Arizona

Arizona's Instrument to Measure Standards

2015
Technical Report

Submitted to the
Arizona Department of Education
October 2015

ALWAYS LEARNING PEARSON

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 Fall 2014 and Spring 2015 Arizona's Instrument to Measure Standards (AIMS) assessments for the development of tests, analysis of data, calibration, scoring, and scaling. This document also describes the results of the Fall 2014 and Spring 2015 AIMS 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* (American Education Research Association, American Psychological Association, National Council on Measurement in Education, 1999). The *Standards* were revised in 2014, *Standards for Educational and Psychological Testing* (American Education Research Association, American Psychological Association, National Council on Measurement in Education, 2014). The beginning of each part of this technical report will list the different standards addressed in each edition. Part 1 of the technical report addresses 1999 standards 2.7, 3.2, 3.3, 6.3, 6.4, 6.15, and 13.6, and 2014 standards 4.1, 4.2, 7.0, 7.2, 7.4, and 12.9.

A special note concerning the AIMS Administration for 2014-2015

The Fall 2014 AIMS assessments were administered in reading, writing, and mathematics to students in high school who were in Grades 11 and 12 and had not yet obtained a passing score in all three of the content areas. Additionally, students wishing to improve their scores, in any content area, and attain the exceeding category were eligible for this assessment.

Before the Spring 2015 AIMS administration, Arizona Senate Bill 1191 was passed and signed into law by the governor. Senate Bill 1191 provided "for a temporary moratorium on the requirement of obtaining a passing score on standardized tests in order to graduate from high school." This law removed the state graduation requirement that had been in place for reading, writing, and mathematics. The application of this new law meant that the spring 2015 administration of the AIMS tests in reading, writing, and mathematics was voluntary for Arizona high school students. The data collected in this technical report for the spring 2015 administration for reading, writing, and mathematics in high school is based on that pool of students who voluntarily took AIMS.

Furthermore, during the 2014-2015 school year, the new AzMERIT assessments were administered for the first time (in spring 2015) to fulfill the ongoing federal ESEA requirements for annual assessment in ELA and mathematics. As a result, starting in spring 2015, the AIMS assessments in reading, writing, and mathematics were no longer administered to students in grade 3 through grade 8.

Structure of AIMS Technical Report

The Spring 2015 AIMS assessments were designed and developed to provide fair and accurate ability scores that support appropriate, meaningful, and useful educational decisions. In addition to the evidence provided in Part 2 (Involvement of Arizona Educators), additional validity evidence may be found in the following parts as described: Part 3 (Test Design), Part 4 (Test Development), Part 5 (Test Administration), Part 6 (Classical Item Analysis), Part 7 (Calibration, Scaling and Equating), Part 8 (Reliability), and Part 10 (Classification). As the technical report progresses

chapter by chapter, it moves through the phases of the testing cycle. Each part of the technical report details the procedures and processes applied in the creation of AIMS, 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.*

Students in high school began taking AIMS (Form A) in reading, writing, and mathematics in 1999. The AIMS assessments are designed to measure Arizona students' performance on the Arizona content standards. The AIMS Reading test was written to Arizona content standards adopted in March 2003. The AIMS Writing test was written to content standards adopted in June 2004. It was revised to include multiple-choice items along with a written essay in the Spring 2011 AIMS. New performance standards were set for these writing tests in spring 2011. The AIMS Mathematics test assesses content standards adopted in June 2008. Performance standards were set for the AIMS Mathematics test in spring 2010.

Students' test scores on the AIMS high school tests were one component of the high school graduation requirements, and passing scores were required to earn a diploma for students who graduated beginning in spring 2006 till the fall of 2014. As noted above, students in high school were no longer required to pass the AIMS high school tests in order to earn the high school deploma starting in the spring of 2015. The AIMS high school tests in reading and mathematics consist of multiple-choice items. The AIMS high school test in writing consists of a set of multiple choice items and a single prompt essay, which is scored using a holistic six-point rubric (see Appendix C).

The Spring 2015 AIMS tests were also administered in science to students in grade 4, 8, and high school. This was the sixth year that Grades 4, 8, and high school were administered science. These AIMS Science tests remain mandatory for all general education students in these grades. The AIMS Science tests consist of multiple-choice items, which are written entirely by Arizona teachers.

The AIMS assessments are designed to measure Arizona students' performance on the Arizona content standards. All AIMS Science tests are written to Arizona content standards approved by the State Board on May 24, 2004, and updated on March 10, 2005.

Based on the input of Arizona educators's review of the content standards, a design was derived, developed, administered, and scored. 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. A brief content summary of the report is provided below.

Involvement of Arizona Educators

- Part 2 of this report describes the involvement of Arizona educators in test development.
- Several committees met throughout the year in preparation for the 2015 AIMS assessments.

Test Design and Development

- Part 3 of this report describes the test design and the item development process. It provides the content frameworks and the blueprints upon which all of the AIMS tests are based. This section also includes descriptions and the structure of each AIMS test administered in the 2014-2015 academic year.
- Part 4 of this report provides a chronological description of the passage, stimulus, and item development process including modification of specifications, committee passage/stimulus reviews, item content and sensitivity reviews, data analysis and item

selection committees, and customer and contractor reviews to guarantee a quality, error-free product.

Administration

- Part 5 briefly describes test administration, accommodations, security, and the written procedures available to all test administrations and school personnel.
- The accommodations were available to eligible students while testing on AIMS.
- The same accommodations were available for both Fall 2014 and Spring 2015 AIMS.
- Personnel involved in testing administration were asked to sign a security agreement form certifying that all AIMS tests were administered under secure testing conditions.
- In order to ensure standardized testing administration for all students, a *Test Coordinator Manual* was made available to all test coordinators. Also, *Test Administration Directions* were made available to all test administrators.

Data for Operational Analysis

- Part 6 describes the data used for calibration and scaling of the Spring 2015 AIMS and also presents classical test statistics and item analysis statistics.
- In order to ensure valid calibration and scaling, several data cleaning steps occurred.
- The values for Cronbach's alpha were provided as a measure of internal consistency.

Calibration, Scaling, and Equating

- Part 7 reviews calibration, equating, scoring methods, and calibration results. Evaluation of the calibration results includes model-to-item fit.
- Displacement values and other item characteristics were considered for evaluating anchor items.
- Part 7 also shows the relationships between raw scores and scale score through scoring tables.
- Scaling results including the standard error of measurement are also presented.
- For all content areas, scoring tables were established using students' responses to the spring 2015 administration.

Test Results

- Part 8 summarizes information about the results of the spring 2015 administration of AIMS high school. The test results for different ethnic backgrounds and special program membership status are provided.
- Results for AIMS high school assessments are reported by graduating cohort, viz., for students graduating in years from 2015 to 2017. Students in cohort 2018 are included in the high school science results.
- Scale score frequency distributions with three cut scores are also presented.

Validity Evidence

- Part 9 reviews the main validity issues discussed in all prior chapters and provides additional validity evidence supporting the AIMS tests.
- For reading, mathematics, and science, Cronbach's alpha was estimated and is provided as a measure of internal consistency, where for writing, inter-rater position consistency and stratified alpha were estimated and are provided.
- An analysis of differential item functioning is presented.
- Correlations among assessments are presented in the context of construct validity.

Classification

- Part 10 provides information regarding classification consistency and accuracy when students were classified into proficiency categories.
- The cut scores used for classifying proficiency categories were determined during standard setting and adopted by the State Board of Education.

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 for Educational and Psychological Testing* (AERA, APA, NCME, 1999), and standard 4.6 in the 2014 edition.

Typically several committees met throughout the year in preparation for AIMS Writing, Reading, Mathematics, and Science assessments. These committees included teachers, curriculum specialists, and administrators from across the state and were an integral part of both the AIMS test development processes and AIMS results interpretation. However, because Arizona moved to a new assessment for reading, writing, and mathematics starting spring 2015, and because they had developed a sufficient number and quality of items in the Science item bank, they chose to change their process for spring 2015 test development.

The 2015 AIMS called for administering one operational test per grade per content area. Committee meetings focused on the selection of all items to be used. Note that Data Analysis and Item Selection committee meetings were held in summer to construct AIMS for the next year in the past; however the AIMS Science forms for the spring 2015 administration were built by trained ADE staff, most of whom also held Arizona teacher certificates, in the summer of 2014. The Spring 2015 AIMS Reading, Mathematics, and Writing high school tests were reused forms of previously administered tests. Thus, there was no Item Selection committee held in the summer.

PART 3: TEST DESIGN

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

3.1 Content Standards

The AIMS assessments are designed to measure performance on the Arizona content standards adopted in March 2003 for reading, June 2008 for mathematics, June 2004 for writing, and March 2005 for science. These standards are organized by strand, concept, and performance objective. The AIMS Reading and Mathematics test blueprints are based on the concepts and strands of the Arizona content standards, presented in Figures 3.1.1-3.1.2. The AIMS Writing tests were revised in spring 2011 to include multiple-choice items and a writing prompt. The writing tests address the six concepts that are incorporated in Strand 2 of the Writing Standard. Figure 3.1.3 presents the statement of the six concepts in Strand 2.The AIMS Science test blueprints are based on the concepts and strands of the Arizona content standards, presented in Figures 3.1.4 through 3.1.6.

Figure 3.1.1
Arizona Reading Concepts and Strands

Strand 1: Reading Process

- **Concept 1: Print Concepts**
- **Concept 3: Phonics**
- **Concept 4: Vocabulary**
- **Concept 6: Comprehension Strategies**

Strand 2: Comprehending Literary Text

- **Concept 1: Elements of Literature**
- **Concept 2: Historical and Cultural Aspects**

Strand 3: Comprehending Informational Text

- **Concept 1: Expository Text**
- **Concept 2: Functional Text**
- **Concept 3: Persuasive Text**

Figure 3.1.2
Arizona Mathematics Concepts and Strands

Strand 1: Number and Operations

- **Concept 1: Number Sense**
- **Concept 2: Numerical Operations**
- **Concept 3: Estimation**

Strand 2: Data Analysis, Probability and Discrete Math

- **Concept 1: Data Analysis (Statistics)**
- **Concept 2: Probability**
- **Concept 3: Systematic Listing and Counting**
- **Concept 4: Vertex-Edge Graphs**

Strand 3: Patterns, Algebra and Functions

- **Concept 1: Patterns**
- **Concept 2: Functions and Relationships**
- **Concept 3: Algebraic Representations**
- **Concept 4: Analysis of Change**

Strand 4: Geometry and Measurement

- **Concept 1: Geometric Properties**
- **Concept 2: Transformation of Shapes**
- **Concept 3: Coordinate Geometry**
- **Concept 4: Measurement**

Strand 5: Structure and Logic

- **Concept 1: Algorithms and Algorithmic Thinking**
- **Concept 2: Logic, Reasoning, Problem Solving and Proof**

Figure 3.1.3
Arizona Writing Concepts in Strand 2

Trait 1: Ideas and Content

Trait 2: Organization

Trait 3: Voice

Trait 4: Word Choice

Trait 5: Sentence Fluency

Trait 6: Conventions

Figure 3.1.4
Arizona Science Concepts and Strands – Grade 4

Strand 1: Inquiry Process

- **Concept 1: Observations, Questions, and Hypotheses**
- **Concept 2: Scientific Testing (Investigating and Modeling)**
- **Concept 3: Analysis and Conclusions**
- **Concept 4: Communication**

Strand 2: History and Nature of Science

- **Concept 1: History of Science as a Human Endeavor**
- **Concept 2: Nature of Scientific Knowledge**

Strand 3: Science in Personal and Social Perspectives

- **Concept 1: Changes in Environments**
- **Concept 2: Science and Technology in Society**

Strand 4: Life Science

- **Concept 1: Characteristics of Organisms**
- **Concept 2: Life Cycles**
- **Concept 3: Organisms and Environments**
- **Concept 4: Diversity, Adaptation, and Behavior**

Strand 5: Physical Science

- **Concept 1: Properties of Objects and Materials**
- **Concept 2: Position and Motion of Objects**
- **Concept 3: Energy and Magnetism**

Strand 6: Earth and Space Science

- **Concept 1: Properties of Earth Materials**
- **Concept 2: Earth's Processes and Systems**
- **Concept 3: Changes in the Earth and Sky**

Figure 3.1.5
Arizona Science Concepts and Strands – Grade 8

Strand 1: Inquiry Process

- **Concept 1: Observations, Questions, and Hypotheses**
- **Concept 2: Scientific Testing (Investigating and Modeling)**
- **Concept 3: Analysis and Conclusions**
- **Concept 4: Communication**

Strand 2: History and Nature of Science

- **Concept 1: History of Science as a Human Endeavor**
- **Concept 2: Nature of Scientific Knowledge**

Strand 3: Science in Personal and Social Perspectives

- **Concept 1: Changes in Environments**
- **Concept 2: Science and Technology in Society**

Strand 4: Life Science

- **Concept 1: Structure and Function in Living Systems**
- **Concept 2: Reproduction and Heredity**
- **Concept 3: Populations of Organisms in an Ecosystem**
- **Concept 4: Diversity, Adaptation, and Behavior**

Strand 5: Physical Science

- **Concept 1: Properties and Changes of Properties in Matter**
- **Concept 2: Motion and Forces**
- **Concept 3: Transfer of Energy**

Strand 6: Earth and Space Science

- **Concept 1: Structure of the Earth**
- **Concept 2: Earth's Processes and Systems**
- **Concept 3: Earth in the Solar System**

Figure 3.1.6
Arizona Science Concepts and Strands – 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**
- **Concept 2: Nature of Scientific Knowledge**

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 assessments were designed in accordance with the following blueprints in Tables 3.2.1 through 3.2.6. Tables 3.2.1, 3.2.2, and 3.2.6 show the blueprints for only high school in reading, mathematics, and writing, respectively, as only the high school level was administered in reading, mathematics, and writing in spring 2015. Further discussion of item selection to match the blueprints is included in Part 4 of this report.

Table 3.2.1
AIMS Blueprint for Reading

AIMS Reading Blueprint (beginning Spring 2005)

Grade	HS	Strand 1						Strand 2		Strand 3		
		Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6	Concept 1	Concept 2	Concept 1	Concept 2	Concept 3
	% of test	0%	0%	0%	7%	0%	7%	26%	7%	22%	15%	15%
	% of strand on test	15%						33%		52%		

Arizona Department of Education 1 of 1 January 10, 2006

The table has been abridged to show only the high school percentages, as testing in grades 3 through 8 has been suspended.

Source: http://www.azed.gov/assessment/files/2014/06/reading-blueprint-1-10-06.pdf

Table 3.2.2
AIMS Blueprint for Mathematics

AIMS Mathematics Blueprint (beginning with the 2010 Assessments)

Strand/Concept	HS
1. Number and Operations	**6%**
1.1 Number Sense	6%
1.2 Numerical Operations	
1.3 Estimation	
2. Data Analysis/Prob/Discrete	**14%**
2.1 Data Analysis (Statistics)	5%
2.2 Probability	5%
2.3 Systematic Listing and Counting	5%
2.4 Vertex-Edge Graphs	
3. Patterns/Algebra/Functions	**33%**
3.1 Patterns	5%
3.2 Functions and Relationships	7%
3.3 Algebraic Representations	16%
3.4 Analysis of Change	5%
4. Geometry and Measurement	**33%**
4.1 Geometric Properties	13%
4.2 Transformation of Shapes	5%
4.3 Coordinate Geometry	8%
4.4 Measurement	7%
5. Structure and Logic	**14%**
5.1 Algorithms and Algorithmic Thinking	14%
5.2 Logic, Reasoning, Prob Solving, & Proof	

The table has been abridged to show only the high school percentages, as testing in grades 3 through 8 has been suspended.

Source: http://www.azed.gov/assessment/files/2014/06/aims-mathematics-blueprint-beginning-2010-updated-5-5-11.pdf

Table 3.2.3
AIMS Blueprint for Science Grade 4

AIMS Science
Grade 4 Test Blueprint

Strand/Concept	% of Test	# of Items
Strand 1: Inquiry Process	**33.3%**	
Concept 1: Observations, Questions, and Hypotheses	11.1%	6
Concept 2: Scientific Testing (Investigating and Modeling)	11.1%	6
Concept 3: Analysis and Conclusions	11.1%	6
Concept 4: Communications		
Strand 2: History and Nature of Science	**11.1%**	
Concept 1: History of Science as a Human Endeavor	11.1%	6
Concept 2: Nature of Scientific Knowledge		
Strand 3: Science in Personal and Social Perspectives	**11.1%**	
Concept 1: Changes in Environments	11.1%	6
Concept 2: Science and Technology in Society		
Strand 4: Life Science	**11.1%**	
Concept 1: Characteristics of Organisms	11.1%	6
Concept 3: Organisms and Environments		
Concept 4: Diversity, Adaptations, and Behavior		
Strand 5: Physical Science	**11.1%**	
Concept 3: Energy and Magnetism	11.1%	6
Strand 6: Earth and Space Science	**22.2%**	
Concept 2: Earth's Processes and Systems	11.1%	6
Concept 3: Changes in the Earth and Sky	11.1%	6
According to the Science Standard, the following Strands and Concepts do not have Performance Objectives for Grade 4: **Strand 4: Life Science, Concept 2** (Life Cycles); **Strand 5: Physical Science, Concept 1** (Properties of Objects and Materials) and **Concept 2** (Position and Motion of Objects); **Strand 6: Earth and Space Science, Concept 1** (Properties of Earth Materials).		**54**

Source: http://www.azed.gov/assessment/files/2014/06/science-blueprint-with-item-counts-11-10-09.pdf

Table 3.2.4
AIMS Blueprint for Science Grade 8

AIMS Science
Grade 8 Test Blueprint

Strand/Concept	% of Test	# of Items
Strand 1: Inquiry Process	**34.5%**	
Concept 1: Observations, Questions, and Hypotheses	10.3%	**6**
Concept 2: Scientific Testing (Investigating and Modeling)	6.9%	**4**
Concept 3: Analysis and Conclusions	10.3%	**6**
Concept 4: Communications	6.9%	**4**
Strand 2: History and Nature of Science	**10.3%**	
Concept 1: History of Science as a Human Endeavor Concept 2: Nature of Scientific Knowledge	10.3%	**6**
Strand 3: Science in Personal and Social Perspectives	**10.3%**	
Concept 1: Changes in Environments Concept 2: Science and Technology in Society	10.3%	**6**
Strand 4: Life Science	**13.8%**	
Concept 2: Reproduction and Heredity Concept 4: Diversity, Adaptations, and Behavior	13.8%	**8**
Strand 5: Physical Science	**31.0%**	
Concept 1: Properties and Changes of Properties in Matter	17.2%	**10**
Concept 2: Motion and Forces	13.8%	**8**
According to the Science Standard, the following Strands and Concepts do not have Performance Objectives for Grade 8: **Strand 4: Life Science**, **Concept 1** (Structure and Function in Living Organisms) and **Concept 3** (Populations of Organisms in an Ecosystem); **Strand 5: Physical Science, Concept 3** (Transfer of Energy).		**58**

Source: http://www.azed.gov/assessment/files/2014/06/science-blueprint-with-item-counts-11-10-09.pdf

Table 3.2.5
AIMS Blueprint for Science High School

AIMS Science

High School Test Blueprint

Strand/Concept	% of Test	# of Items
Strand 1: Inquiry Process	**33.8%**	
Concept 1: Observations, Questions, and Hypotheses	9.2%	**6**
Concept 2: Scientific Testing (Investigating and Modeling)	9.2%	**6**
Concept 3: Analysis, Conclusions, and Refinements	9.2%	**6**
Concept 4: Communications	6.2%	**4**
Strand 2: History and Nature of Science	**9.2%**	
Concept 1: History of Science as a Human Endeavor	9.2%	**6**
Concept 2: Nature of Scientific Knowledge		
Strand 3: Science in Personal and Social Perspectives	**10.8%**	
Concept 1: Changes in Environments	10.8%	**7**
Concept 2: Science and Technology in Society		
Concept 3: Human Population Characteristics		
Strand 4: Life Science	**46.2%**	
Concept 1: The Cell	9.2%	**6**
Concept 2: Molecular Basis of Heredity	9.2%	**6**
Concept 3: Interdependence of Organisms	9.2%	**6**
Concept 4: Biological Evolution	9.2%	**6**
Concept 5: Matter, Energy, and Organization in Living Systems (Including Human Systems)	9.2%	**6**
		65

Source: http://www.azed.gov/assessment/files/2014/06/science-blueprint-with-item-counts-11-10-09.pdf

Table 3.2.6
AIMS Blueprint for Writing

AIMS Writing Blueprint May 2010
Strand 2, Concepts 1-6

Grade	10	
Concepts	% MC Items	% of Score
1. Ideas and Content	44%	18%
2. Organization		
3. Voice	33%	13%
4. Word Choice		
5. Sentence Fluency		
6. Conventions	22%	9%
Total Multiple Choice	100%	40%
Extended Response		60%
Totals		100%

The Blueprint was proposed on May 29, 2009 and revised on May 19, 2010.
The table has been abridged to show only the grade 10 percentages, as testing in grades 3 through 8 has been suspended.

Source: http://www.azed.gov/assessment/files/2014/06/writing-blueprint-5-19-10.pdf

3.3 Description of 2015 AIMS Tests

The test blueprints were used with the processes described in detail in Part 4 to develop all AIMS tests administered in 2015. The resulting test configurations are as follows.

3.3.1 Reading for High School

The AIMS Reading test for high school consisted of 54 multiple-choice items developed by Arizona teachers. The raw scores ranged from 0-54, and scale scores were designed to range from 500 to 900. All items on the high school reading test reported to a criterion-referenced score. No norm-referenced items were included on the high school reading test. Ten reading field test items were embedded with the operational items to form a total of 64 reading test items.

Table 3.3.1.1
Spring 2015 AIMS Test Structure of Reading for High School

Grade	RD FT	RD OP	TOTAL ITEMS ON TEST
HS	10	54	54

*The high school reading and writing tests are administered separately. The writing test contains 32 multiple-choice items and 1 prompt for 33 total items.

3.3.2 Writing for High School

The AIMS Writing test form for high school consisted of one extended-response writing prompt and 27 multiple-choice items. The multiple-choice component is weighted 40% and the essay response is weighted 60% in the total score. Responses to the prompt were scored on the holistic six-point rubric (see appendix D). Each essay response received two ratings. Final scores for responses with adjacent ratings were derived by averaging the two ratings. Final scores for responses with discrepant ratings (difference of 2 points) were resolved by a third rater. The raw scores ranged from 0-138, and scale scores were designed to range from 300-700. There were two forms of the high school writing test, A and T. Form T was used as a make-up form administered one week after the administration of Form A. No norm-referenced items were included on the high school writing tests. Five field test items were embedded with the operational items to form a total of 32 multiple-choice items and one prompt.

Table 3.3.2.1
Spring 2015 AIMS Test Structure of Writing for High School

Grade	WR FT	WR OP	TOTAL ITEMS ON TEST
HS	5	27	32

3.3.3 Mathematics for High School

The AIMS Mathematics test form for high school consisted of 85 multiple-choice items developed by Arizona teachers. The raw scores ranged from 0-85, and scale scores were designed to range from 300 to 700. All items on the high school mathematics test reported to a criterion-referenced score. New performance standards were set in spring 2010. No norm-referenced items were included in the high school mathematics test. Fifteen field test items were embedded with the operational items to form a total of 100 test items.

Table 3.3.3.1
Spring 2015 AIMS Test Structure of Mathematics for High School

Grade	MA FT	MA OP	TOTAL ITEMS ON TEST
HS	15	85	100

3.3.4 Science for Grades 4, 8, and High School

The 2014 AIMS Science tests consisted of one operational form with 54 multiple-choice items on the grade 4 test, 58 multiple-choice items on the grade 8 test, and 65 multiple-choice items on the high school test. All multiple-choice items were developed by Arizona teachers. Ten field test items written to the Arizona standards were embedded with the operational items at each grade level. The scale scores for each test range from 200 to 800 and all items on each test reported to a criterion-referenced score. No norm-referenced items were included on any of the science tests. Table 3.3.10.1 displays the structure of the science tests.

Table 3.3.4.1
Spring 2015 AIMS Test Structure of Science

Grade	SC FT	SC OP	TOTAL ITEMS ON TEST	Anchor
4	N/A	54	54	21
8	N/A	58	58	23
HS	N/A	65	65	18

*Grades 4, 8, and HS science each had no field test items on spring 2015 tests.

3.3.5 AIMS Score Ranges

Raw score and scale score ranges of 2015 AIMS Reading, Mathematics, and Writing assessments for high school and AIMS Science in grades 4, 8, and high school are presented in Table 3.3.5.1.

Table 3.3.5.1
Raw Score and Scale Score ranges of 2015 AIMS Assessments

Content	Grade	Raw Score Range	Scale Score range
Reading	HS	0-54	500-900
Writing	HS	0-138	300-700
Mathematics	HS	0-85	300-700
Science	4	0-54	200-800
	8	0-58	200-800
	HS	0-65	200-800

PART 4: TEST DEVELOPMENT

Part 4 of the technical report provides a summary of the test development activities that occurred during the 2014-2015 contract year. Information is provided relating to the following topics as they pertain to AIMS:

- a discussion of the AIMS test development and editing process;
- a description of the use of previously created AIMS item specifications;
- a description of the AIMS item editing procedures;
- a description of the data analysis committee procedures;
- a description of the AIMS item selection committee meetings; and

A comprehensive, multi-segment development process guides the development of assessment materials. The following section outlines this process in general terms. The remainder of Part 4 provides details of how these processes were implemented in Arizona. This section of the technical report addresses the following AERA/APA/NCME *Standards* from the 1999 edition: 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 new edition of *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

4.1 AIMS Test Development and Editing Process

4.1.1 Test Development Process

Test development for the 2015 test administration began with the planning meeting held in Phoenix, January 16-18, 2013. During this meeting, the project deliverables were defined, such as number of forms, answer documents, test administration manuals, test coordinator manuals, test interpretation guides, and materials to support special accommodations, including Braille and large print books. The actual test form design was unchanged from the previous year. The ancillary materials were modified and all modifications were discussed and shared among all team members to ensure understanding. In the meeting, it was decided that the Spring 2015 AIMS Reading and Mathemetics tests for high school would be reused forms of the spring 2014 tests. It was also decided that the multiple choice portion of Spring 2015 AIMS Writing tests for high school would be the same as the spring 2012 test while the essay portion of the writing test would be the same as the spring 2013 administration. Thus, no Item Selection Committee was held for the Spring 2015 AIMS Reading, Mathematics, and Writing tests for high school.

4.1.2 Documents and Materials Development

Following definition of project deliverables, Pearson's entire test development team reviewed the blueprints, item specifications, and the *ADE Style Guide* to ensure that the 2015 assessment would meet all of the required, previously-developed criteria.

4.1.3 Item Writing for Science

The no new items were developed for field testing in the Spring 2015 AIMS Science assessments since there were sufficient items of sufficient quality in the AIMS item bank.

4.1.4 Quality Reviews

ADE and Pearson personnel implemented a series of quality review checks at various stages of production to ensure all AIMS materials were error free.

ADE first reviewed each component at a relatively early stage of forms 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. Answer keys were checked. All changes were approved in writing by ADE.

A smooth AIMS test administration requires that all test materials, including test books, answer documents, and directions to students and test coordinators align with each other. Therefore, Pearson and ADE conducted a review of all materials as the second quality check. A side benefit of this review was the detection of possible revisions required on any unclear field test items.

Prior to creation of proofs (blueline stage), Pearson performed a Final Forms review. The purpose of the Final Forms review was to ensure that all publishable products met ADE's high quality standards and expectations.

After Pearson conducted their Final Forms review, all test forms were again submitted to ADE for review. All final forms and documents were reviewed and approved by ADE content specialists.

4.2 Pool of Items Used for Test Construction

4.2.1 Item Specifications

The item specifications were developed by Pearson and ADE in May 2009. The item specifications provide a definition of what is tested by each Performance Objective (PO) and, where needed, provide clarification of the PO statements, the content limits, and the stimulus and response attribute descriptions. Taken together, these 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.

4.2.2 Data Analysis

AIMS Data Analysis was conducted for Science in June 2014. Primary responsibility for conducting this workshop rested with ADE. The primary purpose of the Data Analysis meeting was to examine the item data generated for field tested items within the Spring 2014 AIMS Science test. Each item was assigned a status code to be included with the item information in the item bank, and determine each item's eligibility for possible selection as an operational item starting in spring 2015.

ADE staff were trained on how to interpret basic statistical concepts related to item data including *p*-values, Rasch values, infit/outfit, point biserial correlations, response distributions and ethnic and gender differential item functioning (DIF) flags, omit rates, and population counts.

Items that measured the content they were intended to measure and whose statistics were within acceptable limits were assigned Item Available (IA) status. These items were eligible for selection as operational items. Throughout the meeting, content was stressed as the deciding factor over statistics for items to attain IA status. Across all grades in Science, approximately 87% of the items received IA status.

Items whose statistics indicated a fixable problem and that defined where the items could be improved were assigned Re-Field Test (RFT) status. These items would be revised during future item writing workshops and would be re-field tested in future assessments. None of items reviewed was coded RFT.

Items whose statistics indicated they would not function fairly and reliably were rejected and assigned Do Not Use (DNU) status. These items were removed from consideration as operational items. Across the content and grade levels, about 13% of the items were assigned DNU status.

Table 4.1 shows the number and portion of items classified into each category during Data Analysis by grade level.

Table 4.1
Items Given Special Codes

Content Area	Grade	Items Reviewed	Items Assigned IA * Status		Items Assigned RFT* Status		Items Assigned DNU* Status	
Science	4	40	36	90%	0	0%	4	10%
	8	40	34	85%	0	0%	6	15%
	HS	40	34	85%	0	0%	6	15%
Science Total		**120**	**104**	**87%**	**0**	**0%**	**16**	**13%**

Note:* Item Available (IA) - Re-field Test (RFT) - Do Not Use (DNU) ** For reading, since going forward, no further item development was expected, no items were marked as RFT where for Mathematics, the RFT items were identified for use in field-test slots in spring 2014.

4.2.3 AIMS Item Selection

AIMS Item Selection meeting for science was conducted by ADE staff in July 2014. The primary purpose of the Item Selection meeting was to select items to place on test forms for the spring 2015 operational test that would produce valid and reliable scores using the items from the 2014 field test administration that had been designated as "item accepted" (IA) as well as using items from previous test administrations. Two sets of criteria primarily guided the selection of AIMS items: content representation and statistical requirements. In addition, the committee members were encouraged to select items with high-level DOKs in order to help prepare students for assessments based on the Arizona Science Standard.

All of the items in the item bank that were available and eligible for selection as operational items in spring 2014 were displayed in grade level and content area item pool tables. With minor exceptions, the pool consisted of items field tested in 2008 through 2013. The items field tested in spring 2014 were also available in the data analysis materials. The item pool tables for the science committee were arranged by Performance Objective. All tables could also be sorted according to any of the columns, making them extremely useful tools for searching for items with specific characteristics. These items formed the pool for item selection. Item images could be viewed electronically via the item bank. The meeting room was equipped with a laptop with access to the item bank and a projection screen so that the entire group could view items at the same time.

Each entry on the table contained identification numbers, content alignment information (Strand, Concept, Performance Objective), the most recent test administration, and the most current statistical information about that item (*p*-value, Rasch values, point biserial, differential item functioning summary flags, Rasch model fit statistics, and the percent of students who omitted the item). Participants were given training to interpret these statistics and statistical guidelines for test selection. These guidelines included a target difficulty level for each test. Specifically, a target mean and range of selected item *p*-values, as well as a suggested distribution for the item *p*-values was provided for each grade/subject combination. Careful adherence to the specified distribution of *p*-values guaranteed students a reasonable opportunity to do well on a test that would be neither too easy nor too hard.

In addition to selecting items within specific *p*-values ranges, committee members were also asked to select items with item discriminations that indicate that getting the item correct is

reasonably correlated with performance on the entire test (i.e., preferably item correlations greater than 0.3) and do not exhibit the potential for item bias (i.e., the items should not be flagged using various differential item functioning statistics).

Content considerations were addressed by the test blueprints. Careful adherence to the blueprints guaranteed the tests would validly measure the construct of science as represented in the Arizona Science Standard, maintain consistency, link to instruction, and allow for selection of items from different performance objectives within each concept. Substantial variance from the test blueprint could alter the test alignment and thus the validity of the scores being reported. Items were selected to represent the significant content categories specified in the test blueprint in the same proportion as the content categories represented in the test blueprint.

Prior to the Item Selection Committee meeting, ADE selected an anchor set of items upon which the operational forms would be constructed. The anchor set consisted of items that had been operational at least the previous year (during the spring 2014 test administration). Regardless of the grade, each anchor set was carefully selected to meet statistical criteria and to proportionally represent the blueprint. Anchor sets were finalized by ADE prior to the item selection workshop.

To facilitate the selection process and to guarantee that the proper number and proportion of items would be selected, participants were provided with item pool tables and item replacement tables. Table 4.2 shows a sample of an item pool table and the available data considered by the Item Selection Committee in its selection of replacement items. An analysis of differential item functioning is performed for every administration. The latest values are included in the item pool tables for each grade/content area and provided to participants in the Item Selection Committee. Table 4.3 is a sample portion of the Item Replacement Table used by the participants to note their replacement requirements for grade 4 Science and to capture proposed items to be used on the spring 2015 assessment. This sample table shows the portion relevant to Strand 1 Concept 1 only. The entire table included all strands and concepts. This sample table shows the portion of columns relevant to spring 2014 and spring 2015. The information in the first column shows the blueprint requirements for Strand 1, Concept 1 – six of the 54 operational items that should be covered by items from Strand 1, Concept 1 in the grade 4 Science test.

