EXHIBIT V

Washington Office of Superintendent of
PUBLIC INSTRUCTION

WASHINGTON ACCESS TO INSTRUCTION & MEASUREMENT (WA-AIM)

Technical Report
2021–2022

October 2022

Prepared by

DRC

for

The Office of the Superintendent of Public Instruction
P.O. Box 47200

Olympia, WA 98504-7200

COPYRIGHT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

The Washington Access to Instruction & Measurement (WA-AIM) program was implemented in fall 2014 as Washington's alternate assessment based on alternate academic achievement standards (AA-AAAS) for students with the most significant cognitive disabilities.

The 2021–2022 administration was the eighth year of the WA-AIM program. This chapter starts with an overview of the COVID-19 situation in the 2021–2022 administration. Afterwards, the intended purpose and the structure of this report are introduced. The remaining chapter provides general information of the WA-AIM assessment program, including its intended uses and score interpretation, target population, overall assessment design, reported scores, and standard setting.

1.1 Context

The ongoing COVID-19 pandemic continued to disrupt the educational experiences of many Washington students in school year 2021–2022. Most students returned to in-person instruction in fall 2022. However, many districts and schools struggled with intermittent school closures, bussing issues, school staffing shortages, and frequent staff and student absences due to localized COVID-19 outbreaks. Additionally, Washington saw a decline in public school enrollment, and an increase in the number of students enrolling in local school district online instructional alternative learning offerings.

1.2 This Technical Report

The main purpose of this technical report is to document 1) the assessment activities that have occurred in the administration and related procedural validity evidence, 2) characteristics and test performance of students who have participated in the 2021–2022 WA-AIM assessment, and 3) test analyses and test reliability and validity evidence based on collected data.

There are eight chapters in the technical report. Chapter 1 (*Introduction*) provides a background for the 2021–2022 WA-AIM assessment and presents general information to help readers understand the assessment, including its intended uses and score

interpretations, target student population, overall assessment design, reported achievement levels and scores, and related standard setting activities.

Chapter 2 (*WA-AIM Assessment Development*) documents past WA-AIM assessment development activities. Chapter 3 (*Test Administration*) summarizes information related to test administration such as the test window, the data collection platform configurations and procedures, test administration materials and educator training, test accommodation and adaptation, and supporting quality control measures. Major enhancements that have been implemented in the online administration system starting in the 2020–2021 WA-AIM are also summarized in this chapter. Chapter 4 (*Data Auditing and Score Reporting*) focuses on post-assessment activities that relate to scoring and reporting. Chapter 5 (*Feedback Loop for Assessment Improvement*) describes a feedback loop that has been implemented to continuously improve the WA-AIM assessment based on user and Alternate Assessment Auditing feedback.

Chapter 6 (*Test Data Analyses*) presents analysis results of empirical test data from the 2021–2022 WA-AIM administration, including test participation, total score distributions, achievement level distributions, reliability indices, classification accuracy and consistency indices, and subgroup statistics.

Lastly, Chapter 7 (*Fairness in Testing*) and Chapter 8 (*Reliability and Validity*) summarize evidence related to test fairness, reliability/precision, and validity, and describe how individual chapters in this report combine to form an overall validity argument.

Given the iterative nature of an educational program design process (Willis, 1995; Crawford, 2004), it is not unusual to see findings from test evaluation being used to inform continued improvement of the test program and validation of its uses. It is expected that the technical information presented in this report will be reviewed and that constructive discussions will be held on a regular basis with various educational stakeholders and technical advisors with the purpose of promoting progressive development of and enhancement to the assessment system, leading to optimal benefits for Washington educators and students.

1.3 WA-AIM Intended Uses and Score Interpretations

Statewide testing is part of a comprehensive system intended to ensure all public school students, no matter where enrolled, receive a quality education. Washington offers a comprehensive assessment system through which students are tested by the state to assess their progress toward grade-level standards as they move through elementary, middle, and high school. Additionally, in high school, state assessments can be used to determine whether a student has mastered a minimum set of skills required for graduation. An overview of Washington's testing system can be found at http://www.k12.wa.us/assessment/StateTesting/.

Given the legislative context within which the entire statewide assessment system sits, the WA-AIM assessment is governed by the same laws and rules that govern the state's general assessments. Federal legislation, including the Individuals with Disabilities Education Act (IDEA) of 2004 and the Elementary and Secondary Education Act (ESEA) of 2015 (the Every Student Succeeds Act, abbreviated as ESSA), requires that students with disabilities have access to the general curriculum, with appropriate accommodations where necessary, and that each student be assessed on the same general curriculum standards as all other students.

For students who are unable to participate in regular assessments, even with accommodations as indicated in their respective Individualized Education Programs (IEPs), a state must develop and implement an alternate assessment based on AA-AAAS. The AA-AAAS is typically designed with a reduction in academic breadth, depth, and complexity that acknowledge students' disabilities while maintaining linkage to the same general academic standards taught to all students.

The WA-AIM assessment was designed as the AA-AAAS to the general assessment for students with the most significant cognitive disabilities. It serves as Washington's federal and state accountability assessment for reporting of student progress toward state grade-level standards; additionally, it can be used to fulfill a high school student's Certificate of Individual Achievement (CIA) which is one available pathway used for purposes of graduation.

The assessment intends to measure student knowledge and skills in the content areas of mathematics and English language arts (ELA) at grades 3 through 8 and high school and in the content area of science at grades 5, 8, and high school.

It should be noted that due to circumstances related to the COVID-19 pandemic, participation rates in the 2021–2022 WA-AIM, although higher than those from the 2020–2021 administration, were not as high as those in a typical administration year (see more details about the student participation in Section 6.1, Chapter 6). In addition to lower than typical test participation rates, aggregated test performance results from the 2021–2022 WA-AIM should be interpreted in the context of several factors, including possible disrupted learning during previous and current school years, differences in instruction delivery (e.g., online, in-person, or hybrid model), and potential overrepresentation of certain demographic groups and underrepresentation of other groups in the tested students. As such, any comparison of the group test results from the 2021–2022 administration should be made with caution.

Additionally, with the continued disruptions and interruptions due to COVID-19 over the last two administrations, caution is urged when comparing the 2021–2022 data to any previous administrations. Readers of this report are encouraged to consider the context provided in Section 1.1 when interpreting the data and results and when thinking of the data presented as a baseline for future administrations to be compared to.

1.4 WA-AIM Target Population and Participation Eligibility

The ESEA requires the participation of all students enrolled in grades where state-level testing is mandated. State and federal requirements have been aligned so that all students must participate fully and meaningfully in the state-level assessments. For a very small percentage of students, participation in the statewide assessment program is achieved by participating in the WA-AIM assessment. Specifically, students with the most significant cognitive disabilities who are working toward alternate academic achievement standards, as documented in their IEP, are eligible to take the WA-AIM, and thus constitute the intended testing population for the assessments.

For purposes of the WA-AIM, students with the most significant cognitive disabilities are defined as those students who require intensive or extensive levels of direct support that is not of a temporary or transient nature. Students with the most significant cognitive disabilities also require specially designed instruction to acquire, maintain or generalize skills in multiple settings in order to successfully transfer skills to natural settings including the home, school, workplace, and community. In addition, these students score at least two (2) standard deviations below the mean on standardized, norm-referenced assessments for adaptive behavior and intellectual functioning.

Students who participate in the WA-AIM represent a diverse population having in most instances severe and/or multiple disabilities impeding cognitive capacity and performance expected of nominal K–12 children. A student may experience compounding effects of limited cognitive processing of academic information with impacted modes of communication. In the majority of circumstances, these students' academic instruction occurs in self-contained classrooms, and only about one-fourth of the students receive academic instruction in regular education settings.

In general, the decision as to how a student with a disability participates in the state's accountability system is made by the student's IEP team. The IEP team, including a student's parents/guardians, determines on an individual basis how a child with an IEP participates in state assessment. This determination should be made at every annual IEP review. The IEP team must determine if a student will participate in the state general assessment, with or without accommodations, or the state alternate assessment.

If the IEP team determines that the general assessment, even with accommodations, may not be the appropriate means of measuring a particular child's knowledge and skills, the team must discuss the participation criteria for the alternate assessment (WA-AIM). Only those students meeting the criteria and factors for the alternate assessment should participate in the WA-AIM. When considering whether students with disabilities should participate in the WA-AIM, the IEP team is required to use the criteria developed by the OSPI.

IEP team guidance for determining whether or not a student is eligible to participate in an alternate assessment can be found in OSPI's *Guidance for IEP Teams: Student Participation in Statewide Assessments for Accountability and Graduation*.

For a student to be considered as having a significant cognitive disability and therefore, appropriate for consideration as a candidate for an alternate assessment a student must:

- have documented cognitive and adaptive behavior disabilities that are both at least two or more standard deviations below the mean and that are demonstrated in school, work, home, and community environments even with program modifications, adaptations, and accommodations;
- be eligible for special education under one or more of the existing categories of disabilities under IDEA (e.g., intellectual disabilities, multiple disabilities, traumatic brain injury, autism) and have an IEP in effect at the time of the decision and during the duration of the assessment;
- require extensive direct and individualized instruction and/or extensive supports in and across multiple settings to acquire, maintain and generalize academic and functional skills necessary for application in school, work, home, and community environments.
- The student's need or extensive direct individualized instruction is not temporary or transient;
- be learning content that is linked to (derived from) the K-12 Learning Standards, that have been appropriately broken into a continuum of access points in order to provide the student with entry points of varying levels of complexity to show their knowledge and skills aligned to the K-12 Learning Standards; and
- need substantial supports to achieve gains in the grade and age-appropriate academic and functional curriculum and require substantially adapted materials and customized methods of accessing information in alternative ways to acquire, maintain, and generalize skills across multiple settings.

There are other issues that may affect a student's educational experience and his/her ability to learn and show what he/she knows that are not appropriate to consider during the decision-

making process for the alternate assessment. ***The following criteria may not be used for alternate assessment participation decisions:***

- poor attendance, excessive or extended absences
- disability related to visual or auditory disabilities, emotional-behavioral disabilities, specific learning disabilities, or speech and language impairment
- lack of access to quality instruction in core standards
- social, cultural, linguistic, or economic differences for the WA-AIM; however cultural and linguistic differences should not be used as sole exclusionary factors for eligibility to participate in the WIDA Alternate ACCESS
- below average reading or achievement levels
- displays of behaviors or emotional distress during testing
- expectations of poor performance, non-proficiency, or the pre-determined or anticipated impact of the student's performance on the school/district on-grade level assessment scores
- an administrative decision
- the student's disability category, educational placement, type of instruction, and/or amount of time receiving special education services.

1.5 Overall WA-AIM Assessment Design

The WA-AIM assessment is built off of Access Point Frameworks, which have been designed to connect Washington's learning standards in mathematics, ELA, and science in such a way that students with the most significant cognitive disabilities have multiple access points to the standards. The frameworks specify five standards at each grade and content area, with three access points: *Less Complex*, *Intermediate Complex*, and *More Complex*, where *Intermediate Complex* defines the "minimal mastery level" knowledge and skills of the grade-level standard.

The WA-AIM assessment measures all five standards from the Access Point Frameworks at each assessed grade and content area. Educators are required to assess each standard at a chosen access point for their student with a corresponding Performance Task that consists of five unique, dichotomously scored (score 0 for incorrect responses, and score 1 for correct

responses) items, which may take a variety of formats, including multiple-choice (MC), constructed-response (CR), and performance. Educators can only use items from the state-provided item library (see more details about the item library in Section 2.4 of Chapter 2, and in Sections 3.4 and 3.5 of Chapter 3).

The WA-AIM design encourages a baseline measure, targeted academic instruction over time, with the final measure being used for state and federal reporting purposes. The baseline measure is used to determine the appropriate access point entry level for measuring the student's attainment of related content skills and knowledge across a school year. Once the access point entry level for each standard is determined from the baseline measure, it is recommended that a minimum of six weeks of targeted academic instruction follow before administration of the final measure. Only results from the final measure are used for state reporting purpose.

When administering final assessments educators have the choice of using a pre-built form for each standard and access point or building their own forms with items from the state-provided item library. Pre-built forms contain five items and stimulus and fully meet the Requirements and Restrictions defined in the Performance Task for that standard and access point. Educator created forms are built by educator self-selection of items from the item pool available for that standard and access point. The educator must ensure the five items selected in totality meet all Requirements and Restriction. Educator created forms are taken completely through review by the DRC's Alternate Assessment Auditing team (see more details in Sections 4.1 and 4.2 of Chapter 4) to ensure adherence to the Requirements and Restriction.

OSPI regularly communicates with the DRC's Alternate Assessment Auditing team to consider feedback from data review activities for possible incorporation into design and development of documents and trainings to use in teacher professional development, focused on continuing efforts to minimize impacts from teacher errors on student scores.

The option of using an educator created form is available for ELA and Math. This option is not available for science due to the multi-dimensionality of the Washington State K–12 Science

Standards and the need for all five items to work dependently to fully measure performance expectations in science.

All items associated with each Performance Task can be adapted to meet each individual student's learning style and preferred mode of receptive and expressive communication. Educators are encouraged to present the items in styles that most closely resemble how daily instructional materials are presented to the student. Typical adaptations and ideas are listed at the beginning of each grade-level set of Performance Tasks. Local supporting materials and administrations of the Performance Tasks must comply with the Performance Task Requirements and Restrictions.

To ensure integrity of educator-submitted data, an observer attestation is required along with the submission. The attestation must be completed by an educational professional (administrator, paraprofessional, educator, or service provider) who is not the educator administering the assessment. The observer must observe the student performing the task and verify that the student independently generated the answers as documented in the educator-submitted data into the data collection platform.

During the process of designing and developing the Access Point Frameworks, Performance Task specifications, and WA-AIM assessment administration requirements and procedures, attention was given to methods of increasing accessibility, reducing unintentional bias, ensuring meaningful coverage of the general education learning standards, and better standardizing assessment tasks and administration protocols. More details on the WA-AIM assessment development are presented in Chapter 2 of this report.

Washington educators were involved throughout the development of key components of the assessment and its standards, such as reviews of the Access Point Frameworks and Performance Task specifications, drafting and reviews of the Alternate Achievement Level Descriptors (AALDs), participating in a weighting study in which score weights were derived for varying access points based on expert judgment and empirical data, and participating in standard setting activities.

To improve the assessment, the OSPI has been actively gathering feedback from stakeholders and providing targeted teacher support based on the feedback. For example, a feedback loop has been implemented to continuously improve the WA-AIM assessment and its administration based on feedback a) from schools and districts and b) from the DRC Alternate Assessment Auditing team review of teacher submissions of student assessment records. Actions following OSPI's review of collected feedback have led to enhancements to the assessment systems such as those listed in Section 3.5, Chapter 3.

1.6 Achievement Levels and Reported Scores

The WA-AIM assessment reports four alternate achievement levels in each content area: Level 1, Level 2, Level 3, and Level 4, where Level 4 represents the highest level of knowledge, skills, and understandings. Level 3 and above has been used as the criterion of "meeting standard" in the state accountability system. Table 1.2 provides the WA-AIM cut scores across content areas and grade levels. Information about past standard setting activities can be found in Section 1.7 of this chapter.

Table 1.1. WA-AIM Cut Scores, All Content Areas and Grade Levels

Content	Grade	Cut Scores		
		Level 2	***Level 3***	***Level 4***
ELA	3	109	124	150
	4	107	125	158
	5	108	129	162
	6	110	125	159
	7	108	123	154
	8	110	123	150
	10	110	126	162
Mathematics	3	108	129	161
	4	106	126	161
	5	106	120	153
	6	109	131	160
	7	109	124	163
	8	112	133	162
	10	108	120	146
Science	5	111	129	169
	8	110	127	163
	11	111	128	164

In addition to alternate achievement levels, WA-AIM also reports content area test scores and number correct standard-level scores weighted by access point level.

To obtain content area test scores, the WA-AIM assessment uses weighted raw scores and grade-specific scaling. Specifically, a raw score (range: 0–5) is calculated for each standard as a sum of student observed scores across five dichotomously scored items (score 0 for incorrect responses, and score 1 for correct responses) in the administered Performance Task at the given standard. Then, the raw score for each standard is weighted according to the access point (complexity level) of the administered Performance Task. Based on a weighting study (see more details in the 2018–2019 WA-AIM Technical Report), the following set of weight values for access points are used, which apply to all grades and content areas.

- *Less Complex*: 0.7
- *Intermediate Complex*: 1.7
- *More Complex*: 4.0

The sum of the weighted scores across standards (five standards in total per test) results in an overall scale score that ranges from 0 to 100. For reporting, the scale score is rounded to integer and added to by 100 to place the reported score on a 100–200 scale.

For example, if a student was assessed at the *Intermediate Complex* access point for a given standard and received item scores of 0, 1, 1, 1, 0 in the tested Performance Task at that standard, the student's raw score at that standard is 3 (0 + 1 + 1 + 1 + 0 = 3). The raw score is then multiplied by 1.7 (the weight for *Intermediate Complex*), which results in a weighted standard-level score of 5.1 (1.7 × 3 = 5.1). Suppose the student's weighted scores on the other four standards are 8.0, 2.1, 8.5, and 16, respectively. The sum of the standard scores for that student would be 39.7 (5.1 + 8.0 + 2.1 + 8.5 + 16 = 39.7). After rounded to 40 and added to by 100, the student's total test score for reporting is 140 (40 + 100 = 140).

Overall, there are five major types of scores calculated for each student:

- Item score (0–1)*
- Raw score on each standard (0–5) based on the sum of item scores
- Weighted raw score on each standard (0–20)
- Content area total score (100–200)
- Content area achievement level (Level 1, 2, 3, or 4)

* *Irrespective of the item type (multiple choice, constructed response, etc.), each item is dichotomously scored.*

Note that any invalidated Performance Task assessment from data auditing (see more details about data auditing in Sections 4.1 and 4.2, Chapter 4) is marked non-scorable (NS) and results in a raw score of zero for the corresponding standard in the student's total score calculation. Similarly, if an item score is invalidated, the invalidation results in a score of zero on that item.

1.7 Standard Setting

The initial set of WA-AIM cut scores were established in 2015 following the first administration of the WA-AIM assessment. With the addition of the high school science assessment to the WA-

AIM in the 2015–2016 school year, a Standard Setting study was conducted in 2016 to set cut scores for high school science.

In the 2017–2018 school year, new content standards were established for science at grades 5, 8, and high school (grade 11), and the WA-AIM has been aligned to these new science standards starting from that year.

Also starting in the 2017–2018 school year, the high school assessments of ELA and mathematics have been shifted from grade 11 to grade 10. For ELA, this shift in grade level also came with a change to the content standards: the high school assessment of ELA now aligns to the content standards for grades 9–10 instead of grades 11–12. For mathematics, the assessment continues to measure the same content standards across grades 9–12.

Because of the change in content standards for grades 5, 8, and high school science, and for high school ELA, a Standard Setting study was conducted in spring 2018 to set cut scores for those assessments.

Although the content standards did not change for high school mathematics and the same assessment was used, the test administration has shifted to grade 10 instead of grade 11. Deference to sound technical process led to OSPI's decision to conduct a standards validation to review the cut scores for high school mathematics and determine whether the cut scores would remain valid for continued use.

At the end of each standard setting workshop, a summary of the workshop participant recommendations was submitted to the OSPI for consideration. After internal reviews, the OSPI presented the recommended cut scores from the standard setting workshop to the Washington State Board of Education for adoption. The standard setting design and results were also presented to the Washington's National Technical Advisory Committee (NTAC).

This section briefly summarizes the three standard setting studies that were conducted in 2015, 2016, and 2018, respectively, followed by a presentation of the final set of cut scores that have

been applied to the WA-AIM administration since spring 2018. Detailed information regarding the standard setting study procedures, study participants, and associated materials is presented in stand-alone WA-AIM standard setting technical reports (refer to the OSPI website to access state assessment technical reports.

1.7.1 2015 Standard Setting

The 2015 WA-AIM standard setting study consisted of four activities, including one (Teacher Achievement Level [TAL] Study) that occurred prior to the standard setting workshop.

1) **TAL study (based on the contrasting groups procedure)**. Before the workshop, DRC (formerly CTB) conducted an online study coinciding with the end of the 2014–2015 assessment window where special educators from across the state studied the alternate achievement level descriptors (AALDs) for each achievement level and decided which level best described each of the students in their classroom.

2) **Standard setting workshop**. On July 14–16, 2015, educators from across the state of Washington convened for the WA-AIM standard setting. The OSPI, in collaboration with DRC, convened 76 participants for a multiphase workshop. Over the course of three days of discussions and deliberation, the educators recommended cut scores for the WA-AIM defining four achievement levels: Level 1, Level 2, Level 3, and Level 4, where Level 4 represents the highest level of knowledge, skills, and understandings.

a. **Judgmental policy capturing workshop** (also known as a profile sorting workshop). The initial phase of the standard setting workshop had educators consider the Access Point Frameworks, Performance Task item examples, and AALDs. Using this information, they considered the content-based expectations for students in each achievement level and then examined 100 sets of test scores from Washington students who took the WA-AIM in each grade and content area combination. Participants sorted these 100 profiles into achievement levels based on the content knowledge demonstrated by students through their assigned scores.

b. **Synthesis discussion**. As a complete group, participants at each table examined the recommended cut scores from the profile sorting process, using the results of the TAL study to validate and build context around these results. The table leaders from the profile sorting then convened to examine the recommended cut scores for reasonableness and for consistency across grades within each content area. The table leaders in ELA and science, incorporating comments from the entire committee, confirmed that the results of the profile sorting process reflected the content-based expectations from the process; the mathematics team, though reflecting similar thinking, recommended adjustments to the cut scores at specific grade levels to promote better articulation across all grades.

c. **AALD refinement**. A portion of the committee convened for the last activity of the standard setting workshop to review the AALDs. Using their learnings from throughout the standard setting workshop, the team recommended refinements to the AALDs to make them clear, well-articulated, and useful for educators throughout the state.

1.7.2 2016 High School Science Standard Setting

The 2016 WA-AIM high school science standard setting was aimed to establish WA-AIM high school science cut scores to reflect academic expectations as described in the Access Point Frameworks and AALDs for students with the most significant cognitive disabilities. The study followed a similar approach to that of the 2015 WA-AIM standard setting, and may be seen as a continuation of the 2015 work.

1.7.3 2018 Standard Setting for Science and High School ELA and Mathematics

The 2018 standard setting and standards validation processes followed the steps as described below.

1) **TAL study**. A total of 187 special education practitioners reviewed the AALDs and rated the performance of students in their classrooms across all three content areas that were part of the standard setting or standards validation studies. A summary of the TAL study can be found in the 2017–2018 WA-AIM Technical Report.

2) **Standard setting workshop**. On July 24–26, 2018, educators from across the state of Washington convened for the WA-AIM standard setting. The OSPI, in collaboration with DRC, convened 30 participants for a multiphase workshop. Over the course of three days of discussions and deliberation, the educators recommended cut scores for WA-AIM science assessments at grades 5, 8, and high school, and for WA-AIM high school ELA assessment. Educators also validated the cut scores for WA-AIM high school mathematics.

a. **Judgmental policy capturing workshop**. The workshop engaged 30 Washington educators to review the AALDs and to consider the scoring patterns of students in each achievement level on the WA-AIM. Participants engaged in two rounds of discussions and decision-making to recommend cut scores for the tests that reflect the types of knowledge and skills expected of students in each achievement level. a. For high school mathematics, seven educators reviewed the existing cut scores using the judgmental policy capturing process. This same process was used to recommend cut scores for the test in 2015. Participants saw the achievement level that various students earned, based on the existing cut scores; and using those students' scoring patterns, they determined whether the existing cut scores were still valid for continued use.

b. **Synthesis discussion.** Participants came together to review their recommended cut scores, the associated impact data, and the results of the teacher achievement level study. Participants noted that they had spent three days of careful study and reflection on the WA-AIM AALDs and test, and they believed their recommended cut scores were defensible and reflected the knowledge and skill expected of students in each achievement level.

c. **AALD refinement**. Participants refined the AALDs, making suggestions to enhance the clarity and usefulness of the AALDs for educators in the field.

Chapter 2. WA-AIM Assessment Development

In this chapter, historical background of the WA-AIM development is introduced, followed by a description of three phases that the WA-AIM development has undergone. The three phases include development activities prior to, during, and after the 2017–2018 WA-AIM administration. The WA-AIM assessment design is provided in Section 1.5, Chapter 1.

2.1 Historical Background of the WA-AIM Development

In 2011, the OSPI adopted new college- and career-ready learning standards in the areas of English language arts (ELA) and mathematics. With the adoption of new standards, the OSPI was required to develop new Alternate Assessments based on Alternate Academic Achievement Standards (AA-AAAS) that align to the new standards. In October 2013, the OSPI, assisted by Measured Progress, began development of the WA-AIM assessment. The intention was to have the new alternate assessment ready for initial use in the 2014–2015 school year, the same year as the state's transition to new general assessments aligned to college- and career-ready standards.

The development of the ELA and mathematics WA-AIM assessments was based on the Common Core Essential Elements (CCEEs), which were authored by the Dynamic Learning Maps (DLM) Alternate Assessment System Consortium (2013a, 2013b).

In June 2017 the Washington State Legislature took action to move the high school English language arts and mathematics assessments administration from grade 11 to grade 10. (Refer to 28A.655.061 of the Revised Code of Washington.)

Once this action was signed into law, the OSPI began a review of the high school English language arts and mathematics WA-AIM assessment to determine if and what changes needed to occur based on an earlier administration year.

To determine what changes needed to occur, the OSPI reviewed the current standards and Essential Elements measured on the WA-AIM to see if there were significant differences in the essential knowledge, skills, and abilities between grade 10 and 11 standards.

For mathematics, both the Common Core State Standards (CCSSs) and the related CCEEs are grouped at the high school level by domain and are applicable to grades 9–12. Due to this, it was determined that the content of the high school Mathematics WA-AIM did not need to change, but we would convene a panel of educators to conduct a math achievement level validation, to be held with science and high school ELA achievement level setting.

For ELA, the CCSSs and the related CCEEs are banded for grades 9–10 and then for grade 11–12. Federal legislation requires that students participate in grade-level assessments aligned to grade-level standards. Due to this requirement, the OSPI determined that portions of the high school ELA WA-AIM would need to be rewritten to fulfill this requirement.

For science, although adopting the Next Generation Science Standards (NGSS) in 2013, Washington's state science assessments continued to align to Washington's previous K–12 Science Learning Standards (adopted in 2009) through the 2016–2017 school year. During the 2016–2017 school year, the OSPI began development work on the new Science WA-AIM aligned to the Next Generation Science Standards (NGSS). The new assessments were implemented starting with the 2017–2018 administration.

To reduce teacher burden and to standardize assessment content, OSPI contracted DRC in 2019 for item development to provide educators with fifteen items for each standard and access point for use in the WA-AIM final assessment.

The bulk of the item content was deployed for use during the 2020–2021 WA-AIM administration. Some passage-based ELA items and additional science sets were unable to be made available for the 2020–2021 administration but have been made available for the 2021–2022 administration.

The new item development has led to an expanded, standardized item library. With the item library available, educators administering the WA-AIM are no longer allowed or able to create their own item content, and all final assessments must use items from the library that was developed through OSPI and DRC.

2.2 WA-AIM Development: Prior to the 2017–2018 Administration

The alternate assessment design for the WA-AIM began with the development of alternate standards, the Access Point Frameworks. The Access Point Frameworks connect to the general learning standards to allow students with the most significant cognitive disabilities to access an assessment connected to the grade-level academic content.

After development of the Access Point Frameworks, Performance Tasks were developed to give educators the means to measure a student's knowledge of the standards. During administration of the assessment, students are allowed to use the mode of communication preferable to each to demonstrate their mastery of the knowledge and skills of the standards.

Washington educators participated in the development of both the Access Point Frameworks and the Performance Task specifications.

2.2.1 Target Standards and Standard Elements

In the content areas of mathematics and ELA, access points were expanded from the CCEEs, which were authored by the DLM Alternate Assessment System Consortium (2013a, 2013b). The CCEEs provide specific statements of the content and skills that are linked to the Common Core State Standards (CCSS) grade-level specific expectations for students with the most significant cognitive disabilities.

More information about the DLM project is at http://dynamiclearningmaps.org/. To facilitate Washington educators' understanding and use of the CCEEs, the OSPI also provided an introduction of the CCEEs at https://www.k12.wa.us/sites/default/files/public/specialed/pubdocs/ccee-ccss-math.pdf (for mathematics) and https://www.k12.wa.us/sites/default/files/public/specialed/pubdocs/ccee-ccss-ela.pdf (for ELA).

Prior to the 2017–2018 school year, the state science assessments were aligned to the 2009

version of the *Washington's K–12 Science Learning Standards*. In those assessments, the access points in science are expanded from the Essential Academic Learning Requirements (EALRs) and associated performance expectations for the prioritized standards. EALRs define what all students should know and be able to do at each grade level, and there are four EALRs in the science standards. EALRs 1, 2, and 3 (labeled as *System*, *Inquiry*, and *Applicatio*n, respectively) describe crosscutting concepts and abilities that characterize the nature and practice of science and technology, while EALR 4 (*The Domain of Sciences*) describes what all students should know and be able to do in the domains of Life, Physical, and Earth & Space Science. There is one Big Idea each for EALRs 1, 2, and 3. In EALR 4, nine Big Ideas were identified: three in Life Science, three in Earth & Space Science, and three in Physical Science. Each Big Idea is a single important concept that begins in the early grades and builds toward an adult-level understanding. A detailed description of the 2009 science standards and the standards' components is provided at http://www.k12.wa.us/science/pubdocs/WAScienceStandards.pdf.

Target standards and standard elements were selected from each content area and grade, which then served as the building foundation of the WA-AIM access points. The selection for mathematics and ELA was guided by test blueprints of the Smarter Balanced assessment (Washington's general assessments for accountability) to ensure that the WA-AIM assessment would measure a student's academic skills while promoting access to the general education curriculum. Measured Progress content specialists, in cooperation with the OSPI, intended the selection for each content area to provide broad academic coverage at each grade level and across grades. For ELA, the selection consists of five strands for grades 3 through 8 and high school, encompassing reading, writing, speaking, and listening. For mathematics, five domains were selected at each grade level. For science, five Big Ideas were selected from the Washington's K–12 Science Learning Standards (2009) for grades 5, 8, and high school, with all four EALRs covered.

The State adopted the Next Generation Science Standards (NGSS) in 2013. During the 2016–2017 school year, the OSPI began development work on the new Science WA-AIM aligned to

the NGSS. The new assessments were implemented starting with the 2017–2018 administration. Information regarding the new science development is presented in Section 2.3.

2.2.2 Access Point Frameworks Development

Access points serve to connect to the robust state standards and were determined through educator vetting to represent manageable content for assessing students. In the Access Point Frameworks, for each standard assessed, a continuum of three access points was developed, representing three levels of complexity for each content area and each grade: *More Complex*, *Intermediate Complex*, and *Less Complex*. The *Intermediate Complex* access point was developed to demonstrate the "minimal mastery level" knowledge and skills for that grade-level standard.

In ELA and mathematics, linkage between the CCSS, CCEEs, and the Access Point Frameworks is present, with the access points defining the knowledge and skills measured with the assessment at varying complexity levels. Similarly, in science, linkage between EALRs, Big Ideas, and the corresponding performance expectations in the Access Point Frameworks is articulated, with each access point describing a specific performance expectation at its given complexity level.

To illustrate the Access Point Frameworks, below is a screenshot of the access points on one standard from the framework document on grade 3 ELA. The complete WA-AIM Access Point Frameworks are available at http://www.k12.wa.us/Assessment/WA-AIM/Frameworks.aspx.

Figure 2.1. Example Access Point from the Access Point Frameworks

ENGLISH LANGUAGE ARTS				**Grade 3**
Strand: Reading: Literature				
Substrand: Key Ideas and Details				
Common Core State Standard	Common Core Essential Element	**ACCESS POINTS Built on Three Levels of Complexity**		
		More Complex ◀·····◀·····◀·····▶·····▶·····▶ Less Complex		
RL.3.1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.	EE.RL.3.1 Answer who and what questions to demonstrate understanding of details in a text.	**Student will use details to answer questions about the plot and characters in a text.**	**Student will answer questions about characters in a text.**	**Student will identify text details, such as character, in a text.**

The access point authoring started with ELA and mathematics, following an iterative process between Measured Progress content and special education specialists and the OSPI. Thorough consideration was given to the number of levels on the access point continuum to ensure accessibility to a wide range of the intended students while maintaining fidelity to the knowledge and skills of the academic content, even at the lowest levels of complexity.

Drafting of the access points in science followed that of ELA and mathematics. A small group of Washington stakeholders, including both content experts and special education panelists, reviewed the *Washington's K–12 Science Learning Standards* and then extended each selected standard into access points with three levels of complexity. This activity was facilitated by the OSPI in February 2014 in Olympia, Washington. Once the access points were drafted, the OSPI and Measured Progress reviewed the draft documents and edited as necessary for clarity and consistency in language and expectations and for vertical alignment.

The access points went through multiple iterations of review during their development including review by educator committees comprising content experts, general educators, and special educators, as well as Local Educational Agency (LEA) administrators and OSPI staff. The review committees carefully considered issues of academic intent, accessibility, and bias and sensitivity.

Access Point Frameworks Committee Review

The OSPI coordinated recruitment of stakeholder committees that met for a two-day meeting in SeaTac, Washington, in February 2014. The intent of this meeting was to gather stakeholder input on the WA-AIM Access Point Frameworks. There were a total of 25 panelists that represented special and general education practitioners as well as districts and Educational Service Districts (ESDs). Content expertise was balanced between those who have a deep understanding of the special education population for whom this assessment is designed and those who have a deep understanding of the content and measurable skills and knowledge embedded in the standards. Panelists were chosen by the OSPI with the intention to remain consistent throughout subsequent development meeting(s) and provide consistency in the overall process and content interpretation.

OSPI and Measured Progress, panelists were presented with an overview of the development process, the format of the materials, and the intention of the committee work. Following the opening session, panelists moved into their assigned content work groups; using a standardized template, each group was asked to follow the same basic steps for their work.

Two expected outcomes were communicated to the work groups. The first outcome was that each content-specific group was to review the access points assigned to all grades (grades 3 through 8 and high school for mathematics and ELA; grades 5 and 8 and high school for science). The review focused on curricular congruence and alignment, developmental applicability for students with the most significant cognitive disabilities, and the consideration of the wide range of abilities of students with the most significant cognitive disabilities. The second outcome was for the groups to have in-depth discussions, responding to the access points designed to align to the content standards and making recommendations.

Step 1: Introductions and material review. The panelists in each content work group introduced themselves and indicated which region they were representing. A content specialist reviewed the expectations for their work and identified a note taker to record key points of their discussions and recommendations. Panelists were asked to familiarize themselves with the layout of the content standards documents and the already approved WA-AIM Blueprints, with the purpose of building foundational knowledge for review of the Access Point Frameworks.

Step 2: Review of content area access points. Using *Content Area Review Checklist*s as guides, each group considered the standards, CCEEs, and performance expectations being assessed and the corresponding access points. Groups began at the lower grades and worked through all grades over a span of two days. Specific review criteria were established in three main areas: accessibility, academics, and bias and sensitivity.

With the focus on accessibility, the following aspects were considered. The access points needed to provide three distinct levels of complexity, allowing for a wide range of learners to enter into the standard, with varying modes of communication. Another accessibility consideration

concerned the *Less Complex* access point. Groups were asked to evaluate whether these access points were the least complex that they could be while still being linked to the CCEEs.

Panelists were also asked to evaluate the academic strength of all access points, answering questions regarding the maintenance of academic intent through all three levels as well as the emphasis on academic learning.

Bias and sensitivity were also considered to ensure that the access points emphasized academic learning and not life experience, that they were age and grade appropriate, and that they would not contain non-curricular issues that may offend or dismay students, or district students from academic content.

Step 3: Review of the group work. Within each content area group, facilitating content specialists encouraged finalizing recommendations and then debriefed the review process prior to reconvening of the large group. The OSPI and Measured Progress facilitated a whole-group wrap-up session. The session summarized the work that was accomplished, outlined anticipated next steps, and discussed plans for an expanded June meeting to review the next stage in development. Following the work group meetings, an extensive review of the draft documents was conducted by the OSPI and Measured Progress.

2.2.3 Performance Task Development

While the Access Point Frameworks define the knowledge, skills, and understandings being assessed, the Performance Tasks measure actual student attainment of the skills and understandings. The Performance Tasks authoring was an iterative process between Measured Progress content and special education specialists and the OSPI. Each access point is coupled to an associated Performance Task, with each Performance Task constructed with the following components: Requirements, Item Examples, and Restrictions. Among these, Requirements and Restrictions are intended to standardize the assessment operation and administration of each Performance Task, and Item Examples are intended as teacher support to provide models of item development standardization.

Specifically, the Requirements outline the necessary and expected components of each Performance Task for a valid measurement. All Requirements must be met for a Performance Task to be an accurate measure of the access point. The provided Item Examples include administrator directions, stimuli, answer choices, graphics, and/or text. Though not affecting most Performance Tasks, Restrictions (i.e., specific constraints that must be upheld during the administration) are detailed, if needed; and specific manipulatives and tools that are prohibited are outlined, as such use would change the skill being assessed in the access point, thus invalidating the results to be interpreted from student scores. The comprehensive structure of the Performance Tasks was predicated on a commitment to provide the field with strong guidance, clear directions, and clearer standardization.

All examples and items presented in the Performance Tasks are allowed to be adapted to meet each individual student's learning style and preferred mode of receptive and expressive communication. Teachers are encouraged to present the Performance Task components in styles that most closely resemble how daily instructional materials are presented to the student. Typical adaptations and ideas are listed at the beginning of each grade-level set of Performance Tasks.

Figure 2.2 shows an example Performance Task from grade 6 mathematics. A complete set of the WA-AIM Performance Tasks are available at http://www.k12.wa.us/Assessment/WA-AIM/Frameworks.aspx.

Figure 2.2. Example Performance Task from Grade 6 Mathematics

6.RP.1 Ratio and Proportional Relationships-Understand ratio concepts and use ratio reasoning to solve problems — **Grade 6**

ACCESS POINT (More Complex) 6.RP.1.M. Student will generate a ratio based on a model or a real-world situation.

PERFORMANCE TASK

Requirements:

- Every performance task must have at least five unique items/questions.
- Task must include five items that are based on:
  - a model
  - a real-world situation
  - or a combination of both
- Ratios must have both numbers greater than or equal to 2.
- Given ratios should be part to part.

Restrictions:

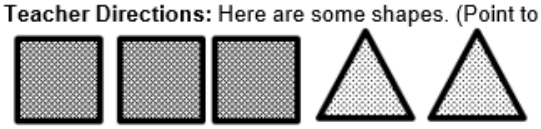
- Items must not be multiple-choice.

Note to Teacher: Both quantities used in the ratios MUST be greater than or equal to 2.

Example Items

Item 1:

Teacher Directions: Here are some shapes. (Point to the shapes.)



What is the ratio of squares to triangles? (Provide the student with a ratio template:___ to _____.)

Item 2:

Teacher Directions: There are seven students at a table. There are three girls and four boys. Here are some number cards. (Place number cards 1, 2, 3, 4, and 5 on the work surface.) Use these numbers to write the ratio of girls to boys. (Provide the student with a ratio template:___ to_.)

Answer Key (for teacher use only)
Item 1: 3:2
Item 2: 3:4

Performance Tasks Committee Review

The OSPI coordinated recruitment of stakeholder committees that convened for two days in June 2014 in SeaTac, Washington. A total of 38 panelists participated in the review, representing special and general education practitioners from districts and ESDs.

For this meeting, multiple groups were formed by grade spans and content areas. The work groups were assembled with the purpose of reviewing the Performance Task Requirements associated with each access point. The panelists who reviewed science were tasked with reviewing Performance Task Requirements written to the access points generated from the *Washington's K–12 Science Learning Standards*, while the panelists for ELA and mathematics worked with the CCSS and CCEEs. Each Performance Task was designed to measure an observable student action related to the specific knowledge, skills, and understandings from a target access point. Work groups edited and refined the draft Performance Tasks aligned to the access points.

Two expected outcomes communicated to panelists dictated the group work. The first outcome was for the work groups to review the Performance Tasks and focus on content centrality, performance fidelity, and developmental applicability for students with the most significant cognitive disabilities and to consider all variations of these factors among students in the expected population. The second outcome was for groups to have in-depth discussions about the Performance Task Requirements, and to make comments and final recommendations for edits and changes.

During the opening session facilitated by the OSPI and Measured Progress, panelists were presented with an overview of the process and the format of the materials. Following the opening session, panelists moved into their assigned content groups; using a standardized template, each group was asked to follow the same basic steps for their work.

Step 1: Introductions and material review. A content specialist reviewed the expectations for their work and identified a note taker to record key points of their discussions and decisions.

Panelists were asked to familiarize themselves with the approved WA-AIM Blueprints and Access Point Frameworks.

Step 2: Review of content area Performance Task requirements. Using a *Performance Task Requirements Review Checklist* as a guide, each group considered the access points and the corresponding Performance Task Requirements. The following areas were the focus of their review: accessibility, academics, and bias and sensitivity. Accessibility was addressed by evaluating whether at least one Performance Task allowed access for learning to be measured for a broad continuum of students, whether there were any specific accessibility concerns with any single Performance Task, and whether students using varying modes of communication could access the Performance Tasks. The academics criteria centered on whether the Performance Task Requirements related to the access point in terms of content and skills. Bias and sensitivity review took into account the need for the Performance Task Requirements to measure academic learning and not life experiences, be age and grade appropriate, and not contain any non-curricular issues that may offend or dismay students, or distract students from academic content.

Step 3: Review of the group work. Within each grade span and content area group, content specialists facilitated the compilation of final recommendations and debriefed the group regarding the review process prior to adjourning the meeting. Following the work group meetings, an extensive review of the draft documents was conducted by the OSPI and Measured Progress.

2.2.4 Standards Alignment

The WA-AIM Access Point Frameworks serve as conceptual linkage between the WA-AIM assessment and general education academic standards. As noted in Section 2.3.2, the WA-AIM Access Point Frameworks for mathematics and ELA are connected to the CCSS through CCEEs selected by committees of Washington educators. Similarly, the science Access Point Frameworks are mapped to the performance expectations of the EALRs, the foundations of the state science standards and the underpinnings of the general state science assessments, through the selection of Big Ideas by committees of Washington educators. These connections were

intended to provide access to general education standards for students with the most significant cognitive disabilities.

