Indiana's Alternate Measure

2020–2021

Volume 1
Summary of Technical Report

ACKNOWLEDGMENTS

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

Major contributors to this technical report include the following staff from Cambium Assessment, Inc. (CAI): Stephan Ahadi, Yuan Hong, Hyesuk Jang, Hashim Evans, Anne Mogilnicki, Cherise Lesesne, Celine Bryan, and Gabriel Martinez. Major contributors from the Indiana Department of Education (IDOE) include the assessment director, assistant assessment director, and program leads.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

1. INTRODUCTION

The Indiana's Alternate Measure (I AM) 2020–2021 technical report documents and provides transparency on all methods used in item development, test construction, psychometric analyses, standard setting, score reporting, summaries of student assessment results, and supporting evidence for intended uses and interpretations of the test scores. The technical report is presented as five separate, self-contained volumes that cover the following topics:

1. *Summary of Technical Report*. This annually updated volume provides a general overview of the tests administered to students each year.
2. *Test Development*. This volume details the procedures used to construct test forms and summarizes the I AM item bank and its development process.
3. *Test Administration*. This volume describes the methods used to administer all available test forms, security protocols, and modifications or accommodations.
4. *Evidence of Reliability and Validity*. This volume provides an array of reliability and validity evidence that supports the intended uses and interpretations of the test scores.
5. *Score Interpretation Guide*. This volume describes the score types reported along with the appropriate inferences and intended uses of each score type.

The Indiana Department of Education (IDOE) communicates the quality of I AM by making these technical reports accessible to the public.

1.1 BACKGROUND AND HISTORICAL CONTEXT OF THE TEST

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).

1.2 PURPOSE AND INTENDED USES OF INDIANA'S ALTERNATE MEASURE

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 immediate feedback to educators and parents, which can be used to inform instructional strategies that remediate or enrich instruction. A variety of reporting metrics allows achievement to be monitored at both student and aggregated levels and growth over time to be measured at both student and group levels.

I AM draws all items from the I AM item bank, which is a custom item bank created and maintained by Indiana. I AM content standards in ELA, Mathematics, Science, and Social Studies are aligned with knowledge and skills that are essential for competitive employment and post-secondary education. For the Spring 2021 I AM administration, Cambium Assessment Inc. (CAI) and IDOE worked together to ensure that the items in

the test forms constructed for all grades uniquely measured students' mastery of the Indiana Alternate Academic Standards (Content Connectors) in ELA, Mathematics, Science, and Social Studies.

Table 1 outlines the required participation criteria for I AM.

Table 1: 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.
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, synthesize, demonstrate, and transfer skills across multiple settings.
Goals listed in the Individual Education Plan (IEP) for this student are linked to the enrolled grade level Alternate Achievement Standards (Indiana Content Connectors).

The purpose and intended use of I AM are described on the IDOE web page.

1.2.1 Participants in the Development and Analysis of Indiana's Alternate Measure

IDOE manages the Indiana state assessment program with the assistance of Indiana educators, a Technical Advisory Committee (TAC), and several vendors (listed in the following paragraphs). IDOE fulfills the various state requirements for implementing I AM while meeting or exceeding the guidelines established in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, 2014).

Indiana Department of Education

The Office of Student Assessment oversees all aspects of the alternate assessment program, including coordination with other IDOE offices, Indiana public schools, and vendors.

Indiana Educators

Indiana educators participate in most aspects of the conceptualization and development of I AM. Educators participate in the development of the academic standards, clarification

of how these standards will be assessed, test design, and committee reviews of test items and passages.

Technical Advisory Committee

The State Board of Education convenes a TAC panel three times per year to discuss psychometric, test development, administrative, and policy issues relevant to current and future Indiana assessments. This committee is composed of nationally recognized assessment experts and highly experienced practitioners from independent and higher-education institutions.

Advisory Group

Indiana also utilizes a corporation-based advisory group, Assessment Implementation Advisory Group, that meets monthly to advise on implementation practices. This group consists of educators, technology coordinators, and test coordinators.

Cambium Assessment, Inc.

CAI was selected as the vendor for I AM through the state-mandated competitive procurement process. In Spring 2018, CAI (formerly American Institutes for Research) became the primary party responsible for developing test content, building test forms, conducting psychometric analyses, administering assessments, scoring test forms, and reporting test results for I AM. Additionally, CAI is responsible for developing and maintaining the I AM item bank, which is used for the I AM test construction.

Human Resources Research Organization

The Human Resources Research Organization (HumRRO) conducted independent verifications of scoring activities for the Spring 2021 assessments.

1.3 Available Test Formats and Special Versions

I AM is administered as an online, stage-adaptive assessment using mainly multiple-choice (MC) item types (see Section 3.2.1 of Volume 2). 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 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. English learners can take the Spanish-language version of the I AM Mathematics, Science, and Social Studies assessments; these forms are the same tests as the English forms but translated into Spanish. The items are translated by a third-party vendor that provides

professional translation services. Test developers also evaluate forms by researching and testing various response options to ensure that scores obtained using the Spanish-language version or other alternative modes of administration will be comparable to those earned on the standard online test that adheres to the same blueprint.

1.4 STUDENT PARTICIPATION

Table 1 identifies criteria required for student participation in the I AM assessment. All students in Indiana public or accredited nonpublic schools who meet the requirements stated in Table 1 are required to participate in their graded level I AM assessments to meet state accountability measures. 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. Table 2 shows the number of students tested and the number of students reported in the Spring 2021 I AM administration by grade and subject area. Tables 3–6 present the distribution of students, in counts and percentages, by subgroups for ELA, Mathematics, Science, and Social Studies, respectively. The subgroup categories reported here are gender, primary disability, and race/ethnicity.

Table 2: Number of Students Who Participated in the I AM Spring 2021 Administration

ELA			Mathematics			Science			Social Studies		
Grade	Number Tested	Number Reported	Grade	Number Tested	Number Reported	Grade	Number Tested	Number Reported	Grade	Number Tested	Number Reported
3	624	565	3	624	568						
4	690	627	4	684	629	4	678	622			
5	725	671	5	720	666				5	714	660
6	819	749	6	812	746	6	806	740			
7	857	808	7	854	804						
8	1022	971	8	1022	966						
10	1006	951	10	994	944	Biology	1026	963			

Table 3: Distribution of Demographic Characteristics of Tested Population: ELA

Grade	Metric	All Students	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
3	*N*	624	219	405	216	391	157	450	100	84	385
	%	100	35.10	64.90	34.62	62.66	25.16	72.12	16.03	13.46	61.70
4	*N*	690	221	469	237	437	201	473	97	93	430
	%	100	32.03	67.97	34.35	63.33	29.13	68.55	14.06	13.48	62.32
5	*N*	725	262	463	227	489	216	500	113	86	467
	%	100	36.14	63.86	31.31	67.45	29.79	68.97	15.59	11.86	64.41
6	*N*	819	284	535	257	549	218	588	114	118	514
	%	100	34.68	65.32	31.38	67.03	26.62	71.79	13.92	14.41	62.76
7	*N*	857	267	590	246	595	207	634	146	100	545
	%	100	31.16	68.84	28.70	69.43	24.15	73.98	17.04	11.67	63.59
8	*N*	1022	364	658	301	694	285	710	164	135	637
	%	100	35.62	64.38	29.45	67.91	27.89	69.47	16.05	13.21	62.33
10	*N*	1006	379	627	258	722	260	720	156	114	674
	%	100	37.67	62.33	25.65	71.77	25.84	71.57	15.51	11.33	67.00

Table 4: Distribution of Demographic Characteristics of Tested Population: Mathematics

Grade	Metric	All Students	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
3	*N*	624	218	406	215	394	158	451	100	85	384
	%	100	34.94	65.06	34.46	63.14	25.32	72.28	16.03	13.62	61.54
4	*N*	684	220	464	231	434	200	465	96	91	427
	%	100	32.16	67.84	33.77	63.45	29.24	67.98	14.04	13.30	62.43
5	*N*	720	260	460	223	484	215	492	112	87	462
	%	100	36.11	63.89	30.97	67.22	29.86	68.33	15.56	12.08	64.17
6	*N*	812	279	533	257	541	218	580	112	116	510
	%	100	34.36	65.64	31.65	66.63	26.85	71.43	13.79	14.29	62.81
7	*N*	854	267	587	246	593	207	632	145	100	543
	%	100	31.26	68.74	28.81	69.44	24.24	74.00	16.98	11.71	63.58
8	*N*	1022	367	655	302	693	285	710	164	136	636
	%	100	35.91	64.09	29.55	67.81	27.89	69.47	16.05	13.31	62.23
10	*N*	994	374	620	255	710	257	708	151	109	672
	%	100	37.63	62.37	25.65	71.43	25.86	71.23	15.19	10.97	67.61

Table 5: Distribution of Demographic Characteristics of Tested Population: Science

Grade	Metric	All Students	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
4	*N*	678	218	460	228	429	200	457	95	91	423
	%	100	32.15	67.85	33.63	63.27	29.50	67.40	14.01	13.42	62.39
6	*N*	806	277	529	254	537	214	577	109	115	508
	%	100	34.37	65.63	31.51	66.63	26.55	71.59	13.52	14.27	63.03
Biology	*N*	1026	374	652	269	735	267	737	152	122	685
	%	100	36.45	63.55	26.22	71.64	26.02	71.83	14.81	11.89	66.76

Table 6: Distribution of Demographic Characteristics of Tested Population: Social Studies

Grade	Metric	All Students	Female	Male	Autism	Non-Autism	Moderate and Severe Intellectual Disability	Non-Moderate and Severe Intellectual Disability	African American	Hispanic	White
5	*N*	714	258	456	219	478	213	484	110	87	459
	%	100	36.13	63.87	30.67	66.95	29.83	67.79	15.41	12.18	64.29

2. SUMMARY OF OPERATIONAL PROCEDURES

2.1 ADMINISTRATION PROCEDURES

The Spring 2021 I AM administration window for all subjects opened on April 5, 2021, and closed on May 14, 2021. Key personnel included the Corporation Test Coordinators (CTCs), School Test Coordinators (STCs), and TAs who proctored the test. A *Test Administrator's Manual (TAM)* and *Test Coordinator's Manual (TCM)* were 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.

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 features 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 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 which ensure equitable access to instructional and assessment content and generate valid assessment results for students who need such accommodations. Embedded accommodations (e.g., color contrast, print size) are provided digitally through instructional or assessment technology; non-embedded designated features (e.g., assistive technology to magnify/enlarge) are non-digital. Students who require 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).

The 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, American Sign Language (ASL) interpreter, streamlined mode, assistive technology (e.g., adaptive keyboards, touchscreen, switches), calculation devices, and multiplication tables.

Tables 7–14 list the number of students who are recorded in the Test Information Distribution Engine (TIDE) as receiving each accommodation during the Spring 2021 test administration.

Table 7: Total Students with Allowed Embedded and Non-Embedded Accommodations: ELA

Accommodations	Grade						
	3	4	5	6	7	8	10
Embedded Accommodations							
Permissive Mode	99	114	98	124	120	141	124
Streamlined Mode	22	21	20	34	29	46	20
Non-Embedded Accommodations							
Alternate Indicator of Response	448	518	502	537	498	605	361
Print Booklet	28	15	29	33	23	35	27
Large Print Booklet	14	8	12	21	8	16	19
Read Aloud Script for Paper Booklet	33	20	33	35	24	40	29
Braille Booklet	1						1
Read Aloud to Self	4	5	7	11	5	8	6
Human Reader for All Items, Including Reading Comprehension	183	219	180	221	231	270	280
ASL Interpreter	2	2		2	7	7	11
Student Provided with Additional Breaks	529	598	608	664	656	791	707
Student Provided Access to Own Resources	96	117	122	161	141	177	143
Special Request							
Non-Standard Accommodation (approved by IDOE)							

Table 8: Total Students with Allowed Embedded and Non-Embedded Designated Features: ELA

Designated Features	Grade						
	3	4	5	6	7	8	10
Embedded Designated Features							
Translations							
Masking	623	690	725	819	856	1022	1004
Mouse Pointer	4		4	4	1	6	3
Print Size	3	4	3	4	2	5	4
Color Contrast	3	2	2	1	2		
Non-Embedded Designated Features							
Color Acetate Film for Paper Assessment			1	5	1	1	3
Assistive Technology to Magnify/Enlarge	19	17	24	28	22	27	45
Access to Sound Amplification System	6	17	6	7	12	14	10
Special Furniture or Equipment for Viewing Test	77	54	58	58	42	53	52
Special Lighting Conditions	19	23	18	20	15	19	23
Time of Day for Testing Altered	96	114	127	129	121	162	99

Table 9: Total Students with Allowed Embedded and Non-Embedded Accommodations: Mathematics

Accommodations	Grade						
	3	4	5	6	7	8	10
Embedded Accommodations							
Permissive Mode	98	113	96	121	120	141	121
Streamlined Mode	23	21	18	33	29	45	19
Non-Embedded Accommodations							
Alternate Indicator of Response	447	513	498	532	497	604	356
Handheld/Adaptive Calculator	190	257	350	476	620	764	770
Multiplication Table	122	175	236	299	341	354	214
Print Booklet	29	15	29	31	23	35	28
Large Print Booklet	14	8	12	21	8	16	19
Read Aloud Script for Paper Booklet	33	20	34	33	23	40	31
Hundreds Chart	233	289	301	332	322	342	164
Braille Booklet	1						1
Read Aloud to Self	4	4	7	11	6	8	6
Bilingual Word to Word Dictionary	1	2	1	4	3	4	10
ASL Interpreter	2	2		2	7	7	10
Student Provided with Additional Breaks	527	592	603	660	655	790	697
Student Provided Access to Own Resources	98	117	121	162	145	178	141
Special Request							
Non-Standard Accommodation (approved by IDOE)							

Table 10: Total Students with Allowed Embedded and Non-Embedded Designated Features: Mathematics

Designated Features	Grade						
	3	4	5	6	7	8	10
Embedded Designated Features							
Translations							1
Masking	623	684	720	811	852	1022	992
Mouse Pointer	4	1	4	4	1	4	4
Print Size	4	4	4	4	3	3	5
Color Contrast	2	1	2	1	2		
Non-Embedded Designated Features							
Color Acetate Film for Paper Assessment			1	5		1	3
Assistive Technology to Magnify/Enlarge	18	17	23	25	22	27	43
Access to Sound Amplification System	6	17	6	8	12	14	10
Special Furniture or Equipment for Viewing Test	76	54	55	54	43	53	49
Special Lighting Conditions	19	22	17	18	15	19	23
Time of Day for Testing Altered	94	114	125	126	122	161	97

Table 11: Total Students with Allowed Embedded and Non-Embedded Accommodations: Science

Accommodations	Grade		
	4	6	Biology
Embedded Accommodations			
Permissive Mode	113	121	133
Streamlined Mode	21	33	21
Non-Embedded Accommodations			
Alternate Indicator of Response	510	530	414
Paper Test Booklet	13	31	37
Large Print Booklet	9	20	21
Read Aloud Script for Paper Booklet	19	32	39
Braille Booklet			1
Read Aloud to Self	4	11	6
Bilingual Word to Word Dictionary	2	4	10
ASL Interpreter	2	2	7
Student Provided with Additional Breaks	586	659	710
Student Provided Access to Own Resources	119	163	130
Special Request			
Non-Standard Accommodation (approved by IDOE)			

Table 12: Total Students with Allowed Embedded and Non-Embedded Designated Features: Science

Designated Features	Grade		
	4	6	Biology
Embedded Designated Features			
Translations			
Masking	678	805	1026
Mouse Pointer	1	3	5
Print Size	4	4	2
Color Contrast	1	1	
Non-Embedded Designated Features			
Color Acetate Film for Paper Assessment		5	
Assistive Technology to Magnify/Enlarge	17	27	33
Access to Sound Amplification System	17	9	9
Special Furniture or Equipment for Viewing Test	52	56	55
Special Lighting Conditions	22	19	20
Time of Day for Testing Altered	113	127	104

Table 13: Total Students with Allowed Embedded and Non-Embedded Accommodations: Social Studies

Accommodations	Grade
	5
Embedded Accommodations	
Permissive Mode	95
Streamlined Mode	19
Non-Embedded Accommodations	
Alternate Indicator of Response	493
Print Booklet	29
Large Print Booklet	12
Read Aloud Script for Paper Booklet	34
Braille Booklet	
Read Aloud to Self	7
Bilingual Word to Word Dictionary	1
ASL Interpreter	
Student Provided with Additional Breaks	599
Student Provided Access to Own Resources	121
Special Request	
Non-Standard Accommodation (approved by IDOE)	

Table 14: Total Students with Allowed Embedded and Non-Embedded Designated Features: Social Studies

Designated Features	Grade
	5
Embedded Designated Features	
Translations	
Masking	714
Mouse Pointer	4
Print Size	3
Color Contrast	2
Non-Embedded Designated Features	
Color Acetate Film for Paper Assessment	1
Assistive Technology to Magnify/Enlarge	24
Access to Sound Amplification System	6
Special Furniture or Equipment for Viewing Test	56
Special Lighting Conditions	17
Time of Day for Testing Altered	124

3. ITEM BANK AND TEST CONSTRUCTION

3.1 OVERVIEW OF ITEM DEVELOPMENT

All operational items used on the I AM test forms are drawn from the I AM item bank. Volume 2 is a separate, stand-alone report containing complete details on the I AM item bank. The I AM item bank is comprised of Indiana-developed items. The parameter estimation of I AM items is described in Section 5 of the present volume, Item Calibration and Establishing the I AM Scale.

3.2 TEST DESIGN

The I AM assessments were designed to be stage adaptive, with all students administered common test segments in Part 1 and subsequent routing to Part 2 tiered segments on the basis of Part 1 ability estimates.

Part 1 consisted of the following four sections:

- Segment 1: 2 practice items confirming the student's participation in the practice test
- Segment 2: 3 operational items of varying complexity
- Segment 3: 17 operational items of varying complexity
- Segment 4: 6 embedded field-test (EFT) items in an EFT block

Students who completed the first five questions (i.e., Segments 1 and 2) but had No Response (NR) for all five questions were recorded as No Mode of Communication (NMC). In this case, the TDS provided a pop-up message to show that students were identified as NMC and stopped their tests. Segments 3 and 4 of the assessment are administered to students as a single section of items, providing a seamless student testing experience which aids to obtain high-quality responses to both the EFT and operational items.

In Part 2, more targeted items were administered to the student on the basis of his or her performance in Part 1. There were three stage-adaptive segments of Part 2: Form A (low complexity), Form B (moderate complexity), and Form C (high complexity). Segment 5 corresponded to Form A, Segment 6 corresponded to Form B, and Segment 7 corresponded to Form C. Each form in Part 2 included 12 operational items.

Each I AM test included 32 operational items that were used for scoring as well as 2 practice items and 6 EFT items.

3.3 FIELD TESTING

The Spring 2021 I AM test forms contained new field-test items. The EFT slots (in paper-and-pencil tests) or segments (in online tests) were located with fixed positions across all subjects, such that item location and motivation effects, if they existed, did not propagate

into the estimates of the item parameters. To obtain high-quality responses to the EFT items, it was necessary that students remain unaware of which items were operational and which were EFT.

The EFT items were administered by using one of the EFT blocks, which included six field-test items. Table 15 shows the number of EFT blocks and items per grade. Three EFT blocks for grades 3–6 and four EFT blocks for grades 7, 8, and 10 were constructed. For the online assessments, one of the EFT blocks was randomly administered to each of the students. For the paper tests, one of the EFT blocks was fixed for all students in each of the grade and subject-area tests.

Table 15: Number of EFT Blocks and Items

Grade	Number of EFT Blocks	Number of EFT Items (6 items in each block)
3	3	18
4	3	18
5	3	18
6	3	18
7	4	24
8	4	24
10	4	24

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.

3.4 OPERATIONAL FORM CONSTRUCTION

Operational test form development (see Volume 2) includes an array of item types used to measure the Content Connectors. Table 16 describes the item types in the I AM pool, and Table 17 shows the number of items by item type used in the operational forms. In the operational forms, only multiple-choice items were administered except for grade 6 science, which had one multiple-select item. Descriptions and examples of each of the item types are also provided in Volume 2.

Table 16: 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.

* Note that the abbreviations MC, MS, and MI correlate to the attributes used in CAI's Item Tracking System (ITSx).

Table 17: Operational Items by Item Type and Grade

Subject	Item Type	3	4	5	6	7	8	10	Biology
ELA	MC	44	44	44	44	44	44	44	
Mathematics	MC	43	44	44	44	44	44	44	
Science	MC		43		42				44
Science	MS				1				
Social Studies	MC			44					

4. CLASSICAL ANALYSES OVERVIEW

4.1 CLASSICAL ITEM ANALYSES

Classical item statistics are based on the classical test theory framework and have been widely applied to examine whether test items function as intended. The criteria for flagging and reviewing items are provided in Table 18, and a description of the statistics is provided in the following subsections. All of the field test items administered in Spring 2021 I AM were MC items. The flagged items from the field tests were reviewed in the item data review. Except for the replacement of one rejected item in grade 4 ELA, IDOE elected to re-field test all field-test items administered in Spring 2021 in Spring 2022. Classical item statistics, the number of flagged items, and differential item functioning (DIF) categories for field-test items are provided in Appendices A.

Table 18: Thresholds for Flagging Items in Classical Item Analysis

Analysis Type	Flagging Criteria
Item Discrimination	Biserial correlation statistic is less than .25 for MC items.
Distractor Analysis	Biserial correlation statistic is greater than .00 for MC item distractors. Proportion of students responding to a distractor exceeds the proportion responding to a keyed response for MC items.
Item Difficulty (MC items)	Proportion correct value is less than .25 or greater than .95 for MC items.
Inverted Mean Total Score	Mean total score for a lower score point exceeds the mean total score for a higher score point for multi-point non-MC items.

4.1.1 Item Discrimination

The item discrimination index indicates the extent to which each item differentiates between those test takers who possessed the skills being measured and those who did not. In general, the higher the value, the better the item was able to differentiate between high- and low-achieving students. The discrimination index was calculated as the correlation between the item score and the ability estimate for students.

4.1.2 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.

4.1.3 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. Conventional item *p*-values are summarized in Section 4.3, Item Analyses Results.

4.2 DIFFERENTIAL ITEM FUNCTIONING ANALYSIS

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 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 2021.

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 χ^2 DIF statistics for balancing the stability and sensitivity of the DIF scoring category selection. The analysis program computes the MH χ^2 value, the conditional odds ratio, and the MH-delta for dichotomous items; the GMH χ^2 and the standardized mean difference (SMD) are computed for polytomous items.

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

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

where $k = \{1, 2, \ldots K\}$ 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

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

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 $var(a_k)$, a $(T-1) \times (T-1)$ variance matrix, are calculated analogously to the corresponding elements in MH χ^2 , in stratum k.

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

$$SMD = \sum_{k} \quad p_{FK} m_{FK} - \sum_{k} \quad p_{FK} m_{RK},$$

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 19. 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, or female) or negative DIF (i.e., –A, –B, or –C), signifying that the item favored the reference group (e.g., White or 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 19: DIF Classification Rules

Dichotomous Items	
Category	*Rule*
C	MH_{X^2} is significant, and $\left\|\hat{\Delta}_{MH}\right\| \geq 1.5$.
B	MH_{X^2} is significant, and $1 \leq \left\|\hat{\Delta}_{MH}\right| < 1.5$.
A	MH_{X^2} is not significant, or $\left\|\hat{\Delta}_{MH}\right| < 1$.

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 those 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 A. Only the items that met the minimum counts ($n = 50$) for both focal and reference groups were included in the DIF analysis.

5. ITEM CALIBRATION AND ESTABLISHING THE I AM SCALE

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.

5.1 ITEM RESPONSE THEORY MODELS

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:

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

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 (θ_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

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

5.2 IRT Analyses Results

IDOE elected to readminister the Spring 2021 I AM field-test items in the field tests in Spring 2022. The IRT analysis of the Spring 2021 field-test items was postponed until after the Spring 2022 I AM administration. Summaries of the IRT parameters for the operational items are provided in the following section.

5.2.1 IRT Summaries

The IRT statistical properties of the I AM operational test forms are summarized in Tables 20–23.

Table 20: 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	−1.07	−0.92	−0.37	0.09	0.29	0.52	1.11
4	44	−1.57	−1.06	−0.59	−0.02	0.25	0.65	1.60
5	44	−1.77	−1.29	−0.68	−0.23	0.17	1.02	1.44
6	44	−1.55	−1.16	−0.40	0.00	0.36	0.80	1.00
7	44	−1.53	−1.01	−0.35	−0.04	0.38	0.57	1.59
8	44	−1.57	−1.19	−0.73	−0.21	0.09	0.84	1.43
10	44	−1.54	−1.39	−0.71	−0.16	0.14	0.88	1.09

Table 21: Operational Item Parameter Five-Point Summary and Range: Mathematics

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
3	43	−1.37	−1.13	−0.53	−0.05	0.29	0.95	1.38
4	44	−2.37	−0.74	−0.37	−0.13	0.30	1.17	2.00
5	44	−1.31	−0.80	−0.48	−0.10	0.36	0.74	0.80
6	44	−1.66	−1.09	−0.62	−0.05	0.38	0.85	1.66
7	44	−1.64	−0.68	−0.37	0.03	0.44	0.70	0.99
8	44	−1.17	−0.84	−0.46	−0.17	0.12	0.94	1.06
10	44	−1.20	−0.85	−0.37	0.08	0.49	0.97	1.04

Table 22: Operational Item Parameter Five-Point Summary and Range: Science

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
4	43	−1.51	−1.15	−0.50	0.03	0.51	1.07	1.85

Grade	Total	Min	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Max
6	43	−2.23	−1.15	−0.46	0.02	0.43	0.92	1.23
Biology	44	−1.62	−1.34	−0.69	0.10	0.33	1.26	1.38

Table 23: 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	−0.94	−0.44	−0.07	0.30	0.52	0.59

5.2.2 I AM Test Characteristic Curves (2021)

Another way to view the technical properties of the I AM test forms is via the test characteristic curves (TCCs). These plots are displayed in Appendix B with TCCs corresponding to the three stage-adaptive forms.

6. SCORING AND REPORTING

The I AM assessment measures the knowledge and skills students are expected to develop and demonstrate in the context of Indiana 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. This section describes the methodologies that the I AM assessment implements to estimate students' abilities and to convert those estimates into scores for reporting.

6.1 MAXIMUM LIKELIHOOD ESTIMATION

Students' abilities were estimated using maximum likelihood estimation (MLE). Parameter estimates were calibrated using the Rasch model for dichotomously scored items and Masters' partial credit model for polytomous items.

6.1.1 Likelihood Function

The likelihood function for generating the MLEs is based on a mixture of item types, including 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 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:

$$\frac{\partial lnL(\theta)^{MC}}{\partial \theta} = \sum_{i=1}^{N} \left\{ x_i - \left[\frac{1}{1 + exp[-(\theta - b_i)]} \right] \right\}$$

$$\frac{\partial lnL(\theta)^{CR}}{\partial \theta} = \sum_{i=1}^{N} \left\{ x_i - \left[\frac{\sum_{j=1}^{m_i} j exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right] \right\}$$

$$\frac{\partial^2 lnL(\theta)^{MC}}{\partial \theta^2} = -\sum_{i=1}^{N} \left(1 - \left[\frac{1}{1 + exp[-(\theta - b_i)]} \right] \right) \left[\frac{1}{1 + exp[-(\theta - b_i)]} \right]$$

$$\frac{\partial^2 lnL(\theta)^{CR}}{\partial \theta^2} = \sum_{i=1}^{N} \left[\left(\frac{\sum_{j=1}^{m_i} j exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right)^2 - \left(\frac{\sum_{j=1}^{m_i} j^2 exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})}{1 + \sum_{j=1}^{m_i} exp \sum_{k=1}^{x_i} (\theta - \delta_{ki})} \right) \right]$$

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

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

where

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

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

and where θ_t denotes the estimated θ at iteration t.

6.1.3 Standard Errors of Estimates

The standard error of the MLE is estimated by

$$SE(\theta_j) = \frac{1}{\sqrt{I(\theta_j)}},$$

where $I(\theta_j) = -\left(\frac{\partial^2 lnL(\theta)}{\partial^2 \theta} \middle| \theta = \theta_j \right)$ is the Fisher information at the MLE and is calculated as follows:

$$\frac{\partial^2 log([p(\theta)]^{Z_i}[q(\theta)]^{1-Z_i})}{\partial^2 \theta} = -D^2 p_i(\theta) q_i(\theta),$$

where

$$q_i = 1 - p_i.$$

In general, the second derivate for the *i*th Masters' partial credit model item is

$$\frac{\partial^2 log\big(P(\theta)\big)}{\partial \theta^2} = D^2 \frac{\left[\sum_{j=1}^{m_i} j exp\big(\sum_{k=1}^{j} D(\theta - b_{ki})\big)\right]^2}{\left[1 + \sum_{j=1}^{m_i} exp \sum_{k=1}^{j} D(\theta - b_{ki})\right]^2} - D^2 \frac{\left[\sum_{j=1}^{m_i} j^2 exp\big(\sum_{k=1}^{j} D(\theta - b_{ki})\big)\right]}{\left[1 + \sum_{j=1}^{m_i} exp \sum_{k=1}^{j} (\theta - b_{ki})\right]}.$$

6.1.4 Extreme Case Handling

When students answer all items correctly or all items incorrectly, the likelihood function is unbounded and an MLE cannot be generated. For I AM scoring, the 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.

The truncated lowest-observable theta scores (LOT) and highest-observable theta scores (HOT) and the associated scale scores for each grade and subject area are provided in Table 24.

Table 24: Theta and Corresponding Scaled-Score Limits for Extreme Ability Estimates

Subject	Grade	Lowest-Observable Theta	Highest-Observable Theta	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

6.1.5 Standard Error of LOT/HOT Scores

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 6.1.3, Standard Error of Estimates, $\hat{\theta}$ = LOT or HOT, and difficulties (*b*) are for the administered items.

6.2 TRANSFORMING THETA SCORES TO REPORTING SCALE SCORES

For the Spring 2021 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 IRT ability estimate using the scaling constants a and b shown in Table 25:

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

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

Table 25: Scaling Constants on the Reporting Metric

Subject	Grade	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 is provided in Appendix C.

6.3 OVERALL PERFORMANCE CLASSIFICATION

Each student was assigned an overall performance category in accordance with his or her overall scale score. Tables 26–29 provide the scale score range of performance standards for ELA, Mathematics, Science, and Social Studies. The lower bound of Level 3 (At Proficiency) marks the minimum cut score for proficiency.

Table 26: Proficiency Levels for ELA by Grade

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
3	1300–1463	1464–1481	1482–1700
4	1300–1478	1479–1497	1498–1700
5	1300–1474	1475–1488	1489–1700
6	1300–1466	1467–1486	1487–1700
7	1300–1485	1486–1497	1498–1700
8	1300–1464	1465–1490	1491–1700
10	1300–1467	1468–1505	1506–1700

Table 27: Proficiency Levels for Mathematics by Grade

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
3	2300–2462	2463–2473	2474–2700
4	2300–2461	2462–2478	2479–2700
5	2300–2459	2460–2470	2471–2700
6	2300–2461	2462–2477	2478–2700
7	2300–2466	2467–2477	2478–2700
8	2300–2463	2464–2474	2475–2700
10	2300–2470	2471–2484	2485–2700

Table 28: Proficiency Levels for Science by Grade

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
4	3300–3475	3476–3496	3497–3700
6	3300–3465	3466–3488	3489–3700
Biology	3300–3471	3472–3502	3503–3700

Table 29: Proficiency Levels for Social Studies by Grade

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
5	4300–4488	4489–4499	4500–4700

6.4 REPORTING CATEGORY SCORES

Reporting category scores were computed using all items for scoring within each reporting category for categories that have at least a minimum of seven items in the blueprint. The reporting category scores were computed as a percent-correct score for each student and an average percent-correct score for aggregate reporting. Summaries of the scores for each reporting category by demographic groups as well as for all students is provided in Appendix D.

7. 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 configuration
- 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.

7.1 QUALITY ASSURANCE IN TEST CONSTRUCTION

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 the FormBuilder software to help construct operational forms.

Immediately upon generation of a test form, the FormBuilder 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 particular 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
- The answer options are lettered with A, B, and C
- 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 matches 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.
- The 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 criteria listed above. 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 uploads the item list into FormBuilder. After operational forms were defined in FormBuilder, all test maps, key files, and conversion tables were produced directly from FormBuilder to eliminate the possibility of human error in the construction of these important files. Test maps, key files, conversion tables, and other critical documents were generated directly from information maintained in 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 opportunities for errors.

Test maps 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 FormBuilder was customized to I AM.

As a further layer of QA for printed test booklets, both during the blueline 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 FormBuilder for content and item ordering to ensure that no changes to the form were introduced prior to printing.

7.2 Quality Assurance in Test Configuration

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 the scoring specifications were applied accurately. The simulated data file is scored independently by two programmers, following the 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, cut scores, and the item information (cut scores, item attributes, item parameters, passage information, etc.). The accuracy of the information in the configuration file is checked and confirmed numerous times independently by multiple staff members prior to the testing window.

7.3 Quality Assurance in Computer-Delivered Test Production

7.3.1 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. 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 the state for testing purposes.
4. The final system is deployed to a staging environment accessible to the Indiana Department of Education (IDOE) for user acceptance testing (UAT) and final review.

7.3.2 Web Approval of Content During Development

The ITSx 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 see it. Layout changes can be made after this process in two ways:

1. Content can be revised and re-approved for web display; and
2. Online style sheets can change 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. Below, 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.

7.3.3 Platform Review

Platform review is a process in which each item is checked to ensure that it is displayed appropriately on each tested 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 platforms that are significantly different from one another.

Platform review is conducted by an internal team within CAI. The team leader projects the item as it was web approved in ITSx, and team members, each behind a different platform, look at the same item to see that it renders as expected.

7.3.4 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 user acceptance testing (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 test scenarios and constant support during the UAT period. Identified issues will be 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 will sign off to open the administration window at the conclusion of UAT activities.

Deployments to the production environment all 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 tested on the testing and staging servers. Upon completion of each deployment project, management approves the deployment log.

During the year, some changes may be required to the production system. 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. If any proposed change will affect client functionality or pose risk to 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 USBE with an opportunity to interact with the exact test with which the students will interact.

7.3.5 Functionality and Configuration

The items, both in themselves 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; and
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 is 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.

7.4 Quality Assurance in Data Preparation

When a student responds to test questions online, his or her 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. Our 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 programmer lead 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 ITS.

CAI's TDS has a real-time quality-monitoring component built in. As students test, data flow through our Quality Monitor (QM) software. QM 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. QM scores the test, recalculates performance-level designations, calculates subscores, compares item parameters to the reference item parameters in the bank, and conducts a host of other checks.

QM also aggregates data to detect problems that become apparent only in the aggregate. For example, QM 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 that are done for data review, but (a) they are done on operational data, and (b) they are conducted in real time so that our psychometricians can catch and correct any problems before they have an opportunity to do any harm.

Data pass directly from the QM 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 ODE and their QA contractor. CAI psychometricians ensure that data in the extract files match the DOR prior to delivery to the ODE.

7.5 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.

7.5.1 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 verify the accuracy of the Reporting System, we merge item response data with the demographic information taken either from previous year assessment data, or if current year enrollment data is available by the time simulated data files are created, we can verify online reporting using current year testing information. By populating the simulated data files with real school information, it is possible to verify that special school types and special districts are being handled properly in the Reporting System.

Specifications for generating simulated data files are included in the analysis output student data file specifications document submitted to IDOE each year. Review of all simulated data is scheduled to be completed prior to the opening of the test administration, so that the integrity of item administration, data capture, and item and test scoring and reporting can be verified before the system goes live.

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 test 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 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 item response theory (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 to generate reports based on all items in the pool.

Item* p-*Value. For multiple-choice items, the proportion of students selecting each response option is computed; for constructed-response, performance, and technology items, the proportion of student responses classified at each score point is computed. For multiple-choice items, if the keyed response is not the modal response, the item is also flagged. Although the correct response is not always the modal response, keyed

response options flagged for both low biserial correlations and non-modal response 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 observed and expected score given the student's ability is computed for each item. The residuals for each are averaged across all students, and the average residual is used to flag an item.

7.5.2 Quality Assurance in Reporting

Scores for the I AM online assessments are assigned by automated systems in real time. For machine-scored portions of assessments, 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, 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 the Reporting System and Database of Record (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 System, 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 the Reporting System, 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 MLEs using the procedures described within this report. Additionally, HumRRO provides replication of the psychometric scoring process. Scores are approved and published by IDOE only when all three independent systems match.

8. REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999, rev. 2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (ETS Research Report No. 91-47). Princeton, NJ: Educational Testing Service.

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). Hillsdale, NJ: Lawrence Erlbaum.

Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 531–578). Westport, CT: American Council on Education/Praeger.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–173.

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.

Muniz, J., Hambleton, R. K., & Xing, D. (2001). Small sample studies to detect flaws in item translations, *International Journal of Testing, 1*(2), 115–135.

Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning, *Educational Research and Evaluation, 19*(2–3), 170–187.

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician*, *40*(2):106–108.

van der Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

Indiana's Alternate Measure

2020-2021

Volume 2
Test Development

ACKNOWLEDGMENTS

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

Major contributors to this technical report include the following staff from American Institutes for Research: Stephan Ahadi, Hyesuk Jang, Yuan Hong, Gabriel Martinez, Hashim Evans, Cherise Lesesne, and Celine Bryan. Major contributors from the Indiana Department of Education include the Assessment Director, Assistant Assessment Director, and Program Leads.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

Indiana's Alternate Measure (I AM) is the summative accountability assessment for students with significant cognitive disabilities in grades 3–8 and high school. The I AM assessment is a stage-adaptive assessment. Performance on Part 1 of the assessment determines placement into one of three forms for Part 2: Form A (low complexity); Form B (moderate complexity); or Form C (high complexity). The assessment measures student achievement and growth according to Indiana's Alternate Standards, called Content Connectors, which are aligned to and derived from the Indiana Academic Standards (IAS). I AM is assessing the following:

- English/Language Arts: Grades 3–8 and 10;
- Mathematics: Grades 3–8 and 10;
- Science: Grades 4 and 6, and Biology End-of-Course Assessment (ECA) (high school); and
- Social Studies: Grade 5.

