**Arizona**

**Arizona’s Instrument to Measure Standards Alternate**

**AIMS A**

**2016**

**Technical Report**

Copyright © 2017 Arizona Department of Education. All rights reserved. Only State of Arizona educators and citizens may copy, download, and/or print the document, located online at http://www.azed.gov. Any other use or reproduction of this document, in whole or in part, requires written permission of the Arizona Department of Education.

Arizona Department of Education

Assessment Section  
1535 West Jefferson  
Phoenix, AZ 85007

# Foreword

The technical information herein is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999, 2014).

# Table of Contents

[Foreword 1](#_Toc506295715)

[Table of Contents 2](#_Toc506295716)

[Figures and Tables 4](#_Toc506295717)

[Part 1: Executive Summary 5](#_Toc506295718)

[Part 2: Involvement of Arizona Educators at All Levels 8](#_Toc506295719)

[Part 3: Test Design 9](#_Toc506295720)

[3.1 Content Standards 9](#_Toc506295721)

[3.2 Test Blueprints 12](#_Toc506295722)

[3.3 Description of AIMS A 2016 Science Tests 12](#_Toc506295723)

[Part 4: Test Development 14](#_Toc506295724)

[4.1.1 AIMS A Test Development and Editing Process 14](#_Toc506295725)

[4.1.2 Blueprint Development 14](#_Toc506295726)

[4.1.3 Item Writing and Editing 14](#_Toc506295727)

[4.1.4 Item Specifications and Review Procedures 14](#_Toc506295728)

[4.1.5 Test Construction Process 15](#_Toc506295729)

[4.1.6 Quality Reviews 15](#_Toc506295730)

[Part 5: Test Administration 16](#_Toc506295731)

[5.1 Adaptations 16](#_Toc506295732)

[5.1.1 Overview of Adaptations 16](#_Toc506295733)

[5.2 Test Security 18](#_Toc506295734)

[5.3 Test Administration 19](#_Toc506295735)

[Part 6: Data for Operational Analysis 20](#_Toc506295736)

[6.1 Data 20](#_Toc506295737)

[6.2 Descriptive Statistics by Test 20](#_Toc506295738)

[6.3 Classical Item Analysis 22](#_Toc506295739)

[Part 7: Calibration, Equating, and Scaling 26](#_Toc506295740)

[7.1.1 Calibration Methods 26](#_Toc506295741)

[7.1.1 Calibration Models 26](#_Toc506295742)

[7.1.2 Calibration Software 27](#_Toc506295743)

[7.2 Calibration Results 27](#_Toc506295744)

[7.2.1 IRT Item Statistics 27](#_Toc506295745)

[7.3 Equating 31](#_Toc506295746)

[7.4 Scaling and Standard Error of Measurement 33](#_Toc506295747)

[7.4.1 Scaling Software 33](#_Toc506295748)

[Part 8: Test Results 38](#_Toc506295749)

[8.1 Data 38](#_Toc506295750)

[8.1.1 AIMS A State Test Results 38](#_Toc506295751)

[Part 9: Reliability and Validity Evidence 45](#_Toc506295752)

[9.1 Reliability 45](#_Toc506295753)

[9.1.1 Measures of Internal Consistency 45](#_Toc506295754)

[9.2 Validity 46](#_Toc506295755)

[9.2.1 Correlations of AIMS A Science to Other Assessments 47](#_Toc506295756)

[Part 10: Classification 49](#_Toc506295757)

[10.1 Standard Setting Technical Documentation 49](#_Toc506295758)

[10.2 Standard Error of Measurement at Cut Scores 49](#_Toc506295759)

[References 51](#_Toc506295760)

[APPENDIX A: AIMS A Eligibility Criteria 54](#_Toc506295761)

[APPENDIX B: Item Writer Selection Criteria 60](#_Toc506295762)

[APPENDIX C: 2016 AIMS A Monitoring Review 63](#_Toc506295763)

[APPENDIX D: Example Item Specification Card 65](#_Toc506295764)

# Figures and Tables

[Figure 3.1.1 Arizona Alternate Science Strands and Concepts – Grades 4, 8, and High School 10](#_Toc514415486)

[Table 3.2.1 AIMS A Science Blueprints 12](#_Toc514415487)

[Table 3.3.1 2016 AIMS A Science Test Structure 13](#_Toc514415488)

[Table 5.1.1 2016 AIMS A Science Adaptations Provided 17](#_Toc514415489)

[Figure 5.2.1 2016 AIMS A Test Security Agreement 18](#_Toc514415490)

[Table 6.2.1 2016 AIMS A Science Classical Test Analysis Statistics 21](#_Toc514415491)

[Table 6.2.2 2016 AIMS A Science Raw Score Test Analysis 21](#_Toc514415492)

[Table 6.3.1 2016 AIMS A Science Classical Item Analysis - Grade 4 23](#_Toc514415493)

[Table 6.3.2 2016 AIMS A Science Classical Item Analysis - Grade 8 24](#_Toc514415494)

[Table 6.3.3 2016 AIMS A Science Classical Item Analysis – High School 25](#_Toc514415495)

[Table 7.2.1.1 Weighted and Unweighted Flagged Items 28](#_Toc514415496)

[Table 7.2.1.2 2016 AIMS A Science IRT Item Statistics - Grade 4 29](#_Toc514415497)

[Table 7.2.1.3 2016 AIMS A Science IRT Item Statistics - Grade 8 30](#_Toc514415498)

[Table 7.2.1.4 2016 AIMS A Science IRT Item Statistics - High School 31](#_Toc514415499)

[Table 7.3.1 Spring 2016 AIMS A Anchor Items 32](#_Toc514415500)

[Table 7.3.2 Content Representation of 2016 Anchor Sets 32](#_Toc514415501)

[Table 7.3.3 Rasch Difficulty Representation of 2016 Anchor Sets 32](#_Toc514415502)

[Table 7.4.1 AIMS A Transformation Constants for Science Established 2009 33](#_Toc514415503)

[Table 7.4.2 2016 AIMS A Science Raw Score to Scale Score - Grade 4 35](#_Toc514415504)

[Table 7.4.3 2016 AIMS A Science Raw Score to Scale Score - Grade 8 36](#_Toc514415505)

[Table 7.4.4 2016 AIMS A Science Raw Score to Scale Score - High School 37](#_Toc514415506)

[Table 8.1.1.1 2016 AIMS A Science State Test Results Grades 4, 8, and High School 39](#_Toc514415507)

[Table 8.1.1.2 2016 AIMS A Science Frequency Distribution - Grade 4 42](#_Toc514415508)

[Table 8.1.1.3 2016 AIMS A Science Frequency Distribution - Grade 8 43](#_Toc514415509)

[Table 8.1.1.4 2016 AIMS A Science Frequency Distribution - High School 44](#_Toc514415510)

[Table 9.1. 1 2016 AIMS A Science Internal Consistency 45](#_Toc514415511)

[Table 9.2.1.1 Correlations among AIMS A Science and MSAA ELA and Mathematics – Grade 4 48](#_Toc514415512)

[Table 9.2.1.2 Correlations among AIMS A Science and MSAA ELA and Mathematics – Grade 8 48](#_Toc514415513)

[Table 10.1.1 AIMS A Scale Score Ranges by Performance Level Set in 2009 49](#_Toc514415514)

[Table 10.1.2 2016 AIMS A Science Standard Error of Measurement at Cut Scores 50](#_Toc514415515)

# 

# Part 1: Executive Summary

This document provides information regarding processes and procedures implemented in the 2016 Spring Arizona’s Instrument to Measure Standards Alternate (AIMS A) assessments for the development of tests, analysis of data, scoring, and scaling. This document also describes the results of the 2016 Spring AIMS A assessments. The technical information in this report is intended for those who evaluate tests, interpret scores, or use test results in making educational decisions.

This document also provides information relevant to the *Standards for Educational and Psychological Testing* (*Standards*, American Education Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), 1999, 2014). Each part of this technical report addresses different standards. The standards addressed by each part are listed at the beginning of each part. Part 1 of the Technical Report addresses standards 2.7, 3.2, 3.3, 6.3, 6.4, 6.15, and 13.6 from the 1999 *Standards* (AERA, APA, NCME) as well as standards 4.1, 4.2, 7.0, 7.2, 7.4, and 12.9 from the 2014 edition.

Arizona includes all students with disabilities in state-wide assessments with or without accommodations, however, a small percentage of students are unable to participate in these assessments even with accommodations. Arizona’s Instrument to Measure Standards Alternate (AIMS A) is an alternate assessment based on alternate achievement standards that was specifically developed to assess students with significant cognitive disabilities (SCDs) as prescribed by Title I of the Elementary and Secondary Education Act (ESEA) and the Individuals with Disabilities in Education Act (IDEA). The AIMS A Science test measures student ability on grade-level alternate academic standards; these standards are based on the Arizona Academic Standards, however, the breadth, depth, and complexity has been reduced as delineated in federal laws covering this population (IDEA 1412 (a) (16)).

Arizona has established eligibility criteria for students to qualify for an Alternate Assessment. Individualized Education Program (IEP) teams have been trained to utilize the AIMS A eligibility form and flow chart available at [www.azed.gov](http://www.azed.gov) to identify students with significant cognitive disabilities who would be eligible to take AIMS A. (A copy of the eligibility form can be found in Appendix A.) Students who are tested with AIMS A are students who function at developmental and instructional levels significantly below those students who are assessed with the general standardized state assessment, AIMS. In order to be considered for AIMS A Science, students must meet all three of the criteria below.

1. Evidence of Significant Cognitive Disability: Disability is determined by empirical evidence (formal testing results, multidisciplinary evaluation team results, etc.).
2. Intensity of instruction: It is difficult for the student to acquire, maintain, generalize, and apply academic skills across environments even with extensive/intensive, pervasive, frequent and individualized instruction in multiple settings.
3. Curricular Outcomes: Goals and objectives in the student’s IEP focus on enrolled grade level Alternate Arizona Academic Standards in science and grade level Arizona College and Career Ready Standards for ELA and mathematics.

Children with SCDs are a unique population of students with extremely diverse abilities as well as limitations. Kleinert, Browder, and Towles-Reeves (2005) characterized students with SCDs as students who have:

* varied levels of symbolic communication
* issues attending to salient features of stimuli
* difficulty with memory
* limited motor response repertoire
* difficulty generalizing learned information or skills
* difficulty with meta-cognition
* difficulty with skill synthesis
* sensory deficits and
* special health care needs.

IDEA, Section 1412 (a) (16), mandates that students in special education participate in the regular state assessments. If students in special education need accommodations, accommodations are provided as long as they still produce valid scores for individuals. Using non-standard accommodations, like a calculator or reading the reading passages, would invalidate the assessment and would not produce valid scores that in turn cannot be aggregated with other scores that are valid. However, alternate assessments based on alternate achievement standards are designed specifically for students with SCDs and these students require specialized instruction (Flowers, C. & Browder, D., 2004). Substantial modifications and adaptations are made to the curriculum so that students with SCDs can access the information and demonstrate what they know (Lehr, C., & Thurlow, M., 2003). Instructional adaptation strategies, like accommodations, should be implemented during daily instruction. Only those adaptations and instructional strategies used consistently during instructional activities should be made available to the students with SCDs being assessed with AIMS A. When administering AIMS A, test administrators are trained to utilize best practice strategies, adaptations, and assistive technology to ensure students have access to and are able to demonstrate what they know. Implementing adaptations specifically to meet a student’s individual needs promotes participation and progress in the general curriculum (Kleinert, H. and Kearns Farmer, J. 2001).

Items on the Multiple-Choice and performance tasks sections of AIMS A represent the essential fundamentals taught to students with significant cognitive disabilities. The Kentucky Statewide Alternate Assessment Project (1999) suggests that states create alternate assessments that mirror the elements of daily classroom instruction. Arizona’s teachers receive regular training on implementing the use of instructional adaptations as long as they allow the student to demonstrate their knowledge or respond to AIMS A items presented during the assessment administration. Teachers are trained not to influence the students’ response. While this is not an exhaustive list of adaptations, teachers are encouraged to support students’ access by utilizing any of the following (Kleinert, H. and Kearns Farmer, J. 2001; Denham, A, 2006):

* Visual/verbal cueing;
* Varied level of independence;
* Hand-over-hand assistance on performance tasks;
* Re-reading questions/passages;
* Manipulatives such as number line, calculator, clocks, or counters;
* Communication devices;
* Use symbols, pictures, or tactile objects that represent concepts.

AIMS A test administration procedures support the inclusion of assistive technology, prompting, and scaffolding to help students with SCDs demonstrate what they know. The mandatory online training for Test Coordinators, conducted by ADE Staff, emphasizes these strategies which are designed to support student achievement and success.

Assistive technology (AT) as defined by IDEA is “any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve the functional capabilities of a child with a disability.” AT has become a necessary component in ensuring academic success for some students with disabilities. Effective use of AT in daily instruction allows students to access the curriculum, facilitates testing accommodations, and helps improve the performance of students who are struggling (Satterfield & Satterfield, 2009). AIMS A allows for the use of AT as an adaptation to support student access to the online assessment and to demonstrate their knowledge.

AIMS A assesses science in Grades 4, 8, and High School. AIMS A consists of two item types for each of the content areas: Multiple-Choice items (presented to the student online) and Performance Tasks. The Multiple-Choice items include a stem and three possible answer choices. For Multiple-Choice items a score of 0 is assigned for an incorrect response and a score of 4 is assigned for a correct response. The values for these score assignments were established to allow for equal weighting of the Multiple-Choice items to the Performance Task items which are scored via a 0-4 point rubric. The Performance Tasks are standardized, constructed response items which are scored on standardized data sheets based on that 0-4 rubric. The AIMS A assessment system’s design, administration, content, and scoring were developed based on the input of, and in participation with, Arizona educators. The present Technical Report documents all aspects of the testing cycle in the subsequent chapters. The structure of the present Technical Report mirrors the testing cycle.

# Part 2: Involvement of Arizona Educators at All Levels

Part 2 of the Technical Report addresses the involvement of Arizona educators in test development. This part of the Technical Report addresses standard 3.5 of the *Standards* (AERA, APA, NCME, 1999) and standard 4.6 in the 2014 edition (AERA, APA, NCME).

Several committees met previously for the development of AIMS A Science, which includes items administered in 2016. Those committees included special education teachers, general education teachers, curriculum specialists, and other related service professionals (i.e., school psychologists and administrators). The committee participants were selected from across the state and were an integral part of the AIMS A Science test development processes and AIMS A Science results interpretation. In addition to these committees, internal teams, consisting of various Arizona Department of Education specialists and administrators, were called upon to conduct reviews to support quality assurance.

Work during the 2015 season of item/test development for 2016 AIMS A Science included a review of and modifications to existing AIMS A Science items by internal Arizona staff, all of whom had expertise in both item writing and the Arizona’s content requirements and extensive experience working with students with significant cognitive disabilities. The resulting test, with the modifications, was reviewed and approved by five external consultants. These consultants included special education, EL, and general education teachers and administrators, who had worked on Alternate Assessment Item Writing Committees in the past.

# Part 3: Test Design

## Content Standards

Part 3 of the Technical Report provides information regarding test design. The following standards from the 1999 *Standards* (AERA, APA, NCME) are addressed: 1.2, 1.6, 3.1, 3.2, 3.3, 3.11, 6.4, 6.15, 13.3, and 13.5 as well as standards 1.1, 1.11, 4.0, 4.1, 4.2, 4.12, 7.0, 7.2, 12.4, and 12.8 from the 2014 edition (AERA, APA, NCME).

AIMS A assessment is designed to measure performance on the Arizona Alternate Content Standards adopted by the Arizona State Board of Education in May 2006 for Grades 4, 8, and High School for Science. These standards are organized by strand, concept, and performance objective. Performance Objectives are specific tasks and skills that students are expected to know and be able to do. Only the strand and concept level are described below, and scores are only reported at the strand level. The AIMS A Science test blueprints are based on the concepts and strands of the Arizona Alternate Content Standards.

Figure 3.1.1 Arizona Alternate Science Strands and Concepts – Grades 4, 8, and High School

|  |  |
| --- | --- |
| **Science Grade 4** | **Science Grade 8** |
| **Strand 1: Inquiry Process**  **Concept 1: Observations, Questions, and Hypotheses**  **Concept 2: Scientific Testing (Investigating and Modeling)**  **Concept 3: Analysis and Conclusions**  **Concept 4: Communication**  **Strand 2: History and Nature of Science**  **Concept 1: History of Science as a Human Endeavor**  **Strand 3: Science in Personal and Social Perspectives**  **Concept 1: Changes in Environments**  **Concept 2: Science and Technology in Society**  **Strand 4: Life Science**  **Concept 1: Characteristics of Organisms**  **Concept 3: Organisms and Environments**  **Concept 4: Diversity, Adaptation, and Behavior**  **Strand 5: Physical Science**  **Concept 3: Energy and Magnetism**  **Strand 6: Earth and Space Science**  **Concept 2: Earth’s Processes and Systems**  **Concept 3: Changes in the Earth and Sky** | **Strand 1: Inquiry Process**  **Concept 1: Observations, Questions, and Hypotheses**  **Concept 2: Scientific Testing (Investigating and Modeling)**  **Concept 3: Analysis and Conclusions**  **Concept 4: Communication**  **Strand 2: History and Nature of Science**  **Concept 1: History of Science as a Human Endeavor**  **Strand 3: Science in Personal and Social Perspectives**  **Concept 1: Changes in Environments**  **Concept 2: Science and Technology in Society**  **Strand 4: Life Science**  **Concept 2: Reproduction and Heredity**  **Concept 4: Diversity, Adaptation, and Behavior**  **Strand 5: Physical Science**  **Concept 1: Properties and Changes of Properties in Matter**  **Concept 2: Motion and Forces** |

|  |  |
| --- | --- |
| **Science High School** | |
| **Strand 1: Inquiry Process**  **Concept 1: Observations, Questions, and Hypotheses**  **Concept 2: Scientific Testing (Investigating and Modeling)**  **Concept 3: Analysis, Conclusions, and Refinements**  **Concept 4: Communication**  **Strand 2: History and Nature of Science**  **Concept 1: History of Science as a Human Endeavor**  **Strand 3: Science in Personal and Social Perspectives**  **Concept 1: Changes in Environments**  **Concept 2: Science and Technology in Society**  **Concept 3: Human Population Characteristics**  **Strand 4: Life Science**  **Concept 1: The Cell**  **Concept 2: Molecular Basis of Heredity**  **Concept 3: Interdependence of Organisms**  **Concept 4: Biological Evolution**  **Concept 5: Matter, Energy, and Organization in Living Systems (Including Human Systems)** | **Strand 5: Physical Science**  **Concept 1: Structure and Properties of Matter**  **Concept 2: Motions and Forces**  **Concept 3: Conservation of Energy and Increase in Disorder**  **Concept 4: Chemical Reactions**  **Concept 5: Interactions of Energy and Matter**  **Strand 6: Earth and Space Science**  **Concept 1: Geochemical Cycles**  **Concept 2: Energy in the Earth System (Both Internal and External)**  **Concept 3: Origin and Evolution of the Earth System**  **Concept 4: Origin and Evolution of the Universe** |

## Test Blueprints

A test blueprint designates the percentage of items that should measure each strand and concept. All AIMS A Science assessments were designed in accordance with the following blueprints. Further discussion of item selection to match the blueprints is included in Part 4 of this report.