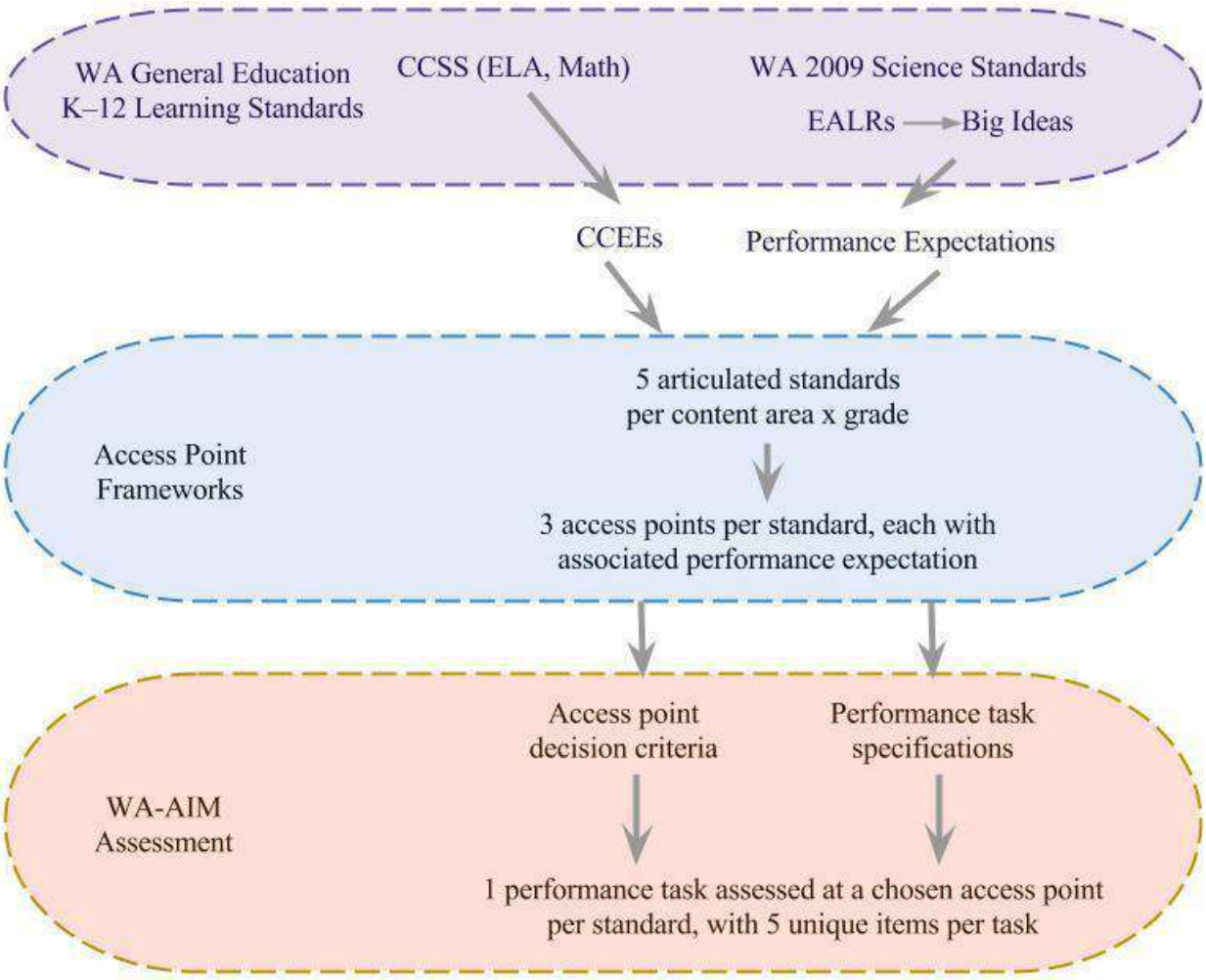
The selections of CCSS/CCEEs and Big Ideas were intended to provide a broad academic coverage of corresponding general education standards. Details about the selection process are presented in Section 2.3.1.

The selections resulted in a total of five standards (named "strands" in ELA, "domains" in mathematics, and "EALRs" in science) for each grade and content area in the Access Point Frameworks. Each standard was associated with a specific CCEE (in ELA and mathematics) or performance expectation (in science). The CCEE/performance expectation was then mapped by three access points (*Less Complex*, *Intermediate Complex*, and *More Complex*), with corresponding performance expectations articulated at each access point, which are intended to provide students with multiple entry points for accessing grade-level content. It should be noted that the *Intermediate Complex* access point was designed to be anchored to minimal mastery-level expectations from the general education standards.

The WA-AIM assessment requires measuring students on all five standards in the Access Point Frameworks for their given grade and content area. Depending on each student's needs and instructional entry point, a student could be assessed at varying access points across the five standards. A Performance Task with five unique items is expected to be used for assessment at each access point, with the task content and administration adhering to OSPI-standardized Performance Task requirements and restrictions.

Figure 2.3 summarizes the conceptual linkage between the general education standards, the Access Point Frameworks, and the WA-AIM assessment.

Figure 2.3. Conceptual Linkage Between Standards and Assessment in the WA-AIM



To empirically examine the linkage, an alignment study was conducted in late 2016 with twenty-two subject matter experts serving as panelists. The panelists assessed selected dimensions of alignment for each component of the WA-AIM assessment system, including the Washington State K–12 Learning Standards, CCEE/performance expectations, Access Point Frameworks, Performance Tasks, and a sampling of the assessments created by classroom teachers. Additional documents were reviewed as part of the experts' alignment considerations, including the *Educators' Directions for Administration*, *Generating the Score*, and the *Parent Guide*.

The considerations of the panelists were directed in two major areas. First, the panelists were directed to evaluate the overall alignment of the standards (CCEE/performance expectations) to the Washington State K–12 Learning Standards (Alignment A). Secondly, the panelists were to evaluate the overall alignment between the assessment, the Performance Tasks, and the Access Point Frameworks (Alignment C). Once these alignments were completed, panelists were to consider the overall alignment between the assessment and the standards (Alignment B).

The study found that out of 108 pre-defined criteria areas, 104 were rated with high alignment, 3 with moderate alignment, and 1 with moderate-to-low alignment. Details on the study method and results can be found in the Washington Access to Instruction and Measurement (WA-AIM) Final Alignment Study Technical Report that was submitted to the OSPI in January 2017.

Information regarding standards alignment for the new science assessments is presented in Section 2.3.1.

2.2.5 Alternate Achievement Level Descriptors

In November 2014, initial drafts of Alternate Achievement Level Descriptors (AALDs) were developed by a group of Washington general and special educators facilitated by Dr. Jan Sheinker and the OSPI. Because Washington had already developed Performance Task item examples for each access point, the access points guided the development of achievement descriptors for the alternate assessment. The stakeholder groups considered the alignment of the access points across and within grade levels, the clarity and conciseness of language, and the need for concise examples to be incorporated into descriptors. Because there were three access points and the OSPI wished to develop four achievement levels for the descriptors to be consistent with the general assessments, a Level 1 descriptor was added by the participants. The lowest level describes the parts of the Level 2 descriptor that the lowest group of students are likely to be able to do. Following the development of draft AALDs, the OSPI conducted an interdepartmental review of the drafts. The review included representatives from various divisions within the agency, including Teaching and Learning, Title One Federal Programs, and Migrant Bilingual Education.

In July 2015 at the conclusion of standard setting, the AALDs were reviewed and finalized based on the experience and outcomes of the standard-setting evolution. Immediately following administration of the WA-AIM, the AALDs had been used in a Teacher Achievement Level Study (TAL) where teachers estimated into which achievement level their students would be placed. Some members of the standard setting panel participated in the TAL. The standard setting panel worked with the AALDs during the profile-sorting process at the standard setting meeting. The entire standard setting panel was then asked to record comments that affected the clarity and usefulness of the AALDs in standard setting and identify any needed adjustments to the AALDs, given standard setting panel decisions in the profile-sorting process. The recommendations were used by a subgroup at the end of the standard setting evolution to craft proposed revisions for OSPI consideration.

Dr. Sheinker examined the recommendations for alignment with the CCEEs (for ELA and mathematics) and performance expectations (for science), access points, Performance Task item examples, and standard setting outcomes. She made recommendations to the OSPI to accept the proposed revisions or cited reasons for the OSPI to reconsider the proposed revisions. In most cases, Dr. Sheinker recommended accepting the proposed revisions from the standard setting panel and subgroup except where doing so would contradict other elements of the assessment system and final outcomes of standard setting, thus misrepresenting the achievement of students within each achievement level. The OSPI conducted an interdepartmental review of the proposed revisions and consultant recommendations, and the final descriptors were approved by the OSPI.

In May 2017, following administration and scoring of the first administration of the WA-AIM high school science assessment, DRC personnel conducted a standard setting to set cut scores and determine final AALDs. As in the previous standard setting, some members of the standard setting panel had experience using the draft AALDs in a TAL conducted prior to the standard-setting meeting. The TAL asked educators to estimate into which achievement level the students to whom they had just administered the WA-AIM high school science assessment would be placed.

Dr. Sheinker and the AALD Review Panel, in this case the full standard setting panel, used the same process previously described to refine the high school science AALDs per standard-setting results. Using the same process as described above, Dr. Sheinker examined the score distributions within each achievement level based on the final cut scores determined by the standard setting panel. Given the distribution of scores for each performance expectation across achievement levels, Dr. Sheinker did not recommend the movement either higher or lower of any achievement descriptor from the draft AALDs. Dr. Sheinker asked the AALD Review panel to consider what they had seen in their own profile sorting and the standard-setting information that resulted in the final cut scores. Panelists concurred that no movement of descriptors was warranted.

Some concerns were raised with regard to the vertical alignment of the performance expectation for the "System" domain between grade 5 and high school. Some panelists suggested that additional or revised Performance Task Item Examples that better represent the intended difference in difficulty between the expectations at the two grade levels could mitigate the concern. No change in the language for the descriptor was agreed upon, in part due to the constraints of the access points from which the descriptor emerged. Most panelists agreed that the difference could be better represented in revised Performance Task Item Examples than in changes to descriptor language.

Panelists felt that some descriptors required clarification of language to more accurately reflect what students did. They provided specific recommendations for refinement of the AALDs. Dr. Sheinker concurred with the panelists' recommendations and had no additional recommendations for refinements. After reviewing the recommendations, the OSPI accepted all proposed refinements. The OSPI also indicated that the recommended adjustments to the supporting information for Performance Tasks would be made to better represent the intended differences in difficulty between achievement at grade 5 and high school. The final descriptors are those approved by the OSPI.

Together, the system of standards and descriptors of the WA-AIM is designed to allow students with the most significant cognitive disabilities to demonstrate progress toward performance

expectations in ELA, mathematics, and science. The AALDs at each achievement level describe what students within that achievement range demonstrate on the assessment.

A detailed description of the AALD development and its review committees can be found in an in-depth AALD report submitted to the OSPI.

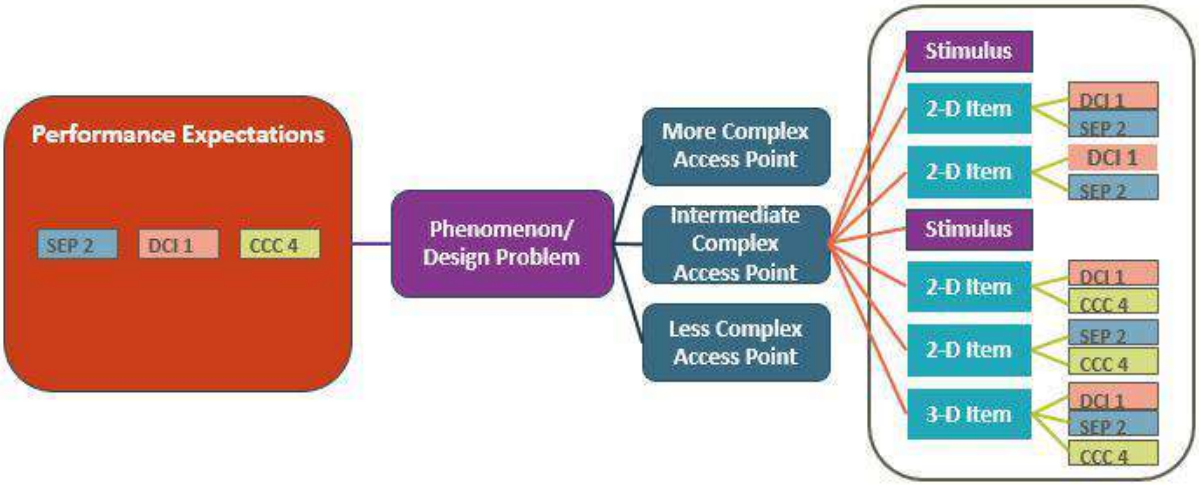
Information regarding the AALD development for the new science and high school ELA assessments is presented in Section 2.3.

2.3 WA-AIM Development: the 2017–2018 Administration

2.3.1 New Development for Science

During the 2016–2017 school year, the OSPI began development work on the new Science WA-AIM aligned to the NGSS. The overall development goals of the new Science WA-AIM were to adhere to the administration procedures of the current WA-AIM assessment system, but strategically incorporate the idea of multi-dimensional standards. To ensure adherence to the multi-dimensionality of the NGSS, the OSPI determined the new assessment should allow for item clusters aligned to a central phenomenon, as well as allow items to be grouped around stimulus. Figure 2.4 shows the overarching development framework of the Science WA-AIM.

Figure 2.4. WA-AIM Development Framework



The development process involved engagement with Washington educators who either worked with students with the most significant cognitive disabilities or general educators with experience in the new science standards, as well as feedback loops between the following OSPI divisions and offices: Assessment Development, Select Assessments, and Learning and Teaching.

The final WA-AIM Science Access Point Frameworks and Performance Tasks are available at http://www.k12.wa.us/Assessment/WA-AIM/Frameworks.aspx.

The development process and related meetings are outlined in Table 2.1.

Table 2.1. Science Development Activities

Date	Activity	Outcome
October 2013		Adoption of the Next Generation Science Standards
October 2016	Performance Expectation Selection	Group selected five Performance Expectations per grade band (3-5; middle school, high school)
December 2016	Access Point Development	Group developed an Essential Concept and three Access Points for each standard
March 2017	Performance Task Development	Group developed the Requirements, Restrictions, and Example Items for each of the Access Points
May 2017	Bias and Sensitivity Review	Group reviewed Access Points, Performance Tasks, Example Items, and supporting graphics for bias, sensitivity, and content issues.
January 2018	Alternate Achievement Level Descriptor Development	Group drafted AALDs for four reporting levels (1–4) at grades 5, 8, and 11
April 2018	Auditor training material review meeting	OSPI and vendor selected exemplars for use in Alternate Assessment Auditing.
July 2018	Achievement Level Setting and AALD Refinement Workshop	Group recommended cut-scores that define each of the four reporting levels and refined the AALDs.
August 2018	Alignment Study	Group made independent judgments on the alignment between various components of the WA-AIM
August 2018	OSPI Achievement Level Setting Recommendations presented to State Board of Education	SBE adopts OSPI recommendations

Performance Expectation Selection

The overall outcome of this activity was to have the group select the five NGSS Performance Expectations per grade level to be measured. Twenty-three educators from around the state participated in this activity, representing special education and general content educators with a background in the NGSS. The group was broken into three smaller grade band groups: elementary, middle school, and high school. The performance expectation meeting materials can be found in the 2017–2018 WA-AIM Technical Report.

Activity 1: General Orientation to the Next Generation Science Standards and WA-AIM

The group was taken through extensive training on the three dimensions of the NGSS. It was important for the group to have a foundational understanding that the new standards were now three dimensional, shifting focus away from content only standards of the past, and focusing on three dimensions: Scientific and Engineering Practices (SEP), Disciplinary Core Ideas (DCI), and Cross Cutting Concepts (CCC).

The OSPI also presented foundational information on the WA-AIM assessment design, and learner characteristics of students taking the WA-AIM.

Activity 2: Prioritization of Scientific and Engineer Practices and Cross Cutting Concepts

Each educator group was asked to come to consensus prioritizing both the SEPs and CCCs each grade band group felt were important in science instruction while considering the student population who participates in the WA-AIM. Below are the eight Scientific and Engineering Practices identified as essential for all students to learn and describes in detail in the NGSS science framework (National Research Council, 2012):

1) Asking Questions (for science) and defining problems (for engineering)
2) Developing and using models
3) Planning and carrying out investigation
4) Analyzing and interpreting data
5) Using mathematics, information and computer technology, and computational thinking
6) Constructing explanations (for science) and designing solutions (for engineering)
7) Engaging in argument from evidence

8) Obtaining, evaluating, and communicating information

Table 2.2 outlines the SEPs prioritized by each grade band group.

Table 2.2. Prioritized SEPs

Elementary	Middle School	High School
1) Asking Questions and defining problems	2) Developing and using models	2) Developing and using models
2) Developing and using models	4) Analyzing and interpreting data	4) Analyzing and interpreting data
4) Analyzing and interpreting data	6) Constructing explanation and designing solutions	6) Constructing explanation and designing solutions

The National Research Council (2012) lists the following seven Crosscutting Concepts:

1) Patterns
2) Cause and effect: Mechanism and explanation
3) Scale, proportion, and quantity
4) Systems and system models
5) Energy and matter
6) Structure and function
7) Stability and change

Table 2.3 outlines the CCCs prioritized by each grade band group:

Table 2.3. Prioritized CCCs

Elementary	Middle School	High School
1) Patterns	1) Patterns	1) Patterns
2) Cause and Effect	2) Cause and Effect	4) Systems and systems models
3) Scale, proportion and quantity	4) Systems and systems models	7) Stability and change

Activity 3: Application of Priorities to the Disciplinary Core Ideas

The establishment of prioritized SEPs and CCCs was used to narrow the breadth of Performance Expectations available to be chosen for inclusion in the WA-AIM.

Table 2.4 models the application of this process.

Table 2.4. Applied Prioritization on Performance Expectations

		Science and Engineering Practices							
		1. Asking questions and defining problems	2. Developing and using models	3. Planning and carrying out investigations	4. Analyzing and interpreting data	5. Using mathematics...	6. Constructing explanations and designing solutions	7. Engaging in argument from evidence	8. Obtaining, evaluating, and communicating information
Crosscutting Concepts	1. Patterns		MS-ESS1-1		MS-ESS2-3 MS-ESS3-2 MS-LS4-1 MS-LS4-3 MS-PS1-2	MS-PS4-1	MS-LS2-2 MS-LS4-2		
	2. Cause and effect: Mechanism and explanation	MS-PS2-3	MS-LS3-2 MS-PS1-4	MS-ESS2-5 MS-PS2-5	MS-LS2-1	MS-LS4-6	MS-ESS3-1 MS-ESS3-3 MS-LS1-5 MS-LS4-4	MS-ESS3-4 MS-LS1-4	MS-LS1-8 MS-LS4-5
	3. Scale, proportion, and quantity		MS-PS1-1	MS-LS1-1 MS-PS3-4	MS-ESS1-3 MS-PS3-1		MS-ESS1-4 MS-ESS2-2		
	4. Systems and system models		MS-ESS1-2 MS-ESS2-6 MS-PS3-2				MS-PS2-1	MS-LS1-3 MS-PS2-4	
	5. Energy and matter. Flows, cycles, and conservation		MS-ESS2-4 MS-LS1-7 MS-LS2-3 MS-PS1-5				MS-LS1-6 MS-PS1-6 MS-PS3-3	MS-PS3-5	
	6. Structure and function		MS-LS1-2 MS-LS3-1 MS-PS4-2						MS-PS1-3 MS-PS4-3
	7. Stability and change	MS-ESS3-5	MS-ESS2-1	MS-PS2-2				MS-LS2-4 MS-LS2-5	
	None	MS-ETS1-1*	MS-ETS1-4		MS-ETS1-3			MS-ETS1-2	

In Table 2.4, any Performance Expectation tied to a non-prioritized SEP or CCC was removed from consideration for inclusion on the WA-AIM for that grade band.

Activity 4: Performance Expectation Selection

Once the priorities were applied, each grade band worked to choose the five standards to be used on the WA-AIM. Each group was given the directive that they must have at least one Performance Expectation from Life Sciences, Earth and Space Sciences, and Physical Sciences. Including a Performance Expectation from Engineering, Technology, and the Application of Science was left up to each grade band's judgment.

The final Performance Expectations are shown in the WA-AIM Test Blueprint in Table 2.5:

Table 2.5. WA-AIM Test Blueprint

Domain	DCI	Grade 5	Grade 8	High School
Life Sciences	1: From Molecules to Organisms	3-LS1-1		
	2: Ecosystems		MS-LS21	HS-LS2-5
	3: Heredity			
	4: Biology Evolution			
Physical Sciences	1: Matter and Its Interactions	5-PS1-1		HS-PS1-5
	2: Motion and Stability	3-PS2-3		
	3: Energy		MS-PS3-3	
	4: Waves and Their Applications			
Earth and Space Sciences	1: Earth's Place in the Universe	5-ESS1-2	MS-ESS1-1	
	2: Earth's Systems		MS-ESS2-6	HS-ESS2-2
	3. Earth and Human Activity			HS-ESS3-4
Engineering, Technology, and the Application of Science		3–5-ETS1-1	MS-ETS1-3	HS-ETS1-2

Activity 5: Draft Essential Concept

Once the Performance Expectations were selected, each grade band group broke into 2–3 person sub-groups to begin drafting Essential Concepts. Essential Concepts serve as the initial reduction in depth of the Performance Expectation outlining the key concept to be measured on the WA-AIM. The Essential Concepts were required to maintain the three dimensionalities of the original Performance Expectation, although allowed to be reduced in depth.

Products from this activity were then reviewed and revised by OSPI's alternate assessment team, science assessment development team, and science Learning and Teaching team.

Access Point Development

The overall outcomes of this activity were to have the group revise and finalize the Essential Concepts for each Performance Expectation, to develop the three Access Points- Less, Intermediate, and More for each Performance Expectation, and to draft the Requirements and Restrictions for each Performance Task. Twenty-two educators from around the state participated in this activity, representing special education and general educators with a background in the NGSS. Most of this group also participated in the Performance Expectation Selection meeting. The group was broken into three smaller grade band groups: elementary, middle school, and high school. The access point meeting materials can be found in the 2017–2018 WA-AIM Technical Report.

Activity 1: General Orientation

The general orientation served to remind the participants the key components of the Next Generation Science Standards, the WA-AIM design and administration processes, and learner characteristics of students who take the WA-AIM. Participants were also provided an overview of the Performance Expectation Selection Process and drafting of the Essential Concepts.

Activity 2: Finalize Essential Concepts

The large group was taken through the draft Essential Concepts and large group consensus was required to ensure each Essential Concept contained the key concept of the original Performance Expectation, while maintaining the three dimensionalities intended of the NGSS.

Activity 3: Draft Access Points and Performance Task Requirements and Restrictions

Each grade level group was tasked with building the Access Points which allow three differentiated levels of access into the Essential Concept aligned to the original Performance Expectation. Groups were required to begin with the Intermediate Access Point, as, by design, the Intermediate Access Point is designed to have the most direct alignment to the Essential Concept (and through design, direct alignment to the original Performance Expectation). Through facilitation by a table leader, each grade band group worked to write the Access Points, Requirements and Restriction for one of the five standards. These were then reviewed and critiqued in amongst the whole group. Once table leaders were confident in their group's

understanding of the desired outcomes and process to be followed, grade band sub-groups of 2–3 were assigned a different Performance Expectation to build out the Access Points, Requirements, and Restrictions for the remaining standards at the assigned grade-band.

Throughout the drafting process, cross-grade band groups were assembled to provide feedback and critique the drafts of other grade band groups. This also served as calibration of the whole group process.

Products from this activity were then reviewed and revised by OSPI's alternate assessment team, science development assessment development team, and science Learning and Teaching unit.

Performance Task Development

The overall outcomes of this activity were to have the group review and approve all Access Point Frameworks, the associated Performance Tasks for each Access Point, and develop a complete set of five example items adhering to the Requirements and Restrictions drafted during the Access Point Development activity. Twenty-seven educators from around the state participated in this activity, representing special education and general educators with a background in the NGSS. Most of this group also participated in the Performance Expectation Selection meeting. The group was broken into three smaller grade band groups: elementary, middle school, and high school. The performance task meeting materials can be found in the 2017–2018 WA-AIM Technical Report.

Activity 1: General Orientation

The general orientation served to remind the participants of previous development meetings, provide a refresher on the core components of the NGSS standards, WA-AIM administration, and the learner characteristics of students who take the WA-AIM.

Activity 2: Review and Finalize Access Point Frameworks and Associated Performance Tasks

An entire group review occurred for one selected Access Point Framework at each grade band. The general structure of the Access Point Frameworks was discussed. The group then discussed

questions and concerns. The large group was then put in grade level groups of elementary, middle school, and high school to follow a similar process as that of the large group review of the Access Point Frameworks. Grade level groups discussed questions and concerns then reported out to the larger group.

Activity 3: Performance Task Writing Training

To prepare the participants for writing example items which adhere to the Requirements and Restrictions of the Performance Tasks associated with each Access Points, the participants were trained on the following topics:

- Sharing of a Performance Task from the Regular Assessment
- Identifying a Phenomenon and Topic
- Meeting Requirements and Restrictions
- Item (any activity generating a student response) and question (multiple choice, multi-select, constructed response, etc.)
- Range of item/question types.
- Evaluation criteria

Activity 4: Writing Example Items

Participants were then placed into smaller groups of 2–3 specific to their grade band. Each pair/trio had at least one special education educator and one general educator with background in the NGSS. Each group brainstormed phenomena specific to their assigned Access Point Framework. The phenomena were shared with whole group to solicit feedback on the phenomena. Once phenomena were chosen for each Access Point Framework, groups began working on development of five example items for each Performance Task associated with each Access Point of the Access Point Framework. Groups were required to begin item drafting at the Intermediate Access Point, since this is the Access Point most directly linked to the Essential Concept (and through design, direct alignment to the original Performance Expectation). Once groups drafted their example items for the Intermediate Performance Tasks, the entire group reviewed, critiqued, and provided feedback. Groups then had work time to apply feedback and revise their Performance Task example items. Once the Intermediate Performance Task example items were set, the groups began to draft the Performance Task example items for the

More and Less Access Points. These also went through whole group review with time following for each group to apply necessary edits and revisions.

During this development meeting, the OSPI required their selected graphics vendor be in attendance. The purpose was so the graphic artists and the development panelists could collaborate during drafting of the example items to ensure supporting graphics were consistent with the items writer's visions and expectations.

Products from this activity were then reviewed and revised by OSPI's alternate assessment team, science assessment development team, and science Learning and Teaching unit.

Bias and Sensitivity Review

The overall outcome of this activity was to conduct a review of the new WA-AIM NGSS Science Assessments—at the elementary, middle school, and high school levels—by independent committees of WA educators and stakeholders for bias and sensitivity. Twenty-five educators from around the state participated in this activity, representing special education, general education, teacher mentors, parents, vision and hearing specialists, who also represented diverse ethnic and cultural backgrounds. No member of this group had participated in any of three previous development activities.

The group was asked to review all final Performance Tasks and make consensus judgments on the following criteria:

1. Content offensive to any group?
2. Content that is different or unfamiliar to different groups?
3. Language that might be offensive to any group?
4. Language and vocabulary that might be more familiar to some groups than others?
5. Language or content that may generate an emotional response by any group and interfere with the ability to demonstrate knowledge or understanding?
6. Material that reinforces stereotypes-language, images, social/occupational roles, and /or behaviors and characteristics?

7. Material that shows a lack of sensitivity to the way a group has been represented over time?
8. Language, content, or context that is not accessible or not widely familiar to any group?
9. Material that portrays one or more people with disabilities in a negative or stereotypical manner?
10. Material that addresses a wide range of abilities and skills, ensuring that students with diverse needs receive opportunities to demonstrate competence on the same standard?

Groups made judgements independently then debriefed as a grade level team discussing and coming to consensus on each criterion as they related to each of the following Performance Task elements:

- Requirements
- Restrictions
- Adaptations
- Teacher Directions
- Graphics
- Item 1
- Item 2
- Item 3
- Item 4
- Item 5

Results of the Bias and Sensitivity Review can be found in the 2017–2018 WA-AIM Technical Report.

Edits and revisions were made to any element that did not meet the Bias and Sensitivity criteria. During debriefing, table leaders recorded the group's suggestions for revision. These suggestions were reviewed and implemented. Any product requiring revision was then reviewed and revised by OSPI's alternate assessment team, science assessment development team, and science Learning and Teaching unit.

Alternate Achievement Level Descriptor Drafting and Refinement

The overall goal of this activity was to draft initial Alternate Achievement Level Descriptors for all accountability reporting levels (levels 1–4). Twenty-six educators from around the state participated in this activity, representing special education, general education, teacher mentors, parents, vision and hearing specialists.

After cut-score recommendations were solidified from the Achievement Level Setting, the meeting participants also reviewed the WA-AIM Science AALDs and made suggested refinements.

A complete description of this process and outcomes is presented in a Science AALD report as part of a stand-alone WA-AIM Standard Setting Technical Report.

Auditor Training Material Review Meeting

The overall goal of this activity was the OSPI and their Auditing vendor to come to agreement on teacher submitted evidence to be used during Alternate Assessment Auditing training. The purpose was to identify high and low anchors for each Access Point, and to select training and validation papers in preparing auditors for reviewing teacher submitted data.

The agenda and a sample of auditor training materials used in the review meeting can be found in the 2017–2018 WA-AIM Technical Report.

Achievement Level Setting

The overall goal of the Achievement Level Setting activities was to set new achievement levels that define the four reporting levels used for accountability. Spring 2018 was the first operational use of the new WA-AIM Science Assessment aligned to the NGSS. Twenty-one educators representing special educators and general educators engaged in the Profile Sorting process to determine final recommendations of appropriate cut-scores to be presented to the Washington State Board of Education for approval.

A complete description of the process and outcomes is presented in a stand-alone standard setting technical report. A brief summary of the achievement level setting (Standard Setting) activities is provided in Section 1.7.

OSPI Achievement Level Setting Recommendations presented to State Board of Education (SBE)

On August 9, 2018, the OSPI presented the recommendations from the Achievement Level Setting workshop to the Washington State Board of Education (SBE). The SBE approved the recommendations.

Science Alignment Study

The purpose of this study was to evaluate the degree to which the scores from the WA-AIM can be interpreted in relation to the standards. The Links for Academic Learning alignment method was utilized. A complete description of the alignment study and outcomes can be located in a stand-alone WA-AIM Science Alignment Study Technical Report.

2.3.2 Changes to High School ELA and Mathematics

In June 2017 the Washington State Legislature took action to move the high school English language arts and mathematics assessments administration from grade 11 to grade 10. (Refer to 28A.655.061 of the Revised Code of Washington.)

Once this action was signed into law, the OSPI began a review of the high school English language arts and mathematics WA-AIM assessment to determine if any adjustments to the assessments would be needed based on administration occurring at an earlier grade level.

The English language arts and mathematics WA-AIM is based on the Essential Elements (Maps, 2013) developed by the Dynamic Learning Maps consortium. To determine what changes needed to occur, the OSPI reviewed the current standards and Essential Elements measured on the WA-AIM to see if there were significant differences in the essential knowledge, skills, and abilities between grade 10 and 11 standards.

For math, both the Common Core State Standards and the related Essential Elements are grouped at the high school level by domain and are applicable to grades 9–12. Due to this, it was determined that the content of the high school Mathematics WA-AIM did not need to change, but it was decided to convene a panel of educators to conduct a math achievement level validation, which was held concurrently with science and high school ELA achievement level setting.

For English language arts, the Common Core State Standards and the related Essential Elements are banded for grades 9–10 and for grades 11–12. Federal legislation requires that students participate in grade-level assessments aligned to grade-level standards. Due to this requirement, along with the shift in grade-level administration, the OSPI determined that portions of the high school English Language Arts WA-AIM would need to be rewritten to meet the intent of the Federal legislation.

High School English Language Arts Development

The overall development goals of the new High School English Language Arts WA-AIM were to adhere to the administration procedures and protocols of the existing High School ELA WA-AIM while making as minimal change to content as necessary. The final WA-AIM High School ELA Access Point Frameworks and Performance Tasks are available at http://www.k12.wa.us/Assessment/WA-AIM/Frameworks.aspx.

Table 2.6 shows the overarching development framework of the high school ELA WA-AIM.

Table 2.6. English Language Arts Activities

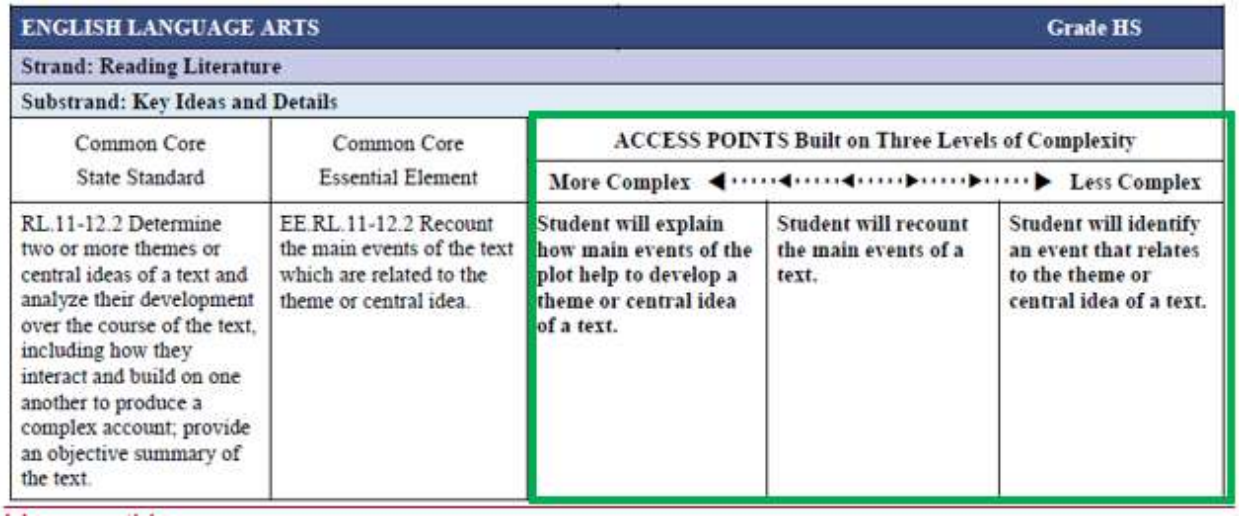
Date	Activity	Outcome
June 2017	Legislative Action on RCW 28A.655.061	Moved administration of ELA and mathematics assessments from Grade 11 to Grade 10
July/August 2017	OSPI Agency Review of CCSS and Essential Elements	Determined which HS ELA standards needed to have new Access Point Frameworks and Performance Tasks developed
October 2017	Access Point Development	Group adopted Dynamic Learning Maps Essential Elements and developed three Access Points for each standard
February 2018	Alternate Achievement Level Descriptor Development	Group drafted AALDs for four reporting levels (1–4) at grade 10
April 2018	Auditor training material review meeting	OSPI and vendor selected exemplars for use in Alternate Assessment Auditing.
July 2018	Achievement Level Setting and AALD Refinement Workshop	Group recommended cut-scores that define each of the four reporting levels and refined the AALDs.
August 2018	OSPI Achievement Level Setting Recommendations presented to State Board of Education	SBE adopts OSPI recommendations

OSPI Agency Review of CCSS, Essential Elements, and Access Point

The purpose of this activity was to determine the linkage between the measured grade 11 ELA strand and sub-strand Essential Elements and the corresponding grade 10 strand and sub-strand Essential Elements. This activity also involved a review of the Access Points developed to measure the 11–12 ELA standards to determine if any of the Access Points and their related Performance Tasks would still be viable in measuring the 9–10 Essential Elements.

The OSPI conducted a crosswalk between the 11–12 Access Point Frameworks and the correlated 9–10 standards to determine whether the previous Access Points would sufficiently measure the 9–10 standard.

In other words, would this:



ENGLISH LANGUAGE ARTS				Grade HS
Strand: Reading Literature				
Substrand: Key Ideas and Details				
Common Core State Standard	Common Core Essential Element	ACCESS POINTS Built on Three Levels of Complexity		
		More Complex ◀·····◀·····◀·····▶·····▶·····▶ Less Complex		
RL.11-12.2 Determine two or more themes or central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to produce a complex account; provide an objective summary of the text.	EE.RL.11-12.2 Recount the main events of the text which are related to the theme or central idea.	**Student will explain how main events of the plot help to develop a theme or central idea of a text.**	**Student will recount the main events of a text.**	**Student will identify an event that relates to the theme or central idea of a text.**

Measure this:

EE.RL.9-10.2 Recount events related to the theme or central idea, including details about character and setting.

The determinations and content by standard follow:

EE.RL.9-10.2 Recount events related to the theme or central idea, including details about character and setting. Previous Access Points would not fully measure the new standard.

EE.RI.9-10.3 Determine the logical connections between individuals, ideas, or events in a text. The OSPI content team had concerns with this standard. They felt the Essential Element

(EE) did not capture the true intent behind this standard as the EE lost the focus on author intent. Additionally, it was felt this was a complex skill for students without disabilities to learn, and even harder to measure on the regular assessment. Due to this reasoning, it was determined this standard would be replaced for the grade 10 WA-AIM.

EE.W.9-10.1 Write claims about topics or texts. a. Introduce a topic or text and write one claim and one counterclaim about it. b. not applicable c. not applicable d. not applicable e. not applicable. Previous Access Points would not fully measure the new standard.

EE.SL.9-10.4 Present an argument on a topic with logically organized claims, reasons, and evidence. Previous Access Points would not fully measure the new standard.

EE.W.9-10.7 Conduct research projects to answer questions posed by self and others using multiple sources of information. The state learning standard and the Essential Element for this standard have no language changes between the 9–10 and the 11–12 standard. Due to this it was determined the current Access Points fully measure the new standard. No adjustments were needed.

Table 2.7 shows the new High School ELA Standards and essential elements that would be applied to adjustments to the high school ELA WA-AIM assessment.

Table 2.7. New High School ELA Standards and Essential Elements

Washington K-12 Learning Standard	Essential Element
RL.9-10.2 Determine a theme or central idea of a text and analyze in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provide an objective summary of the text.	EE.RL.9-10.2 Recount events related to the theme or central idea, including details about character and setting.
RI.9-10.1 Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.	EE.RI.9-10.1 Determine which citations demonstrate what the text says explicitly as well as inferentially.

Washington K-12 Learning Standard	Essential Element
W.9-10.1 Write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence. a. Introduce precise claim(s), distinguish the claim(s) from alternate or opposing claims, and create an organization that establishes clear relationships among claim(s), counterclaims, reasons, and evidence. b. Develop claim(s) and counterclaims fairly, supplying evidence for each while pointing out the strengths and limitations of both in a manner that anticipates the audience's knowledge level and concerns. c. Use words, phrases, and clauses to link the major sections of the text, create cohesion, and clarify the relationships between claim(s) and reasons, between reasons and evidence, and between claim(s) and counterclaims. d. Establish and maintain a formal style and objective tone while attending to the norms of conventions of the discipline in which they are writing. e. Provide a concluding statement or section that follows from and supports the arguments presented.	EE.W.9-10.1 Write claims about topics or texts. a) Introduce a topic or text and write one claim and one counterclaim about it. b) not applicable c) not applicable d) not applicable e) not applicable
W.9-10.7 Conduct short as well as more sustained research projects to answer a question (including a self-generated question) or solve a problem; narrow or broaden the inquiry when appropriate; synthesize multiple sources on the subject, demonstrating understanding of the subject under investigation.	EE.W.9-10.7 Conduct research projects to answer questions posed by self and others using multiple sources of information.
SL.9-10.4 Present information, findings, and supporting evidence clearly, concisely, and logically such that listeners can follow the line of reasoning and the organization, development, substance, and style are appropriate to purpose, audience, and task.	EE.SL.9-10.4 Present an argument on a topic with logically organized claims, reasons, and evidence.

Access Point Framework and Performance Task Development Meeting

The OSPI led a group of seven expert teachers in special education and English language arts to draft the Access Points for More, Intermediate, and Less for each new standard on the assessment.

Activity 1: General Orientation

The group was taken through a brief orientation about why a change was needed, background information on the WA-AIM assessment design, and learner characteristics of the students who take the WA-AIM

Activity 2: Access Point Refinement and Confirmation

In groups of two or three, the participants began with the Access Point Frameworks based on the 11–12 standards with the CCSS and Essential Elements replaced with the corresponding 9– 10 CCSS and Essential Element. Participants used the former Access Points as a launching off point for development of new Access Points to measure the 9–10 standards. Since the Essential Elements for most standards were prerequisite skills to the 11–12 standards, most groups were able to refine and adjust the previous Access Points. Once each group had drafted the Access Points for their assigned standard, the whole group reviewed and worked toward consensus on the final product.

Activity 3: Performance Task Writing

In groups of two, participants began drafting Requirements and Restrictions for their assigned Access Points. Each group started with the Performance Task intended at the Intermediate Access Point. Once the Intermediate Level Requirements and Restrictions were drafted, the whole group reviewed the product and provided feedback. Groups revised Intermediate Level Requirements and Restrictions based the group feedback. This process was repeated for the More and Less Access Points Performance Tasks.

Activity 4: Item Writing

Once all Requirements and Restrictions for each Access Point were finalized, each group then wrote example items for all Access Point Performance Task. Like in Activity 3, the smaller groups began drafting at the Intermediate Access Point, a whole-group review was conducted, and the smaller groups applied revisions from feedback received. This process was repeated for the More and Less Access Point Performance Tasks.

Forms used in the sample review processes can be found in the 2017–2018 WA-AIM Technical Report.

After review for content and bias during the meeting, the OSPI WA-AIM, ELA Learning and Teaching, and ELA assessment development staff reviewed the final drafts.

Alternate Achievement Level Descriptor Drafting and Refinement

The overall goal of this activity was to draft initial Alternate Achievement Level Descriptors for all accountability reporting levels (levels 1–4). Nine educators from around the state participated in this activity, representing special education, and vision and hearing specialists.

After cut-score recommendations were solidified, the meeting participants also reviewed these WA-AIM High School English Language Arts AALDs and made suggested refinements.

A complete description of this process and outcomes is presented in the WA-AIM ELA high-school AALD report as part of a stand-alone WA-AIM Standard Setting Technical Report.

Auditor Training Material Review Meeting

The overall goal of this activity was the OSPI and their Auditing vendor to come to agreement on teacher submitted evidence to be used during Alternate Assessment Auditing training. The purpose was to identify high and low anchors for each Access Point and to select training and validation papers in preparing auditors for rating student submissions.

The agenda and a sample of auditor training materials used in the review meeting can be found in the 2017–2018 WA-AIM Technical Report.

Achievement Level Setting

The overall goal of this activity was to set new achievement levels that define the four reporting levels used for accountability. Spring 2018 was the first operational use of the new WA-AIM High School ELA Assessment aligned to the grade 9–10 CCSS and Essential Elements. Eight

educators representing special educators and general educators engaged in the Profile Sorting process to determine final recommendations of appropriate cut-scores to be presented to the Washington State Board of Education for approval.

A complete description of the process and outcomes is presented in a stand-alone standard setting technical report, and a brief summary of the achievement level setting (Standard Setting) activities is provided in Section 1.7.

OSPI Achievement Level Setting Recommendations presented to State Board of Education (SBE)

On August 9, 2018, the OSPI presented the recommendations from the Achievement Level Setting workshop to the Washington State Board of Education (SBE). The SBE approved the recommendations.

2.4 WA-AIM Development: 2019–2022

To reduce teacher burden and to standardize assessment content, OSPI contracted DRC for item development to provide educators with fifteen items for each standard and access point for use in the WA-AIM final assessment. Educators administering the WA-AIM are no longer allowed or able to create their own item content and all final assessments must use items developed through OSPI and DRC.

The bulk of the item content was deployed for use during the 2020–2021 WA-AIM administration. Some items were unable to be made available for the 2020–2021 administration but were made available for the 2021–2022 administration.

The item and test development process requires a cohesive development approach blending what may appear to be discrete processes into a single, seamless development cycle. Those discrete processes included the review of the *WA-AIM Access Points Frameworks and Performance Task*, item writing, item editing, passage and/or stimulus creation and passage adaptations, item reviews (by the OSPI and by Washington educators), and data reviews.

DRC's model for the WA-AIM development follows the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) since items are developed to reflect the range of cognitive ability inherent in the standards, resulting in reliable and instructionally valid tests. In addition, the item and test development process, as stated, adheres to the Principles of Universal Design, and it reflects a clear understanding of how items and test forms must lend themselves to accessibility by diverse groups of students and must function appropriately across a broad range of test administration accommodations.

The following sections provide a summary of the major (new) item and test development activities that occurred to develop the WA-AIM that were administered through 2022. Table 2.8 provides the development timeline.

