

Annual Technical Report

Indiana's Alternate Measure **(*I AM*)**

2024–2025

ACKNOWLEDGMENTS

This technical report was produced on behalf of the Indiana Department of Education (IDOE). Requests for additional information concerning this technical report or the associated appendices should be directed to IDOE at INassessments@doe.in.gov.

Major contributors to this technical report include the following staff from Cambium Assessment, Inc. (CAI): Stephan Ahadi, Elizabeth Xiaoxin Wei, Guanlan Xu, Maryam Pezeshki, Kevin Clayton, Christina Sneed, Jessica Singh, Katherine Mullahy, Suzanne Huston, and Gabriel Martinez. Major contributors from IDOE include the assessment director, assistant assessment director, and program leads.

TABLE OF CONTENTS

Executive Summary	x
1. Introduction and Background	1
1.1 Purposes of the Assessment.....	1
1.2 Background of the Assessments	1
1.2.1 DEVELOPMENT OF INDIANA ALTERNATE ACADEMIC STANDARDS	2
1.2.2 I AM ITEM POOL CONSTRUCTION.....	2
1.3 Overview of the Report.....	2
2. Validity of Test Score Interpretations.....	4
2.1 Validity Evidence	4
2.1.1 CONTENT STANDARDS	6
2.2 Evidence Based on Test Content.....	10
2.2.1 REVIEW PROCESS FOR ITEMS APPEARING IN I AM OPERATIONAL TEST ADMINISTRATION	10
2.2.2 INDEPENDENT ALIGNMENT STUDY.....	12
2.3 Evidence for Interpretation of Performance Standards.....	13
2.4 Evidence Based on Internal Structure	15
2.4.1 CORRELATION AMONG REPORTING CATEGORY SCORES	16
2.4.2 LOCAL INDEPENDENCE.....	20
2.4.3 CONVERGENT AND DISCRIMINANT VALIDITY	22
2.5 Fairness and Accessibility	26
2.5.1 FAIRNESS IN CONTENT	26
2.5.2 STATISTICAL FAIRNESS IN ITEM STATISTICS	26
2.6 Summary of Validity of Test Score Interpretations.....	27
3. Summary of the Summative Test Administration	28
3.1 Student Population and Participation.....	28
3.2 Summary of Operational Procedures	35
3.2.1 ADMINISTRATION PROCEDURES.....	35
3.2.2 DESIGNATED FEATURES AND ACCOMMODATIONS	35
3.3 Summary of Overall Student Performance	45
3.4 Student Performance by Subgroup	47
3.5 Reliability	52
3.5.1 MARGINAL RELIABILITY	52
3.5.2 STANDARD ERROR OF MEASUREMENT	54
3.5.3 STUDENT CLASSIFICATION RELIABILITY	58
3.5.4 CLASSIFICATION ACCURACY	60
3.5.5 CLASSIFICATION CONSISTENCY.....	61
3.5.6 CLASSIFICATION ACCURACY AND CONSISTENCY ESTIMATES.....	63
3.5.7 RELIABILITY FOR SUBGROUPS IN THE POPULATION	66
3.5.8 REPORTING CATEGORY RELIABILITY	66
3.5.9 RELIABILITY FOR ACCOMMODATED TESTERS.....	69
4. Item Development and Test Construction	71
4.1 Test Design and Test Specifications	71
4.1.1 I AM BLUEPRINT DEVELOPMENT.....	71
4.1.2 TEST DESIGN.....	74
4.1.3 ITEM SPECIFICATIONS.....	76

4.1.4	TARGET BLUEPRINTS.....	84
4.1.5	BLUEPRINT MATCH.....	96
4.1.6	TEST FORM ASSEMBLY	100
4.2	Item Development Process	101
4.2.1	SUMMARY OF ITEM SOURCES.....	101
4.2.2	DEVELOPMENT OF NEW ITEMS	101
4.3	Item Review.....	102
4.3.1	ITEM REVIEW PROCESSES.....	102
4.3.2	COMMITTEE REVIEW OF ITEM POOL	104
4.3.3	FIELD TESTING.....	106
4.3.4	STRATEGY FOR POOL EVALUATION AND REPLENISHMENT	107
4.4	Item Statistics	107
4.4.1	CLASSICAL STATISTICS.....	107
4.4.2	ITEM RESPONSE THEORY STATISTICS.....	108
4.4.3	ANALYSIS OF DIFFERENTIAL ITEM FUNCTIONING.....	109
4.5	Item Banks.....	112
4.5.1	ESTABLISHING THE ITEM BANKS	113
4.5.2	ITEM BANK MAINTENANCE	114
5.	Test Administration.....	117
5.1	Testing Options	117
5.1.1	ADMINISTRATIVE ROLES.....	118
5.1.2	ONLINE ADMINISTRATION	119
5.1.3	ACCOMMODATED TEST ADMINISTRATION.....	122
5.1.4	ALLOWABLE RESOURCES FOR ONLINE TESTING	123
5.2	Training and Information for Test Coordinators and Administrators	126
5.2.1	MANUALS AND USER GUIDES	127
5.3	Test Security.....	130
5.3.1	STUDENT-LEVEL TESTING CONFIDENTIALITY.....	131
5.3.2	MAINTAINING TEST SECURITY	131
5.3.3	ONLINE MANAGEMENT SYSTEM	133
5.4	Tracking and Resolving Test Irregularities	135
6.	Scaling and Equating	137
6.1	Item Response Theory Procedures.....	137
6.1.1	CALIBRATION OF I AM ITEM BANKS	137
6.1.2	ESTIMATING STUDENT ABILITY USING MAXIMUM LIKELIHOOD ESTIMATION	137
6.1.3	CALIBRATING FIELD-TEST ITEMS ONTO THE I AM SCALE.....	139
6.2	I AM Reporting Scale (Scale Scores).....	141
6.2.1	OVERALL PERFORMANCE	141
6.2.2	REPORTING CATEGORY PERFORMANCE	142
6.2.3	RULES FOR ZERO AND PERFECT SCORES.....	143
6.2.4	RULES FOR SCORING AND REPORTING OF INCOMPLETE TEST ADMINISTRATIONS	143
7.	Performance Standards	145
7.1	Standard-Setting Procedures	145
7.2	Recommended Proficiency Standards	148

8. Reporting and Interpreting I AM Scores	153
8.1 Overview of I AM Score Reports	153
8.2 Reporting System for Students and Educators	157
8.3 Interpretation of Reported Scores	163
8.3.1 SCALE SCORE	163
8.3.2 STANDARD ERROR OF MEASUREMENT	163
8.3.3 PERFORMANCE LEVELS	164
8.3.4 AGGREGATED SCORE	164
8.3.5 PERFORMANCE CATEGORY FOR REPORTING CATEGORIES	164
8.4 Appropriate Uses for Scores and Reports	165
9. Quality Assurance Procedures	167
9.1 Quality Assurance in Item Development and Test Construction	167
9.2 Quality Assurance in Computer-Delivered Test Production	169
9.2.1 PRODUCTION OF CONTENT	169
9.2.2 WEB APPROVAL OF CONTENT DURING DEVELOPMENT	169
9.2.3 PLATFORM REVIEW	170
9.2.4 USER ACCEPTANCE TESTING AND FINAL REVIEW	170
9.2.5 FUNCTIONALITY AND CONFIGURATION	172
9.3 Quality Assurance in Data Preparation	172
9.4 Quality Assurance in Item Analyses and Equating	173
9.5 Quality Assurance in Scoring and Reporting	174
9.5.1 QUALITY ASSURANCE IN TEST SCORING	174
9.5.2 QUALITY ASSURANCE IN REPORTING	175
10. References	177

TABLES

Table 1: Number of Items for Each Reporting Category, ELA.....	6
Table 2: Number of Items for Each Reporting Category, Mathematics	7
Table 3: Number of Items for Each Reporting Category, Science	8
Table 4: Number of Items for Each Reporting Category, Social Studies	9
Table 5: Estimated Percentage of Students Meeting I AM and Benchmark Proficient Standards	14
Table 6: Correlation Matrix Among Reporting Categories, ELA	16
Table 7: Correlation Matrix Among Reporting Categories, Mathematics	17
Table 8: Correlation Matrix Among Reporting Categories, Science	19
Table 9: Correlation Matrix Among Reporting Categories, Social Studies	19
Table 10: Q3 Statistics, ELA	21
Table 11: Q3 Statistics, Mathematics	21
Table 12: Q3 Statistics, Science	21
Table 13: Q3 Statistics, Social Studies	22
Table 14: Correlation Matrix Among Reporting Categories, Grade 3	22
Table 15: Correlation Matrix Among Reporting Categories, Grade 4	23
Table 16: Correlation Matrix Among Reporting Categories, Grade 5	23
Table 17: Correlation Matrix Among Reporting Categories, Grade 6	24
Table 18: Correlation Matrix Among Reporting Categories, Grade 7	25
Table 19: Correlation Matrix Among Reporting Categories, Grade 8	25
Table 20: Correlation Matrix Among Reporting Categories, Grade 10	25
Table 21: Participation Criteria for I AM.....	28
Table 22: Number of Students Participating in I AM, ELA	30
Table 23: Number of Students Participating in I AM, Mathematics	30
Table 24: Number of Students Participating in I AM, Science	30
Table 25: Number of Students Participating in I AM, Social Studies	31
Table 26: Distribution of Demographic Characteristics of Tested Population, ELA	31
Table 27: Distribution of Demographic Characteristics of Tested Population, Mathematics	32
Table 28: Distribution of Demographic Characteristics of Tested Population, Science	34
Table 29: Distribution of Demographic Characteristics of Tested Population, Social Studies	34
Table 30: Total Students with Allowed Embedded and Non-Embedded Accommodations: ELA	37
Table 31: Total Students with Allowed Embedded and Non-Embedded Designated Features: ELA	38
Table 32: Total Students with Allowed Embedded and Non-Embedded Accommodations: Mathematics	39
Table 33: Total Students with Allowed Embedded and Non-Embedded Designated Features: Mathematics	40
Table 34: Total Students with Allowed Embedded and Non-Embedded Accommodations: Science	41
Table 35: Total Students with Allowed Embedded and Non-Embedded Designated Features: Science	42

Table 36: Total Students with Allowed Embedded and Non-Embedded Accommodations: Social Studies	43
Table 37: Total Students with Allowed Embedded and Non-Embedded Designated Features: Social Studies	44
Table 38: 2024–2025 Percentage of Students in Proficiency Levels, ELA	45
Table 39: 2024–2025 Percentage of Students in Proficiency Levels, Mathematics	46
Table 40: 2024–2025 Percentage of Students in Proficiency Levels, Science	46
Table 41: 2024–2025 Percentage of Students in Proficiency Levels, Social Studies	47
Table 42: Marginal Reliability, ELA	53
Table 43: Marginal Reliability, Mathematics	53
Table 44: Marginal Reliability, Science	54
Table 45: Marginal Reliability, Social Studies	54
Table 46: Average Standard Error of Measurement by Performance Level, ELA	56
Table 47: Average Standard Error of Measurement by Performance Level, Mathematics	57
Table 48: Average Standard Error of Measurement by Performance Level, Science ...	58
Table 49: Average Standard Error of Measurement by Performance Level, Social Studies	58
Table 50: Decision Accuracy and Consistency Indices for Performance Standards, ELA	63
Table 51: Decision Accuracy and Consistency Indices for Performance Standards, Mathematics	64
Table 52: Decision Accuracy and Consistency Indices for Performance Standards, Science	65
Table 53: Decision Accuracy and Consistency Indices for Performance Standards, Social Studies	66
Table 54: Marginal Reliability Coefficients for ELA Reporting Categories	66
Table 55: Marginal Reliability Coefficients for Mathematics Reporting Categories	67
Table 56: Marginal Reliability Coefficients for Science Reporting Categories	69
Table 57: Marginal Reliability Coefficients for Social Studies Reporting Categories	69
Table 58: Marginal Reliability Coefficients for Accommodated vs. Non-Accommodated Students	70
Table 59: Summary of How Each Step of Development Supports Claim Validity	77
Table 60: I AM Quantitative Passage Specifications	78
Table 61: I AM Qualitative Passage Specifications	79
Table 62: Sample ELA Specifications for Grade 3	82
Table 63: Blueprint Percentage of Items Assessing Each Reporting Category, ELA	85
Table 64: Blueprint Percentage of Items Assessing Each Reporting Category, Mathematics	86
Table 65: Blueprint Percentage of Items Assessing Each Reporting Category, Science	87
Table 66: Blueprint Percentage of Items Assessing Each Reporting Category, Social Studies	87
Table 67: DIF Classification Rules	112
Table 68: Operational Item Counts by Source	113
Table 69: I AM Item Types and Descriptions	114

Table 70: Number of Field-Test Items in 2024–2025, ELA.....	115
Table 71: Number of Field-Test Items in 2024–2025, Mathematics	115
Table 72: Number of Field-Test Items in 2024–2025, Science.....	116
Table 73: Number of Field-Test Items in 2024–2025, Social Studies.....	116
Table 74: Universal Tools, Designated Features, and Accommodations Available in Spring 2025	123
Table 75: User Guides and Manuals	128
Table 76: Examples of Test Irregularities and Test Security Violations.....	136
Table 77: Number of Students Used in Field-Test Calibrations.....	139
Table 78: Operational Item Parameter Five-Point Summary and Range: ELA.....	140
Table 79: Operational Item Parameter Five-Point Summary and Range: Mathematics	140
Table 80: Operational Item Parameter Five-Point Summary and Range: Science.....	141
Table 81: Operational Item Parameter Five-Point Summary and Range: Social Studies	141
Table 82: Scaling Constants on the Reporting Metric	142
Table 83: Theta and Scaled Score Limits for Extreme Ability Estimates	143
Table 84: Final Recommended Performance Standards.....	148
Table 85: Percentage of Students at Each Performance Level Based on Final Recommended Performance Standards.....	149
Table 86: Estimated Percentage of Students Meeting <i>I AM</i> and Benchmark Proficient Standards	150
Table 87: <i>I AM</i> Scale Score Ranges Based on Final Performance Standards.....	151
Table 88: Indiana Score Reports Summary	155
Table 89: Indiana List of Subgroups by Category	156

FIGURES

Figure 1: Average Scale Score by Subgroup, ELA	48
Figure 2: Average Scale Score by Subgroup, Mathematics	49
Figure 3: Average Scale Score by Subgroup, Science.....	51
Figure 4: Average Scale Score by Subgroup, Social Studies.....	51
Figure 5: Sample Test Information Function	56
Figure 6: I AM Test Design 2024–2025.....	75
Figure 7: Dashboard: District Level	157
Figure 8: Detailed Dashboard: District Level	158
Figure 9: Subject Detail Page for ELA: District View	159
Figure 10: Reporting Category Detail Page for ELA: District Level	160
Figure 11: Student Performance on Test Report: Performance by Roster	160
Figure 12: Student Performance on Test Report: Performance by Roster with Expanded Reporting Category Section.....	161
Figure 13: Student Individual Score Report for ELA.....	162

APPENDICES

Appendix 2-A: I AM Fixed-Form Content Review Checklist
Appendix 3-A: Distribution of Scale Scores and Standard Deviations
Appendix 3-B: Percentage of Students in Performance Levels for Overall and by Subgroup
Appendix 3-C: Distribution of Reporting Category Scores by Subgroup
Appendix 3-D: Standard Error of Measurement Curves by Subgroup
Appendix 3-E: Standard Error of Measurement Curves by Reporting Category
Appendix 3-F: Marginal Reliability Coefficients for Overall and by Subgroup
Appendix 4-A: Language, Accessibility, Bias, and Sensitivity Guidelines and Checklist
Appendix 4-B: English/Language Arts Blueprints
Appendix 4-C: Mathematics Blueprints
Appendix 4-D: Science Blueprints
Appendix 4-E: Social Studies Blueprints
Appendix 4-F: Item Review Checklist
Appendix 4-G: Field-Test Item Classical StatisticsU
Appendix 4-H: Field-Test Item Parameters
Appendix 4-I: Field-Test Item Differential Item Functioning

Appendix 5-A: *I AM Test Administrator's Manual Grades 3–8 and 10*

Appendix 5-B: Accessibility and Accommodations Information for Statewide Assessments

Appendix 5-C: *Online Test Delivery System (TDS) User Guide*

Appendix 5-D: *Test Information Distribution Engine (TIDE) User Guide*

Appendix 5-E: Test Administrator Certification Course

Appendix 5-F: *Indiana Assessments Policy Manual*

Appendix 5-G: Understanding *I AM* Webinar

Appendix 5-H: *I AM* Educator Brochure

Appendix 5-I: Accessibility and Accommodations Implementation and Setup Module

Appendix 5-J: First Year Training for New *I AM* TAs Webinar

Appendix 5-K: *Released Items Repository Quick Guide*

Appendix 5-L: *I AM 2023–2024 Released Items Repository Scoring Guide*

Appendix 5-M: *Centralized Reporting System (CRS) User Guide*

Appendix 5-N: *Assistive Technology Manual*

Appendix 5-O: *I AM Online Paper and Testing Script*

EXECUTIVE SUMMARY

This executive summary provides an overview of validity evidence of Indiana’s Alternate Measure (*I AM*) to support a validity argument regarding the uses of and inferences for the *I AM* assessments, as well as a summary of the *I AM* program and its Spring 2025 test administration.

Overview of Validity Evidence

The intended uses for *I AM* test scores include school accountability, feedback about student and class performance, evaluation of performance gaps between groups, and diagnosis of individual student strengths and opportunities for improvement. Evidence for the validity of test score interpretations is imperative to support claims that *I AM* test scores can fulfill their intended purposes. *I AM* scores help evaluate the effectiveness with which Indiana corporations and schools teach students the Indiana Alternate Academic Standards, or Content Connectors, and evaluate individual students’ performance by the end of each school year.

The items used in *I AM* tests are aligned to the Indiana Alternate Academic Standards, or Content Connectors. Items are identified and reviewed during test form construction. *I AM* test blueprints specify the range with which each of the content strands and standards will be covered in each test administration. *I AM* test blueprints also link Indiana’s academic standards to the *I AM* content-based test score interpretations. *I AM* items are developed to measure specific constructs and intellectual processes; therefore, evidence described in this report that test takers have engaged in relevant performance strategies to answer the items correctly supports the validity of the test scores.

I AM assessments report test scores as an overall performance measure in each subject area and provide scores for various reporting categories as indicators of strand-specific performance. Consequently, it is important to collect validity evidence on the intended measurement structure. The validity evidence regarding the selected measurement model and structures for reporting *I AM* assessments have been provided in this technical report. Based on the analysis of how well the measurement’s underlying structure matches empirical research, the results indicated that it is reasonable to report an overall score in a subject area in addition to individual scores for each reporting category.

Interpretation of *I AM* test scores depends on how they relate to performance standards, which define the extent to which students have achieved the expectations defined in Indiana’s Alternate Academic Standards (Content Connectors). *I AM* test scores are reported with respect to three proficiency levels, demarcating the degree to which Indiana students participating in *I AM* have achieved the learning expectations defined by Indiana’s academic standards. The standardized and rigorous procedures that Indiana

educators, serving as standard-setting panelists, followed to recommend performance standards in the standard-setting process after the Spring 2019 test administration provided central and strong evidence to support the validity of test score interpretations regarding performance standards.

Summary of the Assessment Program

The *I AM* assessment measures the knowledge and skills students are expected to develop and demonstrate in the context of Indiana’s Alternate Academic Standards, or Content Connectors, in ELA, Mathematics, Science, and Social Studies.

I AM assessments were created using items from several sources. To meet blueprint and test design requirements, items developed and field-tested specifically for *I AM* were combined with legacy items that align to the Indiana Content Connectors for the 2024–2025 operational *I AM* assessments. Item development efforts, both by CAI and by IDOE, support the goal of high-quality items through rigorous development processes managed and tracked by a content development platform that ensures every item flows through the correct sequence of reviews and captures every comment and change to the item. The blueprint design and test construction also follow rigorous procedures to support the validity of the claims that *I AM* assessments are designed to support.

I AM assessments, as assessment instruments, have established test administration procedures that support useful interpretations of score results, as specified in Standard 6.0 of the Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Various test administration-related evidence for the validity of assessment results are presented in this report, including testing procedures, accommodations, Test Administrator (TA) training and resources, and test security procedures implemented for *I AM*.

I AM scores are provided to corporations and schools through the Indiana Centralized Reporting System (CRS). The CRS is designed to assist stakeholders in reviewing and downloading test results and in understanding and using them appropriately. It provides information on student performance and aggregated summaries at several levels—state, corporation, school, and roster. Assessment results on student performance on the test can be used to help teachers or schools make decisions on how to support students’ learning. Aggregate score reports on the educator and school levels provide information

about the strengths and opportunities for improvement for students and can be used to improve teaching and student learning.

Finally, quality assurance procedures are enforced throughout all stages of *I AM* test development, configuration, administration, and scoring and reporting. These procedures ensure the accuracy and integrity of the test scores as well as strengthen the validity of score interpretation.

Chapter Overview

Chapter 1 begins with an introduction and background of the assessment, offering a brief but important overview of the assessment's purpose. Chapter 2 provides a review of validity evidence evaluated to date. Chapter 3 presents the results of the 2024–2025 *I AM* test administration, which provides summaries of the test-taking student population and their performance on the assessments. In addition, this chapter describes administration-specific evidence for the reliability of the *I AM* assessments, including internal consistency reliability, standard errors of measurement (SEMs), and the reliability of performance-level classifications. Chapter 4 describes the design and development of the *I AM* assessments, including Indiana's Alternate Academic Standards (Content Connectors), which define the content domain to be assessed by *I AM*; the development of test specifications, including blueprints, that ensure the breadth of the content domain is sampled adequately by the assessments; and test development procedures that ensure alignment of test forms with the blueprint specifications. Chapter 5 discusses the test administration procedures, including eligibility for participation in *I AM* assessments; testing conditions, including accessibility tools and accommodations; systems security for assessments administered online; and test security procedures for all test administrations.

Chapter 6 describes the procedures used to scale and equate the *I AM* assessments for scoring and reporting. Chapter 7 outlines the procedures used to identify and adopt performance standards for the *I AM* assessments. Chapter 8 provides a description of the score reporting system and the interpretation of test scores. Finally, Chapter 9 provides an overview of the quality assurance (QA) processes CAI uses to ensure that all test development, administration, scoring, and reporting activities are conducted with fidelity to the developed procedures.

1. INTRODUCTION AND BACKGROUND

1.1 PURPOSES OF THE ASSESSMENT

I AM is a criterion-referenced test that applies principles of evidence-centered design to yield overall and reporting category-level test scores at the student level and other levels of aggregation that reflect student achievement of Indiana's Alternate Academic Standards, or Content Connectors. *I AM* supports instruction and student learning by providing feedback to educators and parents about students' overall proficiency on Indiana's academic standards, which can be used to support instructional next steps. *I AM* also provides aggregate scores which can be used by educators to monitor the effectiveness of instructional strategies and educational programming.

I AM, as an assessment instrument, has established test administration procedures that support useful interpretations of score results, as specified in Standard 6.0 of the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

1.2 BACKGROUND OF THE ASSESSMENTS

I AM was constructed to measure student achievement in English/Language Arts (ELA), Mathematics, Science, and Social Studies relative to the Indiana Alternate Academic Standards, or Content Connectors. *I AM* was first administered to students in Spring 2019, replacing the Indiana Standards Tool for Alternate Reporting (ISTAR).

The transition from ISTAR to *I AM* occurred with the Spring 2019 administration. In Spring 2019, the *I AM* assessments began in the format of a combination of an operational field test and an embedded field test for the purpose of establishing the item bank and setting the performance standards of the new Indiana alternate assessment. The first implementation of an operational test, which was scheduled to take place in Spring 2020, was canceled due to the COVID-19 pandemic. In Spring 2021, initial administration of the current *I AM* test design, including an operational test and an embedded field test, was conducted using the *I AM* item bank established in Spring 2019.

The number of participants in *I AM* ranges from approximately 800 to more than 1,100 students per grade, and the number of participants increases from elementary school to high school. The number of students participating in *I AM* decreased in Spring 2021, which was expected due to the pandemic and Indiana's efforts to reduce the number of students classified as having significant cognitive disabilities based on the U.S. Department of Education's (USDE) 1% threshold. In the post-pandemic era, student participation gradually increased from Spring 2022 to Spring 2025. Although the number of students participating in *I AM* has varied among recent administrations, enrollment-based participation rates have been higher (i.e., 97–99%) in the post-pandemic years. Data for past administrations (i.e., 2020–2021, 2021–2022, 2024–2025) are provided for

overall and by demographic subgroups in Chapter 3, Summary of the Summative Test Administration, with the most recent administration for Spring 2025.

1.2.1 DEVELOPMENT OF INDIANA ALTERNATE ACADEMIC STANDARDS

In June 2018, the Indiana State Board of Education approved the adoption of new Content Connectors for ELA, Mathematics, Science, and Social Studies. Various stakeholders planned, designed, and facilitated the review, revision, and development of the Content Connectors. These alternate academic standards are designed to measure the knowledge and skills of students with significant cognitive disabilities. A systematic process was followed to ensure assessment content aligned appropriately to Indiana’s academic standards and was readily available to teachers, families, and students across Indiana. Alternate standards are necessary to ensure all students have access to grade-level-aligned content and to achieve educational accountability.

1.2.2 *I AM* ITEM POOL CONSTRUCTION

For *I AM* assessments to yield valid and reliable assessment scores and proficiency-level classifications, the *I AM* assessment blueprints guide the *I AM* item pool development. The *I AM* item pool consists of three source types: legacy operational items from the Indiana Standards Tool for Alternate Reporting (ISTAR), custom *I AM* items developed by CAI in 2018–2019, and custom items developed by IDOE in 2020–2022. With these new, custom items being field-tested in the spring administration of each year (excluding 2020), the operational pool size for each assessment has increased since 2019. In addition, a subset of the legacy ISTAR items was reformatted to better match *I AM* style and re-field-tested in Spring 2024 and Spring 2025.

1.3 OVERVIEW OF THE REPORT

This technical report documents the evidence that supports claims made for how *I AM* assessment scores may be interpreted. While *I AM* is designed as a school accountability assessment and *I AM* results inform the state’s calculations for school accountability, the primary and foremost purpose of this report is to reflect and support validity expectations of *I AM* data and reporting. Therefore, after Chapter 1 provides an overview of the purpose and intended uses of the assessment, Chapter 2 provides a review of validity evidence evaluated to date to support the intended uses and interpretations of the assessment. Because evidence for the validity of test score interpretations will accrue over time, this chapter will be expanded as further evidence is collected.

Chapter 3 presents the results of the 2024–2025 *I AM* test administration. This chapter provides summaries of the test-taking student population and their performance on the assessments. In addition, these sections describe administration-specific evidence for the reliability of *I AM* assessments, including internal consistency reliability, standard errors of measurement (SEMs), and the reliability of performance-level classifications.

The remaining chapters are organized in chronological order and document technical details of test development, administration, scoring, and reporting activities. Chapter 4 of this technical report describes the design and development of *I AM* assessments, including Indiana’s Alternate Academic Standards, which define the content domain to be assessed by *I AM*; the development of test specifications, including blueprints, that ensure the breadth of the content domain is adequately sampled by the assessments; and test development procedures that ensure alignment of test forms with blueprint specifications. *I AM* is administered as an online, stage-adaptive assessment for ELA and Mathematics for grades 3–8 and 10, Science for grades 4, 6, and Biology, and Social Studies for grade 5. Students who are unable to participate in the online administration are administered the test in a paper-and-pencil format as an accommodation. For the 2024–2025 school year, paper-and-pencil versions of the assessments were available to students whose educational record indicated that need. Chapter 4 describes the item development process and the sequence of reviews that each item must pass through before being eligible for *I AM* test administration.

Chapter 5 discusses test administration procedures, including eligibility for participation in *I AM* assessments; testing conditions, including accessibility tools and accommodations; systems security for assessments administered online; and test security procedures for all test administrations.

Chapter 6 describes the procedures used to scale and equate *I AM* assessments for scoring and reporting. Chapter 7 outlines the procedures used to identify and adopt performance standards for the *I AM* assessments. Chapter 8 provides a description of the score reporting system and the interpretation of test scores.

Finally, Chapter 9 provides an overview of the quality assurance (QA) processes CAI uses to ensure that all test development, administration, scoring, and reporting activities are conducted with fidelity to the developed procedures.

2. VALIDITY OF TEST SCORE INTERPRETATIONS

2.1 VALIDITY EVIDENCE

The term *validity* refers to the degree to which test score interpretations are supported by evidence, and it speaks directly to the legitimate uses of test scores. Establishing the validity of test score interpretations is the most fundamental component of test design and evaluation. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) provide a framework for evaluating whether claims based on test score interpretations are supported by evidence. Within this framework, the standards describe the range of evidence that may be brought to support the validity of test score interpretations.

The first source of validity evidence is the relationship between test content and the intended test construct. For test score inferences to support a validity claim, the items should be representative of the content domain, and the content domain should be relevant to the proposed interpretation of test scores. To determine content representativeness, diverse panels of content experts conduct alignment studies in which experts review individual items and rate them based on how well they match the test specifications or cognitive skills required for a particular construct. Test scores can be used to support an intended validity claim when they contain minimal construct-irrelevant variance.

The second source of validity evidence is based on “the fit between the construct and the detailed nature of performance or response actually engaged in by examinees” (AERA, APA, & NCME, 2014). This evidence is collected by surveying test takers about their performance strategies or responses to particular items. Because items are developed to measure specific constructs and intellectual processes, evidence that test takers have engaged in relevant performance strategies to answer the items correctly supports the validity of the test scores.

The third source of validity evidence is based on the internal structure: the degree to which the relationships among test items and test components relate to the construct on which the proposed test scores are interpreted. Differential item functioning (DIF), which determines whether specific items may function differently for subgroups of test takers, is one method of analyzing the internal structure of tests. Other possible analyses to examine internal structure are dimensionality assessment, goodness-of-model-fit to data, and reliability analysis.

A fourth source of validity evidence is the relationship of the test scores to external variables. The *Standards* (AERA, APA, & NCME, 2014) divide this source of evidence into three parts: convergent and discriminant evidence, test-criterion relationships, and validity generalization. Convergent evidence supports the relationship between the test and other measures intended to assess similar constructs; conversely, discriminant evidence distinguishes the test from other measures intended to assess different

constructs. A multi-trait multi-method matrix can be used to analyze both convergent and discriminant evidence. Additionally, test-criterion relationships indicate how accurately test scores predict criterion performance. The degree of accuracy mainly depends on the purpose of the test, such as classification, diagnosis, or selection. Test-criterion evidence is also used to investigate predictions of favoring different groups. Due to construct underrepresentation or construct-irrelevant components, the relation of test scores to a relevant criterion may differ from one group to another. Furthermore, validity generalization is related to whether the evidence is situation-specific or can be generalized across different settings and times. For example, sampling errors or range restrictions may need to be considered to determine whether the conclusions of a test can be assumed for the larger population.

The fifth source of validity evidence is that the intended and unintended consequences of test use should be included in the test validation process. Determining the validity of the test should depend upon evidence directly related to the test; external factors should not influence this process. For example, if an employer administers a test to determine the hiring rates for different groups of people and the results indicate an unequal distribution of skills related to the measurement construct, that would not necessarily imply a lack of test validity. However, if the unequal distribution of scores is, in fact, due to an unintended, confounding aspect of the test, that would interfere with the test's validity. Test use should align with the test's intended purpose.

Supporting a validity argument requires multiple sources of validity evidence. This then allows for an evaluation of whether sufficient evidence has been presented to support the intended uses and interpretations of the test scores. Thus, determining test validity first requires an explicit statement regarding the intended uses of the test scores and, subsequently, evidence that the scores can be used to support these inferences.

The kinds of evidence required to support the validity of test score interpretations depend on the claims made for how test scores may be interpreted. Moreover, the standards make it explicit that validity is an attribute not of tests but rather of test score interpretations. Thus, the test itself is not assessed for validity; instead, the intended interpretation and use of test scores are evaluated.

There are several intended uses for *I AM* test scores, including school accountability, feedback about student and class performance, evaluation of performance gaps between groups, and diagnosis of individual student strengths and weaknesses. Each of these intended uses requires claims to be made about the interpretation of test scores, and the strength of those claims rests on the validity evidence supporting them. Some validity evidence will be central to all of the claims, including evidence showing that test items and administrations align with Indiana's Alternate Academic Standards. Other evidence may target more specific claims. Validity evidence should therefore be evaluated with respect to the claim that it is purported to support.

Determining whether the test measures the intended construct is central to evaluating the validity of test score interpretations. Such an evaluation in turn requires a clear definition of the measurement construct. For *I AM* assessments, the definition of the measurement construct is provided by Indiana's Alternate Academic Standards.

Because directly measuring student achievement against each benchmark in Indiana’s Alternate Academic Standards would result in an impractically long test, each test administration is designed to measure a representative sample of the content domain defined by Indiana’s Alternate Academic Standards. The test blueprints represent a policy statement about the relative importance of content strands and standards in addition to meeting important measurement goals (e.g., sufficient items to report strand performance levels reliably). Because the test blueprint determines how student achievement of Indiana’s Alternate Academic Standards is evaluated, alignment of test blueprints with the content standards is critical. The *I AM* assessment blueprints describe the content to be covered and the allocations for Reporting Categories and Content Connectors.

To assemble the Spring 2025 test forms, CAI content specialists selected operational items to represent the blueprint for each grade and subject. Content specialists and senior reviewers ensured the set of operational items selected met the quality criteria described on the *I AM* Fixed Form Construction Checklist (refer to Appendix 2-A, *I AM* Fixed-Form Content Review Checklist).

2.1.1 CONTENT STANDARDS

I AM was aligned to the ELA, Mathematics, Science, and Social Studies Content Connectors adopted in June 2018. *I AM* Content Connectors are available for review on the Content Connectors page of the Indiana Department of Education (IDOE) website. Note that the Spring 2025 *I AM* assessment was aligned to the 2018 Content Connectors. Blueprints were developed to ensure that the assessment and items were aligned to the prioritized Content Connectors that they were intended to measure.

Table 1 through Table 4 present the reporting categories by grade and test, as well as the number of items measuring each category used for the reporting category scores. For ELA (grades 6, 7, 8, 10) and Mathematics, there are items included in the overall score, but not any reporting category score. A complete description of the blueprint and test form construction process can be found in Chapter 4, Item Development and Test Construction.

Table 1: Number of Items for Each Reporting Category, ELA

Grade	Reporting Category	Number of Items
3	Key Ideas and Textual Support/Vocabulary (KITS)	8
3	Reading Foundations (RF)	9
3	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	8
3	Writing (W)	7
4	Key Ideas and Textual Support/Vocabulary (KITS)	12–13
4	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	11–12
4	Writing (W)	7–8

Grade	Reporting Category	Number of Items
5	Key Ideas and Textual Support/Vocabulary (KITS)	14
5	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	9
5	Writing (W)	9
6	Key Ideas and Textual Support/Vocabulary (KITS)	11
6	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	11
6	Writing (W)	8
7	Key Ideas and Textual Support/Vocabulary (KITS)	13–14
7	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	8–10
7	Writing (W)	7–8
8	Key Ideas and Textual Support/Vocabulary (KITS)	13–14
8	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	8–10
8	Writing (W)	7–8
10	Key Ideas and Textual Support/Vocabulary (KITS)	12
10	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	10–11
10	Writing (W)	8

Table 2: Number of Items for Each Reporting Category, Mathematics

Grade	Reporting Category	Number of Items
3	Algebraic Thinking and Data Analysis (ATDA)	7–8
3	Computation (C)	8
3	Geometry and Measurement (GM)	7
3	Number Sense (NS)	7–8
4	Algebraic Thinking and Data Analysis (ATDA)	7
4	Computation (C)	7–8
4	Geometry and Measurement (GM)	7
4	Number Sense (NS)	7
5	Algebraic Thinking (AT)	7–8
5	Computation (C)	7–8
5	Geometry and Measurement, Data Analysis, and Statistics (GMDAS)	8
5	Number Sense (NS)	8
6	Algebra and Functions (AF)	8

Grade	Reporting Category	Number of Items
6	Computation (C)	7
6	Geometry and Measurement, Data Analysis, and Statistics (GMDAS)	7
6	Number Sense (NS)	9
7	Algebra and Functions (AF)	9
7	Data Analysis, Statistics, and Probability (DASP)	7–8
7	Geometry and Measurement (GM)	7
7	Number Sense and Computation (NSC)	7–8
8	Algebra and Functions (AF)	9–10
8	Data Analysis, Statistics, and Probability (DASP)	7
8	Geometry and Measurement (GM)	7
8	Number Sense and Computation (NSC)	7–8
10	Equations and Inequalities (Linear and Systems) (EI)	7–8
10	Functions (Linear and Nonlinear) (F)	7–8
10	Geometry and Measurement (GM)	7
10	Number Sense and Data Analysis (NSDA)	8

Table 3: Number of Items for Each Reporting Category, Science

Grade	Reporting Category	Number of Items
4	Analyzing, Interpreting, and Computational Thinking (AICT)	7–8
4	Explaining Solutions, Reasoning, and Communicating (ESRC)	7–8
4	Investigating (I)	7
4	Questioning and Modeling (QM)	9–10
6	Analyzing, Interpreting, and Computational Thinking (AICT)	7–8
6	Explaining Solutions, Reasoning, and Communicating (ESRC)	7–8
6	Investigating (I)	8–10
6	Questioning and Modeling (QM)	8
Biology	Analyzing Data and Mathematical Thinking (ADMT)	13–14
Biology	Communicating Explanations and Evaluating Claims Using Evidence (CEEC)	7–8
Biology	Developing and Using Modeling to Describe Structure and Function (UM)	10–11

Table 4: Number of Items for Each Reporting Category, Social Studies

Grade	Reporting Category	Number of Items
5	Civics and Government/History (US_FOUND_CGH)	17
5	Economics (US_FOUND_ECON)	7
5	Geography (US_FOUND_GEO)	8

2.2 EVIDENCE BASED ON TEST CONTENT

Determining whether the test measures the intended construct is central to evaluating the validity of test score interpretations. Such an evaluation in turn requires a clear definition of the measurement construct. For *I AM* assessments, the tests are constructed to measure student proficiency on the Indiana Content Connectors in ELA, Mathematics, Science, and Social Studies. The test was developed using principles of evidence-centered design and adherence to the principles of universal design to ensure all students have access to the test content.

The primary purpose of *I AM* is to yield test scores at the student level and other levels of aggregation that reflect student performance relative to the Indiana Content Connectors. These scores, which are estimates of student achievement and proficiency measured by assessment, are used to explain how well students performed against such expectations for student learning as specified in the Indiana Alternate Academic Standards.

Several processes are in place to ensure *I AM* fully aligns to the Indiana Content Connectors, including a rigorous item development process, adherence to test blueprints, consideration of cognitive complexity, and standard setting based on content standards. These processes include the Indiana State Board of Education, IDOE, test developers, and educator and stakeholder committees.

Ensuring the alignment of test items to their intended content standards establishes a critical link between the expectations for student achievement articulated in Indiana's Alternate Academic Standards with the *I AM* item content. The *I AM* test blueprints, in turn, specify the range with which each of the content strands and standards will be covered in each test administration and complete the link between Indiana's Alternate Academic Standards and the *I AM* content-based test score interpretations. A complete description of the test development process, including the assessment development process and mapping *I AM* assessments to the Content Connectors, can be found in Chapter 4, Item Development and Test Construction.

2.2.1 REVIEW PROCESS FOR ITEMS APPEARING IN *I AM* OPERATIONAL TEST ADMINISTRATION

This section describes the item review procedures used to ensure item accuracy and alignment with Indiana's Alternate Academic Standards. All items developed by CAI follow a standard item review process whereby item reviews proceed initially through a series of internal CAI reviews before items are deemed eligible for review by external content experts. Most of the CAI content staff members responsible for conducting internal reviews are former classroom teachers who hold degrees in education and/or their respective content areas. Each item passes through the following five internal review steps before it is designated as eligible for review by IDOE content specialists:

1. Preliminary Review, conducted by a group of CAI content-area experts
2. Content Review 1, performed by a Level 3–4 CAI content specialist

3. Accessibility Review, performed by a former special education teacher to ensure items are as accessible as possible to students across a wide spectrum of cognitive and physical disabilities
4. Edit Review 1, in which a copy editor checks the item for correct grammar and usage
5. Senior Content Review, conducted by a Level 4–5 lead content expert

At every stage of the item review process, beginning with the preliminary review, CAI's test developers analyze each item to ensure the following:

- The item aligns with Content Connector.
- The item matches the item specifications for the skill being assessed.
- The item is based on a quality idea (i.e., it assesses something worthwhile in a reasonable way).
- The item is properly aligned to the Links for Academic Learning (LAL) Depth of Knowledge (DOK) level.
- The vocabulary used in the item is appropriate for the grade and subject matter.
- The item considers language accessibility and is fair to all students.
- The content is accurate and straightforward.
- The graphic and stimulus materials are necessary to answer the question.
- The stimulus is clear, concise, and succinct (i.e., it contains enough information to make clear what is being asked, is stated positively, and does not rely on negatives—such as no, not, none, never—unless absolutely necessary).

Based on their reviews of each item, test developers may accept the item and classification as written, revise the item, or reject the item outright.

Items passing through the internal review process are sent to IDOE for review. At this stage, items may be further revised in accordance with any edits or changes requested by IDOE or rejected outright. Items at the IDOE review level pass through three external reviews in which committees of Indiana educators and stakeholders assess each item's accuracy, alignment to the intended standard, and DOK level, as well as item fairness and language sensitivity. All items considered for inclusion in the *I AM* item pools are initially reviewed as follows:

- IDOE State (client) reviews to ensure that items are eligible for Content and Fairness Committee Review. At this stage, IDOE can request edits to wording, scoring, or alignment or Depth of Knowledge (DOK) updates. A CAI director reviews all IDOE-requested edits in light of the item specifications to determine how requested edits will be applied.

- Indiana Content and Fairness Committee (CFC) Review ensures that each item is reviewed for content validity, grade-level appropriateness, alignment to the Content Connectors, and accessibility and fairness. All custom- and educator-authored Indiana development was taken to CFC Review, which combines the functions of CAI's Content Advisory Committee and the Language Accessibility, Bias, and Sensitivity (LABS) Committee.
- After IDOE- and IDOE committee-recommended edits have been applied, experts implement accessibility markups (e.g., text-to-speech). Accessibility markup is embedded into each item as part of the item development process rather than as a post-hoc process applied to completed test forms.

Items successfully passing through these committee review processes are then field-tested to ensure that they behave as intended when administered to students. Despite conscientious item development, some items perform differently than expected when administered to students. Using the item statistics gathered in field testing to review item performance is an important step in constructing valid and equivalent operational test forms.

Classical item analyses ensure that items function as intended with respect to the underlying scales. Classical item statistics are designed not only to evaluate item difficulty and the relationship of each item to the overall scale (item discrimination) but also to identify items that may exhibit a bias across subgroups (differential item functioning [DIF] analyses).

Items flagged for review based on their statistical performance must pass a three-stage review to be included in the final item pool from which operational forms are created. In the first stage of this review, a team of psychometricians reviews all flagged items to ensure that the data are accurate and properly analyzed, response keys are correct, and that there are no other obvious problems with the items.

IDOE then convenes the data review committee to evaluate flagged field-test items in the context of each item's statistical performance. Based on their review of each item's performance, IDOE decides if a flagged item is rejected or deemed eligible for inclusion in operational test administrations.

2.2.2 INDEPENDENT ALIGNMENT STUDY

An independent alignment study was conducted November 6–8 in 2019 by a third-party vendor, edCount. The study documented the following findings:

- The blueprints for all four content-area assessments met expectations for Domain Concurrence and Balance of Representation.
- The Performance-Level Descriptors (PLDs) for all four content area assessments met expectations for Domain Concurrence and Differentiation.

- All test forms were well-aligned to the Content Connectors in terms of both content and performance expectations, with the exception of the Biology assessment, which met the criteria for “somewhat aligned” in the area of performance centrality.

