Technical Report

Spring 2019

MI-Access

Michigan's Alternate Assessment Program

This document has been formatted to be ADA compliant.

TABLE OF CONTENTS

Executive Summary

MI-Access is Michigan's alternate assessment program for students who have the most significant cognitive disabilities and whose Individualized Education Program (IEP) teams have determined that general assessments, even with accommodations, are not appropriate. MI-Access assessments are based on Michigan's alternate content expectations: Essential Elements with Michigan Range of Complexity for English language arts (ELA) and mathematics, Extended Grade Level Content Expectations for social studies, and Extended Benchmarks for science. These alternate content expectations are aligned to the Michigan K-12 content standards for each content area.

MI-Access is tested at three levels:

- Functional Independence (FI)—for students whose instruction is aligned closest to the "High" range of complexity on the alternate content expectations
- Supported Independence (SI)—for students whose instruction is aligned closest to the "Medium" range of complexity on the alternate content expectations
- Participation (P)—for students whose instruction is aligned closest to the "Low" range of complexity on the alternate content expectations

While the three "levels" of MI-Access are designed for specific populations of students within the universe of students "with significant cognitive disabilities," altogether the levels of MI-Access represent only those Michigan students with the most significant cognitive disabilities whose IEP teams have determined that, based on the students' disabilities, progress toward the general content standards is neither possible nor measurable using M-STEP, the state's standard assessment.

This technical report addresses all phases of the testing cycle with the intention of providing evidence that supports the validity of the MI-Access alternate assessment program. All subsequent chapters of this report constitute evidence for the validity argument that MI-Access was developed with rigor, implemented with fidelity, and validated psychometrically.

E.1 MDE Office of Educational Assessment and Accountability (OEAA)

The Michigan Department of Education (MDE) Office of Educational Assessment and Accountability (OEAA) has the responsibility of carrying out the requirements in state and federal statutes and rules for statewide assessments. The office oversees the planning, scheduling, and implementation of all major assessment activities and supervises MDE's testing contractors (Data Recognition Corporation [DRC] and Measurement Incorporated). In addition, OEAA staff, in collaboration with outside contractors, conducts quality control activities for every aspect of the development and administration of the assessment program. For additional details on these groups, refer to Appendix C of this report. The OEAA also actively monitors the security provisions of the assessment program.

E.2 Michigan Testing Contractors

Data Recognition Corporation is MDE's item development contractor. DRC is responsible for providing test development content leads who work in conjunction with OEAA's content leads to develop test items. MI-Access FI is delivered primarily through DRC's online test engine but also through some paper/pencil testing. DRC test development staff are responsible for rendering test items according to OEAA's style guide. Each item is reviewed by both DRC and OEAA content leads to ensure every student is presented with properly formatted test items that are clear and engaging and to ensure the content of each item replicates how the item appears in the item bank. MI-Access SI and P levels are scored by assessment administrators using a standard rubric, with student scores being entered into a secure DRC online answer portal.

Measurement Incorporated is Michigan's contractor for paper/pencil materials, handscoring, and reporting. Measurement Incorporated is responsible for the development, distribution, and collection of all paper/pencil test materials and for monitoring test security. MI-Access SI and P, FI accommodated testing materials, and the FI Expressing Ideas portion of the FI ELA test are delivered in paper/pencil form. Measurement Incorporated hand scores all the FI Expressing Ideas constructed-response (CR) test questions, using Michigan-provided rubrics. Once testing is complete, Measurement Incorporated is responsible for developing and providing student results.

Assessment and Evaluation Services (AES) contracts for independent third-party validation of psychometric work (see Chapter 7 and Appendix G).

E.3 Michigan's Assessment System

Michigan's assessment system is a comprehensive, standards-based system. All students in grades 3–8 and 11 are required to take Michigan's standards-based accountability assessments. Michigan's accountability assessments are listed in Table E-1 and are described in more detail in section 3.3 of this report.

Table E-1. Michigan's Accountability Assessments

Test	Content	Grades
M-STEP	Mathematics	3–8
M-STEP	ELA	3–8
M-STEP (field testing year)	Science	5, 8, 11
M-STEP	Social Studies	5, 8, 11
SAT	Mathematics	11
SAT	ELA	11
MI-Access (alternate assessment)	Mathematics	3–8, 11
MI-Access (alternate assessment)	ELA	3–8, 11
MI-Access (alternate assessment)	Science	4, 7, 11
MI-Access FI (alternate assessment)	Social Studies	5, 8, 11
WIDA	Listening	1–12
WIDA	Reading	K–12
WIDA	Speaking	K–12
WIDA	Writing	1–12

E.4 Overview of This Report

Subsequent chapters of this technical report document the major activities of the testing cycle. This report provides comprehensive details that confirm that the processes and procedures applied in the MI-Access program adhere to appropriate professional standards and practices of educational assessment. Ultimately, this report serves to document evidence that valid inferences about Michigan student performance can be derived from the MI-Access assessments.

Each chapter of this report details the procedures and processes applied in the MI-Access administration and the results of the administration. Each chapter also highlights the meaning and significance of the procedures, processes, and results in terms of validity and the relationship to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). A brief overview of the contents of this report is described below.

Chapter 1, "Background of Spring 2019 MI-Access," describes the background and history of MI-Access.

Chapter 2, "Uses of Test Scores," describes the use of the assessment scores and touches on the validity arguments this technical report intends to address.

Chapter 3, “Test Design and Item Development,” describes the involvement of Michigan educators in the item and assessment development process, which formed an important part of the validity of MI-Access. The knowledge, expertise, and professional judgment offered by Michigan educators ultimately ensured that the content of MI-Access formed an adequate and representative sample of appropriate content and that the content formed a legitimate basis upon which to derive valid conclusions about student performance. Chapter 3 thus addresses Standard 4.6 of the *Standards* (AERA, APA, & NCME, 2014, p. 87). It shows that the assessment design process, and the participation of Michigan educators in that process, provides a solid rationale for having confidence in the content and design of MI-Access as a tool from which to derive valid inferences about Michigan student performance. This chapter also addresses AERA, APA, and NCME (2014) *Standards* 1.1, 1.11, 4.0, 4.1, 4.2, 4.12, 7.2, 8.4, 12.4, and 12.8.

Chapters 4 and 5, “Test Administration Plan” and “Test Delivery and Administration,” describe the processes, procedures, and policies that guided the administration of MI-Access. These include accommodations, security measures, and written procedures provided to assessment administrators and school personnel. These chapters address AERA, APA, and NCME (2014) *Standards* 4.15, 4.16, 6.1, 6.2, 6.3, 6.4, 6.6, 6.7, and 6.10.

Chapter 6, “Scoring,” explains the procedures used for scoring MI-Access autoscored items and handscored items. This chapter adheres to AERA, APA, and NCME *Standards* 4.18, 4.20, 6.8, and 6.9.

Chapter 7, “Operational Data Analyses,” describes the data used for calibration and scaling. For content areas for which they are appropriate, raw-score results and a classical item analysis were provided, which served as a foundation for subsequent analyses. This chapter also describes the calibration and scaling processes, procedures, and results. Some references to introductory and advanced discussions of item response theory (IRT) are provided. This chapter thereby demonstrates adherence to AERA, APA, and NCME (2014) *Standards* 1.8, 5.2, 5.13, and 5.15.

Chapter 8, “Test Results,” presents scale-score results and achievement-level information. Scale-score results provide a basic quantitative reference to student performance as derived through the IRT models that were applied. This chapter thus addresses AERA, APA, and NCME (2014) *Standards* 5.1, 6.10, 7.0, and 12.18.

Chapter 9, “Performance-Level Setting,” provides background on the standard-setting activities and functions to address *Standards* 5.21 and 5.22 of the *Standards* (AERA, APA, & NCME, 2014).

Chapter 10, “Fairness,” addresses validity evidence, specifically with respect to issues of bias. This chapter demonstrates adherence to AERA, APA, and NCME (2014) *Standards* 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6.

The first half of Chapter 11, “Reliability and Evidence of Construct-Related Validity,” demonstrates adherence to the AERA, APA, and NCME (2014) *Standards* through several analyses of the reliability of the 2019 MI-Access. It presents information on reliability and precision by reporting results on reliability, standard error of measurement (SEM), conditional standard error of measurement (CSEM), and classification consistency and accuracy. The first half of Chapter 11 thereby addresses AERA, APA, and NCME (2014) *Standards* 2.0, 2.3, 2.13, and 2.19. The second half of Chapter 11 addresses validity evidence, including assessment content, response processes, issues of bias, dimensionality analysis, relations to other assessments, and consequences of assessment use. It demonstrates adherence to AERA, APA, and NCME (2014) *Standards* 3.16 and 4.3. Chapter 11 ends with a section addressing the development of validity arguments for MI-Access.

MDE and its testing vendors maintained an unwavering focus on the gathering of validity evidence in support of MI-Access throughout the development, administration, analysis, and reporting of the 2019 MI-Access administration.

Chapter 1: Background of Spring 2019 MI-Access

1.1 Background of MI-Access

MI-Access is Michigan's alternate assessment system and is designed for with the most significant cognitive disabilities and whose Individualized Education Program (IEP) teams have determined that general assessments, even with accommodations, are not appropriate, based on the assessment selection guidelines for this assessment. The three MI-Access assessments are described below.

- Functional Independence (FI) assessments are for students whose instruction is aligned closest to the "High" range of complexity on the alternate content expectations. With guidance, this population of students (within the overall definition of students with the most significant cognitive disabilities) can typically identify basic personal strengths and limitations, as well as access resources, strategies, and supports to help maximize a level of independence.
- Supported Independence (SI) assessments are for students whose instruction is aligned closest to the "Medium" range of complexity on the alternate content expectations. This population of students (within the overall definition of students with the most significant cognitive disabilities) requires ongoing support in one or more major life roles and may have disability-related impacts on the ability to generalize and/or transfer learning.
- Participation (P) assessments are for students whose instruction is aligned closest to the "Low" range of complexity on the alternate content expectations. This population of students (within the overall definition of students with the most significant cognitive disabilities) is expected to require extensive ongoing support in adulthood to participate in most major life roles and faces significant disability-related impacts on the ability to generalize and transfer learning.

Students may take MI-Access FI assessments for only some content areas while taking the M-STEP assessment for other areas, although this distinction is not typical. Students may also take assessments of different MI-Access levels in different content areas, as a student with a significant cognitive disability might function differently in one content area than another. For example, a student's instruction might align to the high range of complexity in one area but to the medium range of complexity in other areas. Each student's IEP team determines the appropriate level of instruction and assessment based on the state guidelines for participation in the alternate assessment.

MI-Access satisfies the federal requirement that all students with disabilities be assessed at the state level.

1.2 Alternate Content Expectations

All students deserve a quality educational experience with challenging expectations that will prepare them for life and careers. To ensure that students with the most significant cognitive disabilities have that same opportunity in a manner that respects their abilities, Michigan developed alternate academic content expectations that adjust the depth, breadth, and complexity of the general content standards at high, medium, and low levels. These provide a range of expectations to meet the range of student abilities.

Michigan's alternate content expectations were developed in collaboration with state leaders, local educators, and national consortia. Development included experts in the content areas and in the instruction of students with disabilities. Alternate content expectations were reviewed by rounds of committees, submitted for public comment, and approved by MDE leadership.

Michigan's alternate content expectations are the Essential Elements with Michigan Range of Complexity for English language arts (ELA) and mathematics, Extended Grade Level Content Expectations for social studies, and Extended Benchmarks for science. The complete alternate content expectations are available online.[1]

1.3 Purpose and Design of the MI-Access Assessments

The alternate assessments determine students' progress toward college and career readiness in four content areas—ELA, mathematics, social studies, and science—based on alternate content and achievement expectations. These assessments are given at the end of the school year.

The alternate assessments accurately measure student achievement (i.e., how much students know at the end of the year) to inform program evaluation and school, district, and state accountability systems.

The MI-Access FI assessment is administered primarily (91%) online; however, each student takes at least the Expressing Ideas portion of the assessment in paper/pencil form and may take more or all of the assessment in paper/pencil form, based on what is instructionally appropriate and needed for accommodations. The SI and P assessments consist of selected-response items and activity-based observation items, with an online interface for administrators to submit student responses.

The blueprints for all content areas can be found in Chapter 3, section 3.3 of this report.

[1] https://www.michigan.gov/mde/0,4615,7-140-22709_28463-429725--,00.html

Chapter 2: Uses of Test Scores

Validity is an overarching component of MI-Access. The following excerpt is from the *Standards for Educational and Psychological Testing* (hereafter the *Standards*) (AERA, APA, & 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. (p. 22)

As stated in the *Standards*, the validity of a testing program hinges on the use of the test scores. Validity evidence that supports the uses of MI-Access scores is provided in this technical report. In this chapter, some possible uses of the test scores are examined.

As the *Standards* notes, "validation is the joint responsibility of the test developer and the test user." (AERA, APA, & NCME, 2014, p. 13).

The subsequent chapters of this technical report provide additional evidence for these uses and technical support for some of the interpretations and uses of test scores. The information in Chapters 3 through 11 also provides a firm foundational claim that the MI-Access assessments measure what they are intended to measure. However, this technical report cannot anticipate all possible interpretations and uses of MI-Access 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 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, the purpose of the test must first be understood. The intended uses of MI-Access scores include the following:

- identifying Michigan students' strengths, weaknesses, and growth between academic years
- communicating expectations for all students
- evaluating school-, district-, and/or state-level programs
- informing stakeholders (teachers, school administrators, district administrators, Michigan Department of Education [MDE] staff members, parents, and the public) on progress toward meeting state academic performance standards and meeting the requirements of the state's accountability program

This technical report refers to the use of the test-level scores (scale scores and performance levels), sub-scores, and performance indicators.

2.2 Test-Level Scores

At the Functional Independence (FI) level, an overall scale score is reported. For the Supported Independence (SI) and Participation (P) levels, a score reflecting points earned out of points possible, based on student performance on the entire test, is reported. In addition, an associated performance level is reported. The FI scores indicate, in varying ways, a student's performance in English language arts (ELA) accessing print and using language/expressing ideas, mathematics, science, or social studies. Likewise, the SI and P scores indicate a student's performance in ELA, mathematics, and science. Test-level scores are reported at four reporting levels: state, school district, school, and student.

Items on the MI-Access test forms were developed by Michigan educators in conjunction with the MDE Office of Educational Assessment and Accountability (OEAA) and Data Recognition Corporation (DRC). See Chapter 3 for an explanation of the item development and review process.

The following sections discuss two types of test-level scores that are reported to indicate a student's performance on MI-Access: 1) the scale score and 2) its associated level of performance.

2.2.1 Scale Scores

A scale score indicating a student's total performance is determined for each content area on MI-Access FI. The overall scale score for a content area quantifies the performance being measured by the test. In other words, the scale score represents the student's level of performance, where higher scale scores indicate higher levels of performance on the test and lower scale scores indicate lower levels of performance.

Scale scores are not comparable across grade levels or content areas. Scores are scaled within grade levels, so even if the same numbers are used in different grades, it does not mean that the scales form a single "vertical scale." MI-Access is a standards-based test that assesses the alternate content expectations for each grade, so a very high score on grade 4 expectations does not provide a valid estimate of how that student performs on grade 5 expectations.

For MI-Access SI and P, students are observed responding to assessment prompts and activities and are scored based on an observation rubric that does not yield a scale score. The student's overall reported scores are the points earned by the student out of the total points possible.

2.2.2 Levels of Performance

A student's performance on MI-Access is reported on one of the three levels: Emerging Toward the Performance Standard, Attained the Performance Standard, and Surpassed the Performance Standard. The cut scores for the MI-Access assessments were established in collaboration between MDE and Michigan educators. Standard setting was conducted in 2015 for science and social studies and in 2017 for English language arts and mathematics.

MI-Access performance levels reflect the performance standards and abilities intended by the Michigan legislature, Michigan teachers, Michigan citizens, and MDE, relative to the alternate content expectations. Descriptions of each performance level in terms of what a student should know and be able to do are provided by MDE and are referenced in the MI-Access performance-level descriptors.[1]

2.2.3 Use of Test-Level Scores

MI-Access performance levels provide summary evidence of student performance. Classroom teachers may use these scores as evidence of student performance 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 scores are valid and reliable indicators of student performance.

2.3 Use of Sub-scores

Sub-scores are scores on important domain areas within each content area. The sub-scores correspond to claims, strands, and disciplines. For ELA and mathematics, the reporting categories are called claims; for science, the reporting categories are called strands; and for social studies, the reporting categories are called disciplines. These reporting categories are primary structural elements in test blueprints and item development.

The purpose of reporting sub-scores on MI-Access is to show the relationship between the overall performance being measured and the skills shown by the individual students in each of the areas delimited by the claims, strands, or disciplines. Teachers may use a student's sub-scores as indicators of strengths and weaknesses.

However, the sub-scores are best corroborated by other evidence, such as homework, class participation, diagnostic test scores, or observations. Chapter 11 of this technical report provides evidence of content validity and reliability that supports the use of the claim, strand, and discipline sub-scores. Chapter 11 also provides evidence of construct-related validity that further supports the use of these sub-scores.

2.3.1 ELA Claims

Claim #1 – Reading and Reading Comprehension

- Students can comprehend text in increasingly complex ways.

Claim #2 – Writing: Text Types and Purposes

- Students can produce writing for a range of purposes and audiences.

Claim #3 - Communication and Language

- Students can communicate for a range of purposes and audiences.

[1] https://www.michigan.gov/mde/0,4615,7-140-22709_28463-429725--,00.html

ELA Claims (continued)

Claim #4 - Research and Inquiry

- Students can investigate topics and present information.

2.3.2 Mathematics Claims

Claim #1 - Number Sense

- Students demonstrate increasingly complex understanding of number sense.

Claim #2 - Geometry

- Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.

Claim #3 - Measurement, Data Analysis

- Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures.

Claim #4 - Problem Solving

- Students solve increasingly complex mathematical problems, making productive use of algebra and functions.

2.3.3 Science Strands

Strand: Constructing New Scientific Knowledge (CN)

- All students will design and conduct investigations using appropriate methodology and technology.

Strand: Reflecting on Scientific Knowledge (RO)

- All students will analyze claims for their scientific merit and explain how scientists decide what constitutes scientific knowledge.

Strand: Using Life Science Knowledge

- CELLS (CE) All students will apply an understanding of cells to the functioning of multi-cellular organisms, including how cells grow, develop, and reproduce.
- ORGANIZATION OF LIVING THINGS (OR) All students will use classification systems to describe groups of living things.
- HEREDITY (HE) All students will investigate and explain how characteristics of living things are passed on through generations.
- EVOLUTION (EV) All students will explain how scientists construct and scientifically test theories concerning the origin of life and evolution of species.
- ECOSYSTEMS (EC) All students will explain how parts of an ecosystem are related and how they interact.

Science Strands (continued)

Strand: Using Physical Science Knowledge

- MATTER AND ENERGY (ME) All students will explain what the world around us is made of.
- CHANGES IN MATTER (CM) All students will investigate, describe, and analyze ways in which matter changes.
- MOTION OF OBJECTS (MO) All students will describe how things around us move, explain why things move as they do, and demonstrate and explain how we control the motions of objects.
- WAVES AND VIBRATIONS (WV) All students will describe sounds and sound waves.

Strand: Using Earth Science Knowledge

- GEOSPHERE (GE) All students will describe the earth's surface.
- HYDROSPHERE (HY) All students will describe the characteristics of water and demonstrate where water is found on earth.
- ATMOSPHERE AND WEATHER (AW) All students will investigate and describe what makes up weather and how it changes from day to day, from season to season and over long periods of time.
- SOLAR SYSTEM, GALAXY AND UNIVERSE (SS) All students will compare and contrast our planet and sun to other planets and star systems.

2.3.4 Social Studies Disciplines

Discipline: Beginnings to 1620

- American Indian Life in the Americas
- European Exploration
- Three World Interactions

Discipline: Colonization and Settlement (1585–1763)

- European Struggle for Control of North America
- European Slave Trade and Slavery in Colonial America
- Life in Colonial America

Discipline: Revolution and the New Nation (1754–1800s)

- Causes of the American Revolution
- The American Revolution and Its Consequences
- Creating New Government(s) and a New Constitution

Discipline: Public Discourse, Decision Making, Citizen Involvement

- Identifying and Analyzing Public Issues
- Decision Making
- Persuasive Communication About a Public Issue
- Citizen Involvement

Social Studies Disciplines (continued)

Discipline: Expansion and Reform (1792–1861)

- Challenges to an Emerging Nation
- Regional and Economic Growth
- Reform Movements

Discipline: Civil War, Reconstruction, and Development of United States (1850–1930)

- The Coming of the Civil War
- Civil War
- Reconstruction
- America in the Last Half of the 19th Century

Discipline: World History and Geography

- Expanding and Intensified Hemispheric Interactions (300–1500 CE/AD)
- The Emergence of the First Global Age (15th–18th centuries)
- An Age of Global Revolutions (18th century–1914)
- Global Crisis and Achievement (1900–1945)
- The Cold War and Its Aftermath: The 20th Century Since 1945

Discipline: United States History and Geography (USHG)

- The Development of an Industrial, Urban, and Global United States (1870–1930)
- The Great Depression and World War II (1920–1945)
- Post-World War II United States (1945–1989)
- America in a New Global Age

Discipline: Economics

- The Market Economy
- The National Economy of the United States of America

Discipline: Civics

- Conceptual Foundations of Civic and Political Life
- Origins and Foundations of Government of the United States of America
- Structure and Functions of Government in the United States of America
- The United States of America and World Affairs
- Citizenship in the United States of America

Chapter 3: Test Design and Item Development

3.1 Overview

This chapter is particularly relevant to AERA, APA, & NCME (2014) *Standards* 4.0, 4.1, and 4.7, which are from Chapter 4, "Test Design and Development," of the AERA, APA, & NCME (2014) *Standards*. 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.

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. (p. 85)

The purpose of this chapter is to document the test design and item development process used for MI-Access. In this chapter, the steps taken to create MI-Access are described, from the development of test specifications to the selection of operational items.

Guidelines for bias and sensitivity issues, accessibility and accommodations, and style help item developers and reviewers ensure consistency and fairness across the item bank. The specifications and guidelines were reviewed by school districts, higher education representatives, and other stakeholders. The item specifications describe the evidence to be elicited to guide the development of items that measure student performance relative to the target.

The assessment blueprints describe the content of the alternate assessments for grades 3–8 and 11 that were administered in the 2018–19 school year and describe how that content was assessed. The test blueprints for the alternate assessment reflected the depth and breadth of the performance expectations of Michigan's alternate content expectations. The test blueprints that were subsequently developed into fixed form test maps.

3.1.1 A Brief Description of Content Structure for ELA: Accessing Print and Using Language/Expressing Ideas, Mathematics, Science, and Social Studies

MI-Access content in English language arts (ELA), mathematics, science, and social studies is defined by the knowledge and skills identified in the Michigan alternate content expectations. These expectations were developed in consultation and collaboration with educators and the general public, representing consensus on the essential content for Michigan learners. The alternate content expectations are grade level or grade band specific, and, as developed, aligned to the Michigan K-12 content standards for a given content area for the corresponding grade level or grade band. 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 MI-Access test specifications evolve from the tension between the constraints of the assessment program and the benefits sought from the examination of students. These benefits and constraints mix scientific rigor with policy considerations.

The MI-Access test specifications consist of a blueprint and test maps for each grade level and content area. The 2019 MI-Access test selection specifications were finalized by the Michigan Department of Education (MDE) and its psychometricians and vendors in 2018.

The key structural aspect, the test blueprint, represents a compromise among many constraints, including the availability of items from field-testing and results of multiple reviews by content specialists. Test design includes such elements as the number and types of items for each of the scores reported. The 2019 MI-Access operational forms matched the test blueprints that were intended for all MI-Access content area alternate assessments.

3.2 Test Blueprints

Test specifications and blueprints define the knowledge, skills, and abilities intended to be measured on each student's test event. A blueprint also specifies how skills are sampled from a set of content standards (e.g., the Michigan alternate content expectations). Other important factors, such as Extended Depth of Knowledge (EDOK), are also specified. Specifically, a test blueprint is a formal document that guides the development and assembly of an assessment event/form by explicating the following types of essential information:

- content (claims/strands/disciplines and assessment targets) that is included for each assessed content area and grade across various levels of the system (student, classroom, school, district, and state levels)
- the relative emphasis of content expectations, generally indicated as the number of items or percentage of points per claim/strand/discipline and assessment target
- the item types used or required, which communicate to item developers how to measure each claim/strand/discipline and assessment target and communicate learning expectations to teachers and students
- EDOK, indicating the complexity of item types for each claim/strand/discipline and assessment target

The test blueprint is an essential guide for both assessment developers and for curriculum and instruction. For assessment developers, the blueprint and related test-specification documents define how the test will ensure coverage of the full breadth and depth of content and how it will maintain fidelity to the intent of the Michigan alternate content expectations on which the assessments are based. Full content alignment is necessary to ensure that educational stakeholders can make valid, reliable, and unbiased inferences about student, classroom, school, district, and state performance. At the instructional level, the test blueprint provides a guide to the relative importance of competing content demands and suggests how the content is demonstrated, as indicated by item type and EDOK. In summary, an assessment blueprint provides clear development specifications and signals to the broader education community both the full complexity of the standards and how performance on these standards is substantiated.

3.2.1 Test Specifications

AERA, APA, and 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). (p. 85)

The purpose of MI-Access is discussed in sections 1.2 and 1.3 of Chapter 1 of this report. MI-Access tests the knowledge and skills that are identified within Michigan's standards-based accountability system. This framework, in turn, is based on prior consensus among MDE staff, Michigan educators, and experienced content-area experts that the framework represents content that is important for teachers to teach and for students to learn. MI-Access aligns to Michigan's alternate content expectations in ELA, mathematics, science, and social studies, designed for students with the most significant cognitive disabilities.

In accordance with these purposes, AERA, APA, and 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. (p. 89)

Item and test development are guided by sets of specifications. Details on these specifications for all MI-Access assessments can be found within this chapter. All MI-Access assessments are developed by content experts at the MDE using content developed by Michigan teachers.

A general description of development activities applying to all Michigan-created assessments (including MI-Access) is provided below. The Office of Educational Assessment and Accountability (OEAA) staff, contractors, and Michigan educators work together to develop these state assessments. Specifically, the development cycle includes the following steps:

- Item writer training
- Item development
- Item review
- Field-testing

- Field-test data review (item level)
- Operational test construction

3.2.2 Item Writer Training

Once item specifications are finalized, Michigan's item development contractor uses customized materials approved by the OEAA to train item writers to write items specifically for MI-Access. Item writer training can last anywhere from three to five days and is conducted by contractor staff in conjunction with the OEAA test development staff. The process of item writing includes cycle(s) of feedback from contractor and OEAA staff. It can take between four to eight weeks for an item to move from initial assignment to accepted status. All item writers are Michigan educators who have curriculum and instruction expertise for the grade level and content area for which they are writing, as well as experience instructing students for whom MI-Access is intended. In addition, prospective item writers are required to submit three original test items aligned to grade-specific content expectations, which the OEAA test development staff review and potentially approve for item authoring. Michigan's item writers possess relevant degrees and experience, and many have previous specific experience in item writing for MI-Access.

3.2.3 Item Development

Item development is discussed in this section in compliance with the AERA, APA, and NCME (2014) *Standards*. Standard 4.7 states the following:

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

For MI-Access ELA, mathematics, science, and social studies items, Michigan item writers draft test items in accordance with item specifications approved by the OEAA test development staff, following the best practices for the field. Contractor staff review items internally and then share them with OEAA test development staff for an additional review. Sections 3.2.6 and 3.3 of this report discuss how the items are selected for field-testing or operational use.

The internal review consists of determining whether the item meets the following criteria:

Skill:

- Item measures one skill level.
- Item measures skill in manner consistent with specifications.
- Item assesses an appropriate (realistic) level of skill.
- Item makes clear the skill to be employed.

Content:

- Item measures one primary academic content expectation.
- Item measures the academic content expectation in a manner consistent with specifications.
- Item taps the appropriate (important) aspect of content associated with the academic content expectation.
- Item makes clear the benchmark or problem to be solved.

Relevance:

- Item is not contrived.
- Item is appropriate for the grade level to be tested.
- Item groups reflect instructional emphasis.

Accuracy:

- Item is factually accurate.
- Multiple-choice (MC) items contain only one correct or best response.
- If item pertains to disputed content, context for correct answer is clearly defined.
- Item is worded unambiguously.
- Item contains no extraneous material, except as required by the content expectation.
- Vocabulary is grade-level appropriate or deemed appropriate for the population of students being assessed and is clear.
- Item contains no errors in grammar, spelling, or mechanics.
- Item responses are parallel and related to the stem.
- Item responses are independent.
- Item contains no clues or irrelevant distracters.
- Directions for responding to a constructed-response (CR) item are clear.
- CR item and rubric match.
- CR rubric is clear and easy to apply.
- Item is clearly and conveniently placed on the page.
- Physical arrangement of item is consistent with the OEAA style guide.
- Keys for sets of multiple-choice (MC) items are balanced (for example, equal numbers of A, B, and/or C response options).

Bias:

- Item is free of racial, socioeconomic, and gender stereotypes.
- Item contains no material known or suspected to give advantage to any group.
- Item is free of insensitive language.
- Item sets that identify race or gender either directly or indirectly are balanced with reference to race and gender.
- Item content and format are accessible to students with varying types of disabilities.
- Item content and format are accessible to students with limited English proficiency.

3.2.4 Graphics Creation

MDE has an internal team of media designers who use the graphic descriptions submitted by the item writers through Michigan's Item Bank System (IBS) to create the pictures, graphs, maps, and other artwork needed for online test items. MDE and DRC staff review and approve the completed artwork in preparation for the item review.

3.2.5 Item Review

Continuing from Standard 4.7 (above), AERA, APA, and NCME (2014) Standard 3.2 is particularly relevant to fairness in item development:

> 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. (p. 64)

The Bias and Sensitivity Review Committees (BSCs) are composed of representatives from various backgrounds whose purpose is to screen the items for racial, socioeconomic, gender, and other sensitivity issues. This follows AERA, APA, and NCME (2014) Standard 3.1, which states the following:

> **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. (p. 63)

Panels of educators reviewed items, item stimuli, and paper/pencil documents for accessibility, bias/sensitivity, and content. (Item stimuli include the reading passages used on the ELA and social studies assessments and the figures and graphics used on the ELA, mathematics, science, and social studies assessments.) During the accessibility reviews, panelists identified issues that could negatively affect a student's ability to access stimuli and items or to elicit valid evidence about an assessment target. During the BSC review, panelists identified content in stimuli and items that could negatively affect a student's ability to produce a correct response because of the student's background.

After the BSC review, all MI-Access items were reviewed by Michigan educators in a Content Advisory Committee (CAC). The content review focused on developmental appropriateness and alignment of stimuli, items, and tasks to the content specifications and appropriate depths of knowledge. Panelists in the content review also checked the accuracy of the content, answer keys, and scoring materials.

Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised to address the issues identified by the panelists or marked as Do Not Use (DNU) in the Michigan IBS.

Contractor staff trains the CAC and BSC participants using OEAA-approved materials and facilitates the committee meetings under the leadership of the OEAA test development staff. All newly written test items are typically reviewed first by the BSC and then by the CAC.

An item rejected by the BSC might or might not get passed on to the CAC for review. Each review is led by experienced contractor staff, with test development staff in attendance, using the following prescribed guidelines to indicate the final status of each item:

- **Accept:** The criteria outlined in the review were met in all areas (skill, content, relevance, accuracy, and bias), and the item appears suitable for field-testing.

- **Revise:** One or more of the criteria have not been met or the item needs minor changes to make it acceptable. Reviewers provide recommendations on changes to be made to the item that will make the item suitable for field-testing.
- **Reject:** Several category conditions have not been met, are suspect, or need radical changes to make the item acceptable. In such cases, the item might be vague or ambiguous, inappropriate, or not clearly related to the text or the standard. Without extensive revisions, the item is unlikely to be salvaged. Reviewers provide comments to explain why the item should be rejected.

Items that pass bias/sensitivity and content reviews are eligible for field-testing.

3.2.6 Field-Testing

Before an item can be used on an operational test or added to the operational item pool, it must be field-tested. The OEAA uses two approaches to administer field-test items: embed field-test items in an operational administration or embed field-test items in a stand-alone field-test administration. Items that pass bias/sensitivity and content review are eligible for field-testing.

The OEAA embeds FT items in multiple forms of operational fixed-form assessments. Administering field-test items this way ensures that they are randomly distributed, allowing for a large representative sample of responses to be gathered under operational conditions for each item. Enough field-test items are administered annually to replenish and improve the item pools.

When MDE implements testing at new grade levels, for new content areas, or for revised academic standards, it is necessary to conduct a separate stand-alone field test to obtain performance data. When stand-alone field-testing is required, MDE requests volunteer participation from school districts.

In 2019, all items field-tested on the MI-Access assessments were embedded into operational fixed-forms.

3.2.7 Range-Finding

After the student responses to the field-tested CR items are collected, a range-finding is conducted to determine scoring guidelines and score-point ranges for the different score points for each field-tested CR item. This information is then used in the preparation of materials to guide the handscoring of student responses to the item, which is done by a trained team of readers, as described in Chapter 6 of this report.

Note: For MI-Access FI ELA, the Expressing Ideas portion is the only area in which CR items are administered. The Expressing Ideas portion is administered in paper/pencil format independently of the rest of the ELA assessment to eliminate barriers for students as they respond, based on the allowable types of responses on the scoring rubric.

3.2.8 Data Review

After field-testing, the results are analyzed by MDE psychometric staff. Contractor staff and test development staff convene data review committee meetings with Michigan educators. Significant effort goes into ensuring that these committee members represent the state demographically with respect to ethnicity, gender, school district size, and geographical region. These committees receive training on interpreting the psychometric data compiled for each field-test item from the OEAA psychometric staff. Content experts (usually teachers) and group facilitators apply this training to the data review process. During these data review meetings, participants review the items with field-test statistics. Data provided to the data review committees are separated by BSC and CAC.

The data that are reviewed during BSC include the following:

- N-count
- adjusted *p*-value (that is, the adjusted item mean in the range of 0–1 for all items)
- Differential Item Functioning (DIF) flag (for FI tests)
- favored group
- percentage of students who choose each option, omit a response, and/or submit multiple marks (in paper/pencil tests)
- option-total correlation
- omit-total correlation

The data that are reviewed during CAC include the following:

- overall N-count
- adjusted *p*-value
- difficulty flag
- item-total correlation
- item-total flag
- percentage of students who choose each option, omit providing a response, and/or submit multiple marks (in paper/pencil tests)
- option-total correlation
- omit-total correlation

As mentioned above, specific directions are provided on the use of the statistical information and how to use Michigan's IBS. BSC members evaluate each test item for fairness issues with respect to culture, ethnicity, gender, geographic location, and economic status, using the data listed above for this group. CAC members evaluate each test item regarding alignment to the alternate academic content expectations, grade-level appropriateness, and level of EDOK, using the data information listed above for this group. Both committees then recommend that the item either be accepted, revised for additional field-testing, or rejected.

New items that survive all reviews and field-testing are saved in the Michigan IBS as "Ready for Operational," meaning they are now eligible for operational use.

3.3 Operational Test Construction

The OEAA test development staff build test maps that meet the test specifications (blueprint and psychometric specifications) inside Michigan's IBS. All test maps are reviewed for the correct answer key, accurate content expectation, and appropriate statistic/psychometric information for each item. In addition, comparability of the overall test across forms and across adjacent years is also examined. Corresponding details for the four content areas are presented below.

3.3.1 English Language Arts

MI-Access English language arts (ELA) assessments are based on Michigan's ELA alternate content expectations. The ELA assessment consists of four claims: Reading and Reading Comprehension, Writing and Sharing Ideas, Communication and Language, and Research and Inquiry. These are divided into two sections of the assessment: "Accessing Print and Using Language" (APUL) and "Expressing Ideas" (EI). The assessment is administered in grades 3–8 and 11.

The ELA assessment structure is summarized in Tables 3-1 through 3-22.

Table 3-1. ELA Overall Structure: Number of Items

Assessment Name	Operational Items per Form	Embedded Field Test Items per Form	Total Items per Form
MI-Access Functional Independence	31	12	43
MI-Access Supported Independence	15	5	20
MI-Access Participation	10	5	15

Blueprint specifications by claim/score reporting category are provided in the next section. The blueprint specifications for MI-Access SI and MI-Access P specify the total number of items per claim and total number of items by item type; however, there is flexibility within those parameters from year to year on the distribution of items across item type per claim.

The following tables specify what was true for the assessments in the 2019 testing cycles. Operational coverage by claim is the same from test cycle to test cycle, but coverage for field test items changes from cycle to cycle based on inventory needs. There were three forms for the FI assessments, and two forms for the SI and P assessments. The MI-Access SI and P assessments had three embedded field-test selected-response (SR) items per form and two embedded field-test activity-based observation (ABO) items per form. The field test designations below show the coverage across all forms for 2019.

