Arizona

Arizona's Instrument to Measure Standards

2017
Technical Report

Submitted to the
Arizona Department of Education
October 2017

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

FOREWORD ..... **iii**

TABLE OF CONTENTS ..... **iv**

FIGURES AND TABLES ..... **vi**

PART 1: EXECUTIVE SUMMARY **1**

PART 2: INVOLVEMENT OF ARIZONA EDUCATORS AT ALL LEVELS **4**

PART 3: TEST DESIGN **5**

3.1 CONTENT STANDARDS 5
3.2 TEST BLUEPRINTS 7
3.3 DESCRIPTION OF 2017 AIMS TESTS 11
3.3.1 Science for Grades 4, 8, and High School *11*
3.3.2 AIMS Score Ranges *11*

PART 4: TEST DEVELOPMENT **12**

4.1 AIMS TEST DEVELOPMENT AND EDITING PROCESS 12
4.1.1 Test Development Process *12*
4.1.2 Documents and Materials Development *12*
4.1.3 Item Writing for Science *12*
4.1.4 Quality Reviews *12*
4.2 POOL OF ITEMS USED FOR TEST CONSTRUCTION 13
4.2.1 Item Specifications *13*
4.2.2 Data Analysis *13*
4.2.3 AIMS Item Selection *14*
4.3 CUSTOMER APPROVALS 19
4.3.1 Item Selection Approval *19*
4.3.2 Test Book Approvals *19*
4.3.3 FTP Site *19*
4.3.4 Final Forms Review (Pearson) *19*
4.3.5 ADE Quality Review *20*
4.3.6 Final Sign-off *20*

PART 5: TEST ADMINISTRATION **21**

5.1 ACCOMMODATIONS 21
5.1.1 Overview of Accommodations *21*
5.1.2 Descriptions of Universal and Standard Accommodations *22*
5.1.3 Determining if a Student Needs a Testing Accommodation *22*
5.1.4 Reporting Results of Assessments Taken with Accommodations *24*
5.2 TEST SECURITY 24
5.3 TEST ADMINISTRATION 27

PART 6: CLASSICAL ITEM ANALYSIS **29**

6.1 DATA 29
6.2 DESCRIPTIVE STATISTICS BY TEST 29
6.3 CLASSICAL ITEM ANALYSIS 30

PART 7: CALIBRATION, SCALING AND EQUATING **37**

7.1 ENSURING VALID RECORDS IN CALIBRATION SAMPLE 37
7.2 CALIBRATION METHODS 37
7.2.1 Calibration Model *37*

7.2.2 Calibration Software *38*
7.3 CALIBRATION RESULTS 38
7.3.1 IRT Item Statistics *38*
7.4 SCALING METHODS 42
7.4.1 Science *42*
7.5 EQUATING 42
7.5.1 Science *42*
7.5.2 Scoring and Standard Error of Measurement *45*

PART 8: TEST RESULTS 49

8.1 DATA 49
8.1.1 AIMS State Test Results *49*
8.2 LONGITUDINAL DATA 56

PART 9: VALIDITY EVIDENCE 59

9.1 RELIABILITY 59
9.1.1 Measures of Internal Consistency *59*
9.2 VALIDITY 64
9.2.1 Differential Item Functioning *65*

PART 10: CLASSIFICATION 68

10.1 STANDARD SETTING TECHNICAL DOCUMENTATION 68
10.2 CLASSIFICATION CONSISTENCY AND ACCURACY 68
10.2.1 Classification Consistency *69*
10.2.2 Classification Accuracy *71*
10.2.3 Classification Consistency and Accuracy Results *71*

REFERENCES 72

APPENDIX A: COMMITTEE MEMBER SELECTION CRITERIA 74

APPENDIX B: DIF Results 75

FIGURES AND TABLES

TABLE 3.1.1 ARIZONA SCIENCE CONCEPTS AND STRANDS – GRADE 4 5
TABLE 3.1.2 ARIZONA SCIENCE CONCEPTS AND STRANDS – GRADE 8 6
TABLE 3.1.3 ARIZONA SCIENCE CONCEPTS AND STRANDS – HIGH SCHOOL 7
TABLE 3.2.1 AIMS BLUEPRINT FOR SCIENCE GRADE 4 8
TABLE 3.2.2 AIMS BLUEPRINT FOR SCIENCE GRADE 8 9
TABLE 3.2.3 AIMS BLUEPRINT FOR SCIENCE HIGH SCHOOL 10
TABLE 3.3.1.1 SPRING 2017 AIMS TEST STRUCTURE OF SCIENCE 11
TABLE 3.3.2.1 RAW SCORE AND SCALE SCORE RANGES OF 2017 AIMS ASSESSMENTS 11
TABLE 4.1 ITEMS GIVEN SPECIAL CODES 14
FIGURE 4.1 SAMPLE GRADE 4 SCIENCE ITEM POOL TABLE 16
FIGURE 4.2 SAMPLE GRADE 4 SCIENCE ITEM REPLACEMENT TABLE 17
FIGURE 4.3 SAMPLE P-VALUE TARGET TABLE AND GRAPH 18
TABLE 4.2 NUMBER OF SCIENCE ITEMS SELECTED BY COMMITTEE 19
FIGURE 5.2.1 SPRING 2017 AIMS TEST SECURITY AGREEMENT FOR SUPERINTENDENTS/CHARTER REPRESENTATIVES AND DISTRICT TEST COORDINATORS 25
FIGURE 5.2.2 SPRING 2017 AIMS TEST SECURITY AGREEMENT FOR ALL SCHOOL/DISTRICT/CHARTER PERSONNEL 26
TABLE 6.2.1 SPRING 2017 AIMS SCIENCE CLASSICAL TEST ANALYSIS STATISTICS 29
TABLE 6.3.1 SPRING 2017 AIMS CLASSICAL ITEM ANALYSIS SCIENCE GRADE 4 31
TABLE 6.3.2 SPRING 2017 AIMS CLASSICAL ITEM ANALYSIS SCIENCE GRADE 8 33
TABLE 6.3.3 SPRING 2017 AIMS CLASSICAL ITEM ANALYSIS SCIENCE GRADE 10 35
TABLE 7.3.1.1 SPRING 2017 AIMS IRT ITEM STATISTICS SCIENCE GRADE 4 39
TABLE 7.3.1.2 SPRING 2017 AIMS IRT ITEM STATISTICS SCIENCE GRADE 8 40
TABLE 7.3.1.3 SPRING 2017 AIMS IRT ITEM STATISTICS SCIENCE GRADE HS 41
TABLE 7.5.1.1 SPRING 2017 AIMS SCIENCE ANCHOR ITEMS 42
TABLE 7.5.1.2 REPRESENTATION OF CONTENT BY 2017 SCIENCE ANCHOR SETS 43
TABLE 7.5.1.3 REPRESENTATION OF DIFFICULTY BY 2017 SCIENCE ANCHOR SETS 44
TABLE 7.5.2.1 SPRING 2017 AIMS RAW SCORE TO SCALE SCORE TABLE SCIENCE GRADE 4 46
TABLE 7.5.2.2 SPRING 2017 AIMS RAW SCORE TO SCALE SCORE TABLE SCIENCE GRADE 8 47
TABLE 7.5.2.3 SPRING 2017 AIMS RAW SCORE TO SCALE SCORE TABLE SCIENCE HIGH SCHOOL 48
TABLE 8.1.1.2 SPRING 2017 AIMS STATE TEST RESULTS SCIENCE GRADES 4 AND 8 50
TABLE 8.1.1.3 SPRING 2017 AIMS STATE TEST RESULTS SCIENCE HIGH SCHOOL 51
TABLE 8.1.1.4 SPRING 2017 AIMS FREQUENCY DISTRIBUTION SCIENCE CRT GRADE 4 52
TABLE 8.1.1.5 SPRING 2017 AIMS FREQUENCY DISTRIBUTION SCIENCE CRT GRADE 8 53
TABLE 8.1.1.6 SPRING 2017 AIMS FREQUENCY DISTRIBUTION SCIENCE CRT HIGH SCHOOL COHORT 19 54
TABLE 8.1.1.7 SPRING 2017 AIMS FREQUENCY DISTRIBUTION SCIENCE CRT HIGH SCHOOL COHORT 20 55
TABLE 8.2.1 LONGITUDINAL COMPARISON OF SCALE SCORES IN SCIENCE 57
TABLE 8.2.2 LONGITUDINAL COMPARISON OF PERFORMANCE LEVEL DISTRIBUTION IN SCIENCE 58
TABLE 9.1.1.1 SPRING 2017 AIMS SCIENCE INTERNAL CONSISTENCY 60
TABLE 9.1.1.2 SPRING 2017 AIMS STRAND/CONCEPT INTERNAL CONSISTENCY SCIENCE GRADE 4 61
TABLE 9.1.1.3 SPRING 2017 AIMS STRAND/CONCEPT INTERNAL CONSISTENCY SCIENCE GRADE 8 62
TABLE 9.1.1.4 SPRING 2017 AIMS STRAND/CONCEPT INTERNAL CONSISTENCY SCIENCE HIGH SCHOOL 63
TABLE 9.2.1.1 DIFFERENTIAL ITEM FUNCTIONING FLAG CATEGORIES 66
TABLE 9.2.1.2 DIF STATISTICS FOR ITEMS EXHIBITING STRONG DIF 67
TABLE 10.1.1 SPRING 2017 AIMS SCIENCE FINAL SCALE SCORE RANGES BY PERFORMANCE LEVEL 68
TABLE 10.2.1 SPRING 2017 AIMS STANDARD ERROR OF MEASUREMENT AT CUT SCORES 69
TABLE 10.2.3.1 SPRING 2017 AIMS CLASSIFICATION CONSISTENCY AND ACCURACY 71
TABLE B.1 SPRING 2017 AIMS DIFFERENTIAL ITEM FUNCTIONING SCIENCE CRT GRADE 4 75
TABLE B.2 SPRING 2017 AIMS DIFFERENTIAL ITEM FUNCTIONING SCIENCE CRT GRADE 8 79
TABLE B.3 SPRING 2017 AIMS DIFFERENTIAL ITEM FUNCTIONING SCIENCE CRT HIGH SCHOOL 83

PART 1: EXECUTIVE SUMMARY

This document provides information regarding processes and procedures implemented in the Spring 2017 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 Spring 2017 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 (the Executive Summary) 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.

Structure of AIMS Technical Report

The Spring 2017 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.*

The Spring 2017 AIMS tests were 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. Science tests remain mandatory for all students in these grades. 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 Science testing. 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' 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. 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 and the work they performed to help prepare the 2017 AIMS Science 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 2016-2017 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, security, and the written procedures available to all test administrations and school personnel and the accommodations that were available to eligible students while testing on Spring 2017 AIMS Science. This section also describes instituted procedures to ensure the security and standardization of test administrations.

Data for Operational Analysis

- Part 6 describes the data used for calibration and scaling of the Spring 2017 AIMS Science and presents classical test statistics and item analysis statistics. This section includes steps taken to ensure the valid calibration and scaling of these tests as well as the resulting measures of internal consistency.

Calibration, Scaling, and Equating

- Part 7 reviews calibration, equating, scoring methods, and calibration results. This section includes considerations for the evaluation of the calibration results and anchor items. It also presents the relationships between raw scores and scale score through scoring tables and scaling results including the standard error of measurement.

Test Results

- Part 8 summarizes information about the results of the Spring 2017 AIMS Science administration. The test results for different ethnic backgrounds and special program membership status are provided. Students in cohorts 2019 and 2020 are included

separately in the high school science results. The results presented include frequency distributions and longitudinal comparisons of scale scores.

Validity Evidence

- Part 9 reviews the main validity issues discussed in all prior chapters and provides additional validity evidence supporting the AIMS Science tests. The evidence presented includes the results of an analysis of differential item functioning.

Classification

- Part 10 provides information regarding classification consistency and accuracy when students were classified into proficiency categories. These analyses used cut scores that 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 in preparation for AIMS 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. The criteria on selecting the Arizona educators for the committee meetings are presented in Appendix A. However, starting spring 2015, because ADE had developed a sufficient number and quality of items in the Science item bank, they chose to change their process for the development of the spring assessments.

The Spring 2017 AIMS Science called for administering one operational test for grade 4, 8, and high school in science. All items available for placement on an operational test had been previously field-tested, and passed through multiple educator committees, including development, bias and content, and data analysis meetings, prior to the start of the development of these tests. The AIMS Science tests for the spring 2017 administration were built during the summer of 2014 to match the blueprint, difficulty distribution, and include as many higher Depth-of-Knowledge (DOK) items as possible. Item selections for these tests were performed by trained ADE staff, most of whom also held Arizona teacher certificates.

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 2005 for science. These standards are organized by strand, concept, and performance objective. The AIMS Science test blueprints are based on the concepts and strands of the Arizona content standards, presented in Figures 3.1.1 through 3.1.3.

Table 3.1.1
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

Table 3.1.2
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

TABLE 3.1.3
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. AIMS assessments in science were designed in accordance with blueprints provided in Tables 3.2.1 through 3.2.3. Further discussion of item selection to match the blueprints is included in Part 4 of this report.

Table 3.2.1
AIMS Blueprint for Science Grade 4

AIMS Science
Grade 4 Test Blueprint

Strand/Concept	% of Test
Strand 1: Inquiry Process	**33.3%**
Concept 1: Observations, Questions, and Hypotheses	11.1%
Concept 2: Scientific Testing (Investigating and Modeling)	11.1%
Concept 3: Analysis and Conclusions	11.1%
Concept 4: Communications	
Strand 2: History and Nature of Science	**11.1%**
Concept 1: History of Science as a Human Endeavor	11.1%
Concept 2: Nature of Scientific Knowledge	
Strand 3: Science in Personal and Social Perspectives	**11.1%**
Concept 1: Changes in Environments	11.1%
Concept 2: Science and Technology in Society	
Strand 4: Life Science	**11.1%**
Concept 1: Characteristics of Organisms	11.1%
Concept 3: Organisms and Environments	
Concept 4: Diversity, Adaptations, and Behavior	
Strand 5: Physical Science	**11.1%**
Concept 3: Energy and Magnetism	11.1%
Strand 6: Earth and Space Science	**22.2%**
Concept 2: Earth's Processes and Systems	11.1%
Concept 3: Changes in the Earth and Sky	11.1%

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).

Source: https://cms.azed.gov/home/GetDocumentFile?id=584eddeaaadebe050c573ef7

Table 3.2.2
AIMS Blueprint for Science Grade 8

AIMS Science
Grade 8 Test Blueprint

Strand/Concept	% of Test
Strand 1: Inquiry Process	**34.5%**
Concept 1: Observations, Questions, and Hypotheses	10.3%
Concept 2: Scientific Testing (Investigating and Modeling)	6.9%
Concept 3: Analysis and Conclusions	10.3%
Concept 4: Communications	6.9%
Strand 2: History and Nature of Science	**10.3%**
Concept 1: History of Science as a Human Endeavor	10.3%
Concept 2: Nature of Scientific Knowledge	
Strand 3: Science in Personal and Social Perspectives	**10.3%**
Concept 1: Changes in Environments	10.3%
Concept 2: Science and Technology in Society	
Strand 4: Life Science	**13.8%**
Concept 2: Reproduction and Heredity	13.8%
Concept 4: Diversity, Adaptations, and Behavior	
Strand 5: Physical Science	**31.0%**
Concept 1: Properties and Changes of Properties in Matter	17.2%
Concept 2: Motion and Forces	13.8%

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).

Source: https://cms.azed.gov/home/GetDocumentFile?id=584ede02aadebe050c573efb

Table 3.2.3
AIMS Blueprint for Science High School

AIMS Science
High School Test Blueprint

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

Source: https://cms.azed.gov/home/GetDocumentFile?id=584ede25aadebe050c573eff

3.3 Description of 2017 AIMS Tests

The test blueprints were used with the processes described in detail in Part 4 to develop all AIMS tests administered in 2017. The resulting test configurations are as follows.

3.3.1 Science for Grades 4, 8, and High School

The 2017 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 items on each test were developed by Arizona teachers, were operational, and reported to a criterion-referenced score. Since item development and associated field-testing (FT) had been halted starting with the Spring 2014 administration, no field-test items were included on any of the science tests. The scale scores for each test range from 200 to 800. Table 3.3.1.1 displays the structure of the science tests.

Table 3.3.1.1
Spring 2017 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 the spring 2017 tests.

3.3.2 AIMS Score Ranges

Raw score and scale score ranges of 2017 AIMS Science in grades 4, 8, and high school are presented in Table 3.3.2.1.

Table 3.3.2.1
Raw Score and Scale Score ranges of 2017 AIMS Assessments

Content	Grade	Raw Score Range	Scale Score range
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 development activities that occurred for the Spring 2017 AIMS Science tests. 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 2017 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.

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 2017 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 2017 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.

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

The most recent AIMS Data Analysis workshop 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 race/ethnicity 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 the June 2014 Data Analysis workshop 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)

4.2.3 AIMS Item Selection

The Item Selection meeting for Spring 2017 AIMS Science was conducted by ADE staff in July 2014. The purpose of the Item Selection meeting was to select items to place on test forms that would produce valid and reliable scores using items from previous test administrations as well as items from the 2014 field test administration that had been designated as "item accepted" (IA). 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 that most reflect the expectation of skills represented within 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. Figure 4.1 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 2017 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 2016 and spring 2017. 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 2017 New Operational Items include all of the AZ items covering Strand 1 Concept 1 that were in the spring 2016 test. The set of columns labeled Spring 2017 New Operational Items show the items that were retained from the spring 2016 or prior administrations (highlighted in blue). These retained items were designated as anchor items. During item selection for spring 2017, 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 4.4 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 4.2 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.

