Arizona

Arizona's Instrument to Measure Standards Alternate

AIMS A

2015

Technical Report

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

Figures and Tables

Part 1: Executive Summary

This document provides information regarding processes and procedures implemented in the 2015 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 2015 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). AIMS A 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 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 responds 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, B. and Satterfield, P., 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 to4-point rubric. The Performance Tasks are standardized, constructed response items which are scored on standardized data sheets based on that 0 to 4-point 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 in previous years in preparation for the 2015 AIMS A Science assessments. These 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 test development processes and AIMS A results interpretation. In addition to these external committees, internal teams, consisting of various Arizona Department of Education specialists and administrators, were called upon to conduct reviews to support quality assurance. The test development committee and internal team meetings included:

- Item Review in which the internal team reviewed each item administered in 2014. The members made notes on the items including clarity of content, overall appearance, size of font and graphics, punctuation, and grammar.
- Item Gap Analysis in which the internal team reviewed the current academic standards. The internal team reviewed the item bank. From this analysis a gap was identified and a plan developed for the Item Writing committees. The plan identified which standards and concepts needed items to be developed and field tested subsequent administrations;
- Item Writing in which educators wrote Multiple-Choice items, and Performance Tasks aligned to the alternate content standards for possible use in the spring of 2015 as field test items.
- Content and Bias Review in which educators reviewed Multiple-Choice items, and Performance Tasks for content, bias, and sensitivity. Items that passed these reviews were eligible for inclusion on the 2015 AIMS A assessment;
- External Consultant Final Document Review in which external consultants (special education and general education teachers, school psychologists, and special education directors) were hired to review all final test documents that were assembled and placed on the ADE development site prior to the administration of AIMS A. The external consultants attended a face to face meeting with the Alternate Assessment unit to review all Multiple-Choice and performance items in a display similar to what the students would see when presented the items. Team notes were made to reflect changes that needed to be implanted (i.e., spelling errors or items not fitting on the page correctly);
- ADE Internal Team in which the internal team (AIMS A coordinator, specialist, project specialist, director, and deputy associate superintendent) reviewed the documents returned by the external consultants. Decisions were made based on the feedback to make edits and revisions. A final internal review of every item was conducted prior to the test administration.

Part 3: Test Design

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

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**

3.2 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%

3.3 Description of AIMS A 2015 Science Tests

The test blueprints were used with the processes described in Part 4 to develop all AIMS A Science tests administered in 2015. 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 2015 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
2015 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 2015 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).

4.1 AIMS A Test Development and Editing Process

4.1.1 Blueprint Development

The development of the 2015 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.

4.1.2 Item Writing and Editing

The development of the 2015 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. The new items were then constructed in response to an internal review of the test map and a thorough gap analysis. After the item writing workshops, the new 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.

4.1.3 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.

4.1.4 Test Construction Process

Test construction for the 2015 test administration began with an internal review of the item statistics for the items used in the 2014 administration to identify, for replacement, items that were performing less than optimally. A maximum of 30 operational items were chosen to be administered for 2015. 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.

4.1.5 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 conducted 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).

5.1 Adaptations

5.1.1 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 2015 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 2015 AIMS A Science assessments; however, this is not an exhaustive list of adaptations that could be utilized.

Table 5.1.1
2015 AIMS A Science Adaptations Provided

	Number of Students Using Adaptation			
Adaptation	Grade 4	Grade 8	High School	Total
Adaptive calculators	82	160	128	370
Alphabet line	303	213	128	644
Graph paper	69	107	40	216
Highlight or mark key phrases, words, or letters	307	309	259	875
Line drawings	127	136	68	331
Magnifier	46	22	22	90
Manipulatives	617	557	437	1611
None	60	72	115	247
Number line	452	390	239	1081
Other	167	179	143	489
Picture/Object system	321	303	229	853
Read passages or any test item/describe graphics	776	772	593	2141
Sign language	103	77	44	224
Switch	68	72	43	183
Symbolic/Picture system	298	263	200	761
Use of objects	406	351	259	1016
Total Used	**4202**	**3983**	**2947**	**11132**

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

5.2 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 2015 AIMS A Test Security Agreement

Arizona's Instrument to Measure Standards
AIMS A Test Security / Testing Ethics Agreement 2015

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.
    a. Limit access to persons with a responsible, professional interest in the test's security.
    b. Names of all persons having access to the materials will be kept on file by the special education director.
    c. 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.
        i. Building administrators will maintain signed agreements of building staff.
        ii. Special Education Directors will maintain signed agreements of building administrators.
2. I will keep all test materials secure, limiting access to Test Administrators.
    a. Test materials will be kept secure until they are actually distributed to students.
    b. 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.
    a. No content of the test will be disclosed or allowed to be disclosed.
    b. 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 2015 must also:

- participate in training activities prior to administering the AIMS A;
- review *AIMS A Test Administration Directions* for 2015 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 2015 AIMS A test will also sign a Test Security Agreement.

Signed By: _____

Printed Name: _____

Title: _____

School: _____

5.3 Test Administration

To ensure standardized testing administration for all students, the *AIMS A Test Coordinator Manual* was made available to all special education directors for the spring 2015 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 Test Administration Directions* was made available to all test administrators for the spring 2015 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 TAs 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 2015 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.

6.1 Data

AIMS A has one test window spanning six weeks. The 2015 assessments were administered between February 15 and March 31, 2015. 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.

6.2 Descriptive Statistics by Test

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
2015 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	1,016	60	36.66	14.48	0.86	60	40.67	15.30	0.95
8	999	60	38.07	15.93	0.85	60	40.82	14.88	0.96
HS	891	60	41.67	15.74	0.86	60	41.30	15.95	0.97

Table 6.2.2 presents the standard Lertap analysis statistics of the raw scores for 2015 AIMS A assessment for each grade tested.

Table 6.2.2
2015 AIMS A Science Raw Score Test Analysis

	Grade		
	4	8	10
Number Tested	1,016	999	891
Minimum RS	0	0	0
Median	83.0	85.0	92.0
Mean	77.3	78.9	83.0
Maximum	120	120	120
Std. Deviation	27.5	28.3	29.4
Variance	757.1	800.5	866.5
Range	120	120	120
Interquartile Range	38	40	41
Skewness	-0.863	-0.912	-1.017
Kurtosis	0.294	0.224	0.327
Min. Possible	0.00	0.00	0.00
Max. Possible	120.00	120.00	120.00
# No Response	20	21	23
% No Response	2.0%	2.1%	2.6%

6.3 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
2015 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	989	0.70	70%	0.32	0.42	17%	-0.11	10%	-0.30	3%
2	983	0.66	66%	0.56	0.72	18%	-0.36	13%	-0.26	3%
3	983	0.68	68%	0.46	0.60	10%	-0.24	19%	-0.28	3%
4	982	0.71	71%	0.49	0.65	12%	-0.24	14%	-0.30	3%
5	981	0.62	62%	0.43	0.55	19%	-0.19	15%	-0.28	3%
6	984	0.72	72%	0.48	0.65	10%	-0.28	15%	-0.27	3%
7	985	0.75	75%	0.53	0.72	9%	-0.29	13%	-0.32	3%
8	982	0.46	46%	0.28	0.35	27%	-0.13	23%	-0.15	3%
9	983	0.65	65%	0.50	0.64	11%	-0.28	21%	-0.29	3%
10	983	0.67	67%	0.47	0.62	12%	-0.22	17%	-0.31	3%
11	982	0.44	44%	0.17	0.21	13%	-0.20	40%	0.00	3%
12	984	0.53	53%	0.45	0.56	21%	-0.31	23%	-0.17	3%
13	979	0.29	29%	0.14	0.19	28%	-0.07	39%	0.00	4%
14	982	0.61	61%	0.50	0.63	18%	-0.30	18%	-0.25	3%
15	980	0.66	66%	0.35	0.46	14%	-0.16	16%	-0.22	4%