Table 3-2. ELA Structure for FI Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas Forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-3. ELA Structure for SI Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	4	2	3	1
MI-Access SI	Writing and Sharing Ideas	1	2	1	1
MI-Access SI	Communication and Language	2	1	1	1
MI-Access SI	Research and Inquiry	2	1	1	1

Table 3-4. ELA Structure for P Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	1	1
MI-Access P	Communication and Language	1	1	1	1
MI-Access P	Research and Inquiry	1	1	1	1

Table 3-5. ELA Structure for FI Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-6. ELA Structure for SI Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	4	2	3	1
MI-Access SI	Writing and Sharing Ideas	2	1	1	1
MI-Access SI	Communication and Language	2	1	2	0
MI-Access SI	Research and Inquiry	1	2	0	2

Table 3-7. ELA Structure for P Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	0
MI-Access P	Writing and Sharing Ideas	0	2	0	2
MI-Access P	Communication and Language	2	0	1	1
MI-Access P	Research and Inquiry	1	1	2	1

Table 3-8. ELA Structure for FI Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas Forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	0	1	0	2
MI-Access FI	Communication and Language	5	0	9	0
MI-Access FI	Research and Inquiry	5	0	3	0

Table 3-9. ELA Structure for SI Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	5	1	4	0
MI-Access SI	Writing and Sharing Ideas	0	3	0	2
MI-Access SI	Communication and Language	2	1	1	1
MI-Access SI	Research and Inquiry	2	1	1	1

Table 3-10. ELA Structure for P Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	0	1
MI-Access P	Communication and Language	1	1	1	0
MI-Access P	Research and Inquiry	1	1	2	2

Table 3-11. ELA Structure for FI Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas forms
MI-Access FI	Reading and Reading Comprehension	20	0	211	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-12. ELA Structure for SI Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	5	1	4	0
MI-Access SI	Writing and Sharing Ideas	2	1	0	1
MI-Access SI	Communication and Language	2	1	1	1
MI-Access SI	Research and Inquiry	0	3	1	2

Table 3-13. ELA Structure for P Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	1	1
MI-Access P	Communication and Language	2	0	1	1
MI-Access P	Research and Inquiry	0	2	1	1

Table 3-14. ELA Structure for FI Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas Forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-15. ELA Structure for SI Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	4	2	3	2
MI-Access SI	Writing and Sharing Ideas	3	0	2	0
MI-Access SI	Communication and Language	2	1	1	1
MI-Access SI	Research and Inquiry	0	3	0	1

Table 3-16. ELA Structure for PI Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	1	1
MI-Access P	Communication and Language	1	1	1	1
MI-Access P	Research and Inquiry	1	1	1	1

Table 3-17. ELA Structure for FI Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas Forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-18. ELA Structure for SI Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	5	1	3	1
MI-Access SI	Writing and Sharing Ideas	0	3	0	1
MI-Access SI	Communication and Language	3	0	2	1
MI-Access SI	Research and Inquiry	1	2	1	1

Table 3-19. ELA Structure for P Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	1	1
MI-Access P	Communication and Language	1	1	1	1
MI-Access P	Research and Inquiry	1	1	1	1

Table 3-20. ELA Structure for FI Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Operational CR per form	Embedded FT MC across 3 forms	Embedded FT CR across 2 Expressing Ideas Forms
MI-Access FI	Reading and Reading Comprehension	20	0	21	0
MI-Access FI	Writing and Sharing Ideas	2	1	3	2
MI-Access FI	Communication and Language	4	0	6	0
MI-Access FI	Research and Inquiry	4	0	3	0

Table 3-21. ELA Structure for SI Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Reading and Reading Comprehension	4	2	2	2
MI-Access SI	Writing and Sharing Ideas	2	1	1	0
MI-Access SI	Communication and Language	2	1	1	2
MI-Access SI	Research and Inquiry	1	2	2	0

Table 3-22. ELA Structure for P Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Reading and Reading Comprehension	3	1	3	1
MI-Access P	Writing and Sharing Ideas	1	1	1	1
MI-Access P	Communication and Language	1	1	1	1
MI-Access P	Research and Inquiry	1	1	1	1

3.3.2 Mathematics

MI-Access mathematics assessments are based on Michigan's alternate content expectations for mathematics. MI-Access mathematics consists of four claims: Number Sense, Geometry, Measurement, Data and Analysis, and Problem Solving. The assessment is administered in grades 3–8 and 11.

The mathematics assessment structure is summarized in Tables 3-23 through 3-44.

Table 3-23. Mathematics Overall Structure: Number of Items

Assessment Name	Operational Items per Form	Embedded Field Test Items per Form	Total Items per Form
MI-Access Functional Independence	24	10	34
MI-Access Supported Independence	15	5	20
MI-Access Participation	10	5	15

Blueprint specifications by claim/reporting level are provided in the next section. The blueprint specifications for MI-Access SI and MI-Access P specify total number of items per claim and total number of items by item type; however, there is flexibility within those parameters from year to year on the distribution of items across item type per claim.

Operational coverage by claim is the same from test cycle to test cycle, however coverage for field test items change from cycle to cycle based on inventory needs. There were three forms for the FI assessments, and two forms for the SI and P assessments. The MI-Access SI and P assessments had three embedded field-test selected-response (SR) items per form and two embedded field-test activity-based observation (ABO) items per form. The field test designations below show the coverage across all forms for 2019.

Table 3-24. Mathematics Structure for FI Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 form
MI-Access FI	Number Sense	7	9
MI-Access FI	Geometry	4	6
MI-Access FI	Measurement, Data, and Analysis	7	9
MI-Access FI	Problem Solving	6	6

Table 3-25. Mathematics Structure for SI Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	3	2	2	2
MI-Access SI	Geometry	1	1	1	1
MI-Access SI	Measurement, Data, and Analysis	3	2	2	0
MI-Access SI	Problem Solving	2	1	1	1

Table 3-26. Mathematics Structure for P Grade 3: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	2	1	3	1
MI-Access P	Geometry	2	0	1	1
MI-Access P	Measurement, Data, and Analysis	1	2	1	1
MI-Access P	Problem Solving	1	1	1	1

Table 3-27. Mathematics Structure for FI Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	7	9
MI-Access FI	Geometry	4	6
MI-Access FI	Measurement, Data, and Analysis	8	9
MI-Access FI	Problem Solving	5	6

Table 3-28. Mathematics Structure for SI Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	4	1	0	2
MI-Access SI	Geometry	1	1	1	1
MI-Access SI	Measurement, Data, and Analysis	3	2	3	1
MI-Access SI	Problem Solving	2	3	2	0

Table 3-29. Mathematics Structure for P Grade 4: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR cross 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	2	1	3	1
MI-Access P	Geometry	1	1	1	1
MI-Access P	Measurement, Data, and Analysis	2	1	1	1
MI-Access P	Problem Solving	1	1	1	1

Table 3-30. Mathematics Structure for FI Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	12	15
MI-Access FI	Geometry	4	5
MI-Access FI	Measurement, Data, and Analysis	6	7
MI-Access FI	Problem Solving	2	3

Table 3-31. Mathematics Structure for SI Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	4	3	3	2
MI-Access SI	Geometry	2	1	1	1
MI-Access SI	Measurement, Data, and Analysis	2	1	1	1
MI-Access SI	Problem Solving	1	1	1	0

Table 3-32. Mathematics Structure for P Grade 5: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	3	1	2	2
MI-Access P	Geometry	1	1	1	1
MI-Access P	Measurement, Data, and Analysis	2	1	1	1
MI-Access P	Problem Solving	0	1	2	0

Table 3-33. Mathematics Structure for FI Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	11	12
MI-Access FI	Geometry	4	6
MI-Access FI	Measurement, Data, and Analysis	4	6
MI-Access FI	Problem Solving	5	62

Table 3-34. Mathematics Structure for SI Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	4	2	3	1
MI-Access SI	Geometry	1	2	0	2
MI-Access SI	Measurement, Data, and Analysis	3	0	2	0
MI-Access SI	Problem Solving	1	2	1	1

Table 3-35. Mathematics Structure for P Grade 6: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR per form	Embedded FT ABO per form
MI-Access P	Number Sense	2	2	3	1
MI-Access P	Geometry	2	0	1	1
MI-Access P	Measurement, Data, and Analysis	0	2	1	1
MI-Access P	Problem Solving	2	0	1	1

Table 3-36. Mathematics Structure for FI Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC across 3 forms	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	10	12
MI-Access FI	Geometry	8	12
MI-Access FI	Measurement, Data, and Analysis	3	3
MI-Access FI	Problem Solving	3	31

Table 3-37. Mathematics Structure for SI Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	4	2	1	2
MI-Access SI	Geometry	3	2	2	2
MI-Access SI	Measurement, Data, and Analysis	1	1	1	0
MI-Access SI	Problem Solving	1	1	2	0

Table 3-38. Mathematics Structure for P Grade 7: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	3	0	2	2
MI-Access P	Geometry	1	2	2	0
MI-Access P	Measurement, Data, and Analysis	1	1	1	1
MI-Access P	Problem Solving	1	1	1	1

Table 3-39. Mathematics Structure for FI Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	6	6
MI-Access FI	Geometry	8	11
MI-Access FI	Measurement, Data, and Analysis	2	24
MI-Access FI	Problem Solving	8	9

Table 3-40. Mathematics Structure for SI Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	3	1	2	1
MI-Access SI	Geometry	3	2	1	2
MI-Access SI	Measurement, Data, and Analysis	1	1	1	0
MI-Access SI	Problem Solving	2	2	2	1

Table 3-41. Mathematics Structure for P Grade 8: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	2	1	1	1
MI-Access P	Geometry	2	1	3	1
MI-Access P	Measurement, Data, and Analysis	2	0	1	1
MI-Access P	Problem Solving	0	2	1	1

Table 3-42. Mathematics Structure for FI Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational MC per form	Embedded FT MC across 3 forms
MI-Access FI	Number Sense	3	6
MI-Access FI	Geometry	3	3
MI-Access FI	Measurement, Data, and Analysis	7	9
MI-Access FI	Problem Solving	11	12

Table 3-43. Mathematics Structure for SI Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access SI	Number Sense	2	1	1	1
MI-Access SI	Geometry	1	2	1	1
MI-Access SI	Measurement, Data, and Analysis	2	1	2	1
MI-Access SI	Problem Solving	4	2	2	1

Table 3-44. Mathematics Structure for P Grade 11: Number of Items by Claim and Item Type

Assessment Name	Claim/Score Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Number Sense	1	1	1	1
MI-Access P	Geometry	1	1	2	0
MI-Access P	Measurement, Data, and Analysis	1	1	1	1
MI-Access P	Problem Solving	3	1	2	2

3.3.3 Social Studies

MI-Access social studies assessments are based on Michigan's alternate content expectations for social studies. This assessment is administered in grades 5, 8, and 11. Currently, the social studies alternate content expectations and assessment are offered only at the FI level. Most students participating in the SI and P levels of MI-Access engage in social studies topics that are specific to their immediate world (home, school, and community); therefore, assessments for these levels are customized at the local level.

- The MI-Access social studies assessment for FI grade 5 consists of four disciplines (32 operational items and 8 embedded field-test items):
  - United States History and Geography (USHG): Beginnings to 1620
  - USHG: Colonization/Settlement
  - USHG: Revolution/New Nation
  - Public Discourse/Citizenship

- The MI-Access social studies assessment for grade 8 consists of four disciplines (33 operational items and 9 embedded field-test items):
  - USHG: Revolution/New Nation
  - USHG: Expansion/Reform
  - USHG: Civil War, Reconstruction, and Development of the United States
  - Public Discourse/Citizenship

- The MI-Access social studies assessment for grade 11 consists of four disciplines (41 operational items and 11 embedded field-test items):
  - USHG
  - World History and Geography
  - Civics
  - Economics

The social studies assessment structure is summarized in Table 3-45.

Table 3-45. Social Studies Structure for Grades 5, 8, and 11

Grade	Discipline	Number of Operational Items per form	Number of Embedded Field Test items across 3 forms
5	USHG: Beginnings to 1620	8	3
5	USHG: Colonization/Settlement	9	7
5	USHG: Revolution/New Nation	10	11
5	Public Discourse/Citizenship	5	3
8	USHG: Revolution/New Nation	7	11
8	USHG: Expansion/Reform	11	4
8	USHG: Civil War, Reconstruction and Development of the United States	10	5
8	Public Discourse/Citizenship	5	7
11	World History and Geography	10	3
11	USHG	13	13
11	Civics	13	6
11	Economics	5	11

3.3.4 Science

MI-Access science assessments are based on Michigan's science extended benchmarks. The assessment is administered in grades 4, 7, and 11. The MI-Access science assessment in all three grades consists of five strands:

- Constructing New Scientific Knowledge
- Reflecting on New Scientific Knowledge
- Using Life Science
- Using Physical Science
- Using Earth Science

The science assessment structure is summarized in Tables 3-46 through 3-55.

Table 3-46. Science Overall Structure: Number of Items

Assessment Name	Operational Items per form	Embedded FT per form	Total Items per form
MI-Access FI – Grade 4	35	8	43
MI-Access SI – Grade 4	17	5	22
MI-Access P – Grade 4	15	5	20
MI-Access FI – Grade 7	40	10	50
MI-Access SI – Grade 7	17	5	22
MI-Access P – Grade 7	15	5	20
MI-Access FI – Grade 11	45	10	55
MI-Access SI – Grade 11	17	5	22
MI-Access P – Grade 11	15	5	20

Blueprint specifications by strand/reporting category are provided in the next section. Embedded field-test items vary from strand to strand, year to year, based on inventory needs. The tables on the following pages report what was field tested in 2019.

Table 3-47. Science Structure for FI Grade 4: Number of Items by Strand /Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational MC per form	Embedded FT MC across 2 forms
MI-Access FI	Constructing New Scientific Knowledge	2	1
MI-Access FI	Reflecting on New Scientific Knowledge	2	1
MI-Access FI	Using Life Science	13	6
MI-Access FI	Using Physical Science	12	4
MI-Access FI	Using Earth Science	6	4

Table 3-48. Science Structure for SI Grade 4: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational SR per form	Embedded FT SR across 2 forms
MI-Access SI	Constructing New Scientific Knowledge	1	1
MI-Access SI	Reflecting on New Scientific Knowledge	1	1
MI-Access SI	Using Life Science	7	2
MI-Access SI	Using Physical Science	3	2
MI-Access SI	Using Earth Science	5	4

Table 3-49. Science Structure for P Grade 4: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Constructing New Scientific Knowledge	0	1	0	1
MI-Access P	Reflecting on New Scientific Knowledge	1	0	0	0
MI-Access P	Using Life Science	4	1	2	1
MI-Access P	Using Physical Science	3	2	2	1
MI-Access P	Using Earth Science	1	2	2	1

Table 3-50. Science Structure for FI Grade 7: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational MC per form	Embedded FT MC across 2 forms
MI-Access FI	Constructing New Scientific Knowledge	2	1
MI-Access FI	Reflecting on New Scientific Knowledge	2	1
MI-Access FI	Using Life Science	14	6
MI-Access FI	Using Physical Science	14	6
MI-Access FI	Using Earth Science	8	6

Table 3-51. Science Structure for SI Grade 7: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational SR per form	Embedded FT SR across 2 forms
MI-Access SI	Constructing New Scientific Knowledge	1	1
MI-Access SI	Reflecting on New Scientific Knowledge	1	1
MI-Access SI	Using Life Science	7	2
MI-Access SI	Using Physical Science	3	2
MI-Access SI	Using Earth Science	5	4

Table 3-52. Science Structure for P Grade 7: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Constructing New Scientific Knowledge	1	0	1	1
MI-Access P	Reflecting on New Scientific Knowledge	0	1	0	0
MI-Access P	Using Life Science	4	1	2	2
MI-Access P	Using Physical Science	2	3	2	0
MI-Access P	Using Earth Science	2	1	1	1

Table 3-53. Science Structure for FI Grade 11: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational MC per form	Embedded FT MC across 2 forms
MI-Access FI	Constructing New Scientific Knowledge	2	1
MI-Access FI	Reflecting on New Scientific Knowledge	2	1
MI-Access FI	Using Life Science	14	6
MI-Access FI	Using Physical Science	15	6
MI-Access FI	Using Earth Science	12	6

Table 3-54. Science Structure for SI Grade 11: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/ Reporting Category	Operational SR per form	Embedded FT SR across 2 forms
MI-Access SI	Constructing New Scientific Knowledge	1	1
MI-Access SI	Reflecting on New Scientific Knowledge	1	0
MI-Access SI	Using Life Science	7	3
MI-Access SI	Using Physical Science	3	2
MI-Access SI	Using Earth Science	5	4

Table 3-55. Science Structure for P Grade 11: Number of Items by Strand/Reporting Category and Item Type

Assessment Name	Strand/Reporting Category	Operational SR per form	Operational ABO per form	Embedded FT SR across 2 forms	Embedded FT ABO across 2 forms
MI-Access P	Constructing New Scientific Knowledge	0	1	1	0
MI-Access P	Reflecting on New Scientific Knowledge	1	0	0	1
MI-Access P	Using Life Science	4	2	3	1
MI-Access P	Using Physical Science	3	2	1	1
MI-Access P	Using Earth Science	1	1	1	1

3.3.5 Accommodations

Michigan is committed to ensuring all students, including English Learners and students with disabilities, have access to a wide array of tools across MI-Access. Sections 4.1 through 4.3 of this report detail the universal tools, designated supports, and accommodations Michigan provides. Paper/pencil accommodated versions of the tests are available in unified English braille, contracted braille, and enlarged print. MI-Access accommodated assessments are administered during the same testing window as standard operational tests.

3.4 Sources of Items and Metadata

3.4.1 ELA, Mathematics, Science, and Social Studies

The item development process for MI-Access utilizes the Michigan IBS as its main resource. The IBS contains items that have been developed and reviewed by Michigan teachers using processes described earlier in the chapter. The Michigan IBS is a secure, web-based application that allows users to create contexts and test items. It leads users through all the steps of the item development process, including context review, item review, and data review.

3.5 Import into DRC INSIGHT Test Engine

MI-Access FI is administered through the DRC INSIGHT test engine. The test items must be imported into INSIGHT from the IBS. Once the items are loaded into INSIGHT, they can be rendered for review in the identical formatting structure in which a student would see the item on a test. After the items have been formatted and rendered, they can be assembled into online test forms based on the sequence and information provided in the test maps.

3.6 Psychometric Review during Assessment Construction

Content specialists and psychometricians from MDE followed psychometric guidelines and targets for operational forms construction. The foremost guideline was for item content to match the test blueprint. Item flagging criteria (discussed below) were used to guide the assessment construction. Items with flags were avoided when possible.

Details for psychometric reviews are described below.

3.6.1 MI-Access Item Statistics Flagging Criteria

The psychometric review of the items on the fixed form was conducted by the MDE psychometrics team. MDE flagged items based on the following content criteria:

- The following items were flagged based on item difficulty and score distribution:
  - items with a low average item score or a low proportion obtaining the correct choice (i.e., adjusted *p*-value less than 0.33 for MC items, or adjusted *p*-value less than 0.10 for constructed-response (CR) and/or multi-point items)
  - items with a high average item score or a high proportion obtaining the correct choice (i.e., adjusted *p*-value greater than 0.90)

- The following items were flagged based on item discrimination:
  - items with a low item-total correlation (less than 0.20)
  - items with a higher mean criterion score for students in a lower score-point category

- The following MC items were flagged:
  - items where higher-ability students (those in the top 20% of the overall score) selected a distractor more often than they selected the key
  - items with a higher criterion score mean for students choosing a distractor than the mean for those choosing the key
  - items with a positive correlation between a distractor and the total score

Items were also classified into three Differential Item Functioning (DIF) (for corresponding details, see Chapter 10) categories. These were A, B, or C for MC items and AA, BB, or CC for CR items. As shown in the Chapter 10 DIF analysis result tables, the focus group was indicated by a positive value (such as C+ or CC+) and the reference group was noted with a negative value (such as C- or CC-). DIF comparison was not done if the sample size for either group was less than 30 students. For MI-Access FI assessments, items in the B or BB categories were flagged for moderate DIF and items categorized as C or CC were flagged for significant DIF.

DIF was evaluated for the following subgroup comparisons (focal – reference) for FI tests:

- Gender: Female – Male
- Race/Ethnicity: Black – White
- Economically Disadvantaged: Yes – No
- Accommodation: Yes – No

For MI-Access, all field-test items were reviewed by the data review panels regardless of whether an item was flagged. Items that were not flagged for content or bias statistical issues were eligible for use in the operational pools. Flagged items became eligible for the operational item pools if they were approved by the data review panel and the final review of the MI-Access content leads.

3.6.2 MI-Access Test Map Psychometric Review

For MI-Access test map development, the following analyses were carried out for psychometric review (note that the listed analyses are routine annual procedures):

1. Content expectation distribution check: This check is to ensure that operational items on each form have the desired content coverage (i.e., the reporting categories are the same as depicted in the test blueprint), and within each reporting category, the content standards have as much variety as possible. Moreover, across years, the distribution of content expectations or content strands is the same.

2. Item position check: For FI tests, equating items and common items (non-equating items that appear on multiple forms or across years) must appear in the same test positions across forms. Moreover, to control for possible position effect on item parameter estimation, equating items are checked to make sure they are within ±2 positions from the previous year's positions; for non-equating common operational items, differences in position across years are within ±5.

3. Across-year comparability check: For this check, distributions of item difficulty and item discrimination (*p*-values and adjusted item-total correlations) (see Chapter 7 for details) are checked across adjacent years for unique items to make sure they are comparable.

4. Across-mode comparability check for FI: Comparability of equating items and other operational items, including repeated operational items and unique operational items across mode (paper/pencil versus online), is checked using the same approaches as mentioned above in the across-year comparability check. Specifically, the MDE psychometrics team conducted the following:

    a. a content coverage homogeneity test (to make sure that equating items and other operational items have comparable content coverage)
    b. a comparability check of distributions of item difficulty and adjusted item-total correlation

    These analyses are conducted to make sure that the equating items function as a miniature test if possible—that is, they represent both the content and the statistics of the overall test.

5. Item key distribution check: This check involves all items on the test (operational and field-test items). Only MC items for FI and SR items for SI and P are involved in this check. For this check, the desired result is for all three key options to appear relatively equally on each test map, with no same-key option appearing three times consecutively. Although it is desirable to have unique field-test items on each form, if a field-test item must be repeated on multiple forms, a check is carried out to ensure that it appears in the same test position across forms and modes.

6. Overall operational item set quality check: This check ensures that no operational items have problematic flags. Specifically, DIF results are checked to make sure that, if possible, no equating operational items have “B” or “C” DIF flags. All operational items that appear on the final form are scrutinized to make sure that there are no bias or sensitivity issues involved. Moreover, adjusted item-total correlations, item statistics flags, and IRT item parameters for FI are also checked to determine whether items are free of concerns. Items are flagged if any of the following conditions is met: the key option-total correlation is negative, distractor option-total correlation is positive, omit-total correlation is positive, or key option percentage is not the highest. Item statistics are checked to ensure that the adjusted p-value should be within the normal range of >0.3 and <0.9; adjusted item-total correlation should be >=0.2; and there are no item statistics flags for equating items for FI.

The above test map review procedures occur throughout the entire process of test map development. At the very earliest stage—usually after MDE has finished the previous school year’s statistics analysis and the IBS statistics are ready for use for the current year’s tests—the lead psychometrician provides the content leads with the current year’s test map statistical targets for each content area by grade level. These targets include the mean adjusted *p*-value and mean adjusted item total correlation for equating items, non-equating common items, and all operational items combined for FI. These targets also include the mean adjusted *p*-value and mean adjusted item total correlation for operational items for SI and P. Next, the content leads select the equating items for FI (this step is skipped for P and SI), and the lead psychometrician reviews the statistical targets and the proposed equating items based on the procedures described above in procedures 1–6). After the MDE content leads finish the test map in the IBS and the lead psychometrician is notified to review the test map, the above procedures are implemented.

If any issues are found, the identified problems are documented and communicated to the content leads. Content leads then revise and resubmit the test map for another round of review. This iterative process continues until all issues have been resolved or the imperfect items are proven to be the best selections given various constraints, such as content coverage considerations and the need to avoid possible clueing.

3.7 Item Types Included

MI-Access FI uses traditional MC items on all test forms and CR items in ELA Expressing Ideas. MI-Access SI and P use “selected response” MC items with three options for SI and two options for P, along with activity-based observation items. Technology-enhanced items were not used for this assessment in 2019.

3.8 Field-Test Selection and Administration

3.8.1 Field-Test Item Selection

The OEAA content leads are tasked with selecting field-test items. The blueprints specify the number of field-test items by grade level and content area. The content leads work within Michigan's IBS to monitor the number of operational items available for each content expectation. Where there are gaps in the numbers available, content leads may decide to field-test items assessing that content expectation. The content leads also monitor the number of items that may be overexposed and need replacement as one way to select field-test items.

Responses on field-test items do not contribute to a student's score on the operational tests. The specific locations of the embedded items in the assessment are not disclosed. These data are free from the effects of differential student motivation that might characterize stand-alone field-test designs since the items are answered by students taking operational tests under standardized test administration procedures.

3.8.2 Field-Test Administration

MI-Access assessments consist entirely of MDE-developed operational and embedded field-test items for all grade levels and content areas.

The operational item set is the same across all online forms in a grade level, appearing in the same test positions. The remaining form positions are used for field-test items, which are unique to each form. The online forms in each grade are randomly administered to the student population.

For all content areas, the paper/pencil forms share the equating items with the online forms. Details on constructing forms are found in sections 3.9 and 3.10.

3.9 Online Form Building and Rendering Process

3.9.1 Overview of Rendering Process

DRC and MDE follow a very rigorous rendering process for all items on the 2019 MI-Access assessments. Using the web-based application LeanKit, DRC and MDE monitor the progress of each grade and content batch. The process begins right after the import of items from Michigan's IBS. All parts of the rendering process are completed at least one month prior to the start of testing to ensure time for User Acceptance Testing (UAT) of all grades and content areas. Figure 3-1 below shows the entire process for MI-Access FI items that are imported from the Michigan IBS.

Figure 3-1. Rendering Process of Michigan-Built Items

Requirements are established and reviewed with MDE prior to importing. The requirements include the QTI 2.2 import specs between the IBS and DRC's IDEAS system and the specific rules when importing each item. Detailed rendering requirements are also documented and reviewed.

3.9.2 Form Preparation and Rendering in INSIGHT

For all fixed forms, after the individual items are formatted and rendered, online test forms are assembled in the INSIGHT test engine based on the sequence and information provided in the test maps created by MDE. The test maps provide test-form data, item-form sequence location, and metadata (content standard, depth of knowledge, item position, p-value, item response theory parameters, answer key, and points possible) for each test form for each test type (program, content area, and grade level). DRC applies the appropriate styles and formatting to the fixed forms based on the previously set style and formatting guidelines.

The assembled fixed forms are then reviewed by content leads at DRC and MDE in a UAT setting to ensure that the forms match the exact design and data displayed in the test maps and that the forms, features, and functionality of INSIGHT appear and operate correctly. The UAT is conducted using the same INSIGHT test delivery system as the students use so the forms appear and function just as the students see them. The forms include features such as the online tools provided for each item, test directions, help files, calculators, and reference materials. Detailed information on student tools can be found in Chapter 4 of this report. UAT is an end-to-end process that validates every step, from student test registration to testing to data transfers to scoring data.

3.10 Paper/Pencil Form Building and Review Process

MI-Access FI testing is administered online 91% of the time, with paper/pencil tests available where they are instructionally appropriate, necessary for accommodations, or technologically necessary. Michigan offers the following accommodations delivered through paper/pencil assessments for students with disabilities and for English Learners: enlarged print, braille, and audio supports, such as audio CDs, human read-aloud, and live translations to a student's native language (for mathematics, science, and social studies). The MI-Access SI and P assessments are administered to students and scored by assessment administrators using SR and ABO item format items. Booklets and student-level picture cards are developed for use by assessment administrators in delivering assessments to students.

The MI-Access assessments are developed by OEAA's content leads using Michigan's IBS. The content leads review each item in the test map to check for text and/or graphic errors, clueing, correct answer keys, and a balance of answer keys. Once the test map is approved by the content lead, the psychometric lead reviews the test map in a similar way as detailed above for online forms but with more focus on comparability of paper/pencil forms to their online counterparts.

Once the test maps are approved by both the content lead and the psychometric lead, the composition unit creates one item per page (a "one-per") for review by both the OEAA content lead and the OEAA editor. A one-per is created for each item on the test map, showing how each item will appear in a test booklet. Content leads ensure the one-per matches the item in the IBS, which is the source of truth for each item. The item as it appears on the one-per must also follow OEAA's style guide and be free of errors.

After the content lead approves the one-pers, they are reviewed by OEAA's editor. Once the editor approves the one-pers, the OEAA's composition unit assembles the test booklets. There are several rounds of reviews conducted by OEAA content leads, OEAA assessment specialists, and OEAA's editor. Once the initial test booklets are approved, they are posted for printing by Measurement Incorporated. The paper/pencil test maps are provided to Measurement Incorporated for use in creating braille and enlarged print forms, a function subcontracted through the American Printing House for the Blind.

3.11 Summary

In summary, this chapter explicates the procedures used in the development of the MI-Access assessments. The efforts by MDE and its vendors address multiple best practices of the test industry. They are related to the following AERA, APA, and 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 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.

Chapter 4: Test Administration Plan

Chapter 4 reviews the test administration process for both the online and paper/pencil administrations of the MI-Access assessments. In 2019, MI-Access Functional Independence (FI) was administered online 91% of the time and on paper/pencil 9% of the time. MI-Access Supported Independence (SI) and Participation (P) are administered using paper/pencil versions of the test, and the student responses are entered using a DRC online answer document portal. Detailed information on supports, accommodations, test materials, and training and test security practices can be found throughout this chapter.

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" (p. 111). Chapter 4 of this report examines how test administration procedures implemented for MI-Access strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

The online platform components of eDIRECT and INSIGHT, which were necessary for all online test administrations, are discussed in section 4.4. The web-based application known as eDIRECT was used for all test preparation and test monitoring, while INSIGHT was the online test delivery system used by students when taking online assessments.

4.1 Universal Tools, Designated Supports, and Accommodations

A variety of testing tools are available across all grades, content areas, and modes of testing so that all students have the ability to fully demonstrate their knowledge and skills on the statewide assessments. The variety of tools offered attempts to ensure that a student's opportunity to demonstrate knowledge on a test is not negatively impacted by the student's disability or English language proficiency.

The Michigan Department of Education (MDE) categorizes tools into three levels: universal tools, designated supports, and accommodations. Universal tools can be used by students at their own discretion. Use of a designated support requires an educator to identify that support type for a student because of an instructional need. Tools listed as accommodations require that a student has an Individualized Education Program (IEP) or 504 Plan and that the need to use that support is identified within that document.

Regardless of the level of the tool type, MDE requires educators to make decisions about use on an individual basis. The decision for use should be based on the individual student's instructional needs for each content area. Some tools may be classified as nonstandard, in which case the use of those tools by students may result in invalid test scores. School districts may contact MDE if an IEP or 504 team wants to use an accommodation that is not on the approved list. MDE will consider allowing that accommodation for the current administration and in future administrations pending literature and research reviews and discussions with MDE's assessment content leads.

MDE's policies related to the use of accommodations are in compliance with AERA, APA, and NCME (2014) Standard 6.2, which 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. (p. 115)

Additional information about Michigan's accommodations framework and a list of which accommodations are considered allowable and valid for students to use can be found in the "Student Supports and Accommodations" table.[1]

4.1.1 Educator Guidelines

Many of the allowable designated supports and accommodations require educators to perform an action for the student or on behalf of the student. For example, a student needing a scribe may be provided one as long as the scribe is using the guidelines for scribing outlined in MDE's *Scribing Protocol*. Additional documents exist to ensure educators are providing these supports and accommodations in a consistent and reliable manner.

4.1.2 Accommodations Use Monitoring

MDE monitors Designated Supports and Accommodations used by students to ensure high reliability and validity of test results. Data audits include verification that students receiving Accommodations on the assessment had an Individualized Education Program or 504 plan. In the event that students received accommodations without an IEP or 504 plan, schools are contacted and asked to verify the use of Accommodations and make a plan to improve their process for future student use of Designated Supports and Accommodations. Starting with the next operational assessment, interviews will be conducted with schools after assessment monitoring to verify the decision-making processes used in providing Designated Supports and Accommodations to students for use on the assessment.

4.2 Online Accommodations

Appropriate accommodations, designated supports, and universal tools were available for students to use while taking the 2019 MI-Access FI assessment online. These accommodations and supports were required to be documented in the student's IEP, while universal tools were available to all students in the INSIGHT online test engine.

There were no embedded online accommodations used for the Spring 2019 MI-Access. An embedded online accommodation is one that is built into the test engine. There were accommodations available for online testing outside the test engine as follows.

- Directions provided by test administrator using American Sign Language (ASL) or Signed Exact English (SEE)
- Signing of test content in ASL or SEE—except for text designated as Do Not Read Aloud
- Use of abacuses
- Use of counters, coins, base-ten blocks, or other manipulatives for solving mathematics problems
- Use of an alternative communication device—that is, a computer with alternative access for an alternate response mode, such as a switch, alternative keyboard, eye-gaze motion sensor, voice recognition software, head or mouth pointer, or specialized trackball or mouse—when such tools successfully interacted with the test engine

The one embedded online designated support available for the MI-Access FI assessments is masking.

The non-embedded designated supports available for the online MI-Access FI assessments are listed below:

- Scribe (for non-writing items, using the *Scribing Protocol*)
- Noise buffers (e.g., ear mufflers, white noise, and/or other equipment to block external sounds)
- Auditory amplification devices or special sound systems
- Visual aids (e.g., closed-circuit television and magnification devices)
- Non-electronic bilingual word-to-word dictionary
- Augmentative/alternative communication devices (e.g., picture/symbol communication boards and speech-generating devices)

Text-to-speech was available to all students at all grades as a universal tool. Students or test administrators could control the volume and speed of this feature at any time. Items were scripted to provide alternate text for graphics, tables, and specific item elements that would violate the item construct if they were read aloud. The table below provides a list of the available embedded universal tools that were provided within the INSIGHT system by grade and content area.

Table 4-1. Available Tools for MI-Access in INSIGHT

Assessment	Grade	Text-to-Speech	Pointer	Highlighter	Magnifier	4-Function Calculator (Item-Level)
ELA	3	x	x	x	x	
ELA	4	x	x	x	x	
ELA	5	x	x	x	x	
ELA	6	x	x	x	x	
ELA	7	x	x	x	x	
ELA	8	x	x	x	x	
ELA	11	x	x	x	x	
Mathematics	3	x	x	x	x	x
Mathematics	4	x	x	x	x	x
Mathematics	5	x	x	x	x	x
Mathematics	6	x	x	x	x	x
Mathematics	7	x	x	x	x	x
Mathematics	8	x	x	x	x	x
Mathematics	11	x	x	x	x	x
Science	4	x	x	x	x	
Science	7	x	x	x	x	
Science	11	x	x	x	x	
Social Studies	5	x	x	x	x	
Social Studies	8	x	x	x	x	
Social Studies	11	x	x	x	x	

Figure 4-1 presents more details for DRC INSIGHT student tools.

Figure 4-1. DRC INSIGHT Student Tools

Some tools are available only on certain fixed forms or in certain content areas.

TOOL	DESCRIPTION/FUNCTION
Navigation Tools	
	**Back and Next**—Move to the next question or a previous question. (**Back** is only available in CAT within passage and listening sets.)
	**Go To Question**—Jump to any item or passage set on the test by choosing the item from a drop-down list (only available in fixed forms).
	**Pause**—Pause the test for a short period of time (e.g., restroom break) and resume upon return.
	**Flag**—Mark a question for review at a later point (only available in fixed forms).
	**Test Review**—Review and change answers by section and indicate whether the test is ready to be scored (only available in fixed forms).
Standard Test-Taking Tools (available at all times)	
	**Pointer**—Select, change, or unselect an answer option; select other user tools; and navigate through the test. When moved over an answer choice, the pointer converts to a pencil image.
	**Highlighter**—Highlight a portion of text or a graphic and remove highlights.
	**Magnifier**—Magnify/enlarge a portion of the screen (i.e., object, image, or text) by two times for better viewing.
	**Help**—The Help Library provides information on tool usage, test directions, helpful hints, and other topics. Also includes a "What's This?" feature that allows a student to access contextual help for a specific tool or button.
	**Sticky Note**—Creates and places a small note in which a student can type a short message for later reference (multiple notes can be created for each item or passage).
	**Calculator**—Basic four-function and scientific options are available as required, either individually or together.
	**Click to Enlarge**—Allows for large graphics by using a thumbnail image of the graphic that can be enlarged for viewing. Student can interact with the test item and other tools simultaneously.
Accommodations Tools (determined at the student level)	
	**Audio/Video tools**—Includes a **Text-to-Speech Synthesizer** that allows all test-related information (e.g., test directions, questions and answers, formula sheets) to be read aloud to the student. VSL fixed forms provide video for **sign language administration**.
	**Display Options**—Can be made available for all students or just those with a specific accommodation, such as **Color Overlays**, that allows a student to change the background color for text, graphics, and response areas.

4.3 Paper/Pencil Accommodations

Dozens of accommodations, designated supports, and universal tools are available for the MI-Access assessments, as listed in the "Supports and Accommodations" table. The list below shows the designated support and accommodation information that is tracked (bubbled in) on each content area's booklet for MI-Access FI. This is not a full list of allowable designated supports and accommodations; it is a list of only what MDE considers the most frequently used designated supports and accommodations. It does not include universal tools available to all students for paper/pencil assessments.

- Contracted braille
- Enlarged print/Use of word processor (Expressing Ideas section only)
- Read aloud
- Audio CD
- Scribe

4.4 Online Test Platform

The secure web-based test engine DRC INSIGHT Online Learning System was loaded on computers that students accessed for all online assessments, including MI-Access FI. Test items and forms could be accessed only by using a valid test ticket. It was suggested that automatic updates be turned to "Enable" so that the software updated as needed without manual updates. From the INSIGHT landing page, students had access to the test via the "Test Sign In" link and to the sample item sets via the "Online Tools Training" link.

DRC's client portal, eDIRECT, was used to manage the test setup functions of student assessments and to provide the installable downloads. The custom browser software was downloaded from eDIRECT and installed on student testing devices. The secure browser could be installed on computers individually or downloaded to a central location, copied, and distributed to multiple computers simultaneously using common network distribution tools. Everything needed for testing was found within the secure browser, eliminating the need for districts to coordinate updates to third-party software.