In June of 2018, the Indiana State Board of Education approved the adoption of new Content Connectors for English/Language Arts (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 appropriately aligned to the IAS and was readily available to teachers, parents, and students across the state. Alternate standards are necessary to ensure all students have access to grade-level-aligned content and to achieve educational accountability.

1.1 CLAIM STRUCTURE

The I AM assessments are designed to support the following claims about proficiency:

Level 1 – Below Proficiency

Indiana students below proficiency have not met current grade-level Content Connectors. Students may require significant support to develop the knowledge, application, and skills necessary to be on track for post-secondary education or competitive integrated employment.

Level 2 – Approaching Proficiency

Indiana students approaching proficiency have nearly met current grade-level Content Connectors by demonstrating some basic knowledge, application, and skills. Students may require some support in order to be on track for post-secondary education or competitive integrated employment.

Level 3 – At Proficiency

Indiana students at proficiency have met current grade-level Content Connectors by demonstrating essential knowledge, application, and skills necessary to be on track for postsecondary education or competitive integrated employment.

1.2 Organization of This Volume

This volume is organized into three sections:

- An overview of the I AM blueprint development process that supports the validity of the claims I AM assessments are designed to support.
- An overview of the item development process that supports the validity of the claims I AM assessments are designed to support.
- An overview of the test construction process for the I AM assessments that supports the validity of the claims I AM assessments are designed to support.

2. I AM Assessment Blueprints

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 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 in this volume in Section 1.1, Claim Structure.

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.

2.1 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 I AM assessments. Blueprints are the assessment design specifications that ensure assessment scores support the performance-level descriptors (PLDs) described in Section 1.1, Claim Structure. 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 is meaningful and appropriate for students with significant cognitive disabilities. To meet these requirements, the 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 Connect Connectors for the I AM population. Participants then disseminated 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, the 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 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 determined 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 demonstrating 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 was 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:

- Measure the breadth and depth of Indiana Content Connectors, aligned to and derived from the IAS;
- Provide weight to the Content Connectors and reporting categories as identified by educators;
- Produce accurate and precise test scores and performance-level classifications;
- Meet required item count limits; and
- Remain consistent related to measurable content across test administrations.

2.1.1 ELA Blueprints

The I AM blueprints developed for ELA grades 3–8 and 10 are provided in Appendix A: I AM 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 the 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 enough 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 (i.e., 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.

2.1.2 Mathematics Blueprints

The I AM blueprints developed for Mathematics grades 3–8 and 10 are provided in Appendix B: I AM 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 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.

The I AM Mathematics blueprints also include Content Connectors in a category that is reported as an aggregate score. The items assessing those 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 on each I AM Mathematics assessment is 32.

2.1.3 Science Blueprints

The I AM blueprints developed for Science grades 4 and 6 and Biology are provided in Appendix C: I AM 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 the 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 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 on each I AM Science assessment is 32.

2.1.4 Social Studies Blueprint

The I AM blueprint developed for Social Studies grade 5 is provided in Appendix D: I AM Social Studies Blueprint. 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 the 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 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 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 each on the I AM Social Studies assessment is 32.

2.1.5 Test Length

All I AM assessments include 32 operational items. The 2021 I AM test design also included 6 embedded field-test (EFT) items placed into fixed positions at the end of Part 1 (segment 4). All test forms contained fixed operational items. 1 shows the number of operational items, the number of EFT items, and the total number of items administered on each 2021 assessment.

Table 1: Observed Spring 2021 Test Length by Grade and Subject

Subject	Grades	Number of Practice Items	Number of Operational Items	Number of EFT Items Per Test	Total Items Per Test
ELA	3–8, 10	2	32	6	40
Mathematics	3–8, 10	2	32	6	40
Science	4 & 6 & Biology ECA	2	32	6	40
Social Studies	5	2	32	6	40

2.1.6 Reporting Category Percentages

The blueprint is designed to support reporting at the Reporting Category level in addition to the overall test score. Individual scores for each Reporting Category provide information to help identify areas in which a student may have had difficulty.

Tables 2–5 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 2: Blueprint Percentage of Items Assessing Each Reporting Category in ELA

Grade	Reporting Category			
	Key Ideas and Textual Support/Vocabulary	***Structural Elements and Organization/Connection of Ideas/Media Literacy***	***Writing***	***Reading Foundations***
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
	Key Ideas and Textual Support/Vocabulary	***Structural Elements and Organization/Connection of Ideas/Media Literacy***	***Writing***	***Speaking and Listening (Aggregate Only)***
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 3: Blueprint Percentage of Items Assessing Each Reporting Category in Mathematics

Grade	Reporting Category				
	Algebraic Thinking and Data Analysis	***Computation***	***Geometry and Measurement***	***Number Sense***	***Process Standards (Aggregate Only)***
3	22–25%	22–25%	22–25%	22–25%	6–12%

4	22–25%	22–25%	22–25%	22–25%	6–12%
	Algebraic Thinking	***Computation***	***Geometry and Measurement, Data Analysis, and Statistics***	***Number Sense***	***Process Standards (Aggregate Only)***
5	22–25%	22–25%	22–25%	25–28%	3–12%
	Algebra and Functions	***Computation***	***Geometry and Measurement, Data Analysis, and Statistics***	***Number Sense***	***Process Standards (Aggregate Only)***
6	25–28%	22–25%	22–25%	25–28%	3–12%
	Algebra and Functions	***Data Analysis, Statistics, and Probability***	***Geometry and Measurement***	***Number Sense and Computation***	***Process Standards (Aggregate Only)***
7	25–28%	22–25%	22–25%	22–25%	3–6%
8	28–31%	22–25%	22–25%	22–25%	3–6%
	Equations and Inequalities (Linear and Systems)	***Functions (Linear and Non-linear)***	***Geometry and Measurement***	***Number Sense and Data Analysis***	***Process Standards (Aggregate Only)***
10	22–25%	22–25%	22–25%	22–25%	3–12%

Table 4: Blueprint Percentage of Items Assessing Each Reporting Category in Science

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

Table 5: Blueprint Percentage of Items in Assessing Each Reporting Category in Social Studies

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

3. I AM ITEM POOL AND DEVELOPMENT

In order 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 field tested in 2019, and newly-developed, custom EFT items.

3.1 OPERATIONAL ITEMS

In order to support blueprint and test design requirements as new items for the I AM 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 2020-2021 I AM assessments. Items were also evaluated and selected for alignment to the new 2018 I AM item specifications when possible. However, because variance exists between the item specifications in use when the legacy operational items were developed and the new 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 2021 operational use to achieve better alignment of the new item specifications for I AM assessments. More information about how legacy operational items were equated onto the I AM scale can be found in Section 5.2, Establishing the I AM Bank, of the Spring 2019 Technical Report, volume 1.

3.2 FIELD-TEST ITEMS

In order to begin growing the I AM operational pool, CAI and IDOE developed new items for field testing based on blueprint needs and 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 IAS. A second analysis was completed after 2019 I AM testing. Based on these analyses, CAI created I AM item development plans and developed new items that targeted the depth and breadth of coverage required by the test blueprints, with the intent to grow the pool over time. Beginning in 2020, IDOE created I AM item development plans and worked with Indiana Educators to develop new 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 Item Tracking System (ITS), an auditable content-development tool with a built-in workflow and 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.

3.2.1 Item Types

The majority of I AM items are Multiple-Choice (MC). Five operational I AM Science items are of a different type: Four are Multiple-Select (MS) items and one is a Table Match item (MI).

Table 6 Table 1lists the I AM item types in the I AM item bank and provides a brief description of each. Examples of each item type can be found in Appendix E: Item Type Examples.

Table 6: I AM Item Types and Descriptions

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

*Note that the abbreviations MC, MS, and MI correlate to the attributes used in CAI's Item Tracking System (ITS).

3.2.2 Underlying Principles Guiding Item Development

I AM item development is based on the needs formalized by the I AM assessment blueprints and is guided by detailed item specifications. The specifications, discussed in Section 3.2.4, 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 has received Web Content Accessibility Guidelines (WCAG) 2.0 AA certification, 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 measure 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 the items can support the claims that will be based on them. Table 7 describes how each step contributes to these goals. Each step in the process is discussed in more detail below the table.

Table 7: Summary of How Each Step of Development Supports the Validity of Claims

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 items that appear to function differently for subsequent review.	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 function.	

3.2.3 I AM 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 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 Reading[1];
- ATOS® Readability Formula;
- Flesch-Kincaid Grade Level; and
- 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 takes into account the most important predictors of text complexity, which are average sentence length, average word length, and word difficulty level. The results are provided in 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 composed 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 8 provides the quantitative specifications for I AM passages by grade for word count, Lexile range, Flesch-Kincaid range, and ATOS range.

Table 8: I AM Quantitative Passage Specifications

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

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

Table 9 **Error! Reference source not found.**provides the qualitative specifications for passages by tier.

Table 9: I AM Qualitative Passage Specifications

Tier 1	Tier 2	Tier 3
● **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.	● **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.	● **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.

3.2.4 I AM 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 3-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 done in the grade-level panels. Each break-out group presented their work for the full subject-area panel to review for consistency across the subject area. Modifications were made by the note-takers 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 (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 give 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 10 presents a sample ELA specification for one grade 3 Content Connector.

Table 10: 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 Standard	**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	Items must be passage based. Tier 1 and 2 items should avoid the word "best" in the stem. Tier 1 items should contain picture support in answer choices when possible to aid comprehension. Tier 2 items can contain picture support in answer choices. Tier 3 items should not contain picture support. Tier 1 distractors should demonstrate clearly incorrect understanding of events or details in the passage. Tier 2 distractors should be possible misunderstanding of events or details in the passage or unrelated details or events in the passage. Text complexity will increase with Tiers.
Recommended Response Mechanisms	Multiple Choice (MC) Table Match (TM) Multi-Select (MS)
Construct-Relevant Vocabulary	main idea
Cognitive Complexity	4
Evidence Statements	
Evidence Statements	**Tier 1** Students can identify a key detail in the text.
	Tier 2 Students can identify an explicitly stated main idea of the text.
	Tier 3 Students can determine the main idea of a text.
Accessibility and Accommodation Considerations	
Stimulus Graphic Limitations	Stimulus graphics will be limited to clear photos, illustrations, diagrams, tables, and charts that directly relate to the passage topic.

	Information contained within stimulus graphics is ineligible for assessment unless specifically prescribed by Content Connector and/or evidence statements.
Visual and Auditory Considerations	Graphics will be provided in formats that are accessible to students to understand or process information. Graphics that do not contribute to the student's understanding should not be included.
Sample Item	
Tier 3	[Stimulus: Passage about the history of telephones] Which sentence tells the main idea? A. No one uses telephones anymore. B. Telephones are a lot bigger than they used to be. **C. Telephones have changed a lot over the years.**

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 2020-2021 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 increases, with ongoing efforts being made to align I AM administrations solely to the 2018 I AM item specifications.

3.2.5 Training of Item Writers

All CAI item writers developing 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 G: Language, Accessibility, Bias, and Sensitivity Guidelines and Checklist.

3.3 INTERNAL REVIEW

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 our test developers used is included here as Appendix F: Item Review Checklist. The I AM Internal Review cycle includes five levels:

- Preliminary Review;
- Content Review 1;
- Special Education Review;
- Edit Review 1; and
- Senior Review 1.

3.3.1 Preliminary Review

Items are first written independently by test developers. For the 2020-2021 I AM administration, ELA items associated with literary, or nonfiction, passages were written after the passages had been reviewed and approved by educators (Passage Review is described in more detail in Section 3.4.2).

After items are written by test developers, the items undergo Preliminary Review. The 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 LAL 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.

3.3.2 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.

3.3.3 Accessibility Review

The accessibility 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 the Content Review 1 review level for revision.

3.3.4 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 consistency of style 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 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.

3.3.5 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.

3.4 Review by State Personnel and Stakeholder Committees

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.

3.4.1 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.

3.4.2 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 would forgo passage review as a separate step preceding item development. Passage Review is important for long passages with numerous associated items to make sure 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 2020-2021 I AM administration.

3.4.3 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.

A summary of the 2019 I AM Content and Fairness Committee Review meeting appears in Appendix H: 2019 I AM Content and Fairness Committee Review Summary.

3.5 Markup for Translation and Accessibility Features

After all approved state- and committee-recommended edits have been applied, the items are considered “locked” and ready for accessibility markup. 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.

Accessibility markups, whether for translations or text-to-speech, follow similar processes. A trained expert enters the markup. A second expert reviews the work and recommends changes if necessary. If there is disagreement, a third expert is engaged to resolve the conflict.

Currently, all I AM items are tagged with text-to-speech. I AM Mathematics, Science, and Social Studies are also tagged with Spanish translations.

3.6 FIELD TESTING

Newly developed I AM items are field tested as embedded field-test items in the I AM assessment. Field testing is described in detail in Volume 1, Section 3.3, of this technical report.

3.7 POST-FIELD-TEST REVIEW

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.

3.7.1 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.

3.7.2 Item Data Review

Volume 1**,** Sections 4.1 and 4.2, of this technical report describe 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; and
- Differential Item Functioning.

Item Quality and Performance

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 MS and table match items are flagged for item quality and performance if the correlation with test score is less than 25%.

Item Difficulty

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.

I AM MS and table match items are flagged for item difficulty if the percentage of students achieving ANY score point is greater than 95%.

Differential Item Functioning (DIF)

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.

Data Review Committee meetings were not conducted in 2020-2021 owing to the impact of COVID-19. Instead, CAI flagged items field tested in the Spring 2021 I AM administration and IDOE staff reviewed the item statistics. One item was rejected during this review and all other items will be field tested again in Spring 2022.

3.8 Strategy for Pool Evaluation and Replenishment

IDOE seeks to release approximately six items per grade and subject from the pool 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 in grades 3–6 supports the development and testing of 18 field-test items per year (six items each in three forms). The current I AM student population in grades 7–10 supports the development and testing of 24 field-test items per year (six items each in four 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; and
3. Tabulations of content coverage to identify gaps in the pool.

4. I AM Assessment Construction

4.1 Assessment Form Construction

During Fall 2019, CAI psychometricians and content experts worked with IDOE content specialists and leadership to build assessments for the Spring 2020 administration. After that test administration was cancelled due to the COVID-19 pandemic, IDOE instructed CAI to hold these assessments for administration in Spring 2021.

In this section, the processes used for assessment construction are described in order to support the claim that they are technically sound and consistent with current professional standards. These processes include the utilization of a structured test design plan and collaborative participation from all parties.

4.1.1 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).

Table 11 illustrates the I AM test design for forms in each grade and subject.

Table 11: I AM Test Design 2020-2021

I AM Test Design 2020-2021			
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	**Key**		
item 17	**Tier 1 item**		
item 18	**Tier 2 item**		
item 19	**Tier 3 item**		
item 20			

Part 1 is administered to all students. On both online and paper 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 Table 1 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 Table 11, 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 Table 1. For example, in grade 3 math Form C, a Tier 3 item was replaced by a Tier 1 item from Form A 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.

4.2 TEST FORM ASSEMBLY

As discussed previously, the I AM assessment blueprints describe the content to be covered and the allocations for Reporting Categories and Content Connectors. To assemble the 2021 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 (Appendix I).

4.2.1 Role of the CAI Content Team

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 OFT 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; and
- Revision of the forms to incorporate feedback from IDOE.

4.2.2 Role of the CAI Technical Team

The CAI technical team, which included psychometricians and statistical support associates, prepared the item bank by updating ITS with current item statistics and providing test construction training to the internal content team. During test construction, at least one psychometrician facilitated with each content area.

The technical team performed the following tasks:

- Preparing item bank statistics and updating of 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; and
- Creating statistical summary and materials for IDOE review.

4.2.3 Role of the IDOE

The 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.

4.3 BLUEPRINT MATCH

The I AM assessment blueprints are designed to support reporting at the Reporting Category level in addition to the overall test score. Individual scores for each Reporting Category provide information to help identify areas in which a student may have had difficulty.

Tables 12-16 provide the percentage of test items assessing each reporting category that appeared on the Spring 2021 forms.

Table 12: Observed Spring 2021 Percentage of Items Assessing Each Reporting Category in ELA

Grade	Reporting Category			
	Key Ideas and Textual Support/Vocabulary	***Structural Elements and Organization/Connection of Ideas/Media Literacy***	***Writing***	***Reading Foundations***
3	31%	25%	22%	22%
4	41%	34%	25%	N/A
5	44%	28%	28%	N/A
	Key Ideas and Textual Support/Vocabulary	***Structural Elements and Organization/Connection of Ideas/Media Literacy***	***Writing***	***Speaking and Listening (Aggregate Only)***

6	38%	34%	25%	3%
7	44%	28%	25%	3%
8	38%	31%	25%	6%
10	38%	34%	25%	3%

Table 13: Observed Spring 2021 Percentage of Items Assessing Each Reporting Category in Mathematics

Grade	Reporting Category				
	Algebraic Thinking and Data Analysis	***Computation***	***Geometry and Measurement***	***Number Sense***	***Process Standards (Aggregate Only)***
3	22%	22-25%	22-25%	25%	6%
4	22%	22-25%	22%	22%	9-12%
	Algebraic Thinking	***Computation***	***Geometry and Measurement, Data Analysis, and Statistics***	***Number Sense***	***Process Standards (Aggregate Only)***
5	22-25%	22%	25%	25-28%	3%
	Algebra and Functions	***Computation***	***Geometry and Measurement, Data Analysis, and Statistics***	***Number Sense***	***Process Standards (Aggregate Only)***
6	25%	22%	22%	28%	3%
	Algebra and Functions	***Data Analysis, Statistics, and Probability***	***Geometry and Measurement***	***Number Sense and Computation***	***Process Standards (Aggregate Only)***
7	25%	22%	22%	25%	6%
8	31%	22%	22%	22%	3%
	Equations and Inequalities (Linear and Systems)	***Functions (Linear and Non-linear)***	***Geometry and Measurement***	***Number Sense and Data Analysis***	***Process Standards (Aggregate Only)***
10	22-25%	25%	22%	22-25%	6%

Table 14: Observed Spring 2021 Percentage of Items Assessing Each Reporting Category in Science

Grade	Reporting Category			
	Analyzing, Interpreting, and Computational Thinking	***Explaining Solutions, Reasoning, and Communicating***	***Investigating***	***Questioning and Modeling***
4	22-25%	22-25%	22%	28-31%
6	22-25%	22%	28-31%	25%
	Analyzing Data and Mathematical Thinking	***Communicating Explanations and Evaluating Claims Using Evidence***	***Developing and Using Modeling to Describe Structure and Function***	***N/A***
Biology	41-44%	25%	31-34%	N/A

Table 15: Observed Spring 2021 Percentage of Items Assessing Each Reporting Category in Social Studies

Grade	Reporting Category		
	Civics and Government/History	***Economics***	***Geography***
5	54%	18%	28%

Table 16: Observed Spring 2021 Percentage of Items Assessing Each Reporting Category in Social Studies (post-IDR)

Grade	Reporting Category		
	Civics and Government/History	***Economics***	***Geography***
5	53%	25%	22%

In every case, the percentages across reporting categories on the Spring 2021 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. The Item Development Plans that determined blueprint coverage of the Spring 2021 I AM field test pool are provided in Appendices A-D for ELA, Mathematics, Science, and Social Studies, respectively.

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.

5. REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior, 19*(2), 135–145.

Fisher, D., Frey, N., & Lapp, D. (2012). *Text complexity: Raising rigor in reading*. Newark, DE: International Reading Association.

Flowers, C., Wakeman, S., Browder, D.M., & Karvonen, M. (2009). Links for academic learning (LAL): A conceptual model for investigating alignment of alternate assessments based on alternate achievement standards. *Educational Measurement: Issues and Practice*, 28(1), 25–37.

Freebody, P., & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, *15*(3), 19–39.

Gillioz, C., Gygax, P., & Tapiero, I. (2012). Individual differences and emotional inferences during reading comprehension. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66*(4), 239–250.

Kucer, S. B. (2010). Going beyond the author: What retellings tell us about comprehending narrative and expository texts. *Literacy, 45*(2), 62–69.

Long, D. L., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language, 42*(4), 545–570.

McConaughy, S. (1985). Good and poor readers' comprehension of story structure across different input and output modalities. *Reading Research Quarterly, 20*(2), 219–232. doi:10.2307/747757

Rapp, D. N., & Mensink, M. C. (2011). Focusing effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 141–164). Charlotte, NC: Information Age Publishing.

Rich, S. S., & Taylor, H. A. (2000). Not all narrative shifts function equally. *Memory & Cognition, 28*(7), 1257–1266.

Riding, R. J., & Taylor, E. M. (1976). Imagery performance and prose comprehension in seven-year-old children. *Educational Studies, 2*(1), 21–22.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience, 25*(5), 762–776.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). A causal model of sentence recall: Effects of familiarity, concreteness, comprehensibility, and interestingness. *Journal of Reading Behavior, 25*(1), 5–16.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43.

Sparks, J. R., & Rapp, D. N. (2011). Readers reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 230–247.

Taylor, S.E. (1989). *EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies.* Orlando, FL: Steck-Vaughn.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/Synthesis44.html.

Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states*. Washington, DC: Council of Chief State School Officers.

Indiana's Alternate Measure

2020–2021

Volume 3
Test Administration

ACKNOWLEDGMENTS

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

Major contributors to this technical report include the following staff from Cambium Assessment, Inc. (CAI) Stephan Ahadi, Hyesuk Jang, Yuan Hong, Hashim Evans, Celine Bryan and Cherise Lesesne. Major contributors from the Indiana Department of Education include the Assessment Director, Assistant Assessment Director, and Program Leads.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

1. INTRODUCTION

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 2020-2021, both stages of all tests were administered online just as they were during the first year of the administration. Spanish was also offered as an accommodation for the online administration, where the Spanish translation was stacked above the English content. 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 2020-2021 required the braille accommodation.

As specified in Standard 6.0 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), assessment instruments are required to have established test administration procedures to support useful interpretations of score results. This volume of the I AM Technical Report provides details on the testing procedures, accommodations, Test Administrator (TA) training and resources, 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 the 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

2. TESTING PROCEDURES AND TEST WINDOWS

Administering the 2020–2021 I AM 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 test window was April 5 through May 14, 2021.

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 I) and the *2020–2021 Indiana Accessibility and Accommodations Guidance* (Appendix S). Throughout the 2020-2021 school year, the TAM was available on the Indiana Assessment Portal and the *2020–2021 Indiana Accessibility and Accommodations Guidance* was available on the IDOE website.

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

All students were required to take a subject-specific practice test within the operational test environment prior to taking the Spring 2021 I AM operational assessment. Students who were administered the paper I AM form completed the practice test items in the paper-and-pencil test booklet. The practice tests contained sample test items and helped 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. The I AM RIR was deployed on October 5, 2020, allowing students access to the items online for seven months prior to the opening of the test window.

Before a student could enter the operational test environment to participate in the subject-specific practice test and operational test, TAs had to complete the Learner Characteristics Inventory (LCI) for each student. Details regarding the LCI are available in the *I AM Online Test Delivery System (TDS) User Guide* (Appendix O) and the *I AM Test Administrator's Manual* (TAM) (Appendix I).

I AM is a stage-adaptive assessment administered in two segments, where a student's answers in Segment 1 determined the next group of items presented to the student in Segment 2. Each I AM assessment included 32 operational items that were used for scoring, as well as 6 embedded field-test (EFT) items. The student's total score is based on performance from both segments of the assessment.

The I AM assessments were untimed and were delivered to students individually. Students could start and finish one segment of an assessment in a single day or over the

course of multiple days, if needed. TAs were advised, however, that students could not complete Segment 1 and begin Segment 2 on the same day.

2.1 ELIGIBLE STUDENTS

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 Corporation Test Coordinator (CTC).
- **English Learners (ELs).** All ELs participated in statewide assessments. ELs who were enrolled in school in the United States for less than one year could be exempt for one administration from the I AM ELA assessments if a student's EL team agreed that exemption was appropriate.
- **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 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.
- Goals listed in the IEP for this student are linked to the enrolled grade-level Alternate Achievement Standards (Indiana Content Connectors).

2.2 TESTING ACCOMMODATIONS

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 font 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.

The I AM assessments provided three categories of assessment features to students. These included universal tools, designated features, and accommodations. Section 3.2 in Volume 1 of this technical report lists the allowed accommodations and the number of

students who were provided with accommodations during the Spring 2021 test administration.

Error! Reference source not found. provides a list of universal tools, designated features, and accommodations that were offered in the Spring 2021 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 *I AM Online Test Delivery System (TDS) User Guide*, available through the Indiana Assessment Portal (and included as Appendix O), provides instructions on how to access and use these features.

Table 1: Universal Tools, Designated Features, and Accommodations Available in Spring 2021

Universal Tools (for all students)	Designated Features	Accommodations (available per IEP)
Embedded/Online		
Online calculator Expandable passages Highlighter Masking Strikethrough Text-to-Speech (required) Zoom in and zoom out for text and graphics Line Reader	Color contrast Masking Mouse pointer (size and color) Print size (zoom in and zoom out) Spanish Translation (stacked)	Permissive mode to use assistive technology devices Streamlined mode
Non-Embedded		
Headphones or noise buffers to block out distractions Low-tech assistive writing instrument Preferential seating Scratch paper, including lined or graph paper Handheld calculator for paper assessment Student tested individually	Color acetate film for paper assessment Assistive technology to magnify/ enlarge text and images Access to sound amplification Special furniture or equipment for viewing test Special lighting conditions Time of day for testing altered	Alternate indicator of response Calculator Multiplication table Paper test booklet Large print test booklet Read-aloud script for paper-and-pencil test booklet Hundreds chart Interpreter for American Sign Language Braille test booklet 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 the Test Information Distribution Engine (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 School Test Coordinator (STC) were responsible for ensuring that arrangements for accommodations had been made before the test administration dates. IDOE provided a separate accessibility policy manual, *2020–2021 Indiana Assessments Policies Manual*, included as Appendix R of this technical report; the current manual is available on the IDOE Assessment Website at https://www.in.gov/doe/files/2021-2022-Indiana-Assessments-Policy-Manual.pdf 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 Paper test booklet

- Large print test booklet
- Braille test booklet

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 D of this report).

The Accessibility and Accommodations Guidance 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 *2020–2021 Indiana Assessments Policies Manual* (Appendix R) and the *2020–2021 Indiana Accessibility and Accommodations Guidance* (Appendix S).

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

3. ADMINISTRATOR TRAINING

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 the next section.

3.1 VIRTUALTEST ADMINISTRATOR TRAINING

TAs were required to attend one half-day training session in Indiana before administering the I AM. Before the Spring 2021 assessment administration, over a period of three weeks, CAI collaborated with IDOE to conduct online training sessions on the 2020–2021 test administration. These training sessions 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 Online Reporting System (ORS). During the training session, CAI used video vignettes, which included Indiana educators and students, to illustrate important concepts. Appendix P includes the PowerPoint presentation used during each training session.

All test administration personnel were required to attend online training sessions. 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.

3.2 COMPUTER-BASED ADMINISTRATION

TAMs and guides were available online for school and corporation staff. The *I AM Online Test Delivery System (TDS) User Guide* (Appendix O) was designed to familiarize TAs with TDS and contains 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 *I AM Online TDS User Guide* (Appendix O) and the *TIDE User Guide* (Appendix 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 *I AM Online TDS User Guide* (Appendix O) provides instructions for creating test sessions, monitoring sessions, verifying student information, assigning test accommodations, and starting, pausing, and submitting tests. The *Technology Setup for Online Testing Quick Guide* (Appendix A) as well as the *Additional Configurations and Troubleshooting Guide* (Appendix B) provide 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 were summarized in the following sections.

Roles and Responsibilities in the Online Testing Systems

CTCs, STCs, and TAs each had specific roles and responsibilities in the online testing systems. See the *I AM Test Administrator's Manual (TAM)* (Appendix I) and the *I AM Test Coordinator's Manual (TCM)* (Appendix J) 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 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 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 test 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

TAs administered the I AM assessments to students. Prior to administration of I AM, TAs completed the LCI (see Appendix R) for each student and administered a practice test session.

TAs were responsible for attending an in-person training, 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 I). 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.

3.3 TEST ADMINISTRATION RESOURCES

The list of training resources for the Spring 2021 I AM test administration is provided in this section. These materials were all available online on the Indiana Assessment Portal. (PDFs of these six resources have also been included in this technical report as Appendices C, E, F, G, and Q, respectively.)

1. **Technology Requirements for Online Testing Webinar Module.** This module provides technology requirements for CTCs and STCs to ensure their testing devices are set up properly before testing.
2. **Test Information Distribution Engine (TIDE) Webinar Module.** This module provides a general overview of the CAI system, TIDE, and the features applicable to educators before, during, and after testing.
3. **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 2021 assessment.
4. **I AM Educator Brochure.** This brochure provides an overview of the new I AM assessment to prepare educators for the Spring 2021 assessment.
5. **Online Reporting System (ORS) Webinar Module.** This module provides a general overview of the ORS, where student scores (including individual scores and aggregate scores) are displayed after students complete the I AM assessments.

Table 2 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 A, B, D, I, J, K, L, O, and P respectively.)

Table 2: User Guides and Manuals

Resource	Description
Technology Setup for Online Testing Quick Guide	This document explains in four steps how to set up technology in Indiana corporations and schools.
Additional Configurations and Troubleshooting Guide for Windows, Mac, Chrome OS, and Linux	This document contains additional configurations and troubleshooting for a school or corporation's network and workstations for Windows, Mac, Chrome and Linux devices.
Test Information Distribution Engine (TIDE) User Guide	This user guide describes the tasks performed in TIDE for I AM assessments.
I AM Test Administrator's Manual (TAM)	This manual provides information on the policies and procedures surrounding the I AM assessments, as well as an overview of the specific roles and responsibilities required before, during, and after testing.
I AM Test Coordinator's Manual (TCM)	This provides an overview of I AM test administration activities intended for Test Coordinators.
Released Item Repository Quick Guide	This quick guide provides an overview of how to administer the I AM RIR.
Released Item Repository Scoring Guides	These answer keys provide information on the items included in the RIR for each tested grade and content area.
I AM Online Test Delivery System (TDS) User Guide	This user guide supports TAs who manage testing for students participating in the LCI and the I AM practice and operational tests.
Online Reporting System (ORS) User Guide	This user guide provides an overview of the different features available to educators to support viewing student scores for the I AM assessments.

Department Resources and Support

In addition to the resources listed in Table 2, IDOE provided the following resources for corporations:

- A weekly newsletter was distributed via email to CTCs from the IDOE Office of Assessment every Monday. 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, or "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.

- General information about the assessments (such as dates of test windows for all state-administered assessments) was posted on the IDOE Office of Assessment website. The Accessibility and Accommodations Guidance in the I AM Policy and Guidance section of the IDOE website was designed to address questions pertaining to accommodations and overall accessibility.

- The *2020–2021 Indiana Assessments Policies Manual* (Appendix R) was also posted on the IDOE Office of Assessment website. This manual discussed 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 discussed formal security and integrity training for school and corporation personnel as well as the different aspects surrounding test security.

- The *2020-2021 Indiana Accessibility and Accommodations Guidance* (Appendix S) 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, parents, 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 deployed on October 5, 2020, and remained available all year.

There were no newly released items for the Fall 2020 RIR deployment because the I AM Spring 2020 operational assessment was not administered due to the COVID-19 pandemic. The RIR were items that were previously released from the Spring 2019 operational I AM assessment. An answer key for each grade and content area from the 2019-2020 released items (Appendix P) accompanied 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 2021 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.

4. TEST SECURITY PROCEDURES

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. If non-embedded accessibility supports are used, assessment security can become an issue when other test formats are used (e.g., large print) 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.

Some of the same considerations for test security apply to embedded accessibility supports. For example, ensuring that only authorized personnel have access to the test and that test materials are kept confidential is critical in technology-based assessments. In addition, it is important to guarantee that students cannot access any unauthorized programs, the Internet, saved data, or computer shortcuts while they are taking the assessment. In most cases, any specially required hardware devices and appropriate applications, such as switches, should be compatible with computer-delivered assessments. Prior to testing, educators should check device compatibility and make appropriate adjustments if necessary.

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 is appropriate/necessary

TAs were trained on test security procedures via an online TA Certification Course as well as training webinars hosted by both CAI and IDOE. Before testing, educators were required to sign the Test Security and Integrity Agreement, acknowledging that they would adhere to test security procedures. Both CAI and IDOE are committed to ensuring that, going forward, the test security policies and procedures are clearly presented in the user guides and TAMs.

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 training for everyone involved in test administration, outlining appropriate incident-reporting procedures, detecting test irregularities, and planning for investigation and handling of test security violations.

Furthermore, every school corporation or other test administration location that administers tests under the Indiana assessment system must have a locally developed, written test security policy that is shared with staff. While IDOE does not require school board approval of this policy, corporations should follow local level practices to determine if this policy needs to be approved by the local school board. The corporation’s test security policy must

- specify that secure test materials should not be delivered to school buildings more than one week (preferably less) in advance of test administration;
- specify that teachers and other school staff members are not allowed access to secure materials (except for the TAM) more than four hours in advance of the test administration; and
- describe the entity's plan for ensuring the security of assessment materials during testing and storage of all secure assessment materials before, during, and after testing. All test materials should be stored at a central location under lock and key.

If a TA has reason to believe that a violation in test security has occurred, he or she should notify the STC immediately. The STC should then notify the CTC.

If he or she has reason to believe that a lapse in test security has occurred, the CTC must do the following:

- Submit a Testing Irregularity Report to the IDOE Office of Student Assessment
- Submit a Test Irregularity request in TIDE
- Securely transmit relevant evidence of irregularities via secure file transfer protocol after the incident occurred
- Maintain the confidentiality of all evidence and documentation related to test security investigations

If IDOE has reason to believe that a violation in testing security has occurred, it has an obligation to investigate the incident as soon as possible. Additionally, IDOE receives data forensic information from CAI after testing has concluded. Following a review of the data forensic analysis, IDOE contacts corporations where there may be a test security issue. Corporations and schools are required to comply with IDOE's requests for documentation and information relevant to their initial investigation. IDOE may involve the school corporation or conduct a separate investigation.

If the IDOE determined that an irregularity in test administration or security had occurred, IDOE notified the CTC in writing, indicating the status of the case. Depending on the severity of the incident and its potential impact to the IDOE program, actions may have included but were not limited to the following:

- Invalidation of student scores
- A requirement for corporations to complete documentation and conduct interviews to gather more details regarding any test sessions identified as concerns
- IDOE's requirement of additional action steps taken by the corporations

4.1 Security of Test Materials

Before test materials were finalized, items went through a data review process where IDOE reviewed the statistics of previously administered items to ensure items performed

as without bias and were psychometrically sound. Items were accessed directly from CAI's secure Item Tracking System (ITS) for data review; thus, no printed copies needed to be transported to meetings. Once items passed through item data review, IDOE securely reviewed paper test booklets and the online test through user acceptance testing (UAT), both of which went through multiple reviews.

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.

The security of all test materials was required before, during, and after test administration. Under no circumstances were students permitted to assist in either preparing secure materials before testing or in organizing and returning materials after testing. After any administration, initial or make-up, secure materials (e.g., test booklets and TA scripts) 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 to be removed from the school's campus overnight. 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.

Printed test booklets were shipped to each Indiana school corporation one week prior to the start of the test window. Corporations were required to return printed test materials to the vendor one week after the end of the test window. Due to the fact that these materials were in corporations for six weeks, the security of the test booklets was a critical responsibility.

CTCs were therefore required to develop, implement, and assess procedures for the secure storage, administration, and delivery of standardized test materials back to testing vendors by established deadlines. Failure by a school corporation or its employees to securely store, administer, and return all secure test materials by established deadlines was considered an integrity breach under 511 IAC 5-5-3, which may have resulted in an action under Indiana Code 20-28-5-7. While student responses would not be scored, schools were required to immediately return student answer booklets found more than one week after the pick-up date.

It is considered a testing security violation for an individual to fail to follow security procedures set forth by IDOE, and no individual was permitted to

- use another staff member's username and/or password to access vendor systems or administer tests;
- use a student's login information to access operational assessments;
- review test questions prior to, during, or after test administration;
- give test takers access to test questions prior to testing;

- copy, reproduce, or use in any manner any portion of any secure assessment, for any reason;
- alter student answer documents (paper-and-pencil or online) prior to, during, or after testing;
- 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;
- make answer keys available to test takers;
- score student responses on the test locally before submitting the assessment for scoring to the test contractor, as designated by IDOE; or
- participate in, direct, aid, counsel, assist, encourage, or fail to report any of the acts prohibited in this section.

All special document test materials (print, large print, braille) were treated as secure documents, and processes were in place to protect them from loss, theft, and reproduction of any kind.

To access the I AM tests, an Indiana Secure Browser was required. The Indiana Secure Browser provided a secure environment for student testing by disabling hot keys, copy, and screen capture capabilities and preventing access to the desktop (Internet, email, and other files or programs installed on school machines). Users could not access other applications from within the Indiana Secure Browser, even if they knew the keystroke sequences. Students were also unable to print from the Indiana Secure Browser. During testing, the desktop was locked down. The Indiana Secure Browser was designed to ensure test security by prohibiting access to external applications or navigation away from the test. See the *I AM Online TDS User Guide* in Appendix O for further details.

