items are avoided, whenever possible:

a. *p*-value ≤ 0.10 or ≥ 0.90
b. Omit rates ≥ 5%
c. Poor Fit statistics (Q1)
d. Significant DIF statistics: If an item with DIF has 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).

Constructed-response items were selected first such that the test characteristic curves for these sets were aligned with the test characteristic curves of the online field test item pool. These items were designated to serve as anchor items on the 2014–15 test forms.

Production of the 2015 operational test forms and ancillary materials commenced in October 2014. Items were ordered and placed into online test forms in preparation for online operational testing, and a process of reviews between DRC/CTB and DESE ensued until final approvals were in place in March 2015, prior to the beginning of the Spring 2015 operational test.

3.6 Universal Design

Grade-level assessments that are universally designed allow participation of the widest possible range of students, resulting in more valid inferences about students' performance. Universally designed grade-level assessments may reduce the need for accommodations by reducing or eliminating access barriers associated with the tests themselves. Table 3.3 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 2015 Science test forms in compliance with AERA, APA, & NCME (2014) Standard 3.1, which states the following:

> Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

Universal design requires that grade-level assessments 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 the Science MAP tests, the assessments include 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 print test books and through the web formatting of the online test forms. The page specifications 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 and online test forms. Written instructions 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.

3.7 Accommodations and Designated Supports

AERA, APA, & NCME (2014) Standard 3.9 states the following:

> Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs. (67)

In compliance with Standard 3.9, the MAP assessments were developed to be inclusive of special needs students. Students with disabilities or students who are English Language Learners were provided test administration accommodation and/or designated supports based on their Individualized Education Plan (IEP). More information on accommodations and designated supports can be found in Section 4.4.2 of Chapter 4. Accommodation and designated supports code definitions can be found in the *Test Administration Manual* presented in Appendix A.

Braille and large print versions were constructed for each grade/content area to enable visually challenged students to participate in the Science MAP testing. Specific recommendations on how to transcribe items into Braille were provided by an independent Braille expert, who collaborated with the Braille publisher to produce the Braille version of the MAP and teacher's notes that accompany the Braille forms. DESE conducted a review meeting with a committee of teachers in February 2015 to ensure that both the Braille and large print versions of the 2015 Science MAP assessment would be accessible to Missouri's visually challenged students. DESE and the teacher committee made recommendations, as needed, for how to further revise 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, because some items represent concepts that are simply not appropriate for students who take the Braille form. In Science it was necessary to omit items from the Braille version due to excessive difficulty associated with the Braille transcription. Table 3.4 lists the items that were omitted from the 2015 Braille forms.

3.8 Standards and Content Specifications

AERA, APA, & NCME (2014) Standard 4.12 states the following:

> Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications. (89)

The Science MAP assessed version 2.0 of the Missouri GLEs. Prior to selecting the operational tests, In compliance with Standard 4.12, CTB and DESE performed an in-depth comparison of the version 2.0 GLEs against the former version in place since 2006. This comparison was conducted beginning in early 2008 through the approval of the 2010–2013 MAP test specifications. The analysis included an alignment of the entire MAP item pool to the version 2.0 GLEs, which was reviewed and approved by DESE. The alignment study conducted by HumRRO indicated that there were some alignment deficiencies in the 2013 MAP test forms for Science (range-of-knowledge) which was mainly attributed to a large number of GLEs at each grade level. Recommendations were made to broaden the scope of item development so more GLEs could be tested; increase the

cognitive complexity of new test items; and reduce the number of Science GLEs so a greater proportion can be tested each year.

DESE followed these recommendations in development of the 2015 test forms. HumRRO was contracted again to review and analyze the alignment of the 2015 Science assessments for Grades 5 and 8 to the 2009 revision of the Missouri Learning Standards for Science.

To conduct the content alignment review, HumRRO applied the Webb (2005) alignment method. Dr. Norman Webb developed a procedure to evaluate alignment of the assessment to the content standards using four statistics. These statistics indicate how well an assessment covers the content standards in terms of content breadth and depth. The alignment indicators include:

- Categorical concurrence – determines the degree of overall content coverage by the assessment for each content strand.
- Range-of-knowledge correspondence – indicates the specific content expectations (e.g., standard, benchmark) assessed within each strand.
- Balance-of-knowledge representation – provides a statistical index reflecting the distribution of assessed content within each strand (i.e., how evenly the content is assessed.)
- Depth-of-knowledge consistency – compares the cognitive complexity ratings of the items with the complexity ratings of each content standard.

The alignment evaluation involved a comparison of the 2015 MAP Grade-Level Science Assessment to the Missouri Learning Standards. The content alignment evaluation involved a review by current and recently retired Missouri educators highly familiar with the content standards and the assessment.

The results of the independent alignment review provide generally positive support for the content validity of the MAP Grade-Level Science Assessment based on several outcomes. First, panelists found that the included items assessed a level of cognitive complexity that was at or above the cognitive complexity level of the Grade Level Expectation (GLE) associated with each item for both forms. Second, items were distributed evenly over content expectations. Additional analyses found that the Science items do measure the intended content and panelists' levels indicate that almost all of the items in each strand assess students at the appropriate cognitive complexity (using Webb Depth-of-Knowledge levels). These results demonstrate that the forms are properly aligned with the Missouri Learning Standards and the items on the forms reflect a balanced representation of the standards at a level of complexity appropriate for the grade.

However, it should be noted that due to the Science MAP assessments being grade-span assessments, i.e. the GLEs from three grade levels and not just one were included in the assessment, the range-of-knowledge results imply a restricted range of content assessed by the assessments. These findings stem from more GLEs than items being available for the assessment. This is a direct result of the Science assessments being grade-span assessments, and the way in which the Webb indicators do not account for the state's intentions/emphasis of content particularly for a grade span assessment (HumRRO, 2014).

3.9 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the Science MAP grade-level assessments. The efforts by DESE, University of

Iowa, and DRC/CTB in developing the Science MAP are in alignment with multiple best practices of the test industry but, in particular, support the following AERA, APA, & NCME (2014) Standards:

- Standard 3.1—Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population.
- Standard 3.2—Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.
- Standard 3.9—Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.
- Standard 4.0—Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.
- Standard 4.1—Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).
- Standard 4.7—The procedures used to develop, review, and try out items and to select items from the item pool should be documented.
- Standard 4.12—Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.

Table 3.1: 2015 Science MAP Test Blueprint: Target Score Points by GLE Strand

Science	Grade	
GLE Strand	**5**	**8**
Matter and Energy	8	8
Force and Motion	6	5
Characteristics of Living Organisms	6	6
Interactions of Organisms	7	5
Earth Systems	7	8
The Universe	6	6
Scientific Inquiry	15	17
Science, Technology, and Human Activity	5	5

Table 3.2: Science MAP 2015 GLE Strand, Item Type, and Point Distributions

Grade	GLE Strand	SR Items	CR/PE Items	Total Items	SR Points	CR/PE Points	Total Points	% of Total Points
5	Matter and Energy	3	2	5	3	4	7	12%
	Force and Motion	1	2	3	1	4	5	8%
	Characteristics of Living Organisms	3	2	5	3	4	7	12%
	Interactions of Organisms	5	2	7	5	4	9	15%
	Earth Systems	4	2	6	4	4	8	13%
	The Universe	2	2	4	2	4	6	10%
	Scientific Inquiry		9	9		14	14	23%
	Science, Technology, and Human Activity		2	2		4	4	7%
	Total	18	23	41	18	42	60	100%
8	Matter and Energy	4	2	6	4	4	8	13%
	Force and Motion		2	2		4	4	7%
	Characteristics of Living Organisms	2	2	4	2	4	6	10%
	Interactions of Organisms	2	2	4	2	4	6	10%
	Earth Systems	5	2	7	5	4	9	15%
	The Universe	2	2	4	2	4	6	10%
	Scientific Inquiry		9	9		17	17	28%
	Science, Technology, and Human Activity	1	2	3	1	4	5	8%
	Total	16	23	39	16	45	61	100%

Note: Data in this table reflect item and point distribution for one test form in each grade. Item number or point distribution per GLE strand may vary slightly across test forms in a given grade, but all forms are aligned with the test blueprint.

Table 3.3: Elements of the 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, tables, figures, illustrations, and response formats.

Table 3.4: Items Omitted from the MAP Spring 2015 Braille Version

Grade	Content Area	Type	Session	Item
5	Science	CR	3	2
8	Science	CR	3	6

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* (2014), "[t]he usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions" (111). 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 (2014) Standards 4.15, 4.16, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7 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 Science assessments 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 grade-level assessments is a collaborative effort between DESE and DRC/CTB. DESE conveys to districts the purpose of the assessments and the importance of test administration being consistent with test industry standards. The tests and the consistent standards of administration must also 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 district test coordinators who, in turn, convey test administration training to schools within their districts. DESE conducts quality assurance visits during testing to ensure district adherence to the standardized administration of the tests.

The district test coordinators are responsible for the schools within their districts. They disseminate information to each school, offer assistance with test administration, and serve as the liaisons between DESE and their districts. DESE also provides assistance with and interpretation of the assessment data and test results.

The Assistant Director of Assessment trained the district test coordinators in the following components of assessment administration: the *Test Administration 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 the last administration in Spring 2014.

Appendix A of this report contains DESE's presentations on the *Test Administration Manual*. During these presentations, the Assistant Director of Assessment walked the district test coordinators and other department staff through an annotated version of the *Test Administration Manual.* The district test coordinators, in turn, used this information to train staff within their districts.

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 (2014) Standards related to test administration procedures.

For the Spring 2015 test administration, DRC/CTB produced one administration manual: the *Test Administration Manual*. DESE Curriculum and Assessment staff reviewed, provided feedback, and gave final approval for this manual.

The *Test Administration Manual* is common to all grades and content areas. It provides detailed instructions for administering the Missouri Assessment Program Guide-Level Assessments. The manual includes instructions for test preparation, scripts for administering the tests (including links to secure listening scripts for accommodated versions of the assessments), and post-test administration procedures. Information included in the *Test Administration Manual* is listed below.

1.0 Overview of Important Information for the MAP Grade-Level Assessments
- 1.1 This Test Administration Manual
- 1.2 Glossary of Terms
- 1.3 About the Tests
- 1.4 Schedule of Important Dates for Spring 2015
- 1.5 Test Administration Policies
- 1.6 Scheduling the Tests
- 1.7 Accommodations and Special Populations
- 1.8 Tutorials and Practice Tests

2.0 Before Online Testing
- 2.1 Advance Announcements and Preparation
- 2.2 User Roles
- 2.3 Test Security
- 2.4 eDIRECT and INSIGHT
- 2.5 Assessment Materials for Students/Administrators

3.0 During Online Testing
- 3.1 Specific Administration Information
- 3.2 Moving a Student During an Assessment

4.0 After Online Testing
- 4.1 Submitting All Tests/Close of Testing Window
- 4.2 Reporting Test Invalidations
- 4.3 How to Handle Student Absences
- 4.4 Securely Destroy Materials
- 4.5 Individual Student Reports

5.0 Large Print, Braille, and Paper-and-Pencil Editions
- 5.1 Before Testing
- 5.2 During Testing
- 5.3 After Testing

Appendix A: Item Types
Appendix B: Handling Student Transfers and Changes in Testing Status
Appendix C: Contaminated Test Materials

Appendix D: Test Book Accountability Form
Appendix E: INSIGHT Keyboard Shortcuts and Icons

This section presents the AERA, APA, & NCME (2014) Standards relevant to test administration and how information in the MAP *Test Administration Manual* addresses these Standards.

Standard 4.15 The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented. (90)

In compliance with Standard 4.15, the MAP *Test Administration Manual* provides 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 *Test Administration Manual* describe the following: general rules of online testing; pause rules; scheduling the tests; recommended order of test administration; classroom activity information; assessment duration, timing, and sequencing information; and the materials that the examiner and students need for testing.

Standard 4.16 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 materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions. (90)

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.

Tutorials and practice tests are provided in each content area to familiarize students with how to navigate the online system and practice with the item types and the functionality of the testing environment.

Standard 6.1 Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user. (114)

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 administration manual. It should be noted that adhering to the test schedule is also one critical component. The *Test Administration Manual* includes instructions for scheduling the test within the state testing window of March 30–May 22, 2015. The *Test Administration Manual* contains the

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

Standard 6.3 Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user. (115)

Department staff administer reports on testing concerns, which cover a wide range of improper activities that may occur during testing, including the following: copying and reviewing assessment questions with students; cueing students during testing either verbally or with written materials on the classroom walls; cueing students nonverbally, such as by tapping or nodding the head; using a calculator on parts of the test where it is not allowed; allowing students to correct or complete answers after tests have been submitted; splitting sessions into two parts; ignoring the standardized directions in the online assessment; 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 Individualized Education Plan (IEP); allowing accommodations for students who do not have an IEP; allowing students to use dictionaries on parts of the assessment; 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 issues. 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 Quality Assurance —Grade-Level Assessment—Self-Monitoring Report is sent for the superintendent to use for replying to the allegation. A sample district report is shown in Figure 4.2.

Standard 6.4 The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance. (116)

Section 2.3 in the *Test Administration Manual* overviews the following steps that teachers should take to prepare for computer-based testing for administering the MAP online test:

- Determine the layout of the physical computer lab.
- 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 covered, removed, or placed out of the students' view.

Standard 6.6 Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means. (116)

The *Test Administration Manual* presents instructions for post-test activities to ensure that online tests are submitted and printed test materials are handled properly to ensure the integrity of student information and test scores. Detailed instructions guide test examiners in submitting all online test records. For students who were administered a large print or Braille version of the Science MAP, examiners are instructed to transcribe students' responses from the large print test or Braille test

book into the online testing system (INSIGHT) exactly as the responses appear in the large print or Braille test book.

Standard 6.7 Test users have the responsibility of protecting the security of test materials at all times. (117)

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 Administration Manual* instructs test coordinators in procedures for organizing and packing materials and returning them to DRC/CTB for secure inventory purposes. DESE curriculum and assessment staff has opportunities to review, provide feedback, and give final approval. The purpose of the instructions is to ensure that secure test materials are properly accounted for and organized properly for return shipment.

4.2.2 Security Forms

As soon as large print and Braille test books are received by a district, the district test coordinator ensures 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 test books to DRC, school and district test coordinators are required to complete and submit a *Test Book Accountability Form* that details the number of test books or printed test forms returned. This form also requires that districts/schools document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books. A sample *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 the MAP score reports from the 2015 administration. More detail about the guide can be found in Chapter 7. It is also presented in Appendix B.

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 Administration Manual* (see Appendix A, p.28).

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 2015 test was administered to students within the state testing window of March 30–May 22, 2015. 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 *Test Administration Manual* provided examiners with timing guidelines for the assessments. For the MAP's custom sessions, examiners were instructed to allow students to complete the assessment if they were making adequate progress. The timing schedule of the MAP is presented in Table 4.1.

4.4.2 Tools, Designated Supports, and Accommodations

Universal tools, designated supports, and accommodations are allowed on the MAP. These types of student aids are described below.

- Universal tools are available to all students based on student preference and selection. Some tools, such as a ruler and a digital notepad, are embedded in the online system, while others, such as a physical thesaurus and scratch paper, are external to the system. The availability of particular universal tools varies by item.
- Designated supports are accessibility features of the assessments available for use by any student for whom the need has been indicated by a team of educators knowledgeable about the student.
- Accommodations are changes in procedures or materials that increase equitable access during the Science MAP assessments. Assessment accommodations allow students to access assessment content to show what they know and can do. Accommodations are available for students with documented Individualized Education Programs (IEPs) or 504 Plans.

Accommodations may be used with students: who qualify under the Individuals with Disabilities Education Act (IDEA) and have an IEP; who qualify under Section 504 of the Americans with Disabilities Act and have a Section 504 plan; or who are identified as English Language Learner (ELL) students. 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 (2014) Standard 6.2 states the following:

> When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing. (115)

In compliance with this, the grade-specific MAP *Test Administration Manual* contains the list of universal tools, designated supports, and accommodations permissible for the MAP assessments. The tables of universal tools, designated supports, and accommodations are presented in the *Test Administration Manual* and are shown in Tables 4.2 through 4.4, respectively. Note that if a specific accommodation is not on the list of accommodations in the *Test Administration Manual*, the

accommodation may still be permitted. Detailed information regarding testing accommodations can be found on the DESE website at http://dese.mo.gov/college-career-readiness/assessment.

Braille and large print forms are provided for students who are visually challenged.

Table 4.5 summarizes the numbers of reportable students for whom designated supports or accommodations were indicated by a teacher for the 2015 Science MAP. The analyses in Table 4.5 are based on census data and include only students with indicated designated supports or accommodations who received a scale score on the Science MAP.

In 2015, the separate setting (designated support) and having the test read aloud (designated support) were the most frequently used for Science 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 designated supports or accommodations, and providing guidance on appropriate interpretations of the test results. These communication and training efforts by DESE and the ancillary information developed by DRC are in alignment with multiple best practices of the testing industry but, in particular, support the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- Standard 4.15—The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented.
- Standard 4.16—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 materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.
- Standard 6.1—Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user.
- Standard 6.2—When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing.
- Standard 6.3—Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user.
- Standard 6.4—The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance.

- Standard 6.6—Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means.
- Standard 6.7—Test users have the responsibility of protecting the security of test materials at all times.

Table 4.1: Science MAP Administration Schedule Timing Guidelines by Session (Time in Minutes)*

Grade	Session	Science
5	1	45–55
	2	20–25
	3	45–65
8	1	45–55
	2	20–25
	3	45–65

*All times are estimates and all sessions are untimed.
** Listed time excludes untimed 30-minute classroom activity administered prior to performance tasks.

Table 4.2: MAP Universal Tools

Universal Tools

- The following is a list of universal tools for the Grade-Level Assessments.
- These tools are available to all students.

Tool	Format	Description
Break (Pause)	Online	The system allows all students to pause the assessment for up to 20 minutes. There is no limit on the amount of times a student may use this tool.
	Any	All students may take breaks of up 20 minutes as needed.
Calculator (For calculator-allowed items only)	Online	The system allows all students, on items where calculator use is allowed, to have access to an embedded digital calculator.
	Any	All students may have access, on items where calculator use is allowed, to a physical calculator.
English Dictionary	Online	The system allows all students access to an embedded English dictionary for use on the writing performance task.
	Any	All students may have access to a physical English dictionary for use on the writing performance task.
Highlighter	Online	The system allows all students to have access to a highlighter for marking desired text, questions, and answers.
	Any	All students may have access to a physical highlighter.
Keyboard Navigation	Online	The system allows all students to navigate through the text by using the keyboard.
Mark for Review	Online	The system allows all students to mark an item for review.
Notepad (Scratch paper)	Online	The system allows all students to use a digital notepad (called "Sticky Notes") to make notes about an item.
	Paper	All students may have access to physical scratch paper to make notes about an item. Physical scratch paper should be collected and destroyed immediately upon the conclusion of the testing session, except during the ELA and Mathematics performance tasks.
Protractor	Online	The system allows all students to use an embedded protractor on specific items where appropriate.
	Paper	All students may have access to a physical protractor for use on specific items where appropriate.
Ruler	Online	The system allows all students to use an embedded ruler on specific items where appropriate.
	Paper	All students may have access to a physical ruler for use on specific items where appropriate.
Spell Check	Online	The system allows students to use an embedded spell check feature on specific items where appropriate. NOTE: This feature must be manually turned on to be activated in the system.
Strikethrough (Called "Cross Off")	Online	The system allows all students to cross out answer options.
Thesaurus	Any	All students may have access to a physical thesaurus during the writing performance task.
Writing Tools	Online	The system allows all students to use selected writing tools on specific items where appropriate. The tools include the ability to bold text, italicize text, create bullets points. There is also an undo/redo feature.
Zoom (Called "Magnifier")	Online	The system allows all students to zoom in or zoom out on text or graphics to make them appear larger or smaller than the default size.
	Paper	All students may have access to devices that allow them to change the size of text, formulas, tables, graphics, etc.

Table 4.3: MAP Designated Supports

Designated Supports for Students Participating in Science Assessments

- The following is a list of designated supports for the Grade-Level Assessments.
- These supports are available to students when deemed appropriate by a team of educators.
- These supports are available to ELL students.

Support	Format	Description	Code
Bilingual Dictionary	Any	ELL students may have access to a physical bilingual dictionary for use on the writing performance task.	S431
Color Contrast	Online	The system allows students to adjust background or font color based on student needs or preferences.	S101
	Paper	Students may have the test presented to them printed in different colors based on student needs or preferences.	S102
Color Overlay	Paper	Students may have a color transparency placed over the test presented to them based on student needs or preferences.	S103
Magnification	Online—Not Embedded	The system allows students to use assistive technology devices to change the size of text, formulas, tables, graphics, etc., beyond the capabilities of the zoom tool.	S105
Masking	Online	The system allows students to block off content that is not of immediate need or that may be distracting by using an embedded masking tool.	S106
	Paper	Students may use a masking tool to block off content that is not of immediate need or that may be distracting.	S107
Read-Aloud (For all items in any subject, excluding ELA reading passages)	Online	The system allows items in mathematics and English language arts to be read aloud to the student via embedded text-to-speech technology. The student can control the speed and volume of the voice.	S041
	Online— Not Embedded	Students may use assistive technology text-to-speech software to allow all items in any subject, not including ELA reading passages, to be read aloud.	S042
	Any	Students may have items in mathematics, science, and English language arts read aloud to them by a trained reader. Reading aloud of ELA reading passages requires an IEP or 504 Plan.	S043
	Any	ELL students may have items in mathematics, science, and English language arts read aloud to them in their native language by a trained translator. Reading aloud of ELA reading passages requires an IEP or 504 Plan.	S111
Scribe (For all items in any subject, excluding ELA writing)	Any	Students may dictate their responses to a trained scribe, who must follow the administration guidelines. Scribing of ELA writing requires an IEP or 504 Plan.	S351
Separate Setting	Any	Students may be allowed to test in a separate setting from other students. This includes testing individually or testing as part of a smaller group.	S501
Translation	Online	The system allows ELL students to use stacked Spanish translations on selected construct-irrelevant math items.	S108
	Any	ELL students may have test directions for math, science, and social studies translated. ELL students may respond to any assessment in their native language. The responses must be translated and then transcribed by a trained scribe, who must follow the administration guidelines. ELL students taking the paper-based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment. This glossary can be translated locally.	S109

Table 4.4: MAP Accommodations for Students with Disabilities

Accommodations for Students with Disabilities Participating in Science Assessments

- The following is a list of accommodations for the Grade-Level Assessments.
- The accommodations must appear in an IEP or a 504 Plan to be allowed.
- These accommodations are available to ELL students.

Accommodation	Format	Description	Invalidates	Code
Abacus	Any	Students may have access to an abacus.		A391
Alternate Response Options	Any	Students may respond to items using an alternate option, including, but not limited to: Adapted Keyboards, StickyKeys, MouseKeys, FilterKeys, Adapted Mouse, Touch Screen, Head Wand, Switches.		A441
American Sign Language (ASL) (For math and science items)	Any	Students may have math, science, social studies items and ELA listening items translated into ASL.		A052
Braille	Paper	Students with visual impairments may access the assessment via a Braille version. Tactile overlays and graphics tools may be used to assist the student in accessing the content.		A012
Calculator GRADES 4–8 ONLY (For non-calculator-allowed items only)	Any	All students in grades 4–8 may have access, on items where calculator use is not allowed, to a physical calculator.		A393
Large Print	Paper	Students with visual impairments may access the assessment via a Large Print version.		A021
Multiplication Table GRADES 4–8	Any	Students in grades 4–8 may have access to a single-digit multiplication table.		A395
Paper-Based Assessment	Paper	Students may have access to a paper-based version of the assessment.		A102
Specialized Calculator (For calculator-allowed items only)	Any	Students may have access, on items where calculator use is allowed, to a specialized calculator, including talking calculators or Braille calculators, when appropriate.		A396

Table 4.5: Number and Percent of Students Receiving Accommodations or Designated Supports, Science MAP 2015 Regular Edition, Braille, and Large Print

Accommodation or Support	Science Grade 5		Science Grade 8	
	Freq.	Pct.	Freq.	Pct.
Braille (A)	3	0.00%	8	0.01%
Large Print (A)	40	0.06%	27	0.04%
Signing of assessment (ASL) (online or paper) (A)	13	0.02%	8	0.01%
Paper based assessment (A)	216	0.33%	175	0.26%
Use of specialized calculator (A)	21	0.03%	16	0.02%
Alternate response options (A)	13	0.02%	4	0.01%
Use of bilingual dictionary (S)	94	0.14%	216	0.32%
Support color contrasting text (S)	470	0.71%	342	0.51%
Support color chooser (S)	593	0.89%	365	0.55%
Support color contrast (S)	5	0.01%	4	0.01%
Color overlay (S)	19	0.03%	2	0.00%
Magnification (S)	247	0.37%	327	0.49%
Masking (online) (S)	314	0.47%	505	0.76%
Masking (paper) (S)	15	0.02%	3	0.00%
Read-aloud embedded technology (S)	10939	16.47%	9000	13.53%
Read-aloud assistive technology (S)	264	0.40%	290	0.44%
Read-Aloud Translator (S)	146	0.22%	126	0.19%
Scribe (S)	1528	2.30%	496	0.75%
Separate setting (S)	8845	13.32%	6625	9.96%
Support translated test directions (S)	108	0.16%	113	0.17%
Read-Aloud - trained reader (S)	2875	4.33%	1836	2.76%
Abacus (A)	30	0.05%	17	0.03%
Calculator (Grades 4-8) (A)	1563	2.35%	3056	4.59%
Multiplication table (Grades 4-8) (A)	1030	1.55%	402	0.60%
Non-accommodation special case - paper based assessment	7	0.01%	45	0.07%

(A) Indicates an accommodation.
(S) Indicates a support.

Figure 4.1: Sample Script from *Test Administration Manual*

3.1 Specific Administration Information

1. The TE distributes the Student Test Tickets.

You should have received Student Test Tickets for this testing session from your DTC or STC. Before beginning, ensure that you have all of the correct test tickets for the students who will be testing. Note the Test Name and read it aloud where the script states [Test Name].

If students are starting a new session:

SAY **You are about to take (the) [Test Name].**

If students are resuming a session:

SAY **You are about to continue (the) [Test Name].**

I will now hand out a Test Ticket to each of you. When you receive your Test Ticket, check that your name appears on the ticket. If your name does not appear, raise your hand.

Distribute test tickets to each student, ensuring that each student is given the correct ticket with his or her name printed on it. Contact your STC or DTC if a needed ticket is missing.

2. The TE directs students to the test sign-in page.

SAY **Now select the "MO Online Assessments" icon that appears on your screen.**

Students using a laptop or desktop workstation should double click on the icon. Students using a Chromebook or iPad should tap on the icon. Help students if they have trouble activating the icon.

3. The TE instructs students to log in.

SAY **At the top of your screen you should see "Missouri Department of Elementary & Secondary Education." Below that, you will see links for the Online Tools Training and Test Sign in for the MAP Grade-Level Assessment Summative test. Please select "Test Sign In."**

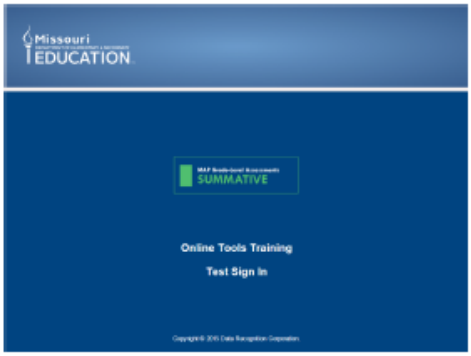


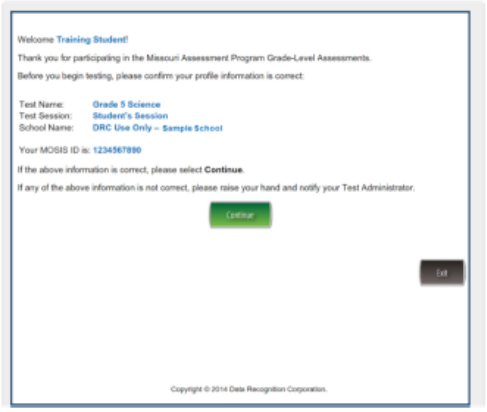
Figure 4.1: Sample Script from *Test Administration Manual* (cont.)

SAY **This is the Login screen. Type your username and password from your Test Ticket into the correct boxes on the screen. Then select "Sign In."**

Test Ticket information is unique to each student and each session. Assist students as needed; TEs may have to help students type in this information. After the login, make sure all students are on the correct screen. Wait for all students to reach this page.

SAY **This is the Welcome screen. Please check that your name appears at the top of the screen. Check that the test name is [Test Name]. Then check that your school, MOSIS ID, and other information are correct. If everything is correct, select "Continue." If your information is not correct, please raise your hand.**

If a student's information is incorrect, the TE should contact the STC and/or the DTC.



SAY **You are now on the screen that shows the name of the test you are scheduled to take. If you do not see this, please raise your hand. Please select the test link that is shown.**

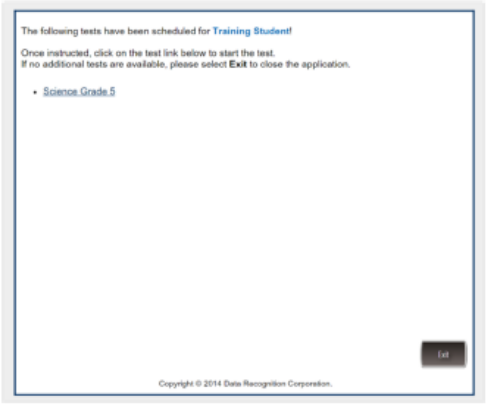


Figure 4.1: Sample Script from *Test Administration Manual* (cont.)

SAY Select the NEXT arrow to continue.

SAY The following screens contain the test directions for the test you are taking today. Please read the directions carefully. If you have any questions about the directions, raise your hand. You can find the directions during your test by clicking the HELP button in the top right corner.

SAY During the test, you may see a page with no test questions. Follow the directions on the page to continue taking the test.

If you are unsure of an answer, provide what you think is the best answer; there is no penalty for guessing. If you would like to review that answer at a later time, mark the item for review by clicking the FLAG at the bottom of the screen before going on to the next question. Flagging the item will remind you to go back and decide whether or not you want to change the answer.

You may PAUSE at any point in the test by clicking PAUSE after answering an item. The PAUSE button is used to stop the test. Please raise your hand if you need a break and ask me before you click PAUSE. After pausing, a timer will appear on your screen. After your break, click on the RESUME button to continue. If you pause for more than 20 minutes, you will need to log back in.

Your answers need to be your own work. Please keep your eyes on your own test and remember that there should be no talking.

Read aloud the following paragraph if students are taking Part 1 (Session 1) of an ELA performance task.

SAY Use your scratch paper to take notes you want to keep for Part 2, the essay portion, of this performance task. Any notes you take online using Sticky Notes will not be saved for Part 2.

SAY When you are ready to begin your test, click BEGIN THE TEST.

Figure 4.2: Sample District Report Form

MISSOURI DEPARTMENT OF ELEMENTARY AND SECONDARY EDUCATION
OFFICE OF COLLEGE AND CAREER READINESS – ASSESSMENT SECTION

QUALITY ASSURANCE – GRADE-LEVEL AND END-OF-COURSE ASSESSMENTS
SPRING 2015

DIRECTIONS

A Department of Elementary and Secondary Education (Department) employee or designee (On-Site visit) **OR** District Test Coordinator (District Self-Monitoring) completes this form at an onsite or self-monitoring Quality Assurance (QA) visit during the Spring 2015 assessment window. This QA visit **MUST** occur during the district testing window.

The End-of-Course visit **MUST** be for Algebra I, Biology or English II.

Please complete all questions on this form. In addition to completing the questions on this form, the QA visit will include a classroom observation.

After the visit or self-monitoring has occurred, the Department employee or designee (On-Site Visit) **or** District Test Coordinator (District Self-Monitoring) will submit the QA form to the Department by accessing the form electronically at https://www.surveymonkey.com/s/GLA2015. The questions on that site mirror those on this form. Forms must be entered electronically by June 12, 2015 at the latest.

Questions: Contact the Assessment Section at 573-751-3545 or email assessment@dese.mo.gov.

Important: District Test Coordinators (DTCs) should continue to report testing irregularities or concerns immediately to the Department. Please contact the Assessment Section at 573-751-3545.

ABOUT THE VISIT

As part of the No Child Left Behind (NCLB) Act required monitoring process, the Department uses this document as a tool to monitor and strengthen statewide administration of the Missouri Assessment Program's Assessments. The questions are designed to focus attention and help districts examine important areas of assessment training, administration, and test security.

The following are components of the self-monitoring and quality assurance processes:

- documentation of assessment trainings;
- interviews with District Test Coordinators, Special Education Director, Test Examiners, Individualized Education Program (IEP) team members, and school administrators;
- review of documents; and
- classroom visit.

DISTRICT INFORMATION

NAME OF PERSON FILLING OUT THIS FORM:	
DATE OF VISIT:	
SCHOOL DISTRICT NAME:	
COUNTY-DISTRICT CODE:	
BUILDING NAME:	
BUILDING CODE:	
DISTRICT TEST COORDINATOR'S NAME:	
TEST EXAMINER'S NAME:	
GRADE LEVEL OR EOC CONTENT:	

The Department of Elementary and Secondary Education does not discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs and activities. Inquiries related to Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator – Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number 573-526-4757 or TTY 800-735-2966; email civilrights@dese.mo.gov.

MO 500-3068 Page 1

Figure 4.3: Sample Test Book Accountability Form

MAP 2015 GRADE-LEVEL ASSESSMENTS – TEST BOOK ACCOUNTABILITY FORM

District Name: _____ **District #:** _____

School Name: _____ **School #:** _____

This form provides start-to-finish accountability for the Grade-Level test materials 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. 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 and printed PDF tests.
3. Return the completed form to your District Coordinator.

District Coordinator:

4. Complete a copy of this form for District Overage, providing any additional information on the back as required.
5. 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 or PRINTED	GR 3	GR 4	GR 5	GR 6	GR 7	GR 8
(1)	Number of LP/Braille test books listed on packing list	+	+	+	+	+	+
(2)	LP/Braille test books missing from shipment (Fill out A on back)	–	–	–	–	–	–
(3)	Extra LP/Braille books received in shipment (Fill out B on back)	+	+	+	+	+	+
(4)	Additional LP/Braille books from district office (Fill out C on back)	+	+	+	+	+	+
(5)	Printed PDF tests (Fill out D on back)	+	+	+	+	+	+
(6)	**Total test books received and printed (Add lines 1, 3, 4 and 5; then subtract line 2)**						
	TEST BOOKS and PRINTED PDF TESTS RETURNED	**GR 3**	**GR 4**	**GR 5**	**GR 6**	**GR 7**	**GR 8**
(7)	Number of LP/Braille tests administered	+	+	+	+	+	+
(8)	Number of unused LP/Braille test books	+	+	+	+	+	+
(9)	Number of printed PDF tests	+	+	+	+	+	+
(10)	**Total test books returned and printed (Sum of lines 7–9)**						
	TEST BOOKS and PRINTED PDF TESTS NOT RETURNED	**GR 3**	**GR 4**	**GR 5**	**GR 6**	**GR 7**	**GR 8**
(11)	LP/ Braille test books securely destroyed (Fill out E on back)	+	+	+	+	+	+
(12)	LP/Braille test books unaccounted for (Fill out E on back)	+	+	+	+	+	+
(13)	Printed PDF tests unaccounted for (Fill out E on back)	+	+	+	+	+	+
(14)	**Total test books and printed PDF tests not returned (Sum of lines 11–13)**						

I confirm that Line 6 = Line 10 + Line 14.

School Test Coordinator: _____ Date Faxed: _____

Print Name: _____

A. Security barcode numbers of LP/Braille test books missing from shipment

B. Security barcode numbers of extra LP/Braille test books (not on packing list) received in shipment

C. Security barcode numbers of additional LP/Braille test books from district office

D. Security barcode numbers of printed PDF tests

E. Security barcode numbers of lost or securely destroyed LP/Braille test books or printed PDF tests	Explanation

CHAPTER 5: SCORING OF CONSTRUCTED-RESPONSE AND TECHNOLOGY-ENHANCED ITEMS

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

Chapter 5 adheres to AERA, APA, & NCME Standards 4.18, 4.20, 6.8, and 6.9. Each of these Standards will be presented in the pertinent section of this chapter. Standard 4.18 provides some general guidance for Chapter 5:

> Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays. (91)

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

5.1 Constructed-Response Scoring Process

Constructed-response items were scored by human raters who were trained by DRC/CTB.

5.1.1 Selection of Scoring Evaluators

AERA, APA, & NCME (2014) Standard 4.20 specifies the following:

> The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring. (92)

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.

DRC/CTB strives to develop a highly qualified, experienced core of evaluators so that the integrity of all projects is appropriately maintained.

Recruitment
The MAP 2015 project was staffed with a large number of returning evaluators and team leaders who had previous experience with the MAP and other handscoring projects. In addition, DRC/CTB worked with Stafforward (a company specializing in staffing practice areas such as clerical and administrative, call centers, accounting, healthcare, scientific and light-industry) to recruit new team leaders and evaluators for employment. Recruitment sources included advertisements online and in newspapers in Indianapolis, Indiana, and nearby areas.

DRC/CTB requires that all evaluators and team leaders possess a bachelor's degree or higher. Stafforward carefully screened all new applicants and required them to produce either a transcript or a copy of the degree. Stafforward 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. Stafforward 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.

Stafforward 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 Stafforward 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 DRC/CTB business ethics and security procedures.

5.1.2 Handscoring Training Process

AERA, APA, & NCME (2014) Standard 6.9 specifies the following:

> Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected. (118)

Training Material Development
All materials necessary for scoring Science were developed by DRC/CTB. These materials included the scoring guides and training papers used to complete the handscoring of constructed-response items.

Missouri Science operational items have been previously field tested and the training materials used during the field test were used for training readers to score the operational items. During the previous field testing, 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 field test materials were previously presented during a 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 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 a specific Rater Item Block consisting of one item to be scored. Evaluators were trained using the following steps:

- Reviewing constructed-response items
- Reviewing rubrics
- Reviewing anchor papers
- Explaining scoring strategies, followed by a question-and-answer period
- Scoring a training set, followed by sharing established scores
- Qualifying Round 1
- Qualifying Round 2 (if necessary)
- Explaining condition codes and sensitive paper procedures
- Explaining unscannable image procedures

All evaluators were trained and qualified using the same procedures and criteria. 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% exact agreement qualification
- 3-point item: 85% exact agreement qualification
- 2-point item: 95% exact agreement qualification
- 1-point item: 100% exact agreement qualification

Qualification tests consisted of 10 papers. Evaluators were given 2 attempts to qualify on an item. If an evaluator did not achieve the targeted exact percentage on the first qualification attempt (or had a non-adjacent score), they were re-trained and were allowed to attempt a second qualification round. Readers failing both qualification attempts were not allowed to score that particular item, but may have been allowed to train and qualify for scoring a different item.

5.1.3 Monitoring the Scoring Process

AERA, APA, & NCME (2014) Standard 6.8 states the following:

> Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented. (118)

Section 5.1.3 explains the monitoring procedures that DRC/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 in this report in order to preserve the integrity of the items for use in future Science 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.

In addition to the checkset process, DRC/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 and provide feedback and counseling as appropriate.

Approximately 10% of all responses 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.

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 Technology-Enhanced Item Scoring Process

All technology-enhanced items were processed through DRC's autoscoring engine and scored according to the assigned scoring rules. DRC ensured that all rubrics and scoring rules were verified for accuracy before scoring any technology-enhanced items. DRC established an adjudication process for technology-enhanced items and any gridded responses to verify that correct answers were identified. DRC's technology-enhanced scoring quality process included the following:

- A scoring rubric was created for each technology-enhanced item. It was as simple as describing the one and only correct answer for dichotomously scored items (scored as either right or wrong). If partial credit was possible, the rubric described in detail the type of response that could receive credit for each score point.
- The information from the scoring rubric was entered into the scoring system within the item banking system so that the truth resided in one place, along with the item image and other metadata. This scoring information designated specific information that varied by item type. For example, for a drag-and-drop item, the information included which objects are to be placed in which drop region to receive credit.
- The information was then verified by another autoscoring expert.
- After testing started, reports were generated that showed every response, how many students gave that response, and the score the scoring system provided.
- The scoring was then checked against the scoring rubric using two levels of verification.
- If any discrepancies were found, the scoring information was modified and verified again. Scoring was then re-run. This checking and modification process continued until no other issues were found.
- As a final check, a final report was run that showed all student responses, along with their frequencies and received scores.

In case of Braille, large print or paper-and-pencil non-accommodated form administration, student responses to paper-and-pencil technology enhanced-equivalent items were transcribed (entered) into the online system by a test examiner.

5.3 Multiple-Choice and Multi-Select Item Scoring Process

Responses to multiple-choice and multi-select items were captured during the online test administration. In case of Braille, large print or paper-and-pencil form administration, student responses to these items were transcribed into the online system by a test examiner.

5.4 Inter-Rater Reliability

Approximately 10% of the papers in Science were scored independently by a second reader. The statistics for the inter-rater reliability were calculated for all items at all grades. To determine the reliability of scoring, the percentage of perfect agreement and adjacent agreement between the two readers was examined.

For each item, a quadratic weighted kappa statistic was calculated to reflect the level of improvement beyond the chance level in the consistency of scoring. These quadratic weighted kappa values are presented in Table 5.1. To aid in the interpretation of kappa statistic, 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

A total of 78 items were scored by human readers across all test forms and both Science grade levels. As shown in Table 5.1, raters demonstrated at or above 89% perfect and adjacent agreement for all Science items. The quadratic weighted kappa values indicate that there was good or very good inter-rater agreement for all Science items except for eight items. Two Grade 8 items on the transcribed form showed fair agreement and one Grade 8 item on the same form showed moderate agreement. In addition, two Grade 8 items and three Grade 5 items on regular forms demonstrated moderate inter-rater agreements as measured by quadratic weighted kappa.