Technology coordinators installed local caching servers (a testing site manager (TSM) or Central Office Services (COS) Service Device) to manage the content (test content, responses, and audio files) and regulate traffic between testing sites and Data Recognition Corporation's (DRC's) servers. The System Readiness Check helped troubleshoot any issues that might occur during INSIGHT installation or while INSIGHT was running. This application is installed when INSIGHT is installed and performs a series of tests that can be used to diagnose and prevent or correct most errors.

The Load Simulation Tool was also available for sites to use for preplanning purposes. The software was used by technology coordinators to perform load simulation tests that helped estimate the amount of time it would take to download tests and upload responses based on the number of students testing at the same time, the current amount of network traffic, the amount of available bandwidth, and other site-specific factors.

The local caching software featured Load Balancing for monitoring content caching availability. Load Balancing solutions also allowed a district to quickly add or remove content servers when required without reconfiguring testing clients or redirecting or reassigning addresses. This tool also supplied an easier method to manage the distribution of testers between servers; each testing client was not dependent on a single server having enough capacity.

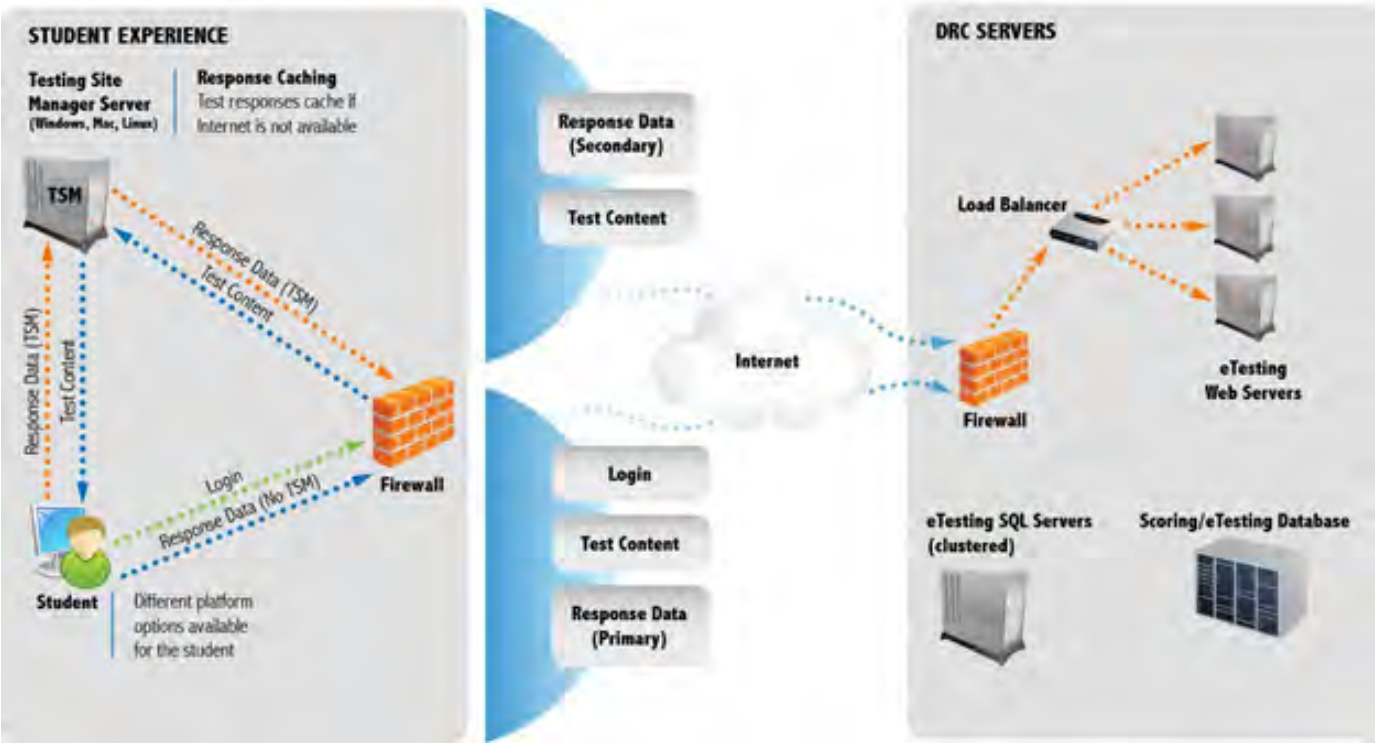
Prior to an assessments' operational use, DRC's quality assurance staff performed full system-level tests in an independent test environment that simulated the production configuration. Tests were run on all supported computer platforms and browsers and included a comprehensive review of system functionality, usability, reliability, security, and overall performance. Test content was also validated during this process.

Multiple methods were used to ensure secure data transfer, including encryption technologies and Secure Sockets Layer protocol through Secure Hypertext Transfer Protocol Secure. Test content was encrypted at the host server and remained encrypted throughout all network transmissions; content was decrypted only after the student login was validated. Decrypted test content on a student workstation was stored in memory only during each test session. After the session ended (that is, the test was completed or the student logged out), computer memory was purged to ensure the security of test content.

During testing, responses were sent to a DRC server each time the student navigated away from an item or clicked the Next button to submit an answer. Responses were saved automatically every 45 seconds during testing, when the student navigated away from an item, or when the student answered a selected-response item, whichever came first. If the student took longer than 45 seconds to answer an item, the incomplete response was submitted at 45-second intervals until the student completed the item. When the student returns to the test after a break or interruption, the student is returned to the point at which the student left off to avoid having to navigate through all previously answered questions.

Figure 4-2 illustrates the secure transfer of online test responses between the student and DRC.

Figure 4-2. Architecture of the Student Testing Experience



4.5 Test Administrator Training

On March 6, 2019, DRC, in conjunction with MDE, held a WebEx training presentation with district and school building coordinators and test administrators. The presentation included pertinent information for all MI-Access online testing. The presentation was recorded and posted to eDIRECT for Michigan users to reference throughout the testing window.

MDE held a New Assessment Coordinator Preconference Workshop for both paper/pencil and online M-STEP administrations at the 2019 Michigan School Testing Conference on February 12, 2019. This presentation provided detailed information for new assessment coordinators administering both the paper/pencil assessment and the Online assessment. This training was structured into before-, during-, and after-testing activities and included the following:

- Before Testing
  - Universal Tools, Designated Supports, and Accommodations
  - Pre-identification of students
  - Materials ordering
  - Providing training to test administrators and proctors
  - Scratch paper and calculator policies
  - How to prepare students for testing (MI-Access tutorials, Online Tools Training (OTTs))
  - Off-Site testing requirements and requests
  - eDIRECT training
  - Test security and the Assessment Integrity Guide (AIG)
  - Test materials and handling of secure materials

    - Test schedules and test session setup
    - How to address a testing irregularity

- During Testing
    - Test directions
    - Testing irregularities
    - Active monitoring during testing
    - Materials allowed/not allowed in a test session

- After Testing
    - Materials return
    - Preliminary reports
    - Data files
    - Final reports

MDE also provided a PowerPoint presentation that discussed what administrators should do before, during, and after MI-Access administration. This presentation was available on the MI-Access web page in the ”Assessment Training and Resources” section. MDE also held an “Update on MDE MI-Access Assessments” breakout session specifically for those involved with either coordination or administration of MI-Access.

4.6 Test Security

4.6.1 Overview

The primary goal of test security is to protect the integrity of the assessment and to assure that results are accurate and meaningful. The MDE Office of Educational Assessment and Accountability (OEAA) uses four test security goals to maintain the integrity of the Michigan’s assessment system. These goals include the following:

- To provide secure assessments that result in valid and reliable scores
- To adhere to high professional test administration standards
- To maintain consistency across all testing occasions and sites
- To protect the investment of resources, time, and energy

4.6.1.1 Prevention

Prevention of breaches in test security includes standards and best practices for test integrity and security aspects of the design, development, operation, and administration of MI-Access, both paper/pencil and online test administrations, to prevent irregularities from occurring. Operational and administrative security policies and procedures apply to both online and paper/pencil test administrations. Online student-facing testing (MI-Access FI) uses DRC’s INSIGHT Online Learning System. This is a secure browser that locks a student into the testing environment, preventing access to other applications or websites. The software must be installed on each device used for testing. Test content is held securely in a TSM, which is an encrypted local cache. The TSM also provides backup response storage in the event of network issues. All students are assigned to test sessions and require an individual test ticket for every online test session. For the SI and P assessments, a test session with test tickets is assigned

for the entry of scoring information online. Each ticket has a username and a unique password. Access to test tickets is controlled through DRC's eDIRECT site, and eDIRECT access is controlled through locally administered permissions in the OEAA Secure Site.

For the paper/pencil test administration, the OEAA and its vendor, Measurement Incorporated, design forms to assist the district and building assessment coordinators with the successful receipt and return of test materials. These forms provide security and accountability during fulfillment and distribution, test administration, and collection processes. Secure packaging and distribution of materials for MI-Access are provided to ensure prompt, accurate, and secure delivery of test materials to districts and schools. All materials that contain test questions (including other materials such as picture cards) or student responses are considered secure materials and must be handled in a way that maintains their security before, during, and after testing. As part of professional test administration practices, the OEAA provides test security resources for state, district, and school personnel to use in the prevention of testing irregularities. These include the *Assessment Integrity Guide* (AIG), test administration manuals (TAMs), online and paper/pencil administration directions, test security training modules, and incident reporting procedures.

All school staff members involved in testing are required to be trained in test administration and security prior to the opening of the assessment window. Training resources are available on a statewide basis. Districts and schools can customize trainings by role and location, using state-provided materials and including local plans.

The AIG is intended to be used by districts and schools in the fair and appropriate administration of state assessments. It includes guidelines on the expected professional conduct of educators who administer state assessments to ensure proper test administration and academic integrity.

Four assessment security training modules are available as a supplement to the AIG. The modules are intended to be used as an online training program for district and building assessment coordinators, test administrators, and test proctors. These modules explain why test security is important, describe different staff roles in test administration, and detail how to plan for and handle incidents that compromise test security.

Each assessment has a TAM that helps the staff administering the assessment understand how the administration process works, when specific assessment activities take place, what the roles of school personnel are in the administration process, and how to use available supports and accommodations. Test administrators have online and paper/pencil test directions to follow when administering MI-Access.

District assessment coordinators are required to file an incident report in the case of any testing irregularity. The incident reports are filed on the OEAA Secure Site. The test security specialist and other MDE assessment administrative staff review the incidents and determine what the required remediation will be through the use of internal and independent investigations.

4.6.1.2 Detection

Detection practices include guidelines for assessment monitoring, testing, and reporting of irregularities. Detection resources and practices include the AIG, incident reporting, random/targeted test administration monitoring, social media monitoring, and data forensic analysis. Districts are instructed to monitor test sessions for proper test administration and to enforce the policies and guidelines in the AIG to promote fair, approved, and standardized practices.

The OEAA uses random and targeted assessment monitoring to ensure the security and confidentiality of state assessments and to ensure testing personnel adhere to proper procedures. Targeted assessment monitoring is used when schools have had a previous irregularity or show unusual results from previous state assessment data analyses. Random assessment monitoring uses a sample of schools that are randomly selected for quality and integrity checks. Specific requirements of assessment monitoring are described in the *Assessment Observation Requirements Document* created with Measurement Incorporated. The AIG details the process for monitoring district and school personnel.

Internet and media monitoring occurs during testing windows. The goal of this monitoring is to combat breaches and any disclosure of secure assessment materials. These monitoring activities include monitoring comments on the internet for test items captured and shared, either from testing computer screens or from paper/pencil test booklets. Social media sites are also monitored for posts discussing or exposing test material. Requirements for social media monitoring are documented in the *Social Media Monitoring Requirements Document* created with Measurement Incorporated. The AIG details the process for monitoring the social media sites of district and school personnel.

During and after online and paper/pencil test administrations, the OEAA conducts multiple analyses on student assessment results. These statistical analyses help in the flagging of potential testing irregularities. The types of data forensic analyses used in Spring 2019 included unusual score gains and losses, online right-to-wrong changes, and paper/pencil erasure. Analyses performed on Spring 2019 data will provide a baseline for data forensics in future years.

4.6.1.3 Investigation and Remediation

District assessment coordinators are required to notify the OEAA as soon as they are made aware of an alleged or suspected violation or misadministration of MI-Access. Testing irregularities are reported to the OEAA via an online incident report form. The MI-Access TAM and AIG provide an incident reporting guide for districts and schools.

The OEAA also has a phone and online “tip line” for reporting of unethical behavior. Reports can be made anonymously. This provides a means for school staff members to report test integrity issues within their chain of command when they do not feel comfortable reporting the issues to their chain of command.

All incident reports and supporting documentation are reviewed by MDE, and a determination is made regarding the disposition of each incident. If the OEAA determines that the irregularity caused no consequences affecting security, validity, or fraud, and that the school took appropriate actions to correct the situation, the OEAA may consider the issue resolved and the case is logged and closed. If the OEAA determines that questions remain regarding the security, validity, or authenticity of the test administration, the OEAA will request either a school self-investigation or, if the problem is considered potentially severe, an independent investigation.

After investigations have taken place, the OEAA will create a summary report of the findings. Determination of the investigation is provided in the report.

Remediation of the incidents reported and investigated differ based on the severity of a confirmed allegation or misadministration. Minor mistakes receive recommendations of best practices. Isolated security incidents or negligence provide good candidates for targeted monitoring the next year. Individual student tests tainted by misadministration are typically invalidated. More serious incidents can lead to invalidating entire classes of tests, required retraining of the testing staff, or barring staff from participating in statewide testing. When possible, remediation happens within the testing window so that students can be retested if appropriate.

4.6.2 Online Test Security Practices

Test security is essential for obtaining reliable and valid scores for accountability purposes. All district assessment coordinators, building assessment coordinators, test administrators, proctors, and other staff who participate in MI-Access or handle secure assessment materials are required to receive the proper training for their role. Security training is provided through the AIG, MI-Access TAMs, and the test security training modules.

Test security training includes proper protocol to be followed before, during, and after test administration. The AIG, TAM, and the test administration directions embedded in the FI TAMs provide necessary information on the distribution, collection, and return of secure testing materials. The AIG provides information on self-monitoring of assessment administration practices, incident reporting, and monitoring conducted by the OEAA.

Each district is required to self-monitor the test administration practices within their district. Incident reporting by district assessment coordinators is required when there is any type of misadministration or problem with test administration. The OEAA monitors all test administrations. Each person involved in test administration is required to sign the OEAA Assessment Security and Confidentiality Agreement. Security training includes the handling and chain of custody for secure materials.

DRC's online test platform, INSIGHT, is a secure web browser that is downloaded to students' machines. Once launched, INSIGHT goes into "lockdown" mode and prevents students from accessing any other programs. The INSIGHT software is only accessible from 7:00 a.m. to 4:00 p.m. EST and is locked during all other times.

MDE has approved some testing sites to have an alternate INSIGHT availability window to test students at atypical hours; these sites are able to test via INSIGHT until 10:00 p.m. EST. On these sites—just like on similar testing sites— all student test tickets and student test rosters are considered secure materials and must be stored securely by test administrators when not in use.

DRC also provides MDE with online forensic telemetry data via a secure table data load. The table below references the data that are captured and sent to MDE on a weekly basis during the testing windows.

Table 4-2. INSIGHT Forensic Data

Attribute of Forensic Data	Description
Test Interrupted Stopped Flag	Test was interrupted/stopped
Test Interrupted Stopped Count	Number of times the test was interrupted/stopped
Total Item Time	Total time spent on an item
Item Visit Count	Total number of times the item was visited
Wrong to Right	Item's response was changed from wrong to right (within or across item visits)
Wrong to Right Count	Total number of times the item's response was changed from wrong to right (within or across item visits)
Right to Wrong	Item's response was changed from right to wrong (within or across item visits).
Right to Wrong Count	Total number of times the item's response was changed from right to wrong (within or across item visits)
Wrong to Wrong	Item's response was changed from wrong to wrong (within or across item visits).
Wrong to Wrong Item Count	Total number of times the item's response was changed from wrong to wrong (within or across item visits)
Total Enters Net Total Exits	Records total enters are greater than or less than total exits.

4.6.3 Paper/Pencil Test Security Practices

Test security is essential for obtaining reliable and valid scores for accountability purposes. All district assessment coordinators, building assessment coordinators, test administrators, proctors, and other staff who participate in MI-Access or handle secure assessment material are required to receive the proper training for their role. Security training is provided through the AIG, MI-Access TAM, and the test security training modules.

Test security training includes proper protocol to be followed before, during, and after test administration. The AIG and TAM provide necessary information on the distribution, collection, and return of secure testing materials. The AIG provides information on self-monitoring of assessment administration practices, incident reporting, and monitoring conducted by the OEAA.

Each district is required to self-monitor the test administration practices within its district. Incident reporting by district assessment coordinators is required when there is any type of misadministration or problem with test administration. The OEAA monitors all test administrations.

Each person involved in test administration is also required to sign the OEAA Assessment Security and Confidentiality Agreement. Security training includes the handling and chain of custody for secure materials. All materials that contain test questions or student responses are considered secure materials and must be handled in a way that maintains their security before, during, and after testing. Paper/pencil secure materials include the following:

- test booklets (for paper/pencil testing)
- assessment administrator test booklets (for SI and P assessments)
- student picture cards (for SI and P assessments)
- answer documents (for FI paper/pencil testing)
- accommodation materials
- scratch paper

Test materials are delivered about two weeks before the test cycle begins. Packaging lists are used to document orders. Schools are instructed to retain all secure materials in one secure, locked location within the school. During the test administration window, all secure materials must be securely stored unless being used for test administration. Building assessment coordinators are required to carry out the building-level duties related to the distribution, security, and collection of test materials. The test administrator is responsible for distributing and collecting test booklets, answer sheets, assessment administrator booklets, student picture cards, scratch paper, and accommodation materials used during administration and for delivering them to the building coordinator after each test session.

The OEAA provides training and guidance materials for local test administrators who have the duty of ensuring a secure testing environment. Before and during test administration, test administrators arrange the testing environment so that all visual cues are covered or removed.

For MI-Access FI, each student will have a test booklet. Assessment administrators will retain the answer document (which contains an individual barcode containing necessary test and student information) and will transfer the student responses from the test booklet to the answer document. Test administrators must be familiar with the test directions in the MI-Access FI TAM that must be read and followed. The test administrator is required to remain in the testing room at all times. Students are not permitted to access any electronic devices used for communication, capturing images, or data storage. Lists of professional and prohibited test security practices are available in the AIG.

For the MI-Access SI and P assessments, assessment administrators will use the assessment administrator test booklets, picture cards and/or objects, or other materials that are familiar to the student. Primary and shadow assessment administrators will use these materials, along with the scoring documents that correspond to each grade level and content area, to administer and score the assessment.

Schools are required to return all secure materials. The exceptions to this are scratch paper and the scoring documents used by the primary and shadow assessment administrators when observing and scoring the items. Scratch paper is to be destroyed after each testing session. Once the scores are entered online, the scoring documents are kept on file at the school with the security agreements. Schools are provided a return kit for secure materials.

When returned materials arrive at Measurement Incorporated, the boxes are scanned, logged, and checked against the material tracking information for each school or district. Boxes and all their contents are scanned, repackaged, and warehoused. All discrepancies between the secure materials sent and returned are noted and become part of the report to inform schools/districts of any missing materials. Several rounds of contact are attempted to account for every piece of missing secure materials. Schools with excess missing materials may receive targeted monitoring in future years to check local controls.

Measurement Incorporated makes scanned images of documents available to the OEAA and retains warehoused documents for the length of records retention. Paper/pencil documents are reviewed for secure disposal five years after the end of testing, with the written permission of the OEAA director. Electronic files are kept in a highly secure location with off-site backup. Files include, but are not limited to, scanned images, scanned scored files, import and export files, and all student testing data. All electronic files are available to the OEAA, and no student testing data are deleted without written permission from the OEAA director.

4.7 Summary of M-Access Administration Best Practices

The elements discussed in previous sections not only align with MDE prevention practices that help maintain the integrity of the assessment but also adhere to the testing practices and AERA, APA, & NCME (2014) *Standards* relevant to test administration. The previous sections also demonstrate how information in the MDE trainings and manuals addresses the following 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. (p. 90)

The MI-Access TAM and AIG provide instructions for before-, during-, and after-testing activities, with sufficient detail and clarity to support reliable test administrations by qualified test administrators. To ensure uniform administration conditions throughout the state, instructions in the TAM and AIG describe the following: general rules of online testing; pause rules; scheduling of 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 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. (p. 114)

To ensure the usefulness and interpretability of test scores and to minimize sources of construct-irrelevant variance, it is essential that the MI-Access is administered according to the prescribed TAM and AIG.

MDE's protocol, discussed in section 4.6 of this report stresses incident reporting and adheres to the following standards:

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

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. (p. 116)

Standard 6.7 Test users have the responsibility of protecting the security of test materials at all times. (p. 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 accordingly noted. Detailed information about test security procedures are presented in section 4.6.

4.8 Test Materials

A list of available test materials can be found below in Table 4-3.

Table 4-3. MI-Access Paper/Pencil Test Materials

Material Description	Product Type
Blank Labels	Ancillary
FedEx Return Air Bills	Ancillary
Instruction for Materials Return	Ancillary
OEAA Security Compliance Form	Ancillary
Outgoing Box Labels	Ancillary
Packing List Enclosed Label	Ancillary
Picture Card Security Reminder Sheet (SI & P)	Ancillary
Pre-ID Labels (FI)	Ancillary
Return Kit Cover Sheet	Ancillary
Special Handling Envelopes	Ancillary
Answer Document, by content area and grade (FI)	Answer Document
Assessment Administrator Booklet for Braille (AABB), by content area and grade (FI)	Braille
Braille Kit, by content area and grade (Answer Document, Braille Test Book, AABB, and Kit Cover Sheet) (FI)	Braille
CD Kit, by content area and grade (Audio CD, Test Booklet(s), Answer Document(s), Kit Cover Sheet) (FI)	CD
Enlarged Print Kit, by content area and grade (Answer Document, Enlarged Print Test Book, Test Booklet, Kit Cover Sheet) (FI)	Enlarged Print
Accessing Print Listening Script, by grade (FI)	Listening Script
Picture Cards, by grade (SI & P)	Picture Cards

Material Description	Product Type
Scoring Documents, by content area and grade (SI & P)	Scoring Document
Assessment Administrator Booklets, by grade (SI & P)	Test Booklet
Test Booklet, by content area and grade (FI)	Test Booklet

4.9 Summary

In summary, the overall purpose of each test administration workshop and the ancillary materials is to keep districts informed about policies and procedures related to testing in general and to the MI-Access program in particular. The information imparted is clearly related to maintaining the integrity of the administration of MI-Access, 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 MDE and its test vendors are in alignment with multiple best practices of the testing industry but are particularly related to the following standards (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 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 6.6—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.

Chapter 5: Test Delivery and Administration

5.1 Online Administration Details

In Spring 2019, the Michigan Department of Education (MDE), in conjunction with Data Recognition Corporation (DRC), delivered 91% of MI-Access Functional Independence (FI) assessments online via DRC's online testing platform, INSIGHT. During that testing period, 614 Michigan school districts administered MI-Access FI online to 1,777 Michigan schools.

MI-Access FI English language arts (ELA) (Accessing Print and Using Language), mathematics, science, and social studies were administered as fixed forms, just as they were in Spring 2019.

The Spring 2019 MI-Access FI was administered to enrolled students in grades 3–8 and 11. Table 5-1 presents the content areas tested by grade.

Table 5-1. Content Areas Tested by Grade

Grade Tested	Content Areas Tested
Grade 3	ELA and Mathematics
Grade 4	ELA, Mathematics, and Science
Grade 5	ELA, Mathematics, and Social Studies
Grade 6	ELA and Mathematics
Grade 7	ELA, Mathematics, and Science
Grade 8	ELA, Mathematics, and Social Studies
Grade 11	ELA, Mathematics, Science, and Social Studies

The number of students tested online for the Spring 2019 MI-Access FI can be found in Table 5-2 below.

Table 5-2. Number of Students Tested Online, MI-Access Functional Independence

Grade	Content	Online Students Tested
3	ELA (Accessing Print and Using Language)	932
4	ELA (Accessing Print and Using Language)	1,089
5	ELA (Accessing Print and Using Language)	1,142
6	ELA (Accessing Print and Using Language)	1,353
7	ELA (Accessing Print and Using Language)	1,301
8	ELA (Accessing Print and Using Language)	1,288
11	ELA (Accessing Print and Using Language)	1,165
3	Mathematics	936
4	Mathematics	1,103
5	Mathematics	1,184
6	Mathematics	1,415

Grade	Content	Online Students Tested
7	Mathematics	1,382
8	Mathematics	1,351
11	Mathematics	1,237
4	Science	952
7	Science	1,311
11	Science	1,246
5	Social Studies	1,127
8	Social Studies	1,344
11	Social Studies	1,249

5.1.1 Online Administration Reports

Prior to administering the 2019 assessments, DRC and MDE outlined requirements for all online administration reporting. Administration reports were delivered to MDE daily or weekly based on the established requirements. Table 5-3 shows the types of administration reports that were delivered to MDE during the 2019 MI-Access FI testing window.

Table 5-3. Online Administration Reports

Report Name	Delivery Frequency	Description of Report
After-Hours Report	Daily throughout the testing window	Shows online tests that have test login times and/or stop times within the defined after-hours time
Accommodations-Supports Report	Daily throughout the testing window	Shows assigned accommodations and supports at the student level
Form Distribution Report	Weekly throughout the testing window	Shows fixed-form assignments for monitoring equal distribution of fixed forms per grade and content area
Testing Times Report	Daily throughout the testing window	Daily summary of testing times to allow MDE to monitor how long students take to complete tests
Cumulative Student Status	Daily throughout the testing window	Status of student testing by site; allows MDE to monitor how students are progressing with testing by grade and content area

5.1.2 Online User Manuals and Reference Documents

To help assist with the administration of the online MI-Access FI assessments, DRC and MDE created numerous manuals and documents. . These include the test administration manuals (which includes test directions by grade), the *Technology User Guide*, and many additional reference documents.

5.2 Paper/Pencil Administration Details

MDE delivered MI-Access Supported Independence (SI) and Participation (P) entirely as paper/pencil assessments, with an online answer portal for schools to submit answers. MDE delivered MI-Access FI as paper/pencil tests for students in school that applied and were approved for a waiver of online testing and for individual students at the school's discretion.

Online testing waivers were available for the following reasons:

- Buildings were not technologically ready.
- Buildings were under construction or had otherwise disrupted technological environment.
- Locations were testing in a center-based program.
- Locations were testing in a juvenile justice facility.
- Buildings had other instructional reasons.

Individual students with accommodations that required a paper/pencil assessment were also administered the paper/pencil test, as well as any student for whom the instructional team considered paper/pencil testing more appropriate.

The paper/pencil test was available in enlarged print and in both contracted and uncontracted braille versions. The paper/pencil test also included support options such as live translation and read aloud, as described in Chapter 4 of this report.

There were three forms for each FI test, including the braille form. These forms are listed in the table below. For MI-Access SI and P, there were two forms, with each form serving as an emergency form for the other.

Table 5-4. Paper/Pencil Test Forms by Content Area

Content Area	Paper/Pencil Forms Available
ELA	Form 1 – administered to all students testing paper/pencil
	Form 2 of online test – Emergency form
	Braille form (FI only)
Mathematics	Form 1 – administered to all students testing paper/pencil
	Form 2 of online test – Emergency form
	Braille form (FI only)
Science	Form 1 – administered to all students testing paper/pencil
	Form 2 of online test – Emergency form
	Braille form (FI only)
Social Studies (FI only)	Form 1 – administered to all students testing paper/pencil
	Form 2 of online test – Emergency form
	Braille form

The MI-Access FI paper/pencil test was provided for the same grades and content areas that had online counterparts (see Table 5-1). All tests for MI-Access SI and P were paper/pencil test formats, composed of selected response items (using picture cards) and activity-based observations. The grade levels and content areas match Table 5-1 for these levels with the exception of social studies, for which students were administered a locally determined assessment.

The number of students tested using paper/pencil MI-Access assessments can be found in the table below. All MI-Access FI students took the Expressing Ideas portion of the ELA assessment on paper; the FI ELA counts in Table 5-5 represent the students who took the entire ELA assessment, not just the Expressing Ideas portion, on paper.

Table 5-5. Numbers of MI-Access Students Tested with Paper/Pencil

Grade	Content Area	Number of Students Tested with Paper/Pencil FI	Number of Students Tested with Paper/Pencil SI	Number of Students Tested with Paper/Pencil P
3	ELA	109	483	487
4	ELA	109	448	406
5	ELA	107	394	396
6	ELA	100	467	363
7	ELA	118	460	328
8	ELA	128	449	334
11	ELA	138	467	375
3	Mathematics	107	479	488
4	Mathematics	108	446	406
5	Mathematics	109	390	396
6	Mathematics	107	469	364
7	Mathematics	124	458	330
8	Mathematics	132	452	335
11	Mathematics	138	467	376
4	Science	99	446	407
7	Science	111	455	324
11	Science	139	467	374
5	Social Studies	107	NA	NA
8	Social Studies	130	NA	NA
11	Social Studies	140	NA	NA

5.3 eDIRECT

5.3.1 Michigan Users

DRC uses the MDE Secure Site to pull and load Michigan users to eDIRECT, based on Secure Site Test Cycle IDs. For the 2018–19 school year, the MI-Access FI Test Cycle ID was 177. Users are identified by their Security Role IDs and pulled into eDIRECT according to the established requirements. The mapping of users from the Secure Site to eDIRECT can be found below in Table 5-6.

Table 5-6. Mapping of Building Users from Secure Site to eDIRECT

Security Role ID	eDIRECT Role and Permission Set
17 – Public School Administrator	School
20 – District Administrator	School
40 – Public Online Test Administrator	School
31 – Nonpublic School Administrator	School
41 – Private School Online Test Administrator	School
42 – District Test Administrator	School
45 – State	State
38 – District Technology Coordinator	District Technology Coordinator
39 – School Technology	District Technology Coordinator
43 – Public School Technology	District Technology Coordinator
44 – Private School Technology	District Technology Coordinator

All users are identified by the site code(s) they have access to within eDIRECT. Users are only able to access student and test information by using their site permissions in the MDE Secure Site.

5.3.2 Administrative Functions

Online administration is managed through the DRC eDIRECT client portal, which provides tiered, secure access to all required administrative functions. Within eDIRECT, users manage student information and create test sessions.

Student information for MI-Access FI is imported into eDIRECT via automatic loading of data. DRC utilizes the MDE Secure Site to pull new and updated student records for import into eDIRECT. Student data is pulled three times a day so that any new student records or updated student records are loaded in a timely manner. Building users are able to view all the demographic information associated with the students from the Secure Site before placing them in test sessions for test tickets.

Once the student data is loaded into the Test Setup application within eDIRECT, users organize students into test sessions. Test sessions can be created by content area, class, grade, or school. Through Test Setup, users can also update student accommodation information, print test tickets, and monitor student testing status.

The student login ticket contains unique login credentials used by the student to access the testing software. For a selected test session, users can download and print a PDF document containing instructions, a roster of student tickets, and the actual test tickets. Student test tickets are considered secure materials, and test administrators are required to keep printed tickets in a predetermined, locked, secure storage area.

5.3.3 Online Testing Resources

eDIRECT houses an assortment of testing resources available to the district and school users and to the technology coordinators. The INSIGHT installables and requirements are maintained on eDIRECT, as are all technology guides and information necessary for setting up schools' computers and servers.

Video tutorials containing mini chapters on how to use eDIRECT applications are available to help users familiarize themselves with the different administrative applications within eDIRECT. An eDIRECT user guide is also available for reference.

For more information on MDE-specific online testing resources, visit the OEAA website.[1]

5.4 MDE Secure Site

The MDE Secure Site is a web-based application used for state assessments and accountability. The primary functions of the Secure Site include pre-identification of students for both paper/pencil and online assessments; ordering paper/pencil tests, including accommodated versions of the assessments; incident reporting; review of accountable students and test verification; and retrieval of data score files and score reports. The Secure Site also supports requests for online testing waivers for schools unable to administer that mode of testing and requests for students testing off-site.

The Secure Site is available only to authorized district and school personnel with sign-on credentials. The MDE Secure Site training page[2] includes a complete list of Secure Site functions and how to use them.

5.5 Return Material Processing

Each box of materials shipped to schools contains a box list, showing each item in the box. Each order contains a packing list, which shows a complete list of items, quantities, and box location for the entire order. When an order contains secure materials, a security list is also included, which shows a complete list of secure items and the associated shrink-wrapped pack barcodes.

At the end of testing, all MI-Access scorable and non-scorable testing materials are to be returned to Measurement Incorporated for processing, via FedEx Express Saver.

[1] http://www.michigan.gov/oeaa/
[2] https://www.michigan.gov/mde/0,4615,7-140-22709_57003---,00.html

When boxes of returned materials arrive at Measurement Incorporated, the warehouse team scans the boxes into the Measurement Incorporated tracking system database, where they are checked against the tracking numbers that are assigned to each school. FedEx also scan each of its tracking barcodes to record each box as it was delivered to Measurement Incorporated. This provides immediate information on the number of boxes received and points of origin of the boxes. Once this procedure is completed, the boxes are opened, and all materials are sorted.

Scorable and non-scorable materials are securely scanned in using Measurement Incorporated's Security Barcode Check-In Application. This application allows Measurement Incorporated IT Operations to scan the security identifier on individual secure materials or the security identifier located on the outside of an intact pack of shrink-wrapped documents, using Measurement Incorporated's automated security scanning process. Scanning the security identifier on the shrink-wrapped pack is equivalent to scanning all the individual security identifiers included in the pack and is more efficient than scanning each individual test booklet in the shrink-wrapped pack.

As each security identifier is securely scanned, it is checked against the original list of identifiers that were entered into the Measurement Incorporated database. Any discrepancies are noted, and a security report is generated for MDE.

For scorable answer documents, the same scanning process that captured the security identifier information also captures information from the student pre-ID label, the bubbled demographic information on the answer document cover, the bubbled student responses, and images of constructed responses to be sent on to handscoring.

All loose (individual) test booklets are securely scanned into the Measurement Incorporated database by Measurement Incorporated IT Operations using Measurement Incorporated's automated security scanners.

Warehouse personnel securely scan all returned accommodated materials, using a human-operated computer station equipped with a barcode reader; these materials are entered into the ObjectTracker database.

The accommodated materials include CDs, braille test booklets, assessment administrator braille booklets, and enlarged print test booklets. Although they are not accommodated materials, ELA Listening Scripts for MI-Access FI and picture cards for SI and P are also scanned.

After all returned secure materials are checked in, Measurement Incorporated's IT team prepares the initial security report data by comparing the security barcodes of checked-in materials with the barcodes of all secure materials.

The initial missing materials and security report data are provided to MDE in a spreadsheet. All schools that were sent materials by Measurement Incorporated are included in the summary, regardless of whether the schools are active or inactive entities.

For public school districts that are missing secure materials, security reports are shipped to district coordinators to be further distributed to building coordinators.

For public school academies and nonpublic schools that are missing secure materials, a security report is shipped to each building coordinator.

Missing materials reported as destroyed or never received are not included on the security report sent to the district or school. Missing materials reported as lost remain on the security report, and the comment “Reported Lost” is added to the comment section of the security report.

FedEx Ground Package Returns Program labels are provided in case any secure materials need to be returned. Schools that find no additional secure materials are directed to return the summaries of missing secure materials and any additional information.

The Measurement Incorporated IT team updates the security report data using the spreadsheet of issues reported to the call center, which includes materials that were lost, destroyed, or never received. This spreadsheet is maintained by the Measurement Incorporated management team. MDE staff forwards to the Measurement Incorporated management team any information collected via phone calls or incident reports regarding materials that were lost, destroyed, or never received.

If a summary of missing secure materials is accompanied by a corresponding explanation letter, the two are stapled together. All summaries of missing secure materials are checked in using the district/building code barcode and are filed in order by assessment, district code, and building code. Any returned secure materials are checked in by security barcode and are stored with the other secure materials.

After the initial response window ends and the returned letters and secure materials are processed, the IT team refreshes the security report data for each assessment, indicating schools that responded with newly returned secure materials and/or letters and schools that did not respond. Follow-up security reports are generated.

A second round of cover letters and security reports is sent to districts and schools that still have outstanding missing materials and have not returned a letter or a security report with comments. This procedure is the same as used for the first round of security reports. Schools that return a letter, materials, or both in the first round are not included in the second round.

Measurement Incorporated checks in and files any returned summaries of missing secure materials, secure materials, and additional information received. When MDE determines that schools have had sufficient time to respond, Measurement Incorporated generates and provides to MDE a final missing materials report.

The final security report spreadsheet sent from Measurement Incorporated to MDE includes all schools and districts that were tested. The Excel filter feature is used to list schools that still have outstanding missing materials. The “Returned Letter or Additional Items or Both” column reflects letters and items returned in response to both the initial round and the second round of security reports.

Tables 5-7 through 5-10 show MI-Access material information. The numbers in the Table 5-7 are (and are expected to be) higher than the number of students testing on paper/pencil. Each student needs at least two secure materials for testing; additionally, some secure accommodated materials are needed for students testing online. The numbers of SI and P materials are shown in one table (Table 5-10) because test booklets cover all content areas in each grade.