This section also provides information regarding how DRC item and test developers engaged Washington educators in the process and followed rigorous procedures to develop and subsequently select items to be administered on the WA-AIM ELA, mathematics, and science assessments.

Development Timeline

Specific item and performance task development activities can be found in Table 2.8. Information regarding each step in the development process can also be found in the sections that follow.

Table 2.8. Development Timeline for the *WA-AIM* Administered in 2020–2022

Process Task	Description	Period
1. Item and Test Development Planning Meeting	DRC met with the OSPI to confirm understanding of the scope and timeline for all stages of the item and test development process.	July 24-25, 2019
2. Review Item Specifications	DRC met internally to review the guiding documentation related to the *WA-AIM*; DRC developed assessment documents for each grade and content area.	Fall 2019
3. Passage Development Specifications	Passage specifications were developed and included passage length (word count) and readability guidelines	March, 2020
4. Item Writing	Items were written by DRC and entered into the Washington Item Banking System (IDEAS).	Fall 2019-Summer 2021
5. Editorial and Content Review of the Items, and Graphics Creation	DRC item and test development specialists and editors, including bias, fairness, and sensitivity experts, reviewed and edited items as needed. Items were prepared for item review.	Fall 2019-Summer 2021
6. Item Review Meeting with Washington Educators	DRC facilitated the item review meeting with educators; items were reviewed for content, alignment to standards, bias, fairness, and sensitivity, etc. (Note: the item specifications were also reviewed, and suggested edits made as needed.)	November 7, 2019 April 6-8 & 13-15, 2020 August 3, 2020 October 5-13, 2020 June 29-July1, 2021 July 19-23, 2021
7. Items Selected for the Prebuilt Forms	DRC provided documents to enable OSPI to select the items for the Prebuilt forms.	Winter 2020
8. Administration of the 2021 *WA-AIM*	The 2021 testing started.	Spring 2021 (12/7/20–6/11/21)
9. 30 Math Items Developed and placed in the WA-AIM System	3.NF.1; 15 items –Intermediate; 15 items–More. 30 Items reviewed by Washington educators. Items added to the WA-AIM system in December 2021.	12/6/21–12/17/21
10. Administration of the 2022 WA-AIM	The 2022 testing started.	Spring 2022 (1/31/22–5/6/22)

Process Task 1: Item and Test Development Planning Meeting

Prior to the start of any WA-AIM item and development work, DRC's item and test development staff met with the OSPI to discuss the (new) item development plans for the WA-AIM, including the review of item specifications, and the development of the testing plan. The meeting included plans for the complete development cycle (e.g., review item specifications; item review; construction of pre-built forms).

Process Task 2: Development of Test Designs, Blueprints, and Draft Item Specifications

A critical part of the evidence that supports the use of the WA-AIM for its intended purposes is based on test content and the extent to which the content domain is represented in the test. According to the Standards, content-based evidence "can include logical or empirical analyses of the adequacy with which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores" (AERA, APA, & NCME, 2014, p. 14). Hence, documentation of the content domain, how the content is sampled and represented, and the alignment of items to the content must be well articulated.

The first steps in the development of the WA-AIM involved the review of item specifications (*WA-AIM Access Point Frameworks and Performance Tasks*) for each grade and content area of the WA-AIM. The test item specifications, including style guides, served to guide the entire item and test development process and provided consistency throughout the development of the WA-AIM.

Item Specifications

DRC item and test development staff also created draft item specifications and style guides to guide the item development for the WA-AIM Assessment. The item specifications were reviewed by the OSPI and DRC prior to item development. The item specifications for each grade and content area included the following:

- *WA-AIM Access Point Frameworks and Performance Task* domains for which items would be reviewed
- *WA-AIM Access Point Frameworks* Requirements and Restrictions reviewed
- Assessment limits and content constraints

- Range and balance of the items to be developed, including but not limited to the following: *WA-AIM Access Point Frameworks* Requirements and Restrictions
- Specifications for each item type, including artwork and graphics specifications
- Information regarding item/technical quality, including style
- Other, as required by the program

Process Task 3: Passage Development Process

Selecting and Adapting Passages

OSPI researched and selected permissioned passages that were to be used for the WA-AIM. OSPI sent a digital copy of each passage with notes about specific sections and/or pages including such information as starting and ending points and the access point(s) that the passage was to assess. DRC edited the passage to meet the WA-AIM passage specifications, including making adaptations. Adaptations made would allow accessibility to students with significant cognitive disabilities while maintaining the author's original message. Adaptations could include shortened text without the use of ellipses and/or brackets for deleted or changed text, vocabulary word replacement to lower the readability level, simplified sentences, addition of graphics to support comprehension, etc. DRC returned each adapted passage to OSPI for approval. Once DRC received approval of each passage, item writing could commence.

Approved Passage Specifications

Passage Specifications		
GRADE	**PASSAGE LENGTH**	**READABILITY**
3	25–50 words	.5–1.0
4	50–75 words	1.5–2.0
5	50–75 words	2.5–3.0
6	75–100 words	3.5–4.0
7	75–100 words	4.5–5.0
8	100–150 words	5.5–6.0
10	100–150 words	7.5–8.0

Passages must

- be clear and of fine quality,
- be engaging,
- include a title,
- be rich in content to support well-developed questions,
- include art as needed,
- must be of grade-level interest while maintaining the appropriateness of students with significant cognitive disabilities, and
- be free of sensitivity and bias issues.

Process Task 4: Item Writing

The items of the WA-AIM program in all grades and content areas were written by DRC item and content specialists who have experience in writing items for alternate assessments. The items were written to cover a range of subject matter and a range of difficulty, with the goal of meeting the numbers of items requested by OSPI (15 items per access point and standard).

The DRC item writers are trained on how to write items to meet quality expectations, including how best to write items to adhere to the Principles of Universal Design and to be free of issues of bias, fairness, and sensitivity. DRC's item and test development staff developed the training materials and conducted the training with its staff. Their training included a brief introduction of the purpose of the WA-AIM, including preliminary information regarding the test designs and the blueprints. Their training also included a presentation of the draft item specifications. During the training, examples of items were also provided. It has been the experience of DRC's item and test development staff that educators need to be aware of the reasons items might be rejected. Staff members writing the items were provided with item writing templates, the *WA-AIM Access Point Frameworks*, and other supportive materials needed to complete the task of writing items. Other supportive materials included, for example, information regarding how to avoid issues of bias, fairness, and sensitivity, and information regarding how to best adhere to the Principles of Universal Design.

When writing the items, the item writers used an item-writing template generated with WA-AIM Item Cards prepared with DRC-proprietary software. The item-writing templates includes codes to identify the content area, standard being measured, grade level, content category, item type and answer key (for multiple-choice items).

Using the item-writing templates, all items written by item writers were automatically entered into the DRC Item Development and Educational Assessment System (IDEAS), a comprehensive, secure online item banking system. IDEAS accommodates item writing, item viewing and reviewing, and item tracking and versioning for the Washington WA-AIM. DRC's item development staff used IDEAS to manage the transition of each item from its developmental stage (initial writing by the writer) to its approval for use. The system supports an extensive item history that includes item form, item-level notes, and content domain categories and subcategories.

Process Task 5: Editorial and Content Review of the Items, and Graphics Creation

As part of the WA-AIM item development process, each item was also reviewed by senior-level item and test development content specialists and editors at DRC. Item and test development specialists and editors evaluated each item to make sure that it measured the intended *WA-AIM Access Point Frameworks*. They also assessed each item to make certain that it was appropriate for the intended grade and that, if relevant, it provided and cued only one correct answer. In addition, the difficulty level, other features such as graphics, language demand, and distractors were also evaluated. Other elements considered in this internal DRC item review process included, but were not limited to, adherence to the Principles of Universal Design and freedom from issues of bias, grammar/punctuation, and technical quality. Adherence to the WA-AIM item specifications were also important considerations for the internal item reviews conducted by DRC senior-level alternate assessment specialists and editors. DRC consulted the OSPI regarding any general issues or concerns (e.g., style, format, interpretation of a given *WA-AIM Access Point Frameworks*) and about edits to specific items.

Note: Item writers adhered to the *WA-AIM Access Point Frameworks* as they drafted and revised items. Throughout the item development and review process, the alignment between each item and the associated standard was checked during each editing phase. All test items were carefully reviewed for content and style by DRC test development specialists. During all item reviews, careful attention was paid to verifying that each item measured the intended state-mandated *WA-AIM Access Point Frameworks.* If there was any misalignment, the item was edited to achieve greater alignment, or the item was realigned.

Graphics Creation

As a part of the development process and subsequent internal review of the items, DRC graphic specialists ensured that created item art could be reproduced clearly and accurately when items were displayed both in print and electronically. During this process, the item specifications and style guides were reviewed to identify any potential display requirements that may have presented challenges in the display environment. Display tolerances can be impacted by line thickness, percent screening for shading, and specialized fonts and symbols. These were defined in the early stages of the item and development process to help guide the delineation of style requirements and specifications.

For the WA-AIM at all grades and content areas, the item art was produced using vector graphics that allow for scalar adjustments without the breakdown of image clarity that is common with lower quality bitmapped formats. DRC's multitiered quality assurance process consisting of item and test development specialists, editors, and graphic artists makes certain item art is carefully compared to the original format or the original item throughout the item and test development and production process. The display of high-quality art in tests does not end with art production and the application of Universal Design principles. The medium for display and the conversion or transformation of the artist's work to this medium is also given careful consideration.

Process Task 6: Item Review Meeting with Washington Educators

All newly developed test items were submitted for review to the OSPI content team as well as to content committees consisting of Washington educators. The primary responsibility of the content committees was to evaluate the newly developed items with regard to quality and content classification, including, but not limited to, grade-level appropriateness, estimated difficulty, and source of challenge. “Source of challenge” issues refer to items where the cognitive demand is focused on an unintended content, concept, or skill (Webb, 2002). In addition, source of challenge may be attributed if the reason that an answer could be given results from a cultural bias, an inappropriate reading level, or a flawed graphic in an item, or it may be attributed if an item requires specialized knowledge outside the intended content to answer. Source of challenge could result in a student who has mastered the intended content or skill answering the item incorrectly or a student who has not mastered the intended content or skill answering the item correctly. Committee members were also asked to note any items with an issue related to source of challenge and to suggest revisions to remove the source of challenge issue. They also suggested revisions and made recommendations for reclassification or realignment of items. In some cases, when the committee recommended that an item not be assessed for a given reason, the committee was asked to suggest a replacement item and/or reviewed a suggested replacement item provided by the facilitators. The committee also reviewed the items for adherence to the Principles of Universal Design, including language demand, and issues of bias, fairness, and sensitivity.

Item reviews with Washington educators were held in 2019, 2020, and 2021. Committee members were selected by the OSPI. The meetings commenced with a welcome by the OSPI and DRC. This was followed by an overview of the test development process by DRC who also provided training on the procedures and forms to be used for content item reviews. See Appendix A for the item review training presentation.

DRC content item and test development specialists facilitated the reviews and were assisted by representatives from the OSPI. Committee members, grouped by grade level and content area, worked through, and reviewed the items for quality and content, including but not limited to the following considerations:

- Alignment to the given *WA-AIM Access Point Framework*
- Content Limits as guided by the item specifications (classified as Yes or No)
- Grade-Level Appropriateness (classified as At Grade Level, Below Grade Level, or Above Grade Level)
- Depth of Knowledge (classified as Recall, Skills and Concepts, or Extended Thinking)
- Correct Answer (classified as Yes or No)
- Quality of Distractors (classified as Yes or No)
- Graphics (classified as Yes or No) in regard to appropriateness
- Appropriate Language Demand (classified as Yes or No)

The members then came to a consensus and assigned a status to each item: Approved, Accepted with Revision, or Rejected.

As stated, members of the committees were also trained to review items for bias, fairness, and sensitivity and for adherence to the Principles of Universal Design. Each member noted bias, fairness, and/or sensitivity comments, if any, on the tracking sheets and on the item, if needed, for clarification. Committee members individually categorized any concerns as related to ageism, disability, ethnicity/culture, gender, region, religion, socioeconomic status, or stereotyping. These categories provided the framework through which recommendations for modification or rejection of items occurred during the subsequent committee consensus process. The committees then discussed each of the issues as a group and came to a consensus as to which issues should be presented to represent the view of the committee.

Process Task 7: Items Selected for Prebuilt Forms

OSPI was provided with documentation of items available for placement on Prebuilt Forms available for assessment. The items for these forms were chosen to meet the restrictions and requirements outlined in the *WA-AIM Access Point Frameworks*.

Process Task 8: Administration of the 2021 and 2022 Test Forms

The 2021 WA-AIM Spring testing window was available from December 7, 2020 through June 11, 2021. The WA-AIM Fall 2021 testing window started on September 27, 2021 and went through November 23, 2021.

The 2022 WA-AIM Spring testing window was available from January 31, 2022 through May 6, 2022. The WA-AIM Fall 2022 testing window started on September 26, 2022 and went through November 22, 2022.

Chapter 3. Test Administration

This chapter provides information on the test window, followed by a summary of the WA-AIM administration procedures and materials, as well as associated training materials and activities. Information on post-test assessment survey is also provided. The online WA-AIM system that has been utilized to facilitate the WA-AIM administrations is described, and major enhancements that have been implemented in the system are summarized. Afterwards, information on WA-AIM accommodations and adaptions is presented. Quality control measures regarding trustworthiness and security of collected data are discussed towards the end of the chapter.

3.1 Assessment Window

The WA-AIM assessment is structured as a baseline and final measure and is administered in a one-on-one setting using Performance Tasks, with student performance reported in an online customized WA-AIM system. Mathematics and ELA were assessed at grades 3 through 8 and 10; science was assessed at grades 5, 8, and 11.

The 2022 Spring Administration was scheduled from January 31, 2022 to May 6, 2022. District test coordinators were given permissions/access to the DRC INSIGHT Portal Registration and Student Performance Data Applications late in the day on January 28, 2022. (As a reminder, the Item and Form Management Application is available year-round for instructional and baseline assessment use.) The assessment protocol recommended that there be at least six weeks of instruction between the baseline measure Performance Task administration and the final measure.

In addition to the annual spring assessment window, the WA-AIM Fall 2022 administration/testing window was open to the Fall High School Retake(s) students. The WA-AIM Fall 2022 testing window started on September 26, 2022 and went through November 22, 2022.

3.2 Administration Procedures and Materials

The WA-AIM administration procedures are contained within a variety of source documents including the WA-AIM Access Point Frameworks, Performance Tasks, DRC INSIGHT Portal Users' Guide and the Test Administration Manual (TAM). The WA-AIM Access Point Frameworks provide instructional and planning guidance, and the Performance Task specifications set the parameters by which a student's knowledge and the skills identified in the Frameworks are measured.

Designed as a baseline and final measure, the WA-AIM assessment requires students to be assessed twice, with a recommended minimum of six weeks of instruction between the baseline and final Performance Task administrations.

The baseline measure serves to identify the appropriate instructional and assessment access point for each standard for each student. For each of the five standards in a content area, teachers make judgments as to the appropriate access point at which to assess each student on a standard-by-standard basis. This judgment is further refined with the requirement that if the student performance was >75% on the baseline measure at the *Less Complex* or *Intermediate Complex* access point, the baseline measure was to be re-administered at a higher access point, specifically the next level of complexity. The *Intermediate Complex* access point is considered the "anchor" or target. Beginning with *Intermediate Complex*, teachers are required to consider each student's prior knowledge of the concept as well as the student's skills.

An observer attestation is required to be completed by an educational professional (administrator, paraprofessional, educator, or service provider) who is not the teacher administering the assessment. The observer attestation is designed to strengthen the procedural validity of the assessment and is integral to administration of each Performance Task. The observer must observe the student performing the task and verify that the student independently generated the answers as documented in the performance scoring section.

DRC and the OSPI worked closely to deliver the 508-compliant WA-AIM DRC INSIGHT Portal User Guide, WA-AIM system contextual Help, and the WA-AIM Training Modules. OSPI

created the Test Administration Manual (TAM). The content of the TAM was organized in a logical way, providing detail and instructions for each step of administration. The DRC INSIGHT Portal User Guide is posted on OSPI's website https://www.k12.wa.us/student-success/testing/state-testing/assessment-students-cognitive-disabilities-wa-aim/access-point-frameworks-and-performance-tasks and was also included as a link within INSIGHT (General Information section), which facilitated user access to relevant training materials.

Training modules (YouTube videos) were developed to provide a step-by-step guide to accessing and navigating the WA-AIM system. The following modules are posted on the OSPI website (located at the link provided). The modules are specific to actions required by District and/or School Assessment Coordinators. https://www.k12.wa.us/student-success/testing/state-testing/assessment-students-cognitive-disabilities-wa-aim/trainings

Training Modules for District & School Test Coordinators

WAMS

- Pre-identification

INSIGHT – User Management

- Adding Single User
- Adding Multiple Users
- Updating User Permissions

Student Management

- Review Student Details and Student Demographics

Student Performance Data

- Monitor
- Progress/Status

Training Modules for Test Administrators

These modules are designed for educators who administer the WA-AIM to students.

General

These modules provide the necessary background information on the WA-AIM

- What are Alternate Assessments
- WA-AIM Participation Criteria

- WA-AIM Overview
- WA-AIM Components and Materials
- WA-AIM Access Point Frameworks
- WA-AIM Performance Tasks
- Engagement Rubric Eligibility

INSIGHT

The following modules help users navigate and use the DRC INSIGHT Portal.

- INSIGHT Overview
- Accessing INSIGHT Online Help

Student Management

- Review Student Details and Demographics

Item & Form Management

- Preview Forms
- Preview Items
- Creating Baseline and Instructional Forms
- Creating Final Forms
- Modifying

Registration

- Final Forms
- ER Forms
- Modifying Student Registration

Student Performance Data

- Complete a Standard Assessment
- Complete Engagement Rubric Form
- Completing a Student Characteristic Survey

3.3 Assessment Survey

With each WA-AIM testing window, a WA-AIM Teacher Feedback Survey is available. DRC and OSPI use the survey to gather feedback from Test Administrators who participated in the WA-AIM Administration. The survey is an instrument to gather feedback on experiences with the DRC INSIGHT Portal and WA-AIM system applications (Item and Form Management, Registration, Student Performance Data), manuals and training, online resources and DRC Customer Service support.

During the 2022 WA-AIM Spring administration, DRC sent two survey reminder emails (May 2, 2022 and May 9, 2022) to encourage those Test Administrators who may not have responded to the survey, to participate and provide feedback. Approximately 648 Test Administrators submitted their feedback. See Appendix C for the 2022 Spring WA-AIM Teacher Feedback Survey.

3.4 Assessment Data Collection

A new online WA-AIM system has been implemented since December 7, 2020. DRC customized the WA-AIM online system, using DRC INSIGHT (an online test and student management system), for three major administration activities:

1. Item and Form Management (INSIGHT Item Bank),
2. Registration, and
3. Student Performance Data.

These INSIGHT applications supported the WA-AIM administrations. District Assessment Coordinators (DACs) were pre-populated within INSIGHT using a master file provided by the OSPI. The account-creation process after this initial upload followed a tiered approach whereby district test coordinators created school coordinator accounts and school coordinators created teacher accounts. The maintenance of accounts was managed at the district and school levels. This role-based access to INSIGHT ensured that only the appropriate personnel had access to certain data and features, as requested by the OSPI.

In the online WA-AIM system, Test Administrators/Teachers can see only students that are enrolled in their schools within the system. Students eligible to take the WA-AIM must be identified through the Washington Assessment Management System (WAMS) by the District Assessment Coordinator (DAC). Students identified in WAMS are loaded to Student Management in the DRC INSIGHT Portal and display in Registration.

Item and Form Management (INSIGHT Item Bank)

Item and Form Management within the DRC INSIGHT Portal is where Test Administrators manage and create forms used for WA-AIM assessments.

OSPI and DRC partnered to create a minimum of one pre-built, Public Final form for each standard and access point. Teachers and Test Administrators may also create their own forms from the library of available items in Item and Form Management. Students must be registered for and administered five Final forms per Content Area, one per standard and access point.

Teachers may opt to use either pre-built Public, Final forms or Final forms they've created themselves for a student's final assessments, or a combination of pre-built and teacher-created Final forms. Users are not allowed to edit Public forms.

Final forms must consist of five items that fully measure the standard and access point following Performance Task Requirements and Restriction. Students are administered one Final form for each standard.

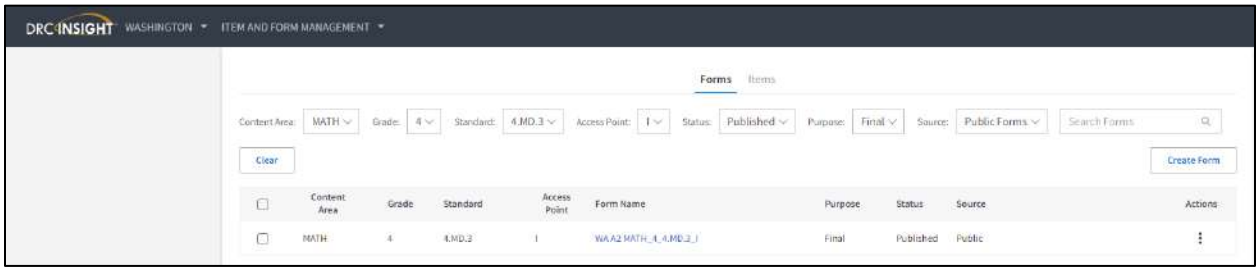
The number of forms registered to a student are as follows:

- Grades 3, 4, 7, and 10 = 10 total Final forms (5 ELA forms, 5 Math forms)
- Grades 5, 8 = 15 total Final forms (5 ELA forms, 5 Math forms, and 5 Science forms)
- Grade 11 = 5 Final forms (5 Science forms)

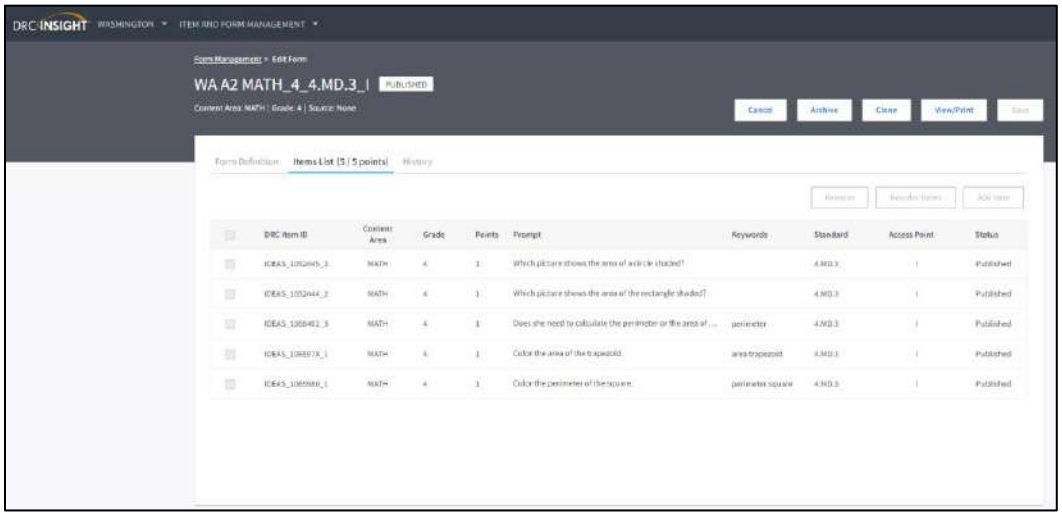
Test Administrators can view Published and Public forms before assigning them to a registration to determine whether or not the Form Purpose, content, and Items List is appropriate for assessing the student.

Test Administrators may find and view Published/Public forms by selecting from the system drop-down menu:

- Content Area, Grade, Standard, Access Point, Status, Purpose, and Source



Select the Form Name and the Test Administrator/User can see Form Definition and the Items List:



See Appendix B for an example of Final Public Form.

Test Administrators/teachers may also use stand-alone items available in Item and Form Management to create a teacher-created form. Teacher-created forms are submitted to hand-scoring to verify that the five items selected fulfill the Requirement of Performance Task. Pre-built forms do not require this hand-scoring process.

Registration:

Teachers/Test Administrators interact with the system to Register a student to a final assessment form. They select their site, create a unique Registration name for the form, add their student(s), add the form that they would like to administer to this student(s) (by grade, content area, standard, access point level) and select Register. A confirmation message within the system indicates that the Registration was successful. The registration appears under Registration Management. With the first registration for the student, a Student Characteristics Survey (SCS) form is also created automatically.

> The Student Characteristics Survey (SCS) is an informal questionnaire required to be completed for each student participating in the WA-AIM. This information is used to provide context for WA-AIM score interpretation when communicating results to various stakeholders. This information also helps OSPI identify trends and/or future supports. The information collected in the SCS DOES NOT impact a student's final score total or reporting level and is only used at an aggregate level. When completing the SCS teacher should use their best judgement at the time the SCS is being completed. It is understood that the information provided by the Test Administrator represents a single moment in time perception by the teacher and is not a definitive statement on the student.

The forms selected are now available in the Student Performance Data application as assessments, where Test Administrators enter student results after administering the Performance Tasks.

Test Administrators use the Registration application to associate students with Final forms. Students can be registered for pre-built Final forms, Final forms created by Test Administrators in Item and Form Management, or a combination of the two types of Final forms. Once registrations are created, Test Administrators use the Student Performance Data application to enter the student's assessment results.

Student Performance Data:

Student Performance Data is where Test Administrators enter student results for assessments associated with Registrations. Test Administrators must complete the Student Characteristics Survey (SCS) for each student before they can enter results for any other type of assessment.

Assessments are categorized as either Not Started, In Progress, or Ready to Submit. The list below describes the information regarding each of these statuses:

- **Not Started**
    - Displays a list of assessments associated with Registrations that Test Administrators have not yet started.
    - Once a user clicks Begin Assessment its status immediately changes to In Progress, even without responding to a question or saving or exiting.
- **In Progress**
    - Displays a list of assessments associated with Registrations that Test Administrators have started but have not yet completed.
    - Can be edited regardless of assessment type.
    - Once a user clicks Complete Assessment its status immediately changes to Ready to Submit.
- **Ready to Submit**
    - Displays a list of assessments that have been completed by the Test Administrator.
    - Student Characteristic Surveys in Ready to Submit status cannot be edited. All other assessment types can still be edited.
    - Assessments with a Ready to Submit status are scored at the end of the testing window. Users do not need to take additional action within INSIGHT to submit tests for scoring. All assessments are submitted to scoring by DRC at the end of the assessment testing window.

Test Administrators/Teachers enter in one response per item (5 items within the Form) as shown in Figure 3.1. They input the student's response (correct, incorrect, or no response) for each item on the form. Figure 3.2 shows the Adaptations drop-down selection by item, the input text

box for the Observer, and a Details Panel that is readily available for quick reference for the teacher (including the Teachers Directions, Access Point, Requirements and Restrictions).

Figure 3.1. DRC INSIGHT Student Performance Data

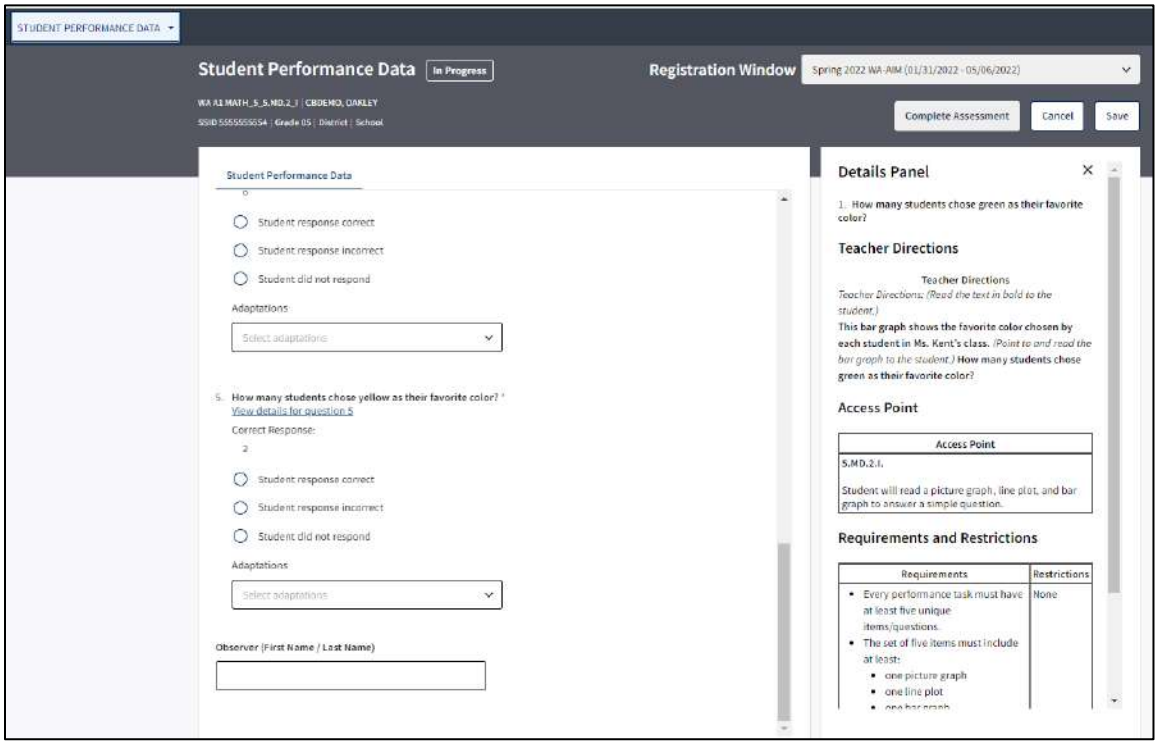
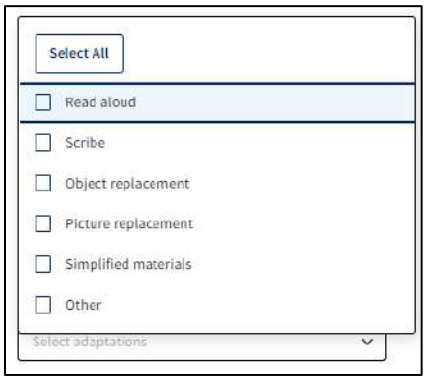


Figure 3.2. Adaptations Drop-down (by Item)



3.5 Major Enhancements to the WA-AIM System

The new WA-AIM online administration system that has been implemented starting in the 2020–2021 WA-AIM administration has four major enhancements, as described below.

3.5.1 Item library built-out

A total of 15 items per standard and access point are provided in the administration system from which teachers choose when building a form for a given performance task.

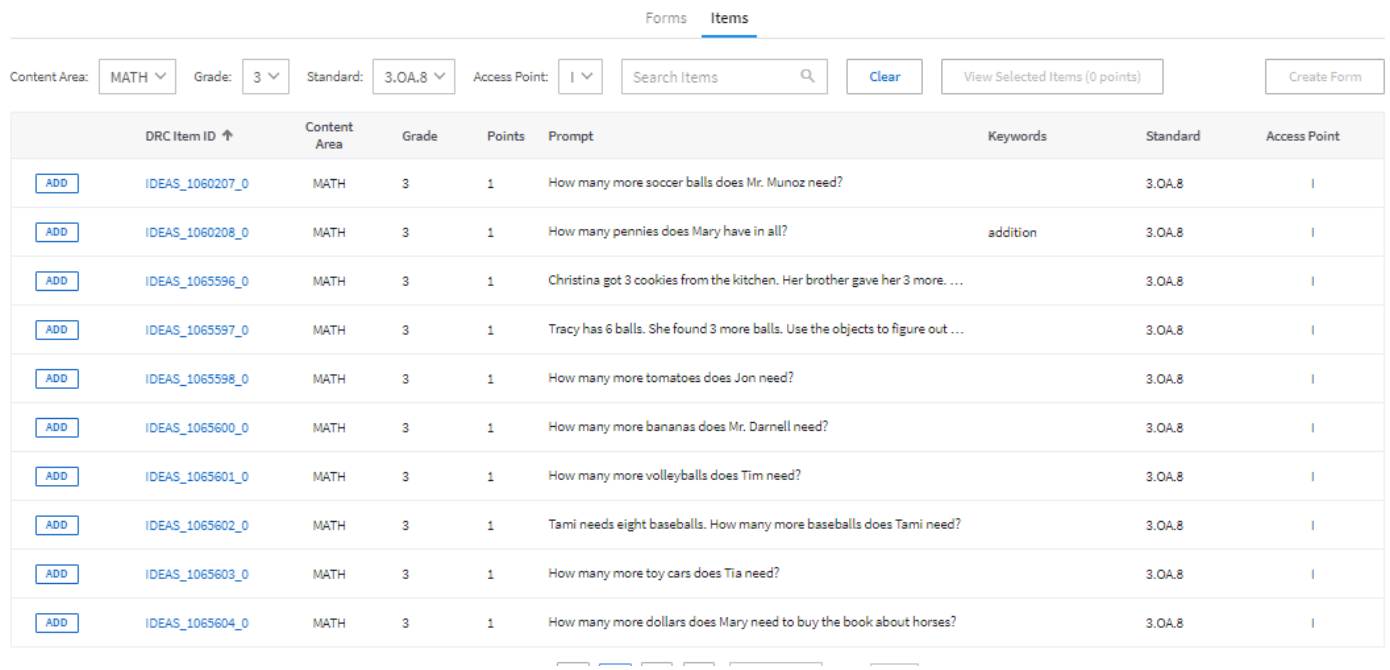
Overview: 15 items per standard and access point available for each performance task

Purpose:

- Address teacher created item content
- Standardize the item content

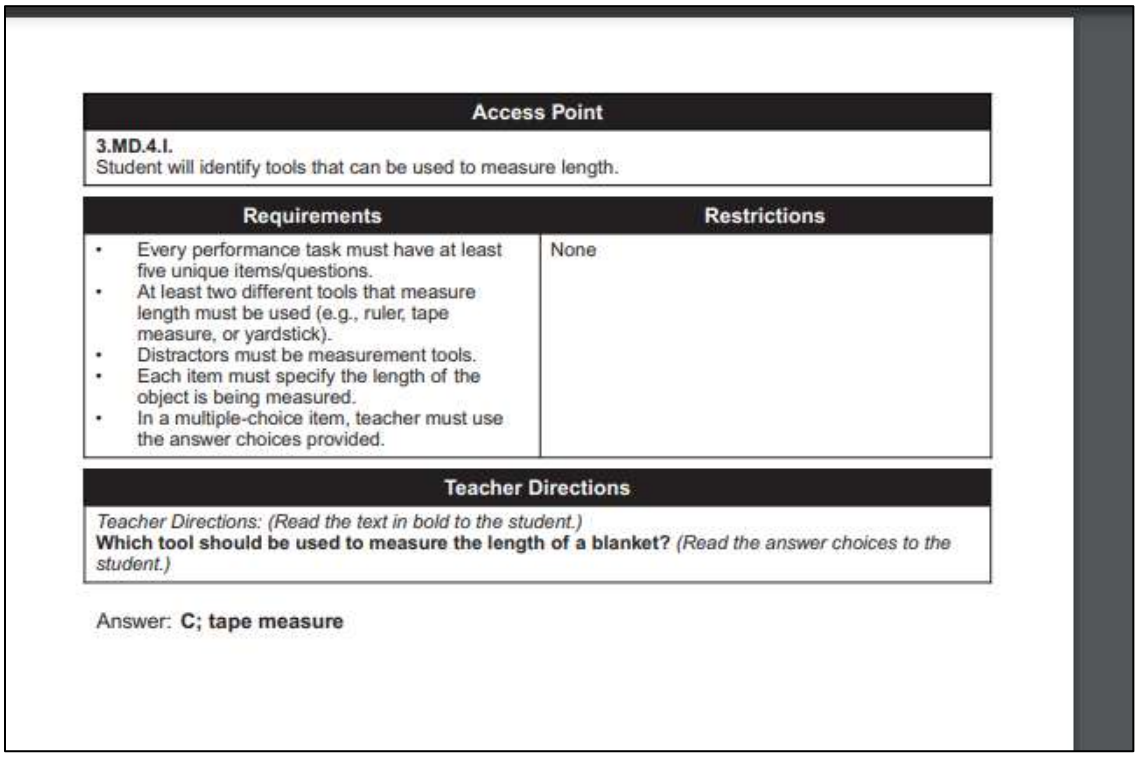
Example:

- Filter to find items



	DRC Item ID	Content Area	Grade	Points	Prompt	Keywords	Standard	Access Point
ADD	IDEAS_1060207_0	MATH	3	1	How many more soccer balls does Mr. Munoz need?		3.OA.8	I
ADD	IDEAS_1060208_0	MATH	3	1	How many pennies does Mary have in all?	addition	3.OA.8	I
ADD	IDEAS_1065596_0	MATH	3	1	Christina got 3 cookies from the kitchen. Her brother gave her 3 more. …		3.OA.8	I
ADD	IDEAS_1065597_0	MATH	3	1	Tracy has 6 balls. She found 3 more balls. Use the objects to figure out …		3.OA.8	I
ADD	IDEAS_1065598_0	MATH	3	1	How many more tomatoes does Jon need?		3.OA.8	I
ADD	IDEAS_1065600_0	MATH	3	1	How many more bananas does Mr. Darnell need?		3.OA.8	I
ADD	IDEAS_1065601_0	MATH	3	1	How many more volleyballs does Tim need?		3.OA.8	I
ADD	IDEAS_1065602_0	MATH	3	1	Tami needs eight baseballs. How many more baseballs does Tami need?		3.OA.8	I
ADD	IDEAS_1065603_0	MATH	3	1	How many more toy cars does Tia need?		3.OA.8	I
ADD	IDEAS_1065604_0	MATH	3	1	How many more dollars does Mary need to buy the book about horses?		3.OA.8	I

Teacher Material:



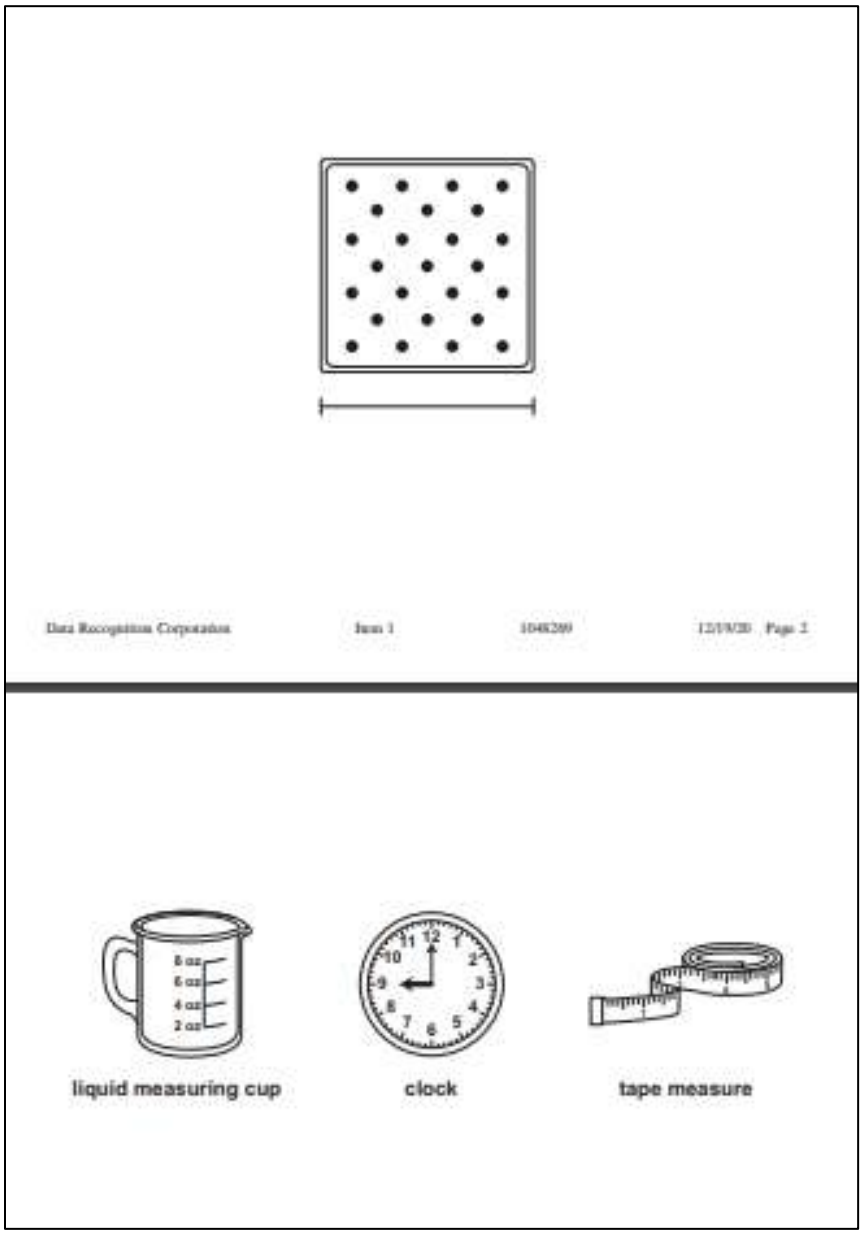
Access Point
3.MD.4.I. Student will identify tools that can be used to measure length.

Requirements	Restrictions
• Every performance task must have at least five unique items/questions. • At least two different tools that measure length must be used (e.g., ruler, tape measure, or yardstick). • Distractors must be measurement tools. • Each item must specify the length of the object is being measured. • In a multiple-choice item, teacher must use the answer choices provided.	None

Teacher Directions
Teacher Directions: (Read the text in bold to the student.) **Which tool should be used to measure the length of a blanket?** *(Read the answer choices to the student.)*

Answer: **C; tape measure**

Student Material:



3.5.2. Form based system

It is specified in the system that each measured standard must use one form with five items, and teachers can choose to use a pre-built form provided by the system or set up a teacher-created form using items already documented in the library. Any teacher-created form goes to handscoring to verify that the five items selected fulfill the Requirement of Performance Task. Pre-built forms do not require this handscoring process.