5.5 Summary

The information presented in this chapter summarizes the scoring procedures for different types of items and steps taken by DRC/CTB to ensure accuracy in the technology-enhanced item scoring and handscoring process. The inter-rater reliability statistics presented in Section 5.4 demonstrate that the items are scored reliably. These efforts by DRC/CTB follow multiple best practices of the testing industry and support AERA, APA, & NCME (2014) Standards 4.18 4.20, 6.8, and 6.9:

- Standard 4.18—Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays.

- Standard 4.20—The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring.
- Standard 6.8—Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented.
- Standard 6.9—Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected.

Table 5.1: Inter-rater Reliability, Science

Grade	Session	Form	Item #	Score Range	% Perfect	% Adjacent	% Perfect & Adjacent*	Quadratic Weighted Kappa
5	1	CA2	1	0–2	73.2	24.3	97.5	0.75
	1	CA2	3	0–2	78.2	18.6	96.8	0.73
	1	CA2	4	0–2	77.6	20.8	98.4	0.78
	1	CA2	7	0–2	84.1	15.5	99.6	0.80
	1	CA2	8	0–2	63.0	30.7	93.7	0.60
	1	CA2	10	0–2	81.2	17.8	99.0	0.82
	1	CA2	11	0–2	86.1	12.5	98.6	0.88
	1	CA2	12	0–2	82.9	16.9	99.8	0.80
	1	CA2	13	0–2	89.6	9.9	99.5	0.87
	1	CA2	14	0–2	91.6	8.0	99.6	0.90
	1	CA2, CA3	2	0–2	85.2	14.2	99.4	0.85
	1	CA2, CA3	5	0–2	93.0	6.2	99.2	0.92
	1	CA2, CA3	6	0–2	81.2	17.8	99.0	0.81
	1	CA2, CA3	9	0–2	79.2	19.3	98.5	0.76
	1	CA3	1	0–2	82.5	17.0	99.5	0.84
	1	CA3	3	0–2	90.5	8.9	99.4	0.86
	1	CA3	4	0–2	97.2	2.3	99.5	0.97
	1	CA3	7	0–2	98.5	1.5	100.0	0.97
	1	CA3	8	0–2	82.6	17.1	99.7	0.82
	1	CA3	10	0–2	83.8	15.4	99.2	0.79
	1	CA3	11	0–2	91.6	8.1	99.7	0.91
	1	CA3	12	0–2	92.2	7.4	99.6	0.92
	1	CA3	13	0–2	85.0	14.4	99.4	0.83
	1	CA3	14	0–2	88.5	11.4	99.9	0.88
	3	CA2, CA3	33	0–2	98.8	0.5	99.3	0.97
	3	CA2, CA3	34	0–4	82.9	11.9	94.8	0.91
	3	CA2, CA3	35	0–1	90.5	9.1	99.6	0.82
	3	CA2, CA3	36	0–1	96.7	3.1	99.8	0.94
	3	CA2, CA3	37	0–2	72.5	24.9	97.4	0.72
	3	CA2, CA3	38	0–1	84.8	14.7	99.5	0.65
	3	CA2, CA3	39	0–1	78.4	20.8	99.2	0.50
	3	CA2, CA3	40	0–1	75.1	24.3	99.4	0.51
	3	CA2, CA3	41	0–1	99.2	0.8	100.0	0.99
8	1	CA2, CA3	5	0–2	88.6	11.2	99.8	0.90
	1	CA2, CA3, CT2	10	0–2	95.2	4.4	99.6	0.96
	1	CA2, CT2	1	0–2	77.2	18.4	95.6	0.76
	1	CA2, CT2	2	0–2	74.3	22.9	97.2	0.75
	1	CA2, CT2	3	0–2	88.9	10.3	99.2	0.89

Table 5.1: Inter-rater Reliability, Science (cont.)

Grade	Session	Form	Item #	Score Range	% Perfect	% Adjacent	% Perfect & Adjacent*	Quadratic Weighted Kappa
	1	CA2, CT2	4	0–2	94.5	4.7	99.2	0.94
	1	CA2, CT2	6	0–2	83.4	14.1	97.5	0.83
	1	CA2, CT2	7	0–2	79.8	16.4	96.2	0.82
	1	CA2, CT2	8	0–2	80.4	18.4	98.8	0.81
	1	CA2, CT2	9	0–2	85.8	13.3	99.1	0.89
	1	CA2, CT2	11	0–2	91.8	7.7	99.5	0.93
	1	CA2, CT2	12	0–2	85.0	14.4	99.4	0.81
	1	CA2, CT2	13	0–2	86.9	11.8	98.7	0.87
	1	CA2, CT2	14	0–2	87.4	12.4	99.8	0.85
	1	CA3	1	0–2	80.3	18.6	98.9	0.80
	1	CA3	2	0–2	77.6	20.4	98.0	0.76
	1	CA3	3	0–2	77.9	20.8	98.7	0.75
	1	CA3	4	0–2	87.0	12.5	99.5	0.89
	1	CA3	5	0–2	94.4	4.7	99.1	0.95
	1	CA3	6	0–2	77.6	19.4	97.0	0.80
8	1	CA3	7	0–2	79.8	19.3	99.1	0.80
	1	CA3	8	0–2	92.8	7.2	100.0	0.60
	1	CA3	9	0–2	77.7	20.9	98.6	0.81
	1	CA3	11	0–2	79.2	19.2	98.4	0.80
	1	CA3	12	0–2	76.2	22.3	98.5	0.78
	1	CA3	13	0–2	76.6	21.9	98.5	0.74
	1	CA3	14	0–2	80.2	18.3	98.5	0.78
	3	CA2, CA3	31, 30	0–2	92.7	6.6	99.3	0.94
	3	CA2, CA3	32, 31	0–2	95.0	4.1	99.1	0.97
	3	CA2, CA3	33, 32	0–1	73.4	26.3	99.7	0.47
	3	CA2, CA3	34, 33	0–1	99.7	0.3	100.0	0.99
	3	CA2, CA3	35, 34	0–4	77.8	19.6	97.4	0.92
	3	CA2, CA3	36, 35	0–1	88.8	10.8	99.6	0.78
	3	CA2, CA3	37, 36	0–2	80.5	16.0	96.5	0.82
	3	CA2, CA3	38, 37	0–3	63.8	25.3	89.1	0.71
	3	CA2, CA3	39, 38	0–1	86.2	13.3	99.5	0.72

Table 5.1: Inter-rater Reliability, Science (cont.)

Grade	Session	Form	Item #	Score Range	% Perfect	% Adjacent	% Perfect & Adjacent*	Quadratic Weighted Kappa
	3	CT2	31	0–2	88.9	3.7	92.6	0.98
	3	CT2	32	0–2	64.0	36.0	100.0	0.63
	3	CT2	33	0–1	83.3	16.7	100.0	0.28
	3	CT2	34	0–1	88.5	11.5	100.0	0.77
8	3	CT2	35	0–1	96.2	0.0	96.2	1.00
	3	CT2	36	0–4	80.0	20.0	100.0	0.83
	3	CT2	37	0–1	88.0	12.0	100.0	0.36
	3	CT2	38	0–2	81.5	14.8	96.3	0.83
	3	CT2	39	0–2	65.4	26.9	92.3	0.55

* 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 items worth 2 or more points, percent discrepant also refers to the cases where the assigned score varied by more than 1 point.

CHAPTER 6: OPERATIONAL DATA ANALYSES

This chapter of the Technical Report describes the analyses that occurred on the Science operational data. These analyses include a classical item analysis and examination of the raw scores and an item response theory (IRT) analysis involving calibrating, scaling, and linking. All of these analyses were conducted using the calibration sample and some were replicated using census data for reporting purposes.

In this section, we present the classical item statistics, including aggregate raw score statistics and individual item-level statistics. Next, we discuss the IRT models used for calibrating the data and address the purpose of data calibration and scaling for each content area. The calibration samples are presented next, followed by the data calibration results, including the model-data fit for the Missouri data. If the IRT models fit the empirical item response distributions for the population for which generalizations (i.e., Missouri students) are made, 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 the Science tests are presented.

Chapter 6 demonstrates adherence in the MAP program to AERA, APA, & NCME (2014) Standards 1.8, 4.14, 5.2, 5.13, and 5.15. Each standard will be explicated within the appropriate section of this chapter.

6.1 Classical Item Statistics

In this section, we present summary test statistics for each form and grade of Science MAP. This is followed by item-level statistics for each grade level. These statistics were produced using census data.

6.1.1. Test-Level Statistics

Table 6.1 presents the number of items and score points on each test form, 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 Science. Note that the Grade 5 transcribed form (Braille, large print, or paper-and-pencil form) was the same as the regular CA2 form and as such the test statistics were computed using the combined data of both forms.

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

6.1.2. Item-Level Statistics

Tables 6.2 and 6.3 present the item statistics for each item included in all test forms by grade for Science. The tables include test form, test session, item number, *p*-value, item-total correlation (R_{it}), omit rates and adjusted N count for each item by grade and content area. As stated in the previous section, the transcribed form in Grade 5 included the same items as the regular form CA2. Therefore item statistics for items included in these two

forms were computed using the combined data. The transcribed form in Grade 8 contained nine unique items that were not repeated in either CA2 or CA3 regular form. All other Grade 8 items in the transcribed form were the same as items in the regular CA2 form. The statistics for common items across the transcribed form and the regular forms were computed on the combined across forms data. The statistics for the nine items unique to the transcribed form were computed using the data for that form. These items were administered to approximately 250 students and the item statistics should be interpreted with caution.

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 shown in Table 6.1 the average *p*-values associated with the Science regular forms range from 0.53 (Grade 8, form CA2) to 0.69 (Grade 5, form CA3). The average *p*-value for the Science Grade 8 transcribed (CT2) form is 0.36.

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 the performance of students in the center of the distribution. Having a range of *p*-values also helps to prevent floor and ceiling effects so that the test does not have large numbers of students at the minimum or maximum possible scores. The items included in regular forms had *p*-values ranging from 0.30 to 0.92 (see Tables 6.2 and 6.3). The nine unique items included in the Grade 8 transcribed form had *p*-values ranging from 0.15 to 0.69 (see Table 6.3). Overall, this broad range of *p*-values, which indicates the items measure well throughout the range of skills at a given grade, 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 excluded from the total score. It indicates how well an item differentiates between low- and high-achieving students. In general, items with correlations below 0.20 are said to be poorly discriminating. Except for two Grade 5 items and three Grade 8 items, all the items in the Science MAP had item-test correlations above this threshold. Items with an item-total correlation below the 0.20 threshold were further analyzed to ensure that the items were correctly keyed.

Omit Rates: The omit rate for each item indicates 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 Standard 4.14 of the AERA, APA, & NCME (2014) *Standards*. This standard is concerned with speededness of a test:

> For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure. (90)

The results in this section show that, overall, student test scores are not adversely affected by the rate at which students complete the test. The results presented in Tables 6.2 and 6.3 show that the omit rates for all items included in regular forms were less than 5% suggesting that the majority of students were able to complete the test in the prescribed amount of time. All of the nine items that were unique to the transcribed Grade 8 form were omitted by more than 5% of test takers. However, as stated earlier, this form was administered only to a small number of students requiring specific testing accommodations (Braille, large print, or paper-and-pencil form) and the item statistics reflect only performance of this small group of students and not of the entire population of Missouri students. In general, students have ample time to complete all sections of the test.

6.2. 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) for Science items contained in regular test forms. 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

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

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

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, 2002). 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.

Using the 3PL/2PPC model for estimation of Science items parameters was consistent with the past methodology implemented for this content area. Item parameters estimated after the 2014–15 Science test administration and equated to the existing Missouri Science scales were used to score Missouri students who took these tests.

Items unique to the transcribed Grade 8 form were not calibrated due to the small number of students who took these items. Instead, field test parameters for these items were used in student scoring.

6.3. Calibration Sample

In this section we describe the calibration sample in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) *Standards*:

> The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics. (25)

Science test data were analyzed using calibration samples. In order to accommodate the reporting schedule and necessity to conduct the cut point validation for Science, samples of Grades 5 and 8 data were acquired shortly after the test scoring started. The samples were drawn from the pool of students who tested and were scored early in the test administration window. The Science calibration samples were selected to be representative of the Missouri student population in a given grade in regard to gender and race/ethnicity distribution. Table 6.4 shows the representativeness of the Science calibration samples compared to the census data. This table demonstrates that the Science calibration sample was representative of the state.

6.4. Calibration and Scaling

6.4.1. Data Calibration

Science data were calibrated and scaled after the 2014–15 test administration. The 3PL/2PPC IRT models were used to estimate item parameters for Science Grades 5 and 8. The test forms in each grade level shared common items and were calibrated concurrently at that grade level. In a process of item calibration, the number of estimation cycles was set to 80 with the convergence criterion of 0.001 for all content areas. The maximum value of *a*-parameter was set to 3.0, and the range for *b*-parameter was set between –7.5 and 7.5. For all items, the estimated *a*- and *b*-parameters were within the prescribed parameter ranges. It should be noted that there were a number of items with the default value for the *c*-parameter on the Science tests. When the PARDUX (Burket, 2002) program used to calibrate the items encounters difficulty estimating the *c*-parameter, it assigns a default *c*-parameter value of 0.20.

6.4.2. 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. The fit index for item *i* is

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

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

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

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

No items were flagged for poor fit in Grade 5 and four items were flagged for poor fit in Grade 8. Table 6.5 shows the chi-square statistic and the *Z*-statistic for each flagged item.

The average percentage correct 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. The item characteristic curves for these items are presented in Figures 6.1 through 6.4. 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.[2] 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.

Each of the flagged items was examined more closely by studying its item characteristic curve (ICC) at each nonzero 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 of these regions of the underlying ability distribution where there are relatively few students. 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. All items flagged for poor fit in Grade 8 Science test were retained and contributed to student scores.

It is important to notice that while items may be flagged for misfit, these flags may not 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.

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 5.2 of the AERA, APA, & NCME (2014) *Standards*, which states the following:

> The procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly. (102)

The Science 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 Science assessment data were performed using PARDUX (Burket, 2002), which is proprietary software developed by CTB/McGraw-Hill. PARDUX is

[2] For constructed-response items, there will be one graph for overall items fit and one graph for each score level. For example, a 2-point item will have four graphs: overall fit and fit for 0, 1, and 2 score points.

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).

6.4.3. Linking Methods

DRC/CTB used a common-item, non-equivalent groups design to link the current year's assessment to the established MAP Science scales. The constructed-response items administered to Missouri students in the past served as the anchor set, and the non-equivalent groups are comprised of approximately 5,000 student records in Grade 5 data and approximately 4,400 student records in Grade 8 data. After the initial IRT item calibration, item parameters were linked to the MAP Science scales using the Stocking & Lord (1983) equating procedure.

Standard 5.13 of the AERA, APA, & NCME (2014) *Standards* states the following:

> 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 were established and on the accuracy of the equating functions. (105)

The Stocking & Lord (1983) procedure minimizes the mean squared difference between the two test characteristics curves (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 TCC 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 the Stocking and Lord procedure (Kolen & Brennan, 1995;

Michaelides & Haertel, 2004). The estimation of the SEE is beyond the scope of this report.

6.4.4. Anchor Items

AERA, APA, & NCME (2014) Standard 5.15 requires information about the anchors, stating the following:

> 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 in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented. (105)

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

The Stocking and Lord (1983) procedure, also called the test characteristic curve (TCC) method for which the mathematical equation was provided in Section 6.4.3 (Linking Methods), 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.

Differential item functioning was evaluated by examining previous (input) and transformed (estimated) item parameters. Items with an absolute difference of parameters greater than two times the root mean square deviation were flagged. These differences were also monitored by plotting input and estimated item parameters.

Item Response Theory (IRT) Item-Ability Regression Curves

Differences between the item-ability regression curves of the anchor items in 2014–15 Science test administration were compared to previous calibrations. The differences between the curves are evaluated using the following statistics:

- UnWtd Mean = Average signed difference in estimated probability.
- UnWtd Mean Abs Dif = Average Absolute (unsigned) difference in estimated probability.
- UnWtd RMSD = Root mean squared difference.
- 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.

Both unweighted and weighted versions of these statistics were 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, that is, the frequency distribution of estimated student abilities during the calibration.

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) was identified. For MaxAbsDIFPC, large differences are those greater than +.15, and moderate differences are all differences between +.125 and +.15.

Removal of Anchor Items

One of the key requirements of anchor items in deriving valid reliable linking results is that the anchor items form a miniature of the test, in terms of content coverage or test blueprint. While dropping an anchor item flagged based solely on statistical criteria has its simplicity, this option may change the content coverage and invalidate results. Before an anchor item may be dropped from an anchor set, the item characteristics, adequacy of the content coverage, and impact to the size of the anchor set must be evaluated.

As stated above, an item is removed from the anchor set only if it adversely affects the quality of scaling, not the desirability of the results. As such, DRC/CTB does not consider how the removal of an item affects the overall mean scale score or the impact data (percentage 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 equating set under the following conditions:

1. An item may be a candidate for removal when it is flagged for large differences on four of the seven statistics (listed above) considered when examining the differences between the IRT regression curves.
2. 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.
3. 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, though within 5% is preferred.
4. 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 set may not be less than 20% of the total test length.

Results of Anchor Evaluation

Two items on the Grade 5 Science test were flagged using four or more of the statistics used to examine ICC differences using the IRT Anchor Regression Curve. One of those items, (item #2 in Session 3 on CA2, CA3, and CT2 forms) was recommended for removal from the anchor set due to content concerns (the stimulus and graphing were presented in a different way in the online format than in the paper-and-pencil format). The second flagged item (item #7 in Session 3 on CA2, CA3, and CT2 forms) was not recommended for removal because the student experience between the online and paper and pencil presentation was not determined to be practically different. None of the remaining anchor items had content concerns nor were any anchors flagged on more than four of the statistical criteria. The item-ability regression statistics for Grade 5 anchor items are presented in Table 6.6. The item characteristic curves for the flagged items are displayed in Figures 6.5 and 6.6. In these figures the dashed red line is the ICC curve before equating (based on input parameters) and the solid blue line is the ICC curve after equating (based on new parameter estimates).

In Grade 8, three items were flagged using four or more of the statistics examined with the IRT Regression Curves (item #9 in Session 1, Form CA2 and CT2; item #3 in Session 3 on both CA2 and CA3 forms; and item #6 in Session 1 on Form CA3). All three items were determined to have no content issues or to be affected by the administration mode (online versus paper-and-pencil); therefore, those items remained part of the anchor set. The item-ability regression statistics for Grade 8 anchor items are presented in Table 6.7. The item characteristic curves for the flagged items are displayed in Figures 6.7 through 6.9.

Table 6.8 provides results for the TCC method. This table summarizes the following information for each grade content area: grade level, number of iterations, scaling constants (M1 and M2), and quadratic loss function (F).

Please note that the actual TCCs are used to assess the quality of the linking results. The TCCs for Science Grade 5 are presented in Figure 6.10 (with all anchor items) and in Figure 6.11 (with one anchor item removed). The TCCs for Science Grade 8 are presented in Figures 6.12. The red dashed TCC lines in the plots are the TCCs for the input anchor items. The blue lines are the TCCs from the 2015 MAP parameter estimates transformed to the 2014 MAP scale. The closer the two TCCs are to each other at all ability levels, the more confidence we have in the equating result. In all three cases, the input and estimate TCCs overlay each other, making the two curves indistinguishable.

6.4.5. Vertical Properties of Science Scale

The Science scale is unique to Missouri and was developed after the first Science operational test administration in 2008. This scale has been developed by utilizing the vertical scale properties of the standardized achievement test *TerraNova* (CTB/McGraw-Hill, 2003). Although the Science tests no longer include *TerraNova* items, the Missouri Science item pool is on the *TerraNova* scale customized for Missouri.

Evidence of the validity of the Science MAP growth scales is provided by the increase of the scale score at selected percentiles as grade level increases. Figure 6.13 displays the scale scores for several points on the score distributions for both Science grade levels. These scale scores indicate the growth, or change, in score by grade at the 10th, 25th, 50th, 75th, and 90th, percentiles. Ideally, the scale score associated with each percentile will increase from one grade to another. Figure 6.13 shows that there is an upward progression of scale scores across the two Science grades.

Figure 6.14 shows the TCCs by grade for the Science MAP. The TCCs were generated using equated item parameters from the 2015 test administration. The TCCs are based on combined parameters across all test forms in a given grade. Figure 6.14 shows that the Grade 8 test is more difficult than the Grade 5 test.

6.4.6. 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. In addition, although maximum likelihood estimates are available for students with extreme scores other than zero or perfect, occasionally these estimates have standard errors of measurement that are very large, and differences between these extreme values have little meaning. Therefore, scores are established for these students based on a rational but necessarily non-maximum likelihood procedure. These values, which are set separately by grade, are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). Table 6.9 shows the LOSS and HOSS values used for each grade of Science MAP tests.

6.5. Item-Pattern Scoring

The 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. For additional information on the technical details of the item-pattern scoring, readers can also refer to Yen & Candell (1991).

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 years so that test results may be appropriately compared across years. The data analyses undertaken by DRC/CTB is in alignment with multiple best practices

of the testing industry but, in particular, support the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- Standard 1.8—The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics.
- Standard 4.14—For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure.
- Standard 5.2—The procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly.
- Standard 5.13—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 were established and on the accuracy of the equating functions.
- Standard 5.15—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 in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented.

Table 6.1: MAP Means, Standard Deviations for Raw Scores, *p*-values, Item-Total Correlation (R_{it}): Science 2015

Grade	Form	Total Items	Total Points	Mean Raw Score (SD)	Mean *p*-value (SD)	Mean R_{it} (SD)
5	CA2	41	60	35.38 (10.51)	0.61 (0.18)	0.38 (0.08)
	CA3	41	60	40.34 (9.29)	0.69 (0.16)	0.35 (0.07)
8	CA2	39	61	30.87 (12.59)	0.53 (0.12)	0.42 (0.11)
	CA3	38	60	32.14 (11.12)	0.56 (0.12)	0.38 (0.12)
	CT2	39	60	19.39 (10.90)	0.36 (0.14)	0.40 (0.13)

Table 6.2: Item Statistics Science Grade 5

Science						
Form	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Adj. N**
CA2	1	1	0.50	0.49	0.31	39855
CA2,CA3	1,1	2,2	0.60	0.34	0.14	66222
CA2	1	3	0.31	0.41	0.81	39551
CA2	1	4	0.54	0.42	0.31	39814
CA2,CA3	1,1	5,5	0.66	0.34	0.11	66253
CA2,CA3	1,1	6,6	0.46	0.32	0.15	66162
CA2	1	7	0.54	0.52	0.16	39971
CA2	1	8	0.56	0.42	0.54	39612
CA2,CA3	1,1	9,9	0.53	0.38	0.52	65714
CA2	1	10	0.54	0.47	0.37	39813
CA2	1	11	0.54	0.36	0.53	39630
CA2	1	12	0.74	0.45	0.36	39831
CA2	1	13	0.75	0.41	0.40	39753
CA2	1	14	0.64	0.39	0.34	39823
CA2	2	1	0.92	0.35	0.05	40087
CA2	2	2	0.46	0.10	0.08	40077
CA2	2	3	0.66	0.37	0.08	40077
CA2,CA3	2,2	4,4	0.82	0.31	0.06	66338
CA2,CA3	2,2	5,5	0.89	0.33	0.07	66336
CA2,CA3	2,2	6,6	0.46	0.28	0.06	66338
CA2,CA3	2,2	7,7	0.85	0.45	0.07	66335
CA2	2	8	0.45	0.27	0.08	40076
CA2,CA3	2,2	9,9	0.52	0.35	0.07	66332
CA2	2	10	0.78	0.31	0.09	40072
CA2	2	11	0.91	0.40	0.08	40075
CA2,CA3	2,2	12,8	0.82	0.34	0.06	66339
CA2	2	13	0.86	0.30	0.10	40068
CA2	2	14	0.64	0.49	0.09	40071
CA2	2	15	0.86	0.34	0.11	40066
CA2,CA3	2,2	16,16	0.85	0.36	0.07	66336
CA2	2	17	0.54	0.39	0.13	40058
CA2,CA3	2,2	18,18	0.43	0.23	0.09	66323
CA2,CA3	3,3	1,1	0.91	0.29	0.06	66160
CA2,CA3	3,3	2,2	0.36	0.50	0.16	66247
CA2,CA3	3,3	3,3	0.44	0.40	0.54	65658

Table 6.2: Item Statistics Science Grade 5 (cont.)

Science						
Form	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Adj. N**
CA2,CA3	3,3	4,4	0.60	0.54	0.16	66103
CA2,CA3	3,3	5,5	0.64	0.42	0.47	65570
CA2,CA3	3,3	6,6	0.70	0.33	0.40	65873
CA2,CA3	3,3	7,7	0.30	0.35	0.45	65664
CA2,CA3	3,3	8,8	0.54	0.41	0.61	65532
CA2,CA3	3,3	9,9	0.43	0.43	0.39	66096
CA3	1	1	0.62	0.29	0.05	26244
CA3	1	3	0.62	0.38	0.07	26239
CA3	1	4	0.67	0.35	0.03	26249
CA3	1	7	0.77	0.28	0.06	26253
CA3	1	8	0.64	0.45	0.08	26230
CA3	1	10	0.76	0.44	0.06	26240
CA3	1	11	0.62	0.42	0.21	26140
CA3	1	12	0.65	0.46	0.26	26124
CA3	1	13	0.79	0.36	0.13	26179
CA3	1	14	0.76	0.47	0.07	26239
CA3	2	1	0.73	0.18	0.04	26262
CA3	2	2	0.79	0.36	0.06	26256
CA3	2	3	0.80	0.24	0.05	26258
CA3	2	10	0.91	0.33	0.05	26260
CA3	2	11	0.92	0.30	0.04	26261
CA3	2	12	0.81	0.38	0.05	26260
CA3	2	13	0.83	0.34	0.06	26257
CA3	2	14	0.74	0.36	0.07	26253
CA3	2	15	0.76	0.36	0.05	26259
CA3	2	17	0.81	0.32	0.09	26249

Table 6.3: Item Statistics Science Grade 8

Science						
Form	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Adj. N**
CA2, CT2	1,1	1,1	0.69	0.39	1.07	38427
CA2, CT2	1,1	2,2	0.47	0.41	0.64	38558
CA2, CT2	1,1	3,3	0.43	0.43	0.41	38704
CA2, CT2	1,1	4,4	0.74	0.44	1.13	38375
CA2, CT2	1,1	5,5	0.45	0.56	0.21	38856
CA2, CT2	1,1	6,6	0.48	0.46	0.60	38629
CA2, CT2	1,1	7,7	0.46	0.52	0.95	38434
CA2, CT2	1,1	8,8	0.37	0.49	1.96	37961
CA2, CT2	1,1	9,9	0.46	0.54	1.31	38315
CA2,CA3, CT2	1,1,1	10,10,10	0.45	0.51	2.07	64616
CA2, CT2	1,1	11,11	0.36	0.57	0.67	38587
CA2, CT2	1,1	12,12	0.37	0.37	0.83	38463
CA2, CT2	1,1	13,13	0.74	0.51	1.16	38310
CA2, CT2	1,1	14,14	0.59	0.48	0.55	38654
CA2,CA3, CT2	2,2,2	1,1,1	0.41	0.30	0.08	66380
CA2,CA3, CT2	2,2,2	2,2,2	0.47	0.09	0.09	66372
CA2, CT2	2,2	3,3	0.60	0.57	0.12	38916
CA2,CA3, CT2	2,2,2	4,4,4	0.73	0.40	0.11	66356
CA2, CT2	2,2	5,5	0.73	0.42	0.14	38910
CA2, CT2	2,2	6,6	0.62	0.33	0.13	38913
CA2, CT2	2,2	7,7	0.66	0.40	0.11	38922
CA2, CT2	2,2	8,8	0.60	0.36	0.11	38921
CA2, CT2	2,2	9,9	0.77	0.33	0.11	38920
CA2, CT2	2,2	10,10	0.54	0.40	0.12	38917
CA2, CT2	2,2	11,11	0.50	0.39	0.16	38902
CA2, CT2	2,2	12,12	0.65	0.38	0.12	38916
CA2, CT2	2,2	13,13	0.58	0.26	0.12	38919
CA2, CT2	2,2	14,14	0.54	0.26	0.15	38906
CA2,CA3, CT2	2,2,2	15,14,15	0.59	0.32	0.13	66343
CA2,CA3, CT2	2,2,2	16,15,16	0.52	0.34	0.14	66339
CA2,CA3	3,3	1,1	0.47	0.42	0.70	65398
CA2,CA3	3,3	2,2	0.49	0.57	0.74	65266
CA2,CA3	3,3	3,3	0.52	0.42	0.56	65484
CA2,CA3	3,3	4,4	0.79	0.39	0.38	65816
CA2,CA3	3,3	5,5	0.43	0.58	0.69	65633

Table 6.3: Item Statistics Science Grade 8 (cont.)

Science						
Form	**Session**	**Item**	***p*-value**	**R_{it}**	**Omit Rate**	**Adj. N**
CA2,CA3	3,3	6,6	0.58	0.52	0.68	65318
CA2,CA3	3,3	7,7	0.41	0.43	3.99	62628
CA2,CA3	3,3	8,8	0.40	0.52	1.07	65019
CA2,CA3	3,3	9,9	0.42	0.40	1.32	64748
CA3	1	1	0.60	0.49	0.47	27300
CA3	1	2	0.66	0.49	0.78	27205
CA3	1	3	0.49	0.53	0.44	27302
CA3	1	4	0.48	0.57	0.64	27269
CA3	1	5	0.69	0.40	0.75	27204
CA3	1	6	0.44	0.57	3.27	26431
CA3	1	7	0.43	0.46	0.89	27175
CA3	1	8	0.50	0.29	0.19	27394
CA3	1	9	0.46	0.49	2.33	26745
CA3	1	11	0.42	0.36	0.63	27240
CA3	1	12	0.59	0.47	1.51	26935
CA3	1	13	0.68	0.34	1.03	27109
CA3	1	14	0.59	0.35	0.89	27150
CA3	2	3	0.62	0.33	0.09	27441
CA3	2	5	0.41	0.16	0.07	27446
CA3	2	6	0.79	0.35	0.08	27444
CA3	2	7	0.73	0.27	0.08	27444
CA3	2	8	0.59	0.22	0.09	27442
CA3	2	9	0.61	0.26	0.08	27445
CA3	2	10	0.57	0.21	0.09	27441
CA3	2	11	0.80	0.32	0.09	27442
CA3	2	12	0.53	0.24	0.10	27438
CA3	2	13	0.39	0.13	0.09	27442
CT2	3	1	0.38	0.54	5.14	231
CT2	3	2	0.16	0.50	5.93	238
CT2	3	3	0.18	0.44	4.74	235
CT2	3	4	0.69	0.35	5.93	238
CT2	3	5	0.39	0.49	13.04	219
CT2	3	6	0.28	0.62	11.07	224
CT2	3	7	0.15	0.41	8.30	230
CT2	3	8	0.24	0.42	11.07	221
CT2	3	9	0.25	0.47	20.55	198

Table 6.4: Summary of Calibration and Census Data: Science

	Calibration Sample		Census Data		Difference
	N	%	N	%	(Calib. % − Census %)
Science, Grade 5					
All Students	4929		66412		
Gender					
Male	2524	51.21%	33965	51.14%	0.07%
Female	2405	48.79%	32447	48.86%	−0.07%
Race/Ethnicity					
White	3657	74.19%	48288	72.71%	1.48%
Black	788	15.99%	10688	16.09%	−0.10%
Hispanic	260	5.27%	3874	5.83%	−0.56%
Asian/Pacific Islander	92	1.86%	1490	2.24%	−0.38%
American Indian/Alaska Native	21	0.43%	296	0.45%	−0.02%
Other	111	2.25%	1776	2.67%	−0.42%
Science, Grade 8					
All Students	4397		66526		
Gender					
Male	2328	52.95%	33726	50.70%	2.25%
Female	2069	47.05%	32800	49.30%	−2.25%
Race/Ethnicity					
White	3220	73.23%	49009	73.67%	−0.44%
Black	744	16.92%	10684	16.06%	0.86%
Hispanic	214	4.87%	3618	5.44%	−0.57%
Asian/Pacific Islander	102	2.32%	1418	2.13%	0.19%
American Indian/Alaska Native	20	0.45%	286	0.43%	0.02%
Other	97	2.21%	1511	2.27%	−0.06%

Table 6.5: Item Fit Statistics for Misfitting Items: Science

Content	Grade	Form	Item (Session)	Chi-Square	DF	Total N	Z	Ob-served	Pre-dicted	Obs.-Pred.
SC	8	CA2, CT	3 (1)	87.57	17	2630	12.10	0.4002	0.3991	0.0011
SC	8	CA2, CA3	7 (3)	128.85	17	4397	19.18	0.3734	0.3757	−0.0022
SC	8	CA3	3 (1)	46.27	17	1767	5.02	0.4802	0.4830	−0.0028
SC	8	CA3	5 (1)	97.49	17	1767	13.80	0.6590	0.6587	0.0003

Table 6.6: Statistics Comparing IRT Item-Ability Regression Curves, Science, Grade 5

Anchor Item Position	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	711808	0.0274	0.0207	0.0543	0.0207	0.0425	0.0400	0.0400
2	711847	0.0194	0.0169	0.0285	0.0119	0.0176	0.0155	0.0149
3	711863	0.0791	0.0649	0.1390	−0.0586	0.1123	0.1071	−0.1069
4	711817	0.0274	0.0242	0.0449	−0.0108	0.0216	0.0174	−0.0086
5	711880	0.0366	0.0325	0.0576	−0.0270	0.0499	0.0483	−0.0477
7	711870	0.0406	0.0337	0.0745	0.0043	0.0466	0.0424	−0.0120
8	711812	0.0189	0.0153	0.0340	0.0109	0.0176	0.0137	0.0117
9	711891	0.0363	0.0319	0.0533	−0.0317	0.0468	0.0457	−0.0457
10	711893	0.0348	0.0303	0.0562	0.0303	0.0484	0.0473	0.0473
11	711886	0.0630	0.0517	0.1002	−0.0439	0.0605	0.0523	−0.0516
12	711859	0.0989	0.0695	0.1798	−0.0651	0.0662	0.0451	−0.0437
13	711881	0.0265	0.0234	0.0429	−0.0179	0.0367	0.0356	−0.0351
14	711820	0.0210	0.0174	0.0377	−0.0127	0.0308	0.0293	−0.0287
33	711936	0.0823	0.0533	0.1896	0.0269	0.0437	0.0345	−0.0215
34	**711918**	**0.1351**	**0.1021**	**0.2647**	**0.1021**	**0.1862**	**0.1676**	**0.1676**
35	711920	0.0446	0.0346	0.0848	−0.0346	0.0678	0.0647	−0.0647
36	711921	0.0058	0.0035	0.0138	−0.0035	0.0083	0.0068	−0.0068
37	711924	0.0151	0.0126	0.0257	0.0126	0.0215	0.0209	0.0209
38	711939	0.0626	0.0559	0.1041	−0.0170	0.0816	0.0768	−0.0664
39	**711926**	**0.1032**	**0.0764**	**0.2016**	**−0.0764**	**0.1466**	**0.1326**	**−0.1326**
40	711927	0.0487	0.0369	0.0985	0.0296	0.0719	0.0640	0.0617
41	711896	0.0296	0.0215	0.0626	−0.0174	0.0411	0.0345	−0.0323
42	711890	0.0378	0.0316	0.0609	0.0203	0.0280	0.0220	0.0189
43	711809	0.0982	0.0928	0.1429	0.0928	0.1140	0.1121	0.1121
44	711867	0.0173	0.0147	0.0291	0.0147	0.0241	0.0234	0.0234
45	711841	0.0930	0.0678	0.2004	0.0491	0.0502	0.0345	0.0234
46	711858	0.0408	0.0356	0.0661	0.0356	0.0573	0.0560	0.0560
47	711836	0.0387	0.0310	0.0695	0.0307	0.0554	0.0533	0.0532
48	711810	0.0843	0.0730	0.1378	−0.0730	0.1148	0.1117	−0.1117
49	711776	0.0227	0.0170	0.0444	−0.0170	0.0351	0.0332	−0.0332
50	711822	0.0423	0.0350	0.0711	−0.0350	0.0435	0.0398	−0.0398
51	711873	0.0193	0.0167	0.0357	−0.0082	0.0274	0.0258	−0.0242

Note: flagged items are indicated in **bold** print.

Table 6.7: Statistics Comparing IRT Item-Ability Regression Curves, Science, Grade 8

Anchor Item Position	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
1	711892	0.0371	0.0297	0.0685	−0.0297	0.0512	0.0483	−0.0483
2	711894	0.0663	0.0546	0.1199	−0.0405	0.0832	0.0725	−0.0677
3	711897	0.0319	0.0253	0.0577	0.0253	0.0467	0.0447	0.0447
4	711866	0.0238	0.0182	0.0459	−0.0122	0.0184	0.0134	−0.0068
5	711916	0.0261	0.0191	0.0526	0.0191	0.0388	0.0354	0.0354
6	711869	0.0604	0.0451	0.1229	−0.0393	0.0892	0.0794	−0.0778
7	711901	0.0237	0.0168	0.0509	−0.0154	0.0354	0.0314	−0.0312
8	711857	0.0200	0.0162	0.0389	−0.0108	0.0278	0.0249	−0.0237
9	**711878**	**0.0857**	**0.0625**	**0.1735**	**0.0625**	**0.1217**	**0.1083**	**0.1083**
10	711900	0.0529	0.0426	0.0882	0.0426	0.0658	0.0611	0.0611
11	711902	0.0317	0.0203	0.0742	0.0203	0.0472	0.0396	0.0396
12	711862	0.0386	0.0358	0.0523	−0.0094	0.0313	0.0268	−0.0214
13	711834	0.0098	0.0083	0.0173	0.0013	0.0128	0.0119	0.0087
14	711889	0.0580	0.0466	0.1145	−0.0304	0.0837	0.0759	−0.0713
31	711917	0.0280	0.0231	0.0475	0.0228	0.0388	0.0369	0.0369
32	711923	0.0388	0.0248	0.0935	0.0248	0.0633	0.0557	0.0557
33	**711919**	**0.0814**	**0.0620**	**0.1557**	**0.0620**	**0.1199**	**0.1116**	**0.1116**
34	711933	0.0190	0.0151	0.0377	0.0106	0.0287	0.0268	0.0260
35	711944	0.0706	0.0536	0.1408	0.0536	0.1065	0.0996	0.0996
36	711940	0.0206	0.0133	0.0491	−0.0115	0.0297	0.0240	−0.0228
37	711945	0.0523	0.0382	0.1044	−0.0382	0.0768	0.0697	−0.0697
38	711942	0.0302	0.0204	0.0734	−0.0122	0.0467	0.0393	−0.0352
39	711941	0.0342	0.0257	0.0673	−0.0257	0.0512	0.0476	−0.0476
40	711899	0.0203	0.0143	0.0422	0.0133	0.0266	0.0222	0.0219
41	711914	0.0605	0.0426	0.1320	0.0364	0.0802	0.0651	0.0620
42	711844	0.0193	0.0147	0.0380	−0.0121	0.0247	0.0213	−0.0203
43	711831	0.0468	0.0327	0.1048	−0.0265	0.0673	0.0558	−0.0518
44	711872	0.0108	0.0099	0.0159	0.001	0.0119	0.0109	0.0062
45	**711846**	**0.0733**	**0.0462**	**0.1755**	**−0.0451**	**0.1197**	**0.1055**	**−0.1054**
46	711885	0.0242	0.0221	0.0353	0.0010	0.0240	0.0214	0.0101
47	711895	0.0891	0.0770	0.1607	0.0770	0.0487	0.0388	0.0388
48	711904	0.0489	0.0382	0.0919	0.0382	0.0697	0.0647	0.0647
49	711882	0.0332	0.0287	0.0560	−0.0211	0.0425	0.0391	−0.0385
50	711839	0.0330	0.0267	0.0663	−0.0168	0.0467	0.0413	−0.0370
51	711898	0.0379	0.0350	0.0592	−0.0115	0.0452	0.0418	−0.0341
52	711845	0.0735	0.0635	0.1161	−0.0383	0.0622	0.0509	−0.0439

Note: flagged items are indicated in **bold** print.

