TECHNICAL REPORT

PENNSYLVANIA
KEYSTONE EXAMS

2017 Keystone Technical Report: Algebra 1, Biology, and Literature

Provided by Data Recognition Corporation

This document has been formatted to be ADA compliant.

TABLE OF CONTENTS

PREFACE: AN OVERVIEW OF THE ASSESSMENTS

THE KEYSTONE EXAMS FROM 2008 TO PRESENT

COMPREHENSIVE GRADUATION COMPETENCY ASSESSMENT PROGRAM

In 2008, the Commonwealth of Pennsylvania initiated a comprehensive graduation competency assessment program. The goals of this program include the following:

- To provide a system that is aligned, focused, standards-based, accurate, universally applicable, and publicly accessible
- To develop, produce, distribute, administer (both online and in paper-and-pencil), collect, score, analyze, track, and report results of graduation competency assessments for ten high school–level content areas: Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History, with each area or course comprised of modules containing unique content
- To provide graduation competency testing opportunities for students three times each school year—spring, summer, and fall—with students permitted to retake modules until proficiency is achieved on each module
- To report graduation competency results under accelerated timelines
- To ensure validity and reliability of the assessment systems through technically sound test development and psychometric practices, detailed statistical analyses and research studies, and well-documented processes and quality procedures

The Keystone Exams, as the graduation competency assessments are named, are just one component of Pennsylvania's system of high school graduation requirements. Keystone Exams are designed to help school districts guide students toward meeting state standards—standards aligned with expectations for success in college and the workplace. In order to receive a diploma, students are also required to meet local district credit and attendance requirements and to complete a culminating project, along with any additional district requirements.

For graduating classes, students are to demonstrate successful completion of secondary-level course work in Algebra I, Biology, and Literature, in which the Keystone Exam served as the final course exam. Students' Keystone Exam scores count for at least one-third of the final course grades.

Based upon Chapter 4 regulations, each Keystone Exam is designed in modules that reflect distinct, related academic content common to the traditional progression of course work. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module based upon other specific requirements.

ASSESSMENT ACTIVITIES OCCURRING FROM 2010 TO PRESENT

The first assessment activities took place in the 2010–2011 school year. Prior to November 2010, there were no Keystone Exams assessment events. The table below outlines the field tests and operational exams administered during the 2010–11 school year.

Following the development of Assessment Anchors and Eligible Content, exams were developed for the initial field test in 2010 and were subsequently administered as operational exams in 2011. Additional exams, which were based on the Assessment Anchors and Eligible Content developed in 2009 and 2010, were developed for the initial field test in 2011. Detailed information about the operational exam activities that occurred during the 2010–2011 school year is in the *Keystone Exams Spring 2011 Algebra I, Biology, and Literature Technical Report.*

Field Test and Operational Exams during the 2015–16 School Year

Exam	Assessment Activity	Date
Algebra I	Initial Stand-Alone Field Test	Fall 2010 (November)
Algebra I	Inaugural Operational Exam Administration	Spring 2011 (May)
Algebra II	Initial Stand-Alone Field Test	Spring 2011 (May)
Biology	Initial Stand-Alone Field Test	Fall 2010 (November)
Biology	Inaugural Operational Exam Administration	Spring 2011 (May)
English Composition	Initial Stand-Alone Field Test	Spring 2011 (May)
Geometry	Initial Stand-Alone Field Test	Spring 2011 (May)
Literature	Initial Stand-Alone Field Test	Fall 2010 (November)
Literature	Inaugural Operational Exam Administration	Spring 2011 (May)

Following a one-year program hiatus in 2012, the field test items embedded in the Spring 2011 operational forms were used to construct the forms for the next four administrations (spring, summer, winter, and possible breach). The table below outlines exams administered during the 2012–13 school year. Detailed information about the operational exam activities that occurred during the 2012–2013 school year is in the *Keystone Exams Spring 2013 Algebra I, Biology, and Literature Technical Report.*

Operational Exams during the 2012–13 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Algebra I	Operational Exam Administration	Spring 2013 (May)
Algebra I	Operational Retest Exam Administration	Summer 2013 (August)
Biology	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Biology	Operational Exam Administration	Spring 2013 (May)
Biology	Operational Retest Exam Administration	Summer 2013 (August)
Literature	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Literature	Operational Exam Administration	Spring 2013 (May)
Literature	Operational Retest Exam Administration	Summer 2013 (August)

Some of the field test items embedded in the Spring 2013 operational forms were used to construct the forms for the next four administrations (spring, summer, winter, and possible breach). The core items on the 2013–2014 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter 3. The table below outlines exams administered during the 2013–14 school year. Detailed information about the operational exam activities that occurred during the 2013–2014 school year is in the *Keystone Exams Spring 2014 Algebra I, Biology, and Literature Technical Report.*

Operational Exams during the 2013–14 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Algebra I	Operational Exam Administration	Spring 2014 (May)
Algebra I	Operational Retest Exam Administration	Summer 2014 (August)
Biology	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Biology	Operational Exam Administration	Spring 2014 (May)
Biology	Operational Retest Exam Administration	Summer 2014 (August)
Literature	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Literature	Operational Exam Administration	Spring 2014 (May)
Literature	Operational Retest Exam Administration	Summer 2014 (August)

Some of the field test items embedded in the Spring 2014 operational forms were used to construct the forms for the next year's administrations (spring, summer, winter). The core items on the 2014–2015 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2014–15 school year.

Operational Exams during the 2014–15 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Algebra I	Operational Exam Administration	Spring 2015 (May)
Algebra I	Operational Retest Exam Administration	Summer 2015 (August)
Biology	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Biology	Operational Exam Administration	Spring 2015 (May)
Biology	Operational Retest Exam Administration	Summer 2015 (August)
Literature	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Literature	Operational Exam Administration	Spring 2015 (May)
Literature	Operational Retest Exam Administration	Summer 2015 (August)

Some of the field test items embedded in the Spring 2015 operational forms were used to construct the forms for the next year's administrations (spring, summer, winter). The core items on the 2015–2016 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2015-16 school year.

Operational Exams during the 2015–16 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Algebra I	Operational Exam Administration	Spring 2016 (May)
Algebra I	Operational Retest Exam Administration	Summer 2016 (August)
Biology	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Biology	Operational Exam Administration	Spring 2016 (May)
Biology	Operational Retest Exam Administration	Summer 2016 (August)
Literature	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Literature	Operational Exam Administration	Spring 2016 (May)
Literature	Operational Retest Exam Administration	Summer 2016 (August)

Some of the field test items embedded in the Spring 2016 operational forms were used to construct the forms for the next year's administrations (spring, summer, winter). The core items on the 2016–2017 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2016-17 school year.

Operational Exams during the 2016–17 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Algebra I	Operational Exam Administration	Spring 2017 (May)
Algebra I	Operational Retest Exam Administration	Summer 2017 (July-August)
Biology	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Biology	Operational Exam Administration	Spring 2017 (May)
Biology	Operational Retest Exam Administration	Summer 2017 (July-August)
Literature	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Literature	Operational Exam Administration	Spring 2017 (May)
Literature	Operational Retest Exam Administration	Summer 2017 (July-August)

CHAPTER ONE: BACKGROUND OF THE KEYSTONE EXAMS

This brief overview of the Pennsylvania Keystone Exams summarizes the history of the program's development process, intent and purpose, and recent changes.

ASSESSMENT HISTORY IN PENNSYLVANIA

Pennsylvania's involvement in statewide assessment actually began in the 1969–1970 school year with a purely school-based assessment known as *Educational Quality Assessment* (EQA), which continued through the 1987–1988 school year. A state-mandated student competency testing program called *Testing for Essential Learning and Literacy Skills* (TELLS) also operated from the school years of 1984–1985 through 1990–1991. Also in 1990, the state initiated an on-demand writing assessment.

The Pennsylvania System of School Assessment (PSSA) program was instituted in 1992 as a school evaluation model with reporting at the school level only. The PSSA initially measured performance in the content areas of mathematics and reading at grades 5, 8, and 11, and in writing at grades 6 and 9. Starting in 1994, as part of Chapter 5 regulations, the PSSA added student-level reports. In 1999, as part of Chapter 4 regulations, the State Board of Education adopted the Pennsylvania Academic Standards for mathematics and for reading, writing, speaking, and listening. Proficiency levels for Advanced, Proficient, Basic, and Below Basic were defined in 2000. In 2001 and 2004, the reading and mathematics assessments underwent various content enhancements to improve alignment to the 1999 Academic Standards. Grade 11 was added to the writing assessment in 2001. Then, in 2004-2005, the PSSA Assessment Anchors and Eligible Content were developed to clarify content structure and improve articulation between assessment and instruction. In addition, in 2005, the grade 6 and 9 writing assessments were moved to grades 5 and 8. By 2006, the operational mathematics and reading assessments incorporated grades 3 through 8 and 11. In 2007, the PSSA and the PSSA Assessment Anchors and Eligible Content underwent additional content enhancements. In 2008, science was added to the PSSA as an operational assessment. Starting with the 2013 field test, PSSA began a multiyear transition to a new set of standards called the Pennsylvania Core Standards. Detailed information about the operational exam activities that occurred during the 2013–2014 school year is in the *2014 PSSA Technical Report.*

THE KEYSTONE EXAMS

In 2008, the Commonwealth of Pennsylvania initiated a comprehensive graduation competency assessment program. As a key piece of this program, the Keystone Exams are designed to assess proficiency in various subject areas, including Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History. The Keystone Exams are just one component of Pennsylvania's high school graduation requirements. Students must also earn state-specified credits, fulfill the state's service learning and attendance requirements, and complete any additional local school system requirements to receive a Pennsylvania high school diploma.

The stated goals of the Keystone program are to

- provide for a system that is aligned, focused, standards-based, accurate, universally applicable, and publicly accessible.
- develop, produce, distribute, administer (both online and in paper-and-pencil), collect, score, analyze, track, and report results of graduation competency assessments for ten high school–level content areas: Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History, with each area or course composed of modules containing unique content.
- provide graduation competency testing opportunities for students three times each school year—spring, summer, and fall—with students permitted to retake modules until proficiency is achieved in each module.
- report graduation competency results under accelerated timelines.
- ensure validity and reliability of the assessment systems through technically sound test development and psychometric practices, detailed statistical analyses and research studies, and well-documented processes and quality procedures.

GRADUATION REQUIREMENTS AND THE KEYSTONE EXAMS

Based upon Chapter 4 regulations, each Keystone Exam is designed in modules that reflect distinct, related academic content common to the traditional progression of coursework. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module based on the requirements detailed below.

If a student is unable to meet the requirements in § 4.24(b)(1)(iv)(A) (relating to high school graduation requirements) after two attempts on a Keystone Exam, the student may supplement a Keystone Exam score with satisfactory completion of a project-based assessment. Points earned through satisfactory performance on one or more project modules related to the Keystone Exam module or modules that the student did not pass shall be added to the student's highest Keystone Exam score.

A student may qualify to participate in one or more project-based assessments if the student has met all of the following conditions:

1. The student has taken the course.
2. The student was unsuccessful in achieving a score of Proficient on the Keystone Exam after at least two attempts.
3. The student has met the district's attendance requirements for the course.
4. The student has participated in a satisfactory manner in supplemental instructional services under § 4.24(i).

KEYSTONE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

In 2009, the state initiated development of test designs and test blueprints for the Keystone Exams based on Pennsylvania Keystone Course Standards. Committees of Pennsylvania educators met in 2009, 2010, and 2011 to write, review, and approve Assessment Anchors and Eligible Content statements and sample exam items. To provide initial focus, each test blueprint committee was presented with materials specific to the exam in question, including a basic blueprint structure, the Pennsylvania State Standards, and draft Eligible Content statements based on the standards. The results from the initial committee work were evaluated by national, state, and local subject experts, and following revisions, they were ultimately validated by another committee of Pennsylvania educators. Following committee approval, the Keystone Assessment Anchors and Eligible Content statements for literacy, mathematics, and science were approved by the State Board of Education in September 2010.

- Mathematics
    - The first committee meetings took place in April 2009, where initial drafts of the test blueprints were developed for Algebra I, Algebra II, and Geometry.
    - A follow-up committee meeting for the three mathematics exams was held in August 2009.
- Literacy
    - The first committee meetings took place in April 2009, where initial drafts of the test blueprints were developed for English Composition and Literature.
    - A follow-up committee meeting for the two literacy exams was held in November 2009.
- Science
    - The first committee meetings took place in October 2009, where the initial draft of the test blueprint was developed for Biology.
    - A follow-up committee meeting for Biology was held in January 2010.
    - In addition, in January 2010, the initial draft of the test blueprint was developed for Chemistry.
    - Chemistry was part of a follow-up committee meeting held in late January 2010.

- Social Studies
    - The first committee meetings took place in November 2010, where initial drafts of the test blueprints were developed for Civics and Government, U.S. History, and World History.
    - A follow-up committee meeting for the Civics and Government exam was held in October 2011.
    - A follow-up committee meeting for U.S. History and World History remains unscheduled pending further decisions about the future of these Keystone exams.

WAVE IMPLEMENTATION OF THE EXAMS

The implementation plan for the Keystone Exams envisioned the ten Keystone Exams becoming operational through a series of waves. The initial wave included Algebra I, Biology, and Literature. These first three exams were field tested in fall 2010 and reached operational status with the spring 2011 administration. The second wave included Algebra II, English Composition, and Geometry; these were field tested in spring 2011. English Composition is scheduled to reach operational status at a future date. Civics and Government is projected to reach operational status following English Composition. The implementation of the five remaining courses, Algebra II, Geometry, Chemistry, U.S. History, and World History, is currently unscheduled. The Pennsylvania Department of Education continues to evaluate the implementation schedule. Table 1–1 reflects the implementation plans as of September 2017.

Table 1–1. Keystone Exams Wave Implementation Plan

Wave	Exam(s)	Initial Field Test	First Operational
1	Algebra I, Biology, Literature	Fall 2010	Spring 2011
2	English Composition	Spring 2011	TBD
2	Algebra II, Geometry	Spring 2011	Not Scheduled
3	Civics and Government	TBD	Not Scheduled
4	Chemistry, U.S. History, World History	TBD	Not Scheduled

MODE OF DELIVERY FOR THE EXAMS

One key feature of the Keystone Exams is the dual mode of delivery of the testing materials that is available to districts. In addition to the traditional paper-and-pencil format, the Keystone Exams are also available in a computer-based online format using test-delivery software.

While exam materials are still available in the traditional format (two pieces of exam materials—a test book and a separate answer book [or, in the case of English Composition, a single test/answer book]), districts are given the option to administer the exams using computer-based online testing software instead of the paper-and-pencil format.

For more information about how the online exams were developed in concert with the traditional paper-and-pencil format, see Chapter Three.

MULTIPLE TESTING OPPORTUNITIES

Another key feature of the Keystone Exams is the multiple testing opportunities provided to students. Main administrations in both spring and winter provide options for students completing course work at various times of the year and accommodate both traditional and block scheduling. In addition, a summer retest opportunity is also available. More information about the spring, winter, and summer administrations can be found in Chapter Seven.

PERFORMANCE LEVELS FOR THE KEYSTONE EXAMS

The State Board approved a set of criteria defining Advanced, Proficient, Basic, and Below Basic levels of performance for the Keystone Exams. More information about these Performance Level Descriptors (PLDs) is found in Chapter Thirteen.

OPERATIONAL TEST DESIGN INFORMATION

The test definition of each of the operational Keystone Exams, including information about exam-specific test designs, test blueprints, test layouts, item types, and other exam elements, is detailed in Chapter Three.

CHAPTER TWO: TEST DEVELOPMENT OVERVIEW OF THE KEYSTONE EXAMS

KEYSTONE BLUEPRINT/ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The Keystone Test Blueprints—known as the Keystone Exams Assessment Anchors and Eligible Content—are based on Pennsylvania Keystone Course Standards and the Pennsylvania Core Standards. Prior to the development of the Assessment Anchors, multiple groups of Pennsylvania educators convened to create a set of standards for each of the Keystone Exams. Derived from a review of existing standards, these Enhanced Standards (Course Standards) focus on what students need to know and be able to do in order to be ready for college and career.

Although the Keystone Course Standards indicate what students should know and be able to do, Assessment Anchors are designed to indicate the parts of the Keystone Course Standards (Instructional Standards) that will be assessed on the Keystone Exams. Based on recommendations from Pennsylvania educators, the Assessment Anchors were designed as a tool to improve the articulation of curricular, instructional, and assessment practices. The Assessment Anchors clarify what is expected and focus the content of the standards into what is assessable on a large-scale exam. The Assessment Anchor documents also serve to communicate Eligible Content—the range of knowledge and skills from which the Keystone Exams are designed.

The Keystone Exams Assessment Anchors and Eligible Content have been designed to hold together, or anchor, the state assessment system and curricular/instructional practices in schools by following these design parameters:

- **Clear:** The Assessment Anchors are easy to read and user friendly; they clearly detail which standards are assessed on the Keystone Exams.
- **Focused:** The Assessment Anchors identify a core set of standards that can be reasonably assessed on a large-scale assessment; this keeps educators from having to guess which standards are critical.
- **Rigorous:** The Assessment Anchors support the rigor of the state standards by assessing higher order and reasoning skills.
- **Manageable:** The Assessment Anchors define the standards in a way that can be easily incorporated into a course to prepare students for success.

The Assessment Anchors and Eligible Content are organized into cohesive blueprints, each structured with a common labeling system. This framework is organized by increasing levels of detail: first, Module (Reporting Category); second, Assessment Anchor; third, Anchor Descriptor; and fourth, Eligible Content statement. The common format of this outline is followed across the Keystone Exams.

A description of each level in the labeling system for the Keystone Exams is as follows:

- **Module:** The Assessment Anchors are organized into two thematic modules for each of the Keystone Exams, and these modules serve as the Reporting Categories for the Keystone Exams. The module title appears at the top of each page in the Assessment Anchor document. The module level is also important because the Keystone Exams are built using a module format, with each of the Keystone Exams divided into two equally sized test modules. Each module is made up of two or more Assessment Anchors.
- **Assessment Anchor:** The Assessment Anchor appears in the shaded bar across the top of each Assessment Anchor table in the Assessment Anchor document. The Assessment Anchors represent categories of subject matter that anchor the content of the Keystone Exams. Each Assessment Anchor is part of a module and has one or more Anchor Descriptors unified under it.
- **Anchor Descriptor:** Below each Assessment Anchor in the Assessment Anchor document is a specific Anchor Descriptor. The Anchor Descriptor level details the scope of content covered by the Assessment Anchor. Each Anchor Descriptor is part of an Assessment Anchor and has one or more Eligible Content unified under it.

- **Eligible Content:** The column to the right of the Anchor Descriptor in the Assessment Anchor document contains the Eligible Content statements. The Eligible Content is the most specific description of the content that is assessed in the Keystone Exams. This level is considered the assessment limit. It helps educators identify the range of content covered on the Keystone Exams.
- **Enhanced Standard:** In the column to the right of each Eligible Content statement is a code representing one or more Enhanced Standards that correlate to the Eligible Content statement. Some Eligible Content statements include annotations that clarify the scope of an Eligible Content.
- **Notes:** There are three types of notes included in the Assessment Anchor document.

  "e.g." ("for example")—sample approach, but not a limit to the Eligible Content
  "i.e." ("that is")—specific limit to the Eligible Content
  "Note"—content exclusions or definable range of the Eligible Content

The Assessment Anchor's coding is read like an outline. The coding includes the Subject (Exam), Reporting Category/Module, Assessment Anchor, Anchor Descriptor, and Eligible Content. Each exam has two modules. Each module has two or more Assessment Anchors. Each of the Assessment Anchors has one or more Anchor Descriptors, and each Anchor Descriptor has at least one Eligible Content (generally more than one). The Assessment Anchors form the basis of the test design for the exams undergoing test development. In turn, this hierarchy is the basis for organizing the total module and exam scores (based on the core [common] portions).

Table 2–1. Sample Keystone Assessment Anchor Coding

Sample Code	Subject (Exam)	Reporting Category (Module)	Assessment Anchor (AA)	Anchor Descriptor (AD)	Eligible Content (EC)
A1.1.1.2.1	**A1** Algebra I	**1** Operations and Linear Equations & Inequalities	**1** Linear Equations	**2** Write, solve, and/or graph linear equations using various methods.	**1** Write, solve, and/or apply a linear equation (including problem situations).
BIO.A.2.1.1	**BIO** Biology	**A** Cells and Cell Processes	**2** The Chemical Basis for Life	**1** Describe how the unique properties of water support life on Earth	**1** Describe the unique properties of water and how these properties support life on Earth (e.g., freezing point, high specific heat, cohesion).
L.F.2.4.1	**L** Literature	**F** Fiction	**2** Analyzing and Interpreting Literature—Fiction	**4** Use appropriate strategies to interpret and analyze the universal significance of literary fiction.	**1** Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.

The complete set of Assessment Anchors and Eligible Content can be referenced at PDE's Standards Aligned System (SAS) website: http://www.pdesas.org/Standard

HIGH-LEVEL TEST DESIGN CONSIDERATIONS

The Keystone Exams employ two types of test items (questions): multiple choice and constructed response. These item types assess different levels of knowledge and provide different information about achievement. Psychometrically, multiple-choice items are very useful and efficient tools for collecting information about a student's academic achievement. Constructed-response performance tasks generally generate fewer scorable points than multiple-choice items generate in the same amount of testing time; however, they provide tasks that are more realistic and sample eligible content that best lends itself to this item type. Furthermore, well-constructed scoring guides have made it possible to include constructed-response tasks in large-scale assessments, and trained scorers apply the scoring guides to efficiently score large numbers of student responses in a highly reliable way. The design of the Keystone Exams attempts to achieve a reasonable balance between the two item types.

Table 2–2. Keystone Exams High-Level Design Considerations

Exam	MC as Percentage of Core Points	CR as Percentage of Core Points	Number of Points per MC	Number of Points per CR	Number of Modules	Number of Assessment Anchors	Number of Eligible Content
Algebra I	60	40	1	4	2	6	33
Biology	73	27	1	3	2	8	38
Literature	65	35	1	3	2	4	56

DEPTH OF KNOWLEDGE

The goal of each Keystone Exam is for each item to be of sufficient rigor, or Webb's Depth of Knowledge (DOK) Level 3. Webb's DOK was created by Norman Webb of the Wisconsin Center for Education Research. Webb's definition of depth of knowledge is the degree or complexity of knowledge that the content curriculum standards and expectations require. Therefore, when reviewing items for depth of knowledge, the item is reviewed to determine whether it is as demanding cognitively as what the actual content curriculum standard expects. In the case of the Pennsylvania Keystone items, the item meets the criterion if the DOK of the item is in alignment with the DOK of the Assessment Anchor as defined by the Eligible Content. Webb's DOK includes four levels, from the lowest (basic recall) level to the highest (extended thinking) level.

In some specific cases, DOK level 2 was allowed when the cognitive intent of an Eligible Content was level 2. For more information on DOK, see Chapter Three and Appendix A.

ONLINE TESTING DESIGN CONSIDERATIONS

The Keystone Exams were designed from the beginning to provide a dual mode of test delivery, using traditional paper-and-pencil forms and using computer-based online forms. The computer-based online testing environment (called INSIGHT) is designed to provide a testing experience that mirrors the elements of traditional paper-and-pencil-based test delivery. This includes not only standard ancillary testing materials available in or with the printed forms, like formula sheets, periodic tables, scoring guidelines, and response spaces, but also analogs of the mechanical elements of response generation not necessarily associated with a computer-screen interface. These elements include line guides, rulers, screen highlighters, magnifiers, equation-building software, online calculators and graphing tools, and keyboard shortcuts.

Consideration of other components of online testing—like item layout, passage layout, font, screen resolution, navigation tools, and other interface mechanisms—all played a role in the overall design constraints, with some considerations having a more meaningful impact on specific exams. For more information on how the online test design impacted the overall test design considerations, see the sections below under each exam. See also Chapter Twenty for more information on a study comparing the use of both modes of delivery.

Online testing also provides an opportunity to utilize software to generate scores for student responses. In cases where responses to questions invoke numerical strings or equations, online responses can be scored through the use of lookup tables. Lookup tables are automated scoring rubrics that contain common correct and incorrect responses. When a response does not match a record in the lookup table, a human scorer is used to adjudicate the score. Operational autoscoring was only used for the Algebra I Exam; see below for more information on its use in Algebra I. For more information on scoring, see Chapter Eight.

ALGEBRA I

The Keystone Algebra I Exam has two reporting categories: Module 1, Operations and Linear Equations & Inequalities, and Module 2, Linear Functions and Data Organizations. Both modules include three Assessment Anchors. Module 1 has 18 Eligible Content, and Module 2 has 15 Eligible Content. Each module corresponds to specific content aligned to statements and specifications included in the course-specific Assessment Anchor documents. The Algebra I content included in the Keystone Algebra I multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Algebra I constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Algebra I constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

ALGEBRA I ONLINE CONSIDERATIONS

Students taking the computer-based online delivery of the Algebra I exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a standard four-function calculator, a scientific calculator, a graphing tool (similar, but not identical to, a graphing calculator), a ruler (available in metric and English units), a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. In addition, an equation builder—which allows students to generate complex equations not normally possible with a standard keyboard—is also made available with all constructed-response items. Also, if the constructed-response item requests that the student draw, label, or otherwise change a graph, special graph-drawing tools are provided for on-screen graph generation. The Algebra I general scoring guideline and formula sheets are also available to students.

Layout of both the multiple-choice and constructed-response items is optimized for minimal screen manipulation (minimal scrolling required to see graphics or text that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy.

ALGEBRA I MULTIPLE-CHOICE ITEMS

Sixty percent of the possible points on the Algebra I Exam are derived from multiple-choice items. This item type is especially efficient for measuring a broad range of content. Each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, incorrect application of an algorithm, or computational errors.

Algebra I multiple-choice items are intended to take about one and a half minutes of response time per item. They are used to assess a variety of skill levels, including problem solving. Algebra I items involving application emphasize the requirement to carry out some mathematical process to find an answer rather than simply recalling information from memory.

ALGEBRA I CONSTRUCTED-RESPONSE ITEMS

Constructed-response items (tasks) require that students read a problem description and develop an appropriate solution. Algebra I constructed-response items are designed to take about ten minutes of response time per item. Most of the constructed-response items have several components in the overall task that may enable students to enter or begin the problem at different places. In some items, each successive component is designed to assess progressively more difficult skills or higher knowledge levels. Certain components may ask students to explain their reasoning for applying particular operations or for arriving at certain conclusions. The types of tasks utilized do not necessarily require computations. Students may also be asked to perform such tasks as constructing a graph, shading some portion of a figure, or listing object combinations that meet specified criteria.

Constructed-response tasks are especially useful for measuring students' problem-solving skills in Algebra. They offer the opportunity to present real-life situations that necessitate that the students solve problems using mathematics abilities learned in the classroom. Students must read the task carefully, identify the necessary information, devise a method of solution, perform the calculations, enter the solution directly in the answer document, and, when required, offer an explanation. This provides insight into the students' mathematical knowledge, abilities, and reasoning processes.

The constructed-response Algebra items are scored on a 0–4 point scale using an item-specific scoring guideline. The item-specific scoring guideline outlines the requirements for each score point. Item-specific scoring guidelines are based on the Algebra I General Description of Scoring Guidelines. The general guidelines describe a hierarchy of responses, which represent the five score levels. See Appendix B or these portals:

- https://pa.drcedirect.com [Click on "Documents" under the "General Information" tab.]
- www.education.state.pa.us [Click on the green check mark and select "Keystone Exams."]

The Algebra I Keystone Exam includes two types of constructed-response items: Scaffolded Constructed Response Items (SCR) and Extended Constructed Response Items (ECR). Both types are scored on the same 0–4 point scale using the same Algebra I General Description of Scoring Guidelines as the base. SCR items are constructed to generally elicit four distinct responses (a response may contain more than one answer blank), and each response has the potential to earn a discrete number of score points (generally just one [1] score point per response). In turn, the four distinct responses are generally organized into four sections, with each labeled as a "Part" within an SCR. The next table shows a generic (nonauthentic) illustration of the application of this concept.

Table 2–3. Generic Example [Nonauthentic] Showing Concept of Four Distinct Responses

Stem	Part A	Part B	Part C	Part D
Presents a numerical distribution	In the answer spaces, write the list of numbers from least to greatest.	Write the mean in an answer blank.	Write the median in an answer blank.	Write the mode in an answer blank.
4 points	1 distinct point even though students enter more than one number	1 distinct point with one distinct entry	1 distinct point with one distinct entry	1 distinct point with one distinct entry

- SCR items do not require narrative, explanation, or "show all your work" responses.
- Most SCR item responses lend themselves to automatic scoring; however, **not all items can be automatically scored exclusively with the use of lookup tables.** The full application of Assessment Anchors and Eligible Content sometimes requires item construction that is incompatible with lookup tables.

In familiar and probably the most descriptive terms, Algebra I ECR items—in form, format, and scoring provisions—adhere to the philosophy of PSSA OE item format. Like SCR items, development is based on the item qualities that best measure the skills and concepts with which the item aligns.

- ECR items intentionally elicit narrative, explanation of reasoning, "explain why . . .", and/or "show your work" responses.
- In contrast to SCR items, in which DOK level 3 cognitive engagement is inferred from student responses, ECR items (through explanations and recorded work) often provide direct evidence of DOK level 3 engagement. This aspect of ECR items is intentionally included during development. Following initial development, the ECR item will be approved by PDE as accepted by the review committee, or PDE and DRC will collaborate in amending the item.

BIOLOGY

The Keystone Biology Exam has two reporting categories: Module 1[A], Cells and Cell Processes; and Module 2[B], Continuity and Unity of Life. Both modules have four Assessment Anchors. Module A has 16 Eligible Content, and Module B has 22 Eligible Content. Each module corresponds to specific content aligned to statements and specifications included in the course-specific assessment anchor documents. The Biology content included in the Keystone Biology multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Biology constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Biology constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

BIOLOGY ONLINE CONSIDERATIONS

Students taking the computer-based online delivery of the Biology Exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. The Biology general scoring guideline and a periodic table are also provided to students.

Layout of both the multiple-choice and constructed-response items is optimized for minimal screen manipulation (minimal scrolling to see graphics or text that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy.

BIOLOGY MULTIPLE-CHOICE ITEMS

Seventy-three percent of the possible points on the Biology Exam are derived from multiple-choice items. Multiple-choice items are especially efficient for measuring a broad range of content. Each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, or incorrect application of a biological principle.

Biology multiple-choice items are intended to take about one and a quarter minutes of response time per item. They are used to assess a variety of skill levels, including the application of Biology content. Biology items involving application emphasize the requirement to utilize science content to find an answer rather than simply recalling information from memory.

BIOLOGY CONSTRUCTED-RESPONSE ITEMS

Constructed-response items (tasks) require students to read a description of a Biology problem and to develop an appropriate solution. Biology constructed-response items are designed to take about eight minutes of response time per item. Constructed-response tasks are especially useful for measuring students' skills in biology. These tasks may present real-life situations that require students to solve problems using science abilities learned in the classroom. Students must read a task carefully, identify the necessary information, devise a method of solution, enter the solution directly into the answer document, and when required, offer an explanation. This provides insight into students' science knowledge, abilities, and reasoning processes.

The constructed-response science items are scored on a 0–3 point scale with an item-specific scoring guideline, and each task is carefully constructed with a scoring guideline reflecting the task requirements. The general guidelines describe a hierarchy of responses, which represents the four score levels. Each item-specific scoring guideline outlines the requirements at each score point, and each item-specific scoring guideline is based on the Biology General Description of Scoring Guidelines. See Appendix B or these portals:

- *https://pa.drcedirect.com* [Click on "Documents" under the "General Information" tab.]
- *www.education.state.pa.us* [Click on the green check mark and select "Keystone Exams."]

LITERATURE

The Keystone Literature Exam has two reporting categories: Module 1, Fiction; and Module 2, Nonfiction. Both modules have two Assessment Anchors. Module 1 has 25 Eligible Content, and Module 2 has 33 Eligible Content. The Literature Exam employs two types of test items, multiple-choice and constructed-response, and the content included aligns with content included in the Eligible Content statements. The items are designed to measure students' comprehension of the content contained in the literature passages. Each module corresponds to specific content aligned to statements and specifications included in the course-specific Assessment Anchor documents. The Literature content included in the Keystone Literature multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Literature constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Literature constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

LITERATURE ONLINE CONSIDERATIONS

Students taking the computer-based online delivery of the Literature Exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. The Literature general scoring guideline is also provided to students.

Layout of passages, multiple-choice items, and constructed-response items is optimized for minimal screen manipulation (minimal scrolling to see text and graphics that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy. In addition, the amount of space devoted to the passage compared to the amount of space devoted to the exam questions was also optimized.

LITERATURE MULTIPLE-CHOICE ITEMS

Sixty-five percent of the possible points on the Literature Exam are derived from multiple-choice items. Literature multiple-choice items are intended to take about one minute of response time per item. They are designed to measure how well students comprehend the overall meaning of a passage or make basic inferences about it. At times, asking students to choose a preferred answer is the best way to determine whether they have gleaned certain information from a story. Such information may include central idea, setting, or main events and their sequence.

Each Literature multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent some kind of misinterpretation, predisposition, unsound reasoning, or casual reading.

LITERATURE CONSTRUCTED-RESPONSE ITEMS

Constructed-response items (tasks) are designed to address comprehension of text in ways that multiple-choice items cannot. Literature constructed-response items are designed to take about five minutes of response time per item. A short written response allows students to prepare an answer and summarize using supporting details or examples derived from the text.

The Literature constructed-response items are scored on a 0–3 point scale using an item-specific scoring guideline. Each task is text-dependent and is carefully constructed with the scoring guideline reflecting the task requirements. All item-specific scoring guidelines are based on the Literature General Description of Scoring Guidelines. The general guidelines describe a hierarchy of responses, which represent the four score levels. See Appendix B or these portals:

- *https://pa.drcedirect.com* [Click on “Documents” under the “General Information” tab.]
- *www.education.state.pa.us* [Click on the green check mark and select “Keystone Exams.”]

LITERATURE PASSAGES

One of the key requirements of the Keystone Literature Exam is that students should be able to read and comprehend both literature and informational texts of sufficient text complexity and quality as required by the Assessment Anchors and Eligible Content. For example, the Literature Keystone Assessment Anchors and Eligible Content require students to engage with appropriately complex literary fiction, literary nonfiction, and informational works. Passage genres include, but are not limited to, the following: stories; excerpts from novels, biographies, and autobiographies; letters; dramas; poems; myths from diverse cultures and different time periods; texts in history/social studies, science, and other disciplines; seminal U.S. documents; the classics of American, British, and world literature; and current articles and editorials.

TEXT COMPLEXITY

Text complexity involves three components: matching reader to text and task, qualitative evaluation of the text, and quantitative evaluation of the text.

<u>MATCHING READER TO TEXT AND TASK</u>

A number of factors are taken into consideration when deciding whether a passage will be placed in the pool for possible use on the Keystone Literature Exam. The factors include, but are not limited to, the following:

- Are the conceptual load, vocabulary, syntactic patterns, sentence length, and clarity appropriate for the grade level?
- Does the passage stand the test of time as an example of literary fiction, literary nonfiction, and/or informational text, and is it judged by the committee of Pennsylvania educators as having sufficient quality?
- Is the passage “rich” enough to generate a variety of items?

- Do the passages represent a range of reading levels appropriate to the grade level?
- Do the passages lend themselves well to measuring the Keystone Assessment Anchors and Eligible Content, including text structures and elements?
- Are the passages free of issues of bias, fairness, and/or sensitivity?
- Does the pool of passages represent diversity in the areas of gender, culture, ethnicity, urban/rural status, socioeconomic status, physical differences, and age?

QUALITATIVE EVALUATION OF THE TEXT

Evaluating the text complexity of a passage is essentially a judgmental process by individuals familiar with the classroom context and what is linguistically appropriate at a given grade level. All Keystone passages to be included in the pool of passages for possible use on the Keystone Literature Exam are reviewed and approved by PDE and the Pennsylvania Reading Content Committee (a committee of Pennsylvania educators). The passages are reviewed by Pennsylvania educators to judge whether each passage meets the criteria outlined above. All potential passages are also reviewed by the Pennsylvania Bias, Fairness, and Sensitivity Committee.

QUANTITATIVE EVALUATION OF THE TEXT

Each readability program uses different methods to determine the readability for a particular passage (e.g., syllables, sentence length, number of words, vocabulary lists). Each readability formula is designed for a particular grade range of materials. When using the various readability formulas, a wide range of readability levels may be identified for a particular passage. Some readability formulas are better suited to a particular grade level. If a particular formula being used is outside of the intended range, then the results may be unreliable.

Readability of the Keystone Literature Exam passages has been determined using several of the most widely accepted readability formulas. These formulas are not used in a rigid way, but rather more informally to provide for several “snapshots” of a passage. The readability formulas used for the passages that appear on the Keystone Literature Exam are the Dale-Chall Formula, the Flesch Grade Level Formula, and the Fry Graph.

CHAPTER THREE: ITEM AND TEST DEVELOPMENT PROCESSES

GENERAL KEYSTONE TEST DEVELOPMENT PROCESSES

The 2017 Keystone Exams continued to use the core-to-core biennial overlap. Approximately 30% to 50% of the operational points in each module overlap with items used operationally 2 years prior. The 2017 Keystone Exam cores were made up of items that had appeared on the Spring 2015, Summer 2015, and/or Winter 2015/2016 cores. The remainder of the operational 2017 exams were made up of items that were field tested on the Spring 2016 Keystone Exams embedded field test administration. Table 3–1 is a graphic representation of the basic process flow and overlap of the development cycles.

Table 3–1. General Development and Usage Cycle of the Algebra I, Biology, and Literature Keystone Exams

Admin Year	Events Occuring in 2009	Events Occuring in 2010	Events Occuring in 2011	Events Occuring in 2012	Events Occuring in 2013	Events Occuring in 2014	Events Occuring in 2015	Events Occuring in 2016	Events Occuring in 2017	Events Occuring in 2018*
2010–2011	Initial Item Dev. for Fall 2010 FT	Fall 2010 Stand-alone FT; New Item Dev. for Spring 2011 FT	Data Rev. of Fall 2010 FT; Spring 2011 Oper. & Embedded FT	Data Rev. of Spring 2011 FT	Biennial Core-to-Core Overlap; Deferred Due to 2012 Hiatus	Biennial Core-to-Core Overlap (2011); core included as a portion of the 2014 core)				
2011–2012	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins	No Oper. Admins
2012–2013			New Item Dev.for Spring 2013 FT	Winter 2012/13 Admin	Spring 2013 Oper. & Embedded FT; Data Review of Spring 2013 FT; Summer 2013 Admin.		Biennial Core-to-Core Overlap (2013 core included as a portion of the 2015 core)			
2013–2014					Winter 2013/14 Admin; New Item Dev.for Spring 2014 FT	Spring 2014 Oper. & Embedded FT; Data Review of Spring 2014 FT; Summer 2014 Admin		Biennial Core-to-Core Overlap (2014 core included as a portion of the 2015 core)		

Table 3–1. General Development and Usage Cycle of the Algebra I, Biology, and Literature Keystone Exams

Admin Year	Events Occuring in 2009	Events Occuring in 2010	Events Occuring in 2011	Events Occuring in 2012	Events Occuring in 2013	Events Occuring in 2014	Events Occuring in 2015	Events Occuring in 2016	Events Occuring in 2017	Events Occuring in 2018*
2014–2015						Winter 2014/15 Admin; New Item Dev. for Spring 2015 FT	Spring 2015 Oper. & Embedded FT; Data Review of Spring 2015 FT; Summer 2015 Admin		Biennial Core-to-Core Overlap (2014 core included as a portion of the 2015 core)	
2015-2016							Winter 2015/16 Admin; New Item Dev. for Spring 2016 FT	Spring 2016 Oper. & Embedded FT; Data Review of Spring 2016 FT; Summer 2016 Admin		Biennial Core-to-Core Overlap (2014 core included as a portion of the 2015 core)
2016-2017								Winter 2016/17 Admin; New Item Dev. for Spring 2017 FT	Spring 2017 Oper. & Embedded FT; Data Review of Spring 2017 FT; Summer 2017 Admin	
2017-2018*									Winter 2017/18 Admin; New Item Dev. for Spring 2018 FT	Spring 2018 Oper. & Embedded FT; Data Review of Spring 2018 FT; Summer 2018 Admin

*Projected/scheduled tasks and activities

GENERAL TEST DEFINITION

The plan for the Keystone Exam was developed through the collaborative efforts of the Pennsylvania Department of Education (PDE) and Data Recognition Corporation (DRC). The exams are presented online or in two printed testing materials, a test book and a separate answer book. The test book contains multiple-choice (MC) items. The answer book contains scannable pages for multiple-choice responses, constructed-response (CR) items with response spaces, and demographic data collection areas. All MC items are worth 1 point. Algebra I CR items receive a maximum of 4 points (on a scale of 0–4), and all Biology and Literature CR items receive a maximum of 3 points (on a scale of 0–3). In spring 2017, each test form contained operational (core) items (identical on all forms) along with embedded field test items.

CORE-TO-CORE OVERLAP ITEMS

The operational items consist of a set of core items taken by all students. Starting in 2014 these core items included core-to-core overlapping items, which are items that also appeared on the core form of the administration two years before. The overlap connects the spring and summer administrations of year *x* and the winter administration of year *x*+1, with the year *x*+2 spring and summer and year *x*+3 winter administrations. The first biennial core-to-core overlap from the spring 2011 and winter 2011–2012 core was scheduled to begin with the spring 2013 administration. However, when the program was placed on hiatus during the 2011–2012 school year, the overlap was moved to the spring 2014 administration.

ALGEBRA I TEST DEFINITIONS

The Spring 2017 Algebra I Keystone Exam was composed of 20 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Algebra I for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of MC and CR items in the form

Table 3–2. Algebra I Test Plan (Spring 2017) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	18	3	5	1	23	4
2	18	3	5	1	23	4
Total	36	6	10	2	46	8

Table 3–3. Algebra I Test Plan (Spring 2017) per 20 Operational Forms

Module	Core per 20 Forms MC Items	Core per 20 Forms CR Items	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Total per 20 Forms Core & FT MC Items	Total per 20 Forms Core & FT CR Items
1	18	3	100	20	118	23
2	18	3	100	20	118	23
Total	36	6	200	40	236	46

The operational (core) portions of the Spring 2017, Summer 2017, and the Winter 2016/2017 administrations came from the same sources. Therefore 30% to 50% of the 2016/2017 Winter, Spring and Summer cores overlap with the Spring 15, Summer 15, and Winter 14/15 cores. The remaining core items that appeared on the 2016/2017 forms were field tested on prior administrations. Although each spring administration includes embedded field test items, the summer, winter, and breach forms do not include any embedded field test items due to the lower *n*-counts for these administrations. However, summer, winter, and breach forms still include the same number of items that appear in the spring administration. Instead of field test items, the slots in these exams are filled by placeholder (PH) items. Table 3–4 displays the design for the Algebra I Summer, Winter, and Breach operational forms.

Table 3–4. Algebra I Test Plan (2017 Summer, Winter, and Breach) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	18	3	5	1	23	4	1	3
2	18	3	5	1	23	4	1	3
Total	36	6	10	2	46	8	1	3

Since an individual student's score is based solely on the operational (or core) items, the total number of operational points is 60 for Algebra I. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 4 points each) portions of the exam as follows:

Table 3–5. Algebra I Core Points

Category	Module 1 MC Items	Module 1 CR Items	Module 2 MC Items	Module 2 CR Items	Total MC Items	Total CR Items
Total Points	30 (50%)		30 (50%)		60 (100%)	
Core Items	18	3	18	3	36	6
Core Points	18	12	18	12	36	24

The Algebra I Exam results will be reported in two categories based on the two modules of the Algebra I Exam. The code letters for these Assessment Anchor categories are

1. Operations and Linear Equations & Inequalities
2. Linear Functions and Data Organization

The distribution of Algebra I items into these two categories is shown in the following table.

Table 3–6. Algebra I Module and Anchor Distribution

Algebra I Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	30	50%	3	18
2	30	50%	3	15

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Algebra I operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Algebra I operational exam, see Chapter Seven.

BIOLOGY TEST DEFINITIONS

The Spring 2017 Biology Keystone Exam was composed of 20 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Biology for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of MC and CR items in the form

Table 3–7. Biology Test Plan (Spring 2017) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	24	3	8	1	32	4
2	24	3	8	1	32	4
Total	48	6	16	2	64	8

Table 3–8. Biology Test Plan (Spring 2017) per 20 Operational Forms

Module	Core per 20 Forms MC Items	Core per 20 Forms CR Items	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Total per 20 Forms Core & FT MC Items	Total per 20 Forms Core & FT CR Items
1	24	3	160	20	184	23
2	24	3	160	20	184	23
Total	48	6	320	40	368	46

The operational (core) portions of the Spring 2017, Summer 2017, and the Winter 2016/2017 administrations came from the same sources. Therefore 30% to 50% of the 2016/2017 Winter, Spring and Summer cores overlap with the Spring 15, Summer 15, and Winter 14/15 cores. The remaining core items that appeared on the 2016/2017 forms were field tested on prior administrations. Although each spring administration includes embedded field test items, the summer, winter, and breach forms do not include any embedded field test items due to the lower *n*-counts for these administrations. However, summer, winter, and breach forms still include the same number of items that appear in the spring administration. Instead of field test items, the slots in these exams are filled by placeholder (PH) items. Table 3–9 displays the design for the Biology Summer, Winter, and Breach operational forms.

Table 3–9. Biology Test Plan (2017 Summer, Winter, and Breach) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	24	3	8	1	32	4	1	3
2	24	3	8	1	32	4	1	3
Total	48	6	16	2	64	8	1	3

Since an individual student's score is based solely on the operational (or core) items, the total number of operational points is 66 for Biology. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 3 points each) portions of the exam as follows:

Table 3–10. Biology Core Points

Category	Module 1 MC Items	Module 1 CR Items	Module 2 MC Items	Module 2 CR Items	Total MC Items	Total CR Items
Total Points	33 (50%)		33 (50%)		66 (100%)	
Core Items	24	3	24	3	48	6
Core Points	24	9	24	9	48	18

The Biology Exam results will be reported in two categories based on the two modules of the Biology Exam.

1. Cells and Cell Processes
2. Continuity and Unity of Life

The distribution of Biology items into these two categories is shown in the following table.

Table 3–11. Biology Module and Anchor Distribution

Biology Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	33	50%	4	16
2	33	50%	4	22

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Biology operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Biology operational exam, see Chapter Seven.

LITERATURE TEST DEFINITIONS

The Spring 2017 Literature Keystone Exam was composed of 20 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Literature for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core passages
- Number of unique core MC items
- Number of unique core CR items
- Number of embedded field test passages
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of passages, MC items, and CR items in the form

Table 3–12. Literature Test Plan (Spring 2017) per Operational Form

Module	Core per Form Passages	Core per Form MC Items	Core per Form CR Items	Field Test per Form Passages	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Passages	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	2	17	*3	1	6	1	3	23	4
2	2	17	*3	1	6	1	3	23	4
Total	4	34	6	2	12	2	6	46	8

*For each module, one core passage has two CRs and one core passage has one CR.

Table 3–13. Literature Test Plan (Spring 2017) per 20 Operational Forms

Module	Core per 20 Forms Passages	Core per 20 Forms MC Items	Core per 20 Forms CR Items	Field Test per 20 Forms Passages	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Total per 20 Forms Passages	Total per 20 Forms Core & FT MC Items	Total per 20 Forms Core & FT CR Items
1	2	17	*3	12	120	20	10	137	23
2	2	17	*3	12	120	20	10	137	23
Total	4	34	6	24	240	40	20	274	46

*For each module, one core passage has two CRs and one core passage has one CR.

The operational (core) portions of the Spring 2017, Summer 2017, and the Winter 2016/2017 administrations came from the same sources. Therefore 30% to 50% of the 2016/2017 Winter, Spring and Summer cores overlap with the Spring 15, Summer 15, and Winter 14/15 cores. The remaining core items that appeared on the 2016/2017 forms were field tested on prior administrations. Although each spring administration includes embedded field test items, the summer, winter, and breach forms do not include any embedded field test items due to the lower *n*-counts for these administrations. However, Summer, Winter, and breach forms still include the same number of items that appear in the Spring administration. Instead of field test items, the slots in these exams are filled by placeholder (PH) items. Table 3–14 displays the design for the Literature Summer, Winter, and Breach operational forms.

Table 3–14. Literature Test Plan (2017 Summer, Winter, and Breach) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	2	17	*3	1	6	1	1	3
2	2	17	*3	1	6	1	1	3
Total	4	34	6	2	12	2	1	3

*For each module, one core passage has two CRs and one core passage has one CR.

Since an individual student's score is based solely on the operational (or core) items, the total number of operational points is 52 for Literature. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 3 points each) portions of the exam as follows:

Table 3–15. Literature Core Points

Category	Module 1 Passages	Module 1 MC Items	Module 1 CR Items	Module 2 Passages	Module 2 MC Items	Module 2 CR Items	Total Passages	Total MC Items	Total CR Items
Total Points	26 (50%)			26 (50%)			52 (50%)		
Core Items	2	17	3	2	17	3	4	34	6
Core Points	NA	17	9	NA	17	9	NA	34	18

The Literature Exam results will be reported in two broad categories based on the two modules of the Literature Exam.

1. Fiction Literature
2. Nonfiction Literature

The distribution of Literature items into these two categories is shown in the following table.

Table 3–16. Literature Module and Anchor Distribution

Literature Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	26	50%	2	25
2	26	50%	2	31

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Literature operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Literature operational exam, see Chapter Seven.

ITEM DEVELOPMENT CONSIDERATIONS

Alignment to the Keystone Assessment Anchors and Eligible Content, course-level appropriateness (as specified by PDE), depth of knowledge (DOK), item/task level of complexity, estimated difficulty level, relevancy of context, rationale for distractors, style, accuracy, and correct terminology were major considerations in the item development process. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) and *Universal Design* (Thompson, Johnstone, & Thurlow, 2002) guided the development process. In addition, *Fairness in Testing: Training Manual for Issues of Bias, Fairness, and Sensitivity* (DRC, 2010) was used for developing items. All items were reviewed for fairness by bias and sensitivity committees and for content by Pennsylvania educators and field specialists.

BIAS, FAIRNESS, AND SENSITIVITY OVERVIEW

At every stage of the item and test development process, DRC employs procedures that are designed to ensure that items and tests meet Standard 7.4 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.

To meet Standard 7.4, DRC uses a series of internal quality steps. DRC provides specific training for test developers, item writers, and reviewers on how to write, review, revise, and edit items related to issues of bias, fairness, and sensitivity (as well as based on technical quality). Training also includes an awareness of and sensitivity to issues of cultural diversity. In addition to providing *internal* training in reviewing items in order to eliminate potential bias, DRC also provides *external* training to the review panels of minority experts, teachers, and other stakeholders.

DRC's guidelines for bias, fairness, and sensitivity include instruction concerning how to eliminate language, symbols, words, phrases, and content that might be considered offensive by members of racial, ethnic, gender, or other groups. Areas of bias that are specifically targeted include, but are not limited to, stereotyping, gender, region/geography, ethnic/cultural, socioeconomics/class, religion, experience, and biases against a particular age group (ageism) or persons with disabilities. DRC catalogues topics that should be avoided and maintains balance in gender and ethnic emphasis within the pool of available items and passages.

See the sections below in this chapter for more information about the Bias, Fairness, and Sensitivity Review meetings conducted for the Keystone Exams.

UNIVERSAL DESIGN OVERVIEW

The principles of universal design were incorporated throughout the item development process to allow participation of the widest possible range of students in the Keystone Exams. The following checklist was used as a guideline:

- Items measure what they are intended to measure.
- Items respect the diversity of the assessment population.
- Items have a clear format for text.
- Stimuli and items have clear pictures and graphics.
- Items have concise and readable text.
- Items allow changes to other formats, such as Braille, without changing meaning or difficulty.
- The arrangement of the items on the test has an overall appearance that is clean and well organized.

A more extensive description of the application of the principles of universal design is provided in Chapter Four.

DEPTH-OF-KNOWLEDGE OVERVIEW

An important element in statewide graduation exams is the alignment between the overall assessment system and the state's standards. A methodology developed by Norman Webb (1999, 2006) offers a comprehensive model that can be applied to a wide variety of contexts. With regard to the alignment between standards statements and the assessment instruments, Webb's criteria include five categories, one of which deals with content. Within the content category is a useful set of levels for evaluating DOK. According to Webb (1999), "depth-of-knowledge consistency between standards and assessments indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards" (Webb, 1999, pp. 7–8). The four levels of cognitive complexity (i.e., DOK) are as follows:

- Level 1: Recall
- Level 2: Application of Skill/Concept
- Level 3: Strategic Thinking
- Level 4: Extended Thinking

DOK levels were incorporated into the item writing and review process, and items were coded with respect to the level they represented. The default DOK for the Keystone Exams is Level 3. The DOK level for CR items must be Level 3. The DOK level for MC items must also be Level 3; however, in some specific cases, Level 2 is allowed when the cognitive intent of an Eligible Content is Level 2. DOK Level 1 and DOK Level 4 are not included on the Keystone Exams. For more information on DOK (and a comparison of DOK to Bloom's Taxonomy), see Appendix A.

PASSAGE READABILITY OVERVIEW

Evaluating the readability of a passage is essentially a judgment by individuals familiar with the classroom context and what is linguistically appropriate (PDE recommends that the Literature Keystone Exam be administered at grade 10). Although various readability indices were computed and reviewed, it is recognized that such methods measure different aspects of readability and are often fraught with particular interpretive liabilities. Thus, the commonly available readability formulas were not used in a rigid way, but more informally to provide for several snapshots of a passage that senior test development staff considered, along with experience-based judgments in guiding the passage-selection process. In addition, passages were reviewed by committees of Pennsylvania educators who evaluated each passage for readability and grade-level appropriateness. For more information on Literature passages, see Chapter Two and the literature passage-selection process described below.

TEST ITEM READABILITY OVERVIEW

Careful attention was given to the readability of the items to make certain that the assessment focus of the item did not shift based on the difficulty of reading the item. Subject/course areas such as Algebra I or Biology contain many content-specific vocabulary terms. As a result, readability formulas were not used. However, wherever it was practicable and reasonable, every effort was made to keep the vocabulary at or one level below the course level for non-Literature exams. There was a conscious effort made to ensure that each question was evaluating a student's ability to build toward mastery of the course standards rather than evaluating the student's reading ability. Resources used to verify the vocabulary level were the *EDL Core Vocabularies* and the *Children's Writer's Word Book*.

In addition, every test question is brought before several different committees composed of Pennsylvania educators who are course-level/grade-level experts in the content field in question. They review each question from the perspective of the students they teach, determine the validity of the vocabulary used, and work to minimize the level of reading required.

Vocabulary was also addressed at the Bias, Fairness, and Sensitivity Review, although the focus was on how certain words or phrases may represent possible sources of bias or issues of fairness or sensitivity. See the sections that follow in this chapter for more information about the Bias, Fairness, and Sensitivity Review meetings conducted for the Keystone Exams.

ITEM AND TEST DEVELOPMENT CYCLE

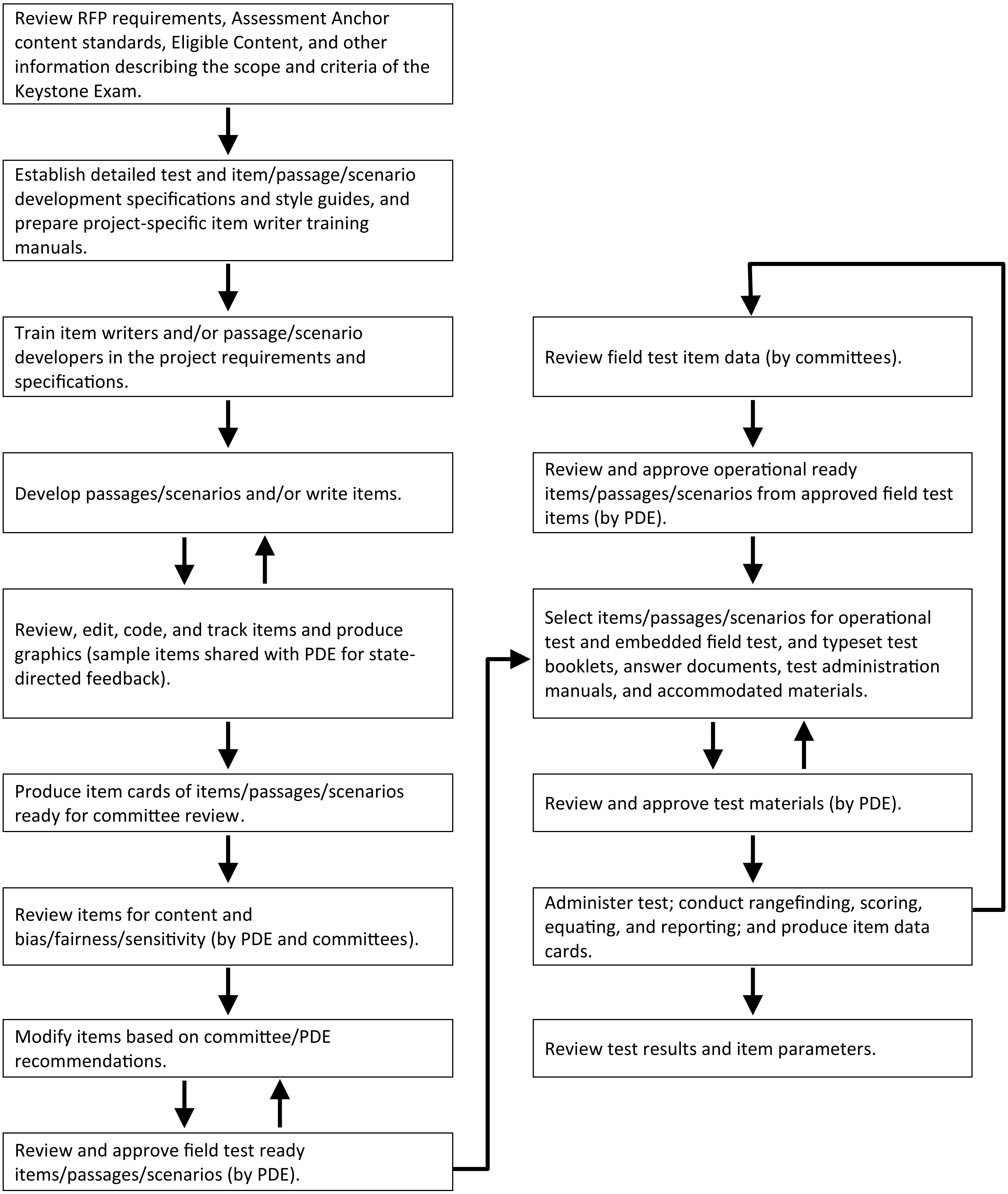
The item development process for items followed a logical cycle and time line, which are outlined in the table and figure on the next pages. On the front end of the schedule, tasks were generally completed with the goal of presenting field test candidate items to committees of Pennsylvania educators. On the back end of the schedule, all tasks led to the field test data review and operational test construction. This process represents a typical life cycle for an embedded Keystone field test event, not a stand-alone field test event or an accelerated development cycle.

The process flowchart, also shown below, illustrates the interrelationship among the steps in the primary cycle that occurs in a normal process of development (i.e., when the items for field testing are primarily from new development, as opposed to being selected from an existing item bank). In addition, a detailed process table describing the item and test development processes also appears in Appendix C.

Table 3–17. General Item and Test Development Life Cycle for Spring Keystone Administrations

Cycle	Steps in Development Life Cycle	Time Line	Approximate Window	
Primary	Development Planning	Summer/Fall	Month 1–4	Jul–Oct
Primary	Literature Passage Selection	Summer/Fall	Month 1–6	Jul–Dec
Primary	Item Writer Training	Fall	Month 5	Nov
Primary	Initial Item Authoring	Fall/Winter	Month 5–9	Nov–Mar
Primary	Internal Reviews and PDE Reviews	Fall to Spring	Month 6–12	Dec–Jun
Primary	Bias, Fairness, and Sensitivity Review	Summer	Month 13	Jul
Primary	New Item Content Committee Review (PA Educators)	Summer	Month 13	Jul
Primary	Post-Review Resolution and Cleanup	Summer	Month 13–14	Jul–Aug
Primary	Build Field Test Forms	Summer/Fall	Month 15–16	Sep–Oct
Primary	Internal Form Reviews and PDE Reviews	Summer/Fall	Month 15–18	Sep–Dec
Primary	Final Form and Printer Proof Approvals	Fall/Winter	Month 18–19	Dec–Jan
Primary	Ancillary and Accommodated Form Development	Fall/Winter	Month 18–20	Dec–Feb
Primary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 19–22	Jan–Apr
Primary	Field Test Administration	Spring	Month 23	May
Primary	Material/Data Processing, Rangefinding, and Scoring	Spring/Summer	Month 23–26	May–Aug
Primary	Field Test Item Data Review (PA Educators)	Summer	Month 27	Sept
Primary	Select Operational Items	Summer/Fall	Month 27–28	Sep–Oct
Primary	Build Operational Forms	Fall	Month 28–29	Oct–Nov
Primary	Internal Form Reviews and PDE Reviews	Fall	Month 29–30	Nov–Dec
Primary	Final Form and Printer Proof Approvals	Fall/Winter	Month 30–31	Dec–Jan
Primary	Ancillary and Accommodated Form Development	Fall/Winter	Month 31–33	Jan–Mar
Primary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 31–33	Jan–Mar
Primary	Operational Test Administration	Spring	Month 35	May
Primary	Material/Data Processing and Scoring	Spring/Summer	Month 35–36	May–Jun
Primary	Score Reporting	Summer	Month 35–39	May–Sep
Secondary	Select Biennial Core-to-Core Overlap Items (Operational Items)	Summer/Fall	Month 51–52	Sep–Oct
Secondary	Build Operational Forms	Fall/Winter	Month 52–53	Oct–Nov
Secondary	Internal Form Reviews and PDE Reviews	Winter	Month 53–54	Nov–Dec
Secondary	Final Form and Printer Proof Approvals	Winter	Month 54–55	Dec–Jan
Secondary	Ancillary and Accommodated Form Development	Winter/Spring	Month 55–57	Jan–Mar
Secondary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 56–58	Feb–Apr
Secondary	Second Operational Test Administration	Spring	Month 59	May
Secondary	Material/Data Processing and Scoring	Spring/Summer	Month 59–60	May–Jun
Secondary	Score Reporting	Summer	Month 59–63	May–Sep
Tertiary	Release Core-to-Core Overlap Items in Samplers	Fall	Month 63	Sep

Figure 3–1. DRC Item and Test Development Primary Cycle



GENERAL ITEM AND TEST DEVELOPMENT PROCESS

This section describes the processes which lead up to an operational exam. These processes were used to develop the entire pool of items that appeared in the field test and operational administrations.

ITEM DEVELOPMENT PLANNING MEETING

Prior to the start of any item development work, DRC's test development staff meets with PDE's assessment office to discuss the test development plans for the next administration, including the test blueprint, the field test plan (including development counts), procedures, time lines, etc. With a complete development cycle lasting about three years (from item authoring through field test, data review, and operational usage), the initial planning begins well in advance of the anticipated administration.

ITEM WRITER TRAINING

Item writers were selected and trained for the subject areas of Algebra I, Biology, and Literature. Qualified writers—either hired independently by the testing vendor, DRC, or through subcontractors like Victory— were college graduates with teaching experience and a demonstrated base of knowledge in the content area. Many of these writers were content assessment specialists and curriculum specialists. The writers were trained individually and had previous experience in writing MC and CR items. Prior to developing items for the Keystone Exams, the cadre of item writers was trained with regard to the following areas:

- Keystone Assessment Anchors and Eligible Content
- Webb's levels of cognitive complexity, DOK
- Subject-specific general scoring guidelines
- Specific and general guidelines for item writing
- Bias, fairness, and sensitivity guidelines
- Principles of universal design
- Item quality technical style guidelines
- Reference information
- Sample items

LITERATURE PASSAGE SELECTION

The task of searching for passages was conducted by DRC professionals with classroom experience in reading/language arts. These professionals also underwent specialized training (provided by DRC) in the characteristics of acceptable passages. Guidelines for passage selection included appropriate length, text structure, density, and vocabulary. A judgment was also made about whether the reading level required by a particular passage was at the independent level—that is, where the average student should be able to read 90 percent of words in the text independently. Passage finders were given the task of searching for a specified number of passages for each genre. Generally, they looked for at least twice as many passages as were needed. Passages acquired were either authentic (permissioned), in that they were culled from published materials, or commissioned by experienced authors. See Chapter Two for more information on the types of passages used on the Literature Keystone Exam.

For permissioned passages, approval to reprint was secured from the publishers. Passages underwent an internal review by several test development content editors to judge their merit with regard to the following criteria:

- Passages have interest value for students.
- Passages are appropriate in terms of vocabulary and language characteristics.
- Passages are free of bias, fairness, and sensitivity issues.
- Passages represent different cultures.
- Passages are from a variety of sources.

- Passages are able to stand the test of time.
- Passages are sufficiently rich to generate a variety of MC and CR items.
- Passages are complete with all necessary permissions documentation.
- Passages avoid dated subject matter unless a relevant historical context is provided.
- Passages should not require students to have extensive background knowledge in a certain discipline or area to understand a text.

Once through the internal review process, the passages deemed potentially acceptable were reviewed by the Reading Content Committee and the Bias, Fairness, and Sensitivity Committee for final approval.

ITEM AUTHORING AND TRACKING

Initially, items are generated with software-prepared Keystone Exams Item Cards, which allow for preliminary sorting and reviewing. Although very similar, the Keystone Exams Item Card for Multiple-Choice Items differs from the Keystone Exams Item Card for Constructed-Response Items in that the former has a location at the bottom of the card for comments regarding the distractors. Blank examples of these two cards are shown in Appendix D. In both instances, a column against the right margin includes codes to identify the subject area, grade, content categories, passage information (in the case of reading), item type, DOK (cognitive complexity), estimated difficulty, answer key (for MC items), and calculator use (for mathematics items).

All items undergoing field testing were entered into the DRC Item Development and Educational Assessment System (IDEAS), which is a comprehensive, secure, online item banking system. It accommodates item writing, item viewing and reviewing, and item tracking and versioning. IDEAS manages the transition of an item from its developmental stage to its approval for use in a test form (for both print and online delivery). The system supports item history records that include item usage within a form, item-level notes, content categories and subcategories, item statistics from both classical and Rasch item analyses, and classifications derived from analyses of differential item functioning (DIF). A sample IDEAS Item Card is presented in Appendix D.

INTERNAL REVIEWS AND PDE REVIEWS

To ensure that the items produced were sufficient in number and adequately distributed across subcategories and levels of difficulty, item writers were informed of the required quantities of items. As items were written, an item authoring card was completed. It contained information about the item, such as subject, content category, and subcategories. Based on the item writer's classroom teaching experience, his/her knowledge of the content area curriculum, and the cognitive demands required by the item, estimates were recorded for level of cognitive complexity and difficulty level. Items were written to provide for a range of difficulty and for cognitive complexity focused on DOK Level 3.

As part of the item construction process, each item was reviewed by content specialists and editors at DRC. Content specialists and editors evaluated each item to make sure that it measured the intended Eligible Content and/or Assessment Anchor. They also assessed each item to make certain that it was appropriate for the intended grade and that it provided and cued only one correct answer (MC items only). In addition, the difficulty level, DOK, graphics, language demand, and distractors were also evaluated. Other elements considered in this process included, but were not limited to, universal design, bias, source of challenge, grammar/punctuation, and Keystone style.

Following this internal process, items were reviewed by content specialists at PDE, who then consulted with DRC about any general issues or concerns (e.g., style, format, interpretation of Assessment Anchors and Eligible Content) and about edits to specific items. Following PDE's review, the items were prepared for the content review meetings conducted with Pennsylvania educators.

ITEM CONTENT REVIEWS IN AUGUST 2016

Prior to the 2017 field testing, all newly developed test items were submitted to content committees for review. The content committees consisted of Pennsylvania educators from school districts throughout the Commonwealth of Pennsylvania, some with postsecondary university affiliations. The primary responsibility of the content committee was to evaluate items with regard to quality and content classification, including grade-level (course) appropriateness, estimated difficulty, DOK, and source of challenge. They also suggested revisions and made recommendations for reclassification of items. In some cases when an item was deleted, the committee suggested a replacement item and/or reviewed a suggested replacement item provided by the facilitators. The committee also reviewed the items for adherence to the principles of universal design, including language demand and issues of bias, fairness, and sensitivity.

With source of challenge, items were identified where the cognitive demand was focused on an unintended content, concept, or skill (Webb, 2002). Source of challenge may be a contributing factor if the reason that an answer could be given results from a cultural bias, an inappropriate reading level, a flawed graphic in an item, or if an item requires specialized, noncontent-related knowledge to answer. Source of challenge could result in a student who has mastered the intended content or skill answering the item incorrectly or a student who has not mastered the intended content or skill answering the item correctly. Committee members were asked to note any items with a source of challenge and to suggest revisions to remove the source of challenge.

The content review meetings were held on August 1–5, 2016, for Algebra I, Biology, and Literature. Committee members were approved by PDE, and PDE-approved invitations were sent to them by DRC. PDE also selected internal staff members for attendance. The meeting commenced with a welcome by PDE and DRC. This was followed by an overview of the test development process by DRC. PDE, along with DRC, also provided training on the procedures and forms to be used for item content review.

DRC content assessment specialists facilitated the reviews and were assisted by representatives of PDE. Committee members, grouped by exam, worked through and reviewed the items for quality and content, as well as for the following categories:

- Assessment Anchor alignment
- Content limits
- Grade-level (course-level) appropriateness
- Difficulty level
- DOK
- Appropriate source of challenge
- Correct answer
- Quality of distractors
- Graphics in regards to appropriateness
- Appropriate language demand
- Freedom from bias

The members then came to consensus and assigned a status to each item: Approved, Accepted with Revision, or Rejected. All comments were recorded, and a master rating sheet was completed. Committee facilitators recorded the committee consensus on the Item Review Rating Sheet. A sample form and rating criteria may be found in Appendix E.

Security was addressed by adhering to a strict set of procedures. Items in binders were distributed for committee review by number and signed for by each member on a daily basis. All attendees, with the exception of PDE staff, were required to sign a confidentiality agreement. All materials not in use at any time were stored in a locked room. Secure materials that did not need to be retained after the meetings were deposited in secure barrels, the contents of which were shredded.

BIAS, FAIRNESS, AND SENSITIVITY REVIEWS IN AUGUST 2016

Prior to the 2016 field testing, all newly developed test items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. This review took place from August 8–11, 2016, for Algebra I, Biology, and Literature. The committee's primary responsibility was to evaluate items with regard to bias, fairness, and sensitivity issues. They also made recommendations for changes or deletion of items in order to remove the potential for issues of bias, fairness, and/or sensitivity. Included in the review were proposed reading passages. An expert, multiethnic committee composed of men and women was trained by a DRC test development lead to review items for bias, fairness, and sensitivity issues. Training materials included a manual developed by DRC (DRC, 2016). Members of the committee also had expertise with special needs students and English Language Learners (ELL). PDE staff members were also trained and participated in the review. All items were read by a cross-section of committee members. Each member noted bias, fairness, and/or sensitivity comments on tracking sheets and on the item, if needed, for clarification. Committee members individually categorized any concerns as related to ageism, disability, ethnicity/culture, gender, region, religion, socioeconomics, or stereotypes. These categories were the framework through which recommendations for modification or rejection of items occurred during the subsequent committee consensus process. The committee discussed each of the issues as a group and came to consensus as to which decisions should represent the view of the committee. All consensus comments were then compiled, and the suggested actions on these items were recorded and submitted to PDE. This review followed the same security procedures as outlined above. The table below shows the gender and race/ethnicity of the members of the bias committee who reviewed the Keystone items and passages for bias, fairness, and sensitivity.

Table 3–18. Demographic Composition of the 2016 Keystone Bias, Fairness, and Sensitivity Committee

The results from the 2016 Bias, Fairness, and Sensitivity Committee reviews are summarized in the next table.

Member #	Gender	Race/Ethnicity
1.	Female	Asian American
2.	Female	Asian American
3.	Male	Asian American
4.	Female	Native American
5.	Female	Caucasian American
6.	Female	Caucasian American
7.	Male	Caucasian American
8.	Male	Caucasian American
9.	Male	African American
10.	Female	African American
11.	Female	African American
12.	Female	Hispanic American
Totals	8 Females, 4 Males	1 Hispanic American, 3 Asian Americans, 4 Caucasian Americans, 1 Native American, 3 African Americans

Table 3–19A. Number of Items—Bias, Fairness, and Sensitivity Committee Review: Algebra I

Date	Total Reviewed	Accepted As Is	Accepted with Revision	Rejected
August 2016	270 Items	251 Items	19 Items	0 Items

Table 3–19B. Number of Items—Bias, Fairness, and Sensitivity Committee Review: Biology

Date	Total Reviewed	Accepted As Is	Accepted with Revision	Rejected
August 2016	20 Scenarios, 540 Items	20 Scenarios, 540 Items	0 Scenarios, 0 Items	0 Scenarios, 0 Items

Table 3–19L. Number of Items—Bias, Fairness, and Sensitivity Committee Review: Literature

Date	Total Reviewed	Accepted As Is	Accepted with Revision	Rejected
August 2016	22 Passages, 329 Items	22 Passages, 329 Items	0 Passages, 0 Items	0 Passages, 0 Items

Table 3–19T. Number of Items—Bias, Fairness, and Sensitivity Committee Review: Total

Date	Total Reviewed	Accepted As Is	Accepted with Revision	Rejected
August 2016	20 Scenarios, 22 Passages, 1,139 Items	20 Scenarios, 22 Passages, 1,120 Items	0 Scenarios, 0 Passages, 19 Items	0 Scenarios, 0 Passages, 0 Items

CHAPTER FOUR: UNIVERSAL DESIGN PROCEDURES APPLIED TO THE KEYSTONE EXAMS TEST DEVELOPMENT PROCESS

UNIVERSAL DESIGN

Universally designed assessments allow participation of the widest possible range of students and contribute to valid inferences about participating students. Principles of Universal Design are based on the premise that each child in school is a part of the population to be tested and that testing results should not be affected by disability, gender, race, or English language ability (Thompson, Johnstone, & Thurlow, 2002). At every stage of the item and test development process, procedures were employed to ensure that items and subsequent tests (in both print and online delivery methods) were designed and developed using the elements of universally designed assessments established by the National Center on Educational Outcomes (NCEO).

Federal legislation addresses the need for universally designed assessments. The No Child Left Behind Act (Elementary and Secondary Education Act) requires that each state must "provide for the participation in [statewide] assessments of all students" [Section 1111(b)(3)(C)(ix)(I)]. Both Title I and IDEA regulations call for universally designed assessments that are accessible and valid for all students, including English Language Learners and students with disabilities. The benefits of universally designed assessments not only apply to these groups of students, but to all individuals with wide-ranging characteristics.

DRC's test development team was trained in the elements of Universal Design as they relate to developing large-scale statewide assessments. Team leaders were trained directly by NCEO, and other team members were subsequently trained by team leaders. Committees involved in content review included some members who were familiar with the unique needs of students with disabilities and English Language Learners. Likewise, some members of the Bias, Fairness, and Sensitivity Committee were conversant with these issues. What follows are the Universal Design guidelines that were followed during all stages of the item development process for the Keystone Exams.

ELEMENTS OF UNIVERSALLY DESIGNED ASSESSMENTS

After a review of research relevant to the assessment development process and the Principles of Universal Design (Connell et al., 1997), NCEO has produced seven elements of Universal Design as they apply to assessments (Thompson, Johnstone, & Thurlow, 2002). These elements served to guide item development for the Keystone Exams.

- **Inclusive Assessment Population**

  The target population includes students attending Commonwealth schools who participate in one or more of the graduation competency exams.

- **Precisely Defined Constructs**

  An important function of well-designed assessments is that the assessments actually measure what they are intended to measure. The Keystone Exams Assessment Anchor Content Standards (Assessment Anchors) provided clear descriptions of the constructs to be measured on each of the exams. Universally designed assessments must remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers.

- **Accessible, Nonbiased Items**

  DRC conducted both internal and external reviews of items and test specifications to ensure that they did not create barriers due to lack of sensitivity to disability, culture, or other subgroups. Items and test specifications were developed by a team who understood the varied characteristics of items that might create difficulties for any group of students. Accessibility is incorporated as a primary dimension of test specifications, so accessibility was woven into the fabric of the test rather than being added after the fact.

- **Amenable to Accommodations**

  Even though items on universally designed assessments are accessible for most students, there are some students who continue to need accommodations. This essential element of a universally designed assessment requires that the exam is compatible with accommodations and a variety of widely used adaptive equipment and assistive technology (see also the section on Assessment Accommodations later in this chapter).

- **Simple, Clear, and Intuitive Instructions and Procedures**

  Assessment instructions should be easy to understand regardless of a student's experience, knowledge, language skills, or current concentration level. Questions that are posed using complex language can invalidate the test if students cannot understand how they are expected to respond to a question. To meet this guideline, directions and questions were prepared in simple, clear, and understandable language that underwent multiple reviews.

- **Maximum Readability and Comprehensibility**

  A variety of guidelines exist to ensure the maximum readability and comprehensibility of a test. These features go beyond what is measured by readability formulas. Readability and comprehensibility are affected by many factors, including student background, sentence difficulty, and text organization. All of these features were considered as item text was developed.

  Plain language is a concept now being highlighted in research on assessments. Plain language has been defined as language that is straightforward and concise. The following strategies for editing text to produce plain language were used during the editing process of the new Keystone Exam items:

  - Reduction of excessive length
  - Use of common words
  - Avoidance of ambiguous words
  - Avoidance of irregularly spelled words
  - Avoidance of proper names
  - Avoidance of inconsistent naming and graphic conventions
  - Avoidance of unclear signals about directing attention

- **Maximum Legibility**

  Legibility is the physical appearance of text, the way that the shapes of letters and numbers enable people to read text easily. Bias can result when tests contain physical features that interfere with a student's focus on or understanding of the constructs that test items are assessing. A style guide was developed and utilized that included dimensions of style consistent with Universal Design.

GUIDELINES FOR UNIVERSALLY DESIGNED ITEMS

All test items written and reviewed adhered closely to the following guidelines for Universal Design. Item writers and reviewers used a checklist during the item development process to ensure that each aspect was followed. For more information on the checklist, see the Universal Design Overview section in Chapter Three of this report.

1. **Items measure what they are intended to measure.** Item writing training included making certain that writers and reviewers had a clear understanding of Pennsylvania's Academic Standards and the Keystone Assessment Anchors. During all phases of test development, items were presented with content-standard information to ensure that each item reflected the intended Assessment Anchor. Careful consideration of the content standards was important in determining which skills involved in responding to an item were extraneous and which were relevant. With certain types of items, an additional skill was necessary, such as the Algebra I test, which requires the student to read.
2. **Items respect the diversity of the assessment population.** To develop items that avoid content that might unfairly advantage or disadvantage any student subgroup, item writers, test developers, and reviewers were trained to write and review items to avoid issues of bias, fairness, and sensitivity. Training also included an awareness of and sensitivity to issues of cultural and regional diversity.

3. **Items have a clear format for text.** Decisions about how items were presented to students must allow for maximum readability for all students. Appropriate fonts and point sizes were employed with minimal use of italics, which are far less legible and are read considerably more slowly than standard typeface. Captions, keys, and legends were at least a 12-point size, while footnotes and sentence numbers used a 10-point font.[1] Legibility was enhanced by sufficient spacing between letters, words, and lines. Blank space was used around paragraphs and between columns and staggered right margins.
4. **Stimuli and items have clear pictures and graphics.** When pictures and graphics were used, they were designed to provide essential information in a clear and uncluttered manner. Illustrations were placed directly next to the information to which they referred, and labels were used when possible. Sufficient contrast between the background and text, with minimal use of shading, increased readability for students with visual impairments. Color was not used to convey important information.
5. **Items have concise and readable text.** Linguistic demands of stimuli and items can interfere with a student's ability to demonstrate knowledge of the construct being assessed. During item writing and review, the following guidelines were used:
    - Simple, clear, commonly used words were used whenever possible.
    - Extraneous text was omitted.
    - Vocabulary and sentence complexity were appropriate for the grade level being assessed.
    - Technical terms and abbreviations were used only if they were related to the content being measured.
    - Definitions and examples were clear and understandable.
    - Idioms were avoided unless idiomatic speech was being assessed.
    - Questions to be answered were clearly identifiable.
6. **Items allow changes to format without changing meaning or difficulty.** A Braille version was available for each operational exam. Attention was given to using items that allow for Braille. Specific accommodations were permitted, such as signing to a student, the use of oral presentation under specified conditions, and the use of various assistive technologies. A Spanish version of the Algebra I and Biology exams was available for use by English Language Learners who would benefit from this accommodation and who were in U.S. schools for less than three years.
7. **The test has an overall appearance that is clean and organized.** Information was organized in a left-right, top-bottom format. Images, pictures, and text that may not be necessary (e.g., sidebars, overlays, callout boxes, shading, visual crowding caused by excess information) and that could be potentially distracting to students were avoided. Also avoided were purely decorative features that did not serve a purpose.

ITEM DEVELOPMENT

DRC worked closely with the Pennsylvania Department of Education to ensure that the Keystone Exams complied with nationally recognized principles of Universal Design. The implementation of accommodations on large-scale statewide assessments for students with disabilities was supported in the development of the Keystone Exams. In addition to the principles of Universal Design as described in the Pennsylvania Technical Report, DRC applied to each exam the standards for test accessibility as described in *Tests Access: Making Tests Accessible for Students with Visual Impairments—A Guide for Test Publishers, Test Developers, and State Assessment Personnel* (Allman, 2004).

[1] While font size follows specific requirements during online setup of an exam, the screen resolution used at the local level can impact whether the effective font size is visible to the student.

To this end, DRC embraces the following precepts:

- Test directions are worded to allow for alternate responses to constructed-response items.
- During item and bias reviews, committee members are made aware of the Principles of Universal Design and of issues that may adversely affect students with disabilities. The goal is to make certain that the Keystone Exams are bias free for all students. With the goal of ensuring that the Keystone Exams are accessible to the widest range of diverse student populations, PDE instructs DRC to limit item types that are difficult to format in Braille and that may become distorted when published in large print. DRC is instructed to limit the following on the Keystone Exams.
  - Algebra I: Complicated tessellations; charts or graphs that extend beyond one page
  - Literature: Graphics and illustrations that are not germane to the content presented
  - All exams: Unnecessary boxes and framing of text, unless enclosing the text provides necessary context for the student; use of italics (limited to only when it is absolutely necessary, such as with variables)

ITEM FORMAT

For all Keystone Exams (both online and print), DRC formats the items to maximize accessibility for all students by using text that is in an easily readable size and font style. DRC limits shading, graphics, charts, and the number of items per page so that there is sufficient white space on each page. Whenever possible, DRC ensures that graphics, pictures, diagrams, charts, and tables are positioned on the page with the associated test items. DRC uses high contrast for text and background when possible to convey pertinent information. Tests are published on dull-finish paper to avoid the glare encountered on glossy paper. DRC pays close attention to the binding of the exam books to ensure that they lie flat for two-page viewing and ease of reading and handling.

DRC ensures consistency across Keystone Exams by following these Principles of Universal Design:

- High contrast and clarity is used to convey detailed information.
- Typically, shading is avoided; when necessary for content purposes, 10-percent screens are used as the standard.
- Overlaid print on diagrams, charts, and graphs is avoided.
- Charts, graphs, diagrams, and tables are clearly labeled with titles and with short descriptions when applicable.
- Only relevant information is included in diagrams, pictures, and graphics.
- Symbols used in keys and legends are meaningful and provide reasonable representations of the topics they depict.
- Pictures that require physical measurement are true to size.

ASSESSMENT ACCOMMODATIONS

While universally designed assessments provide for participation of the widest range of students, many students require accommodations in order to participate in the regular assessment. Clearly, the intent of providing accommodations for students is to make certain that students are not unfairly disadvantaged during testing and that the accommodations used during instruction, if appropriate, are made available as students take the test. The literature related to assessment accommodations is still evolving and often focuses on state policies regulating accommodations rather than on providing empirical data that supports the reliability and validity of the use of accommodations. On a yearly basis, the Pennsylvania Department of Education examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. Three accommodations manuals for Pennsylvania assessments titled *Accommodations Guidelines for Students with IEPs and Students with 504 Plans, Accommodations Guidelines for Students without IEPs and 504 Plans*, and *Accommodations Guidelines for English Language Learners* were developed for use with the Keystone Exams. The manuals can be accessed by going to www.pdesas.org/Assessment/Keystone and selecting the corresponding document under the Test Accommodations section.

In addition, Spanish-language versions, translated from the original English versions were made available for both the Algebra I Exam and the Biology Exam. The Spanish-translation editions of the exams are discussed in Chapter Six.

CHAPTER FIVE: FIELD TEST LEADING TO THE SPRING 2017 CORE

FIELD TEST OVERVIEW

Approximately 50% of the core items appearing on the Spring 2017 Pennsylvania Keystone Exams came from the field test (nonlinking) items on the Spring 2016 Keystone Exams. The remaining core items were part of the biennial core-to-core overlap. For more information about the core-to-core overlapping items, please see Chapter Three. The purpose of administering field test items is to obtain statistics for them so they can be reviewed before becoming operational (core). Based on this statistical review, many of the field test items appearing in the Spring 2016 Keystone Exams were selected for use as common (core) items in the 2017 Keystone Exams.

As shown in Table 5–1, the overall Keystone Exams Field Test Plan uses a given spring administration to augment the pool of items available for use in core positions in the subsequent administration cycle starting with the next spring administration. Each spring field test is designed to yield up to three cores worth of items to fill the core administrations of the subsequent spring, summer, and winter cores.

Table 5–1. General Pattern Showing Path from Field Test to Core and to Core-to-Core Overlap for a Given Keystone Exam

	Initial Core Use Spring	Initial Core Use Summer	Initial Core Use Winter	Core-to-Core Overlap Use Spring	Core-to-Core Overlap Use Summer	Core-to-Core Overlap Use Winter
Spring Year X	X+1	X+1	X+2	X+3	X+3	X+4
Spring Year X+1	X+2	X+2	X+3	X+4	X+4	X+5
Spring Year X+2	X+3	X+3	X+4	X+5	X+5	X+6

See Chapter Three (concerning the test definitions for the individual Keystone Exams) for more details about the number of embedded field test (EFT) items appearing in a Keystone Exam within a typical operational form setting.

SPRING 2016 KEYSTONE EXAMS EMBEDDED FIELD TEST

For 2016, the embedded field test (in spring) was designed to yield enough items to construct portions of the following operational forms: spring 2017, summer 2017, winter 2017/2018, and a possible breach form for 2017/2018. The next tables describe the embedded field test plans for the Keystone Exams in the spring of 2016.

SPRING 2016 ALGEBRA I KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2016 Algebra I Keystone Exam was composed of 20 forms. All of the forms contained core items that were identical for all students and sets of generally unique items. The following two tables display the design for Algebra I for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of MC and CR items in the form

Table 5–2. Algebra I Test Plan (Spring 2016) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	18	3	5	1	23	4
2	18	3	5	1	23	4
Total	36	6	10	2	46	8

Table 5–3. Algebra I Test Plan (Spring 2016) per 20 Operational Forms

Module	Core per 20 Forms MC Items	Core per 20 Forms CR Items	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Field Test per 20 Forms Core & FT MC Items	Total per 20 Forms Core & FT CR Items
1	18	3	100	20	118	23
2	18	3	100	20	118	23
Total	36	6	200	40	236	46

SPRING 2016 BIOLOGY KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2016 Biology Keystone Exam was composed of 20 forms. All of the forms contained core items that were identical for all students and sets of generally unique items. The following two tables display the design for Biology for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of MC and CR items in the form

Table 5–4. Biology Test Plan (Spring 2016) per Operational Form

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	24	3	8	1	32	4
2	24	3	8	1	32	4
Total	48	6	16	2	64	8

Table 5–5. Biology Test Plan (Spring 2016) per 20 Operational Forms

Module	Core per 20 Forms MC Items	Core per 20 Forms CR Items	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Field Test per 20 Forms Core & FT MC Items	Total per 20 Forms Core & FT CR Items
1	24	3	160	20	184	23
2	24	3	160	20	184	23
Total	48	6	320	40	368	46

SPRING 2016 LITERATURE KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2016 Literature Keystone Exam was composed of 20 forms. All of the forms contained common items that were identical for all students and sets of generally unique items. The following two tables display the design for Literature for forms 1 through 20. The column entries for these tables denote the following:

- Number of unique core passages
- Number of unique core MC items
- Number of unique core CR items
- Number of embedded field test passages
- Number of embedded MC field test items
- Number of embedded CR field test items
- Total number of passages, MC items, and CR items in the form

Table 5–6. Literature Test Plan (Spring 2016) per Operational Form

Module	Core Per Form Passages	Core Per Form MC Items	Core Per Form CR Items	Field Test per Form Passages	Field Test per Form MC Items	FIeld Test Per Form CR Items	Total per Form Core & FT Passages	Total per Form Core & FT MC Items	Total per Form Core & FT CR ITEMS
1	2	17	3	1	6	1	3	23	4
2	2	17	3	1	6	1	3	23	4
Total	4	34	6	2	12	2	6	46	8

Table 5–7. Literature Test Plan (Spring 2016) per 20 Operational Forms

Module	Core Per 20 Forms Passages	Core Per 20 Forms MC Items	Core Per 20 Forms CR Items	Field Test per 20 Forms Passages	Field Test per 20 Forms MC Items	Field Test per 20 Forms CR Items	Total per 20 Forms Core & FT Passages	Total per 20 Form Core & FT MC Items	Total per 20 Form Core & FT CR ITEMS
1	2	17	3	10	120	20	14	137	23
2	2	17	3	10	120	20	14	137	23
Total	4	34	6	20	240	40	28	274	46

STATISTICAL ANALYSES AND RESULTS

All field test items were analyzed statistically following conventional item analysis methods. For MC items, traditional, or classical, item statistics included the corrected point-biserial correlation (Pt. Bis.) for the correct and incorrect responses (distractors), the percent correct (*p*value), and the percentage selecting incorrect responses. For CR items, the statistical indices included the item-total test correlation, the point-biserial correlation for each score category or level, the percentage in each score category, and the percentage of nonscorable responses.