2.3 EVIDENCE FOR INTERPRETATION OF PERFORMANCE STANDARDS

Alignment of test content to Indiana’s Alternate Academic Standards ensures that test scores can serve as valid indicators of the degree to which students have achieved the detailed learning expectations. However, the interpretation of *I AM* test scores rests fundamentally on how test scores relate to performance standards, which define the extent to which students have achieved the expectations defined in Indiana’s Alternate Academic Standards. For *I AM*, scale scores are mapped onto three performance levels (Level 1—Below Proficiency, Level 2—Approaching Proficiency, and Level 3—At Proficiency), demarcating the degree to which *I AM* students have achieved the learning expectations defined by Indiana’s Alternate Academic Standards. The cut score establishing the At Proficiency level of performance is the most critical since it indicates that students are meeting grade-level expectations for the knowledge and skills necessary for competitive employment and post-secondary education. Procedures used to adopt performance standards for the *I AM* assessments are therefore central to the validity of test score interpretations.

Following the operational administration of the *I AM* assessments in 2018–2019, a standard-setting workshop was conducted to recommend a set of performance standards to the IDOE for reporting student performance of Indiana’s Alternate Academic Standards. This section describes the standardized and rigorous procedures that Indiana educators, serving as standard-setting panelists, followed to recommend performance standards. The workshops employed the Bookmark procedure, a widely used method in which standard-setting panelists use their expert knowledge of the Indiana Academic Standards and student achievement to map the Performance-Level Descriptors (PLDs) adopted by the IDOE onto an ordered-item book (OIB) based on operational test forms administered to students in Spring 2019. Chapter 7, Performance Standards, explains the standard-setting procedures in more detail.

Panelists were also provided with contextual information to help inform their primarily content-driven cut-score recommendations. The decision to provide panelists with contextual benchmark information was discussed during a meeting with the Indiana State Board of Education (SBOE) and Indiana’s Technical Advisory Committee (TAC) (and confirmed by the policy committee). The assessments consist of ELA and Mathematics assessments in grades 3–8 and 10; Science assessments in grade 4, grade 6, and Biology; and a Social Studies assessment in grade 5.

Panelists recommending performance standards for the ELA and Mathematics grades 3–8 and 10 assessments were provided with the approximate location of relevant performance standards from the most recent (2015) administration of a multistate assessment (created by the National Center and State Collaborative [NCSC]) of students with significant intellectual disabilities. The performance standards for the alternate

assessments were also considered in relationship to the performance standards for the general education assessment for the general population (the Indiana Learning Evaluation Assessment Readiness Network [*I LEARN*]). Panelists were asked to consider the location of these benchmark locations when making their content-based cut-score recommendations. When panelists used benchmark information to locate performance standards that converged across assessment systems, the validity of test score interpretations was bolstered.

Following the recommendations of final performance standards and moderation sessions to ensure articulation of recommended cut scores across grade levels, the recommended cut scores were presented to a stakeholder panel for review and comment.

Based on the recommended cut scores, Table 5 shows the estimated percentage of students meeting the *I AM* proficient standard for each assessment in Spring 2019. Table 5 also shows the national percentages of students who meet the NCSC and *I LEARN* proficient standards. NCSC is delivered only in ELA and Mathematics. As Table 5 indicates, the performance standards recommended for *I AM* assessments are consistent with relevant NCSC and *I LEARN* benchmarks.

Table 5: Estimated Percentage of Students Meeting *I AM* and Benchmark Proficient Standards

Subject	Grade	<i>I AM</i> At Proficiency	NCSC Proficient*	<i>I LEARN</i> At Proficiency
ELA	3	45	51	46
	4	45	56	45
	5	51	58	47
	6	50	63	47
	7	50	56	49
	8	49	64	50
	10	49	70	50**
Mathematics	3	59	73	58
	4	48	53	53
	5	48	57	47
	6	47	58	46
	7	47	68	41
	8	42	61	37
	10	32	57	37**
Science	4	41		46
	6	48		47
	Biology	43		39

Subject	Grade	I AM At Proficiency	NCSC Proficient*	ILEARN At Proficiency
Social Studies	5	35		45

*NCSC Science and Social Studies were not included because NCSC did not include those subjects.

** Because *ILEARN* was not administered in grade 10, the grade 10 benchmarking activities used the data from the *ILEARN* grade 8 assessments.

2.4 EVIDENCE BASED ON INTERNAL STRUCTURE

While the blueprints ensure that the full range of the intended measurement construct is represented in each test administration, tests may also inadvertently measure attributes that are not relevant to the construct of interest. For example, when a high level of English language proficiency is necessary to access content in mathematics and science items, language proficiency may unnecessarily limit the student's ability to demonstrate achievement in those subject areas. Although such tests may measure achievement of relevant mathematics and science content standards, they may also measure construct-irrelevant variation in language proficiency, limiting the universality of test score interpretations for some student populations.

In this section, we explore the internal structure of the *I AM* assessment using the scores provided at the reporting category level. The relationship of the subscores is just one indicator of the test dimensionality. In ELA, Mathematics, Science, and Social Studies, there are three to four reporting categories that differ in some cases by grade (see Table 1 through Table 4 for reporting category information). Evidence is needed to verify that scores for each reporting category provide useful information on student performance.

It may not be reasonable to expect that the reporting category scores are completely orthogonal—this would suggest that there are no relationships among reporting category scores and would make the justification of a unidimensional item response theory (IRT) model difficult, although we could then easily justify reporting these separate scores. On the contrary, if the reporting categories were perfectly correlated, we could justify a unidimensional model, but we could not justify the reporting of separate scores.

One pathway to explore the internal structure of the test is to explore observed correlations between the subscores. However, as each reporting category is measured with a small number of items, the standard errors of the observed scores within each reporting category are typically larger than the standard error of the total test score. Disattenuating for measurement error could offer some insight into the theoretical true score correlations. Both observed and disattenuated correlations between the subscores for the test or at grade level are provided in the following sections. The theta estimates of each subscore were used for the correlations.

2.4.1 CORRELATION AMONG REPORTING CATEGORY SCORES

Table 6 through Table 9 present the correlation matrix of the reporting category scores for each subject area.

In some instances, the observed correlations were lower than one might expect. However, as previously noted, the correlations were subject to a larger standard error of measurement (SEM) at the strand level, given the limited number of items from which the scores were derived. Consequently, over-interpretation of these correlations as either high or low should be made cautiously.

The correction for attenuation indicates what the correlation would be if reporting category scores could be measured with perfect reliability. The observed correlation between two reporting category scores with measurement errors can be corrected for attenuation as

$$r_{x'y'} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$$

where $r_{x'y'}$ is the correlation between x and y corrected for attenuation, r_{xy} is the observed correlation between x and y , r_{xx} is the reliability coefficient for x , and r_{yy} is the reliability coefficient for y . When corrected for attenuation, the correlations among reporting scores are quite high, indicating that the assessments measure a common underlying construct.

The observed correlations, disattenuated correlations, and reliability of the reporting category are provided in Table 6 through Table 9 for each subject area. The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the disattenuated correlation. The diagonal line indicates the reliability of each reporting category. In ELA, the observed correlations among the reporting categories ranged from 0.28 to 0.63. For Mathematics, the correlations were between -0.06 and 0.46. In Science, the correlations among reporting categories ranged from 0.28 to 0.63. In Social Studies, the correlations among reporting categories ranged from 0.48 to 0.55. Disattenuated correlation is capped if the correlation is greater than 1. These values suggest that internal structure validity evidence is supported with attenuated correlations greater than 0.72 for ELA, 0.43 for Mathematics (except for 0.14 in grade 5), 0.72 for Science, and 0.95 for Social Studies.

Table 6: Correlation Matrix Among Reporting Categories, ELA

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Key Ideas and Textual Support/Vocabulary (Cat1)	8	0.49	0.26	0.38	0.35
	Reading Foundations (Cat2)	9	0.79	0.23	0.23	0.32
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat3)	8	0.81	0.72	0.45	0.34

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
	Writing (Cat4)	7	1.00*	1.00*	1.00*	0.19
4	Key Ideas and Textual Support/Vocabulary (Cat1)	12–13	0.63	0.50	0.61	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11–12	0.87	0.53	0.45	
	Writing (Cat3)	7–8	1.00*	0.91	0.46	
5	Key Ideas and Textual Support/Vocabulary (Cat1)	14	0.62	0.55	0.46	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	9	0.99	0.51	0.50	
	Writing (Cat3)	9	0.83	1.00*	0.50	
6	Key Ideas and Textual Support/Vocabulary (Cat1)	11	0.61	0.51	0.36	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11	0.90	0.52	0.42	
	Writing (Cat3)	8	0.72	0.91	0.41	
7	Key Ideas and Textual Support/Vocabulary (Cat1)	13–14	0.61	0.61	0.46	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	8–10	1.00*	0.54	0.44	
	Writing (Cat3)	7–8	0.88	0.89	0.45	
8	Key Ideas and Textual Support/Vocabulary (Cat1)	11–13	0.61	0.58	0.53	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	10–11	1.00*	0.49	0.51	
	Writing (Cat3)	7–8	0.97	1.00*	0.48	
10	Key Ideas and Textual Support/Vocabulary (Cat1)	12	0.64	0.63	0.62	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	10–11	1.00*	0.58	0.55	
	Writing (Cat3)	8	1.00*	1.00*	0.51	

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 7: Correlation Matrix Among Reporting Categories, Mathematics

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Algebraic Thinking and Data Analysis (Cat1)	7–8	0.37	0.44	0.29	0.39
	Computation (Cat2)	8	1.00*	0.38	0.24	0.32
	Geometry and Measurement (Cat3)	7	0.98	0.82	0.24	0.41

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
	Number Sense (Cat4)	7–8	0.98	0.78	1.00*	0.43
4	Algebraic Thinking and Data Analysis (Cat1)	7	0.27	0.37	0.26	0.34
	Computation (Cat2)	7–8	1.00*	0.40	0.36	0.46
	Geometry and Measurement (Cat3)	7	0.98	1.00*	0.25	0.28
	Number Sense (Cat4)	7	1.00*	1.00*	0.89	0.39
5	Algebraic Thinking (Cat1)	7–8	0.34	0.04	0.06	-0.06
	Computation (Cat2)	7–8	0.14	0.25	0.32	0.25
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	8	0.17	1.00*	0.42	0.31
	Number Sense (Cat4)	8	NA	1.00*	1.00*	0.18
6	Algebra and Functions (Cat1)	8	0.20	0.26	0.23	0.29
	Computation (Cat2)	7	1.00*	0.32	0.21	0.27
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	7	0.99	0.73	0.26	0.14
	Number Sense (Cat4)	9	1.00*	0.79	0.45	0.38
7	Algebra and Functions (Cat1)	9	0.24	0.30	0.22	0.17
	Data Analysis, Statistics, and Probability (Cat2)	7–8	1.00*	0.26	0.18	0.14
	Geometry and Measurement (Cat3)	7	0.74	0.55	0.39	0.30
	Number Sense and Computation (Cat4)	7–8	0.84	0.64	1.00*	0.18
8	Algebra and Functions (Cat1)	9–10	0.24	0.19	0.12	0.19
	Data Analysis, Statistics, and Probability (Cat2)	7	0.99	0.15	0.24	0.08
	Geometry and Measurement (Cat3)	7	0.59	1.00*	0.17	0.05
	Number Sense and Computation (Cat4)	7–8	1.00*	0.68	0.43	0.09
10	Equations and Inequalities (Linear and Systems) (Cat1)	7–8	0.24	0.35	0.18	0.31
	Functions (Linear and Nonlinear) (Cat2)	7–8	1.00*	0.35	0.20	0.35
	Geometry and Measurement (Cat3)	7	1.00*	1.00*	0.04	0.22
	Number Sense and Data Analysis (Cat4)	8	1.00*	1.00*	1.00*	0.35

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 8: Correlation Matrix Among Reporting Categories, Science

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
4	Analyzing, Interpreting, and Computational Thinking (Cat1)	7–8	0.27	0.36	0.35	0.41
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7–8	1.00*	0.30	0.28	0.46
	Investigating (Cat3)	7	1.00*	0.93	0.31	0.31
	Questioning and Modeling (Cat4)	9–10	1.00*	1.00*	0.72	0.60
6	Analyzing, Interpreting, and Computational Thinking (Cat1)	7–8	0.49	0.43	0.37	0.40
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7–8	0.90	0.47	0.48	0.42
	Investigating (Cat3)	8–10	0.84	1.00*	0.40	0.41
	Questioning and Modeling (Cat4)	8	0.87	0.93	1.00*	0.43
Biology	Analyzing Data and Mathematical Thinking (Cat1)	13–14	0.69	0.59	0.63	
	Communicating Explanations and Evaluating Claims Using Evidence (Cat2)	7–8	1.00*	0.47	0.55	
	Developing and Using Modeling to Describe Structure and Function (Cat3)	10–11	0.96	1.00*	0.62	

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 9: Correlation Matrix Among Reporting Categories, Social Studies

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3
5	Civics and Government/History (Cat1)	17	0.57	0.48	0.55
	Economics (Cat2)	7	0.95	0.46	0.55
	Geography (Cat3)	8	1.00*	1.00*	0.44

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

2.4.2 LOCAL INDEPENDENCE

The validity of the application of IRT depends greatly on meeting the underlying assumptions of the models. One such assumption is local independence, which means that for a given proficiency estimate, the (marginal) likelihood is maximized, assuming the probability of correct responses is the product of independent probabilities over all items (Chen & Thissen, 1997):

When local independence is not met, there are issues of multidimensionality that are unaccounted for in the modeling of the data (Bejar, 1980). In fact, Lord (1980) noted that “local independence follows automatically from unidimensionality” (as cited in Bejar [1980], p. 5). From a dimensionality perspective, there may be nuisance factors that are influencing relationships among certain items after accounting for the intended construct of interest. These nuisance factors can be influenced by a number of testing features, such as speediness, fatigue, item chaining, and item or response formats (Yen, 1993).

Yen’s Q_3 statistic (Yen, 1984) was used to measure local independence, which was derived from the correlation between the performances of two items. Simply, the Q_3 statistic is the correlation among IRT residuals and is computed using the equation

$$d_{ij} = u_{ij} - T_i(\hat{\theta}_j),$$

where u_{ij} is the item score of the j th test taker for item i , $T_i(\hat{\theta}_j)$ is the estimated true score for item i of test taker j , which is defined as

$$T_i(\hat{\theta}_j) = \sum_{l=1}^m y_{il} P_{il}(\hat{\theta}_j),$$

where y_{il} is the weight for response category l , m is the number of response categories, and $P_{il}(\hat{\theta}_j)$ is the probability of response category l to item i by test taker j with the ability estimate $\hat{\theta}_j$.

The pairwise index of local dependence Q_3 between item i and item i' is

$$Q_{3ii'} = r(d_i, d_{i'}),$$

where r refers to the Pearson product-moment correlation.

When there are n items, $n(n-1)/2$, Q_3 statistics will be produced. The Q_3 values are expected to be small. Table 10 through Table 13 present summaries of the distributions of Q_3 statistics—minimum, 5th percentile, median, 95th percentile, and maximum values from each grade and subject. The results show that about 90% of the items, between the 5th and 95th percentiles for most grades and subjects, were around or smaller than a critical value of 0.2 (Chen & Thissen, 1997), except for Mathematics grade 8, which has the value of 0.21.

Table 10: Q3 Statistics, ELA

Grade	Q3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
3	-0.59	-0.23	-0.04	0.18	0.48
4	-0.36	-0.19	-0.04	0.15	0.38
5	-0.34	-0.19	-0.05	0.16	0.35
6	-0.27	-0.17	-0.05	0.15	0.35
7	-0.31	-0.16	-0.05	0.15	0.35
8	-0.36	-0.19	-0.05	0.16	0.35
10	-0.31	-0.18	-0.04	0.11	0.36

Table 11: Q3 Statistics, Mathematics

Grade	Q3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
3	-0.41	-0.21	-0.05	0.17	0.38
4	-0.51	-0.20	-0.05	0.16	0.32
5	-0.38	-0.24	-0.06	0.20	0.55
6	-0.32	-0.21	-0.05	0.18	0.37
7	-0.54	-0.23	-0.05	0.19	0.59
8	-0.60	-0.22	-0.06	0.21	0.56
10	-0.31	-0.17	-0.04	0.13	0.38

Table 12: Q3 Statistics, Science

Grade	Q3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
4	-0.38	-0.23	-0.06	0.20	0.44
6	-0.33	-0.18	-0.04	0.16	0.32
Biology	-0.30	-0.16	-0.04	0.13	0.36

Table 13: Q3 Statistics, Social Studies

Grade	Q3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
5	-0.34	-0.19	-0.06	0.17	0.36

2.4.3 CONVERGENT AND DISCRIMINANT VALIDITY

According to Standard 1.14 of *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), it is necessary to provide evidence of convergent and discriminant validity. Convergent evidence supports the relationship between measures assessing the same construct, while discriminant evidence distinguishes the test from other measures assessing different constructs. It is a part of validity evidence demonstrating that assessment scores are related as expected with the criterion and other variables for all student groups. However, a second, independent test measuring the same constructs as ELA, Mathematics, Science, and Social Studies in Indiana, which could easily allow for a cross-test set of correlations, was not available. Therefore, the correlations between subscores within and across assessments were examined alternatively.

The *a-priori* expectation is that subscores within the same subject (e.g., ELA) will correlate more positively than subscore correlations across subjects (e.g., ELA and Mathematics). These correlations are based on a small number of items (e.g., typically around 7 to 11); as a consequence, the observed score correlations will be smaller in magnitude as a result of the very large measurement error at the subscore level. For this reason, both the observed score and the disattenuated correlations are provided.

Observed and disattenuated subscore correlations were calculated both within the content area and across subjects and grades. Table 14 through Table 20 show the observed or disattenuated subscore correlations among two or three subjects: tables of grades 3, 7, and 8 include ELA and Mathematics; tables of grades 4, 6, and 10 include ELA, Mathematics, and Science; and tables of grade 5 include ELA, Mathematics, and Social Studies. In general, the pattern is consistent with the *a-priori* expectation that subscores within an assessment correlate more highly than correlations among assessments measuring a different construct.

Table 14: Correlation Matrix Among Reporting Categories, Grade 3

Subject	Reporting Category	ELA				Mathematics			
		Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	0.49	0.26	0.38	0.35	0.34	0.26	0.20	0.34
	RF (Cat2)	0.79	0.23	0.23	0.32	0.11	0.12	0.15	0.24
	SECM (Cat3)	0.81	0.72	0.45	0.34	0.13	0.09	0.02	0.12

Subject	Reporting Category	ELA				Mathematics			
		Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
	W (Cat4)	1.00*	1.00*	1.00*	0.19	0.21	0.14	0.11	0.19
Mathematics	ATDA (Cat1)	0.79	0.37	0.32	0.78	0.37	0.44	0.29	0.39
	C (Cat2)	0.60	0.40	0.21	0.53	1.00*	0.38	0.24	0.32
	GM (Cat3)	0.59	0.65	0.06	0.51	0.98	0.82	0.24	0.41
	NS (Cat4)	0.73	0.77	0.27	0.66	0.98	0.78	1.00*	0.43

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 15: Correlation Matrix Among Reporting Categories, Grade 4

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	0.63	0.50	0.61	0.30	0.40	0.35	0.33	0.39	0.45	0.30	0.63
	SECM (Cat2)	0.87	0.53	0.45	0.17	0.30	0.30	0.15	0.32	0.33	0.34	0.41
	W (Cat3)	1.00*	0.91	0.46	0.29	0.38	0.30	0.34	0.37	0.42	0.27	0.56
Mathematics	ATDA (Cat1)	0.73	0.44	0.83	0.27	0.37	0.26	0.34	0.21	0.19	0.11	0.31
	C (Cat2)	0.80	0.65	0.89	1.00*	0.40	0.36	0.46	0.28	0.29	0.17	0.39
	GM (Cat3)	0.88	0.82	0.87	0.98	1.00*	0.25	0.28	0.21	0.31	0.16	0.34
	NS (Cat4)	0.66	0.33	0.79	1.00*	1.00*	0.89	0.39	0.24	0.21	0.06	0.37
Science	AICT (Cat1)	0.95	0.84	1.00*	0.78	0.85	0.79	0.74	0.27	0.36	0.35	0.41
	ESRC (Cat2)	1.00*	0.82	1.00*	0.68	0.82	1.00*	0.61	1.00*	0.30	0.28	0.46
	I (Cat3)	0.68	0.85	0.73	0.37	0.48	0.58	0.18	1.00*	0.93	0.31	0.31
	QM (Cat4)	1.00*	0.72	1.00*	0.77	0.79	0.87	0.76	1.00*	1.00*	0.72	0.60

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 16: Correlation Matrix Among Reporting Categories, Grade 5

Subject	Reporting Category	ELA			Mathematics				Social Studies		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	0.62	0.55	0.46	0.11	0.32	0.44	0.22	0.53	0.54	0.55

Subject	Reporting Category	ELA			Mathematics				Social Studies		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
	SECM (Cat2)	0.99	0.51	0.50	0.05	0.32	0.45	0.22	0.46	0.51	0.48
	W (Cat3)	0.83	1.00*	0.50	0.15	0.29	0.4	0.16	0.39	0.39	0.40
	AT (Cat1)	0.23	0.12	0.35	0.34	0.04	0.06	-0.06	0.2	-0.01	0.13
Mathematics	C (Cat2)	0.8	0.88	0.81	0.14	0.25	0.32	0.25	0.24	0.30	0.28
	GMDAS (Cat3)	0.87	0.98	0.87	0.17	1.00*	0.42	0.31	0.41	0.42	0.43
	NS (Cat4)	0.65	0.75	0.54	--	1.00*	1.00*	0.18	0.18	0.24	0.25
Social Studies	CGH (Cat1)	0.89	0.85	0.74	0.45	0.64	0.86	0.55	0.57	0.48	0.55
	ECON (Cat2)	1.00*	1.00*	0.81	--	0.90	0.95	0.84	0.95	0.46	0.55
	GEO (Cat3)	1.00*	1.00*	0.85	0.35	0.83	1	0.90	1.00*	1.00*	0.44

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 17: Correlation Matrix Among Reporting Categories, Grade 6

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	0.61	0.51	0.36	0.27	0.37	0.17	0.41	0.48	0.47	0.36	0.40
	SECM (Cat2)	0.90	0.52	0.42	0.24	0.33	0.22	0.30	0.42	0.42	0.32	0.44
	W (Cat3)	0.72	0.91	0.41	0.24	0.27	0.23	0.24	0.38	0.34	0.31	0.33
Mathematics	AF (Cat1)	0.77	0.73	0.83	0.20	0.26	0.23	0.29	0.24	0.29	0.26	0.25
	C (Cat2)	0.84	0.8	0.73	1.00*	0.32	0.21	0.27	0.32	0.28	0.25	0.24
	GMDAS (Cat3)	0.44	0.61	0.71	0.99	0.73	0.26	0.14	0.20	0.21	0.22	0.30
	NS (Cat4)	0.84	0.69	0.61	1.00*	0.79	0.45	0.38	0.34	0.40	0.37	0.26
Science	AICT (Cat1)	0.88	0.84	0.84	0.76	0.82	0.57	0.8	0.49	0.43	0.37	0.40
	ESRC (Cat2)	0.88	0.84	0.77	0.96	0.72	0.60	0.94	0.90	0.47	0.48	0.42
	I (Cat3)	0.73	0.71	0.77	0.91	0.70	0.68	0.95	0.84	1.00*	0.40	0.41
	QM (Cat4)	0.78	0.92	0.79	0.85	0.63	0.9	0.65	0.87	0.93	1.00*	0.43

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 18: Correlation Matrix Among Reporting Categories, Grade 7

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	0.61	0.61	0.46	0.31	0.19	0.37	0.26
	SECM (Cat2)	1.00*	0.54	0.44	0.25	0.15	0.36	0.21
	W (Cat3)	0.88	0.89	0.45	0.25	0.23	0.35	0.28
Mathematics	AF (Cat1)	0.82	0.68	0.77	0.24	0.30	0.22	0.17
	DASP (Cat2)	0.46	0.41	0.67	1.00*	0.26	0.18	0.14
	GM (Cat3)	0.76	0.78	0.83	0.74	0.55	0.39	0.30
	NSC (Cat4)	0.77	0.68	0.98	0.84	0.64	1.00*	0.18

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 19: Correlation Matrix Among Reporting Categories, Grade 8

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	0.61	0.58	0.53	0.22	0.25	0.16	0.17
	SECM (Cat2)	1.00*	0.49	0.51	0.27	0.21	0.07	0.20
	W (Cat3)	0.97	1.00*	0.48	0.17	0.21	0.02	0.14
Mathematics	AF (Cat1)	0.57	0.78	0.51	0.24	0.19	0.12	0.19
	DASP (Cat2)	0.83	0.74	0.79	0.99	0.15	0.24	0.08
	GM (Cat3)	0.5	0.25	0.08	0.59	1.00*	0.17	0.05
	NSC (Cat4)	0.73	0.97	0.7	1.00*	0.68	0.43	0.09

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

Table 20: Correlation Matrix Among Reporting Categories, Grade 10

Subject	Reporting Category	ELA			Mathematics				Science		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	0.64	0.63	0.62	0.35	0.48	0.22	0.37	0.67	0.58	0.61
	SECM (Cat2)	1.00*	0.58	0.55	0.31	0.46	0.20	0.37	0.62	0.52	0.55

Subject	Reporting Category	ELA			Mathematics				Science		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
	W (Cat3)	1.00*	1.00*	0.51	0.33	0.46	0.22	0.35	0.55	0.49	0.52
Mathematics	EI (Cat1)	0.90	0.84	0.95	0.24	0.35	0.18	0.31	0.37	0.35	0.35
	F (Cat2)	1.00*	1.00*	1.00*	1.00*	0.35	0.20	0.35	0.52	0.45	0.46
	GM (Cat3)	1.00*	1.00*	1.00*	1.00*	1.00*	0.04	0.22	0.20	0.18	0.20
	NSDA (Cat4)	0.79	0.82	0.84	1.00*	1.00*	1.00*	0.35	0.42	0.39	0.38
Science	ADMT (Cat1)	1.00*	0.98	0.94	0.90	1.00*	1.00*	0.85	0.69	0.59	0.63
	CEEC (Cat2)	1.00*	0.99	1.00*	1.00*	1.00*	1.00*	0.96	1.00*	0.47	0.55
	UM (Cat3)	0.97	0.91	0.92	0.89	0.98	1.00*	0.82	0.96	1.00*	0.62

Note: The top triangle of the matrix shows the observed correlation, and the bottom triangle shows the attenuated correlation. The diagonal line indicates the reliability of each report category. Dissattenuated values greater than 1.00 are reported as 1.00*.

2.5 FAIRNESS AND ACCESSIBILITY

2.5.1 FAIRNESS IN CONTENT

The principles of universal design of assessments provide guidelines for test design to minimize the impact of construct-irrelevant factors in assessing student achievement. Universal design removes barriers to provide access for the widest range of students possible. Seven principles of universal design are applied in the process of test development (Thompson, Johnstone, & Thurlow, 2002). They include the following:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

CAI content experts received extensive training on the principles of universal design and applied these principles in the development of all test materials. In the review process, adherence to the principles of universal design was verified by Indiana content specialists.

2.5.2 STATISTICAL FAIRNESS IN ITEM STATISTICS

Analysis of the content alone is not sufficient to determine the fairness of an assessment. Rather, it must be accompanied by statistical processes. While a variety of item statistics were reviewed during form building to evaluate the item quality, one notable statistic that

was utilized was differential item functioning (DIF). Items were classified into three categories (A, B, or C) for DIF, ranging from “no evidence of DIF” to “severe DIF.” Furthermore, items were categorized positively (i.e., +A, +B, or +C), signifying that the item favored the focal group (e.g., African American/Black, Hispanic, Female), or negatively (i.e., –A, –B, or –C), signifying that the item favored the reference group (e.g., White, Male). Items were flagged if their DIF statistics indicated the “C” category for any group. A DIF classification of “C” indicates that the item shows significant DIF and should be reviewed for potential content bias, differential validity, or other issues that may reduce item fairness. Items were reviewed by the Bias and Sensitivity Committee regardless of whether the DIF statistic favored the focal or reference group. The details surrounding this review of items for bias are further described in Chapter 4, Item Development and Test Construction.

DIF analyses were conducted for all items to detect potential item bias from a statistical perspective across major ethnic and gender groups. These DIF analyses were performed for the following groups:

- Male/Female
- White/African American
- White/Hispanic
- Autism/Other
- Moderate and Severe Intellectual Disability/Other

The purpose of these analyses is to identify items that may have favored students in one group (focal group) over students of similar ability in another group (reference group).

2.6 SUMMARY OF VALIDITY OF TEST SCORE INTERPRETATIONS

Evidence for the validity of test score interpretations is strengthened as evidence supporting test score interpretations accrues. In this sense, the process of seeking and evaluating evidence for the validity of test score interpretation is ongoing. Nevertheless, sufficient evidence exists to support the principal claims for the test scores, including that *I AM* test scores indicate the degree to which students have achieved Indiana’s Alternate Academic Standards at each grade level and that students scoring at the Proficient level of achievement are consistent with national benchmarks that indicate they are on track to demonstrate the knowledge and skills aligned to Indiana’s Alternate Academic Standards. These claims are supported by evidence of a test development process that ensures alignment of test content to Indiana’s Alternate Academic Standards and evidence that the structural model described by Indiana’s Alternate Academic Standards and implemented in the *I AM* assessments is sound.

3. SUMMARY OF THE SUMMATIVE TEST ADMINISTRATION

I AM is administered as an online, stage-adaptive assessment using multiple-choice (MC) item types. Students who are unable to participate in the online administration are administered the test in a paper-and-pencil format as an accommodation. This format is available in regular print, large print, and uncontracted braille. The paper-and-pencil format includes the same operational items as the online assessment. Students participating in the computer-based *I AM* assessment use text-to-speech (TTS) to hear the item stimulus, stem, and answer choices. Similarly, Test Administrators (TAs) use a script to read the item stimulus, stem, and answer choices to students who participate in the paper-and-pencil format or to students participating online who need a human reader.

Students participating in the computer-based *I AM* assessment can use standard online testing features in the Test Delivery System (TDS), which include a selection of font colors and sizes and the ability to zoom in, zoom out, and highlight text. Students can take *I AM* with or without accommodations. Test developers also evaluate forms by researching and testing various response options to ensure that scores obtained using alternative modes of administration will be comparable to those earned on the standard online test that adheres to the same blueprint.

The following tests were available in the 2024–2025 administration:

- English/Language Arts (ELA) grades 3–8 and 10
- Mathematics grades 3–8 and 10
- Science grades 4, 6, and Biology
- Social Studies grade 5

3.1 STUDENT POPULATION AND PARTICIPATION

Table 21 identifies criteria required for student participation in the *I AM* assessments. All students in Indiana public or accredited nonpublic schools who meet the requirements outlined in Table 21 are required to participate in their grade-level *I AM* assessments to meet state accountability measures.

Table 21: Participation Criteria for *I AM*

Participation Criteria
Review of student record indicates a disability that significantly impacts intellectual functioning and adaptive behavior. Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life.

Participation Criteria

The student requires extensive, repeated, individualized instruction and support that are not of a temporary nature.

The student uses substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, synthesize, demonstrate, and transfer skills across multiple settings.

Goals listed in the Individualized Education Program (IEP) for this student are linked to the enrolled grade-level Alternate Academic Standards (Indiana Content Connectors).

Students in grades 3–8 and 10 may participate in the ELA and Mathematics state assessments; students in grades 4 and 6 and high school may participate in the Science state assessments; and students in grade 5 may participate in the Social Studies state assessment. Tables 22–25 show the number of students tested and the number of students reported in the Spring 2025 *I AM* administration by grade and subject area. The number of students tested and reported for historical administrations (i.e., 2020–2021 through 2023–2024) is also provided to show the trend in student participation. It is important to note that participation based on enrollment is high (i.e., 97–99%) in the post-pandemic years. The decrease in the number tested and reported is due to lower enrollment.

Table 22: Number of Students Participating in I AM, ELA

		G3	G4	G5	G6	G7	G8	G10
SP25	Number Tested	954	999	1018	1047	1036	1064	1180
	Number Reported	847	925	939	979	963	1007	1114
SP24	Number Tested	901	940	949	923	938	1021	1207
	Number Reported	816	862	884	857	879	963	1139
SP23	Number Tested	810	843	828	834	887	1026	1135
	Number Reported	734	774	751	787	841	957	1066
SP22	Number Tested	770	741	761	805	938	1057	963
	Number Reported	700	680	700	742	871	983	897

Table 23: Number of Students Participating in I AM, Mathematics

		G3	G4	G5	G6	G7	G8	G10
SP25	Number Tested	948	996	1016	1040	1033	1060	1171
	Number Reported	843	916	932	968	964	1003	1101
SP24	Number Tested	896	935	946	920	939	1018	1208
	Number Reported	808	855	881	853	880	961	1141
SP23	Number Tested	804	840	823	830	879	1018	1134
	Number Reported	727	764	748	779	832	950	1066
SP22	Number Tested	767	738	758	805	933	1048	959
	Number Reported	701	676	696	744	873	977	892

Table 24: Number of Students Participating in I AM, Science

		G4	G6	Biology
SP25	Number Tested	992	1035	1211
	Number Reported	910	963	1138
SP24	Number Tested	929	911	1246
	Number Reported	849	847	1174
SP23	Number Tested	832	818	1167
	Number Reported	760	767	1094
SP22	Number Tested	730	793	945
	Number Reported	669	731	864

Table 25: Number of Students Participating in I AM, Social Studies

		G5
SP25	Number Tested	1008
	Number Reported	924
SP24	Number Tested	941
	Number Reported	870
SP23	Number Tested	815
	Number Reported	741
SP22	Number Tested	752
	Number Reported	692

Tables 26–29 present the distribution of subgroups of students in percentages. The subgroup categories reported are gender, primary disability, and race/ethnicity. The percentage of participation by subgroup seems to be consistent from 2020–2021 to 2024–2025.

Table 26: Distribution of Demographic Characteristics of Tested Population, ELA

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G3	SP25	954	31.24	68.76	48.74	50.84	16.98	82.60	18.24	15.20	55.35
	SP24	901	33.41	66.59	44.95	54.50	20.87	78.58	14.43	16.43	57.82
	SP23	810	29.75	69.14	44.57	54.32	21.36	77.53	17.65	14.57	56.54
	SP22	770	30.78	68.31	41.69	57.79	21.95	77.53	17.14	14.03	58.70
G4	SP25	999	33.33	66.67	45.45	53.75	21.12	78.08	14.81	16.52	58.16
	SP24	940	30.53	69.47	45.00	54.15	22.55	76.60	17.77	15.00	57.34
	SP23	843	31.79	67.38	40.45	58.84	23.37	75.92	16.37	14.59	57.89
	SP22	741	34.68	64.64	34.82	64.24	25.10	73.95	17.27	12.96	60.73
G5	SP25	1018	32.91	67.09	43.91	55.40	22.30	77.01	18.47	14.05	56.48
	SP24	949	33.40	66.60	40.25	58.80	22.34	76.71	16.54	14.44	59.11
	SP23	829	35.59	63.33	34.86	64.17	25.93	73.10	16.65	14.96	58.87
	SP22	761	32.19	67.02	35.35	63.60	27.46	71.48	14.45	14.19	60.58

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G6	SP25	1047	33.43	66.57	41.55	57.50	21.59	77.46	16.91	14.61	59.12
	SP24	923	34.45	65.55	35.97	63.71	25.24	74.43	17.55	14.95	57.64
	SP23	834	31.77	66.55	34.17	64.39	26.98	71.58	15.23	15.95	57.07
	SP22	805	33.29	66.09	33.54	66.34	27.58	72.30	15.16	13.04	63.73
G7	SP25	1036	34.56	65.44	35.81	63.71	22.88	76.64	17.28	13.90	59.07
	SP24	938	33.48	66.52	35.07	64.18	27.08	72.17	15.57	16.20	58.74
	SP23	889	34.31	64.45	33.63	65.69	25.53	73.79	15.75	12.60	62.32
	SP22	938	34.43	65.03	30.49	68.44	25.91	73.03	14.18	13.97	62.58
G8	SP25	1064	34.12	65.88	35.24	64.19	25.66	73.78	14.76	16.07	60.06
	SP24	1022	34.74	65.26	33.46	65.95	24.66	74.76	15.26	12.52	63.21
	SP23	1026	33.04	65.59	29.82	69.10	25.73	73.20	13.74	13.55	62.96
	SP22	1057	32.26	66.60	28.57	70.58	23.37	75.78	17.22	11.92	62.06
G10	SP25	1180	34.58	65.42	29.92	69.32	21.95	77.29	15.00	13.81	62.12
	SP24	1207	35.05	64.95	30.49	67.77	21.13	77.13	17.56	12.59	61.47
	SP23	1137	32.89	65.70	31.13	67.55	23.31	75.37	17.50	12.49	59.72
	SP22	963	34.58	64.07	28.35	70.72	24.92	74.14	16.93	13.29	61.47

Table 27: Distribution of Demographic Characteristics of Tested Population, Mathematics

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G3	SP25	948	31.22	68.78	48.95	50.63	17.09	82.49	18.14	15.19	55.49
	SP24	896	33.48	66.52	44.87	54.46	20.98	78.35	14.40	16.41	57.81
	SP23	804	29.85	69.03	44.15	54.60	21.64	77.11	17.66	14.18	56.84
	SP22	767	30.90	68.58	41.85	57.63	22.29	77.18	17.21	14.34	58.80
	SP25	996	33.33	66.67	45.38	53.82	20.98	78.21	14.76	16.67	58.03

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G4	SP24	935	30.37	69.63	45.24	53.90	22.35	76.79	17.75	14.76	57.54
	SP23	840	31.79	67.74	40.48	58.69	23.21	75.95	16.07	14.40	58.69
	SP22	738	34.55	64.91	34.82	64.36	25.20	73.98	17.34	13.14	60.43
G5	SP25	1016	32.87	67.13	43.80	55.51	22.34	76.97	18.41	14.07	56.50
	SP24	946	33.51	66.49	40.38	58.88	22.41	76.85	16.38	14.48	59.20
	SP23	824	35.80	63.47	34.83	64.20	26.09	72.94	16.87	14.44	59.34
	SP22	758	31.93	67.55	35.49	63.46	27.44	71.50	15.04	14.12	60.29
G6	SP25	1040	33.37	66.63	41.44	57.60	21.63	77.40	16.92	14.62	59.23
	SP24	920	34.46	65.54	35.87	63.91	25.33	74.46	17.39	15.00	57.72
	SP23	830	31.93	66.75	34.58	64.10	27.35	71.33	15.06	16.02	57.47
	SP22	805	33.42	65.59	33.54	66.34	27.33	72.55	15.03	12.92	63.48
G7	SP25	1033	34.56	65.44	35.72	63.79	22.94	76.57	17.33	13.84	59.05
	SP24	939	33.55	66.45	34.93	64.32	27.16	72.10	15.55	16.08	58.89
	SP23	879	34.81	64.51	34.24	65.07	25.60	73.72	15.81	12.74	62.57
	SP22	933	34.62	64.95	30.33	68.60	26.05	72.88	14.26	13.93	62.81
G8	SP25	1060	34.15	65.85	35.19	64.43	25.75	73.87	14.72	16.04	60.19
	SP24	1018	34.68	65.32	33.69	65.82	24.75	74.75	15.13	12.57	63.36
	SP23	1018	32.91	65.62	30.35	68.57	25.54	73.38	13.65	13.46	63.06
	SP22	1048	32.63	66.70	28.44	70.80	23.57	75.67	17.18	11.93	62.31
G10	SP25	1171	34.50	65.50	30.06	69.17	21.86	77.37	15.03	13.92	61.91
	SP24	1208	34.93	65.07	30.63	67.63	21.03	77.24	17.47	12.50	61.67
	SP23	1134	33.16	66.05	31.48	67.20	23.28	75.40	17.55	12.61	60.14
	SP22	959	34.83	63.92	28.57	70.49	25.13	73.93	17.10	13.66	61.31

Table 28: Distribution of Demographic Characteristics of Tested Population, Science

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G4	SP25	992	33.27	66.73	45.36	53.83	20.97	78.23	14.82	16.23	58.27
	SP24	929	30.14	69.86	45.32	53.82	22.50	76.64	17.87	14.75	57.48
	SP23	832	31.25	67.31	40.14	58.89	23.44	75.60	15.75	14.54	58.05
	SP22	730	34.38	64.25	34.66	64.66	25.34	73.97	17.53	13.01	59.59
G6	SP25	1035	33.33	66.67	41.45	57.58	21.64	77.39	17.00	14.69	59.03
	SP24	911	34.80	65.20	35.78	64.00	25.25	74.53	17.12	14.93	57.96
	SP23	818	32.15	66.38	34.11	64.67	27.51	71.27	15.16	16.01	57.09
	SP22	793	33.42	65.70	33.54	66.46	27.36	72.64	14.63	13.11	63.81
Biology	SP25	1211	36.09	63.91	30.06	68.70	22.21	76.55	15.28	14.37	61.44
	SP24	1247	35.85	64.15	29.27	69.85	21.57	77.55	17.48	13.95	60.95
	SP23	1167	32.39	66.92	30.16	68.98	24.85	74.29	17.91	12.17	60.58
	SP22	945	35.34	64.13	28.57	70.48	24.34	74.71	18.73	12.80	60.21

Table 29: Distribution of Demographic Characteristics of Tested Population, Social Studies

Grade	Year	N	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
G5	SP25	1009	32.80	67.20	43.81	55.50	22.20	77.11	18.33	14.07	56.49
	SP24	941	33.58	66.42	40.17	58.98	22.53	76.62	16.37	14.45	59.19
	SP23	816	35.54	63.24	34.56	64.34	25.86	73.04	16.67	14.71	58.70
	SP22	752	31.78	67.29	35.24	63.70	27.79	71.14	15.03	13.70	60.24

3.2 SUMMARY OF OPERATIONAL PROCEDURES

3.2.1 ADMINISTRATION PROCEDURES

The Spring 2025 *I AM* test administration window for all subjects opened on March 31, 2025, and closed on May 9, 2025. Key personnel included the Corporation Test Coordinators (CTCs), School Test Coordinators (STCs), and TAs who proctored the test. A *Test Administrator's Manual (TAM)* was provided so that personnel administering statewide assessments could maintain both standardized testing conditions and test security.

The CAI Secure Browser was required to access the *I AM* assessments. The online browser provided a secure environment for student testing by disabling the hot keys, copy, and screen capture capabilities and preventing access to the desktop (i.e., Internet, email, and other files or programs installed on networked machines). During the online assessment, students could pause a test, review previously answered questions, and modify their responses. If the test was paused for more than 10 days, the test opportunity expired. To reopen the test, the STC was required to submit a test irregularity request.

3.2.2 DESIGNATED FEATURES AND ACCOMMODATIONS

Three types of accessibility supports are discussed within this document:

1. Both embedded (digitally provided) and non-embedded (non-digitally or locally provided) universal tools that are available to all students as they access instructional or assessment content
2. Designated features that are available to students for whom the need has been identified by an informed educator or team of educators
3. Accommodations that are available to students for whom there is documentation on an Individualized Education Program (IEP) or Individual Learning Plan (ILP)

Scores achieved by students using designated features are included for federal accountability purposes. All educators making decisions on the use of these features are trained in the process and understand the range of designated features available.

Accommodations involve changes in procedures or materials that ensure equitable access to instructional and assessment content and generate valid assessment results for students who need such accommodations. Embedded accommodations (e.g., Streamline Format, Permissive Mode) are provided digitally through instructional or assessment technology and are available within the Test Delivery System (TDS). Non-embedded accommodations (e.g., Print Booklets, Bilingual Word-to-Word Dictionary) are provided by schools and are available outside of TDS.

CAI also supports embedded and non-embedded designated features on *I AM* assessments. Embedded designated features (e.g., Color Contrast, Print Size) are available within TDS, and non-embedded designated features (e.g., Human Reader) are provided by schools. Students who require third-party assistive technology must have

Permissive Mode turned on to allow the assistive technology to function in conjunction with the secure testing environment. These accommodations are generally available for students whose eligibility has been documented on an IEP or ILP. State-approved accommodations do not compromise the learning expectations, constructs, or grade-level standards. Such accommodations help students with a need that has been documented in an IEP or ILP to generate valid outcomes on the assessments, enabling them to fully demonstrate what they know and are able to do. From the psychometric perspective, the purpose of providing accommodations is to “increase the validity of inferences about students with disabilities by offsetting specific disability-related, construct-irrelevant impediments to performance” (Koretz & Hamilton, 2006, p. 562).