Performance Tasks

		Score 0		Score 1		Score 2		Score 3		Score 4	
Item	***N***	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}
16	1,016	9%	-0.62	12%	-0.27	14%	-0.07	22%	0.06	44%	0.54
17	1,016	10%	-0.65	11%	-0.30	10%	-0.08	18%	0.05	52%	0.59
18	1,016	10%	-0.60	18%	-0.23	21%	0.03	25%	0.24	25%	0.35
19	1,016	9%	-0.62	13%	-0.31	14%	-0.10	19%	0.09	45%	0.56
20	1,016	10%	-0.62	16%	-0.27	19%	0.01	26%	0.24	29%	0.39
21	1,016	8%	-0.68	9%	-0.32	10%	-0.11	17%	0.01	56%	0.62
22	1,016	8%	-0.62	12%	-0.32	14%	-0.12	23%	0.11	44%	0.54
23	1,016	9%	-0.65	12%	-0.31	13%	-0.07	22%	0.13	44%	0.51
24	1,016	7%	-0.64	8%	-0.32	8%	-0.15	14%	-0.05	63%	0.65
25	1,016	10%	-0.67	9%	-0.27	11%	-0.09	20%	0.04	50%	0.58
26	1,016	9%	-0.67	11%	-0.31	12%	-0.09	21%	0.11	47%	0.55
27	1,016	10%	-0.63	14%	-0.30	10%	-0.10	14%	0.06	51%	0.62
28	1,016	12%	-0.65	20%	-0.26	19%	0.05	20%	0.21	29%	0.46
29	1,016	12%	-0.63	21%	-0.25	20%	0.04	16%	0.19	31%	0.47
30	1,016	13%	-0.58	22%	-0.24	20%	0.07	22%	0.24	24%	0.39

Table 6.3.2
2015 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	959	0.77	77%	0.56	0.77	11%	-0.27	8%	-0.32	4%
2	960	0.74	74%	0.52	0.70	14%	-0.30	8%	-0.26	4%
3	964	0.58	58%	0.46	0.58	14%	-0.17	25%	-0.30	4%
4	960	0.64	64%	0.44	0.56	18%	-0.21	15%	-0.24	4%
5	962	0.58	58%	0.37	0.47	22%	-0.11	16%	-0.27	4%
6	956	0.62	62%	0.52	0.66	14%	-0.30	20%	-0.24	4%
7	960	0.64	64%	0.58	0.74	15%	-0.28	17%	-0.34	4%
8	958	0.50	50%	0.36	0.45	18%	-0.24	27%	-0.11	4%
9	957	0.77	77%	0.50	0.69	8%	-0.19	11%	-0.31	4%
10	961	0.68	68%	0.60	0.78	13%	-0.32	14%	-0.33	4%
11	958	0.40	40%	0.29	0.37	19%	-0.19	36%	-0.06	4%
12	956	0.74	74%	0.47	0.64	11%	-0.16	11%	-0.31	4%
13	958	0.72	72%	0.54	0.72	8%	-0.18	15%	-0.37	4%
14	958	0.62	62%	0.53	0.68	19%	-0.21	15%	-0.35	4%
15	959	0.51	51%	0.47	0.59	31%	-0.23	14%	-0.25	4%

Performance Tasks

		Score 0		Score 1		Score 2		Score 3		Score 4	
Item	***N***	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}
16	999	9%	-0.62	12%	-0.27	14%	-0.07	22%	0.06	44%	0.54
17	999	10%	-0.65	11%	-0.30	10%	-0.08	18%	0.05	52%	0.59
18	999	10%	-0.60	18%	-0.23	21%	0.03	25%	0.24	25%	0.35
19	999	9%	-0.62	13%	-0.31	14%	-0.10	19%	0.09	45%	0.56
20	999	10%	-0.62	16%	-0.27	19%	0.01	26%	0.24	29%	0.39
21	999	8%	-0.68	9%	-0.32	10%	-0.11	17%	0.01	56%	0.62
22	999	8%	-0.62	12%	-0.32	14%	-0.12	23%	0.11	44%	0.54
23	999	9%	-0.65	12%	-0.31	13%	-0.07	22%	0.13	44%	0.51
24	999	7%	-0.64	8%	-0.32	8%	-0.15	14%	-0.05	63%	0.65
25	999	10%	-0.67	9%	-0.27	11%	-0.09	20%	0.04	50%	0.58
26	999	9%	-0.67	11%	-0.31	12%	-0.09	21%	0.11	47%	0.55
27	999	10%	-0.63	14%	-0.30	10%	-0.10	14%	0.06	51%	0.62
28	999	12%	-0.65	20%	-0.26	19%	0.05	20%	0.21	29%	0.46
29	999	12%	-0.63	21%	-0.25	20%	0.04	16%	0.19	31%	0.47
30	999	13%	-0.58	22%	-0.24	20%	0.07	22%	0.24	24%	0.39

Table 6.3.3
2015 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	870	0.85	85%	0.56	0.86	4%	-0.24	9%	-0.41	2%
2	869	0.67	67%	0.43	0.56	11%	-0.26	19%	-0.27	2%
3	866	0.54	54%	0.37	-0.20	22%	-0.26	21%	-0.14	3%
4	867	0.68	68%	0.62	0.80	14%	-0.38	16%	-0.33	3%
5	867	0.74	74%	0.60	0.82	11%	-0.30	12%	-0.40	3%
6	866	0.66	66%	0.54	0.70	18%	-0.28	13%	-0.34	3%
7	867	0.51	51%	0.30	0.38	25%	-0.13	21%	-0.19	3%
8	865	0.77	77%	0.53	0.73	10%	-0.28	10%	-0.31	3%
9	864	0.74	74%	0.65	0.88	11%	-0.42	12%	-0.33	3%
10	869	0.73	73%	0.51	0.69	12%	-0.33	12%	-0.28	2%
11	865	0.72	72%	0.45	0.59	11%	-0.15	14%	-0.35	3%
12	866	0.68	68%	0.62	0.81	12%	-0.37	16%	-0.35	3%
13	865	0.77	77%	0.55	0.77	10%	-0.22	11%	-0.41	3%
14	867	0.66	66%	0.33	0.43	13%	-0.21	18%	-0.18	3%
15	868	0.71	71%	0.45	0.59	10%	-0.14	17%	-0.55	3%

Performance Tasks

		Score 0		Score 1		Score 2		Score 3		Score 4	
Item	***N***	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}	**%**	r_{pb}
16	891	9%	-0.62	12%	-0.27	14%	-0.07	22%	0.06	44%	0.54
17	891	10%	-0.65	11%	-0.30	10%	-0.08	18%	0.05	52%	0.59
18	891	10%	-0.60	18%	-0.23	21%	0.03	25%	0.24	25%	0.35
19	891	9%	-0.62	13%	-0.31	14%	-0.10	19%	0.09	45%	0.56
20	891	10%	-0.62	16%	-0.27	19%	0.01	26%	0.24	29%	0.39
21	891	8%	-0.68	9%	-0.32	10%	-0.11	17%	0.01	56%	0.62
22	891	8%	-0.62	12%	-0.32	14%	-0.12	23%	0.11	44%	0.54
23	891	9%	-0.65	12%	-0.31	13%	-0.07	22%	0.13	44%	0.51
24	891	7%	-0.64	8%	-0.32	8%	-0.15	14%	-0.05	63%	0.65
25	891	10%	-0.67	9%	-0.27	11%	-0.09	20%	0.04	50%	0.58
26	891	9%	-0.67	11%	-0.31	12%	-0.09	21%	0.11	47%	0.55
27	891	10%	-0.63	14%	-0.30	10%	-0.10	14%	0.06	51%	0.62
28	891	12%	-0.65	20%	-0.26	19%	0.05	20%	0.21	29%	0.46
29	891	12%	-0.63	21%	-0.25	20%	0.04	16%	0.19	31%	0.47
30	891	13%	-0.58	22%	-0.24	20%	0.07	22%	0.24	24%	0.39

Part 7: Calibration, Equating, and Scaling

Part 7 of the Technical Report describes the scaling procedures and results for the 2015 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).

7.1 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.