In general, more-capable students are expected to respond correctly to easy items and less-capable students are expected to respond incorrectly to difficult items. If either of these situations does not occur, the item will be reviewed by DRC test development staff and committees of Pennsylvania educators to determine the nature of the problem and the characteristics of the students affected. The primary way of detecting such conditions is through the point-biserial correlation coefficient for MC items and the item-total correlation for CR items. In each case the statistic will be positive if the total-test mean score is higher for the students who respond correctly to MC items (or attain a higher CR item score) and negative when the reverse is true. A detailed explanation of item statistics based on the classical test theory (CTT) can be found in Chapter Eleven.

DIFFERENTIAL ITEM FUNCTIONING

The differential item functioning (DIF) analysis was also conducted on all the field test items. Differential item functioning occurs when examinees with the same ability level but different group memberships do not have the same probability of answering an item correctly. This pattern of results may suggest the presence of item bias. As a statistical concept, however, DIF can be differentiated from item bias, which is a content issue that can arise when an item presents negative group stereotypes, uses language that is more familiar to one subpopulation than to another, or is presented in a format that disadvantages certain learning styles. While the source of item bias is often plain to trained judges, DIF may have no clear cause. However, studying how DIF arises and how it presents itself can help to detect and correct for it.

DIF DETECTION PROCEDURES

For MC items, the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959) for detecting DIF is a commonly used technique in educational testing. It does not depend on the application or the fit of any specific measurement model. However, it does have significant philosophical overlap with the Rasch model since it uses a test's total score to organize the analysis.

The procedure as implemented by DRC contrasts a focal group with a reference group. While it makes no practical difference in the analysis which group is defined as the focal group, the group most apt to be disadvantaged by a biased measurement is typically defined as the focal group. In these analyses, the focal group was female for gender-based DIF, black and Hispanic for ethnicity-based DIF, and computer-based-test (CBT) group for the test administration mode-based DIF; reference groups were male, white, and paper-and-pencil test (PPT) group respectively. The MH statistic for each item is computed from a contingency table. It has two groups (focal and reference) and two outcomes (right or wrong). The ability groups are defined by the test's score distribution for the total examinee populations.

The basic MH statistic is a single degree of freedom chi-square that compares the observed number in each cell to the expected number. The expected counts are computed to ensure that the analysis is not confounded with differences in the achievement level of the two groups.

For CR items, a comparable statistic is computed based on the standardized mean difference (SMD) (Dorans, Schmitt, & Bleistein, 1992), computed as the differences in mean scores for the focal and reference groups if both groups had the same score distribution.

To assist the review committees in interpreting the analyses, the items are assigned a severity code based on the magnitude of the MH statistic (Zwick & Erickan, 1989) and the effect size for the SMD. Items classified as A+ or A- have little or no statistical indication of DIF. Items classified as B+ or B- have some indication of DIF but are acceptable for future use. Items classified as C+ or C- have strong evidence of DIF and should be reviewed and possibly rejected from the eligible item pool. The plus sign indicates that the item favors the focal group and a minus sign indicates that the item favors the reference group.

LIMITATIONS OF STATISTICAL DETECTION

No statistical procedure should be used as a substitute for rigorous, hands-on reviews by content and bias specialists. The statistical results can help organize the review so the effort is concentrated on the most problematic cases. Further, no items should be automatically rejected simply because a statistical method flagged them or accepted because they were not flagged.

Statistical detection of DIF is an inexact science. There have been a variety of methods proposed for detecting DIF, but no one statistic can be considered either necessary or sufficient. Different methods are more or less successful depending on the situation. No analysis can guarantee that a test is free of bias, but almost any thoughtful analysis will uncover the most flagrant problems.

A fundamental shortcoming of all statistical methods used in DIF evaluation is that all are intrinsic to the test being evaluated. If a test is unbiased overall but contains one or two DIF items, any method will locate the problems. If, however, all items on the test show consistent DIF to the disadvantage of a given subpopulation, a statistical analysis of the items will not be able to separate DIF effects from true differences in achievement.

CRITERIA USED TO FLAG ITEMS

Item statistics are used as a means of detecting items that deserve closer scrutiny rather than as a mechanism for automatic retention or rejection. Toward this end, a set of criteria was used as a screening tool to identify items needing a closer review by committees of Pennsylvania educators.

For all the items, the following criteria were used to flag items:

1. *p*-value less than 0.3 or greater than 0.9
2. Item-total correlation less than 0.25
3. Gender (male vs. female), ethnicity (white vs. black or Hispanic), and/or test administration mode (PPT vs. CBT) DIF code of C+ or C-

For an MC item to be flagged, the following two additional criteria were also used:

4. Point-biserial correlation for any incorrect response greater than 0.0
5. Percentage responding to any incorrect responses greater than the percent correct

For a CR item to be flagged, the following additional criterion was used:

6. Score proportion less than 0.05

The intent of the above criteria is to flag everything that should be reviewed. For this purpose, the preference is to over-identify rather than under-identify the outliers. Any of these flags should cause the item to be reviewed by content experts, but there are many reasons the experts might want to keep an item in spite of the statistics.

RESULTS AND OBSERVATIONS

Details of the samples used for the spring 2016 field test item analysis can be found in Chapter 9 of the *2016 Pennsylvania Keystone Exams Technical Report* (Pennsylvania Department of Education, 2016). Overall, the samples used to analyze the field test items embedded in 20 forms were equivalent, so the classical statistics for all the field test items across forms can be compared.

This section focuses on reporting the number (*N*) and percentage (%) of items flagged by different criteria (see Tables 5–8 to 5–10). For the DIF analysis, the number and percentage of items were provided not only for the C- and C+ bias codes, which were used as the criteria to flag items, but also for the bias codes A-, A+, B-, and B+.

Table 5–8. Summary of Items Flagged by the CTT-Based Statistics

Item Type	Flagging Criterion*	Alg. I Total N	Alg. I N	Alg. I %	Bio. Total N	Bio. N	Bio. %	Lit. Total N	Lit. N	Lit. %
MC	1	240	43	17.9	384	17	4.4	288	13	4.5
MC	2	240	53	22.1	384	73	19.0	288	56	19.4
MC	4	240	56	23.3	384	70	18.2	288	40	13.9
MC	5	240	54	22.5	384	37	9.6	288	18	6.3
CR	6	47**	38	80.9	47**	9	19.1	48	0	0.0

* See section Criteria Used to Flag Items for what 1–6 stands for.
** One Algebra I and one Biology CR items were not scored.

Table 5–9. DIF Summary — MC Items

Reference Group	Focal Group	Bias Code	Alg. I (Total N=240) N	Alg. I (Total N=240) %	Biology (Total N=384) N	Biology (Total N=384) %	Literature (Total N=288) N	Literature (Total N=288) %
Male	Female	A-	113	47.1	236	61.5	131	45.5
Male	Female	A+	115	47.9	138	35.9	144	50.0
Male	Female	B-	9	3.8	8	2.1	6	2.1
Male	Female	B+	3	1.3	1	0.3	6	2.1
Male	Female	C-	0	0.0	1	0.3	0	0.0
Male	Female	C+	0	0.0	0	0.0	1	0.3
White	Black	A-	170	70.8	268	69.8	219	76.0
White	Black	A+	66	27.5	104	27.1	47	16.3
White	Black	B-	3	1.3	12	3.1	21	7.3
White	Black	B+	0	0.0	0	0.0	0	0.0
White	Black	C-	1	0.4	0	0.0	1	0.3
White	Black	C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	163	67.9	272	70.8	201	69.8
White	Hispanic	A+	76	31.7	99	25.8	56	19.4
White	Hispanic	B-	1	0.4	13	3.4	27	9.4
White	Hispanic	B+	0	0.0	0	0.0	0	0.0
White	Hispanic	C-	0	0.0	0	0.0	4	1.4
White	Hispanic	C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	79	32.9	177	46.1	89	30.9
PPT	CBT	A+	155	64.6	206	53.6	191	66.3
PPT	CBT	B-	0	0.0	0	0.0	2	0.7
PPT	CBT	B+	5	2.1	1	0.3	5	1.7
PPT	CBT	C-	0	0.0	0	0.0	0	0.0
PPT	CBT	C+	1	0.4	0	0.0	1	0.3

Table 5–10. DIF Summary — CR Items

Reference Group	Focal Group	Bias Code	Alg. I (Total N=46) N	Alg. I (Total N=46) %	Biology (Total N=48) N	Biology (Total N=48) %	Literature (Total N=48) N	Literature (Total N=48) %
Male	Female	A-	17	36.2	12	25.5	0	0.0
Male	Female	A+	26	55.3	32	68.1	7	14.6
Male	Female	B-	0	0.0	0	0.0	0	0.0
Male	Female	B+	3	6.4	3	6.4	25	52.1
Male	Female	C-	1	2.1	0	0.0	0	0.0
Male	Female	C+	0	0.0	0	0.0	16	33.3
White	Black	A-	31	66.0	28	59.6	33	68.8
White	Black	A+	4	8.5	4	8.5	14	29.2
White	Black	B-	10	21.3	6	12.8	0	0.0
White	Black	B+	0	0.0	0	0.0	1	2.1
White	Black	C-	2	4.3	9	19.1	0	0.0
White	Black	C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	30	63.8	31	66.0	33	68.8
White	Hispanic	A+	8	17.0	8	17.0	9	18.8
White	Hispanic	B-	8	17.0	5	10.6	6	12.5
White	Hispanic	B+	0	0.0	0	0.0	0	0.0
White	Hispanic	C-	1	2.1	3	6.4	0	0.0
White	Hispanic	C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	27	57.4	22	46.8	35	72.9
PPT	CBT	A+	16	34.0	24	51.1	8	16.7
PPT	CBT	B-	4	8.5	0	0.0	5	10.4
PPT	CBT	B+	0	0.0	1	2.1	0	0.0
PPT	CBT	C-	0	0.0	0	0.0	0	0.0
PPT	CBT	C+	0	0.0	0	0.0	0	0.0

REVIEW OF ITEMS WITH DATA

In the preceding section on statistical analysis of item data, it was stated that content-area test development specialists used certain statistics from item and DIF analyses of the spring 2016 embedded field test to identify items for further review. Specific flagging criteria for this purpose were specified in the previous section. Items not identified for this review were those that had good statistical characteristics and, consequently, were regarded as statistically acceptable. Likewise, items of extremely poor statistical quality were regarded as unacceptable and needed no further review. However, there were some items that DRC content-area test development specialists and DRC psychometric specialists regarded as needing further review by a committee of Pennsylvania educators. The intent was to capture all items that needed a closer look; thus, the criteria employed tended to over-identify rather than under-identify items.

The review of the items with data was conducted by more than 30 Pennsylvania educators (teachers and PDE staff) broken out into exam-based committees. The review took place on September 7 and 8, 2016. In these sessions, committee members were first trained by a representative from DRC's psychometrics staff with regard to the statistical indices used in item evaluation. This training was followed by a discussion with examples concerning reasons an item might be retained regardless of the statistics. The committee review process involved a brief exploration of possible reasons for the statistical profile of an item (e.g., possible bias, grade appropriateness, instructional issues) and a decision regarding acceptance. DRC content-area test development specialists facilitated the review of the items. Each committee reviewed the pool of flagged embedded field test (EFT) items and made recommendations on each item. The results of the committee reviews are show in the table below. Further discussion on how this information was used is covered in Chapter Six.

Table 5–11. Spring 2016 Keystone Exam Data Review Results

Exam	Module	Number of Items in Spring 2016 EFT	Number of Items*	% of Field Test*	Number of Items**	% of Field Test**	Number of Items***	% of Field Test***
Algebra I	1	144	61	42.36%	20	13.89%	20	13.89%
Algebra I	2	144	68	47.22%	14	9.72%	15	10.42%
Biology	4	214	69	32.24%	0	0.00%	2	0.93%
Biology	2	216	47	21.76%	0	0.00%	3	1.39%
Literature	1	168	48	28.57%	8	4.76%	11	6.55%
Literature	2	168	40	23.81%	7	4.17%	9	5.36%
	Total	1054	333	31.59%	49	4.65%	60	5.69%

*Flagged Items in Spring 2016 EFT Examined at Sept 2016 Data Review Committee
**Flagged Items in Spring 2016 EFT Rejected by Sept 2016 Data Review Committee
*** Items Classified as "Rejected" from Spring 2016 EFT (all sources: Data Review Committee, PDE, and DRC)

CHAPTER SIX: OPERATIONAL FORMS CONSTRUCTION FOR 2017 ADMINISTRATIONS

FINAL SELECTION OF ITEMS AND KEYSTONE FORMS CONSTRUCTION

Approximately 50% of the items that made up the Winter 2016/2017, Spring 2017, and Summer 2017 operational forms emerged from the Spring 2016 embedded field test. The remaining operational (core) items were part of the biennial core-to-core overlap. For more information about the core-to-core overlapping items, please see Chapter Three. Prior to being placed on the operational tests, these items had undergone multiple reviews, including the following:

- Reviews by Data Recognition Corporation (DRC) content-area test development specialists and curriculum specialists to ensure that all items were properly aligned with content standards
- Formal bias, fairness, and sensitivity review by the Bias, Fairness, and Sensitivity Committee, which consisted of a multiethnic group of men and women having expertise with special-needs students and English Language Learners (ELL)
- Formal review by the content committees consisting of Pennsylvania educators, including teachers as well as district personnel
- Pennsylvania Department of Education (PDE) review
- Item data review by members of the PDE subject-area teacher committees

The item and bias reviews are detailed in Chapter Three. The results of the data review are summarized in Chapter Five.

The end product of the above process was an item status designation for each field test item. All items having an item status code of Accepted/Operational Ready were candidates to be selected for the 2017

Keystone Exams. To have an item status code of Accepted/Operational Ready meant that the item met the following criteria:

- Appropriately aligned with its designated Keystone Assessment Anchor Content Standard (Assessment Anchor) and subclassifications
- Acceptable in terms of bias/fairness/sensitivity issues, including differential item functioning (for gender and ethnicity)
- Acceptable in terms of psychometric standards, including a special review of flagged items

Next, all relevant information regarding the acceptable items, including associated graphics, was entered into the item banking system known as IDEAS (Item Development and Education Assessment System). From IDEAS and other database sources, Microsoft Excel files were created for each exam. These files contained all relevant content codes and statistical characteristics. IDEAS also created an item card displaying each acceptable item, any associated graphic, and all relevant exam codes and item statistics for use by the subject-area test development specialists and psychometric services staff.

DRC test development specialists reviewed the test design blueprint, including the number of items per strand for each content-area test. Psychometricians provided content-area test development specialists with an overview of the psychometric guidelines for forms construction.

Senior DRC content-area test development specialists reviewed all items in the operational pool to make an initial selection (pull) for common (core) positions according to test blueprint requirements and psychometric guidelines. Changes to items were not encouraged since alterations could affect how an item would perform in subsequent testing.

For these common items, this meant that the combination of multiple-choice (MC) and constructed-response (CR) items would yield the appropriate range of points while tapping an appropriate variety of the Assessment Anchors and related Eligible Content within each Reporting Category (module). Items selected in the first round were examined with regard to how well they fit together as a set. Of particular concern were the following:

- One item providing cues as to the correct answer to another item
- Context redundancy (e.g., mathematics items with a sports context)
- Presence of clang (distractors not unique from one another)
- Diversity of names and artwork for gender and ethnicity

A core-building software tool known as PerForm was used in concert with performance data and metadata from IDEAS to aid in the organization and communication of the pulled data. PerForm automatically tabulates the statistical characteristics of the proposed core, updating instantly whenever item swaps were performed. Using PerForm, the first round of items was then evaluated for statistical features such as an acceptable point-biserial correlation and whether correct answers were distributed equally—that is, whether approximately 25 percent of correct answers appeared in each of the four possible positions (A, B, C, or D). Selected items that were deemed psychometrically less advantageous in contrast to the overall psychometric characteristics of the core resulted in a search by the senior reviewer for suitable replacements. At this point, the second round of items was analyzed. If necessary, this iterative process between content-based selections and statistical properties continued in an effort to reach the best possible balance.

Once the recommendations were finalized for the core items, they were submitted to PDE for review. Department staff provided feedback, which could be in the form of approval or recommendations for replacement of certain items. Any item replacement was accomplished by the collective effort of the test development specialists, psychometricians, and PDE staff until final PDE approval was given. See Appendix F for the Keystone Exams Tally Sheets.

Following final approval by PDE, test development specialists developed print and online forms based on the approved core and approved embedded field test items. Both modes of delivery were built using IDEAS. Highly skilled test development specialists and editors used specialized checklists to verify accuracy of layout and formatting in both modes of delivery. Following final approval to print, the documents were prepared for the printing presses. See Chapter 8 for more information about the steps following the final approval to print.

SPECIAL FORMS USED WITH THE OPERATIONAL 2017 KEYSTONE EXAMS

SPANISH TRANSLATION

Starting with the operational exams in spring 2011, school personnel had the option of allowing Spanish-speaking students who had been enrolled in schools in the United States for less than three years to respond to a Spanish version of the Keystone Exams for Algebra I and Biology. The original translation of the items and the *Directions for Administration Manual* was initiated by Language Services Associates and completed/verified by Exact Communications. These companies use translators with varying cultural and regional backgrounds to create the Spanish versions. The translations were then reviewed and verified by DRC's internal Spanish group. As part of the internal review, a Spanish style guide is maintained to document Spanish word choice from administration-to-administration and across exams within an administration.

Following PDE's approval of the translation, the translated text was typeset into print delivery forms. The test book is constructed with a side-by-side format with the English text and Spanish-translated text on facing pages. The Spanish-translated text is on the left-hand side followed by the original English text on the right-hand (facing) side. Each CR item covered either two or four pages in the answer book, depending on the length of the original English-language item. In the case of four-page open-ended items, the first set of facing pages of an item was presented in Spanish. The second set of facing pages of an item was presented in the original English.

Those students using this accommodated version are permitted to write their answers on either the English language pages or on the translated Spanish language pages. Their answers can be written in English, Spanish, or a combination of both Spanish and English because all pages are evaluated and scored, and the highest possible scores from those combinations are recorded for the students.

On a yearly basis, the PDE examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. Three accommodations manuals for Pennsylvania assessments titled *Accommodations Guidelines for Students with IEPs and Students with 504 Plans, Accommodations Guidelines for Students without IEPs or 504 Plans*, and *Accommodations Guidelines for English Language Learners* were developed for use with the Keystone Exams. The manuals can be accessed by going to *www.pdesas.org/Assessment/Keystone* and selecting the corresponding document under the Test Accommodations section. For more information about the general on-screen testing aids available to students taking the online mode of delivery, see Chapter Two.

AUDIO

For students requiring an auditory presentation accommodation, a text-to-speech synthesizer is available to students taking the Algebra I and/or Biology Exams using the online mode of test delivery. For each operational exam, one form was selected for the creation of the audio version. Special scripts are crafted, writing out each item, distractor, graphic, and directions to utilize the rich, synthesized voice features while accounting for specific nuances of the intended sounds. The resulting audio information is provided to students receiving the accommodation. Since additional software is required to generate the vocalization from the scripted text and since headphones are required to minimize disruptions within a computer lab setting, local school personnel generally must preplan to use the audio version in order to ensure that the student has a properly equipped computer and a proper setting.

BRAILLE, LARGE PRINT, AND VIDEO SIGN LANGUAGE

Students were able to respond to test materials that were available in Braille, large print, or Video Sign Language. At each grade level assessed, one form was selected for the creation of these accommodations.

The large print edition is a replication of the standard print form; 8.5X11 standard form is enlarged to an 11X17 page format to achieve a font size of approximately 18-point. A side-by-side verification is completed between the standard print and large print forms to ensure that the integrity of all formatting and graphics is maintained on the large print forms.

For Braille production, the final selected form is delivered to American Printing House for the Blind (APH) via APH's secure website. APH ensures that all tests are translated correctly and accurately by using a translator and a validator. After all Braille booklets are printed, APH conducts a quality assurance step to ensure all items are bound in order and directions are included. All Braille booklets are shipped from APH to DRC via UPS.

DRC applies a security barcode to each large print and Braille booklet for purposes of shipping, distributing, and collecting the materials. This security barcode is used with DRC's Operations Materials Management System (Ops MMS).

School personnel were directed to transcribe all student answers (SR and CR) into scannable answer documents exactly as the student responded. No alterations or corrections of student work were permitted, and the transcribed answer document had to have the same form designation as the Braille and large print version.

DRC utilizes Victory Productions for the production of Sign Language Videos. The items are passed to Victory Productions via a secure ftp site. Two to three different interpreters are used to interpret and validate the translations during video recording. After the interpretations are recorded and returned to DRC via a secure ftp site, DRC loads these videos in the online test engine. When school personnel assign the specific sign language accommodation, the student will be able to play each video next to the item.

CHAPTER SEVEN: TEST ADMINISTRATION PROCEDURES

SECTIONS, SESSIONS, TIMING, AND LAYOUT OF THE KEYSTONE EXAMS

The design for most Keystone Exams utilizes separate test books and answer books. An answer book is used to respond to the multiple-choice (MC), evidence based selected response (EBSR), and constructed-response (CR) items and to collect demographic information. The MC items and all stimulus text are placed within the test book. One exam uses a single consumable book. When a single scannable answer book is utilized, the contents of the answer book and the test book are combined into one integrated book. The table below identifies the exam material format for each 2017 Keystone Exam.

Table 7–1. Book Type by Exam

Exam	Test Book	Answer Book	Single Consumable Book
Algebra I	✓	✓	
Biology	✓	✓	
Literature	✓	✓	

Generally, a separate test book and answer book are used to separate the MC items and the CR items. For passage-based exams, like Literature, the separate exam materials allow the students to reference stimulus materials at the same time that a response to a CR item is composed. In addition, since all student responses must be scanned for scoring and storage purposes, a separate answer book limits the volume of data that must be stored.

SECTIONS AND SESSIONS

Each operational Keystone Exam is organized around two equally sized test modules; the focus of each is on two or more specific, thematically linked Assessment Anchors and Eligible Content. The content in each module remains separate, and items measuring the Eligible Content in a module appear only in that module. The module design is identical in the print (paper-and-pencil) and online modes of delivery.

Each exam section is administered in an exam session. Local districts must schedule the two modules as two separate exam sessions (morning and afternoon or two separate days), and an individual module must be completed in one exam session.

Each test session is to be completed within a prescribed testing window. The testing windows below reflect both online and paper-based administrations in the 2016–2017 school year. The testing windows also include all make-up testing. Schools were able to choose one of the two testing windows (“waves”) for the winter administration. Two windows were provided to accommodate different semester end dates for schools with block scheduling.

Table 7–2. Winter 2016/2017 Operational Keystone Exam Testing Windows

Exam	Wave 1 Dates	Wave 2 Dates
Algebra I	December 5-16, 2016	January 9-23, 2017
Biology	December 5-16, 2016	January 9-23, 2017
Literature	December 5-16, 2016	January 9-23, 2017

Table 7–3. Spring 2017 Operational Keystone Exam Testing Window

Exam	Dates
Algebra I	May 15-26, 2017
Biology	May 15-26, 2017
Literature	May 15-26, 2017

Table 7–4. Summer 2017 Operational Keystone Exam Testing Window

Exam	Dates
Algebra I	July 31-August 4, 2017
Biology	July 31-August 4, 2017
Literature	July 31-August 4, 2017

TIMING

In general, the estimated testing times allow 1–2 minutes per MC item on the Keystone Exams, depending on the exam. The CR items are estimated to take approximately 5–10 minutes per item, also depending on the exam. Each stimulus passage on the Literature exam is estimated to take about 10 minutes to read. There was no difference in the timing for online and print forms of delivery.

Test administrators were instructed that each section (module) in a form should be scheduled as a separate exam session. Exam modules were not to have been scheduled back-to-back in the morning (or in the afternoon). Instead, the exam modules were to be divided across two days or divided across the morning and afternoon of the same day.

Since not all students are expected to finish the exam sections at the same time, test administrators are advised to use the flexibility of the time limits to the students' advantage. For example, test administrators manage the testing time so that students do not feel rushed while they are taking any assessment section, and no student is penalized because he or she works slowly. It is also stressed to test administrators that a student should not be given an opportunity to waste time. Students are told to close their exam materials when they have finished the section of the exam in which they have been working. Students who finish early are allowed to sit quietly or read for pleasure until all students have finished. Students with special requirements and/or abilities (i.e., physical, visual, auditory, or learning disabilities as defined by their IEP or service contracts) and students who just work slowly may require extended time. Special assessment situations are arranged for these students. When all students in a testing session indicate that they have finished an exam section, test administrators end the section.

Scheduled extended time is provided by a test administrator, and students are allowed to request extended time if they indicate that they have not completed the task. Such requests are granted if the test administrator finds the request to be educationally valid. Test administrators are advised that not permitting ample time for students to complete the assessment might impact the students' and school's performances.

As a general guideline, however, when all students indicate that they have finished a section, that section is closed. Students requiring time beyond the majority of the student population are allowed to continue immediately following the regularly scheduled session in another setting. When such accommodations are made, school personnel ensure that students are monitored at all times to prevent sharing of information. Students are not permitted to continue a section of the assessment after a significant lapse of time from the original session.

Table 7–5. Keystone Testing Load and Duration by Exam

Exam	Total No. of MC Items per Form per Administration	Total No. of CR Items per Form per Administration	Total Estimated Testing Time per Form (in minutes)	Total Estimated Administration Time per Form (in minutes)
Algebra I	46	8	150	170–180
Biology	64	8	144	164–174
Literature	46	8	146	166–176

Table 7–6. Keystone Testing Load and Duration by Type per Unit (in minutes)

Exam	Administration Tasks	Stimulus Passages	MC Points per Minute [PPM]*	CR [PPM]*	Estimated Overall PPM**
Algebra I	24	–	1.5 [0.670]	10 [0.400]	0.400
Biology	24	–	1.25 [0.800]	8 [0.375]	0.458
Literature	24	–	1 [1.000]	5 [0.600]	0.356

*Based on rates per item type
**Based on total testing time

Prior to beginning the exam, students were asked to verify that they understood the *Code of Conduct for Test Takers* by marking the circle in the exam. Additionally, an Attention statement was added to the beginning of the exams to notify students of the penalties incurred if exam materials are copied.

LAYOUT

The layout of the operational Keystone Exams follows a general sequence regardless of the exam. Each exam is divided into thematically linked sets of content called modules. Within each module, there are core (common) items and field test items. Both core and field test items are represented though MC and CR items.

Stimulus material (like passages), text for MC items, answer options, and any stimulus materials associated with MC items or answer options appear in the test book. Answer bubbles, text for CR items, and associated response spaces appear in the answer book.

Within a non-passage-based module (like Algebra I and Biology), the sequencing of items follows this pattern:

1st: Approximately half of the MC items

2nd: Half of the CR items

3rd: Remaining half of the MC items

4th: Remaining CR items

Within a passage-based module (like Literature), the sequencing of items follows this pattern:

1st: Stimulus Passage X

2nd: MC items associated with Passage X

3rd: CR items associated with Passage X

4th: Stimulus Passage Y

5th: MC items associated with Passage Y

6th: CR items associated with Passage Y

7th: Stimulus Passage Z

8th: MC items associated with Passage Z

9th: CR items associated with Passage Z

Regardless of sequencing pattern, the field test items appear in the relative middle of each module, and item sequencing is self-contained within a module.

For more information about the test layout of the operational Keystone Exams, see Appendix G.

SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS

There were two shipments sent out by DRC for the Keystone Exams operational assessments:

- Shipment one contained the *Handbook for Assessment Coordinators* and the *Directions for Administration Manual*s for each subject at schools participating in the Algebra I, Biology, and Literature Keystone Exams. Shipment one was delivered four weeks prior to the start of the test window.
- Shipment two contained the administrative materials (e.g., Return Shipping labels, District/School labels, Do Not Score labels, Student Precode labels) and secure materials (e.g., consumable test/answer books) for each subject at schools participating in the Algebra I, Biology, and Literature Keystone Exams. Shipment two was delivered two weeks prior to the start of the test window.

DRC ensured that all exam materials were assembled correctly prior to shipping. DRC operations staff used the automated Operations Materials Management System (Ops MMS) to assign secure materials to a school at the time of ship out. This system used barcode technology to provide an automated quality check between items requested for a site and items shipped to a site. A shipment box manifest was produced for and placed in each box shipped. DRC operations staff double-checked all box contents with the box manifest prior to sealing the box for shipping to ensure accurate delivery of materials. DRC operations staff performed lot acceptance sampling on both shipments. Districts and schools were selected at random and examined for correct and complete packaging and labeling. This sampling represented a minimum of 10 percent of all shipping sites.

DRC's materials management system, along with the systems of shippers, allowed DRC to track materials from DRC's warehouse facility to receipt at the district, school, or testing site. All DRC shipping facilities, materials processing facilities, and storage facilities are secure. Access is restricted by security code. Non-DRC personnel are escorted by a DRC employee at all times. Only DRC inventory control personnel have access to stored secure materials. DRC employees are trained in and made aware of the high level of security that is required.

DRC used United Parcel Service (UPS) to deliver the secure materials to the testing sites.

ONLINE TESTING

Online administration is managed through the DRC eDIRECT client portal that provides tiered, secure access to all required administrative functions. Within eDIRECT, users manage student information and create test sessions.

Student information from the Pennsylvania Information Management System (PIMS) is imported into eDIRECT via file transfer or LEAs upload student directly into eDIRECT. From here, LEAs are able to view all of the demographic information associated with the students from PIMS before placing them in test sessions for test tickets.

Once the student data is loaded into Test Setup, users organize students into test sessions. Test sessions can be created by class, grade, or school. Through Test Setup, users can also update student accommodation information, print test tickets, and monitor student testing status.

The student login ticket contains unique login credentials used by the student to access the testing software. For a selected test session, users can download and print a PDF document containing instructions, a roster of student tickets being printed, and the actual test tickets. Student test tickets are considered secure materials and LEAs are required to keep printed tickets in a predetermined, locked, secure storage area.

The web-based test engine, DRC INSIGHT Online Learning System, is downloaded onto computers that students will access during the assessment. Test items and forms can only be accessed using a valid test ticket. During testing, responses are sent to a DRC server each time the student navigates away from an item or clicks the Next button to submit an answer. The system is configured to allow students to review answers before submitting their test.

TEST SECURITY MEASURES

Test security is essential to obtaining reliable and valid scores for accountability purposes. Test Security Certifications were required to be signed by each building Principal, School Assessment Coordinator, District Assessment Coordinator, Test Administrator, and Proctor after the assessment is administered. All signed Certifications were returned to the Chief School Administrator who must retain the Certifications for three years. The purpose of the Certifications was to serve as a tool to document that the individuals responsible for administering the assessments both understood and acknowledged the importance of test security and accountability. Additional details can be found in the *Handbook for Assessment Coordinators*. A screen shot of the Test Administrator Certificate is provided in Figure 7-1.

Figure 7-1. Test Administrator and Proctor Test Security Certification

pennsylvania
DEPARTMENT OF EDUCATION

2017 PSSA Test Security Certification

(Test Administrator and Proctor)

District: ______________________________

School: ______________________________

AUN: ______________________________

Maintaining the security and integrity of all assessment materials, preventing any dishonest or fraudulent behavior in the administration and handling of the assessment, and promoting a fair and equitable testing environment are essential in order to obtain reliable and valid student scores. In that regard, I certify the following:

Prior to the administration of the assessment, I completed the Pennsylvania State Test Administration Training, and I understand that the assessment materials are secure, confidential, and proprietary documents owned by the Pennsylvania Department of Education.

I have not reviewed, discussed, disseminated, described, or otherwise revealed the contents of the assessment to anyone. I have not removed any assessment materials from the school building unless I was specifically authorized to administer the assessment to a student on homebound instruction. I have not kept, copied, reproduced, released, or used any assessment, assessment question, specific assessment content, or examinee response to any item or any section of the secure assessment in any manner that is inconsistent with the instructions provided by or through the Pennsylvania Department of Education. I have not provided any examinee with an answer to an assessment question or in any way influenced an examinee's response to any assessment question. I have not in any manner altered or caused the alteration of any examinee response, assessment booklet, or papers used by examinees.

I understand that any breach in assessment security could result in the invalidation of assessment results, professional discipline, and/or criminal prosecution.

I understand that false statements herein are made subject to the penalties of 18 Pa.C.S. § 4904.

______________________	______________________	______________________
Administrator/Proctor Name	Administrator/Proctor Signature	Date of Signature

SAMPLE MANUALS

Copies of the *Handbook for Assessment Coordinators* and the *Directions for Administration Manual*s can be found on the PDE website at www.education.pa.gov.

TESTING WINDOW ASSESSMENT ACCOMMODATIONS

The *Accommodations Guidelines* was developed by PDE for use with the PSSA. This manual can be found on the PDE website at www.education.pa.gov. Additional information regarding assessment accommodations can be found in Chapter Four of this report.

CHAPTER EIGHT: PROCESSING AND SCORING

RECEIPT OF MATERIALS

Receipt of Pennsylvania Keystone Exams' test materials began five days after the start of the test window. DRC's Operations Materials Management System (Ops MMS) was utilized to receive assessment materials securely, accurately, and efficiently. This system features innovative automation and advanced barcode scanners. Captured data was organized into reports, which provided timely information with respect to suspected missing material.

The first step in Ops MMS was Box Receipt. When a shipment arrived at DRC, the boxes were removed from the carrier's truck and passed under a barcode reader, which read the barcode printed on the return label and identified the district and school. The number of boxes was immediately compared to what was picked up at the district. The data collected in this process was stored in the Ops MMS database. After the barcode data was captured, the boxes were placed on a pallet and assigned a corresponding pallet number.

Once the Box Receipt process was completed, the Materials Separation phase began. Warehouse personnel opened the boxes and sorted materials by grade, subject, and status (used and/or unused booklets) into scanning boxes. Every booklets' security barcode and precode barcode were hand scanned to link each document to the original box. As the booklets were sorted, Ops MMS guided the floor operator to the box in which to place the document. Ops MMS kept count and record of the materials placed in each box. This count remained correlated to the box as an essential quality-control step throughout the secure booklet processing and provided a target number for all steps of the check-in process. Once a box was closed, an MMS Processing Label was placed on that box.

Once labeled, the sorted and counted boxes proceeded to Quality Assurance, where a secure booklet check-in operator used a hand scanner to scan the MMS Processing Label. This procedure identified the material type and quantity parameters for what Ops MMS should expect within a box. The box contents were then loaded into the stream feeder.

The documents were fed past oscillating scanners that captured both the security code and precode from the booklets. A human operator monitored an Ops MMS screen that displayed scan errors, an ordered accounting of what was successfully scanned, and the document count for each box. The system ensured that all material within the box matched the information obtained from the original hand-scanning process.

When all materials were scanned and the correct document count was confirmed, the box was sealed and placed on a pallet. If the correct document count was not confirmed, or if the operator encountered difficulties with material scanning, the box and its contents were delivered to an exception-handling station for resolution.

This check-in process occurred immediately upon receipt of materials; therefore, DRC provided feedback to districts and schools regarding any missing materials based on actual receipt versus expected receipt. Sites that had 100 percent of their materials missing after the date they were due to DRC were contacted, and any issues were resolved.

Throughout the process of secure booklet check-in, DRC project management ran a daily Missing Materials Report. Every site that was missing any number of booklets was contacted by DRC. Results of these correspondences were recorded for inclusion in the final Missing Materials Report if the missing booklets were not returned by the testing site. DRC produced the Missing Materials Report for PDE upon completion of secure booklet check-in. The report listed all schools in each participating district, along with security barcodes for any booklets not returned to DRC.

After scannable materials (used answer booklets) were processed through booklet check-in, the materials became available to the DRC Document Processing log-in staff for document log-in. The booklets were logged in using the following process:

- A DRC scannable barcode batch header was scanned, and a batch number was assigned to each box of booklets.

- The DRC box label barcode was scanned into the system to link the box and booklets to the newly created batch and to create a Batch Control Sheet.
- The DRC box label barcode number and the number of booklets in the box were printed on the Batch Control Sheet for document-tracking purposes. All booklets linked to the box barcode were assigned to the batch number and tracked through all processing steps. As booklets were processed, DRC staff dated and initialed the Batch Control Sheet to indicate that proper processing and controls were observed.

Before the booklets were scanned, all batches went through a quality inspection to ensure batch integrity and correct document placement.

After a quality check-in at the DRC Document Processing log-in area, the spines were cut off the scannable documents, and the pages were sent to DRC's Imaging and Scoring System.

SCANNING OF MATERIALS

Customized scanning programs for all scannable documents were prepared to read the books and to format the scanned information electronically. Before materials arrived, all image-scanning programs went through a quality review process that included scanning of mock data from production books to ensure proper data collection.

DRC's image scanners were calibrated using a standard deck of scannable pages with 16 known levels of gray. On a predefined page location, the average pixel darkness was compared to the standard calibration to determine the level of gray. Marks with an average darkness level of 4 or above on a scale of 16 (0 through F) were determined to be valid responses, per industry standards. If multiple marks were read for a single item and the difference between the grayscale reads was greater than four levels, the lighter mark was discarded. If the multiple marks had fewer than four levels of grayscale difference, the response was flagged and forwarded to an editor for resolution.

DRC's image scanners read selected-response, demographic, and identification information. The image scanners also used barcode readers to read preprinted barcodes from a label on the book.

The scannable documents were automatically fed into the image scanners where predefined processing criteria determined which fields were to be captured electronically. Open-ended (OE) response images were separated out for image-based scoring.

During scanning, a unique serial number was printed on each sheet of paper. This serial number was used to ensure document integrity and to maintain sequencing within a batch of books.

A monitor randomly displayed images, and the human operator adjusted or cleaned the scanner when the scanned image did not meet DRC's strict quality standards for image clarity.

All images passed through a process and a software clean-up program that despeckled, deskewed, and desmeared the images. A random sample of images was reviewed for image quality approval. If any document failed to meet image quality standards, the document was returned for rescanning.

Page-scan verification was performed to ensure that all predefined portions of the booklets were represented in their entirety in the image files. If a page was missing, the entire book was flagged for resolution.

After each batch was scanned, books were processed through a computer-based editing program to detect potential errors as a result of smudges, multiple marks, and omissions in predetermined fields. Marks that did not meet the predefined editing standards were routed to editors for resolution.

Experienced DRC Document Processing editing staff reviewed all potential errors detected during scanning and made necessary corrections to the data file. The imaging system displayed each suspected error. The editing staff then inspected the image and made any necessary corrections using the unique serial number printed on the document during scanning.

Upon completion of editing, quality control reports were run to ensure that all detected potential errors were reviewed again and a final disposition was determined.

Before batches of books were extracted for scoring, a final edit was performed to ensure that all requirements for final processing were met. If a batch contained errors, it was flagged for further review before being extracted for scoring and reporting.

During this processing step, the actual number of documents scanned was compared to the number of books assigned to the box during book receipt. Count discrepancies between book receipt and books scanned were resolved at this time.

Once all requirements for final processing were met, the batch was released for scoring and student level processing.

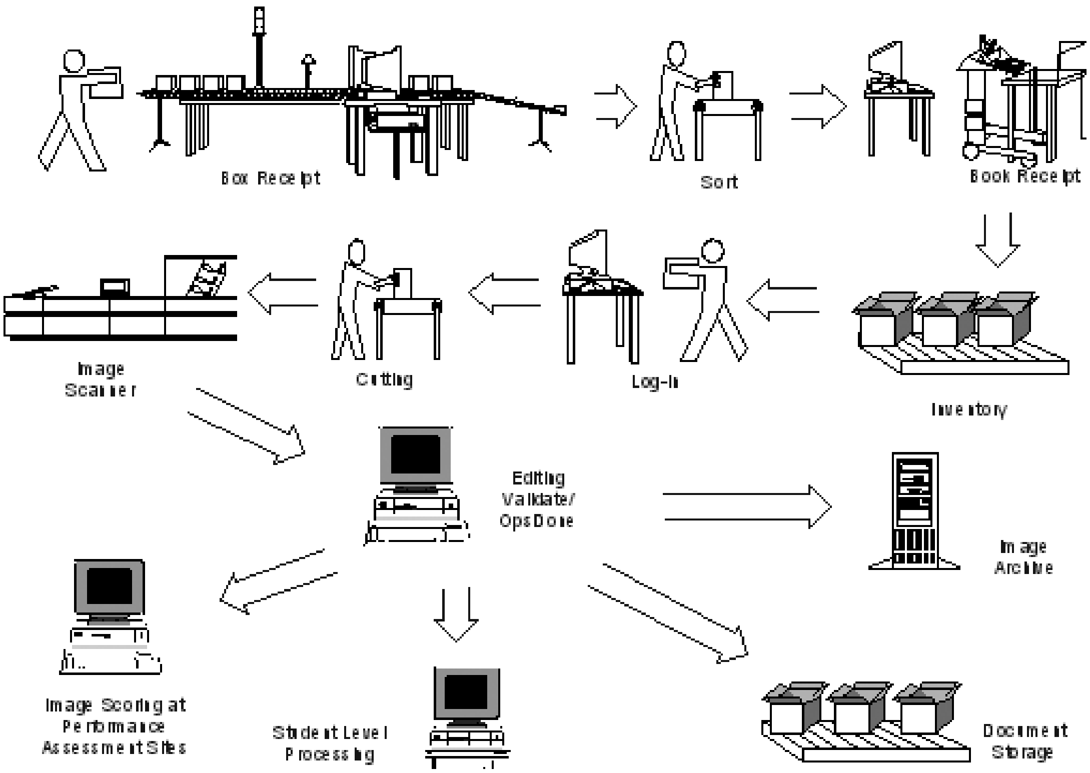
Table 8–1 shows the number of answer books received through book check-in, the number of books that contained student responses that were scanned and scored, the number of test books received, and the total number of books received for the Algebra I, Biology, and Literature Keystone Exams.

Table 8–1. Counts of 2017 Keystone Exams Materials Received: Algebra I, Biology, and Literature

Exam	Answer Books Received	Used Answer Books Received	Test Books Received	Total Books Received	Total Books Shipped
Algebra I (Winter)	45,792	53,281	45,812	91,604	91,628
Biology (Winter)	35,024	39,293	35,034	70,058	70,070
Literature (Winter)	31,095	34,666	31,107	62,202	62,218
Algebra I (Spring)	225,644	131,003	225,640	451,284	451,330
Biology (Spring)	183,488	133,013	183,486	366,974	366,984
Literature (Spring)	173,761	125,717	173,761	347,522	347,534
Algebra I (Summer)	1,839	1,136	1,840	3,679	3,726
Biology (Summer)	1,451	710	1,451	2,901	2,923
Literature (Summer)	1,173	308	1,173	2,346	2,368

Figure 8–1 illustrates the production workflow for DRC’s Ops MMS and Image Scanning and Scoring System from receipt of materials through all processing of materials and the presentation of scanned images for scoring.

Figure 8–1. Workflow System



MATERIALS STORAGE

Upon completion of processing, student response documents were boxed for security purposes and final storage as follows:

- Project-specific box labels were created containing unique customer and project information, material type, batch number, pallet/box number, and the number of boxes for a given batch.
- Boxes were stacked on pallets that were labeled with the project information and a list of the pallet's contents before delivery to the Materials Distribution Center for final secure storage.

Materials will be destroyed one year after the contract year ends with PDE written approval.

ONLINE TESTING

The DRC INSIGHT test engine runs on a custom web browser that is designed to ensure a fully secure environment during testing. The secure browser "locks down" the student's testing device, preventing the student from accessing the desktop, the Internet, and other external programs. For non-secure testing such as practice and training sessions, students can use the Online Tools Training (OTT) environment, which runs on a standard web browser.

The custom browser software is downloaded from eDIRECT and installed onto student testing devices. The secure browser can be installed on computers individually, or it can be downloaded to a central location, copied, and distributed to multiple computers simultaneously using common network distribution tools. Everything needed for testing is found within the secure browser, eliminating the need for districts to coordinate updates to third-party software.

Prior to operational use, DRC's quality assurance staff will perform full system-level tests in an independent test environment that simulates the production configuration. Tests are run on all supported computer platforms and browsers and include comprehensive review of system functionality, usability, reliability, security, and overall performance. Test content is also validated during this process.

Multiple methods are used to ensure secure data transfer, including encryption technologies and Secure Sockets Layer (SSL) protocol through Hypertext Transfer Protocol Secure (HTTPS). Test content is encrypted at the host server, and remains encrypted throughout all network transmissions; content is decrypted only once the student login is validated. Decrypted test content on the student workstation is stored only in memory during each test session. Once the session is ended (the test is completed or the student logs out), computer memory is purged to ensure security of test content is maintained.

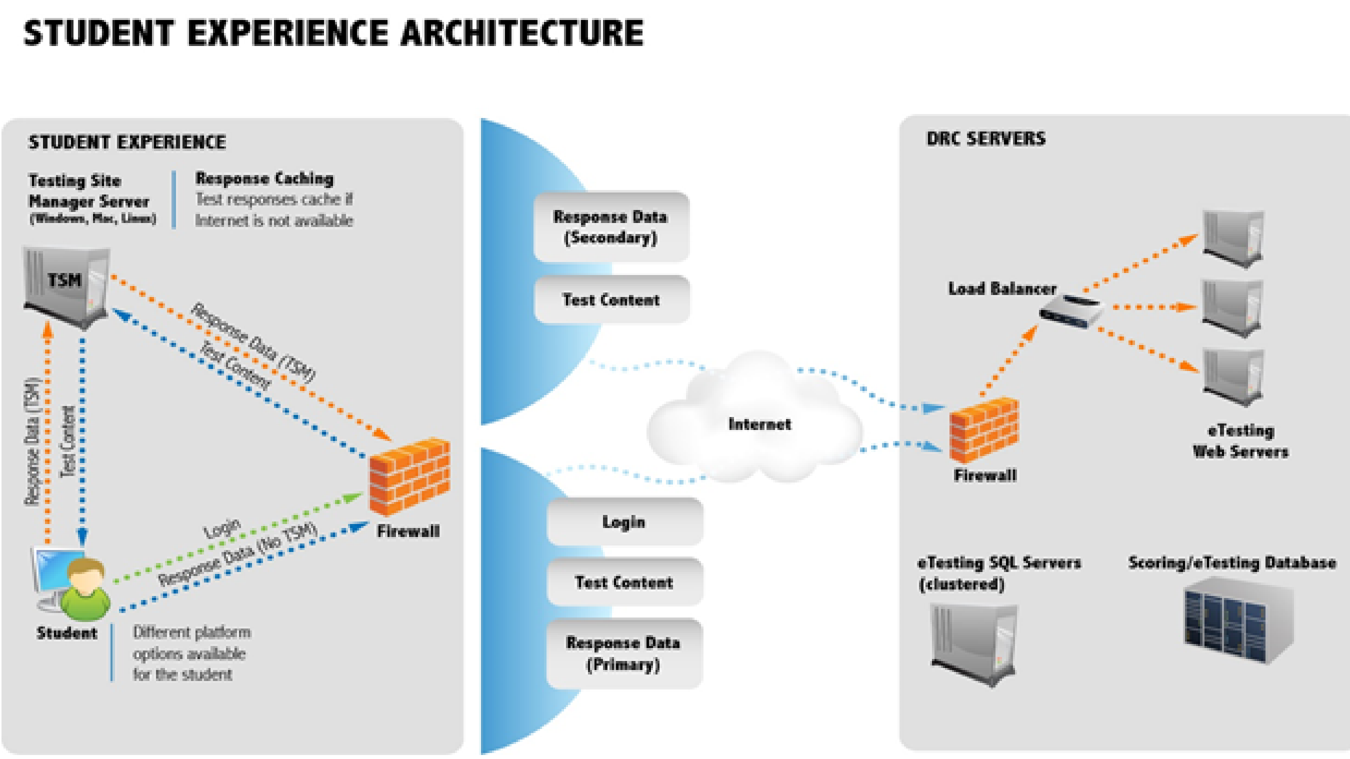
Responses are saved automatically every 45 seconds during testing, or when the student navigates away from an item or answers a selected-response item (whichever comes first). If a particular question takes the student longer than 45 seconds to answer, then the partial, incomplete responses are submitted at 45-second intervals until the student completes the item. This auto-save helps safeguard against students losing their work on longer items, such as constructed-response items. When the student returns to the test after a break or interruption, the student is returned to the point that they left off without having to navigate through all previously answered questions.

Table 8–2. Counts of 2017 Keystone Exams Online Assessments

Grade/Subject	Total Online Assessments Completed
Algebra I (Winter)	5,653
Biology (Winter)	6,171
Literature (Winter)	4,921
Algebra I (Spring)	14,156
Biology (Spring)	17,568
Literature (Spring)	13,246
Algebra I (Summer)	378
Biology (Summer)	280
Literature (Summer)	152

Figure 8–2 illustrates the secure transfer of online test responses between the student and DRC.

Figure 8–2. Architecture of the Student Testing Experience



SCORING MULTIPLE-CHOICE ITEMS

For both online and paper-and-pencil modes, the scoring process included the scoring of multiple-choice (MC) items against the answer key and the aggregation of raw scores from the OE responses. A student's raw score is the actual number of points achieved for tested elements of an assessment. From the raw scores, the scale scores were calculated.

The student file was scored against the final and approved MC answer key. Items were scored as right, wrong, omitted, or double-gridded (more than one answer was bubbled for an item). Sections of the exam were evaluated as a whole, and an attempt status was determined for each student for each subject. The score program defined all data elements for reporting at the student level.

RANGEFINDING

After student answer documents were received and processed, DRC's Performance Assessment Services (PAS) staff assembled groups of responses that exemplified the different score points for each subject. The score point ranges were represented by the following scoring guidelines:

- 0–3 item-specific scoring guidelines for Literature
- 0–4 item-specific scoring guidelines for Algebra 1 (some items were divided into separate parts that were scored on a 0 – 1, 0 – 2, or 0 – 3 point scale, but the sum of the parts always resulted in an overall score of 0 – 4 for each item).
- 0–3 item-specific scoring guidelines for Biology

Responses were pulled from the embedded field test portion of the Keystone Exams for each subject. Once examples of all score points were selected for each item, sets were assembled for rangefinding and copies were made for each rangefinding participant. Rangefinding committees consisted of Pennsylvania educators, PDE staff members, DRC Test Development staff, and DRC Performance Assessment Services staff. The Algebra I and Biology rangefinding meetings were held from July 10-13 at the Sheraton Harrisburg-Hershey, Harrisburg. The Literature rangefinding meetings were held from July 10-14, also at the Sheraton Harrisburg-Hershey.

Each rangefinding meeting began in a joint session with a review of the history of the assessment as well as discussion of the purpose of the rangefinding meeting and the role rangefinding plays within the item development process. The session then broke into subject/grade-specific committees. Sets of student responses were presented to the committees, one item at a time. Each committee initially reviewed and scored student responses as a group to ensure that everyone was interpreting the scoring guidelines consistently. Committee members then went on to score responses independently. For each student response, committee members' scores were discussed until a consensus was reached. Only those responses for which there was strong agreement among committee members were chosen for inclusion in training materials for DRC raters.

Discussions of student responses included the mandatory use of scoring guideline language. This ensured that committee members remained focused on the specific requirements of each score level. DRC PAS staff took notes addressing how and why the committees arrived at score point decisions, and this information was used by the scoring directors in rater training.

DRC and PDE discussed scoring guideline edits suggested by the rangefinding committees. Changes approved by PDE were incorporated into the scoring guidelines by DRC Test Development staff. The edited scoring guidelines were used in the preparation of materials and the training of raters.

RATER RECRUITMENT/QUALIFICATIONS

DRC retains a number of raters from year to year; the overall return rate in 2017 was 50%. This pool of experienced raters was drawn from to staff the scoring of the 2017 Keystones. To complete the rater staffing for this project, recruiting events were held and applications for rater positions were screened by DRC's recruiting staff. Candidates were personally interviewed by DRC staff. In addition, each candidate was required to provide an on-demand writing sample, an on-demand math sample, references, and proof of a four-year college degree. In this screening process, preference was given to candidates with previous experience scoring large-scale assessments and degrees emphasizing expertise in the subjects being scored. In some locations, staffing partners were used to augment hiring using the same practices as those employed by DRC. The rater pool consisted of educators and other professionals with content-specific backgrounds. These individuals were valued for their content-specific knowledge, but they were required to set aside their own biases about student performance and accept the scoring standards of the Keystone Exams.

LEADERSHIP RECRUITMENT/QUALIFICATIONS

Scoring directors and team leaders were selected from a pool of employees who displayed expertise as raters and leaders on previous DRC projects. These individuals had strong backgrounds in mathematics, English language arts, or science, and demonstrated organizational, leadership, communication, and management skills. All scoring directors had previous leadership experience working on large scale assessments. All scoring directors, team leaders, and raters were required to sign confidentiality agreements before handling secure materials.

Each room of raters was assigned a scoring director. All handscoring activities were led by a scoring director for the duration of the project. Scoring directors assisted in rangefinding, worked with supervisors to create training materials, conducted team leader training, and were responsible for training the raters. The scoring director made sure that reports were available and interpreted those reports for the raters. The scoring director also supervised the team leaders. Scoring directors were monitored by the project managers.

Team leaders assisted the scoring director with rater training by answering individual questions that raters may not have felt comfortable asking in a large group. Once raters were qualified, team leaders were responsible for monitoring and maintaining the accuracy and workload of each team member. Ongoing monitoring identified those individuals having difficulty scoring accurately. These raters received one-on-one retraining from the team leader or scoring director. Any rater who could not be successfully retrained had his/her scores purged and was released from the project.

TRAINING

As part of preparation for the 2017 Keystone Exams, DRC's PAS staff assembled the PDE-approved scoring guidelines and scored student responses approved by rangefinding committees into sets used for training raters. The item-specific scoring guidelines served as the raters' constant reference. Responses that were relevant in terms of the scoring concepts they illustrated were annotated and included in an anchor set. The full range of each score point was clearly represented and annotated in the anchor set, which was used for reference by raters throughout the project.

Training sets and qualifying sets contained student responses consensus-scored by rangefinding committee members. Raters were instructed on how to apply the scoring guidelines and were required to demonstrate a clear comprehension of each anchor set by performing well on the associated training materials. Responses were selected for training to show raters the range of each score point (e.g., high, mid, and low 2s). Examples of 0s were also included for all items. This process helped raters recognize the various ways that a student could respond in order to earn each score point outlined and defined in the item-specific scoring guidelines.

The scoring director conducted a team leader training session before training the raters. This session followed the same procedures as rater training, but standards were more stringent due to the extra responsibilities required of team leaders. During team leader training, all Keystone materials were reviewed and discussed. Team leaders were required to annotate all of their training materials with committee justifications from the rangefinding meetings. To facilitate scoring consistency, it was imperative that all team leaders imparted the same rationale for each response. Once the team leaders were qualified, leadership responsibilities were reviewed and team assignments were given. A ratio of one team leader per 7-10 raters ensured sufficient monitoring rates for team members.

Rater training began with the scoring director providing an intensive review of the scoring guidelines and anchor papers. Next, raters practiced by independently scoring the responses in the training sets. After each training set, the scoring director led a thorough discussion of the responses. Once the scoring guidelines, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Raters who failed to achieve at least 70 percent exact agreement on the first qualifying set were given additional, individual training. Raters who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential raters in DRC's imaging system and released from the project.

The 2017 assessment included the opportunity for students to respond in Spanish to Algebra 1 and Biology items. Training for the raters for the Spanish language responses was conducted at Tri-Lin Integrated Services and overseen by a DRC scoring director, who is a Spanish language speaker with a strong mathematics and science background, and has worked closely with the Keystone exams since they were first administered. All Spanish raters were bilingual and hired specifically to score the Spanish portion of the assessment and were required to meet the same standards set for raters of the English language version of the assessment.

Table 8–3. Qualification Rates for 2017 Keystone Open-Ended Response Items - Winter

Subject	% Qualifying On First Attempt	% Qualifying On Second Attempt	% That Did Not Qualify
Algebra I	99	1	0
Biology	96	3	1
Literature	100	0	0

Table 8–4. Qualification Rates for 2017 Keystone Open-Ended Response Items - Spring

Subject	% Qualifying On First Attempt	% Qualifying On Second Attempt	% That Did Not Qualify
Algebra I	96	4	0
Biology	91	7	2
Literature	98	2	0

Table 8–5. Qualification Rates for 2017 Keystone Open-Ended Response Items - Summer

Subject	% Qualifying On First Attempt	% Qualifying On Second Attempt	% That Did Not Qualify
Algebra I	100	0	0
Biology	91	9	0
Literature	100	0	0

HANDSCORING PROCESS

Student responses were scored independently. All responses were scored once, and ten percent of the responses were scored a second time. The data collected from the ten-percent double-read portion was used to calculate the exact and adjacent agreement rates in the Scoring Summary Reports. The responses that were used for the ten percent read behind were randomly chosen by the imaging system at the item level. Additional read behinds by the team leaders and scoring directors were done to further ensure reliability.

Raters scored the imaged student responses on PC monitors at scoring locations in Sharonville, Ohio; Columbus, Ohio; Plymouth, Minnesota; Woodbury, Minnesota; King of Prussia, Pennsylvania; Jacksonville, Florida, and San Antonio, Texas.

In all locations, raters were seated at tables with individual imaging stations. Image distribution was controlled, ensuring that student images were sent only to designated groups of raters qualified to score those items. Imaged student responses were electronically separated for routing to individual raters by item. Raters were only provided with student responses for items that they were qualified to score. Scores were keyed into DRC's imaging system.

To handle possible alerts (i.e., student responses indicating potential issues related to students' safety and well-being that sometimes require attention at the state or local level), DRC's imaging system allows raters to forward responses needing attention to the scoring director. These alerts are reviewed by project management, who then notifies the students' schools and PDE of the occurrences. PDE does not receive any identifying information about the students. At no time in the alerts process do raters, or other DRC handscoring staff, acquire any knowledge concerning a student's personal identity.

HANDSCORING VALIDITY PROCESS

One of the training tools PAS utilized to ensure rater accuracy was the validity process. The goal of the validity process is to ensure that scoring standards are maintained. Specifically, the objective is to make sure that raters score student responses in a manner consistent with statewide standards both within a single administration of the Keystones and across consecutive administrations. In scoring the 2017 Keystone Exams, scoring consistency was maintained, in part, through the validity process.

The validity process began with the selection of scored responses. Forty validity papers were selected for each core open-ended (OE) item. These 40 papers were drawn from a pool of exemplars (responses that are representative of a particular score point and have been verified by the scoring director). The scores on validity responses are considered true scores.

The validity papers were then implemented to test rater accuracy. The responses were selected within the imaging system and dispersed intermittently to the raters. By the end of the project, raters had scored all 40 validity papers for any items they were qualified to score. Raters were unaware when they were being dealt pre-scored validity responses and assumed that they were scoring live student responses. This helped bolster the internal validity of the process. All raters who received validity papers had already successfully completed the training/qualifying process.

The scores that the raters assigned to the validity papers were compared to the true scores in order to determine the validity of the raters' scores. For each item, the percentage of exact agreement as well as the percentage of high and low scores was computed. This data was accessed through the Validity Item Detail Report. The same sort of data was also computed for each specific rater. This data was accessed through the Validity Reader Detail Report. Both of these may be run as daily or cumulative reports.

The Validity Reader Detail Report was used to identify particular raters for retraining. If a rater on a certain day generated a lower rate of agreement on a group of validity papers, it was immediately apparent in the Validity Reader Detail Report. A lower rate of agreement was defined as anything below 70 percent exact agreement with the true scores. Any time a rater's validity agreement rate fell below 70 percent, the scoring director was cued to examine that rater's scoring. First, the scoring director attempted to ascertain what kind of validity papers the rater was scoring incorrectly. This was done to determine whether there was any sort of a trend (e.g., trending low on the 1–2 line). Once the source of the low agreement rate was determined, the rater was retrained. If it was determined that the rater had been scoring live responses inaccurately, then his/her scores were purged for that day, and the responses were re-circulated and scored by other raters.

The cumulative Validity Item Detail Report was utilized to identify potential room-wide trends in need of correction. For instance, if a particular validity response with a true score of 3 was given a score of 2 by a significant number of raters within the room, that trend would be revealed in the Validity Item Detail Report. To correct a trend of this sort, the scoring director would look for student responses similar to the validity response being scored incorrectly. Once located, these responses would be used in room-wide re-training, usually in the form of an annotated handout or a short set of responses without printed scores given to raters as a recalibration test.

Validity was employed on all core Algebra 1, Biology, and Literature OE items. Each 40-paper validity set was formulated to mirror the score point distribution that the item generated during its previous administration. Each validity set included at least five examples of each score point. Examples of different types of responses were included to ensure that raters were tested on the full spectrum of response types.

The exact rater agreement rate generated during the validity process was often higher than the inter-rater agreement rate for the same item. The reason for this discrepancy has to do with how validity sets are formulated. The 40 validity responses for each item are intended to cover the full breadth of each score point. For example, each validity set contains examples of high, mid, and low 2s. This scope ensures that the validity process is truly valid in terms of addressing the complete spectrum of response types. However, certain types of responses are generally not included in validity sets. These include line responses (i.e., examples of score points that are so close to the adjacent score point that raters are instructed to consult with a supervisor before assigning a score) and responses that, because of poor word choice/writing, are difficult to understand. The reason for these exclusions is that confusing/line/illegible papers often do not impart a teachable lesson. Since these types of responses are usually unique, any potential lesson the response might teach would apply only to that particular response. Conversely, the responses in validity sets are chosen because they represent common response types and teach lessons that can be applied to other similar papers. Due to this distinction, validity sets often generate a slightly higher agreement rate than is typically generated during operational scoring.

QUALITY CONTROL

Rater accuracy was monitored throughout the scoring session by means of daily and on-demand reports. These reports ensured that an acceptable level of scoring accuracy was maintained throughout the project. Inter-rater reliability was tracked and monitored with multiple quality control reports that were reviewed by quality assurance analysts. These reports and other quality control documents were generated at the scoring centers, where they were reviewed by the scoring directors, team leaders, and project managers. The following reports and documents were used during the scoring of the open-ended items:

The Scoring Summary Report (includes two related reports).

- The Reader Monitor Report monitored how often raters were in exact agreement with one another and ensured that an acceptable agreement rate was maintained. This report provided daily and cumulative exact and adjacent inter-rater agreement on the ten percent that was double read.
- The Score Point Distribution Report monitored the percentage of responses given each of the score points. For example, the Algebra 1 daily and cumulative reports showed what percentage of 0s, 1s, 2s, 3s, and 4s a rater had given to all the responses scored at the time the report was produced. It also indicated the number of responses read by each rater so that production rates could be monitored.

The Item Status Report monitored the progress of handscoring. This report tracked each response and indicated the status (e.g., not read, complete, awaiting supervisor review, etc.). This report ensured that all responses were scored by the end of the project.

The Reader Score Report identified all responses scored by an individual rater. This report was useful if any responses needed rescoring because of possible rater drift.

The Validity Reports (addressed in detail on previous page) tracked how raters performed by comparing pre-scored responses to raters' scores for the same responses. If a rater's scoring fell below the 70 percent determined agreement rate, remediation occurred. Raters who did not retrain to the required level of agreement were released from the project.

The Read-Behind Log was used by the team leader/scoring director to monitor individual rater reliability. Team leaders read randomly-selected, scored items from each team member. If the team leader disagreed with a rater's score, remediation occurred. This proved to be a very effective type of feedback because it was done with live items scored by a particular rater.

Recalibration Sets were used throughout the scoring sessions to ensure accuracy by comparing each rater's scores with the true scores on a pre-selected set of responses. Recalibration sets helped to refocus raters on Pennsylvania scoring standards. This check made sure there was no change in the scoring pattern as the project progressed. Raters failing to achieve 70 percent agreement with the recalibration true scores were given additional training to achieve the highest degree of accuracy possible. Raters who were unable to recalibrate were released from the project. The process for creating and administering recalibration sets was similar to the one used for training sets.

Table 8–6. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Winter 2017

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/NS
Alg. 1	1	616748	A	0-2	99	1	98	65	1	24			9
Alg. 1	1	616748	B	0-2	99	1	100	46	15	29			9
Alg. 1	1	700872		0-4	93	7	95	35	40	6	3	1	16
Alg. 1	1	672277	A	0-2	97	3	98	55	14	17			13
Alg. 1	1	672277	B	0-1	100	0	100	71	16				13
Alg. 1	1	672277	C	0-1	100	0	100	38	49				13
Alg. 1	2	701634	A	0-1	99	1	99	10	81				8
Alg. 1	2	701634	B	0-1	97	3	99	47	44				8
Alg. 1	2	701634	C	0-1	99	1	99	74	17				8
Alg. 1	2	701634	D	0-1	97	3	99	62	29				8
Alg. 1	2	644984	A	0-1	99	1	100	38	52				10
Alg. 1	2	644984	B	0-1	100	0	100	78	12				10
Alg. 1	2	644984	C	0-1	100	0	99	89	1				10
Alg. 1	2	644984	D	0-1	100	0	100	81	9				10
Alg. 1	2	644984	MU	0-1	100	0	100	90	0				10
Alg. 1	2	704112	A	0-1	99	1	99	55	33				12
Alg. 1	2	704112	B	0-1	100	0	100	74	14				12
Alg. 1	2	704112	C	0-1	100	0	100	77	10				12
Alg. 1	2	704112	D	0-1	98	2	100	72	16				12
Bio.	1	703365		0-3	87	13	96	21	32	25	8		14
Bio.	1	703537		0-3	91	9	92	58	14	8	5		15
Bio.	1	703528		0-3	91	9	95	43	20	15	6		16
Bio.	2	607815		0-3	82	17	85	22	34	19	10		15
Bio.	2	703605		0-3	85	15	91	37	28	14	4		13
Bio.	2	705264		0-3	87	12	88	19	40	20	9		12
Lit.	1	643730		0-3	82	18	86	24	33	26	5		13
Lit.	1	643178		0-3	89	11	95	12	32	37	4		15
Lit.	1	643179		0-3	85	15	91	14	34	27	8		17
Lit.	2	644041		0-3	81	19	78	15	19	37	16		14
Lit.	2	677768		0-3	85	15	92	4	31	43	8		13
Lit.	2	644767		0-3	89	11	90	9	35	32	5		19

Notes: B = blank; NS = non-scorable. Algebra I responses received a total of 0–4 points. For some Algebra I items, readers applied a single score of 0, 1, 2, 3, or 4. Many of the Algebra I items, however, were divided into separate parts that were scored on a 0–1, 0–2, or 0–3 point scale, although the sum of the parts always resulted in an overall score of 0–4 for each item. For example, an Algebra I item might have a part A, part B, part C, and part D, each of which was scored on a 0–1 point scale, which results in a summed 0–4 point total score for each response. Furthermore, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) if they did not receive points for any of the individual parts.

Table 8–7. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Spring 2017

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/ NS
Alg. 1	1	724679	A	0-1	98	2	99	27	66				7
Alg. 1	1	724679	B	0-2	99	1	100	28	65				7
Alg. 1	1	724679	C	0-1	100	0	99	90	3				7
Alg. 1	1	724679	D	0-1	99	1	100	44	49				7
Alg. 1	1	730202	A	0-1	99	1	99	68	23				9
Alg. 1	1	730202	B	0-1	99	1	97	81	10				9
Alg. 1	1	730202	C	0-1	100	0	100	69	21				9
Alg. 1	1	730202	D	0-1	100	0	99	72	18				9
Alg. 1	1	674404		0-4	88	12	88	23	28	12	9	15	12
Alg. 1	2	730208		0-4	84	16	92	11	25	24	22	10	8
Alg. 1	2	666526	A	0-1	99	1	99	36	56				8
Alg. 1	2	666526	B	0-1	99	1	98	53	39				8
Alg. 1	2	666526	C	0-1	100	0	100	85	8				8
Alg. 1	2	666526	D	0-1	100	0	100	88	4				8
Alg. 1	2	739459		0-4	90	10	90	16	20	42	10	1	10
Bio.	1	736837		0-3	88	10	97	40	11	21	19		9
Bio.	1	679989		0-3	83	16	87	29	25	21	14		11
Bio.	1	735487		0-3	76	23	88	23	29	28	9		12
Bio.	2	736555		0-3	96	4	98	29	21	27	14		9
Bio.	2	736552		0-3	86	14	86	16	29	33	12		9
Bio.	2	736553		0-3	89	11	95	28	21	18	16		17
Lit.	1	735336		0-3	77	23	86	12	30	40	9		10
Lit.	1	643181		0-3	80	20	85	24	20	31	14		11
Lit.	1	643182		0-3	79	20	84	10	22	42	13		12
Lit.	2	683334		0-3	76	24	85	8	24	44	13		11
Lit.	2	742911		0-3	84	15	88	9	24	48	7		12
Lit.	2	742912		0-3	81	19	85	11	30	36	10		13

Notes: B = blank; NS = non-scorable. Algebra I responses received a total of 0–4 points. For some Algebra I items, readers simply applied a single score of 0, 1, 2, 3, or 4. Many of the Algebra I items, however, were divided into separate parts that were scored on a 0–1, 0–2, or 0–3 point scale, although the sum of the parts always resulted in an overall score of 0–4 for each item. For example, an Algebra I item might have a part A, part B, part C, and part D, each of which was scored on a 0–1 point scale, which results in a summed 0–4 point total score for each response. Furthermore, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) if they did not receive points for any of the individual parts.

Table 8–8. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Summer 2017

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/ NS
Alg. 1	1	724699		0-4	92	8	98	14	44	18	17	3	3
Alg. 1	1	681303	A	0-1	99	1	99	18	79				3
Alg. 1	1	681303	B	0-1	100	0	100	79	18				3
Alg. 1	1	681303	C	0-1	100	0	100	96	1				3
Alg. 1	1	681303	D	0-1	100	0	99	96	1				3
Alg. 1	1	701631		0-4	91	9	96	43	42	9	1	1	3
Alg. 1	2	734692		0-4	98	2	94	10	34	26	27	0	3
Alg. 1	2	714265	A	0-1	99	1	99	89	8				3
Alg. 1	2	714265	B	0-1	100	0	100	43	53				3
Alg. 1	2	714265	C	0-1	99	1	100	85	12				3
Alg. 1	2	714265	D	0-1	100	0	100	72	25				3
Alg. 1	2	678819		0-4	93	7	97	20	43	22	8	2	5
Bio.	1	736839		0-3	87	12	97	25	45	22	3		4
Bio.	1	742851		0-3	96	4	95	31	42	20	4		3
Bio.	1	741445		0-3	92	8	98	24	29	27	16		3
Bio.	2	737711		0-3	97	3	96	35	37	20	4		5
Bio.	2	703003		0-3	84	15	93	20	32	26	18		4
Bio.	2	677890		0-3	81	19	85	34	44	12	6		4
Lit.	1	643960		0-3	81	19	90	8	33	44	11		4
Lit.	1	703918		0-3	78	22	93	14	50	30	2		3
Lit.	1	703919		0-3	87	13	85	13	55	24	2		6
Lit.	2	683634		0-3	73	27	85	13	39	40	4		3
Lit.	2	683411		0-3	79	21	95	14	50	31	1		3
Lit.	2	683414		0-3	85	15	91	8	53	32	3		4

Notes: B = blank; NS = non-scorable. Algebra I responses received a total of 0–4 points. For some Algebra I items, readers applied a single score of 0, 1, 2, 3, or 4. Many of the Algebra I items, however, were divided into separate parts that were scored on a 0–1, 0–2, or 0–3 point scale, although the sum of the parts always resulted in an overall score of 0–4 for each item. For example, an Algebra I item might have a part A, part B, part C, and part D, each of which was scored on a 0–1 point scale, which results in a summed 0–4 point total score for each response. Furthermore, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) if they did not receive points for any of the individual parts.

CHAPTER NINE: DESCRIPTION OF DATA SOURCES

This section describes the filtering process and data sources used for the various analysis procedures discussed in the remaining sections of this report. Psychometric analyses were conducted at several points for the Winter 2016/2017, Spring 2017, and Summer 2017 Pennsylvania Keystone Exams in Algebra I, Biology, and Literature: 1) key verification analyses for quality-control purposes; 2) post-equating check; 3) item analysis and calibration of field test items embedded in the spring forms; and 4) analyses for this technical report.

STUDENT FILTERING CRITERIA

Students' records included in all the psychometric analyses needed to meet at least the following psychometric analyses criteria:

- Module 1 Attempted Status = 1 (1 = the student attempted a minimum of five items in Module 1)
- Module 2 Attempted Status = 1 (1 = the student attempted a minimum of five items in Module 2)
- Module 1 Invalidated = N (N = the student's score was not invalidated)
- Module 2 Invalidated = N (N = the student's score was not invalidated)
- Student Duplication Status = N (N = no duplication)
- Module 1 Form Number = Module 2 Form Number
- Module_1_Form_Name ≠ Form (exclude the VSL form)
- Module_2_Form_Name ≠ Form (exclude the VSL form)

For each specific analysis conducted at different times, additional criteria might be needed to filter students. For example, the following criteria were used in addition to the ones listed above for the post-equating check, since the analyses were conducted during the scoring window:

- Module 1 Complete Status = 01
- Module 2 Complete Status = 01

The value 01 represents the response string which includes scores on the multiple-choice (MC) and constructed-response (CR) operational items. When the analyses were conducted by using the final data files, these criteria were no longer necessary since all operational CR items had been scored.

Item analysis and calibration of embedded field test items were conducted using the first-time testers only (i.e., retester = N). The classical item statistics for the field test items analyzed by using the first-time testers were more comparable to the results of the Spring 2011 Keystone Exams, which were given to the first-time test takers. Students who took form 1 and with accommodations were removed from the analyses so the results can be comparable to other forms which did not provide accommodations.

Because a large number of students took the Keystone Exams, only a representative sample of students' responses on field test CR items was scored within each content area. For the item analysis of field test CR items, the following additional criteria were used to select only those who were sampled for hand-scoring:

- Module 1 CR Score Sample = Y
- Module 2 CR Score Sample = Y

For the analyses such as reliability analyses that used the final data files, the following filtering criteria were used as well. Any student with an exclusion code 00 received a scale score.

- Module 1 Exclusion Code = 00
- Module 2 Exclusion Code = 00

KEY VERIFICATION DATA

The key verification data are mentioned only for completeness, as no formal results are provided in this technical document. A key verification is often conducted early in the scoring process to ensure the keys for the MC items are applied correctly. The data files used for the key verification analysis are usually (but not always) based on the student data from early-return schools. The sample representativeness is not required for this internal quality check. Available student data typically suffices as long as there is reasonable variability in the total-test scores of students. The details about the sample sizes for the winter, spring, and summer administrations can be found in Table 9–1.

CALIBRATION OF OPERATIONAL TEST DATA

The post-equating check data included all students who met the inclusion criteria and were scored by 02/20/2017, 06/29/2017, and 08/29/2016 for the winter, spring, and summer administrations respectively. Note that the students included in the post-equating check data included those who had testing accommodations.

FINAL DATA

The final data files were used to conduct item analyses for the operational items and analyses conducted for Chapters 16–19 in this technical report. The final data contained students' responses to both the MC and CR items. All students' responses included in the analyses met the filtering criteria. The final sample sizes (or *N* counts) can be found in the column labeled "Final" in Table 9–1.

Table 9–1. Data Source *N*-Counts

Administration	Content Area	Key Verification	Post-Equating Check	Final
Winter	Algebra I	11,074	55,510	55,500
Winter	Biology	7,555	41,924	41,916
Winter	Literature	8,000	36,552	36,551
Spring	Algebra I	156,136	167,097	167,097
Spring	Biology	129,611	142,245	142,262
Spring	Literature	125,568	130,310	130,328
Summer	Algebra I	1,454	1,448	1,454
Summer	Biology	960	942	960
Summer	Literature	446	444	446

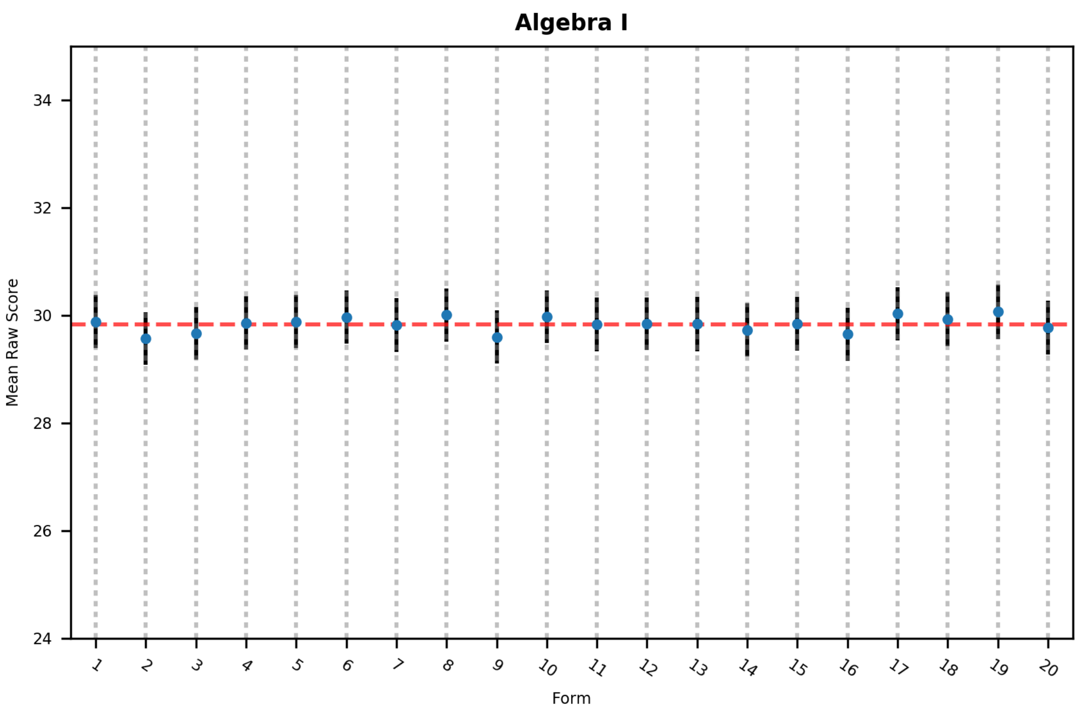
SPIRALING OF FORMS

During the administration of Keystone Exams, test forms were spiraled at the student level. The goal of spiraling is to achieve equivalent samples of students across forms so the classical statistics (e.g., *p*-value and point-biserial correlation) for all the field test items can be compared. Given that the field test items were embedded in the spring administration only, the equivalence of samples was checked for the spring administration instead of all administrations. When spiraling achieves randomly equivalent samples, the forms will have equal means (within sampling error) over the operational items.

Appendix H provides summary statistics for all the spring forms for each content area exam. The tables provide the form number (Form), number of students (*N*), test length in items (L), total points (Pts.), minimum (Min) score, maximum (Max) score, mean (Mean) score, median (Med) score, and standard deviation (SD). The extent to which the mean raw scores across forms are similar indicates the extent to which the student populations taking each form are of approximately equal ability. This equivalence of ability distributions across forms is the desired outcome of spiraling and allows for optimum analysis of the embedded field test items.

In Figure 9–1, the form mean raw scores are plotted (circle-shaped marker) with standard error of mean lines. For each form, the standard error of mean was computed by taking the standard deviation of all student scores (assumed as the population standard deviation divided by the square root of the form *n*-count). The mean score across all forms is indicated by the red horizontal broken line. If the three standard error band captures the horizontal line, then that suggests only random differences exist between the form mean and the population mean. This is true for all forms in all content areas.

Figure 9–1. Form Mean Scores with +/- Three Standard Error (SE) Bands



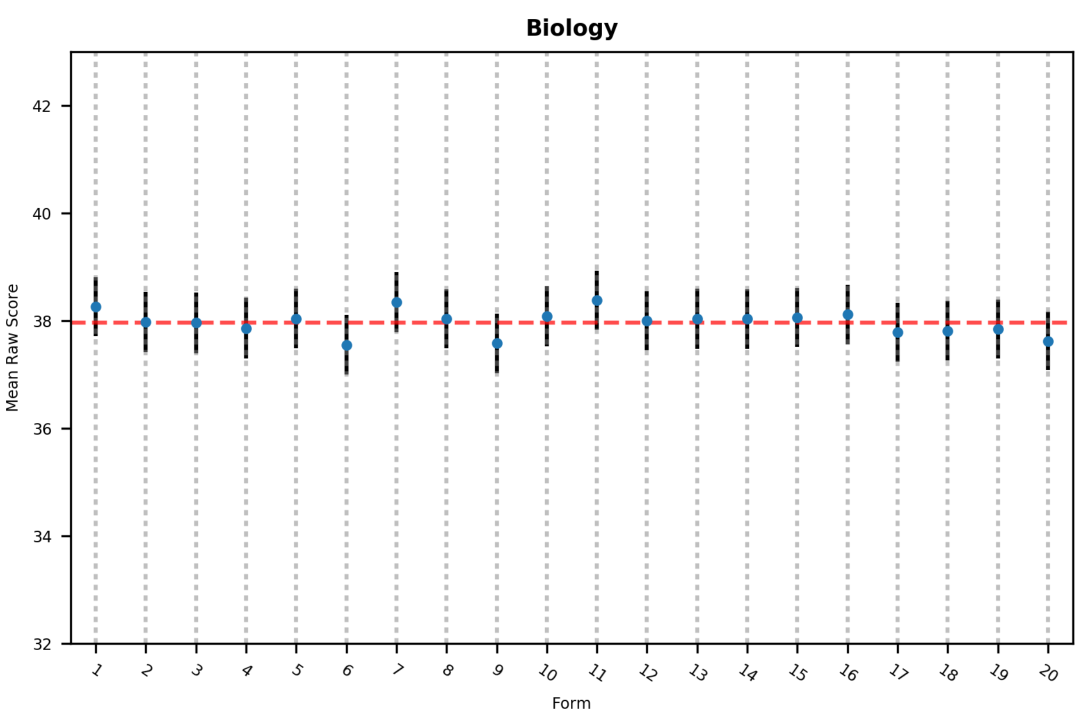
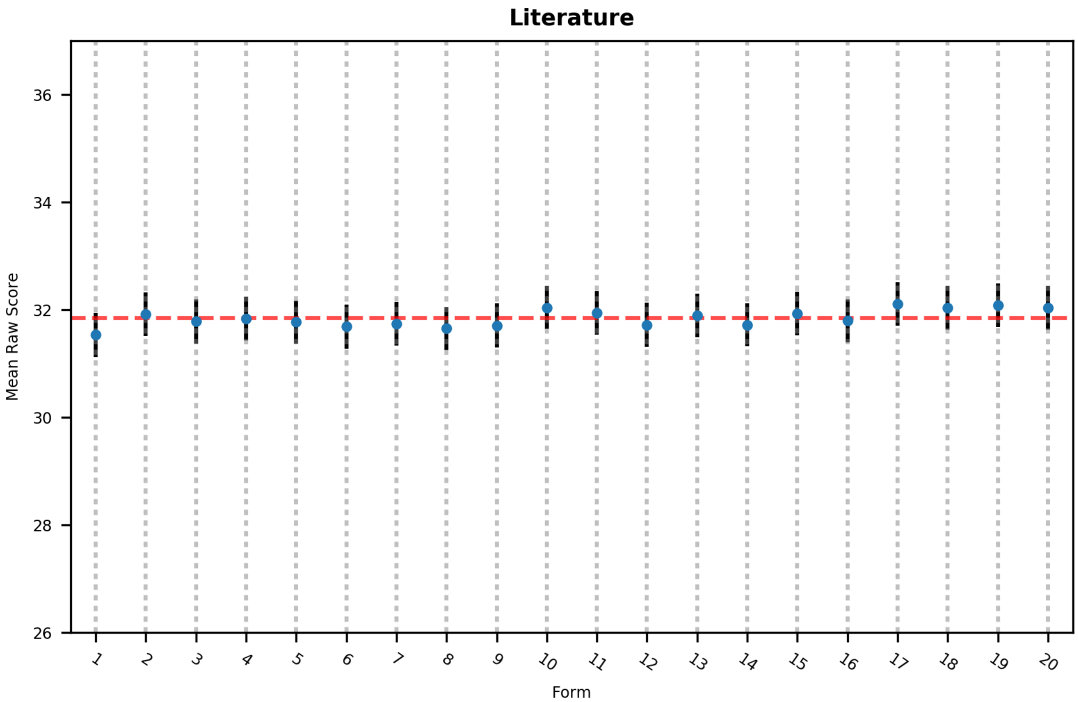


Figure 9–1 (continued). Form Mean Scores with +/- Three Standard Error (SE) Bands



CHAPTER TEN: SUMMARY DEMOGRAPHIC AND ACCOMMODATION DATA FOR SPRING 2017 KEYSTONE EXAMS

ASSESSED STUDENTS

Students assessed on the Keystone Exams include students from public schools who are required to participate by virtue of being in the graduating class of 2019, students in a school district planning to use the Keystone Exams to meet graduation requirements, and students enrolled in Algebra I, Biology, or Literature during the 2016–2017 school year. The operational Keystone Exams were administered in both paper-and-pencil test (PPT) and computer-based test (CBT) formats.

Statistical tables and associated commentary embedded in this chapter are based on data from the Spring 2017 test administration of the Keystone Exams. There were two other administrations during the school year, each of which involved fewer students than the spring. One occurred during Winter 2016/2017 and the other in the Summer 2017. Tables summarizing results from these two administrations can be found in Appendix I.

Results for this chapter are presented in sets of tables for the three Keystone Exams administered in Spring 2017 (Algebra I, Biology, and Literature). Accompanying each numbered table is a letter (A, B, or L) to designate the content area. Tables 10–1A through 10–1L provides a summary of tests processed and scored, which are displayed separately by student grade level. The first two rows present the number processed for each administration mode (PPT and CBT). The total number of tests processed is presented on the third row. The fourth row shows the number and percentage of students with a Keystone Exam score, while the fifth row presents the number and percentage not receiving a score. Please note that the percent of students assessed (received a total score) is typically in the high 90s across grade levels.

Table 10–1A. Students Assessed on the Spring 2017 Keystone Exam: Algebra I

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	86	372	7,853	32,628	60,025	37,120	18,464	248	156,796
Total number of CBT processed (Number)	0	27	779	2,881	4,771	3,610	2,017	71	14,156
Total number of tests processed (Number)	86	399	8,632	35,509	64,796	40,730	20,481	319	170,952
Total number of tests processed with a score (Number)	67	398	8,607	35,293	62,557	38,422	18,211	277	163,832
Total number of tests processed with a score (Percent)	77.9	99.7	99.7	99.4	96.5	94.3	88.9	86.8	95.8
Total number of tests processed without a score (Number)	19	1	25	216	2,239	2,308	2,270	42	7,120
Total number of tests processed without a score (Percent)	22.1	0.3	0.3	0.6	3.5	5.7	11.1	13.2	4.2

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Table 10–1B. Students Assessed on the Spring 2017 Keystone Exam: Biology

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	48	212	46,082	62,469	19,954	261	129,026
Total number of CBT processed (Number)	0	2	5,476	9,091	2,883	116	17,568
Total number of tests processed (Number)	48	214	51,558	71,560	22,837	377	146,594
Total number of tests processed with a score (Number)	38	212	50,275	68,639	20,129	321	139,614
Total number of tests processed with a score (Percent)	79.2	99.1	97.5	95.9	88.1	85.1	95.2
Total number of tests processed without a score (Number)	10	2	1,283	2,921	2,708	56	6,980
Total number of tests processed without a score (Percent)	20.8	0.9	2.5	4.1	11.9	14.9	4.8

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Table 10–1L. Students Assessed on the Spring 2017 Keystone Exam: Literature

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	41	58	8,017	94,572	17,821	165	120,674
Total number of CBT processed (Number)	0	0	232	10,417	2,533	64	13,246
Total number of tests processed (Number)	41	58	8,249	104,989	20,354	229	133,920
Total number of tests processed with a score (Number)	32	56	7,596	101,779	17,953	191	127,607
Total number of tests processed with a score (Percent)	78.0	96.6	92.1	96.9	88.2	83.4	95.3
Total number of tests processed without a score (Number)	9	2	653	3,210	2,401	38	6,313
Total number of tests processed without a score (Percent)	22.0	3.4	7.9	3.1	11.8	16.6	4.7

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

REASONS FOR STUDENT NON-ASSESSMENT

As observed from the bottom row of Table 10–1, a small percent of students were not assessed. Although there are a variety of reasons for this, the major ones pertain to:

- Extended absence from school that continued beyond the assessment window.
- Absence without make-up for at least one section of a test.
- Failure to meet the attempt criteria on one or more test modules and no exclusion code marked by school personnel. The attempt criteria required a minimum of five items to be completed in each module.
- Medical emergency.
- Parental request due to a religious reason.
- Other reasons.