Table 6.8: Anchor Evaluation Results: TCC Method: Science

Grade	Equating Run	Number of Anchors	Number of Iterations	F Value	M1	M2	Number of Anchors Removed
5	1	32	13	0.427748	31.2721	669.3987	0
5	2	31	16	0.192101	29.9942	666.9863	1
8	1	36	5	0.226833	31.3381	696.7150	0

Table 6.9: LOSS and HOSS Values by Grade: Science

Grade	Science	
	LOSS	HOSS
5	470	855
8	540	895

Figure 6.1: Item Characteristic Curve for Grade 8 Science, Form CA2 and CT2, Item 3, Session 1

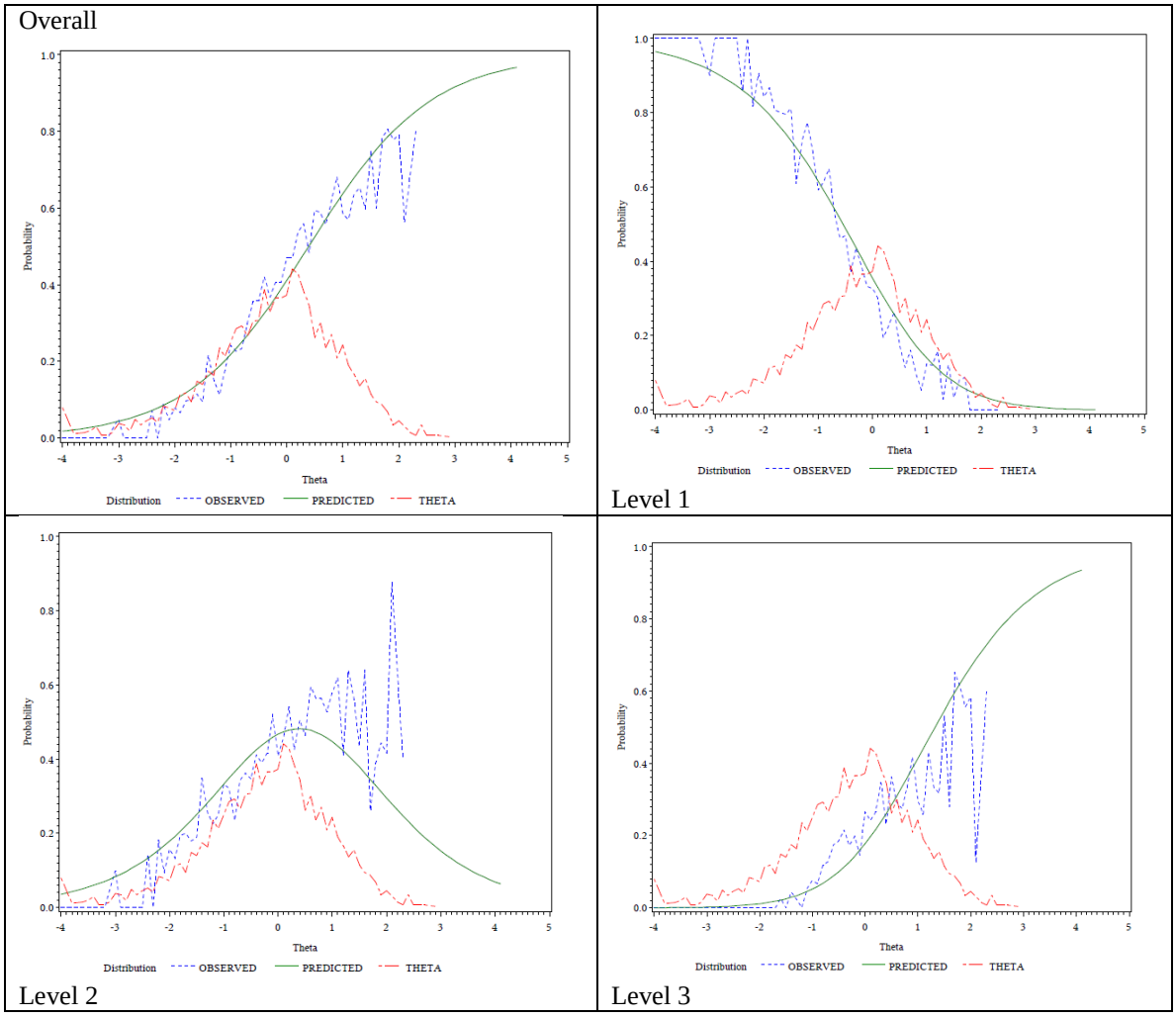


Figure 6.2: Item Characteristic Curve for Grade 8 Science, Form CA2 and CA3, Item 7, Session 3

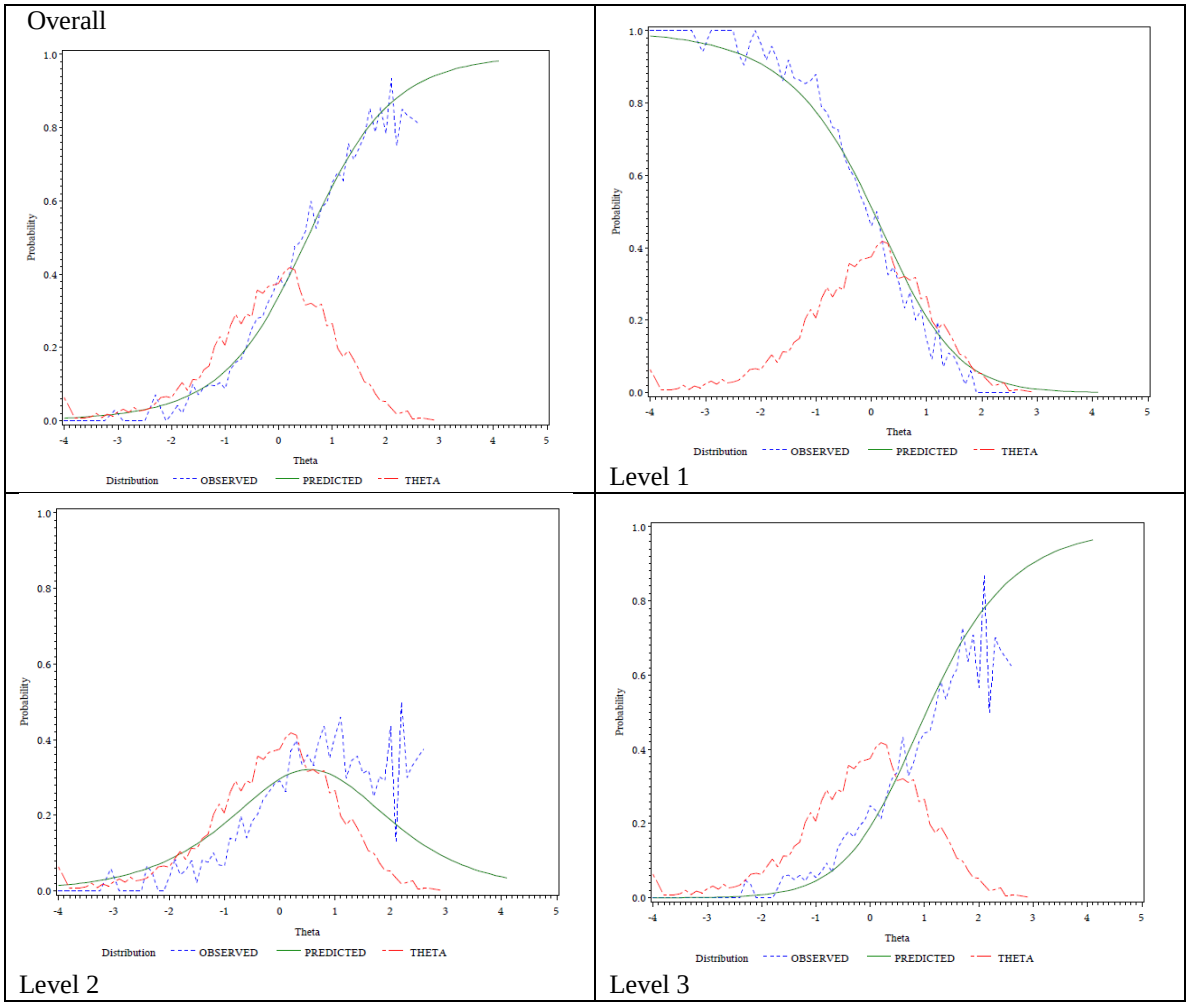


Figure 6.3: Item Characteristic Curve for Grade 8 Science, Form CA3, Item 3, Session 1

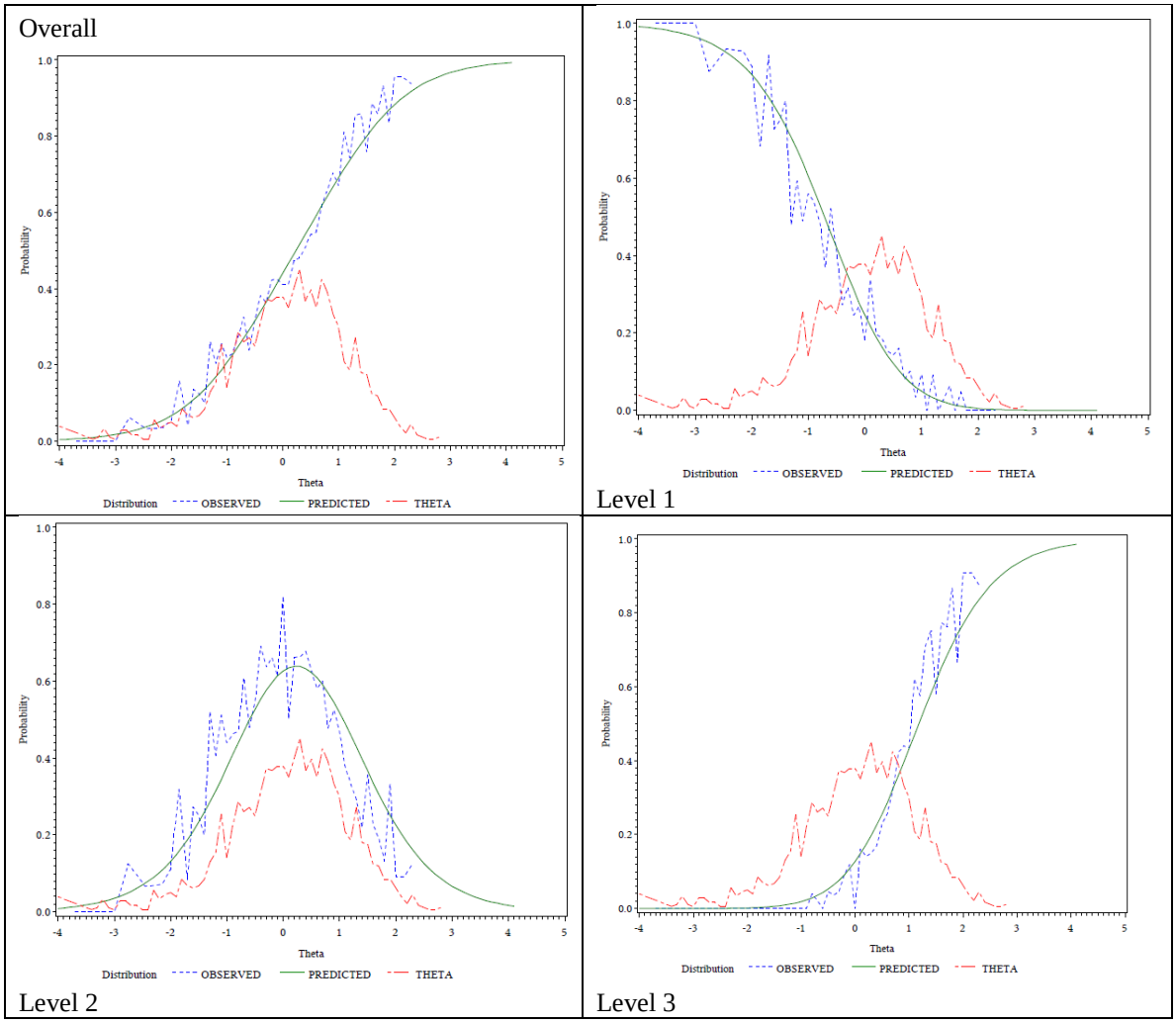


Figure 6.4: Item Characteristic Curve for Grade 8 Science, Form CA3, Item 5, Session 1

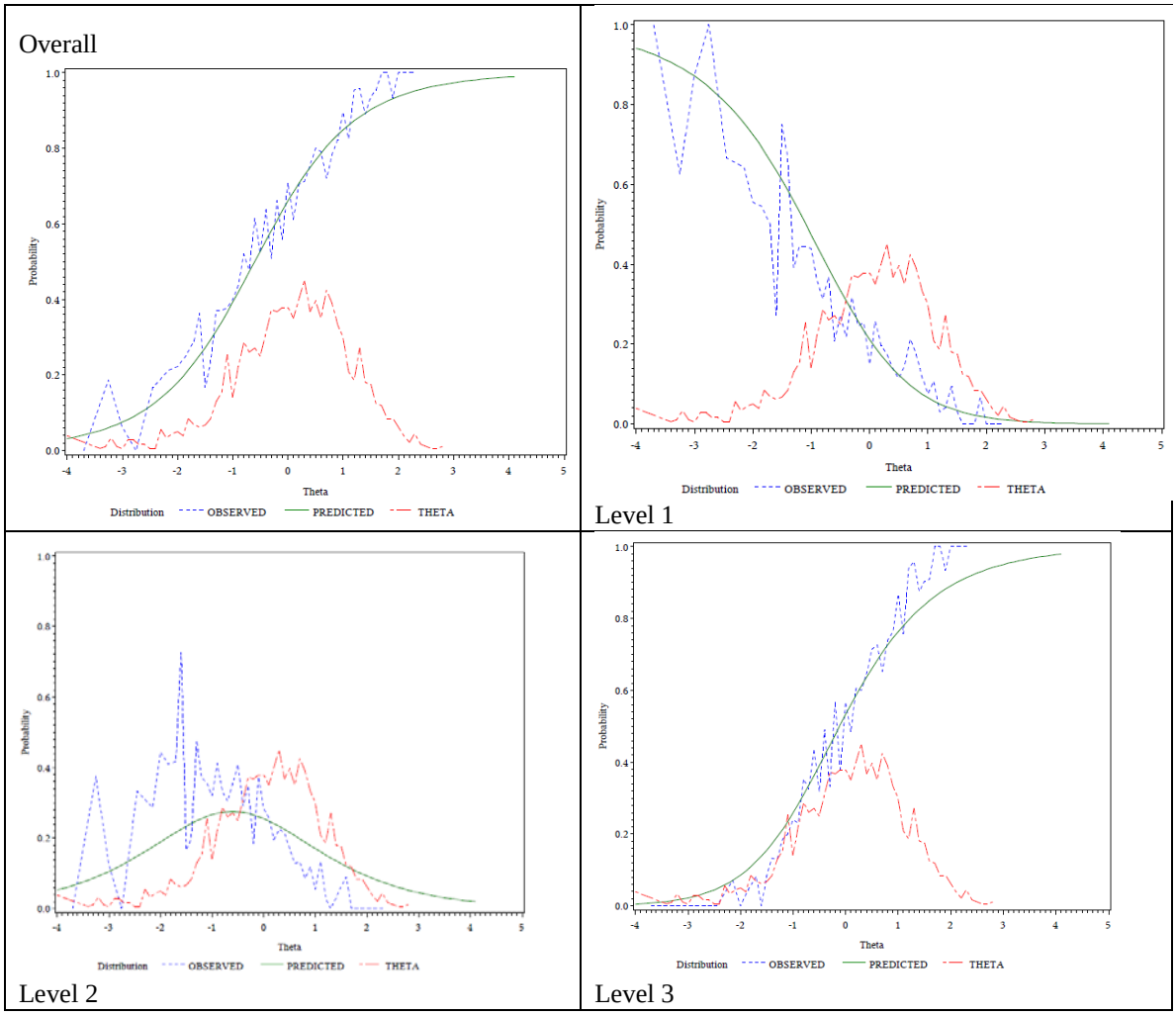


Figure 6.5: Before and after Equating Item Characteristic Curve for Grade 5, Forms CA2 and CA3, item 2, Session 3 (Removed from Equating)

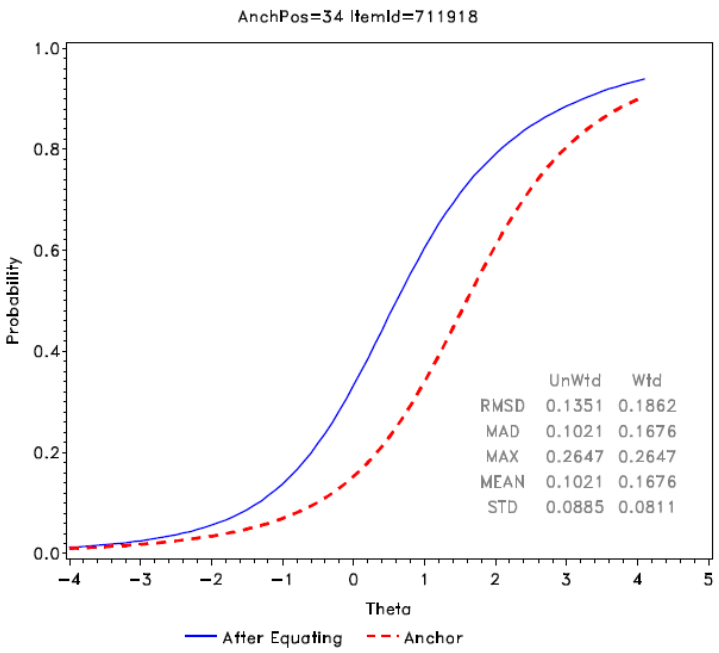


Figure 6.6: Before and after Equating Item Characteristic Curve for Grade 5, Forms CA2 and CA3, item 7, Session 3.

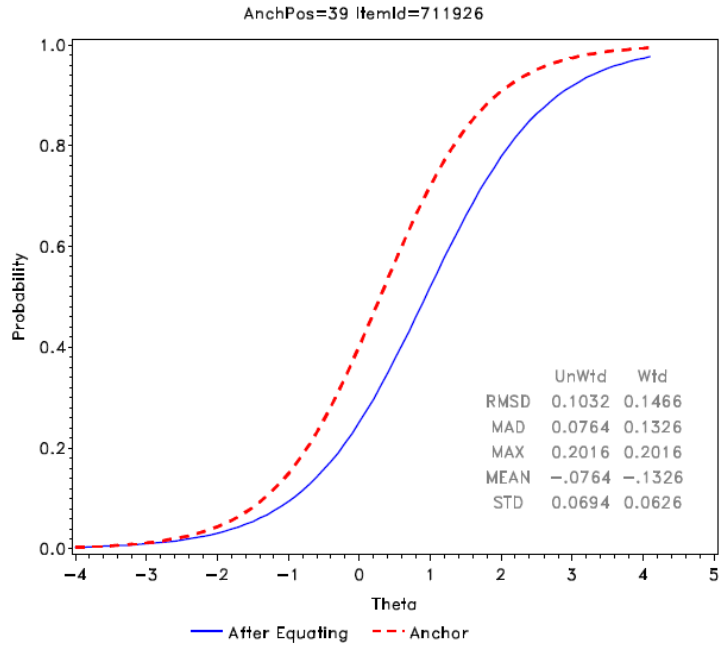


Figure 6.7: Before and after Equating Item Characteristic Curve for Grade 8, Form CA2 and CT2, Item 9, Session 1

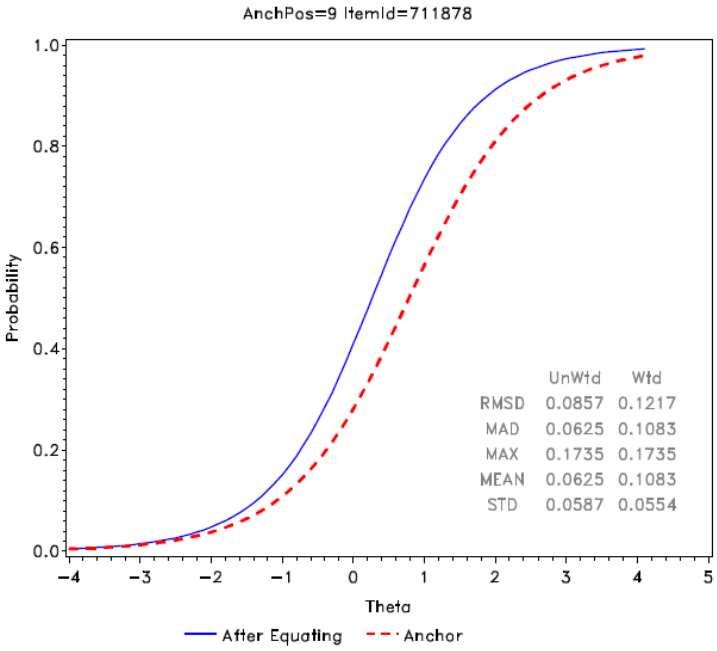


Figure 6.8: Before and after Equating Item Characteristic Curve for Grade 8, Forms CA2 and CA3, Item 3, Session 3

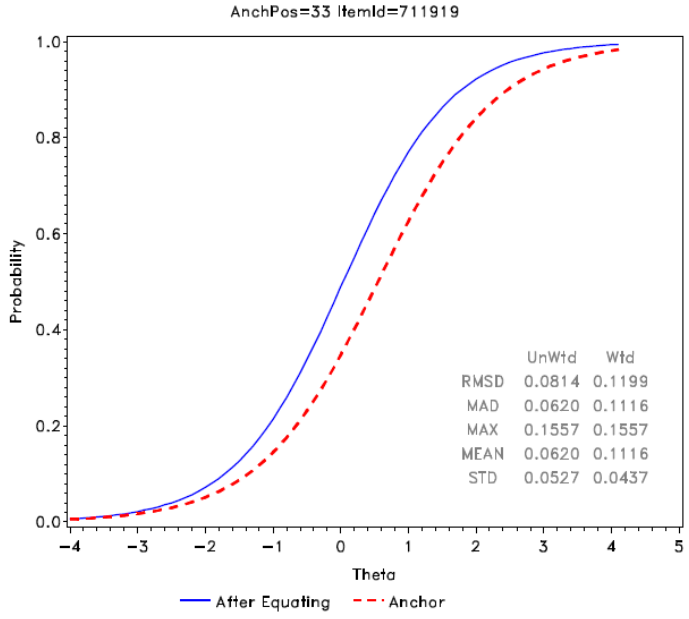


Figure 6.9: Before and after Equating Item Characteristic Curve for Grade 8, From CA3, Item 6, Session 1

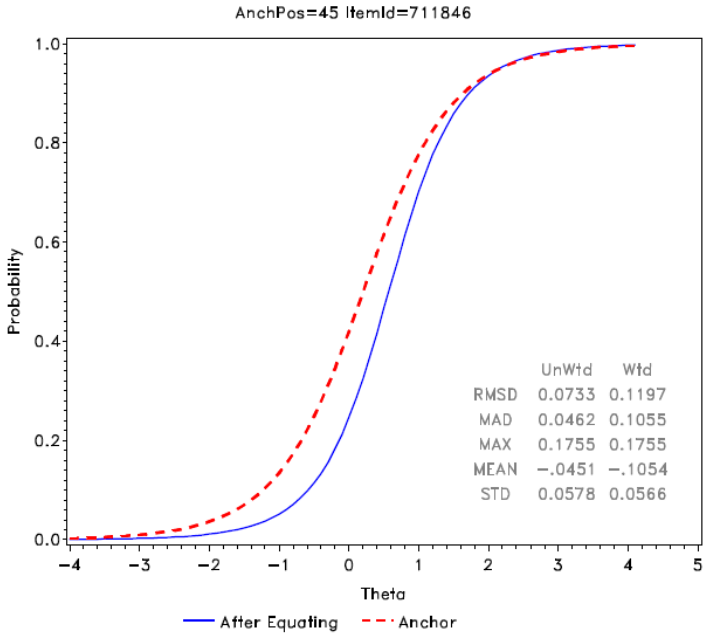


Figure 6.10: Science Grade 5 Anchor TCCs (with all Anchors Included)

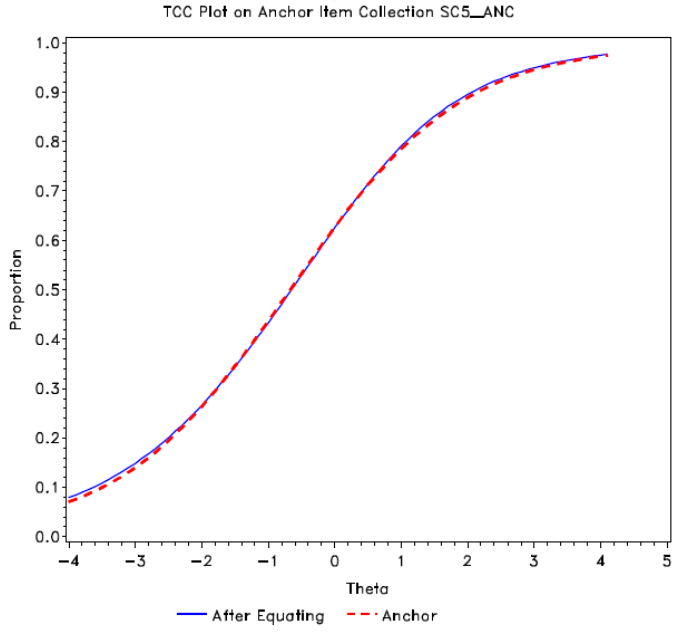


Figure 6.11: Science Grade 5 Anchor TCCs (with One Anchor Removed)

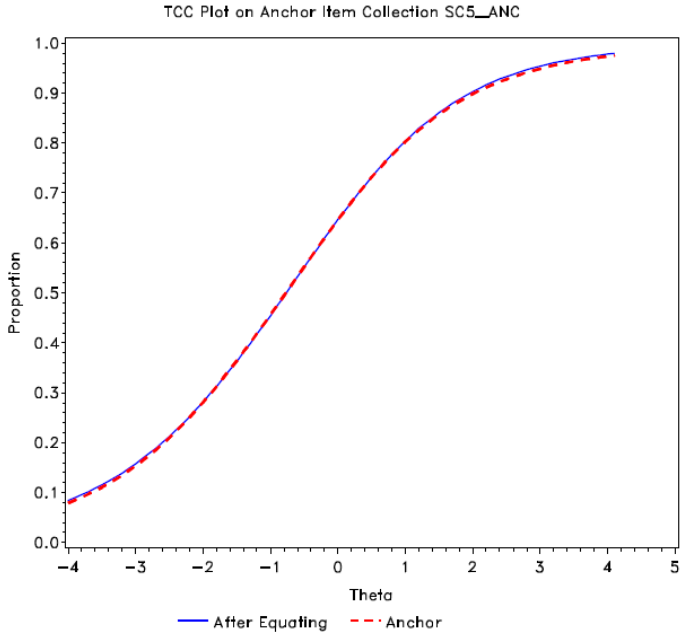


Figure 6.12: Science Grade 8 Anchor TCCs

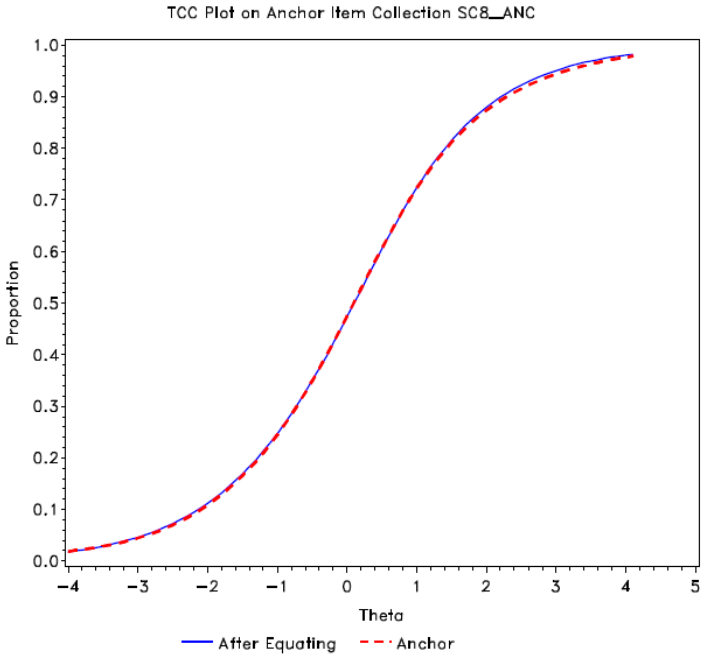


Figure 6.13: Cross-Grade Articulation of Scale Scores at Selected Percentiles, Science MAP

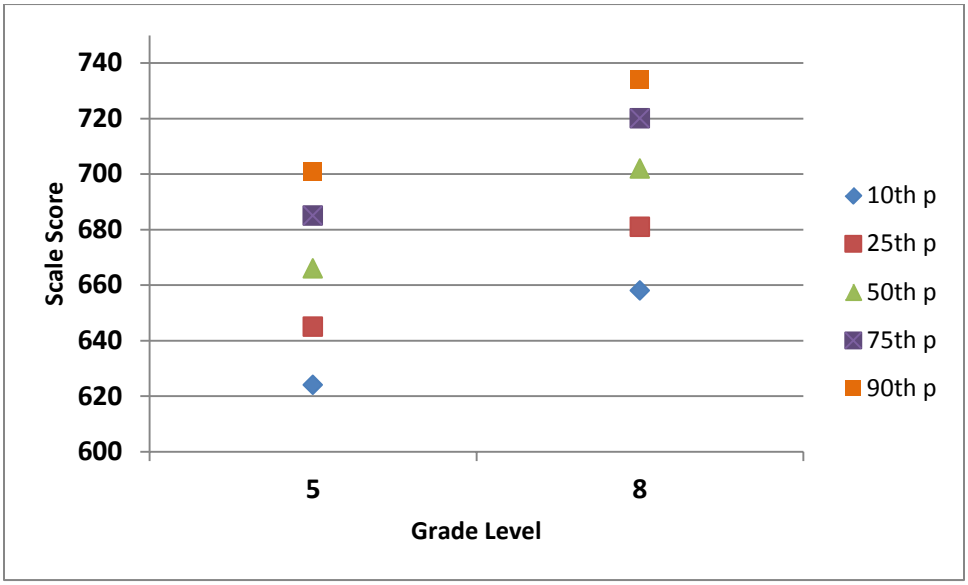
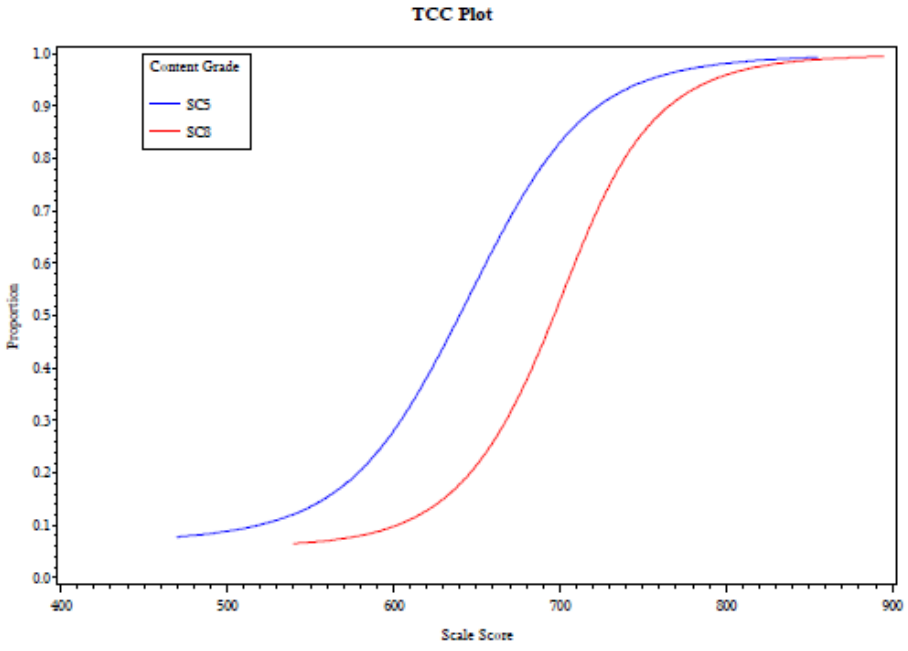


Figure 6.14: Science Test Characteristic Curves by Grade



CHAPTER 7: TEST RESULTS

This chapter of the Technical Report contains information on the results of the Spring 2015 administration of the Science MAP. The scale score results are presented here. Achievement-level information is also provided. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement: *Below Basic*, *Basic*, *Proficient*, and *Advanced.*

While the scale score provides an essential quantitative reference to student achievement, the achievement-level information plainly outlines the meaning of the scores to parents, students, and educators. When combined, scale scores and achievement levels provide a comprehensive set of tools to assess Missouri student achievement in Science.

This chapter also provides descriptions of the score reports, data structure, and interpretive guide. The AERA, APA, & NCME (2014) *Standards* addressed in Chapter 7 are 5.1, 6.10, 7.0, and 12.18. Each standard will be presented in the pertinent section of this chapter.

Results presented below are based on Missouri student 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 tables in this chapter presented as evidence of reliability and validity of the scores from the Science MAP assessments and should not be used for state accountability purposes.

7.1 Student Participation

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

- Gender: Female and Male
- Race/Ethnicity: White, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, and Other
- Accommodations/Designated Supports: Students receiving testing accommodations or designated supports

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 attempted to take the online test or received a test book. These participation rates are summarized in Tables 7.1 to 7.10. Each table shows 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 (teacher-invalidated student are excluded). The Accountable column shows the total number of students who attempted to take an online test or received a test book. These include students who should have received a Science MAP scale score but who did not take the test and could not be assigned a scale score.

7.2 Current Administration Data

The Science MAP assessments were administered to students in Grades 5 and 8. Table 7.11 provides a summary of the scale scores based on the state population for the 2015 administration of the Science assessments.

7.3 Cross-Year, Cross-Sectional Comparisons

It is often desirable to examine the scores of students across time and monitor group performance. This is possible if the test content and the construct measured by the test are comparable from year to year and if the scores are reported on the same scale in multiple years. The test content and the constructs measured by the test remained the same for Science Grades 5 and 8 assessments between the 2008 and 2015 administrations and the 2015 test scores for Science continued to be reported on the same scales as in the previous years. Table 7.12 shows the state-level means for all grades from 2007–08 through 2014–15 for Science.

As shown in Table 7.12, the mean scale score for Grade 5 remained stable (a difference of less than 1 scale score point), and the Grade 8 mean scale score decreased by approximately 4 scale score points between the 2013–14 and the 2014–15 administrations. The overall mean scale score trends for Grades 5 and 8 have been stable across the eight years of the assessment administration with the year-to-year differences ranging from 0 to less than 4 scale score points in any year and in each grade level.

Table 7.13 shows the percentage of students in each achievement level from 2007–08 through 2014–15 on the Science test. In Grade 5, the percentage of students at or above *Proficient* remained stable from 2013–14 to 2014–15 (a difference less than a half of a percentage point), and in Grade 8, the percentage of students at or above *Proficient* decreased by three percent. Again, the overall trend of student proficiency data for Science had been stable across the eight years of the assessment administration with the year-to-year differences in the percentage of students at or above *Proficient* ranging from zero to four percent for Grade 5 and from about half percent to three percent for Grade 8 in any year.

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 (2014) Standard 6.10 states the following:

> When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used. (119)

Standard 5.1 is related in that it states the following:

> Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations. (102)

In compliance with Standards 6.10 and 5.1, interpretations related to the test scores are disseminated in two ways: (1) the individual score report and (2) the *Guide to Interpreting Results* (developed collaboratively by DRC/CTB and DESE staff). The *Guide to Interpreting Results* which includes samples of score reports is presented in Appendix B.

In addition to providing interpretation, it is important that the information is understandable by the target audience. Standard 7.0 of the AERA, APA, & NCME (2014) *Standards* states the following:

> Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores. (125)

In support of Standard 7.0, the *Guide to Interpreting Results* is 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 the 2014–2015 administration, DRC/CTB reported the MAP assessments through the Missouri MAP Online Reporting System. The MAP online reporting system was delivered on the PRISM platform and is a browser-based system designed to deliver online interactive reporting to authorized users at the state and district level for the Missouri public schools.

7.4.1 Description of Each Type of Report

In this section, descriptions for the following reports are provided: Student Roster, Individual Student Report, and Student Score Label. In addition, the Missouri Comprehensive Data System is briefly discussed.

In compliance with AERA, APA, & NCME (2014) Standard 12.18, the MAP score reports provide clear information about individual student achievement and groups of students. Standard 12.18 states the following:

> In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports. (200)

Student Roster

Available from the Missouri Online Reporting System is a Student Roster which displays a list of students based on the specific report filter options selected such as test administration, grade, school, district, gender, race/ethnicity, and examiner. Total test scores and achievement levels indictors are displayed in a table-type format for the content area chosen. Percent correct is reported for the GLE Strands. Upon selecting a student from the roster list their Student Individual Report will open. A PDF of the data displayed can be printed. A sample Student Roster report is provided in Appendix C, Figure C1.

Individual Student Report

The Individual Student Report (ISR) is another type of report available through the Missouri Online Reporting System. The Individual Student Reports are provided to schools to be sent home to the parents. On the left side of the page, student identifying information and the overview of the performance, including the student's MAP scale score results for a given content area, are shown. In the middle of the page, a bar graph along with the student's scale score is shown, along with the achievement level associated with that scale score, for a given content area. This is followed by a brief explanation of what the achievement level means. When a student does not receive a scale score, then his or her achievement level will be labeled "Level Not Determined" (LND). Invalidated students are assigned the lowest obtainable scale score (LOSS) for a given content area and the *Below Basic* achievement level. The ISR also contains brief explanation of the meaning of the content area achievement level indicators.

On the right side of the page the content area achievement level descriptors and scale score ranges for each achievement level are listed. A sample ISR report is provided in Appendix C, Figure C2.

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 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–one for each of the content areas administered. The label lists the student's scale score and achievement level for the content area. DRC/CTB provided multiple labels per student submitted for scoring. The labels are provided in print only. A sample Student Score Label report is provided in Appendix C, Figure C3.

Missouri Comprehensive Data System

Schools and districts are able to access summary level reports through the online Missouri Comprehensive Data System (MCDS). The MCDS allows school district personnel with appropriate permissions to access MAP data at a variety of levels, and to request on-demand, customized reports that are configured and disaggregated in ways that best meet their needs for such activities as evaluating programs, revising curriculum,

and improving teaching and learning. Users access the MCDS from the Data Management tab on DESE's home page (http://dese.mo.gov/). From there, they access the data portal directly through the MCDS link. Each school and/or district is assigned a user name and password so that it can access the site.

7.5 Data Structures

A data file referred to as General Research File (GRF) was provided to DESE by DRC/CTB. 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 the state level GRF is included in Appendix C. Note that the GRF file included information for the three tested content areas: English Language Arts/Literacy, Mathematics, and Science.

7.6 Interpreting Test Results

The student's correct responses to the assessment questions are used to derive a MAP scale score. The scale score describes achievement on a continuum which ranges from 470 to 895 for Science. Science scores can be compared within a grade level in a given test administration and across administrations. The scores can be compared across different forms administered in the same grade level. If an assessment scores are reported on a vertical scale, scale scores from adjacent grades may be compared. Since Science assessment is administered in Grades 5 and 8 which are non-adjacent grades, despite the fact that Science scales have underlying vertical scale properties, comparing the scores across grades should be done with caution. Scale scores cannot be compared across content areas. For example, it is appropriate to compare Science scale scores for two students in the same grade but it is not appropriate to compare their Science scores with their Mathematics scores. The Science MAP scale scores determine the student's achievement level. Student performance can be reported in terms of four performance, or achievement, levels that describe a pathway to proficiency and college and career readiness. Each achievement level represents standards of performance for each assessed content area (ELA, Mathematics, and Science). Achievement-level scores provide a description of what students can do in terms of the content and skills assessed, as described in the Missouri Learning Standards and Grade Level Expectations.

The information on score interpretation is also included in the *Guide to Interpreting Results* (see Appendix B) which was written for Missouri teachers and administrators who receive score reports from the 2014–15 administration of the MAP. This guide has four sections. The first section presents an overview of key terms and test related concepts. The second section discusses assessment terms and types of scores that will be presented on the score reports. The third section presents the achievement-level descriptors for all grade/content areas. Finally, the fourth section presents sample score reports.

7.7 Summary

In summary, the overall purpose of reporting test results is to communicate information on 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 DRC/CTB are in alignment with multiple best practices of the testing industry but, in particular, support the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- Standard 5.1—Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations.
- Standard 6.10—When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used.
- Standard 7.0—Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores.
- Standard 12.18—In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports.

Table 7.1: Participation Rates: All Students

Grade	Accountable in Science	Percent Reportable in Science
5	66412	99.95%
8	66526	99.86%

Table 7.2: Participation Rates: Males

Grade	Accountable in Science	Percent Reportable in Science
5	33965	99.95%
8	33726	99.81%

Table 7.3: Participation Rates: Females

Grade	Accountable in Science	Percent Reportable in Science
5	32447	99.95%
8	32800	99.90%

Table 7.4: Participation Rates: White

Grade	Accountable in Science	Percent Reportable in Science
5	48288	99.95%
8	49009	99.86%

Table 7.5: Participation Rates: Black

Grade	Accountable in Science	Percent Reportable in Science
5	10688	99.96%
8	10684	99.81%

Table 7.6: Participation Rates: Hispanic

Grade	Accountable in Science	Percent Reportable in Science
5	3874	99.97%
8	3618	99.86%

Table 7.7: Participation Rates: Asian/Pacific Islander

Grade	Accountable in Science	Percent Reportable in Science
5	1490	100.00%
8	1418	99.93%

Table 7.8: Participation Rates: American Indian/Alaska Native

Grade	Accountable in Science	Percent Reportable in Science
5	296	100.00%
8	286	100.00%

Table 7.9: Participation Rates: Other Race/Ethnicity

Grade	Accountable in Science	Percent Reportable in Science
5	1776	99.83%
8	1511	99.87%

Table 7.10: Participation Rates: Students Receiving Accommodations or Designated Supports

Grade	Accountable in Science	Percent Reportable in Science
5	15856	99.94%
8	13119	99.64%

Table 7.11: State-Level Scale Score Statistics: Science

Grade	N	Mean SS	S.D. SS	Scale Scores by Percentiles				
				10th	25th	50th	75th	90th
5	66,381	664	31.7	624	645	666	685	701
8	66,430	698	31.7	658	681	702	720	734

Table 7.12: Comparison of State-Level Means, 2008 through 2015 Census Data: Science

Grade	Year	N	Mean SS	S.D. SS
5	2008	65,586	661.64	31.52
	2009	67,118	662.22	30.40
	2010	66,558	664.76	32.48
	2011	67,196	666.04	33.43
	2012	66,492	667.99	34.23
	2013	65,850	667.54	33.03
	2014	65,935	664.06	30.50
	2015	66,381	664.00	31.72
8	2008	67,209	694.36	30.67
	2009	66,702	695.65	30.94
	2010	66,101	698.28	31.07
	2011	65,828	700.05	30.98
	2012	66,724	700.18	31.92
	2013	66,418	699.92	31.71
	2014	66,912	701.94	29.53
	2015	66,430	698.19	31.67

Table 7.13: Comparison of Percentage of Students in Each Achievement Level, Science 2008 through 2015 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
	2010	66,730	0.3	10.4	40.5	29.6	19.3	48.9
	2011	67,461	0.4	10.0	39.1	29.5	21.0	50.5
	2012	66,675	0.3	9.8	38.5	27.2	24.3	51.4
	2013	65,980	0.2	9.6	39.0	28.1	23.1	51.3
	2014	66,153	0.4	9.0	43.3	31.5	15.9	47.3
	2015	66,411	0.0	10.6	42.4	28.8	18.1	47.0
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
	2010	66,463	0.5	16.4	35.1	38.4	9.6	48.0
	2011	66,205	0.6	15.7	33.7	38.6	11.4	50.0
	2012	67,037	0.5	16.1	33.8	37.0	12.6	49.6
	2013	66,710	0.4	15.7	33.8	38.4	11.6	50.0
	2014	67,168	0.4	12.8	35.0	40.5	11.4	51.9
	2015	66,524	0.1	16.7	34.3	39.1	9.8	48.9

CHAPTER 8: ACHIEVEMENT-LEVEL SETTING

In this chapter, we briefly describe the Science MAP achievement-level setting, and we present the cut scores established and the achievement-level descriptors derived from the achievement-level setting.