Figure 4.1
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

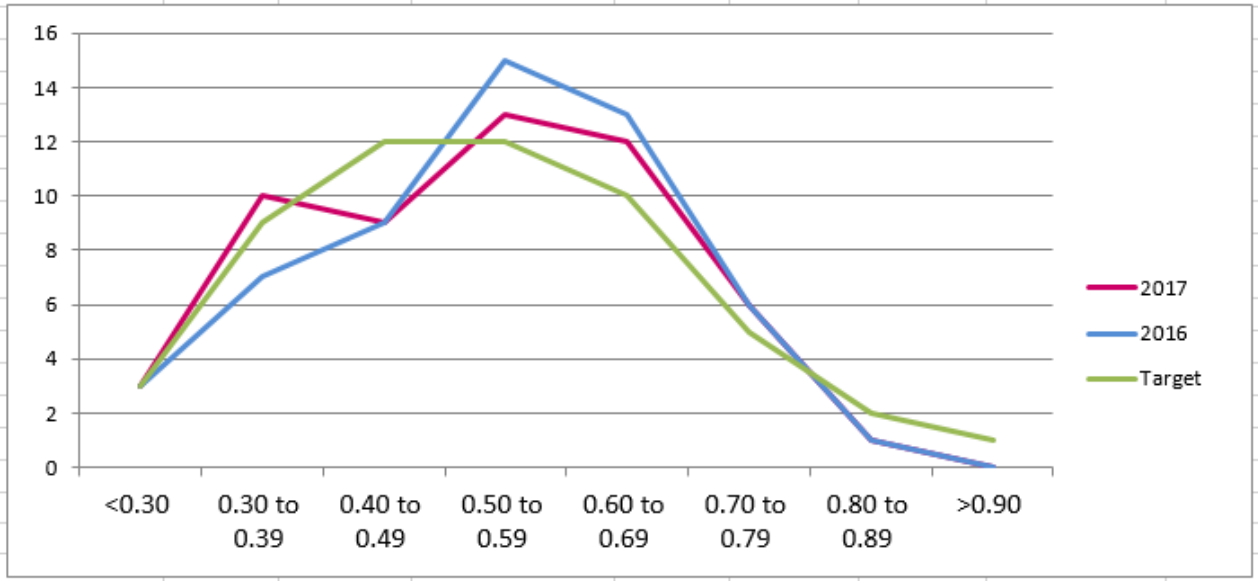
Row	N Count	Rasch	PVal	Flag PVal	PT Bis.	Flag PTBIS	Male vs Female Bias Flag	Non-Hispanic vs Hispanic Bias Flag	White vs Black Bias Flag	White vs Hispanic Bias Flag	White vs AmIn Bias Flag	White vs Asian Bias Flag	White vs Hawi/PaclsIr Bias Flag	White vs Multiracial Bias Flag	Dist A	Dist B	Dist C	Dist 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

Figure 4.2
Sample Grade 4 Science Item Replacement Table

AZ AIMS Grade 4 Spring 17 Operational Item Replacement Plan for Science																	
# of Items Required per Blueprint	Strand	Concept	Spring 16 - New Operational Items Selections								Spring 17 - 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	1.1.1	3514444	0	0.399	1.3943	0.373	1	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		1.1.1	3514504	0	0.519	0.816	0.354	1		1.1.1	3514504	0	0.519	0.816	0.354	1
	1	1		1.1.3	44114434	Electricity and Magnetism	0.674	0.1057	0.532	2		1.1.3	44114434	Electricity and Magnetism	0.674	0.1057	0.532	2
	1	1		1.1.2	44114447	Volcanoes	0.736	-0.2494	0.545	1		1.1.2	44114447	Volcanoes	0.736	-0.2494	0.545	1
	1	1		1.1.2	44114318	0	0.445	1.2724	0.335	2		1.1.1	44134442	0	0.585	0.4906	0.368	3

Figure 4.3
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	7	1	0	0	**18**
Actual - new selections		3	9	6	7	5	5	1	0	**36**
NRT/Dual Purpose		0	0	0	0	0	0	0	0	**0**
TOTAL	2017	3	10	9	13	12	6	1	0	**54**
Actual	2016	3	7	9	15	13	6	1	0	**54**



2017

DOK Level	Target N	Actual N
1	4	9
2	36	34
3	14	11
4	0	0
Total	54	54

DOK	2016
1	10
2	36
3	8
4	0
Total	54

Table 4.2
Number of Science Items Selected by Committee

Content Area	Grade	Total Items	Anchor Items		Total Selected	
Science	4	54	22	41%	32	59%
	8	58	24	41%	34	59%
	HS	65	20	31%	45	69%
Science Total		**177**	**66**	**37%**	**111**	**63%**

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 ADE 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 via this site 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 the Spring 2017 AIMS Science testing. 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 the Spring 2017 AIMS Science grades 4, 8, and high school tests. Accommodations are included if there is no 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 (EL) 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 are excluded from AIMS testing and encouraged to take the AIMS A Science assessment.

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 (EL) is a student whose primary home language is other than English, who scores below the proficient level on the AZELLA. Fluent English Proficient (FEP) is a term that is used to refer to a former EL student who has scored at the proficient level of the AZELLA.

For detailed information on testing accommodations, please see *AIMS Testing Accommodations: Guidelines* on the Arizona Department of Education website.

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 assessments, 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 Science are limited to those listed insection 5.1.3 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 ELs 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.

- Have answers transferred from a test book into an answer document
- Record or dictate multiple-choice responses to a scribe

EL/FEP

For students who were eligible to receive a standard accommodation due to their classification as an EL student or as a FEP (Year 1 or Year 2) student.

- More breaks and/or several shorter sessions
- Simplified language for the scripted directions in English
- Read aloud in English the science test items, as needed upon student request
- Provide a word-for-word published, paper translation dictionary
- 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.

- Place marker used

- More breaks and/or several shorter sessions
- Test at a different time of day
- Simplify language for the scripted directions in English
- Read aloud or sign the directions that students read on their own
- Read aloud in English or sign the science test items
- Large print edition of test
- Have answers transferred from the test book into an answer document
- Record or dictate multiple-choice responses to a scribe
- 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 Science assessments will count as having tested for accountability purposes. Their AIMS results will be included in aggregate results at the school, district, and state level on reports provided by the testing contractor.

Students who receive standard testing accommodations while participating in AIMS Science assessments 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 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 2017 AIMS Test security agreement for Superintendents/Charter Representatives and District Test Coordinators

School Year 2016-2017
Assessment Test Security Agreement
For Superintendents/Charter Representatives and District Test Coordinators

As Superintendent/Charter Representative or District Test Coordinator, I acknowledge that all state Assessment Tests (AIMS Science, AIMS A Science, AzMERIT 3-8, AzMERIT EOC, MSAA and AZELLA Placement and AZELLA ReassessmentTests) are secure tests and I agree to the following conditions concerning the security of the state Assessment 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 Assessment Test Coordinator Information Sheet for School Year 2016-2017 must be on file with the Assessment Section of the Arizona Department of Education (ADE).
   b. The designated District Test Coordinator(s) must complete all pre-test trainings provided by ADE for each of the test administrations in which that the district will be participating.
2. All necessary security precautions shall be in place to safeguard test materials.
   a. Access to paper test books, answer documents, test booklets, paper based assessments, online tests, and all other secure ancillary documents is restricted.
   b. All persons having access to the secure test materials, other than students to whom the tests are administered, shall sign a School Year 2016-2017 State Assessment Test Security Agreement which will 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.
   c. A list of students who tested must be kept on file, with the names of who the test administrator(s) and test proctors(s) who were in the test room during the test administration.
   d. All secure test materials including secure ancillary test materials shall be kept under lock and key **except during actual test times when distributed to students.**
      i. Secure test materials shall be delivered to test administrators no sooner than the date of testing.
      ii. Students shall not be permitted to remove test materials including scratch paper from the testing room except under supervision of staff.
   e. All secure student documents shall not be examined, read, or reviewed by anyone other than the student unless in compliance with the appropriate Administration Directions.
      i. No secure test materials shall be used for instruction before or after test administration.
      ii. No content or items of the test shall be disclosed nor allowed to be discussed or disclosed.
      iii. No student response or notations (including stray marks) on a student test booklet or answer document can be changed or erased and will be submitted for scoring exactly as completed by student.
      iv. No reporting of any students' answer choices based on previous experience outside the test administration.
   f. Upon completion of testing, all test materials, including student data sheets and/or secure testing materials including the appropriate Manuals and Administration Directions shall be returned to the designated District Test Coordinator.
3. All Usernames and passwords used for state assessments are unique to individuals and shall not be shared.
4. The district superintendent or charter representative shall develop, distribute, and enforce disciplinary procedures for the violation of test security by staff.
5. *Test Preparation and Administration Practices,* the guidelines approved by the State Board of Education in January 2003 and updated December 2007, shall be followed.
6. All instructions in the Coordinator Manuals and Administration Directions for each state assessment, which include reading the directions to students exactly as scripted, 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 State Assessment tests for School Year 2016-2017, will also sign an Assessment Test Security Agreement.

District/Charter Name: _____ **District Entity #:**_____

Superintendent/Charter Representative:

Printed Name:_____ Signature:_____ Date:_____

Achievement District Test Coordinator:

Printed Name:_____ Signature:_____ Date:_____

Alternative Assessment District Test Coordinator:

Printed Name:_____ Signature:_____ Date:_____

AZELLA District Test Coordinator:

Printed Name:_____ Signature:_____ Date:_____

Email: Testing@azed.gov

Figure 5.2.2
Spring 2017 AIMS Test security agreement for all school/district/charter personnel

Achievement Tests
(AIMS Science and AzMERIT)
School Year 2016-2017 Test Security Agreement

I acknowledge that all Achievement Tests are secure tests and agree to the following conditions of use to ensure the security of the test. For this document Achievement Tests refers to AIMS Science, AzMERIT 3-8, and AzMERIT EOC.

1. I shall take necessary precautions to safeguard test materials.
   a. I shall sign an Achievement Test Security Agreement for School Year 2016-2017.
   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 Achievement 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. I shall not erase or change any student responses or any marks (including stray marks) on a scorable test booklet or answer document.
   g. If test books and answer documents are distributed to me, I shall return all AzMERIT 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 Achievement Tests for school year 2016-2017 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 appropriate Test Administration Directions prior to administering the test.
6. I shall follow all instructions in the appropriate 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 Achievement Tests, will also sign a Test Security Agreement.

Signed By: _____ Date: _____

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 spring 2017 administration. The manual included the following topics:

- Responsibilities of the Achievement 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
  - Test Coordinator's Materials Checklist
  - Inventorying Test Materials
  - Precautions
  - Inspecting and Organizing Test Materials
  - Assembling Scorable Test Materials
  - Assembling Nonscorable Test Materials
  - Materials Retrieval
- State Board of Education Rule
- Contact Information
- Material Packing Diagrams
- Checklist for Packing and Shipping Test Materials

Test Administration Directions were made available to all test administrators for the spring 2017 assessments. The *Test Administration Directions* included the following topics:

- Overview for the Spring 2017 Administration of AIMS
  - Test Administrator Responsibilities
  - Students to Be Tested
  - Test Administration Schedule
  - Test Materials
- Before Testing Guidelines
  - Training and Test Security
  - Preparing the Room for Testing

- During Testing Guidelines
    - Reading the Scripted Directions
    - Student Identification Information
    - Pre-ID Labels and Demographic Data Grid
    - Monitoring Testing
    - Precautions
    - 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 to the spring test administration. 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 the AIMS Science grade 4, 8, and high school tests computed from the data used for calibration and scaling. 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.

6.1 Data

Arizona had one test window for operational testing in spring 2017. The AIMS Science tests for grade 4 and 8, and high school were administered between March 13 and April 21.

6.2 Descriptive Statistics by Test

Table 6.2.1 presents descriptive statistics by grade level which are computed with the calibration samples. 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 Science tests. The item-to-total correlation is computed as a point biserial correlation. The point biserial correlation reported is the correlation of the item scores and the total test score.

Table 6.2.1

Spring 2017 AIMS Science Classical Test Analysis Statistics

Grade	N	Max RS Obtained	RS M	RS SD	P-value M	rpb M	Internal Consistency
4	87346	54	29.65	10.00	0.55	0.35	0.89
8	83431	58	33.93	10.91	0.58	0.37	0.91
HS	80405	65	31.36	12.17	0.48	0.35	0.91

6.3 Classical Item Analysis

Classical item analysis was conducted for each Science test. 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), 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 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 the correlation between the item and total score as if the item scores were normally distributed rather than binary.

Table 6.3.1
Spring 2017 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	87346	0.78	0.3	0.36	0.08	4.31	-0.133426	7.8	-0.184793	10.22	-0.100345	77.59	0.2570756
2	87346	0.7	0.34	0.39	0.12	70.26	0.2980103	13.81	-0.13332	7.16	-0.205976	8.65	-0.13142
3	87346	0.76	0.41	0.51	0.18	8.34	-0.230948	3.97	-0.193658	11.25	-0.172444	76.26	0.3680652
4	87346	0.59	0.48	0.56	0.16	25.34	-0.281533	58.83	0.4437048	9.14	-0.173289	6.53	-0.185225
5	87346	0.67	0.41	0.47	0.17	9.79	-0.254123	6.58	-0.170579	66.89	0.3656845	16.57	-0.145219
6	87346	0.49	0.35	0.38	0.14	20.8	-0.048207	49.37	0.3036504	22.97	-0.156097	6.72	-0.265559
7	87346	0.63	0.4	0.46	0.14	13.61	-0.212633	10.11	-0.069476	62.77	0.3602972	13.38	-0.235412
8	87346	0.63	0.44	0.51	0.15	8.85	-0.226227	8.91	-0.221256	62.89	0.3954192	19.21	-0.161268
9	87346	0.44	0.24	0.24	0.14	28.15	0.0244545	8.44	-0.186083	44.31	0.188932	18.97	-0.135358
10	87346	0.42	0.38	0.43	0.13	16.71	-0.162294	42	0.3366245	11.23	-0.09859	29.93	-0.162499
11	87346	0.6	0.45	0.52	0.05	60.3	0.4081114	13.23	-0.110038	7.08	-0.26301	19.34	-0.24021
12	87346	0.39	0.38	0.42	0.11	12.89	-0.13997	38.89	0.3329579	4.15	-0.118488	43.97	-0.184906
13	87346	0.43	0.37	0.41	0.18	21.22	-0.106813	24.92	-0.140455	10.56	-0.185969	43.12	0.326432
14	87346	0.46	0.42	0.48	0.17	26.91	-0.163881	19.19	-0.154993	7.68	-0.206543	46.06	0.3787615
15	87346	0.62	0.43	0.49	0.17	61.64	0.3845351	8.51	-0.262581	10.7	-0.17904	18.99	-0.148276
16	87346	0.75	0.52	0.67	0.19	8.03	-0.274574	6.49	-0.261534	10.5	-0.238161	74.79	0.489713
17	87346	0.33	0.3	0.33	0.2	26.23	-0.136305	20.05	-0.045023	20.75	-0.103857	32.78	0.2558888
18	87346	0.63	0.38	0.43	0.19	8.63	-0.228437	62.61	0.3335834	9.54	-0.239034	19.03	-0.068282
19	87346	0.53	0.49	0.56	0.21	20.59	-0.301093	12.12	-0.212592	52.67	0.4494546	14.42	-0.094163
20	87346	0.74	0.46	0.57	0.27	11.76	-0.179224	6.31	-0.256518	8.12	-0.237386	73.55	0.4207659
21	87346	0.68	0.5	0.61	0.23	11.91	-0.251166	68.2	0.4665867	14.89	-0.245099	4.77	-0.226113
22	87346	0.74	0.54	0.68	0.21	74.04	0.5037557	9.33	-0.232033	5.37	-0.255798	11.04	-0.303213
23	87346	0.55	0.39	0.43	0.23	10.21	-0.164675	15.64	-0.118424	18.95	-0.199706	54.98	0.3443812
24	87346	0.42	0.33	0.37	1.75	17	-0.130876	21.44	-0.075604	42.2	0.2947408	17.61	-0.169903
25	87346	0.36	0.2	0.2	0.15	17.63	-0.19325	8.82	-0.131695	37.36	0.074148	36.04	0.156544
26	87346	0.6	0.37	0.42	0.27	59.88	0.329285	10.26	-0.206357	10.4	-0.232162	19.19	-0.07014
27	87346	0.32	0.18	0.17	0.49	25.4	-0.059173	27.63	-0.002381	31.9	0.1305735	14.58	-0.096348
28	87346	0.61	0.36	0.4	0.06	2.98	-0.202547	5.41	-0.227008	30.49	-0.143172	61.06	0.311266
29	87346	0.73	0.54	0.68	0.1	9.35	-0.286828	11.62	-0.245225	6.42	-0.25741	72.51	0.5049952
30	87346	0.48	0.34	0.37	0.21	7.93	-0.242443	5.24	-0.202815	38.73	-0.073453	47.91	0.2933729
31	87346	0.44	0.35	0.39	0.21	33.38	-0.06985	43.71	0.3080192	9.52	-0.244597	13.19	-0.141819
32	87346	0.38	0.41	0.47	0.2	17.23	-0.21476	33.27	-0.116936	38.29	0.365752	11.01	-0.132705
33	87346	0.44	0.47	0.54	0.24	10.17	-0.148637	43.61	0.4298132	40.16	-0.233677	5.81	-0.228904
34	87346	0.64	0.49	0.58	0.18	64.49	0.4539285	10.69	-0.253033	17.92	-0.203145	6.73	-0.242802
35	87346	0.61	0.47	0.55	0.24	8.18	-0.288859	8.05	-0.237255	22.29	-0.157019	61.23	0.429849
36	87346	0.61	0.26	0.27	0.26	17.29	-0.011974	61.19	0.2144325	5.41	-0.167075	15.85	-0.16972
37	87346	0.37	0.28	0.3	0.19	36.84	0.2334352	38.44	-0.088474	6.3	-0.20584	18.23	-0.050583
38	87346	0.47	0.37	0.41	0.26	24.92	-0.223905	12.27	0.0035419	47.31	0.3285471	15.25	-0.189784
39	87346	0.42	0.37	0.41	0.08	42.24	0.3245746	19.13	-0.13966	14.87	-0.160783	23.69	-0.113269
40	87346	0.75	0.33	0.39	0.14	7.98	-0.088903	11.42	-0.185213	75.17	0.2863585	5.29	-0.180718

Note. 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 2017 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	87346	0.32	0.36	0.42	0.29	16.7	-0.194941	21.66	-0.116524	31.99	0.3218613	29.37	-0.064504
42	87346	0.64	0.4	0.46	0.19	8.74	-0.233873	9.57	-0.186028	17.99	-0.137354	63.51	0.3610121
43	87346	0.55	0.41	0.46	0.21	9.88	-0.270535	6.98	-0.18541	28.02	-0.117377	54.9	0.3635729
44	87346	0.33	0.27	0.29	0.23	53.53	-0.069496	32.55	0.2237432	9.22	-0.139073	4.48	-0.144725
45	87346	0.81	0.48	0.64	0.18	6.41	-0.247985	81.11	0.4437371	6.97	-0.252854	5.33	-0.213246
46	87346	0.47	0.31	0.33	0.27	14.12	-0.170008	13.85	-0.217226	46.58	0.2617781	25.18	0.0086826
47	87346	0.24	0.23	0.26	0.27	23.88	0.1857596	12.92	-0.055188	13.04	-0.181996	49.89	0.0012461
48	87346	0.61	0.38	0.42	0.28	7.95	-0.239774	61.05	0.330895	10.53	-0.257802	20.2	-0.042499
49	87346	0.54	0.48	0.55	0.45	20.79	-0.127008	54.31	0.4377307	13.63	-0.266746	10.81	-0.239801
50	87346	0.54	0.44	0.49	0.53	19.25	-0.106999	53.68	0.3917977	15.16	-0.233172	11.37	-0.217458
51	87346	0.53	0.44	0.5	0.52	20.65	-0.107656	15.26	-0.220123	11.02	-0.244791	52.54	0.4002949
52	87346	0.55	0.48	0.55	0.64	20.58	-0.196954	10.28	-0.226076	55.21	0.4408348	13.28	-0.206763
53	87346	0.65	0.35	0.39	0.82	7.98	-0.131859	64.93	0.3067244	10.21	-0.183764	16.06	-0.146948
54	87346	0.68	0.47	0.56	1.06	68.44	0.4320278	8.22	-0.226953	8.12	-0.226926	14.17	-0.212982