The number of students without a total test score for each of these reasons is provided in Tables 10–2A through 10–2L. Associated with this number is the percent of the total of non-assessed students in each column (grade level) attributed to a particular reason.

Table 10–2A. Counts/Percentages of Students without Scores on the Spring 2017 Keystone Exam: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	3	0	4	47	586	513	468	19	1,640
Extended absence from school (Percent)	15.8	0.0	16.0	21.8	26.2	22.2	20.6	45.2	23.0
Absent without make-up (Number)	1	0	0	14	416	431	490	3	1,355
Absent without make-up (Percent)	5.3	0.0	0.0	6.5	18.6	18.7	21.6	7.1	19.0
Non-attempt (Number)	7	0	2	25	645	660	496	6	1,841
Non-attempt (Percent)	36.8	0.0	8.0	11.6	28.8	28.6	21.9	14.3	25.9
Medical emergency (Number)	0	1	10	61	136	97	75	0	380
Medical emergency (Percent)	0.0	100.0	40.0	28.2	6.1	4.2	3.3	0.0	5.3
Parental request - Chapter 4 (Number)	1	0	7	35	121	167	185	0	516
Parental request - Chapter 4 (Percent)	5.3	0.0	28.0	16.2	5.4	7.2	8.1	0.0	7.2
Parental request - Other reasons (Number)	0	0	1	16	58	93	152	1	321
Parental request - Other reasons (Percent)	0.0	0.0	4.0	7.4	2.6	4.0	6.7	2.4	4.5
Other reasons (Number)	7	0	1	18	277	347	404	13	1,067
Other reasons (Percent)	36.8	0.0	4.0	8.3	12.4	15.0	17.8	31.0	15.0
Total not assessed (Number)	19	1	25	216	2,239	2,308	2,270	42	7,120

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Table 10–2B. Counts/Percentages of Students without Scores on the Spring 2017 Keystone Exam: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	1	1	409	694	553	21	1,679
Extended absence from school (Percent)	10.0	50.0	31.9	23.8	20.4	37.5	24.1
Absent without make-up (Number)	2	0	175	548	515	4	1,244
Absent without make-up (Percent)	20.0	0.0	13.6	18.8	19.0	7.1	17.8
Non-attempt (Number)	1	0	275	798	546	18	1,638
Non-attempt (Percent)	10.0	0.0	21.4	27.3	20.2	32.1	23.5
Medical emergency (Number)	0	0	81	186	74	0	341
Medical emergency (Percent)	0.0	0.0	6.3	6.4	2.7	0.0	4.9
Parental request - Chapter 4 (Number)	1	1	91	190	214	2	499
Parental request - Chapter 4 (Percent)	10.0	50.0	7.1	6.5	7.9	3.6	7.1
Parental request - Other reasons (Number)	0	0	55	112	240	1	408
Parental request - Other reasons (Percent)	0.0	0.0	4.3	3.8	8.9	1.8	5.8
Other reasons (Number)	5	0	197	393	566	10	1,171
Other reasons (Percent)	50.0	0.0	15.4	13.5	20.9	17.9	16.8
Total not assessed (Number)	10	2	1,283	2,921	2,708	56	6,980

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Table 10–2L. Counts/Percentages of Students without Scores on the Spring 2017 Keystone Exam: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	1	279	833	517	13	1,643
Extended absence from school (Percent)	0.0	50.0	42.7	26.0	21.5	34.2	26.0
Absent without make-up (Number)	0	1	80	561	455	3	1,100
Absent without make-up (Percent)	0.0	50.0	12.3	17.5	19.0	7.9	17.4
Non-attempt (Number)	3	0	140	810	554	4	1,511
Non-attempt (Percent)	33.3	0.0	21.4	25.2	23.1	10.5	23.9
ELL in first year in U.S. schools (Number)	0	0	2	52	41	0	95
ELL in first year in U.S. schools (Percent)	0.0	0.0	0.3	1.6	1.7	0.0	1.5
Medical emergency (Number)	0	0	13	243	72	0	328
Medical emergency (Percent)	0.0	0.0	2.0	7.6	3.0	0.0	5.2
Parental request - Chapter 4 (Number)	1	0	14	225	168	2	410
Parental request - Chapter 4 (Percent)	11.1	0.0	2.1	7.0	7.0	5.3	6.5
Parental request - Other reasons (Number)	0	0	13	133	156	2	304
Parental request - Other reasons (Percent)	0.0	0.0	2.0	4.1	6.5	5.3	4.8
Other reasons (Number)	5	0	112	353	438	14	922
Other reasons (Percent)	55.6	0.0	17.2	11.0	18.2	36.8	14.6
Total not assessed (Number)	9	2	653	3,210	2,401	38	6,313

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

DEMOGRAPHIC CHARACTERISTICS OF STUDENTS RECEIVING TEST SCORES

COMPOSITION OF SAMPLE USED IN SUBSEQUENT TABLES

The following state summary statistic data analyses were completed using the final individual student data file containing records from the Spring 2017 administration, which was provided to the Pennsylvania Department of Education in July 2017. State summary statistics were based on students who received a total test score on the Spring 2017 administration with the exception of students who attended non-public schools or those who were home schooled. Also, students who took the non-Keystone Proficient assessments were excluded.

Demographic data for students taking the Keystone Exams is presented separately for each course (Tables 10–3A, 10–3B, 10–3L). Results for accommodations received were collected separately by course and are presented in separate tables as well. For example, tables involving accommodations for Biology are found in Tables 10–4B, 10–5B, 10–6B, and 10–7B. Similar data from the Winter 2016/2017 and Summer 2017 test administrations can be found in Appendix I.

COLLECTION OF STUDENT DEMOGRAPHIC INFORMATION

Data for analyses involving demographic characteristics were obtained primarily from information supplied by school district personnel through the Pennsylvania Information Management System (PIMS) and subsequently transmitted to DRC. Some data such as accommodation information are recorded by school personnel directly on the student answer document (PPT) or in eDIRECT Test Setup (CBT) at the time a Keystone Exam is administered.

DEMOGRAPHIC CHARACTERISTICS

Frequency data for each demographic category is presented in Tables 10–3A through 10–3L. Data is presented by grade level with PPT and CBT formats combined into a single composite. Shown at the bottom of the appropriate table is the number of assessed students contributing to summary statistics on which the column percentages are based.

Table 10–3A. Demographic Characteristics of Students taking the Spring 2017 Keystone Exam: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	11	116	4,044	18,676	30,826	18,539	8,487	105	80,804
Female (Percent)	16.4	29.1	47.0	52.9	49.3	48.3	46.6	37.9	49.3
Male (Number)	27	281	4,562	16,605	31,714	19,872	9,710	172	82,943
Male (Percent)	40.3	70.6	53.0	47.0	50.7	51.7	53.3	62.1	50.6
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	9	42	98	82	33	1	265
American Indian/Alaskan Native (not Hispanic) (Percent)	0.0	0.0	0.1	0.1	0.2	0.2	0.2	0.4	0.2
Asian (not Hispanic) (Number)	6	138	1,034	1,926	1,896	812	422	5	6,239
Asian (not Hispanic) (Percent)	9.0	34.7	12.0	5.5	3.0	2.1	2.3	1.8	3.8
Black or African American (not Hispanic) (Number)	16	2	232	2,174	11,363	7,957	4,059	79	25,882
Black or African American (not Hispanic) (Percent)	23.9	0.5	2.7	6.2	18.2	20.7	22.3	28.5	15.8
Hispanic (any race) (Number)	3	10	263	1,756	7,765	5,994	2,995	63	18,849
Hispanic (any race) (Percent)	4.5	2.5	3.1	5.0	12.4	15.6	16.4	22.7	11.5
Multi-Racial (not Hispanic) (Number)	1	14	198	800	1,678	1,000	491	12	4,194
Multi-Racial (not Hispanic) (Percent)	1.5	3.5	2.3	2.3	2.7	2.6	2.7	4.3	2.6
White (not Hispanic) (Number)	12	232	6,867	28,550	39,673	22,532	10,176	117	108,159
White (not Hispanic) (Percent)	17.9	58.3	79.8	80.9	63.4	58.6	55.9	42.2	66.0
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	1	3	30	62	26	19	0	141
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0.0	0.3	0.0	0.1	0.1	0.1	0.1	0.0	0.1
IEP (not gifted) (Number)	2	12	180	1,085	8,842	9,393	5,333	113	24,960
IEP (not gifted) (Percent)	3.0	3.0	2.1	3.1	14.1	24.4	29.3	40.8	15.2
Student exited IEP in last 2 years (Number)	1	12	218	675	1,071	549	258	1	2,785
Student exited IEP in last 2 years (Percent)	1.5	3.0	2.5	1.9	1.7	1.4	1.4	0.4	1.7
Title I (Number)	11	14	522	5,212	16,478	11,559	5,244	186	39,226
Title I (Percent)	16.4	3.5	6.1	14.8	26.3	30.1	28.8	67.1	23.9
Title III served (Number)	2	0	9	131	2,278	2,259	1,402	36	6,117
Title III served (Percent)	3.0	0.0	0.1	0.4	3.6	5.9	7.7	13.0	3.7
Title III not served (Number)	0	0	0	3	10	10	11	0	34
Title III not served (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Migrant student (Number)	0	0	0	2	17	17	10	0	46
Migrant student (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
ELL (enrolled after 5/27/16) (Number)	1	0	4	32	521	369	344	5	1,276
ELL (enrolled after 5/27/16) (Percent)	1.5	0.0	0.0	0.1	0.8	1.0	1.9	1.8	0.8

Table 10–3A (continued). Demographic Characteristics of Students taking the Spring 2017 Keystone Exam: Algebra

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
ELL (enrolled on or before 5/27/16) (Number)	2	0	4	116	1,869	2,013	1,131	33	5,168
ELL (enrolled on or before 5/27/16) (Percent)	3.0	0.0	0.0	0.3	3.0	5.2	6.2	11.9	3.2
Exited ESL/bilingual program and in first year of monitoring (Number)	0	1	12	89	174	139	99	0	514
Exited ESL/bilingual program and in first year of monitoring (Percent)	0.0	0.3	0.1	0.3	0.3	0.4	0.5	0.0	0.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	2	20	98	163	114	75	0	472
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0.0	0.5	0.2	0.3	0.3	0.3	0.4	0.0	0.3
Former ELL no longer monitored (Number)	1	6	97	719	1,567	950	359	0	3,699
Former ELL no longer monitored (Percent)	1.5	1.5	1.1	2.0	2.5	2.5	2.0	0.0	2.3
Foreign exchange student (Number)	0	0	0	2	3	7	6	0	18
Foreign exchange student (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Economically disadvantaged (Number)	6	20	1,026	8,973	31,898	22,157	10,723	207	75,010
Economically disadvantaged (Percent)	9.0	5.0	11.9	25.4	51.0	57.7	58.9	74.7	45.8
Historically Underperforming Subgroup (Number)	11	30	1,182	9,713	36,097	26,230	13,091	233	86,587
Historically Underperforming Subgroup (Percent)	16.4	7.5	13.7	27.5	57.7	68.3	71.9	84.1	52.9
Enrollment in school of residence after 10/1/16 (Number)	8	1	44	395	3,132	2,144	1,401	55	7,180
Enrollment in school of residence after 10/1/16 (Percent)	11.9	0.3	0.5	1.1	5.0	5.6	7.7	19.9	4.4
Enrollment in district of residence after 10/1/16 (Number)	5	1	31	316	2,312	1,670	1,158	51	5,544
Enrollment in district of residence after 10/1/16 (Percent)	7.5	0.3	0.4	0.9	3.7	4.3	6.4	18.4	3.4
Enrollment as PA resident after 10/1/16 (Number)	2	1	16	158	1,309	899	655	16	3,056
Enrollment as PA resident after 10/1/16 (Percent)	3.0	0.3	0.2	0.4	2.1	2.3	3.6	5.8	1.9
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	2	195	1,436	3,711	26,747	9,218	3,251	66	44,626
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	3.0	49.0	16.7	10.5	42.8	24.0	17.9	23.8	27.2

Table 10–3A (continued). Demographic Characteristics of Students taking the Spring 2017 Keystone Exam: Algebra

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	1	15	226	1,199	8,270	3,666	2,039	65	15,481
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	1.5	3.8	2.6	3.4	13.2	9.5	11.2	23.5	9.4
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Court/agency placed (Number)	0	0	0	2	30	44	36	30	142
Court/agency placed (Percent)	0.0	0.0	0.0	0.0	0.0	0.1	0.2	10.8	0.1
Number of assessed students (Number)	67	398	8,607	35,293	62,557	38,422	18,211	277	163,832

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Table 10–3B. Demographic Characteristics of Students Taking the Spring 2017 Keystone Exam: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	9	92	25,737	33,411	9,657	133	69,039
Female (Percent)	23.7	43.4	51.2	48.7	48	41.4	49.4
Male (Number)	10	120	24,528	35,217	10,469	188	70,532
Male (Percent)	26.3	56.6	48.8	51.3	52.0	58.6	50.5
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	73	119	35	1	228
American Indian/Alaskan Native (not Hispanic) (Percent)	0.0	0.0	0.1	0.2	0.2	0.3	0.2
Asian (not Hispanic) (Number)	1	14	2,616	2,032	518	9	5,190
Asian (not Hispanic) (Percent)	2.6	6.6	5.2	3.0	2.6	2.8	3.7
Black or African American (not Hispanic) (Number)	6	1	5,488	10,510	4,075	100	20,180
Black or African American (not Hispanic) (Percent)	15.8	0.5	10.9	15.3	20.2	31.2	14.5
Hispanic (any race) (Number)	1	4	3,514	7,887	3,259	54	14,719
Hispanic (any race) (Percent)	2.6	1.9	7.0	11.5	16.2	16.8	10.5
Multi-Racial (not Hispanic) (Number)	0	3	1,159	1,564	525	12	3,263
Multi-Racial (not Hispanic) (Percent)	0.0	1.4	2.3	2.3	2.6	3.7	2.3
White (not Hispanic) (Number)	11	190	37,372	46,450	11,689	145	95,857
White (not Hispanic) (Percent)	28.9	89.6	74.3	67.7	58.1	45.2	68.7
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	40	57	24	0	121
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0.0	0.0	0.1	0.1	0.1	0.0	0.1
IEP (not gifted) (Number)	1	6	4,864	11,749	4,776	126	21,522
IEP (not gifted) (Percent)	2.6	2.8	9.7	17.1	23.7	39.3	15.4
Student exited IEP in last 2 years (Number)	0	6	834	891	246	2	1,979
Student exited IEP in last 2 years (Percent)	0.0	2.8	1.7	1.3	1.2	0.6	1.4
Title I (Number)	7	45	8,905	14,650	5,699	183	29,489
Title I (Percent)	18.4	21.2	17.7	21.3	28.3	57.0	21.1
Title III served (Number)	0	1	801	2,371	1,368	28	4,569
Title III served (Percent)	0.0	0.5	1.6	3.5	6.8	8.7	3.3
Title III not served (Number)	0	0	4	11	15	0	30
Title III not served (Percent)	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Migrant student (Number)	0	0	9	27	14	1	51
Migrant student (Percent)	0.0	0.0	0.0	0.0	0.1	0.3	0.0
ELL (enrolled after 5/27/16) (Number)	0	1	162	377	352	3	895
ELL (enrolled after 5/27/16) (Percent)	0.0	0.5	0.3	0.5	1.7	0.9	0.6
ELL (enrolled on or before 5/27/16) (Number)	0	0	676	2,134	1,089	27	3,926
ELL (enrolled on or before 5/27/16) (Percent)	0.0	0.0	1.3	3.1	5.4	8.4	2.8
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	113	262	121	0	496

Table 10–3B (continued). Demographic Characteristics of Students Taking the Spring 2017 Keystone Exam: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Percent)	0.0	0.0	0.2	0.4	0.6	0.0	0.4
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	116	178	102	0	396
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0.0	0.0	0.2	0.3	0.5	0.0	0.3
Former ELL no longer monitored (Number)	0	1	1,178	1,727	467	1	3,374
Former ELL no longer monitored (Percent)	0.0	0.5	2.3	2.5	2.3	0.3	2.4
Foreign exchange student (Number)	0	0	2	9	10	0	21
Foreign exchange student (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Economically disadvantaged (Number)	4	25	17,584	32,168	11,252	233	61,266
Economically disadvantaged (Percent)	10.5	11.8	35.0	46.9	55.9	72.6	43.9
Historically Underperforming Subgroup (Number)	5	30	20,067	37,513	13,374	267	71,256
Historically Underperforming Subgroup (Percent)	13.2	14.2	39.9	54.7	66.4	83.2	51.0
Enrollment in school of residence after 10/1/16 (Number)	3	2	1,356	2,595	1,455	59	5,470
Enrollment in school of residence after 10/1/16 (Percent)	7.9	0.9	2.7	3.8	7.2	18.4	3.9
Enrollment in district of residence after 10/1/16 (Number)	3	2	1,116	2,072	1,214	56	4,463
Enrollment in district of residence after 10/1/16 (Percent)	7.9	0.9	2.2	3.0	6.0	17.4	3.2
Enrollment as PA resident after 10/1/16 (Number)	0	2	607	1,081	684	18	2,392
Enrollment as PA resident after 10/1/16 (Percent)	0.0	0.9	1.2	1.6	3.4	5.6	1.7
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	0	11	17,201	13,427	3,421	107	34,167
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	0.0	5.2	34.2	19.6	17.0	33.3	24.5
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	7	4,560	5,195	2,209	86	12,057
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0.0	3.3	9.1	7.6	11.0	26.8	8.6
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Court/agency placed (Number)	0	0	18	43	40	30	131
Court/agency placed (Percent)	0.0	0.0	0.0	0.1	0.2	9.3	0.1
Number of assessed students (Number)	38	212	50,275	68,639	20,129	321	139,614

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Table 10–3L. Demographic Characteristics of Students taking the Spring 2017 Keystone Exam: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	4	36	3,665	50,419	7,535	61	61,720
Female (Percent)	12.5	64.3	48.2	49.5	42.0	31.9	48.4
Male (Number)	5	20	3,929	51,347	10,414	130	65,845
Male (Percent)	15.6	35.7	51.7	50.4	58.0	68.1	51.6
American Indian/Alaskan Native (not Hispanic) (Number)	0	1	14	146	27	0	188
American Indian/Alaskan Native (not Hispanic) (Percent)	0.0	1.8	0.2	0.1	0.2	0.0	0.1
Asian (not Hispanic) (Number)	1	1	286	4,029	503	5	4,825
Asian (not Hispanic) (Percent)	3.1	1.8	3.8	4.0	2.8	2.6	3.8
Black or African American (not Hispanic) (Number)	5	5	1,851	13,271	3,830	67	19,029
Black or African American (not Hispanic) (Percent)	15.6	8.9	24.4	13.0	21.3	35.1	14.9
Hispanic (any race) (Number)	0	1	774	9,193	2,681	37	12,686
Hispanic (any race) (Percent)	0.0	1.8	10.2	9.0	14.9	19.4	9.9
Multi-Racial (not Hispanic) (Number)	0	5	187	2,291	454	9	2,946
Multi-Racial (not Hispanic) (Percent)	0.0	8.9	2.5	2.3	2.5	4.7	2.3
White (not Hispanic) (Number)	2	43	4,477	72,754	10,435	73	87,784
White (not Hispanic) (Percent)	6.3	76.8	58.9	71.5	58.1	38.2	68.8
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	1	0	5	80	18	0	104
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	3.1	0.0	0.1	0.1	0.1	0.0	0.1
IEP (not gifted) (Number)	1	0	1,085	14,109	5,283	94	20,572
IEP (not gifted) (Percent)	3.1	0.0	14.3	13.9	29.4	49.2	16.1
Student exited IEP in last 2 years (Number)	0	1	96	1,338	230	1	1,666
Student exited IEP in last 2 years (Percent)	0.0	1.8	1.3	1.3	1.3	0.5	1.3
Title I (Number)	1	51	2,072	19,162	4,723	129	26,138
Title I (Percent)	3.1	91.1	27.3	18.8	26.3	67.5	20.5
Title III served (Number)	0	0	233	2,169	1,263	17	3,682
Title III served (Percent)	0.0	0.0	3.1	2.1	7.0	8.9	2.9
Title III not served (Number)	0	0	0	13	16	0	29
Title III not served (Percent)	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Migrant student (Number)	0	0	3	30	9	1	43
Migrant student (Percent)	0.0	0.0	0.0	0.0	0.1	0.5	0.0
ELL (enrolled after 5/27/16) (Number)	0	0	19	190	164	2	375
ELL (enrolled after 5/27/16) (Percent)	0.0	0.0	0.3	0.2	0.9	1.0	0.3
ELL (enrolled on or before 5/27/16) (Number)	0	0	225	2,110	1,182	19	3,536
ELL (enrolled on or before 5/27/16) (Percent)	0.0	0.0	3.0	2.1	6.6	9.9	2.8
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	13	294	116	0	423

Table 10–3L (continued). Demographic Characteristics of Students taking the Spring 2017 Keystone Exam: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Percent)	0.0	0.0	0.2	0.3	0.6	0.0	0.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	15	246	104	0	365
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0.0	0.0	0.2	0.2	0.6	0.0	0.3
Former ELL no longer monitored (Number)	0	0	189	2,514	354	0	3,057
Former ELL no longer monitored (Percent)	0.0	0.0	2.5	2.5	2.0	0.0	2.4
Foreign exchange student (Number)	0	0	0	21	9	0	30
Foreign exchange student (Percent)	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Economically disadvantaged (Number)	1	52	3,733	40,988	10,157	151	55,082
Economically disadvantaged (Percent)	3.1	92.9	49.1	40.3	56.6	79.1	43.2
Historically Underperforming Subgroup (Number)	2	52	4,161	47,612	12,499	171	64,497
Historically Underperforming Subgroup (Percent)	6.3	92.9	54.8	46.8	69.6	89.5	50.5
Enrollment in school of residence after 10/1/16 (Number)	1	1	375	3,219	1,347	47	4,990
Enrollment in school of residence after 10/1/16 (Percent)	3.1	1.8	4.9	3.2	7.5	24.6	3.9
Enrollment in district of residence after 10/1/16 (Number)	1	1	271	2,558	1,082	44	3,957
Enrollment in district of residence after 10/1/16 (Percent)	3.1	1.8	3.6	2.5	6.0	23.0	3.1
Enrollment as PA resident after 10/1/16 (Number)	1	0	146	1,359	567	12	2,085
Enrollment as PA resident after 10/1/16 (Percent)	3.1	0.0	1.9	1.3	3.2	6.3	1.6
Enrollment in school of residence after 10/1/15 but on/ before 10/1/16 (Number)	0	2	2,981	17,510	3,050	65	23,608
Enrollment in school of residence after 10/1/15 but on/ before 10/1/16 (Percent)	0.0	3.6	39.2	17.2	17.0	34.0	18.5
Enrollment in district of residence after 10/1/15 but on/ before 10/1/16 (Number)	0	2	986	6,098	1,860	61	9,007
Enrollment in district of residence after 10/1/15 but on/ before 10/1/16 (Percent)	0.0	3.6	13.0	6.0	10.4	31.9	7.1
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Court/agency placed (Number)	0	0	39	57	35	26	157
Court/agency placed (Percent)	0.0	0.0	0.5	0.1	0.2	13.6	0.1
Number of assessed students (Number)	32	56	7,596	101,779	17,953	191	127,607

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

PARTICIPATION BY ADMINISTRATION MODE

The number of students assessed by the two administration modes, paper/pencil test (PPT) or a computer-based test (CBT), was as follows (see Tables 10–4A through 10–4L):

- PPT: Algebra I (149,962), Biology (122,614), and Literature (114,714)
- CBT: Algebra I (13,870), Biology (17,000), and Literature (12,893)

TEST ACCOMMODATIONS PROVIDED

School personnel supplied information regarding accommodations that a student may have received while taking the Keystone Exams. Accommodations are classified in terms of presentation, response, setting, and timing to enable students to better manage disabilities that hinder their ability to learn and respond to assessments. An accommodations manual entitled, Accommodations Guidelines: Keystone Exams and PSSA (PDE, revised 12/12/2016), was updated for use with the 2017 PSSA and Keystone Exams. This manual may be found on the PDE website at www.education.pa.gov. You can find the document by typing the manual title in the search box. A glossary of accommodation terms as applied to the Keystone Exams is provided in Table 10–10 at the end of this chapter.

The frequency with which accommodations were utilized for PPT and CBT formats is summarized separately for each course exam in Tables 10–4A through 10–7L. Tabled values are based on all students whose scores contributed to state summary statistics in a given Keystone Exam. Because of the very small incidence of usage of many accommodations, combined with the fact that a number of accommodations are primarily accessed by only one of the two administration modes, meaningful comparisons between modes are rather limited. In the following tables, an NA denotes those instances in which a particular accommodation does not apply to one of the testing modes.

PRESENTATION ACCOMMODATIONS RECEIVED

Presentation accommodations are those that provide alternate ways for students to access and process printed instructional material and assessments. These include auditory, tactile, visual, and combined auditory/visual modes of presentation. The number of presentation accommodations provided in the 2017 Keystone Exams varied by content area and test administration mode as follows:

- PPT: Algebra I, 12; Biology, 12; and Literature, 7
- CBT: Algebra I, 13; Biology, 13; and Literature, 7

As depicted in Tables 10–4A through 10–4L, the actual frequencies were quite low. The most notable exceptions, applicable to Algebra I and Biology only, were “All items/questions read aloud” and “Some items/questions read aloud.” Among accommodations specific to CBT, the use of audio was the most frequent. Although included in the table data, the Spanish version (Algebra I and Biology; PPT only) is not included in the counts listed above.

RESPONSE ACCOMMODATIONS RECEIVED

Response accommodations permit students to complete assignments, tests, and activities in different ways and to solve or organize problems using some type of assistive device or organizer. The number of response accommodations provided on the Spring 2017 Keystone Exams varied by subject as follows:

- PPT: Algebra I, 12; Biology, 12; and Literature, 9
- CBT: Algebra I, 9; Biology, 9; and Literature, 6

The frequency with which these accommodations were utilized is summarized in Tables 10–5A through 10–5L. The actual frequencies are quite low, representing less than one-tenth of one percent of assessed students in nearly all instances, regardless of administration mode.

SETTING ACCOMMODATIONS RECEIVED

Setting accommodations permit a change in the location in which a student receives instruction or participates in an assessment. In the Spring 2017 Keystone Exam administration, there were four categories of setting accommodations, which applied to both administration modes and to each course exam. As depicted in Tables 10–6A through 10–6L, the most common accommodation was small group setting for both PPT and CBT modes of administration, although the percentage of usage was somewhat higher for PPT.

TIMING ACCOMMODATIONS RECEIVED

Timing accommodations involve a change in the allowable length of time to complete assignments or assessments, including the way in which time is organized. There were four categories of timing accommodations, which applied to both administration modes and to each course exam. As depicted in Tables 10–7A through 10–7L, the most common accommodation was extended time for both PPT and CBT administration modes with slightly higher percentages for PPT than CBT in Algebra I and Literature.

Table 10–4A. Incidence of Presentation Accommodations Received on the Spring 2017 Keystone Exam: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	3	N/A	3
Braille format (Percent)	0	N/A	0
Large print format (Number)	78	N/A	78
Large print format (Percent)	.1	N/A	0
Computer Assistive Technology (Number)	4	N/A	4
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	1,007	92	1,099
Some test items/questions read aloud (Percent)	.7	.7	.7
All test items/questions read aloud (Number)	606	121	727
All test items/questions read aloud (Percent)	.4	.9	.4
Test items/questions signed (Number)	22	0	22
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	66	3	69
Test items/questions interpreted for ELL student (Percent)	0	0	0
Amplification device (Number)	16	1	17
Amplification device (Percent)	0	0	0
Magnification device (Number)	4	2	6
Magnification device (Percent)	0	0	0
Color overlay (Number)	33	N/A	33
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	64	5	69
Other (per Accommodations Guidelines) (Percent)	0	0	0
Spanish version (Number)	1,514	N/A	1,514
Spanish version (Percent)	1	N/A	.9
Audio (Number)	N/A	715	715
Audio (Percent)	N/A	5.2	.4
Color Chooser (Number)	N/A	22	22
Color Chooser (Percent)	N/A	.2	0
Contrasting Text Chooser (Number)	N/A	21	21
Contrasting Text Chooser (Percent)	N/A	.2	0
Reverse Contrast – only Spring and Summer (Number)	N/A	21	21
Reverse Contrast – only Spring and Summer (Percent)	N/A	.2	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	2	2
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	149,962	13,870	163,832

Table 10–4B. Incidence of Presentation Accommodations Received on the Spring 2017 Keystone Exam: Biology

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	6	N/A	6
Braille format (Percent)	0	N/A	0
Large print format (Number)	72	N/A	72
Large print format (Percent)	.1	N/A	.1
Computer Assistive Technology (Number)	6	N/A	6
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	765	124	889
Some test items/questions read aloud (Percent)	.6	.7	.6
All test items/questions read aloud (Number)	870	227	1,097
All test items/questions read aloud (Percent)	.7	1.3	.8
Test items/questions signed (Number)	7	0	7
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	30	2	32
Test items/questions interpreted for ELL student (Percent)	0	0	0
Amplification device (Number)	8	1	9
Amplification device (Percent)	0	0	0
Magnification device (Number)	3	3	6
Magnification device (Percent)	0	0	0
Color overlay (Number)	42	N/A	42
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	39	9	48
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Spanish version (Number)	1,099	N/A	1,099
Spanish version (Percent)	.9	N/A	.8
Audio (Number)	N/A	1,052	1,052
Audio (Percent)	N/A	6.2	.8
Color Chooser (Number)	N/A	25	25
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	23	23
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast – only Spring and Summer (Number)	N/A	22	22
Reverse Contrast – only Spring and Summer (Percent)	N/A	.1	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	2	2
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	122,614	17,000	139,614

Table 10–4L. Incidence of Presentation Accommodations Received on the Spring 2017 Keystone Exam: Literature

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	7	N/A	7
Braille format (Percent)	0	N/A	0
Large print format (Number)	69	N/A	69
Large print format (Percent)	.1	N/A	.1
Computer Assistive Technology (Number)	6	N/A	6
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	14	1	15
Amplification device (Percent)	0	0	0
Magnification device (Number)	4	3	7
Magnification device (Percent)	0	0	0
Color overlay (Number)	38	N/A	38
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	42	5	47
Other (per Accommodations Guidelines) (Percent)	0	0	0
Color Chooser (Number)	N/A	24	24
Color Chooser (Percent)	N/A	.2	0
Contrasting Text Chooser (Number)	N/A	23	23
Contrasting Text Chooser (Percent)	N/A	.2	0
Reverse Contrast – only Spring and Summer (Number)	N/A	20	20
Reverse Contrast – only Spring and Summer (Percent)	N/A	.2	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Number of assessed students (Number)	114,714	12,893	127,607

Table 10–5A. Incidence of Response Accommodations Received on the Spring 2017 Keystone Exam: Algebra I

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	76	0	76
Test administrator marked multiple-choice responses at student's direction (Percent)	.1	0	0
Test administrator scribed open-ended responses at student's direction (Number)	84	1	85
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	.1
Test administrator transcribed student responses (Number)	142	0	142
Test administrator transcribed student responses (Percent)	.1	0	.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	18	0	18
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	32	0	32
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	24	N/A	24
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	4	N/A	4
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	139	11	150
Translation dictionary for ELL student (Percent)	.1	.1	.1
Other (per Accommodations Guidelines) (Number)	53	2	55
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	149,962	13,870	163,832

Table 10–5B. Incidence of Response Accommodations Received on the Spring 2017 Keystone Exam: Biology

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	52	0	52
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	63	0	63
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	0
Test administrator transcribed student responses (Number)	138	0	138
Test administrator transcribed student responses (Percent)	.1	0	.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	13	0	13
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	37	0	37
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	37	N/A	37
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	6	N/A	6
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	1	0	1
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	124	12	136
Translation dictionary for ELL student (Percent)	.1	.1	.1
Other (per Accommodations Guidelines) (Number)	41	1	42
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	122,614	17,000	139,614

Table 10–5L. Incidence of Response Accommodations Received on the Spring 2017 Keystone Exam: Literature

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	45	1	46
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	80	0	80
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	.1
Test administrator transcribed student responses (Number)	168	1	169
Test administrator transcribed student responses (Percent)	.1	0	.1
Keyboard, word processor, or computer (Number)	73	N/A	73
Keyboard, word processor, or computer (Percent)	.1	N/A	.1
Brailler/Notetaker (Number)	8	N/A	8
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	14	3	17
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	114,714	12,893	127,607

Table 10–6A. Incidence of Setting Accommodations Received on the Spring 2017 Keystone Exam: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	56	0	56
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	357	13	370
One-on-one setting (Percent)	.2	.1	.2
Small group setting (Number)	11,277	967	12,244
Small group setting (Percent)	7.5	7	7.5
Other (per Accommodations Guidelines) (Number)	66	3	69
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	149,962	13,870	163,832

Table 10–6B. Incidence of Setting Accommodations Received on the Spring 2017 Keystone Exam: Biology

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	54	0	54
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	309	9	318
One-on-one setting (Percent)	.3	.1	.2
Small group setting (Number)	9,853	1,291	11,144
Small group setting (Percent)	8	7.6	8
Other (per Accommodations Guidelines) (Number)	37	5	42
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	122,614	17,000	139,614

Table 10–6L. Incidence of Setting Accommodations Received on the Spring 2017 Keystone Exam: Literature

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	49	0	49
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	258	6	264
One-on-one setting (Percent)	.2	0	.2
Small group setting (Number)	8,969	1,049	10,018
Small group setting (Percent)	7.8	8.1	7.9
Other (per Accommodations Guidelines) (Number)	24	6	30
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	114,714	12,893	127,607

Table 10–7A. Incidence of Timing Accommodations Received on the Spring 2017 Keystone Exam: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	14,973	980	15,953
Extended time (Percent)	10	7.1	9.7
Frequent breaks (Number)	523	125	648
Frequent breaks (Percent)	.3	.9	.4
Changed test schedule (Number)	282	6	288
Changed test schedule (Percent)	.2	0	.2
Other (per Accommodations Guidelines) (Number)	22	0	22
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	149,962	13,870	163,832

Table 10–7B. Incidence of Timing Accommodations Received on the Spring 2017 Keystone Exam: Biology

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	3,973	679	4,652
Extended time (Percent)	3.2	4	3.3
Frequent breaks (Number)	427	180	607
Frequent breaks (Percent)	.3	1.1	.4
Changed test schedule (Number)	193	1	194
Changed test schedule (Percent)	.2	0	.1
Other (per Accommodations Guidelines) (Number)	3	1	4
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	122,614	17,000	139,614

Table 10–7L. Incidence of Timing Accommodations Received on the Spring 2017 Keystone Exam: Literature

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	11,764	698	12,462
Extended time (Percent)	10.3	5.4	9.8
Frequent breaks (Number)	409	150	559
Frequent breaks (Percent)	.4	1.2	.4
Changed test schedule (Number)	200	1	201
Changed test schedule (Percent)	.2	0	.2
Other (per Accommodations Guidelines) (Number)	6	1	7
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	114,714	12,893	127,607

ACCOMMODATION RATE FOR NON-IEP AND IEP STUDENTS

A comparison between students without an IEP (non-IEP Students) and those with an IEP (IEP Students) with regard to having received an accommodation is provided in Table 10–8. In this data, accommodated means that a student received one or more of the total number of accommodations available for a given course; however, this varies somewhat with administration mode. The category of non-accommodated indicates that a student did not receive any accommodations during testing.

The general pattern of findings provided in Table 10–8 reveals a consistent and substantially higher percentage of IEP Students receiving an accommodation, in contrast to non-IEP Students. This same pattern holds true regardless of test administration mode of Keystone Exam. The comparisons between administration modes revealed that the accommodation rates for IEP students taking a PPT are close to those taking a CBT.

Table 10–8A. Accommodation Rate for Non-IEP and IEP Students on the Spring 2017 Keystone Exams: Algebra I

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	127,412	11,460	138,872
Non-Accommodated (Number)	113,635	10,745	124,380
Non-Accommodated (Percent)	89.2	93.8	89.6
Accommodated (Number)	13,777	715	14,492
Accommodated (Percent)	10.8	6.2	10.4
IEP Students (Number)	22,550	2,410	24,960
Non-Accommodated (Number)	11,521	1,199	12,720
Non-Accommodated (Percent)	51.1	49.8	51
Accommodated (Number)	11,029	1,211	12,240
Accommodated (Percent)	48.9	50.2	49

Table 10–8B. Accommodation Rate for Non-IEP and IEP Students on the Spring 2017 Keystone Exams: Biology

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	104,274	13,818	118,092
Non-Accommodated (Number)	100,504	13,494	113,998
Non-Accommodated (Percent)	96.4	97.7	96.5
Accommodated (Number)	3,770	324	4,094
Accommodated (Percent)	3.6	2.3	3.5
IEP Students (Number)	18,340	3,182	21,522
Non-Accommodated (Number)	8,930	1,518	10,448
Non-Accommodated (Percent)	48.7	47.7	48.5
Accommodated (Number)	9,410	1,664	11,074
Accommodated (Percent)	51.3	52.3	51.5

Table 10–8L. Accommodation Rate for Non-IEP and IEP Students on the Spring 2017 Keystone Exams: Literature

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	96,635	10,400	107,035
Non-Accommodated (Number)	86,277	10,034	96,311
Non-Accommodated (Percent)	89.3	96.5	90
Accommodated (Number)	10,358	366	10,724
Accommodated (Percent)	10.7	3.5	10
IEP Students (Number)	18,079	2,493	20,572
Non-Accommodated (Number)	9,106	1,406	10,512
Non-Accommodated (Percent)	50.4	56.4	51.1
Accommodated (Number)	8,973	1,087	10,060
Accommodated (Percent)	49.6	43.6	48.9

THE INCIDENCE OF ACCOMMODATIONS AND IEP AND ELL STATUS

As noted in Table 10–8, students with an IEP received an accommodation of some type far more often than non-IEP students. Certain accommodations with very low frequencies are specific to particular disabilities while others are far more common and may also apply to students classified as English Language Learners (ELL). Because the accommodations with the largest frequencies can potentially supply the most stable data when separated out for subgroup analysis, those in most common use were selected for display in Tables 10–9A through 10–9L. The most frequently occurring accommodations for assessed students were:

- Some test items/questions read aloud (Algebra I and Biology only)
- All test items/questions read aloud (Algebra I and Biology only)
- Small group setting
- Extended time
- Frequent breaks

Coding for IEP is dichotomous, as students are classified IEP and non-IEP. For purposes of this analysis, an English Language Learner (ELL) is an assessed student classified ELL and enrolled in a U.S. school on or before May 27, 2016. All other assessed students, including those who have exited an ESL/bilingual program and are in the first or second year of monitoring, are regarded as non-ELL.

Customarily, a considerably larger percentage of IEP students receive a given accommodation than non-IEP students. Although less frequent, certain accommodations also have a high frequency rate for ELL students. To separate out the effect of being classified IEP or ELL, four possible combinations are presented in Tables 10–9A through 10–9L. These include general education students (who are neither IEP nor ELL), students who are IEP but non-ELL, students who are ELL but non-IEP, and students who are both IEP and ELL. The bottom row for each administration mode provides the total number of assessed students in each of the four classifications.

For purposes of descriptively comparing the four IEP/ELL subgroups with respect to whether a subgroup displayed a larger percentage rate than another subgroup, a choice was made to use a difference of five or more percentage points as a criterion for judging importance. In many instances, the percentage difference between subgroups was of little practical significance (from zero to only several percentage points).

Although the separate presentation of data for PPT and CBT modes provides an impression of overall findings, the much smaller *n*-counts and accommodation rate by students taking a CBT renders an administration mode comparison less meaningful. Nevertheless, it is possible to make some cautious observations when sufficient *n*-counts and consistency are present as noted in the summary of findings below.

<u>SUBGROUP COMPARISONS FOR PPT ADMINISTRATION MODE</u>

Subgroup comparisons were regarded as viable for the PPT administration. There was little differentiation across subgroups for the two accommodations involving items/questions read aloud (Algebra I and Biology) and for frequent breaks (Algebra I, Biology, and Literature). Small group setting was by far the most prevalent accommodation for the IEP/non-ELL subgroup followed by the IEP/ELL and ELL/non-IEP subgroups. This pattern was consistent across all three course exams. Another consistent pattern was observed for extended time, which was more prevalent for the IEP/non-ELL, ELL/non-IEP, and IEP/ELL subgroups than for the General Education subgroup.

<u>SUBGROUP COMPARISONS FOR CBT ADMINISTRATION MODE</u>

For the CBT administration the ELL/Non-IEP and IEP/ELL subgroup *n*-counts were only 183 and 33 for Algebra I, 252 and 66 for Biology, and 134 and 32 for Literature, respectively. Consequently, only the General Education and IEP/non-ELL subgroups had a sufficient sample size to support reasonable comparisons. A consistent pattern noted for all three course exams was the greater prevalence of small group setting, extended time, and frequent breaks by the IEP/non-ELL subgroup than for the General Education subgroup.

<u>COMPARISONS BETWEEN PPT AND CBT</u>

The only subgroups for which comparisons between PPT and CBT administration modes were deemed reasonable based on sample size were within the General Education and IEP/non-ELL subgroups. The findings are summarized below.

- The General Education subgroup displayed a very low incidence of accommodations, typically less than one percent, in nearly all instances for both PPT and CBT administrations. The accommodation students mostly received is extended time.
- For the IEP/non-ELL subgroup, small group setting was the only accommodation for which PPT administration consistently exceeded CBT by more than five percentage points in all three course exams. The instances in which students tested by CBT exceeded those responding by PPT were extended time and frequent breaks.

Table 10–9A. Incidence of IEP and ELL Students Receiving Selected Accommodations on the Spring 2017 Keystone Exam: Algebra I

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	45	876	66	20
PPT - Some test items/questions read aloud (Percent)	0	4	1.2	2.2
PPT - All test items/questions read aloud (Number)	9	564	11	22
PPT - All test items/questions read aloud (Percent)	0	2.6	.2	2.4
PPT - Small group setting (Number)	944	9,654	392	287
PPT - Small group setting (Percent)	.8	44.6	7.4	31.3
PPT - Extended time (Number)	11,933	2,561	411	68
PPT - Extended time (Percent)	9.8	11.8	7.7	7.4
PPT - Frequent breaks (Number)	42	474	1	6
PPT - Frequent breaks (Percent)	0	2.2	0	.7
PPT - Number assessed (Number)	122,101	21,633	5,311	917
CBT - Some test items/questions read aloud (Number)	7	84	1	0
CBT - Some test items/questions read aloud (Percent)	.1	3.5	.5	0
CBT - All test items/questions read aloud (Number)	6	113	1	1
CBT - All test items/questions read aloud (Percent)	.1	4.8	.5	3
CBT - Small group setting (Number)	40	901	18	8
CBT - Small group setting (Percent)	.4	37.9	9.8	24.2
CBT - Extended time (Number)	613	345	19	3
CBT - Extended time (Percent)	5.4	14.5	10.4	9.1
CBT - Frequent breaks (Number)	3	122	0	0
CBT - Frequent breaks (Percent)	0	5.1	0	0
CBT - Number assessed (Number)	11,277	2,377	183	33
Total - Some test items/questions read aloud (Number)	52	960	67	20
Total - Some test items/questions read aloud (Percent)	0	4	1.2	2.1
Total - All test items/questions read aloud (Number)	15	677	12	23
Total - All test items/questions read aloud (Percent)	0	2.8	.2	2.4
Total - Small group setting (Number)	984	10,555	410	295
Total - Small group setting (Percent)	.7	44	7.5	31.1
Total - Extended time (Number)	12,546	2,906	430	71
Total - Extended time (Percent)	9.4	12.1	7.8	7.5
Total - Frequent breaks (Number)	45	596	1	6
Total - Frequent breaks (Percent)	0	2.5	0	.6
Total - Number assessed (Number)	133,378	24,010	5,494	950

Table 10-9B. Incidence of IEP and ELL Students Receiving Selected Accommodations on the Spring 2017 Keystone Exam: Biology

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	35	683	32	15
PPT - Some test items/questions read aloud (Percent)	0	3.9	.8	2.1
PPT - All test items/questions read aloud (Number)	36	791	9	34
PPT - All test items/questions read aloud (Percent)	0	4.5	.2	4.8
PPT - Small group setting (Number)	736	8,589	298	230
PPT - Small group setting (Percent)	.7	48.7	7.9	32.5
PPT - Extended time (Number)	2,337	1,343	254	39
PPT - Extended time (Percent)	2.3	7.6	6.7	5.5
PPT - Frequent breaks (Number)	34	385	0	8
PPT - Frequent breaks (Percent)	0	2.2	0	1.1
PPT - Number assessed (Number)	100,478	17,633	3,796	707
CBT - Some test items/questions read aloud (Number)	5	114	1	4
CBT - Some test items/questions read aloud (Percent)	0	3.7	.4	6.1
CBT - All test items/questions read aloud (Number)	9	209	1	8
CBT - All test items/questions read aloud (Percent)	.1	6.7	.4	12.1
CBT - Small group setting (Number)	64	1,185	21	21
CBT - Small group setting (Percent)	.5	38	8.3	31.8
CBT - Extended time (Number)	223	429	17	10
CBT - Extended time (Percent)	1.6	13.8	6.7	15.2
CBT - Frequent breaks (Number)	6	168	0	6
CBT - Frequent breaks (Percent)	0	5.4	0	9.1
CBT - Number assessed (Number)	13,566	3,116	252	66
Total - Some test items/questions read aloud (Number)	40	797	33	19
Total - Some test items/questions read aloud (Percent)	0	3.8	.8	2.5
Total - All test items/questions read aloud (Number)	45	1,000	10	42
Total - All test items/questions read aloud (Percent)	0	4.8	.2	5.4
Total - Small group setting (Number)	800	9,774	319	251
Total - Small group setting (Percent)	.7	47.1	7.9	32.5
Total - Extended time (Number)	2,560	1,772	271	49
Total - Extended time (Percent)	2.2	8.5	6.7	6.3
Total - Frequent breaks (Number)	40	553	0	14
Total - Frequent breaks (Percent)	0	2.7	0	1.8
Total - Number assessed (Number)	114,044	20,749	4,048	773

Table 10–9L. Incidence of IEP and ELL Students Receiving Selected Accommodations on the Spring 2017 Keystone Exam: Literature

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Small group setting (Number)	693	7,850	215	211
PPT - Small group setting (Percent)	.7	45.1	7	32.2
PPT - Extended time (Number)	9,167	2,209	318	70
PPT - Extended time (Percent)	9.8	12.7	10.3	10.7
PPT - Frequent breaks (Number)	36	366	2	5
PPT - Frequent breaks (Percent)	0	2.1	.1	.8
PPT - Number assessed (Number)	93,546	17,423	3,089	656
CBT - Small group setting (Number)	38	983	18	10
CBT - Small group setting (Percent)	.4	39.9	13.4	31.3
CBT - Extended time (Number)	308	368	18	4
CBT - Extended time (Percent)	3	15	13.4	12.5
CBT - Frequent breaks (Number)	6	142	0	2
CBT - Frequent breaks (Percent)	.1	5.8	0	6.3
CBT - Number assessed (Number)	10,266	2,461	134	32
Total - Small group setting (Number)	731	8,833	233	221
Total - Small group setting (Percent)	.7	44.4	7.2	32.1
Total - Extended time (Number)	9,475	2,577	336	74
Total - Extended time (Percent)	9.1	13	10.4	10.8
Total - Frequent breaks (Number)	42	508	2	7
Total - Frequent breaks (Percent)	0	2.6	.1	1
Total - Number assessed (Number)	103,812	19,884	3,223	688

GLOSSARY OF ACCOMMODATION TERMS

Table 10–10 provides a brief description of accommodation terms as used in the PSSA and Keystone Exams. Accommodation data was supplied by school personnel as noted in the left column of the table. The right column contains an explanation derived from the PDE publication, *2017 Accommodations Guidelines* (PDE, revised 12/12/2016). This manual may be found on the PDE website at www.education.state.pa.us. You can find the document by typing the manual title in the search box.

Table 10–10. Glossary of Accommodation Terms as Applied in the 2017 PSSA and 2016–2017 Keystone Exams

Type of Testing Accommodation	Explanation
Student used the following Online Presentation Accommodations	
Braille format	Students may use a Braille format of the test. Answers must then be transcribed into the answer booklet without alteration.
Large print format	Students with visual impairments may use a large print format. Answers must then be transcribed into the answer booklet without alteration.
Magnification device	Devices to magnify print may be used for students with visual impairments and/or print disabilities.
Color overlay	Students with visual impairments may place a color overlay on a printed page of the test document to make text more readable.
Computer assistive technology (e.g., electronic screen reader) (PDE approval required)	Students with severe visual disabilities that prevent them from accessing instructional material or performing the skill may use computer assistive technology; however, PDE must approve the program and functions prior to the test window.
Test items/questions/ prompt/text-dependent analysis signed	Deaf/hearing impaired students may receive test directions from a qualified interpreter. Signing is also permitted for PSSA ELA writing section multiple choice items, essay prompts, and text-dependent analysis questions and all items in PSSA mathematics and science and for Keystone Algebra and Biology.
Test items/questions/ prompt/text-dependent analysis interpreted for ELL	A qualified interpreter may translate directions or clarify instructions for the assessments. The interpreter may translate but not define specific words or test questions on the PSSA mathematics, science, ELA writing section multiple choice items, essay prompts, and text-dependent analysis questions and Keystone Algebra and Biology exams.
Some or all test items/ questions/prompt/text-dependent analysis read aloud	Students unable to decode text visually may have items/questions read aloud for PSSA ELA writing section multiple choice items, essay prompts, and text-dependent analysis questions and all items in PSSA mathematics and science and for Keystone Algebra and Biology; however, words may not be defined.
Amplification device	In addition to using hearing aids, an amplification device to enhance clarity may be required.
Other (PDE approval required)	Other presentation accommodations indicated in the *Accommodation Guidelines* may be provided; however, PDE approval is required prior to the test window.
Spanish version for PSSA (Math and Science) and Keystone (Algebra and Biology)	Students whose first language is Spanish and who have been enrolled in U.S. schools for fewer than three years may take this version.
Student used the following Online Presentation Accommodations	
Audio	The online test form reads permissible test directions and items for a student unable to decode text. The accommodation must be marked within the test engine system. The accommodation is available on PSSA mathematics, science, ELA writing section multiple choice items, essay prompts, and text-dependent analysis questions and Keystone Algebra and Biology exams.
Video sign language (per accommodations guidelines)	Eligible students who use a sign language accommodation during instructional periods may use a VSL on the PSSA mathematics and science assessments, or Keystone Algebra I and Biology.

Type of Testing Accommodation	Explanation
Color chooser or contrasting text chooser	The use of this accommodation enables a visually impaired student to change the background color or text color to make text more readable.
Refreshable Braille	This accommodation allows students to use a screen reader to produce a Braille translation output.
Student used the following Response Accommodations	
Brailler/Note taker (per *Accommodations Guidelines*)	Students using this device as part of their regular instructional program may use it on the assessments; however, without thesaurus, spelling, or grammar checker.
Test administrator scribed **open-ended** responses at student's direction	A test administrator may record word-for-word exactly what a student dictated directly into the test booklet. This includes MC and OE responses Keystone Algebra, Biology, and Literature tests and PSSA mathematics and science.
Test administrator marked **multiple-choice** responses at student's direction	A test administrator may mark an answer booklet at the direction of a student (e.g., a student may point to an MC answer with the test administrator marking the response in the answer booklet).
Test administrator transcribed student responses (per *Accommodations Guidelines*)	A test administrator may transcribe (copy) a student's written, typed, or keyed response into a standard answer booklet.
Qualified Interpreter translated, transcribed, and/or scribed student's **signed** responses	A qualified interpreter may interpret a student's **signed** responses into written English for Keystone Algebra and Biology exams, and PSSA mathematics and science assessments. Interpreters are not permitted to make corrections or change the meaning of the response.
Qualified Interpreter translated, transcribed, and/or scribed ELL student responses	A qualified interpreter may interpret a student's non-English oral responses into written English for Keystone Algebra and Biology exams, and PSSA mathematics and science assessments. Interpreters are not permitted to make corrections or change the meaning of the response.
Augmentative communication device	Students with severe communication difficulties may use a special device to convey responses, which must be transcribed into the answer booklet by the test administrator.
Keyboarding, word processor, or computer (per *Accommodations Guidelines*)	This is an allowable accommodation as a typing function only for students with the identified need. Online test should be considered for students who prefer/need to type open-ended responses. Supports such as dictionaries, thesauri, spell checkers, and grammar checkers must be turned off. Answers must then be transcribed into the answer booklet without alteration.
Translation dictionary for ELL student	A word-to-word dictionary that translates native language to English (or vice versa) without word definitions or pictures is allowed on any portion of the Keystone Algebra and Biology exams, and PSSA mathematics and science tests.
Computer assistive technology (e.g., electronic screen reader) (PDE approval required)	Students with blindness or extremely low vision may use dictate text into a computer. Responses must be transcribed verbatim into student's regular answer booklet.
Other (per *Accommodations Guidelines* or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.
Student used the following Setting Accommodations	
Hospital/home testing	A student who is confined to a hospital or to home during the testing window may be tested in that environment.
One-on-one setting	One-on-one settings are necessitated in certain instances, such as to reduce distraction or in the use of certain devices. A separate room may be used to reduce distraction.

Type of Testing Accommodation	Explanation
Small group setting	Some students may require a test setting with fewer students or a setting apart from all other students to minimize distraction.
Other (per *Accommodations Guidelines* or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.
Student used the following	**Timing Accommodations**
Extended time	Extended time may be allotted for each section of the test as a planned accommodation to enable students to finish.
Frequent breaks	Frequent breaks (breaks within a test section) may be scheduled for the completion of each test section; however, a test section must be completed within one school day.
Changed test schedule	Students whose disabilities prevent them from following a regular, planned test schedule may follow an individual schedule that enables test completion.
Other (per *Accommodations Guidelines* or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.

CHAPTER ELEVEN: CLASSICAL ITEM STATISTICS

This chapter provides an overview of the two most familiar item-level statistics obtained from classical (traditional) item analysis: item difficulty and item discrimination. The following results pertain not only to the operational Keystone Exams items but also to the embedded field test items. Other statistics such as Rasch item statistics and test-level statistics are discussed in Chapter Twelve and Chapter Seventeen, respectively.

ITEM-LEVEL STATISTICS

Appendix J provides classical item statistics for all items (i.e., operational and embedded field test items) in the Algebra I, Biology, and Literature Exams. Results are organized by administration and then by content area. These statistics represent the item characteristics most often used to determine whether an item functioned properly and/or how a group of students performed on a particular item. The item statistics in Appendix J include *N*, the number of students taking the test form for which there are valid test scores; *p*-values (denoted as PVal) for multiple-choice (MC) items and item means (denoted as Mean) for constructed-response (CR) items (indicators of item difficulty); proportions of students who chose each response option for MC items (denoted as P(A), P(B), P(C), P(D)) or gained each score point for CR items (denoted as P(0), P(1), P(2), P(3), and/or P(4)); proportions of students who did not respond to an item (denoted as P(-) for MC items and P(B) for CR items); item-total correlations (denoted as Total, indicators of item discrimination); item-total correlations for each response option for MC items (denoted as PT(A), PT(B), PT(C), and PT(D)); and gained score point for CR items (denoted as PT(1), PT(2), PT(3), and PT(4)).

Appendix J also provides the Rasch measurement-based statistics in columns Rasch, Infit, and Outfit. Detailed explanations of these statistics can be found in Chapter Twelve. The differential item functioning (DIF) analysis on the embedded field test items is provided in the last three columns. The detailed explanation of DIF codes can be found in Chapter Five.

ITEM DIFFICULTY

At the most general level, an item's difficulty is indicated by its mean score in some specified group (e.g., grade level).

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

In the mean score formula above, the individual item scores (x_i) are summed and then divided by the total number of students (n). For MC items, student scores are represented by 0s and 1s (0 = wrong, 1 = right). With 0/1 scoring, the equation above also represents the number of students correctly answering the item divided by the total number of students. So, this is also the *proportion correct* for the item, or as it is better known, the *p*value. In theory, *p*-values can range from 0.00[1] to 1.00 on the proportion-correct scale. For example, if an item has a *p*-value of 0.89, it means 89 percent of the students answered the item correctly. Additionally, this value might also suggest that the item is relatively easy and/or the students who attempted the item are relatively high achievers. In other words, item difficulty and student ability are somewhat confounded.

For CR items, mean scores can range from the minimum possible score (usually zero) to the maximum possible score (e.g., four points in the case of Algebra I CR items and three points in the case of Biology and Literature CR items). Sometimes a *pseudo p*-value is provided for a CR item by dividing the mean item score by the maximum possible item score.

The minimum and maximum extremes of the difficulty scale are virtually never seen in applied practice. However, understanding what those values are helps illustrate that relatively lower values correspond to more difficult items and that relatively higher values correspond to easier items. (Because of this, some assert that this index would be better referred to as the item's *easiness*.)

Item difficulty is an important consideration for the Keystone Exams because of the ranging achievement levels of students in Pennsylvania (Below Basic, Basic, Proficient, and Advanced). Items that are either very hard or very

[1] For MC items with four response options, pure random guessing would lead to an expected *p*-value of 0.25.

easy provide little information about student differences in achievement. However, an item answered correctly by a high percentage of students would suggest that the knowledge or skill the item taps has been mastered by most students. Conversely, an item answered correctly by a low percentage of students would suggest that few students have mastered the knowledge or skill the item taps. So, on a criteria-referenced test like the Keystone Exams, a test development goal is to include a wide range of item difficulties.

Utilizing the proportion of students who chose each option can be helpful for verifying keys. For example, if a large proportion of students chose a distractor instead of the key answer, it may, but not always, indicate that the key is not correct. The proportion of students omitting or not reaching an item is useful for identifying issues related to testing time and item/test layout. Keystone Exams are not speed tests. Therefore, students should have enough time to take the exams. An omit proportion greater than 0.05 for a single item could be an indication that students were not given enough time to take the test or an indication of an item/test layout problem. For example, some students might accidentally skip an item that follows a lengthy stem.

ITEM DISCRIMINATION

At the most general level, item discrimination[2] indicates an item's ability to differentiate between high and low achievers. It is expected that students with high ability (i.e., those who perform well on the Keystone Exams overall) would be more likely to answer any given Keystone Exams item correctly, while students with low ability (i.e., those who perform poorly on the Keystone Exams overall) would be more likely to answer the same item incorrectly. For the Keystone Exams, Pearson's product-moment correlation coefficient between item scores and test scores is used to indicate discrimination. As commonly practiced, Data Recognition Cooperation (DRC) removes the item score from the total score so that the resulting correlations will not be spuriously high. The correlation coefficient can range from -1.0 to +1.0. If the aforementioned expectation is met (high-scoring students tend to get the item right while low-scoring students do not), the correlation between the item score and the total test score will be both positive and noticeably large in its magnitude (i.e., well above zero) indicating that the item is a good discriminator between high- and low-ability students.

Item-total correlation for each option is another indicator of an item's ability to differentiate between high and low achievers. It is expected that students with high ability would be less likely to choose any distractors, while students with low ability would be more likely to choose a distractor. In other words, the item-total correlations for the distractors are expected to be negative.

In summary, the correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly.[3] In other words, students who did well on the total test tended to do well on the item as well. However, an interaction can exist between item discrimination and item difficulty. Items answered correctly (or incorrectly) by a large proportion of examinees (i.e., items with extreme *p*-values) can have reduced power to discriminate and thus can have lower correlations.

Discrimination is an important consideration for the operational Keystone Exams because the use of more discriminating items on a test is associated with more reliable test scores. This in turn means that score estimates will be more precise (i.e., there will be smaller confidence intervals around the scores) and, perhaps more importantly, that more accurate performance level placements will be made. The issues of reliability, confidence intervals, and performance level classifications are further discussed in Chapter Eighteen.

---

[2] As noted earlier, the discrimination index for dichotomous MC items is typically referred to as the *point-biserial correlation coefficient*. For CR items, the term *item-test correlation* is sometimes used.

[3] It is legitimate to view the point-biserial correlation as a standardized mean. A positive value indicates that students who chose that response had a higher mean score than the average score; a negative value indicates that students who chose that response had a lower mean score than the average score.

SCATTER PLOTS OF ITEM DISCRIMINATION AND DIFFICULTY

Figure 11–1 contains a series of scatter plots showing item discrimination (i.e., item-total correlation on *y*-axis) vs. item difficulty (i.e., *p*-value on *x*-axis) for the operational items in each content area exam by test administration. These plots provide information about item discrimination and difficulty in a single visual image for each Keystone Exam. This is because the *x*- and *y*-axes visually represent many important distributional indices:

- The minimum and maximum values are listed.
- Mean and median scores are indicated by the red dash lines.
- The first and third quartile (Q1 and Q3) are indicated by the red lines.
- Marginal histogram indicates the density of the individual data points.

It should be noted that pseudo *p*-values are used for CR items in these plots. Of course, the bivariate relationship between discrimination and difficulty is also presented. One does not usually expect any type of trend here. However, as noted earlier, it is often the case that items with extreme difficulties can have lower discrimination values, so this can be revealed in such a plot.

Figure 11–1. Scatter Plots of Item Discrimination and Difficulty

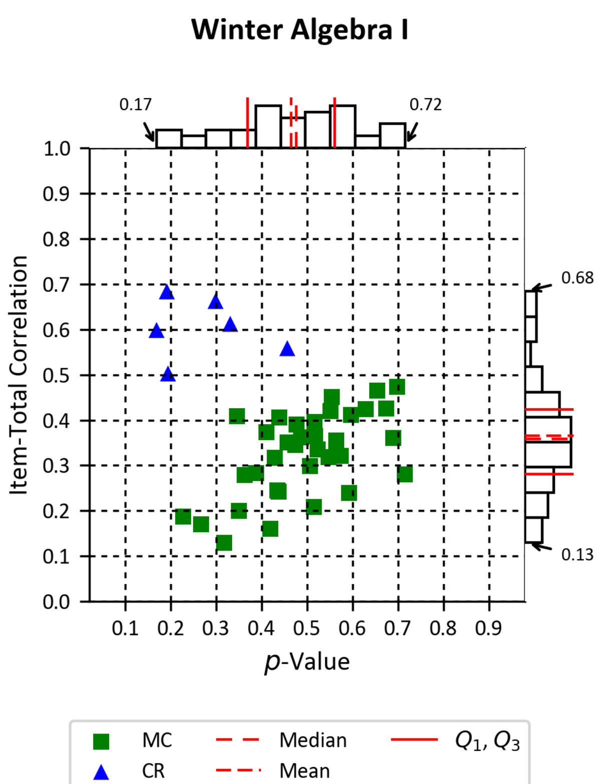
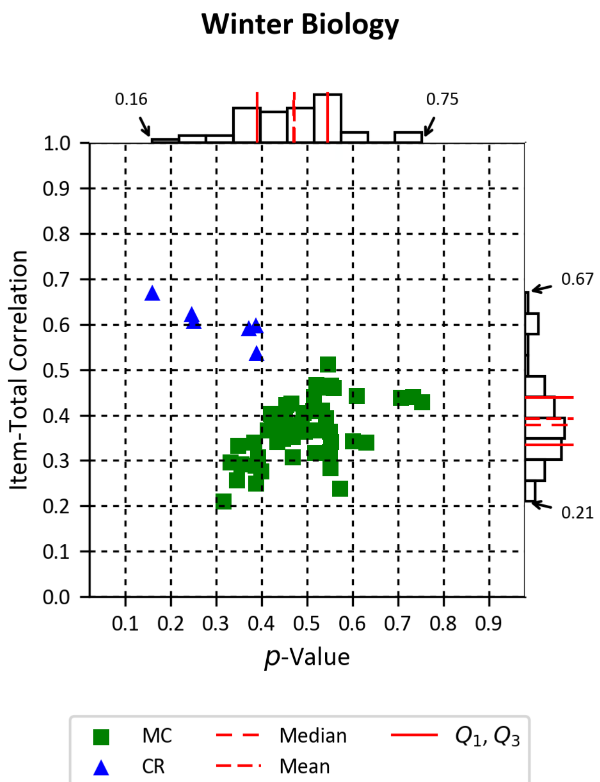


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



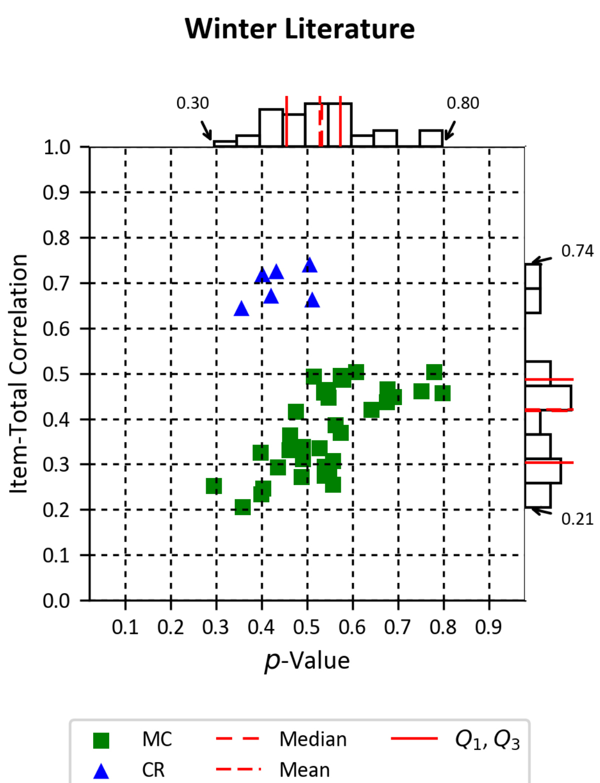
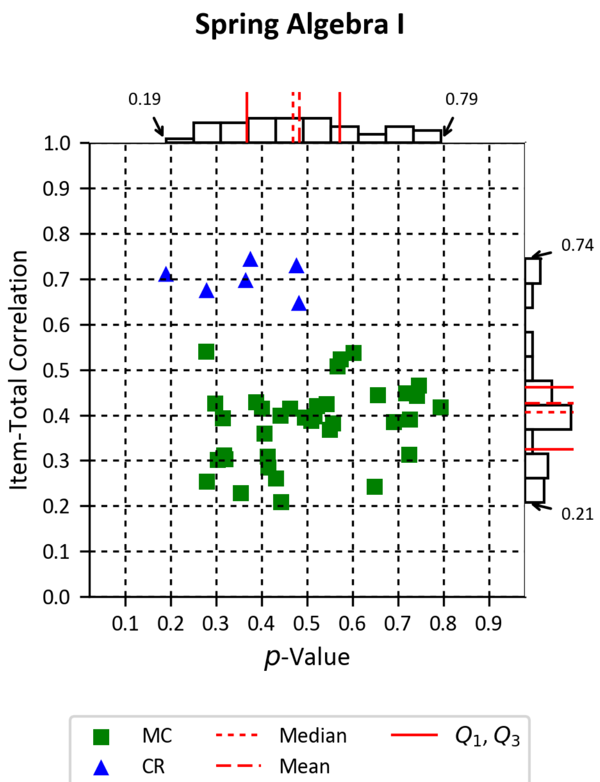


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



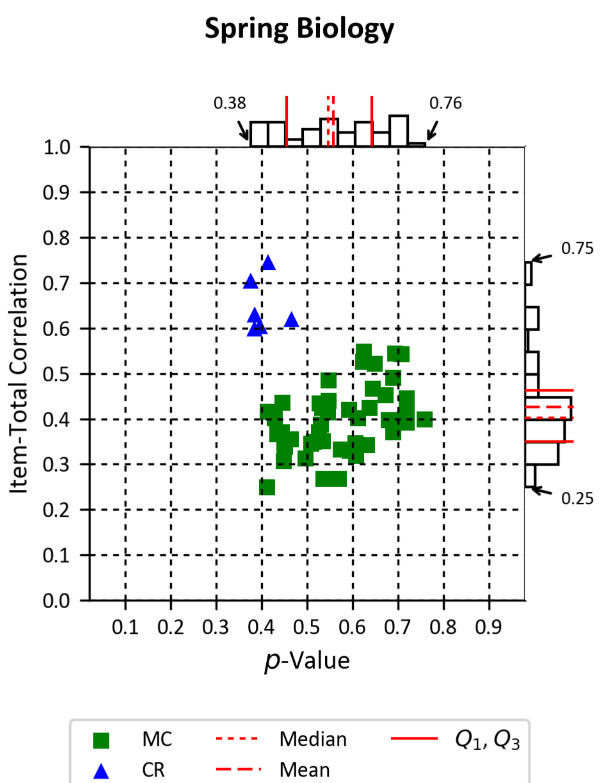
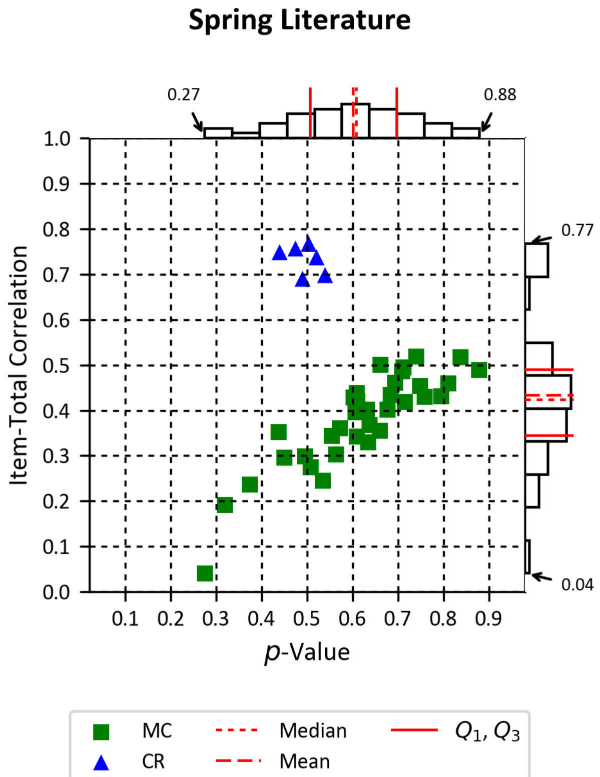


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



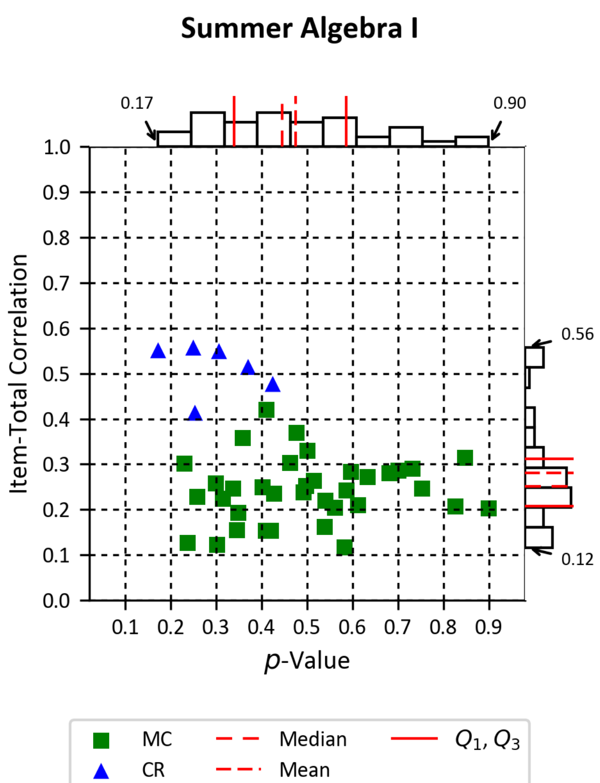
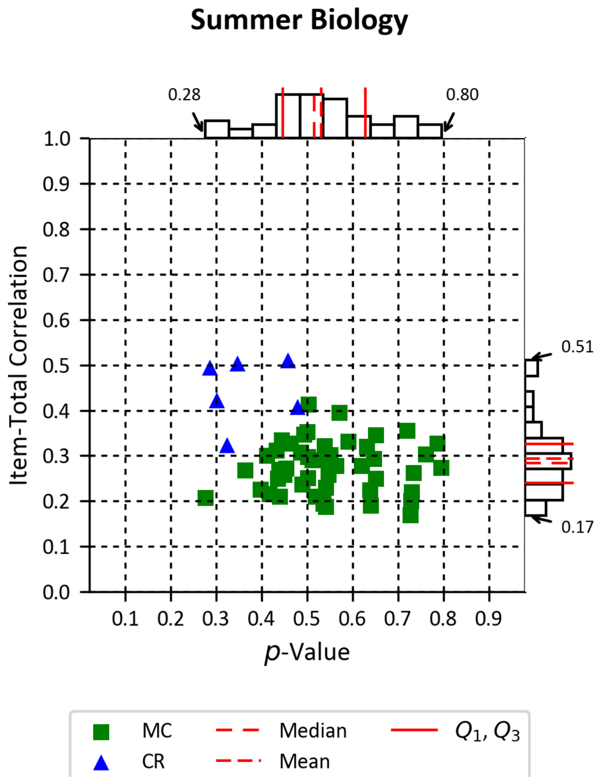
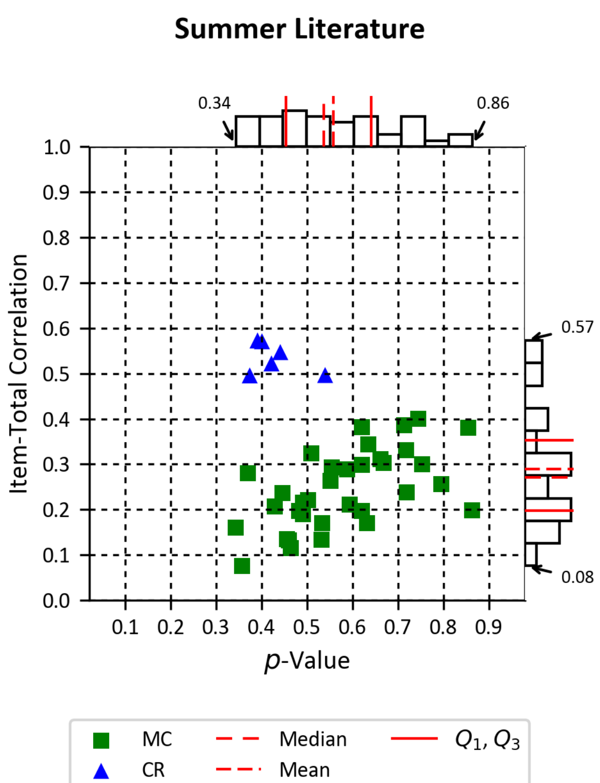


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty





OBSERVATIONS AND INTERPRETATIONS

Table 11–1 provides the mean and median *p*-values and median[4] item-total correlations for the operational MC and CR items in each content area. The mean *p*-value for the operational MC items ranged from about 0.49 to 0.62 with the standard deviation (SD) ranging from 0.10 to 0.18, while the mean *p*-values for the CR items ranged from about 0.27 to 0.49 with the standard deviation ranging from 0.03 to 0.11. The median item-test correlations ranged from 0.24 to 0.40 and 0.46 to 0.74 for the MC and CR items, respectively. The CR correlations tended to be higher than the MC correlations, which is not surprising because the CR items include more score points.

It is impossible to make global conclusions about the overall test quality from these item statistics alone. With that caveat in mind, the results presented in this chapter indicate that the item difficulties and discriminations were in expected and acceptable ranges.

Table 11–1. Mean and Median Statistics for Operational MC and CR Items

Administration	Content Area	MC Items Mean p-Value	MC Items SD p-Value	MC Items Median p-Value	MC Items Median I-T Corr.	CR Items Mean p-Value	CR Items SD p-Value	CR Items Median p-Value	CR Items Median I-T Corr.
Winter	Algebra I	0.50	0.12	0.51	0.34	0.27	0.11	0.25	0.61
Winter	Biology	0.49	0.10	0.49	0.37	0.30	0.10	0.31	0.60
Winter	Literature	0.54	0.11	0.54	0.37	0.44	0.06	0.43	0.70
Spring	Algebra I	0.50	0.15	0.50	0.39	0.36	0.11	0.37	0.70
Spring	Biology	0.58	0.10	0.57	0.40	0.40	0.03	0.39	0.62
Spring	Literature	0.62	0.14	0.63	0.40	0.49	0.04	0.50	0.74
Summer	Algebra I	0.50	0.18	0.49	0.24	0.30	0.09	0.28	0.53
Summer	Biology	0.55	0.12	0.54	0.27	0.37	0.08	0.34	0.46
Summer	Literature	0.58	0.14	0.57	0.24	0.43	0.06	0.41	0.53

Note: I-T Corr. is the item-total test score correlation; SD represents the standard deviation.

[4] Given that the value of the item-total correlation coefficient is not a linear function of the magnitude of the relation between the item and total test scores, the median instead of the mean of the item-total correlation was calculated for this statistic.

CHAPTER TWELVE: RASCH ITEM CALIBRATION

The particular item response theory (IRT) model used for the Keystone Exams is based on the work of Georg Rasch. Rasch models have had a long-standing presence in applied testing programs and have been the methodology continually used to calibrate the Pennsylvania System of School Assessment (PSSA) items in recent history. Consequently, this model was chosen for the Keystone Exams. IRT has several advantages over classical test theory, so it has become the standard procedure for analyzing item response data in large-scale assessments. However, IRT models make a number of strong assumptions related to dimensionality, local independence, and model-data fit. Resulting inferences derived from any application of IRT rest strongly on the degree to which the underlying assumptions are met.

This chapter outlines the procedures used for calibrating the operational Keystone Exams items. Generally, item calibration is the process of assigning a difficulty-parameter estimate to each item on an assessment so that they are placed on a common scale. This chapter briefly introduces the Rasch model, reports the results from evaluations of the adequacy of the Rasch assumptions, and summarizes the Rasch item statistics for the Keystone Exams in Algebra I, Biology, and Literature.

DESCRIPTION OF THE RASCH MODEL

The Rasch partial credit model (RPCM) (Wright & Masters, 1982) was used to calibrate Keystone Exams items because both multiple-choice (MC) and constructed-response (CR) items were part of the assessment. The RPCM extends the Rasch model (Rasch, 1960) for dichotomous (0, 1) items so that it accommodates the polytomous CR items. Under the RPCM, for a given item *i* with m_i score categories, the probability of person *n* scoring *x* ($x = 0, 1, 2, \ldots m_i$) is given by:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^{x} (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{x} (\beta_n - \delta_{ij})}, \quad x = 0, 1, \ldots, m_i$$

where represents a student's proficiency (ability) level, and is the step difficulty of the *j*th step on item *i*. For dichotomous MC items, the RPCM reduces to the standard Rasch model and the single step difficulty is referred to as the item's difficulty. The Rasch model predicts the probability of person *n* getting item *i* correct as follows:

$$\Phi_{ni}\left(X = 1 \mid \beta_n\right) = \frac{\exp\left(\beta_n - \delta_{ij}\right)}{1 + \exp\left(\beta_n - \delta_{ij}\right)}.$$

The Rasch model places both student ability and item difficulty (estimated in terms of log-odds or logits) on the same continuum. When the model assumptions are met, it also provides person ability estimates that are independent of the items employed in the assessment, and, conversely, estimates item difficulty independently of the sample of examinees. (As noted in Chapter Eleven, interpretation of item *p*-values confounds item difficulty and student ability.)

SOFTWARE AND ESTIMATION ALGORITHM

Item calibration was implemented via WINSTEPS computer program (Linacre & Wright, 2013), which employs unconditional (UCON), joint-maximum-likelihood estimation (JMLE).

SAMPLE CHARACTERISTICS

The characteristics of calibration samples are reported in Chapter Nine. These samples only include the students who attempted the tests. All omits (no response) and multiple responses (more than one response selected) were scored as incorrect answers (coded as 0s) for calibration.

CHECKING RASCH ASSUMPTIONS

Because the Rasch model was the basis of all calibration and equating analyses associated with the Keystone Exams, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met and how well the model fits the test data. Therefore, it is important to check these assumptions. This section evaluates the dimensionality of the data, local item independence, and model-data fit at the item level. Though a variety of methods are available for assessing these issues, the Rasch analyses and criteria available from WINSTEPS were used here. It should be noted that only operational items were analyzed since they are the basis of student scores.

Given that Keystone Exams use a pre-equating design (see details in Chapter Fifteen), calibrations with and without anchoring all the item parameter estimates were conducted to check the item difficulty stability. After reviewing the analyses results for the winter, spring, and summer administrations, a decision was made to use the item difficulty estimated from the field test data to generate the raw-to-scaled-score conversion tables. In this chapter, the adequacy of the Rasch calibration assumptions was checked with all the item difficulties anchored to the pre-equated values.

UNIDIMENSIONALITY

Rasch models assume that one dominant dimension determines the difference in students' performances. WINSTEPS provides results from a principal components analysis (PCA) that can be used to assess the unidimensionality assumption. Different from standard applications of PCA, WINSTEPS conducts its PCA on the response residuals, not the original observations. That is, the primary dimension from the Rasch model is removed first and then the residual variance is analyzed. The purpose of the analysis is to verify whether any other dominant components exist among the residuals (i.e., they account for a practically significant amount of residual variance). If any other dimensions are found, the unidimensionality assumption would be violated.

For Keystone Exams, the standardized residuals were used to conduct the PCA because simulation studies indicate that it gives the most accurate reflection of secondary dimensions in the items (Linacre, 2013). Table 12–1 presents the PCA results by administration for each content area. The results include the eigenvalues and variance explained by each component. As can been seen from the table, the eigenvalues for the first component are much larger than those for the rest of the components. The first component explained about 23.1 to 45.8 percent of the total variances. The rest of the components explained only a small percentage of variance. These results suggest that each of the Keystone Exams essentially measure a single dominant dimension.

Table 12–1. Results from PCA of Residuals in WINSTEPS

Administration/Content Area	Component	Eigenvalue	Variance Explained
Winter Algebra I	1	28.8	40.7%
Winter Algebra I	2	1.7	2.4%
Winter Algebra I	3	1.5	2.1%
Winter Algebra I	4	1.2	1.8%
Winter Algebra I	5	1.2	1.7%
Winter Biology	1	24.9	31.6%
Winter Biology	2	1.8	2.3%
Winter Biology	3	1.3	1.7%
Winter Biology	4	1.3	1.6%
Winter Biology	5	1.2	1.5%
Winter Literature	1	20.7	34.1%
Winter Literature	2	2.4	3.9%
Winter Literature	3	1.6	2.6%
Winter Literature	4	1.3	2.2%
Winter Literature	5	1.2	2.1%
Spring Algebra I	1	35.4	45.8%
Spring Algebra I	2	2.1	2.7%
Spring Algebra I	3	1.4	1.8%
Spring Algebra I	4	1.3	1.6%
Spring Algebra I	5	1.2	1.5%
Spring Biology	1	28.7	34.7%
Spring Biology	2	1.7	2.1%
Spring Biology	3	1.4	1.7%
Spring Biology	4	1.3	1.5%
Spring Biology	5	1.2	1.5%
Spring Literature	1	24.4	37.8%
Spring Literature	2	2.6	4.0%
Spring Literature	3	1.4	2.2%
Spring Literature	4	1.2	1.9%
Spring Literature	5	24.4	37.8%
Summer Algebra I	1	22.9	35.2%
Summer Algebra I	2	1.8	2.7%
Summer Algebra I	3	1.6	2.4%
Summer Algebra I	4	1.4	2.2%
Summer Algebra I	5	1.3	2.0%
Summer Biology	1	16.2	23.1%
Summer Biology	2	1.7	2.4%
Summer Biology	3	1.6	2.3%

Table 12–1 (continued). Results from PCA of Residuals in WINSTEPS

Administration/Content Area	Component	Eigenvalue	Variance Explained
Summer Biology	4	1.6	2.2%
Summer Biology	5	1.5	2.2%
Summer Literature	1	13.6	25.4%
Summer Literature	2	2.7	5.0%
Summer Literature	3	1.7	3.1%
Summer Literature	4	1.6	2.9%
Summer Literature	5	1.5	2.8%

LOCAL INDEPENDENCE

Local independence (LI) is a fundamental assumption of IRT. No relationship should exist between examinees' responses to different items after accounting for the abilities measured by a test. In formal statistical terms, a test *X* that is composed of items X_1, X_2,…X_I is locally independent with rest to the latent variable δ_n if, for all $x = (x_1, x_2,…x_I)$ and δ_n,

$$P_n\left(\mathbf{X}=\mathbf{x} \mid \delta_n\right)=\prod_{i=1}^{I} P\left(X_i=x_i \mid \delta_n\right).$$

This formula essentially states that the probability of any pattern of responses across all items (*x*), after conditioning on the abilities (δ_n) measured by the test, should be equal to the product of the conditional probabilities across each item (cf. the multiplication rule for independent events where the joint probabilities are equal to the product of the associated marginal probabilities).

The equation above shows the condition after satisfying the *strong form* of local independence. A *weak form* of local independence (WLI) was proposed by McDonald (1979). The distinction is important, since many indicators of local dependency are actually framed by WLI. The requirement here would be for the conditional covariance of all pairs of item responses, conditioned on the abilities, to be equal to zero. When this assumption is met, the joint probability of responses to an item pair, conditioned on abilities, is the product of the probabilities of responses to these two items, as shown below. (This is a *weaker* form because higher-order dependencies among items are allowed.) Based on the WLI, the following expression can be derived:

$$P\left(X_i=x_i, X_j=x_j \mid \delta_n\right)=P\left(X_i=x_i \mid \delta_n\right) P\left(X_j=x_j \mid \delta_n\right).$$

Marais and Andrich (2008) pointed out that local item dependence in the Rasch model can occur in two ways that some may not distinguish. The first way occurs when the assumption of unidimensionality is violated. Here, other nuisance dimensions besides a dominant dimension also determine students' performance (this can be called *trait dependence*). The second violation occurs when responses to an item depend on responses to another. This is a violation of statistical independence and can be called *response dependence*. Many people treat the assumptions of *unidimensionality* and *local independence* as one phenomenon and believe that once unidimensionality holds, that local independence also holds. By distinguishing the two sources of local dependence, one can see that while local independence can be related to unidimensionality, the two are different assumptions, and therefore, require different tests.

Residual item correlations provided in WINSTEPS for each item pair were used to assess the local dependence among the Keystone Exams items. In general, these residuals are computed as follows. First, expected item performance based on the Rasch model is determined using ability and item parameter estimates. Next, deviations (residuals) between the examinees' expected and observed performance is determined for each item. Finally, for each item pair, a correlation between the respective deviations is computed.

Two types of residual correlations are available in WINSTEPS: raw and standardized residuals. It should be noted that the raw score residual correlation essentially corresponds to Yen's Q_3 index (Yen, 1993), a popular LI statistic. The expected value for the Q_3 statistic is approximately -1/(k-1) when no local dependence exists, where k is test length. Thus, the expected Q_3 values should be approximately -0.026 or larger for the Keystone Exams (since Literature is the shortest test with 40 items). Index values that are greater than 0.20 indicate a degree of local dependence that probably should be examined by test developers (Chen & Thissen, 1997). Since the two residual correlations are very similar, the default *standardized residual correlation* in WINSTEPS was used for these analyses. Table 12–2 shows the summary statistics—mean, SD, minimum (Min), maximum (Max), and several percentiles (P_{10}, P_{25}, P_{50}, P_{75}, P_{90})—for all the residual correlations for each test. The total number of item pairs (N) and the number of pairs with residual correlations greater than 0.20 are also reported in this table. The mean residual correlations were slightly negative and the values were -0.02 after rounding. The vast majority of the correlations were very small, suggesting local item independence generally holds for the Keystone Exams in Algebra I, Biology, and Literature.

Table 12–2. Summary of Item Residual Correlations

Administration	Content Area	Stats N	Stats Mean	Stats SD	Stats Min	Stats P10	Stats P25	Stats P50	Stats P75	Stats P90	Stats Max	Stats >0.20
Winter	Algebra I	861	-0.02	0.03	-0.09	-0.06	-0.04	-0.02	-0.01	0.01	0.15	0
Winter	Biology	1431	-0.02	0.02	-0.09	-0.04	-0.03	-0.02	-0.01	0.01	0.10	0
Winter	Literature	780	-0.02	0.04	-0.12	-0.08	-0.05	-0.02	0.00	0.02	0.29	3
Spring	Algebra I	861	-0.02	0.03	-0.12	-0.06	-0.04	-0.02	0.00	0.02	0.09	0
Spring	Biology	1431	-0.02	0.02	-0.10	-0.04	-0.03	-0.02	-0.01	0.01	0.12	0
Spring	Literature	780	-0.02	0.05	-0.15	-0.09	-0.04	-0.02	0.00	0.02	0.25	2
Summer	Algebra I	861	-0.02	0.04	-0.14	-0.07	-0.05	-0.02	0.00	0.02	0.11	0
Summer	Biology	1431	-0.02	0.04	-0.13	-0.07	-0.04	-0.02	0.01	0.03	0.19	0
Summer	Literature	780	-0.02	0.06	-0.20	-0.10	-0.07	-0.02	0.02	0.05	0.30	7

ITEM FIT

WINSTEPS provides two item-fit statistics (infit and outfit) for evaluating the degree to which the Rasch model predicts the observed item responses. Each fit statistic can be expressed as a mean square (MnSq) statistic or on a standardized metric (Zstd with mean = 0 and variance = 1). MnSq values are more oriented toward practical significance, while Zstd values are more oriented toward statistical significance. Though both are informative, the Zstd values are very likely too sensitive to the large sample sizes observed on the Keystone Exams. In this situation it is recommended that the Zstd values be ignored if the MnSq values are acceptable (Linacre, 2009).