Table 3.2.1   
AIMS A Science Blueprints

|  | **GRADE 4** | |  | | **GRADE 8** | |  | | **HIGH SCHOOL** | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Strand** | **POs** | **Percent of Test** | | **POs** | | **Percent of Test** | | **POs** | | **Percent of Test** | |
| **Strand 1** | 10 | 30% | | 16 | | 47% | | 12 | | 27% | |
| **Strand 2** | 4 | 13% | | 5 | | 27% | | 5 | | 13% | |
| **Strand 3** |
| **Strand 4** | 12 | 57% | | 6 | | 27% | | 20 | | 60% | |
| **Strand 5** |
| **Strand 6** |
| **TOTAL** | 24 | 100% | | 25 | | 100% | | 22 | | 100% | |

## Description of AIMS A 2016 Science Tests

The test blueprints were used with the processes described in Part 4 to develop all AIMS A Science tests administered in 2016. All viable items were used to as closely as possible match the blueprint.

The AIMS A Science consisted of 15 multiple-choice items and 15 performance tasks developed by Arizona teachers. All items were scored on a basis of 4 raw score points per item. The raw scores ranged from 0-120 and scale scores were designed to range from 1000 to 1500. All items on the Science tests reported to a criterion-referenced score. All Science tests included 10 embedded field test items. The structure of the 2016 AIMS A Science test is presented in Table 3.3.1. The scale score ranges that were established in 2009 are presented with information about the standard setting process in Table 10.1.1.

Table 3.3.1   
2016 AIMS A Science Test Structure

|  |  |  |  |
| --- | --- | --- | --- |
|  | Number of Items | Multiple-Choice | Performance Tasks |
| **Grade 4** |  |  |  |
| Strand 1- Inquiry Process | 9 | 4 | 5 |
| Strands 2 & 3- History, Nature, Personal and Social | 3 | 2 | 1 |
| Strands 4, 5 & 6 - Science Content | 18 | 9 | 9 |
| **Total** | 30 | 15 | 15 |
| **Grade 8** |  |  |  |
| Strand 1- Inquiry Process | 14 | 6 | 8 |
| Strands 2 & 3-History, Nature, Personal and Social | 8 | 5 | 3 |
| Strands 4, 5 & 6 - Science Content | 8 | 4 | 4 |
| **Total** | 30 | 15 | 15 |
| **High School** |  |  |  |
| Strand 1- Inquiry Process | 8 | 1 | 7 |
| Strands 2 & 3- History, Nature, Personal and Social | 4 | 2 | 2 |
| Strands 4, 5 & 6- Science Content | 18 | 12 | 6 |
| **Total** | 30 | 15 | 15 |

# Part 4: Test Development

Part 4 of the Technical Report provides a summary of the test development activities that occurred in preparation for the spring 2016 AIMS A.

A comprehensive, multi-segment development process guides the development of assessment materials. The following section outlines this process in general terms and addresses the following standards from the 1999 *Standards* (AERA, APA, NCME): 1.6, 3.1, 3.5, 3.6, 3.7, 3.9, 3.11, 3.16, 6.4, 6.15, 7.3, 7.4, 7.7, 13.3, and 13.5 and standards 1.11, 3.2, 3.6, 4.0, 4.6, 4.7, 4.8, 4.10, 4.12, 7.0, 7.2, 12.4, 12.8 in the 2014 *Standards* (AERA, APA, NCME).

## AIMS A Test Development and Editing Process

### Blueprint Development

The development of the 2016 AIMS A assessment blueprint was derived from the 2009 blueprint and input received from the field and the Technical Advisory Committee (TAC) about the length and structure of the assessment. The length of the test was increased slightly in 2010 to allow for field-testing items.

### Item Writing and Editing

The development of the 2016 AIMS A assessments involved many educators, content specialists, and professionals from across Arizona and ADE collaborating in an effort to ensure that all newly developed items closely matched the Arizona Alternate Content Standards and the item specifications. The Arizona teachers and education professionals selected to serve on item writing committees all possessed content and assessment expertise, many of whom also had special education expertise. These committee members were selected for their ability to be creative while adhering to the test blueprint, detailed item specifications, and content limits. The participants received a considerable amount of professional development prior to writing items. Items from the previous administration were reviewed and clarified. New Multiple-Choice items were developed by Arizona teachers using a template to capture all requirements and supporting information such as strand, concept, performance objective, and content reference documentation. New Performance Tasks were constructed and reviewed by committees of special educators and content specialists. These new items were constructed in response to an internal review of the test map and a thorough gap analysis. After the item writing workshops were concluded, test items were edited and revised by in-house content specialists, assessment specialists, and research scientists for content appropriateness and standards match and were modified to match Arizona’s AIMS A Format Style Guide.

### Item Specifications and Review Procedures

Prior to item writing, ADE reviewed the Item Specifications. The Item Specifications are living documents and need to be constantly reviewed. The purpose of the review and revision was to provide further clarity for how AIMS A will measure students’ understanding of the alternate content standards. This is based on feedback from previous item writing workshops and best practices utilized in the development of AIMS A items. ADE staff reviewed the definition of what is being tested by each Performance Objective (PO) and where needed, clarified the PO statements, the content limits, and the stimulus and response attribute descriptions. Taken together, these revisions further help to inform instruction by explaining in detail what each PO means at each grade level and by describing how each PO is to be tested.

The resulting documents were used during item writing, and refinements and inputs were implemented. During item writing, it became clear that the Item Specifications would continue to require clarification and refinement in order to assure varied PO coverage within the test blueprint each year. More and varied illustrative samples for each PO need to be created each year and adapted from prior assessment items that truly reflect the item specification components and clearly test the PO. These Item Specifications will continue to be refined continuously where needed.

### Test Construction Process

Test construction for the 2016 test administration began with an internal review of the item statistics for the items used in the 2015 administration to identify, for replacement, items that were performing less than optimally. A maximum of 30 operational items were chosen to be administered for 2016. Each grade and content area was administered the same number of items. Each test form contained 15 Multiple-Choice items and 15 Performance Tasks, plus 5 field-test items of each type.

### Quality Reviews

ADE personnel implemented a series of quality review checks at various stages of production to assure all AIMS A materials were as error free as possible. ADE first reviewed each component at a relatively early stage of screen production. Items were compared to the way they were presented to the content/bias review committee to be sure no unauthorized changes had been introduced. In addition to the ADE personnel conducting the quality review checks, external consultants were acquired to conduct a thorough review of all items. During this review period, they provided comments for any suggested changes or improvement to items, instructions, materials, and online system usability. A smooth AIMS A test administration requires that all test materials, including online test, Data Sheets, Performance Task Materials, and directions to test administrators are in alignment. A final quality review of all forms and documents were reviewed and approved by ADE personnel.

# Part 5: Test Administration

Part 5 of the Technical Report describes administration procedures, including accommodations, security, and written procedures available to test administrators and school personnel. The following standards from the 1999 *Standards* (AERA, APA, NCME) are addressed: 1.13, 3.3, 3.19, 3.20, 3.21, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 6.11, 6.15, 9.1, 10.1, and 10.2 as well as standards 1.10, 3.1, 3.9, 4.2, 4.5, 4.15, 4.16, 4.21, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 7.0, 7.8 in the 2014 *Standards* (AERA, APA, NCME).

## Adaptations

### Overview of Adaptations

Some students taking the general assessment (AIMS) are allowed accommodations. Accommodations are specific practices and procedures that provide students with equitable access during instruction and assessment. Students with Significant Cognitive Disabilities (SCDs) require much more intensive instructional support which is provided through instructional adaptations. Significant adaptations and best practice strategies are necessary to develop an instructional environment to meet the unique abilities of students with SCDs. Instructional adaptation strategies, like accommodations, should be implemented during daily instruction. Only those adaptations and instructional strategies used consistently during instructional activities should be made available to the students with SCDs being assessed on AIMS A. Table 5.1.1 presents the adaptations (accommodations) provided to students during the 2016 administration.

Students identified as having a SCD are dismissed from ELL programs based on the IEP team decisions. This is in accordance with Federal and State mandates that the IEP team decisions need to be documented in the student’s IEP. This documentation drives the educational program and all services for the student and supersedes Arizona Revised Statutes and Arizona Administrative Code.

Multiple-Choice Items and Performance Tasks include text with reduced cognitive loads and are supported with graphics as appropriate. Test administrators adhere to the accommodation and adaption guidance when administering the test. To further encourage appropriate access to AIMS A, so that all students with SCDs can demonstrate their knowledge, guidance is also provided in the test instructions to utilize verbal and non-verbal support, objects, pictures, symbol systems, and manipulatives.

Any instructional adaptations or strategies can be used to support students with SCDs as long as the students indicate the response choices. Table 5.1.1 presents the number of adaptations provided to students on the 2016 AIMS A Science assessments; however, this is not an exhaustive list of adaptations that could be utilized.

Table 5.1.1   
2016 AIMS A Science Adaptations Provided

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Number of Students Using Adaptation** | | | | |
| **Adaptation** | | Grade 4 | Grade 8 | | | High School |
| Adaptive calculators | 88 | | | 157 | 147 | |
| Alphabet line | 263 | | | 205 | 150 | |
| Graph paper | 70 | | | 114 | 59 | |
| Highlight or mark key phrases, words, or letters | 288 | | | 288 | 329 | |
| Line drawings | 127 | | | 128 | 99 | |
| Magnifier | 55 | | | 51 | 17 | |
| Manipulatives | 720 | | | 564 | 455 | |
| Number line | 462 | | | 397 | 296 | |
| Other | 120 | | | 125 | 116 | |
| Picture/Object system | 327 | | | 299 | 244 | |
| Read passages or any test item/describe graphics | 862 | | | 788 | 693 | |
| Sign language | 111 | | | 90 | 48 | |
| Switch | 92 | | | 75 | 50 | |
| Symbolic/Picture system | 327 | | | 246 | 203 | |
| Use of objects | 401 | | | 331 | 249 | |
| **Total Used** | **4372** | | | **3932** | **3259** | |

Note: Students may and do use multiple adaptations on the assessment. Students may be counted in multiple cells within a column.

## Test Security

All AIMS A tests were administered under secure testing conditions. Figure 5.2.1 presents the security agreement signed by personnel involved with testing administration.

Figure 5.2.1 2016 AIMS A Test Security Agreement

**Arizona’s Instrument to Measure Standards**

**AIMS A Test Security / Testing Ethics Agreement 2016**

I acknowledge that AIMS A is a secure test, and I agree to the following conditions of use to ensure the security of the test:

1. I will take necessary precautions to safeguard test materials.

1. Limit access to persons with a responsible, professional interest in the test’s security.
2. Names of all persons having access to the materials will be kept on file by the special education director.
3. All persons having access to the AIMS A test materials (other than students to whom the test is administered) will sign the test security agreement.
4. Building administrators will maintain signed agreements of building staff.
5. Special Education Directors will maintain signed agreements of building administrators.

2. I will keep all test materials secure, limiting access to Test Administrators.

1. Test materials will be kept secure until they are actually distributed to students.
2. In no case will students be permitted to remove test materials from the room where testing takes place except under supervision of staff.

3. I will not report students’ answer choices based on previous experience outside the testing window.

4. I will attend training and properly administer all sections of AIMS A.

5. I will not examine the AIMS A to determine the content beyond the requirements to administer the test.

1. No content of the test will be disclosed or allowed to be disclosed.
2. No test item will be discussed at any time.

6. After completing the test administration, I will store all testing materials, including student data sheets, in a secure area.

7. I will not use any test materials for instruction before or after test administration.

8. I understand the district superintendent or charter operator will develop, distribute, and enforce disciplinary procedures for the violation of test security by district or agency staff.

Individuals that will be administering the AIMS A for 2016 must also:

* participate in training activities prior to administering the AIMS A;
* review *AIMS A Test Administration Directions* for 2016 prior to test date;
* follow *AIMS A Test Administration Directions*; and
* secure all AIMS A test materials upon completion of testing, including all student data sheets.

**By signing my name to this document, I am assuring my district/charter and the Arizona Department of Education that I will abide by the above conditions and that anyone I supervise who will have access to the 2016 AIMS A test will also sign a Test Security Agreement.**

**Signed By:**

**Printed Name:**

**Title:**

**School:**

## Test Administration

To ensure standardized testing administration for all students, the *AIMS A Science Test Coordinator Manual* was made available to all special education directors for the spring 2016 administration. The manual included the following topics:

* Federal Guidance
* Test Coordinator’s Responsibilities
* Student Eligibility Requirements
* Testing Timeline
* Procedures During Test Administration
* Procedures Following Test Administration
* Test Security.

A separate document called the *AIMS A Science Test Administration Directions* was made available to all test administrators for the spring 2016 assessments. It included the following:

* Test Administrator Responsibilities
* Test Administration Guidelines
* Testing Time Guidelines
* Information about Accommodations and Adaptations
* Access of Test Materials
* Directions on Entering Students into the System
* Detailed Scripts for Administration of Each Part of Each Test
* Detailed Instructions for Using the Performance Task Scoring Rubric
* Procedures Following Test Administration.

Online training modules were presented to AIMS A test coordinators across the state. All Public Education Agencies with AIMS A eligible students are required have an AIMS A Test Coordinator complete the mandatory online training before access to the AIMS A application system would be granted to the agency. The Test Coordinator has the responsibility of training all TA’s prior to allowing access to the AIMS A application system. The annual training PowerPoints are maintained for easy reference on ADE’s Assessment website.

# Part 6: Data for Operational Analysis

Part 6 of the Technical Report describes the data that were used for calibrating and scaling of the 2016 Spring AIMS A. This part also presents classical test statistics and item analysis statistics for each grade level. Addressed in this part of the technical report are the following standards from the 1999 *Standards* (AERA, APA, NCME): 1.5, 1.13, 2.4, 2.8, 3.18, 6.5, and 7.1. The standards from the 2014 *Standards* (AERA, APA, NCME) addressed by this chapter are: 1.8, 1.10, 2.19, 3.6, 4.14, and 7.4.

## Data

AIMS A has one test window spanning six weeks. The 2016 assessments were administered between February 15 and March 31, 2016. All results presented, except for calibration, included all students who sat for the test. For calibration, operational analysis of Science tests excluded only a small number of students who did not respond to any item. This cleaning process, designed to ensure valid calibration results, is described below.

The ADE Information Technology (IT) department, which hosts the online test and publishes the results, provided data including student responses to Multiple-Choice items (A, B, C or NR, meaning No Response), and the performance scores for each item (0, 1, 2, 3, 4). Multiple-Choice items where the student did not respond (NR) were coded within the raw score portion of the datafile as -2. These were then recoded as Omits for descriptive statistics and 0’s for calibration and score calculation.

The only cleaning process employed was to remove the few students per grade who did not respond to any items (Omits for all Multiple-Choice items and 0’s on all Performance Tasks). These students, with extreme scores, are eliminated within the WINSTEPS Item Response Theory (IRT) estimation in standard practice, Arizona, however, explicitly eliminates them prior to calibration.

Details on calibration are included in Part 7: Calibration, Equating, and Scaling.

## Descriptive Statistics by Test

The descriptive statistics presented within this section include all students regardless of responding status. Table 6.2.1 presents descriptive statistics by test (content area and grade level) which are computed with the population data in Reading, Mathematics, and Science. The table identifies the test, grade, number of students (N), the maximum obtainable raw score (Max RS), the mean raw score (Mean RS), the standard deviation of the raw score (SD RS), and Cronbach’s alpha as a measure of internal consistency by item type, Multiple-Choice (MC), and Performance Task (PT). It should be noted that the accuracy of the reliability coefficient for the Multiple-Choice portion of the test in some grade is relatively low. This may be due to the large number of non-responders in the data set, however in most grades and across all PT sections, reliability is in acceptable ranges (greater than .80).

Table 6.2.1   
2016 AIMS A Science Classical Test Analysis Statistics

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade** | **N** | **MC**  **Max RS** | **MC**  **Mean RS** | **MC**  **SD RS** | **MC**  **Reliability (alpha)** | **PT**  **Max RS** | | **PT**  **Mean RS** | **PT**  **SD RS** | **PT**  **Reliability (alpha)** |
| 4 | 1,033 | 60 | 36.01 | 14.45 | 0.86 | 60 | 40.46 | | 15.32 | 0.95 |
| 8 | 994 | 60 | 37.63 | 15.31 | 0.85 | 60 | 41.20 | | 14.72 | 0.96 |
| HS | 932 | 60 | 41.74 | 15.98 | 0.86 | 60 | 40.22 | | 16.65 | 0.97 |

Note: MC = Multiple-Choice, PT = Performance Task, Max RS = Maximum raw score, Mean RS = Mean raw score, SD RS = Raw score standard deviation.

Table 6.2.2 presents the standard Lertap analysis statistics of the raw scores for 2016 AIMS A assessment for each grade tested.

Table 6.2.2   
2016 AIMS A Science Raw Score Test Analysis

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Grade** | | |
|  | **4** | **8** | **HS** |
| Number Tested | 1,033 | 994 | 932 |
| Minimum | 0 | 0 | 0 |
| Median | 82.0 | 86.0 | 92.0 |
| Mean | 76.5 | 78.8 | 82.0 |
| Maximum | 120 | 120 | 120 |
| Std. Deviation | 27.8 | 27.6 | 30.1 |
| Variance | 771.0 | 761.4 | 908.4 |
| Range | 120 | 120 | 120 |
| Interquartile Range | 39 | 37 | 43 |
| Skewness | -0.839 | -0.919 | -0.976 |
| Kurtosis | 0.146 | 0.297 | 0.143 |
| Min. Possible | 0.00 | 0.00 | 0.00 |
| Max. Possible | 120.00 | 120.00 | 120.00 |
| # No Response | 20 | 21 | 23 |
| % No Response | 1.9% | 2.1% | 2.5% |

## Classical Item Analysis

Classical item analyses were conducted for each grade (4, 8, and High School). Tables 6.3.1 – 6.3.3 present item statistics for each Science test. Note that operational items are reported in sequence without embedded field test items. The tables show the number of students (N), the item difficulty (*p*-value), point biserial correlation (*r*pb) and biserial correlation (*r*bi) for dichotomous items, percentage of students responding to, and point biserial for the key and each distractor, and the percentage of students who omitted a Multiple-Choice item (% Omit). The point biserial correlation (*r*pb) reported is the correlation of the item and the total scores of the other items on the test. The biserial correlation (*r*bi) is a statistical measure indicating the strength of the relationship between the right answer for each item relative to the total number of correct answers for all other items on the test. It is arrived at by comparing how well students did answering one item, relative to how well they did answering all the items. These coefficients answer this question: How did the students who selected an item option do on the criterion measure? If they did well on the criterion, both (*r*pb) and (*r*bi) will be “high,” where “high” may be taken as anything over 0.30 for (*r*pb), and anything over 0.40 for (*r*bi). A low point-biserial implies that students who get the item correct tend to do poorly on the overall test, and students who get the item wrong tend to do well on the test, each of which indicates an anomaly.