Table 5-7. Number of Secure MI-Access FI Materials Shipped

Grade	ELA	Mathematics	Science	Social Studies
FI 3	2,755	388	NA	NA
FI 4	3,205	435	408	NA
FI 5	3,200	439	NA	432
FI 6	3,300	374	NA	NA
FI 7	3,306	444	411	NA
FI 8	3,416	494	NA	492
FI 11	3,453	567	557	562

Table 5-8. Number of Secure MI-Access FI Materials Returned

Grade	ELA	Mathematics	Science	Social Studies
FI 3	2,671	386	NA	NA
FI 4	3,103	430	406	NA
FI 5	3,087	428	NA	425
FI 6	3,189	367	NA	NA
FI 7	3,194	435	403	NA
FI 8	3,258	478	NA	484
FI 11	3,232	549	539	544

Table 5-9. Number of Secure MI-Access FI Materials Not Returned

Grade	ELA	Mathematics	Science	Social Studies
FI 3	84	2	NA	NA
FI 4	102	5	2	NA
FI 5	113	11	NA	7
FI 6	111	7	NA	NA
FI 7	112	9	8	NA
FI 8	158	16	NA	8
FI 11	221	18	18	18

Table 5-10. Number of Secure MI-Access SI and P Materials

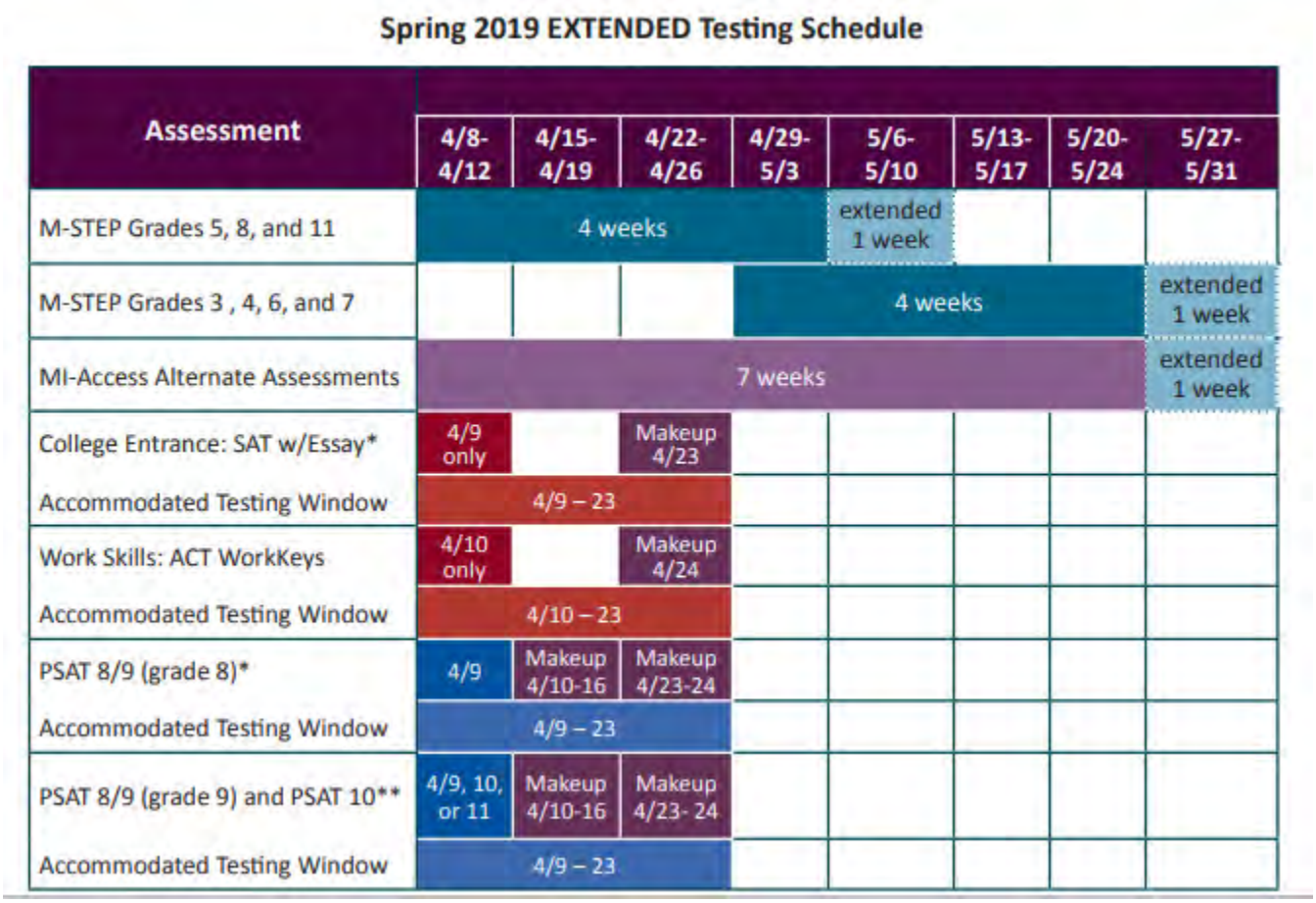
Grade	Shipped	Returned	Not Returned
SI 3	968	934	34
SI 4	927	833	94
SI 5	782	712	70
SI 6	810	782	28
SI 7	793	556	237
SI 8	755	709	46
SI 11	725	478	247
P 3	898	883	15
P 4	807	780	27
P 5	723	700	23
P 6	713	693	20
P 7	634	606	28
P 8	610	597	13
P 11	689	672	17

5.6 Testing Window and Length of Assessment

The testing window for the 2019 operational MI-Access assessments began Monday, April 8, 2019, and was scheduled through Friday, May 24, 2019. Due to inclement weather that interrupted instruction statewide, a cabinet-level decision was made to extend the testing window by one week to Friday, May 31, 2019 (see Figure 5-1) All online and paper/pencil assessments were administered in this time frame; there were no specific makeup windows for online assessments.

The Spring 2019 MI-Access assessments were not timed and were paced by students. Schools scheduled test sessions and determined the appropriate amount of time for students to spend testing in a single session. Any students needing more time were able to complete the test in a later test session during the four-week grade-level testing window. Further information on test session timing is provided on pages 10–11 of the 2018–2019 Guide to State Assessments.

Figure 5-1. Calendar of Extended Testing



Chapter 6: Scoring

Chapter 6 shows how MI-Access scoring adhered to the AERA, APA, & NCME standards. Standard 4.18 provides some general guidance for Chapter 6:

> 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. (p. 91)

Chapter 6 explains the procedures used for scoring multiple-choice (MC), selected-response (SR), and activity-based observation (ABO) items, as well as handscoring constructed-response (CR) items. To preserve the integrity of the items for future use, the scoring criteria used for each item are not presented in this chapter.

6.1 Online Scoring

The online scoring process for MI-Access FI includes the scoring of MC items, in which students choose only one correct answer from choices A–C. The items are scored against a scoring key that was prepared and validated before the start of each testing window. Responses to MC items were captured during the online test administration, and items were scored as "right," "wrong," or "blank" (not answered). Additional answer key checks were conducted during the testing windows to ensure that the items were scored based on the provided key.

6.2 Handscoring

Measurement Incorporated performed all required scoring of paper/pencil constructed-response items. For the MI-Access Functional Independence (FI) English language arts (ELA) Expressing Ideas portion, these items included written text and/or drawn response items for grades 3–8 and 11. MI-Access FI items were scored by readers working in Taylor, Michigan.

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. (p. 92)

Sections 6.2.1 through 6.2.5 explain how scorers are selected and trained for the MI-Access FI handscoring process. Sections 6.2.6 and 6.2.7 describe how the scorers are monitored throughout the MI-Access FI handscoring process.

6.2.1 Security

All Measurement Incorporated scoring rooms are designated secure areas with stringent security regulations that are vigorously enforced. Measurement Incorporated routinely implements several measures to help safeguard the security of student responses while they are in Measurement Incorporated's possession and to maintain the confidentiality of student identity.

In the scoring rooms, the use of cellphones, tablets, MP3 players, laptops, or recording or photographic equipment is prohibited. The copying of materials for anything other than the training purposes expressly permitted by the Michigan Department of Education (MDE) is prohibited.

All buildings that house student responses—including Measurement Incorporated headquarters, scoring centers, and warehouses—utilize an electronic security system during nonbusiness hours.

All readers scoring remotely are required to work from a private, password-protected environment. No free or public Wi-Fi can be used. Readers can access a project website (VSC Score) only from a secure, password-protected network. Readers cannot access any project website from a public computer or a public network, such as a wireless network at a hotel or restaurant. While in VSC Score, readers are unable to take screenshots or access email or other applications. Maintaining a secure workstation is a condition for employment for all remote employees.

Before receiving any training materials, all scoring project staff are required to sign a confidentiality and proprietary agreement; the agreement indicates that no participant in training and/or scoring may reveal any specific information about the test or about the criteria and methods for scoring to any person as part of the contractual obligation to score student responses.

At scoring centers, all training materials remain on the premises during a project and are collected at the end of each workday to be secured. All materials are collected and accounted for at the end of the scoring project.

Readers who score remotely access training materials from an online resource library. The software does not allow readers to print or download data.

No identifying student information is provided on the images sent to readers via VSC Score software.

Readers do not have the ability to access training materials or student responses unless they and their team leader are logged on to the system.

Violation of any portion of the Measurement Incorporated security policy results in termination.

6.2.2 Measurement Incorporated Reader and Team Leader Hiring

Measurement Incorporated recruits, interviews, and hires a pool of readers to ensure ample staff for scoring projects.

All readers must have a minimum of a bachelor's degree. The names, demographics, educational backgrounds, and experience (including scoring experience) of all readers can be provided to MDE by Measurement Incorporated. Reader degrees are verified before the applicants are interviewed. Applicants must provide either an official transcript with a seal (no copies accepted), an official letter from a registrar's office (which would be mailed to the site manager), or access to a third-party company such as Parchment or Student Clearing House. Reader applicants can also bring their original diploma with a seal when they come for an interview.

Team leaders are selected and recruited from Measurement Incorporated's experienced reader staff. Each team leader supervises a group of 10–12 readers during live scoring.

6.2.3 Preparation of Training Materials for MI-Access FI

Three sets of student responses were used in training readers and team leaders:

- Anchor sets consisted of typical student responses at each score point, with examples of a response that would barely earn that point, a median response for that point, and a high response within that point without quite reaching the next point. These sets were used to show readers and team leaders how the rubric was applied to each response.
- Training sets consisted of atypical student responses and were used to further demonstrate application of the rubric to actual student responses.
- Qualifying sets consisted of student responses similar to those in the anchor and training sets. These sets were used for readers to demonstrate their understanding of the application of the rubric to student responses.
- Measurement Incorporated scoring directors used MDE-approved training materials. Anchor sets consisted of three responses at each score point. Each response was annotated to explain how the rubric criteria were applied. Training sets contained 5–10 papers. There was a training set for each trait for analytic scoring and a training set that combined the traits. The responses in each of these sets were arranged in random score-point order, and all score points were represented.

6.2.4 Training and Qualifying Reader and Team Leader

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. (p. 118)

Readers and team leaders were trained by the scoring director on the scoring criteria approved by MDE and were required to achieve qualifying standards set by MDE. Readers were divided into teams consisting of one team leader and 10–15 readers.

The scoring director presented the items and anchor sets and then discussed each score point as readers and team leaders took notes. Following the presentation of these anchor sets, readers and team leaders scored a training set and then one or two qualifying sets.

Readers and team leaders were provided a copy of anchor sets, training sets, and qualifying sets. Readers and team leaders were required to refer to the anchor sets and their notes when scoring training sets and qualifying sets.

Readers and team leaders scored the qualifying sets and submitted their scores. The percentage of correct scores was recorded. After each set was completed, the scoring director discussed the set with the group.

If a particular response or type of response generated numerous questions across teams, the scoring director discussed the problem with the group or posted a note to chat to ensure that everyone heard the same explanation.

Once the group finished discussing the first qualifying set, the readers and team leaders scored the next set. Training continued until all training sets and qualifying sets were scored and discussed.

Readers were required to demonstrate their ability to score accurately by attaining the qualifying agreement percentage approved by MDE before they gained access to actual student responses.

Any reader or team leader unable to meet the qualifying standards set by MDE was released. Tables 6-1 and 6-2 provide the number of qualifying sets per item and the qualifying standard.

Table 6-1. Qualifying Sets

Content	Number of Qualifying Sets per Item
FI Expressing Ideas	2

Table 6-2. Qualifying Standards

Score-Point Range	Qualifying Standard (Exact Agreement)
0–4	70%; no non-adjacent scores

6.2.5 Virtual Scoring Center

Measurement Incorporated used its VSC Score system for the image-based scoring of paper/pencil responses and for the scoring of online responses transferred to Measurement Incorporated from Data Recognition Corporation.

Readers and team leaders accessed the VSC Score system through a secure web-based interface with the use of a unique user ID and password. Each team leader and reader was assigned a unique number for easy identification of his or her scoring work throughout the scoring session. VSC Score enabled readers and team leaders to score only those items that they were trained and qualified to score.

Each CR item was randomly assigned to be read by one reader. A random sample (10%) of all student responses was then randomly assigned to a second reader. VSC Score managed readers' individual workloads and allowed readers to review and submit their scores.

Readers were trained on how to use the VSC Score performance assessment scoring system—how to assign scores; how to adjust the image for legibility; and how to "flag" responses that were atypical from the anchor sets, training sets, and qualifying sets for review by the team lead and scoring director.

Readers logged in and checked out a scoring set of student responses. This scoring set was generated by randomly selecting student responses from the pool of unscored student responses. A reader evaluated the first response, entered the score by clicking the appropriate value on the scoring toolbar, and clicked the "submit" button. The next response in the scoring set then appeared for the reader to score and submit. This process continued until all responses in the set had been scored. After scoring all responses in a set, the reader had the option to review any of the responses and modify the scores before submitting them to the system.

Once the scores were submitted, the set was checked in and responses were routed to other qualified readers as necessary. The requirements for subsequent readings were defined in the system during setup, and a student response was not marked as complete until the requisite number of independent readers had scored the response.

When a reader had a question about a response, he or she could transfer the image (along with the question and/or comments) from the current scoring set to a review set, which was assigned to a team leader. The team leader could forward the question to the scoring director, submit the appropriate score, or return the response to the reader with comments. This procedure was used whenever a reader had scoring concerns or encountered apparent non-scorable responses. Readers could mark completely blank responses as non-scorable, but otherwise only scoring directors or the project director could assign a non-scorable condition code to a student response.

6.2.6 Quality Control and Reliability of Scoring

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. (p. 118)

Section 6.2.6 explains the monitoring procedures that Measurement Incorporated uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics are available for all CR items and specify the criteria for scoring these items. These rubrics will not be presented in this report in order to preserve the integrity of the items for use in future test forms.

MDE reader production and reliability statistics, including reader training results, were available to MDE via a suite of VSC reports, which could be accessed online using secure credentials supplied to MDE staff.

Detailed Reader Status Reports were generated for each scoring project, utilizing a comprehensive system for collecting and analyzing score data. Daily analyses of the Reader Status Reports alerted management personnel to individual or group retraining needs.

After the readers' scores were submitted in the VSC Score system, the data was uploaded into the primary Scoring Resource Center servers. The scores were then validated and processed.

Updated real-time reports that showed both daily and cumulative data (project-to-date data) were available 24 hours a day via a secure website. The reports included data on the number of responses scored by each reader, the percentage of responses scored that day in exact agreement or adjacent agreement with a second reader, and the total number of responses scored at each score point.

For MI-Access FI CR scoring, a random sample of 10% of all student responses were scored a second time to generate agreement data.

Readers were required to consistently demonstrate the ability to assign scores according to the rubric and anchor papers that were introduced during training. Their scoring accuracy was under scrutiny using validity responses that were included daily with the actual student responses (for details, see section 6.2.7).

If questionable reader reliability indications were found, the affected responses were scored again.

The monitoring and retraining process was sustained throughout the project to promote strict adherence to MDE-approved scoring criteria and consistency throughout the scoring effort.

Scoring directors and team leaders provided consistent monitoring of the scoring patterns of each reader throughout the project, responded to questions, spot-checked (read behind) reader scoring, provided feedback, and counseled readers who were having difficulty with the criteria.

Scoring directors continued to look for atypical types of responses that were not covered in the initial training and presented further instruction about handling these types of responses when necessary.

6.2.7 Validity

Measurement Incorporated used validity responses, similar to the student responses found in the qualifying sets, during live scoring to monitor readers' accuracy in scoring. Preselected validity responses were approved by MDE. Scoring directors also had the ability to select live responses as validity responses, which were also subject to MDE approval. The true scores for these responses were entered into a validity database.

Validity responses were randomly incorporated into readers' sets each day of the project. Team leaders reviewed the validity results and provided feedback to the readers.

A validity report was generated, which included the response identification number, the scores assigned by the readers, and the "true" scores. Measurement Incorporated provided MDE with daily and project-to-date summaries of what percentages of papers scored by readers matched the validity checks or were high or low at each score point. Of the responses that a reader scored, 5% were validity papers. These responses appeared to the reader daily throughout the entire scoring project. The validity standards can be found in Table 6-3.

Table 6-3. Validity Standards

Score-Point Range	Validity Standard (Exact Agreement)
0–1	90%
0–2	80%
0–3	80%
0–4	70%

6.2.8 Alerts

Measurement Incorporated implemented a formal process for notifying MDE when student responses reflected a possibly dangerous situation for the student. Such situations could include responses indicating endangerment, abuse, or psychological and/or emotional difficulties.

Measurement Incorporated also alerted MDE if there appeared to be possible instances of teacher or proctor interference or student collusion with other students.

Measurement Incorporated took immediate action following a scoring alert.

6.3 Observation-Based Scoring

The MI-Access Supported Independence (SI) and Participation (P) assessments were scored by a primary and a shadow assessment administrator. The administrators observed a student's response to either an SR prompt, for which the student chooses between pictures or objects, or an ABO item, for which the student responds to an assessment prompt within a routine or instructional activity. MDE offered detailed online training on the administration and scoring of the SI and P assessments, available year-round on the MI-Access web page (https://www.michigan.gov/mi-access).

The primary assessment administrator (PAA) started the testing with an administrator assessment booklet and picture cards for each student being tested. The PAA and the shadow assessment administrator (SAA) used scoring documents provided in the assessment materials. The scoring documents were used to tally the student responses during administration; the responses were then transferred to the online answer document after testing was complete. Each scoring document contained the rubric in the header of the sheet. This was designed for easy reference during the observations.

6.3.1 SI/P Selected Response Items

SR items have three components:

- the item stem (or question)
- the scoring focus (a short statement that links the item to the Extended Grade Level Content Expectation, Extended High School Context Expectation, or Extended Benchmark being measured)
- picture answer choices

The P items had two picture answer choices, and the SI items had three picture answer choices. The assessment administrator test booklet and the P/SI test administration manual (TAM) gave specific instructions on how this type of item was to be administered. In some cases, based on item construct, parts of an item were prohibited from being read. The do-not-read guidance was outlined on the inside front cover of the assessment administrator test booklet.

Picture cards followed specific presentation styles. On the P assessments, there were two answer choices. Both picture cards were presented at the same time in one order, then immediately presented to the student again with the positions of the cards reversed. The student needed to respond correctly according to the rubric (see section 6.3.3) both times to receive full points for the item. Varying options for presenting the item were outlined in the P/SI TAM to account for varying student abilities and limitations.

6.3.2 SI/P Activity-Based Observation Items

ABO items, which were used on portions of the P/SI ELA and mathematics assessments and on portions of the P science assessments, were designed to reflect activities that typically take place in the classroom and with which students are most likely to be familiar.

For example, if an ELA word recognition item required a student to identify one or two words associated with a lunchtime routine, the item was observed as the student helped prepare a meal. If a mathematics item required the student to identify a missing object, the item was observed as the student took part in a table-setting routine with a necessary utensil missing. In this way, the assessment item was integrated into—or became part of—the normal instructional routine. With ABOs, assessment administrators were asked to present items the same way they would during a routine instructional activity for the students.

6.3.3 Scoring Selected-Response and Activity-Based Observation Items

Both item formats—SR and ABO—were scored using a standardized scoring rubric. During the assessment, the PAA recorded his or her scores or condition codes on the MI-Access PAA scoring document, while the SAA simultaneously and independently recorded his or her scores or condition codes on the MI-Access SAA Scoring Document. Once all the items had been administered, the PAA recorded the PAA and SAA score points and/or condition codes on the online student answer document.

6.3.3.1 Participation Scoring Rubric (3-Point Rubric)

The scoring rubric for the P assessments has three score points and three condition codes. The rubric is based on a student responding correctly and takes into consideration the amount of assistance the student requires to engage in the item. This is done to allow the regular instructional activity to continue and to avoid administering the item outside the context of a routine or instructional activity. Figure 6-1 below details the P score points and condition codes. Additionally, Figure 6-2 shows how to apply the rubric during assessment administration. The student's score for an item is the sum of the score given by the PAA and SAA.

6.3.3.2 Supported Independence Scoring Rubric (2-Point Rubric)

The scoring rubric for MI-Access SI is similar to the P scoring rubric except it has only two score points with the same three condition codes. The SI rubric is based on the student responding correctly and takes into consideration the amount of assistance the student requires to engage in the item. Again, this is done to allow the regular instructional activity to continue and to avoid administering the item outside the context of a routine or instructional activity. Figure 6-1 shows the SI score points and condition codes. Additionally, Figure 6-3 shows how the rubric is applied during assessment administration. Both the PAA and the SAA observe and score the student independently and simultaneously.

Figure 6-1. MI-Access P/SI Scoring Rubrics

Participation Score Point/Condition Code	**Supported Independence** Score Point/Condition Code	**Response**
3	2	Responds correctly with no assessment administrator assistance
2	1	Responds correctly after assessment administrator provides verbal/physical cues
1	Not Allowed in SI	Responds correctly after assessment administrator provides modeling, short of hand-over-hand assistance
A	A	Incorrect response
B	B	Resists/Refuses
C	C	Assessment administrator provides step-by-step directions and/or hand-over-hand assistance

Figure 6-2. Participation Scoring Rubric Flowchart

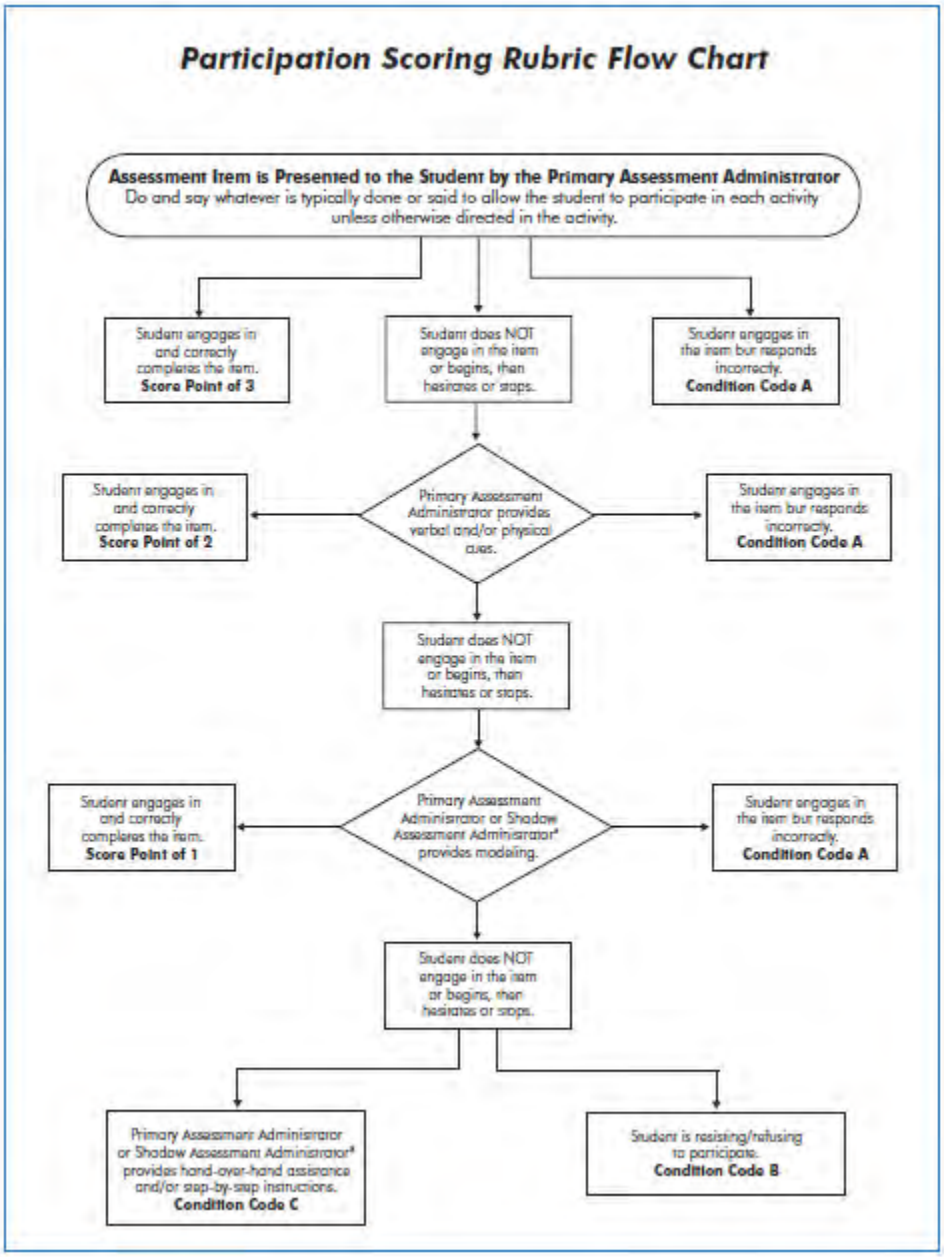
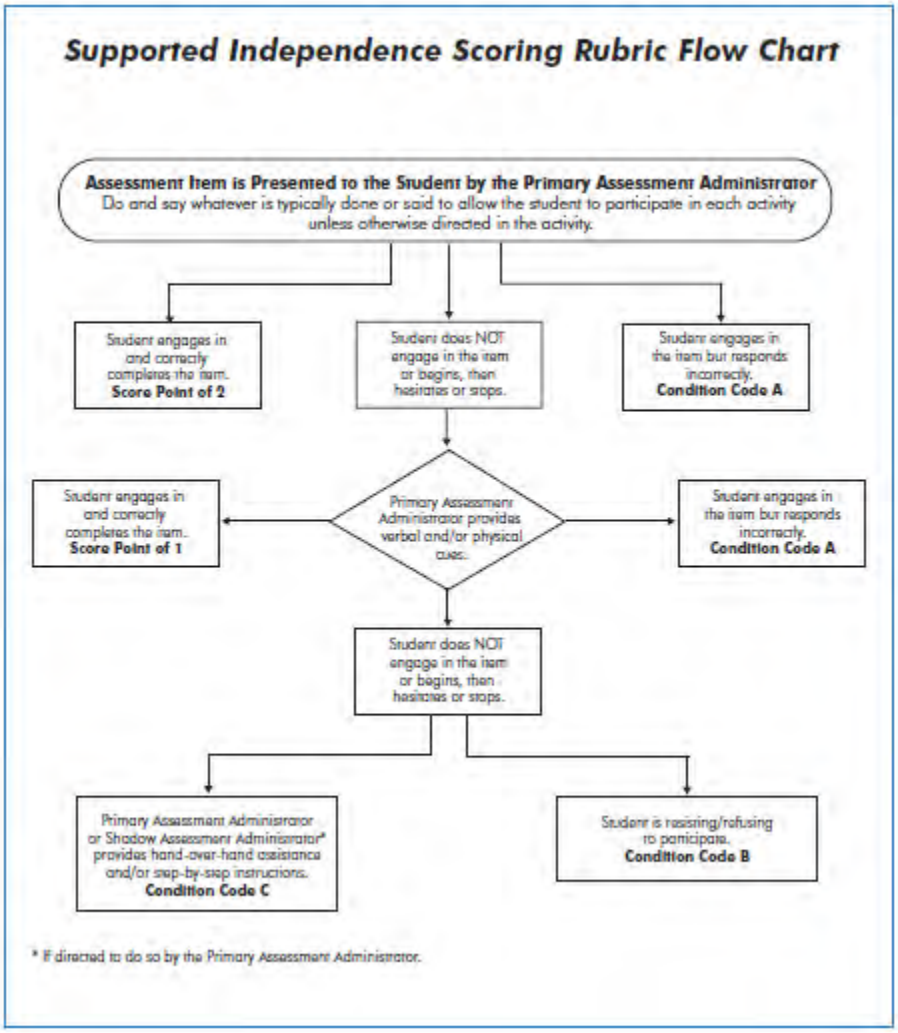


Figure 6-3. Supported Independence Scoring Rubric Flow Chart



6.4 Summary

The information presented in this chapter summarizes the scoring procedures for different types of items and the steps taken by DRC and Measurement Incorporated to ensure accuracy in scoring each item type for MI-Access. The reliability statistics presented in sections 6.2.7 and 6.3 demonstrate that the items were scored reliably. These efforts follow multiple best practices of the testing industry and are particularly related to 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.

Chapter 7: Operational Data Analyses

This chapter describes the analyses conducted with the operational (OP) data. Item/test analyses from both the Classical Test Theory (CTT) and the item response theory (IRT) frameworks are used (when appropriate) and reported here.

This chapter demonstrates adherence of MI-Access to AERA, APA, & NCME (2014) *Standards* 1.8, 5.2, 5.13, and 5.15. Each standard will be explicated within the appropriate section of this chapter. Standard 7.2 provides general guidance that is relevant to this chapter:

> The population for whom a test is intended and specifications for the test should be documented. (p. 126)

Chapter 3 of this report presents the test specifications. Information regarding reported data is discussed in detail in Chapter 8.

7.1 Operational Analysis of MI-Access

MI-Access is composed of three levels with different ranges of complexity and difficulty: Functional Independence (FI), Supported Independence (SI), and Participation (P). In other words, the three groups of students constitute the population for MI-Access.

Because only FI tests are scaled and scored using an IRT model, this chapter will report the operational analysis of the results based on the IRT model and the results based on the CTT for all FI tests. For MI-Access SI and P, only the CTT-based analysis will be provided. The FI results appear first, followed by SI and P.

7.1.1 Test-Level Analysis

This section presents the test-level summary statistics, the minimum observed score point (Min), and the maximum possible points (Max). The Max is equivalent to the number of operational items for MI-Access FI mathematics, science, and social studies because all items for these content areas are dichotomously scored. For FI ELA, there is one Expressing Ideas (EI) constructed response (CR) item (with score points ranging from 0 to 4). The total score reflects the summation of thirty Accessing Print and Using Language (APUL) multiple-choice (MC) OP items plus one EI CR OP item.

Since the OP items are the same across the online forms, the statistics for the online mode in Tables 7-1 through 7-4 represent all the students who took any online test form.

Table 7-1 provides the FI ELA raw score descriptive statistics, which include the number (N) of students taking a certain mode of test (either online fixed form or paper/pencil form), the raw score (Mean), the standard deviation (SD), and the minimum (Min) and maximum (Max) score points earned. For FI ELA APUL, there are three online OP forms and one paper/pencil form. For EI, there are two paper/pencil forms. The FI ELA test scores combine FI APUL and EI scores (i.e., an online form APUL with a paper/pencil form EI and a paper/pencil form APUL with a paper/pencil form EI). The mean raw score for FI ELA ranged from about 21 to 25 points.

Table 7-1. Test-Level Descriptive Statistics by Mode: FI ELA Raw Score

Grade	N	Mode	Mean	SD	Min	Max
3	901	Online	21.95	5.98	2	34
3	109	Paper	22.03	6.25	2	34
4	1059	Online	22.18	5.71	3	34
4	108	Paper	20.80	6.42	6	33
5	1129	Online	22.71	5.86	6	33
5	106	Paper	23.31	5.70	8	33
6	1340	Online	22.84	5.97	6	34
6	98	Paper	22.57	5.97	10	34
7	1292	Online	23.18	5.63	4	34
7	118	Paper	23.20	6.02	9	33
8	1280	Online	24.38	6.06	6	34
8	127	Paper	24.42	5.77	6	33
11	1120	Online	24.39	5.70	7	34
11	137	Paper	24.71	5.74	8	34

Table 7-2 provides the FI mathematics raw score descriptive statistics, which include the number (N) of students taking a certain mode of test (either online fixed form or paper/pencil form), the raw score (Mean), the standard deviation (SD), and the minimum (Min) and maximum (Max) score points earned. For MI-Access FI mathematics, there are three online fixed forms and one paper/pencil form. The mean raw score for FI mathematics ranged from about 13 to 16 points.

Table 7-2. Test-Level Descriptive Statistics by Mode: FI Mathematics Raw Score

Grade	N	Mode	Mean	SD	Min	Max
3	936	Online	15.15	4.88	2	24
3	107	Paper	15.72	4.76	4	24
4	1103	Online	14.69	4.56	2	24
4	107	Paper	15.25	4.83	6	24
5	1185	Online	15.34	4.88	2	24
5	108	Paper	15.10	4.93	3	24
6	1415	Online	14.26	4.26	4	24
6	105	Paper	15.08	4.74	7	24
7	1382	Online	14.74	4.23	1	24
7	123	Paper	14.80	4.81	1	23
8	1351	Online	14.74	4.94	0	24
8	131	Paper	15.49	5.12	3	24
11	1238	Online	13.68	4.87	1	24
11	137	Paper	14.28	4.88	2	24

Table 7-3 provides the FI science raw score descriptive statistics, which include the number (N) of students taking a certain mode of test (either online fixed form or paper/pencil form), the raw score (Mean), the standard deviation (SD), and the minimum (Min) and maximum (Max) score points earned. MI-Access FI science was administered to grades 4, 7, and 11, with two online fixed forms and one paper/pencil form. The mean raw score for FI science ranged from about 21 to 28 points.

Table 7-3. Test-Level Descriptive Statistics by Mode: FI Science Raw Score

Grade	N	Mode	Mean	SD	Min	Max
4	937	Online	21.33	6.66	6	35
4	98	Paper	21.41	6.98	9	34
7	1297	Online	24.85	6.62	7	40
7	110	Paper	26.40	6.80	8	38
11	1205	Online	27.80	8.19	8	45
11	138	Paper	27.55	8.49	10	44

Table 7-4 provides the FI social studies raw score descriptive statistics, which include the number (N) of students taking a certain mode of test (either online fixed form or paper/pencil form), the raw score (Mean), the standard deviation (SD), and the minimum (Min) and maximum (Max) score points earned. MI-Access FI social studies was administered to grades 5, 8, and 11, with three online fixed forms and one paper/pencil form. The mean raw score for FI social studies ranged from about 17 to 24 points.

Table 7-4. Test-Level Descriptive Statistics by Mode: FI Social Studies Raw Score

Grade	N	Mode	Mean	SD	Min	Max
5	1118	Online	17.46	6.14	1	32
5	106	Paper	18.27	5.54	7	28
8	1323	Online	18.17	6.04	5	33
8	129	Paper	18.09	6.80	1	32
11	1206	Online	22.42	7.49	4	41
11	139	Paper	23.62	8.23	4	41

Tables 7-5 through 7-8 present the FI scale score descriptive statistics, which include the mean scale score, standard deviation, and minimum and maximum scale score points earned by content area, grade, and mode. Like Tables 7-1 through 7-4, Tables 7-5 through 7-8 present the statistics for all the students who took any online test form.

As shown in these tables, mean scale scores across the two modes are generally very similar within a grade level for FI ELA, science, but with some differences for some grades for FI mathematics and social studies.

Table 7-5. Test-Level Descriptive Statistics by Mode: FI ELA Scale Score

Grade	N	Mode	Mean	SD	Min	Max
3	901	Online	2304.57	18.58	2234	2383
3	109	Paper	2305.23	21.09	2234	2383
4	1059	Online	2413.68	18.05	2349	2491
4	108	Paper	2410.00	19.97	2366	2469
5	1129	Online	2513.99	19.83	2463	2576
5	106	Paper	2515.96	19.88	2471	2576
6	1340	Online	2621.74	21.38	2569	2700
6	98	Paper	2620.90	21.24	2583	2700
7	1292	Online	2720.44	19.89	2658	2800
7	118	Paper	2721.19	21.69	2678	2779
8	1280	Online	2825.26	22.79	2768	2900
8	127	Paper	2824.57	20.86	2768	2878
11	1120	Online	3181.05	32.83	3097	3298
11	137	Paper	3183.66	34.81	3103	3298

Table 7-6. Test-Level Descriptive Statistics by Mode: FI Mathematics Scale Score

Grade	N	Mode	Mean	SD	Min	Max
3	936	Online	2304.57	18.58	2234	2383
3	107	Paper	2305.23	21.09	2234	2383
4	1103	Online	2413.68	18.05	2349	2491
4	107	Paper	2410.00	19.97	2366	2469
5	1185	Online	2513.99	19.83	2463	2576
5	108	Paper	2515.96	19.88	2471	2576
6	1415	Online	2621.74	21.38	2569	2700
6	105	Paper	2620.90	21.24	2583	2700
7	1382	Online	2720.44	19.89	2658	2800
7	123	Paper	2721.19	21.69	2678	2779
8	1351	Online	2825.26	22.79	2768	2900
8	131	Paper	2824.57	20.86	2768	2878
11	1238	Online	3181.05	32.83	3097	3298
11	137	Paper	3183.66	34.81	3103	3298

Table 7-7. Test-Level Descriptive Statistics by Mode: FI Science Scale Score

Grade	N	Mode	Mean	SD	Min	Max
4	937	Online	2400.57	18.51	2360	2476
4	98	Paper	2403.80	19.44	2373	2456
7	1297	Online	2700.12	15.79	2659	2777
7	110	Paper	2701.70	16.47	2661	2741
11	1205	Online	3101.66	18.11	3061	3182
11	138	Paper	3100.25	19.70	3065	3159

Table 7-8. Test-Level Descriptive Statistics by Mode: FI Social Studies Scale Score

Grade	N	Mode	Mean	SD	Min	Max
5	1118	Online	2492.59	17.66	2425	2573
5	106	Paper	2494.92	14.57	2465	2525
8	1323	Online	2796.62	18.14	2757	2883
8	129	Paper	2795.53	20.81	2723	2858
11	1206	Online	3097.61	17.49	3051	3184
11	139	Paper	3101.01	22.14	3050	3184

For MI-Access SI and P tests, which contain selected-response (SR) and activity-based observation (ABO) items, each item has a primary rater's score and a secondary (shadow) rater's score, as described in Chapter 6. For both SI and P, the reported raw scores reflect the summation of the two raters' scores. For SI, the possible raw scores for each item range from 0 to 4. For P, the possible raw scores for each item range from 0 to 6. Tables 7-9 through 7-14 provide the test-level descriptive statistics for both SI and P by content area and grade level. The mean raw scores ranged approximately from 39 to 42 for SI ELA, from 39 to 41 for SI mathematics, and from roughly 45 to 50 for SI science. The mean raw scores ranged approximately from roughly 31 to 35 for P ELA, from roughly 31 to 35 for P mathematics, and from roughly 50 to 54 for P science.