Both infit and outfit MnSq are the average of standardized residual variance (the difference between the observed score and the Rasch estimated score divided by the square root of the Rasch model variance). The difference is that the outfit statistic gives all examinees equal weight in computing the fit and tends to be affected more by unexpected responses far from the person, item, or rating scale category measure (i.e., it is more sensitive to outlying, off-target, low-information responses). The infit statistic is weighted by the examinee locations relative to item difficulty and tends to be affected more by unexpected responses close to the person, item, or rating scale category measure (i.e., informative, on-target responses). Some feel that extreme infit values are a greater threat to the measurement process than extreme outfit values since most tests intend to measure the on-target population rather than extreme outliers.

The expected MnSq value is 1.0, and it can range from 0 to infinity. Deviation in excess of the expected value can be interpreted as noise or lack of fit between the items and the model. Values lower than the expected value can be interpreted as item redundancy or overfitting items (too predictable, too much redundancy), and values greater than the expected value indicate underfitting items (too unpredictable, too much noise). Rules of thumb regarding practically significant MnSq values vary. More conservative users might prefer items with MnSq values that range from 0.8 to 1.2. Others believe reasonable test results can be achieved with values from 0.5 to 1.5. In the results below, values outside of 0.7 to 1.3 are given practical importance.

Table 12–3 presents the summary statistics of infit and outfit mean square statistics for the Keystone Exams in Algebra I, Biology, and Literature, including the mean, SD, and minimum and maximum values. The number of items within the range of [0.7, 1.3] is also reported in Table 12–3. As can been seen, the mean values for both fit statistics were close to 1.00 for all the exams. Most of the items had fit statistics falling in the range of [0.7, 1.3].

Table 12–3. Summary of Infit and Outfit Mean Square Statistics

Admin	Content Area	N	Infit Mean Square Mean	Infit Mean Square SD	Infit Mean Square Min	Infit Mean Square Max	Outfit Mean Square [0.7, 1.3]	Outfit Mean Square Mean	Outfit Mean Square SD	Outfit Mean Square Min	Outfit Mean Square Max	Outfit Mean Square [0.7, 1.3]
Winter	Algebra I	42	1.03	0.14	0.72	1.42	40	1.05	0.20	0.71	1.66	36
Winter	Biology	54	1.01	0.10	0.70	1.26	54	1.01	0.14	0.58	1.42	52
Winter	Literature	40	1.02	0.18	0.66	1.50	35	1.04	0.25	0.64	1.74	34
Spring	Algebra I	42	1.03	0.12	0.76	1.33	41	1.06	0.19	0.68	1.46	35
Spring	Biology	54	1.01	0.10	0.76	1.21	54	1.01	0.14	0.72	1.32	53
Spring	Literature	40	1.01	0.16	0.58	1.31	36	1.02	0.24	0.57	1.54	30
Summer	Algebra I	42	0.97	0.14	0.65	1.36	40	0.98	0.17	0.65	1.42	37
Summer	Biology	54	0.99	0.07	0.80	1.14	54	0.99	0.09	0.74	1.24	54
Summer	Literature	40	0.99	0.17	0.57	1.49	35	1.00	0.20	0.57	1.68	35

RASCH ITEM STATISTICS

As noted earlier, the Rasch model expresses item difficulty (and student ability) in units referred to as *logits*, rather than using the percent-correct metric. In the simplest case, a logit is a transformed *p*-value with the average *p*-value becoming a logit of zero. In this form, logits resemble *z*-scores or standard normal deviates; a very difficult item might have a logit of +4.0 and a very easy item might have a logit of –4.0. However, they have no formal relationship to the normal distribution.

The logit metric has several mathematical advantages over *p*-values. Logits have an interval scale, meaning that two items with logits of 0.0 and +1.0, respectively, are the same distance apart as two items with logits of +3.0 and +4.0. Logits are not dependent on the ability level of the students. For example, a test form can have a mean logit of zero, whether the average item *p*-value for the student sample is 0.8 or 0.3.

The standard Rasch calibration procedure arbitrarily sets the mean difficulty of the items on any form at zero. Under normal circumstances where all students are administered the same set of items, any item with a *p*-value lower than the average item on the form receives a positive logit and any item with a *p*-value higher than the average receives a negative logit. Consequently, the logits for any calibration relate to an arbitrary origin defined by the center of items on that form. Logits for both item difficulties and student abilities are placed on the same scale and relate to the same mean item difficulty.

There are a number of other choices that could be made for centering the item difficulties. Rather than using all the items, the origin could be defined by content. For the Keystone Exams, all test forms in a particular content area share the same operational item set. All items on each form can then be easily adjusted to a single origin by defining the origin as the mean of the operational items. With this done, the origins for all the forms will be statistically equal. For example, items on any two forms that are equally difficult will now have statistically equal logit difficulties.

Appendix J reports the item statistics including classical and Rasch logit difficulties for all the operational items and the field test items embedded in the spring forms. Table 12–4 summarizes the Rasch logit difficulties of the operational items on each test for each administration. The mean of MC item difficulty was no longer equal to zero as it was for the 2011 administration. This is because all the item parameter estimates were anchored to the pre-equated values. The mean item difficulties for MC items were smaller than those for the CR items. Table 12–4 also shows the mean standard errors (SE) of the item difficulties, which were relatively small, suggesting that items were calibrated with very small errors. The minimum (Min) and maximum (Max) values and standard deviations (SD) suggest the Keystone Exams items covered a relatively wide range of difficulties.

Table 12–4. Summary of Rasch Item Difficulties

Administration/Content Area	Item Types	N	Mean Item Difficulty	Mean SE	SD	Min	Max
Winter Algebra I	All	42	0.20	0.01	0.77	-1.08	2.16
Winter Algebra I	MC	36	0.04	0.01	0.66	-1.08	1.41
Winter Algebra I	CR	6	1.20	0.01	0.67	0.39	2.16
Winter Biology	All	54	-0.01	0.01	0.59	-1.33	1.29
Winter Biology	MC	48	-0.12	0.01	0.51	-1.33	0.97
Winter Biology	CR	6	0.86	0.01	0.43	0.26	1.29
Winter Literature	All	40	0.37	0.01	0.70	-1.48	1.54
Winter Literature	MC	34	0.31	0.01	0.73	-1.48	1.54
Winter Literature	CR	6	0.73	0.01	0.32	0.26	1.23
Spring Algebra I	All	42	0.59	0.01	0.88	-1.32	2.05
Spring Algebra I	MC	36	0.45	0.01	0.86	-1.32	1.68
Spring Algebra I	CR	6	1.45	0.00	0.49	0.95	2.05
Spring Biology	All	54	0.12	0.01	0.56	-1.10	1.05
Spring Biology	MC	48	0.03	0.01	0.52	-1.10	1.03
Spring Biology	CR	6	0.86	0.00	0.19	0.50	1.05
Spring Literature	All	40	0.34	0.01	0.81	-1.90	1.63
Spring Literature	MC	34	0.23	0.01	0.83	-1.90	1.63
Spring Literature	CR	6	0.92	0.00	0.24	0.68	1.32
Summer Algebra I	All	42	0.56	0.06	0.88	-1.55	2.43
Summer Algebra I	MC	36	0.40	0.06	0.83	-1.55	1.50
Summer Algebra I	CR	6	1.55	0.03	0.49	1.10	2.43
Summer Biology	All	54	0.12	0.07	0.55	-1.08	1.35
Summer Biology	MC	48	0.03	0.07	0.50	-1.08	0.96
Summer Biology	CR	6	0.88	0.04	0.38	0.33	1.35
Summer Literature	All	40	0.40	0.10	0.77	-1.52	1.98
Summer Literature	MC	34	0.26	0.11	0.75	-1.52	1.98
Summer Literature	CR	6	1.16	0.07	0.26	0.68	1.44

ITEM DIFFICULTY-STUDENT ABILITY MAP

The distributions of the Rasch item logits (item difficulty estimates) are shown on the item difficulty-student ability maps presented in Figure 12–1. In each item-student map, markers on the right-hand side represent student ability values, whereas markers on the left-hand side represent item difficulty parameter estimates. As noted earlier, the Rasch model enables placement of both items and students on the same scale. Consequently, one can easily visualize information about how the difficulty of the test items related to the ability distribution of students who took the test. The students located in the upper-right quadrant of any given plot have relatively higher ability. Items in the lower-left quadrant are relatively easier.[1] The three dashed lines (from bottom to top) represent the performance cuts: below basic/basic, basic/proficient, and proficient/advanced. The common pattern seen across all maps was that the item difficulties were comparable to the student ability levels. It is also important to understand where the items are providing more accurate measurement. This issue is addressed more fully in Chapter Eighteen (see Figure 18–2).

[1] Of course, high-ability students have higher probabilities of correctly answering easier items. Similarly, low-ability students (in lower-right quadrant of any given plot) have lower probabilities of answering harder items (in upper-left quadrant).

Figure 12–1. Item Difficulty-Student Ability Maps

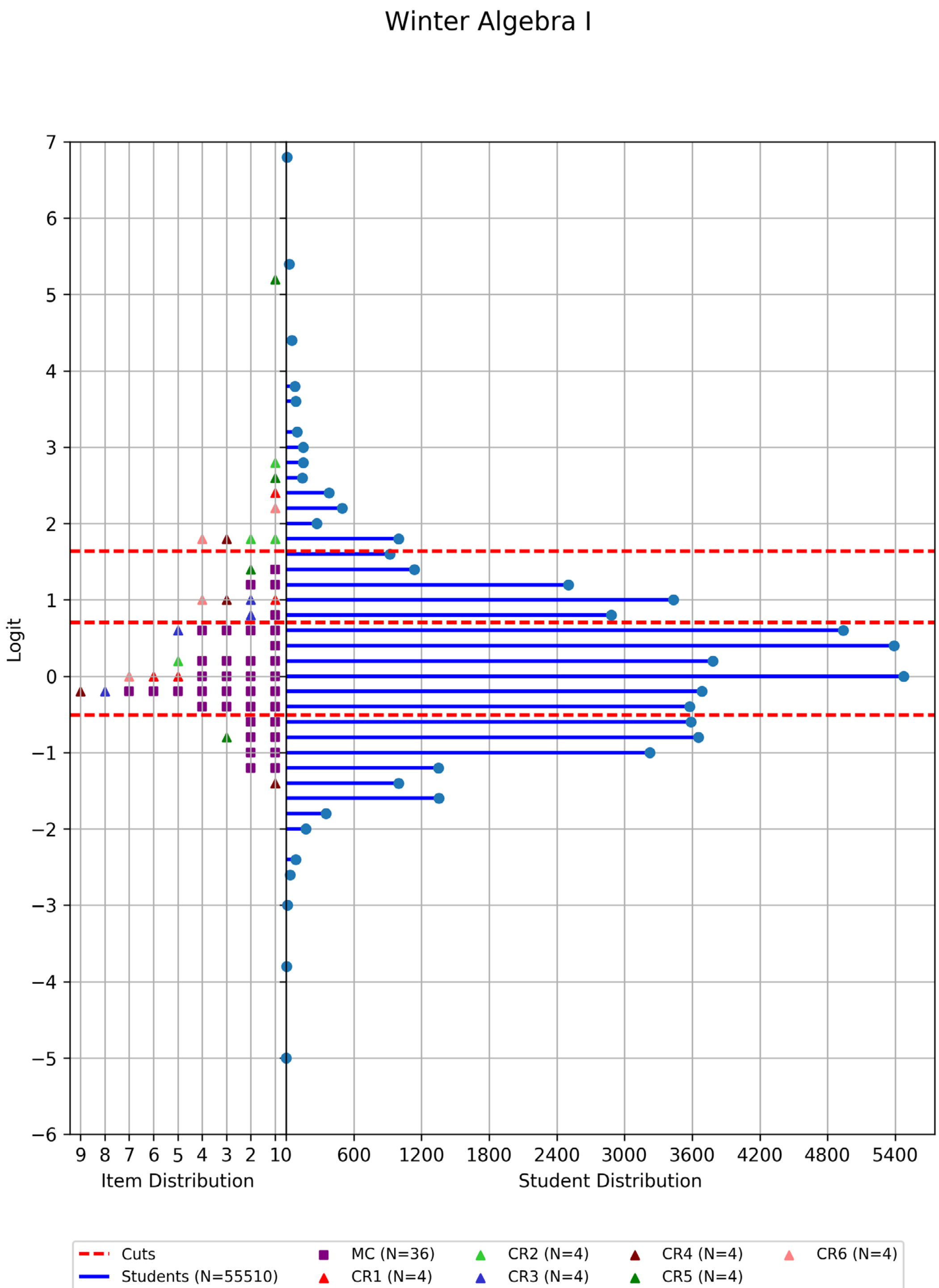


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

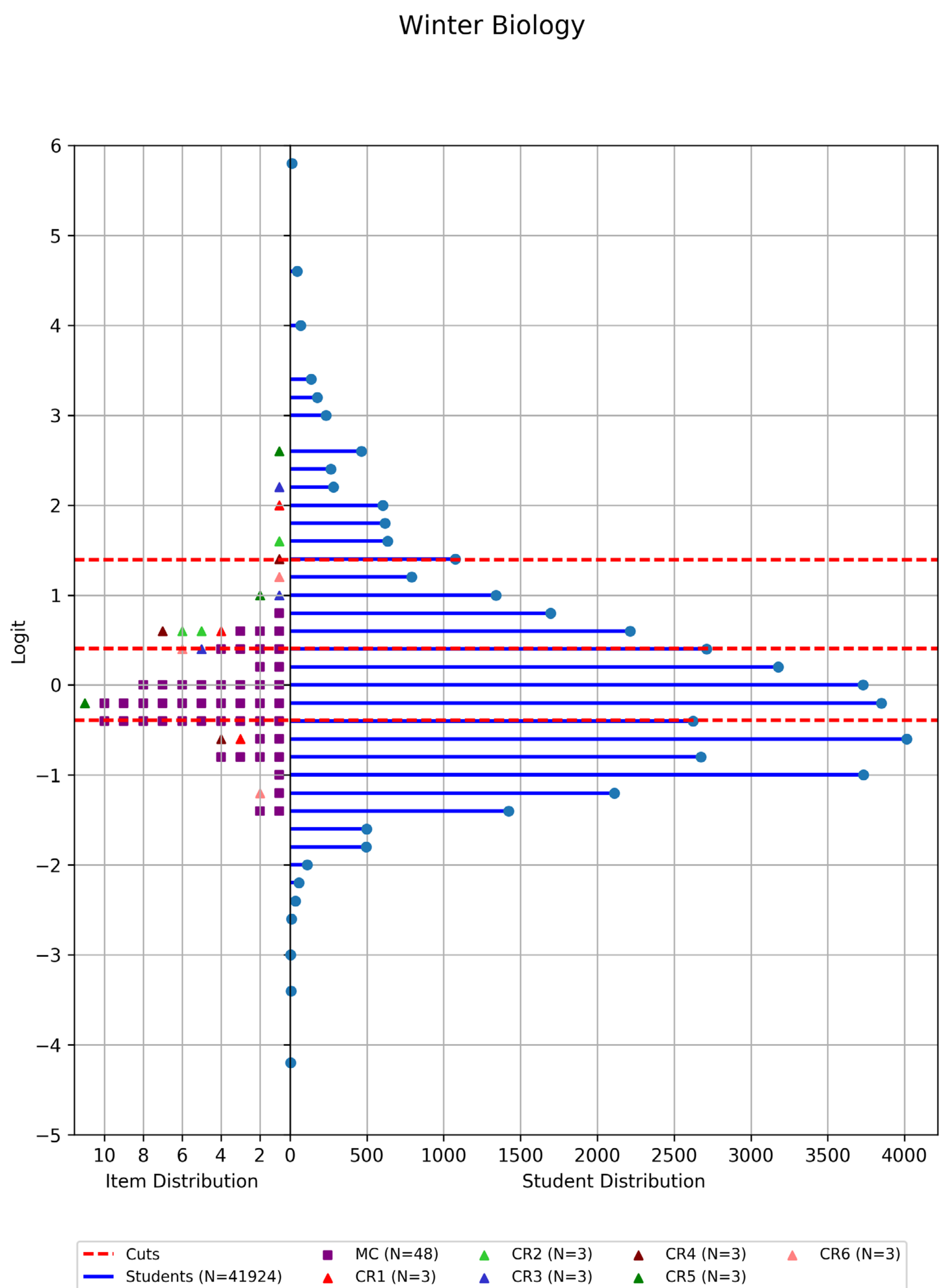


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

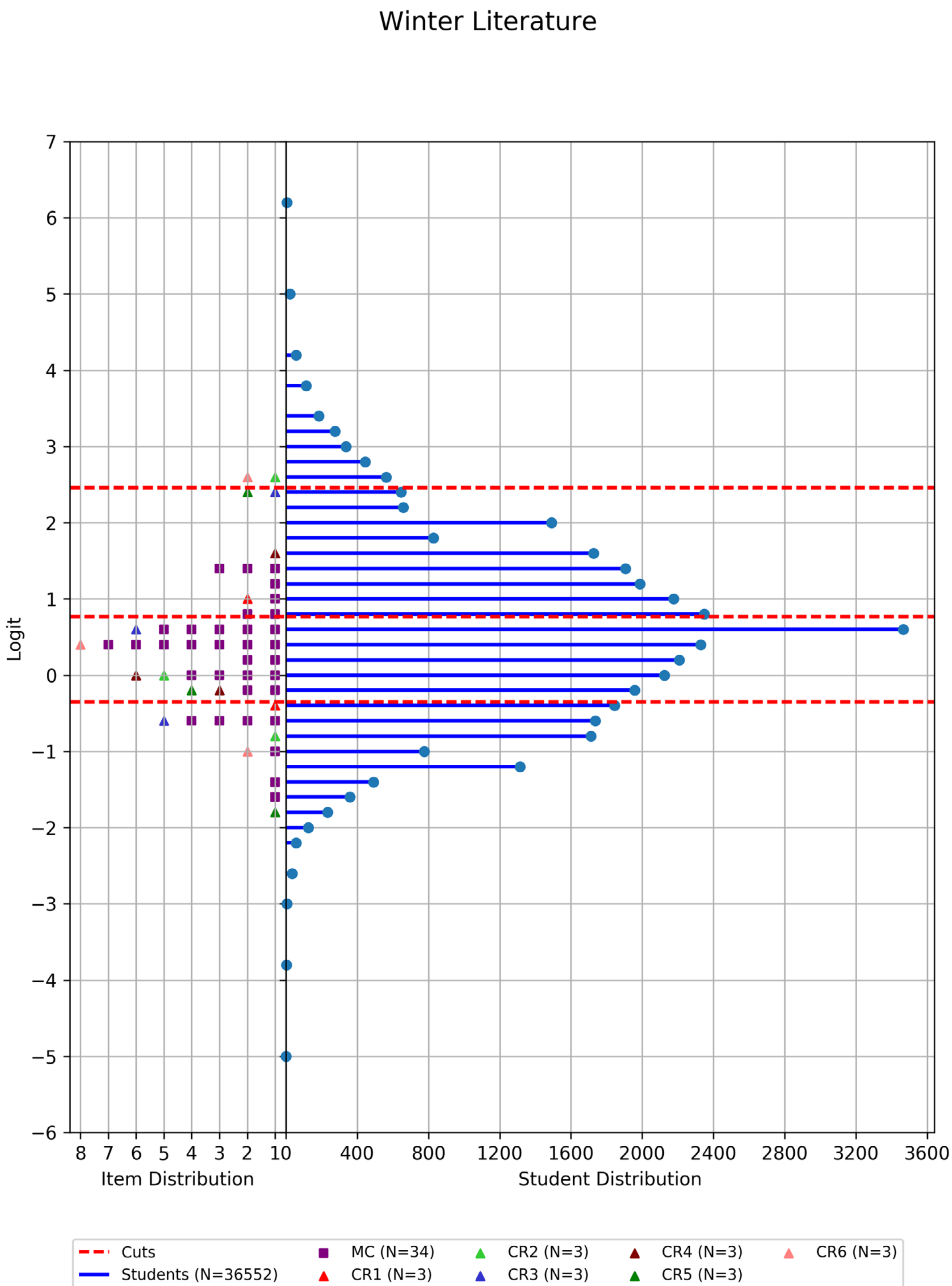


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

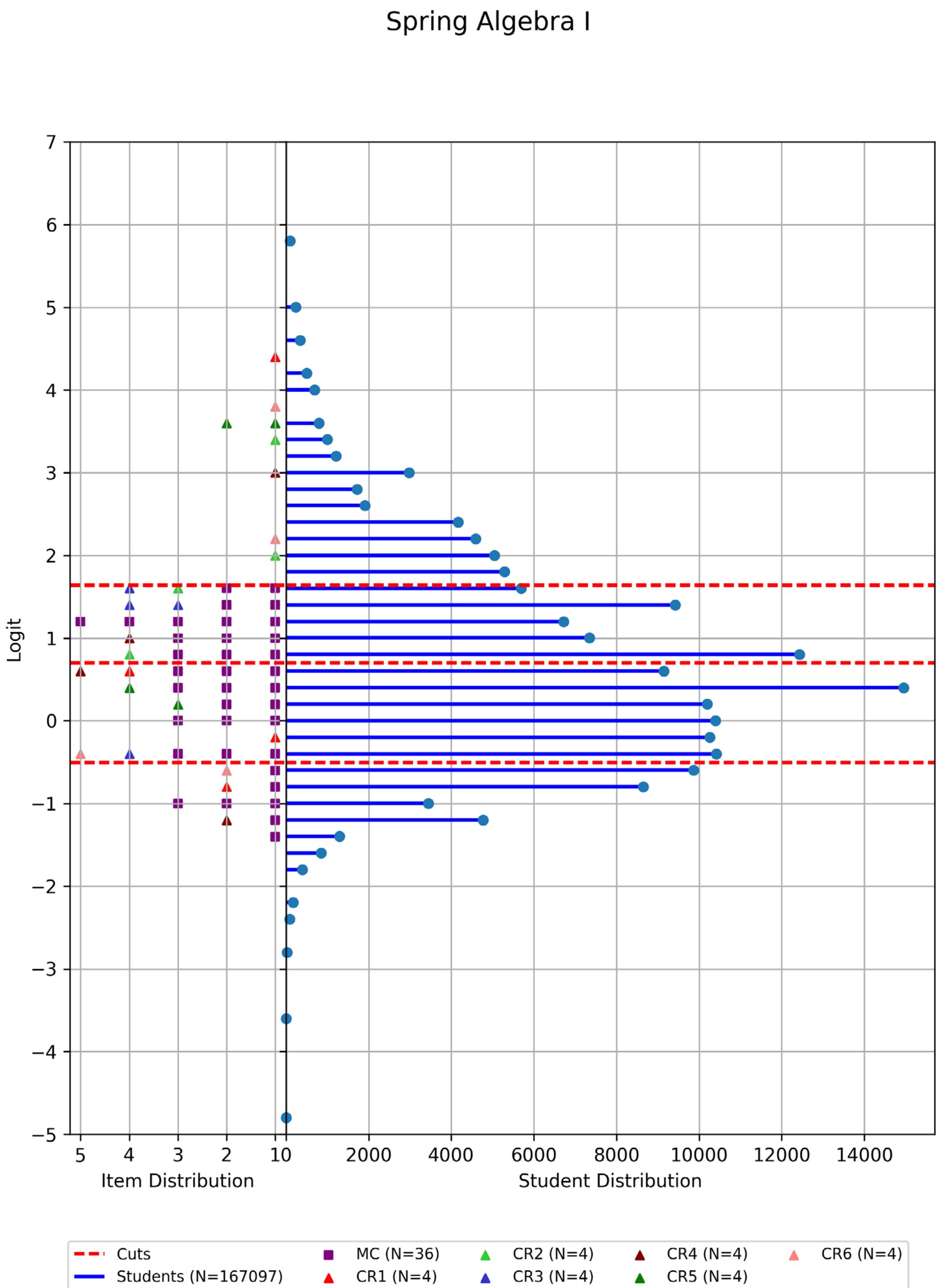


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

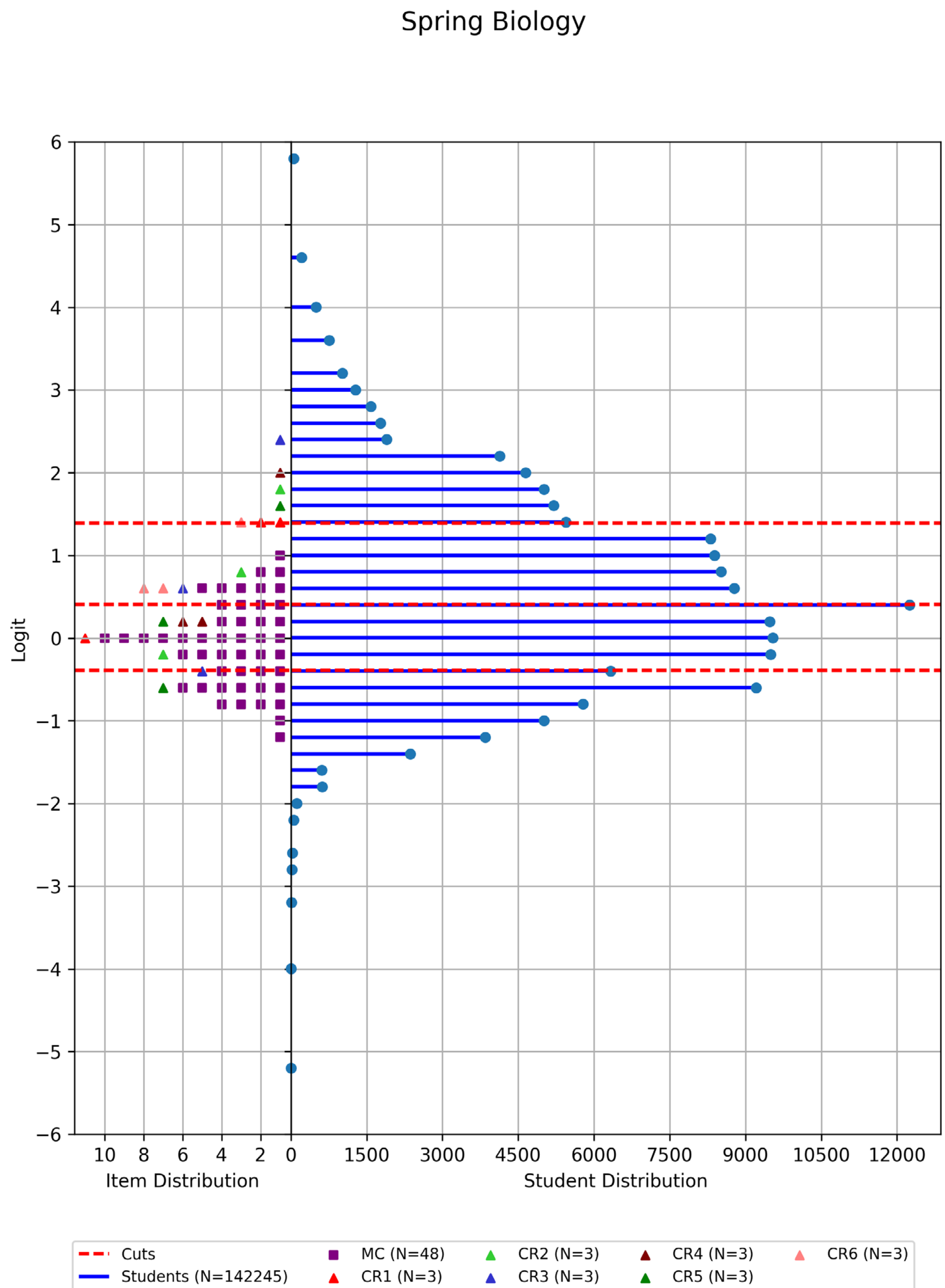


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

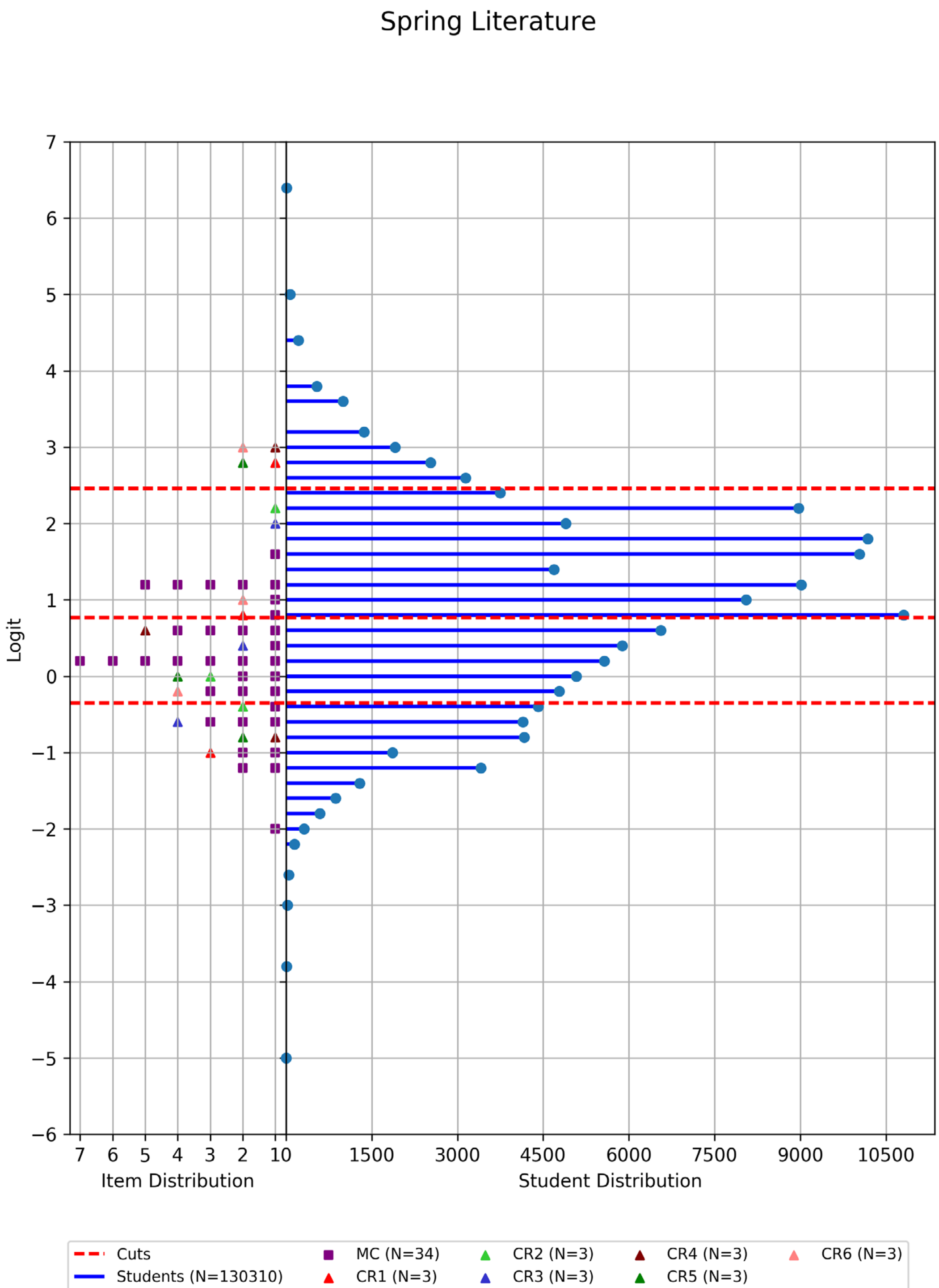


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

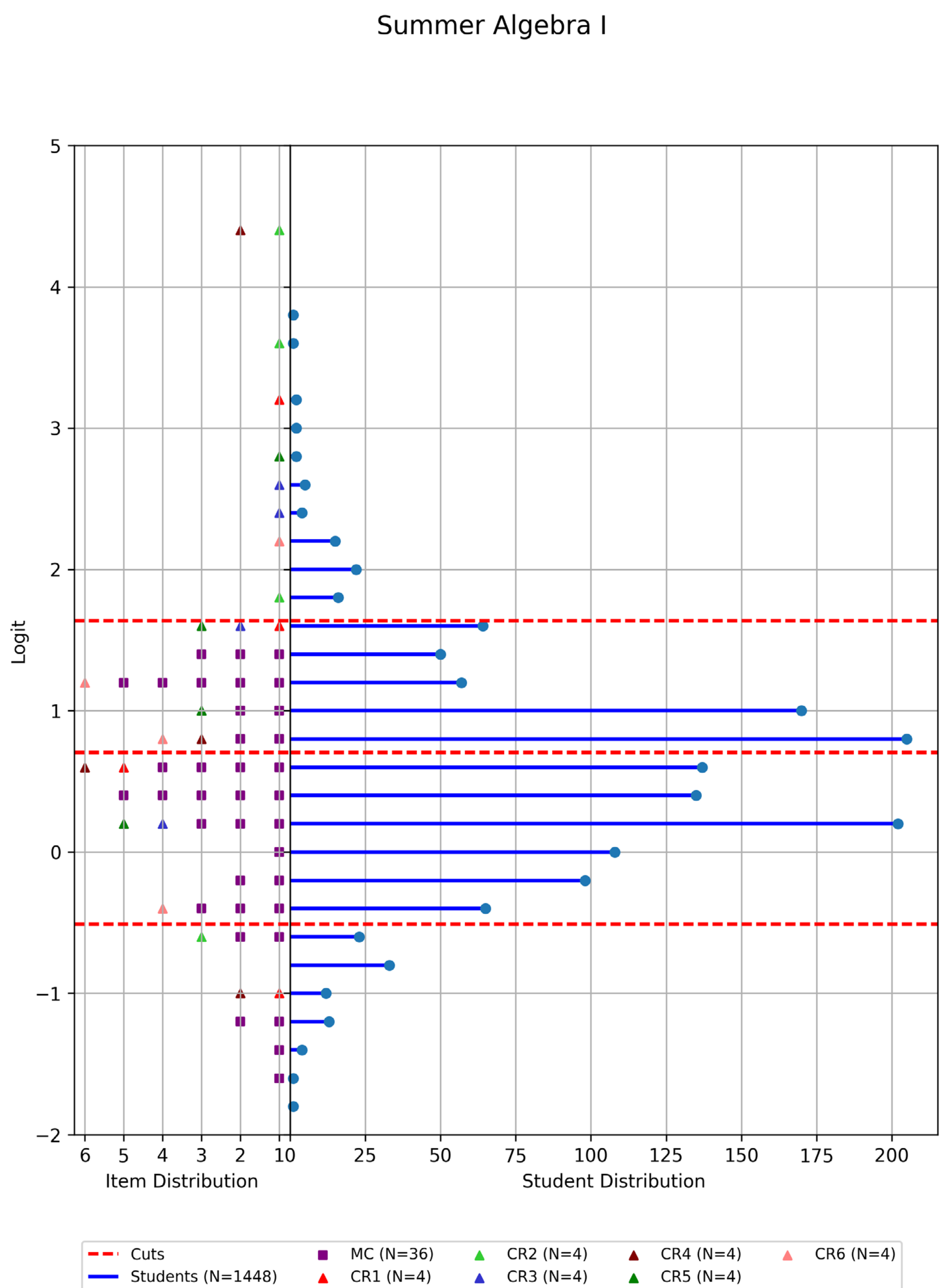


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

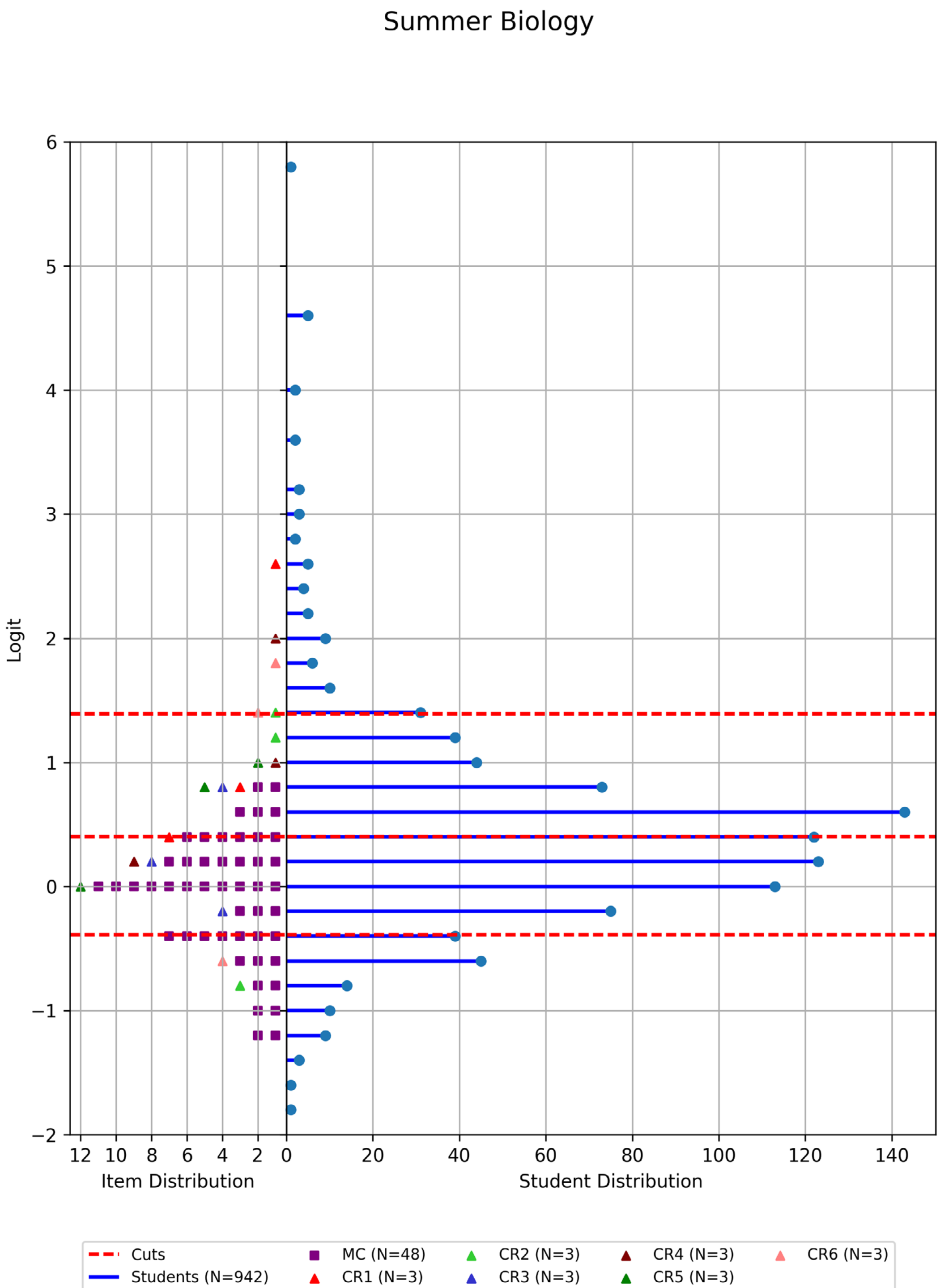
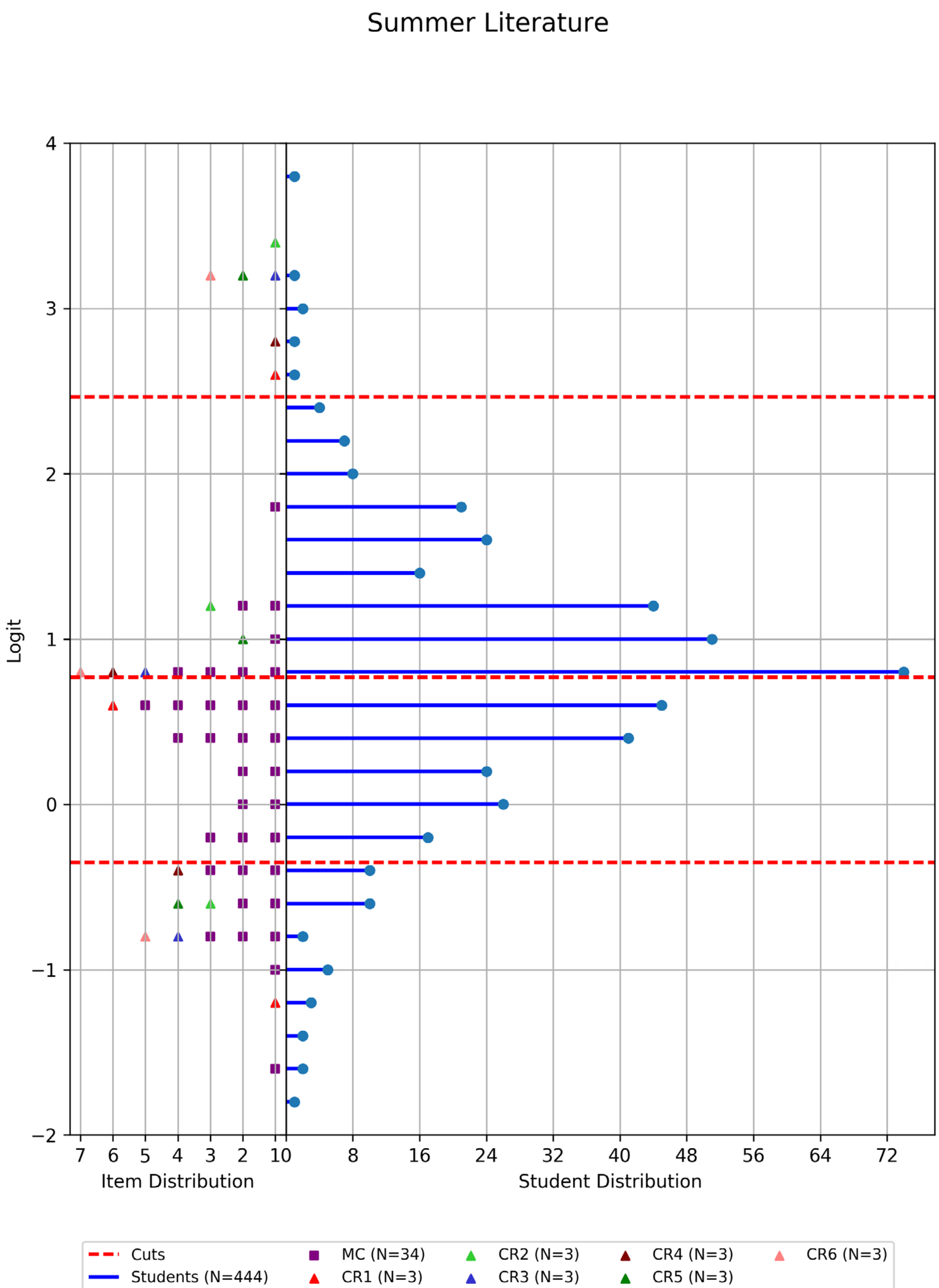


Figure 12–1 (continued). Item Difficulty-Student Ability Maps



CHAPTER THIRTEEN: STANDARD SETTING

STANDARD SETTING AND PERFORMANCE LEVEL DESCRIPTORS

The Keystone Performance Level Descriptors (PLDs) are paragraphs that describe the knowledge and skills expected at different performance levels with respect to the content standards (Pennsylvania Keystone Exams Assessment Anchor Content Standards and Eligible Content) for each of the Keystone Exams. Descriptors must be clearly written to ensure that all stakeholders have a common understanding of what describes expected performance at the various levels (i.e., Below Basic, Basic, Proficient, and Advanced). PLDs were developed, reviewed, and finalized by the PDE/QRT[1] and committees of Pennsylvania educators as required by the Chapter 4 Regulations. After the development and final review by PDE/QRT and Pennsylvania educators, the descriptors were prepared for use during the standard setting workshop. During this meeting, the descriptors were used to guide the standard setting process for each of the Keystone Exams. They were instrumental to the validity and defensibility of the standard setting process.

The standard setting for the Algebra I, Biology, and Literature Keystone Exams was conducted by Data Recognition Corporation (DRC) using a Bookmark procedure (Lewis, Mitzel, & Green, 1996) during a workshop held in Harrisburg, Pennsylvania, June 23–24, 2011. After the standard setting event, the descriptors were finalized. Along with the recommended cut scores, final PLDs for each of the Keystone Exams were submitted to the Pennsylvania Board of Education for final approval.

Below is a summary of the process that was used to guide the development of the Keystone Exams PLDs and a summary of the methodology and results of the standard setting workshop. Additional details about the standard setting event can be found in the *Keystone Standard Setting Technical Report* (Pennsylvania Department of Education, 2011).

DEVELOPMENT OVERVIEW FOR THE PERFORMANCE LEVEL DESCRIPTORS

The Keystone Exams PLDs were developed by Pennsylvania educators during two meetings. The goal of the first meeting was to have Pennsylvania educators review the general Pennsylvania Policy Definitions that describe, at a high level, performance expected for each level and complete an in-depth analysis of the Keystone Exams Assessment Anchors and Eligible Content in order to create a bulleted list describing, in detail, what students are expected to know and be able to do at each performance level. The goal of the second meeting was to have committees of Pennsylvania educators review the Pennsylvania Policy Definitions again and draft general descriptors (paragraphs) that build upon and/or summarize the information from the bulleted lists of what students are expected to know and be able to do at each performance level.

Guiding documents were prepared for each meeting. The guiding documents included the following:

- PowerPoint training presentations
- Meeting agendas
- Assessment Anchors and Eligible Content documents
- Policy definitions
- Other relevant materials as needed to help guide the work of the committees

All meeting materials were submitted to PDE/QRT for review and approval before each Keystone Exams meeting following an agreed-upon development schedule. The following section provides specific information concerning each meeting.

[1] The PDE/QRT includes the representatives from the Pennsylvania Department of Education, members of the Quality Review Team, and/or others appointed by the Quality Review Team.

ROLE OF FACILITATORS AND OBSERVERS FOR THE MEETINGS

The role of the facilitators was to ensure that a fair and orderly consensus process was followed for each meeting, that the committee members' work was adequately documented, and that the process stayed on schedule. The facilitators developed the agenda, prepared all meeting materials such as the PowerPoint training presentations and the task-guiding documents, and provided the initial training on the development of the specific descriptors (meeting 1) and the general descriptors (meeting 2). PDE/QRT members supported the facilitation process and/or served as observers of the process.

The facilitators also served as a resource, answering questions pertaining to the content of the standards (Assessment Anchor Content Standards and Eligible Content) and the documents developed to guide the process. Facilitators also summarized the results of each meeting, finalized the results, and prepared the specific descriptors/bulleted lists (meeting 1) and the general descriptors (meeting 2) for PDE/QRT review and approval.

PERFORMANCE LEVEL DESCRIPTORS MEETING 1

CREATING SPECIFIC LISTS DESCRIBING WHAT STUDENTS SHOULD KNOW AND BE ABLE TO DO AT EACH PERFORMANCE LEVEL

The first PLD meeting for Algebra I, Biology, and Literature Exams was held May 18–19, 2010, in Harrisburg, Pennsylvania. The purpose of the first meeting was to guide Pennsylvania educators in understanding the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature for what the Commonwealth of Pennsylvania determined students should know and be able to do for a given Keystone Exam subject. Committee members applied this understanding to the development of a bulleted list of specific determinations as to the level of knowledge and skills deemed necessary for each performance level. The section below describes the process used in the first meeting.

TRAINING

Pennsylvania educators received general training on how to develop specific PLDs, including training on how to describe student performance in relation to the Keystone Exams Assessment Anchors and Eligible Content. The training also provided educators with a general overview of the Standards Aligned System (SAS) and the high-level plan for the Keystone Exams. Definitions of key terms (e.g., Assessment Anchor Content Standards, Eligible Content, Performance Level Descriptors) were provided along with information on the background and purpose of the Keystone Exams. Keystone Exams content-specific materials (e.g., Assessment Anchor Content Standards, Eligible Content, other guiding documents) were distributed. The PDE/QRT also provided information on the policy definitions for existing Pennsylvania assessments.

ANALYZING THE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT AND THE GENERAL POLICY DEFINITIONS FOR PENNSYLVANIA ASSESSMENTS

Following the introductory training, educators were divided into groups according to each Keystone Exam. Each group focused specifically on the task at hand—developing the specific PLDs for a given Keystone Exam. Committee members were informed of the format of the specific descriptors (bulleted list) and the number of proposed performance levels for each Keystone Exam (Below Basic, Basic, Proficient, and Advanced). Committee members were then given time to familiarize themselves with the policy definitions and the Assessment Anchors and the Eligible Content for a given Keystone Exam. They were provided with PDE/QRT-approved guiding documents to facilitate the process. Beginning with Proficient, committee members were asked to draft, in bulleted-list format, each performance level for Basic, Proficient, and Advanced, making sure to consider the knowledge and skills required or deemed necessary for each performance level. Note: Educators were not asked to create a specific descriptor for Below Basic.

DRAFTING SPECIFIC DESCRIPTORS

Outlined below is the sequence of steps taken to develop specific descriptors. The sequence was not always followed exactly. For example, some steps occurred simultaneously; other steps were repeated as needed or reordered as necessary.

1. The committee began with the development of the bulleted list for Proficient to serve as a model for the work during the remainder of the development process. As a formative first task using the Assessment Anchors and Eligible Content and the Pennsylvania Policy Definition for Proficient, the committee was asked to discuss, deliberate, and reach consensus on its initial bulleted list of the knowledge and skills needed to be considered Proficient. During this process, members were encouraged to consult all available resources and guiding documents. Particular emphasis was placed on the alignment of the knowledge and skills necessary for Proficient performance with what students are expected to know and be able to do as defined by the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.
2. Once the committee drafted a bulleted list of the knowledge and skills needed to describe Proficient performance based upon the Assessment Anchors and Eligible Content, a group discussion took place. In reviewing the bulleted list for Proficient, the educators were specifically asked to determine whether all members agreed that the list included the appropriate knowledge and skills from the Assessment Anchors and Eligible Content to describe the Proficient performance level and that all Assessment Anchors and Eligible Content were sufficiently addressed.
3. The results of the discussion were summarized, and suggested revisions were made. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Committee consensus was reached.
4. Following development of the bulleted list of the knowledge and skills needed for the Proficient performance level as determined by the committee, the committee began the development of the bulleted lists describing the specific knowledge and skills needed for Basic and Advanced. To complete the task, the committee members followed the procedures analogous to those used to develop the specific bulleted list for the Proficient performance level. These procedures included, as a formative first task, the committee's use of the Assessment Anchors and Eligible Content and the Pennsylvania Policy Definitions (e.g., Basic, Advanced) to discuss, deliberate, and reach consensus on its initial bulleted list of the knowledge and skills needed for Basic and then Advanced. This order of development—Proficient first, followed by Basic and then Advanced—was followed throughout the remainder of the process.

Once the initial drafts of the bulleted lists for Basic, Proficient, and Advanced were developed, a group discussion took place. To guide the discussion, the following questions were used to evaluate each specific descriptor (bulleted list) for a given performance level (Basic, Proficient, Advanced):

- Is the description of the performance level appropriate? If not, what revisions need to be made?
- Is the description of the specific Keystone Exam inappropriate because the list of knowledge and skills included in the description of the performance level is too demanding? If so, what revisions need to be made?
- Is the description inappropriate because the knowledge and skills included in the description of the performance level is inconsistent with the expectation of the high standards as reflected in the Policy Definition? If so, what revisions need to be made?
- Is the description inappropriate because the knowledge and skills included in the description of the performance level might be too easy? If so, what revisions need to be made?

5. The results of the discussion were summarized, and suggested revisions were listed. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Depending upon the degree of concurrence, the facilitators proposed revisions based on the committee members' feedback to the specific descriptors (bulleted lists) for each descriptor. Committee consensus was reached.
6. Once consensus was reached, the bulleted lists or specific descriptions for each performance level were reviewed once again to confirm that all Assessment Anchors and Eligible Content were sufficiently addressed for each performance level and that the lists showed a clear progression from one performance level to the

next. The results of the discussion were summarized, and suggested revisions were listed. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Depending upon the degree of concurrence, the facilitators proposed revisions to the lists for each descriptor based on the committee members' feedback. Committee consensus was reached.

7. Following completion of the committee's work, the specific PLDs or bulleted lists of the knowledge and skills needed for each descriptor were collected. The bulleted lists were prepared for final review by the PDE/QRT. Upon approval by the PDE/QRT, the bulleted lists of the knowledge and skills describing each performance level were posted on the PDE website for additional review and feedback.

PERFORMANCE LEVEL DESCRIPTORS MEETING 2

CREATING GENERAL DESCRIPTIVE PARAGRAPHS DESCRIBING WHAT STUDENTS SHOULD KNOW AND BE ABLE TO DO AT EACH PERFORMANCE LEVEL

The second meeting for Algebra I, Biology, and Literature Exams took place in Harrisburg, Pennsylvania, on April 27–28, 2011. The second meeting built upon the work completed at the first meeting. The purpose of the second meeting was to guide the committee of Pennsylvania educators in developing general PLDs (paragraphs) for each of the performance levels (Basic, Proficient, and Advanced). These paragraphs were clearly written to ensure all stakeholders have a common understanding of what describes expected performance at the various levels. The paragraphs were not to be as specific as the bulleted lists but were to be aligned to the bulleted lists. In order to complete the task, the educators reviewed the Pennsylvania Policy Definitions for the performance levels.

Table 13–1. Pennsylvania Policy Definitions

Level	Description
Advanced	The Advanced Level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
Proficient	The Proficient Level reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
Basic	The Basic Level reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. This work is approaching satisfactory performance, but has not been reached. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient Level.
Below Basic	The Below Basic Level reflects inadequate academic performance. Below Basic work indicates little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient Level.

The committee members then reviewed the specific bulleted list describing the knowledge and skills for Proficient based upon the Assessment Anchors and Eligible Content to determine whether the list of knowledge and skills provided in the bulleted list was still in alignment with the Policy Definition for Proficient. This review by the committee also included an in-depth analysis of the Assessment Anchors and Eligible Content. The section below describes, in detail, the process used in the second meeting.

TRAINING

Pennsylvania educators received general training on how to develop general descriptors (paragraphs) that describe performance at the various levels, including training on how to describe student performance in relation to the Keystone Exams Assessment Anchors and Eligible Content. The training also included providing Pennsylvania educators with a general overview of the SAS and the high-level plan for the Keystone Exams. Definitions of key terms (e.g., Assessment Anchor Content Standard, Eligible Content, specific and general Performance Level Descriptors) were provided along with information on the background and purpose of the Keystone Exams. A review of the Pennsylvania Policy Definitions was also included in the training, including a discussion of how the policy definition for Proficient relates to what it means to be Proficient on a given Keystone Exam. Content-specific materials (e.g., Policy Definitions, Assessment Anchor Content Standards and Eligible Content, specific descriptors or bulleted lists from the first meeting, other guiding documents) were also distributed.

ANALYZING THE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT AND THE POLICY DEFINITION FOR PROFICIENT

Following the introductory training, Pennsylvania educators were divided into groups according to Keystone Exam. Each group focused specifically on the task at hand—developing the general PLD paragraphs (Basic, Proficient, and Advanced) for a given Keystone Exam. To begin the process, educators reviewed the Pennsylvania Policy Definition for Proficient.

DRAFTING GENERAL DESCRIPTOR PARAGRAPHS

Once the committee reviewed the bulleted list for alignment to the Policy Definition for Proficient, committee members were asked to describe, in general terms, the knowledge and skills deemed necessary for each performance level (Basic, Proficient, and Advanced), beginning with Proficient. As a formative first task, committee members were instructed to refer to the bulleted list of the knowledge and skills required or deemed necessary for each performance level. Outlined below is the sequence of steps for the process used to develop the general PLD paragraphs. The sequence was not always followed exactly. For example, some steps occurred simultaneously; other steps were repeated as needed or reordered as necessary.

The committee began with the development of the general descriptor paragraph for the Proficient performance level. This general descriptor served as a model for the committee's work during the remainder of the development process. Using the Assessment Anchors and Eligible Content, the specific descriptors (bulleted list) for Proficient, and the Pennsylvania Policy Definition for Proficient, the committee was asked to discuss, deliberate, and reach consensus on a written description of the knowledge and skills needed for Proficient. During the process, members were encouraged to consult all available resources and guiding documents. Particular emphasis was placed on the alignment of the knowledge and skills necessary for the Proficient performance descriptor to the Assessment Anchors and Eligible Content for the given Keystone Exam.

Note: In order to help guide educators in the development of the general descriptor paragraph for Proficient, samples of descriptor paragraphs for Algebra I, Biology, and Literature (e.g., Georgia, North Carolina) were provided. The committee members were encouraged to approach the task by noting how the sample general descriptors must provide the right words to define performance—having a balance between keeping the description of Proficient general enough yet not as specific as the bulleted list. Committee members were also encouraged not to focus too heavily upon style, grammar, and mechanics at this stage. In other words, committee members were not to serve as "wordsmiths."

1. Once an initial draft paragraph summarizing the knowledge and skills needed to describe Proficient performance was developed, a group discussion took place. Committee members were asked to review the draft paragraph and determine whether the paragraph provided a clear description of what it means to be Proficient on a given Keystone Exam and the Policy Definition for Proficient. The goal of the discussion was to reach consensus.
2. Following development of the general paragraph describing Proficient, the committee began the development of the general paragraph describing Basic performance and the general paragraph describing Advanced performance. To complete the task, the committee members followed the procedures analogous to those used to develop the general paragraph describing Proficient performance on a given Keystone Exam. This process included, as a formative first task, using the Assessment Anchors and Eligible Content, the specific descriptors (bulleted lists), and the Pennsylvania Policy Definitions for a given level (e.g., Basic, Advanced) and discussing, deliberating, and reaching consensus on the knowledge and skills needed for Basic and the knowledge and skills needed for Advanced. This order of development—Proficient first, followed by Basic and then Advanced—was followed throughout the remainder of the process.
3. Once the initial draft paragraphs were developed for the other performance levels, a group discussion took place. In reviewing the state of development of the general PLD paragraphs at this stage, the committee members were asked to consider the following questions:
4. Does each paragraph clearly summarize the knowledge and skills required for a given performance level (Basic, Proficient, and Advanced)? If not, what revisions need to be made?

5. Does each paragraph provide for an appropriate description of the performance level? In other words, does each paragraph provide an overview or summary of the knowledge and skills appropriate for a given performance level? If not, what revisions need to be made?
6. Does any paragraph provide information that should not be included in the description of the performance level? If so, what revisions need to be made?
7. Is there information in any PLD paragraph that does not align well with the Pennsylvania Policy Definitions for a given performance level? If so, what revisions need to be made?
8. Do any paragraphs include information that might be inconsistent with the knowledge and skills defined by the Assessment Anchors and Eligible Content? If so, what revisions need to be made?
9. Does any paragraph include information describing performance that might be too demanding or too easy? If so, what revisions need to be made?
10. The results of the discussion were summarized, and revisions to each general PLD paragraph were made. Committee consensus was reached.
11. Once consensus was reached, the paragraphs describing performance at each level were reviewed again by the committee to confirm the following:
12. The PLD paragraphs show a clear progression from one performance level to the next level.
13. The PLD paragraphs are consistent with the Pennsylvania Policy Definitions.
14. The PLD paragraphs are aligned to the Assessment Anchors and Eligible Content.
15. The results of the discussion were summarized, suggested revisions were made, and committee members' feedback was incorporated into the paragraphs. Committee consensus was reached.
16. Following completion of the committee's work, the general PLD paragraphs were provided to PDE/QRT for final review and feedback. Upon approval by PDE/QRT, the general PLD paragraphs were used to guide the standard setting process.

STANDARD SETTING

A major purpose in the design of the standard setting workshop for the Keystone Exams is to establish procedures to set the performance cuts for the newly developed exams and, at the same time, adhere to the framework required by federal guidelines (USED, 2004) for setting performance levels. Federal guidelines (USED, 2004: Sect 2) specify that the setting of performance standards must involve the following elements:

- Formal adoption of performance categories that comprise at least three levels
- Pluralistic representation by education stakeholders, to include, for example, members of the public, school teachers and administrators, special education teachers, etc.
- Performance standards based primarily on expert judgment regarding content-based expectations of student achievement, but including the consideration of student impact data
- Descriptions of the competencies associated with each performance level

Accordingly, the standard setting workshop is designed to satisfy the following goals:

- A defensible and federally acceptable standard setting methodology that emphasizes a content-based approach for recommending the new performance standards
- The incorporation of PLDs developed by Pennsylvania educators into the standard setting process. (The larger goal around the incorporation of PLDs into the process is to help ensure the alignment of Pennsylvania's content standards to performance expectations as established by the recommended cut scores.)

The panelists were informed that the results from this meeting would be presented to the Board for review and possible adoption.

PANELIST RECRUITMENT

PDE selected committee members for the Algebra I, Biology, and Literature standard setting workshop mostly from members who participated in the May 2010 and April 2011 Performance Level Advisory committees. These committee members were selected as the starting pool because they represented the diversity of the Commonwealth of Pennsylvania, had a mix of teaching and committee experience, and, most importantly, were familiar with the PLDs of the Keystone Exams. From this list, PDE selected a subset of 25 members for Algebra I, 25 members for Biology, and 23 members for Literature to serve as eligible candidates. DRC, in collaboration with PDE and its Technical Advisory Committee (TAC), established a target of 15 to 20 participants for each of the Keystone Exams in Algebra I, Biology, and Literature.

Between March and June 2011, a great effort was made to recruit enough panelists to meet the target number of participants. In accordance with federal guidelines for representative committees and TAC's recommendation of recruiting a few committee members with higher education experience, the following background factors were applied in the recruitment decision:

- Gender
- Ethnicity
- Grade level and higher education experience
- Content expertise
- Geographic location
- Specializations
- Experience in developing state academic standards, state assessments, and other related activities

However, due to the unavailability of and the cancellation by some committee members, a total of 15, 13, and 11 panelists attended the standard setting workshop for Algebra I, Biology, and Literature, respectively. Table 13–2 contains the summary information about the characteristics of the selected panelists for each content area based on their self-reported responses to the Participant Survey. As can be seen from this table, there were committee members who considered themselves minority in the Algebra I and Literature groups. There were also committee members with administration and/or teaching experience in higher education, special education, and/or individualized education plan (IEP); those with experience working in different regions; and those with different lengths of teaching experience.

Table 13–2. Self-Reported Demographic Composition of Panelists by Content Area

Demographic Information	Algebra 1 Number	Algebra 1 Percent	Biology Number	Biologyl Percent	Literature Number	Literature Percent
Gender: Male	9	60.0%	5	38.5%	5	45.5%
Gender: Female	6	40.0%	8	61.5%	6	54.5%
Ethnicity: Asian	1	6.7%	0	0.0%	0	0.0%
Ethnicity: American Indian	0	0.0%	0	0.0%	0	0.0%
Ethnicity: Black	1	6.7%	0	0.0%	1	9.1%
Ethnicity: Latino	0	0.0%	0	0.0%	0	0.0%
Ethnicity: Multi-Race	0	0.0%	0	0.0%	0	0.0%
Ethnicity: White	13	86.7%	13	100.0%	10	90.9%
Role: Classroom Teacher	8	53.3%	9	69.2%	4	36.4%
Role: Educator	3	20.0%	0	0.0%	0	0.0%
Role: Higher Education Educator	3	20.0%	1	7.7%	4	36.4%
Role: Other	1	6.7%	3	23.1%	3	27.3%
Special Education: Yes	7	46.7%	7	53.8%	4	36.4%
Special Education: No	2	13.3%	4	30.8%	3	27.3%
Special Education: N/A	6	40.0%	2	15.4%	4	36.4%
LEP: Yes	4	26.7%	5	38.5%	2	18.2%
LEP: No	4	26.7%	6	46.2%	5	45.5%
LEP: N/A	6	40.0%	2	15.4%	4	36.4%
LEP: Missing	1	6.7%	0	0.0%	0	0.0%
Region: Urban	3	20.0%	2	15.4%	2	18.2%
Region: Suburban	7	46.7%	5	38.5%	5	45.5%
Region: Rural	4	26.7%	6	46.2%	3	27.3%
Region: Other	1	6.7%	0	0.0%	1	9.1%
Experience: Less than 10 years	0	0.0%	5	38.5%	0	0.0%
Experience: 10–20 years	4	26.7%	2	15.4%	4	36.4%
Experience: 20–30 years	8	53.3%	4	30.8%	3	27.3%
Experience: More than 30 years	3	20.0%	2	15.4%	4	36.4%

MATERIALS PREPARATION

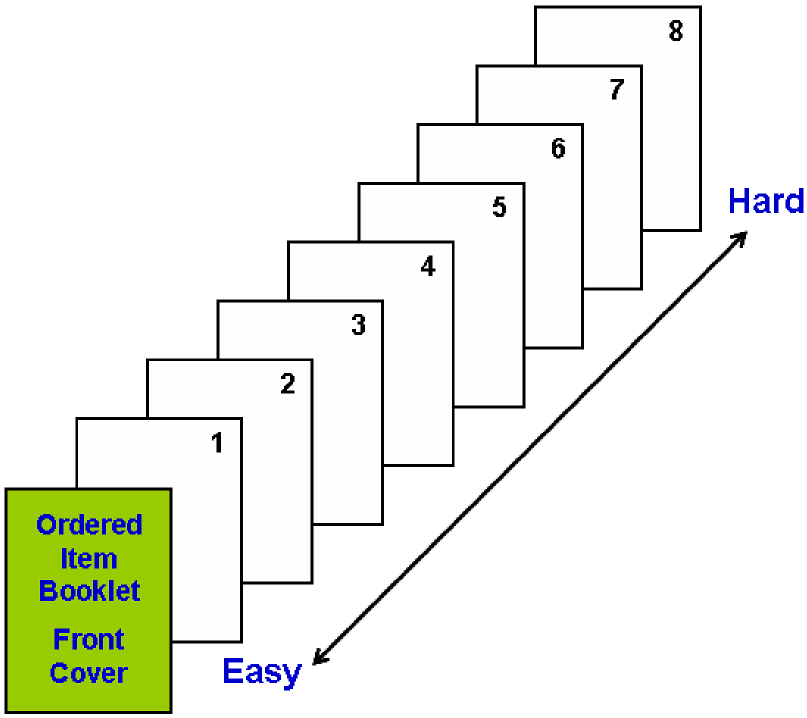
Workshop materials were developed and printed by DRC. The following is a list of materials that were available to panelists during the workshop:

- Item Map
- Item Separation Map
- Ordered Item Booklet (OIB)
- Passages
- Scoring Rubrics
- 2011 Operational Test Form
- PLDs
- Content Standards
- Participant Rating Form
- Participant Survey
- Readiness Form
- Evaluation Form
- Adhesive bookmarks, pens, highlighters, etc.

Item Map. The item map is a summary document displaying relevant information regarding each item. It contains the OIB page number, the original test sequence, item type, key, and content standard. The item map is ordered by difficulty in the same manner as the ordered item booklet. The item separation map is a graphical display of the relative difficulty of each item.

Ordered Item Booklet. The ordered item booklet is composed of all the operational items included in the test given to students in Spring 2011. Items are ordered from the easiest to the hardest. Each page contains an item and a page number. For constructed-response (CR) items, each score point with a sample response has a unique location in the OIB. A visual illustration of the OIB is provided in Figure 13–1.

Figure 13–1. Illustration of Ordered Item Booklet



To ensure there was no item difficulty gap for the items in an OIB, a few field test items were added to the OIBs. Table 13–3 shows the number of items supplemented into the OIBs by content area.

Table 13–3. Number of Score Points in OIB and Number of Items Supplemented

Exam	Number of Score Points in OIB	Number of Items upplemented
Algebra I	63	3
Biology	69	3
Literature	54	2

Details of all other materials can be found in the *Keystone Standard Setting Technical Report* (Pennsylvania Department of Education, 2011).

DATA PREPARATION

In Bookmark standard setting (Lewis et al., 1996), the locations of items are typically rescaled to produce better alignment with the task of asking panelists what a student should know and be able to do. A probability of 0.67 is often used to find the corresponding item location during rescaling because this probability aligns better with the likelihood panelists use to make their judgment on whether a borderline student should answer the item correctly or receive a score point or higher. For Keystone Exams, the multiple-choice (MC) items were calibrated using the familiar form of the dichotomous Rasch model. The CR items were calibrated using another model in the Rasch family, Master's partial-credit model (Wright & Masters, 1982). The latter model parameterizes each threshold needed to obtain the maximum score on the task. Consequently, there is one item difficulty parameter for each of the $n - 1$ score transitions (0/1, 1/2, etc.), or thresholds. Using the equated item parameters, the locations of items were rescaled to a response probability of 0.67 (i.e., RP=0.67). For MC items, the item locations were found by solving

$$\Phi_{ni} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

for the value of β_n that gives Φ_{ni} = 0 .67. Φ_{ni} = is the probability that person *n* scores 1 on item *i*; β_n is the ability of person *n*; and δ_i is the difficulty of item *i*.

For CR items, the probability of person *n* scoring *x* on item *i* is

$$\pi_{nix} = \frac{\exp\sum_{j=0}^{x}(\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i}\exp\sum_{j=0}^{x}(\beta_n - \delta_{ij})}, \quad x = 0, 1, \ldots, mi$$

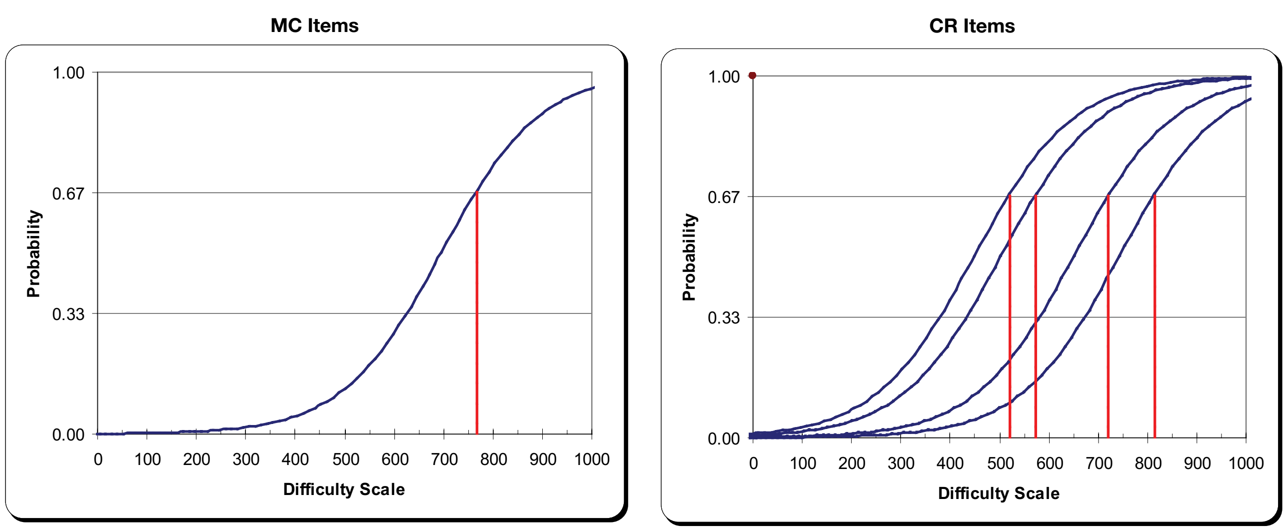
where m_i is the number of thresholds and, for notational convenience,

$$\exp\sum_{j=0}^{0}(\beta_n - \delta_j) = 1.$$

This equation expresses the probability of person *n* scoring *x* on the m_i threshold of item *i* as a function of the person's measure (β_n) and the threshold difficulties (δ_{ij}) of the m_i thresholds for item *i*. The observation *x* is a count of the successfully completed item thresholds. The item location for a score point is determined by finding the β_n for the person who has a 0.67 probability of earning this score point or higher.

The figure below shows how the difficulty values of MC items and score values for the CR items were treated in determining their respective OIB placements. For an MC item (left plot), the difficulty is the point on the scale at which the examinees have a 0.67 probability of answering the item correctly. For the CR item (right plot), the four illustrated values (e.g., on a 0–1000 scale) indicate where the examinees have a 0.67 probability of earning a particular score point or higher. The item difficulty for the MC item is 768, and the four threshold values for the CR items are 521, 575, 723, and 815. The value of 521 is the location on the scale where examinees have a 0.67 probability of earning a score of 1 or higher (i.e., 2, 3, or 4). The value of 575 is where examinees have a 0.67 probability of earning a score of 2 or higher (i.e., 3 or 4). The value of 723 is where examinees have a 0.67 probability of earning a score of 3 or 4.

Figure 13–2. Example of Obtaining Item Difficulties for MC and CR Items



TRAINING

The overall training was conducted the first morning of the workshop. Participants were informed that they were to

- be responsible for all secure materials,
- verify their individual placements for each round of judgments, and
- participate in a discussion as a large group.

Content-specific training was conducted after content area groups assembled in different rooms. These training materials included the following:

- Item Map
- Item Separation Map
- OIBs
- Training Rubrics and/or Passages
- PLDs
- Rating Form

Panelists were told that the process includes iterations (rounds) of individual judgments, group discussions, and opportunities to revise judgments. In addition, impacts were presented (percentage of students in each performance level) based on the large groups' results and external data.

BOOKMARK PROCEDURE

DRC utilized a Bookmark method to set the performance standards. Bookmark is one in a broad category of methods commonly referred to as item mapping that focus on items rather than examinees. To begin the process, participants were asked to visualize the knowledge and skills of a student who is at the borderline between two performance levels based on the PLDs. Thereafter, participants were given an ordered item booklet (with items ordered from easiest to most difficult) and asked to assess whether this borderline student has a reasonably high probability of answering each item correctly. “Reasonably high” was defined as 0.67. In addition, an item map was presented that contained the response key, the content objective, and the item sequence in the test booklets. An item separation map was also presented that showed the relative difficulty of each item. Panelists were given a rating form to record their individual placements for all performance levels in each round. Before each round, panelists were asked to fill out a readiness form in order to proceed.

Round 1. The Bookmark procedure proceeded in three rounds. Round 1 began following the review and discussion of PLDs facilitated by a DRC test development specialist. Participants then reviewed the OIBs independently. During this review, they were asked to determine what academic knowledge, skills, and competencies were required for a barely Proficient, Basic, or Advanced student to respond correctly to each successively more difficult item.

Training by the overall psychometric lead during the bookmark placement session emphasized the following points:

- The bookmark represents a judgment of the divide between items that a student at the borderline of a performance level should master and those that are not necessary to master.
- Bookmark placement should not be thought of as separating two items but rather two groups of items. In other words, a placement should not hinge on distinctions drawn for adjacent items without some compelling reason, such as a large gap in content difficulty.
- Students with a scaled score at a given cut score should have approximately a 0.67 probability of correctly responding to a MC item or receiving a certain score point and higher for a CR item at the cut score. These same students should have a higher probability of success on easier items (before the bookmark placement) and a lower probability of success on harder items (after the bookmark placement).

- While placing their bookmarks, panelists should consider what students should know and be able to do in the context of the skills implied by the PLDs and the item content.
- Panelists could start with placing the Basic/Proficient cut point, next the Below Basic/Basic cut point, and finally the Proficient/Advanced cut point.

Panelists were asked to record their bookmark placements on the rating form after they filled out a readiness form, which indicated they had completed the training and understood the standard setting process and their roles. Panelists' judgments were entered into a spreadsheet program. The median ratings of all panelists were calculated. The median placements were treated as the recommended cut scores. In addition, the standard errors associated with the recommended bookmark placements were calculated and associated impact data were determined.

Round 2. Round 2 started with a discussion of Round 1 results. The individual panelists' Round 1 bookmark placements, the median bookmark placements, and the percentage of students in each performance level were presented. Panelists were instructed to verify the ratings entered into the program as correct. A large-group discussion followed. The panelists compared their results with others by considering questions such as why they made their Round 1 placements at the locations where they did and what skills and knowledge were required to answer the items. After that, the impact data, based on the median bookmark placement from Round 1 (using the Spring 2011 operational test score distributions), were provided to help panelists frame the effects of their judgments. During Round 2 discussion, there was no attempt by the facilitators to reach consensus.

After Round 2 discussion, panelists were asked to make a second set of bookmark placements. Before they revised their Round 1 placements, they were asked to fill out the readiness form to make sure they understood how to adjust their placements (if they desired to do so) based on Round 1 information. The judgments were entered into the spreadsheet program to calculate the median cut scores for each table and the full panel. The associated impact data were also calculated.

Round 3. Round 3 began with a discussion of Round 2 results. The process followed in Round 2 was used. More specifically, the individual panelist's Round 2 bookmark placements, the median bookmark placements from Round 2, and the percentage of students in each performance level were presented. Panelists were instructed to verify the ratings entered into the program as correct. A table discussion followed. Panelists compared their results with others by considering questions such as why they made their Round 2 placements at the locations where they did and what skills and knowledge were required to answer the questions. The impact data, based on the median bookmark placement from Round 2, were provided to help panelists frame the effects of their judgments.

The Keystone Exams are one component of Pennsylvania's new system of high school graduation requirements. Because of the high-stakes consequences, the TAC strongly recommended bringing in external impact data to provide panelists with a reference outside of the Keystone Exams. The intent was to achieve reasonableness of results rather than to use the external data in a directive manner. DRC investigated Pennsylvania students' performance on the Pennsylvania System of School Assessment (PSSA), National Assessment of Educational Progress (NAEP), and Student Achievement Test (SAT) and presented external data as shown in Figures 13–3A to 13–3C before panelists made their Round 3 judgments. The panelists were informed of the following points:

- The PSSA and NAEP results were based on students' performance in 2009. The PSSA results were from grades 6–8 and 11. The NAEP results were from grade 8.
- All students in grades 6–8 and 11 in Pennsylvania took the PSSA. A sample that represents the Pennsylvania grade 8 students took the NAEP tests.
- The SAT results were based on the performance of students who took the SAT in 2010 or prior years.
- About 99% of students in the 2010 SAT data file indicated their expected graduation dates were in 2010; most of these students were in grade 11 in 2009. Therefore, the 2010 SAT data and the 2009 PSSA data were matched.
- Based on the matched sample, it was found that students with higher PSSA scores were more likely to take the SAT. To represent the full population in terms of demographics and PSSA scores, the matched sample was weighted by students' demographics and PSSA scores when calculating the impacts.

Figure 13–3A. External Impact Data: Algebra I

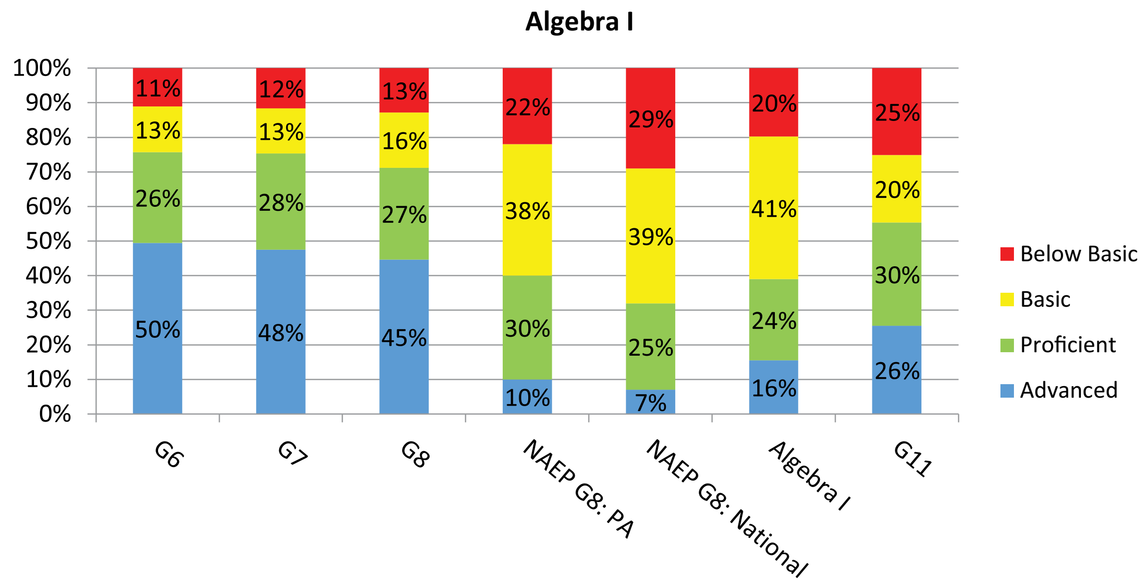


Table 13–4A. Number of Score Points in OIB and Number of Items Supplemented: Algebra I

Performance Level	PSSA G6	PSSA G7	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Alg. I	PSSA G11	College Ready Yes Projected	SAT: College Ready Yes: PA	SAT: College Ready Yes: National
Below Basic	11.1%	11.6%	12.8%	22.0%	29.0%	19.8%	25.1%	0.9%	51.7%	54.0%
Basic	13.2%	13.1%	16.0%	38.0%	39.0%	41.2%	19.5%	7.8%	51.7%	54.0%
Proficient	26.2%	27.8%	26.6%	30.0%	25.0%	23.5%	29.8%	42.2%	51.7%	54.0%
Advanced	49.5%	47.5%	44.7%	10.0%	7.0%	15.5%	25.5%	91.2%	51.7%	54.0%
Below Basic + Basic	24.3%	24.7%	28.8%	60.0%	68.0%	61.0%	44.8%	4.0%	51.7%	54.0%
Proficient + Advanced	**75.7%**	**75.3%**	**71.3%**	**40.0%**	**32.0%**	**39.0%**	**55.3%**	**64.9%**	**51.7%**	**54.0%**
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	38.1%	51.7%	54.0%
Total N	128,421	132,803	135,909	3,600	161,700	93,703	135,676	61,118	65,426	N/A

Figure 13–3B. External Impact Data: Biology

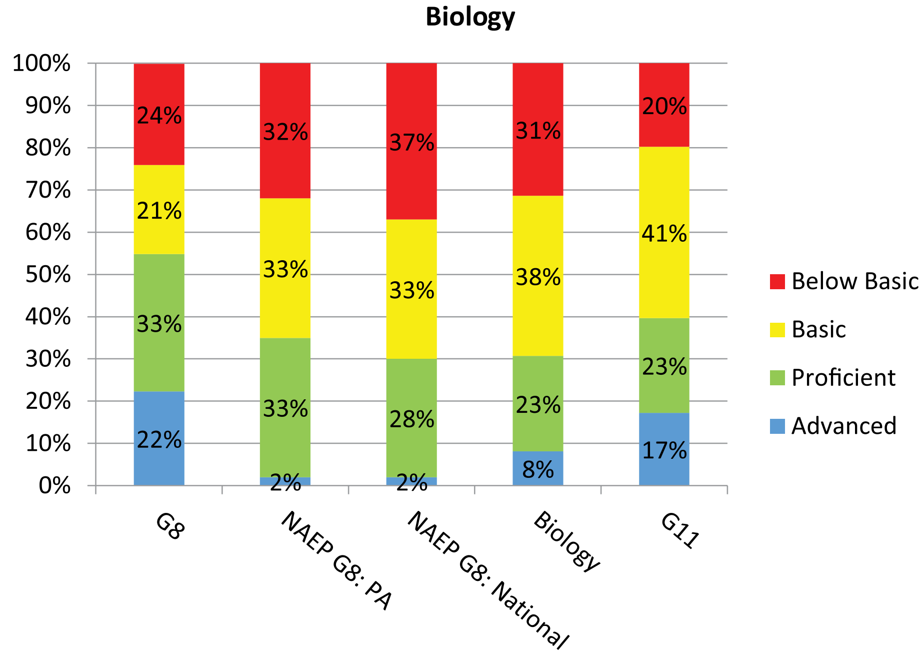


Table 13–4B. Number of Score Points in OIB and Number of Items Supplemented: Biology

Performance Level	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Biology	PSSA G11	College Ready Yes Projected	SAT Math: College Ready Yes: PA	SAT Math: College Ready Yes: National
Below Basic	24.0%	32.0%	37.0%	31.4%	19.8%	1.4%	38.7%	43.0%
Basic	21.1%	33.0%	33.0%	37.9%	40.5%	9.1%	38.7%	43.0%
Proficient	32.5%	33.0%	28.0%	22.6%	22.5%	45.9%	38.7%	43.0%
Advanced	22.3%	2.0%	2.0%	8.1%	17.2%	86.6%	38.7%	43.0%
Below Basic + Basic	45.1%	65.0%	70.0%	69.3%	60.3%	6.6%	38.7%	43.0%
Proficient + Advanced	**54.8%**	**35.0%**	**30.0%**	**30.7%**	**39.7%**	**63.8%**	**38.7%**	**43.0%**
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	29.4%	38.7%	43.0%
Total N	134,969	3,600	151,100	46,394	131,534	60,311	65,426	N/A

Figure 13–3L. External Impact Data: Literature

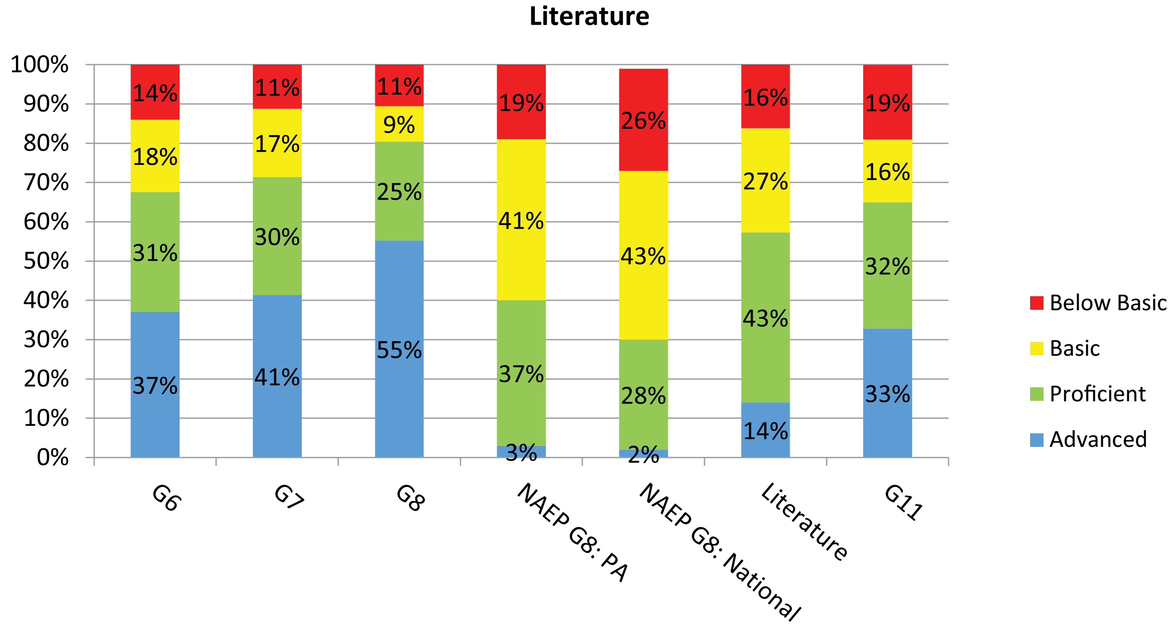


Table 13–4L. Number of Score Points in OIB and Number of Items Supplemented: Literature

Performance Level	PSSA G6	PSSA G7	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Literature	PSSA G11	College Ready Yes: Projected	SAT Math: College Ready Yes: PA	SAT Math: College Ready Yes: National
Below Basic	14.0%	11.2%	10.6%	19.0%	26.0%	16.2%	19.0%	1.2%	46.4%	50.0%
Basic	18.4%	17.4%	8.9%	41.0%	43.0%	26.5%	16.0%	4.7%	46.4%	50.0%
Proficient	30.5%	30.0%	25.2%	37.0%	28.0%	43.3%	32.1%	24.8%	46.4%	50.0%
Advanced	37.1%	41.4%	55.3%	3.0%	2.0%	14.0%	32.9%	76.7%	46.4%	50.0%
Below Basic + Basic	32.4%	28.6%	19.5%	60.0%	69.0%	42.7%	35.1%	2.7%	46.4%	50.0%
Proficient + Advanced	**67.6%**	**71.4%**	**80.5%**	**40.0%**	**30.0%**	**57.3%**	**64.9%**	**51.1%**	**46.4%**	**50.0%**
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	34.5%	46.4%	50.0%
Total N	128,284	132,641	135,739	3,500	160,900	42,292	135,470	61,081	65,426	N/A

The Keystone Exams and PSSA results were presented to the panelists first. Panelists were encouraged to compare the impact data and discuss whether the results for the Keystone Exams were reasonable. The NAEP results were added next, and the SAT results were introduced last for comparison and discussion. While discussing the external data, panelists were reminded that all these tests were created for different purposes and might cover different content standards.

Before panelists provided their final judgments, they were instructed to fill out the readiness form to make sure they understood how to adjust their placements (if they desired to do so) based on the Round 2 information and external impact data. After their individual bookmark placements, panelists filled out the evaluation form. The judgments were entered into the spreadsheet program to calculate the median placements for the full panel. The associated impact data were also calculated. The Round 3 results were presented to the panelists for their information after the lunch break.

PANELISTS' RECOMMENDATIONS

Table 13–4 provides a summary of each round's median, minimum, and maximum ratings (i.e., bookmark page numbers) of the group.