Overview: Each standard must use 1 form for measurement which contain the five items. There are two options:

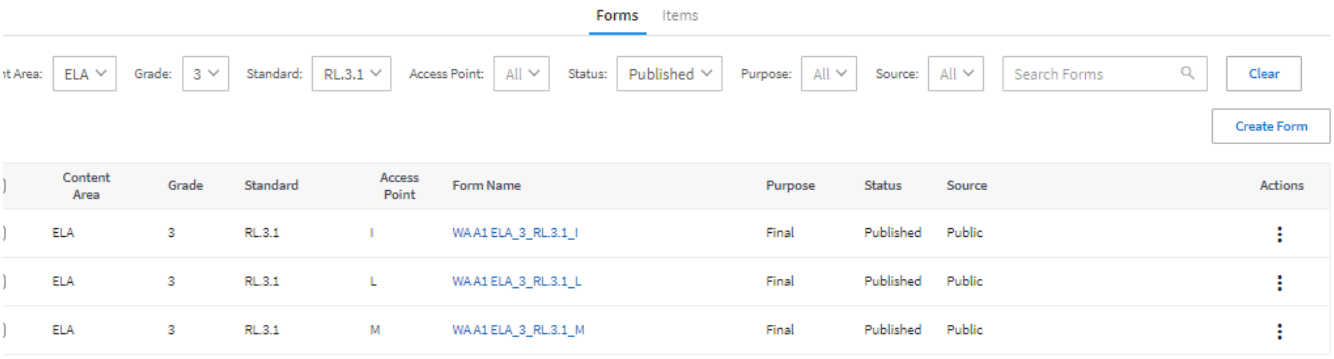
- Pre-built: Items selected to fulfill Requirements of Performance Task
- Teacher Created: Items from the library must be used. Goes to handscoring to verify the five items selected fulfill Requirement of Performance Task

Purpose:

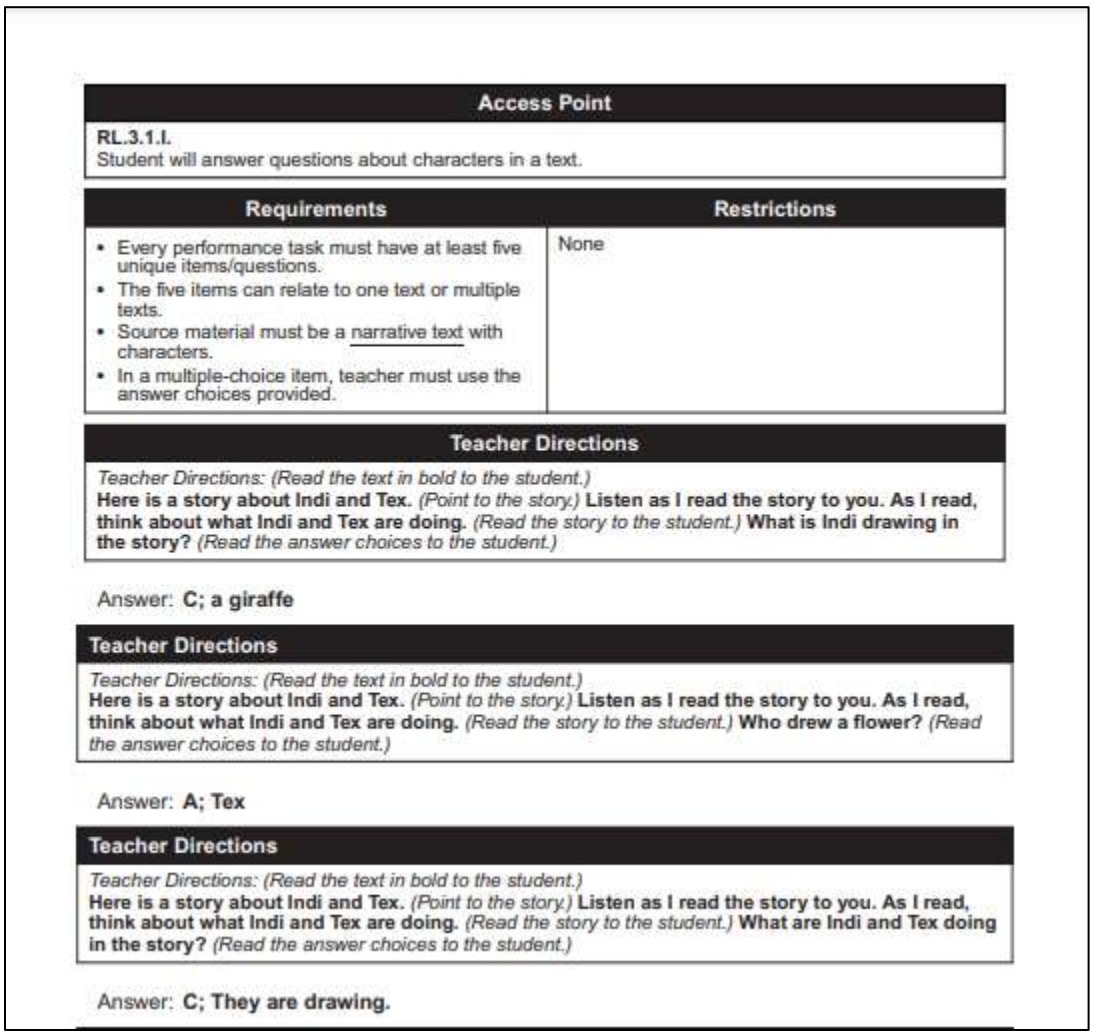
- Address item point value at an access point level (e.g. items worth multiple points)
- User ease
- Evaluate teacher preference

Example:

- Filter to find forms



Teacher Materials:



Access Point
RL.3.1.I. Student will answer questions about characters in a text.

Requirements	Restrictions
• Every performance task must have at least five unique items/questions. • The five items can relate to one text or multiple texts. • Source material must be a narrative text with characters. • In a multiple-choice item, teacher must use the answer choices provided.	None

Teacher Directions
Teacher Directions: (Read the text in bold to the student.) **Here is a story about Indi and Tex.** *(Point to the story.)* **Listen as I read the story to you. As I read, think about what Indi and Tex are doing.** *(Read the story to the student.)* **What is Indi drawing in the story?** *(Read the answer choices to the student.)*

Answer: **C; a giraffe**

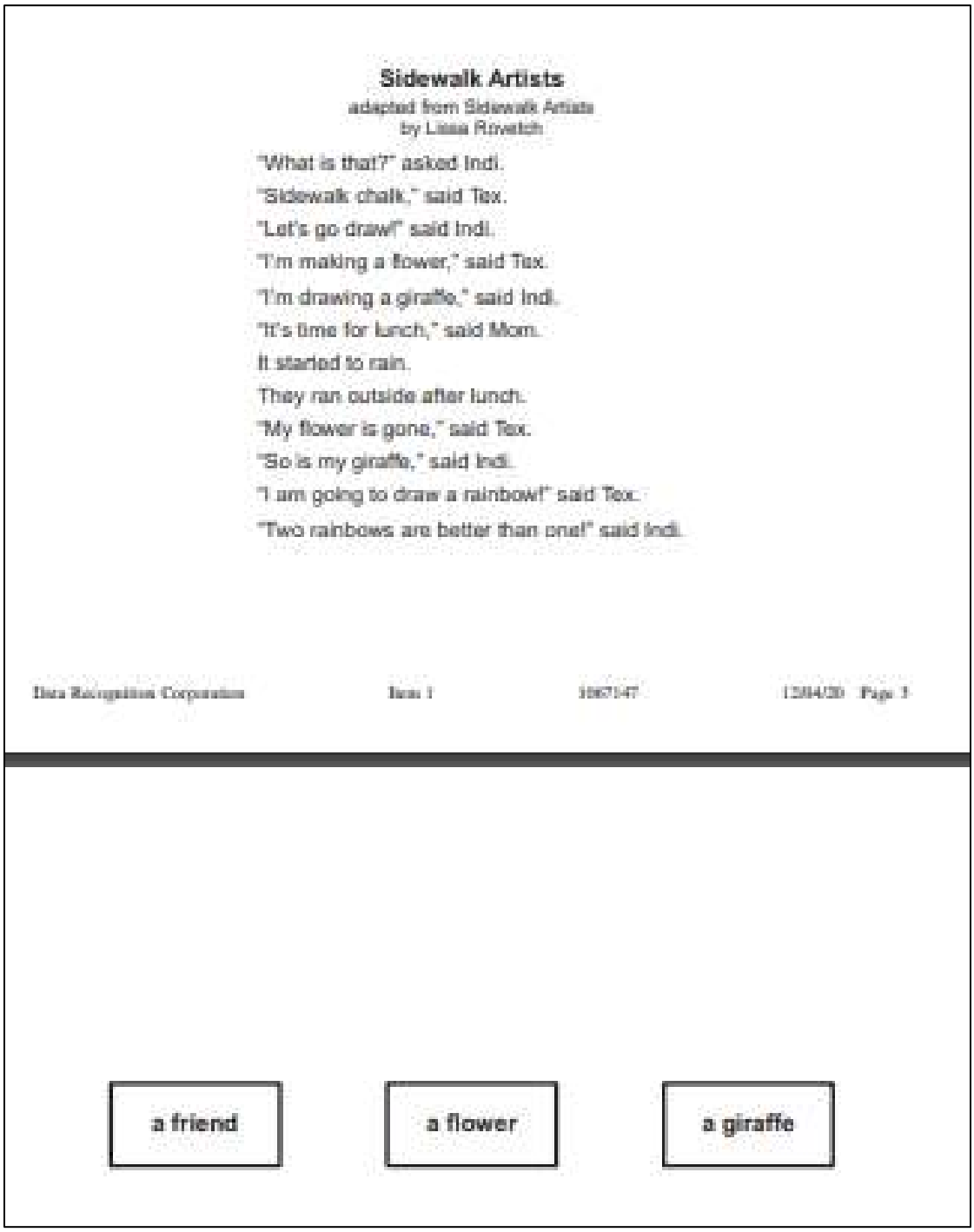
Teacher Directions
Teacher Directions: (Read the text in bold to the student.) **Here is a story about Indi and Tex.** *(Point to the story.)* **Listen as I read the story to you. As I read, think about what Indi and Tex are doing.** *(Read the story to the student.)* **Who drew a flower?** *(Read the answer choices to the student.)*

Answer: **A; Tex**

Teacher Directions
Teacher Directions: (Read the text in bold to the student.) **Here is a story about Indi and Tex.** *(Point to the story.)* **Listen as I read the story to you. As I read, think about what Indi and Tex are doing.** *(Read the story to the student.)* **What are Indi and Tex doing in the story?** *(Read the answer choices to the student.)*

Answer: **C; They are drawing.**

Student Materials:



Sidewalk Artists

adapted from Sidewalk Artists
by Lissa Rovetch

"What is that?" asked Indi.
"Sidewalk chalk," said Tex.
"Let's go draw!" said Indi.
"I'm making a flower," said Tex.
"I'm drawing a giraffe," said Indi.
"It's time for lunch," said Mom.
It started to rain.
They ran outside after lunch.
"My flower is gone," said Tex.
"So is my giraffe," said Indi.
"I am going to draw a rainbow!" said Tex.
"Two rainbows are better than one!" said Indi.

Data Recognition Corporation Item 1 1067147 1384020 Page 3

a friend

a flower

a giraffe

3.5.3. Entry of student performance data

All item content directly renders within the WA-AIM online administration system; there is no additional entry requirements for teachers.

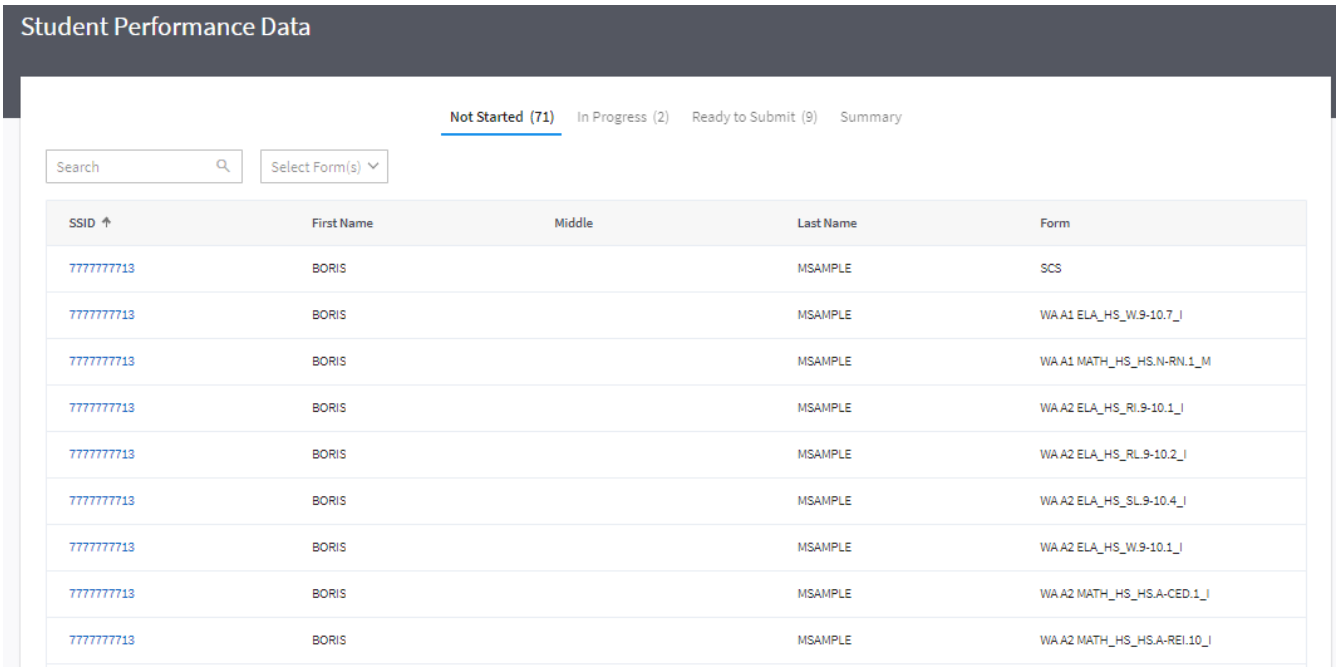
Overview: Item content directly rendered within INSIGHT system

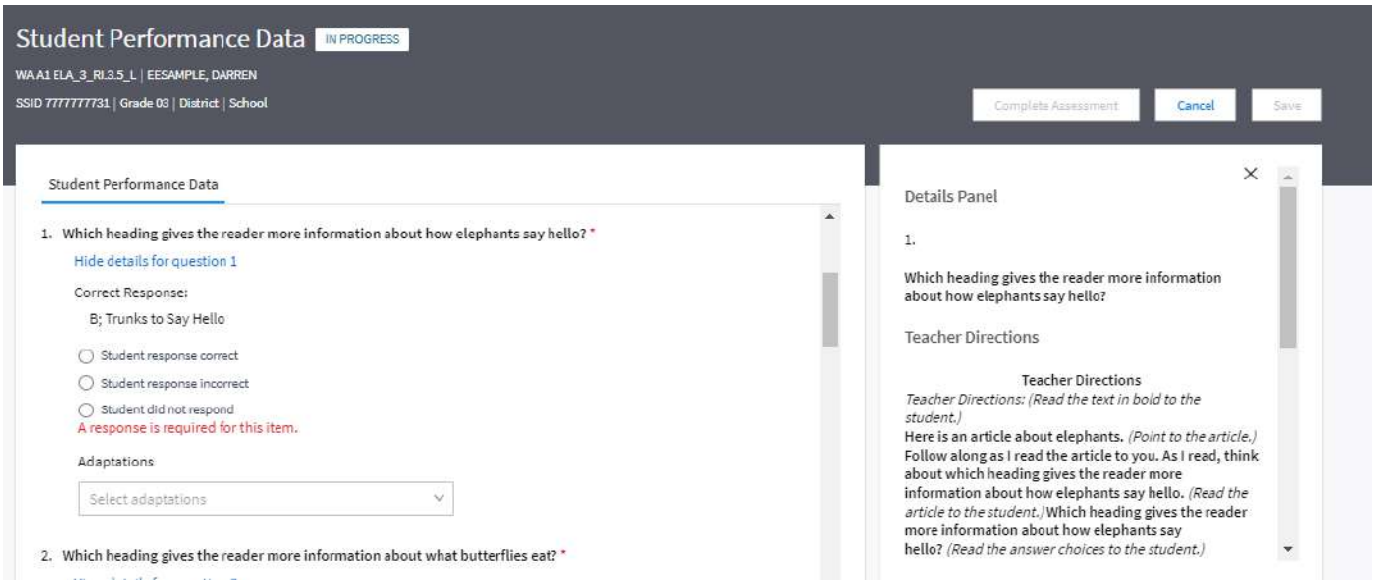
Purpose:

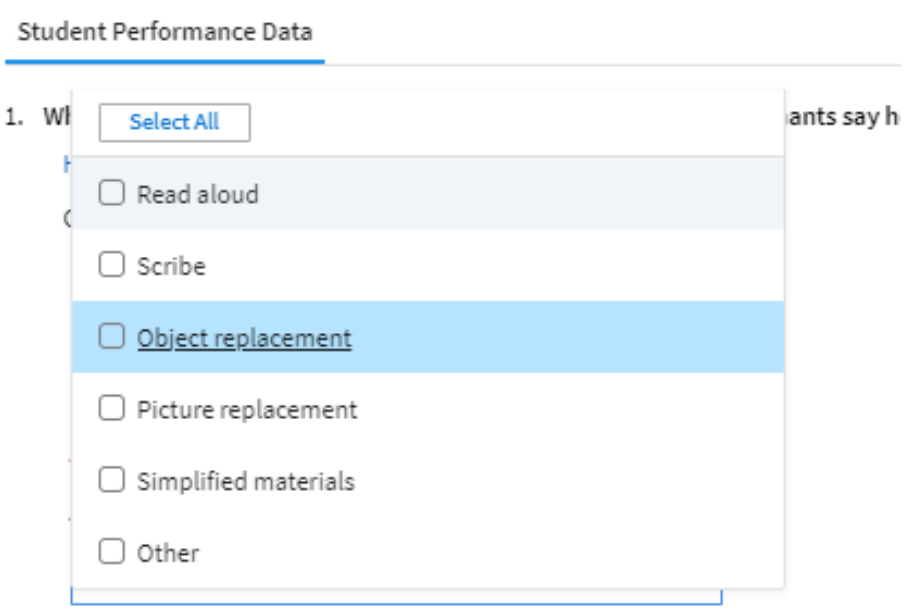
- Teacher data entry
- Standardize the scoring expectation for items
- Reduce data entry time required by teachers
- Collect item accommodations/adaptations used

Example:

- All forms associated to a student
- All forms listed independently







3.5.4. Accommodation and adaptation

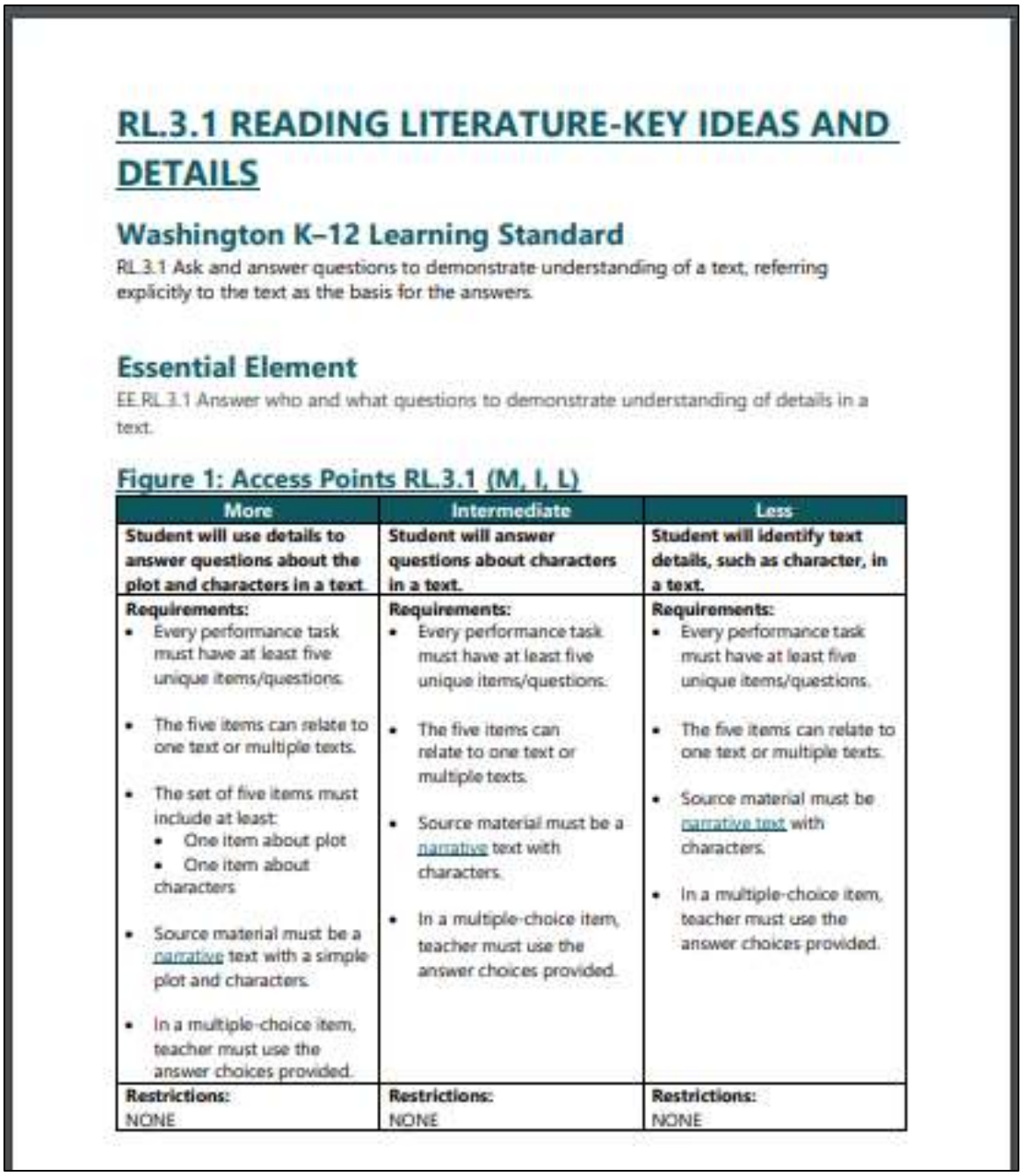
Accommodation and adaptation guidelines have been added to Performance Task documents. The INSIGHT Student Performance Data application also allows collecting accommodation and adaptation use information at the item level. The collection information is intended to inform future item development activities.

Overview: Accommodation and adaptation allowances and suggestions added to Performance Task documents

Purpose:

- Define allowable accommodation and/or adaptation for each standard and access point
- Information will help drive future item development activities

Example:



RL.3.1 READING LITERATURE-KEY IDEAS AND DETAILS

Washington K–12 Learning Standard

RL.3.1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.

Essential Element

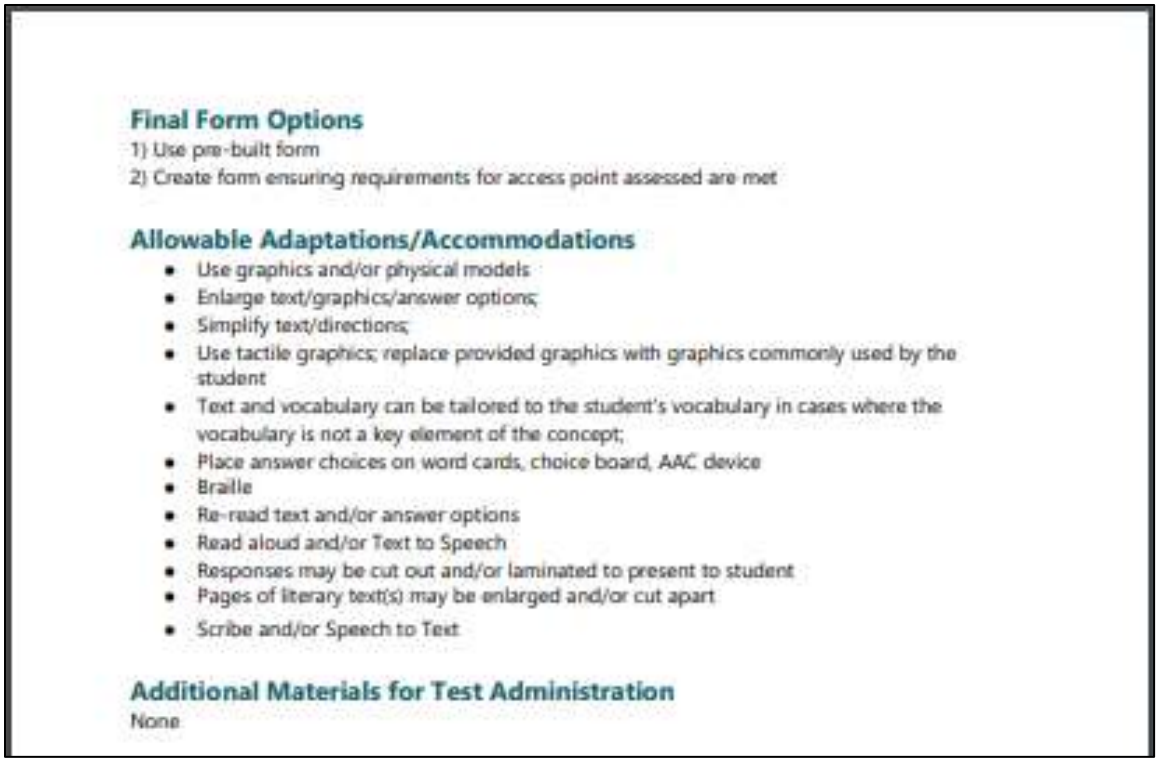
EE.RL.3.1 Answer who and what questions to demonstrate understanding of details in a text.

Figure 1: Access Points RL.3.1 (M, I, L)

More	Intermediate	Less
Student will use details to answer questions about the plot and characters in a text.	**Student will answer questions about characters in a text.**	**Student will identify text details, such as character, in a text.**

Requirements:
• Every performance task must have at least five unique items/questions.
• The five items can relate to one text or multiple texts.
• The set of five items must include at least:
  • One item about plot
  • One item about characters
• Source material must be a narrative text with a simple plot and characters.
• In a multiple-choice item, teacher must use the answer choices provided.
	Requirements:
	• Every performance task must have at least five unique items/questions.
	• The five items can relate to one text or multiple texts.
	• Source material must be a narrative text with characters.
	• In a multiple-choice item, teacher must use the answer choices provided.
		Requirements:
		• Every performance task must have at least five unique items/questions.
		• The five items can relate to one text or multiple texts.
		• Source material must be narrative text with characters.
		• In a multiple-choice item, teacher must use the answer choices provided.

Restrictions: NONE	**Restrictions:** NONE	**Restrictions:** NONE



3.6 Accommodation and Adaptations

As noted in the previous section, accommodation and adaptation guidelines have been added to Performance Task documents. OSPI worked with Washington educators who were very familiar with the administration of the WA-AIM and WA-AIM performance tasks to review and develop a list of allowable adaptations and accommodations for each performance task. This work occurred during the fall of 2020.

The INSIGHT Student Performance Data application also allows collection of adaptation and accommodation use information at the item level. The collection information is intended to inform future item development activities.

All examples and items presented in the Performance Tasks are allowed to be adapted to meet each individual student's learning style and preferred mode of receptive and expressive communication. Teachers are encouraged to present the Performance Task components in styles that most closely resemble how daily instructional materials are presented to the student. Below

are typical adaptations and ideas for presenting the Performance Tasks. **This is not an all-inclusive or exhaustive list**.

- Use graphics and/or physical models
- Enlarge text/graphics
- Simplify text/directions
- Use tactile graphics
- Use pictorial/word/object representations for numbers and graph parts
- Written material may be read aloud (unless the Performance Task specifically requires the student to read)
- Reenactments or computer simulations may be used to represent scenarios
- For items that require the student to do physical tasks, teacher may do the physical tasks if directed by the student (i.e. MS PS3-3)
- Replace provided graphics with graphics commonly used by student
- Teacher can use real-life objects when asking questions
- Text and vocabulary can be tailored to the student's vocabulary in cases where the vocabulary is not a key element of the concept
- Braille
- Dictation/Scribe/Speech to Text
- Sign language
- Re-read text and/or answer options
- Read aloud and/or Text to Speech
- Responses may be cut out and/or laminated to present to student
- Manipulatives
- Number line

Each Performance Task lists typical adaptations/accommodations relevant to the Performance Task. This is not a comprehensive list but outlines typical accommodations/adaptations used for each Performance Task.

3.7. Quality Control

In this section, quality control measures in the WA-AIM administration are summarized in two aspects: 1) quality of assessment tasks, administrations, and submitted data, which ensures fair and valid test administrations across students and instructional settings, and 2) data security, which guards against any breach of student data leading to unintended consequences to assessment stakeholders.

Quality of assessment tasks, administrations, and submitted data. The WA-AIM assessment supports individualized adaptation of the assessment for each unique student and encourages assessment practices that are meaningful and instructionally useful. Given the high-stakes nature of the assessment score uses, multiple measures are taken to ensure quality of the task design and development, administration, and submitted data in the WA-AIM assessments. Below is a list of some key measures.

- Use of the Access Point Frameworks where the measured standards are clearly articulated and presented in a consistent format across grades and content areas that is easy for educators to understand and follow.
- Use of Performance Task specifications and a standardized item library.
- Specifications and standardization of key components of the assessment administration, such as decision criteria for the access point and requirements for choosing and administering Performance Tasks.
- Requirements for and specifications of the observer attestation, which served as a local verification mechanism.
- Utilization of an online platform to facilitate development, organization, recording, and submission of assessment data, which reduces chances of clerical errors in data management and allows educators to focus more on assessment task development and administration.
- Collection of a comprehensive body of assessment data for independent data auditing by DRC’s Alternate Assessment Auditing and the OSPI.

- Availability of on-demand and online trainings as well as training and administration materials, which have helped educators understand and develop familiarity with the assessment requirements.

Data security. A secure FTP client program has been used for transfer of all confidential documents and test data between the OSPI and DRC. All data submitted to the WA-AIM online platform are stored in a secure online system for at least seven years after completion of scoring; this corresponds to elements of Washington's overall assessment-record retention policies. DRC ensures that clients' data always remain confidential and secure. DRC practices adhere to the federal Family Educational Rights and Privacy Act (FERPA) regulations for the security and confidentiality of student data, and their systems provide data privacy safeguards throughout every step of the assessment process. While FERPA provides a foundation for DRC's data privacy policy, DRC views FERPA as providing a baseline set of requirements. DRC works with state clients to meet FERPA as well as state-specific requirements and policies for securing student data. For more information specific to the various aspects of DRC data security protocols and processes, refer to Appendix D.

Chapter 4. Data Auditing and Score Reporting

This chapter starts with a description of the data under review and the related data auditing process which was implemented during the 2021–2022 WA-AIM administration for an independent check of teacher-submitted assessment data. Different from a typical handscoring process where raters assign scores based on student performance, the WA-AIM data auditing checks teacher adherence to the Performance Task specifications, and to the recording and submission protocols. Decisions from the data auditing may lead to invalidation of teacher-submitted data at the item or standard level in final score reporting.

The types of score reports and their delivery methods are introduced, and associated quality-control measures are summarized toward the end of the chapter.

4.1 Types of Data Under Review

Teacher-submitted data for students in the 2021–2022 WA-AIM administration were reviewed by DRC's Alternate Assessment Auditing team if at least one of three conditions existed (teacher created form, form with adaptations marked as other, engagement rubric form). All high school records and all engagement rubric forms were double-blind reviewed. In addition, 25% of records from grades 3 through 8 were double-blind reviewed.

All of the data under review is submitted data required of teachers (see Section 3.4 for a list of the required data components in teacher submission) for the purpose of identifying any standard-level or item-level violation that relates to adherence to the Performance Task specifications, and to the recording and submission protocols. The specific data auditing directions given to auditors appear below.

Data Review Directions Given to WA-AIM Auditors

Standard-Level Review

Non-score A (Absent):

* Review the requirements and restrictions
    a. Verify access point level auditing and read the access point description for required skills.
    b. Read the access point-specific Performance Task requirements.
    c. Read the teacher's directions.
    d. Verify response entered.
    e. Check for teacher's adaptations used.
    f. Compare teacher directions and adaptations with access point text, requirements, and restrictions.
    g. If the documentation indicates that the student wasn't present or didn't take the test, assign non-score A. Note that a refusal to participate should not be scored a non-score A. Instead, refusals are assigned an N3 violation (see below).

Non-score I (Requirements or Restrictions Not Met):

* Review the requirements and restrictions
    a. Verify access point level auditing and read the access point description for required skills.
    b. Read the access point-specific Performance Task requirements.
    c. Read the teacher's directions.
    d. Verify response entered.
    e. Check for teacher's adaptations used.
    f. Compare teacher directions and adaptations with access point text, requirements, and restrictions.
    g. If the requirements were not met, assign non-score I.
    h. If the restrictions were violated, assign non-score I.

Item-Level Review

If no standard-level violations are found, proceed to reviewing each of the five items separately:

1. Look at the response entered for the first item. Look at the documentation submitted. The teacher documentation should corroborate the response entered.
    a. If the documentation supports the response entered, mark Y and proceed to the next item.

b. If the documentation does not support the response entered, mark N. Then select one of the following reason codes and move on to the next item:
   N1. "Documentation set-up error"
   N2. "Mismatch between response entered and teacher documentation"
   N3. "Teacher documentation too limited to support response entered"

2. More about reason codes.

NOTE: The following are general rules. In all cases, refer to the specific auditing notes for the standard/access point level for further instructions.

Assign code N1 when there is evidence of a problem with the item set-up. For example, some items have requirements for the type of source material (e.g., informational versus literary test), or the item types (e.g., subtraction and addition, positive numbers and negative numbers, questions about plot and questions about characters, etc.).

Assign code N2 when there is clear evidence that the response entered was not entered correctly based on the teacher documentation. For example, if the teacher documentation states that the student answered the question correctly, but the student response is entered as incorrect, N-2 should be assigned.

Assign code N3 when the teacher documentation does not have enough information to support the response entered. Examples of some common N-3 violations are key-bashing and refusals. Note: Refusals are different from the student not being present to take the test. When students are not present to take the test, a non-score A is assigned.

Additional Notes on Item-Level Violations:

- If have both N1 & N2 at the individual item level, assign N1.
- If have both N2 & N3 at the individual item level, assign N3.
- If have N1, N2 & N3 at the individual item level, assign N1.

If there were one or more flags from the standard-level review, the assessed standard would be invalidated and marked as non-scorable (NS). When one or more flags were noted from the item-level review, the assessed item would be invalidated. More detailed information on the WA-AIM scoring can be found in Section 1.6.

Information on agreement between auditors 1 and 2 was used to gather inter-auditor consistency data. The auditor agreement statistics for the 2021–2022 WA-AIM administration are presented in Section 4.2.4. Auditing decision comparisons between auditors 1 and 2 were also one of the measures in place for checks of data auditing quality and to identify the potential need for additional training among the Alternate Assessment Auditing staff.

4.2 Data Auditing Process

A series of activities were arranged to support the WA-AIM data auditing. Below is a summary of the key events.

Supervisor Training, February/March 2021 & November 2021 (Indianapolis)

- An overview of remote work policies and procedures.
- An overview of the WA-AIM program.
- Review of the WA-AIM Access Point Frameworks and Performance Tasks (complexity levels, Performance Task requirements, sample items, etc.).
- Auditing platform (ScoreBoard) overview (auditing by standard instead of auditing everything at once for each student, tools in ScoreBoard to use during auditing, how to assign auditing decisions arrived during the auditing process (A, I, Y, N1, N2, N3) in ScoreBoard, etc.).
- Overview of processes used to monitor the auditing decisions applied and quickly identify areas of auditor drift (training and qualification reports, auditing summary reports, validity set reports, dual-audit reports, conducting dual audits).
- Review of data auditing rules.

DRC Final Preparation of Training Materials, February 2022 (Indianapolis)

- Entered all training materials into training system.
- DRC's Alternate Assessment Auditing team selected validity sets in ScoreBoard based on customer-approved philosophies established from the review of auditor training materials.

Data Auditing (Indianapolis)

- Data auditing occurred in May-June. Auditing took place remotely by trained and highly experienced supervisors, team leaders and scorers. DRC's remote work is designed to very closely emulate the work done in physical scoring locations, applying the same metrics.
- Team leader and scorer training and qualification checks for the first submission window occurred in May/June 2022.
- Applying final data auditing decisions in ScoreBoard.
- Dual audits.
- High school third-audit resolution.

4.2.1 Review of Auditor Training Materials

Thoroughly vetted training materials are the foundation on which accurate data audits are built. DRC's Alternate Assessment Auditing team created training materials with annotations that clarified the intent of each data auditing decision assigned to submitted data. Updated training materials to reflect the new submission platform were developed in February 2022. These materials were presented to the OSPI for review/approval. A description of how auditor training materials were reviewed and developed is given below.

The purpose of reviewing auditor training materials involves examining a randomly selected set of student records representative of those submitted for the assessment. The review focuses on identifying trends or issues that should be addressed in the WA-AIM data auditing procedures. In addition, data entries to be used in auditor training, practice, and qualifier samples are identified.

Hundreds of student records from districts across the state of Washington were considered for inclusion while developing auditor training materials. Before submitting auditor training materials to OSPI to review, DRC's Alternate Assessment Auditing team reviewed and confirmed the viability of the data records for evaluation. Student records represented a variety of districts and assessed standards.

DRC Alternate Assessment Auditing supervisors developed and reviewed updated auditor training materials. The materials were then sent to OSPI for review and approval. Alternate Assessment Auditing supervisors were selected from a pool of highly qualified scoring specialists who have had supervisory experience with both WA-AIM and other DRC projects; these identified specialists were selected to act as supervisors during operational data audits.

Participants reviewing auditor training materials reviewed and trial-audited presented data samples using the same processes and procedures to be applied during operational data audits. Based on the samples reviewed and the trial data audits, the group came to consensus on changes to or clarification of the data auditing process auditor guidelines that would provide greater understanding and make the process easier to follow.

4.2.2 Auditor Qualifications and Training

Selected auditors for the WA-AIM data review, at a minimum, hold a bachelor's degree. All supervisors who led the auditing process had multiple years of WA-AIM auditing experience.

Prior to the actual data auditing, DRC's Alternate Assessment Auditing team created training materials. The process included several presorting steps and subsequent iterative/consensus processes in order to achieve ever-increasing agreement and precision. Looking at materials multiple times ensured that all factors had been considered, which led to a consistent data auditing approach. Individuals looked at materials independently and then, through group review, reached consensus on data audits and selection. When a subset of data records for a grade and content area were selected and assigned a status as good anchors, training, qualifying, or validity set examples, they were consolidated into training formats. Auditing guides (consisting of the associated standards for Performance Tasks with access point level

requirements and restrictions, and annotated anchors) served as a process foundation, setting the course for all subsequent training and data audits. In addition, auditors received specialized ScoreBoard training before the start of auditing in order to become familiar and comfortable with the system.

Validation is a critical task in the assessment training process. It ensures that auditors have internalized the data auditing guidelines before they begin the process. Reports from ScoreBoard show when auditors have drifted from auditing philosophies. All auditors must achieve a minimum of 80% exact agreement on the qualifying round following the completion of training. If validation does not occur on the first attempt, further training occurs prior to taking an additional qualifying round. Only those who successfully qualified were allowed to audit submitted assessments.

4.2.3 Data auditing Procedure

ScoreBoard distributed assessments electronically to auditors. Student work from the same class was automatically routed to multiple auditors. Assessments were automatically routed to auditors until all audits were complete. For grades 3 through 8, 25% of the students were selected for a dual audit. For high school, all students received dual audits. All students with engagement rubrics also received dual audits. At grades 3 through 8, review results from auditor 1 were considered the final decisions. Auditors cannot tell if they are the first or second auditor, and they also cannot access auditing decisions given by other auditors. Auditing decision comparisons between auditors 1 and 2 were used as one of the measures to facilitate interim checks of data-audit quality and to identify the potential need of additional training among auditing staff.

At high school, due to the high stakes associated with state graduation requirements, the OSPI conducted third-audit reviews for student records with differences in decisions between auditors 1 and 2. OSPI decisions from third-auditor reviews were entered into ScoreBoard as the final decisions.

Throughout the course of the data auditing process, calibration sets of pre-audited data records (validity sets) were administered daily to all auditors to monitor data-audit accuracy and to maintain a consistent focus on the established rubric and data auditing guidelines. Live assessments are used as validity sets during auditing, so auditors have no way of knowing when the assessment they are auditing is a validity set. Validity set reports from ScoreBoard show how the whole group and/or individual auditors are maintaining consistency to the established processes. If auditors' validity set results fall below the qualifying level, those auditors are removed from live data audits and given additional training and another qualifying (validation) round. Auditors unable to re-qualify are dismissed. If auditors re-qualify, but fail to maintain standards later, they are also dismissed. When auditors are dismissed, their work is reviewed carefully to see whether it needs to be redone by another auditor.

Supervisors or team leaders audited a random selection of assessments audited (akin to backreading in other scoring processes). The auditing decisions were compared, and if they agreed, feedback was offered which enhanced the auditor's confidence and ability to assign decisions quickly and accurately. However, if an individual was straying from the standard established in the training and validation samples, the aberrant data audit was detected, and dual audits were conducted more frequently for any auditor whose data auditing appeared inconsistent. In addition, if data audits are found to be outside acceptable parameters, they are reviewed to see whether submissions need to be re-evaluated by another auditor.

Auditing Summary Reports from ScoreBoard provide information on how each auditor is doing with respect to auditor agreement. This information is helpful in monitoring inter-auditor reliability. Auditing Summary reports were posted daily for the OSPI.

4.2.4 Auditor Agreement

As noted in the previous section, there were 100% dual audits at high school and 25% dual audits at grades 3 through 8 for test data that were routed to data auditing in the 2022 WA-AIM administration. The dual-audit approach was used to promote auditor reliability through monitoring.

Auditor agreement statistics reported in this section were obtained using results of the high school and grades 3 through 8 dual audits. A summary of the agreement results by grade and content area for the spring and fall administrations is presented in Table 4.1, with separate statistics on standard-level and item-level agreement.

The auditor agreements in Table 4.1 were defined considering agreement impact on student scores. At the standard level, an agreement was recorded when two auditors agreed on the scorable quality of a given standard; at the item level, the auditor agreement was determined by two auditors agreeing on the same number of items (out of a total of five items) on a given standard regarding the item scorable quality.

As shown in Table 4.1, perfect agreement (100%) was observed at the standard level across grades and content areas. At the item level, perfect or near perfect agreement (> 98%) was observed.

Table 4.1. Auditor Agreement, Standard Level and Item Level
Content Area: ELA
Spring 2022

Grade	N of Dual Audits	Standard-Level Agreement		Item-Level Agreement	
		Agreed %	*Not Agreed %*	*Agreed %*	*Not Agreed %*
3	95	100.00	0.00	98.94	1.06
4	93	100.00	0.00	100.00	0.00
5	86	100.00	0.00	100.00	0.00
6	79	100.00	0.00	100.00	0.00
7	94	100.00	0.00	100.00	0.00
8	80	100.00	0.00	100.00	0.00
10	277	100.00	0.00	99.64	0.36

Content Area: Mathematics

Spring 2022

Grade	N of Dual Audits	Standard-Level Agreement		Item-Level Agreement	
		Agreed %	*Not Agreed %*	*Agreed %*	*Not Agreed %*
3	86	100.00	0.00	100.00	0.00
4	62	100.00	0.00	100.00	0.00
5	73	100.00	0.00	100.00	0.00
6	71	100.00	0.00	100.00	0.00
7	88	100.00	0.00	100.00	0.00
8	79	100.00	0.00	100.00	0.00
10	258	100.00	0.00	100.00	0.00

Content Area: Science

Not applicable, as only pre-built forms are allowed

4.3 Score Reporting

4.3.1 Types of Score Reports

Three types of reports were prepared for the 2022 WA-AIM administration. These reports included a) Individual Student Report, b) Student Roster Report, and c) Student Data File (SDF). The three reports are described below.

Individual Student Report (ISR). ISRs were produced for all students for whom an answer record was received from the WA-AIM system whether or not all portions of the test were taken. The report contained the student's score information in each content area and content category (standard). The report provided a total score for the content area, alternate achievement levels in graphic display, the score range of each alternate achievement level, the weighted content category score, the complexity level for each content category, and static narrative statements describing the different sections of the report. On the back of the report were grade-specific Alternate Achievement Level Definitions.

Student Roster Report. The Student Roster Report contained a list of all students in a group (class). The report provided a total score for the content area, alternate achievement level

scores, the weighted content category score, the complexity level for each content category, and a static purpose statement describing the report.

Student Data File (SDF). The SDF was a file of student data sorted by district, school, grade, and student last name. Each student record contained detailed information about the student's demographic profiles, test record flags (e.g., invalidation flags), and obtained scores and alternate achievement levels for each content area and content category. The SDF was produced for the entire state as one file which included final data for each of the WA-AIM test windows.

Parents may access an interpretation guide to help understand information presented in an ISR. The guide, Washington Access to Instruction & Measurement Score Interpretation Guide for Families and Educators, is attached in Appendix E. An ISR mock-up is also included in the guide.