Table 7-9. Test-Level Descriptive Statistics by Grade: SI ELA Raw Score

Grade	N	Mean	SD	Min	Max
3	483	40.67	13.22	0	60
4	448	39.89	12.80	0	60
5	394	38.98	14.99	0	60
6	467	41.46	13.06	0	60
7	460	39.90	12.50	0	60
8	449	41.67	12.94	2	60
11	467	39.92	13.17	0	60

Table 7-10. Test-Level Descriptive Statistics by Grade: SI Mathematics Raw Score

Grade	N	Mean	SD	Min	Max
3	479	39.07	14.57	0	60
4	446	36.63	14.41	0	60
5	390	36.05	13.36	0	60
6	469	33.50	13.93	0	60
7	458	35.29	12.90	2	60
8	452	35.23	13.10	4	60
11	467	41.15	14.64	0	60

Table 7-11. Test-Level Descriptive Statistics by Grade: SI Science Raw Score

Grade	N	Mean	SD	Min	Max
4	446	47.96	15.43	0	68
7	455	45.32	15.09	4	68
11	467	50.01	14.84	0	68

Table 7-12. Test-Level Descriptive Statistics by Grade: P ELA Raw Score

Grade	N	Mean	SD	Min	Max
3	487	32.55	16.59	0	60
4	406	35.22	17.17	0	60
5	396	33.08	16.44	0	60
6	363	31.43	16.26	0	60
7	328	32.14	16.44	0	60
8	334	34.08	15.46	0	60
11	375	32.01	17.66	0	60

Table 7-13. Test-Level Descriptive Statistics by Grade: P Mathematics Raw Score

Grade	N	Mean	SD	Min	Max
3	488	31.07	17.38	0	60
4	406	30.53	16.84	0	60
5	396	30.80	15.99	0	60
6	364	31.21	16.79	0	60
7	330	31.61	17.13	0	60
8	335	34.91	15.85	0	60
11	376	30.70	17.36	0	60

Table 7-14. Test-Level Descriptive Statistics by Grade: P Science Raw Score

Grade	N	Mean	SD	Min	Max
4	407	53.59	26.78	0	90
7	324	50.39	25.30	0	90
11	374	50.67	27.36	0	90

7.2 Item-Level Analysis

This section presents various item-level statistics for all OP items on the Spring 2019 MI-Access. Specifically, item difficulty and adjusted item-total correlations defined by the CTT are reported here.

MI-Access FI mathematics, science, and social studies items on the Spring 2019 MI-Access tests were dichotomously scored, and the *p*-value was computed as an indicator for item difficulty. The *p*-value equals the proportion of students who answered an item correctly. A high *p*-value means that an item is easy, and a low *p*-value means that an item is difficult. For MC items, the *p*-value and the adjusted *p*-value are exactly the same, in order to be consistent across all content areas, the adjusted *p*-value is used instead of the *p*-value for reporting the item difficulty indicator here. For FI ELA, because there was one EI CR OP item with scores ranging from 0 to 4, an adjusted *p*-value is used as an indicator for item difficulty. The adjusted *p*-value shows the percentage of points the students obtained. It is calculated by dividing the

item mean by the maximum points possible.

The adjusted item-total correlation is an index of the association between students' performance on an item and their performance on the test as a whole; however, the item of interest is excluded from the total raw score. A high adjusted item-total correlation is desired because high correlations indicate that students with high scores on all other test items (i.e., students with high ability) tend to answer the item correctly, while students with low scores on all other test items (i.e., students with low ability) tend to answer the item incorrectly.

The item-level descriptive statistics by content area, grade, and mode for all OP items on the Spring 2019 MI-Access FI appear below in Tables 7-15 through 7-22. As shown in these tables, the average difficulty indicator (adjusted *p*-value or *p*-value) and discrimination indicator (item-total correlation) across modes within a content area by grade level are generally very similar, although for the discrimination indicator, there appeared to be some differences across the two mode tests. One possible reason for this is that sample sizes for the paper/pencil tests are very small, ranging from about one hundred to about two hundred for FI tests.

Table 7-15. Item-Level Descriptive Statistics by Mode: FI ELA Adjusted *p*-Value

Grade	N OP Items	Mode	Mean	SD	Min	Max
3	31	Online	0.66	0.13	0.39	0.88
3	31	Paper	0.67	0.12	0.39	0.86
4	31	Online	0.67	0.13	0.44	0.88
4	31	Paper	0.63	0.11	0.45	0.81
5	31	Online	0.69	0.13	0.39	0.89
5	31	Paper	0.71	0.12	0.47	0.91
6	31	Online	0.69	0.14	0.44	0.89
6	31	Paper	0.69	0.13	0.37	0.89
7	31	Online	0.70	0.14	0.46	0.88
7	31	Paper	0.70	0.12	0.46	0.90
8	31	Online	0.73	0.11	0.53	0.93
8	31	Paper	0.74	0.11	0.53	0.89
11	31	Online	0.73	0.11	0.53	0.92
11	31	Paper	0.74	0.12	0.53	0.91

Table 7-16. Item-Level Descriptive Statistics by Mode: FI Mathematics *p*-Value

Grade	N OP Items	Mode	Mean	SD	Min	Max
3	24	Online	0.64	0.14	0.36	0.88
3	24	Paper	0.65	0.14	0.36	0.87
4	24	Online	0.62	0.11	0.38	0.84
4	24	Paper	0.64	0.10	0.50	0.89
5	24	Online	0.64	0.10	0.45	0.82
5	24	Paper	0.63	0.12	0.38	0.83
6	24	Online	0.60	0.13	0.38	0.89
6	24	Paper	0.63	0.12	0.43	0.89
7	24	Online	0.62	0.15	0.36	0.86
7	24	Paper	0.62	0.12	0.37	0.83
8	24	Online	0.62	0.11	0.43	0.83
8	24	Paper	0.65	0.09	0.51	0.83
11	24	Online	0.58	0.12	0.36	0.86
11	24	Paper	0.59	0.12	0.42	0.81

Table 7-17. Item-Level Descriptive Statistics by Mode: FI Science *p*-Value

Grade	N OP Items	Mode	Mean	SD	Min	Max
4	35	Online	0.61	0.12	0.41	0.94
4	35	Paper	0.61	0.11	0.37	0.88
7	40	Online	0.62	0.16	0.35	0.90
7	40	Paper	0.66	0.14	0.32	0.92
11	45	Online	0.62	0.12	0.39	0.88
11	45	Paper	0.61	0.12	0.43	0.86

Table 7-18. Item-Level Descriptive Statistics by Mode: FI Social Studies *p*-Value

Grade	N OP Items	Mode	Mean	SD	Min	Max
5	32	Online	0.55	0.09	0.36	0.73
5	32	Paper	0.57	0.13	0.30	0.77
8	33	Online	0.55	0.09	0.34	0.73
8	33	Paper	0.55	0.08	0.38	0.68
11	41	Online	0.55	0.10	0.39	0.76
11	41	Paper	0.58	0.11	0.40	0.82

Table 7-19. Item-Level Descriptive Statistics by Mode: FI ELA Adjusted Item-Total Correlation

Grade	N OP Items	Mode	Mean	SD	Min	Max
3	31	Online	0.35	0.07	0.19	0.48
3	31	Paper	0.37	0.12	0.09	0.55
4	31	Online	0.33	0.07	0.17	0.46
4	31	Paper	0.36	0.11	0.05	0.57
5	31	Online	0.36	0.08	0.16	0.47
5	31	Paper	0.35	0.12	0.08	0.57
6	31	Online	0.37	0.07	0.20	0.48
6	31	Paper	0.37	0.12	0.17	0.63
7	31	Online	0.35	0.09	0.14	0.50
7	31	Paper	0.37	0.10	0.19	0.59
8	31	Online	0.39	0.07	0.25	0.52
8	31	Paper	0.37	0.12	0.21	0.68
11	31	Online	0.36	0.08	0.20	0.48
11	31	Paper	0.36	0.10	0.15	0.57

Table 7-20. Item-Level Descriptive Statistics by Mode: FI Mathematics Adjusted Item-Total Correlation

Grade	N OP Items	Mode	Mean	SD	Min	Max
3	24	Online	0.35	0.07	0.18	0.49
3	24	Paper	0.36	0.10	0.16	0.51
4	24	Online	0.29	0.06	0.17	0.42
4	24	Paper	0.34	0.06	0.23	0.48
5	24	Online	0.34	0.08	0.11	0.48
5	24	Paper	0.35	0.11	0.15	0.56
6	24	Online	0.27	0.06	0.12	0.38
6	24	Paper	0.33	0.12	0.09	0.59
7	24	Online	0.28	0.08	0.14	0.48
7	24	Paper	0.34	0.10	0.11	0.52
8	24	Online	0.35	0.08	0.20	0.47
8	24	Paper	0.38	0.10	0.20	0.55
11	24	Online	0.32	0.06	0.19	0.41
11	24	Paper	0.34	0.07	0.19	0.48

Table 7-21. Item-Level Descriptive Statistics by Mode: FI Science Adjusted Item-Total Correlation

Grade	N OP Items	Mode	Mean	SD	Min	Max
4	35	Online	0.34	0.08	0.18	0.46
4	35	Paper	0.37	0.10	0.16	0.56
7	40	Online	0.30	0.08	0.16	0.46
7	40	Paper	0.32	0.10	-0.00	0.51
11	45	Online	0.34	0.08	0.15	0.49
11	45	Paper	0.35	0.10	0.11	0.62

Table 7-22. Item-Level Descriptive Statistics by Mode: FI Social Studies Adjusted Item-Total Correlation

Grade	N OP Items	Mode	Mean	SD	Min	Max
5	32	Online	0.32	0.07	0.17	0.43
5	32	Paper	0.30	0.10	0.08	0.48
8	33	Online	0.30	0.08	0.12	0.46
8	33	Paper	0.36	0.10	0.18	0.57
11	41	Online	0.32	0.08	0.12	0.46
11	41	Paper	0.37	0.08	0.10	0.52

Tables 7-23 through 7-34 present the item-level descriptive statistics (the mean item difficulty and item discrimination indicator, i.e., the mean adjusted *p*-value and the mean adjusted item total correlation, respectively) by content area and grade for all OP items on the Spring 2019 MI-Access SI and P.

Table 7-23. Item-Level Descriptive Statistics: SI ELA Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
3	15	0.68	0.09	0.55	0.87
4	15	0.66	0.11	0.46	0.85
5	15	0.65	0.09	0.47	0.83
6	15	0.69	0.11	0.48	0.85
7	15	0.67	0.11	0.47	0.85
8	15	0.70	0.11	0.49	0.87
11	15	0.67	0.09	0.51	0.82

Table 7-24. Item-Level Descriptive Statistics: SI Mathematics Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
3	15	0.65	0.10	0.50	0.81
4	15	0.61	0.10	0.44	0.75
5	15	0.60	0.13	0.42	0.78
6	15	0.56	0.11	0.40	0.79
7	15	0.59	0.15	0.37	0.84
8	15	0.59	0.15	0.36	0.85
11	15	0.69	0.11	0.49	0.86

Table 7-25. Item-Level Descriptive Statistics: SI Science Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
4	17	0.71	0.10	0.52	0.89
7	17	0.67	0.10	0.52	0.83
11	17	0.74	0.10	0.51	0.86

Table 7-26. Item-Level Descriptive Statistics: SI ELA Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
3	15	0.45	0.10	0.18	0.56
4	15	0.43	0.08	0.25	0.55
5	15	0.52	0.03	0.46	0.58
6	15	0.46	0.07	0.32	0.56
7	15	0.41	0.10	0.22	0.55
8	15	0.46	0.08	0.35	0.58
11	15	0.44	0.09	0.26	0.55

Table 7-27. Item-Level Descriptive Statistics: SI Mathematics Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
3	15	0.50	0.06	0.40	0.63
4	15	0.46	0.06	0.32	0.52
5	15	0.41	0.07	0.27	0.51
6	15	0.43	0.06	0.36	0.58
7	15	0.40	0.07	0.30	0.53
8	15	0.42	0.08	0.31	0.59
11	15	0.53	0.08	0.35	0.62

Table 7-28. Item-Level Descriptive Statistics: SI Science Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
4	17	0.48	0.08	0.24	0.58
7	17	0.43	0.08	0.30	0.53
11	17	0.48	0.10	0.24	0.61

Table 7-29. Item-Level Descriptive Statistics: P ELA Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
3	10	0.54	0.07	0.42	0.64
4	10	0.59	0.06	0.50	0.66
5	10	0.55	0.05	0.45	0.65
6	10	0.52	0.04	0.45	0.58
7	10	0.54	0.07	0.45	0.66
8	10	0.57	0.07	0.48	0.66
11	10	0.53	0.07	0.38	0.66

Table 7-30. Item-Level Descriptive Statistics: P Mathematics Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
3	10	0.52	0.07	0.39	0.60
4	10	0.51	0.07	0.44	0.64
5	10	0.51	0.05	0.42	0.59
6	10	0.52	0.07	0.42	0.60
7	10	0.53	0.05	0.48	0.61
8	10	0.58	0.06	0.51	0.68
11	10	0.51	0.09	0.29	0.60

Table 7-31. Item-Level Descriptive Statistics: P Science Adjusted *P*-Value

Grade	N	Mean	SD	Min	Max
4	15	0.60	0.06	0.48	0.67
7	15	0.56	0.07	0.42	0.66
11	15	0.56	0.07	0.46	0.68

Table 7-32. Item-Level Descriptive Statistics: P ELA Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
3	10	0.50	0.04	0.43	0.59
4	10	0.54	0.07	0.43	0.63
5	10	0.51	0.04	0.46	0.56
6	10	0.49	0.07	0.40	0.58
7	10	0.50	0.07	0.41	0.64
8	10	0.47	0.08	0.34	0.59
11	10	0.57	0.07	0.50	0.69

Table 7-33. Item-Level Descriptive Statistics: P Mathematics Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
3	10	0.53	0.05	0.42	0.59
4	10	0.51	0.06	0.41	0.60
5	10	0.47	0.05	0.37	0.53
6	10	0.51	0.05	0.45	0.57
7	10	0.53	0.06	0.42	0.62
8	10	0.50	0.06	0.34	0.54
11	10	0.54	0.06	0.44	0.62

Table 7-34. Item-Level Descriptive Statistics: P Science Adjusted Item-Total Correlation

Grade	N	Mean	SD	Min	Max
4	15	0.60	0.06	0.47	0.71
7	15	0.56	0.07	0.40	0.65
11	15	0.62	0.05	0.50	0.70

7.2.1 IRT Statistics for MI-Access FI ELA, Mathematics, Science, and Social Studies

The Rasch partial credit model (RPCM) (Wright and Masters, 1982) was used to calibrate MI-Access FI ELA, mathematics, science, and social studies items and to derive the scale scores. This model was used because of the flexibility of the RPCM to accommodate both MC and CR items. The RPCM extends the Rasch model (Rasch, 1960) for dichotomous (0, 1) items so that it accommodates the polytomously scored item data. Under the RPCM, for a given item i with *mi* score categories, the probability of person n scoring x (x = 0, 1, 2,... m_i) is given by

$$P_{ni}(X = x) = \frac{\exp\sum_{j=0}^{x}(\theta_n - D_{ij})}{\sum_{k=0}^{m_i}\exp\sum_{j=0}^{k}(\theta_n - D_{ij})}, \quad (7\text{-}1)$$

where θ_n represents a student's proficiency (ability) level and D_{ij} is the step difficulty of the j^{th} step on item i.

For dichotomous MC items, the RPCM reduces to the standard Rasch model and the single step difficulty is referred to as the item's difficulty.

The Rasch model predicts the probability of person n getting item i correct and is mathematically expressed as follows:

$$P_{ni}(X = 1) = \frac{\exp(\theta_n - D_{ij})}{1 + \exp(\theta_n - D_{ij})}. \quad (7\text{-}2)$$

7.2.2 Item Calibration for MI-Access FI ELA, Mathematics, Science, and Social Studies

A fixed parameter Rasch calibration approach was used to put all items onto the base scale. The IRT software used was WINSTEPS version 3.92.1 (Linacre, 2015). The following is an overview of the annual item calibration, equating, and scaling procedures for FI assessments:

- A WINSTEPS free calibration was conducted with all online OP item data for each content area and grade combination. For FI ELA, the data from online APUL and paper/pencil EI were used for calibration.

- After each free calibration, the obtained item parameters for the anchor (equating) items were compared with their banked values. During the mean-mean equating process, stability of the anchor (equating) items was checked. Outliers of the anchor items (i.e., a displacement with adjusted absolute value greater or equal to 0.5) were identified. The outliers were removed from the anchor (equating) item list. This process was done in iteration with some constraints; that is, the anchor item with the largest differential value was dropped first, then a second WINSTEPS free calibration was conducted to examine the outliers again and drop the second largest outlier item from the anchor list, and so on. The Office of Educational Assessment and Accountability (OEAA) MI-Access psychometrician then made the decision on the final anchor item set and shared the results with the National Center for Research on Evaluation, *Standards*, and Student Testing, University of California, Los Angeles (CRESST), which serves as an independent third party to validate the psychometric work done by the OEAA MI-Access psychometrician. See Appendix G: MI-Access Psychometric Verification Report for Spring 2019 for the detailed description of the psychometric verification work by CRESST.
- CRESST conducted an independent anchor item stability check and compared its conclusion with that of the OEAA MI-Access psychometrician.
- After the OEAA MI-Access psychometrician and CRESST agreed on the final anchor item set, the OEAA MI-Access psychometrician ran the mean-mean equating process to obtain the equating constants for each content by grade test, which was verified by CRESST.
- After the OEAA MI-Access psychometrician and CRESST agreed on the equating constants, the OEAA psychometrician applied the equating constants to the free calibration item parameters, thus transforming the free calibration scale to the item bank base scale. These equating constants were also applied to the WINSTEPS raw-to-theta-score tables, which were later used to generate the OP raw-to-scale-score tables.
- CRESST conducted a validation check on all the equated item parameters and the theta values and verified the OEAA MI-Access psychometrician's results.
- A fixed item parameter calibration method was then used to calibrate the paper/pencil forms to put all the items onto the item bank base scale.

7.2.3 Anchor Item Evaluation

A statistical method (adjusted absolute value against the mean difference) was used to identify the outlier or unstable anchor (equating) items, as mentioned in the above section. Specifically, the procedure was as follows:

- After each free calibration, the item parameter values were placed side-by-side with the item bank base parameter value.
- The mean of the current year free calibration set of the anchor (equating) items was calculated.
- The mean of the item bank base parameter values was calculated.
- The mean difference of the free calibration and the item bank base anchor item values was calculated.

- The adjusted value for each anchor item was calculated by adding the mean difference to each of the free run item parameter estimates.
- The absolute difference value (ADV) was calculated by subtracting the adjusted value from the bank base value.
- Items with ADV ≥ 0.5 were identified as outliers or unstable anchor items.
- The anchor item with the largest ADV was removed and a second WINSTEPS free calibration run was conducted to examine the rest of the anchor items.
- The process was repeated until all unstable anchor items were removed.

7.2.4 Evidence of Model Fit for FI ELA, Mathematics, Science, and Social Studies

An important assumption of IRT models, including the Rasch model, is scale unidimensionality. The OEAA has conducted exploratory factor analyses (multifactor vs. single factor) and model selection analyses. Although the model selection index tends to prefer more complex models, taking model parsimony into consideration and using the root mean square error of approximation (RMSEA) value criterion (close to 0), the OEAA found that the RMSEA results (ranging from 0.017 to 0.05 for FI ELA, from 0.019 to 0.06 for FI mathematics, from 0.015 to 0.045 for FI science, and from 0.020 to 0.05 for FI social studies) show evidence to support the use of (single factor) Rasch model item parameter calibration. For more details about the factor analysis, refer to Chapter 11 of this report.

In addition, the OEAA computed the IRT (WINSTEPS) item model fit/misfit and flagged the number of items and categories for FI tests. WINSTEPS provides two item fit statistics (infit and outfit) for evaluating the degree to which the Rasch model predicts the observed item responses. Each fit statistic can be expressed as a mean square (MnSq) statistic. Both infit and outfit MnSq (MSQIN and MSQOUT) are the average of standardized residual variance (i.e., the difference between the observed score and the Rasch estimated score divided by the square root of the Rasch model variance). The difference between the two values is that the MSQOUT gives all student responses equal weight. The MSQIN gives more weight to student response pattern, in which the student ability is closer to the item difficulty.

The average MSQIN and MSQOUT values are 1.0 and can range from 0.0 to infinity. Deviation in excess of the expected value can be interpreted as noise or lack of fit between the items and the model. Values lower than the expected value can be interpreted as item redundancy or overfitting items (too predictable, too much redundancy), and values greater than the expected value indicate underfitting items (too unpredictable, too much noise).

Rules of thumb regarding “practically significant” MnSq values vary. Items were flagged for model misfit by using MSQIN and MSQOUT. Values of MSQIN and MSQOUT are flagged using the following criterion:

- If MSQIN/MSQOUT > 2, then the MSQIN/MSQOUT flag indicates that the item has a high degree of misfit (MH).
- If the MSQIN/MSQOUT is between 1.5 and 2, then the MSQIN/MSQOUT flag indicates that the item has a moderate degree of misfit (MM).
- If MSQIN is below 0.5 and MSQOUT is below 1.5, then MSQINFL flag indicates that the item is too predicative (TP);

- If MSQOUT is below 0.5 and MSQIN is below 1.5, then MSQOUTFL flag indicates that the item is too predictive (TP).

Table 7-35 summarizes the item model fit and number of flagged items and categories for FI tests by content area, mode, and grade level for the Spring 2019 administration. As shown in the table, very few items were flagged as outliers (flagged for item misfit).

Table 7-35. FI IRT Item Model Fit and Flags by Content Area, Mode, and Grade Level

Content Area	Grade	Mode	N of OP Items	MSQIN	MSQINFL Type	MSQOUT	MSQOUTFL Type
ELA	3	Online	31	0		0	
ELA	3	Paper	31	0		0	
ELA	4	Online	31	0		0	
ELA	4	Paper	31	0		0	
ELA	5	Online	31	0		0	
ELA	5	Paper	31	0		1	MH
ELA	6	Online	31	0		0	
ELA	6	Paper	31	0		0	
ELA	7	Online	31	0		0	
ELA	7	Paper	31	0		0	
ELA	8	Online	31	0		0	
ELA	8	Paper	31	0		0	
ELA	11	Online	31	0		0	
ELA	11	Paper	31	0		0	
Mathematics	3	Online	24	0		0	
Mathematics	3	Paper	24	0		1	MM
Mathematics	4	Online	24	0		0	
Mathematics	4	Paper	24	0		0	
Mathematics	5	Online	24	0		0	
Mathematics	5	Paper	24	0		0	
Mathematics	6	Online	24	0		0	
Mathematics	6	Paper	24	0		0	
Mathematics	7	Online	24	0		0	
Mathematics	7	Paper	24	0		0	
Mathematics	8	Online	24	0		0	
Mathematics	8	Paper	24	0		0	
Mathematics	11	Online	24	0		0	
Mathematics	11	Paper	24	0		1	MM
Science	4	Online	35	0		0	

Content Area	Grade	Mode	N of OP Items	MSQIN	MSQINFL Type	MSQOUT	MSQOUTFL Type
Science	4	Paper	35	0		0	
Science	7	Online	40	0		0	
Science	7	Paper	40	1		2	MM (1), MH (1), TP (1)
Science	11	Online	45	0		0	
Science	11	Paper	45	0		1	MM
Social Studies	5	Online	32	0		0	
Social Studies	5	Paper	32	0		0	
Social Studies	8	Online	33	0		0	
Social Studies	8	Paper	33	0		0	
Social Studies	11	Online	41	0		0	
Social Studies	11	Paper	41	0		0	

7.2.5 Test Characteristic Curves and Conversion Tables

7.2.5.1 Test Characteristic Curves

The test characteristic curve (TCC) is the graphical representation of the test characteristic function (TCF), which is the expected raw total score given theta. For FI mathematics, science, and social studies, as all items are dichotomously scored, the expression of TCF is as follows (adapted from Yen & Fitzpatrick, 2006, p. 125):

$$E(X.|\theta) = \sum_{i=1}^{n} E(X_i|\theta) = \sum_{i=1}^{n} P_i(\theta) \quad (7\text{-}3)$$

For FI ELA, there is one EI CR item, so the TCF is the expected raw total score given theta, which contains the summation of expected raw scores for all APUL MC items and the step scores for the EI CR item.

The TCCs for MI-Access FI ELA, mathematics, science, and social studies by content area and grade level are provided in Appendix D. These graphs were created by using the WINSTEPS item parameter estimates for the online form OP items from the post-administration calibration in 2019. The OEAA also compared two mode TCCs using the separate mode data. The results show that for all the content areas by grade-level, the two modes' TCCs overlapped, indicating the two modes scale comparably.

7.2.5.2 FI Raw-to-Scale-Score Conversion Tables

The creation of the FI raw-to-scale-score conversion table involved the following steps:

- After completing equating and obtaining the equating constant for each content area by grade-level calibration, the equating constants to the theta values in the raw-to-theta tables from WINSTEPS free run were applied (after removing all unstable anchor items).
- The scaling constants (see Chapter 9.5), slope (A) and intercept (B) were applied to the theta values and conditional standard errors of the theta to get the scale scores and conditional error of measurement (CSEM) for each raw score:
  - Scale score=B+A*theta
  - CSEM=A*theta_SE

The scaling constants, slope and intercept, were obtained from MI-Access standard-setting outcomes. For FI ELA and mathematics, they were obtained from 2017 standard-setting outcomes. For FI science and social studies, they were obtained from 2015 standard-setting outcomes (refer to Performance-Level Standard Setting in Chapter 9 for details).

The tables in Appendix F present the raw-to-scale-score conversion tables by content area and grade level for the 2019 MI-Access FI assessments, which were used for OP reporting. No paper/pencil data were available for calibration when the conversion tables were created; therefore, a policy decision was made to apply the raw-to-scale-score conversion tables obtained from the online form to the corresponding content by grade level paper/pencil form for scale score generation. Since online and pencil/paper form test maps are designed using the same blueprint and the majority (74%–80%) of OP items on the two tests are the same, the assumption is that there is comparability between the two tests. As indicated by the overlaid TCCs in Appendix D, the evidence seems to support the mode scale comparability.

7.2.6 IRT Statistics

Tables 7-36 through 7-39 present the IRT item difficulty (b-parameter) descriptive statistics (mean item difficulty [BPar_Mean], minimum item difficulty [BPar_Min], maximum item difficulty [BPar_Max], and total number of OP items in the test [N]) by mode and grade level for FI tests. As shown in these tables, the average item difficulty is generally similar with some variations for some grades across the two modes within the same content area and grade level.

Table 7-36. Item Difficulty Statistics for FI ELA by Mode and Grade

Grade	Mode	BPar_Mean	BPar_Min	BPar_Max	N
3	Online	-0.153	-1.678	1.312	31
3	Paper	-0.248	-1.542	1.001	31
4	Online	0.196	-1.258	1.434	31
4	Paper	0.195	-0.914	1.344	31
5	Online	0.181	-1.242	1.833	31
5	Paper	-0.140	-1.667	1.171	31
6	Online	0.302	-1.155	1.766	31
6	Paper	-0.038	-1.470	1.715	31
7	Online	0.209	-1.083	1.538	31
7	Paper	0.125	-1.379	1.475	31
8	Online	0.185	-1.525	1.376	31
8	Paper	-0.081	-1.240	1.133	31
11	Online	-0.068	-1.676	1.100	31
11	Paper	-0.145	-1.564	1.117	31

Table 7-37. Item Difficulty Statistics for FI Mathematics by Mode and Grade

Grade	Mode	BPar_Mean	BPar_Min	BPar_Max	N
3	Online	-0.143	-1.777	1.394	24
3	Paper	-0.073	-1.502	1.610	24
4	Online	0.096	-1.178	1.286	24
4	Paper	-0.261	-1.966	0.474	24
5	Online	0.096	-0.944	1.145	24
5	Paper	0.031	-0.955	1.343	24
6	Online	0.025	-1.799	1.134	24
6	Paper	-0.014	-1.719	0.953	24
7	Online	-0.289	-1.761	1.050	24
7	Paper	-0.246	-1.524	0.686	24
8	Online	-0.020	-1.323	0.980	24
8	Paper	-0.063	-1.323	0.980	24
11	Online	-0.013	-1.685	1.146	24
11	Paper	-0.087	-1.324	0.692	24

Table 7-38. Item Difficulty Statistics for FI Science by Mode and Grade

Grade	Mode	BPar_Mean	BPar_Min	BPar_Max	N
4	Online	-0.061	-2.519	0.991	35
4	Paper	0.108	-1.678	0.888	35
7	Online	-0.117	-1.902	1.240	40
7	Paper	-0.239	-1.902	1.445	40
11	Online	-0.169	-1.811	1.010	45
11	Paper	-0.243	-1.811	0.719	45

Table 7-39. Item Difficulty Statistics for FI Social Studies by Mode and Grade

Grade	Mode	BPar_Mean	BPar_Min	BPar_Max	N
5	Online	0.067	-0.861	0.960	32
5	Paper	0.113	-0.861	0.960	32
8	Online	-0.099	-0.995	0.939	33
8	Paper	-0.144	-0.995	0.690	33
11	Online	0.100	-1.009	0.855	41
11	Paper	0.088	-1.009	0.855	41

7.3 Summary

In summary, the overall purpose of the OP data analysis is to ensure that the test items, as well as the overall test, are functioning appropriately. The analyses also help maintain the test scale across years so that test results may be appropriately compared across years. The data analyses undertaken by MDE (with contractor support from Measurement Incorporated) are in alignment with multiple best practices of the assessment industry; in particular, they are related to the following standards from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- 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.

Chapter 8: Test Results

This chapter of the technical report contains information on the results of the Spring 2019 administration of the MI-Access and provides descriptions of the score reports, data structure, and interpretive guide. The AERA, APA, and NCME (2014) *Standards* addressed in Chapter 8 include 5.1, 6.10, and 7.0. Each standard will be presented in the pertinent section of this chapter.

8.1 Student Participation

The Spring 2019 MI-Access was administered to Michigan students in four content areas: English language arts (ELA), mathematics, science, and social studies. The social studies test was administered only as a Functional Independence (FI) assessment. For the purposes of this technical report, "percent valid" is the percentage of students who received a valid score given the total number of students registered to take the online or paper/pencil test. Student participation information is reported for all students and for the following demographic subgroups:

- Gender: Female and Male
- Race/Ethnicity: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Two or More Races, and White
- Economically Disadvantaged: Yes, No
- English Language Learners: Yes, No
- Students Used Standard Accommodations: Yes, No

"Participation rate" measures something different for alternate assessment than for general assessment. The decision to take an alternate assessment comes from a student's local Individualized Education Program (IEP) team, with guidance, but not control, from the Michigan Department of Education (MDE). There is no state tracking of IEP content. This decision is locally controlled, can change at an IEP team's discretion, and can differ among content areas. Thus, if a student who should take MI-Access is never registered for *any* assessment, MDE knows that student did not take an assessment but not which assessment the student should have taken. That lack of testing would appear in the general assessment count of students with disabilities who did not participate in assessment.

In this chapter, Tables 8-1a through 8-4f show valid tests as a percentage of MI-Access FI tests taken. Tables 7-5 through 7-14 in Chapter 7 show the numbers of MI-Access Supported Independence (SI) and Participation (P) tests taken; nonparticipation in an SI or P test is factored into the scoring rubric and would still result in a valid, scored test, potentially one with a "0" score.

8.2 Current Administration Data Scale Score Summaries

MI-Access SI and P scores represent the number of points earned out of the number of points possible but do not yield a scale score.

8.3 Description of Reports

Score reports are the primary means of communicating test scores to relevant district personnel (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 provided interpretations appropriate to the audience. The interpretations described in simple language what the test covered, what scores represent, the precision/reliability of the scores, and how scores are intended to be used. (p. 119)

Standard 5.1 is also addressed:

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

This section outlines the array of reports that were produced and provided for the 2019 MI-Access administration. Scale score, raw score, and points earned information can be found in section 8.3.1, and information pertaining to each type of report can be found in section 8.3.2.

8.3.1 Scale Scores

FI scale scores are statistical conversions of raw score points and are the results of a linear transformation of the underlying ability distributions. Since scale scores are produced after equating and scaling, they permit comparison of assessment results across different test administrations within a particular grade and content area.

Each year, new test forms are developed. These new forms never contain exactly the same questions as the previous forms. To have a fair comparison across years for different cohorts, it is necessary to have a scale score that shares the same meaning across different administrations.

Scale scores are not comparable across grade levels and across subject areas. Details of the development of MI-Access scale scores are described in Chapter 9, section 9.5.

Scale scores are stable because they allow for students' scores to be reported on the same scale regardless of which year the students took the assessment and which form of the assessment the students took. Schools can use scale scores to compare the performances of groups of students across years. These comparisons can then be used to assess the impact of changes or differences in instruction or curriculum. The scale scores can be used to determine whether students are demonstrating the same skill and ability across cohorts within a grade level and content area.

8.3.1.1 Raw Score

In addition to scale scores, sub-content raw scores are reported in the score reports. These scores are the sum of raw points earned in each content category. Total raw scores are also reported. Several values that are derived from the raw scores are added to assist in interpreting them: maximum possible score points, percentage correct, and aggregate averages (for school- and district-level reports).

8.3.1.2 Points Earned

The MI-Access SI and P reports do not use a scale score to display results; instead they use raw scores. These raw scores are displayed as earned points versus points possible. The total earned points are displayed.

8.3.2 Score Reports

MI-Access score reports comprise student-level data reports and aggregate data reports. Brief descriptions of these reports are provided below. More extensive descriptions with samples are included in the *Spring 2019 Interpretive Guide to MI-Access Reports*.

8.3.2.1 Student-Level Data Reports and Data Files

- The *Student Record Labels* provide a summary of student performance levels for individual students. The labels include district and school information, student demographic information, M-Access FI administration cycle information, and overall student performance level for tested content areas. Student Record Labels are provided for inclusion in a student's CA60 (or Cumulative Student Record) folder. In late summer, the labels are printed and shipped to the school in which the student tested. These labels are available on the Secure Site if the school needs to print additional copies.

- The *Individual Student Report* (ISR) provides information about student performance by content area. Each student will have a separate ISR for each content assessed. The report is divided into three main sections:
    - Student demographic information
    - Overall content performance and detailed claim data for ELA and mathematics
    - Strand/discipline and content expectation data for science and social studies

- *Parent Reports* are printed and shipped to schools for distribution to parents. The *parent report* provides information about student performance in tested content areas. This report includes four main sections:
    - Superintendent letter
    - Overall performance level and scale score
    - Detailed claim data for ELA and mathematics and strand/discipline data for science and social studies
    - Definitions for parents and performance-level descriptors

- The *Student Roster* allows users to view student scale scores and claim performance data for ELA and mathematics or discipline data for science and social studies by content area and grade. The report is divided into four main sections:
  - An alphabetical listing of the selected students
  - Overall content performance in table format
  - Overall content performance in graphical format
  - Claim data for ELA and mathematics and strand/discipline data for science and social studies

- The *Student Overview* provides summary information about student performance in all tested content areas in the selected grade. For each selected student, the following data are displayed for each tested content area in both graphical and table format: scale score, margin of error, points earned, performance level, and claim/strand/discipline performance.

8.3.2.2 Aggregate Data Reports and Data Files

- The *Expectation/Scoring Focus Analysis Report* provides the percentage of points earned by grade, the content area expectations in each discipline (for science and social studies), and the number of students scoring in each of four quartiles. The report is intended to provide an overview of performance by content expectation.

  The report displays the number of students assessed in each expectation/scoring focus because not all students were assessed on every expectation. The report also displays the average percentage of points earned and the number of students scoring in one of four bands of quartiles: 0%–25%, 26%–50%, 51%–75%, and 76%–100% points earned out of all possible points.

- The *Demographic Report* provides a comparison of students by grade and content area, aggregated across selected demographic groups and showing the percentage of students proficient at each level (Emerging Toward the Performance Standard, Attained the Performance Standard, and Surpassed the Performance Standard). The demographic report is available at the school, district, intermediate school district (ISD), and state levels.

  After a user selects a grade to view online, all tested content areas for that grade are displayed in alphabetical order. The report is divided into three main sections:
  - Overall performance-level percentages for the selected students in the grade and content area, displayed in graphical format
  - Demographic subgroup performance-level data, displayed in table format
  - Performance-level percentages for a selected demographic subgroup, displayed in graphical format

- The *Comprehensive Report* provides a comparison of students by grade and content area, aggregated across schools and districts and showing the percentage of students performing at each level (Emerging Toward the Performance Standard, Attained the Performance Standard, and Surpassed the Performance Standard). The *Comprehensive Report* is available at the ISD and district levels. After a user selects a grade to view online, all tested content areas for that grade are displayed in alphabetical order. The report is divided into three main sections:
  - Overall performance-level percentages for the selected students in the grade and content area, displayed in graphical format
  - Entity performance-level data for each school (compiled in a *District Report*) or district (compiled in an ISD report), displayed in table format
  - Performance-level percentages, displayed in graphical format

- The *Student Data File* contains detailed individual student data in an Excel file. This data includes school information, student demographic data, test administration data, and student performance data. The *Student Data File* is provided for schools to use as a data resource for school- or district-level data reviews. Schools or districts can use the *Student Data File* to manipulate and evaluate data in ways that support school improvement goals or other data-based decision-making purposes.