The set of columns labeled Spring 2015 New Operational Items include all of the AZ items covering Strand 1 Concept 1 that were in the spring 2014 test. The set of columns labeled Spring 2015 New Operational Items show the items that were retained from the spring 2014 or prior administrations (highlighted in blue). These retained items were designated as anchor items. During item selection for spring 2014, the participants' tasks were to retain anchor items, if possible, and select items to fill in any gaps in blueprint coverage. As the participants considered each option based on content and difficulty, they could refer to the Item Pool Table to determine if the statistical considerations were being met and to the item bank to see the actual items.

As selections were made, they were recorded on item replacement tables. These tables were loaded onto computers and projected for group discussion. These tables provided a running record of the selections and further helped to guarantee blueprint coverage. Table 4.43 shows a sample of the *p*-value target distribution table and graph used by the committees. Note that this table and graph are displayed as if items were in the process of being selected. These tables were completed for all selections and were subject to approval by both ADE and Pearson's content and psychometric departments.

Table 4.5 shows the numbers of AIMS Science items that were selected for each grade. All selections were approved by Pearson content and psychometric staff and ADE staff.

Table 4.2
Sample Grade 4 Science Item Pool Table

Page 1

Row	AZID	Subject	Grade	Status	Stimulus Title	Strand	Concept	Perf. Obj.	DOK	Year 2006	Year 2007	Year 2008	Year 2009	Year 2010	Year 2011	Year 2012	Year 2013	Year 2014	Recent Year	Item No.
1	44144025	Science	4	New	Circuit Study	5	3	2	2									FT	2014	7
2	44144005	Science	4	New	Soil Erosion	6	2	3	2									FT	2014	59
3	44144047	Science	4	New		1	1	1	2									FT	2014	6
4	44144049	Science	4	New		1	1	2	2									FT	2014	6
5	44144051	Science	4	New		1	1	2	2									FT	2014	6
6	44144055	Science	4	New		1	1	2	2									FT	2014	7
7	44144041	Science	4	New		1	1	2	4											
8	44144054	Science	4	New		1	1	3	2											
9	44144043	Science	4	New		1	1	3	2									FT	2014	7
10	44144046	Science	4	New		1	1	3	2									FT	2014	7

Page 2

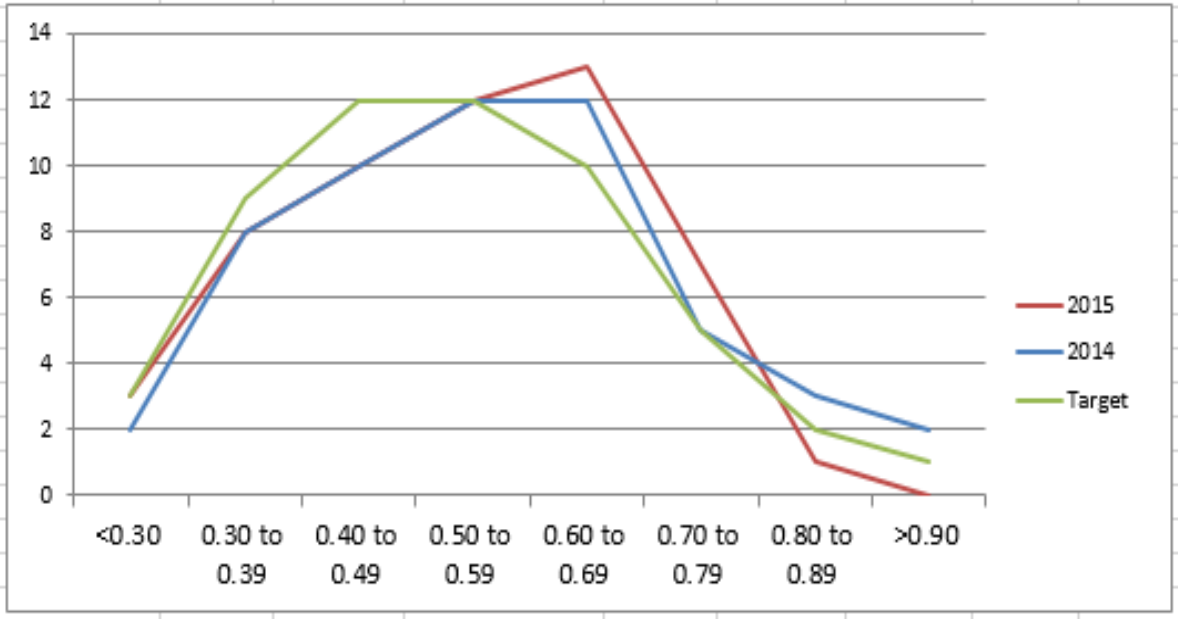
Row	N Count	Rasch	PVal	Flag PVal	PT Bis.	Flag PTBIS	Male vs Female Bias Flag	Non-Hispanic vs Hispanic Bias Flag	White vs Black Bias Flag	White vs Hispanic Bias Flag	White vs AmIn Bias Flag	White vs Asian Bias Flag	White vs Hawi/PaclsIr Bias Flag	White vs Multiracial Bias Flag	Dist A	Dist B	Dist C	Dist D	Omit
1	20638	1.116	0.46		0.24	*	A	A	A		A	A	A	A	38.3	45.8	9.5	6.4	0.0
2	20339	-0.411	0.76		0.50		A	A	A		A	A	A	A	76.1	7.5	7.4	8.8	0.1
3	20500	1.850	0.30	*	0.28	*	A	A	A		A	A	A	A	13.5	12.4	29.7	44.4	0.0
4	20340	1.759	0.33		0.21	*	A	A	A		A	A	A	A	15.6	34.2	16.9	33.3	0.0
5	20638	2.455	0.21	*	0.13	*	A	A	A		A	A	A	A	12.2	7.8	58.9	21.1	0.0
6	20339	1.205	0.45		0.26	*	A	A	A		A	A	A	A	15.2	14.7	25.1	44.9	0.1
7																			
8																			
9	20500	-0.283	0.72		0.44		A	A	A		A	A	A	A	4.9	5.4	72.4	17.3	0.0
10	20340	-0.268	0.74		0.52		A	A	A		A	A	A	A	74.0	14.8	4.5	6.6	0.0

Table 4.3
Sample Grade 4 Science Item Replacement Table

AZ AIMS Grade 4 Spring 15 Operational Item Replacement Plan for Science																	
# of Items Required per Blueprint	Strand	Concept	Spring 14 - New Operational Items Selections								Spring 15 - New Operational Items Selections							
			Actual # of Items	PO	AZID	Passg ID	P-VALUE	Rasch	PtBis	DOK	Actual # of Items	PO	AZID	Passg ID	P-VALUE	Rasch	PtBis	DOK
6	1	1	6	1.1.2	3514479	0	0.543	0.4563	0.395	2	6	1.1.1	3514444	0	0.399	1.3943	0.373	1
	1	1		1.1.3	3514583	0	0.62	0.5167	0.402	3		1.1.3	3514583	0	0.62	0.5167	0.402	3
	1	1		1.1.3	44114434	Electricity and Magnetism	0.674	0.1057	0.532	2		1.1.1	3514504	0	0.519	0.816	0.354	1
	1	1		1.1.1	3514504	0	0.519	0.816	0.354	1		1.1.3	44114434	Electricity and Magnetism	0.674	0.1057	0.532	2
	1	1		1.1.1	3514444	0	0.399	1.3943	0.373	1		1.1.2	44114447	Volcanoes	0.736	-0.2494	0.545	1
	1	1		1.1.1	3514592	0	0.494	0.9322	0.386	2		1.1.3	44104325	0	0.575	0.624	0.451	3

Table 4.4
Sample *P*-Value Target Table and Graph

		<0.30	0.30 to 0.39	0.40 to 0.49	0.50 to 0.59	0.60 to 0.69	0.70 to 0.79	0.80 to 0.89	>0.90	Total Number of Items
Pct of items for target		6%	17%	22%	22%	19%	9%	4%	2%	
Target Totals		**3**	**9**	**12**	**12**	**10**	**5**	**2**	**1**	**54**
Actual - Anchors		0	1	3	6	8	2	0	0	**20**
Actual - new selections		3	7	7	6	5	5	1	0	**34**
NRT/Dual Purpose		0	0	0	0	0	0	0	0	**0**
TOTAL	2015	3	8	10	12	13	7	1	0	**54**
Actual	2014	2	8	10	12	12	5	3	2	**54**



2015		
DOK	Target	Actual
Level	N	N
1	6	10
2	36	33
3	12	11
4	0	0
Total	54	54

DOK	2014
1	7
2	37
3	10
4	0
Total	54

Table 4.5
Number of Science Items Selected by Committee

Content Area	Grade	Total Items	Anchor Items		Total Selected	
Science	4	54	20	37%	34	73%
	8	58	24	41%	34	59%
	HS	65	20	31%	45	69%
Science Total		**177**	**61**	**34%**	**116**	**66%**

4.3 Customer Approvals

Approvals from ADE staff were obtained during several phases of development: during selection of the items, after forms were created, at the completion of the QA reviews, and when pre-press test books were available. Each is described below.

4.3.1 Item Selection Approval

ADE staff members were given the item replacement tables. Approval was verbal. The item selection tables were then reviewed by Pearson's research scientist. Psychometric evaluation of the test selection was the main focus of this review. Recommended changes were discussed with and approved by ADE.

4.3.2 Test Book Approvals

At the test book phase of development, items had been arranged into test book format. That is, they were no longer treated as individual items, but appeared in page layouts as they would appear in the final, printed test books. By this point, all content issues were resolved. The focus of this approval was on format and presentation issues, rather than on content issues. Formal approval was given. Desired changes were communicated via PDF markup and the Development Tracking Form, which included a description of the change, a justification, and space for the customer to grant or deny approval. Formal sign-off of test books by ADE was achieved via the use of signed electronic Final Proof Approval Forms.

4.3.3 FTP Site

A secure FTP site had been established by ADE for transfer of electronic documents (annotated test books, test book reviews, etc.) that need to be reviewed by ADE staff. After careful review by ADE staff, corrections and edits were transmitted to Pearson for inclusion/revision of the test documents.

4.3.4 Final Forms Review (Pearson)

The Final Forms review provided an opportunity for Pearson staff members who had not previously seen the test materials to review them. This review helped assure that test books, answer documents, and test administration directions all work in concert. In addition, this review helped in detecting errors, inconsistencies, cosmetic errors, and key verifications. Items with problems identified during the Final Forms review were annotated. Pearson staff resolved all

comments and made necessary corrections prior to releasing the materials.

4.3.5 ADE Quality Review

After Pearson reviewed and edited test documents, ADE staff conducted a final review of forms to determine if all edits had been accomplished properly.

4.3.6 Final Sign-off

A final, formal approval (blueline stage) was given as test books became available for printing. A copy of the test book was sent for ADE to review and to provide formal approval.

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 for all AIMS testing for the 2014-2015 school year. The following 1999 AERA/APA/NCME *Standards* (AERA, APA, NCME, 1999) are addressed: 1.13, 3.3, 3.19, 3.20, 3.21, 3.24, 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. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this part of the technical report are 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.

5.1 Accommodations

Accommodations were made available for all of the Fall 2014 and Spring 2015 AIMS tests, including AIMS Reading, Mathematics, and Writing in high school, and AIMS Science grades 4, 8, and high school. All of the AIMS tests allow some of the same accommodations but exclude others if there is evidence that the accommodation changes the construct that is being assessed. All statistics include students who have received accommodations.

Arizona statutes (A.R.S. §15-741 and §15-755), the Individuals with Disabilities Education Act (IDEA) (300.160), and the Elementary and Secondary Education Act (ESEA) (§1111) mandate that all students who are educated with public funds must participate in state assessment, including all students with disabilities and all students identified as English Language Learners.

For the purposes of assessment, a Special Education student is eligible to receive services under the Individuals with Disabilities Education Act and has an Individualized Education Program (IEP); and a 504 student is eligible under Section 504 of the Rehabilitation Act of 1973 and has a 504 Accommodation Plan.

Students with disabilities who have an IEP, or who have a 504 plan, may be considered for both universal test administration conditions and standard accommodations (described in section 5.1.1). Also, students identified as English Language Learner (ELL) and students who have been identified as Fluent English Proficient (FEP) for no more than two years may be considered for universal test administration conditions and standard accommodations.

Students with significant cognitive disabilities and whose current Individualized Education Program (IEP) designates them as eligible for an alternate assessment, AIMS A, are excluded from AIMS testing.

The Arizona English Language Learner Assessment (AZELLA), a language proficiency assessment, is given to determine a student's proficiency in English and respective instructional placement. An English Language Learner (ELL) is a student whose native language is other than English, who scores below the proficient level on the AZELLA, and is placed into a language program. Fluent English Proficient (FEP) is a term that is used to refer to a former ELL student who has scored at the proficient level of the AZELLA.

For detailed information on testing accommodations, please see *AIMS Testing Accommodations: Guidelines for School Year 2014-2016* on the Arizona Department of Education website at the following location: http://www.azed.gov/assessment/files/2014/08/testing-accommodations-2014-2016.pdf .

5.1.1 Overview of Accommodations

Accommodations are specific practices and procedures that provide students with equitable access during instruction and assessment. Accommodations are made in order to provide a student equal access to learning and equal opportunity to demonstrate what is known. They are intended to reduce or even eliminate the effects of a student's disability.

Accommodations can be changes in the presentation, response, setting, and timing/scheduling of educational activities. There should be a direct connection between a student's disability, special education need or language need and the accommodation(s) provided to the student during educational activities, including assessment.

Students should receive the same accommodations for classroom instruction, classroom assessments, district assessment, and state assessments. No accommodations should be provided during assessments that are not also provided during instruction. However, not all accommodations appropriate for instruction are appropriate for use during a standardized state assessment. The accommodations available to students while testing on AIMS assessments in high school and AIMS Science are limited to those listed in later sections of this document.

Accommodations may not provide verbal or other clues or suggestions that hint at or give away the correct response to the student. Therefore, it is not permissible to simplify, paraphrase, explain, or eliminate any test item, prompt, or multiple-choice option. Additionally, accommodations provided for one student may not impede or impact other students in the testing room. It is the responsibility of the testing administrator to see that each student, who qualifies for testing accommodations, receives appropriate accommodations while also ensuring that other students, who do not receive accommodations, are not affected.

5.1.2 Descriptions of Universal and Standard Accommodations

Arizona offers two levels of accommodations to students participating in state assessments: universal test administration conditions and standard accommodations.

Universal Test Administration Conditions are specific testing situations and conditions that may be offered to **any** student in order to provide him/her a comfortable and distraction-free testing environment. Universal test administration conditions may be included in a student's IEP or 504 plan as a required "accommodation"; however, for Arizona state testing purposes, these are not considered testing accommodations and are not limited to only students with IEPs or 504 plans.

Standard Accommodations are provisions made in how a student accesses and demonstrates learning that do not substantially change the instructional level, the content, or the performance criteria. For students with disabilities, standard accommodations are intended to reduce or even eliminate the effects of a student's disability. For English Language Learners and FEP Year 1 and Year 2 students, standard accommodations are intended to allow students the opportunity to demonstrate their content knowledge even though the student is not functioning at grade level in English.

During the assessment, all accommodations for assessment identified in a student's IEP or 504 plan must be made available. However, students may choose not to use the accommodation(s).

5.1.3 Determining if a Student Needs a Testing Accommodation

When students need accommodations in how they learn or demonstrate learning, they are likely to need accommodations in how they are assessed. Conversely, if students do not need accommodations in how they learn or demonstrate learning, they will not need accommodations in how they are assessed. Therefore, no accommodation can be put in place for an assessment that is not already used regularly in the classroom.

To determine if a student will need testing accommodations to participate in state assessments, the following questions were asked:

- Does the student use accommodations during daily instruction?
- If the student uses accommodations during daily instruction, does the student need accommodations in order to participate in the state assessment?
- If so, which testing accommodations are necessary and appropriate for the student?

It is important to annually re-consider the types of accommodations used for students, particularly as they gain more skills. The following is a list of the specific testing accommodations available to students while participating in a state assessment.

Universal Test Administration Conditions

- Testing in a small group, testing one-on-one, testing in a separate location or in a study carrel
- Being seated in a specific location within the testing room or being seated at special furniture
- Having the test administered by a familiar test administrator
- Using a special pencil or pencil grip
- Using devices that allow the student to see the test: glasses, contacts, magnification, special lighting, and color overlays
- Using devices that allow the student to hear the test directions: hearing aids and amplification
- Wearing noise buffers after the scripted directions have been read
- Having the scripted directions included in the *Test Administration Directions* repeated (at student request) and having questions about the scripted directions or the directions that students read on their own answered.

Standard Accommodations

Injury

For students who were eligible to receive a standard accommodation due to an injury.

1 = Have answers transferred from a test book into an answer document
2 = Record or dictate multiple-choice responses to a scribe (not available for writing)
3 = Use assistive technology with spell check, grammar check, and predict ahead functions turned off (not available for reading, mathematics, or science)

ELL/FEP

For students who were eligible to receive a standard accommodation due to their classification as an English Language Learner student or as a Fluent English Proficient (Year 1 or Year 2) student.

4 = More breaks and/or several shorter sessions
5 = Simplified language for the scripted directions in English
6 = Read aloud in English the writing prompt, mathematics test items, or science test items, as needed upon student request
7 = Provide a word-for-word published, paper translation dictionary
8 = Exact oral translation of the scripted directions or the directions that students read on their own as needed upon student request

IEP/504

For students who were eligible to receive a standard accommodation due to their IEP or 504 plan.

9 = Place marker used
10 = More breaks and/or several shorter sessions
11 = Test at a different time of day
12 = Simplify language for the scripted directions in English
13 = Read aloud or sign the directions that students read on their own
14 = Read aloud in English or sign the writing prompt, mathematics test items, or science test items
15 = Large print edition of test
16 = For a student who is blind, use of an abacus for mathematics test items
17 = For a student who is blind, use of an electronic dictionary and thesaurus with grammar check, spell check, encyclopedia, and internet access turned off (not available for reading, mathematics, or science)
18 = For student who is blind, Braille writers
19 = Have answers transferred from the test book into an answer document
20 = Record or dictate multiple-choice responses to a scribe (not available for writing)
21 = Use assistive technology with spell check, grammar check, and predict ahead functions turned off (not available for reading, mathematics, or science)
22 = For the mathematics sections, use of a personal whiteboard which can be seen by only the student and is erased after every problem (not available for reading, writing, or science)
Braille = use of a Braille edition of the test

5.1.4 Reporting Results of Assessments Taken with Accommodations

The use of standard accommodations results in scores that are considered valid for comparison and accountability purposes. Students who received standard accommodations on AIMS assessments in high school and AIMS Science will count as having tested for federal accountability (AYP) purposes. Their AIMS results will be included in aggregate results at the school, district, and state level on the paper reports provided by the testing contractor.

Students who receive standard testing accommodations while participating in AIMS assessments in high school and AIMS Science must have their accommodations appropriately identified on their

answer document as directed in the corresponding *Test Administration Directions*. It is not necessary to identify students who received universal test administration conditions while participating in the AIMS assessments in high school or AIMS Science assessments.

5.2 Test Security

All AIMS tests were administered under secure testing conditions. Figure 5.2.1 includes the security agreement signed by the superintendent/charter representative and district test coordinator involved with the testing administration. Figure 5.2.2 includes the security agreement signed by personnel involved with the testing administration.

District test coordinators are responsible for establishing and enforcing test security procedures that comply with the Test Security Agreement, the State Board of Education Rule regarding test security, and Test Security guidance provided in the Pre-Test Workshop package and included in the *AIMS Test Administration Directions.*

Figure 5.2.1
Spring 2015 AIMS Test security agreement for Superintendents/Charter Representatives and District Test Coordinators

AIMS HS and AIMS Science
School Years 2014-2015 Test Security Agreement
For Superintendents/Charter Representatives and District Test Coordinators

As Superintendent/Charter Representative or District Test Coordinator, I acknowledge that AIMS HS and AIMS Science are secure tests and agree to the following conditions of use to ensure the security of the tests.

1. Superintendents and Charter Representatives are responsible for all testing activities within their district/charter. Superintendents and Charter Representatives are allowed to designate a District Test Coordinator to act on their behalf.
   a. An accurate Test Coordinator Information Sheet for school year 2013-2014 must be on file with the Assessment Section of the Arizona Department of Education (ADE).
   b. If the designated District Test Coordinator is unable to attend a School Year 2014-2015 Pre-Test Workshop for AIMS HS and AIMS Science, the superintendent or charter representative is the only substitute permitted to attend in his/her place.
2. All necessary security precautions shall be in place to safeguard test materials.
   a. Access to test books and answer documents shall be restricted.
   b. The names of all persons having access to the test books and answer documents shall be kept on file by the designated district test coordinator.
   c. All persons having access to the AIMS HS and AIMS Science test materials, other than students to whom the tests are administered, shall sign a School Year 2014-2015 test security agreement. Signed test security agreements shall be kept on file for 6 years.
      i. Building administrators shall maintain the agreements signed by building staff.
      ii. Superintendents/charter representatives shall maintain the agreements signed by building administrators.
      iii. The Assessment Section of ADE shall maintain the agreements signed by superintendents and charter representatives.
   d. All test books and answer documents shall be kept under lock and key except during actual test times.
      i. Test books and answer documents shall be delivered to test administrators no sooner than the date of testing.
      ii. Test books and answer documents shall be kept secure until they are distributed to students.
      iii. Students shall not be permitted to remove test material from the testing room except under supervision of staff.
   e. The AIMS HS and AIMS Science tests shall not be examined, read, or reviewed.
      i. No content of the test shall be disclosed nor allowed to be disclosed.
      ii. No test item shall be discussed at any time.
      iii. No student responses shall be examined, read, or reviewed.
   f. Upon completion of testing, all AIMS HS and AIMS Science test materials shall be returned to the designated district test coordinator.
3. The district superintendent or charter representative shall develop, distribute, and enforce disciplinary procedures for the violation of test security by staff.
4. *Test Preparation and Administration Practices,* guidelines approved by the State Board of Education in January 2003 and updated December 2007, shall be followed.
5. **All instructions in the *Test Coordinator's Manual* and the *Test Administration Directions*, which include reading the directions to students exactly as scripted in the *Test Administration Directions*, shall be followed.**

By signing my name to this document, I am assuring the Arizona Department of Education that I will abide by the above conditions and that anyone I supervise, who will have access to the AIMS HS and Science tests for School Year 2014-2015, will also sign a Test Security Agreement.

Superintendent/Charter Representative Signature: _____

Printed Name: _____ Title: _____

District Test Coordinator Signature: _____

Printed Name: _____

District/Charter: _____ Entity #: _____

Address: _____

City, State, Zip: _____

Fax: 602-542-5467 or Email: marypat.wood@azed.gov

Figure 5.2.2
Spring 2015 AIMS Test security agreement for all school/district/charter personnel

AIMS HS and AIMS Science Tests
School Year 2014-2015 Test Security Agreement

I acknowledge that AIMS HS and AIMS Science are secure tests and agree to the following conditions of use to ensure the security of the test.

1. I shall take necessary precautions to safeguard test materials.
   a. I shall sign a School Year 2014-2015 test security agreement.
   b. Access to test books and answer documents is restricted. I shall not attempt to gain access to test materials beyond that which is granted to me by my school/district test coordinator, superintendent, or charter representative.
   c. If test books and answer documents are distributed to me, I shall keep them under lock and key except during actual test times.
   d. I shall not permit students to remove test material from the testing room except under the supervision of staff.
   e. I shall not examine, read, or review the AIMS HS and AIMS Science tests.
      i. I shall not disclose, nor allow to be disclosed, the content of the test.
      ii. I shall not discuss any test item at any time.
      iii. I shall not examine, read, or review any student responses.
   f. If test books and answer documents are distributed to me, I shall return all AIMS HS and AIMS Science test materials to the school/district test coordinator immediately upon the completion of testing.
2. I understand that the district superintendent or charter representative will develop, distribute, and enforce disciplinary procedures for the violation of test security by staff.

Individuals who will administer or proctor AIMS HS and AIMS Science for School Year 2014-2015 must also agree to the following conditions to ensure the correct administration of the tests.

3. I shall participate in training activities prior to administering the tests.
4. I shall follow *Test Preparation and Administration Practices*, the guidelines approved by the State Board of Education in January 2003 and updated in December 2007.
5. I shall review the ***Test Administration Directions*** prior to administering the test.
6. I shall follow all instructions in the ***Test Administration Directions*** including **reading the directions to students exactly as scripted**.

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 AIMS HS and AIMS Science tests, will also sign a Test Security Agreement.

Signed By: _____

Printed Name: _____

Title: _____ School: _____

Please return signed copy as per instructions from your school/district test coordinator.
Signed copies will be maintained by school/district administrators for 6 years.

5.3 Test Administration

In order to ensure a standardized testing administration for all students, a *Test Coordinator's Manual* was made available to all test coordinators for the fall 2014 and spring 2015 administrations. The manual included the following topics:

- Responsibilities of the District Test Coordinator
  - Before Testing
  - During Testing
  - After Testing
- Procedures for Test Administration
  - Students to Be Tested
  - Test Administration Schedules
  - Required Test Materials
  - Test Security
  - Student Identification Information
  - Arrangements Prior to Test Administration
- Procedures for Handling Test Materials (before, during, and after testing)
  - Receiving Test Materials
  - Inventorying Test Materials
  - Precautions
  - Inspecting and Organizing Test Materials
  - Assembling Scorable Test Materials
  - Assembling Nonscorable Test Materials
  - Materials Retrieval
- State Board of Education Rule
- Important Dates for spring 2015 Testing

Test Administration Directions were made available to all test administrators for the fall 2014 and spring 2015 assessments. The *Test Administration Directions* included the following topics:

- Overview for the Administration of AIMS
  - Test Administrator Responsibilities
  - Students to Be Tested
  - Test Administration Schedule
  - Test Materials
  - Precautions
- Before Testing Guidelines
  - Training and Test Security
  - Preparing the Room for Testing
- During Testing Guidelines
  - Reading the Scripted Directions

  - Student Identification Information
  - Monitoring Testing
  - Use of Resources
  - Use of Unacceptable Resources
  - Disruptive Students and Students Who Leave During Testing
  - Detailed Scripts for Administration of Each Part of Each Test
- After Testing
  - Inspecting Test Materials
  - Completing Student Identification Information
  - Transferring Student Responses
  - Returning Test Materials to the Test Coordinator

For specific information related to test administration, refer to the *Test Coordinator's Manual* and/or the *Test Administration Directions*.

Pre-Test Workshops were conducted online prior for each test administration, fall and spring. Every district test coordinator is required to view a 3-session online Pre-Test Workshop. The Pre-Test Workshop encompasses training related to test administration which includes test security, accommodations, test coordinator responsibility, and test schedule. Materials handling is included in these online workshops, covering ordering, receiving, preparing for retrieval, and the retrieval of test materials.

PART 6: CLASSICAL ITEM ANALYSIS

Part 6 presents classical test statistics and item analysis statistics for each content area and grade level computed from the data used for calibration and scaling. The NRT components of AIMS tests were no longer administered in spring 2015. Addressed in this part of the technical report are the following 1999 AERA/APA/NCME *Standards*: 1.5, 1.13, 2.4, 2.8, 3.18, 6.5, and 7.1. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this chapter are: 1.8, 1.10, 2.19, 3.6, 4.14, and 7.4. (See Appendix A for this information for the Fall 2014 AIMS administration.)

6.1 Data

Arizona had two test windows for operational testing in spring 2015. The Writing and Reading tests for high school were administered on February 23 and 24, respectively. The AIMS Mathematics tests for high school were administered on February 25. The AIMS Science tests for grade 4 and 8, and high school were administered between March 16 and April 24.

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 calibration samples in reading, mathematics, science, and writing. The table shows the number of students (N), the maximum obtained raw score (Max RS), the raw score mean (RS M), the raw score standard deviation (RS SD), the average *p*-value (P-Value M), the average item-to-total correlation (rpb M) and the estimate of internal consistency. Cronbach's alpha is the measure of internal consistency used for the AIMS Reading, Mathematics and Science tests. Stratified alpha is used to estimate the internal consistence reliability of the AIMS Writing tests. The item-to-total correlation is computed as a point biserial correlation for dichotomous items and as a Pearson product-moment correlation for polytomous items. The point biserial correlation reported is the correlation of the item scores and the total test score. The Pearson product-moment correlation reported is the correlation of the polytomous item and the total test score.

Table 6.2.1
Spring 2015 AIMS Classical Test Analysis Statistics

	Content	Grade	Prompt	N	Max RS Obtained	RS M	RS SD	P-value M	rpb M	Internal Consistency
CRT										
	Science									
		4		84113	54	29.95	9.94	0.55	0.35	0.89
		8		82248	58	33.17	10.72	0.57	0.36	0.90
		HS		80038	65	32.38	11.81	0.50	0.34	0.91

Note: CRT = Criterion-referenced test.

6.3 Classical Item Analysis

Classical item analysis was conducted for all Science grades. The classical item analysis statistics for Spring 2015 AIMS Reading, Mathematics, and Writing in high school were not generated since the tests were pre-equated for this administration. Please refer to the AIMS 2014 technical report (Arizona Department of Education, 2014) for the classical item analysis statistics in the high school reading and mathemetics since the spring 2015 tests were reused forms of the spring 2014 tests. Please also refer to the 2012 AIMS Technical Report (Arizona Department of Education, 2012) and 2013 AIMS Technical Report (Arizona Department of Education, 2013) for the classical item analysis statistics for the Writing multiple-choice items and an extended-response item in high school, respectively since the multiple-choice items and extended-response item were reused items from the spring 2012 and spring 2013 administrations, respectively. Tables 6.3.1—6.3.3 present item statistics for the spring science tests. The tables show the number of students (N), the item difficulty (P-Value), point biserial correlation (rpb) and biserial correlation (rbi) for dichotomous items, item-to-total Pearson product-moment correlation (r) for polytomous items, percentage of students who omitted the item (% Omit), and the percentage of students responding to and point biserial for each response option. The keyed response has a percent responding that matches the *p*-value and a positive point biserial correlation while the incorrect response options have a negative point biserial correlation. The item-to-total Pearson product-moment correlation reported is the Pearson product moment correlation of the item and total score on the test. The point biserial correlation (rpb) reported is the correlation between student performance on an item and the total score on a test. The biserial correlation (rbi) is an adjusted point-biserial correlation intended to estimate the value of a Pearson correlation between the item and total score as if the item scores were normally distributed rather than binary.