The first standard setting to establish cut scores for the Science MAP was held in 2008 (refer to the *Missouri Assessment Program Bookmark Standard Setting Technical Report 2008 for Missouri Achievement-Level Setting Grades 5, 8, and 11 Science* [2008]). The reporting scales and the cut scores for Grades 5 and 8 remained unchanged throughout assessment administrations. Given that items not previously field tested in Missouri (developed by the University of Iowa) were included in the 2014–15 Science tests, a validation of Science cut points, after which the existing cut scores were upheld, was performed after the 2014–15 assessment administration.

The AERA, APA, & NCME (2014) Standards addressed in Chapter 8 are 5.21 and 5.22, which will be presented in the pertinent section of this chapter.

A short description of the cut point validation procedure for Science, during which the participants upheld the existing cut scores, is provided in this section of the report. Details of this procedure and results may be found in the *Missouri Assessment Program Cut Point Validation Technical Report Grades 5 and 8 Science* (2015).

Both Technical Reports address AERA, APA, and NCME Standard 5.21:

> When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly. (107)

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

8.1 Cut Point Validation for Grades 5 and 8 Science

The Missouri Department of Elementary and Secondary Education (DESE) partnered with McGraw-Hill Education CTB to conduct a cut point validation (CPV) for the Grades 5 and 8 Science tests of the MAP. The workshop was held in Columbia, Missouri, on June 11–12, 2015. A modification of the Bookmark Standard Setting Procedure (BSSP) (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) was implemented to validate the cut points for the assessments. Workshop participants worked individually and in concert to consider cut points associated with four achievement levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*, with *Advanced* representing the highest level of knowledge, skills, and abilities. As mentioned earlier in

this chapter, the existing cut points for these tests were established in 2008 following a standard setting with Missouri educators using the BSSP (CTB/McGraw-Hill, 2008). A total of 18 participants from across the state of Missouri participated in the workshop. The workshop was facilitated by researchers and content experts from McGraw-Hill Education CTB. During the workshop, participants considered the test items, achievement level descriptors, Grade Level Expectations, and test data. Following the workshop, DESE considered participants' recommendations. After consulting with McGraw-Hill Education CTB and with members of its technical advisory committee, DESE determined that participants' recommendations were consistent with the existing cut points. Accordingly, DESE found that CPV participants validated the existing cut points, and that the existing cut points (and associated achievement level descriptors) are suitable for continued use. Details and outcomes of the Science Cut Point Validation workshop can be found in the *Missouri Assessment Program Cut Point Validation Technical Report for Grades 5 and 8 Science* submitted to DESE in September 2015.

At both workshops, the panelists' work involved multiple rounds of discussion and voting. The process of the cut point validation adhered to the AERA, APA, & NCME (2014) Standard 5.22 which states the following:

> When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way. (108)

8.2 Cut Scores

In this section, we present the cut scores for each grade/content area of MAP. Table 8.1 shows the cut scores for Grades 5 and 8 Science.

8.3 Achievement-Level Descriptors

The short achievement-level descriptors that were adopted by DESE for the reporting purposes are presented on page 6 of the *Guide to Interpreting Results* (see Appendix B.)

8.4 Summary

This chapter presented a brief overview of the cut point validation procedure for Science. This procedure is addressed in more detail in the relevant Technical Report.

The standard setting undertaken by McGraw-Hill Education CTB supports the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- Standard 5.21—When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.
- Standard 5.22—When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the

judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way.

Table 8.1: 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, DRC/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 (2014) Standards 1.13, 1.21, 2.0, 2.3, 2.13, 2.14, 2.15, 2.16, and 2.19. Each standard will be discussed in the pertinent section of this chapter.

9.1 Minimization of Construct-Irrelevant Variance and Construct Underrepresentation

Minimization of construct-irrelevant variance and construct underrepresentation 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) item 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 underrepresentation occurs when the content of the assessment does not reflect the full range of content that the assessment is expected to cover. 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 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 (2014) *Standards* indicate the following:

> The term *reliability* has been used in two ways in the measurement literature. First, the term has been used to refer to the reliability coefficients of classical test theory, defined as the correlation between scores on two equivalent forms of the test, presuming that taking one form has no effect on performance on the second form. Second, the term has been used in a more general sense, to refer to the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported (e.g., in terms of standard errors, reliability coefficients per se, generalizability coefficients, error/tolerance ratios, item response theory(IRT) information functions, or various indices of classification consistency). (33)

In accordance with the AERA, APA, & NCME (2014) *Standards* and in developing and maintaining tests of the highest quality, DRC/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 (2014) *Standards* that this chapter addresses. These include Standards 2.0, 2.3, 2.13, and 2.19, each articulated below.

Standard 2.0 Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use. (42)

Standard 2.3 For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported. (43)

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 2.13 The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score. (45)

The raw score-based SEM is discussed in Section 9.2.2 and is reported in raw score units. The conditional SEM is discussed in Section 9.2.3 and is presented in scale score units. Note that the SEM associated with any type of score is not reported for Science MAP.

Standard 2.19 Each method of quantifying the reliability/precision of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select test takers for reliability/precision analyses and the descriptive statistics on these samples, subject to privacy obligations where applicable, should be reported. (47)

Section 9.2 discusses different ways of measuring test reliability, including reliability of raw scores and test form SEM, IRT-based conditional SEM, and decision consistency of

achievement-level classifications. These statistics were computed based on Missouri student census data.

9.2.1 Test Reliability

The reliability of raw scores by test form 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 the variance of the total 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 are, 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 Science MAP are reported in Table 9.1. These reliability coefficients were computed using Missouri student census data. The reliability coefficients for Science ranged from 0.86 to 0.91 for all forms. These results indicate acceptable reliability coefficients for Science MAP tests.

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:

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

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 SEM at the test form level was computed in raw score metric and is also presented in Table 9.1.

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 (2014) Standard 2.14, which states the following:

> When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score. (46)

In further compliance with Standard 2.14, the CSEM of each cut score is reported in Table 9.2.

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 Science test forms 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 9.2 and range from 8 to 11 scale score points.

Figure 9.1 displays the CSEM curves for each Science grade. 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. Figure 9.1 demonstrates 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

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 (2014) Standard 2.16, which states as follows:

> When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure. (46)

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

DRC/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).

Note that the classification consistency and classification accuracy analyses were conducted using data for regular test forms only. It is not recommended to perform the classification consistency and classification accuracy analyses on the Transcribed forms due to a low number of students taking these forms which may negatively affect the stability of the results, thus making the results interpretation difficult. Classification consistency and classification accuracy conditioned on performance level (Table 9.3) and on cut score (Table 9.4) are presented for the 2015 Science MAP in this section of the

report. As shown in Table 9.3, classification accuracy conditioned on achievement level range from 0.63 to 0.88, and classification consistency conditioned on achievement level range from 0.54 to 0.78. 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, test reliability and associated SEM, and student score distribution.

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 9.4 shows the classification accuracy and consistency estimates when conditioned on MAP cut scores. The classification accuracy statistics are at or above 0.90 for all test forms and all cut points, and the classification consistency statistics are at or above 0.86 for all test forms and all cut points. These results suggest that consistent and accurate performance level classifications are being made for students in Missouri based on the Science 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 Science test is designed to measure a single overall construct—Science achievement; therefore, the items comprising the Science MAP should only measure Science, not Mathematics, 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) 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 four items were flagged for poor model/data fit across all Science tests.

9.3 Principal Components Analysis

As another measure of construct validity, DRC/CTB examined the unidimensionality of each Science 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 measure a single content domain. For example, Science items should measure Science ability and not Reading skills. Standard 1.13 of the AERA, APA, & NCME (2014) *Standards* states the following:

> If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided. (26–27)

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. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other factors.

For both grades of Science MAP, 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 Science MAP tests exhibit first principal components accounting for more than 16% of the test variance (see Table 9.5). To further investigate the unidimensionality of the Science tests, the ratio of the first eigenvalue to the second eigenvalue was explored (also in Table 9.5). These ratios show that the first eigenvalue is at least five times as large as the second eigenvalue for all Science tests. This substantial difference in magnitude indicates that one factor appears to be dominant and that the 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 GLE Strands

Three sets of analyses were conducted at the GLE Strand level for Science in another attempt to assess the construct validity of MAP. First, correlation coefficients that measure the relationship between the Strand scores were computed. Second, the reliability of each GLE Strand was computed. Finally, the SEM was computed for each GLE Strand.

9.4.1 Correlations among GLE Strands Subscores

In this section, we report the strength of the interrelationships among the Claims or GLE Strands by computing correlation between them. Table 9.6 reports the uncorrected

Pearson product-moment (PPM) correlation coefficients and the PPM corrected for attenuation (CAPPM). The PPM among the GLE Strand subscores is presented below the diagonal portion of the matrix, and the CAPPM is presented above the diagonal portion of the matrix.

The uncorrected PPM in Table 9.6 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. In most cases, the PPM coefficients show that performance on one GLE Strand is moderately to strongly related to performance on another GLE Strand within the same grade. The exception was the correlation of 0.29 between Impact of Science, Technology and Human Activity, and Force and Motion GLE strands in one of the Grade 5 forms. As noted above, the value of the correlation coefficients will be affected by the limited number of items measuring each GLE Strand. So, caution should be used when comparing the PPM coefficients measuring the relationships between GLE Strands to those measuring the relationships between content areas (Table 9.8). We expect to see a more modest relationship (smaller correlation coefficients) reported between the GLE Strands as a consequence of the lower number of items measuring each of the reporting categories. The PPM between two GLE Strand subscores may be artificially low because of measurement error.

AERA, APA, & NCME (2014) Standard 1.21, states the following:

> 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. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates. (29)

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 GLE Strands, r_{xx} is the reliability of one of those GLE Strands, and r_{yy} is the reliability for the other GLE Strand.

Across all tables, the CAPPM indicate strong relationships between the GLE Strands. In some cases, the CAPPM is greater than 1.00. "Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed" (Schumacker, 1996). The strong relationships suggested by the CAPPM in Table 9.6 are further evidence of the validity of the test construct. Since the overall content area is comprised of the GLE Strands subscores and the content area is expected to measure a single dimension, we would expect that these subscores are also highly related.

9.4.2 Reliability of GLE Strands

Raw score summary statistics (mean and standard deviation), Cronbach's (1951) coefficient alpha, and SEM were computed for each of the GLE Strands for each test form using the census data. These statistics are presented in Table 9.7. Reliability indices, such as Cronbach's coefficient alpha (and resulting SEM), are a function of the number of test items. It is expected that coefficient alpha would be lower for a GLE Strand assessed by a small number of items compared to a GLE Strand assessed by a larger number of items.

9.4.3 Standard Error of Measurement of GLE Strands

In this chapter, we also report the SEM associated with each of the GLE Strands in Table 9.7. These SEMs are reported in the raw score correct 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 Science MAP tests, correlations were computed between the ELA, Mathematics, and Science scale scores for students who took the Science test and at least one other test in 2015. These correlations are based on the census data and the results are shown in Table 9.8. The correlation coefficients ranged from 0.75 (between Science and Mathematics in Grade 8) to 0.80 (between ELA and Science in Grade 8). The correlation coefficients suggest that individual student scores for ELA and Science, and Mathematics and Science are highly related. Despite high correlations, the tests are not perfectly related to each other, suggesting that different constructs are being tapped; however, the test scores do appear highly related to one another, suggesting they may be tapping into a similar knowledge base or general underlying ability.

9.6 Summary

In summary, the analyses of the internal structure of the test can indicate the degree to which the relationship among test items and test components conform to the test construct which in turn provide basis for test score interpretation. This chapter of the report includes reliability analysis results indicating that the Science MAP tests produce scores that would be relatively stable if the test were administered repeatedly under similar conditions. The assumption that the Science MAP tests were unidimensional (that is the grade level test measured one primary dimension) was confirmed through principal component analysis. In addition, the divergent validity of the Science MAP tests was evaluated through the correlations computed between the ELA, Mathematics, and Science scale scores. These analyses conducted by DRC/CTB are in alignment with

multiple best practices of the testing industry but, in particular, support the following *Standards for Educational and Psychological Testing* (2014):

- Standard 1.13—If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided.
- Standard 1.21—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. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates.
- Standard 2.0—Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use.
- Standard 2.3—For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.
- Standard 2.13—The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score.
- Standard 2.14—When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.
- Standard 2.16—When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure.
- Standard 2.19—Each method of quantifying the reliability/precision of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select test takers for reliability/precision analyses and the descriptive statistics on these samples, subject to privacy obligations where applicable, should be reported.

Table 9.1: Reliability in Science

Grade	Form	Number of Items	Number of Score Points	Cronbach's Alpha	SEM	N-Count
5	CA2	41	60	0.89	3.56	40104
	CA3	41	60	0.86	3.49	26268
8	CA2	39	61	0.91	3.82	38704
	CA3	38	60	0.88	3.78	27464
	CT2	39	60	0.90	3.46	253

Table 9.2: Conditional Standard Error of Measurement at the Basic, Proficient, & Advanced Cut Scores: Science

Grade	Basic		Proficient		Advanced	
	Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM
5	626	10	669	10	692	11
8	671	10	703	8	735	9

Table 9.3: Decision Accuracy and Consistency Conditioned on Level of Achievement: Science

Grade	Form	Accuracy				Consistency			
		Below Basic	Basic	Prof.	Adv.	Below Basic	Basic	Prof.	Adv.
5	CA2	0.86	0.82	0.65	0.84	0.78	0.77	0.55	0.70
	CA3	0.83	0.82	0.63	0.83	0.74	0.77	0.54	0.73
8	CA2	0.86	0.78	0.83	0.70	0.77	0.72	0.73	0.55
	CA3	0.88	0.76	0.82	0.83	0.78	0.73	0.76	0.70

Table 9.4: Decision Accuracy and Consistency at Achievement Cut Points: Science

Grade	Form	Accuracy			Consistency		
		Below Basic/ Basic	Basic/ Prof.	Prof./ Adv.	Below Basic/ Basic	Basic/ Prof.	Prof./ Adv.
5	CA2	0.96	0.90	0.92	0.94	0.86	0.89
	CA3	0.97	0.90	0.90	0.96	0.86	0.86
8	CA2	0.94	0.91	0.95	0.92	0.87	0.94
	CA3	0.96	0.91	0.94	0.94	0.89	0.92

Table 9.5: Principal Component Analysis for Science

Grade	Form	Components	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
5	CA2	First Component	8.07	19.68	19.68
		Second Component	1.38	3.37	23.05
		Ratio (First/Second)	5.83		
	CA3	First Component	6.85	16.71	16.71
		Second Component	1.33	3.23	19.94
		Ratio (First/Second)	5.17		
8	CA2	First Component	9.19	23.56	23.56
		Second Component	1.22	3.13	26.69
		Ratio (First/Second)	7.52		
	CA3	First Component	7.66	20.15	20.15
		Second Component	1.26	3.32	23.47
		Ratio (First/Second)	6.07		

Table 9.6: Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among GLE Strands: Science

Grade	Form	No.	GLE Strand	Number of Items	1	2	3	4	5	6	7	8
5	CA2	1	Matter and Energy	5	.	1.00	0.96	0.98	1.00	0.96	0.92	0.89
		2	Force and Motion	3	0.44	.	0.92	0.93	0.97	0.95	0.94	0.85
		3	Living Organisms	5	0.45	0.36	.	1.03	0.99	0.98	0.85	0.92
		4	Ecosystems	7	0.52	0.42	0.50	.	1.03	0.97	0.86	0.93
		5	Earth's Systems	6	0.51	0.42	0.46	0.54	.	0.95	0.91	0.90
		6	The Universe	4	0.45	0.37	0.41	0.47	0.44	.	0.88	0.93
		7	Scientific Inquiry	9	0.55	0.47	0.46	0.53	0.54	0.47	.	0.80
		8	Impact of Sci., Tech. and Human Activity	2	0.39	0.32	0.37	0.42	0.39	0.37	0.41	.
	CA3	1	Matter and Energy	5	.	0.83	0.95	0.93	0.96	0.89	0.80	0.92
		2	Force and Motion	3	0.32	.	0.80	0.75	0.79	0.79	0.72	0.77
		3	Living Organisms	5	0.34	0.31	.	0.97	0.98	0.94	0.83	0.95
		4	Ecosystems	7	0.40	0.35	0.41	.	0.97	0.92	0.82	0.97
		5	Earth's Systems	5	0.36	0.32	0.37	0.44	.	0.93	0.81	0.98
		6	The Universe	5	0.38	0.37	0.40	0.47	0.42	.	0.82	0.93
		7	Scientific Inquiry	9	0.39	0.38	0.40	0.47	0.41	0.47	.	0.80
		8	Impact of Sci., Tech. and Human Activity	2	0.33	0.29	0.33	0.41	0.37	0.40	0.38	.

Table 9.6: Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among GLE Strands: Science (cont.)

Grade	Form	No.	GLE Strand	Number of Items	1	2	3	4	5	6	7	8
8	CA2	1	Matter and Energy	6	.	1.05	0.94	0.93	0.98	0.98	0.92	0.96
		2	Force and Motion	2	0.50	.	0.91	0.94	0.99	1.02	1.00	0.92
		3	Living Organisms	4	0.56	0.42	.	0.93	0.97	0.93	0.86	0.94
		4	Ecosystems	4	0.53	0.43	0.52	.	0.97	0.97	0.86	1.01
		5	Earth's Systems	7	0.59	0.47	0.57	0.55	.	1.00	0.88	0.96
		6	The Universe	4	0.50	0.41	0.46	0.47	0.50	.	0.91	0.96
		7	Scientific Inquiry	9	0.63	0.54	0.58	0.55	0.60	0.52	.	0.87
		8	Impact of Sci., Tech. and Human Activity	3	0.39	0.30	0.38	0.39	0.39	0.33	0.40	.
	CA3	1	Matter and Energy	6	.	0.95	1.02	0.92	0.99	1.00	0.93	1.04
		2	Force and Motion	2	0.40	.	0.93	0.92	0.95	0.94	0.88	1.05
		3	Living Organisms	4	0.49	0.37	.	0.98	1.01	0.95	0.87	1.04
		4	Ecosystems	4	0.39	0.32	0.39	.	1.00	0.87	0.80	1.03
		5	Earth's Systems	6	0.46	0.37	0.45	0.39	.	0.95	0.82	1.08
		6	The Universe	4	0.55	0.43	0.49	0.40	0.48	.	0.85	1.02
		7	Scientific Inquiry	9	0.57	0.45	0.50	0.41	0.46	0.56	.	0.93
		8	Impact of Sci., Tech. and Human Activity	3	0.44	0.37	0.41	0.36	0.42	0.46	0.47	.
	CT2	1	Matter and Energy	6	.	1.44	1.01	0.94	1.11	0.92	0.90	1.07
		2	Force and Motion	2	0.49	.	1.34	1.44	1.32	1.56	1.35	1.45
		3	Living Organisms	4	0.59	0.44	.	1.03	1.12	1.00	0.91	1.10
		4	Ecosystems	4	0.55	0.48	0.59	.	1.08	0.98	0.78	1.18
		5	Earth's Systems	7	0.61	0.41	0.60	0.58	.	1.00	0.86	1.02
		6	The Universe	4	0.39	0.38	0.42	0.41	0.39	.	1.01	1.09
		7	Scientific Inquiry	9	0.62	0.52	0.61	0.52	0.54	0.49	.	0.92
		8	Impact of Sci., Tech. and Human Activity	3	0.45	0.34	0.45	0.48	0.39	0.32	0.44	.

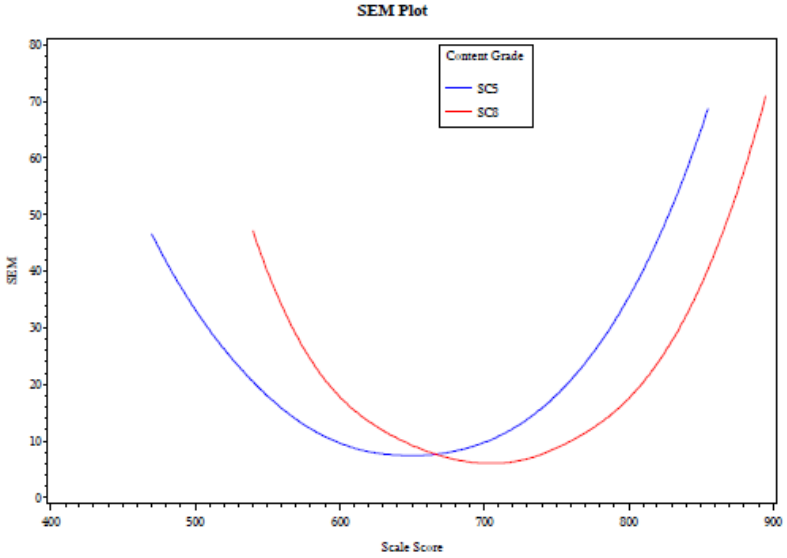
Table 9.7: Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Science GLE Strands

Grade	Form	GLE Strand	Number of Items	Number of Score Points	N Count	Mean	Std. Dev.	Cronbach's Alpha	SEM
5	CA2	1	5	7	40104	4.12	1.77	0.52	1.23
		2	3	5	40104	2.12	1.35	0.37	1.07
		3	5	7	40104	4.70	1.51	0.42	1.15
		4	7	9	40104	5.93	1.89	0.55	1.27
		5	6	8	40104	4.35	1.83	0.50	1.29
		6	4	6	40104	4.21	1.41	0.42	1.07
		7	9	14	40088	7.20	3.35	0.69	1.85
		8	2	4	40104	2.75	1.01	0.38	0.80
	CA3	1	5	7	26268	4.81	1.46	0.36	1.17
		2	3	5	26268	3.11	1.46	0.42	1.11
		3	5	7	26268	4.91	1.40	0.35	1.13
		4	7	9	26268	6.87	1.64	0.51	1.14
		5	5	7	26268	4.65	1.42	0.40	1.10
		6	5	7	26268	4.96	1.59	0.52	1.10
		7	9	14	26260	7.93	3.14	0.65	1.86
		8	2	4	26268	3.10	1.07	0.35	0.86
8	CA2	1	6	8	38704	4.20	2.13	0.61	1.33
		2	2	4	38704	1.47	1.15	0.38	0.91
		3	4	6	38704	3.01	1.70	0.58	1.10
		4	4	6	38704	3.95	1.53	0.54	1.04
		5	7	9	38704	4.58	2.28	0.59	1.46
		6	4	6	38704	3.36	1.52	0.43	1.15
		7	9	17	38664	7.54	4.42	0.77	2.12
		8	3	5	38704	2.77	1.41	0.28	1.20
	CA3	1	6	8	27464	4.45	1.98	0.51	1.39
		2	2	4	27464	1.99	1.24	0.35	0.99
		3	4	6	27464	3.15	1.53	0.45	1.13
		4	4	6	27464	3.68	1.45	0.35	1.17
		5	6	8	27464	4.37	1.50	0.43	1.13
		6	4	6	27464	3.16	1.70	0.59	1.09
		7	9	17	27436	8.36	4.19	0.73	2.16
		8	3	5	27464	2.98	1.34	0.35	1.08
	CT2	1	6	8	253	2.74	1.99	0.60	1.26
		2	2	4	253	0.71	0.79	0.19	0.71
		3	4	6	253	1.84	1.51	0.57	0.99
		4	4	6	253	2.77	1.69	0.58	1.10
		5	7	9	253	3.02	1.93	0.50	1.36
		6	4	6	253	2.28	1.37	0.30	1.15
		7	9	16	253	4.11	3.64	0.78	1.71
		8	3	5	253	1.92	1.40	0.29	1.18

Table 9.8: Inter-Correlation of Science Scale Scores with English Language Arts/Literacy and Mathematics Scale Scores

Grade	ELA/SC	MA/SC
5	0.79	0.77
8	0.80	0.75

Figure 9.1: CSEM Curves Science, Grades 5 and 8



CHAPTER 10: FAIRNESS

As noted in the *Standards* (AERA, APA, & NCME, 2014), 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 Science MAP. It should be noted that differences in test performance among subgroups do 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 (2014) Standards 3.1 through 3.6. These standards are from Chapter 3 of the AERA, APA, & NCME (2014) *Standards*, "Fairness in Testing." Each of these standards will be presented as will be the way the standard is addressed in this chapter.

Standard 3.6 Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (65)

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, DRC/CTB has several steps that are followed in item development and selections as is explained in Section 10.1 of this chapter. In addition, DESE conducts content and bias reviews on items as explained in Chapter 3. These practices adhere to Standard 3.3:

> Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test. (64)

DRC/CTB conducts Differential Item Functioning (DIF) studies following the operational administration of Science MAP. Typically items are evaluated for possible DIF in the field test phase of the test development and items flagged for DIF are typically further examined for possible bias. Also, Section 10.2 of this chapter explains the steps taken to evaluate MAP items through the use of DIF in order to adhere with this standard.

In addition, standardized test administration and extensive training of test scores for Science MAP comply with Standards 3.4 and 3.5:

Standard 3.4 Test takers should receive comparable treatment during the test administration and scoring process. (65)

Standard 3.5 Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population. (65)

Section 10.1 of this chapter is also directly relevant to Standards 3.1 and 3.2.

Standard 3.1 Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

Standard 3.2 Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

In this section, we explain the steps taken by DRC/CTB to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Section 3.3 of Chapter 3 discusses the item review conducted for Science MAP items. This review is also critical in fulfilling Standards 3.1 and 3.2.

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. DRC/CTB used *Guidelines for Bias-Free Publishing* (Macmillan/McGraw-Hill, 1993a) and *Reflecting Diversity: Multicultural Guidelines for Educational Publishing Professionals* (Macmillan/McGraw-Hill, 1993b) to guide them through development of Science items included in the 2014–15 assessments. Test developers reviewed all 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 typically 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 (DIF) Statistics

After administering the test, an empirical approach known as differential item functioning (DIF) was used to examine the items. The DIF statistics indicate the degree to which members of a particular subgroup 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. It should be noted, though, that all items included on the MAP have been thoroughly reviewed for content and bias by Missouri educators and DRC/CTB content experts ensure that they do not tap knowledge or specific ability irrelevant to the construct the test intends to measure. Therefore DIF flags 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). Items are not necessarily suppressed from operational scoring if they are flagged for DIF.

The position of DRC/CTB 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 such biases, DRC/CTB strives to minimize the role of 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 follows (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 k^{th} 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:

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

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:

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

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 $1.0 \leq |\text{MH D-DIF}| < 1.5$
- Large DIF: Significant MH chi-square statistic ($p < 0.05$) and $|\text{MH D-DIF}| \geq 1.5$

For constructed-response items, an effect size (ES) statistic based on the MH chi-square is 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 Science MAP raw score. SMD is computed as follows (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. Table 10.1 shows the DIF results for the following subgroups:

- ***Gender***: Focal group is females; reference group is males.

- ***Race/Ethnicity***: Focal groups are students whose race/ethnicity is reported as Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, or Other; reference group is students whose race/ethnicity is reported as White.

- ***Accommodations/Designated Supports***: Focal group is students who had one or more testing accommodations or designated supports indicated by a teacher; 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 minimum case count for the focal group was set at 200 and the minimum case count for the reference group was set at 400. The DIF analyses are not performed for subgroups of fewer than 200 students. In these cases, the statistical procedures do not have sufficient power to detect differences should they exist.

Table 10.1 summarizes the number of DIF flags by grade and test form for each focal group that included at least 200 students for Science assessments. The analyses were conducted by test form. As noted previously, multiple test forms were administered within one grade and content area in 2014–15 MAP. Consequently, the number of American Indian/Alaska Native students taking each form was smaller than 200 and no DIF was performed for this group on any test form.

For example, consider Grade 5 Science, form CA2 (see Table 10.1). In this form, one item was flagged for DIF for the female subgroup and it exhibited moderate negative DIF. Three items were flagged for DIF for the Asian/Pacific Islander subgroup. All three items exhibited moderate negative DIF. One item was flagged for the Black subgroup and one item was flagged for the Hispanic subgroup: in each case the flagged item displayed moderate negative DIF. Lastly, one item was flagged for the accommodated subgroup and it exhibited moderate positive DIF.

Again, any items included on the Science MAP (including those items flagged for DIF) have been thoroughly reviewed for content and bias by Missouri teachers, DESE staff, and DRC/CTB Content experts and deemed appropriate for inclusion in the MAP assessments.

10.3 Evaluating Bias through Impact Analysis

The impact of achievement testing on minority subgroups can be determined and reported in the form of average scores and also in terms of test score reliability. Tables 10.2 and 10.3 present the number of students, scale score means and standard deviations, effect size (Cohen's *d*), and test form reliability statistics (coefficient alpha, see Chapter 9) for the various subgroups of interest.

10.3.1 Reliability

Tables 10.4 and 10.5 show the regular test form reliability coefficients and SEM by student race/ethnicity, gender, and whether or not students are using any testing accommodations or designated supports. The reliability coefficients ranged from 0.83 to 0.93. This analysis shows that the test reliability is of acceptable magnitude for all of the subgroups. Test reliability by subgroup of students was not computed for transcribed forms due to low number of students taking these forms.

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) offers 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 10.2 and 10.3. There is a large difference between the mean Science test scores of Black students compared to White students in both grades, where Black students underperform White students. There is a medium difference between mean Science test scores of Hispanic students compared to White students in Grade 5 and a small difference in Grade 8, where Hispanic students underperform White students. There is a small difference between the mean Science test scores of American Indian/Alaska Native students compared to White students in both grades, where American Indian/Alaska Native students underperform White students. In addition, there is a large difference between the mean Science test scores of students not needing testing accommodations or designated supports and students using testing accommodation or designated supports in both grades, where students not using testing accommodations or designated supports outperform students using testing accommodations or designated supports.

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 supports multiple best practices of the testing industry, and in particular is related to the following *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- Standard 3.1—Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population.
- Standard 3.2—Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.

- Standard 3.3—Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test.
- Standard 3.4—Test takers should receive comparable treatment during the test administration and scoring process.
- Standard 3.5—Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population.
- Standard 3.6—Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws.

Table 10.1: 2015 MAP DIF Statistics: Number of Flagged Items, Science

Science		Grade	5		8	
		Form	CA2	CA3	CA2	CA3
		# of Items	41	41	39	38
Group	**DIF Magnitude**	**DIF Direction**	**Number of Flagged Items**			
Female	Moderate	Negative	1	1	3	1
		Positive	0	0	2	1
	Large	Negative	0	0	0	0
		Positive	0	0	1	1
Asian/ Pacific Islander	Moderate	Negative	3	2	1	1
		Positive	0	1	1	0
	Large	Negative	0	0	0	0
		Positive	0	0	0	1
Black	Moderate	Negative	1	4	1	0
		Positive	0	1	1	2
	Large	Negative	0	0	0	0
		Positive	0	0	0	0
Hispanic	Moderate	Negative	1	1	0	0
		Positive	0	1	0	0
	Large	Negative	0	0	0	0
		Positive	0	0	0	0
Other	Moderate	Negative	0	0	0	0
		Positive	0	0	0	0
	Large	Negative	0	0	0	0
		Positive	0	0	0	0
Accommo-dations/ Designated Supports	Moderate	Negative	0	0	0	1
		Positive	1	0	0	0
	Large	Negative	0	0	0	0
		Positive	0	0	0	0

Table 10.2: Impact Analysis, Grade 5 Science

Category	Group	N	Mean	Std. Dev.	Effect Size
Race/ Ethnicity	White (not Hispanic)	48265	669.31	29.04	
	Asian/Pacific	1490	673.06	34.10	-0.13
	Black (not Hispanic)	10684	642.41	32.97	0.90
	Hispanic	3873	654.24	32.04	0.51
	Am. Indian/Alaska	296	662.08	29.23	0.25
	Other	1773	663.49	30.67	0.20
Gender	Male	33949	664.09	32.37	
	Female	32432	663.90	31.02	0.01
Accommodations / Designated Supports	No	50534	669.94	28.51	
	Yes	15847	645.04	33.91	0.83

Table 10.3: Impact Analysis, Grade 8 Science

Category	Group	N	Mean	Std. Dev.	Effect Size
Race/ Ethnicity	White (not Hispanic)	48941	703.73	28.30	
	Asian/Pacific	1417	709.10	34.97	-0.19
	Black (not Hispanic)	10664	674.38	34.04	1.00
	Hispanic	3613	689.73	31.82	0.49
	Am. Indian/Alaska	286	695.47	28.41	0.29
	Other	1509	697.49	31.30	0.22
Gender	Male	33663	697.82	32.84	
	Female	32767	698.57	30.41	-0.02
Accommodations / Designated Supports	No	53358	703.69	28.08	
	Yes	13072	675.77	35.38	0.94

Table 10.4: Grade 5 Science Reliability and SEM by Subgroup

Form	Category	Group	N Count	Cronbach's Alpha	SEM
CA2	Race/Ethnicity	White (not Hispanic)	28592	0.87	3.54
		Asian/Pacific Islander	933	0.90	3.52
		Black (not Hispanic)	6751	0.88	3.59
		Hispanic	2563	0.88	3.59
		Am. Indian/Alaska N.	180	0.86	3.52
		Other	1086	0.87	3.56
	Gender	Male	20940	0.89	3.56
		Female	19164	0.88	3.55
	Accommodations/ Design. Supports	No	25424	0.85	3.53
		Yes	14680	0.89	3.57
CA3	Race/Ethnicity	White (not Hispanic)	19668	0.83	3.44
		Asian/Pacific Islander	557	0.83	3.40
		Black (not Hispanic)	3931	0.88	3.64
		Hispanic	1310	0.86	3.55
		Am. Indian/Alaska N.	116	0.88	3.46
		Other	687	0.87	3.53
	Gender	Male	13002	0.86	3.47
		Female	13266	0.86	3.50
	Accommodations/ Design. Supports	No	25104	0.85	3.48
		Yes	1165	0.89	3.62

Table 10.5: Grade 8 Science Reliability and SEM by Subgroup

Form	Category	Group	N Count	Cronbach's Alpha	SEM
CA2	Race/Ethnicity	White (not Hispanic)	28116	0.90	3.83
		Asian/Pacific Islander	847	0.93	3.74
		Black (not Hispanic)	6474	0.90	3.66
		Hispanic	2234	0.90	3.79
		Am. Indian/Alaska N.	172	0.89	3.82
		Other	864	0.90	3.81
	Gender	Male	19956	0.91	3.77
		Female	18748	0.90	3.85
	Accommodations/ Design. Supports	No	27060	0.89	3.85
		Yes	11646	0.91	3.68
CA3	Race/Ethnicity	White (not Hispanic)	20668	0.86	3.78
		Asian/Pacific Islander	559	0.88	3.65
		Black (not Hispanic)	4128	0.88	3.65
		Hispanic	1355	0.87	3.81
		Am. Indian/Alaska N.	113	0.86	3.79
		Other	640	0.89	3.77
	Gender	Male	13544	0.89	3.74
		Female	13922	0.88	3.81
	Accommodations/ Design. Supports	No	26260	0.88	3.79
		Yes	1205	0.90	3.58

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. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

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. (2002). 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.

Cizek, G. J. & Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.

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). *TerraNova 3rd Edition Technical Addendum: Forms E and F*. Monterey, CA: Author.

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

DRC/CTB (2015). *Missouri Assessment Program Cut Point Validation Technical Grades 5 and 8 Science.* 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.

Human Resources Research Organization [HumRRO] (2014). *Independent Alignment Review of the Science Missouri Assessment Program (MAP)*. Report No. 70, prepared for Missouri Department of Elementary and Secondary Education.

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.

Lewis D.M., Mitzel, H. C., Green, D. R. (1996). Standard Setting: A Bookmark Approach. In D. R. Green (Chair), *IRT-Based Standard-Setting Procedures Utilizing Behavioral Anchoring*. Symposium presented at the 1996 Council of Chief State School Officers 1996 National Conference on Large Scale Assessment, Phoenix, AZ.

Lewis, D. M., Mitzel, H. C., Mercado, R. L, & Schultz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations*. New York: Routledge.

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.

McGraw-Hill Education CTB (2015). *Guide to Interpreting Results.* Monterey, CA: Author.

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.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.

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

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

Webb, N. L. (2005). *Webb alignment tool: Training manual*. Madison, WI: Wisconsin Center for Education Research. Available: http://www.wcer.wisc.edu/WAT/index.aspx.

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.

Table of Contents

Appendix A
Test Administration Manual

Missouri Assessment Program

Grade-Level Assessments
Test Administration Manual

English Language Arts / Literacy
Grades
3, 4, 5, 6, 7, 8

Mathematics
Grades
3, 4, 5, 6, 7, 8

Science
Grades
5, 8

Spring 2015
February 27, 2015

Missouri Department of Elementary and Secondary Education

2707075

NOTICE OF NON-DISCRIMINATION

It is the policy of the Missouri Department of Elementary and Secondary Education not to discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs or employment practices as required by Title VI and VII of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination Act of 1975 and Title II of the Americans with Disabilities Act of 1990.

Inquiries related to Department employment practices may be directed to the Jefferson State Office Building, Human Resources Director, 8th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 751-9619 or TYY (800) 735-2966. Inquiries related to Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator–Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 526-4757 or TTY (800) 735-2966, email civilrights@dese.mo.gov.

Anyone attending a meeting of the State Board of Education who requires auxiliary aids or services should request such services by contacting the Executive Assistant to the State Board of Education, Jefferson State Office Building, 205 Jefferson Street, Jefferson City, MO 65102-0480; telephone number (573) 751-4446 or TTY (800) 735-2966.

Inquiries or concerns regarding civil rights compliance by school districts or charter schools should be directed to the local school district or charter school Title IX/non-discrimination coordinator. Inquiries and complaints may also be directed to the Office for Civil Rights, Kansas City Office, U.S. Department of Education, 8930 Ward Parkway, Suite 2037, Kansas City, MO 64114; telephone number (816) 268-0550; FAX: (816) 823-1404; TDD: (877) 521-2172.

Developed and published by CTB/McGraw-Hill LLC, 20 Ryan Ranch Road, Monterey, California 93940-5703.

This *Test Administration Manual* is NOT a secure document. All administrators should read this manual before administering the test.

Table of Contents

February 27, 2015 (Version 2) Updates

Page 8: Changed ELA PT "sessions" to "parts"
Pages 15 and 17: Moved Spell Check from the Designated Supports table to the Universal Tools table
Page 17: Specified that stacked translations are in Spanish
Page 23: Added the date the paper-and-pencil tests are available
Pages 26–30: Clarified information and roles regarding test security
Page 32: Clarified permitted materials for ELL students
Page 45: Corrected the name of the system that generates the paper-and-pencil tests
Page 57: Added a paper-based assessment name and clarified return shipping instructions
Page 64: Updated the instructions regarding the Missing Test Materials Form

1.0 OVERVIEW OF IMPORTANT INFORMATION FOR THE MAP GRADE-LEVEL ASSESSMENTS

1.1 This Test Administration Manual

The purpose of this *Test Administration Manual* is to provide detailed instructions for administering the Missouri Assessment Program Grade-Level Assessments. The manual includes instructions for test preparation, scripts for administering the tests (including links to secure listening scripts for accommodated versions of the assessments), and post-test administration procedures. District Test Coordinators (DTCs), School Test Coordinators (STCs), and Test Examiners (TEs) should thoroughly read the manual and attend training before administering the tests.

1.2 Glossary of Terms

Accommodations	Changes in procedures or materials that increase equitable access during the MAP Grade-Level Assessments. Assessment accommodations allow students to access assessment content to show what they know and can do. Accommodations are available for students with documented Individualized Education Programs (IEPs) or 504 Plans.
Break	Provides an opportunity to pause the test for up to 20 minutes.
Classroom Activity	A short, teacher-led activity designed to introduce students to the context and contextual vocabulary in the performance task (PT) to ensure that students are not disadvantaged in demonstrating the skills the task intends to assess.
Designated Supports	Accessibility features of the assessments available for use by any student for whom the need has been indicated by a team of educators knowledgeable about the student.
eDIRECT	The administrative side of the platform—the Missouri Assessment Program Portal—from which district personnel will manage the assessments.
INSIGHT	INSIGHT is the secure, browser-based test engine for the MAP Grade-Level Assessments.
Item	A test question or stimulus presented to a student to elicit a response.
Pause	Action taken by a student or Test Examiner (TE) to temporarily halt the test during any part of the test, as needed.
Performance Event	A performance event comprises Session 3 of the MAP Grade-Level Science Assessment. It is designed to provide students with an opportunity to demonstrate their ability to apply their knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario.

Performance Task (PT)	A PT is an English Language Arts (ELA) and mathematics item type designed to provide students with an opportunity to demonstrate their ability to apply their knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario. It is a required portion of the test for grades 5 and 8.
Segment	A part of a test within the test delivery system. Segments separate items from others if the eligible tools are different (i.e., the mathematics test may have two segments—one segment that allows calculator use and another segment that does not).
Session	A specific part of a test assigned to a specific student, which is grouped by Test Examiner according to the precode file.
Stimulus/Stimuli	Material or materials used in the test context, which form the basis for assessing the knowledge and skills of students. Many items/tasks for the assessments include a stimulus along with a set of questions to which the student responds. Examples of stimuli include, but are not limited to, traditional reading passages/texts viewed on a computer screen, images with audio presentations, and simulated web pages.
Universal Tools	Universal tools are available to all students based on student preference and selection. Some tools, such as a ruler and a digital notepad, are embedded in the online system, while others, such as a physical thesaurus and scratch paper, are external to the system. The availability of particular universal tools varies by item.