Note. 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 2017 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	83431	0.84	0.26	0.35	0.02	9.42	-0.139117	84.18	0.2297134	2.87	-0.137809	3.51	-0.10946
2	83431	0.73	0.37	0.45	0.06	73.35	0.33714	6.43	-0.147763	12.07	-0.23427	8.09	-0.133415
3	83431	0.72	0.34	0.41	0.06	6.3	-0.223575	9.88	-0.123915	71.51	0.3074862	12.24	-0.144554
4	83431	0.81	0.44	0.6	0.04	5.26	-0.192247	80.92	0.4118921	5.32	-0.210797	8.46	-0.256868
5	83431	0.5	0.54	0.64	0.04	50.26	0.5107765	24.64	-0.231207	22.11	-0.314113	2.95	-0.149928
6	83431	0.82	0.39	0.53	0.04	7.18	-0.220188	2.14	-0.185036	82.35	0.361232	8.28	-0.19568
7	83431	0.81	0.39	0.51	0.04	4.64	-0.230324	5.09	-0.174298	9.3	-0.181034	80.93	0.3549859
8	83431	0.88	0.4	0.61	0.04	2.49	-0.155522	6.49	-0.26275	87.99	0.3774921	2.99	-0.197128
9	83431	0.6	0.28	0.3	0.07	18.48	-0.04263	11.6	-0.228526	60.14	0.2377211	9.71	-0.089861
10	83431	0.6	0.39	0.44	0.05	17.19	-0.106558	10.13	-0.192612	60.44	0.3459568	12.2	-0.216321
11	83431	0.78	0.33	0.42	0.02	12.46	-0.143551	77.86	0.2978755	5.42	-0.191283	4.24	-0.163354
12	83431	0.7	0.42	0.51	0.04	16.51	-0.155862	4.48	-0.218058	8.93	-0.254753	70.04	0.3835795
13	83431	0.43	0.29	0.31	0.08	8.22	-0.16952	31.89	-0.08924	43.13	0.2432936	16.67	-0.086655
14	83431	0.48	0.38	0.43	0.06	28.97	-0.096323	13.22	-0.166806	9.39	-0.246354	48.36	0.3443604
15	83431	0.29	0.14	0.13	0.09	22.26	-0.040261	20.83	-0.050885	27.47	-0.016211	29.35	0.0980504
16	83431	0.63	0.43	0.5	0.04	11.42	-0.187647	9.88	-0.202487	63.38	0.3927332	15.29	-0.191891
17	83431	0.59	0.45	0.52	0.08	18.8	-0.101976	6.85	-0.24683	14.84	-0.277527	59.43	0.4092446
18	83431	0.68	0.47	0.57	0.06	67.79	0.4378274	10.71	-0.257096	9.43	-0.215911	12	-0.190304
19	83431	0.73	0.45	0.57	0.08	8.99	-0.188391	7.98	-0.260611	72.68	0.42107	10.26	-0.2073
20	83431	0.31	0.41	0.5	0.14	30.95	0.3787254	18.15	-0.112669	26.63	-0.236904	24.13	-0.062931
21	83431	0.58	0.31	0.33	0.07	11.69	-0.241735	15.18	-0.128901	57.93	0.264689	15.13	-0.018671
22	83431	0.68	0.37	0.44	0.09	20.34	-0.143621	67.82	0.3367365	6.94	-0.233115	4.82	-0.187355
23	83431	0.51	0.48	0.55	0.09	17.17	-0.206402	51.24	0.4384602	26.59	-0.20457	4.91	-0.235349
24	83431	0.38	0.44	0.52	0.28	4.46	-0.155012	24.57	-0.061658	32.28	-0.298022	38.41	0.4069691
25	83431	0.43	0.49	0.58	0.04	23.49	-0.233007	15.64	-0.261211	43.26	0.4590772	17.58	-0.088736
26	83431	0.43	0.37	0.41	0.06	29.49	-0.062531	42.67	0.3258379	11.92	-0.205577	15.85	-0.180708
27	83431	0.21	0.2	0.22	0.12	20.78	0.1589434	12.92	-0.182697	17.2	-0.072167	48.97	0.0480982
28	83431	0.46	0.38	0.43	0.08	21.5	-0.118898	45.56	0.3408275	17.22	-0.200015	15.65	-0.124807
29	83431	0.47	0.45	0.52	0.1	13.78	-0.178559	46.94	0.4160858	17.3	-0.209679	21.88	-0.161408
30	83431	0.7	0.53	0.66	0.04	10.55	-0.29416	6.79	-0.267947	12.35	-0.215641	70.27	0.5006162
31	83431	0.51	0.38	0.43	0.09	18.88	-0.268568	50.82	0.3420207	19.48	-0.0753	10.73	-0.116301
32	83431	0.81	0.48	0.66	0.07	4.54	-0.240742	81.26	0.456107	10.13	-0.278107	4	-0.223014
33	83431	0.72	0.48	0.6	0.1	5.33	-0.201022	9.42	-0.232513	72.06	0.4519466	13.09	-0.26526
34	83431	0.45	0.4	0.45	0.26	44.95	0.3584737	8.87	-0.176674	8.4	-0.185725	37.51	-0.157901
35	83431	0.49	0.37	0.41	0.1	10.97	-0.25918	48.51	0.3267357	21.55	-0.062176	18.87	-0.144811
36	83431	0.64	0.48	0.57	0.06	3.47	-0.253965	5.24	-0.256415	64.01	0.4478359	27.22	-0.249908
37	83431	0.48	0.33	0.36	0.09	25.44	-0.035675	17.89	-0.215898	48.09	0.2903648	8.49	-0.16771
38	83431	0.59	0.4	0.46	0.12	9.62	-0.262168	6.97	-0.27407	23.97	-0.069214	59.32	0.359878
39	83431	0.49	0.4	0.45	0.13	38.1	-0.159412	48.87	0.3619999	5.82	-0.269792	7.08	-0.157059
40	83431	0.63	0.5	0.6	0.15	11.03	-0.226032	21.6	-0.274314	63.08	0.4683117	4.14	-0.211048

Note. 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 2017 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	83431	0.73	0.49	0.61	0.03	4.58	-0.257423	14.14	-0.243325	72.55	0.4569564	8.7	-0.231566
42	83431	0.79	0.5	0.66	0.06	5.75	-0.26406	8.07	-0.267568	78.51	0.4657112	7.61	-0.213982
43	83431	0.51	0.45	0.52	0.12	13.8	-0.270311	10.46	-0.274968	51.48	0.4121139	24.15	-0.066554
44	83431	0.8	0.54	0.73	0.08	6.7	-0.272527	7.78	-0.279989	5.8	-0.269544	79.64	0.5127133
45	83431	0.45	0.28	0.3	0.1	41.13	0.0008315	6.71	-0.235314	44.8	0.2369492	7.26	-0.228615
46	83431	0.52	0.44	0.5	0.14	7.55	-0.240701	52.24	0.3976819	7.56	-0.25705	32.51	-0.142982
47	83431	0.66	0.43	0.51	0.1	6.88	-0.230206	21.2	-0.167277	5.59	-0.25272	66.23	0.3908706
48	83431	0.59	0.52	0.61	0.16	16.56	-0.276714	8.34	-0.192836	15.6	-0.226571	59.34	0.4858185
49	83431	0.6	0.39	0.45	0.13	19.85	-0.052014	11.31	-0.290316	60.32	0.3514357	8.39	-0.213083
50	83431	0.57	0.48	0.56	0.18	10.85	-0.245565	57.17	0.4421514	17.96	-0.220535	13.84	-0.166603
51	83431	0.3	0.32	0.37	0.16	43.42	-0.096107	30.46	0.2808618	13.48	-0.078291	12.49	-0.166021
52	83431	0.57	0.42	0.48	0.19	9.11	-0.214793	10.11	-0.309101	57.07	0.3786477	23.51	-0.075968
53	83431	0.53	0.39	0.44	0.18	9.02	-0.274318	25.29	-0.055083	53.25	0.3483526	12.26	-0.216889
54	83431	0.57	0.48	0.55	0.26	9.89	-0.270978	13.63	-0.250409	18.91	-0.127888	57.31	0.4391692
55	83431	0.34	0.32	0.36	0.09	26.99	-0.018315	15.84	-0.145223	23.09	-0.167488	33.99	0.2781453
56	83431	0.78	0.52	0.69	0.1	78.42	0.4926587	8.66	-0.286775	7.2	-0.253822	5.62	-0.243337
57	83431	0.47	0.22	0.22	0.14	11.87	-0.170876	46.75	0.1784733	31.45	0.0576237	9.79	-0.203534
58	83431	0.52	0.38	0.42	0.14	52.19	0.3382861	9.38	-0.240368	28.52	-0.040748	9.76	-0.270801

Note. 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 2017 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	80405	0.55	0.36	0.4	0.15	26.12	-0.1252	9.5	-0.196769	9.06	-0.160266	55.17	0.3193539
2	80405	0.57	0.51	0.6	0.04	5.91	-0.061855	27.93	-0.391582	9.07	-0.1627	57.06	0.4788506
3	80405	0.38	0.39	0.45	0.24	33.26	-0.173574	18.2	-0.104147	10.39	-0.159903	37.9	0.3520946
4	80405	0.41	0.28	0.31	0.14	14.15	-0.118646	29.72	-0.038342	41.04	0.2438515	14.95	-0.171172
5	80405	0.37	0.22	0.23	0.13	19.64	-0.15968	32.98	0.0494266	10.18	-0.1526	37.07	0.178834
6	80405	0.43	0.37	0.42	0.11	42.65	0.3331648	9	-0.059898	17.09	-0.205423	31.14	-0.151733
7	80405	0.28	0.41	0.51	0.13	19.65	0.0135128	31.04	-0.211574	27.62	0.3820322	21.56	-0.190385
8	80405	0.57	0.38	0.43	0.1	9.33	-0.091646	13.82	-0.16203	20.2	-0.217942	56.55	0.3432993
9	80405	0.65	0.37	0.43	0.11	18.98	-0.133702	6.18	-0.195529	65.07	0.3318952	9.66	-0.198274
10	80405	0.57	0.43	0.5	0.16	21.86	-0.185993	4.88	-0.15663	56.72	0.3959804	16.37	-0.230792
11	80405	0.56	0.37	0.42	0.05	14.93	-0.152447	55.9	0.3300461	8.56	-0.213171	20.57	-0.123355
12	80405	0.29	0.29	0.34	0.13	13.06	-0.111386	23.08	0.0495415	34.25	-0.21383	29.48	0.2590874
13	80405	0.73	0.43	0.54	0.06	13.39	-0.216369	73.05	0.4010005	3.56	-0.183001	9.95	-0.234571
14	80405	0.56	0.22	0.23	0.08	8.25	-0.190704	29.57	0.0196898	6.49	-0.188158	55.61	0.18087
15	80405	0.75	0.31	0.38	0.06	74.6	0.2778409	13.54	-0.12111	9.53	-0.189824	2.27	-0.158756
16	80405	0.62	0.39	0.45	0.07	62.43	0.3523507	5.37	-0.170712	25.66	-0.18602	6.48	-0.20658
17	80405	0.32	0.3	0.34	0.11	29.92	-0.107498	32.04	0.2627746	21.82	-0.05598	16.11	-0.136713
18	80405	0.49	0.41	0.47	0.06	6.01	-0.170858	49.43	0.372512	16.8	-0.275694	27.69	-0.095017
19	80405	0.39	0.36	0.41	0.09	12.07	-0.20762	16.49	-0.208922	32.01	-0.030459	39.35	0.3262732
20	80405	0.49	0.51	0.6	0.1	15.04	-0.21715	21.32	-0.256208	14.46	-0.162362	49.08	0.479564
21	80405	0.7	0.31	0.36	0.09	12.12	-0.213468	5.93	-0.103663	11.97	-0.094527	69.9	0.2723876
22	80405	0.49	0.32	0.35	0.14	18.23	-0.059358	25.22	-0.194096	49.39	0.2785036	7.03	-0.125063
23	80405	0.67	0.47	0.57	0.13	14.96	-0.235957	7.35	-0.248886	66.57	0.4351833	10.99	-0.178832
24	80405	0.56	0.39	0.45	0.18	12.41	-0.151803	10.98	-0.228465	20.46	-0.136916	55.97	0.3563066
25	80405	0.43	0.42	0.49	0.06	18.96	-0.164083	17.3	-0.190201	43.29	0.3891112	20.39	-0.140232
26	80405	0.39	0.41	0.48	0.1	35.09	-0.11823	39.03	0.3780223	19.4	-0.220438	6.38	-0.166913
27	80405	0.48	0.42	0.48	0.11	5.16	-0.146125	31.75	-0.284008	47.69	0.382117	15.28	-0.072961
28	80405	0.27	0.34	0.41	0.12	26.88	0.0094884	26.15	-0.154915	20.25	-0.17451	26.61	0.3032365
29	80405	0.32	0.29	0.33	0.14	32.29	0.2491908	14.48	-0.190767	15.69	-0.285691	37.4	0.1127437
30	80405	0.53	0.39	0.44	0.12	18.49	-0.136566	11.82	-0.14024	16.29	-0.207214	53.29	0.35056
31	80405	0.45	0.4	0.46	0.1	14.39	-0.182362	45.45	0.364096	28.24	-0.134768	11.83	-0.175153
32	80405	0.51	0.4	0.45	0.13	31.43	-0.143302	11.15	-0.219082	50.8	0.3613096	6.5	-0.182794
33	80405	0.45	0.43	0.5	0.18	44.76	0.3982663	23.4	-0.177677	18.62	-0.192791	13.05	-0.141514
34	80405	0.42	0.32	0.35	0.14	33.28	-0.043592	10.38	-0.204749	14.32	-0.158531	41.89	0.2808466
35	80405	0.32	0.24	0.26	0.18	30.99	0.0011422	32.33	0.198673	23.55	-0.058342	12.95	-0.204545
36	80405	0.75	0.45	0.58	0.1	13.94	-0.286062	75.41	0.4230658	5.29	-0.203714	5.26	-0.166797
37	80405	0.62	0.45	0.53	0.17	10.89	-0.189578	12.68	-0.226903	61.6	0.4136942	14.65	-0.187708
38	80405	0.49	0.44	0.51	0.13	10.72	-0.229519	12.73	-0.188313	27.08	-0.158807	49.34	0.4089222
39	80405	0.43	0.38	0.44	0.08	17.04	-0.17202	43.18	0.3464922	16.69	-0.168451	23.01	-0.104791
40	80405	0.56	0.44	0.51	0.19	16.03	-0.184226	14.97	-0.240052	55.77	0.4044427	13.03	-0.140838

Note. 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 2017 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	80405	0.33	0.31	0.35	0.14	33.45	-0.061811	22.18	-0.184112	33.35	0.2694673	10.88	-0.068659
42	80405	0.38	0.34	0.39	0.17	27.15	-0.01377	13.93	-0.229384	38.27	0.306442	20.48	-0.156958
43	80405	0.44	0.45	0.52	0.14	17.98	-0.135598	22.66	-0.169923	14.89	-0.228673	44.33	0.4120298
44	80405	0.42	0.35	0.39	0.11	13.35	-0.150233	32.91	-0.047014	11.65	-0.252296	41.97	0.3124333
45	80405	0.31	0.35	0.41	0.17	31.07	0.0845607	11.77	-0.218986	26.32	-0.259684	30.67	0.3161984
46	80405	0.54	0.43	0.49	0.13	22.42	-0.161678	18.16	-0.216682	53.52	0.3939476	5.77	-0.194654
47	80405	0.48	0.51	0.6	0.14	48.31	0.4780186	9.49	-0.171543	20.7	-0.265608	21.37	-0.197329
48	80405	0.41	0.33	0.37	0.19	14.28	-0.119868	22	-0.06769	41.48	0.2916059	22.05	-0.177607
49	80405	0.29	0.26	0.29	0.17	15.29	-0.215194	20.96	-0.144256	28.99	0.2193535	34.59	0.0770527
50	80405	0.59	0.49	0.57	0.18	59.47	0.452363	9	-0.227886	20.24	-0.271534	11.11	-0.151176
51	80405	0.35	0.29	0.33	0.22	35.03	0.2536769	12.15	-0.253283	7.43	-0.251622	45.17	0.0558573
52	80405	0.54	0.42	0.49	0.22	10.32	-0.219131	9.17	-0.211451	54.2	0.3863843	26.08	-0.147112
53	80405	0.27	0.32	0.38	0.1	43.37	-0.136462	19.32	-0.117048	27.34	0.2815657	9.87	-0.039111
54	80405	0.43	0.41	0.48	0.13	43.24	0.3780284	24.41	-0.175139	20.46	-0.162824	11.76	-0.143707
55	80405	0.56	0.56	0.67	0.16	55.99	0.5330008	10.15	-0.249187	26.65	-0.329221	7.05	-0.17012
56	80405	0.64	0.48	0.57	0.16	9.49	-0.191189	16.04	-0.248234	63.52	0.4445206	10.79	-0.214599
57	80405	0.46	0.36	0.4	0.24	46.4	0.3222483	26.77	-0.060088	17.75	-0.179469	8.85	-0.230341
58	80405	0.39	0.44	0.51	0.23	39.16	0.4039442	12.97	-0.221826	20.38	-0.150319	27.26	-0.139246
59	80405	0.44	0.38	0.43	0.24	44.05	0.3418816	15.7	-0.213183	12.09	-0.258603	27.91	-0.017217
60	80405	0.55	0.53	0.63	0.3	8.75	-0.242646	20.14	-0.274136	15.94	-0.189644	54.87	0.4989854
61	80405	0.69	0.44	0.53	0.25	11.4	-0.223625	69.06	0.4080622	13.86	-0.221907	5.43	-0.178189
62	80405	0.46	0.35	0.39	0.23	46.16	0.3142674	26.02	-0.115909	10.42	-0.205767	17.16	-0.113596
63	80405	0.5	0.43	0.49	0.25	6.55	-0.231122	28.06	-0.132348	15.26	-0.219884	49.88	0.3918379
64	80405	0.49	0.42	0.48	0.25	21.91	-0.145604	49.1	0.3805857	18.26	-0.201251	10.48	-0.170087
65	80405	0.55	0.43	0.5	0.23	15.85	-0.159497	13.1	-0.206378	54.9	0.3952627	15.92	-0.187382

Note. 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 2017 AIMS Science assessments. Each grade level was calibrated and scaled with calibration samples that typically consisted close to 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.

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.

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 2019 or 2020 (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 Science tests. Each grade-level test was calibrated separately. All calibration activities were replicated by ADE staff as an added quality control check.

7.2.1 Calibration Model

The AIMS Science assessments are composed of multiple-choice items. Historically, the AIMS 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.

7.2.2 Calibration Software

Parameter estimation for items on the science tests in grade 4, 8, and high school was implemented using WINSTEPS 3.90.0 (Linacre, 2015). WINSTEPS uses joint maximum likelihood estimation (JMLE) as described by Wright and Masters (1982).

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. These tables contain each item's Rasch difficulty, standard error of the difficulty (SE), weighted mean-square (MNSQ infit), and unweighted mean-square (MNSQ outfit).

All items for all AIMS tests converged during calibration using typical procedures for WINSTEPS software. Typically in IRT, Rasch difficulty values range from -3.00 to +3.00 with positive values indicating that the item is relatively difficult and negative values indicating that it is relatively easy. Standard error of estimates for the Rasch difficulty measures indicated that the parameters were well estimated. Model-to-item data fit was monitored using MNSQ infit and MNSQ outfit statistics, which indicate the degree of accuracy and predictability with which the data fits the model (Linacre, 2002). The MNSQ infit is sensitive to unexpected responses at or near the item's calibrated level; whereas, MNSQ outfit is sensitive to unexpected responses away from the item's calibrated level. Typically for MNSQ infit, values less than 0.6 and greater than 1.4 indicate misfit, where values greater than 1.4 indicate misfit for MNSQ outfit (Wright & Linacre, 1994). No item was flagged as having misfit as indicated by either MNSQ infit or MNSQ outfit.