Table 6.3.1
Spring 2015 AIMS Classical Item Analysis
Science Grade 4

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
1	83989	0.79	0.25	0.36	0.03	4.31	-0.13	7.21	-0.18	9.28	-0.10	79.17	0.25
2	83989	0.71	0.29	0.38	0.07	71.07	0.29	13.26	-0.13	7.12	-0.21	8.48	-0.12
3	83989	0.76	0.36	0.49	0.12	8.43	-0.23	4.16	-0.19	11.16	-0.16	76.13	0.36
4	83989	0.59	0.43	0.54	0.10	25.39	-0.27	58.86	0.43	9.07	-0.18	6.57	-0.18
5	83989	0.67	0.36	0.45	0.13	10.60	-0.25	6.79	-0.17	66.54	0.36	15.94	-0.13
6	83989	0.50	0.30	0.37	0.10	18.17	-0.05	50.30	0.30	24.38	-0.14	7.04	-0.27
7	83989	0.63	0.35	0.45	0.09	13.63	-0.21	10.45	-0.07	62.83	0.35	13.00	-0.23
8	83989	0.63	0.40	0.51	0.08	9.46	-0.25	7.80	-0.22	63.21	0.40	19.45	-0.15
9	83989	0.46	0.20	0.24	0.10	28.04	0.01	8.26	-0.18	45.86	0.20	17.74	-0.14
10	83989	0.73	0.41	0.54	0.09	73.28	0.41	8.55	-0.16	6.58	-0.27	11.51	-0.23
11	83989	0.49	0.33	0.41	0.06	12.68	-0.21	49.28	0.33	23.02	-0.11	14.96	-0.14
12	83989	0.34	0.25	0.32	0.08	12.15	-0.16	33.64	0.25	4.11	-0.14	50.03	-0.07
13	83989	0.42	0.31	0.39	0.09	23.36	-0.09	25.00	-0.13	9.22	-0.21	42.34	0.31
14	83989	0.47	0.26	0.33	0.11	14.39	-0.22	46.80	0.26	14.74	-0.12	23.97	-0.02
15	83989	0.48	0.22	0.27	0.09	15.35	-0.05	25.11	-0.05	11.72	-0.22	47.73	0.22
16	83989	0.60	0.44	0.55	0.09	19.78	-0.20	11.74	-0.24	59.66	0.44	8.72	-0.21
17	83989	0.71	0.30	0.40	0.13	8.96	-0.13	10.76	-0.12	9.21	-0.21	70.94	0.30
18	83989	0.62	0.32	0.40	0.13	9.13	-0.23	61.53	0.32	9.80	-0.23	19.42	-0.05
19	83989	0.32	0.22	0.28	0.15	31.22	-0.08	17.27	-0.05	19.47	-0.12	31.89	0.22
20	83989	0.52	0.45	0.55	0.18	20.94	-0.30	13.00	-0.21	51.67	0.45	14.21	-0.09
21	83989	0.73	0.43	0.56	0.21	12.37	-0.18	6.41	-0.26	8.26	-0.25	72.75	0.43
22	83989	0.66	0.43	0.55	0.20	12.96	-0.24	65.83	0.43	15.92	-0.22	5.09	-0.21
23	83989	0.53	0.34	0.42	0.14	11.46	-0.14	15.51	-0.12	19.83	-0.19	53.05	0.34
24	83989	0.42	0.29	0.35	0.86	18.21	-0.12	20.73	-0.08	42.11	0.29	18.10	-0.17
25	83989	0.76	0.49	0.66	0.16	76.10	0.49	9.22	-0.25	5.20	-0.26	9.32	-0.27
26	83989	0.54	0.41	0.50	0.20	54.23	0.41	13.82	-0.12	9.28	-0.27	22.47	-0.20
27	83989	0.35	0.26	0.33	0.29	22.84	-0.10	15.25	-0.13	35.08	0.26	26.54	-0.08
28	83989	0.63	0.30	0.37	0.03	2.79	-0.19	5.31	-0.22	29.32	-0.14	62.55	0.30
29	83989	0.71	0.51	0.66	0.08	9.80	-0.29	12.20	-0.24	6.61	-0.26	71.31	0.51
30	83989	0.46	0.29	0.36	0.16	8.27	-0.23	6.21	-0.20	38.97	-0.06	46.39	0.29
31	83989	0.31	0.17	0.22	0.13	29.42	-0.05	27.44	-0.03	31.11	0.17	11.91	-0.13
32	83989	0.39	0.35	0.44	0.15	16.74	-0.21	32.33	-0.10	38.71	0.35	12.07	-0.14
33	83989	0.44	0.43	0.52	0.14	9.91	-0.15	43.97	0.43	40.65	-0.24	5.34	-0.22
34	83989	0.63	0.45	0.56	0.15	63.31	0.45	10.10	-0.25	19.42	-0.20	7.02	-0.24
35	83989	0.62	0.43	0.54	0.14	8.51	-0.28	7.60	-0.24	21.38	-0.16	62.36	0.43
36	83989	0.61	0.24	0.30	0.17	16.72	-0.02	60.74	0.24	5.95	-0.17	16.42	-0.17
37	83989	0.35	0.20	0.26	0.14	34.98	0.20	40.60	-0.06	6.68	-0.21	17.60	-0.04
38	83989	0.58	0.37	0.46	0.19	5.70	-0.24	58.13	0.37	27.17	-0.12	8.81	-0.27
39	83989	0.42	0.32	0.40	0.06	41.91	0.32	19.50	-0.13	15.09	-0.16	23.44	-0.12
40	83989	0.74	0.30	0.41	0.11	8.02	-0.11	12.02	-0.19	74.19	0.30	5.66	-0.18

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

(table continues)

Table 6.3.1 (continued)
Spring 2015 AIMS Classical Item Analysis
Science Grade 4 (continued)

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
41	83989	0.32	0.29	0.38	0.20	16.41	-0.21	20.19	-0.11	32.23	0.29	30.97	-0.04
42	83989	0.64	0.36	0.45	0.13	8.69	-0.23	9.46	-0.18	18.03	-0.14	63.69	0.36
43	83989	0.55	0.37	0.46	0.14	8.65	-0.27	6.80	-0.18	29.56	-0.14	54.85	0.37
44	83989	0.24	0.20	0.27	0.16	24.44	0.20	12.99	-0.06	13.80	-0.18	48.60	-0.01
45	83989	0.39	0.27	0.34	0.20	32.18	-0.03	38.65	0.27	11.44	-0.25	17.53	-0.10
46	83989	0.81	0.45	0.63	0.15	6.86	-0.26	80.89	0.45	6.60	-0.25	5.50	-0.21
47	83989	0.49	0.42	0.51	0.17	24.35	-0.18	18.67	-0.19	8.05	-0.21	48.76	0.42
48	83989	0.59	0.40	0.49	0.20	58.97	0.40	9.48	-0.26	11.30	-0.21	20.05	-0.13
49	83989	0.71	0.52	0.67	0.22	10.28	-0.29	7.96	-0.28	10.26	-0.24	71.29	0.52
50	83989	0.55	0.39	0.48	0.44	16.71	-0.08	54.60	0.39	15.94	-0.24	12.31	-0.23
51	83989	0.53	0.40	0.49	0.45	21.92	-0.12	14.71	-0.21	10.33	-0.25	52.58	0.40
52	83989	0.57	0.45	0.55	0.60	19.46	-0.21	9.80	-0.23	56.87	0.45	13.27	-0.21
53	83989	0.58	0.23	0.29	0.82	9.10	-0.10	58.20	0.23	10.92	-0.15	20.96	-0.09
54	83989	0.67	0.42	0.53	1.02	67.29	0.42	8.67	-0.23	8.36	-0.23	14.67	-0.20

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

Table 6.3.2
Spring 2015 AIMS Classical Item Analysis
Science Grade 8

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
1	82141	0.84	0.23	0.34	0.03	9.90	-0.15	83.51	0.23	2.83	-0.13	3.74	-0.11
2	82141	0.73	0.34	0.45	0.07	73.37	0.34	6.72	-0.15	12.01	-0.23	7.83	-0.14
3	82141	0.64	0.24	0.30	0.04	9.14	-0.25	11.07	-0.02	16.24	-0.10	63.51	0.24
4	82141	0.62	0.46	0.58	0.05	61.84	0.46	18.23	-0.19	13.04	-0.29	6.84	-0.21
5	82141	0.79	0.39	0.54	0.05	2.66	-0.18	2.61	-0.19	78.70	0.39	15.98	-0.27
6	82141	0.61	0.18	0.22	0.06	9.06	-0.10	4.34	-0.21	25.14	-0.03	61.41	0.18
7	82141	0.78	0.39	0.54	0.05	5.15	-0.25	5.94	-0.20	10.80	-0.19	78.06	0.39
8	82141	0.88	0.38	0.61	0.05	2.51	-0.17	7.05	-0.27	87.80	0.38	2.59	-0.18
9	82141	0.63	0.25	0.31	0.09	16.27	-0.08	11.12	-0.22	62.63	0.25	9.90	-0.08
10	82141	0.62	0.34	0.43	0.05	17.15	-0.12	9.62	-0.19	62.31	0.34	10.87	-0.21
11	82141	0.68	0.31	0.39	0.01	9.77	-0.18	67.54	0.31	11.17	-0.10	11.51	-0.18
12	82141	0.69	0.39	0.51	0.06	16.12	-0.16	4.70	-0.22	9.79	-0.26	69.33	0.39
13	82141	0.76	0.31	0.42	0.04	12.61	-0.13	76.45	0.31	6.62	-0.22	4.28	-0.17
14	82141	0.44	0.23	0.29	0.07	8.03	-0.16	29.94	-0.07	43.51	0.23	18.45	-0.10
15	82141	0.48	0.34	0.42	0.09	29.43	-0.08	12.73	-0.17	10.00	-0.25	47.75	0.34
16	82141	0.65	0.39	0.50	0.07	11.75	-0.19	8.64	-0.20	64.96	0.39	14.59	-0.19
17	82141	0.61	0.41	0.51	0.06	18.42	-0.11	6.33	-0.24	14.36	-0.27	60.83	0.41
18	82141	0.29	0.14	0.19	0.14	22.22	-0.05	23.13	-0.08	25.58	-0.02	28.93	0.14
19	82141	0.72	0.41	0.54	0.07	9.52	-0.18	7.53	-0.25	72.16	0.41	10.71	-0.21
20	82141	0.30	0.35	0.45	0.09	29.93	0.35	17.86	-0.09	27.13	-0.21	24.99	-0.07
21	82141	0.43	0.36	0.44	0.12	23.76	-0.12	12.02	-0.21	42.50	0.36	21.61	-0.14
22	82141	0.48	0.30	0.37	0.10	6.16	-0.21	35.87	-0.06	9.55	-0.24	48.31	0.30
23	82141	0.51	0.43	0.53	0.15	17.24	-0.21	51.18	0.43	26.66	-0.20	4.78	-0.22
24	82141	0.35	0.38	0.48	0.15	4.37	-0.15	27.89	-0.04	32.42	-0.29	35.17	0.38
25	82141	0.67	0.37	0.47	0.05	18.06	-0.15	66.83	0.37	9.78	-0.25	5.28	-0.19
26	82141	0.42	0.34	0.42	0.09	29.35	-0.09	42.18	0.34	12.99	-0.20	15.39	-0.17
27	82141	0.41	0.44	0.55	0.13	24.11	-0.21	16.45	-0.25	41.03	0.44	18.28	-0.09
28	82141	0.43	0.33	0.41	0.09	20.15	-0.10	43.45	0.33	17.40	-0.21	18.90	-0.11
29	82141	0.48	0.41	0.51	0.09	14.39	-0.18	47.68	0.41	16.09	-0.21	21.75	-0.16
30	82141	0.70	0.48	0.63	0.05	11.11	-0.29	6.48	-0.25	11.88	-0.21	70.49	0.48
31	82141	0.51	0.36	0.44	0.09	19.97	-0.26	51.22	0.36	18.05	-0.10	10.68	-0.12
32	82141	0.81	0.44	0.63	0.04	4.24	-0.24	81.11	0.44	9.56	-0.27	5.04	-0.21
33	82141	0.50	0.21	0.26	0.09	8.41	-0.20	10.37	-0.17	50.15	0.21	30.97	0.00
34	82141	0.70	0.45	0.58	0.07	5.85	-0.22	11.26	-0.23	69.98	0.45	12.84	-0.24
35	82141	0.43	0.35	0.44	0.17	43.29	0.35	9.29	-0.17	8.27	-0.19	38.97	-0.15
36	82141	0.49	0.29	0.36	0.06	25.49	-0.03	16.23	-0.22	48.71	0.29	9.51	-0.17
37	82141	0.22	0.15	0.20	0.07	22.28	0.15	13.48	-0.16	16.25	-0.07	47.93	0.04
38	82141	0.57	0.34	0.42	0.14	9.56	-0.26	7.45	-0.28	25.41	-0.04	57.44	0.34
39	82141	0.49	0.35	0.43	0.22	11.19	-0.25	48.55	0.35	22.05	-0.09	17.99	-0.15
40	82141	0.60	0.46	0.58	0.28	12.64	-0.23	23.63	-0.27	59.69	0.46	3.77	-0.18

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

(table continues)

Table 6.3.2 (continued)
Spring 2015 AIMS Classical Item Analysis
Science Grade 8

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
41	82141	0.54	0.42	0.52	0.03	31.82	-0.23	53.84	0.42	6.45	-0.26	7.85	-0.16
42	82141	0.73	0.43	0.57	0.04	4.40	-0.24	13.78	-0.22	73.18	0.43	8.61	-0.23
43	82141	0.51	0.39	0.48	0.09	13.00	-0.26	10.59	-0.27	51.16	0.39	25.15	-0.05
44	82141	0.79	0.51	0.70	0.09	6.79	-0.27	7.70	-0.28	6.31	-0.26	79.11	0.51
45	82141	0.47	0.23	0.29	0.09	39.23	-0.01	6.47	-0.22	47.12	0.23	7.09	-0.22
46	82141	0.52	0.40	0.49	0.06	7.76	-0.25	51.85	0.40	7.74	-0.26	32.59	-0.14
47	82141	0.78	0.49	0.68	0.10	77.70	0.49	8.68	-0.29	7.71	-0.25	5.81	-0.25
48	82141	0.60	0.49	0.61	0.07	16.65	-0.28	8.56	-0.21	14.26	-0.22	60.46	0.49
49	82141	0.59	0.33	0.41	0.13	20.76	-0.02	10.99	-0.28	59.21	0.33	8.92	-0.23
50	82141	0.51	0.50	0.61	0.07	51.45	0.50	21.66	-0.20	22.50	-0.30	4.32	-0.19
51	82141	0.52	0.33	0.41	0.13	9.60	-0.27	26.37	-0.05	51.88	0.33	12.02	-0.19
52	82141	0.56	0.40	0.49	0.17	9.96	-0.25	10.32	-0.28	55.63	0.40	23.92	-0.09
53	82141	0.79	0.41	0.58	0.16	5.30	-0.21	79.36	0.41	6.98	-0.21	8.20	-0.24
54	82141	0.61	0.45	0.57	0.31	9.02	-0.28	10.85	-0.26	18.37	-0.16	61.44	0.45
55	82141	0.31	0.27	0.35	0.05	43.56	-0.10	30.62	0.27	13.26	-0.09	12.51	-0.14
56	82141	0.34	0.28	0.36	0.09	25.38	0.00	15.44	-0.15	25.23	-0.18	33.86	0.28
57	82141	0.53	0.23	0.29	0.12	12.64	-0.19	52.95	0.23	25.54	0.01	8.75	-0.20
58	82141	0.51	0.34	0.42	0.11	51.40	0.34	10.04	-0.25	28.25	-0.04	10.20	-0.26

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

Table 6.3.3
Spring 2015 AIMS Classical Item Analysis
Science Grade 10

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
1	79966	0.55	0.30	0.37	0.19	26.31	-0.10	9.75	-0.19	9.10	-0.16	54.65	0.30
2	79966	0.57	0.48	0.60	0.04	5.85	-0.06	27.60	-0.39	9.11	-0.17	57.40	0.48
3	79966	0.45	0.32	0.39	0.10	34.10	-0.09	13.99	-0.22	7.28	-0.15	44.53	0.32
4	79966	0.42	0.21	0.27	0.19	14.67	-0.11	29.66	-0.03	41.68	0.21	13.80	-0.16
5	79966	0.39	0.17	0.22	0.14	15.01	-0.15	37.85	0.02	8.06	-0.16	38.94	0.17
6	79966	0.39	0.26	0.33	0.13	6.38	-0.15	39.17	0.26	3.10	-0.14	51.23	-0.13
7	79966	0.27	0.26	0.34	0.10	42.52	-0.15	27.48	0.26	7.59	-0.16	22.31	0.00
8	79966	0.59	0.32	0.40	0.10	7.92	-0.09	13.18	-0.15	20.09	-0.21	58.70	0.32
9	79966	0.66	0.32	0.40	0.12	19.09	-0.12	5.86	-0.20	65.53	0.32	9.40	-0.20
10	79966	0.57	0.39	0.48	0.14	21.64	-0.18	5.09	-0.17	57.30	0.39	15.82	-0.22
11	79966	0.57	0.33	0.41	0.06	14.32	-0.14	56.59	0.33	8.74	-0.22	20.29	-0.13
12	79966	0.30	0.25	0.33	0.16	12.39	-0.09	22.88	0.03	34.24	-0.21	30.34	0.25
13	79966	0.74	0.40	0.53	0.04	12.96	-0.22	74.39	0.40	3.02	-0.17	9.58	-0.23
14	79966	0.56	0.17	0.22	0.09	8.18	-0.19	29.43	0.03	6.23	-0.19	56.07	0.17
15	79966	0.75	0.26	0.35	0.07	74.80	0.26	14.57	-0.10	8.50	-0.19	2.05	-0.17
16	79966	0.65	0.36	0.46	0.08	65.10	0.36	5.21	-0.18	23.75	-0.20	5.85	-0.21
17	79966	0.53	0.38	0.47	0.07	6.45	-0.16	52.93	0.38	16.84	-0.29	23.72	-0.10
18	79966	0.52	0.46	0.57	0.11	12.55	-0.22	19.67	-0.26	15.46	-0.16	52.21	0.46
19	79966	0.41	0.36	0.45	0.13	20.31	-0.13	14.72	-0.16	41.02	0.36	23.82	-0.16
20	79966	0.33	0.26	0.34	0.21	27.67	-0.15	32.62	0.26	24.50	-0.03	15.00	-0.12
21	79966	0.84	0.37	0.57	0.08	4.49	-0.21	5.65	-0.20	84.41	0.37	5.37	-0.20
22	79966	0.52	0.29	0.36	0.15	19.28	-0.10	20.06	-0.21	51.86	0.29	8.64	-0.08
23	79966	0.69	0.46	0.59	0.11	14.13	-0.25	6.99	-0.26	69.49	0.46	9.27	-0.20
24	79966	0.76	0.37	0.50	0.08	11.69	-0.24	4.33	-0.17	7.99	-0.17	75.92	0.37
25	79966	0.56	0.35	0.44	0.06	11.99	-0.16	10.36	-0.24	21.65	-0.12	55.94	0.35
26	79966	0.39	0.36	0.45	0.08	35.37	-0.12	39.24	0.36	18.94	-0.20	6.37	-0.16
27	79966	0.30	0.26	0.34	0.13	30.32	0.26	14.52	-0.24	16.55	-0.28	38.47	0.14
28	79966	0.55	0.33	0.41	0.14	18.48	-0.14	11.43	-0.14	14.94	-0.18	55.01	0.33
29	79966	0.47	0.36	0.45	0.15	12.88	-0.18	46.86	0.36	29.05	-0.14	11.06	-0.18
30	79966	0.51	0.37	0.45	0.13	10.83	-0.21	24.41	-0.12	14.00	-0.19	50.63	0.37
31	79966	0.48	0.36	0.45	0.09	4.86	-0.15	31.17	-0.27	47.55	0.36	16.32	-0.06
32	79966	0.50	0.33	0.41	0.12	33.65	-0.14	9.89	-0.19	50.07	0.33	6.28	-0.18
33	79966	0.46	0.37	0.46	0.18	46.24	0.37	22.23	-0.15	18.62	-0.20	12.73	-0.14
34	79966	0.29	0.28	0.37	0.14	24.96	-0.01	28.04	-0.15	18.27	-0.15	28.59	0.28
35	79966	0.63	0.48	0.61	0.13	6.56	-0.23	15.66	-0.28	14.18	-0.21	63.47	0.48
36	79966	0.76	0.43	0.59	0.11	13.88	-0.30	76.39	0.43	4.96	-0.21	4.67	-0.16
37	79966	0.38	0.24	0.31	0.18	35.44	-0.07	17.07	-0.16	37.58	0.24	9.73	-0.09
38	79966	0.28	0.24	0.31	0.15	31.44	0.01	28.28	0.24	26.32	-0.08	13.81	-0.22
39	79966	0.56	0.41	0.50	0.07	16.60	-0.19	14.61	-0.25	56.35	0.41	12.37	-0.13
40	79966	0.40	0.32	0.40	0.13	28.51	-0.03	14.04	-0.23	39.72	0.32	17.61	-0.16

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

(table continues)

Table 6.3.3 (continued)
Spring 2015 AIMS Classical Item Analysis
Science Grade 10

Item	N	P-Value	rpb	rbi	% Omit	Option A		Option B		Option C		Option D	
						%	rpb	%	rpb	%	rpb	%	rpb
41	79966	0.60	0.43	0.53	0.16	11.16	-0.20	15.53	-0.25	59.87	0.43	13.28	-0.16
42	79966	0.51	0.41	0.51	0.13	9.16	-0.22	13.15	-0.20	26.65	-0.17	50.91	0.41
43	79966	0.44	0.34	0.42	0.15	17.57	-0.15	43.82	0.34	16.26	-0.17	22.20	-0.11
44	79966	0.40	0.29	0.36	0.12	14.53	-0.12	32.75	-0.05	12.25	-0.24	40.35	0.29
45	79966	0.29	0.29	0.39	0.16	33.59	0.10	11.16	-0.22	26.27	-0.25	28.82	0.29
46	79966	0.54	0.37	0.46	0.12	22.88	-0.15	17.43	-0.20	54.10	0.37	5.47	-0.19
47	79966	0.49	0.47	0.58	0.13	49.15	0.47	9.25	-0.17	20.35	-0.27	21.11	-0.19
48	79966	0.28	0.22	0.29	0.21	14.98	-0.21	22.34	-0.14	27.68	0.22	34.79	0.07
49	79966	0.59	0.35	0.44	0.15	15.24	-0.15	59.18	0.35	15.13	-0.23	10.30	-0.12
50	79966	0.61	0.46	0.57	0.12	61.02	0.46	7.81	-0.23	20.57	-0.28	10.48	-0.16
51	79966	0.57	0.40	0.50	0.16	14.22	-0.15	14.56	-0.23	56.86	0.40	14.20	-0.19
52	79966	0.41	0.28	0.34	0.23	14.08	-0.09	21.27	-0.06	40.98	0.28	23.44	-0.19
53	79966	0.51	0.37	0.46	0.08	6.38	-0.25	32.18	-0.12	10.85	-0.21	50.51	0.37
54	79966	0.58	0.53	0.66	0.11	57.57	0.53	9.47	-0.25	26.38	-0.34	6.47	-0.17
55	79966	0.50	0.33	0.41	0.14	14.05	-0.23	7.31	-0.25	50.28	0.33	28.22	-0.05
56	79966	0.47	0.34	0.42	0.15	46.55	0.34	15.25	-0.24	10.31	-0.26	27.74	-0.01
57	79966	0.45	0.35	0.43	0.16	45.24	0.35	24.60	-0.07	20.90	-0.18	9.09	-0.22
58	79966	0.40	0.41	0.51	0.17	39.93	0.41	12.94	-0.22	20.06	-0.15	26.90	-0.14
59	79966	0.32	0.27	0.35	0.18	42.35	-0.10	17.23	-0.11	32.20	0.27	8.04	-0.12
60	79966	0.37	0.14	0.18	0.30	9.52	-0.18	37.31	0.14	31.29	-0.07	21.58	0.05
61	79966	0.68	0.40	0.52	0.18	11.78	-0.23	68.32	0.40	13.96	-0.21	5.76	-0.17
62	79966	0.46	0.30	0.38	0.17	45.68	0.30	25.36	-0.12	9.90	-0.21	18.89	-0.10
63	79966	0.46	0.38	0.47	0.18	45.93	0.38	16.69	-0.20	25.35	-0.15	11.85	-0.15
64	79966	0.42	0.34	0.42	0.19	41.57	0.34	24.43	-0.18	20.87	-0.11	12.94	-0.13
65	79966	0.49	0.39	0.48	0.18	20.53	-0.15	48.81	0.39	17.28	-0.19	13.21	-0.19

Note. Item number is not the item number in test booklet due to the nature of the dual purpose assessment and embedded field test items. This test included multiple-choice items only. The statistics presented in this table are based on a calibration sample, which was near census for this administration.

PART 7: CALIBRATION, SCALING AND EQUATING

Part 7 of the technical report describes calibration and scaling procedures and results for the Spring 2015 AIMS assessments. All grade levels and content areas were calibrated and scaled with calibration samples that typically consisted of the entire student population. Part 7 of this report addresses the following AERA/APA/NCME *Standards* from the 1999 edition: 1.13, 2.1, 2.2, 2.14, 4.1, 4.2, 4.3, 6.4, 6.5, and 13.6. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this chapter are: 1.10, 2.3, 2.13, 2.14, 5.1, 5.2, 5.3, 7.2, 7.4, and 12.9. Also note that the Fall 2014 AIMS assessment is described in Appendix A.

7.1 Ensuring Valid Records in Calibration Sample

In order to ensure valid calibration results, several data cleaning steps occurred upon receipt of raw data from the scanning and scoring processes. These steps allowed for calibration to be conducted on valid student responses at the targeted grade level. Records for students taking all field test forms of the tests were included.

The cleaning process removed the following records from the calibration datasets for each content area and grade level:

- records with invalid tests noted by a special invalidation code obtained from ADE and marked on the answer document;
- records with non-valid attempts noted by less than one response in any of the test sessions;
- records for Bureau of Indian Affairs schools, juvenile corrections centers, state hospital schools, private schools, and home schooled students;
- records for students in cohorts other than 2016 (high school tests only);
- records which indicated the student took a test other than their grade level test; and
- duplicate records (score sheets were double scanned or students indicated as taking the test more than one time).

7.2 Calibration Methods

Item Response Theory (IRT) models were used in the item calibration for all AIMS Reading, Writing, Mathematics, and Science tests. All tests were calibrated separately by grade and content area. All calibration activities were replicated by ADE staff as an added quality control check.

7.2.1 Calibration Model

The AIMS Mathematics, Reading, and Science assessments are composed of multiple-choice items. Historically, the AIMS Mathematics, Reading, and Science tests have been developed and calibrated using the Rasch Model. The Rasch model (Rasch, 1960; Wright, 1977) can be conceptualized as a one-parameter IRT (1PL) model in which item difficulty and student ability are estimated on the same scale. The Rasch model defines a multiple-choice 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 item difficulty.

The AIMS Writing test in high school is composed of multiple-choice items and one extended-response essay scored on a six-point scale. The Rasch model is not applicable in this case, so the Partial Credit Model (Masters, 1982; Wright & Stone, 1979) is used to create the scale. The Partial Credit Model is an extension of the Rasch model designed for use with items that have multiple response categories. The PCM reduces to the Rasch model for items with only two response categories, such as multiple-choice items. For an item involving m_i score categories, the general expression for the probability of scoring x on item i is given by:

$$P_{xi} = \exp \sum_{j=0}^{x} (\theta - D_{ij}) / \sum_{k=0}^{m_i} \left[\exp \sum_{j=0}^{k} (\theta - D_{ij}) \right] \qquad (6.1)$$

where $x = 0, 1, ..., m_i$, and by definition, $\sum_{j=0}^{0} (\theta - D_{ij}) = 0$.

The above equation gives the probability of scoring x on the i-th test item as a function of ability (θ) and the difficulty (D_{ij}) of the m_i steps of the task. According to this model, the probability of an examinee scoring in a particular category is the sum of the logit (log-odds) differences between θ and D_{ij} of all the completed steps, divided by the sum of the differences of all the steps of a task.

7.2.2 Calibration Software

Parameter estimation for items on the science tests in grade 4, 8, and high school was implemented using WINSTEPS 3.71 (Linacre, 2011). WINSTEPS uses joint maximum likelihood estimation (JMLE) as described by Wright and Masters (1982). Item parameters for the reading, mathematics, and writing assessments in high school were obtained from the previous administrations in which items were administered to create the score tables.

7.3 Calibration Results

7.3.1 IRT Item Statistics

Item statistics resulting from calibration of the AIMS science tests for grades 4, 8, and high school are presented in Tables 7.3.1.1 through 7.3.1.3. The IRT item statistics for the reading, mathematics, and writing assessments for high school were not generated since the tests were pre-equated for the spring 2015 administration. Please refer to the AIMS 2014 Technical Report (Arizona Department of Education, 2014) for the IRT item analysis statistics in the reading and mathemetics assessments for high school because the spring 2015 tests were reused forms from spring 2014. Please also refer to the 2012 AIMS Technical Report (Arizona Department of Education, 2012) and 2013 AIMS Technical Report (Arizona Department of Education, 2013) for the IRT item statistics for the writing multiple-choice items and an extended-response item in high school, respectively since the multiple-choice items and extended-response item were reused items from the spring 2012 and spring 2013 administrations, respectively. All items for all AIMS tests converged during calibration using typical procedures for WINSTEPS software. Standard error of estimates for the Rasch difficulty measures indicated that the parameters were well estimated. Model-to-item data fit was monitored using weighted mean-square (MNSQ) and unweighted MNSQ

statistics, which indicate 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 MNSQ and unweighted mean-square is referred to as outfit MNSQ. The infit MNSQ is sensitive to unexpected responses at or near the item's calibrated level; whereas, outfit MNSQ 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 MNSQ indicate misfit, and values greater than 1.4 for outfit MNSQ indicate misfit (Wright & Linacre, 1994). No item was flagged as having misfit as indicated by weighted MNSQ or unweighted MNSQ.

Table 7.3.1.1
Spring 2015 AIMS IRT Item Statistics
Science Grade 4

Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit
1	-0.64	0.01	1.02	1.08	28	0.42	0.01	1.03	1.02
2	-0.20	0.01	1.05	1.08	29	-0.16	0.01	0.84	0.74
3	-0.44	0.01	0.95	0.89	30	1.16	0.01	1.08	1.10
4	0.49	0.01	0.93	0.89	31	1.87	0.01	1.15	1.28
5	0.10	0.01	0.98	0.97	32	1.47	0.01	0.99	1.03
6	0.89	0.01	1.06	1.08	33	1.28	0.01	0.94	0.94
7	0.52	0.01	0.97	0.97	34	0.38	0.01	0.89	0.85
8	0.27	0.01	0.95	0.92	35	0.34	0.01	0.92	0.89
9	1.12	0.01	1.15	1.20	36	0.25	0.01	1.13	1.18
10	-0.27	0.01	0.91	0.88	37	1.66	0.01	1.14	1.20
11	0.95	0.01	1.02	1.03	38	0.53	0.01	0.99	0.96
12	1.73	0.01	1.08	1.16	39	1.39	0.01	1.04	1.08
13	1.18	0.01	1.04	1.06	40	-0.33	0.01	1.00	1.03
14	1.07	0.01	1.09	1.12	41	1.80	0.01	1.03	1.11
15	1.03	0.01	1.13	1.17	42	0.15	0.01	1.01	1.00
16	0.45	0.01	0.92	0.88	43	0.67	0.01	0.99	0.98
17	-0.14	0.01	1.01	1.06	44	2.26	0.01	1.10	1.27
18	0.42	0.01	1.01	1.04	45	1.47	0.01	1.07	1.12
19	1.82	0.01	1.10	1.21	46	-0.83	0.01	0.90	0.74
20	0.84	0.01	0.91	0.89	47	0.98	0.01	0.95	0.94
21	-0.24	0.01	0.90	0.84	48	0.48	0.01	0.96	0.96
22	0.14	0.01	0.91	0.89	49	-0.16	0.01	0.83	0.73
23	0.82	0.01	1.02	1.02	50	0.69	0.01	0.97	0.94
24	1.13	0.01	1.05	1.07	51	0.68	0.01	0.97	0.95
25	-0.44	0.01	0.85	0.71	52	0.59	0.01	0.92	0.88
26	0.71	0.01	0.96	0.94	53	0.48	0.01	1.12	1.15
27	1.65	0.01	1.07	1.13	54	0.13	0.01	0.91	0.87

Table 7.3.1.2
Spring 2015 AIMS IRT Item Statistics
Science Grade 8

Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit
1	-1.07	0.01	1.03	1.14	30	-0.09	0.01	0.84	0.77
2	-0.38	0.01	0.99	1.01	31	0.91	0.01	1.02	1.03
3	0.16	0.01	1.11	1.25	32	-1.06	0.01	0.97	0.82
4	0.25	0.01	0.91	0.87	33	0.82	0.01	1.16	1.22
5	-0.71	0.01	0.93	0.83	34	-0.18	0.01	0.91	0.83
6	0.27	0.01	1.18	1.31	35	1.16	0.01	1.02	1.03
7	-0.78	0.01	0.97	0.93	36	0.89	0.01	1.09	1.13
8	-1.29	0.01	0.80	0.64	37	2.33	0.01	1.13	1.50
9	0.21	0.01	1.10	1.19	38	0.37	0.01	1.04	1.09
10	0.23	0.01	1.02	1.04	39	0.90	0.01	1.03	1.03
11	-0.05	0.01	1.04	1.07	40	0.57	0.01	0.90	0.87
12	-0.14	0.01	0.95	0.92	41	0.64	0.01	0.95	0.92
13	-0.57	0.01	1.00	1.03	42	-0.36	0.01	0.91	0.85
14	1.15	0.01	1.14	1.19	43	0.67	0.01	0.99	0.99
15	1.05	0.01	1.04	1.05	44	-0.97	0.01	0.95	0.73
16	0.04	0.01	0.98	0.97	45	0.97	0.01	1.14	1.18
17	0.23	0.01	0.97	0.96	46	0.74	0.01	0.98	0.97
18	1.92	0.01	1.19	1.39	47	-0.65	0.01	0.84	0.69
19	-0.38	0.01	0.96	0.92	48	0.37	0.01	0.87	0.84
20	1.99	0.01	1.04	1.10	49	0.53	0.01	1.02	1.04
21	1.20	0.01	1.02	1.04	50	0.76	0.01	0.88	0.85
22	0.91	0.01	1.08	1.10	51	0.74	0.01	1.05	1.06
23	0.71	0.01	0.95	0.93	52	0.46	0.01	0.99	0.98
24	1.64	0.01	0.98	1.02	53	-0.76	0.01	0.90	0.80
25	-0.01	0.01	0.98	0.98	54	0.27	0.01	0.92	0.87
26	1.13	0.01	1.02	1.05	55	1.82	0.01	1.05	1.19
27	1.27	0.01	0.92	0.93	56	1.64	0.01	1.07	1.14
28	1.17	0.01	1.04	1.08	57	0.69	0.01	1.14	1.17
29	0.98	0.01	0.96	0.96	58	0.76	0.01	1.04	1.04

Table 7.3.1.3
Spring 2015 AIMS IRT Item Statistics
Science Grade HS

Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit
1	0.20	0.01	1.05	1.09	34	1.59	0.01	1.04	1.11
2	0.20	0.01	0.87	0.84	35	-0.14	0.01	0.86	0.82
3	0.76	0.01	1.03	1.04	36	-0.73	0.01	0.83	0.73
4	0.90	0.01	1.12	1.16	37	1.10	0.01	1.10	1.13
5	1.04	0.01	1.16	1.22	38	1.60	0.01	1.07	1.18
6	1.03	0.01	1.08	1.11	39	0.20	0.01	0.95	0.92
7	1.65	0.01	1.06	1.15	40	1.00	0.01	1.02	1.05
8	0.09	0.01	1.01	1.01	41	0.04	0.01	0.92	0.88
9	-0.24	0.01	0.99	1.06	42	0.46	0.01	0.95	0.94
10	0.16	0.01	0.96	0.94	43	0.80	0.01	1.01	1.02
11	0.19	0.01	1.01	1.01	44	0.97	0.01	1.06	1.08
12	1.49	0.01	1.08	1.12	45	1.57	0.01	1.04	1.09
13	-0.74	0.01	0.91	0.86	46	0.31	0.01	0.98	0.97
14	0.22	0.01	1.15	1.21	47	0.61	0.01	0.90	0.87
15	-0.75	0.01	1.00	1.18	48	1.64	0.01	1.10	1.19
16	-0.22	0.01	0.96	0.96	49	0.07	0.01	0.98	1.02
17	0.37	0.01	0.97	0.96	50	-0.24	0.01	0.95	0.90
18	0.40	0.01	0.90	0.87	51	0.18	0.01	0.95	0.93
19	0.93	0.01	0.99	1.00	52	0.94	0.01	1.07	1.09
20	1.36	0.01	1.07	1.12	53	0.48	0.01	0.98	0.98
21	-1.41	0.01	0.89	0.72	54	0.30	0.01	0.82	0.78
22	0.45	0.01	1.05	1.09	55	0.49	0.01	1.02	1.03
23	-0.49	0.01	0.89	0.81	56	0.67	0.01	1.01	1.02
24	-0.82	0.01	0.92	0.90	57	0.78	0.01	1.01	1.02
25	0.23	0.01	0.99	0.99	58	1.09	0.01	0.96	0.98
26	1.02	0.01	0.99	1.00	59	1.38	0.01	1.07	1.12
27	1.49	0.01	1.06	1.14	60	1.12	0.01	1.19	1.27
28	0.27	0.01	1.01	1.04	61	-0.43	0.01	0.93	0.88
29	0.65	0.01	0.99	0.99	62	0.62	0.01	1.04	1.04
30	0.48	0.01	0.99	0.98	63	0.70	0.01	0.98	0.97
31	0.62	0.01	0.99	0.98	64	0.91	0.01	1.01	1.02
32	0.48	0.01	1.01	1.01	65	0.46	0.01	0.96	0.96
33	0.86	0.01	1.01	1.01					

7.4 Scaling Methods

7.4.1 Reading

In 2005, a scale of measurement was determined for each of the AIMS Reading tests. The AIMS Reading test for high school was placed on a scale that ranged from 500 to 900 with an approximate mean of 700. The standard deviation was set to 50 for high school reading. A detailed description concerning the development of the scale of measurement for the AIMS Reading tests can be found in section 7.2.1 of the *2005 AIMS Technical Report* which can be obtained by contacting the Arizona Department of Education. A report detailing the procedures used to set performance standards on the reading tests is available at http://www.azed.gov/assessment/files/2014/05/az-aims-reading-math-ss-report-2005.pdf.