1.3 About the Tests

- The Missouri State Board of Education identified the following purposes for the MAP Grade-Level Assessments:
    - Measuring and reflecting student mastery toward post-secondary readiness
    - Identifying students' strengths and weaknesses
    - Communicating expectations for all students
    - Serving as the basis for state and national accountability plans
    - Evaluating programs
    - Providing professional development for teachers
- The MAP Grade-Level Assessments are designed to adapt testing to the needs of Missouri districts, schools, teachers, and students, while meeting state and federal requirements.

  The MAP Grade-Level Assessments are based on the Missouri Learning Standards, which align to college and career readiness standards. The 2015 assessments will include traditional selected-response and constructed-response items, as well as performance events, performance tasks, and innovative technology enhanced items designed to elicit student evidence in new ways. See Appendix A: Item Types.

  The Missouri Department of Elementary and Secondary Education (DESE) uses the information obtained through the MAP Grade-Level Assessments to monitor the progress of Missouri's students in meeting the Missouri Learning Standards, to inform the public and the state legislature about students' performance, and to help make informed decisions about educational issues.
- CTB is collaborating with the Data Recognition Corporation (DRC) and DESE to deliver Missouri's Spring 2015 Grade-Level Assessments. Missouri educators will use DRC's eDIRECT online platform for enrollment and test administrator processes and INSIGHT (DRC's online delivery system) for test delivery. CTB will provide handscoring and reporting services. These cooperative efforts and systems comprise a fully integrated assessment platform to meet the needs of DESE staff, educators, students, and other Missouri stakeholders.
- The Spring 2015 MAP Grade-Level Assessments include the following:

  English Language Arts/Literacy Assessment for Grades 3–8
  Mathematics Assessment for Grades 3–8
  Science Assessment for Grades 5 and 8
- The English Language Arts and Mathematics Assessments for Grades 3–8 include a component containing selected-response, short constructed-response, and technology enhanced items. This component is divided into two sessions that may be administered in two sittings.
- In addition to the first component, the English Language Arts and Mathematics Assessments for Grades 5 and 8 include a second component containing a performance task. The ELA performance task consists of two parts, which is administered in two sessions. The Mathematics performance task is administered in one session. Both ELA and Mathematics performance tasks are preceded by a Classroom Activity, which is administered as an independent session.

- The Science Assessments consist of three sessions. The first session contains constructed-response items, the second session contains selected-response items, and the third session contains a performance event.
- All MAP Grade-Level Assessments are available only in INSIGHT, the secure online browser, unless a Large Print, Braille, or paper-and-pencil edition is required by the student as an accommodation.

1.4 Schedule of Important Dates for Spring 2015

Precode File Due to DESE	Precode Data Available in eDIRECT	MAP Grade-Level Assessment Test Window
January 30, 2015	March 5, 2015	March 30, 2015–May 22, 2015

Event	Schedule
District Test Coordinators receive welcome email and login information to eDIRECT.	January 7, 2015
District Test Coordinators provide grade-content test windows, purchase order numbers, and Large Print and Braille orders.	Must be provided through eDIRECT Enrollments between January 12, 2015–February 20, 2015. See the eDIRECT User Guide for detailed instructions. The deadline for ordering additional Large Print and Braille testing materials is May 11, 2015. Purchase Orders must be faxed to CTB at 1-888-282-0526 by February 20, 2015.
School Test Coordinators and School Information Technology Coordinators coordinate the installation of INSIGHT on all student workstations and complete a site certification.	Site certification must be completed before the statewide administration window. The Statewide Readiness Test (SRT) window is February 10, 2015–February 27, 2015.
School Test Coordinators verify that all student accommodations and status codes are recorded.	Starting March 5, 2015, indicate through eDIRECT Test Setup any accommodations and designated supports that will be used for each student. All accommodations and designated supports **must** be marked prior to testing. See the eDIRECT User Guide for detailed instructions.
District Test Coordinators contact CTB to schedule pickup of Large Print, Braille, and paper-and-pencil test books.	The deadline is May 26, 2015. Materials must be picked up no later than May 29, 2015.
Test results and Individual Student Reports (ISR) are available online via eDIRECT.	ISRs for ELA and Mathematics are available no later than the close of business on the 10th business day after each district content area testing window closes. ISRs for Science are available July 1, 2015. In future years, ISRs for all content areas will be available 10 business days after the completion of testing.

1.5 Test Administration Policies

General Rules of Online Testing

Middle school students taking the Algebra I EOC Assessment should not be precoded for or administered the Grade-Level Mathematics Assessment.

Starting in 2015, students in grades 3, 4, 6, and 7 will take online tests for English Language Arts and Mathematics consisting of selected-response (SR), constructed-response (CR), and technology enhanced (TE) items. Students in grades 5 and 8 will also take performance tasks (PTs) for ELA and mathematics, as well as online science tests. The SR, CR, and TE items component and the PT component will be presented as separate tests. Students may not return to a test once it has been completed and submitted. Basic online testing parameters:

- Within each test there may be segments. For example, the grades 6 through 8 mathematics tests include a segment with an embedded calculator available and another segment where the embedded calculator is not allowed and is unavailable for testing. A student may not return to a segment once it has been completed and submitted.
- Some items include multiple parts over more than one page. Students may need to use the vertical scroll bar to view an entire item on a page.
- Students may mark items for review and return to those items within a session (or segment for tests with segments).

The "Pause" feature allows a student to pause a test, either to take a short break of up to 20 minutes or to continue testing at a later time.

Pause Rules

The INSIGHT system includes a "Pause" feature that allows a student to pause a test, either to take a short break of up to 20 minutes or to continue testing at a later time as indicated by the district's testing schedule. While the test is paused, a large count-down timer displays in the INSIGHT system on the student's computer. This allows the Test Examiner to easily monitor which students have activated the feature and how much time remains in their break. If a student does not resume testing before 20 minutes elapses, then the student is logged out of the test and is required to log back in to the test using the login and password from his or her Test Ticket. Students may also choose to exit the test from the Pause screen.

During the assessments:

- If a test is paused for 20 minutes or more, the student can return to the section and continue entering his or her responses. The student may also review and change previously answered items. The student is not permitted to return to items in a different segment.
- Any highlighted text and sticky notes **will** be saved when a test is paused regardless of how long the assessment is paused.
- In the event of a technical issue (e.g., power outage or network failure), students will be logged out and the test will automatically be paused. Student responses will not be lost, and students may move to a different device. The students will need to log in again upon resuming the test.

Test Timeout (Due to Inactivity)

As a security measure, students are automatically logged out of the test after 20 minutes of inactivity. *Activity* is defined as selecting an answer or navigation option in the assessment (e.g., clicking [Next] or [Back] or using the quick navigation drop-down list to move to another item). Moving the mouse or clicking on an empty space on the screen is not considered activity. Test timeout occurs when the test is not paused.

Any highlighted text and sticky notes will be saved when a test is paused regardless of how long the test is paused.

If a student starts the test near the end of the testing window, the student must finish before the district administration window officially closes. The assessment will automatically end at 8 P.M. on the last day of the scheduled district administration window, even if the student has not finished.

1.6 Scheduling the Tests

The table below lists rough estimates of the time it will take most students to complete each component of the online MAP Grade-Level Assessments. These times do not include time needed to start computers, load secure browsers, and log in students. Nor do they include time needed for students to complete the INSIGHT Tutorials.

These times also do not account for breaks. This information is for scheduling purposes only, as the assessments are **untimed**.

Content Area	Grades	Test Component	Estimated Time hrs : mins
English Language Arts/ Literacy	3, 4, 6, 7	SR, CR, and TE Items (Sessions 1 and 2)*	1:30
	5, 8	SR, CR, and TE Items (Sessions 1 and 2)*	1:30
		Classroom Activity**	:30
		Performance Task (PT) (Parts 1 and 2)**	2:00
Mathematics	3, 4	SR, CR, and TE Items (Sessions 1 and 2)*	1:30
	5	SR, CR, and TE Items (Sessions 1 and 2)*	1:30
		Classroom Activity***	:30
		Performance Task (PT)	1:00
	6, 7	SR, CR, and TE Items (Sessions 1 and 2)*	2:00
	8	SR, CR, and TE Items (Sessions 1 and 2)*	2:00
		Classroom Activity***	:30
		Performance Task (PT)	1:00
Science	5, 8	Session 1	:45–:55
		Session 2	:20–:25
		Session 3	:45–:65

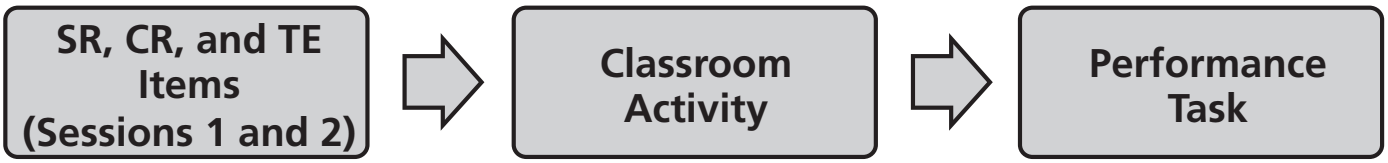
**Testing time for each session is approximately half of the testing time shown (e.g., ELA Sessions 1 and 2 are approximately 45 minutes each).*

***Testing time for each part is approximately one hour.*

****Classroom Activities are administered prior to the performance task and are designed to fit into a thirty-minute window; however, the time within the window will vary due to the complexity of the topic and individual student needs.*

Recommended Order of Test Administration (ELA and Mathematics Grades 5 and 8 Only)

For grades 5 and 8, the ELA and Mathematics Assessments are comprised of two components (tests): SR, CR, and TE items and a PT. All PTs must be preceded by the administration of a Classroom Activity. It is recommended that students take the two components on separate days. It is also recommended that students begin with the SR, CR, and TE items, followed by the Classroom Activity, and then the PT. Districts/Schools may opt to administer in a different order if needed; however, the Classroom Activity, which is designed to introduce the PT, must occur prior to the PT.



Classroom Activity (ELA and Mathematics Grades 5 and 8 Only)

The purpose of the Classroom Activities is to introduce students to the context of a performance task so they are not disadvantaged in demonstrating the skills the task intends to assess. Classroom Activities do not address content information; instead, they focus on vocabulary and key contextual topics. The Classroom Activity is designed to be an introduction and not an assessment.

Guidelines for administering the Classroom Activity for ELA or Mathematics are as follows:

- Classroom Activities should be administered by a teacher. It is preferable—but not essential—that the teacher or TE administering the Classroom Activity has content knowledge in the area of assessment.
- The teacher/TE should be able to record information—including any tables, graphics, formulas, or other information contained in the Classroom Activity materials—for students to see, such as on a chalkboard or dry-erase board. Computers, projectors, and other technology are allowed, but not required, for the Classroom Activity. Recorded information should not be available when students participate in the PT. When the PT is being administered, content from the Classroom Activity should not be available (i.e., do not put any content from the Classroom Activity on the board, in handouts, etc.).
- Students may take notes during the Classroom Activity, but the notes may not be used during the PT. Notes must be collected before proceeding to the PT and stored in a secure location until securely shredded.
- There should be no more than a three-day lapse between the Classroom Activity and the PT administration. Inadvertently administering the PT before or without the Classroom Activity constitutes a testing irregularity.
- Classroom Activities should only be administered to students once and are designed to be completed in approximately thirty minutes.
- The Classroom Activity should not be supplemented with any other content that the administrator may think is helpful. Supplementing the Classroom Activity may detract from the intended purpose of the Classroom Activity.

The Classroom Activity should not be supplemented with any other content.

All students in the same grade within a school will receive the same Classroom Activity with one exception. Students taking the Large Print, Braille, or paper-and-pencil edition may be assigned a different Classroom Activity, in which case, scheduling accommodations will have to be made to provide those students the correct Classroom Activity.

- Consider the accommodations that should be provided to students in the class that would normally be provided during instruction. See "Classroom Activity" in Section 5.1 of this manual for information on student resources that can be provided during the Classroom Activity.
- In the event a student is absent during the Classroom Activity, a make-up session must be scheduled. The Classroom Activity may be recorded; however, the make-up session should provide students with an experience similar to that of his or her peers. To the greatest extent possible, the make-up session should provide students an opportunity to interact with the teacher or TE and his or her peers.

Classroom Activities are located on the **Documents** page of eDIRECT, https://mo.drcedirect.com. The Classroom Activity for students taking a paper-based assessment may differ from the Classroom Activity for students taking the online assessment.

1. From the eDIRECT homepage, log in using your eDIRECT credentials.
2. In the left navigation pane, under **General Information**, select **Documents**.
3. In the main page on the **Documents** tab,
   a. Choose "Summative Grade-Level Assessments Spring 2015" from the Administration drop-down.
   b. Choose "Classroom Activities" from the Document Type drop-down.
   c. Click "Show Documents." A list of all available Classroom Activities will appear in the grid.

Duration and Timing Information

The scheduling/rules for each assessment are included in Tables 1, 2, and 3. Note that the duration, timing, break/pause rules, and session recommendations vary for each content area and component.

Table 1: Assessment Sequence—ELA

ELA	SR, CR, and TE Items (all grades)	Classroom Activity (grades 5 and 8)	Performance Task (PT) (grades 5 and 8)
Number and Duration of Sessions	The SR, CR, and TE items are presented in two sessions. **Recommendations:** • Administer in two sessions corresponding to Sessions 1 and 2. Administer in no more than six sessions (rare/ extreme). • Session durations range from 40–60 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 30 minutes. • Should occur one to three days prior to PT. • Should NOT occur on the same day as the ELA PT.	The PT is presented in two parts (sessions). **Recommendations:** • Administer in two sessions corresponding to Parts 1 and 2 of the PT. • Session durations range from 60–120 minutes.
Breaks Within Sessions	Breaks can be provided during the test sessions using the software's pause feature. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens.	NA	A student can take breaks during Parts 1 and 2. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens within the same session.
Total Duration	• **Recommendation:** Student completes this component within five days of starting.	NA	• **Recommendation:** Students complete Part 1 in one test session and Part 2 the next school day.

Table 2: Assessment Sequence—Mathematics

Mathematics	SR, CR, and TE Items (all grades)	Classroom Activity (grades 5 and 8)	Performance Task (PT) (grades 5 and 8)
Number and Duration of Sessions	The SR, CR, and TE items are presented in two sessions. **Recommendations:** • Administer in two sessions corresponding to Sessions 1 and 2. • Session durations range from 40–60 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 30 minutes. • Should occur as close to the PT as is feasible, and no more than three days prior to the PT. • MAY occur on the same day as the PT.	**Recommendations:** • Administer in one session. • Session duration ranges from 40–120 minutes.
Breaks Within Sessions	Breaks can be provided during the test sessions using the software's pause feature. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens within the same segment.	NA	A student can take breaks during the PT test sessions. Mathematics PT items are presented on a single screen. Following a break, the student will have access to the same items.
Total Duration	• **Recommendation:** Student completes this component within five days of starting it.	NA	• **Recommendation:** Student completes the PT in one day.

Table 3: Assessment Sequence—Science

Science	Sessions 1	Session 2	Session 3 (Performance Event)
Number and Duration of Sessions	• Administer in one session. • Session duration ranges from 45–55 minutes.	• Administer in one session. • Session duration ranges from 20–25 minutes.	• Administer in one session. • Session duration ranges from 45–65 minutes.
Breaks Within Sessions	Breaks can be provided during the test session using the software's pause feature. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens.	Breaks can be provided during the test session using the software's pause feature. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens.	Breaks can be provided during the test session using the software's pause feature. If the test is paused for more than 20 minutes, the student will be able to go back to items on the previous screens.
Total Duration	• **Recommendation:** Student completes Session 1 in a single session.	• **Recommendation:** Student completes Session 2 in a single session.	• **Recommendation:** Student completes Session 3 in a single session.

Additional Administration Recommendations

- For the performance tasks, students may be best served by sequential, uninterrupted time that may exceed the time allotted in a student's schedule.
- Minimize the amount of time between beginning and completing each test within a content area.

Important reminders:

- The test can be spread out over multiple days as needed.
- The Classroom Activity must be completed prior to administration of the PT. Administering the PT before the Classroom Activity is considered a testing irregularity.

1.7 Accommodations and Special Populations

Updated Accommodations Procedures/Codes

The accommodations for the MAP Grade-Level Assessments have changed starting with the Spring 2015 Grade-Level administration. What we previously knew as accommodations has now been split into three areas: Universal Tools, Designated Supports, and Accommodations.

- Universal Tools are available to all students taking a Grade-Level Assessment.
- Designated Supports are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in a student's IEP/504 Plan.

For Special Education students, the IEP team should choose **all** of the designated supports and accommodations that a student will receive.

Some designated supports and accommodations are only for ELL students.

Prior to testing, Test Examiners should log in to eDIRECT to check and set accommodations for students from the Edit Student window. See the eDIRECT User Guide for detailed instructions. It is recommended that districts keep local documentation of Designated Supports.

Table 4: Universal Tools

- The following is a list of universal tools for the Grade-Level Assessments.
- These tools are available to all students.

Tool	Format	Description
Break (Pause)	Online	The system allows all students to pause the assessment for up to 20 minutes. There is no limit on the amount of times a student may use this tool.
	Any	All students may take breaks of up 20 minutes as needed.
Calculator (For calculator-allowed items only)	Online	The system allows all students, on items where calculator use is allowed, to have access to an embedded digital calculator.
	Any	All students may have access, on items where calculator use is allowed, to a physical calculator.
English Dictionary	Online	The system allows all students access to an embedded English dictionary for use on the writing performance task.
	Any	All students may have access to a physical English dictionary for use on the writing performance task.
Glossary (Grades 3–8 Math and ELA only)	Online	The system allows all students to access an embedded glossary, which shows grade- and context-appropriate definitions of specific construct-irrelevant terms. This tool is not available for grades 5 and 8 Science assessments.
Highlighter	Online	The system allows all students to have access to a highlighter for marking desired text, questions, and answers.
	Any	All students may have access to a physical highlighter.

Table 4: Universal Tools continued

Tool	Format	Description
Keyboard Navigation	Online	The system allows all students to navigate through the text by using the keyboard.
Mark for Review	Online	The system allows all students to mark an item for review.
Notepad (Scratch paper)	Online	The system allows all students to use a digital notepad (called "Sticky Notes") to make notes about an item.
	Paper	All students may have access to physical scratch paper to make notes about an item. Physical scratch paper should be collected and destroyed immediately upon the conclusion of the testing session, except during the ELA and Mathematics performance tasks.
Protractor	Online	The system allows all students to use an embedded protractor on specific items where appropriate.
	Paper	All students may have access to a physical protractor for use on specific items where appropriate.
Ruler	Online	The system allows all students to use an embedded ruler on specific items where appropriate.
	Paper	All students may have access to a physical ruler for use on specific items where appropriate.
Spell Check	Online	The system allows students to use an embedded spell check feature on specific items where appropriate. **NOTE:** This feature must be manually turned on to be activated in the system.
Strikethrough (Called "Cross Off")	Online	The system allows all students to cross out answer options.
Thesaurus	Any	All students may have access to a physical thesaurus during the writing performance task.
Writing Tools	Online	The system allows all students to use selected writing tools on specific items where appropriate. The tools include the ability to bold text, italicize text, create bullets points. There is also an undo/redo feature.
Zoom (Called "Magnifier")	Online	The system allows all students to zoom in or zoom out on text or graphics to make them appear larger or smaller than the default size.
	Paper	All students may have access to devices that allow them to change the size of text, formulas, tables, graphics, etc.

Table 5: Designated Supports

- The following is a list of designated supports for the Grade-Level Assessments.
- These supports are available to students when deemed appropriate by a team of educators.
- These supports are available to ELL students.

Support	Format	Description	Code
Bilingual Dictionary	Any	ELL students may have access to a physical bilingual dictionary for use on the writing performance task.	S431
Color Contrast	Online	The system allows students to adjust background or font color based on student needs or preferences.	S101
	Paper	Students may have the test presented to them printed in different colors based on student needs or preferences.	S102
Color Overlay	Paper	Students may have a color transparency placed over the test presented to them based on student needs or preferences.	S103
Glossary (Grades 3–8 Math and ELA only)	Paper	All students taking the paper-based, Braille, or Large Print Assessment may have access to a specific glossary, to be included with the assessment. This support is not available for grades 5 and 8 Science assessments.	S104
Magnification	Online—Not Embedded	The system allows students to use assistive technology devices to change the size of text, formulas, tables, graphics, etc., beyond the capabilities of the zoom tool.	S105
Masking	Online	The system allows students to block off content that is not of immediate need or that may be distracting by using an embedded masking tool.	S106
	Paper	Students may use a masking tool to block off content that is not of immediate need or that may be distracting.	S107

Table 5: Designated Supports continued

Support	Format	Description	Code
Read-Aloud (For all items in any subject, excluding ELA reading passages)	Online	The system allows items in mathematics and English language arts to be read aloud to the student via embedded text-to-speech technology. The student can control the speed and volume of the voice.	S041
	Online—Not Embedded	Students may use assistive technology text-to-speech software to allow all items in any subject, not including ELA reading passages, to be read aloud.	S042
	Any	Students may have items in mathematics, science, and English language arts read aloud to them by a trained reader. Reading aloud of ELA reading passages requires an IEP or 504 Plan.	S043
	Any	ELL students may have items in mathematics, science, and English language arts read aloud to them in their native language by a trained translator. Reading aloud of ELA reading passages requires an IEP or 504 Plan.	S111
Scribe (For all items in any subject, excluding ELA writing)	Any	Students may dictate their responses to a trained scribe, who must follow the administration guidelines. Scribing of ELA writing requires an IEP or 504 Plan.	S351
Separate Setting	Any	Students may be allowed to test in a separate setting from other students. This includes testing individually or testing as part of a smaller group.	S501
Translation	Online	The system allows ELL students to use stacked Spanish translations on selected construct-irrelevant math items.	S108
	Any	ELL students may have test directions for math, science, and social studies translated. ELL students may respond to any assessment in their native language. The responses must be translated and then transcribed by a trained scribe, who must follow the administration guidelines. ELL students taking the paper-based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment. This glossary can be translated locally.	S109

Table 6: Accommodations for Students with Disabilities

- The following is a list of accommodations for the Grade-Level Assessments.
- The accommodations must appear in an IEP or a 504 Plan to be allowed.
- These supports are available to ELL students.

Accommodation	Format	Description	Code
Abacus	Any	Students may have access to an abacus.	A391
Alternate Response Options	Any	Students may respond to items using an alternate option, including, but not limited to: Adapted Keyboards, StickyKeys, MouseKeys, FilterKeys, Adapted Mouse, Touch Screen, Head Wand, Switches.	A441
American Sign Language (ASL) (For math and science items and ELA listening items)	Online	The system allows students to access math items and ELA listening items by viewing ASL video.	A051
	Any	Students may have math, science, social studies items and ELA listening items translated into ASL.	A052
Braille	Paper	Students with visual impairments may access the assessment via a Braille version. Tactile overlays and graphics tools may be used to assist the student in accessing the content.	A012
INVALIDATION Calculator GRADE 3 ONLY (For non-calculator-allowed items only) *INVALIDATION*	Any	All students in grade 3 may have access, on items where calculator use is not allowed, to a physical calculator. **NOTE:** Use of this will result in invalidation—Student will receive lowest obtainable scale score (LOSS).	A392
Calculator GRADES 4–8 ONLY (For non-calculator-allowed items only)	Any	All students in grades 4–8 may have access, on items where calculator use is not allowed, to a physical calculator.	A393
Large Print	Paper	Students with visual impairments may access the assessment via a Large Print version.	A021
INVALIDATION Multiplication Table GRADE 3 ONLY *INVALIDATION*	Any	Students in grade 3 may have access to a single-digit multiplication table. **NOTE:** Use of this will result in invalidation—Student will receive lowest obtainable scale score (LOSS).	A394

Table 6: Accommodations for Students with Disabilities continued

Accommodation	Format	Description	Code
Multiplication Table GRADES 4–8	Any	Students in grades 4–8 may have access to a single-digit multiplication table.	A395
Paper-Based Assessment	Paper	Students may have access to a paper-based version of the assessment.	A102
INVALIDATION Read-Aloud GRADES 3–5 ONLY (ELA reading passages) *INVALIDATION*	Any	Students in grades 3–5 may have English language arts reading passages read aloud to them by a trained reader. **NOTE:** Use of this will result in invalidation—Student will receive lowest obtainable scale score (LOSS).	A041
	Online—Not Embedded	Students in grades 3–5 may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud. **NOTE:** Use of this will result in invalidation—Student will receive lowest obtainable scale score (LOSS).	A042
	Any	ELL students in grades 3–5 may have English language arts reading passages read aloud to them in their native language by a trained translator. **NOTE:** Use of this will result in invalidation—Student will receive lowest obtainable scale score (LOSS).	A111
Read-Aloud GRADES 6–8 ONLY (ELA reading passages)	Online—Not Embedded	Students may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud.	A044
	Any	Students may have English language arts reading passages read aloud to them by a trained reader.	A045
	Any	ELL students may have English language arts reading passages read aloud to them in their native language by a trained translator.	A112
Read-Aloud (ELA reading passages)	Paper	Blind students in any grade who do not yet have adequate Braille skills may have ELA reading passages read aloud.	A046
Scribe (For ELA writing)	Any	Students may dictate their responses to a trained scribe, who must follow the administration guidelines.	A351
Specialized Calculator (For calculator-allowed items only)	Any	Students may have access, on items where calculator use is allowed, to a specialized calculator, including talking calculators or Braille calculators, when appropriate.	A396

Inclusion of Special Populations

All students, including, but not limited to, the following groups of students, must participate in the required MAP Grade-Level Assessments.

- **Missouri Virtual Instruction Program (MoVIP):** Missouri students enrolled in MoVIP are required to participate in the MAP Grade-Level Assessments. For further inquiries regarding MoVIP participation, contact the MoVIP Section at 573-751-2453.
- **Homebound Students:** Homebound students must be tested, either at home or at the school, at the discretion of the district. If the student can come to the school, the student may take the test online. If the student cannot come to the school, the student may take the test online using a district device. If, for any reason, the student cannot take the test online, then the student may take a paper-and-pencil edition of the test. (See instructions in the Large Print, Braille, and Paper-and-Pencil Editions section of this manual.) Test Examiners of homebound students should receive training in the administration of the MAP Grade-Level Assessments. Test Examiners are responsible for ensuring the security of the tests and transcribing student responses into INSIGHT for paper-and-pencil tests.
- **IEP Students:** Students with disabilities, as classified under the Individuals with Disabilities Education Act (IDEA), have an Individualized Education Program (IEP). All decisions regarding a student's participation in the MAP Grade-Level Assessments are made by the student's IEP team and documented in the IEP. All students, including those students with an IEP, must take the MAP Grade-Level Assessments that are required for accountability purposes. For more information about the MAP-A, including eligibility criteria, see http://dese.mo.gov/college-career-readiness/assessment/map-a. The IEP team has the responsibility and authority to determine designated supports and accommodations needed to ensure accessibility to the MAP Grade-Level Assessments.
- **IAP/504 Students:** Students with an Individual Accommodation Program (IAP) are considered disabled under Section 504 of the 1973 Rehabilitation Act. These students are not served under IDEA and are not documented with a particular designation for the MAP Grade-Level Assessments. However, professionals knowledgeable about IAP students' disabilities and their educational needs will make decisions about designated supports and accommodations for these students as they would with IEP students. All IAP/504 accommodations should be marked in the same manner as the IEP student accommodations.
- **English Language Learner (ELL) Students:** Students who have been in the United States 12 cumulative months or fewer at the time of the test administration may be exempt from the English Language Arts Assessment. ELL students must participate in all other required assessments regardless of the length of time they have been in the United States.

Further Information on Special Populations

For further questions regarding special populations, contact the DESE Assessment Section at 573-751-3545 or the Special Education Section at 573-751-5739. Accommodation code definitions can be found in this section of the *Test Administration Manual*.

Optional Populations

The following student groups MAY participate in MAP Grade-Level Assessments:

- **Foreign Exchange Students:** Foreign exchange students are allowed, but not required, to take the MAP Grade-Level Assessments at the discretion of the district.
- **Homeschooled Students:** Homeschooled students may take part in the MAP Grade-Level Assessments at the discretion of the district. Homeschooled students participating in the MAP Grade-Level Assessments will take the assessment(s) online at the local school with district-approved procedures in place during the school's testing window. When a homeschooled student is entered into eDIRECT, the "Homeschool" box on the Testing Codes screen must be checked. The MOSIS ID field should be left blank. Individual Student Reports containing the homeschooled student's assessment scores will be created and posted to eDIRECT. District Test Coordinators must collect contact information from the parents of homeschooled students so that DTCs can notify the parents when reports become available.
- **Private School Students:** Private school students may also participate in the MAP Grade-Level Assessments. A representative from the private school must contact the MAP Service Line at 1-800-544-9868. Private schools must uphold the same standardized administration procedures and security measures that Missouri public schools uphold.

Special Circumstances

Some students may require special arrangements for testing. Please refer to the following guidelines for students requiring a change in test setting, test format, or test administration.

- **Designated Supports and Accommodations:** Prior to testing, be sure to consider any additional planning that may be required to administer the test using students' designated supports and/or accommodations. Designated supports/accommodations that require particular attention include, but are not necessarily limited to:
    - **Use of a Translator:** District staff may read Mathematics and Science Assessments and English items to students in their native language. Read aloud of English reading passages in a student's native language is allowed only if specified in a student's IEP or 504 Plan. For all assessments, ELL students may give their responses orally or in writing in their native language. Their responses must be translated into English and transcribed into INSIGHT.

      Refer to Tables 5 and 6 in this section for the appropriate support/accommodation codes to use when a test is being translated. The translation and transcription must be an accurate interpretation of the student's responses.

      Translators must be trained in administering the Grade-Level Assessments. Translators for students taking the online assessments will not have the opportunity to read and review the test before test administration. If needed, translators for students taking the Large Print, Braille, or paper-and-pencil edition of the assessments may have access to printed student test books in a secure environment to read and review before the test administration. Please see Section 5.0 for instructions regarding administering the Large Print, Braille, and paper-and-pencil editions of the tests.

Additional guidelines for use of a scribe are located on the DESE website at http://dese.mo.gov/sites/default/files/asmt-scribing-guidelines.pdf. Refer to Tables 5 and 6 in this section for appropriate support/accommodation codes for scribing.

- **Use of a Scribe:** Students with an Individualized Education Program (IEP) or Section 504 Plan must have a scribing accommodation specified within the plan if it is to be used for writing items (short text and full-write essay item types) for the English Language Arts portion of Missouri's Grade-Level Assessments. Scribing is considered to be a designated support for all other content areas and item types.

  Scribes may be teachers, teacher aides, teacher assistants, or other school personnel who are appropriately trained and qualified. Translators for ELL students may also act as scribes. Parents, school volunteers, peer tutors, and other students may NOT act as scribes on Missouri's Grade-Level Assessments.

- **Paper-and-Pencil Test Accommodation:** See the Large Print, Braille, and Paper-and-Pencil Forms section in this manual for instructions concerning the paper-and-pencil accommodation procedures.

- **Large Print and Braille:** See the Large Print, Braille, and Paper-and-Pencil Forms section in this manual for instructions concerning Large Print and Braille procedures.

- **Students Testing Out of District:** Students receiving services in off-site placements (other districts, private agencies, correctional facilities, etc.) must be tested. They may be tested in those placements if necessary, or they may come to the school of residence if possible. The DTC from the district where the student resides must make arrangements for the student to test in the serving district/agency.

  Out-of-district students may take the online or the paper-and-pencil edition of the MAP Grade-Level Assessment. If the student takes the paper-and-pencil edition, his or her responses must be transcribed into INSIGHT. The DTC from the district of residence has several responsibilities in this process.

The DTC must contact the off-site district/agency prior to the first day of the district of residence's testing window to make arrangements:

- If the student is testing online at a school, arrange for the the student's Test Tickets to be available through eDIRECT.
- If the student is testing online at an off-site location, arrange for the student to take the test using a district device.
- If the student is taking a paper-and-pencil edition of the assessment, follow the administration instructions in Section 5.0 of the manual.

DTCs have access to eDIRECT before the INSIGHT site opens and may securely download and securely print the MAP Grade-Level Assessments as appropriate. Print copies of the assessments will have a barcode. Barcoded printed assessments must be returned to CTB after the tests have been transcribed into the test delivery system (INSIGHT).
NOTE: Paper-and-pencil tests cannot be printed until students are available in the system on March 5th.

1.8 Tutorials and Practice Tests

Tutorials

The Tutorials provide step-by-step video instructions on how to navigate the online system and give detailed explanations about the key features of the software. **The Tutorials should be reviewed at least once by Test Examiners who will supervise any of the MAP Grade-Level Assessments and by students in advance of their first test day. Allow students to repeat the Tutorials as often as desired and needed.**

Students should review the Tutorial before completing the Online Tools Training (OTT). It has been proven beneficial for schools to schedule a Tutorial session for students immediately before at least one OTT session.

If computer lab availability is limited, the Tutorials may be presented to school personnel and students in a classroom using an LCD projector and a single Internet connection.

The Tutorials can be accessed via the Online Tutorials desktop icon once the testing software has been installed. The Tutorials may also be accessed through eDIRECT.

Instructions for Accessing the Tutorials through eDIRECT

1. Navigate to eDIRECT, https://mo.drcedirect.com. (Login is not required.)
2. Under **Test Setup** select **General Information.**
3. Select the **Test Tutorials** tab.
4. Select the **Play Tutorial** action button.
5. Select **Play All** or choose from different sections within the tutorial.

The Tutorials walk students through the software and tools that are available. In the Tutorial, the student can move forward as directed or jump around if desired. A menu at the left of the page allows the student to select specific sections for review.

Online Tools Training

In preparation for the test and to expose students to the various item-response types in each content area (see Appendix A for item types), it is highly recommended that all students access the Online Tools Training (OTT) for each content area. Each OTT is designed to provide students and educators with an opportunity to quickly familiarize themselves with the software and navigational tools that they will use on the MAP Grade-Level Assessments.

The OTT for each content area includes a variety of item response types. The OTTs also include a comprehensive reflection of embedded universal tools, designated supports, and accommodations. The OTTs should also be provided to students with any non-embedded universal tools, designated supports, and accommodations as allowed on the operational assessments.

The OTTs can be accessed via the INSIGHT desktop icon once the testing software has been installed. Nonaccommodated versions of the OTTs can be publicly accessed using the Google Chrome browser at https://wbte.drcedirect.com/MO/portals/mo.

List of INSIGHT Keyboard Shortcuts and Icons

During online testing, all students may have access to a printed list of the keyboard shortcuts and icons available in INSIGHT. The list may be printed from Appendix E or may be accessed on the **Documents** page of eDIRECT, https://mo.drcedirect.com.

Science Practice Items

Additional practice items for Science include constructed-response items and a performance event for grades 5 and 8. They allow students to practice with the item types and the functionality of the testing environment that they will experience during summative testing. The Science practice items can be accessed via the INSIGHT desktop icon once the testing software has been installed. Scoring materials for the practice items are available on the **Documents** page of eDIRECT (login required). After a practice test is closed, student responses are no longer available in the online system.

2.0 BEFORE ONLINE TESTING

2.1 Advance Announcements and Preparation

Parents and guardians should be informed of the district MAP Grade-Level Assessment schedule so they can help ensure their students are present on the day of testing (without scheduled appointments or vacation days during the testing window) and prepared with the proper materials that may not be provided by the district.

In addition to completing the applicable content for the grade level, students should have experience using the specific device on which they will be taking the assessments. Students taking the assessments on a desktop or laptop computer should know how to use a mouse and keyboard. Instead of a mouse, students may use the embedded touchpad in the keyboard of a laptop. Students taking the assessments on iPads should know how to use a touchscreen (and/or stylus, if applicable). Touch interfaces are not supported on other devices for Spring 2015. It is strongly recommended, but not required, that students taking the assessments on tablet devices have access to (and know how to use) an external keyboard. Students should review the INSIGHT Online Tools Training (OTT) for the MAP Grade-Level Assessment they will be taking. OTTs are for Test Examiners and students to become familiar with the format and functionality of the online test. The OTTs provide a preview of the item types included in the MAP Grade-Level Assessments. Item types are listed and described in Appendix A.

MAP Grade-Level Assessments are available on the following devices:

Desktop Computers

Laptops

Netbooks

Chromebooks

iPads

Students should be familiar with the device on which they will be taking the assessment prior to testing.

2.2 User Roles

The District Test Coordinator (DTC) is responsible for training all School Test Coordinators (STCs) on testing procedures. If a district does not have STCs, the DTC performs the role of the STC. While the training of Test Examiners may be delegated to each building's STC, the DTC is responsible for ensuring that all Test Examiners are well-prepared and trained. Training includes special education teachers, proctors, translators, and Test Examiners who are administering the MAP Grade-Level Assessments to homebound or out-of-district students.

District Test Coordinator Responsibilities

All DTCs are responsible for the following:

- Attend all trainings provided by DESE and CTB.
- Stay abreast of all communication regarding the MAP Grade-Level Assessments.

DTCs must train all STCs, Test Examiners, and other responsible district and/or school staff.

The DTC is responsible for updating the district's testing schedule if it changes.

- Train all STCs, Test Examiners, and other responsible district and/or school staff.
- Maintain the district's testing schedule and be prepared to provide it to DESE upon request. If the district's testing schedule changes in any way, the DTC is responsible for updating this information in eDIRECT until February 20, 2015, and by contacting the MAP Service Line after that time. Inform district staff of the testing schedule so that distractions such as PA announcements, lawn maintenance, or fire drills are avoided at the time of test administration.
- Update student demographic information to correct any errors and ensure these corrections are also made in the local student information systems and MOSIS. See Appendix B in this manual for instructions on how to handle student transfers.

CTB's dedicated MAP Service Line
1-800-544-9868
7:30 A.M. to 6:30 P.M. Central Time, Monday–Friday

- Communicate with CTB and DESE on behalf of the district. The STC should contact the DTC if help is needed. If the DTC is unable to answer a question, he or she will contact CTB's dedicated MAP Service Line.
- Ensure the DTC's email account allows receipt of all communication from DESE's, CTB's, and DRC's email domains (@dese.mo.gov, @ctb.com, and @datarecognitioncorp.com).
- Verify with the STCs that INSIGHT has been installed and certified on all applicable workstations for the current statewide window.

Any Test Examiner who needs to set or check accommodations will need an eDIRECT account. Other Test Examiners do not need an eDIRECT account, as logging into eDIRECT is not required to start a test.

- After verifying each building's security, ensure that STCs have access to eDIRECT and secure test administration materials.
- Enter Test Examiners into eDIRECT in order to generate their eDIRECT logins (for Test Examiners needing an eDIRECT login).
- Ensure test security is maintained by restricting Test Examiner access to the MAP Grade-Level Assessments and other secure testing materials before and after testing.
- Transcribe Large Print, Braille, and paper-and-pencil edition responses into INSIGHT (in districts where this role is not assigned to the Test Examiner).

School Test Coordinator Responsibilities

All STCs are responsible for the following:

- Attend all trainings provided by the DTC, DESE, and CTB.

- Stay abreast of all communication from the DTC regarding the MAP Grade-Level Assessments.
- Train all Test Examiners on MAP Grade-Level Assessment procedures.
- Review the Tutorial and the Online Tools Training (OTT) prior to testing and ensure that Test Examiners and students have an opportunity to review both the Tutorial and OTT prior to testing.
- Work with the District Technology Coordinator (if applicable) to ensure INSIGHT has been installed and certified on all applicable workstations.
- Verify the accuracy of student and Test Examiner information in eDIRECT for the school and update as needed. Confirm that any appropriate student accommodation codes are marked in **Test Setup** in eDIRECT.
- Communicate with the DTC regarding the school's testing schedule prior to testing. If the school's testing schedule changes in any way, the STC is responsible for updating the DTC.
- Ensure that all Test Examiners are knowledgeable about permitted and prohibited materials (see Section 2.5 Assessment Materials for Students/Administrators).
- Provide login information to Test Examiners as soon as possible to allow the Test Examiners adequate time to prepare for administering the tests.
- Ensure that each Test Examiner has the following:
    - eDIRECT login information
    - Student Test Tickets for each test session
    - Classroom Activity materials
    - The appropriate quantity of Large Print and Braille test books or access to paper-and-pencil editions as required per content area
    - Any required ancillary testing materials
- Ensure test security is maintained by restricting Test Examiner access to the MAP Grade-Level Assessments and other secure testing materials before and after testing.
- Validate that testing procedures are followed as written in this *Test Administration Manual*. Printed copies of the manual should be destroyed at the building level after the final district content testing window has closed.

STCs must train all Test Examiners on MAP Grade-Level Assessment procedures.

Test Examiners must ensure that all grade-level testing materials are secure at all times. Although this manual is not considered secure, it contains links to secure test materials.

Both written and verbal discussion of specific MAP Grade-Level Assessment items breach the security and integrity of the test.

NOTE: Students may use their own calculators if the calculators meet the "permitted materials" guidelines (page 32), or the DTC may provide calculators per district practice.

Administrators and Test Examiners are responsible for reporting any intentional or unintentional unethical behavior by students or staff members to district administration and/or to the DESE Assessment Section at 573-751-3545 or assessment@dese.mo.gov.

Test Examiner Responsibilities

All Test Examiners are responsible for the following:

- Ensure **all** grade-level testing materials are secure at all times. **Both written and verbal discussion of specific MAP Grade-Level Assessment items breach the security and integrity of the test**. Discussion between Test Examiners, proctors, translators, or any district staff regarding test items is not permitted.
- Ensure any ancillary testing materials or tools are available or provided, such as:
  - a dictionary and a thesaurus for the full-write essay portion of an ELA performance task
  - scratch and graph paper
  - calculators for the calculator-allowed portions of the mathematics assessments
  - Braille paper (if provided)
- After testing is complete
  - Check that tests have been submitted.
  - Check that tests are closed in the system.
  - Collect the Large Print, Braille, and/or paper-and-pencil materials from the students, and prepare materials for return to the STC.
  - Transcribe Large Print, Braille, and paper-and-pencil edition responses into INSIGHT.
  - Contact the STC for guidance regarding the handling of any contaminated test materials. (See Appendix C in this manual.)
  - Collect all draft, scratch, grid, graph, or Braille paper and return all used materials to the DTC/STC for secure shredding.