7.1.1 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

$$P_i(\theta) = \frac{\exp[(\theta - b_i)]}{1 + \exp[(\theta - b_i)]},$$

where b_i 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

$$P_{ix}(\theta_j) = \frac{\exp \sum_{l=0}^{x} (\theta_j - D_{il})}{\sum_{k=0}^{m_i} [\exp \sum_{l=0}^{k} (\theta_j - D_{il})]}.$$

Here, x = 0, 1, …, m_i and D_{il} is a step difficulty for score l and is defined as

$$\sum_{i=0}^{0} (\theta_j - D_{il}) \equiv 0,$$

and can be decomposed as

$$D_{il} = b_i + h_{il},$$

where b_i is the overall difficulty for item i and h_{il} is the threshold for score point l (Embretson & Reise, 2000).

7.1.2 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).

7.2 Calibration Results

7.2.1 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, nine 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		**2.62**		
2	Grade 4	3		**1.64**		
3	Grade 4	5		**1.43**		
4	Grade 4	6		**4.88**		
5	Grade 4	8	**1.44**	**9.90**	**0.27**	
6	Grade 4	9		**1.41**		
7	Grade 4	10		**3.65**		
8	Grade 4	11	**1.67**	**2.84**	**0.15**	
9	Grade 4	12		**1.55**		
10	Grade 4	13	**1.59**	**9.90**	**0.12**	**.30**
11	Grade 4	14		**9.26**		
12	Grade 4	15		**4.34**		
13	Grade 8	1		**4.91**		
14	Grade 8	2		**4.30**		
15	Grade 8	3		**1.57**		
16	Grade 8	4		**9.90**		
17	Grade 8	5	**1.51**	**2.06**	**0.29**	
18	Grade 8	8	**1.46**	**2.14**		
19	Grade 8	11	**1.57**	**3.11**	**0.23**	
20	HS	2		**4.80**		
21	HS	3	**1.58**	**1.99**		
22	HS	5		**1.48**		
23	HS	6		**1.85**		
24	HS	7	**1.77**	**2.68**	**0.25**	
25	HS	8		**1.92**		
26	HS	10		**1.54**		
27	HS	11		**1.87**		
28	HS	13		**1.73**		
29	HS	14	**1.63**	**2.49**		
30	HS	15		**4.00**		

Table 7.2.1.2
2015 AIMS A Science IRT Item Statistics - Grade 4

Item	Rasch Measure	SE	INFIT	OUTFIT	PT. BIS	*p*-value
1	0.1424	0.0214	1.36	**2.62**	0.32	0.71
2	0.2144	0.0207	0.98	1.34	0.54	0.67
3	0.1714	0.0211	1.12	**1.64**	0.47	0.69
4	0.1295	0.0216	1.14	1.04	0.47	0.72
5	0.2785	0.0202	1.23	**1.43**	0.40	0.63
6	0.1068	0.0219	1.05	**4.88**	0.50	0.73
7	0.0146	0.0233	1.13	1.11	0.51	0.76
8	0.5278	0.0196	**1.44**	**9.90**	**0.27**	0.47
9	0.2320	0.0205	1.13	**1.41**	0.46	0.66
10	0.1890	0.0209	1.12	**3.65**	0.47	0.68
11	0.5724	0.0197	**1.67**	**2.84**	**0.15**	0.44
12	0.4267	0.0196	1.16	**1.55**	0.43	0.54
13	0.8100	0.0211	**1.59**	**9.90**	**0.12**	**0.30**
14	0.3053	0.0200	1.07	**9.26**	0.48	0.62
15	0.2089	0.0207	1.33	**4.34**	0.35	0.67
16	-0.1403	0.0280	0.78	0.76	0.65	0.70
17	-0.2471	0.0292	0.85	0.76	0.70	0.74
18	0.1496	0.0260	0.77	0.81	0.52	0.60
19	-0.1463	0.0280	0.77	0.74	0.69	0.70
20	0.0814	0.0263	0.71	0.71	0.59	0.63
21	-0.3493	0.0305	0.80	0.68	0.73	0.76
22	-0.1603	0.0282	0.68	0.66	0.70	0.71
23	-0.1545	0.0281	0.80	0.74	0.63	0.71
24	-0.4773	0.0325	0.90	0.72	0.71	0.80
25	-0.2437	0.0291	0.81	0.73	0.70	0.74
26	-0.2255	0.0289	0.78	0.70	0.69	0.73
27	-0.1688	0.0283	0.91	0.83	0.67	0.71
28	0.1685	0.0259	0.73	0.70	0.63	0.60
29	0.1839	0.0258	0.77	0.76	0.62	0.59
30	0.2706	0.0256	0.77	0.77	0.56	0.56

Table 7.2.1.3
2015 AIMS A Science IRT Item Statistics - Grade 8

Item	Rasch Measure	SE	INFIT	OUTFIT	PT. BIS	*p*-value
1	0.0884	0.0242	1.07	**4.91**	0.50	0.79
2	0.0464	0.0249	1.28	**4.30**	0.51	0.76
3	0.4539	0.0205	1.17	**1.57**	0.44	0.59
4	0.3502	0.0211	1.30	**9.90**	0.39	0.65
5	0.4115	0.0207	**1.51**	**2.06**	**0.29**	0.59
6	0.3820	0.0209	1.21	1.34	0.44	0.63
7	0.3502	0.0211	1.04	1.07	0.53	0.65
8	0.5777	0.0202	**1.46**	**2.14**	0.30	0.51
9	0.0911	0.0242	1.10	0.86	0.50	0.78
10	0.2669	0.0219	0.99	1.00	0.55	0.70
11	0.7399	0.0204	**1.57**	**3.11**	**0.23**	0.41
12	0.0784	0.0244	1.29	1.17	0.46	0.76
13	0.1814	0.0228	1.11	1.28	0.50	0.74
14	0.3837	0.0209	1.12	1.19	0.49	0.63
15	0.5617	0.0202	1.19	1.40	0.43	0.53
16	-0.2518	0.0311	0.74	0.66	0.68	0.79
17	-0.2210	0.0307	0.81	0.74	0.65	0.76
18	0.0870	0.0278	0.77	0.75	0.59	0.66
19	0.0059	0.0284	0.83	0.83	0.61	0.69
20	-0.2399	0.0309	0.78	0.68	0.67	0.76
21	-0.4969	0.0349	0.95	0.74	0.71	0.82
22	-0.2183	0.0307	0.80	0.71	0.68	0.78
23	0.3669	0.0266	0.85	0.84	0.47	0.56
24	-0.2352	0.0309	0.70	0.62	0.73	0.78
25	-0.1948	0.0304	0.88	0.79	0.70	0.76
26	0.2864	0.0268	0.79	0.80	0.46	0.59
27	0.4851	0.0265	0.90	0.93	0.37	0.52
28	-0.0095	0.0285	0.82	0.79	0.64	0.69
29	0.0575	0.0280	0.74	0.71	0.64	0.67
30	0.2190	0.0271	0.86	0.84	0.53	0.62