4.3.2 Report Delivery

The SDF was delivered to the OSPI in electronic format. Individual Student Reports and Student Roster Reports were posted by district and school to the DRC INSIGHT Portal. Printed reports were also distributed for the 2022 administrations.

Starting with the spring 2017 administration, the ISRs and Student Roster Reports have been made available online on INSIGHT. State, district, and school users can access score reports for their district and school from INSIGHT.

4.4 Quality Control

DRC's data management system was used to provide valid, reliable, and cost-efficient data auditing. The quality assurance groups within DRC's technology department and performance assessment services were both charged with reviewing data and reports during all stages of the process. The technology quality assurance team verified the accuracy of all reporting programs before the programs were made operational. The auditing quality assurance team verified the

accuracy of report information during the data auditing process. After all data were entered into the scoring system and all reporting programs were completed, a sample of reports were submitted to the scoring quality assurance group, which reviewed the sample reports to verify the accuracy and correct presentation of all data.

Systematic quality assurance checks were in place throughout the data auditing process to ensure the accuracy of reports. Prior to delivering any electronic files, all reports were given a final, extensive quality check. This final review was conducted by multiple DRC departments. The reports were verified for accuracy and correct format, and to ensure that they matched the detailed requirements outlined for each report.

Chapter 5. Feedback Loop for Assessment Improvement

This chapter explains a feedback loop that has been implemented to continuously improve the WA-AIM assessment based on feedback from WA-AIM users and from the DRC's Alternate Assessment Auditing team in review of assessment records.

5.1 User Feedback Survey

As part of WA-AIM's feedback loop for improving the assessment, the OSPI has conducted a voluntary User Feedback Survey (see an example in Appendix C) at each test administration to solicit feedback on the experience of administering the WA-AIM. The survey collected feedback in the following areas:

- Administration Materials (OSPI website, Test Administration Manual, guidelines, trainings, etc.)
- Administration Platform and Supports (ease of use, user guides, etc.)
- Data Collection Platform (what users liked, suggestions for improvement)
- Help Desk and Customer Care (experience and response)
- Standards and Access Points (what needs clarification)

Feedback through the survey and other methods (e.g., email, letters, and phone calls) in previous years pointed to same areas in need of improvement. These continuing areas for improvement include:

Time

Many responses included frustration with the amount of time it takes to create assessment materials, administer the assessment, and document the assessment.

In response, the OSPI has made the following changes:

- Creating more example items which can be administered to students.
- Eliminating requiring baseline documentation in the Data Collection Platform.

Furthermore, OSPI addressed these concerns through item development efforts during the 2018–2021 school years. DRC was contracted for development to provide educators with fifteen items for each standard and access point for use in the final assessment.

The bulk of the item content was deployed for use during the 2020–2021 WA-AIM administration. Some passage-based ELA items and additional science sets were unable to be made available for the 2020–2021 administration but were made available for the 2021–2022 administration.

The new item development has led to an expanded, standardized item library. With the item library available, educators administering the WA-AIM are no longer allowed or able to create their own item content, and all final assessments must use items from the library that was developed through OSPI and DRC.

Ease of Use of Materials, Guides, Trainings

Many responses indicated being overwhelmed by the materials, navigating the various guides, and the length of training modules.

In response, the OSPI made the following changes:

- Re-formatting of Performance Tasks to include clear identification of items
- Adding answer keys for all example items
- Updating user guides for better user navigation
- Creating shorter, more specific training modules that users can access on-demand

In December 2018 the OSPI released a request for proposal (RFP) for administration of the WA-AIM to begin with the spring 2020 administration. Included in the RFP were requests to address the areas in need of improvement.

Further efforts were made for the 2020–2021 administration including:

- Added applications within INSIGHT allowing this to become a one-stop place for the assessment administration.
- Functionality changes in INSIGHT to allow better control of access to specific INSIGHT application during test administration
- Further redesign of the Performance Task document to shorten their length and increase their reference usability

5.2 Alternate Assessment Auditing Notes

Annually, the DRC Alternate Assessment Auditing supervisor and team leads provide to the OSPI summary information of patterns of nonscorable standards or items observed during the auditing process. The feedback has been used to inform improvement in test development, educator training, and test administration.

Chapter 6. Test Data Analyses

Test-level statistics are presented in this chapter, including test participation, total score distributions, achievement level distributions, reliability indices, classification accuracy and consistency indices, and subgroup statistics. The statistics are intended to provide summative information on Washington students' performance on the 2021–2022 WA-AIM assessment and to present test-level empirical evidence regarding fairness, reliability and validity of the assessment.

In addition to test-level statistics, difficulty and discrimination data at the content standard level are also provided, as well as the distribution of access points at each standard and the distribution of raw score points by access point and standard.

As mentioned in Chapter 1, the WA-AIM assessment is designed as a measure of five standards within a content area. Each standard is assessed by a Performance Task that aligns to the student's access point (three possible access points at each standard). The access point is chosen by the teacher for a given student on a given standard, and each Performance Task has five dichotomously scored (score 0 for incorrect responses, and score 1 for correct responses) items. This design is consistent across all grades and content areas. In the design, the *Intermediate Complex* access point is anchored to minimal mastery-level expectations linked with the general education standards. Given that the assessment design, scoring algorithm, and access-point weights are the same across grades and content areas in the WA-AIM, it can be viewed that regardless of the assessed grade and content area, if a student achieved full points on the *Intermediate Complex* access point for all five standards (which would result in a final total test score of 143), the student would meet minimal mastery-level expectations for that grade and content area.

With that said, WA-AIM scores are reported on grade-specific scales; therefore, any inferences made from cross-grade comparisons should be conducted with caution and triangulated with additional sources of evidence.

Another consideration when interpreting the reported statistics is that the WA-AIM assessment is administered to a relatively small testing population as compared to that on a typical large-scale state assessment. Readers are recommended to interpret reported reliability/precision estimates with those factors in mind.

In addition, due to circumstances related to the COVID-19 pandemic, participation rates in the 2021–2022 WA-AIM were close to but not as high as those in a typical administration year (see more details about the student participation in Section 6.1, Chapter 6). Accordinlgy, aggregated test performance results from the 2021–2022 WA-AIM could be impacted by a variety of factors, such as lower than typical test participation rates, potential overrepresentation of certain demographic groups and underrepresentation of other groups in the tested students, possible disrupted learning during previous and current school years, and differences in instruction delivery (e.g., online, in-person, or hybrid model). As such, any comparison of the group test results from the 2021–2022 administration should be made with caution.

6.1 Test Participation

In the 2021–2022 administration, 599 to 761 students by assessment grade received valid test scores at grades 3 through 8 and high school. The number of the students by grade is summarized in Table 6.1. Across grades, these numbers are greater than those from the 2020–2021 administration and are approximately 80% (grade 6) to 98% (grade 11) of those in a normal year (e.g., 2018–2019).

Table 6.1. Number of Students with Valid Test Scores

Grade	N of Tested Students in 2021–2022	*Reference:* N of Tested Students in 2020–2021	*Reference:* N of Tested Students in 2018–2019
3	720	650	821
4	761	605	812
5	647	596	783
6	599	563	747
7	622	569	730
8	611	539	682
10	653	517	747
11	613	475	628

6.2 Demographics of the Participants

Tables 6.2 describes demographic distributions of the participating students who obtained valid scores in each content area of the 2021–2022 WA-AIM. In addition, demographics from the 2018–2019 and the 2020–2021 WA-AIM administrations are included for reference. The demographics include

- gender (*Female* vs. *Male*)
- income status (*Low Income* vs. *Non-Low Income*)
- Bilingual status (*Bilingual* vs. *Non-Bilingual*)
- race and ethnicity (i.e., *American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Caucasian or White, Native Hawaiian or Other Pacific Islander*, and *Of More than One Race/Multi-Racial*).

Table 6.2. Demographics of Tested Students

ELA

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	**Percent**	**Valid N**	**Percent**	**Valid N**	**Percent**
3	Female	214	29.72%	226	34.93%	243	29.78%
	Male	506	70.28%	421	65.07%	573	70.22%
	Low Income	445	61.81%	396	61.21%	532	65.28%
	Non-Low Income	275	38.19%	251	38.79%	283	34.72%
	Bilingual	167	23.19%	150	23.18%	156	19.14%
	Non-Bilingual	553	76.81%	497	76.82%	659	80.86%
	Hispanic/Latino	219	30.42%	175	27.05%	228	27.98%
	Non-Hispanic/Latino	501	69.58%	472	72.95%	587	72.02%
	American Indian or Alaska Native	9	1.25%	9	1.39%	11	1.35%
	Asian	62	8.61%	43	6.65%	55	6.75%
	Black or African American	62	8.61%	56	8.66%	59	7.24%
	Hispanic or Latino	219	30.42%	175	27.05%	228	27.98%
	Caucasian or White	286	39.72%	295	45.60%	360	44.17%
	Native Hawaiian or Other Pacific Islander	9	1.25%	7	1.08%	11	1.35%
	Of More than One Race/Multi-Racial	73	10.14%	62	9.58%	91	11.17%
4	Female	269	35.44%	170	28.33%	253	31.51%
	Male	490	64.56%	430	71.67%	550	68.49%
	Low Income	494	65.09%	359	59.83%	532	66.25%
	Non-Low Income	265	34.91%	241	40.17%	271	33.75%
	Bilingual	191	25.16%	146	24.33%	170	21.17%
	Non-Bilingual	568	74.84%	454	75.67%	633	78.83%
	Hispanic/Latino	214	28.19%	193	32.17%	226	28.14%
	Non-Hispanic/Latino	545	71.81%	407	67.83%	577	71.86%
	American Indian or Alaska Native	12	1.58%	5	0.83%	14	1.74%
	Asian	63	8.30%	45	7.50%	52	6.48%
	Black or African American	59	7.77%	37	6.17%	45	5.60%
	Hispanic or Latino	214	28.19%	193	32.17%	226	28.14%
	Caucasian or White	327	43.08%	253	42.17%	388	48.32%
	Native Hawaiian or Other Pacific Islander	9	1.19%	8	1.33%	13	1.62%
	Of More than One Race/Multi-Racial	75	9.88%	59	9.83%	65	8.09%
5	Female	188	29.24%	184	31.45%	252	32.56%
	Male	455	70.76%	400	68.38%	522	67.44%
	Low Income	416	64.70%	349	59.66%	484	62.53%
	Non-Low Income	227	35.30%	236	40.34%	290	37.47%
	Bilingual	160	24.88%	121	20.68%	153	19.77%
	Non-Bilingual	483	75.12%	464	79.32%	621	80.23%
	Hispanic/Latino	213	33.13%	172	29.40%	228	29.46%
	Non-Hispanic/Latino	430	66.87%	413	70.60%	546	70.54%
	American Indian or Alaska Native	8	1.24%	5	0.85%	7	0.90%

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
	Asian	44	6.84%	48	8.21%	54	6.98%
	Black or African American	52	8.09%	41	7.01%	59	7.62%
	Hispanic or Latino	213	33.13%	172	29.40%	228	29.46%
	Caucasian or White	255	39.66%	261	44.62%	351	45.35%
	Native Hawaiian or Other Pacific Islander	9	1.40%	7	1.20%	10	1.29%
	Of More than One Race/Multi-Racial	62	9.64%	51	8.72%	65	8.40%
6	Female	170	28.48%	175	31.36%	237	32.07%
	Male	426	71.36%	383	68.64%	502	67.93%
	Low Income	371	62.14%	346	62.01%	479	64.91%
	Non-Low Income	226	37.86%	212	37.99%	259	35.09%
	Bilingual	116	19.43%	131	23.48%	138	18.70%
	Non-Bilingual	481	80.57%	427	76.52%	600	81.30%
	Hispanic/Latino	169	28.31%	179	32.08%	208	28.18%
	Non-Hispanic/Latino	428	71.69%	379	67.92%	530	71.82%
	American Indian or Alaska Native	5	0.84%	11	1.97%	17	2.30%
	Asian	41	6.87%	32	5.73%	50	6.78%
	Black or African American	46	7.71%	32	5.73%	63	8.54%
	Hispanic or Latino	169	28.31%	179	32.08%	208	28.18%
	Caucasian or White	277	46.40%	256	45.88%	338	45.80%
	Native Hawaiian or Other Pacific Islander	3	0.50%	12	2.15%	7	0.95%
	Of More than One Race/Multi-Racial	56	9.38%	36	6.45%	55	7.45%
7	Female	201	32.52%	195	34.45%	253	35.19%
	Male	417	67.48%	371	65.55%	466	64.81%
	Low Income	397	64.24%	338	59.72%	451	62.73%
	Non-Low Income	221	35.76%	228	40.28%	268	37.27%
	Bilingual	157	25.40%	135	23.85%	125	17.39%
	Non-Bilingual	461	74.60%	431	76.15%	594	82.61%
	Hispanic/Latino	195	31.55%	175	30.92%	196	27.26%
	Non-Hispanic/Latino	423	68.45%	391	69.08%	523	72.74%
	American Indian or Alaska Native	14	2.27%	6	1.06%	12	1.67%
	Asian	44	7.12%	38	6.71%	56	7.80%
	Black or African American	40	6.47%	43	7.60%	57	7.94%
	Hispanic or Latino	195	31.55%	175	30.92%	196	27.30%
	Caucasian or White	271	43.85%	248	43.82%	323	44.99%
	Native Hawaiian or Other Pacific Islander	12	1.94%	14	2.47%	10	1.39%
	Of More than One Race/Multi-Racial	42	6.80%	42	7.42%	64	8.91%
8	Female	212	35.04%	185	34.71%	218	33.08%
	Male	393	64.96%	348	65.29%	441	66.92%
	Low Income	384	63.47%	333	62.48%	379	57.51%
	Non-Low Income	221	36.53%	200	37.52%	280	42.49%
	Bilingual	137	22.64%	110	20.64%	106	16.08%
	Non-Bilingual	468	77.36%	423	79.36%	553	83.92%

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
	Hispanic/Latino	186	30.74%	164	30.77%	157	23.82%
	Non-Hispanic/Latino	419	69.26%	369	69.23%	502	76.18%
	American Indian or Alaska Native	8	1.32%	12	2.25%	8	1.21%
	Asian	36	5.95%	35	6.57%	64	9.71%
	Black or African American	45	7.44%	43	8.07%	43	6.53%
	Hispanic or Latino	186	30.74%	164	30.77%	157	23.82%
	Caucasian or White	265	43.80%	242	45.40%	329	49.92%
	Native Hawaiian or Other Pacific Islander	13	2.15%	5	0.94%	11	1.67%
	Of More than One Race/Multi-Racial	52	8.60%	32	6.00%	47	7.13%
	Female	213	32.82%	181	35.84%	250	33.83%
	Male	436	67.18%	324	64.16%	489	66.17%
	Low Income	382	58.86%	257	50.89%	405	54.88%
	Non-Low Income	267	41.14%	248	49.11%	333	45.12%
	Bilingual	123	18.95%	87	17.23%	103	13.96%
	Non-Bilingual	526	81.05%	418	82.77%	635	86.04%
	Hispanic/Latino	175	26.96%	120	23.76%	158	21.41%
10	Non-Hispanic/Latino	474	73.04%	385	76.24%	580	78.59%
	American Indian or Alaska Native	12	1.85%	5	0.99%	14	1.90%
	Asian	62	9.55%	43	8.51%	59	7.99%
	Black or African American	49	7.55%	36	7.13%	44	5.96%
	Hispanic or Latino	175	26.96%	120	23.76%	158	21.41%
	Caucasian or White	291	44.84%	249	49.31%	401	54.34%
	Native Hawaiian or Other Pacific Islander	6	0.92%	12	2.38%	4	0.54%
	Of More than One Race/Multi-Racial	54	8.32%	40	7.92%	58	7.86%

Mathematics

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
3	Female	214	29.76%	224	34.84%	242	29.73%
	Male	505	70.24%	419	65.16%	572	70.27%
	Low Income	445	61.89%	391	60.81%	530	65.19%
	Non-Low Income	274	38.11%	252	39.19%	283	34.81%
	Bilingual	167	23.23%	147	22.86%	158	19.43%
	Non-Bilingual	552	76.77%	496	77.14%	655	80.57%
	Hispanic/Latino	219	30.46%	171	26.59%	228	28.04%
	Non-Hispanic/Latino	500	69.54%	472	73.41%	585	71.96%
	American Indian or Alaska Native	9	1.25%	9	1.40%	11	1.35%
	Asian	61	8.48%	43	6.69%	56	6.89%
	Black or African American	62	8.62%	56	8.71%	58	7.13%
	Hispanic or Latino	219	30.46%	171	26.59%	228	28.04%
	Caucasian or White	286	39.78%	295	45.88%	357	43.91%
	Native Hawaiian or Other Pacific Islander	9	1.25%	7	1.09%	11	1.35%
	Of More than One Race/Multi-Racial	73	10.15%	62	9.64%	92	11.32%
4	Female	269	35.68%	168	28.14%	254	31.75%
	Male	485	64.32%	429	71.86%	546	68.25%
	Low Income	489	64.85%	357	59.80%	527	65.87%
	Non-Low Income	265	35.15%	240	40.20%	273	34.12%
	Bilingual	190	25.20%	142	23.79%	170	21.25%
	Non-Bilingual	564	74.80%	455	76.21%	630	78.75%
	Hispanic/Latino	213	28.25%	193	32.33%	223	27.88%
	Non-Hispanic/Latino	541	71.75%	404	67.67%	577	72.12%
	American Indian or Alaska Native	12	1.59%	5	0.84%	14	1.75%
	Asian	63	8.36%	43	7.20%	51	6.38%
	Black or African American	59	7.82%	37	6.20%	44	5.50%
	Hispanic or Latino	213	28.25%	193	32.33%	223	27.88%
	Caucasian or White	323	42.84%	252	42.21%	390	48.75%
	Native Hawaiian or Other Pacific Islander	9	1.19%	8	1.34%	13	1.63%
	Of More than One Race/Multi-Racial	75	9.95%	59	9.88%	65	8.13%
5	Female	188	29.19%	184	31.13%	250	32.43%
	Male	456	70.81%	406	68.70%	521	67.57%
	Low Income	415	64.44%	353	59.73%	481	62.39%
	Non-Low Income	229	35.56%	238	40.27%	290	37.61%
	Bilingual	160	24.84%	121	20.47%	153	19.84%
	Non-Bilingual	484	75.16%	470	79.53%	618	80.16%
	Hispanic/Latino	212	32.92%	173	29.27%	228	29.57%
	Non-Hispanic/Latino	432	67.08%	418	70.73%	543	70.43%
	American Indian or Alaska Native	8	1.24%	5	0.85%	7	0.91%
	Asian	46	7.14%	48	8.12%	53	6.87%
	Black or African American	52	8.07%	43	7.28%	59	7.65%

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
	Hispanic or Latino	212	32.92%	173	29.27%	228	29.57%
	Caucasian or White	255	39.60%	264	44.67%	350	45.40%
	Native Hawaiian or Other Pacific Islander	9	1.40%	6	1.02%	10	1.30%
	Of More than One Race/Multi-Racial	62	9.63%	52	8.80%	64	8.30%
6	Female	169	28.55%	176	31.48%	237	32.33%
	Male	422	71.28%	383	68.52%	496	67.67%
	Low Income	369	62.33%	349	62.43%	476	65.03%
	Non-Low Income	223	37.67%	210	37.57%	256	34.97%
	Bilingual	115	19.43%	131	23.43%	138	18.85%
	Non-Bilingual	477	80.57%	428	76.57%	594	81.15%
	Hispanic/Latino	168	28.38%	178	31.84%	208	28.42%
	Non-Hispanic/Latino	424	71.62%	381	68.16%	524	71.58%
	American Indian or Alaska Native	5	0.84%	11	1.97%	17	2.32%
	Asian	41	6.93%	33	5.90%	49	6.69%
	Black or African American	45	7.60%	32	5.72%	63	8.61%
	Hispanic or Latino	168	28.38%	178	31.84%	208	28.42%
	Caucasian or White	275	46.45%	254	45.44%	334	45.63%
	Native Hawaiian or Other Pacific Islander	3	0.51%	12	2.15%	7	0.96%
	Of More than One Race/Multi-Racial	55	9.29%	39	6.98%	54	7.38%
7	Female	200	32.36%	194	34.52%	255	35.37%
	Male	418	67.64%	368	65.48%	466	64.63%
	Low Income	396	64.08%	335	59.61%	454	62.97%
	Non-Low Income	222	35.92%	227	40.39%	267	37.03%
	Bilingual	155	25.08%	132	23.49%	127	17.61%
	Non-Bilingual	463	74.92%	430	76.51%	594	82.39%
	Hispanic/Latino	195	31.55%	172	30.60%	198	27.46%
	Non-Hispanic/Latino	423	68.45%	390	69.40%	523	72.54%
	American Indian or Alaska Native	14	2.27%	6	1.07%	12	1.66%
	Asian	44	7.12%	38	6.76%	58	8.04%
	Black or African American	39	6.31%	43	7.65%	58	8.04%
	Hispanic or Latino	195	31.55%	172	30.60%	198	27.46%
	Caucasian or White	272	44.01%	246	43.77%	321	44.52%
	Native Hawaiian or Other Pacific Islander	12	1.94%	14	2.49%	10	1.39%
	Of More than One Race/Multi-Racial	42	6.80%	43	7.65%	64	8.88%
8	Female	211	34.93%	184	34.65%	223	33.63%
	Male	393	65.07%	347	65.35%	440	66.37%
	Low Income	382	63.25%	331	62.34%	381	57.47%
	Non-Low Income	222	36.75%	200	37.66%	282	42.53%
	Bilingual	137	22.68%	110	20.72%	106	15.99%
	Non-Bilingual	467	77.32%	421	79.28%	557	84.01%
	Hispanic/Latino	184	30.46%	163	30.70%	158	23.83%
	Non-Hispanic/Latino	420	69.54%	368	69.30%	505	76.17%

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	**Percent**	**Valid N**	**Percent**	**Valid N**	**Percent**
	American Indian or Alaska Native	8	1.32%	12	2.26%	8	1.21%
	Asian	36	5.96%	35	6.59%	64	9.65%
	Black or African American	45	7.45%	43	8.10%	43	6.49%
	Hispanic or Latino	184	30.46%	163	30.70%	158	23.83%
	Caucasian or White	266	44.04%	241	45.39%	332	50.08%
	Native Hawaiian or Other Pacific Islander	13	2.15%	6	1.13%	11	1.66%
	Of More than One Race/Multi-Racial	52	8.61%	31	5.84%	47	7.09%
	Female	213	32.87%	185	35.85%	251	34.34%
	Male	435	67.13%	331	64.15%	480	65.66%
	Low Income	381	58.80%	262	50.78%	402	55.07%
	Non-Low Income	267	41.20%	254	49.22%	328	44.93%
	Bilingual	123	18.98%	90	17.44%	101	13.84%
	Non-Bilingual	525	81.02%	426	82.56%	629	86.16%
	Hispanic/Latino	175	27.01%	125	24.22%	155	21.23%
10	Non-Hispanic/Latino	473	72.99%	391	75.78%	575	78.77%
	American Indian or Alaska Native	12	1.85%	5	0.97%	14	1.92%
	Asian	63	9.72%	44	8.53%	58	7.95%
	Black or African American	49	7.56%	36	6.98%	43	5.89%
	Hispanic or Latino	175	27.01%	125	24.22%	155	21.23%
	Caucasian or White	290	44.75%	254	49.22%	399	54.66%
	Native Hawaiian or Other Pacific Islander	6	0.93%	12	2.33%	4	0.55%
	Of More than One Race/Multi-Racial	53	8.18%	40	7.75%	57	7.81%

Science

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
5	Female	187	29.50%	177	31.49%	240	32.61%
	Male	447	70.50%	384	68.33%	496	67.39%
	Low Income	409	64.51%	334	59.43%	455	61.82%
	Non-Low Income	225	35.49%	228	40.57%	281	38.18%
	Bilingual	159	25.08%	115	20.46%	144	19.57%
	Non-Bilingual	475	74.92%	447	79.54%	592	80.43%
	Hispanic/Latino	211	33.28%	167	29.72%	217	29.48%
	Non-Hispanic/Latino	423	66.72%	395	70.28%	519	70.52%
	American Indian or Alaska Native	7	1.10%	4	0.71%	7	0.95%
	Asian	46	7.26%	45	8.01%	48	6.52%
	Black or African American	51	8.04%	41	7.30%	54	7.34%
	Hispanic or Latino	211	33.28%	167	29.72%	217	29.48%
	Caucasian or White	247	38.96%	249	44.31%	337	45.79%
	Native Hawaiian or Other Pacific Islander	9	1.42%	6	1.07%	10	1.36%
	Of More than One Race/Multi-Racial	63	9.94%	50	8.90%	63	8.56%
8	Female	204	34.81%	185	35.31%	223	33.79%
	Male	382	65.19%	339	64.69%	437	66.21%
	Low Income	373	63.65%	329	62.79%	377	57.21%
	Non-Low Income	213	36.35%	195	37.21%	282	42.79%
	Bilingual	134	22.87%	109	20.80%	105	15.93%
	Non-Bilingual	452	77.13%	415	79.20%	554	84.07%
	Hispanic/Latino	182	31.06%	164	31.30%	156	23.67%
	Non-Hispanic/Latino	404	68.94%	360	68.70%	503	76.33%
	American Indian or Alaska Native	7	1.19%	12	2.29%	8	1.21%
	Asian	36	6.14%	34	6.49%	64	9.71%
	Black or African American	43	7.34%	42	8.02%	44	6.68%
	Hispanic or Latino	182	31.06%	164	31.30%	156	23.67%
	Caucasian or White	255	43.52%	237	45.23%	331	50.23%
	Native Hawaiian or Other Pacific Islander	13	2.22%	5	0.95%	11	1.67%
	Of More than One Race/Multi-Racial	50	8.53%	30	5.73%	45	6.83%
11	Female	222	36.27%	180	37.89%	190	31.05%
	Male	390	63.73%	295	62.11%	422	68.95%
	Low Income	322	52.61%	277	58.32%	366	60.30%
	Non-Low Income	290	47.39%	198	41.68%	241	39.70%
	Bilingual	115	18.79%	92	19.37%	82	13.51%
	Non-Bilingual	497	81.21%	383	80.63%	525	86.49%
	Hispanic/Latino	139	22.71%	116	24.42%	129	21.25%
	Non-Hispanic/Latino	473	77.29%	359	75.58%	478	78.75%
	American Indian or Alaska Native	6	0.98%	7	1.47%	19	3.13%
	Asian	63	10.29%	43	9.05%	50	8.24%

Grade	Subgroup	2021–2022		*Reference:* 2020–2021		*Reference:* 2018–2019	
		Valid N	Percent	Valid N	Percent	Valid N	Percent
	Black or African American	42	6.86%	21	4.42%	40	6.59%
	Hispanic or Latino	139	22.71%	116	24.42%	129	21.25%
	Caucasian or White	301	49.18%	260	54.74%	322	53.05%
	Native Hawaiian or Other Pacific Islander	10	1.63%	3	0.63%	8	1.32%
	Of More than One Race/Multi-Racial	51	8.33%	25	5.26%	39	6.43%

6.3 Student Characteristics of the Participants

The WA-AIM Student Characteristics Survey (SCS) was modified from the Learner Characteristics Inventory (LCI) survey (Kearns, Kleinert, Kleinert, & Towles-Reeves, 2006) and is intended to gather information about student characteristics from the following perspectives:

- Communication
- Expressive and receptive communication
- Use of alternative communication system(s)
- Attendance/health status
- Instructional time per week
- Teacher estimated student alternate achievement level
- WA-AIM representation of students' IEP goals/objectives
- English learner status and related questions on language use and support

The survey was administered online to educators during the testing window of the 2021–2022 WA-AIM and was required for all participating students. A high-level summary of the SCS survey results on the tested students is presented in Tables 6.3 through 6.16. Available statistics from the 2018–2019 and 2020–2021 WA-AIM SCS results are also included in the tables for reference.

Table 6.3. SCS on Tested Students: Communication

Communication	2021–2022	*Reference:* 2020–2021	*Reference:* 2018–2019
Does not alert to others	1.4%	1.3%	1.6%
Alerts to Others	14.0%	15.5%	15.0%
Initiates and sustains social interactions	48.3%	46.2%	46.6%
Responds with social interaction, but does not initiate or sustain social interactions	36.3%	37.0%	36.8%

Table 6.4. SCS on Tested Students: Primary Mode of Expressive Communication

Expressive Communication	2021–2022	*Reference:* 2020–2021	*Reference:* 2018–2019
Student communicates primarily through cries, facial expressions, change in muscle tone, etc., but no clear use of objects/textures, regularized gestures, pictures, signs, etc., to communicate	5.4%	5.4%	5.1%
Student uses intentional communication, but not at a symbolic level; uses understandable communication through such modes as gestures, pictures, objects/textures, points, etc., to clearly express a variety of intentions	29.9%	25.4%	24.9%
Student uses symbolic language to communicate; uses verbal or written words, signs, Braille, or language-based augmentative systems to request, initiate, and respond to questions, describe things or events, and express refusal	64.7%	69.2%	70.0%

Table 6.5. SCS on Tested Students: Augmentative/Alternative Communication

Augmentative/Alternative Expressive Communication	2021–2022	*Reference:* 2020–2021	*Reference:* 2018–2019
Yes	27.6%	24.5%	22.4%
No	53.2%	54.6%	42.3%
Does Not Apply	19.2%	20.9%	35.3%

Table 6.6. SCS on Tested Students: Receptive Communication

Receptive Communication	2021–2022	*Reference:* 2020–2021	*Reference:* 2018–2019
Has uncertain response to sensory stimuli (e.g., sound/voice; sight/gesture; touch; movement; smell)	1.3%	1.5%	1.3%
Alerts to sensory input from another person (auditory, visual, touch, movement) BUT requires actual physical assistance to follow simple directions	6.7%	7.5%	5.5%
Requires additional cues (e.g., gestures, pictures, objects, or demonstrations/models) to follow 1-2 step directions	43.6%	44.2%	43.0%
Independently follows 1-2 step directions presented through words (e.g. words may be spoken, signed, printed, or any combination) and does NOT need additional cues	48.4%	46.9%	50.2%

Table 6.7. SCS on Tested Students: Attendance/Health Status

Attendance/Health Status	2021–2022	*Reference:* 2020–2021	*Reference:* 2018–2019
Attends approximately 50% or less of school days; absences primarily due to health issues	2.8%	2.4%	1.8%
Attends approximately 75% of school days; absences primarily due to health issues	14.1%	11.1%	12.1%
Attends at least 90% of school days	79.8%	82.7%	84.3%
Attends highly irregularly or receives homebound instruction due to issues other than health	3.1%	3.3%	1.7%
Receives homebound instruction due to health issues	0.3%	0.5%	0.1%

Table 6.8. SCS on Tested Students: Instructional Time Per Week

2021–2022

	Percent				
Instructional Time Per Week	***< 1 hr***	***1–3 hrs***	***4–6 hrs***	***7–9 hrs***	***≥ 10 hrs***
Reading	2.8%	23.2%	60.9%	10.8%	2.3%
Writing	5.3%	36.9%	51.2%	6.0%	0.7%
Math	3.2%	22.6%	64.7%	8.1%	1.4%
Science	45.1%	38.3%	15.4%	1.1%	0.1%

Reference: 2020–2021

	Percent				
Instructional Time Per Week	***< 1 hr***	***1–3 hrs***	***4–6 hrs***	***7–9 hrs***	***≥ 10 hrs***
Reading	3.2%	28.4%	58.1%	7.9%	2.4%
Writing	6.2%	40.2%	47.8%	4.6%	1.2%
Math	3.4%	28.6%	60.7%	5.9%	1.4%
Science	51.0%	36.9%	11.3%	0.6%	0.2%

Reference: 2018–2019

Instructional Time Per Week	Percent			
	0–3 hrs	***4–6 hrs***	***7–9 hrs***	***≥ 10 hrs***
Reading/Writing	11.5%	57.0%	21.2%	10.3%
Math	13.8%	68.5%	13.8%	3.8%
Science	70.5%	25.8%	2.8%	0.9%
Behavioral	41.2%	38.5%	11.7%	8.6%
Daily Living	34.8%	43.4%	13.2%	8.5%

Table 6.9. SCS on Tested Students: Estimated Alternate Achievement Level

2021–2022

Content	Percent				
	Level 1	***Level 2***	***Level 3***	***Level 4***	***Not Applicable/Not Assessed***
ELA	51.5%	28.30%	13.9%	1.7%	4.6%
Math	47.7%	29.10%	16.0%	2.8%	4.5%
Science	44.1%	11.70%	3.7%	0.4%	40.1%

Reference: 2020–2021

Content	Percent				
	Level 1	***Level 2***	***Level 3***	***Level 4***	***Not Applicable/Not Assessed***
ELA	47.1%	31.2%	16.0%	2.6%	3.1%
Math	42.8%	32.2%	18.3%	3.7%	3.0%
Science	45.0%	13.8%	4.8%	0.4%	36.0%

Table 6.10. SCS on Tested Students: WA-AIM Representation of IEP Goals/Objectives

2021–2022

Content	Representation		
	Limited	***Somewhat Represented***	***Well Represented***
ELA	31.90%	48.8%	19.3%
Math	33.10%	48.4%	18.5%
Science	77.70%	18.1%	4.2%

Reference: 2020–2021

Content	Representation		
	Limited	***Somewhat Represented***	***Well Represented***
ELA	32.8%	50.5%	16.8%
Math	34.4%	49.9%	15.7%
Science	80.5%	15.6%	3.9%

Table 6.11. SCS on Tested Students: English Learner Status

English Learner Status	2021–2022	*Reference:* 2020–2021
Yes	24.1%	23.6%
No	74.2%	74.8%
I don’t know	1.6%	1.6%

Table 6.12. SCS on Tested Students: Settings Where English Learners Use English

Settings Where English Learners Use English (Multi-Select)	2021–2022 (Identified as English Learners in the Survey Only; N Count)	2021–2022 (N Count)	*Reference:* 2020–2021 (N Count)
Home	721	3371	2926
School	1217	3896	3368
Community	795	3332	2810
Other	43	326	346
I don’t know	51	76	54
Not applicable	24	1277	1102

Table 6.13. SCS on Tested Students: Settings Where English Learners Use a Language Other Than English

Settings Where English Learners Use a Language Other Than English (Multi-Select)	2021–2022 (Identified as English Learners in the Survey Only; N Count)	2021–2022 (N Count)	*Reference:* 2020–2021 (N Count)
Home	953	1254	1056
School	170	261	203
Community	454	566	485
Other	27	42	47
I don’t know	165	354	326
Not applicable	169	3655	3157

Table 6.14. SCS on Tested Students: English Language Acquisition Specialist

English Learners: English Language Acquisition Specialist on IEP Team	**2021–2022 (Identified as English Learners in the Survey Only)**	**2021–2022**	***Reference:* 2020–2021**
Yes	40%	10.5%	8.9%
No	49%	31.7%	33.6%
I don’t know	3%	1.2%	2.0%
Not applicable	8%	56.5%	55.5%

Table 6.15. SCS on Tested Students: Hours Per Week on English Language Development Instruction

Hours Per Week on English Language Development Instruction	**2021–2022 (Identified as English Learners in the Survey Only)**	**2021–2022**	***Reference:* 2020–2021**
0 hours	31%	17.7%	18.4%
Less than 1 hour	4%	6.4%	5.8%
1–2 hours	22%	3.2%	3.8%
2–3 hours	12%	1.1%	1.7%
3–4 hours	4%	2.0%	1.8%
More than 4 hours	6%	2.9%	2.7%
I don’t know	9%	1.6%	1.6%
Not applicable	12%	65.1%	64.1%

Table 6.16. SCS on Tested Students: English Language Development Service

English Language Development Service	2021–2022 (Identified as English Learners in the Survey Only)	2021–2022	*Reference:* 2020–2021
Services planned and delivered by special education teacher	45%	16.0%	18.7%
Services planned and delivered by an English language development acquisition specialist	11%	2.9%	2.2%
Services planned and/or delivered collaboratively between special education teacher and English language development acquisition specialist	16%	4.5%	4.5%
Services planned and delivered by an educator with dual certification or training (special education and English language acquisition)	3%	0.9%	0.4%
Not applicable	26%	75.8%	74.1%

6.4 Content Score Distributions

Table 6.17 summarizes the following statistics of the reported total scores for each content area by grade. Note that the test scale was designed from 100 (the lowest possible score) to 200 (the highest possible score) at each grade and content area.

- the number of valid scores (N)
- the minimum (Min.) and maximum (Max.) observed scores
- the median (Median)
- the mean (Mean) and standard deviation (SD)
- the skewness (Skew.) and kurtosis (Kurt.) of the score distribution

Student WA-AIM scores generally spread across the full test scale (100–200). Across grades and content areas, the score distributions were positively skewed (which means fewer students at the higher end of the test scale). The median total test scores ranged from 118 to 122 in ELA, from 121 to 126 in mathematics, and from 113 to 116 in science.

As described at the beginning of this chapter, a total test score of 143 on the WA-AIM, regardless of the tested grade and content area, could be roughly interpreted as the student meeting minimal mastery-level expectations across standards for the tested grade and content area. Observed score distributions in the 2021–2022 WA-AIM suggest that the tests are challenging for the tested students.

Table 6.17. Test Score Distributions, by Grade and Content Area

Content	Grade	Valid N	Min.	Max.	Median	Mean	SD	Skew.	Kurt.
ELA	3	720	100	200	119	124.38	19.92	1.29	1.59
	4	760	100	200	119	125.86	21.90	1.32	1.38
	5	643	100	200	122	126.57	21.51	1.49	2.41
	6	597	100	200	118	126.10	23.26	1.27	1.04
	7	618	100	200	118	127.01	24.62	1.17	0.47
	8	606	100	200	118	126.06	22.62	1.29	1.09
	10	650	100	200	119	124.58	18.41	1.67	3.68
Mathematics	3	719	100	200	122	129.08	23.14	1.29	1.09
	4	755	100	200	122	129.27	23.53	1.27	1.05
	5	644	100	200	122	129.39	25.00	1.28	0.89
	6	592	100	200	124	133.09	27.58	0.98	-0.09
	7	618	100	200	126	132.98	26.10	0.83	-0.34
	8	605	100	200	121	132.06	26.52	1.06	0.02
	10	649	100	200	124	129.48	20.60	1.35	1.85
Science	5	634	100	200	113	118.58	18.81	2.04	4.56
	8	587	100	200	113	119.77	19.86	1.86	3.31
	11	613	100	200	116	122.03	16.27	1.88	5.09

Table 6.18 presents information on student performance in 2015–2019, 2021 and 2022, including the mean test score (Mean), standard deviation (SD), and the pass rate (the percentage of students classified at Level 3 or above).

Table 6.18. Student Performance in 2015–2019, 2021 and 2022

ELA

Grade	Category	2015	2016	2017	2018	2019	2021	2022
3	Mean	129.50	140.03	136.26	134.72	134.08	127.24	124.38
	SD	21.99	26.46	22.70	22.40	22.03	22.82	19.92
	Pass Rate (%)	48.97	64.68	63.84	61.93	60.54	45.59	42.78
4	Mean	127.08	137.80	139.35	133.78	134.59	126.93	125.86
	SD	20.35	25.34	24.10	22.21	23.45	22.85	21.90
	Pass Rate (%)	43.50	62.67	66.75	59.23	60.02	43.60	41.97
5	Mean	129.20	139.53	141.47	136.98	136.51	128.25	126.57
	SD	20.18	22.83	23.57	22.27	23.40	22.07	21.51
	Pass Rate (%)	42.82	65.70	68.35	58.37	57.23	42.40	37.48
6	Mean	122.51	134.88	138.68	134.76	135.96	131.68	126.10
	SD	18.07	25.82	25.07	25.19	25.33	25.90	23.26
	Pass Rate (%)	33.06	55.03	63.81	56.41	60.76	51.25	41.20
7	Mean	117.58	129.33	133.19	133.85	132.97	128.23	127.01
	SD	13.68	22.27	23.42	23.85	23.51	24.50	24.62
	Pass Rate (%)	26.11	51.84	58.18	60.00	59.80	47.18	42.55
8	Mean	120.52	132.08	135.76	130.69	134.95	127.44	126.06
	SD	16.81	23.50	23.75	21.48	23.63	21.51	22.62
	Pass Rate (%)	32.59	55.71	64.16	56.85	61.76	49.35	41.42
10*	Mean				122.48	123.81	125.29	124.58
	SD				17.07	17.51	18.15	18.41
	Pass Rate (%)				34.07	36.27	43.16	40.62
11*	Mean	113.59	124.66	125.12	123.08			
	SD	11.55	16.45	15.70	18.21			
	Pass Rate (%)	14.95	49.18	50.37	35.75			

Mathematics

Grade	Category	2015	2016	2017	2018	2019	2021	2022
3	Mean	134.92	145.96	144.57	141.64	140.34	131.80	129.08
	SD	26.58	28.67	28.04	27.61	26.53	23.63	23.14
	Pass Rate (%)	46.36	65.04	64.04	59.00	59.09	45.41	38.66
4	Mean	135.41	141.80	142.69	138.45	138.22	129.71	129.27
	SD	25.68	28.23	26.03	26.16	26.09	24.10	23.53
	Pass Rate (%)	52.77	62.35	69.94	61.53	62.50	43.98	45.17
5	Mean	130.55	139.83	141.09	137.54	136.92	131.54	129.39
	SD	23.69	25.80	25.78	25.41	25.99	25.21	25.00
	Pass Rate (%)	52.87	71.79	73.80	67.97	66.15	58.04	54.04
6	Mean	131.37	142.60	149.08	146.02	144.23	139.91	133.09
	SD	25.25	29.11	29.89	29.90	29.59	30.83	27.58
	Pass Rate (%)	40.95	57.56	68.51	60.99	59.62	52.95	43.75
7	Mean	121.71	135.39	139.32	142.42	139.16	133.57	132.98
	SD	20.43	25.84	27.22	28.10	28.13	27.42	26.10
	Pass Rate (%)	32.61	57.49	62.60	67.85	60.47	54.10	52.92
8	Mean	123.59	136.40	139.72	138.84	137.58	135.37	132.06
	SD	18.19	25.56	26.04	25.92	26.08	26.61	26.52
	Pass Rate (%)	27.12	48.65	55.36	53.26	51.28	46.14	36.86
10*	Mean				132.13	131.05	132.09	129.48
	SD				20.93	20.47	21.71	20.60
	Pass Rate (%)				67.88	65.25	63.95	55.47
11*	Mean	117.48	134.57	133.85	133.25			
	SD	17.92	24.44	22.97	21.16			
	Pass Rate (%)	28.79	63.17	65.54	70.42			

Science

Grade	Category	2015	2016	2017	2018***	2019	2021	2022
5	Mean	129.01	142.25	144.39	123.72	124.23	119.63	118.58
	SD	22.93	26.77	26.83	19.05	20.19	19.52	18.81
	Pass Rate (%)	41.72	67.28	69.03	33.38	33.43	20.64	20.66
8	Mean	124.27	136.92	139.20	121.21	121.24	121.10	119.77
	SD	18.56	24.88	23.74	18.93	18.90	20.16	19.86
	Pass Rate (%)	34.88	58.32	67.44	31.39	30.61	27.86	24.36
11**	Mean		136.94	136.46	123.27	125.54	122.14	122.03
	SD		24.61	22.28	15.52	19.77	16.00	16.27
	Pass Rate (%)		62.70	65.79	33.07	38.24	30.95	30.18

Note.

** WA-AIM high school ELA and mathematics were initially tested at grade 11; the two assessments have been administered at grade 10 for accountability starting from 2018, with changed standards for ELA and no change in standards for mathematics.*

*** WA-AIM high school science (tested at grade 11) was first administered in 2016.*

**** WA-AIM science assessments aligned to new science standards were first administered in 2018.*

6.5 Achievement Level Distributions

The WA-AIM assessment reports four achievement levels in each content area: Level 1, Level 2, Level 3, and Level 4, where Level 4 represents the highest level of knowledge, skills, and understandings. Level 3 and above has been used as the criterion of "meeting standard" in the state accountability system.