Table 13–4. Summary of Panelists' Ratings for Each Round

Exam	Round	Bookmark Page Number Median	Bookmark Page Number Min.	Bookmark Page Number Max.	Basic/ Proficient Median	Basic/ Proficient Min.	Basic/ Proficient Max.	Proficient/ Advanced Median	Proficient/ Advanced Min.	Proficient/ Advanced Max.
Alg. I	1	11	6	19	28	14	42	45	33	56
Alg. I	2	11	6	13	26	17	33	42	40	46
Alg. I	3	11	10	12	26	18	30	46	41	46
Bio.	1	9	4	15	26	20	30	56	43	62
Bio.	2	8	7	14	24	21	30	54	50	60
Bio.	3	8	7	12	22	20	30	54	50	60
Lit.	1	8	5	14	27	12	34	47	38	52
Lit.	2	9	8	15	23	15	34	46	38	48
Lit.	3	9	8	15	25	17	34	48	38	48

CUT POINTS AND STANDARD ERRORS

Each bookmark page number is associated with a bookmark difficulty (i.e., logit value). The logit cut is the bookmark difficulty corresponding to the median OIB page number minus one. The logit cut and the standard error (SE) of median logit based on panelists' Round 1 rating were used to establish the 1 and 2 SE confidence intervals. By bracketing the median cut score by 2 SEs, the 95% confidence interval was identified; the confidence interval can be used to estimate the effects of false positives (passing students who may not actually have sufficient knowledge and skills) or false negatives (failing students who do have sufficient knowledge and skills). PDE can use these standard errors to identify the appropriate cut score by taking into consideration the variance in the human judgments. Table 13–5 summarizes the logit cuts associated with Round 3 median ratings, median +/-1 SE, and median +/-2 SE. The corresponding impacts (percentages in performance level) are provided in this table as well. Note that BB represents Below Basic; B represents Basic; P represents Proficient; and A represents Advanced.

Table 13–5. Summary of Logit Cuts and Impacts

Exam	Stats	Logit Cut BB/B	Logit Cut B/P	Logit Cut P/A	Percentage in Performance Level (%) BB	Percentage in Performance Level (%) B	Percentage in Performance Level (%) P	Percentage in Performance Level (%) A	Percentage in Performance Level (%) P+A
Alg. 1	Median-2SE	-0.7273	0.4291	1.3694	14.7	37.3	30.2	17.8	48.0
Alg. 1	Median-1SE	-0.6181	0.5659	1.5041	17.2	37.8	31.6	13.4	45.0
Alg. 1	Median	-0.5090	0.7027	1.6388	19.8	41.2	27.5	11.5	39.0
Alg. 1	Median+1SE	-0.3999	0.8395	1.7735	22.4	44.4	23.5	9.7	33.2
Alg. 1	Median+2SE	-0.2907	0.9763	1.9082	25.2	47.0	19.7	8.1	27.8
Bio.	Median-2SE	-0.5977	0.3098	1.2500	22.5	38.9	27.9	10.7	38.6
Bio.	Median-1SE	-0.4933	0.3564	1.3205	28.3	33.1	29.3	9.3	38.6
Bio.	Median	-0.3888	0.4029	1.3910	31.4	32.7	27.8	8.1	35.9
Bio.	Median+1SE	-0.2843	0.4494	1.4615	34.5	32.2	25.2	8.1	33.3
Bio.	Median+2SE	-0.1799	0.4960	1.5320	40.7	26.0	26.4	6.9	33.3
Lit.	Median-2SE	-0.6561	0.2338	1.7014	12.4	20.9	46.3	20.4	66.7
Lit.	Median-1SE	-0.5545	0.4116	1.9551	14.2	25.3	46.5	14.0	60.5
Lit.	Median	-0.4530	0.5894	2.2088	16.2	26.5	48.4	8.9	57.3
Lit.	Median+1SE	-0.3515	0.7672	2.4625	18.3	31.4	43.6	6.7	50.3
Lit.	Median+2SE	-0.2499	0.9450	2.7162	20.5	36.5	38.2	4.8	43.0

FINAL RESULTS

After reviewing the results in Table 13–5 and considering panelists' discussions at the standard setting workshop, PDE recommended using the logit cut scores associated with the median of panelists' Round 3 ratings for Algebra I and Biology. For Literature, PDE recommended the logits cuts associated with the Round 3 median plus 1 SE.

To avoid negative values on the logit scale, the scaling constants were determined next to linearly convert the logit values to scaled scores. The scaled cut scores for each performance level were obtained by linearly transforming the logit cuts. Details of the scaling process can be found in Chapter Fourteen. A brief description is below.

For Keystone Exams, the linear transformation from logits or Rasch measures to scaled scores was established by anchoring the logit cut for Basic/Proficient to a scaled score 1500 and fixing the slope constant to 50. The intercept constant was calculated next based on the known values 1500, 50, and the logits cut for Basic/Proficient for each content area. In addition, the bottom of the scale was truncated at the lowest obtainable scaled score (LOSS), 1200. The top of the scaled scores was truncated at the highest obtainable scaled score (HOSS), 1800. The recommended scaled score cuts and the corresponding impacts were provided to the Board on July 20, 2011, for approval. Table 13–6 presents the final scaling constants and the Board-approved scaled-score ranges for each performance level.

Table 13–6. Summary of Scaled-Score Ranges and Scaling Constants

Exam	Performance Level Below Basic	Performance Level Basic	Scaling Constants Proficient	Scaling Constants Advanced	Scaling Constants Slope	Scaling Constants Intercept
Algebra I	1200–1438	1439–1499	1500–1545	1546–1800	50	1464.365
Biology	1200–1459	1460–1499	1500–1548	1549–1800	50	1479.355
Literature	1200–1443	1444–1499	1500–1583	1584–1800	50	1461.140

The Keystone Exams are reported by total and modules. Although the panelists made recommendations based on the total test only, the Basic/Proficient cut for the total test is applied directly in setting the passing cut score for each module. In this case, the passing scaled score cut at module level is 1500.

CHAPTER FOURTEEN: SCALING

Scaling is used to transform test score values (i.e., raw scores) onto a scale that can be interpreted by users easily and correctly. Raw scores cannot be used to compare students' achievement across administrations because they depend on the difficulty of the tests. The same student can score higher on an easy test than on a difficult test. To overcome the limitation of raw scores, the scaled scores are introduced to report students' achievement in Algebra I, Biology, and Literature. This chapter describes the two major steps to convert a raw score to a scaled score (SS) and some key considerations for establishing the score scale for Keystone Exams.

RAW SCORES TO RASCH ABILITY ESTIMATES

The pre-equated item parameter estimates for the operational items (further discussed in Chapters Twelve and Fifteen) were used to obtain Rasch person ability estimates and asymptotic standard errors of measurement for each possible raw score value for the overall test, as well as each module. The generation of this raw score-to-Rasch ability was accomplished through application of the fundamental formulas in the Rasch measurement model. The combination of both dichotomously scored multiple-choice (MC) items as well as polytomously scored constructed-response (CR) items requires the use of a partial-credit model (RPCM) (Wright & Masters, 1982). The Newton-Raphson iterative procedure is used to obtain precise ability estimates:

$$b_r^{(t+1)} = b_r^t - \frac{r - \sum_i^L \sum_{k=1}^m kP_{rik}^{(t)}}{-\sum_i^L \left[\sum_{k=1}^m k^2 P_{rik}^{(t)} - \left(\sum_{k=1}^m kP_{rik}^{(t)} \right)^2 \right]} \quad r = 1, \ldots, M - 1,$$

where b_r^t is the estimated ability of the student with score r after t iterations, k is the number of thresholds, L is the number of items, $M = \sum_i^L m_i$, and $P_{rik}^{(t)}$ is the probability, π_{nix}, defined earlier in Chapter Twelve:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^{x} (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{x} (\beta_n - \delta_{ij})} \quad , \quad x = 0, 1, \ldots, m_i.$$

The asymptotic standard error of measurement (SEM) was estimated from the denominator of the final iteration:

$$SE(b_r) = \left[\sum_i^L \left[\sum_{k=1}^m k^2 P_{rik}^{(t)} - \left(\sum_{k=1}^m kP_{rik}^{(t)} \right)^2 \right] \right]^{-1/2} .$$

The Rasch ability estimates and the corresponding SEMs are then transformed to scaled scores and SEMs of scaled scores as discussed in the following section.

ZERO AND PERFECT SCORES

A direct ability estimate for zero (no points earned) or perfect (all points earned) raw scores can't be achieved. Thus, a default procedure for estimating such extreme scores was used for the Keystone Exams. Essentially, a fractional raw score (a value less than one, e.g., 0.3) was added to zero scores and subtracted from perfect scores to determine the corresponding logit values for these extreme scores.

RASCH ABILITY ESTIMATES TO SCALED SCORES

Generally, scaled scores are preferred over Rasch ability estimates for reporting purposes. One issue is that Rasch ability estimates are on a scale that includes negative and decimal values. By transforming the Rasch ability estimates to scaled scores, all reported values can become positive integers, which makes more sense to parents and students. Since Rasch ability estimates are comparative after equating (discussed further in Chapter Fifteen) to the base administration/year, the transformed scaled scores have a common scale across administrations, even though the corresponding raw scores may differ.

Scaled scores are usually obtained through some linear transformation of Rasch ability estimates. Before the linear equation is established for each content area, a few points were considered for the Keystone Exams:

- Avoid scales that might be confused with scores for other types of assessment, for example:
    - Scaled scores ranging from 0 to 100 (because this might be confused with percentage correct scores or percentile ranks)
    - Scaled scores ranging from 200 to 800 (because this might be confused with SAT scores)
    - Scaled scores with similar ranges as the ones for the Pennsylvania System of School Assessment (PSSA) or Classroom Diagnostic Tools (CDT)
- Avoid scales similar to raw scores from a base form.
- Avoid scales that might suggest the scores are more precise than they actually are (i.e., suggesting more precision than can actually be supported by the test scores).
- Avoid scales with negative numbers and decimals.

In terms of industry standard practice, a common perspective is that scaled scores should facilitate score interpretation while at the same time minimize misinterpretation and unwarranted inferences. Often this is done by incorporating some kind of meaning to the scores[1] (Peterson, Kolen, and Hoover, 1989). The incorporation of content meaning is one way to facilitate score interpretation. This might be done in several different ways. For example, the current PSSA scaled scores, like those of many other state assessments, try to input some content meaning by having the PSSA performance level cut scores have known values on the scaled-score metric. Such an approach appears to make good sense given the purposes of a criterion-referenced test like the PSSA.

As a result, a scaled-score range of 1200 to 1800 and the Proficient scaled score cut, 1500, for all content areas were selected as the start point to establish the scales for the Keystone Exams in Algebra I, Biology, and Literature. It is worth noting that, although careful considerations were given to the selection of these values, they are completely arbitrary. For example, the label of 1500 could have been called 100 or any other value or letter without affecting any of the relationships among schools, administrations, students, or items. In other words, changing the scale would simply be changing the labels on the axis of a graph without moving any of the points.

[1] Not everyone agrees with this sentiment. Some have argued the opposite point, that is, any attempt to add meaning to test scores actually predisposes the scores to be misinterpreted (Angoff, 1984).

LINEAR TRANSFORMATION FORMULAS

The scaled scores for the Keystone Exams are obtained through a linear transformation of the Rasch ability estimates ($\hat{\beta}$). Specifically,

$SS = m\,\hat{\beta} + b,$

where *m* is the slope and *b* is the intercept. The linear transformation for the Keystone Exams is derived by anchoring the Proficient cut (i.e., Rasch ability estimate) recommended by the panelists at the standard- setting workshop to the scaled score 1499.5 (i.e., 1500 after rounding), and then set the slope of the line. There could be many lines with different slopes going through the anchor point. However, the slope of the line has influence over the variability of the scaled scores. For Keystone Exams, the slope of 50 was chosen because it results in desired scaled score standard deviation. Once the scaled score, slope, and Rasch ability estimate are determined, the intercept *b* can be derived by the equation above. The final slopes and intercepts for deriving scaled scores for the Keystone Exams are provided in Table 14–1.

Table 14–1. Scaling Constants by Content Area

Exam	Scaling Constants Slope	Scaling Constants Intercept
Algebra I	50	1,464.365
Biology	50	1,479.355
Literature	50	1,461.140

ROUNDING

The linearly transformed scaled scores are always rounded to the nearest integer value for reporting purposes. Values greater than or equal to 0.50 are rounded up. Values less than 0.50 are rounded down.

LOWEST OBTAINABLE SCALED SCORES

The Keystone Exams in Algebra I, Biology, and Literature have a lowest obtainable scaled score (LOSS) of 1200. Any derived scaled score less than 1200 is truncated to this minimum value. The selection of a LOSS is mainly based on two considerations: 1) extreme low scaled scores may have an impact on the average of the scaled scores at school/district level and 2) score truncation makes sense from a score precision perspective given measurement errors at the extremes are large. The LOSS value 1200 is established by giving consideration to *chance* performance over the MC items (e.g., if 40 four-option MCs were on a test, approximately 10 points might be earned on guessing alone) and considering the percentage of students who would be awarded the LOSS values.

HIGHEST OBTAINABLE SCALED SCORES

A highest obtainable scale score (HOSS), 1800, is set for the Keystone Exams for the same reasons described for the LOSS value. However, unlike the LOSS value, which is set initially by giving consideration to guessing over MC items, it is somewhat more difficult to determine what rules should be applied to establish the HOSS. Based on the empirical results, the value 1800 corresponds to a logit value (or Rasch ability estimate) that ranged from 6 to 7, and 0 percent of students received this score.

RAW-TO-SCALED-SCORE TABLES

The final raw-to-scaled-score conversion tables can be found in Appendix K. Note that only the raw-to-scaled-score tables for each single administration were reported. In other words, these tables cannot be used to look for a student's best scaled score if it is combined from two different administrations. The conditional standard error of measurement (CSEM, see Chapter Eighteen for detailed discussion) and corresponding 1 CSEM confidence intervals are also provided in these tables.

CHAPTER FIFTEEN: EQUATING

Equating is a statistical process that is used to adjust scores on test forms so that scores on the forms can be used interchangeably (Kolen & Brennan, 2004), even though the test forms consist of different items. In large-scale testing programs, it is a common practice to have different item sets appear in different test forms across administrations. Students' raw scores (or number-correct scores) cannot be compared between forms or administrations because they depend on the difficulty of the items in a form. The same student can score higher on an easy test than on a difficult test.

To make meaningful comparisons of test scores across administrations, various equating models and procedures have been developed in the literature. For example, in terms of design, there are randomly equivalent groups design and common-item non-equivalent groups design. In terms of testing model, it can be classified as either classical test theory-based equating model or modern test theory-based (e.g., Rasch model or item response theory) equating model. In terms of when the equating is conducted in the assessment cycle, it can be classified as pre-equating or post-equating. The following sections will focus on the equating design and analyses for the winter, spring, and summer Keystone Exams administered in 2015–2016.

PRE- VS. POST-EQUATING

As with other Pennsylvania assessment programs, the Rasch model is used to guide the test design, form construction, calibration, scaling, and equating of the Keystone Exams. The key element of equating test forms using the Rasch model is to place the item parameters from different administrations on the same scale. This is also referred to as item equating. Once the item parameters from different operational test forms are on the same scale, the Newton Raphson procedure can be used to convert number-correct scores to scaled scores as described in Chapter Fourteen. As a result, the scaled scores can be compared across forms with different items.

A common practice in many K–12 large-scale assessment programs is to have all the items field tested before they go operational. Once the field test items' difficulties are placed on the base scale or common metric, in theory, one should not expect the Rasch item difficulties for these items to change, except within a reasonable range of measurement error, after they are administered in an operational test providing the Rasch model fits the data. Based on this theoretical advantage of using Rasch models, equating can be conducted using the item parameters calibrated from field test data. This statistical procedure is referred to as pre-equating. In contrast, post-equating involves the use of Rasch item difficulties calibrated from the data of the operational test to be equated.

Although, in theory, the two equating procedures should provide identical results when the model fits the data, each has its own advantages and disadvantages. The use of pre-equating can facilitate the operational process in terms of rapid score reporting, more time for quality control, and more flexibility in the assessment. One successful application of pre-equating is for computer-adaptive tests where test questions are tailored to the student's achievement as the test progresses. This allows for providing scores immediately after students finish the test. However, a variety of issues need to be considered when using pre-equating in practice. For example, students may not be motivated to take the field tests, especially standalone field tests, which may make the items appear harder in the field test than in the operational test (Eignor, 1985; Eignor & Stocking, 1986; Stocking & Eignor, 1986; Kolen & Harris, 1990). Other concerns for the field test items include item context, item position, and sample size. In contrast, the use of post-equating, when applicable, does not have the same motivational concerns as with pre-equating. Also, post-equating uses post-administration data and is sometimes considered to yield more accurate analysis results, given that the number of students who take the operational tests is usually large. On the other hand, when the reporting window is extremely tight, as is the case with some graduation or end-of-course exams in various states, post-equating has to occur within a very short time, and hence it leaves less time for the equating analyses and quality control.

EQUATING DESIGN FOR KEYSTONE EXAMS

The Keystone Exams, like many other graduation or end-of-course exams, require a quick turnaround of testing results. After the exams are administered, the bulk of the time is consumed by various data-processing steps. As a result, the equating analyses must be produced in a short period of time, which puts the quality of final analysis results under great risk. In addition, the requirement that a student's final score is the combination of the two highest module scores from any operational test (see Chapter 16 for details) increases the complexity of equating analyses and score reporting for future administrations. To control the quality of post-administration processing and guarantee the accuracy of students' reports, pre-equating, one of the most promising applications of Rasch model or item response theory (see Lord, 1980, chap. 13), was proposed and implemented for the Keystone Exams.

To implement the pre-equating model in the Keystone Exams, more efforts have been made to enhance the accuracy of pre-equating results based on the findings from the literature. For example, to address the concerns regarding students' motivation to take field tests, it was decided that no Rasch item difficulty estimates from stand-alone field tests can be used to pre-equate test forms. Instead, all the field test items have to be embedded in an operational test before their Rasch item difficulty estimates can be used. This is based on the assumption that students should be equally motivated to take the operational and embedded field test items, especially when they are not aware of which item is a field test item. To minimize item context and item position effects (i.e., lack of motivation and fatigue), field test items were interspersed within the operational sections. With this design, students have a lesser chance of knowing the field test item positions. Fatigue effects due to field test items being placed in the last section of the operational test can be mitigated in this design as well.

To improve the accuracy of the Rasch item difficulties estimated from the field test data but used as the values for the operational items, Data Recognition Cooperation (DRC) scored as many students' responses to the field test items as possible, given that increasing sample size can increase the estimation accuracy. More specifically, DRC scored all students' responses to the multiple-choice (MC) items and approximately 2,000 students' responses to the constructed-response (CR) items.

POST-EQUATING CHECK ANALYSES

Although extra care has been taken to guarantee the success of pre-equating during the test design, form construction, and calibration of embedded field test items, is the pre-equated result (e.g., raw-to-scaled-score table) still valid given the sample change and item sequence change from the field test positions to operational test positions?

After the operational testing data was collected for the winter, spring, and summer 2016–2017 administrations, post-equating check analyses (with the exception of Literature administered in summer) were conducted to validate the raw-to-scaled-score tables generated using the pre-equated item parameter estimates. The post-equating check analysis conducted at item level evaluated the item difficulty estimate stability. The analysis conducted at form level investigated whether or not the raw-to-scaled-score tables had changed significantly.

ANALYSES AT ITEM LEVEL

To conduct the evaluation of item difficulty parameter estimate stability, the operational items were calibrated using WINSTEPS. There were two approaches used in running WINSTEPS. The first one was that the item parameters for all operational items were anchored to the bank values (also referred to as old values). WINSTEPS provided the displacements between the anchor values and the values that would have been estimated from the current data. The items with displacement value of 0.5 or larger were further investigated as outliers. The second approach was to calibrate the item parameters freely in WINSTEPS. The newly calibrated values (referred to as new values) were equated to the bank scale by adjusting the new item parameter estimates by the difference of the means between the old bank values and the new values. These adjusted values are referred to as equated values. Tables L–1 to L–9 in Appendix L present the item sequence change; *n*-count; old, new, and equated item difficulty estimates (i.e., logit); the corresponding standard error of measurement (SEM); and displacement. A scatter plot of the old and equated values was plotted to check for outlier items. Outliers were identified as those items where the perpendicular distance to the line was greater than or equal to 1.96 standard deviations (see Figures 15–1 to 15–9). As can be seen from Appendix L and the figures below, most of the items had stable item difficulty estimates; most outliers were flagged consistently by both the scatter plot and displacement. Table 15–1 summarizes the outliers flagged by both criteria. These items were reviewed by DRC content specialists, but no obvious reasons were found to explain the item difficulty change.

Table 15–1. Summary of Items Flagged by both the Scatter Plot and Displacement

Administration	Content Area	Item IDs
Winter	Algebra I	674514, 641481
Winter	Biology	703537
Winter	Literature	641567
Spring	Algebra I	NA
Spring	Biology	674080
Spring	Literature	734611
Summer	Algebra I	702548
Summer	Biology	739695
Summer	Literature	683424

Figure 15–1. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Winter

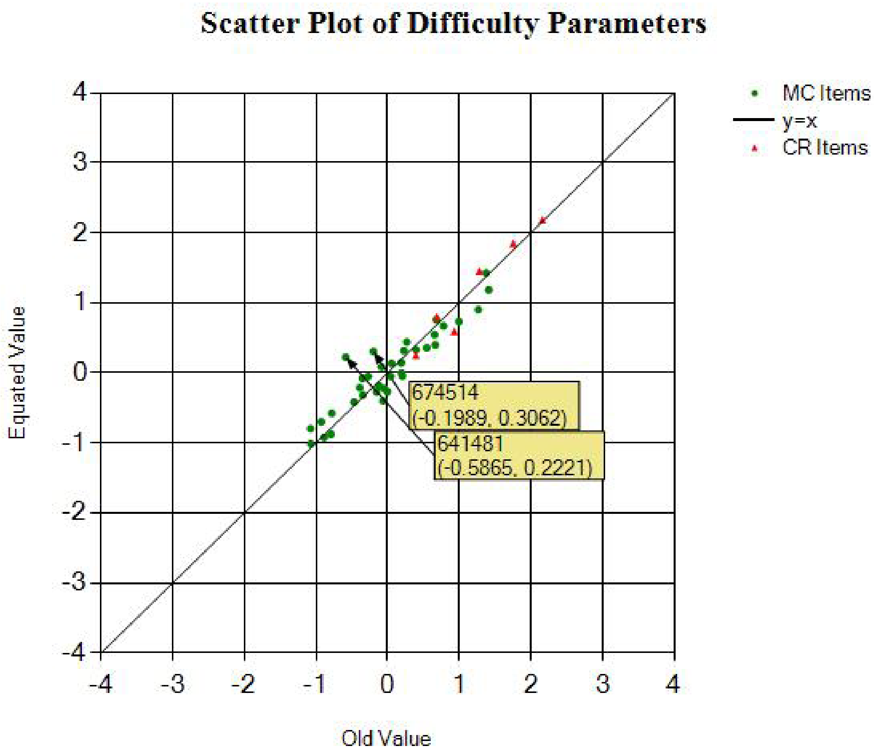


Figure 15–2. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Winter

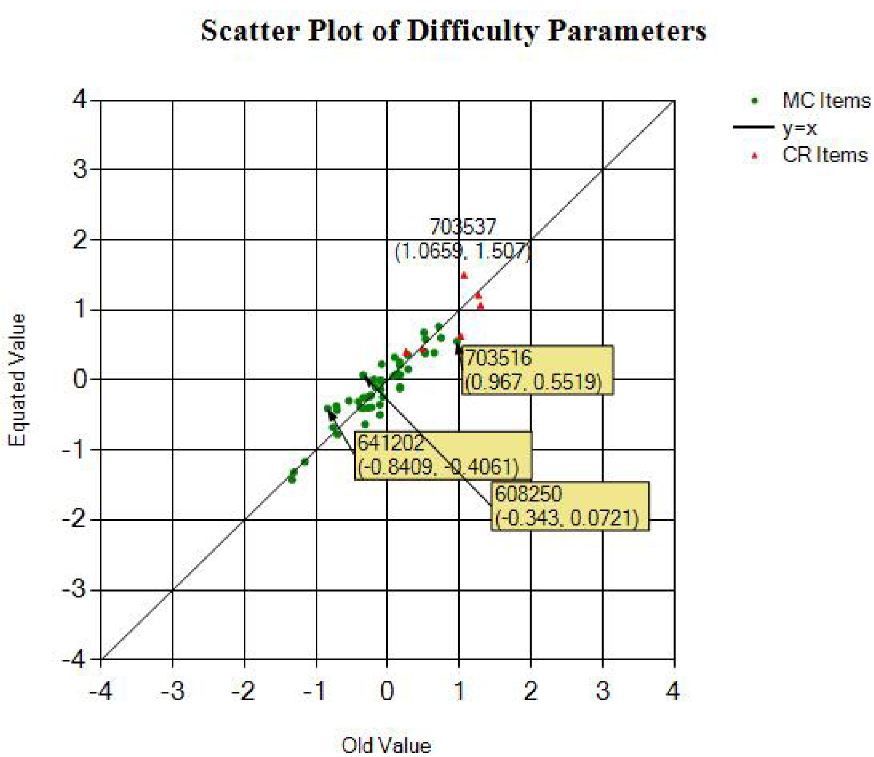


Figure 15–3. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Winter

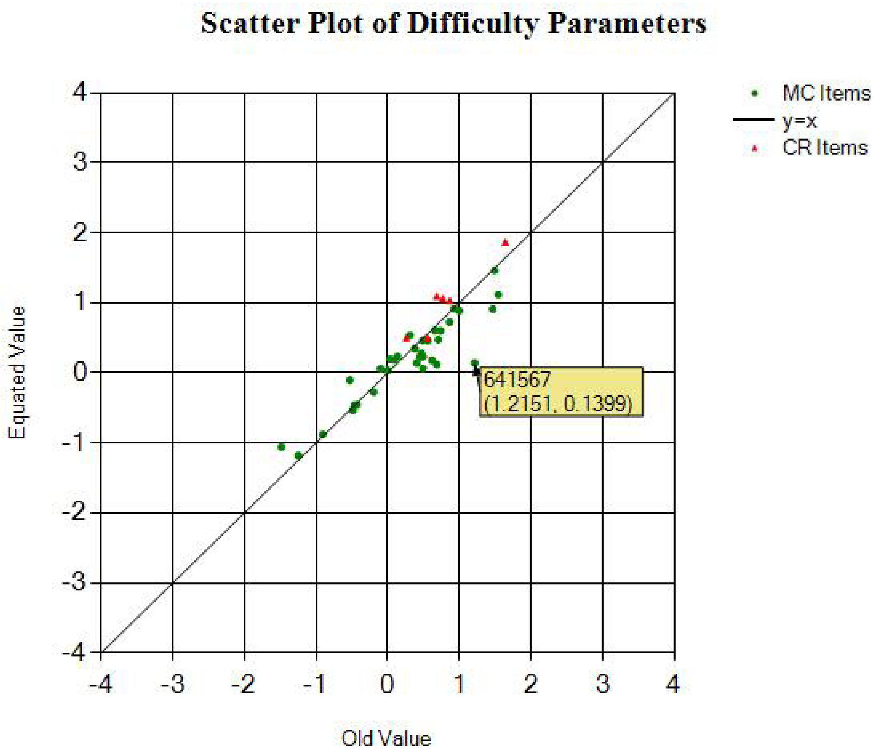


Figure 15–4. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Spring

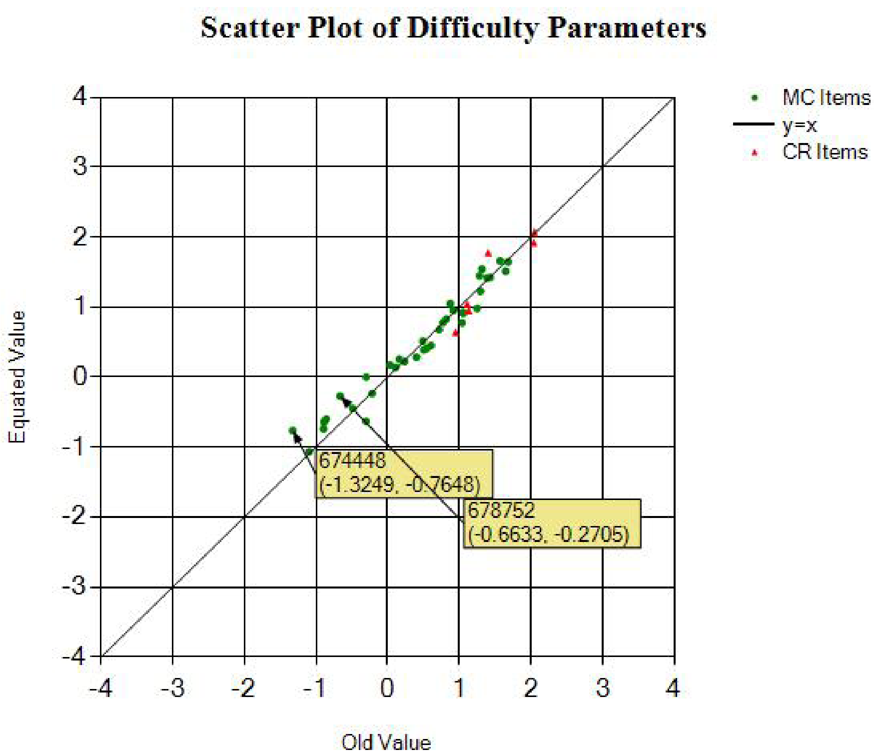


Figure 15–5. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Spring

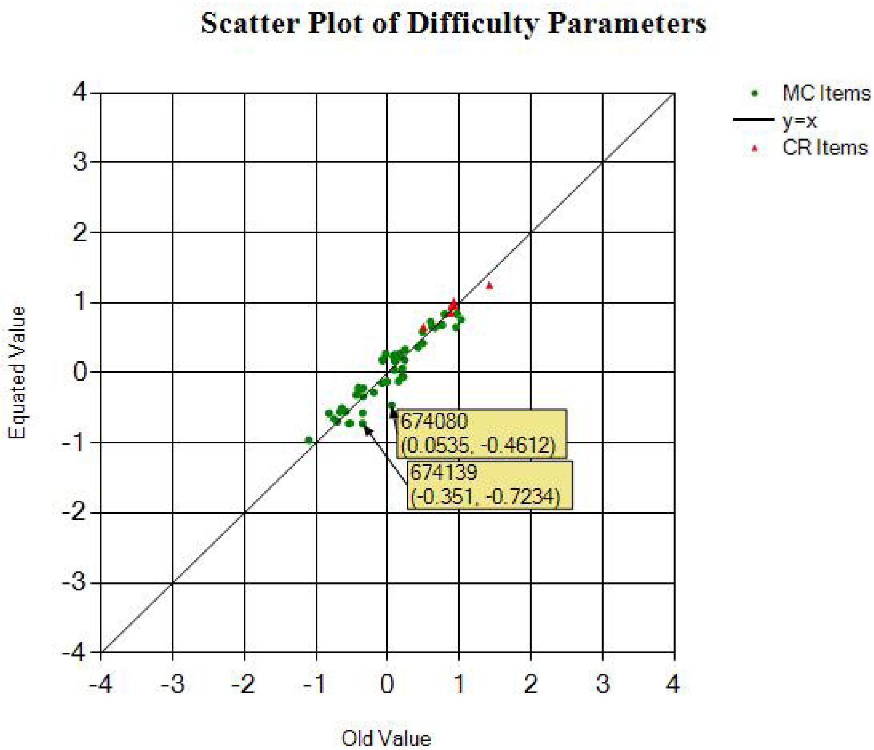


Figure 15–6. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Spring

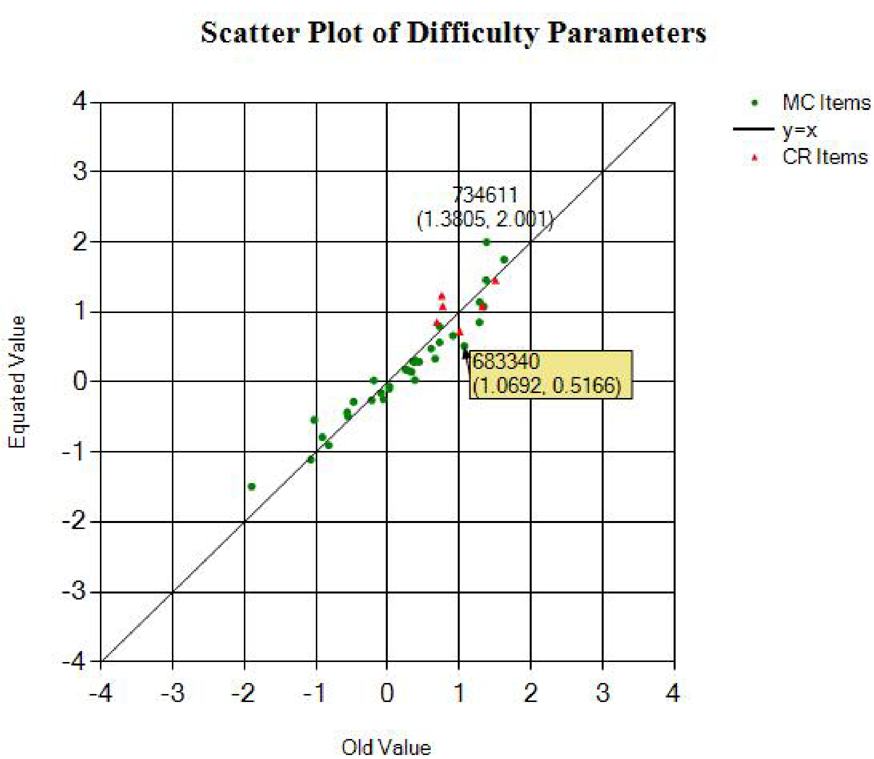


Figure 15–7. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Summer

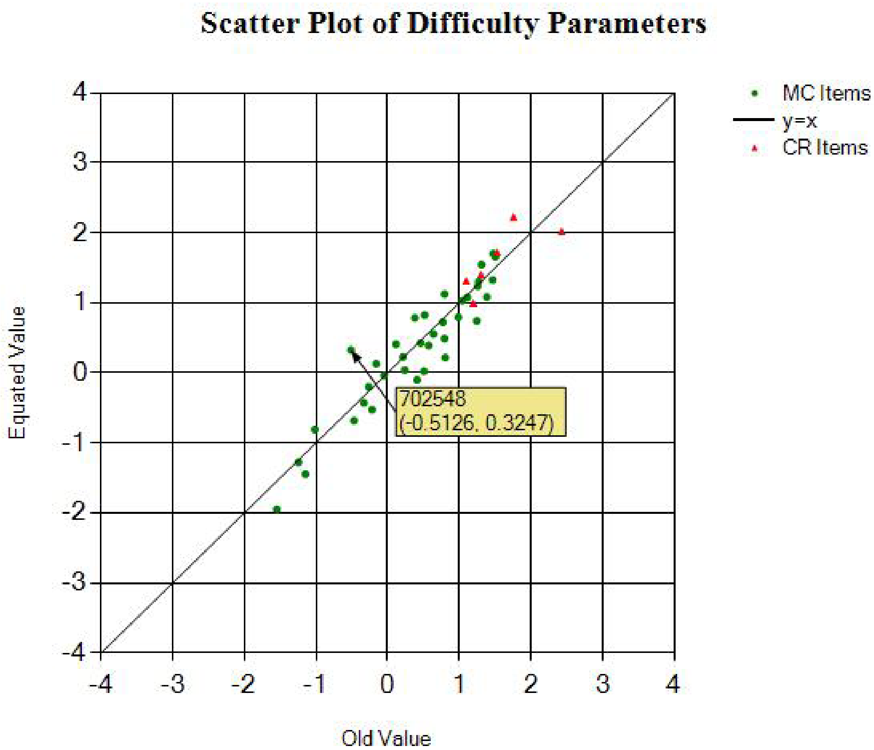


Figure 15–8. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Summer

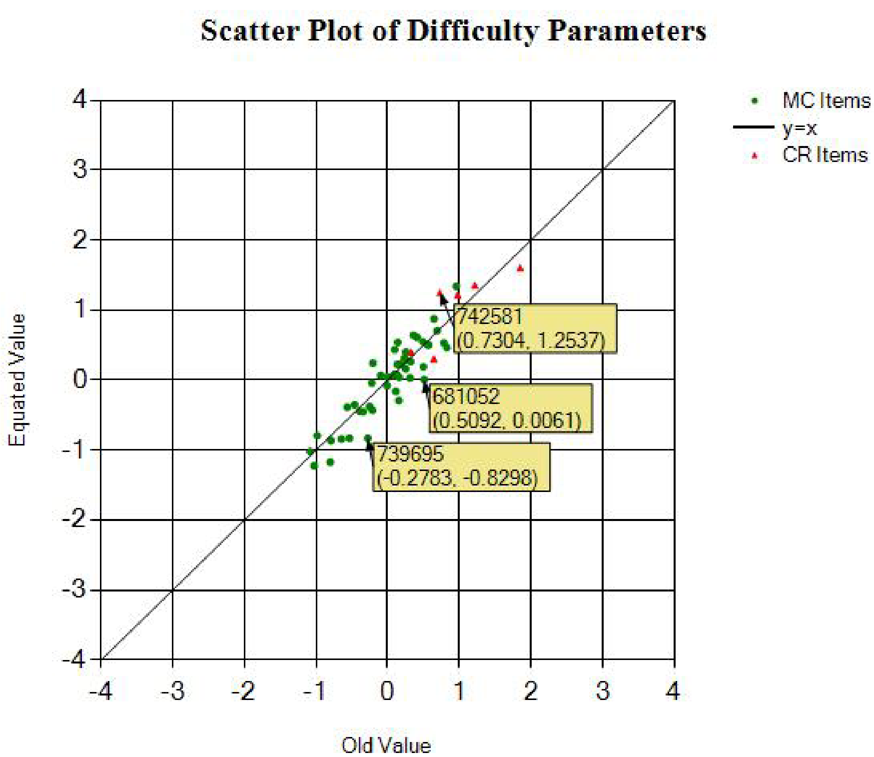
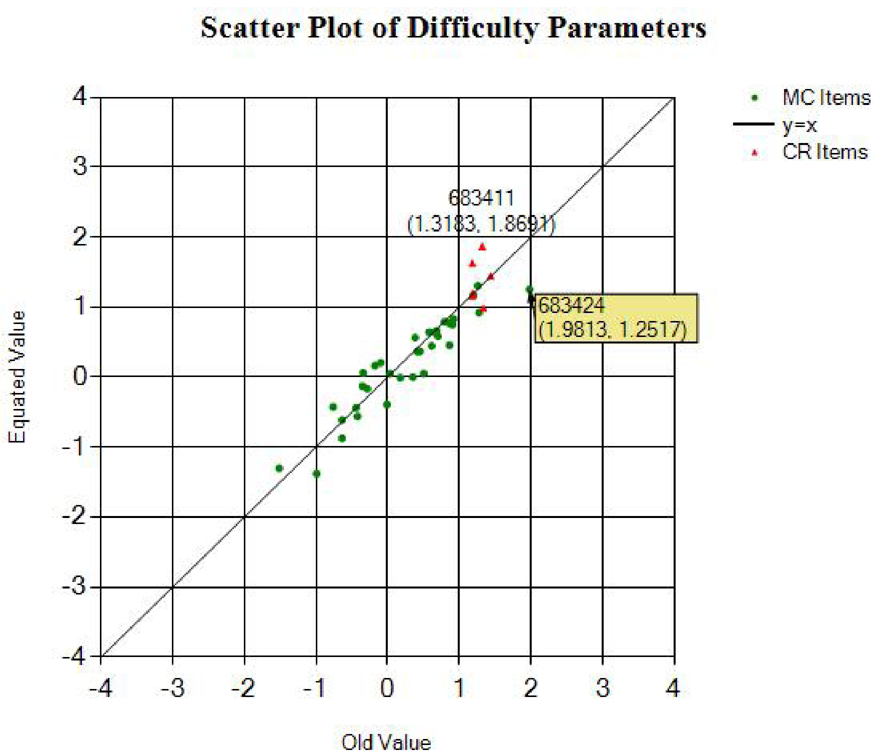


Figure 15–9. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Summer



ANALYSES AT FORM LEVEL

At the form level, the analyses focused on the comparison of pre- and post-equated raw-to-scaled score tables. The outliers, if any, were either kept or removed from the post-equating analyses. Tables L–10 to L–18 in Appendix L contain the raw-to-scaled-score tables produced under different conditions. The three performance level cuts are shown by the thicker lines. As can be seen, the post-equated scaled scores, whether outliers were removed or not, were very close or identical to the pre-equated scaled scores at each raw score point. The differences were within the standard error of measurement. The raw cut scores were the same or within one score point difference.

To summarize, both the item-level and form-level post-equating check analyses results indicate that the raw-to-scaled-score tables produced by using the pre-equated item difficulty parameter estimates can be used to score students.

EQUATING FOR THE EMBEDDED FIELD TEST ITEMS

Field test items were embedded in the spring operational forms to guarantee enough items would be available for future form construction. Equating was needed to place these embedded field test items onto the base or common scale established in Spring 2011. The equating was accomplished by running the calibration of field test items with item parameters of operational items fixed/anchored to the bank values using WINSTEPS. The final Rasch item difficulty estimates can be found in Appendix J.

CHAPTER SIXTEEN: SCORES AND SCORE REPORTS

This chapter provides information about the scores provided for the Pennsylvania Keystone Exams (e.g., scaled scores, performance levels, and module scores), how the scores are presented on score reports, and appropriate and inappropriate uses of the scores.

SCORING

Keystone Exams items include both multiple-choice (MC) and constructed-response (CR) items. Each correct response to an MC item receives a score of 1. Incorrect responses receive a score of 0. Scores on CR items range from 0 to 4, depending on the content area. Table 16–1 summarizes the types of items used in each content-area exam.

Table 16–1. Item Types Used by Content Area

Exam	Item Type MC (1 point)	Item Type CR (3 point)	Item Type CR (4 point)
Algebra I	■		■
Biology	■	■	
Literature	■	■	

DESCRIPTION OF TOTAL-TEST SCORES

Different types of scores have been developed for Keystone Exams reporting. Since the underlying properties of these scores are not necessarily the same, the particular scores used depend on the purposes for which the test has been given. The following types of scores are provided for reporting overall performance on each Keystone Exam:

- Raw scores
- Scaled scores
- Performance levels

RAW SCORES

A raw score (or number-correct score) is the number of points a student earned over all the operational MC and CR items. By itself, the raw score has very limited utility. One limitation is that it can only be interpreted with reference to the total number of items on a specific exam (e.g., a raw score of 15 on a 20-item exam is different from a raw score of 15 on a 30-item exam). In addition, raw scores depend on the difficulty of test items across test forms (e.g., a raw score of 15 on a test with 20 easy items is different from a raw score of 15 on a test with 20 difficult items). Because the difficulty of the items on a test can change from administration to administration, raw scores should not be compared across administrations.

SCALED SCORES

Scaled scores were introduced in Chapter Fourteen. In the simplest sense, a scaled score is a transformed number-correct score. The specifics of the transformation processes for the Keystone Exams were also discussed in Chapter Fourteen. When all students take the same test items, as with the operational items on the Keystone Exams, the more points the student earns, the higher the associated scaled score will be.

The value of switching to the more abstract scaled-score metric is that it produces more general, interpretable, and equitable results. As noted above, a raw score of 30 is meaningless unless the maximum raw score is known. The difficulty of the test items was also mentioned as an additional challenge with interpreting raw scores. Number-correct scores are transformed to scaled scores to remove the effects of test length and item difficulty. (Strictly speaking, transformation of number-correct scores to percent-correct scores would also remove the effect of test length, but it would do nothing to adjust for the difficulty of the items.)

Another advantage of scaled scores is that they lend themselves to interpretations at what is referred to as an interval level, while raw scores do not. Interval-level scales allow an interpretation of a scaled score difference of 5 points to be the same whether the scores are 1295 vs. 1300 or 1445 vs. 1450. Raw-score differences, in this context, cannot be interpreted in this manner and are thus neither generalizable nor equitable.

A scaled score of 1500—or any other value for a particular content-area exam, such as Algebra I—should have the same absolute meaning in the current administration as it had in previous administrations when test scores are properly equated across administrations. More importantly, a significant increase in the scaled score from the previous administration to the current administration means that student performance improved[1]; it does not say anything about whether this administration's exam is easier or harder than last administration's exam. To make these interpretations requires no information about the length or the difficulty of the exam in either administration, although these variables are essential for the process of deriving the scaled scores.

There is considerable auxiliary information presented in this report that might aid in further contextualizing Keystone Exams scaled scores:

- Chapter Fourteen provides information on the development of the Keystone Exams scaled-score system, including transformation formulas, rounding rules, and general scale characteristics (e.g., minimum values).
- Chapter Seventeen provides total-test score statistics. In particular, Table 17–2 lists the scaled score means and standard deviations for the testing results.

PERFORMANCE LEVELS

Keystone Exams results are also reported using four performance levels: Below Basic, Basic, Proficient, and Advanced. The cut scores on the scaled-score metric (i.e., the lowest possible scaled score to enter the Basic, Proficient, and Advanced levels) were presented earlier in this report. However, the information is repeated below (Table 16–2) for convenience.

Table 16–2. Scaled Score Cuts for Each Performance Level by Content Area

Exam	Min	Scaled Score Cuts BB/B	Scaled Score Cuts B/P	Scaled Score Cuts P/A	Max
Algebra I	1,200	1,439	1,500	1,546	1,800
Biology	1,200	1,460	1,500	1,549	1,800
Literature	1,200	1,444	1,500	1,584	1,800

Note: BB = Below Basic; B = Basic; P = Proficient; and A = Advanced

Performance Level Descriptors (PLDs) are another way to attach meaning to the scaled-score metric. They associate precise quantitative ranges of scaled scores with verbal, qualitative descriptions of student status. While much less precise, the qualitative description of the levels is one way for parents and teachers to interpret the student scores. They are also useful in assessing the status of the school. The Pennsylvania General PLDs developed by Pennsylvania Department of Education (PDE) and teacher panels are given below. These are also included on student score reports.

[1] This example is not an endorsement of conducting a trend analysis with just two years of results. Further, small differences may not be statistically or practically significant.

- **Advanced:** Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
- **Proficient:** Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
- **Basic:** Marginal academic performance indicating work approaching, but not yet reaching, satisfactory performance. Performance indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. The student may need additional opportunities and/or increased student academic commitment to achieve the Proficient Level.
- **Below Basic:** Inadequate academic performance indicating little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient Level.

DESCRIPTION OF MODULE SCORES

Each of the Keystone Exams in Algebra I, Biology, and Literature contains two modules. A module score describes performance of a student, school, or district on a particular module (content standard defined in the exam). The following types of scores are provided for Keystone Exams at module level:

- Raw scores
- Scaled scores
- Performance levels

MODULE RAW SCORES

Raw scores at module and assessment anchor levels were reported in different summary reports. As described earlier, a raw score is the number of points a student earned over all the operational MC and CR items; it depends on the difficulty and length of the test form; and it should not be compared across administrations. In the summary reports, the school, district, and/or state median points earned were reported at module and assessment anchor levels. These raw scores can provide some diagnostic information when they are compared with the minimum estimated points needed to pass. The latter is calculated by summing the probabilities of a barely proficient student answering the items included in a module or assessment anchor correctly. The sum is rounded up to the nearest integer. The probability is derived using the Rasch models discussed in Chapter Twelve.

MODULE SCALED SCORES

The module scaled scores were provided in the individual student report. For the Keystone Exams, the module scaled score represents a student's achievement on each module. They can be compared across administrations because they are statistically equated. However, it is not advisable to compare scores across modules because each module contains varying item content and difficulty. This variation is also the reason the total scaled score is not the average of the two modules' scaled scores.

MODULE PERFORMANCE LEVELS

Based on the testing results at the module level, students can be classified as Passed or Not Passed. The derived scaled score cut is 1500 for both modules. This cut score is determined by panelists' recommendations for the proficient cut of the corresponding total test. Note that a student who does not pass a module can still be Proficient or above on the total test if the student performs very well on the other module. If a student is not proficient on the total test but passes one module, although it is recommended that this student take both modules during retesting, the student can choose to take just the non-passed module because the final score is based on the highest combination of module scores.

APPROPRIATE SCORE USE

INDIVIDUAL STUDENTS

Scaled scores on the Keystone Exams indicate a student's achievement with respect to the Keystone Exams Assessment Anchors and Eligible Content. Scaled scores are primarily used to determine student performance level classifications (i.e., a criterion-referenced inference). Scaled scores that are based on Rasch models are typically assumed to be of the interval type, so comparisons may be made on differences in scaled scores. If this assumption holds, then it would be safe to infer for Algebra I that the ability difference between 1410 and 1420 represents the same ability difference that separates 1550 and 1560. Scaled scores can also be used to compare the performance of an individual student to the performance of a similar demographic or subgroup at a school or district. Test score standard errors (discussed in Chapter Eighteen) should be considered.

GROUPS OF STUDENTS

Test results can be used to evaluate performance over time. Mean scaled scores can be compared across administrations within the same content area to indicate whether a student's performance is improving across years. Generally, such trend analyses benefit from using mean results from as many test administrations as possible. Different cohorts of students are used (i.e., the same student or students are not tracked across grade levels). All scores can be analyzed within the same content area for any single administration to determine which demographic or program group had, for example, the highest average performance or the highest percentage of students at or above Proficient.

Module scores can help evaluate academic areas for relative strengths or weaknesses. These module scores provide information to identify areas where further diagnosis is warranted. Generalizations from test results may be made to the specific content domain represented by the academic standards measured in the Keystone Exams. However, all instruction and program evaluations should include as much information from other sources as possible to provide a complete picture of performance.

CAUTIONS FOR SCORE USE

EXTREME ERROR FOR EXTREME SCORES

Student scores toward the minimum or maximum ends of the score range will have very large standard errors of measurement (SEM) and, therefore, should be viewed very cautiously. The maximum scaled score only provides a very rough estimate of a student's ability. For example, if a student achieved the maximum scaled score, for example, 1776 for Biology in the winter administration, it cannot be determined whether this student could have achieved an even higher scaled score. If the test were 10 items longer, a different estimate might have been obtained. Similarly, if the items in a new test are more difficult than the items on a previous administration, the maximum scaled score would likely be higher on the new test because it would take a greater level of achievement to answer the items correctly. In this manner, extreme scaled scores may vary from one administration to the next even if the number of test items does not change. The fluctuation of extreme scaled scores complicates the comparisons of students with scaled scores at the extreme ends of the score distribution. To minimize confusion and potential misinterpretation, the minimum and maximum scaled scores possible on the Keystone Exams have been fixed (see Table 16–2) so they do not change between administrations.

UNIQUE SCALE FOR EACH CONTENT AREA

Scaling was conducted for each content-area exam separately. Therefore, the scaled scores should be interpreted only within each content area. The scaled scores are not status indicators in the same sense as percentile ranks (or scales that are essentially transformations of percentile ranks) and therefore cannot be used to profile relative strengths and weaknesses across content areas. As an example, the scaled scores of 1450 in Algebra I and 1400 in Biology gained by a student do not necessarily imply that the student performed better in Algebra I than in Biology.

USING KEYSTONE EXAMS RESULTS FOR OTHER PURPOSES

Other uses or inferences based on Keystone Exams results may or may not be valid as the validity evidence and arguments provided in Chapter Nineteen may not necessarily support other score uses and interpretations. According to the *Standards for Educational and Psychological Tests* (AERA, APA, & NCME, 1999), if a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary. Finally, a universal caveat for any test's result is that it should not be used for placement and educational planning alone. Instead, other information about the student (e.g., other test performance data) should be included.

REPORT DEVELOPMENT

Several months prior to the first release of reports for the Keystone Exams, PDE and DRC conducted focus groups with Pennsylvania educators and parents/guardians. In the focus groups, educators and parents/guardians provided feedback on report mock-ups for the Keystone Exams. Feedback from the focus groups was used to inform the design and content of the Keystone Exams individual and summary reports. The focus groups targeted educator and parent/guardian constituencies in three geographic regions of the state—the Pittsburgh area, the Harrisburg area, and the Philadelphia area.

Two preliminary educator groups were convened in Harrisburg on November 15 and 17, 2010. These groups, totaling 34 educators, reviewed the student report and provided feedback using both a survey and group discussion. Substantive changes to the individual student report were made on the basis of these meetings, with two different versions of the report emerging from these reviews. These two groups did not review the summary reports.

A second set of focus groups were conducted in December 2010 to review the updated reports. For the December meetings there were 35 panelists (22 educators & 13 parents) for six focus groups in Pittsburgh (December 3), Harrisburg (December 6), and King of Prussia (December 7). The three educator groups reviewed the two versions of the student report and the one version of the school summary report. The three parent groups reviewed the two versions of the student report.

Feedback from these two focus groups was taken into consideration during final report development. For more information about the focus groups, please refer to the *Keystone Exams Score Report Focus Group Findings* (Pennsylvania Department of Education, 2011).

REPORTS

The following score reports are provided to students, schools, and districts for the Keystone Exams in Algebra I, Biology, and Literature:

- Individual student report
- School summary report
- District summary report
- State summary report
- Report interpretation guide

PARENT LETTER

Parent letters were delivered to Pennsylvania districts when district files were posted after each Keystone administration. This score report provided parents and students with their first glimpse of performance on the 2016/2017 Keystone Exams. This report provides results at the student level.

INDIVIDUAL STUDENT REPORT

A student report is provided for all students who took the Keystone Exams. Two copies of the individual student report for all Keystone Exams were sent to each school district and charter school for distribution to parents, teachers, guidance counselors, and/or principals. School districts and charter schools may publish the results of the Keystone Exams school-level reports. This report is a two-page color document that provides the types of scores explained earlier in this chapter. Screenshots of the two pages from a sample individual student report are provided in Figures 16–1 and 16–2.

Figure 16–1. Page 1 of the Individual Student Report

PENNSYLVANIA
KEYSTONE EXAMS

Student Report

Student Name:	SAMPLE STUDENT
PA Student ID:	*****12345
School:	SAMPLE HS
District:	SAMPLE SD
Test Date:	Spring 2011
Grade:	09

Content Area:
Algebra I

Student's Keystone Exam Result

		Goal Range	
Below Basic	Basic	Proficient	Advanced
		✓	

Dear Family:

This report provides information about your child's performance on a Pennsylvania test known as the Keystone Exam. Your child took this Keystone Exam in May 2011. On this page, you can see your child's overall performance – below basic, basic, proficient or advanced.

On this report, you will find specific information about your child's performance on the Algebra I Keystone Exam. It displays your child's Highest Total Test Scale Score to Date for Module 1 and Module 2. Module 1 assesses Operations and Linear Equations and Inequalities, and Module 2 assesses Linear Functions and Data Organization. No previous scores will be displayed because the May 2011 administration marks the first time this test was given.

For detailed information about how the Keystone Exams are being integrated into the Pennsylvania graduation requirements, please visit the Pennsylvania Department of Education's Standards Aligned System website at www.pdesas.org, or contact your child's school.

Sincerely,
Ronald J. Tomalis
Ronald J. Tomalis
Secretary of Education

About the Keystone Exams

The Keystone Exams are tests students take at the end of specific high school level courses, including for 2010-11: Algebra I, Biology and Literature. They are offered in both paper/pencil and online formats.

Keystone Exams are one component of Pennsylvania's new system of high school graduation requirements affecting students in the class of 2015 and beyond. These tests were developed by Pennsylvania educators and are aligned to the standards adopted by the Pennsylvania State Board of Education. The results help students, parents and educators understand how well we are meeting rigorous expectations for student achievement in core subject areas. In future years, under Pennsylvania's new system of graduation requirements, Keystone Exam results will help determine whether or not a student has mastered the standards associated with earning a high school diploma.

For more information about the Keystone Exams, please visit the Pennsylvania Department of Education's Standards Aligned System website at www.pdesas.org (select "Assessments" and then "Keystone Exams").

www.pdesas.org

pennsylvania
DEPARTMENT OF EDUCATION

Page 1

Figure 16–2. Page 2 of the Individual Student Report

Performance Level on Total Test

Highest Total Test Scale Score to Date

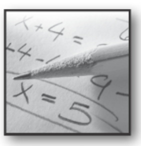
Student's total test scale score is indicated by the (↑) . If this student were to test again under similar circumstances, his/her score would likely remain in the following range: **1520-1550.**

1200	1439	1500	1546	1800

Below Basic	Basic	Proficient	Advanced
Inadequate academic performance that indicates little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.	Marginal academic performance, work approaching, but not yet reaching, satisfactory performance. Performance indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content. The student may need additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.	Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content.	Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content.

Highest Total Test Scale Score to Date - 1535

Algebra I - SAMPLE STUDENT



	Module 1 Operations and Linear Equations & Inequalities			Module 2 Linear Functions and Data Organization			Total Test [1]	
	Result	Scale Score	Test Date	Result	Scale Score	Test Date	Scale Score	Performance Level
Highest Total Test Scale Score to Date [2]	Passed	1524	Spring 2011	Passed	1547	Spring 2011	1535	Proficient

[1] The scale score for the Total Test reflects the highest score computed as the combination of the two highest module scores to date. Given that modules contain varying item content and difficulty, the total scale score may not equal the average of the modules.

[2] Students who do not score Proficient on the Total Test may choose to take the test more than once.

www.pdesas.org

pennsylvania DEPARTMENT OF EDUCATION

Page 2

SUMMARY REPORTS

Summary reports are provided at the school, district, and state levels. These reports contain summary information about the percentage of students in each of the four performance levels. Raw scores are also provided by assessment anchor to allow schools or districts to identify strengths and weaknesses at the content-strand level.

REPORT INTERPRETATION GUIDE

A report interpretation guide is provided to help parents and other Keystone Exams stakeholders better understand test result information presented in the individual student report. The report interpretation guide can be found on the PDE SAS website (www.pdesas.org/Assessment/Keystone).

CHAPTER SEVENTEEN: OPERATIONAL TEST STATISTICS

This chapter presents various summary statistics for the total-test scores based on the final data file described in Chapter Nine. Related information covered elsewhere in this report includes the item-level statistics that were presented in Chapters Eleven (classical item statistics) and Twelve (Rasch item statistics). The reader is referred to these chapters for additional consideration as item difficulty distributions can affect total score distributions.

PERFORMANCE LEVEL STATISTICS

Table 17–1 presents performance level percentages by test administration, content area, and student type. As can be seen from the table, the overall percentage in each performance level varied from administration to administration, depending on the ratio of the first-time testers and retesters. In general, retesters had a lower percentage of students in the Proficient and Advanced levels than the first-time testers did.

Table 17–1A. Performance Level Percentages: All Testers

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	55,500	18.2	56.3	20.8	4.7
Winter	Biology	41,916	28.1	43.8	19.6	8.5
Winter	Literature	36,551	13.9	43.6	38.5	4.0
Spring	Algebra I	167,097	16.2	42.1	23.4	18.3
Spring	Biology	142,262	19.9	32.0	28.4	19.7
Spring	Literature	130,328	12.5	30.0	49.3	8.3
Summer	Algebra I	1,454	3.6	56.2	35.4	4.9
Summer	Biology	960	4.5	43.3	44.9	7.3
Summer	Literature	446	3.1	42.8	52.5	1.6

Table 17–1B. Performance Level Percentages: First-time Testers

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	11,107	23.3	29.9	26.6	20.2
Winter	Biology	13,219	24.4	20.1	29.4	26.1
Winter	Literature	15,799	13.6	22.9	54.6	9.0
Spring	Algebra I	118,704	16.1	32.2	26.4	25.3
Spring	Biology	115,473	17.6	25.8	32.4	24.2
Spring	Literature	110,966	11.5	24.0	54.8	9.7
Summer	Algebra I	197	14.7	46.2	21.3	17.8
Summer	Biology	63	6.3	9.5	25.4	58.7
Summer	Literature	16	0.0	12.5	62.5	25.0

Table 17–1C. Performance Level Percentages: Retesters

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	44,393	16.9	62.9	19.4	0.8
Winter	Biology	28,697	29.9	54.7	15.1	0.4
Winter	Literature	20,752	14.1	59.4	26.3	0.1
Spring	Algebra I	48,393	16.5	66.4	16.0	1.1
Spring	Biology	26,789	29.6	58.7	11.5	0.3
Spring	Literature	19,362	18.2	63.9	17.8	0.1
Summer	Algebra I	1,257	1.8	57.8	37.5	2.9
Summer	Biology	897	4.3	45.7	46.3	3.7
Summer	Literature	430	3.3	44.0	52.1	0.7

SCALED SCORES

Table 17–2 provides the scaled score means and standard deviations by test administration, content area, and student type. As can be seen from the table, in most of the cases, first-time testers had a higher average scaled score than retesters did.

Table 17–2. Means and Standard Deviations of Scaled Scores

Administration	Content Area	All Testers		First-Time Testers		Retesters	
		Mean	SD	Mean	SD	Mean	SD
Winter	Algebra I	1477.5	42.0	1493.2	64.9	1473.5	32.8
Winter	Biology	1484.9	45.2	1508.6	63.5	1474.0	27.5
Winter	Literature	1494.8	49.1	1513.7	59.2	1480.5	33.3
Spring	Algebra I	1495.2	56.6	1504.4	61.8	1472.5	31.0
Spring	Biology	1504.4	52.8	1511.6	54.7	1473.0	25.7
Spring	Literature	1508.7	54.8	1514.8	55.5	1473.7	32.9
Summer	Algebra I	1495.6	30.3	1492.2	54.1	1496.1	24.5
Summer	Biology	1505.5	35.3	1572.7	76.0	1500.8	24.4
Summer	Literature	1504.2	31.5	1555.6	50.4	1502.3	29.0

RAW SCORES

SUMMARY STATISTICS

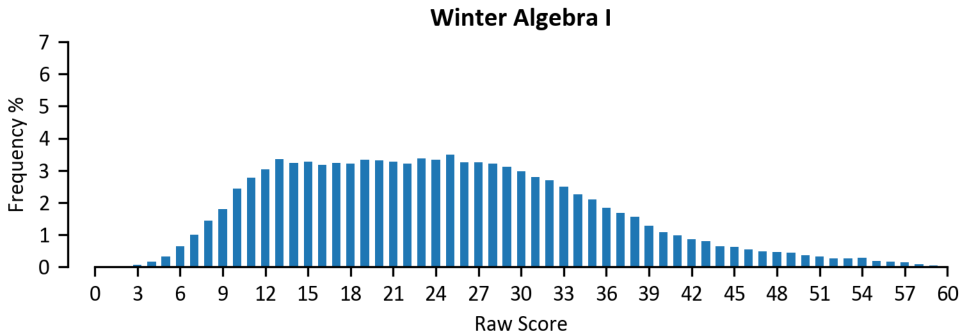
The reader is referred to Appendix M to review summary statistics for the operational raw scores. The statistics reported include number of points possible (Pts.), number of items (Len.), number of students tested (*N*), mean number of score points received (Mean), standard deviation of test scores (SD), reliability (r), and traditional standard error of measurement (SEM).

SCORE DISTRIBUTIONS

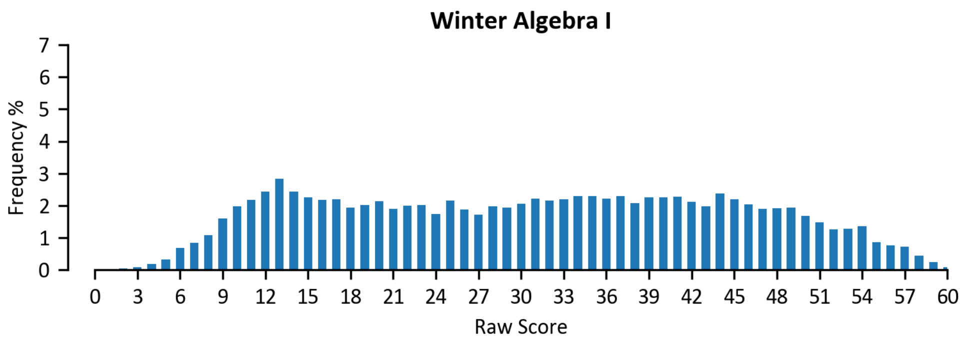
Raw score distributions are provided in Figure 17–1. As can be seen from the graphs, overall, the retesters scored lower than the first-time testers.

Figure 17–1. Raw Score Distributions

All Testers



First Time Testers



Retesters

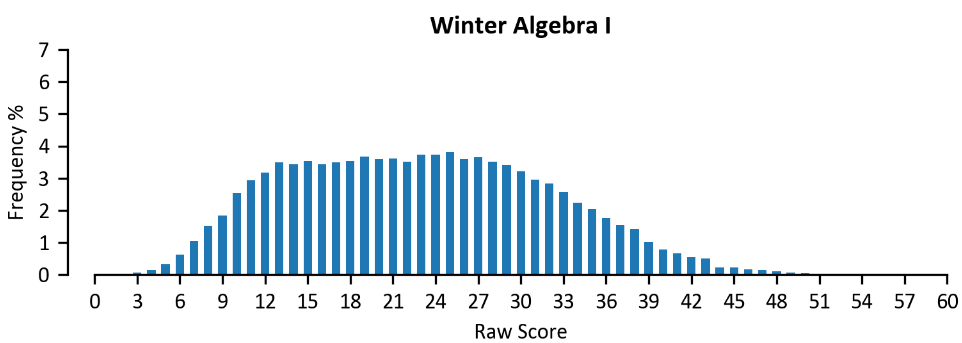
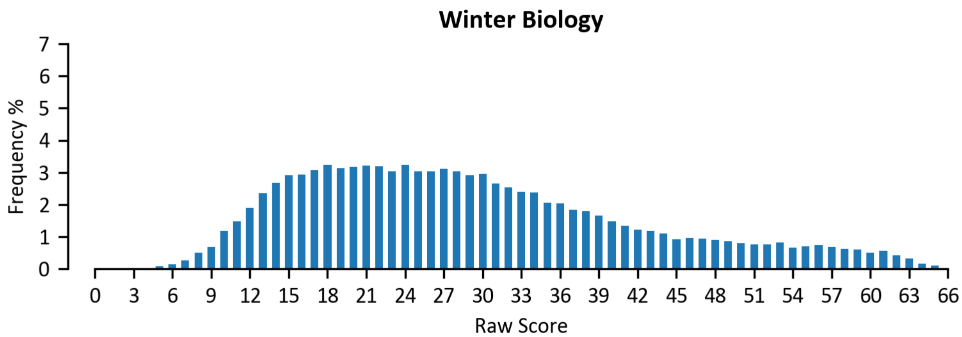
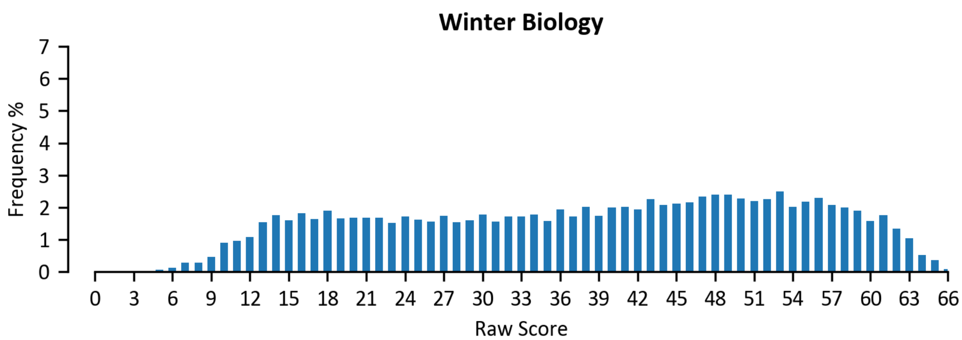


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

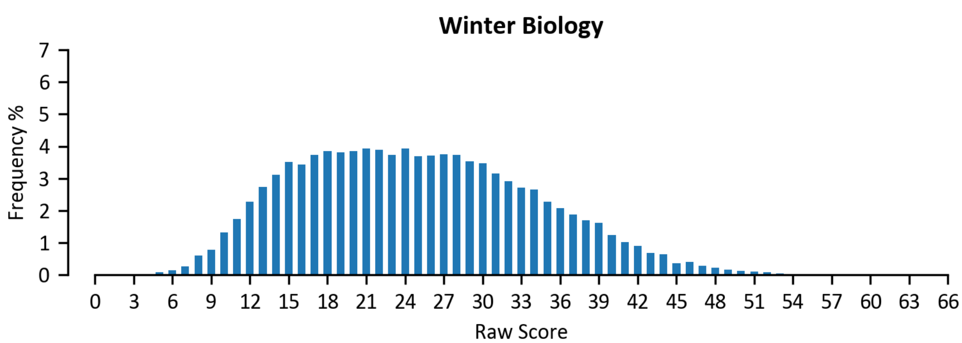
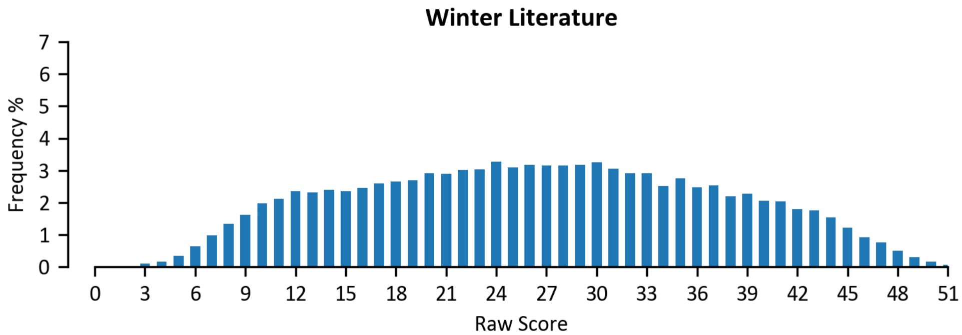
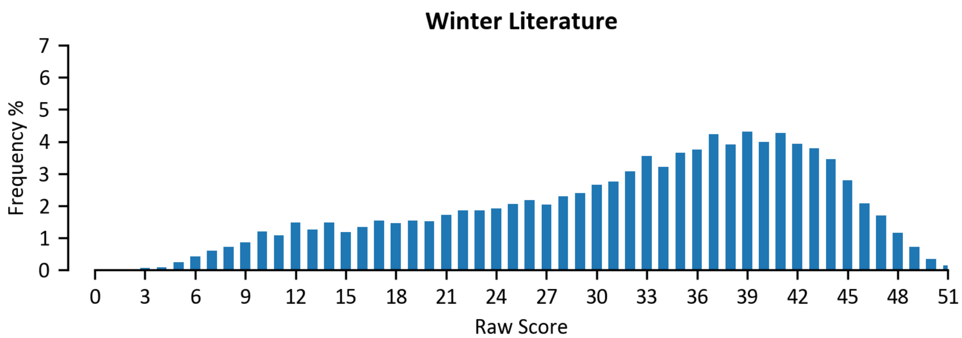


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

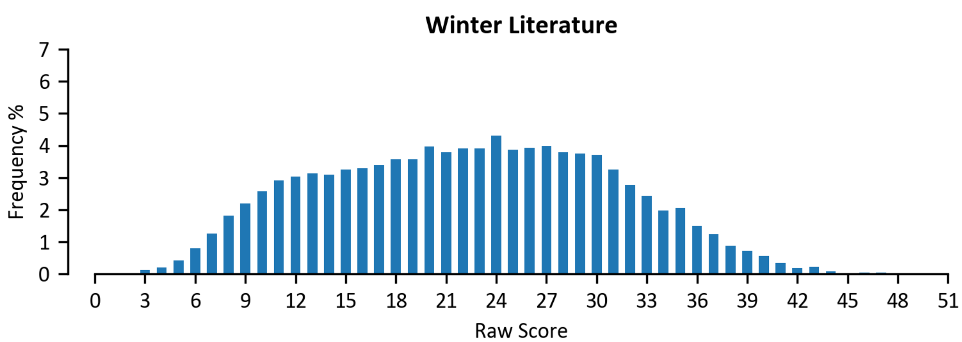
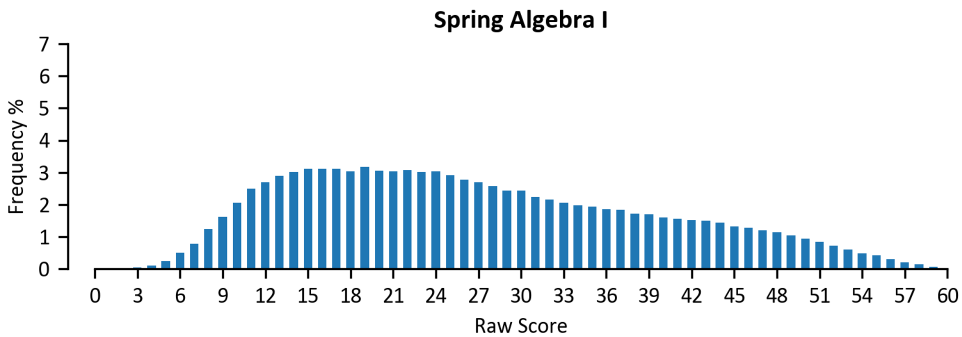
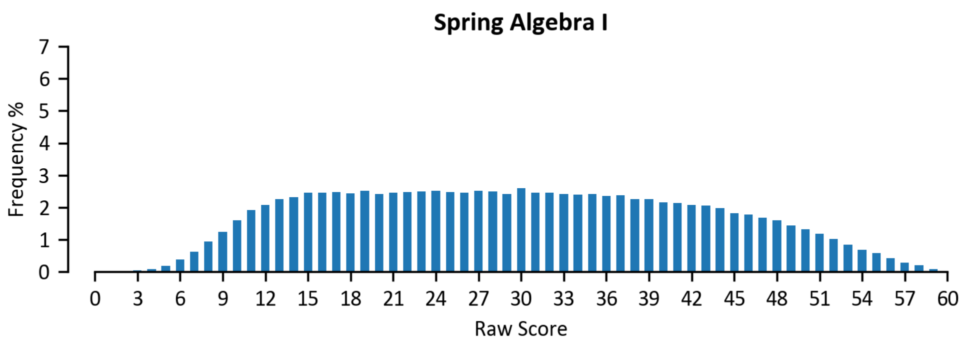


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

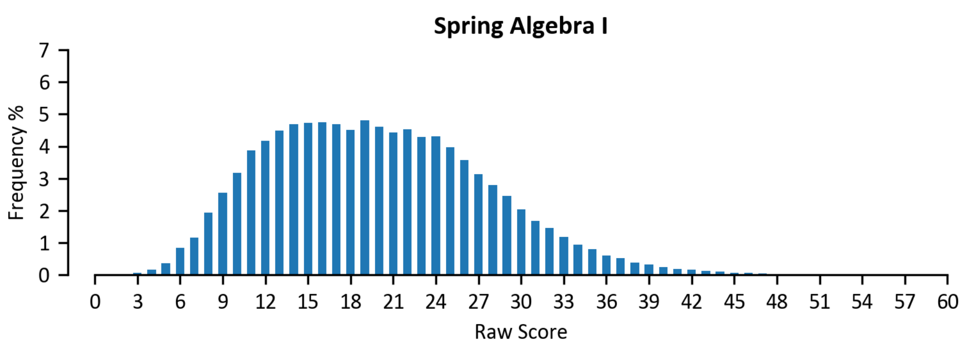
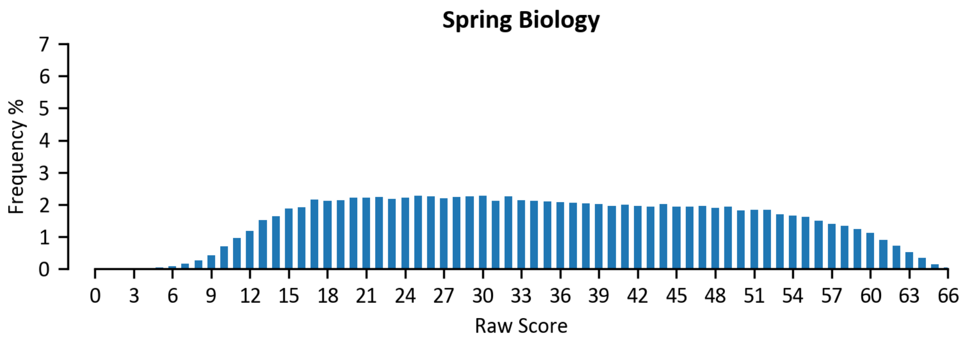
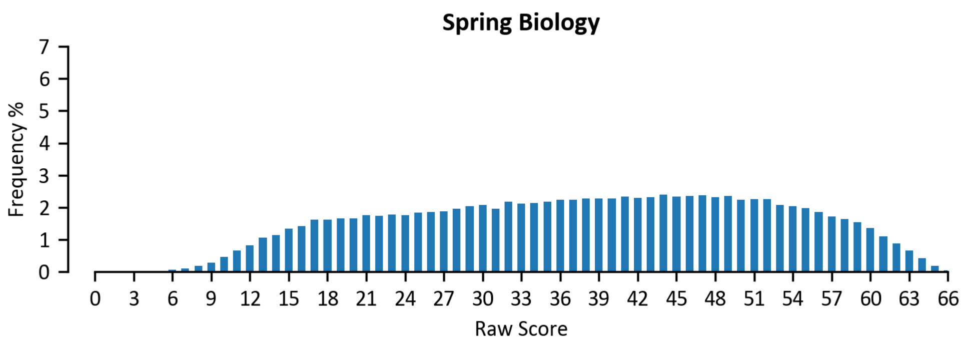


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

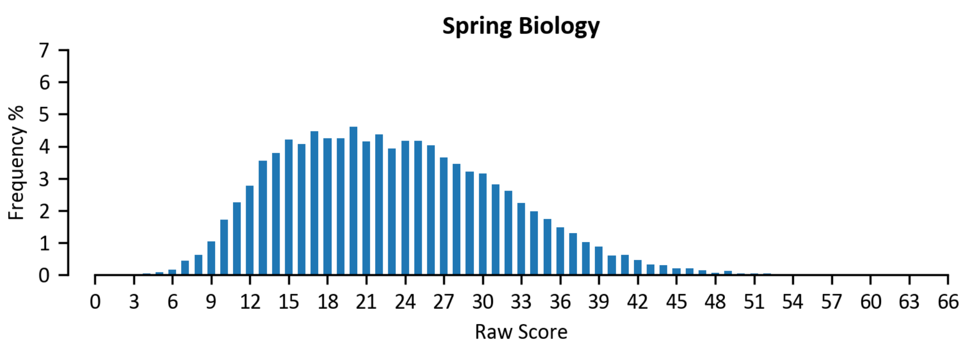
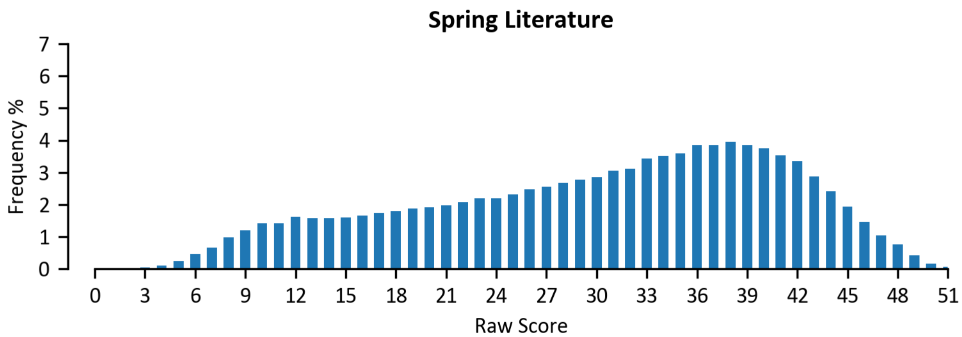
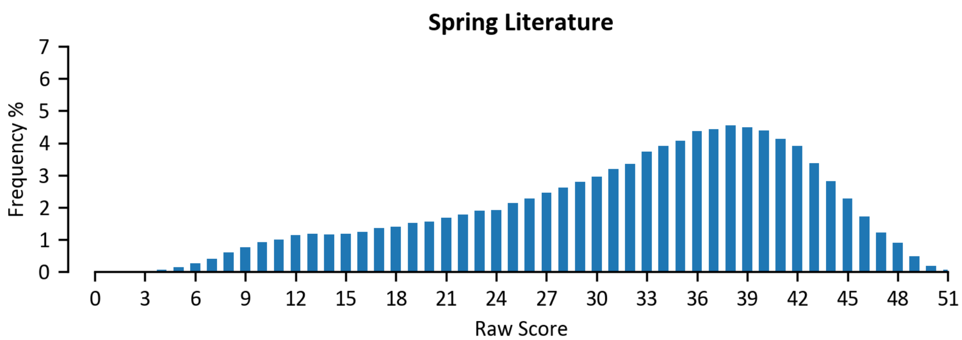


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

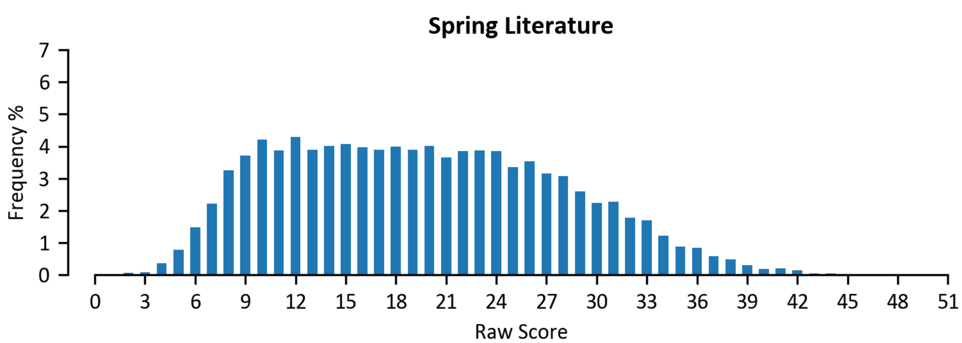
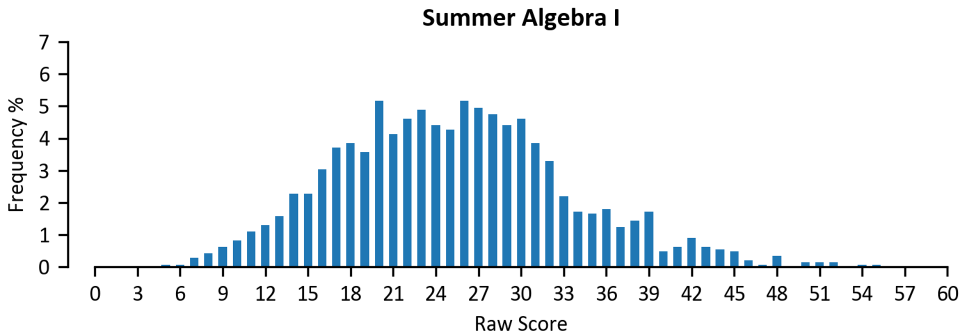
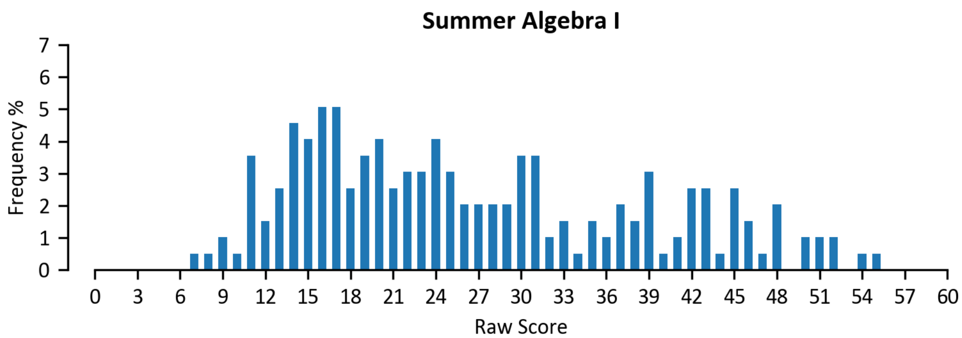


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

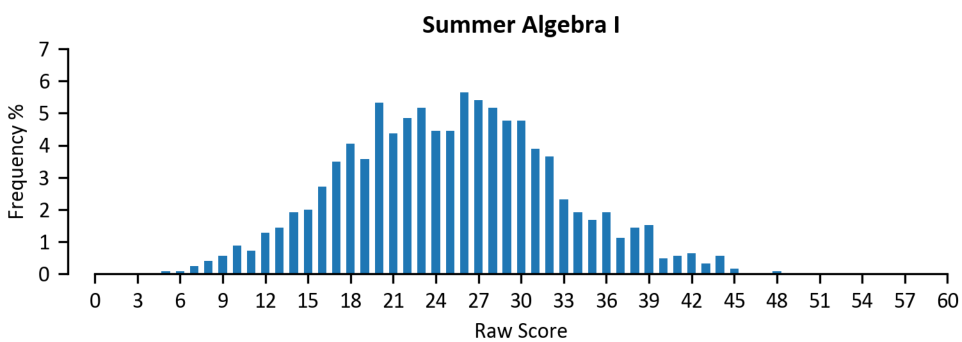
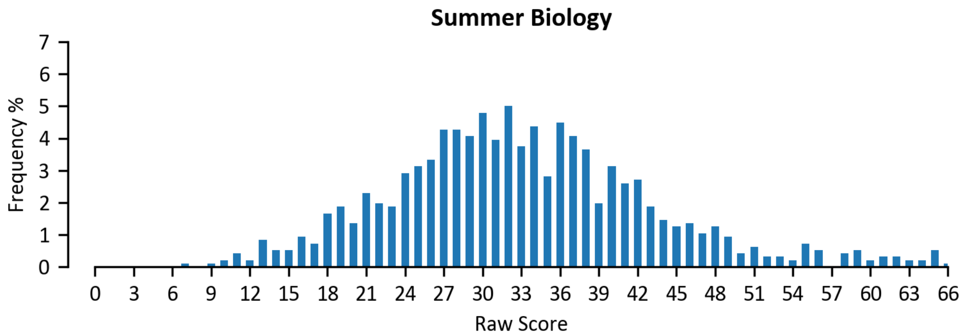
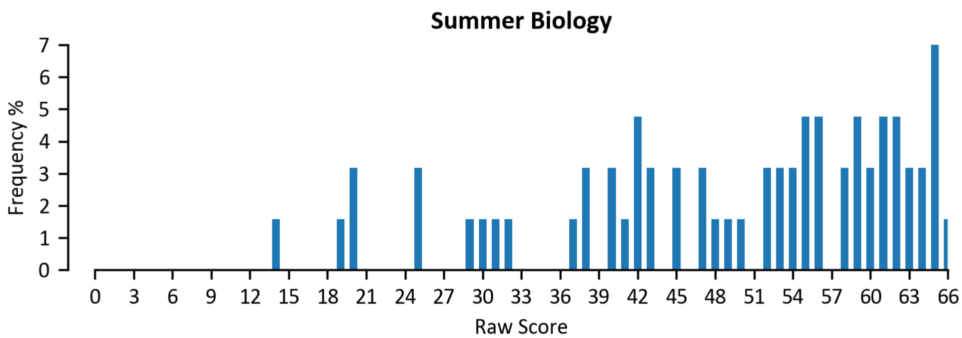


Figure 17–1 (continued). Raw Score Distributions

All Testers



First Time Testers



Retesters

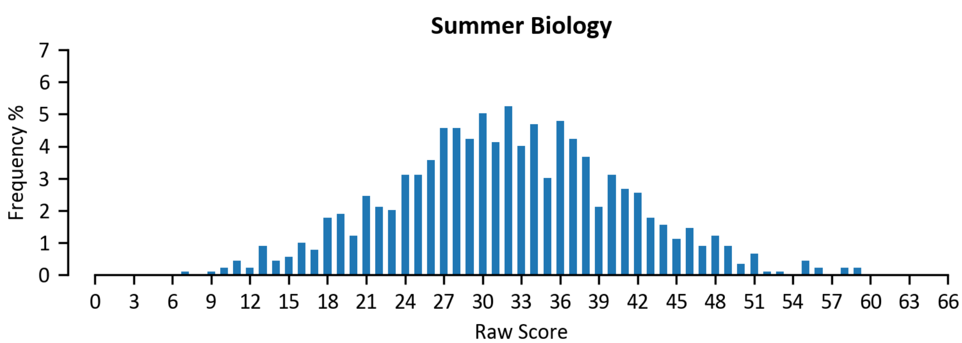
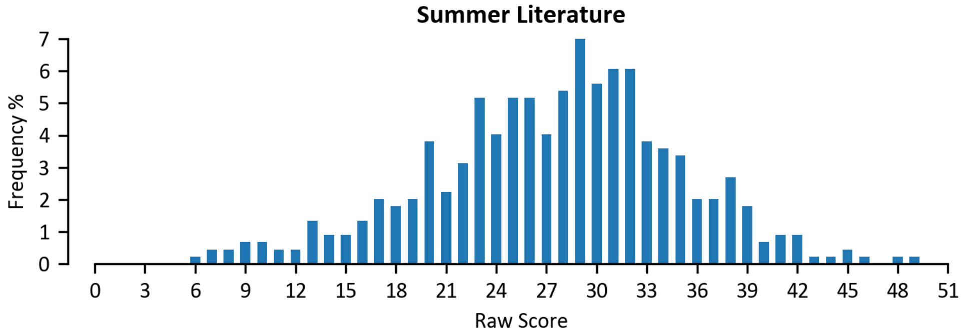
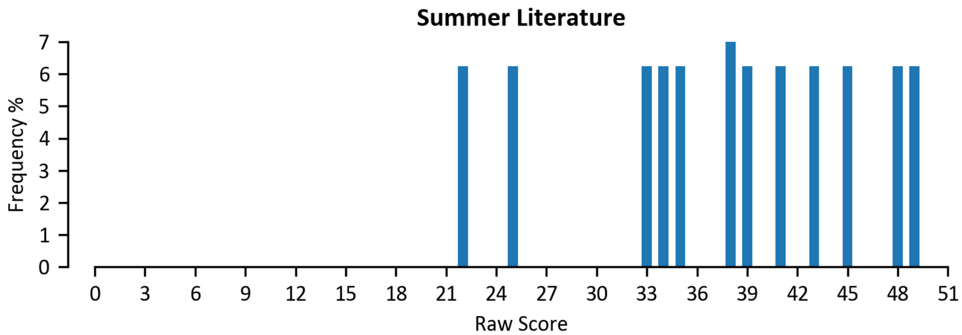


Figure 17–1 (continued). Raw Score Distributions

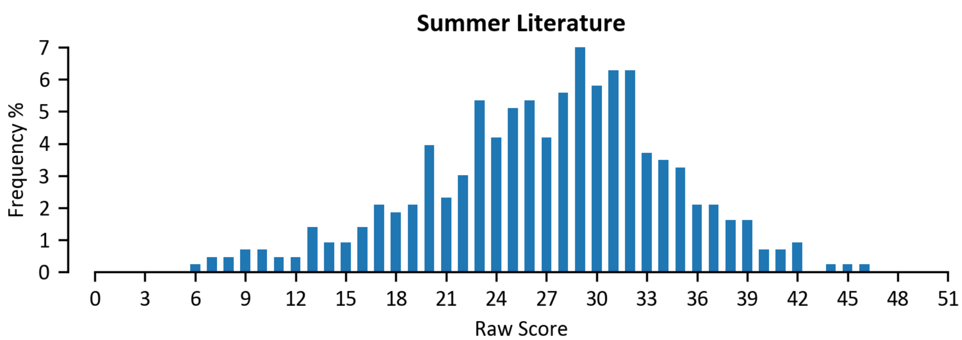
All Testers



First Time Testers



Retesters



CHAPTER EIGHTEEN: RELIABILITY

This chapter addresses the reliability of Pennsylvania Keystone Exams test scores. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), reliability refers to

> the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group (p. 25).