Table 7.3.1.1
Spring 2017 AIMS IRT Item Statistics
Science Grade 4

Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit
1	-0.52	0.01	1.03	1.08	28	0.42	0.01	1.03	1.03
2	-0.20	0.01	1.07	1.11	29	-0.21	0.01	0.84	0.73
3	-0.43	0.01	0.94	0.88	30	1.16	0.01	1.08	1.10
4	0.51	0.01	0.92	0.89	31	1.24	0.01	1.05	1.07
5	0.10	0.01	0.97	0.95	32	1.51	0.01	0.98	1.01
6	0.89	0.01	1.06	1.08	33	1.28	0.01	0.94	0.93
7	0.52	0.01	0.96	0.96	34	0.38	0.01	0.88	0.83
8	0.27	0.01	0.96	0.94	35	0.34	0.01	0.94	0.92
9	1.21	0.01	1.16	1.21	36	0.25	0.01	1.15	1.21
10	1.32	0.01	1.02	1.03	37	1.58	0.01	1.12	1.16
11	0.44	0.01	0.95	0.93	38	1.07	0.01	1.03	1.06
12	1.48	0.01	1.01	1.05	39	1.39	0.01	1.03	1.08
13	1.18	0.01	1.03	1.05	40	-0.36	0.01	1.01	1.06
14	1.13	0.01	0.98	0.99	41	1.84	0.01	1.00	1.09
15	0.37	0.01	0.96	0.97	42	0.15	0.01	1.01	1.02
16	-0.34	0.01	0.85	0.72	43	0.67	0.01	1.00	0.98
17	1.80	0.01	1.08	1.15	44	1.81	0.01	1.10	1.21
18	0.42	0.01	0.99	1.03	45	-0.83	0.01	0.90	0.74
19	0.81	0.01	0.92	0.89	46	1.10	0.01	1.09	1.12
20	-0.27	0.01	0.90	0.85	47	2.32	0.01	1.10	1.32
21	0.03	0.01	0.88	0.83	48	0.40	0.01	1.01	1.04
22	-0.30	0.01	0.84	0.72	49	0.73	0.01	0.93	0.90
23	0.82	0.01	1.01	1.01	50	0.69	0.01	0.97	0.95
24	1.13	0.01	1.06	1.07	51	0.68	0.01	0.97	0.95
25	1.62	0.01	1.18	1.27	52	0.69	0.01	0.92	0.89
26	0.46	0.01	1.02	1.03	53	0.48	0.01	1.00	0.98
27	1.84	0.01	1.19	1.33	54	0.13	0.01	0.89	0.84

Table 7.3.1.2
Spring 2017 AIMS IRT Item Statistics
Science Grade 8

Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Item	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit
1	-1.13	0.01	1.04	1.20	30	-0.09	0.01	0.83	0.76
2	-0.37	0.01	1.00	1.07	31	0.91	0.01	1.05	1.07
3	-0.26	0.01	1.03	1.11	32	-1.06	0.01	0.95	0.79
4	-0.87	0.01	0.90	0.80	33	-0.30	0.01	0.90	0.83
5	0.84	0.01	0.88	0.84	34	1.10	0.01	1.03	1.05
6	-0.98	0.01	0.94	0.87	35	0.92	0.01	1.07	1.08
7	-0.78	0.01	0.90	0.88	36	0.15	0.01	0.93	0.87
8	-1.29	0.01	0.79	0.64	37	0.94	0.01	1.11	1.16
9	0.21	0.01	1.17	1.32	38	0.37	0.01	1.02	1.08
10	0.33	0.01	1.03	1.06	39	0.90	0.01	1.02	1.03
11	-0.66	0.01	1.01	1.08	40	0.20	0.01	0.91	0.87
12	-0.18	0.01	0.97	0.95	41	-0.32	0.01	0.89	0.83
13	1.19	0.01	1.15	1.21	42	-0.70	0.01	0.87	0.72
14	1.05	0.01	1.06	1.07	43	0.67	0.01	0.98	0.98
15	1.93	0.01	1.25	1.53	44	-0.97	0.01	0.92	0.70
16	0.04	0.01	1.02	1.02	45	1.02	0.01	1.15	1.20
17	0.23	0.01	1.00	0.99	46	0.74	0.01	0.99	0.98
18	-0.06	0.01	0.93	0.91	47	0.03	0.01	0.98	0.96
19	-0.38	0.01	0.94	0.90	48	0.37	0.01	0.90	0.87
20	1.99	0.01	1.03	1.09	49	0.53	0.01	1.02	1.03
21	0.46	0.01	1.12	1.18	50	0.49	0.01	0.94	0.90
22	-0.06	0.01	1.02	1.06	51	1.86	0.01	1.06	1.23
23	0.71	0.01	0.95	0.95	52	0.46	0.01	1.01	1.01
24	1.64	0.01	1.01	1.07	53	0.69	0.01	1.04	1.06
25	1.18	0.01	0.92	0.92	54	0.56	0.01	0.94	0.90
26	1.13	0.01	1.05	1.09	55	1.66	0.01	1.09	1.17
27	2.48	0.01	1.13	1.59	56	-0.70	0.01	0.84	0.68
28	1.17	0.01	1.06	1.10	57	1.01	0.01	1.22	1.28
29	0.98	0.01	0.97	0.97	58	0.74	0.01	1.05	1.06

Table 7.3.1.3
Spring 2017 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	1.02	1.07	34	0.86	0.01	1.08	1.09
2	0.20	0.01	0.87	0.85	35	1.35	0.01	1.14	1.22
3	1.06	0.01	1.01	1.02	36	-0.73	0.01	0.85	0.75
4	0.90	0.01	1.11	1.14	37	-0.08	0.01	0.93	0.89
5	1.10	0.01	1.17	1.23	38	0.50	0.01	0.95	0.95
6	0.83	0.01	1.03	1.04	39	0.80	0.01	1.01	1.02
7	1.62	0.01	0.96	1.00	40	0.18	0.01	0.95	0.92
8	0.16	0.01	1.00	1.00	41	1.30	0.01	1.08	1.13
9	-0.25	0.01	0.98	1.05	42	1.04	0.01	1.05	1.08
10	0.15	0.01	0.96	0.93	43	0.74	0.01	0.95	0.95
11	0.19	0.01	1.02	1.02	44	0.86	0.01	1.05	1.07
12	1.51	0.01	1.10	1.14	45	1.44	0.01	1.04	1.07
13	-0.74	0.01	0.92	0.89	46	0.31	0.01	0.96	0.95
14	0.21	0.01	1.15	1.21	47	0.61	0.01	0.90	0.87
15	-0.77	0.01	0.98	1.13	48	0.88	0.01	1.07	1.09
16	-0.12	0.01	0.98	0.99	49	1.54	0.01	1.12	1.22
17	1.37	0.01	1.09	1.14	50	-0.24	0.01	0.96	0.92
18	0.50	0.01	0.99	0.98	51	1.21	0.01	1.10	1.15
19	0.99	0.01	1.04	1.05	52	0.27	0.01	0.97	0.96
20	0.52	0.01	0.89	0.86	53	1.64	0.01	1.06	1.13
21	-0.51	0.01	1.01	1.10	54	0.80	0.01	0.99	0.98
22	0.45	0.01	1.07	1.13	55	0.30	0.01	0.83	0.79
23	-0.49	0.01	0.95	0.90	56	-0.18	0.01	0.90	0.86
24	0.23	0.01	0.99	0.98	57	0.78	0.01	1.05	1.07
25	0.79	0.01	0.98	0.97	58	1.09	0.01	0.98	1.00
26	1.00	0.01	0.99	0.99	59	0.76	0.01	1.02	1.03
27	0.58	0.01	0.98	0.97	60	0.24	0.01	0.86	0.84
28	1.68	0.01	1.02	1.12	61	-0.43	0.01	0.91	0.84
29	1.36	0.01	1.09	1.16	62	0.62	0.01	1.04	1.04
30	0.32	0.01	1.00	1.03	63	0.48	0.01	0.97	0.97
31	0.69	0.01	1.00	1.01	64	0.46	0.01	0.98	0.98
32	0.48	0.01	1.00	0.99	65	0.24	0.01	0.96	0.95
33	0.86	0.01	0.99	0.98					

7.4 Scaling Methods

7.4.1 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 https://cms.azed.gov/home/GetDocumentFile?id=58506dfeaadebe050c5745c8. 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.

7.5 Equating

7.5.1 Science

The 2017 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 2015 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. Table 7.5.1.1 presents the number of operational and anchor items for each grade. Table 7.5.1.2 shows the content representation for the 2017 anchor items compared to the 2017 operational form. Table 7.5.1.3 presents descriptive statistics (IRT and classical) for the 2017 anchor item difficulties and the 2017 operational form.

Table 7.5.1.1
Spring 2017 AIMS Science Anchor Items

Grade	Operational	Anchor
4	54	22
8	58	24
HS	65	20

Table 7.5.1.2
Representation of Content by 2017 Science Anchor Sets

		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	
4th																				
All	N	6	6	5	1	3	3	4	2	4	0	0	2		0	0	6	6	6	54
	Pct	11.11	11.11	9.26	1.85	5.56	5.56	7.41	3.7	7.41	0	0	3.7		0	0	11.11	11.11	11.11	100
Anchor	N	3	2	2		1	2	2	1	2					0	0	3	2	2	22
	Pct	13.64	9.09	9.09		4.55	9.09	9.09	4.55	9.09					0	0	13.64	9.09	9.09	100
8th																				
All	N	6	4	6	4	4	2	1	5	0	2	0	6		10	8				58
	Pct	10.34	6.9	10.34	6.9	6.9	3.45	1.72	8.62	0	3.45	0	10.34		17.24	13.79				100
Anchor	N	2	2	2	1	1	2	0	3	0	0	0	3		4	4				24
	Pct	8.33	8.33	8.33	4.17	4.17	8.33	0	12.5	0	0	0	12.5		16.67	16.67				100
HS																				
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	3	2	1	1	1	1	3		2	2	1	2	1						20
	Pct	15	10	5	5	5	5	15		10	10	5	10	5						100

Table 7.5.1.3
Representation of Difficulty by 2017 Science Anchor Sets

Grade	Statistic	IRT Difficulty Entire 2017 Test	IRT Difficulty All Anchors Items	P-Value Entire 2017 Test	P-Value All Anchors Items
4	N	54	22	54	22
	Mean	0.6900	0.5600	0.5500	0.5800
	Std Dev	0.7100	0.5200	0.1400	0.1000
	Min	-0.8300	-0.8300	0.2400	0.4200
	Max	2.3200	1.3900	0.8100	0.8100
8	N	58	24	58	24
	Mean	0.3900	0.3900	0.5800	0.5800
	Std Dev	0.8600	0.8400	0.1600	0.1500
	Min	-1.2900	-1.2900	0.2100	0.3100
	Max	2.4800	1.9900	0.8800	0.8800
HS	N	65	20	65	20
	Mean	0.5700	0.2200	0.4800	0.5500
	Std Dev	0.6100	0.5100	0.1200	0.1000
	Min	-0.7700	-0.7400	0.2700	0.3900
	Max	1.6800	1.0900	0.7500	0.7500

A fixed-parameter equating was implemented within WINSTEPS in order to link the 2017 science tests to the operational reporting scale. This is implemented by constraining the 2017 parameter estimates for the common anchor items to equal the final parameter estimates obtained in the most recent 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 0.3 or less than -0.3 was dropped from the anchor item set. If more than one item was observed with a displacement statistic greater than 0.3 or less than -0.3, then only the first item with the largest absolute displacement value was dropped from the anchor set. The displacement values of the remaining anchor items were reestimated 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 0.3 or less than -0.3 was repeated until all displacements were acceptable. All items with displacement values greater than 0.3 or less than -0.3 were noted to be carried over for removal from the anchor set for next year.

2. 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.

3. 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 only one item from the anchor sets for grade 4, 8, and high school. This one item was within the grade 8 test leaving 3 items aligned to the same strand and concept as anchors, 38% of the operational items so aligned.

7.5.2 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). SEM is calculated as follows

$$SEM = \sigma_x (1 - r)^{1/2},$$

where σ_x is a standard deviation of total score X, and r is a reliability coefficient such as Cronbach's alpha (Crocker & Algina, 1986). SEM and Cronbach's alpha for the total group as well as subgroups are presented in Table 9.1.1.1.

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 7.5.2.1 through 7.5.2.3 present raw score to scale score conversion tables and IRT conditional standard errors of measurement for 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.

Table 7.5.2.1
Spring 2017 AIMS Raw Score to Scale Score Table
Science Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	70	28	506	14
1	296	50	29	510	14
2	331	36	30	514	14
3	353	29	31	518	14
4	368	26	32	522	14
5	381	23	33	527	14
6	391	22	34	531	15
7	400	20	35	535	15
8	408	19	36	540	15
9	415	18	37	544	15
10	422	18	**38**	**549**	**15**
11	428	17	39	554	16
12	434	17	40	559	16
13	440	16	41	564	16
14	445	16	42	570	17
15	450	16	43	576	17
16	455	15	44	582	18
17	460	15	45	589	19
18	**464**	**15**	46	596	19
19	469	15	47	604	20
20	473	15	48	613	22
21	477	14	49	624	24
22	481	14	50	636	26
23	486	14	51	651	30
24	490	14	52	673	36
25	494	14	53	708	50
26	498	14	54	800	70
27	**502**	**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. The complete list of AIMS Science score cuts and ranges is presented in Table 10.1.1.

Table 7.5.2.2
Spring 2017 AIMS Raw Score to Scale Score Table
Science Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	69	30	495	14
1	280	49	31	499	14
2	315	35	**32**	**503**	**14**
3	336	29	33	506	14
4	352	25	34	510	14
5	364	23	35	514	14
6	374	21	36	518	14
7	383	20	37	523	14
8	391	19	38	527	14
9	398	18	39	531	14
10	405	18	**40**	**535**	**15**
11	411	17	41	540	15
12	417	16	42	544	15
13	422	16	43	549	15
14	428	16	44	554	16
15	433	15	45	560	16
16	437	15	46	565	16
17	442	15	47	571	17
18	446	15	48	577	18
19	451	14	49	584	18
20	455	14	50	591	19
21	459	14	51	599	20
22	463	14	52	608	21
23	467	14	53	618	23
24	471	14	54	631	26
25	**475**	**14**	55	646	29
26	479	14	56	667	35
27	483	14	57	702	49
28	487	14	58	800	69
29	491	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. The complete list of AIMS Science score cuts and ranges is presented in Table 10.1.1.

Table 7.5.2.3
Spring 2017 AIMS Raw Score to Scale Score Table
Science High School

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	200	74	33	495	13
1	267	53	34	499	13
2	305	38	**35**	**502**	**13**
3	327	31	36	506	14
4	343	27	37	509	14
5	356	25	38	513	14
6	367	23	39	516	14
7	376	21	40	520	14
8	384	20	41	524	14
9	391	19	42	527	14
10	398	18	43	531	14
11	404	18	44	535	14
12	410	17	**45**	**539**	**14**
13	416	17	46	543	15
14	421	16	47	547	15
15	426	16	48	552	15
16	430	16	49	556	15
17	435	15	50	561	16
18	439	15	51	566	16
19	444	15	52	571	17
20	448	15	53	576	17
21	452	14	54	582	18
22	456	14	55	588	18
23	459	14	56	595	19
24	463	14	57	602	20
25	467	14	58	611	21
26	471	14	59	620	23
27	474	14	60	630	25
28	**478**	**14**	61	643	27
29	481	14	62	659	31
30	485	14	63	681	38
31	488	13	64	719	53
32	492	13	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. The complete list of AIMS Science score cuts and ranges is presented in Table 10.1.1.

PART 8: TEST RESULTS

8.1 Data

Part 8 of this technical report contains information about the results of Spring 2017AIMS Science in grades 4, 8, and 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.

Results presented below are based on population data contained within the final electronic data files and gone through the same clean-up process as calibration data in Part 7. 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.

Test results are presented in Tables 8.1.1.2 and 8.1.1.3. For each grade, scale score means and standard deviations as well as the percentages of students in each performance level are reported 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 and 8.1.1.3, scale score frequency distributions are displayed in Tables 8.1.1.4 through 8.1.1.7. The information for each grade is contained within 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. Cohort 20 is defined as the group of students that expect to graduate in 2020 and typically includes grade 9 students. Cohort 19 is defined as the group of students that expect to graduate in 2019 and typically includes 10th grade students.