- The *Comma-separated File* (CSV) contains student performance data used in the selected report. This data includes school information, student population, demographic group, and student performance data. The CSV is provided for schools to use as a data resource for school- or district-level data reviews. Schools or districts can use the CSV to evaluate data in ways that support school improvement goals or other data-based decision-making purposes.

8.4 Interpretive Guide to MI-Access Reports

For the Spring 2019 MI-Access, MDE produced individual and aggregate reports for students, schools, districts, and the state. The information provided in these reports can be interpreted and used in a variety of ways. In addition to providing interpretation, it is important that the information can be understood by the target audience. Standard 7.0 of the AERA, APA, and 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. (p. 125)

To aid in interpretation, MDE prepared the *Spring 2019 MI-Access Interpretive Guide to Reports* for Michigan parents, teachers, and administrators. The *Spring 2019 MI-Access Interpretive Guide to Reports* can be found in Appendix B of this technical report.

8.5 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 MDE and its contractors are in alignment with multiple best practices of the testing industry; in particular, they are related to the following standards in the *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.

Table 8-1a. MI-Access FI Test Completion Rates by Grade: English Language Arts—All Students

	Grade	3	4	5	6	7	8	11
All Students	Total Valid	1,041	1,198	1,249	1,453	1,419	1,416	1,303
All Students	Total Tested	1,110	1,277	1,343	1,551	1,523	1,518	1,440
All Students	Percent Valid	93.78	93.81	93.00	93.68	93.17	93.28	90.49

Table 8-1b. MI-Access FI Test Completion Rates by Grade: English Language Arts—Gender

	Grade	3	4	5	6	7	8	11
Female	Total Valid	366	413	404	480	474	476	447
Female	Total Tested	381	434	439	523	504	511	492
Female	Percent Valid	96.06	95.16	92.03	91.78	94.05	93.15	90.85
Male	Total Valid	675	785	845	973	945	940	856
Male	Total Tested	729	843	904	1,028	1,019	1,007	948
Male	Percent Valid	92.59	93.12	93.47	94.65	92.74	93.35	90.30

Table 8-1c. MI-Access FI Test Completion Rates by Grade: English Language Arts—Race/Ethnicity

	Grade	3	4	5	6	7	8	11
American Indian/Alaska Native	Total Valid	12	NULL	NULL	13	NULL	14	NULL
American Indian/Alaska Native	Total Tested	13	NULL	NULL	13	NULL	16	NULL
American Indian/Alaska Native	Percent Valid	92.31	NULL	NULL	100	NULL	87.50	NULL
Asian	Total Valid	18	18	17	10	12	18	20
Asian	Total Tested	19	18	18	10	13	18	22
Asian	Percent Valid	94.74	100	94.44	100	92.31	100	90.91
Black/African American	Total Valid	276	376	343	418	427	396	340
Black/African American	Total Tested	301	403	368	456	463	442	385
Black/African American	Percent Valid	91.69	93.30	93.21	91.67	92.22	89.59	88.31
Hispanic/Latino	Total Valid	92	97	108	125	104	121	97
Hispanic/Latino	Total Tested	97	104	117	133	108	125	104
Hispanic/Latino	Percent Valid	94.85	93.27	92.31	93.98	96.30	96.80	93.27
Two or More Races	Total Valid	52	68	54	69	70	57	47
Two or More Races	Total Tested	55	69	56	74	73	61	51
Two or More Races	Percent Valid	94.55	98.55	96.43	93.24	95.89	93.44	92.16
White	Total Tested	591	632	721	815	796	807	790
White	Percent Valid	625	675	776	862	854	853	868
White	Total Valid	94.56	93.63	92.91	94.55	93.21	94.61	91.01

Table 8-1d. MI-Access FI Test Completion Rates by Grade: English Language Arts—Economically Disadvantaged

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	775	941	929	1,119	1,081	994	877
Yes	Total Tested	827	1,003	1,001	1,198	1,156	1,074	977
Yes	Percent Valid	93.71	93.82	92.81	93.41	93.51	92.55	89.76
No	Total Valid	266	257	320	334	338	422	426
No	Total Tested	283	274	342	353	367	444	463
No	Percent Valid	93.99	93.80	93.57	94.62	92.10	95.05	92.01

Table 8-1e. MI-Access FI Test Completion Rates by Grade: English Language Arts—English Language Learners

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	77	81	92	99	96	102	81
Yes	Total Tested	78	86	100	104	97	104	84
Yes	Percent Valid	98.72	94.19	92.00	95.19	98.97	98.08	96.43
No	Total Valid	964	1,117	1,157	1,354	1,323	1,314	1,222
No	Total Tested	1,032	1,191	1,243	1,447	1,426	1,414	1,356
No	Percent Valid	93.41	93.79	93.08	93.57	92.78	92.93	90.12

Table 8-1f. MI-Access FI Test Completion Rates by Grade: English Language Arts—Students Used Standard Accommodations

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	25	37	48	25	35	27	17
Yes	Total Tested	28	38	49	25	38	29	18
Yes	Percent Valid	89.29	97.37	97.96	100	92.11	93.10	94.44
No	Total Valid	1,016	1,161	1,201	1,428	1,384	1,389	1,286
No	Total Tested	1,082	1,239	1,294	1,526	1,485	1,489	1,422
No	Percent Valid	93.90	93.70	92.81	93.58	93.20	93.28	90.44

Table 8-2a. MI-Access FI Test Completion Rates by Grade: Mathematics—All Students

	Grade	3	4	5	6	7	8	11
All Students	Total Valid	1,043	1,211	1,293	1,522	1,506	1,483	1,375
All Students	Total Tested	1,051	1,223	1,302	1,536	1,518	1,497	1,402
All Students	Percent Valid	99.24	99.02	99.31	99.09	99.21	99.06	98.07

Table 8-2b. MI-Access FI Test Completion Rates by Grade: Mathematics—Gender

	Grade	3	4	5	6	7	8	11
Female	Total Valid	371	417	420	522	509	494	474
Female	Total Tested	375	419	427	525	512	502	482
Female	Percent Valid	98.93	99.52	98.36	99.43	99.41	98.41	98.34
Male	Total Valid	672	794	873	1,000	997	989	901
Male	Total Tested	676	804	875	1,011	1,006	995	920
Male	Percent Valid	99.41	98.76	99.77	98.91	99.11	99.40	97.93

Table 8-2c. MI-Access FI Test Completion Rates by Grade: Mathematics—Race/Ethnicity

	Grade	3	4	5	6	7	8	11
American Indian/Alaska Native	Total Valid	12	NULL	NULL	11	11	15	NULL
American Indian/Alaska Native	Total Tested	13	NULL	NULL	11	11	16	NULL
American Indian/Alaska Native	Percent Valid	92.31	NULL	NULL	100	100	93.75	NULL
Asian	Total Valid	16	17	17	11	11	18	21
Asian	Total Tested	16	17	17	11	12	18	22
Asian	Percent Valid	100	100	100	100	91.67	100	95.45
Black/African American	Total Valid	290	394	357	444	452	426	361
Black/African American	Total Tested	295	397	360	451	458	431	370
Black/African American	Percent Valid	98.31	99.24	99.17	98.45	98.69	98.84	97.57
Hispanic/Latino	Total Valid	94	100	108	132	107	124	102
Hispanic/Latino	Total Tested	94	102	110	133	107	125	103
Hispanic/Latino	Percent Valid	100	98.04	98.18	99.25	100	99.20	99.03
Two or More Races	Total Valid	52	64	59	72	73	59	49
Two or More Races	Total Tested	52	64	59	72	73	59	50
Two or More Races	Percent Valid	100	100	100	100	100	100	98.00
White	Total Valid	579	630	745	849	851	838	833
White	Total Tested	581	636	749	855	856	845	848
White	Percent Valid	99.66	99.06	99.47	99.30	99.42	99.17	98.23

Table 8-2d. MI-Access FI Test Completion Rates by Grade: Mathematics—Economically Disadvantaged

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	781	958	966	1,164	1,147	1,047	930
Yes	Total Tested	786	967	972	1,176	1,156	1,060	949
Yes	Percent Valid	99.36	99.07	99.38	98.98	99.22	98.77	98.00
No	Total Valid	262	253	327	358	359	436	445
No	Total Tested	265	256	330	360	362	437	453
No	Percent Valid	98.87	98.83	99.09	99.44	99.17	99.77	98.23

Table 8-2e. MI-Access FI Test Completion Rates by Grade: Mathematics—English Language Learners

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	77	80	95	104	90	104	84
Yes	Total Tested	78	82	96	105	91	105	84
Yes	Percent Valid	98.72	97.56	98.96	99.05	98.90	99.05	100
No	Total Valid	966	1,131	1,198	1,418	1,416	1,379	1,291
No	Total Tested	973	1,141	1,206	1,431	1,427	1,392	1,318
No	Percent Valid	99.28	99.12	99.34	99.09	99.23	99.07	97.95

Table 8-2f. MI-Access FI Test Completion Rates by Grade: Mathematics—Students Used Standard Accommodations

	Grade	3	4	5	6	7	8	11
Yes	Total Valid	NULL	11	NULL	NULL	NULL	NULL	NULL
Yes	Total Tested	NULL	11	NULL	NULL	NULL	NULL	NULL
Yes	Percent Valid	NULL	100	NULL	NULL	NULL	NULL	NULL
No	Total Valid	1,037	1,200	1,287	1,519	1,498	1,477	1,370
No	Total Tested	1,045	1,212	1,296	1,533	1,510	1,491	1,397
No	Percent Valid	99.23	99.01	99.31	99.09	99.21	99.06	98.07

Table 8-3a. MI-Access FI Test Completion Rates by Grade: Science—All Students

	Grade	4	7	11
All Students	Total Valid	1,051	1,422	1,385
All Students	Total Tested	1,060	1,428	1,411
All Students	Percent Valid	99.15	99.58	98.16

Table 8-3b. MI-Access FI Test Completion Rates by Grade: Science—Gender

	Grade	4	7	11
Female	Total Valid	354	468	481
Female	Total Tested	356	468	488
Female	Percent Valid	99.44	100	98.57
Male	Total Valid	697	954	904
Male	Total Tested	704	960	923
Male	Percent Valid	99.01	99.38	97.94

Table 8-3c. MI-Access FI Test Completion Rates by Grade: Science —Race/Ethnicity

	Grade	4	7	11
American Indian/Alaska Native	Total Valid	NULL	10	NULL
American Indian/Alaska Native	Total Tested	NULL	10	NULL
American Indian/Alaska Native	Percent Valid	NULL	100	NULL
Asian	Total Valid	15	12	22
Asian	Total Tested	15	12	22
Asian	Percent Valid	100	100	100
Black/African American	Total Valid	350	428	368
Black/African American	Total Tested	352	431	377
Black/African American	Percent Valid	99..43	99.30	97.61
Hispanic/Latino	Total Valid	92	102	103
Hispanic/Latino	Total Tested	93	102	104
Hispanic/Latino	Percent Valid	98.92	100	99.04
Two or More Races	Total Valid	59	71	49
Two or More Races	Total Tested	59	71	50
Two or More Races	Percent Valid	100	100	98.00
White	Total Valid	529	798	834
White	Total Tested	534	801	849
White	Percent Valid	99.06	99.63	98.23

Table 8-3d. MI-Access FI Test Completion Rates by Grade: Science—Economically Disadvantaged

	Grade	4	7	11
Yes	Total Valid	830	1082	937
Yes	Total Tested	836	1086	956
Yes	Percent Valid	99.28	99.63	98.01
No	Total Valid	221	340	448
No	Total Tested	224	342	455
No	Percent Valid	98.66	99.42	98.46

Table 8-3e. MI-Access FI Test Completion Rates by Grade: Science—English Language Learners

	Grade	4	7	11
Yes	Total Valid	71	90	84
Yes	Total Tested	72	90	84
Yes	Percent Valid	98.61	100	100
No	Total Valid	980	1,332	1,301
No	Total Tested	988	1,338	1,327
No	Percent Valid	99.19	99.55	98.04

Table 8-3f. MI-Access FI Test Completion Rates by Grade: Science —Students Used Standard Accommodations

	Grade	4	7	11
Yes	Total Valid	13	NULL	NULL
Yes	Total Tested	13	NULL	NULL
Yes	Percent Valid	100	NULL	NULL
No	Total Valid	1,038	1,418	1,382
No	Total Tested	1,047	1,424	1,408
No	Percent Valid	99.14	99.58	98.15

Table 8-4a. MI-Access FI Test Completion Rates by Grade: Social Studies—All Students

	Grade	5	8	11
All Students	Total Valid	1,234	1,474	1,389
All Students	Total Tested	1,242	1,478	1,414
All Students	Percent Valid	99.36	99.73	98.23

Table 8-4b. MI-Access FI Test Completion Rates by Grade: Social Studies—Gender

	Grade	5	8	11	7	8
Female	Total Valid	400	495	481	561	583
Female	Total Tested	406	498	487	565	584
Female	Percent Valid	98.52	99.40	98.77	99.29	99.83
Male	Total Valid	834	979	908	99.40	97.93
Male	Total Tested	836	980	927	1115	1086
Male	Percent Valid	99.76	99.90	97.95	99.55	99.91

Table 8-4c. MI-Access FI Test Completion Rates by Grade: Social Studies—Race/Ethnicity

	Grade	5	8	11
American Indian/Alaska Native	Total Valid	NULL	14	NULL
American Indian/Alaska Native	Total Tested	NULL	14	NULL
American Indian/Alaska Native	Percent Valid	NULL	100	NULL
Asian	Total Valid	15	19	22
Asian	Total Tested	15	19	22
Asian	Percent Valid	100	100	100
Black/African American	Total Valid	346	433	368
Black/African American	Total Tested	349	434	376
Black/African American	Percent Valid	99.14	99.77	97.87
Hispanic/Latino	Total Valid	105	119	103
Hispanic/Latino	Total Tested	107	119	104
Hispanic/Latino	Percent Valid	98.13	100	99.04
Two or More Races	Total Valid	54	57	49
Two or More Races	Total Tested	54	57	50
Two or More Races	Percent Valid	100	100	98.00
White	Total Valid	707	830	838
White	Total Tested	710	833	853
White	Percent Valid	99.58	99.64	98.24

Table 8-4d. MI-Access FI Test Completion Rates by Grade: Social Studies—Economically Disadvantaged

	Grade	5	8	11	6	7	8	11
Yes	Total Valid	926	1,043	941	1197	1143	1166	951
Yes	Total Tested	931	1,046	960	1313	1266	994	1,081
Yes	Percent Valid	99.46	99.71	98.02	91.17	90.28	94.49	89.76
No	Total Valid	308	431	448	347	334	338	454
No	Total Tested	311	432	454	368	422	92.01	367
No	Percent Valid	99.04	99.77	98.68	94.62	93.81	94.66	90.8

Table 8-4e. MI-Access FI Test Completion Rates by Grade: Social Studies—English Language Learners

	Grade	5	8	11
Yes	Total Valid	90	102	84
Yes	Total Tested	91	102	84
Yes	Percent Valid	98.90	1	1
No	Total Valid	1,144	1,372	1,305
No	Total Tested	1,151	1,376	1,330
No	Percent Valid	99.39	99.71	98.12

Table 8-4f. MI-Access FI Test Completion Rates by Grade: Social Studies—Students Used Standard Accommodations

	Grade	5	8	11
Yes	Total Valid	NULL	NULL	NULL
Yes	Total Tested	NULL	NULL	NULL
Yes	Percent Valid	NULL	NULL	NULL
No	Total Valid	1,229	1,468	1,385
No	Total Tested	1,237	1,472	1,410
No	Percent Valid	99.35	99.73	98.23

Chapter 9: Performance-Level Setting

This chapter briefly describes the MI-Access performance-level setting and presents the cut scores established and the performance-level descriptors created for the performance levels.

9.1 Performance-Level Setting for ELA, Mathematics, Science, and FI Social Studies

The Michigan Department of Education (MDE) in collaboration with Measurement Incorporated conducted performance-level standard settings on MI-Access English language arts (ELA), mathematics, science, and Functional Independence (FI) social studies.

The standard-setting meetings for ELA and Mathematics occurred in June and July of 2017, when MDE realigned the grade-based alternate content expectations to the Essential Elements based on the Dynamic Learning Maps (University of Kansas Research Center, 2013a, 2013b). The standard-setting meetings for MI-Access Science and FI Social Studies occurred in June and July of 2015, when MDE made changes to the Science and FI Social Studies tests, including changes in test length and form numbers.

The test content expectations for Science were based on the Michigan Extended Benchmarks of Science, and the test blueprints for FI Social Studies were based on the Michigan Extended Grade Level Expectations and the Extended High School Content Expectations. As the most recent example, the 2017 meetings are outlined in sections 9.2 through 9.4, with further details and discussion in Appendix E.

9.2 Selection and Constitution of the Standard-Setting Panels

MDE recruited panelists for the standard-setting event. All active members of the database of educators who participate as item writers or committee members (bias and sensitivity committees, content area committees, or range-finding committees) were invited to apply. In addition, school principals and special education supervisors were encouraged to nominate teachers. Finally, a call went out through the MDE “Spotlight on Student Assessment” newsletter for educators to apply.

MDE received more applicants than there were spaces on the educator panels. Candidates were matched to panels based on the level of assessment their students currently took. Then, the panelists were prioritized—first by location in the state and then by years of experience—to get a diverse representation of experience and to ensure a broad coverage of panelists from across the state.

While some panels had multiple panelists from within a single ISD, no panel had more than one panelist from the same local educational agency.

Table 9-1 summarizes the locations from which panelists for the Supported Independence (SI) and Participation (P) groups came. There were 26 Intermediate School Districts and 42 ISD or local districts represented, as well as one public school academy management provider and one higher education representative on these committees. Table 9-2 summarizes the locations

from which panelists for the Functional Independence groups came. There were 32 Intermediate School Districts and 64 ISD or local districts represented.

The following terms are abbreviated in Tables 9-1 and 9-2: Intermediate School District (ISD), Educational Service Agency (ESA), Educational Service District (ESD), and Regional Education School District (RESD).

Table 9-1. Summary of Locations of Panelists for Supported Independence and Participation Standard Setting, 2017

ISD/ESA/ESD/RESA	Local District or PSA
Allegan Area ESA	Allegan Area ESA
Alpena-Montmorency-Alcona ESD	Alpena-Montmorency-Alcona ESD
Charlevoix-Emmet ISD	Public Schools of Petoskey
Dickson-Iron ISD	Dickson-Iron ISD
Eastern Upper Peninsula ISD	Eastern Upper Peninsula ISD
Genesee ISD	Genesee ISD
Gratiot-Isabella RESD	Gratiot-Isabella RESD
Huron ISD	Huron ISD
Ingham ISD	Haslett Public Schools Lansing Public Schools Mason Public Schools
Ionia ISD	Belding Area Schools
Jackson ISD	Jackson ISD Jackson Public Schools Northwest Community Schools Vandercook Lake Public Schools
Kalamazoo RESA	Kalamazoo RESA
Kent ISD	Lowell Area Schools
Lenawee ISD	Lenawee ISD
Lapeer ISD	Lapeer Community Schools
Lewis Cass ISD	Lewis Cass ISD
Livingston ESA	Livingston Educational Service Agency
Montcalm Area ISD	Montcalm Area ISD
Mecosta-Osceola ISD	Mecosta-Osceola ISD
Oakland Schools	Lake Orion Community Schools Oak Park Schools Troy School District Walled Lake Consolidated Schools

ISD/ESA/ESD/RESA	Local District or PSA
Other	CS Partners (Education Service Provider) Alma College
Ottawa Area ISD	Coopersville Area Public Schools Grand Haven Area Public Schools
Saginaw ISD	Carrolton Public Schools Chesaning Union Schools Saginaw ISD Saginaw Public Schools
St. Joseph County ISD	St. Joseph County ISD
Washtenaw ISD	Ann Arbor Public Schools Washtenaw ISD
Wayne RESA	Dearborn Public Schools Detroit Public Schools Community District Garden City School District Grosse Pointe Public Schools Wayne-Westland Community Schools Wyandotte Public Schools

Table 9-2. Summary of Locations of Panelists for Functional Independence Standard Setting, 2017

ISD/ESA/ESD/RESA	Local District or PSA
Allegan Area ESA	Otsego Public Schools
Bay-Arenac ISD	Bay-Arenac ISD Bay City Public Schools
Berrien RESA	Lakeshore Public Schools
Calhoun ISD	Lakeview School District
Eastern Upper Peninsula ISD	DeTour Area Schools Eastern Upper Peninsula ISD
Eaton RESA	Potterville Public Schools
Genesee ISD	Flushing Community Schools Greater Heights Academy Linden Community Schools
Ingham ISD	Lansing Charter Academy Waverly Community Schools
Jackson ISD	Jackson Public Schools Vandercook Lake Public Schools Western School District

ISD/ESA/ESD/RESA	Local District or PSA
Kalamazoo RESA	Comstock Public Schools
Kent ISD	East Grand Rapids Public Schools
	Grand Rapids Public Schools
	Kentwood Public Schools
Lenawee ISD	Adrian Public Schools
	Onsted Community Schools
Lapeer ISD	Almont Community Schools
Livingston ESA	Brighton Area Schools
	Pinckney Community Schools
Macomb ISD	Chippewa Valley Schools
	Fraser Public Schools
	Lakeview Public Schools
	Utica Community Schools
	VanDyke Public Schools
Manistee ISD	Michigan Great Lakes Virtual Academy
Montcalm Area ISD	Greenville Public Schools
	Tri County Area Schools
Midland County ESA	Midland Public Schools
Muskegon ISD	Montague Area Schools
	Orchard View Schools
Newaygo County RESA	Freemont Public Schools
	Newaygo Public Schools
Oakland Schools	Berkley School District
	Troy Public Schools
	West Bloomfield School District
Ottawa Area ISD	Jenison Public Schools
Sanilac ISD	Sanilac ISD
Saginaw ISD	Freeland Community Schools
	Saginaw Public Schools
Shiawassee RESD	Corunna Public Schools
	Morrice Area Schools
St. Clair RESA	Capac Community Schools
	Landmark Academy
	Memphis Community Schools
	Yale Public Schools
St. Joseph County ISD	Colon Community Schools

ISD/ESA/ESD/RESA	Local District or PSA
Van Buren ISD	Bloomingdale Public School District
	Gobles Public Schools
	South Haven Public Schools
Washtenaw ISD	Ypsilanti Community Schools
Wayne RESA	Detroit Public Schools Community School District
	Lincoln Park Public Schools
	Livonia Public Schools
	South Redford School District
	University Preparatory Academy
West Shore ESD	Baldwin Community Schools
Wexford-Missaukee ISD	Cadillac Area Public Schools
	Marion Public Schools

9.3 Performance-Level Descriptors (PLDs)

In the spring of 2016, the MI-Access assessment programs for English language arts (ELA) and for mathematics were realigned to measure the current alternate content expectations in these areas. MI-Access measures the Essential Elements with Michigan Range of Complexity for ELA and Mathematics. This change required that a new standard setting take place for these content areas.

Standard setting is the methodology used to define levels of achievement or proficiency and the cut scores corresponding to those levels. For MI-Access, this process helped determine the cut scores that separate the reported performance levels of "Emerging Toward the Performance Standard," "Attained the Performance Standard," and "Surpassed the Performance Standard."

In the summer of 2017, a standard-setting process was completed for MI-Access ELA and mathematics. This process included over 140 educators from across the state of Michigan as described in section 9.2. The process involved the use of PLDs. Organized by reported performance levels (Emerging Toward the Performance Standard, Attained the Performance Standard, and Surpassed the Performance Standard), the PLDs describe what a student at each level should be able to do relative to the content expectations being measured. The PLDs used for the MI-Access standard-setting process in 2017 can be found on the MDE website.

9.4 Standard-Setting Methods and Procedures

The bookmark method (Lewis, Mitzel, & Green, 1996; Cizek & Bunch, 2007; Lewis, Mitzel, Mercado, & Schulz, 2012) was utilized for setting MI-Access Functional Independence (FI) ELA, Mathematics, Science, and Social Studies performance standards. MDE created the ordered item booklets (OIBs), which included RP 67 statistics and other necessary documents that accompanied the OIBs. Measurement Incorporated, an administration contractor, selected facilitators, conducted the training workshops, and facilitated the panel session meetings and the vertical articulation meetings. Three rounds of bookmark panel sessions were conducted.

A vertical articulation session concluded the meetings, in which selected grade-level panel members from each content area reviewed and revised the panel-recommended cut scores when they deemed it necessary.

For MI-Access SI and P, the body of work method (Cizek & Bunch, 2007; Kingston & Tiemann, 2012) was used with one round of range-finding and two rounds of pinpointing by the eight panels. Similarly, a cross-grade-level articulation session concluded the meetings, in which representatives from each content area and grade level reviewed and revised the cut scores recommended by the table panels. MDE provided the body of work documents, such as students' score distributions, picture cards, and test items, and all other related files and documents.

For more details regarding the MI-Access performance-level standard settings, refer to Measurement Incorporated's MI-Access Standard Setting Final Report (2015) and *MI-Access Standard Setting Final Report* (2017) in Appendix E.

9.5 Scale Scores

This section presents the slopes and intercepts for transforming thetas to scale scores, as well as the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS) for various MI-Access FI content areas. For SI and P, only raw scores were utilized in reporting, and cut scores based on the raw score points were derived from the standard-setting meetings. SI and P cut scores are presented in the next section, "MI-Access Supported Independence and Participation Cut Scores."

In creating FI scaling constants (slopes and intercepts), MDE fixed the LOSS and HOSS and ran a linear regression. MDE transformed the theta metric results onto a four-digital scale, which is consistent with the previous MI-Access FI scales and is easier and more meaningful to interpret for stakeholders. After obtaining the slopes (As), intercepts (Bs), and raw-to-theta conversion table (from the WINSTEPS calibration run), MDE applied the following formula to derive the scale score:

$$Scale\ score = (theta * slope) + intercept$$

More information regarding FI scaling and raw-to-scale-score conversion tables can be found in Chapter 7.

Table 9-3 presents the FI scaled cut scores derived from the standard-setting meetings, the scaling constants (slopes and intercepts) that transform the theta (or the raw score) to scale scores, and the LOSS and HOSS for each content area and grade level.

Table 9-3. FI Scaling Constants, Performance-Level Cut Scores, LOSS, and HOSS

Subject	Grade	B	A	LOSS	HOSS	Cut1	Cut2
ELA	3	2291.51628	16.61544	2200	2400	2300	2319
ELA	4	2393.75425	17.07504	2300	2500	2400	2423
ELA	5	2492.01440	17.44896	2400	2600	2499	2519
ELA	6	2596.15967	17.48863	2500	2700	2607	2626
ELA	7	2695.97419	17.98885	2600	2800	2698	2713
ELA	8	2796.46326	17.70695	2700	2900	2807	2821
ELA	11	3144.22115	28.84615	3000	3300	3151	3175
Mathematics	3	2299.03113	21.80787	2200	2400	2312	2344
Mathematics	4	2400.69428	21.32651	2300	2500	2410	2430
Mathematics	5	2499.51075	21.50306	2400	2600	2518	2543
Mathematics	6	2599.79136	21.58196	2500	2700	2611	2629
Mathematics	7	2699.42309	21.97561	2600	2800	2704	2730
Mathematics	8	2801.93852	20.66543	2700	2900	2810	2831
Mathematics	11	3149.66487	32.97428	3000	3300	3153	3185
Science	4	2390.735758	17.52848	2300	2500	2400	2412
Science	7	2690.97248	16.88619	2600	2800	2700	2716
Science	11	3093.11551	17.7841	3000	3200	3100	3118
Social Studies	5	2486.77337	17.77462	2400	2600	2500	2511
Social Studies	8	2793.07675	19.1168	2700	2900	2800	2810
Social Studies	11	3090.86026	18.11266	3000	3200	3100	3113

Notes: Cut1 = Level 2 (Attained) cut score and Cut2 = Level 3 (Surpassed) cut score.
ELA and mathematics cut scores are based on the Spring 2017 standard-setting results.
Science and Social Studies cut scores are based on the Spring 2015 standard setting results.

9.6 MI-Access Supported Independence and Participation Cut Scores

As mentioned above, for MI-Access SI and P, no IRT scaling was utilized and only raw scores were reported. Therefore, cut scores based on raw score points were derived from the standard setting meetings. Tables 9-4 and 9-5 present the ELA, mathematics, and science cut scores for SI and P, respectively.

Table 9-4. Supported Independence Performance-Level Cut Scores

Program	Subject	Grade	Cut1	Cut2
SI	ELA	3	28	43
SI	ELA	4	31	44
SI	ELA	5	30	46
SI	ELA	6	31	46
SI	ELA	7	31	46
SI	ELA	8	33	45
SI	ELA	11	35	46
SI	Mathematics	3	35	47
SI	Mathematics	4	34	45
SI	Mathematics	5	31	46
SI	Mathematics	6	32	44
SI	Mathematics	7	30	45
SI	Mathematics	8	30	46
SI	Mathematics	11	33	47
SI	Science	4	32	55
SI	Science	7	33	55
SI	Science	11	45	57

Notes: Cut1 = Level 2 (Attained) cut score and Cut2 = Level 3 (Surpassed) cut score.
ELA and mathematics cut scores are based on the Spring 2017 standard-setting results.
Science cut scores are based on the Spring 2015 standard-setting results.

Table 9-5. Participation Performance-Level Cut Scores

Program	Subject	Grade	Cut1	Cut2
P	ELA	3	31	45
P	ELA	4	32	43
P	ELA	5	28	42
P	ELA	6	29	41
P	ELA	7	28	45
P	ELA	8	27	43
P	ELA	11	34	46
P	Mathematics	3	33	47
P	Mathematics	4	32	47
P	Mathematics	5	32	46
P	Mathematics	6	31	44
P	Mathematics	7	27	43
P	Mathematics	8	28	43
P	Mathematics	11	31	46
P	Science	4	46	72
P	Science	7	44	72
P	Science	11	48	75

Notes: Cut1 = Level 2 (Attained) cut score and Cut2 = Level 3 (Surpassed) cut score.
ELA and mathematics cut scores are based on the Spring 2017 standard-setting results.
Science cut scores are based on the Spring 2015 standard-setting results.

9.7 Summary

This chapter presented a brief overview of the process for performance-level setting used by MI-Access for derivation of the MI-Access ELA, mathematics, science, and FI social studies cut scores. It also presented an overview of the methods and procedures used for FI scaling and scale scores, as well as SI and P reporting scores.

The standard settings undertaken by MI-Access support the following standards in the *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.

Chapter 10: Fairness

As noted in the *Standards* (AERA, APA, & NCME, 2014), there are varying definitions of fairness. This chapter examines test performance among varying subgroups assessed by MI-Access and fairness as it relates to minimizing bias on a test.

Differences in test performance among subgroups do not mean that a test is unfair—it simply means that groups performed differently 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 the students in the subgroup.

This chapter is particularly relevant to AERA, APA, & NCME (2014) *Standards* 3.1 through 3.6, found in Chapter 3, "Fairness in Testing," of the AERA, APA, & NCME (2014) *Standards*. Each of these standards will be presented below.

> **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. (p. 65)

There is no specific research on MI-Access 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, the Michigan Department of Education (MDE) and its supporting contractors follow several steps in the item development and selection processes as explained in section 10.1 of this chapter. In addition, MDE and Data Recognition Corporation (DRC) have conducted content and bias reviews on items, as explained in Chapter 3. These practices adhere to Standard 3.3:

> **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. (p. 64)

MDE conducts annual differential item functioning (DIF) studies following each administration of MI-Access. Typically, items are evaluated for possible DIF in the field-test phase of the test development process, and items flagged for DIF are typically further examined for possible bias. During test development, MDE follows procedures to minimize the inclusion of items that may potentially favor one demographic group over another. Section 10.2 of this chapter explains the steps taken to evaluate MI-Access items through the use of DIF to adhere to this standard.

In addition, standardized test administration and training of test administrators for MI-Access comply with *Standards* 3.4 and 3.5:

> **Standard 3.4** Test takers should receive comparable treatment during the test administration and scoring process. (p. 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. (p. 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. (p. 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. (p. 64)

Section 10.1 below explains the steps taken by MDE and DRC to minimize the use of words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Chapter 3 discusses content considerations during development and bias reviews for items included in MI-Access. These reviews are 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 are used to minimize bias are summarized below.

First, careful attention is 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 follow several published guidelines for reducing or eliminating bias.

Michigan educators, as item writers, and MDE staff, as item reviewers and test developers, follow documented bias and sensitivity guidelines to help ensure that the items are fair for all groups of test takers, despite differences in characteristics. These characteristics include, but are not limited to, disability status, ethnic group, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status. Test developers review all items included in MI-Access and other testing materials with these guidelines in mind.

Careful attention is given to item statistics (if available) throughout the test development process. As part of the test assembly process, attempts are made to avoid using or reusing items with poor statistics. Additional steps to reduce bias, including the use of content and bias committees comprised of Michigan educators, are described in more detail in Chapter 3 of this report. For MI-Access, all items—not only items that have DIF flags—are reviewed.

The goal of fairness in assessment is to ensure that test materials are as free as possible from unnecessary barriers to the success of diverse groups of students.

10.2 Evaluating Bias through Differential Item Functioning (DIF)

An empirical approach known as DIF is used to examine items after they have been administered. The DIF statistics indicate the degree to which members of a particular subgroup performed better or worse than expected on each item as compared to members of the reference group. 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). The DIF procedures and results are described in this section. Note that items are not necessarily suppressed from operational scoring if they are flagged for DIF. Due to small sample sizes for Supported Independence (SI) and Participation (P), DIF analysis is only done for Functional Independence (FI) assessments in English language arts (ELA), mathematics, science, and social studies.

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. Additionally, schools have been assigned the tasks of developing certain basic cognitive skills in students and supporting the development of these skills equitably among all students. Therefore, there is a need for tests that measure the skills and bodies of knowledge that are common to all learners. The test developers’ task is to create 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).

To lessen such biases, MDE 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 form construction processes to lessen the influence of these elements for large numbers of students (including the use of content and bias review committees). Unfortunately, in some cases, extraneous 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 test administration. DIF statistics are used to quantify differences in item performance between two groups after controlling for examinees' overall achievement level. For MI-Access FI, DIF is conducted for ELA, mathematics, science, and social studies using very similar procedures. Section 10.3 below provides DIF results for the following subgroups:

- ***Gender:*** The focal group is female; the reference group is male.
- ***Race/Ethnicity:*** The focal groups are students whose race/ethnicity is reported as African American or Black, Hispanic or Latino, or Asian; the reference group is students whose race/ethnicity is reported as White. However, due to the constraint of the sample size, race/ethnicity DIF for MI-Access FI involves only White and African American/Black students.
- ***Socioeconomic status:*** The focal group is students who are identified as economically disadvantaged (EconDis); the reference group is all others.
- ***Students with/without accommodations:*** The focal group is students who used test accommodation; the reference group is those students who did not use test accommodation.

10.3 DIF Statistics

Two commonly used DIF statistics were applied to MI-Access FI items and are described here. They are 1) the Mantel-Haenszel (MH) statistic (Mantel & Haenszel, 1959) for dichotomously scored items and an extension of the $MH\chi^2$ (Mantel, 1963) for polytomously scored items and 2) the standardized mean difference (SMD) effect size (ES) for polytomously scored items (Dorans & Schmitt, 1991).

For dichotomously scored items, such as multiple-choice (MC) items, the MH statistic is computed as follows (Camilli & Shepard, 1994):

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

where $VAR(A_j) = \frac{n_{Rj} n_{Fj} m_{1j} m_{0j}}{T_j^2 (T_j - 1)}$ and $E(A_j) = \frac{n_{Rj} m_{1j}}{T_j}$. (10.1)

In Equation 10.1, $A_j - E(A_j)$ represents the difference between the observed number and the expected number of correct responses on the item by the reference group members who have the *j*th score on the matching variable;[1] n_{Rj} and n_{Fj} represent the number of examinees in the reference and focal groups, respectively, for the *j*th score on the matching variable; m_{1j} represents the total number of examinees (both reference and focal) with the *j*th score on the matching variable and with a correct response on the current item; and m_{0j} represents the total number of examinees with the *j*th score on the matching variable and with an incorrect response on the current item. The $MH\chi^2$ is evaluated against the standard χ^2 critical with one degree of freedom.

[1] The total observed score is used as the matching variable for DIF analysis here.

The $MH\chi^2$ does not indicate the strength of association of the relationship between item performance and group membership. The MH odds ratio can be computed to estimate the strength of this association. The resulting estimate represents the relative likelihood of success on a particular item for members of two different groups of examinees (Camilli, 2006). This odds ratio thus provides an estimate of ES with a value of 1.0, indicating no DIF. A value greater than 1.0 indicates that, on average, the reference group members performed better than comparable focal group members did. A value less than 1.0 indicates that, on average, the reference group members performed worse than comparable focal group members did.

The odds of a correct response (proportion passing divided by proportion failing) is *P/Q* (i.e., *P/[1-P]*). The MH odds ratio is simply the odds of a correct response of the reference group divided by the odds of a correct response of the focal group. The formula for its estimation is as follows (Camilli & Shepard, 1994, p. 116):

$$\hat{\alpha}_{MH} = \frac{\sum_{j=1}^{S} A_j D_j / T_j}{\sum_{j=1}^{S} B_j C_j / T_j}, \quad (10.2)$$

where $S = K - 1$ and represents the actual number of 2×2 contingency tables (assuming the tables have at least 1 person in each cell); *K* represents the number of items on the test; and *j* signifies the *j*th score on the matching variable and runs from 0 to *K*.[2] For the *j*th score category, A_j represents the number of reference group members with a correct response, B_j represents the number of reference group members with an incorrect response, C_j represents the number of focal group members with a correct response, and D_j represents the number of focal group members with an incorrect response. T_j represents the total number of examinees who have the *j*th score on the matching variable.