TAs and STCs in Indiana are responsible for ensuring that accommodations are updated before the test administration dates. The available accommodation options for eligible students include braille booklets, Interpreter for Sign Language, Streamline Format, Alternate Indication of Response (e.g., adaptive keyboards, touchscreen, switches), calculation devices, and multiplication tables.

Tables 30–37 list the number of students who are recorded in the Test Information Distribution Engine (TIDE) as receiving each accommodation during the Spring 2025 test administration.

Table 30: Total Students with Allowed Embedded and Non-Embedded Accommodations: ELA

Accommodations	Grade						
	3	4	5	6	7	8	10
Embedded Accommodations							
Permissive Mode	205	176	184	188	167	173	124
Streamline Format	37	38	24	40	53	37	33
Non-Embedded Accommodations							
Alternate Indication of Response	751	784	745	751	695	692	572
Print Booklet	184	300	345	407	396	360	289
Large Print Booklet	28	19	16	26	14	12	14
Braille Booklet	10	11	6	3	9	16	8
Read Aloud to Self	380	484	521	536	472	410	291
Bilingual Word-to-Word Dictionary	2	1	11	1	13	9	34
Interpreter for Sign Language							
Scribe	1		2		4	3	1
Sign Language Interpreter for Directions and All Items Including Items Testing Reading Comprehension	5	3	1	4	5		3
Student Provided Access to Own Resources	88	116	115	147	159	193	170

Table 31: Total Students with Allowed Embedded and Non-Embedded Designated Features: ELA

Designated Features	Grade						
	3	4	5	6	7	8	10
Embedded Designated Features							
Masking	954	999	1018	1046	1036	1064	1179
Mouse Pointer	2					1	1
Text Size	4	2		1	1	2	2
Color Contrast							
Text-to-Speech Tracking	1	2	1		1		2
Non-Embedded Designated Features							
Color Acetate Film for Paper Assessment	1			1			1
Assistive Technology to Magnify/Enlarge	18	11	27	26	23	26	27
Access to Sound Amplification System	11	14	7	12	12	16	15
Special Furniture or Equipment for Viewing Test	102	96	72	83	64	57	59
Special Lighting Conditions	38	25	25	33	18	26	26
Time of Day for Testing Altered	121	128	117	131	114	125	135
Human Reader for All Items Including Reading Comprehension	186	170	147	149	93	107	112

Table 32: Total Students with Allowed Embedded and Non-Embedded Accommodations: Mathematics

Accommodations	Grade						
	3	4	5	6	7	8	10
Embedded Accommodations							
Permissive Mode	203	175	184	187	166	170	120
Streamline Format	37	38	22	40	54	37	33
Non-Embedded Accommodations							
Alternate Indication of Response	746	783	745	747	691	690	566
Multiplication Table	184	304	345	405	393	361	288
Print Booklet	27	19	16	25	15	12	14
Large Print Booklet	9	11	6	3	9	16	8
Hundreds Chart	380	487	520	533	468	411	290
Braille Booklet	746	783	745	747	691	690	566
Read Aloud to Self	184	304	345	405	393	361	288
Bilingual Word-to-Word Dictionary	2	1	11	1	13	9	34
Interpreter for Sign Language							
Scribe	1		2		4	3	1
Sign Language Interpreter for Directions and All Items Including Items Testing Reading Comprehension	5	3	1	4	5	1	3
Student Provided Access to Own Resources	88	116	115	145	158	193	169

Table 33: Total Students with Allowed Embedded and Non-Embedded Designated Features: Mathematics

Designated Features	Grade						
	3	4	5	6	7	8	10
Embedded Designated Features							
	948	996	1016	1040	1033	1060	1170
Masking	1	1				1	1
Mouse Pointer	3	2		3	1	1	2
Text Size	2	2	1	1	1		2
Color Contrast							
Text-to-Speech Tracking	948	996	1016	1040	1033	1060	1171
Non-Embedded Designated Features							
Color Acetate Film for Paper Assessment	1			1			1
Assistive Technology to Magnify/Enlarge	17	11	27	26	23	27	27
Access to Sound Amplification System	11	14	7	12	12	16	15
Special Furniture or Equipment for Viewing Test	100	94	72	83	65	57	57
Special Lighting Conditions	37	25	25	33	18	27	25
Time of Day for Testing Altered	120	128	120	132	114	123	132
Human Reader for All Items Including Reading Comprehension	185	169	148	149	93	106	109

Table 34: Total Students with Allowed Embedded and Non-Embedded Accommodations: Science

Accommodations	Grade		
	4	6	Biolog y
Embedded Accommodations			
Permissive Mode	176	185	146
Streamline Format	38	38	41
Non-Embedded Accommodations			
Alternate Indication of Response	779	743	615
Multiplication Table	303	405	321
Print Booklet	19	25	13
Large Print Booklet	11	4	11
Hundreds Chart	485	533	332
Braille Booklet			
Read Aloud to Self			
Bilingual Word-to-Word Dictionary	1	1	33
Interpreter for Sign Language			
Scribe			
Sign Language Interpreter for Directions and All Items Including Items Testing Reading Comprehension	3	3	2
Student Provided Access to Own Resources	117	144	190

Table 35: Total Students with Allowed Embedded and Non-Embedded Designated Features: Science

Designated Features	Grade		
	4	6	Biology
Embedded Designated Features			
	992	1035	1211
Masking			2
Mouse Pointer	2	3	4
Text Size	2		2
Color Contrast			
Text-to-Speech Tracking	992	1035	1211
Non-Embedded Designated Features			
Color Acetate Film for Paper Assessment		1	2
Assistive Technology to Magnify/Enlarge	11	26	34
Access to Sound Amplification System	14	11	18
Special Furniture or Equipment for Viewing Test	94	83	64
Special Lighting Conditions	26	33	31
Time of Day for Testing Altered	127	130	128
Human Reader for All Items Including Reading Comprehension	167	146	97

Table 36: Total Students with Allowed Embedded and Non-Embedded Accommodations: Social Studies

Accommodations	Grade
	5
Embedded Accommodations	
Permissive Mode	181
Streamline Format	22
Non-Embedded Accommodations	
Alternate Indication of Response	738
Print Booklet	345
Large Print Booklet	16
Braille Booklet	6
Read Aloud to Self	519
Bilingual Word-to-Word Dictionary	2
Interpreter for Sign Language	
Scribe	11
Sign Language Interpreter for Directions and All Items Including Items Testing Reading Comprehension	1
	115
Student Provided Access to Own Resources	

Table 37: Total Students with Allowed Embedded and Non-Embedded Designated Features: Social Studies

Designated Features	Grade
	5
Embedded Designated Features	
	1009
Masking	
Mouse Pointer	
Text Size	1
Color Contrast	
Text-to-Speech Tracking	1009
Non-Embedded Designated Features	
Color Acetate Film for Paper Assessment	27
Assistive Technology to Magnify/Enlarge	7
Access to Sound Amplification System	71
Special Furniture or Equipment for Viewing Test	25
Special Lighting Conditions	120
Time of Day for Testing Altered	145
Human Reader for All Items Including Reading Comprehension	27

3.3 SUMMARY OF OVERALL STUDENT PERFORMANCE

The 2024–2025 state summary results for the average scale scores and the percentage of students in each proficiency level by grade and content area are presented in Table 38 to Table 41. In terms of both average scale scores and percentages at or above proficiency, student performances in Spring 2025 show comparable results from 2020–2021 to 2024–2025.

Table 38: 2024–2025 Percentage of Students in Proficiency Levels, ELA

Grade	Admin	Number Reported	Scale Score Mean	Scale Score SD	% Below Proficiency	% Approaching Proficiency	% At Proficiency
G3	SP25	847	1475.23	36.83	34.59	23.73	41.68
	SP24	816	1476.64	37.84	33.95	24.51	41.54
	SP23	734	1476.97	35.30	32.83	24.52	42.64
	SP22	700	1477.48	36.20	32.29	23.43	44.29
G4	SP25	925	1490.41	43.89	47.03	16.43	36.54
	SP24	862	1487.95	41.25	46.29	19.14	34.57
	SP23	774	1490.32	45.80	45.09	16.80	38.11
	SP22	680	1488.37	45.60	42.35	17.94	39.71
G5	SP25	939	1492.33	42.66	34.61	16.40	48.99
	SP24	884	1495.34	42.34	32.47	16.74	50.79
	SP23	751	1493.17	41.15	32.36	19.04	48.60
	SP22	700	1489.16	45.00	45.71	10.57	43.71
G6	SP25	979	1489.11	41.93	32.48	22.37	45.15
	SP24	857	1492.17	38.45	24.39	27.42	48.19
	SP23	787	1483.69	42.37	34.18	24.78	41.04
	SP22	742	1489.03	43.85	29.25	22.37	48.38
G7	SP25	963	1503.47	45.13	42.58	5.30	52.13
	SP24	879	1502.35	43.02	36.75	14.68	48.58
	SP23	841	1503.37	44.32	34.84	14.63	50.54
	SP22	871	1505.94	50.49	37.77	12.06	50.17
G8	SP25	1007	1493.09	44.85	24.33	31.78	43.89
	SP24	963	1495.43	42.57	23.05	30.43	46.52
	SP23	957	1498.96	49.68	22.78	28.11	49.11
	SP22	983	1497.04	48.92	26.86	27.87	45.27
G10	SP25	1114	1516.41	54.47	16.70	26.66	56.64
	SP24	1139	1515.20	50.98	16.86	28.01	55.14
	SP23	1066	1510.02	53.65	21.95	29.36	48.69
	SP22	897	1502.91	54.55	26.09	30.21	43.70

Table 39: 2024–2025 Percentage of Students in Proficiency Levels, Mathematics

Grade	Admin	Number Reported	Scale Score Mean	Scale Score SD	% Below Proficiency	% Approaching Proficiency	% At Proficiency
G3	SP25	843	2476.01	39.26	30.96	11.03	58.01
	SP24	808	2481.57	36.67	25.37	12.25	62.38
	SP23	727	2477.91	35.64	30.54	16.78	52.68
	SP22	701	2477.91	39.94	32.95	9.84	57.20
G4	SP25	916	2475.94	39.91	31.00	22.27	46.72
	SP24	855	2477.57	35.41	30.53	21.99	47.49
	SP23	764	2480.33	38.96	29.32	20.42	50.26
	SP22	676	2476.27	41.12	33.73	25.30	40.98
G5	SP25	932	2472.79	30.64	34.66	17.49	47.85
	SP24	881	2477.33	30.61	27.24	18.27	54.48
	SP23	748	2470.87	28.51	32.22	25.67	42.11
	SP22	696	2470.23	31.26	34.91	19.83	45.26
G6	SP25	968	2478.61	30.51	28.72	24.69	46.59
	SP24	853	2477.36	30.89	30.36	24.62	45.02
	SP23	779	2476.30	31.50	33.76	20.41	45.83
	SP22	744	2476.84	35.89	31.18	26.61	42.20
G7	SP25	964	2474.79	30.39	38.07	10.48	51.45
	SP24	880	2474.18	29.24	39.32	9.20	51.48
	SP23	832	2474.14	30.54	35.82	18.99	45.19
	SP22	873	2475.51	28.54	38.60	10.42	50.97
G8	SP25	1003	2469.28	26.69	39.38	15.15	45.46
	SP24	961	2471.79	27.69	38.81	14.05	47.14
	SP23	950	2468.85	28.72	43.16	12.42	44.42
	SP22	977	2470.01	27.24	39.30	14.53	46.16
G10	SP25	1101	2475.54	32.46	43.78	16.80	39.42
	SP24	1141	2473.57	31.26	46.45	17.00	36.55
	SP23	1066	2472.44	31.32	48.59	16.79	34.62
	SP22	892	2473.35	27.04	41.93	27.24	30.83

Table 40: 2024–2025 Percentage of Students in Proficiency Levels, Science

Grade	Admin	Number Reported	Scale Score Mean	Scale Score SD	% Below Proficiency	% Approaching Proficiency	% At Proficiency
G4	SP25	910	3484.01	38.77	45.71	20.44	33.85
	SP24	849	3488.06	36.15	39.46	24.85	35.69
	SP23	760	3489.47	39.17	39.74	22.37	37.89
	SP22	669	3489.07	39.72	44.54	20.93	34.53
G6	SP25	963	3482.78	41.02	32.29	27.52	40.19
	SP24	847	3481.51	44.05	35.89	25.38	38.72
	SP23	767	3475.60	41.79	40.81	23.99	35.20
	SP22	731	3485.51	36.98	28.73	23.53	47.74
Biology	SP25	1138	3500.13	51.19	34.45	21.27	44.29
	SP24	1174	3501.84	50.65	32.79	21.81	45.40

Grade	Admin	Number Reported	Scale Score Mean	Scale Score SD	% Below Proficiency	% Approaching Proficiency	% At Proficiency
	SP23	1094	3495.97	46.86	32.82	30.53	36.65
	SP22	864	3495.26	45.93	34.49	26.16	39.35

Table 41: 2024–2025 Percentage of Students in Proficiency Levels, Social Studies

Grade	Admin	Number Reported	Scale Score Mean	Scale Score SD	% Below Proficiency	% Approaching Proficiency	% At Proficiency
G5	SP25	924	4485.64	39.80	63.64	4.76	31.60
	SP24	870	4487.63	39.50	59.08	5.06	35.86
	SP23	741	4483.40	43.74	62.75	5.26	31.98
	SP22	692	4481.25	42.52	65.17	9.10	25.72

3.4 STUDENT PERFORMANCE BY SUBGROUP

The 2024–2025 state summary results for the average scale scores and the percentage of students in each proficiency level by grade and by content area were calculated for several subcategories—including female, male, autism, non-autism, moderate and severe intellectual disability, non-moderate and severe intellectual disability, African American, Hispanic/Latino, and White.

Distribution of scale scores by subgroups, along with historical statistics, are presented in Appendix 3-A, Distribution of Scale Scores and Standard Deviations. The percentage of students in performance levels for overall and by subgroup, along with historical statistics, are presented in Appendix 3-B, Percentage of Students in Performance Levels for Overall and by Subgroup. In addition, the summary of scale scores by subgroup for each reporting category, along with historical statistics, are provided in Appendix 3-C, Distribution of Reporting Category Scores by Subgroup.

Figures 1–4 display the average scale scores, overall and by subgroup, for the 2024–2025 administration as well as for historical administrations. As shown in the figures, students with autism and moderate or severe disabilities, and Hispanic and African American students, had relatively lower average scale scores across administrations.

Figure 1: Average Scale Score by Subgroup, ELA

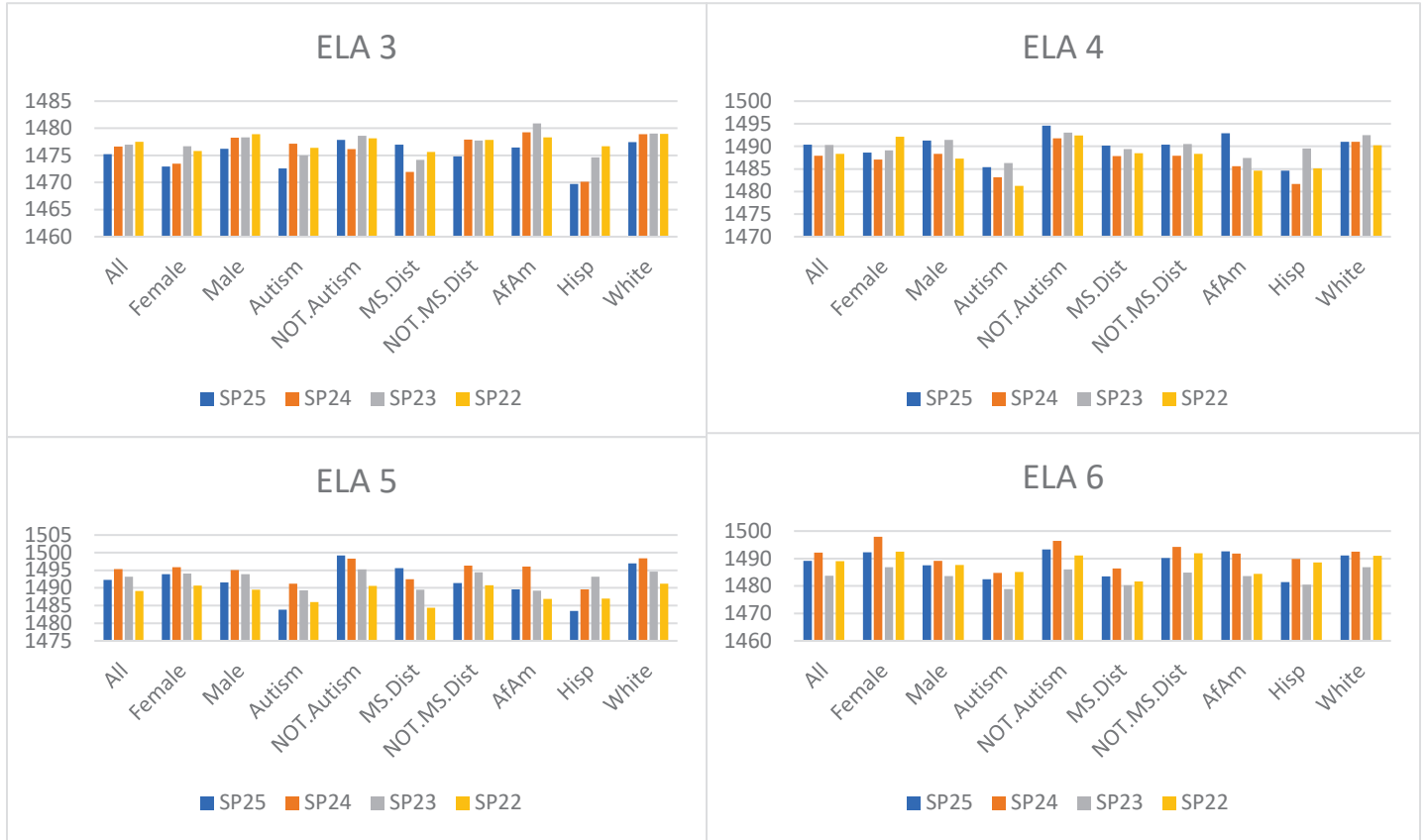
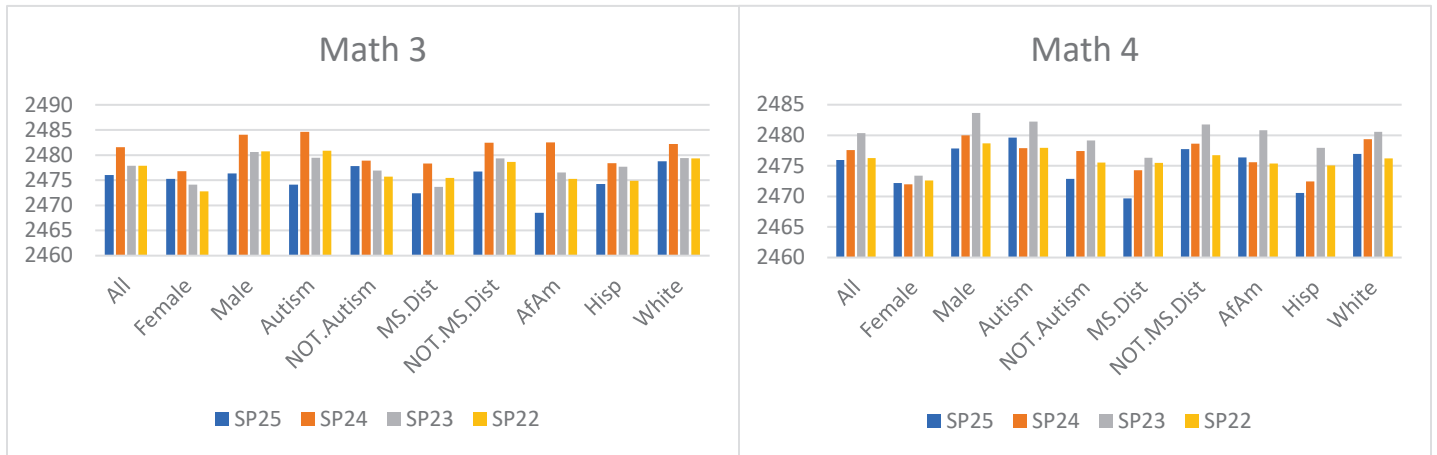
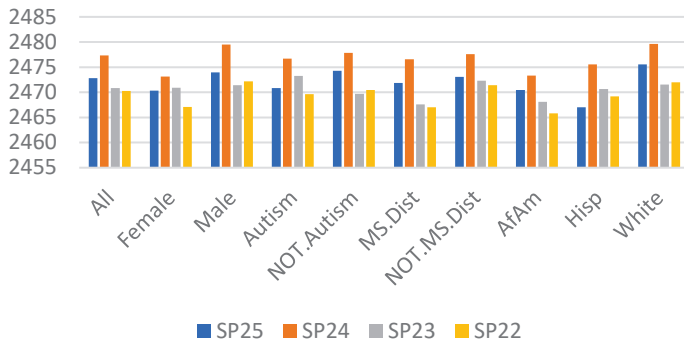




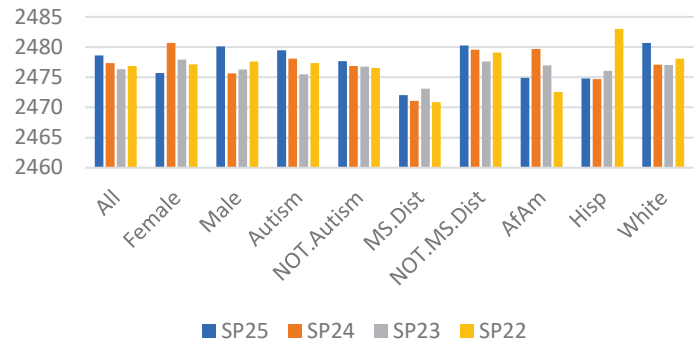
Figure 2: Average Scale Score by Subgroup, Mathematics



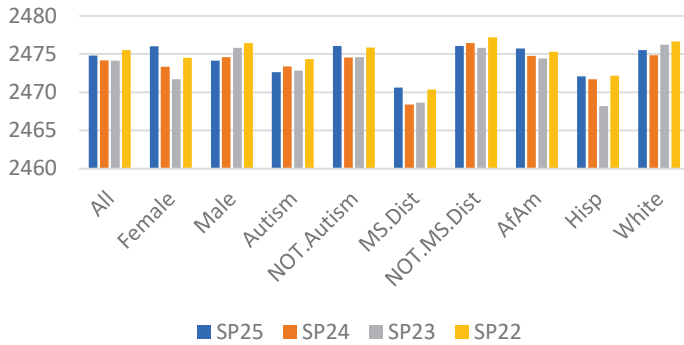
Math 5



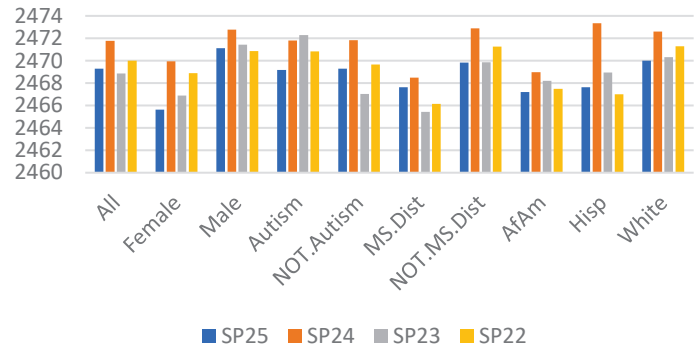
Math 6



Math 7



Math 8



Math 10

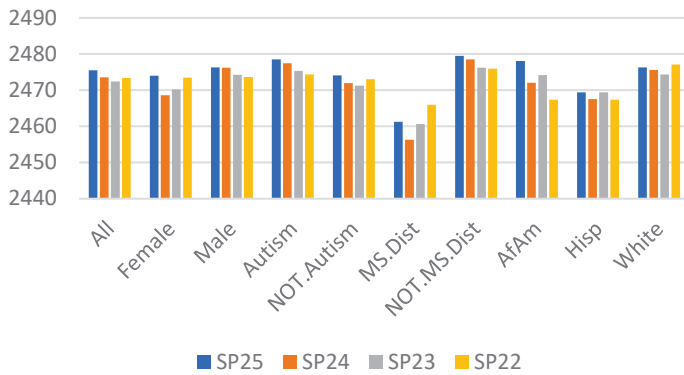


Figure 3: Average Scale Score by Subgroup, Science

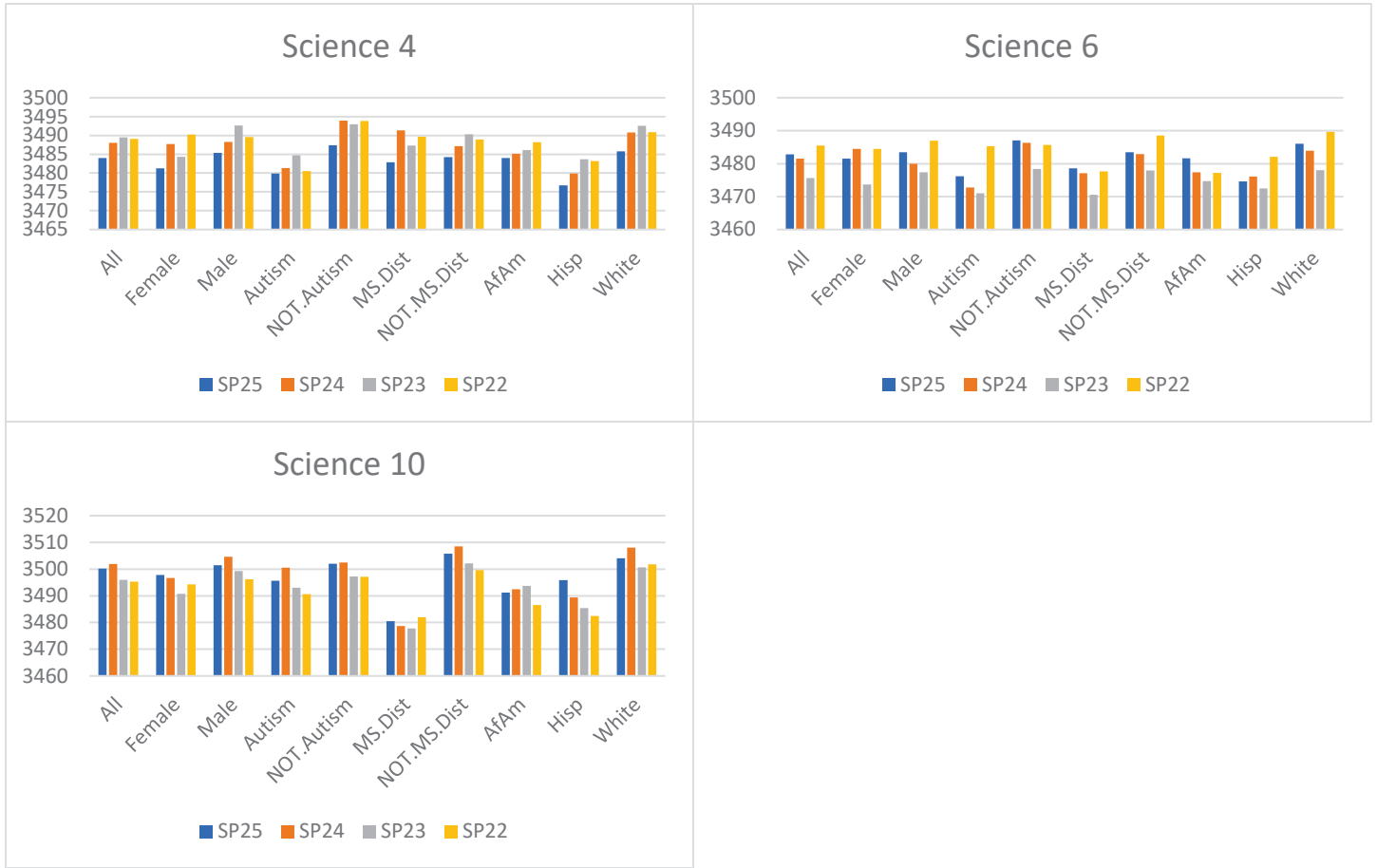
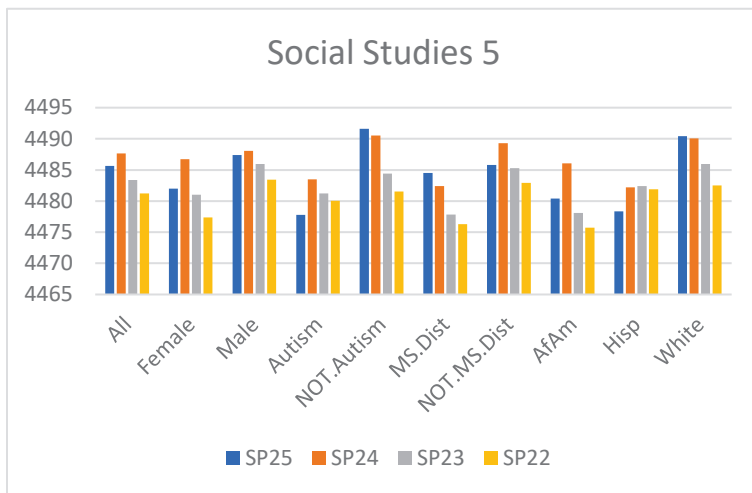


Figure 4: Average Scale Score by Subgroup, Social Studies



3.5 RELIABILITY

Test score reliability is traditionally estimated using both classical and item response theory (IRT) approaches. Classical estimates of test reliability, such as Cronbach’s alpha, provide an index of the internal consistency reliability of the test or the likelihood that a student would achieve the same score in an equivalently constructed test form. While classical indicators provide a single estimate of the reliability of test forms, the precision of test scores varies with respect to the information value of the test at each location. For example, most fixed-form assessments target test information near important cut scores or near the population mean so that test scores are most precise in targeted locations. Because stage-adaptive design targets test information near the student’s ability level in each tier, the precision of test scores may increase, especially for lower- and higher-ability students. The precision of individual test scores is critically important to valid test score interpretation and is provided along with test scores as part of all student-level reporting.

3.5.1 MARGINAL RELIABILITY

While measurement error is conditional on test information, it is nevertheless desirable to provide a single index of a test’s internal consistency reliability. Such an index is provided by the marginal reliability coefficient, which considers the varying measurement errors across the ability range. Marginal reliability is a measure of the overall reliability of an assessment based on the average conditional standard errors, which are estimated at different points on the ability scale for all students. The marginal reliability coefficients are nearly identical or close to the coefficient alpha.

Marginal reliability ($\bar{\rho}$) is defined as

$$\bar{\rho} = [\sigma^2 - \left(\frac{\sum_{i=1}^N CSEM_i^2}{N} \right)] / \sigma^2,$$

where N is the number of students, $CSEM_i$ is the conditional standard error of measurement of the scaled score for student i , and σ^2 is the variance of the scaled score. The higher the reliability coefficient, the greater the precision of the test.

Table 42 to Table 45 present the number of students, marginal reliability coefficients, mean and standard deviation of scale scores, and average standard error of measurement for the total scale scores for the 2024–2025 administration as well as for historical administrations. The marginal reliability coefficients for ELA, Science, and Social Studies range from 0.71 to 0.84, which is similar to other statewide standardized tests. In upper-grade Mathematics, the marginal reliability coefficients are relatively lower than in other assessments. While the marginal reliability coefficients of lower grades in Mathematics had a similar level to other subjects, which is about 0.74, other grades including grades 5–8, and 10 showed the lower marginal reliability coefficients of 0.48–0.64. This is expected due to the small standard deviations of test scores. As seen in Tables 42 to 45, grades 5–8, and 10 mathematics have a smaller standard deviation of scale scores ranging from 26.7 to 32.5, while other subject and grade tests have a standard deviation from 37 to 54, with most over 40.

Table 42: Marginal Reliability, ELA

Grade	Admin	N	Marginal Reliability	Mean	SD	Mean SEM
3	SP25	847	0.704	1475.231	36.826	19.604
	SP24	816	0.719	1476.641	37.836	19.612
	SP23	734	0.686	1476.967	35.305	19.432
	SP22	700	0.704	1477.479	36.197	19.412
4	SP25	925	0.789	1490.405	43.886	19.724
	SP24	862	0.769	1487.947	41.254	19.524
	SP23	774	0.803	1490.319	45.804	19.873
	SP22	680	0.807	1488.372	45.603	19.812
5	SP25	939	0.782	1492.329	42.659	19.637
	SP24	884	0.778	1495.340	42.344	19.600
	SP23	751	0.771	1493.170	41.146	19.491
	SP22	700	0.799	1489.160	44.997	19.886
6	SP25	979	0.776	1489.113	41.927	19.564
	SP24	857	0.744	1492.170	38.446	19.298
	SP23	787	0.775	1483.687	42.369	19.715
	SP22	742	0.792	1489.032	43.851	19.755
7	SP25	963	0.803	1503.465	45.134	19.863
	SP24	879	0.783	1502.346	43.025	19.807
	SP23	841	0.793	1503.367	44.322	19.865
	SP22	871	0.833	1505.937	50.485	20.285
8	SP25	1007	0.796	1493.092	44.852	19.971
	SP24	963	0.783	1495.432	42.573	19.677
	SP23	957	0.824	1498.955	49.683	20.398
	SP22	983	0.822	1497.042	48.917	20.281
10	SP25	1114	0.838	1516.407	54.473	21.244
	SP24	1139	0.826	1515.204	50.977	20.859
	SP23	1066	0.841	1510.018	53.655	20.887
	SP22	897	0.844	1502.905	54.552	20.978

Table 43: Marginal Reliability, Mathematics

Grade	Admin	N	Marginal Reliability	Mean	SD	Mean SEM
3	SP25	843	0.735	2476.008	39.261	19.752
	SP24	808	0.713	2481.574	36.669	19.364
	SP23	727	0.700	2477.912	35.641	19.392
	SP22	701	0.749	2477.914	39.937	19.744
4	SP25	916	0.741	2475.937	39.914	19.945
	SP24	855	0.691	2477.566	35.409	19.559
	SP23	764	0.744	2480.332	38.957	19.594
	SP22	676	0.759	2476.266	41.117	19.960
5	SP25	932	0.591	2472.788	30.644	19.413
	SP24	881	0.598	2477.335	30.610	19.252
	SP23	748	0.531	2470.870	28.510	19.464
	SP22	696	0.603	2470.234	31.263	19.497
6	SP25	968	0.594	2478.615	30.512	19.343
	SP24	853	0.599	2477.358	30.888	19.418
	SP23	779	0.604	2476.302	31.498	19.701

Grade	Admin	N	Marginal Reliability	Mean	SD	Mean SEM
7	SP22	744	0.692	2476.840	35.895	19.802
	SP25	964	0.592	2474.788	30.392	19.220
	SP24	880	0.568	2474.176	29.238	19.155
	SP23	832	0.597	2474.141	30.545	19.244
	SP22	873	0.543	2475.513	28.543	19.198
8	SP25	1003	0.479	2469.277	26.686	19.107
	SP24	961	0.522	2471.786	27.690	19.036
	SP23	950	0.547	2468.849	28.724	19.170
	SP22	977	0.512	2470.014	27.244	18.950
10	SP25	1101	0.641	2475.537	32.456	19.344
	SP24	1141	0.614	2473.568	31.265	19.335
	SP23	1066	0.613	2472.435	31.318	19.379
	SP22	892	0.473	2473.346	27.038	19.497

Table 44: Marginal Reliability, Science

Grade	Admin	N	Marginal Reliability	Mean	SD	Mean SEM
4	SP25	910	0.734	3484.011	38.765	19.725
	SP24	849	0.708	3488.064	36.146	19.453
	SP23	760	0.747	3489.471	39.168	19.571
	SP22	669	0.751	3489.067	39.724	19.637
6	SP25	963	0.765	3482.781	41.023	19.756
	SP24	847	0.787	3481.506	44.049	20.019
	SP23	767	0.768	3475.597	41.794	19.992
	SP22	731	0.718	3485.512	36.978	19.586
Biology	SP25	1138	0.835	3500.127	51.192	20.451
	SP24	1174	0.832	3501.837	50.647	20.417
	SP23	1094	0.810	3495.967	46.864	20.150
	SP22	864	0.804	3495.256	45.931	20.091

Table 45: Marginal Reliability, Social Studies

Grade	Admin	N	Marginal Reliability	Mean	SD	Mean SEM
5	SP25	924	0.755	4485.637	39.802	19.515
	SP24	870	0.756	4487.631	39.503	19.450
	SP23	741	0.791	4483.404	43.737	19.785
	SP22	692	0.779	4481.249	42.522	19.769

3.5.2 STANDARD ERROR OF MEASUREMENT

Within the IRT framework, measurement error varies across the range of abilities. The amount of precision is indicated by the test information at any given point of a distribution.

The inverse of the test information function (TIF) represents the standard error of measurement (SEM). The SEM is equal to the inverse square root of information. The larger the measurement error, the less test information is being provided. The amount of test information provided is at its maximum for students toward the center of the distribution, unlike students with more extreme scores. Conversely, measurement error is minimal for the part of the underlying scale at the middle of the test distribution and greater for scaled values farther away from the middle.

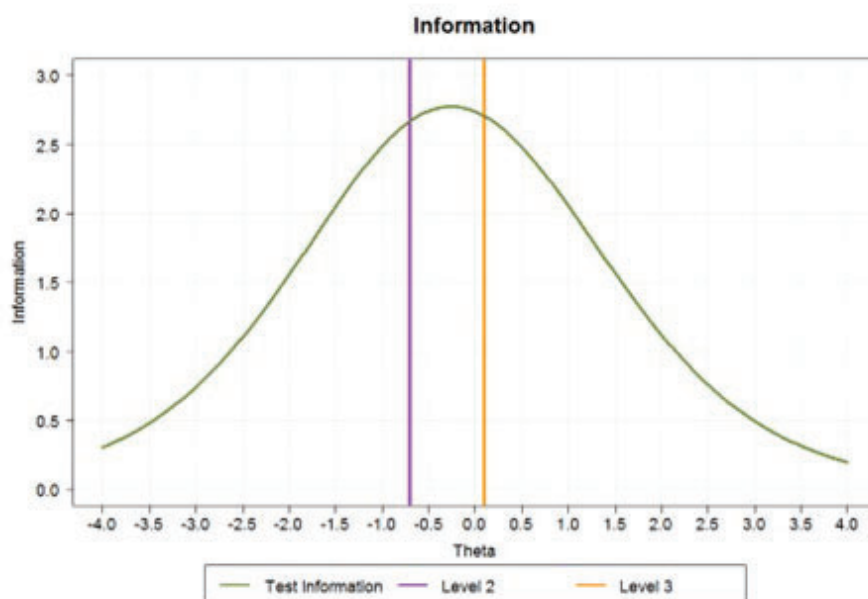
Within the IRT framework, measurement error varies across the range of abilities as a result of the test, providing varied information across the range of abilities as displayed by the TIF. The TIF describes the amount of information provided by the test at each score point along the ability continuum. The inverse of the TIF is characterized as the conditional measurement error at each score point. For instance, if the measurement error is large, then less information is being provided by the assessment at the specific ability level.

Figure 5 displays a sample TIF with two vertical lines indicating the performance cut scores. The graphic shows that this test information is maximized in the middle of the score distribution, meaning it provides the most precise scores in this range. The test provides less information about test takers at the tails, where the curve is lower, relative to the center.

Computing these TIFs is useful to evaluate where the test is maximally informative. In IRT, the TIF is based on the estimates of the item parameters in the test, and the formula used for the *I AM* is calculated as:

$$TIF(\theta_s) = \sum_{i=1}^{N_{PCM}} \left(\frac{\sum_{h=1}^{m_i} h^2 \exp(\sum_{l=1}^h (\theta_s - b_{il}))}{1 + \sum_{h=1}^{m_i} \exp(\sum_{l=1}^h (\theta_s - b_{il}))} - \left(\frac{\sum_{h=1}^{m_i} h \exp(\sum_{l=1}^h (\theta_s - b_{il}))}{1 + \sum_{h=1}^{m_i} \exp(\sum_{l=1}^h (\theta_s - b_{il}))} \right)^2 \right),$$

where N_{PCM} is the number of items that are scored using partial credit model (PCM) items, i indicates item i ($i \in \{1, 2, \dots, N\}$), m_i is the maximum possible score of the item, s indicates student s , and θ_s is the ability of student s .

Figure 5: Sample Test Information Function

The SEM for estimated student ability (theta score) is the square root of the reciprocal of the TIF:

$$se(\theta_s) = \frac{1}{\sqrt{TIF(\theta_s)}}$$

It is typically more useful to consider the inverse of the TIF rather than the TIF itself, as the SEMs are more useful for score interpretation.

SEM plots are presented in Appendix 3-D, Standard Error of Measurement Curves by Subgroup and Appendix 3-E, Standard Error of Measurement Curves by Reporting Category. Vertical lines in the plots represent the Approaching Proficiency and At Proficiency performance category cut scores, respectively.

Table 46 to Table 49 provide the results of the average standard errors for each performance level. Generally, the average standard error is largest in the Below Proficiency level, which can be expected given a shortage of very easy items in the item pools to better measure low-performing students.

Table 46: Average Standard Error of Measurement by Performance Level, ELA

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
3	SP25	21.911	18.515	18.310	19.604
	SP24	21.858	18.521	18.421	19.612
	SP23	21.464	18.551	18.374	19.432
	SP22	21.318	18.634	18.435	19.412
4	SP25	19.719	18.109	20.456	19.724
	SP24	19.760	18.110	19.991	19.524
	SP23	20.159	18.111	20.311	19.873

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
5	SP22	20.159	18.374	20.091	19.812
	SP25	20.457	18.442	19.457	19.637
	SP24	20.526	18.444	19.389	19.600
	SP23	20.163	18.468	19.445	19.491
	SP22	20.035	18.489	20.067	19.886
6	SP25	20.296	18.483	19.574	19.564
	SP24	20.207	18.453	19.319	19.298
	SP23	20.977	18.472	19.415	19.715
	SP22	20.619	18.676	19.731	19.755
7	SP25	19.718	18.505	20.119	19.863
	SP24	20.163	18.611	19.900	19.807
	SP23	20.089	18.561	20.088	19.865
	SP22	19.618	18.384	21.244	20.285
8	SP25	21.134	18.729	20.226	19.971
	SP24	20.672	18.657	19.851	19.677
	SP23	21.016	18.775	21.040	20.398
	SP22	20.024	18.578	21.482	20.281
10	SP25	21.747	18.889	22.205	21.244
	SP24	21.054	18.880	21.804	20.859
	SP23	21.075	18.746	22.093	20.887
	SP22	20.957	18.753	22.529	20.978

Table 47: Average Standard Error of Measurement by Performance Level, Mathematics

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
3	SP25	22.299	18.794	18.575	19.752
	SP24	21.492	18.808	18.608	19.364
	SP23	20.676	18.849	18.822	19.392
	SP22	21.230	18.907	19.032	19.744
4	SP25	22.220	18.958	18.905	19.945
	SP24	21.264	18.950	18.745	19.559
	SP23	21.298	18.887	18.887	19.594
	SP22	21.325	18.914	19.483	19.960
5	SP25	20.717	18.973	18.630	19.413
	SP24	20.741	18.956	18.607	19.252
	SP23	20.846	19.125	18.614	19.464
	SP22	20.876	18.981	18.659	19.497
6	SP25	20.739	18.966	18.681	19.343
	SP24	20.892	18.946	18.681	19.418
	SP23	20.953	19.161	19.018	19.701
	SP22	21.322	19.058	19.148	19.802
7	SP25	20.529	18.652	18.366	19.220
	SP24	20.313	18.654	18.361	19.155
	SP23	20.651	18.596	18.401	19.244
	SP22	20.286	18.789	18.457	19.198
8	SP25	20.357	18.524	18.218	19.107
	SP24	20.158	18.506	18.271	19.036
	SP23	20.278	18.511	18.277	19.170
	SP22	19.924	18.512	18.258	18.950

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
10	SP25	20.144	18.556	18.792	19.344
	SP24	20.109	18.539	18.723	19.335
	SP23	20.165	18.542	18.681	19.379
	SP22	20.650	18.810	18.537	19.497

Table 48: Average Standard Error of Measurement by Performance Level, Science

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
4	SP25	20.509	18.664	19.308	19.725
	SP24	20.065	18.667	19.323	19.453
	SP23	20.010	18.647	19.657	19.571
	SP22	20.021	18.578	19.785	19.637
6	SP25	20.919	18.705	19.541	19.756
	SP24	21.157	18.706	19.824	20.019
	SP23	21.110	18.821	19.496	19.992
	SP22	20.813	19.043	19.116	19.586
Biology	SP25	20.386	18.707	21.338	20.451
	SP24	20.093	18.713	21.470	20.417
	SP23	20.482	18.807	20.972	20.150
	SP22	20.415	18.799	20.666	20.091

Table 49:

Table 50: Average Standard Error of Measurement by Performance Level, Social Studies

Grade	Admin	Below Proficiency	Approaching Proficiency	At Proficiency	Overall
5	SP25	19.334	18.187	20.080	19.515
	SP24	19.344	18.187	19.803	19.450
	SP23	19.700	18.208	20.210	19.785
	SP22	19.587	18.296	20.751	19.769

3.5.3 STUDENT CLASSIFICATION RELIABILITY

When student performance is reported in terms of performance categories, a reliability index is computed in terms of the probabilities of consistent classification of students as specified in Standard 2.16 in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). This index considers the consistency of classifications for the percentage of test takers who would, hypothetically, be classified in the same category on a second *I AM* administration, using either the same form or an alternate, equivalent form.