4.2 INVESTIGATING TEST IRREGULARITIES

Throughout the test 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. 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, and 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 could monitor testing anomalies throughout the test 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 test window.

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

4.3 GUIDELINES FOR TEST INVALIDATION

During the test window, TAs were required to immediately report any test incidents (e.g., disruptive students, loss of Internet connectivity, student improprieties) to the STC. A test incident could include testing that was interrupted for an extended period due to a local technical malfunction or severe weather. STCs notified CTCs of any test irregularities that were reported. CTCs were responsible for submitting requests for test invalidations to IDOE via TIDE. IDOE made the final decision on whether to approve the requested test invalidation. CTCs could track the status and final decisions of requested test invalidations in TIDE.

4.4 CAI'S SYSTEM SECURITY

CAI has built-in security controls in all of 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 resides 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. REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). Standards for Educational and Psychological Testing.

Indiana's Alternate Measure

2020–2021

Volume 4
Evidence of Reliability and Validity

ACKNOWLEDGMENTS

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

Major contributors to this technical report include the following staff from American Institutes for Research: Stephan Ahadi, Yuan Hong, Hyesuk Jang, Hashim Evans, Anne Mogilnicki, Cherise Lesesne, and Celine Bryan. Major contributors from the Indiana Department of Education include the Assessment Director, Assistant Assessment Director, and Program Leads.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION AND OVERVIEW OF RELIABILITY AND VALIDITY EVIDENCE

Indiana's Alternate Measure (I AM) is an alternate assessment to measure student achievement and proficiency related to Indiana Alternate Achievement Standards or Content Connectors. I AM is designed for the students with significant cognitive disabilities to provide the summative accountability assessment for ELA, Mathematics, and Science and Social Studies. The I AM assessments replaced the Indiana Standards Tool for Alternate Reporting (ISTAR) from 2018-2019 school year. It is delivered as an online assessment, and online accommodated and paper-and-pencil versions are also available to students whose Individualized Education Program (IEP) or Individual Learning Plan (ILP) indicated such a need. Table 1 displays the complete list of test administration methods for the 2020–2021 school year.

Table 1: Test Administration

Subject	Administration*	Grade
ELA	Online/Paper-and-pencil tests	3–8, and 10
Mathematics	Online/Paper-and-pencil tests	3–8, and 10
Science	Online/Paper-and-pencil tests	4, 6, and high school
Social Studies	Online/ Paper-and-pencil tests	5

* Stacked Spanish is available for the mathematics, science, and social studies online assessments. Printed braille forms are also available. Full descriptions of available accommodations are listed in Volume 3, Section 1.2, Testing Accommodations. The number of students who were provided with accommodations is presented in Volume 1, Section 2.2, Designated Features and Accommodations.

With the implementation of these tests, both validity and reliability evidence is necessary to support appropriate inferences of student academic performance from the I AM scores. This volume provides empirical evidence about the validity and reliability of the 2020–2021 I AM, given its intended uses.

The I AM assessments for students with significant cognitive disabilities are designed for measuring student achievement according to Indiana's Alternate Academic Standards (Indiana Content Connectors). Measuring the wide variation of intellectual ability is one of the challenges in developing alternate assessments. To consider the variability of performance within this population, the I AM assessments utilize the attemptedness status of students' engagement with and the stage-adaptive tests. To determine whether a student is sufficiently engaged to produce a valid test score, students are categorized as according to a defined proficiency level, No Mode of Communication (NMC) or Undetermined (UND), according to the number of times they do not respond to and also the total number of responded items to the first five items. The tests assigned as NMC or UND are not scored. The stage-adaptive tests route students to one of three complexity levels in Part 2 based on their performance in Part 1. Part 2 is also constructed to target different levels of cognitive processing based on results from Part 1 to yield greater measurement precision across the distribution of student achievement levels observed

among the I AM student population. Details of I AM test design are described in Volume 1, Section 3.2 Test Design.

The purpose of this volume is to provide empirical evidence to support a validity and reliability argument regarding the uses and inferences for I AM. This volume addresses the following:

- **Content Validity.** Evidence is provided in Section 3, Evidence of Content Validity, to show that test forms were constructed to measure the Indiana Content Connectors with a sufficient number of items targeting each area of the blueprint.
- **Internal Structure Validity.** Evidence is provided in this volume regarding the internal relationships among the subscale scores to support their use and to justify the Item Response Theory (IRT) measurement model. This type of evidence includes observed and disattenuated Pearson correlations among reporting categories per grade. Additionally, local item independence, an assumption of unidimensional IRT, was tested using the Q_3 statistic.
- **Test Fairness.** Fairness is statistically analyzed in Section 5, Fairness in Content, using differential item functioning (DIF) in tandem with content alignment reviews by specialists.
- **Reliability.** Section 6, Reliability, addresses the marginal reliability estimates for each test. The reliability estimates are presented by grade and subject, as well as by demographic subgroups. This section also includes conditional standard errors of measurement (CSEMs) and classification accuracy and consistency results by grade and subject.

1.1 VALIDITY

Validity refers to the degree to which "evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 11). Messick (1989) defines validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment." Both of these definitions emphasize evidence and theory to support the inferences and interpretations of test scores. *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) suggests five sources of validity evidence that can be used in evaluating a proposed interpretation of test scores. When validating test scores, these sources of evidence should be carefully considered.

The first source of evidence for validity is the relationship between the test content and the intended test construct (see Section 3.2, Alignment of I AM Test Forms to the Content Connectors and Benchmarks). In order 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 (for details, see Volume 2 of this technical report, Test Development). Test scores can be used to support an intended validity claim when they contain minimal construct-irrelevant variance. For example, a Mathematics item targeting a specific Mathematics skill that requires advanced reading proficiency and vocabulary has a high level of construct-irrelevant variance. Thus, the intended construct of measurement is confounded, which impedes the validity of the test scores. Statistical analyses, such as factor analysis or multi-dimensional scaling, are also used to evaluate content relevance. Evidence based on test content is a crucial component of validity, because construct underrepresentation or irrelevancy could result in unfair advantages or disadvantages to one or more group of test takers.

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, p. 15). This evidence is collected by surveying test takers about their performance strategies or responses to particular items. Because items are developed to measure particular constructs and intellectual processes, evidence that test takers have engaged in relevant performance strategies to correctly answer the items supports the validity of the test scores.

The third source of evidence for validity is based on internal structure, which is the degree to which the relationships among test items and test components relate to the construct on which the proposed test scores are interpreted. DIF, which determines whether particular items may function differently for subgroups of test takers, is one method used for analyzing the internal structure of tests (see Volume 1, Section 4.2, Differential Item Functioning Analysis). Other possible analyses used to examine internal structure are dimensionality assessment, goodness-of-model-fit to data, and reliability analysis (for details, see Section 4, Evidence on Internal-External Structure, and Section 6, Reliability).

A fourth source of evidence for validity is the relationship of test scores to external variables. *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) divides this source of evidence into three parts: (1) convergent and discriminant evidence, (2) test-criterion relationships, and (3) validity generalization. Convergent evidence supports the relationship between the test and other measures intended to assess similar constructs. Conversely, discriminant evidence delineates the test from other measures intended to assess different constructs. To analyze both convergent and discriminant evidence, a multi-trait-multimethod matrix can be used. Additionally, test-criterion relationships indicate how accurately test scores predict criterion performance. The degree of accuracy depends mainly upon 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, one may need to consider sampling errors or range restrictions to determine whether the conclusions of a test can be assumed for the larger population.

Fifth, 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; this process should not be influenced by external factors. For example, if an employer administers a test to determine hiring rates for different groups of people, an unequal distribution of skills related to the measurement construct does not necessarily imply a lack of validity for the test. However, if the unequal distribution of scores is in fact due to an unintended, confounding aspect of the test, this would affect the test's validity. As described in this volume, as well as in Volume 1, test use should align with the intended purpose of the test.

Supporting a validity argument requires multiple sources of validity evidence. This then allows for one to evaluate if sufficient evidence has been presented to support the intended uses and interpretations of the test scores. Thus, determining the validity of a test 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.

1.2 RELIABILITY

Reliability refers to consistency in test scores. Reliability can be defined as the degree to which individuals' deviation scores remain relatively consistent over repeated administrations of the same test or alternate test forms (Crocker & Algina, 1986). For example, if a person repeatedly takes the same or parallel tests, he or she should receive consistent results. The reliability coefficient refers to the ratio of true score variance to observed score variance:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2}.$$

Test score reliability is traditionally estimated using both classical and 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 along the append. 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 student's ability level in each tier, the precision of test scores may increase, especially for lower- and higher-ability students. 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.

2. PURPOSE OF INDIANA'S ALTERNATE MEASURE ASSESSMENTS

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. Those scores, which are estimates of student achievement and proficiency measured by assessment, are used to explain how well students performed against such expectations.

A scale score is the student's overall numeric score. Scale scores can be used to illustrate students' current level of performance and to compare the 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. 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.

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 6.4, Precision at Cut Scores). Performance-Level Descriptors (PLDs) are descriptions of content-area knowledge and skills that students at each performance level are expected to possess. Thus, performance levels can be interpreted based on PLDs. Students performing on the I AM at Level 3 are considered on track to demonstrate progress toward mastery of the knowledge and skills necessary for competitive employment and post-secondary education.

I AM supports instruction and student learning by measuring student performance and providing feedback to educators and parents. Assessments can be used as an indicator to determine whether the students with significant cognitive disabilities in Indiana are ready with the knowledge and skills that are essential for post-secondary education or competitive integrated employment.

I AM assessments also provide evidence in the requirements for state and federal accountability systems. Test scores can be employed to evaluate students' learning progress and help teachers to improve their instruction, which in turn will have a positive effect on student learning over time.

The tests are constructed to measure student proficiency on the Indiana Content Connectors in English/Language Arts (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. Volume 2, Test Development, describes the Indiana Content Connectors and test blueprints in more detail. The I AM test scores are useful indicators for understanding individual students' academic performance on the Indiana Content Connectors and whether students are progressing in their performance over time. Additionally, individual test scores can be used for measuring reliability of the test which can be found in Section 3, Reliability.

I AM is a criterion-referenced test designed to measure student performance on the Indiana Content Connectors in ELA, Mathematics, Science, and Social Studies. As a comparison, norm-referenced tests are designed to compare or rank all students to one another.

With the overall scale score, percent-correct scores at the reporting category (domain) level were provided for each student to indicate student performance in different content areas of the test relative to the other areas and to the district and state. These scores serve as useful feedback for teachers to tailor their instruction, provided that they are viewed with the usual caution that accompanies use of reporting category scores. Thus, we must examine the reliability coefficients for these test scores and the validity of the test scores to support practical use across the state. Volume 5 of this technical report series is the score interpretation guide and provides details on all scores generated and their appropriate uses and limitations.

3. EVIDENCE OF CONTENT VALIDITY

This section demonstrates that the knowledge and skills assessed by the I AM were representative of the Content Connectors of the larger knowledge domain. We describe the Content Connectors for I AM and discuss the assessment development process, mapping I AM assessments to the Content Connectors. A complete description of the test development process can be found in Volume 2, *Test Development*.

3.1 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. Blueprints were developed to ensure that the assessment and items were aligned to the prioritized Content Connectors that they were intended to measure.

Table 2–Table 5 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 the 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 Volume 2, Section 4 (I AM Assessment Construction), of the I AM technical report.

Table 2: Number of Items for Each Reporting Category (ELA)

Grade	Reporting Category	Number of Items
3	Key Ideas and Textual Support/Vocabulary (KITS)	10
3	Reading Foundations (RF)	7
3	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	8
3	Writing (W)	7
4	Key Ideas and Textual Support/Vocabulary (KITS)	13
4	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	11
4	Writing (W)	8
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)	12
6	Structural Elements and Organization/Connection of Ideas/Media Literacy (SECM)	11
6	Writing (W)	8

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

Table 3: Number of Items for Each Reporting Category (Mathematics)

Grade	Reporting Category	Number of Items
3	Algebraic Thinking and Data Analysis (ATDA)	7
3	Computation (C)	7–8
3	Geometry and Measurement (GM)	7–8
3	Number Sense (NS)	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
5	Geometry and Measurement, Data Analysis, and Statistics (GMDAS)	8
5	Number Sense (NS)	8–9
6	Algebra and Functions (AF)	8
6	Computation (C)	7
6	Geometry and Measurement, Data Analysis, and Statistics (GMDAS)	7
6	Number Sense (NS)	9

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

Table 4: 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
6	Investigating (I)	9–10
6	Questioning and Modeling (QM)	8
10	Analyzing Data and Mathematical Thinking (ADMT)	13–14
10	Communicating Explanations and Evaluating Claims Using Evidence (CEEC)	8
10	Developing and Using Modeling to Describe Structure and Function (UM)	10–11

Table 5: 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)	8
5	Geography (US_FOUND_GEO)	7

4. EVIDENCE ON INTERNAL-EXTERNAL STRUCTURE

In this section, we explore the internal structure of Indiana's Alternate Measure (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 2 through Table 5 for reporting category information). Students were provided with their percentage of correct answers based on each reporting category. Evidence is needed to verify that scores for each 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 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 test or at grade level are provided in the following sections. The theta estimates of each subscore were used for the correlations.

4.1 CORRELATIONS AMONG REPORTING CATEGORY SCORES

Table 6–Table 9 present the observed correlation matrix of the reporting category scores for each subject area. In ELA, the correlations among the reporting categories ranged from 0.2–0.68. For Mathematics, the correlations were between -0.11–0.41. In Science, the correlations among reporting categories ranged from 0.27–0.63. In Social Studies, the correlations among reporting categories ranged from 0.50–0.59.

In some instances, these 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.

Table 10–Table 13 display disattenuated correlations. Disattenuated values greater than 1 are reported as 1.00*. The overall average disattenuated correlation was 0.95 for ELA, 0.88 for Mathematics, 0.92 for Science, and 1.00 for Social Studies. The values suggest that validity evidence of internal structure is supported.

Table 6: Observed Correlation Matrix Among Reporting Categories (ELA)

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Key Ideas and Textual Support/Vocabulary (Cat1)	10	1.00	0.30	0.40	0.30
	Reading Foundations (Cat2)	7		1.00	0.31	0.20
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat3)	8			1.00	0.27
	Writing (Cat4)	7				1.00
4	Key Ideas and Textual Support/Vocabulary (Cat1)	13	1.00	0.42	0.59	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	0.35	
	Writing (Cat3)	8			1.00	
5	Key Ideas and Textual Support/Vocabulary (Cat1)	14	1.00	0.53	0.57	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	9		1.00	0.48	
	Writing (Cat3)	9			1.00	
6	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	0.56	0.52	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	0.46	
	Writing (Cat3)	8			1.00	
7	Key Ideas and Textual Support/Vocabulary (Cat1)	14	1.00	0.61	0.57	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	9		1.00	0.56	
	Writing (Cat3)	8			1.00	
8	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	0.64	0.61	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	10		1.00	0.54	
	Writing (Cat3)	8			1.00	
10	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	0.65	0.68	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	0.59	
	Writing (Cat3)	8			1.00	

Table 7: Observed Correlation Matrix Among Reporting Categories (Mathematics)

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Algebraic Thinking and Data Analysis (Cat1)	7	1.00	0.35	0.30	0.41
	Computation (Cat2)	7–8		1.00	0.25	0.35
	Geometry and Measurement (Cat3)	7–8			1.00	0.29
	Number Sense (Cat4)	8				1.00
4	Algebraic Thinking and Data Analysis (Cat1)	7	1.00	0.29	0.29	0.23
	Computation (Cat2)	7–8		1.00	0.28	0.34
	Geometry and Measurement (Cat3)	7			1.00	0.31
	Number Sense (Cat4)	7				1.00
5	Algebraic Thinking (Cat1)	7–8	1.00	0.04	0.16	-0.11
	Computation (Cat2)	7		1.00	0.31	0.24
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	8			1.00	0.15
	Number Sense (Cat4)	8–9				1.00
6	Algebra and Functions (Cat1)	8	1.00	0.32	0.30	0.36
	Computation (Cat2)	7		1.00	0.34	0.38
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	7			1.00	0.35
	Number Sense (Cat4)	9				1.00
7	Algebra and Functions (Cat1)	8	1.00	0.26	0.18	0.12
	Data Analysis, Statistics, and Probability (Cat2)	7		1.00	0.19	0.21
	Geometry and Measurement (Cat3)	7			1.00	0.09
	Number Sense and Computation (Cat4)	8				1.00
8	Algebra and Functions (Cat1)	10	1.00	0.16	0.14	0.21
	Data Analysis, Statistics, and Probability (Cat2)	7		1.00	0.20	0.20
	Geometry and Measurement (Cat3)	7			1.00	0.16
	Number Sense and Computation (Cat4)	7				1.00
10	Equations and Inequalities (Linear and Systems) (Cat1)	7–8	1.00	0.19	-0.01	0.23
	Functions (Linear and Non-linear) (Cat2)	8		1.00	0.05	0.24
	Geometry and Measurement (Cat3)	7			1.00	0.14
	Number Sense and Data Analysis (Cat4)	7–8				1.00

Table 8: Observed 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	1.00	0.30	0.31	0.34
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7–8		1.00	0.32	0.48
	Investigating (Cat3)	7			1.00	0.38
	Questioning and Modeling (Cat4)	9–10				1.00
6	Analyzing, Interpreting, and Computational Thinking (Cat1)	7–8	1.00	0.27	0.42	0.38
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7		1.00	0.43	0.33
	Investigating (Cat3)	9–10			1.00	0.37
	Questioning and Modeling (Cat4)	8				1.00
Biology	Analyzing Data and Mathematical Thinking (Cat1)	13–14	1.00	0.63	0.59	
	Communicating Explanations and Evaluating Claims Using Evidence (Cat2)	8		1.00	0.56	
	Developing and Using Modeling to Describe Structure and Function (Cat3)	10–11			1.00	

Table 9: Observed Correlation Matrix Among Reporting Categories (Social Studies)

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3
5	Civics and Government/History (Cat1)	17	1.00	0.57	0.59
	Economics (Cat2)	8		1.00	0.50
	Geography (Cat3)	7			1.00

Table 10: Disattenuated Correlation Matrix Among Reporting Categories (ELA)

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Key Ideas and Textual Support/Vocabulary (Cat1)	10	1.00	0.88	0.73	1.00*
	Reading Foundations (Cat2)	7		1.00	1.00	1.00*
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat3)	8			1.00	1.00*
	Writing (Cat4)	7				1.00
4	Key Ideas and Textual Support/Vocabulary (Cat1)	13	1.00	0.72	0.99	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	0.69	
	Writing (Cat3)	8			1.00	
5	Key Ideas and Textual Support/Vocabulary (Cat1)	14	1.00	0.96	0.96	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	9		1.00	0.92	
	Writing (Cat3)	9			1.00	
6	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	0.97	0.96	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	0.97	
	Writing (Cat3)	8			1.00	
7	Key Ideas and Textual Support/Vocabulary (Cat1)	14	1.00	1.00*	0.98	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	9		1.00	1.00*	
	Writing (Cat3)	8			1.00	
8	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	1.00*	1.00	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	10		1.00	1.00*	
	Writing (Cat3)	8			1.00	
10	Key Ideas and Textual Support/Vocabulary (Cat1)	12	1.00	1.00*	1.00*	
	Structural Elements and Organization/Connection of Ideas/Media Literacy (Cat2)	11		1.00	1.00*	
	Writing (Cat3)	8			1.00	

Table 11: Disattenuated Correlation Matrix Among Reporting Categories (Mathematics)

Grade	Reporting Category	Number of Items	Cat1	Cat2	Cat3	Cat4
3	Algebraic Thinking and Data Analysis (Cat1)	7	1.00	0.99	0.96	1.00*
	Computation (Cat2)	7–8		1.00	0.88	0.93
	Geometry and Measurement (Cat3)	7–8			1.00	0.87
	Number Sense (Cat4)	8				1.00
4	Algebraic Thinking and Data Analysis (Cat1)	7	1.00	1.00*	1.00*	1.00*
	Computation (Cat2)	7–8		1.00	0.80	0.87
	Geometry and Measurement (Cat3)	7			1.00	0.87
	Number Sense (Cat4)	7				1.00
5	Algebraic Thinking (Cat1)	7–8	1.00	0.21	0.54	N/A
	Computation (Cat2)	7		1.00	1.00*	1.00*
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	8			1.00	0.55
	Number Sense (Cat4)	8–9				1.00
6	Algebra and Functions (Cat1)	8	1.00	1.00*	0.99	1.00*
	Computation (Cat2)	7		1.00	1.00*	1.00*
	Geometry and Measurement, Data Analysis, and Statistics (Cat3)	7			1.00	0.97
	Number Sense (Cat4)	9				1.00
7	Algebra and Functions (Cat1)	8	1.00	1.00*	0.73	0.67
	Data Analysis, Statistics, and Probability (Cat2)	7		1.00	0.96	1.00*
	Geometry and Measurement (Cat3)	7			1.00	0.37
	Number Sense and Computation (Cat4)	8				1.00
8	Algebra and Functions (Cat1)	10	1.00	0.60	0.89	1.00*
	Data Analysis, Statistics, and Probability (Cat2)	7		1.00	1.00*	0.98
	Geometry and Measurement (Cat3)	7			1.00	1.00*
	Number Sense and Computation (Cat4)	7				1.00
10	Equations and Inequalities (Linear and Systems) (Cat1)	7–8	1.00	0.84	N/A	0.83
	Functions (Linear and Non-linear) (Cat2)	8		1.00	N/A	1.00*
	Geometry and Measurement (Cat3)	7			1.00	N/A
	Number Sense and Data Analysis (Cat4)	7–8				1.00

Table 12: Disattenuated 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	1.00	1.00*	1.00*	1.00*
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7–8		1.00	0.85	1.00*
	Investigating (Cat3)	7			1.00	0.85
	Questioning and Modeling (Cat4)	9–10				1.00
6	Analyzing, Interpreting, and Computational Thinking (Cat1)	7–8	1.00	0.56	0.96	0.82
	Explaining Solutions, Reasoning, and Communicating (Cat2)	7		1.00	1.00*	0.78
	Investigating (Cat3)	9–10			1.00	0.95
	Questioning and Modeling (Cat4)	8				1.00
Biology	Analyzing Data and Mathematical Thinking (Cat1)	13–14	1.00	1.00*	0.96	
	Communicating Explanations and Evaluating Claims Using Evidence (Cat2)	8		1.00	1.00*	
	Developing and Using Modeling to Describe Structure and Function (Cat3)	10–11			1.00	

Table 13: Disattenuated Correlation Matrix Among Reporting Categories (Social Studies)

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

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):

$$L(\theta) = \int \prod_{i=1}^{I} Pr\ (\theta)\ f(\theta) d\theta$$

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 examinee for item *i*, $T_i(\hat{\theta}_j)$ is the estimated true score for item *i* of examinee *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 examinee *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 14–Table 17 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 of grades and subjects, were around or smaller than a critical value of 0.2 for $|Q_3|$ (Chen & Thissen, 1997), except for a few grades in Mathematics which have the value ranging 0.22 to 0.30 for $|Q_3|$.

Table 14: ELA Q3 Statistic

Grade	Q_3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
3	-0.575	-0.216	-0.040	0.190	0.456
4	-0.319	-0.193	-0.041	0.159	0.307
5	-0.385	-0.197	-0.038	0.159	0.471
6	-0.308	-0.179	-0.038	0.123	0.324
7	-0.387	-0.184	-0.038	0.135	0.361
8	-0.270	-0.171	-0.043	0.115	0.306
10	-0.432	-0.188	-0.039	0.112	0.453

Table 15: Mathematics Q3 Statistic

Grade	Q_3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
3	-0.428	-0.203	-0.049	0.178	0.645
4	-0.469	-0.222	-0.041	0.191	0.557
5	-0.686	-0.300	-0.053	0.247	0.762
6	-0.379	-0.198	-0.050	0.163	0.524
7	-0.725	-0.243	-0.041	0.216	1.000
8	-0.616	-0.220	-0.051	0.178	0.631
10	-0.457	-0.200	-0.044	0.149	0.415

Table 16: Science Q3 Statistic

Grade	Q_3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
4	-0.351	-0.219	-0.044	0.175	0.353
6	-0.340	-0.205	-0.045	0.159	0.375
Biology	-0.248	-0.166	-0.037	0.129	0.352

Table 17: Social Studies Q3 Statistic

Grade	Q_3 Distribution				
	Minimum	5th Percentile	Median	95th Percentile	Maximum
5	-0.292	-0.194	-0.039	0.142	0.335

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 evidence. It is a part of validity evidence demonstrating that assessment scores are related as expected with 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 permit 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 content area and across subjects and grades. Each correlation table shows 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 between assessments measuring a different construct.

Table 18: Grade 3 Observed Score Correlations

Subject	Reporting Category	ELA				Mathematics			
		Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.30	0.40	0.30	0.39	0.29	0.32	0.38
	RF (Cat2)		1.00	0.31	0.20	0.13	0.15	0.16	0.14
	SECM (Cat3)			1.00	0.27	0.14	0.17	0.14	0.10
	W (Cat4)				1.00	0.22	0.16	0.19	0.23
Mathematics	ATDA (Cat1)					1.00	0.35	0.30	0.41
	C (Cat2)						1.00	0.25	0.35
	GM (Cat3)							1.00	0.29
	NS (Cat4)								1.00

Table 19: Grade 4 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.42	0.59	0.31	0.35	0.37	0.35	0.29	0.44	0.39	0.53
	SECM (Cat2)		1.00	0.35	0.25	0.24	0.25	0.10	0.23	0.27	0.33	0.32
	W (Cat3)			1.00	0.28	0.34	0.34	0.33	0.25	0.47	0.34	0.53
Mathematics	ATDA (Cat1)				1.00	0.29	0.29	0.23	0.19	0.26	0.24	0.29
	C (Cat2)					1.00	0.28	0.34	0.21	0.30	0.26	0.38
	GM (Cat3)						1.00	0.31	0.22	0.28	0.26	0.39
	NS (Cat4)							1.00	0.17	0.21	0.15	0.37
Science	AICT (Cat1)								1.00	0.30	0.31	0.34
	ESRC (Cat2)									1.00	0.32	0.48
	I (Cat3)										1.00	0.38
	QM (Cat4)											1.00

Table 20: Grade 5 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics				Social Studies		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	1.00	0.53	0.57	0.14	0.25	0.34	0.15	0.54	0.53	0.52
	SECM (Cat2)		1.00	0.48	0.02	0.24	0.39	0.20	0.46	0.41	0.43
	W (Cat3)			1.00	0.13	0.23	0.41	0.12	0.52	0.43	0.49
Mathematics	AT (Cat1)				1.00	0.04	0.16	-0.11	0.24	0.17	0.24
	C (Cat2)					1.00	0.31	0.24	0.24	0.20	0.26
	GMDAS (Cat3)						1.00	0.15	0.40	0.29	0.37
	NS (Cat4)							1.00	0.09	0.02	0.15
Social Studies	CGH (Cat1)								1.00	0.57	0.59
	ECON (Cat2)									1.00	0.50
	GEO (Cat3)										1.00

Table 21: Grade 6 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.56	0.52	0.36	0.41	0.32	0.41	0.55	0.36	0.44	0.46
	SECM (Cat2)		1.00	0.46	0.30	0.33	0.36	0.36	0.43	0.36	0.39	0.37
	W (Cat3)			1.00	0.32	0.38	0.32	0.35	0.41	0.33	0.38	0.40
Mathematics	AF (Cat1)				1.00	0.32	0.30	0.36	0.27	0.31	0.26	0.32
	C (Cat2)					1.00	0.34	0.38	0.35	0.23	0.29	0.31
	GMDAS (Cat3)						1.00	0.35	0.28	0.31	0.31	0.36
	NS (Cat4)							1.00	0.31	0.28	0.34	0.37
Science	AICT (Cat1)								1.00	0.27	0.42	0.38
	ESRC (Cat2)									1.00	0.43	0.33
	I (Cat3)										1.00	0.37
	QM (Cat4)											1.00

Table 22: Grade 7 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.61	0.57	0.17	0.27	0.39	0.14
	SECM (Cat2)		1.00	0.56	0.13	0.25	0.41	0.09
	W (Cat3)			1.00	0.23	0.30	0.41	0.16
Mathematics	AF (Cat1)				1.00	0.26	0.18	0.12
	DASP (Cat2)					1.00	0.19	0.21
	GM (Cat3)						1.00	0.09
	NSC (Cat4)							1.00

Table 23: Grade 8 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.64	0.61	0.19	0.30	0.22	0.24
	SECM (Cat2)		1.00	0.54	0.20	0.26	0.17	0.19
	W (Cat3)			1.00	0.15	0.35	0.18	0.22
Mathematics	AF (Cat1)				1.00	0.16	0.14	0.21
	DASP (Cat2)					1.00	0.20	0.20
	GM (Cat3)						1.00	0.16
	NSC (Cat4)							1.00

Table 24: Grade 10 Observed Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	1.00	0.65	0.68	0.30	0.29	0.02	0.33	0.64	0.60	0.63
	SECM (Cat2)		1.00	0.59	0.27	0.26	0.05	0.30	0.57	0.51	0.49
	W (Cat3)			1.00	0.31	0.27	-0.02	0.32	0.60	0.53	0.61
Mathematics	EI (Cat1)				1.00	0.19	-0.01	0.23	0.36	0.31	0.30
	F (Cat2)					1.00	0.05	0.24	0.28	0.28	0.25
	GM (Cat3)						1.00	0.14	0.01	0.01	0.03
	NSDA (Cat4)							1.00	0.40	0.30	0.30
Science	ADMT (Cat1)								1.00	0.63	0.59
	CEEC (Cat2)									1.00	0.56
	UM (Cat3)										1.00

Table 25: Grade 3 Disattenuated Score Correlations

Subject	Reporting Category	ELA				Mathematics			
		Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.88	0.73	1.00*	0.81	0.64	0.82	0.74
	RF (Cat2)		1.00	1.00	1.00*	0.49	0.61	0.72	0.48
	SECM (Cat3)			1.00	1.00*	0.32	0.41	0.40	0.23
	W (Cat4)				1.00	1.00	0.78	1.00*	0.99
Mathematics	ATDA (Cat1)					1.00	0.99	0.96	1.00*
	C (Cat2)						1.00	0.88	0.93
	GM (Cat3)							1.00	0.87
	NS (Cat4)								1.00

Table 26: Grade 4 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.72	0.99	1.00*	0.68	0.80	0.68	0.91	0.84	0.79	0.85
	SECM (Cat2)		1.00	0.69	1.00*	0.55	0.63	0.22	0.87	0.61	0.78	0.59
	W (Cat3)			1.00	1.00*	0.75	0.84	0.72	0.90	1.00*	0.79	0.97
Mathematics	ATDA (Cat1)				1.00	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
	C (Cat2)					1.00	0.80	0.87	0.87	0.78	0.70	0.81
	GM (Cat3)						1.00	0.87	1.00*	0.81	0.79	0.92
	NS (Cat4)							1.00	0.71	0.52	0.41	0.78
Science	AICT (Cat1)								1.00	1.00*	1.00*	1.00*
	ESRC (Cat2)									1.00	0.85	1.00*
	I (Cat3)										1.00	0.85
	QM (Cat4)											1.00

Table 27: Grade 5 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics				Social Studies		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	1.00	0.96	0.96	0.37	0.70	0.73	0.42	0.87	0.95	0.94
	SECM (Cat2)		1.00	0.92	0.07	0.76	0.94	0.64	0.84	0.84	0.90
	W (Cat3)			1.00	0.36	0.71	0.93	0.36	0.91	0.83	0.96
Mathematics	AT (Cat1)				1.00	0.21	0.54	N/A	0.64	0.48	0.70
	C (Cat2)					1.00	1.00*	1.00*	0.71	0.63	0.84
	GMDAS (Cat3)						1.00	0.55	0.87	0.70	0.90
	NS (Cat4)							1.00	0.25	0.08	0.47
Social Studies	CGH (Cat1)								1.00	1.00*	1.00*
	ECON (Cat2)									1.00	1.00*
	GEO (Cat3)										1.00

Table 28: Grade 6 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	0.97	0.96	0.84	0.96	0.69	0.79	0.92	0.66	0.90	0.87
	SECM (Cat2)		1.00	0.97	0.80	0.87	0.89	0.81	0.83	0.77	0.91	0.80
	W (Cat3)			1.00	0.92	1.00*	0.85	0.84	0.83	0.75	0.95	0.94
Mathematics	AF (Cat1)				1.00	1.00*	0.99	1.00*	0.72	0.89	0.83	0.93
	C (Cat2)					1.00	1.00*	1.00*	0.92	0.66	0.92	0.93
	GMDAS (Cat3)						1.00	0.97	0.68	0.82	0.91	0.99
	NS (Cat4)							1.00	0.69	0.68	0.88	0.92
Science	AICT (Cat1)								1.00	0.56	0.96	0.82
	ESRC (Cat2)									1.00	1.00*	0.78
	I (Cat3)										1.00	0.95
	QM (Cat4)											1.00

Table 29: Grade 7 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	1.00*	0.98	0.48	0.99	0.87	0.42
	SECM (Cat2)		1.00	1.00*	0.42	1.00*	1.00*	0.29
	W (Cat3)			1.00	0.72	1.00*	1.00*	0.52
Mathematics	AF (Cat1)				1.00	1.00*	0.73	0.67
	DASP (Cat2)					1.00	0.96	1.00*
	GM (Cat3)						1.00	0.37
	NSC (Cat4)							1.00

Table 30: Grade 8 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics			
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4
ELA	KITS (Cat1)	1.00	1.00*	1.00	0.47	0.66	0.81	0.78
	SECM (Cat2)		1.00	1.00*	0.57	0.64	0.73	0.71
	W (Cat3)			1.00	0.41	0.86	0.75	0.81
Mathematics	AF (Cat1)				1.00	0.60	0.89	1.00*
	DASP (Cat2)					1.00	1.00*	0.98
	GM (Cat3)						1.00	1.00*
	NSC (Cat4)							1.00

Table 31: Grade 10 Disattenuated Score Correlations

Subject	Reporting Category	ELA			Mathematics				Science		
		Cat1	Cat2	Cat3	Cat1	Cat2	Cat3	Cat4	Cat1	Cat2	Cat3
ELA	KITS (Cat1)	1.00	1.00*	1.00*	0.72	0.82	N/A	0.77	0.98	1.00*	1.00*
	SECM (Cat2)		1.00	1.00*	0.70	0.79	N/A	0.78	0.95	0.97	0.87
	W (Cat3)			1.00	0.78	0.80	N/A	0.82	0.99	1.00	1.00*
Mathematics	EI (Cat1)				1.00	0.84	N/A	0.83	0.86	0.85	0.76
	F (Cat2)					1.00	N/A	1.00*	0.77	0.88	0.73
	GM (Cat3)						1.00	N/A	N/A	N/A	N/A
	NSDA (Cat4)							1.00	0.95	0.82	0.76
Science	ADMT (Cat1)								1.00	1.00*	0.96
	CEEC (Cat2)									1.00	1.00*
	UM (Cat3)										1.00

5. 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 performance. 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:

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 leadership.

5.1 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 quality of items, one notable statistic that was utilized was DIF. Items were classified into three categories (A, B, or C) for DIF, ranging from "no evidence of DIF" to "severe DIF," according to the DIF classification convention illustrated in Volume 1, Section 4.2 Differential Item Functioning Analysis. 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, or Female), or negatively (i.e., –A, –B, or –C), signifying that the item favored the reference group (e.g., White or 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 the reference group. The details surrounding this review of items for bias is further described in Volume 2 of this technical report, *Test Development*.

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

- Male/Female
- White/African-American

- White/Hispanic
- Autism/Other
- Moderate and severe intellectual disability/Other

A detailed description of the DIF analysis that was performed is presented in Volume 1, Section 4.2, Differential Item Functioning Analysis.

6. RELIABILITY

6.1 MARGINAL RELIABILITY

Marginal reliability is a measure of the overall reliability of the test based on the average CSEMs, estimated at different points on the performance scale for all students. The marginal reliability coefficients are nearly identical, or close to, the coefficient *alpha*. For our analysis, the marginal reliability coefficients were computed using operational items.

Within the IRT framework, measurement error varies across the range of ability. 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 SEM, which is equal to the inverse square root of information. The larger the SEM, 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, as opposed to students with more extreme scores. Conversely, measurement error is minimal for the part of the underlying scale that is at the middle of the test distribution and greater on scaled values farther away from the middle.

The marginal reliability of a test is computed by integrating θ out of the TIF as follows:

$$\rho = \frac{\sigma_\theta^2 - \bar{\sigma}_e^2}{\sigma_\theta^2}$$

where σ_θ^2 is the true score variance of student ability estimate (θ). The marginal measurement error variance ($\bar{\sigma}_e^2$) can be estimated as the average of squared standard error of θ across all test takers. The marginal measurement error variance ($\bar{\sigma}_e^2$) is computed as:

$$\bar{\sigma}_e^2 = \int_{-\infty}^{\infty} \frac{1}{I(\theta)} g(\theta) d\theta \ = \frac{\sum_{i=1}^{\mathrm{N}} CSEM_i^2}{N},$$

where $g(\theta)$ is a density function and *N* is the number of students.

Table 32 presents the marginal reliability coefficients for all students by test. The marginal reliability coefficients for all grades of ELA, Science, and Social Studies range from 0.72–0.85, which is similar to other statewide standardized tests. In mathematics, the marginal reliability coefficients are relatively lower than in other subjects. While the marginal reliability coefficients of lower grades in Mathematics had a similar level to other subjects, ranging from 0.68–0.70, other grades including grades 5, 7, 8, and 10 showed the lower marginal reliability coefficients of 0.48–0.51. This is expected due to the small standard deviations of theta scores. As seen in the marginal reliability equation the small variance of student ability estimates is associated with low marginal reliability at the same level of marginal measurement error variance. The standard deviations of theta and scales scores are provided in Table 33. The table shows that grades 5, 7, 8, and 10 mathematics have a smaller standard deviation of theta scores ranging from 0.52 to 0.56 while other subject and grade tests have a standard deviation from 0.7 to 1.1. Figures 2–5 also show

that the range of observed scale scores for these mathematics grades is narrower and fewer test records received higher scores.