2.3 Test Security

Test security and ethical testing practices continue to be of PARAMOUNT importance. A test security policy must be in place for each district and charter school. The test security policy should be placed in the District's Assessment Plan, which is locally board approved annually. The accurate assessment of student achievement is a critical component of the educational process in Missouri. It is the responsibility of everyone involved in the assessment process to understand the security measures in place to avoid any intentional or unintentional unethical behavior by students

or staff members. Administrators and Test Examiners are responsible for reporting any of these behaviors to district administration and/or to the DESE Assessment Section at 573-751-3545 or assessment@dese.mo.gov.

Preparing for computer-based testing includes determining the layout of the physical computer lab, training for the teachers and staff, and preparing the students. Although DESE does not provide specific requirements for a computer lab, the lab must be set up with test security in mind. Workstations must have adequate space between them so that students are not able to view one another's screens.

Instructional materials must be removed or covered, including, but not limited to, information that might assist students in answering questions that is displayed on bulletin boards, chalkboards or dry-erase boards, or on charts (e.g., wall charts that contain literary definitions, maps, mathematics formulas, etc.).

District and School Test Coordinators, Test Examiners, translators, proctors, and any other district and/or staff who have testing responsibilities must follow test security procedures. The tests **must not** be read, scored, reviewed, photocopied, duplicated, scanned, transported by students, or made accessible to personnel not responsible for testing. **Both written and/or verbal discussion of specific MAP Grade-Level Assessment items breach the security and integrity of the test and may result in an invalidation or loss of scores for accountability purposes.**

Translators and transcribers who read student test items and answers must maintain test security at all times. Test items or answers must not be discussed with anyone at any time. When hard-copy editions of the test are not in use, they must be stored in a secure, locked location outside of the classroom. Large Print, Braille, and paper-and-pencil editions of the tests must be transcribed into INSIGHT and shipped back to CTB following the procedures in Section 5.3 in this manual once testing is complete.

Test security and ethics also include standardized training for all District and School Test Coordinators, Test Examiners, translators, proctors, and any district and/or school staff who have responsibilities in testing. Training webinars from DESE and manuals (including this manual) are provided for training purposes at http://dese.mo.gov/college-career-readiness/assessment/grade-level. This *Test Administration Manual* is also available on the **Documents** page of eDIRECT.

Both written and/or verbal discussion of specific MAP Grade-Level Assessment items breach the security and integrity of the test and may result in an invalidation or loss of scores for accountability purposes.

Large Print, Braille, and paper-and-pencil editions of the tests must be transcribed into INSIGHT and shipped back to CTB following the procedures in Section 5.3 in this manual once testing is complete.

Only this *Test Administration Manual* may be reviewed before testing, NOT the secure tests. Only translators may review secure test material prior to test administration.

2.4 eDIRECT and INSIGHT

eDIRECT hosts the Missouri Assessment Portal, https://mo.drcedirect.com.

Two online systems support the MAP Grade-Level Assessments: eDIRECT and INSIGHT.

eDIRECT hosts the Missouri Assessment Portal. Through this system, Missouri educators are able to:

- Review documentation and training for the MAP Grade-Level Assessments.
- Download secure materials for administering the MAP Grade-Level Assessments.
- Download software for administering the MAP Grade-Level Assessments.
- Provide enrollment information, including orders for Large Print and Braille test books.
- View and update student data prior to testing, including indicating any accommodations or designated supports that will be used.
- Place students into test sessions and print Student Test Tickets.

The **Documents** page of eDIRECT contains manuals, trainings, and secure administration materials. Secure materials require login to access, while non-secure materials are publicly available. To access the page, click **Documents** under the General Information menu in the upper left portion of the eDIRECT Home Page. Click the Show Documents button to display the available materials.

Details are provided in the eDIRECT User Guide, which is available on the **Documents** page of eDIRECT.

INSIGHT is the test engine for the MAP Grade-Level Assessments.

INSIGHT is the secure browser-based test engine through which students take the MAP Grade-Level Assessments and that provides students with an engaging test experience. Technology coordinators download the INSIGHT client software to the devices that will be used for testing.

Details are provided in the DRC INSIGHT Technology User Guide, which is available on the **Documents** page of eDIRECT.

2.5 Assessment Materials for Students/ Administrators

This section concerns all materials required, permitted but not provided, or prohibited while taking Grade-Level Online Assessments.

Required Materials

- A workstation with Internet access, a monitor, a mouse, and a keyboard for each student, or a tablet device with Internet access if a student will be testing on a tablet. Devices must have INSIGHT properly loaded and certified.
- Student Test Tickets
- The resources in Tables 7 and 8

Table 7: Additional Required Resources for ELA and Mathematics

Content Area	SR, CR, and TE Items	Classroom Activity	Performance Task (PT)
ELA	• Headphones are required for the listening portion of the ELA assessment for all grade levels and for students requiring text-to-speech. • Scratch paper should be provided for note taking if necessary.	NA	• Headphones are required for some performance tasks and for students requiring text-to-speech. • Scratch paper should be provided for note taking if necessary.
Mathematics	• Headphones are required for students requiring text-to-speech and for students requiring Audio Glossaries. • Scratch paper is required for all grades. • Graph paper is also required for grades 6 and above. • An embedded calculator will be available for some mathematics items in grades 6 and above.	NA	• Headphones are required for students requiring text-to-speech. • Graph paper is also required for grades 6 and above. • Scratch paper is required for all grades. • An embedded calculator will be available for all mathematics PT items in grade 8.

Table 8: Additional Required Resources for Science

Content Area	Sessions 1 and 3	Session 2
Science	• Headphones are required for students requiring text-to-speech. • Graph paper is required. • Scratch paper is required.	• Headphones are required for students requiring text-to-speech. • Scratch paper is required.

Permitted Materials for Accommodations, Universal Tools, and Designated Supports

- Scratch paper and grid/graph paper are allowable for all assessments even if not required.
- An English dictionary and a thesaurus may be available for the full-write essay portion (Part 2) of an ELA performance task. ELL students may use an English, a non-English, and a bilingual dictionary and thesaurus as needed during Part 2 of an ELA performance task.
- A physical calculator can be accessed for calculator-allowed items for the Mathematics assessments.
    - For grade 6 Mathematics assessments, a four-function calculator with square root and percentage functions is permitted. (This type of calculator is permitted for grades 3–5 as an accommodation only, as the assessments include no calculator-allowed items.)
    - For grades 7 and 8 Mathematics assessments, a scientific calculator with exponents, trigonometry, and logarithmic functionalities is permitted.
    - Test Examiners are responsible for ensuring and verifying that any calculator with the ability to store functions and equations, e.g., a scientific calculator, has the memory cleared before and after each mathematics assessment.
    - Calculators cannot have Internet connectivity or be able to connect to anyone inside or outside the classroom during testing.
    - Students cannot use a calculator on a laptop or other portable computer, pocket organizer, cell phone, device with a typewriter-style keyboard, electronic writing pad, or pen-input device unless a particular assistive device is required for a student and is specified on his or her IEP.
    - No calculators with QWERTY keyboards are allowed.

Prohibited Materials

- Electronic devices, including any portable device that can connect to the Internet or to anyone inside or outside of the classroom, must not be accessible during the testing sessions. Such items include, but are not limited to:
    - cellular/mobile phones
    - electronic music players
    - digital cameras
    - handheld scanners
    - portable gaming devices
    - any device that can connect to the Internet
- If students are allowed to enter the testing room with cell phones, the phones **must** be collected prior to testing and returned at the end of the testing session. Students are not allowed to have cell phones in their pockets, purses, or backpacks during testing.

Assessment Materials and Training for Test Examiners

- *Test Administration Manual*
- Test Examiner training provided online by DESE
- Student logins (obtained from the School Test Coordinator)

  NOTE: All materials distributed to the students with usernames and passwords must be collected before the students leave the testing area.
- Extra pencils and a supply of scratch and graph paper
- Classroom Activity materials

3.0 DURING ONLINE TESTING

Use the following information and script to assist students with the login procedures.

The Test Examiner (TE) should verify the security of the testing environment prior to beginning a test session. TEs must ensure that students do not have access to prohibited devices and materials during testing.

To ensure that all students are tested under the same conditions, the TE should adhere strictly to the script for administering the test. These instructions can be found in the boxes in bold on the following pages. When asked, the TE should answer questions raised by students but should never help the class or individual students with specific test items. Except for single words, no test items can be read to any student for any content area, unless specified as an accommodation or designated support.

Please remember that the script must be followed exactly and used each time a test is administered. If the class is resuming a test and the TE is sure that all students are able to log in without hearing the login directions again, the TE may skip the italicized portions of the directions for the login section.

All directions that a TE needs to read to students are indicated by the word "SAY" and are in boxes so they stand out from the regular text. They should be read exactly as they are written, using a natural tone and manner. If the TE makes a mistake in reading a direction, the TE should stop and say, "I made a mistake. Listen again." Then the direction should be reread.

The TE should try to maintain a natural classroom atmosphere during the test administration. Before each test begins, he or she should encourage students to do their best.

RECOMMENDATION: Consider printing this section to be used on the day of testing for each portion of each content area test. Remember that the SR, CR, and TE items component and PT component are considered two unique tests and, as such, adherence to the process that follows is needed when initiating both tests.

The TE should adhere strictly to the script for administering the test.

Except for single words, no test items can be read to any student for any content area, unless specified as an accommodation as listed in the Usability, Accessibility, and Accommodations Guidelines.
The TE may pronounce one word in a sentence for a student upon request.

If students are beginning Part 2 of the ELA performance task, the TE should distribute the students' notes retained from Part 1. If students are resuming the Mathematics performance task, the TE should distribute the students' notes and grid/graph paper retained from the previous testing session of the same component.

Any time a student logs in to the testing system, the TE should follow this script. This includes logging in to complete either session of the SR, CR, and TE items, the PT, or any session of the Science Assessment.

3.1 Specific Administration Information

1. The TE distributes the Student Test Tickets.

You should have received Student Test Tickets for this testing session from your DTC or STC. Before beginning, ensure that you have all of the correct test tickets for the students who will be testing. Note the Test Name and read it aloud where the script states [Test Name].

If students are starting a new session:

☐ **You are about to take (the) [Test Name].**

If students are resuming a session:

☐ **You are about to continue (the) [Test Name].**

I will now hand out a Test Ticket to each of you. When you receive your Test Ticket, check that your name appears on the ticket. If your name does not appear, raise your hand.

Distribute test tickets to each student, ensuring that each student is given the correct ticket with his or her name printed on it. Contact your STC or DTC if a needed ticket is missing.

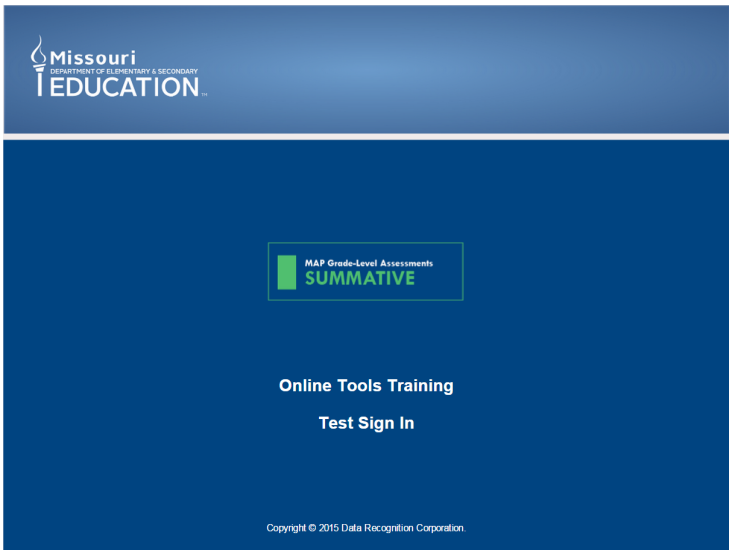
2. The TE directs students to the test sign-in page.

☐ **Now select the "MO Online Assessments" icon that appears on your screen.**

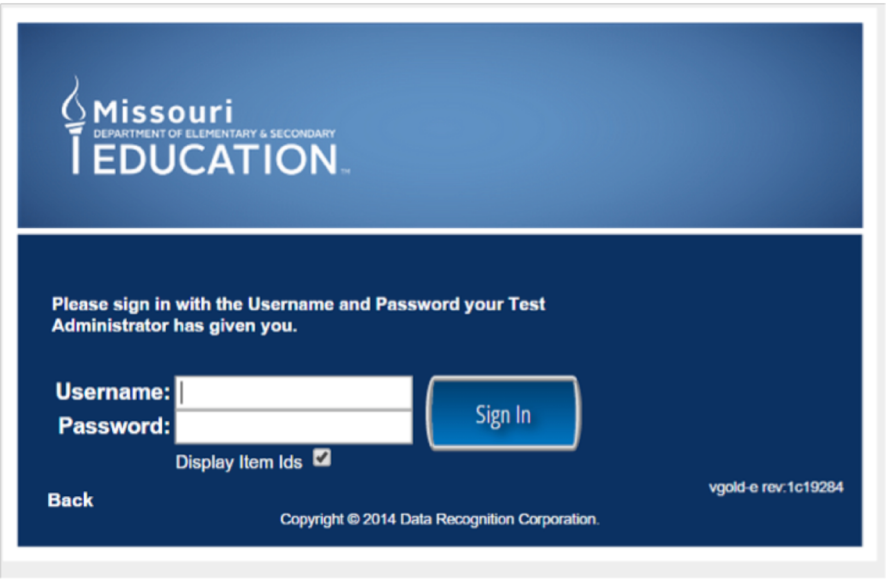
Students using a laptop or desktop workstation should double click on the icon. Students using a Chromebook or iPad should tap on the icon. Help students if they have trouble activating the icon.

3. The TE instructs students to log in.

At the top of your screen you should see "Missouri Department of Elementary & Secondary Education." Below that, you will see links for the Online Tools Training and Test Sign in for the MAP Grade-Level Assessment Summative test. Please select "Test Sign In."



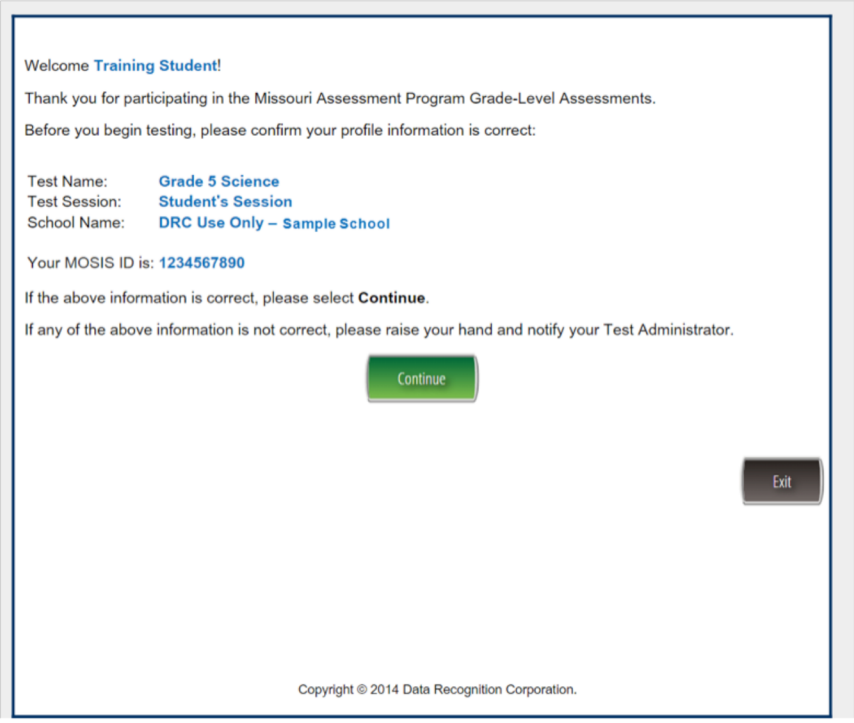
This is the Login screen. Type your username and password from your Test Ticket into the correct boxes on the screen. Then select "Sign In."



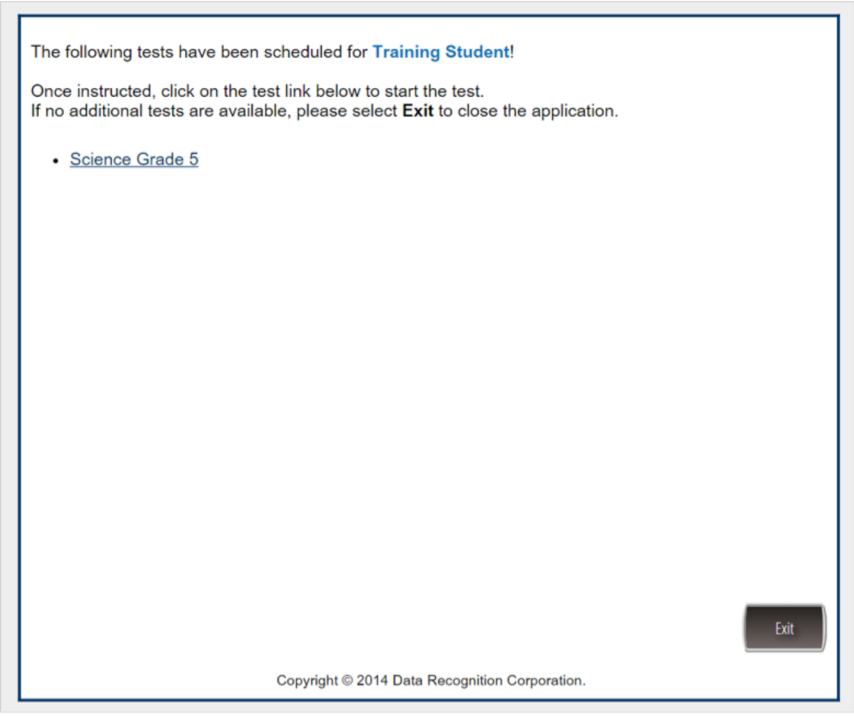
Test Ticket information is unique to each student and each session. Assist students as needed; TEs may have to help students type in this information. After the login, make sure all students are on the correct screen. Wait for all students to reach this page.

This is the Welcome screen. Please check that your name appears at the top of the screen. Check that the test name is [Test Name]. Then check that your school, MOSIS ID, and other information are correct. If everything is correct, select "Continue." If your information is not correct, please raise your hand.

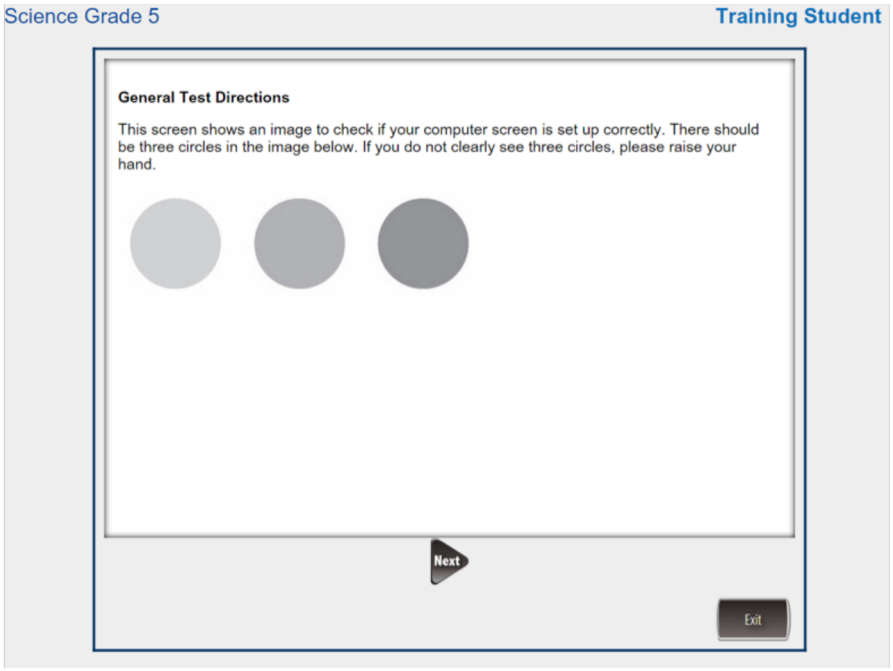
If a student's information is incorrect, the TE should contact the STC and/or the DTC.



You are now on the screen that shows the name of the test you are scheduled to take. If you do not see this, please raise your hand. Please select the test link that is shown.



Select the NEXT arrow to continue.



The following screens contain the test directions for the test you are taking today. Please read the directions carefully. If you have any questions about the directions, raise your hand. You can find the directions during your test by clicking the HELP button in the top right corner.

During the test, you may see a page with no test questions. Follow the directions on the page to continue taking the test.

If you are unsure of an answer, provide what you think is the best answer; there is no penalty for guessing. If you would like to review that answer at a later time, mark the item for review by clicking the FLAG at the bottom of the screen before going on to the next question. Flagging the item will remind you to go back and decide whether or not you want to change the answer.

You may PAUSE at any point in the test by clicking PAUSE after answering an item. The PAUSE button is used to stop the test. Please raise your hand if you need a break and ask me before you click PAUSE. After pausing, a timer will appear on your screen. After your break, click on the RESUME button to continue. If you pause for more than 20 minutes, you will need to log back in.

Students may PAUSE at any point in the test by clicking PAUSE after answering an item. The PAUSE button is used to stop the test. Students must raise their hands if they need a break and ask the TE before clicking PAUSE. After pausing, students must click on the RESUME button to continue. If students pause for more than 20 minutes, they will need to log back in.

Session 1 of a grade 6, 7, or 8 Mathematics SR, CR, and TE items component contains two segments. The first segment contains items that do not allow calculators. The second segment contains calculator-allowed items. Students may have access to physical calculators after they have submitted the first segment.

Your answers need to be your own work. Please keep your eyes on your own test and remember that there should be no talking.

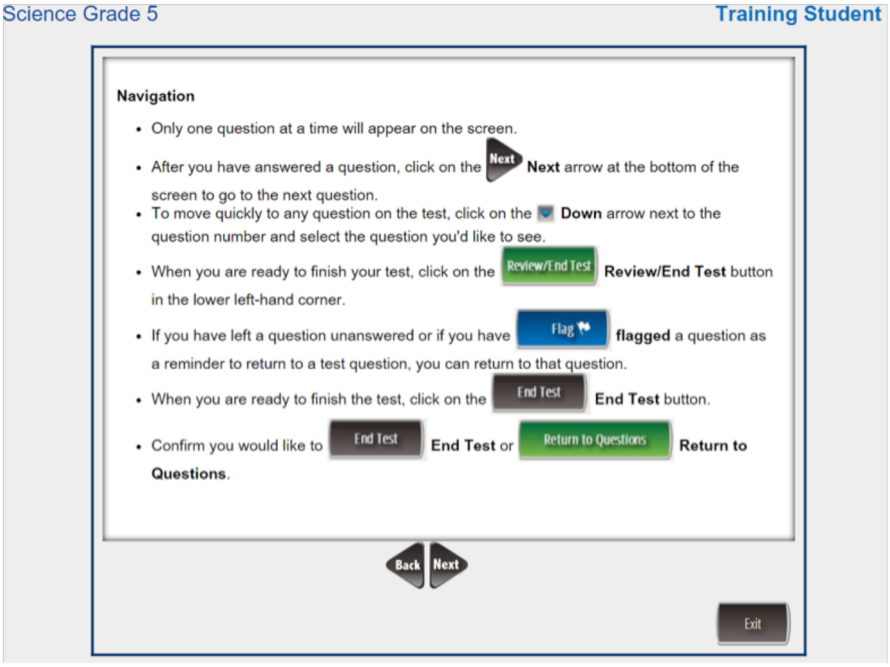
Read aloud the following paragraph if students are taking Session 1 of a grade 6, 7, or 8 Mathematics SR, CR, and TE items component.

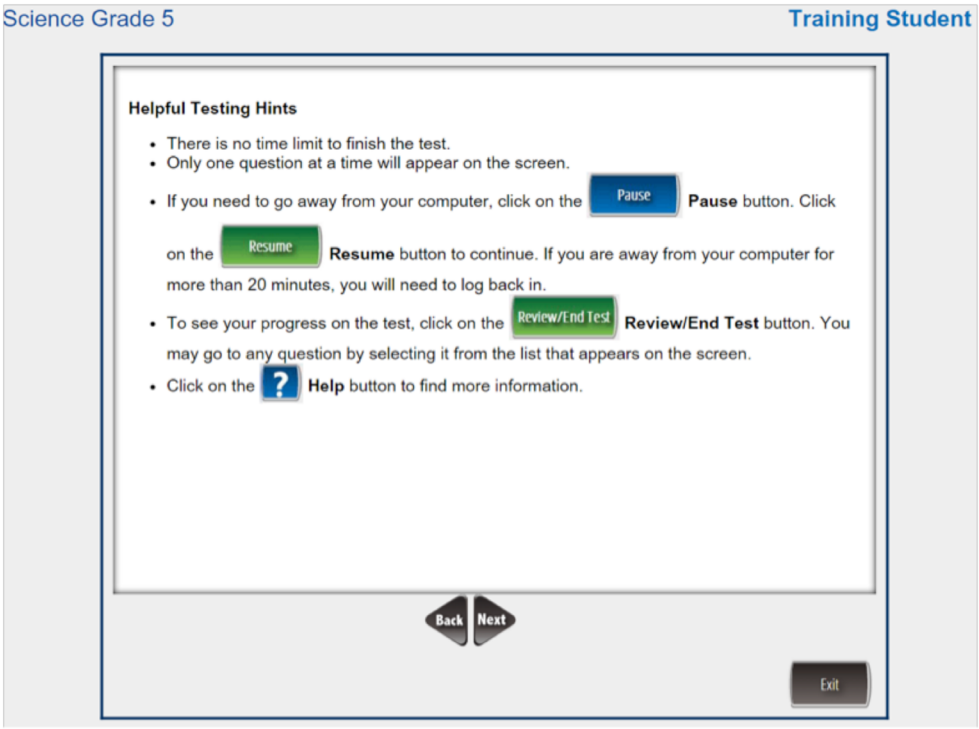
☐ **Please keep in mind that this test is divided into two segments. When you get to the end of a segment, you will be prompted to submit your answers before moving on. Once you submit your answers and move on to the next segment, you will not be able to return to the previous segment. After you have submitted the first segment, you may begin using a calculator.**

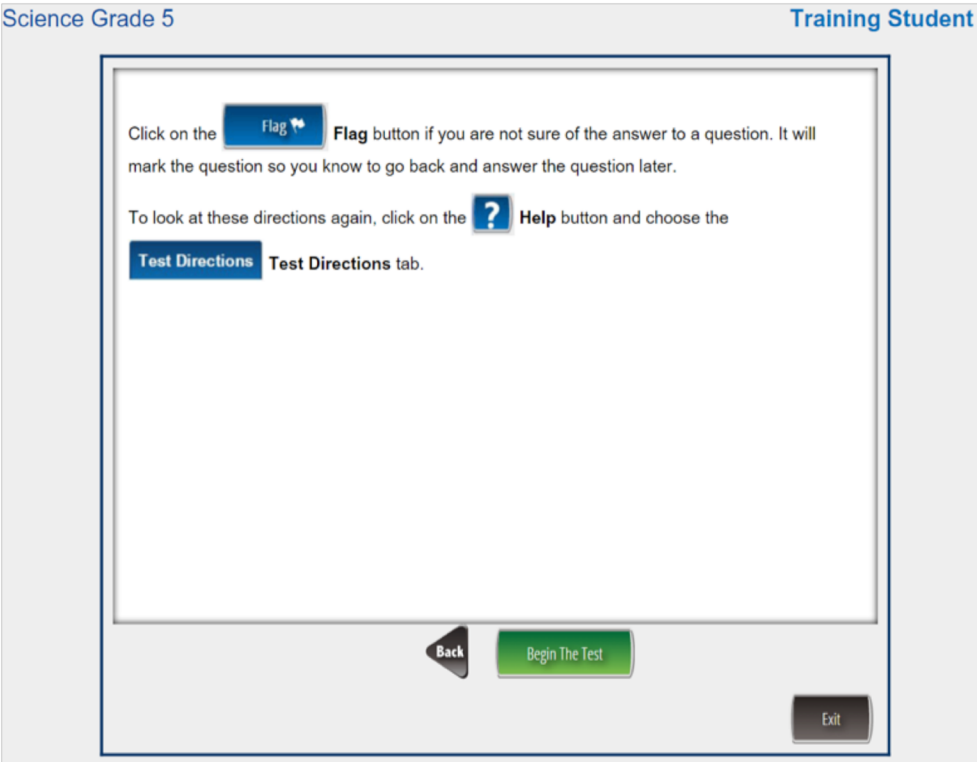
Read aloud the following paragraph if students are taking Part 1 (Session 1) of an ELA performance task.

☐ **Use your scratch paper to take notes you want to keep for Part 2, the essay portion, of this performance task. Any notes you take online using Sticky Notes will not be saved for Part 2.**

☐ **When you are ready to begin your test, click BEGIN THE TEST.**







4. The TE monitors student progress.

Monitoring Test Progress

Once students have started their tests, the TE should circulate through the room to ensure that all conditions of test security are maintained. If the TE witnesses or suspects the possibility of a test security incident, the STC and DTC should be contacted immediately in accordance with the security guidance provided in this manual.

If the TE notices that a student is off task, the TE may say the following statement to the student, verbatim, to keep him or her focused.

☐ **It is important that you do your best. Do you need to pause the test and take a break?**

If a student asks for assistance either in answering an item or manipulating an item type, the TE should let the student know that he or she should try his or her best, but that the TE cannot help answer an item.

☐ **I can't help you with your test. Check the HELP button to read the directions.**

The TE may remind the student to reread the instructions for that item.

5. The TE ends the test session.

When there are approximately ten minutes left in the test session, the TE should give students a brief warning.

If students will continue this portion of the test at a later time, read aloud the following two scripts:

☐ **We are nearing the end of this test session. Please review any completed or marked items now. You will be able to finish the test at another time.**

At the end of the session:

☐ **This test session is now over. Click PAUSE, then click EXIT, and then click YES, EXIT. You will be able to finish at another time. I will now collect any scratch paper or other material.**

If students are completing this portion of the test, read aloud the following two "SAY" scripts:

☐ **We are nearing the end of this test session. Please review any completed or marked items now. Do not submit your test unless you have answered all of the questions.**

After answering the last item in each session, the student will press the Review/End Test button at the bottom left-hand corner of the screen. The student is then presented with a screen prompting him or her to review answers (marked and unmarked) for all items prior to submitting the test. At that point, the student can either click the Return to Questions button to answer previously unanswered items, or can press End Test to submit the test. If a student needs additional testing time, direct him or her to pause the test and then exit so testing can continue at another time.

This test session is now over. When you have finished, click END TEST. I will now collect any scratch paper or other material.

TEs should collect any scratch paper (and graph paper for grades 6 and above).

Testing Over Multiple Sessions or Days

For some tests, students may be best served by sequential, uninterrupted time that may exceed the time in the regular class schedule. It is recommended that the ELA PT be administered in two sessions corresponding with Part 1 and Part 2. Each part requires a separate Test Ticket. Students can be provided breaks within each part; however, once a student begins Part 2, he or she will not be able to review or revise items in Part 1. For this reason, it is recommended that students complete Part 1 in one test session; Part 2 would ideally be delivered the next school day. For the Mathematics PT, it is recommended that it be administered in one test session of 40–120 minutes.

If the TE intends to administer a session over the course of multiple days for a student or group of students, TEs may ask students to pause and exit after they reach a designated point. For most tests, there is nothing built into the system to prevent students from progressing from one section of the test to another. In those cases, the TE should give the students clear directions on when to pause. For example, TEs may designate a certain amount of time for testing. This guidance may be written on a dry-erase board, chalkboard, or another place that students can easily see. Students will receive a notification when they reach the end of a segment within Session 1 of the grades 6–8 Mathematics SR, CR, and TE items component.

3.2 Moving a Student During an Assessment

Occasionally a student must be moved to a new location to continue testing. In order for the student to continue his or her test, complete the following steps:

1. Pause and end the student's online assessment. To do so, select the "Pause" button, then select the "Exit" button, and then select the "Yes, Exit" button. (Once the student exits the test, the workstation becomes immediately available for other use.)
2. Escort the student to the new location.
3. Using the login and password from the student's Test Ticket, log the student in to his or her assessment at the new workstation to complete the assessment.

4.0 AFTER ONLINE TESTING

4.1 Submitting All Tests/Close of Testing Window

After all testing for a grade level/content area is completed, the DTC/STC should review the Testing Status for each student in eDIRECT and communicate with Test Examiners to resolve any tests that appear as "In Progress." The DTC should also check the Testing Site Manager (if used) to ensure that there are no unsent responses. If all testing is completed for a grade level/content area prior to the end of the district's designated testing window for that grade level/content area, the DTC has the option to close that testing window early. To close a grade level/content window early, the DTC must contact CTB's MAP Service Line. Please note, only the DTC can request to close a district's testing window. It is very important that the DTC ensure that all testing for the grade level/content area is completed prior to closing a testing window. Once a testing window has been closed, scoring for that grade level/content area begins and the window cannot be re-opened for any reason. If the DTC does not request to close a testing window early, the window will close automatically at 8 P.M. on the end date that the DTC entered into eDIRECT when the testing window was set.

4.2 Reporting Test Invalidations

Neither a student's behavior during testing nor the judgment of a student's effort during testing can invalidate a student's test.

A MAP Grade-Level Assessment should be invalidated if a student is discovered cheating. To do so, select the "Teacher Invalidation" bubble for the affected content area in eDIRECT. (See the eDIRECT User Guide for instructions.) Cheating is the only time the "Teacher Invalidation" code is used. This code invalidates all sessions of the content area.

If the "Teacher Invalidation" bubble is used due to cheating, adhere to the following process:

1. The STC and the Test Examiner agree that a particular student's test should be invalidated.
2. A district invalidation letter on district letterhead and signed by the superintendent is faxed to Accountability Data at 573-522-6384.
3. The district invalidation fax should include the following information:
   - a. Student Name
   - b. MOSIS ID
   - c. Date of Birth
   - d. Grade
   - e. School Name
   - f. County District Code
   - g. District Name
   - h. School Code
   - i. Content Area
   - j. The reason the testing session is being invalidated/description of the incident
4. The district files a copy of the fax for its records and future reference.

4.3 How to Handle Student Absences

If a student is absent for any or all of the MAP Grade-Level Assessments and unable to test in make-up sessions, then mark the student as absent in eDIRECT. (ELLs in-country less than one year and being exempted from the ELA assessments are also treated as absences in eDIRECT.)

4.4 Securely Destroy Materials

Federal law—the Family Educational Rights and Privacy Act—prohibits the release of any student's personally identifiable information. Any printed materials must be securely stored and then shredded.

The STC or DTC should destroy the following materials at the building level:

- Printed copies of the *Test Administration Manual* should be destroyed after the final district content testing window has closed.
- Classroom Activity materials should be destroyed after the applicable grade and content area testing window has closed. Electronic files must be deleted.
- Large Print, Braille, and paper-and-pencil administration materials (i.e., manuals, printed pages from manuals, and glossary resource sheets) should be destroyed after the final district content testing window has closed. Electronic files must be deleted.

Scratch paper and graph paper must be kept in a securely locked room or locked cabinet that can be opened only with a key or keycard by staff responsible for test administration. All test materials must remain secure at all times. Scratch paper and graph paper must be collected and inventoried at the end of each test session and then given to the School Test Coordinator to securely destroy. DO NOT keep scratch paper for future test sessions except as noted below for performance tasks (PTs).

Use of Scratch Paper on Performance Tasks

The only exception to the requirement governing the destruction of printed materials and scratch paper is when notes are used during the ELA and Mathematics PTs.

During the ELA PT, students **must** use scratch paper to take notes during Part 1 in order for those notes to be available during Part 2. During Part 2, students are not able to return to the items in Part 1 or to the notes on the embedded universal tool, Sticky Notes, taken during Part 1. TEs should tell students to write their names (or some appropriate identifying information) on each piece of scratch paper, collect the scratch paper at the completion of Part 1 of the ELA PT, and securely store it for students' use during Part 2 of the ELA PT.

Likewise, the Mathematics PT may extend beyond one test session. When this happens, TEs should tell students to write their names on the scratch paper (and graph paper), collect the paper used in the first session, and securely store it for students' use in the subsequent test session.

The retention of scratch paper is only allowed during the PTs. Following the conclusion of the PT, all scratch paper and graph paper must be collected and inventoried and then given to the School Test Coordinator to securely destroy.

4.5 Individual Student Reports

Individual Student Reports (ISRs) are available in PRISM. A link to PRISM is in eDIRECT in the left-hand navigation pane. ISRs for ELA and Mathematics are available no later than the close of business on the tenth business day after each district content area testing window closes. ISRs for Science are available July 1, 2015.

5.0 LARGE PRINT, BRAILLE, AND PAPER-AND-PENCIL EDITIONS

Large Print, Braille, and paper-and-pencil editions of the MAP Grade-Level Assessments will be available for students with designated IEPs or special circumstances for spring 2015 testing. Large Print and Braille forms may be ordered online via eDIRECT during the enrollment period January 12, 2015, to February 20, 2015. Paper-and-pencil editions can be generated from eDIRECT (after students are registered for such an accommodation). Unique identification numbers will be used to produce barcodes that will be imprinted onto the paper-and-pencil editions. After testing, student responses for Large Print, Braille, and paper-and-pencil editions must be entered into the INSIGHT system and all test materials must be collected for return to CTB for processing and storage.

For additional information regarding Large Print and Braille forms, refer to the Large Print and Braille Kit and follow the instructions in the Braille Omit Return Instruction Sheet. Also, see http://dese.mo.gov/college-career-readiness/assessment/grade-level.

5.1 Before Testing

Paper-and-Pencil Materials

For special circumstances that require students to test on paper, a paper-and-pencil edition print feature is a part of the test delivery system. To activate the paper-and-pencil edition print function, Test Examiners will access the Test Setup feature in eDIRECT to mark the applicable accommodation and code for students who require the paper version of the test. Using the information collected during the precode and enrollment processes, the administration component of the online testing system will generate a unique barcode number for a paper-and-pencil edition prior to local printing. Depending on the printed accommodation needed for a particular student, the unique barcode number will then become embedded into the electronic version on each page of the paper-and-pencil form. During local printing, the embedded barcode number will print along with each page of the paper-and-pencil edition. Each barcode number will be unique to a student for the purposes of linking the printed form to the student's record in the master database. Barcode numbers will be recorded and associated with each student's record.

For specific instructions regarding how to generate a paper-and-pencil edition, see the Test Setup section of the eDIRECT User Guide, available on the **Documents** page of eDIRECT, https://mo.drcedirect.com.

Unless a student's IEP requires a paper-based accommodation, districts will be charged a processing fee of $15 for each paper-and-pencil PDF form of the test that is printed per content area.

Once the PDF downloads, it is available for printing on the local network printer.

A Test Examiner may print a paper-and-pencil edition to administer an oral reading accommodation or to transcribe Braille responses into the paper-and-pencil edition test book. It must be printed for a specific student.

The Test Examiner should become familiar with the directions for administering a paper-and-pencil edition. The paper-and-pencil edition of the test is secure and should be treated as such.

Large Print and Braille Materials

Large Print and Braille forms can be ordered online via eDIRECT. Material orders must be placed between January 12, 2015, and February 20, 2015. DTCs should order all Large Print and Braille materials through the Enrollments tab in eDIRECT. See the eDIRECT User Guide for enrollment instructions.

Test Examiners or Test Coordinators must transcribe students' responses into INSIGHT.

Large Print and Braille testing materials are packaged by building and shipped to the district's office address (or the shipping address indicated by the district during the registration process). The materials shipped to the district are based on the content-specific test window entered during registration.

District Test Coordinator

For every building administering a Large Print, Braille, or paper-and-pencil assessment, the DTC needs to make one copy of the Test Book Accountability Form for the STC. The Test Book Accountability Form is included in the District Test Coordinator Kit (TCK) and can be copied from Appendix D of this manual or printed from the **Documents** page of eDIRECT. Complete the following steps for each building before distributing copies to the STC:

1. Confirm the box count of the Large Print and Braille testing materials shipment from CTB (e.g., Box 1 of 5 through Box 5 of 5).
2. Verify the security barcode numbers of the test books against the packing list.
3. Record the number of test books listed on the packing list and the number of paper-and-pencil tests that were downloaded and printed on the Test Book Accountability Form.
4. Report any discrepancies to CTB's dedicated MAP Service Line at 1-800-544-9868 between the hours of 7:30 A.M. and 6:30 P.M. Central Time, Monday–Friday.

School Test Coordinator

After receiving the Test Book Accountability Form from the DTC, complete the following steps:

1. Verify that security barcode numbers printed on the Large Print and Braille test books match the numbers listed on the packing list (located in Box 1 of the building's shipment).
2. Confirm that the proper accommodation code is marked in eDIRECT.
3. Complete the Test Book Accountability Form, following the directions on the form.
4. Document any Large Print and Braille security barcode discrepancies.
5. Notify the DTC of any discrepancies immediately.
6. If any student is taking a MAP Grade-Level Assessment out of district/building, or if the student is homebound, note the barcode number of the test book before delivering it to the testing site to ensure proper accounting of all test books when they are returned to the district.
7. Ensure all test books have been accounted for before they are shipped to CTB.
8. Follow the procedures in Appendix C of this manual for any contaminated test materials.
9. Maintain the Test Book Accountability Form during the test administration, retain a copy for school records, and return the original with the testing materials to the DTC.

Test Examiner

Count the number of books received and assign each test book to a student. Write the student's name and MOSIS ID on the front of each test book.

Document this information in preparation for returning the test books to the STC.

Duration and Timing Information

The scheduling/rules for each component of the Large Print and Braille assessments are included in Tables 9, 10, and 11. Note that the duration, timing, break/pause rules, and session recommendations vary for each content area and component. This information is for scheduling purposes only, as the assessments are untimed.