Frisbie (2005) highlighted several elements of this definition. First, reliability is a property of test scores, not of a test itself. Many may appreciate this distinction, but in casual usage, individuals frequently make reference to a reliable test. While reliability concerns test scores (and not the test specifically), it's important to appreciate the fact that test scores can be affected by characteristics of the instrument. For example, all other things being equal, tests with more items/points tend to be more reliable than tests with fewer items/points. Second, reliability coefficients are group specific. Reliabilities tend to be higher in populations that are more heterogeneous and lower in populations that are more homogeneous. Consequently, both test length and population heterogeneity should be considered when evaluating reliability.

There are other reliability considerations that may be less evident from the definition above yet are still important for test users to understand. While freedom from measurement error is highlighted in the definition, reliability is specifically concerned with random sources of error. Indeed, the degree of inconsistency due to random error sources is what determines reliability: less consistency is associated with lower reliability and more consistency is associated with higher reliability. Of course, systematic error sources also exist. These can artificially increase reliability and decrease validity. (Validity is further discussed in Chapter Nineteen.)

Another noteworthy issue is that multiple sources of error exist (e.g., the day of testing, the items used, the raters who score the items). However, most widely used reliability indices only reflect a single type of error. Consequently, it is important for test users to understand which specific type of error is being considered in a reliability study, and equally, if not more importantly, which types are not.

Understanding the distinction between relative error and absolute error is important because many reliability indices only reflect relative error. Relative error is of interest whenever the relative ordering of individuals with respect to their test performance is of interest. When specific score values are considered important (e.g., if cut scores are used), then absolute error is of interest, too. Generally, there is more error variance when considering the absolute scores of examinees, which, in turn, suggests lower reliability. Understanding examinee rank-order stability is also important; however, such stability might be well achieved even when the specific score values are considerably different.

As the above discussion suggests, reliability is a complex, nonunitary notion that cannot be adequately represented by a single number. There are several reliability indices available, and these may not provide the same results (Frisbie, 2005). The remainder of this chapter covers the following:

- Reliability coefficients and their interpretation
- Unconditional and conditional standard errors of measurement
- Decision consistency
- Rater agreement

RELIABILITY INDICES

As shown below, the reliability coefficient expresses the consistency of test scores as the ratio of true score variance to total score variance. The total variance contains two components: variance in true scores and variance due to the imperfections in the measurement process. Put differently, total variance equals true score variance plus error variance.[1]

$$\rho_X^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

Reliability coefficients indicate the degree to which differences in test scores reflect true differences in the attribute being tested rather than random fluctuations. Total test score variance (i.e., individual differences) is partly due to real differences in the attribute (true variance) and partly due to random error in the measurement process (error variance).

Reliability coefficients range from 0.0 to 1.0. If all test score variance were true, the index would equal 1.0. The index would be 0.0 if none of the test score variance were true. Such scores would be pure random noise—that is, all measurement error. If the index had a value of 1.0, scores would be perfectly consistent—that is, contain no measurement error. Although values of 1.0 are never achieved in practice, it is clear that larger coefficients are more desirable as they indicate that test scores are less influenced by random error. (How big is big enough and how small is too small are issues considered in a later section.)

As noted in the introduction, there are several different indices that can be used to estimate this ratio. One approach is referred to as internal consistency, which is derived from analyzing the performance consistency of individuals over the items within a test. As discussed below, these internal consistency indices do not take into account other sources of error, such as day-to-day variations (e.g., student health, testing environment) or rater inconsistency.

COEFFICIENT ALPHA

Although a number of reliability indices exist, perhaps the most frequently reported for achievement tests is coefficient alpha. Consequently, this index is the one reported for the Keystone Exams (see the column with title “r” in Appendix M). Alpha indicates the internal consistency over the responses to a set of items measuring an underlying trait, in this case, academic achievement, in content areas such as algebra, biology, and literature.

Alpha is an internal consistency index. It can be conceptualized as the extent to which an exchangeable set of items from the same domain would result in a similar rank ordering of students. Note that relative error is reflected in this index. Variation in student performance from one sample of items to the next should be of particular concern for any achievement test user. Consider two hypothetical vocabulary tests intended for the same group of students. Each test contains different sets of unique words that are believed to be randomly equivalent, perhaps like the ones shown below:

[1] A Covariance term is not required as true scores and error are assumed to be uncorrelated in classical test theory.

Table 18–1. Two Hypothetical Vocabulary Tests

Test One	Test Two
Abase	Abate
Boon	Bilk
Capricious	Circuitous
Deface	Debase
....	....
Zealous	Zenith

If a representative group of students could take both of these tests, the correlation between the scores obtained would represent the parallel-forms reliability of the test scores. However, such data-collection designs are impractical in large-scale settings, and experimental confounds like fatigue and practice effects are likely to affect the results. Internal-consistency reliability indices arose in part to provide reliability measures using the data from just a single test administration. So, if students only took Test One and the coefficient alpha index for those test scores were high, this would suggest that Test Two would provide a very similar rank ordering of the students if they had taken it instead. If coefficient alpha were low, dissimilar rank orderings would likely be observed—again, relative-error variance is reflected in alpha.

FORMULA

Consider the following data matrix representing the scores of persons (rows) on items (columns):

Table 18–2. Person × Item Score (*Xpi*) Infinite (Population-Universe) Matrix

Person	Item			
	1	2	. . . I	. . . k
1	*Y*11	*Y*12	. . . *Y*1*i*	. . . *X*1*k*
2	*Y*21	*Y*22	. . . *Y*2*i*	. . . *X*2*k*
. . . . . .				
P	*Yp*1	*Yp*2	. . . *Ypi*	. . . *Xpk*
. . . . . .				
N	*YN*1	*YN*2	. . . *YNi*	. . . *XNk*

Note: Adapted from Cronbach and Shavelson (2004).

Then, a general computational formula for alpha is as follows:

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

where N is the number of parts (items or testlets), σ_X^2 is the variance of the observed total test scores, and $\sigma_{Y_i}^2$ is the variance of part i.

FURTHER INTERPRETATIONS

RULES OF THUMB

Which reliability values are considered high enough? Which values are considered too low? Although frequently asked for, any rules of thumb for interpreting the magnitude of reliability indices are mostly arbitrary. Another approach is to research the reliabilities from similar testing instruments to see what values are commonly observed. For the Keystone Exams, comparisons to tests of similar lengths that were administered to similar student populations from other large-scale assessment programs would be relevant. For many other state assessment programs, reliabilities in the low 0.90s are usually the highest ever observed, and reliabilities in the high 0.80s are very common.

The lower a given reliability coefficient, the greater the potential for over-interpretation of the associated results. As suggested earlier, there is no firm guideline regarding how low is too low. However, as an informative point of reference, a reliability coefficient of 0.50 would mean that there is as much error variance as true-score variance in the scores.

IS ALPHA A LOWER LIMIT TO RELIABILITY?

According to Brennan (1998), the conventional wisdom that coefficient alpha is a lower limit to reliability is based largely on a misunderstanding. In reflecting on the 50th anniversary of his seminal 1951 article, Cronbach—in Cronbach and Shavelson (2004)—expressed similar misgivings about this conventional wisdom:

> one could argue that alpha was almost an unbiased estimate of the desired reliability. . . the *almost* in the preceding sentence refers to a small mathematical detail that causes the alpha coefficient to run a trifle lower than the desired value. This detail is of no consequence and does not support the statement made frequently in textbooks or in articles that alpha is a lower value to the reliability coefficient. That statement is justified by reasoning that starts with the definition of the desired coefficient as the expected consistency among measurements that had a higher degree of parallelism than the random parallel concept implied.

The assumptions for three common parallelism models are presented in Table 18–3. Alpha's assumptions come from the Essentially-Tau-Equivalent model, which does not require equal means or equal variances across test parts. Based on this, Brennan (1998) asserts that the lower-limit issue, as conceptualized by many, provides an answer to a question that is of minimal importance. Reframed differently, the goal of selecting a reliability coefficient is not to find the one that provides the highest coefficient, but the one that most accurately reflects the test data under study.

It is important to note that there are factors encountered in practice that may legitimately make coefficient alpha an underestimate of reliability. However, there are also factors that might make coefficient alpha an overestimate of reliability. Both possibilities are discussed further below and generally arise when the Essentially-Tau-Equivalent assumptions are strained.

Table 18–3. Summary of Expectations/Observable Relationships for Different Parallelism Models

Relationship	Degree of Measurement Parallelism* Classically Parallel	Degree of Measurement Parallelism* Essentially-Tau Equivalent	Degree of Measurement Parallelism* Congeneric
Content Similarity	Yes	Yes	Yes
Equal Means across Parts	Yes	No	No
Equal Variances across Parts	Yes	No	No
Equal Covariances across Parts	Yes	Yes	No
Equal Covariances with other Variables	Yes	Yes	No

*Note: Other models exist but are not considered here due to their limited application in practice.

BIASES THAT MIGHT MAKE ALPHA AN UNDERESTIMATE OF RELIABILITY

There are factors that might negatively bias coefficient alpha, making the apparent reliability lower than it may actually be. In practice, two situations frequently encountered that might cause this include tests that are composed of mixed item types (e.g., MC and CR items) and tests that include a planned stratification of the test items according to topics or subdomains.

Although both situations strictly violate the assumptions used in deriving the coefficient alpha (i.e., the tests are not based on equal part lengths in the former case and are not randomly parallel in the latter case), neither necessarily guarantees that the reliability will be markedly lower. In the latter case, reliability will be underestimated only when strand items are homogeneous enough for the average covariance within strata to exceed the average covariance between strata. Although both are potential influences for the Keystone Exams, the total test score reliabilities (i.e., r) reported in Appendix M ranged from 0.84 to 0.94, indicating highly consistent test scores for these instruments.

BIASES THAT MIGHT MAKE ALPHA AN OVERESTIMATE OF RELIABLITY

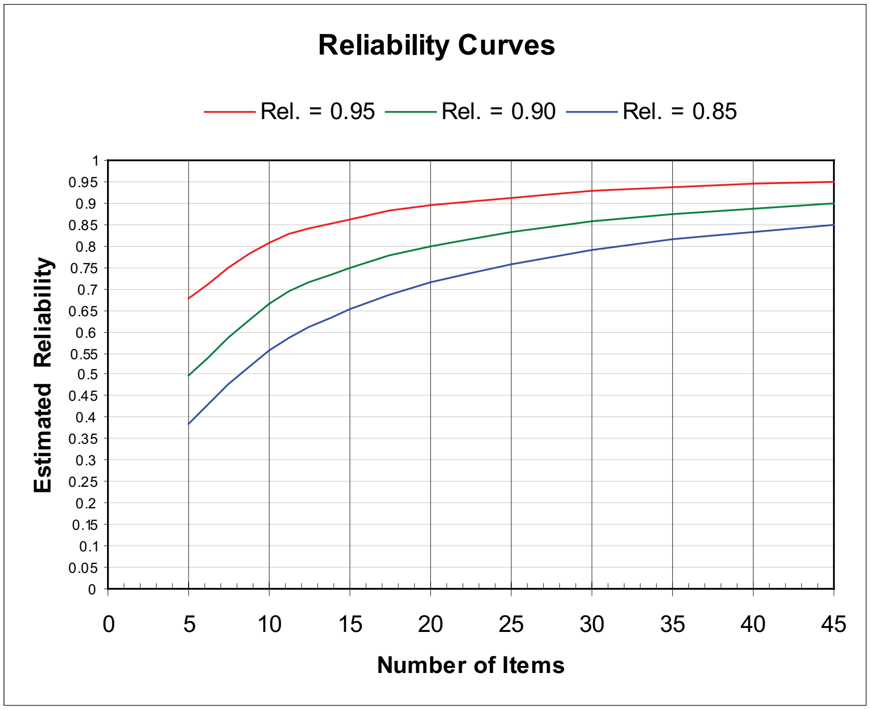
As emphasized in earlier sections, coefficient alpha only takes into account measurement error that arises from the selection of items used on a particular test form. There are other sources of random inaccuracy. One is due to the occasion of testing. Examples of other various random conditions that might affect students on any particular testing occasions include illness, fatigue, and anxiety. Also, when a test includes CR items, another source of random fluctuation can be the CR item scorers. In a sense, alpha may be positively biased because it does not take into account these other important sources of random error. Actually, any internal consistency reliability index might understate the overall problem of measurement error because they all ignore such sources of random error.

Another positive bias can occur when items are associated (clustered) with a common stimulus. Item bundles and testlets are other frequently used terms for this situation. One concrete example is when multiple reading comprehension items are associated with a common passage selection. Again, such a situation does not guarantee that the reliability estimate will be markedly affected, but the potential exists.

MODULE SCORE RELIABILITY

As noted in the introduction, reliabilities tend to be higher with an increase in test length and lower with a decrease in test length. Figure 18–1 illustrates this relationship for a hypothetical 45-point test with three total score reliabilities: 0.95, 0.90, and 0.85. As an example, the curve for reliability equal to 0.90 suggests that a 10-item module would be expected to have a score reliability of just over 0.65. The use of the Spearman-Brown prophecy formula assumes all items are exchangeable, which in practice, they may not be. While such a chart may not perfectly model actual module correlations, the intent is only to illustrate the substantial impact that limited numbers of module items can have on module score reliability.

Figure 18–1. Example of the Relationship Between Test Length and Reliability



As can be seen in Appendix M, the reliability coefficients at the module level were always lower than those at the total test level. This is more likely because the number of items at the module level is half of the number of items in the total test.

STANDARD ERROR OF MEASUREMENT

The reliability coefficient is a unit-free indicator that reflects the degree to which scores are free of measurement error. It always ranges between 0.0 and 1.0 regardless of the test's scale. Reliability coefficients best reflect the extent to which measurement inconsistencies may be present or absent in a group. However, they are not that useful for helping users interpret test scores. The standard error of measurement (SEM) is another indicator of test score precision that is better suited for determining the effect of measurement inconsistencies on the scores obtained by individual examinees. This is particularly so for conditional SEMs (CSEM) discussed further below.

TRADITIONAL STANDARD ERROR OF MEASUREMENT

A precise, theoretical interpretation of the SEM (see Appendix M) is somewhat unwieldy. A beginning point for understanding the concept is as follows. If everyone being tested had the same true score,[2] there would still be some variation in observed scores due to imperfections in the measurement process, such as random differences in attention during instruction or concentration during testing or the sampling of test items. The standard error is defined as the standard deviation[3] of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents very important information for test score users.

The SEM formula is provided below.

$$SEM = SD\sqrt{1 - reliability}$$

It indicates that the value of the SEM depends on both the reliability coefficient and the standard deviation of test scores. If the reliability were equal to 0.00 (the lowest possible value), the SEM would be equal to the standard

[2] True score is the score the person would receive if the measurement process were perfect.

[3] The standard deviation of a distribution is a measure of the dispersion of the observations. For the normal distribution, about 16 percent of the observations are more than one standard deviation above the mean.

deviation of the test scores. If test reliability were equal to 1.00 (the highest possible value), the SEM would be 0.0. In other words, a perfectly reliable test has no measurement error (Harvill, 1991). Additionally, the value of the SEM takes the group variation (i.e., score standard deviation) into account. Consider that an SEM of 3.0 on a 10-point test would be very different from an SEM of 3.0 on a 100-point test.

TRADITIONAL SEM CONFIDENCE INTERVALS

The SEM is an index of the random variability in test scores in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the precision of individual tests scores. SEMs help place reasonable limits (Gulliksen, 1950) around observed scores through construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, *X*, and adding and subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between ±1 SEM about two-thirds of the time.[4] For ±2 SEM confidence intervals, the percentage increases to about 95 percent.

FURTHER INTERPRETATIONS

ONE SEM FOR ALL TEST SCORES

The SEM approach described above only provides a single numerical estimate for constructing the confidence intervals for examinees regardless of their score levels. In reality, however, such confidence intervals vary according to one's score. Consequently, care should be taken when using the SEM for students with extreme scores. An alternate approach that conditions the SEM on a student's score estimate is described in the next sections.

GROUP SPECIFIC

As noted in the introduction, reliabilities are group specific. The same is true for SEMs because both score reliabilities and score standard deviations vary across groups.

RAW SCORE METRIC

The SEM approach is calculated using raw scores, and as such, the resulting confidence interval bands are on the raw score metric. Error bands on the scaled-score metric are considered in the next section.

TYPE OF ERROR REFLECTED

The interpretation of the SEM should be driven by the type of score reliability that underpins it. So, the Keystone Exams SEMs involve the same source of error relevant to internal consistency indices. As noted earlier, a precise technical explanation of the SEM (and resulting confidence intervals) can be unwieldy. Because of this, score users are often provided less complex interpretations.

One simpler description sometimes used is that a confidence interval represents the possible score range that one would observe if a student could be tested twice with the same instrument. Taking the same test on a different day implies the only source of random error being considered is related to the occasion of testing—such as a student might be sleepier one day than another, might be sick, or might not have eaten a good breakfast. There is a reliability index that captures this source of random error, and it is referred to as the test-retest reliability coefficient. This is not the type of reliability computed for the Keystone Exams. When internal consistency reliability estimates are used, such an explanation blurs the fact that random error based on the occasion of testing is not considered.

When SEMs are derived from internal consistency reliability estimates, a better approach is to describe the confidence interval as providing reasonable bounds for the range of scores that a student might receive if he or she took an equivalent version of the test. (That is, the student took a test that covered exactly the same content but included a different set of items.) As an example, if the Algebra I score was 1450 and the SEM band was 1435 to 1465, then a student would be likely to receive a score somewhere between 1435 and 1465 if he or she took a different version of the test.

---

[4] Some prefer the following interpretation: if a student were tested an infinite number of times, the ±1 SEM confidence intervals constructed for each score would capture the student's true score 68 percent of the time.

RESULTS AND OBSERVATIONS

Coefficient alpha results and associated (traditional) SEMs for various Keystone Exam scores are documented in Appendix M. Values were derived using the final data file (see Chapter Nine). The results are organized by administration and then content area. Each table also breaks out the modules and groups of interest such as the total student population (overall), gender, ethnicity, English language learner (ELL), students with an individualized education plan (IEP), and the economically disadvantaged (ED). The statistics reported include the number of points possible (Pts.), number of items (Len.), number of students tested (N), mean number of score points received (Mean), standard deviation of test scores (SD), reliability (r), and traditional standard error of measurement (SEM).

Note that these tables report the standard deviations of observed scores. Assuming normally distributed scores, one would expect about two-thirds of the observations to be within one standard deviation of the mean. An estimate of the standard deviation of the true scores can be computed as

$$\hat{\sigma}_T = \sqrt{\hat{\sigma}_x^2 - \hat{\sigma}_x^2 (1 - \hat{\rho}_{\boldsymbol{x}})}$$

The overall test score reliability values are high (with a value of 0.84 or higher) for Algebra I, Biology, and Literature. The reliabilities at the module level are relatively low. This is most likely due to the fact that each module contains fewer items. It was also noted that reliabilities tend to go up in value with an increase in population heterogeneity and go down in value with a decrease in more homogeneous populations. Once again, there is no firm guideline regarding how low is too low. The lower a given reliability coefficient, the greater the potential for over-interpretation. As a point of reference, a reliability coefficient of 0.50 would suggest that there is as much error variance as true-score variance in the scores. It should be noted that the reliability of group mean scores (e.g., school or district means) tends to be higher than that of individual scores, suggesting interpretation of strand scores at these aggregate levels is likely reasonable.

RASCH CONDITIONAL STANDARD ERRORS OF MEASUREMENT

The CSEM also indicates the degree of measurement error but does so in scaled-score units and varies as a function of a student's actual scaled score. Therefore, the CSEM may be especially useful in characterizing measurement precision in the neighborhood of a score level used for decision making—such as cut scores for identifying students who meet a performance standard.

Technically, when a Rasch model is applied, the CSEM at any given point on the ability continuum is defined as the reciprocal of the square root of the test information function derived from the Rasch scaling model:

$$CSEM(\hat{\beta}_n) = \frac{1}{\sqrt{I(\hat{\beta}_n)}}$$

where $CSEM(\hat{\beta}_n)$ is conditional standard error of measurement and $I(\hat{\beta}_n)$ is test information function. Test information depends on the sum of the corresponding information functions for the test items. Item information depends on each item's difficulty and conditional item score variance. The formula above utilizes the Rasch ability β_n metric. The conditional standard error on the scaled-score (SS) metric is determined simply by multiplying the $CSEM(\hat{\beta}_n)$ by the slope (multiplicative constant, *m*) of the linear transformation equation used to convert the Rasch ability estimates to scaled scores.

$$CSEM(\mathrm{SS}) = CSEM(\hat{\beta}_n) * m$$

Chapter Fourteen provides the linear transformation formulas for each of the Keystone Exams.

RASCH CSEM CONFIDENCE INTERVALS

CSEMs also allow statements regarding the precision of individual tests scores. And like SEMs, they help place reasonable limits around observed scaled scores through construction of an approximate score band. The confidence intervals are constructed by adding and subtracting a multiplicative factor of the CSEM and may be interpreted as described in the earlier section.

FURTHER INTERPRETATIONS

DIFFERENT CSEMS FOR DIFFERENT TEST SCORES

The CSEM approach provides different numerical estimates for constructing the confidence intervals for examinees depending on their specific score levels. The magnitude of the CSEM values is U-shaped, with larger CSEM values associated with lower and higher scores.

GROUP SPECIFIC

Assuming reasonable model-data fit—as explored in Chapter Twelve—the Rasch-based CSEMs (conditioned on score level) should not vary across groups.

SCALED-SCORE METRIC

The CSEM and associated confidence interval bands are on the scaled-score metric.

TYPE OF ERROR REFLECTED

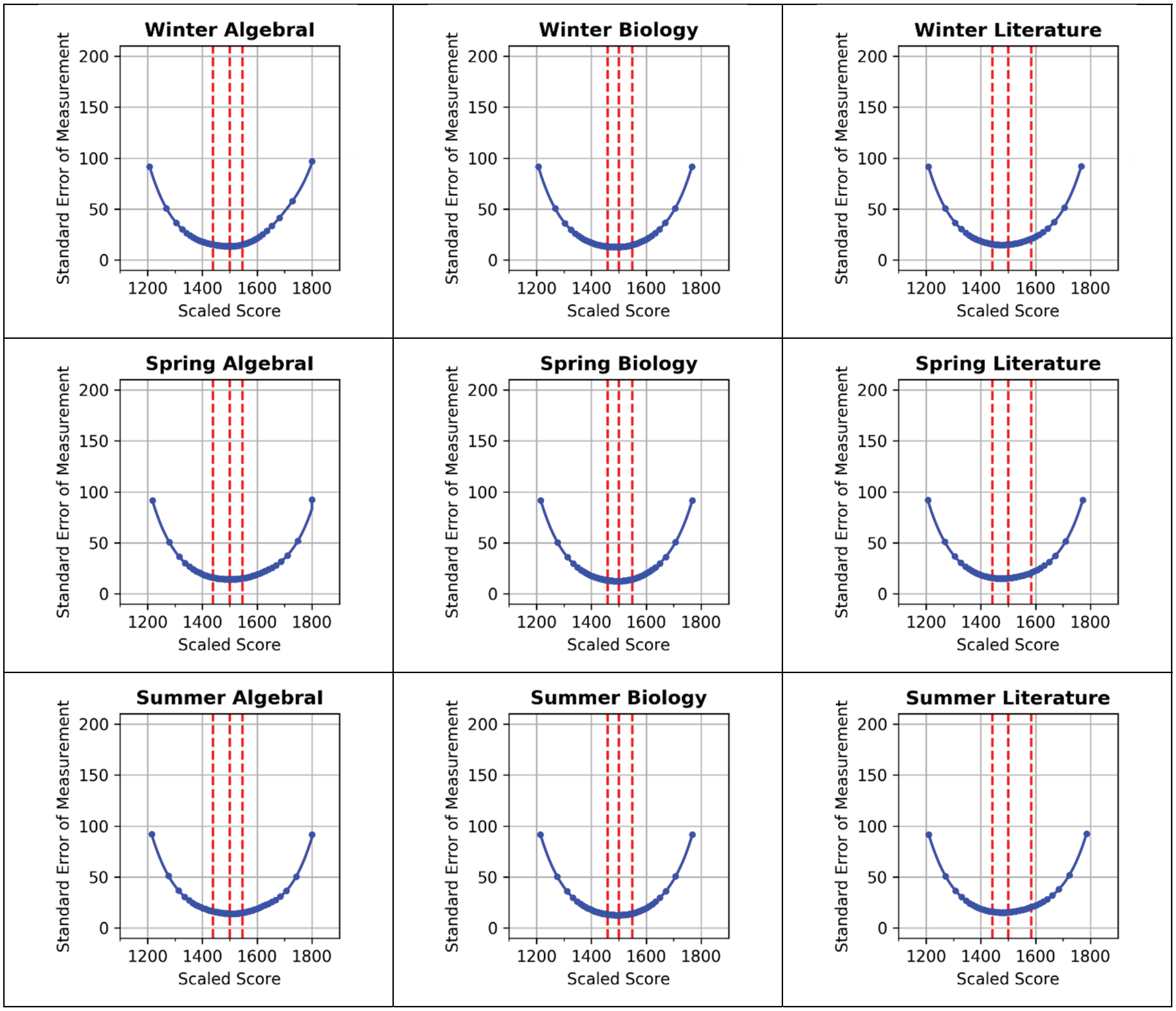
The CSEMs documented on the Keystone Exams score reports are the Rasch-based conditional standard errors of measurement described above. These are provided by the program WINSTEPS described in Chapter Twelve. As noted earlier, these CSEMs are based on the concept of statistical information. For the purpose of providing a simpler explanation of CSEMs to test score users, the earlier description of SEMs framed using the idea of internal consistency reliability was provided in the Keystone Exams score report interpretive guide.[5] Score report content is considered in greater detail in Chapter Sixteen.

RESULTS AND OBSERVATIONS

Figure 18–2 shows the Rasch CSEMs associated with each scaled-score level. (This information is also provided in tabular form in Appendix K.) Values were derived using the calibration data file described in Chapter Ten. The values are fairly consistent across a noticeably large range of the scaled scores, as demonstrated by the relatively flat bottoms of most plots. The values increase at both extremes (i.e., at smaller and larger scaled scores) giving these figures their typical U-shaped pattern. The three red-dashed lines represent the Basic, Proficient, and Advanced scaled score cuts, respectively, moving from lower to higher scaled-score values. CSEM values at the cut score lines are associated with smaller values, indicating more precise measurement occurs at these cuts.

[5] Because Rasch CSEMs are based on statistical information, it is questionable whether they account for error variance due to items. However, it seems difficult to construct a simple explanation of Rasch CSEMs for the general public.

Figure 18–2. Conditional Standard Error Plots for Each Administration and Content Area



RELIABILITY OF PERFORMANCE LEVEL CLASSIFICATION DECISIONS

Student performance on the Keystone Exams is classified into one of four achievement levels using the cut scores described in Chapter Thirteen. The reliability of the classification decisions can be assessed by two statistics: decision accuracy and decision consistency.

DECISION ACCURACY

Decision accuracy describes the extent to which performance level classification decisions based on the administered test form would agree with the decisions that would be made on the basis of a perfectly reliable test (i.e., if it was possible to know each examinee's true score). Decision accuracy answers the question: How does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known?

DECISION CONSISTENCY

Decision consistency describes the extent to which classification decisions based on the administered test form would agree with the decisions made if a parallel alternate form had been administered. Decision consistency answers the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?

Since the true scores are unknown and it is not feasible to repeat the Keystone Exams in order to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model needs to be imposed on the data in order to project the consistency of classifications solely using data from the available administration (Hambleton and Novick, 1973). Although a number of procedures are available, two well-known methods were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995), utilizing specific true score models. These approaches are fairly complex, and the cited sources contain details regarding the statistical models used to calculate the decision accuracy and consistency from a single administration.

For Keystone Exams, given that the two approaches provide similar results, true scores and single-form scores on forms parallel to the one actually given are estimated following the Livingston and Lewis (1995) method. The decision accuracy is estimated using an estimated joint distribution of reported performance-level classifications on the current form of the exam and the performance-level classifications based on the true score. Decision consistency is estimated using an estimated joint distribution of reported performance-level classifications on the current form of the exam and performance-level classifications on the parallel alternate form. In each case, the proportion of performance-level classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the joint distribution. Reliability of classification at each performance-level cut score is estimated by collapsing the joint distribution at the passing score boundary into a 2-by-2 table and summing the two entries.

Several factors might affect the classification decision accuracy and consistency. One important factor is the reliability of the scores. All other things being equal, more reliable test scores tend to result in more similar reclassifications. Another factor is the location of the cut score in the score distribution. More consistent classifications are observed when the cut scores are located away from the mass of the score distribution. For example, when scores are close to being normally distributed, the mass is concentrated in the middle of the distribution, and thus, classifications tend to become more consistent when cut scores go up from 70 percent to 80 percent, or, alternatively, go down from 30 percent to 20 percent. The number of performance levels is also a consideration. Consistency indices for four performance levels should be lower than for those based on two categories. This is not surprising since classification using four levels would allow more opportunity to change achievement levels. Hence, there would be more classification errors with four achievement levels, resulting in lower consistency indices.

The results—derived using the program *BB-Class* (Brennan, 2004)—for the overall accuracy and consistency across all four performance levels as well as for the dichotomies created by the three cut scores are presented in Table 18–4.

Across all administrations and content areas, the overall decision accuracy ranged from 0.75 to 0.82 and the decision consistency ranged from 0.66 to 0.75. Dichotomous decisions have the higher accuracy and consistency values than the overall. The decision accuracy of the Basic/Proficient cut scores ranged from 0.89 to 0.95 and the decision consistency ranged from 0.85 to 0.93. These results indicate that at least 89% of students meeting or exceeding the Proficient cut score would receive the same classification if their true scores were known. If a parallel test were administered, at least 85% or more of students meeting or exceeding the Proficient cut score would be classified in the same way.

Table 18–4. Reliability of Performance-Level Classification Decisions

Administration	Content Area	Statistics	Overall	Below Basic/ Basic	Basic/Proficient	Proficient/ Advanced
Winter	Algebra I	Accuracy	0.80	0.98	0.89	0.93
Winter	Algebra I	Consistency	0.72	0.97	0.85	0.90
Winter	Biology	Accuracy	0.82	0.97	0.91	0.94
Winter	Biology	Consistency	0.75	0.96	0.87	0.92
Winter	Literature	Accuracy	0.81	0.97	0.93	0.92
Winter	Literature	Consistency	0.74	0.96	0.90	0.88
Spring	Algebra I	Accuracy	0.78	0.95	0.91	0.92
Spring	Algebra I	Consistency	0.69	0.93	0.88	0.89
Spring	Biology	Accuracy	0.80	0.95	0.93	0.93
Spring	Biology	Consistency	0.72	0.92	0.90	0.90
Spring	Literature	Accuracy	0.80	0.93	0.95	0.92
Spring	Literature	Consistency	0.73	0.91	0.93	0.89
Summer	Algebra I	Accuracy	0.79	0.97	0.93	0.89
Summer	Algebra I	Consistency	0.71	0.96	0.90	0.84
Summer	Biology	Accuracy	0.75	0.97	0.89	0.90
Summer	Biology	Consistency	0.66	0.95	0.85	0.86
Summer	Literature	Accuracy	0.80	0.98	0.95	0.87
Summer	Literature	Consistency	0.72	0.98	0.92	0.81

RATER AGREEMENT

Because CR items are included on the Keystone Exams, another source of random error is related to the scorers of those items. Frisbie (2005) noted that “test score reliability differs from scorer reliability” and that “the need for one kind of estimate cannot be satisfied by the other.” Additionally, the data most easily obtainable that captures this information comes from the “10 percent read behinds” collected during the scoring process. Partly because of the way these data are obtained and reported (i.e., it’s **not** a ratio of true score variance over observed score variance), the term *rater agreement* is used here, not *rater reliability* or *inter-rater reliability* as these terms are somewhat misleading.

The rater agreements for the Keystone Exams are presented in Tables 18–5 to 18–7. In addition, the percentages awarded to each score point are also presented in these tables. As the table shows, the exact inter-rater agreement percentages ranged from 73 to 100 percent. Overall, Algebra I has the highest exact agreements while Literature has the lowest exact agreement. The percentages of exact and adjacent agreement for all content areas are 100 or close to 100.

Table 18–5. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Winter

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/ NS
Algebra 1	1A	99	1	100	65	1	24	NA	NA	9
Algebra 1	1B	99	1	100	46	15	29	NA	NA	9
Algebra 1	2	93	7	100	35	40	6	3	1	16
Algebra 1	3A	97	3	100	55	14	17	NA	NA	13
Algebra 1	3B	100	0	100	71	16	NA	NA	NA	13
Algebra 1	3C	100	0	100	38	49	NA	NA	NA	13
Algebra 1	4A	99	1	100	10	81	NA	NA	NA	8
Algebra 1	4B	97	3	100	47	44	NA	NA	NA	8
Algebra 1	4C	99	1	100	74	17	NA	NA	NA	8
Algebra 1	4D	97	3	100	62	29	NA	NA	NA	8
Algebra 1	5A	99	1	100	38	52	NA	NA	NA	10
Algebra 1	5B	100	0	100	78	12	NA	NA	NA	10
Algebra 1	5C	100	0	100	89	1	NA	NA	NA	10
Algebra 1	5D	100	0	100	81	9	NA	NA	NA	10
Algebra 1	5MU	100	0	100	90	0	NA	NA	NA	10
Algebra 1	6A	99	1	100	55	33	NA	NA	NA	12
Algebra 1	6B	100	0	100	74	14	NA	NA	NA	12
Algebra 1	6C	100	0	100	77	10	NA	NA	NA	12
Algebra 1	6D	98	2	100	72	16	NA	NA	NA	12
Biology	1	87	13	100	21	32	25	8	NA	14
Biology	2	91	9	100	58	14	8	5	NA	15
Biology	3	91	9	100	43	20	15	6	NA	16
Biology	4	82	17	99	22	34	19	10	NA	15
Biology	5	85	15	100	37	28	14	4	NA	13
Biology	6	87	12	99	19	40	20	9	NA	12
Literature	1	82	18	100	24	33	26	5	NA	13
Literature	2	89	11	100	12	32	37	4	NA	15
Literature	3	85	15	100	14	34	27	8	NA	17
Literature	4	81	19	100	15	19	37	16	NA	14
Literature	5	85	15	100	4	31	43	8	NA	13
Literature	6	89	11	100	9	35	32	5	NA	19

Note: Some of the Algebra I CR items were scored by part. For example, 1A in the second column means part A of item 1. B/NS in the last column represents blank/non-scorable. NA means not applicable.

Table 18–6. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Spring

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/NS
Algebra 1	1A	98	2	100	27	66	NA	NA	NA	7
Algebra 1	1B	99	1	100	28	65	NA	NA	NA	7
Algebra 1	1C	100	0	100	90	3	NA	NA	NA	7
Algebra 1	1D	99	1	100	44	49	NA	NA	NA	7
Algebra 1	2A	99	1	100	68	23	NA	NA	NA	9
Algebra 1	2B	99	1	100	81	10	NA	NA	NA	9
Algebra 1	2C	100	0	100	69	21	NA	NA	NA	9
Algebra 1	2D	100	0	100	72	18	NA	NA	NA	9
Algebra 1	3	88	12	100	23	28	12	9	15	12
Algebra 1	4	84	16	100	11	25	24	22	10	8
Algebra 1	5A	99	1	100	36	56	NA	NA	NA	8
Algebra 1	5B	99	1	100	53	39	NA	NA	NA	8
Algebra 1	5C	100	0	100	85	8	NA	NA	NA	8
Algebra 1	5D	100	0	100	88	4	NA	NA	NA	8
Algebra 1	6	90	10	100	16	20	42	10	1	10
Biology	1	88	10	98	40	11	21	19	NA	9
Biology	2	83	16	99	29	25	21	14	NA	11
Biology	3	76	23	99	23	29	28	9	NA	12
Biology	4	96	4	100	29	21	27	14	NA	9
Biology	5	86	14	100	16	29	33	12	NA	9
Biology	6	89	11	100	28	21	18	16	NA	17
Literature	1	77	23	100	12	30	40	9	NA	10
Literature	2	80	20	100	24	20	31	14	NA	11
Literature	3	79	20	99	10	22	42	13	NA	12
Literature	4	76	24	100	8	24	44	13	NA	11
Literature	5	84	15	99	9	24	48	7	NA	12
Literature	6	81	19	100	11	30	36	10	NA	13

Note: Some of the Algebra I CR items were scored by part. For example, 1A in the second column means part A of item 1. B/NS in the last column represents blank/non-scorable. NA means not applicable.

Table 18–7. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Summer

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/NS
Algebra 1	1	92	8	100	14	44	18	17	3	3
Algebra 1	2A	99	1	100	18	79	NA	NA	NA	3
Algebra 1	2B	100	0	100	79	18	NA	NA	NA	3
Algebra 1	2C	100	0	100	96	1	NA	NA	NA	3
Algebra 1	2D	100	0	100	96	1	NA	NA	NA	3
Algebra 1	3	91	9	100	43	42	9	1	1	3
Algebra 1	4	98	2	100	10	34	26	27	0	3
Algebra 1	5A	99	1	100	89	8	NA	NA	NA	3
Algebra 1	5B	100	0	100	43	53	NA	NA	NA	3
Algebra 1	5C	99	1	100	85	12	NA	NA	NA	3
Algebra 1	5D	100	0	100	72	25	NA	NA	NA	3
Algebra 1	6	93	7	100	20	43	22	8	2	5
Biology	1	87	12	99	25	45	22	3	NA	4
Biology	2	96	4	100	31	42	20	4	NA	3
Biology	3	92	8	100	24	29	27	16	NA	3
Biology	4	97	3	100	35	37	20	4	NA	5
Biology	5	84	15	99	20	32	26	18	NA	4
Biology	6	81	19	100	34	44	12	6	NA	4
Literature	1	81	19	100	8	33	44	11	NA	4
Literature	2	78	22	100	14	50	30	2	NA	3
Literature	3	87	13	100	13	55	24	2	NA	6
Literature	4	73	27	100	13	39	40	4	NA	3
Literature	5	79	21	100	14	50	31	1	NA	3
Literature	6	85	15	100	8	53	32	3	NA	4

Note: Some of the Algebra I CR items were scored by part. For example, 1A in the second column means part A of item 1. B/NS in the last column represents blank/non-scorable. NA means not applicable.

CHAPTER NINETEEN: VALIDITY

As defined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999, referred to as the *Standards*), validity is "the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (p. 9). The *Standards* provides a framework for describing the sources of evidence that should be considered when evaluating validity. These sources include evidence based on test content, response processes, the internal structure of the test, the relationships between test scores and other variables, and the consequences of testing. In addition, when Rasch models are used to analyze assessment data, validity considerations related to those processes should also be explored.

The validity process involves the collection of a variety of evidence to support the proposed test score interpretations and uses. The entire technical report describes the technical aspects of the Keystone Exams in support of their score interpretations and uses. Each of the previous chapters contributes important evidence components that pertain to score validation: test development, test administration, test scoring, item analysis, Rasch calibration, scaling, equating, score reporting, and reliability. This chapter summarizes and synthesizes the evidence based on the framework of the *Standards*. The purposes and intended uses of the Keystone Exams are reviewed first, and then each type of validity evidence is addressed in turn.

PURPOSES AND INTENDED USES OF THE KEYSTONE EXAMS

The *Standards* emphasize that validity pertains to how test scores are used. To help contextualize the evidence that will be presented below, the purposes of the Pennsylvania Keystone Exams will be reviewed first. The Keystone Exams, which began in 2010–2011 for Algebra I, Biology, and Literature, are one component of Pennsylvania's new system of high school graduation requirements for students in the class of 2019 and beyond. Students take the exams toward the end of specific courses. The Keystone Exams results help school districts guide students toward meeting state standards. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module, provided that they meet the requirements detailed below.

- The student has taken the course.
- The student was unsuccessful in achieving a score of Proficient or Advanced on the Keystone Exam after at least two attempts.
- The student has met the district's attendance requirements for the course.
- The student has participated in a satisfactory manner in supplemental instructional services.

EVIDENCE BASED ON TEST CONTENT

Test content validity evidence for the Keystone Exams rests greatly on establishing a link between each piece of the assessment (i.e., the items) and what students should know and be able to do as prescribed by the Keystone Exams Assessment Anchors and Eligible Content. The Keystone Exams are intended to measure the knowledge and skills described in the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.

Lane (1999) suggests taking the following steps to support the content validity of an assessment. In the case of Keystone Exams, one should

- Evaluate the degree to which the test specifications represent and align with the knowledge and skills described in the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.
- Evaluate the alignment between the Keystone Exams items and test specifications to ensure representativeness.
- Evaluate the extent to which the curriculum aligns with the Assessment Anchors and Eligible Content.
- Conduct content reviews of the Keystone Exams items using a panel of content experts to see whether items measure the intended construct or are the sources of construct-irrelevant variance.
- Conduct fairness reviews of the items to avoid issues related to a specific subpopulation.

- Evaluate procedures for administration and scoring such as the appropriateness of instructions to examinees, time limit for the assessment, and training of raters.
- Submit operational tests to third-party independent reviews.

Chapters Two through Eight of this report present a considerable amount of evidence related to test content. As described in these chapters, all the items were developed and aligned with the Keystone Exams Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature. After development, items underwent multiple rounds of content and bias reviews. After being field tested, they were reviewed with respect to their statistical properties. Items selected for the operational assessment had to pass content, psychometric, and PDE reviews. Tests were administrated according to standardized procedures with allowable accommodations.

Some of the efforts made to ensure content validity are summarized below.

- DRC used Webb's (1999) Depth of Knowledge (DOK) model to ensure the Keystone Exams items aligned with the Assessment Anchors and Eligible Content and the Academic Content Standards in terms of both content and cognitive levels.
- DRC established detailed test and item/passage development specifications and ensured the items were sufficient in number and adequately distributed across content, levels of cognitive complexity, and levels of difficulty.
- DRC selected qualified item writers and provided training to help ensure they wrote high-quality items.
- All newly developed items were first reviewed by content specialists and editors at DRC to make sure they measured the intended Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature. Appropriateness for the intended students was also considered, as well as DOK, graphics, grammar/punctuation, language demand, and distractor reasonableness.
- Prior to field testing, the test items were submitted to content committees (composed of Pennsylvania educators) for review using, but not limited to, the following categories:
    - Overall quality and clarity
    - Anchor, Eligible Content, and/or standard alignment
    - Grade-level appropriateness
    - Difficulty level
    - DOK
    - Appropriate sources of challenge (e.g., unintended content and skills)
    - Correct answer
    - Quality of distractors
    - Graphics
    - Appropriate language demand
    - Freedom from bias
- The items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. This committee reviewed items for issues related to diversity, gender, and other pertinent factors.
- Items passing all the prior hurdles were tried out as embedded field test items in the operational test. Several statistical analyses were conducted on the field test data including classical item analyses, distractor analyses, and differential item functioning (DIF). Items were again carefully reviewed by DRC staff and a committee of Pennsylvania teachers with respect to their statistical characteristics. DIF was used to detect test items that might bias test scores for particular groups. Empirical investigation of DIF strengthens the validity evidence related to score interpretations for students in particular groups by eliminating potential sources of construct-irrelevant variance.

- The Keystone Exams were administered according to standardized procedures with allowable accommodations. Students were given ample time to complete the tests (i.e., there were no speediness issues).
- As described in Chapter Eight, the raters for constructed-response (CR) items were carefully recruited and well trained. Their scoring was monitored throughout the scoring session to ensure that an acceptable level of scoring accuracy was maintained.

EVIDENCE BASED ON RESPONSE PROCESS

Response-process evidence is used to examine the extent to which the cognitive skills and processes employed by students match those identified in the test developer's defined construct domains for all students and for each subgroup. Think-aloud procedures or cognitive labs can be used to collect this type of evidence. In addition, when an assessment includes CR items, an examination of the extent to which the raters interpret and apply the scoring criteria accurately when assigning scores to students' responses on CR items also adds response-process validity evidence.

For the operational Keystone Exams offered in winter, spring, and summer, no cognitive lab studies were conducted to collect the response-process evidence. Rather, for all the Keystone Exams, well-organized scorer training and subsequent monitoring of rating accuracy helped ensure that raters strictly followed the scoring criteria and that no features unrelated to the rubric significantly affected their scoring.

EVIDENCE BASED ON INTERNAL STRUCTURE

As described in the *Standards* (1999), internal-structure evidence refers to the degree to which the relationships between test items and test components conform to the construct on which the proposed test interpretations are based. For each Keystone Exam, one total test score as well as module scores were reported (see Chapter Sixteen for more information about the Keystone Exams scores). Several dimensionality studies were conducted in order to provide internal-structure evidence relating to the use of both types of scores.

ITEM-TEST CORRELATIONS

Item-test correlations are provided and discussed in Chapter Eleven. All values were positive and of acceptable magnitude.

DIFFERENTIAL ITEM FUNCTIONING (DIF)

DIF analyses with respect to gender, ethnicity, and test administration mode help address construct-irrelevant variance, which represents an important threat to the validity of achievement tests. As noted in Chapter Five, field test items were screened and reviewed for DIF. Only items approved by teacher committees were eligible for operational use. DIF analyses were conducted on the operational items again to monitor the bias code change. As can be seen in Table 19–1, there were a very few items where the bias code changed from A/B (i.e., A+, A-, B+, and B-) to C (i.e., C+ or C-). Given that most items selected to build the operational forms had no C-level DIF, the Keystone Exams can be considered valid from this perspective.

Table 19–1. Summary of Bias Code Change from Field Test to Operational Test

Administration	Content Area	Change	Male/ Female	White/ Black	PPT/CBT
Winter	Algebra I	C →A/B	0	0	0
Winter	Algebra I	A/B → C	0	0	1
Winter	Biology	C → A/B	0	0	0
Winter	Biology	A/B → C	0	0	0
Winter	Literature	C → A/B	1	0	0
Winter	Literature	A/B → C	0	0	0
Spring	Algebra I	C → A/B	0	0	0
Spring	Algebra I	A/B → C	0	0	0
Spring	Biology	C → A/B	0	1	0
Spring	Biology	A/B → C	0	0	0
Spring	Literature	C → A/B	1	0	0
Spring	Literature	A/B → C	0	0	0
Summer	Algebra I	C → A/B	0	0	0
Summer	Algebra I	A/B → C	0	1	0
Summer	Biology	C → A/B	0	0	0
Summer	Biology	A/B → C	0	0	0
Summer	Literature	C → A/B	0	0	0
Summer	Literature	A/B → C	1	2	3

Note: PPT represents the paper-and-pencil-based test, and CBT represents the computer-based test.

DIMENSIONALITY

Dimensionality analyses were conducted for the winter, spring, and summer Keystone Exams using WINSTEPS's principle components analyses on response residuals for each content area. Results are shown in Chapter Twelve. The principal component analysis results provided evidence that each of the three Keystone Exams was essentially unidimensional, supporting the validity of using the total scores to estimate student's overall ability in each subject area.

MODULE CORRELATIONS

Correlations and disattenuated correlations among module scores for the Keystone Exams are presented below. Values were derived from the Keystone Exams final data files (see Chapter Nine). These data can also provide information on score dimensionality that is part of internal-structure evidence. All Keystone Exams have two modules. The intercorrelations between the modules within the content areas were positive and ranged from 0.65 to 0.88. The intercorrelations between modules in different content areas ranged from 0.43 to 0.75, which were relatively small as expected.

Table 19–2. Correlations among Algebra I, Biology, and Literature Modules

Administration	Content Area	Module	Algebra I Module 1	Algebra I Module 2	Biology Module 1	Biology Module 2	Literature Module 1	Literature Module 2
Winter	Algebra I	Module 1	-					
Winter	Algebra I	Module 2	**0.79**	-				
Winter	Biology	Module 1	0.51	0.51	-			
Winter	Biology	Module 2	0.55	0.54	**0.83**	-		
Winter	Literature	Module 1	0.52	0.51	0.65	0.68	-	
Winter	Literature	Module 2	0.54	0.53	0.65	0.69	**0.82**	-
Spring	Algebra I	Module 1	-					
Spring	Algebra I	Module 2	**0.85**	-				
Spring	Biology	Module 1	0.63	0.65	-			
Spring	Biology	Module 2	0.66	0.68	**0.86**	-		
Spring	Literature	Module 1	0.55	0.57	0.72	0.75	-	
Spring	Literature	Module 2	0.55	0.57	0.72	0.75	**0.84**	-
Summer	Algebra I	Module 1	-					
Summer	Algebra I	Module 2	**0.69**	-				
Summer	Biology	Module 1	0.40	0.43	-			
Summer	Biology	Module 2	0.53	0.54	**0.71**	-		
Summer	Literature	Module 1	0.49	0.45	0.51	0.64	-	
Summer	Literature	Module 2	0.51	0.57	0.56	0.68	**0.69**	-

The correlations in Table 19–2 are based on the observed module scores. These observed-score correlations are weakened by existing measurement error contained within each module. As a result, disattenuated correlations could provide an estimate of the relationships among modules if there were no measurement error. (An important caveat is explained further below.) The disattenuated correlation coefficients R_{12} can be computed by using the formula (Spearman, 1904; Spearman, 1910) below:

$$R_{12} = \frac{r_{12}}{\sqrt{r_{11} r_{22}}},$$

Where r_{12} is the observed correlation, and r_{11} and r_{22} are the reliabilities for Module 1 and Module 2. Disattenuated correlations very near 1.00 suggest that the same or very similar constructs are being measured. Values somewhat less than 1.00 suggest that different modules are measuring slightly different aspects of the same construct. Values markedly less than 1.00 suggest the modules reflect different constructs.

Table 19–3 shows the corresponding disattenuated correlations for each Keystone Exam. Given that none of these modules had perfect reliabilities (see Chapter Eighteen), the disattenuated module correlations are higher than their observed score counterparts.

Table 19–3. Disattenuated Correlations among Algebra I, Biology, and Literature Modules

Administration	Content Area	Module	Algebra I Module 1	Algebra I Module 2	Biology Module 1	Biology Module 2	Literature Module 1	Literature Module 2
Winter	Algebra I	Module 1	-					
Winter	Algebra I	Module 2	**0.99**	-				
Winter	Biology	Module 1	0.63	0.62	-			
Winter	Biology	Module 2	0.67	0.66	**0.97**	-		
Winter	Literature	Module 1	0.64	0.62	0.78	0.80	-	
Winter	Literature	Module 2	0.66	0.65	0.78	0.81	**0.97**	-
Spring	Algebra I	Module 1	-					
Spring	Algebra I	Module 2	**1.02**	-				
Spring	Biology	Module 1	0.75	0.76	-			
Spring	Biology	Module 2	0.78	0.79	**1.00**	-		
Spring	Literature	Module 1	0.66	0.67	0.84	0.87	-	
Spring	Literature	Module 2	0.66	0.67	0.84	0.87	**1.00**	-
Summer	Algebra I	Module 1	-					
Summer	Algebra I	Module 2	**0.98**	-				
Summer	Biology	Module 1	0.54	0.60	-			
Summer	Biology	Module 2	0.70	0.74	**0.93**	-		
Summer	Literature	Module 1	0.71	0.67	0.72	0.90	-	
Summer	Literature	Module 2	0.69	0.80	0.74	0.89	**0.98**	-

The within-content-area correlations were high (e.g., above 0.93), suggesting that the within-content-area modules might be measuring essentially the same construct. This, in turn, suggests that the within-content-area module scores might not provide unique information about the strengths or weaknesses of many of the students.

On a fairly consistent basis, the correlations among the modules within each content area are higher than the correlations among modules across different content areas. In general, within-content-area module correlations are at or higher than 0.93, while across-content-area module correlations range from 0.54 to 0.90.

It should be noted that some caution is needed when interpreting the disattenuated results because the reliabilities used to calculate the disattenuated correlations are subject to both upward and downward biases. Consequently, some of the values in the table above may be higher or lower than they should be, depending on which bias prevails for any given pair of module scores. When the reliabilities are lower than they should be, the disattenuated correlations will be inflated and in some instances can appear higher than the theoretical correlation maximum value of 1.00.

EXPLORATORY FACTOR ANALYSIS

In order to further explore the internal structure of the Keystone Exams, an exploratory factor analysis (EFA) of the module scores across all the Keystone Exams content areas was conducted. The Keystone Exams final data file (see Chapter Nine) was used to create the observed correlation matrices shown in Table 19–2, which, in turn, were used in the EFA. In the Statistical Package for the Social Sciences (SPSS), Principle Axis Factor extraction was utilized with an oblique rotation (Promax) of the initial factor solution to improve interpretability. Oblique rotations allow for correlated factors, which seemed more appropriate for the Keystone Exams because of a priori expectations that academic achievement across subject areas should be correlated.

Table 19–4 presents the eigenvalues and the explained variance for the extracted factors for the Keystone Exams in Algebra I, Biology, and Literature. The scree plots of the eigenvalues and the first six factors can be found in Figure

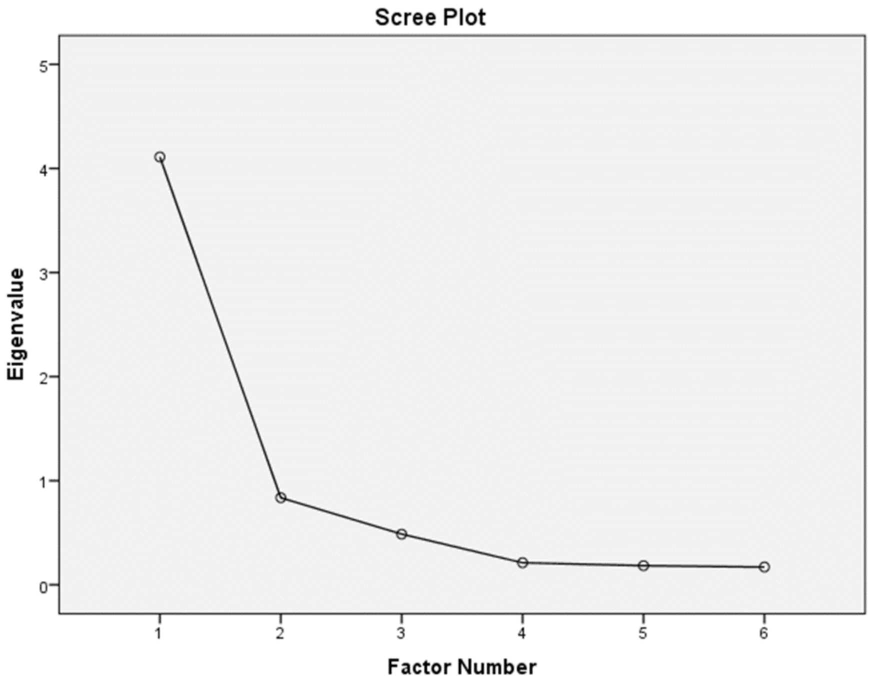
19–1. The first factor accounts for 63.44 to 74.35 percent of the total variance, while the second factor explains 12.45 to 13.95 percent of the total variance. Only the first factor had an eigenvalue greater than 1.0, typically suggesting a one-factor solution using the Kaiser criterion. However, based on the belief that there should be three distinct factors (one for each content area), a three-factor solution was further explored.

Table 19–4. Eigenvalues and Explained Variance for Algebra I, Biology, and Literature Modules

Administration	Factor	Eigenvalue	Variance Explained (%)
Winter	1	4.11	68.51
Winter	2	0.84	13.95
Winter	3	0.49	8.11
Winter	4	0.21	3.53
Winter	5	0.18	3.06
Winter	6	0.17	2.85
Spring	1	4.46	74.35
Spring	2	0.75	12.45
Spring	3	0.36	5.96
Spring	4	0.16	2.59
Spring	5	0.15	2.42
Spring	6	0.13	2.23
Summer	1	3.81	63.44
Summer	2	0.80	13.32
Summer	3	0.53	8.89
Summer	4	0.35	5.85
Summer	5	0.26	4.32
Summer	6	0.25	4.18

Figure 19–1. Scree Plot for Algebra I, Biology, and Literature Modules

Winter



Spring

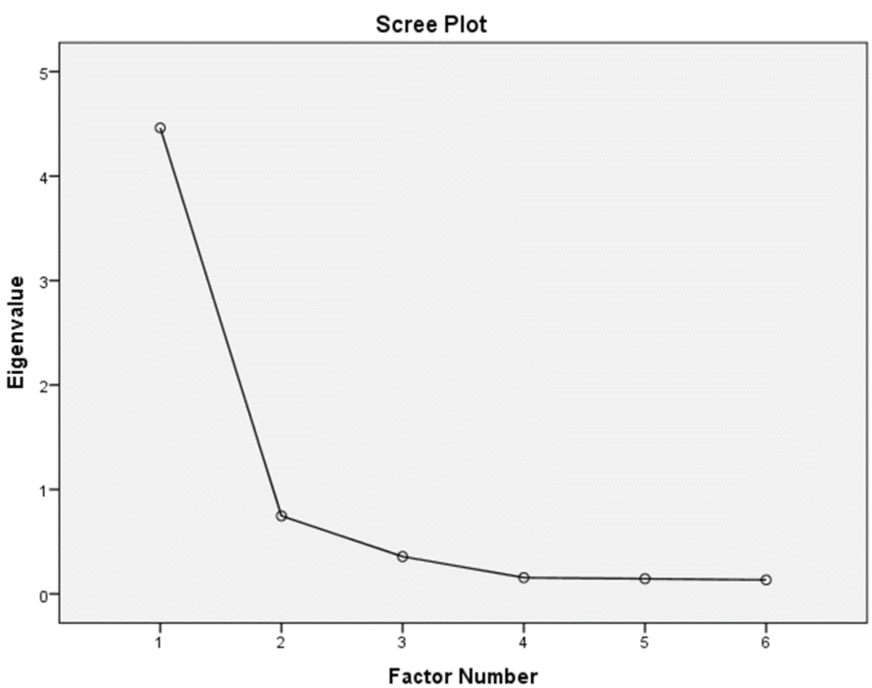
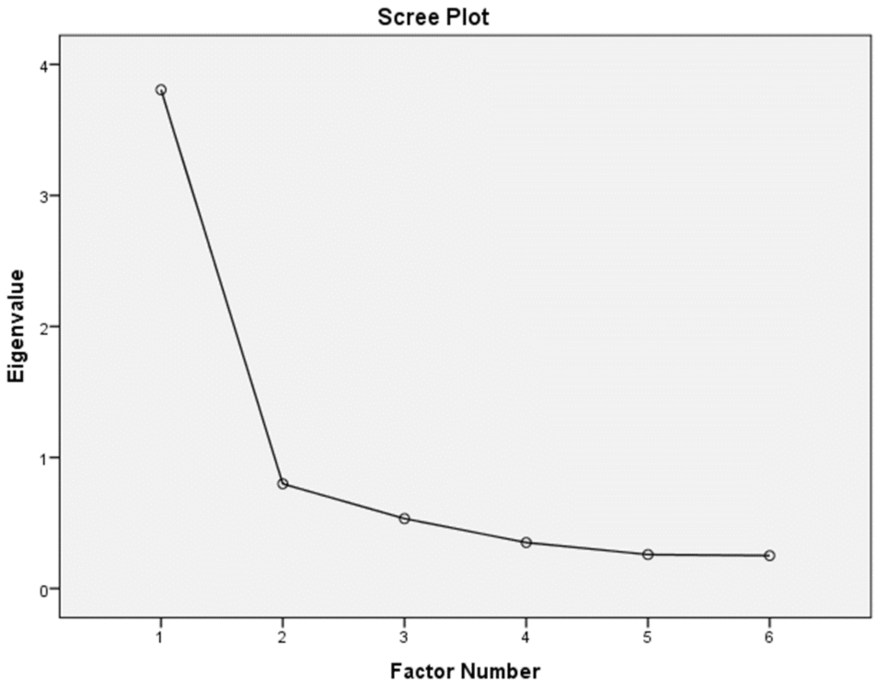


Figure 19–1 (continued). Scree Plot for Algebra I, Biology, and Literature Modules

Summer



The pattern loadings resulting from the three-factor solution are presented in Table 19–5. The pattern loadings have *simple structure*, which shows that the two Algebra I, Biology, and Literature modules clearly loaded on three different factors. The respective factor loadings were quite high. The factor correlation matrix showed that the correlations among the three latent factors are very close to the observed correlations (see Table 19–5) but lower than the disattenuated correlations.

Table 19–5. Pattern Matrix and Factor Correlation

Administration	Content Area	Module	Factor 1	Factor 2	Factor 3	Correlation
Winter	Algebra I	Module 1	**0.87**	0.01	0.02	Correlation (F1, F2) = 0.63
Winter	Algebra I	Module 2	**0.88**	0.01	0.00	Correlation (F1, F2) = 0.63
Winter	Biology	Module 1	-0.01	**0.87**	0.04	Correlation (F1, F3) = 0.63
Winter	Biology	Module 2	0.03	**0.85**	0.06	Correlation (F1, F3) = 0.63
Winter	Literature	Module 1	-0.01	0.06	**0.86**	Correlation (F2, F3) = 0.77
Winter	Literature	Module 2	0.03	0.04	**0.85**	Correlation (F2, F3) = 0.77
Spring	Algebra I	Module 1	0.01	**0.91**	0.01	Correlation (F1, F2) = 0.63
Spring	Algebra I	Module 2	0.02	**0.90**	0.03	Correlation (F1, F2) = 0.63
Spring	Biology	Module 1	0.09	0.03	**0.82**	Correlation (F1, F3) = 0.80
Spring	Biology	Module 2	0.12	0.06	**0.79**	Correlation (F1, F3) = 0.80
Spring	Literature	Module 1	**0.87**	0.01	0.04	Correlation (F2, F3) = 0.73
Spring	Literature	Module 2	**0.87**	0.01	0.05	Correlation (F2, F3) = 0.73
Summer	Algebra I	Module 1	**0.73**	-0.03	0.11	Correlation (F1, F2) = 0.69
Summer	Algebra I	Module 2	**0.94**	0.02	-0.08	Correlation (F1, F2) = 0.69
Summer	Biology	Module 1	-0.04	**0.88**	-0.06	Correlation (F1, F3) = 0.67
Summer	Biology	Module 2	0.05	**0.74**	0.14	Correlation (F1, F3) = 0.67
Summer	Literature	Module 1	-0.05	-0.03	**0.92**	Correlation (F2, F3) = 0.78
Summer	Literature	Module 2	0.15	0.20	**0.54**	Correlation (F2, F3) = 0.78

Taken as a whole, all the internal structure evidence presented generally indicates that related elements of each of the Keystone Exams are correlated in the intended manner. Different Keystone Exams seem to measure different constructs. Additionally, the modules *within* each content area have stronger relationships than the *across* content area modules. This further supports using a total score to report students' performances in the different content areas.

The module scores present more of a mixed message. Since the modules in each content area were designed to measure distinct components of the content area, it is reasonable to expect that the inter-content module correlations should be positive and strong but, ideally, not extremely high. However, the disattenuated correlations imply that some modules are essentially measuring the same constructs for most of the students. Consequently, there may be less support for providing results for some module scores beyond the total score. While there is content rationale underlying the creation of the module scores, the empirical correlations illustrate that caution is required when using the module scores as a way to identify individual student's strengths and weaknesses. Certainly, instructional programs should not be based on module score information alone, but rather in conjunction with other sources of evidence available (e.g., teacher observations, other exam performance).

EVIDENCE BASED ON RELATIONSHIPS WITH OTHER VARIABLES

As described in the *Standards* (AERA, APA, & NCME, 1999), "Evidence based on relationships with other variables addresses questions about the degree to which relationships are consistent with the construct underlying the proposed interpretations" (p. 13). This category of evidence refers to external structure evidence and has been classified as three types of evidence: *convergent, discriminant*, and *criterion-related*. Convergent evidence is provided by relationships among students' performances on different assessments intended to measure a similar construct. Discriminant evidence is provided by relationships among students' performances on different tests intended to measure different constructs. Criterion-related evidence, either predictive or concurrent, is provided by relationships between students' test scores and their performances on a criterion measure (Cronbach, 1971; Messick, 1989).

The correlations among students' test scores on different Keystone Exams including Algebra I, Biology, and Literature are shown in Table 19–6 to provide some discriminant validity evidence. In this table, both the observed and disattenuated correlations (in the parentheses) are reported.

Table 19–6. Correlations Among Students' Performances

Administration	Content Area	Algebra I	Biology
Winter	Biology	0.62 (0.69)	-
Winter	Literature	0.61 (0.68)	0.74 (0.81)
Spring	Biology	0.74 (0.81)	-
Spring	Literature	0.64 (0.70)	0.80 (0.86)
Summer	Biology	0.57 (0.68)	-
Summer	Literature	0.61 (0.74)	0.71 (0.85)

Each Keystone Exam assessment measures a different construct, so the correlations among them were not expected to be extremely high. The values in this table are consistent with this expectation. As can be seen, the correlations among the Keystone Exams ranged from 0.57 to 0.80.

External evidence for the Keystone Exams is examined by using students' scores on the 2017 Pennsylvania System of School Assessment (PSSA) as the external criteria. The final Algebra I, Biology and Literature data files were merged with the PSSA mathematics, science, and reading data using students' PAsecureIDs. Then the correlations between students' scores on the Keystone Exams and on the PSSA were calculated as one piece of external evidence. This analysis was attempted for all administrations of the Keystone Exams. However, only enough students were obtained for the spring administration. Table 19–7 summarizes the sample sizes and correlations by grade and content area after the file merging of the spring Keystone Exams and the PSSA.

Table 19–7. Number of Students with Both Spring Keystone Exams and PSSA Scores

Content Area	Grade 7 N	Grade 7 Correlation	Grade 8 N	Grade 8 Correlation
Algebra I/Mathematics	8,400	0.79	34,024	0.85
Biology/Science	NA	NA	197	0.81
Literature/Reading	NA	NA	57	0.69

The correlations within the same content area ranged from 0.69 to 0.85. These results suggest the Keystone Exams measured something similar but not identical to the corresponding PSSA tests. The results also provide external evidence in support of the Keystone Exams as a valid measure of students' achievement.

The collection of external evidence relating to the Keystone Exams is an ongoing process once the data are collected in the future. Other criterion-related evidence can be evaluated by the relationships between the Keystone Exams and criterion variables such as the Scholastic Aptitude Test (SAT), the American College Testing (ACT), or students' Grade Point Average (GPA) in their first college course.

EVIDENCE BASED ON CONSEQUENCES OF TESTS

Based on the *Standards* (AERA, APA, & NCME, 1999), evidence of the consequences of implementing an assessment program is an additional source of validity information. Both positive and negative (intended and unintended) consequences of score-based inferences must be investigated to fully evaluate the pool of validity evidence.

Lane and Stone (2002) summarized the general *intended* consequences for state assessments and accountability programs:

- Student, teacher, and administrator motivation and effort
- Curriculum and instruction practices (including content and strategies)
- Improved learning for all students
- Content and format of classroom assessments
- Professional development support
- Use and nature of test-preparation activities
- Student, teacher, administrator, and public awareness and beliefs about the assessment, criteria for judging performance, and the use of assessment results

Evidence for the improvement of student learning can be seen by looking at the increasing percentage of students who scored Proficient or Advanced across administrations. Table 19–8 provides the percentages of students who scored Proficient or Advanced by administration and content area. Values are derived from the first-time test takers for the purpose of comparison. The values for the Summer administrations were not provided because most students were retesters. For Keystone Exams, because of the change of student population across administrations, extra care should be given while drawing any conclusions.

Table 19–8. Percentages of Students at Proficient or Advanced Across Administrations

Administration	Algebra I Number	Algebra I Percent	Biology Number	Biology Percent	Literature Number	Literature Percent
Spring 2011	94,697	38.6	46,979	35.7	42,808	49.9
Winter 2012/2013	177,302	54.8	138,506	41.9	138,379	66.4
Spring 2013	157,811	47.6	134,995	47.2	117,830	63.1
Winter 2013/2014	21,621	44.4	19,672	47.3	19,795	56.3
Spring 2014	124,954	51.5	119,274	52.9	113,477	60.9
Winter 2014/2015	14,194	48.2	15,054	49.6	16,323	59.5
Spring 2015	121,255	50.1	115,936	58.0	114,387	67.2
Winter 2015/2016	12,036	46.4	15,064	54.5	16,507	65.7
Spring 2016	120,104	50.6	116,345	56.8	112,361	65.0
Winter 2016/2017	11,107	46.8	13,219	55.5	15,799	63.5
Spring 2017	118,704	51.7	115,473	56.6	110,966	64.5

Lane and Stone (2002) also summarized the possible unintended outcomes:

- Narrowing of curriculum and instruction to focus only on the specific standards assessed and ignoring the broader construct reflected in the specified standards
- Use of test-preparation materials that are closely linked to the assessment without making changes to instruction
- Use of unethical test-preparation materials or administration procedures
- Differential performance gains for subgroups of students
- Inappropriate or unfair uses of test scores, such as questionable practices in reassignment of teachers or principals
- For some students, decreased confidence and motivation to learn and to perform well on the assessment because of past experiences with assessments

As noted above, one important piece of consequential evidence pertains to the use of assessment results. As shown in Chapter Sixteen, there are several different types of scores and score reports used for the Keystone Exams. The extent to which various groups of users (e.g., students, teachers) interpret these scores and reports appropriately affects the validity of subsequent uses of these results. Chapter Sixteen is intended to provide accurate and clear test score and report information with the hope that this will help users avoid unintended uses and interpretations of the Keystone Exams results. Nevertheless, evidence pertaining to other consequences of the Keystone Exams needs continued research.

EVIDENCE RELATED TO THE USE OF RASCH MODEL

Since the Rasch model is the basis of all calibration, scaling, and equating analyses associated with the Keystone Exams, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met, as well as the fit between the model and the test data. As discussed in Chapter Twelve, the underlying assumptions of Rasch models were essentially met for all the Keystone Exams data, indicating the appropriateness of using the Rasch models to analyze the Keystone Exams data.

VALIDITY EVIDENCE SUMMARY

Validity evidence related to test content was reviewed earlier in this chapter. On the whole, the early chapters of this technical report show that a strong link can be established between each Keystone Exams item and its associated Eligible Content. Details regarding how the operational Keystone Exams were assembled to reflect the state content standards and detailed information regarding educator reviews (including content, bias, and sensitivity reviews) are presented in Chapter Six.

Module score intercorrelations were also presented in this chapter. In general, within-content-area modules (e.g., Algebra I) were correlated more highly with themselves than they did with other content-area modules (e.g., Literature). Consequently, this provides some favorable evidence regarding the internal and external relationships between the tests' components.

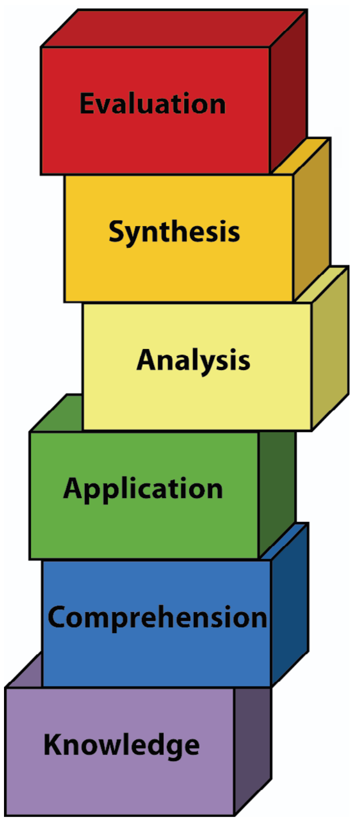
Validity of score inferences is bolstered when test scores are consistent. Here, the reliabilities of the total test scores (presented in Chapter Eighteen) were good, with many in the low 0.90s and upper 0.80s.

As reported in Chapter Five, DIF with respect to gender, ethnicity, and test administration mode helps address construct-irrelevant variance, which represents an important threat to the validity of inferences made from achievement test scores. As noted in that chapter, field test items are screened and reviewed for DIF. Only items approved by data review committees are eligible for operational use.

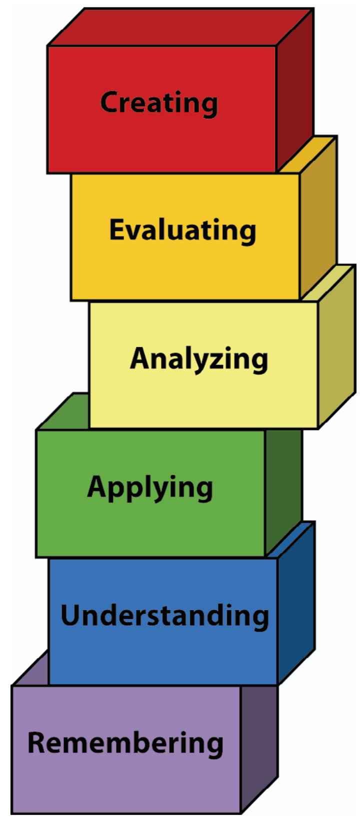
APPENDIX A: UNDERSTANDING DEPTH OF KNOWLEDGE AND COGNITIVE COMPLEXITY

One of the steps in the item review process involves Pennsylvania educators' review of items for cognitive complexity or the nature of thinking. One model for classifying thinking into cognitive levels of complexity is Bloom's Taxonomy. Bloom's Taxonomy was first presented in 1956 through the publication, The Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain. This taxonomy identifies six levels within the cognitive domain, from the simple recall or recognition of facts, at the lowest level, through increasingly more complex levels, to the highest level which is classified as evaluation.

During the late 1990s, the original Bloom's Taxonomy was revised (Anderson and Krathwohl, 2001). In the 2001 version of Bloom's Taxonomy, the names of the six major cognitive process categories or levels were revised to indicate action (verbs) rather than non-action (nouns) as noted in the graphic below.



Bloom's Taxonomy (1956)



Revised Taxonomy (2001)

More recently, Webb's Depth-of-Knowledge Levels have also been used in the review of items for cognitive demand. Webb's Depth of Knowledge was created by Norman Webb from the Wisconsin Center for Education Research. Webb's definition of depth of knowledge is the degree or complexity of knowledge that the content curriculum standards and expectations require. Therefore, when reviewing items for depth of knowledge, the item is reviewed to determine whether or not it is as demanding cognitively as what the actual content curriculum standard expects. In the case of the Pennsylvania Keystone items, the item meets the criterion if the depth of knowledge of the item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

Webb's Depth of Knowledge includes four levels, from the lowest (basic recall) to the highest (extended thinking). Verb examples that represent each level in Webb's Depth of Knowledge can be found in the information that follows. However, verbs alone do not describe the depth of knowledge. Rather, depth of knowledge also focuses upon how well the students need to know the content before they can respond to a given item.

Because Bloom's Taxonomy (1956) is very familiar to many teachers, information comparing Bloom's Taxonomy and Webb's Depth of Knowledge is provided to Pennsylvania educators during the review of the Keystone items. The comparison serves as a "bridge" for teachers to understand Webb's Depth of Knowledge as compared to Bloom's Taxonomy.

ALGEBRA I DEPTH OF KNOWLEDGE

DEPTH OF KNOWLEDGE GUIDELINES FOR REVIEW OF ALGEBRA I, ALGEBRA II, AND GEOMETRY ITEMS

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom's Taxonomy and Webb's Depth of Knowledge for mathematics (Algebra I, Algebra II, and Geometry). Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

DEFINITIONS OF WEBB'S DEPTH OF KNOWLEDGE

Level 1 (*Recall*) requires the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics, a one-step, well-defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify Level 1 include "identify," "recall," "recognize," "use," and "measure." Verbs such as "describe" and "explain" could be classified at different levels, depending on what is to be described and explained.

Level 2 (*Skill/Concept*) requires the engagement of some mental processing beyond a habitual response. A Level 2 item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of objects or phenomena and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different levels depending on the object of the action. For example, interpreting information from a simple graph, or reading information from the graph, are also at Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is at Level 3. Level 2 activities are not limited only to number skills, but may involve visualization skills and probability skills. Other Level 2 activities include noticing or describing non-trivial patterns; explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (*Strategic Thinking*) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is at Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be at Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and deciding which concepts to apply in order to solve a complex problem.

Level 4 (*Extended Thinking*) requires complex reasoning, planning, developing, and thinking most likely over an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student is to conduct a river study that requires taking into consideration a number of variables, this would be a Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include designing *and* conducting experiments and projects; developing and proving conjectures; making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Level 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g., performance-based tasks; portfolios; research studies/projects).

(Webb, N. 1997, 1999, 2002, 2005, 2006)

Table A–1. Bloom's Taxonomy – Algebra I

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers, or recalls appropriate previously learned information.	define; identify; name; select; state; order; (involves a one-step problem)
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	convert; estimate; explain; express; factor; generalize; give example; identify; indicate; locate; picture; (involves two or more steps)
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; choose; compute; employ; interpret; graph; modify; operate; plot; practice; solve; use; (involves three or more steps)
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	compare; contrast; correlate; differentiate; discriminate; examine; infer; maximize; minimize; prioritize; subdivide; test
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	arrange; collect; construct; design; develop; formulate; organize; set up; prepare; plan; propose; create experiment and record data
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	appraise; assess; defend estimate; evaluate; judge; predict; rate; validate; verify

Table A–2. Webb's Depth of Knowledge – Algebra I

Categories	Definition	Example of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	define; identify; name; select; state; order; one step
Basic Application of Skill/ Concept	Student uses information, conceptual knowledge, and procedures.	apply; choose; compute; employ; interpret; graph; modify; operate; plot; practice; solve; use; two or more steps
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	compare; contrast; correlate; differentiate; discriminate; examine; infer; maximize; minimize; prioritize; subdivide; test
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	arrange; collect; construct; design; develop; formulate; organize; set up; prepare; plan; propose; create experiment and record data

*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

BIOLOGY DEPTH OF KNOWLEDGE

Note: "Knowledge" can refer both to content knowledge and knowledge of scientific processes. This meaning of knowledge is consistent with the National Science Education Standards (NSES), which terms "Science as Inquiry" as its first Content Standard.

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom's Taxonomy and Webb's Depth of Knowledge for biology. Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

DEFINITIONS OF WEBB'S DEPTH OF KNOWLEDGE

Level 1 (*Recall*) requires the recall of information, such as a fact, definition, term, or a simple procedure, as well as performance of a simple science process or procedure. Level 1 only requires students to demonstrate a rote response, use a well-known formula, follow a set procedure (like a recipe), or perform a clearly defined series of steps. A "simple" procedure is well defined and typically involves only one step. Verbs such as "identify," "recall," "recognize," "use," "calculate," and "measure" generally represent cognitive work at the recall level. Simple word problems that can be directly translated into and solved by a formula are considered Level 1. Verbs such as "describe" and "explain" could be classified at different depth-of-knowledge levels, depending on the complexity of what is to be described and explained.

A student answering a Level 1 item either knows the answer or does not: that is, the item does not need to be "figured out" or "solved." In other words, if the knowledge necessary to answer an item automatically provides the answer to it, then the item is at Level 1. If the knowledge needed to answer the item is not automatically provided in the stem, the item is at least at Level 2. Some examples that represent but do not constitute all Level 1 performance are as follows:

- Recall or recognize a fact, term, or property.
- Represent in words or diagrams a scientific concept or relationship.
- Provide or recognize a standard scientific representation for simple phenomenon.
- Perform a routine procedure, such as measuring length.

Level 2 (*Skills and Concepts*) requires the engagement of some mental processing beyond recalling. The content knowledge or process involved is **more complex** than in Level 1. Items require students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply **more than one step**. For example, to compare data requires first identifying characteristics of the objects or phenomena and then grouping or ordering the objects. Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different depth-of-knowledge levels, depending on the complexity of the action. For example, interpreting information from a simple graph, which requires reading information from the graph, is a Level 2. An item that requires interpretation from a complex graph, such as making decisions regarding features of the graph that need to be considered and how information from the graph can be aggregated, is at Level 3. Some examples that represent but do not constitute all of Level 2 performance are as follows:

- Specify and explain the relationship between facts, terms, properties, or variables.
- Describe and explain examples and non-examples of science concepts.

- Select a procedure according to specified criteria and perform it.
- Formulate a routine problem, given data and conditions.
- Organize, represent, and interpret data.

Level 3 (*Strategic Thinking*) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. The cognitive demands at Level 3 are complex and abstract. The complexity does not result only from the fact that there could be multiple answers, a possibility for both Levels 1 and 2, but because the multi-step task requires more demanding reasoning. In most instances, requiring students to explain their thinking is at Level 3; requiring a very simple explanation or a word or two should be at Level 2. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Experimental designs in Level 3 typically involve more than one dependent variable. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems. Some examples that represent but do not constitute all Level 3 performance are as follows:

- Identify research questions and design investigations for a scientific problem.
- Solve non-routine problems.
- Develop a scientific model for a complex situation.
- Form conclusions from experimental data.

Level 4 (*Extended Thinking*) requires high cognitive demands and complexity. Students are required to make several connections—relate ideas within the content area or among content areas—and have to select or devise one approach among many alternatives to solve the problem. Many on-demand assessment instruments will not include any assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated in such a way as to expect students to perform extended thinking. “Develop generalizations of the results obtained and the strategies used and apply them to new problem situations,” is an example of a grade 8 objective that is a Level 4. Many, but not all, performance assessments and open-ended assessment activities requiring significant thought will be Level 4.

Level 4 involves complex reasoning, experimental design and planning, and probably will require an extended period of time either for the science investigation required by an objective, or for carrying out the multiple steps of an assessment item. However, the extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student is asked to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2 activity. However, if the student conducts a river study that requires taking into consideration a number of variables, this would be a Level 4. Some examples that represent but do not constitute all Level 4 performance are as follows:

- Based on data provided from a complex experiment that is novel to the student, deduct the fundamental relationship between several controlled variables.
- Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions.

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Level 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g. performance-based tasks, portfolios, research studies/projects).

(Webb, N. 1997, 1999, 2002, 2005, 2006)

Table A–3. Bloom's Taxonomy – Biology

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers, or recalls appropriate previously learned information.	identify; recall; observe; recognize; use; calculate; measure; order
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	explain; interpret; describe; classify; identify; recognize; predict
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; classify; experiment; interpret; use; order; calculate
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	analyze; order; explain; classify; arrange; compare; contrast; infer; calculate; categorize; examine; experiment; question; test
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	combine; arrange; rearrange; modify; invent; design; construct; organize; predict; infer; conclude; create; experiment and record data
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	evaluate; measure; explain; compare; summarize; predict; test decide; rate; conclude

Table A–4. Webb's Depth of Knowledge – Biology

Categories	Definition	Examples of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	identify; recall; observe; recognize; use; calculate; measure; order
Basic Application of Skill/ Concept	Student uses information, conceptual knowledge, and procedures.	explain; interpret; describe; classify; identify; order; recognize; predict; apply; use; calculate; organize; estimate; observe; collect; and display data
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	analyze; order; explain; classify; arrange; compare; contrast; infer; interpret; calculate; categorize; examine; experiment; question; predict; evaluate; test
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	combine; arrange; rearrange; propose; evaluate; modify; invent; design; construct; organize; predict; infer; conclude; evaluate; create; experiment and record data

*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

LITERATURE DEPTH OF KNOWLEDGE

Note: The levels are based on Valencia and Wixson (2000, pp. 909–935).

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom's Taxonomy and Webb's Depth of Knowledge for literature. Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

DEFINITIONS OF WEBB'S DEPTH OF KNOWLEDGE

Level 1 requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent but do not constitute all Level 1 performance are as follows:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage.

Level 2 requires the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis of inference is required. Some important concepts are covered, but not in a complex way. Content curriculum standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Some examples that represent but do not constitute all Level 2 performance are as follows:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a selection.
- Identify and summarize the major events in a narrative.

Level 3 requires deeper knowledge. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Content curriculum standards and items (Assessment Anchors as defined by the Eligible Content) at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent but do not constitute all Level 3 performance are as follows:

- Explain or recognize how the author's purpose affects the interpretation of a selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

Level 4 requires higher-order thinking and deep knowledge. The content curriculum standard or item at this level will probably require an extended activity, with extended time provided for completing it. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking. Students take information from at least one passage of a text and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts. Some examples that represent but do not constitute all Level 4 performance are as follows:

- Analyze and synthesize information from more than one source.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Leve 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g. performance-based tasks, portfolios, research studies/projects).

(Webb, N. 2005; Valencia and Wixson, 2000)

Table A–5. Bloom's Taxonomy – Literature

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers, or recalls appropriate previously learned information.	define; identify; name; recall; recognize; select; tell
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	describe; distinguish; explain; identify; indicate; interpret; locate; recognize; restate; summarize
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; choose; demonstrate; determine; interpret; inform; select; show; use
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	analyze; characterize; compare; contrast; discriminate; distinguish; explain; infer
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	compose; create; develop; formulate; generalize; organize
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	assess; conclude; convince; defend; evaluate; explain; justify; predict; prove; support

Table A–6. Webb's Depth of Knowledge – Literature

Categories	Definition	Examples of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	define; identify; locate; name; recall; recognize; sequence; tell
Basic Application of Skill/Concept	Student uses information, conceptual knowledge, and procedures.	apply; compare; comprehend; identify; describe; determine; infer; interpret; predict; summarize; use
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	analyze; cite evidence; compare; contrast; draw conclusions; explain; generalize; infer; interpret; evaluate; recognize; summarize; support
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	describe and illustrate; evaluate; examine and explain; analyze; synthesize

*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

APPENDIX B: GENERAL SCORING GUIDELINES

ALGEBRA I

ALGEBRA I CONSTRUCTED-RESPONSE QUESTIONS

GENERAL DESCRIPTION OF SCORING GUIDELINES

4 Points

- The response demonstrates a *thorough* understanding of the mathematical concepts and procedures required by the task.
- The response provides correct answer(s) with clear and complete mathematical procedures shown and a correct explanation, as required by the task. Response may contain a minor "blemish" or omission in work or explanation that does not detract from demonstrating a *thorough* understanding.

3 Points

- The response demonstrates a *general* understanding of the mathematical concepts and procedures required by the task.
- The response and explanation (as required by the task) are mostly complete and correct. The response may have minor errors or omissions that do not detract from demonstrating a *general* understanding.

2 Points

- The response demonstrates a *partial* understanding of the mathematical concepts and procedures required by the task.
- The response is somewhat correct with *partial* understanding of the required mathematical concepts and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

1 Point

- The response demonstrates a *minimal* understanding of the mathematical concepts and procedures required by the task.

0 Points

- The response has no correct answer and *insufficient* evidence to demonstrate any understanding of the mathematical concepts and procedures required by the task.

BIOLOGY CONSTRUCTED-RESPONSE QUESTIONS

GENERAL DESCRIPTION OF SCORING GUIDELINES

3 Points

- The response demonstrates a *thorough* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response provides a clear, complete, and correct response as required by the task(s). The response may contain a minor blemish or omission in work or explanation that does not detract from demonstrating a *thorough* understanding.

2 Points

- The response demonstrates a *partial* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response is somewhat correct with *partial* understanding of the required scientific content, concepts, and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

1 Point

- The response demonstrates a *minimal* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response is somewhat correct with *minimal* understanding of the required scientific content, concepts, and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

0 Points

- The response provides *insufficient* evidence to demonstrate any understanding of the scientific content, concepts, and/or procedures as required by the task(s).
- The response may show only information copied or rephrased from the question or *insufficient* correct information to receive a score of 1.

LITERATURE CONSTRUCTED-RESPONSE QUESTIONS

GENERAL DESCRIPTION OF SCORING GUIDELINES

3 Points

- The response provides a clear, complete, and accurate answer to the task.
- The response provides relevant and specific information from the passage.