The marginal reliability by each demographic subgroup are presented in Appendix A, Reliability Coefficients. In Appendix A, demographic subgroups include Female, Male, Autism, Non-Autism, Moderate and Severe Intellectual Disability, Non-Moderate and Severe Intellectual Disability, African American, Hispanic, and White.

Table 32: Marginal Reliability Coefficients

Test	Number of Students	Number of Operational Items	Marginal Reliability
ELA 3	565	32	0.73
ELA 4	627	32	0.79
ELA 5	671	32	0.79
ELA 6	749	32	0.80
ELA 7	808	32	0.83
ELA 8	971	32	0.84
ELA 10	951	32	0.85
Mathematics 3	568	32	0.70
Mathematics 4	629	32	0.68
Mathematics 5	666	32	0.51
Mathematics 6	746	32	0.68
Mathematics 7	804	32	0.48
Mathematics 8	966	32	0.51
Mathematics 10	944	32	0.51
Science 4	622	32	0.72
Science 6	740	32	0.76
Biology	963	32	0.82
Social Studies 5	660	32	0.79

6.2 Test Information Curves and Standard Error of Measurement

Within the IRT framework, measurement error varies across the range of ability as a result of the assessment, providing varied information across the range of ability 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 CSEM at each score point. For instance, if the SEM is large, then less information is being provided by the assessment at the specific ability level.

Figure 1 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

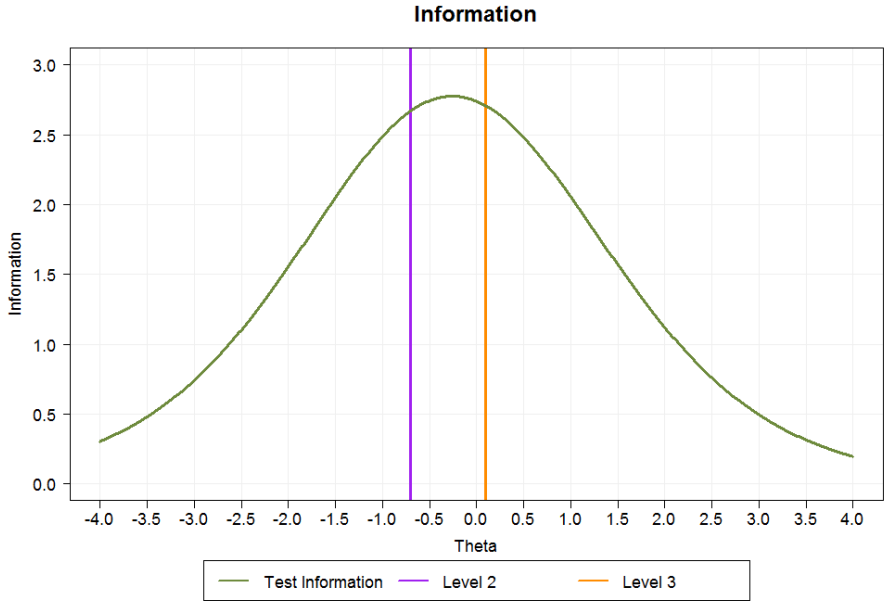
provides less information about test takers at the tails, where the curve is lower, relative to the center.

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:

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

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

Figure 1: 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. For this reason, SEM plots are presented in Figure 2 through Figure 5 for ELA, Mathematics, Science, and Social Studies respectively, instead of the TIFs. These plots are based on the scaled scores reported in 2021. Vertical lines represent two performance category cut scores.

When the maximum likelihood estimate (MLE) is used for score estimation, it is also common to compute the SEM from the numerically differentiated Hessian, which approximates as follows:

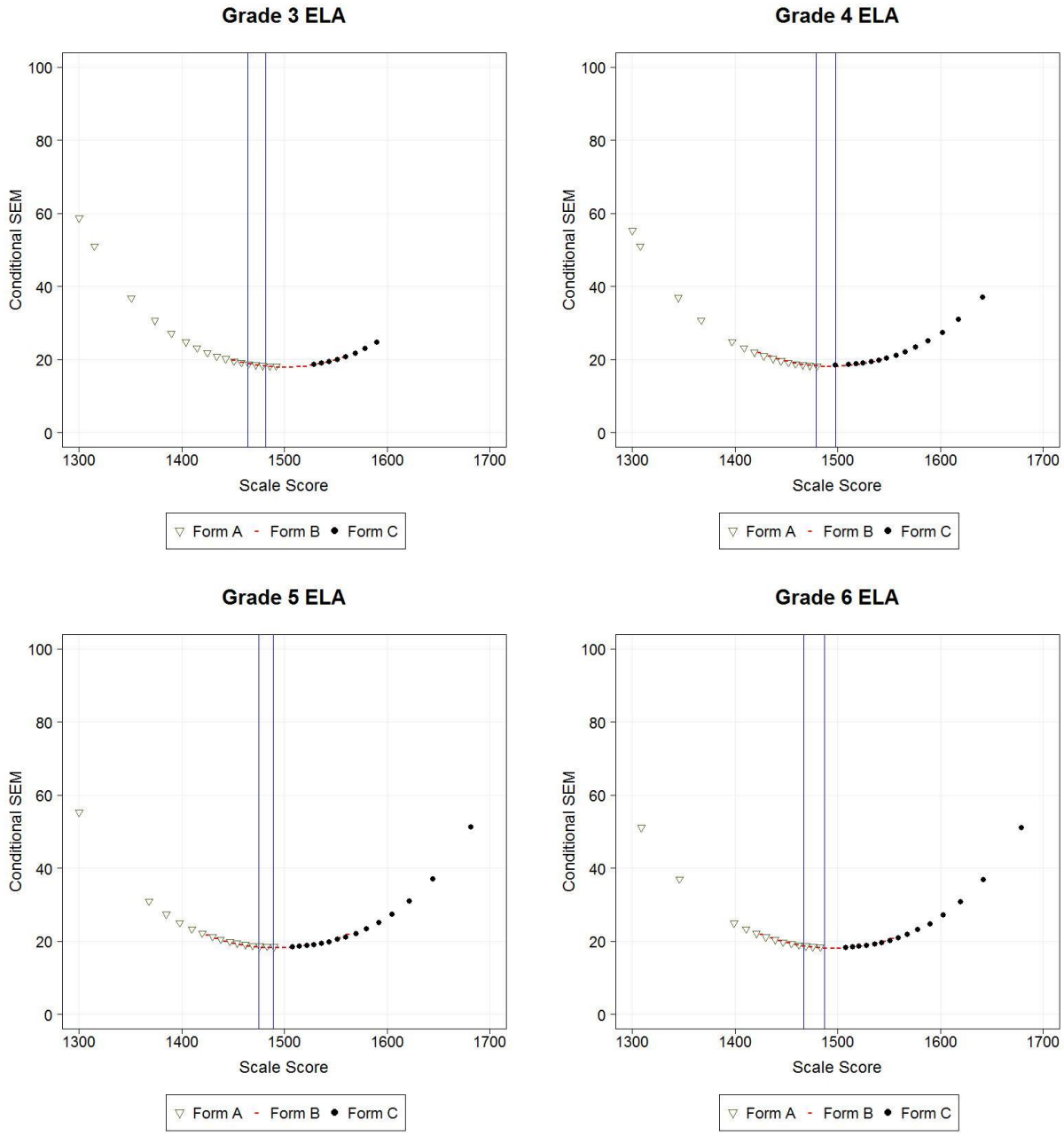
$$se(\hat{\theta}) = \frac{1}{\sqrt{-\left(\frac{\partial^2 lnL(\hat{\theta})}{\partial^2 \theta}\right)}},$$

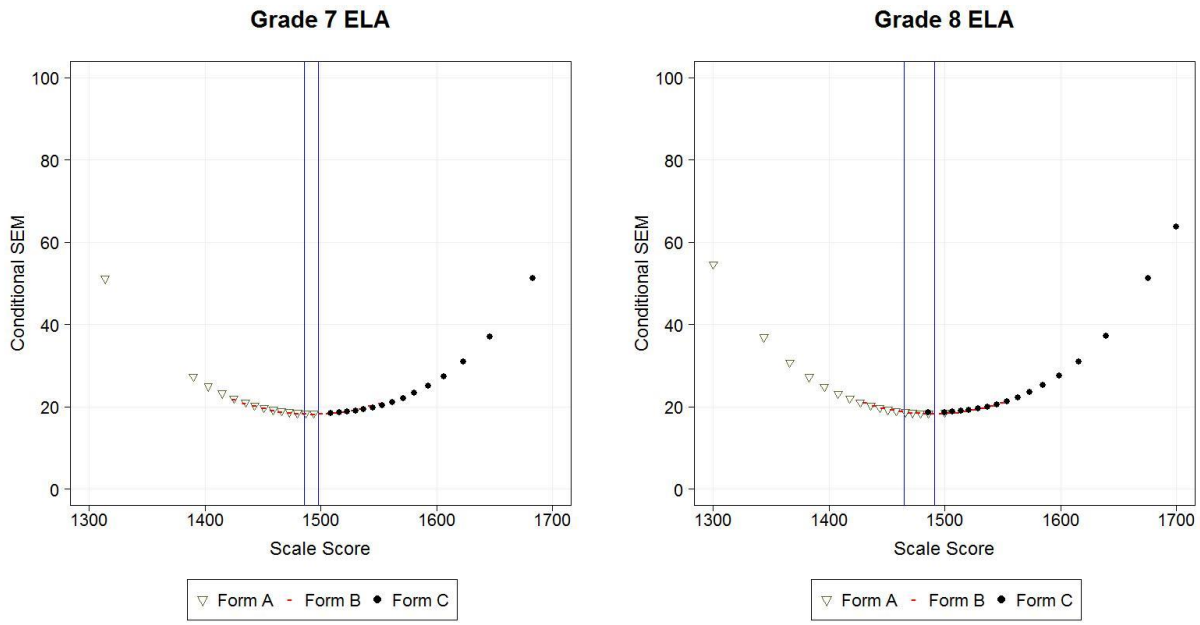
where

$$\frac{\partial^2 lnL(\hat{\theta})}{\partial^2 \theta} = \sum_{i=1}^{N_{PCM}} \left(\left(\frac{\sum_{j=1}^{m_i} j exp\left(\sum_{k=1}^{j}(\hat{\theta} - b_{ik})\right)}{1 + \sum_{j=1}^{m_i} exp\left(\sum_{k=1}^{j}(\hat{\theta} - b_{ik})\right)} \right)^2 - \frac{\sum_{j=1}^{m_i} j^2 exp\left(\sum_{k=1}^{j}(\hat{\theta} - b_{ik})\right)}{1 + \sum_{j=1}^{m_i} exp\left(\sum_{k=1}^{j}(\hat{\theta} - b_{ik})\right)} \right),$$

where N_{PCM} is the number of items that are scored using PCM items. Figure 2–Figure 5 (also the CSEM in Appendix B, Conditional Standard Error of Measurement) are based on the Hessian estimates. Vertical lines in the plots represent the Approaching Proficiency and At Proficiency performance category cut scores respectively.

Figure 2: Conditional Standard Errors of Measurement (ELA)




Grade 7 ELA
Conditional SEM
Scale Score
Form A
Form B
Form C
Grade 8 ELA
Conditional SEM
Scale Score
Form A
Form B
Form C

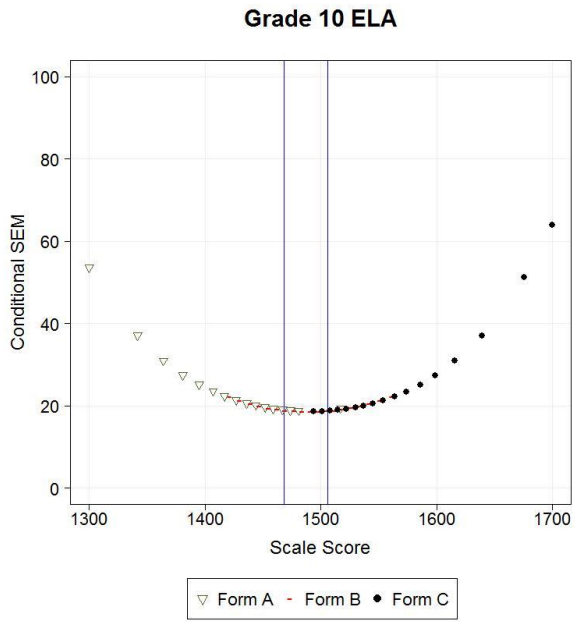
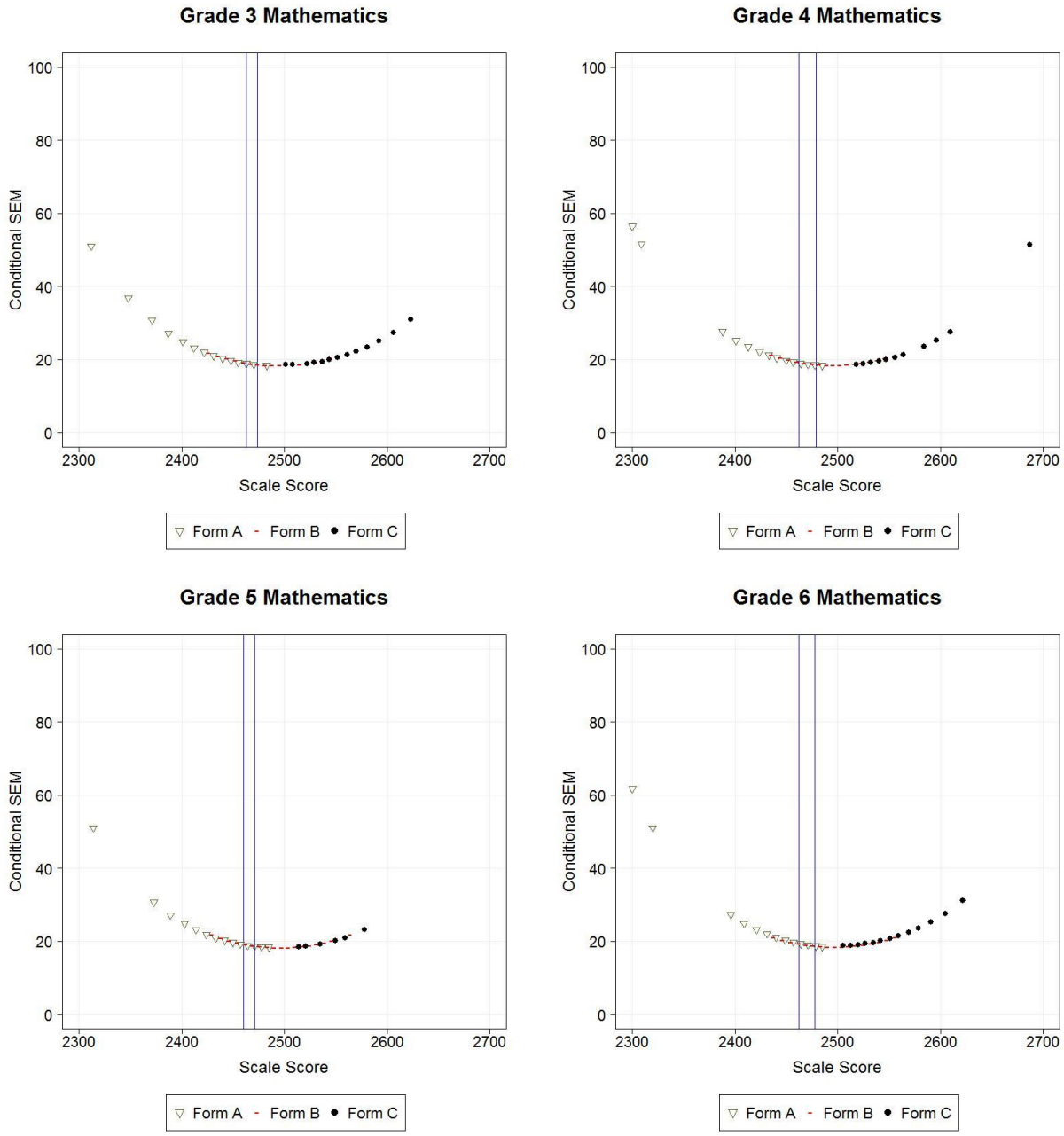
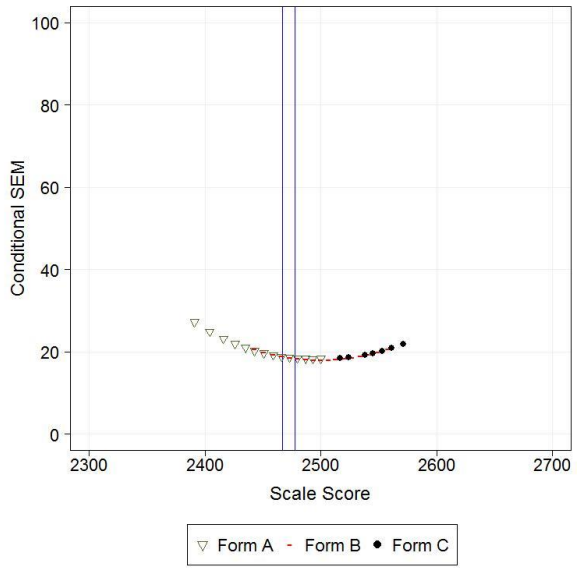

Grade 10 ELA
Conditional SEM
Scale Score
Form A
Form B
Form C

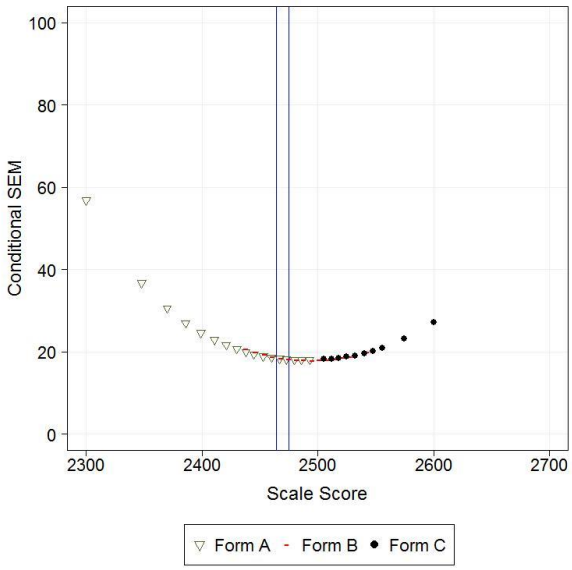
Figure 3: Conditional Standard Errors of Measurement (Mathematics)



Grade 7 Mathematics



Grade 8 Mathematics



Grade 10 Mathematics

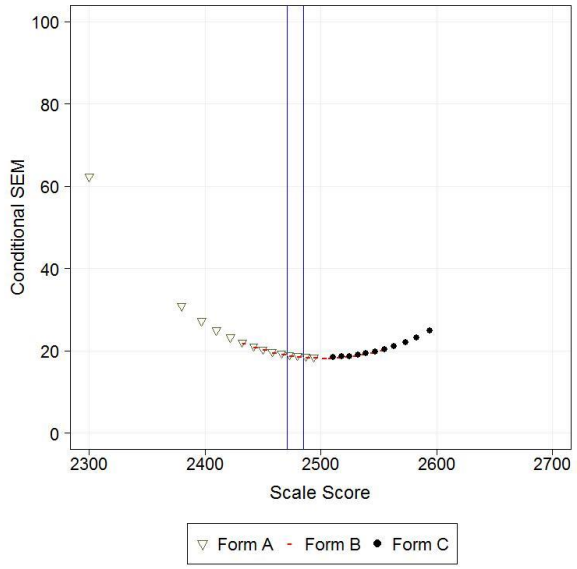


Figure 4: Conditional Standard Errors of Measurement (Science)

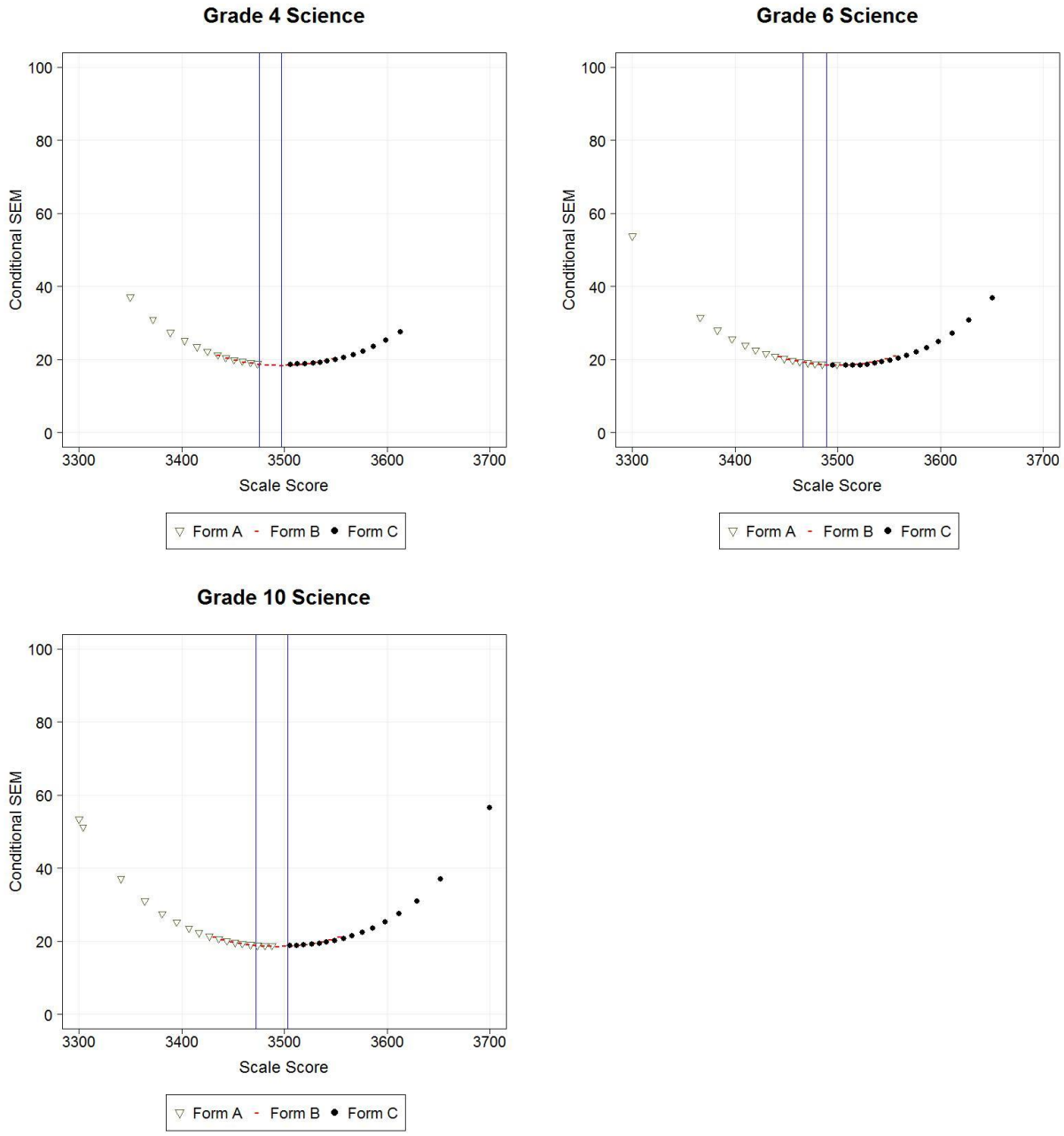
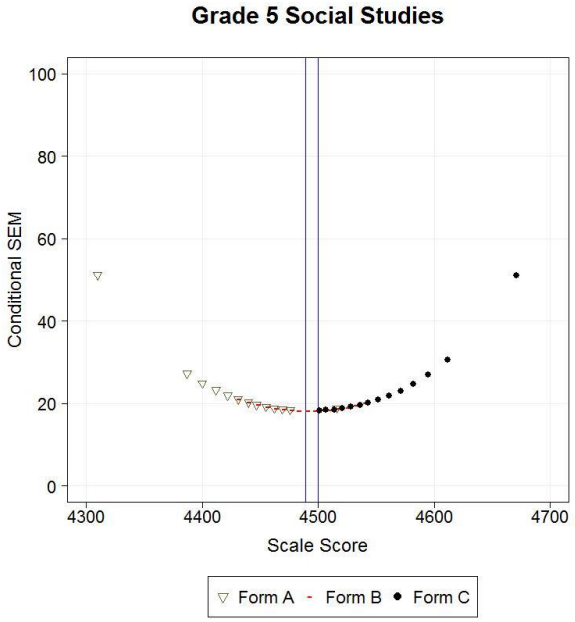


Figure 5: Conditional Standard Errors of Measurement (Social Studies)



Generally, the relationship between CSEM and scale score is U-shaped, with large CSEM values at the lower and upper ends of the scale score range and the smallest CSEMs in the middle, approximately at the Approaching Proficiency and At Proficiency cut scores. The middle section of the scale scores has most of the measurement information, and therefore, the CSEM remains low. The higher CSEMs at the lower and upper ends indicate a lack of easier or harder items compared to student ability.

Reliability coefficients by demographic subgroups are also presented in Appendix A, Reliability Coefficients. Appendix B, Conditional Standard Error of Measurement, includes scale-score-by-scale-score CSEMs and corresponding performance levels for each scale score.

6.3 RELIABILITY OF PERFORMANCE CLASSIFICATION

When students complete an I AM assessment, they are placed into performance levels depending upon their observed scaled score. The cut scores for student classification into the different performance levels were determined after the I AM standard-setting process. A complete description of the standard-setting process can be found in Volume 6 of 2018-2019 I AM Technical Report, *Standard Setting Report*.

6.3.1 Classification Accuracy

Misclassification probabilities are computed for all performance-level standards (i.e., for the cut scores between levels 1 and 2 and the cut scores between and levels 2 and 3). The performance-level cut score between level 2 and level 3 is of primary interest because this cut score is used to classify students as Approaching Proficiency or At Proficiency. Students with observed scores far from the level 3 cut score are expected to

be classified more accurately as At Proficiency or Approaching Proficiency than students with scores near this cut score.

This report estimates classification reliabilities using two different methods: one based on observed abilities and a second based on estimating a latent posterior distribution for the true scores.

Two approaches for estimating classification probabilities are provided. The first is an observed-score approach to computing misclassification probabilities and is designed to explore the following research questions:

1. What is the overall classification accuracy index (CAI) of the total test?
2. What is the classification accuracy rate index for each individual performance cut score within the test?

The second approach computes misclassification probabilities using an IRT-based method for students scoring at each score point. This approach is designed to explore the following research questions:

1. What is the probability that the student's true score is below the cut score point?
2. What is the probability that the student's true score is above the cut score point?

Both approaches yield student-specific classification probabilities that can be aggregated to form overall misclassification rates for the test.

For these analyses, we used students from the Spring 2019 I AM population data files that included the status of reported scores. Table 33 provides the sample size, mean, and standard deviation of the observed theta scores. The theta scores are based on the MLEs obtained from CAI's scoring engine.

Table 33: Descriptive Statistics by Test

Test	Sample Size	Average Theta	Standard Deviation of Theta	Average Scale Score	Standard Deviation of Scale Score
ELA 3	565	-0.49	0.77	1475	39
ELA 4	627	-0.25	0.88	1487	44
ELA 5	671	-0.15	0.88	1492	44
ELA 6	749	-0.21	0.91	1489	45
ELA 7	808	0.19	0.99	1509	49
ELA 8	971	-0.16	1.05	1491	53
ELA 10	951	0.23	1.13	1511	57
Mathematics 3	568	-0.50	0.72	2475	36
Mathematics 4	629	-0.46	0.70	2477	35

Test	Sample Size	Average Theta	Standard Deviation of Theta	Average Scale Score	Standard Deviation of Scale Score
Mathematics 5	666	-0.59	0.55	2470	28
Mathematics 6	746	-0.39	0.70	2480	35
Mathematics 7	804	-0.45	0.52	2477	26
Mathematics 8	966	-0.67	0.55	2466	27
Mathematics 10	944	-0.45	0.56	2477	28
Science 4	622	-0.25	0.75	3487	37
Science 6	740	-0.28	0.80	3485	40
Biology	963	-0.04	0.95	3497	48
Social Studies 5	660	-0.29	0.86	4485	43

The observed score approach (Rudner, 2001) implemented to assess classification accuracy is based on the probability that the true score, θ, for student j is within performance level $l = 1,2,\cdots,L$. This probability can be estimated from evaluating the integral

$$p_{jl} = P\,r\left(c_{lower} \leq \theta_j < c_{upper} \middle| \hat{\theta}_j, \hat{\sigma}_j^2\right) = \int_{c_{lower}}^{c_{upper}} f\left(\hat{\theta}_j, \hat{\sigma}_j^2\right) d\theta_j,$$

where c_{upper} and c_{lower} denote the score corresponding to the upper and lower limits of the performance level, respectively. $\hat{\theta}_j$ is the ability estimate of the *j*th student with a SEM of $\hat{\sigma}_j$, and using the asymptotic property of normality of the MLE, $\hat{\theta}_j$, we take $f(\cdot)$ as asymmetrically normal. Therefore, the above probability can be estimated by

$$p_{jl} = \Phi\left(\frac{c_{upper} - \hat{\theta}_j}{\hat{\sigma}_j}\right) - \Phi\left(\frac{c_{lower} - \hat{\theta}_j}{\hat{\sigma}_j}\right),$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (CDF). The expected number of students at level *l* based on students from observed level *v* can be expressed as

$$E_{vl} = \sum_{pl_i \,\epsilon\, v} p_{jl},$$

where pl_j is the *j*th student's performance level, the values of E_{vl} are the elements used to populate the matrix E, a 3×3 matrix of conditionally expected numbers of students to score within each performance-level bin based on their true scores. The overall CAI of the test can then be estimated from the diagonal elements of the matrix

$$CAI = \frac{tr(E)}{N},$$

where $N = \sum_{v=1}^{3} N_v, N_v$ is the observed number of students scoring in performance level v. The classification accuracy index for the individual cut score, p, ($CAIC_p$) is estimated by forming square partitioned blocks of the matrix E and taking the summation over all elements within the block as follows:

$$CAIC_p = \frac{\left(\sum_{v=1}^{p}\sum_{l=1}^{p} E_{vl} + \sum_{v=p+1}^{3}\sum_{l=p+1}^{3} E_{vl}\right)}{N}.$$

The *p*th cut score is $p(p = 1, 2)$.

The IRT-based approach (Guo, 2006) makes use of student-level item response data from the 2019 I AM assessment administration. For the *j*th student, we can estimate a posterior probability distribution for the latent true score and from this estimate, the probability that a true score is above the cut score as

$$p(\theta_j \geq c) = \frac{\int_c^{\infty} p(\theta_j) f(\mu, \sigma) d\theta_j}{\int_{-\infty}^{\infty} p(\theta_j) f(\mu, \sigma) d\theta_j},$$

where c is the cut score required for passing in the same assigned metric, θ_j is true ability in the true-score metric, z_j is the item score, μ is the mean, and σ is the standard deviation of the population distribution. The function $p(\theta_j)$ is the probability of the particular pattern of responses given the theta, and $f(\theta)$ is the density of the proficiency θ in the population.

Similarly, we can estimate the probability that a true score is below the cut score as

$$p(\theta_j < c) = \frac{\int_{-\infty}^{c} p(\theta_j) f(\mu, \sigma) d\theta_j}{\int_{-\infty}^{\infty} p(\theta_j) f(\mu, \sigma) d\theta_j}.$$

From these misclassification probabilities, we can estimate the overall false positive rate (FPR) and false negative rate (FNR) of the test. The FPR is expressed as the proportion of individuals who scored above the cut score based on their observed score, but their true score would otherwise have classified them as below the cut score. The FNR is expressed as the proportion of individuals who scored below the cut score based on their observed score, but otherwise would have been classified as above the cut score based on their true scores. These rates are estimated as follows:

$$FPR = \frac{\sum_{j \in \hat{\theta}_j \geq c} p(\theta_j < c)}{N}$$

$$FNR = \frac{\sum_{j \in \hat{\theta}_j < c} p(\theta_j \geq c)}{N}$$

Table 34–Table 37 provide the overall CAI and the classification accuracy index for the individual cut scores $(CAIC)$ based on the observed score approach (Rudner, 2001). The overall classification accuracy of the test ranges from 0.719–0.808 for ELA, from 0.665–

0.721 for Mathematics, and from 0.723–0.770 for Science. The overall classification accuracy of Social Studies grade 5 is 0.818. There is no industry standard, but these numbers suggest that misclassification would not be frequent in the population data.

The overall cut-score accuracy rates are much higher, denoting that the degree to which we can reliably differentiate students between adjacent performance levels is typically from 0.836–0.903 for ELA, Science, and Social Studies, and from 0.793–0.829 for Mathematics.

Table 34: Classification Accuracy Index (ELA)

Grade	Overall Accuracy Index	Cut Score Accuracy Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3
3	0.719	0.854	0.836
4	0.764	0.863	0.874
5	0.761	0.850	0.862
6	0.735	0.846	0.866
7	0.808	0.879	0.884
8	0.759	0.870	0.879
10	0.802	0.897	0.903

Table 35: Classification Accuracy Index (Mathematics)

Grade	Overall Accuracy Index	Cut Score Accuracy Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3
3	0.721	0.820	0.815
4	0.695	0.829	0.822
5	0.672	0.793	0.795
6	0.687	0.823	0.826
7	0.710	0.809	0.799
8	0.694	0.797	0.798
10	0.665	0.793	0.807

Table 36: Classification Accuracy Index (Science)

Grade	Overall Accuracy Index	Cut Score Accuracy Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3
4	0.723	0.836	0.861
6	0.727	0.856	0.850
Biology	0.770	0.884	0.879

Table 37: Classification Accuracy Index (Social Studies)

Grade	Overall Accuracy Index	Cut Score Accuracy Index	
		Cut 1 and Cut 2	Cut 2 and Cut 3
5	0.818	0.878	0.893

Table 38–Table 41 provide the FPR, FNR and accuracy index based on the IRT-based method (Lee, Hanson, & Brennan, 2002; Guo, 2006). In ELA, Science, and Social Studies, the FPR and FNR rates for the level 1 and level 2 cut scores are around 4–10%, and the rates for the level 2 and level 3 cut scores are around 4–9%. In Mathematics, the rates are around 6–12% at the cut scores between levels 1 and 2 and 7–12% at the cut scores between levels 2 and 3.

Table 38: False Classification Rates (ELA)

	1/2 cut			2/3 cut		
Grade	**FPR**	**FNR**	**Accuracy**	**FPR**	**FNR**	**Accuracy**
3	0.091	0.054	0.855	0.068	0.093	0.839
4	0.078	0.056	0.866	0.069	0.054	0.877
5	0.063	0.084	0.853	0.053	0.083	0.865
6	0.096	0.057	0.847	0.057	0.074	0.869
7	0.067	0.051	0.882	0.059	0.055	0.887
8	0.065	0.063	0.872	0.035	0.084	0.881
10	0.062	0.040	0.898	0.044	0.051	0.905

Table 39: False Classification Rates (Mathematics)

	1/2 cut			2/3 cut		
Grade	**FPR**	**FNR**	**Accuracy**	**FPR**	**FNR**	**Accuracy**
3	0.104	0.075	0.821	0.104	0.078	0.817
4	0.095	0.075	0.830	0.078	0.097	0.826
5	0.111	0.094	0.795	0.082	0.118	0.800
6	0.118	0.058	0.824	0.068	0.102	0.830
7	0.085	0.104	0.811	0.109	0.089	0.802
8	0.104	0.097	0.799	0.111	0.088	0.801
10	0.105	0.099	0.795	0.074	0.114	0.812

Table 40: False Classification Rates (Science)

	1/2 cut			2/3 cut		
Grade	**FPR**	**FNR**	**Accuracy**	**FPR**	**FNR**	**Accuracy**
4	0.098	0.064	0.838	0.074	0.062	0.864
6	0.089	0.054	0.857	0.077	0.071	0.852
Biology	0.054	0.061	0.886	0.053	0.066	0.881

Table 41: False Classification Rates (Social Studies)

	1/2 cut			2/3 cut		
Grade	**FPR**	**FNR**	**Accuracy**	**FPR**	**FNR**	**Accuracy**
5	0.051	0.068	0.881	0.038	0.066	0.896

6.3.2 Classification Consistency

Classification accuracy refers to the degree to which a student's true score and observed score would fall within the same performance level (Rudner, 2001). Classification consistency refers to the degree to which test takers are classified into the same performance level, assuming the test is administered independently twice (Lee, Hanson, and Brennan, 2002)—that is, the percentages of students who are consistently classified in the same performance levels on two equivalent test forms. In reality, the true ability is unknown, and students do not take an alternate, equivalent form; therefore, classification consistency is estimated based on students' item scores and the item parameters, and the assumed underlying latent ability distribution.

The classification consistency index for the individual cut score, c, ($CICC_c$) was estimated using the following equation:

$$CICC_c = \frac{\sum_j \{p^2(\theta_j \geq c) + p^2(\theta_j < c)\}}{N}$$

Classification consistency with classification accuracy results based on the IRT-based method (Lee, Hanson, & Brennan, 2002) are presented in Table 42 and Table 43. In cut score 1/2 and cut score 2/3 results, all accuracy values in ELA, Science, and Social Studies are higher than 0.84, and consistency values are higher than 0.78. In Mathematics, all accuracy values are higher than 0.80 and consistency values are higher than 0.72. In all performance levels, classification accuracy is slightly higher than classification consistency. Classification consistency rates can be lower than classification accuracy because the consistency is based on two tests with SEMs, while the accuracy is based on one test with an SEM and the true score. The accuracy and consistency rates for each performance level are higher for the levels with a smaller SEM.