Table 7.2.1.4
2015 AIMS A Science IRT Item Statistics - High School

Item	Rasch Measure	SE	INFIT	OUTFIT	PT. BIS	*p*-value
1	-0.3573	0.0333	1.28	0.94	0.53	0.86
2	0.1450	0.0235	1.37	**4.80**	0.42	0.68
3	0.3989	0.0222	**1.58**	**1.99**	0.33	0.54
4	0.1130	0.0238	0.98	0.83	0.61	0.68
5	0.0011	0.0252	1.09	**1.48**	0.54	0.75
6	0.1791	0.0232	1.16	**1.85**	0.51	0.67
7	0.4423	0.0222	**1.77**	**2.68**	**0.25**	0.52
8	-0.0933	0.0267	1.21	**1.92**	0.51	0.78
9	0.0187	0.0249	0.93	0.64	0.62	0.75
10	0.0286	0.0248	1.21	**1.54**	0.49	0.74
11	0.0553	0.0245	1.38	**1.87**	0.41	0.73
12	0.1271	0.0237	0.94	0.89	0.62	0.69
13	-0.0680	0.0263	1.16	**1.73**	0.53	0.78
14	0.1778	0.0233	**1.63**	**2.49**	0.30	0.67
15	0.0790	0.0242	1.39	**4.00**	0.40	0.72
16	-0.4046	0.0336	0.78	0.63	0.76	0.80
17	-0.2309	0.0312	0.71	0.64	0.72	0.76
18	-0.2322	0.0312	0.79	0.69	0.72	0.74
19	-0.0086	0.0291	0.76	0.71	0.64	0.66
20	-0.2819	0.0318	0.77	0.68	0.74	0.75
21	0.0481	0.0287	0.79	0.82	0.56	0.65
22	0.0133	0.0289	0.78	0.74	0.65	0.66
23	0.0962	0.0284	0.88	0.97	0.56	0.65
24	-0.2081	0.0309	0.82	0.71	0.71	0.73
25	-0.3400	0.0326	0.86	0.73	0.73	0.77
26	-0.2293	0.0312	0.77	0.69	0.75	0.73
27	0.2364	0.0278	0.72	0.72	0.57	0.58
28	-0.3262	0.0324	0.83	0.70	0.75	0.76
29	0.1057	0.0283	0.84	0.81	0.58	0.63
30	0.1192	0.0283	0.84	0.80	0.60	0.61

7.3 Equating

The 2015 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 2014 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 7.3.1 presents the number of anchor items for each grade and subject area. Table 7.3.2 shows the content representation for the 2015 anchor items compared to the 2015 operational form for Science. Table 7.3.3 presents descriptive statistics for the 2015 anchor item difficulties and the 2015 operational form.

Table 7.3.1
Spring 2015 AIMS A Anchor Items

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

Table 7.3.2
Content Representation of 2015 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	23%

Table 7.3.3
Rasch Difficulty Representation of 2015 Anchor Sets

Grade	Statistic	Test	Anchor
4	M_b	0.096	0.009
	SD_b	0.283	0.242
	MIN_b	-0.477	-0.477
	MAX_b	0.810	0.305
8	M_b	0.153	0.055
	SD_b	0.294	0.273
	MIN_b	-0.497	-0.252
	MAX_b	0.740	0.562
HS	M_b	-0.013	-0.041
	SD_b	0.214	0.216
	MIN_b	-0.405	-0.405
	MAX_b	0.442	0.236

Note: M_b = Mean Rasch difficulty, SD_b = Standard Deviation of the Rasch difficulty, MIN_b = Minimum Rasch difficulty, MAX_b = Maximum Rasch difficulty.

A fixed-parameter equating process was used within Winsteps to link the 2015 AIMS A assessments to their operational scale. This was implemented by constraining the 2015 item parameter estimate of the anchor items to be equal to the final estimates obtained in the 2014 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 -0.50 is considered significant and

cause for an anchor to be removed from the anchor set. Arizona uses the more conservative criterion of 0.30 and -0.30 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 2015 AIMS A calibration, no anchor items displayed a displacement statistic greater than 0.30 or less than -0.30.

7.4 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)

7.4.1 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	MI	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
2015 AIMS A Science Raw Score to Scale Score - Grade 4

Raw Score	Scale Score	*SEM*	Raw Score	Scale Score	*SEM*
0	1000	447	61	1259	11
1	1000	100	62	1260	11
2	1000	71	63	1261	11
3	1018	58	64	1263	11
4	1046	49	65	1264	11
5	1067	44	66	1265	11
6	1085	39	67	1266	11
7	1099	36	68	1268	11
8	1111	33	69	1269	11
9	1121	31	70	1270	11
10	1130	29	71	1272	11
11	1138	27	72	1273	11
12	1145	26	73	1274	12
13	1152	25	74	1276	12
14	1157	23	75	1277	12
15	1163	22	76	1278	12
16	1167	22	77	1280	12
17	1172	21	78	1281	12
18	1176	20	79	1282	12
19	1180	19	80	1284	12
20	1184	19	81	1285	12
21	1187	18	82	1287	12
22	1190	18	83	1288	12
23	1193	17	84	1290	12
24	1196	17	85	1291	12
25	1199	16	86	1293	13
26	1201	16	87	1294	13
27	1204	16	88	1296	13
28	1206	15	89	1298	13
29	1209	15	90	1299	13
30	1211	15	91	1301	13
31	1213	14	92	1303	13
32	1215	14	93	1305	14
33	1217	14	94	1306	14
34	1219	14	95	1308	14
35	1221	14	96	1310	14
36	1223	13	97	1312	15
37	1224	13	98	1315	15
38	1226	13	99	1317	15
39	1228	13	100	1319	15
40	1229	13	101	1322	16
41	1231	13	102	1324	16
42	1232	12	103	1327	17
43	1234	12	104	1330	17
44	1236	12	105	1333	18
45	1237	12	106	1336	18
46	1238	12	107	1340	19
47	1240	12	108	1344	20
48	1241	12	109	1348	21
49	1243	12	110	1353	22
50	1244	12	111	1358	24
51	1246	12	112	1364	25
52	1247	12	113	1371	28
53	1248	12	114	1379	31
54	1250	12	115	1390	34
55	1251	11	116	1404	40
56	1252	11	117	1423	48
57	1254	11	118	1453	63
58	1255	11	119	1500	94
59	1256	11	120	1500	446
60	1257	11			

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

Table 7.4.3
2015 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	1013	59	63	1258	10
3	1047	48	64	1259	10
4	1071	42	65	1260	10
5	1090	37	66	1261	10
6	1105	34	67	1262	10
7	1117	31	68	1263	10
8	1127	28	69	1264	10
9	1136	26	70	1266	10
10	1144	25	71	1267	10
11	1151	23	72	1268	10
12	1158	22	73	1269	10
13	1163	21	74	1270	10
14	1168	20	75	1271	10
15	1173	19	76	1272	10
16	1177	18	77	1274	10
17	1181	18	78	1275	10
18	1184	17	79	1276	10
19	1188	16	80	1277	10
20	1191	16	81	1278	10
21	1194	15	82	1280	10
22	1197	15	83	1281	10
23	1199	15	84	1282	10
24	1202	14	85	1283	10
25	1204	14	86	1285	10
26	1206	14	87	1286	11
27	1209	13	88	1287	11
28	1211	13	89	1289	11
29	1213	13	90	1290	11
30	1214	12	91	1292	11
31	1216	12	92	1293	11
32	1218	12	93	1295	11
33	1220	12	94	1296	12
34	1221	12	95	1298	12
35	1223	11	96	1299	12
36	1225	11	97	1301	12
37	1226	11	98	1303	12
38	1228	11	99	1305	13
39	1229	11	100	1307	13
40	1230	11	101	1309	13
41	1232	11	102	1311	14
42	1233	11	103	1314	14
43	1234	10	104	1316	15
44	1236	10	105	1319	15
45	1237	10	106	1322	16
46	1238	10	107	1325	16
47	1240	10	108	1328	17
48	1241	10	109	1332	18
49	1242	10	110	1336	19
50	1243	10	111	1341	21
51	1244	10	112	1346	22
52	1246	10	113	1353	24
53	1247	10	114	1361	27
54	1248	10	115	1370	30
55	1249	10	116	1383	35
56	1250	10	117	1401	42
57	1251	10	118	1428	54
58	1252	10	119	1479	80
59	1253	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
2015 AIMS A Science Raw Score to Scale Score - High School