2 Points

- The response provides a partial answer to the task.
- The response provides limited information from the passage and may include inaccuracies.

1 Point

- The response provides a minimal answer to the task.
- The response provides little or no information from the passage and may include inaccuracies.

OR

- The response relates minimally to the task.

0 Points

- The response is totally incorrect or irrelevant or contains insufficient information to demonstrate comprehension.

APPENDIX C: ITEM AND TEST DEVELOPMENT PROCESS FOR THE KEYSTONE EXAMS

Table C–1. Item and Test Development Process for the Keystone Exams

Step	Description
1. Review Guiding Documentation	Each year item and test development specialists meet internally to review all guiding documentation related to the Keystone Exams. Documentation reviewed includes the test design blueprints, the Keystone Assessment Anchors and Eligible Content, the test item specifications, the test style specifications (style guide), and all test content descriptions.
2. Meet with PDE to Confirm Understanding of Program	The goal of the meeting each year is to ensure that item and test development teams have a clear understanding of PDE's vision for test development. A successful development cycle requires a clear understanding of Pennsylvania's content-area test specifications and of any unique interpretations of the Keystone Assessment Anchors (if any).
3. Create Preliminary Test Item Development Plan	Item and test development specialists generate a preliminary development plan which includes an overview of the program, the internal and external (PDE) review and approval processes, a projected schedule for development of test items—including the number of test items to be developed for review by PDE and subsequent review by the committees of Pennsylvania educators. Item and test development specialists also generate strategies for securing passages and developing passage-based items, etc.
4. Meet with PDE to Finalize Test Item Development Plan	Over the course of the meeting, item and test development specialists verify all steps in the development process including timelines and schedules for test item/test development.
5. Analyze Item Bank	Existing test items in the current Keystone Exams Item Bank are reviewed for technical psychometric quality as well as for their match to the Assessment Anchors. During this phase, test development specialists also make a tally of the test items by Assessment Anchor—including test development specialists' best thinking regarding the number of usable test items in the existing item bank. A tally is also made of the number of usable passages, as well as other stimulus prompts in the bank, including science scenarios.
6. Refine Test Item Development Plan to Include Writers and Subcontractors	Item and test development specialists identify the writers who will write the test items (test development specialists or other professional item writers, subcontractors, etc.), the estimated number of writers needed, the qualifications of writers, and the approximate number of test items to be submitted by each source.
7. Train Item Writers	Item and test development specialists train item writers, as needed. Item writers who have written for the Keystone Exams in the past receive updated information, as needed.
8. Write and Review Items	Test items are written by item writers after training is complete, and feedback is provided by the item and test development specialists to item writers on a regular basis. As test items are written, they are reviewed and edited in a series of internal reviews. Item and test development specialists review and edit items to include, but not limited to, the following: match to Assessment Anchor/ Eligible Content, relevance to purpose, accuracy of content, item difficulty, interest level, depth of knowledge and cognitive complexity, adherence to the principles of Universal Design, and freedom from issues of bias/fairness/sensitivity. At the same time, the process of procuring permissions also begins, including securing permissions for passages, art, etc.
9. Enter Test Items into Database	Upon acceptance from item writers, test items are entered into the item management system, IDEAS (Item Development and Educational Assessment System). Item data stored in the system database includes, but is not limited to, the following: readability, cognitive level, estimated level of difficulty, alignment to assessment anchors, and correlation to stimulus passages.
10. Prepare Item Set for Sample Item Review by PDE	Item and test development specialists prepare a subset of the items for review by PDE.
11. PDE Conducts Sample Item Review	After a subset of the items is submitted to PDE for review, PDE reviews the items and provides feedback to item and test development teams via a conference call. Items are revised per PDE feedback.

Step	Description
12. Continue to Write and Review Items	The remaining items are written, and feedback is provided by the item and test development specialists to item writers on a regular basis. Items are entered into the item management system, IDEAS (Item Development and Educational Assessment System) (See step 8 and step 9).
13. Review Items Prior to Test Item Review and Validation Sessions	Prior to New Item Content Review, all items are submitted to PDE for review. Item and test development specialists incorporate all PDE feedback, and PDE-requested edits to items are made.
14. Prepare for Test Item Review Sessions (the New Item Content Review and the Bias, Fairness, and Sensitivity Review)	Item and test development specialists prepare all items and stimulus passages for review by the New Item Content Review Committee (consisting of Pennsylvania educators) and by the separate Bias, Fairness, and Sensitivity Committee (consisting of a panel of experts). Item and test development specialists also prepare training materials needed for training committee members to review items for content or for bias, fairness, and sensitivity issues. All training materials and other ancillary materials (e.g. agendas, presentations, etc.) are also developed and then submitted to PDE for review and approval. Invitations are also sent to Pennsylvania educators and national experts from PDE-approved committee lists.
15. Conduct Test Item Review Sessions (the New Item Content Review and the Bias, Fairness, and Sensitivity Review)	Committees of Pennsylvania educators and national experts review items in two meetings: one addressing item content and quality, the other addressing bias, fairness, and sensitivity. PDE, with support from item and test development specialists, presents training on how to review new test items for content considerations or bias/fairness/sensitivity issues. At the New Item Content Review, suggested edits to test items are made and/or replacement test items are written during the actual item review so that both the committee and the PDE are able to observe changes to the test items and approve the test items during the committee review process. At the Bias, Fairness, and Sensitivity Review, experts in bias, fairness, and sensitivity review all test items and passages and come to a consensus about any issues that are noted. At both meetings the results are carefully documented.
16. Conduct Item Review Resolution and Cleanup	Following the conclusion of the New Item Content Review Committee meetings, PDE re-examines the consensus changes suggested by the committee members during the New Item Content Review Committee meetings. DRC item and test development specialists then record all of PDE's follow-up decisions and changes. During this cleanup process, PDE either accepts the changes as requested by the committee, or PDE rejects the decision of the committee. If a committee decision is rejected, PDE provides an alternate decision for DRC to implement. During this cleanup process, PDE also interprets the report from the Bias, Fairness, and Sensitivity Committee meetings and subsequently applies changes to test items and passages. DRC item and test development specialists then apply the changes to the test items and passages per PDE's decisions.
17. Submit Field Test Items for Final Sign-Off	PDE-approved changes are applied to the items, non-permissioned passages, prompts, etc. (Changes reflect PDE's arbitration of the committee decisions.) Once all revisions to the items, non-permissioned passage text, and/or the art used by test items and passages are completed, the test items are submitted to PDE for final review and sign-off. (Changes requested to permissioned passages are sought from the publisher of record, and, if approved by the copyright holders, changes are implemented.) [PDE's approval process for field test items generally occurs simultaneously with PDE's approval of the core test forms. See step 25.]
	To follow the path for new field test items, skip to step 22, or to follow the chronological test development path, continue with step 18.
18. Review Results of the Field Test	Following the administration of a field test form and the subsequent rangefinding and field test scoring processes for field test items, performance data for all field test items are analyzed by DRC psychometricians and test development specialists. Test item performance data that meet certain triggering criteria are flagged for additional reviews by test development specialists. Flagged field test items with extreme performance data are considered psychometrically unusable and are removed from future operational consideration. Field test items with marginal performance data are prepared for the Field Test Item Data Review meeting.

Step	Description
19. Prepare for Field Test Item Data Review	Test development specialists prepare all items and stimulus passages for review by the Field Test Item Data Review Committee (which consists of Pennsylvania educators). Psychometricians also prepare training materials needed for training committee members to review items for their performance. All training materials and other ancillary materials (e.g. agendas, presentations, etc.) are submitted to PDE for review and approval. Invitations are also sent to Pennsylvania educators from PDE-approved committee lists.
20. Conduct Field Test Item Data Review	Committees of Pennsylvania educators review the performance data of flagged field test items. Psychometricians present training on how to review field test items based on their performance data. At the Item Data Review, committee members examine the performance of the items and determine whether the field test item is technically sound and appropriate for use on an operational Keystone Exams test. Since test items cannot be modified at the Field Test Item Data Review, the committee can either accept an item as is or the committee can reject the item.
21. Conduct Field Test Item Data Review Reconciliation	Following the conclusion of the Field Test Item Data Review Committee meetings, PDE re-examines the consensus decisions (accept or reject) suggested by the committee members during the Field Test Item Data Review Committee meetings. Test development specialists record all of PDE's follow-up decisions and changes. During this cleanup process, PDE either accepts the decisions of the data review committee, or PDE rejects the decisions of the data review committee. If a committee decision is not accepted, PDE provides an alternate decision for test development specialists to implement. All PDE-approved changes to the test items status (accepted or rejected) are incorporated into the Item Development and Educational Assessment System, IDEAS.
22. Select Items to Fill Core, Field Test, and Equating Block Positions in Core and Field Test Forms	After the PDE-approved changes to the new field test items is completed AND the results of the prior field test have been finalized following data review, test development specialists collaborate with psychometricians to follow the Test Design Blueprints and build requirements to make the initial selection of items for core and field-test positions for all test forms. In later administrations, core-to-core linking items will also be selected during this step.
23. Review Core and Equating Block Selections	After test content and psychometric requirements have been achieved for core, the core items are provided to PDE for review and approval. Any changes to the content of the core requested by PDE are balanced with psychometric requirements until all core positions are approved by PDE, test development specialists, and psychometricians.
24. Construct Test Forms	Items, passages, and test components are assembled into forms using the form construction and typesetting function of DRC's Item Development and Educational Assessment System, IDEAS. Forms are reviewed internally for style and formatting requirements.
25. Review Typeset Forms	After forms are constructed in IDEAS, draft hard copies of the forms are produced and presented to PDE for review and approval. Any changes to the content of the core requested by PDE are balanced with psychometric requirements until all core positions are approved by PDE, test development specialists, and psychometricians. PDE also re-reviews all field test items appearing in the test forms. DRC applies changes to the field test items as required.
26. Print Test Forms	Following PDE's approval of the test forms, DRC completes a series of final proofing of all test forms. Final forms (along with ancillary materials) are then approved for printing.
27. Assemble Documentation of Test Materials	Metadata for each test item and form is documented and proofed, including: grade, form, session/section, item sequence, reporting category, Assessment Anchor, Descriptor, Eligible Content, number of points, item type, number of answer options, item usage, stimulus ID, etc.
	To follow the path for new field test items, return to step 18.

APPENDIX D: ITEM AND DATA REVIEW CARD EXAMPLES

ITEM REVIEW CARD EXAMPLE

Standard: Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.

1. [illegible]

A. [illegible]
B. [illegible]
C. [illegible]
D. [illegible]

PA Keystone Item Card

Item ID	[illegible]
Content Area	Science
Course	Biology
Scenario ID	
Scenario Title	
Grade	HS
KAACS Standards	BIO.A.2.3.2
Item Type	Multiple Choice
Points	1
Depth of Knowledge	2
Est Difficulty	Medium
Key	[illegible]

Data Recognition Corporation 06423477 07/22/13 Page 4

DATA REVIEW CARD EXAMPLE

Standard: Compare and/or order any real numbers (rational and irrational may be mixed).

PA Keystone Data Card

Item ID	[illegible]
Content Area	Mathematics
Course	Algebra I
Passage ID	
Passage Title	
Grade	HS
Standards	KAACS: A1.1.1.1.1
Item Type	Multiple Choice
Points	1
Calculator	Yes
Depth of Knowledge	2
Est Difficulty	High
Key	[illegible]
Focus	

1. [illegible]

A. [illegible]

B. [illegible]

C. [illegible]

D. [illegible]

A: [illegible]
B: [illegible]
C: [illegible]
D: [illegible]

Data Recognition Corporation [illegible] 08/19/13 Page 3

PA3 - Data Card continued

Administration

Form Name	Use Function	Rptg Flag	Seq	Period	Year	Session	Calc	Model/Ext	Grade
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	1	Yes		HS

Traditional Statistics

N	P-Val	Mean	Item Total Corr
[illegible]	0.34		0.10

Fit Statistics

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit
9.9	9.9	1.28	1.18				

IRT Statistics

Label	Final	Final S.E.	Preliminary	Preliminary S.E.
Location	1.39	0.02		

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Threshold
A*	0.34	0.10		
B	0.24	0.11		
C	0.25	-0.22		
D	0.17	0.01		
MULTS	0.00			
OMITS	0.00			

DIF Analysis

Category	Bias Code	Num Value	N - Ref	N - Focal
MALEFEMALE	A-	-0.13	4709	4550
PAPERONLINE	A+	0.15	8242	1029
WHITEBLACK	A-	-0.23	6812	1245
WHITEHISPANIC	A-	-0.16	6812	726

APPENDIX E: ITEM RATING SHEET

Item Rating Sheet

Reviewer Signature: ____________________

Content Area: ____________ **Module:** ____________

August 2013		Content Alignment	Rigor Level Alignment				Technical Design			Universal Design		STATUS
Harrisburg, PA		Standards	Grade	Difficulty	Depth of Knowledge	Source of Challenge	Correct Answer	Distractors	Graphics	Language Demand	Bias	Acceptance Status
Unique ID	Pg #	—Higher —Lower —None	—Above —At —Below	—High —Medium —Low	—Recall —Application —Strategic Thinking	—Yes —No	—Yes —No	—Yes —No —N/A	—Yes —No —N/A	—Yes —No	—Yes —No	— Approved as is — Accepted with suggested revisions — Dissenting View

The purpose of this form is to provide guidelines to the item review process in terms of item characteristics that are essential in building a fair and balanced assessment. Use these guidelines in conjunction with the Item Rating Sheet when recording your feedback on individual items.

Content Alignment		Options
Standards, Anchors, Eligible Content	Does the content of the item align with the Standard/Anchor/Eligible Content? Each item was written to assess a particular Standard/Anchor/ Eligible Content statement which is indicated on the individual Item Card. Consider the degree to which the item is, in fact, aligned with the indicated eligible content. In making this judgment, it is important to consider whether the **content** is aligned (e.g., do the eligible content and the item both deal with fractions) and whether the required **performance** is aligned (e.g., if the eligible content calls for a comparison to be made, is this reflected in the item).	**HIGHER**—Aligns to the higher level of the EC **LOWER**—Aligns to the lower level of the EC **NONE**—No alignment with EC

Rigor Level Alignment		Options
Grade	Is the item grade-level appropriate? Is the content consistent with the experiences of a student at the grade level assessed? Is the challenge level appropriate for the grade?	**ABOVE** Grade Level **AT** Grade Level **BELOW** Grade Level
Difficulty	Do you agree with the item's difficulty rating? Item Difficulty is indicated as Low, Medium, and High. Is your rating in agreement with the difficulty rating on the Item Form?	**HIGH** **MEDIUM** **LOW**
Depth of Knowledge	Depth of Knowledge is based on the alignment work of Norman Webb. Rate each item based on the cognitive demand, using the following levels: 1. Recall – ***Recall*** of a fact, information, or procedure. 2. Basic Application of Skill or Concept – ***Use*** of information, conceptual knowledge, procedures, two or more steps, etc. 3. Strategic Thinking – Requires reasoning, developing a plan or sequence of steps; has some complexity; more than one possible answer. 4. Extended Thinking – Requires an investigation, time to think and process multiple conditions of the problem or task, and more than 10 minutes to do non-routine manipulations. (This level is generally not assessed in on-demand assessments.)	**4** = Extended Thinking **3** = Strategic Thinking **2** = Basic Application **1** = Recall

Rigor Level Alignment		Options
Source of Challenge	Is the source of challenge appropriately targeted to the content? The hardest part of the item (i.e., source of challenge) should be the content that is targeted. For example, in mathematics, the mathematics should be the major source of challenge rather than the wording or graphic. Students should not give an incorrect answer to a mathematics item because the reading level is too high or a graphic is flawed. Conversely, students should not give correct answers for reasons such as prior knowledge that make the answer to the question obvious (e.g., if the question asks which country has the largest population and students are to read a graph that includes China, there is no need to read the graph to answer the question).	**Y** = Yes **N** = No

Technical Design		Options
Correct Answer	Is there one clear, correct answer option? There should be no other answer that "could" be correct. CAUTION: This does not mean that "good" distractors are unfair.	**Y** = Yes **N** = No
Distractors	Are distractors fair and appropriate? Distractors that are appropriate offer students reasonable choices that can be arrived at by making common errors. There should be no distractors that make no sense at all. It should be possible to examine each option and to reason how a student with some deficiency in knowledge or skill could choose it. The distractors should be formatted according to acceptable standards of test construction (e.g., a phrase that is common to each distractor should be in the stem).	**Y** = Yes **N** = No **N/A** = OE items do not have distractors
Graphics	Are the graphics clear and accurate?	**Y** = Yes **N** = No **N/A** = No graphic

Universal Design		Options
Language Demand	Is language clear, well-formatted, and precise? Does the item use correct terminology for the content area? In order for all students to enter into the questions of the assessment, they must be able to understand them. If the items are formatted poorly, use unnecessarily complex words or phrases, or use figures or layouts that are difficult to understand, some students will give incorrect answers due to these factors rather than the content that is being assessed.	**Y** = Yes **N** = No
Bias	Is the item free of bias? All students will not be able to enter into the assessment if bias considerations are not resolved. Does the item contain clear bias problems? *A thorough, independent bias review* (separate from this meeting) *will be completed for all items*.	**Y** = Yes **N** = No

Status		Options
Acceptance Status	This is an overall judgment about the item. Based on the consensus of the committee, indicate whether the item was approved without revision to the content of the item or whether the item was accepted by the committee after revision of the content of the item. If there is a dissenting view (opposed to the committee consensus), record a	—**Approved** as is —**Accepted** with suggested revisions

	brief explanation of the dissenting view on the back of the Item Rating Sheet.	**—Dissenting View**

NOTES:

- If you leave a box blank on the Item Rating Sheet, it will be recorded to indicate that you did not have any specific feedback for that item or issue.
- If you object to the consensus of the committee, please note this on the item rating sheet and then record a brief explanation of the dissenting view on the back of the Item Rating Sheet.
- <u>**Do NOT remove any items from the item binder at any time.**</u>
- You must sign your item rating sheet.

APPENDIX F: TALLY SHEETS

ALGEBRA I-WINTER 2016/2017										
Keystone Exam										**Algebra I**
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items			Points: Core Points		
					MC	SCR	ECR	MC	SCR	ECR
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	0	1	0	0	4
	1	1	1	Compare and/or order any real numbers.	1	0	0	1	0	0
	1	1	2	Simplify square roots.	1	0	0	1	0	0
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1	0	0	1	0	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	0	1	0	0
	1	4	1	Use estimation to solve problems.	1	0	0	1	0	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	0	0	0	0	0	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1	0	0	1	0	0
	1	5	3	Simplify/reduce a rational algebraic expression.	0	0	0	0	0	0
	Total For Assessment Anchor A1.1.1				6	0	1	6	0	4
	2			**Linear Equations**	0	1	0	0	4	0
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	0	1	0	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1	0	0	1	0	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	2	0	0	2	0	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1	0	0	1	0	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.2				6	1	0	6	4	0
	3			**Linear Inequalities**	0	1	0	0	4	0
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	0	1	0	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1	0	0	1	0	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	2	0	0	2	0	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1	0	0	1	0	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.3				6	1	0	6	4	0
Total For Reporting Category A1.1					18	2	1	18	8	4

Keystone Exam **Algebra I**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	SCR	ECR	Points: Core Points MC	SCR	ECR
A1.2: Linear Functions & Data Organizations	1			**Functions**	0	1	0	0	4	0
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	2	0	0	2	0	0
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1	0	0	1	0	0
	1	1	3	Identify the domain or range of a relation.	1	0	0	1	0	0
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	1	0	0	1	0	0
	1	2	2	Translate from one representation of a linear function to another.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.1				6	1	0	6	4	0
	2			**Coordinate Geometry**	0	1	0	0	4	0
	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	0	1	0	0
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	2	0	0	2	0	0
	2	1	3	Write or identify a linear equation when given...	1	0	0	1	0	0
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1	0	0	1	0	0
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.2				6	1	0	6	4	0
	3			**Data Analysis**	0	0	1	0	0	4
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1	0	0	1	0	0
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	1	0	0	1	0	0
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1	0	0	1	0	0
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	2	0	0	2	0	0
	3	3	1	Find probabilities for compound events.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.3				6	0	1	6	0	4
Total For Reporting Category A1.2					18	2	1	18	8	4

ALGEBRA I-SPRING 2017										
Keystone Exam				**Algebra I**						
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items			Points: Core Points		
					MC	SCR	ECR	MC	SCR	ECR
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	1	0	0	4	0
	1	1	1	Compare and/or order any real numbers.	1	0	0	1	0	0
	1	1	2	Simplify square roots.	0	0	0	0	0	0
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1	0	0	1	0	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	0	1	0	0
	1	4	1	Use estimation to solve problems.	0	0	0	0	0	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1	0	0	1	0	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1	0	0	1	0	0
	1	5	3	Simplify/reduce a rational algebraic expression.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.1				6	1	0	6	4	0
	2			**Linear Equations**	0	0	1	0	0	4
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	0	1	0	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1	0	0	1	0	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	2	0	0	2	0	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1	0	0	1	0	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.2				6	0	1	6	0	4
	3			**Linear Inequalities**	0	1	0	0	4	0
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	0	1	0	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1	0	0	1	0	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	2	0	0	2	0	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.3				6	1	0	6	4	0
Total For Reporting Category A1.1					18	2	1	18	8	4

Keystone Exam **Algebra I**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	SCR	ECR	Points: Core Points MC	SCR	ECR
A1.2: Linear Functions & Data Organizations	1			**Functions**	0	0	1	0	0	4
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	2	0	0	2	0	0
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1	0	0	1	0	0
	1	1	3	Identify the domain or range of a relation.	1	0	0	1	0	0
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	1	0	0	1	0	0
	1	2	2	Translate from one representation of a linear function to another.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.1				6	0	1	6	0	4
	2			**Coordinate Geometry**	0	0	1	0	0	4
	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	0	1	0	0
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	2	0	0	2	0	0
	2	1	3	Write or identify a linear equation when given...	1	0	0	1	0	0
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1	0	0	1	0	0
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.2				6	0	1	6	0	4
	3			**Data Analysis**	0	1	0	0	4	0
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1	0	0	1	0	0
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	1	0	0	1	0	0
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1	0	0	1	0	0
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	2	0	0	2	0	0
	3	3	1	Find probabilities for compound events.	1	0	0	1	0	0
	Total For Assessment Anchor A1.2.3				6	1	0	6	4	0
Total For Reporting Category A1.2					18	1	2	18	4	8

ALGEBRA I-SUMMER 2017										
Keystone Exam					**Algebra I**					
					Items			Points		
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Number of Core Items			Core Points		
					MC	SCR	ECR	MC	SCR	ECR
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	1	0	0	4	0
	1	1	1	Compare and/or order any real numbers.	0	0	0	0	0	0
	1	1	2	Simplify square roots.	1	0	0	1	0	0
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	0	0	0	0	0	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	0	1	0	0
	1	4	1	Use estimation to solve problems.	2	0	0	2	0	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1	0	0	1	0	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	0	0	0	0	0	0
	1	5	3	Simplify/reduce a rational algebraic expression.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.1				6	1	0	6	4	0
	2			**Linear Equations**	0	0	1	0	0	4
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	0	1	0	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1	0	0	1	0	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1	0	0	1	0	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	2	0	0	2	0	0
	Total For Assessment Anchor A1.1.2				6	0	1	6	0	4
	3			**Linear Inequalities**	0	0	1	0	0	4
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	0	1	0	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	2	0	0	2	0	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1	0	0	1	0	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	0	1	0	0
	Total For Assessment Anchor A1.1.3				6	0	1	6	0	4
Total For Reporting Category A1.1					18	1	2	18	4	8

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items SCR	Items: Number of Core Items ECR	Points: Core Points MC	Points: Core Points SCR	Points: Core Points ECR
A1.2: Linear Functions & Data Organizations	1			**Functions**	0	1	0	0	4	0
A1.2: Linear Functions & Data Organizations	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	2	0	0	2	0	0
A1.2: Linear Functions & Data Organizations	1	1	3	Identify the domain or range of a relation.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	1	2	2	Translate from one representation of a linear function to another.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	Total For Assessment Anchor A1.2.1				6	1	0	6	4	0
A1.2: Linear Functions & Data Organizations	2			**Coordinate Geometry**	0	0	1	0	0	4
A1.2: Linear Functions & Data Organizations	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	2	1	3	Write or identify a linear equation when given...	2	0	0	2	0	0
A1.2: Linear Functions & Data Organizations	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	Total For Assessment Anchor A1.2.2				6	0	1	6	0	4
A1.2: Linear Functions & Data Organizations	3			**Data Analysis**	0	0	1	0	0	4
A1.2: Linear Functions & Data Organizations	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	2	0	0	2	0	0
A1.2: Linear Functions & Data Organizations	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	3	3	1	Find probabilities for compound events.	1	0	0	1	0	0
A1.2: Linear Functions & Data Organizations	Total For Assessment Anchor A1.2.3				6	0	1	6	0	4
Total For Reporting Category A1.2					18	1	2	18	4	8

BIOLOGY-WINTER 2016/2017

Keystone Exam | **Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.A: Basic Biological Principles	1			**Basic Biological Principles**	0	0	0	0
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	2	0	2	0
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	3	0	3	0
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	2	0	2	0
	Total For Assessment Anchor BIO.A.1				7	0	7	0
	2			**The Chemical Basis for Life**	0	0	0	0
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	3	0	3	0
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	1	0	1	0
	2	2	2	Describe how biological macromolecules form from monomers.	1	0	1	0
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	1	0	1	0
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	0	1	0	3
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	1	0	1	0
	Total For Assessment Anchor BIO.A.2				7	1	7	3
	3			**Bioenergetics**	0	0	0	0
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	1	1	1	3
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	1	0	1	0
	3	2	2	Describe the role of ATP in biochemical reactions.	1	0	1	0
	Total For Assessment Anchor BIO.A.3				3	1	3	3
	4			**Homeostasis and Transport**	0	0	0	0
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	3	0	3	0
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	2	0	2	0
	4	1	3	Describe how membrane-bound cellular organelles.	1	0	1	0
	4	2	1	Explain how organisms maintain homeostasis.	1	1	1	3
	Total For Assessment Anchor BIO.A.4				7	1	7	3
Total For Reporting Category BIO.A					24	3	24	9

Keystone Exam **Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.B: Cell Growth and Reproduction	1			**Cell Growth and Reproduction**	0	0	0	0
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	2	0	2	0
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	2	0	2	0
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.	2	0	2	0
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1	0	1	0
	Total For Assessment Anchor BIO.B.1				7	0	7	0
	2			**Genetics**	0	0	0	0
	2	1	1	Describe and/or predict observed patterns of inheritance.	1	0	1	0
	2	1	2	Describe processes that can alter composition or number of chromosomes.	1	0	1	0
	2	2	1	Describe how the processes of transcription and translation are similar in all organisms.	1	0	1	0
	2	2	2	Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	1	0	1	0
	2	3	1	Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	1	0	1	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1	1	1	3
	Total For Assessment Anchor BIO.B.2				6	1	6	3
	3			**Theory of Evolution**	0	0	0	0
	3	1	1	Explain how natural selection can impact allele frequencies of a population.	2	0	2	0
	3	1	2	Describe the factors that can contribute to the development of new species.	1	0	1	0
	3	1	3	Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	1	0	1	0
	3	2	1	Interpret evidence supporting the theory of evolution.	1	0	1	0
	3	3	1	Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.	0	1	0	3
	Total For Assessment Anchor BIO.B.3				5	1	5	3
	4			**Ecology**	0	0	0	0
	4	1	1	Describe the levels of ecological organization.	1	0	1	0
	4	1	2	Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	1	0	1	0
	4	2	1	Describe how energy flows through an ecosystem.	1	1	1	3
	4	2	2	Describe biotic interactions in an ecosystem.	1	0	1	0
	4	2	3	Describe how matter recycles through an ecosystem.	0	0	0	0
	4	2	4	Describe how ecosystems change in response to natural and human disturbances.	1	0	1	0
	4	2	5	Describe the effects of limiting factors on population dynamics and potential species extinction.	1	0	1	0
	Total For Assessment Anchor BIO.B.4				6	1	6	3
Total For Assessment Anchor BIO.B					24	3	24	9

BIOLOGY-SPRING 2017

Keystone Exam **Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.A: Basic Biological Principles	1			**Basic Biological Principles**	0	0	0	0
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	2	1	2	3
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2	0	2	0
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	2	0	2	0
	Total For Assessment Anchor BIO.A.1				6	1	6	3
	2			**The Chemical Basis for Life**	0	0	0	0
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	1	0	1	0
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	2	0	2	0
	2	2	2	Describe how biological macromolecules form from monomers.	1	0	1	0
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	1	0	1	0
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	2	0	2	0
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	0	0	0	0
	Total For Assessment Anchor BIO.A.2				7	0	7	0
	3			**Bioenergetics**	0	0	0	0
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	2	0	2	0
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	1	1	1	3
	3	2	2	Describe the role of ATP in biochemical reactions.	2	0	2	0
	Total For Assessment Anchor BIO.A.3				5	1	5	3
	4			**Homeostasis and Transport**	0	0	0	0
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	2	0	2	0
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	2	1	2	3
	4	1	3	Describe how membrane-bound cellular organelles.	1	0	1	0
	4	2	1	Explain how organisms maintain homeostasis.	1	0	1	0
	Total For Assessment Anchor BIO.A.4				6	1	6	3
Total For Reporting Category BIO.A					24	3	24	9

Keystone Exam **Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.B: Cell Growth and Reproduction	1			**Cell Growth and Reproduction**	0	0	0	0
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	2	1	2	3
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	2	0	2	0
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.	1	0	1	0
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1	0	1	0
	Total For Assessment Anchor BIO.B.1				6	1	6	3
	2			**Genetics**	0	0	0	0
	2	1	1	Describe and/or predict observed patterns of inheritance.	1	1	1	3
	2	1	2	Describe processes that can alter composition or number of chromosomes.	1	0	1	0
	2	2	1	Describe how the processes of transcription and translation are similar in all organisms.	1	0	1	0
	2	2	2	Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	1	0	1	0
	2	3	1	Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	0	0	0	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	2	0	2	0
	Total For Assessment Anchor BIO.B.2				6	1	6	3
	3			**Theory of Evolution**	0	0	0	0
	3	1	1	Explain how natural selection can impact allele frequencies of a population.	2	0	2	0
	3	1	2	Describe the factors that can contribute to the development of new species.	1	0	1	0
	3	1	3	Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	0	0	0	0
	3	2	1	Interpret evidence supporting the theory of evolution.	1	0	1	0
	3	3	1	Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.	1	1	1	3
	Total For Assessment Anchor BIO.B.3				5	1	5	3
	4			**Ecology**	0	0	0	0
	4	1	1	Describe the levels of ecological organization.	0	0	0	0
	4	1	2	Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	1	0	1	0
	4	2	1	Describe how energy flows through an ecosystem.	1	0	1	0
	4	2	2	Describe biotic interactions in an ecosystem.	1	0	1	0
	4	2	3	Describe how matter recycles through an ecosystem.	2	0	2	0
	4	2	4	Describe how ecosystems change in response to natural and human disturbances.	0	0	0	0
	4	2	5	Describe the effects of limiting factors on population dynamics and potential species extinction.	2	0	2	0
	Total For Assessment Anchor BIO.B.4				7	0	7	0
Total For Assessment Anchor BIO.B					24	3	24	9

BIOLOGY-SUMMER 2017								
Keystone Exam								**Biology**
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.A: Basic Biological Principles	1			**Basic Biological Principles**	0	0	0	0
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	2	1	2	3
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2	0	2	0
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	1	0	1	0
	Total For Assessment Anchor BIO.A.1				5	1	5	3
	2			**The Chemical Basis for Life**	0	0	0	0
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	2	0	2	0
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	2	0	2	0
	2	2	2	Describe how biological macromolecules form from monomers.	1	0	1	0
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	1	0	1	0
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	1	0	1	0
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	1	0	1	0
	Total For Assessment Anchor BIO.A.2				8	0	8	0
	3			**Bioenergetics**	0	0	0	0
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	2	0	2	0
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	1	0	1	0
	3	2	2	Describe the role of ATP in biochemical reactions.	2	1	2	3
	Total For Assessment Anchor BIO.A.3				5	1	5	3
	4			**Homeostasis and Transport**	0	0	0	0
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	2	1	2	3
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	2	0	2	0
	4	1	3	Describe how membrane-bound cellular organelles.	1	0	1	0
	4	2	1	Explain how organisms maintain homeostasis.	1	0	1	0
	Total For Assessment Anchor BIO.A.4				6	1	6	3
Total For Reporting Category BIO.A					24	3	24	9

Keystone Exam **Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
BIO.B: Cell Growth and Reproduction	1			**Cell Growth and Reproduction**	0	0	0	0
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	2	0	2	0
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	2	0	2	0
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.	0	1	0	3
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1	0	1	0
	Total For Assessment Anchor BIO.B.1				5	1	5	3
	2			**Genetics**	0	0	0	0
	2	1	1	Describe and/or predict observed patterns of inheritance.	2	1	2	3
	2	1	2	Describe processes that can alter composition or number of chromosomes.	1	0	1	0
	2	2	1	Describe how the processes of transcription and translation are similar in all organisms.	1	0	1	0
	2	2	2	Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	1	0	1	0
	2	3	1	Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	0	0	0	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	0	0	0	0
	Total For Assessment Anchor BIO.B.2				5	1	5	3
	3			**Theory of Evolution**	0	0	0	0
	3	1	1	Explain how natural selection can impact allele frequencies of a population.	2	0	2	0
	3	1	2	Describe the factors that can contribute to the development of new species.	2	0	2	0
	3	1	3	Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	2	0	2	0
	3	2	1	Interpret evidence supporting the theory of evolution.	1	0	1	0
	3	3	1	Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.	1	0	1	0
	Total For Assessment Anchor BIO.B.3				8	0	8	0
	4			**Ecology**	0	0	0	0
	4	1	1	Describe the levels of ecological organization.	1	0	1	0
	4	1	2	Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	1	1	1	3
	4	2	1	Describe how energy flows through an ecosystem.	1	0	1	0
	4	2	2	Describe biotic interactions in an ecosystem.	1	0	1	0
	4	2	3	Describe how matter recycles through an ecosystem.	1	0	1	0
	4	2	4	Describe how ecosystems change in response to natural and human disturbances.	1	0	1	0
	4	2	5	Describe the effects of limiting factors on population dynamics and potential species extinction.	0	0	0	0
	Total For Assessment Anchor BIO.B.4				6	1	6	3
Total For Assessment Anchor BIO.B					24	3	24	9

LITERATURE-WINTER 2016/2017

Keystone Exam **Literature**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
L.F: Fiction	1			**Reading for Meaning—Fiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	2	0	2	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	0	0	0	0
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	1	0	1	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	0	0	0	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	0	0	0	0
	1	2	4	Draw conclusions about connotations of words.	1	0	1	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	1	0	1	0
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.	1	0	1	0
	Total For Assessment Anchor L.F.1				6	0	6	0
	2			**Analyzing and Interpreting Literature—Fiction**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1	2	1	6
	2	1	2	Cite evidence from a text to support generalizations.	1	0	1	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	1	0	1	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	0	0	0	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	0	0	0	0
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.	0	0	0	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1	0	1	0
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:	2	0	2	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	0	0	0	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	0	0	0	0
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:	0	1	0	3
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	2	0	2	0
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.	1	0	1	0
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	2	0	2	0
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.	0	0	0	0
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.	0	0	0	0
	Total For Assessment Anchor L.F.2				11	3	11	9
Total For Reporting Category L.F					17	3	17	9

Keystone Exam **Literature**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
L.N: Nonfiction	1			**Reading for Meaning—Nonfiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	0	0	0	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.	1	0	1	0
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.	0	0	0	0
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	0	0	0	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	0	0	0	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1	0	1	0
	1	2	4	Draw conclusions about connotations of words.	0	0	0	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	2	0	2	0
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.	0	0	0	0
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.	0	1	0	3
	Total For Assessment Anchor L.N.1				4	1	4	3
	2			**Data Analysis**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	2	0	2	0
	2	1	2	Cite evidence from a text to support generalizations.	1	0	1	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	0	0	0	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1	0	1	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	3	1	3	3
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:	0	0	0	0
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:	0	0	0	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:	0	0	0	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:	0	0	0	0
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:	1	0	1	0
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	0	0	0	0
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	0	1	0	3
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.	0	0	0	0
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.	1	0	1	0
	2	4	4	Make connections between a text and the content of graphics and charts.	0	0	0	0
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.	0	0	0	0
	2	5	1	Differentiate between fact and opinion.	0	0	0	0
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.	2	0	2	0
	2	5	3	Distinguish essential from nonessential information.	1	0	1	0
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	1	0	1	0
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.	0	0	0	0
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.	0	0	0	0
	Total For Assessment Anchor L.N.2				13	2	13	6
Total For Reporting Category L.N					17	3	17	9

LITERATURE-SPRING 2017

Keystone Exam **Literature**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
L.F: Fiction	1			**Reading for Meaning—Fiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	0	1	0	3
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	1	0	1	0
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	1	0	1	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	0	0	0	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1	0	1	0
	1	2	4	Draw conclusions about connotations of words.	1	0	1	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	0	0	0	0
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.	0	0	0	0
	Total For Assessment Anchor L.F.1				4	1	4	3
	2			**Analyzing and Interpreting Literature—Fiction**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	2	0	2	0
	2	1	2	Cite evidence from a text to support generalizations.	0	0	0	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	1	0	1	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1	0	1	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	1	1	1	3
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.	0	0	0	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1	1	1	3
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:	0	0	0	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	0	0	0	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	0	0	0	0
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:	1	0	1	0
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	1	0	1	0
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.	0	0	0	0
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	4	0	4	0
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.	1	0	1	0
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.	0	0	0	0
	Total For Assessment Anchor L.F.2				13	2	13	6
Total For Reporting Category L.F					17	3	17	9

Keystone Exam **Literature**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
L.N: Nonfiction	1			**Reading for Meaning—Nonfiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	0	0	0	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.	0	0	0	0
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.	0	0	0	0
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	1	0	1	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	1	0	1	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1	0	1	0
	1	2	4	Draw conclusions about connotations of words.	0	0	0	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	1	0	1	0
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.	1	0	1	0
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.	1	0	1	0
	Total For Assessment Anchor L.N.1				6	0	6	0
	2			**Data Analysis**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	2	1	2	3
	2	1	2	Cite evidence from a text to support generalizations.	2	0	2	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	1	0	1	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	0	0	0	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	1	0	1	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:	0	1	0	3
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:	1	0	1	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:	1	0	1	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:	0	0	0	0
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:	0	0	0	0
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	0	0	0	0
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	0	0	0	0
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.	0	0	0	0
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.	1	0	1	0
	2	4	4	Make connections between a text and the content of graphics and charts.	0	0	0	0
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.	0	0	0	0
	2	5	1	Differentiate between fact and opinion.	0	0	0	0
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.	1	0	1	0
	2	5	3	Distinguish essential from nonessential information.	0	0	0	0
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	1	0	1	0
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.	0	1	0	3
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.	0	0	0	0
	Total For Assessment Anchor L.N.2				11	3	11	9
Total For Reporting Category L.N					17	3	17	9

LITERATURE-SUMMER 2017								
Keystone Exam								**Literature**
					Items		Points	
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Number of Core Items		Core Points	
					MC	CR	MC	CR
L.F: Fiction	1			**Reading for Meaning—Fiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	0	0	0	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	1	1	1	3
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	2	0	2	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	0	0	0	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	0	0	0	0
	1	2	4	Draw conclusions about connotations of words.	0	0	0	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	0	0	0	0
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.	0	0	0	0
	Total For Assessment Anchor L.F.1				3	1	3	3
	2			**Analyzing and Interpreting Literature—Fiction**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1	0	1	0
	2	1	2	Cite evidence from a text to support generalizations.	1	0	1	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	0	0	0	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1	0	1	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	0	0	0	0
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.	1	0	1	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1	0	1	0
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:	1	0	1	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	1	0	1	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	1	0	1	0
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:	1	1	1	3
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	1	0	1	0
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.	0	0	0	0
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	2	1	2	3
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.	2	0	2	0
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.	0	0	0	0
	Total For Assessment Anchor L.F.2				14	2	14	6
Total For Reporting Category L.F					17	3	17	9

Keystone Exam								Literature
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items: Number of Core Items MC	Items: Number of Core Items CR	Points: Core Points MC	Points: Core Points CR
L.N: Nonfiction	1			**Reading for Meaning—Nonfiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	0	0	0	0
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	0	0	0	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.	0	0	0	0
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.	1	0	1	0
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	0	0	0	0
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	2	0	2	0
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	0	0	0	0
	1	2	4	Draw conclusions about connotations of words.	0	0	0	0
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	1	0	1	0
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.	0	0	0	0
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.	2	0	2	0
	Total For Assessment Anchor L.N.1				6	0	6	0
	2			**Data Analysis**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	0	0	0	0
	2	1	2	Cite evidence from a text to support generalizations.	2	0	2	0
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	2	0	2	0
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	0	0	0	0
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.	0	0	0	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:	0	0	0	0
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:	0	0	0	0
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:	0	0	0	0
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:	0	1	0	3
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:	1	0	1	0
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	1	0	1	0
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	0	1	0	3
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.	0	0	0	0
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.	1	0	1	0
	2	4	4	Make connections between a text and the content of graphics and charts.	0	0	0	0
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.	0	0	0	0
	2	5	1	Differentiate between fact and opinion.	1	0	1	0
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.	0	0	0	0
	2	5	3	Distinguish essential from nonessential information.	1	0	1	0
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	0	0	0	0
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.	1	0	1	0
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.	1	1	1	3
	Total For Assessment Anchor L.N.2				11	3	11	9
Total For Reporting Category L.N					17	3	17	9

APPENDIX G: KEYSTONE EXAMS MODULE LAYOUT PLANS

Table G–1A. Winter 2016/2017, Spring 2017, and Summer 2017 Algebra I Keystone Exams Section Layout Plan

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	23	18—Operational (Core) Items; 5—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	75	85–90
2	23	18—Operational (Core) Items; 5—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	75	85–90

Table G–1B. Winter 2016/2017, Spring 2017, and Summer 2017 Biology Keystone Exams Section Layout Plan

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	32	24—Operational (Core) Items; 8—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	72	82–87
2	32	24—Operational (Core) Items; 8—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	72	82–87

Table G–1L. Winter 2016/2017, Spring 2017, and Summer 2017 Literature Keystone Exams Section Layout Plan

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	23	17—Operational (Core) Items; 6—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	73	83–88
2	23	17—Operational (Core) Items; 6—Embedded Field Test Items	4	3—Operational (Core) Items; 1—Embedded Field Test Items	73	83–88

APPENDIX H: MEAN RAW SCORES BY FORM

Table H–1. Mean Raw Scores by Form

Column Heading	Definition
Form	Form
N	Number of students
L	Length
Pts	Points possible
Min	Minimum
Max	Maximum
Mean	Mean
Med	Median
SD	Standard deviation

ALGEBRA I: SPRING

Table H–2. Algebra I Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	117,262	42	60	1	60	29.8	29	12.6
1	5,891	42	60	2	60	29.9	30	12.6
2	5,903	42	60	2	60	29.6	29	12.5
3	5,920	42	60	2	60	29.7	29	12.5
4	5,883	42	60	3	60	29.9	29	12.6
5	5,817	42	60	3	60	29.9	30	12.6
6	5,849	42	60	2	59	30.0	29	12.6
7	5,797	42	60	2	60	29.8	29	12.6
8	5,903	42	60	3	60	30.0	30	12.7
9	5,862	42	60	1	60	29.6	29	12.5
10	5,891	42	60	3	60	30.0	30	12.5
11	5,877	42	60	2	60	29.8	30	12.7
12	5,845	42	60	3	60	29.8	29	12.4
13	5,874	42	60	3	60	29.8	30	12.8
14	5,918	42	60	2	60	29.7	29	12.6
15	5,849	42	60	3	60	29.8	29	12.8
16	5,846	42	60	3	60	29.6	29	12.5
17	5,876	42	60	4	60	30.0	30	12.6
18	5,841	42	60	3	60	29.9	30	12.7
19	5,775	42	60	2	60	30.1	29	12.8
20	5,845	42	60	2	60	29.8	29	12.7

BIOLOGY: SPRING

Table H–3. Biology Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	113,838	54	66	1	66	38.0	39	13.9
1	5,791	54	66	2	66	38.3	39	13.9
2	5,681	54	66	3	66	38.0	39	14.1
3	5,602	54	66	5	66	38.0	39	14.1
4	5,692	54	66	2	66	37.9	38	14.1
5	5,675	54	66	1	66	38.0	39	13.9
6	5,729	54	66	4	66	37.6	38	14.0
7	5,658	54	66	3	66	38.3	39	14.0
8	5,734	54	66	2	66	38.0	39	13.8
9	5,717	54	66	3	66	37.6	38	13.9
10	5,682	54	66	3	66	38.1	39	14.1
11	5,697	54	66	2	66	38.4	39	13.7
12	5,723	54	66	1	66	38.0	39	13.8
13	5,635	54	66	4	65	38.0	39	14.0
14	5,630	54	66	3	66	38.0	39	13.8
15	5,686	54	66	4	66	38.1	39	13.9
16	5,660	54	66	4	66	38.1	39	13.9
17	5,714	54	66	5	66	37.8	38	13.9
18	5,693	54	66	3	65	37.8	39	13.9
19	5,696	54	66	4	66	37.8	38	13.8
20	5,743	54	66	3	66	37.6	38	13.8

LITERATURE: SPRING

Table H–4. Literature Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	110,904	40	52	0	52	31.8	34	10.1
1	5,554	40	52	3	51	31.5	34	10.1
2	5,563	40	52	3	51	31.9	34	10.1
3	5,537	40	52	2	51	31.8	34	10.2
4	5,526	40	52	1	51	31.8	34	10.0
5	5,609	40	52	2	51	31.8	34	10.0
6	5,543	40	52	2	52	31.7	34	10.2
7	5,549	40	52	3	52	31.7	34	10.1
8	5,610	40	52	3	52	31.7	34	10.0
9	5,553	40	52	0	52	31.7	34	10.2
10	5,517	40	52	3	52	32.0	34	9.9
11	5,550	40	52	0	52	31.9	34	10.0
12	5,522	40	52	2	51	31.7	34	10.1
13	5,484	40	52	2	51	31.9	34	10.0
14	5,526	40	52	3	51	31.7	34	10.0
15	5,526	40	52	1	51	31.9	34	10.1
16	5,565	40	52	2	52	31.8	34	10.0
17	5,572	40	52	2	52	32.1	34	10.1
18	5,487	40	52	3	52	32.0	34	10.1
19	5,561	40	52	1	52	32.1	34	10.1
20	5,550	40	52	2	51	32.0	34	10.1

APPENDIX I: DEMOGRAPHIC AND ACCOMMODATION DATA

WINTER 2016/2017

Students Assessed on the Winter 2016/2017 Keystone: Algebra I

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	10	0	28	623	9,596	23,583	17,175	393	51,408
Total number of CBT processed (Number)	0	0	6	105	774	2,721	1,907	140	5,653
Total number of tests processed (Number)	10	0	34	728	10,370	26,304	19,082	533	57,061
Total number of tests processed with a score (Number)	10	0	34	722	9,923	25,208	17,785	461	54,143
Total number of tests processed with a score (Percent)	100	0	100	99.2	95.7	95.8	93.2	86.5	94.9
Total number of tests processed without a score (Number)	0	0	0	6	447	1,096	1,297	72	2,918
Total number of tests processed without a score (Percent)	0	0	0	.8	4.3	4.2	6.8	13.5	5.1

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Students Assessed on the Winter 2016/2017 Keystone: Biology

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	9	1	1,956	15,029	19,732	399	37,126
Total number of CBT processed (Number)	0	2	166	2,742	3,088	173	6,171
Total number of tests processed (Number)	9	3	2,122	17,771	22,820	572	43,297
Total number of tests processed with a score (Number)	7	3	2,000	17,008	21,159	540	40,717
Total number of tests processed with a score (Percent)	77.8	100	94.3	95.7	92.7	94.4	94
Total number of tests processed without a score (Number)	2	0	122	763	1,661	32	2,580
Total number of tests processed without a score (Percent)	22.2	0	5.7	4.3	7.3	5.6	6

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Students Assessed on the Winter 2016/2017 Keystone: Literature

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	
Total number of PPT processed (Number)	8	1	411	11,550	20,287	282	32,539
Total number of CBT processed (Number)	0	0	33	1,800	2,968	120	4,921
Total number of tests processed (Number)	8	1	444	13,350	23,255	402	37,460
Total number of tests processed with a score (Number)	7	1	357	12,884	21,659	351	35,259
Total number of tests processed with a score (Percent)	87.5	100	80.4	96.5	93.1	87.3	94.1
Total number of tests processed without a score (Number)	1	0	87	466	1,596	51	2,201
Total number of tests processed without a score (Percent)	12.5	0	19.6	3.5	6.9	12.7	5.9

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Winter 2016/2017 Keystone: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	0	196	305	352	17	870
Extended absence from school (Percent)	0	0	0	0	43.8	27.8	27.1	23.6	29.8
Absent without make-up (Number)	0	0	0	0	58	157	245	11	471
Absent without make-up (Percent)	0	0	0	0	13	14.3	18.9	15.3	16.1
Non-attempt (Number)	0	0	0	1	134	401	409	31	976
Non-attempt (Percent)	0	0	0	16.7	30	36.6	31.5	43.1	33.4
Medical emergency (Number)	0	0	0	1	17	40	38	4	100
Medical emergency (Percent)	0	0	0	16.7	3.8	3.6	2.9	5.6	3.4
Parental request - Chapter 4 (Number)	0	0	0	1	11	23	68	1	104
Parental request - Chapter 4 (Percent)	0	0	0	16.7	2.5	2.1	5.2	1.4	3.6
Parental request - Other reasons (Number)	0	0	0	1	5	56	79	0	141
Parental request - Other reasons (Percent)	0	0	0	16.7	1.1	5.1	6.1	0	4.8
Other reasons (Number)	0	0	0	2	26	114	106	8	256
Other reasons (Percent)	0	0	0	33.3	5.8	10.4	8.2	11.1	8.8
Total not assessed (Number)	0	0	0	6	447	1,096	1,297	72	2,918

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Winter 2016/2017 Keystone: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	1	0	61	210	434	9	715
Extended absence from school (Percent)	50	0	50	27.5	26.1	28.1	27.7
Absent without make-up (Number)	0	0	17	112	313	2	444
Absent without make-up (Percent)	0	0	13.9	14.7	18.8	6.3	17.2
Non-attempt (Number)	0	0	33	279	446	11	769
Non-attempt (Percent)	0	0	27	36.6	26.9	34.4	29.8
Medical emergency (Number)	0	0	2	30	37	1	70
Medical emergency (Percent)	0	0	1.6	3.9	2.2	3.1	2.7
Parental request - Chapter 4 (Number)	0	0	0	36	115	1	152
Parental request - Chapter 4 (Percent)	0	0	0	4.7	6.9	3.1	5.9
Parental request - Other reasons (Number)	0	0	0	19	134	0	153
Parental request - Other reasons (Percent)	0	0	0	2.5	8.1	0	5.9
Other reasons (Number)	1	0	9	77	182	8	277
Other reasons (Percent)	50	0	7.4	10.1	11	25	10.7
Total not assessed (Number)	2	0	122	763	1,661	32	2,580

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Winter 2016/2017 Keystone: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	63	161	403	10	637
Extended absence from school (Percent)	0	0	72.4	34.5	25.3	19.6	28.9
Absent without make-up (Number)	0	0	10	60	255	8	333
Absent without make-up (Percent)	0	0	11.5	12.9	16	15.7	15.1
Non-attempt (Number)	1	0	11	130	412	19	573
Non-attempt (Percent)	100	0	12.6	27.9	25.8	37.3	26
ELL in first year in U.S. schools (Number)	0	0	0	0	22	0	22
ELL in first year in U.S. schools (Percent)	0	0	0	0	1.4	0	1
Medical emergency (Number)	0	0	0	19	48	3	70
Medical emergency (Percent)	0	0	0	4.1	3	5.9	3.2
Parental request - Chapter 4 (Number)	0	0	0	6	90	0	96
Parental request - Chapter 4 (Percent)	0	0	0	1.3	5.6	0	4.4
Parental request - Other reasons (Number)	0	0	0	6	120	0	126
Parental request - Other reasons (Percent)	0	0	0	1.3	7.5	0	5.7
Other reasons (Number)	0	0	3	84	246	11	344
Other reasons (Percent)	0	0	3.4	18	15.4	21.6	15.6
Total not assessed (Number)	1	0	87	466	1,596	51	2,201

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Winter 2016/2017 Keystone: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	3	0	13	384	5,156	12,675	8,440	211	26,882
Female (Percent)	30	0	38.2	53.2	52	50.3	47.5	45.8	49.7
Male (Number)	4	0	21	338	4,767	12,532	9,344	250	27,256
Male (Percent)	40	0	61.8	46.8	48	49.7	52.5	54.2	50.3
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	2	21	49	33	0	105
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	.3	.2	.2	.2	0	.2
Asian (not Hispanic) (Number)	1	0	14	76	284	620	328	10	1,333
Asian (not Hispanic) (Percent)	10	0	41.2	10.5	2.9	2.5	1.8	2.2	2.5
Black or African American (not Hispanic) (Number)	2	0	2	34	1,184	5,265	3,649	119	10,255
Black or African American (not Hispanic) (Percent)	20	0	5.9	4.7	11.9	20.9	20.5	25.8	18.9
Hispanic (any race) (Number)	1	0	0	38	1,083	3,716	2,734	41	7,613
Hispanic (any race) (Percent)	10	0	0	5.3	10.9	14.7	15.4	8.9	14.1
Multi-Racial (not Hispanic) (Number)	0	0	0	21	248	628	441	13	1,351
Multi-Racial (not Hispanic) (Percent)	0	0	0	2.9	2.5	2.5	2.5	2.8	2.5
White (not Hispanic) (Number)	3	0	18	551	7,098	14,913	10,579	276	33,438
White (not Hispanic) (Percent)	30	0	52.9	76.3	71.5	59.2	59.5	59.9	61.8
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	5	16	18	2	41
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	.1	.1	.1	.4	.1
IEP (not gifted) (Number)	1	0	2	13	718	4,421	4,493	128	9,776
IEP (not gifted) (Percent)	10	0	5.9	1.8	7.2	17.5	25.3	27.8	18.1
Student exited IEP in last 2 years (Number)	0	0	2	9	152	369	270	1	803
Student exited IEP in last 2 years (Percent)	0	0	5.9	1.2	1.5	1.5	1.5	.2	1.5
Title I (Number)	3	0	1	56	1,933	7,564	5,026	133	14,716
Title I (Percent)	30	0	2.9	7.8	19.5	30	28.3	28.9	27.2
Title III served (Number)	0	0	0	1	139	1,185	1,038	20	2,383
Title III served (Percent)	0	0	0	.1	1.4	4.7	5.8	4.3	4.4
Title III not served (Number)	0	0	0	0	1	2	5	0	8
Title III not served (Percent)	0	0	0	0	0	0	0	0	0
Migrant student (Number)	0	0	0	0	3	13	15	0	31
Migrant student (Percent)	0	0	0	0	0	.1	.1	0	.1
ELL (enrolled after 1/20/16) (Number)	0	0	1	1	15	127	165	0	309
ELL (enrolled after 1/20/16) (Percent)	0	0	2.9	.1	.2	.5	.9	0	.6
ELL (enrolled on or before 1/20/16) (Number)	0	0	1	1	134	1,102	909	23	2,170
ELL (enrolled on or before 1/20/16) (Percent)	0	0	2.9	.1	1.4	4.4	5.1	5	4

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	1	17	115	91	2	226
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	.1	.2	.5	.5	.4	.4
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	22	88	72	1	183
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	.2	.3	.4	.2	.3
Former ELL no longer monitored (Number)	0	0	0	9	257	703	386	3	1,358
Former ELL no longer monitored (Percent)	0	0	0	1.2	2.6	2.8	2.2	.7	2.5
Foreign exchange student (Number)	0	0	0	0	0	4	12	1	17
Foreign exchange student (Percent)	0	0	0	0	0	0	.1	.2	0
Economically disadvantaged (Number)	1	0	1	115	3,908	13,705	10,032	261	28,023
Economically disadvantaged (Percent)	10	0	2.9	15.9	39.4	54.4	56.4	56.6	51.8
Historically Underperforming Subgroup (Number)	2	0	5	127	4,264	15,747	12,033	320	32,498
Historically Underperforming Subgroup (Percent)	20	0	14.7	17.6	43	62.5	67.7	69.4	60
Enrollment in school of residence after 10/1/16 (Number)	4	0	0	2	72	433	325	18	854
Enrollment in school of residence after 10/1/16 (Percent)	40	0	0	.3	.7	1.7	1.8	3.9	1.6
Enrollment in district of residence after 10/1/16 (Number)	4	0	0	2	47	281	235	15	584
Enrollment in district of residence after 10/1/16 (Percent)	40	0	0	.3	.5	1.1	1.3	3.3	1.1
Enrollment as PA resident after 10/1/16 (Number)	4	0	0	2	20	126	96	3	251
Enrollment as PA resident after 10/1/16 (Percent)	40	0	0	.3	.2	.5	.5	.7	.5
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	8	26	3,427	6,611	3,426	98	13,596
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	23.5	3.6	34.5	26.2	19.3	21.3	25.1
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	5	21	971	2,177	1,896	65	5,135
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	14.7	2.9	9.8	8.6	10.7	14.1	9.5
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	1	3	19	38	31	92
Court/agency placed (Percent)	0	0	0	.1	0	.1	.2	6.7	.2
Number of assessed students (Number)	10	0	34	722	9,923	25,208	17,785	461	54,143

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Winter 2016/2017 Keystone: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	2	2	1,033	8,443	10,247	270	19,997
Female (Percent)	28.6	66.7	51.7	49.6	48.4	50	49.1
Male (Number)	3	1	967	8,564	10,911	270	20,716
Male (Percent)	42.9	33.3	48.4	50.4	51.6	50	50.9
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	5	24	24	2	55
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.3	.1	.1	.4	.1
Asian (not Hispanic) (Number)	1	1	120	518	403	13	1,056
Asian (not Hispanic) (Percent)	14.3	33.3	6	3	1.9	2.4	2.6
Black or African American (not Hispanic) (Number)	0	0	171	2,566	3,969	118	6,824
Black or African American (not Hispanic) (Percent)	0	0	8.6	15.1	18.8	21.9	16.8
Hispanic (any race) (Number)	0	0	186	1,726	3,211	50	5,173
Hispanic (any race) (Percent)	0	0	9.3	10.1	15.2	9.3	12.7
Multi-Racial (not Hispanic) (Number)	0	0	52	383	512	21	968
Multi-Racial (not Hispanic) (Percent)	0	0	2.6	2.3	2.4	3.9	2.4
White (not Hispanic) (Number)	4	2	1,465	11,779	13,016	335	26,601
White (not Hispanic) (Percent)	57.1	66.7	73.3	69.3	61.5	62	65.3
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	1	11	20	1	33
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.1	.1	.1	.2	.1
IEP (not gifted) (Number)	0	0	202	2,786	4,751	122	7,861
IEP (not gifted) (Percent)	0	0	10.1	16.4	22.5	22.6	19.3
Student exited IEP in last 2 years (Number)	0	0	32	251	251	3	537
Student exited IEP in last 2 years (Percent)	0	0	1.6	1.5	1.2	.6	1.3
Title I (Number)	0	0	234	3,080	5,704	124	9,142
Title I (Percent)	0	0	11.7	18.1	27	23	22.5
Title III served (Number)	0	0	29	390	1,130	15	1,564
Title III served (Percent)	0	0	1.5	2.3	5.3	2.8	3.8
Title III not served (Number)	0	0	1	0	6	0	7
Title III not served (Percent)	0	0	.1	0	0	0	0
Migrant student (Number)	0	0	0	0	17	0	17
Migrant student (Percent)	0	0	0	0	.1	0	0
ELL (enrolled after 1/20/16) (Number)	0	0	2	29	148	0	179
ELL (enrolled after 1/20/16) (Percent)	0	0	.1	.2	.7	0	.4
ELL (enrolled on or before 1/20/16) (Number)	0	0	29	385	1,024	19	1,457
ELL (enrolled on or before 1/20/16) (Percent)	0	0	1.5	2.3	4.8	3.5	3.6
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	2	51	120	2	175

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.1	.3	.6	.4	.4
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	6	45	91	3	145
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	.3	.3	.4	.6	.4
Former ELL no longer monitored (Number)	0	0	54	406	526	6	992
Former ELL no longer monitored (Percent)	0	0	2.7	2.4	2.5	1.1	2.4
Foreign exchange student (Number)	0	0	0	5	11	1	17
Foreign exchange student (Percent)	0	0	0	0	.1	.2	0
Economically disadvantaged (Number)	2	1	655	7,355	11,825	297	20,135
Economically disadvantaged (Percent)	28.6	33.3	32.8	43.2	55.9	55	49.5
Historically Underperforming Subgroup (Number)	2	1	761	8,758	13,900	353	23,775
Historically Underperforming Subgroup (Percent)	28.6	33.3	38.1	51.5	65.7	65.4	58.4
Enrollment in school of residence after 10/1/16 (Number)	2	0	18	175	312	24	531
Enrollment in school of residence after 10/1/16 (Percent)	28.6	0	.9	1	1.5	4.4	1.3
Enrollment in district of residence after 10/1/16 (Number)	2	0	10	115	221	21	369
Enrollment in district of residence after 10/1/16 (Percent)	28.6	0	.5	.7	1	3.9	.9
Enrollment as PA resident after 10/1/16 (Number)	2	0	4	60	87	5	158
Enrollment as PA resident after 10/1/16 (Percent)	28.6	0	.2	.4	.4	.9	.4
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	1	0	914	3,275	3,659	92	7,941
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	14.3	0	45.7	19.3	17.3	17	19.5
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	171	1,006	1,877	66	3,120
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	8.6	5.9	8.9	12.2	7.7
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	3	14	36	30	83
Court/agency placed (Percent)	0	0	.2	.1	.2	5.6	.2
Number of assessed students (Number)	7	3	2,000	17,008	21,159	540	40,717

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Winter 2016/2017 Keystone: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	1	1	166	6,212	9,236	153	15,769
Female (Percent)	14.3	100	46.5	48.2	42.6	43.6	44.7
Male (Number)	2	0	191	6,672	12,421	198	19,484
Male (Percent)	28.6	0	53.5	51.8	57.3	56.4	55.3
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	3	23	33	0	59
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.8	.2	.2	0	.2
Asian (not Hispanic) (Number)	0	0	1	405	538	10	954
Asian (not Hispanic) (Percent)	0	0	.3	3.1	2.5	2.8	2.7
Black or African American (not Hispanic) (Number)	0	0	101	1,570	3,887	85	5,643
Black or African American (not Hispanic) (Percent)	0	0	28.3	12.2	17.9	24.2	16
Hispanic (any race) (Number)	1	0	75	1,440	3,072	37	4,625
Hispanic (any race) (Percent)	14.3	0	21	11.2	14.2	10.5	13.1
Multi-Racial (not Hispanic) (Number)	0	0	21	235	486	13	755
Multi-Racial (not Hispanic) (Percent)	0	0	5.9	1.8	2.2	3.7	2.1
White (not Hispanic) (Number)	2	1	156	9,203	13,618	206	23,186
White (not Hispanic) (Percent)	28.6	100	43.7	71.4	62.9	58.7	65.8
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	8	20	0	28
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	.1	.1	0	.1
IEP (not gifted) (Number)	0	0	73	1,778	5,721	124	7,696
IEP (not gifted) (Percent)	0	0	20.4	13.8	26.4	35.3	21.8
Student exited IEP in last 2 years (Number)	0	0	3	185	293	1	482
Student exited IEP in last 2 years (Percent)	0	0	.8	1.4	1.4	.3	1.4
Title I (Number)	1	1	194	2,037	5,193	100	7,526
Title I (Percent)	14.3	100	54.3	15.8	24	28.5	21.3
Title III served (Number)	0	0	20	278	1,168	8	1,474
Title III served (Percent)	0	0	5.6	2.2	5.4	2.3	4.2
Title III not served (Number)	0	0	0	1	8	0	9
Title III not served (Percent)	0	0	0	0	0	0	0
Migrant student (Number)	0	0	0	1	19	0	20
Migrant student (Percent)	0	0	0	0	.1	0	.1
ELL (enrolled after 1/20/16) (Number)	0	0	0	13	68	0	81
ELL (enrolled after 1/20/16) (Percent)	0	0	0	.1	.3	0	.2
ELL (enrolled on or before 1/20/16) (Number)	0	0	20	277	1,133	15	1,445
ELL (enrolled on or before 1/20/16) (Percent)	0	0	5.6	2.1	5.2	4.3	4.1
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	1	37	139	5	182

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.3	.3	.6	1.4	.5
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	33	93	3	129
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	.3	.4	.9	.4
Former ELL no longer monitored (Number)	0	0	9	283	502	7	801
Former ELL no longer monitored (Percent)	0	0	2.5	2.2	2.3	2	2.3
Foreign exchange student (Number)	0	0	0	2	13	0	15
Foreign exchange student (Percent)	0	0	0	0	.1	0	0
Economically disadvantaged (Number)	0	0	255	4,999	11,706	217	17,177
Economically disadvantaged (Percent)	0	0	71.4	38.8	54	61.8	48.7
Historically Underperforming Subgroup (Number)	0	0	272	5,846	14,274	266	20,658
Historically Underperforming Subgroup (Percent)	0	0	76.2	45.4	65.9	75.8	58.6
Enrollment in school of residence after 10/1/16 (Number)	4	0	16	161	381	13	575
Enrollment in school of residence after 10/1/16 (Percent)	57.1	0	4.5	1.2	1.8	3.7	1.6
Enrollment in district of residence after 10/1/16 (Number)	4	0	10	109	266	12	401
Enrollment in district of residence after 10/1/16 (Percent)	57.1	0	2.8	.8	1.2	3.4	1.1
Enrollment as PA resident after 10/1/16 (Number)	3	0	7	49	108	2	169
Enrollment as PA resident after 10/1/16 (Percent)	42.9	0	2	.4	.5	.6	.5
Enrollment in school of residence after 10/1/15 but on/ before 10/1/16 (Number)	0	0	123	2,072	3,969	86	6,250
Enrollment in school of residence after 10/1/15 but on/ before 10/1/16 (Percent)	0	0	34.5	16.1	18.3	24.5	17.7
Enrollment in district of residence after 10/1/15 but on/ before 10/1/16 (Number)	0	0	64	812	2,010	51	2,937
Enrollment in district of residence after 10/1/15 but on/ before 10/1/16 (Percent)	0	0	17.9	6.3	9.3	14.5	8.3
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	2	15	44	27	88
Court/agency placed (Percent)	0	0	.6	.1	.2	7.7	.2
Number of assessed students (Number)	7	1	357	12,884	21,659	351	35,259

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Incidence of Presentation Accommodations Received on the Winter 2016/2017 Keystone: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	2	N/A	2
Braille format (Percent)	0	N/A	0
Large print format (Number)	19	N/A	19
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	261	32	293
Some test items/questions read aloud (Percent)	.5	.6	.5
All test items/questions read aloud (Number)	259	80	339
All test items/questions read aloud (Percent)	.5	1.5	.6
Test items/questions signed (Number)	6	0	6
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	25	0	25
Test items/questions interpreted for ELL student (Percent)	.1	0	0
Amplification device (Number)	5	0	5
Amplification device (Percent)	0	0	0
Magnification device (Number)	6	1	7
Magnification device (Percent)	0	0	0
Color overlay (Number)	1	N/A	1
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	5	12	17
Other (per Accommodations Guidelines) (Percent)	0	.2	0
Spanish version (Number)	615	N/A	615
Spanish version (Percent)	1.3	N/A	1.1
Audio (Number)	N/A	274	274
Audio (Percent)	N/A	5	.5
Color Chooser (Number)	N/A	41	41
Color Chooser (Percent)	N/A	.7	.1
Contrasting Text Chooser (Number)	N/A	21	21
Contrasting Text Chooser (Percent)	N/A	.4	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	48,671	5,472	54,143

Incidence of Presentation Accommodations Received on the Winter 2016/2017 Keystone: Biology

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	2	N/A	2
Braille format (Percent)	0	N/A	0
Large print format (Number)	14	N/A	14
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	2	N/A	2
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	281	56	337
Some test items/questions read aloud (Percent)	.8	.9	.8
All test items/questions read aloud (Number)	325	63	388
All test items/questions read aloud (Percent)	.9	1.1	1
Test items/questions signed (Number)	5	0	5
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	16	0	16
Test items/questions interpreted for ELL student (Percent)	0	0	0
Amplification device (Number)	2	0	2
Amplification device (Percent)	0	0	0
Magnification device (Number)	1	1	2
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	2	6	8
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Spanish version (Number)	381	N/A	381
Spanish version (Percent)	1.1	N/A	.9
Audio (Number)	N/A	293	293
Audio (Percent)	N/A	4.9	.7
Color Chooser (Number)	N/A	19	19
Color Chooser (Percent)	N/A	.3	0
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	34,753	5,964	40,717

Incidence of Presentation Accommodations Received on the Winter 2016/2017 Keystone: Literature

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	1	N/A	1
Braille format (Percent)	0	N/A	0
Large print format (Number)	20	N/A	20
Large print format (Percent)	.1	N/A	.1
Computer Assistive Technology (Number)	2	N/A	2
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	1	0	1
Amplification device (Percent)	0	0	0
Magnification device (Number)	3	1	4
Magnification device (Percent)	0	0	0
Color overlay (Number)	1	N/A	1
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	4	7	11
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Color Chooser (Number)	N/A	28	28
Color Chooser (Percent)	N/A	.6	.1
Contrasting Text Chooser (Number)	N/A	12	12
Contrasting Text Chooser (Percent)	N/A	.3	0
Number of assessed students (Number)	30,462	4,797	35,259

Incidence of Response Accommodations Received on the Winter 2016/2017 Keystone: Algebra I

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	17	0	17
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	19	1	20
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	34	0	34
Test administrator transcribed student responses (Percent)	.1	0	.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	3	0	3
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	8	0	8
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	6	N/A	6
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	1	N/A	1
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	1	0	1
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	64	2	66
Translation dictionary for ELL student (Percent)	.1	0	.1
Other (per Accommodations Guidelines) (Number)	7	0	7
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	48,671	5,472	54,143

Incidence of Response Accommodations Received on the Winter 2016/2017 Keystone: Biology

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	14	0	14
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	23	0	23
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	.1
Test administrator transcribed student responses (Number)	27	0	27
Test administrator transcribed student responses (Percent)	.1	0	.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	2	0	2
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	3	0	3
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	7	N/A	7
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	1	N/A	1
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	46	2	48
Translation dictionary for ELL student (Percent)	.1	0	.1
Other (per Accommodations Guidelines) (Number)	13	0	13
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	34,753	5,964	40,717

Incidence of Response Accommodations Received on the Winter 2016/2017 Keystone: Literature

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	20	0	20
Test administrator marked multiple-choice responses at student's direction (Percent)	.1	0	.1
Test administrator scribed open-ended responses at student's direction (Number)	28	0	28
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	.1
Test administrator transcribed student responses (Number)	49	0	49
Test administrator transcribed student responses (Percent)	.2	0	.1
Keyboard, word processor, or computer (Number)	13	N/A	13
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	0	N/A	0
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	3	0	3
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	30,462	4,797	35,259

Incidence of Setting Accommodations Received on the Winter 2016/2017 Keystone: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	5	0	5
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	97	0	97
One-on-one setting (Percent)	.2	0	.2
Small group setting (Number)	4,117	500	4,617
Small group setting (Percent)	8.5	9.1	8.5
Other (per Accommodations Guidelines) (Number)	38	11	49
Other (per Accommodations Guidelines) (Percent)	.1	.2	.1
Number of assessed students (Number)	48,671	5,472	54,143

Incidence of Setting Accommodations Received on the Winter 2016/2017 Keystone: Biology

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	2	0	2
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	112	0	112
One-on-one setting (Percent)	.3	0	.3
Small group setting (Number)	3,378	454	3,832
Small group setting (Percent)	9.7	7.6	9.4
Other (per Accommodations Guidelines) (Number)	25	7	32
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Number of assessed students (Number)	34,753	5,964	40,717

Incidence of Setting Accommodations Received on the Winter 2016/2017 Keystone: Literature

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	4	0	4
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	89	0	89
One-on-one setting (Percent)	.3	0	.3
Small group setting (Number)	3,286	445	3,731
Small group setting (Percent)	10.8	9.3	10.6
Other (per Accommodations Guidelines) (Number)	22	6	28
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Number of assessed students (Number)	30,462	4,797	35,259

Incidence of Timing Accommodations Received on the Winter 2016/2017 Keystone: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	4,453	572	5,025
Extended time (Percent)	9.1	10.5	9.3
Frequent breaks (Number)	231	94	325
Frequent breaks (Percent)	.5	1.7	.6
Changed test schedule (Number)	70	0	70
Changed test schedule (Percent)	.1	0	.1
Other (per Accommodations Guidelines) (Number)	2	1	3
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	48,671	5,472	54,143

Incidence of Timing Accommodations Received on the Winter 2016/2017 Keystone: Biology

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	1,814	369	2,183
Extended time (Percent)	5.2	6.2	5.4
Frequent breaks (Number)	205	79	284
Frequent breaks (Percent)	.6	1.3	.7
Changed test schedule (Number)	67	0	67
Changed test schedule (Percent)	.2	0	.2
Other (per Accommodations Guidelines) (Number)	1	0	1
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	34,753	5,964	40,717

Incidence of Timing Accommodations Received on the Winter 2016/2017 Keystone: Literature

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	3,225	344	3,569
Extended time (Percent)	10.6	7.2	10.1
Frequent breaks (Number)	196	74	270
Frequent breaks (Percent)	.6	1.5	.8
Changed test schedule (Number)	45	0	45
Changed test schedule (Percent)	.1	0	.1
Other (per Accommodations Guidelines) (Number)	4	1	5
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	30,462	4,797	35,259

Accommodation Rate for Non-IEP and IEP Students on the Winter 2016 Keystone Exams: Algebra I

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	39,975	4,392	44,367
Non-Accommodated (Number)	36,212	4,017	40,229
Non-Accommodated (Percent)	90.6	91.5	90.7
Accommodated (Number)	3,763	375	4,138
Accommodated (Percent)	9.4	8.5	9.3
IEP Students (Number)	8,696	1,080	9,776
Non-Accommodated (Number)	4,576	520	5,096
Non-Accommodated (Percent)	52.6	48.1	52.1
Accommodated (Number)	4,120	560	4,680
Accommodated (Percent)	47.4	51.9	47.9

Accommodation Rate for Non-IEP and IEP Students on the Winter 2016 Keystone Exams: Biology

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	28,083	4,773	32,856
Non-Accommodated (Number)	26,612	4,591	31,203
Non-Accommodated (Percent)	94.8	96.2	95
Accommodated (Number)	1,471	182	1,653
Accommodated (Percent)	5.2	3.8	5
IEP Students (Number)	6,670	1,191	7,861
Non-Accommodated (Number)	3,396	632	4,028
Non-Accommodated (Percent)	50.9	53.1	51.2
Accommodated (Number)	3,274	559	3,833
Accommodated (Percent)	49.1	46.9	48.8

Accommodation Rate for Non-IEP and IEP Students on the Winter 2016 Keystone Exams: Literature

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	23,838	3,725	27,563
Non-Accommodated (Number)	21,293	3,582	24,875
Non-Accommodated (Percent)	89.3	96.2	90.2
Accommodated (Number)	2,545	143	2,688
Accommodated (Percent)	10.7	3.8	9.8
IEP Students (Number)	6,624	1,072	7,696
Non-Accommodated (Number)	3,314	615	3,929
Non-Accommodated (Percent)	50	57.4	51.1
Accommodated (Number)	3,310	457	3,767
Accommodated (Percent)	50	42.6	48.9

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2016/2017 Keystone: Algebra I

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	3	247	6	5
PPT - Some test items/questions read aloud (Percent)	0	3	.3	1.4
PPT - All test items/questions read aloud (Number)	9	236	3	11
PPT - All test items/questions read aloud (Percent)	0	2.8	.1	3.1
PPT - Small group setting (Number)	321	3,585	104	107
PPT - Small group setting (Percent)	.8	43	5.1	30.5
PPT - Extended time (Number)	3,141	1,082	196	34
PPT - Extended time (Percent)	8.3	13	9.6	9.7
PPT - Frequent breaks (Number)	9	213	3	6
PPT - Frequent breaks (Percent)	0	2.6	.1	1.7
PPT - Number assessed (Number)	37,926	8,345	2,049	351
CBT - Some test items/questions read aloud (Number)	2	22	8	0
CBT - Some test items/questions read aloud (Percent)	0	2.1	11.1	0
CBT - All test items/questions read aloud (Number)	1	78	0	1
CBT - All test items/questions read aloud (Percent)	0	7.3	0	14.3
CBT - Small group setting (Number)	20	464	13	3
CBT - Small group setting (Percent)	.5	43.2	18.1	42.9
CBT - Extended time (Number)	340	217	13	2
CBT - Extended time (Percent)	7.9	20.2	18.1	28.6
CBT - Frequent breaks (Number)	1	92	0	1
CBT - Frequent breaks (Percent)	0	8.6	0	14.3
CBT - Number assessed (Number)	4,320	1,073	72	7
Total - Some test items/questions read aloud (Number)	5	269	14	5
Total - Some test items/questions read aloud (Percent)	0	2.9	.7	1.4
Total - All test items/questions read aloud (Number)	10	314	3	12
Total - All test items/questions read aloud (Percent)	0	3.3	.1	3.4
Total - Small group setting (Number)	341	4,049	117	110
Total - Small group setting (Percent)	.8	43	5.5	30.7
Total - Extended time (Number)	3,481	1,299	209	36
Total - Extended time (Percent)	8.2	13.8	9.9	10.1
Total - Frequent breaks (Number)	10	305	3	7
Total - Frequent breaks (Percent)	0	3.2	.1	2
Total - Number assessed (Number)	42,246	9,418	2,121	358

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2016/2017 Keystone: Biology

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	9	257	8	7
PPT - Some test items/questions read aloud (Percent)	0	4	.6	3
PPT - All test items/questions read aloud (Number)	8	304	4	9
PPT - All test items/questions read aloud (Percent)	0	4.7	.3	3.8
PPT - Small group setting (Number)	243	2,978	88	69
PPT - Small group setting (Percent)	.9	46.3	6.8	29.4
PPT - Extended time (Number)	1,009	689	99	17
PPT - Extended time (Percent)	3.8	10.7	7.6	7.2
PPT - Frequent breaks (Number)	13	184	3	5
PPT - Frequent breaks (Percent)	0	2.9	.2	2.1
PPT - Number assessed (Number)	26,781	6,435	1,302	235
CBT - Some test items/questions read aloud (Number)	2	47	7	0
CBT - Some test items/questions read aloud (Percent)	0	4	8.1	0
CBT - All test items/questions read aloud (Number)	2	61	0	0
CBT - All test items/questions read aloud (Percent)	0	5.2	0	0
CBT - Small group setting (Number)	16	424	11	3
CBT - Small group setting (Percent)	.3	36	12.8	23.1
CBT - Extended time (Number)	155	200	10	4
CBT - Extended time (Percent)	3.3	17	11.6	30.8
CBT - Frequent breaks (Number)	2	76	0	1
CBT - Frequent breaks (Percent)	0	6.5	0	7.7
CBT - Number assessed (Number)	4,687	1,178	86	13
Total - Some test items/questions read aloud (Number)	11	304	15	7
Total - Some test items/questions read aloud (Percent)	0	4	1.1	2.8
Total - All test items/questions read aloud (Number)	10	365	4	9
Total - All test items/questions read aloud (Percent)	0	4.8	.3	3.6
Total - Small group setting (Number)	259	3,402	99	72
Total - Small group setting (Percent)	.8	44.7	7.1	29
Total - Extended time (Number)	1,164	889	109	21
Total - Extended time (Percent)	3.7	11.7	7.9	8.5
Total - Frequent breaks (Number)	15	260	3	6
Total - Frequent breaks (Percent)	0	3.4	.2	2.4
Total - Number assessed (Number)	31,468	7,613	1,388	248

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2016/2017 Keystone: Literature

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Small group setting (Number)	200	2,928	90	68
PPT - Small group setting (Percent)	.9	45.8	7.4	29.2
PPT - Extended time (Number)	2,178	919	106	22
PPT - Extended time (Percent)	9.6	14.4	8.7	9.4
PPT - Frequent breaks (Number)	6	179	5	6
PPT - Frequent breaks (Percent)	0	2.8	.4	2.6
PPT - Number assessed (Number)	22,615	6,391	1,223	233
CBT - Small group setting (Number)	12	415	12	6
CBT - Small group setting (Percent)	.3	39.2	21.4	42.9
CBT - Extended time (Number)	121	206	14	3
CBT - Extended time (Percent)	3.3	19.5	25	21.4
CBT - Frequent breaks (Number)	2	71	0	1
CBT - Frequent breaks (Percent)	.1	6.7	0	7.1
CBT - Number assessed (Number)	3,669	1,058	56	14
Total - Small group setting (Number)	212	3,343	102	74
Total - Small group setting (Percent)	.8	44.9	8	30
Total - Extended time (Number)	2,299	1,125	120	25
Total - Extended time (Percent)	8.7	15.1	9.4	10.1
Total - Frequent breaks (Number)	8	250	5	7
Total - Frequent breaks (Percent)	0	3.4	.4	2.8
Total - Number assessed (Number)	26,284	7,449	1,279	247

SUMMER 2017

Students Assessed on the Summer 2017 Keystone: Algebra I

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	3	0	7	134	476	410	74	18	1,122
Total number of CBT processed (Number)	0	0	8	67	138	139	26	0	378
Total number of tests processed (Number)	3	0	15	201	614	549	100	18	1,500
Total number of tests processed with a score (Number)	3	0	15	195	599	531	94	15	1,452
Total number of tests processed with a score (Percent)	100	0	100	97	97.6	96.7	94	83.3	96.8
Total number of tests processed without a score (Number)	0	0	0	6	15	18	6	3	48
Total number of tests processed without a score (Percent)	0	0	0	3	2.4	3.3	6	16.7	3.2

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Students Assessed on the Summer 2017 Keystone: Biology

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	2	29	241	316	92	20	700
Total number of CBT processed (Number)	0	1	88	155	36	0	280
Total number of tests processed (Number)	2	30	329	471	128	20	980
Total number of tests processed with a score (Number)	2	30	322	463	125	15	957
Total number of tests processed with a score (Percent)	100	100	97.9	98.3	97.7	75	97.7
Total number of tests processed without a score (Number)	0	0	7	8	3	5	23
Total number of tests processed without a score (Percent)	0	0	2.1	1.7	2.3	25	2.3

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Students Assessed on the Summer 2017 Keystone: Literature

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	
Total number of PPT processed (Number)	0	0	28	193	77	7	305
Total number of CBT processed (Number)	0	2	9	114	27	0	152
Total number of tests processed (Number)	0	2	37	307	104	7	457
Total number of tests processed with a score (Number)	0	2	36	300	100	7	445
Total number of tests processed with a score (Percent)	0	100	97.3	97.7	96.2	100	97.4
Total number of tests processed without a score (Number)	0	0	1	7	4	0	12
Total number of tests processed without a score (Percent)	0	0	2.7	2.3	3.8	0	2.6

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Summer 2017 Keystone: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	0	0	1	0	0	1
Extended absence from school (Percent)	0	0	0	0	0	5.6	0	0	2.1
Absent without make-up (Number)	0	0	0	0	2	0	1	0	3
Absent without make-up (Percent)	0	0	0	0	13.3	0	16.7	0	6.3
Non-attempt (Number)	0	0	0	6	11	16	5	3	41
Non-attempt (Percent)	0	0	0	100	73.3	88.9	83.3	100	85.4
Medical emergency (Number)	0	0	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	0	2	1	0	0	3
Other reasons (Percent)	0	0	0	0	13.3	5.6	0	0	6.3
Total not assessed (Number)	0	0	0	6	15	18	6	3	48

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Summer 2017 Keystone: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	1	0	0	0	1
Extended absence from school (Percent)	0	0	14.3	0	0	0	4.3
Absent without make-up (Number)	0	0	2	0	0	0	2
Absent without make-up (Percent)	0	0	28.6	0	0	0	8.7
Non-attempt (Number)	0	0	4	8	3	5	20
Non-attempt (Percent)	0	0	57.1	100	100	100	87
Medical emergency (Number)	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	0	0	0	0
Other reasons (Percent)	0	0	0	0	0	0	0
Total not assessed (Number)	0	0	7	8	3	5	23

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Counts of Students without Scores on the Summer 2017 Keystone: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	0	1	0	1
Extended absence from school (Percent)	0	0	0	0	25	0	8.3
Absent without make-up (Number)	0	0	1	1	0	0	2
Absent without make-up (Percent)	0	0	100	14.3	0	0	16.7
Non-attempt (Number)	0	0	0	5	3	0	8
Non-attempt (Percent)	0	0	0	71.4	75	0	66.7
ELL in first year in U.S. schools (Number)	0	0	0	0	0	0	0
ELL in first year in U.S. schools (Percent)	0	0	0	0	0	0	0
Medical emergency (Number)	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	1	0	0	1
Other reasons (Percent)	0	0	0	14.3	0	0	8.3
Total not assessed (Number)	0	0	1	7	4	0	12

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Summer 2017 Keystone: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	1	0	5	95	307	272	40	9	729
Female (Percent)	33.3	0	33.3	48.7	51.3	51.2	42.6	60	50.2
Male (Number)	1	0	10	100	289	255	53	6	714
Male (Percent)	33.3	0	66.7	51.3	48.2	48	56.4	40	49.2
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	0	0	1	0	0	1
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	0	0	.2	0	0	.1
Asian (not Hispanic) (Number)	0	0	4	9	22	11	3	1	50
Asian (not Hispanic) (Percent)	0	0	26.7	4.6	3.7	2.1	3.2	6.7	3.4
Black or African American (not Hispanic) (Number)	0	0	0	21	63	119	14	1	218
Black or African American (not Hispanic) (Percent)	0	0	0	10.8	10.5	22.4	14.9	6.7	15
Hispanic (any race) (Number)	0	0	0	15	52	41	7	0	115
Hispanic (any race) (Percent)	0	0	0	7.7	8.7	7.7	7.4	0	7.9
Multi-Racial (not Hispanic) (Number)	0	0	1	3	10	18	0	0	32
Multi-Racial (not Hispanic) (Percent)	0	0	6.7	1.5	1.7	3.4	0	0	2.2
White (not Hispanic) (Number)	2	0	10	146	446	334	67	12	1,017
White (not Hispanic) (Percent)	66.7	0	66.7	74.9	74.5	62.9	71.3	80	70
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	2	0	1	0	3
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	.3	0	1.1	0	.2
IEP (not gifted) (Number)	0	0	1	6	82	90	25	5	209
IEP (not gifted) (Percent)	0	0	6.7	3.1	13.7	16.9	26.6	33.3	14.4
Student exited IEP in last 2 years (Number)	0	0	0	2	6	7	1	0	16
Student exited IEP in last 2 years (Percent)	0	0	0	1	1	1.3	1.1	0	1.1
Title I (Number)	0	0	1	6	33	148	10	0	198
Title I (Percent)	0	0	6.7	3.1	5.5	27.9	10.6	0	13.6
Title III served (Number)	0	0	0	0	3	7	2	0	12
Title III served (Percent)	0	0	0	0	.5	1.3	2.1	0	.8
Title III not served (Number)	0	0	4	22	87	59	9	0	181
Title III not served (Percent)	0	0	26.7	11.3	14.5	11.1	9.6	0	12.5
Migrant student (Number)	0	0	0	0	1	0	0	0	1
Migrant student (Percent)	0	0	0	0	.2	0	0	0	.1
ELL (enrolled after 8/5/16) (Number)	0	0	0	0	1	3	2	0	6
ELL (enrolled after 8/5/16) (Percent)	0	0	0	0	.2	.6	2.1	0	.4
ELL (enrolled on or before 8/5/16) (Number)	0	0	0	0	2	6	1	0	9
ELL (enrolled on or before 8/5/16) (Percent)	0	0	0	0	.3	1.1	1.1	0	.6

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	1	1	2	0	0	4
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	.5	.2	.4	0	0	.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	1	1	1	0	0	3
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	.5	.2	.2	0	0	.2
Former ELL no longer monitored (Number)	0	0	0	0	8	5	0	0	13
Former ELL no longer monitored (Percent)	0	0	0	0	1.3	.9	0	0	.9
Foreign exchange student (Number)	0	0	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	0	1	19	99	201	30	3	353
Economically disadvantaged (Percent)	0	0	6.7	9.7	16.5	37.9	31.9	20	24.3
Historically Underperforming Subgroup (Number)	0	0	2	25	173	249	52	7	508
Historically Underperforming Subgroup (Percent)	0	0	13.3	12.8	28.9	46.9	55.3	46.7	35
Enrollment in school of residence after 10/1/16 (Number)	0	0	0	13	42	23	2	1	81
Enrollment in school of residence after 10/1/16 (Percent)	0	0	0	6.7	7	4.3	2.1	6.7	5.6
Enrollment in district of residence after 10/1/16 (Number)	0	0	0	4	20	20	2	1	47
Enrollment in district of residence after 10/1/16 (Percent)	0	0	0	2.1	3.3	3.8	2.1	6.7	3.2
Enrollment as PA resident after 10/1/16 (Number)	0	0	0	2	15	10	1	0	28
Enrollment as PA resident after 10/1/16 (Percent)	0	0	0	1	2.5	1.9	1.1	0	1.9
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	1	4	101	23	4	0	133
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	6.7	2.1	16.9	4.3	4.3	0	9.2
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	1	4	21	24	7	0	57
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	6.7	2.1	3.5	4.5	7.4	0	3.9
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0	0	0
Number of assessed students (Number)	3	0	15	195	599	531	94	15	1,452