Table 42: Classification Accuracy and Consistency (Cut Score 1 and Cut Score 2)

Grade	Accuracy	Consistency
ELA 3	0.855	0.798
ELA 4	0.866	0.815
ELA 5	0.853	0.798
ELA 6	0.847	0.794
ELA 7	0.882	0.834
ELA 8	0.872	0.822
ELA 10	0.898	0.861
Mathematics 3	0.821	0.758
Mathematics 4	0.830	0.766
Mathematics 5	0.795	0.722
Mathematics 6	0.824	0.763
Mathematics 7	0.811	0.742
Mathematics 8	0.799	0.725
Mathematics 10	0.795	0.721
Science 4	0.838	0.780
Science 6	0.857	0.800
Biology	0.886	0.841
Social Studies 5	0.881	0.832

Table 43: Classification Accuracy and Consistency (Cut Score 2 and Cut Score 3)

Grade	Accuracy	Consistency
ELA 3	0.839	0.780
ELA 4	0.877	0.827
ELA 5	0.865	0.811
ELA 6	0.869	0.814
ELA 7	0.887	0.840
ELA 8	0.881	0.836
ELA 10	0.905	0.867
Mathematics 3	0.817	0.753
Mathematics 4	0.826	0.760
Mathematics 5	0.800	0.726
Mathematics 6	0.830	0.766

Grade	Accuracy	Consistency
Mathematics 7	0.802	0.727
Mathematics 8	0.801	0.733
Mathematics 10	0.812	0.740
Science 4	0.864	0.811
Science 6	0.852	0.796
Biology	0.881	0.832
Social Studies 5	0.896	0.853

6.4 PRECISION AT CUT SCORES

Table 44–Table 47 present the mean CSEM at each performance level by grade and subject. The tables include performance-level cut scores and associated CSEM. If no observed scale score equals the cut scores, the CSEM of the observed score closest to the cut score is provided. The I AM assessment scores are somewhat more precise for test scores near the middle of the scale, especially around the At Proficiency performance standard cut score. The tables below also show that test scores remain precise even for students in the lowest and highest performance levels.

Table 44: Performance Levels and Associated Conditional Standard Error of Measurement (ELA)

Grade	Performance Level	*N*	Mean CSEM	Cut Score (Scale Score)	CSEM at Cut Scores
3	1	193	21.78		
	2	136	18.64	1464	18.79
	3	236	18.51	1482	18.39
4	1	262	20.23		
	2	107	18.39	1479	18.47
	3	258	19.73	1498	18.40
5	1	268	20.01		
	2	93	18.48	1475	18.66
	3	310	20.07	1489	18.53
6	1	230	20.47		
	2	181	18.68	1467	18.94
	3	338	20.03	1487	18.38
7	1	282	19.61		
	2	85	18.38	1486	18.40
	3	441	21.02	1498	18.40
8	1	322	20.61		
	2	257	18.57	1465	18.70
	3	392	21.87	1491	18.45
10	1	214	20.69		
	2	260	18.77	1468	19.00
	3	477	23.10	1506	18.78

Table 45: Performance Levels and Associated Conditional Standard Error of Measurement (Mathematics)

Grade	Performance Level	*N*	Mean CSEM	Cut Score (Scale Score)	CSEM at Cut Scores
3	1	204	20.73		
	2	71	18.89	2463	18.85
	3	293	18.92	2474	18.70
4	1	209	20.95		
	2	141	18.91	2462	19.21
	3	279	19.01	2479	18.46
5	1	248	20.36		
	2	127	18.99	2460	19.22
	3	291	18.55	2471	18.57
6	1	209	21.15		
	2	199	19.06	2462	19.36
	3	338	19.09	2478	18.69
7	1	314	19.92		
	2	77	18.68	2467	18.93
	3	413	18.34	2478	18.41
8	1	446	20.05		
	2	119	18.52	2464	18.60
	3	401	18.25	2475	18.31
10	1	404	20.41		
	2	207	18.82	2471	18.99
	3	333	18.62	2485	18.56

Table 46: Performance Levels and Associated Conditional Standard Error of Measurement (Science)

Grade	Performance Level	*N*	Mean CSEM	Cut Score (Scale Score)	CSEM at Cut Scores
4	1	238	20.22		
	2	140	18.75	3476	18.90
	3	244	19.36	3497	18.56
6	1	224	21.02		
	2	182	19.03	3466	19.35
	3	334	19.29	3489	18.69
Biology	1	307	20.62		
	2	239	18.80	3472	19.06
	3	417	20.52	3503	18.78

Table 47: Performance Levels and Associated Conditional Standard Error of Measurement (Social Studies)

Grade	Performance Level	*N*	Mean CSEM	Cut Score (Scale Score)	CSEM at Cut Scores
5	1	388	19.64		
	2	62	18.30	4489	18.29
	3	210	20.33	4500	18.22

SUMMARY

This report is intended to provide a collection of validity and reliability evidence that supports appropriate inferences from the observed test scores. The overall results can be summarized as follows:

- **Content Validity.** Evidence is provided to support the assertion that content coverage on each test was consistent with the assessment specifications of the blueprint.
- **Internal Structural Validity.** Evidence is provided to support the selection of the measurement model, the tenability of local independence, and the reporting of an overall score and subscores at the reporting category levels.
- **Reliability.** Various measures of reliability are provided at the aggregate and subgroup levels, showing that the reliability of all assessments is in line with acceptable industry standards.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *The Standards for Educational and Psychological Testing.* Washington, DC: Author.

Bejar, I. I. (1980). Biased assessment of program impact due to psychometric artifacts. *Psychological Bulletin, 87*(3), 513–524.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chen, F., Bollen, K. A., Paxton, P., Curran, P., & Kirby, J. 2001. Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*, 468–508.

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.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213–220.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

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

Feldt, L. S., & Brennan, R. (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 105–146). New York: Macmillan.

Feldt, L. S., & Qualls, A. L. (1996). Bias in coefficient alpha arising from heterogeneity of test content. *Applied Measurement in Education, 9*, 277–286.

Guo, F. (2006). Expected classification accuracy using the latent distribution. *Practical Assessment, Research & Evaluation, 11*(6).

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*(3), 381–389.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika, 2*(3), 151–160.

Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education, 12*, 237–255.

Lee, W., Hanson, B., & Brennan, R. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26(4)*, 412–432.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13–103). New York: Macmillan.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.

Muthén, B. O., du Toit, S.H.C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460.

Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education, 8*, 111–120.

Raju, N. S. (1977). A generalization of coefficient alpha. *Psychometrika, 42*, 549–565.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment, Research & Evaluation, 7*(14).

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 2002, from http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html.

van Driel, Otto P. (1978). On various causes of improper solutions in maximum likelihood Factor analysis. *Psychometrika, 43*, 225–43.

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*, 125–145.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213.

Yoon, B., & Young, M. J. (2000). *Estimating the reliability for test scores with mixed item formats: Internal consistency and generalizability.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Indiana's Alternate Measure

2020–2021

Volume 5
Score Interpretation Guide

ACKNOWLEDGMENTS

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

Major contributors to this technical report include the following staff from Cambium Assessment, Inc. (CAI): Stephan Ahadi, Yuan Hong, Hyesuk Jang, Hashim Evans, Anne Mogilnicki, Grace (Sungeun) Chung, Cherise Lesesne, and Celine Bryan. Major contributors from the Indiana Department of Education (IDOE) include the assessment director, the assistant assessment director, and program leads.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. I AM SCORE REPORTS

In Spring 2021, pursuant to Chapter 5 of Indiana Code 20-32-5, "Indiana Statewide Testing for Educational Progress," the following Indiana's Alternate Measure (I AM) assessments were administered to Indiana students: English/Language Arts (ELA) and Mathematics for grades 3–8 and 10; Science for grades 4 and 6, and Biology for high school; and Social Studies for grade 5.

The purpose of this volume is to describe the information available from the scores reported for the 2020-2021 I AM assessments, and to define appropriate uses and inferences that can be drawn from them. This volume also documents the features of the score reports provided through the Indiana Online Reporting System (ORS), which is designed to assist stakeholders in reviewing, downloading, and appropriately interpreting test results.

1.1 OVERVIEW OF I AM SCORE REPORTS

I AM assessments were administered in Spring 2021. Scores from each Spring 2021 assessment were provided to corporations and schools through the ORS on June 10, 2021. The ORS provides information on student performance and aggregated summaries at several levels—state, corporation, school, and roster.

The ORS is a web-based application that provides I AM results to users at various levels. Assessment results are available to users on the basis of their roles and the privileges granted to each authenticated user. There are four types of access: (1) state, (2) corporation, (3) school, and (4) teacher. 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, schools can access data only for the students in their school, and corporations can access data for all schools and students in their corporation.

Users have the following types of access to the ORS:

- 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 ORS is password protected; users can access data at their assigned level and below. For example, an STC can access the school report of students for their school but not for another school.

1.2 Overall Scores and Reporting Categories

Students receive a single scale score for each subject assessed if there is a valid score to report. The validity of a score is determined using attempted rules, which define a set of parameters under which a student's attemptedness may be counted. All students begin the assessment with two practice items. Next, students are presented with three operational items. Educators document if the students are able to engage with these first five items. Students that are not able to engage are marked as 'no response.' If the student has Mark as No Response (NR) for the first two practice items and first three operational items, a student's score will be assigned as No Mode of Communication (NMC). For the students not assigned as NMC, if the student fails to respond to five test items in segments 1 and 2, they will be assigned as Undetermined (UND). Students are counted as 'attempted' if they are not assigned NMC or UND. Attempted students will be scored and provided a proficiency-level designation. Normally, a student takes an assessment in the Test Delivery System (TDS) and then submits it. Once the assessment is completed in the TDS, the assessment score is reported in the ORS. However, assessments may also be manually invalidated before reaching the ORS if assessment irregularities occur (e.g., cheating, unscheduled interruptions).

A student's score is based only on the operational items on the assessment. A scale score is an estimate of students' knowledge and skills measured by the assessment and is used to describe how well a student performed on an assessment. The scale score is transformed from a theta score, which is estimated based on item response theory (IRT) models as described in Volume 1, Annual Technical Report. Lower scale scores indicate that the student does not possess sufficient knowledge and skills measured by the assessment. Conversely, higher scale scores indicate that the student has more proficient knowledge and skills measured by the assessment. Interpretation of scale scores is more meaningful when the scale scores are used along with performance levels and Performance-Level Descriptors (PLDs).

A student's scale score determines his or her overall performance level. Performance levels on an assessment correlate with proficiency categories into which students fall on the basis of their scale scores. For I AM, scale scores are mapped onto three performance levels:

- Level 1: Below Proficiency
- Level 2: Approaching Proficiency
- Level 3: At Proficiency

PLDs set out content-area knowledge and skills that students at each performance level are expected to possess; and cut scores, unique to each grade and subject, are determined by using cut points listed in Section 2.4, Cut Scores; additional details can be found on the Indiana Department of Education (IDOE) web page.

Performance levels can be interpreted on the basis of PLDs, which represent a more descriptive analysis of a student's abilities. Generally, students performing on I AM at Level 3 have met current grade-level Content Connectors by demonstrating essential knowledge, application, and skills to be on track for post-secondary education or competitive integrated employment.

In addition to an overall score, students receive reporting category scores. Reporting categories represent distinct groups of knowledge within each grade and subject. For I AM, students' performance in each reporting category is reported using a percent-correct score for each student and an average percent-correct score for aggregate reporting. Tables 1 through 4 display the reporting categories by grade and subject.

Table 1: Reporting Categories for ELA

Grade	Reporting Category
3	1. Key Ideas and Textual Support/Vocabulary 2. Reading Foundations 3. Structural Elements and Organization/Connection of Ideas/Media Literacy 4. Writing
4	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing
5	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing
6	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing
7	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing
8	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing
10	1. Key Ideas and Textual Support/Vocabulary 2. Structural Elements and Organization/Connection of Ideas/Media Literacy 3. Writing

Table 2: Reporting Categories for Mathematics

Grade	Reporting Category
3	1. Algebraic Thinking and Data Analysis 2. Computation 3. Geometry and Measurement 4. Number Sense
4	1. Algebraic Thinking and Data Analysis 2. Computation 3. Geometry and Measurement 4. Number Sense
5	1. Algebraic Thinking 2. Computation 3. Geometry and Measurement, Data Analysis, and Statistics 4. Number Sense
6	1. Algebra and Functions 2. Computation 3. Geometry and Measurement, Data Analysis, and Statistics 4. Number Sense
7	1. Algebra and Functions 2. Data Analysis, Statistics, and Probability 3. Geometry and Measurement 4. Number Sense and Computation
8	1. Algebra and Functions 2. Data Analysis, Statistics, and Probability 3. Geometry and Measurement 4. Number Sense and Computation
10	1. Equations and Inequalities (Linear and Systems) 2. Functions (Linear and Non-Linear) 3. Number Sense and Data Analysis 4. Geometry and Measurement

Table 3: Reporting Categories for Science

Grade	Reporting Category
4	1. Analyzing, Interpreting, and Computational Thinking 2. Explaining Solutions, Reasoning, and Communicating 3. Investigating 4. Questioning and Modeling
6	1. Analyzing, Interpreting, and Computational Thinking 2. Explaining Solutions, Reasoning, and Communicating 3. Investigating 4. Questioning and Modeling
Biology	1. Analyzing Data and Mathematical Thinking 2. Communicating Explanations and Evaluating Claims Using Evidence 3. Developing and Using Modeling to Describe Structure and Function

Table 4: Reporting Categories for Social Studies

Grade	Reporting Category
5	1. Civics and Government/History 2. Economics 3. Geography

1.3 ONLINE REPORTING SYSTEM

The ORS generates a set of online score reports that describe student performance for students, parents, educators, and other stakeholders. The online score reports are produced after the assessments are submitted by the students and processed into the ORS. In addition to each individual student's score report, the ORS produces aggregate score reports for teachers, schools, corporations, and states. 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.

Furthermore, 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 the school performance can be compared with the corporation performance and the state performance. If a teacher is selected, the summary results for the school, corporations, and state above the teacher are also provided for comparison purposes. Table 5 (in Section 1.4, Available Reports on the Indiana Online Reporting System) lists the following types of online reports: student, roster, teacher, school, and corporation.

1.4 AVAILABLE REPORTS ON THE I AM ONLINE REPORTING SYSTEM

The hierarchical structure of the Indiana ORS enables authorized users to view reports at their own level and at any lower level 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. CTCs can view the reports and data for all schools and students in their corporations.

Table 5 summarizes the types of score reports that are available in the ORS 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 ORS user guide, *Indiana State Assessment Online Reporting System User Guide*.

Table 5: 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	Text/CSV files containing overall and reporting category scale scores and performance levels along with demographic information		✓	✓	✓	✓	

1.5 REPORTING BY SUBGROUP

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 subgroup. For example, when the "Gender" subgroup is selected, the ORS 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 will have 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 6 presents the types of subgroups and subgroup categories provided in the ORS.

Table 6: Indiana List of Subgroups by Category

Subgroup	Subgroup Category
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	Special Education
	Not Special Education
Section 504 Plan	Section 504 Plan
	Not Section 504 Plan
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

1.6 REPORTS

1.6.1 Summary Performance Report

The home page allows authorized users to log in to the ORS and select "Score Reports," which contains summaries of student performance across grades and subjects. State personnel are able to view state summaries, corporation personnel see corporation summaries, school personnel see school summaries, and teachers see student summaries. State users can view a summary of students' performance within each corporation, as well. The Summary Performance Report

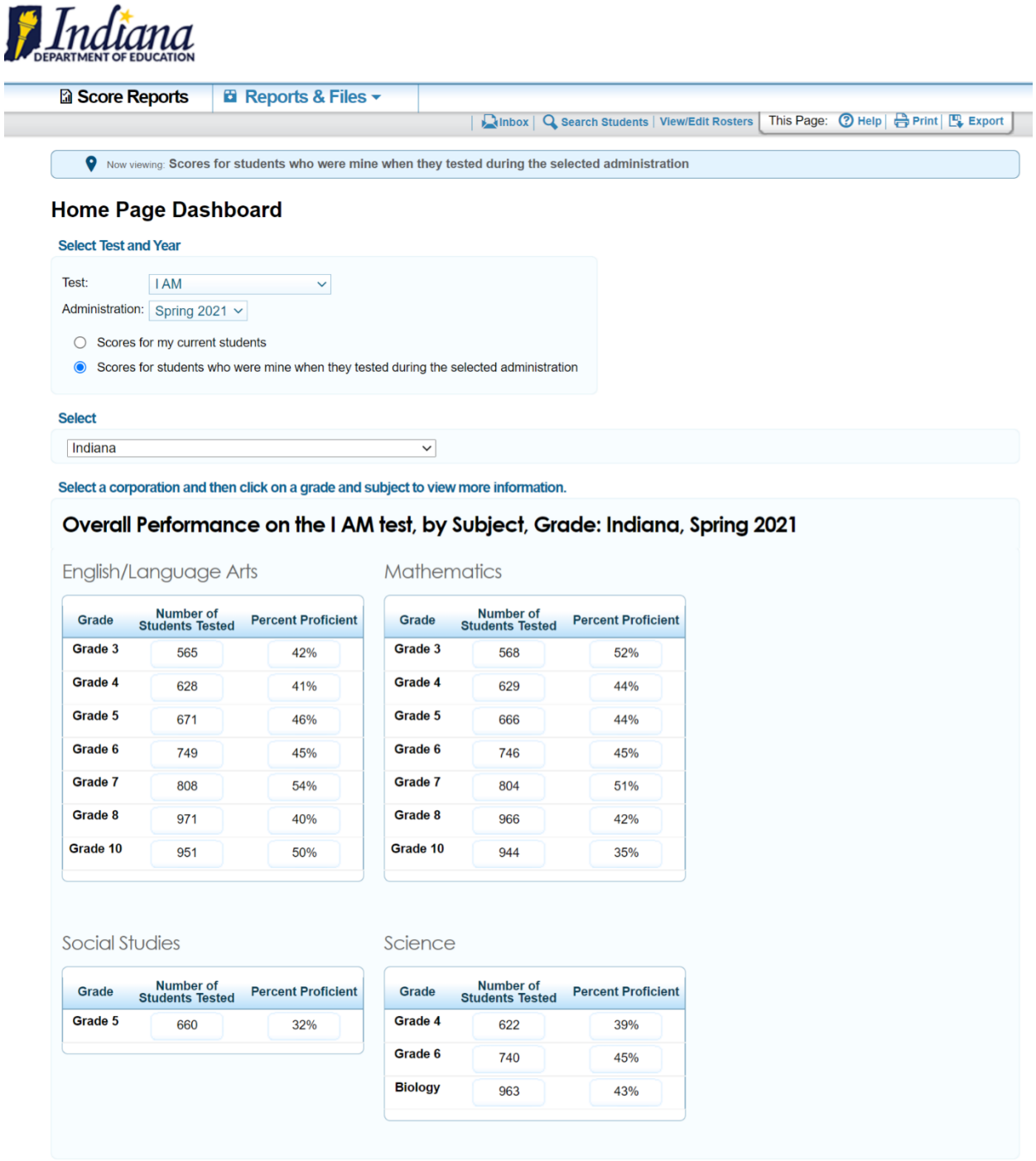
- displays summary data separated by grade and subject;
- bases the level of aggregation on a user's role; and
- reports the number of students assessed and percentage proficient.

The Summary Performance Report provides summaries of student performance, including the

- number of students assessed; and
- percentage proficient.

Figure 1 and Figure 2 present sample Summary Performance Reports at the state and corporation level.

Figure 1: Sample State Summary Performance Report



Indiana DEPARTMENT OF EDUCATION

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export

Now viewing: Scores for students who were mine when they tested during the selected administration

Home Page Dashboard

Select Test and Year

Test: I AM

Administration: Spring 2021

- Scores for my current students
- Scores for students who were mine when they tested during the selected administration

Select

Indiana

Select a corporation and then click on a grade and subject to view more information.

Overall Performance on the I AM test, by Subject, Grade: Indiana, Spring 2021

English/Language Arts

Grade	Number of Students Tested	Percent Proficient
Grade 3	565	42%
Grade 4	628	41%
Grade 5	671	46%
Grade 6	749	45%
Grade 7	808	54%
Grade 8	971	40%
Grade 10	951	50%

Mathematics

Grade	Number of Students Tested	Percent Proficient
Grade 3	568	52%
Grade 4	629	44%
Grade 5	666	44%
Grade 6	746	45%
Grade 7	804	51%
Grade 8	966	42%
Grade 10	944	35%

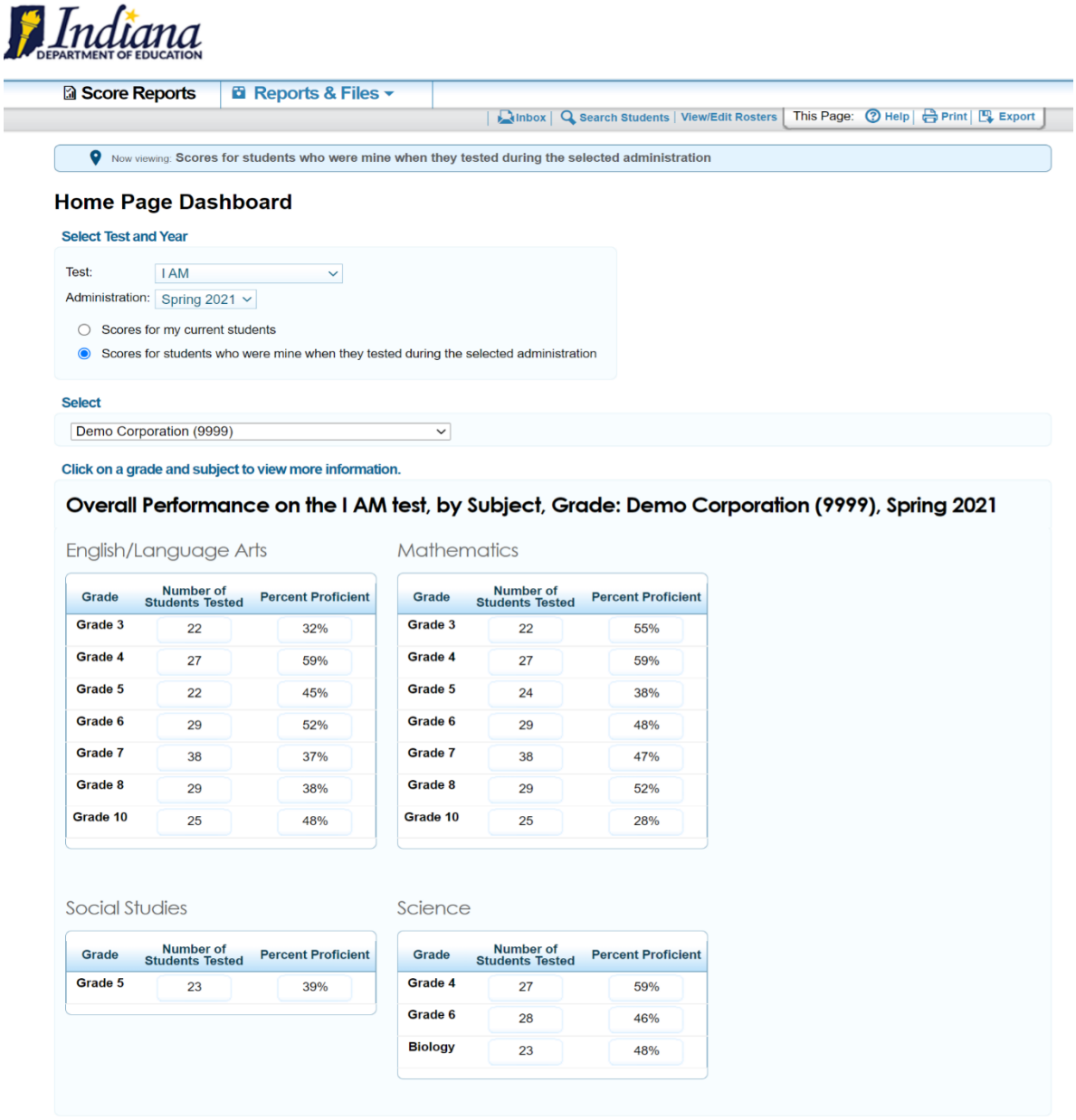
Social Studies

Grade	Number of Students Tested	Percent Proficient
Grade 5	660	32%

Science

Grade	Number of Students Tested	Percent Proficient
Grade 4	622	39%
Grade 6	740	45%
Biology	963	43%

Figure 2: Corporation-Level Summary Performance Report



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export

Now viewing: Scores for students who were mine when they tested during the selected administration

Home Page Dashboard

Select Test and Year

Test: I AM

Administration: Spring 2021

- Scores for my current students
- Scores for students who were mine when they tested during the selected administration

Select

Demo Corporation (9999)

Click on a grade and subject to view more information.

Overall Performance on the I AM test, by Subject, Grade: Demo Corporation (9999), Spring 2021

English/Language Arts

Grade	Number of Students Tested	Percent Proficient
Grade 3	22	32%
Grade 4	27	59%
Grade 5	22	45%
Grade 6	29	52%
Grade 7	38	37%
Grade 8	29	38%
Grade 10	25	48%

Mathematics

Grade	Number of Students Tested	Percent Proficient
Grade 3	22	55%
Grade 4	27	59%
Grade 5	24	38%
Grade 6	29	48%
Grade 7	38	47%
Grade 8	29	52%
Grade 10	25	28%

Social Studies

Grade	Number of Students Tested	Percent Proficient
Grade 5	23	39%

Science

Grade	Number of Students Tested	Percent Proficient
Grade 4	27	59%
Grade 6	28	46%
Biology	23	48%

The Corporation Summary Report is similar to the State Summary Report, except that in the Corporation Summary Report, summary data are displayed for all students in the selected corporation who have completed the selected assessment with a valid reported score.

1.6.2 Aggregate-Level Subject Report

Detailed summaries of student performance within a grade subject area are available within the Aggregate-Level Subject Report. The Aggregate-Level Subject Report presents results for the aggregate unit as well as results for the state and any higher-level aggregate units. For example, a school Aggregate-Level Subject Report will also contain

the summary results of the state and school corporation so that school performance can be compared with the above aggregate levels.

The Aggregate-Level Subject Report provides the aggregate summaries on a specific subject area, including the

- number of students;
- percentage proficient;
- number of students in each performance level; and
- percentage of students in each performance level.

The summaries are also presented for overall students and by subgroups. Figure 3 presents an example of Aggregate-Level Subject Reports for grade 8 ELA at the corporation level without subgroups. Figure 4 highlights grade 8 Mathematics at the corporation level when a user selects a subgroup of gender. Figures 5 and 6 present Science and Social Studies subject reports at corporation level.

Figure 3: Corporation Aggregate-Level Subject Report, Grade 8 ELA

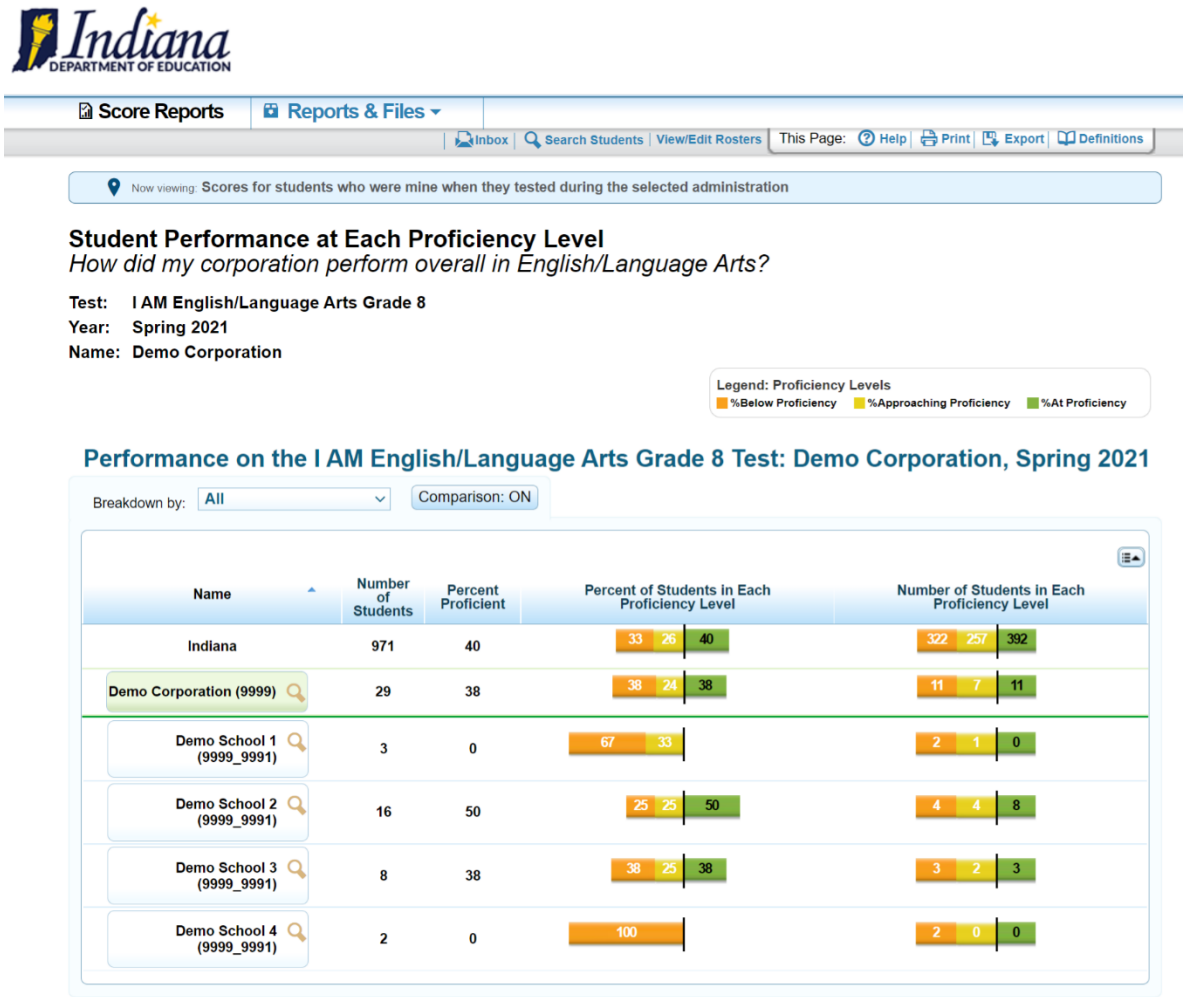
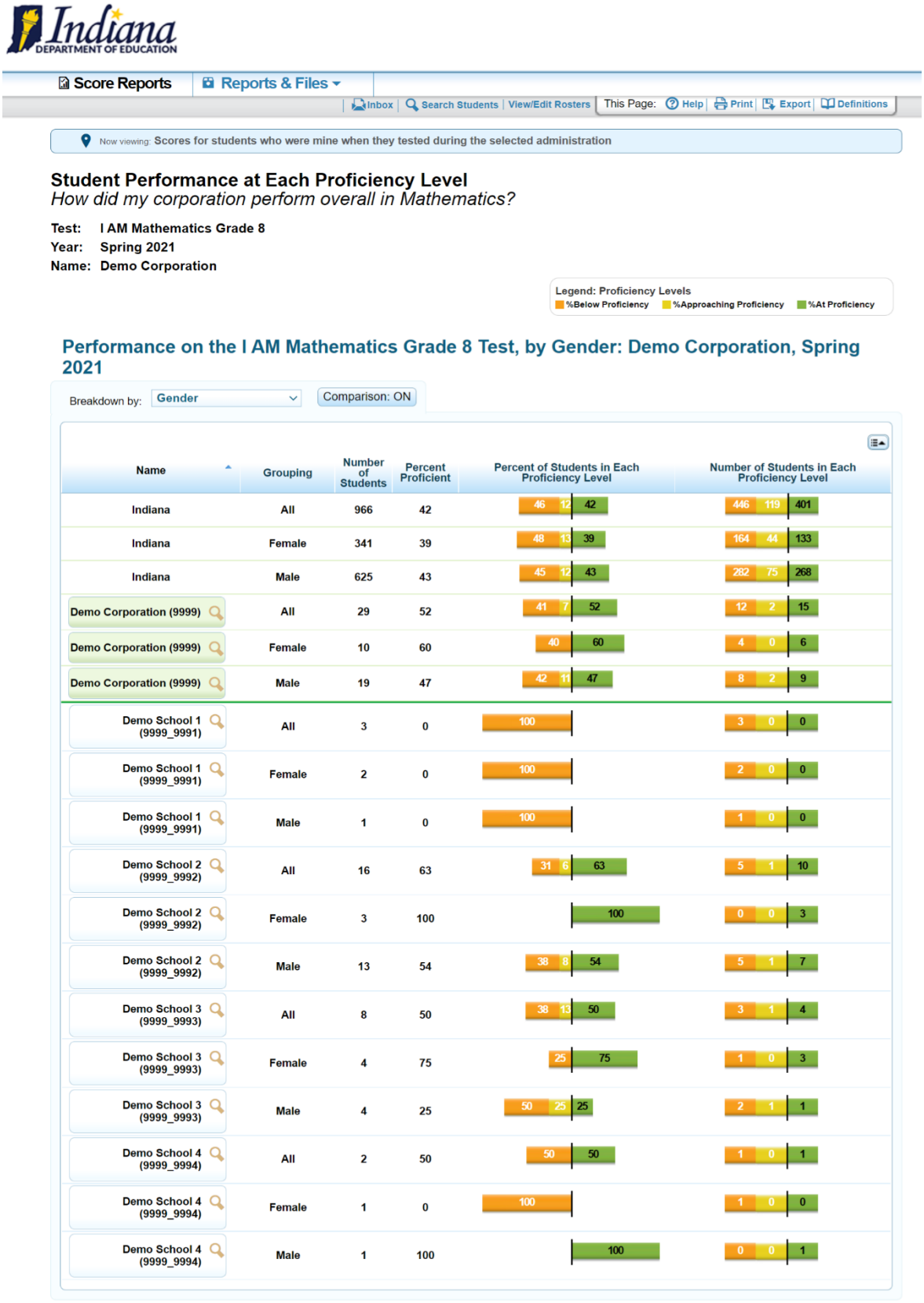


Figure 4: Corporation Aggregate-Level Subject Report, Grade 8 Mathematics by Gender



Name	Grouping	Number of Students	Percent Proficient	Percent of Students in Each Proficiency Level	Number of Students in Each Proficiency Level
Indiana	All	966	42	46 12 42	446 119 401
Indiana	Female	341	39	48 13 39	164 44 133
Indiana	Male	625	43	45 12 43	282 75 268
Demo Corporation (9999)	All	29	52	41 7 52	12 2 15
Demo Corporation (9999)	Female	10	60	40 60	4 0 6
Demo Corporation (9999)	Male	19	47	42 11 47	8 2 9
Demo School 1 (9999_9991)	All	3	0	100	3 0 0
Demo School 1 (9999_9991)	Female	2	0	100	2 0 0
Demo School 1 (9999_9991)	Male	1	0	100	1 0 0
Demo School 2 (9999_9992)	All	16	63	31 6 63	5 1 10
Demo School 2 (9999_9992)	Female	3	100	100	0 0 3
Demo School 2 (9999_9992)	Male	13	54	38 8 54	5 1 7
Demo School 3 (9999_9993)	All	8	50	38 13 50	3 1 4
Demo School 3 (9999_9993)	Female	4	75	25 75	1 0 3
Demo School 3 (9999_9993)	Male	4	25	50 25 25	2 1 1
Demo School 4 (9999_9994)	All	2	50	50 50	1 0 1
Demo School 4 (9999_9994)	Female	1	0	100	1 0 0
Demo School 4 (9999_9994)	Male	1	100	100	0 0 1