7.4.2 Science

A scale of measurement was determined for science using spring 2008 operational test results and cut scores were determined during standard setting meetings. A detailed description concerning the development of the scale of measurement can be found in Appendix B of the *2008 AIMS Technical Report* which can be obtained from the Arizona Department of Education. A report detailing the procedures used to set performance standards on the science tests is available at http://www.azed.gov/assessment/files/2014/05/aims2008sciencerevisedstandardsettingtechnicalreport.pdf. The AIMS science scales for grades 4, 8, and high school ranged from 200 to 800. The science scales are not on a vertical scale. Each grade has its own unique scale so that the scale scores for different grades can NOT be compared.

7.4.3 Mathematics

A new scale was created for the mathematics tests in 2010 using the operational test results. The tests were built to the 2008 content standards using a new test blueprint. Performance standards were also set on the new tests in spring 2010. The procedures used to set performance standards are described in a report available from the Arizona Department of Education, and on the ADE website at http://www.azed.gov/assessment/files/2014/05/azaimsmathssreport2010.pdf.

The high school test was scaled to have a mean of 500 and a standard deviation of 50. The scale scores range from 300 to 700.

7.4.4 Writing

New writing tests were implemented for high school in spring 2011. The tests consist of multiple-choice items and an extended-response essay. Performance standards were also set on the new tests in spring 2011.The test for each grade was scaled such that scores range from 300 to 700 and have a mean of 500 and a standard deviation of 50. A detailed description concerning the development of the scale of measurement can be found in Section 7.5.2 of this *Technical Report* and in the writing standard setting report available at http://www.azed.gov/assessment/files/2014/05/az-aims-writing-ss-report-final.pdf.

7.5 Equating

7.5.1 Reading, Mathematics, and Science

The 2015 AIMS Science tests were equated and placed on the operational AIMS scale using a common-item, non-equivalent groups design. A set of anchor items was selected from the 2014 and previous operational assessments before the item selection workshop. The anchor items were selected with two principles in mind. First, the subset of anchor items should represent the content covered by the full AIMS assessment. Second, the subset of anchor items should be representative of the distribution of item difficulties for the full assessment. The only items eligible to be considered for anchor items were AIMS items. No dual purpose or NRT items served as anchors. Note that the AIMS Reading and Mathematics tests for high school were pre-equated for the spring 2015 administration, in which the score tables were generated by using the most recent IRT statistics in the item bank.

Table 7.5.1.1 presents the number of anchor/common items for each grade/subject area. Table 7.5.1.2 show the content representation for the 2015 anchor items compared to the 2015 operational form. Table 7.5.1.3 presents descriptive statistics for the 2015 anchor/common item difficulties and the 2014 operational form.

Table 7.5.1.1
Spring 2015 AIMS Anchor Items

Content	Grade	CRT Total	Anchor
Science	4	54	21
	8	58	23
	HS	65	18

Table 7.5.1.2
Representation of Content by 2015 Anchor Sets, Science

		Strand																		
		1				2		3		4					5			6		Total
		Concept				Concept		Concept		Concept					Concept			Concept		
		1	2	3	4	1	2	1	2	1	2	3	4	5	1	2	3	2	3	
SC04 All	N	6	6	6		3	3	3	3	3	0	1	2		0	0	6	6	6	54
	Pct	11.11	11.11	11.11		5.56	5.56	5.56	5.56	5.56	0	1.85	3.70		0	0	11.11	11.11	11.11	100
Anchor	N	3	2	1		1	2	2	1	2					0	0	3	2	2	21
	Pct	14.29	9.52	4.76		4.76	9.52	9.52	4.76	9.52					0	0	14.29	9.52	9.52	100
SC08 All	N	6	4	6	4	4	2	2	4	0	3	0	5		10	8				58
	Pct	10.34	6.90	10.34	6.90	6.90	3.45	3.45	6.90	0	5.17	0	8.62		17.24	13.79				100
Anchor	N	2	2	2	1	0	2	0	3	0	0	0	3		4	4				23
	Pct	8.70	8.70	8.70	4.35	0	8.70	0	13.04	0	0	0	13.04		17.39	17.39				100
SCHS All	N	6	6	6	4	4	2	7		6	6	6	6	6						65
	Pct	9.23	9.23	9.23	6.15	6.15	3.08	10.77		9.23	9.23	9.23	9.23	9.23						100
Anchor	N	2	2	1	1	1	1	3		2	2	1	2							18
	Pct	11.11	11.11	5.56	5.56	5.56	5.56	16.67		11.11	11.11	5.56	11.11							100

Table 7.5.1.3
Representation of Difficulty by 2015 Anchor Sets, Science

			Difficulty Parameter		P-Value	
Content	Grade	Statistic	Entire 2015 Test	All Anchor Items	Entire 2015 Test	All Anchor Items
SC	4	N	54	21	54	21
		Mean	0.6400	0.5700	0.5500	0.5700
		Std Dev	0.7200	0.5300	0.1400	0.1000
		Min	-0.8300	-0.8300	0.2400	0.4200
		Max	2.2600	1.3900	0.8100	0.8100
SC	8	N	58	23	58	23
		Mean	0.4500	0.3700	0.5700	0.5900
		Std Dev	0.8200	0.8400	0.1500	0.1500
		Min	-1.2900	-1.2900	0.2200	0.3000
		Max	2.3300	1.9900	0.8800	0.8800
SC	HS	N	65	18	65	18
		Mean	0.5100	0.2200	0.5000	0.5600
		Std Dev	0.6700	0.5400	0.1300	0.1000
		Min	-1.4100	-0.7400	0.2700	0.4000
		Max	1.6500	1.0900	0.8400	0.7600

A fixed-parameter equating was implemented within WINSTEPS in order to link the 2015 science tests to the operational reporting scale. This is implemented by constraining the 2015 parameter estimates for the common anchor items to equal the final parameter estimates obtained in the original AIMS calibration analyses. The displacement statistic, which estimates the difference between the fixed parameter and the estimate had the item parameter not been constrained, was evaluated for each anchor item. Displacement statistic greater than 0.5 or less than -0.5 are considered significant in the Rasch literature and caused the anchor item to be removed from the anchor set. The following procedure was used to examine anchor item performance and determine whether to remove anchor items that exhibited significant displacement statistics from the annual equating:

1. All anchor items with displacement statistics greater than 0.3 or less than -0.3 were flagged. Any anchor item with displacement statistic greater than 0.5 or less than -0.5 was dropped from the anchor item set. If more than one item was observed with a displacement statistic greater than 0.5 or less than -0.5, then only the first item with the largest displacement value was dropped from the anchor set. The displacement values of the remaining anchor items were re-estimated by equating the test again using the remaining anchor items. This process of equating and dropping the anchor item with the largest displacement greater than 0.5 or less than -0.5 was repeated until all displacements were acceptable. All items with displacement values greater than 0.5 or less than -0.5 were noted to be carried over for removal from the anchor set for next year.

2. Anchor items with displacement statistics greater than 0.3 and less than 0.5 or less than -0.3 and greater than -0.5 were investigated using infit MNSQ, outfit MNSQ, item difficulty, and point-biserial correlation. If either infit MNSQ or outfit MNSQ was flagged for the anchor item, the

removal of this item was considered. Also, if any item showed point-biserial correlation less than 0.3, the removal of this item was considered.

3. Whenever an anchor item was removed, content and difficulty representativeness of the remaining anchor set was examined. In instances where more than one anchor item was considered for removal for a given content and grade, the content strand and difficulty level of the item was considered to prevent removal of more than one item from the same content strand and difficulty level.

4. If more than one item was removed from the same content strand, a note was made to address the problem in the setup of anchors for the succeeding year's assessment.

This procedure resulted in removing from the anchor sets for grade 4, 8, and high school a total of four science items.

7.5.2 Writing

New writing assessments were implemented in grades 5, 6, 7, and high school in spring 2011, but only writing for high school was administered in spring 2015. The assessments are composed of multiple-choice items and an essay prompt. The high school test contains 27 AIMS multiple-choice items. The essay prompt is scored holistically using a 6-point scoring rubric. At the high school level, the essay prompt is scored by two readers with identical or adjacent scores being summed to produce a final essay score. Non-adjacent scores are reviewed and scored by a supervisor whose score becomes the final score. The AIMS test scores are constructed such that the 27 multiple-choice items account for approximately 40% of the total test score and the score on the essay response counts for approximately 60% of the total test score. The high school writing multiple-choice items are embedded in a single test along with the prompt, while the elementary tests combine the writing multiple-choice along with the reading multiple-choice items. These items are administered in three sessions with the prompt administered during a separate session.

7.5.2.1 **Weighting the Multiple-Choice and Essay Components**

The writing test at the high school level consists of 27 multiple-choice items and an essay prompt. The essay is scored by two raters with exact or adjacent scores being summed for a final essay score ranging from 2 to 12 points. In order to achieve the desired weighting of the multiple-choice and essay components, the multiple-choice score is multiplied by 2 and the essay score is multiplied by 7. In this way, there are 2*27 + 7*12, or 138 total points for the test with the multiple-choice items accounting for 54/138, or 39 percent of the total score, and the essay accounting for 84/138, or 61 percent of the total score.

7.5.2.2 **Calibration**

Calibration of the writing tests is performed using WINSTEPS, version 3.71. A two-step process is used in which the multiple-choice items are calibrated first without including the essay score. Then the multiple-choice item parameters from the initial calibration are held fixed and the essay is calibrated. During this second step, the IWEIGHT command is used with a weight of 7 for the essay response for the high school test, the IWEIGHT command is used with a weight of 2 for the multiple-choice items and a weight of 7 for the essay response.

7.5.2.3 **Standard Setting**

Performance standards were determined separately for the multiple-choice and essay components of the writing tests. The Item Mapping method of standard setting was used to identify the standard on the theta scale for each performance level for the multiple-choice component. The performance

standards were set for the essay component such that the theta value for each performance level corresponded approximately to a specific raw score on the essay. For Grades 5, 6, and 7, the value of theta corresponding raw scores of 3, 4, and 5.5 were used as the cut points for Approaches, Meets, and Exceeds, respectively. For the high school test, the value of theta corresponding raw scores of 5, 7, and 11 were used as the cut points for Approaches, Meets, and Exceeds, respectively. The final theta cut points were determined by weighting the multiple-choice theta at each cut point by 40% and the essay theta at each cut by 60% and combining them into a final theta cut point for each performance level.

7.5.3 Scoring and Standard Error of Measurement

Item response theory makes available two types of scoring: number-correct and item-pattern. With number-correct scoring, the value of theta corresponding to each number-correct score (or raw score) is converted to a scale score. Item-pattern scoring produces a scale score, taking into account not only how many items were answered correctly but also which items and the characteristics of those items. For groups of 25 or more students, the two methods produce tau-equivalent results (Yen, 1984.) Tau-equivalent means that examinees are expected to receive the same score on average between the two methods. Number-correct scoring was used to derive scales scores for the AIMS tests.

Typically, a test score is obtained from a single observation of performance 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 or fatigue. 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 7.5.3.1 through 7.5.3.6 present raw score to scale score conversion tables and IRT conditional standard errors of measurement for all AIMS Reading, Mathematics, and Writing assessments for high school and Science grades 4, 8, and high school assessments. The values in bold represent the scale score with the smallest value greater than or equal to the established cut score for each grade level and content area. The "greater than" rule is evoked when the actual scale score is not observed in any given table. Note that the scores for writing in high school (Tables 7.5.3.3) range from 0 to 138. This is because of the differential weighting of the multiple-choice items and the essay described above.

Table 7.5.3.1
Spring 2015 AIMS Raw Score to Scale Score Table
Mathematics High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	59	43	486	10
1	300	42	**44**	**488**	**10**
2	315	30	45	490	10
3	332	25	46	493	10
4	345	22	47	495	10
5	355	20	48	497	10
6	364	18	49	500	10
7	371	17	50	502	10
8	378	16	51	504	10
9	384	15	52	507	10
10	389	15	53	509	10
11	394	14	54	512	10
12	398	14	55	514	10
13	403	13	56	517	10
14	407	13	57	519	10
15	410	12	58	522	10
16	414	12	59	524	10
17	418	12	60	527	11
18	421	12	61	530	11
19	424	11	62	533	11
20	427	11	63	536	11
21	430	11	**64**	**539**	**11**
22	433	11	65	542	11
23	436	11	66	545	11
24	439	11	67	548	12
25	442	11	68	551	12
26	445	11	69	555	12
27	447	10	70	558	12
28	450	10	71	562	13
29	452	10	72	566	13
30	455	10	73	571	14
31	457	10	74	575	14
32	460	10	75	580	15
33	462	10	76	586	15
34	465	10	77	591	16
35	467	10	78	598	17
36	469	10	79	605	18
37	**472**	**10**	80	614	20
38	474	10	81	624	22
39	476	10	82	637	25
40	479	10	83	655	30
41	481	10	84	685	42
42	483	10	85	700	59

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

Table 7.5.3.2
Spring 2015 AIMS Raw Score to Scale Score Table
Reading High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	500	61	28	700	13
1	509	44	29	704	13
2	541	31	30	708	13
3	560	26	31	712	13
4	574	23	32	716	13
5	585	21	33	719	13
6	594	19	34	723	13
7	602	18	35	727	13
8	610	17	36	732	13
9	616	17	37	736	14
10	623	16	38	740	14
11	**628**	**15**	39	745	14
12	634	15	40	749	14
13	639	15	41	754	15
14	644	14	42	759	15
15	648	14	43	764	15
16	653	14	44	770	16
17	657	14	**45**	**776**	**17**
18	662	13	46	783	17
19	666	13	47	790	18
20	670	13	48	798	19
21	**674**	**13**	49	808	21
22	678	13	50	819	23
23	682	13	51	833	26
24	685	13	52	852	31
25	689	13	53	883	44
26	693	13	54	900	61
27	697	13			

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

Table 7.5.3.3
Spring 2015 AIMS Raw Score to Scale Score Table
Writing High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	55	35	383	9	70	452	9	105	539	12
1	300	55	36	385	9	71	454	9	106	542	12
2	300	55	37	387	9	72	456	9	107	545	12
3	300	55	38	389	9	73	458	9	108	549	12
4	300	55	39	391	9	74	460	9	109	553	12
5	300	55	40	393	9	75	462	9	110	556	12
6	300	55	41	395	9	76	464	9	111	560	12
7	300	55	42	397	9	77	466	9	112	563	12
8	300	55	43	399	9	78	469	9	113	567	12
9	300	55	44	401	9	79	471	9	114	571	12
10	300	55	45	403	9	80	473	9	115	575	12
11	300	55	46	405	9	81	475	9	116	579	12
12	300	55	47	407	9	82	477	9	117	583	13
13	300	55	48	409	9	83	479	9	**118**	**587**	**13**
14	300	55	49	411	9	**84**	**481**	**9**	119	591	13
15	300	38	50	413	9	85	484	9	120	595	13
16	300	26	51	415	9	86	486	9	121	600	13
17	311	21	52	417	9	87	488	10	122	605	14
18	321	18	53	419	9	88	490	10	123	610	14
19	329	16	54	421	9	89	493	10	124	615	14
20	335	15	55	423	9	90	495	10	125	620	15
21	341	14	56	425	9	91	498	10	126	626	15
22	345	13	57	427	9	92	500	10	127	631	15
23	349	12	58	429	9	93	503	10	128	638	16
24	353	12	59	430	9	94	505	10	129	644	16
25	357	11	60	432	9	95	508	10	130	651	17
26	360	11	**61**	**434**	**9**	96	511	10	131	659	18
27	363	11	62	436	9	97	514	11	132	667	19
28	366	10	63	438	9	98	517	11	133	676	20
29	368	10	64	440	9	99	519	11	134	687	22
30	371	10	65	442	9	100	523	11	135	700	24
31	373	10	66	444	9	101	526	11	136	700	29
32	376	10	67	446	9	102	529	11	137	700	40
33	378	10	68	448	9	103	532	11	138	700	56
34	381	9	69	450	9	104	535	11			

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

Table 7.5.3.4
Spring 2015 AIMS Raw Score to Scale Score Table
Science Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	70	**28**	**504**	**14**
1	294	50	29	508	14
2	329	36	30	512	14
3	350	29	31	516	14
4	366	26	32	520	14
5	378	23	33	524	14
6	389	22	34	528	15
7	398	20	35	533	15
8	406	19	36	537	15
9	413	18	37	542	15
10	420	18	**38**	**547**	**15**
11	426	17	39	552	16
12	432	17	40	557	16
13	437	16	41	562	16
14	442	16	42	567	17
15	448	16	43	573	17
16	452	15	44	580	18
17	457	15	45	586	19
18	**462**	**15**	46	594	19
19	466	15	47	602	20
20	470	15	48	611	22
21	475	14	49	621	24
22	479	14	50	634	26
23	483	14	51	649	30
24	487	14	52	671	36
25	491	14	53	706	50
26	495	14	54	800	70
27	499	14			

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

Table 7.5.3.5
Spring 2015 AIMS Raw Score to Scale Score Table
Science Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	69	30	498	14
1	284	49	**31**	**502**	**14**
2	319	35	32	505	14
3	340	29	33	509	14
4	355	25	34	513	14
5	367	23	35	517	14
6	378	21	36	521	14
7	387	20	37	525	14
8	395	19	38	529	14
9	402	18	**39**	**533**	**14**
10	408	18	40	538	14
11	415	17	41	542	15
12	420	16	42	547	15
13	426	16	43	552	15
14	431	16	44	556	16
15	436	15	45	562	16
16	441	15	46	567	16
17	445	15	47	573	17
18	450	15	48	579	17
19	454	14	49	585	18
20	459	14	50	593	19
21	463	14	51	601	20
22	467	14	52	609	21
23	471	14	53	620	23
24	**475**	**14**	54	632	25
25	479	14	55	647	29
26	482	14	56	668	35
27	486	14	57	703	49
28	490	14	58	800	69
29	494	14			

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

Table 7.5.3.6
Spring 2015 AIMS Raw Score to Scale Score Table
Science High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	74	33	493	14
1	262	53	34	496	14
2	300	38	**35**	**500**	**14**
3	322	31	36	503	14
4	338	27	37	507	14
5	351	25	38	511	14
6	362	23	39	514	14
7	372	21	40	518	14
8	380	20	41	522	14
9	387	19	42	525	14
10	394	18	43	529	14
11	401	18	44	533	14
12	407	17	**45**	**537**	**15**
13	412	17	46	541	15
14	417	16	47	546	15
15	422	16	48	550	15
16	427	16	49	554	16
17	432	15	50	559	16
18	436	15	51	564	16
19	441	15	52	569	17
20	445	15	53	575	17
21	449	14	54	580	18
22	453	14	55	587	18
23	457	14	56	593	19
24	460	14	57	601	20
25	464	14	58	609	21
26	468	14	59	618	23
27	472	14	60	629	25
28	**475**	**14**	61	642	27
29	479	14	62	658	31
30	482	14	63	680	38
31	486	14	64	717	53
32	489	14	65	800	74

Note. SEM is the standard error of measurement for the scale score.
Note. Cut scores for Approaches the Standard, Meets the Standard, and Exceeds the Standard are in boldface.

PART 8: TEST RESULTS

8.1 Data

Part 8 of this technical report contains information about the results of the spring 2015 administration of AIMS Science in grades 4, 8, and high school and AIMS Reading, Mathematics, and Writing assessments for high school. The 1999 AERA/APA/NCME *Standards* addressed in Part 8 include: 1.5, 4.3, 4.5, 4.6, 4.7, 6.5, 7.1, 7.10, 13.15, and 13.19. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this chapter are: 1.10, 5.1, 5.2, 5.3, 5.8, 5.9, 7.2, 7.4, and 12.9. Please note that the corresponding information for the Fall 2014 AIMS administration can be found in Appendix A.

Results presented below are based on population data contained within the final electronic data files. The results presented 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 final results typically use more detailed school-level information than is used to conduct research analyses. The results in the following tables are presented as evidence of reliability and validity of the AIMS assessments and should not be used for state accountability purposes.

8.1.1 AIMS State Test Results

The AIMS test results for Science for grades 4, 8, and high school are not on a vertical scale and therefore the scale scores across grades can not be compared. For each grade, the lowest obtainable scale score (LOSS) on the science tests is 200, and the highest obtainable scale score (HOSS) is 800. The AIMS Reading, Mathematics, and Writing assessments for high school are each on a separate scale and each has different LOSS and HOSS. The LOSS and HOSS values for each grade/subject can be found in Table 8.1.1.1.

Table 8.1.1.1
Spring 2015 AIMS LOSS and HOSS Table

Content	Grade	Loss	Hoss
Reading	HS	500	900
Mathematics	HS	300	700
Writing	HS	300	700
Science	4	200	800
	8	200	800
	HS	200	800

Test results for each grade level and content area test follow in Tables 8.1.1.2 through 8.1.1.6. For each grade, scale score means and standard deviations as well as the percentages of students in each performance level are presented for the state as a whole and disaggregated into various demographic groups.

In addition to the descriptive statistics presented in Tables 8.1.1.2 through 8.1.1.6, scale score frequency distributions are also presented in Tables 8.1.1.7 through 8.1.1.13. Each grade and content area is presented in a separate table. These tables show the scale score, frequency (Freq), cumulative frequency (Cum Freq), percentage (%), and cumulative percentage (Cum %).

Results for AIMS assessments for high school are reported by graduating cohort for Science only. Starting spring 2015, the Reading, Writing, and Mathematics data will no longer be broken out by cohort for high school because students are not longer required to take those tests. Cohort 18 is defined as the group of students that expect to graduate in 2018 and typically includes grade 9 students. The results for Cohort 18 are reported only for science because grade 9 science scores are used in the state accountability system. Cohort 17 is defined as the group of students that expect to graduate in 2017 and typically includes 10th grade students.

Table 8.1.1.2
Spring 2015 AIMS State Test Results
Mathematics High School

		Scale Score		% at Performance Level			
Group	N	M	SD	FFBS	AS	MS	ES
Total	4286	470.36	27.98	53	24	20	3
Hispanic	2485	468.80	24.44	54	25	19	1
Non Hispanic	1501	473.66	32.72	51	23	21	5
Ethnic Background							
White	1736	475.23	30.61	48	24	24	4
Black or African American	333	464.23	22.82	61	26	13	1
Asian	69	485.84	43.23	39	29	17	14
American Indian or Alaskan Native	592	463.19	23.40	64	21	15	1
Native Hawaiian or Other Pacific Islander	22	469.86	31.20	45	27	23	5
Multiple Indication	66	480.24	48.97	59	17	14	11
Special Program Membership							
English Learner Program	202	461.93	24.73	67	19	13	1
Special Education	571	454.62	21.98	78	13	8	0
Low SES	2705	467.60	24.27	56	24	18	1
Migrant	3	*	*	*	*	*	*

Note. FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. These results should not be used for accountability purposes. Statistics for subgroups with less than 11 students are omittted in compliance with FERPA regulations and replaced with an '*'.

Table 8.1.1.3
Spring 2015 AIMS State Test Results
Reading High School

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Total	1478	693.06	54.06	4	42	44	10
Hispanic	727	685.41	48.01	3	46	45	6
Non Hispanic	654	702.61	58.67	4	36	45	14
Ethnic Background							
White	663	709.89	57.84	3	30	51	16
Black or African American	109	675.92	42.40	5	51	41	3
Asian	65	714.60	60.77	3	34	42	22
American Indian or Alaskan Native	204	668.41	40.09	6	59	33	2
Native Hawaiian or Other Pacific Islander	11	689.82	55.81	9	45	36	9
Multiple Indication	34	716.09	58.21	0	35	44	21
Special Program Membership							
English Learner Program	172	657.29	20.36	5	73	22	0
Special Education	272	654.67	28.50	8	69	22	1
Low SES	868	680.12	45.20	4	51	41	4
Migrant	1	*	*	*	*	*	*

Note. FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. These results should not be used for accountability purposes. Statistics for subgroups with less than 11 students are omittted in compliance with FERPA regulations and replaced with an '*'

Table 8.1.1.4
Spring 2015 AIMS State Test Results
Writing High School

		Scale Score		% at Performance Level			
Group	N	M	SD	FFBS	AS	MS	ES
Total	3595	468.38	51.94	23	45	27	4
Hispanic	1967	459.00	42.49	25	51	22	2
Non Hispanic	1348	484.14	61.14	19	38	35	8
Ethnic Background							
White	1430	484.08	58.92	19	38	35	7
Black or African American	223	457.85	38.87	27	46	27	0
Asian	155	480.08	73.73	30	31	26	14
American Indian or Alaskan Native	355	452.10	35.80	27	55	17	0
Native Hawaiian or Other Pacific Islander	26	473.62	53.20	27	35	35	4
Multiple Indication	62	489.45	64.05	15	37	39	10
Special Program Membership							
English Learner Program	368	426.41	31.22	58	38	4	0
Special Education	628	435.98	30.71	46	46	8	0
Low SES	2239	457.99	42.88	27	50	21	2
Migrant	2	*	*	*	*	*	*

Note. FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. These results should not be used for accountability purposes. Statistics for subgroups with less than 11 students are omittted in compliance with FERPA regulations and replaced with an '*'

Table 8.1.1.5
Spring 2015 AIMS State Test Results
Science Grades 4 and 8

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Grade 4							
Total	83905	513.86	46.53	13	29	32	26
Hispanic	37420	498.75	41.24	18	37	31	14
Non Hispanic	46031	526.34	46.93	8	23	33	36
Ethnic Background							
White	66593	516.44	46.49	12	28	33	28
Black or African American	4752	499.16	41.22	17	38	31	14
Asian	2523	540.86	47.51	4	17	31	48
American Indian or Alaskan Native	6023	487.47	36.71	23	46	24	8
Native Hawaiian or Other Pacific Islander	309	510.68	42.68	11	30	39	20
Multiple Indication	2932	517.50	45.01	10	27	35	28
Special Program Membership							
English Learner Program	7670	467.08	28.58	43	46	11	1
Special Education	10882	484.29	41.97	32	38	20	10
Low SES	47898	499.08	41.40	18	37	31	14
Migrant	458	482.69	37.33	30	43	21	6
Grade 8							
Total	82042	513.06	48.08	22	20	24	34
Hispanic	36009	498.45	42.14	29	24	25	22
Non Hispanic	45762	524.63	49.30	16	16	24	45
Ethnic Background							
White	66182	515.36	47.72	20	19	25	36
Black or African American	4840	496.46	43.10	32	24	23	21
Asian	2457	545.39	53.30	9	10	20	61
American Indian or Alaskan Native	5538	487.37	38.82	39	27	21	13
Native Hawaiian or Other Pacific Islander	290	511.27	46.26	22	19	25	33
Multiple Indication	2219	515.6	46.52	18	20	26	36
Special Program Membership							
English Learner Program	2552	452.32	25.98	81	14	4	1
Special Education	8579	470.43	38.56	60	20	12	8
Low SES	43965	497.64	42.12	30	24	24	21
Migrant	457	486.22	41.26	41	25	19	15

Note. FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. These results should not be used for accountability purposes. Science results are not on a vertical scale.

Table 8.1.1.6
Spring 2015 AIMS State Test Results
Science High School

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Cohort 17							
Total	50767	484.29	44.73	45	20	21	14
Hispanic	23919	472.48	38.53	56	21	17	7
Non Hispanic	26226	495.22	47.21	35	20	25	20
Ethnic Background							
White	39434	486.94	44.93	42	21	22	15
Black or African American	3243	472.87	39.76	55	21	17	8
Asian	1237	506.96	53.48	30	17	24	30
American Indian or Alaskan Native	3793	463.97	35.12	66	18	11	4
Native Hawaiian or Other Pacific Islander	224	478.38	39.24	50	22	20	8
Multiple Indication	1474	489.90	43.26	40	21	25	15
Special Program Membership							
English Learner Program	1201	437.08	24.13	93	4	2	0
Special Education	5286	451.80	34.36	79	11	7	3
Low SES	27655	471.71	38.72	57	21	16	7
Migrant	454	453.83	29.70	76	17	6	1
Cohort 18							
Total	28869	504.20	49.28	29	19	26	26
Hispanic	10404	485.80	42.32	42	22	23	13
Non Hispanic	18052	514.88	49.97	22	17	28	34
Ethnic Background							
White	23736	505.25	48.91	28	19	27	27
Black or African American	1419	486.02	43.95	42	21	25	13
Asian	1150	533.71	52.17	13	13	28	47
American Indian or Alaskan Native	836	476.85	40.06	54	19	17	10
Native Hawaiian or Other Pacific Islander	100	488.29	40.95	41	13	33	13
Multiple Indication	817	504.63	50.41	29	19	25	27
Special Program Membership							
English Learner Program	249	438.40	25.10	92	4	3	0
Special Education	1906	458.04	40.46	74	11	9	6
Low SES	11344	484.48	43.33	44	21	22	13
Migrant	21	468.71	39.96	67	19	0	14

Note. FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. These results should not be used for accountability purposes. Science results are not on a vertical scale.