Students can be misclassified in one of two ways. Students who are truly below a proficiency cut point but are classified based on the assessment as being above the cut point are considered to be *false positives*. Similarly, students who are truly above a

proficiency cut point but are classified as being below the cut point are considered to be *false negatives*.

Decision accuracy refers to the agreement between the classifications based on the form taken and the classifications that would be made based on the test taker's true scores. *Decision consistency* refers to the agreement between the classifications based on the form actually taken and the classifications that would be made based on an alternate form, that is, the percentages of students who are consistently classified in the same proficiency levels on two equivalent administrations of the test.

For a fixed-form test, the consistency of classifications is estimated on single-form test scores from a single test administration based on the true-score distribution that is estimated by fitting a bivariate beta-binomial model or a four-parameter beta model (Huynh, 1979; Livingston & Lewis, 1995). For the spring 2019 administration and all future administrations, the consistency classification is based on all sets of items administered across students because each student takes one of three stage-adaptive forms.

The classification index can be examined for decision accuracy and decision consistency. Decision accuracy refers to the agreement between the classifications based on the form actually taken and the classifications that would be made based on the test takers' true scores, if their true scores could somehow be known. Decision consistency refers to the agreement between the classifications based on the form (adaptively administered items) actually taken and the classifications that would be made based on an alternate, equivalently constructed test form or test administration (e.g., another set of adaptively administered items given the same ability)—that is, the percentages of students who are consistently classified in the same performance levels on two equivalent test administrations.

The true score is an expected value of the test score with measurement error. For a student with estimated ability $\hat{\theta}$ and associated standard error $se(\hat{\theta})$, we can assume that $\hat{\theta}$ follows a normal distribution with mean of true ability θ and standard deviation of $se(\hat{\theta})$, that is, $\hat{\theta} \sim N(\theta, se(\hat{\theta})^2)$. The probability of the true score at or above the cut score θ_c is estimated as

$$P(\theta \geq \theta_c) = P\left(\frac{\theta - \hat{\theta}}{se(\hat{\theta})} \geq \frac{\theta_c - \hat{\theta}}{se(\hat{\theta})}\right) = P\left(\frac{\hat{\theta} - \theta}{se(\hat{\theta})} < \frac{\hat{\theta} - \theta_c}{se(\hat{\theta})}\right) = \Phi\left(\frac{\hat{\theta} - \theta_c}{se(\hat{\theta})}\right),$$

where $\Phi(\cdot)$ is the cumulative function of standard normal distribution. Similarly, the probability of the true score being below the cut score is estimated as

$$P(\theta < \theta_c) = 1 - \Phi\left(\frac{\hat{\theta} - \theta_c}{se(\hat{\theta})}\right).$$

3.5.4 CLASSIFICATION ACCURACY

Instead of assuming a normal distribution, we can directly estimate the probability of consistent classification using the likelihood function. The likelihood function of the achievement attribute, designated θ , given a student's item scores, represents the likelihood of the student's ability at that theta value. Integrating the likelihood values over the range of theta at and above the cut score (with proper normalization) represents the probability of the student's latent ability or the true score being at or above that cut point. If a student's estimated theta is below the cut score, the probability of *at or above* the cut score is an estimate of the chance that this student is misclassified as below the cut score, and 1 minus that probability is the estimate of the chance that the student is correctly classified as below the cut score. Using this logic, we can define various classification probabilities.

The probability of a student with true ability θ being classified at or above the cut score θ_c , given the student's item scores $\mathbf{x} = (x_1, \dots, x_N)$, can be estimated as

$$P(\theta \geq \theta_c | \mathbf{x}) = \frac{\int_{\theta_c}^{+\infty} L(\theta | \mathbf{x}) d\theta}{\int_{-\infty}^{+\infty} L(\theta | \mathbf{x}) d\theta},$$

where the likelihood function is

$$L(\theta | \mathbf{x}) = \prod_{i=1}^N P(x_i | \theta),$$

and $P(x_i | \theta)$ is calculated from the Rasch model or partial credit model based on the estimated item parameters.

Similarly, we can estimate the probability of being below the cut score as:

$$P(\theta < \theta_c | \mathbf{x}) = \frac{\int_{-\infty}^{\theta_c} L(\theta | \mathbf{x}) d\theta}{\int_{-\infty}^{+\infty} L(\theta | \mathbf{x}) d\theta}$$

Mathematically, we have

$$N_{11} = \sum_{i \in N_1} P(\theta_i \geq \theta_c | \mathbf{x}),$$

$$N_{01} = \sum_{i \in N_1} P(\theta_i < \theta_c | \mathbf{x}),$$

$$N_{10} = \sum_{i \in N_0} P(\theta_i \geq \theta_c | \mathbf{x}), \text{ and}$$

$$N_{00} = \sum_{i \in N_0} P(\theta_i < \theta_c | \mathbf{x})$$

where N_1 consists of the students with estimated $\hat{\theta}_i$ being at and above the cut score, and N_0 contains the students with estimated $\hat{\theta}_i$ being below the cut score. The accuracy index is then computed as:

$$\frac{N_{11} + N_{00}}{N_1 + N_0}.$$

In Exhibit A, accurate classifications occur when the decision made based on the true score agrees with the decision made based on the form taken. Misclassifications, false positives, and false negatives occur when students' true-score classifications differ from their observed-score classifications (e.g., a student whose true score results in an At Proficiency level classification but is classified incorrectly as Approaching Proficiency). N_{11} represents the expected number of students who are truly above the cut score; N_{01} represents the expected number of students falsely above the cut score; N_{00} represents the expected number of students truly below the cut score; and N_{10} represents the number of students falsely below the cut score.

Exhibit A: Classification Accuracy

		Classification on a Form Actually Taken	
		At or Above the Cut Score	Below the Cut Score
Classification on True Score	At or Above the Cut Score	N_{11} (Truly above the cut score)	N_{10} (False negative)
	Below the Cut Score	N_{01} (False positive)	N_{00} (Truly below the cut score)

3.5.5 CLASSIFICATION CONSISTENCY

To estimate the consistency, we assume students are tested twice independently; hence, the probability of the student being classified as at or above the cut score θ_c in both tests can be estimated as

$$P(\theta_1 \geq \theta_c, \theta_2 \geq \theta_c) = P(\theta_1 \geq \theta_c)P(\theta_2 \geq \theta_c) = \left(\frac{\int_{\theta_c}^{+\infty} L(\theta|x)d\theta}{\int_{-\infty}^{+\infty} L(\theta|x)d\theta} \right)^2.$$

Similarly, the probability of consistency for at or above the cut score is estimated as

$$P(\theta_1 \geq \theta_c, \theta_2 \geq \theta_c | x) = \left(\frac{\int_{\theta_c}^{+\infty} L(\theta|x)d\theta}{\int_{-\infty}^{+\infty} L(\theta|x)d\theta} \right)^2.$$

The probability of consistency for below the cut score is estimated as

$$P(\theta_1 < \theta_c, \theta_2 < \theta_c | \mathbf{x}) = \left(\frac{\int_{-\infty}^{\theta_c} L(\theta | \mathbf{x}) d\theta}{\int_{-\infty}^{+\infty} L(\theta | \mathbf{x}) d\theta} \right)^2.$$

The probability of inconsistency is estimated as

$$P(\theta_1 \geq \theta_c, \theta_2 < \theta_c | \mathbf{x}) = \frac{\int_{\theta_c}^{+\infty} L(\theta | \mathbf{x}) d\theta \int_{-\infty}^{\theta_c} L(\theta | \mathbf{x}) d\theta}{\left[\int_{-\infty}^{+\infty} L(\theta | \mathbf{x}) d\theta \right]^2}, \text{ and}$$

$$P(\theta_1 < \theta_c, \theta_2 \geq \theta_c | \mathbf{x}) = \frac{\int_{-\infty}^{\theta_c} L(\theta | \mathbf{x}) d\theta \int_{\theta_c}^{+\infty} L(\theta | \mathbf{x}) d\theta}{\left[\int_{-\infty}^{+\infty} L(\theta | \mathbf{x}) d\theta \right]^2}.$$

The consistent index is computed as

$$\frac{N_{11} + N_{00}}{N},$$

where

$$N_{11} = \sum_{i \in N} P(\theta_{i,1} \geq \theta_c, \theta_{i,2} \geq \theta_c | \mathbf{x}),$$

$$N_{01} = \sum_{i \in N} P(\theta_i < \theta_c, \theta_{i,2} \geq \theta_c | \mathbf{x}),$$

$$N_{10} = \sum_{i \in N} P(\theta_i \geq \theta_c, \theta_{i,2} < \theta_c | \mathbf{x}),$$

$$N_{00} = \sum_{i \in N} P(\theta_i < \theta_c, \theta_{i,2} < \theta_c | \mathbf{x}), \text{ and}$$

$$N = N_{11} + N_{10} + N_{01} + N_{00}.$$

As shown in Exhibit B, consistent classification occurs when two forms agree on the classification of a student as either *at or above* or *below* the performance standard, whereas inconsistent classification occurs when the decisions made by the forms differ.

Exhibit B: Classification Consistency

		Classification on the Second Form Taken	
		At or Above the Cut Score	Below the Cut Score
Classification on the First Form Taken	At or Above the Cut Score	N_{11} (Consistently Above the Cut)	N_{10} (Inconsistent)
	Below the Cut Score	N_{01} (Inconsistent)	N_{00} (Consistently Below the Cut)

3.5.6 CLASSIFICATION ACCURACY AND CONSISTENCY ESTIMATES

The analysis of the classification index was performed for test scores in the 2024–2025 administration. Tables 50 to 53 present the decision accuracy and consistency indices. Accuracy classifications are slightly higher than the consistency classifications in all performance standards. The consistency classification rate can be somewhat lower than the accuracy rate because consistency assumes two test scores, both of which include measurement error, while the accuracy rate assumes a single test score and the true score, which does not include measurement error. The classification index ranged from 0.67% to 0.82% for accuracy, and from 0.59% to 0.75% for consistency across all grades and subjects. The accuracy and consistency rates for each performance standard are greater for the performance standards associated with smaller standard errors. The better the test is targeted to the student's ability, the higher the classification index.

Table 51: Decision Accuracy and Consistency Indices for Performance Standards, ELA

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
3	SP25	0.849	0.834	0.788	0.772
	SP24	0.849	0.839	0.787	0.779
	SP23	0.838	0.829	0.776	0.764
	SP22	0.858	0.843	0.800	0.782
4	SP25	0.856	0.885	0.802	0.837
	SP24	0.849	0.876	0.793	0.823
	SP23	0.864	0.885	0.813	0.838
	SP22	0.862	0.879	0.809	0.831
5	SP25	0.850	0.850	0.794	0.791
	SP24	0.852	0.853	0.799	0.795
	SP23	0.846	0.844	0.791	0.786
	SP22	0.850	0.871	0.790	0.816
6	SP25	0.846	0.868	0.790	0.812
	SP24	0.849	0.846	0.792	0.784
	SP23	0.834	0.860	0.776	0.803
	SP22	0.854	0.862	0.801	0.808
7	SP25	0.864	0.875	0.812	0.825
	SP24	0.853	0.857	0.795	0.800
	SP23	0.854	0.866	0.801	0.809
	SP22	0.878	0.891	0.830	0.845

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
8	SP25	0.867	0.853	0.815	0.799
	SP24	0.870	0.860	0.820	0.808
	SP23	0.874	0.873	0.826	0.822
	SP22	0.875	0.877	0.826	0.829
10	SP25	0.913	0.889	0.878	0.845
	SP24	0.906	0.887	0.868	0.841
	SP23	0.901	0.895	0.862	0.850
	SP22	0.890	0.902	0.849	0.860

Table 52: Decision Accuracy and Consistency Indices for Performance Standards, Mathematics

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
3	SP25	0.849	0.829	0.789	0.768
	SP24	0.841	0.822	0.783	0.760
	SP23	0.839	0.826	0.776	0.762
	SP22	0.836	0.825	0.776	0.765
4	SP25	0.841	0.827	0.780	0.766
	SP24	0.840	0.828	0.779	0.766
	SP23	0.848	0.837	0.792	0.778
	SP22	0.833	0.833	0.772	0.772
5	SP25	0.798	0.793	0.728	0.716
	SP24	0.819	0.799	0.755	0.729
	SP23	0.798	0.799	0.729	0.728
	SP22	0.801	0.799	0.729	0.725
6	SP25	0.809	0.807	0.741	0.738
	SP24	0.807	0.804	0.737	0.733
	SP23	0.802	0.813	0.734	0.743
	SP22	0.825	0.835	0.765	0.771
7	SP25	0.816	0.804	0.747	0.736
	SP24	0.821	0.808	0.752	0.740

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
8	SP23	0.822	0.808	0.754	0.739
	SP22	0.811	0.802	0.743	0.734
	SP25	0.792	0.781	0.718	0.708
	SP24	0.800	0.794	0.729	0.722
10	SP23	0.797	0.798	0.726	0.727
	SP22	0.799	0.796	0.727	0.724
	SP25	0.826	0.844	0.760	0.783
	SP24	0.822	0.840	0.752	0.779
	SP23	0.821	0.848	0.753	0.785
	SP22	0.782	0.807	0.707	0.733

Table 53: Decision Accuracy and Consistency Indices for Performance Standards, Science

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
4	SP25	0.837	0.862	0.777	0.806
	SP24	0.841	0.856	0.780	0.797
	SP23	0.844	0.868	0.786	0.812
	SP22	0.841	0.872	0.784	0.819
6	SP25	0.865	0.868	0.813	0.815
	SP24	0.860	0.875	0.807	0.823
	SP23	0.853	0.877	0.797	0.826
	SP22	0.859	0.849	0.805	0.791
Biology	SP25	0.872	0.899	0.825	0.858
	SP24	0.874	0.899	0.826	0.855
	SP23	0.861	0.892	0.809	0.849
	SP22	0.869	0.881	0.817	0.834

Table 54: Decision Accuracy and Consistency Indices for Performance Standards, Social Studies

Grade	Admin	Cut Score Accuracy Index		Cut Score Consistency Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3	Cut 1 and Cut 2	Cut 2 and Cut 3
5	SP25	0.865	0.892	0.811	0.845
	SP24	0.872	0.889	0.820	0.844
	SP23	0.876	0.899	0.825	0.856
	SP22	0.877	0.904	0.826	0.862

3.5.7 RELIABILITY FOR SUBGROUPS IN THE POPULATION

The 2024–2025 marginal reliability results for each of the identified subgroups (gender, ethnicity [White, African American, and Hispanic], and Primary Disability [Autism, Non-Autism, Moderate and Severe Intellectual Disability, and Non-Moderate and Severe Intellectual Disability]) were calculated. The marginal reliability coefficients for subgroups, along with historical statistics, are provided in Appendix 3-F, Marginal Reliability Coefficients for Overall and by Subgroup. As the appendix indicates, reliabilities are consistent across subgroups, indicating that the *I AM* assessments measure a common underlying achievement dimension across all subgroups. Where reliability estimates are attenuated, there is an associated decrease in variance within the subgroup population, indicating that the decrease in reliability is likely due to a restriction in range.

3.5.8 REPORTING CATEGORY RELIABILITY

The marginal reliability coefficients and the measurement errors are computed for the reporting categories. Table 54 through Table 57 present the marginal reliability coefficients for reporting categories.

Table 55: Marginal Reliability Coefficients for ELA Reporting Categories

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
3	Key Ideas and Textual Support/Vocabulary	8	1470.145	63.879	1315	1655	0.488
	Reading Foundations	9	1474.63	46.961	1339	1616	0.226
	Structural Elements and Organization/Connection of Ideas/Media Literacy	8	1471.758	63.515	1328	1662	0.453
	Writing	7	1476.857	50.429	1336	1661	0.192
4	Key Ideas and Textual Support/Vocabulary	12–13	1489.349	56.708	1300	1675	0.628

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
	Structural Elements and Organization/Connection of Ideas/Media Literacy	11–12	1489.384	51.082	1318	1687	0.532
	Writing	7–8	1493.954	64.269	1328	1661	0.458
5	Key Ideas and Textual Support/Vocabulary	14	1493.166	51.665	1300	1700	0.617
	Structural Elements and Organization/Connection of Ideas/Media Literacy	9	1496.728	58.961	1329	1677	0.507
	Writing	9	1485.934	57.409	1306	1675	0.499
6	Key Ideas and Textual Support/Vocabulary	11	1486.268	61.179	1307	1680	0.609
	Structural Elements and Organization/Connection of Ideas/Media Literacy	11	1487.781	52.506	1313	1683	0.522
	Writing	8	1491.944	55.942	1336	1676	0.412
7	Key Ideas and Textual Support/Vocabulary	13–14	1504.174	50.678	1312	1700	0.611
	Structural Elements and Organization/Connection of Ideas/Media Literacy	8–10	1507.242	69.266	1300	1664	0.545
	Writing	7–8	1501.234	63.317	1341	1686	0.452
8	Key Ideas and Textual Support/Vocabulary	11–13	1490.312	59.379	1300	1689	0.612
	Structural Elements and Organization/Connection of Ideas/Media Literacy	10–11	1492.098	51.200	1316	1690	0.495
	Writing	7–8	1495.658	64.326	1334	1670	0.477
10	Key Ideas and Textual Support/Vocabulary	12	1522.499	72.372	1300	1677	0.640
	Structural Elements and Organization/Connection of Ideas/Media Literacy	10–11	1516.226	61.058	1338	1699	0.582
	Writing	8	1514.159	64.155	1329	1682	0.506

Table 56: Marginal Reliability Coefficients for Mathematics Reporting Categories

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
3	Algebraic Thinking and Data Analysis	7–8	2475.925	57.972	2323	2653	0.373
	Computation	8	2475.707	55.711	2332	2660	0.378

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
	Geometry and Measurement	7	2475.459	55.148	2341	2677	0.236
	Number Sense	7–8	2466.985	66.436	2329	2660	0.434
4	Algebraic Thinking and Data Analysis	7	2470.921	55.379	2326	2673	0.271
	Computation	7–8	2474.192	60.048	2329	2666	0.403
	Geometry and Measurement	7	2469.092	55.557	2335	2670	0.254
	Number Sense	7	2477.146	64.364	2335	2659	0.394
5	Algebraic Thinking	7–8	2474.239	57.718	2340	2668	0.341
	Computation	7–8	2469.572	51.208	2324	2658	0.253
	Geometry and Measurement, Data Analysis, and Statistics	8	2470.663	56.697	2300	2660	0.415
	Number Sense	8	2467.343	51.187	2347	2677	0.177
6	Algebra and Functions	8	2474.702	46.166	2327	2665	0.200
	Computation	7	2474.104	57.426	2334	2655	0.323
	Geometry and Measurement, Data Analysis, and Statistics	7	2482.535	55.481	2345	2675	0.256
	Number Sense	8	2474.702	46.166	2327	2665	0.200
7	Algebra and Functions	9	2475.067	51.373	2322	2666	0.379
	Data Analysis, Statistics, and Probability	9	2474.687	45.990	2330	2622	0.236
	Geometry and Measurement	7–8	2472.570	49.831	2320	2662	0.263
	Number Sense and Computation	7	2473.629	59.715	2312	2651	0.390
8	Algebra and Functions	9–10	2468.038	43.358	2316	2666	0.236
	Data Analysis, Statistics, and Probability	7	2465.808	49.159	2329	2652	0.154
	Geometry and Measurement	7	2463.018	51.558	2339	2591	0.165
	Number Sense and Computation	7–8	2467.967	50.199	2342	2677	0.087
	Equations and Inequalities (Linear and Systems)	9–10	2468.038	43.358	2316	2666	0.236
10	Functions (Linear and Nonlinear)	7–8	2469.640	54.306	2325	2670	0.243
	Geometry and Measurement	7–8	2481.278	53.587	2318	2645	0.352
	Number Sense and Data Analysis	7	2471.415	46.573	2328	2672	0.045

Table 57: Marginal Reliability Coefficients for Science Reporting Categories

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
4	Analyzing, Interpreting, and Computational Thinking	7–8	3486.978	49.540	3332	3668	0.271
	Explaining Solutions, Reasoning, and Communicating	7–8	3481.473	55.334	3338	3683	0.299
	Investigating	7	3490.293	55.593	3321	3672	0.306
	Questioning and Modeling	9–10	3475.690	63.957	3305	3677	0.596
6	Analyzing, Interpreting, and Computational Thinking	7-8	3477.627	61.764	3300	3651	0.486
	Explaining Solutions, Reasoning, and Communicating	7-8	3480.765	67.277	3325	3654	0.472
	Investigating	8-10	3480.758	53.160	3328	3681	0.396
	Questioning and Modeling	8	3486.655	59.170	3330	3658	0.433
Biology	Analyzing Data and Mathematical Thinking	13-14	3504.849	63.511	3300	3697	0.692
	Communicating Explanations and Evaluating Claims Using Evidence	7-8	3498.985	64.702	3329	3676	0.470
	Developing and Using Modeling to Describe Structure and Function	10-11	3495.420	62.734	3300	3684	0.622

Table 58: Marginal Reliability Coefficients for Social Studies Reporting Categories

Grade	Reporting Category	Number of Items	Mean	SD	Min	Max	Marginal Reliability
5	Civics and Government/History	17	4486.920	41.474	4300	4700	0.567
	Economics	7	4484.819	64.291	4313	4648	0.456
	Geography	8	4481.529	59.479	4311	4657	0.439

3.5.9 RELIABILITY FOR ACCOMMODATED TESTERS

Internal consistency reliabilities are also calculated for accommodated paper-and-pencil test administrations. Given the small number of students for any accommodated test, all accommodated test administrations are collapsed into a single category for the reliability analysis.

Table 58 shows the marginal reliabilities for accommodated versus non-accommodated test administrations. Note that the number of accommodated testers for some assessments was very small, limiting the generalizability of the results. Nevertheless, the reliability of most accommodated test administrations was comparable to that of non-accommodated test administrations, indicating that, like the non-accommodated assessments, accommodated test administrations result in test scores of similar precision as non-accommodated test administrations. Some accommodated tests, including grades 6 and 10 mathematics and grade 4 science, showed low reliabilities due to small variances in scale scores from the limited population.

Table 59: Marginal Reliability Coefficients for Accommodated vs. Non-Accommodated Students

Grade	Accommodated		Non-Accommodated	
	N	Reliability	N	Reliability
ELA				
3	30	0.378	817	0.709
4	26	0.735	899	0.790
5	19	0.767	920	0.783
6	32	0.606	947	0.779
7	18	0.717	945	0.803
8	27	0.880	980	0.790
10	19	0.774	1095	0.839
Mathematics				
3	30	0.694	813	0.737
4	26	0.708	890	0.742
5	19	0.722	913	0.584
6	31	0.599	937	0.594
7	19	0.232	945	0.592
8	26	0.630	977	0.474
10	19	0.510	1082	0.642
Science				
4	26	0.762	884	0.734
6	31	0.723	932	0.766
Biology	22	0.808	1116	0.835
Social Studies				
5	18	0.578	906	0.756

4. ITEM DEVELOPMENT AND TEST CONSTRUCTION

4.1 TEST DESIGN AND TEST SPECIFICATIONS

The *I AM* assessments are designed to measure student achievement of the Indiana Content Connectors. The Indiana Content Connectors were designed as an extension of the Indiana Academic Standards (IAS) and were adopted by the Indiana State Board of Education to measure the knowledge and skills of students with significant cognitive disabilities. To ensure that the *I AM* assessments appropriately measure the knowledge and skills of the *I AM* student population, assessment blueprints were constructed to represent the range of content defined in the Indiana Content Connectors. This ensures the assessments result in accurate classifications of student achievement. The *I AM* assessments are designed to support the claims about proficiency described at the outset of this chapter.

This section describes the development of *I AM* assessment blueprints that yield valid and reliable assessment scores and proficiency-level classifications to indicate whether students have demonstrated the knowledge and skills associated with the Indiana Content Connectors. The details in this section support the claim that the blueprints are technically sound and consistent with current professional standards.

4.1.1 *I AM* BLUEPRINT DEVELOPMENT

Cambium Assessment, Inc. (CAI) worked closely with the Indiana Department of Education (IDOE) to create blueprints that guide the development process for the *I AM* assessments. Blueprints are the assessment design specifications that ensure assessment scores support the Performance-Level Descriptors (PLDs) described in Chapter 7.1, Standard-Setting Procedures. Blueprints specify the proportionality of how *I AM* assesses the Indiana Content Connectors, including the relative range of each Content Connector on the assessment as represented in the minimum and maximum number of items to be administered to each student.

CAI and IDOE recruited Indiana educators to inform *I AM* blueprint development in June 2018. These educators represented different regions of the state, diverse student populations, and content and accessibility expertise. Panels of content and special education educators serving students with significant cognitive disabilities were convened at each grade level, where they recommended the priorities and associated item ranges used within the blueprints. Educators also considered the vertical articulation of content across grades 3–10.

The *I AM* assessments must provide a valid assessment of the Content Connectors. They were designed as part of a system of assessments with the Indiana Learning Evaluation Assessment Readiness Network (*ILEARN*) and should work alongside *ILEARN* to provide similar data that are meaningful and appropriate for students with significant cognitive disabilities. To meet these requirements, *I AM* assessment blueprints were constructed to include the range of content defined in the IAS, as represented on *ILEARN*, but aligned

with the Content Connectors that are appropriate for the *I AM* student population to achieve the result of the accurate classification of student achievement.

The workshop began with a large-group session to orient participants to the workshop objectives and review the agenda activities to meet those objectives. IDOE oriented participants to the standardized process to be followed and detailed IDOE expectations around their participation.

During the large-group session, discussion emphasized that blueprints that reflect the breadth of the subject-area content domains, cognitive complexity, and vertical articulation across grades must be developed to ensure assessments align to the IAS Content Connectors for the *I AM* population. Participants then broke up into grade-level groups.

In order to design blueprints that would yield valid and reliable assessment scores and proficiency-level classifications able to indicate whether students demonstrate the knowledge and skills associated with the Content Connectors, blueprint meeting participants began by reviewing the Content Connectors and identifying key evidence that demonstrated proficiency in each Content Connector.

Next, using the *ILEARN* reporting categories created by Indiana educators during the *ILEARN* workshops in February 2018, CAI and IDOE presented two documents for each content area to the participants:

- 1) A completed *ILEARN* blueprint for the content area and grade, with the percentages and item minimums/maximums for the reporting categories and IAS for reference.
- 2) A draft *I AM* blueprint for the content area and grade, with all percentages and item minimums/maximums for the reporting categories and Content Connectors left blank. Participants filled in the blank spaces to prioritize and determine the critical importance of each standard for the *I AM* student population.

Because grade 10 blueprints for English/Language Arts (ELA) and Mathematics were not constructed by the *ILEARN* committees, participants used the *ILEARN* blueprints developed for grades 7 and 8 ELA and Mathematics as a reference point for the *I AM* grade 10 discussions. Grade 10 workshop participants were given wide latitude to change the blueprint based on their discussions during workshop sessions.

Grade 10 ELA and Mathematics workshop participants received the following:

- 1) A completed *ILEARN* blueprint for the content area for grades 7 and 8, with the percentages and item minimums/maximums for the reporting categories and IAS for reference.
- 2) A list of all Content Connectors in general blueprint form without reporting categories, prioritization, percentages, or item minimums/maximums listed. Participants determined reporting categories, assigned Content Connector priority, and identified critical importance for the *I AM* student population at grade 10.

Within each subject-area and grade-level panel, panelists worked independently to classify each reporting category as either critically important (3), important (2), or less important (1) to demonstrate mastery of the Content Connectors at that grade level.

Panelists discussed and rationalized their priorities and came to a consensus about the weights of each reporting category. Once weights were determined, percentages were assigned by reporting category.

Next, subject-area panels convened to review the system of weighted reporting categories across the grade-level panels. The goal of the subject-area panel meeting was to ensure any shifts across grades were thoughtful and intentional.

The next step was to classify the Content Connectors according to the relevance of the content being assessed within each of the reporting categories. Panelists worked in subject-area and grade-level groups to indicate which Content Connectors best informed the reporting category and which provided less information for the reporting category.

Panelists first worked independently in Google Polls to classify each Content Connector as either (3) a standard that best informs the reporting category, (2) a standard that provides some information for the reporting category, or (1) a standard that provides little information for the reporting category to demonstrate mastery of the reporting category. After making individual, initial classifications, CAI staff tabulated the scores using Google Polls to show areas of consensus and areas of disagreement in real time. Where a majority of voters agreed (e.g., 4 out of 6 panelists) on a Content Connector's classification, that classification was assigned to the Content Connector. Where there was disagreement about the priority of a standard, panelists further discussed and rationalized their prioritization/classification until they came to a consensus. The panel came to a majority decision about each classification in a draft blueprint.

Next, all grade-level panels convened as one subject-area group to review the prioritized Content Connectors that emerged from the grade-level panels. The overall purpose of the subject-area group meeting was to ensure that any shifts in the importance of Content Connectors across grade levels were thoughtful and intentional.

Panels re-evaluated the previous proportions based on the review of individual Content Connectors, working toward the end goal of final blueprint percentages and determination of reporting category weights.

Following the close of the workshop, CAI worked to incorporate the panelists' feedback in the development of public-facing blueprints for *I AM* assessments. Blueprints were presented for IDOE review prior to a follow-up webinar with workshop participants.

Subject-area panels were reconvened via this follow-up webinar during the week of June 25, 2018. A separate webinar was held for each subject area to review the draft blueprints and ensure they matched the intent of the individual committees. A guided review of the draft blueprints illustrated how each of the blueprint elements was generated from the panelists' feedback based on requirements of the assessment system, reporting framework, and their rating of the Content Connectors and reporting category weights. Subject-area panels evaluated whether revisions should be made to the proposed grade-level blueprints in order to better meet IDOE's assessment goals.

At the conclusion of each webinar, participants confirmed that the recommended blueprints satisfied the requirements for *I AM* and that the *I AM* blueprints developed during the June 2018 meetings achieved the following:

- Measured the breadth and depth of Indiana Content Connectors, aligned to, and derived from the IAS
- Provided weight to the Content Connectors and reporting categories as identified by educators
- Produced accurate and precise test scores and performance-level classifications
- Met required item count limits
- Remained consistent related to measurable content across test administrations

4.1.2 TEST DESIGN

I AM is a stage-adaptive assessment administered in segments. In Part 1, all students take the same assessment form (20 operational items), which measures a range of cognitive complexities. Performance on this first set of items determines the next set of items received in one of three Part 2 forms (each containing 12 operational items): Form A (low complexity); Form B (moderate complexity); or Form C (high complexity). Each form is associated with an item complexity Tier: 1, 2, or 3, respectively.

Each Part 2 form (Form A, Form B, or Form C) contains unique items associated with that form and its tier, as well as items from adjacent tiers. For example, a student who receives Form C will see both Tier 2 and Tier 3 items, while a student who receives Form A will receive only Tier 1 and Tier 2 items. Performance on items from both parts is combined for the final summative scale scores. The overall scale scores for Indiana students align with three proficiency levels (Below Proficiency, Approaching Proficiency, and At Proficiency).

Figure 6 illustrates the *I AM* test design for forms in each grade and subject.

Figure 6: I AM Test Design 2024–2025

Part 1	Part 2		
	Form A	Form B	Form C
item 1	item 21	item 21	item 30
item 2	item 22	item 22	item 31
item 3	item 23	item 23	item 32
item 4	item 24	item 30	item 36
item 5	item 25	item 31	item 37
item 6	item 26	item 32	item 38
item 7	item 27	item 33	item 39
item 8	item 28	item 34	item 40
item 9	item 29	item 35	item 41
item 10	item 30	item 36	item 42
item 11	item 31	item 37	item 43
item 12	item 32	item 38	item 44
item 13			
item 14			
item 15			
item 16			
item 17			
item 18			
item 19			
item 20			

Key
Tier 1 item
Tier 2 item
Tier 3 item

Part 1 is administered to all students. On both online and paper-and-pencil tests, the 20 operational items in Part 1 are separated into two segments. The first segment contains three operational items that allow for early stopping, while the second segment contains the remaining 17 items. Performance in Part 1 determines placement into one of the three Part 2 forms. As the Part 2 stage-adaptive design in Figure 6 shows, item complexities are indicated by color: blue for low complexity, pink for moderate complexity, and green for high complexity. Form A is relatively less difficult, Form C is relatively more difficult, and each of these forms contains nine low-complexity or high-complexity items, respectively. Form B has six items with medium complexity.

Parts 1 and 2 have a combined total of 32 operational items on each form. As shown in Figure 6, 44 unique operational items are generally needed for form building. This is due

to the cross-tier linking pattern in the Part 2 forms. Each Part 2 form contains unique items and items from adjacent tiers. Due to pool constraints and the priority given to meeting blueprint, there were some exceptions in meeting the design in Part 2 of Figure 6. For example, in grade 4 Mathematics Form A, a Tier 3 item was placed in a Tier 1 slot to prioritize meeting blueprint. It should be noted that operational items in Part 2 were assigned to forms based on *a-priori* complexity and item specifications, not item difficulty.

4.1.3 ITEM SPECIFICATIONS

I AM item development is based on the needs formalized by the I AM assessment blueprints and is guided by detailed item specifications, which describe the interaction types that can be used, provide guidelines for targeting the appropriate cognitive engagement, offer suggestions for controlling item difficulty, and offer sample items.

Items are written with the goal that virtually every item will be accessible to all students within the designated population, either by itself or in conjunction with accessibility tools such as text-to-speech, translations, or assistive technologies. This goal is supported by the delivery of the items on CAI's Test Delivery System (TDS), which offers a wide array of accessibility tools and is compatible with most assistive technologies.

Item development supports the goal of high-quality items through rigorous development processes, which are managed and tracked by a content development platform that ensures every item flows through the correct sequence of reviews and captures every comment and change to the item.

Developers seek to ensure that the items meet the standards in a fair and meaningful way by engaging educators and other stakeholders at each step of the item development process. Educators evaluate the alignment of items to the standards and item specifications and offer guidance and suggestions for improvement. They also participate in the review of items for accessibility and fairness.

Combined, these principles and the processes that support them have led to an item pool that measures the standards with fidelity and does so in a way that minimizes construct-irrelevant variance and barriers to access. The details of these processes follow.

The process is guided by passage and item specifications, and includes

- selection and training of item writers;
- writing and internal review of items;
- review by state personnel and stakeholder committees;
- markup for translation and accessibility features;
- field testing; and
- post field-test reviews.

Each of these steps has a role in ensuring that the items can support the claims that will be based on them. Table 59 describes how each step contributes to these goals. Each step in the process is discussed in more detail in the table.

Table 60: Summary of How Each Step of Development Supports Claim Validity

Item Development Step	Supports Alignment to the Standards	Reduces Construct-Irrelevant Variance Through Universal Design	Expands Access Through Linguistic and Other Supports
Passage and item specifications	Specifies item types, passage topics, content limits, Depth of Knowledge (DOK), and guidelines for meeting tier requirements	Avoids the use of any item types with accessibility constraints; provides language guidelines	
Selection and training of item writers	Ensures that item writers have the background to understand the unique needs of the alternate student population, as well as specific details related to standards and specifications	Training in language accessibility and fairness prevents the introduction of unnecessary barriers	
Writing and internal review of items	Checks content and tier alignment; evaluates and improves overall quality	Eliminates editorial issues; flags and removes bias and accessibility issues	
Markup for translation and accessibility features		Adds text-to-speech to reduce barriers	Adds text-to-speech and Spanish translations
Review by state personnel and stakeholder committees	Checks content and tier alignment; evaluates and improves overall quality	Flags sensitivity issues	
Field testing	Provides statistical check on quality; flags issues	Flags for subsequent review items that appear to function differently	May reveal usability or implementation issues with markup
Post-field-test reviews	Final, more focused check on flagged items	Final, more focused review on items flagged for differential item functioning	

Passage Specifications

I AM English/Language Arts (ELA) development begins with passage specifications. Detailed passage specifications ensure that all passages align to the correct grade level and provide sufficient complexity and appropriate subject matter.

Passage specifications for the Indiana Standards Tool for Alternate Reporting (ISTAR) were developed by educators in the Summer of 2017. These passage specifications were used to review passages for the *I AM* assessment by educator stakeholders in collaboration with IDOE content experts and CAI content experts during a Passage

Review workshop in August 2018. At the end of this workshop, participants affirmed through an end-of-workshop survey that the ISTAR passage specifications included passages that are appropriate for the *I AM* student population and were therefore appropriate for continued use as *I AM* passage specifications.

Using the following tools and resources, passages for the *I AM* ELA assessments are evaluated quantitatively for content and vocabulary:

- Lexile® Framework for Reading¹
- ATOS® Readability Formula
- Flesch-Kincaid Grade Level
- EDL Core Vocabularies

The Lexile® Framework for Reading was developed by MetaMetrics, Inc., and employs a scientific formula to calculate the Lexile level of a text based on the semantic and syntactic elements of that text.

The ATOS® Readability Formula considers the most important predictors of text complexity, which are average sentence length, average word length, and word difficulty level. The results are provided on a grade-level scale.

The Flesch-Kincaid Grade Level measures sentence length by the average number of words in a sentence and word length by the average number of syllables in a word to provide the U.S. grade level in which an average student would be able to understand the text.

The EDL Core Vocabularies resource is used for all grades to determine the readability of vocabulary words. The EDL is comprised of words introduced in reading instruction and found on frequency lists. This resource is used to determine what vocabulary to assess in each grade level.

Table 60 provides the quantitative specifications for *I AM* passages by grade for word count, Lexile range, Flesch-Kincaid range, and ATOS range.

Table 61: *I AM* Quantitative Passage Specifications

<i>I AM</i> Grade(s)	Max Word Count	Lexile Range	Flesch-Kincaid Range	ATOS Range
3	250	300–740	1.5–2.0	1.5–2.8
4–5	280	300–820	1.5–5.7	2.0–4.8
6–8	300	300–925	2.0–6.5	2.5–6.0
10	350	400–1050	2.3–7.0	2.8–6.5

Each *I AM* passage is also evaluated qualitatively. The complexity of the passages is reduced through the three tiers, from most complex (Tier 3) to least complex (Tier 1). It is

¹ Lexile ® measures are the intellectual property of MetaMetrics, Incorporated

assumed that students have experience with text in their grade spans or those of earlier grade spans.

Table 61 provides the qualitative specifications for passages by tier.

Table 62: I AM Qualitative Passage Specifications

Tier 1	Tier 2	Tier 3
<ul style="list-style-type: none"> • Passage topic is grade and age appropriate. • Sentences are short and use primarily simple structure, with concrete language and clearly connected pronouns. • Passage is comprised of high-frequency, commonly used vocabulary. • Topic is directly stated and supported with concrete details. • Dialogue is either not used or limited, with no more than one or two people speaking in brief interactions. • Illustrations are used to support the concepts in the passage (typically, 2–3 throughout text, appearing before any associated text). • Text features have simple information with limited detail. • Figurative language, if assessed, is simple. • Assessed vocabulary is two or more grades below the assessed grade. 	<ul style="list-style-type: none"> • Passage topic is grade and age appropriate. • Sentences may include compound subjects and predicates and introductory phrases. • Passage is comprised of mostly high frequency, commonly used vocabulary and some basic subject-specific vocabulary. • Topic may be directly stated or require simple inferences. • Dialogue is limited, with two people speaking in brief interactions. • Images are sometimes used to support the concepts in the passage (typically one right below title). • Text features have information with few details. • Figurative language, if assessed, is simple. • Assessed vocabulary is two or more grades below the assessed grade. 	<ul style="list-style-type: none"> • Passage topic is grade and age appropriate. • Sentences may be a mix of simple and compound structures, as well as some complex constructions. • Passage includes some common expressions, controlled vocabulary, and some subject-specific language. • Topic may include more inferential concepts and themes with multiple characters. • Dialogue may include two or more people speaking. • Images are sometimes used to support the concepts in the passage (typically one right below title). • Text features have information with complex ideas. • Figurative language, if assessed, is simple. • Assessed vocabulary is two or more grades below the assessed grade.

These quantitative and qualitative specifications help test developers create passages that will support appropriate difficulty. The specifications are used in subsequent reviews by IDOE and panelists during committee reviews.

Item Specifications

Item specifications guide the I AM item development process. In July 2018, Indiana educators met to develop item specifications for the new 2018 Content Connectors for ELA, Mathematics, Science, and Social Studies.

The I AM item specifications were designed to provide guidance on how to construct valid and reliable items aligned to the Content Connectors. They were developed specifically for the I AM student population to ensure that the I AM assessments provide a valid

assessment of the Content Connectors and align with the *I AM* assessment blueprints. This allows the *I AM* assessments to provide an accurate classification of student achievement.

Using evidence statements, educators analyzed the Content Connectors for various dimensions outlined on the item specification templates.

The workshop began with a large-group session to orient participants to the workshop objectives and review the agenda activities to meet those objectives. IDOE oriented and standardized the participants in IDOE expectations.

The large-group session focused on helping panelists understand that, to ensure assessments align to the Content Connectors, item specifications must be developed that reflect the breadth of the subject-area content domains, cognitive complexity, and vertical articulation across grades.

Next, subject-area panels convened. Each subject-area group completed two item specification templates as preparation and training for the grade-level work that followed. Discussion was guided by CAI facilitators and IDOE.

In grade-level groups, the participants worked in smaller three-member groups to develop the item specifications for all Content Connectors assessed on the *I AM* blueprints for their grade and subject area. Item specifications were completed based on educator discussions by CAI facilitators and IDOE. The small groups were given a designated number of item specifications to complete before reconvening with the larger group.

At designated checkpoints, participants completed peer reviews of the sections they had developed to that point. This was critical to ensure that grade-level expectations were met, that each grade/grade-band working group was consistent in their approach to writing item specifications, and that grade-level-specific content limits were respected.

Following the initial completion of item specifications by grade-level panels, the entire subject area reconvened to review the work performed in the grade-level panels. Each breakout group presented their work for the full subject-area panel to review for consistency across the subject area. Modifications were made by the notetakers to match the panelists' discussions. A CAI/IDOE content-matter expert facilitated.

Following the close of the workshop, CAI reviewed the teacher-crafted item specifications to ensure completeness, rigor, and accuracy. As part of that process, CAI developed any missing sample items as necessary, which were included in the final item specification drafts that were reviewed and approved by IDOE.

Specifications for all assessed grades and subjects include the following:

- **Reporting Category.** This is the blueprint reporting category that the Content Connector is a part of for the *I AM* assessments.
- **Content Connector.** This includes the language and the coding used for the Content Connector (Indiana's alternate standards, aligned to and derived from IAS).