Table 8.1.1.2
Spring 2017 AIMS State Test Results
Science Grades 4 and 8

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Grade 4							
Total	87350	514.96	46.76	13	27	35	25
Hispanic	40149	500.46	41.20	18	34	34	14
Non-Hispanic	46751	527.57	47.64	9	20	36	35
Race							
White	68325	517.31	46.70	12	26	36	27
Black or African American	5367	498.90	42.42	20	33	33	14
Asian	2661	540.04	48.23	6	15	33	46
American Indian or Alaskan Native	5632	490.01	37.04	23	40	30	7
Native Hawaiian or Other Pacific Islander	449	507.21	41.79	15	31	36	18
Multiple Indication	3842	522.63	46.15	10	23	37	30
Special Program Membership							
English Learner Program	8426	469.51	29.57	43	43	13	1
Special Education	11121	485.69	41.83	32	36	22	10
Low SES	45423	500.74	41.48	18	34	34	14
Grade 8							
Total	83398	514.00	50.85	23	19	24	34
Hispanic	36851	498.27	43.81	31	24	25	21
Non-Hispanic	45942	526.83	52.51	16	15	24	45
Race							
White	66285	516.39	50.24	21	19	25	36
Black or African American	5056	496.98	45.86	33	22	25	20
Asian	2530	550.53	61.15	9	11	18	62
American Indian or Alaskan Native	5442	487.12	40.25	40	26	21	13
Native Hawaiian or Other Pacific Islander	361	508.83	47.69	21	22	29	28
Multiple Indication	2569	518.62	49.46	19	18	25	38
Special Program Membership							
English Learner Program	3519	453.04	27.97	79	15	5	1
Special Education	8584	469.76	39.34	61	20	11	7
Low SES	40370	498.39	44.17	31	23	24	22

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, and students attending state hospital schools 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 8.1.1.3
Spring 2017 AIMS State Test Results
Science High School

		Scale Score		% at Performance Level			
	N	M	SD	FFBS	AS	MS	ES
Cohort 19							
Total	44191	483.30	46.99	50	18	18	13
Hispanic	20975	471.58	40.40	61	18	14	7
Non-Hispanic	22769	494.32	49.98	40	19	21	19
Race							
White	34928	485.33	47.15	48	19	19	14
Black or African American	2850	469.79	40.94	62	17	13	7
Asian	1141	514.81	58.57	27	18	23	32
American Indian or Alaskan Native	2729	463.56	35.53	70	16	10	4
Native Hawaiian or Other Pacific Islander	191	478.03	42.85	54	20	16	9
Multiple Indication	1363	486.58	45.33	46	20	21	14
Special Program Membership							
English Learner Program	1740	438.79	23.28	94	4	2	0
Special Education	4799	452.11	34.81	82	9	6	3
Low SES	18585	468.16	37.81	64	18	13	6
Cohort 20							
Total	36104	498.93	49.68	36	19	23	22
Hispanic	14017	481.97	42.10	50	21	19	11
Non-Hispanic	21616	510.06	51.07	28	18	26	29
Race							
White	28640	500.62	49.24	35	19	24	22
Black or African American	1979	479.77	41.17	52	19	19	10
Asian	1404	536.06	55.19	13	14	24	48
American Indian or Alaskan Native	1236	472.69	41.29	60	17	15	8
Native Hawaiian or Other Pacific Islander	149	489.52	46.93	43	19	22	15
Multiple Indication	1278	498.61	48.59	36	19	24	20
Special Program Membership							
English Learner Program	427	442.55	24.88	91	6	2	0
Special Education	2244	457.16	38.92	78	10	7	5
Low SES	14202	480.27	40.81	51	21	19	10

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, and students attending state hospital schools 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 8.1.1.4
Spring 2017 AIMS Frequency Distribution
Science CRT Grade 4

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	2	0	2	0	28	506	2881	3.3	40541	46.41
1	296	0	0	2	0	29	510	2778	3.18	43319	49.59
2	331	0	0	2	0	30	514	2871	3.29	46190	52.88
3	353	0	0	2	0	31	518	2859	3.27	49049	56.15
4	368	6	0.01	8	0.01	32	522	2746	3.14	51795	59.3
5	381	11	0.01	19	0.02	33	527	2749	3.15	54544	62.44
6	391	27	0.03	46	0.05	34	531	2793	3.2	57337	65.64
7	400	71	0.08	117	0.13	35	535	2773	3.17	60110	68.82
8	408	145	0.17	262	0.3	36	540	2670	3.06	62780	71.87
9	415	243	0.28	505	0.58	37	544	2576	2.95	65356	74.82
10	422	459	0.53	964	1.1	38	**549**	**2544**	2.91	**67900**	**77.73**
11	428	716	0.82	1680	1.92	39	554	2485	2.84	70385	80.58
12	434	901	1.03	2581	2.95	40	559	2412	2.76	72797	83.34
13	440	1310	1.5	3891	4.45	41	564	2215	2.54	75012	85.88
14	445	1522	1.74	5413	6.2	42	570	2170	2.48	77182	88.36
15	450	1872	2.14	7285	8.34	43	576	2002	2.29	79184	90.65
16	455	2072	2.37	9357	10.71	44	582	1810	2.07	80994	92.72
17	460	2223	2.54	11580	13.26	45	589	1632	1.87	82626	94.59
18	**464**	**2318**	**2.65**	**13898**	**15.91**	46	596	1304	1.49	83930	96.08
19	469	2498	2.86	16396	18.77	47	604	1106	1.27	85036	97.35
20	473	2607	2.98	19003	21.76	48	613	821	0.94	85857	98.29
21	477	2519	2.88	21522	24.64	49	624	634	0.73	86491	99.02
22	481	2590	2.97	24112	27.6	50	636	417	0.48	86908	99.49
23	486	2717	3.11	26829	30.71	51	651	258	0.3	87166	99.79
24	490	2759	3.16	29588	33.87	52	673	123	0.14	87289	99.93
25	494	2648	3.03	32236	36.9	53	708	47	0.05	87336	99.98
26	498	2652	3.04	34888	39.94	54	800	14	0.02	87350	100
27	**502**	**2772**	**3.17**	**37660**	**43.11**						

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 8.1.1.5
Spring 2017 AIMS Frequency Distribution
Science CRT Grade 8

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	0	0	0	0	30	495	2357	2.83	32444	38.9
1	280	0	0	0	0	31	499	2377	2.85	34821	41.75
2	315	0	0	0	0	32	503	**2456**	**2.94**	**37277**	**44.7**
3	336	0	0	0	0	33	506	2502	3	39779	47.7
4	352	0	0	0	0	34	510	2538	3.04	42317	50.74
5	364	1	0	1	0	35	514	2523	3.03	44840	53.77
6	374	8	0.01	9	0.01	36	518	2499	3	47339	56.76
7	383	15	0.02	24	0.03	37	523	2592	3.11	49931	59.87
8	391	38	0.05	62	0.07	38	527	2444	2.93	52375	62.8
9	398	101	0.12	163	0.2	39	531	2541	3.05	54916	65.85
10	405	184	0.22	347	0.42	**40**	**535**	**2609**	**3.13**	**57525**	**68.98**
11	411	290	0.35	637	0.76	41	540	2587	3.1	60112	72.08
12	417	481	0.58	1118	1.34	42	544	2416	2.9	62528	74.98
13	422	621	0.74	1739	2.09	43	549	2349	2.82	64877	77.79
14	428	803	0.96	2542	3.05	44	554	2346	2.81	67223	80.61
15	433	1011	1.21	3553	4.26	45	560	2152	2.58	69375	83.19
16	437	1235	1.48	4788	5.74	46	565	2104	2.52	71479	85.71
17	442	1401	1.68	6189	7.42	47	571	1898	2.28	73377	87.98
18	446	1571	1.88	7760	9.3	48	577	1787	2.14	75164	90.13
19	451	1592	1.91	9352	11.21	49	584	1704	2.04	76868	92.17
20	455	1702	2.04	11054	13.25	50	591	1470	1.76	78338	93.93
21	459	1901	2.28	12955	15.53	51	599	1289	1.55	79627	95.48
22	463	1911	2.29	14866	17.83	52	608	1095	1.31	80722	96.79
23	467	2059	2.47	16925	20.29	53	618	895	1.07	81617	97.86
24	471	2012	2.41	18937	22.71	54	631	669	0.8	82286	98.67
25	**475**	**2110**	**2.53**	**21047**	**25.24**	55	646	536	0.64	82822	99.31
26	479	2238	2.68	23285	27.92	56	667	354	0.42	83176	99.73
27	483	2258	2.71	25543	30.63	57	702	158	0.19	83334	99.92
28	487	2279	2.73	27822	33.36	58	800	64	0.08	83398	100
29	491	2265	2.72	30087	36.08						

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 8.1.1.6
Spring 2017 AIMS Frequency Distribution
Science CRT High School Cohort 19

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	0	0	0	0	33	495	1057	2.39	29368	66.46
1	267	0	0	0	0	34	499	989	2.24	30357	68.69
2	305	1	0	1	0	**35**	**502**	**990**	**2.24**	**31347**	**70.94**
3	327	1	0	2	0	36	506	980	2.22	32327	73.15
4	343	1	0	3	0.01	37	509	868	1.96	33195	75.12
5	356	7	0.02	10	0.02	38	513	786	1.78	33981	76.9
6	367	9	0.02	19	0.04	39	516	789	1.79	34770	78.68
7	376	13	0.03	32	0.07	40	520	757	1.71	35527	80.39
8	384	52	0.12	84	0.19	41	524	716	1.62	36243	82.01
9	391	102	0.23	186	0.42	42	527	731	1.65	36974	83.67
10	398	173	0.39	359	0.81	43	531	674	1.53	37648	85.19
11	404	286	0.65	645	1.46	44	535	663	1.5	38311	86.69
12	410	436	0.99	1081	2.45	**45**	**539**	**580**	**1.31**	**38891**	**88.01**
13	416	651	1.47	1732	3.92	46	543	582	1.32	39473	89.32
14	421	864	1.96	2596	5.87	47	547	549	1.24	40022	90.57
15	426	1088	2.46	3684	8.34	48	552	482	1.09	40504	91.66
16	430	1371	3.1	5055	11.44	49	556	455	1.03	40959	92.69
17	435	1542	3.49	6597	14.93	50	561	409	0.93	41368	93.61
18	439	1619	3.66	8216	18.59	51	566	438	0.99	41806	94.6
19	444	1671	3.78	9887	22.37	52	571	357	0.81	42163	95.41
20	448	1710	3.87	11597	26.24	53	576	345	0.78	42508	96.19
21	452	1648	3.73	13245	29.97	54	582	278	0.63	42786	96.82
22	456	1614	3.65	14859	33.62	55	588	255	0.58	43041	97.4
23	459	1603	3.63	16462	37.25	56	595	223	0.5	43264	97.9
24	463	1511	3.42	17973	40.67	57	602	203	0.46	43467	98.36
25	467	1542	3.49	19515	44.16	58	611	207	0.47	43674	98.83
26	471	1383	3.13	20898	47.29	59	620	172	0.39	43846	99.22
27	474	1352	3.06	22250	50.35	60	630	125	0.28	43971	99.5
28	**478**	**1317**	**2.98**	**23567**	**53.33**	61	643	96	0.22	44067	99.72
29	481	1274	2.88	24841	56.21	62	659	71	0.16	44138	99.88
30	485	1182	2.67	26023	58.89	63	681	38	0.09	44176	99.97
31	488	1161	2.63	27184	61.51	64	719	11	0.02	44187	99.99
32	492	1127	2.55	28311	64.07	65	800	4	0.01	44191	100

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 8.1.1.7
Spring 2017 AIMS Frequency Distribution
Science CRT High School Cohort 20

Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %	Raw Score	Scale Score	Freq	%	Cum. Freq.	Cum. %
0	200	1	0	1	0	33	495	1026	2.84	18987	52.59
1	267	0	0	1	0	34	499	968	2.68	19955	55.27
2	305	0	0	1	0	**35**	**502**	**896**	**2.48**	**20851**	**57.75**
3	327	1	0	2	0.01	36	506	888	2.46	21739	60.21
4	343	1	0	3	0.01	37	509	899	2.49	22638	62.7
5	356	1	0	4	0.01	38	513	840	2.33	23478	65.03
6	367	5	0.01	9	0.02	39	516	854	2.37	24332	67.39
7	376	8	0.02	17	0.05	40	520	803	2.22	25135	69.62
8	384	16	0.04	33	0.09	41	524	791	2.19	25926	71.81
9	391	28	0.08	61	0.17	42	527	783	2.17	26709	73.98
10	398	72	0.2	133	0.37	43	531	800	2.22	27509	76.19
11	404	117	0.32	250	0.69	44	535	770	2.13	28279	78.33
12	410	219	0.61	469	1.3	**45**	**539**	**740**	**2.05**	**29019**	**80.38**
13	416	292	0.81	761	2.11	46	543	668	1.85	29687	82.23
14	421	440	1.22	1201	3.33	47	547	695	1.92	30382	84.15
15	426	530	1.47	1731	4.79	48	552	652	1.81	31034	85.96
16	430	634	1.76	2365	6.55	49	556	628	1.74	31662	87.7
17	435	775	2.15	3140	8.7	50	561	548	1.52	32210	89.21
18	439	866	2.4	4006	11.1	51	566	576	1.6	32786	90.81
19	444	949	2.63	4955	13.72	52	571	511	1.42	33297	92.23
20	448	986	2.73	5941	16.46	53	576	454	1.26	33751	93.48
21	452	1033	2.86	6974	19.32	54	582	382	1.06	34133	94.54
22	456	1084	3	8058	22.32	55	588	349	0.97	34482	95.51
23	459	992	2.75	9050	25.07	56	595	347	0.96	34829	96.47
24	463	1068	2.96	10118	28.02	57	602	314	0.87	35143	97.34
25	467	1023	2.83	11141	30.86	58	611	262	0.73	35405	98.06
26	471	1005	2.78	12146	33.64	59	620	223	0.62	35628	98.68
27	474	975	2.7	13121	36.34	60	630	173	0.48	35801	99.16
28	**478**	**1049**	**2.91**	**14170**	**39.25**	61	643	144	0.4	35945	99.56
29	481	963	2.67	15133	41.92	62	659	76	0.21	36021	99.77
30	485	927	2.57	16060	44.48	63	681	56	0.16	36077	99.93
31	488	955	2.65	17015	47.13	64	719	22	0.06	36099	99.99
32	492	946	2.62	17961	49.75	65	800	5	0.01	36104	100

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 2017 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, mean scale score (M) and standard deviation (SD), as well as the scale score values at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile ranking (P10 – P90) and the percentage of students scoring within each performance level for the AIMS Science 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 these tables.

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
	2016	85917	514.4	47.6	451	478	515	550	578
	2017	87350	515.0	46.8	455	481	514	549	576
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
	2016	82475	512.6	48.8	449	477	512	546	578
	2017	83398	514.0	50.9	451	475	510	549	577
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
	2016(Cohort 18)	46427	482.2	44.2	432	449	476	508	542
	2016(Cohort 19)	33922	499.4	48.4	441	465	493	530	565
	2017(Cohort 19)	44191	483.3	47.0	430	448	474	509	547
	2017(Cohort 20)	36104	498.9	49.7	439	459	495	531	566

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), and attending state hospital schools (not in 2005) 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
	2016	85917	15	25	34	25
	2017	87350	13	27	35	25
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
	2016	82475	22	19	25	35
	2017	83398	23	19	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
	2016(Cohort 18)	46427	48	20	19	12
	2016(Cohort 19)	33922	33	20	25	22
	2017(Cohort 19)	44191	50	18	18	13
	2017(Cohort 20)	36104	36	19	23	22

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), and attending state hospital schools (not in 2005) 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 2017 AIMS Science assessments in grades 4, 8, and high school. All data presented in this section were computed using population test data available in the final electronic data files gone through the same clean-up process as the calibration data in Part 7. The following AERA/APA/NCME *Standards* from the 1999 edition are addressed: 1.5, 1.7, 2.1, 2.4, 2.10, 2.11, 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, p. 25) 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, p. 33) 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, defined 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 replications 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 2017 AIMS Science assessments is an estimate of its internal consistency.

9.1.1 Measures of Internal Consistency

For tests consisting of only constructed response or multiple-choice items, such as AIMS Science tests, 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.

Reliability estimates (Alpha) for the Spring 2017 AIMS Science assessments, for all students as well as for the various subgroups, are presented in Table 9.1.1.1. Note that a high degree of internal consistency is evident for all three tests.

Table 9.1.1.1
Spring 2017 AIMS Science Internal Consistency

Subgroup	N	Alpha	SEM
Grade 4			
All Students	87350	0.89	3.32
Hispanic	40149	0.87	3.29
Non-Hispanic	46751	0.90	3.14
White	68325	0.89	3.30
Black/African American	5367	0.87	3.36
Asian	2661	0.90	3.09
American Indian	5632	0.83	3.41
Hawaii/Pacific Islander	449	0.87	3.34
Multiple Indicators	3842	0.89	3.24
Female	42825	0.89	3.23
Male	44426	0.90	3.23
EL	8426	0.73	3.39
SPED	11121	0.87	3.32
Low SES	45423	0.87	3.31
Grade 8			
All Students	83398	0.89	3.32
Hispanic	36851	0.91	3.25
Non-Hispanic	45942	0.91	3.24
White	66285	0.89	3.43
Black/African American	5056	0.92	3.06
Asian	2530	0.87	3.39
American Indian	5442	0.89	3.41
Hawaii/Pacific Islander	361	0.90	3.32
Multiple Indicators	2569	0.90	3.30
Female	40974	0.92	3.21
Male	42243	0.72	3.44
EL	3519	0.86	3.40
SPED	8584	0.89	3.35
Low SES	40370	0.89	3.32
HS			
All Students	80295	0.91	3.65
Hispanic	34992	0.88	3.67
Non-Hispanic	44385	0.92	3.54
White	63568	0.91	3.64
Black/African American	4829	0.88	3.69
Asian	2545	0.93	3.44
American Indian	3965	0.86	3.62
Hawaii/Pacific Islander	340	0.90	3.60
Multiple Indicators	2641	0.91	3.55
Female	39853	0.90	3.71
Male	40241	0.92	3.56
EL	2167	0.60	3.57
SPED	7043	0.84	3.61
Low SES	32787	0.87	3.69

Presented in Tables 9.1.1.2 through 9.1.1.4 are number of items, mean and standard deviation (STD) of the raw scores, and the internal consistency reliability estimates (Alpha) at the science strand and concept level.

Table 9.1.1.2
Spring 2017 AIMS Strand/Concept Internal Consistency
Science Grade 4

Strand	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry	18	87350	10.16	3.78	0.75
Concept 1: Observations, Questions, and Hypotheses	6	87350	3.68	1.58	0.55
Concept 2: Scientific Testing (Investigating and Modeling)	6	87350	3.48	1.49	0.48
Concept 3/4: Analysis and Conclusions/ Communication	6	87350	3.00	1.55	0.49
2. History and Nature of Science	6	87350	3.97	1.66	0.63
Concept 1/2: History of Science as a Human Endeavor/ Nature of Scientific Knowledge	6	87350	3.97	1.66	0.63
3. Science in Personal and Social Perspectives	6	87350	3.42	1.60	0.52
Concept 1/2: Changes in Environments/Science and Technology in Society	6	87350	3.42	1.60	0.52
4. Life Science	6	87350	3.59	1.53	0.56
Concept 1/3/4: Characteristics of Organisms/Organisms and Environments/ Diversity, Adaptation, and Behavior	6	87350	3.59	1.53	0.56
5. Physical Science	6	87350	3.02	1.57	0.51
Concept 3: Energy and Magnetism	6	87350	3.02	1.57	0.51
6. Earth and Space Science	12	87350	5.48	2.35	0.54
Concept 2: Earth's Processes and Systems	6	87350	2.53	1.52	0.48
Concept 3: Changes in the Earth and Sky	6	87350	2.96	1.35	0.30

Table 9.1.1.3
Spring 2017 AIMS Strand/Concept Internal Consistency
Science Grade 8

Strand	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry	20	83398	12.53	3.93	0.77
Concept 1: Observations, Questions, and Hypotheses	6	83398	4.01	1.53	0.56
Concept 2: Scientific Testing (Investigating and Modeling)	4	83398	2.31	1.01	0.31
Concept 3: Analysis, Conclusions, and Refinements	6	83398	3.20	1.50	0.44
Concept 4: Communication	4	83398	3.02	1.06	0.47
2. History and Nature of Science	6	83398	3.23	1.65	0.57
Concept 1/2: History of Science as a Human Endeavor/ Nature of Scientific Knowledge	6	83398	3.23	1.65	0.57
3. Science in Personal and Social Perspectives	6	83398	3.53	1.63	0.56
Concept 1/2: Changes in Environments/Science and Technology in Society	6	83398	3.53	1.63	0.56
4. Life Science	8	83398	4.68	2.09	0.65
Concept 2/4: Reproduction and Heredity/Diversity, Adaptation, and Behavior	8	83398	4.68	2.09	0.65
5. Physical Science	18	83398	9.96	3.46	0.73
Concept 1: Properties and Changes of Properties in Matter	10	83398	5.27	2.06	0.56
Concept 2: Motion and Forces	8	83398	4.69	1.88	0.60

Table 9.1.1.4
Spring 2017 AIMS Strand/Concept Internal Consistency
Science High School

Strand	Number of Items	N	Raw Score Mean	Raw Score STD	Alpha
1. Scientific Inquiry	22	80295	11.40	4.85	0.81
Concept 1: Observations, Questions, and Hypotheses	6	80295	3.11	1.68	0.57
Concept 2: Scientific Testing (Investigating and Modeling)	6	80295	3.05	1.65	0.56
Concept 3: Analysis, Conclusions, and Refinements	6	80295	3.22	1.58	0.51
Concept 4: Communication	4	80295	2.01	1.13	0.42
2. History and Nature of Science	6	80295	3.27	1.50	0.46
Concept 1/2: History of Science as a Human Endeavor/Nature of Scientific Knowledge	6	80295	3.27	1.50	0.46
3. Science in Personal and Social Perspectives	7	80295	3.05	1.68	0.50
Concept 1/2/3: Changes in Environments/Science and Technology in Society/Human Population Characteristics	7	80295	3.05	1.68	0.50
4. Life Science	30	80295	13.65	5.76	0.82
Concept 1: The Cell	6	80295	2.37	1.50	0.46
Concept 2: Molecular Basis of Heredity	6	80295	3.04	1.63	0.55
Concept 3: Interdependence of Organisms	6	80295	2.88	1.65	0.57
Concept 4: Biological Evolution	6	80295	2.93	1.41	0.38
Concept 5: Matter, Energy, and Organization in Living Systems (Including Human Systems)	6	80295	2.42	1.53	0.48

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, 2014, p. 11). The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for 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 2017 AIMS Science 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 tests, 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 evidence of 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 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 2017 AIMS Science tests.