Raw Score	Scale Score	*SEM*	Raw Score	Scale Score	*SEM*
0	1000	339	61	1248	8
1	1009	75	62	1249	8
2	1059	52	63	1250	8
3	1088	42	64	1251	8
4	1107	35	65	1252	8
5	1122	31	66	1253	8
6	1133	28	67	1254	8
7	1142	25	68	1255	8
8	1150	23	69	1256	8
9	1157	22	70	1257	8
10	1163	20	71	1258	9
11	1168	19	72	1259	9
12	1172	18	73	1260	9
13	1176	17	74	1261	9
14	1180	16	75	1262	9
15	1183	16	76	1263	9
16	1186	15	77	1264	9
17	1189	14	78	1265	9
18	1192	14	79	1266	9
19	1194	13	80	1267	9
20	1196	13	81	1268	9
21	1199	13	82	1269	9
22	1201	12	83	1270	9
23	1203	12	84	1271	9
24	1205	12	85	1272	9
25	1206	12	86	1273	9
26	1208	11	87	1274	9
27	1210	11	88	1276	10
28	1211	11	89	1277	10
29	1213	11	90	1278	10
30	1214	11	91	1279	10
31	1216	10	92	1281	10
32	1217	10	93	1282	10
33	1219	10	94	1283	10
34	1220	10	95	1285	11
35	1221	10	96	1286	11
36	1222	10	97	1288	11
37	1224	10	98	1290	11
38	1225	9	99	1291	11
39	1226	9	100	1293	12
40	1227	9	101	1295	12
41	1228	9	102	1297	12
42	1229	9	103	1299	13
43	1231	9	104	1301	13
44	1232	9	105	1304	14
45	1233	9	106	1306	14
46	1234	9	107	1309	15
47	1235	9	108	1312	16
48	1236	9	109	1315	16
49	1237	9	110	1319	17
50	1238	9	111	1323	18
51	1239	9	112	1328	20
52	1240	9	113	1334	22
53	1241	9	114	1341	24
54	1242	9	115	1349	27
55	1243	9	116	1360	31
56	1244	8	117	1375	37
57	1245	8	118	1399	48
58	1245	8	119	1443	71
59	1246	8	120	1500	338
60	1247	8			

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

Part 8: Test Results

8.1 Data

Part 8 of this Technical Report contains information about the results of the 2015 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.

8.1.1 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 2015, 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
2015 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	1017	1282.16	66.29	5%	12%	68%	15%	2%	0%
Ethnic Background									
White	323	1279.87	67.42	6%	14%	66%	14%	3%	0%
Black	74	1280.00	71.17	5%	7%	73%	15%	3%	0%
Hispanic	494	1282.74	63.04	5%	12%	69%	14%	2%	0%
American Indian	74	1287.27	68.70	5%	12%	60%	23%	0%	0%
Asian	24	1265.25	85.51	8%	4%	79%	8%	0%	0%
Hawaiian Pacific Islander	3	*	*	*	*	*	*	*	*
Multiracial	25	1305.20	74.42	0%	12%	56%	32%	0%	0%
Other	0	*	*	*	*	*	*	*	*
Gender									
Male	674	1283.34	65.88	5%	12%	67%	15%	2%	0%
Female	343	1279.84	67.13	5%	12%	69%	14%	3%	0%
Need									
Autism	316	1283.35	54.47	2%	17%	68%	13%	0%	0%
DD	42	1308.93	31.54	0%	0%	81%	19%	0%	0%
ED	4	*	*	*	*	*	*	*	*
EDP	3	*	*	*	*	*	*	*	*
HI	3	*	*	*	*	*	*	*	*
MD	21	1268.19	78.42	10%	10%	67%	14%	5%	0%
MDSSI	62	1203.35	89.20	31%	32%	36%	2%	11%	0%
MIID	297	1311.27	39.91	0%	2%	74%	23%	0%	0%
MOID	122	1258.49	56.54	7%	22%	67%	4%	2%	0%
OHI	16	1309.81	30.09	0%	0%	75%	25%	0%	0%
OI	68	1233.16	101.50	18%	18%	60%	4%	12%	0%
SID	14	1189.86	105.97	36%	21%	43%	0%	7%	0%
SLD	38	1325.39	24.78	0%	0%	66%	34%	0%	0%
SLI	7	*	*	*	*	*	*	*	*
VI	3	*	*	*	*	*	*	*	*
Other	1	*	*	*	*	*	*	*	*
SES									
Free/Reduced Lunch	644	1285.67	60.74	5%	11%	69%	15%	1%	0%
No Lunch Assistance	373	1276.09	74.63	6%	14%	65%	15%	3%	0%
Other	0	*	*	*	*	*	*	*	*
Migrant									
Non-Migrant	1002	1282.08	66.28	5%	12%	68%	15%	2%	0%
Migrant	15	1287.47	69.44	7%	13%	73%	7%	0%	0%
Other	0	*	*	*	*	*	*	*	*
ELL									
Non-ELL	980	1281.28	66.89	5%	12%	67%	15%	2%	0%
ELL	37	1305.43	42.19	0%	5%	78%	16%	0%	0%
Other	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	1000	1277.55	65.77	5%	15%	59%	21%	3%	0%
Ethnic Background									
White	381	1274.90	66.35	5%	18%	58%	19%	3%	0%
Black	78	1283.19	40.55	1%	17%	67%	15%	0%	0%
Hispanic	446	1280.35	61.21	5%	11%	60%	24%	4%	0%
American Indian	50	1266.68	78.25	12%	16%	48%	24%	6%	0%
Asian	19	1274.63	82.19	5%	21%	53%	21%	5%	0%
Hawaiian Pacific Islander	3	*	*	*	*	*	*	*	*
Multiracial	23	1271.87	31.87	4%	17%	70%	9%	0%	0%
Other	0	*	*	*	*	*	*	*	*
Gender									
Male	622	1282.71	58.06	4%	14%	59%	24%	2%	0%
Female	378	1269.07	69.07	7%	16%	60%	17%	5%	0%
Need									
Autism	258	1280.50	48.63	2%	20%	62%	17%	1%	0%
DD	0	*	*	*	*	*	*	*	*
ED	7	*	*	*	*	*	*	*	*
EDP	9	*	*	*	*	*	*	*	*
HI	4	*	*	*	*	*	*	*	*
MD	23	1283.13	31.91	4%	9%	78%	9%	0%	0%
MDSSI	75	1201.83	96.54	27%	32%	39%	3%	17%	0%
MIID	320	1303.50	31.52	0%	4%	61%	34%	0%	0%
MOID	132	1263.16	38.04	5%	21%	70%	5%	1%	0%
OHI	24	1312.79	21.82	0%	0%	54%	46%	0%	0%
OI	82	1243.30	82.07	15%	23%	56%	6%	10%	0%
SID	29	1201.52	96.35	21%	35%	45%	0%	21%	0%
SLD	33	1342.06	44.29	0%	0%	30%	70%	0%	0%
SLI	4	*	*	*	*	*	*	*	*
VI	0	*	*	*	*	*	*	*	*
Other	0	*	*	*	*	*	*	*	*
SES									
Free/Reduced Lunch	587	1287.34	52.78	4%	11%	59%	26%	2%	0%
No Lunch Assistance	413	1263.64	72.53	7%	20%	59%	14%	5%	0%
Other	0	*	*	*	*	*	*	*	*
Migrant									
Non-Migrant	994	1277.44	62.92	5%	15%	59%	21%	3%	0%
Migrant	6	*	*	*	*	*	*	*	*
Other	0	*	*	*	*	*	*	*	*
ELL									
Non-ELL	979	1277.01	62.88	5%	15%	59%	21%	3%	0%
ELL	21	1302.95	52.67	0%	5%	67%	29%	0%	0%
Other	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	892	1276.12	89.78	5%	18%	52%	25%	2%	0%
Ethnic Background									
White	343	1282.76	54.06	4%	16%	52%	28%	2%	0%
Black	70	1275.93	51.75	4%	19%	53%	24%	1%	0%
Hispanic	380	1273.71	61.02	5%	21%	52%	23%	2%	0%
American Indian	56	1264.68	75.14	7%	20%	52%	21%	4%	0%
Asian	21	1248.00	78.16	14%	19%	52%	14%	0%	0%
Hawaiian Pacific Islander	4	*	*	*	*	*	*	*	*
Multiracial	18	1274.83	79.12	6%	6%	67%	22%	6%	0%
Other	0	*	*	*	*	*	*	*	*
Gender									
Male	555	1276.68	60.84	5%	18%	52%	25%	1%	0%
Female	337	1275.20	58.06	5%	18%	53%	23%	2%	0%
Need									
Autism	219	1275.95	49.58	2%	22%	58%	18%	0%	0%
DD	0	*	*	*	*	*	*	*	*
ED	19	1310.84	27.31	0%	0%	47%	53%	0%	0%
EDP	9	*	*	*	*	*	*	*	*
HI	3	*	*	*	*	*	*	*	*
MD	12	1286.00	38.13	0%	17%	58%	25%	0%	0%
MDSSI	51	1201.20	95.89	31%	31%	35%	2%	16%	0%
MIID	298	1299.91	38.69	0%	4%	58%	37%	0%	0%
MOID	146	1259.68	31.77	2%	36%	57%	5%	0%	0%
OHI	21	1318.19	32.27	0%	0%	43%	57%	0%	0%
OI	60	1237.10	75.07	20%	35%	33%	12%	7%	0%
SID	19	1162.95	99.86	42%	53%	5%	0%	16%	0%
SLD	29	1325.52	32.38	0%	0%	31%	69%	0%	0%
SLI	1	*	*	*	*	*	*	*	*
VI	1	*	*	*	*	*	*	*	*
Other	4	*	*	*	*	*	*	*	*
SES									
Free/Reduced Lunch	539	1281.06	54.13	3%	19%	50%	28%	1%	0%
No Lunch Assistance	353	1268.57	66.87	8%	18%	55%	19%	3%	0%
Other	0	*	*	*	*	*	*	*	*
Migrant									
Non-Migrant	886	1276.04	59.86	5%	18%	52%	25%	2%	0%
Migrant	6	*	*	*	*	*	*	*	*
Other	0	*	*	*	*	*	*	*	*
ELL									
Non-ELL	879	1275.66	59.96	5%	19%	52%	24%	2%	0%
ELL	13	1306.92	35.04	0%	0%	54%	46%	0%	0%
Other	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
2015 AIMS A Science Frequency Distribution - Grade 4