The corresponding null hypothesis is that the odds of getting the item correct are equal for the two groups (the odds ratio is equal to 1):

$$H_0: \alpha_{MH} = 1 \quad (10.3)$$

To make the odds ratio symmetrical around zero with its range located in the interval $-\infty$ to $+\infty$, the odds ratio is transformed into a log-odds ratio as follows (Camilli & Shepard, 1994, p. 116):

$$\hat{\lambda}_{MH} = \log(\alpha_{MH}) \quad (10.4)$$

The natural logarithm transformation of this odds ratio is symmetrical around zero, where 0 indicates no DIF. This DIF measure is a signed index, where a positive value represents DIF in favor of the reference group and a negative value indicates DIF in favor of the focal group.

The variance of the log-odds ratio estimate ($V_{\hat{\lambda}}$) is computed as follows (Camilli & Shepard, 1994, p. 121):

$$V_{\hat{\lambda}} = \frac{\sum_{j=1}^{S} T_j^{-2} (A_j D_j + \alpha_{MH} B_j C_j)[A_j + D_j + \alpha_{MH}(B_j + C_j)]}{2(\sum_{j=1}^{S} A_j D_j / T_j)^2}. \quad (10.5)$$

[2] Although the value of the matching variable runs from 0 to *K*, the all correct (*K*) and all incorrect (0) score categories are not included in the DIF analysis in order to avoid having a denominator equal to 0.

The terms included in Equation 10.5 correspond to those presented for Equation 10.2. In practice, a standardized MH log-odds ratio is computed by dividing the estimate $\hat{\lambda}_{MH}$ by the estimated standard error. According to Penfield (2007, p. 16), "A value greater than 2.0 or less than -2.0 may be considered evidence of the presence of DIF."

In addition, once $\hat{\lambda}_{MH}$ is obtained using Equation 10.4, the delta statistic (MH D-DIF) can be computed as follows:

$$\text{MH D-DIF} = -2.35 \times \hat{\lambda}_{MH} \quad (10.6)$$

For polytomously scored items, an extension of the $MH\chi^2$ procedure was computed (Mantel, 1963). The statistic is computed as follows (Zwick, Donaghue, & Grima, 1993):

$$Mantel\ \chi^2 = \frac{(\sum_k F_k - \sum_k E(F_k))^2}{\sum_k VAR(F_k)}, \quad (10.7)$$

where F_k is the sum of scores for the focal group at the kth level of the matching variable and is defined as

$$F_k = \sum_t y_t n_{Ftk}, \quad (10.8)$$

the expectation of F_k under the hypothesis of no association is

$$E(F_k) = \frac{n_{F+k}}{n_{++k}} \sum_t y_t\, n_{+tk}, \quad (10.9)$$

and the variance of F_k under the assumption of no association is

$$\text{Var}(F_k) = \frac{n_{R+k} n_{F+k}}{n_{++k}^2 (n_{++k} - 1)} \left\{ \left(n_{++k} \sum_t y_t^2 n_{+tk} \right) - \left(\sum_t y_t\, n_{+tk} \right)^2 \right\}. \quad (10.10)$$

Using the Mantel approach for ordered categories, the data are organized into a $2 \times T \times K$ contingency table, where T is the number of response categories and K is the number of levels of the matching variable. $y_1, y_2, \ldots, y_T$ represent the T scores that can be obtained on the item, and n_{Rtk} and n_{Ftk} represent the number of examinees in the reference and focal groups, respectively, who are at the kth level of the matching variable and received an item score of y_t. The "+" denotes summation over a particular index (e.g., n_{R+k} denotes the total number of reference group members at the kth level of the matching variable). Under the null hypothesis of no association, the Mantel statistic has a chi-square distribution with one degree of freedom. For dichotomous items, the Mantel statistic reduces to the MH statistic (without the continuity correction).

In addition to the MH statistic, an ES was calculated by dividing the SMD statistics by the overall (focal and reference groups combined) standard deviation (SD) of the item scores: ES = SMD/SD. The SMD compares the mean of the reference and focal groups, adjusting for the distribution of reference and focal group members on the matching variable (Zwick et al., 1993), which for these analyses is the MI-Access FI raw score. SMD is computed as follows (Zwick et al., 1993):

$SMD = \sum_k p_{Fk\,(m_{Fk}-m_{Rk})}$ (10.11)

where p_{Fk} is the proportion of the focal group members at the kth level of the matching variable and m_{Fk} and m_{Rk} indicate mean item score for the focal group and the reference group at the kth level of the matching variable, respectively.

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.

10.3.1 Flagging Criteria and Results for FI ELA, Mathematics, Science, and Social Studies

For FI assessments, due to the sample size requirement, DIF was only computed with an n count equal or larger than 30 for both focal and reference groups. If either the focal group or the reference group n count is less than 30, then DIF is not computed.

The following flagging criteria, adapted from Penfield (2007), were used:

- Negligible DIF (a): if either MH common log-odds ratio ($\hat{\lambda}_{MH}$) is not significantly different from zero or $|\hat{\lambda}_{MH}| < 0.426$
- Moderate DIF (b): if $\hat{\lambda}_{MH}$ is significantly different from zero and $|\hat{\lambda}_{MH}| > 0.426$ and either (a) $|\hat{\lambda}_{MH}| < 0.638$ or (b) $|\hat{\lambda}_{MH}|$ is not significantly greater than 0.426
- Large DIF (c): if $|\hat{\lambda}_{MH}|$ is significantly greater than 0.426 and $|\hat{\lambda}_{MH}| > 0.638$

The following flagging criteria were used for polytomously scored items, based on Penfield (2007):

- AA: if either the Liu-Agresti cumulative common log-odds ratio ($\hat{\alpha}_{LA}$) is not significantly different from zero or $|\hat{\alpha}_{LA}| < 0.426$
- BB: if $\hat{\alpha}_{LA}$ is significantly different from zero and $|\hat{\alpha}_{LA}| \geq 0.426$ and either (a) $|\hat{\alpha}_{LA}| \leq 0.638$ or (b) $|\hat{\alpha}_{LA}|$ is not significantly greater than 0.426
- CC: if $|\hat{\alpha}_{LA}|$ is significantly greater than 0.426 and $|\hat{\alpha}_{LA}| > 0.638$

A positive MH D-DIF or ES value indicates that the item favors the focal group, while a negative value indicates that the item favors the reference group instead.

Table 10-1 shows the item counts for DIF analyses based on the Spring 2019 MI-Access FI administration. Tables 10-2 through 10-6 summarize the number of items having moderate or large DIF flags (b, c, bb, or cc) by mode and grade for each focal/reference group meeting the minimum n count.

For example, in the FI grade 3 ELA Accessing Print and Using Language (APUL) online assessment, only one item on the paper test (approximately 3.3% of all eligible items) was flagged for significant DIF for gender, which favors female group, while 2 items (approximately 6.7%) were flagged for economically disadvantaged/non-economically disadvantaged groups, with one item favoring the reference group (non-economically disadvantaged) and one item favoring the focal group (economically disadvantaged).

Table 10-1. Item Counts Used in Differential Item Functioning Analyses: FI ELA (APUL, EI), Mathematics, Science, and Social Studies

Content Area	Grade	Test Mode	N Items	Female/ Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
ELA: Accessing Print & Using Language	3	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	3	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	4	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	4	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	5	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	5	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	6	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	6	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	7	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	7	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	8	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	8	Paper	30	30	30	30	30
ELA: Accessing Print & Using Language	11	Online	30	30	30	30	30
ELA: Accessing Print & Using Language	11	Paper	30	30	30	30	30

Content Area	Grade	Test Mode	N Items	Female/ Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
ELA: Expressing Ideas	3	Paper	1	1	1	1	1
ELA: Expressing Ideas	4	Paper	1	1	1	1	1
ELA: Expressing Ideas	5	Paper	1	1	1	1	1
ELA: Expressing Ideas	6	Paper	1	1	1	1	1
ELA: Expressing Ideas	7	Paper	1	1	1	1	1
ELA: Expressing Ideas	8	Paper	1	1	1	1	1
ELA: Expressing Ideas	11	Paper	1	1	1	1	1
Mathematics	3	Online	24	24	24	24	24
Mathematics	3	Paper	24	24	24	24	24
Mathematics	4	Online	24	24	24	24	24
Mathematics	4	Paper	24	24	24	24	24
Mathematics	5	Online	24	24	24	24	24
Mathematics	5	Paper	24	24	24	24	24
Mathematics	6	Online	24	24	24	24	24
Mathematics	6	Paper	24	24	24	24	24
Mathematics	7	Online	24	24	24	24	24
Mathematics	7	Paper	24	24	24	24	24
Mathematics	8	Online	24	24	24	24	24
Mathematics	8	Paper	24	24	24	24	24
Mathematics	11	Online	24	24	24	24	24
Mathematics	11	Paper	24	24	24	24	24
Science	4	Online	35	35	35	35	35
Science	4	Paper	35	35	35	35	35
Science	7	Online	40	40	40	40	40
Science	7	Paper	40	40	40	40	40
Science	11	Online	45	45	45	45	45
Science	11	Paper	45	45	45	45	45
Social Studies	5	Online	32	32	32	32	32
Social Studies	5	Paper	32	32	32	32	32
Social Studies	8	Online	33	33	33	33	33
Social Studies	8	Paper	33	33	33	33	33
Social Studies	11	Online	41	41	41	41	41
Social Studies	11	Paper	41	41	41	41	41

Table 10-2. Number of Differential Item Functioning Flagged Items: FI Accessing Print and Using Language (APUL)

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
3	Online	b-	0	2	1	0
3	Online	b+	0	1	0	0
3	Online	c-	0	0	0	0
3	Online	c+	0	0	0	0
3	Paper	b-	0	1	0	--
3	Paper	b+	0	1	1	--
3	Paper	c-	0	0	0	--
3	Paper	c+	1	0	0	--
4	Online	b-	0	0	0	1
4	Online	b+	0	1	1	1
4	Online	c-	0	0	0	0
4	Online	c+	0	0	0	0
4	Paper	b-	1	0	--	0
4	Paper	b+	0	0	--	0
4	Paper	c-	0	0	--	0
4	Paper	c+	0	0	--	0
5	Online	b-	1	1	1	1
5	Online	b+	--	1	1	1
5	Online	c-	--	0	0	0
5	Online	c+	--	0	0	0
5	Paper	b-	--	1	--	0
5	Paper	b+	--	0	--	0
5	Paper	c-	--	0	--	0
5	Paper	c+	--	0	--	0
6	Online	b-	0	0	0	1
6	Online	b+	0	1	1	0
6	Online	c-	0	0	0	0
6	Online	c+	0	0	0	0
6	Paper	b-	0	0	--	0
6	Paper	b+	0	1	--	0
6	Paper	c-	0	0	--	0
6	Paper	c+	1	0	--	0

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
7	Online	b-	1	0	0	0
7	Online	b+	0	1	1	0
7	Online	c-	0	0	0	0
7	Online	c+	0	0	0	0
7	Paper	b-	0	0	--	2
7	Paper	b+	0	0	--	0
7	Paper	c-	0	0	--	0
7	Paper	c+	0	0	--	0
8	Online	b-	1	1	1	0
8	Online	b+	0	1	2	1
8	Online	c-	0	1	0	0
8	Online	c+	0	0	0	0
8	Paper	b-	1	0	1	0
8	Paper	b+	1	0	0	0
8	Paper	c-	0	0	0	0
8	Paper	c+	0	0	0	0
11	Online	b-	3	1	0	0
11	Online	b+	0	1	3	0
11	Online	c-	0	0	0	0
11	Online	c+	0	0	0	0
11	Paper	b-	1	1	0	1
11	Paper	b+	0	0	1	0
11	Paper	c-	0	0	0	0
11	Paper	c+	0	0	0	0

Note: "--" indicates that sample size for either the reference group or the focal group is too small (< 30), and thus, no DIF statistics and categories are computed.

Table 10-3. Number of Differential Item Functioning Flagged Items: FI Expressing Ideas (EI)

Grade	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
3	bb-	0	0	0	0
3	bb+	0	0	0	0
3	cc-	0	0	0	0
3	cc+	0	0	0	0
4	bb-	0	0	0	0
4	bb+	1	0	0	0
4	cc-	0	0	0	0
4	cc+	0	0	0	0
5	bb-	0	0	0	0
5	bb+	0	0	0	0
5	cc-	0	0	0	0
5	cc+	0	0	0	0
6	bb-	0	0	0	0
6	bb+	0	0	0	0
6	cc-	0	0	0	0
6	cc+	1	0	0	0
7	bb-	0	0	0	1
7	bb+	1	0	0	0
7	cc-	0	0	0	0
7	cc+	0	0	0	0
8	bb-	0	0	0	0
8	bb+	0	0	0	0
8	cc-	0	0	0	0
8	cc+	0	0	0	0
11	bb-	0	0	0	0
11	bb+	0	0	0	0
11	cc-	0	0	0	0
11	cc+	0	0	0	0

Table 10-4. Number of Differential Item Functioning Flagged Items: FI Mathematics

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
3	Online	b-	0	0	1	0
3	Online	b+	0	2	2	0
3	Online	c-	0	0	0	0
3	Online	c+	0	0	0	1
3	Paper	b-	0	0	0	0
3	Paper	b+	0	1	1	0
3	Paper	c-	0	0	0	0
3	Paper	c+	0	0	0	0
4	Online	b-	0	2	0	0
4	Online	b+	2	1	1	1
4	Online	c-	0	0	0	0
4	Online	c+	0	0	0	0
4	Paper	b-	0	0	--	1
4	Paper	b+	0	0	--	1
4	Paper	c-	0	0	--	0
4	Paper	c+	0	0	--	0
5	Online	b-	0	0	1	0
5	Online	b+	0	1	0	1
5	Online	c-	0	0	0	0
5	Online	c+	0	0	0	0
5	Paper	b-	1	0	--	0
5	Paper	b+	2	0	--	0
5	Paper	c-	0	0	--	0
5	Paper	c+	0	0	--	0
6	Online	b-	0	0	0	1
6	Online	b+	1	1	0	1
6	Online	c-	0	0	0	0
6	Online	c+	0	0	0	0
6	Paper	b-	0	0	--	0
6	Paper	b+	0	0	--	0
6	Paper	c-	0	0	--	0
6	Paper	c+	0	0	--	0

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
7	Online	b-	0	0	0	0
7	Online	b+	1	1	1	0
7	Online	c-	0	0	0	0
7	Online	c+	0	0	0	0
7	Paper	b-	0	1	0	0
7	Paper	b+	0	1	0	0
7	Paper	c-	0	0	0	0
7	Paper	c+	0	0	0	0
8	Online	b-	1	0	0	0
8	Online	b+	0	0	1	1
8	Online	c-	0	0	0	0
8	Online	c+	0	0	0	0
8	Paper	b-	1	0	0	0
8	Paper	b+	1	0	0	0
8	Paper	c-	0	0	0	0
8	Paper	c+	0	0	0	0
11	Online	b-	0	1	0	0
11	Online	b+	3	0	0	0
11	Online	c-	0	0	0	0
11	Online	c+	0	0	0	0
11	Paper	b-	0	0	0	0
11	Paper	b+	0	0	0	0
11	Paper	c-	0	0	0	0
11	Paper	c+	0	0	0	0

Notes: "--" indicates that sample size for either the reference group or the focal group is too small (i.e., < 30), and thus, no DIF statistics and categories are computed. For FI Mathematics online tests, there is no "Standard Accommodation" function(s) defined, and therefore, no such data were collected.

Table 10-5. Number of Differential Item Functioning Flagged Items: FI Science

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
4	Online	b-	1	1	0	NA
4	Online	b+	1	1	1	NA
4	Online	c-	0	1	0	NA
4	Online	c+	0	0	0	NA
4	Paper	b-	0	0	--	--
4	Paper	b+	2	0	--	--
4	Paper	c-	1	1	--	--
4	Online	b-	1	1	0	NA
7	Online	b-	2	2	1	NA
7	Online	b+	3	0	1	NA
7	Online	c-	0	0	0	NA
7	Online	c+	0	0	0	NA
7	Paper	b-	0	0	--	1
7	Paper	b+	0	1	--	0
7	Paper	c-	0	0	--	0
7	Paper	c+	1	1	--	0
11	Online	b-	1	2	1	NA
11	Online	b+	1	1	1	NA
11	Online	c-	0	0	0	NA
11	Online	c+	0	0	0	NA
11	Paper	b-	0	1	1	0
11	Paper	b+	0	1	0	1
11	Paper	c-	0	0	0	1
11	Paper	c+	0	0	0	0

Notes: "--" indicates that sample size for either the reference group or the focal group is too small (< 30), and thus, no DIF statistics and categories are computed. For FI Science online tests, there is no "Standard Accommodation" function(s) defined, and therefore, no such data were collected.

Table 10-6. Number of Differential Item Functioning Flagged Items: FI Social Studies

Grade	Test Mode	DIF Category	Female/Male	Black or African American/White	Economically Disadvantaged/ Non-Disadvantaged	With Accommodations/ Without Accommodations
5	Online	b-	0	0	0	NA
5	Online	b+	1	0	1	NA
5	Online	c-	0	0	0	NA
5	Online	c+	0	0	0	NA
5	Paper	b-	1	0	--	0
5	Paper	b+	0	1	--	0
5	Paper	c-	0	1	--	0
5	Paper	c+	0	0	--	0
8	Online	b-	0	2	1	NA
8	Online	b+	1	3	0	NA
8	Online	c-	0	0	0	NA
8	Online	c+	0	0	0	NA
8	Paper	b-	0	0	0	1
8	Paper	b+	0	0	0	0
8	Paper	c-	0	0	0	1
8	Paper	c+	0	0	0	0
11	Online	b-	1	1	1	NA
11	Online	b+	1	1	0	NA
11	Online	c-	0	0	0	NA
11	Online	c+	0	0	0	NA
11	Paper	b-	0	0	1	2
11	Paper	b+	1	0	0	1
11	Paper	c-	0	0	0	0
11	Paper	c+	0	0	0	0

Notes: "--" indicates that sample size for either the reference group or the focal group is too small (< 30), and thus, no DIF statistics and categories are computed. For FI Social Studies online tests, there is no "Standard Accommodation" function(s) defined, and therefore, no such data were collected.

10.4 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of MI-Access. The information in this chapter supports multiple best practices of the testing industry and in particular is related to 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 3.3—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.

Chapter 11: Reliability and Evidence of Construct-Related Validity

This chapter presents evidence supporting construct-related validity. Part of the test validity argument is that scores must be consistent and precise enough to be useful for the intended purposes. The concepts of reliability and precision are examined through analysis of measurement error in simulated and operational conditions.

This chapter demonstrates the adherence to AERA, APA, & NCME (2014) *Standards* 2.0, 2.3, 2.13, 2.14, 2.16, and 2.19. Each standard will be discussed in the pertinent section of this chapter.

11.1 Reliability

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

The AERA, APA, & NCME (2014) *Standards* says:

> 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). (p. 33)

In the development and maintenance of tests of the highest quality, the reliability of each MI-Access assessment has been calculated in accordance with the AERA, APA, & NCME (2014) *Standards*.

This chapter addresses several specific AERA, APA, & NCME (2014) standards. These include *Standards* 2.0, 2.3, 2.13, and 2.19; each is articulated below.

> **Standard 2.0**—Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use. (p. 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. (p. 43)

The total score reliabilities are discussed in section 11.1. The overall standard errors of measurement (SEMs) and conditional standard errors of measurement (CSEMs) are presented in sections 11.1.4 and 11.1.5.

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

The SEM based on scale scores and the CSEM based on scale scores are discussed below in sections 11.1.4 and 11.1.5.

> **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. (p. 47)

11.1.1 Reliability and Standard Error of Measurement

According to the classical true score theory, which is a fundamental component of the classical test theory (CTT), an observed score is a sum of two parts—a random component of true score (T) and a random component of error score (E), or mathematically, $X = T + E$ (McDonald, 1999). This model has the following properties: 1) the expected error score is zero, 2) the correlation between the true score and the error score is zero, and 3) the correlation between the error scores on different but parallel forms is zero (Lord & Novick, 1968).

Based on this model, a student's observed test score is an imprecise estimate of the student's actual ability because a portion of that score is attributable to random error. A fundamental theoretical quantity in test theory, the *reliability coefficient* of observed scores, is defined as the ratio of the variance of true scores to the variance of observed scores. Tests are therefore most reliable when the proportion of observed score variance that may be attributed to error variance is minimalized. According to McDonald (1999), test-retest methods, parallel or alternate-form methods, and internal analysis are the three recognized methods for estimating the reliability coefficient.

Due to practical difficulties in applying the first two above-mentioned methods, only the internal consistency reliability approach is described here. Estimates of internal consistency reliability involve "dividing the test into two or more constituent parts and in some way estimating reliability from the consistency of performance across these part-tests" (Haertel, 2006, p. 71).

11.1.2 Cronbach's Coefficient Alpha

Historically, various internal consistency reliability estimates have been proposed. However, the most widely used for fixed forms is Cronbach's (1951) coefficient alpha (Haertel, 2006). Using sample statistics, it is computed as follows (adapted from Haertel, 2006, p. 74):

$$\alpha = \frac{I}{I-1}\left(1 - \frac{\sum_{i=1}^{I} {s_i}^2}{{s_X}^2}\right) \quad (11.1)$$

where l represents the number of items on the test, S_i^2 represents the sample variance of item i, and S_X^2 represents the sample variance of the total raw score.

The use of coefficient alpha has several theoretical advantages (Haertel, 2006). First, since it equals the mean of all possible split-half reliability coefficients, which is another estimate of internal consistency reliability that involves the division of the total test into two "parallel" sub-tests, the use of coefficient alpha avoids the arbitrary choice of a split or division. Second, it is mathematically equivalent to one of the lower bounds of the theoretical reliability coefficient. The implication of this is that the theoretical reliability coefficient is higher than the observed coefficient alpha.

11.1.3 Standard Error of Measurement

SEM is related to reliability and is calculated with sample statistics as follows (Hays, 1994, p. 617):

$$\mathrm{SEM}(X) = S_X\sqrt{1 - r_{XX'}} \quad (11.2)$$

where $\mathrm{SEM}(X)$ represents the estimated SEM of the observed test score X, S_X denotes the estimated standard deviation (SD) (sample SD) of the observed score, and $r_{XX'}$ represents the estimated reliability coefficient of a test. In this report, the observed coefficient alpha is used as the estimated reliability coefficient for social studies.

According to Equation 11.2, the SEM is inversely related to the reliability of a test: For any SD of the observed score, the SEM decreases when the reliability coefficient increases. Thus, when an SEM is small, there can be more confidence in the accuracy, or precision, of the observed test scores.

11.1.4 Observed Reliability and SEM for MI-Access

Cronbach's coefficient alpha as the internal consistency reliability index was calculated using the Spring 2019 MI-Access administration data. The results for Functional Independence (FI) tests are presented in Tables 11-1 through 11-4.

As all operational items across FI online fixed form tests are the same, and the same raw-to-scale-score tables were used for all online forms, reliability and related statistics are reported for the combined online forms as compared with the paper/pencil form tests.

The results for Supported Independence (SI) are shown in Table 11-5 and the results for P are displayed in Table 11-6. For SI and P, only paper/pencil form tests were administered and all the operational items were the same across forms; therefore, one set of CTT-based internal consistency reliability statistics were computed for each assessment.

As shown in the FI tables, Cronbach's coefficient alpha values are very similar within the same context of content area by mode and grade level. For FI, Cronbach's coefficient alpha values range from 0.73 to about 0.88, indicating high moderate to strong internal consistency reliability. For SI, Cronbach's coefficient alphas range from 0.80 to 0.87. For P, Cronbach's coefficient alpha values range from 0.79 to 0.92 across all the content areas and grade levels, indicating relatively high moderate to strong internal consistency reliability.

Table 11-1. FI ELA Internal Consistency Reliability with Raw Score Mean and SEM by Mode and Grade Level

Subject	Grade	Mode	N	Mean	SD	SEM	Alpha
ELA	3	Online	901	21.95	5.98	2.44	0.83
ELA	3	Paper	109	22.03	6.25	2.44	0.85
ELA	4	Online	1059	22.18	5.71	2.43	0.82
ELA	4	Paper	108	20.80	6.42	2.49	0.85
ELA	5	Online	1129	22.71	5.86	2.34	0.84
ELA	5	Paper	106	23.31	5.70	2.31	0.84
ELA	6	Online	1340	22.84	5.97	2.33	0.85
ELA	6	Paper	98	22.57	5.97	2.35	0.85
ELA	7	Online	1292	23.18	5.63	2.32	0.83
ELA	7	Paper	118	23.20	6.02	2.33	0.85
ELA	8	Online	1280	24.38	6.06	2.26	0.86
ELA	8	Paper	127	24.42	5.77	2.30	0.84
ELA	11	Online	1120	24.39	5.70	2.30	0.84
ELA	11	Paper	137	24.71	5.74	2.28	0.84

Table 11-2. FI Mathematics Internal Consistency Reliability with Raw Score Mean and SEM by Mode and Grade Level

Subject	Grade	Mode	N	Mean	SD	SEM	Alpha
Mathematics	3	Online	906	15.37	4.79	2.07	0.81
Mathematics	3	Paper	107	15.72	4.76	2.05	0.81
Mathematics	4	Online	1079	14.87	4.42	2.18	0.76
Mathematics	4	Paper	107	15.25	4.83	2.14	0.80
Mathematics	5	Online	1167	15.48	4.78	2.12	0.80
Mathematics	5	Paper	108	15.10	4.93	2.12	0.81
Mathematics	6	Online	1395	14.36	4.21	2.20	0.73
Mathematics	6	Paper	105	15.08	4.74	2.14	0.80
Mathematics	7	Online	1369	14.82	4.16	2.15	0.73
Mathematics	7	Paper	123	14.80	4.81	2.14	0.80
Mathematics	8	Online	1336	14.84	4.87	2.14	0.81
Mathematics	8	Paper	131	15.49	5.12	2.11	0.83
Mathematics	11	Online	1191	13.97	4.71	2.19	0.78
Mathematics	11	Paper	137	14.28	4.88	2.17	0.80

Table 11-3. FI Science Internal Consistency Reliability with Raw Score Mean and SEM by Mode and Grade Level

Subject	Grade	Mode	N	Mean	SD	SEM	Alpha
Science	4	Online	937	21.33	6.66	2.61	0.85
Science	4	Paper	98	21.41	6.98	2.61	0.86
Science	7	Online	1297	24.85	6.62	2.74	0.83
Science	7	Paper	110	26.40	6.80	2.71	0.84
Science	11	Online	1205	27.80	8.19	2.95	0.87
Science	11	Paper	138	27.55	8.49	2.95	0.88

Table 11-4. FI Social Studies Internal Consistency Reliability with Raw Score Mean and SEM by Mode and Grade Level

Subject	Grade	Mode	N	Mean	SD	SEM	Alpha
Social Studies	5	Online	1118	17.46	6.14	2.59	0.82
Social Studies	5	Paper	106	18.27	5.54	2.57	0.79
Social Studies	8	Online	1323	18.17	6.04	2.65	0.81
Social Studies	8	Paper	129	18.09	6.80	2.62	0.85
Social Studies	11	Online	1206	22.42	7.49	2.93	0.85
Social Studies	11	Paper	139	23.62	8.23	2.85	0.88

Table 11-5. SI ELA, Mathematics, and Science Internal Consistency Reliability with Raw Score Mean and SEM by Grade Level

Subject	Grade	N	Mean	SD	SEM	Alpha
ELA	3	483	40.67	13.22	5.47	0.83
ELA	4	448	39.89	12.80	5.54	0.81
ELA	5	394	38.98	14.99	5.40	0.87
ELA	6	467	41.46	13.06	5.37	0.83
ELA	7	460	39.90	12.50	5.64	0.80
ELA	8	449	41.67	12.94	5.32	0.83
ELA	11	467	39.92	13.17	5.62	0.82
Mathematics	3	479	39.07	14.57	5.47	0.86
Mathematics	4	446	36.63	14.41	5.80	0.84
Mathematics	5	390	36.05	13.36	5.97	0.80
Mathematics	6	469	33.50	13.93	5.92	0.82
Mathematics	7	458	35.29	12.90	5.82	0.80
Mathematics	8	452	35.23	13.10	5.73	0.81
Mathematics	11	467	41.15	14.64	5.18	0.87
Science	4	446	47.96	15.43	5.83	0.86

Subject	Grade	N	Mean	SD	SEM	Alpha
Science	7	455	45.32	15.09	6.26	0.83
Science	11	467	50.01	14.84	5.69	0.85

Table 11-6. P ELA, Mathematics, and Science Internal Consistency Reliability with Raw Score Mean and SEM by Grade Level

Subject	Grade	N	Mean	SD	SEM	Alpha
ELA	3	487	32.55	16.59	7.02	0.82
ELA	4	406	35.22	17.17	6.75	0.85
ELA	5	396	33.08	16.44	6.88	0.82
ELA	6	363	31.43	16.26	7.05	0.81
ELA	7	328	32.14	16.44	7.03	0.82
ELA	8	334	34.08	15.46	7.00	0.79
ELA	11	375	32.01	17.66	6.62	0.86
Mathematics	3	488	31.07	17.38	7.01	0.84
Mathematics	4	406	30.53	16.84	7.07	0.82
Mathematics	5	396	30.80	15.99	7.16	0.80
Mathematics	6	364	31.21	16.79	6.99	0.83
Mathematics	7	330	31.61	17.13	6.92	0.84
Mathematics	8	335	34.91	15.85	6.82	0.82
Mathematics	11	376	30.70	17.36	6.79	0.85
Science	4	407	53.59	26.78	8.09	0.91
Science	7	324	50.39	25.30	8.35	0.89
Science	11	374	50.67	27.36	7.94	0.92

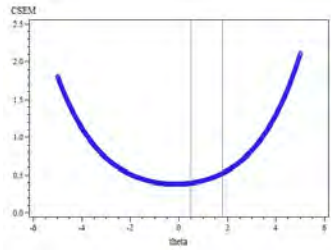
11.1.5 SEM for FI Tests

In addition to the CTT-based reliability and SEM presented in the previous section, the item response theory (IRT) CSEM was calculated for FI ELA, mathematics, science, and social studies. Related numerical information can be found in corresponding conversion tables reported in Appendix F). These CSEM graphs are presented in Figures 11-1 through 11-4 below.

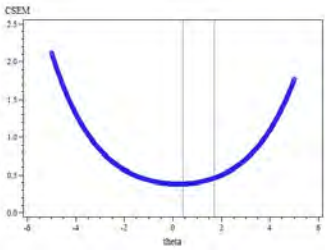
As shown in these figures, in most cases, the CSEMs are the lowest at level 1 and level 2 cut scores (the first vertical line, which indicates the cut between Emerging and Attained). In some cases, they are not the lowest at the Emerging/Attained cut. This might be due to the relatively small sample sizes for FI tests. Also note, these CSEM curves are generated using the post-administration estimated item parameters.

Figure 11-1. IRT-Based CSEM Curves for FI English Language Arts by Grade

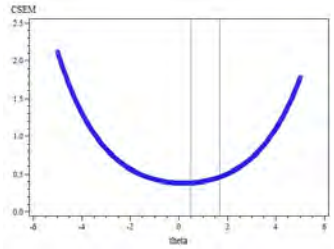
Conditional Standard Errors of Measurement for FI Grade 3 English Language Arts



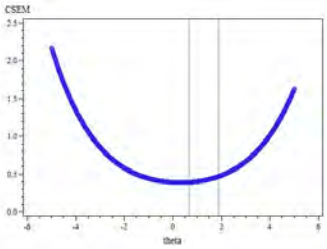
Conditional Standard Errors of Measurement for FI Grade 4 English Language Arts



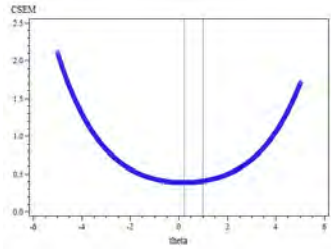
Conditional Standard Errors of Measurement for FI Grade 5 English Language Arts



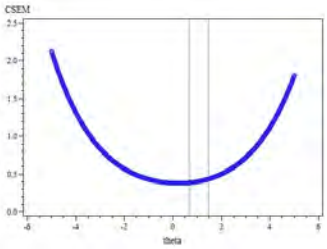
Conditional Standard Errors of Measurement for FI Grade 6 English Language Arts



Conditional Standard Errors of Measurement for FI Grade 7 English Language Arts



Conditional Standard Errors of Measurement for FI Grade 8 English Language Arts



Conditional Standard Errors of Measurement for FI Grade 11 English Language Arts

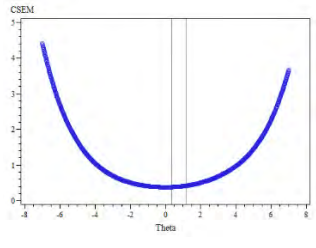
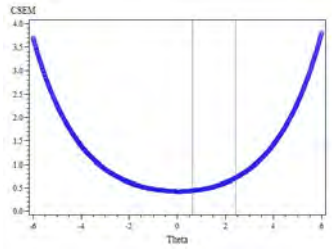
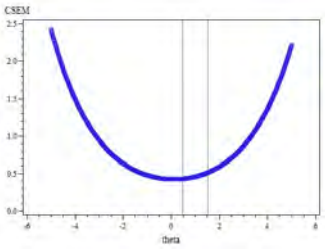


Figure 11-2. IRT-Based CSEM Curves for FI Mathematics by Grade

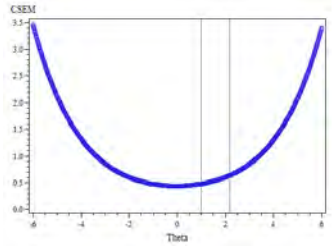
Conditional Standard Errors of Measurement for FI Grade 3 Mathematics



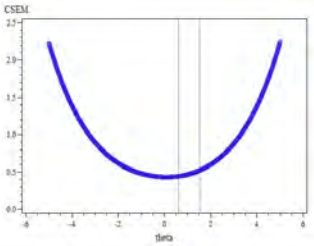
Conditional Standard Errors of Measurement for FI Grade 4 Mathematics



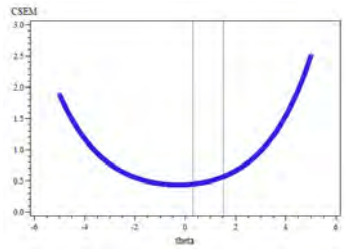
Conditional Standard Errors of Measurement for FI Grade 5 Mathematics



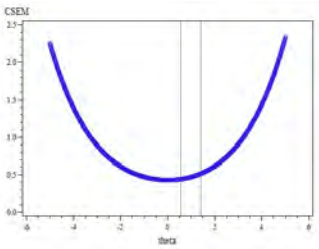
Conditional Standard Errors of Measurement for FI Grade 6 Mathematics



Conditional Standard Errors of Measurement for FI Grade 7 Mathematics



Conditional Standard Errors of Measurement for FI Grade 8 Mathematics



Conditional Standard Errors of Measurement for FI Grade 11 Mathematics

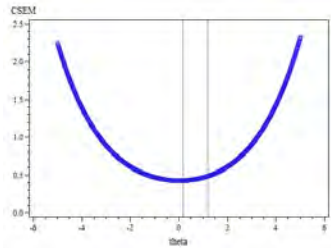
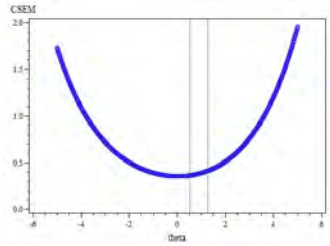


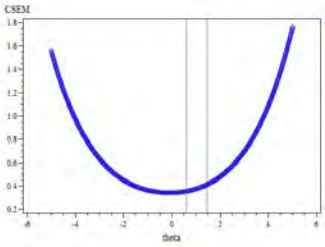
Figure 11-3 IRT-Based CSEM Graphs for FI Science by Grade

Conditional Standard Errors of Measurement for FI Grade 4 Science

Conditional Standard Errors of Measurement for FI Grade 4 Science



Conditional Standard Errors of Measurement for FI Grade 7 Science



Conditional Standard Errors of Measurement for FI Grade 11 Science

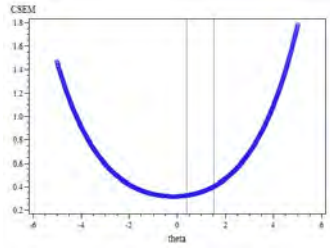
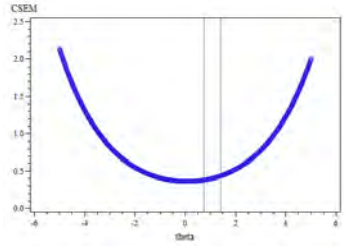
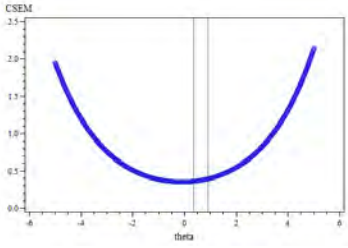


Figure 11-4. IRT-Based CSEM Graphs for FI Social Studies by Grade

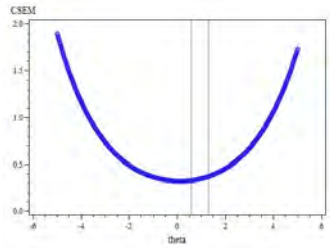
Conditional Standard Errors of Measurement for FI Grade 5 Social Studies



Conditional Standard Errors of Measurement for FI Grade 8 Social Studies



Conditional Standard Errors of Measurement for FI Grade 11 Social Studies



11.1.6 Inter-Rater Reliability of FI Expressing Ideas Handscoring

FI ELA: Expressing Ideas (EI) has one operational constructed-response (CR) item and two field-test CR items on each grade level test, which were hand scored. A second rating was done on a sample of the papers. Table 11-7 below presents the inter-rater reliability in terms of exact agreement and adjacent agreement rates. As shown in the table, there is a very high exact agreement rate plus adjacent agreement rate between the two ratings (ranging from 97.6% to 100%), thus showing very high inter-rater reliability.