Figure 5: Corporation Aggregate-Level Subject Report, Grade 6 Science

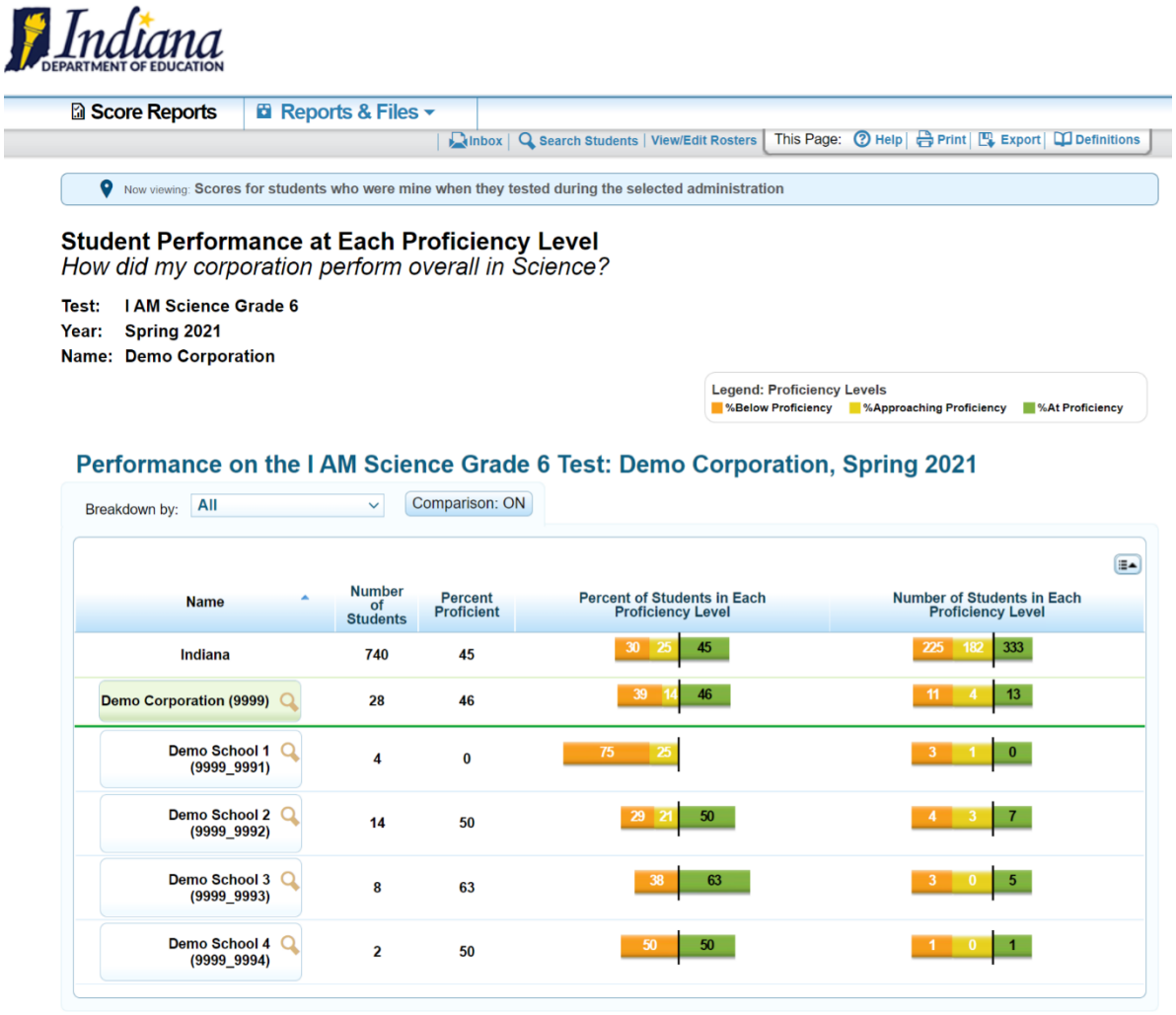
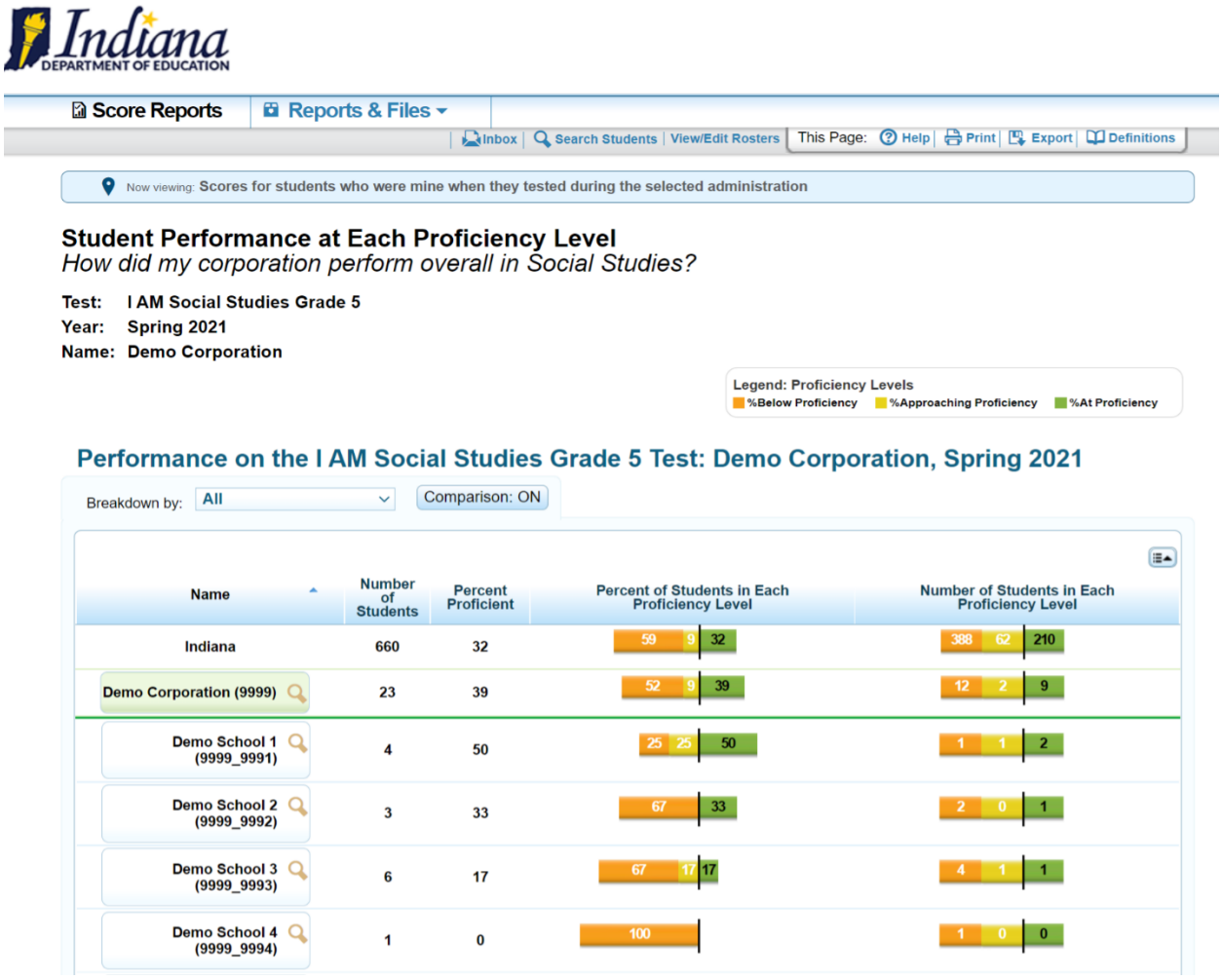


Figure 6: Corporation Aggregate-Level Subject Report, Grade 5 Social Studies



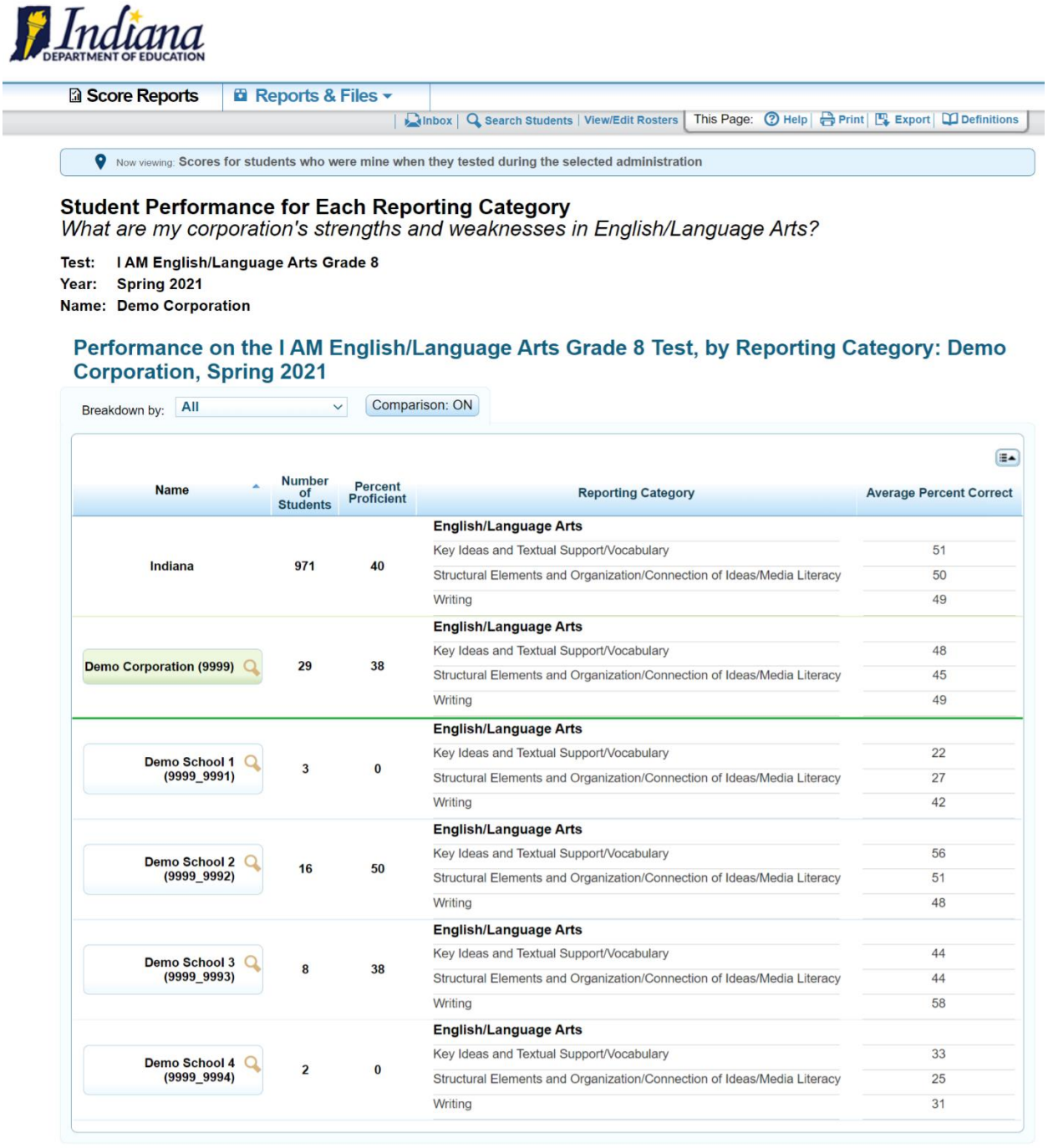
1.6.3 Aggregate-Level Reporting Category Report

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 the

- number of students;
- percent proficient; and
- average percent correct for each reporting category.

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 state and the aggregate unit above the selected aggregate. In addition, summaries can be presented for all students within an aggregate and by students within a defined subgroup. Figure 7 through Figure 10 present examples of the Corporation Aggregate-Level Reporting Category Report for I AM.

Figure 7: Corporation Aggregate-Level Reporting Category Report for Grade 8 ELA



Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: **Scores for students who were mine when they tested during the selected administration**

Student Performance for Each Reporting Category

What are my corporation's strengths and weaknesses in English/Language Arts?

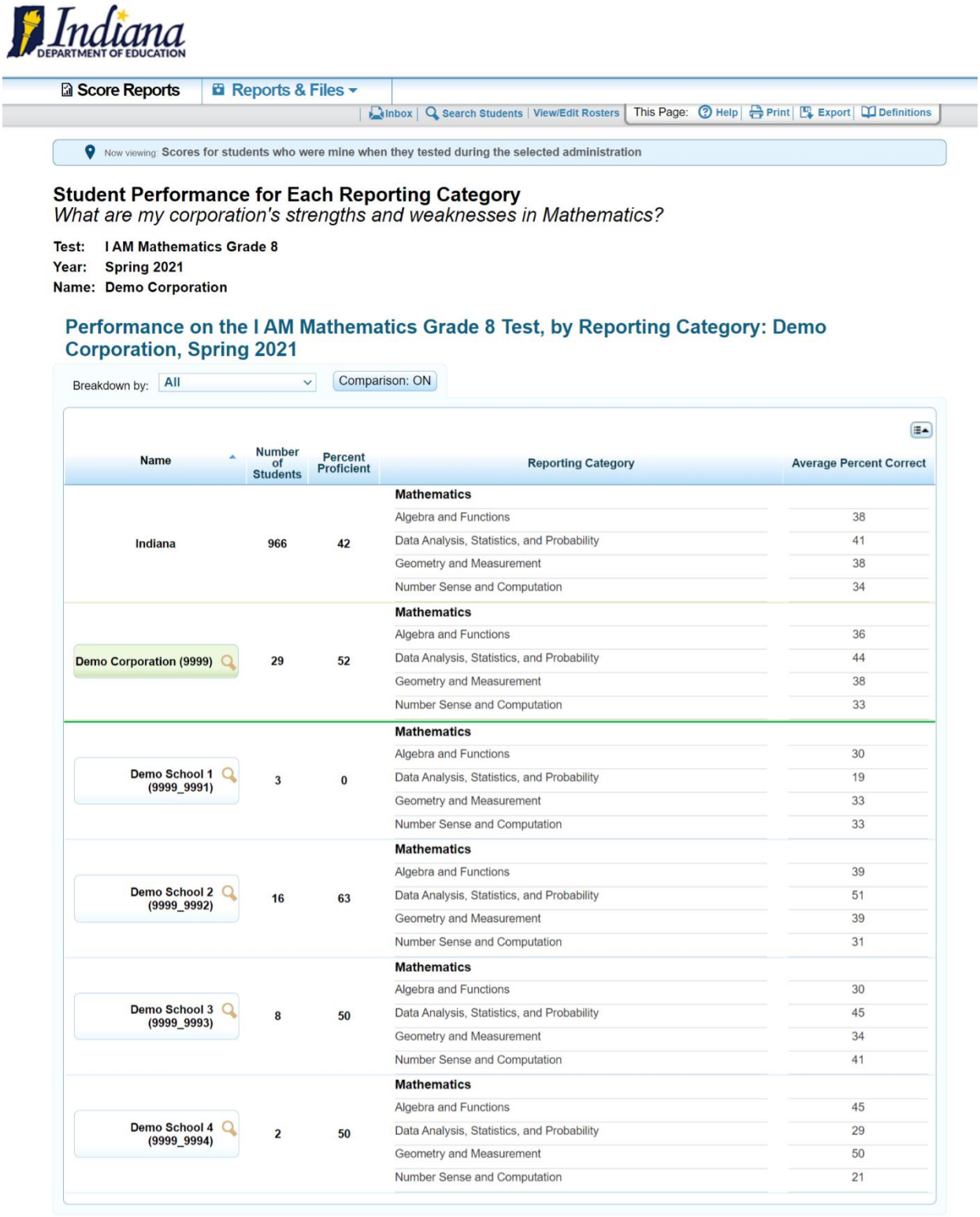
Test: I AM English/Language Arts Grade 8
Year: Spring 2021
Name: Demo Corporation

Performance on the I AM English/Language Arts Grade 8 Test, by Reporting Category: Demo Corporation, Spring 2021

Breakdown by: All | Comparison: ON

Name	Number of Students	Percent Proficient	Reporting Category	Average Percent Correct
Indiana	971	40	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	51
			Structural Elements and Organization/Connection of Ideas/Media Literacy	50
			Writing	49
Demo Corporation (9999)	29	38	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	48
			Structural Elements and Organization/Connection of Ideas/Media Literacy	45
			Writing	49
Demo School 1 (9999_9991)	3	0	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	22
			Structural Elements and Organization/Connection of Ideas/Media Literacy	27
			Writing	42
Demo School 2 (9999_9992)	16	50	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	56
			Structural Elements and Organization/Connection of Ideas/Media Literacy	51
			Writing	48
Demo School 3 (9999_9993)	8	38	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	44
			Structural Elements and Organization/Connection of Ideas/Media Literacy	44
			Writing	58
Demo School 4 (9999_9994)	2	0	**English/Language Arts**	
			Key Ideas and Textual Support/Vocabulary	33
			Structural Elements and Organization/Connection of Ideas/Media Literacy	25
			Writing	31

Figure 8: Corporation Aggregate-Level Reporting Category Report for Grade 8 Mathematics



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance for Each Reporting Category

What are my corporation's strengths and weaknesses in Mathematics?

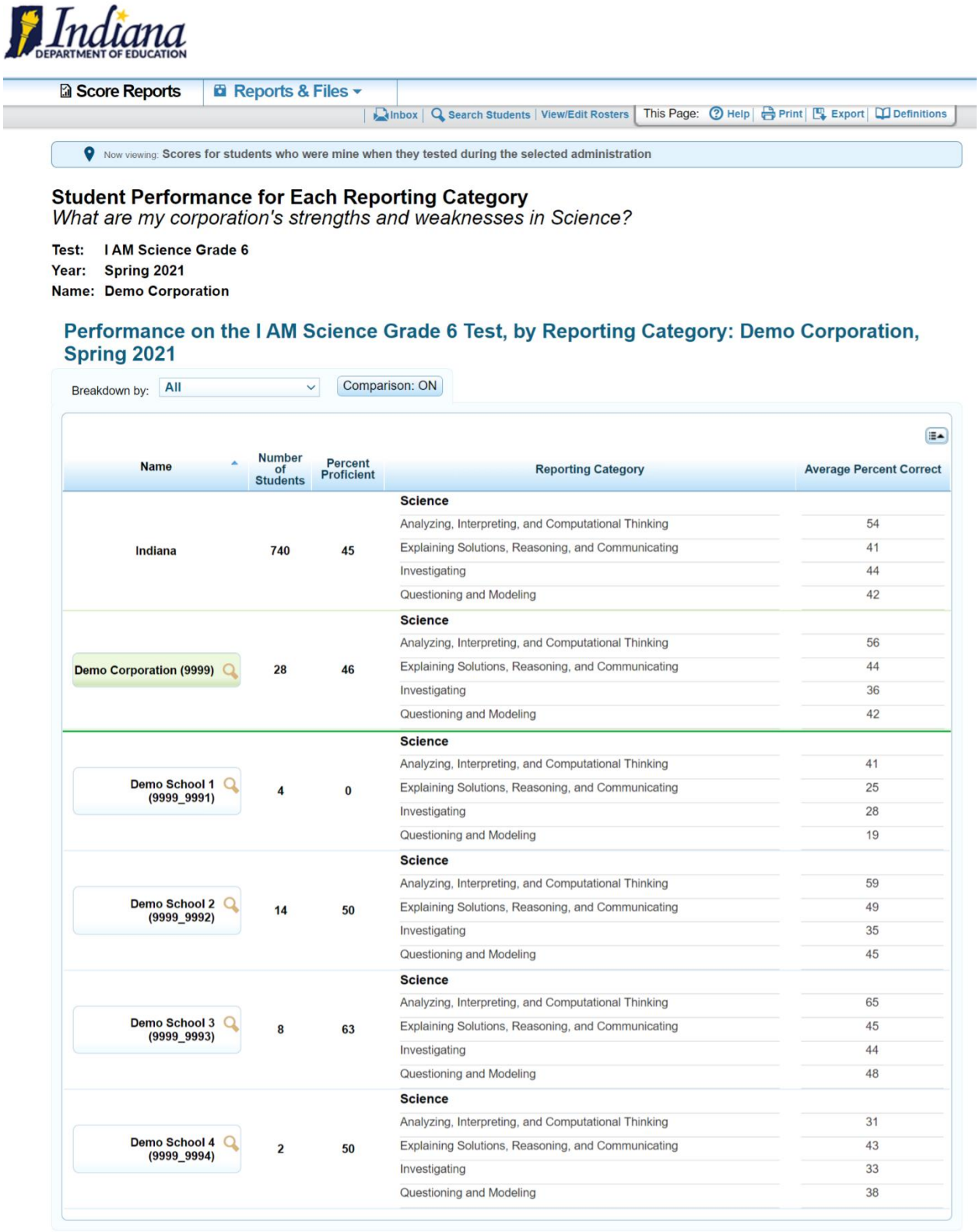
Test: I AM Mathematics Grade 8
Year: Spring 2021
Name: Demo Corporation

Performance on the I AM Mathematics Grade 8 Test, by Reporting Category: Demo Corporation, Spring 2021

Breakdown by: All | Comparison: ON

Name	Number of Students	Percent Proficient	Reporting Category	Average Percent Correct
Indiana	966	42	**Mathematics**	
			Algebra and Functions	38
			Data Analysis, Statistics, and Probability	41
			Geometry and Measurement	38
			Number Sense and Computation	34
Demo Corporation (9999)	29	52	**Mathematics**	
			Algebra and Functions	36
			Data Analysis, Statistics, and Probability	44
			Geometry and Measurement	38
			Number Sense and Computation	33
Demo School 1 (9999_9991)	3	0	**Mathematics**	
			Algebra and Functions	30
			Data Analysis, Statistics, and Probability	19
			Geometry and Measurement	33
			Number Sense and Computation	33
Demo School 2 (9999_9992)	16	63	**Mathematics**	
			Algebra and Functions	39
			Data Analysis, Statistics, and Probability	51
			Geometry and Measurement	39
			Number Sense and Computation	31
Demo School 3 (9999_9993)	8	50	**Mathematics**	
			Algebra and Functions	30
			Data Analysis, Statistics, and Probability	45
			Geometry and Measurement	34
			Number Sense and Computation	41
Demo School 4 (9999_9994)	2	50	**Mathematics**	
			Algebra and Functions	45
			Data Analysis, Statistics, and Probability	29
			Geometry and Measurement	50
			Number Sense and Computation	21

Figure 9: Corporation Aggregate-Level Reporting Category Report for Grade 6 Science



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance for Each Reporting Category

What are my corporation's strengths and weaknesses in Science?

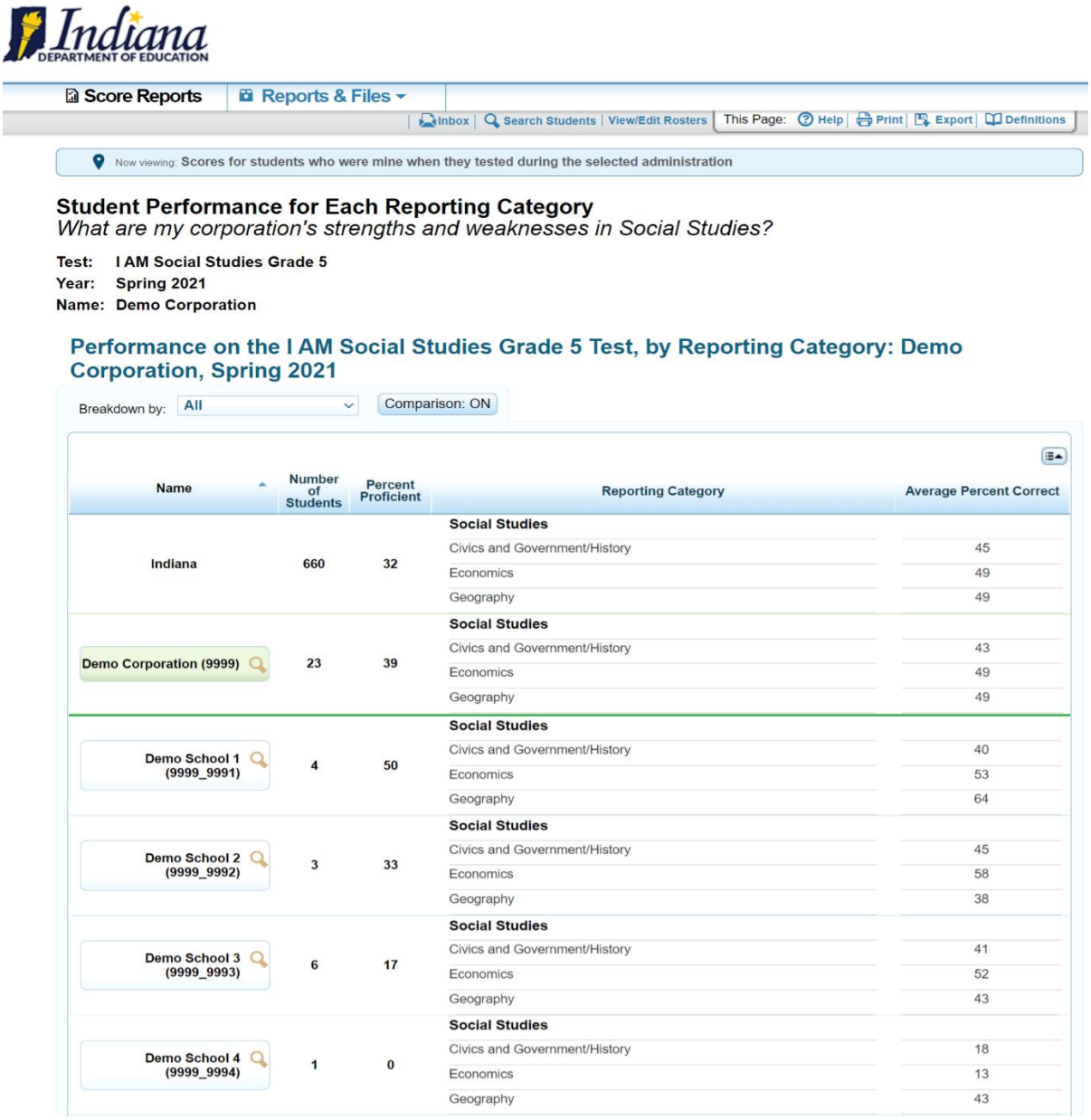
Test: I AM Science Grade 6
Year: Spring 2021
Name: Demo Corporation

Performance on the I AM Science Grade 6 Test, by Reporting Category: Demo Corporation, Spring 2021

Breakdown by: All | Comparison: ON

Name	Number of Students	Percent Proficient	Reporting Category	Average Percent Correct
Indiana	740	45	**Science**	
			Analyzing, Interpreting, and Computational Thinking	54
			Explaining Solutions, Reasoning, and Communicating	41
			Investigating	44
			Questioning and Modeling	42
Demo Corporation (9999)	28	46	**Science**	
			Analyzing, Interpreting, and Computational Thinking	56
			Explaining Solutions, Reasoning, and Communicating	44
			Investigating	36
			Questioning and Modeling	42
Demo School 1 (9999_9991)	4	0	**Science**	
			Analyzing, Interpreting, and Computational Thinking	41
			Explaining Solutions, Reasoning, and Communicating	25
			Investigating	28
			Questioning and Modeling	19
Demo School 2 (9999_9992)	14	50	**Science**	
			Analyzing, Interpreting, and Computational Thinking	59
			Explaining Solutions, Reasoning, and Communicating	49
			Investigating	35
			Questioning and Modeling	45
Demo School 3 (9999_9993)	8	63	**Science**	
			Analyzing, Interpreting, and Computational Thinking	65
			Explaining Solutions, Reasoning, and Communicating	45
			Investigating	44
			Questioning and Modeling	48
Demo School 4 (9999_9994)	2	50	**Science**	
			Analyzing, Interpreting, and Computational Thinking	31
			Explaining Solutions, Reasoning, and Communicating	43
			Investigating	33
			Questioning and Modeling	38

Figure 10: Corporation Aggregate-Level Reporting Category Report for Grade 5 Social Studies



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance for Each Reporting Category

What are my corporation's strengths and weaknesses in Social Studies?

Test: I AM Social Studies Grade 5
Year: Spring 2021
Name: Demo Corporation

Performance on the I AM Social Studies Grade 5 Test, by Reporting Category: Demo Corporation, Spring 2021

Breakdown by: All | Comparison: ON

Name	Number of Students	Percent Proficient	Reporting Category	Average Percent Correct
Indiana	660	32	**Social Studies**	
			Civics and Government/History	45
			Economics	49
			Geography	49
Demo Corporation (9999)	23	39	**Social Studies**	
			Civics and Government/History	43
			Economics	49
			Geography	49
Demo School 1 (9999_9991)	4	50	**Social Studies**	
			Civics and Government/History	40
			Economics	53
			Geography	64
Demo School 2 (9999_9992)	3	33	**Social Studies**	
			Civics and Government/History	45
			Economics	58
			Geography	38
Demo School 3 (9999_9993)	6	17	**Social Studies**	
			Civics and Government/History	41
			Economics	52
			Geography	43
Demo School 4 (9999_9994)	1	0	**Social Studies**	
			Civics and Government/History	18
			Economics	13
			Geography	43

1.6.4 Student-Level Subject Report

The Student-Level Subject Report lists all students who belong to the selected aggregate level, such as a school, and reports the following measures for each student:

- Overall subject scale score
- Overall subject performance level

Figure 11 through Figure 14 demonstrate examples of the Student-Level Subject Report for I AM.

Figure 11: Student-Level Subject Report for Grade 8 ELA

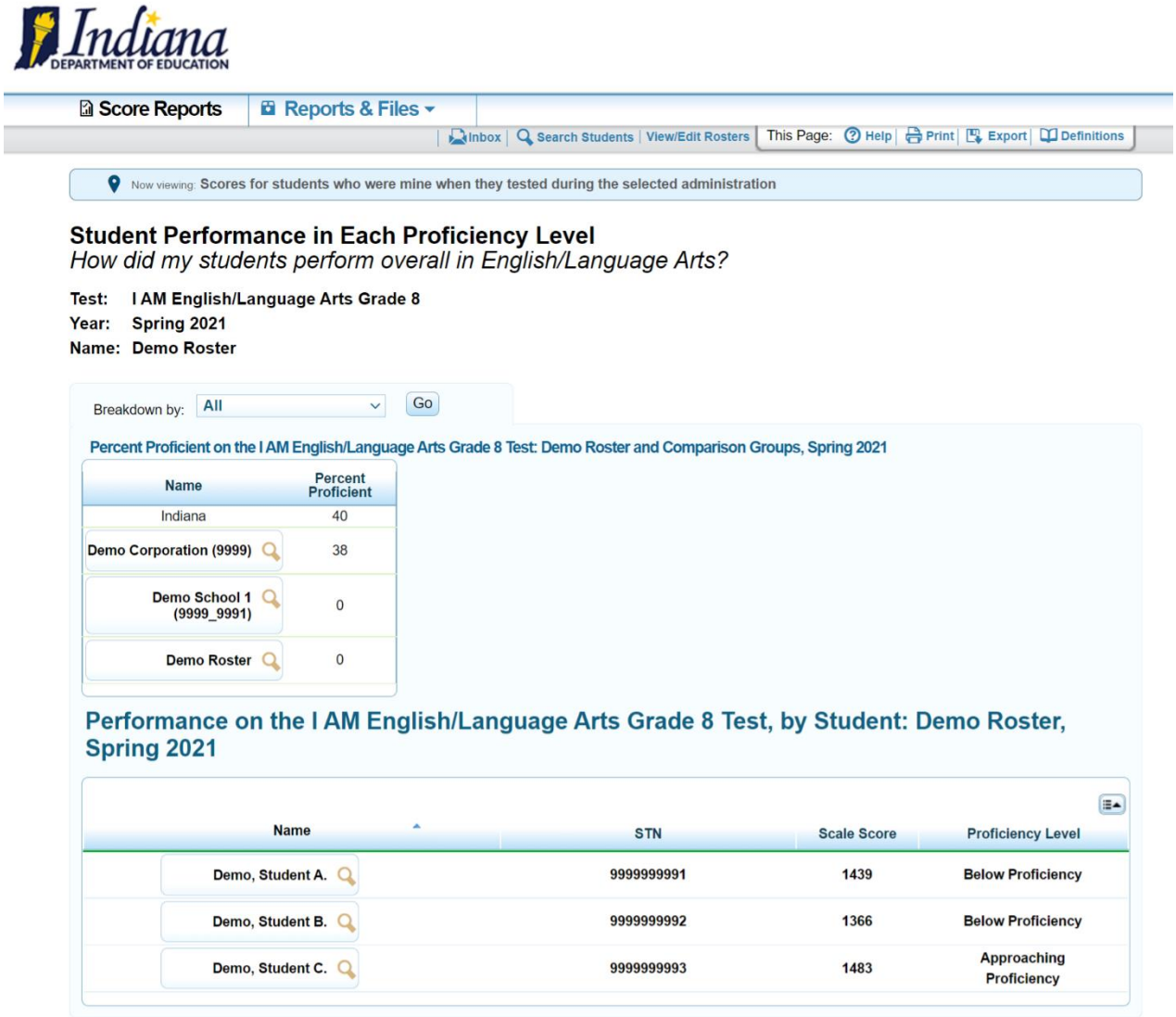


Figure 12: Student-Level Subject Report for Grade 8 Mathematics

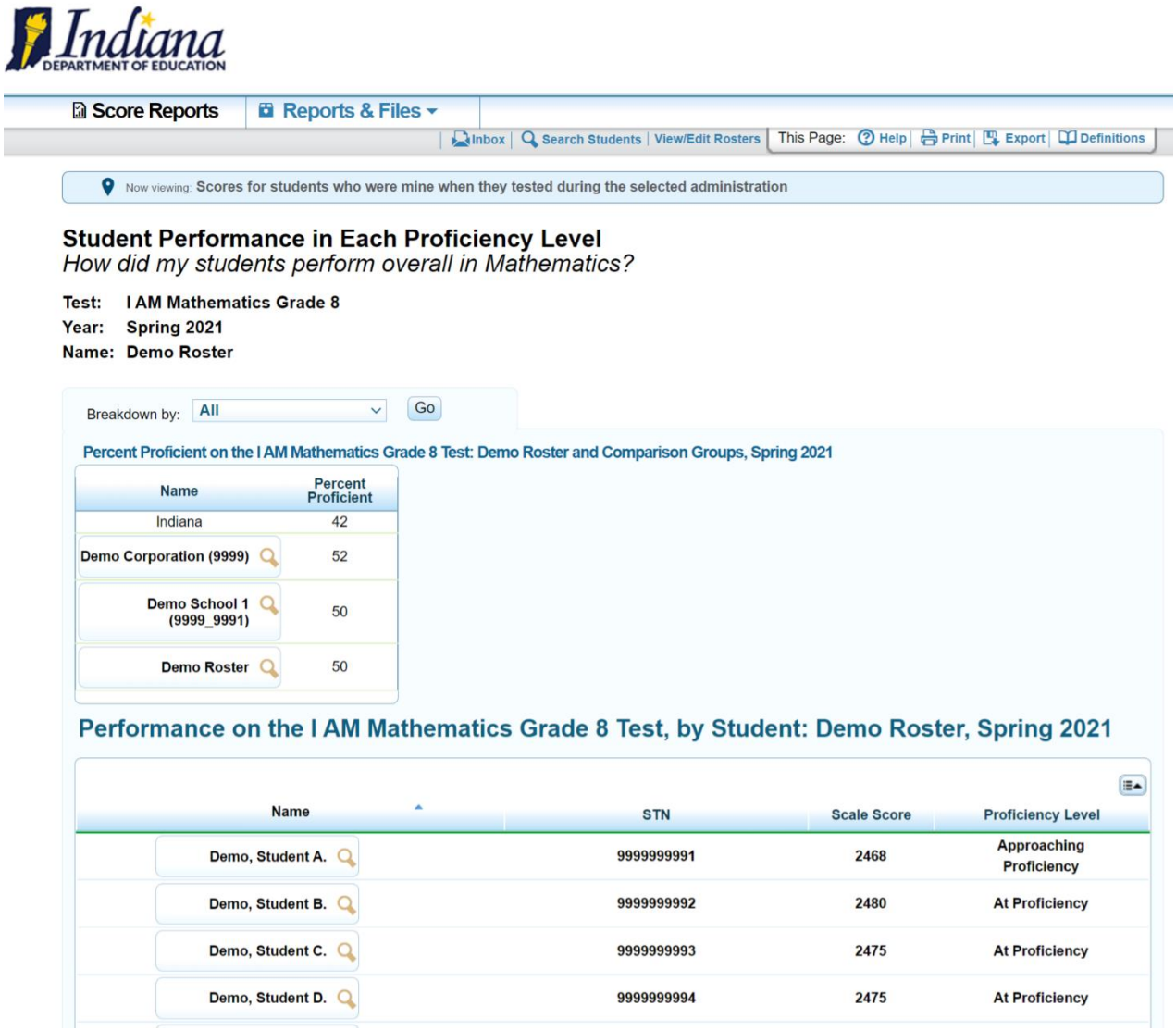


Figure 13: Student-Level Subject Report for Grade 6 Science

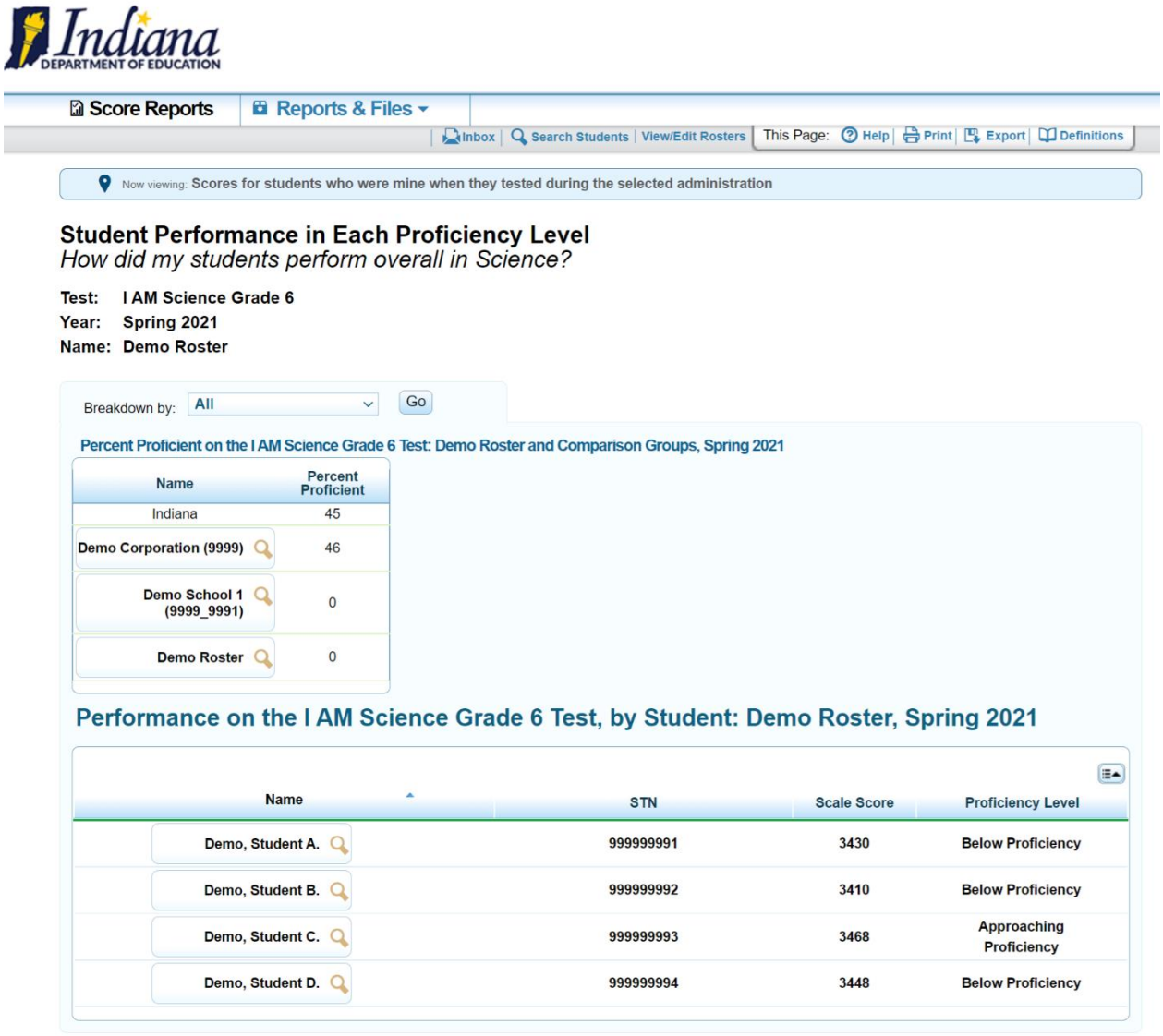
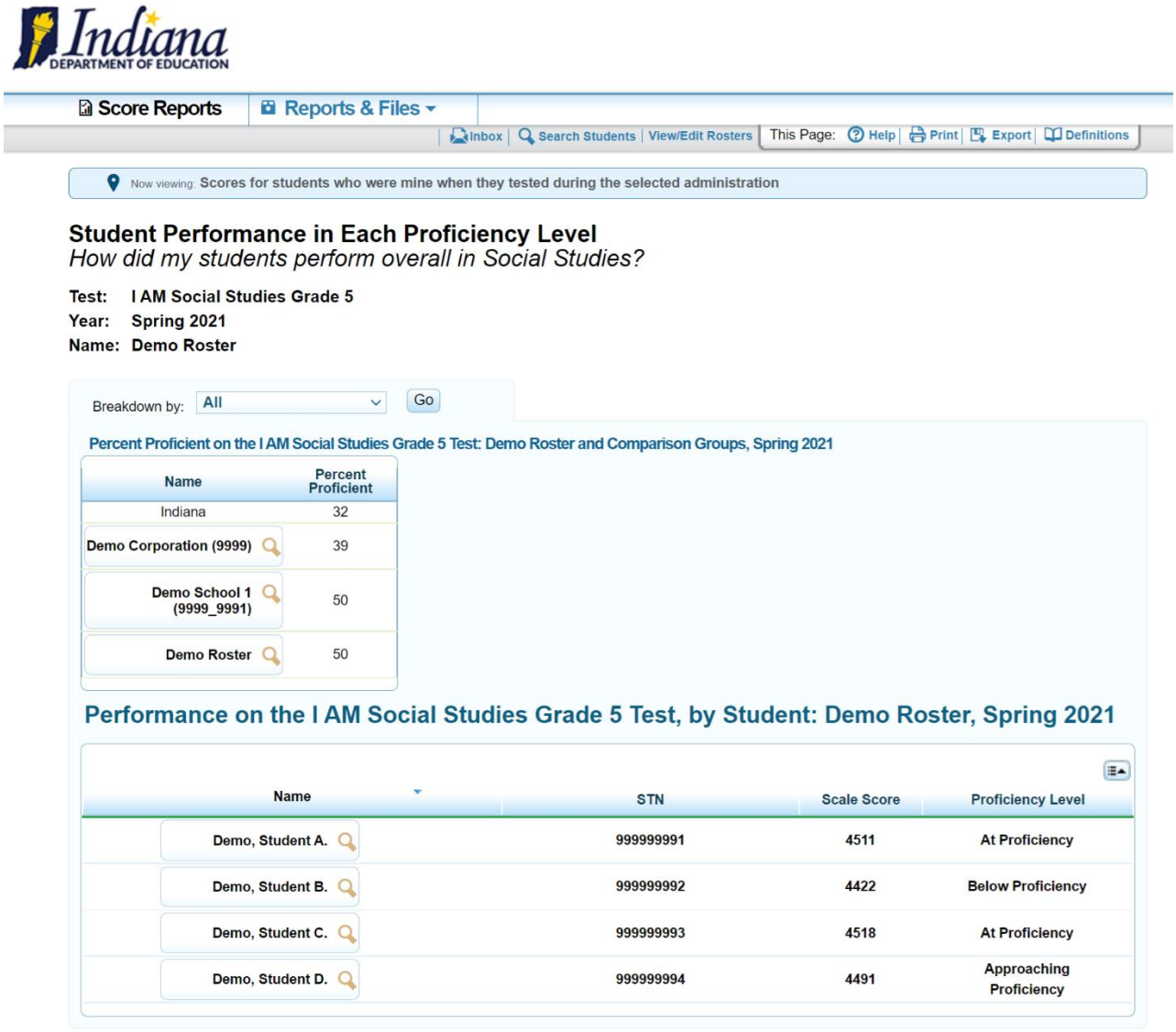


Figure 14: Student-Level Subject Report for Grade 5 Social Studies



1.6.5 Student-Level Reporting Category Report

The Student-Level Reporting Category Report lists all students who belong to the selected aggregate level, such as a school, and reports the following measures for each student:

- Overall subject scale score
- Overall subject performance level
- Reporting category percent correct

Figure 15 through Figure 18 displays this information for I AM.