Table 9: Assessment Sequence for Large Print Braille*, and Paper-and-Pencil—ELA

ELA	Session 1 (all grades)	Session 2 (all grades)	Classroom Activity (grades 5 and 8)	Session 3 (PT) (grades 5 and 8)
Number and Duration of Sessions	**Recommendations:** • Administer in one session. • Approximate session duration: 90–120 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 15–30 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 30 minutes or less. • Should occur one to three days prior to the PT. • Should NOT occur on the same day as the PT.	The Performance Task is presented in two parts. **Recommendations:** • Administer in two sessions corresponding to Parts 1 and 2 of the PT. • Approximate session durations: 35–40 minutes for Part 1 and 70–85 minutes for Part 2.
Breaks Within Sessions	A student may be provided breaks within a test session as needed.	A student may be provided breaks within a test session as needed.	NA	The PT is presented in two parts. Students can take breaks between Parts 1 and 2; however, once a student moves to Part 2, he or she will be unable to review or revise items in Part 1.
Total Duration	90–120 minutes	15–30 minutes	30 minutes or less	Approximate session durations: 35–40 minutes for Part 1 and 70–85 minutes for Part 2.

*Braille administration times will likely be longer than the times indicated here.

Table 10: Assessment Sequence for Large Print Braille*, and Paper-and-Pencil—Mathematics

Mathematics	Session 1 (all grades)	Session 2 (all grades)	Classroom Activity (grades 5 and 8)	Session 3 (PT) (grades 5 and 8)
Number and Duration of Sessions	**Recommendations:** • Administer in one session. • Approximate session duration: 15–90 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 15–90 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 30 minutes. • Should occur as close to the PT as is feasible, and no more than three days prior to the PT. • MAY occur on the same day as the PT.	**Recommendations:** • Administer in one session. • Approximate session duration: 15–90 minutes.
Breaks Within Sessions	A student may be provided breaks within a test session as needed.	A student may be provided breaks within a test session as needed.	NA	A student may be provided breaks within a test session as needed.
Total Duration	15–90 minutes	15–90 minutes	Less than 30 minutes	**Recommendation:** • Student completes the PT in one day.

*Braille administration times will likely be longer than the times indicated here.

Table 11: Assessment Sequence for Large Print Braille*, and Paper-and-Pencil—Science

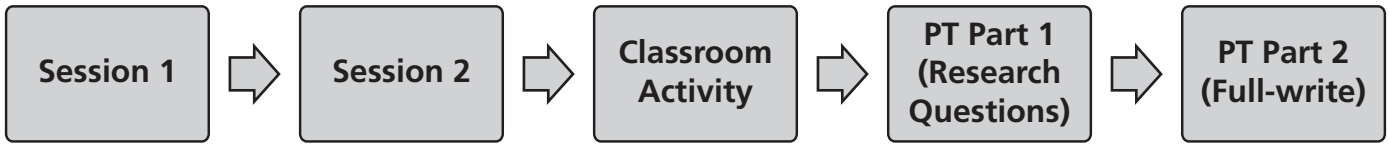
Science	Session 1	Session 2	Session 3 (Performance Event)
Number and Duration of Sessions	**Recommendations:** • Administer in one session. • Approximate session duration: 45–55 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 20–25 minutes.	**Recommendations:** • Administer in one session. • Approximate session duration: 45–65 minutes.
Breaks Within Sessions	A student may be provided breaks within a test session as needed.	A student may be provided breaks within a test session as needed.	A student may be provided breaks within a test session as needed.
Total Duration	45–55 minutes	20–25 minutes	45–65 minutes

*Braille administration times will likely be longer than the times indicated here.

Recommended Order of Test Administration

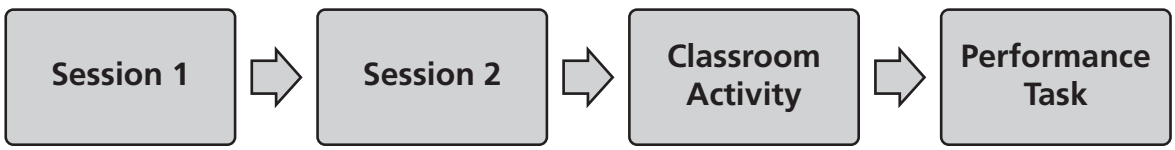
The assessments are comprised of two sessions for grades 3, 4, 6, and 7. The assessments are comprised of three sessions for each content area for grades 5 and 8. The third session for grades 5 and 8 is the performance task (PT) or performance event (PE). The ELA PT consists of two parts. The Mathematics PT and the Science PE each consist of one part. All ELA and Mathematics PTs must be preceded by the administration of a Classroom Activity.

Recommended Order of Test Administration for ELA



Students may take the non-PT portions of the test (Sessions 1 and 2) and Parts 1 and 2 of PT on separate days. For ELA, the order of administration should be Session 1 and Session 2, followed by the Classroom Activity, PT Part 1, and then PT Part 2. Districts/Schools may opt to administer in a different order if needed; however, the Classroom Activity, which is designed to introduce the PT, **must** occur prior to the PT.

Recommended Order of Test Administration for Mathematics



Students may take the non-PT portions of the test (Sessions 1 and 2) and PT portion of the test on separate days. For mathematics, the order of administration should be Session 1 and Session 2, followed by the Classroom Activity, and then the PT. Districts/Schools may opt to administer in a different order if needed; however, the Classroom Activity, which is designed to introduce the PT, **must** occur prior to the PT.

Recommended Order of Test Administration for Science

Classroom Activity

The purpose of the Classroom Activities is to introduce students to the context of a performance task so they are not disadvantaged in demonstrating the skills the task intends to assess. Classroom Activities do not address content information; instead, they focus on vocabulary and key contextual topics. The Classroom Activity is designed to be an introduction and not an assessment.

Students with designated IEPs are allowed to have accommodations, and English learners should have access to language supports that they regularly use during classroom instruction. The information noted in Table 13 provides Test Examiners with options that may be implemented during the activity as needed for students to have appropriate access to the information included in the Classroom Activity.

Overall Strategies for the Classroom Activity are as follows:

- Test Examiners may employ the same strategies for the Classroom Activity that they use during classroom instruction to attend to the diversity of their individual student needs.
- Test Examiners can employ formative practices and professional judgment to determine whether or not individual students require additional support or scaffolding to meet the objectives of the Classroom Activity.
- Test Examiners can read and reread aloud any text included in the Classroom Activity.
- Test Examiners may employ assistive technologies that are typically available during classroom instruction.
- The additional supports and strategies described in Table 13 may be made available to any student based on the student's individual needs and are not limited to particular impairments or to students who have Individualized Education Programs (IEPs) or 504 Plans.
- Test Examiners may adjust any Classroom Activity to allow for different instructional settings (e.g., individual student make-up activity, remote learning environment).
- Test Examiners may employ more than one suggested strategy listed in Table 13 to meet individual student's needs.
- These strategies are not mutually exclusive.

Table 13: Accessibility Guidelines for Classroom Activities

Student Need Category	Guidance for Accessibility
Visual Impairments	• Reading Materials: All materials that are required to be read by a student may be read aloud to the student. • Pictures, Figures, Drawings, and Photographs: Descriptions may be read to students. In addition, teachers can provide students with further explanation of the descriptions. These explanations may clarify the description without adding additional content. • Graphs: Further descriptions or repetition of descriptions may be read if necessary for a student. These explanations may clarify the description without adding additional content. • Venn Diagrams: Venn diagrams may be described to the student. In addition, a teacher may use a different chart, diagram format, or graphic organizer.
Reading Impairments	• Reading Materials: All materials that are required to be read by students may be read aloud to the student. • Writing Activities: All activities that require the student to write may allow for an oral response or the use of technology usually used by the student in a classroom environment.
Physical Impairments	• Kinesthetic Activities: If a student cannot participate in a kinesthetic activity, the student may be asked to describe the activity orally. • Activities Requiring Movement: Tasks such as moving around the room or coming up to the board can be modified to allow the teacher or other students to interact with the student or allow for the student to respond orally. • Writing Activities: If helpful to a student, all activities that require the student to write may allow for an oral response or the use of technology usually used by the student in a classroom environment.
Hearing Impairments	• Activities Requiring Listening: Listening activities may be presented in American Sign Language (ASL). For activities that require students to describe sounds, such as those from a thunderstorm, a sound may be described by the student to respond to how it feels and looks. • Activities Requiring Oral Responses: Oral responses may be provided via sign language or in writing.
Expressive Language Impairments	• Activities Requiring Oral Responses: Oral responses may be provided in writing, using a communication device, or any other means that the student uses to communicate.

Student Need Category	Guidance for Accessibility
English Language Learners	• Reading Materials: All materials that are required to be read by students may be read aloud to the student. • Writing Activities: All activities that require the student to write may allow for an oral response. • Visual Supports: If helpful to a student, vocabulary and key contextual topics may be supplemented with visual supports. • Flexible Grouping: Teachers may administer the Classroom Activity in flexible groups based on English language proficiency. • Activities Requiring Oral Responses: Oral responses may be provided in writing. • Students may use an English, a non-English, and a bilingual dictionary and thesaurus as needed.
Separate Setting	• Group activities may be tailored to occur between a single student and his or her educator where the educator and student share discussion and work. • Activities between student(s) and an educator may be conducted online or via a telephone connection. • All student-facing information included in a Classroom Activity should be presented to students working in a separate setting.

Classroom Activities are located on the **Documents** page of eDIRECT, https://mo.drcedirect.com. Also on the **Documents** page is a lookup table indicating which Classroom Activity should be administered for the Large Print, Braille, and paper-and-pencil editions.

1. From the eDIRECT homepage, log in using your eDIRECT credentials.
2. In the left navigation pane, under **General Information**, select **Documents**.
3. In the main page on the **Documents** tab,
    a. Choose "Summative Grade-Level Assessments Spring 2015" from the Administration drop-down.
    b. Choose "Classroom Activities" from the Document Type drop-down.
    c. Click "Show Documents." A list of all available Classroom Activities and the lookup table will appear in the grid.

5.2 During Testing

This section provides an overview of preparing the testing environment, guidelines for test administration, and directions for accessing specific scripts for administering the Large Print, Braille, and paper-and-pencil editions. Test Examiners should become familiar with this section well in advance of the start of testing.

The scripts are secure; do not print or allow unauthorized persons to access them. Maintaining the security of all test materials is crucial to obtaining valid and reliable test

results. Therefore, test materials must be kept in locked storage, except during actual test administration. It is the responsibility of all individuals who administer the test to follow security procedures.

Before administering the assessment, make sure that you have the following materials available for students:

- A test book for each student
- At least two sharpened No. 2 pencils
- Blank scratch paper for each student
- Blank grid/graph paper for all Science Assessments and for the Mathematics Assessments for grades 6 and above
- An English dictionary and a thesaurus for the full-write essay portion (Part 2) of an ELA performance task
- Any additional materials appropriate for the student that are noted as "Any" or "Paper" in Tables 4, 5, and 6 of this manual
- A calculator for the calculator-allowed portion of the Mathematics Assessment (Calculators must meet the guidelines below.)
    - For grade 6 Mathematics Assessments, a four-function calculator with square root and percentage functions is permitted. (This type of calculator is permitted for grades 3–5 as an accommodation only, as the assessments include no calculator-allowed items.)
    - For grades 7 and 8 Mathematics Assessments, a scientific calculator with exponents, trigonometry, and logarithmic functionalities is permitted.
    - DESE does not provide, endorse, or recommend a list of calculator brands or types that students are permitted to use. Test Examiners should follow their own district's general education policy for the types of calculators permitted during district-administered quizzes, benchmark tests, common assessments, chapter/unit tests, and final exams.
    - Calculators cannot contain stored equations or functions at the time of the MAP Grade-Level Mathematics Assessments. Test Examiners are responsible for ensuring and verifying that calculators that have the ability to store functions and equations, e.g., a scientific calculator, have the memory cleared before and after each Mathematics Assessment.
    - Calculators cannot have Internet connectivity or be able to connect to anyone inside or outside the classroom during testing. Students cannot use a calculator on a laptop or other portable computer, pocket organizer, cell phone, device with a typewriter-style keyboard, electronic writing pad, or pen-input device unless a particular assistive device is required for a student and is specified on his or her IEP.
    - No calculators with QWERTY keyboards are allowed.

NOTE: If students are allowed to bring cell phones into the testing room, the cell phones must be collected and kept in a central area until testing is completed.

Specific Directions for Administering the Braille Form

The directions in this manual also apply to the administration of the Braille version of the English Language Arts Summative Assessment. Additional Braille instructions are as follows:

- References to specific page numbers in the student test book may be incorrect for the Braille version. To supply the correct page numbers and other references, Test Examiners should review—prior to testing—all test materials that accompany the Braille test book.
- The student's name, Test Examiners, school, and system must be printed on the front cover of each Braille test book.
- Because extra time may be needed for administering the Braille version, it is recommended that students be tested individually or in a small-group setting.
- When a Braille student responds by pointing to the answers or giving a verbal response in English only, the Test Examiner is permitted during the course of test administration to fill in student responses in the student test book. When a Braille student responds by using a Braillewriter or marking answers in the test book, the procedures for transcribing student responses detailed in the 5.3 "After Testing" section of this manual should be followed. In each instance, the Test Examiner must provide written affirmation to the School Test Coordinator that student responses have been completed in the student test book with accuracy. Under no circumstances should a student's answer be altered or edited—to do so is a direct violation of test security.

Scripts for Administering the Large Print, Braille, and Paper-and-Pencil Editions

The specific scripts for administering the Large Print, Braille, and paper-and-pencil editions of each assessment are located on the **Documents** page of eDIRECT, https://mo.drcedirect.com.

1. From the eDIRECT homepage, log in using your eDIRECT credentials.
2. In the left navigation pane, under **General Information**, select **Documents**.
3. In the main page on the **Documents** tab,
   a. Choose "Summative Grade-Level Assessments Spring 2015" from the Administration drop-down.
   b. Choose "Scripts" from the Document Type drop-down.
   c. Click "Show Documents." A list of all available scripts will appear in the grid.

5.3 After Testing

Assemble Materials for Return and for Entry into INSIGHT

After testing has been completed, prepare materials to be returned to the School Test Coordinator. Check test books to make sure there are no sticky notes, staples, pins, paper clips, or tape of any kind on any pages. Check to make sure that no scratch or graph paper was left inside test books. Remove any extraneous material.

Transcription of Large Print, Braille, and Paper-and-Pencil Editions

After testing, student responses for Large Print, Braille, and paper-and-pencil editions **must** be transcribed into the INSIGHT testing software before the district's test window closes. It is recommended that transcription occur as soon after testing as possible. To transcribe responses requires the Test Examiner or other designated and authorized district or school

personnel to log in to INSIGHT using the student's Test Ticket. Follow these steps to transcribe student answers:

1. In eDIRECT Test Setup, ensure that the student has been assigned the appropriate accommodation:
   a. Paper-Based Assessment
   b. Paper-Based Braille
   c. Paper-Based Large Print
   d. Non-Accommodation Special Case—Paper-Based Assessment
2. In eDIRECT Test Setup, assign the student to a test session and print their Test Ticket. Retain the Test Ticket rather than distributing it to the student.
3. After the student has completed the test on paper, use a test machine that has the INSIGHT client software installed and use the student's Test Ticket to log into the student's test.
4. Begin transcribing student responses. Once you have finished, select End Test and Submit. The Test Examiner should then return all printed test materials to the STC.

Transcribe the student's responses as faithfully and as completely as possible using the following guidelines.

- Do not transcribe erased or crossed out words or marks.
- If a student's response consists of incomprehensible squiggles, marks, etc., which clearly are not words or word fragments, then leave the item blank.
- If a student's response is wholly or partly illegible, enter "ILLEGIBLE" for the entire response or for the part where applicable.
- If 50% or more of a student's response is written in any language other than English, then note "WRITTEN IN ANOTHER LANGUAGE" where applicable.
- If part of a student's response cannot be entered into INSIGHT, then leave that part blank.
- If no part of a student's response can be entered, then leave the entire item blank.
- Additional clarifying notes may be entered as needed if the item type allows text entry.

Arrange for the Return Shipment of Large Print, Braille, and Paper-and-Pencil Test Books to CTB

All secure Large Print, Braille, and paper-and-pencil test books must be returned to CTB via FedEx. Shipping Return labels are provided in the District Test Coordinator Kit (TCK). If the DTC does not have shipping labels or shipping boxes, please contact the MAP Service Line at 1-800-544-9868.

CTB is responsible for all return shipping costs for the Large Print, Braille, and paper-and-pencil test books; however, the DTC must make shipping arrangements at least 24 hours in advance of package pickup.

NOTE: DTCs MUST use CTB boxes to return Large Print, Braille, and Paper-and-Pencil test books. Braille and Large Print Assessments are shipped to the district in a kit that includes boxes and labels necessary for returning testing materials. Paper-and-pencil test books may be returned in the same shipping boxes with Braille and Large Print test books. If the district downloaded paper-and-pencil test books, but did not order any Braille or Large Print test books, the DTC must call for boxes and shipping labels to return the paper-and-pencil test books.

Organize Materials for the District Test Coordinator

Instructions for the School Test Coordinator

Make sure that all Large Print, Braille, and paper-and-pencil testing materials are received from each Test Examiner in the school. Contact any Test Examiner who delays returning student testing materials.

Follow these guidelines for packaging testing materials for the DTC:

1. **Obtain Boxes**

   Test materials must be returned in the CTB boxes with aqua shading. Reuse the boxes in which the Large Print and Braille testing materials arrived. If the DTC does not have CTB boxes or needs additional boxes, please contact the MAP Service Line at 1-800-544-9868.

   Prior to packing test materials, securely tape the bottom of each box to prevent breakage. Reinforce all bottom seams, following an "H" pattern.

2. **Package Materials**

   Place the following materials in boxes in the order specified below, with the first items listed on the top in Box 1.

   - Paper-and-pencil test books
   - Braille test books
   - Large Print test books

3. **Affix Shipping Labels**
   - Affix the **white shipping labels** to the boxes. Labels should be placed on the side of the box in the white space marked "PLACE CTB BARCODE RETURN LABEL HERE." Do not place the label on the top of the box.
   - **Number each set of boxes separately for each school** (e.g., "1 of X," "2 of X," etc., where "X" is the total number of boxes per school).
   - Complete all of the information requested on the labels.

   Return shipping labels are scannable and cannot be photocopied. If more return shipping labels are needed, contact the CTB dedicated MAP Service Line at 1-800-544-9868.

4. **Send Materials to the District Test Coordinator**
   - Do not seal the boxes of test books.
   - The DTC will review the contents of each box.

Package and Ship Testing Materials

Instructions for the District Test Coordinator

Make sure that all testing materials are received from each school in the district. Contact any STC who delays returning school testing materials. Verify that the STC followed the instructions in this *Test Administration Manual*.

If a box from an STC is received without a return shipping label on it, affix one of the blank District return shipping labels that were provided in the DTC's Package. Fill out the School information on the label to ensure correct processing.

Do **not** return the following to CTB:

- *Test Administration Manuals*
- Classroom Activity materials (must be **securely** destroyed by district)
- test administration scripts and glossary resource sheets for the Large Print, Braille, or paper-and-pencil editions (must be **securely** destroyed by district)
- scratch and/or grid paper used for the English Language Arts, Mathematics, and Science Assessments (must be **securely** destroyed by district)
- contaminated test materials (must be **securely** destroyed by district; see Appendix C in this manual)
- unused return shipping labels
- Test Book Accountability forms (keep for your records)

Check all materials from the STCs to ensure they have correctly followed the procedure described in this manual.

1. **Add Packing Material**

   To avoid damage caused when materials shift during transit, add sufficient packing material to fill all voids and hold documents firmly in place. We strongly recommend using crumpled, recycled paper for this purpose. Do **not** use foam packing "peanuts" or "popcorn."

2. **Seal Boxes**

   Seal each box securely with packing tape to reinforce the top and side seams of the boxes. This will prevent damage to the boxes and subsequent loss of test materials.

3. **Schedule Testing Material Pickup**

   The DTC will schedule the pickup of MAP Grade-Level Assessment testing materials through a secure ctb.com pickup site. Contact CTB via the ctb.com site no later than May 26, 2015, to schedule your pickup date. Please allow 1–3 days for pickup of your test materials. All materials must be picked up no later than May 29, 2015.

   Test materials must be returned via the secure ctb.com pickup site in order to ensure secure tracking of materials.

   Materials must be returned in a single shipment unless prior arrangements are made with CTB.

   Instructions for scheduling the pickup of MAP Grade-Level Assessment testing materials:

   1. Go to http://programs.ctb.com/MAP.
   2. Enter your district number, contact name, and email address.
   3. Verify the pickup address and enter the number of boxes to be picked up.
   4. Click submit.

5. Print the FedEx shipping PDF label from the Confirmation page on standard 8.5 x 11 paper.
6. Fold the shipping label in half and securely tape it to your box in the location marked Carrier label. You will receive a pickup confirmation email. The email contains the pickup confirmation number and FedEx phone number.

If you have any questions regarding the pickup of materials, call FedEx at the number provided on your confirmation email for assistance.

Store boxes in a protected area while waiting for FedEx pickup.

4. **Fax Test Book Accountability Forms to CTB**

   After you have confirmed that you have received completed, signed Test Book Accountability Forms from each school, fax them to CTB at the fax number listed on the form.

5. **Questions**

   For answers to any questions regarding the return procedures described in this manual, call the CTB dedicated MAP Service Line at 1-800-544-9868.

Securely Destroy Other Materials

See Section 4.4 and Appendix C in this manual for details regarding the destruction of materials not returned to CTB.

Appendix A: Item Types

As students engage with the MAP Grade-Level Assessments, they will be asked test questions that require them to use technology to respond in several ways, some of which may be new to students. The following table lists the different item types and briefly describes each one.

Content Area	Type of Item	Brief Description of How to Respond
ELA, Mathematics, and Science	Selected Response (also known as Multiple Choice, single correct response)	Select the radio button corresponding to one of four options. To deselect an option, select a different radio button. Select only one option.
	Short Text (also known as Constructed Response)	Respond via keyboard entry into text box (no text formatting). This item type offers the ability to edit previously entered text.
ELA and Mathematics	Multiple Choice, multiple correct responses	Mark a checkbox corresponding to an option. To deselect an option, click on the checkbox that is already marked. Mark one or more options.
	Matching Tables (with a variation True/False or Yes/No)	Select a checkbox corresponding to an option in a table cell. To deselect an option, select a checkbox that is already marked.
Mathematics and Science	Drag-and-Drop	Click and drag an object to the appropriate location in the response area.
	Table Fill In	Respond via keyboard entry into table cells or drag/drop objects into table cells.
ELA Only	Two-part multiple choice, with evidence-based response (EBSR)	This item type has two parts. Each part may consist of one of three item types: Multiple Choice, single correct response; Multiple Choice, multiple correct responses; and Hot Text, Select Text. See those item types for descriptions of how to respond.
	Hot Text, Select Text	Highlight an option by selecting it. To deselect an option, click on it to remove the highlighting. Select one or more options.
	Hot Text, Reorder Text	Select text and then click and drag text to a new area.
	Essay	Respond via keyboard entry using text formatting buttons. This item type offers the ability to edit previously entered text.

Content Area	Type of Item	Brief Description of How to Respond
Mathematics Only	Hot Spot	Select targeted areas in the response area.
	Equation/Numeric	Select buttons representing numbers and mathematic symbols to create a numeric response or equation.
	Graphing	Plot points and/or draw lines in the response area.
Science Only	Bar Graphing	Click targeted areas in the response area and respond via keyboard entry into response fields.
	Line Graphing	Plot points and/or draw lines in the response area. Respond via keyboard entry into response fields.
	Build a Table	Respond via keyboard to make entries into table fields.

Appendix B: Handling Student Transfers and Changes in Testing Status

Students Who Move Before or During the MAP Grade-Level Assessment Administration

If…	*then…*
a student needs to be moved into a different test session in the same school:	Edit the student's profile by moving the student to a new test session.*
a new student moves into the district:	Add the new student in eDIRECT. Then assign the student to the appropriate test session(s).* **NOTE:** If the DTC is unable to add the new student, the DTC must contact the MAP Service Line.
a student moves out of the district prior to or during the district test administration window:	Remove the student from any test session in eDIRECT. Do **not** log into the test and do **not** mark any status code(s) for the student.*
a student moves from one building to another building within the same district:	The DTC should edit the student's information in eDIRECT before the student begins testing so that the student's scores report to the correct building. The DTC must move the student to a different test session in eDIRECT.*

*See the eDIRECT User Guide, available on the **Documents** page of eDIRECT, https://mo.drcedirect.com.

Please contact the CTB dedicated MAP Service Line at 1-800-544-9868 if there are any questions regarding moving a student within a school or district.

Appendix C: Contaminated Test Materials

Test materials are considered *contaminated* due to: a) a student health issue that affects the test book itself (blood, fluids, etc.) or b) contact with any potentially hazardous material. If test materials are contaminated, the Test Examiner should notify the School Test Coordinator for instructions for handling the contaminated materials since **all** printed testing material must be accounted for. The DTC, or STC, or TE is responsible for transcribing the answers into the online system, and then the contaminated test materials must be securely destroyed at the test site by the DTC or STC. A Missing Test Materials Form must be completed and faxed to CTB and DESE to account for the contaminated test materials. The form may be accessed on the **Documents** page of eDIRECT, https://mo.drcedirect.com, or on the DESE website at http://dese.mo.gov/college-career-readiness/assessment/grade-level.

The STC should provide the DTC with the following information for inclusion on the form:

- an explanation of what happened to the test book
- security barcode number (write or cut-and-paste it onto the letter). This is the code beginning with two letters, followed by six numbers, printed vertically below the barcode on the front book cover. Be sure to use this number and not the number from the student barcode label.
- school name
- school code
- student's name
- grade level
- test book edition (Large Print, Braille, or paper-and-pencil)

MAP 2015 GRADE-LEVEL ASSESSMENTS – TEST BOOK ACCOUNTABILITY FORM

District Name: _____ **District #:** _____

School Name: _____ **School #:** _____

This form provides start-to-finish accountability for the Grade-Level test materials 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. 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 and printed PDF tests.
3. Return the completed form to your District Coordinator.

District Coordinator:

4. Complete a copy of this form for District Overage, providing any additional information on the back as required.
5. 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 or PRINTED	GR 3	GR 4	GR 5	GR 6	GR 7	GR 8
(1)	Number of LP/Braille test books listed on packing list	+	+	+	+	+	+
(2)	LP/Braille test books missing from shipment (Fill out A on back)	–	–	–	–	–	–
(3)	Extra LP/Braille books received in shipment (Fill out B on back)	+	+	+	+	+	+
(4)	Additional LP/Braille books from district office (Fill out C on back)	+	+	+	+	+	+
(5)	Printed PDF tests (Fill out D on back)	+	+	+	+	+	+
(6)	**Total test books received and printed (Add lines 1, 3, 4 and 5; then subtract line 2)**						
	TEST BOOKS and PRINTED PDF TESTS RETURNED	**GR 3**	**GR 4**	**GR 5**	**GR 6**	**GR 7**	**GR 8**
(7)	Number of LP/Braille tests administered	+	+	+	+	+	+
(8)	Number of unused LP/Braille test books	+	+	+	+	+	+
(9)	Number of printed PDF tests	+	+	+	+	+	+
(10)	**Total test books returned and printed (Sum of lines 7–9)**						
	TEST BOOKS and PRINTED PDF TESTS NOT RETURNED	**GR 3**	**GR 4**	**GR 5**	**GR 6**	**GR 7**	**GR 8**
(11)	LP/ Braille test books securely destroyed (Fill out E on back)	+	+	+	+	+	+
(12)	LP/Braille test books unaccounted for (Fill out E on back)	+	+	+	+	+	+
(13)	Printed PDF tests unaccounted for (Fill out E on back)	+	+	+	+	+	+
(14)	**Total test books and printed PDF tests not returned (Sum of lines 11–13)**						

I confirm that Line 6 = Line 10 + Line 14.

School Test Coordinator: _____ Date Faxed: _____

Print Name: _____

A. Security barcode numbers of LP/Braille test books missing from shipment	B. Security barcode numbers of extra LP/Braille test books (not on packing list) received in shipment	C. Security barcode numbers of additional LP/Braille test books from district office

D. Security barcode numbers of printed PDF tests	E. Security barcode numbers of lost or securely destroyed LP/Braille test books or printed PDF tests	Explanation

Appendix E: INSIGHT Keyboard Shortcuts and Icons

The following list contains the keyboard shortcuts and icons available in INSIGHT. All students may have access to a printed copy of this list during online testing.

INSIGHT Function	Keyboard Shortcut		
	Desktop	Chromebook	iPad
Transfers the focus from one button to the next (from left to right). The focus is indicated by a red box that appears around the selected tool or function button when the Tab key is pressed.	Tab	Tab	N/A
Transfers the focus from one button to the next (from right to left). The focus is indicated by a red box that appears around the selected tool or function button when the Shift key and Tab key are pressed.	Shift + Tab	Shift + Tab	N/A
Activates the tool or function highlighted by the red box. Pressing the Enter key or Space Bar a second time deactivates the tool or function (with the exception of tools that keep the focus, such as Sticky Notes).	Enter/Space Bar	Enter/Space Bar	N/A
Selects the highlighted test question from the Review/End Test page Selects the Sign In button after a Username and Password are entered Selects Continue from the Student Verification Page Selects the Go To Page number within the quick navigation drop-down menu	Enter	N/A	N/A
Closes the Magnifier and "?" [Help] button when activated. If the red box is activated and the Esc key is pressed while on the tool bar without having any tools activated, the red box will move to the Pointer button.	Esc	Esc	N/A

INSIGHT Function	Keyboard Shortcut		
	Desktop	Chromebook	iPad
Selects an answer option (i.e., ABCD) on a multiple-choice question when only one set of "ABCD bubbles" exists. Entering one of the letters fills or unfills the letter bubble before the answer option. Both uppercase and lowercase letters can be used	ABCD, abcd	ABCD, abcd	N/A
Exits the online testing system from each page that has an Exit button	Alt + X	Alt + X	N/A
Moves any pop-up tool, such as the "?" [Help] button around the screen. (Does not work with Sticky Notes.)	CTRL + Right Arrow CTRL + Left Arrow CTRL + Up Arrow CTRL + Down Arrow	CTRL + Right Arrow CTRL + Left Arrow CTRL + Up Arrow CTRL + Down Arrow	N/A
Rotates the active tool +/– 1 degree	CTRL + plus [+] CTRL + minus [–]	CTRL + plus [+] CTRL + minus [–]	N/A
Moves the cursor up and down through a list of choices (such as questions on the Review/End Test screen)	Up/Down Arrows	Up/Down Arrows	N/A
Switches between multiple active pop-up tools on the screen	CTRL + Tab	CTRL + Tab	N/A
Activates the Review/End Test button and moves the user to the Review page of the test	Alt + R	Alt + R	Option + R
Activates the Pause button and pauses the test	Alt + P	Alt + P	Option + P
Activates the Flagged button and marks an item as flagged, or removes flag from an item	Alt + F	Alt + F	Option + F
Activates the Back button and moves the student back a question	Alt + B	Alt + B	N/A
Activates the Next button and moves the student forward a question	Alt + N	Alt + N	N/A

Appendix E: INSIGHT Keyboard Shortcuts and Icons continued

INSIGHT Calculator Function	Keyboard Shortcut		
	Desktop	Chromebook	iPad
Clears the calculator screen	Alt + Delete	Alt + Delete	N/A
Works as a shortcut key for subtracting on all calculators	–	–	
Works as a shortcut key for factorial on the Scientific Calculator/Graphing Tool	!	!	
Works as a shortcut key for using open parenthesis on the Scientific Calculator/Graphing Tool	(	(	
Works as a shortcut key for using closed parenthesis on the Scientific Calculator/Graphing Tool	)	)	
Works as a shortcut key for multiplying on all calculators	*	*	
Works as a shortcut key for dividing on all calculators	/	/	
Works as a shortcut key for squaring on the Scientific Calculator/Graphing Tool	@	@	
Works as a shortcut key for adding on all calculators	+	+	
Work as shortcut keys for numeric entry on all calculators	0–9	0–9	
Works as a backspace on all calculators	Backspace	Backspace	
Works as a delete function on all calculators	Delete	N/A	
Works as a shortcut to take a number to a specific power on the Scientific Calculator/ Graphing Tool	^	^	
Works as the negate key on the Basic Calculator	‚	‚	

INSIGHT Audio (TTS) Function	Keyboard Shortcut		
	Desktop	Chromebook	iPad
Activates the Options button and opens or closes the Audio settings selection pop-up window	Alt + A	Alt + A	Option + A
Activates the Options button and opens or closes the Color Chooser selection pop-up window	Alt + O	Alt + O	Option + O
Activates the Play/Pause button when Audio is active	F8 (Mac — use FUNC F8)	N/A	N/A
Writing Tools			
Undo	CTRL + Z	N/A	CMD + Z
Redo	CTRL + Y	N/A	CMD + Shift + Z
Highlight text to the left	Shift + left arrow	Shift + left arrow	N/A
Highlight text to the right	Shift + right arrow	Shift + right arrow	N/A
Highlight all text	CTRL + A	CTRL + A	CMD + A
Cut highlighted text	CTRL + X	CTRL + X	CMD + X
Copy text from clipboard	CTRL + C	CTRL + C	CMD + C
Paste text from clipboard	CTRL + V	CTRL + V	CMD + V
Move to start of next word	CTRL + right arrow	CTRL + right arrow	Option + right arrow
Move to start of previous word	CTRL + left arrow	CTRL + left arrow	Option + left arrow
Move cursor forward one character	Right Arrow	Right Arrow	Right Arrow
Move cursor backward one character	Left Arrow	Left Arrow	Left Arrow
Delete text (from cursor position) to the end of the line	N/A	N/A	Control + K
Delete text (from cursor position) to the beginning of the line	N/A	N/A	CMD + Delete

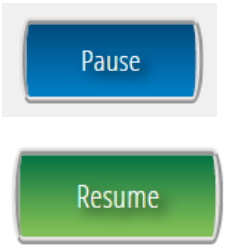
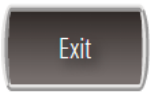
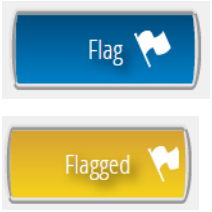
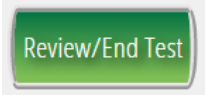
Appendix E: INSIGHT Keyboard Shortcuts and Icons continued

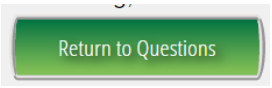
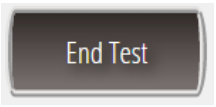
Writing Tools	Keyboard Shortcut		
	Desktop	Chromebook	iPad
Delete the word before the cursor	N/A	N/A	Option + Delete
Jump cursor location to end of text entered	N/A	N/A	CMD + Left Arrow
Jump cursor location to the beginning of text entered	N/A	N/A	CMD + Right Arrow
Jump cursor location to previous start of line	N/A	N/A	Option + Up Arrow
Jump cursor location to next end of line	N/A	N/A	Option + Up Arrow
Apply bold formatting / repeat to turn off	N/A	N/A	CMD + B
Apply italic formatting / repeat to turn off	N/A	N/A	CMD + I
Apply underline formatting / repeat to turn off	N/A	N/A	CMD + U

Appendix E: INSIGHT Keyboard Shortcuts and Icons continued

Tool Icon	Tool Name	Tool Definition
	**Pointer**	The **Pointer** tool is the default tool that is active when you begin. It is used to select answers as well as other tools and features within the online assessment. The **Pointer** will change to a pencil head when moved over a multiple-choice answer bubble. Use it to select your answer. If another tool has been selected, you can return to the **Pointer** tool mode by clicking on the Pointer tool button. This button is at the far left of the tools row.
	**Cross-Off**	The **Cross-Off** tool is used to narrow down the possible answer choices by allowing you to mark answer choices you believe to be incorrect. This tool is only available for multiple-choice items.
	**Highlighter**	The **Highlighter** tool is used to highlight important information.
	**Sticky Note**	The **Sticky Note** allows you to place a short note almost anywhere within the window that contains a question, passage, or scenario. Use a note to mark a special part or to leave a reminder of some important information in that question, passage, or scenario.
	**Magnifier**	The **Magnifier** allows you to enlarge the entire screen. Other tools, including the **Line Guide, Cross-Off, Highlighter,** and **Calculator,** can be used when the **Magnifier** is turned on.
	**Line Guide**	The **Line Guide** tool provides a horizontal line that brings the focus to a single line of text. The **Line Guide** can be used to track a passage or an individual question.
 	**Measurement Tools**	The **Measurement Tools** button allows you to access the ruler, which can be used to measure an object. The ruler can be moved around the screen and can also be rotated.

Appendix E: INSIGHT Keyboard Shortcuts and Icons continued

Tool Icon	Tool Name	Tool Definition
	**Calculator**	The **Calculator** tool may be used to assist with calculations necessary to answer questions on the exam. You will be given a Basic or Scientific calculator.
	**Graphing Tool**	The **Graphing Tool** is designed to graph functions when solved for the "Y" variable and has the ability to give the corresponding "Y" values for given "X" values.
	**Next Button** **Back Button**	The **Next** and **Back** buttons are used to navigate between questions on the test. They are also used to move between pages on multi-page questions. Click on the **Next** button to move forward to the next question or page. Click on the **Back** button to move backward to the previous question or page.
	**Pause** and **Resume**	When the **Pause** button is clicked, the test will be temporarily stopped. The test cannot be paused for more than 20 minutes. A countdown timer will be displayed showing how much longer the test will be paused. At any time during the countdown, the test can be resumed by clicking on the **Resume** button.
	**Exit**	The **Exit** button appears on the Pause Page. Click on **Exit** to close the test. WARNING: If a student exits a test using this button, the test remains incomplete. The student must log in again to complete the test.
	**Flag**	Click on the **Flag** button to mark a test question for review at a later time. When you click on the Flag button, the color of the button will change to yellow to indicate the question is flagged. To unflag a test question, use the **Pointer** tool to click the button again.
	**Review/End Test**	The **Review/End Test** button allows you to see all of the test questions you have flagged for review. The Review Page also shows which questions have been answered and which have not.

Tool Icon	Tool Name	Tool Definition
	**Return to Questions**	The **Return to Questions** button appears on the Review Page. Clicking **Return to Questions** will take the student back to the most recently visited question. The student can then review any questions, and proceed by clicking **Review/End Test** again.
	**End Test**	The **End Test** button appears on the Review Page. Clicking this button will provide a prompt for the student to confirm whether they would like to **Return to Review** or **End Test.** Clicking on the **End Test** button will end the exam.
	**Go to Question**	To quickly navigate to any question, passage, or scenario on the test, click on the down arrow next to the question number in the upper-left corner of the screen. A list of all available test questions and scenarios will appear. Click on the number of the test question, passage, or scenario you want to go to, and that question will appear on the screen. Click on the passage or scenario and you will be taken to the first question that appears with the passage or scenario.

Review Page Key

Key Icon	Key Description
	Unanswered multiple choice item
	Answered multiple choice item
	Blank constructed response item
	Filled constructed response (text has been entered into the response box)
	Flagged item
S	Scenario indicator for Science; example (S1)
P	Passage indicator for ELA; example (P1)

CTB/McGraw-Hill
20 Ryan Ranch Road
Monterey, California 93940-5703
800.538.9547 | www.ctb.com

Appendix B
MAP Grade-Level Assessments: Guide to Interpreting Results

Missouri Assessment Program Grade-Level Assessments

Guide to Interpreting Results

Summative Assessments

English Language Arts/Literacy, Mathematics, and Science

Revised June 2015

This guide has been prepared by CTB to provide an overview for interpreting reports generated from the Missouri Assessment Program (MAP). It is intended to help educators apply MAP data to the needs of individual students and the district as a whole.

Table of Contents

Introduction

Educational Assessment: A Primary Tool

Assessment, or testing, fulfills a vital role in today's educational environment. Assessment results often are a major force in shaping public perceptions about the capabilities of our students and the quality of our schools. As a primary tool for educators and policymakers, assessment is used for many important purposes. Educators use assessment results to help improve teaching and learning and to evaluate programs and schools. Policy decisions are often based, in part, on assessment data. Because of its important role, educational assessment is used in every school, district, and state. It is vital to innovation, higher standards, and educational excellence.

Originally developed in response to Missouri's Outstanding Schools Act of 1993, the **Missouri Assessment Program (MAP)** encompasses several statewide assessments that meet state and federal statutory requirements. **MAP Grade-Level Assessments** are administered to students in grades 3 through 8 to determine their progress toward the Show-Me Standards/Missouri Learning Standards. As directed by the Outstanding Schools Act, the Show-Me Standards were developed by the Missouri Department of Elementary and Secondary Education (DESE), in cooperation with teachers, school administrators, parents, and business professionals throughout the state, to identify the knowledge, skills, and competencies that Missouri students should acquire prior to graduating from high school. For a more detailed explanation of the Show-Me Standards, refer to the DESE website (http://dese.mo.gov/show-me-standards). The Missouri Learning Standards articulate the Show-Me Standards in each content area across the grade levels. MAP Grade-Level Assessment items are aligned with the Missouri Learning Standards, which are available on the DESE website (http://dese.mo.gov/college-career-readiness/curriculum/missouri-learning-standards).

The spring 2015 Grade-Level MAP includes the following required assessments:

English Language Arts (ELA)—Grades 3–8

Mathematics—Grades 3–8

Science—Grades 5 and 8

For students in grades 3, 4, 6, and 7, the ELA and Mathematics assessments require approximately 1½ to 2 hours of test administration time per content area. For students in grades 5 and 8, the assessments require approximately 3 to 4 hours of test administration time per content area for ELA and Mathematics. In addition, students in grades 5 and 8 take a Science assessment requiring an additional 2 to 2½ hours of test administration. All assessments are administered online, unless students require a Braille, Large Print, or paper/pencil form as an accommodation.