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 Cronbach's alpha as a measure for internal consistency.

Part 9 below presents the results of an analysis of Differential Item Functioning (DIF). Complete tables of gender, ethnicity, and race differential functioning of all operational items for the 2017 AIMS Science assessments are presented in Appendix B.

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

Also, note that further evidence in support of the AIMS assessment has been documented in previous AIMS annual 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 2017 AIMS Science tests were developed using procedures to minimize item and test bias and included reviews such as the Content and Sensitivity Reviews and Data Analysis Workshops after each item was field-tested as described in Part 4, Test Development. Expertise in this area is not, however, a substitute for statistical analyses of the items or the continued monitoring of the fairness of items. Thus, an empirical differential item functioning (DIF) approach was used to examine potential item bias on all operational items. DIF studies include systematic item analyses to determine if examinees, in identified groups, 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 the identifiable subgroups of the population are due to extraneous or construct irrelevant information which makes the items unfairly difficult, or easy, for one of the subgroups.

DIF analyses of the Spring 2017 AIMS Science tests were conducted for ethnic/race 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 test was used as the conditioning variable.

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 Mantel-Haenszel delta statistic (ΔMH DIF) was computed for all items. Educational Testing Service (ETS) first developed the ΔMH 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. ΔMH DIF is then computed as:

$$\Delta MH\ DIF = -2.35 \ln(\alpha_{MH})$$

Positive values of ΔMH DIF indicate items that favor the focal group, whereas negative values of ΔMH 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 2017 AIMS Science items that exhibit strong, weak, or no DIF (Zieky, 1993). Table 9.2.1.1 indicates the criteria for each category used for the 2017 AIMS Science 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-Haenszel chi-square not significantly different from zero
B	Weak DIF	Significant Mantel-Haenszel chi-square ($p<0.01$) and $1.0 \leq \lvert\Delta MH\rvert < 1.5$
C	Strong DIF	Significant Mantel-Haenszel chi-square ($p<0.01$) and $\lvert\Delta MH\rvert \geq 1.5$

Another measure, also used to analyze DIF for the Spring 2017 AIMS Science operational items, is the standardized mean difference (SMD; Zwick, Donoghue, & Grima, 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 raw score. Using the Zwick, Donoghue, Grima formulation, SMD is computed as:

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

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 2017 AIMS Science tests are presented in Appendix B. It is important to note that DIF analyses were also conducted on field test items prior to these items being eligible for operationalizarion during 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 Science testsexhibit no DIF or weak DIF. The one item that was flagged for exhibiting strong DIF is summarized in Table 9.2.1.2 with the results for all items used in 2017 presented in Appendix B.

Table 9.2.1.2
DIF Statistics for Items Exhibiting Strong DIF

Content	Grade	Item	Item Type	In favor of/ Against	Group	MH χ2	ΔMH	SMD
Science	8	34	MC	Against	Female	1854.92	-1.55	-0.14

Note: MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta DIF, SMD = Standardized Mean Difference,

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 2017 AIMS Science 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 in 2008. 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 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.

The scale score ranges for each of the four performance level categories and their associated cut scores, along with the lowest possible and highest possible scale scores for the AIMS Science tests are presented below in Table 10.1.1.

Table 10.1.1
Spring 2017 AIMS Science
Final Scale Score Ranges by Performance Level

Grade	LOSS	FFBS	AS Cut	AS	MS Cut	MS	ES Cut	ES	HOSS
4	200	200-460	462	464-498	500	502-544	547	549-800	800
8	200	364-471	473	475-499	500	503-531	532	535-800	800
HS	200	200-474	475	478-499	500	502-535	537	539-800	800

Note: LOSS=Lowest Observable Scale Score, FFBS=Fall Far Below the Standard, AS=Approaches the Standard, MS=Meets the Standard, ES=Exceeds the Standard, HOSS=Highest Observable Scale Score.

10.2 Classification Consistency and Accuracy

This section describes the analyses conducted to estimate classification consistency and accuracy for the Spring 2017 AIMS Science administration in grades 4, 8, and high school. 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 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 2017 AIMS Science assessments.

Classification consistency is most important for students whose ability is near each cut score. Students whose ability is far above or far below the established cut value 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 with their associated conditional standard error of measurement (CSEM) are presented in Table 10.2.2.1. The CSEMs around the Performance Level cuts were lower than those outside of the lowest and highest Performance Level cuts, indicating better measurement precision around the cuts.

Table 10.2.1
Spring 2017 AIMS
Standard Error of Measurement at Cut Scores

Test	Grade	AS		MS		ES	
		Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM
Science	4	462	15	500	14	547	15
Science	8	473	14	500	14	532	15
Science	HS	475	14	500	13	537	14

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

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 Science assessments. 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 $C_{h-1} \leq X \leq C_h$ 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:

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

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

Because obtaining test scores from two independent administrations of AIMS Science 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 $\boldsymbol{\theta}$, 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:

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

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,…, H), and a latent score , $\theta \in \Gamma_w (w$=1, 2,…, 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

Table 10.2.3.1 presents results from the classification consistency and classification accuracy analyses. These results are for classifying students into AIMS' four performance levels. Included in the table for each grade 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 2017 AIMS Science classification consistency and accuracy results are consistent with classification analyses from previous AIMS Science administrations. 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 Science assessments are quite consistent with other testing programs with similar structure and purpose.

Table 10.2.3.1
Spring 2017 AIMS
Classification Consistency and Accuracy

Test	Grade	N	Agreement	Inconsistency	Chance	Kappa	Accuracy
Science	4	87350	0.69	0.31	0.57	0.27	0.77
Science	8	83398	0.70	0.30	0.59	0.26	0.77
Science	HS	80295	0.72	0.28	0.60	0.30	0.79

Note: High school results include students in all cohorts. Results were computed with the IRT method suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000).

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. (2015). *Arizona's Instrument to Measure Standards 2015 technical report*. Iowa City, IA: Pearson.

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

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.

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. (2015). Winsteps® (Version 3.90.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2015. Available from http://www.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.

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

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

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

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: 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 B: DIF Results

Table B.1
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 4

	Reference: Male N= 44424 Focal: Female N= 42823				Reference: Hispanic N= 47198 Focal: Non Hispanic N= 40148				Reference: White N= 68322 Focal: Africa American N= 5367				Reference: White N= 68322 Focal: Native American N= 5632			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	26.05	0.20	0.01	A	39.22	0.26	0.02	A	12.36	-0.27	-0.02	A	13.69	0.29	0.02	A
2	284.63	0.62	0.05	A	1.30	0.04	0.01	A	19.15	-0.32	-0.03	A	0.00	0.00	0.00	A
3	28.37	0.22	0.01	A	37.31	-0.25	-0.02	A	0.06	0.02	0.00	A	36.88	-0.45	-0.04	A
4	9.11	-0.11	-0.01	A	75.53	-0.33	-0.03	A	11.25	0.25	0.02	A	48.49	-0.50	-0.04	A
5	4.97	-0.08	-0.01	A	78.06	-0.33	-0.03	A	4.51	0.16	0.01	A	3.96	-0.14	-0.01	A
6	9.82	0.11	0.01	A	52.32	-0.25	-0.03	A	1.97	0.10	0.01	A	4.65	-0.15	-0.01	A
7	307.55	0.63	0.05	A	8.14	0.11	0.01	A	0.93	-0.07	-0.01	A	4.80	-0.15	-0.01	A
8	71.42	0.31	0.03	A	26.25	-0.19	-0.02	A	1.80	0.10	0.01	A	10.54	-0.23	-0.02	A
9	0.47	-0.02	0.00	A	0.11	-0.01	0.00	A	34.97	-0.42	-0.04	A	81.62	0.61	0.06	A
10	51.95	0.25	0.02	A	14.67	-0.14	-0.01	A	0.58	-0.06	0.00	A	12.86	-0.27	-0.02	A
11	4.08	-0.07	-0.01	A	13.39	-0.14	-0.01	A	13.92	0.28	0.02	A	0.36	-0.04	0.00	A
12	69.24	0.29	0.03	A	41.10	-0.23	-0.02	A	5.00	-0.17	-0.01	A	5.58	-0.18	-0.01	A
13	237.73	0.53	0.05	A	2.12	0.05	0.00	A	0.84	0.07	0.01	A	0.01	-0.01	0.00	A
14	549.82	-0.82	-0.07	A	27.95	-0.19	-0.02	A	4.91	-0.16	-0.01	A	2.96	-0.13	-0.01	A
15	8.96	0.11	0.01	A	17.53	0.16	0.02	A	1.15	-0.08	-0.01	A	3.35	-0.13	-0.01	A
16	88.98	0.41	0.02	A	0.52	-0.03	0.00	A	0.00	0.00	0.00	A	1.21	0.09	0.01	A
17	14.67	0.14	0.01	A	0.47	0.03	0.00	A	4.09	0.15	0.01	A	0.07	0.02	0.00	A
18	0.00	0.00	0.00	A	12.09	-0.13	-0.01	A	13.32	-0.26	-0.02	A	6.18	-0.17	-0.02	A
19	167.03	-0.47	-0.04	A	171.45	-0.49	-0.04	A	0.21	-0.03	0.00	A	10.35	-0.23	-0.02	A
20	212.46	0.59	0.04	A	0.05	0.01	0.00	A	0.56	0.06	0.00	A	5.94	-0.18	-0.02	A
21	186.93	-0.54	-0.04	A	39.42	-0.25	-0.02	A	0.81	-0.07	-0.01	A	3.39	-0.13	-0.01	A
22	0.37	-0.03	0.00	A	84.97	-0.41	-0.03	A	4.63	0.18	0.01	A	15.39	-0.30	-0.02	A
23	125.55	0.39	0.04	A	5.73	0.09	0.01	A	1.70	0.09	0.01	A	0.01	0.01	0.00	A
24	0.04	-0.01	0.00	A	37.71	-0.22	-0.02	A	0.31	0.04	0.00	A	1.86	0.10	0.01	A
25	10.93	0.11	0.01	A	9.45	-0.11	-0.01	A	0.82	0.07	0.01	A	6.70	0.18	0.02	A
26	1.69	0.05	0.00	A	0.42	-0.02	0.00	A	19.67	-0.31	-0.03	A	5.22	0.16	0.02	A
27	0.07	0.01	0.00	A	0.10	0.01	0.00	A	0.72	-0.06	-0.01	A	5.51	0.17	0.02	A
28	14.64	0.13	0.01	A	0.02	0.01	0.00	A	2.08	-0.10	-0.01	A	0.27	-0.04	0.00	A
29	8.89	0.13	0.01	A	11.03	0.14	0.01	A	18.41	-0.35	-0.03	A	6.14	-0.19	-0.02	A
30	59.81	0.26	0.03	A	1.25	-0.04	0.00	A	3.82	-0.14	-0.01	A	16.40	-0.29	-0.03	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.1 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 4

	Reference: White N= 68322 Focal: Asian N= 2661				Reference: White N= 68322 Focal: Hawaii N= 448				Reference: White N= 68322 Focal: Multiple Indicator N= 3842			
Item	MH $\chi 2$	ΔMH	SMD	Flag	MH $\chi 2$	ΔMH	SMD	Flag	MH $\chi 2$	ΔMH	SMD	Flag
1	2.43	0.21	0.01	A	0.02	0.04	0.00	A	1.33	-0.11	-0.01	A
2	4.76	0.27	0.02	A	4.47	-0.51	-0.04	A	10.82	-0.30	-0.02	A
3	2.05	-0.19	-0.01	A	2.22	0.43	0.03	A	8.90	0.32	0.02	A
4	16.19	0.48	0.03	A	0.24	0.12	0.01	A	1.37	0.11	0.01	A
5	6.10	0.30	0.02	A	1.09	0.26	0.02	A	1.38	0.11	0.01	A
6	0.29	-0.05	0.00	A	4.87	-0.53	-0.05	A	0.37	0.05	0.00	A
7	9.41	0.35	0.03	A	8.27	0.73	0.06	A	0.03	0.01	0.00	A
8	1.11	-0.12	-0.01	A	6.74	-0.64	-0.06	A	0.04	0.02	0.00	A
9	0.21	-0.04	0.00	A	1.08	-0.24	-0.02	A	0.74	-0.07	-0.01	A
10	0.19	0.04	0.00	A	3.23	-0.44	-0.04	A	2.00	-0.12	-0.01	A
11	12.36	0.40	0.03	A	0.40	-0.15	-0.01	A	3.58	0.17	0.01	A
12	0.31	-0.06	0.00	A	0.19	0.11	0.01	A	4.18	-0.18	-0.02	A
13	1.42	0.12	0.01	A	2.40	-0.38	-0.03	A	5.03	0.19	0.02	A
14	4.63	-0.22	-0.02	A	0.24	-0.12	-0.01	A	0.05	-0.02	0.00	A
15	1.75	0.15	0.01	A	0.31	-0.14	-0.01	A	0.34	0.05	0.00	A
16	4.32	0.33	0.01	A	0.00	0.02	0.00	A	3.95	0.22	0.01	A
17	10.78	0.33	0.03	A	2.88	-0.44	-0.04	A	0.27	-0.04	0.00	A
18	0.86	-0.10	-0.01	A	0.02	0.04	0.00	A	0.02	-0.01	0.00	A
19	53.78	0.84	0.06	A	0.13	-0.09	-0.01	A	0.00	0.00	0.00	A
20	5.69	0.33	0.02	A	3.59	0.53	0.04	A	0.31	0.06	0.00	A
21	24.02	0.68	0.03	A	0.35	0.15	0.01	A	0.18	0.04	0.00	A
22	5.45	0.37	0.01	A	0.03	-0.05	0.00	A	0.59	0.09	0.00	A
23	0.00	0.00	0.00	A	0.31	0.13	0.01	A	0.74	-0.07	-0.01	A
24	7.07	0.27	0.03	A	0.32	-0.13	-0.01	A	1.06	0.09	0.01	A
25	0.02	0.01	0.00	A	3.10	-0.43	-0.04	A	0.00	0.00	0.00	A
26	20.77	0.51	0.04	A	0.32	0.13	0.01	A	0.15	-0.03	0.00	A
27	49.36	0.68	0.07	A	0.29	0.13	0.01	A	1.42	-0.10	-0.01	A
28	9.59	0.35	0.03	A	0.04	0.05	0.00	A	3.12	0.15	0.01	A
29	7.71	0.43	0.02	A	0.02	0.04	0.00	A	0.00	0.00	0.00	A
30	3.26	-0.18	-0.02	A	2.13	-0.34	-0.03	A	1.57	-0.10	-0.01	A

Note: African Am. = African American, Native Am. = Native American, MH $\chi 2$ = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.1 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 4

	Reference: Male N= 44424 Focal: Female N= 42823				Reference: Hispanic N= 47198 Focal: Non Hispanic N= 40148				Reference: White N= 68322 Focal: Africa American N= 5367				Reference: White N= 68322 Focal: Native American N= 5632			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	0.04	-0.01	0.00	A	29.69	-0.19	-0.02	A	0.07	-0.02	0.00	A	4.00	0.14	0.01	A
32	501.51	-0.81	-0.07	A	109.49	-0.39	-0.03	A	23.73	-0.37	-0.03	A	0.03	-0.01	0.00	A
33	1611.11	-1.47	-0.12	B<	88.59	-0.35	-0.03	A	14.22	0.28	0.02	A	7.24	-0.21	-0.02	A
34	26.05	0.19	0.01	A	7.10	-0.10	-0.01	A	22.14	0.36	0.03	A	18.83	-0.31	-0.03	A
35	17.50	-0.15	-0.01	A	16.85	-0.15	-0.01	A	2.38	-0.12	-0.01	A	1.65	0.09	0.01	A
36	35.76	0.20	0.02	A	0.37	-0.02	0.00	A	0.41	-0.04	0.00	A	1.77	0.09	0.01	A
37	27.70	-0.18	-0.02	A	19.01	0.16	0.01	A	0.07	0.02	0.00	A	3.44	-0.14	-0.01	A
38	25.29	-0.17	-0.02	A	10.92	-0.12	-0.01	A	2.92	0.12	0.01	A	0.00	0.00	0.00	A
39	35.64	0.21	0.02	A	1.24	0.04	0.01	A	3.22	-0.13	-0.01	A	54.72	-0.55	-0.05	A
40	374.51	-0.76	-0.05	A	0.38	-0.02	0.00	A	8.78	-0.23	-0.02	A	6.03	-0.18	-0.02	A
41	224.29	0.55	0.04	A	22.70	-0.18	-0.01	A	12.52	-0.29	-0.02	A	1.55	-0.10	-0.01	A
42	0.61	-0.03	0.00	A	5.62	-0.09	-0.01	A	4.10	-0.15	-0.01	A	2.51	-0.11	-0.01	A
43	177.86	-0.46	-0.04	A	26.42	-0.18	-0.01	A	32.01	-0.41	-0.04	A	0.06	-0.02	0.00	A
44	123.26	-0.39	-0.03	A	52.57	-0.27	-0.02	A	14.06	-0.29	-0.02	A	0.00	0.00	0.00	A
45	42.58	-0.31	-0.02	A	102.26	-0.49	-0.03	A	5.47	-0.21	-0.01	A	1.15	0.09	0.01	A
46	3.04	0.06	0.01	A	1.34	0.04	0.01	A	0.34	0.04	0.00	A	0.68	0.06	0.01	A
47	103.01	-0.39	-0.03	A	46.37	-0.28	-0.02	A	8.49	-0.25	-0.02	A	7.54	-0.24	-0.01	A
48	0.59	-0.03	0.00	A	0.13	-0.01	0.00	A	7.08	-0.19	-0.02	A	2.82	-0.12	-0.01	A
49	224.77	0.54	0.05	A	18.00	-0.16	-0.01	A	2.34	-0.11	-0.01	A	0.93	-0.07	-0.01	A
50	63.95	-0.28	-0.02	A	9.59	-0.11	-0.01	A	0.18	-0.03	0.00	A	0.65	-0.06	-0.01	A
51	3.99	0.07	0.01	A	0.78	-0.03	0.00	A	2.35	-0.11	-0.01	A	3.28	0.13	0.01	A
52	2.82	0.06	0.01	A	0.05	-0.01	0.00	A	5.30	-0.17	-0.01	A	2.03	0.10	0.01	A
53	137.12	0.41	0.04	A	5.74	-0.09	-0.01	A	30.19	-0.39	-0.04	A	21.53	-0.32	-0.03	A
54	93.04	0.37	0.03	A	87.75	0.37	0.03	A	1.02	0.08	0.01	A	0.07	-0.02	0.00	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.1 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 4