- **Indiana Academic Standard.** This includes the language and coding used for the IAS that the Content Connector is aligned to and derived from.
- **Content Limits.** This section denotes grade-level limitations for assessment. Content limits delineate what terms, concepts, or procedures are acceptable at a particular grade level for a particular standard—and, in some cases, what is not acceptable.
- **Recommended Response Mechanisms.** This section identifies the ways in which students may respond to a prompt.
- **Construct-Relevant Vocabulary.** This section lists any key vocabulary that can be used in the item.
- **Cognitive Complexity (Depth of Knowledge/DOK).** This section indicates a number between 1 and 6. The number corresponds to the Links for Academic Learning (LAL) DOK model, which has six cognitive complexity levels to account for the differentiated needs and abilities of the special education population. DOK represents cognitive complexity and is defined for each Content Connector. Items are to match the recommended DOK of the Content Connector to which it is aligned.
- **Evidence Statements.** Because students with significant cognitive disabilities are a diverse population with a variety of needs, *I AM* items are classified into one of three tiers. Generally, Tier 1 items are less complex than Tier 2 items, and Tier 2 items are less complex than Tier 3 items. The *I AM* item specifications include an evidence statement for each tier. Evidence statements describe the knowledge and skills that an assessment item elicits from students.
 - **Tier 1:** Questions and answer choices include low structural-level items with a range of item difficulty and complexity. Graphics are provided for most answer choices along with text, which gives students a visual support to answer the questions.
 - **Tier 2:** Questions and answer choices include medium structural-level items with a range of item difficulty and complexity. They may include more introductory phrases in the questions and fewer graphics in answer choices than in Tier 1. They also include a greater level of complexity in how students respond to the questions than in Tier 1.
 - **Tier 3:** Questions and answer choices include high structural-level items with a range of item difficulty and complexity. There is more text and few to no graphics in the answer choices. There may be more abstract ideas and inferencing. There is more complexity in how students respond to the questions than in Tier 2.
- **Accessibility and Accommodation Considerations.** This section provides guidance regarding graphics, as well as auditory and visual considerations.
- **Sample Item.** In this section, a sample item is provided along with its corresponding tier.

Table 62 presents sample ELA specifications for one grade 3 Content Connector.

Table 63: Sample ELA Specifications for Grade 3

Reporting Category	Key Ideas & Textual Support/Vocabulary
Content Connector	3.RN.2.2.a.1: Determine the main idea of a text.
IAS	3.RN.2.2: Determine the main idea of a text; recount the key details and explain how they support the main idea.
Content Limits	<p>Items must be passage based.</p> <p>Tier 1 and 2 items should avoid the word “best” in the stem.</p> <p>Tier 1 items should contain picture support in answer choices when possible to aid comprehension.</p> <p>Tier 2 items can contain picture support in answer choices.</p> <p>Tier 3 items should not contain picture support.</p> <p>Tier 1 distractors should demonstrate clearly incorrect understanding of events or details in the passage.</p> <p>Tier 2 distractors should be possible misunderstanding of events or details in the passage or unrelated details or events in the passage.</p> <p>Text complexity will increase with tiers.</p>
Recommended Response Mechanisms	<p>Multiple-Choice (MC)</p> <p>Table Match (TM)</p> <p>Multi-Select (MS)</p>
Construct-Relevant Vocabulary	Main idea
Cognitive Complexity	4
Evidence Statements	
Evidence Statements	<p>Tier 1</p> <p>Students can identify a key detail in the text.</p>
	<p>Tier 2</p> <p>Students can identify an explicitly stated main idea of the text.</p>
	<p>Tier 3</p>

	Students can determine the main idea of a text.
Accessibility and Accommodation Considerations	
Stimulus Graphic Limitations	<p>Stimulus graphics will be limited to clear photos, illustrations, diagrams, tables, and charts that directly relate to the passage topic.</p> <p>Information contained within stimulus graphics is ineligible for assessment unless specifically prescribed by Content Connector and/or evidence statements.</p>
Visual and Auditory Considerations	<p>Graphics will be provided in formats that are accessible to students to understand or process information.</p> <p>Graphics that do not contribute to the student’s understanding should not be included.</p>
Sample Item	
Tier 3	<p>[Stimulus: Passage about the history of telephones]</p> <p>Which sentence tells the main idea?</p> <p>A. No one uses telephones anymore.</p> <p>B. Telephones are a lot bigger than they used to be.</p> <p>C. Telephones have changed a lot over the years.</p>

At the time of item specification development, available item types for the Recommended Response Mechanisms section of the *I AM* item specifications included two-, three-, or four-option MC; five-option MS; and table match. For Mathematics only, numeric/equation response was also considered an available item type.

IDOE and CAI conducted a cognitive laboratory study in the fall of 2018 to learn more about how students taking *I AM* interact with different item types. For the *I AM* student population, three-option MC was recommended as the most appropriate response mechanism. Based on the results of this study, *I AM* item specifications were edited to remove references to item types no longer being considered for *I AM*, from evidence statements and sample items. The edits to the evidence statements and sample items were approved by educator committees. Note, however, that additional item types were retained in the Recommended Response Mechanisms section for further consideration based on future studies that may occur.

All newly developed *I AM* items align to the 2018 *I AM* item specifications. Legacy operational items on the 2024–2025 *I AM* assessments were selected for “best fit” to the new 2018 *I AM* Content Connectors and item specifications. However, because legacy operational items were developed prior to the creation of *I AM* item specifications, not all legacy operational items align fully to the *I AM* item specifications. Alignment of operational legacy items to the 2018 *I AM* Content Connectors was deemed sufficient when alignment to the new 2018 *I AM* item specifications was not possible. Future *I AM*

administrations will continue to replace legacy operational items with new *I AM* items as the depth and breadth of the *I AM* pool increase, with ongoing efforts being made to align *I AM* administrations solely to the 2018 *I AM* item specifications.

Training of Item Writers

All CAI item writers who develop *I AM* items have at least a bachelor's degree, and many have teaching experience. All item writers are trained in

- the principles of universal design;
- the avoidance of bias and sensitivity issues;
- language accessibility guidelines; and
- the *I AM* Passage and Item Specifications.

Key material is included as Appendix 4-A, Language, Accessibility, Bias, and Sensitivity Guidelines and Checklist.

4.1.4 TARGET BLUEPRINTS

Summative Target Blueprints

Blueprints specify a range of items to be administered in each reporting category (or strand). The target blueprints include the requirements for the total test length and the minimum and maximum number of operational items for each score reporting category. Individual scores for each reporting category provide information to help identify areas in which a student may have had difficulty.

Tables 63–66 provide the percentage of operational items required in the blueprints by reporting category for each grade level by subject. The percentages represent an acceptable range of item counts.

Table 64: Blueprint Percentage of Items Assessing Each Reporting Category, ELA

Grade	Reporting Category			
	<i>Key Ideas and Textual Support/Vocabulary</i>	<i>Structural Elements and Organization/Connection of Ideas/Media Literacy</i>	<i>Writing</i>	<i>Reading Foundations</i>
3	22–31%	22–25%	22–25%	22–31%
4	34–41%	31–38%	22–25%	N/A
5	34–44%	28–38%	22–28%	N/A
	<i>Key Ideas and Textual Support/Vocabulary</i>	<i>Structural Elements and Organization/Connection of Ideas/Media Literacy</i>	<i>Writing</i>	<i>Speaking and Listening (Aggregate Only)</i>
6	28–38%	25–34%	22–25%	3–6%
7	28–44%	25–34%	22–25%	3–6%
8	28–44%	25–34%	22–25%	3–6%
10	28–38%	25–34%	22–25%	3–6%

Table 65: Blueprint Percentage of Items Assessing Each Reporting Category, Mathematics

Grade	Reporting Category				
	<i>Algebraic Thinking and Data Analysis</i>	<i>Computation</i>	<i>Geometry and Measurement</i>	<i>Number Sense</i>	<i>Process Standards (Aggregate Only)</i>
3	22–25%	22–25%	22–25%	22–25%	6–12%
4	22–25%	22–25%	22–25%	22–25%	6–12%
	<i>Algebraic Thinking</i>	<i>Computation</i>	<i>Geometry and Measurement, Data Analysis, and Statistics</i>	<i>Number Sense</i>	<i>Process Standards (Aggregate Only)</i>
5	22–25%	22–25%	22–25%	25–28%	3–12%
	<i>Algebra and Functions</i>	<i>Computation</i>	<i>Geometry and Measurement, Data Analysis, and Statistics</i>	<i>Number Sense</i>	<i>Process Standards (Aggregate Only)</i>
6	25–28%	22–25%	22–25%	25–28%	3–12%
	<i>Algebra and Functions</i>	<i>Data Analysis, Statistics, and Probability</i>	<i>Geometry and Measurement</i>	<i>Number Sense and Computation</i>	<i>Process Standards (Aggregate Only)</i>
7	25–28%	22–25%	22–25%	22–25%	3–6%
8	28–31%	22–25%	22–25%	22–25%	3–6%
	<i>Equations and Inequalities (Linear and Systems)</i>	<i>Functions (Linear and Nonlinear)</i>	<i>Geometry and Measurement</i>	<i>Number Sense and Data Analysis</i>	<i>Process Standards (Aggregate Only)</i>
10	22–25%	22–25%	22–25%	22–25%	3–12%

Table 66: Blueprint Percentage of Items Assessing Each Reporting Category, Science

Grade	Reporting Category			
	<i>Analyzing, Interpreting, and Computational Thinking</i>	<i>Explaining Solutions, Reasoning, and Communicating</i>	<i>Investigating</i>	<i>Questioning and Modeling</i>
4	22–25%	22–25%	22–25%	25–34%
6	22–25%	22–25%	25–34%	22–25%
	<i>Analyzing Data and Mathematical Thinking</i>	<i>Communicating Explanations and Evaluating Claims Using Evidence</i>	<i>Developing and Using Modeling to Describe Structure and Function</i>	N/A
Biology	40–50%	22–25%	28–37%	N/A

Table 67: Blueprint Percentage of Items Assessing Each Reporting Category, Social Studies

Grade	Reporting Category		
	<i>Civics and Government/History</i>	<i>Economics</i>	<i>Geography</i>
5	50–56%	22–25%	22–25%

In every case, the percentages across reporting categories on the Spring 2025 forms met the required blueprint range.

To ensure the item pool can support blueprint needs, annual item development plans are developed based on a pool analysis against blueprint needs. Blueprints that guided item development plans that determined the Spring 2019 and Spring 2021 *I AM* field-test pools are provided in Appendices 4-B to 4-E for ELA, Mathematics, Science, and Social Studies, respectively. IDOE created item development plans for items that were field-tested in Spring 2022 and Spring 2023. No new items were developed in advance of the Spring 2025 test administration.

Developing and maintaining a robust operational pool aligned to the *I AM* blueprint requirements will allow for future *I AM* assessment administrations to continue to yield valid and reliable test scores and proficiency-level classifications that indicate whether students taking the *I AM* assessment have demonstrated the knowledge and skills associated with the Indiana Content Connectors.

English Language Arts Score-Reporting Categories

The I AM ELA assessments measure students' understanding of the standards at the end of grades 3–8 and 10. These assessments measure students' proficiency in ELA knowledge and skills. I AM individual student reports describe “at proficiency” ELA performance in the following reporting categories:

Grade 3

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain themes/central ideas; retell texts; describe the effect of characters' actions; connect ideas; and explain the meanings/relationships of words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always distinguish points of view in literature and nonfiction. They can explain text features and illustrations; distinguish between fact and opinion; describe facts that support a point; and compare/contrast two stories from the same author/same topic.
- **Writing.** Your student can almost always recognize characteristics of persuasive, informative, and narrative works. They can organize and use evidence to support ideas and use some appropriate writing conventions, such as capitalizing proper nouns and using regular and irregular verbs.

Grade 4

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain themes/main ideas, describe how characters/settings affect the plot, summarize texts, and explain meanings and relationships of common words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always use illustrations and text features to gain meaning in literature/nonfiction. They can compare and contrast firsthand/secondhand accounts, explain organizational structures and how an author supports a claim, and combine information from texts.
- **Writing.** Your student can almost always recognize characteristics of the three main types of compositions. They can organize and use evidence to support ideas and use some appropriate writing conventions, such as capitalizing the first word in quotations and forming possessives.

Grade 5

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual/inferential questions about text with evidence. They can explain themes/main ideas, describe characters and how their actions affect the plot, summarize texts, and explain the meanings/relationships of common words.

- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always find claims/supporting details and explain points of view in literature/nonfiction. They can compare text structures and versions of the same event; and explain how texts fit together and affect the reader.
- **Writing.** Your student can almost always recognize characteristics of the three main types of writing. They can organize and use evidence to support ideas and use some appropriate writing conventions, such as perfect verb tenses and verbs that are often misused (e.g., lie/lay).

Grade 6

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain themes/central ideas, describe characters and identify how they change, summarize texts, and explain meanings and relationships of common words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always explain relationships between individuals and concepts in literature and nonfiction. They can determine points of view or purpose, identify and use text features as intended, and trace complex arguments and claims.
- **Writing.** Your student can almost always recognize characteristics of the three main types of compositions. They can organize and use evidence to support ideas on the same topic, identify complete complex sentences, and use some appropriate writing conventions.

Grade 7

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain themes/central ideas, describe how story elements interact, summarize texts, and explain meanings/relationships of common words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always explain how text structures and features contribute to ideas in literature and nonfiction. They can explain points of view or purpose, describe similarities and differences in historical accounts/historical fiction, and trace arguments and claims.
- **Writing.** Your student can almost always recognize characteristics of the three main writing types. They can organize and use evidence to support ideas; identify complete complex sentences; use some appropriate conventions; and use specific language that contributes to clarity.

Grade 8

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain themes/central ideas and what details show about characters, summarize texts, and explain meanings/relationships of common words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always explain the importance of text structure and specific details in literature and nonfiction. They can explain points of view/purpose and the connection between ideas, describe conflicting information in two texts, and trace arguments/claims.
- **Writing.** Your student can almost always recognize characteristics of the three main types of compositions. They can organize and use evidence to support ideas on the same topic, identify complete complex sentences, and use some appropriate writing conventions.

Grade 10

- **Key Ideas and Textual Support/Vocabulary.** Your student can almost always answer factual and inferential questions about literature and nonfiction. They can explain two themes/central ideas and how characters develop, analyze the connections between ideas, and explain meanings/relationships of common words.
- **Structural Elements and Organization/Connection of Ideas/Media Literacy.** Your student can almost always explain the importance of author/character perspective in literature/nonfiction. They can explain how text structure contributes to meaning, trace claims and supporting evidence, and explain connections between literary works/world documents.
- **Writing.** Your student can almost always recognize characteristics of the three main types of compositions. They can organize and use evidence to support ideas on the same topic, identify complete complex sentences, and use some appropriate writing conventions.

Mathematics Score-Reporting Categories

The *I AM* mathematics assessments measure students' understanding of the standards at the end of grades 3–8 and 10. These assessments measure students' proficiency in mathematical knowledge and skills, and whether they are adept at demonstrating the process standards. *I AM* individual student reports describe “at proficiency” mathematics performance in the following reporting categories:

Grade 3

- **Algebraic Thinking and Data Analysis.** Your student can almost always use pictures and/or manipulatives when solving real-world word problems involving the four operations with numbers up to 100; create models and apply properties for multiplication or division; and organize given data into a graph or line plot.

- **Computation.** Your student can almost always perform multi-digit addition and subtraction up to 100 with regrouping; sort up to 20 objects into groups of five independently; solve mathematical problems using zero and identity properties of multiplication; and find multiplication facts up to 10.
- **Geometry and Measurement.** Your student can almost always identify all solids or attributes shared among shapes; split shapes into halves, thirds, and fourths; measure volume and select measuring tools; calculate areas of rectangles; and find the perimeter of a polygon with more than four sides.
- **Number Sense.** Your student can almost always read, model, and write whole numbers up to 200 in standard and word form; identify numerators and denominators (thirds); locate unit fractions on number lines; compare two fractions using symbols; and round two-digit numbers to the nearest 10.

Grade 4

- **Algebraic Thinking and Data Analysis.** Your student can almost always apply the relationship between adding and multiplying; show verbal multiplication statements as equations; interpret data from tables, bar graphs, and circle graphs; create line plots using data; and solve one- and two-step word problems.
- **Computation.** Your student can almost always add and subtract numbers with sums up to 500; create models to multiply up to two-digit by one-digit numbers and divide up to 50 without remainders; and use models to add and subtract fractions and mixed numbers with like denominators.
- **Geometry and Measurement.** Your student can almost always categorize shapes based on features; identify parallel and perpendicular lines in given models; identify appropriate measurement units and solve problems involving money and time; and find angles in circles and two-dimensional shapes.
- **Number Sense.** Your student can almost always read, write, compare, and round (to the tens or hundreds place) whole numbers up to 500; write tenths as decimals or fractions; show equivalent fractions up to tenths; and compare fractions and decimals to the tenths using symbols and words.

Grade 5

- **Algebraic Thinking.** Your student can almost always locate/graph ordered pairs on a graph and identify the x- and y-axes; solve one-step decimal problems using addition, subtraction, or multiplication to the hundredths place; and solve two-digit multiplication or division word problems.
- **Computation.** Your student can almost always multiply two-digit by two-digit numbers and divide numbers up to 100 without remainders; add or subtract fractions with unlike denominators (fourths, fifths, and tenths); and solve addition or multiplication expressions with parentheses.

- **Geometry and Measurement, Data Analysis, and Statistics.** Your student can almost always answer one-step questions about graphs and find the mode and median of line plots; count the number of sides of a hexagon, trapezoid, and rhombus; and convert measurements of time, such as hours in a day and months in a year.
- **Number Sense.** Your student can almost always compare two fractions or two decimals using the vocabulary "greater than or less than" and using $<$, $>$, or $=$ symbols; round decimals to the nearest whole number; and use models to show percentage as part of 100.

Grade 6

- **Algebra and Functions.** Your student can almost always create equivalent expressions; solve one-step linear equations; write inequalities for real-world problems; plot ordered pairs in all four quadrants; write and solve variable expressions; and analyze variables in proportional relationships.
- **Computation.** Your student can almost always divide using multi-digit numbers; divide with fractions (one step); add and subtract with decimals or fractions; represent and evaluate exponents; and apply order of operations in mathematical expressions.
- **Geometry and Measurement, Data Analysis, and Statistics.** Your student can almost always convert between measurement systems; identify data in statistical questions; collect and graph data; find patterns (range, mean, and mode) among data; solve triangle angle problems; and find area (quadrilaterals) or volume (rectangular prisms).
- **Number Sense.** Your student can almost always find, plot, and compare numbers; describe ratio relationships; solve one-step real-world ratio problems; find greatest common factors or least common multiples; and identify decimal or percentage equivalents (halves, fourths, fifths, and tenths).

Grade 7

- **Algebra and Functions.** Your student can almost always use variables to model and solve two-step, real-world equations or inequalities; find a proportional relationship or unit rate from tables or coordinates; calculate the slope; and graph a line using slope and a point.
- **Data Analysis, Statistics, and Probability.** Your student can almost always draw conclusions from data; find range, median, mean, or mode; compare two similar populations to draw conclusions; make predictions based on probability; and compare results of simple experiments with theoretical probabilities.

- **Geometry and Measurement.** Your student can almost always identify similar polygons; determine an appropriate scale for real-world situations; identify various angles in real-world situations; calculate the area or circumference of circles; and calculate the volume of cylinders.
- **Number Sense and Computation.** Your student can almost always add, subtract, multiply, and divide integers to solve problems; find the distance between rational points on a number line using absolute value; order and compare rational and irrational numbers on a number line; and identify perfect squares.

Grade 8

- **Algebra and Functions.** Your student can almost always recognize when linear equations have one, many, or no solutions and solve two-step equations in context; describe multiple features of linear and nonlinear graphs and functions; and solve systems of linear equations.
- **Data Analysis, Statistics, and Probability.** Your student can almost always graph data on a scatterplot and identify associations between variables; use the line of best fit to find a point that answers a question about the data; and determine the probability of multistage events and the total number of outcomes.
- **Geometry and Measurement.** Your student can almost always describe attributes of three-dimensional objects; use volume formulas; describe the effects of a sequence of transformations on a figure; and use the Pythagorean theorem to determine distance on a coordinate plane.
- **Number Sense and Computation.** Your student can almost always solve two-step problems with rational numbers and scientific notation; round to the hundredths place and estimate the location of irrational numbers on a number line; and solve problems using square roots and integer exponents.

Grade 10

- **Equations and Inequalities (Linear and Systems).** Your student can almost always solve two-step equations with integer coefficients; represent real-world situations with a proportion, graph, inequality, or absolute value; and solve systems of linear equations and inequalities that represent real-world problems.
- **Functions (Linear and Nonlinear).** Your student can almost always describe a function as linear or nonlinear; distinguish between functions and non-functions using tables and graphs; describe the properties of quadratic functions in real-world context; and solve equations using properties of square roots.
- **Geometry and Measurement.** Your student can almost always describe attributes of three-dimensional shapes and use the volume formula; describe the sequence of transformations between two congruent figures and their coordinates; and apply the Pythagorean theorem to determine lengths and distances.

- **Number Sense and Data Analysis.** Your student can almost always interpret bivariate data on scatterplots and two-way tables; use the multiplication counting principle to determine probability outcomes; use factoring to find equivalent expressions and add, subtract, multiply, and divide polynomials.

Science Score-Reporting Categories

The *I AM* science assessments measure students' understanding of the standards at the end of grades 4 and 6, and high school Biology. These assessments measure students' proficiency in science knowledge and skills. *I AM* individual student reports describe "at proficiency" science performance in the following reporting categories:

Grade 4

- **Questioning and Modeling.** Your student can almost always identify a scenario that matches a question, the outcome of a series of events, which two steps are missing in a model, the missing moon phase in a series, the link between the moon and tides, and obtain information to solve a problem.
- **Investigating.** Your student can almost always determine missing parts of an experiment, improve models, describe how erosion changed land, explain how energy relates to speed, identify two simple machines working together, and identify inherited traits for survival.
- **Analyzing, Interpreting, and Computational Thinking.** Your student can almost always identify solutions to a problem, the functions of given devices and technology used in a task, the multiple effects of a cause, how energy transfers from place to place, and explain how different types of fuel can affect the environment.
- **Explaining Solutions, Reasoning, and Communicating.** Your student can almost always develop solutions to reduce the impact of humans on an ecosystem, describe the different ways energy can be created or converted from one form to another, evaluate online resources, and use evidence to make predictions and/or support a claim.

Grade 6

- **Questioning and Modeling.** Your student can almost always describe the link between hardware and software, how gravity or inertia affects the motion of objects in space, use models/data to show the energy flow in a food web, ask a testable question about motion, and identify the constraints of a design.
- **Investigating.** Your student can almost always identify helpful and harmful impacts of technology, predict how changes in an ecosystem affect living and nonliving things, describe how specific organisms relate in an ecosystem, and use data to compare moving objects and planets and moons.

- **Analyzing, Interpreting, and Computational Thinking.** Your student can almost always explain how balance is needed for living things to meet their needs and how potential and kinetic energy can change forms, resolve hardware issues, model how Earth's movements cause seasons and daylight hours, and organize an investigation.
- **Explaining Solutions, Reasoning, and Communicating.** Your student can almost always provide proper feedback to make improvements, list electronic resources for a topic, recognize that some materials reflect or absorb light or sound waves, develop a solution for a problem, find information, and use evidence to support an argument.

Biology

- **Developing and Using Modeling to Describe Structure and Function.** Your student can almost always group proteins, carbohydrates, and lipids based on function, label specialized structures in a cell model, illustrate how matter/energy moves through an ecosystem, and model the steps of protein synthesis using a codon ring/chart.
- **Analyzing Data and Mathematical Thinking.** Your student can almost always show how a limited resource affects a population, how human or natural events change the flow of matter/energy in an ecosystem and describe ways to reduce their impact, interpret data to predict traits of offspring, and evaluate an investigation.
- **Communicating Explanations and Evaluating Claims Using Evidence.** Your student can almost always explain the role of natural selection in how species adapt, how environmental impacts affect population size, use evidence to group organisms based on taxonomic categories, describe the factors affecting evolution, and use tools to make solutions.

Social Studies Score-Reporting Categories

The *I AM* social studies assessment measures students' understanding of the standards at the end of grade 5. The assessment measures students' proficiency in social studies knowledge and skills. *I AM* individual student reports describe "at proficiency" social studies performance in the following reporting categories:

Grade 5

- **Civics and Government/History.** Your student can almost always use sources independently and be an active citizen. Your student can almost always explain early settlements in North America, the key ideas and events of the founding of the United States, and the type of U.S. government.

- **Geography.** Your student can almost always use maps independently to locate places and recognize regions. Your student can almost always use a map to identify geographical features from both today and the past.
- **Economics.** Your student can almost always describe examples of early Native American and colonial cultures' economic activities. Your student can almost always define a market economy and describe and explain factors that make it work.

Accommodated Paper-and-Pencil Form Construction

Students who are unable to participate in the online administration are administered the test in a paper-and-pencil format as an accommodation. The paper-and-pencil format includes the same operational items as the online assessment. For the paper-and-pencil tests, one of the embedded field-test (EFT) blocks is fixed for all students in each of the grade and subject-area tests.

4.1.5 BLUEPRINT MATCH

ELA Blueprints

The *I AM* blueprints developed for ELA grades 3–8 and 10 are provided in Appendix 4-B, English/ Language Arts Blueprints.

The key features of the *I AM* ELA blueprints include reporting categories, reporting category allocations, Content Connectors, Content Connector allocations (number of minimum and maximum items per Content Connector), and total number of operational items.

Reporting Categories

The *I AM* ELA blueprints are organized by reporting category and specify the number of items required for each reporting category, thus ensuring the form contains enough items from each category to elicit sufficient information from the student to justify reporting category-level scores. The *I AM* ELA grade 3 blueprint includes an additional reporting category for Reading Foundations.

Reporting categories comprise a broad domain—or segment—of the subject area identified by educators as containing meaningful sets of interrelated Content Connectors. Reporting categories are broad to allow for individual-level reporting of student performance. In many cases, the reporting category combines two or more related domains, as indicated by educators.

The *I AM* ELA blueprints in grades 6–8 and 10 also include Speaking and Listening Content Connectors that contribute to the student score as a whole.

Reporting Category Allocations

The *I AM* ELA blueprints include the overall percentage of the assessment characterized by each reporting category. For ELA grade 3, educators placed an emphasis on Reading Foundations and literary texts. Blueprints for grades 4 and 5 continue to emphasize literary texts, transitioning to place more emphasis on nonfiction texts in grades 6–8 and 10. On the *I AM* ELA assessment, the focus of reading is on comprehending text. To meet the varied needs of this population, reading is defined broadly to allow for students who require use of appropriate accommodations (e.g., listening to text read aloud).

Content Connectors

The *I AM* ELA blueprints list the code for each Content Connector in each reporting category.

Content Connector Allocations

The *I AM* ELA blueprints also specify the minimum and maximum number of items per Content Connector. A Content Connector with a range that starts at 0 indicates that the Content Connector may not be assessed each year. The item ranges in the blueprint allow each student to experience a wide range of content while still providing flexibility during form construction.

Total Number of Operational Items

The total number of operational items on each *I AM* ELA assessment is 32.

Mathematics Blueprints

The *I AM* blueprints developed for Mathematics grades 3–8 and 10 are provided in Appendix 4-C, Mathematics Blueprints. The blueprints for grades 3–8 were finalized in December 2018. The blueprint for grade 10 was finalized in June 2019.

The key features of the *I AM* Mathematics blueprints include reporting categories, reporting category allocations, Content Connectors, Content Connector allocations (minimum and maximum number of items per Content Connector), and the total number of operational items.

Reporting Categories

The *I AM* Mathematics blueprints are organized by reporting category and specify the number of items required for each reporting category, ensuring that the form contains enough items from that category to elicit sufficient information from the student to justify reporting category-level scores.

Reporting categories comprise a broad domain, or segment, of the subject area, identified by educators as containing meaningful sets of interrelated Content Connectors. Reporting categories are broad to allow for individual-level reporting of student performance. In many cases, a reporting category combines two or more related domains, as indicated by educators.

The *I AM* Mathematics blueprints also include Content Connectors in a category that is reported as an aggregate score. The items assessing these Content Connectors will contribute to the student score as a whole.

Reporting Category Allocations

The *I AM* Mathematics blueprints include the overall percentage of the assessment characterized by each reporting category. For Mathematics, educators determined that all reporting categories should have equal emphasis in grades 3 and 4. For grades 5 and 6, educators placed an emphasis on Number Sense and transitioned to place more focus on Algebra and Functions in grades 7–8. Educators determined that all reporting categories should have equal emphasis for grade 10.

Content Connectors

The *I AM* Mathematics blueprints list the code of each Content Connector in each reporting category.

Content Connector Allocations

The *I AM* Mathematics blueprints specify the minimum and maximum number of items per Content Connector. A Content Connector with a range that starts at 0 indicates that the Content Connector may not be assessed each year. The item ranges in the blueprint allow each student to experience a wide range of content while still providing flexibility during form construction.

Total Number of Operational Items

The total number of operational items on each *I AM* Mathematics assessment is 32.

Science Blueprints

The *I AM* blueprints developed for Science grades 4 and 6 and Biology are provided in Appendix 4-D, Science Blueprints. The blueprints for grade 6 and Biology were finalized in December 2018. The Biology blueprint was finalized in June 2019.

The key features of the *I AM* Science blueprints include reporting categories, reporting category allocations, Content Connectors, Content Connector allocations (minimum and maximum number of items per Content Connector), and total number of operational items.

Reporting Categories

The *I AM* Science blueprints are organized by reporting category and specify the number of items required for each reporting category, ensuring that the form contains enough items from that category to elicit enough information from the student to justify reporting category-level scores.

Reporting categories comprise a broad domain, or segment, of the subject area, identified by educators as containing meaningful sets of interrelated Content Connectors. Reporting categories are broad to allow for individual-level reporting of student performance. In many cases, a reporting category combines two or more related domains, as indicated by educators.

Reporting Category Allocations

The *I AM* Science blueprints include the overall percentage of the assessment characterized by each reporting category. For grade 4 Science, educators determined that Questioning and Modeling was of greatest priority. For grade 6 Science, educators placed an emphasis on Investigating. In the Biology End-of-Course Assessment (ECA), educators determined that Analyzing Data and Mathematical Thinking should receive the greatest emphasis.

Content Connectors

The *I AM* Science blueprints list the code of each Content Connector in each reporting category.

Content Connector Allocations

The *I AM* Science blueprints also specify the minimum and maximum number of items per Content Connector. A Content Connector with a range that starts at 0 indicates that the Content Connector may not be assessed each year. The item ranges in the blueprint allow each student to experience a wide range of content while still providing flexibility during form construction.

Total Number of Operational Items

The total number of operational items on each *I AM* Science assessment is 32.

Social Studies Blueprints

The *I AM* blueprint developed for Social Studies grade 5 is provided in Appendix 4-E, Social Studies Blueprints. The Social Studies grade 5 blueprint was finalized in June 2019.

The key features of the *I AM* Social Studies blueprint include reporting categories, reporting category allocations, Content Connectors, Content Connector allocations (minimum and maximum number of items per Content Connector), and total number of operational items.

Reporting Categories

The *I AM* Social Studies blueprint is organized by reporting category and specifies the number of items required for each reporting category, ensuring that the form contains enough items from that category to elicit sufficient information from the student to justify reporting category-level scores.

Reporting categories comprise a broad domain, or segment, of the subject area, identified by educators as containing meaningful sets of interrelated Content Connectors. Reporting categories are broad to allow for individual-level reporting of student performance. In many cases, a reporting category combines two or more related domains, as indicated by educators.

Reporting Category Allocations

The *I AM* Social Studies blueprint includes the overall percentage of the assessment characterized by each reporting category. For grade 5 Social Studies, educators placed an emphasis on Civics and Government/History.

Content Connectors

The *I AM* Social Studies blueprint lists the code of each Content Connector in each reporting category.

Content Connector Allocations

The blueprint also specifies the minimum and maximum number of items per Content Connector. A Content Connector with a range that starts at 0 indicates that the Content Connector may not be assessed each year. The item ranges in the blueprint allow each student to experience a wide range of content while still providing flexibility during form construction.

Total Number of Operational Items

The total number of operational items on the *I AM* Social Studies assessment is 32.

4.1.6 TEST FORM ASSEMBLY

CAI ELA, Mathematics, Science, and Social Studies content teams were responsible for the initial form construction and subsequent revisions. CAI content teams performed the following tasks:

- Selection of the operational items
- Selection of the field-test items
- Revision of the operational item sets according to feedback from senior CAI content staff
- Revision of the operational item sets according to feedback from the CAI technical team
- Revision of the operational item sets according to feedback from IDOE
- Assistance in the generation of materials for IDOE review
- Revision of the forms to incorporate feedback from IDOE

The CAI technical team, which included psychometricians and statistical support associates, prepared the item bank by updating the Item Tracking System (ITS) with current item statistics and providing test construction training to the internal content team.

The technical team performed the following tasks:

- Preparing item bank statistics and updating CAI’s ITS
- Creating the master data sheets (MDS) for each grade and subject
- Providing feedback on the statistical properties of initial item selections
- Providing feedback on the statistical properties of each subsequent item selection

IDOE assessment and content specialists reviewed and approved selected items and forms provided by CAI. Feedback provided by IDOE was addressed in subsequent rounds by CAI until all *I AM* forms were approved by IDOE.

4.2 ITEM DEVELOPMENT PROCESS

All custom Indiana development followed a very similar review process. This process was managed by CAI’s ITS, which is an auditable content-development tool that enforces rigorous workflow and captures every change to, and comment about, each item. Reviewers, including internal CAI reviewers and stakeholders in committee meetings, reviewed items in ITS as they would appear to the student, with all accessibility features and tools.

4.2.1 SUMMARY OF ITEM SOURCES

Operational items used on *I AM* test forms were drawn from legacy ISTAR items and Indiana custom-developed items.

4.2.2 DEVELOPMENT OF NEW ITEMS

New items are generally developed each year to be added to the operational item pool after field testing. Several factors play into the development of new items; the item development team conducts a gap analysis for distributions of items across multiple dimensions, such as item counts, item types, item difficulty, and numbers in each strand or benchmark.

All CAI item writers who developed *I AM* items have at least a bachelor’s degree, and many bring teaching experience. All item writers are trained in

- the principles of universal design,

- the appropriate use of item types, and
- the *I AM* item specifications.

Key materials include:

- CAI’s Language Accessibility, Bias, and Sensitivity (LABS) Guidelines, which include a focus on Linguistic Complexity (Appendix 4-A);
- Indiana item specifications; and
- a training presentation (using Microsoft PowerPoint) for the appropriate use of item types.

4.3 ITEM REVIEW

During and after each operational test administration, a series of quality assurance reports is generated and used to evaluate whether operational items are performing as intended. These reports serve as a key check for the early detection of potential problems with item scoring, including incorrect designation of a keyed response or other scoring errors, as well as potential breaches of test security that may be indicated by changes in the difficulty of test items. Flagged items are reviewed by psychometricians and content experts. Details can be found in Chapter 9, Quality Assurance Procedures.

4.3.1 ITEM REVIEW PROCESSES

CAI’s *I AM* assessment development structure utilizes highly effective units of test developers organized around each content area. Unit directors oversee team leaders who work with team members to ensure item quality and adherence to best practices. All team members, including item writers, are content-area experts. Teams include senior content specialists who review items prior to client review and provide training and feedback for all content-area team members.

CAI items go through a rigorous, multiple-level Internal Review process before they are sent to External Review. Staff members are trained to review items for both content and accessibility throughout the entire process. A sample item review checklist that CAI test developers used is included in this technical report as Appendix 4-F, Item Review Checklist. The *I AM* Internal Review cycle includes five levels, including:

- Preliminary Review
- Content Review 1
- Accessibility Review

- Edit Review 1
- Senior Review 1

Preliminary Review

Items are first written independently by test developers. After items are written by test developers, the items undergo Preliminary Review. Preliminary Review is conducted by team leads or senior content staff. During the Preliminary Review process, test developers, either individually or as a group, analyze items to ensure the following:

- The item aligns with the academic standard.
- The item matches the item specifications for the skill being assessed.
- The item is based on a quality idea (i.e., it assesses something worthwhile in a reasonable way).
- The item is properly aligned to Links for Academic Learning (LAL) Depth of Knowledge (DOK) level.
- The vocabulary used in the item is appropriate for the grade and subject matter.
- The item considers language accessibility and is fair to all students.
- The content is accurate and straightforward.
- The graphic and stimulus materials are necessary to answer the question.
- The stimulus is clear, concise, and succinct (i.e., it contains enough information to make clear what is being asked, it is stated positively, and it does not rely on negatives—such as no, not, none, never—unless absolutely necessary).

At the conclusion of the Preliminary Review, items that were accepted as written or revised during this review move on to Content Review 1. Items that were rejected during this review do not move on.

Content Review 1

Content Review 1 is conducted by a senior content specialist who was not part of the Preliminary Review. This reviewer carefully examines each item based on all the criteria identified for the Preliminary Review. He or she also ensures that the revisions made during the Preliminary Review did not introduce errors or content inaccuracies. This reviewer approaches the item both from the perspective of potential clients as well as his or her own experience in test development. If substantive changes are deemed to be necessary, this reviewer rejects the item or sends the item back to a test developer with the requested changes and then reviews the item again.

Accessibility Review

During Accessibility Review, the reviewer examines and revises items to make sure they not only meet the content standards but are also as accessible as possible to students across a wide spectrum of cognitive and physical disabilities. If the accessibility reviewer has concerns about the accessibility of an item, the item gets sent back to Content Review 1 for revision.

Edit Review 1

During Edit Review 1, editors have four primary tasks.

First, editors perform basic line editing for correct spelling, punctuation, grammar, and mathematical and scientific notation, ensuring style consistency across items.

Second, editors ensure that all items are accurate in content. Editors compare reading passages against the items to make sure that all information is internally consistent across stimulus materials and items, including names, facts, or cited lines of text that appear in the item. Editors ensure the key is correct and that all information in the item is accurate. For Mathematics items, editors perform all calculations to ensure accuracy.

Third, editors review all material for fairness and language accessibility issues.

Finally, editors confirm that items reflect the accepted guidelines for good item construction. In all items, they look for language that is simple, direct, and free of ambiguity with minimal verbal difficulty. Editors confirm that a problem or task and its stem are clearly defined and concisely worded with no unnecessary information. For multiple-choice (MC) items, editors check that options are parallel (to the extent possible) in structure and fit logically and grammatically with the stem. They also confirm that the key accurately and correctly answers the question as posed, is not inappropriately obvious, and is the only correct answer to an item among the distractors.

Senior Review 1

By the time an *I AM* item arrives at Senior Review 1, it has been thoroughly vetted by both content reviewers and editors. Senior reviewers (i.e., senior content specialists) look back at the item's entire review history, ensuring that all the issues identified in that item have been adequately addressed. Senior reviewers verify the overall content of each item, confirming its accuracy, alignment to the standard, and consistency with the expectations for the highest quality.

4.3.2 COMMITTEE REVIEW OF ITEM POOL

All *I AM* items have been through an exhaustive external review process. *I AM* items in the item bank are reviewed by IDOE content experts and then reviewed again and approved by a stakeholder committee that evaluates content, accessibility, bias/fairness, and sensitivity.

State Review

After items have been developed in the *I AM* item bank, state content experts review all items prior to committee review. At this stage in the review process, states can request edits, such as wording edits, scoring edits, or alignment/DOK updates. A CAI content lead reviews and implements these requested edits and ensures the resulting items are aligned to *I AM* Content Connectors and item specifications. At this stage, items are ready for committee review.

Passage Review

For the 2018–2019 *I AM* administration, there was a separate review and acceptance process for passages that preceded item development. During the 2018 ELA Passage Review, passages were reviewed against the *I AM* Passage Specifications, which include criteria for passage quality, quantitative metrics for readability and grade-level appropriateness, accessibility, fairness, sensitivity, and bias.

Committees were designed to include two subject-matter experts, two administrators or instructional coaches, and two special education teachers or accessibility specialists. Committee members accepted passages as they appeared or recommended revisions based on a quality criteria checklist.

After the 2018–2019 *I AM* administration, IDOE and CAI agreed that content development for future *I AM* assessments would forgo passage review as a separate step preceding item development. Passage Review is important for long passages with numerous associated items to ensure that the passage is acceptable before beginning work on developing associated items. With alternate assessments, however, passages are short with typically only 3–5 associated items. It was therefore deemed more conducive to develop the passage while developing the items, which resulted in simultaneous development and review of the passages and items field-tested in the 2022–2023 *I AM* administration.

Content and Fairness Committee Review

During the Content and Fairness Committee Review, items are reviewed for content validity, grade-level appropriateness, and alignment to the content standards and item specifications. Committee members are typically grade-level and subject-matter experts or may be accessibility specialists or corporation-/school-level administrators. During this review, committee members also review the items for bias, fairness, sensitivity, and accessibility.

Committee members either accept items as they appear or recommend revisions based on a quality criteria checklist.

4.3.3 FIELD TESTING

Newly developed *I AM* items are field-tested as embedded field-test items in the *I AM* assessment. The details of field testing are described in Chapter 4.5, Item Banks, of this technical report.

Following field testing, items are subject to additional reviews. These include key verification, for items that are key-scored, and data review, for items that failed standard flagging criteria.

Each of these processes is discussed in the following sections.

Key Verification

Key verification is a simple process by which we create a frequency table of response frequencies and the scores they received. These are reviewed by qualified content staff to ensure only correct responses receive a score.

Item Data Review

Chapter 4.4, Item Statistics, describes in detail the statistical flags that send items to item data review. These flags are designed to highlight potential content weaknesses, miskeys, or possible bias issues.

I AM items that are field-tested are flagged for review in the following areas:

- Item Quality and Performance
- Item Difficulty
- Differential Item Functioning

I AM MC items are flagged for item quality and performance if the correlation for the key is less than 25% and/or if the correlation for the distractor(s) is greater than 0.

I AM MC items are flagged for item difficulty if the percentage of students selecting the key is less than 25% or greater than 95% and/or if students select an incorrect option more often than they select the key.

To evaluate DIF, CAI evaluates the likelihood of correct responses between students in different groups who were matched on ability. With fair items, students of the same ability should have the same likelihood of responding correctly, regardless of group membership. When items are flagged for DIF, groups matched on ability have different likelihoods of responding correctly based on group membership only.

CAI flagged items field-tested in the Spring 2023 *I AM* administration, and IDOE staff reviewed the item statistics. Twenty-two items were rejected during this review, and all other items were either promoted to the operational pool or flagged by IDOE to “hold for potential release.”

4.3.4 STRATEGY FOR POOL EVALUATION AND REPLENISHMENT

IDOE seeks to release items for each grade and subject each year for use in Indiana's Released Items Repository (RIR). To grow the operational pool each year, IDOE intends to develop items to be included in six field-test slots on each content-area form. The total number of items on the field-test forms on each year's assessments from which these six items will be randomly selected for any one student is based on what the anticipated student population can support in order to ensure that each field-test item is administered to at least 200 students. The current *I AM* student population supports the development and testing of 12 field-test items per year (six items each in two forms).

The general strategy for item development planning gathers information from three sources, including:

1. Characteristics of released items to be replaced
2. Characteristics of legacy items to be replaced
3. Tabulations of content coverage to identify gaps in the pool

4.4 ITEM STATISTICS

The item analyses included classical item statistics and item calibrations using the Rasch model for ELA, Mathematics, Science, and Social Studies. Classical item statistics are designed to evaluate item difficulty and the relationship of each item to the overall scale (item discrimination) and to identify items that may exhibit bias across subgroups (DIF analyses).

4.4.1 CLASSICAL STATISTICS

Classical item statistics are based on the classical test theory framework and have been widely applied to examine whether test items function as intended. A description of the statistics and the criteria for flagging and reviewing items is provided in the following subsections. All field-test items administered in Spring 2025 were MC items. The flagged items from the field tests were reviewed in the item data review.

Item Discrimination

The item discrimination index indicates the extent to which each item differentiates between those test takers who possess the skills being measured and those who do not. In general, the higher the value, the better the item can differentiate between high- and low-achieving students. The discrimination index is calculated as the correlation between the item score and the student's IRT-based ability estimate.

Item Difficulty

Items that were either extremely difficult or extremely easy were flagged for review but were not necessarily removed if they were grade-level appropriate and aligned with the test specifications. For MC items, the proportion of students in the sample selecting the correct answer (the p -value) was computed in addition to the proportion of students selecting incorrect responses.

Distractor Analysis

Distractor analysis for MC items was used to identify items that may have had marginal distractors, ambiguous correct responses, the wrong key, or more than one correct answer that attracted high-scoring students. For MC items, the correct response should have been the option most frequently selected by high-scoring students. The discrimination value of the correct response should have been substantial and positive, and the discrimination values for distractors should have been lower and, generally, negative.