Raw Score	Scale Score	FREQ	%	CUML %	Raw Score	Scale Score	FREQ	%	CUML %
0	1000	5	0.5%	0.5%	61	1259	8	0.8%	24.6%
1	1000	0	0.0%	0.5%	62	1260	16	1.6%	26.2%
2	1000	2	0.2%	0.7%	63	1261	20	2.0%	28.2%
3	1018	2	0.2%	0.9%	64	1263	9	0.9%	29.1%
4	1046	3	0.3%	1.2%	65	1264	9	0.9%	30.0%
5	1067	4	0.4%	1.6%	66	1265	16	1.6%	31.5%
6	1085	0	0.0%	1.6%	67	1266	10	1.0%	32.5%
7	1099	1	0.1%	1.7%	68	1268	9	0.9%	33.4%
8	1111	6	0.6%	2.3%	69	1269	11	1.1%	34.5%
9	1121	1	0.1%	2.4%	70	1270	14	1.4%	35.9%
10	1130	0	0.0%	2.4%	71	1272	7	0.7%	36.6%
11	1138	1	0.1%	2.5%	72	1273	18	1.8%	38.4%
12	1145	4	0.4%	2.9%	73	1274	13	1.3%	39.7%
13	1152	1	0.1%	3.0%	74	1276	6	0.6%	40.3%
14	1157	0	0.0%	3.0%	75	1277	5	0.5%	40.8%
15	1163	2	0.2%	3.2%	76	1278	12	1.2%	42.0%
16	1167	1	0.1%	3.3%	77	1280	14	1.4%	43.4%
17	1172	0	0.0%	3.3%	78	1281	12	1.2%	44.6%
18	1176	2	0.2%	3.5%	79	1282	11	1.1%	45.7%
19	1180	0	0.0%	3.5%	80	1284	10	1.0%	46.7%
20	1184	5	0.5%	4.0%	81	1285	14	1.4%	48.1%
21	1187	1	0.1%	4.1%	82	1287	13	1.3%	49.4%
22	1190	1	0.1%	4.2%	83	1288	5	0.5%	49.9%
23	1193	3	0.3%	4.5%	84	1290	9	0.9%	50.7%
24	1196	2	0.2%	4.7%	85	1291	9	0.9%	51.6%
25	1199	1	0.1%	4.8%	86	1293	17	1.7%	53.3%
26	1201	0	0.0%	4.8%	87	1294	21	2.1%	55.4%
27	1204	4	0.4%	5.2%	88	1296	20	2.0%	57.4%
28	1206	6	0.6%	5.8%	89	1298	13	1.3%	58.7%
29	1209	2	0.2%	6.0%	90	1299	10	1.0%	59.7%
30	1211	0	0.0%	6.0%	91	1301	20	2.0%	61.7%
31	1213	2	0.2%	6.2%	92	1303	14	1.4%	63.1%
32	1215	2	0.2%	6.4%	93	1305	15	1.5%	64.6%
33	1217	3	0.3%	6.7%	94	1306	22	2.2%	66.8%
34	1219	1	0.1%	6.8%	95	1308	23	2.3%	69.1%
35	1221	10	1.0%	7.8%	96	1310	17	1.7%	70.7%
36	1223	6	0.6%	8.4%	97	1312	17	1.7%	72.4%
37	1224	3	0.3%	8.7%	98	1315	14	1.4%	73.8%
38	1226	3	0.3%	9.0%	99	1317	16	1.6%	75.4%
39	1228	2	0.2%	9.2%	100	1319	16	1.6%	77.0%
40	1229	0	0.0%	9.2%	101	1322	17	1.7%	78.7%
41	1231	1	0.1%	9.3%	102	1324	21	2.1%	80.8%
42	1232	3	0.3%	9.6%	103	1327	22	2.2%	83.0%
43	1234	5	0.5%	10.0%	104	1330	20	2.0%	85.0%
44	1236	3	0.3%	10.3%	105	1333	11	1.1%	86.1%
45	1237	5	0.5%	10.8%	106	1336	18	1.8%	87.9%
46	1238	6	0.6%	11.4%	107	1340	14	1.4%	89.3%
47	1240	2	0.2%	11.6%	108	1344	11	1.1%	90.3%
48	1241	6	0.6%	12.2%	109	1348	14	1.4%	91.7%
49	1243	6	0.6%	12.8%	110	1353	16	1.6%	93.3%
50	1244	7	0.7%	13.5%	111	1358	10	1.0%	94.3%
51	1246	9	0.9%	14.4%	112	1364	12	1.2%	95.5%
52	1247	12	1.2%	15.6%	113	1371	8	0.8%	96.3%
53	1248	8	0.8%	16.4%	114	1379	12	1.2%	97.5%
54	1250	8	0.8%	17.2%	115	1390	8	0.8%	98.3%
55	1251	15	1.5%	18.7%	116	1404	4	0.4%	98.7%
56	1252	12	1.2%	19.9%	117	1423	3	0.3%	99.0%
57	1254	6	0.6%	20.5%	118	1453	5	0.5%	99.5%
58	1255	9	0.9%	21.4%	119	1500	2	0.2%	99.7%
59	1256	14	1.4%	22.8%	120	1500	3	0.3%	100.0%
60	1257	10	1.0%	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
2015 AIMS A Science Frequency Distribution - Grade 8