Table 8.1.1.7
Spring 2015 AIMS Frequency Distribution
Mathematics High School

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	300	0	0.00	0	0.00	43	486	127	2.82	3516	77.99
1	300	0	0.00	0	0.00	**44**	**488**	**127**	**2.82**	**3643**	**80.81**
2	315	0	0.00	0	0.00	45	490	106	2.35	3749	83.16
3	332	0	0.00	0	0.00	46	493	95	2.11	3844	85.27
4	345	0	0.00	0	0.00	47	495	84	1.86	3928	87.13
5	355	1	0.02	1	0.02	48	497	59	1.31	3987	88.44
6	364	0	0.00	1	0.02	49	500	57	1.26	4044	89.71
7	371	0	0.00	1	0.02	50	502	43	0.95	4087	90.66
8	378	0	0.00	1	0.02	51	504	49	1.09	4136	91.75
9	384	0	0.00	1	0.02	52	507	35	0.78	4171	92.52
10	389	0	0.00	1	0.02	53	509	33	0.73	4204	93.26
11	394	0	0.00	1	0.02	54	512	23	0.51	4227	93.77
12	398	2	0.04	3	0.07	55	514	25	0.55	4252	94.32
13	403	4	0.09	7	0.16	56	517	32	0.71	4284	95.03
14	407	3	0.07	10	0.22	57	519	22	0.49	4306	95.52
15	410	14	0.31	24	0.53	58	522	12	0.27	4318	95.79
16	414	11	0.24	35	0.78	59	524	26	0.58	4344	96.36
17	418	27	0.60	62	1.38	60	527	10	0.22	4354	96.58
18	421	44	0.98	106	2.35	61	530	14	0.31	4368	96.89
19	424	39	0.87	145	3.22	62	533	14	0.31	4382	97.20
20	427	50	1.11	195	4.33	63	536	11	0.24	4393	97.45
21	430	70	1.55	265	5.88	**64**	**539**	**14**	**0.31**	**4407**	**97.76**
22	433	87	1.93	352	7.81	65	542	13	0.29	4420	98.05
23	436	89	1.97	441	9.78	66	545	12	0.27	4432	98.31
24	439	119	2.64	560	12.42	67	548	8	0.18	4440	98.49
25	442	94	2.09	654	14.51	68	551	10	0.22	4450	98.71
26	445	113	2.51	767	17.01	69	555	10	0.22	4460	98.94
27	447	154	3.42	921	20.43	70	558	4	0.09	4464	99.02
28	450	151	3.35	1072	23.78	71	562	9	0.20	4473	99.22
29	452	145	3.22	1217	27.00	72	566	6	0.13	4479	99.36
30	455	185	4.10	1402	31.10	73	571	6	0.13	4485	99.49
31	457	135	2.99	1537	34.09	74	575	5	0.11	4490	99.60
32	460	171	3.79	1708	37.89	75	580	2	0.04	4492	99.65
33	462	172	3.82	1880	41.70	76	586	1	0.02	4493	99.67
34	465	175	3.88	2055	45.59	77	591	1	0.02	4494	99.69
35	467	205	4.55	2260	50.13	78	598	2	0.04	4496	99.73
36	469	166	3.68	2426	53.82	79	605	4	0.09	4500	99.82
37	**472**	**174**	**3.86**	**2600**	**57.68**	80	614	3	0.07	4503	99.89
38	474	187	4.15	2787	61.82	81	624	0	0.00	4503	99.89
39	476	151	3.35	2938	65.17	82	637	1	0.02	4504	99.91
40	479	163	3.62	3101	68.79	83	655	2	0.04	4506	99.96
41	481	155	3.44	3256	72.23	84	685	2	0.04	4508	100.00
42	483	133	2.95	3389	75.18	85	700	0	0.00	4508	100.00

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.8
Spring 2015 AIMS Frequency Distribution
Reading High School

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	500	2	0.13	2	0.13	28	700	22	1.42	1008	65.20
1	509	1	0.06	3	0.19	29	704	16	1.03	1024	66.24
2	541	0	0.00	3	0.19	30	708	17	1.10	1041	67.34
3	560	0	0.00	3	0.19	31	712	15	0.97	1056	68.31
4	574	1	0.06	4	0.26	32	716	22	1.42	1078	69.73
5	585	2	0.13	6	0.39	33	719	18	1.16	1096	70.89
6	594	2	0.13	8	0.52	34	723	13	0.84	1109	71.73
7	602	4	0.26	12	0.78	35	727	14	0.91	1123	72.64
8	610	6	0.39	18	1.16	36	732	16	1.03	1139	73.67
9	616	19	1.23	37	2.39	37	736	18	1.16	1157	74.84
10	623	26	1.68	63	4.08	38	740	34	2.20	1191	77.04
11	**628**	**48**	**3.10**	**111**	**7.18**	39	745	29	1.88	1220	78.91
12	634	49	3.17	160	10.35	40	749	31	2.01	1251	80.92
13	639	62	4.01	222	14.36	41	754	35	2.26	1286	83.18
14	644	62	4.01	284	18.37	42	759	33	2.13	1319	85.32
15	648	82	5.30	366	23.67	43	764	33	2.13	1352	87.45
16	653	74	4.79	440	28.46	44	770	38	2.46	1390	89.91
17	657	77	4.98	517	33.44	**45**	**776**	**35**	**2.26**	**1425**	**92.17**
18	662	75	4.85	592	38.29	46	783	29	1.88	1454	94.05
19	666	63	4.08	655	42.37	47	790	29	1.88	1483	95.92
20	670	60	3.88	715	46.25	48	798	26	1.68	1509	97.61
21	**674**	**46**	**2.98**	**761**	**49.22**	49	808	14	0.91	1523	98.51
22	678	52	3.36	813	52.59	50	819	14	0.91	1537	99.42
23	682	40	2.59	853	55.17	51	833	6	0.39	1543	99.81
24	685	42	2.72	895	57.89	52	852	0	0.00	1543	99.81
25	689	35	2.26	930	60.16	53	883	3	0.19	1546	100.00
26	693	34	2.20	964	62.35	54	900	0	0.00	1546	100.00
27	697	22	1.42	986	63.78						

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.9
Spring 2015 AIMS Frequency Distribution
Writing High School

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	300	0	0	0	0	37	387	5	0.13	78	2.07
1	300	0	0	0	0	38	389	20	0.53	98	2.61
2	300	0	0	0	0	39	391	9	0.24	107	2.85
3	300	0	0	0	0	40	393	27	0.72	134	3.56
4	300	0	0	0	0	41	395	8	0.21	142	3.78
5	300	0	0	0	0	42	397	38	1.01	180	4.79
6	300	0	0	0	0	43	399	4	0.11	184	4.89
7	300	0	0	0	0	44	401	44	1.17	228	6.06
8	300	0	0	0	0	45	403	17	0.45	245	6.52
9	300	0	0	0	0	46	405	40	1.06	285	7.58
10	300	0	0	0	0	47	407	16	0.43	301	8.01
11	300	0	0	0	0	48	409	53	1.41	354	9.41
12	300	0	0	0	0	49	411	23	0.61	377	10.03
13	300	0	0	0	0	50	413	50	1.33	427	11.36
14	300	0	0	0	0	51	415	17	0.45	444	11.81
15	300	0	0	0	0	52	417	49	1.3	493	13.11
16	300	0	0	0	0	53	419	31	0.82	524	13.94
17	311	0	0	0	0	54	421	50	1.33	574	15.27
18	321	1	0.03	1	0.03	55	423	41	1.09	615	16.36
19	329	0	0	1	0.03	56	425	39	1.04	654	17.39
20	335	1	0.03	2	0.05	57	427	42	1.12	696	18.51
21	341	0	0	2	0.05	58	429	70	1.86	766	20.37
22	345	5	0.13	7	0.19	59	430	49	1.3	815	21.68
23	349	0	0	7	0.19	60	432	69	1.84	884	23.51
24	353	3	0.08	10	0.27	**61**	**434**	**45**	**1.2**	**929**	**24.71**
25	357	1	0.03	11	0.29	62	436	66	1.76	995	26.46
26	360	6	0.16	17	0.45	63	438	62	1.65	1057	28.11
27	363	0	0	17	0.45	64	440	90	2.39	1147	30.51
28	366	1	0.03	18	0.48	65	442	52	1.38	1199	31.89
29	368	4	0.11	22	0.59	66	444	107	2.85	1306	34.73
30	371	9	0.24	31	0.82	67	446	53	1.41	1359	36.14
31	373	5	0.13	36	0.96	68	448	113	3.01	1472	39.15
32	376	1	0.03	37	0.98	69	450	53	1.41	1525	40.56
33	378	6	0.16	43	1.14	70	452	124	3.3	1649	43.86
34	381	10	0.27	53	1.41	71	454	56	1.49	1705	45.35
35	383	2	0.05	55	1.46	72	456	100	2.66	1805	48.01
36	385	18	0.48	73	1.94	73	458	58	1.54	1863	49.55

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

(table continued)

Table 8.1.1.9 (Continued)
Spring 2015 AIMS Frequency Distribution
Writing High School

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
74	460	103	2.74	1966	52.29	107	545	25	0.66	3453	91.84
75	462	69	1.84	2035	54.12	108	549	12	0.32	3465	92.15
76	464	100	2.66	2135	56.78	109	553	19	0.51	3484	92.66
77	466	76	2.02	2211	58.80	110	556	8	0.21	3492	92.87
78	469	81	2.15	2292	60.96	111	560	26	0.69	3518	93.56
79	471	69	1.84	2361	62.79	112	563	12	0.32	3530	93.88
80	473	68	1.81	2429	64.60	113	567	15	0.40	3545	94.28
81	475	51	1.36	2480	65.96	114	571	14	0.37	3559	94.65
82	477	56	1.49	2536	67.45	115	575	18	0.48	3577	95.13
83	479	58	1.54	2594	68.99	116	579	24	0.64	3601	95.77
84	**481**	**61**	**1.62**	**2655**	**70.61**	117	583	6	0.16	3607	95.93
85	484	54	1.44	2709	72.05	**118**	**587**	**26**	**0.69**	**3633**	**96.62**
86	486	52	1.38	2761	73.43	119	591	6	0.16	3639	96.78
87	488	37	0.98	2798	74.41	120	595	43	1.14	3682	97.93
88	490	58	1.54	2856	75.96	121	600	3	0.08	3685	98.01
89	493	27	0.72	2883	76.68	122	605	16	0.43	3701	98.43
90	495	55	1.46	2938	78.14	123	610	4	0.11	3705	98.54
91	498	26	0.69	2964	78.83	124	615	9	0.24	3714	98.78
92	500	49	1.30	3013	80.13	125	620	11	0.29	3725	99.07
93	503	23	0.61	3036	80.74	126	626	0	0.00	3725	99.07
94	505	58	1.54	3094	82.29	127	631	4	0.11	3729	99.18
95	508	14	0.37	3108	82.66	128	638	2	0.05	3731	99.23
96	511	54	1.44	3162	84.10	129	644	10	0.27	3741	99.49
97	514	10	0.27	3172	84.36	130	651	0	0.00	3741	99.49
98	517	46	1.22	3218	85.59	131	659	4	0.11	3745	99.60
99	519	6	0.16	3224	85.74	132	667	4	0.11	3749	99.71
100	523	48	1.28	3272	87.02	133	676	0	0.00	3749	99.71
101	526	8	0.21	3280	87.23	134	687	4	0.11	3753	99.81
102	529	49	1.30	3329	88.54	135	700	0	0.00	3753	99.81
103	532	13	0.35	3342	88.88	136	700	6	0.16	3759	99.97
104	535	32	0.85	3374	89.73	137	700	0	0.00	3759	99.97
105	539	17	0.45	3391	90.19	138	700	1	0.03	3760	100.00
106	542	37	0.98	3428	91.17						

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.10
Spring 2015 AIMS Frequency Distribution
Science Grade 4

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	0	0.00	0	0.00	**28**	**504**	**2673**	**3.18**	**37839**	**44.99**
1	294	0	0.00	0	0.00	29	508	2784	3.31	40623	48.30
2	329	0	0.00	0	0.00	30	512	2635	3.13	43258	51.43
3	350	2	0.00	2	0.00	31	516	2658	3.16	45916	54.59
4	366	2	0.00	4	0.00	32	520	2829	3.36	48745	57.95
5	378	9	0.01	13	0.02	33	524	2719	3.23	51464	61.18
6	389	23	0.03	36	0.04	34	528	2675	3.18	54139	64.36
7	398	56	0.07	92	0.11	35	533	2775	3.30	56914	67.66
8	406	123	0.15	215	0.26	36	537	2720	3.23	59634	70.90
9	413	223	0.27	438	0.52	37	542	2637	3.14	62271	74.03
10	420	374	0.44	812	0.97	**38**	**547**	**2583**	**3.07**	**64854**	**77.10**
11	426	614	0.73	1426	1.70	39	552	2524	3.00	67378	80.10
12	432	870	1.03	2296	2.73	40	557	2385	2.84	69763	82.94
13	437	1172	1.39	3468	4.12	41	562	2279	2.71	72042	85.65
14	442	1466	1.74	4934	5.87	42	567	2079	2.47	74121	88.12
15	448	1701	2.02	6635	7.89	43	573	1900	2.26	76021	90.38
16	452	1912	2.27	8547	10.16	44	580	1780	2.12	77801	92.50
17	457	2057	2.45	10604	12.61	45	586	1610	1.91	79411	94.41
18	**462**	**2209**	**2.63**	**12813**	**15.23**	46	594	1310	1.56	80721	95.97
19	466	2227	2.65	15040	17.88	47	602	1132	1.35	81853	97.31
20	470	2290	2.72	17330	20.60	48	611	862	1.02	82715	98.34
21	475	2405	2.86	19735	23.46	49	621	577	0.69	83292	99.02
22	479	2542	3.02	22277	26.48	50	634	419	0.50	83711	99.52
23	483	2528	3.01	24805	29.49	51	649	224	0.27	83935	99.79
24	487	2509	2.98	27314	32.47	52	671	127	0.15	84062	99.94
25	491	2607	3.10	29921	35.57	53	706	46	0.05	84108	99.99
26	495	2616	3.11	32537	38.68	54	800	5	0.01	84113	100.00
27	499	2629	3.13	35166	41.81						

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.11
Spring 2015 AIMS Frequency Distribution
Science Grade 8

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	0	0.00	0	0.00	30	498	2341	2.85	34030	41.37
1	284	0	0.00	0	0.00	**31**	**502**	**2391**	**2.91**	**36421**	**44.28**
2	319	0	0.00	0	0.00	32	505	2449	2.98	38870	47.26
3	340	0	0.00	0	0.00	33	509	2530	3.08	41400	50.34
4	355	1	0.00	1	0.00	34	513	2525	3.07	43925	53.41
5	367	2	0.00	3	0.00	35	517	2473	3.01	46398	56.41
6	378	4	0.00	7	0.01	36	521	2561	3.11	48959	59.53
7	387	37	0.04	44	0.05	37	525	2492	3.03	51451	62.56
8	395	51	0.06	95	0.12	38	529	2507	3.05	53958	65.60
9	402	94	0.11	189	0.23	**39**	**533**	**2532**	**3.08**	**56490**	**68.68**
10	408	174	0.21	363	0.44	40	538	2497	3.04	58987	71.72
11	415	314	0.38	677	0.82	41	542	2338	2.84	61325	74.56
12	420	472	0.57	1149	1.40	42	547	2468	3.00	63793	77.56
13	426	680	0.83	1829	2.22	43	552	2216	2.69	66009	80.26
14	431	865	1.05	2694	3.28	44	556	2222	2.70	68231	82.96
15	436	1070	1.30	3764	4.58	45	562	2116	2.57	70347	85.53
16	441	1300	1.58	5064	6.16	46	567	1949	2.37	72296	87.90
17	445	1504	1.83	6568	7.99	47	573	1793	2.18	74089	90.08
18	450	1541	1.87	8109	9.86	48	579	1592	1.94	75681	92.02
19	454	1792	2.18	9901	12.04	49	585	1461	1.78	77142	93.79
20	459	1910	2.32	11811	14.36	50	593	1243	1.51	78385	95.30
21	463	1950	2.37	13761	16.73	51	601	1069	1.30	79454	96.60
22	467	2044	2.49	15805	19.22	52	609	927	1.13	80381	97.73
23	471	2100	2.55	17905	21.77	53	620	724	0.88	81105	98.61
24	**475**	**2198**	**2.67**	**20103**	**24.44**	54	632	472	0.57	81577	99.18
25	479	2281	2.77	22384	27.22	55	647	363	0.44	81940	99.63
26	482	2193	2.67	24577	29.88	56	668	190	0.23	82130	99.86
27	486	2405	2.92	26982	32.81	57	703	95	0.12	82225	99.97
28	490	2367	2.88	29349	35.68	58	800	23	0.03	82248	100.00
29	494	2340	2.85	31689	38.53						

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.12
Spring 2015 AIMS Frequency Distribution
Science High School Cohort 17

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	4	0.01	4	0.01	33	493	1430	2.81	32043	62.86
1	262	2	0.00	6	0.01	34	496	1369	2.69	33412	65.55
2	300	2	0.00	8	0.02	**35**	**500**	**1320**	**2.59**	**34732**	**68.14**
3	322	1	0.00	9	0.02	36	503	1191	2.34	35923	70.47
4	338	1	0.00	10	0.02	37	507	1182	2.32	37105	72.79
5	351	1	0.00	11	0.02	38	511	1104	2.17	38209	74.96
6	362	6	0.01	17	0.03	39	514	1112	2.18	39321	77.14
7	372	17	0.03	34	0.07	40	518	1041	2.04	40362	79.18
8	380	38	0.07	72	0.14	41	522	1021	2.00	41383	81.18
9	387	73	0.14	145	0.28	42	525	933	1.83	42316	83.01
10	394	111	0.22	256	0.50	43	529	915	1.79	43231	84.81
11	401	250	0.49	506	0.99	44	533	843	1.65	44074	86.46
12	407	382	0.75	888	1.74	**45**	**537**	**805**	**1.58**	**44879**	**88.04**
13	412	585	1.15	1473	2.89	46	541	758	1.49	45637	89.53
14	417	783	1.54	2256	4.43	47	546	669	1.31	46306	90.84
15	422	1029	2.02	3285	6.44	48	550	673	1.32	46979	92.16
16	427	1208	2.37	4493	8.81	49	554	587	1.15	47566	93.31
17	432	1443	2.83	5936	11.64	50	559	554	1.09	48120	94.40
18	436	1575	3.09	7511	14.73	51	564	467	0.92	48587	95.32
19	441	1690	3.32	9201	18.05	52	569	417	0.82	49004	96.13
20	445	1721	3.38	10922	21.43	53	575	417	0.82	49421	96.95
21	449	1776	3.48	12698	24.91	54	580	326	0.64	49747	97.59
22	453	1784	3.50	14482	28.41	55	587	281	0.55	50028	98.14
23	457	1747	3.43	16229	31.84	56	593	232	0.46	50260	98.60
24	460	1722	3.38	17951	35.22	57	601	212	0.42	50472	99.01
25	464	1715	3.36	19666	38.58	58	609	170	0.33	50642	99.35
26	468	1680	3.30	21346	41.88	59	618	121	0.24	50763	99.58
27	472	1643	3.22	22989	45.10	60	629	94	0.18	50857	99.77
28	**475**	**1609**	**3.16**	**24598**	**48.26**	61	642	46	0.09	50903	99.86
29	479	1563	3.07	26161	51.32	62	658	31	0.06	50934	99.92
30	482	1495	2.93	27656	54.25	63	680	29	0.06	50963	99.98
31	486	1464	2.87	29120	57.13	64	717	9	0.02	50972	99.99
32	489	1493	2.93	30613	60.05	65	800	3	0.01	50975	100.00

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

Table 8.1.1.13
Spring 2015 AIMS Frequency Distribution
Science High School Cohort 18

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	0	0.00	0	0.00	33	493	756	2.60	13061	44.94
1	262	0	0.00	0	0.00	34	496	785	2.70	13846	47.64
2	300	1	0.00	1	0.00	**35**	**500**	**793**	**2.73**	**14639**	**50.37**
3	322	1	0.00	2	0.01	36	503	755	2.60	15394	52.97
4	338	0	0.00	2	0.01	37	507	759	2.61	16153	55.58
5	351	1	0.00	3	0.01	38	511	780	2.68	16933	58.26
6	362	2	0.01	5	0.02	39	514	826	2.84	17759	61.11
7	372	4	0.01	9	0.03	40	518	805	2.77	18564	63.88
8	380	13	0.04	22	0.08	41	522	758	2.61	19322	66.48
9	387	19	0.07	41	0.14	42	525	704	2.42	20026	68.91
10	394	41	0.14	82	0.28	43	529	751	2.58	20777	71.49
11	401	61	0.21	143	0.49	44	533	725	2.49	21502	73.98
12	407	114	0.39	257	0.88	**45**	**537**	**691**	**2.38**	**22193**	**76.36**
13	412	168	0.58	425	1.46	46	541	637	2.19	22830	78.55
14	417	241	0.83	666	2.29	47	546	659	2.27	23489	80.82
15	422	347	1.19	1013	3.49	48	550	648	2.23	24137	83.05
16	427	376	1.29	1389	4.78	49	554	607	2.09	24744	85.14
17	432	455	1.57	1844	6.34	50	559	595	2.05	25339	87.19
18	436	516	1.78	2360	8.12	51	564	514	1.77	25853	88.96
19	441	581	2.00	2941	10.12	52	569	532	1.83	26385	90.79
20	445	629	2.16	3570	12.28	53	575	471	1.62	26856	92.41
21	449	605	2.08	4175	14.37	54	580	415	1.43	27271	93.83
22	453	628	2.16	4803	16.53	55	587	356	1.22	27627	95.06
23	457	698	2.40	5501	18.93	56	593	310	1.07	27937	96.13
24	460	723	2.49	6224	21.42	57	601	286	0.98	28223	97.11
25	464	731	2.52	6955	23.93	58	609	281	0.97	28504	98.08
26	468	739	2.54	7694	26.47	59	618	204	0.70	28708	98.78
27	472	774	2.66	8468	29.14	60	629	153	0.53	28861	99.30
28	**475**	**761**	**2.62**	**9229**	**31.76**	61	642	77	0.26	28938	99.57
29	479	758	2.61	9987	34.36	62	658	69	0.24	29007	99.81
30	482	759	2.61	10746	36.97	63	680	37	0.13	29044	99.93
31	486	778	2.68	11524	39.65	64	717	16	0.06	29060	99.99
32	489	781	2.69	12305	42.34	65	800	3	0.01	29063	100.00

Note: Freq. = Frequency, Cum = Cumulative. Students with no valid attempt, invalidation or off-grade are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections facilities, and students attending hospital schools are not included in this summary.

8.2 Longitudinal Data

The spring 2008 administration represents the baseline year for the AIMS Science assessment. In this section, the spring 2015 results are presented along with results back to 2008 to provide longitudinal information. Tables 8.2.1 and 8.2.2 include scale score descriptive statistics and performance level distributions for the AIMS Science administrations.

In this section, the spring 2015 results are presented along with the longitudinal results going back to the base administration for each subject. Tables 8.2.1 - 8.2.2 include scale score descriptive statistics and performance level distributions for the AIMS administration from each year. Caution should be taken when interpreting year-to-year or grade-to-grade comparisons, as slight differences in exclusion rules, changes in the manner in which accommodations were identified, and changes in the manner in which high school results were separated may result in different student population characteristics reported in Tables 8.2.1-8.2.2.

Starting spring 2015, the longitudinal data for Reading, Mathematics, and Writing in high school is no longer presented because students are no longer required to take these tests.

Table 8.2.1
Longitudinal Comparison of Scale Scores in Science

			Scale Score		Percentiles				
Grade	Year	N	M	SD	P10	P25	P50	P75	P90
4	2008	80296	501.8	50.2	436	466	503	536	567
	2009	81724	508.2	50.5	443	475	508	540	567
	2010	80982	513.8	52.7	446	478	515	547	583
	2011	81934	534.8	61.7	455	492	536	575	615
	2012	81892	518.9	57.6	448	478	514	554	589
	2013	83028	513.4	51.9	445	477	511	549	581
	2014	83408	513.5	46.6	457	480	510	546	574
	2015	84113	513.8	46.5	452	479	512	547	573
8	2008	79482	500.6	50.0	435	463	498	534	568
	2009	78703	506.4	50.0	439	471	506	539	571
	2010	79293	510.4	51.5	446	473	508	545	578
	2011	79409	517.7	47.6	454	484	521	551	578
	2012	80019	519.3	47.9	456	487	521	553	581
	2013	81485	516.7	43.1	459	486	518	544	571
	2014	82470	516.7	45.7	459	483	516	546	573
	2015	82248	513.0	48.1	454	479	509	547	573
HS	2008 (Cohort 10)	45286	477.3	50.1	414	440	475	510	543
	2009 (Cohort 11)	51195	475.8	49.7	410	439	477	508	541
	2010(Cohort 12)	53671	479.1	51.8	414	442	474	512	545
	2011(Cohort 13)	54610	484.6	58.3	407	443	484	524	559
	2011(Cohort 14)	19392	523.7	58.8	446	488	524	559	596
	2012(Cohort 14)	53344	487.0	62.6	403	441	487	528	569
	2012(Cohort 15)	21142	526.3	65.4	441	487	528	569	603
	2013(Cohort 15)	52650	485.7	56.0	414	442	482	521	562
	2013(Cohort 16)	24094	517.3	59.0	438	475	517	556	591
	2014(Cohort 16)	50096	487.2	52.9	421	448	484	522	555
	2014(Cohort 17)	26254	514.5	53.0	445	477	514	550	582
	2015(Cohort 17)	50975	484.2	44.7	432	453	479	514	546
	2015(Cohort 18)	29063	504.2	49.3	441	468	500	537	569

Note: Students without a valid attempt, invalidation, off-grade, a non-standard accommodation (not in 2008), home-schooled students, attending Bureau of Indian Affairs schools, attending juvenile corrections centers (not in 2005), attending state hospital schools (not in 2005), and who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. Caution should be used when interpreting results across years, as exclusion rules differ slightly and high school identification of grade versus cohort may result in different student population characteristics.

Table 8.2.2
Longitudinal Comparison of Performance Level Distribution in Science

			% at Performance Level			
Grade	Year	N	FFBS	AS	MS	ES
4	2008	80296	22	25	35	18
	2009	81724	17	26	36	21
	2010	80982	17	22	33	28
	2011	81934	12	17	29	43
	2012	81892	16	21	31	32
	2013	83028	17	25	32	26
	2014	83408	12	29	36	22
	2015	84113	13	29	32	26
8	2008	79482	31	20	22	28
	2009	78703	26	19	23	32
	2010	79293	23	18	25	34
	2011	79409	17	17	27	39
	2012	80019	18	15	28	40
	2013	81485	16	18	29	37
	2014	82470	18	20	24	38
	2015	82248	22	20	24	34
HS	2008 (Cohort 10)	45286	49	19	20	12
	2009 (Cohort 11)	51195	50	18	22	11
	2010 (Cohort 12)	53671	50	16	21	14
	2011 (Cohort 13)	54610	43	15	23	18
	2011 (Cohort 14)	19392	19	12	27	41
	2012(Cohort 14)	53344	41	17	21	21
	2012(Cohort 15)	21142	20	14	23	43
	2013(Cohort 15)	52650	44	17	21	18
	2013(Cohort 16)	24094	23	15	25	36
	2014(Cohort 16)	50096	44	17	21	18
	2014(Cohort 17)	26254	24	16	27	33
	2015(Cohort 17)	50975	45	20	21	14
	2015(Cohort 18)	29063	29	19	26	26

Note: Students without a valid attempt, invalidation, off-grade, a non-standard accommodation (not in 2008), home-schooled students, attending Bureau of Indian Affairs schools, attending juvenile corrections centers (not in 2005), attending state hospital schools (not in 2005), and who already met expectations in a previous test administration are not included in this summary. These results are not final results and are presented here for purposes of addressing reliability and validity. Caution should be used when interpreting results across years, as exclusion rules differ slightly and high school identification of grade versus cohort may result in different student population characteristics.

PART 9: VALIDITY EVIDENCE

Part 9 of the technical report provides evidence supporting the reliability and validity of the 2015 AIMS Science assessments in grades 4, 8, and high school and the AIMS Reading, Mathematics, and Writing assessments for high school. All data presented in this section were computed using population test data available in the final electronic data files. The following AERA/APA/NCME *Standards* from the 1999 edition 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. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this chapter are: 1.8, 1.9, 2.3, 2.7, 2.8, 2.19, 3.3, 3.6, 4.4, 5.19 and 7.4.

9.1 Reliability

AERA/APA/NCME *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) refer to reliability as the "consistency of [a measure] when the testing procedure is repeated on a population of individuals or groups." The 2014 edition of AERA/APA/NCME *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) indicates that "The term *reliability* has been used in two ways in the measurement literature. First, the term has been used to refer to the reliability coefficients of classical test theory, refined as the correlation between scores on two equivalent forms of the test, presuming that taking one form has no effect on performance on the second form. Second, the term has been used in a more general sense, to refer to the consistency of scores across relications of a testing procedure, regardless of how this consistency is estimated or reported (e.g., in terms of standard errors, reliability coefficients per se, generalizability coefficient, error/tolerance ratios, item response theory (IRT) information functions, or various indices of classification consistency)".

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 Spring 2015 AIMS assessments was estimated in two ways: internal consistency for all multiple-choice tests and reliability of hand scoring for all writing tests.

9.1.1 Measures of Internal Consistency

For tests consisting of only constructed response or multiple-choice items, Cronbach's alpha is a frequently used 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.

For tests consisting of a mixture of item types with differing numbers of maximum points, it is more appropriate to use stratified alpha (Feldt and Brennan, 1989). The stratified alpha is a weighted average of Cronbach's alpha for the different components of the test. This is the measure of internal consistency reported for the writing tests. The formula for calculating stratified alpha is

$$stratified\ \hat{\alpha} = 1 - \frac{\sum \sigma_{x_j}^2 \left(1 - \alpha_j\right)}{\sigma_x^2}$$

where $\sigma_{x_j}^2$ is the variance of component *j*, α_j is the reliability of component *j*, and σ_x^2 is the variance of the total test.

Reliability estimates for the multiple-choice tests administered as part of the Spring 2015 AIMS Science assessment are presented in Table 9.1.1.1. Note that a high degree of internal consistency is evident for all CRT tests. Reliability estimates for the AIMS Reading, Mathematics, and Writing assessments for high school were not generated because the tests were pre-equated for the spring 2015 administration. Please refer to the 2014 AIMS Technical Report (Arizona Department of Education, 2014) for the IRT item analysis statistics in the AIMS Reading and Mathemetics assessments in high school because the spring 2015 tests were reused forms from spring 2014. Please also refer to the 2012 AIMS Technical Report (Arizona Department of Education, 2012) for reliability estimates for the writing multiple-choice portion since the multiple-choice items were reused items from the spring 2012 administration.

In addition to the total test score, raw scores are reported at the Arizona strand and concept level. Internal consistency reliability estimates at the strand and concept level for science are reported in Tables 9.1.1.2 through 9.1.1.4.

Table 9.1.1.1
Spring 2015 AIMS Internal Consistency for Science

Subgroup	Value	N	Alpha
Grade 4			
	Hispanic	37420	0.87
	Non-Hisp	46031	0.89
Ethnicity	White	66593	0.89
Ethnicity	Black	4752	0.87
Ethnicity	Asian	2523	0.89
Ethnicity	Am Indian	6023	0.83
Ethnicity	Hawaii	309	0.88
Ethnicity	Multiple	2932	0.88
Gender	Female	41331	0.89
Gender	Male	42538	0.90
ELL	Yes	7670	0.72
SPED	Yes	10882	0.87
Low SES	Yes	47898	0.87
Migrant	Yes	458	0.84
TOTAL		83905	0.89
Grade 8			
	Hispanic	36009	0.88
	Non-Hisp	45762	0.91
Ethnicity	White	66182	0.90
Ethnicity	Black	4840	0.88
Ethnicity	Asian	2457	0.91
Ethnicity	Am Indian	5538	0.86
Ethnicity	Hawaii	290	0.90
Ethnicity	Multiple	2219	0.90
Gender	Female	40385	0.89
Gender	Male	41604	0.91
ELL	Yes	2552	0.67
SPED	Yes	8579	0.85
Low SES	Yes	43965	0.88
Migrant	Yes	457	0.88
TOTAL		82042	0.90
HS			
	Hispanic	34323	0.87
	Non-Hisp	44278	0.91
Ethnicity	White	63170	0.91
Ethnicity	Black	4662	0.88
Ethnicity	Asian	2387	0.92
Ethnicity	Am Indian	4629	0.84
Ethnicity	Hawaii	324	0.87
Ethnicity	Multiple	2291	0.90
Gender	Female	39217	0.90
Gender	Male	40161	0.91
ELL	Yes	1450	0.61
SPED	Yes	7192	0.83
Low SES	Yes	38999	0.87
Migrant	Yes	475	0.78
TOTAL		79636	0.91

Table 9.1.1.2
Spring 2015 AIMS Strand/Concept Internal Consistency
Science Grade 4

Strand	Concept	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry		18	84113	10.38	3.67	0.74
	Concept 1: Observations, Questions, and Hypotheses	6	84113	3.60	1.64	0.58
	Concept 2: Scientific Testing (Investigating and Modeling)	6	84113	3.65	1.39	0.40
	Concept 3/4: Analysis and Conclusions/Communication	6	84113	3.13	1.51	0.46
2. History and Nature of Science		6	84113	3.92	1.63	0.61
	Concept 1/2: History of Science as a Human Endeavor/Nature of Scientific Knowledge	6	84113	3.92	1.63	0.61
3. Science in Personal and Social Perspectives		6	84113	3.46	1.62	0.53
	Concept 1/2: Changes in Environments/Science and Technology in Society	6	84113	3.46	1.62	0.53
4. Life Science		6	84113	3.73	1.56	0.56
	Concept 1/3/4: Characteristics of Organisms/Organisms and Environments/Diversity, Adaptation, and Behavior	6	84113	3.73	1.56	0.56
5. Physical Science		6	84113	3.03	1.58	0.52
	Concept 3: Energy and Magnetism	6	84113	3.03	1.58	0.52
6. Earth and Space Science		12	84113	5.44	2.38	0.56
	Concept 2: Earth's Processes and Systems	6	84113	2.51	1.53	0.49
	Concept 3: Changes in the Earth and Sky	6	84113	2.93	1.37	0.32

Table 9.1.1.3
Spring 2015 AIMS Strand/Concept Internal Consistency
Science Grade 8

Strand	Concept	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry		20	82248	11.99	3.95	0.75
	Concept 1: Observations, Questions, and Hypotheses	6	82248	3.55	1.57	0.51
	Concept 2: Scientific Testing (Investigating and Modeling)	4	82248	2.33	1.01	0.32
	Concept 3: Analysis, Conclusions, and Refinements	6	82248	3.32	1.52	0.47
	Concept 4: Communication	4	82248	2.79	1.08	0.39
2. History and Nature of Science		6	82248	2.96	1.66	0.56
	Concept 1/2: History of Science as a Human Endeavor/Nature of Scientific Knowledge	6	82248	2.96	1.66	0.56
3. Science in Personal and Social Perspectives		6	82248	3.39	1.57	0.50
	Concept 1/2: Changes in Environments/Science and Technology in Society	6	82248	3.39	1.57	0.50
4. Life Science		8	82248	5.01	1.99	0.64
	Concept 2/4: Reproduction and Heredity/Diversity, Adaptation, and Behavior	8	82248	5.01	1.99	0.64
5. Physical Science		18	82248	9.83	3.49	0.73
	Concept 1: Properties and Changes of Properties in Matter	10	82248	5.21	2.08	0.57
	Concept 2: Motion and Forces	8	82248	4.62	1.88	0.59

Table 9.1.1.4
Spring 2015 AIMS Strand/Concept Internal Consistency
Science High School

Strand	Concept	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry		22	80038	11.43	4.64	0.79
	Concept 1: Observations, Questions, and Hypotheses	6	80038	3.23	1.65	0.56
	Concept 2: Scientific Testing (Investigating and Modeling)	6	80038	2.80	1.54	0.46
	Concept 3: Analysis, Conclusions, and Refinements	6	80038	3.35	1.59	0.53
	Concept 4: Communication	4	80038	2.05	1.12	0.38
2. History and Nature of Science		6	80038	3.50	1.55	0.50
	Concept 1/2: History of Science as a Human Endeavor/Nature of Scientific Knowledge	6	80038	3.50	1.55	0.50
3. Science in Personal and Social Perspectives		7	80038	3.07	1.67	0.50
	Concept 1/2/3: Changes in Environments/Science and Technology in Society/Human Population Characteristics	7	80038	3.07	1.67	0.50
4. Life Science		30	80038	14.37	5.62	0.81
	Concept 1: The Cell	6	80038	2.38	1.50	0.45
	Concept 2: Molecular Basis of Heredity	6	80038	3.12	1.65	0.57
	Concept 3: Interdependence of Organisms	6	80038	3.45	1.59	0.57
	Concept 4: Biological Evolution	6	80038	2.95	1.40	0.38
	Concept 5: Matter, Energy, and Organization in Living Systems (Including Human Systems)	6	80038	2.48	1.45	0.40

9.1.2 Reliability of Constructed-Response Scoring

For constructed response items, the consistency with which two raters assign scores to student responses is typically determined by inter-rater agreement. In scoring the essay, each student essay is randomly assigned to a rater. All essays are scored by two raters at the high school level. Since different raters scored different essays, the inter-rater statistics computed here do not measure the degrees of agreement or disagreements between the same two raters across the entire set of essay responses. Therefore, it is more accurate to describe the inter-rater agreement reported in this section as inter-rater position reliability.