The criteria used for flagging based on the classical statistics are as follows:

- Biserial correlation statistic is less than 0.25.
- Biserial correlations for MC item distractors are greater than 0.00.
- Proportion correct value is less than 0.25 or greater than 0.95.
- The proportion of students responding to a distractor exceeds the proportion responding to the keyed response.

The classical item statistics for the field-test items are presented in Appendix 4-G, Field-Test Item Classical Statistics.

4.4.2 ITEM RESPONSE THEORY STATISTICS

Item response theory (IRT; van der Linden & Hambleton, 1997) is used to calibrate all items and derive scores for all I AM items. IRT is a general framework that models test responses resulting from an interaction between students and test items.

IRT encompasses many related measurement models that allow for varied assumptions about the nature of the data. Simple unidimensional models are the most common models used in grades K–12 operational testing programs, and items are often calibrated using a sample of students from within a state population.

Calibration is the process by which the statistical relationship between student responses and the underlying measurement construct is estimated. Traditional item response models assume a single underlying trait and assume that items are independent given that underlying trait. In other words, the models assume that given the value of the underlying trait, knowing the response to one item provides no information about

responses to other items. This basic simplifying assumption allows the likelihood function of these models to take the relatively simple form of a product over items for a single student:

$$L(Z) = \prod_{j=1}^n P(z_j|\theta),$$

where Z represents the vector of item responses, and θ represents a student's true ability.

Traditional item response models differ only in the form of the function $P(Z)$. The one-parameter model (also known as the Rasch model) is used to calibrate dichotomously scored I AM items and takes the form

$$P(x_j = 1|\theta_k, b_j) = \frac{e^{(\theta_k - b_j)}}{1 + e^{(\theta_k - b_j)}} = P_{j1}(\theta_k).$$

The b parameter is often called the *location* or *difficulty* parameter; the greater the value of b , the greater the difficulty of the item. The one-parameter model assumes that the probability of a correct response approaches zero as proficiency ($\theta_k - b_j$) decreases toward negative infinity. In other words, the one-parameter model assumes that no guessing occurs. In addition, the one-parameter model assumes that all items are equally discriminating.

For items that have multiple, ordered response categories (i.e., partial credit items), I AM items are calibrated using the Rasch family Masters' (1982) partial credit model. Under Masters' model, the probability of a response in category i for an item with m_j categories can be written as

$$P(x_j = i|\theta_k, b_{j0} \dots b_{jm_j-1}) = \frac{e^{\sum_{v=0}^i (\theta_k - b_{jv})}}{\sum_{g=0}^{m_j-1} e^{\sum_{v=0}^g (\theta_k - b_{jv})}}.$$

Field-test item calibration is conducted using IRTPRO 6.0. IRTPRO implements the method of Maximum Likelihood (ML) for item parameter estimation. The item parameter estimates of the field-test items are presented in Appendix 4-H, Field-Test Item Parameters.

4.4.3 ANALYSIS OF DIFFERENTIAL ITEM FUNCTIONING

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999, 2014) provides a guideline for when sample sizes permitting subgroup differences in performance should be examined, and appropriate actions should be taken to ensure that differences in performance are not attributable to construct-irrelevant factors. To identify such potential problems, all I AM items were evaluated in terms of DIF statistics based on the analyses made before the item bank was established and also after I AM was administered in Spring 2025.

DIF analyses were performed for the following groups:

- Male/Female
- White/African American
- White/Hispanic
- Autism/Other
- Moderate and Severe Intellectual Disability/Other

DIF refers to items that appear to function differently across identifiable groups, typically across different demographic groups. Identifying DIF was important because it provided a statistical indicator that an item could contain either cultural or another type of bias. DIF-flagged items were further examined by content experts, who were asked to re-examine each flagged item to decide whether the item should have been excluded from the pool due to bias. Not all items that exhibit DIF are biased; characteristics of the education system may also lead to DIF. For example, if schools in certain areas are less likely to offer rigorous mathematics classes, students at those schools might perform more poorly on Mathematics items than would be expected, given their proficiency in other types of items. In this example, it is not the item that exhibits bias but the instruction. However, DIF can indicate bias, so all items were evaluated for DIF.

A generalized Mantel–Haenszel (MH) procedure was applied to calculate DIF. The generalizations include (1) adaptation to polytomous items and (2) improved variance estimators to render the test statistics valid under complex sample designs. In this procedure, each student's raw score on the operational items on a given test is used as the ability-matching variable. That score is divided into 10 intervals in order to compute the MH DIF statistics for balancing the stability and sensitivity of the DIF scoring category selection. The analysis program computes the MH value, the conditional odds ratio, and the MH-delta for dichotomous items; the GMH and the standardized mean difference (SMD) are computed for polytomous items.

The MH chi-square statistic (Holland & Thayer, 1988) is calculated as

$$MH\chi^2 = \frac{(|\sum_k n_{R1k} - \sum_k E(n_{R1k})| - 0.5)^2}{\sum_k var(n_{R1k})},$$

where $k = \{1, 2, \dots, K\}$ for the strata, n_{R1k} is the number of correct responses for the reference group in stratum k , and 0.5 is a continuity correction. The expected value is calculated as

$$E(n_{R1k}) = \frac{n_{+1k}n_{R+k}}{n_{++k}},$$

where n_{+1k} is the total number of correct responses, n_{R+k} is the number of students in the reference group, and n_{++k} is the number of students in stratum k . The variance is calculated as

$$var(n_{R1k}) = \frac{n_{R+k}n_{F+k}n_{+1k}n_{+0k}}{n_{++k}^2(n_{++k} - 1)},$$

where n_{F+k} is the number of students in the focal group, n_{+1k} is the number of students with correct responses, and n_{+0k} is the number of students with incorrect responses in stratum k .

The MH conditional odds ratio is calculated as

$$\alpha_{MH} = \frac{\sum_k \frac{n_{R1k}n_{F0k}}{n_{++k}}}{\sum_k \frac{n_{R0k}n_{F1k}}{n_{++k}}}.$$

The MH-delta (Δ_{MH}) (Holland & Thayer, 1988) is then defined as

$$\Delta_{MH} = -2.35 \ln(\alpha_{MH}).$$

The MH statistic generalizes itself to polytomous items (Somes, 1986) and is defined as

$$GMH\chi^2 = \left(\sum_k a_k - \sum_k E(a_k) \right)' \left(\sum_k \text{var}(a_k) \right)^{-1} \left(\sum_k a_k - \sum_k E(a_k) \right),$$

where a_k is a $(T-1) \times 1$ vector of item response scores, corresponding to the T response categories of a polytomous item (excluding one response); $E(a_k)$ and $\text{var}(a_k)$, a $(T-1) \times (T-1)$ variance matrix, are calculated analogously to the corresponding elements in $MH\chi^2$, in stratum k .

The SMD (Dorans & Schmitt, 1991) is defined as

$$SMD = \sum_k p_{FK} m_{FK} - \sum_k p_{RK} m_{RK},$$

where

$$p_{FK} = \frac{n_{F+k}}{n_{F++}}$$

is the proportion of the focal group students in stratum k ,

$$m_{FK} = \frac{1}{n_{F+k}} \left(\sum_t a_t n_{Ftk} \right)$$

is the mean item score for the focal group in stratum k , and

$$m_{RK} = \frac{1}{n_{R+k}} \left(\sum_t a_t n_{Rtk} \right)$$

is the mean item score for the reference group in stratum k .

Items were classified into three categories (A, B, or C) for DIF, ranging from no evidence of DIF to severe DIF. DIF classification rules are shown in Table 67. Items were also indicated as positive DIF (i.e., +A, +B, or +C), signifying that the item favored the focal

group (e.g., African American, Hispanic, female) or negative DIF (i.e., –A, –B, or –C), signifying that the item favored the reference group (e.g., White, male). If the DIF statistics fell into the “C” category for any group, the item showed significant DIF and was reviewed for potential content bias or differential validity, whether the DIF statistic favored the focal or the reference group. Content experts reviewed all items flagged based on DIF statistics. They were encouraged to discuss these items and were asked to decide whether each item should be excluded from the pool of potential items given its performance.

Table 68: DIF Classification Rules

Dichotomous Items	
Category	Rule
C	is significant, and 1.5.
B	is significant, and <1.5.
A	is not significant, or <1.

Because of the unreliability of the DIF statistics when calculated with small samples, caution must be used in evaluating DIF classifications for items where focal or reference groups contain fewer than 200 students (Mazor, Clauser, & Hambleton, 1992; Camilli & Shepard, 1994; Muniz, Hambleton, & Xing, 2001; Sireci & Rios, 2013). Because these sample sizes are not tenable for the alternate assessment program, CAI used a much smaller threshold ($n = 50$), which, although it may not have the power to detect real differences between subgroups, provides at least some opportunity to flag and evaluate items for possible bias. DIF summaries are provided only for field-test items and can be found in Appendix 4-I, Field-Test Item Differential Item Functioning (DIF). Only the items that met the minimum counts ($n = 50$) for both focal and reference groups were included in the DIF analysis.

4.5 ITEM BANKS

The *I AM* item pool consists of three source types: legacy operational items from the Indiana Standards Tool for Alternate Reporting (ISTAR), custom *I AM* items field-tested in 2019, and embedded field-test (EFT) items. The *I AM* item banks support a stage-adaptive assessment for ELA, Mathematics, Science, and Social Studies. Summaries of item inventories are provided in this section.

Table 68 provides the count of items, by source, available for the 2024–2025 *I AM* assessments.

Table 69: Operational Item Counts by Source

Subject and Grade	# Legacy Items	# Custom Items	Total # of Items
ELA 3	22	65	87
ELA 4	27	54	81
ELA 5	24	58	82
ELA 6	28	47	75
ELA 7	23	48	71
ELA 8	23	59	82
ELA 10	23	63	86
Mathematics 3	22	55	77
Mathematics 4	22	74	96
Mathematics 5	22	70	92
Mathematics 6	25	58	83
Mathematics 7	17	82	99
Mathematics 8	21	73	94
Mathematics 10	20	76	96
Science 4	26	61	87
Science 6	18	67	85
Biology	29	64	93
Social Studies 5	19	75	94

4.5.1 ESTABLISHING THE ITEM BANKS

ELA, Mathematics, Science, and Social Studies

To support blueprint and test design requirements as new items for the *I AM* item pool were developed and field-tested, legacy operational items that aligned to the new Indiana Content Connectors and that met *I AM* blueprint needs were retained for operational use on the 2024–2025 *I AM* assessments. Items were also evaluated and selected for alignment to the 2018 *I AM* item specifications when possible. However, because the item specifications in use when the legacy operational items were developed differ from the *I AM* item specifications, full alignment of the legacy operational items to the new *I AM* item specifications was not possible. Where possible given pool constraints, legacy operational items were replaced with custom *I AM* items for operational use to achieve better alignment of the new item specifications for *I AM* assessments.

To begin growing the *I AM* operational pool, CAI and IDOE developed new items for field testing based on blueprint needs that fully aligned to the new Content Connectors and item specifications.

CAI completed a preliminary legacy operational pool analysis in June 2018 based on metadata indicating alignment to the Indiana Academic Standards (IAS). A second analysis was completed after 2019 *I AM* testing. Based on these analyses, CAI created *I AM* item development plans and created new, custom *I AM* items that targeted the depth and breadth of coverage required by the test blueprints, with the intent to grow the item pool over time. Beginning in 2020 and through 2022, IDOE created *I AM* item development plans and worked with Indiana educators to develop additional, custom *I AM* items that were needed.

I AM field-test item development was a rigorous, structured process that engaged stakeholders at critical junctures. This process was managed by CAI's ITS, an auditable content-development tool with a built-in workflow that captures every item change and comment. When reviewers and stakeholders inspect items in ITS, they can see the items as they will appear to the student, with all accessibility features and tools available.

Item Bank Composition

Table 69 lists the ELA, Mathematics, Science, and Social Studies item types and provides a brief description of each.

Table 70: *I AM* Item Types and Descriptions

Response Type*	Description
Multiple-Choice (MC)	Student selects one correct answer from three options.
Multiple-Select (MS) (Science only)	Student selects all correct answers from several options.
Table-Match (MI) (Science only)	Student checks a box to indicate whether information in a column header matches information in a row.

Most of the *I AM* items are MC items. There are five Science items of the MS or MI item types, but none is currently in operational use, at IDOE's request.

4.5.2 ITEM BANK MAINTENANCE

ELA, Mathematics, Science, and Social Studies

To maintain the *I AM* item banks, new items are developed and field-tested in the spring administration of each year. They are then calibrated and analyzed following the procedures described in Section 4.4.2, Item Response Theory Statistics. The embedded field-test (EFT) slots (in paper-and-pencil tests) or segments (in online tests) were located with fixed positions across all subjects. The EFT items were administered by using one of the EFT blocks, which included six field-test items. For the online assessments, one of the EFT blocks was randomly administered to each of the students. For the paper-and-pencil tests, one of the EFT blocks was fixed for all students in each of the grade and subject-area tests. The field-test engine randomly sampled a field-test block for each individual test administration. This randomization ensured that (1) each item block was

seen by a representative sample of Indiana students, and (2) every item block was as likely as every other item block to appear in a class or school, minimizing clustering effects.

The Spring 2024 I AM field-test blocks contained linked legacy ISTAR items and existing field-test pool items in the I AM item bank. One EFT block was constructed for all grades and subject-area tests with the exception of Mathematics grade 10, which had different EFT blocks between the online and paper tests. Table 70 through Table 73 show the number of legacy ISTAR items and existing field-test pool items used for the EFT blocks per grade and subject.

Table 71: Number of Field-Test Items in 2024–2025, ELA

Grade	Legacy ISTAR Items	Existing Field-Test Pool Items	Total Field-Test Items
3	0	6	6
4	5	16	21
5	5	16	21
6	0	23	23
7	4	19	23
8	4	17	21
10	2	20	22

Table 72: Number of Field-Test Items in 2024–2025, Mathematics

Grade	Legacy ISTAR Items	Existing Field-Test Pool Items	Total Field-Test Items
3	6	8	14
4	1	9	10
5	0	10	10
6	4	13	17
7	2	15	17
8	5	17	22
10	0	17	17

Table 73: Number of Field-Test Items in 2024–2025, Science

Grade	Legacy ISTAR Items	Existing Field-Test Pool Items	Total Field-Test Items
4	0	12	12
6	0	9	9
Biology	2	23	25

Table 74: Number of Field-Test Items in 2024–2025, Social Studies

Grade	Legacy ISTAR Items	Existing Field-Test Pool Items	Total Field-Test Items
5	0	10	10

5. TEST ADMINISTRATION

The State of Indiana implemented a new online assessment for students with significant cognitive disabilities for operational use beginning with the 2018–2019 school year. Referred to as Indiana’s Alternate Measure (*I AM*), this assessment program replaced the Indiana Standards Tool for Alternate Reporting (ISTAR) in English/Language Arts (ELA), Mathematics, Science, and Social Studies. *I AM* is a two-stage adaptive assessment that comprises ELA and Mathematics assessments for grades 3–8 and 10, Science assessments for grades 4 and 6, a Biology end-of-course assessment, and a Social Studies assessment for grade 5.

In 2024–2025, both stages of all *I AM* tests were administered online just as they were during the first year of the administration. Standard print and large print accommodations were available for students who could not access the assessment online. Braille was offered as an accommodation for print booklets; however, very few students taking *I AM* in 2024–2025 required the braille accommodation.

As specified in Standard 6.0 of the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), assessment instruments are required to have established test administration procedures to support useful interpretations of score results. This chapter provides details on the Test Administrator (TA) training and resources, accommodations, testing procedures, and test security procedures implemented for *I AM*. Specifically, it provides the following test administration-related evidence for the validity of the assessment results:

- A description of the student population that takes the *I AM* assessment
- A description of the training and documentation provided to TAs to follow standardized administration procedures
- A description of available test accommodations intended to remove barriers that otherwise would interfere with a student’s ability to take a test
- A description of the test security process to mitigate loss, theft, and test content reproduction of any kind
- A description of Cambium Assessment Inc.’s (CAI’s) Quality Monitor (QM) system and the test irregularity investigation process to detect cheating, monitor real-time item quality, and evaluate test integrity

5.1 TESTING OPTIONS

Administering the 2024–2025 *I AM* assessments required coordination, detailed specifications, and proper training. In addition to these efforts, several individuals were involved in the administration process, from those setting up testing environments to those administering the tests. Without the proper training and coordination of these

individuals, the standardization of test administration could have been compromised. The Indiana Department of Education (IDOE) worked with CAI to develop and provide the training and documentation necessary for the successful administration of *I AM* under standardized conditions within all testing environments. The *I AM* testing window was March 31 through May 9, 2025.

The accommodations available for eligible students participating in the *I AM* assessments are described in both the *I AM Test Administrator's Manual* (TAM) (Appendix 5-A, *I AM Test Administrator's Manual* Grades 3–8 and 10) and the *Indiana Accessibility and Accommodations Information for Statewide Assessments* (Appendix 5-B, *Accessibility and Accommodations Information for Statewide Assessments*). Throughout the 2024–2025 school year, the TAM was available on the [Indiana Assessment Portal](#), and the *Indiana Accessibility and Accommodations Information for Statewide Assessments* was available on the [IDOE](#) website.

For eligible students participating in the computer-based *I AM* ELA, Mathematics, Science, and Social Studies assessments, the accommodations made available are described in the *Online Test Delivery System (TDS) User Guide* (Appendix 5-C), which was accessible before and during testing through the [Indiana Assessment Portal](#).

All students were required to take subject-specific practice test items within the operational test environment prior to taking the Spring 2025 *I AM* operational assessment. Students who were administered the paper-and-pencil *I AM* form completed the practice test items in the paper-and-pencil test booklet. The practice tests contained sample test items designed to help students become familiar with the test system's functionality, if applicable, and item types. Indiana alternate assessment students and TAs also had the opportunity to interact with released, non-secure items on a public-facing [Released Item Repository \(RIR\)](#) assessment that is available on the [Indiana Assessment Portal](#). No new *I AM* RIR tests were released for the 2024–2025 school year. *I AM* RIR tests from 2018–2019 through 2022–2023 were available on the Indiana Assessment Portal for the entire 2024–2025 school year.

I AM is a stage-adaptive assessment administered in two parts, where a student's answers in Part 1 determine the next group of items presented to the student in Part 2. The student's total score is based on performance from both parts of the assessment. Each Spring 2025 *I AM* assessment included 32 operational items that were used for scoring and six embedded field-test (EFT) items, some of which were placeholder items.

The *I AM* assessments were untimed and were delivered to students individually. Students could start and finish one part of an assessment in a single day or over the course of multiple days, if needed. TAs were advised to monitor student engagement and cognitive load and pause the test when needed.

5.1.1 ADMINISTRATIVE ROLES

Corporation Test Coordinators (CTCs), School Test Coordinators (STCs), and Test Administrators (TAs) each had specific roles and responsibilities in the online testing

systems. See the *I AM Test Administrator's Manual (TAM)* (Appendix 5-A) for their specific responsibilities before, during, and after testing.

Corporation Test Coordinators

CTCs were responsible for coordinating testing at the corporation level, ensuring that the STCs in each school were appropriately trained and aware of policies and procedures, and that they were trained to use CAI's systems.

School Test Coordinators

Before each administration, STCs and CTCs were required to verify that student eligibility was correct in the Test Information Distribution Engine (TIDE) and that any accommodations or test settings were correct. To participate in a computer-based online test, students were required to appear as eligible for that test in TIDE. See the *TIDE User Guide* (Appendix 5-D) for more information.

STCs were responsible for ensuring that testing at their schools was conducted in accordance with test security and other policies and procedures established by IDOE. STCs worked with technology coordinators to ensure that computers and devices were prepared for testing and technical issues were resolved to ensure a smooth testing experience for the students. During the testing window, STCs monitored testing progress, ensured that all students participated as appropriate, and handled testing issues as necessary by contacting the CAI Help Desk.

Test Administrators

To be certified as an *I AM* TA, educators needed to complete an online Test Administrator Certification Course and pass an associated quiz (Appendix 5-E). TAs administered the *I AM* assessment to students, as well as RIR tests, prior to the operational assessment.

TAs were also responsible for reviewing necessary user manuals and user guides to prepare the testing environment and ensuring that students did not have access to books, notes, or electronic devices. They were required to administer the *I AM* assessment following the directions found in the *I AM Test Administrator's Manual (TAM)* (Appendix 5-A) and the *I AM Online & Paper Testing Script* (Appendix 5-O). Any deviation in test administration was required to be reported by TAs to the STC, who was to report it to the CTC. Then, if necessary, the CTC was to report it to IDOE. TAs also ensured that only the resources allowed for specific tests were available and no additional resources were used during administration of the *I AM* assessments.

5.1.2 ONLINE ADMINISTRATION

The *Online Test Delivery System (TDS) User Guide* (Appendix 5-C) provided instructions for creating test sessions; monitoring sessions; verifying student information; assigning test accommodations; and starting, pausing, and submitting tests. The *Technology Guide* found on the Indiana Assessment Portal provided information about hardware, software, and network configurations to run CAI's various testing applications.

Personnel involved with statewide assessment administration played an important role in ensuring the validity of the assessment by maintaining both standardized administration conditions and test security.

Test Participation

Students with significant cognitive disabilities who met the criteria to participate in the alternate assessments, as defined by Title 20 of the Indiana Code and federal law, participated in *I AM*.

Students eligible to participate in *I AM* were required to take the assessments appropriate for the grade level/subject in which they were receiving instruction. These students represented the following groups:

- **Public School Students, including Charter School Students.** Indiana public school and charter school students who met the participation criteria to participate in the alternate assessment and were enrolled in tested grade levels/subjects were required to participate in *I AM*.
- **Private School Students.** Indiana private school students who met the participation criteria to participate in the alternate assessment and were enrolled in tested grade levels/subjects were required to participate in *I AM*.
- **Accredited Nonpublic School Students.** Indiana students who attended accredited nonpublic schools and who met the participation criteria to participate in the alternate assessment and were enrolled in tested grade levels/subjects were required to participate in *I AM*.
- **Choice School Students.** Indiana Choice school students who met the participation criteria to participate in the alternate assessment and were enrolled in tested grade levels/subjects were required to participate in *I AM*.
- **Home Education Program Students.** Students who met the participation criteria to participate in the alternate assessment and who received instruction at home and were registered appropriately with their corporation office as Home Education Program students were eligible to participate in statewide assessments. If parents or guardians identified an *I AM* assessment as a selected measure of their child's annual progress, students could participate in an *I AM* administration, as directed by the CTC.
- **English Learners (ELs).** All ELs participated in statewide assessments.
- **Students with Disabilities.** Indiana has established procedures to ensure the inclusion for testing of all public elementary and secondary school students with disabilities. Federal and state law require that all students participate in the state testing system. In Indiana, a student on an Individualized Education Program (IEP) participates under one of these four general options:
 1. Indiana Learning Evaluation Assessment Readiness Network (*ILEARN*) without accommodations

2. *ILEARN* with approved accommodations
3. *I AM* without accommodations
4. *I AM* with approved accommodations

A student's Case Conference Committee (CCC) determined, based on the criteria provided and the student's individual and unique needs, whether a student with disabilities participated in general education assessments with or without testing accommodations, or in the alternate assessment with or without accommodations. A student was eligible to participate in *I AM* in lieu of *ILEARN* if the CCC determined the student met the following criteria:

- Review of student record indicates a disability that significantly impacts intellectual functioning and adaptive behavior. Adaptive behavior is defined as essential for a person to live independently and function safely in daily life.
- The student requires extensive, repeated, individualized instruction and support that is not of a temporary nature.
- The student uses substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate, and transfer skills across multiple settings.

Scheduling Make-Up Testing and Test Completion Sessions

After a test had been paused for 20 minutes, the student could no longer view or modify responses from that testing session. Students could not view or change prior answers during a make-up session. A make-up or completion session was provided only to finish the remaining portions of the test.

Test Irregularities

On rare occasions, a non-standard situation arose during test administration. Three ways to account for irregularities were provided. Steps for dealing with test irregularities are outlined in more detail in the sections on Appeals or Appeal Requests in the *TIDE User Guide*.

- **Reset a Test.** Resetting a test eliminates all responses for a student. When that student logged in to the test again, the test would start over. Resetting could only be implemented in situations where the test could not be appropriately completed as is (e.g., two students accidentally log in to each other's test, a student requiring braille was not given the accommodation). A test could never be reset to give a student a second opportunity.
- **Grace Period Extension.** Extending a test's grace period gives a student access to his or her previous responses. This extension could be granted if a test session was interrupted unexpectedly (e.g., fire drill, lockdown). The grace period extension could not be applied if the test session ended normally or if the student was given time to review his or her answers before logging out of a test.

- **Invalidate a Test.** Tests could be invalidated when a student’s performance was not an accurate measure of his or her ability (e.g., the student cheated, used inappropriate materials). If a test was invalidated, the student was not given another opportunity to take the test. Invalidating a test required the approval of a local education agency (LEA)-level user.
- **Reopen a Test.** Reopening a test changed the test’s status from completed or reported to paused. This capability was useful if a student accidentally submitted a test before reviewing it. After the test was reopened, a student could resume testing. A test was not reopened once a student saw a score.
- **Reopen a Test Segment.** Reopening a test segment allowed a student to return to a prior segment in cases where the student moved to the next segment in error. After the test segment was reopened, a student could return to the prior segment and complete his or her work.

5.1.3 ACCOMMODATED TEST ADMINISTRATION

The *I AM* assessments make available to students three categories of assessment tools and supports, which may be embedded or non-embedded in the Test Delivery System (TDS): universal tools, designated features, and accommodations.

Universal tools are available in TDS to all students taking *I AM* assessments. Table 74 lists these features. During the tests, students must use the embedded text-to-speech feature to hear test content read aloud (unless the student is assigned a Human Reader designated feature). Students can zoom in and zoom out to increase or decrease the size of text and images, highlight items and passages (or sections of items and passages), cross out response options by using the strikethrough function, and use an online or handheld and/or adaptive calculator for all Mathematics and Science items.

Designated features, such as the ability to select an alternate background and font color, mouse pointer size and color, and text size before testing, as well as a Human Reader, are available for use by any student for whom the need has been indicated by an educator, or team of educators, with parent/guardian and student.

Accommodations are supports provided to students with disabilities enrolled in public schools with current IEPs or Section 504 Plans, as well as to students identified as ELs. All Indiana state assessments have appropriate accommodations available to make test content accessible to students with disabilities and ELs, including ELs with disabilities. The accommodations available for eligible students participating in the *I AM* assessments are described in the *I AM TAM* (Appendix 5-A), which was accessible to schools before and during testing in the [Resources](#) section of the Indiana Assessment Portal. A comprehensive list of accommodations available for eligible students with IEPs, Section 504 Plans, or Individual Learning Plans participating in online assessments is given in the *Test Information Distribution Engine (TIDE) User Guide* (Appendix 5-D) and IDOE’s Accessibility and Accommodations Information for Statewide Assessments.

5.1.4 ALLOWABLE RESOURCES FOR ONLINE TESTING

Students participating in the computer-based *I AM* were able to use the standard online testing features in the Test Delivery System (TDS). Before testing, TAs were able to select an alternate background and font color, mouse pointer size and color, and text size. During the assessments, students could zoom in and zoom out to increase or decrease the size of text and images, highlight items and passages (or sections of items and passages), cross out response options by using the strikethrough or masking function, or use the online basic Desmos calculator.

All *I AM* assessments had appropriate accommodations available to make these options accessible to students with significant cognitive disabilities who required additional accommodations, per the student's IEP. Online accommodations included permissive mode (to use assistive technology) and streamlined mode. As an accommodation, students could also participate in *I AM* by using a standard print paper-and-pencil test booklet, a large print test booklet, or a braille test booklet. During the 2024–2025 school year, UEB Contracted braille was available.

The *I AM* assessments provided three categories of assessment features to students. These included universal tools, designated features, and accommodations. Section 3.2.2, Designated Features and Accommodations, lists the allowed accommodations and the number of students who were provided with accommodations during the Spring 2025 test administration.

Table 74 provides a list of universal tools, designated features, and accommodations that were offered in the Spring 2025 administration. Universal tools are accessibility features of the TDS that are delivered either digitally (i.e., embedded) or separately (i.e., non-embedded). Designated features for *I AM* are those supports that are available for use by any student for whom the need has been indicated by an educator (or team of educators with parent/guardian and student). The *Online Test Delivery System (TDS) User Guide*, available through the [Indiana Assessment Portal](#) (and included as Appendix 5-C), provides instructions on how to access and use these features.

Table 75: Universal Tools, Designated Features, and Accommodations Available in Spring 2025

Universal Tools (for all students)	Designated Features	Accommodations (available per IEP)
<i>Embedded/Online</i>		
Online calculator for all mathematics items Online calculator for all science items Expandable passages Highlighter Masking Strikethrough Text-to-Speech (required)	Color contrast Masking Mouse pointer (size and color) Text size (zoom in and zoom out) Text-to-speech tracking	Permissive mode to use assistive technology devices Streamlined mode

Universal Tools (for all students)	Designated Features	Accommodations (available per IEP)
Zoom in and zoom out for text and graphics Line reader		
Non-Embedded		
Headphones or noise buffers to block out distractions Low-tech assistive writing instrument Preferential seating Scratch paper, including lined or graph paper Student tested individually Adaptive and/or handheld calculator for all mathematics items Adaptive and/or handheld calculator for all science items	Color acetate film for paper assessment Assistive technology to magnify/enlarge text and images Access to sound amplification Human Reader for all items including reading comprehension Special furniture or equipment for viewing test Special lighting conditions Time of day for testing altered	Alternate indication of response Bilingual word-to-word dictionary Multiplication table Paper test booklet Large print test booklet Hundreds chart Interpreter for American Sign Language Braille test booklet (Uncontracted) Read aloud to self Student provided access to own resources

IDOE also collected information about non-standard accommodation requests under a Special Requests section in TIDE. These special requests required IDOE approval.

Students participating in *I AM* who required computer-based accommodations (e.g., permissive mode) were provided the opportunity to participate in practice activities for the statewide assessments with appropriate allowable accommodations. Computer-based test settings and accommodations were required to be identified in TIDE before starting a test session. Some settings and accommodations could not be changed after a student started the test.

If a student used any accommodations during the test administration, this information was recorded by the TA in his or her required administration information.

Guidelines recommended for making accommodation decisions included the following:

- Accommodations should facilitate an accurate demonstration of what the student knows or can do.
- Accommodations should not provide the student with an unfair advantage or negate the validity of a test; accommodations must not change the underlying skills that are being measured by the test.

- Accommodations must be the same or nearly the same as those needed and used by the student in completing daily classroom instruction and routine assessment activities.
- Accommodations must be necessary for enabling the student to demonstrate knowledge, ability, skill, or mastery.

Students with disabilities not enrolled in public schools or receiving services through public school programs who required accommodations to participate in a test administration were permitted access to accommodations if the following information was provided:

- Evidence that the student had been found eligible as a student with a disability as defined by the Individuals with Disabilities Education Act (IDEA)
- Documentation that the requested accommodations had been regularly used for instruction

Available Accommodations

The TA and the STC were responsible for ensuring that arrangements for accommodations had been made before the test administration dates. IDOE provided a separate accessibility policy manual, *Indiana Assessments Policy Manual*, included as Appendix 5-F of this technical report; the current manual is available on the IDOE Assessment Website at <https://www.in.gov/doe/students/assessment/indiana-assessments-policy-manual/> as a supplement to the TAMs, for individuals involved in administering assessments to students with accommodations.

For eligible students with IEPs who participated in *I AM* paper-based assessments, the following accommodations were available:

- Standard print test booklet
- Large print test booklet
- Braille test booklet (UEB Uncontracted)

For eligible students with IEPs who participate in computer-based *I AM* assessments, a comprehensive list of accommodations is included in the *Test Information Distribution Engine (TIDE) User Guide* (Appendix 5-D of this report).

The Accessibility and Accommodations Information for Statewide Assessments provides information about the available tools, supports, and accommodations that were available to students taking the *I AM* assessments. For further information, please refer to both the *Indiana Assessments Policy Manual* (Appendix 5-F) and the *Accessibility and Accommodations Information for Statewide Assessments* (Appendix 5-B).

IDOE monitors test administration in corporations and schools to ensure that appropriate assessments, with or without accommodations, are administered for all students with disabilities and ELs and are consistent with Indiana's policies for accommodations.

5.2 TRAINING AND INFORMATION FOR TEST COORDINATORS AND ADMINISTRATORS

IDOE established and communicated to its educators and key personnel involved with the *I AM* assessment administration a clear, standardized procedure for the administration process, including giving students access to accommodations. Key personnel involved with the *I AM* administration included CTCs, Corporation Information Technology Coordinators (CITCs), STCs, and TAs. The roles and responsibilities of staff involved in testing are further detailed in Section 5.1.1, Administrative Roles.

First year *I AM* TAs were required to attend a one-hour virtual training session before administering the *I AM*. Before the Spring 2025 assessment administration, CAI collaborated with IDOE to conduct an online training session on the 2024–2025 test administration. This training session provided an overview of the alternate assessment and the online systems used during test administration. These online systems included the *I AM* Portal, the TDS, TIDE, and the Centralized Reporting System (CRS). During the training session, CAI used video vignettes, which included Indiana educators and students, to illustrate important concepts. Appendix 5-J includes the PowerPoint presentation used during each training session.

All CTCs were required to attend online training sessions hosted by IDOE to ensure assessments in their building(s) were administered with fidelity. The CTCs were then required to provide training to all TAs within their corporation. Additionally, test administration personnel were required to take the *I AM* Test Administrator (TA) Certification Course. The TA Certification Course included a short quiz at the end of the course that TAs were required to pass before being able to administer the *I AM* assessment.

IDOE conducted two Q&A sessions following the CAI presentations and prior to the *I AM* test window.

TAMs and guides were available online for school and corporation staff. The *Online Test Delivery System (TDS) User Guide* (Appendix 5-C) was designed to familiarize TAs with TDS and included tips and screen captures throughout the text. The user guide contained

- steps to take prior to accessing the system and logging in;
- navigation instructions for the TA Interface application;
- details about the Student Interface, used by students for online testing;
- instructions for using the training sites available for TAs and students; and
- information on Indiana Secure Browser features and keyboard shortcuts.

The *User Support* sections of both the *Online TDS User Guide* (Appendix 5-C) and the *TIDE User Guide* (Appendix 5-D) provide instructions to address possible technology

challenges during test administration. The CAI Help Desk collaborated with IDOE to provide support to Indiana schools as they administered the state assessment.

The *Online Test Delivery System User Guide* (Appendix 5-C) provides instructions for creating test sessions, monitoring sessions, verifying student information, assigning test accommodations, and starting, pausing, and submitting tests. The *Technology Guide*, located on the Indiana Assessment Portal, provides information about hardware, software, and network configurations to run CAI's various testing applications.

Personnel involved with statewide assessment administration played an important role in ensuring the validity of the assessment by maintaining both standardized administration conditions and test security. Their roles and responsibilities are summarized in Section 5.1.1, Administrative Roles.

5.2.1 MANUALS AND USER GUIDES

The list of webinars and training resources for the Spring 2025 *I AM* test administration is provided in this section. These materials were available online on the [Indiana Assessment Portal](#). (PDFs of these five resources have also been included in this technical report as Appendices 5-E, 5-G, 5-H, 5-I, and 5-J, respectively.)

1. **Test Administrator Certification Course:** All educators who administered the *I AM* assessment were required to complete the online TA Certification Course and quiz.
2. **Understanding Indiana's Alternate Measure (*I AM*) Webinar Module:** This online module walks Indiana educators through the new *I AM* assessments to prepare educators for the Spring 2023 assessment.
3. ***I AM* Educator & Family Brochure:** This brochure provides an overview of the new *I AM* assessment to prepare educators for the Spring 2025 assessment.
4. **Accessibility and Accommodations Implementation and Setup Module:** This online module provides information on accessibility and accommodations available for use on the *I AM* assessments.
5. **First Year Training for New *I AM* TAs Webinar:** This webinar provides important information for the administration of *I AM* for first-year *I AM* TAs.

Table 75 presents the list of available user guides and manuals related to the *I AM* administration. These materials were all available on the [Indiana Portal](#). (PDFs of these six publications have also been included in this technical report as Appendices 5-D, 5-A, 5-K, 5-L, 5-C, 5-M, 5-N, and 5-B, respectively.)

Table 76: User Guides and Manuals

Resource	Description
<i>Test Information Distribution Engine (TIDE) User Guide</i>	This user guide describes the tasks performed in TIDE for <i>I AM</i> assessments.
<i>I AM Test Administrator's Manual (TAM)</i>	This manual provides information on the policies and procedures surrounding the <i>I AM</i> assessments, as well as an overview of the specific roles and responsibilities required before, during, and after testing.
<i>Released Item Repository Quick Guide</i>	This quick guide provides an overview of how to administer the <i>I AM</i> RIR tests.
<i>Released Item Repository Scoring Guides</i>	These answer keys provide information on the items included in the RIR for each tested grade and content area.
<i>Online Test Delivery System (TDS) User Guide</i>	This user guide supports TAs who manage testing for students participating in the <i>I AM</i> RIR tests and operational tests.
<i>Centralized Reporting System (CRS) User Guide</i>	This user guide provides an overview of the different features available to educators to support viewing student scores and downloadable score data files for the <i>I AM</i> assessments.
<i>Assistive Technology Manual</i>	This manual provides an overview of the embedded and non-embedded assistive technology tools that can be used to help students with special accessibility needs complete online tests in the TDS. It includes lists of supported devices and applications for each type of assistive technology that students may need, as well as setup instructions for the assistive technologies that require additional configuration in order to work with TDS.
<i>Accessibility and Accommodations Information for Statewide Assessments</i>	The accessibility manual establishes the guidelines for the selection, administration, and evaluation of accessibility supports for instruction and assessment of all students, including students with disabilities, ELs, ELs with disabilities, and students without an identified disability or EL status.

Department Resources and Support

In addition to the resources listed in Table 75, IDOE provided the following resources for corporations:

- A weekly newsletter was distributed via email to CTCs from the IDOE Office of Assessment most Mondays. The newsletter was titled, “*I AM* Assessment Update” and included information on new announcements relevant to the *I AM* assessment, reminders of upcoming milestones, and a planning-ahead section that included important dates in the *I AM* program. The IDOE Office of Assessment contact information was also available at the end of each weekly newsletter so that corporations could contact IDOE directly with any questions.
- A weekly newsletter was distributed via email to educators from the IDOE Office of the State Superintendent of Public Instruction or the Secretary of Education every Friday. The newsletter was titled, “An Update from the Department of Education” and included information on new announcements relevant to the *I AM* assessment, as well as updates from other offices in the IDOE. Access to various

social media platforms, as well as information on accessing previous weekly updates, was also available at the end of each weekly newsletter.

- Communications via newsletter from either the Office of Assessment or the Office of the State Superintendent of Public Instruction took place on an “as needed” basis. These messages generally addressed specific issues that needed to be communicated quickly to administrators and teachers in the field or information that the IDOE wanted to ensure was clearly outlined due to its importance to the *I AM* program.
- The Office of Student Assessment required that all Corporation Test Coordinators complete an *I AM* Pretest Workshop prior to the *I AM* assessment window. The *I AM* pretest workshop was combined with *ILEARN* and *IREAD-3*. Two question and answer sessions were hosted by the *I AM* assessment specialist prior to the assessment window.
- General information about the assessments (such as dates of testing windows for all state-administered assessments) was posted on the [IDOE Office of Assessment website](#). The Accessibility and Accommodations Information for Statewide Assessments in the *I AM* Policies and Guidance section of the IDOE website was designed to address questions pertaining to accommodations and overall accessibility.
- The *Indiana Assessments Policy Manual* (Appendix 5-F) was also posted on the [IDOE Office of Assessment website](#). This manual discusses CTC and STC responsibilities regarding IDOE communication and monitoring of test administration. The manual provided guidance on students opting out of an assessment and specific categories of students; descriptions on the various roles of personnel involved in test administration; and what needs to be done before, during, and after test administration. The manual also discusses formal security and integrity training for school and corporation personnel as well as the different aspects surrounding test security.
- The *Accessibility and Accommodations Information for Statewide Assessments* (Appendix 5-B) was also posted on the [IDOE Office of Assessment website](#). This manual includes the guidelines for the selection, administration, and evaluation of accessibility supports for instruction and assessment of all students, including students with disabilities, ELs, ELs with disabilities, and students without an identified disability or EL status.

***I AM* Released Item Repository**

The *I AM* RIR is a collection of non-secure items that are available to the public via the [Indiana Assessment Portal](#) and are intended to allow students, families, and educators access to content that will be similar to what the student will encounter when taking the *I AM* assessments. The *I AM* RIR was redeployed on July 24, 2024, and remained available all year.

For the 2024–2025 school year, RIR tests from 2018–2019 through 2023–2024 were available on the Indiana Assessment portal. Due to item bank restraints, no new RIR tests were released for the 2024–2025 school year. An answer key for each applicable grade and content area accompanied all years of the RIR, which provided educators the opportunity to see how their students were performing on the assessment and where educators might focus efforts to improve student performance before the administration of the *I AM* assessment.

***I AM* Practice Test Items**

The purpose of the practice test items is to familiarize students with the system, functionality, and item types that will be on the *I AM* operational test. Historically, students taking the *I AM* on paper or online were also required to take the practice test prior to taking the operational *I AM* assessment. During the Spring 2025 administration, the required practice test items were delivered to students as the first two items of the paper-and-pencil test booklets and the online test.

The Indiana Assessment Portal provided a list of supported web browsers and their versions. CAI's TDS delivers the operational test, including the practice test items, through a secure mode of the test delivery engine.

5.3 TEST SECURITY

Test security involves maintaining the confidentiality of test questions and answers and is critical in ensuring the integrity of a test and the validity of test results. Indiana has developed an appropriate set of policies and procedures to prevent test irregularities and ensure test result integrity. These include maintaining the security of test materials, assuring adequate trainings for everyone involved in test administration, outlining appropriate incident-reporting procedures, detecting test irregularities, and planning for investigation and handling of test security violations.

All personnel who administered *I AM* assessments were required to complete the online TA Certification Course accessible through the *I AM* page of the Indiana Assessment Portal. TDS was configured so that personnel could not administer tests without first completing the TA Certification Course. Access to the course was limited to the following roles: CTC, Co-Op, CITC, NPSTC, STC, and TA.

The test security procedures for *I AM* included the following:

- Procedures to ensure security of test materials
- Procedures to investigate test irregularities
- Guidelines to determine if test invalidation was appropriate/necessary

5.3.1 STUDENT-LEVEL TESTING CONFIDENTIALITY

To support these policies and procedures, IDOE leveraged security measures within CAI systems. For example, students taking the online assessments were logged out of a test within the CAI Secure Browser after 20 minutes of inactivity.

In developing the *I AM TAM* (Appendix 5-A), IDOE and CAI ensured that all test security procedures were available to everyone involved in test administration. Each manual included protocols for reporting any deviations in test administration.

If IDOE determined that an irregularity in test administration or security occurred, it acted based upon approved procedures including, but not limited to, invalidation of student scores.

5.3.2 MAINTAINING TEST SECURITY

Before test materials were finalized, test items and performance tasks went through multiple reviews, including reviews by various committees. Maintaining security of all test content was of high priority before, during, and after committee meetings. Printed copies of items and performance task content were not provided to educator participants. Any secure materials created or distributed during the meetings were collected and destroyed following the meetings.

All test items, test materials, and student-level testing information were deemed secure and were required to be appropriately handled. Secure handling protects the integrity, validity, and confidentiality of assessment questions, prompts, and student results. Any deviation in test administration was required to be reported to protect the validity of the assessment results.

Secure handling of all test materials was required before, during, and after test administration. After any administration, initial or make-up test session, secure materials (e.g., scratch paper) were required to be returned immediately to the STC and placed in locked storage. Secure materials were never to be left unsecured and were not permitted to remain in classrooms or be removed from the school's campus overnight. Secure materials were not allowed to be discarded in the trash. In addition, any monitoring software that might have allowed test content on student workstations to be viewed or recorded on another computer or device during testing had to be disabled.