Table 6.3.1   
2016 AIMS A Science Classical Item Analysis - Grade 4

**Multiple-Choice**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **Correct** | | | **Distractor 1** | | **Distractor 2** | |  |
| **Item** | ***N*** | ***p*-value** | **%** | ***r*pb** | ***r*bi** | **%** | ***r*pb** | **%** | ***r*pb** | **% Omit** |
| 1 | 998 | 0.69 | 69% | 0.38 | 0.50 | 17% | -0.13 | 11% | -0.30 | 3% |
| 2 | 999 | 0.66 | 66% | 0.57 | 0.73 | 16% | -0.35 | 15% | -0.29 | 3% |
| 3 | 992 | 0.66 | 66% | 0.49 | 0.63 | 13% | -0.25 | 17% | -0.26 | 4% |
| 4 | 993 | 0.67 | 67% | 0.52 | 0.67 | 13% | -0.26 | 16% | -0.29 | 4% |
| 5 | 993 | 0.58 | 58% | 0.38 | 0.47 | 23% | -0.15 | 15% | -0.24 | 4% |
| 6 | 995 | 0.73 | 73% | 0.45 | 0.61 | 10% | -0.23 | 14% | -0.26 | 4% |
| 7 | 993 | 0.72 | 72% | 0.52 | 0.70 | 7% | -0.23 | 17% | -0.33 | 4% |
| 8 | 991 | 0.47 | 47% | 0.26 | 0.32 | 29% | -0.06 | 20% | -0.17 | 4% |
| 9 | 992 | 0.62 | 62% | 0.49 | 0.63 | 13% | -0.25 | 21% | -0.27 | 4% |
| 10 | 993 | 0.68 | 68% | 0.45 | 0.58 | 13% | -0.24 | 15% | -0.23 | 4% |
| 11 | 994 | 0.42 | 42% | 0.18 | 0.22 | 12% | -0.17 | 42% | -0.01 | 4% |
| 12 | 991 | 0.53 | 53% | 0.41 | 0.51 | 22% | -0.30 | 21% | -0.10 | 4% |
| 13 | 991 | 0.27 | 27% | 0.13 | 0.18 | 26% | -0.10 | 44% | 0.05 | 4% |
| 14 | 994 | 0.62 | 62% | 0.47 | 0.59 | 17% | -0.28 | 17% | -0.21 | 4% |
| 15 | 995 | 0.68 | 68% | 0.37 | 0.48 | 14% | -0.13 | 15% | -0.25 | 4% |

**Performance Tasks**

|  |  | **Score 0** | | **Score 1** | | **Score 2** | | **Score 3** | | **Score 4** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | ***N*** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** |
| 16 | 1,033 | 7% | -0.52 | 15% | -0.40 | 14% | -0.12 | 23% | 0.12 | 42% | 0.53 |
| 17 | 1,033 | 8% | -0.60 | 12% | -0.37 | 10% | -0.14 | 18% | 0.04 | 52% | 0.61 |
| 18 | 1,033 | 9% | -0.58 | 19% | -0.29 | 23% | 0.05 | 25% | 0.23 | 25% | 0.36 |
| 19 | 1,033 | 9% | -0.63 | 13% | -0.36 | 14% | -0.07 | 20% | 0.11 | 44% | 0.56 |
| 20 | 1,033 | 10% | -0.59 | 18% | -0.29 | 18% | 0.03 | 26% | 0.26 | 27% | 0.37 |
| 21 | 1,033 | 8% | -0.63 | 11% | -0.33 | 11% | -0.17 | 15% | 0.03 | 56% | 0.63 |
| 22 | 1,033 | 8% | -0.57 | 13% | -0.37 | 13% | -0.12 | 21% | 0.15 | 45% | 0.52 |
| 23 | 1,033 | 9% | -0.62 | 13% | -0.36 | 14% | -0.09 | 20% | 0.16 | 44% | 0.53 |
| 24 | 1,033 | 8% | -0.63 | 9% | -0.38 | 8% | -0.17 | 13% | -0.02 | 62% | 0.68 |
| 25 | 1,033 | 6% | -0.59 | 13% | -0.39 | 11% | -0.11 | 19% | 0.05 | 50% | 0.57 |
| 26 | 1,033 | 7% | -0.59 | 14% | -0.38 | 14% | -0.12 | 19% | 0.11 | 46% | 0.56 |
| 27 | 1,033 | 9% | -0.61 | 16% | -0.36 | 12% | -0.09 | 15% | 0.08 | 48% | 0.62 |
| 28 | 1,033 | 12% | -0.62 | 22% | -0.27 | 18% | 0.07 | 20% | 0.22 | 28% | 0.44 |
| 29 | 1,033 | 12% | -0.61 | 22% | -0.29 | 18% | 0.05 | 18% | 0.19 | 31% | 0.49 |
| 30 | 1,033 | 12% | -0.59 | 22% | -0.26 | 21% | 0.05 | 21% | 0.25 | 25% | 0.41 |

Table 6.3.2   
2016 AIMS A Science Classical Item Analysis - Grade 8

**Multiple-Choice**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **Correct** | | | **Distractor 1** | | **Distractor 2** | |  |
| **Item** | ***N*** | ***p*-value** | **%** | ***r*pb** | ***r*bi** | **%** | ***r*pb** | **%** | ***r*pb** | **% Omit** |
| 1 | 967 | 0.74 | 74% | 0.51 | 0.68 | 12% | -0.23 | 11% | -0.38 | 3% |
| 2 | 968 | 0.75 | 75% | 0.50 | 0.68 | 13% | -0.30 | 9% | -0.31 | 3% |
| 3 | 967 | 0.59 | 59% | 0.47 | 0.59 | 15% | -0.22 | 24% | -0.32 | 3% |
| 4 | 967 | 0.63 | 63% | 0.45 | 0.57 | 19% | -0.22 | 15% | -0.32 | 3% |
| 5 | 970 | 0.56 | 56% | 0.36 | 0.46 | 24% | -0.13 | 18% | -0.31 | 2% |
| 6 | 968 | 0.59 | 59% | 0.49 | 0.62 | 13% | -0.27 | 26% | -0.30 | 3% |
| 7 | 966 | 0.62 | 62% | 0.49 | 0.63 | 18% | -0.23 | 17% | -0.34 | 3% |
| 8 | 966 | 0.49 | 49% | 0.35 | 0.44 | 19% | -0.24 | 29% | -0.16 | 3% |
| 9 | 965 | 0.77 | 77% | 0.45 | 0.62 | 9% | -0.19 | 11% | -0.33 | 3% |
| 10 | 967 | 0.71 | 71% | 0.60 | 0.80 | 12% | -0.37 | 14% | -0.35 | 3% |
| 11 | 964 | 0.39 | 39% | 0.22 | 0.28 | 19% | -0.17 | 39% | -0.06 | 3% |
| 12 | 970 | 0.75 | 75% | 0.42 | 0.57 | 14% | -0.23 | 9% | -0.29 | 2% |
| 13 | 968 | 0.70 | 70% | 0.51 | 0.67 | 11% | -0.19 | 16% | -0.40 | 3% |
| 14 | 966 | 0.63 | 63% | 0.46 | 0.59 | 19% | -0.20 | 16% | -0.35 | 3% |
| 15 | 965 | 0.50 | 50% | 0.45 | 0.57 | 31% | -0.26 | 15% | -0.25 | 3% |

**Performance Tasks**

|  |  | **Score 0** | | **Score 1** | | **Score 2** | | **Score 3** | | **Score 4** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | ***N*** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** |
| 16 | 994 | 7% | -0.52 | 15% | -0.40 | 14% | -0.12 | 23% | 0.12 | 42% | 0.53 |
| 17 | 994 | 8% | -0.60 | 12% | -0.37 | 10% | -0.14 | 18% | 0.04 | 52% | 0.61 |
| 18 | 994 | 9% | -0.58 | 19% | -0.29 | 23% | 0.05 | 25% | 0.23 | 25% | 0.36 |
| 19 | 994 | 9% | -0.63 | 13% | -0.36 | 14% | -0.07 | 20% | 0.11 | 44% | 0.56 |
| 20 | 994 | 10% | -0.59 | 18% | -0.29 | 18% | 0.03 | 26% | 0.26 | 27% | 0.37 |
| 21 | 994 | 8% | -0.63 | 11% | -0.33 | 11% | -0.17 | 15% | 0.03 | 56% | 0.63 |
| 22 | 994 | 8% | -0.57 | 13% | -0.37 | 13% | -0.12 | 21% | 0.15 | 45% | 0.52 |
| 23 | 994 | 9% | -0.62 | 13% | -0.36 | 14% | -0.09 | 20% | 0.16 | 44% | 0.53 |
| 24 | 994 | 8% | -0.63 | 9% | -0.38 | 8% | -0.17 | 13% | -0.02 | 62% | 0.68 |
| 25 | 994 | 6% | -0.59 | 13% | -0.39 | 11% | -0.11 | 19% | 0.05 | 50% | 0.57 |
| 26 | 994 | 7% | -0.59 | 14% | -0.38 | 14% | -0.12 | 19% | 0.11 | 46% | 0.56 |
| 27 | 994 | 9% | -0.61 | 16% | -0.36 | 12% | -0.09 | 15% | 0.08 | 48% | 0.62 |
| 28 | 994 | 12% | -0.62 | 22% | -0.27 | 18% | 0.07 | 20% | 0.22 | 28% | 0.44 |
| 29 | 994 | 12% | -0.61 | 22% | -0.29 | 18% | 0.05 | 18% | 0.19 | 31% | 0.49 |
| 30 | 994 | 12% | -0.59 | 22% | -0.26 | 21% | 0.05 | 21% | 0.25 | 25% | 0.41 |

Table 6.3.3   
2016 AIMS A Science Classical Item Analysis – High School

**Multiple-Choice**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **Correct** | | | **Distractor 1** | | **Distractor 2** | |  |
| **Item** | ***N*** | ***p*-value** | **%** | ***r*pb** | ***r*bi** | **%** | ***r*pb** | **%** | ***r*pb** | **% Omit** |
| 1 | 899 | 0.85 | 85% | 0.59 | 0.90 | 5% | -0.27 | 6% | -0.34 | 4% |
| 2 | 900 | 0.66 | 66% | 0.47 | 0.61 | 12% | -0.25 | 18% | -0.25 | 3% |
| 3 | 901 | 0.55 | 55% | 0.28 | -0.12 | 18% | -0.19 | 23% | -0.08 | 3% |
| 4 | 900 | 0.69 | 69% | 0.61 | 0.81 | 12% | -0.33 | 16% | -0.34 | 3% |
| 5 | 899 | 0.74 | 74% | 0.60 | 0.81 | 10% | -0.27 | 13% | -0.37 | 4% |
| 6 | 900 | 0.67 | 67% | 0.57 | 0.74 | 17% | -0.28 | 13% | -0.34 | 3% |
| 7 | 899 | 0.52 | 52% | 0.26 | 0.33 | 24% | -0.12 | 20% | -0.12 | 4% |
| 8 | 899 | 0.75 | 75% | 0.57 | 0.77 | 10% | -0.29 | 12% | -0.30 | 4% |
| 9 | 899 | 0.76 | 76% | 0.68 | 0.92 | 11% | -0.42 | 10% | -0.31 | 4% |
| 10 | 899 | 0.72 | 72% | 0.52 | 0.69 | 12% | -0.28 | 13% | -0.26 | 4% |
| 11 | 896 | 0.73 | 73% | 0.46 | 0.62 | 10% | -0.17 | 13% | -0.28 | 4% |
| 12 | 896 | 0.68 | 68% | 0.64 | 0.83 | 11% | -0.38 | 17% | -0.30 | 4% |
| 13 | 899 | 0.78 | 78% | 0.51 | 0.72 | 10% | -0.20 | 9% | -0.33 | 4% |
| 14 | 898 | 0.65 | 65% | 0.44 | 0.57 | 14% | -0.19 | 18% | -0.26 | 4% |
| 15 | 898 | 0.69 | 69% | 0.51 | 0.67 | 9% | -0.20 | 18% | -0.32 | 4% |

**Performance Tasks**

|  |  | **Score 0** | | **Score 1** | | **Score 2** | | **Score 3** | | **Score 4** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | ***N*** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** | **%** | **rpb** |
| 16 | 932 | 7% | -0.52 | 15% | -0.40 | 14% | -0.12 | 23% | 0.12 | 42% | 0.53 |
| 17 | 932 | 8% | -0.60 | 12% | -0.37 | 10% | -0.14 | 18% | 0.04 | 52% | 0.61 |
| 18 | 932 | 9% | -0.58 | 19% | -0.29 | 23% | 0.05 | 25% | 0.23 | 25% | 0.36 |
| 19 | 932 | 9% | -0.63 | 13% | -0.36 | 14% | -0.07 | 20% | 0.11 | 44% | 0.56 |
| 20 | 932 | 10% | -0.59 | 18% | -0.29 | 18% | 0.03 | 26% | 0.26 | 27% | 0.37 |
| 21 | 932 | 8% | -0.63 | 11% | -0.33 | 11% | -0.17 | 15% | 0.03 | 56% | 0.63 |
| 22 | 932 | 8% | -0.57 | 13% | -0.37 | 13% | -0.12 | 21% | 0.15 | 45% | 0.52 |
| 23 | 932 | 9% | -0.62 | 13% | -0.36 | 14% | -0.09 | 20% | 0.16 | 44% | 0.53 |
| 24 | 932 | 8% | -0.63 | 9% | -0.38 | 8% | -0.17 | 13% | -0.02 | 62% | 0.68 |
| 25 | 932 | 6% | -0.59 | 13% | -0.39 | 11% | -0.11 | 19% | 0.05 | 50% | 0.57 |
| 26 | 932 | 7% | -0.59 | 14% | -0.38 | 14% | -0.12 | 19% | 0.11 | 46% | 0.56 |
| 27 | 932 | 9% | -0.61 | 16% | -0.36 | 12% | -0.09 | 15% | 0.08 | 48% | 0.62 |
| 28 | 932 | 12% | -0.62 | 22% | -0.27 | 18% | 0.07 | 20% | 0.22 | 28% | 0.44 |
| 29 | 932 | 12% | -0.61 | 22% | -0.29 | 18% | 0.05 | 18% | 0.19 | 31% | 0.49 |
| 30 | 932 | 12% | -0.59 | 22% | -0.26 | 21% | 0.05 | 21% | 0.25 | 25% | 0.41 |

# Part 7: Calibration, Equating, and Scaling

Part 7 of the Technical Report describes the scaling procedures and results for the 2016 AIMS A Science assessments. Each grade level was scaled with calibration samples that typically consisted of the entire student population with a very few students excluded from the analysis because they did not respond to any question. These exclusionary rules were explained in Section 6.1, Data. Part 7 of this report addresses the following standards from the 1999 *Standards* (AERA, APA, NCME): 1.13, 2.1, 2.2, 2.14, 4.1, 4.2, 4.3, 6.4, 6.5, and 13.6, as well as standards 1.10, 2.3, 2.13, 2.14, 5.1, 5.2, 5.3, 7.2, 7.4, and 12.9 from the 2014 *Standards* (AERA, APA, NCME).

## Calibration Methods

Item Response Theory (IRT) models were used in the item calibration for the AIMS A Science tests. Tests were calibrated separately by grade. As an added quality control check, all calibration activities were independently conducted by two ADE staff members.

### Calibration Models

The AIMS A Science, criterion-referenced assessments, are comprised of multiple-choice items and performance task items. All items contributing to the AIMS A scores were calibrated using the Rasch (or Rasch family) models to create the scale scores. The Rasch model (Rasch, 1960; Wright, 1977) can be conceptualized as a one-parameter IRT model in which item difficulty and student ability are estimated on the same scale. The Rasch model defines a dichotomous item in terms of one parameter: item difficulty. In the Rasch model, the probability that a student with an ability estimate (*θ*) responds correctly to item *i* is

,

where  is the difficulty parameter for item *i*.

Similarly, for polytomous items (performance tasks where multiple score points are available), the Rasch family’s Masters’ partial credit model was used. Under Masters’ model, which was designed to calibrate items with multiple, ordered response categories, the probability that student *j* scores *x* on item *i* which has a maximum possible point value of *m* (*k*=*m*+1 possible response categories) can be expressed as

****.

Here, *x* = 0, 1, …, *mi* and *Dil* is a step difficulty for score *l* and is defined as

,

and can be decomposed as

,

where *bi* is the overall difficulty for item *i* and *hil* is the threshold for score point *l* (Embretson & Reise, 2000).

### Calibration Software

Parameter estimation for items on the tests using the Rasch model was implemented using Winsteps 3.73.0 (Linacre, 2011). Winsteps uses joint maximum likelihood estimation (JMLE) as described by Wright and Masters (1982).

## Calibration Results

### IRT Item Statistics

All items for the science tests converged during calibration using typical procedures for Winsteps software. Standard error (SE) of estimates for the Rasch difficulty measures indicated that the parameters were well estimated. Model to item data fit was monitored using weighted and unweighted mean-square statistics, which indicated the degree of accuracy and predictability with which the data fits the model (Linacre, 2002). In Winsteps and Rasch literature, weighted mean square is also referred to as infit and unweighted mean square is referred to as outfit. The infit statistic is sensitive to unexpected responses at or near the item’s calibrated level, whereas outfit statistic is sensitive to unexpected responses away from the item’s calibrated level. Typically, values less than 0.6 and greater than 1.4 for infit indicate misfit, and values greater than 1.4 for outfit indicate misfit (Wright & Linacre, 1994). Of the 90 operational items used in the three tests, eleven items were flagged as having misfit as indicated by infit and 30 items were flagged as having misfit as indicated by outfit. All items that were flagged for infit were also flagged for outfit. It should be noted that the amount of difference between the limits and actual measure was as little as 0.01. The items that were flagged for both infit and outfit along with low point biserial (PT.BIS) statistics and *p*-values are summarized in Table 7.2.1.1. Statistics resulting from calibration of the AIMS A Science tests for all operational items are presented in tables 7.2.1.2 through 7.2.1.4.