Figure 15: Student-Level Reporting Category Report for Grade 8 ELA

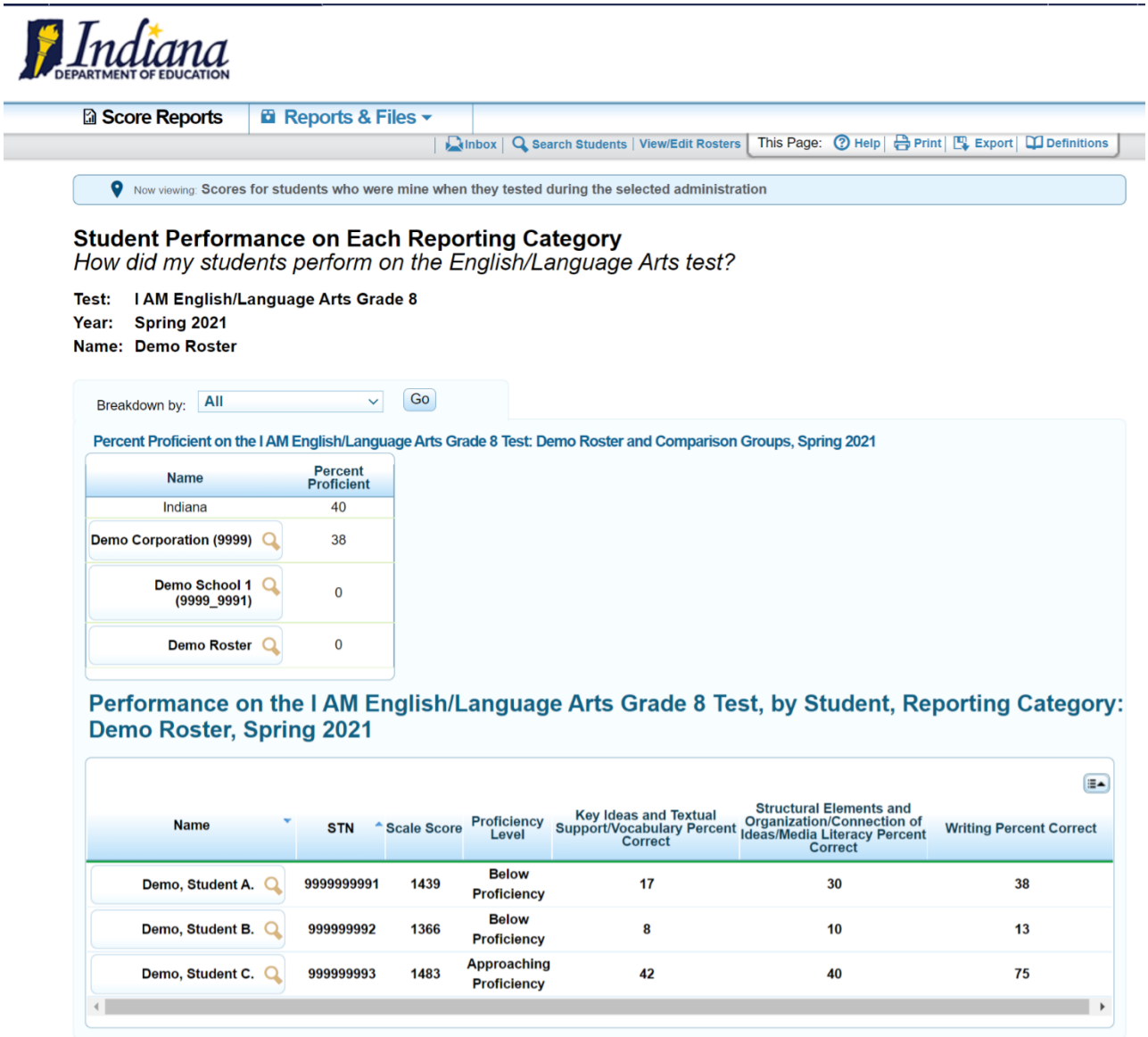
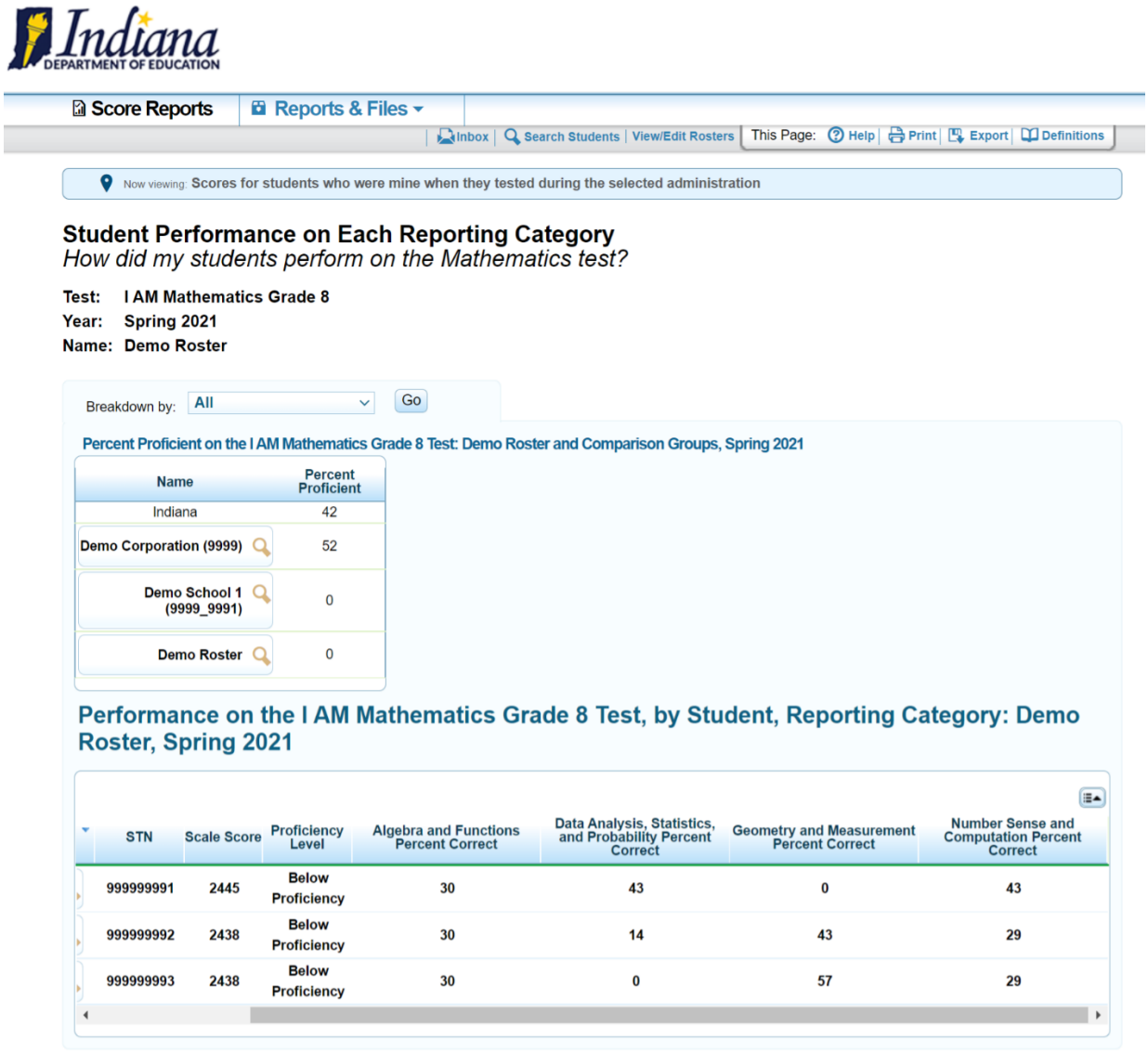


Figure 16: Student-Level Reporting Category Report for Grade 8 Mathematics



Indiana DEPARTMENT OF EDUCATION

Score Reports | Reports & Files | Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance on Each Reporting Category

How did my students perform on the Mathematics test?

Test: I AM Mathematics Grade 8
Year: Spring 2021
Name: Demo Roster

Breakdown by: All Go

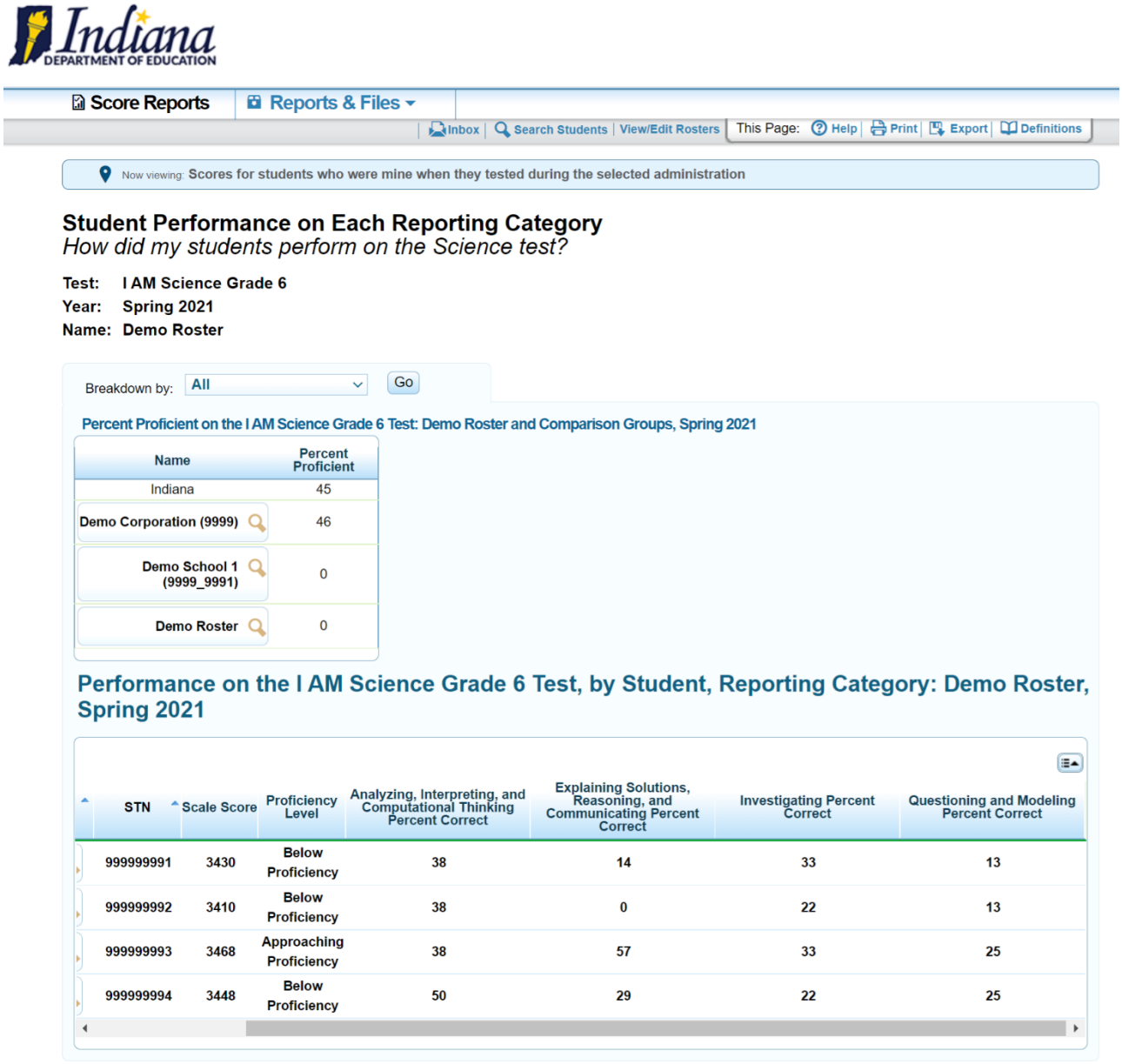
Percent Proficient on the I AM Mathematics Grade 8 Test: Demo Roster and Comparison Groups, Spring 2021

Name	Percent Proficient
Indiana	42
Demo Corporation (9999)	52
Demo School 1 (9999_9991)	0
Demo Roster	0

Performance on the I AM Mathematics Grade 8 Test, by Student, Reporting Category: Demo Roster, Spring 2021

STN	Scale Score	Proficiency Level	Algebra and Functions Percent Correct	Data Analysis, Statistics, and Probability Percent Correct	Geometry and Measurement Percent Correct	Number Sense and Computation Percent Correct
999999991	2445	Below Proficiency	30	43	0	43
999999992	2438	Below Proficiency	30	14	43	29
999999993	2438	Below Proficiency	30	0	57	29

Figure 17: Student-Level Reporting Category Report for Grade 6 Science



Indiana DEPARTMENT OF EDUCATION

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance on Each Reporting Category

How did my students perform on the Science test?

Test: I AM Science Grade 6
Year: Spring 2021
Name: Demo Roster

Breakdown by: All Go

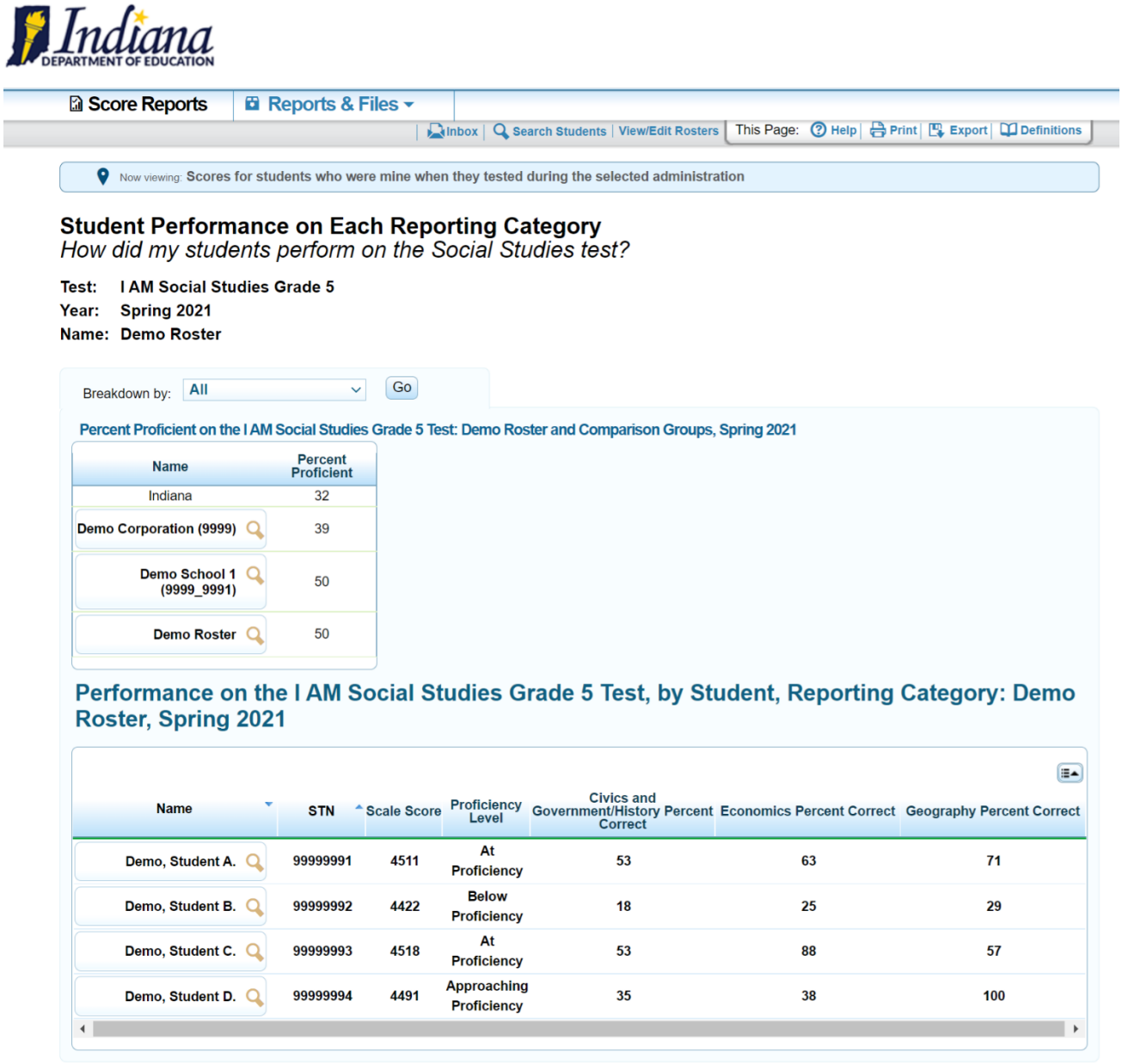
Percent Proficient on the I AM Science Grade 6 Test: Demo Roster and Comparison Groups, Spring 2021

Name	Percent Proficient
Indiana	45
Demo Corporation (9999)	46
Demo School 1 (9999_9991)	0
Demo Roster	0

Performance on the I AM Science Grade 6 Test, by Student, Reporting Category: Demo Roster, Spring 2021

STN	Scale Score	Proficiency Level	Analyzing, Interpreting, and Computational Thinking Percent Correct	Explaining Solutions, Reasoning, and Communicating Percent Correct	Investigating Percent Correct	Questioning and Modeling Percent Correct
999999991	3430	Below Proficiency	38	14	33	13
999999992	3410	Below Proficiency	38	0	22	13
999999993	3468	Approaching Proficiency	38	57	33	25
999999994	3448	Below Proficiency	50	29	22	25

Figure 18: Student-Level Reporting Category Report for Grade 5 Social Studies



Indiana DEPARTMENT OF EDUCATION

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance on Each Reporting Category

How did my students perform on the Social Studies test?

Test: I AM Social Studies Grade 5
Year: Spring 2021
Name: Demo Roster

Breakdown by: All Go

Percent Proficient on the I AM Social Studies Grade 5 Test: Demo Roster and Comparison Groups, Spring 2021

Name	Percent Proficient
Indiana	32
Demo Corporation (9999)	39
Demo School 1 (9999_9991)	50
Demo Roster	50

Performance on the I AM Social Studies Grade 5 Test, by Student, Reporting Category: Demo Roster, Spring 2021

Name	STN	Scale Score	Proficiency Level	Civics and Government/History Percent Correct	Economics Percent Correct	Geography Percent Correct
Demo, Student A.	99999991	4511	At Proficiency	53	63	71
Demo, Student B.	99999992	4422	Below Proficiency	18	25	29
Demo, Student C.	99999993	4518	At Proficiency	53	88	57
Demo, Student D.	99999994	4491	Approaching Proficiency	35	38	100

1.6.6 Individual Student Report

When a student receives a valid test score, an individual student report (ISR) can be generated in the ORS. The ISR contains the following measures for that student:

- Overall subject scale score
- Overall subject performance level
- Percent proficiency for a student's state, corporation, and school
- Percent correct in each reporting category

The top section of the report includes the following key student information:

- Name
- Scale score
- Performance level

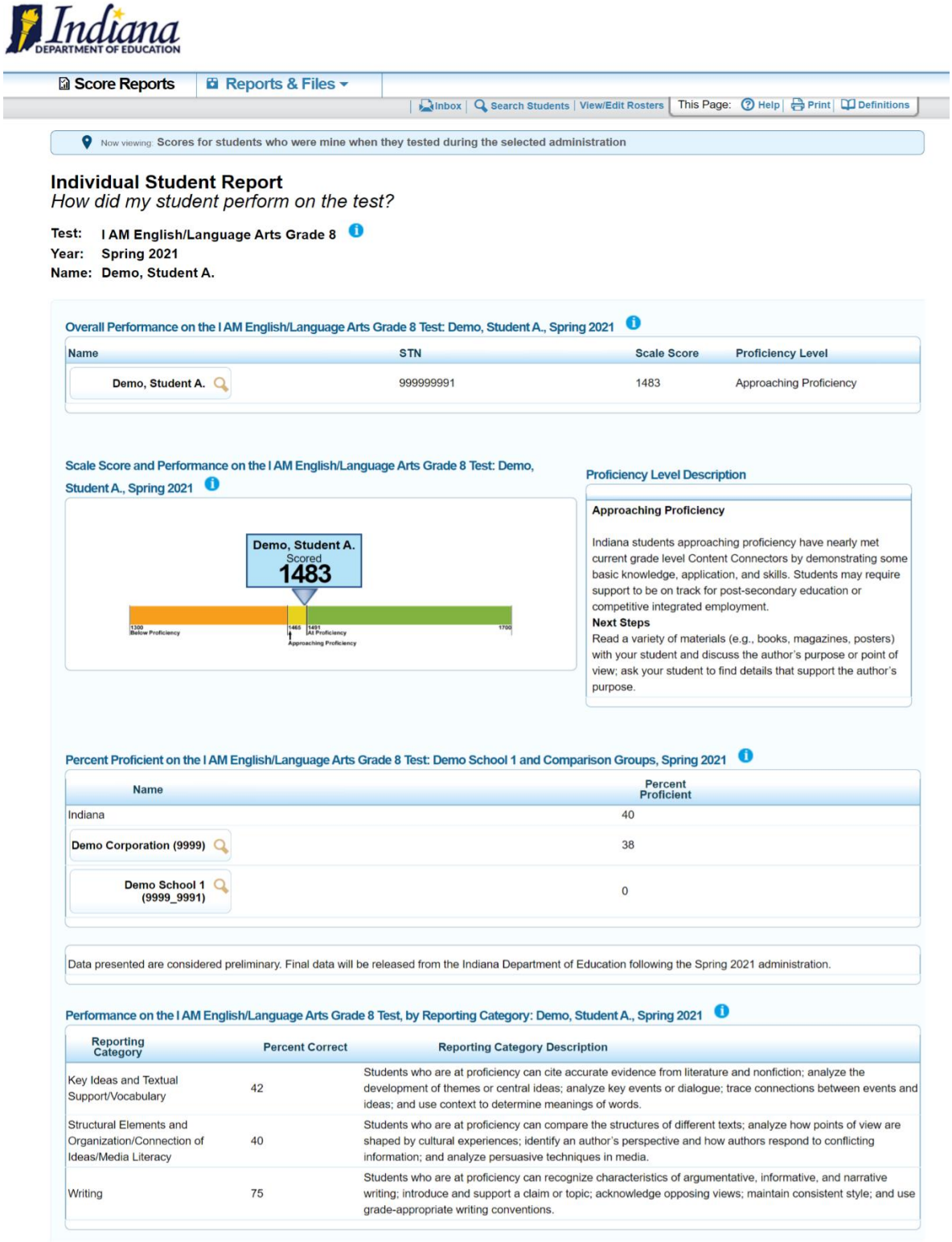
The middle section includes the following data:

- Bar chart display of the student's scale score
- PLDs with cut scores at each performance level
- Average aggregated scale scores at the state, corporation, and school levels

The bottom section of the report contains detailed information on student performance in each reporting category.

Figure 19 through Figure 22 present examples of ISRs for an I AM assessment.

Figure 19: Individual Student Report for Grade 8 ELA



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Individual Student Report

How did my student perform on the test?

Test: I AM English/Language Arts Grade 8
Year: Spring 2021
Name: Demo, Student A.

Overall Performance on the I AM English/Language Arts Grade 8 Test: Demo, Student A., Spring 2021

Name	STN	Scale Score	Proficiency Level
Demo, Student A.	999999991	1483	Approaching Proficiency

Scale Score and Performance on the I AM English/Language Arts Grade 8 Test: Demo, Student A., Spring 2021

Proficiency Level Description

Approaching Proficiency

Indiana students approaching proficiency have nearly met current grade level Content Connectors by demonstrating some basic knowledge, application, and skills. Students may require support to be on track for post-secondary education or competitive integrated employment.

Next Steps

Read a variety of materials (e.g., books, magazines, posters) with your student and discuss the author's purpose or point of view; ask your student to find details that support the author's purpose.

Percent Proficient on the I AM English/Language Arts Grade 8 Test: Demo School 1 and Comparison Groups, Spring 2021

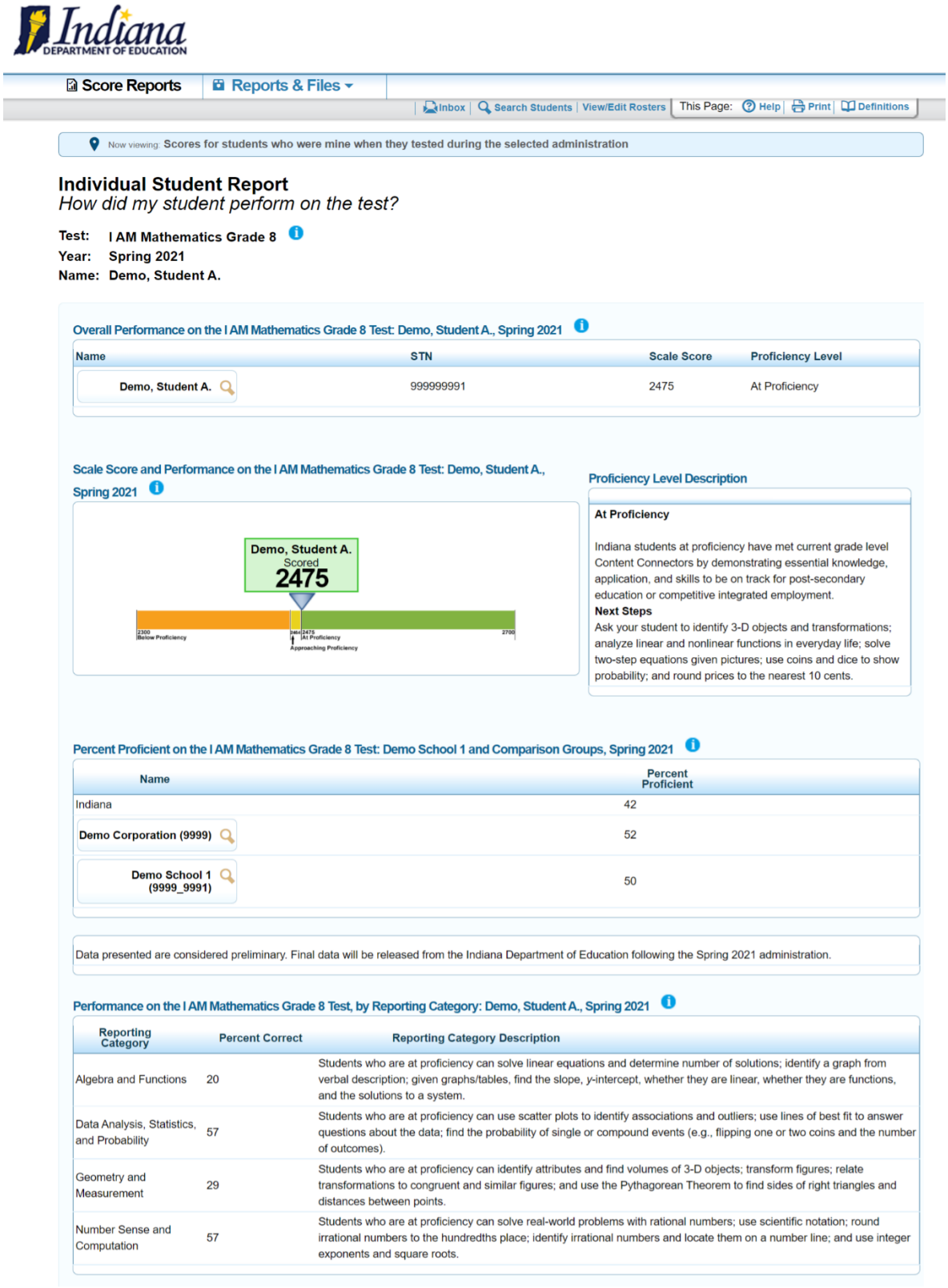
Name	Percent Proficient
Indiana	40
Demo Corporation (9999)	38
Demo School 1 (9999_9991)	0

Data presented are considered preliminary. Final data will be released from the Indiana Department of Education following the Spring 2021 administration.

Performance on the I AM English/Language Arts Grade 8 Test, by Reporting Category: Demo, Student A., Spring 2021

Reporting Category	Percent Correct	Reporting Category Description
Key Ideas and Textual Support/Vocabulary	42	Students who are at proficiency can cite accurate evidence from literature and nonfiction; analyze the development of themes or central ideas; analyze key events or dialogue; trace connections between events and ideas; and use context to determine meanings of words.
Structural Elements and Organization/Connection of Ideas/Media Literacy	40	Students who are at proficiency can compare the structures of different texts; analyze how points of view are shaped by cultural experiences; identify an author's perspective and how authors respond to conflicting information; and analyze persuasive techniques in media.
Writing	75	Students who are at proficiency can recognize characteristics of argumentative, informative, and narrative writing; introduce and support a claim or topic; acknowledge opposing views; maintain consistent style; and use grade-appropriate writing conventions.

Figure 20: Individual Student Report for Grade 8 Mathematics



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Individual Student Report

How did my student perform on the test?

Test: I AM Mathematics Grade 8
Year: Spring 2021
Name: Demo, Student A.

Overall Performance on the I AM Mathematics Grade 8 Test: Demo, Student A., Spring 2021

Name	STN	Scale Score	Proficiency Level
Demo, Student A.	999999991	2475	At Proficiency

Scale Score and Performance on the I AM Mathematics Grade 8 Test: Demo, Student A., Spring 2021

Proficiency Level Description

At Proficiency

Indiana students at proficiency have met current grade level Content Connectors by demonstrating essential knowledge, application, and skills to be on track for post-secondary education or competitive integrated employment.

Next Steps

Ask your student to identify 3-D objects and transformations; analyze linear and nonlinear functions in everyday life; solve two-step equations given pictures; use coins and dice to show probability; and round prices to the nearest 10 cents.

Percent Proficient on the I AM Mathematics Grade 8 Test: Demo School 1 and Comparison Groups, Spring 2021

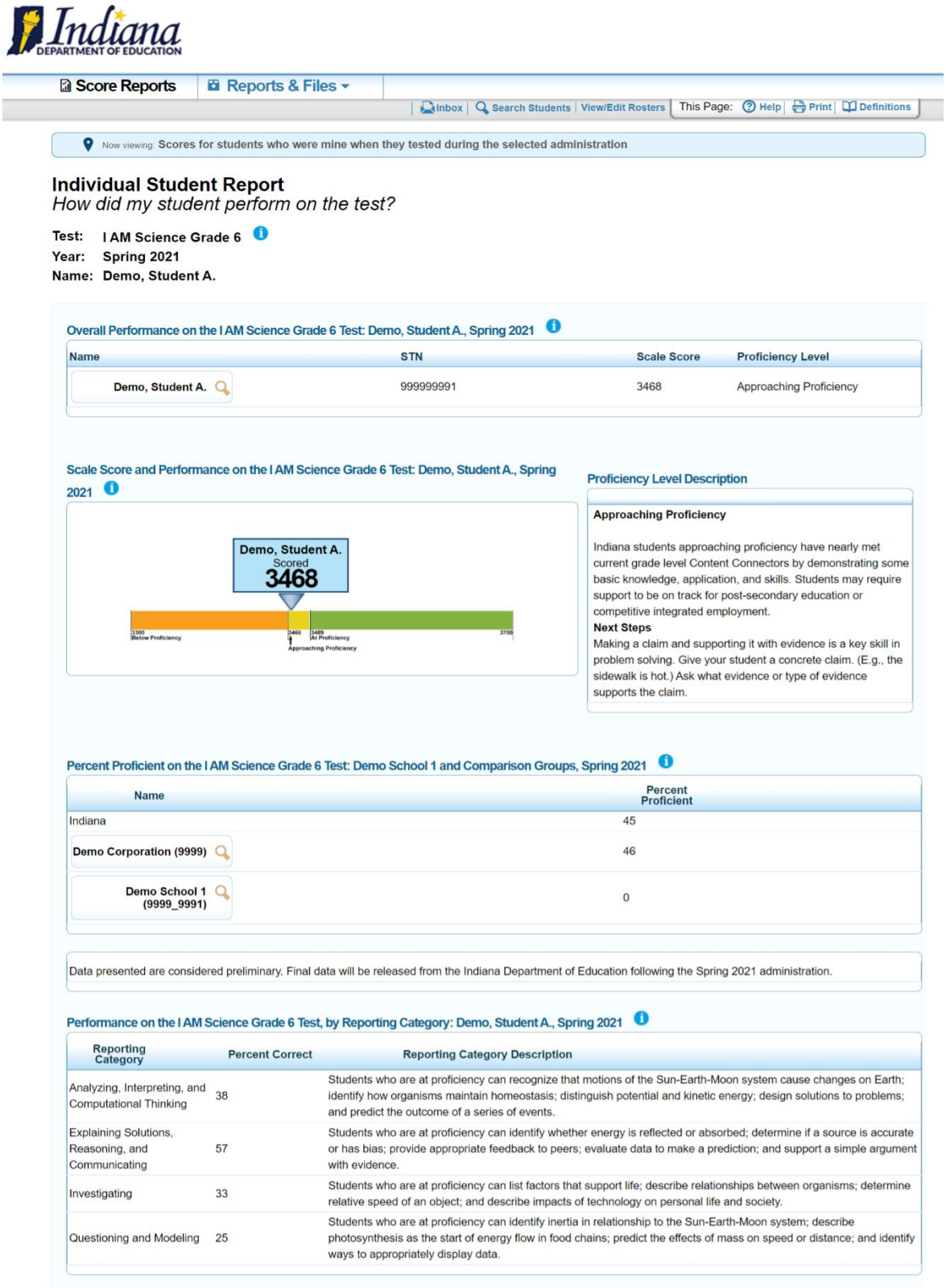
Name	Percent Proficient
Indiana	42
Demo Corporation (9999)	52
Demo School 1 (9999_9991)	50

Data presented are considered preliminary. Final data will be released from the Indiana Department of Education following the Spring 2021 administration.