	Reference: White N= 68322 Focal: Asian N= 2661				Reference: White N= 68322 Focal: Hawaii N= 448				Reference: White N= 68322 Focal: Multiple Indicator N= 3842			
Item	MH $\chi 2$	ΔMH	SMD	Flag	MH $\chi 2$	ΔMH	SMD	Flag	MH $\chi 2$	ΔMH	SMD	Flag
31	3.35	-0.18	-0.02	A	0.00	-0.01	0.00	A	0.85	-0.08	-0.01	A
32	0.03	0.02	0.00	A	0.54	-0.19	-0.02	A	0.13	-0.03	0.00	A
33	20.50	0.49	0.04	A	1.55	0.32	0.03	A	10.08	0.28	0.02	A
34	1.73	0.16	0.01	A	1.99	-0.36	-0.03	A	2.03	0.14	0.01	A
35	2.01	0.17	0.01	A	2.40	-0.39	-0.03	A	5.19	0.21	0.02	A
36	16.42	-0.40	-0.04	A	0.75	-0.20	-0.02	A	0.67	-0.07	-0.01	A
37	0.01	0.01	0.00	A	0.49	0.17	0.02	A	0.21	-0.04	0.00	A
38	6.91	0.27	0.02	A	0.54	-0.18	-0.02	A	7.61	0.23	0.02	A
39	0.09	-0.03	0.00	A	0.08	-0.07	-0.01	A	1.09	-0.09	-0.01	A
40	0.09	0.04	0.00	A	0.01	-0.03	0.00	A	0.49	0.07	0.00	A
41	0.53	0.08	0.01	A	0.64	0.21	0.02	A	0.75	0.08	0.01	A
42	6.72	0.30	0.02	A	0.35	0.14	0.01	A	2.10	0.13	0.01	A
43	0.76	0.09	0.01	A	0.13	0.09	0.01	A	3.52	-0.16	-0.01	A
44	2.00	0.14	0.01	A	0.26	-0.13	-0.01	A	0.16	-0.03	0.00	A
45	0.66	0.14	0.00	A	4.37	0.68	0.04	A	4.64	0.27	0.01	A
46	2.72	0.16	0.02	A	0.17	0.09	0.01	A	0.05	0.02	0.00	A
47	75.34	-0.98	-0.08	A	0.40	-0.17	-0.01	A	11.66	-0.32	-0.02	A
48	1.10	-0.11	-0.01	A	0.66	0.20	0.02	A	0.10	-0.03	0.00	A
49	12.88	0.41	0.03	A	0.06	0.06	0.01	A	2.92	0.15	0.01	A
50	1.46	-0.13	-0.01	A	0.13	-0.09	-0.01	A	0.78	-0.08	-0.01	A
51	3.59	0.21	0.02	A	1.45	-0.30	-0.03	A	0.63	-0.07	-0.01	A
52	0.02	-0.02	0.00	A	0.01	0.03	0.00	A	0.11	0.03	0.00	A
53	1.45	-0.13	-0.01	A	0.27	-0.13	-0.01	A	2.42	-0.14	-0.01	A
54	9.58	0.41	0.02	A	2.14	-0.38	-0.03	A	0.81	-0.09	-0.01	A

Note: African Am. = African American, Native Am. = Native American, MH $\chi 2$ = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.2
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 8

	Reference: Male N= 42265 Focal: Female N= 40985				Reference: Hispanic N= 46562 Focal: Non Hispanic N= 36869				Reference: White N= 66303 Focal: Africa American N= 5059				Reference: White N= 66303 Focal: Native American N= 5442			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	45.85	-0.31	-0.02	A	7.30	-0.13	-0.01	A	4.75	-0.20	-0.01	A	5.68	-0.20	-0.01	A
2	77.86	-0.35	-0.03	A	9.55	-0.13	-0.01	A	0.02	-0.01	0.00	A	0.17	-0.03	0.00	A
3	15.48	0.15	0.01	A	0.68	0.03	0.00	A	0.84	-0.07	-0.01	A	13.42	0.27	0.02	A
4	24.52	0.23	0.01	A	0.56	0.04	0.00	A	51.25	-0.62	-0.04	A	2.80	0.14	0.01	A
5	36.59	0.23	0.02	A	34.36	-0.23	-0.02	A	0.54	0.06	0.00	A	31.59	-0.45	-0.03	A
6	95.45	0.45	0.02	A	1.08	-0.05	0.00	A	17.57	-0.37	-0.02	A	4.84	-0.18	-0.01	A
7	0.00	0.00	0.00	A	12.95	-0.16	-0.01	A	0.00	0.00	0.00	A	0.03	0.01	0.00	A
8	194.33	-0.77	-0.03	A	27.48	-0.30	-0.02	A	17.50	-0.43	-0.02	A	63.06	-0.71	-0.04	A
9	550.55	0.81	0.08	A	0.42	0.02	0.00	A	6.89	-0.19	-0.02	A	0.42	-0.04	0.00	A
10	72.13	-0.31	-0.03	A	54.54	-0.27	-0.02	A	13.90	-0.27	-0.03	A	14.69	-0.27	-0.03	A
11	5.94	0.10	0.01	A	13.78	0.16	0.01	A	0.86	-0.08	-0.01	A	0.12	0.03	0.00	A
12	377.11	0.76	0.06	A	31.37	0.22	0.02	A	0.04	-0.02	0.00	A	9.11	0.22	0.02	A
13	1.53	0.04	0.00	A	1.01	0.04	0.01	A	3.12	-0.13	-0.01	A	1.51	0.09	0.01	A
14	7.95	0.10	0.01	A	1.91	0.05	0.01	A	2.12	0.11	0.01	A	11.30	0.24	0.02	A
15	42.49	0.24	0.02	A	2.73	0.06	0.01	A	0.80	0.07	0.01	A	0.76	0.07	0.01	A
16	16.23	0.15	0.01	A	0.26	-0.02	0.00	A	6.98	-0.20	-0.02	A	1.92	-0.10	-0.01	A
17	97.95	0.36	0.03	A	5.95	-0.09	-0.01	A	0.08	0.02	0.00	A	3.42	-0.13	-0.01	A
18	25.85	0.20	0.02	A	157.68	-0.50	-0.04	A	8.02	0.23	0.02	A	26.35	-0.38	-0.03	A
19	379.79	0.80	0.05	A	37.27	0.26	0.02	A	1.96	0.11	0.01	A	1.12	0.08	0.01	A
20	116.24	-0.42	-0.03	A	122.67	-0.45	-0.03	A	6.78	-0.22	-0.01	A	33.17	-0.52	-0.03	A
21	27.40	0.18	0.02	A	28.44	0.19	0.02	A	11.19	-0.24	-0.02	A	9.82	0.22	0.02	A
22	65.76	0.30	0.02	A	3.39	0.07	0.01	A	24.39	-0.37	-0.03	A	4.32	0.15	0.01	A
23	73.99	-0.32	-0.03	A	87.91	-0.36	-0.03	A	8.84	-0.23	-0.02	A	23.38	-0.36	-0.03	A
24	0.00	0.00	0.00	A	88.41	-0.37	-0.03	A	0.05	0.02	0.00	A	28.11	-0.44	-0.03	A
25	19.69	-0.17	-0.01	A	165.49	-0.50	-0.04	A	3.10	0.14	0.01	A	28.52	-0.44	-0.03	A
26	10.93	-0.12	-0.01	A	76.90	-0.32	-0.03	A	6.25	0.19	0.02	A	5.64	-0.18	-0.01	A
27	59.38	-0.32	-0.02	A	48.54	-0.31	-0.01	A	0.54	0.07	0.00	A	24.08	-0.47	-0.03	A
28	72.33	-0.30	-0.03	A	28.89	-0.20	-0.02	A	22.27	-0.35	-0.03	A	0.74	0.06	0.01	A
29	489.44	-0.81	-0.07	A	289.30	-0.64	-0.05	A	20.64	-0.35	-0.03	A	22.08	-0.35	-0.03	A
30	450.57	0.89	0.06	A	63.39	0.34	0.02	A	2.72	-0.14	-0.01	A	10.57	0.26	0.02	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.2 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 8

	Reference: White N= 66303 Focal: Asian N= 2530				Reference: White N= 66303 Focal: Hawaii N= 361				Reference: White N= 66303 Focal: Multiple Indicator N= 2569			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	28.90	-0.76	-0.03	A	2.60	-0.51	-0.03	A	0.02	0.02	0.00	A
2	16.07	-0.51	-0.03	A	1.91	-0.40	-0.03	A	0.51	0.08	0.01	A
3	3.60	-0.23	-0.01	A	0.66	-0.23	-0.02	A	0.06	0.03	0.00	A
4	7.53	-0.46	-0.02	A	0.06	0.08	0.00	A	1.81	0.18	0.01	A
5	20.80	0.57	0.04	A	0.75	-0.24	-0.02	A	0.59	-0.08	-0.01	A
6	2.17	-0.24	-0.01	A	12.27	-1.09	-0.07	B<	0.75	0.12	0.01	A
7	6.15	-0.38	-0.01	A	0.26	-0.17	-0.01	A	11.67	0.47	0.02	A
8	14.79	-0.76	-0.02	A	0.29	-0.22	-0.01	A	1.10	0.18	0.01	A
9	0.00	0.00	0.00	A	0.69	0.22	0.02	A	2.31	-0.15	-0.01	A
10	13.34	-0.40	-0.03	A	4.65	-0.57	-0.05	A	0.02	-0.02	0.00	A
11	3.61	-0.26	-0.01	A	5.72	0.85	0.05	A	0.34	-0.07	0.00	A
12	9.10	0.42	0.02	A	1.37	-0.34	-0.03	A	1.26	0.13	0.01	A
13	8.24	0.29	0.03	A	1.18	-0.28	-0.03	A	1.85	-0.14	-0.01	A
14	39.40	0.70	0.06	A	6.48	0.67	0.06	A	4.32	0.21	0.02	A
15	20.82	0.46	0.04	A	0.94	0.27	0.02	A	0.67	0.09	0.01	A
16	0.46	-0.08	-0.01	A	0.49	0.19	0.02	A	3.24	0.20	0.02	A
17	3.52	0.22	0.02	A	0.76	-0.23	-0.02	A	0.06	0.03	0.00	A
18	2.05	0.19	0.01	A	6.20	-0.73	-0.06	A	2.98	0.20	0.01	A
19	17.46	0.64	0.03	A	0.54	-0.23	-0.02	A	0.00	0.01	0.00	A
20	25.36	0.54	0.05	A	0.10	-0.09	-0.01	A	1.73	-0.14	-0.01	A
21	16.53	0.44	0.04	A	0.16	-0.10	-0.01	A	5.04	-0.22	-0.02	A
22	7.85	0.35	0.02	A	0.02	0.04	0.00	A	2.53	-0.17	-0.01	A
23	40.26	-0.72	-0.05	A	1.57	-0.33	-0.03	A	0.82	0.10	0.01	A
24	16.90	0.46	0.04	A	1.26	-0.32	-0.03	A	1.01	-0.11	-0.01	A
25	30.79	0.67	0.05	A	0.90	0.26	0.02	A	2.07	0.16	0.01	A
26	0.15	-0.04	0.00	A	2.74	-0.45	-0.04	A	2.19	0.15	0.01	A
27	17.52	0.47	0.04	A	3.38	-0.63	-0.04	A	8.62	-0.36	-0.02	A
28	0.06	0.03	0.00	A	0.06	0.07	0.01	A	0.60	-0.08	-0.01	A
29	8.29	-0.32	-0.02	A	2.51	-0.43	-0.04	A	0.63	-0.08	-0.01	A
30	0.96	0.15	0.01	A	5.22	0.74	0.05	A	0.28	-0.07	0.00	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.2 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 8

	Reference: Male N= 42265 Focal: Female N= 40985				Reference: Hispanic N= 46562 Focal: Non Hispanic N= 36869				Reference: White N= 66303 Focal: Africa American N= 5059				Reference: White N= 66303 Focal: Native American N= 5442			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	37.00	-0.21	-0.02	A	9.75	-0.11	-0.01	A	0.94	0.07	0.01	A	0.55	0.05	0.01	A
32	574.86	1.16	0.06	B>	0.12	0.02	0.00	A	0.34	-0.05	0.00	A	0.00	0.00	0.00	A
33	0.85	0.04	0.00	A	9.76	0.13	0.01	A	0.32	-0.05	0.00	A	0.77	0.07	0.01	A
34	1854.92	-1.55	-0.14	C<	14.67	-0.14	-0.01	A	1.79	-0.10	-0.01	A	2.97	-0.13	-0.01	A
35	27.15	-0.18	-0.02	A	34.37	-0.21	-0.02	A	8.29	0.21	0.02	A	0.13	-0.03	0.00	A
36	96.06	-0.38	-0.03	A	122.79	-0.43	-0.03	A	9.37	-0.24	-0.02	A	24.99	-0.36	-0.03	A
37	25.01	0.17	0.02	A	77.12	-0.31	-0.03	A	11.65	0.25	0.02	A	0.14	-0.03	0.00	A
38	188.13	0.49	0.04	A	72.74	0.32	0.03	A	1.59	-0.09	-0.01	A	1.28	0.08	0.01	A
39	0.75	0.03	0.00	A	2.66	0.06	0.01	A	4.13	-0.15	-0.01	A	0.04	0.01	0.00	A
40	9.30	-0.12	-0.01	A	2.33	0.06	0.00	A	0.24	-0.04	0.00	A	2.94	-0.13	-0.01	A
41	1.35	-0.05	0.00	A	21.64	-0.20	-0.01	A	12.55	-0.29	-0.02	A	0.61	0.06	0.01	A
42	518.10	1.05	0.06	B>	8.63	0.14	0.01	A	0.19	-0.04	0.00	A	3.23	-0.15	-0.01	A
43	0.00	0.00	0.00	A	24.33	-0.18	-0.01	A	21.94	-0.35	-0.03	A	17.51	-0.31	-0.03	A
44	175.80	0.65	0.03	A	9.60	0.15	0.01	A	1.49	-0.11	-0.01	A	3.28	0.15	0.01	A
45	56.49	-0.26	-0.03	A	39.11	-0.22	-0.02	A	10.38	-0.23	-0.02	A	0.09	-0.02	0.00	A
46	100.63	-0.36	-0.03	A	0.04	-0.01	0.00	A	10.42	-0.24	-0.02	A	1.10	0.08	0.01	A
47	11.04	0.13	0.01	A	21.25	0.18	0.02	A	17.47	-0.32	-0.03	A	2.21	0.11	0.01	A
48	11.74	-0.13	-0.01	A	76.16	-0.34	-0.03	A	14.23	0.30	0.02	A	13.07	-0.27	-0.02	A
49	227.92	0.54	0.05	A	67.87	0.31	0.03	A	0.33	0.04	0.00	A	3.63	0.14	0.01	A
50	0.17	0.02	0.00	A	0.40	-0.02	0.00	A	0.17	0.03	0.00	A	0.03	-0.01	0.00	A
51	168.37	0.49	0.04	A	61.46	-0.31	-0.02	A	1.92	0.11	0.01	A	0.07	0.02	0.00	A
52	13.68	-0.13	-0.01	A	89.60	-0.35	-0.03	A	0.18	0.03	0.00	A	0.46	0.05	0.01	A
53	47.97	0.24	0.02	A	1.51	0.04	0.01	A	8.70	0.22	0.02	A	12.96	-0.26	-0.02	A
54	37.03	-0.23	-0.02	A	2.67	-0.06	0.00	A	0.80	-0.07	-0.01	A	2.49	-0.12	-0.01	A
55	1.29	0.04	0.00	A	1.37	-0.04	0.00	A	3.52	0.15	0.01	A	1.93	-0.11	-0.01	A
56	24.85	0.23	0.01	A	18.88	0.21	0.01	A	9.67	-0.28	-0.02	A	0.02	0.01	0.00	A
57	139.29	-0.40	-0.04	A	17.34	-0.14	-0.01	A	5.32	-0.16	-0.02	A	15.04	-0.27	-0.03	A
58	48.19	-0.24	-0.02	A	91.38	-0.34	-0.03	A	35.34	-0.43	-0.04	A	9.80	-0.22	-0.02	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.2 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT Grade 8

	Reference: White N= 66303 Focal: Asian N= 2530				Reference: White N= 66303 Focal: Hawaii N= 361				Reference: White N= 66303 Focal: Multiple Indicator N= 2569			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
31	1.24	0.12	0.01	A	0.68	-0.22	-0.02	A	1.14	0.11	0.01	A
32	1.88	0.26	0.01	A	0.37	-0.22	-0.01	A	1.36	0.17	0.01	A
33	24.66	0.79	0.03	A	5.36	0.72	0.05	A	0.00	0.01	0.00	A
34	12.97	0.39	0.03	A	0.42	-0.17	-0.02	A	0.89	0.10	0.01	A
35	42.92	0.70	0.06	A	0.15	-0.10	-0.01	A	3.22	0.18	0.02	A
36	1.52	0.16	0.01	A	0.79	-0.25	-0.02	A	0.76	0.10	0.01	A
37	8.77	0.31	0.03	A	0.13	0.09	0.01	A	1.58	0.12	0.01	A
38	8.80	0.34	0.03	A	0.26	-0.14	-0.01	A	0.11	0.03	0.00	A
39	12.39	0.40	0.03	A	0.17	0.11	0.01	A	0.23	-0.05	0.00	A
40	53.62	1.03	0.05	B>	0.68	0.25	0.02	A	0.68	-0.09	-0.01	A
41	4.70	-0.31	-0.01	A	0.13	0.11	0.01	A	0.02	0.02	0.00	A
42	0.96	0.17	0.01	A	0.05	0.08	0.00	A	0.73	0.12	0.01	A
43	0.00	0.00	0.00	A	1.02	-0.28	-0.02	A	0.20	0.05	0.00	A
44	3.67	-0.36	-0.01	A	0.13	0.13	0.01	A	0.78	0.13	0.01	A
45	4.51	0.22	0.02	A	1.37	-0.30	-0.03	A	3.41	-0.18	-0.02	A
46	5.79	0.28	0.02	A	1.59	0.34	0.03	A	0.96	-0.10	-0.01	A
47	0.62	0.10	0.01	A	1.18	0.31	0.03	A	4.81	-0.24	-0.02	A
48	4.76	0.28	0.02	A	0.12	0.10	0.01	A	6.43	0.29	0.02	A
49	9.83	0.36	0.03	A	0.74	0.24	0.02	A	0.00	0.00	0.00	A
50	1.30	0.14	0.01	A	1.10	0.29	0.02	A	6.13	0.27	0.02	A
51	85.08	0.97	0.09	A	2.12	0.43	0.03	A	0.01	0.01	0.00	A
52	1.22	-0.12	-0.01	A	2.23	0.39	0.04	A	12.89	0.38	0.03	A
53	25.80	0.57	0.05	A	0.63	0.20	0.02	A	0.29	-0.05	0.00	A
54	10.76	0.41	0.03	A	0.59	0.22	0.02	A	0.01	0.01	0.00	A
55	59.23	0.80	0.08	A	0.05	-0.06	-0.01	A	0.84	-0.09	-0.01	A
56	2.34	-0.27	-0.01	A	2.33	-0.51	-0.03	A	1.34	0.17	0.01	A
57	0.84	-0.09	-0.01	A	0.52	-0.18	-0.02	A	0.18	-0.04	0.00	A
58	16.28	0.44	0.04	A	0.28	-0.14	-0.01	A	2.38	0.16	0.01	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.3
Spring 2017 AIMS Differential Item Functioning
Science CRT High School