The 2021–2022 WA-AIM cut scores and associated achievement level percentages (impact data) by grade and content area are presented in Table 6.19. A summary of the percentage of students at or above Level 3 in 2015–2019, 2021 and 2022 are presented in Table 6.18 of Section 6.4.

As shown in Table 6.18, the percentage of students at or above Level 3 varied across grades and content areas in the 2021–2022 administration, ranging from a low of 21% (grade 5 science) to a high of 55% (grade 10 mathematics). Across grades, the average percentage at or above Level 3 was 41% for ELA, 47% for mathematics, and 25% for science.

Table 6.19. Cut Scores and Associated Impact Data, 2021–2022 WA-AIM

Content	Grade	Cut Scores			Achievement Levels				
		Level 2	*Level 3*	*Level 4*	*Level 1*	*Level 2*	*Level 3*	*Level 4*	Proficient (Level 3 or Above)
ELA	3	109	124	150	21.53%	35.69%	31.81%	10.97%	42.78%
	4	107	125	158	15.66%	42.37%	31.71%	10.26%	41.97%
	5	108	129	162	16.17%	46.35%	30.79%	6.69%	37.48%
	6	110	125	159	27.14%	31.66%	29.98%	11.22%	41.20%
	7	108	123	154	21.36%	36.08%	27.18%	15.37%	42.55%
	8	110	123	150	25.41%	33.17%	27.89%	13.53%	41.42%
	10	110	126	162	17.69%	41.69%	35.85%	4.77%	40.62%
Mathematics	3	108	129	161	9.32%	52.02%	26.70%	11.96%	38.66%
	4	106	126	161	5.96%	48.87%	33.38%	11.79%	45.17%
	5	106	120	153	8.70%	37.27%	38.82%	15.22%	54.04%
	6	109	131	160	16.89%	39.36%	25.00%	18.75%	43.75%
	7	109	124	163	17.64%	29.45%	35.44%	17.48%	52.92%
	8	112	133	162	22.15%	40.99%	19.17%	17.69%	36.86%
	10	108	120	146	4.62%	39.91%	41.76%	13.71%	55.47%
Science	5	111	129	169	42.90%	36.44%	17.35%	3.31%	20.66%
	8	110	127	163	35.43%	40.20%	18.23%	6.13%	24.36%
	11	111	128	164	19.25%	50.57%	26.92%	3.26%	30.18%

6.6 Internal Test Consistency

With pre-built forms available for teachers to use starting with the 2020–2021 WA-AIM administration (see more details in Sections 3.4 and 3.5 of Chapter 3), it is possible to estimate internal test consistency from the following perspectives, with the purpose to provide a more complete picture on the WA-AIM internal test consistency.

- Cronbach's alpha and the associated standard error of measurement using performance task scores from all tested students
- G coefficient obtained from the $i{:}p$ design using item scores from all tested students
- Cronbach's alpha and the associated standard error of measurement using data from students who took pre-built forms

6.6.1 Cronbach's Alpha Based on Performance Task Scores

Classical Test Theory (CTT)–based statistics, such as Cronbach's alpha (1951), are typically used to estimate internal test reliability, particularly for raw score–based assessments that are

comprised of a fixed set of test items. In a more traditional assessment instrument with independently scored items and with the same items being taken by all the students, the Cronbach's alpha coefficient can be interpreted as the proportion of observed variance accounted for by the CTT model. An alpha coefficient toward the high end is taken to mean that the parts of the test are likely eliciting very similar student performance—that is, the subscore units are consistent with one another and suggest a reliable assessment.

A related measure of reliability is an estimate of the degree of measurement error in students' total score on a test, or classical standard error of measurement (SEM). It represents the number of score points about which a given score can vary due to assessment errors. The lower the SEM, the lower the variability and the higher the reliability.

It should be noted that reliability estimates such as the SEM are dependent not only on the measurement properties of a test but also on the statistical distribution of the studied student group. The greater the variance of the test scores in the student group, the higher the SEM.

In the case of the WA-AIM assessment, a student receives a Performance Task score based on five test items per standard that the teacher has selected from a state-provided item pool (see more details in Sections 3.4 and 3.5 of Chapter 3). The Performance Task score is then weighted based on the access point chosen for the student for the assessed standard. The weighted Performance Task score on the standard is intended to be comparable across students regardless of the varying access points that students are assessed on.

To calculate the Cronbach's alpha for the WA-AIM assessment, the weighted Performance Task score on each content standard is used as the subscore unit, and the alpha estimate can be viewed as an indication of the strength of association among the standard scores. As the Cronbach's alpha measures tau-equivalency of test components (e.g., items), the obtained alpha value is usually regarded as a lower bound of the internal test consistency (Cortina, 1993).

The obtained alpha coefficients, along with associated SEMs and total test score standard deviations (SDs), are presented in Table 6.20a. The alpha values in 2022 have a mean of .92 and

standard deviation of .01, ranging from .88 to .94. The observed SEMs were relatively small in magnitude. The estimates were around or below 7% of the total length of the test scale and were around or below one third of the estimated total score standard deviations.

Standard-level statistics, such as difficulty and discrimination indices, of the 2021–2022 WA-AIM assessment can be found in Section 6.9 (*Standard Statistics*) of this chapter.

6.6.2 G Coefficient

Generalizability theory (G theory) can be viewed as a comprehensive extension of the classical test theory (CTT). The focus of G theory is on isolating and estimating the relative magnitude of specific sources of measurement error (p.2, Shavelson & Webb, 1991).

Feedback from the 2016 peer review of the WA-AIM assessment questioned item variance and suggested an $i{:}p$ G-theory approach for estimating internal consistency. With the $i{:}p$ design, item score variance, rather than standard score variance (which was examined using the Cronbach's alpha approach), within each student, becomes the source of measurement error being estimated. In a G-theory study, the generalizability coefficient (the G coefficient) for relative (norm-referenced) decisions is analogous to Cronbach's alpha.

Table 6.20a presents the G coefficient obtained from the $i{:}p$ design, along with the Cronbach's alpha estimate based on performance task scores, for each grade and content area. Results show that the G coefficient is consistently higher than the Cronbach's alpha value, with a mean of .96 and standard deviation of .01, ranging from .94 to .97.

6.6.2 Cronbach's Alpha Based on Pre-Built Form Data

The Cronbach's alpha estimates for tests that used pre-built forms are provided in Table 6.20b. A summary of the alpha estimates is also presented in Table 6.20a. Note that only tests that had pre-built forms across all tested standards and had more than 50 student records were included in the analysis. The sample size for each of the tests, ranging from 53 to 336, is much smaller compared with that for estimating alpha based on performance task scores from the tested

population. The limited sample sizes may negatively impact the alpha estimates. With that said, most of the alpha values for the fixed-form tests are above 0.80, which suggests strong internal reliability of the WA-AIM assessments.

Table 6.20. WA-AIM Internal Test Consistency

a. Summary of Internal Test Consistency Indices by Content Area and Grade

		CTT Approach Based on Performance Task Scores			G Coefficient	Alpha Based on Pre-Built Form Data*	
Content	**Grade**	***Alpha***	***SEM***	***SD***	**Coefficient**	***Min.***	***Max.***
ELA	3	0.88	6.87	19.92	0.94	0.81	0.89
	4	0.89	7.37	21.90	0.95	0.87	0.88
	5	0.91	6.45	21.51	0.96	0.82	0.86
	6	0.90	7.29	23.26	0.96	0.87	0.89
	7	0.92	6.76	24.62	0.96	0.88	0.94
	8	0.91	6.75	22.62	0.96	0.89	0.91
	10	0.92	5.09	18.41	0.96	0.85	0.94
Mathematics	3	0.91	6.76	23.14	0.96	0.84	0.91
	4	0.91	7.00	23.53	0.96	0.81	0.91
	5	0.93	6.69	25.00	0.96	0.86	0.91
	6	0.94	6.95	27.58	0.97	0.90	0.91
	7	0.92	7.34	26.10	0.96	0.81	0.89
	8	0.92	7.27	26.52	0.97	0.75	0.91
	10	0.92	5.85	20.60	0.96	0.84	0.92
Science	5	0.92	5.22	18.81	0.96	0.87	0.88
	8	0.93	5.09	19.86	0.96	0.89	0.92
	11	0.93	4.43	16.27	0.96	0.83	0.87

* Only tests that had pre-built forms across all tested standards and had more than 50 student records were included.

b. Alpha Estimates for Tests with Pre-Built Forms on All Standards

Content	**Grade**	**Access Point Pattern Across Standards***	**N**	**Alpha**
ELA	3	IIIII	141	0.81
ELA	3	LLLLL	237	0.89
ELA	4	IIIII	133	0.87
ELA	4	LLLLL	224	0.88
ELA	5	IIIII	154	0.82
ELA	5	LLLLL	187	0.86
ELA	6	IIIII	137	0.89
ELA	6	LLLLL	201	0.89
ELA	6	MMMMM	54	0.87
ELA	7	IIIII	124	0.94
ELA	7	LLLLL	183	0.88
ELA	7	MMMMM	75	0.88

Content	Grade	Access Point Pattern Across Standards*	N	Alpha
ELA	8	IIIII	115	0.91
ELA	8	LLLLL	196	0.89
ELA	8	MMMMM	58	0.90
ELA	10	IIIII	268	0.87
ELA	10	LLLLL	336	0.85
ELA	10	MMMMM	53	0.94
Math	3	IIIII	153	0.84
Math	3	LLLLL	226	0.91
Math	3	MMMMM	74	0.87
Math	4	IIIII	167	0.81
Math	4	LLLLL	232	0.89
Math	4	MMMMM	74	0.91
Math	5	IIIII	151	0.86
Math	5	LLLLL	185	0.86
Math	5	MMMMM	66	0.91
Math	6	IIIII	147	0.90
Math	6	LLLLL	188	0.91
Math	6	MMMMM	74	0.91
Math	7	IIIII	136	0.89
Math	7	LLLLL	169	0.89
Math	7	MMMMM	91	0.81
Math	8	IIIII	127	0.87
Math	8	LLLLL	188	0.91
Math	8	MMMMM	88	0.75
Math	10	IIIII	320	0.86
Math	10	LLLLL	318	0.84
Math	10	MMMMM	65	0.92
Science	5	IIIII	118	0.87
Science	5	LLLLL	267	0.88
Science	8	IIIII	106	0.92
Science	8	LLLLL	270	0.89
Science	11	IIIII	232	0.87
Science	11	LLLLL	314	0.83

*The pattern shows the access point levels (L = *Less Complex*; I = *Intermediate*; M = *More Complex*) across standards for the associated test. Only tests that had pre-built forms across all tested standards and had more than 50 student records were included.

6.7 Classification Accuracy and Consistency

While related to reliability, the accuracy and consistency of classifying students into performance categories are even more important statistics in a standards-based reporting

framework (Livingston & Lewis, 1995). Decision accuracy and consistency (DAC) can usually be computed with the data currently available for most alternate assessments.

In DAC, accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated because errorless test scores do not exist. Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete and parallel forms of the test are given to the same group of students. In operational test programs, however, such a design is usually impractical. Instead, techniques have been developed to estimate both the accuracy and consistency of classification decisions based on a single administration of a test.

The Livingston and Lewis (1995) technique based on the beta-binomial model was used for the 2029–2021 WA-AIM assessment data. The technique is easily adaptable to all types of testing formats. The accuracy and consistency estimates make use of "true scores" in the CTT sense. A true score is the score that would be obtained if a test had no measurement error. Of course, true scores cannot be observed and so must be estimated. In the Livingston and Lewis method, estimated true scores are used to categorize students into their "true" classifications.

To obtain DAC statistics, a two-by-two contingency table of accuracy was created for each content area and grade, where cell [*i*, *j*] represented the estimated proportion of students whose true score fell into classification i (where $i = 1$ or 2) and whose observed score fell into classification *j* (where $j = 1$ or 2). The sum of the diagonal entries (i.e., the proportion of students whose true and observed classifications matched) signified overall accuracy.

To calculate consistency, true scores were used to estimate the joint distribution of classifications on two independent, parallel test forms. Following statistical adjustments per Livingston and Lewis (1995), a new two-by-two contingency table was created for each content area and grade and populated by the proportion of students who would be categorized into each combination of classifications according to the two (hypothetical) parallel test forms. Cell [*i*, *j*] of this table

represented the estimated proportion of students whose observed score on the first form would fall into classification *i* (where *i* = 1 or 2) and whose observed score on the second form would fall into classification *j* (where *j* = 1 or 2). The sum of the diagonal entries (i.e., the proportion of students categorized by the two forms into exactly the same classification) signified overall consistency.

Classification accuracy (or consistency) conditional on a performance level is computed as the ratio between the proportion of correct (or alternative) classifications at the selected level and the proportion of all student performance classified into that level.

The classification accuracy (or consistency) index conditional on a cut point is computed as the sum of the proportions of correct classifications around a selected cut point. For example, if the accuracy index at the cut point between Levels 2 and 3 equals 0.96, this means that 96% of student performance was correctly classified either above or below the particular cut point.

Additionally, false positive and false negative rates can be examined. A false positive rate conditional on a cut point describes the percentage of students who were classified above the cut point by their scores but fell below the cut point by their true or alternative scores. Similarly, a false negative rate conditional on a cut point shows the percentage of students with assigned performance levels below the cut point whose true or alternative levels were above the cut point.

It should be noted that Livingston and Lewis (1995) discussed two versions of the accuracy and consistency tables. A standard version performs calculations for forms parallel to the form taken. An “adjusted” version adjusts the results of one form to match the observed score distribution obtained in the data. The reported statistics use the standard version for two reasons: (1) this “unadjusted” version can be considered a smoothing of the data, thereby decreasing the variability of the results, and (2) for results dealing with the consistency of two parallel forms, the unadjusted tables are symmetrical, indicating that the two parallel forms have the same statistical properties. This second reason is consistent with the notion of forms that are parallel; that is, it is more intuitive and interpretable for two parallel forms to have the same statistical distribution.

Another way to measure consistency is to use Cohen's coefficient Kappa (1960), which assesses the proportion of consistent classifications after removing the proportion of consistent classifications that would be expected by chance. Because the Kappa is corrected for chance, its values are lower than other consistency estimates.

Empirical data plots and model fit statistics such as the Likelihood Ratio Chi-Square were inspected to determine the density function to use for estimating the true score distribution in the Livingston and Lewis procedure. Two models were considered: the two-parameter beta and the four-parameter beta. In the WA-AIM assessment context, observed student scores were spread across the scale, and conceptually it is expected to see students obtaining the lowest and the highest scores on the scale given the wide spectrum of learner characteristics of the WA-AIM test population. From the examination of empirical data, the two-parameter beta was found to be more sensitive to score distributions at the two ends of the test scale and therefore was used in the estimation.

In the WA-AIM assessment, each student was classified into one of four performance levels. Students at or above Level 3 are regarded as meeting or exceeding on-grade standards.

Tables 6.21 and 6.22 present the classification accuracy and consistency results on total scores by grade and content area. In Table 6.21, overall accuracy (Acc.) and consistency (Con.) indices, including kappa, as well as the percentages of accuracy and consistency classifications conditional on performance levels, are provided. Table 6.22 provides a summary of classification accuracy and consistency results conditional on cut points. For each cut point, the report shows the percentages of accurate/consistent classifications (% Correct), false positive rates, and false negative rates.

As shown in Table 6.21, the overall classification accuracy and consistency estimates are at or above 0.74 and 0.65, respectively, across grades and content areas. The reported kappa values range from 0.52 to 0.66. As the kappa estimate removes the probability of agreement by chance, it is expected to have a lower value than the overall consistency estimate. Kappa values below 0.40 generally indicate poor agreement.

When conditional on the performance level, Level 3 and Level 4 generally show similar or higher accuracy and consistency estimates compared to the overall estimates.

Table 6.22 shows high classification consistency and accuracy (≥ 0.85) conditional on the Level 2/Level 3 cut point across grades and content areas, which supports using Level 3 or above for high-stakes purposes. In addition, the accuracy and consistency indices at the Level 3/Level 4 cut point are high (≥ 0.91). The false positive/negative rates are low (≤ 0.07) across all three cut points.

Table 6.21. Classification Accuracy and Consistency, Conditional on Performance Level

ELA

	Overall			Level 1		Level 2		Level 3		Level 4	
Grade	***Acc.***	***Con.***	***Kappa***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***
3	0.74	0.65	0.52	0.84	0.73	0.65	0.55	0.73	0.65	0.82	0.71
4	0.75	0.66	0.52	0.82	0.70	0.68	0.59	0.77	0.69	0.82	0.69
5	0.79	0.71	0.59	0.84	0.74	0.76	0.68	0.80	0.73	0.83	0.71
6	0.76	0.68	0.56	0.87	0.78	0.61	0.50	0.77	0.69	0.83	0.72
7	0.78	0.70	0.60	0.88	0.80	0.65	0.55	0.77	0.69	0.87	0.79
8	0.76	0.68	0.56	0.87	0.79	0.60	0.49	0.75	0.66	0.86	0.77
10	0.82	0.75	0.63	0.87	0.78	0.75	0.67	0.86	0.80	0.82	0.69

Note. Acc.=Accuracy; Con.=Consistency.

Mathematics

	Overall			Level 1		Level 2		Level 3		Level 4	
Grade	***Acc.***	***Con.***	***Kappa***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***
3	0.79	0.70	0.59	0.85	0.74	0.75	0.67	0.78	0.71	0.84	0.74
4	0.79	0.70	0.58	0.82	0.69	0.74	0.66	0.80	0.73	0.84	0.74
5	0.79	0.71	0.60	0.84	0.74	0.66	0.56	0.80	0.73	0.88	0.81
6	0.80	0.72	0.63	0.87	0.79	0.74	0.65	0.75	0.66	0.88	0.82
7	0.79	0.70	0.59	0.86	0.76	0.63	0.53	0.81	0.75	0.86	0.77
8	0.79	0.71	0.61	0.88	0.81	0.71	0.61	0.74	0.65	0.87	0.79
10	0.79	0.70	0.59	0.83	0.71	0.67	0.57	0.80	0.73	0.87	0.80

Note. Acc.=Accuracy; Con.=Consistency.

Science

	Overall			Level 1		Level 2		Level 3		Level 4	
Grade	***Acc.***	***Con.***	***Kappa***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***	***Acc.***	***Con.***
5	0.84	0.77	0.65	0.91	0.86	0.73	0.64	0.84	0.77	0.81	0.65
8	0.84	0.77	0.66	0.91	0.86	0.73	0.64	0.83	0.77	0.84	0.72
11	0.84	0.77	0.66	0.88	0.81	0.79	0.72	0.86	0.80	0.80	0.64

Note. Acc.=Accuracy; Con.=Consistency.

Table 6.22. Classification Accuracy and Consistency, Conditional on Cut Point

Content Area: ELA

		Level 1/Level 2			Level 2/Level 3			Level 3/Level 4		
Grade	**Type**	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***
3	Acc.	0.90	0.04	0.06	0.90	0.05	0.05	0.95	0.03	0.02
	Con.	0.86	0.07	0.07	0.85	0.07	0.07	0.93	0.04	0.04
4	Acc.	0.90	0.04	0.06	0.90	0.05	0.05	0.96	0.03	0.02
	Con.	0.86	0.07	0.07	0.86	0.07	0.07	0.94	0.03	0.03
5	Acc.	0.91	0.03	0.05	0.91	0.04	0.04	0.97	0.02	0.01
	Con.	0.88	0.06	0.06	0.88	0.06	0.06	0.95	0.02	0.02
6	Acc.	0.90	0.04	0.06	0.91	0.05	0.05	0.96	0.03	0.02
	Con.	0.86	0.07	0.07	0.87	0.07	0.07	0.94	0.03	0.03
7	Acc.	0.92	0.03	0.05	0.92	0.04	0.04	0.95	0.03	0.02
	Con.	0.88	0.06	0.06	0.88	0.06	0.06	0.93	0.03	0.03
8	Acc.	0.91	0.04	0.06	0.91	0.04	0.05	0.95	0.03	0.02
	Con.	0.87	0.07	0.07	0.87	0.06	0.06	0.92	0.04	0.04
10	Acc.	0.92	0.03	0.05	0.92	0.04	0.04	0.98	0.01	0.01
	Con.	0.89	0.06	0.06	0.89	0.06	0.06	0.97	0.01	0.01

Note. Acc.=Accuracy; Con.=Consistency.

Content Area: Mathematics

		Level 1/Level 2			Level 2/Level 3			Level 3/Level 4		
Grade	**Type**	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***
3	Acc.	0.92	0.03	0.05	0.91	0.04	0.04	0.96	0.03	0.02
	Con.	0.89	0.06	0.06	0.88	0.06	0.06	0.94	0.03	0.03
4	Acc.	0.92	0.03	0.05	0.91	0.04	0.05	0.96	0.03	0.02
	Con.	0.89	0.06	0.06	0.87	0.06	0.06	0.94	0.03	0.03
5	Acc.	0.92	0.03	0.05	0.92	0.04	0.04	0.95	0.03	0.02
	Con.	0.89	0.06	0.06	0.88	0.06	0.06	0.93	0.04	0.04
6	Acc.	0.93	0.03	0.04	0.93	0.04	0.04	0.95	0.03	0.02
	Con.	0.90	0.05	0.05	0.90	0.05	0.05	0.93	0.04	0.04
7	Acc.	0.92	0.03	0.05	0.91	0.04	0.05	0.95	0.03	0.02
	Con.	0.89	0.05	0.05	0.88	0.06	0.06	0.93	0.04	0.04
8	Acc.	0.92	0.03	0.05	0.92	0.04	0.04	0.95	0.03	0.02
	Con.	0.89	0.06	0.06	0.89	0.05	0.05	0.93	0.04	0.04
10	Acc.	0.93	0.02	0.04	0.91	0.04	0.05	0.94	0.03	0.03
	Con.	0.91	0.05	0.05	0.88	0.06	0.06	0.91	0.04	0.04

Note. Acc.=Accuracy; Con.=Consistency.

Content Area: Science

		Level 1/Level 2			Level 2/Level 3			Level 3/Level 4		
Grade	**Type**	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***	***% Correct***	***False Positive***	***False Negative***
5	Acc.	0.91	0.04	0.05	0.94	0.03	0.03	0.99	0.01	0.00
	Con.	0.87	0.06	0.06	0.91	0.04	0.04	0.99	0.01	0.01
8	Acc.	0.92	0.04	0.05	0.94	0.03	0.03	0.98	0.01	0.01
	Con.	0.88	0.06	0.06	0.91	0.04	0.04	0.98	0.01	0.01
11	Acc.	0.92	0.03	0.05	0.93	0.04	0.03	0.99	0.01	0.00
	Con.	0.88	0.06	0.06	0.90	0.05	0.05	0.99	0.01	0.01

Note. Acc.=Accuracy; Con.=Consistency.

6.8 Subgroup Statistics

To examine subgroup performance and potential relationship between subgroup categories and test performance on the WA-AIM, subgroup summary statistics (N, median, mean, and SD) and reliability statistics (coefficient alphas and associated SEMs) are presented in Tables 6.23 through 6.25 for each grade and content area. The subgroups of interest include

- gender (*Female* vs. *Male*)
- income status (*Low Income* vs. *Non-Low Income*)
- English learner (EL) status (*EL* vs. *Non-EL*)
- race and ethnicity (i.e., *American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Caucasian or White, Native Hawaiian or Other Pacific Islander,* and *Of More than One Race/Multi-Racial*).

Test data for subgroups with ten or less students were not reported for the reason of data confidentiality.

Generally, the sample coefficient alpha obtained from larger samples tends to produce a more accurate estimate of the population coefficient alpha, and very small sample sizes may result in misleading alpha values. Therefore, reliability statistics are not reported for subgroups with 50 or less students. In addition, readers should interpret with caution reported reliability statistics for subgroups with a small sample size (N < 200).

Overall, the alpha reliability coefficients for ELA across reported subgroups ranged from 0.77 to 0.94. For mathematics, the reliability coefficients ranged from 0.84 to 0.94. For science, the reliability coefficients ranged from 0.84 to 0.94. The test reliability was of reasonable range for reported subgroups, taking into consideration of their sample sizes.

Independent t-tests were performed to detect statistical significance ($p < .05$) of subgroup score difference. Because the t-test informs us only whether a subgroup effect exists and is also known for being sensitive to sample size, Cohen's d, a standardized effect size measurement that is independent of sample size, is reported to evaluate the size of the subgroup effect. To interpret the effect size, the following criteria suggested by Cohen (1988) are used. Cohen's d absolute values that are less than .20 are regarded as trivial.

Cohen's d absolute value ([Cohen's d|): small ≥ .20, medium ≥ .50, large ≥ .80

Note that due to relatively small sample sizes of certain ethnicity categories such as "American Indian or Alaska Native" and "Asian", detailed ethnicity categories were not included in the t-test or Cohen's d analyses; rather, all non-Hispanic categories were grouped into the "Non-Hispanic/Latino" category and compared with the Hispanic/Latino group.

The t-test and Cohen's d statistics are presented in Table 6.26. The t-test statistics indicate significant score differences ($p < .05$) on the following grade and content area combinations for each associated subgroup pair; however, the effect size is trivial or small across all tests and subgroup pairs.

Female vs. Male

- ELA grade 5
- Math grade 5

Low Income vs. Non-Low Income

- Science grade 5

EL vs. Non-EL

- ELA grade 7
- Math grade 7
- Science grade 11

Hispanic/Latino vs. Non-Hispanic/Latino

- Math grade 6
- Math grade 10
- Science grade 5
- Science grade 11

To investigate whether subgroups impact student standard-level scores, differential item functioning (DIF) was performed. DIF analyses have the technical advantage of supporting a systematic item analysis to determine whether examinees with the same underlying level of ability have the same probability of getting the item correct. In this case, the "items" of interest are WA-AIM tested standards.

Typical DIF procedures such as Mantel-Haenszel and SIBTEST were not used in this case as they would require a moderately long matching test to be valid, and the WA-AM administered only five performance tasks (one performance task per standard) to each student.

Rather, the Standardized Mean Difference (SMD) statistic (Dorans & Schmitt, 1991) and effect size (ES) statistic were used descriptively to identify standards that demonstrate the most evidence of DIF.

The SMD statistic has been widely used in DIF analyses as a descriptive measure. It calculates differences in item (in this case, the tested WA-AIM standard) mean scores between the focal and reference groups for each possible score of the matching variable (in this case, the WA-AIM total test score), and then the weighted average of these differences as the standardized mean difference for that item (the WA-AIM tested standard), where the relative frequency of the focal group at each matching score serves as the weighting function. A negative SMD value implies

that the focal group has a lower mean standard-level score than the reference group, whereas a positive value implies that the focal group has a higher mean standard-level score than the reference group, conditioned on the matching test score.

The ES is obtained by dividing the SMD statistic by the standard deviation of the item (the WA-AIM tested standard). A rough criterion to describe DIF for polytomous items has been to flag any item with an effective size of at least .25 (absolute value) ($|ES| \geq .25$) as large DIF for attention, and between .17 and .25 ($.17 \leq |ES| < .25$) as moderate DIF. The SMD and ES statistics for subgroup pairs are presented in Table 6.27.

Note that in DIF analyses, sufficient sample sizes in both focal and reference groups are necessary in order to detect differences in performance across groups matched on ability. As in the t-test analyses, detailed ethnicity categories were not included in the DIF analyses; instead, the "Non-Hispanic/Latino" group was compared with the Hispanic/Latino group for greater sample sizes. Readers should interpret with caution reported ESs where the focal or reference group sample size is less than 200.

As shown in Table 6.27, only one standard was flagged with large DIF. The standard was *The Number System / Real Number System* at high school Mathematics, and the flagged DIF was on ELL status. Nevertheless, high school Mathematics showed that the total test score difference on the subgroup pair (*ELL* vs. *Non-ELL*) was not significant and small (see Table 6.26).

Overall, the subgroup statistics suggest that some subgroups of interest may have performed differently on the WA-AIM at some grade and content area combinations, but the total test score difference was small or trivial in all cases, and no significant total test score difference was associated with a particular tested standard being flagged for DIF at the given grade and content area combination.

Table 6.23. Subgroup Score Statistics, ELA

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
3	Female	214	121	125.57	20.29	0.88	6.93
	Male	506	118	123.88	19.76	0.88	6.83
	Low Income	445	118	124.02	19.82	0.89	6.61
	Non-Low Income	275	120	124.96	20.11	0.87	7.14
	EL	167	121	123.56	18.36	0.87	6.54
	Non-EL	553	118	124.63	20.38	0.88	6.97
	Hispanic/Latino	219	122	125.47	20.05	0.88	7.04
	Non-Hispanic/Latino	501	118	123.91	19.87	0.88	6.79
	American Indian or Alaska Native	9	_	_	_	_	_
	Asian	62	120	122.76	19.10	0.89	6.21
	Black or African American	62	115	118.90	14.39	0.83	5.93
	Hispanic or Latino	219	122	125.47	20.05	0.88	7.04
	Caucasian or White	286	118	124.89	20.80	0.90	6.62
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	73	118	125.44	20.96	0.84	8.30
4	Female	269	118	125.46	21.72	0.90	6.91
	Male	490	119	126.10	22.02	0.88	7.60
	Low Income	494	119	126.78	22.97	0.89	7.46
	Non-Low Income	265	118	124.17	19.69	0.87	7.14
	EL	191	118	125.49	21.51	0.87	7.75
	Non-EL	568	119	126.00	22.05	0.89	7.23
	Hispanic/Latino	214	119	125.38	21.34	0.87	7.76
	Non-Hispanic/Latino	545	119	126.06	22.14	0.89	7.21
	American Indian or Alaska Native	12	111	110.75	5.17	_	_
	Asian	63	117	123.62	19.37	0.89	6.42
	Black or African American	59	117	126.10	23.22	0.90	7.30
	Hispanic or Latino	214	119	125.38	21.34	0.87	7.76
	Caucasian or White	327	120	127.21	22.99	0.89	7.52
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	75	119	124.84	19.66	0.88	6.81
5	Female	188	118	123.12	18.88	0.91	5.54
	Male	455	124	127.99	22.37	0.91	6.78
	Low Income	416	122	126.91	21.94	0.92	6.35
	Non-Low Income	227	123	125.94	20.72	0.90	6.62
	EL	160	120	125.56	21.70	0.93	5.59
	Non-EL	483	124	126.90	21.45	0.90	6.70
	Hispanic/Latino	213	124	127.83	23.78	0.94	5.96
	Non-Hispanic/Latino	430	122	125.94	20.28	0.89	6.67
	American Indian or Alaska Native	8	_	_	_	_	_
	Asian	44	115	119.98	16.57	_	_

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
	Black or African American	52	122	128.83	23.53	0.92	6.86
	Hispanic or Latino	213	124	127.83	23.78	0.94	5.96
	Caucasian or White	255	123	126.39	21.07	0.90	6.51
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	62	128	128.16	16.84	0.77	8.06
6	Female	170	116	123.78	21.35	0.87	7.68
	Male	426	119	126.85	23.70	0.91	7.12
	Low Income	371	116	125.62	23.51	0.91	6.89
	Non-Low Income	226	120	126.88	22.86	0.88	7.91
	EL	116	118	125.41	22.04	0.90	7.09
	Non-EL	481	117	126.26	23.56	0.90	7.32
	Hispanic/Latino	169	119	127.80	24.77	0.92	7.15
	Non-Hispanic/Latino	428	117	125.42	22.62	0.89	7.34
	American Indian or Alaska Native	5	_	_	_	_	_
	Asian	41	115	122.98	22.66	_	_
	Black or African American	46	114	120.52	20.69	_	_
	Hispanic or Latino	169	119	127.80	24.77	0.92	7.15
	Caucasian or White	277	119	127.10	23.54	0.89	7.73
	Native Hawaiian or Other Pacific Islander	3	_	_	_	_	_
	Of More than One Race/Multi-Racial	56	117	122.95	19.39	0.89	6.56
7	Female	201	119	127.23	24.63	0.93	6.57
	Male	417	117	126.90	24.64	0.92	6.85
	Low Income	397	118	127.37	23.80	0.92	6.83
	Non-Low Income	221	116	126.36	26.07	0.94	6.57
	EL	157	114	123.62	23.12	0.92	6.56
	Non-EL	461	118	128.17	25.03	0.93	6.83
	Hispanic/Latino	195	119	126.96	24.34	0.93	6.60
	Non-Hispanic/Latino	423	117	127.04	24.77	0.92	6.83
	American Indian or Alaska Native	14	114	116.00	8.03	_	_
	Asian	44	110	118.57	23.39	_	_
	Black or African American	40	119	125.75	21.00	_	_
	Hispanic or Latino	195	119	126.96	24.34	0.93	6.60
	Caucasian or White	271	118	128.37	25.11	0.92	7.09
	Native Hawaiian or Other Pacific Islander	12	117	127.17	23.81	_	_
	Of More than One Race/Multi-Racial	42	120	132.19	28.99	_	_
8	Female	212	117	126.57	23.27	0.91	7.08
	Male	393	118	125.73	22.28	0.92	6.46
	Low Income	384	118	126.18	22.58	0.91	6.78
	Non-Low Income	221	118	125.76	22.74	0.92	6.58
	EL	137	119	125.26	19.74	0.87	7.19
	Non-EL	468	118	126.25	23.41	0.92	6.53

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
	Hispanic/Latino	186	120	127.83	22.36	0.89	7.31
	Non-Hispanic/Latino	419	117	125.22	22.71	0.92	6.39
	American Indian or Alaska Native	8	_	_	_	_	_
	Asian	36	114	116.33	13.43	_	_
	Black or African American	45	114	120.36	20.26	_	_
	Hispanic or Latino	186	120	127.83	22.36	0.89	7.31
	Caucasian or White	265	118	126.12	23.33	0.92	6.46
	Native Hawaiian or Other Pacific Islander	13	134	137.23	24.78	_	_
	Of More than One Race/Multi-Racial	52	115	126.44	23.39	0.91	7.21
	Female	213	118	125.56	20.25	0.94	4.79
	Male	436	120	124.15	17.44	0.91	5.20
	Low Income	382	119	124.51	17.83	0.91	5.23
	Non-Low Income	267	120	124.76	19.23	0.94	4.87
	EL	123	117	121.97	16.07	0.90	5.09
	Non-EL	526	120	125.23	18.87	0.93	5.09
	Hispanic/Latino	175	118	122.63	16.28	0.92	4.65
10	Non-Hispanic/Latino	474	120	125.34	19.09	0.92	5.25
	American Indian or Alaska Native	12	131	134.00	24.11	_	_
	Asian	62	118	122.03	16.02	0.93	4.36
	Black or African American	49	118	122.20	17.00	_	_
	Hispanic or Latino	175	118	122.63	16.28	0.92	4.65
	Caucasian or White	291	121	126.65	20.23	0.93	5.43
	Native Hawaiian or Other Pacific Islander	6	_	_	_	_	_
	Of More than One Race/Multi-Racial	54	117	122.94	16.32	0.89	5.38

Note.
*Test data for subgroups of ten or less students are not reported for the purpose of data confidentiality.
** Reliability statistics are not reported for subgroups with 50 or less students.

Table 6.24. Subgroup Score Statistics, Mathematics

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
3	Female	214	124	128.93	23.18	0.93	6.25
	Male	505	121	129.14	23.15	0.91	6.96
	Low Income	445	123	129.02	22.67	0.91	6.79
	Non-Low Income	274	121	129.18	23.92	0.92	6.68
	EL	167	124	128.71	22.30	0.92	6.40
	Non-EL	552	121	129.19	23.41	0.91	6.86
	Hispanic/Latino	219	124	131.37	24.40	0.91	7.28
	Non-Hispanic/Latino	500	121	128.07	22.52	0.92	6.51
	American Indian or Alaska Native	9	_	_	_	_	_
	Asian	61	124	128.05	22.42	0.93	6.04
	Black or African American	62	118	122.94	16.49	0.85	6.45
	Hispanic or Latino	219	124	131.37	24.40	0.91	7.28
	Caucasian or White	286	121	128.99	23.41	0.92	6.57
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	73	124	129.32	23.49	0.92	6.75
4	Female	269	120	127.89	23.00	0.92	6.43
	Male	485	124	130.07	23.83	0.91	7.28
	Low Income	489	122	130.19	24.07	0.91	7.11
	Non-Low Income	265	122	127.63	22.50	0.91	6.79
	EL	190	122	129.53	23.78	0.90	7.37
	Non-EL	564	122	129.21	23.48	0.91	6.85
	Hispanic/Latino	213	122	129.40	23.91	0.91	7.29
	Non-Hispanic/Latino	541	122	129.25	23.42	0.91	6.87
	American Indian or Alaska Native	12	116	115.92	5.23	_	_
	Asian	63	122	127.71	19.26	0.84	7.80
	Black or African American	59	119	126.24	22.66	0.93	6.20
	Hispanic or Latino	213	122	129.40	23.91	0.91	7.29
	Caucasian or White	323	124	130.86	24.63	0.92	6.89
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	75	122	127.61	22.75	0.90	7.12
5	Female	188	118	125.22	21.08	0.93	5.71
	Male	456	122	131.10	26.27	0.93	7.05
	Low Income	415	122	130.28	25.58	0.93	6.75
	Non-Low Income	229	121	127.76	23.87	0.92	6.59
	EL	160	120	128.25	24.80	0.94	6.29
	Non-EL	484	122	129.76	25.08	0.93	6.81
	Hispanic/Latino	212	125	131.86	27.04	0.94	6.60
	Non-Hispanic/Latino	432	121	128.17	23.87	0.92	6.72
	American Indian or Alaska Native	8	_	_	_	_	_
	Asian	46	117	122.50	21.87	_	_

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
	Black or African American	52	121	128.02	24.47	0.93	6.61
	Hispanic or Latino	212	125	131.86	27.04	0.94	6.60
	Caucasian or White	255	121	128.88	24.28	0.93	6.64
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	62	122	130.84	23.22	0.88	7.93
6	Female	169	120	129.72	25.80	0.92	7.21
	Male	422	127	134.28	28.03	0.94	6.84
	Low Income	369	124	132.46	27.42	0.94	6.95
	Non-Low Income	223	127	134.15	27.88	0.94	6.90
	EL	115	128	134.17	28.26	0.93	7.40
	Non-EL	477	124	132.83	27.44	0.94	6.84
	Hispanic/Latino	168	131	136.86	29.78	0.94	7.06
	Non-Hispanic/Latino	424	122	131.60	26.55	0.93	6.90
	American Indian or Alaska Native	5	_	_	_	_	_
	Asian	41	120	128.32	26.13	_	_
	Black or African American	45	117	127.67	28.65	_	_
	Hispanic or Latino	168	131	136.86	29.78	0.94	7.06
	Caucasian or White	275	124	132.70	26.61	0.93	7.24
	Native Hawaiian or Other Pacific Islander	3	_	_	_	_	_
	Of More than One Race/Multi-Racial	55	126	132.98	26.07	0.94	6.18
7	Female	200	126	131.32	23.75	0.91	7.16
	Male	418	127	133.77	27.15	0.93	7.42
	Low Income	396	128	133.33	25.27	0.91	7.60
	Non-Low Income	222	124	132.35	27.57	0.94	6.85
	EL	155	122	128.94	23.90	0.91	7.23
	Non-EL	463	127	134.33	26.69	0.92	7.39
	Hispanic/Latino	195	132	134.46	26.36	0.92	7.45
	Non-Hispanic/Latino	423	126	132.29	25.99	0.92	7.29
	American Indian or Alaska Native	14	119	123.79	14.54	_	_
	Asian	44	118	120.70	18.72	_	_
	Black or African American	39	121	131.21	24.15	_	_
	Hispanic or Latino	195	132	134.46	26.36	0.92	7.45
	Caucasian or White	272	126	134.04	26.81	0.92	7.56
	Native Hawaiian or Other Pacific Islander	12	118	134.83	36.36	_	_
	Of More than One Race/Multi-Racial	42	130	136.24	25.73	_	_
8	Female	211	120	131.18	26.24	0.92	7.26
	Male	393	122	132.57	26.71	0.93	7.27
	Low Income	382	121	131.80	26.11	0.92	7.21
	Non-Low Income	222	122	132.58	27.30	0.93	7.36
	EL	137	121	131.14	24.85	0.91	7.39
	Non-EL	467	122	132.36	27.03	0.93	7.24

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
	Hispanic/Latino	184	124	134.33	26.42	0.92	7.53
	Non-Hispanic/Latino	420	120	131.10	26.56	0.93	7.11
	American Indian or Alaska Native	8	_	_	_	_	_
	Asian	36	116	119.50	13.12	_	_
	Black or African American	45	117	124.73	22.11	_	_
	Hispanic or Latino	184	124	134.33	26.42	0.92	7.53
	Caucasian or White	266	122	132.74	27.55	0.93	7.32
	Native Hawaiian or Other Pacific Islander	13	137	144.62	30.46	_	_
	Of More than One Race/Multi-Racial	52	119	132.17	27.99	0.94	6.99
	Female	213	123	129.43	21.21	0.93	5.51
	Male	435	126	129.54	20.33	0.91	6.01
	Low Income	381	126	129.02	18.73	0.89	6.13
	Non-Low Income	267	122	130.21	23.03	0.94	5.44
	EL	123	120	126.89	17.35	0.89	5.67
	Non-EL	525	126	130.12	21.26	0.92	5.85
	Hispanic/Latino	175	124	126.45	16.24	0.86	6.02
10	Non-Hispanic/Latino	473	125	130.64	21.91	0.93	5.79
	American Indian or Alaska Native	12	127	136.17	29.77	_	_
	Asian	63	124	129.60	21.59	0.93	5.51
	Black or African American	49	118	125.53	18.23	_	_
	Hispanic or Latino	175	124	126.45	16.24	0.86	6.02
	Caucasian or White	290	128	132.21	22.84	0.93	6.06
	Native Hawaiian or Other Pacific Islander	6	_	_	_	_	_
	Of More than One Race/Multi-Racial	53	120	127.11	18.14	0.91	5.46

Note.
*Test data for subgroups of ten or less students are not reported for the purpose of data confidentiality.
** Reliability statistics are not reported for subgroups with 50 or less students.