Raw Score	Scale Score	FREQ	%	CUML %	Raw Score	Scale Score	FREQ	%	CUML %
0	1000	1	0.1%	0.1%	61	1256	13	1.3%	1.3%
1	1000	1	0.1%	0.1%	62	1257	9	0.9%	0.9%
2	1013	0	0.0%	0.0%	63	1258	8	0.8%	0.8%
3	1047	0	0.0%	0.0%	64	1259	5	0.5%	0.5%
4	1071	3	0.3%	0.3%	65	1260	10	1.0%	1.0%
5	1090	0	0.0%	0.0%	66	1261	10	1.0%	1.0%
6	1105	1	0.1%	0.1%	67	1262	5	0.5%	0.5%
7	1117	0	0.0%	0.0%	68	1263	8	0.8%	0.8%
8	1127	0	0.0%	0.0%	69	1264	11	1.1%	1.1%
9	1136	1	0.1%	0.1%	70	1266	12	1.2%	1.2%
10	1144	0	0.0%	0.0%	71	1267	7	0.7%	0.7%
11	1151	0	0.0%	0.0%	72	1268	13	1.3%	1.3%
12	1158	0	0.0%	0.0%	73	1269	11	1.1%	1.1%
13	1163	0	0.0%	0.0%	74	1270	7	0.7%	0.7%
14	1168	0	0.0%	0.0%	75	1271	9	0.9%	0.9%
15	1173	2	0.2%	0.2%	76	1272	11	1.1%	1.1%
16	1177	6	0.6%	0.6%	77	1274	8	0.8%	0.8%
17	1181	0	0.0%	0.0%	78	1275	12	1.2%	1.2%
18	1184	1	0.1%	0.1%	79	1276	9	0.9%	0.9%
19	1188	1	0.1%	0.1%	80	1277	11	1.1%	1.1%
20	1191	4	0.4%	0.4%	81	1278	11	1.1%	1.1%
21	1194	5	0.5%	0.5%	82	1280	12	1.2%	1.2%
22	1197	2	0.2%	0.2%	83	1281	19	1.9%	1.9%
23	1199	2	0.2%	0.2%	84	1282	17	1.7%	1.7%
24	1202	3	0.3%	0.3%	85	1283	15	1.5%	1.5%
25	1204	0	0.0%	0.0%	86	1285	15	1.5%	1.5%
26	1206	3	0.3%	0.3%	87	1286	7	0.7%	0.7%
27	1209	2	0.2%	0.2%	88	1287	15	1.5%	1.5%
28	1211	7	0.7%	0.7%	89	1289	15	1.5%	1.5%
29	1213	2	0.2%	0.2%	90	1290	13	1.3%	1.3%
30	1214	3	0.3%	0.3%	91	1292	12	1.2%	1.2%
31	1216	2	0.2%	0.2%	92	1293	22	2.3%	2.3%
32	1218	2	0.2%	0.2%	93	1295	15	1.5%	1.5%
33	1220	6	0.6%	0.6%	94	1296	14	1.4%	1.4%
34	1221	2	0.2%	0.2%	95	1298	18	1.8%	1.8%
35	1223	3	0.3%	0.3%	96	1299	22	2.3%	2.3%
36	1225	6	0.6%	0.6%	97	1301	18	1.8%	1.8%
37	1226	4	0.4%	0.4%	98	1303	12	1.2%	1.2%
38	1228	5	0.5%	0.5%	99	1305	14	1.4%	1.4%
39	1229	4	0.4%	0.4%	100	1307	23	2.4%	2.4%
40	1230	2	0.2%	0.2%	101	1309	14	1.4%	1.4%
41	1232	4	0.4%	0.4%	102	1311	14	1.4%	1.4%
42	1233	5	0.5%	0.5%	103	1314	24	2.5%	2.5%
43	1234	7	0.7%	0.7%	104	1316	23	2.4%	2.4%
44	1236	3	0.3%	0.3%	105	1319	18	1.8%	1.8%
45	1237	9	0.9%	0.9%	106	1322	17	1.7%	1.7%
46	1238	4	0.4%	0.4%	107	1325	18	1.8%	1.8%
47	1240	10	1.0%	1.0%	108	1328	23	2.4%	2.4%
48	1241	3	0.3%	0.3%	109	1332	18	1.8%	1.8%
49	1242	6	0.6%	0.6%	110	1336	16	1.6%	1.6%
50	1243	8	0.8%	0.8%	111	1341	14	1.4%	1.4%
51	1244	4	0.4%	0.4%	112	1346	12	1.2%	1.2%
52	1246	9	0.9%	0.9%	113	1353	7	0.7%	0.7%
53	1247	4	0.4%	0.4%	114	1361	12	1.2%	1.2%
54	1248	5	0.5%	0.5%	115	1370	10	1.0%	1.0%
55	1249	7	0.7%	0.7%	116	1383	5	0.5%	0.5%
56	1250	12	1.2%	1.2%	117	1401	7	0.7%	0.7%
57	1251	11	1.1%	1.1%	118	1428	5	0.5%	0.5%
58	1252	2	0.2%	0.2%	119	1479	4	0.4%	0.4%
59	1253	11	1.1%	1.1%	120	1500	2	0.2%	0.2%
60	1255	5	0.5%	0.5%					

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
2015 AIMS A Science Frequency Distribution - High School

Raw Score	Scale Score	FREQ	%	CUML %	Raw Score	Scale Score	FREQ	%	CUML %
0	1000	6	0.7%	0.7%	61	1248	5	0.6%	0.6%
1	1009	0	0.0%	0.0%	62	1249	4	0.5%	0.5%
2	1059	1	0.1%	0.1%	63	1250	6	0.7%	0.7%
3	1088	1	0.1%	0.1%	64	1251	7	0.8%	0.8%
4	1107	3	0.3%	0.3%	65	1252	7	0.8%	0.8%
5	1122	0	0.0%	0.0%	66	1253	7	0.8%	0.8%
6	1133	1	0.1%	0.1%	67	1254	9	1.0%	1.0%
7	1142	0	0.0%	0.0%	68	1255	6	0.7%	0.7%
8	1150	2	0.2%	0.2%	69	1256	10	1.1%	1.1%
9	1157	0	0.0%	0.0%	70	1257	8	0.9%	0.9%
10	1163	1	0.1%	0.1%	71	1258	6	0.7%	0.7%
11	1168	0	0.0%	0.0%	72	1259	3	0.3%	0.3%
12	1172	3	0.3%	0.3%	73	1260	7	0.8%	0.8%
13	1176	0	0.0%	0.0%	74	1261	9	1.0%	1.0%
14	1180	1	0.1%	0.1%	75	1262	3	0.3%	0.3%
15	1183	2	0.2%	0.2%	76	1263	6	0.7%	0.7%
16	1186	5	0.6%	0.6%	77	1264	9	1.0%	1.0%
17	1189	3	0.3%	0.3%	78	1265	4	0.5%	0.5%
18	1192	1	0.1%	0.1%	79	1266	7	0.8%	0.8%
19	1194	1	0.1%	0.1%	80	1267	6	0.7%	0.7%
20	1196	2	0.2%	0.2%	81	1268	9	1.0%	1.0%
21	1199	1	0.1%	0.1%	82	1269	11	1.3%	1.3%
22	1201	4	0.5%	0.5%	83	1270	13	1.5%	1.5%
23	1203	1	0.1%	0.1%	84	1271	9	1.0%	1.0%
24	1205	2	0.2%	0.2%	85	1272	11	1.3%	1.3%
25	1206	4	0.5%	0.5%	86	1273	12	1.4%	1.4%
26	1208	0	0.0%	0.0%	87	1274	10	1.1%	1.1%
27	1210	2	0.2%	0.2%	88	1276	13	1.5%	1.5%
28	1211	6	0.7%	0.7%	89	1277	10	1.1%	1.1%
29	1213	0	0.0%	0.0%	90	1278	2	0.2%	0.2%
30	1214	1	0.1%	0.1%	91	1279	12	1.4%	1.4%
31	1216	4	0.5%	0.5%	92	1281	13	1.5%	1.5%
32	1217	3	0.3%	0.3%	93	1282	6	0.7%	0.7%
33	1219	0	0.0%	0.0%	94	1283	16	1.8%	1.8%
34	1220	1	0.1%	0.1%	95	1285	10	1.1%	1.1%
35	1221	2	0.2%	0.2%	96	1286	16	1.8%	1.8%
36	1222	1	0.1%	0.1%	97	1288	13	1.5%	1.5%
37	1224	2	0.2%	0.2%	98	1290	18	2.0%	2.0%
38	1225	4	0.5%	0.5%	99	1291	14	1.6%	1.6%
39	1226	2	0.2%	0.2%	100	1293	15	1.7%	1.7%
40	1227	5	0.6%	0.6%	101	1295	20	2.3%	2.3%
41	1228	1	0.1%	0.1%	102	1297	16	1.8%	1.8%
42	1229	3	0.3%	0.3%	103	1299	19	2.2%	2.2%
43	1231	4	0.5%	0.5%	104	1301	13	1.5%	1.5%
44	1232	4	0.5%	0.5%	105	1304	22	2.5%	2.5%
45	1233	3	0.3%	0.3%	106	1306	21	2.4%	2.4%
46	1234	5	0.6%	0.6%	107	1309	21	2.4%	2.4%
47	1235	2	0.2%	0.2%	108	1312	13	1.5%	1.5%
48	1236	6	0.7%	0.7%	109	1315	28	3.2%	3.2%
49	1237	3	0.3%	0.3%	110	1319	19	2.2%	2.2%
50	1238	5	0.6%	0.6%	111	1323	18	2.0%	2.0%
51	1239	6	0.7%	0.7%	112	1328	19	2.2%	2.2%
52	1240	5	0.6%	0.6%	113	1334	18	2.0%	2.0%
53	1241	4	0.5%	0.5%	114	1341	15	1.7%	1.7%
54	1242	8	0.9%	0.9%	115	1349	23	2.6%	2.6%
55	1243	5	0.6%	0.6%	116	1360	13	1.5%	1.5%
56	1244	9	1.0%	1.0%	117	1375	9	1.0%	1.0%
57	1245	8	0.9%	0.9%	118	1399	14	1.6%	1.6%
58	1245	6	0.7%	0.7%	119	1443	2	0.2%	0.2%
59	1246	13	1.5%	1.5%	120	1500	7	0.8%	0.8%
60	1247	10	1.1%	1.1%					