Table 7.2.1.1   
Weighted and Unweighted Flagged Items

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Grade** | **Item** | **INFIT** | **OUTFIT** | **PT. BIS** | ***p*-value** |
| 1 | Grade 4 | 1 |  | **9.90** |  |  |
| 2 | Grade 4 | 2 |  | **2.50** |  |  |
| 3 | Grade 4 | 5 |  | **9.90** |  |  |
| 4 | Grade 4 | 6 |  | **7.43** |  |  |
| 5 | Grade 4 | 8 | **1.51** | **2.03** | **0.25** |  |
| 6 | Grade 4 | 10 |  | **1.49** |  |  |
| 7 | Grade 4 | 11 | **1.66** | **9.90** | **0.15** |  |
| 8 | Grade 4 | 12 |  | **1.53** |  |  |
| 9 | Grade 4 | 13 | **1.63** | **5.01** | **0.08** | **0.27** |
| 10 | Grade 4 | 14 |  | **1.73** |  |  |
| 11 | Grade 4 | 15 |  | **3.42** |  |  |
| 12 | Grade 8 | 1 |  | **3.30** |  |  |
| 13 | Grade 8 | 3 |  | **2.00** |  |  |
| 14 | Grade 8 | 4 |  | **1.58** |  |  |
| 15 | Grade 8 | 5 | **1.49** | **1.83** |  |  |
| 16 | Grade 8 | 6 |  | **1.59** |  |  |
| 17 | Grade 8 | 8 | **1.42** | **1.84** |  |  |
| 18 | Grade 8 | 11 | **1.59** | **4.33** | **0.17** |  |
| 19 | Grade 8 | 12 | **1.42** | **5.59** |  |  |
| 20 | Grade 8 | 13 |  | **2.07** |  |  |
| 21 | Grade 8 | 15 |  | **1.57** |  |  |
| 22 | HS | 2 | **1.40** | **2.00** |  |  |
| 23 | HS | 3 | **1.95** | **3.05** | **0.18** |  |
| 24 | HS | 6 |  | **1.49** |  |  |
| 25 | HS | 7 | **1.89** | **4.57** | **0.19** |  |
| 26 | HS | 8 |  | **1.91** |  |  |
| 27 | HS | 10 |  | **4.97** |  |  |
| 28 | HS | 11 |  | **1.97** |  |  |
| 29 | HS | 14 | **1.46** | **1.92** |  |  |
| 30 | HS | 15 |  | **2.07** |  |  |

Table 7.2.1.2   
2016 AIMS A Science IRT Item Statistics - Grade 4

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | **Rasch Measure** | | **SE** | | **INFIT** | | **OUTFIT** | | **PT. BIS** | | ***p*-value** | |
| 1 | 0.1761 | 0.0212 | | 1.29 | | **9.90** | | 0.37 | | 0.70 | |
| 2 | 0.2144 | 0.0208 | | 0.98 | | **2.50** | | 0.56 | | 0.67 | |
| 3 | 0.2168 | 0.0208 | | 1.15 | | 1.15 | | 0.46 | | 0.68 | |
| 4 | 0.2046 | 0.0209 | | 1.04 | | 0.99 | | 0.52 | | 0.68 | |
| 5 | 0.3628 | 0.0198 | | 1.35 | | **9.90** | | 0.33 | | 0.59 | |
| 6 | 0.0931 | 0.0223 | | 1.15 | | **7.43** | | 0.47 | | 0.74 | |
| 7 | 0.0146 | 0.0235 | | 1.26 | | 1.12 | | 0.51 | | 0.73 | |
| 8 | 0.5330 | 0.0195 | | **1.51** | | **2.03** | | **0.25** | | 0.48 | |
| 9 | 0.2957 | 0.0202 | | 1.11 | | 1.24 | | 0.47 | | 0.63 | |
| 10 | 0.1904 | 0.0211 | | 1.18 | | **1.49** | | 0.44 | | 0.69 | |
| 11 | 0.6092 | 0.0196 | | **1.66** | | **9.90** | | **0.15** | | 0.43 | |
| 12 | 0.4387 | 0.0196 | | 1.25 | | **1.53** | | 0.39 | | 0.54 | |
| 13 | 0.8705 | 0.0213 | | **1.63** | | **5.01** | | **0.08** | | **0.27** | |
| 14 | 0.3053 | 0.0201 | | 1.16 | | **1.73** | | 0.43 | | 0.63 | |
| 15 | 0.2089 | 0.0209 | | 1.28 | | **3.42** | | 0.37 | | 0.69 | |
| 16 | -0.1730 | 0.0281 | | 0.73 | | 0.69 | | 0.65 | | 0.71 | |
| 17 | -0.3128 | 0.0296 | | 0.78 | | 0.70 | | 0.71 | | 0.75 | |
| 18 | 0.1157 | 0.0262 | | 0.74 | | 0.77 | | 0.53 | | 0.61 | |
| 19 | -0.1463 | 0.0278 | | 0.74 | | 0.70 | | 0.69 | | 0.70 | |
| 20 | 0.0814 | 0.0263 | | 0.79 | | 0.84 | | 0.55 | | 0.62 | |
| 21 | -0.3701 | 0.0303 | | 0.83 | | 0.74 | | 0.71 | | 0.77 | |
| 22 | -0.1603 | 0.0279 | | 0.74 | | 0.70 | | 0.67 | | 0.72 | |
| 23 | -0.1558 | 0.0279 | | 0.75 | | 0.69 | | 0.68 | | 0.70 | |
| 24 | -0.4773 | 0.0318 | | 0.83 | | 0.68 | | 0.73 | | 0.80 | |
| 25 | -0.3076 | 0.0295 | | 0.76 | | 0.70 | | 0.68 | | 0.75 | |
| 26 | -0.2255 | 0.0286 | | 0.76 | | 0.70 | | 0.68 | | 0.72 | |
| 27 | -0.1581 | 0.0279 | | 0.77 | | 0.71 | | 0.73 | | 0.70 | |
| 28 | 0.1687 | 0.0260 | | 0.77 | | 0.75 | | 0.61 | | 0.59 | |
| 29 | 0.1592 | 0.0260 | | 0.78 | | 0.76 | | 0.63 | | 0.59 | |
| 30 | 0.2706 | 0.0258 | | 0.78 | | 0.77 | | 0.58 | | 0.57 | |

Table 7.2.1.3   
2016 AIMS A Science IRT Item Statistics - Grade 8

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | **Rasch Measure** | | **SE** | | **INFIT** | | **OUTFIT** | | **PT. BIS** | | ***p*-value** | |
| 1 | 0.0884 | 0.0242 | | 1.17 | | **3.30** | | 0.52 | | 0.75 | |
| 2 | 0.0464 | 0.0250 | | 1.26 | | 1.22 | | 0.51 | | 0.77 | |
| 3 | 0.4334 | 0.0204 | | 1.11 | | **2.00** | | 0.47 | | 0.60 | |
| 4 | 0.3622 | 0.0208 | | 1.26 | | **1.58** | | 0.40 | | 0.64 | |
| 5 | 0.4115 | 0.0205 | | **1.49** | | **1.83** | | 0.31 | | 0.57 | |
| 6 | 0.4400 | 0.0203 | | 1.19 | | **1.59** | | 0.43 | | 0.60 | |
| 7 | 0.3760 | 0.0207 | | 1.17 | | 1.38 | | 0.45 | | 0.63 | |
| 8 | 0.5962 | 0.0199 | | **1.42** | | **1.84** | | 0.30 | | 0.50 | |
| 9 | 0.1008 | 0.0240 | | 1.19 | | 1.27 | | 0.46 | | 0.78 | |
| 10 | 0.2669 | 0.0217 | | 0.91 | | 1.02 | | 0.57 | | 0.72 | |
| 11 | 0.7597 | 0.0203 | | **1.59** | | **4.33** | | **0.17** | | 0.39 | |
| 12 | 0.0784 | 0.0244 | | **1.42** | | **5.59** | | 0.42 | | 0.76 | |
| 13 | 0.2319 | 0.0221 | | 1.10 | | **2.07** | | 0.49 | | 0.71 | |
| 14 | 0.3725 | 0.0208 | | 1.22 | | 1.38 | | 0.42 | | 0.64 | |
| 15 | 0.5617 | 0.0199 | | 1.18 | | **1.57** | | 0.42 | | 0.51 | |
| 16 | -0.2518 | 0.0307 | | 0.70 | | 0.62 | | 0.73 | | 0.78 | |
| 17 | -0.2532 | 0.0308 | | 0.74 | | 0.66 | | 0.73 | | 0.76 | |
| 18 | 0.0434 | 0.0278 | | 0.81 | | 0.78 | | 0.60 | | 0.67 | |
| 19 | -0.0658 | 0.0287 | | 0.79 | | 0.78 | | 0.63 | | 0.70 | |
| 20 | -0.2013 | 0.0301 | | 0.75 | | 0.68 | | 0.70 | | 0.74 | |
| 21 | -0.5073 | 0.0347 | | 0.95 | | 0.73 | | 0.73 | | 0.82 | |
| 22 | -0.2183 | 0.0303 | | 0.70 | | 0.63 | | 0.72 | | 0.78 | |
| 23 | 0.2918 | 0.0266 | | 0.78 | | 0.79 | | 0.51 | | 0.58 | |
| 24 | -0.2352 | 0.0305 | | 0.73 | | 0.65 | | 0.69 | | 0.79 | |
| 25 | -0.1948 | 0.0300 | | 0.83 | | 0.74 | | 0.72 | | 0.78 | |
| 26 | 0.3152 | 0.0265 | | 0.75 | | 0.75 | | 0.51 | | 0.57 | |
| 27 | 0.4418 | 0.0264 | | 0.86 | | 0.89 | | 0.39 | | 0.53 | |
| 28 | 0.0075 | 0.0281 | | 0.83 | | 0.83 | | 0.60 | | 0.68 | |
| 29 | 0.0604 | 0.0277 | | 0.75 | | 0.72 | | 0.64 | | 0.66 | |
| 30 | 0.1959 | 0.0269 | | 0.84 | | 0.82 | | 0.54 | | 0.62 | |

Table 7.2.1.4   
2016 AIMS A Science IRT Item Statistics - High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | **Rasch Measure** | | **SE** | | **INFIT** | | **OUTFIT** | | **PT. BIS** | | ***p*-value** | |
| 1 | -0.3573 | 0.0324 | | 1.12 | | 0.60 | | 0.54 | | 0.87 | |
| 2 | 0.1773 | 0.0233 | | **1.40** | | **2.00** | | 0.41 | | 0.68 | |
| 3 | 0.3841 | 0.0222 | | **1.95** | | **3.05** | | **0.18** | | 0.57 | |
| 4 | 0.1130 | 0.0238 | | 1.04 | | 1.15 | | 0.58 | | 0.71 | |
| 5 | -0.0010 | 0.0251 | | 1.05 | | 1.12 | | 0.55 | | 0.76 | |
| 6 | 0.1791 | 0.0232 | | 1.18 | | **1.49** | | 0.50 | | 0.69 | |
| 7 | 0.4311 | 0.0221 | | **1.89** | | **4.57** | | **0.19** | | 0.54 | |
| 8 | -0.0933 | 0.0264 | | 1.30 | | **1.91** | | 0.49 | | 0.77 | |
| 9 | -0.0342 | 0.0255 | | 0.97 | | 0.69 | | 0.60 | | 0.78 | |
| 10 | 0.0504 | 0.0244 | | 1.27 | | **4.97** | | 0.46 | | 0.74 | |
| 11 | 0.0287 | 0.0247 | | 1.37 | | **1.97** | | 0.41 | | 0.75 | |
| 12 | 0.1267 | 0.0237 | | 1.05 | | 1.35 | | 0.57 | | 0.70 | |
| 13 | -0.0680 | 0.0260 | | 1.24 | | 1.39 | | 0.45 | | 0.80 | |
| 14 | 0.2030 | 0.0231 | | **1.46** | | **1.92** | | 0.38 | | 0.66 | |
| 15 | 0.1154 | 0.0238 | | 1.30 | | **2.07** | | 0.45 | | 0.71 | |
| 16 | -0.4046 | 0.0331 | | 0.74 | | 0.61 | | 0.77 | | 0.79 | |
| 17 | -0.2309 | 0.0309 | | 0.70 | | 0.62 | | 0.72 | | 0.75 | |
| 18 | -0.1755 | 0.0304 | | 0.72 | | 0.65 | | 0.74 | | 0.72 | |
| 19 | 0.0609 | 0.0287 | | 0.72 | | 0.68 | | 0.67 | | 0.64 | |
| 20 | -0.2648 | 0.0313 | | 0.76 | | 0.70 | | 0.73 | | 0.74 | |
| 21 | 0.1049 | 0.0284 | | 0.77 | | 0.81 | | 0.58 | | 0.63 | |
| 22 | -0.0568 | 0.0294 | | 0.70 | | 0.66 | | 0.68 | | 0.68 | |
| 23 | 0.0962 | 0.0285 | | 0.78 | | 0.81 | | 0.63 | | 0.64 | |
| 24 | -0.2328 | 0.0310 | | 0.78 | | 0.70 | | 0.72 | | 0.73 | |
| 25 | -0.2986 | 0.0317 | | 0.80 | | 0.68 | | 0.76 | | 0.75 | |
| 26 | -0.1562 | 0.0302 | | 0.78 | | 0.72 | | 0.74 | | 0.71 | |
| 27 | 0.2364 | 0.0280 | | 0.73 | | 0.76 | | 0.59 | | 0.57 | |
| 28 | -0.3571 | 0.0324 | | 0.85 | | 0.71 | | 0.74 | | 0.77 | |
| 29 | 0.1610 | 0.0282 | | 0.77 | | 0.74 | | 0.60 | | 0.61 | |
| 30 | 0.1192 | 0.0284 | | 0.79 | | 0.79 | | 0.63 | | 0.61 | |

## Equating

The 2016 AIMS A Mathematics, Reading, and Science assessments were equated and placed on their respective operational AIMS A scale using a common-item, non-equivalent groups design. A set of anchor items was selected from the 2016 operational assessments prior to running Winsteps calibration. The anchor items were selected with two principles in mind. First, the subset of anchor items should represent the content covered by the final AIMS A assessment. Second, the subset of anchor items should be representative of the distribution of item difficulties for the full assessment. Table 7.4.1 presents the number of anchor items for each grade and subject area. Table 7.4.2 shows the content representation for the 2016 anchor items compared to the 2016 operational form for Science. Table 7.4.3 presents descriptive statistics for the 2016 anchor item difficulties and the 2016 operational form.

Table 7.3.1   
Spring 2016 AIMS A Anchor Items

|  |  |  |
| --- | --- | --- |
| Grade | Operational Total | Anchor |
| 4 | 30 | 10 |
| 8 | 30 | 10 |
| HS | 30 | 10 |

Table 7.3.2   
Content Representation of 2016 Anchor Sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade | Strand | # Items | # Anchors | % Anchors |
| 4 | 1 | 9 | 3 | 10% |
|  | 2 & 3 | 3 | 1 | 3% |
|  | 4, 5, & 6 | 18 | 6 | 20% |
| 8 | 1 | 14 | 5 | 17% |
|  | 2 & 3 | 8 | 3 | 10% |
|  | 4 & 5 | 8 | 2 | 7% |
| HS | 1 | 8 | 3 | 10% |
|  | 2 & 3 | 4 | 1 | 3% |
|  | 4, 5, & 6 | 18 | 6 | 20% |

Table 7.3.3   
Rasch Difficulty Representation of 2016 Anchor Sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade | Statistic | Test | Anchor | |
| 4 | Mb | 0.101 | | 0.009 |
|  | SDb | 0.302 | | 0.242 |
|  | MINb | -0.477 | | -0.477 |
|  | MAXb | 0.871 | | 0.305 |
| 8 | Mb | 0.152 | | 0.055 |
|  | SDb | 0.298 | | 0.273 |
|  | MINb | -0.507 | | -0.252 |
|  | MAXb | 0.760 | | 0.562 |
| HS | Mb | -0.005 | | -0.041 |
|  | SDb | 0.213 | | 0.216 |
|  | MINb | -0.405 | | -0.405 |
|  | MAXb | 0.431 | | 0.236 |

Note: Mb = Mean Rasch difficulty, SDb = Standard Deviation of the Rasch difficulty, MINb = Minimum Rasch difficulty, MAXb = Maximum Rasch difficulty.

A fixed-parameter equating process was used within Winsteps to link the 2016 AIMS A assessments to their operational scale. This was implemented by constraining the 2016 item parameter estimate of the anchor items to be equal to the final estimates obtained in the 2016 AIMS A calibration analysis. The displacement statistic, which estimates the difference between the fixed difficulty parameter and the new estimate of that parameter, if it had not been constrained, was evaluated for each anchor item. Within the Rasch literature, a displacement statistic greater than 0.50 or less than -.50 is considered significant and cause for an anchor to be removed from the anchor set. Arizona uses the more conservative criterion of .30 and -.30 to remove items from usage within the anchor set for the current calibration.

During calibration, when one or more anchors are flagged for displacement the one item with the highest absolute value is removed from the anchor set and the calibration of all items is rerun. This process is repeated until all anchor items have a displacement value between -.30 and .30. If more than one anchor item was removed from the same content strand, a replacement from the rest of the operational items used on the test is sought. For 2016 AIMS A calibration, no anchor items displayed a displacement statistic greater than .30 or less than -.30.

## Scaling and Standard Error of Measurement

A raw score to scale score table was determined for each of the Spring 2016 AIMS A Science tests. The scale of measurement was determined for each test using spring 2009 operational test results and cut scores from the subsequent standard setting. The desired AIMS A scales for Grades 4, 8, and High School ranged from 1000 to 1500. Like the AIMS Science tests, AIMS A scales are not on a vertical scale. Each grade has its own unique scale within the 1000-1500 range. The scale scores for different grades cannot be compared.

Item response theory makes available number-correct scoring. Number-correct scoring was used to derive scales scores for the AIMS A tests. With number-correct scoring, a student’s number-correct score (or raw score) is converted to a scale score through the use of transformation constants. These constants were calculated for each test and each grade. A direct linear transformation was then applied in Excel (Microsoft Corporation, 2010) to transform the logit value generated in the score file provided by Winsteps to the necessary scale score. The formula utilized for calculating the M1 and M2 values was as follows:

M1 = Desired SD/Logit SD

M2 = Desired Mean/(Logit Mean \* M1)

### Scaling Software

Excel (Microsoft Corporation, 2010) was used to compute final scale scores and associated standardized errors.

Table 7.4.1   
AIMS A Transformation Constants for Science Established 2009

|  |  |  |
| --- | --- | --- |
| Grade | M1 | M2 |
| 4 | 100.00000000000000 | 1240 |
| 8 | 83.33333333333330 | 1235 |
| High School | 75.75757575757580 | 1245 |

The desired mean for all tests was set to 1250 with a standard deviation of 25. The transformation constants were calculated based on that mean and standard deviation.

Typically, a test score is obtained from a single observation of behavior and represents an estimate of the trait being measured. As an estimate, an observed test score contains some measurement error and does not perfectly reflect an individual’s true score. The degree of measurement error in a test score can be estimated using a statistic called the standard error of measurement (*SEM*).

A student’s exact true score cannot be known. The true score is defined as the average test score that would result if the test could be administered repeatedly without the effects of practice, fatigue, or learning. The standard error of measurement is an estimate of the standard deviation of an individual’s observed scores from these repeated administrations. For practical purposes, this statistic can be used to obtain a range within which a student’s true score is likely to fall. Using item response theory, the standard error of measurement can be calculated for every possible scale score.

Tables 7.4.2 through 7.4.4 present raw score to scale score conversion tables and IRT conditional *SEM* for the AIMS A Science tests.