For the high school writing prompts, each student paper was scored by two independent raters for all students. The statistics for the prompt are presented in Table 9.1.2.1 and were calculated using the scores from both raters. The two scores for each essay were used in the analyses to calculate rater agreement.

The raw score means, raw score standard deviations, and percentage of agreement between the first and second rater were computed. Perfect agreement is defined as trait scores that are exactly the same between the first and second rating. Adjacent agreement is defined as trait scores differing by one point between the first and second rating. Discrepant cases include records in which scores from the first and the second rating differed by more than one point. In addition, Cohen's kappa and intraclass correlation are provided as indices of agreement between the first and second rating.

Cohen's kappa (Cohen, 1960) is commonly used to summarize the agreement between raters and is computed as (Brennan & Prediger, 1981):

$$\kappa = \frac{\sum P_{ii} - \sum P_{i\cdot} P_{\cdot i}}{1 - \sum P_{i\cdot} P_{\cdot i}},$$

where $\sum P_{ii}$ is the observed proportion of agreement and $\sum P_{i\cdot} P_{\cdot i}$ is the chance proportion of agreement.

Intraclass correlation is defined by Shrout and Fleiss (1979) as "the correlation between one measurement on a target and another measurement obtained on that target." In the context of the Spring 2015 AIMS Writing test, the "target" is the student response, and each measurement was obtained by a randomly assigned rater to that response. Therefore, ICC(1,1) was used to estimate intraclass correlation. ICC(1,1) is estimated as (Shrout & Fleiss, 1979):

$$ICC(1,1) = \frac{BMS - WMS}{BMS + (k-1)WMS},$$

where *BMS* is between-targets mean square, *WMS* is within-targets mean square, and *k* is the number of raters rating each target.

Table 9.1.2.1
Spring 2015 AIMS Inter-Rater Position Consistency
High School

		MAX	Rater 1		Rater 2		% Agreement			Wgt	Interclass
	N	Points	M	SD	M	SD	Perfect	Adjacent	Discrepant	Kappa	Correlation
Form A	3760	12	6.38	1.89	6.50	1.88	62.29	36.38	1.33	0.46	0.77

Note: All student responses were rated by the two raters. Only students receiving a valid score or a condition codes of Blank, Illegible, or off Topic are included in this analysis.

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). The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for particular purposes or uses. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process.

The Spring 2015 AIMS tests were designed and developed to provide fair and accurate ability scores that support appropriate, meaningful, and useful educational decisions. In addition to the evidence provided in Part 2 (Involvement of Arizona Educators), additional validity evidence may be found in the following parts as described: Part 3 (Test Design), Part 4 (Test Development), Part 5 (Test Administration), Part 6 (Classical Item Analysis), Part 7 (Calibration, Scaling and Equating), Part 9.1 (Reliability), 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, 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 Chapters 2 through 10. Below is a brief review.

Part 2 of the technical report described the involvement of Arizona educators, ADE, and Pearson 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. The knowledge, expertise, and professional judgment offered by Arizona educators ultimately ensured that the content of AIMS 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 addressed the issue of test form development. Part 3 provided a general discussion of test book creation and editing process, the process of selecting operational test items, the content distribution of embedded field test items, and the process of obtaining ADE approvals. The test design process and the participation of Arizona educators in the process of test selection, including item content and bias review, provide a solid rationale for having confidence in the content and design of AIMS as a tool from which to derive valid inferences about Arizona student performance.

Part 5 of the technical report described the process, procedures, and policies that guided the administration of the AIMS, including accommodations, security, and the written procedures provided to test administrators and school personnel.

Part 6 described classical data analysis of the Spring 2015 AIMS.

Part 7 of the technical report described the calibration, scaling 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 dealt with the test results, longitudinal comparisons, score distributions and performance levels.

Part 9 above dealt with inter-rater position consistency, interclass correlation, rater agreement, and alpha reliabilities and internal consistency. It described Cronbach's alpha as a measure for

internal consistency for reading, mathematics, science, and inter-rater position consistency for writing.

Part 9 below presents an analysis of DIF (Differential Item Functioning) complete with tables of gender and ethnic functioning of all operational items for the 2015 AIMS assessments.

Part 10 of the technical report will describe a detailed analysis of classification consistency and classification accuracy.

Additional evidence to support the validity of the 2015 AIMS assessments is provided by the following:

- Any items that displayed differential item functioning for subgroups of ethnicity and gender were identified.
- Correlations between scores on the 2015 AIMS tests for each grade level as construct validity were presented.

Also note that further evidence in support of the AIMS assessment has been documented in previous AIMS technical reports and standard setting technical reports.

9.2.1 Differential Item Functioning

Because test scores can have many sources of variation, the test publishers' task is to develop assessments that measure the intended abilities and skills without introducing extraneous elements or construct irrelevant variance. When tests measure something other than what they are intended to measure, test scores will reflect these unintended skills and knowledge, as well as what is purportedly assessed by the test. If this occurs, these tests can be called biased (Angoff, 1993; Camilli & Shepard, 1994; Green, 1975). One of the factors that may render test scores to be biased is differing cultural and socioeconomic experiences.

The Spring 2015 AIMS tests were developed using procedures to minimize item and test bias and included reviews such as the Content and Sensitivity Reviews described in Part 4, Test Development. Expertise in this area is not, however, a substitute for statistical analyses of the items. Thus, an empirical differential item functioning (DIF) approach was used to examine potential item bias. DIF studies include systematic item analyses to determine if examinees with the same underlying level of ability have the same probability of correctly responding to the item. Items identified with DIF are further examined to determine if item performance differences between identifiable subgroups of the population are due to extraneous or construct irrelevant information which makes the items unfairly difficult for one of the subgroups.

DIF analyses of the Spring 2015 AIMS tests were conducted for ethnic subgroups and gender. In order to compute DIF, students must be matched on ability level using a conditioning variable. For these analyses, raw score on the CRT test in the content area of interest was used as the conditioning variable. Note that DIF analyses were conducted on all multiple-choice science items. The DIF analyses for AIMS Reading, Mathematics, and Writing in high school were not conducted since the tests were pre-equated for the spring 2015 administration. Please refer to the 2014 AIMS Technical Report (Arizona Department of Education, 2014) for the DIF statistics for reading and mathemetics in high school because the spring 2015 tests were reused forms from spring 2014. Please also refer to the 2012 AIMS Technical Report (Arizona Department of Education, 2012) for the DIF statistics for the writing multiple-choice items in high school because the multiple-choice items were reused items from the spring 2012 administration.

The Mantel-Haenszel chi-square statistic was used to identify DIF in multiple-choice items. The Mantel-Haenszel statistic was first recommended by Holland and Thayer (1988), is frequently used,

and is efficient in terms of statistical power (Clauser & Mazor, 1998). The Mantel-Haenszel statistic is computed as (Zwick, Donoghue, & Grima, 1993):

$$\text{Mantel } \chi^2 = \frac{\left(\sum_k F_k - \sum_k E(F_k)\right)^2}{\sum_k \text{Var}(F_k)},$$

where F_k is the sum of scores for the focal group at the *k*th level of the matching variable. Note that the Mantel-Haenszel statistic is sensitive to *N* such that larger sample sizes increase the value of chi square.

In addition to the Mantel-Haenszel chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. Educational Testing Service (ETS) first developed the MH-D DIF statistic. To compute delta, alpha (the odds ratio) is first computed as:

$$\alpha_{MH} = \frac{\sum_{k=1}^{K} N_{r1k} N_{f0k} / N_k}{\sum_{k=1}^{K} N_{f1k} N_{r0k} / N_k},$$

where N_{r1k} is the number of correct responses in the reference group at ability level *k*, N_{f0k} is the number of incorrect responses in the focal group at ability level *k*, N_k is the total number of responses, N_{f1k} is the number of correct responses in the focal group at ability level *k*, and N_{r0k} is the number of incorrect responses in the reference group at ability level *k*. MH-D DIF is then computed as:

$$\text{MH-D DIF} = -2.35 \ln(\alpha_{MH}).$$

Positive values of MH-D DIF indicate items that favor the focal group, whereas negative values of MH-D DIF indicate items that favor the reference group.

The Mantel-Haenszel chi-square statistic and the delta statistic were used in combination to identify the Spring 2015 AIMS items that exhibit strong, weak, or no DIF (Zieky, 1993). Table 9.2.1.1 indicates the criteria for each category used for the 2015 AIMS DIF analysis. An alpha level of .01 was used for all Mantel-Haenszel statistics. Note that the criteria are very lenient given very large sample sizes and the number of DIF statistics computed. In other words, a large number of items will be placed in categories B and C given the critical value. For reference, the critical value for the chi-square statistic to be significant at $p<0.01$ is 6.635, at $p<0.001$ the critical value is 10.827, and at $p<0.0005$ the critical value is 12.116.

Table 9.2.1.1
Differential Item Functioning Flag Categories

Category	Description	Criterion
A	No DIF	Mantel-Haenzel chi-square not significantly different than zero
B	Weak DIF	Significant Mantel-Haenzel chi-square ($p<0.01$) and $\lvert\text{MH D-DIF}\rvert < 1.5$
C	Strong DIF	Significant Mantel-Haenzel chi-square ($p<0.01$) and $\lvert\text{MH D-DIF}\rvert \geq 1.5$

Another measure of DIF, also presented here for the Spring 2015 AIMS operational items, is the standardized mean difference (SMD; Zwick et al., 1993). The SMD is an effect size index of DIF, which is relatively easy to interpret. The SMD compares the means of the reference and focus groups, adjusting for the distribution of reference and focal group members on the conditioning variable, which for these analyses is the CRT raw score. SMD is computed as (Zwick et al., 1993):

$$SMD = \sum_{k} p_{Fk} \left(m_{Fk} - m_{Rk} \right),$$

where p_{Fk} is the proportion of the focal group members at the *k*th level of the matching variable, m_{Fk} is the mean item response of the focal group at the *k*th level and m_{Rk} is the mean item response of the reference group at the *k*th level . A negative SMD value indicates an item on which the focal group has a lower mean than the reference group. A positive SMD value indicates an item on which the reference group has a lower mean than the focal group.

Mantel-Haenszel chi-square statistic, MH-D DIF, SMD, and flag category results for all items in the Spring 2015 AIMS tests are presented in Tables 9.2.1.2 through 9.2.1.4. It is important to note that DIF analyses are also conducted on field test items prior to form construction. Very few AIMS items are identified as exhibiting strong DIF in field testing. All items exhibiting strong DIF are investigated for possible sources of differential functioning by Pearson and ADE staff and such items are avoided in form construction. Not surprisingly, the vast majority of items on the operational AIMS exhibit no DIF or weak DIF. Items that were flagged for exhibiting strong DIF are summarized in Table 9.2.1.5. There were a total of 2 such items.

Table 9.2.1.2
Spring 2015 AIMS Differential Item Functioning
Science Grade 4

	Reference: Male N= 42551				Reference: Hispanic N= 46530				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557			
	Focal: Female N= 41346				Focal: Non Hispanic N= 37459				Focal: Africa American N= 4763				Focal: Native American N= 6036				Focal: Asian N= 2520				Focal: Hawaii N= 310				Focal: Multiple Indicator N= 29			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	80.46	0.37	0.02	A	54.73	0.32	0.02	A	8.71	-0.25	-0.02	A	16.22	0.31	0.02	A	0.10	0.04	0.00	A	0.65	0.28	0.02	A	0.24	0.06	0.00	A
2	226.00	0.57	0.04	A	0.01	0.00	0.00	A	2.70	-0.13	-0.01	A	1.85	0.09	0.01	A	0.72	0.11	0.01	A	0.02	-0.04	0.00	A	0.01	-0.01	0.00	A
3	51.80	0.30	0.02	A	42.25	-0.27	-0.02	A	0.02	-0.01	0.00	A	20.14	-0.32	-0.03	A	0.65	-0.11	-0.01	A	0.48	-0.23	-0.02	A	0.96	0.11	0.01	A
4	6.99	-0.10	-0.01	A	48.81	-0.27	-0.02	A	3.69	0.15	0.01	A	111.49	-0.74	-0.07	A	11.95	0.43	0.03	A	0.04	0.06	0.01	A	0.09	-0.03	0.00	A
5	9.30	-0.11	-0.01	A	21.29	-0.18	-0.02	A	10.29	0.25	0.02	A	24.20	-0.33	-0.03	A	4.93	0.27	0.02	A	0.40	0.19	0.02	A	2.73	-0.17	-0.01	A
6	6.77	0.09	0.01	A	37.47	-0.22	-0.02	A	5.60	-0.18	-0.02	A	10.71	-0.22	-0.02	A	9.32	-0.31	-0.03	A	0.00	0.01	0.00	A	0.14	0.04	0.00	A
7	185.79	0.50	0.04	A	0.62	-0.03	0.00	A	2.23	0.12	0.01	A	1.67	-0.09	-0.01	A	19.13	0.52	0.04	A	2.30	0.47	0.04	A	1.43	0.12	0.01	A
8	126.25	0.42	0.03	A	47.03	-0.26	-0.02	A	3.39	0.14	0.01	A	2.37	-0.11	-0.01	A	0.19	0.05	0.00	A	0.04	0.06	0.01	A	1.23	0.12	0.01	A
9	17.59	0.14	0.01	A	12.24	-0.12	-0.01	A	23.99	-0.37	-0.04	A	96.99	0.64	0.07	A	4.61	0.21	0.02	A	0.35	0.16	0.02	A	0.28	-0.05	0.00	A
10	21.06	0.19	0.01	A	0.23	-0.02	0.00	A	2.51	0.13	0.01	A	9.46	-0.22	-0.02	A	0.26	0.07	0.00	A	1.95	0.49	0.03	A	0.55	-0.08	-0.01	A
11	13.66	0.13	0.01	A	1.72	-0.05	0.00	A	0.53	0.05	0.01	A	0.16	-0.03	0.00	A	0.22	-0.05	0.00	A	3.30	-0.51	-0.05	A	0.03	-0.02	0.00	A
12	44.20	0.24	0.02	A	27.34	-0.20	-0.01	A	2.74	-0.13	-0.01	A	12.11	-0.26	-0.02	A	0.23	0.05	0.00	A	6.02	-0.77	-0.06	A	0.71	0.08	0.01	A
13	183.04	0.47	0.04	A	11.00	0.12	0.01	A	0.00	0.00	0.00	A	1.14	-0.08	-0.01	A	1.96	0.14	0.01	A	2.56	0.47	0.04	A	0.44	0.06	0.01	A
14	162.23	-0.43	-0.04	A	0.49	0.02	0.00	A	9.55	-0.23	-0.02	A	17.77	0.29	0.03	A	4.42	-0.21	-0.02	A	0.18	0.12	0.01	A	0.27	0.05	0.00	A
15	101.73	0.34	0.03	A	0.07	-0.01	0.00	A	0.87	-0.07	-0.01	A	2.14	0.10	0.01	A	1.56	0.12	0.01	A	0.27	-0.14	-0.01	A	0.00	0.00	0.00	A
16	30.70	-0.21	-0.02	A	189.58	-0.52	-0.04	A	59.96	-0.60	-0.05	A	3.01	-0.12	-0.01	A	11.39	-0.39	-0.03	A	3.84	-0.57	-0.05	A	0.09	0.03	0.00	A
17	16.80	0.16	0.01	A	0.63	0.03	0.00	A	7.22	0.22	0.02	A	9.74	0.22	0.02	A	13.09	0.46	0.03	A	0.72	0.26	0.02	A	4.09	0.22	0.02	A
18	0.02	-0.01	0.00	A	3.22	-0.07	-0.01	A	0.31	-0.04	0.00	A	13.64	-0.25	-0.02	A	1.93	-0.15	-0.01	A	1.10	-0.30	-0.03	A	0.30	-0.05	0.00	A
19	8.52	0.11	0.01	A	1.22	0.04	0.01	A	1.05	-0.08	-0.01	A	3.33	0.13	0.01	A	25.32	0.52	0.05	A	1.68	0.37	0.03	A	1.21	-0.11	-0.01	A
20	190.14	-0.51	-0.04	A	142.97	-0.45	-0.03	A	3.59	-0.15	-0.01	A	60.90	-0.56	-0.05	A	17.58	0.49	0.04	A	2.84	0.48	0.04	A	0.40	-0.06	-0.01	A
21	150.94	0.51	0.03	A	0.37	0.03	0.00	A	0.07	0.02	0.00	A	13.27	-0.26	-0.02	A	11.89	0.50	0.02	A	5.08	0.78	0.05	A	0.00	0.00	0.00	A
22	149.60	-0.47	-0.04	A	106.66	-0.41	-0.04	A	1.27	-0.09	-0.01	A	1.65	0.09	0.01	A	26.67	0.69	0.04	A	2.52	0.50	0.04	A	0.00	0.00	0.00	A
23	63.99	0.28	0.03	A	0.02	0.00	0.00	A	0.02	0.01	0.00	A	4.39	-0.14	-0.01	A	0.46	0.07	0.01	A	0.08	-0.08	-0.01	A	0.70	-0.08	-0.01	A
24	0.01	0.00	0.00	A	53.59	-0.27	-0.02	A	0.02	0.01	0.00	A	13.19	0.25	0.02	A	14.03	0.39	0.04	A	0.13	-0.10	-0.01	A	1.84	0.13	0.01	A
25	1.16	0.05	0.00	A	79.54	-0.41	-0.03	A	6.07	0.22	0.01	A	7.00	-0.20	-0.02	A	10.64	0.56	0.02	A	1.08	0.40	0.02	A	0.12	0.04	0.00	A
26	1.33	0.04	0.00	A	2.75	0.06	0.01	A	5.52	0.18	0.02	A	5.87	-0.17	-0.01	A	34.87	0.70	0.05	A	2.68	-0.48	-0.04	A	1.14	0.11	0.01	A
27	1.62	0.05	0.00	A	18.29	-0.16	-0.01	A	0.10	0.02	0.00	A	5.74	-0.18	-0.01	A	0.34	0.06	0.01	A	0.36	0.17	0.02	A	1.96	-0.14	-0.01	A
28	6.13	0.09	0.01	A	0.25	-0.02	0.00	A	3.13	-0.13	-0.01	A	0.83	-0.06	-0.01	A	2.90	0.19	0.01	A	1.01	0.29	0.03	A	0.34	-0.06	0.00	A
29	12.12	0.15	0.01	A	4.65	0.09	0.01	A	1.95	-0.12	-0.01	A	6.22	-0.18	-0.01	A	1.38	0.18	0.01	A	3.15	-0.59	-0.04	A	0.11	-0.04	0.00	A
30	33.69	0.20	0.02	A	2.38	-0.06	0.00	A	15.53	-0.30	-0.03	A	2.52	-0.11	-0.01	A	5.05	-0.23	-0.02	A	0.64	-0.22	-0.02	A	0.20	0.04	0.00	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

(table continues)

Table 9.2.1.2 (continued)
Spring 2015 AIMS Differential Item Functioning
Science Grade 4

	Reference: Male N= 42551				Reference: Hispanic N= 46530				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557				Reference: White N= 66557			
	Focal: Female N= 41346				Focal: Non Hispanic N= 37459				Focal: Africa American N= 4763				Focal: Native American N= 6036				Focal: Asian N= 2520				Focal: Hawaii N= 310				Focal: Multiple Indicator N= 29			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	0.30	0.02	0.00	A	0.17	-0.02	0.00	A	0.31	-0.04	0.00	A	0.00	0.00	0.00	A	26.27	0.51	0.05	A	1.55	0.37	0.03	A	1.60	0.12	0.01	A
32	513.92	-0.83	-0.07	A	112.28	-0.40	-0.03	A	20.39	-0.37	-0.03	A	0.32	0.04	0.00	A	0.00	0.01	0.00	A	1.19	0.32	0.03	A	0.31	-0.06	0.00	A
33	1257.20	-1.32	-0.11	B<	49.93	-0.27	-0.02	A	2.68	0.13	0.01	A	1.57	-0.09	-0.01	A	7.74	0.30	0.03	A	0.53	0.22	0.02	A	2.18	0.15	0.01	A
34	4.44	0.08	0.01	A	26.36	-0.20	-0.01	A	12.67	0.29	0.02	A	16.71	-0.28	-0.02	A	5.28	0.30	0.02	A	0.00	-0.01	0.00	A	4.83	0.23	0.02	A
35	5.03	-0.08	-0.01	A	28.83	-0.21	-0.02	A	9.70	-0.25	-0.02	A	1.02	0.07	0.01	A	0.66	0.10	0.01	A	7.70	-0.82	-0.07	A	0.88	0.10	0.01	A
36	35.17	0.20	0.02	A	6.43	-0.09	-0.01	A	6.39	-0.19	-0.02	A	0.65	0.05	0.01	A	2.97	-0.18	-0.02	A	0.03	-0.05	0.00	A	3.48	-0.17	-0.02	A
37	28.72	-0.19	-0.02	A	24.68	0.18	0.02	A	0.28	0.04	0.00	A	2.50	-0.12	-0.01	A	1.07	0.10	0.01	A	2.22	0.42	0.04	A	3.39	-0.18	-0.02	A
38	21.72	0.17	0.01	A	37.34	0.23	0.03	A	5.39	0.18	0.02	A	0.29	-0.04	0.00	A	18.79	0.50	0.04	A	3.74	-0.55	-0.05	A	0.35	-0.06	-0.01	A
39	25.35	0.18	0.02	A	3.61	0.07	0.01	A	2.06	-0.11	-0.01	A	35.52	-0.43	-0.04	A	0.03	-0.02	0.00	A	0.65	-0.24	-0.02	A	1.81	-0.13	-0.01	A
40	260.48	-0.64	-0.05	A	10.09	-0.13	-0.01	A	0.22	0.04	0.00	A	10.22	-0.22	-0.02	A	5.38	0.31	0.02	A	1.19	0.36	0.03	A	0.14	-0.04	0.00	A
41	296.89	0.64	0.05	A	0.43	0.03	0.00	A	0.16	-0.03	0.00	A	8.53	-0.23	-0.02	A	10.59	0.34	0.03	A	0.02	-0.05	0.00	A	2.06	-0.15	-0.01	A
42	14.89	-0.14	-0.01	A	28.85	-0.20	-0.02	A	5.30	-0.18	-0.02	A	9.07	-0.20	-0.02	A	7.15	0.32	0.02	A	0.01	0.03	0.00	A	0.43	-0.07	-0.01	A
43	190.42	-0.49	-0.04	A	42.14	-0.24	-0.02	A	16.21	-0.31	-0.03	A	8.75	-0.20	-0.02	A	0.27	0.06	0.00	A	0.13	0.11	0.01	A	9.46	-0.30	-0.03	A
44	94.12	-0.38	-0.03	A	39.78	-0.26	-0.02	A	18.65	-0.39	-0.03	A	10.78	-0.28	-0.02	A	24.74	-0.56	-0.05	A	2.03	-0.48	-0.03	A	11.82	-0.37	-0.03	A
45	2.89	0.06	0.01	A	18.58	-0.16	-0.01	A	2.18	-0.11	-0.01	A	0.17	0.03	0.00	A	0.73	-0.09	-0.01	A	0.04	0.06	0.01	A	1.06	-0.10	-0.01	A
46	41.62	-0.31	-0.02	A	84.13	-0.44	-0.03	A	4.05	0.19	0.01	A	0.00	0.00	0.00	A	0.26	0.09	0.00	A	1.01	0.37	0.02	A	3.27	0.24	0.01	A
47	270.98	-0.60	-0.05	A	39.50	-0.24	-0.02	A	1.62	-0.10	-0.01	A	1.19	-0.08	-0.01	A	3.90	-0.21	-0.02	A	0.08	-0.09	-0.01	A	0.07	0.03	0.00	A
48	0.00	0.00	0.00	A	10.68	0.12	0.01	A	0.53	0.06	0.01	A	0.07	-0.02	0.00	A	8.18	0.33	0.02	A	0.02	0.04	0.00	A	0.08	0.03	0.00	A
49	46.52	0.29	0.02	A	0.05	-0.01	0.00	A	1.46	0.10	0.01	A	0.06	-0.02	0.00	A	13.40	0.58	0.02	A	0.16	0.14	0.01	A	0.13	0.04	0.00	A
50	72.89	-0.31	-0.03	A	54.36	-0.27	-0.02	A	0.51	-0.05	0.00	A	0.62	0.05	0.01	A	0.89	-0.11	-0.01	A	0.44	-0.19	-0.02	A	0.83	0.09	0.01	A
51	9.65	0.11	0.01	A	3.01	-0.06	0.00	A	0.45	-0.05	0.00	A	1.32	0.08	0.01	A	4.81	0.25	0.02	A	0.00	0.01	0.00	A	2.94	-0.17	-0.01	A
52	2.49	0.06	0.00	A	19.05	-0.17	-0.01	A	1.26	-0.09	-0.01	A	0.20	0.03	0.00	A	3.46	0.23	0.02	A	0.02	-0.04	0.00	A	0.01	-0.01	0.00	A
53	115.67	0.37	0.04	A	14.21	-0.13	-0.01	A	9.15	-0.22	-0.02	A	10.61	-0.21	-0.02	A	3.54	-0.19	-0.02	A	0.35	-0.17	-0.02	A	0.68	-0.08	-0.01	A
54	123.52	0.43	0.03	A	51.78	0.29	0.02	A	6.21	0.20	0.02	A	0.29	0.04	0.00	A	11.54	0.46	0.03	A	5.03	0.73	0.05	A	0.05	0.02	0.00	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

Table 9.2.1.3
Spring 2015 AIMS Differential Item Functioning
Science Grade 8

	Reference: Male N= 41610				Reference: Hispanic N= 46083				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155			
	Focal: Female N= 40415				Focal: Non Hispanic N= 36058				Focal: Africa American N= 4850				Focal: Native American N= 5544				Focal: Asian N= 2459				Focal: Hawaii N= 293				Focal: Multiple Indicator N= 22			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	78.72	-0.41	-0.02	A	24.03	-0.23	-0.01	A	4.37	-0.19	-0.01	A	7.88	-0.23	-0.02	A	14.80	-0.56	-0.02	A	0.60	-0.28	-0.02	A	0.10	0.05	0.00	A
2	110.00	-0.42	-0.03	A	2.36	-0.06	-0.01	A	0.60	-0.06	-0.01	A	0.25	-0.04	0.00	A	17.58	-0.53	-0.03	A	0.92	0.33	0.02	A	0.14	0.05	0.00	A
3	421.90	0.73	0.07	A	54.95	-0.27	-0.03	A	5.62	-0.18	-0.02	A	1.28	0.08	0.01	A	84.40	-0.94	-0.08	A	0.63	0.24	0.02	A	4.93	0.25	0.02	A
4	65.37	0.31	0.02	A	5.14	0.09	0.01	A	6.18	0.20	0.02	A	42.81	-0.49	-0.04	A	8.78	0.40	0.02	A	2.97	-0.54	-0.04	A	0.02	0.02	0.00	A
5	78.30	-0.39	-0.02	A	43.22	-0.29	-0.02	A	0.74	0.07	0.01	A	16.98	-0.32	-0.03	A	2.22	-0.24	-0.01	A	0.36	-0.21	-0.01	A	3.11	0.25	0.01	A
6	54.25	-0.25	-0.03	A	80.21	0.32	0.03	A	2.28	-0.11	-0.01	A	6.73	0.18	0.02	A	14.47	-0.39	-0.04	A	1.62	-0.36	-0.04	A	3.57	-0.20	-0.02	A
7	1.32	-0.05	0.00	A	29.80	-0.24	-0.02	A	1.41	0.10	0.01	A	0.12	0.03	0.00	A	0.28	-0.08	0.00	A	0.07	0.10	0.01	A	0.16	0.06	0.00	A
8	158.69	-0.70	-0.03	A	12.08	-0.20	-0.01	A	37.12	-0.61	-0.03	A	125.62	-0.98	-0.06	A	1.43	-0.26	-0.01	A	1.55	-0.54	-0.02	A	0.55	-0.13	0.00	A
9	872.63	1.05	0.10	B>	0.19	-0.02	0.00	A	2.44	-0.12	-0.01	A	1.83	-0.09	-0.01	A	3.06	0.20	0.02	A	0.33	0.18	0.02	A	1.65	0.14	0.01	A
10	121.58	-0.40	-0.04	A	24.79	-0.19	-0.02	A	23.94	-0.36	-0.03	A	20.20	-0.32	-0.03	A	10.97	-0.37	-0.03	A	1.19	-0.33	-0.03	A	0.00	0.00	0.00	A
11	118.76	-0.41	-0.03	A	23.16	-0.18	-0.02	A	3.63	-0.14	-0.01	A	0.00	0.00	0.00	A	0.90	-0.11	-0.01	A	1.18	0.34	0.03	A	0.62	0.09	0.01	A
12	338.26	0.72	0.05	A	57.63	0.31	0.02	A	0.01	-0.01	0.00	A	4.47	0.15	0.01	A	20.98	0.65	0.03	A	1.23	0.37	0.03	A	0.06	-0.03	0.00	A
13	22.20	0.19	0.01	A	10.03	0.13	0.01	A	0.14	-0.03	0.00	A	0.44	-0.05	0.00	A	3.67	0.28	0.01	A	0.02	0.05	0.00	A	5.47	-0.29	-0.02	A
14	1.12	0.04	0.00	A	7.78	0.10	0.01	A	1.62	-0.09	-0.01	A	0.92	-0.07	-0.01	A	18.96	0.44	0.04	A	0.34	-0.16	-0.02	A	1.20	-0.12	-0.01	A
15	6.64	-0.09	-0.01	A	0.29	-0.02	0.00	A	2.90	0.13	0.01	A	6.68	0.18	0.02	A	37.49	0.68	0.06	A	4.12	0.61	0.06	A	0.07	-0.03	0.00	A
16	26.37	0.19	0.02	A	0.03	0.01	0.00	A	17.49	-0.32	-0.03	A	2.34	0.11	0.01	A	0.11	0.04	0.00	A	5.32	-0.74	-0.06	A	11.62	-0.40	-0.03	A
17	50.32	0.26	0.02	A	0.18	-0.02	0.00	A	1.72	-0.10	-0.01	A	1.70	-0.09	-0.01	A	2.46	-0.18	-0.01	A	0.00	0.01	0.00	A	0.73	-0.10	-0.01	A
18	48.04	0.26	0.02	A	11.82	0.13	0.01	A	0.01	-0.01	0.00	A	1.16	0.08	0.01	A	2.20	0.16	0.01	A	1.50	-0.39	-0.03	A	0.07	-0.03	0.00	A
19	510.40	0.92	0.06	A	24.84	0.21	0.01	A	3.08	0.15	0.01	A	1.99	0.11	0.01	A	15.54	0.59	0.03	A	0.05	0.08	0.01	A	0.05	0.03	0.00	A
20	89.56	-0.37	-0.03	A	96.22	-0.40	-0.03	A	16.04	-0.35	-0.02	A	56.39	-0.67	-0.04	A	16.95	0.45	0.04	A	0.24	-0.16	-0.01	A	4.06	-0.24	-0.02	A
21	4.88	-0.08	-0.01	A	26.64	-0.19	-0.02	A	12.42	0.27	0.02	A	10.47	-0.25	-0.02	A	25.71	0.54	0.05	A	0.62	0.24	0.02	A	3.30	0.20	0.02	A
22	0.20	-0.02	0.00	A	243.32	-0.56	-0.05	A	0.18	0.03	0.00	A	4.09	-0.14	-0.01	A	0.40	0.07	0.01	A	0.14	-0.11	-0.01	A	2.06	-0.15	-0.01	A
23	54.04	-0.27	-0.02	A	78.05	-0.33	-0.03	A	24.67	-0.39	-0.03	A	6.56	-0.19	-0.02	A	14.28	-0.43	-0.03	A	0.08	0.08	0.01	A	2.98	0.20	0.02	A
24	2.09	0.06	0.01	A	43.70	-0.26	-0.02	A	0.01	-0.01	0.00	A	41.01	-0.54	-0.04	A	13.72	0.41	0.04	A	0.02	0.04	0.00	A	0.02	-0.02	0.00	A
25	73.58	0.33	0.03	A	4.75	0.09	0.01	A	14.57	-0.29	-0.03	A	5.94	0.18	0.02	A	13.48	0.47	0.03	A	0.13	0.11	0.01	A	1.33	-0.14	-0.01	A
26	13.67	-0.13	-0.01	A	75.37	-0.32	-0.02	A	5.26	0.18	0.02	A	10.33	-0.24	-0.02	A	0.82	0.10	0.01	A	0.23	0.14	0.01	A	1.17	0.12	0.01	A
27	11.06	-0.13	-0.01	A	420.94	-0.81	-0.06	A	8.17	0.23	0.02	A	54.61	-0.61	-0.04	A	51.09	0.85	0.06	A	1.24	-0.37	-0.03	A	0.36	0.07	0.01	A
28	62.87	-0.28	-0.03	A	28.92	-0.20	-0.02	A	8.12	-0.22	-0.02	A	1.01	-0.07	-0.01	A	9.15	0.32	0.03	A	0.00	-0.01	0.00	A	1.18	0.12	0.01	A
29	328.46	-0.67	-0.06	A	195.86	-0.53	-0.04	A	15.74	-0.31	-0.03	A	18.38	-0.32	-0.03	A	13.49	-0.40	-0.03	A	0.78	-0.27	-0.02	A	1.63	-0.14	-0.01	A
30	338.43	0.77	0.05	A	55.12	0.32	0.02	A	4.37	-0.17	-0.01	A	18.63	0.33	0.03	A	12.55	0.55	0.02	A	5.97	0.87	0.06	A	2.08	-0.19	-0.01	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