	Reference: Male N= 40327 Focal: Female N= 39875				Reference: Hispanic N= 45354 Focal: Non Hispanic N= 35051				Reference: White N= 63627 Focal: Africa American N= 4843				Reference: White N= 63627 Focal: Native American N= 3970			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	53.49	0.26	0.02	A	0.28	-0.02	0.00	A	23.86	-0.37	-0.03	A	8.91	-0.25	-0.02	A
2	146.68	0.47	0.04	A	3.27	0.07	0.00	A	2.87	-0.14	-0.01	A	6.26	-0.22	-0.02	A
3	56.90	-0.28	-0.02	A	18.70	-0.17	-0.01	A	9.34	0.24	0.02	A	2.93	0.15	0.01	A
4	63.54	-0.28	-0.03	A	0.02	0.00	0.00	A	4.11	-0.15	-0.01	A	1.07	0.09	0.01	A
5	0.09	0.01	0.00	A	144.08	0.44	0.04	A	13.46	0.27	0.02	A	71.51	0.68	0.07	A
6	277.20	-0.60	-0.06	A	8.69	-0.11	-0.01	A	6.62	0.20	0.02	A	0.64	0.07	0.01	A
7	0.07	0.01	0.00	A	111.90	-0.45	-0.03	A	12.92	0.31	0.02	A	0.59	0.08	0.01	A
8	87.30	-0.34	-0.03	A	0.86	-0.03	0.00	A	13.26	-0.27	-0.03	A	0.29	-0.04	0.00	A
9	7.08	0.10	0.01	A	12.03	-0.13	-0.01	A	19.73	-0.34	-0.03	A	1.39	-0.10	-0.01	A
10	64.04	0.30	0.03	A	0.71	0.03	0.00	A	4.49	0.16	0.01	A	0.01	0.01	0.00	A
11	2.83	-0.06	-0.01	A	0.52	-0.03	0.00	A	2.38	-0.11	-0.01	A	12.50	-0.29	-0.03	A
12	9.25	-0.12	-0.01	A	39.38	-0.25	-0.02	A	3.42	0.15	0.01	A	2.31	0.14	0.01	A
13	20.89	0.19	0.01	A	37.05	-0.26	-0.02	A	4.02	-0.17	-0.01	A	77.10	-0.77	-0.06	A
14	1.54	-0.04	0.00	A	25.62	0.18	0.02	A	0.62	0.06	0.01	A	12.28	0.28	0.03	A
15	195.91	0.57	0.04	A	2.66	-0.07	0.00	A	12.72	-0.29	-0.02	A	33.25	-0.50	-0.04	A
16	228.69	-0.56	-0.05	A	184.14	-0.51	-0.05	A	45.06	-0.50	-0.05	A	18.69	-0.36	-0.03	A
17	68.57	-0.31	-0.03	A	16.72	-0.16	-0.01	A	2.16	0.12	0.01	A	9.31	-0.28	-0.02	A
18	27.04	-0.19	-0.02	A	0.42	-0.02	0.00	A	0.08	0.02	0.00	A	3.66	-0.16	-0.01	A
19	20.18	0.16	0.01	A	31.18	-0.21	-0.02	A	1.26	0.09	0.01	A	0.01	-0.01	0.00	A
20	1467.18	-1.48	-0.12	B<	174.15	-0.52	-0.04	A	22.33	-0.38	-0.03	A	0.02	0.01	0.00	A
21	300.28	-0.66	-0.05	A	37.64	-0.24	-0.02	A	37.92	-0.47	-0.04	A	0.01	0.01	0.00	A
22	6.48	0.09	0.01	A	20.01	0.16	0.01	A	0.00	0.00	0.00	A	21.76	0.38	0.04	A
23	248.73	-0.63	-0.05	A	3.54	0.08	0.00	A	4.00	-0.16	-0.01	A	0.10	-0.03	0.00	A
24	583.07	0.88	0.08	A	8.88	0.11	0.01	A	0.36	-0.05	0.00	A	4.78	0.18	0.02	A
25	382.10	0.72	0.06	A	141.19	-0.45	-0.03	A	21.52	-0.36	-0.03	A	75.99	-0.77	-0.06	A
26	32.85	-0.21	-0.02	A	16.00	-0.15	-0.01	A	13.96	-0.30	-0.02	A	8.04	-0.25	-0.02	A
27	2.39	-0.06	0.00	A	1.62	-0.05	0.00	A	4.61	0.16	0.01	A	4.30	-0.18	-0.02	A
28	162.84	-0.51	-0.04	A	17.10	-0.17	-0.01	A	0.14	-0.03	0.00	A	32.40	0.53	0.04	A
29	333.76	-0.68	-0.06	A	63.77	-0.31	-0.02	A	0.01	-0.01	0.00	A	9.68	-0.28	-0.02	A
30	45.13	-0.24	-0.02	A	7.68	-0.10	-0.01	A	2.41	-0.12	-0.01	A	0.60	-0.06	-0.01	A
31	24.48	0.18	0.02	A	9.63	0.12	0.01	A	6.12	0.19	0.02	A	0.33	-0.05	0.00	A
32	129.46	0.41	0.04	A	3.66	0.07	0.01	A	0.28	0.04	0.00	A	1.03	0.08	0.01	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.3 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT High School

	Reference: White N= 63627 Focal: Asian N= 2547				Reference: White N= 63627 Focal: Hawaii N= 340				Reference: White N= 63627 Focal: Multiple Indicator N= 2644			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
1	2.56	-0.17	-0.01	A	0.47	-0.19	-0.02	A	0.34	-0.06	-0.01	A
2	8.42	0.36	0.02	A	0.68	0.24	0.02	A	3.48	-0.20	-0.02	A
3	71.90	0.94	0.08	A	8.98	0.84	0.07	A	0.35	-0.06	-0.01	A
4	89.55	0.99	0.09	A	0.12	0.09	0.01	A	3.77	-0.19	-0.02	A
5	59.48	0.77	0.08	A	1.65	-0.36	-0.03	A	0.01	-0.01	0.00	A
6	14.93	0.40	0.04	A	0.35	0.16	0.01	A	7.41	0.27	0.03	A
7	154.11	1.39	0.11	B>	5.25	0.67	0.05	A	0.86	-0.10	-0.01	A
8	0.13	0.04	0.00	A	0.85	-0.24	-0.02	A	0.63	-0.08	-0.01	A
9	12.44	0.43	0.03	A	0.00	-0.01	0.00	A	0.11	0.03	0.00	A
10	16.56	0.49	0.03	A	0.29	0.15	0.01	A	0.77	-0.09	-0.01	A
11	1.07	0.11	0.01	A	0.00	0.00	0.00	A	1.01	-0.10	-0.01	A
12	24.09	0.49	0.05	A	0.96	0.28	0.02	A	0.17	0.04	0.00	A
13	0.09	0.04	0.00	A	8.87	-0.91	-0.07	A	1.48	0.14	0.01	A
14	2.66	0.16	0.02	A	1.85	-0.36	-0.04	A	0.04	0.02	0.00	A
15	5.50	-0.29	-0.02	A	2.52	-0.50	-0.04	A	0.13	0.04	0.00	A
16	11.22	-0.38	-0.03	A	4.37	-0.60	-0.05	A	0.00	0.00	0.00	A
17	20.88	0.47	0.04	A	0.03	-0.04	0.00	A	9.26	0.31	0.03	A
18	25.90	0.57	0.05	A	0.93	0.26	0.02	A	0.13	0.04	0.00	A
19	1.23	0.11	0.01	A	2.93	0.46	0.04	A	0.49	0.07	0.01	A
20	0.08	-0.03	0.00	A	0.00	0.01	0.00	A	3.72	0.21	0.02	A
21	22.42	0.61	0.04	A	0.27	0.16	0.01	A	0.33	0.06	0.00	A
22	5.51	-0.23	-0.02	A	2.21	-0.41	-0.04	A	0.68	-0.08	-0.01	A
23	1.05	-0.13	-0.01	A	1.98	-0.43	-0.03	A	1.54	0.14	0.01	A
24	36.95	0.69	0.05	A	0.12	-0.09	-0.01	A	0.24	0.05	0.00	A
25	18.88	-0.47	-0.04	A	6.83	-0.75	-0.06	A	7.16	-0.27	-0.02	A
26	1.37	-0.13	-0.01	A	0.03	-0.05	0.00	A	0.27	0.05	0.00	A
27	3.75	0.21	0.02	A	0.23	-0.13	-0.01	A	0.65	-0.08	-0.01	A
28	57.51	0.82	0.07	A	7.96	0.81	0.06	A	0.72	-0.10	-0.01	A
29	0.83	0.09	0.01	A	1.09	-0.31	-0.03	A	0.83	0.09	0.01	A
30	2.49	0.17	0.01	A	1.10	-0.28	-0.03	A	0.34	0.06	0.01	A
31	9.16	0.32	0.03	A	0.94	-0.27	-0.02	A	0.91	-0.10	-0.01	A
32	32.94	0.64	0.05	A	0.06	0.07	0.01	A	2.89	-0.17	-0.02	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.3 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT High School

	Reference: Male N= 42265 Focal: Female N= 40985				Reference: Hispanic N= 46562 Focal: Non Hispanic N= 36869				Reference: White N= 66303 Focal: Africa American N= 5059				Reference: White N= 66303 Focal: Native American N= 5442			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
33	15.77	-0.15	-0.01	A	6.30	-0.10	0.00	A	0.19	-0.03	0.00	A	1.71	-0.11	-0.01	A
34	65.57	0.29	0.03	A	17.40	0.15	0.01	A	14.98	0.29	0.03	A	12.48	0.29	0.03	A
35	1.43	0.04	0.00	A	0.95	0.04	0.01	A	0.03	0.01	0.00	A	3.34	0.16	0.01	A
36	49.15	0.31	0.02	A	0.60	-0.03	0.00	A	74.96	-0.72	-0.05	A	4.83	0.20	0.02	A
37	0.04	0.01	0.00	A	0.50	0.03	0.01	A	0.40	-0.05	0.00	A	7.50	0.23	0.02	A
38	18.81	0.16	0.01	A	103.17	-0.38	-0.03	A	9.50	-0.23	-0.02	A	13.44	-0.31	-0.03	A
39	0.00	0.00	0.00	A	21.26	-0.17	-0.01	A	12.74	-0.28	-0.02	A	35.76	0.49	0.05	A
40	2.43	0.06	0.01	A	7.50	0.10	0.01	A	0.52	-0.05	0.00	A	5.78	-0.20	-0.02	A
41	36.85	-0.22	-0.02	A	106.61	-0.40	-0.03	A	4.28	0.16	0.01	A	1.14	-0.10	-0.01	A
42	7.01	-0.10	-0.01	A	19.83	-0.17	-0.01	A	0.00	0.00	0.00	A	3.34	-0.16	-0.01	A
43	7.86	-0.10	-0.01	A	6.83	0.10	0.01	A	5.58	-0.19	-0.02	A	8.49	-0.25	-0.02	A
44	60.48	-0.28	-0.03	A	4.20	-0.08	-0.01	A	16.93	-0.32	-0.03	A	2.32	0.13	0.01	A
45	5.86	0.09	0.01	A	5.80	-0.10	-0.01	A	3.43	0.15	0.01	A	7.52	-0.27	-0.02	A
46	192.58	0.51	0.05	A	105.91	-0.38	-0.03	A	0.74	-0.07	-0.01	A	0.38	0.05	0.00	A
47	140.10	-0.45	-0.04	A	433.60	-0.81	-0.06	A	1.45	-0.09	-0.01	A	6.54	-0.23	-0.02	A
48	2.97	-0.06	0.00	A	11.46	-0.12	-0.01	A	5.49	-0.18	-0.02	A	1.09	-0.09	-0.01	A
49	92.50	-0.36	-0.03	A	1.81	-0.05	-0.01	A	2.83	0.14	0.01	A	0.30	0.05	0.00	A
50	169.10	0.50	0.04	A	0.37	-0.02	0.00	A	0.25	0.04	0.00	A	18.75	-0.37	-0.03	A
51	214.62	-0.53	-0.05	A	13.23	0.14	0.01	A	0.06	-0.02	0.00	A	6.74	-0.23	-0.02	A
52	7.31	-0.10	-0.01	A	26.40	-0.19	-0.02	A	4.62	-0.16	-0.01	A	1.76	-0.11	-0.01	A
53	3.98	-0.08	-0.01	A	16.34	-0.17	-0.01	A	0.10	0.03	0.00	A	0.10	0.03	0.00	A
54	8.34	0.11	0.01	A	29.87	-0.21	-0.02	A	21.24	-0.36	-0.03	A	42.61	-0.58	-0.05	A
55	298.41	-0.70	-0.05	A	231.49	-0.62	-0.05	A	4.83	-0.18	-0.01	A	5.86	-0.22	-0.02	A
56	687.57	1.04	0.08	B>	22.75	0.19	0.02	A	0.06	0.02	0.00	A	7.81	0.24	0.02	A
57	5.51	-0.08	-0.01	A	4.20	-0.08	-0.01	A	0.59	-0.06	-0.01	A	45.85	-0.58	-0.05	A
58	21.71	-0.18	-0.01	A	53.15	-0.29	-0.02	A	0.16	-0.03	0.00	A	15.96	-0.37	-0.03	A
59	56.97	0.27	0.03	A	6.26	-0.09	-0.01	A	0.78	-0.07	-0.01	A	4.12	0.17	0.02	A
60	2.15	0.06	0.00	A	8.96	0.12	0.01	A	0.14	-0.03	0.00	A	11.67	-0.31	-0.02	A
61	396.44	0.80	0.06	A	5.18	0.09	0.01	A	1.06	0.08	0.01	A	13.55	-0.32	-0.03	A
62	2.01	-0.05	0.00	A	2.19	-0.05	0.00	A	6.15	0.19	0.02	A	3.14	-0.15	-0.01	A
63	267.14	0.60	0.05	A	20.00	-0.17	-0.01	A	0.18	0.03	0.00	A	1.28	-0.10	-0.01	A
64	0.05	-0.01	0.00	A	33.38	-0.22	-0.02	A	22.82	-0.36	-0.03	A	1.95	0.12	0.01	A
65	91.96	0.35	0.03	A	1.99	-0.05	0.00	A	0.41	0.05	0.00	A	0.12	-0.03	0.00	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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 B.3 (continued)
Spring 2017 AIMS Differential Item Functioning
Science CRT High School

	Reference: White N= 66303 Focal: Asian N= 2530				Reference: White N= 66303 Focal: Hawaii N= 361				Reference: White N= 66303 Focal: Multiple Indicator N= 2569			
Item	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag	MH χ2	ΔMH	SMD	Flag
33	13.19	0.39	0.03	A	1.04	-0.30	-0.03	A	0.15	0.04	0.00	A
34	35.21	0.61	0.06	A	0.00	0.00	0.00	A	0.73	0.08	0.01	A
35	85.42	0.94	0.09	A	0.13	0.10	0.01	A	3.43	-0.19	-0.02	A
36	6.80	0.43	0.02	A	0.10	0.11	0.01	A	0.63	-0.10	-0.01	A
37	0.30	-0.07	0.00	A	0.10	0.09	0.01	A	0.31	0.06	0.00	A
38	3.02	-0.19	-0.02	A	0.01	-0.02	0.00	A	0.64	0.08	0.01	A
39	1.31	-0.12	-0.01	A	0.01	0.02	0.00	A	0.62	0.08	0.01	A
40	8.20	0.34	0.02	A	0.00	-0.02	0.00	A	0.86	-0.10	-0.01	A
41	21.19	0.47	0.04	A	0.14	-0.11	-0.01	A	0.15	0.04	0.00	A
42	0.88	0.10	0.01	A	0.06	0.07	0.01	A	0.00	0.00	0.00	A
43	24.19	0.55	0.04	A	0.22	0.13	0.01	A	3.80	-0.21	-0.02	A
44	18.94	-0.44	-0.04	A	2.40	-0.44	-0.04	A	0.01	0.01	0.00	A
45	0.76	-0.09	-0.01	A	2.29	0.44	0.04	A	0.60	0.08	0.01	A
46	0.98	-0.11	-0.01	A	0.13	-0.10	-0.01	A	9.78	0.32	0.03	A
47	0.23	0.06	0.00	A	0.97	0.28	0.02	A	4.11	0.21	0.02	A
48	4.75	0.22	0.02	A	0.34	0.16	0.01	A	7.22	0.26	0.03	A
49	0.63	-0.08	-0.01	A	0.26	0.15	0.01	A	2.88	0.18	0.01	A
50	13.76	-0.45	-0.03	A	2.09	-0.40	-0.03	A	2.60	-0.17	-0.01	A
51	13.15	0.37	0.04	A	11.63	-1.02	-0.08	B<	9.03	0.30	0.03	A
52	26.03	0.58	0.04	A	3.58	-0.53	-0.05	A	0.76	0.09	0.01	A
53	60.44	0.80	0.07	A	0.71	0.26	0.02	A	1.49	0.13	0.01	A
54	0.03	0.02	0.00	A	0.00	-0.01	0.00	A	0.00	0.01	0.00	A
55	4.99	-0.29	-0.02	A	0.00	-0.02	0.00	A	0.81	-0.10	-0.01	A
56	13.57	0.50	0.03	A	0.38	-0.18	-0.01	A	0.68	-0.09	-0.01	A
57	12.98	0.38	0.03	A	0.18	0.11	0.01	A	1.65	-0.13	-0.01	A
58	18.09	0.47	0.04	A	0.11	-0.10	-0.01	A	0.01	0.01	0.00	A
59	3.44	0.20	0.02	A	3.65	0.51	0.05	A	0.97	-0.10	-0.01	A
60	0.06	0.03	0.00	A	1.28	0.33	0.03	A	6.78	-0.28	-0.02	A
61	27.51	0.76	0.04	A	1.08	0.31	0.02	A	1.83	-0.15	-0.01	A
62	3.57	0.20	0.02	A	0.10	-0.08	-0.01	A	2.51	-0.16	-0.01	A
63	18.04	0.47	0.04	A	0.15	0.11	0.01	A	2.76	0.17	0.02	A
64	37.01	-0.63	-0.06	A	0.31	0.15	0.01	A	0.01	-0.01	0.00	A
65	2.19	0.17	0.01	A	0.01	0.03	0.00	A	0.65	-0.08	-0.01	A

Note: African Am. = African American, Native Am. = Native American, MH χ2 = Mantel_Haenszel Chi-Square, ΔMH = Mantel-Haenszel Delta 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.