It was considered a testing security violation for authorized corporation or school personnel to fail to follow security procedures set forth by the IDOE, and no individual was permitted to do the following:

- Read, copy, share, or view the passages or test items before, during, or after testing
- Explain the passages or test items to students
- Change or otherwise interfere with student responses (print books or online) to test items

- Copy or read student responses (unless transcribing paper responses into TDS)
- Cause achievement of schools to be inaccurately measured or reported
- Use another staff member's username and/or password to access vendor systems or administer tests
- Share or post actual or paraphrased test items/content or student responses in a public forum, social media, text, or email
- Comment on test content in a public forum, social media, text, or email
- Take pictures, snapshots, or videos of assessment materials
- Deviate from the prescribed administration procedures specified in the TAM
- Score student responses on the test locally before submitting the assessment for scoring to the test contractor, as designated by IDOE
- Participate in, direct, aid, counsel, assist, encourage, or fail to report any of the acts prohibited in this section

All accommodated assessment books (regular print, large print, and braille) were treated as secure documents, and processes were in place to protect them from loss, theft, and reproduction of any kind.

If non-embedded accessibility supports are used, assessment security can become an issue when other test formats are used (e.g., regular print, large print, braille print books) or when someone other than the student is allowed to see the test (e.g., interpreter, reader, scribe). To ensure test security and confidentiality, TAs were required to keep testing materials in a secure place to prevent unauthorized access. TAs were required to maintain the confidentiality of all test content and had to refrain from sharing information or revealing test content, and returned all materials as instructed after administration.

A secure browser was required to access the online *I AM* tests. The CAI Secure Browser provided a secure environment for student testing by disabling hot keys, copy, and screen capture capabilities and preventing access to the desktop (e.g., Internet, email, other files or programs installed on school machines). Users could not access other applications from within the CAI Secure Browser, even if they knew the keystroke sequences.

Some test security considerations applied to embedded accessibility supports. For example, ensuring that only authorized personnel had access to the test and that test materials were kept confidential was critical in technology-based assessments. In addition, it was important to guarantee that students could not access any unauthorized programs, the Internet, saved data, or computer shortcuts while they were taking the assessment. In most cases, any specially required hardware devices and appropriate applications, such as switches, should have been compatible with computer-delivered assessments. Prior to testing, educators should have checked device compatibility and made appropriate adjustments, if necessary.

The CAI Secure Browser was designed to ensure test security by prohibiting access to external applications or navigation away from the test. Review Appendix A of the *Online Test Delivery System (TDS) User Guide* for further details.

5.3.3 ONLINE MANAGEMENT SYSTEM

CAI has built-in security controls in all its data stores and transmissions. Unique user identification is a requirement for all systems and interfaces. All of CAI's systems encrypt data at rest and in transit. IDOE data reside on servers at Rackspace, CAI's hosting provider. Rackspace maintains 24-hour surveillance of both the interior and exterior of its facilities. Staff at both CAI and Rackspace receive formal training in security procedures to ensure that they know the procedures and implement them properly.

Hardware firewalls and intrusion detection systems protect our networks from intrusion. CAI's systems maintain security and access logs that are regularly audited for login failures, which may indicate intrusion attempts. All of CAI's secure websites and software systems enforce role-based security models that protect individual privacy and confidentiality in a manner consistent with the Family Educational Rights and Privacy Act (FERPA).

CAI's systems implement sophisticated, configurable privacy rules that can limit access to data to only appropriately authorized personnel. CAI maintains logs of key activities and indicators, including data backup, server response time, user accounts, system events and security, and load test results.

5.3.3.1 Secure System Design

CAI has developed a custom single sign-on application that is made available in Indiana's secure portal. This application is used to support access to CAI's systems in accordance with Indiana's user ID and password policy. Authorized users can log in to Indiana's single sign-on using their current user IDs and passwords and can be redirected to CAI's portal, where they have access to CAI's secure applications such as TIDE, TDS, and CRS. Nightly backups protect the data. The server backup agents send alerts to notify system administration staff in the event of a backup error, at which time they will inspect the error to determine whether the backup was successful, or they will need to rerun the backup. The system can withstand failure of almost any component with little or no interruption of service.

CAI's hosting provider, Rackspace, has redundant power generators that can continue to operate for up to 60 hours without refueling. With multiple refueling contracts in place, these generators can operate indefinitely. Rackspace partners with nine different network providers, providing multiple, redundant data routes. Every installation is served by multiple servers, any one of which can take over for an individual test upon failure of another.

CAI's architecture ensures data are recoverable at all times. Each disk array is internally redundant, with multiple disks containing each data element. Immediate recovery from failure of any individual disk is performed by accessing the redundant data on another

disk. CAI maintains support and maintenance agreements through our hosting provider for all hardware used by our systems.

5.3.3.2 System Security Components

CAI has built-in security controls in all its data stores and transmissions. Unique user identification is a requirement for all systems and interfaces. All of CAI's systems encrypt data at rest and in transit.

Physical Security

Indiana data reside on servers at Rackspace, CAI's hosting provider. Rackspace maintains 24-hour surveillance of both the interior and exterior of its facilities. All access is keycard controlled, and sensitive areas require biometric scanning.

Secure data are processed at CAI facilities and are accessed from CAI machines. CAI's servers are housed in a secure, climate-controlled location with access codes required for entry. Access to our servers is limited to our network engineers, all of whom, like all CAI employees, have undergone rigorous background checks.

Staff at both CAI and Rackspace receive formal training in security procedures to ensure that they know the procedures and implement them properly. CAI and Rackspace protect data from accidental loss through redundant storage, backup procedures, and secure off-site storage.

Network Security

Hardware firewalls and intrusion detection systems protect our networks from intrusion. They are installed and configured to prevent access to services other than hypertext transfer protocol secure (HTTPS) for our secure sites.

CAI's systems maintain security and access logs that are regularly audited for login failures, which may indicate intrusion attempts.

Software Security

All of CAI's secure websites and software systems enforce role-based security models that protect individual privacy and confidentiality in a manner consistent with Indiana's privacy laws, FERPA, and other federal laws.

CAI's systems implement sophisticated, configurable privacy rules that can limit access to data to only appropriately authorized personnel. Different states interpret FERPA differently, and our system is designed to support these interpretations flexibly. CAI has worked with IDOE to maintain data security according to its specifications.

CAI maintains logs of key activities and indicators, including data backup, server response time, user accounts, system events and security, and load test results. In addition, CAI runs automated functional tests of our TDS every morning, and logs from these runs are available for at least one week from the time of the run.

CAI psychometricians monitor the quality and performance of test administrations statewide through a series of quality assurance (QA) reports. The QA reports provide information on item behavior, blueprint match rates, and item exposure rates, and also provide cheating analysis reports.

5.4 TRACKING AND RESOLVING TEST IRREGULARITIES

Throughout the testing window, TAs were required to report breaches of protocol and testing irregularities to the appropriate STC, who was responsible for relaying the report to IDOE. Table 76 provides examples of test irregularities and security violations. Online test invalidation requests were submitted, as appropriate, through the *Test Irregularities* module under *Administering Tests* in CAI's TIDE.

CAI's QM system gathered data used to detect irregularities, monitored real-time item function, and evaluated test integrity. Every completed test ran through the QM system, and any anomalies (such as unscored or missing items, unexpected test lengths, or other unlikely issues) were flagged. Immediate notification went to CAI psychometricians and the project team through quality assurance (QA) reports. The forensic analysis report from the QM system flagged unlikely patterns of behavior in testing administrations aggregated at the following levels: test administration, TA, and school.

CAI psychometricians were able to monitor testing anomalies throughout the testing window. A variety of evidence was collected for the evaluation. These included unusual changes in test scores across administrations, much shorter or longer item response times as compared to the state average, and item response patterns using the person-fit index. The flagging criteria used for these analyses were configurable and could be changed by the user. The analyses used to detect the testing anomalies could be run anytime within the testing window.

If any unexpected results were identified, the lead psychometrician alerted the project manager immediately to resolve any issues.

Table 77: Examples of Test Irregularities and Test Security Violations

Description
Student(s) making distracting gestures/sounds or talking during the test session that creates a disruption in the test session for other students
Student(s) leaving the test room without authorization
TA or Test Coordinator leaving related instructional materials on the walls in the testing room
Student(s) accessing or using unauthorized electronic equipment (e.g., cell phones, smart watches, iPods, electronic translators) during testing
Disruptions to a test session such as a fire drill, school-wide power outage, earthquake, or other acts
TA or Test Coordinator failing to ensure administration and supervision of the assessments by qualified, trained personnel
TA giving incorrect instructions
TA or Test Coordinator giving out his or her username/password (via email or otherwise), including to other authorized users
TA allowing students to continue testing beyond the close of the testing window
TA or teacher coaching or providing any other type of assistance to students that may affect their responses. This includes both verbal cues (e.g., interpreting, explaining, paraphrasing the test items or prompts) and nonverbal cues (e.g., voice inflection, pointing, nodding head) to the correct answer. This also includes leading students through instructional strategies such as think-aloud or reminding students of a recent lesson on a topic.
TA providing students with unallowable materials or devices during test administration or allowing inappropriate designated features and/or accommodations during test administration
TA providing a student access to another student's work/responses
TA or Test Coordinator modifying student responses or records at any time
TA using another staff member's username and/or password to access vendor systems or administer tests
TA using a student's login information to access operational tests, when testing is not taking place and the student is not present

6. SCALING AND EQUATING

6.1 ITEM RESPONSE THEORY PROCEDURES

6.1.1 CALIBRATION OF I AM ITEM BANKS

The embedded field-test design, in conjunction with the stage administration of operational tests, produces item response data in a sparse data matrix. The items in the sparse data matrix were concurrently calibrated by grade and content area, with parameter estimates for operational items fixed to their bank values and field-test items calibrated under that constraint. The field-test items are calibrated using the IRTPRO software, version 6.0. In each calibration, the parameters of the operational items were fixed to their bank values, and the item parameters of the field-test items, as well as the mean and variance of each group, were estimated.

6.1.2 ESTIMATING STUDENT ABILITY USING MAXIMUM LIKELIHOOD ESTIMATION

6.1.2.1 Maximum Likelihood Estimation

The I AM assessments are scored using maximum likelihood estimation (MLE). MLEs are useful since an estimate of a person's ability can be obtained after one item has been answered correctly and one item has been answered incorrectly. With number-correct scoring, the test must be completed before an assessment of ability can be computed. This “early” estimate of ability is what allows tests to be adaptive.

However, when all the items administered at a specific point in the test have been answered correctly or incorrectly, the estimate of ability goes to positive or negative infinity, respectively, or the highest or lowest score. This has implications for determining what constitutes a completed test. Theoretically, with maximum likelihood scoring, the student could answer the first item correctly, quit the test, and receive the maximum score. To avoid this, the definition for a complete test needs to be based on something in addition to a minimum number of items attempted, as is often the case with number-correct scored tests.

Ability estimates were generated using pattern scoring, a method that scores students depending on how they answer individual items.

The likelihood function for generating the MLEs is based on a mixture of item types, including multiple-choice (MC, typically worth one point) and non-MC (often worth more than one point but scored for integer partial credit), and can therefore be expressed as

$$L(\theta) = L(\theta)^{MC} L(\theta)^{CR},$$

where

$$L(\theta)^{MC} = \prod_{i=1}^N \left[\frac{1}{1 + \exp[-D(\theta - b_i)]} \right]^{x_i} \left[1 + \frac{1}{1 + \exp[-D(\theta - b_i)]} \right]^{1-x_i}$$

and

$$L(\theta)^{CR} = \prod_{i=1}^N \frac{\exp \sum_{k=1}^{x_i} D(\theta - \delta_{ki})}{\sum_{j=1}^{m_i} \exp \sum_{k=1}^j D(\theta - \delta_{ki})},$$

where b_i is the location (i.e., difficulty) parameter, x_i is the observed response to the item, i indexes item, δ_{ki} is the k^{th} step for item i with m total categories, and D is the scaling constant equal to 1.

We subsequently find the optimal point to maximize the log-likelihood as the student's theta (i.e., MLE) given the set of items administered to the student.

6.1.2.2 Derivatives

Finding the MLE requires an iterative method, such as Newton–Raphson iterations. Because the log-likelihood is a monotonic function of the likelihood, the following derivatives based on the log-likelihood function (with Rasch constraints) are used:

$$\begin{aligned} \frac{\partial \ln L(\theta)^{MC}}{\partial \theta} &= \sum_{i=1}^N \left\{ x_i - \left[\frac{1}{1 + \exp[-(\theta - b_i)]} \right] \right\} \\ \frac{\partial \ln L(\theta)^{CR}}{\partial \theta} &= \sum_{i=1}^N \left\{ x_i - \left[\frac{\sum_{j=1}^{m_i} j \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right] \right\} \\ \frac{\partial^2 \ln L(\theta)^{MC}}{\partial \theta^2} &= - \sum_{i=1}^N \left(1 - \left[\frac{1}{1 + \exp[-(\theta - b_i)]} \right] \right) \left[\frac{1}{1 + \exp[-(\theta - b_i)]} \right] \\ \frac{\partial^2 \ln L(\theta)^{CR}}{\partial \theta^2} &= \sum_{i=1}^N \left[\left(\frac{\sum_{j=1}^{m_i} j \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right)^2 - \left(\frac{\sum_{j=1}^{m_i} j^2 \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right) \right] \end{aligned}$$

Hence, the estimated MLE is found via the following maximization routine:

$$\theta_{t+1} = \theta_t - \frac{\frac{\partial \ln L(\theta_t)}{\partial \theta_t}}{\frac{\partial^2 \ln L(\theta_t)}{\partial \theta_t^2}},$$

where

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial \ln L(\theta)^{MC}}{\partial \theta} + \frac{\partial \ln L(\theta)^{CR}}{\partial \theta},$$

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} = \frac{\partial^2 \ln L(\theta)^{MC}}{\partial \theta^2} + \frac{\partial^2 \ln L(\theta)^{CR}}{\partial \theta^2},$$

and where θ_t denotes the estimated θ at iteration t .

6.1.2.3 Standard Errors of Measurement

The standard error of the MLE is estimated by

$$se(\hat{\theta}) = \frac{1}{\sqrt{TIF(\hat{\theta}_s)}},$$

where $TIF(\hat{\theta}_s)$ is the test information for student s . The test information is calculated as

$TIF(\theta^2) = -\frac{d^2 l(\theta)}{d\theta^2}$ where $\frac{d^2 l(\theta)}{d\theta^2}$ is defined in the previous section on derivatives. Note that the calculation of the standard error of estimate depends on the unique set of items that each student answers and their estimate of θ . Different students have different SEMs, even if they have the same raw score and/or theta estimate.

6.1.3 CALIBRATING FIELD-TEST ITEMS ONTO THE I AM SCALE

Following the Spring 2019 I AM assessments, item response theory (IRT) calibrations and linking were completed that placed all items within a grade and subject on the same scale. For the calibrations of the Spring 2025 field-test items, the operational items, excluding the linked legacy items, were anchored to their bank values, and field-test item parameters were estimated. Table 77 displays the total number of students contributing to the calibration and the average sample size per item. The number of field-test items calibrated and the item parameter five-point summary and range are provided in Appendix 4-H, Field-Test Item Parameters.

Table 78: Number of Students Used in Field-Test Calibrations

Subject	Grade	Total Number of Students Used	Mean Sample Size per Item
ELA	3	813	803
ELA	4	895	889
ELA	5	919	911
ELA	6	939	935
ELA	7	943	941
ELA	8	975	970
ELA	10	1086	1080
Mathematics	3	808	797
Mathematics	4	889	884

Subject	Grade	Total Number of Students Used	Mean Sample Size per Item
Mathematics	5	911	906
Mathematics	6	935	933
Mathematics	7	940	937
Mathematics	8	974	970
Mathematics	10	1074	1072
Science	4	884	878
Science	6	930	929
Science	Biology	1106	1103
Social Studies	5	905	902

In Spring 2025, all assessments were pre-equated. The IRT statistical properties of the Spring 2025 / AM operational test forms are summarized in Tables 78–81.

Table 79: Operational Item Parameter Five-Point Summary and Range: ELA

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
3	44	-1.06	-0.78	-0.31	0.06	0.27	0.50	0.58
4	46	-1.18	-0.90	-0.41	0.02	0.17	0.31	0.63
5	45	-1.77	-0.63	-0.38	-0.07	0.30	0.96	1.44
6	44	-1.05	-0.73	-0.30	0.00	0.28	0.69	0.85
7	45	-1.53	-1.07	-0.28	0.23	0.57	1.04	1.21
8	45	-1.55	-1.01	-0.48	0.00	0.43	0.98	1.43
10	45	-1.54	-1.24	-0.46	0.03	0.54	0.93	1.21

Table 80: Operational Item Parameter Five-Point Summary and Range: Mathematics

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
3	44	-0.66	-0.58	-0.30	-0.01	0.17	0.62	0.97
4	44	-1.00	-0.75	-0.35	0.00	0.46	0.98	1.20
5	43	-1.68	-0.74	-0.38	-0.02	0.34	0.72	0.80
6	44	-0.98	-0.84	-0.45	-0.07	0.34	0.63	1.06
7	44	-1.64	-0.69	-0.35	-0.12	0.36	0.64	0.93

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
8	44	-1.03	-0.68	-0.30	-0.15	0.13	0.68	1.06
10	44	-1.20	-1.02	-0.47	-0.13	0.34	0.68	1.04

Table 81: Operational Item Parameter Five-Point Summary and Range: Science

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
4	44	-1.51	-1.00	-0.37	0.00	0.33	0.95	1.26
6	45	-2.23	-1.13	-0.40	-0.08	0.25	0.68	0.88
Biology	44	-1.37	-1.12	-0.68	-0.01	0.33	1.02	1.09

Table 82: Operational Item Parameter Five-Point Summary and Range: Social Studies

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
5	44	-1.75	-0.91	-0.33	-0.10	0.31	0.58	0.71

6.2 I AM REPORTING SCALE (SCALE SCORES)

6.2.1 OVERALL PERFORMANCE

For the Spring 2025 administration, the *I AM* scale scores were reported for each student who took the ELA, Mathematics, Science, and Social Studies assessments. The scale scores were based on the operational items presented to the student and did not include any field-test items.

The scale score is the linear transformation of the item response theory (IRT) ability estimate using the scaling constants a and b shown in Table 82:

$$SS = a * \theta + b$$

Scale scores are reported and compared as integers, with their decimal digits rounded down.

Table 83: Scaling Constants on the Reporting Metric

Subject	Grades	Slope (a)	Intercept (b)
ELA	3–8 & 10	50	1500
Mathematics	3–8 & 10	50	2500
Science	4, 6, & Biology	50	3500
Social Studies	5	50	4500

Summaries of the *I AM* scale scores for each test by demographic groups, as well as for all students, are provided in Appendix 3-A, Distribution of Scale Scores and Standard Deviations.

6.2.2 REPORTING CATEGORY PERFORMANCE

For reporting categories, the classification indicator of the performance level is reported for each student at the reporting category level and for aggregate reporting.

Theta scores of each reporting category were calculated using MLE based on the items contained in a reporting category. The transformed scale score and standard error of measurement (SEM) were used for determining the classification of reporting category scores. The same rules for scoring all correct and all incorrect cases were applied to reporting category scores. The difference between the proficiency cut score and the reporting category score plus or minus one SEM of the reporting category is used to determine the student's relative strengths and weaknesses within the reporting category. The specific rules for mastery are as follows:

- Below Proficiency: if $\text{round}(SS_{rc} + 1 * SE(SS_{rc}), 0) < SS_p$
- Near Proficiency: if $\text{round}(SS_{rc} + 1 * SE(SS_{rc}), 0) \geq SS_p$ and $\text{round}(SS_{rc} - 1 * SE(SS_{rc}), 0) < SS_p$, a strength or weakness is indeterminable
- At Proficiency: if $\text{round}(SS_{rc} - 1 * SE(SS_{rc}), 0) \geq SS_p$

where SS_{rc} is the student's scale score on a reporting category; SS_p is the proficiency scale score cut (Level 3 cut); and $SE(SS_{rc})$ is the standard error of the student's scale score on the reporting category. The round function (i.e., $\text{round}(SS_{rc} + 1 * SE(SS_{rc}), 0)$) in the classification rules indicates that the values calculated from scale score and SEM are rounded down to the integers, which is the same with the overall scale score transformation.

Summaries of the scores for each reporting category by demographic groups as well as for all students are provided in Appendix 3-C, Distribution of Reporting Category Scores by Subgroup.

6.2.3 RULES FOR ZERO AND PERFECT SCORES

In IRT maximum likelihood ability estimation methods, zero and perfect scores are assigned the ability of minus and plus infinity. For all the tests, the extreme student ability estimates are truncated to the lowest observable scores (LOT/LOSS) or the highest observable scores (HOT/HOSS). Note that LOSS = lowest observable scale score and HOSS = highest observable scale score; LOT = lowest observable theta and HOT = highest observable theta. Estimated theta values lower than the LOT or higher than the HOT will be truncated to the LOT and HOT values and will be assigned the LOSS and HOSS associated with the LOT and HOT. For *I AM* scoring, extreme cases were handled according to the following guidelines:

- Score all incorrect and all correct cases by either adding or subtracting 0.3 to/from an item score.
- Generate MLE for every other case and apply the following rule:
 - a. If MLE is lower than -4 , assign theta to -4 .
 - b. If MLE is higher than 4 , assign theta to 4 .

Table 83 gives the LOT, LOSS, HOT, and HOSS for the *I AM* assessments. The standard error for LOT and HOT was computed using the LOT and HOT ability estimates derived from the administered items. For example, in the formula discussed in Section 6.1.2.3, Standard Errors of Measurement, $\hat{\theta}$ = LOT or HOT, and difficulties (b) are for the administered items.

Table 84: Theta and Scaled Score Limits for Extreme Ability Estimates

Subject	Grade	Lowest-Observable Theta (LOT)	Highest-Observable Theta (HOT)	Lowest-Observable Scale Score (LOSS)	Highest-Observable Scale Score (HOSS)
ELA	3–8 & 10	–4	4	1300	1700
Mathematics	3–8 & 10	–4	4	2300	2700
Science	4, 6, & Biology	–4	4	3300	3700
Social Studies	5	–4	4	4300	4700

6.2.4 RULES FOR SCORING AND REPORTING OF INCOMPLETE TEST ADMINISTRATIONS

Reporting for each of the subject-area test administrations (ELA, Mathematics, Science, and Social Studies) is based both on an attemptedness criterion and on whether the test administration is completed. All operational items are included in the evaluation of test records for attemptedness, or whether students attempted or completed a test. Field-test items are excluded.

The attemptedness flag in the student data file includes four values: P (UND: Undetermined), E (NMC: No Mode of Communication), Y (Attempted), and N (Invalidated). Students who do not complete the first five questions in Segments 1 and 2 are assigned as UND. Students who complete the first five questions in Segments 1 and 2 but have No Response (NR) for those five questions are assigned as NMC. In this case, TDS provides the pop-up message to show students are identified as NMC and stop their tests. For students who complete the first five questions and have NR in fewer than five items in Segments 1 and 2, the test is counted as “attempted.” Attempted tests will be scored. For the tests attempted, if an operational item in a Part 2 Form is taken, items without responses in the routed form will be scored as ‘0’. If no operational item in the Part 2 Form is taken, items without responses in Part 1 and all items in Form A, the easiest form, will be scored as ‘0’. Items with “No Response” will be scored as ‘0’. Tests that are invalidated are assigned as N, and the score report shows Invalidated.

7. PERFORMANCE STANDARDS

The first operational administration of the *I AM* assessments took place in Spring 2019 for all grades and subjects. Following the close of the test administration windows, one hundred educators from Indiana convened at the Sheraton Indianapolis Hotel at Keystone Crossing in Indianapolis, Indiana, from July 22 through 24 of that year, with the purpose of completing three rounds of standard setting to recommend two performance standards (cut scores) for the *I AM* assessments in each content area.

This chapter briefly describes the procedures used by educators to recommend standards and resulting proficiency standards. Details of the panels, procedures, and outcomes are documented in the Spring 2019 *I AM* technical report.

7.1 STANDARD-SETTING PROCEDURES

Student achievement on *I AM* is classified into three performance levels: Below Proficiency, Approaching Proficiency, and At Proficiency. Interpretation of the *I AM* test scores rests fundamentally on how test scores relate to proficiency standards that define the extent to which students have achieved the expectations identified in the Indiana Academic Standards. The cut score establishing the Proficient level of performance is the most critical because it indicates that students are meeting grade-level expectations for achievement of Indiana's Alternate Academic Standards, that they are prepared to benefit from instruction at the next grade level, and that they are on track to pursue post-secondary education or enter the workforce. Procedures used to adopt proficiency standards for the *I AM* assessments are therefore central to the validity of test score interpretations.

Procedures

Following the first operational administration of the *I AM* assessments in Spring 2019, a standard-setting workshop was conducted to recommend to the State Board of Education (SBOE) a set of proficiency standards for reporting student achievement of the Indiana Academic Standards. The workshop consisted of a series of standardized and rigorous procedures that the Indiana educators serving as standard-setting panelists followed to recommend proficiency standards. The workshops employed the Bookmark procedure, a widely used method where standard-setting panelists used their expert knowledge of Indiana's Alternate Academic Standards and student achievement to map the Performance-Level Descriptors (PLDs) adopted by the IDOE to an ordered-item booklet (OIB) based on the first operational test form administered. The Bookmark procedure was implemented in three rounds, providing panelists with feedback and benchmark information prior to Round 2, and panelist feedback, benchmark, and impact data prior to Round 3.

Following discussion of panelist feedback, panelists were presented with benchmark data, performance standards comparable to other important assessment systems, and a

multistate assessment (created by the National Center and State Collaborative [NCSC]) of students with significant intellectual disabilities. The IDOE’s policy committee also recommended that the performance standards for the alternate assessment be considered in relationship to the performance standards for the general education assessment for the general population (the Indiana Learning Evaluation Assessment Readiness Network [ILEARN]). To facilitate comparisons of Indiana performance standards with other national benchmarks, panelists were provided with the locations of performance standards from these other assessment systems in their OIBs. In particular, performance standard locations for the following assessments were provided as part of the panelists’ OIB review:

- NCSC ELA and Mathematics performance standards in grades 3–8 and 10
- ILEARN performance standards in ELA and Mathematics in grades 3–8, Science in grades 4, 6, and Biology and Social Studies in grade 5

When panelists can use benchmark information to locate proficiency standards that converge across assessment systems, the validity of test score interpretations is bolstered.

Panelists were also provided with feedback about the vertical articulation of their recommended proficiency standards so that they could view how the locations of their recommended cut scores for each grade-level assessment related to the cut-score recommendations at other grade levels. This approach allowed panelists to view their cut-score recommendations as a coherent system of proficiency standards and further reinforced the interpretation of test scores as indicating not only achievement of current grade-level standards but also preparedness to benefit from instruction in the subsequent grade level.

Performance-Level Descriptors (PLDs)

A prerequisite to standard setting is to determine the nature of the categories, or performance levels, into which students are classified. The three performance-level categories for the *I AM* are “Below Proficiency,” “Approaching Proficiency,” and “At Proficiency.” These categories, or performance levels, are associated with PLDs. PLDs define the content-area knowledge and skills that students at each performance level are expected to demonstrate. PLDs link the assessment content to the IAS. There are multiple types of PLDs (Egan, Schneider, & Ferrara, 2012), including the following:

1. *Policy PLDs*: Policy PLDs articulate the overall claims about a student’s performance at each performance level. Policy PLDs are used by policymakers to broadly articulate the goals and rigor for the state’s performance standards. The *I AM* Policy PLDs 2018–2019 can be found [here](#).
2. *Range PLDs*: A description of what students should know and be able to do throughout the range of each performance level. For example, the Range PLD for Approaching Proficiency describes what students know and can do at that level all the way to just below the At Proficiency cut score. The Range PLDs for the *I AM* can also be found [here](#).
3. *Target PLDs*: Sometimes called “Threshold” or “Just Barely” PLDs, these are created during the standard-setting workshop and are used only for standard setting. Target PLDs

describe what a student just barely scoring at the entry point of each performance level knows and can do.

On July 25, 2018, the IDOE worked with the seven-person Indiana stakeholder panel to make recommendations for *I AM* Policy PLDs. The IDOE led the *I AM* Policy PLD meeting, and CAI (formerly the American Institutes for Research [AIR]) staff were present at the meeting in the role of notetakers to document the process and the committee wording for the Policy PLDs. Policy PLDs define, at a broad policy level, the goals and rigor of the *I AM* assessment. The IDOE provided panelists with background on the *I AM* development process and on the purpose and role of PLDs within the assessment system. The IDOE discussed example PLDs from national and state alternate assessments, including NCSC, Dynamic Learning Maps (DLM), and several states. During the Policy PLD meeting, the panel drafted the following Policy PLDs: Below Proficiency, Approaching Proficiency, and At Proficiency.

On September 11–13, 2018, Indiana educators convened to develop the Range PLDs for each content area and grade level included in the *I AM* assessments. During the meeting, educators reviewed Policy PLDs and created Range PLDs. With the goal of reinforcing the alignment to *ILEARN* and ensuring a cohesive system of assessments, the IDOE invited the same policy panel that met on May 15, 2018, to develop *ILEARN* Policy PLDs to the extent possible. The goal of the *I AM* PLD meeting was to connect the content of the general assessment to the content of the alternate assessment for students with significant cognitive disabilities. The PLDs describe student performance at the following levels: Below Proficiency, Approaching Proficiency, and At Proficiency.

Participants in the standard-setting workshop primarily worked with the Range PLDs and Target PLDs.

Panelists used the PLDs to develop a representation of students who are “just barely” described by each of the PLDs. During this training task, panelists learned that while PLDs are written to characterize typical members of each performance level, their bookmark placements would be directed toward characterizing and identifying the most minimally qualified members of each performance level. Characterizing a student as “just barely” meeting the performance standard is not an intuitive judgment, and panelists worked to identify the minimum characteristics of student achievement for entry into each performance level. Each panel produced a “just barely” PLD to help guide their discussions and bookmark placements. To develop a common understanding among panelists, each panel was asked to do the following:

- Review and parse PLDs
- Discuss characteristics of students classified near thresholds of performance standards
- Identify the characteristics that distinguish students “just above” the performance standard from those “just below”
- Determine what evidence was necessary to conclude that a student possessed the minimum knowledge and skills needed to meet the performance standard

- Summarize knowledge and skills of students who “just barely” meet each performance standard, or are “just barely” described by each PLD

These discussions yielded common descriptions of students “just barely” characterized by each PLD within each room.

7.2 RECOMMENDED PROFICIENCY STANDARDS

Panelists were tasked with recommending two proficiency standards (Approaching Proficient and Proficient) that resulted in three performance levels (Below Proficiency, Approaching Proficiency, and At Proficiency). As panelists discussed the reasons for their bookmark placements in the context of feedback from other panelists and impact data, variability often decreased across rounds. In general, there was considerable consistency in the placement of performance standards across rounds.

The final recommended performance standards for each assessment, grade, and performance standard are presented in Table 84 along with the projected impact each performance standard would have on Indiana public school students tested in 2019. The final recommended OIB page numbers are the median bookmarks of each panel following Round 3 bookmark placement, and subsequent moderation.

Following the standard-setting workshop, panelist recommendations were submitted to IDOE; IDOE formally adopted the standards in July 2019.

Table 85: Final Recommended Performance Standards

Grade	Performance Level	OIB Page	RP50	Estimated Percentage of Students At or Above Performance Standard
ELA 3	Approaching Proficiency	7	-0.72	60%
	At Proficiency	12	-0.37	45%
ELA 4	Approaching Proficiency	13	-0.42	60%
	At Proficiency	20	-0.05	45%
ELA 5	Approaching Proficiency	11	-0.51	65%
	At Proficiency	17	-0.21	51%
ELA 6	Approaching Proficiency	9	-0.67	65%
	At Proficiency	16	-0.26	50%
ELA 7	Approaching Proficiency	10	-0.28	63%
	At Proficiency	18	-0.04	50%
ELA 8	Approaching Proficiency	11	-0.71	71%
	At Proficiency	19	-0.18	49%
ELA 10	Approaching Proficiency	13	-0.64	79%
	At Proficiency	27	0.12	49%

Grade	Performance Level	OIB Page	RP50	Estimated Percentage of Students At or Above Performance Standard
Mathematics 3	Approaching Proficiency	6	-0.75	71%
	At Proficiency	10	-0.52	59%
Mathematics 4	Approaching Proficiency	7	-0.76	68%
	At Proficiency	12	-0.42	48%
Mathematics 5	Approaching Proficiency	6	-0.81	66%
	At Proficiency	10	-0.58	48%
Mathematics 6	Approaching Proficiency	8	-0.75	66%
	At Proficiency	14	-0.43	47%
Mathematics 7	Approaching Proficiency	8	-0.65	59%
	At Proficiency	11	-0.45	47%
Mathematics 8	Approaching Proficiency	6	-0.71	55%
	At Proficiency	10	-0.50	42%
Mathematics 10	Approaching Proficiency	8	-0.58	55%
	At Proficiency	16	-0.29	32%
Science 4	Approaching Proficiency	12	-0.49	57%
	At Proficiency	19	-0.07	41%
Science 6	Approaching Proficiency	11	-0.69	71%
	At Proficiency	19	-0.21	48%
Biology	Approaching Proficiency	15	-0.55	67%
	At Proficiency	22	0.06	43%
Social Studies 5	Approaching Proficiency	13	-0.22	41%
	At Proficiency	17	-0.01	35%

Table 85 shows the estimated percentage of students classified at each performance level based on final panelist-recommended standards for the overall student population across grade levels and courses.

Table 86: Percentage of Students at Each Performance Level Based on Final Recommended Performance Standards

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
ELA 3	40%	16%	45%
ELA 4	40%	14%	45%
ELA 5	35%	14%	51%
ELA 6	35%	16%	50%

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
ELA 7	37%	14%	50%
ELA 8	29%	22%	49%
ELA 10	21%	30%	49%
Mathematics 3	29%	12%	59%
Mathematics 4	32%	21%	48%
Mathematics 5	34%	19%	48%
Mathematics 6	34%	19%	47%
Mathematics 7	41%	12%	47%
Mathematics 8	45%	14%	42%
Mathematics 10	45%	22%	32%
Science 4	43%	15%	41%
Science 6	29%	23%	48%
Biology	33%	24%	43%
Social Studies 5	59%	6%	35%

Table 86 shows the estimated percentage of students meeting the *I AM* proficient standard for each assessment in Spring 2019. It also shows the national percentages of students that meet the NCSC and *ILEARN* proficient standards. Since NCSC is only delivered in ELA and mathematics, the percentages in science and social studies were not provided. As Table 86 indicates, the performance standards recommended for *I AM* assessments are consistent with relevant NCSC and *ILEARN* proficient benchmarks.

Table 87: Estimated Percentage of Students Meeting *I AM* and Benchmark Proficient Standards

Grade	<i>I AM</i>	NCSC	<i>ILEARN</i>
ELA 3	45	51	46
ELA 4	45	56	45
ELA 5	51	58	47
ELA 6	50	63	47
ELA 7	50	56	49
ELA 8	49	64	50
ELA 10	49	70	50*
Mathematics 3	59	73	58
Mathematics 4	48	53	53
Mathematics 5	48	57	47
Mathematics 6	47	58	46

Grade	<i>I AM</i>	NCSC	<i>ILEARN</i>
Mathematics 7	47	68	41
Mathematics 8	42	61	37
Mathematics 10	32	57	37*
Science 4	41		46
Science 6	48		47
Biology	43		39
Social Studies 5	35		45

*Because *ILEARN* was not administered in grade 10, the grade 10 benchmarking activities used the data from the *ILEARN* grade 8.

ELA, Mathematics, Science, and Social Studies assessments were reported on a separate within-test scale. Applying the *I AM* scale score transformations to the performance standards recommended by the workshop panels results in the system of scale score ranges for each of the *I AM* performance-level classifications identified in Table 87.

Table 88: *I AM* Scale Score Ranges Based on Final Performance Standards

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
ELA 3	1300–1463	1464–1481	1482–1700
ELA 4	1300–1478	1479–1497	1498–1700
ELA 5	1300–1474	1475–1488	1489–1700
ELA 6	1300–1466	1467–1486	1487–1700
ELA 7	1300–1485	1486–1497	1498–1700
ELA 8	1300–1464	1465–1490	1491–1700
ELA 10	1300–1467	1468–1505	1506–1700
Mathematics 3	2300–2462	2463–2473	2474–2700
Mathematics 4	2300–2461	2462–2478	2479–2700
Mathematics 5	2300–2459	2460–2470	2471–2700
Mathematics 6	2300–2461	2462–2477	2478–2700
Mathematics 7	2300–2466	2467–2477	2478–2700
Mathematics 8	2300–2463	2464–2474	2475–2700
Mathematics 10	2300–2470	2471–2484	2485–2700
Science 4	3300–3475	3476–3496	3497–3700
Science 6	3300–3465	3466–3488	3489–3700
Biology	3300–3471	3472–3502	3503–3700

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
Social Studies 5	4300–4488	4489–4499	4500–4700

8. REPORTING AND INTERPRETING I AM SCORES

The purpose of this chapter is to describe the information available from the scores reported for the 2024–2025 I AM assessments, and to define appropriate uses and inferences that can be drawn from them. This chapter also documents the features of the score reports provided through the online Centralized Reporting System (CRS), which is designed to assist stakeholders in reviewing, downloading, and appropriately interpreting test results.

8.1 OVERVIEW OF I AM SCORE REPORTS

Scores from each Spring 2025 assessment were provided to corporations and schools through the CRS beginning on April 7, 2025, for the preliminary scores, and on July 1, 2025, for the final scores. The CRS provides information on student performance and aggregated summaries at several levels—state, corporation, school, and roster.

Centralized Reporting System

The CRS generates a set of online score reports that describe student performance for students, families, educators, and other stakeholders. The online score reports are produced after the assessments are submitted by the students and processed into the CRS. In addition to each individual student's score report, the CRS produces aggregate score reports for teachers, schools, corporations, and the state. The timely accessibility of aggregate score reports helps users monitor student performance in each subject and grade area, evaluate the effectiveness of instructional strategies, and inform the adoption of strategies to improve student learning and teaching during the school year.

To facilitate comparisons, each aggregate report contains the summary results for the selected aggregate unit, as well as all aggregate units above the selected aggregate in the hierarchy. For example, if a school is selected, the summary results of the corporations to which the school belongs and the summary results of the state are also provided so that school performance can be compared with corporation performance and state performance. If a teacher is selected, summary results for the school, corporations, and state above the teacher are also provided for comparison purposes. Table 88 lists the following types of online score reports: student, roster, teacher, school, and corporation.

When the state produces reports that the public can access, such as school- and corporation-level means or percentage proficient overall disaggregated by subgroup, suppression rules are intended to protect privacy for disaggregated reporting. IDOE implements a minimum group size of 10 for publishing those results disaggregated by subgroup.

CRS is designed to help educators and parents answer questions about how well students have performed on ELA, Mathematics, Science, and Social Studies

assessments. CRS is the online tool that provides educators and other stakeholders with timely, relevant score reports. It has been designed with multiple stakeholders, including those who are not technical measurement experts, to ensure that the score reports are easy to read and understand. This is achieved by using simple language so that users can understand assessment results quickly and make inferences about student achievement. CRS is also designed to present student performance in a uniform format. For example, similar colors are used for groups of similar elements, such as performance levels, throughout the design. This design strategy allows readers to compare similar elements and avoid comparing dissimilar elements.

The CRS is a web-based application that provides *I AM* results to users at various levels. Assessment results are available to users based on their roles and the access privileges granted to each authenticated user. There are four types of access: (1) state, (2) corporation, (3) school, and (4) teacher (roster). Users at each level are granted drill-down access to reports in the system in accordance with their assigned role. This means that teachers can access data only for their roster(s) of students, each school can access data only for the students in that school, and corporations can access data for all schools and students in that corporation.

Users have the following types of access to the CRS:

- State users can access all state, corporation, school, teacher, and student data.
- Co-Op Corporation Administrators (Co-Ops) and Corporation Test Coordinators (CTCs) have access to all data for their corporations and for the schools and students in their corporations.
- School Test Coordinators (STCs) and Principals (PR) have access to all data for their school and for the students in their school.
- Test Administrators (TAs) can access all aggregate data for their roster(s) and the students within their roster(s).

Access to the CRS is password protected; users can access data at their assigned access level and below. For example, an STC can access the school report of students for their school but not for another school.

Available Reports on the *I AM* Centralized Reporting System

The hierarchical structure of the Indiana CRS enables authorized users to view reports at their own level and at any lower level(s) of aggregation. For example, an STC can view only the reports and data for his or her own school and for the students at the school. A CTC can view the reports and data for all schools and students in their corporation.

Table 88 summarizes the types of score reports that are available in the CRS and the levels at which the reports can be viewed. A description of each report is also provided. Data files are also accessible for corporations to download.

For detailed information on available reports and features, educators can refer to the How to Navigate Reports section of the *Centralized Reporting System (CRS) User Guide*.

Table 89: Indiana Score Reports Summary

Report	Description	Level of Availability					
		State	Corporation	School	Teacher	Roster	Student
Summary Performance	Summary of performance (to date) across grades and subjects or courses for the current administration	✓	✓	✓	✓	✓	
Aggregate-Level Subject Report	Summary of overall performance for a subject and a grade for all students in the defined level of aggregation	✓	✓	✓	✓	✓	
Aggregate-Level Reporting Category Report	Summary of overall performance on each reporting category for a given subject and grade across all students within the selected level of aggregation	✓	✓	✓	✓	✓	
Student-Level Subject Report	List of all students who belong to a school, teacher, or roster with their associated subject or course scores for the current administration			✓	✓	✓	
Student-Level Reporting Category Report	List of all students who belong to a school, teacher, or roster with their associated reporting category performance for the current administration			✓	✓	✓	
Individual Student Report (ISR)	Detailed information about a selected student's performance in a specified subject or course; includes overall subject and reporting category results		✓	✓	✓		✓
Data Files	Excel/CSV files containing overall and reporting category scale scores and performance levels along with demographic information		✓	✓	✓	✓	

The aggregate score reports provide overall student results by default but can at any time be analyzed by subgroups based on demographic data. When used on aggregate-level reports, an additional level of analysis will be provided by aggregating students based on subgroups. For example, when the “Gender” subgroup is selected, the CRS will display aggregate results for all students, male students, and female students. When used on student-level reports, subgroups can instead filter individual results. For example, a user has the option to select “Male” or “Female” after the “Gender” subgroup is selected.

Users can see student assessment results by any subgroup at any time by selecting the desired subgroup from the “Breakdown By” drop-down list. Table 89 presents the types of subgroups and subgroup categories provided in the CRS.

Table 90: Indiana List of Subgroups by Category

Subgroup	Subgroup Category
Race/Ethnicity	White
	Black/African American
	Hispanic
	Asian
	American Indian/Alaska Native
	Native Hawaiian/Other Pacific Islander
	Multiracial/Two or More Races
Gender	Male
	Female
Special Education	Yes
	No
Section 504 Plan	Yes
	No
Identified English Learner	Yes
	No
Home Language	English
	Arabic
	Burmese
	Mandarin
	Spanish
	Vietnamese
Grade	Grade 3
	Grade 4
	Grade 5
	Grade 6
	Grade 7
	Grade 8
	Grade 9
	Grade 10
	Grade 11
	Grade 12

Detailed information about the online score reports and instructions on how to navigate the online score reporting system can be found in the *Centralized Reporting System User Guide*, located via a help button on the CRS and posted in the Resources section of the assessment portal.