(table continues)

Table 9.2.1.3 (continued)
Spring 2015 AIMS Differential Item Functioning
Science Grade 8

	Reference: Male N= 41610				Reference: Hispanic N= 46083				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155				Reference: White N= 66155			
	Focal: Female N= 40415				Focal: Non Hispanic N= 36058				Focal: Africa American N= 4850				Focal: Native American N= 5544				Focal: Asian N= 2459				Focal: Hawaii N= 293				Focal: Multiple Indicator N= 22			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	77.97	-0.31	-0.03	A	12.62	-0.13	-0.01	A	0.31	0.04	0.00	A	0.29	-0.04	0.00	A	0.04	0.02	0.00	A	0.07	0.08	0.01	A	2.18	0.16	0.01	A
32	592.99	1.18	0.06	B>	0.04	0.01	0.00	A	0.17	-0.04	0.00	A	1.59	0.11	0.01	A	1.11	-0.19	-0.01	A	0.60	0.30	0.02	A	0.05	0.04	0.00	A
33	49.35	-0.24	-0.02	A	156.28	0.44	0.04	A	2.87	-0.12	-0.01	A	6.43	0.18	0.02	A	7.66	-0.28	-0.03	A	0.98	0.28	0.03	A	0.30	0.06	0.01	A
34	10.73	0.13	0.01	A	0.00	0.00	0.00	A	0.17	-0.03	0.00	A	3.13	0.13	0.01	A	36.39	0.96	0.04	A	5.16	-0.74	-0.05	A	0.07	0.03	0.00	A
35	1785.67	-1.53	-0.14	C<	30.62	-0.21	-0.02	A	3.28	-0.14	-0.01	A	0.49	-0.05	0.00	A	17.35	0.45	0.04	A	0.01	-0.03	0.00	A	0.00	0.00	0.00	A
36	69.19	0.29	0.03	A	93.93	-0.35	-0.03	A	7.64	0.20	0.02	A	0.53	-0.05	0.00	A	9.80	0.33	0.03	A	4.61	0.62	0.06	A	3.40	0.20	0.02	A
37	34.60	-0.24	-0.02	A	7.17	-0.11	-0.01	A	0.11	-0.03	0.00	A	2.92	-0.15	-0.01	A	8.85	0.33	0.03	A	0.52	-0.26	-0.02	A	2.27	-0.19	-0.01	A
38	187.80	0.49	0.04	A	63.50	0.30	0.03	A	20.15	-0.34	-0.03	A	5.48	0.17	0.02	A	19.08	0.50	0.04	A	0.05	-0.06	-0.01	A	0.02	0.02	0.00	A
39	70.66	-0.30	-0.03	A	28.20	-0.19	-0.02	A	0.22	0.04	0.00	A	0.24	0.04	0.00	A	6.07	0.27	0.02	A	0.18	0.13	0.01	A	1.38	0.13	0.01	A
40	14.35	-0.15	-0.01	A	1.05	0.04	0.00	A	6.23	-0.20	-0.02	A	0.45	0.05	0.00	A	42.95	0.89	0.05	A	1.08	-0.33	-0.03	A	0.96	-0.12	-0.01	A
41	51.37	-0.27	-0.02	A	13.77	-0.14	-0.01	A	0.33	-0.04	0.00	A	0.00	0.00	0.00	A	11.19	0.41	0.03	A	0.41	0.20	0.02	A	0.40	0.07	0.01	A
42	4.62	-0.09	-0.01	A	38.86	-0.26	-0.02	A	5.10	-0.18	-0.01	A	1.04	0.08	0.01	A	0.10	0.05	0.00	A	0.52	0.26	0.02	A	0.79	-0.11	-0.01	A
43	11.07	0.12	0.01	A	37.74	-0.23	-0.02	A	48.16	-0.53	-0.05	A	20.34	-0.33	-0.03	A	0.09	0.03	0.00	A	0.03	-0.05	0.00	A	3.40	-0.21	-0.02	A
44	228.65	0.73	0.04	A	26.36	0.25	0.01	A	1.23	-0.10	-0.01	A	3.86	0.17	0.01	A	0.01	-0.02	0.00	A	1.66	0.52	0.03	A	0.00	0.01	0.00	A
45	31.86	-0.19	-0.02	A	26.68	-0.18	-0.02	A	0.14	-0.03	0.00	A	1.44	-0.08	-0.01	A	1.60	0.13	0.01	A	1.07	-0.29	-0.03	A	2.03	-0.15	-0.01	A
46	103.00	-0.37	-0.03	A	0.46	-0.03	0.00	A	11.82	-0.26	-0.02	A	0.25	-0.04	0.00	A	17.69	0.50	0.04	A	0.00	0.01	0.00	A	0.02	0.01	0.00	A
47	6.54	0.12	0.01	A	13.10	0.17	0.01	A	2.50	-0.14	-0.01	A	0.00	0.00	0.00	A	0.15	-0.07	0.00	A	0.18	-0.17	-0.01	A	1.77	-0.20	-0.01	A
48	0.40	0.02	0.00	A	111.18	-0.42	-0.03	A	3.95	0.16	0.01	A	23.28	-0.36	-0.03	A	0.42	0.09	0.01	A	5.28	0.77	0.06	A	2.76	0.20	0.01	A
49	216.63	0.53	0.05	A	89.30	0.35	0.03	A	0.15	0.03	0.00	A	0.87	0.07	0.01	A	16.57	0.47	0.04	A	0.00	0.02	0.00	A	0.12	-0.04	0.00	A
50	69.84	0.32	0.03	A	0.49	-0.03	0.00	A	0.00	0.00	0.00	A	51.27	-0.55	-0.04	A	28.90	0.69	0.04	A	0.30	-0.19	-0.01	A	4.62	-0.26	-0.02	A
51	37.44	0.22	0.02	A	3.37	0.07	0.01	A	0.23	0.04	0.00	A	33.26	-0.41	-0.04	A	18.11	0.47	0.04	A	0.00	-0.02	0.00	A	1.69	0.14	0.01	A
52	15.58	-0.14	-0.01	A	83.92	-0.34	-0.03	A	6.27	0.19	0.02	A	0.78	0.06	0.01	A	0.41	-0.07	-0.01	A	0.50	-0.20	-0.02	A	3.80	0.22	0.02	A
53	111.36	0.48	0.03	A	0.22	0.02	0.00	A	1.67	0.11	0.01	A	4.25	0.17	0.01	A	0.27	0.09	0.00	A	1.57	-0.45	-0.03	A	1.60	0.18	0.01	A
54	20.68	-0.17	-0.01	A	10.87	-0.13	-0.01	A	0.71	0.07	0.01	A	7.60	-0.20	-0.02	A	21.27	0.63	0.04	A	0.26	-0.17	-0.01	A	3.54	0.22	0.02	A
55	104.86	0.39	0.03	A	8.64	-0.12	0.00	A	2.56	0.13	0.01	A	0.05	0.02	0.00	A	36.91	0.65	0.06	A	0.17	0.13	0.01	A	2.41	-0.18	-0.01	A
56	0.18	-0.02	0.00	A	4.20	0.08	0.01	A	0.00	0.00	0.00	A	10.45	-0.25	-0.02	A	62.99	0.83	0.08	A	2.03	0.42	0.04	A	0.71	-0.10	-0.01	A
57	89.75	-0.32	-0.03	A	32.74	-0.20	-0.01	A	15.17	-0.28	-0.03	A	12.23	-0.24	-0.02	A	0.26	0.05	0.01	A	1.36	-0.33	-0.03	A	4.12	-0.21	-0.02	A
58	34.78	-0.21	-0.02	A	118.21	-0.39	-0.03	A	13.12	-0.27	-0.02	A	2.35	-0.11	-0.01	A	3.78	0.21	0.02	A	1.04	-0.30	-0.03	A	4.25	-0.22	-0.02	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

Table 9.2.1.4
Spring 2015 AIMS Differential Item Functioning
Science High School

	Reference: Male N= 40281 Focal: Female N= 39369				Reference: Hispanic N= 45531 Focal: Non Hispanic N= 34435				Reference: White N= 63299 Focal: Africa American N= 4679				Reference: White N= 63299 Focal: Native American N= 4637				Reference: White N= 63299 Focal: Asian N= 2390				Reference: White N= 63299 Focal: Hawaii N= 326				Reference: White N= 63299 Focal: Multiple Indicator N= 22			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	129.73	0.40	0.04	A	0.79	0.03	0.00	A	37.53	-0.46	-0.04	A	3.95	-0.15	-0.01	A	0.28	0.06	0.01	A	3.06	-0.47	-0.05	A	0.54	-0.08	-0.01	A
2	139.62	0.46	0.04	A	0.01	0.00	0.00	A	0.85	0.08	0.01	A	18.74	-0.35	-0.03	A	17.83	0.54	0.03	A	0.13	0.10	0.01	A	0.31	0.06	0.01	A
3	85.16	-0.33	-0.03	A	14.13	-0.14	-0.01	A	6.18	0.19	0.02	A	0.01	0.01	0.00	A	21.59	0.51	0.05	A	0.35	-0.16	-0.02	A	1.35	-0.12	-0.01	A
4	39.83	-0.22	-0.02	A	0.68	-0.03	0.00	A	0.53	-0.05	-0.01	A	6.64	0.20	0.02	A	98.11	1.04	0.10	B>	0.02	0.04	0.00	A	0.45	0.07	0.01	A
5	7.37	-0.09	-0.01	A	42.63	0.24	0.03	A	0.45	0.05	0.00	A	58.45	0.57	0.06	A	63.89	0.81	0.08	A	1.85	0.36	0.04	A	0.42	-0.07	-0.01	A
6	4.19	-0.07	-0.01	A	70.17	-0.31	-0.03	A	2.20	0.11	0.01	A	0.68	0.07	0.01	A	1.96	0.14	0.01	A	1.33	0.31	0.03	A	0.42	0.07	0.01	A
7	0.09	-0.01	0.00	A	119.28	0.45	0.03	A	8.74	0.25	0.02	A	2.20	0.13	0.01	A	30.54	0.59	0.05	A	0.09	-0.10	-0.01	A	4.08	-0.24	-0.02	A
8	71.58	-0.30	-0.03	A	6.00	0.09	0.01	A	34.13	-0.44	-0.04	A	19.53	-0.33	-0.03	A	0.76	-0.10	-0.01	A	0.94	0.28	0.03	A	3.86	-0.21	-0.02	A
9	13.61	0.14	0.01	A	0.17	-0.02	0.00	A	33.24	-0.45	-0.04	A	4.46	-0.16	-0.01	A	1.98	0.17	0.01	A	3.13	-0.49	-0.04	A	1.00	-0.11	-0.01	A
10	86.88	0.34	0.03	A	3.20	0.07	0.01	A	6.54	0.20	0.02	A	2.54	-0.12	-0.01	A	23.33	0.60	0.04	A	0.24	0.14	0.01	A	0.47	0.08	0.01	A
11	6.23	-0.09	-0.01	A	1.50	-0.05	0.00	A	0.09	0.02	0.00	A	3.20	-0.13	-0.01	A	0.83	0.10	0.01	A	1.98	-0.40	-0.04	A	0.66	-0.09	-0.01	A
12	10.92	-0.12	-0.01	A	50.97	-0.28	-0.02	A	0.51	0.06	0.00	A	4.78	0.19	0.01	A	14.95	0.41	0.04	A	0.14	0.11	0.01	A	0.18	0.05	0.00	A
13	1.19	0.05	0.00	A	53.53	-0.32	-0.02	A	11.40	-0.29	-0.02	A	28.05	-0.44	-0.04	A	0.59	-0.11	-0.01	A	3.39	-0.58	-0.04	A	2.45	0.21	0.01	A
14	7.75	-0.10	-0.01	A	14.49	0.14	0.01	A	5.15	0.17	0.02	A	39.61	0.47	0.05	A	3.06	0.18	0.02	A	0.01	0.03	0.00	A	0.07	0.03	0.00	A
15	145.67	0.49	0.04	A	21.79	-0.19	-0.01	A	0.01	-0.01	0.00	A	1.08	-0.08	-0.01	A	1.41	-0.15	-0.01	A	0.01	-0.03	0.00	A	0.11	0.04	0.00	A
16	326.09	-0.69	-0.06	A	111.58	-0.41	-0.04	A	39.47	-0.49	-0.04	A	39.81	-0.49	-0.05	A	7.03	-0.31	-0.02	A	3.77	-0.57	-0.05	A	1.03	-0.12	-0.01	A
17	36.81	-0.22	-0.02	A	4.44	0.08	0.01	A	4.37	0.16	0.01	A	11.35	-0.27	-0.02	A	14.40	0.44	0.04	A	0.02	0.04	0.00	A	2.74	-0.18	-0.02	A
18	1554.74	-1.52	-0.12	C<	173.67	-0.51	-0.04	A	27.13	-0.42	-0.04	A	2.96	0.14	0.01	A	0.21	-0.06	0.00	A	0.28	-0.15	-0.01	A	3.37	0.21	0.02	A
19	227.99	0.56	0.05	A	71.17	-0.32	-0.02	A	57.66	-0.60	-0.05	A	138.22	-1.00	-0.08	B<	15.63	-0.43	-0.04	A	2.43	-0.43	-0.04	A	8.86	-0.33	-0.03	A
20	18.61	-0.16	-0.01	A	22.74	-0.19	-0.01	A	6.42	0.20	0.02	A	2.72	-0.14	-0.01	A	32.63	0.60	0.06	A	0.25	-0.15	-0.01	A	8.24	0.31	0.03	A
21	176.66	-0.68	-0.03	A	7.91	-0.14	-0.01	A	47.20	-0.66	-0.04	A	1.39	-0.11	-0.01	A	2.37	-0.29	-0.01	A	2.76	-0.62	-0.03	A	0.23	-0.08	0.00	A
22	13.49	0.13	0.01	A	29.14	0.20	0.02	A	0.40	0.05	0.00	A	25.17	0.38	0.04	A	0.65	0.09	0.01	A	5.13	0.63	0.06	A	0.03	0.02	0.00	A
23	256.69	-0.67	-0.05	A	5.67	0.10	0.01	A	0.05	0.02	0.00	A	2.86	-0.14	-0.01	A	5.42	-0.32	-0.02	A	0.96	-0.29	-0.02	A	1.53	0.16	0.01	A
24	404.61	-0.87	-0.06	A	77.14	-0.38	-0.03	A	55.67	-0.63	-0.05	A	8.58	0.25	0.02	A	4.72	0.32	0.02	A	0.55	0.25	0.02	A	1.66	0.17	0.01	A
25	679.61	0.95	0.09	A	6.54	0.10	0.01	A	0.23	0.04	0.00	A	4.69	0.16	0.02	A	26.58	0.59	0.05	A	0.17	-0.12	-0.01	A	0.01	-0.01	0.00	A
26	3.17	-0.07	-0.01	A	14.42	-0.15	-0.01	A	8.19	-0.23	-0.02	A	4.66	-0.18	-0.01	A	0.50	0.08	0.01	A	0.09	-0.09	-0.01	A	0.48	0.08	0.01	A
27	425.25	-0.78	-0.06	A	49.24	-0.28	-0.02	A	0.72	-0.07	0.00	A	1.14	-0.09	-0.01	A	0.02	-0.01	0.00	A	8.21	-0.94	-0.07	A	0.23	-0.05	0.00	A
28	27.89	-0.19	-0.02	A	35.08	-0.22	-0.02	A	3.57	-0.14	-0.01	A	0.42	-0.05	0.00	A	0.70	0.09	0.01	A	0.16	0.11	0.01	A	3.53	0.20	0.02	A
29	21.91	0.17	0.02	A	3.27	0.07	0.01	A	10.57	0.25	0.02	A	2.90	0.13	0.01	A	19.68	0.49	0.04	A	2.60	-0.47	-0.04	A	1.86	0.15	0.01	A
30	0.00	0.00	0.00	A	3.19	-0.07	-0.01	A	39.83	0.49	0.04	A	21.05	-0.36	-0.03	A	38.92	0.71	0.06	A	3.27	0.51	0.05	A	2.45	0.17	0.02	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

(table continues)

Table 9.2.1.4 (continued)
Spring 2015 AIMS Differential Item Functioning
Science High School

	Reference: Male N= 40281				Reference: Hispanic N= 45531				Reference: White N= 63299				Reference: White N= 63299				Reference: White N= 63299				Reference: White N= 63299				Reference: White N= 63299			
	Focal: Female N= 39369				Focal: Non Hispanic N= 34435				Focal: Africa American N= 4679				Focal: Native American N= 4637				Focal: Asian N= 2390				Focal: Hawaii N= 326				Focal: Multiple Indicator N= 22			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	6.29	-0.09	-0.01	A	1.26	-0.04	0.00	A	12.35	0.27	0.02	A	19.22	-0.35	-0.03	A	18.69	0.48	0.04	A	0.02	0.03	0.00	A	0.48	0.07	0.01	A
32	89.05	0.34	0.03	A	5.58	0.09	0.01	A	7.89	0.21	0.02	A	0.01	0.01	0.00	A	31.19	0.62	0.05	A	0.39	-0.17	-0.02	A	0.81	0.09	0.01	A
33	36.17	-0.22	-0.02	A	6.32	-0.09	-0.01	A	4.00	-0.15	-0.01	A	2.06	-0.11	-0.01	A	0.74	0.09	0.01	A	1.77	0.38	0.03	A	0.37	0.07	0.01	A
34	84.37	-0.36	-0.03	A	0.50	-0.03	0.00	A	1.16	-0.09	-0.01	A	9.44	0.27	0.02	A	18.70	0.48	0.04	A	0.68	-0.26	-0.02	A	0.12	0.04	0.00	A
35	7.70	-0.11	-0.01	A	2.68	0.07	0.01	A	0.12	0.03	0.00	A	44.73	-0.55	-0.05	A	8.39	-0.38	-0.02	A	3.12	-0.53	-0.04	A	3.76	-0.24	-0.02	A
36	36.09	0.27	0.02	A	0.72	-0.04	0.00	A	54.22	-0.65	-0.05	A	1.42	-0.10	-0.01	A	9.59	0.51	0.02	A	1.07	0.34	0.02	A	0.95	-0.13	-0.01	A
37	0.72	-0.03	0.00	A	48.65	-0.26	-0.02	A	0.55	0.06	0.01	A	0.03	0.01	0.00	A	9.63	0.32	0.03	A	1.67	-0.37	-0.03	A	0.28	0.06	0.01	A
38	11.08	-0.13	-0.01	A	0.80	-0.04	0.00	A	2.91	0.14	0.01	A	0.06	0.02	0.00	A	62.96	0.85	0.08	A	0.22	-0.14	-0.01	A	0.01	-0.01	0.00	A
39	1.96	0.05	0.00	A	15.96	0.15	0.02	A	4.09	-0.16	-0.01	A	6.72	-0.20	-0.02	A	8.72	0.35	0.03	A	0.17	-0.12	-0.01	A	2.76	-0.19	-0.02	A
40	6.75	-0.09	-0.01	A	70.59	-0.32	-0.02	A	4.57	-0.17	-0.01	A	6.22	-0.20	-0.02	A	3.32	-0.20	-0.02	A	2.11	0.40	0.04	A	0.06	-0.03	0.00	A
41	0.22	0.02	0.00	A	0.05	0.01	0.00	A	1.27	-0.09	-0.01	A	0.04	-0.02	0.00	A	0.26	0.06	0.00	A	1.79	-0.40	-0.03	A	2.44	0.18	0.01	A
42	22.93	0.18	0.02	A	117.31	-0.41	-0.03	A	42.06	-0.51	-0.04	A	8.52	-0.23	-0.02	A	0.04	-0.02	0.00	A	2.19	-0.41	-0.04	A	2.33	-0.17	-0.01	A
43	0.53	0.03	0.00	A	38.40	-0.23	-0.02	A	15.07	-0.30	-0.03	A	38.45	0.48	0.04	A	6.22	-0.27	-0.02	A	0.01	-0.03	0.00	A	0.27	0.05	0.01	A
44	121.99	-0.40	-0.04	A	45.21	-0.25	-0.02	A	44.89	-0.53	-0.05	A	5.86	0.19	0.02	A	30.42	-0.57	-0.05	A	4.15	-0.57	-0.05	A	0.29	-0.06	-0.01	A
45	1.22	0.04	0.00	A	34.16	-0.24	-0.01	A	4.68	0.18	0.01	A	3.61	-0.17	-0.01	A	1.30	-0.12	-0.01	A	0.04	-0.06	0.00	A	1.76	0.15	0.01	A
46	284.68	0.62	0.06	A	172.05	-0.49	-0.04	A	1.15	-0.08	-0.01	A	2.68	0.12	0.01	A	14.91	-0.42	-0.04	A	0.14	-0.11	-0.01	A	3.82	0.21	0.02	A
47	136.23	-0.45	-0.04	A	503.74	-0.87	-0.07	A	0.01	0.01	0.00	A	4.19	-0.17	-0.01	A	0.22	-0.06	0.00	A	0.27	0.15	0.01	A	8.14	0.32	0.03	A
48	117.07	-0.42	-0.03	A	39.04	-0.25	-0.02	A	0.11	-0.03	0.00	A	2.03	0.12	0.01	A	0.58	-0.08	-0.01	A	0.73	-0.27	-0.02	A	0.31	0.06	0.01	A
49	237.41	0.57	0.05	A	32.30	0.22	0.02	A	0.04	0.01	0.00	A	0.02	-0.01	0.00	A	8.98	0.35	0.03	A	0.13	0.11	0.01	A	0.19	-0.05	0.00	A
50	96.43	0.38	0.03	A	0.75	0.03	0.00	A	0.57	-0.06	-0.01	A	11.92	-0.27	-0.02	A	17.06	-0.51	-0.03	A	1.58	-0.36	-0.03	A	0.00	0.00	0.00	A
51	44.31	0.25	0.02	A	0.69	-0.03	0.00	A	0.04	-0.02	0.00	A	10.48	-0.25	-0.02	A	0.38	0.07	0.01	A	3.00	-0.49	-0.04	A	1.05	-0.11	-0.01	A
52	0.72	-0.03	0.00	A	4.67	-0.08	0.00	A	3.06	-0.13	-0.01	A	0.08	-0.02	0.00	A	3.36	0.19	0.02	A	1.61	0.35	0.03	A	0.94	-0.10	-0.01	A
53	323.76	0.65	0.06	A	61.43	-0.29	-0.02	A	5.94	0.19	0.02	A	3.61	-0.15	-0.01	A	30.71	0.63	0.05	A	5.07	0.63	0.06	A	3.19	0.19	0.02	A
54	256.97	-0.66	-0.05	A	243.26	-0.64	-0.05	A	11.54	-0.28	-0.02	A	36.67	-0.51	-0.04	A	0.51	-0.10	-0.01	A	0.62	-0.25	-0.02	A	0.15	0.05	0.00	A
55	2.89	0.06	0.01	A	22.25	-0.17	-0.01	A	0.05	0.02	0.00	A	0.51	0.06	0.01	A	0.30	-0.06	-0.01	A	1.93	-0.38	-0.04	A	0.08	0.03	0.00	A
56	67.30	0.29	0.03	A	9.24	-0.11	-0.01	A	0.12	0.03	0.00	A	2.15	0.11	0.01	A	1.33	-0.12	-0.01	A	0.51	0.20	0.02	A	0.18	-0.05	0.00	A
57	3.71	-0.07	-0.01	A	0.73	0.03	0.01	A	0.91	0.07	0.01	A	9.36	-0.24	-0.02	A	26.22	0.56	0.05	A	0.14	0.11	0.01	A	0.02	-0.01	0.00	A
58	48.96	-0.27	-0.02	A	39.26	-0.25	-0.02	A	8.87	-0.24	-0.02	A	4.15	-0.17	-0.01	A	16.98	0.47	0.04	A	7.72	-0.87	-0.07	A	1.14	-0.12	-0.01	A
59	38.58	-0.23	-0.02	A	0.89	-0.04	0.00	A	1.29	0.09	0.01	A	1.93	0.12	0.01	A	23.85	0.52	0.05	A	0.29	-0.16	-0.01	A	0.69	0.09	0.01	A
60	0.58	0.03	0.00	A	32.80	0.21	0.02	A	0.09	-0.02	0.00	A	0.28	0.04	0.00	A	1.96	0.14	0.01	A	0.47	0.19	0.02	A	0.29	0.06	0.01	A
61	361.33	0.76	0.06	A	0.56	0.03	0.00	A	12.91	0.30	0.02	A	20.86	-0.36	-0.03	A	36.45	0.86	0.05	A	0.13	0.11	0.01	A	0.00	0.01	0.00	A
62	0.11	0.01	0.00	A	2.36	-0.06	0.00	A	9.79	0.23	0.02	A	0.57	0.06	0.01	A	2.39	0.17	0.02	A	0.02	-0.03	0.00	A	1.18	-0.11	-0.01	A
63	108.27	0.38	0.03	A	0.30	0.02	0.00	A	1.67	-0.10	-0.01	A	12.45	-0.28	-0.02	A	26.91	0.59	0.05	A	0.03	-0.05	0.00	A	1.06	0.11	0.01	A
64	9.25	0.11	0.01	A	23.82	-0.18	-0.01	A	1.40	-0.09	-0.01	A	33.86	-0.48	-0.04	A	0.01	0.01	0.00	A	1.23	-0.32	-0.03	A	5.10	-0.24	-0.02	A
65	3.20	0.07	0.01	A	50.07	-0.27	-0.02	A	3.79	-0.15	-0.01	A	3.19	0.14	0.01	A	20.14	-0.48	-0.04	A	2.01	-0.40	-0.04	A	0.66	-0.09	-0.01	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel-Haenszel Chi-Square, MH = Delta (MH-D DIF), SMD = Standardized Mean Difference, A=No DIF, B=Weak DIF, C=Strong DIF, < favors reference group, > favors focal group. Item number does not indicate test booklet location due to field test items and NRT items.

Table 9.2.1.5
DIF Statistics for Items Exhibiting Strong DIF

Content	Grade	Item #	Item Type	In favor of / Against	Group	MH Delta	MH SMD
Science	8	35	MC	Against	Female	1785.67	-1.53 -0.14
Science	HS	18	MC	Against	Female	1554.74	-1.52 -0.12

9.2.2 Correlations among AIMS Assessments

Correlations were examined between scale scores on Spring 2015 AIMS tests in high school as the high school students were only a group of students who took different AIMS tests this year. Note that data used for the calculation of correlation included records with valid scale scores in all content areas and tests in each grade level. Sample sizes are therefore slightly lower than presented in other parts of this technical report.

In addition, because students in high school had different testing windows for the reading, writing, mathematics, and science tests, data merging was necessary to match reading and writing records with mathematics and science records. Once valid records were selected for each of the high school content areas and records with duplicate student identification numbers were removed, data was merged based on student identification number. Table 9.2.2.1 presents the resulting reduction in N size for each grade level. Scale score means and distributions prior to and after the merge were compared to ensure that the match did not substantially alter the data. Only slight changes in mean and standard deviation occurred after the merge, and density plots illustrated that the shape of the distribution prior to and after merging matched reasonably well. Therefore, correlations presented for the high school reading, writing, and mathematics and science are based on the matched data, with a total N size of 59.

All correlations are presented in Tables 9.2.2.2. Caution is needed to interpret the results because the number of students taking different AIMS tests was quite low compared to the previous administrations. Correlations were high between tests designed to measure dissimilar constructs such as math, writing, and science.

Table 9.2.2.1
Spring 2015 Matched N Counts for Correlations Calculations

Grade	N Count
HS	59

Table 9.2.2.2
Spring 2015 AIMS Correlations among Assessments
High School

Test	1	2	3	4
1. RD	1	0.74	0.87	0.87
2. WR	0.74	1	0.78	0.81
3. MA	0.87	0.78	1	0.86
4. SC	0.87	0.81	0.86	1

Note: N size will be less than presented in other parts of this Technical Report due to 1) missing or invalid test records in some, but not all, content areas and 2) matching reading, writing, mathematics, and science records according to student identification number. The N-count for this table is 59.

PART 10: CLASSIFICATION

Part 10 of this technical report provides information regarding classifying students into proficiency categories. The following AERA/APA/NCME *Standards* from the 1999 edition are covered in this part: 1.5, 1.7, 2.2, 2.14, 2.15, 4.9, 4.19, 4.20, 4.21, and 6.5. The 2014 AERA/APA/NCME *Standards* (AERA, APA, NCME, 2014) addressed by this chapter are: 1.8, 1.9, 2.13, 2.14, 2.16, 5.5, 5.21, 5.22, 5.23, and 7.4.

Scores from the Spring 2015 AIMS 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. Arizona educators made recommendations for cut scores for each category in the standard setting workshops. Analyses were conducted to examine the consistency and accuracy with which students were assigned to performance categories.

10.1 Standard Setting Technical Documentation

Standard setting for the AIMS Reading tests was conducted in early May 2005, using the bookmark standard setting procedure. All technical documentation regarding the standard setting is available in the bookmark standard setting technical report, available from the ADE at http://www.azed.gov/assessment/files/2014/05/az-aims-reading-math-ss-report-2005.pdf.

Standard setting for the AIMS Science tests was conducted in early June, 2008, using the bookmark standard setting procedure. All technical documentation regarding the standard setting is available in the bookmark standard setting technical report, available from the ADE at http://www.azed.gov/assessment/files/2014/05/aims2008sciencerevisedstandardsettingtechnicalreport.pdf.

Standard setting for the AIMS Mathematics tests was conducted in May and June, 2010, using the bookmark standard setting procedure. All technical documentation regarding the standard setting is available in the bookmark standard setting technical report, available on the ADE website at http://www.azed.gov/assessment/files/2014/05/azaimsmathssreport2010.pdf.

Standard setting for the AIMS Writing tests was conducted in May and June, 2011, using a modified bookmark standard setting procedure. All technical documentation regarding the standard setting is available in the bookmark standard setting technical report, available on the ADE website at http://www.azed.gov/assessment/files/2014/05/az-aims-writing-ss-report-final.pdf.

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

Table 10.1.1
Spring 2015 AIMS
Final Scale Score Ranges by Performance Level

Test	Grade	FFBS	AS	MS	ES
Mathematics	HS	355-469	472-486	488-536	539-685
Reading	HS	500-623	628-670	674-770	776-883
Writing	HS	321-432	434-479	481-583	587-700
Science	4	350-457	462-499	504-542	547-800
Science	8	355-471	475-498	502-529	533-800
Science	HS	200-472	475-496	500-533	537-800

10.2 Classification Consistency and Accuracy

This section describes the analyses conducted to estimate classification consistency and accuracy for the 2015 AIMS grades 4, 8, and high school test administrations. Classification consistency can be defined as the agreement between examinees' performance category classification from two independent administrations of the same test (or two parallel forms of the test). Classification accuracy can be defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston & Lewis, 1995).

In conjunction with internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes pass/fail tests such as the AIMS high school tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance categories. Please see Part 9 of this report for more information on the internal consistency of the AIMS assessments.

For tests such as the AIMS high school assessments, classification consistency is most important for students whose ability is near the pass/fail cut score. Students whose ability is far above or far below the value established for passing are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the standard error of measurement of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. For convenience, the cut scores and their associated standard errors are presented in Table 10.2.3.1. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are most likely to be misclassified. The number of students near the cut scores for each grade and content area can be found in the state scale score frequency distributions presented in Part 8 of this report.

Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000) for the AIMS Reading, Mathematics, and Science assessments. For the AIMS Writing assessment, procedures described by Livingston and Lewis (1995) were used to estimate classification consistency and accuracy. The following description of classification consistency and accuracy is based on the paper by Lee et al. (2002).

10.2.1 Classification Consistency

Assume that θ is a single latent trait measured by a test and denote Φ as a latent random variable. When a test X consists of K items and its maximum number-correct score is N, the marginal probability of the number-correct (NC) score *x* is:

$$P(X = x) = \int P(X = x \mid \Phi = \theta) g(\theta) d\theta, \quad x = 0,1,..., N.,$$

where $g(\theta)$ is the density of θ.

In this report, the marginal distribution $P(X = x)$ is denoted as $f(x)$, and the conditional error distribution $P(X = x \mid \Phi = \theta)$ is denoted as $f(x \mid \theta)$. It is assumed that examinees are classified into one of H mutually exclusive categories on the basis of predetermined H-1 observed score cutoffs, $C_1, C_2, \ldots, C_{H-1}$. Let L_h represent the h^{th} category into which examinees with $C_{h-1} \le X \le C_h$ are classified. $C_0 = 0$ and $C_H =$ the maximum number-correct score. Then, the conditional and marginal probabilities of each category classification are as follows:

$$P(X \in L_h \mid \theta) = \sum_{x=C_{h-1}}^{C_{h-1}} f(x \mid \theta), \quad h = 1, 2, \ldots, \text{H.}$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_{h-1}} f(x \mid \theta) g(\theta) d\theta, \quad h = 1, 2, \ldots, \text{H.}$$

Because obtaining test scores from two independent administrations of AIMS was not feasible due to security, logistic, and cost constraints, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Based on the psychometric model, a symmetric H*H contingency table can be constructed. The elements of H*H contingency table consist of the joint probabilities of the row and column observed category classifications.