Table 7.4.2   
2016 AIMS A Science Raw Score to Scale Score - Grade 4

| **Raw Score** | **Scale Score** | ***SEM*** | **Raw Score** | **Scale Score** | ***SEM*** |
| --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 448 | 61 | 1261 | 11 |
| 1 | 1000 | 101 | 62 | 1262 | 11 |
| 2 | 1000 | 72 | 63 | 1264 | 11 |
| 3 | 1000 | 59 | 64 | 1265 | 11 |
| 4 | 1030 | 51 | 65 | 1266 | 11 |
| 5 | 1053 | 46 | 66 | 1268 | 11 |
| 6 | 1073 | 42 | 67 | 1269 | 11 |
| 7 | 1088 | 38 | 68 | 1270 | 11 |
| 8 | 1102 | 35 | 69 | 1272 | 11 |
| 9 | 1114 | 33 | 70 | 1273 | 11 |
| 10 | 1124 | 31 | 71 | 1274 | 12 |
| 11 | 1133 | 29 | 72 | 1276 | 12 |
| 12 | 1141 | 27 | 73 | 1277 | 12 |
| 13 | 1148 | 26 | 74 | 1278 | 12 |
| 14 | 1154 | 25 | 75 | 1280 | 12 |
| 15 | 1160 | 24 | 76 | 1281 | 12 |
| 16 | 1165 | 23 | 77 | 1282 | 12 |
| 17 | 1170 | 22 | 78 | 1284 | 12 |
| 18 | 1175 | 21 | 79 | 1285 | 12 |
| 19 | 1179 | 20 | 80 | 1287 | 12 |
| 20 | 1183 | 19 | 81 | 1288 | 12 |
| 21 | 1187 | 19 | 82 | 1289 | 12 |
| 22 | 1190 | 18 | 83 | 1291 | 12 |
| 23 | 1193 | 18 | 84 | 1292 | 12 |
| 24 | 1196 | 17 | 85 | 1294 | 12 |
| 25 | 1199 | 17 | 86 | 1296 | 13 |
| 26 | 1202 | 16 | 87 | 1297 | 13 |
| 27 | 1205 | 16 | 88 | 1299 | 13 |
| 28 | 1207 | 16 | 89 | 1301 | 13 |
| 29 | 1210 | 15 | 90 | 1302 | 13 |
| 30 | 1212 | 15 | 91 | 1304 | 13 |
| 31 | 1214 | 15 | 92 | 1306 | 14 |
| 32 | 1216 | 14 | 93 | 1308 | 14 |
| 33 | 1218 | 14 | 94 | 1310 | 14 |
| 34 | 1220 | 14 | 95 | 1311 | 14 |
| 35 | 1222 | 14 | 96 | 1314 | 14 |
| 36 | 1224 | 14 | 97 | 1316 | 15 |
| 37 | 1226 | 13 | 98 | 1318 | 15 |
| 38 | 1228 | 13 | 99 | 1320 | 15 |
| 39 | 1229 | 13 | 100 | 1322 | 16 |
| 40 | 1231 | 13 | 101 | 1325 | 16 |
| 41 | 1233 | 13 | 102 | 1328 | 16 |
| 42 | 1234 | 13 | 103 | 1330 | 17 |
| 43 | 1236 | 13 | 104 | 1333 | 17 |
| 44 | 1238 | 12 | 105 | 1336 | 18 |
| 45 | 1239 | 12 | 106 | 1340 | 19 |
| 46 | 1241 | 12 | 107 | 1343 | 19 |
| 47 | 1242 | 12 | 108 | 1347 | 20 |
| 48 | 1244 | 12 | 109 | 1351 | 21 |
| 49 | 1245 | 12 | 110 | 1356 | 22 |
| 50 | 1246 | 12 | 111 | 1361 | 24 |
| 51 | 1248 | 12 | 112 | 1367 | 25 |
| 52 | 1249 | 12 | 113 | 1374 | 28 |
| 53 | 1251 | 12 | 114 | 1383 | 30 |
| 54 | 1252 | 12 | 115 | 1393 | 34 |
| 55 | 1253 | 12 | 116 | 1407 | 40 |
| 56 | 1255 | 12 | 117 | 1426 | 48 |
| 57 | 1256 | 12 | 118 | 1456 | 62 |
| 58 | 1257 | 11 | 119 | 1500 | 94 |
| 59 | 1259 | 11 | 120 | 1500 | 446 |
| 60 | 1260 | 11 |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; SEM = Standard Error of Measurement.

Table 7.4.3   
2016 AIMS A Science Raw Score to Scale Score - Grade 8

| **Raw Score** | **Scale Score** | ***SEM*** | **Raw Score** | **Scale Score** | ***SEM*** |
| --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 373 | 61 | 1256 | 10 |
| 1 | 1000 | 84 | 62 | 1257 | 10 |
| 2 | 1010 | 60 | 63 | 1258 | 10 |
| 3 | 1044 | 49 | 64 | 1260 | 10 |
| 4 | 1069 | 42 | 65 | 1261 | 10 |
| 5 | 1088 | 37 | 66 | 1262 | 10 |
| 6 | 1103 | 34 | 67 | 1263 | 10 |
| 7 | 1115 | 31 | 68 | 1264 | 10 |
| 8 | 1126 | 29 | 69 | 1265 | 10 |
| 9 | 1135 | 27 | 70 | 1266 | 10 |
| 10 | 1143 | 25 | 71 | 1267 | 10 |
| 11 | 1150 | 24 | 72 | 1268 | 10 |
| 12 | 1157 | 22 | 73 | 1270 | 10 |
| 13 | 1163 | 21 | 74 | 1271 | 10 |
| 14 | 1168 | 20 | 75 | 1272 | 10 |
| 15 | 1172 | 19 | 76 | 1273 | 10 |
| 16 | 1177 | 19 | 77 | 1274 | 10 |
| 17 | 1181 | 18 | 78 | 1275 | 10 |
| 18 | 1184 | 17 | 79 | 1276 | 10 |
| 19 | 1188 | 17 | 80 | 1278 | 10 |
| 20 | 1191 | 16 | 81 | 1279 | 10 |
| 21 | 1194 | 16 | 82 | 1280 | 10 |
| 22 | 1197 | 15 | 83 | 1281 | 10 |
| 23 | 1200 | 15 | 84 | 1283 | 10 |
| 24 | 1202 | 14 | 85 | 1284 | 10 |
| 25 | 1204 | 14 | 86 | 1285 | 10 |
| 26 | 1207 | 14 | 87 | 1286 | 11 |
| 27 | 1209 | 13 | 88 | 1288 | 11 |
| 28 | 1211 | 13 | 89 | 1289 | 11 |
| 29 | 1213 | 13 | 90 | 1291 | 11 |
| 30 | 1215 | 13 | 91 | 1292 | 11 |
| 31 | 1217 | 12 | 92 | 1293 | 11 |
| 32 | 1218 | 12 | 93 | 1295 | 11 |
| 33 | 1220 | 12 | 94 | 1297 | 12 |
| 34 | 1222 | 12 | 95 | 1298 | 12 |
| 35 | 1223 | 12 | 96 | 1300 | 12 |
| 36 | 1225 | 11 | 97 | 1302 | 12 |
| 37 | 1227 | 11 | 98 | 1303 | 12 |
| 38 | 1228 | 11 | 99 | 1305 | 13 |
| 39 | 1229 | 11 | 100 | 1307 | 13 |
| 40 | 1231 | 11 | 101 | 1309 | 13 |
| 41 | 1232 | 11 | 102 | 1311 | 14 |
| 42 | 1234 | 11 | 103 | 1314 | 14 |
| 43 | 1235 | 10 | 104 | 1316 | 14 |
| 44 | 1236 | 10 | 105 | 1319 | 15 |
| 45 | 1238 | 10 | 106 | 1322 | 16 |
| 46 | 1239 | 10 | 107 | 1325 | 16 |
| 47 | 1240 | 10 | 108 | 1328 | 17 |
| 48 | 1241 | 10 | 109 | 1332 | 18 |
| 49 | 1243 | 10 | 110 | 1336 | 19 |
| 50 | 1244 | 10 | 111 | 1340 | 20 |
| 51 | 1245 | 10 | 112 | 1346 | 22 |
| 52 | 1246 | 10 | 113 | 1352 | 24 |
| 53 | 1247 | 10 | 114 | 1359 | 26 |
| 54 | 1248 | 10 | 115 | 1368 | 29 |
| 55 | 1250 | 10 | 116 | 1380 | 34 |
| 56 | 1251 | 10 | 117 | 1397 | 41 |
| 57 | 1252 | 10 | 118 | 1423 | 53 |
| 58 | 1253 | 10 | 119 | 1472 | 79 |
| 59 | 1254 | 10 | 120 | 1500 | 372 |
| 60 | 1255 | 10 |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; SEM = Standard Error of Measurement.

Table 7.4.4   
2016 AIMS A Science Raw Score to Scale Score - High School

| **Raw Score** | **Scale Score** | ***SEM*** | **Raw Score** | **Scale Score** | ***SEM*** |
| --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 339 | 61 | 1249 | 9 |
| 1 | 1000 | 76 | 62 | 1250 | 8 |
| 2 | 1047 | 53 | 63 | 1251 | 8 |
| 3 | 1077 | 43 | 64 | 1252 | 9 |
| 4 | 1098 | 37 | 65 | 1253 | 9 |
| 5 | 1114 | 33 | 66 | 1254 | 9 |
| 6 | 1127 | 29 | 67 | 1255 | 9 |
| 7 | 1137 | 27 | 68 | 1256 | 9 |
| 8 | 1146 | 24 | 69 | 1257 | 9 |
| 9 | 1153 | 23 | 70 | 1258 | 9 |
| 10 | 1159 | 21 | 71 | 1259 | 9 |
| 11 | 1165 | 20 | 72 | 1260 | 9 |
| 12 | 1170 | 19 | 73 | 1261 | 9 |
| 13 | 1174 | 18 | 74 | 1262 | 9 |
| 14 | 1178 | 17 | 75 | 1263 | 9 |
| 15 | 1182 | 16 | 76 | 1264 | 9 |
| 16 | 1185 | 15 | 77 | 1265 | 9 |
| 17 | 1188 | 15 | 78 | 1266 | 9 |
| 18 | 1191 | 14 | 79 | 1267 | 9 |
| 19 | 1193 | 14 | 80 | 1268 | 9 |
| 20 | 1196 | 13 | 81 | 1269 | 9 |
| 21 | 1198 | 13 | 82 | 1270 | 9 |
| 22 | 1200 | 13 | 83 | 1271 | 9 |
| 23 | 1202 | 12 | 84 | 1272 | 9 |
| 24 | 1204 | 12 | 85 | 1274 | 9 |
| 25 | 1206 | 12 | 86 | 1275 | 10 |
| 26 | 1208 | 12 | 87 | 1276 | 10 |
| 27 | 1210 | 11 | 88 | 1277 | 10 |
| 28 | 1211 | 11 | 89 | 1278 | 10 |
| 29 | 1213 | 11 | 90 | 1280 | 10 |
| 30 | 1215 | 11 | 91 | 1281 | 10 |
| 31 | 1216 | 11 | 92 | 1282 | 10 |
| 32 | 1218 | 10 | 93 | 1284 | 10 |
| 33 | 1219 | 10 | 94 | 1285 | 11 |
| 34 | 1220 | 10 | 95 | 1287 | 11 |
| 35 | 1222 | 10 | 96 | 1288 | 11 |
| 36 | 1223 | 10 | 97 | 1290 | 11 |
| 37 | 1224 | 10 | 98 | 1292 | 11 |
| 38 | 1225 | 10 | 99 | 1294 | 12 |
| 39 | 1227 | 10 | 100 | 1295 | 12 |
| 40 | 1228 | 9 | 101 | 1297 | 12 |
| 41 | 1229 | 9 | 102 | 1299 | 13 |
| 42 | 1230 | 9 | 103 | 1302 | 13 |
| 43 | 1231 | 9 | 104 | 1304 | 14 |
| 44 | 1232 | 9 | 105 | 1307 | 14 |
| 45 | 1233 | 9 | 106 | 1309 | 15 |
| 46 | 1234 | 9 | 107 | 1312 | 15 |
| 47 | 1235 | 9 | 108 | 1315 | 16 |
| 48 | 1237 | 9 | 109 | 1319 | 17 |
| 49 | 1238 | 9 | 110 | 1323 | 18 |
| 50 | 1239 | 9 | 111 | 1328 | 19 |
| 51 | 1240 | 9 | 112 | 1333 | 21 |
| 52 | 1241 | 9 | 113 | 1339 | 22 |
| 53 | 1242 | 9 | 114 | 1346 | 25 |
| 54 | 1243 | 9 | 115 | 1355 | 28 |
| 55 | 1244 | 9 | 116 | 1367 | 32 |
| 56 | 1244 | 9 | 117 | 1383 | 38 |
| 57 | 1245 | 9 | 118 | 1408 | 49 |
| 58 | 1246 | 9 | 119 | 1454 | 72 |
| 59 | 1247 | 9 | 120 | 1500 | 338 |
| 60 | 1248 | 9 |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; SEM = Standard Error of Measurement.

# Part 8: Test Results

## Data

Part 8 of this Technical Report contains information about the results of the 2016 spring administration of AIMS A Science. This section provides information on the scores from the AIMS A Science assessments. The standards from the 1999 *Standards* (AERA, APA, NCME) addressed in Part 8 include: 1.5, 4.3, 4.5, 4.6, 4.7, 6.35, 7.1, 7.10, 13.15, and 13.19 as well as standards 1.10, 5.1, 5.2, 5.3, 5.8, 5.9, 7.2, 7.4, and 12.9 from the 2014 edition (AERA, APA, NCME).

Results within this section are based on population data contained within the final electronic data files. The results in this part of the Technical Report may differ slightly from final testing results presented on the Arizona Department of Education website due to slight differences in the application of exclusion rules. Official results typically use more detailed school-level information, such as full academic year enrollment, than is used to conduct research analyses. The results in the following tables are presented as evidence of reliability and validity of the AIMS A assessments and should not be used for state accountability purposes.

### AIMS A State Test Results

The AIMS A test results for Science are each on a scale for Grades 4, 8, and High School that runs from a lowest obtainable scale score (LOSS) of 1000 to a highest obtainable scale score (HOSS) of 1500.

Test results for each grade level follow in Table 8.1.1.1. For each grade and subject, these tables present the number (*N*) of students who took the exam in 2016, the mean scale score (*M*) and standard deviation *(SD*), the percentages of students in each performance level (Falls Far Below the Standard, FFBS; Approaches the Standard, AS; Meets the Standard, MS; and Exceeds the Standard, ES) as well as the percentage of students who either had no response (NR) to any item or had their score invalidated (INV). These descriptive statistics are presented for the state as a whole and disaggregated into various demographic groups.

The scale score frequency distributions for each test are presented in Tables 8.1.1.2 through 8.1.1.4. These tables show the raw score, scale score, number of students scoring each total score (frequency, FREQ), the percent (%) of students scoring each total score, and cumulative percentage (CUML %) which is the percentage of students who scored at or below each total score.

Table 8.1.1.1   
2016 AIMS A Science State Test Results Grades 4, 8, and High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **Scale Score** | | | | **% at Performance Level** | | | | | | | | |  | |  | |
|  | ***N*** | | ***M*** | | ***SD*** | | **FFBS** | | **AS** | | **MS** | | | **ES** | | **NR** | | **INV** | |
| **Grade 4** |  | |  | |  | |  | |  | |  | |  | | |  | |  | |
| Total | 1034 | 1282.60 | | 67.37 | | 5% | | 14% | | 65% | | 16% | | | 2% | | 0% | |
| Ethnic Background |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| White | 328 | 1281.69 | | 68.64 | | 6% | | 16% | | 64% | | 15% | | | 2% | | 0% | |
| Black | 81 | 1272.59 | | 75.48 | | 7% | | 12% | | 70% | | 10% | | | 5% | | 0% | |
| Hispanic | 499 | 1285.15 | | 65.38 | | 5% | | 12% | | 66% | | 17% | | | 2% | | 0% | |
| American Indian | 67 | 1278.21 | | 77.62 | | 6% | | 19% | | 57% | | 18% | | | 5% | | 0% | |
| Asian | 30 | 1276.90 | | 42.60 | | 0% | | 20% | | 67% | | 13% | | | 0% | | 0% | |
| Hawaiian Pacific Islander | 5 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Multiracial | 24 | 1292.79 | | 57.98 | | 4% | | 4% | | 75% | | 17% | | | 0% | | 0% | |
| Other | 0 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Gender |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Male | 674 | 1284.32 | | 67.69 | | 5% | | 14% | | 64% | | 17% | | | 2% | | 0% | |
| Female | 360 | 1279.39 | | 66.75 | | 6% | | 12% | | 67% | | 15% | | | 2% | | 0% | |
| Need |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Autism | 331 | 1286.66 | | 56.69 | | 2% | | 17% | | 67% | | 15% | | | 1% | | 0% | |
| DD | 22 | 1317.09 | | 30.55 | | 0% | | 0% | | 73% | | 27% | | | 0% | | 0% | |
| ED | 7 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| EDP | 2 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| HI | 4 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| MD | 23 | 1252.22 | | 87.74 | | 9% | | 22% | | 61% | | 9% | | | 9% | | 0% | |
| MDSSI | 57 | 1187.88 | | 102.95 | | 37% | | 32% | | 30% | | 2% | | | 16% | | 0% | |
| MIID | 312 | 1310.78 | | 41.02 | | 1% | | 3% | | 72% | | 24% | | | 0% | | 0% | |
| MOID | 135 | 1265.30 | | 42.51 | | 3% | | 20% | | 76% | | 1% | | | 1% | | 0% | |
| OHI | 6 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| OI | 64 | 1244.31 | | 81.70 | | 14% | | 23% | | 56% | | 6% | | | 5% | | 0% | |
| SID | 22 | 1179.82 | | 98.56 | | 36% | | 36% | | 27% | | 0% | | | 9% | | 0% | |
| SLD | 26 | 1342.38 | | 40.78 | | 0% | | 0% | | 35% | | 65% | | | 0% | | 0% | |
| SLI | 18 | 1296.83 | | 45.33 | | 0% | | 17% | | 61% | | 22% | | | 0% | | 0% | |
| VI | 5 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| SES |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Free/Reduced Lunch | 644 | 1286.81 | | 67.34 | | 5% | | 10% | | 66% | | 19% | | | 2% | | 0% | |
| No Lunch Assistance | 390 | 1275.65 | | 66.93 | | 6% | | 19% | | 63% | | 12% | | | 2% | | 0% | |
| Migrant |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-Migrant | 1021 | 1282.51 | | 67.66 | | 5% | | 14% | | 65% | | 16% | | | 2% | | 0% | |
| Migrant | 13 | 1290.00 | | 38.76 | | 0% | | 8% | | 85% | | 8% | | | 0% | | 0% | |
| ELL |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-ELL | 985 | 1281.73 | | 68.45 | | 6% | | 14% | | 64% | | 16% | | | 2% | | 0% | |
| ELL | 49 | 1300.08 | | 36.05 | | 0% | | 8% | | 78% | | 14% | | | 0% | | 0% | |