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 2015 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).

9.1 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 2015 AIMS A assessments was estimated by internal consistency for each section (Multiple-Choice and Performance Tasks) for each test.

9.1.1 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)

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_X^2}\right),$$

where k = number of items, σ_X^2 = the total score variance, and σ_i^2 = the variance of item i.

Reliability estimates for the 2015 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
2015 AIMS A Science Internal Consistency

		Alpha	
Test	**N**	**MC**	**PT**
4	1,016	0.86	0.95
8	999	0.85	0.96
HS	891	0.86	0.97

9.2 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 2015 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 2015 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 2015 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 2015 AIMS A Grade 4 and Grade 8 Science tests and other tests taken by the students in 2015. The results of this analyses, with correlations all 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 2015 AIMS A Science assessments.

Additional evidence to support the validity of the 2015 AIMS A Science assessments is provided by previous AIMS A technical reports available at www.azed.gov.

9.2.1 Correlations of AIMS A Science to Other Assessments

Correlations were examined between scale scores on 2015 AIMS A tests and the federally mandated, Mathematics and ELA NCSC 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 NCSC 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 NCSC 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 all 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 NCSC ELA and Mathematics – Grade 4

Test	Math	Reading	Science
Math (NCSC)	1.000	.732	.685
Reading (NCSC)	.732	1.000	.762
Science (AIMS A)	.685	.762	1.000

N=992

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

Test	Math	Reading	Science
Math (NCSC)	1.000	.699	.679
Reading (NCSC)	.699	1.000	.755
Science (AIMS A)	.679	.755	1.000

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 2015 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.

10.1 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 Science 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.

10.2 Standard Error of Measurement at Cut Scores

The standard error of measurement (SEM) at each of the score cuts is presented in Table 10.2.1. 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 10.2.1
2015 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	18
8	1197	15	1250	10	1315	15
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 (2015). *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.

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 National Center on Educational Outcomes

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.

Beginning in the 2014-2015 school year, Arizona will transition from AIMS A in mathematics and reading, to the NCSC AA-AAS. Arizona will continue to administer the science AIMS A in grades 4, 8, and 10.

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	NCSC 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.

☐ The NCSC Alternate Assessment is not an administered assessment at the student's grade level for this school year.

NCSC Alternate Assessment Participation Decision Documentation

To meet the criteria for the NCSC Alternate Assessment, 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 requires extensive 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 [NCSC Alternate Assessment] if <u>all responses</u> above are marked Yes.

NCSC Alternate Assessment

Participation Decision Documentation

Additional Considerations Not to Use In Reviewing Evidence	
1.	A disability category or label
2.	Poor attendance or extended absences
3.	Native language/social/cultural or economic difference
4.	Expected poor performance on the general education assessment
5.	Academic and other services received
6.	Educational environment or instructional setting
7.	Percent of time receiving special education services
8.	English Language Learner (ELL) status
9.	Low reading level/achievement level
10.	Anticipated disruptive behavior
11.	Impact of test scores on accountability system
12.	Administrator decision
13.	Anticipated emotional duress
14.	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

In order to be considered for AIMS A, students must meet **all** three of the following criteria in all content areas that are tested: Science (Science is **only** for grades 4, 8, and 10).

1. Evidence of a Significant Cognitive Disability

Empirical evidence (formal testing results, multidisciplinary evaluation team results, etc.) of a significant cognitive disability prevents the acquisition of the grade-level Arizona Academic Content Standards. Please note that students with learning disabilities who have overall intellectual and/or adaptive behavior abilities within the average range are **not** students with *most significant cognitive disabilities*. The student functions *like* a student with an intellectual disability (ID) across *all* areas: commensurate abilities in mathematics, reading, *and* writing, adaptive behavior scores, and measures of intellectual abilities.

Check disability category:

☐ MIID ☐ MOID ☐ SID
☐ MD with ID component ☐ MDSSI with ID component ☐ TBI with ID component
☐ Autism with ID component ☐ Other _____

Example 1: An eighth-grade student functioning at second-grade level in reading and writing and at fourth-grade level in mathematics does *not* qualify under criteria 1.
Example 2: A tenth-grade student functioning at the second-grade level in mathematics, reading, and writing, *does* qualify under criteria 1.

The student meets the *Evidence of a SCD* criterion for AIMS A eligibility.

☐ **Yes** ☐ **No**

2. Curricular Outcomes

The student has access to high-quality instruction based on *Alternate* Academic Standards (Science) and the student's IEP goals and objectives focus on enrolled grade-level *Alternate* Academic Standards.

The student meets the *Curricular Outcomes* criterion for AIMS A eligibility.

☐ **Yes** ☐ **No**

3. Intensity of Instruction

Is extremely difficult for the student to acquire, maintain, generalize, and apply academic skills across environments, even with high-quality extensive/intensive, pervasive, frequent, and individualized instruction in multiple settings in all content areas tested.

The student meets the *Intensity of Instruction* criterion for AIMS A eligibility.

☐ **Yes** ☐ **No**

The student is eligible for AIMS A.

☐ **Yes** (*All* responses above are marked **Yes**.)
☐ **No** (*Any* response above is marked **No** and student must participate in AIMS.)

Parent Notification
Alternate Assessment Participation

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

NCSC Alternate Assessment (ELA/Math) and/or AIMS A Science

- ❑ Yes
- ❑ No (student will participate in statewide achievement test or AIMS

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:_____

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:
2015 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, 2015. 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 82.0%.

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 2015. The sampling was done based on special education need, ethnicity, gender, and region. A total of 28 students were selected.

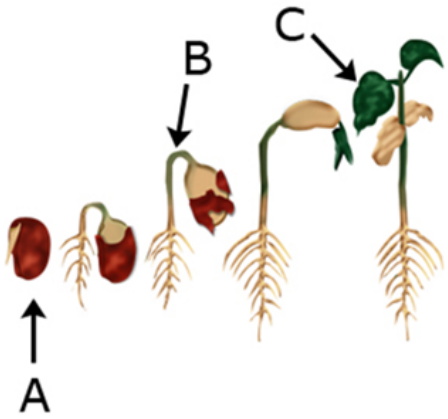
From the committee's suggestions, the following will be continued for the AIMS A 2016 administration.

- Test coordinators will be responsible for completing the AIMS A Science mandatory TC training and for providing training to all TAs who will be administering AIMS A Science.
- TAs will be referred to demonstration videos of performance task administration for clarity on using the rubric.

APPENDIX D:
Example Item Specification Card

AIMS A Item Writing Template

Grade: 4	Item Type: MC	Answer Key: C	DOK Level: 1
Content Area: Science	Strand: 5	Concept: 1	PO: 1



1. Which arrow is pointing to the leaves?

A	B	C