Table 11-7. Inter-Rater Reliability of FI EI Handscored Items

Grade	CR Item	% Perfect +	N Perfect	% Perfect	N Adj	% Adj	N Nonadj	% Nonadj
3	OP	100.0	76	73.1	28	26.9	0	0.0
3	Form 1 FT	100.0	59	86.8	9	13.2	0	0.0
3	Form 2 FT	100.0	28	90.3	3	9.7	0	0.0
4	OP	99.2	93	76.9	27	22.3	1	0.8
4	Form 1 FT	100.0	64	80.0	16	20.0	0	0.0
4	Form 2 FT	100.0	32	76.2	10	23.8	0	0.0
5	OP	100.0	109	86.5	17	13.5	0	0.0
5	Form 1 FT	100.0	46	86.8	7	13.2	0	0.0
5	Form 2 FT	100.0	51	81.0	12	19.0	0	0.0
6	OP	100.0	119	81.0	28	19.0	0	0.0
6	Form 1 FT	98.8	56	66.7	27	32.1	1	1.2
6	Form 2 FT	100.0	53	80.3	13	19.7	0	0.0
7	OP	100.0	107	75.4	35	24.6	0	0.0
7	Form 1 FT	98.5	58	85.3	9	13.2	1	1.5
7	Form 2 FT	97.6	64	76.2	18	21.4	2	2.4
8	OP	100.0	106	74.6	36	25.4	0	0.0
8	Form 1 FT	100.0	47	69.1	21	30.9	0	0.0
8	Form 2 FT	100.0	48	78.7	13	21.3	0	0.0
11	OP	97.8	114	84.4	18	13.3	3	2.2
11	Form 1 FT	100.0	50	74.6	17	25.4	0	0.0
11	Form 2 FT	100.0	44	72.1	17	27.9	0	0.0

11.2 Classification Accuracy and Consistency for MI-Access Assessments

Based on the raw-to-scale-score (R2SS) conversion tables for FI and the raw-to-performance-level (PL) conversion tables for SI and P, student performance in corresponding content areas is classified into one of the three PLs (Emerging Toward the Performance Standard, Attained the Performance Standard, and Surpassed the Performance Standard). Among these, the most important classification is between the Emerging and Attained (Level 1/Level 2) cut. While it is always important to know the reliability of student scores in any examination, it is also important to assess the quality of the decisions, especially with regard to the Attained or not cut. Such evaluation was performed through estimation of the probabilities of accurate and consistent classification of student performance.

Classification accuracy is defined as the extent to which the actual classifications of examinees 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. The magnitude of classification accuracy measures is influenced by key features of the test design, including the number of items, the number of cut scores, reliability, and associated SEM or CSEM.

For MI-Access FI mathematics, science, and social studies, each test under consideration consists only of equally weighted and dichotomously scored items. Procedures from Hanson and Brennan (1990) were applied to derive classification accuracy and classification consistency measures. For FI English language arts (ELA), which contains a CR item, and for SI and P ELA, which contain polytomously scored items, Livingston and Lewis's (1995) more complex procedures that accommodate CR items were used. Moreover, the definitions for accuracy and consistency of decisions presented in Young and Yoon (1998) were adopted here.

Specifically, the accuracy of decisions is the extent to which decisions would agree with those that would be made if each student could somehow be tested with all possible forms of an examination; and the consistency of decisions is the extent to which decisions would agree with those that would be made if each student had taken a parallel form of the examination, equal in difficulty and covering the same content as the form the student actually took (Young & Yoon, 1998). These ideas are shown schematically in Figures 11-5 and 11-6 with reference to a MI-Access test as an example.

In both figures, "Achieves Attained Status" refers to the Attained the Performance Standard (Level 2) and Surpassed the Performance Standard (Level 3) categories on the total raw score and "Does Not Achieve Attained Status" refers to the Emerging Toward the Performance Standard category below the Attained (Level 1/Level 2) cut.

Figure 11-5. Classification Accuracy

		Decision made on a form actually taken	Decision made on a form actually taken
		Does Not Achieve Proficient Status	Achieves Proficient Status
"True status" based on all-forms average	Does Not Achieve Proficient Status	Correct Classification	Misclassification
	Achieves Proficient Status	Misclassification	Correct Classification

Note: Adapted from Young and Yoon (1998)

Figure 11-6. Classification Consistency

		Decision made on the 2nd form taken	Decision made on the 2nd form taken
		Does Not Achieve Proficient Status	Achieves Proficient Status
Decision made on the 1st form taken	Does Not Achieve Proficient Status	Consistent Classification	Inconsistent Classification
	Achieves Proficient Status	Inconsistent Classification	Consistent Classification

Note: Adapted from Young and Yoon (1998)

In Figure 11-5, accurate classification occurs when the decision made on the basis of the form actually taken agrees with the decision made on the basis of the theoretical "all-forms" average. Misclassification occurs, for example, when a student who "Does Not Achieve Attained Status" based on the student's "all-forms" average is classified incorrectly as "Achieves Attained Status."

In Figure 11-6, consistent classification occurs when two possible alternate forms agree on the classification of a student as either "Achieves Attained Status" or "Does Not Achieve Attained Status," whereas inconsistent classification occurs when the decisions made by the forms differ.

The analyses make use of the techniques outlined and implemented by Hanson and Brennan (1990), Brennan (2004), and Livingston and Lewis (1995). Specifically, a four-parameter beta distribution was used to model the true score, and Lord's (1965) two-term approximation to the compound binomial distribution was used to model the conditional error. The BB-CLASS software (Version 1.1) was used to complete these analyses (Brennan, 2004).

Tables 11-8 through 11-17 present the analysis results of decision accuracy and consistency for classifying students at each grade level per test form as "Achieves Attained Status" or "Does Not Achieve Attained Status" based on their respective MI-Access total raw scores. For FI, because the operational items were exactly the same across the online forms, the raw score statistics were very similar across forms and mode (Chapter 7 for details), and the online R2SS tables were used for reporting, the combined classification indexes for FI were reported here.

In addition to classification accuracy and consistency, Tables 11-8 through 11-17 provide information on the proportion of false positives and false negatives (the two types of misclassification). The false positive is the type of misclassification in which students should be classified in the "Does Not Achieve Attained Status" category based on their "all-forms" average but instead end up in the "Achieves Attained Status" category based on the actual form. The false negative is just the opposite—students who should be in the "Achieves Attained Status" category based on their "all-forms" average end up in the "Does Not Achieve Attained Status" category based on the actual form. The sum of the proportion values for accuracy, false positives, and false negatives should be equal to 1.00. Due to rounding, however, the sum of these values in the tables may not be equal to 1.00.

As shown in Tables 11-8 through 11-17, the proportion of false positives (the labeling of a student as Attained the Performance Standard when he or she should be categorized as Emerging Toward the Performance Standard) ranged from 0.12 to 0.24 for FI ELA. This proportion of false positives ranged from 0.17 to 0.21 for FI mathematics, from 0.12 to 0.16 for FI science, and from 0.14 to 0.16 for FI social studies. Moreover, the proportion of false negatives (the labeling of a student as Emerging when the student should be categorized as Attained) ranged from 0.02 to 0.05 for FI ELA, from 0.06 to 0.09 for FI mathematics, from 0.06 to 0.08 for FI science, and from 0.05 to 0.08 for FI social studies. Similar patterns were found for SI and P tests as well.

The last columns in Tables 11-8 through 11-17 report the proportion of students predicted by the model who would be assigned to the same category (either Attained or Emerging) if an alternate form of MI-Access (with similar content coverage and item difficulty as the actual form) had been administered. These values range from 0.63 to 0.82 for FI, from 0.74 to 0.77 for SI, and from 0.72 to 0.75 for P across content area by grade level contexts.

Table 11-8. Classification Accuracy and Consistency on MI-Access Functional Independence English Language Arts Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.83	0.12	0.05	0.77
4	0.81	0.16	0.04	0.77
5	0.80	0.15	0.05	0.76
6	0.75	0.24	0.02	0.76
7	0.81	0.17	0.02	0.82
8	0.78	0.19	0.02	0.79
11	0.76	0.22	0.02	0.78

Table 11-9. Classification Accuracy and Consistency on MI-Access Functional Independence Mathematics Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.78	0.16	0.06	0.71
4	0.72	0.19	0.09	0.64
5	0.76	0.17	0.07	0.69
6	0.71	0.21	0.08	0.63
7	0.73	0.20	0.08	0.64
8	0.74	0.18	0.08	0.67
11	0.74	0.18	0.09	0.65

Table 11-10. Classification Accuracy and Consistency on MI-Access Functional Independence Science Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
4	0.77	0.15	0.08	0.71
7	0.77	0.16	0.07	0.69
11	0.81	0.12	0.06	0.74

Table 11-11. Classification Accuracy and Consistency on MI-Access Functional Independence Social Studies Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
5	0.80	0.15	0.05	0.74
8	0.76	0.16	0.08	0.69
11	0.80	0.14	0.07	0.73

Table 11-12. Classification Accuracy and Consistency on MI-Access Supported Independence English Language Arts Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.84	0.08	0.08	0.77
4	0.83	0.08	0.09	0.76
5	0.83	0.09	0.08	0.76
6	0.84	0.08	0.08	0.78
7	0.83	0.09	0.08	0.76
8	0.82	0.10	0.07	0.76
11	0.82	0.09	0.09	0.75

Table 11-13. Classification Accuracy and Consistency on MI-Access Supported Independence Mathematics Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.82	0.09	0.09	0.75
4	0.81	0.09	0.09	0.74
5	0.82	0.09	0.09	0.75
6	0.82	0.10	0.08	0.75
7	0.82	0.11	0.07	0.75
8	0.81	0.14	0.05	0.76
11	0.82	0.09	0.09	0.75

Table 11-14. Classification Accuracy and Consistency on MI-Access Supported Independence Science Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
4	0.84	0.08	0.07	0.77
7	0.83	0.11	0.06	0.77
11	0.81	0.09	0.10	0.74

Table 11-15. Classification Accuracy and Consistency on MI-Access Participation English Language Arts Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.81	0.09	0.10	0.74
4	0.80	0.09	0.10	0.74
5	0.81	0.09	0.10	0.74
6	0.82	0.09	0.09	0.75
7	0.81	0.09	0.10	0.74
8	0.82	0.09	0.09	0.74
11	0.81	0.09	0.10	0.74

Table 11-16. Classification Accuracy and Consistency on MI-Access Participation Mathematics Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
3	0.82	0.09	0.09	0.75
4	0.81	0.09	0.09	0.74
5	0.79	0.09	0.11	0.72
6	0.81	0.09	0.10	0.75
7	0.82	0.09	0.09	0.75
8	0.81	0.08	0.10	0.74
11	0.81	0.09	0.10	0.74

Table 11-17. Classification Accuracy and Consistency on MI-Access Participation Science Total Raw Score

Grade	Classification Accuracy	False Positive	False Negative	Classification Consistency
4	0.80	0.09	0.11	0.72
7	0.80	0.09	0.11	0.72
11	0.79	0.10	0.11	0.72

11.3 Assumption of Unidimensionality

Another measure of construct validity is unidimensionality. One of the underlying assumptions of the IRT models used to scale MI-Access FI content area tests is that the items being calibrated are unidimensional; that is, items composing FI tests in each grade/content area measure a single content domain. For example, mathematics items should measure mathematics 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. (pp. 26–27)

For MI-Access FI, the Michigan Department of Education (MDE) conducted two analyses to evaluate the unidimensionality assumption with operational items only. The first set was an exploratory factor analysis (EFA) using the Mplus software (Muthén & Muthén, 2012) with the weighted least square mean and variance adjusted (WLSMV) estimator.[1] Barendse, Oort, and Timmerman (2015) found that WLSMV is the preferred estimation method and is recommended to rely on the root mean squared error of approximation (RMSEA) index (in which values less than 0.05 are desired) if the primary interest is in major factors.

[1] WLSMV-weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic that use a full weight matrix" (Muthén and Muthén, 2012, p. 603)

The second set of analyses is a principal component analysis (PCA) using Statistical Analysis System (SAS) software, i.e. SAS Enterprise Guide Version 7.1. For PCA results, the magnitude of the first and second eigenvalues are examined. Both the eigenvalues-greater-than-one rule and the scree plot approach were considered. The RMSEA values for one-factor EFA models and the first two eigenvalues from each PCA model are reported in Tables 11-18 through 11-21.

As shown in Tables 11-18 through 11-21, the dimensionality assessment for FI is examined by mode at each grade level. As seen in these tables, generally speaking, both the EFA and PCA results failed to reject the unidimensionality assumption, which is a supporting piece of evidence for the use of unidimensional IRT models at each content/grade combination for FI tests.

Table 11-18. The First Two Component Eigenvalues and Variance Explained from PCA and RMSEA from 1-Factor EFA for FI ELA

Grade	Mode	RMSEA (1-Factor EFA)	PCA First Eigenvalue	1st Component Variance Explained	PCA Second Eigenvalue	2nd Component Variance Explained
3	Online	0.029	5.6429	0.1820	1.5231	0.0491
3	Paper	0.035	6.2853	0.2028	2.0810	0.0671
4	Online	0.029	5.2174	0.1683	1.5514	0.0500
4	Paper	0.030	6.0745	0.1960	2.0106	0.0649
5	Online	0.028	5.7896	0.1868	1.5177	0.0490
5	Paper	0.024	5.8787	0.1896	2.0367	0.0657
6	Online	0.025	6.0264	0.1944	1.5033	0.0485
6	Paper	0.005	6.1790	0.1993	2.0040	0.0646
7	Online	0.019	5.8158	0.1876	1.3671	0.0441
7	Paper	0.022	6.3071	0.2035	1.7608	0.0568
8	Online	0.021	6.5496	0.2113	1.3368	0.0431
8	Paper	0.018	6.2686	0.2022	1.8809	0.0607
11	Online	0.026	5.8173	0.1877	1.3893	0.0448
11	Paper	0.026	6.0900	0.1965	1.8292	0.0610

Table 11-19. The First Two Component Eigenvalues and Variance Explained from PCA and RMSEA from 1-Factor EFA for FI Math

Grade	Mode	RMSEA (1-Factor EFA)	PCA First Eigenvalue	1st Component Variance Explained	PCA Second Eigenvalue	2nd Component Variance Explained
3	Online	0.026	4.6951	0.1956	1.2433	0.0518
3	Paper	0.041	4.9158	0.2048	2.2022	0.0843
4	Online	0.031	3.7760	0.1573	1.3759	0.0573
4	Paper	0.035	4.4888	0.1870	1.7637	0.0735
5	Online	0.029	4.5127	0.1880	1.2200	0.0508
5	Paper	0.040	4.8239	0.2010	1.6939	0.0706
6	Online	0.039	3.4764	0.1449	1.3822	0.0576
6	Paper	0.041	4.6035	0.1918	2.0071	0.0836
7	Online	0.048	3.6300	0.1513	1.4962	0.0623
7	Paper	0.044	4.5802	0.1908	1.9042	0.0793
8	Online	0.037	4.5893	0.1912	1.2866	0.0536
8	Paper	0.022	5.2123	0.2172	1.5980	0.0666
11	Online	0.027	4.1249	0.1719	1.2747	0.0531
11	Paper	0.027	4.4568	0.1857	1.7108	0.0713

Table 11-20. The First Two Component Eigenvalues and Variance Explained from PCA and RMSEA from 1-Factor EFA for FI Science

Grade	Mode	RMSEA (1-Factor EFA)	PCA First Eigenvalue	1st Component Variance Explained	PCA Second Eigenvalue	2nd Component Variance Explained
4	Online	0.024	5.9002	0.1686	1.5291	0.0437
4	Paper	0.028	6.5246	0.1864	2.3835	0.0681
7	Online	0.019	5.6881	0.1422	1.5345	0.0384
7	Paper	0.024	6.1849	0.1546	2.3254	0.0581
11	Online	0.019	7.1385	0.1586	1.5844	0.0352
11	Paper	0.025	7.7151	0.1714	2.3912	0.0531

Table 11-21. The First Two Component Eigenvalues and Variance Explained from PCA and RMSEA from 1-Factor EFA for FI Social Studies

Grade	Mode	RMSEA (1-Factor EFA)	PCA First Eigenvalue	1st Component Variance Explained	PCA Second Eigenvalue	2nd Component Variance Explained
5	Online	0.027	5.1150	0.1598	1.5003	0.0469
5	Paper	0.022	4.5333	0.1417	2.0928	0.0654
8	Online	0.025	4.8509	0.1470	1.4115	0.0428
8	Paper	0.033	6.1051	0.1850	2.0324	0.0616
11	Online	0.024	6.0691	0.1480	1.5348	0.0374
11	Paper	0.022	7.3663	0.1797	2.2499	0.0549

11.4 Validity Evidence

The *Standards for Educational and Psychological Testing* defines validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests" (AERA, APA, & NCME, 2014).

The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for particular purposes or uses. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process. Every aspect of an assessment provides evidence that either supports or challenges its validity, including design, content specifications, item development, psychometric quality, and inferences made from the results.

The validity of score interpretations for MI-Access is supported by multiple sources of evidence. Chapter 1 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) specifies the following sources of validity evidence that are important to gather and document in order to support validity claims for an assessment:

- Test content
- Response processes
- Internal test structure
- Relation to other variables
- Consequences of test use

It is important to note that these categories are not mutually exclusive. One source of validity evidence often falls into more than one category, as discussed in more detail in this section. The process of gathering evidence of the validity of score interpretations is best characterized as ongoing throughout test development, administration, scoring, reporting, and beyond.

As the technical report has progressed, it has covered the different phases of the testing cycle. Each part of the technical report detailed the procedures and processes applied in Michigan, as well as the corresponding results. Each part also highlighted the meaning and significance of the procedures, processes, and results in terms of validity and their relationship to specific sections

of the *Standards*. The current section now addresses these final issues in validity: test content, response processes, internal test structure, relation to other variables, and consequences of test use.

11.4.1 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 steps 1 through 5 and Chapter 8 for more information on step 6).

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 (for instance, one administration may be timed, but another administration may be untimed), differences in student performance may be partially associated with the different administration conditions. Careful specification of content and review of the items representing that content are the 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 and are designed to ensure that content is appropriately represented.

11.4.2 Evidence Based on Test Content

According to the *Standards*, evidence based on test content "can include logical or empirical analyses of the adequacy with which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores" (AERA, APA, & NCME, 2014). Documentation of the content domains, how the content is sampled and represented, and alignment of items to the content were discussed in Chapter 3 of this report. The documentation showed how test specification documents derived from earlier developmental activities guided the final phases of test development and ultimately yielded the test forms that were administered to students.

Chapter 3 also showed that the participation of Michigan educators in that process provided a solid rationale for having confidence in the content and design of MI-Access as a tool from which to derive valid inferences about Michigan student performance. Particularly for science and social studies, use of classroom teachers also brought into the process the enacted curriculum perspective and the written curriculum perspective. The test development process and the involvement of Michigan educators in that process formed an important part of the validity of the entire MI-Access assessment.

11.4.3 Evidence Based on Response Process

According to the *Standards*, evidence based on response processes "generally comes from analyses of individual responses" (AERA, APA, & NCME, 2014, p. 15). Hence, the best opportunity for detecting and eliminating potential sources of invalidity occurs during the test development process (U.S. Department of Education, 2015).

As described in Chapter 3, all items for MI-Access were carefully reviewed through multiple cycles of the item development process for ambiguity, bias, sensitivity, irrelevance, and inaccuracy to ensure a fit between the construct and the nature of the actual performance.

11.4.4 Evidence Based on Internal Test Structure

According to the *Standards*, evidence based on internal structure reflects "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (AERA, APA, & NCME, 2014, p. 13). Three important sources of internal structure evidence have been addressed within this technical document: measurement invariance, dimensionality, and reliability. The dimensionality investigation mentioned in section 11.3 also provides supporting evidence of the internal test structure.

11.4.5 Evidence Based on Relations to Other Variables

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 being 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 MI-Access mathematics test is designed to measure a single overall construct—mathematics achievement. Therefore, the items composing the MI-Access mathematics test should only measure mathematics—not ELA or social studies.

For MI-Access assessments, this technical report summarizes additional statistics that contribute to item fit and construct validity and reliability, as reported previously in this chapter and in Chapter 7. The internal consistency coefficient (Cronbach's coefficient alpha) reported above is a measure of item homogeneity. 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 FI test items and to report FI 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.

11.4.6 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 being 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 MI-Access, pairwise correlations were computed for FI students' scale scores and P and SI students' raw scores across assessments in multiple subjects. These correlation results are shown in Tables 11-22 through 11-24.

As an example, the correlation coefficients ranged from 0.53 (between ELA and mathematics in FI grade 7) to 0.63 (between ELA and mathematics in FI grades 3 and 4). The correlation coefficients suggest that individual student scores for FI tests are moderately to 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 as highly related to one another, suggesting they may be tapping into a similar knowledge base or general underlying ability. Similar pictures were also seen between subject areas for Supported Independence and Participation tests.

Table 11-22. Inter-Subject Correlation for FI Tests—Correlation between ELA, Mathematics, Science, and Social Studies

Grade	N Count	ELA/Math	ELA/Science	ELA/ Social Studies	Math/Science	Math/ Social Studies	Science/ Social Studies
3	973	0.63	*	*	*	*	*
4	988	0.63	0.66	*	0.70	*	*
5	1145	0.61	*	0.59	*	0.69	*
6	1407	0.57	*	*	*	*	*
7	1317	0.53	0.58	*	0.66	*	*
8	1354	0.56	*	0.56	*	0.63	*
11	1281	0.62	0.63	0.62	0.72	0.70	0.74

Notes: Not all grades have all the content areas. For example, for grades 3 and 6, only ELA and mathematics were administered; for grades 4 and 7, only ELA, mathematics, and science were administered. For grades 5 and 8, only ELA, mathematics and social science were administered.
*Data were not available because no such subject test(s) were administered to those grade students.

Table 11-23. Inter-Subject Correlation for SI Tests—Correlation between ELA, Mathematics, and Science

Grade	N Count	ELA/Math	ELA/Science	Math/Science
3	479	0.70	*	*
4	436	0.76	0.78	0.77
5	388	0.77	*	*
6	465	0.76	*	*
7	452	0.67	0.67	0.71
8	477	0.71	*	*
11	466	0.75	0.73	0.74

Notes: Not all grades have all the content areas. For example, for grades 3 and 6, only ELA and mathematics were administered; for grades 4 and 7, only ELA, mathematics, and science were administered. For P and SI, the social studies assessment was locally administered, and therefore, no statewide social studies tests were administered to P/SI students.
*Data were not available because no such subject test(s) were administered to those grade students.

Table 11-24. Inter-Subject Correlation for P Tests—Correlation between ELA, Mathematics, and Science*

Grade	N Count	ELA/Math	ELA/Science	Math/Science
3	485	0.81	*	*
4	400	0.79	0.81	0.84
5	395	0.75	*	*
6	360	0.78	*	*
7	321	0.78	0.81	0.83
8	333	0.79	*	*
11	373	0.82	0.86	0.83

Notes: Not all grades have all the content areas. For example, for grades 3 and 6, only ELA and mathematics were administered; for grades 4 and 7, only ELA, mathematics, and science were administered. For P and SI, the social studies assessment was locally administered, and therefore, no statewide social studies tests were administered to P/SI students.
*Data were not available because no such subject test(s) were administered to those grade students.

11.4.7 Evidence Based on Consequences of Test Use

The *Standards* incorporates the intended and unintended consequences of test use into the concept of validity. It indicates that information about the consequences of testing does not in and of itself detract from the validity of intended test interpretations (AERA, APA, & NCME, 2014). Rather, according to the *Standards*, a more searching inquiry into the sources of those consequences, given the intended purposes of an assessment, is a basis for evaluating the quality of the validity evidence. The test data alone do not provide sufficient verification of this type of evidence. For this reason, it is not straightforward to measure and collect evidence on the consequential aspects of validity.

To address the intended consequences of MI-Access, the purposes of MI-Access must be specified. MDE has carefully articulated the intended purposes of MI-Access as driving features of the selection of items, the development of tests in each content area, and the implementation of the testing program. The specific purposes associated with MI-Access include the following:

- MI-Access accurately describes both student achievement (how much students know at the end of the year) and student growth (how much students have improved since the previous year) relative to alternate content expectations, to inform program evaluation and school-, district-, and state-accountability systems and to provide valid, reliable, and fair measures of students' progress toward, and attainment of, the knowledge and skills required to be college and career ready.
- MI-Access informs state and federal accountability.
- MI-Access assessments are fair for all students in the intended population, including those with disabilities or limited English proficiency, at all levels of achievement.

11.5 Summary

In summary, Chapter 11 of this report demonstrates the adherence to the AERA, APA, & NCME (2014) *Standards* regarding reliability and construct-related validity. The analyses described above address multiple best practices of the testing industry, and in particular are related to the following *Standards for Educational and Psychological Testing* (2014):

- Standard 2.0—Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use.
- Standard 2.1—The range of replication over which reliability/precision is being evaluated should be clearly stated, along with a rationale for the choice of this designation, given the testing situation.
- 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.

- Standard 4.3—Test developers should document the rationale and supporting evidence for the administration, scoring, and reporting rules used in computer-adaptive, multistage-adaptive, or other tests delivered using computer algorithms to select items. This documentation should include procedures used in selecting items or sets of items for administration, in determining the starting point and termination conditions for the test, in scoring the test, and in controlling item exposure.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (2nd ed.). American Educational Research Association.

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling*, 22(1), 87–101.

Brennan, R. L. (2004). *Manual for BB-CLASS: A computer program that uses the beta-binomial model for classification consistency and accuracy (Version 1.1)*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Camilli, G. (2006). *Test fairness*. In R. L. Brennan (Ed.), Educational Measurement (4th ed., pp. 221–256). Westport, CT: American Council on Education and Praeger.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: SAGE Publications.

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

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.

Cronbach, L. J. (1971). *Test validation*. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, DC: American Council on Education.

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

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.

Haertel, E. H. (2006). *Reliability*. In R. L. Brennan (Ed.), Educational Measurement (4th ed., pp. 65–110). Westport, CT: American Council on Education and Praeger.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27(4)*, 345–359.

Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth.

Kingston, N. M. & Tiemann, G. C. (2012). Setting Performance *Standards* on Complex Assessments: The Body of Work Method. In G. J. Cizek (Ed.) *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd Ed.). New York, NY: Routledge.

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]. Council of Chief State School Officers 1996 National Conference on Large-Scale Assessment, Phoenix, AZ.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), S*etting performance standards: Foundations, methods, and innovations* (pp. 225–254). Routledge.

Linacre, J. M. (2015). WINSTEPS Rasch measurement (Version 3.92.1) [Computer software]. Winsteps.com.

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. (1965). A strong true score theory, with applications. *Psychometrika, 30*, 239–270.

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

Mantel, N. (1963) Chi-Square Tests with One Degree of Freedom: Extensions of the Mantel-Haenszel Procedure. J*ournal of the American Statistical Association, 58*, 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–748.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale, NJ: Erlbaum.

Muthén, B. O., & Muthén, L. K. (2012). Mplus user's guide: Statistical analysis with latent variables. Los Angeles, CA: Muthén & Muthén.

Penfield, R. D. (2007). An approach for categorizing DIF in polytomous items. *Applied Measurement in Education, 20*, 335–355.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

University of Kansas Center for Research (2013a). Dynamic learning maps Essential Elements for English Language Arts. Retrieved December 8th, 2020 from: https://dynamiclearningmaps.org/sites/default/files/documents/ELA_EEs/DLM_Essential_Elements_ELA_(2013)_v4.pdf

University of Kansas Center for Research (2013b). Dynamic learning maps Essential Elements for Mathematics. Retrieved October 31st, 2019 from: https://dynamiclearningmaps.org/sites/default/files/documents/Math_EEs/DLMEssential_Elements_Math_%282013%29_v4.pdf

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

Yen, W. M., & Fitzpatrick, A. R. (2006). *Item response theory*. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 111–153). Westport, CT: Praeger

Young, M. J., & Yoon, B. (1998, April). Estimating the consistency and accuracy of classifications in a standards-referenced assessment. (CSE Technical Report 475). Center for the Study of Evaluation, *Standards*, and Student Testing. Los Angeles, CA: University of California, Los Angeles.

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

Appendix A: Mean-Mean Equating Procedure for MI-Access Functional Independence ELA and Math Tests

1. Calculate the mean of the previous year item parameters (*b*'s) in the item bank.
2. Calculate the mean of the current year item parameters (*b*'s) from the current year calibration run.
3. Subtract the mean of the current year *b* parameters from the mean of the previous year *b* parameters (i.e., item bank values)—this is the constant that will be adjusted or applied to current year *b* parameters.
4. For each of the current year *b* parameters, add the constant from above (3.) to get the adjusted value.
5. Calculate the mean of the adjusted values from (4.).
6. Calculate the difference value for each item (previous year or item bank *b* parameter value - current year adjusted value).
7. Calculate the mean of the difference values—this value should be equal to the mean of the bank anchor parameter values.
8. Check the difference values. If any item difference is greater or equal to the absolute value of 0.5, then remove it from anchor/equating item list and repeat the whole process until all item difference values are less than the absolute value of 0.5.

Appendix B: 2019 MI-ACCESS Phase I Verification Activities

MI-ACCESS Conversion Tables Verification	Start	End	Dependencies	Verification	Conclusion	Status
1. FI Math	**6/3**	**6/19**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Initial OP data calibration • Anchor evaluation • Mean-mean equating • RSS conversion table creation			o OP data and layout files (5/29). o OP item calibration and mean-mean equating item check (6/3). o Final item parameters and RSS tables (6/18).	o Reached the same conclusion on the equating items and equating constants (6/18). o Provided the verification results for OP item parameters and RSS tables (6/19).	• The RSS tables and the equated OP item parameters exactly matched.	
FTP link to Deliverables: /michigan/2019 MI-ACCESS/CRESST Verification/			1) CRESST_Verification_FI_Math_Item_Parameters.xlsx, 2) CRESST_Verification_FI_Math_Raw_to_Scale_Score.xlsx			
2. FI ELA	**7/2**	**7/15**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Initial OP data calibration • Anchor evaluation • Mean-mean equating • RSS conversion table creation			o OP data and layout files (7/1). o OP item calibration and mean-mean equating item check (7/2). o Final item parameters and RSS tables (7/10).	o Reached the same conclusion on the equating item sets and equating constants (7/10). o Provided verification results for OP item parameters and RSS tables (7/15).	• The RSS tables and the equated OP item parameters exactly matched.	
Same FTP link as above			1) CRESST_Verification_FI_ELA_Item_Parameters.xlsx, 2) CRESST_Verification_FI_ELA_OPitem_StepMeasure.xlsx, 3) CRESST_Verification_FI_ELA_Raw_to_Scale_Score.xlsx			
3. FI Science & Social Studies	**6/21**	**7/15**	**MI provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Step 0. Anchor evaluation via fixed parameter calibration • Step 1. Equating online OP items • Step 2. Online FT item calibration • Step 3. Paper-pencil item calibration • Step 4. RSS conversion table creation			o FI SS/SC data (6/20). o Results for Step 0 (anchor evaluation) to Step 3 (PP item calibration) (6/25, 6/27, 7/1, 7/2). o Updated results for Steps 1-2 (7/2), FI SC/SS RSS table (7/9). o Updated Step 3 results and RSS tables (7/11).	o Processed data and discussed analysis procedures (6/21 - 6/25). o Reached the same conclusion on equating item sets and equating constants (6/27). o Provided results for Steps 1 and 2 (7/1, 7/2). o Updated results for Steps 1 and 2 (7/2). o Provided Step 3 verification results (7/3). o Provided the updated Step 3 results and RSS table verification (7/15).	• Differences in WINSTEPS settings (CODES and decimal places) were resolved and the affected results were updated. • Anchor evaluation rule: |displacement|> 0.5 • Item parameters from Steps 1-3 and the RSS tables exactly matched.	
Same FTP link as above			CRESST Verification FI SC SS - Step1 Item Parameters.xlsx, 2) CRESST Verification FI SC SS - Step2 Item Parameters.xlsx, CRESST Verification FI SC SS - Step3 Item Parameters.xlsx, 4) CRESST Verification MI-Access 2019 FI Score Table.xlsx			
4. P/SI ELA, Math, & Science	**7/14**	**7/15**	**MI provided**	**CRESST**	**(Additional technical info, notes)**	Green
• RPL conversion table creation			o RPL table (7/14).	o Provided RPL table verification (7/15).	• The RPL table matched.	
Same FTP link as above			1) CRESST Verification 2019 P SI Score Table.xls			

Appendix C: 2019 MI-ACCESS Phase II Verification Activities

1. FI Math	**7/30**	**8/8**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Final data calibration 1) OL data (one-step approach) 2) PP data (two-step approach) • Classical statistics analysis • DIF analysis			o FI variable list (7/18), o Data and layout file (7/26). o Final item calibration, anchor item check, and related WINSTEPS output files (7/30). o Document on PP data calibration method adjustment (7/30). o Item model fit/misfit files (7/31). o Classical statistics analysis and DIF analysis results (8/1, 8/2). o Updated PP Grade 4 calibration (8/6). o Updated PP classical item statistics (8/7).	o Provided verification results for the final item parameters, item model fit measures, and DIF analyses (8/7). o Provided verification results for classical statistics analyses (8/8).	• An error in the fixed-parameter list was corrected, and Grade 4 calibration was updated. • Errors in the data were fixed, and PP classical item statistics were updated. • All results from Phase II exactly matched, except for a 0.0001 difference in item infit for one OL Grade 5 item.	
FTP link to Deliverables: michigan/2019 MI-ACCESS/ 2019 Phase 2 Files/CRESST Verification Results/ FI Math/"CRESST Verification 2019"			1) FI Math All Grades [Mode] Final Item Files.xlsx, 2) FI Math All Grades [Mode] Final Fit Files.xlsx, 3) FI Math All Grades [Mode] DIF Results.xlsx, 4) FI Math All Grades [Mode] Overall Classical Stats.xlsx, 5) FI Math G3 [Mode] SubGroup Classical Stats.xlsx -- FI Math G11 [Mode] SubGroup Classical Stats.xlsx			
2. FI ELA	**8/5**	**8/12**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Final data calibration 1) OL data (one-step approach) 2) PP data (two-step approach) • Classical statistics analysis • DIF analysis			o Data and layout files (7/27). o Final item calibration, anchor item check, and related WINSTEPS output files (8/5). o Item model fit (8/6). o Updated PP Grade 3 calibration (8/7). o AP classical statistics analysis and DIF analysis results (8/8). o EI classical statistics analysis results (8/9). o Updated AP OL and EI DIF results (8/12).	o Provided verification results for the final item parameters, item model fit measures, as well as, DIF and classical statistics analyses (8/12).	• Errors in DIF results for AP OL Grade 3, and EI Grades 4-5 were fixed. • All results from Phase II exactly matched, except for a few OL Grade 5 and Grade 8 items. They showed 0.001 difference in item parameters.	
FTP link to Deliverables: michigan/2019 MI-ACCESS/ 2019 Phase 2 Files/CRESST Verification Results/ FI ELA/"CRESST Verification 2019"			1) FI ELA All Grades [Mode] Final Item Files.xlsx, 2) FI ELA All Grades [Mode] Final Fit Files.xlsx, 3) FI ELA All Grades Step Measures.xlsx, 4) 4FI AP All Grades [Mode] DIF Results.xlsx, 5) FI EI All Grades DIF Results.xlsx, 6) FI AP All Grades [Mode] Overall Classical Stats.xlsx, 7) FI EI All Grades DIF Results.xlsx, 8) FI AP G3 [Mode] SubGroup Classical Stats.xlsx -- FI AP G11 [Mode] SubGroup Classical Stats.xlsx, 9) FI EI G3 SubGroup Classical Stats.xlsx -- FI EI G11 SubGroup Classical Stats.xlsx			

Appendix C. (continued)

3. FI Science & Social Studies	**7/21**	**8/2**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Classical statistics analysis • DIF analysis			o Classical item statistics and DIF analysis results (7/21). o Updated the item bank statistics results. (7/23, 7/31, 8/1).	o Processed data for item bank statistics and clarified analysis procedures (7/22 - 7/29). o Provided verification results for classical item statistics and DIF analyses (8/2).	• Fixed errors within the item bank statistics results file. • Item bank statistics were not computed for subgroups with *n* <30. • All results from Phase II exactly matched.	
FTP link to Deliverables: michigan/2019 MI-ACCESS/ 2019 Phase 2 Files/CRESST Verification Results			CRESST Verification 2019 FI SC SS Item Bank Stat.xlsx			
4. P/SI ELA, Math, & Science	**7/24**	**8/8**	**MDE provided**	**CRESST**	**(Additional technical info, notes)**	Green
• Classical statistics analysis			o Classical item statistics results (7/24). o Updated the item bank statistics results (8/1).	o Processed data for item bank statistics (7/25 - 8/1). o Provided verification results for classical statistics analyses (8/8).	• Fixed errors within the item bank statistics results file. • Subgroups with *n* <30 were excluded from the analysis. • All results from Phase II exactly matched.	
Same FTP link as above			*CRESST Verification 2019 P SI Item Bank Stat.xlsx*			