Note: FFBS=Falls Far Below the Standard; AS=Approaches the Standard; MS=Meets the Standard; ES=Exceeds the Standard. These results are not accountability results and are presented here for purposes of addressing reliability and validity. They should not be used for accountability purposes. \* To comply with FERPA regulations, results for subgroups of less than 11 are redacted and marked instead with an \*. DD=Developmental Delay, ED=Emotional Disability, EDP=Emotional Disability-Private Placement, HI=Hearing Impairment, MD=Multiple Disabilities, MDSSI=Multiple Disabilities-Severe Sensory Impairment, MIID=Mild Intellectual Disability, MOID=Moderate Intellectual Disability, OHI=Other Health Impairment, OI=Orthopedic Impairment, SID=Severe Intellectual Disability, SLD=Specific Learning Disability, SLI=Speech/Language Impairment, VI=Visual Impairment. (Table continued.)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **Scale Score** | | | | **% at Performance Level** | | | | | | | | |  | |  | |
|  | ***N*** | | ***M*** | | ***SD*** | | **FFBS** | | **AS** | | **MS** | | | **ES** | | **NR** | | **INV** | |
| **Grade 8** |  | |  | |  | |  | |  | |  | |  | | |  | |  | |
| Total | 995 | 1278.27 | | 60.82 | | 5% | | 13% | | 63% | | 19% | | | 2% | | 0% | |
| Ethnic Background |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| White | 369 | 1280.20 | | 67.36 | | 6% | | 12% | | 60% | | 23% | | | 3% | | 0% | |
| Black | 80 | 1279.48 | | 54.24 | | 4% | | 16% | | 56% | | 24% | | | 1% | | 0% | |
| Hispanic | 438 | 1277.61 | | 53.84 | | 3% | | 15% | | 68% | | 14% | | | 1% | | 0% | |
| American Indian | 66 | 1277.05 | | 60.63 | | 6% | | 9% | | 68% | | 17% | | | 3% | | 0% | |
| Asian | 21 | 1251.00 | | 88.51 | | 10% | | 14% | | 62% | | 14% | | | 10% | | 0% | |
| Hawaiian Pacific Islander | 3 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Multiracial | 18 | 1289.78 | | 71.00 | | 17% | | 6% | | 50% | | 28% | | | 0% | | 0% | |
| Other | 0 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Gender |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Male | 618 | 1280.93 | | 59.83 | | 4% | | 14% | | 63% | | 20% | | | 2% | | 0% | |
| Female | 377 | 1273.92 | | 62.23 | | 6% | | 12% | | 65% | | 17% | | | 3% | | 0% | |
| Need |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Autism | 287 | 1279.59 | | 49.51 | | 2% | | 14% | | 70% | | 14% | | | 1% | | 0% | |
| DD | 0 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| ED | 13 | 1326.92 | | 35.62 | | 0% | | 0% | | 39% | | 62% | | | 0% | | 0% | |
| EDP | 4 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| HI | 3 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| MD | 20 | 1279.55 | | 43.62 | | 0% | | 25% | | 65% | | 10% | | | 0% | | 0% | |
| MDSSI | 52 | 1204.77 | | 88.80 | | 31% | | 31% | | 39% | | 0% | | | 12% | | 0% | |
| MIID | 338 | 1303.72 | | 33.51 | | 0% | | 2% | | 69% | | 29% | | | 0% | | 0% | |
| MOID | 140 | 1260.07 | | 33.36 | | 3% | | 29% | | 64% | | 5% | | | 0% | | 0% | |
| OHI | 8 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| OI | 65 | 1229.55 | | 101.49 | | 20% | | 15% | | 59% | | 6% | | | 12% | | 0% | |
| SID | 29 | 1196.07 | | 87.74 | | 28% | | 48% | | 24% | | 0% | | | 14% | | 0% | |
| SLD | 25 | 1330.96 | | 52.70 | | 0% | | 0% | | 40% | | 60% | | | 0% | | 0% | |
| SLI | 10 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| VI | 1 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| SES |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Free/Reduced Lunch | 587 | 1282.93 | | 55.33 | | 3% | | 13% | | 63% | | 21% | | | 1% | | 0% | |
| No Lunch Assistance | 408 | 1271.56 | | 67.45 | | 7% | | 14% | | 64% | | 15% | | | 3% | | 0% | |
| Migrant |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-Migrant | 989 | 1278.30 | | 60.97 | | 5% | | 13% | | 63% | | 19% | | | 2% | | 0% | |
| Migrant | 6 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| ELL |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-ELL | 971 | 1278.04 | | 61.46 | | 5% | | 14% | | 63% | | 19% | | | 2% | | 0% | |
| ELL | 24 | 1287.46 | | 21.45 | | 0% | | 4% | | 88% | | 8% | | | 0% | | 0% | |

Note: FFBS=Falls Far Below the Standard; AS=Approaches the Standard; MS=Meets the Standard; ES=Exceeds the Standard. These results are not accountability results and are presented here for purposes of addressing reliability and validity. They should not be used for accountability purposes. \* To comply with FERPA regulations, results for subgroups of less than 11 are redacted and marked instead with an \*. DD=Developmental Delay, ED=Emotional Disability, EDP=Emotional Disability-Private Placement, HI=Hearing Impairment, MD=Multiple Disabilities, MDSSI=Multiple Disabilities-Severe Sensory Impairment, MIID=Mild Intellectual Disability, MOID=Moderate Intellectual Disability, OHI=Other Health Impairment, OI=Orthopedic Impairment, SID=Severe Intellectual Disability, SLD=Specific Learning Disability, SLI=Speech/Language Impairment, VI=Visual Impairment. (Table continued.)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **Scale Score** | | | | **% at Performance Level** | | | | | | | | |  | |  | |
|  | ***N*** | | ***M*** | | ***SD*** | | **FFBS** | | **AS** | | **MS** | | | **ES** | | **NR** | | **INV** | |
| **High School** |  | |  | |  | |  | |  | |  | |  | | |  | |  | |
| Total | 933 | 1276.44 | | 66.54 | | 5% | | 19% | | 49% | | 27% | | | 3% | | 0% | |
| Ethnic Background |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| White | 331 | 1281.53 | | 61.53 | | 3% | | 18% | | 49% | | 30% | | | 2% | | 0% | |
| Black | 78 | 1282.53 | | 60.53 | | 4% | | 13% | | 55% | | 28% | | | 3% | | 0% | |
| Hispanic | 405 | 1273.36 | | 71.49 | | 6% | | 19% | | 50% | | 26% | | | 4% | | 0% | |
| American Indian | 68 | 1275.82 | | 61.81 | | 7% | | 22% | | 41% | | 29% | | | 2% | | 0% | |
| Asian | 26 | 1263.62 | | 44.98 | | 4% | | 31% | | 54% | | 12% | | | 0% | | 0% | |
| Hawaiian Pacific Islander | 5 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Multiracial | 20 | 1262.75 | | 76.52 | | 5% | | 25% | | 55% | | 15% | | | 5% | | 0% | |
| Other | 0 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Gender |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Male | 559 | 1276.93 | | 67.93 | | 6% | | 18% | | 49% | | 27% | | | 3% | | 0% | |
| Female | 374 | 1275.70 | | 64.48 | | 4% | | 19% | | 50% | | 27% | | | 3% | | 0% | |
| Need |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Autism | 208 | 1279.16 | | 61.14 | | 3% | | 25% | | 45% | | 27% | | | 1% | | 0% | |
| DD | 0 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| ED | 9 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| EDP | 8 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| HI | 2 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| MD | 25 | 1277.68 | | 49.44 | | 4% | | 16% | | 64% | | 16% | | | 0% | | 0% | |
| MDSSI | 49 | 1187.57 | | 102.79 | | 35% | | 41% | | 20% | | 4% | | | 20% | | 0% | |
| MIID | 352 | 1301.30 | | 39.06 | | 0% | | 4% | | 59% | | 37% | | | 0% | | 0% | |
| MOID | 125 | 1260.38 | | 29.82 | | 1% | | 37% | | 58% | | 5% | | | 0% | | 0% | |
| OHI | 8 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| OI | 72 | 1247.49 | | 77.94 | | 14% | | 29% | | 42% | | 15% | | | 4% | | 0% | |
| SID | 30 | 1154.47 | | 113.29 | | 37% | | 53% | | 7% | | 3% | | | 33% | | 0% | |
| SLD | 32 | 1326.38 | | 35.95 | | 0% | | 0% | | 28% | | 72% | | | 0% | | 0% | |
| SLI | 9 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| VI | 1 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| Other | 3 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| SES |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Free/Reduced Lunch | 552 | 1285.91 | | 55.52 | | 3% | | 17% | | 51% | | 30% | | | 1% | | 0% | |
| No Lunch Assistance | 381 | 1262.70 | | 77.90 | | 8% | | 22% | | 48% | | 22% | | | 6% | | 0% | |
| Migrant |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-Migrant | 928 | 1276.48 | | 66.52 | | 5% | | 19% | | 49% | | 27% | | | 3% | | 0% | |
| Migrant | 5 | \* | | \* | | \* | | \* | | \* | | \* | | | \* | | \* | |
| ELL |  |  | |  | |  | |  | |  | |  | | |  | |  | |
| Non-ELL | 922 | 1276.24 | | 66.84 | | 5% | | 19% | | 49% | | 27% | | | 3% | | 0% | |
| ELL | 11 | 1293.09 | | 28.91 | | 0% | | 0% | | 82% | | 18% | | | 0% | | 0% | |

Note: FFBS=Falls Far Below the Standard; AS=Approaches the Standard; MS=Meets the Standard; ES=Exceeds the Standard. These results are not accountability results and are presented here for purposes of addressing reliability and validity. They should not be used for accountability purposes. \* To comply with FERPA regulations, results for subgroups of less than 11 are redacted and marked instead with an \*. DD=Developmental Delay, ED=Emotional Disability, EDP=Emotional Disability-Private Placement, HI=Hearing Impairment, MD=Multiple Disabilities, MDSSI=Multiple Disabilities-Severe Sensory Impairment, MIID=Mild Intellectual Disability, MOID=Moderate Intellectual Disability, OHI=Other Health Impairment, OI=Orthopedic Impairment, SID=Severe Intellectual Disability, SLD=Specific Learning Disability, SLI=Speech/Language Impairment, VI=Visual Impairment.

Table 8.1.1.2   
2016 AIMS A Science Frequency Distribution - Grade 4

| **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** | **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 0 | 0.0% | 0.0% | 61 | 1261 | 8 | 0.8% | 24.6% |
| 1 | 1000 | 1 | 0.1% | 0.1% | 62 | 1262 | 8 | 0.8% | 25.4% |
| 2 | 1000 | 0 | 0.0% | 0.1% | 63 | 1264 | 13 | 1.3% | 26.7% |
| 3 | 1000 | 1 | 0.1% | 0.2% | 64 | 1265 | 12 | 1.2% | 27.9% |
| 4 | 1030 | 6 | 0.6% | 0.8% | 65 | 1266 | 12 | 1.2% | 29.1% |
| 5 | 1053 | 1 | 0.1% | 0.9% | 66 | 1268 | 11 | 1.1% | 30.1% |
| 6 | 1073 | 3 | 0.3% | 1.2% | 67 | 1269 | 15 | 1.5% | 31.6% |
| 7 | 1088 | 1 | 0.1% | 1.3% | 68 | 1270 | 15 | 1.5% | 33.1% |
| 8 | 1102 | 3 | 0.3% | 1.6% | 69 | 1272 | 11 | 1.1% | 34.2% |
| 9 | 1114 | 0 | 0.0% | 1.6% | 70 | 1273 | 11 | 1.1% | 35.3% |
| 10 | 1124 | 0 | 0.0% | 1.6% | 71 | 1274 | 11 | 1.1% | 36.4% |
| 11 | 1133 | 0 | 0.0% | 1.6% | 72 | 1276 | 7 | 0.7% | 37.0% |
| 12 | 1141 | 2 | 0.2% | 1.8% | 73 | 1277 | 18 | 1.8% | 38.8% |
| 13 | 1148 | 2 | 0.2% | 2.0% | 74 | 1278 | 10 | 1.0% | 39.8% |
| 14 | 1154 | 0 | 0.0% | 2.0% | 75 | 1280 | 11 | 1.1% | 40.9% |
| 15 | 1160 | 3 | 0.3% | 2.3% | 76 | 1281 | 14 | 1.4% | 42.3% |
| 16 | 1165 | 2 | 0.2% | 2.5% | 77 | 1282 | 15 | 1.5% | 43.7% |
| 17 | 1170 | 3 | 0.3% | 2.8% | 78 | 1284 | 17 | 1.7% | 45.4% |
| 18 | 1175 | 1 | 0.1% | 2.9% | 79 | 1285 | 10 | 1.0% | 46.4% |
| 19 | 1179 | 1 | 0.1% | 3.0% | 80 | 1287 | 9 | 0.9% | 47.3% |
| 20 | 1183 | 5 | 0.5% | 3.4% | 81 | 1288 | 15 | 1.5% | 48.8% |
| 21 | 1187 | 1 | 0.1% | 3.5% | 82 | 1289 | 15 | 1.5% | 50.2% |
| 22 | 1190 | 1 | 0.1% | 3.6% | 83 | 1291 | 20 | 2.0% | 52.2% |
| 23 | 1193 | 5 | 0.5% | 4.1% | 84 | 1292 | 12 | 1.2% | 53.4% |
| 24 | 1196 | 5 | 0.5% | 4.6% | 85 | 1294 | 12 | 1.2% | 54.6% |
| 25 | 1199 | 0 | 0.0% | 4.6% | 86 | 1296 | 14 | 1.4% | 56.0% |
| 26 | 1202 | 2 | 0.2% | 4.8% | 87 | 1297 | 16 | 1.6% | 57.5% |
| 27 | 1205 | 2 | 0.2% | 5.0% | 88 | 1299 | 11 | 1.1% | 58.6% |
| 28 | 1207 | 0 | 0.0% | 5.0% | 89 | 1301 | 12 | 1.2% | 59.8% |
| 29 | 1210 | 3 | 0.3% | 5.3% | 90 | 1302 | 10 | 1.0% | 60.8% |
| 30 | 1212 | 6 | 0.6% | 5.9% | 91 | 1304 | 10 | 1.0% | 61.8% |
| 31 | 1214 | 4 | 0.4% | 6.3% | 92 | 1306 | 13 | 1.3% | 63.1% |
| 32 | 1216 | 0 | 0.0% | 6.3% | 93 | 1308 | 20 | 2.0% | 65.0% |
| 33 | 1218 | 3 | 0.3% | 6.6% | 94 | 1310 | 16 | 1.6% | 66.6% |
| 34 | 1220 | 6 | 0.6% | 7.2% | 95 | 1311 | 17 | 1.7% | 68.3% |
| 35 | 1222 | 6 | 0.6% | 7.8% | 96 | 1314 | 14 | 1.4% | 69.7% |
| 36 | 1224 | 5 | 0.5% | 8.3% | 97 | 1316 | 21 | 2.1% | 71.7% |
| 37 | 1226 | 4 | 0.4% | 8.7% | 98 | 1318 | 21 | 2.1% | 73.8% |
| 38 | 1228 | 1 | 0.1% | 8.8% | 99 | 1320 | 23 | 2.3% | 76.1% |
| 39 | 1229 | 4 | 0.4% | 9.2% | 100 | 1322 | 20 | 2.0% | 78.0% |
| 40 | 1231 | 6 | 0.6% | 9.8% | 101 | 1325 | 18 | 1.8% | 79.8% |
| 41 | 1233 | 4 | 0.4% | 10.1% | 102 | 1328 | 20 | 2.0% | 81.8% |
| 42 | 1234 | 6 | 0.6% | 10.7% | 103 | 1330 | 18 | 1.8% | 83.5% |
| 43 | 1236 | 2 | 0.2% | 10.9% | 104 | 1333 | 17 | 1.7% | 85.2% |
| 44 | 1238 | 8 | 0.8% | 11.7% | 105 | 1336 | 19 | 1.9% | 87.1% |
| 45 | 1239 | 5 | 0.5% | 12.2% | 106 | 1340 | 21 | 2.1% | 89.2% |
| 46 | 1241 | 14 | 1.4% | 13.6% | 107 | 1343 | 8 | 0.8% | 90.0% |
| 47 | 1242 | 11 | 1.1% | 14.7% | 108 | 1347 | 21 | 2.1% | 92.0% |
| 48 | 1244 | 5 | 0.5% | 15.2% | 109 | 1351 | 11 | 1.1% | 93.1% |
| 49 | 1245 | 6 | 0.6% | 15.8% | 110 | 1356 | 14 | 1.4% | 94.5% |
| 50 | 1246 | 6 | 0.6% | 16.4% | 111 | 1361 | 11 | 1.1% | 95.6% |
| 51 | 1248 | 3 | 0.3% | 16.7% | 112 | 1367 | 13 | 1.3% | 96.8% |
| 52 | 1249 | 8 | 0.8% | 17.4% | 113 | 1374 | 7 | 0.7% | 97.5% |
| 53 | 1251 | 8 | 0.8% | 18.2% | 114 | 1383 | 4 | 0.4% | 97.9% |
| 54 | 1252 | 6 | 0.6% | 18.8% | 115 | 1393 | 6 | 0.6% | 98.5% |
| 55 | 1253 | 7 | 0.7% | 19.5% | 116 | 1407 | 5 | 0.5% | 99.0% |
| 56 | 1255 | 8 | 0.8% | 20.3% | 117 | 1426 | 2 | 0.2% | 99.2% |
| 57 | 1256 | 13 | 1.3% | 21.6% | 118 | 1456 | 1 | 0.1% | 99.3% |
| 58 | 1257 | 9 | 0.9% | 22.5% | 119 | 1500 | 5 | 0.5% | 99.8% |
| 59 | 1259 | 8 | 0.8% | 23.3% | 120 | 1500 | 2 | 0.2% | 100.0% |
| 60 | 1260 | 6 | 0.6% | 23.8% |  |  |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; FREQ = frequency, CUML % = Cumulative percentage of students.