For all grade levels (3 through 8), the MAP Grade-Level Assessments in ELA and Mathematics include multiple item types. **Selected-response items** (also known as multiple-choice) present students with a question followed by three or more response options. **Short-text items** require students to type an appropriate response. **Technology-enhanced items** use innovative technology to allow students to demonstrate their knowledge in ways that are not possible using paper/pencil assessments. For example, the items may include embedded video or audio; they may require students to drag and drop data into a table, click on "hot spots" within a graphic, or indicate their response on a grid. Short-text items are scored by trained readers using specific criteria. Trained readers are always humans, not machines. Some technology-enhanced items (for example, gridded response items) are machine scorable. Others are scored by trained readers.

The ELA and Mathematics assessments in grades 5 and 8 also include a **performance task**. Prior to the performance task, all students participate in a 30-minute classroom activity. Performance tasks require students to provide a series of responses. In ELA, the performance task includes research questions and a full-write essay. The research questions require students to interpret provided informational sources, which are the foundation for the full-write essay. In Mathematics, the performance task may require students to do such things as simulate a study and present and interpret data in a table or graph. Students are required to explain their responses; and often the task allows more than one approach to arrive at a correct response. The ELA full-write essay is scored by trained readers using a 6-point rubric that evaluates purpose and organization, evidence and elaboration, and conventions. The Mathematics performance task is also scored by trained readers using scoring criteria that are specific to each task.

The MAP Grade-Level Assessments in Science include selected-response items, as well as **constructed-response items**, which require students to supply their answer (similar to short-text items), and **performance events**. Similar to the ELA and Mathematics performance tasks, Science performance events require students to provide an extended response, and require students to apply their knowledge and understanding in real-life situations. Like the ELA and Mathematics short-text items and performance tasks, the Science constructed-response items and performance events are scored by trained readers using established scoring criteria.

The Department uses the information obtained through MAP to monitor the progress of Missouri's students toward meeting the Show-Me standards in order to inform the public and the state legislature about student performance and to help make informed decisions about educational issues. The information obtained through MAP provides the academic performance data that drive student services throughout the state. The **MAP reports** provide useful information for determining the performance of individual students, as well as student performance at the classroom, building, and district levels.

Assessment Terms and Types of Scores

Familiarity with the testing terms and the types of scores used in the MAP reports and other components will help you interpret test information accurately and efficiently.

MAP Scale Score

CTB uses the student's correct responses to derive a MAP scale score. The scale score describes achievement on a continuum that in most cases spans the complete range of grades 3–8. These scores range in value from 2100 to 2800 for English Language Arts and Mathematics and from 470 to 895 for Science. Within a content area, scores from adjacent grades may be compared. Scale scores cannot be compared across content areas. For example, it is appropriate to compare a student's grade 5 Mathematics scale score with his or her grade 6 Mathematics scale score. The MAP scale score determines the student's achievement level. The MAP scale score ranges for each achievement level can be found beginning on page 4 of this guide. Within a content area, scale scores can be added, subtracted, and averaged. A student receives a MAP scale score when he or she makes a valid attempt in any content area.

Achievement Levels

Student performance can be reported in terms of four performance, or achievement, levels that describe a pathway to proficiency and college and career readiness. Each achievement level represents standards of performance for each assessed content area (English Language Arts, Mathematics, and Science). Panels drawn from educational, business, and professional communities determined the achievement standards. Achievement-level scores provide a description of what students can do in terms of the content and skills assessed, as described in the Missouri Learning Standards.

Claim-Level Performance

In English Language Arts and Mathematics, student performance can also be reported at the claim level. Each claim is an evidence-based statement about what students know and can do, as demonstrated by their performance on the assessments. Claim performance levels are reported in terms of three levels of proficiency: below, at/near, and above.

Lowest Obtainable Scale Score and Level Not Determined

Within each grade level and content area, a Lowest Obtainable Scale Score (LOSS) is established for students whose scores are below the level expected by guessing. Students with certain accommodations that impact the construct being assessed (e.g., read-aloud of ELA passages for students in grades 3–5) also receive a LOSS.

A student may receive "Level Not Determined" (LND) instead of a MAP scale score. Students that receive LND are not assigned to an achievement level. Students may receive LND for either of the following reasons:

- A student does not attempt any items in one or more content areas of the MAP test.
- A student is absent for all testing sessions for a particular content area.

Standard Error of Measurement

No test provides a perfect measure of a student's ability. This situation is expected because all tests have a known Standard Error of Measurement (SEM). The SEM reports the amount of variability that can be expected in a student's test score due to the inherent imprecision of the test. The SEM for the MAP test will be reported in the 2015 MAP Technical Report.

Achievement-Level Descriptors

English Language Arts, Reporting Achievement-Level Descriptors

Grades 3–5

Below Basic (Level 1)

The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

MAP score ranges:

Grade 3: Below 2367
Grade 4: Below 2416
Grade 5: Below 2442

Basic (Level 2)

The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

MAP score ranges:

Grade 3: 2367–2431
Grade 4: 2416–2472
Grade 5: 2442–2501

Proficient (Level 3)

The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

MAP score ranges:

Grade 3: 2432–2489
Grade 4: 2473–2532
Grade 5: 2502–2581

Advanced (Level 4)

The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

MAP score ranges:

Grade 3: 2490+
Grade 4: 2533+
Grade 5: 2582+

Grades 6–8

Below Basic (Level 1)

The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

Grade 6: Below 2457
Grade 7: Below 2479
Grade 8: Below 2487

Basic (Level 2)

The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

Grade 6: 2457–2530
Grade 7: 2479–2551
Grade 8: 2487–2566

Proficient (Level 3)

The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

Grade 6: 2531–2617
Grade 7: 2552–2648
Grade 8: 2567–2667

Advanced (Level 4)

The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

Grade 6: 2618+
Grade 7: 2649+
Grade 8: 2668+

Mathematics, Reporting Achievement-Level Descriptors

Grades 3–5

Below Basic (Level 1)

The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in future coursework.

MAP score ranges:

- Grade 3: Below 2381
- Grade 4: Below 2411
- Grade 5: Below 2455

Basic (Level 2)

The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in mathematics needed for likely success in future coursework.

MAP score ranges:

- Grade 3: 2381–2435
- Grade 4: 2411–2484
- Grade 5: 2455–2527

Proficient (Level 3)

The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in mathematics needed for likely success in future coursework.

MAP score ranges:

- Grade 3: 2436–2500
- Grade 4: 2485–2548
- Grade 5: 2528–2578

Advanced (Level 4)

The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in mathematics needed for likely success in future coursework.

MAP score ranges:

- Grade 3: 2501+
- Grade 4: 2549+
- Grade 5: 2579+

Grades 6–8

Below Basic (Level 1)

The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

- Grade 6: Below 2473
- Grade 7: Below 2484
- Grade 8: Below 2504

Basic (Level 2)

The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

- Grade 6: 2473–2551
- Grade 7: 2484–2566
- Grade 8: 2504–2585

Proficient (Level 3)

The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

- Grade 6: 2552–2609
- Grade 7: 2567–2634
- Grade 8: 2586–2652

Advanced (Level 4)

The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.

MAP score ranges:

- Grade 6: 2610+
- Grade 7: 2635+
- Grade 8: 2653+

Science, Abbreviated Achievement-Level Descriptors

Grade 5

Below Basic (Level 1)

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.

MAP score range: 470–625

Basic (Level 2)

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; determine the appropriate scientific tool and its function in an investigation; determine how technological advances address problems and enhance life.

MAP score range: 626–668

Proficient (Level 3)

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; interpret data; distinguish between man-made and natural objects; apply problem solving skills to a situation.

MAP score range: 669–691

Advanced (Level 4)

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 and tools in an investigation.

MAP score range: 692–855

Grade 8

Below Basic (Level 1)

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.

MAP score range: 540–670

Basic (Level 2)

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.

MAP score range: 671–702

Proficient (Level 3)

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; use appropriate tools and methods to collect data; describe tools and discoveries that advance scientific knowledge.

MAP score range: 703–734

Advanced (Level 4)

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; apply awareness of the influence of science and technology in society.

MAP score range: 735–895

Sample Reports

Individual Student Report

The Individual Student Report provides information about performance on the MAP Grade-Level Assessments, describing results in terms of four levels of achievement in a content area. For English Language Arts and Mathematics, a student's strength or weakness at the claim level is also reported. This information may be used for instructional planning, as a point of reference during a parent/teacher conference, and for permanent record keeping. Other sources of information, such as classroom performance, should be used along with this report when determining the student's areas of strength or need.

Achievement-level scores describe what students can do in terms of the content and skills assessed by the MAP. Because the English Language Arts and Mathematics Missouri Learning Standards are grounded in expectations for college and career readiness, the MAP Grade-Level Assessments are designed to measure each student's progress toward meeting those expectations. Teachers, students, and parents/guardians can use this information in addition to how the student performs in the classroom to determine what skills and abilities need to be acquired to enable the student to progress to higher achievement levels. A student in the Proficient or Advanced level has met the standard. Students in the Below Basic and Basic levels have typically mastered skills described for their levels on pages 4–6, but need to work on skills in higher levels.

The following pages contain two sample Individual Student Reports—one for English Language Arts and one for Science. Individual Student Reports for Mathematics are similar to those for English Language Arts, so a sample is not provided.

Ⓐ Student Report for:

This area of the report is reserved for the name and biographical data of the student taking the assessment.

Ⓑ How did your child perform?

This is your child's scale score. The scale score is also printed in the left column under "Overview of Performance."

Ⓒ Your child's achievement level is Proficient.

Achievement levels (whether Advanced, Proficient, Basic, or Below Basic) are based on the test score ranges listed beneath each achievement level shown in the right column.

Ⓓ Overview of Performance

The **Scale Score** is derived from student responses to assessment items. It summarizes the overall level of performance attained by your child for a particular content area.

Ⓔ Claim Information (for ELA and Mathematics only)

English Language Arts and Mathematics are comprised of claims. Within each claim, student performance is reported as "Below Standard," "At/Near Standard," or "Above Standard." A description of each claim is provided.

Missouri Grade-Level Assessment Program

2015 Individual Student Report - *Final*

A Student Report for:

Name **WEBBER, PEGGY**

Student ID **1234567890**

Birthdate **MM/DD/YYYY**

Grade **Grade 3**

School **ADAIR CO. ELEM.**

District **ADAIR CO.**

B How did your child perform in English/Language Arts?

Your child received a score of 2452.

One way to measure performance is by achievement levels, which are based on scale scores. Achievement levels describe what your child's score means.

C Your child's achievement level is Proficient.

What does a level of "Proficient" mean?
The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

Advanced
Scores of 2490 and above demonstrate a thorough understanding of the content at this grade level.

Proficient
Scores at 2432-2489 demonstrate an understanding of the content expected at this grade level.

Basic
Scores at 2367-2431 demonstrate a partial understanding of the content expected at this grade level.

Below Basic
Scores at 2366 and below do not demonstrate an understanding of the content expected at this grade level.

D Overview of Performance

Scale Score: 2452

This report provides information about achievement on the Missouri Assessment Program (MAP).

It is the policy of the Missouri Department of Elementary and Secondary Education not to discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs or employment practices as required by Title VI and VII of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination Act of 1975 and title II of the Americans with Disabilities Act of 1990. Inquiries related to the Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator-Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 526-4757 or TTY (800) 735-2966, fax (573) 522-4883, email civilrights@dese.mo.gov.

A single exam can provide only limited information. You should confirm your child's strengths and needs in these topics by reviewing classroom work, standards-based assessments,and your child's progress reports during the year.

For more resources, go to
http://dese.mo.gov/

E

Reading	Writing	Listening	Research
Above Standard	At/Near Standard	Above Standard	Above Standard
Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.	Students can produce effective and well-grounded writing for a range of purposes and audiences.	Students can employ effective speaking and listening skills for a range of purposes and audiences.	Students can engage in research and inquiry to investigate topics, and to analyze, integrate, and present information.

Mc Graw Hill Education | CTB

03/19/2015

Missouri Grade-Level Assessment Program

2015 Individual Student Report - *Final*

A Student Report for:

Name	**WEBBER, PEGGY**
Student ID	**1234567890**
Birthdate	**MM/DD/YYYY**
Grade	**Grade 5**
School	**ADAIR CO. HIGH**
District	**ADAIR CO.**

D Overview of Performance

Scale Score: 679

This report provides information about achievement on the Missouri Assessment Program (MAP).

It is the policy of the Missouri Department of Elementary and Secondary Education not to discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs or employment practices as required by Title VI and VII of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination Act of 1975 and title II of the Americans with Disabilities Act of 1990. Inquiries related to the Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator-Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 526-4757 or TTY (800) 735-2966, fax (573) 522-4883, email civilrights@dese.mo.gov.

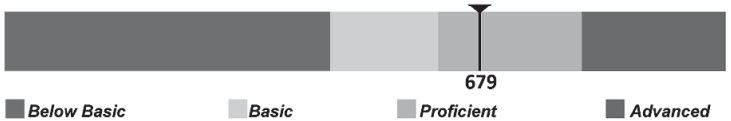
A single exam can provide only limited information. You should confirm your child's strengths and needs in these topics by reviewing classroom work, standards-based assessments,and your child's progress reports during the year.

For more resources, go to
http://dese.mo.gov/

B How did your child perform in Science?
Your child received a score of 679.

One way to measure performance is by achievement levels, which are based on scale scores. Achievement levels describe what your child's score means.

C Your child's achievement level is Proficient.



What does a level of "Proficient" mean?

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; interpret data; distinguish between man-made and natural objects; apply problem solving skills to a situation.

Advanced
Scores of 692 and above demonstrate a thorough understanding of the content at this grade level.

Proficient
Scores at 669–691 demonstrate an understanding of the content expected at this grade level.

Basic
Scores at 626–668 demonstrate a partial understanding of the content expected at this grade level.

Below Basic
Scores at 625 and below do not demonstrate an understanding of the content expected at this grade level.

Mc Graw Hill Education | CTB

06/15/2015

Student Label

MISSOURI ASSESSMENT PROGRAM	Content Area	English Language Arts
Dalbey, Kori A Ⓐ	Achievement Level	**Proficient** Ⓑ
Grade: **8** Test Date: **MM/DD/YY** DOB: **MM/DD/YY** MOSIS State ID: **0123456789**	MAP Scale Score	**2596** Ⓒ

Above is a sample of the MAP student label. The student 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 2015 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—one for each of the content areas administered within a grade.

Ⓐ The left side of the label lists the name and biographical data of the student taking the assessment.

Ⓑ This is the student's Achievement Level (Advanced, Proficient, Basic, or Below Basic).

Ⓒ This is the student's Scale Score for the content area listed at the top of the label.

Notice of Non-Discrimination

It is the policy of the Missouri Department of Elementary and Secondary Education not to discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs or employment practices as required by Title VI and VII of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination Act of 1975, and Title II of the Americans with Disabilities Act of 1990.

Inquiries related to Department employment practices may be directed to the Jefferson State Office Building, Human Resources Director, 8th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 751-9619 or TYY (800) 735-2966. Inquiries related to Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator–Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 526-4757 or TYY (800) 735-2966, email civilrights@dese.mo.gov.

Anyone attending a meeting of the State Board of Education who requires auxiliary aids or services should request such services by contacting the Executive Assistant to the State Board of Education, Jefferson State Office Building, 205 Jefferson Street, Jefferson City, MO 65102-0480; telephone number (573) 751-4446 or TTY (800) 735-2966.

Inquiries or concerns regarding civil rights compliance by school districts or charter schools should be directed to the local school district or charter school Title IX/non-discrimination coordinator. Inquiries and complaints may also be directed to the Office for Civil Rights, Kansas City Office, U.S. Department of Education, 8930 Ward Parkway, Suite 2037, Kansas City, MO 64114; telephone number (816) 268-0550; FAX (816) 823-1404; TDD (877) 521-2172.

CTB

20 Ryan Ranch Road
Monterey, CA 93940−5703

Appendix C
Science MAP Sample Reports and General Research File Layout

Figure C1. Student Roster

Missouri Grade-Level Assessment Program

2015 Individual Student Report - Final

Student Report for:

Name	WEBBER, PEGGY
Student ID	1234567890
Birthdate	MM/DD/YYYY
Grade	Grade 8
School	SALEM MIDDLE
District	DIST ABC

Overview of Performance

Scale Score: 704

This report provides information about achievement on the Missouri Assessment Program (MAP).

It is the policy of the Missouri Department of Elementary and Secondary Education not to discriminate on the basis of race, color, religion, gender, national origin, age, or disability in its programs or employment practices as required by Title VI and VII of the Civil Rights Act of 1964, Title IX of the Education Amendments of 1972, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination Act of 1975 and title II of the Americans with Disabilities Act of 1990. Inquiries related to the Department programs and to the location of services, activities, and facilities that are accessible by persons with disabilities may be directed to the Jefferson State Office Building, Office of the General Counsel, Coordinator-Civil Rights Compliance (Title VI/Title IX/504/ADA/Age Act), 6th Floor, 205 Jefferson Street, P.O. Box 480, Jefferson City, MO 65102-0480; telephone number (573) 526-4757 or TTY (800) 735-2966, fax (573) 522-4883, email civilrights@dese.mo.gov.

A single exam can provide only limited information. You should confirm your child's strengths and needs in these topics by reviewing classroom work, standards-based assessments,and your child's progress reports during the year.

For more resources, go to

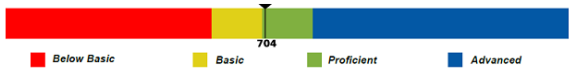
http://dese.mo.gov/

How did your child perform in Science?

Your child received a score of 704.

One way to measure performance is by achievement levels, which are based on scale scores. Achievement levels describe what your child's score means.

Your child's achievement level is Proficient.



What does a level of "Proficient" mean?

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; use appropriate tools and methods to collect data; describe tools and discoveries that advance scientific knowledge.

Advanced
Scores of 735 and above demonstrate a thorough understanding of the content at this grade level.

Proficient
Scores at 703-734 demonstrate an understanding of the content expected at this grade level.

Basic
Scores at 671-702 demonstrate a partial understanding of the content expected at this grade level.

Below Basic
Scores at 670 and below do not demonstrate an understanding of the content expected at this grade level.

01/06/2016

Figure C2. Individual Student Report

Figure C3. Student Label

MISSOURI
ASSESSMENT PROGRAM

WEBBER, PEGGY

Grade: **Grade 8**
Test Date: **MM/DD/YY**
DOB: **MM/DD/YY**
MOSIS State ID: **0123456789**

Content Area	English Language Arts
Achievement Level	**Advanced**
MAP Scale Score	**2731**

MISSOURI
ASSESSMENT PROGRAM

WEBBER, PEGGY

Grade: **Grade 8**
Test Date: **MM/DD/YY**
DOB: **MM/DD/YY**
MOSIS State ID: **0123456789**

Content Area	Science
Achievement Level	**Below Basic**
MAP Scale Score	**777**

MISSOURI
ASSESSMENT PROGRAM

WEBBER, PEGGY

Grade: **Grade 8**
Test Date: **MM/DD/YY**
DOB: **MM/DD/YY**
MOSIS State ID: **0123456789**

Content Area	Mathematics
Achievement Level	**Level Not Determined**
MAP Scale Score	

MISSOURI
ASSESSMENT PROGRAM

Grade:
Test Date:
DOB:
MOSIS State ID:

Content Area
Achievement Level
MAP Scale Score

MO MAP Spring 2015 GRF - *Final*

Start	End	Len	Field	Values	Remarks
		0	Hierarchical Information		
1	2	2	StateCode	MO	Missouri
3	15	13	Test type	MAPSMTSPR2015	Represents the assessment the student will be taking. MAPSMTSPR2015 is MAP Grade-Level Assessment spring Summative.
16	43	28	District Name		Data will come from Node file, not Precode file.
44	49	6	DistCd	0-9, DESE-Assigned Unique 6 digit code representing the district in which the school is located.	Reporting District Code
50	77	28	School Name		Data will come from Node file, not Precode file.
78	81	4	ScCd	0-9; Four digit school code	Reporting School Code
82	83	2	Grade	Grade level as of the time data is being submitted unless otherwise specified. 03 - 08, Must include 2 characters (leading zeros when applicable).	Grade
84	113	30	City	Any character, blank	This is the "SHIP_TO_CITY" in Node file.
114	117	4	Current School Year	YYYY	2015
118	167	50	Blank for Future Use	Filler	Reserved for future use
			Student BIO Information		
168	179	12	CTB Use Only (Unique System-generated ID)	0-11	Unique ID CTB Specific
180	189	10	DRC Student ID (Unique System-generated)	0-9	Unique record number per student assessment instance. This is an INSIGHT internal number that uniquely identifies each student test across multiple administrations in a given contract year.
190	209	20	Local Student ID	0-9	
210	219	10	StateID	0-9; blank State assigned student identifier. See http://dese.mo.gov/MOSIS for more information.	MOSIS ID
220	279	60	Last Name	Legal last name. Supported Characters : A-Z, a-z, 0-9, "-", "_", "'", ".", space	Student Last Name
280	339	60	First Name	Legal first name Supported Characters : A-Z, a-z, 0-9, "-", "_", "'", ".", space	Student First Name

Start	End	Len	Field	Values	Remarks
340	399	60	Middle Name	Legal middle name Supported Characters : A-Z, a-z, 0-9, "-", "_", "'", ".", space	Student Middle Name
400	409	10	Suffix	Legal name suffix. E.g. Jr, Sr Supported Characters : A-Z, a-z, 0-9, "-", "_", "'", ".", space	
410	419	10	Birth Date	mmddyyyy Month = Jan = '01', Feb = '02', Mar = '03', Apr = '04', May = '05', June = '06', July = '07', Aug = '08', Sept = '09', Oct = '10',Nov = '11', Dec = '12' Day = 01 to 31 Year = Each position: 0-9	
420	420	1	Gender	F = Female, M = Male	
421	421	1	RaceEthnicity	A = Asian B = Black H = Hispanic I = Indian (American Indian or Alaskan Native) W = White P = Native Hawaiian or Other Pacific Islander M = Multi-Racial	
422	422	1	Blank	Filler	Data in this field is blanked-out and not included in the GRF file to DESE
423	423	1	Blank	Filler	Data in this field is blanked-out and not included in the GRF file to DESE
424	424	1	State Use 1	Y if marked, Blank if not marked.	
425	425	1	State Use 2	Y if marked, Blank if not marked.	
426	426	1	State Use 3	Y if marked, Blank if not marked.	
427	427	1	State Use 4	Y if marked, Blank if not marked.	
428	428	1	State Use 5	Y if marked, Blank if not marked.	
429	429	1	State Use 6	Y if marked, Blank if not marked.	
430	430	1	State Use 7	Y if marked, Blank if not marked.	
431	431	1	State Use 8	Y if marked, Blank if not marked.	
432	432	1	State Use 9	Y if marked, Blank if not marked.	
433	433	1	State Use 10	Y if marked, Blank if not marked.	
434	435	2	Period	01-10, blank	(Because Fall spans two calendar years, the YY of the two years are used - example Fall 15-16 and Spring 2015).
			CONTENT AREA INFORMATON		

Start	End	Len	Field	Values	Remarks
436	437	2	Content Code	01 = English Language Arts, 02 = Mathematics, 03 = Science	
438	440	3	Content Form	CA1,CA2,CA3,CS1,CT1,CT2, Blank	Form is at the Content level. First character: C= core Second character: A=regular form, S=Spanish, T=Transcription Third character: Numeric identifier
441	443	3	PT Content Form	PA1,PA2, PA3,PS1,PT1, Blank	Form is at the Content level. First character: P= performance task Second character: A=regular form, S=Spanish, T=Transcription Third character: Numeric identifier
444	503	60	EducatorFirstName	Supported Characters : A-Z, a-z, 0-9, "-", "_", " ' ", ".", space	From field in Precode File "Educator" is also known as "Examiner".
504	563	60	EducatorLastName	Supported Characters : A-Z, a-z, 0-9, "-", "_", " ' ", ".", space	From field in Precode File "Educator" is also known as "Examiner".
564	633	70	ExaminerEmail	Up to 50 characters, valid email format (xxx@xxx.xxx), Blank if no email included.	From Precode File. Examiner email is the same as Educator's email.
634	648	15	Content Export Date Time	YYYYMMDDHHMMSS (24 Hr Time Frame in GMT Format)	(24 HOUR FORMAT - IF MORE THAN ONE REORD IN OUR PROCESSING WE KEEP THE LATEST RECEIVED RECORD) NOTE: The first 14 positions will have the DateTime stamp, starting at position 634, and the very last position in the field (15) is blank.
649	658	10	Test Date (MMDDCCYY)	MMDDCCYY	This date is the generic first date of the testing window.
659	659	1	Precode Flag	Y = yes, blank = no; set to Y if student was in a Precode file, if blank student was manually entered.	Only set if the student comes in the precode file.
660	709	50	Blank for Future Use	Filler	Reserved for future use
			Accommodations		
710	710	1	Accommodation Braille edition (A012)	Blank = Not Indicated Y = Indicated	Code A012 refers to Braille administered via Paper format

Start	End	Len	Field	Values	Remarks
711	711	1	Accommodation Large Print edition (A021)	Blank = Not Indicated Y = Indicated	Paper format only
712	712	1	Accommodation Signing of assessment (ASL) for Math and Science and ELA Listening items (A051)	Blank = Not Indicated Y = Indicated	Online
713	713	1	Accommodation Signing of assessment (ASL) for Math and Science and ELA Listening items (A052)	Blank = Not Indicated Y = Indicated	Any format (online or paper)
714	714	1	Accommodation Paper Based Assessment (A102)	Blank = Not Indicated Y = Indicated	Any format
715	715	1	Accommodation Use of Specialized Calculator on items where calculator is allowed, includes talking calculators or Braille calculators. (A396)	Blank = Not Indicated Y = Indicated	Any format
716	716	1	Accommodation Alternate Response Options (such as adapted keyboards, stickykeys, touch screens.) (A441)	Blank = Not Indicated Y = Indicated	Any format
717	717	1	Accommodation Use of Bilingual Dictionary (S431)	Blank= Not Indicated Y = Indicated	
718	718	1	Support Color Contrasting Text (S101)	Blank = Not Indicated Y = Indicated	S101 = Online
719	719	1	Support Color Chooser (S101)	Blank = Not Indicated Y = Indicated	
720	720	1	Support Color Contrast (S102)	Blank = Not Indicated Y = Indicated	S102 = Paper
721	721	1	Accommodation S103 Color Overlay	Blank = Not Indicated Y = Indicated	Paper
722	722	1	Accommodation S105 Magnification	Blank = Not Indicated Y = Indicated	
723	723	1	Accommodation Masking (S106)	Blank = Not Indicated Y = Indicated	S106=Online
724	724	1	Accommodation Masking (S107)	Blank = Not Indicated Y = Indicated	S107=Paper
725	725	1	Accommodation Read-Aloud (For all items in any subject, excluding ELA Reading passages) (S041)	Blank = Not Indicated Y = Indicated	(EXCLUDING ELA READING PASSAGES) S041=Online; The system allows items in mathematics and English Language Arts to be read aloud to the student via embedded text-to-speech technology. The student can control the speed and volume of the voice.

Start	End	Len	Field	Values	Remarks
726	726	1	Accommodation Read-Aloud (For all items in any subject, excluding ELA Reading passages) (S042)	Blank = Not Indicated Y = Indicated	S042= Online- Not Embedded; Student may use assistive technology text-to-speech software to allow all items in any subject, not including ELA reading passage, to be read aloud.
727	727	1	Accommodation Read-Aloud (For all items in any subject, excluding ELA Reading passages) (S111)	Blank = Not Indicated Y = Indicated	S111=Any; Students may have items in mathematics, science, and English language arts to be read aloud to them in their native language by a trained translator. Read Aloud of ELA reading passages requires an IEP or 504 plan
728	728	1	Accommodation Scribe (for all items in any subject, excluding ELA Writing) (S351)	Blank = Not Indicated Y = Indicated	(EXCLUDING ELA WRITING) Any format
729	729	1	Accommodation Separate Setting (S501)	Blank = Not Indicated Y = Indicated	
730	730	1	Support Translated Test Directions (S109)	Blank = Not Indicated Y = Indicated	S109 Any format
731	731	1	Support Translation (Glossary) (S109) Accommodation Translation (Stacked) (S108)	Blank = Not Indicated Y = Indicated	S109 Any format
732	732	1	Support Glossary (Grades 3-8 Math and ELA Only) (S104)	Blank = Not Indicated Y = Indicated	Online - Not Embedded All students taking the paper based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment.
733	733	1	Accommodation Read-Aloud ELA Reading passages (GRADES 3-5 ONLY) (A041) – invalidates ELA	Blank = Not Indicated Y = Indicated	If marked "Y" then ELA will be invalidated. Any format. Students in grades 3-5 may have English Language Arts reading passages read aloud to them by a trained reader. Note: Use of this will result in invalidation - student will receive the lowest obtainable scale score (LOSS).
734	734	1	Accommodation Read-Aloud ELA Reading passages (GRADES 3-5 ONLY) (A042)– invalidates ELA	Blank = Not Indicated Y = Indicated	Online - Not Embedded - Students in grades 3-5 may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud. Note: Use of this will result in invalidation - student will receive the lowest obtainable scale score (LOSS).

Start	End	Len	Field	Values	Remarks
735	735	1	Accommodation Read-Aloud ELA Reading passages (GRADES 3-5 ONLY) (A111)– invalidates ELA	Blank = Not Indicated Y = Indicated	ELL students in grades 3-5 may have English Language Arts reading passagges read aloud to them in their native language by a trained translator. Note: Use of this will result in invalidation - student will receive the lowest obtainable scale score (LOSS).
736	736	1	Accommodation Read-Aloud (For all items in any subject, excluding ELA Reading passages) (S043)	Blank = Not Indicated Y = Indicated	S043=Any; Students may have items in mathematics, science, and English language arts to be read aloud to them by a trained reader. Read Aoud of ELA reading passages requires an IEP or 504 plan.
737	737	1	Accommodation Read-Aloud ELA Reading passages (GRADES 6-8 only) (A044)	Blank = Not Indicated Y = Indicated	A044=Online - Not Embedded; Students may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud.
738	738	1	Accommodation Read-Aloud ELA Reading passages (GRADES 6-8 only) (A045)	Blank = Not Indicated Y = Indicated	A045=Any; Students may have English Language Arts reading passages to be read aloud to them by a trained reader.
739	739	1	Accommodation Read-Aloud ELA Reading passages (GRADES 6-8 only) (A112)	Blank = Not Indicated Y = Indicated	A112=Any ELL students may have English language arts reading passages to be read aloud to them in their native language by a trained translator.
740	740	1	Accommodation A046 Read-Aloud ELA Reading passages (Blind students)	Blank = Not Indicated Y = Indicated	This accommodation is used for Blind students who do not yet have adequate Braille skills. Note this does not invalidate the test.
741	741	1	Accommodation Use of scribe - ELA writing (A351)	Blank = Not Indicated Y = Indicated	Any format
742	742	1	Accommodation Abacus (A391)	Blank = Not Indicated Y = Indicated	
743	743	1	Accommodation Calculator GRADE 3 ONLY (for Non-Calculator Allowed Items Only)– invalidation (A392)	Blank = Not Indicated Y = Indicated	If marked "Y" then Mathematics will be invalidated.
744	744	1	Accommodation Calculator GRADE 4-8 ONLY (for Non-Calculator Allowed Items Only) (A393)	Blank = Not Indicated Y = Indicated	
745	745	1	Accommodation Multiplication Table GRADE 3 ONLY - invalidation (A394)	Blank = Not Indicated Y = Indicated	If marked "Y" then Mathematics will be invalidated.
746	746	1	Accommodation Multiplication Table GRADES 4-8 (A395)	Blank = Not Indicated Y = Indicated	
747	747	1	Non-Accommodation Special Case – Paper Based Assessment	Blank = Not Indicated Y = Indicated	This is for those students who are taking the paper-based assessment but don't have an IEP. They will need to pay for it. This is available for all subjects.

Start	End	Len	Field	Values	Remarks
748	767	20	Blank for Future Use	Filler	
			Teacher Invalidations		"Teacher Invalidations" are populated from eDIRECT
768	768	1	Teacher Invalidation	Blank = No Invalidation marked Y = Invalidated	When invalidation is marked this will invalidate the content area for all sessions.
			Absent		Absent populated from eDIRECT
769	769	1	Absent	Blank = Not Marked, Y = Marked	
			Not Enrolled for Content Area		Not Enrolled populated from eDIRECT
770	819	50	Blank for Future Use	Filler	Blank for future use
			Item Responses and Score Data		
820	840	21	Content Area Title	"English Language Arts", "Mathematics", "Science", blank	
841	940	100	Item Responses for MC	Correct = A, B, C, D (*letter*); Incorrect = 1, 2, 3, 4 (*numeric*); Omit = 0; marked 'Absent'.	Showing up to 100 MC items.
941	1040	100	Item Scores for MC	1 = correct response; 0 = wrong or omitted; blank (blank = field test item when applicable). In all of the following cases: 1) The student started a test, but it was marked with an applicable invalidating Accommodation code 2) The student test was marked 'Teacher Invalidation' (started test or not) 3) The student did not start a test, but was marked 'Absent' then this field is NULL.	
1041	1043	3	Raw Score for MC Items	0-100, blank	0-NNN where NNN is the maximum score of MC; In cases where the student attempted a test, but it was marked with an applicable invalidating Accommodation code the Raw Score is 0 (zero). In cases where the student test was marked 'Teacher Invalidation' (started test or not) the Raw Score is 0 (zero). In cases where the student did not start a test, but was marked 'Absent' the Raw Score is NULL (blanks).
		100	Item Scores CR scores (includes WER)		Total field positions 1044-1143 (100 bytes)

Start	End	Len	Field	Values	Remarks
1044	1123	80	Item Scores CR scores	0-9, A,B,C,D,E Blank, (Trailing blank)	Resolved Final CR score 1 byte (single character) each in positions 1-80 of the overall 100 byte field. For data in the remaining positions 81-100 (see row below).
1124	1143	20	Item Scores CR scores - WER (long write) items	0-9, A,B,C,D,E Blank, (Trailing blank)	This 20 byte section is still part of the 100 byte string of the "Item Scores CR scores" field above and is only separated out for clarity of reading the file. In positions 81 & 82 for ELA Grades 5 & 8 ONLY it will contain the 1 byte 2-point WER (long write) trait score values for 'Conventions', followed by the 1 byte average score of the two 4-point WER trait items, 'Purpose/Organization' and 'Evidence/Elaboration'. This 20 byte section will be blank for all other content areas and grades.
1144	1146	3	Raw Score for CR Items		Total Raw Score for CR Items
1147	1196	50	Item score for TE	0-9, A,B,C,D,E Blank, (Trailing blank)	Resolved Final TE score 1 byte (single character) each.
1197	1199	3	Raw Score for TE Items		
1200	1202	3	Total Raw Score MC, CR, TE	0-100, blank	Raw Score for MC + Raw Score of CR + Raw Score for TE
1203	1203	1	Completion Criteria	N, Y	Y = met, N = not met (Completion Criteria Met = Valid Attempt)
1204	1207	4	Percent Correct	0.0 - 100, blank	Whole Number plus one decimal position (99.9). Total Raw Score / Total Raw Score Possible. Report to the tenth. No decimal point for 100 Rounding rules: we will round up from anything half-way between tenths (for example, 22.25 rounds to 22.3)
1208	1211	4	Scale Score	0000-9999, blank	
1212	1212	1	Content Achievement Level	Values 0, 1-4, Blank Blank = Content Area Not Applicable 0 = if Absent or No Valid Attempt (Level Not Determined) 1 = Below Basic 2 = Basic 3 = Proficient 4 = Advanced.	In cases where the student started a test, but it was marked with an applicable invalidating Accommodation code the Achievement Level is Below Basic (1). In cases where the student test was marked 'Teacher Invalidation' (started test or not) the Achievement Level is Below Basic (1). In cases where the student did not start a test, but was marked 'Absent' the Achievement Level is 0.
1213	1242	30	Blank for Future Use	Filler	Reserved for future use

Start	End	Len	Field	Values	Remarks
1243	1243	1	Objective 1 ELA - Code	1-Reading	Objective score = Claim-level score
1244	1246	3	Objective 1 ELA - Number Correct	0-100, Blank	
1247	1250	4	Objective 1 ELA - Percent Correct	0.0 - 100, blank	Whole Number plus one decimal position (99.9). Total Raw Score / Total Raw Score Possible. Report to the tenth. No decimal point for 100 Rounding rules: we will round up from anything half-way between tenths (for example, 22.25 rounds to 22.3)
1251	1251	1	Objective 1 ELA - Achievement Level	0-3, Blank Blank = Content Area Not Applicable 0 - if Absent or No Valid Attempt (Level Not Determined) 1 - Below standard 2- At/Near standard 3 - Above standard	Smarter Balanced Objective Achievement levels (NOT APPLICABLE TO SCIENCE - blank for Science) In cases where the student started a test, but it was marked with an applicable invalidating Accommodation code the Achievement Level is Below Basic (1). In cases where the student test was marked 'Teacher Invalidation' (started test or not) the Achievement Level is Below Basic (1). In cases where the student did not start a test, but was marked 'Absent' the Achievement Level is 0.
1252	1252	1	Objective 2 ELA - Code	2-Writing	Objective score = Claim-level score
1253	1255	3	Objective 2 ELA - Number Correct	0-100, Blank	
1256	1259	4	Objective 2 ELA - Percent Correct	0.0 - 100, blank	
1260	1260	1	Objective 2 ELA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1261	1261	1	Objective 3 ELA - Code	3-Listening	Objective score = Claim-level score
1262	1264	3	Objective 3 ELA - Number Correct	0-100, Blank	
1265	1268	4	Objective 3 ELA - Percent Correct	0.0 - 100, blank	
1269	1269	1	Objective 3 ELA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1270	1270	1	Objective 4 ELA - Code	4-Research	Objective score = Claim-level score
1271	1273	3	Objective 4 ELA - Number Correct	0-100, Blank	
1274	1277	4	Objective 4 ELA - Percent Correct	0.0 - 100, blank	
1278	1278	1	Objective 4 ELA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1279	1279	1	Objective 1 MA - Code	1-Concepts and Procedures	Objective score = Claim-level score
1280	1282	3	Objective 1 MA - Number Correct	0-100, Blank	
1283	1286	4	Objective 1 MA - Percent Correct	0.0 - 100, blank	
1287	1287	1	Objective 1 MA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1288	1288	1	Objective 2 MA - Code	2-Problem Solving and Modeling and Data Analysis	Objective score = Claim-level score
1289	1291	3	Objective 2 MA - Number Correct	0-100, Blank	
1292	1295	4	Objective 2 MA - Percent Correct	0.0 - 100, blank	

Start	End	Len	Field	Values	Remarks
1296	1296	1	Objective 2 MA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1297	1297	1	Objective 3 MA - Code	3-Communicating Resasoning	Objective score = Claim-level score
1298	1300	3	Objective 3 MA - Number Correct	0-100, Blank	
1301	1304	4	Objective 3 MA - Percent Correct	0.0 - 100, blank	
1305	1305	1	Objective 3 MA - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1306	1306	1	Objective 1 SC - Code	1-Matter and Energy	Objective score = GLE Strand score
1307	1309	3	Objective 1 SC - Number Correct	0-100, Blank	
1310	1313	4	Objective 1 SC - Percent Correct	0.0 - 100, blank	
1314	1314	1	Objective 1 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1315	1315	1	Objective 2 SC - Code	2-Force and Motion	Objective score = GLE Strand score
1316	1318	3	Objective 2 SC - Number Correct	0-100, Blank	
1319	1322	4	Objective 2 SC - Percent Correct	0.0 - 100, blank	
1323	1323	1	Objective 2 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1324	1324	1	Objective 3 SC - Code	3-Living Organisms	Objective score = GLE Strand score
1325	1327	3	Objective 3 SC - Number Correct	0-100, Blank	
1328	1331	4	Objective 3 SC - Percent Correct	0.0 - 100, blank	
1332	1332	1	Objective 3 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1333	1333	1	Objective 4 SC - Code	4-Ecosystems	Objective score = GLE Strand score
1334	1336	3	Objective 4 SC - Number Correct	0-100, Blank	
1337	1340	4	Objective 4 SC - Percent Correct	0.0 - 100, blank	
1341	1341	1	Objective 4 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1342	1342	1	Objective 5 SC - Code	5-Earth's Systems	Objective score = GLE Strand score
1343	1345	3	Objective 5 SC - Number Correct	0-100, Blank	
1346	1349	4	Objective 5 SC - Percent Correct	0.0 - 100, blank	
1350	1350	1	Objective 5 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1351	1351	1	Objective 6 SC - Code	6-The Universe	Objective score = GLE Strand score
1352	1354	3	Objective 6 SC - Number Correct	0-100, Blank	
1355	1358	4	Objective 6 SC - Percent Correct	0.0 - 100, blank	
1359	1359	1	Objective 6 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1360	1360	1	Objective 7 SC - Code	7-Scientific Inquiry	Objective score = GLE Strand score
1361	1363	3	Objective 7 SC - Number Correct	0-100, Blank	
1364	1367	4	Objective 7 SC - Percent Correct	0.0 - 100, blank	
1368	1368	1	Objective 7 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	

Start	End	Len	Field	Values	Remarks
1369	1369	1	Objective 8 SC - Code	8-Impact of Science, Technology, and Human Activity	Objective score = GLE Strand score
1370	1372	3	Objective 8 SC - Number Correct	0-100, Blank	
1373	1376	4	Objective 8 SC - Percent Correct	0.0 - 100, blank	
1377	1377	1	Objective 8 SC - Achievement Level	0-3, Blank (same as Objective 1 ELA – Achievement Level)	
1378	1378	1	Appeal Indicator	Y = Yes, records in this GRF have a appeal rescore change resulting in a higher score/achievement level; X = No score/achievement level change during the appeal/rescore; Blank = Not an appeal student.	• Y = Yes (Score/Level was changed). Records marked with "Y" in this field have an appeal rescore change resulting in a higher score/achievement level; • X = No (No Score/Level change). Records marked with "X" in this field did not have an appeal rescore change, although they were an appeal student; • Blank = Not an appeal student;
1379	1428	50	Blank for Future Use	Filler	Reserved for future use