Table 6.25. Subgroup Score Statistics, Science

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
5	Female	187	112	117.09	17.01	0.92	4.78
	Male	447	113	119.20	19.49	0.92	5.38
	Low Income	409	113	119.88	19.88	0.92	5.49
	Non-Low Income	225	112	116.22	16.46	0.92	4.65
	EL	159	112	117.88	19.17	0.94	4.71
	Non-EL	475	113	118.81	18.70	0.92	5.36
	Hispanic/Latino	211	114	121.06	20.92	0.92	6.00
	Non-Hispanic/Latino	423	112	117.34	17.55	0.93	4.77
	American Indian or Alaska Native	7	_	_	_	_	_
	Asian	46	108	110.85	11.81	_	_
	Black or African American	51	110	115.65	15.40	0.92	4.39
	Hispanic or Latino	211	114	121.06	20.92	0.92	6.00
	Caucasian or White	247	113	119.23	19.62	0.94	4.89
	Native Hawaiian or Other Pacific Islander	9	_	_	_	_	_
	Of More than One Race/Multi-Racial	63	113	117.25	14.36	0.84	5.69
8	Female	204	113	119.98	20.41	0.94	5.03
	Male	382	113	119.66	19.62	0.93	5.11
	Low Income	373	113	120.15	20.01	0.94	5.03
	Non-Low Income	213	112	119.10	19.69	0.93	5.19
	EL	134	113	120.02	19.76	0.93	5.12
	Non-EL	452	113	119.69	19.94	0.94	5.07
	Hispanic/Latino	182	115	120.84	19.18	0.93	5.24
	Non-Hispanic/Latino	404	112	119.29	20.20	0.94	5.01
	American Indian or Alaska Native	7	_	_	_	_	_
	Asian	36	110	110.64	7.53	_	_
	Black or African American	43	111	116.56	19.75	_	_
	Hispanic or Latino	182	115	120.84	19.18	0.93	5.24
	Caucasian or White	255	112	120.88	21.48	0.94	5.33
	Native Hawaiian or Other Pacific Islander	13	115	122.54	18.42	_	_
	Of More than One Race/Multi-Racial	50	111	116.44	15.62	_	_
11	Female	222	117	122.45	16.50	0.93	4.41
	Male	390	116	121.84	16.16	0.92	4.45
	Low Income	322	117	121.94	16.27	0.93	4.22
	Non-Low Income	290	116	122.19	16.30	0.92	4.65
	EL	115	115	117.82	12.59	0.92	3.54
	Non-EL	497	117	123.04	16.87	0.93	4.62
	Hispanic/Latino	139	115	119.27	13.83	0.90	4.44
	Non-Hispanic/Latino	473	117	122.88	16.85	0.93	4.43
	American Indian or Alaska Native	6	_	_	_	_	_

Grade	Subgroup	Valid N	Median	Mean	SD	Alpha	SEM
	Asian	63	116	122.37	16.62	0.93	4.24
	Black or African American	42	118	121.71	13.39	_	_
	Hispanic or Latino	139	115	119.27	13.83	0.90	4.44
	Caucasian or White	301	116	122.41	17.04	0.94	4.32
	Native Hawaiian or Other Pacific Islander	10	_	_	_	_	_
	Of More than One Race/Multi-Racial	51	119	128.20	19.17	0.93	5.17

Note.
*Test data for subgroups of ten or less students are not reported for the purpose of data confidentiality.
** Reliability statistics are not reported for subgroups with 50 or less students.

Table 6.26. Subgroup Pairs T-Test and Cohen's D Statistics

Female vs. Male

Content	Grade	T Value	Degree of Freedom	P Value	|Cohen's D|	Magnitude of Effect Size
ELA	3	1.04	718	0.30	0.08	Trivial
	4	-0.38	757	0.70	0.03	Trivial
	5	-2.63	641	0.01*	0.23	Small
	6	-1.46	594	0.14	0.13	Small
	7	0.16	616	0.88	0.01	Trivial
	8	0.44	603	0.66	0.04	Trivial
	10	0.92	647	0.36	0.08	Trivial
Mathematics	3	-0.11	717	0.91	0.01	Trivial
	4	-1.22	752	0.22	0.09	Trivial
	5	-2.73	642	0.01*	0.24	Small
	6	-1.83	589	0.07	0.17	Small
	7	-1.09	616	0.27	0.09	Trivial
	8	-0.61	602	0.54	0.05	Trivial
	10	-0.07	646	0.95	0.01	Trivial
Science	5	-1.29	632	0.20	0.11	Small
	8	0.19	584	0.85	0.02	Trivial
	11	0.45	610	0.65	0.04	Trivial

** Significant at the .05 level.*

Low Income vs. Non-Low Income

Content	Grade	T Value	Degree of Freedom	P Value	|Cohen's D|	Magnitude of Effect Size
ELA	3	-0.62	718	0.54	0.05	Trivial
	4	1.57	757	0.12	0.12	Small
	5	0.54	641	0.59	0.04	Trivial
	6	-0.75	594	0.45	0.06	Trivial
	7	0.49	616	0.63	0.04	Trivial
	8	0.22	603	0.83	0.02	Trivial
	10	-0.17	647	0.86	0.01	Trivial
Mathematics	3	-0.09	717	0.93	0.01	Trivial
	4	1.42	752	0.16	0.11	Small
	5	1.23	642	0.22	0.10	Small
	6	-0.80	589	0.42	0.07	Trivial
	7	0.44	616	0.66	0.04	Trivial
	8	-0.35	602	0.73	0.03	Trivial
	10	-0.72	646	0.47	0.06	Trivial
Science	5	2.35	632	0.02*	0.20	Small
	8	0.62	584	0.54	0.05	Trivial
	11	-0.18	610	0.85	0.01	Trivial

** Significant at the .05 level.*

EL vs. Non-EL

Content	Grade	T Value	Degree of Freedom	P Value	|Cohen's D|	Magnitude of Effect Size
ELA	3	-0.61	718	0.54	0.05	Trivial
	4	-0.28	757	0.78	0.02	Trivial
	5	-0.69	641	0.49	0.06	Trivial
	6	-0.29	594	0.77	0.03	Trivial
	7	-2.00	616	0.05*	0.19	Small
	8	-0.45	603	0.65	0.04	Trivial
	10	-1.77	647	0.08	0.18	Small
Mathematics	3	-0.24	717	0.81	0.02	Trivial
	4	0.16	752	0.87	0.01	Trivial
	5	-0.66	642	0.51	0.06	Trivial
	6	0.52	589	0.60	0.05	Trivial
	7	-2.23	616	0.03*	0.21	Small
	8	-0.47	602	0.63	0.05	Trivial
	10	-1.57	646	0.12	0.16	Small
Science	5	-0.54	632	0.59	0.05	Trivial
	8	0.17	584	0.87	0.02	Trivial
	11	-3.12	610	0.00*	0.32	Small

** Significant at the .05 level.*

Hispanic/Latino vs. Non-Hispanic/Latino

Content	Grade	T Value	Degree of Freedom	P Value	|Cohen's D|	Magnitude of Effect Size
ELA	3	0.97	718	0.33	0.08	Trivial
	4	-0.39	757	0.70	0.03	Trivial
	5	1.05	641	0.29	0.09	Trivial
	6	1.22	594	0.22	0.11	Small
	7	-0.04	616	0.97	0.00	Trivial
	8	1.31	603	0.19	0.12	Small
	10	-1.67	647	0.10	0.15	Small
Mathematics	3	1.76	717	0.08	0.14	Small
	4	0.08	752	0.94	0.01	Trivial
	5	1.76	642	0.08	0.15	Small
	6	2.17	589	0.03*	0.20	Small
	7	0.96	616	0.34	0.08	Trivial
	8	1.37	602	0.17	0.12	Small
	10	-2.31	646	0.02*	0.20	Small
Science	5	2.36	632	0.02*	0.20	Small
	8	0.87	584	0.38	0.08	Trivial
	11	-2.31	610	0.02*	0.22	Small

** Significant at the .05 level.*

Table 6.27. Standardized Mean Difference (SMD) on Subgroup Pairs by Standard

Female vs. Male (Focal Group: Female)

Content	Grade	Standard	SMD	SD	Effect Size
ELA	3	Reading Literature	0.15	5.45	0.03
		Reading Informational Text	0.15	5.15	0.03
		Reading Foundational Skills	-0.71	5.93	-0.12
		Writing	0.40	3.76	0.11
		Speaking & Listening	0.03	3.78	0.01
	4	Reading Literature	0.25	4.72	0.05
		Reading Informational Text	0.12	5.26	0.02
		Reading Foundational Skills	-0.67	6.36	-0.11
		Writing	-0.15	4.97	-0.03
		Speaking & Listening	0.30	5.22	0.06
	5	Reading Literature	0.11	4.70	0.02
		Reading Informational Text	0.04	5.03	0.01
		Reading Foundational Skills	-0.18	6.48	-0.03
		Writing	0.09	4.42	0.02
		Speaking & Listening	-0.11	4.62	-0.02
	6	Reading Literature	-0.16	5.14	-0.03
		Reading Informational Text	0.06	5.30	0.01
		Writing	0.27	5.60	0.05
		Speaking & Listening	0.28	5.56	0.05
		Language	-0.34	5.76	-0.06
	7	Reading Literature	-0.05	5.65	-0.01
		Reading Informational Text	-0.17	5.23	-0.03
		Writing	-0.23	5.68	-0.04
		Speaking & Listening	0.12	5.70	0.02
		Language	0.42	5.89	0.07
	8	Reading Informational Text	0.42	4.93	0.08
		Reading Informational Text: Integration	0.27	5.08	0.05
		Writing	-0.41	4.77	-0.09
		Speaking & Listening	-0.10	5.48	-0.02
		Language	-0.33	6.14	-0.05
	10	Reading Literature	-0.01	3.71	0.00
		Reading Informational Text	-0.24	4.16	-0.06
		Writing	0.14	4.45	0.03
		Writing-Research to Build & Present Knowledge	0.22	4.27	0.05
		Speaking & Listening	-0.18	4.37	-0.04
Mathematics	3	Geometry	-0.25	6.12	-0.04
		Measurement & Data	0.08	6.22	0.01
		Number & Operations in Base Ten	-0.05	4.63	-0.01
		Number & Operations—Fractions	0.01	4.71	0.00
		Operations & Algebraic Thinking	0.03	5.06	0.01

Content	Grade	Standard	SMD	SD	Effect Size
	4	Geometry	0.01	5.16	0.00
		Measurement & Data	0.04	4.99	0.01
		Number & Operations in Base Ten	0.06	5.78	0.01
		Number & Operations—Fractions	0.04	5.75	0.01
		Operations & Algebraic Thinking	0.07	5.72	0.01
	5	Geometry	-0.04	6.26	-0.01
		Measurement & Data	-0.05	5.52	-0.01
		Number & Operations in Base Ten	0.11	5.41	0.02
		Number & Operations—Fractions	0.34	5.29	0.06
		Operations & Algebraic Thinking	-0.34	5.88	-0.06
	6	Geometry	-0.12	6.56	-0.02
		Expressions & Equations	0.37	5.11	0.07
		The Number System / Real Number System	0.29	6.42	0.05
		Ratios & Proportional Relationships	-0.22	6.69	-0.03
		Statistics & Probability	-0.08	5.93	-0.01
	7	Geometry	-0.34	6.95	-0.05
		Expressions & Equations	0.31	6.13	0.05
		The Number System / Real Number System	0.00	5.61	0.00
		Ratios & Proportional Relationships	-0.08	5.70	-0.01
		Statistics & Probability	0.09	5.53	0.02
	8	Geometry	0.16	5.32	0.03
		Expressions & Equations	-0.52	5.79	-0.09
		Functions	-0.38	5.76	-0.07
		The Number System / Real Number System	0.23	6.57	0.04
		Statistics & Probability	0.33	6.86	0.05
	10	Algebra-Creating Equations	-0.15	4.05	-0.04
		Algebra-Reasoning with Equations & Inequalities	-0.18	5.36	-0.03
		Geometry	0.16	5.30	0.03
		The Number System / Real Number System	-0.02	4.58	0.00
		Statistics & Probability	0.34	4.34	0.08
Science	5	Engineering & Technology-Engineering Design	-0.16	4.88	-0.03
		Life Sciences-From Molecules to Organisms: Structure and Processes	-0.07	5.24	-0.01
		Physical Sciences-Matter and Its Interactions	0.03	3.99	0.01
		Physical Sciences-Motion and Stability: Forces and Interactions	0.15	3.83	0.04
		Earth and Space Sciences-Earth's Place in the Universe	0.00	3.70	0.00
	8	Engineering & Technology-Engineering Design	-0.27	4.35	-0.06
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	-0.26	5.30	-0.05
		Physical Sciences-Energy	0.29	4.28	0.07
		Earth and Space Sciences-Earth's Place in the Universe	0.07	4.34	0.02
		Earth and Space Sciences-Earth's Systems	0.02	4.20	0.00
	11	Engineering & Technology-Engineering Design	0.10	3.64	0.03
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	0.31	3.97	0.08
		Physical Sciences-Matter and Its Interactions	0.02	3.60	0.01

Content	Grade	Standard	SMD	SD	Effect Size
		Earth and Space Sciences-Earth's Systems	-0.08	3.25	-0.02
		Earth and Space Sciences-Earth and Human Activity	-0.34	4.02	-0.08

Low Income vs. Non-Low Income (Focal Group: Low Income)

Content	Grade	Standard	SMD	SD	Effect Size
ELA	3	Reading Literature	0.08	5.45	0.01
		Reading Informational Text	-0.33	5.15	-0.06
		Reading Foundational Skills	-0.51	5.93	-0.09
		Writing	0.67	3.76	0.18*
		Speaking & Listening	0.09	3.78	0.02
	4	Reading Literature	0.02	4.72	0.00
		Reading Informational Text	0.06	5.26	0.01
		Reading Foundational Skills	-0.42	6.36	-0.07
		Writing	0.04	4.97	0.01
		Speaking & Listening	0.36	5.22	0.07
	5	Reading Literature	0.13	4.70	0.03
		Reading Informational Text	-0.03	5.03	-0.01
		Reading Foundational Skills	-0.18	6.48	-0.03
		Writing	-0.15	4.42	-0.03
		Speaking & Listening	0.29	4.62	0.06
	6	Reading Literature	0.09	5.14	0.02
		Reading Informational Text	0.24	5.30	0.04
		Writing	-0.22	5.60	-0.04
		Speaking & Listening	0.17	5.56	0.03
		Language	-0.30	5.76	-0.05
	7	Reading Literature	-0.12	5.65	-0.02
		Reading Informational Text	-0.11	5.23	-0.02
		Writing	-0.64	5.68	-0.11
		Speaking & Listening	-0.03	5.70	-0.01
		Language	0.53	5.89	0.09
	8	Reading Informational Text	-0.16	4.93	-0.03
		Reading Informational Text: Integration	0.07	5.08	0.01
		Writing	0.07	4.77	0.02
		Speaking & Listening	-0.15	5.48	-0.03
		Language	0.28	6.14	0.05
	10	Reading Literature	-0.41	3.71	-0.11
		Reading Informational Text	-0.30	4.16	-0.07
		Writing	0.97	4.45	0.22*
		Writing-Research to Build & Present Knowledge	0.23	4.27	0.05
		Speaking & Listening	-0.21	4.37	-0.05
Mathematics	3	Geometry	0.21	6.12	0.03
		Measurement & Data	0.59	6.22	0.10
		Number & Operations in Base Ten	-0.34	4.63	-0.07
		Number & Operations—Fractions	-0.46	4.71	-0.10
		Operations & Algebraic Thinking	-0.05	5.06	-0.01
	4	Geometry	-0.30	5.16	-0.06

Content	Grade	Standard	SMD	SD	Effect Size
		Measurement & Data	0.11	4.99	0.02
		Number & Operations in Base Ten	0.09	5.78	0.02
		Number & Operations—Fractions	0.00	5.75	0.00
		Operations & Algebraic Thinking	0.14	5.72	0.02
	5	Geometry	0.11	6.26	0.02
		Measurement & Data	-0.14	5.52	-0.02
		Number & Operations in Base Ten	-0.07	5.41	-0.01
		Number & Operations—Fractions	0.16	5.29	0.03
		Operations & Algebraic Thinking	0.02	5.88	0.00
	6	Geometry	0.13	6.56	0.02
		Expressions & Equations	0.47	5.11	0.09
		The Number System / Real Number System	-0.24	6.42	-0.04
		Ratios & Proportional Relationships	-0.40	6.69	-0.06
		Statistics & Probability	0.01	5.93	0.00
	7	Geometry	-0.20	6.95	-0.03
		Expressions & Equations	-0.25	6.13	-0.04
		The Number System / Real Number System	0.56	5.61	0.10
		Ratios & Proportional Relationships	-0.20	5.70	-0.04
		Statistics & Probability	0.09	5.53	0.02
	8	Geometry	0.47	5.32	0.09
		Expressions & Equations	-0.31	5.79	-0.05
		Functions	-0.26	5.76	-0.04
		The Number System / Real Number System	0.22	6.57	0.03
		Statistics & Probability	-0.05	6.86	-0.01
	10	Algebra-Creating Equations	-0.06	4.05	-0.01
		Algebra-Reasoning with Equations & Inequalities	0.14	5.36	0.03
		Geometry	0.58	5.30	0.11
		The Number System / Real Number System	0.20	4.58	0.04
		Statistics & Probability	-0.03	4.34	-0.01
Science	5	Engineering & Technology-Engineering Design	0.18	4.88	0.04
		Life Sciences-From Molecules to Organisms: Structure and Processes	0.09	5.24	0.02
		Physical Sciences-Matter and Its Interactions	-0.11	3.99	-0.03
		Physical Sciences-Motion and Stability: Forces and Interactions	-0.09	3.83	-0.02
		Earth and Space Sciences-Earth's Place in the Universe	-0.07	3.70	-0.02
	8	Engineering & Technology-Engineering Design	0.19	4.35	0.04
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	-0.37	5.30	-0.07
		Physical Sciences-Energy	0.42	4.28	0.10
		Earth and Space Sciences-Earth's Place in the Universe	-0.08	4.34	-0.02
		Earth and Space Sciences-Earth's Systems	-0.02	4.20	0.00
	11	Engineering & Technology-Engineering Design	0.23	3.64	0.06
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	0.09	3.97	0.02
		Physical Sciences-Matter and Its Interactions	0.00	3.60	0.00

Content	Grade	Standard	SMD	SD	Effect Size
		Earth and Space Sciences-Earth's Systems	0.12	3.25	0.04
		Earth and Space Sciences-Earth and Human Activity	-0.25	4.02	-0.06

** Moderate DIF (.17 ≤ |Effect Size| < .25).*

EL vs. Non-EL (Focal Group: EL)

Content	Grade	Standard	SMD	SD	Effect Size
ELA	3	Reading Literature	0.11	5.45	0.02
		Reading Informational Text	-0.16	5.15	-0.03
		Reading Foundational Skills	-0.03	5.93	0.00
		Writing	0.04	3.76	0.01
		Speaking & Listening	0.11	3.78	0.03
	4	Reading Literature	0.34	4.72	0.07
		Reading Informational Text	-0.12	5.26	-0.02
		Reading Foundational Skills	-0.75	6.36	-0.12
		Writing	0.05	4.97	0.01
		Speaking & Listening	0.72	5.22	0.14
	5	Reading Literature	-0.03	4.70	-0.01
		Reading Informational Text	0.33	5.03	0.07
		Reading Foundational Skills	-0.36	6.48	-0.06
		Writing	-0.11	4.42	-0.03
		Speaking & Listening	0.07	4.62	0.02
	6	Reading Literature	0.42	5.14	0.08
		Reading Informational Text	-0.20	5.30	-0.04
		Writing	-0.57	5.60	-0.10
		Speaking & Listening	0.03	5.56	0.01
		Language	-0.07	5.76	-0.01
	7	Reading Literature	0.17	5.65	0.03
		Reading Informational Text	-0.14	5.23	-0.03
		Writing	-0.24	5.68	-0.04
		Speaking & Listening	-0.24	5.70	-0.04
		Language	0.36	5.89	0.06
	8	Reading Informational Text	0.07	4.93	0.01
		Reading Informational Text: Integration	0.19	5.08	0.04
		Writing	-0.33	4.77	-0.07
		Speaking & Listening	0.54	5.48	0.10
		Language	-0.36	6.14	-0.06
	10	Reading Literature	-0.34	3.71	-0.09
		Reading Informational Text	0.06	4.16	0.01
		Writing	-0.75	4.45	-0.17*
		Writing-Research to Build & Present Knowledge	0.14	4.27	0.03
		Speaking & Listening	0.33	4.37	0.07
Mathematics	3	Geometry	-0.44	6.12	-0.07
		Measurement & Data	-0.18	6.22	-0.03
		Number & Operations in Base Ten	0.07	4.63	0.01
		Number & Operations—Fractions	0.51	4.71	0.11
		Operations & Algebraic Thinking	0.00	5.06	0.00
	4	Geometry	-0.01	5.16	0.00

Content	Grade	Standard	SMD	SD	Effect Size
		Measurement & Data	0.48	4.99	0.10
		Number & Operations in Base Ten	-0.36	5.78	-0.06
		Number & Operations—Fractions	0.08	5.75	0.01
		Operations & Algebraic Thinking	-0.04	5.72	-0.01
	5	Geometry	-0.23	6.26	-0.04
		Measurement & Data	0.28	5.52	0.05
		Number & Operations in Base Ten	0.30	5.41	0.06
		Number & Operations—Fractions	-0.12	5.29	-0.02
		Operations & Algebraic Thinking	-0.29	5.88	-0.05
	6	Geometry	0.22	6.56	0.03
		Expressions & Equations	-0.09	5.11	-0.02
		The Number System / Real Number System	0.60	6.42	0.09
		Ratios & Proportional Relationships	-0.35	6.69	-0.05
		Statistics & Probability	-0.28	5.93	-0.05
	7	Geometry	-0.02	6.95	0.00
		Expressions & Equations	-0.05	6.13	-0.01
		The Number System / Real Number System	0.82	5.61	0.15
		Ratios & Proportional Relationships	-0.23	5.70	-0.04
		Statistics & Probability	-0.73	5.53	-0.13
	8	Geometry	0.78	5.32	0.15
		Expressions & Equations	-0.09	5.79	-0.02
		Functions	-0.10	5.76	-0.02
		The Number System / Real Number System	-0.39	6.57	-0.06
		Statistics & Probability	-0.27	6.86	-0.04
	10	Algebra-Creating Equations	0.09	4.05	0.02
		Algebra-Reasoning with Equations & Inequalities	-0.41	5.36	-0.08
		Geometry	0.17	5.30	0.03
		The Number System / Real Number System	1.49	4.58	0.33**
		Statistics & Probability	0.03	4.34	0.01
Science	5	Engineering & Technology-Engineering Design	0.37	4.88	0.08
		Life Sciences-From Molecules to Organisms: Structure and Processes	-0.05	5.24	-0.01
		Physical Sciences-Matter and Its Interactions	0.30	3.99	0.08
		Physical Sciences-Motion and Stability: Forces and Interactions	-0.24	3.83	-0.06
		Earth and Space Sciences-Earth's Place in the Universe	-0.18	3.70	-0.05
	8	Engineering & Technology-Engineering Design	-0.14	4.35	-0.03
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	0.00	5.30	0.00
		Physical Sciences-Energy	0.03	4.28	0.01
		Earth and Space Sciences-Earth's Place in the Universe	0.03	4.34	0.01
		Earth and Space Sciences-Earth's Systems	0.09	4.20	0.02
	11	Engineering & Technology-Engineering Design	-0.02	3.64	-0.01
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	0.10	3.97	0.03
		Physical Sciences-Matter and Its Interactions	-0.02	3.60	-0.01

Content	Grade	Standard	SMD	SD	Effect Size
		Earth and Space Sciences-Earth's Systems	0.34	3.25	0.10
		Earth and Space Sciences-Earth and Human Activity	-0.33	4.02	-0.08

** Moderate DIF (.17 ≤ |Effect Size| < .25).*
*** Large DIF (|ES| ≥ .25).*

Hispanic/Latino vs. Non-Hispanic/Latino (Focal Group: Hispanic/Latino)

Content	Grade	Standard	SMD	SD	Effect Size
ELA	3	Reading Literature	-0.02	5.45	0.00
		Reading Informational Text	-0.03	5.15	-0.01
		Reading Foundational Skills	-0.08	5.93	-0.01
		Writing	0.16	3.76	0.04
		Speaking & Listening	-0.03	3.78	-0.01
	4	Reading Literature	0.23	4.72	0.05
		Reading Informational Text	-0.19	5.26	-0.04
		Reading Foundational Skills	-0.37	6.36	-0.06
		Writing	0.06	4.97	0.01
		Speaking & Listening	0.46	5.22	0.09
	5	Reading Literature	0.21	4.70	0.04
		Reading Informational Text	-0.11	5.03	-0.02
		Reading Foundational Skills	-0.18	6.48	-0.03
		Writing	0.07	4.42	0.01
		Speaking & Listening	0.00	4.62	0.00
	6	Reading Literature	0.41	5.14	0.08
		Reading Informational Text	0.02	5.30	0.00
		Writing	-0.44	5.60	-0.08
		Speaking & Listening	-0.04	5.56	-0.01
		Language	-0.06	5.76	-0.01
	7	Reading Literature	0.82	5.65	0.14
		Reading Informational Text	-0.17	5.23	-0.03
		Writing	-0.25	5.68	-0.04
		Speaking & Listening	-0.39	5.70	-0.07
		Language	-0.09	5.89	-0.01
	8	Reading Informational Text	-0.18	4.93	-0.04
		Reading Informational Text: Integration	-0.03	5.08	-0.01
		Writing	-0.59	4.77	-0.12
		Speaking & Listening	0.17	5.48	0.03
		Language	0.53	6.14	0.09
	10	Reading Literature	-0.20	3.71	-0.05
		Reading Informational Text	0.15	4.16	0.04
		Writing	0.44	4.45	0.10
		Writing-Research to Build & Present Knowledge	0.05	4.27	0.01
		Speaking & Listening	0.05	4.37	0.01
Mathematics	3	Geometry	0.00	6.12	0.00
		Measurement & Data	-0.05	6.22	-0.01
		Number & Operations in Base Ten	0.56	4.63	0.12
		Number & Operations—Fractions	0.06	4.71	0.01
		Operations & Algebraic Thinking	-0.25	5.06	-0.05
	4	Geometry	-0.06	5.16	-0.01

Content	Grade	Standard	SMD	SD	Effect Size
		Measurement & Data	0.21	4.99	0.04
		Number & Operations in Base Ten	-0.04	5.78	-0.01
		Number & Operations—Fractions	0.16	5.75	0.03
		Operations & Algebraic Thinking	-0.17	5.72	-0.03
	5	Geometry	-0.01	6.26	0.00
		Measurement & Data	0.27	5.52	0.05
		Number & Operations in Base Ten	-0.06	5.41	-0.01
		Number & Operations—Fractions	-0.27	5.29	-0.05
		Operations & Algebraic Thinking	0.10	5.88	0.02
	6	Geometry	0.65	6.56	0.10
		Expressions & Equations	0.20	5.11	0.04
		The Number System / Real Number System	-0.06	6.42	-0.01
		Ratios & Proportional Relationships	-0.51	6.69	-0.08
		Statistics & Probability	-0.47	5.93	-0.08
	7	Geometry	0.11	6.95	0.02
		Expressions & Equations	-0.19	6.13	-0.03
		The Number System / Real Number System	0.36	5.61	0.06
		Ratios & Proportional Relationships	0.38	5.70	0.07
		Statistics & Probability	-0.57	5.53	-0.10
	8	Geometry	0.97	5.32	0.18*
		Expressions & Equations	0.21	5.79	0.04
		Functions	-0.77	5.76	-0.13
		The Number System / Real Number System	-0.65	6.57	-0.10
		Statistics & Probability	0.19	6.86	0.03
	10	Algebra-Creating Equations	-0.50	4.05	-0.12
		Algebra-Reasoning with Equations & Inequalities	0.49	5.36	0.09
		Geometry	0.19	5.30	0.04
		The Number System / Real Number System	-0.86	4.58	-0.19*
		Statistics & Probability	0.15	4.34	0.03
Science	5	Engineering & Technology-Engineering Design	0.10	4.88	0.02
		Life Sciences-From Molecules to Organisms: Structure and Processes	0.31	5.24	0.06
		Physical Sciences-Matter and Its Interactions	0.07	3.99	0.02
		Physical Sciences-Motion and Stability: Forces and Interactions	-0.32	3.83	-0.08
		Earth and Space Sciences-Earth's Place in the Universe	-0.12	3.70	-0.03
	8	Engineering & Technology-Engineering Design	-0.09	4.35	-0.02
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	-0.21	5.30	-0.04
		Physical Sciences-Energy	-0.16	4.28	-0.04
		Earth and Space Sciences-Earth's Place in the Universe	0.00	4.34	0.00
		Earth and Space Sciences-Earth's Systems	0.06	4.20	0.01
	11	Engineering & Technology-Engineering Design	0.12	3.64	0.03
		Life Sciences-Ecosystems: Interactions, Energy, and Dynamics	-0.26	3.97	-0.07
		Physical Sciences-Matter and Its Interactions	0.17	3.60	0.05

Content	Grade	Standard	SMD	SD	Effect Size
		Earth and Space Sciences-Earth's Systems	0.31	3.25	0.10
		Earth and Space Sciences-Earth and Human Activity	-0.24	4.02	-0.06

** Moderate DIF (.17 ≤ |Effect Size| < .25).*

6.9 Standard Statistics

As access points may vary across individual students on each tested standard, CTT statistics such as difficulty and discrimination are provided for each standard based on weighted standard scores (after applying access-point weight values). "Difficulty" was defined as the average proportion of weighted points achieved on a standard and was measured by obtaining the average weighted score on a standard and dividing by the maximum score for the standard. By computing the difficulty index as the average proportion of points achieved, the standards are placed on a scale that ranges from 0.0 to 1.0. Although the *p*-value is traditionally described as a measure of difficulty (as it is described here), it is properly interpreted as an easiness index, because larger values indicate easier items. An index of 0.0 indicates that all students received no credit for the standard, and an index of 1.0 indicates that all students received full credit for the standard.

Standards that have either a very high or very low difficulty index are considered to be potentially problematic, because they are either so difficult that few students get them right or so easy that nearly all students get them right. In either case, such standards should be reviewed for appropriateness for inclusion on the assessment. If an assessment were composed entirely of very easy or very hard standards, all students would receive nearly the same scores, and the assessment would not be able to differentiate high-ability students from low-ability students.

It is worth mentioning that using a norm-referenced criterion such as *p*-values to evaluate test items is somewhat contradictory to the purpose of a criterion-referenced assessment like the WA-AIM assessment. Criterion-referenced assessments are primarily intended to provide evidence on student progress relative to a standard rather than to differentiate among students. Thus, the generally accepted criteria regarding classical item statistics are only cautiously applicable to the WA-AIM assessment.

A desirable feature of an item in CTT is that high-ability students perform better on items than low-ability students. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of an item. Within CTT, this item-test correlation is referred to as the item's "discrimination," because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test.

The discrimination index used to evaluate WA-AIM standards was the Pearson product-moment correlation. The theoretical range of this statistic is -1.0 to 1.0. The reported discrimination index can be thought of as a measure of how closely a standard elicits the same student performance assessed by other standards contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency. For the WA-AIM assessment, the test total score was used as the criterion score.

A summary of the difficulty and discrimination statistics of each standard at each grade and content area combination is provided in Tables 6.28 through 6.30. The reported standard difficulty ranges from 0.18 to 0.38 for ELA, 0.25 to 0.41 for mathematics, and 0.14 to 0.25 for science.

The reported discrimination values were high and relatively stable across standards, grades, and content areas. The discrimination values range from 0.75 to 0.91 for ELA, 0.83 to 0.91 for mathematics, and 0.86 to 0.91 for science.

Along with difficulty and discrimination statistics, the mean and standard deviation of student weighted scores at each standard are presented. The mean (on a scale of 0 to 20) is proportional to the reported standard difficulty (on a scale of 0 to 1). A greater value of standard deviation suggests a wider spread of scores on the standard.

Additionally, the distribution of assessed access points at each standard is provided in Tables 6.31 through 6.33. The distribution of access points varied by standard, suggesting a variety in the difficulty of standards or/and in student achievements across standards. To further analyze

student performance on each standard, refer to Tables 6.34 through 6.36 for student score distributions by standard and access point.

Table 6.28. Standard Statistics, ELA

Grade	Strand	N	Mean	SD	Difficulty	Discrimination
3	Reading Literature	717	5.48	5.45	0.27	0.86
	Reading Informational Text	717	5.06	5.15	0.25	0.87
	Reading Foundational Skills	717	6.69	5.93	0.33	0.78
	Writing	713	3.54	3.76	0.18	0.79
	Speaking & Listening	715	3.70	3.78	0.18	0.82
4	Reading Literature	754	4.39	4.72	0.22	0.85
	Reading Informational Text	745	5.55	5.25	0.28	0.88
	Reading Foundational Skills	751	6.73	6.35	0.34	0.75
	Writing	753	4.42	4.96	0.22	0.84
	Speaking & Listening	746	5.10	5.22	0.25	0.85
5	Reading Literature	637	4.65	4.70	0.23	0.89
	Reading Informational Text	638	5.27	5.03	0.26	0.88
	Reading Foundational Skills	634	7.54	6.48	0.38	0.76
	Writing	633	4.62	4.42	0.23	0.87
	Speaking & Listening	637	4.73	4.62	0.24	0.88
6	Reading Literature	581	5.05	5.18	0.25	0.88
	Reading Informational Text	579	5.33	5.33	0.27	0.85
	Writing	581	5.36	5.63	0.27	0.87
	Speaking & Listening	580	5.34	5.59	0.27	0.81
	Language	582	5.69	5.78	0.28	0.83
7	Reading Literature	598	5.14	5.65	0.26	0.91
	Reading Informational Text	604	5.11	5.23	0.26	0.89
	Writing	599	5.22	5.68	0.26	0.85
	Speaking & Listening	603	5.84	5.70	0.29	0.88
	Language	610	6.29	5.89	0.31	0.88
8	Reading Informational Text	599	4.97	4.93	0.25	0.89
	Reading Informational Text: Integration	600	4.95	5.08	0.25	0.89
	Writing	601	4.50	4.81	0.23	0.88
	Speaking & Listening	601	4.44	5.47	0.22	0.86
	Language	599	7.40	6.16	0.37	0.79
10	Reading Literature	639	4.97	3.71	0.25	0.87
	Reading Informational Text	642	4.92	4.16	0.25	0.88
	Writing	639	4.50	4.45	0.23	0.88
	Writing—Research to Build & Present Knowledge	642	5.79	4.27	0.29	0.85
	Speaking & Listening	639	4.72	4.37	0.24	0.90

Table 6.29. Standard Statistics, Mathematics

Grade	Domain	N	Mean	SD	Difficulty	Discrimination
3	Geometry	715	6.70	6.12	0.34	0.86
	Measurement & Data	714	6.63	6.22	0.33	0.88
	Number & Operations in Base Ten	713	5.29	4.63	0.26	0.88
	Number & Operations—Fractions	710	5.30	4.71	0.27	0.85
	Operations & Algebraic Thinking	711	5.35	5.06	0.27	0.87
4	Geometry	752	5.50	5.16	0.28	0.86
	Measurement & Data	752	5.41	4.98	0.27	0.83
	Number & Operations in Base Ten	750	6.40	5.78	0.32	0.88
	Number & Operations—Fractions	752	6.01	5.74	0.30	0.87
	Operations & Algebraic Thinking	748	6.08	5.72	0.30	0.86
5	Geometry	640	7.37	6.26	0.37	0.88
	Measurement & Data	640	5.51	5.52	0.28	0.88
	Number & Operations in Base Ten	638	5.01	5.41	0.25	0.88
	Number & Operations—Fractions	639	5.37	5.29	0.27	0.89
	Operations & Algebraic Thinking	638	6.30	5.88	0.32	0.88
6	Geometry	575	7.35	6.57	0.37	0.91
	Expressions & Equations	588	5.74	5.14	0.29	0.87
	The Number System / Real Number System	582	6.78	6.44	0.34	0.90
	Ratios & Proportional Relationships	578	7.54	6.70	0.38	0.90
	Statistics & Probability	577	6.30	5.95	0.31	0.89
7	Geometry	601	8.16	6.95	0.41	0.87
	Expressions & Equations	608	6.82	6.13	0.34	0.89
	The Number System / Real Number System	602	6.01	5.61	0.30	0.88
	Ratios & Proportional Relationships	608	6.84	5.70	0.34	0.85
	Statistics & Probability	603	5.82	5.53	0.29	0.88
8	Geometry	601	5.52	5.31	0.28	0.85
	Expressions & Equations	601	5.82	5.79	0.29	0.91
	Functions	599	6.00	5.76	0.30	0.89
	The Number System / Real Number System	599	6.88	6.56	0.34	0.88
	Statistics & Probability	599	8.06	6.85	0.40	0.86
10	Algebra—Creating Equations	638	5.12	4.05	0.26	0.87
	Algebra—Reasoning with Equations & Inequalities	636	6.39	5.36	0.32	0.85
	Geometry	639	6.57	5.29	0.33	0.85
	The Number System / Real Number System	639	5.91	4.58	0.30	0.86
	Statistics & Probability	645	5.84	4.34	0.29	0.90

Table 6.30. Standard Statistics, Science

Grade	Performance Expectation	N	Mean	SD	Difficulty	Discrimination
5	Engineering & Technology-Engineering Design	623	4.36	4.88	0.22	0.89
	Life Sciences—From Molecules to Organisms: Structure and Processes	616	4.99	5.24	0.25	0.86
	Physical Sciences—Matter and Its Interactions	620	3.57	3.99	0.18	0.90
	Physical Sciences—Motion and Stability: Forces and Interactions	613	2.86	3.83	0.14	0.87
	Earth and Space Sciences—Earth's Place in the Universe	623	3.20	3.70	0.16	0.88
8	Engineering & Technology—Engineering Design	577	3.78	4.35	0.19	0.90
	Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	578	4.44	5.30	0.22	0.86
	Physical Sciences—Energy	572	3.55	4.28	0.18	0.89
	Earth and Space Sciences—Earth's Place in the Universe	572	4.38	4.34	0.22	0.91
	Earth and Space Sciences—Earth's Systems	580	3.97	4.20	0.20	0.90
11	Engineering & Technology—Engineering Design	605	4.36	3.64	0.22	0.88
	Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	608	4.56	3.97	0.23	0.88
	Physical Sciences—Matter and Its Interactions	602	4.65	3.60	0.23	0.89
	Earth and Space Sciences—Earth's Systems	605	4.05	3.25	0.20	0.89
	Earth and Space Sciences—Earth and Human Activity	601	4.68	4.02	0.23	0.86

Table 6.31. Access Point Distributions by Standard, ELA

Grade	Strand	Valid N	L	I	M	Not Tested
3	Reading Literature	717	45%	35%	20%	0%
	Reading Informational Text	717	47%	34%	18%	0%
	Reading Foundational Skills	717	39%	38%	23%	0%
	Writing	713	58%	33%	8%	1%
	Speaking & Listening	715	53%	36%	10%	1%
4	Reading Literature	754	50%	34%	15%	1%
	Reading Informational Text	745	43%	35%	20%	2%
	Reading Foundational Skills	751	42%	29%	28%	1%
	Writing	753	54%	31%	13%	1%
	Speaking & Listening	746	41%	44%	14%	2%
5	Reading Literature	637	43%	45%	12%	1%
	Reading Informational Text	638	40%	48%	12%	1%
	Reading Foundational Skills	634	35%	36%	27%	1%
	Writing	633	47%	41%	11%	2%
	Speaking & Listening	637	39%	49%	11%	1%
6	Reading Literature	581	41%	40%	16%	3%
	Reading Informational Text	579	43%	37%	17%	3%
	Writing	581	45%	36%	17%	3%
	Speaking & Listening	580	48%	33%	16%	3%
	Language	582	41%	36%	20%	3%
7	Reading Literature	598	43%	35%	19%	3%
	Reading Informational Text	604	42%	40%	16%	2%
	Writing	599	50%	31%	16%	3%
	Speaking & Listening	603	38%	35%	25%	2%
	Language	610	40%	33%	26%	1%
8	Reading Informational Text	599	46%	36%	17%	1%
	Reading Informational Text: Integration	600	44%	39%	16%	1%
	Writing	601	51%	36%	12%	1%
	Speaking & Listening	601	55%	29%	16%	1%
	Language	599	37%	31%	31%	1%
10	Reading Literature	639	40%	51%	7%	2%
	Reading Informational Text	642	49%	38%	11%	1%
	Writing	639	55%	34%	9%	2%
	Writing—Research to Build & Present Knowledge	642	39%	52%	8%	1%
	Speaking & Listening	639	49%	40%	10%	2%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Table 6.32. Access Point Distributions by Standard, Mathematics

Grade	Domain	Valid N	L	I	M	Not Tested
3	Geometry	715	41%	36%	23%	1%
	Measurement & Data	714	42%	34%	24%	1%
	Number & Operations in Base Ten	713	42%	40%	17%	1%
	Number & Operations—Fractions	710	46%	35%	18%	1%
	Operations & Algebraic Thinking	711	41%	39%	19%	1%
4	Geometry	752	42%	41%	17%	0%
	Measurement & Data	752	41%	39%	20%	0%
	Number & Operations in Base Ten	750	37%	38%	24%	1%
	Number & Operations—Fractions	752	42%	38%	20%	0%
	Operations & Algebraic Thinking	748	40%	35%	23%	1%
5	Geometry	640	34%	40%	26%	1%
	Measurement & Data	640	38%	44%	18%	1%
	Number & Operations in Base Ten	638	44%	36%	19%	1%
	Number & Operations—Fractions	639	43%	39%	18%	1%
	Operations & Algebraic Thinking	638	40%	35%	24%	1%
6	Geometry	575	38%	33%	26%	3%
	Expressions & Equations	588	40%	43%	17%	1%
	The Number System / Real Number System	582	38%	34%	26%	2%
	Ratios & Proportional Relationships	578	36%	38%	24%	2%
	Statistics & Probability	577	39%	37%	22%	3%
7	Geometry	601	33%	34%	30%	3%
	Expressions & Equations	608	35%	35%	28%	2%
	The Number System / Real Number System	602	39%	36%	23%	3%
	Ratios & Proportional Relationships	608	35%	36%	28%	2%
	Statistics & Probability	603	40%	32%	26%	2%
8	Geometry	601	41%	36%	22%	1%
	Expressions & Equations	601	44%	35%	20%	1%
	Functions	599	43%	33%	23%	1%
	The Number System / Real Number System	599	41%	32%	26%	1%
	Statistics & Probability	599	37%	33%	29%	1%
10	Algebra—Creating Equations	638	41%	47%	10%	2%
	Algebra—Reasoning with Equations & Inequalities	636	40%	42%	16%	2%
	Geometry	639	37%	42%	19%	2%
	The Number System / Real Number System	639	37%	49%	13%	2%
	Statistics & Probability	645	38%	48%	13%	1%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Table 6.33. Access Point Distributions by Standard, Science

Grade	Performance Expectation	Valid N	L	I	M	Not Tested
5	Engineering & Technology—Engineering Design	623	52%	34%	12%	2%
	Life Sciences—From Molecules to Organisms: Structure and Processes	616	49%	34%	14%	3%
	Physical Sciences—Matter and Its Interactions	620	56%	34%	7%	2%
	Physical Sciences—Motion and Stability: Forces and Interactions	613	59%	29%	8%	3%
	Earth and Space Sciences—Earth's Place in the Universe	623	59%	32%	7%	2%
8	Engineering & Technology—Engineering Design	577	59%	27%	12%	2%
	Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	578	58%	24%	16%	2%
	Physical Sciences—Energy	572	61%	26%	10%	3%
	Earth and Space Sciences—Earth's Place in the Universe	572	52%	35%	11%	3%
	Earth and Space Sciences—Earth's Systems	580	58%	29%	12%	1%
11	Engineering & Technology—Engineering Design	605	49%	44%	6%	1%
	Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	608	50%	41%	8%	1%
	Physical Sciences-Matter and Its Interactions	602	50%	42%	7%	2%
	Earth and Space Sciences—Earth's Systems	605	57%	38%	5%	1%
	Earth and Space Sciences—Earth and Human Activity	601	52%	38%	8%	2%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Table 6.34. Raw Score Distributions by Access Point and Standard, ELA

Grade 3

Strand	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Reading Literature	*L*	14%	18%	17%	25%	13%	12%
Reading Literature	*I*	4%	6%	15%	27%	28%	20%
Reading Literature	*M*	4%	10%	9%	22%	22%	33%
Reading Informational Text	*L*	19%	12%	24%	20%	13%	11%
Reading Informational Text	*I*	4%	7%	13%	26%	34%	15%
Reading Informational Text	*M*	5%	7%	16%	18%	29%	25%
Reading Foundational Skills	*L*	11%	10%	15%	16%	22%	26%
Reading Foundational Skills	*I*	3%	6%	11%	18%	24%	39%
Reading Foundational Skills	*M*	5%	5%	9%	13%	23%	45%
Writing	*L*	17%	15%	26%	19%	16%	8%
Writing	*I*	10%	5%	7%	17%	21%	41%
Writing	*M*	47%	7%	7%	7%	11%	20%
Speaking & Listening	*L*	14%	12%	21%	21%	20%	12%
Speaking & Listening	*I*	5%	15%	19%	22%	24%	14%
Speaking & Listening	*M*	26%	11%	14%	13%	16%	20%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 4