Table 8.1.1.3   
2016 AIMS A Science Frequency Distribution - Grade 8

| **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** | **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 3 | 0.3% | 0.3% | 61 | 1256 | 5 | 0.5% | 21.7% |
| 1 | 1000 | 1 | 0.1% | 0.4% | 62 | 1257 | 16 | 1.6% | 23.3% |
| 2 | 1010 | 0 | 0.0% | 0.4% | 63 | 1258 | 9 | 0.9% | 24.2% |
| 3 | 1044 | 1 | 0.1% | 0.5% | 64 | 1260 | 13 | 1.3% | 25.6% |
| 4 | 1069 | 1 | 0.1% | 0.6% | 65 | 1261 | 6 | 0.6% | 26.2% |
| 5 | 1088 | 0 | 0.0% | 0.6% | 66 | 1262 | 12 | 1.2% | 27.4% |
| 6 | 1103 | 1 | 0.1% | 0.7% | 67 | 1263 | 15 | 1.5% | 28.9% |
| 7 | 1115 | 1 | 0.1% | 0.8% | 68 | 1264 | 16 | 1.6% | 30.6% |
| 8 | 1126 | 2 | 0.2% | 1.0% | 69 | 1265 | 8 | 0.8% | 31.4% |
| 9 | 1135 | 1 | 0.1% | 1.1% | 70 | 1266 | 10 | 1.0% | 32.4% |
| 10 | 1143 | 1 | 0.1% | 1.2% | 71 | 1267 | 8 | 0.8% | 33.2% |
| 11 | 1150 | 0 | 0.0% | 1.2% | 72 | 1268 | 10 | 1.0% | 34.3% |
| 12 | 1157 | 3 | 0.3% | 1.5% | 73 | 1270 | 10 | 1.0% | 35.3% |
| 13 | 1163 | 2 | 0.2% | 1.7% | 74 | 1271 | 13 | 1.3% | 36.6% |
| 14 | 1168 | 0 | 0.0% | 1.7% | 75 | 1272 | 7 | 0.7% | 37.3% |
| 15 | 1172 | 1 | 0.1% | 1.8% | 76 | 1273 | 9 | 0.9% | 38.2% |
| 16 | 1177 | 3 | 0.3% | 2.1% | 77 | 1274 | 11 | 1.1% | 39.4% |
| 17 | 1181 | 1 | 0.1% | 2.2% | 78 | 1275 | 9 | 0.9% | 40.3% |
| 18 | 1184 | 3 | 0.3% | 2.6% | 79 | 1276 | 8 | 0.8% | 41.1% |
| 19 | 1188 | 2 | 0.2% | 2.8% | 80 | 1278 | 13 | 1.3% | 42.4% |
| 20 | 1191 | 2 | 0.2% | 3.0% | 81 | 1279 | 15 | 1.5% | 44.0% |
| 21 | 1194 | 2 | 0.2% | 3.2% | 82 | 1280 | 6 | 0.6% | 44.6% |
| 22 | 1197 | 1 | 0.1% | 3.3% | 83 | 1281 | 12 | 1.2% | 45.8% |
| 23 | 1200 | 1 | 0.1% | 3.4% | 84 | 1283 | 8 | 0.8% | 46.6% |
| 24 | 1202 | 1 | 0.1% | 3.5% | 85 | 1284 | 13 | 1.3% | 48.0% |
| 25 | 1204 | 3 | 0.3% | 3.8% | 86 | 1285 | 20 | 2.0% | 50.0% |
| 26 | 1207 | 3 | 0.3% | 4.1% | 87 | 1286 | 13 | 1.3% | 51.3% |
| 27 | 1209 | 3 | 0.3% | 4.4% | 88 | 1288 | 23 | 2.4% | 53.7% |
| 28 | 1211 | 5 | 0.5% | 4.9% | 89 | 1289 | 17 | 1.7% | 55.4% |
| 29 | 1213 | 3 | 0.3% | 5.2% | 90 | 1291 | 18 | 1.8% | 57.3% |
| 30 | 1215 | 3 | 0.3% | 5.5% | 91 | 1292 | 13 | 1.3% | 58.6% |
| 31 | 1217 | 4 | 0.4% | 5.9% | 92 | 1293 | 11 | 1.1% | 59.7% |
| 32 | 1218 | 4 | 0.4% | 6.3% | 93 | 1295 | 22 | 2.2% | 62.0% |
| 33 | 1220 | 6 | 0.6% | 7.0% | 94 | 1297 | 20 | 2.0% | 64.0% |
| 34 | 1222 | 2 | 0.2% | 7.2% | 95 | 1298 | 19 | 1.9% | 66.0% |
| 35 | 1223 | 3 | 0.3% | 7.5% | 96 | 1300 | 18 | 1.8% | 67.8% |
| 36 | 1225 | 3 | 0.3% | 7.8% | 97 | 1302 | 19 | 1.9% | 69.7% |
| 37 | 1227 | 4 | 0.4% | 8.2% | 98 | 1303 | 16 | 1.6% | 71.4% |
| 38 | 1228 | 4 | 0.4% | 8.6% | 99 | 1305 | 17 | 1.7% | 73.1% |
| 39 | 1229 | 2 | 0.2% | 8.8% | 100 | 1307 | 17 | 1.7% | 74.8% |
| 40 | 1231 | 2 | 0.2% | 9.0% | 101 | 1309 | 25 | 2.6% | 77.4% |
| 41 | 1232 | 5 | 0.5% | 9.5% | 102 | 1311 | 17 | 1.7% | 79.1% |
| 42 | 1234 | 7 | 0.7% | 10.2% | 103 | 1314 | 19 | 1.9% | 81.1% |
| 43 | 1235 | 7 | 0.7% | 10.9% | 104 | 1316 | 24 | 2.5% | 83.5% |
| 44 | 1236 | 7 | 0.7% | 11.7% | 105 | 1319 | 18 | 1.8% | 85.4% |
| 45 | 1238 | 2 | 0.2% | 11.9% | 106 | 1322 | 9 | 0.9% | 86.3% |
| 46 | 1239 | 3 | 0.3% | 12.2% | 107 | 1325 | 18 | 1.8% | 88.1% |
| 47 | 1240 | 9 | 0.9% | 13.1% | 108 | 1328 | 9 | 0.9% | 89.1% |
| 48 | 1241 | 5 | 0.5% | 13.6% | 109 | 1332 | 18 | 1.8% | 90.9% |
| 49 | 1243 | 6 | 0.6% | 14.2% | 110 | 1336 | 13 | 1.3% | 92.2% |
| 50 | 1244 | 3 | 0.3% | 14.5% | 111 | 1340 | 12 | 1.2% | 93.5% |
| 51 | 1245 | 7 | 0.7% | 15.2% | 112 | 1346 | 16 | 1.6% | 95.1% |
| 52 | 1246 | 3 | 0.3% | 15.5% | 113 | 1352 | 9 | 0.9% | 96.0% |
| 53 | 1247 | 6 | 0.6% | 16.2% | 114 | 1359 | 10 | 1.0% | 97.0% |
| 54 | 1248 | 5 | 0.5% | 16.7% | 115 | 1368 | 6 | 0.6% | 97.6% |
| 55 | 1250 | 5 | 0.5% | 17.2% | 116 | 1380 | 6 | 0.6% | 98.3% |
| 56 | 1251 | 9 | 0.9% | 18.1% | 117 | 1397 | 3 | 0.3% | 98.6% |
| 57 | 1252 | 5 | 0.5% | 18.6% | 118 | 1423 | 5 | 0.5% | 99.1% |
| 58 | 1253 | 7 | 0.7% | 19.3% | 119 | 1472 | 5 | 0.5% | 99.6% |
| 59 | 1254 | 11 | 1.1% | 20.4% | 120 | 1500 | 4 | 0.4% | 100.0% |
| 60 | 1255 | 7 | 0.7% | 21.2% |  |  |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; FREQ = frequency, CUML % = Cumulative percentage of students.

Table 8.1.1.4   
2016 AIMS A Science Frequency Distribution - High School

| **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** | **Raw Score** | **Scale Score** | **FREQ** | **%** | **CUML %** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1000 | 2 | 0.2% | 0.2% | 61 | 1249 | 5 | 0.6% | 21.6% |
| 1 | 1000 | 0 | 0.0% | 0.2% | 62 | 1250 | 5 | 0.6% | 22.1% |
| 2 | 1047 | 0 | 0.0% | 0.2% | 63 | 1251 | 8 | 0.9% | 23.0% |
| 3 | 1077 | 0 | 0.0% | 0.2% | 64 | 1252 | 7 | 0.8% | 23.8% |
| 4 | 1098 | 1 | 0.1% | 0.3% | 65 | 1253 | 11 | 1.2% | 25.0% |
| 5 | 1114 | 0 | 0.0% | 0.3% | 66 | 1254 | 6 | 0.7% | 25.7% |
| 6 | 1127 | 0 | 0.0% | 0.3% | 67 | 1255 | 12 | 1.3% | 27.0% |
| 7 | 1137 | 0 | 0.0% | 0.3% | 68 | 1256 | 6 | 0.7% | 27.6% |
| 8 | 1146 | 2 | 0.2% | 0.6% | 69 | 1257 | 7 | 0.8% | 28.4% |
| 9 | 1153 | 0 | 0.0% | 0.6% | 70 | 1258 | 5 | 0.6% | 29.0% |
| 10 | 1159 | 1 | 0.1% | 0.7% | 71 | 1259 | 8 | 0.9% | 29.8% |
| 11 | 1165 | 1 | 0.1% | 0.8% | 72 | 1260 | 4 | 0.4% | 30.3% |
| 12 | 1170 | 2 | 0.2% | 1.0% | 73 | 1261 | 6 | 0.7% | 30.9% |
| 13 | 1174 | 2 | 0.2% | 1.2% | 74 | 1262 | 8 | 0.9% | 31.8% |
| 14 | 1178 | 1 | 0.1% | 1.3% | 75 | 1263 | 6 | 0.7% | 32.5% |
| 15 | 1182 | 1 | 0.1% | 1.4% | 76 | 1264 | 7 | 0.8% | 33.3% |
| 16 | 1185 | 7 | 0.8% | 2.2% | 77 | 1265 | 11 | 1.2% | 34.5% |
| 17 | 1188 | 0 | 0.0% | 2.2% | 78 | 1266 | 8 | 0.9% | 35.4% |
| 18 | 1191 | 0 | 0.0% | 2.2% | 79 | 1267 | 4 | 0.4% | 35.8% |
| 19 | 1193 | 1 | 0.1% | 2.3% | 80 | 1268 | 7 | 0.8% | 36.6% |
| 20 | 1196 | 1 | 0.1% | 2.4% | 81 | 1269 | 8 | 0.9% | 37.4% |
| 21 | 1198 | 1 | 0.1% | 2.5% | 82 | 1270 | 9 | 1.0% | 38.4% |
| 22 | 1200 | 1 | 0.1% | 2.6% | 83 | 1271 | 8 | 0.9% | 39.3% |
| 23 | 1202 | 3 | 0.3% | 3.0% | 84 | 1272 | 9 | 1.0% | 40.3% |
| 24 | 1204 | 2 | 0.2% | 3.2% | 85 | 1274 | 10 | 1.1% | 41.4% |
| 25 | 1206 | 1 | 0.1% | 3.3% | 86 | 1275 | 12 | 1.3% | 42.7% |
| 26 | 1208 | 2 | 0.2% | 3.5% | 87 | 1276 | 7 | 0.8% | 43.5% |
| 27 | 1210 | 6 | 0.7% | 4.2% | 88 | 1277 | 11 | 1.2% | 44.7% |
| 28 | 1211 | 1 | 0.1% | 4.3% | 89 | 1278 | 8 | 0.9% | 45.6% |
| 29 | 1213 | 3 | 0.3% | 4.6% | 90 | 1280 | 9 | 1.0% | 46.6% |
| 30 | 1215 | 2 | 0.2% | 4.8% | 91 | 1281 | 14 | 1.5% | 48.1% |
| 31 | 1216 | 3 | 0.3% | 5.2% | 92 | 1282 | 5 | 0.6% | 48.7% |
| 32 | 1218 | 4 | 0.4% | 5.6% | 93 | 1284 | 15 | 1.7% | 50.3% |
| 33 | 1219 | 2 | 0.2% | 5.8% | 94 | 1285 | 13 | 1.4% | 51.8% |
| 34 | 1220 | 1 | 0.1% | 5.9% | 95 | 1287 | 14 | 1.5% | 53.3% |
| 35 | 1222 | 5 | 0.6% | 6.5% | 96 | 1288 | 18 | 2.0% | 55.3% |
| 36 | 1223 | 2 | 0.2% | 6.7% | 97 | 1290 | 19 | 2.1% | 57.4% |
| 37 | 1224 | 5 | 0.6% | 7.3% | 98 | 1292 | 9 | 1.0% | 58.4% |
| 38 | 1225 | 1 | 0.1% | 7.4% | 99 | 1294 | 19 | 2.1% | 60.5% |
| 39 | 1227 | 8 | 0.9% | 8.3% | 100 | 1295 | 24 | 2.6% | 63.1% |
| 40 | 1228 | 5 | 0.6% | 8.8% | 101 | 1297 | 18 | 2.0% | 65.1% |
| 41 | 1229 | 5 | 0.6% | 9.4% | 102 | 1299 | 13 | 1.4% | 66.5% |
| 42 | 1230 | 8 | 0.9% | 10.2% | 103 | 1302 | 11 | 1.2% | 67.7% |
| 43 | 1231 | 4 | 0.4% | 10.7% | 104 | 1304 | 22 | 2.4% | 70.2% |
| 44 | 1232 | 6 | 0.7% | 11.3% | 105 | 1307 | 20 | 2.2% | 72.4% |
| 45 | 1233 | 5 | 0.6% | 11.9% | 106 | 1309 | 31 | 3.4% | 75.8% |
| 46 | 1234 | 5 | 0.6% | 12.4% | 107 | 1312 | 20 | 2.2% | 78.0% |
| 47 | 1235 | 1 | 0.1% | 12.6% | 108 | 1315 | 22 | 2.4% | 80.4% |
| 48 | 1237 | 8 | 0.9% | 13.4% | 109 | 1319 | 27 | 3.0% | 83.4% |
| 49 | 1238 | 6 | 0.7% | 14.1% | 110 | 1323 | 22 | 2.4% | 85.8% |
| 50 | 1239 | 6 | 0.7% | 14.8% | 111 | 1328 | 21 | 2.3% | 88.1% |
| 51 | 1240 | 5 | 0.6% | 15.3% | 112 | 1333 | 14 | 1.5% | 89.6% |
| 52 | 1241 | 5 | 0.6% | 15.9% | 113 | 1339 | 13 | 1.4% | 91.1% |
| 53 | 1242 | 7 | 0.8% | 16.6% | 114 | 1346 | 24 | 2.6% | 93.7% |
| 54 | 1243 | 10 | 1.1% | 17.7% | 115 | 1355 | 13 | 1.4% | 95.2% |
| 55 | 1244 | 4 | 0.4% | 18.2% | 116 | 1367 | 13 | 1.4% | 96.6% |
| 56 | 1244 | 9 | 1.0% | 19.2% | 117 | 1383 | 7 | 0.8% | 97.4% |
| 57 | 1245 | 3 | 0.3% | 19.5% | 118 | 1408 | 10 | 1.1% | 98.5% |
| 58 | 1246 | 5 | 0.6% | 20.0% | 119 | 1454 | 10 | 1.1% | 99.6% |
| 59 | 1247 | 5 | 0.6% | 20.6% | 120 | 1500 | 4 | 0.4% | 100.0% |
| 60 | 1248 | 4 | 0.4% | 21.0% |  |  |  |  |  |

Note: Blue = Exceeds, Green = Meets, Yellow = Approaches, and Orange = Falls Far Below the Standard; FREQ = frequency, CUML % = Cumulative percentage of students.

# Part 9: Reliability and Validity Evidence

Part 9 of the Technical Report provides evidence supporting the reliability and validity of the 2016 AIMS A Science assessments. All data presented in this section were computed using population test data available in the final electronic data files. The following standards from the 1999 *Standards* (AERA, APA, NCME) are addressed: 1.5, 1.7, 2.1, 2.4, 2.10, 2.13, 3.16, 4.15, 6.5, 7.1, 7.3, and 7.10 as well as standards 1.8, 1.9, 2.3, 2.7, 2.8, 2.19, 3.3, 3.6, 4.4, 5.19 and 7.4 from the 2014 *Standards* (AERA, APA, NCME).

## Reliability

The *Standards* refer to reliability as the “consistency of scores across replications of a testing procedure” (AERA, APA, NCME, 2014, p. 33). A reliable test produces stable scores; that is, very similar score distributions would result if the test were administered repeatedly under similar conditions to the same students without memory or fatigue affecting the scores. Reliability of the 2016 AIMS A assessments was estimated by internal consistency for all tests. It should be noted that due to the large number of non-responders in the sample and the low number of performance task test items the accuracy of the reliability coefficient may be problematic.

### Measures of Internal Consistency

Cronbach’s alpha is a frequently used to measure of internal consistency. Cronbach’s alpha is computed as (Crocker & Algina, 1986)

,

where *k* = number of items,  = the total score variance, and  = the variance of item *i*.

Reliability estimates for the 2016 AIMS A Science assessments are presented in Table 9.1.1. Note that a high degree of internal consistency is evident for all tests.

Table 9.1. 1   
2016 AIMS A Science Internal Consistency

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Alpha** | |
| **Test** | **N** | **MC** | **PT** |
| 4 | 1,033 | 0.86 | 0.95 |
| 8 | 994 | 0.85 | 0.96 |
| HS | 932 | 0.86 | 0.97 |

## Validity

“Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed users of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests” (AERA/APA/NCME, 1999, 2014, p. 11). The purpose of test score validation is not to validate the test itself, but to validate interpretations of the test scores for specific purposes or uses. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization of the assessment and continuing throughout the entire assessment process.

The 2016 AIMS A tests were designed and developed to provide fair and accurate ability scores that support appropriate, meaningful, and useful educational decisions. Evidence of this is also provided in Part 2 (Involvement of Arizona Educators), Part 3 (Test Design), Part 4 (Test Development), Part 5 (Test Administration), Part 6 (Data for Operational Analysis), Part 7 (Calibration, Scaling, and Scoring), Part 8 (Test Results), Part 9 (Validity Evidence), and Part 10 (Classification). As the Technical Report has progressed, chapter by chapter, it has moved through the phases of the testing cycle. Each part of the Technical Report detailed the procedures and processes applied in the creation of AIMS A, as well as their results. Each part also highlights the meaning and significance of the procedures, processes, and results in terms of content and construct validity and the relationship to the *Standards.* Part 9.2 addresses two final issues in validity: the issues of bias and construct validity. The analyses presented here add to the perspectives provided in Parts 2 through 10. The following is a brief review.

Part 2 of the Technical Report describes the involvement of Arizona educators and ADE in the test development process. As indicated in Part 2, the test development process, and the involvement of Arizona educators in that process, formed an important part of the validity of the entire AIMS A. The knowledge, expertise, and professional judgment offered by Arizona educators ultimately ensured that the content of AIMS A formed an adequate and representative sample of appropriate content and that the content formed a legitimate basis upon which to validly derive conclusions about student achievement.

Parts 3 and 4 of the Technical Report address the issue of test form development. These two parts provide a general discussion of test form creation and editing process, the process of selecting operational test items, the content distribution, and the blueprints. The test design process and the participation of Arizona educators in the process of test creation, including item content and bias review, provide a solid rationale for having confidence in the content and design of AIMS A Science as a tool from which to derive valid inferences about the academic performance of students with significant cognitive disabilities in Arizona.

Part 5 of the Technical Report describes the process, procedures, and policies that guided the administration of the AIMS A, including accommodations, security, and the written procedures provided to test administrators and school personnel.

Part 6 of the Technical Report describes classical data analysis of the spring 2016 AIMS A Science. The results presented in this section indicate that, from the classical perspective, the items used to calculate student scores generally function appropriately for the population the tests were designed to assess.

Part 7 of the Technical Report describes the calibration and equating methods, as well as processes and procedures for deriving scale scores from students’ raw scores and the data cleaning steps which ensure valid calibration and scaling. Some references to introductory and advanced discussions of IRT are provided.

Part 8 of the Technical Report describes information about the results of the 2016 spring administration of the AIMS A Science assessments. Importantly, this also describes the results for the many subgroups (e.g., ethnicity/race, primary disability classification, and social economic status, Free/Reduced Lunch). The analyses of these subgroup comparisons, provides evidence that generally, the test is not advantaging or disadvantaging any specific subgroup.