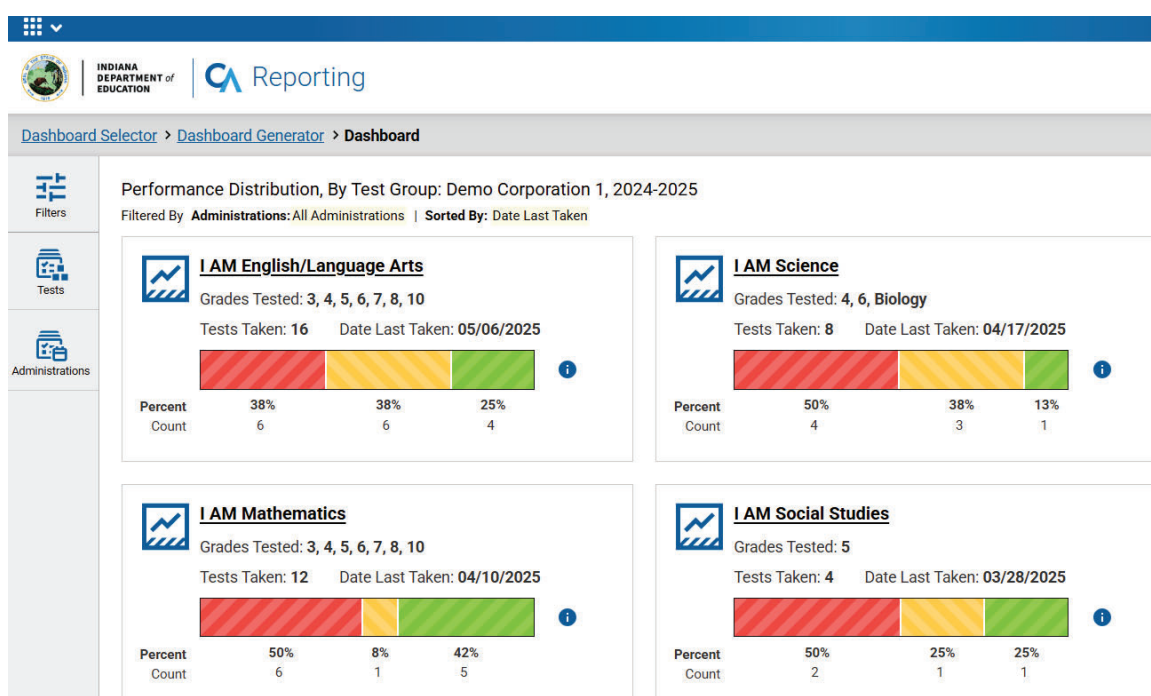
8.2 REPORTING SYSTEM FOR STUDENTS AND EDUCATORS

Dashboard

When users log in to the CRS, the dashboard page shows overall test results for all tests that the students have taken grouped by test family (e.g., Summative ELA). The dashboard summarizes students' performance by test family for ELA, Mathematics, Science, and Social Studies across all grades, including (1) the grades of the students who have tested, (2) the number of tests taken, (3) the test date last taken, and (4) the percentage and counts of students at each achievement level. District personnel see district summaries, school personnel see school summaries, and teachers see summaries of their students.

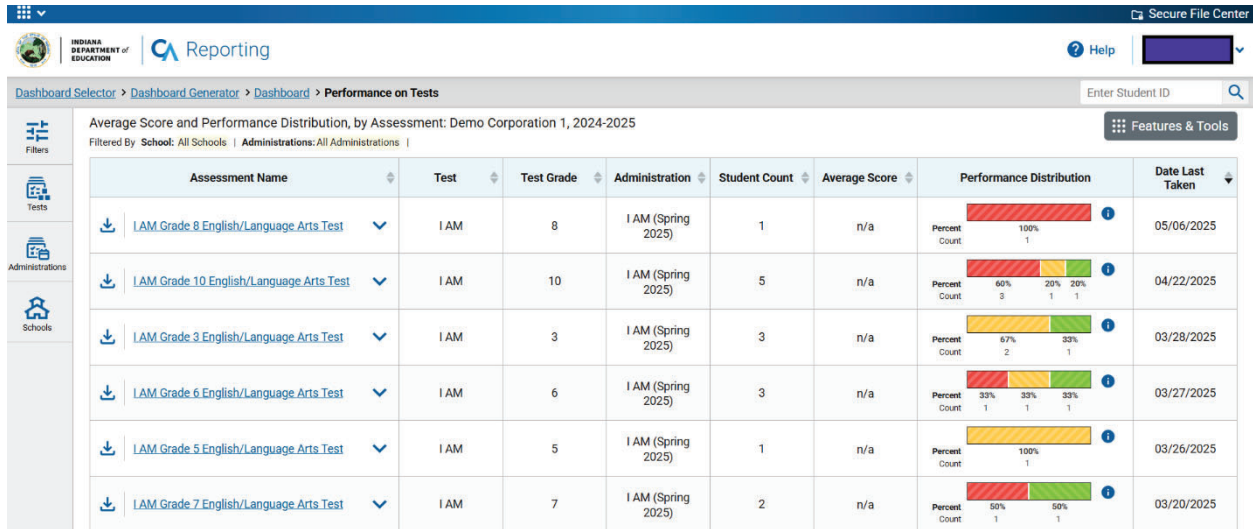
Figure 7 presents an example dashboard page in CRS at the district level.

Figure 7: Dashboard: District Level



Once the user clicks the test family that he or she wants to explore further, the system will take the user to the detailed dashboard, where the results will be displayed by test (e.g., Grade 3 / AM English/Language Arts). The detailed dashboard page will appear by test in each grade. The detailed dashboard summarizes students' performance by test in each grade, including (1) student count, (2) average scale score, and (3) percentage and counts of students at each achievement level.

Figure 8 presents an example detailed dashboard page at the district level.

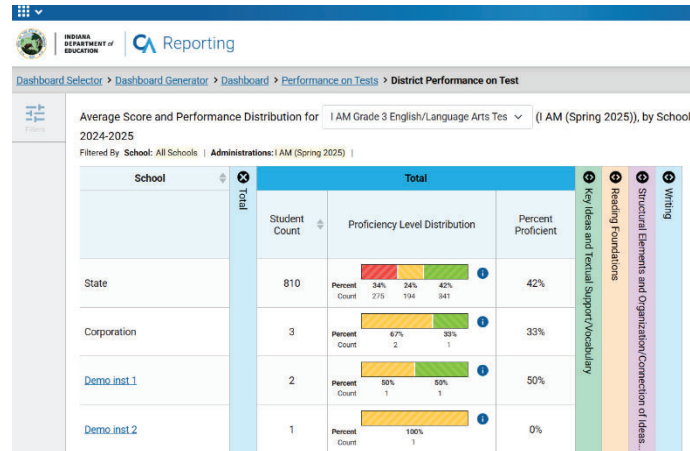
Figure 8: Detailed Dashboard: District Level

Aggregate-Level Subject Detail Page

More detailed summaries of student performance in each grade in a subject area for a selected aggregate level are presented when users select an assessment on the dashboard page. On each aggregate report, the summary report presents the summary results for the selected aggregate unit and the summary results for all aggregate units above the selected aggregate. For example, at the roster level, summaries appear for the teacher, school, and district aggregate. The roster performance can be compared with the above aggregate levels.

The subject detail page provides the aggregate summaries on a specific subject area, including: (1) number of students, (2) percentage proficient, and (3) percentage of students in each performance level. The summaries are also presented for overall students and by subgroup.

Figure 9 presents an example subject detail page for ELA at the district level.

Figure 9: Subject Detail Page for ELA: District View

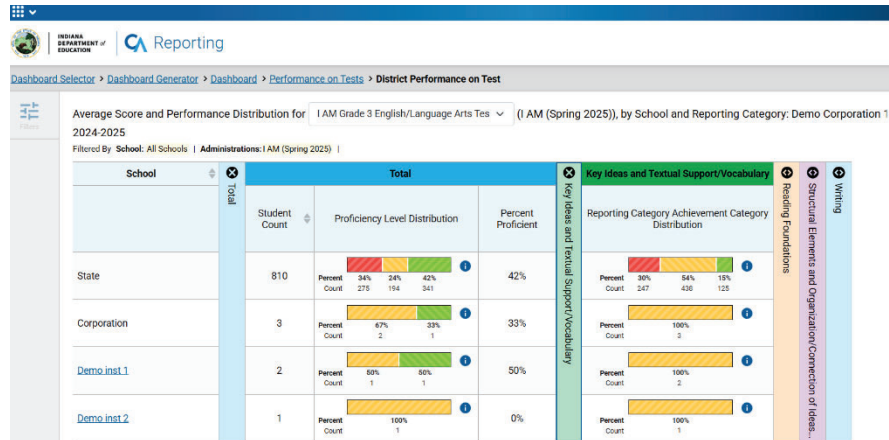
Aggregate-Level Reporting Category Report

The Aggregate-Level Reporting Category Report provides the aggregate summaries on student performance in each reporting category for a particular grade and subject. The summaries on the Aggregate-Level Reporting Category Report include: (1) number and percentage of students in each performance level, (2) percentage proficient, and (3) number and percentage of students in each achievement category for each reporting category.

A performance indicator produces information on how a group of students in a roster, school, or district performed on the standard compared to the proficiency cuts. The performance indicator shows whether performance on this standard for this group was above, no different from, or below what is expected of students at the proficient level.

Similar to the Aggregate-Level Subject Report, this report presents the summary results for the selected aggregate unit as well as the summary results for the aggregate units above the selected aggregate.

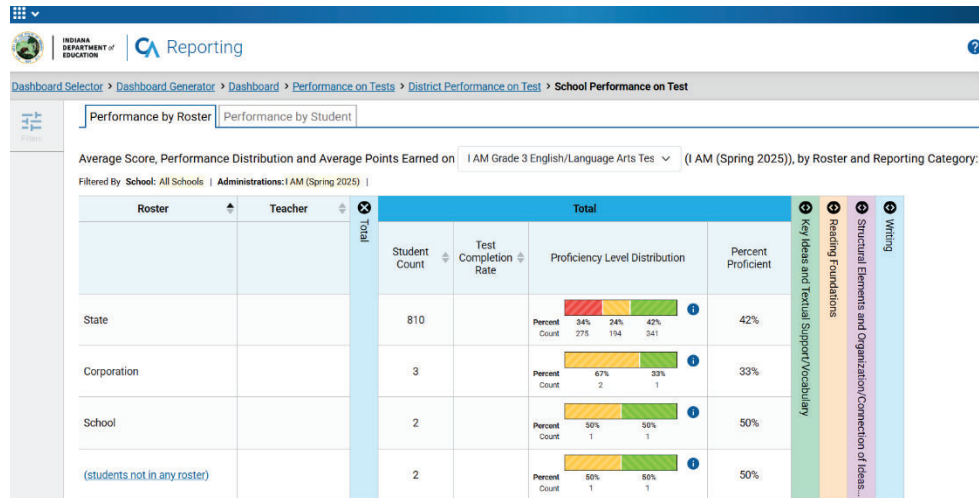
Figure 10 presents an example of the District Aggregate-Level Reporting Category detail page for ELA.

Figure 10: Reporting Category Detail Page for ELA: District Level

Student Performance on Test Report: Performance by Roster

The Student Roster Subject Report lists all students who belong to the selected aggregate level, such as a school, and reports the following measures for each student: (1) number of students, (2) number and percentage of students in each performance level, and (3) percentage proficient.

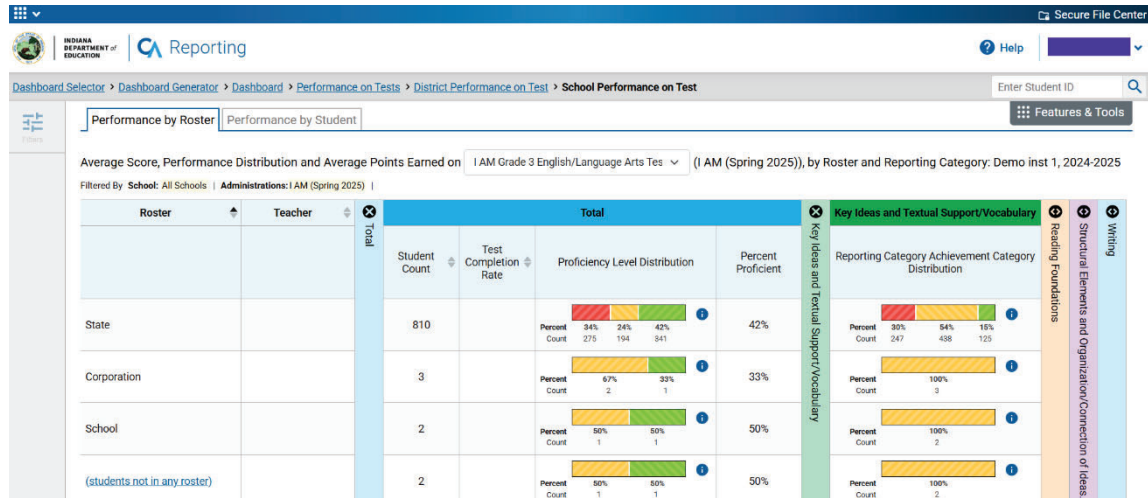
Figure 11 presents an example of the Student Roster Subject Report for ELA.

Figure 11: Student Performance on Test Report: Performance by Roster

Student Performance on Test Report: Performance by Roster with Expanded Reporting Category Section

The Student Roster Reporting Category Report records the reporting category achievement category measures for each student. Figure 12 presents an example of the Student Roster Reporting Category Report for ELA.

Figure 12: Student Performance on Test Report: Performance by Roster with Expanded Reporting Category Section

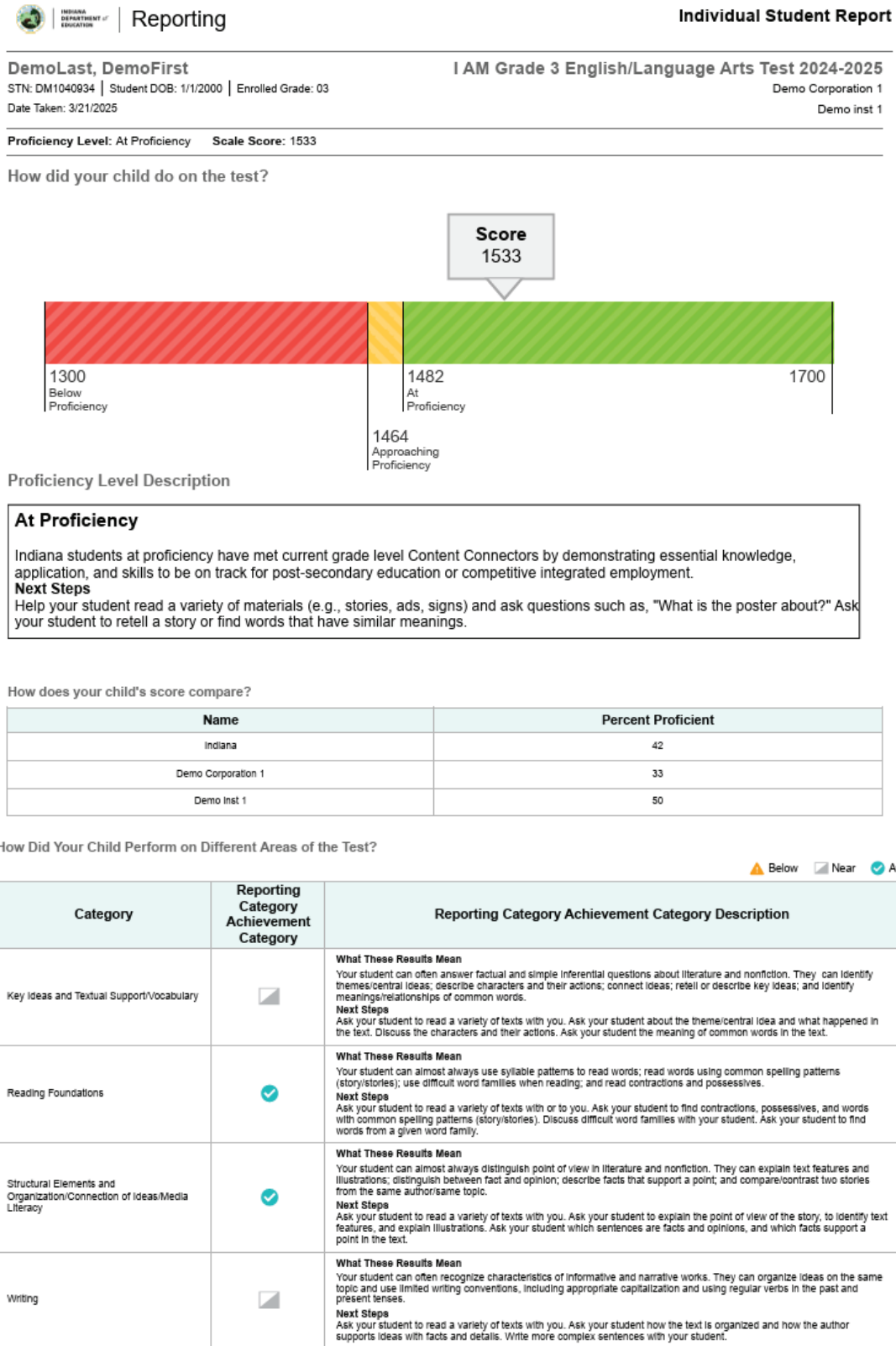


Student Individual Score Report Page

When a student completes a test, an online score report appears on the student detail page in the CRS. The student detail page provides information about individual student performance on the test. It also provides (1) average scale score, (2) performance level for the overall test, and (3) average scale scores for the student's state, district, and school in each subject area.

In the upper-right corner of the page, the student's name, scale score, and performance level are shown. In the "How did your child do on the test?" section, the student's performance is described in detail using a horizontal bar chart. The student scale score is presented in the horizontal bar chart. In the "How does your child's score compare?" section, Percent Proficient for the student's state, district, and school are displayed so that student achievement can be compared with the above aggregate levels. The student's performance in each reporting category is shown in the "How did your child perform on different areas of the test?" section, where the performance is shown graphically followed by a description of the performance. Figure 13 presents an example of a student detail page for ELA.

Figure 13: Student Individual Score Report for ELA



8.3 INTERPRETATION OF REPORTED SCORES

A student's performance on an *I AM* assessment is reported as a scale score and a performance level for the overall assessment, and as a separate performance level for each reporting category. Students' scores and performance levels are summarized at the aggregate level. This section describes how to interpret these scores.

8.3.1 SCALE SCORE

The *I AM* assessment measures the knowledge and skills students are expected to develop and demonstrate in the context of Indiana's Alternate Achievement Standards or Content Connectors. Therefore, scale scores, which are estimates of student achievement and proficiency measured by assessment, are used to explain how well students performed against such expectations. The *I AM* scale of each assessment was developed based on the *I AM* administration in Spring 2019. Details are provided in the *I AM 2018–2019 Technical Report* Volume 1, Section 5.2, Establishing the *I AM* Bank.

A scale score is the student's overall numeric score. Scale scores can be used to illustrate students' current levels of performance and to compare performances across groups of students. Lower scale scores can indicate that the student does not possess sufficient knowledge and skills measured by the assessment. Conversely, higher scale scores can indicate that the student has proficient knowledge and skills measured by the assessment. Tables 42 to 45, Marginal Reliability for ELA, Mathematics, Science and Social Studies, provide the means and standard deviations of the observed scale scores from the Spring 2025 *I AM* population data. When combined across a student population, scale scores can also describe school- and corporation-level changes in performance and reveal gaps in performance among different groups of students.

In addition, scale scores can be averaged across groups of students, allowing educators to use group comparisons. Interpretation of scale scores is more meaningful when the scale scores are used along with performance levels and PLDs. PLDs outline the knowledge and skills that students performing at a given level are expected to demonstrate in each content area and at each grade level for each standard assessed and allow the user to understand the progression of skills expected across the different proficiency levels.

8.3.2 STANDARD ERROR OF MEASUREMENT

A student's score is best interpreted when recognizing that the student's knowledge and skills fall within a score range and are not just precise numbers. A scale score (the observed score on any test) is an estimate of the true score. A test contains items that sample a student's knowledge and skills; if a student takes a similar test several times, the resulting scale scores would vary across administrations, sometimes being a little higher, a little lower, or the same. The standard error of measurement (SEM) represents the precision of the scale score, or the range in which the student would likely score if a

similar test were administered several times. The SEM can be interpreted as the degree of uncertainty of a student's score based on a statistical analysis of the student's answers on a test. When interpreting scale scores, it is recommended to always consider the range of scale scores incorporating the SEM of the scale score, because small differences in scores may not reflect real or meaningful differences in performance. The details of SEM and the graphs of the conditional SEM (CSEM) of each test are provided in Section 3.5, Reliability.

8.3.3 PERFORMANCE LEVELS

For *I AM*, scale scores are mapped onto three performance levels (Level 1—Below Proficiency, Level 2—Approaching Proficiency, and Level 3—At Proficiency) using performance standards (or cut scores; see Section 7.2, Recommended Proficiency Standards). PLDs are descriptions of content-area knowledge and skills that students at each performance level are expected to possess. PLDs are available on the Indiana Department of Education [webpage](#).

8.3.4 AGGREGATED SCORE

Students' scale scores are aggregated at the roster, school, and district levels to represent how a group of students performed on a test. When students' scale scores are aggregated, the aggregated scale scores can be interpreted as an estimate of the knowledge and skills that a group of students possesses. Given that student scale scores are estimates, the aggregated scale scores are also estimates and are subject to measures of uncertainty. In addition to the aggregated scale scores, the percentage of students in each performance level for the overall subject are reported at the aggregate level to represent how well a group of students performed overall.

8.3.5 PERFORMANCE CATEGORY FOR REPORTING CATEGORIES

Students' performance on each reporting category was computed using all items for scoring in categories that have a minimum of seven items in the blueprint. The performance of each reporting category is reported in three performance categories: (1) Below Proficiency, (2) Near Proficiency, and (3) At Proficiency. Students performing at Below Proficiency or At Proficiency can be interpreted as student performances clearly below or at the proficiency cut score for a specific reporting category. Students performing at Near Proficiency can be interpreted as student performances that are close to the proficiency cut score, but where there is not enough information to determine if they are below or at this score.

Unlike the performance level for the overall score, which is determined by comparing an overall scale score against each cut, the performance levels for reporting categories are classified by comparing a SEM range of a subscale score to the proficiency cut. Therefore, performance levels for the reporting category are limited in their diagnostic ability based on the degree of the calculated SEM of the student's scale score for each reporting category. The individual student report (ISR) provides the SEM range of the reporting category scale score for each reporting category along with a proficiency cut.

The means and standard deviations for each reporting category by assessment are also provided in Appendix 3-C, Distribution of Reporting Category Scores by Subgroup.

I AM displays Next Steps information at the reporting category level in CRS. Next Steps information suggests activities educators and parents/guardians may do with their student to help improve their student's knowledge and performance on future assessments. Educators and parents may use the Next Steps feature to better understand student test results and help further support their student.

8.4 APPROPRIATE USES FOR SCORES AND REPORTS

Assessment results can be used to provide information on individual students' performance on the assessment. Overall, assessment results show what students know and can do in certain subject areas and give further information on whether students are on track to demonstrate the knowledge and skills aligned to Indiana's Academic Standards (content standards). Additionally, assessment results can be used to identify students' relative strengths and weaknesses in certain content areas. For example, performance categories for reporting categories can be used to identify an individual student's relative strengths and weaknesses among reporting categories within a content area.

Results on students' performance on the assessment can be used to help teachers or schools make decisions on how to support students' learning. Aggregate score reports on the teacher and school level provide information about students' strengths and weaknesses and can be used to improve teaching and students' learning. For example, a group of students may have performed well overall, but not as well in several reporting categories. In this case, teachers or schools can identify the strengths and weaknesses of their students through the group performance by reporting category and promote instruction on specific areas where student performance is below overall performance.

Furthermore, by narrowing the student performance results by subgroup, teachers and schools can determine what strategies may need to be implemented to improve teaching and students' learning, particularly for students from disadvantaged subgroups. For example, teachers might see students' assessment results by gender and observe that a particular group of students is struggling with literary response and analysis in reading. In addition, assessment results can be used to compare students' performance among different students and different groups. Teachers can evaluate how their students perform compared with other students in schools and corporations by overall scores and reporting category scores.

Although assessment results provide valuable information to understand students' performance, these scores and reports should be used with caution. It is important to note that scale scores are estimates of true scores and hence do not represent a precise measure of student performance. Students' performance on an assessment may vary due to a variety of reasons (e.g., they are not feeling well, they are not feeling motivated). A student's scale score is associated with measurement error, and the SEM is the range in which a student's "true score" is expected to fall. Even though the SEM is not reported in the CRS, when interpreting scale scores, it is important to recognize the uncertainties

associated with them as a result of measurement error and avoid interpreting them as precise numbers. For example, a scale score of 2535 with a SEM of 22 indicates that if the student completed the same test multiple times, the score would likely fall between 2513 and 2557. Scale scores and SEMs will vary based on the test and student.

Moreover, although student scores may be used to help make important decisions about students' placement and retention or teachers' instructional planning and implementation, the assessment results should not be relied on as the only source of information. Given that assessment results provide limited information, other sources of data on student performance, such as classroom assessment and teacher evaluation, should be considered when making decisions on student learning. Finally, when student performance is compared across groups, users must account for group size. The smaller the group, the larger the measurement error related to these aggregated data, thus requiring a more cautious interpretation.

9. QUALITY ASSURANCE PROCEDURES

Quality assurance (QA) procedures are enforced throughout all stages of *I AM* test development, administration, and scoring and reporting. This chapter describes QA procedures associated with the following:

- Test construction
- Test production
- Data preparation
- Equating and scaling
- Scoring and reporting

Because QA procedures pervade all aspects of test development, we note that discussion of QA procedures is not limited to this chapter but is also included in chapters describing all phases of test development and implementation.

9.1 QUALITY ASSURANCE IN ITEM DEVELOPMENT AND TEST CONSTRUCTION

Chapter 4 of this technical report details the item development and test configuration processes. Each form is built to match the detailed test blueprint. The blueprint describes the content to be covered, the type of items that will measure the constructs, and every other content-relevant aspect of the test. CAI's test developers use Workspaces in the Form and Test Engineering System to help construct operational forms.

Immediately upon generation of a test form, Workspace generates a blueprint match report to ensure that all elements of the test blueprint have been satisfied.

The mechanical features of a test—arrangement, directions, and production—are just as important as the quality of the items. Many factors directly affect a student's ability to demonstrate proficiency on the assessment, while others relate to the ability to score the assessment accurately and efficiently. Still others affect the inferences made from the test results.

When the test developer is reviewing a test form for content, in addition to making sure all the benchmark/indicator item requirements are met, test developers must also make sure that the items on the form do not cue each other—that one item does not present material that indicates the answer to another item. This is important to ensure that a student's response on any test item is unaffected by, and is statistically independent of, a response to any other test item. This is called "local independence." Independence is most commonly violated when there is a hint in one item about the answer to another item. In that case, a student's true ability on the second item is not being assessed.

Once the items and passages for the form have been selected and matched against the blueprint, the test developer reviews the form for a variety of additional content considerations, including the following:

- The items are sequentially ordered.
- Each item of the same type is presented in a consistent manner.
- The listing of the options for the multiple-choice items is consistent.
- All graphics are consistently presented.
- All tables and charts have titles and are consistently formatted.
- The number of the answer choice letters should be approximately equal across the form.
- The answer key should be checked by the initial reviewer and one additional independent reviewer.
- All stimuli have items associated with them.
- The topics of items, passages, or stimuli are not too similar to one another.
- There are no errors in spelling, grammar, or accuracy of graphics.
- The wording, layout, and appearance of the item match how the item was field-tested.
- There is gender and ethnic balance.
- Each item and the form have been checked against the appropriate style guide.
- Directions are consistent across items and are accurate.
- All copyrighted materials have up-to-date permissions agreements.
- Word counts are within documented ranges.

After completing the initial build of the form, the test developer hands it off to another content specialist, who conducts a final review of the listed criteria. If the test specialist reviewer finds any issues, the form is sent back for revisions. If the form meets blueprint and complies with all specified criteria, the test developer sends it to the psychometric team for review. When the psychometric team approves the form, the test developer submits the Workspace to the IDOE for review and approval. After operational forms are approved in the Workspace, all test maps, key files, and conversion tables are produced directly from the Form and Test Engineering System to eliminate the possibility of human error in the construction of these important files. Test maps, key files, conversion tables, and other critical documents are generated directly from information maintained in the Item Authoring Tool (IAT). The information stored in IAT is rigorously reviewed by multiple skilled reviewers to protect against errors. Automated production of these critical files (such as key files) virtually eliminates the risk of error.

Test maps can include any item attribute stored in IAT, so that in addition to form-level attributes such as test administration and item position, item attributes such as learning standard, benchmark, indicator, complexity, item release status, point value, weight, keyed response, and more are included in the test maps. The test maps feature in the Workspace is customized to *I AM*.

As a further layer of QA for printed test booklets, both during the blueprint production phase prior to printing and again following the final printing of all test forms, two CAI staff members independently took all test forms. Responses to the test forms were compared to the answer keys for each form to confirm the accuracy of scoring keys. In addition, the printed forms were compared against IAT and the Workspace for content and item ordering to ensure that no changes to the form were introduced prior to printing.

Prior to its implementation in the operational test administration, the CAI scoring engine and the accuracy of data files are checked using a simulated student response data file. The simulated data are used to check whether the student responses entered in the Test Delivery System (TDS) were captured accurately and whether scoring specifications were applied accurately. The simulated data file is scored independently by two programmers, following scoring rules.

In addition to checking the scoring accuracy, the test configuration file is checked thoroughly. For the operational administration, a test configuration file is the key file that contains all specifications for the item selection algorithm, and eventually for the scoring algorithm, such as the test blueprint specification, slopes and intercepts for theta-to-scale score transformation, and the item information (e.g., cut scores, answer keys, item attributes, item parameters, passage information). The accuracy of the information in the configuration file is checked and confirmed numerous times independently by multiple staff members before the testing window opens.

9.2 QUALITY ASSURANCE IN COMPUTER-DELIVERED TEST PRODUCTION

9.2.1 PRODUCTION OF CONTENT

While the online workflow requires some additional steps, it removes a substantial amount of work from the time-critical path, reducing the likelihood of errors. Like a test book, an online system can deliver a sequence of items; however, the online system makes the layout of that sequence algorithmic. The appearance of the item screen can be known with certainty before the final test is configured.

The production of computer-based tests includes four key steps:

1. Final content is previewed and approved in a process called web approval. Web approval packages the item exactly as it will be displayed to the student.
2. The complete test configuration is approved, which gathers the content, form information, display information, and relevant scoring and psychometric information from the item bank and packages it for deployment.
3. Tests are initially deployed to a test site where they undergo platform review, a process during which we ensure that each item displays properly on a large number of platforms representative of those used in Indiana for testing purposes.
4. The final system is deployed to a staging environment accessible to IDOE for user acceptance testing (UAT) and final review.

9.2.2 WEB APPROVAL OF CONTENT DURING DEVELOPMENT

The Item Tracking System (ITS) integrates directly with the TDS display module and displays each item exactly as it will appear to the student. This process is called Web Preview and is tied to specific item review levels. Upon approval at those levels, the

system locks content as it will be displayed to the student, transforming the item representation to the exact representation that will be rendered to the student. No change to the display content can occur without a subsequent Web Preview. This process freezes the display code that will present the item to the student.

Web approval functions as an item-by-item blueline review. It is the final rendering of the item as the student will view it. Layout changes can be made after this process in two ways:

1. Content can be revised and reapproved for web display.
2. Online style sheets can be changed to revise the layout of all items on the test.

Both processes are subject to strict change-control protocols to ensure that accidental changes are not introduced. In the following section, we discuss automated quality control processes during content publication that raise warnings if item content has changed after the most recent web-approved content was generated. The web approval process offers the benefit of allowing final layout review much earlier in the process, reducing the work that must be performed during the very busy period just before tests go live.

9.2.3 PLATFORM REVIEW

Platform review is a process in which each item is checked to ensure that it is displayed appropriately on each testing platform. A platform is a combination of a hardware device and an operating system. In recent years, the number of platforms has proliferated, and platform review now takes place on approximately 15 significantly different platforms.

Platform review is conducted by a team. The team leader projects the item in its web-approved ITS format, and team members, each behind a different platform, look at the same item to gauge whether it renders as expected.

9.2.4 USER ACCEPTANCE TESTING AND FINAL REVIEW

Each release of every one of our systems goes through a complete testing cycle, including regression testing. With each release, and every time we publish a test, the system goes through UAT. During UAT, we provide our client with login information to an identical (though smaller scale) testing environment to which the system has been deployed. We provide recommended testing scenarios and constant support during the UAT period. Identified issues are resolved before the opening of the test administration or noted for future review and resolution if a current resolution is not feasible within the timeline. IDOE signs off on the administration go-live date at the conclusion of UAT activities.

Deployments to the production environment follow specific, approved deployment plans. Teams working together execute the deployment plan. Each step in the deployment plan is executed by one team member and verified by a second. Each deployment undergoes shakeout testing following the deployment. This careful adherence to deployment procedures ensures that the operational system is identical to the system evaluated on the testing and staging servers. Upon completion of each deployment project, management approves the deployment log.

During the year, some changes to the production system may be required. Outside of routine maintenance, no change is made to the production system without approval of the Production Control Board (PCB). The PCB includes the director of CAI's Assessment Program or the chief operating officer, the director of our Computer and Statistical Sciences Center, and the project director. Any request for a change to the production system requires the signature of the system's lead engineer. The PCB reviews risks, test plans, and test results. In addition, if any proposed change will affect client functionality or pose risks to the operation of a client system, the PCB ensures that the client is informed and in agreement with the decision.

The PCB approves a maintenance plan that includes every scheduled change to the system.

Deviations from the maintenance plan must be approved by the PCB, including server or driver patches that differ from those approved in the maintenance plan.

Every bug fix, enhancement, data correction, or new feature must be presented with the results of a quality assurance plan and approved by the PCB.

An emergency procedure is in place that allows rapid response in the event of a time-critical change needed to avert compromise of the system. Under those circumstances, any member of the PCB can authorize the senior engineer to make a change, with the PCB reviewing the change retroactively.

Typically, deployments happen during a maintenance window, and deployments are scheduled at a time that can accommodate full regression testing on the production machines. Any changes to the database or procedures that in any way might affect performance are typically subject to a load test at this time.

Cutover and Parallel Processing

CAI maintains multiple environments to ensure smooth cutover and parallel processing. With a centralized hosting site in Washington, D.C., multiple development environments and a test environment can be maintained. At Rackspace, we maintain a staging environment and the production environment.

The production environment runs independently of the other environments and is changed only with the approval of the PCB. When developing enhancements, they are developed and tested initially on the development and test environments in Washington, D.C., before being deployed to the staging environment in Rackspace.

The staging environment is a scaled-down version of the production environment. It is in this environment that UAT takes place. Only when UAT is complete and the PCB signs off is the production environment updated. In this way, the system continues to function uninterrupted as testing takes place in parallel until a clean cutover takes place.

Prior to deployment, the testing system and content are deployed to a staging server where they are subject to UAT. UAT of the TDS serves both a software evaluation and

content approval role. The UAT period provides IDOE with an opportunity to interact with the exact test with which the students will interact.

9.2.5 FUNCTIONALITY AND CONFIGURATION

The items, both individually and as configured onto the tests, form one type of online product. The delivery of that test can be thought of as an independent service. Here, we document quality assurance procedures for delivering the online assessments.

One area of quality unique to online delivery is the quality of the delivery system. Three activities provide for the predictable, reliable, quality performance of our system. They include:

1. Testing on the system itself to ensure function, performance, and capacity
2. Capacity planning
3. Continuous monitoring

CAI statisticians examine the delivery demands, including the number of tests to be delivered, the length of the testing window, and the historic state-specific behaviors to model the likely peak loads. Using data from the load tests, these calculations indicate the number of each type of server necessary to provide continuous, responsive service, and CAI contracts for service in excess of this amount. Once deployed, our servers are monitored at the hardware, operating system, and software platform levels with monitoring software that alerts our engineers at the first signs that trouble may be ahead. Applications log not only errors and exceptions, but latency (timing) information for critical database calls. This information enables us to know instantly whether the system is performing as designed, or if it is starting to slow down or experience a problem.

In addition, latency data are captured for each assessed student—data about how long it takes to load, view, or respond to an item. All this information is logged, as well, enabling us to automatically identify schools or districts experiencing unusual slowdowns, often before they even notice.

9.3 QUALITY ASSURANCE IN DATA PREPARATION

When a student responds to test questions online, the response to each item is immediately captured and stored in the Database of Record (DOR) at CAI, a repository for all data relevant to a student's testing experience. CAI quality assurance procedures are built on two key principles: automation and replication. Certain procedures can be automated, which removes the potential for human error. Procedures that cannot be reasonably automated are replicated by two independent analysts at CAI.

When data are prepared for psychometric analyses, they undergo two phases: a data preparation phase and a psychometric phase. In the former phase, data are extracted from the DOR and provided to two independent SAS programmers. These two programmers are provided with the client-assigned business rules, and they

independently prepare data files suitable for subsequent psychometric analysis. The data files prepared by the different programmers are formally compared for congruency. Any discrepancies identified are resolved through code review meetings with the lead programmer and the lead psychometrician.

When the two data files match exactly, they are then passed over to two independent psychometricians, who each perform classical and IRT analyses. Any discrepancies are identified and resolved. When all results match from the independent analysts, the final results are uploaded to CAI's Item Tracking System (ITS).

CAI's Test Delivery System (TDS) has a real-time quality-monitoring component built in. As students test, data flow through our Quality Monitor (QM) system. The QM system conducts a series of data integrity checks, ensuring, for example, that the record for each test contains information for each item that was supposed to be on the test, and that the test record contains no data from items that have been invalidated. In addition, the QM scores the test, recalculates performance-level designations, calculates subscores, compares item parameters to the reference item parameters in the item bank, and conducts a host of other checks.

The QM also aggregates data to detect problems that become apparent only in the aggregate. For example, the QM system monitors item statistics and flags items that perform differently operationally than their item parameters predict. This functions as a sort of automated key or rubric check, flagging items where data suggest a potential problem. This automated process is similar to the sorts of checks performed for data review, but they are conducted (a) on operational data, and (b) in real time to allow our psychometricians to catch and correct any problems before they have an opportunity to do any harm.

Data pass directly from the QM system to the DOR, which serves as the repository for all test information, and from which all test information for reporting is pulled. The Data Extract Generator is the tool that is used to pull data from the DOR for delivery to IDOE and their QA contractor. CAI psychometricians ensure that data in the extract files match the DOR prior to delivery to the IDOE.

9.4 QUALITY ASSURANCE IN ITEM ANALYSES AND EQUATING

Prior to operational work, CAI produces simulated datasets for testing software and analysis procedures. The quality assurance procedures are built on two key principles: automation and replication. Certain procedures can be automated, which removes the potential for human error. Procedures that cannot be reasonably automated are independently replicated by two CAI psychometricians. Two psychometricians complete a dry run calibration and linking activities and compare results. The practice runs serve two functions:

1. To verify accuracy of program code and procedures

2. To evaluate the communication and workflow among participants. If necessary, the team will reconcile differences and correct production or verification programs.

Following the completion of these activities and the resolution of questions that arise, analysis specifications are finalized.

9.5 QUALITY ASSURANCE IN SCORING AND REPORTING

CAI implements a series of quality control steps to ensure error-free production of score reports in an online format. The quality of the information produced in the TDS is tested thoroughly before, during, and after the testing window.

9.5.1 QUALITY ASSURANCE IN TEST SCORING

CAI verifies the accuracy of the scoring engine using simulated test administrations. The simulator generates a sample of students with an ability distribution that matches that of the State. The ability of each simulated student is used to generate a sequence of item responses consistent with the underlying ability. Although the simulations were designed to provide a rigorous test of the adaptive algorithm for adaptively administered tests, they also provide a check of the full range of item responses and test scores in fixed-form tests. Simulations are always generated using the production item selection and scoring engine to ensure that verification of the scoring engine is based on a very wide range of student response patterns.

To monitor the performance of the assessment system during the test administration window, a series of quality assurance reports can be generated at any time during the online assessment window. For example, item analysis reports allow psychometricians to ensure that items are performing as intended and serve as an empirical key check through the operational testing window.

The quality assurance reports are generated on a regular schedule. Item analysis reports are evaluated frequently at the opening of the testing window to ensure that items are performing as anticipated. Each time the reports are generated, the lead psychometrician reviews the results. If any unexpected results are identified, the lead psychometrician alerts the content staff and project manager immediately to resolve any issues.

Item Analysis Report

The item analysis report is used to monitor the performance of test items throughout the testing window and serves as a key check for the early detection of potential problems with item scoring, including the incorrect designation of a keyed response or other scoring errors, as well as potential breaches of test security that may be indicated by changes in the difficulty of test items. To examine test items for changes in performance, this report generates classical item analysis indicators of difficulty and discrimination, including proportion correct and biserial/polyserial correlation, as well as IRT-based item fit statistics. The report is configurable and can be produced so that only items with statistics

falling outside a specified range are flagged for reporting or generating reports based on all items in the pool.

Item p-Value. For multiple-choice (MC) items, the proportion of students selecting each response option is computed. If the keyed response is not the modal response, the item is also flagged for MC items. Although the correct response is not always the modal response, keyed response options flagged for both low biserial correlations and non-modal responses are indicative of miskeyed items.

Item Discrimination. Biserial correlations for the keyed response for selected-response items and polyserial correlations for polytomous constructed-response, performance, and technology items are computed. CAI psychometric staff evaluates all items with biserial correlations below a target level, even if the obtained values are consistent with past item performance.

Item Fit. In addition to the item difficulty and item discrimination indices, an item fit index is produced for each item. For each student, a residual between the observed and expected scores given the student's ability is computed for each item. The residuals are averaged across all students, and the average residual is used to flag an item.

9.5.2 QUALITY ASSURANCE IN REPORTING

Scores for the *I AM* online assessments are assigned by automated systems in real time. *I AM* is completely machine-scored, and the machine rubrics are created and reviewed along with the items. The review process locks down the item and rubric when the item is approved for web display (Web Approval).

Once the item scores are sent to the QM system, the records are scored in the test-scoring system that applies the *I AM* scoring rules and assigns scores from the calibrated items, including calculating performance-level indicators, subscale scores, and other features, which then pass automatically to CRS and the DOR. The scoring system is tested extensively prior to deployment, including hand checks of scored tests and large-scale simulations to ensure that point estimates and standard errors are correct.

After passing through the series of validation checks in the QM, data are passed to the DOR, which serves as the centralized location for all student scores and responses, ensuring there is only one place where the official record is stored. Only after scores have passed the QM checks and are uploaded to the DOR are they passed to CRS, which is responsible for presenting individual-level results and calculating and presenting aggregate results.

All student test scores are produced using CAI's scoring engine. Before any scores are released, a second score verification system is used to verify that all test scores match with 100% agreement in all tested grades. This second system is constructed and maintained independently from the main scoring engine and separately estimates marginal maximum likelihood estimations (MLEs) using the procedures described within this report. Additionally, IDOE contracts with a third-party vendor for independent score verification and provides replication of the psychometric scoring process. Scores are

approved and considered final by IDOE only when all three independent systems match and are aligned.

10. REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for Educational and Psychological Testing*. https://www.aera.net/Portals/38/1999%20Standards_revised.pdf
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Bejar, I. I. (1980). Biased assessment of program impact due to psychometric artifacts. *Psychological Bulletin*, 87(3), 513–524. <https://doi.org/10.1037/0033-2909.87.3.513>
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage. <https://us.sagepub.com/en-us/nam/methods-for-identifying-biased-test-items/book3416>
- Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289.
- Dorans, N. J., & Schmitt, A. P. (1991). *Constructed Response and Differential Item Functioning: A Practical Approach*. <https://doi.org/10.1002/j.2333-8504.1991.tb01414.x>
- Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance Level Descriptors: History, Practice, and a Proposed Framework. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 79–106). Routledge.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 129–145). Lawrence Erlbaum Associates. <https://files.eric.ed.gov/fulltext/ED272577.pdf>
- Huynh, H. (1979). Statistical inference for two reliability indices in mastery testing based on the beta-binomial model. *Journal of Educational Statistics*, 4(3), 231–246. <https://doi.org/10.2307/1164672>
- Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). American Council on Education/Praeger.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179–197.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel–Haenszel statistic. *Educational and Psychological Measurement*, 52(2), 443–451. <https://doi.org/10.1177/0013164492052002020>
- Muñiz, J., Hambleton, R. K., & Xing, D. (2001). Small Sample Studies to Detect Flaws in Item Translations. *International Journal of Testing*, 1(2), 115–135. https://doi.org/10.1207/S15327574IJT0102_2
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation : An International Journal on Theory and Practice*, 19(2–3), 170–187. <https://doi.org/10.1080/13803611.2013.767621>
- Somes, G. W. (1986). The generalized Mantel–Haenszel statistic. *The American Statistician*, 40(2), 106–108. <https://doi.org/10.1080/00031305.1986.10475369>
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. <https://nceo.umn.edu/docs/onlinepubs/synth44.pdf>
- Van Der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In *Handbook of modern item response theory* (pp. 1–28). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-2691-6_1
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145. <https://doi.org/10.1177/014662168400800201>