Strand	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Reading Literature	*L*	18%	13%	26%	15%	17%	11%
Reading Literature	*I*	9%	13%	17%	20%	29%	13%
Reading Literature	*M*	7%	10%	14%	21%	27%	21%
Reading Informational Text	*L*	11%	13%	21%	23%	19%	13%
Reading Informational Text	*I*	7%	6%	13%	28%	24%	21%
Reading Informational Text	*M*	3%	6%	15%	21%	32%	22%
Reading Foundational Skills	*L*	9%	9%	21%	18%	21%	22%
Reading Foundational Skills	*I*	11%	8%	13%	18%	24%	24%
Reading Foundational Skills	*M*	1%	6%	10%	14%	31%	37%
Writing	*L*	34%	6%	7%	11%	13%	29%
Writing	*I*	13%	8%	6%	19%	21%	32%
Writing	*M*	12%	12%	9%	13%	20%	35%
Speaking & Listening	*L*	21%	19%	19%	20%	15%	6%
Speaking & Listening	*I*	4%	9%	15%	26%	32%	14%
Speaking & Listening	*M*	5%	8%	8%	16%	14%	49%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 5

Strand	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Reading Literature	*L*	9%	15%	22%	20%	16%	18%
Reading Literature	*I*	7%	15%	20%	24%	22%	13%
Reading Literature	*M*	9%	5%	7%	17%	25%	37%
Reading Informational Text: Integration	*L*	15%	12%	21%	23%	14%	15%
Reading Informational Text: Integration	*I*	4%	8%	14%	21%	25%	28%
Reading Informational Text: Integration	*M*	8%	4%	5%	11%	18%	54%
Reading Foundational Skills	*L*	12%	16%	23%	16%	15%	17%
Reading Foundational Skills	*I*	2%	6%	7%	14%	23%	47%
Reading Foundational Skills	*M*	2%	2%	11%	15%	18%	53%
Writing—Research to Build & Present Knowledge	*L*	7%	11%	14%	22%	24%	22%
Writing—Research to Build & Present Knowledge	*I*	7%	11%	16%	29%	23%	14%
Writing—Research to Build & Present Knowledge	*M*	4%	4%	22%	12%	18%	40%
Speaking & Listening	*L*	14%	18%	19%	16%	17%	15%
Speaking & Listening	*I*	7%	11%	16%	23%	23%	19%
Speaking & Listening	*M*	9%	7%	7%	18%	18%	41%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 6

Strand	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Reading Literature	*L*	13%	16%	24%	22%	13%	11%
Reading Literature	*I*	5%	14%	20%	21%	24%	15%
Reading Literature	*M*	6%	6%	11%	14%	33%	29%
Reading Informational Text: Integration	*L*	12%	12%	22%	20%	22%	13%
Reading Informational Text: Integration	*I*	7%	9%	24%	19%	24%	18%
Reading Informational Text: Integration	*M*	4%	4%	13%	18%	28%	33%
Writing	*L*	29%	10%	13%	15%	14%	19%
Writing	*I*	14%	5%	9%	14%	19%	40%
Writing	*M*	9%	3%	10%	13%	26%	39%
Speaking & Listening	*L*	16%	13%	15%	19%	18%	20%
Speaking & Listening	*I*	15%	5%	4%	13%	16%	48%
Speaking & Listening	*M*	11%	5%	4%	21%	12%	46%
Language	*L*	16%	12%	18%	17%	20%	17%
Language	*I*	6%	13%	18%	21%	19%	23%
Language	*M*	7%	6%	8%	14%	26%	39%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 7

Strand	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Reading Literature	*L*	26%	12%	16%	16%	12%	18%
Reading Literature	*I*	20%	12%	12%	13%	16%	27%
Reading Literature	*M*	6%	4%	9%	23%	27%	30%
Reading Informational Text: Integration	*L*	14%	13%	20%	21%	20%	12%
Reading Informational Text: Integration	*I*	10%	11%	20%	18%	22%	18%
Reading Informational Text: Integration	*M*	4%	4%	10%	21%	29%	31%
Writing	*L*	30%	8%	11%	10%	11%	29%
Writing	*I*	17%	5%	5%	10%	12%	51%
Writing	*M*	10%	3%	10%	9%	25%	42%
Speaking & Listening	*L*	14%	25%	28%	15%	11%	8%
Speaking & Listening	*I*	9%	8%	23%	21%	16%	23%
Speaking & Listening	*M*	2%	6%	15%	29%	25%	23%
Language	*L*	10%	12%	21%	28%	19%	10%
Language	*I*	7%	8%	17%	18%	23%	26%
Language	*M*	2%	6%	13%	22%	28%	30%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 8

Strand	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Reading Literature	*L*	17%	12%	15%	22%	20%	13%
Reading Literature	*I*	4%	14%	20%	19%	25%	19%
Reading Literature	*M*	4%	4%	18%	24%	26%	24%
Reading Informational Text: Integration	*L*	32%	13%	8%	19%	16%	14%
Reading Informational Text: Integration	*I*	2%	14%	20%	21%	22%	21%
Reading Informational Text: Integration	*M*	7%	7%	9%	20%	32%	24%
Writing—Research to Build & Present Knowledge	*L*	13%	14%	18%	24%	20%	11%
Writing—Research to Build & Present Knowledge	*I*	7%	15%	12%	18%	20%	29%
Writing—Research to Build & Present Knowledge	*M*	15%	5%	8%	11%	29%	32%
Speaking & Listening	*L*	14%	15%	19%	23%	18%	11%
Speaking & Listening	*I*	21%	11%	16%	14%	23%	16%
Speaking & Listening	*M*	15%	2%	7%	14%	20%	43%
Language	*L*	14%	9%	17%	16%	16%	29%
Language	*I*	4%	5%	7%	17%	25%	42%
Language	*M*	1%	6%	11%	25%	22%	36%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 10

Strand	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Reading Literature	*L*	5%	7%	22%	22%	19%	25%
Reading Literature	*I*	2%	7%	11%	19%	26%	36%
Reading Literature	*M*	21%	10%	4%	27%	13%	25%
Reading Informational Text	*L*	5%	11%	14%	18%	22%	30%
Reading Informational Text	*I*	4%	5%	16%	21%	29%	27%
Reading Informational Text	*M*	4%	5%	23%	21%	22%	25%
Writing	*L*	4%	10%	20%	21%	25%	20%
Writing	*I*	14%	5%	10%	12%	23%	37%
Writing	*M*	11%	3%	10%	18%	18%	39%
Writing—Research to Build & Present Knowledge	*L*	3%	4%	14%	20%	31%	28%
Writing—Research to Build & Present Knowledge	*I*	3%	4%	7%	14%	22%	49%
Writing—Research to Build & Present Knowledge	*M*	4%	2%	10%	12%	22%	51%
Speaking & Listening	*L*	7%	16%	18%	28%	17%	14%
Speaking & Listening	*I*	7%	4%	13%	14%	23%	39%
Speaking & Listening	*M*	5%	13%	13%	16%	19%	34%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Table 6.35. Raw Score Distributions by Access Point and Standard, Mathematics

Grade 3

Domain	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Geometry	*L*	9%	4%	10%	13%	20%	45%
Geometry	*I*	7%	8%	16%	22%	20%	26%
Geometry	*M*	2%	2%	10%	10%	23%	53%
Measurement & Data	*L*	11%	6%	15%	21%	20%	27%
Measurement & Data	*I*	3%	9%	14%	17%	25%	32%
Measurement & Data	*M*	6%	6%	11%	9%	16%	53%
Number & Operations in Base Ten	*L*	11%	8%	14%	15%	17%	36%
Number & Operations in Base Ten	*I*	5%	10%	15%	21%	27%	22%
Number & Operations in Base Ten	*M*	9%	6%	23%	23%	16%	23%
Number & Operations—Fractions	*L*	7%	6%	21%	21%	19%	27%
Number & Operations—Fractions	*I*	4%	6%	13%	19%	30%	28%
Number & Operations—Fractions	*M*	7%	15%	19%	17%	18%	24%
Operations & Algebraic Thinking	*L*	10%	7%	15%	21%	17%	29%
Operations & Algebraic Thinking	*I*	8%	9%	21%	23%	25%	14%
Operations & Algebraic Thinking	*M*	5%	10%	16%	19%	20%	30%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 4

Domain	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Geometry	*L*	4%	8%	17%	23%	25%	23%
Geometry	*I*	4%	9%	20%	23%	21%	24%
Geometry	*M*	9%	8%	11%	13%	17%	42%
Measurement & Data	*L*	19%	8%	6%	8%	10%	50%
Measurement & Data	*I*	5%	8%	17%	30%	25%	16%
Measurement & Data	*M*	9%	13%	16%	17%	13%	30%
Number & Operations in Base Ten	*L*	9%	11%	18%	17%	20%	25%
Number & Operations in Base Ten	*I*	2%	7%	17%	24%	26%	23%
Number & Operations in Base Ten	*M*	4%	10%	13%	12%	23%	39%
Number & Operations—Fractions	*L*	7%	11%	29%	16%	15%	22%
Number & Operations—Fractions	*I*	3%	6%	17%	26%	26%	22%
Number & Operations—Fractions	*M*	3%	7%	16%	7%	18%	49%
Operations & Algebraic Thinking	*L*	11%	18%	30%	18%	12%	10%
Operations & Algebraic Thinking	*I*	1%	5%	11%	18%	25%	39%
Operations & Algebraic Thinking	*M*	10%	8%	10%	12%	31%	29%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 5

Domain	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Geometry	*L*	7%	11%	13%	16%	13%	40%
Geometry	*I*	3%	7%	11%	20%	28%	31%
Geometry	*M*	2%	5%	10%	12%	14%	57%
Measurement & Data	*L*	11%	16%	19%	24%	16%	14%
Measurement & Data	*I*	8%	11%	13%	19%	25%	24%
Measurement & Data	*M*	13%	4%	11%	9%	18%	45%
Number & Operations in Base Ten	*L*	16%	18%	23%	18%	19%	7%
Number & Operations in Base Ten	*I*	8%	19%	15%	21%	23%	14%
Number & Operations in Base Ten	*M*	5%	8%	11%	22%	24%	31%
Number & Operations—Fractions	*L*	12%	14%	24%	21%	16%	14%
Number & Operations—Fractions	*I*	4%	13%	11%	17%	35%	20%
Number & Operations—Fractions	*M*	5%	10%	13%	18%	15%	39%
Operations & Algebraic Thinking	*L*	9%	8%	19%	19%	17%	28%
Operations & Algebraic Thinking	*I*	5%	10%	15%	26%	22%	22%
Operations & Algebraic Thinking	*M*	5%	6%	11%	15%	25%	37%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 6

Domain	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Geometry	*L*	15%	9%	16%	16%	13%	32%
Geometry	*I*	6%	6%	11%	11%	17%	49%
Geometry	*M*	4%	3%	8%	10%	20%	54%
Expressions & Equations	*L*	10%	11%	18%	18%	21%	21%
Expressions & Equations	*I*	6%	7%	13%	15%	18%	40%
Expressions & Equations	*M*	4%	8%	12%	22%	15%	38%
The Number System / Real Number System	*L*	10%	15%	16%	21%	13%	24%
The Number System / Real Number System	*I*	9%	11%	15%	18%	18%	29%
The Number System / Real Number System	*M*	1%	5%	10%	16%	19%	50%
Ratios & Proportional Relationships	*L*	10%	10%	14%	12%	18%	36%
Ratios & Proportional Relationships	*I*	5%	9%	10%	13%	16%	47%
Ratios & Proportional Relationships	*M*	6%	1%	3%	6%	15%	69%
Statistics & Probability	*L*	13%	12%	20%	15%	21%	19%
Statistics & Probability	*I*	5%	6%	16%	19%	19%	34%
Statistics & Probability	*M*	6%	7%	12%	11%	19%	45%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 7

Domain	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Geometry	*L*	16%	16%	14%	13%	15%	25%
Geometry	*I*	3%	6%	9%	13%	20%	49%
Geometry	*M*	1%	5%	5%	9%	17%	62%
Expressions & Equations	*L*	12%	15%	14%	16%	17%	26%
Expressions & Equations	*I*	7%	12%	15%	16%	19%	32%
Expressions & Equations	*M*	2%	10%	7%	16%	34%	31%
The Number System / Real Number System	*L*	13%	12%	22%	24%	16%	13%
The Number System / Real Number System	*I*	4%	13%	19%	14%	20%	30%
The Number System / Real Number System	*M*	0%	8%	16%	24%	23%	30%
Ratios & Proportional Relationships	*L*	12%	14%	18%	22%	13%	21%
Ratios & Proportional Relationships	*I*	5%	4%	10%	9%	25%	47%
Ratios & Proportional Relationships	*M*	1%	11%	18%	19%	22%	29%
Statistics & Probability	*L*	12%	19%	18%	22%	13%	15%
Statistics & Probability	*I*	6%	15%	15%	15%	23%	27%
Statistics & Probability	*M*	3%	13%	14%	23%	27%	20%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 8

Domain	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Geometry	*L*	8%	5%	10%	17%	26%	34%
Geometry	*I*	11%	16%	18%	25%	14%	17%
Geometry	*M*	5%	13%	11%	21%	22%	28%
Expressions & Equations	*L*	12%	12%	13%	17%	18%	28%
Expressions & Equations	*I*	8%	19%	16%	14%	18%	26%
Expressions & Equations	*M*	2%	4%	8%	16%	32%	38%
Functions	*L*	15%	10%	17%	14%	17%	27%
Functions	*I*	4%	15%	19%	20%	19%	22%
Functions	*M*	1%	5%	9%	21%	35%	29%
The Number System / Real Number System	*L*	19%	15%	16%	14%	10%	26%
The Number System / Real Number System	*I*	4%	11%	8%	23%	25%	30%
The Number System / Real Number System	*M*	1%	6%	5%	15%	19%	53%
Statistics & Probability	*L*	15%	3%	7%	9%	11%	55%
Statistics & Probability	*I*	4%	11%	10%	14%	19%	43%
Statistics & Probability	*M*	2%	4%	6%	6%	18%	65%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 10

Domain	Access Point	Raw Score					
		0	***1***	***2***	***3***	***4***	***5***
Algebra—Creating Equations	*L*	5%	7%	13%	19%	24%	31%
Algebra—Creating Equations	*I*	2%	5%	14%	21%	24%	34%
Algebra—Creating Equations	*M*	25%	2%	8%	19%	20%	27%
Algebra—Reasoning with Equations & Inequalities	*L*	3%	7%	14%	19%	21%	36%
Algebra—Reasoning with Equations & Inequalities	*I*	2%	10%	8%	15%	22%	43%
Algebra—Reasoning with Equations & Inequalities	*M*	1%	5%	5%	15%	25%	50%
Geometry	*L*	3%	4%	10%	16%	20%	47%
Geometry	*I*	3%	7%	14%	15%	22%	39%
Geometry	*M*	2%	7%	7%	14%	31%	38%
The Number System / Real Number System	*L*	3%	2%	12%	16%	25%	42%
The Number System / Real Number System	*I*	3%	4%	13%	15%	25%	40%
The Number System / Real Number System	*M*	13%	7%	7%	10%	24%	38%
Statistics & Probability	*L*	4%	6%	9%	17%	21%	44%
Statistics & Probability	*I*	2%	7%	10%	14%	25%	42%
Statistics & Probability	*M*	3%	16%	16%	19%	13%	33%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Table 6.36. Raw Score Distributions by Access Point and Standard, Science Grade 5

Performance Expectation	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Engineering & Technology—Engineering Design	*L*	21%	15%	20%	18%	13%	13%
Engineering & Technology—Engineering Design	*I*	8%	13%	12%	23%	22%	21%
Engineering & Technology—Engineering Design	*M*	4%	3%	18%	12%	23%	40%
Life Sciences—From Molecules to Organisms: Structure and Processes	*L*	16%	8%	18%	22%	15%	21%
Life Sciences—From Molecules to Organisms: Structure and Processes	*I*	7%	8%	15%	21%	25%	24%
Life Sciences—From Molecules to Organisms: Structure and Processes	*M*	7%	6%	7%	12%	22%	47%
Physical Sciences—Matter and Its Interactions	*L*	13%	15%	27%	20%	14%	11%
Physical Sciences—Matter and Its Interactions	*I*	12%	14%	18%	22%	19%	16%
Physical Sciences—Matter and Its Interactions	*M*	11%	2%	11%	24%	20%	33%
Physical Sciences—Motion and Stability: Forces and Interactions	*L*	19%	11%	23%	25%	12%	10%
Physical Sciences—Motion and Stability: Forces and Interactions	*I*	36%	17%	12%	13%	11%	11%
Physical Sciences—Motion and Stability: Forces and Interactions	*M*	17%	9%	17%	19%	11%	26%
Earth and Space Sciences—Earth's Place in the Universe	*L*	17%	13%	23%	22%	16%	9%
Earth and Space Sciences—Earth's Place in the Universe	*I*	15%	20%	16%	16%	17%	15%
Earth and Space Sciences—Earth's Place in the Universe	*M*	13%	11%	19%	19%	13%	26%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 8

Performance Expectation	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Engineering & Technology—Engineering Design	*L*	18%	16%	19%	15%	19%	13%
Engineering & Technology—Engineering Design	*I*	10%	12%	21%	21%	15%	21%
Engineering & Technology—Engineering Design	*M*	15%	7%	21%	14%	23%	21%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*L*	11%	16%	20%	17%	22%	14%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*I*	15%	15%	18%	22%	15%	15%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*M*	7%	7%	12%	7%	32%	34%
Physical Sciences—Energy	*L*	24%	12%	17%	15%	20%	12%
Physical Sciences—Energy	*I*	16%	13%	10%	14%	21%	26%
Physical Sciences—Energy	*M*	18%	10%	8%	16%	28%	20%
Earth and Space Sciences—Earth's Place in the Universe	*L*	16%	15%	17%	24%	15%	13%
Earth and Space Sciences—Earth's Place in the Universe	*I*	3%	7%	19%	21%	25%	25%
Earth and Space Sciences—Earth's Place in the Universe	*M*	2%	13%	19%	17%	27%	23%
Earth and Space Sciences—Earth's Systems	*L*	14%	11%	17%	20%	22%	15%
Earth and Space Sciences—Earth's Systems	*I*	8%	14%	15%	23%	23%	17%
Earth and Space Sciences—Earth's Systems	*M*	13%	7%	16%	13%	38%	13%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

Grade 11

Performance Expectation	Access Point	Raw Score					
		0	*1*	*2*	*3*	*4*	*5*
Engineering & Technology—Engineering Design	*L*	2%	5%	15%	22%	33%	22%
Engineering & Technology—Engineering Design	*I*	8%	8%	12%	19%	33%	21%
Engineering & Technology—Engineering Design	*M*	26%	8%	10%	5%	15%	36%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*L*	10%	9%	13%	22%	28%	18%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*I*	4%	7%	11%	22%	17%	39%
Life Sciences—Ecosystems: Interactions, Energy, and Dynamics	*M*	18%	6%	6%	16%	31%	22%
Physical Sciences—Matter and Its Interactions	*L*	5%	5%	13%	17%	28%	32%
Physical Sciences—Matter and Its Interactions	*I*	3%	5%	10%	19%	29%	33%
Physical Sciences—Matter and Its Interactions	*M*	17%	17%	20%	2%	10%	34%
Earth and Space Sciences—Earth's Systems	*L*	4%	5%	15%	18%	35%	22%
Earth and Space Sciences—Earth's Systems	*I*	7%	7%	12%	23%	25%	27%
Earth and Space Sciences—Earth's Systems	*M*	7%	21%	14%	11%	18%	29%
Earth and Space Sciences—Earth and Human Activity	*L*	5%	5%	11%	22%	34%	23%
Earth and Space Sciences—Earth and Human Activity	*I*	6%	9%	8%	19%	25%	32%
Earth and Space Sciences—Earth and Human Activity	*M*	0%	17%	6%	23%	17%	38%

Note. L = Less Complex. I = Intermediate Complex. M = More Complex.

6.10 Relationship Between Student Performance and Other Variables

Validity evidence based on relationships with other variables refers to "evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" (AERA, APA, & NCME, 2014, p. 16). In educational testing, such evidence is often gathered through studies of correlations between the test scores and measures of different or similar constructs. As stated in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), relationships between test scores and other measures intended to assess the same or similar constructs provide convergent evidence, whereas relationships between test scores and measures of different constructs provide discriminant evidence (pp. 16–17).

The relationship between the WA-AIM content area scores and other variables was evaluated using the following methods:

- Correlations between the WA-AIM content area scores. Specifically, the correlations between the ELA, mathematics, and science total test scores for students who took more than one content area test in 2022 were computed and examined.
- Correlation between teacher ratings of student performance and observed student achievement levels on each WA-AIM assessment. In this method, Kendall rank correlation coefficient (or Kendall's *tau*) statistic was computed on teacher ratings of student performance in collected SCS responses (see details about the SCS in Section 6.3 of this chapter) and observed student achievement levels in the WA-AIM test data, by grade and content area.
- Tabulation of teacher ratings on the alignment of WA-AIM test standards to student IEP goals by observed student achievement level on each WA-AIM assessment. Same as the teacher ratings of student performance, teacher ratings on the alignment were collected in the 2022 WA-AIM SCS (see details about the SCS in Section 6.3 of this chapter).

6.10.1 Correlations between the WA-AIM content area scores

Table 6.37 shows correlations between content area total test scores for students who had valid test scores on the 2022 WA-AIM. The data from the 2021 WA-AIM are provided for reference.

These 2022 WA-AIM content area score correlations ranged from 0.73 (between mathematics and science at grade 8) to 0.85 (between ELA and mathematics at grade 5). Overall, the correlations between the content area scores for the tested students were found to be moderate to high. The moderate to high correlations indicate that the tests are not perfectly related to one another, suggesting that different constructs are being tapped; however, those constructs are related in that they all involve academic knowledge and skills.

Table 6.37. Correlations between Content Area Total Test Scores

Grade	ELA & Math	ELA & Science	Math & Science
3	0.84		
4	0.83		
5	0.85	0.82	0.77
6	0.83		
7	0.84		
8	0.84	0.78	0.73
10	0.79		

Reference: 2020–2021

Grade	ELA & Math	ELA & Science	Math & Science
3	0.87		
4	0.87		
5	0.85	0.74	0.74
6	0.83		
7	0.84		
8	0.78	0.79	0.68

6.10.2 Correlations between Teacher Ratings and Observed Student Achievement Level

Table 6.38 shows the Kendall rank correlation coefficient statistics (or Kendall's *tau*) on the association of teacher ratings of student performance in collected SCS responses and observed student achievement levels in the 2022 WA-AIM test data, by grade and content area. The data from the 2021 WA-AIM are provided for reference.

Kendall's *tau* at or above 0.21 generally indicates a moderate association, and the value at or above 0.35 indicates a strong association. The Kendall's *tau* in the 2022 WA-AIM is moderate to strong between the teachers' ratings and the observed student achievement levels, with a range of 0.35 to 0.48 in ELA, 0.35 to 0.45 in mathematics, and 0.25 to 0.32 in science.

Table 6.38. Correlations between Teacher Ratings and Observed Student Achievement Levels

	Kendall Rank Correlation Coefficient		
Grade	***ELA***	***Mathematics***	***Science***
3	0.41	0.38	
4	0.48	0.45	
5	0.42	0.45	0.32
6	0.38	0.40	
7	0.43	0.42	
8	0.44	0.41	0.29
HS*	0.35	0.35	0.25

*ELA and mathematics were assessed at grade 10, and science was assessed at grade 11.

Reference: 2020–2021

	Kendall Rank Correlation Coefficient		
Grade	***ELA***	***Mathematics***	***Science***
3	0.48	0.44	
4	0.46	0.45	
5	0.47	0.46	0.30
6	0.50	0.47	
7	0.42	0.44	
8	0.40	0.40	0.33
HS*	0.35	0.35	0.23

*ELA and mathematics were assessed at grade 10, and science was assessed at grade 11.

6.10.3 Tabulation of IEP Goal Alignment by Student Achievement Level

Table 6.39 presents the tabulation of teacher ratings on the alignment of WA-AIM test standards to student IEP goals by observed student achievement level on the 2022 WA-AIM for each grade and content area. The data from the 2021 WA-AIM are provided for reference.

In ELA, 42% to 95% teachers agreed that the grade level academic standards measured on the WA-AIM were somewhat or well represented in the student's IEP goals and/or objectives. Similarly, in mathematics 30% to 87% teachers agreed that the grade level academic standards measured on the WA-AIM were somewhat or well represented in the student's IEP goals and/or objectives. In science the teacher agreement on the alignment ("Somewhat or Well Represented") ranged from 15% to 68%. In general, a greater percentage of teacher agreement on the alignment was observed at higher student achievement levels.

Table 6.39. Teacher Ratings of WA-AIM's IEP Goal Alignment by Student Achievement Level

ELA

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
3	1	54%	32%	14%	46%
	2	29%	48%	23%	71%
	3	17%	57%	26%	83%
	4	14%	62%	24%	86%
4	1	52%	31%	17%	48%
	2	37%	45%	18%	63%
	3	21%	56%	23%	79%
	4	24%	51%	24%	76%
5	1	58%	29%	13%	42%
	2	38%	44%	18%	62%
	3	25%	57%	19%	75%
	4	40%	28%	33%	60%
6	1	41%	40%	19%	59%
	2	33%	55%	13%	67%
	3	20%	58%	22%	80%
	4	12%	63%	25%	88%
7	1	44%	48%	8%	56%
	2	38%	47%	15%	62%
	3	20%	58%	22%	80%
	4	13%	58%	29%	87%
8	1	49%	36%	14%	51%
	2	37%	49%	13%	63%
	3	22%	58%	20%	78%
	4	6%	68%	26%	94%
10	1	48%	38%	15%	52%
	2	35%	44%	21%	65%
	3	29%	51%	20%	71%
	4	5%	50%	45%	95%

Mathematics

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
3	1	63%	21%	16%	37%
	2	30%	49%	21%	70%
	3	20%	60%	19%	80%
	4	16%	55%	29%	84%
4	1	64%	27%	9%	36%
	2	41%	42%	17%	59%
	3	27%	52%	21%	73%
	4	13%	58%	28%	87%
5	1	70%	21%	9%	30%
	2	41%	38%	20%	59%
	3	33%	51%	16%	67%
	4	30%	40%	31%	70%
6	1	46%	39%	15%	54%
	2	39%	45%	15%	61%
	3	20%	61%	18%	80%
	4	14%	63%	23%	86%
7	1	38%	52%	10%	62%
	2	46%	42%	12%	54%
	3	26%	54%	20%	74%
	4	17%	57%	26%	83%
8	1	49%	41%	10%	51%
	2	35%	48%	17%	65%
	3	25%	58%	17%	75%
	4	16%	62%	22%	84%
10	1	40%	33%	28%	60%
	2	42%	38%	20%	58%
	3	31%	53%	16%	69%
	4	25%	51%	24%	75%

Science

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
5	1	85%	12%	4%	15%
	2	79%	15%	6%	21%
	3	78%	18%	4%	22%
	4	71%	10%	19%	29%
8	1	80%	18%	2%	20%
	2	69%	28%	3%	31%
	3	69%	22%	9%	31%
	4	56%	42%	3%	44%
11	1	68%	22%	10%	32%
	2	74%	21%	5%	26%
	3	60%	35%	5%	40%
	4	32%	59%	9%	68%

Reference: 2020–2021

ELA

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
3	1	32%	54%	14%	68%
	2	48%	29%	23%	52%
	3	57%	17%	26%	43%
	4	62%	14%	24%	38%
4	1	31%	52%	17%	69%
	2	45%	37%	18%	55%
	3	56%	21%	23%	44%
	4	51%	24%	24%	49%
5	1	29%	58%	13%	71%
	2	44%	38%	18%	56%
	3	57%	25%	19%	43%
	4	28%	40%	33%	72%
6	1	40%	41%	19%	61%
	2	55%	33%	13%	46%
	3	58%	20%	22%	42%
	4	63%	12%	25%	37%
7	1	48%	44%	8%	52%
	2	47%	38%	15%	53%
	3	58%	20%	22%	42%
	4	58%	13%	29%	42%
8	1	36%	49%	14%	64%
	2	49%	37%	13%	51%
	3	58%	22%	20%	42%
	4	68%	6%	26%	32%
10	1	38%	48%	15%	63%
	2	44%	35%	21%	56%
	3	51%	29%	20%	49%
	4	50%	5%	45%	50%

Mathematics

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
3	1	21%	63%	16%	79%
	2	49%	30%	21%	51%
	3	60%	20%	19%	40%
	4	55%	16%	29%	45%
4	1	27%	64%	9%	73%
	2	42%	41%	17%	58%
	3	52%	27%	21%	48%
	4	58%	13%	28%	42%
5	1	21%	70%	9%	79%
	2	38%	41%	20%	62%
	3	51%	33%	16%	49%
	4	40%	30%	31%	60%
6	1	39%	46%	15%	61%
	2	45%	39%	15%	55%
	3	61%	20%	18%	39%
	4	63%	14%	23%	37%
7	1	52%	38%	10%	48%
	2	42%	46%	12%	58%
	3	54%	26%	20%	46%
	4	57%	17%	26%	43%
8	1	41%	49%	10%	59%
	2	48%	35%	17%	52%
	3	58%	25%	17%	42%
	4	62%	16%	22%	38%
10	1	33%	40%	28%	67%
	2	38%	42%	20%	62%
	3	53%	31%	16%	47%
	4	51%	25%	24%	49%

Science

Grade	WA-AIM Student Achievement Level	Teacher Rating on WA-AIM's IEP Goal Alignment			
		Limited	*Somewhat Represented*	*Well Represented*	*Somewhat or Well Represented*
5	1	12%	85%	4%	88%
	2	15%	79%	6%	85%
	3	18%	78%	4%	82%
	4	10%	71%	19%	90%
8	1	18%	80%	2%	82%
	2	28%	69%	3%	72%
	3	22%	69%	9%	79%
	4	42%	56%	3%	58%
11	1	22%	68%	10%	78%
	2	21%	74%	5%	79%
	3	35%	60%	5%	65%
	4	59%	32%	9%	41%

Chapter 7. Fairness in Testing

7.1 Types of Evidence

Fairness is "central to the validity and comparability of the interpretation of test scores for intended uses" (AERA, APA, & NCME, 2014, p. 63). Tests should be as fair as possible for test takers of different races, gender, ethnic backgrounds, or disability status. Fairness permeates all aspects of testing. The Code of Fair Testing Practices in Education (Joint Committee on Testing Practices [JCTP], 2004) provides guidelines in four critical areas:

- *developing and selecting appropriate tests*
- *administering and scoring tests*
- *reporting and interpreting test results*
- *informing test takers about the nature of the test, test taker rights and responsibilities, the appropriate use of scores, and procedures for resolving challenges to scores*

Similarly, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) includes standards on fairness in four areas:

- *Cluster 1. Test design, development, administration, and scoring procedures that minimize barriers to valid score interpretations for the widest possible range of individuals and relevant subgroups*.
- *Cluster 2. Validity of test score interpretations for intended uses for the intended examinee population*.
- *Cluster 3. Accommodation to remove construct-irrelevant barriers and support valid interpretations of scores for their intended users*.
- *Cluster 4. Safeguards against inappropriate score interpretations for intended uses*.

Standards that are pertinent to the WA-AIM are listed in the following section. Procedural and empirical evidence that addresses each standard is presented throughout this report and summarized in this chapter.

7.2 Summary

In this section, the standards are grouped by relevance of their supporting evidence. Each group of standards is followed by a summary of related evidence for the WA-AIM administration, as well as a list of sections in which detailed information about the evidence is provided.

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

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

(***Standard 3.13***) *A test should be administered in the language that is most relevant and appropriate to the test purpose.*

Summary: Accessibility considerations were built into the design and development of the assessment and its associated components and operations, such as the selection of target general education learning standards, design and development of the Access Point Frameworks and Performance Task specifications, training workshops and materials, and administration procedures and materials, as well as the design and procedures of sampling and participant selection in the weighting study, standard setting studies, and development of AALDs.

Reference sections: Chapters 1, 2 and 3

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

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

*(**Standard 3.9**) Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.*

*(**Standard 3.10**) When test accommodations are permitted, test developers and/or test users are responsible for documenting standard provisions for using the accommodation and for monitoring the appropriate implementation of the accommodation.*

*(**Standard 3.11**) When a test is changed to remove barriers to the accessibility of the construct being measured, test developers and/or users are responsible for obtaining and documenting evidence of the validity of score interpretations for intended use of the changed test, when sample sizes permit.*

Summary: Standardized Performance Task specifications, test items, and administration procedures were developed, thoroughly documented, and communicated to educators through training and administration manuals. The online platform was also set to facilitate standardized form assembly, test administration, documentation and submission of related assessment conditions, provisions, accommodations, and student performance data. Observer attestation forms were specified and required to monitor teacher implementation of requirements. In addition, independent data auditing by DRC Alternate Assessment Auditing team was conducted to ensure adherence to requirements on Performance Tasks, data submission, and documentation.

Reference sections: Chapter 3, and Sections 4.1 and 4.2 of Chapter 4

*(**Standard 3.8**) When tests require the scoring of constructed responses, test developers and/or users should collect and report evidence of the validity of score interpretations for relevant subgroups in the intended population of test takers for the intended use of the test scores.*

*(**Standard 3.15**) Test developers and publishers who claim that a test can be used with examinees from specific subgroups are responsible for providing the necessary information to support appropriate test score interpretations for their intended uses for individuals from these subgroups.*

Summary: Descriptive and inferential statistics on subgroup categories including gender, low-income status, EL status, and race/ethnicity are provided. The test reliability was of reasonable range for reported subgroups, taking into consideration of their sample sizes. The mean score difference was either trivial or small within each subgroup across grades and content areas. In addition, only one large DIF was detected at the standard level (in high school mathematics on ELL status) and the associated total test score difference was not significant and small.

Reference section: Section 6.8 of Chapter 6

Chapter 8. Reliability and Validity

8.1 Types of Evidence

"Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (AERA, APA, & NCME, 2014, p. 11).

The purpose of test validation is not to validate the test itself but to validate interpretations to be made of the test scores for particular purposes or uses. It should be noted that validation is not a quantifiable property but an ongoing process or argument, beginning at initial conceptualization and continuing throughout the assessment process (Kane, 2006, pp. 131–152). Every aspect of an assessment may provide evidence in support of or contrary to its validity, including but not limited to design, content specifications, item development, psychometric quality, and inferences based on the results.

Reliability, though a necessary condition of validity, alone does support the entirety of the validity argument construction. Reliability refers to the consistency of students' test scores on parallel forms or administrations of a test. A reliable test is one that produces scores that are expected to hold relative stability if the test is administered repeatedly under similar conditions. Often, however, it is impractical to administer multiple forms of the test to the same student, thus reliability is estimated on a single administration of the test. This type of reliability, known as internal consistency, provides an estimate of how consistently examinees perform across items within a test during a single test administration (Crocker & Algina, 1986).

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) specifies standards on reliability and validity categorized in eight clusters and three clusters, respectively, as listed below.

Reliability/Precision

- *Cluster 1. Specifications for replications of the testing procedure*
- *Cluster 2. Evaluating reliability/precision*
- *Cluster 3. Reliability/generalizability coefficients*
- *Cluster 4. Factors affecting reliability/precision*
- *Cluster 5. Standard errors of measurement*
- *Cluster 6. Decision consistency*
- *Cluster 7. Reliability/precision of group means*
- *Cluster 8. Documenting reliability/precision*

Validity

- *Cluster 1. Establishing intended uses and interpretations*
- *Cluster 2. Issues regarding samples and settings used in validation*
- *Cluster 3. Specific forms of validity evidence*
    - *(a) Content-oriented evidence*
    - *(b) Evidence regarding cognitive processes*
    - *(c) Evidence regarding internal structure*
    - *(e) Evidence regarding relationships with conceptually related constructs*
    - *(f) Evidence regarding relationships with criteria*
    - *(g) Evidence based on consequences of tests*

8.2 Summary

Reliability and validity evidence for the 2021–22 WA-AIM assessment is described throughout this technical report. Such evidence is summarized in this section, following a similar presentation format to that for Chapter 7 (*Fairness in Testing*) of this report.

*(**Standard 1.1**) The test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The population(s) for which a test is intended should be*

delimited clearly, and the construct or constructs that the test is intended to assess should be described clearly.

Summary: The intended uses, score interpretations, test construct, and target test population were clearly defined in the design and technical documentation of the WA-AIM assessment as well as in various communications to educators such as training workshops, training materials, administration materials, and the OSPI web pages on the WA-AIM.

Reference sections: Sections 1.3–1.4 of Chapter 1, Section 3.2 of Chapter 3

*(**Standard 1.2**) A rationale should be presented for each intended interpretation of test scores for a given use, together with a summary of the evidence and theory bearing on the intended interpretation.*

*(**Standard 1.11**) When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.*

*(**Standard 1.13**) If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided.*

*(**Standard 1.3**) If validity for some common or likely interpretation for a given use has not been evaluated, or if such an interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be strongly cautioned about making unsupported interpretations.*

Summary: The rationale for the intended score interpretation of the WA-AIM assessment is presented and supported by the selection of target general education learning standards, design and development of the Access Point Frameworks and Performance Task specifications, standard setting studies, development of AALDs, results from the weighting study survey and related committee review, as well as internal consistency indices observed from the test data and discrimination power of each standard. Additionally, the observed distributions of content scores and achievement levels, and the difficulty value and access point distribution of each standard, aligned to field expectations of student performance on rigorous college-and-career readiness standards.

Reference sections: Section 1.3 of Chapter 1, and Chapters 2 and 6

*(**Standard 1.14**) When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretations should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.*

*(**Standard 1.15**) When interpretation of performance on specific items, or small subsets of items, is suggested, the rationale and relevant evidence in support of such interpretation should be provided. When interpretation of individual item responses is likely but is not recommended by the developer, the user should be warned against making such interpretations.*

Summary: In addition to content area test scores and associated achievement levels, WA-AIM reports weighted standard scores to educators. The content area test score is a composite score based on weighted standard scores. Weights are applied at the standard level to place students assessed at varying access points for a given standard on the same scale. The weights were derived empirically and with the use of expert judgment. The standards, access points, and corresponding Performance Tasks were designed to be aligned to the Access Point Frameworks for each content area and grade and supported by the administration and training materials. Empirical evidence from internal consistency measures also supports consistency of assessed standards in measuring the intended content area.

Reference sections: Section 1.6 of Chapter 1, Chapters 2 and 3, and Sections 6.6 and 6.7 of Chapter 6

*(**Standard 2.1**) The range of replications over which reliability/precision is being evaluated should be clearly stated, along with a rationale for the choice of this definition, given the testing situation.*

*(**Standard 2.2**) The evidence provided for the reliability/precision of the scores should be consistent with the domain of replications associated with the testing procedures, and with the intended interpretations for use of the test scores.*

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

*(**Standard 2.5**) Reliability estimation procedures should be consistent with the structure of the test.*

*(**Standard 2.6**) A reliability or generalizability coefficient (or standard error) that addresses one kind of variability should not be interpreted as interchangeable with indices that address other kinds of variability, unless their definitions of measurement error can be considered equivalent.*

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

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

*(**Standard 2.14**) When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.*

*(**Standard 2.16**) When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure.*

*(**Standard 2.7**) When subjective judgment enters into test scoring, evidence should be provided on both interrater consistency in scoring and within-examinee consistency over repeated measurements. A clear distinction should be made among reliability data based on (a) independent panels of raters scoring the same performance or products, (b) a single panel scoring successive performances or new products, and (c) independent panels scoring successive performances or new products.*

*(**Standard 2.8**) When constructed-response tests are scored locally, reliability/precision data should be gathered and reported for the local scoring when adequate size samples are available.*

*(**Standard 2.10**) When significant variations are permitted in tests or test administration procedures, separate reliability/precision analyses should be provided for scores produced under each major variation if adequate sample sizes are available.*

Summary: Detailed technical descriptions of the reported reliability and classification indices, along with the rationales, are presented in the report. The reported statistics include test internal consistency indices, discrimination power of each standard, overall classification consistency and accuracy indices, and classification consistency and accuracy conditional on cut scores and on achievement levels. In addition, auditor agreement from data review is reported at both the standard and item levels.

Reference sections: Section 4.2 of Chapter 4, and Sections 6.6 through 6.9 of Chapter 6

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational psychological testing*. Washington, DC: American Educational Research Association.

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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, N.J.: Lawrence Erlbaum.

Cortina J. (1993). What is coefficient alpha: an examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104.

Crawford, C. (2004). Non-linear instructional design model: Eternal, synergistic design and development. *British Journal of Educational Technology*, *35*(4), 413–420.

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

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

Dorans, N. J., & Schmitt, A. P. (1991). Constructed response and differential item functioning: A pragmatic approach. (ETS Research Report 91-47.) Princeton, NJ: Educational Testing Service.

Dynamic Learning Maps Consortium (2013). *Dynamic Learning Maps Essential Elements for English Language Arts*. Lawrence, KS: University of Kansas.

Dynamic Learning Maps Consortium (2013). *Dynamic Learning Maps Essential Elements for Mathematics*. Lawrence, KS: University of Kansas.

Dynamic Learning Maps Consortium (2016). *First Contact Survey*. Lawrence, KS: ATLAS, University of Kansas.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–760.

Kane, M. (2006). Content-related validity evidence in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook on test development* (pp. 131–153). Mahwah, NJ: Lawrence Erlbaum Associates.

Kearns, J., Kleinert, H., Kleinert, J., and Towles-Reeves, E. (2006). Learner Characteristics Inventory. Lexington, KY: University of Kentucky, National Alternate Assessment Center.

Joint Committee on Testing Practices. (2004). Code of Fair Testing Practices in Education. Retrieved July 31, 2013, from http://www.apa.org/science/programs/testing/fair-code.aspx.

Levy, R. & R.J. Mislevy. (2016). *Psychometric graphical modelling*. Boca Raton, FL: CPC Press.

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

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

National Research Council. (2012). *A framework for K-12 science education:* Practices, crosscutting concepts, and core Ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

O'Gorman, T. W. (2004). *Applied adaptive statistical methods: Tests of significance and confidence intervals*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Shavelson, R.J., & Webb, N.M. (1991). *Generalizability Theory: A Premier*. Newbury Park, CA: Sage.

Stigler, S. M. (1977), Do robust estimators work with real data? *The Annals of Statistics, 5*, 1055–1098.

Towles-Reeves, E., Kearns, J, Flowers, C., Hart, L., Kerbel, A., Kleinert, H., Quenemoen, R., & Thurlow, M. (2012). *Learner characteristics inventory project report (A product of the NCSC validity evaluation)*. Minneapolis, MN: University of Minnesota, National Center and State Collaborative.

Ramsey, P. H., & Ramsey, P. P. (2007). Optimal trimming and outlier elimination. *Journal of Modern Applied Statistical Methods*, *6*(2), 355–360.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 583–639.

Willis, J. (1995). A recursive, reflective instructional design model based on constructivist-interpretivist theory. *Educational Technology*, *35*(6), 5–23.

Appendix A. Item Review Training Presentation

Appendix B. Final Public Form Example

Appendix C. 2021–2022 WA-AIM Teacher Feedback Survey

Appendix D. DRC Data Security

Appendix E. Score Interpretation Guide