Part 9 of the Technical Report (above) describes Cronbach’s alpha as a measure for internal consistency. These results indicate that the AIMS A Science assessments produce student scores that are highly reliable.

Part 9 of the Technical Report (below) describes the correlations between student scores on the 2016 AIMS A Grade 4 and Grade 8 Science tests and other tests taken by the students in 2016. The results of this analyses, with correlations generally in the .70 range, are consistent with the expectations given the constructs measured though slightly lower than previous correlations among AIMS A Mathematics, Reading, and Science tests.

Part 10 of the Technical Report describes the cut score classifications as determined by the standard setting and the standard error of measurement at those cuts on the 2016 AIMS A Science assessments.

Additional evidence to support the validity of the 2016 AIMS A Science assessments is provided by previous AIMS A technical reports available at [www.azed.gov](http://www.azed.gov).

### Correlations of AIMS A Science to Other Assessments

Correlations were examined between scale scores on 2016 AIMS A tests and the federally mandated, Mathematics and ELA MSAA tests by grade level. Note that data used for the calculation of correlation included records of students who had a valid scale score on their AIMS Science test and at least one of the two MSAA content tests. Sample sizes for Grades 4 and 8 are therefore slightly lower than presented in other parts of this Technical Report. Since only Grade 11 students sit for the MSAA tests in high school where Grade 10 students are administered AIMS A Science, the lack of matching students precluded correlational analysis for High School AIMS A Science. Spearman rank correlation was used to measure the degree of association between the domains because, unlike the Pearson correlation which assumes normal distribution of both variables, the Spearman correlation test does not claim any assumptions about the distributions. The lack of assumptions is especially important with this population due to the number of non-responsive students.

The correlations for Grades 4 and 8 are presented in Tables 9.2.1.1 and 9.2.1.2, respectively. The patterns of correlation presented in the tables are generally in the .70 range and are consistent with expectations given the constructs measured. However, these values are slightly lower than the same analyses correlating 2014 AIMS A Science to the 2014 AIMS A Mathematics and Reading tests. One plausible reason for this attenuation could be the difference between the item formats in the two testing systems.

Table 9.2.1.1   
Correlations among AIMS A Science and MSAA ELA and Mathematics – Grade 4

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Math** | **Reading** | **Science** |
| **Math (MSAA)** | 1 | .715 | .669 |
| **Reading (MSAA)** | .715 | 1 | .806 |
| **Science (AIMS A)** | .669 | .806 | 1 |

N=992

Table 9.2.1.2   
Correlations among AIMS A Science and MSAA ELA and Mathematics – Grade 8

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Math** | **Reading** | **Science** |
| **Math (MSAA)** | 1 | .680 | .612 |
| **Reading (MSAA)** | .680 | 1 | .743 |
| **Science (AIMS A)** | .612 | .743 | 1 |

N=976

# Part 10: Classification

Part 10 of this Technical Report provides information regarding classifying students into proficiency categories. The following standards from the 1999 *Standards* (AERA, APA, NCME) are covered in this part: 1.5, 1.7, 2.14, 2.15, 4.9, 4.19, 4.20, 4.21, and 6.5 as well as standards 1.8, 1.9, 2.13, 2.14, 2.16, 5.5, 5.21, 5.22, 5.23, and 7.4 from the 2014 edition (AERA, APA, NCME).

Scores from the 2016 AIMS A assessments are used to classify students into one of four performance categories: Falls Far Below the Standard, Approaches the Standard, Meets the Standard, and Exceeds the Standard. This part of the Technical Report provides information regarding classifying students into these four performance categories.

## Standard Setting Technical Documentation

Standard setting for the AIMS A Science tests was conducted in early May 2009 using the Bookmark Standard Setting Procedure. All technical documentation regarding the standard setting is available in the 2009 AIMS A Technical Report.

Final scale score ranges for each of the four performance level categories for the AIMS A tests are presented in Table 10.1.1.

Table 10.1.1   
AIMS A Scale Score Ranges by Performance Level Set in 2009

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade** | **FFBS** | **AS** | **MS** | **ES** |
| 4 | 1000-1187 | 1188-1249 | 1250-1330 | 1331-1500 |
| 8 | 1000-1196 | 1197-1249 | 1250-1314 | 1315-1500 |
| High School | 1000-1196 | 1197-1249 | 1250-1308 | 1309-1500 |

Note: FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard.

## Standard Error of Measurement at Cut Scores

The standard error of measurement (SEM) at each of the score cuts is presented in Table 10.1.2. These SEM values, which are based on both the error at each theta scale and the scale score’s transformation constant (M1, described in Section 7.4), are lowest at the most critical cut (Meets the Standards) which determines proficiency on each assessment. The increase in error at the other two cuts is as expected within the Item Response Theory framework.

Table 10.1.2   
2016 AIMS A Science Standard Error of Measurement at Cut Scores

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade** | **AS** | |  | **MS** | |  | **ES** | |
| **Cut Score** | ***SEM*** |  | **Cut Score** | ***SEM*** |  | **Cut Score** | ***SEM*** |
| 4 | 1188 | 18 |  | 1250 | 12 |  | 1331 | 17 |
| 8 | 1197 | 15 |  | 1250 | 10 |  | 1315 | 14 |
| High School | 1197 | 13 |  | 1250 | 8 |  | 1309 | 15 |

Note: AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard.

# References

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

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

Arizona Department of Education (2009)*. Bookmark Standard Setting Technical Report for Grades 3, 5, 8, and High School Reading, Mathematics, and Science.* Nashville, Tennessee: Stephen Elliott.

Arizona Department of Education (2008). *Special Education Director’s Manual.* Phoenix, AZ.

Arizona Department of Education (2016). *Test Administration Directions.* Phoenix, AZ.

Brennan, R. L. & Prediger, D. J. (1981). Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41,* 687-699.

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

Choi, S. (2005). CalcSEM\_Rasch.sas [Computer program]. Unpublished.

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

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

Denham, A. (2004). *Pathways to learning for students with cognitive challenges: reading, writing and presenting.* Interdisciplinary Human Development Institute, University of Kentucky. Retrieved on June 2, 2009 from <http://www.ihdi.uky.edu/IEI/>

Elliott, S. N. & Braden, J. P. (2001). *Assessing One & All: Facilitating the Meaningful Participation of Students with Disabilities in District and Statewide Assessment Programs.* Reston, VA: Council for Exceptional Children.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mehwah, NJ: Erlbaum.

[Flowers, C. & Browder, D. (2004). *Ten questions that parents should ask about alternate assessments.* [Brochure]. Charlotte, NC: Evaluation of Emerging Alternate Assessment Practices Project.](http://www.naacpartners.org/resources/10_questions_for_Parents.pdf)

Green, D.R. (1975, December). *Procedures for assessing bias in achievement tests*. Presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Individuals with Disabilities Education Act (IDEA), 2004 PL 105-17, 20 U.S.C §§ 1400 et. seq.

Kentucky Statewide Alternate Assessment Project (1999). *Kentucky alternate portfolio teacher’s guide*. Lexington: University of Kentucky, Interdisciplinary Human Development Institute.

Kleinert, H. & Kearns Farmer, J. (2001). *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities.* Baltimore: Paul H. Brookes Publishing Co.

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

Lehr, C. & Thurlow, M. (2003). Putting it all together: Including students with disabilities in assessment and accountability systems (Policy Directions No.16). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved on June 2, 2009 from [NCEO.info](http://education.umn.edu/NCEO/OnlinePubs/Policy16.htm)

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16(2)*, 878.

Linacre, J. M. (2011). WINSTEPS Rasch measurement [Computer software]. Chicago: Winsteps.com.

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

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

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Microsoft Corporation. (2010). Excel 2010 [Computer software]. Bellevue, WA: Microsoft Corp.

Nelson, Larry Richard (2001). *Item analysis for tests and surveys using Lertap 5*. Perth, Western Australia: Curtin University of Technology.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Satterfield, B. & Satterfield, P. (2009). The Marriage of AT and IT. *The ConnSENSE Bulletin: Resources* *for Learning with Technology*. Retrieved on June 2, 2009 from http://Research, articles and programs about AA\Assistive Technology and accommodations research\The Marriage of AT and IT.mht.

Shrout, P. E. & Fleiss, J. L. (1979). *Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 86(2),* 420-428.

Wang, T. W., Kolen, M. J., & Harris, D. J. (2000). Psychometric properties of scale scores and performance levels for performance assessments using polytomous IRT. *Journal of Educational Measurement*, *37*, 141-162.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14(2)*, 97-116.

Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis: Rasch Measurement*. Chicago: MESA Press.

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement*, *21*, 93-111.

Yen, W. M. & Burket, G. R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement, 34(4)*, 293-313.

# APPENDIX A: AIMS A Eligibility Criteria

**Alternate Assessment**

**Participation Guidelines and Eligibility Determination**

The Arizona Department of Education offers an alternate assessment based on alternate achievement standards in compliance with the U.S. Department of Education federal regulations and guidance. A student must have an Individualized Education Program (IEP) to be considered for participation in an alternate assessment. All students must participate in state assessments.

Arizona Alternate Assessment:

ELA and Mathematics Grades 3 through 8 and 11 – *Multi-State Alternate Assessment (MSAA)*

Science Grades 4, 5, and 10 - *AIMS A Science*

IEP Teams will discuss and determine participation and eligibility using the *NCSC Participation Decision Documents* and/or the *AIMS A Science Eligibility Requirements.* Each of these tools should be considered separately as some criteria may be unique.

*Note: IEP teams should consider the testing cycles that will occur during the student’s IEP year.*

**STUDENT NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_STUDENT ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SAIS ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_DATE OF BIRTH\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_GRADE\_\_\_\_\_\_\_\_\_**

**SCHOOL:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_CASE MANAGER\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

| **Grade** | **None** | **MSAA ELA/Math** | **AIMS A Science** |
| --- | --- | --- | --- |
| K, 1st, 2nd | X |  |  |
| 3rd |  | X |  |
| 4th |  | X | X |
| 5th, 6th, 7th |  | X |  |
| 8th |  | X | X |
| 9th | X |  |  |
| 10th (or second year of high school) |  |  | X |
| 11th |  | X |  |
| 12th | X |  |  |

**NOTE: The IEP team must complete the parent notification of alternate assessment participation on page 5.**

**MSAA is not an administered assessment at the student’s grade level for this school year.**

**MSAA Participation Decision**

**Documentation**

**To meet the criteria for MSAA, the student must meet all participation criteria descriptors.**

| **Participation Criteria** | **Participation Criteria Descriptors** | **Sources of Evidence [check if used]** |
| --- | --- | --- |
| 1. The student has a significant cognitive disability  YES  NO | Review of student records indicate a disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior.  *\*Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life.* | Results of Individual Cognitive Ability Test  Results of Adaptive Behavior Skills Assessment  Results of individual and group administered achievement tests  Results of informal assessments  Results of individual reading assessments  Results of district-wide alternate assessments  Results of language assessments including English language learner (ELL) language assessments if applicable |
| 2.The student is learning content linked to (derived from) the Common Core State Standards (CCSS).  YES  NO | Goals and instruction listed in the IEP for this student are linked to the enrolled grade-level CCSS and address knowledge and skills that are appropriate and challenging for this student. | Examples of curriculum, instructional objectives and materials including work samples  Present levels of academic and functional performance, goals and objectives from the IEP  Data from scientific research-based interventions  Progress monitoring data |
| 3. The student requiresextensive direct individualized instruction and substantial supports to achieve measurable gains in the grade-and age-appropriate curriculum.  YES  NO | The student (a) requires extensive, repeated, individualized instruction and support that is not of a temporary or transient nature and (b) uses substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate and transfer skills across academic content. | Examples of curriculum, instructional objectives, and materials including work samples from both school and community based instruction  Teacher collected data and checklists  Present levels of academic and functional performance, goals, and objectives, and post school outcomes from the IEP and the Transition Plan for students age 12 and older |

**The student may participate in MSAA if all responses above are marked YES.**

MSAA Participation Decision Documentation

| *Additional Considerations Not to Use In Reviewing Evidence* |
| --- |
| 1. A disability category or label |
| 1. Poor attendance or extended absences |
| 1. Native language/social/cultural or economic difference |
| 1. Expected poor performance on the general education assessment |
| 1. Academic and other services received |
| 1. Educational environment or instructional setting |
| 1. Percent of time receiving special education services |
| 1. English Language Learner (ELL) status |
| 1. Low reading level/achievement level |
| 1. Anticipated disruptive behavior |
| 1. Impact of test scores on accountability system |
| 1. Administrator decision |
| 1. Anticipated emotional duress |
| 1. Need for accommodations, e.g., assistive technology/AAC to participate in assessment process |
| Evidence shows that the decision for participating in the NCSC Alternate Assessment was not based on the above list. |

IEP Team Statement of Assurance: *Our decision was based on multiple pieces of evidence that, when taken together, demonstrated that the Alternate Assessment is the most appropriate assessment for this students; that his/her academic instruction will be based on the CCCs linked to the CCSS; that the Additional Considerations listed above were not used to make this decision; and that any additional implications of this decision were discussed thoroughly.*

**AIMS A Science is not an administered assessment at the student’s grade level for this school year.**

Part I: AIMS A Science Eligibility Requirements

**Parent Notification**

**Alternate Assessment Participation**

Following IEP team review of participation guidelines, the student is eligible for and will participate in the following assessments:

**Potential Consequences:**

Are there any effects or local policies that would preclude completion requirements for a regular high school diploma for the child participating in testing?

No

Yes

If yes, explain: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Each of us agrees with the alternate assessment participation decisions indicated above.

Parent(s)/Guardian: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Position:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date:\_\_\_\_\_\_\_\_

**MSAA (ELA/Math) and/or AIMS A Science**

Yes

No (student will participate in statewide achievement test or AIMS

# APPENDIX B: Item Writer Selection Criteria

APP AIMS A Committee Participant Selection Criteria

**ARIZONA DEPARTMENT OF EDUCATION**

**PROCEDURE FOR SELECTION OF EDUCATOR COMMITTEES**

**ARIZONA ASSESSMENT SECTION**

Although our database contains over 1000 educators, the Assessment Section is always recruiting new teachers to serve on the committees, and have prevailed upon veteran teachers to become Ambassadors of the Assessment by encouraging their colleagues to apply.

Once Arizona educators are identified and entered into the database, the Assessment Section uses the following procedures for selecting membership for a committee:

* Identify the purpose/function of the committee
* Establish the date and time of the committee
* Determine the criteria for membership on the committee:
  + Content area of expertise
  + Grade level experience
  + Specific skill or knowledge expertise for committee function
  + Prior experience on ADE committees—a minimum 50% of each committee will have prior experience
  + Location of district/school
    - Rural/urban/suburban
    - Approximately 50% of committee members from Maricopa County when appropriate for purpose of committee
  + Ethnicity of school population or committee member
  + SES of school population
  + Number of committees served on recently—a committee member cannot serve on a series of committees used to develop items. Otherwise, they would be passing judgment on their own prior work. (This is a change in procedure)\*
* Review the database for educators that meet the criteria established
* Select committee members based on criteria for particular committee for primary and alternate list
* Invitations are sent to selected committee members on primary list \*\*
* After decline and accept emails are received by established deadline, additional invitations issued to members on alternate list
* Committee meeting held
* Review performance of participants.

\* ADE is concerned that utilizing the same committee members on a series of committees will reduce the input from a variety of educators and have requested that past committee participation be part of the selection process. As the pool of teachers expands, individual members will serve on fewer committees.

\*\* It is not the policy to inform all members in our database of scheduled committee meetings, but only those invited to a particular meeting.

Beginning in April of 2006, all past participants have been invited to update their applications on a yearly basis in order to have the most current information in the database. Also, when Arizona educators participate on a committee, they are asked to review their information and note anything that might have changed. The application identifies the demographics of each committee member: geographic location in Arizona, ethnicity of school/district population and/or committee participant, and a detailed biographical background including participation on AIMS A committees.

In order to replace past participants who have moved, changed positions, or no longer possess the time to serve, the Arizona Department of Education Assessment Division searches in the Committee Database to find individuals that have a desire to participate to serve as a member of the item writing, or content and bias review committee. Participants can at any time submit a committee member application form to the Assessment Division. The ADE is constantly recruiting Arizona educators to serve on the various AIMS A committees as well as encouraging retention of its veteran contributors and recognizing them as excellent Ambassadors of the Assessment.

# APPENDIX C: 2016 AIMS A Monitoring Review

The Individuals with Disabilities Education Act (IDEA) and Title I of the No Child Left Behind Act (NCLB) require the inclusion of all students with disabilities in the State assessment system. Title I further requires that the assessment results for all students be used for system accountability to ensure that the best education possible is provided to all students (Improving the Academic Achievement of the Disadvantaged, 2007).

The Arizona Department of Education (ADE) Assessment and Exceptional Student Services sections monitor the administration of Arizona’s Instrument to Measure Standards Alternate (AIMS A) during the spring testing window. Assessment monitoring is conducted to ensure test validity and reliability and also for continuity in subsequent assessment years. The Individuals with Disabilities Education Act (IDEA) (300.149) requires, and state law (ARS 15-755) authorizes, monitoring and evaluation activities to determine the effectiveness of programs for meeting the educational needs of children with disabilities. These practices help to ensure that programs are carried out and educational results for children with disabilities improve.

Monitoring was conducted by external consultants as the performance tests were administered in person throughout the testing window from February 15 to March 31, 2016. The onsite testing monitors evaluated the environment in which the student was being assessed, as well as the administration of the performance tasks of the assessment. In addition to the AIMS A external consultants observing the administration of the alternate assessment, the external consultants participated in an inter-rater reliability study that more closely examined the performance task scoring rubric as a valid measurement tool for the AIMS A. Data was collected through a random sample of observations. The consultants were trained and reviewed training videos on how to use the performance task scoring rubric. The consultant’s rating was then compared to the test administrator’s rating. The overall inter-rater reliability percentage was 93.6%.

The external consultants evaluated information about the assessment administration, standardized activities, and data collection procedures. Teachers were selected for monitoring based on the students for whom they administered the AIMS A. Schools were randomly selected to be representative of the total population that took AIMS A in 2016. The sampling was done based on special education need, ethnicity, gender, and region. A total of 10 students were selected.

From the committee’s suggestions, the following will be instituted for the AIMS A 2017 administration.

* Each district is required to designate an alternate assessment test coordinator that will participate in the mandatory online training and is responsible to train all staff in their district on the proper administration and scoring of the performance tasks. Including training to address clarification of prompting, modeling, and cueing, based on recommendations from the Alternate Assessment External Consultants. Video demonstrations of the use of the performance tasks scoring rubric can be accessed on the Arizona Department of Education AIMS A web page at [AIMS A Science](http://www.azed.gov/assessment/aims-a/)  under the Videos and Webinars tab.
* The Performance Task will be clarified to include those definitions on prompting, modeling, and cueing provided by the National Alternate Assessment Center.
* Clarification of allowable accommodations will be included in the Test Coordinator Manual and Test Administration Directions.
* Practice tests will be available in the AIMS A Science system.

# APPENDIX D: Example Item Specification Card