Performance on the I AM Mathematics Grade 8 Test, by Reporting Category: Demo, Student A., Spring 2021

Reporting Category	Percent Correct	Reporting Category Description
Algebra and Functions	20	Students who are at proficiency can solve linear equations and determine number of solutions; identify a graph from verbal description; given graphs/tables, find the slope, *y*-intercept, whether they are linear, whether they are functions, and the solutions to a system.
Data Analysis, Statistics, and Probability	57	Students who are at proficiency can use scatter plots to identify associations and outliers; use lines of best fit to answer questions about the data; find the probability of single or compound events (e.g., flipping one or two coins and the number of outcomes).
Geometry and Measurement	29	Students who are at proficiency can identify attributes and find volumes of 3-D objects; transform figures; relate transformations to congruent and similar figures; and use the Pythagorean Theorem to find sides of right triangles and distances between points.
Number Sense and Computation	57	Students who are at proficiency can solve real-world problems with rational numbers; use scientific notation; round irrational numbers to the hundredths place; identify irrational numbers and locate them on a number line; and use integer exponents and square roots.

Figure 21: Individual Student Report for Grade 6 Science



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Individual Student Report

How did my student perform on the test?

Test: I AM Science Grade 6
Year: Spring 2021
Name: Demo, Student A.

Overall Performance on the I AM Science Grade 6 Test: Demo, Student A., Spring 2021

Name	STN	Scale Score	Proficiency Level
Demo, Student A.	999999991	3468	Approaching Proficiency

Scale Score and Performance on the I AM Science Grade 6 Test: Demo, Student A., Spring 2021

Proficiency Level Description

Approaching Proficiency

Indiana students approaching proficiency have nearly met current grade level Content Connectors by demonstrating some basic knowledge, application, and skills. Students may require support to be on track for post-secondary education or competitive integrated employment.

Next Steps

Making a claim and supporting it with evidence is a key skill in problem solving. Give your student a concrete claim. (E.g., the sidewalk is hot.) Ask what evidence or type of evidence supports the claim.

Percent Proficient on the I AM Science Grade 6 Test: Demo School 1 and Comparison Groups, Spring 2021

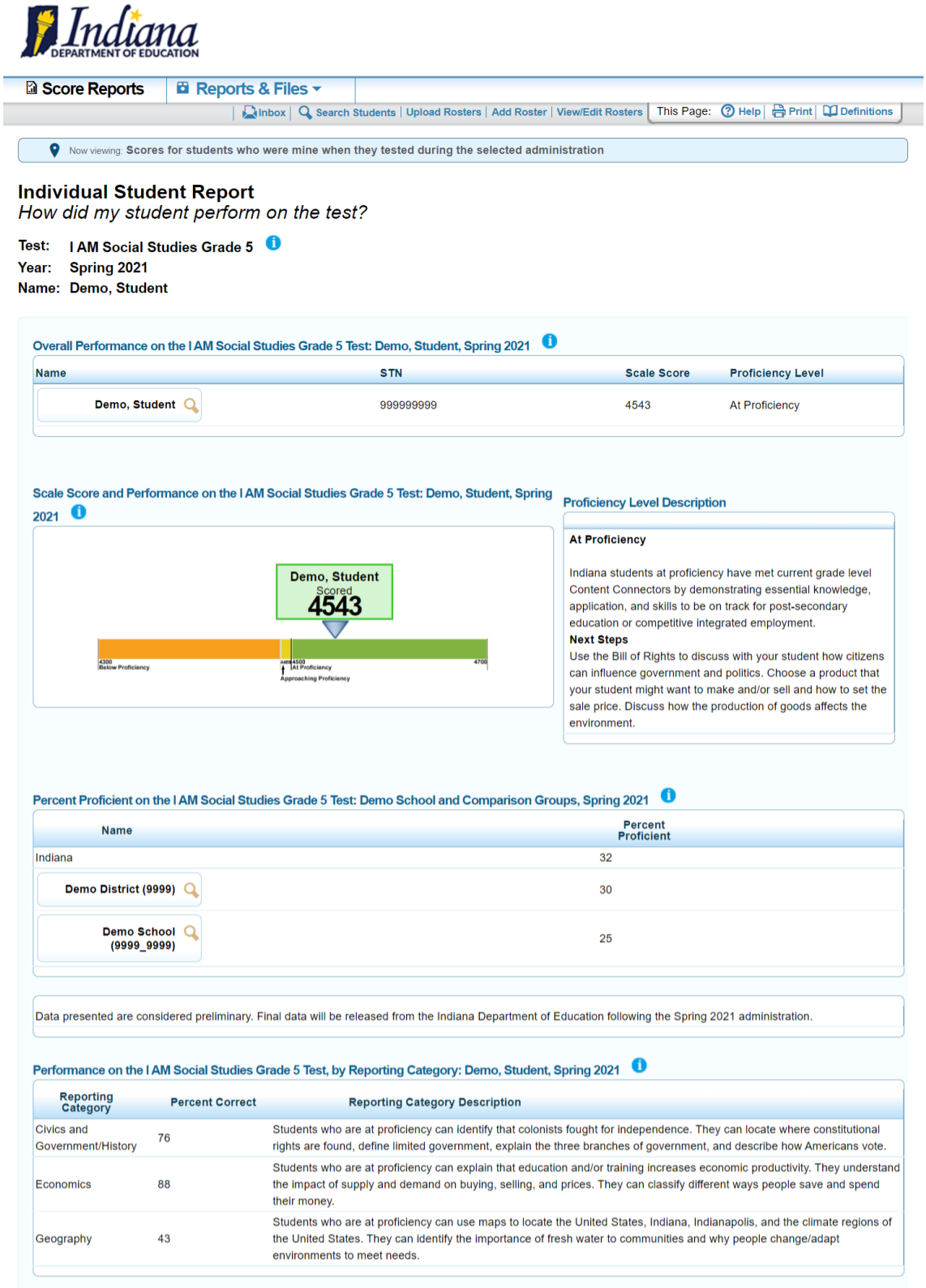
Name	Percent Proficient
Indiana	45
Demo Corporation (9999)	46
Demo School 1 (9999_9991)	0

Data presented are considered preliminary. Final data will be released from the Indiana Department of Education following the Spring 2021 administration.

Performance on the I AM Science Grade 6 Test, by Reporting Category: Demo, Student A., Spring 2021

Reporting Category	Percent Correct	Reporting Category Description
Analyzing, Interpreting, and Computational Thinking	38	Students who are at proficiency can recognize that motions of the Sun-Earth-Moon system cause changes on Earth; identify how organisms maintain homeostasis; distinguish potential and kinetic energy; design solutions to problems; and predict the outcome of a series of events.
Explaining Solutions, Reasoning, and Communicating	57	Students who are at proficiency can identify whether energy is reflected or absorbed; determine if a source is accurate or has bias; provide appropriate feedback to peers; evaluate data to make a prediction; and support a simple argument with evidence.
Investigating	33	Students who are at proficiency can list factors that support life; describe relationships between organisms; determine relative speed of an object; and describe impacts of technology on personal life and society.
Questioning and Modeling	25	Students who are at proficiency can identify inertia in relationship to the Sun-Earth-Moon system; describe photosynthesis as the start of energy flow in food chains; predict the effects of mass on speed or distance; and identify ways to appropriately display data.

Figure 22: Individual Student Report for Grade 5 Social Studies



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | Upload Rosters | Add Roster | View/Edit Rosters | This Page: Help | Print | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Individual Student Report

How did my student perform on the test?

Test: I AM Social Studies Grade 5
Year: Spring 2021
Name: Demo, Student

Overall Performance on the I AM Social Studies Grade 5 Test: Demo, Student, Spring 2021

Name	STN	Scale Score	Proficiency Level
Demo, Student	999999999	4543	At Proficiency

Scale Score and Performance on the I AM Social Studies Grade 5 Test: Demo, Student, Spring 2021

Proficiency Level Description

At Proficiency

Indiana students at proficiency have met current grade level Content Connectors by demonstrating essential knowledge, application, and skills to be on track for post-secondary education or competitive integrated employment.

Next Steps

Use the Bill of Rights to discuss with your student how citizens can influence government and politics. Choose a product that your student might want to make and/or sell and how to set the sale price. Discuss how the production of goods affects the environment.

Percent Proficient on the I AM Social Studies Grade 5 Test: Demo School and Comparison Groups, Spring 2021

Name	Percent Proficient
Indiana	32
Demo District (9999)	30
Demo School (9999_9999)	25

Data presented are considered preliminary. Final data will be released from the Indiana Department of Education following the Spring 2021 administration.

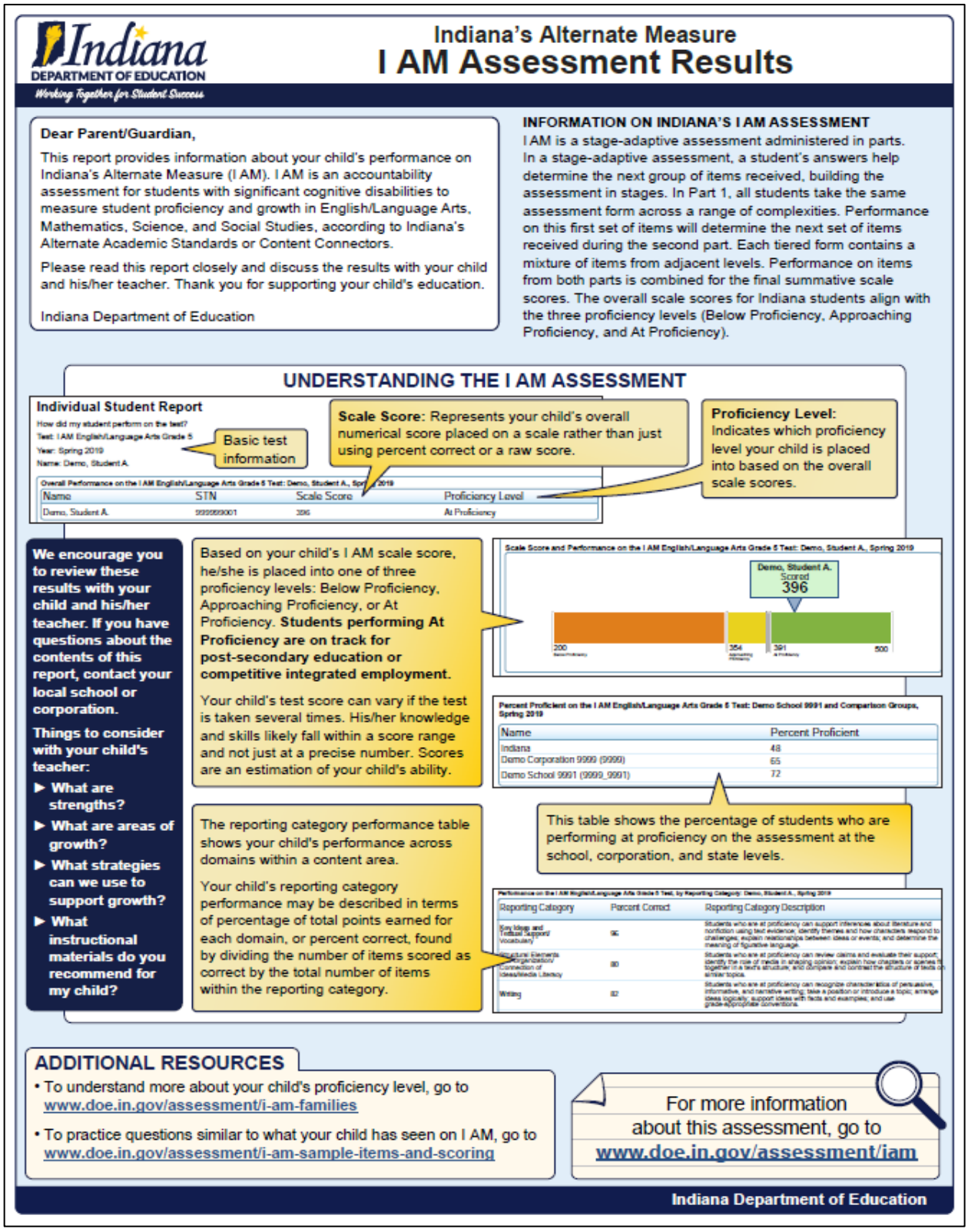
Performance on the I AM Social Studies Grade 5 Test, by Reporting Category: Demo, Student, Spring 2021

Reporting Category	Percent Correct	Reporting Category Description
Civics and Government/History	76	Students who are at proficiency can identify that colonists fought for independence. They can locate where constitutional rights are found, define limited government, explain the three branches of government, and describe how Americans vote.
Economics	88	Students who are at proficiency can explain that education and/or training increases economic productivity. They understand the impact of supply and demand on buying, selling, and prices. They can classify different ways people save and spend their money.
Geography	43	Students who are at proficiency can use maps to locate the United States, Indiana, Indianapolis, and the climate regions of the United States. They can identify the importance of fresh water to communities and why people change/adapt environments to meet needs.

1.6.7 Interpretive Guide

When printing ISRs, users have the option to print a supplemental interpretive guide (called an "Addendum" when printing a Simple ISR) intended as a stand-alone document (see Figure 23) to help teachers, administrators, parents, and students better understand the data presented in the ISR.

Figure 23: Supplemental Interpretive Guide



1.6.8 Reports by Subgroup

At the aggregate level, student performance can be broken down by demographic subgroup, such as gender (Figure 24) or English learner (EL) status (Figure 25).

Figure 24: Corporation Aggregate-Level Subject Report by Gender for Grade 8 ELA

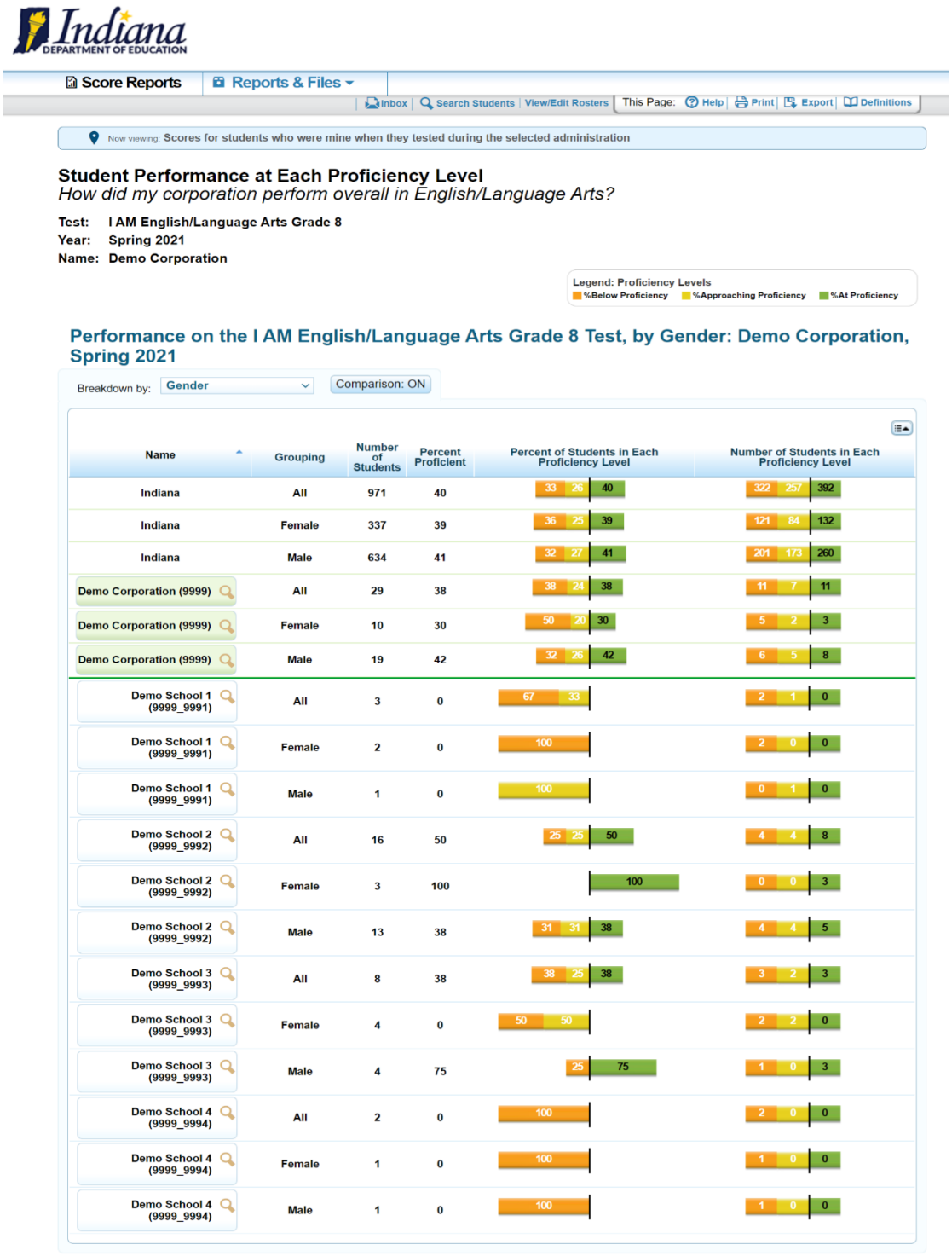
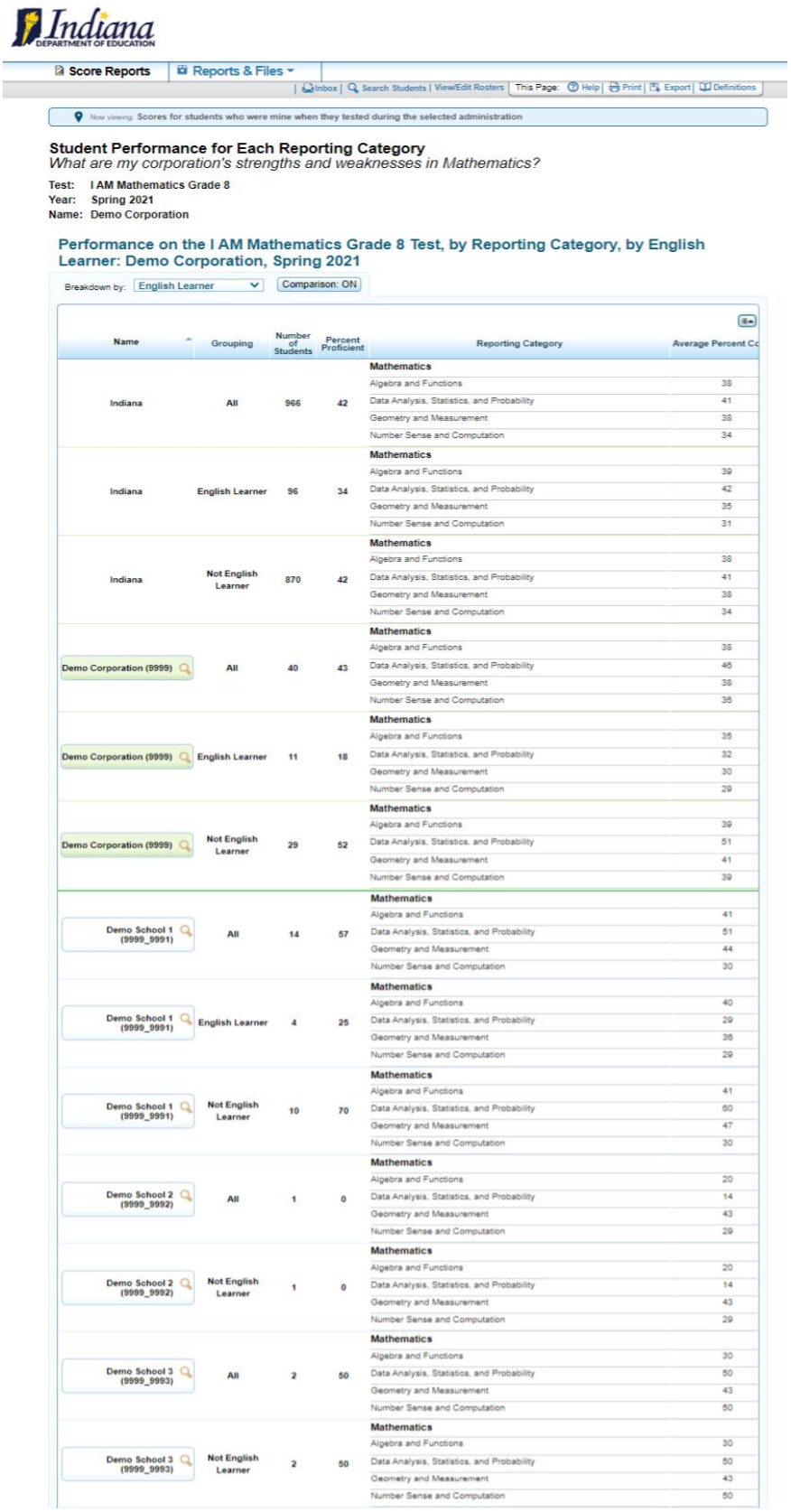


Figure 25: Corporation Aggregate-Level Reporting Category Report by English Learner for Grade 8 Mathematics



Indiana Department of Education

Score Reports | Reports & Files

Inbox | Search Students | View/Edit Rosters | This Page: Help | Print | Export | Definitions

Now viewing: Scores for students who were mine when they tested during the selected administration

Student Performance for Each Reporting Category

What are my corporation's strengths and weaknesses in Mathematics?

Test: I AM Mathematics Grade 8
Year: Spring 2021
Name: Demo Corporation

Performance on the I AM Mathematics Grade 8 Test, by Reporting Category, by English Learner: Demo Corporation, Spring 2021

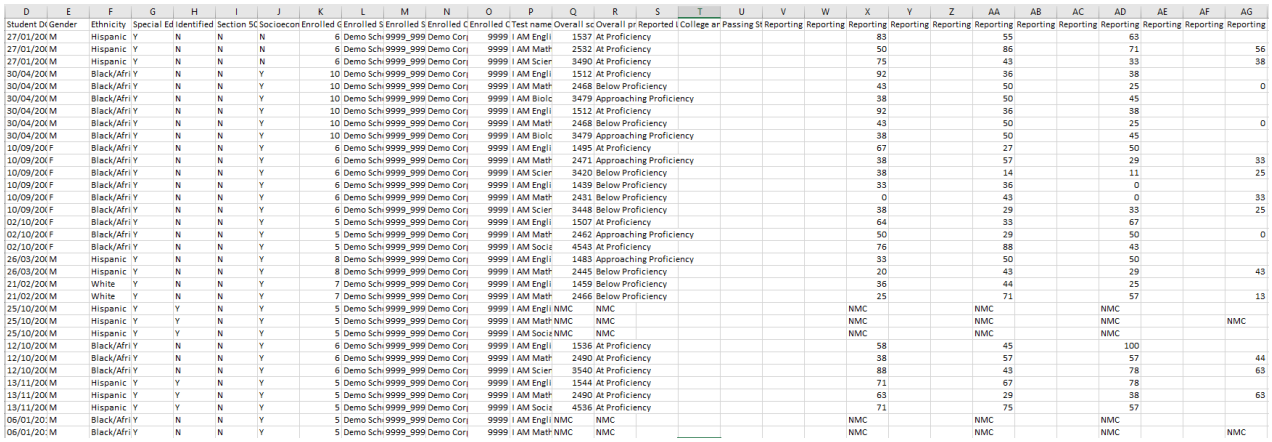
Breakdown by: English Learner | Comparison: ON

Name	Grouping	Number of Students	Percent Proficient	Reporting Category	Average Percent Cc
Indiana	All	966	42	**Mathematics**	
				Algebra and Functions	38
				Data Analysis, Statistics, and Probability	41
				Geometry and Measurement	38
				Number Sense and Computation	34
Indiana	English Learner	96	34	**Mathematics**	
				Algebra and Functions	39
				Data Analysis, Statistics, and Probability	42
				Geometry and Measurement	35
				Number Sense and Computation	31
Indiana	Not English Learner	870	42	**Mathematics**	
				Algebra and Functions	38
				Data Analysis, Statistics, and Probability	41
				Geometry and Measurement	38
				Number Sense and Computation	34
Demo Corporation (9999)	All	40	43	**Mathematics**	
				Algebra and Functions	38
				Data Analysis, Statistics, and Probability	46
				Geometry and Measurement	38
				Number Sense and Computation	36
Demo Corporation (9999)	English Learner	11	18	**Mathematics**	
				Algebra and Functions	35
				Data Analysis, Statistics, and Probability	32
				Geometry and Measurement	30
				Number Sense and Computation	29
Demo Corporation (9999)	Not English Learner	29	52	**Mathematics**	
				Algebra and Functions	39
				Data Analysis, Statistics, and Probability	51
				Geometry and Measurement	41
				Number Sense and Computation	39
Demo School 1 (9999_9991)	All	14	57	**Mathematics**	
				Algebra and Functions	41
				Data Analysis, Statistics, and Probability	51
				Geometry and Measurement	44
				Number Sense and Computation	30
Demo School 1 (9999_9991)	English Learner	4	25	**Mathematics**	
				Algebra and Functions	40
				Data Analysis, Statistics, and Probability	29
				Geometry and Measurement	36
				Number Sense and Computation	29
Demo School 1 (9999_9991)	Not English Learner	10	70	**Mathematics**	
				Algebra and Functions	41
				Data Analysis, Statistics, and Probability	60
				Geometry and Measurement	47
				Number Sense and Computation	30
Demo School 2 (9999_9992)	All	1	0	**Mathematics**	
				Algebra and Functions	20
				Data Analysis, Statistics, and Probability	14
				Geometry and Measurement	43
				Number Sense and Computation	29
Demo School 2 (9999_9992)	Not English Learner	1	0	**Mathematics**	
				Algebra and Functions	20
				Data Analysis, Statistics, and Probability	14
				Geometry and Measurement	43
				Number Sense and Computation	29
Demo School 3 (9999_9993)	All	2	50	**Mathematics**	
				Algebra and Functions	30
				Data Analysis, Statistics, and Probability	50
				Geometry and Measurement	43
				Number Sense and Computation	50
Demo School 3 (9999_9993)	Not English Learner	2	50	**Mathematics**	
				Algebra and Functions	30
				Data Analysis, Statistics, and Probability	50
				Geometry and Measurement	43
				Number Sense and Computation	50

1.6.9 Data File

ORS users have the option to quickly generate a comprehensive data file of their students' scores. Data files (see Figure 26) can be downloaded in Microsoft Excel or CSV format and contain a wide variety of data, including scale and reporting category scores, demographic data, and performance levels. Data files can be useful as a resource for further analysis and can be generated at the corporation, school, teacher, or roster level.

Figure 26: Sample Data File



D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
Student D	Gender	Ethnicity	Special Ed	Identified	Section 50	Socioecon	Enrolled G	Enrolled S	Enrolled S	Enrolled C	Enrolled C	Test name	Overall sc	Overall pr	Reported L	College ar	Passing St	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting	Reporting
27/01/20(	M	Hispanic	Y	N	N	N	6	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1537	At Proficiency						83			55			63			
27/01/20(	M	Hispanic	Y	N	N	N	6	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2532	At Proficiency						50			86			71			56
27/01/20(	M	Hispanic	Y	N	N	N	6	Demo Schi	9999_999	Demo Corj	9999	I AM Scier	3490	At Proficiency						75			43			33			38
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1512	At Proficiency						92			36			38			
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2468	Below Proficiency						43			50			25			0
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Biolc	3479	Approaching Proficiency						38			50			45			
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1512	At Proficiency						92			36			38			
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2468	Below Proficiency						43			50			25			0
30/04/20(	M	Black/Afri	Y	N	N	Y	10	Demo Schi	9999_999	Demo Corj	9999	I AM Biolc	3479	Approaching Proficiency						38			50			45			
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1495	At Proficiency						67			27			50			
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2471	Approaching Proficiency						38			57			29			33
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Scier	3420	Below Proficiency						38			14			11			25
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1439	Below Proficiency						33			36			0			
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2431	Below Proficiency						0			43			0			33
10/09/20(	F	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Scier	3448	Below Proficiency						38			29			33			25
02/10/20(	F	Black/Afri	Y	N	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1507	At Proficiency						64			33			67			
02/10/20(	F	Black/Afri	Y	N	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2462	Approaching Proficiency						50			29			50			0
02/10/20(	F	Black/Afri	Y	N	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Socia	4543	At Proficiency						76			88			43			
26/03/20(	M	Hispanic	Y	N	N	Y	8	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1483	Approaching Proficiency						33			50			50			
26/03/20(	M	Hispanic	Y	N	N	Y	8	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2445	Below Proficiency						20			43			29			43
21/02/20(	M	White	Y	N	N	Y	7	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1459	Below Proficiency						36			44			25			
21/02/20(	M	White	Y	N	N	Y	7	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2466	Below Proficiency						25			71			57			13
25/10/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	NMC	NMC						NMC			NMC			NMC			
25/10/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Math	NMC	NMC						NMC			NMC			NMC			NMC
25/10/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Socia	NMC	NMC						NMC			NMC			NMC			
12/10/20(	M	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1536	At Proficiency						58			45			100			
12/10/20(	M	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2490	At Proficiency						38			57			57			44
12/10/20(	M	Black/Afri	Y	N	N	Y	6	Demo Schi	9999_999	Demo Corj	9999	I AM Scier	3540	At Proficiency						88			43			78			63
13/11/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	1544	At Proficiency						71			67			78			
13/11/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Math	2490	At Proficiency						63			29			38			63
13/11/20(	M	Hispanic	Y	Y	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Socia	4536	At Proficiency						71			75			57			
06/01/20:	M	Black/Afri	Y	N	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Engli	NMC	NMC						NMC			NMC			NMC			
06/01/20:	M	Black/Afri	Y	N	N	Y	5	Demo Schi	9999_999	Demo Corj	9999	I AM Math	NMC	NMC						NMC			NMC			NMC			NMC

2. INTERPRETATION OF REPORTED SCORES

A student's performance on an assessment is reported as a scale score and a performance level for the overall assessment, and as a percent correct score for each reporting category. Students' scores and performance levels are summarized at the aggregate level. This section describes how to interpret these scores.

2.1 SCALE SCORE

The I AM assessment measures the knowledge and skills students are expected to develop and demonstrate in the context of Indiana 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. A scale score is the student's overall numeric score. Scale scores can be used to illustrate students' current level of performance and to compare the 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. 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 comparison. 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 demonstrate in each content area and at each grade level for each standard assessed and allow the user to understand the progression of skill expected across the different proficiency levels. It should be noted that the utility of scale scores is limited when comparing smaller differences among scores (or averaged group scores), particularly when the difference among scores is within the standard error of measurement (SEM). The details of SEM and the graphs of the conditional SEM (CSEM) of each test are provided in Volume 4, Evidence of Reliability and Validity. Furthermore, the scale score of individual students should be cautiously interpreted when comparing two scale scores, because small differences in scores may not reflect real differences in performance.

2.2 PERFORMANCE LEVEL

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 2.4, Cut Scores). PLDs are descriptions of content-area knowledge and skills that students at each performance level are expected to possess. Thus, performance levels can be interpreted based on PLDs. Students performing on the I AM at Level 3 are considered on track to demonstrate progress toward mastery of the knowledge and skills necessary for competitive employment and post-secondary education. Because performance levels are for the classification of the students into a small number of groups, such as those comprising

four or five students, and based on the cut scores, they have limited use for measuring growth. Thus, the performance level is an indicator of whether a student has mastered the required skills for a given level.

PLDs are available on the Indiana Department of Education web page.

2.3 Percent Correct 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 reporting category scores were computed as a percent-correct score for each student and an average percent-correct score for aggregate reporting. Since there may be systematic differences in the difficulty of subscales (e.g., probability versus number sense) as well as in the difficulty of subscales across alternate forms, reporting the raw score metric can be problematic even in fixed-form contexts. From the Spring 2022 I AM assessment the percent correct scores for the reporting category will be replaced with a subscale performance level classification determined by using the scale score and stand error of measurement of the reporting categories on the scale score metric.

2.4 Cut Scores

For all grades and subjects within I AM, scale scores are mapped onto three performance levels: Level 1–Below Proficiency, Level 2–Approaching Proficiency, and Level 3–At Proficiency. For each performance level, there is a minimum and maximum scale score that defines the range of scale scores students within each performance level have achieved. Collectively, these minimum and maximum scale scores are defined as *cut scores*, and they constitute the cutoff points for each performance level. Tables 7 through 10 show the cut scores for I AM.

Table 7: I AM ELA Assessment Proficiency Cut Scores

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
3	1300–1463	1464–1481	1482–1700
4	1300–1478	1479–1497	1498–1700
5	1300–1474	1475–1488	1489–1700
6	1300–1466	1467–1486	1487–1700
7	1300–1485	1486–1497	1498–1700
8	1300–1464	1465–1490	1491–1700
10	1300–1467	1468–1505	1506–1700

Table 8: I AM Mathematics Assessment Proficiency Cut Scores

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
3	2300–2462	2463–2473	2474–2700
4	2300–2461	2462–2478	2479–2700
5	2300–2459	2460–2470	2471–2700
6	2300–2461	2462–2477	2478–2700
7	2300–2466	2467–2477	2478–2700
8	2300–2463	2464–2474	2475–2700
10	2300–2470	2471–2484	2485–2700

Table 9: I AM Science Assessment Proficiency Cut Scores

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
4	3300–3475	3476–3496	3497–3700
6	3300–3465	3466–3488	3489–3700
Biology	3300–3471	3472–3502	3503–3700

Table 10: I AM Social Studies Grade 5 Assessment Proficiency Cut Scores

Grade	Level 1 Below Proficiency	Level 2 Approaching Proficiency	Level 3 At Proficiency
5	4300–4488	4489–4499	4500–4700

2.5 AGGREGATED SCORES

Percentage of students proficient, percentage of students in each proficiency level, and students' percentage correct scores are aggregated at roster, teacher, school, corporation, and state levels to represent how well a group of students performs overall and by reporting category on an assessment. When students' scores are aggregated, these scores can be interpreted as an estimate of knowledge and skills that a group of students possesses. This interpretation makes aggregated scores a powerful tool when comparing student performance across different groups of students, whether it be at a similar level of aggregation (e.g., school to school) or an analysis of a subgroup (e.g., comparing a teacher's roster to the overall school).

2.6 APPROPRIATE USES FOR SCORES AND REPORTS

Assessment results can be used to provide information on individual students' performance on the assessment. Overall, assessment results demonstrate what students know and are able to do in certain subject areas and give further information on whether students are on track to demonstrate the knowledge and skills necessary for competitive employment and post-secondary education. 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 result 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. Student performance on an assessment may vary due to a variety of reasons (e.g., they are not feeling well, or 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 ORS, 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 an 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 take into account the group size. The smaller the group, the larger the measurement error related to these aggregated data, thus requiring a more cautious interpretation.

3. SUMMARY

I AM results are reported online via the ORS at the end of each test window. After the administration window ends, schools have up to two weeks to submit appeal requests to invalidate tests if a student receives an accommodation for which he or she is not eligible or is otherwise impacted by an irregularity that affects the validity of the student's assessment attempt. The Indiana Department of Education (IDOE) Office of Assessment staff review invalidation requests and provide approval for cases that meet the aforementioned criteria. CAI incorporates final invalidation data into a comprehensive student data file (SDF). After generating the student data file, CAI reviews preliminary scores against the scoring specifications to ensure accuracy. Scores are internally replicated by CAI's Statistics Analysis System (SAS) team. The SDF is then provided to an external, third party vendor, HumRRO that also completes score data replication and provides a summary of matches and any evidence of a mismatch in scores. Data replication concludes after both CAI and HumRRO have an overall match in score data.

Score data goes through internal user acceptance testing (UAT) that verifies several test cases reflecting the varying proficiency levels of the I AM student demographic. After such preliminary checks are completed, the I AM results are approved for release. The I AM Spring 2021 assessment scores were released and reported in the ORS on June 10, 2021.

The reporting system is interactive. When educators or administrators log in, they see a summary of data about students for whom they are responsible (e.g., a principal would see students in his or her school; a teacher would see students in his or her class). They can then drill down through various levels of aggregation all the way to individual student reports. The system allows them to tailor the content more precisely, moving from subject area through reporting categories and even to standards-level reports for aggregates. Aggregate reports are available at every level, and authorized users can print or download the reports as well as the data on which the reports are based. ISRs can be produced individually or batched as PDF files.

All authorized users can download reports, including data about students for whom they are responsible, at any time. The various reports may be used to inform stakeholders regarding student performance and instructional strategies.