That two administrations are independent implies that if X_1 and X_2 represent the raw score random variables on the two administrations, then, conditioned on θ, X_1 and X_2 are independent and identically distributed. Consequently, the conditional bivariate distribution of X_1 and X_2 is:

$$f(x_1, \ x_2 \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta).$$

The marginal bivariate distribution of X_1 and X_2 can be expressed as follows:

$$f(x_1, \ x_2) = \int f(x_1, x_2 \mid \theta) f(\theta) d\theta.$$

Consistent classification means that both X_1 and X_2 fall in the same category. The conditional probability of falling in the same category on the two administrations is:

$$P(X_1 \in L_h,\ X_2 \in L_h \mid \theta) = \left[\sum_{x_1 = C_{h-1}}^{C_{h-1}} f(x_1 \mid \theta) \right]^2,$$
$$h = 1, 2, \ldots, \text{H}.$$

The agreement index P, conditional on theta, is obtained by:

$$P(\theta) = \sum_{h=1}^{H} P(X_1 \in L_h,\ X_2 \in L_h \mid \theta).$$

The agreement index (classification consistency) can be computed as:

$$P = \int P(\theta) g(\theta) d(\theta).$$

The probability of consistent classification by chance, P_C, is the sum of squared marginal probabilities of each category classification:

$$P_C = \sum_{h=1}^{H} P(X_1 \in L_h) P(X_2 \in L_h) = \sum_{h=1}^{H} \left[P(X_1 \in L_h) \right]^2.$$

Then, the coefficient kappa (Cohen, 1960) is:

$$k = \frac{P - P_C}{1 - P_C}$$

10.2.2 Classification Accuracy

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ ($h = 1, 2, \ldots, \text{H}$), and a latent score , $\theta \in \Gamma_w$ ($w = 1, 2, \ldots, \text{H}$), an accurate classification is made when $h = w$. The conditional probability of accurate classification is

$$\gamma(\theta) = P(X \in L_w \mid \theta),$$

where w is the category such that $\theta \in \Gamma_w$.

10.2.3 Classification Consistency and Accuracy Results

For convenience, the cut scores and their associated standard errors are presented in Table 10.2.3.1. Table 10.2.3.2 presents results from the classification consistency and classification accuracy analyses. These results are for classifying students into four performance levels. Included in the table for each grade and content area are case counts (N), classification consistency (Agreement), classification inconsistency (Inconsistency), probability of consistent classification by

chance (Chance), Cohen's Kappa (Kappa), and classification accuracy (Accuracy). Inconsistency is defined as 1-agreement.

The 2015 AIMS classification consistency and accuracy results are consistent with classification analyses from the previous AIMS administration. It is important to note that the classification results are dependent on the number of cut scores maintained in a testing program. Moreover, the acceptability of the classification results should be evaluated with respect to the associated stakes of the testing program. The results for the AIMS assessments are quite consistent with other testing programs with similar structure and purpose.

Table 10.2.3.1
Spring 2015 AIMS
Standard Error of Measurement at Cut Scores

Test	Grade	AS		MS		ES	
		Cut Score	SEM	Cut Score	SEM	Cut Score	SEM
Mathematics	HS	471	10	487	10	537	11
Reading	HS	627	15	674	13	773	17
Writing	HS	433	9	480	9	587	13
Science	4	462	15	500	14	547	15
Science	8	473	15	500	14	532	14
Science	HS	475	14	500	14	537	15

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

Table 10.2.3.2
Spring 2015 AIMS
Classification Consistency and Accuracy

Test	Grade	N	Agreement	Inconsistency	Chance	Kappa	Accuracy
Mathematics	HS	4286	0.83	0.17	0.46	0.69	0.88
Reading	HS	1478	0.81	0.19	0.51	0.62	0.86
Writing	HS	3760	0.73	0.27	0.32	0.61	0.81
Science	4	84113	0.69	0.31	0.27	0.57	0.77
Science	8	82248	0.69	0.31	0.26	0.58	0.76
Science	HS	80038	0.70	0.30	0.28	0.59	0.78

Note: High school results include students in all cohorts. Results for Reading, Mathematics, and Science were computed with the IRT method suggested by Lee, Hanson and Brennan (2002) and Wang, Kolen and Harris (2000). Results for Writing were computed using the Livingston-Lewis procedure (1995), implemented with BB-CLASS (Brennan, 2004)

.

PART 11: SCORING OF OPEN-ENDED ITEMS

Arizona's Instrument to Measure Standards (AIMS) assessment requires students to write an essay in response to a specific writing topic or prompt. Students are given two pages to write their response. Writing is administered to students in high school only this year. The Writing assessment for high school is administered in the spring and fall but is no longer a graduation requirement starting spring 2015. The procedure for scoring these writing essays is described in this chapter. This part of the technical report addresses Standards 2.10, 3.22, 3.23, and 5.9 of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and Standards 2.7, 4.18, 4.20, 6.8, and 6.9 in the new edition of *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

11.1 Scoring Process

Outlined below is the scoring process that the AIMS testing contractor follows. The AIMS Writing essays are scored using the distributed model in which scorers are able to train and score at home. Only the PSC (Performance Scoring Center) supervisory staff which monitors the training and scoring is at a regional site. This procedure is used to score all written essays for the AIMS assessments.

11.1.1 Rangefinding and Rubric Review

No Rangefinding was conducted this year since the operational items selected were already reviewed by committees during the previous year's rangefinding.

11.1.2 Recruiting and Training Scorers

Highly qualified scorers are essential to achieving and maintaining a high degree of reliability in scoring students' responses. Thus, the careful selection of professional scorers to evaluate writing essay response items is critical in scoring the AIMS assessments. Scorers are recruited by our Human Resources department and must have extensive experience scoring this type of writing on previous projects. Scorers must also have performed well based on our internal quality metrics of Inter-rater reliability and validity to have achieved a high enough performance rating on previous multi-trait writing to be recruited for the AIMS assessment. Included in our scorer pool is a core group of veteran scorers with experience in working on other holistic writing projects whose insight, flexibility, and dedication have been demonstrated while working on a range of performance assessments.

Scoring supervisors are chosen from the pool of scorers based on demonstrated expertise in all facets of the scoring process, including strong organizational abilities and training skills. Individuals chosen to perform these assignments possess practical skills, leadership abilities, and sensitivity to interpersonal communication requirements. Supervisors also possess the essential capability of assimilating and helping scorers understand the particular scoring requirements of the AIMS.

Upon being hired, scorers sign a confidentiality agreement in which they pledge to keep all information and student responses confidential. Scorers and scoring supervisors are trained to

thoroughly learn the rubric and score responses according to the scoring guidelines developed for the AIMS assessment.

At the beginning of each scoring project, all scoring supervisors and scorers assigned to the project must complete project-specific training and qualification.

11.1.3 Training

Thorough training is vital to the successful completion of any scoring. PSC content specialists and scoring directors follow a series of prescribed steps to ensure training is consistent and of the highest quality. The PSC staff develops its training materials to facilitate learning through visual, auditory and kinesthetic channels. The training for AIMS is conducted using on-line modules designed to take scorers through the background of the assessment and through the rubric and anchor sets for each item. Scorers are then required to take two sets of practice papers and two sets of qualification papers once they complete the item specific modules. Once the scorer completes the practice and qualification sets for that item and qualifies, the scorer takes a calibration set and then is allowed to score live responses for that item. There is a single holistic writing rubric for scoring all grades for the AIMS. A scorer is only required to train and qualify on a particular item.

Prior to scorer training, the PSC subject leaders conduct scoring supervisor training. A primary goal of this session is to ensure scoring supervisors clearly understand the scoring protocols and the training materials. This ensures all responses are scored in a manner consistent with the scores assigned to the anchor papers and according to the intentions of ADE. Scoring supervisors read and discuss the assessment items along with the rubrics used to score them. Scoring supervisors are then required to complete the online training for the item they are assigned. They must qualify for that item in order to be on the project.

The online project training module includes an introduction to the assessment program whose tests are being scored. It is important for scorers to have an understanding of the history and goals of the assessments and the context within which students' responses are evaluated. This gives them a better understanding of what types of responses can be expected. The scorers receive a description of the scoring criteria applied to the responses. Next, the item specific training module presents the writing prompt to be scored and the scoring rubric itself for that item. These online training modules are listed as follows:

Module 1 – Scoring AIMS Writing (Project Specific)

Module 2 – Scoring Grade ____Writing. (Item Specific)

The primary goal of training is to convey to the scorers the decisions made during training paper selection about what type(s) of responses correspond to each score point and to help scorers internalize the scoring protocol so they may effectively apply those decisions. Scorers are better able to comprehend the scoring guidelines in context, so the rubric is presented in conjunction with the anchor papers. Anchor papers are the primary points of reference for scorers as they internalize the rubric. There are three to four anchor papers for each score point value represented in the rubric. The item specific training modules direct scorers' attention to the score point description from the scoring guide, as well as the illustrative anchor papers, thereby enabling scorers to immediately connect the language of the rubric with actual student performance.

After presentation and discussion of the anchor papers, each scorer is shown two practice sets. Practice papers represent each score point and are used during training to help scorers become familiar with applying the rubric. Some papers clearly represent the score point. Others are selected

because they represent borderline responses. Use of these practice sets provides guidance to scorers in defining the line between score points. Training is a continuous process, and scorers are consistently given feedback as they score. Scorers must then pass one of two qualification sets to show a grasp of all the training they have received.

11.1.4 Quality Control

A variety of reports are produced throughout the scoring process to allow scoring supervisory staff to monitor the progress of the project, the reliability of scores assigned, and individual scorers' work. Those reports include:

- Daily and Cumulative Inter-Rater Reliability Reports by Item and Scorer. These reports provide information about how many times scorers were in exact agreement or assigned adjacent scores. The reliability is computed and is monitored daily and cumulatively for the project.

- Pre-scored Validity responses are channeled through the ePEN scoring system and delivered to scorers randomly. Scoring supervisors can get a sense of each reader's grasp of scoring by monitoring their validity agreement rates.

- Daily and Cumulative Frequency Distributions. These reports show how many times each score point has been assigned to the item being scored by readers. The frequency distributions are produced both on a daily basis and cumulatively for the entire scoring project. This report allows scoring supervisors and scoring directors to see whether scorers have a tendency to score consistently high or low.

The most immediate method of monitoring a scorer's performance is through backreading by scoring supervisors. If a scoring supervisor discovers that a scorer is consistently assigning scores other than those the scoring supervisor would assign, he or she can message that scorer using the backreading function and through the instant messaging system in the ePEN scoring system. The validity papers also have annotations that can be reviewed by that scorer after it is scored, using the original anchor papers and training materials. This immediate check and remedial correction also provide an effective guard against scorer drift.

With the help of the individual scorer reliability and validity reports, the scoring staff can closely monitor each scorer's performance. As a distributed project, scorers are monitored for quality using the scorer exception process. Criteria are entered into the ePEN scoring system for inter-rater reliability (IRR), validity and scoring rate. A scorer must meet and maintain the quality metrics established for AIMS in these areas in order to continue scoring the project. If a scorer fails to meet validity, IRR or scoring rate for the first time, the scorer receives a warning and must contact the scoring staff for feedback. If a scorer fails to meet validity a second time, they must then take and pass a 10 paper calibration set in order to continue on the project. Scorers also receive daily calibration sets which they must pass or be targeted for exception. In addition, scorers that have been warned for low IRR or validity are continuously monitored through backreading.

If a scorer fails the targeted calibration for low validity, the scorer is automatically locked out of the system and can no longer score. The scorer receives an e-mail notification that he or she is dismissed from the project. Scorers who have low IRR or a lower or higher than desired scoring rate are closely monitored in backreading and through reports. If in the opinion of the Scoring Director and Content Specialist, these scorers are still performing below acceptable standard after receiving

sufficient feedback and given every reasonable opportunity to improve; they will be manually locked out of the system and notified by email that they are dismissed from the project.

11.1.5 Appeals

The ADE must approve all requests for appeals submitted by district test coordinators. Once PSC receives a request from ADE that a rescore is required for a student, the process outlined below will be followed.

A PSC Scoring Director will use the following procedure to score appeals for an AIMS Writing response:

- Review the scoring rubric, anchor papers, and training materials.
- Read the student's response.
- Score each response based on the scoring rubric and anchor papers.
- Compare rescored response scores to the original scores. For scores that are the same or adjacent, the original score will stand. For scores that are discrepant, the new score will override the original score.
- Determine if the student's response was written in No. 2 pencil.

Pearson will determine the performance level of the rescored response. If the student's rescored performance level (*Falls Far Below, Approaches, Meets,* or *Exceeds*) is the same as the student's original performance level, Pearson will submit an invoice to the district. No paper reports will be generated.

If the student's rescored performance level is different from the student's original performance level **and** the student's response **was not** in No. 2 pencil, Pearson will submit an invoice to the district. Pearson will provide the district with two copies of the revised Student Report.

If the student's rescored performance level is different from the student's original performance level **and** the student's response **was** in No. 2 pencil, Pearson will provide the district with two copies of the revised Student Report at no charge.

No appeals were received for the fall 2014 or spring 2015 administrations.

11.1.6 Security

To ensure security is never compromised, the following safeguards are employed:

- Scoring materials are accessed by scorers on a secure website that is managed by Pearson Scoring Support. Only scorers hired for AIMS are allowed access.
- No scorers are hired who reside in Arizona as per contract requirement.
- Scorers and scoring staff personnel must sign a non-disclosure and confidentiality form in which they agree not to use or divulge any information concerning the tests.
- Any contact with the press is handled through ADE

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. (2104). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Angoff, W. (1993). Perspectives on differential item functioning methodology. In P.W. Holland & H. Warner (Eds.), *Differential item functioning* (pp. 3-24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Arizona Department of Education. (2012). *Arizona's Instrument to Measure Standards 2012 technical report*. Iowa City, IA: Pearson.

Arizona Department of Education. (2013). *Arizona's Instrument to Measure Standards 2013 technical report*. Iowa City, IA: Pearson. Retrieved from http://www.azed.gov/assessment/files/2014/05/aims_tech_report_2013_final.pdf.

Arizona Department of Education. (2014). *Arizona's Instrument to Measure Standards 2014 technical report*. Iowa City, IA: Pearson. Retrieved from http://www.azed.gov/assessment/files/2014/12/aims_tech_report_2014_final.pdf.

Arizona Department of Education. (2005*). Bookmark standard setting technical report for grades 3,5,8, and high school reading and mathematics.* Monterey, CA: CTB/McGraw-Hill.

Arizona Department of Education. (2009). *Bookmark standard setting technical report 2008 for grades 4, 8, and high school science.* Monterey, CA: CTB/McGraw-Hill.

Arizona Department of Education. (2010). *Standard setting report: Mathematics grades 3-8 and high school.* Iowa City, IA: Pearson.

Arizona Department of Education. (2011). *Standard setting report: Writing grades 5, 6, 7 and high school*. Iowa City, IA: Pearson.

Brennan, R. L. (2004). BB-CLASS: A computer program that uses the beta-binomial model for classification consistency and accuracy [Computer program]. Iowa City, IA: The University of Iowa Center for Advanced Studies in Measurement and Assessment.

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.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17,* 31-44.

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.

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

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.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

Linacre, J. M. (2005). WINSTEPS Rasch measurement [Computer software]. Chicago: WINSTEPS.com.

Linacre, J. M. (2011). A User's Guide to WINSTEPS Rasch-Model Computer Programs. 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.

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

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. (1992). *BIGSTEPS Rasch analysis* [Computer software]. Chicago, IL: MESA Press.

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.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago, IL: 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.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 337-348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 26*, 44-66.

APPENDIX A: FALL 2014 AIMS ADMINISTRATION

Overview:

The Fall 2014 AIMS Reading, Writing, and Mathematics assessments were administered to students in high school who were in Grades 11 and 12 and had not yet obtained a passing score on one or more of the content areas. These assessments were written to the Arizona content standards and were designed to measure student performance with respect to these standards. The Fall 2014 AIMS Reading and Mathmatics assessments were entirely multiple-choice tests. The Fall 2014 AIMS Writing assessment consisted of multiple-choice items and a single extended-response essay prompt.

Test Design, Development, and Administration:

The design and development of the Fall 2014 AIMS assessments reflect the same guiding principles that were followed for the Spring 2014 AIMS assessments. Arizona teachers, curriculum specialists, and administrators from across the state were an integral part of the AIMS test development process. More information regarding the committee meetings, the test blueprints, and the test development process is contained in Parts 3 and 4 of this technical report.

The test design and structure of the Fall 2014 AIMS assessments mirrored the structure of the AIMS assessments administered in spring. That is, the AIMS Reading test consisted of 54 multiple-choice items developed by Arizona teachers. Raw scores ranged from 0-54 and scale scores ranged from 500 to 900. Similarly, the AIMS Mathematics test consisted of 85 multiple-choice items developed by Arizona teachers. Raw scores ranged from 0-85 and scale scores ranged from 500 to 900. Finally, the AIMS Writing assessment consisted of one extended-response writing prompt and 27 multiple-choice items. The multiple-choice component is weighted 40% and the essay response is weighted 60% in the total score. Responses to the prompt were scored on the holistic six-point rubric (see appendix D). Each essay response received two ratings. Final scores for responses with adjacent ratings were derived by averaging the two ratings. Final scores for responses with discrepant ratings (difference of 2 points) were resolved by a third rater. The raw scores ranged from 0-138, and scale scores were designed to range from 300-700. There are no norm-referenced items included in the high school AIMS assessments.

The test selection process for the Fall 2014 AIMS was designed such that resulting tests matched the Spring 2014 AIMS in overall test difficulty and content coverage. The test selection process is described in Part 4 of this report.

Test administration procedures for the Fall 2014 AIMS are described in Part 5 of this report.

Scaling and Equating:

The Fall 2014 AIMS administration was designed for students who were retaking the assessment because they had not obtained a passing score on one or more of the content areas. The population of students that retake the AIMS assessments varies from administration to administration in terms of its composition and achievement. Moreover, expediency in the reporting of results to the population of students who are retaking the AIMS assessments was

considered to be a priority. For these reasons, a pre-equating model was adopted for the fall AIMS administration.

Pre-equating takes advantage of an existing bank of previously calibrated and scaled items such that an equated form and an associated number correct to scale score conversion table can be constructed prior to operational administration. The fall 2014 assessments were constructed from items that had been previously administered in the Spring 2014 AIMS administration and had been calibrated and linked to the operational scale of measurement using the Rasch measurement model. Using the existing item parameters for the items selected to be on the Fall 2014 AIMS assessment, the number correct to AIMS scale score conversion tables were created. The raw score to scale score conversions, including the standard error of measurement (SEM) for each scaled score, are presented in Tables A.1 through A.3.

Fall 2014 AIMS Test Results:

The results of the Fall 2014 AIMS are reported in Tables A.4 through A.6. Results presented below are based on population data contained within the final electronic data files. The results presented 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 final results typically use more detailed school-level information than is used to conduct research analyses. The results in the following tables are presented as evidence of reliability and validity of the AIMS assessments and should not be used for state accountability purposes.

Fall 2014 AIMS was the last administration of Fall AIMS Reading, Mathematics, and Writing assessments in high school. The results of the Fall 2014 AIMS are no longer broken out by cohort. Disaggregated results were produced for the various groups by using demographic data on student answer documents.

Table A.1
Fall 2014 AIMS Raw Score to Scale Score Table
Mathematics High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	59	43	483	10
1	300	42	44	485	10
2	315	30	**45**	**488**	**10**
3	333	25	46	490	10
4	345	22	47	492	10
5	355	19	48	494	10
6	364	18	49	497	10
7	371	17	50	499	10
8	378	16	51	501	10
9	383	15	52	504	10
10	389	14	53	506	10
11	393	14	54	508	10
12	398	13	55	511	10
13	402	13	56	513	10
14	406	13	57	516	10
15	410	12	58	518	10
16	413	12	59	521	10
17	417	12	60	524	11
18	420	12	61	526	11
19	423	11	62	529	11
20	426	11	63	532	11
21	429	11	64	535	11
22	432	11	**65**	**538**	**11**
23	435	11	66	541	11
24	438	11	67	544	12
25	440	10	68	548	12
26	443	10	69	551	12
27	446	10	70	555	12
28	448	10	71	559	13
29	451	10	72	563	13
30	453	10	73	567	14
31	456	10	74	572	14
32	458	10	75	576	15
33	460	10	76	582	15
34	463	10	77	588	16
35	465	10	78	594	17
36	467	10	79	602	18
37	470	10	80	610	20
38	**472**	**10**	81	621	22
39	474	10	82	634	25
40	476	10	83	652	30
41	479	10	84	681	42
42	481	10	85	700	59

Note: SEM is the standard error of measurement for the scale score**.**

Table A.2
Fall 2014 AIMS Raw Score to Scale Score Table
Reading High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	500	61	28	684	13
1	500	44	29	688	13
2	525	32	30	692	13
3	544	26	31	696	13
4	558	23	32	699	13
5	569	21	33	703	13
6	579	19	34	707	13
7	587	18	35	711	13
8	594	17	36	715	13
9	601	17	37	719	13
10	607	16	38	724	14
11	613	15	39	728	14
12	618	15	40	733	14
13	623	15	41	738	15
14	**628**	**14**	42	743	15
15	633	14	43	748	15
16	637	14	44	754	16
17	641	14	45	760	16
18	646	13	46	766	17
19	650	13	**47**	**773**	**18**
20	654	13	48	782	19
21	658	13	49	791	21
22	662	13	50	802	23
23	665	13	51	816	26
24	669	13	52	835	31
25	673	13	53	866	44
26	**677**	**13**	54	900	61
27	680	13			

Note: SEM is the standard error of measurement for the scale score.

Table A.3
Fall 2014 AIMS Raw Score to Scale Score Table
Writing High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	54	35	367	10	70	445	9	105	531	11
1	300	54	36	369	10	71	448	9	106	534	11
2	300	54	37	372	10	72	450	9	107	537	11
3	300	54	38	375	10	73	452	9	108	540	11
4	300	54	39	377	10	74	454	9	109	543	11
5	300	54	40	380	10	75	456	9	110	546	11
6	300	54	41	382	10	76	458	9	111	549	11
7	300	54	42	385	10	77	460	9	112	553	11
8	300	54	43	387	10	78	462	9	113	556	11
9	300	54	44	390	10	79	464	9	114	559	11
10	300	54	45	392	10	80	466	9	115	562	11
11	300	54	46	394	10	81	469	9	116	566	11
12	300	54	47	397	9	82	471	9	117	569	12
13	300	54	48	399	9	83	473	9	118	572	12
14	300	54	49	401	9	84	475	9	119	576	12
15	300	38	50	403	9	85	477	9	120	579	12
16	300	26	51	406	9	**86**	**480**	**10**	121	583	12
17	300	21	52	408	9	87	482	10	**122**	**587**	**12**
18	300	18	53	410	9	88	484	10	123	591	13
19	306	17	54	412	9	89	487	10	124	595	13
20	313	15	55	414	9	90	489	10	125	599	13
21	318	14	56	416	9	91	492	10	126	604	14
22	323	14	57	418	9	92	494	10	127	609	14
23	327	13	58	421	9	93	497	10	128	614	14
24	332	12	59	423	9	94	499	10	129	619	15
25	335	12	60	425	9	95	502	10	130	625	16
26	339	12	61	427	9	96	505	10	131	632	17
27	343	12	62	429	9	97	508	10	132	640	18
28	346	11	63	431	9	98	510	10	133	648	19
29	349	11	**64**	**433**	**9**	99	513	11	134	658	21
30	352	11	65	435	9	100	516	11	135	671	24
31	355	11	66	437	9	101	519	11	136	688	29
32	358	11	67	439	9	102	522	11	137	700	40
33	361	11	68	441	9	103	525	11	138	700	56
34	364	10	69	443	9	104	528	11			

Note: SEM is the standard error of measurement for the scale score. The writing scores range from 0 to 138 because the single essay response is scored by two raters on a scale of 1 to 6 on six traits. The trait scores are summed to produce the final score. Scores below 12 are artificial.

Table A.4
Fall 2014 AIMS Results
Mathematics High School

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Total	42690	469.95	25.74	51	24	24	1
Ethnic Background							
Hispanic	23184	468.51	23.83	53	25	22	1
Non Hispanic	18688	471.6	27.59	49	23	26	2
White	28665	471.04	25.7	49	24	25	1
Black	3657	466.28	24.55	57	23	19	1
American Indian	4077	465.06	23.36	59	23	18	0
Asian	715	481.76	37.15	40	22	30	9
Secial Program Membership							
English Learner Program	1558	454.47	22.82	78	13	8	0
Special Education	6844	449.44	20.39	85	10	5	0
SES	25866	467.8	24.14	54	24	21	1
Migrant	520	462.97	21.52	64	21	15	0

Note: FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, off-grade, or a non-standard accommodation are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. The results are not final results and are presented here for purpose of addressing reliability and validity. These results should not be used for accountability purposes.

Table A.5
Fall 2014 AIMS Results
Reading High School

Group	N	Scale Score		% at Performance Level			
		M	SD	FFBS	AS	MS	ES
Total	17624	679.48	39.83	7	42	49	2
Ethnic Background							
Hispanic	9161	673.81	35.36	7	46	46	1
Non Hispanic	7983	685.13	43.24	6	39	51	4
White	11164	681.2	40.87	7	41	49	3
Black	1657	675.98	37.19	7	45	47	1
American Indian	1746	668.87	32.45	9	50	41	0
Asian	518	688.12	46.54	7	40	49	5
Secial Program Membership							
English Learner Program	1409	651.22	24.48	13	72	15	0
Special Education	4268	652.88	28.08	16	63	21	0
SES	10923	672.98	35.64	8	47	44	1
Migrant	303	664.67	28.88	8	57	35	0

Note: FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, off-grade, or a non-standard accommodation are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. The results are not final results and are presented here for purpose of addressing reliability and validity. These results should not be used for accountability purposes.

Table A.6
Fall 2014 AIMS Results
Writing High School

Group	N	Scale Score		% at Performance Level			
		M	SD	FFBS	AS	MS	ES
Total	29369	467.61	42.84	17	46	35	1
Ethnic Background							
Hispanic	15695	462.47	37.49	18	50	31	1
Non Hispanic	12945	472.84	46.61	16	43	39	2
White	19248	469.6	43.2	16	45	37	2
Black	2354	462.02	40.25	20	49	31	1
American Indian	2931	457.37	34.43	21	52	26	0
Asian	759	478.74	59.37	19	39	36	6
Secial Program Membership							
English Learner Program	1659	423.69	33.19	57	39	4	0
Special Education	5753	437.53	33.85	43	47	11	0
SES	18169	461.24	38.02	20	49	30	1
Migrant	445	452	36.06	26	51	23	0

Note: FFBS= Falls Far Below the Standard; AS= Approaches the Standard; MS= Meets the Standard; ES= Exceeds the Standard. Students with no valid attempt, invalidation, off-grade, or a non-standard accommodation are not included in this summary. In addition, home-schooled students, students attending Bureau of Indian Affairs schools, students attending juvenile corrections centers, students attending state hospital schools, and students who already met expectations in a previous test administration are not included in this summary. The results are not final results and are presented here for purpose of addressing reliability and validity. These results should not be used for accountability purposes.

APPENDIX B: COMMITTEE MEMBER SELECTION CRITERIA

AIMS Committee Participant Selection Criteria

ARIZONA DEPARTMENT OF EDUCATION

PROCEDURE FOR SELECTION OF EDUCATOR COMMITTEES
ARIZONA ASSESSMENT SECTION

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.
- 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
- After decline and accept emails are received by established deadline, additional invitations issued to members on alternate list
- Once the committee meeting is held, performance of participants is reviewed.

Recognition of existing AIMS committee participants is an important aspect of retaining our Ambassadors of the Assessment; therefore, after each committee meeting, each participant receives a letter recognizing their excellent contributions to the assessment program and to all Arizona students

APPENDIX C: **AIMS HOLISTIC RUBRIC BASED ON 6 TRAITS OF WRITING**

HOLISTIC RUBRIC BASED ON 6 TRAITS OF WRITING

SCORE POINT 6	SCORE POINT 5	SCORE POINT 4
Response is sophisticated and skillful in written communication, demonstrated by • exceptional clarity, focus, and control in topic development and organization that often show insight. • in-depth and/or creative exploration of the topic using rich, relevant, and credible details. • a strong, perhaps creative, beginning and a satisfying conclusion. • specifically and carefully chosen words that are skillfully crafted into phrases and sentences that enhance meaning. • intentional and committed interaction between the writer and the reader. • effective and/or creative use of a wide range of conventions with few errors.	Response is excellent and skillful in written communication, demonstrated by • clarity, focus, and control in topic development and organization. • a balanced and thorough exploration of the topic using relevant details. • an inviting beginning and a satisfying sense of closure. • a broad range of carefully chosen words crafted into phrases and varied sentences that sound natural. • awareness of the reader and commitment to the audience and topic. • effective use of a wide range of conventions with few errors.	Response is appropriate and acceptable in written communication, demonstrated by • ideas adequately developed with a clear and coherent presentation of ideas with order and structure that can be formulaic. • relevant details that are sometimes general or limited; organization that is clear, but sometimes predictable. • a recognizable beginning and ending, although one or both may be somewhat weak. • effective word choice that is functional and, at times, shows interaction between writer and audience. • somewhat varied sentence structure with good control of simple constructions; a natural sound. • control of standard conventions although a wide range is not used; errors that do not impede readability.

SCORE POINT 3	SCORE POINT 2	SCORE POINT 1
Response is inadequate in written communication, demonstrated by • broad or simplistic ideas that are understood but often ineffective. • attempts at organizing that are inconsistent or ineffective; beginnings and endings that are underdeveloped; repetitive transitional devices. • developmental details that are uneven, somewhat predictable, or leave information gaps; details not always placed effectively in the writing. • reliance on clichés and overused words that do not connect with the reader; limited audience awareness. • monotonous and sometimes misused words; sentences may sound mechanical, although simple constructions are usually correct. • limited control of standard conventions with significant errors.	Response is poor in written communication, demonstrated by • overly simplistic and sometimes unclear ideas that have insufficiently developed details. • sequencing of ideas that is often just a list; missing or ineffective details that require reader inference to comprehend and follow. • missing beginning and/or ending. • repetitive, monotonous, and often misused words awkwardly strung into sentences that are difficult to read because they are either choppy or rambling; many sentences that begin with repetitive noun + verb pattern. • lack of audience awareness. • little control of basic conventions resulting in errors impeding readability.	Response is inferior in written communication, demonstrated by • lack of purpose or ideas and sequencing. • organization that obscures the main point. • an attempt that is too short to offer coherent development of an idea, if it is stated. • extremely limited vocabulary that shows no commitment to communicating a message. • sentences with confusing word order that may not permit oral reading. • severe and frequent errors in conventions.

HOLISTIC RUBRIC BASED ON SIX TRAITS OF WRITING

Score Point 1: The writing skill in a Score Point 1 response is inferior. A paper receiving a SP 1 can be long or short. Length alone is not sufficient reason to score the response a SP 1. One or more bullets in the SP 1 rubric will describe this paper. It may also satisfy a bullet from a higher SP, but the majority of the response attributes are SP 1.

Score Point 2: The writing skill in a Score Point 2 response is poor. Ideas are presented but are not clear or organized. The reader must often reread and infer meaning because of the poor quality of writing. As with the SP 1, this paper may satisfy a bullet from a higher SP, but it fits into the SP 2 rubric more clearly than another.

Score Point 3: The writing in a Score Point 3 response is less than adequate. Ideas are simple; organization is inconsistent; development is uneven. While this paper may meet one or two bullets in a higher score, the majority of the attributes fall in the SP 3 rubric.

Score Point 4: The writing skill in a Score Point 4 response is appropriate and acceptable, demonstrating competent written communication. Clear ideas are organized coherently, although the paper is often formulaic. The message is presented so that most readers can easily understand it. The attributes from higher score points may emerge in some papers, but unless the paper satisfies a majority of bullets above the SP 4, it will earn a SP 4.

Score Point 5: The writing skill in a Score Point 5 response is excellent and shows skill above acceptable. This paper presents ideas clearly, organizes them with care, and uses vocabulary and sentences that demonstrate thoughtful choice and craftsmanship. A SP 5 paper may satisfy a bullet or two in SP 6, but it still does not meet a majority of the criteria in the SP 6. The majority of the SP 5 bullets will be represented.

Score Point 6: The writing skill in a Score Point 6 response is exemplary. The exceptional and sophisticated craftsmanship shows a thoughtful and exacting writer who strives to communicate clearly and creatively. While the paper may not fit every bullet in the SP 6 rubric, it will meet the majority of them. A SP 6 paper is not perfect. Errors may be noticed, but they do not detract from the message.

Condition codes for non-scored papers: BL = Blank; IL = illegible; NE = non-English; and OT = off-topic.

Always score papers by recognizing what the student has done well before looking for errors. Remember that the student is writing to a cold prompt with no assistance from a teacher. Score for the skills demonstrated. An error that is repeated still only counts as one error. For instance, if a student misspells “beautiful” four times, it is evident that he/she cannot spell “beautiful;” it is one error. Or if a student doesn’t put a comma after introductory clauses throughout the paper, it is evident that he/she does not apply this comma rule; it is one error. If a student misuses the word “except” for “accept,” the same applies whether it is done one time or five times. Strive to assign the score that is best represented by the bullets in a score point.