*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Summer 2017 Keystone: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	2	16	150	230	64	8	470
Female (Percent)	100	53.3	46.6	49.7	51.2	53.3	49.1
Male (Number)	0	14	145	232	61	7	459
Male (Percent)	0	46.7	45	50.1	48.8	46.7	48
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	2	0	0	2
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	.4	0	0	.2
Asian (not Hispanic) (Number)	1	11	9	23	5	2	51
Asian (not Hispanic) (Percent)	50	36.7	2.8	5	4	13.3	5.3
Black or African American (not Hispanic) (Number)	0	1	35	49	9	0	94
Black or African American (not Hispanic) (Percent)	0	3.3	10.9	10.6	7.2	0	9.8
Hispanic (any race) (Number)	0	1	37	49	22	0	109
Hispanic (any race) (Percent)	0	3.3	11.5	10.6	17.6	0	11.4
Multi-Racial (not Hispanic) (Number)	0	2	7	10	0	0	19
Multi-Racial (not Hispanic) (Percent)	0	6.7	2.2	2.2	0	0	2
White (not Hispanic) (Number)	1	15	205	327	85	7	640
White (not Hispanic) (Percent)	50	50	63.7	70.6	68	46.7	66.9
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	1	1	2	0	4
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.3	.2	1.6	0	.4
IEP (not gifted) (Number)	0	1	48	74	15	2	140
IEP (not gifted) (Percent)	0	3.3	14.9	16	12	13.3	14.6
Student exited IEP in last 2 years (Number)	0	2	3	8	0	0	13
Student exited IEP in last 2 years (Percent)	0	6.7	.9	1.7	0	0	1.4
Title I (Number)	0	0	8	11	19	0	38
Title I (Percent)	0	0	2.5	2.4	15.2	0	4
Title III served (Number)	0	0	2	2	1	0	5
Title III served (Percent)	0	0	.6	.4	.8	0	.5
Title III not served (Number)	0	0	45	72	7	0	124
Title III not served (Percent)	0	0	14	15.6	5.6	0	13
Migrant student (Number)	0	0	0	0	0	0	0
Migrant student (Percent)	0	0	0	0	0	0	0
ELL (enrolled after 8/5/16) (Number)	0	0	0	1	2	0	3
ELL (enrolled after 8/5/16) (Percent)	0	0	0	.2	1.6	0	.3
ELL (enrolled on or before 8/5/16) (Number)	0	0	2	2	4	0	8
ELL (enrolled on or before 8/5/16) (Percent)	0	0	.6	.4	3.2	0	.8
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	0	0	0	0

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	0	0	0	0
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	1	0	3	0	1	5
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	3.3	0	.6	0	6.7	.5
Former ELL no longer monitored (Number)	0	0	3	8	2	0	13
Former ELL no longer monitored (Percent)	0	0	.9	1.7	1.6	0	1.4
Foreign exchange student (Number)	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	4	64	106	43	1	218
Economically disadvantaged (Percent)	0	13.3	19.9	22.9	34.4	6.7	22.8
Historically Underperforming Subgroup (Number)	0	5	101	164	56	2	328
Historically Underperforming Subgroup (Percent)	0	16.7	31.4	35.4	44.8	13.3	34.3
Enrollment in school of residence after 10/1/16 (Number)	0	0	9	19	3	1	32
Enrollment in school of residence after 10/1/16 (Percent)	0	0	2.8	4.1	2.4	6.7	3.3
Enrollment in district of residence after 10/1/16 (Number)	0	0	8	15	2	1	26
Enrollment in district of residence after 10/1/16 (Percent)	0	0	2.5	3.2	1.6	6.7	2.7
Enrollment as PA resident after 10/1/16 (Number)	0	0	4	7	1	0	12
Enrollment as PA resident after 10/1/16 (Percent)	0	0	1.2	1.5	.8	0	1.3
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	32	46	3	0	81
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	9.9	9.9	2.4	0	8.5
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	1	10	19	4	0	34
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	3.3	3.1	4.1	3.2	0	3.6
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0
Number of assessed students (Number)	2	30	322	463	125	15	957

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Demographic Characteristics of Students taking the Summer 2017 Keystone: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Female (Number)	0	0	15	111	42	3	171
Female (Percent)	0	0	41.7	37	42	42.9	38.4
Male (Number)	0	2	21	189	58	4	274
Male (Percent)	0	100	58.3	63	58	57.1	61.6
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	1	0	0	1
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	.3	0	0	.2
Asian (not Hispanic) (Number)	0	0	0	20	6	1	27
Asian (not Hispanic) (Percent)	0	0	0	6.7	6	14.3	6.1
Black or African American (not Hispanic) (Number)	0	0	1	33	7	0	41
Black or African American (not Hispanic) (Percent)	0	0	2.8	11	7	0	9.2
Hispanic (any race) (Number)	0	0	1	19	16	0	36
Hispanic (any race) (Percent)	0	0	2.8	6.3	16	0	8.1
Multi-Racial (not Hispanic) (Number)	0	1	0	4	1	0	6
Multi-Racial (not Hispanic) (Percent)	0	50	0	1.3	1	0	1.3
White (not Hispanic) (Number)	0	1	34	223	70	5	333
White (not Hispanic) (Percent)	0	50	94.4	74.3	70	71.4	74.8
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	0	0	0
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	0	0	0
IEP (not gifted) (Number)	0	0	2	74	12	3	91
IEP (not gifted) (Percent)	0	0	5.6	24.7	12	42.9	20.4
Student exited IEP in last 2 years (Number)	0	0	0	3	3	0	6
Student exited IEP in last 2 years (Percent)	0	0	0	1	3	0	1.3
Title I (Number)	0	0	0	1	30	0	31
Title I (Percent)	0	0	0	.3	30	0	7
Title III served (Number)	0	0	0	7	0	0	7
Title III served (Percent)	0	0	0	2.3	0	0	1.6
Title III not served (Number)	0	0	1	48	6	0	55
Title III not served (Percent)	0	0	2.8	16	6	0	12.4
Migrant student (Number)	0	0	0	0	2	0	2
Migrant student (Percent)	0	0	0	0	2	0	.4
ELL (enrolled after 8/5/16) (Number)	0	0	0	4	0	0	4
ELL (enrolled after 8/5/16) (Percent)	0	0	0	1.3	0	0	.9
ELL (enrolled on or before 8/5/16) (Number)	0	0	0	5	6	0	11
ELL (enrolled on or before 8/5/16) (Percent)	0	0	0	1.7	6	0	2.5

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	1	0	0	1
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	.3	0	0	.2
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	1	0	1
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	1	0	.2
Former ELL no longer monitored (Number)	0	0	1	5	3	0	9
Former ELL no longer monitored (Percent)	0	0	2.8	1.7	3	0	2
Foreign exchange student (Number)	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	0	4	61	38	1	104
Economically disadvantaged (Percent)	0	0	11.1	20.3	38	14.3	23.4
Historically Underperforming Subgroup (Number)	0	0	5	117	45	3	170
Historically Underperforming Subgroup (Percent)	0	0	13.9	39	45	42.9	38.2
Enrollment in school of residence after 10/1/16 (Number)	0	0	0	6	3	1	10
Enrollment in school of residence after 10/1/16 (Percent)	0	0	0	2	3	14.3	2.2
Enrollment in district of residence after 10/1/16 (Number)	0	0	0	5	3	1	9
Enrollment in district of residence after 10/1/16 (Percent)	0	0	0	1.7	3	14.3	2
Enrollment as PA resident after 10/1/16 (Number)	0	0	0	5	2	0	7
Enrollment as PA resident after 10/1/16 (Percent)	0	0	0	1.7	2	0	1.6
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	5	21	1	0	27
Enrollment in school of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	13.9	7	1	0	6.1
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Number)	0	0	4	11	0	0	15
Enrollment in district of residence after 10/1/15 but on/before 10/1/16 (Percent)	0	0	11.1	3.7	0	0	3.4
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0
Number of assessed students (Number)	0	2	36	300	100	7	445

*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

Incidence of Presentation Accommodations Received on the Summer 2017 Keystone: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	1	N/A	1
Braille format (Percent)	.1	N/A	.1
Large print format (Number)	1	N/A	1
Large print format (Percent)	.1	N/A	.1
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	3	0	3
Some test items/questions read aloud (Percent)	.3	0	.2
All test items/questions read aloud (Number)	2	0	2
All test items/questions read aloud (Percent)	.2	0	.1
Test items/questions signed (Number)	0	0	0
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	0	0	0
Test items/questions interpreted for ELL student (Percent)	0	0	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Spanish version (Number)	1	N/A	1
Spanish version (Percent)	.1	N/A	.1
Audio (Number)	N/A	5	5
Audio (Percent)	N/A	1.3	.3
Color Chooser (Number)	N/A	0	0
Color Chooser (Percent)	N/A	0	0
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast – only Spring and Summer (Number)	N/A	0	0
Reverse Contrast – only Spring and Summer (Percent)	N/A	0	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	1,076	376	1,452

Incidence of Presentation Accommodations Received on the Summer 2017 Keystone: Biology

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	0	N/A	0
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	3	0	3
Some test items/questions read aloud (Percent)	.4	0	.3
All test items/questions read aloud (Number)	3	0	3
All test items/questions read aloud (Percent)	.4	0	.3
Test items/questions signed (Number)	0	0	0
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for ELL student (Number)	0	0	0
Test items/questions interpreted for ELL student (Percent)	0	0	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Spanish version (Number)	0	N/A	0
Spanish version (Percent)	0	N/A	0
Audio (Number)	N/A	4	4
Audio (Percent)	N/A	1.5	.4
Color Chooser (Number)	N/A	1	1
Color Chooser (Percent)	N/A	.4	.1
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast – only Spring and Summer (Number)	N/A	0	0
Reverse Contrast – only Spring and Summer (Percent)	N/A	0	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	682	275	957

Incidence of Presentation Accommodations Received on the Summer 2017 Keystone: Literature

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	0	N/A	0
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Color Chooser (Number)	N/A	2	2
Color Chooser (Percent)	N/A	1.3	.4
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast – only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Number)	N/A	0	0
Refreshable Braille - only Spring and Summer (Percent)	N/A	0	0
Number of assessed students (Number)	294	151	445

Incidence of Response Accommodations Received on the Summer 2017 Keystone: Algebra I

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	2	0	2
Test administrator marked multiple-choice responses at student's direction (Percent)	.2	0	.1
Test administrator scribed open-ended responses at student's direction (Number)	0	0	0
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	3	0	3
Test administrator transcribed student responses (Percent)	.3	0	.2
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	1	N/A	1
Brailler/Notetaker (Percent)	.1	N/A	.1
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	2	0	2
Translation dictionary for ELL student (Percent)	.2	0	.1
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	1,076	376	1,452

Incidence of Response Accommodations Received on the Summer 2017 Keystone: Biology

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	2	0	2
Test administrator marked multiple-choice responses at student's direction (Percent)	.3	0	.2
Test administrator scribed open-ended responses at student's direction (Number)	0	0	0
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	0	0	0
Test administrator transcribed student responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	0	N/A	0
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for ELL student (Number)	0	0	0
Translation dictionary for ELL student (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	682	275	957

Incidence of Response Accommodations Received on the Summer 2017 Keystone: Literature

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	0	0	0
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	1	0	1
Test administrator scribed open-ended responses at student's direction (Percent)	.3	0	.2
Test administrator transcribed student responses (Number)	0	0	0
Test administrator transcribed student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Brailler/Notetaker (Number)	0	N/A	0
Brailler/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Audio recording of student responses (Number)	0	0	0
Audio recording of student responses (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	294	151	445

Incidence of Setting Accommodations Received on the Summer 2017 Keystone: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	2	1	3
One-on-one setting (Percent)	.2	.3	.2
Small group setting (Number)	32	12	44
Small group setting (Percent)	3	3.2	3
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	1,076	376	1,452

Incidence of Setting Accommodations Received on the Summer 2017 Keystone: Biology

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	1	0	1
One-on-one setting (Percent)	.1	0	.1
Small group setting (Number)	22	10	32
Small group setting (Percent)	3.2	3.6	3.3
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	682	275	957

Incidence of Setting Accommodations Received on the Summer 2017 Keystone: Literature

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	2	1	3
One-on-one setting (Percent)	.7	.7	.7
Small group setting (Number)	18	6	24
Small group setting (Percent)	6.1	4	5.4
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	294	151	445

Incidence of Timing Accommodations Received on the Summer 2017 Keystone: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	25	11	36
Extended time (Percent)	2.3	2.9	2.5
Frequent breaks (Number)	3	7	10
Frequent breaks (Percent)	.3	1.9	.7
Changed test schedule (Number)	0	1	1
Changed test schedule (Percent)	0	.3	.1
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	1,076	376	1,452

Incidence of Timing Accommodations Received on the Summer 2017 Keystone: Biology

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	4	10	14
Extended time (Percent)	.6	3.6	1.5
Frequent breaks (Number)	0	6	6
Frequent breaks (Percent)	0	2.2	.6
Changed test schedule (Number)	0	0	0
Changed test schedule (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	682	275	957

Incidence of Timing Accommodations Received on the Summer 2017 Keystone: Literature

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	16	7	23
Extended time (Percent)	5.4	4.6	5.2
Frequent breaks (Number)	0	4	4
Frequent breaks (Percent)	0	2.6	.9
Changed test schedule (Number)	0	0	0
Changed test schedule (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	294	151	445

Accommodation Rate for Non-IEP and IEP Students on the Summer 2017 Keystone Exams: Algebra I

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	914	329	1,243
Non-Accommodated (Number)	893	326	1,219
Non-Accommodated (Percent)	97.7	99.1	98.1
Accommodated (Number)	21	3	24
Accommodated (Percent)	2.3	.9	1.9
IEP Students (Number)	162	47	209
Non-Accommodated (Number)	129	36	165
Non-Accommodated (Percent)	79.6	76.6	78.9
Accommodated (Number)	33	11	44
Accommodated (Percent)	20.4	23.4	21.1

Accommodation Rate for Non-IEP and IEP Students on the Summer 2017 Keystone Exams: Biology

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	573	244	817
Non-Accommodated (Number)	568	241	809
Non-Accommodated (Percent)	99.1	98.8	99
Accommodated (Number)	5	3	8
Accommodated (Percent)	.9	1.2	1
IEP Students (Number)	109	31	140
Non-Accommodated (Number)	88	23	111
Non-Accommodated (Percent)	80.7	74.2	79.3
Accommodated (Number)	21	8	29
Accommodated (Percent)	19.3	25.8	20.7

Accommodation Rate for Non-IEP and IEP Students on the Summer 2017 Keystone Exams: Literature

Student Subgroup Tested	PPT	CBT	Total
Non-IEP Students (Number)	230	124	354
Non-Accommodated (Number)	219	122	341
Non-Accommodated (Percent)	95.2	98.4	96.3
Accommodated (Number)	11	2	13
Accommodated (Percent)	4.8	1.6	3.7
IEP Students (Number)	64	27	91
Non-Accommodated (Number)	43	22	65
Non-Accommodated (Percent)	67.2	81.5	71.4
Accommodated (Number)	21	5	26
Accommodated (Percent)	32.8	18.5	28.6

Incidence of IEP and ELL Students Receiving Accommodations on the Summer 2017 Keystone: Algebra I

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	0	3	0	0
PPT - Some test items/questions read aloud (Percent)	0	1.9	0	0
PPT - All test items/questions read aloud (Number)	0	2	0	0
PPT - All test items/questions read aloud (Percent)	0	1.2	0	0
PPT - Small group setting (Number)	4	27	0	1
PPT - Small group setting (Percent)	.4	16.8	0	100
PPT - Extended time (Number)	16	8	0	1
PPT - Extended time (Percent)	1.8	5	0	100
PPT - Frequent breaks (Number)	0	3	0	0
PPT - Frequent breaks (Percent)	0	1.9	0	0
PPT - Number assessed (Number)	904	161	10	1
CBT - Some test items/questions read aloud (Number)	0	0	0	0
CBT - Some test items/questions read aloud (Percent)	0	0	0	0
CBT - All test items/questions read aloud (Number)	0	0	0	0
CBT - All test items/questions read aloud (Percent)	0	0	0	0
CBT - Small group setting (Number)	2	10	0	0
CBT - Small group setting (Percent)	.6	21.7	0	0
CBT - Extended time (Number)	1	10	0	0
CBT - Extended time (Percent)	.3	21.7	0	0
CBT - Frequent breaks (Number)	2	5	0	0
CBT - Frequent breaks (Percent)	.6	10.9	0	0
CBT - Number assessed (Number)	326	46	3	1
Total - Some test items/questions read aloud (Number)	0	3	0	0
Total - Some test items/questions read aloud (Percent)	0	1.4	0	0
Total - All test items/questions read aloud (Number)	0	2	0	0
Total - All test items/questions read aloud (Percent)	0	1	0	0
Total - Small group setting (Number)	6	37	0	1
Total - Small group setting (Percent)	.5	17.9	0	50
Total - Extended time (Number)	17	18	0	1
Total - Extended time (Percent)	1.4	8.7	0	50
Total - Frequent breaks (Number)	2	8	0	0
Total - Frequent breaks (Percent)	.2	3.9	0	0
Total - Number assessed (Number)	1,230	207	13	2

Incidence of IEP and ELL Students Receiving Accommodations on the Summer 2017 Keystone: Biology

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Some test items/questions read aloud (Number)	0	3	0	0
PPT - Some test items/questions read aloud (Percent)	0	2.8	0	0
PPT - All test items/questions read aloud (Number)	2	1	0	0
PPT - All test items/questions read aloud (Percent)	.4	.9	0	0
PPT - Small group setting (Number)	4	18	0	0
PPT - Small group setting (Percent)	.7	16.5	0	0
PPT - Extended time (Number)	1	3	0	0
PPT - Extended time (Percent)	.2	2.8	0	0
PPT - Frequent breaks (Number)	0	0	0	0
PPT - Frequent breaks (Percent)	0	0	0	0
PPT - Number assessed (Number)	567	109	6	0
CBT - Some test items/questions read aloud (Number)	0	0	0	0
CBT - Some test items/questions read aloud (Percent)	0	0	0	0
CBT - All test items/questions read aloud (Number)	0	0	0	0
CBT - All test items/questions read aloud (Percent)	0	0	0	0
CBT - Small group setting (Number)	2	7	0	1
CBT - Small group setting (Percent)	.8	24.1	0	50
CBT - Extended time (Number)	2	7	0	1
CBT - Extended time (Percent)	.8	24.1	0	50
CBT - Frequent breaks (Number)	2	3	0	1
CBT - Frequent breaks (Percent)	.8	10.3	0	50
CBT - Number assessed (Number)	241	29	3	2
Total - Some test items/questions read aloud (Number)	0	3	0	0
Total - Some test items/questions read aloud (Percent)	0	2.2	0	0
Total - All test items/questions read aloud (Number)	2	1	0	0
Total - All test items/questions read aloud (Percent)	.2	.7	0	0
Total - Small group setting (Number)	6	25	0	1
Total - Small group setting (Percent)	.7	18.1	0	50
Total - Extended time (Number)	3	10	0	1
Total - Extended time (Percent)	.4	7.2	0	50
Total - Frequent breaks (Number)	2	3	0	1
Total - Frequent breaks (Percent)	.2	2.2	0	50
Total - Number assessed (Number)	808	138	9	2

Incidence of IEP and ELL Students Receiving Accommodations on the Summer 2017 Keystone: Literature

Accommodation Received by Administration Mode	General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
PPT - Small group setting (Number)	3	14	0	1
PPT - Small group setting (Percent)	1.4	22.6	0	50
PPT - Extended time (Number)	8	7	0	1
PPT - Extended time (Percent)	3.6	11.3	0	50
PPT - Frequent breaks (Number)	0	0	0	0
PPT - Frequent breaks (Percent)	0	0	0	0
PPT - Number assessed (Number)	221	62	9	2
CBT - Small group setting (Number)	2	4	0	0
CBT - Small group setting (Percent)	1.7	14.8	0	0
CBT - Extended time (Number)	2	5	0	0
CBT - Extended time (Percent)	1.7	18.5	0	0
CBT - Frequent breaks (Number)	1	3	0	0
CBT - Frequent breaks (Percent)	.8	11.1	0	0
CBT - Number assessed (Number)	120	27	4	0
Total - Small group setting (Number)	5	18	0	1
Total - Small group setting (Percent)	1.5	20.2	0	50
Total - Extended time (Number)	10	12	0	1
Total - Extended time (Percent)	2.9	13.5	0	50
Total - Frequent breaks (Number)	1	3	0	0
Total - Frequent breaks (Percent)	.3	3.4	0	0
Total - Number assessed (Number)	341	89	13	2

Table J–1. Item Statistics

Column Heading	Definition
Ref	Reference line number
ID	Item ID
Form	Test form
N	Number of students
PVal	*P*-Value
P( )	Proportion selecting given response (- = blank)
ITCorr	Item total correlation
Corr( )	Correlation of options/points and total test score
Meas	Rasch item difficulty measure estimate
SEM	Standard error of Rasch item difficulty measure estimate
z-Infit	*z* infit statistic
MS-Infit	Mean square infit statistic
z-Outfit	*z* outfit statistic
MS-Outfit	Mean square outfit statistic
M/F	Male/Female DIF code
W/B	White/Black DIF code
W/H	White/Hispanic DIF code
O/P	Online computer-based test/paper-pencil-based test DIF code

ALGEBRA I MULTIPLE-CHOICE ITEMS

Table J–2. Algebra I Multiple-Choice Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	640543	0	55500	0.72	0.07	0.16	0.72	0.05	0.00	0.00	0.28	-0.22	-0.07	0.28	-0.20	-1.07	0.01	3.84	1.02	9.07	1.08
2	640528	0	55500	0.43	0.43	0.14	0.20	0.22	0.00	0.00	0.32	0.32	-0.09	-0.20	-0.10	0.55	0.01	9.90	1.07	9.90	1.11
3	703971	0	55500	0.44	0.39	0.06	0.44	0.11	0.00	0.00	0.24	-0.04	-0.17	0.24	-0.18	0.23	0.01	9.90	1.10	9.90	1.13
4	674501	0	55500	0.35	0.11	0.48	0.07	0.35	0.00	0.00	0.41	-0.23	-0.15	-0.18	0.41	0.68	0.01	-9.90	0.93	-9.90	0.93
5	605092	0	55500	0.36	0.14	0.23	0.36	0.26	0.00	0.00	0.28	-0.14	-0.18	0.28	-0.02	0.78	0.01	9.90	1.09	9.90	1.17
6	640530	0	55500	0.70	0.70	0.04	0.09	0.17	0.00	0.00	0.47	0.47	-0.16	-0.21	-0.34	-0.89	0.01	-9.90	0.82	-9.90	0.75
7	640563	0	55500	0.43	0.14	0.43	0.24	0.17	0.00	0.00	0.25	-0.10	0.25	-0.11	-0.10	0.40	0.01	9.90	1.11	9.90	1.16
8	704018	0	55500	0.57	0.23	0.57	0.14	0.05	0.00	0.00	0.32	-0.17	0.32	-0.14	-0.16	-0.35	0.01	6.38	1.02	5.43	1.03
9	700861	0	55500	0.48	0.09	0.22	0.21	0.48	0.00	0.00	0.39	-0.15	-0.14	-0.23	0.39	0.05	0.01	-9.90	0.97	-9.39	0.95
10	700819	0	55500	0.59	0.14	0.18	0.59	0.07	0.01	0.00	0.24	-0.11	-0.11	0.24	-0.11	-0.07	0.01	9.90	1.09	9.90	1.12
11	700855	0	55500	0.49	0.49	0.22	0.18	0.11	0.01	0.00	0.36	0.36	-0.18	-0.13	-0.16	-0.09	0.01	0.32	1.00	1.71	1.01
12	640597	0	55500	0.63	0.12	0.63	0.17	0.07	0.01	0.00	0.42	-0.22	0.42	-0.23	-0.15	-0.78	0.01	-9.90	0.96	-9.90	0.93
13	640540	0	55500	0.51	0.23	0.09	0.16	0.51	0.01	0.00	0.30	-0.07	-0.22	-0.13	0.30	0.19	0.01	9.90	1.07	9.90	1.07
14	700818	0	55500	0.35	0.21	0.35	0.23	0.20	0.01	0.00	0.20	-0.10	0.20	-0.01	-0.10	1.00	0.01	9.90	1.22	9.90	1.36
15	640579	0	55500	0.27	0.27	0.19	0.30	0.23	0.01	0.00	0.17	0.17	-0.10	-0.04	-0.02	1.41	0.01	9.90	1.23	9.90	1.47
16	704019	0	55500	0.41	0.23	0.17	0.18	0.41	0.01	0.00	0.37	-0.15	-0.11	-0.18	0.37	0.27	0.01	-9.17	0.97	-6.39	0.97
17	674514	0	55500	0.44	0.44	0.20	0.25	0.10	0.01	0.00	0.41	0.41	-0.17	-0.16	-0.17	-0.20	0.01	2.00	1.01	-1.13	0.99
18	700870	0	55500	0.67	0.09	0.12	0.67	0.11	0.01	0.00	0.43	-0.22	-0.22	0.43	-0.19	-1.08	0.01	-3.51	0.98	-7.27	0.94
19	704022	0	55500	0.65	0.65	0.15	0.13	0.07	0.00	0.00	0.46	0.46	-0.24	-0.24	-0.21	-0.93	0.01	-9.90	0.93	-9.90	0.86
20	704003	0	55500	0.69	0.69	0.09	0.17	0.05	0.00	0.00	0.36	0.36	-0.21	-0.18	-0.18	-0.79	0.01	-9.90	0.93	-9.90	0.90
21	641466	0	55500	0.52	0.52	0.21	0.16	0.10	0.00	0.00	0.33	0.33	-0.17	-0.17	-0.10	-0.35	0.01	9.90	1.05	9.90	1.07
22	641477	0	55500	0.42	0.09	0.22	0.27	0.42	0.00	0.00	0.16	-0.18	-0.07	0.01	0.16	0.67	0.01	9.90	1.25	9.90	1.34
23	605102	0	55500	0.52	0.12	0.52	0.17	0.20	0.00	0.00	0.21	-0.08	0.21	-0.11	-0.09	0.20	0.01	9.90	1.16	9.90	1.21
24	696810	0	55500	0.52	0.19	0.52	0.13	0.16	0.00	0.00	0.36	-0.12	0.36	-0.18	-0.20	0.04	0.01	-1.78	0.99	-0.82	1.00
25	702529	0	55500	0.32	0.25	0.21	0.32	0.22	0.00	0.00	0.13	-0.01	-0.12	0.13	-0.01	1.27	0.01	9.90	1.34	9.90	1.66
26	702528	0	55500	0.56	0.15	0.16	0.56	0.13	0.00	0.00	0.33	-0.24	-0.15	0.33	-0.06	-0.16	0.01	1.09	1.00	4.72	1.02
27	641437	0	55500	0.56	0.17	0.22	0.04	0.56	0.00	0.00	0.36	-0.26	-0.10	-0.19	0.36	0.00	0.01	-2.52	0.99	-0.29	1.00
28	641502	0	55500	0.39	0.26	0.16	0.19	0.39	0.01	0.00	0.28	-0.08	-0.15	-0.10	0.28	0.66	0.01	9.90	1.09	9.90	1.13

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
29	703979	0	55500	0.55	0.55	0.16	0.19	0.10	0.01	0.00	0.42	0.42	-0.23	-0.15	-0.19	-0.39	0.01	-9.90	0.95	-9.90	0.93
30	657752	0	55500	0.60	0.11	0.60	0.12	0.17	0.01	0.00	0.41	-0.13	0.41	-0.24	-0.20	-0.47	0.01	-9.90	0.93	-9.90	0.92
31	702537	0	55500	0.52	0.21	0.52	0.09	0.17	0.01	0.00	0.40	-0.18	0.40	-0.19	-0.16	-0.27	0.01	-5.21	0.98	-6.52	0.97
32	678791	0	55500	0.55	0.08	0.10	0.55	0.27	0.01	0.00	0.32	-0.22	-0.25	0.32	-0.04	-0.12	0.01	7.26	1.02	5.78	1.03
33	703976	0	55500	0.47	0.13	0.17	0.47	0.22	0.01	0.00	0.35	-0.13	-0.18	0.35	-0.13	0.19	0.01	3.79	1.01	4.21	1.02
34	641523	0	55500	0.55	0.09	0.09	0.26	0.55	0.01	0.00	0.45	-0.14	-0.22	-0.26	0.45	-0.05	0.01	-9.90	0.90	-9.90	0.88
35	703345	0	55500	0.23	0.15	0.23	0.10	0.52	0.01	0.00	0.19	-0.12	0.19	-0.20	0.06	1.37	0.01	9.90	1.06	9.90	1.31
36	641481	0	55500	0.46	0.21	0.11	0.46	0.22	0.01	0.00	0.35	-0.25	-0.15	0.35	-0.04	-0.59	0.01	9.90	1.20	9.90	1.28

Table J–3. Biology Multiple-Choice Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	643379	0	41916	0.75	0.04	0.05	0.16	0.75	0.00	0.00	0.43	-0.22	-0.22	-0.26	0.43	-1.33	0.01	-9.90	0.82	-9.90	0.72
2	703485	0	41916	0.42	0.29	0.14	0.14	0.42	0.00	0.00	0.40	-0.17	-0.16	-0.18	0.40	-0.08	0.01	-6.89	0.97	-6.46	0.96
3	703515	0	41916	0.39	0.13	0.39	0.12	0.36	0.00	0.00	0.29	-0.09	0.29	-0.21	-0.09	0.65	0.01	9.90	1.17	9.90	1.26
4	703358	0	41916	0.42	0.42	0.23	0.21	0.14	0.00	0.00	0.38	0.38	-0.12	-0.15	-0.22	0.16	0.01	-0.13	1.00	-1.28	0.99
5	680559	0	41916	0.54	0.09	0.18	0.19	0.54	0.00	0.00	0.39	-0.19	-0.20	-0.16	0.39	-0.11	0.01	-2.36	0.99	-5.85	0.96
6	642841	0	41916	0.55	0.55	0.24	0.10	0.11	0.00	0.00	0.36	0.36	-0.16	-0.23	-0.14	-0.38	0.01	0.29	1.00	-3.94	0.97
7	703516	0	41916	0.36	0.22	0.25	0.36	0.18	0.00	0.00	0.29	-0.09	-0.12	0.29	-0.13	0.97	0.01	9.90	1.26	9.90	1.42
8	678540	0	41916	0.49	0.14	0.49	0.23	0.13	0.00	0.00	0.40	-0.19	0.40	-0.18	-0.17	0.17	0.01	4.20	1.02	1.85	1.01
9	642320	0	41916	0.55	0.11	0.22	0.55	0.12	0.00	0.00	0.28	-0.25	-0.06	0.28	-0.12	-0.27	0.01	9.90	1.07	9.90	1.14
10	643772	0	41916	0.63	0.08	0.11	0.18	0.63	0.00	0.00	0.34	-0.22	-0.20	-0.11	0.34	-0.70	0.01	-5.83	0.98	1.20	1.01
11	702614	0	41916	0.57	0.57	0.19	0.13	0.11	0.00	0.00	0.24	0.24	-0.08	-0.15	-0.11	-0.11	0.01	9.90	1.14	9.90	1.17
12	703268	0	41916	0.41	0.20	0.14	0.24	0.41	0.00	0.00	0.37	-0.15	-0.15	-0.16	0.37	0.16	0.01	2.03	1.01	3.30	1.02
13	641270	0	41916	0.51	0.51	0.14	0.13	0.21	0.01	0.00	0.41	0.41	-0.19	-0.16	-0.18	-0.23	0.01	-9.01	0.97	-9.90	0.94
14	680194	0	41916	0.53	0.08	0.11	0.53	0.27	0.01	0.00	0.46	-0.22	-0.23	0.46	-0.20	-0.54	0.01	-9.90	0.94	-9.90	0.89
15	703483	0	41916	0.49	0.49	0.21	0.15	0.14	0.01	0.00	0.37	0.37	-0.12	-0.23	-0.14	0.17	0.01	9.90	1.05	6.60	1.04
16	702072	0	41916	0.33	0.15	0.33	0.28	0.23	0.01	0.00	0.30	-0.15	0.30	-0.23	0.06	0.51	0.01	9.90	1.05	9.90	1.11
17	703493	0	41916	0.56	0.56	0.26	0.10	0.08	0.01	0.00	0.46	0.46	-0.24	-0.23	-0.17	-0.71	0.01	-9.90	0.96	-9.90	0.89
18	642861	0	41916	0.47	0.14	0.21	0.47	0.17	0.01	0.00	0.35	-0.20	-0.15	0.35	-0.09	-0.09	0.01	7.02	1.03	2.60	1.02
19	642314	0	41916	0.52	0.52	0.15	0.16	0.16	0.01	0.00	0.47	0.47	-0.22	-0.23	-0.16	-0.34	0.01	-9.90	0.91	-9.90	0.87
20	643411	0	41916	0.39	0.24	0.39	0.24	0.13	0.01	0.00	0.25	-0.04	0.25	-0.12	-0.13	0.52	0.01	9.90	1.17	9.90	1.24
21	642850	0	41916	0.52	0.09	0.52	0.14	0.25	0.01	0.00	0.32	-0.19	0.32	-0.20	-0.06	-0.07	0.01	9.90	1.07	9.12	1.06
22	702208	0	41916	0.55	0.55	0.22	0.16	0.06	0.01	0.00	0.32	0.32	-0.17	-0.09	-0.19	-0.23	0.01	9.90	1.04	9.54	1.06
23	673866	0	41916	0.38	0.13	0.28	0.38	0.19	0.01	0.00	0.34	-0.15	-0.13	0.34	-0.12	0.51	0.01	9.90	1.07	9.90	1.12
24	702070	0	41916	0.35	0.35	0.20	0.27	0.17	0.01	0.00	0.33	0.33	-0.11	-0.10	-0.17	0.53	0.01	7.93	1.04	9.90	1.08
25	674147	0	41916	0.52	0.21	0.13	0.52	0.15	0.00	0.00	0.37	-0.15	-0.23	0.37	-0.13	-0.28	0.01	2.19	1.01	-1.38	0.99
26	705228	0	41916	0.52	0.52	0.25	0.14	0.09	0.00	0.00	0.44	0.44	-0.25	-0.20	-0.15	-0.25	0.01	-9.90	0.93	-9.90	0.91
27	679253	0	41916	0.60	0.07	0.14	0.60	0.19	0.00	0.00	0.34	-0.17	-0.23	0.34	-0.11	-0.32	0.01	0.88	1.00	1.19	1.01
28	641191	0	41916	0.47	0.22	0.47	0.06	0.25	0.00	0.00	0.31	-0.23	0.31	-0.18	-0.04	-0.10	0.01	9.90	1.07	9.90	1.12
29	703153	0	41916	0.32	0.22	0.32	0.28	0.18	0.00	0.00	0.21	-0.07	0.21	-0.16	0.02	0.71	0.01	9.90	1.15	9.90	1.25

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	678892	0	41916	0.73	0.11	0.73	0.08	0.07	0.00	0.00	0.44	-0.21	0.44	-0.25	-0.22	-1.31	0.01	-9.90	0.85	-9.90	0.79
31	642371	0	41916	0.49	0.22	0.18	0.49	0.12	0.00	0.00	0.38	-0.13	-0.20	0.38	-0.19	-0.18	0.01	-2.35	0.99	0.86	1.01
32	641265	0	41916	0.49	0.49	0.20	0.18	0.12	0.00	0.00	0.36	0.36	-0.15	-0.17	-0.16	-0.09	0.01	3.93	1.01	-0.18	1.00
33	704198	0	41916	0.55	0.55	0.12	0.09	0.23	0.00	0.00	0.34	0.34	-0.19	-0.24	-0.09	-0.34	0.01	4.04	1.01	6.62	1.04
34	678968	0	41916	0.61	0.14	0.09	0.16	0.61	0.00	0.00	0.44	-0.19	-0.24	-0.22	0.44	-0.76	0.01	-9.90	0.92	-9.90	0.86
35	643343	0	41916	0.71	0.07	0.09	0.13	0.71	0.00	0.00	0.44	-0.23	-0.21	-0.23	0.44	-1.16	0.01	-9.90	0.86	-9.90	0.80
36	610863	0	41916	0.45	0.45	0.30	0.16	0.08	0.00	0.00	0.38	0.38	-0.13	-0.20	-0.19	0.17	0.01	4.45	1.02	1.70	1.01
37	704202	0	41916	0.53	0.14	0.20	0.53	0.12	0.01	0.00	0.41	-0.18	-0.18	0.41	-0.19	-0.41	0.01	-9.84	0.96	-7.39	0.95
38	641202	0	41916	0.55	0.14	0.20	0.55	0.10	0.01	0.00	0.47	-0.19	-0.28	0.47	-0.15	-0.84	0.01	-0.79	1.00	-5.43	0.95
39	679206	0	41916	0.47	0.11	0.47	0.16	0.26	0.01	0.00	0.38	-0.21	0.38	-0.20	-0.10	-0.09	0.01	0.33	1.00	-1.47	0.99
40	642860	0	41916	0.45	0.16	0.20	0.17	0.45	0.01	0.00	0.42	-0.14	-0.20	-0.19	0.42	0.08	0.01	-9.21	0.96	-8.62	0.95
41	679964	0	41916	0.43	0.19	0.43	0.22	0.14	0.01	0.00	0.34	-0.14	0.34	-0.11	-0.17	0.29	0.01	9.90	1.07	9.90	1.09
42	704786	0	41916	0.55	0.14	0.16	0.15	0.55	0.01	0.00	0.51	-0.20	-0.23	-0.26	0.51	-0.72	0.01	-9.90	0.92	-9.90	0.86
43	703154	0	41916	0.39	0.17	0.34	0.39	0.08	0.01	0.00	0.31	-0.16	-0.07	0.31	-0.17	0.28	0.01	9.90	1.07	9.90	1.09
44	608250	0	41916	0.45	0.45	0.23	0.16	0.14	0.01	0.00	0.39	0.39	-0.16	-0.23	-0.09	-0.34	0.01	2.48	1.01	9.08	1.06
45	700897	0	41916	0.46	0.15	0.19	0.46	0.19	0.01	0.00	0.43	-0.20	-0.17	0.43	-0.16	-0.19	0.01	-9.90	0.96	-9.90	0.93
46	643401	0	41916	0.35	0.35	0.30	0.19	0.15	0.01	0.00	0.26	0.26	0.04	-0.19	-0.16	0.74	0.01	9.90	1.18	9.90	1.26
47	644022	0	41916	0.40	0.34	0.13	0.13	0.40	0.01	0.00	0.28	0.01	-0.19	-0.20	0.28	0.10	0.01	9.90	1.09	9.90	1.13
48	704256	0	41916	0.45	0.45	0.22	0.19	0.14	0.01	0.00	0.35	0.35	-0.12	-0.14	-0.18	0.13	0.01	9.64	1.04	8.78	1.05

Table J–4. Literature Multiple-Choice Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	641600	0	36551	0.30	0.16	0.22	0.33	0.30	0.00	0.00	0.25	-0.20	-0.03	-0.07	0.25	1.49	0.01	9.90	1.12	9.90	1.28
2	641594	0	36551	0.56	0.16	0.56	0.14	0.14	0.00	0.00	0.31	-0.05	0.31	-0.22	-0.16	0.40	0.01	9.90	1.12	9.90	1.15
3	641611	0	36551	0.53	0.53	0.18	0.15	0.13	0.00	0.00	0.33	0.33	-0.17	-0.18	-0.11	0.47	0.01	9.90	1.09	9.90	1.13
4	641612	0	36551	0.64	0.64	0.16	0.13	0.07	0.00	0.00	0.42	0.42	-0.20	-0.21	-0.23	-0.20	0.01	-6.91	0.97	-8.48	0.93
5	641598	0	36551	0.40	0.17	0.40	0.38	0.05	0.00	0.00	0.25	-0.20	0.25	-0.03	-0.14	1.00	0.01	9.90	1.18	9.90	1.31
6	641595	0	36551	0.55	0.55	0.12	0.12	0.21	0.00	0.00	0.29	0.29	-0.20	-0.19	-0.04	0.62	0.01	9.90	1.17	9.90	1.24
7	641599	0	36551	0.54	0.18	0.19	0.54	0.09	0.00	0.00	0.29	-0.03	-0.18	0.29	-0.23	0.49	0.01	9.90	1.14	9.90	1.19
8	641592	0	36551	0.40	0.11	0.31	0.18	0.40	0.00	0.00	0.23	-0.16	-0.05	-0.11	0.23	1.47	0.01	9.90	1.35	9.90	1.64
9	640877	0	36551	0.68	0.68	0.14	0.12	0.07	0.00	0.00	0.44	0.44	-0.24	-0.23	-0.18	-0.43	0.01	-9.90	0.94	-9.90	0.89
10	640876	0	36551	0.55	0.15	0.13	0.17	0.55	0.00	0.00	0.46	-0.23	-0.20	-0.20	0.46	0.03	0.01	-5.82	0.97	-8.89	0.93
11	640879	0	36551	0.69	0.69	0.14	0.10	0.06	0.00	0.00	0.45	0.45	-0.25	-0.22	-0.20	-0.49	0.01	-9.90	0.91	-9.90	0.88
12	640863	0	36551	0.78	0.06	0.08	0.08	0.78	0.00	0.00	0.50	-0.26	-0.25	-0.27	0.50	-1.48	0.02	6.48	1.06	-8.83	0.85
13	640865	0	36551	0.44	0.24	0.10	0.44	0.21	0.01	0.00	0.29	-0.07	-0.22	0.29	-0.11	0.86	0.01	9.90	1.14	9.90	1.23
14	640881	0	36551	0.49	0.34	0.09	0.49	0.08	0.01	0.00	0.27	-0.02	-0.30	0.27	-0.13	0.71	0.01	9.90	1.18	9.90	1.26
15	640864	0	36551	0.61	0.11	0.17	0.11	0.61	0.01	0.00	0.50	-0.25	-0.22	-0.26	0.50	-0.53	0.01	6.56	1.04	-3.41	0.96
16	640878	0	36551	0.51	0.11	0.18	0.18	0.51	0.01	0.00	0.49	-0.24	-0.24	-0.18	0.49	0.38	0.01	-9.90	0.91	-9.90	0.88
17	640860	0	36551	0.56	0.08	0.56	0.10	0.25	0.01	0.00	0.39	-0.25	0.39	-0.27	-0.08	0.69	0.01	9.90	1.09	9.90	1.13
18	641571	0	36551	0.80	0.07	0.05	0.08	0.80	0.00	0.00	0.46	-0.24	-0.27	-0.23	0.46	-1.25	0.01	-9.90	0.87	-9.90	0.79
19	641570	0	36551	0.49	0.20	0.49	0.22	0.09	0.00	0.00	0.31	-0.18	0.31	-0.08	-0.17	0.49	0.01	9.90	1.11	9.90	1.15
20	641586	0	36551	0.49	0.19	0.09	0.49	0.23	0.00	0.00	0.34	-0.19	-0.28	0.34	-0.03	0.56	0.01	9.90	1.08	9.90	1.12
21	641590	0	36551	0.75	0.06	0.11	0.75	0.08	0.00	0.00	0.46	-0.26	-0.27	0.46	-0.19	-0.91	0.01	-9.90	0.89	-9.90	0.80
22	641588	0	36551	0.46	0.46	0.16	0.22	0.15	0.00	0.00	0.36	0.36	-0.16	-0.17	-0.14	0.74	0.01	9.90	1.06	9.90	1.10
23	641589	0	36551	0.57	0.17	0.19	0.07	0.57	0.00	0.00	0.49	-0.23	-0.22	-0.28	0.49	-0.10	0.01	-9.90	0.94	-9.90	0.90
24	641569	0	36551	0.57	0.24	0.09	0.57	0.09	0.00	0.00	0.37	-0.09	-0.28	0.37	-0.21	0.49	0.01	9.90	1.06	9.90	1.09
25	641591	0	36551	0.68	0.68	0.11	0.10	0.12	0.00	0.00	0.46	0.46	-0.26	-0.25	-0.19	-0.46	0.01	-9.90	0.92	-9.90	0.85
26	641567	0	36551	0.56	0.56	0.13	0.07	0.24	0.00	0.00	0.25	0.25	-0.13	-0.30	0.00	1.22	0.01	9.90	1.50	9.90	1.74
27	642702	0	36551	0.46	0.14	0.46	0.31	0.08	0.00	0.00	0.33	-0.15	0.33	-0.11	-0.19	0.66	0.01	9.90	1.09	9.90	1.15
28	642704	0	36551	0.54	0.13	0.54	0.15	0.18	0.01	0.00	0.46	-0.25	0.46	-0.24	-0.13	0.14	0.01	-7.78	0.96	-6.91	0.95
29	642703	0	36551	0.54	0.09	0.14	0.54	0.22	0.00	0.00	0.27	-0.14	-0.10	0.27	-0.13	0.45	0.01	9.90	1.16	9.90	1.20

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	642707	0	36551	0.48	0.11	0.26	0.48	0.15	0.01	0.00	0.42	-0.20	-0.12	0.42	-0.24	0.31	0.01	3.60	1.02	1.98	1.01
31	642706	0	36551	0.36	0.16	0.25	0.22	0.36	0.01	0.00	0.21	-0.08	-0.09	-0.05	0.21	1.54	0.01	9.90	1.34	9.90	1.60
32	642724	0	36551	0.40	0.23	0.40	0.19	0.17	0.01	0.00	0.33	-0.17	0.33	-0.14	-0.07	0.93	0.01	9.90	1.07	9.90	1.16
33	642705	0	36551	0.55	0.10	0.21	0.14	0.55	0.01	0.00	0.45	-0.17	-0.21	-0.23	0.45	0.09	0.01	-4.67	0.98	-5.28	0.96
34	642723	0	36551	0.58	0.18	0.09	0.14	0.58	0.01	0.00	0.49	-0.18	-0.28	-0.24	0.49	0.00	0.01	-9.90	0.92	-9.90	0.90

Table J–5. Algebra I Multiple-Choice Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	674150	0	167097	0.79	0.79	0.07	0.08	0.06	0.00	0.00	0.42	0.42	-0.25	-0.26	-0.15					-1.10	0.01	-9.90	0.81	-9.90	0.69
2	674468	0	167097	0.31	0.31	0.30	0.20	0.18	0.00	0.00	0.39	0.39	-0.11	-0.17	-0.17					1.28	0.01	-6.94	0.97	2.43	1.02
3	700864	0	167097	0.54	0.54	0.14	0.22	0.09	0.00	0.00	0.42	0.42	-0.15	-0.25	-0.18					0.40	0.01	-5.62	0.97	-6.07	0.95
4	640618	0	167097	0.52	0.28	0.52	0.14	0.06	0.00	0.00	0.42	-0.15	0.42	-0.27	-0.21					0.51	0.01	-3.61	0.98	-5.20	0.96
5	674453	0	167097	0.40	0.20	0.40	0.17	0.23	0.00	0.00	0.41	-0.15	0.41	-0.18	-0.18					1.25	0.01	4.75	1.02	9.90	1.08
6	640603	0	167097	0.56	0.19	0.06	0.56	0.19	0.00	0.00	0.38	-0.20	-0.14	0.38	-0.19					0.24	0.01	5.49	1.03	4.91	1.04
7	674391	0	167097	0.65	0.07	0.13	0.15	0.65	0.00	0.00	0.24	-0.07	-0.24	-0.05	0.24					-0.22	0.01	9.90	1.19	9.90	1.36
8	700863	0	167097	0.32	0.24	0.20	0.24	0.32	0.00	0.00	0.30	-0.05	-0.16	-0.13	0.30					1.39	0.01	9.90	1.11	9.90	1.20
9	674148	0	167097	0.51	0.26	0.16	0.51	0.07	0.00	0.00	0.39	-0.18	-0.21	0.39	-0.14					0.61	0.01	6.08	1.03	3.79	1.03
10	674382	0	167097	0.46	0.10	0.27	0.16	0.46	0.01	0.00	0.41	-0.26	-0.13	-0.16	0.41					0.72	0.01	2.04	1.01	0.25	1.00
11	666551	0	167097	0.72	0.13	0.72	0.09	0.05	0.01	0.00	0.45	-0.21	0.45	-0.26	-0.20					-0.86	0.01	-7.23	0.95	-7.00	0.89
12	674448	0	167097	0.75	0.09	0.75	0.08	0.07	0.01	0.00	0.47	-0.23	0.47	-0.26	-0.21					-1.32	0.02	8.01	1.07	-3.65	0.93
13	736783	0	167097	0.55	0.17	0.11	0.55	0.16	0.01	0.00	0.37	-0.15	-0.22	0.37	-0.13					0.16	0.01	9.90	1.08	7.93	1.07
14	700823	0	167097	0.42	0.14	0.42	0.21	0.22	0.01	0.00	0.29	-0.16	0.29	-0.11	-0.07					1.06	0.01	9.90	1.18	9.90	1.26
15	674433	0	167097	0.28	0.19	0.24	0.28	0.28	0.01	0.00	0.25	-0.06	-0.07	-0.11	0.25					1.68	0.01	9.90	1.12	9.90	1.28
16	681810	0	167097	0.44	0.44	0.27	0.16	0.11	0.01	0.00	0.21	0.21	-0.14	0.01	-0.11					1.04	0.01	9.90	1.27	9.90	1.37
17	696817	0	167097	0.35	0.14	0.26	0.35	0.23	0.01	0.00	0.23	-0.23	-0.12	0.23	0.08					1.29	0.01	9.90	1.19	9.90	1.31
18	674446	0	167097	0.32	0.32	0.13	0.25	0.29	0.01	0.00	0.31	0.31	-0.16	-0.12	-0.06					1.43	0.01	9.90	1.10	9.90	1.17
19	682039	0	167097	0.72	0.72	0.10	0.06	0.12	0.00	0.00	0.31	0.31	-0.20	-0.18	-0.12					-0.30	0.01	-4.07	0.98	-1.81	0.98
20	678771	0	167097	0.74	0.15	0.74	0.08	0.03	0.00	0.00	0.44	-0.27	0.44	-0.25	-0.18					-0.89	0.01	-9.90	0.88	-9.90	0.76
21	702571	0	167097	0.30	0.30	0.25	0.20	0.25	0.00	0.00	0.30	0.30	-0.09	-0.17	-0.06					1.65	0.01	9.90	1.11	9.90	1.29
22	724197	0	167097	0.28	0.08	0.60	0.04	0.28	0.00	0.00	0.54	-0.15	-0.34	-0.18	0.54					1.57	0.01	-9.90	0.80	-9.90	0.82
23	724194	0	167097	0.52	0.08	0.32	0.52	0.08	0.00	0.00	0.40	-0.23	-0.15	0.40	-0.23					0.55	0.01	2.94	1.01	1.78	1.01
24	641531	0	167097	0.73	0.06	0.08	0.73	0.13	0.00	0.00	0.39	-0.23	-0.20	0.39	-0.19					-0.89	0.01	2.28	1.02	-1.52	0.98
25	678752	0	167097	0.66	0.15	0.11	0.09	0.66	0.00	0.00	0.44	-0.22	-0.23	-0.21	0.44					-0.66	0.01	9.50	1.06	-0.03	1.00
26	682132	0	167097	0.43	0.43	0.19	0.18	0.20	0.00	0.00	0.26	0.26	-0.17	-0.13	-0.03					0.82	0.01	9.90	1.22	9.90	1.32
27	712563	0	167097	0.30	0.33	0.21	0.16	0.30	0.00	0.00	0.43	-0.17	-0.17	-0.12	0.43					1.32	0.01	-9.90	0.95	-6.05	0.95
28	724114	0	167097	0.41	0.09	0.23	0.41	0.26	0.01	0.00	0.31	-0.15	-0.04	0.31	-0.20					1.05	0.01	9.90	1.13	9.90	1.19
29	678782	0	167097	0.41	0.39	0.12	0.41	0.07	0.01	0.00	0.36	-0.09	-0.23	0.36	-0.19					0.92	0.01	9.90	1.07	9.90	1.12

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	678717	0	167097	0.69	0.11	0.69	0.12	0.07	0.01	0.00	0.38	-0.18	0.38	-0.22	-0.17					-0.49	0.01	-2.50	0.99	-2.88	0.96
31	724143	0	167097	0.60	0.60	0.17	0.15	0.08	0.01	0.00	0.54	0.54	-0.26	-0.27	-0.23					-0.30	0.01	-9.90	0.89	-9.90	0.80
32	641436	0	167097	0.44	0.08	0.14	0.44	0.33	0.01	0.00	0.40	-0.21	-0.27	0.40	-0.09					0.77	0.01	4.21	1.02	3.76	1.03
33	641453	0	167097	0.57	0.15	0.09	0.17	0.57	0.01	0.00	0.52	-0.22	-0.22	-0.29	0.52					0.11	0.01	-9.90	0.84	-9.90	0.76
34	724120	0	167097	0.57	0.57	0.18	0.18	0.06	0.01	0.00	0.51	0.51	-0.25	-0.26	-0.20					0.03	0.01	-9.90	0.88	-9.90	0.81
35	605156	0	167097	0.39	0.22	0.39	0.25	0.14	0.01	0.00	0.43	-0.17	0.43	-0.18	-0.15					0.87	0.01	-7.36	0.97	-4.62	0.97
36	674156	0	167097	0.50	0.14	0.50	0.13	0.23	0.01	0.00	0.39	-0.21	0.39	-0.23	-0.10					0.49	0.01	6.42	1.03	5.51	1.04
37	700808	1	5891	0.21	0.20	0.19	0.39	0.21	0.00	0.00	0.27	-0.09	-0.10	-0.07	0.27	A+	A-	A-	A-	2.45	0.06	2.09	1.07	5.84	1.40
38	736762	1	5891	0.33	0.21	0.33	0.33	0.14	0.00	0.00	0.23	-0.09	-0.10	0.23	-0.07	A+	A+	A-	A-	1.72	0.05	7.78	1.20	9.36	1.43
39	800172	1	5891	0.33	0.13	0.33	0.24	0.30	0.00	0.00	0.44	0.04	0.44	-0.31	-0.19	A-	A-	A-	A-	1.70	0.05	-2.37	0.94	-0.59	0.98
40	714027	1	5891	0.14	0.20	0.34	0.31	0.14	0.01	0.00	0.07	0.02	-0.04	-0.03	0.07	A-	A+	A+	A-	3.05	0.07	3.89	1.18	9.90	2.41
41	817733	1	5891	0.43	0.11	0.20	0.43	0.26	0.00	0.00	0.40	-0.16	-0.18	0.40	-0.18	A+	A-	A-	A-	1.15	0.05	2.63	1.06	5.16	1.17
42	818791	1	5891	0.58	0.58	0.16	0.20	0.06	0.00	0.00	0.45	0.45	-0.21	-0.26	-0.19	A+	A-	A-	A-	0.38	0.05	-2.28	0.95	-2.47	0.92
43	819635	1	5891	0.68	0.13	0.12	0.07	0.68	0.00	0.00	0.49	-0.41	-0.15	-0.16	0.49	A+	A-	A-	A+	-0.19	0.05	-4.52	0.90	-2.86	0.88
44	818916	1	5891	0.33	0.17	0.34	0.16	0.33	0.00	0.00	0.35	-0.21	-0.23	0.07	0.35	A-	A-	A-	A-	1.71	0.05	4.90	1.12	7.33	1.32
45	817738	1	5891	0.39	0.39	0.28	0.08	0.24	0.00	0.00	0.42	0.42	0.06	-0.27	-0.37	A-	A-	A-	A-	1.36	0.05	-0.10	1.00	0.59	1.02
46	819887	1	5891	0.36	0.32	0.36	0.21	0.11	0.00	0.00	0.14	0.09	0.14	-0.24	-0.01	A-	A-	A-	A+	1.53	0.05	9.90	1.36	9.90	1.55
47	819093	2	5903	0.52	0.52	0.14	0.24	0.09	0.00	0.00	0.38	0.38	-0.23	-0.08	-0.24	A+	A-	A-	A-	0.64	0.05	4.03	1.08	3.77	1.13
48	817731	2	5903	0.27	0.16	0.27	0.22	0.34	0.00	0.00	0.07	0.07	0.07	-0.11	-0.02	A-	A-	A-	A-	2.00	0.06	9.90	1.33	9.90	1.77
49	818802	2	5903	0.56	0.14	0.09	0.56	0.21	0.00	0.00	0.31	-0.13	-0.18	0.31	-0.13	A+	A-	A+	A+	0.46	0.05	5.36	1.11	4.06	1.15
50	819074	2	5903	0.56	0.25	0.09	0.09	0.56	0.00	0.00	0.40	-0.15	-0.20	-0.26	0.40	A+	A-	A-	A+	0.45	0.05	1.45	1.03	0.91	1.03
51	816625	2	5903	0.58	0.04	0.58	0.21	0.16	0.00	0.00	0.36	-0.17	0.36	-0.14	-0.23	A-	A-	A-	A+	0.34	0.05	3.22	1.07	4.21	1.16
52	818266	2	5903	0.36	0.16	0.36	0.34	0.14	0.00	0.00	0.28	-0.17	0.28	-0.04	-0.14	A+	A-	A+	A+	1.50	0.05	5.62	1.14	7.10	1.28
53	724161	2	5903	0.32	0.12	0.15	0.32	0.42	0.00	0.00	0.13	-0.18	-0.09	0.13	0.06	A-	A-	A+	A+	1.73	0.05	9.90	1.34	9.90	1.57
54	818776	2	5903	0.41	0.41	0.39	0.07	0.14	0.00	0.00	0.46	0.46	-0.16	-0.18	-0.29	B-	A+	A-	A+	1.22	0.05	-3.11	0.93	-2.38	0.92
55	820047	2	5903	0.45	0.45	0.09	0.43	0.03	0.00	0.00	0.35	0.35	-0.29	-0.13	-0.15	A-	A-	A+	A+	1.00	0.05	3.32	1.07	2.99	1.10
56	820054	2	5903	0.49	0.07	0.13	0.49	0.31	0.00	0.00	0.50	-0.15	-0.16	0.50	-0.34	A+	A-	A-	A+	0.79	0.05	-4.78	0.91	-3.89	0.88
57	700781	3	5920	0.27	0.30	0.30	0.27	0.12	0.00	0.00	0.21	-0.07	-0.07	0.21	-0.09	A+	A-	A+	A+	2.03	0.06	6.68	1.20	9.63	1.52
58	819627	3	5920	0.38	0.35	0.38	0.19	0.08	0.00	0.00	0.12	0.17	0.12	-0.23	-0.18	A-	A-	A-	A-	1.41	0.05	9.90	1.42	9.90	1.63
59	736806	3	5920	0.27	0.27	0.15	0.24	0.33	0.00	0.00	0.23	0.23	-0.06	-0.16	-0.02	A-	A-	A+	A-	2.00	0.06	6.25	1.18	9.90	1.59

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
60	818770	3	5920	0.49	0.36	0.07	0.08	0.49	0.00	0.00	0.42	-0.18	-0.25	-0.23	0.42	A-	A-	A-	A-	0.83	0.05	0.86	1.02	0.86	1.03
61	820049	3	5920	0.34	0.34	0.22	0.20	0.24	0.00	0.00	0.25	0.25	-0.09	-0.16	-0.04	A-	A-	A-	A+	1.62	0.05	7.71	1.20	9.30	1.39
62	818766	3	5920	0.48	0.09	0.48	0.19	0.25	0.00	0.00	0.29	-0.08	0.29	-0.17	-0.13	B-	A-	A-	A+	0.88	0.05	7.33	1.16	6.56	1.21
63	724199	3	5920	0.32	0.14	0.36	0.32	0.18	0.00	0.00	0.21	-0.28	-0.05	0.21	0.06	A+	A-	A+	A-	1.71	0.05	8.43	1.22	9.90	1.47
64	818391	3	5920	0.44	0.28	0.12	0.44	0.15	0.00	0.00	0.40	-0.18	-0.21	0.40	-0.13	A-	A-	A-	A-	1.06	0.05	0.41	1.01	0.81	1.02
65	818387	3	5920	0.64	0.16	0.64	0.10	0.09	0.00	0.00	0.47	-0.22	0.47	-0.19	-0.30	A-	A-	A-	A+	0.02	0.05	-3.34	0.93	-3.92	0.85
66	724116	3	5920	0.41	0.41	0.16	0.27	0.15	0.00	0.00	0.33	0.33	-0.27	-0.01	-0.16	A-	A-	A-	A+	1.20	0.05	4.32	1.09	4.98	1.16
67	817154	4	5883	0.22	0.12	0.22	0.40	0.26	0.00	0.00	0.02	-0.18	0.02	0.17	-0.07	A-	A+	A+	A+	2.44	0.06	9.90	1.40	9.90	2.28
68	818383	4	5883	0.55	0.21	0.13	0.11	0.55	0.00	0.00	0.59	-0.35	-0.28	-0.19	0.59	A+	A-	A-	A+	0.54	0.05	-8.23	0.84	-6.28	0.79
69	818284	4	5883	0.52	0.04	0.17	0.26	0.52	0.00	0.00	0.41	-0.20	-0.19	-0.21	0.41	A-	A-	A+	A-	0.66	0.05	2.52	1.05	1.35	1.05
70	819091	4	5883	0.49	0.12	0.29	0.49	0.10	0.00	0.00	0.39	-0.15	-0.22	0.39	-0.15	A+	A-	A-	A+	0.82	0.05	3.20	1.07	2.45	1.09
71	724163	4	5883	0.48	0.17	0.48	0.14	0.21	0.00	0.00	0.35	-0.10	0.35	-0.24	-0.12	A+	A-	A+	A+	0.91	0.05	7.20	1.16	6.06	1.22
72	818384	4	5883	0.68	0.10	0.11	0.68	0.11	0.00	0.00	0.39	-0.16	-0.20	0.39	-0.23	A+	A-	A-	A+	-0.14	0.05	-0.95	0.98	-0.45	0.98
73	820456	4	5883	0.49	0.16	0.49	0.10	0.25	0.00	0.00	0.32	-0.17	0.32	-0.12	-0.15	A+	A-	A+	A+	0.83	0.05	5.68	1.12	4.45	1.16
74	817710	4	5883	0.75	0.75	0.11	0.09	0.05	0.00	0.00	0.48	0.48	-0.28	-0.31	-0.15	A-	B-	A-	A+	-0.62	0.06	-6.66	0.82	-4.95	0.71
75	724202	4	5883	0.33	0.15	0.31	0.21	0.33	0.00	0.00	0.42	-0.20	-0.18	-0.10	0.42	A-	A-	A+	A+	1.71	0.05	1.27	1.03	3.10	1.13
76	712492	4	5883	0.43	0.19	0.43	0.14	0.24	0.00	0.00	0.44	-0.09	0.44	-0.28	-0.20	A+	A-	A-	A+	1.16	0.05	0.06	1.00	0.85	1.03
77	818906	5	5817	0.69	0.18	0.69	0.06	0.07	0.00	0.00	0.41	-0.21	0.41	-0.25	-0.19	A+	A-	A-	A-	-0.21	0.05	-0.49	0.99	-0.36	0.98
78	736776	5	5817	0.49	0.16	0.17	0.49	0.18	0.00	0.00	0.29	-0.27	-0.04	0.29	-0.09	A-	A-	A-	A-	0.86	0.05	8.91	1.19	7.07	1.23
79	818771	5	5817	0.10	0.42	0.29	0.19	0.10	0.00	0.00	0.05	0.14	-0.11	-0.07	0.05	A-	A-	A-	A-	3.53	0.08	2.34	1.14	9.90	2.82
80	817701	5	5817	0.76	0.12	0.06	0.76	0.06	0.00	0.00	0.48	-0.26	-0.28	0.48	-0.22	A+	A-	A-	A+	-0.61	0.06	-4.98	0.87	-4.33	0.77
81	736744	5	5817	0.43	0.20	0.43	0.18	0.18	0.00	0.00	0.29	0.00	0.29	-0.18	-0.18	A-	A+	A+	A+	1.13	0.05	6.76	1.15	7.73	1.26
82	818256	5	5817	0.84	0.06	0.84	0.08	0.02	0.00	0.00	0.35	-0.22	0.35	-0.21	-0.15	A-	B-	A-	A+	-1.24	0.07	-2.48	0.91	-0.22	0.98
83	816626	5	5817	0.42	0.42	0.28	0.16	0.14	0.00	0.00	0.29	0.29	0.07	-0.20	-0.28	A+	A-	A-	A+	1.22	0.05	6.95	1.15	6.26	1.21
84	818267	5	5817	0.47	0.04	0.47	0.22	0.26	0.00	0.00	0.28	-0.17	0.28	-0.32	0.07	A+	A+	A+	A-	0.92	0.05	6.73	1.14	6.30	1.20
85	819212	5	5817	0.55	0.30	0.07	0.55	0.08	0.00	0.00	0.42	-0.14	-0.27	0.42	-0.26	A-	A-	A-	A+	0.55	0.05	1.03	1.02	0.40	1.01
86	724128	5	5817	0.27	0.13	0.39	0.27	0.21	0.00	0.00	0.06	-0.12	0.21	0.06	-0.21	A-	A-	A+	A+	2.05	0.06	9.90	1.41	9.90	1.81
87	666569	6	5849	0.41	0.23	0.19	0.17	0.41	0.00	0.00	0.34	-0.03	-0.15	-0.24	0.34	A-	A+	A-	A-	1.26	0.05	4.83	1.11	6.18	1.22
88	817729	6	5849	0.54	0.28	0.54	0.15	0.04	0.00	0.00	0.45	-0.23	0.45	-0.23	-0.20	A-	A+	A-	A+	0.59	0.05	-1.02	0.98	-0.78	0.97
89	818287	6	5849	0.72	0.13	0.72	0.06	0.09	0.00	0.00	0.33	-0.13	0.33	-0.22	-0.18	A+	A-	A-	B+	-0.39	0.06	0.94	1.02	0.66	1.04

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
90	818758	6	5849	0.56	0.20	0.18	0.56	0.06	0.00	0.00	0.30	-0.06	-0.20	0.30	-0.20	A+	A+	A+	A-	0.50	0.05	6.80	1.15	5.53	1.20
91	818261	6	5849	0.57	0.57	0.23	0.09	0.11	0.00	0.00	0.55	0.55	-0.26	-0.23	-0.30	A-	B-	B-	A-	0.42	0.05	-5.39	0.89	-4.48	0.85
92	816628	6	5849	0.63	0.26	0.63	0.05	0.06	0.00	0.00	0.47	-0.28	0.47	-0.25	-0.22	A+	A-	A-	A+	0.10	0.05	-3.50	0.93	-2.41	0.90
93	820056	6	5849	0.61	0.13	0.61	0.14	0.11	0.00	0.00	0.49	-0.10	0.49	-0.28	-0.33	A+	B-	A-	A+	0.21	0.05	-3.30	0.93	-0.72	0.97
94	820057	6	5849	0.60	0.06	0.24	0.10	0.60	0.00	0.00	0.54	-0.17	-0.42	-0.13	0.54	A+	A-	A-	A-	0.26	0.05	-5.14	0.90	-4.10	0.85
95	817159	6	5849	0.65	0.11	0.14	0.65	0.10	0.00	0.00	0.47	-0.25	-0.22	0.47	-0.23	A+	A-	A-	A-	0.00	0.05	-0.98	0.98	-1.23	0.95
96	819102	6	5849	0.46	0.16	0.19	0.46	0.19	0.00	0.00	0.19	-0.14	-0.03	0.19	-0.08	A+	A+	A+	A+	1.02	0.05	9.90	1.35	9.90	1.48
97	818258	7	5797	0.53	0.21	0.07	0.53	0.19	0.00	0.00	0.44	-0.07	-0.16	0.44	-0.38	A-	A-	A+	A-	0.61	0.05	0.18	1.00	-0.47	0.98
98	736719	7	5797	0.73	0.73	0.09	0.11	0.07	0.00	0.00	0.42	0.42	-0.28	-0.17	-0.21	A+	A-	A-	A+	-0.47	0.06	-1.66	0.96	-0.25	0.98
99	819878	7	5797	0.50	0.24	0.50	0.17	0.09	0.00	0.00	0.25	-0.03	0.25	-0.17	-0.17	A-	A+	A-	A-	0.76	0.05	9.90	1.23	8.62	1.30
100	818381	7	5797	0.70	0.06	0.14	0.09	0.70	0.00	0.00	0.45	-0.19	-0.22	-0.27	0.45	A+	A-	A+	A-	-0.32	0.05	-1.63	0.96	-0.09	0.99
101	819221	7	5797	0.59	0.59	0.17	0.13	0.11	0.00	0.00	0.39	0.39	-0.17	-0.19	-0.19	A+	A-	A-	A+	0.29	0.05	2.33	1.05	1.22	1.05
102	820457	7	5797	0.59	0.20	0.09	0.59	0.12	0.00	0.00	0.45	-0.30	-0.23	0.45	-0.10	A-	A+	A+	A+	0.33	0.05	-1.46	0.97	0.51	1.02
103	800480	7	5797	0.28	0.28	0.28	0.23	0.21	0.00	0.00	0.27	0.27	-0.02	-0.22	-0.05	A+	A+	A+	A-	2.00	0.06	4.22	1.12	7.39	1.40
104	724166	7	5797	0.56	0.56	0.16	0.14	0.13	0.00	0.00	0.42	0.42	-0.18	-0.23	-0.17	A+	A-	A+	A+	0.45	0.05	1.00	1.02	0.69	1.02
105	818774	7	5797	0.67	0.12	0.09	0.67	0.12	0.00	0.00	0.53	-0.29	-0.26	0.53	-0.25	A+	A-	A-	A+	-0.12	0.05	-6.57	0.86	-6.09	0.74
106	819225	7	5797	0.25	0.22	0.25	0.24	0.29	0.00	0.00	0.25	-0.08	0.25	-0.07	-0.09	A-	A+	A+	A+	2.22	0.06	4.53	1.15	8.79	1.56
107	712199	8	5903	0.50	0.10	0.24	0.17	0.50	0.00	0.00	0.43	-0.16	-0.26	-0.16	0.43	A+	A-	A-	A-	0.80	0.05	2.59	1.06	3.03	1.10
108	819096	8	5903	0.42	0.09	0.15	0.34	0.42	0.00	0.00	0.27	-0.20	-0.27	0.05	0.27	A+	A-	A+	A-	1.21	0.05	9.26	1.22	8.75	1.33
109	818260	8	5903	0.84	0.04	0.03	0.84	0.08	0.00	0.00	0.41	-0.23	-0.19	0.41	-0.25	A+	B-	A-	A+	-1.28	0.07	-3.96	0.85	-3.66	0.70
110	712396	8	5903	0.66	0.66	0.15	0.09	0.10	0.00	0.00	0.41	0.41	-0.26	-0.16	-0.18	A+	A-	A-	A+	-0.03	0.05	0.26	1.01	-0.12	0.99
111	818907	8	5903	0.76	0.10	0.07	0.76	0.07	0.00	0.00	0.31	-0.14	-0.14	0.31	-0.21	A+	A-	B-	A+	-0.67	0.06	0.53	1.02	3.79	1.26
112	818388	8	5903	0.62	0.12	0.19	0.62	0.06	0.00	0.00	0.27	-0.06	-0.16	0.27	-0.20	A-	A-	A-	A+	0.15	0.05	8.28	1.19	7.38	1.33
113	736737	8	5903	0.29	0.17	0.38	0.15	0.29	0.00	0.00	0.20	-0.04	-0.03	-0.16	0.20	A-	A+	A+	A-	1.95	0.06	9.18	1.27	9.90	1.76
114	817732	8	5903	0.66	0.15	0.66	0.10	0.09	0.00	0.00	0.40	-0.20	0.40	-0.16	-0.25	B-	A-	A-	A+	-0.05	0.05	-0.11	1.00	0.15	1.01
115	820407	8	5903	0.64	0.16	0.08	0.12	0.64	0.00	0.00	0.47	-0.13	-0.29	-0.31	0.47	B+	A-	A-	A+	0.08	0.05	-3.03	0.93	1.00	1.04
116	819083	8	5903	0.86	0.02	0.08	0.86	0.03	0.00	0.00	0.40	-0.15	-0.28	0.40	-0.21	A-	A-	A-	A+	-1.42	0.07	-2.85	0.88	-2.55	0.77
117	736795	9	5862	0.24	0.14	0.28	0.34	0.24	0.00	0.00	0.24	-0.12	-0.12	0.00	0.24	A-	A-	A+	A-	2.24	0.06	4.87	1.16	8.63	1.54
118	800173	9	5862	0.71	0.71	0.12	0.06	0.11	0.00	0.00	0.41	0.41	-0.24	-0.24	-0.16	A+	A-	A-	A+	-0.36	0.05	-0.99	0.98	-1.09	0.94
119	700775	9	5862	0.31	0.13	0.26	0.29	0.31	0.00	0.00	0.22	-0.15	0.00	-0.11	0.22	A-	A-	A+	A-	1.80	0.05	7.98	1.22	8.69	1.41

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
120	818283	9	5862	0.49	0.09	0.11	0.49	0.30	0.00	0.00	0.38	-0.16	-0.25	0.38	-0.14	A+	A-	A-	A-	0.79	0.05	3.83	1.08	2.85	1.09
121	819217	9	5862	0.29	0.29	0.29	0.28	0.14	0.00	0.00	0.10	-0.04	0.10	0.05	-0.14	A+	A-	A-	A-	1.89	0.06	9.90	1.38	9.90	1.78
122	724192	9	5862	0.31	0.29	0.24	0.16	0.31	0.00	0.00	0.46	-0.28	-0.11	-0.11	0.46	A-	A+	A-	A+	1.81	0.05	-2.59	0.93	-1.04	0.96
123	816455	9	5862	0.49	0.49	0.05	0.08	0.38	0.00	0.00	0.32	0.32	-0.22	-0.26	-0.08	A+	A-	A-	A+	0.80	0.05	7.19	1.15	6.39	1.21
124	819222	9	5862	0.62	0.13	0.18	0.62	0.07	0.00	0.00	0.51	-0.27	-0.24	0.51	-0.25	A+	A+	A+	A+	0.16	0.05	-6.22	0.88	-5.66	0.80
125	817712	9	5862	0.46	0.26	0.46	0.19	0.08	0.00	0.00	0.36	-0.13	0.36	-0.19	-0.16	A-	A-	A+	A+	0.94	0.05	4.78	1.10	4.43	1.14
126	818777	9	5862	0.53	0.53	0.11	0.24	0.11	0.00	0.00	0.47	0.47	-0.21	-0.28	-0.16	A+	A-	A+	A+	0.61	0.05	-2.78	0.95	-0.11	1.00
127	819075	10	5891	0.49	0.18	0.12	0.49	0.21	0.00	0.00	0.47	-0.24	-0.19	0.47	-0.20	A-	A-	A-	A+	0.83	0.05	-1.30	0.97	-1.05	0.97
128	816621	10	5891	0.53	0.53	0.15	0.19	0.13	0.00	0.00	0.48	0.48	-0.19	-0.23	-0.24	A+	A-	A-	A-	0.66	0.05	-2.47	0.95	-2.32	0.92
129	800171	10	5891	0.25	0.53	0.25	0.06	0.17	0.00	0.00	0.51	-0.28	0.51	-0.17	-0.11	A-	A-	B-	A+	2.24	0.06	-4.88	0.86	-1.27	0.93
130	800166	10	5891	0.83	0.06	0.08	0.83	0.03	0.00	0.00	0.42	-0.29	-0.21	0.42	-0.19	A+	A-	A-	A+	-1.15	0.06	-2.74	0.90	-1.02	0.92
131	800174	10	5891	0.58	0.08	0.58	0.25	0.08	0.00	0.00	0.46	-0.22	0.46	-0.24	-0.23	A-	B-	A-	A+	0.39	0.05	-1.41	0.97	-1.78	0.94
132	816457	10	5891	0.62	0.13	0.62	0.11	0.13	0.00	0.00	0.37	-0.25	0.37	-0.17	-0.12	A-	A-	A-	A+	0.15	0.05	3.04	1.07	1.73	1.07
133	817711	10	5891	0.75	0.03	0.12	0.75	0.09	0.00	0.00	0.36	-0.19	-0.22	0.36	-0.17	A+	A-	B-	A+	-0.59	0.06	-0.69	0.98	-0.53	0.97
134	724182	10	5891	0.35	0.21	0.24	0.20	0.35	0.00	0.00	0.46	-0.20	-0.23	-0.10	0.46	A-	A-	A-	A-	1.59	0.05	-1.34	0.97	1.77	1.07
135	713816	10	5891	0.45	0.19	0.27	0.45	0.09	0.00	0.00	0.26	-0.07	-0.08	0.26	-0.23	A-	A-	A-	A+	1.05	0.05	9.90	1.23	9.90	1.40
136	724158	10	5891	0.31	0.10	0.27	0.33	0.31	0.00	0.00	0.28	-0.12	-0.29	0.07	0.28	A-	A+	A+	A+	1.85	0.06	6.53	1.18	8.87	1.44
137	818264	11	5877	0.59	0.10	0.59	0.09	0.22	0.00	0.00	0.40	-0.18	0.40	-0.21	-0.20	A+	A-	A+	A+	0.32	0.05	1.91	1.04	1.53	1.06
138	818800	11	5877	0.48	0.34	0.48	0.10	0.09	0.00	0.00	0.40	-0.16	0.40	-0.20	-0.23	A+	A-	A-	A-	0.90	0.05	3.06	1.07	2.87	1.09
139	818910	11	5877	0.70	0.70	0.10	0.15	0.04	0.00	0.00	0.51	0.51	-0.28	-0.29	-0.22	A+	A-	A-	A-	-0.31	0.05	-5.41	0.87	-3.93	0.81
140	818768	11	5877	0.56	0.22	0.15	0.56	0.07	0.00	0.00	0.50	-0.29	-0.24	0.50	-0.16	A-	A-	A-	A+	0.45	0.05	-4.12	0.92	-3.72	0.88
141	736727	11	5877	0.74	0.05	0.06	0.15	0.74	0.00	0.00	0.48	-0.25	-0.26	-0.27	0.48	A+	B-	A-	A-	-0.52	0.06	-6.53	0.83	-4.67	0.76
142	678749	11	5877	0.57	0.07	0.26	0.10	0.57	0.00	0.00	0.42	-0.25	-0.25	-0.10	0.42	A-	A-	A-	A-	0.44	0.05	0.67	1.01	0.60	1.02
143	712561	11	5877	0.23	0.35	0.29	0.23	0.12	0.00	0.00	0.15	0.10	-0.08	0.15	-0.23	A-	A+	A+	A+	2.31	0.06	7.76	1.26	9.90	1.86
144	800485	11	5877	0.37	0.37	0.06	0.47	0.10	0.00	0.00	0.23	0.23	-0.26	0.02	-0.20	A-	A-	A+	A-	1.45	0.05	8.83	1.22	8.12	1.32
145	702466	11	5877	0.39	0.09	0.39	0.45	0.06	0.00	0.00	0.14	-0.12	0.14	0.01	-0.16	A+	A-	A-	A-	1.35	0.05	9.90	1.37	9.90	1.55
146	817709	11	5877	0.34	0.41	0.15	0.34	0.11	0.00	0.00	0.43	-0.16	-0.20	0.43	-0.17	A-	A+	A+	A-	1.64	0.05	-0.92	0.98	3.31	1.14
147	736798	12	5845	0.38	0.38	0.19	0.20	0.23	0.00	0.00	0.34	0.34	-0.14	-0.19	-0.08	A+	A-	A-	A-	1.41	0.05	3.84	1.09	4.81	1.19
148	736723	12	5845	0.34	0.16	0.28	0.22	0.34	0.00	0.00	0.38	-0.06	-0.16	-0.21	0.38	A+	A-	A-	A-	1.62	0.05	2.82	1.07	3.17	1.13
149	700836	12	5845	0.38	0.29	0.38	0.18	0.15	0.00	0.00	0.37	-0.18	0.37	-0.11	-0.15	A+	A-	A-	A-	1.42	0.05	3.07	1.07	5.11	1.20

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
150	819876	12	5845	0.74	0.07	0.12	0.74	0.08	0.00	0.00	0.48	-0.26	-0.25	0.48	-0.25	A+	A-	A-	A+	-0.53	0.06	-5.02	0.87	-4.09	0.77
151	819629	12	5845	0.50	0.24	0.11	0.15	0.50	0.00	0.00	0.34	-0.07	-0.23	-0.19	0.34	A+	A-	A+	A+	0.80	0.05	4.54	1.10	4.44	1.16
152	724159	12	5845	0.65	0.14	0.65	0.16	0.05	0.00	0.00	0.37	-0.25	0.37	-0.13	-0.19	A+	A-	A-	A-	-0.03	0.05	2.98	1.07	3.14	1.15
153	818912	12	5845	0.35	0.20	0.16	0.35	0.29	0.00	0.00	0.43	-0.27	-0.24	0.43	-0.02	A+	A+	A+	A-	1.57	0.05	-0.72	0.98	1.35	1.05
154	818764	12	5845	0.61	0.18	0.61	0.06	0.15	0.00	0.00	0.41	-0.36	0.41	-0.17	-0.06	A-	A-	A-	A-	0.23	0.05	0.12	1.00	0.38	1.02
155	819100	12	5845	0.72	0.72	0.08	0.16	0.04	0.00	0.00	0.46	0.46	-0.25	-0.26	-0.22	A+	A-	A-	A-	-0.38	0.06	-2.69	0.93	-2.24	0.88
156	820459	12	5845	0.39	0.14	0.20	0.27	0.39	0.00	0.00	0.37	-0.19	-0.11	-0.15	0.37	A-	A-	A-	A-	1.38	0.05	4.24	1.10	4.90	1.19
157	712271	13	5874	0.42	0.20	0.12	0.26	0.42	0.00	0.00	0.48	-0.11	-0.24	-0.27	0.48	A+	A-	A+	A-	1.20	0.05	-0.98	0.98	1.10	1.04
158	818769	13	5874	0.86	0.86	0.06	0.05	0.03	0.00	0.00	0.45	0.45	-0.28	-0.24	-0.22	C+	A-	A-	A+	-1.47	0.07	-3.77	0.85	-5.16	0.56
159	819216	13	5874	0.44	0.44	0.12	0.22	0.21	0.00	0.00	0.48	0.48	-0.19	-0.16	-0.26	A+	A-	A-	A-	1.08	0.05	0.06	1.00	0.94	1.03
160	819097	13	5874	0.62	0.08	0.62	0.07	0.22	0.00	0.00	0.27	-0.25	0.27	-0.22	-0.01	A+	A-	A-	A+	0.14	0.05	7.59	1.18	6.78	1.32
161	819072	13	5874	0.84	0.05	0.84	0.06	0.05	0.00	0.00	0.40	-0.24	0.40	-0.25	-0.17	A-	A-	A-	A+	-1.32	0.07	-3.36	0.87	-1.72	0.84
162	819079	13	5874	0.32	0.25	0.32	0.22	0.20	0.00	0.00	0.32	-0.16	0.32	-0.07	-0.11	A+	A-	A-	A-	1.79	0.05	4.24	1.11	5.77	1.28
163	819099	13	5874	0.59	0.06	0.12	0.59	0.23	0.00	0.00	0.49	-0.21	-0.32	0.49	-0.21	A+	A+	A+	A-	0.30	0.05	-2.58	0.94	-2.53	0.90
164	818915	13	5874	0.62	0.04	0.62	0.32	0.02	0.00	0.00	0.33	-0.19	0.33	-0.21	-0.16	A-	A-	A-	A+	0.15	0.05	5.16	1.12	4.73	1.22
165	817735	13	5874	0.56	0.12	0.11	0.56	0.21	0.00	0.00	0.47	-0.24	-0.28	0.47	-0.16	A+	A-	A-	A+	0.46	0.05	-1.07	0.98	-1.13	0.96
166	816629	13	5874	0.78	0.08	0.07	0.07	0.78	0.00	0.00	0.55	-0.30	-0.29	-0.29	0.55	A+	A-	A-	A-	-0.84	0.06	-8.45	0.76	-7.24	0.55
167	818276	14	5918	0.46	0.34	0.06	0.46	0.14	0.00	0.00	0.36	-0.17	-0.21	0.36	-0.14	A-	A-	A+	A+	0.98	0.05	5.10	1.11	4.80	1.16
168	819092	14	5918	0.84	0.03	0.84	0.09	0.04	0.00	0.00	0.47	-0.19	0.47	-0.31	-0.25	A+	B-	B-	A+	-1.23	0.07	-5.55	0.80	-5.89	0.58
169	724165	14	5918	0.39	0.19	0.27	0.15	0.39	0.00	0.00	0.25	-0.11	-0.09	-0.11	0.25	A-	A-	A-	A-	1.34	0.05	8.42	1.20	7.64	1.29
170	818385	14	5918	0.49	0.15	0.49	0.12	0.24	0.00	0.00	0.31	-0.28	0.31	-0.29	0.10	A-	A-	A-	A-	0.80	0.05	7.86	1.17	7.53	1.25
171	818801	14	5918	0.61	0.09	0.11	0.19	0.61	0.00	0.00	0.51	-0.20	-0.23	-0.31	0.51	A+	A-	A-	A+	0.22	0.05	-4.77	0.90	-4.29	0.85
172	724170	14	5918	0.24	0.15	0.20	0.42	0.24	0.00	0.00	0.14	-0.05	-0.06	-0.04	0.14	A+	A-	A+	A+	2.27	0.06	9.00	1.30	9.90	1.84
173	724142	14	5918	0.66	0.07	0.66	0.13	0.14	0.00	0.00	0.27	-0.11	0.27	-0.15	-0.14	A-	A-	A-	A+	-0.06	0.05	4.88	1.12	4.24	1.20
174	818288	14	5918	0.59	0.11	0.06	0.24	0.59	0.00	0.00	0.58	-0.26	-0.26	-0.33	0.58	A+	A-	A-	A+	0.32	0.05	-9.21	0.82	-6.82	0.77
175	818765	14	5918	0.42	0.38	0.10	0.42	0.10	0.00	0.00	0.37	-0.11	-0.19	0.37	-0.23	A+	A-	A-	A+	1.16	0.05	4.03	1.09	4.79	1.16
176	817164	14	5918	0.37	0.37	0.17	0.29	0.17	0.00	0.00	0.40	0.40	-0.19	-0.23	-0.05	A-	A-	A-	A+	1.46	0.05	0.24	1.01	2.41	1.09
177	819631	15	5849	0.39	0.31	0.39	0.15	0.15	0.00	0.00	0.30	-0.04	0.30	-0.16	-0.20	A+	A-	A-	A-	1.35	0.05	9.49	1.23	8.44	1.33
178	820043	15	5849	0.37	0.37	0.15	0.10	0.37	0.00	0.00	0.44	-0.14	-0.16	-0.28	0.44	A+	A-	A-	A+	1.48	0.05	-1.54	0.96	0.29	1.01
179	818386	15	5849	0.51	0.15	0.18	0.51	0.16	0.00	0.00	0.35	-0.19	-0.07	0.35	-0.22	A-	C-	B-	A+	0.72	0.05	4.73	1.10	3.11	1.10

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
180	799608	15	5849	0.70	0.70	0.09	0.08	0.14	0.00	0.00	0.45	0.45	-0.26	-0.24	-0.20	A+	B-	A-	A-	-0.27	0.05	-1.94	0.95	-0.38	0.98
181	820548	15	5849	0.36	0.21	0.36	0.27	0.16	0.00	0.00	0.32	0.02	0.32	-0.17	-0.23	A-	A-	A-	A-	1.56	0.05	5.55	1.14	6.37	1.27
182	819638	15	5849	0.51	0.05	0.20	0.51	0.23	0.00	0.00	0.34	-0.21	-0.25	0.34	-0.05	A+	A+	A-	A+	0.72	0.05	7.83	1.18	6.36	1.22
183	819204	15	5849	0.79	0.10	0.79	0.07	0.04	0.00	0.00	0.51	-0.34	0.51	-0.26	-0.22	A+	A-	A-	A+	-0.88	0.06	-7.23	0.79	-6.30	0.62
184	817707	15	5849	0.42	0.31	0.15	0.42	0.12	0.00	0.00	0.51	-0.32	-0.24	0.51	-0.06	A-	A-	A-	A-	1.21	0.05	-3.89	0.92	-1.53	0.95
185	800484	15	5849	0.47	0.09	0.28	0.47	0.17	0.00	0.00	0.39	-0.15	-0.26	0.39	-0.09	A-	A+	A+	A+	0.93	0.05	4.38	1.10	5.03	1.17
186	818795	15	5849	0.29	0.29	0.17	0.10	0.44	0.00	0.00	0.12	0.12	-0.24	-0.24	0.22	B-	A-	A+	A-	1.93	0.06	9.90	1.33	9.90	1.59
187	819090	16	5846	0.39	0.19	0.21	0.20	0.39	0.00	0.00	0.50	-0.26	-0.19	-0.15	0.50	A-	A-	A-	A-	1.33	0.05	-3.65	0.92	-1.63	0.94
188	818259	16	5846	0.79	0.79	0.11	0.06	0.04	0.00	0.00	0.54	0.54	-0.36	-0.25	-0.25	A+	A-	A-	A-	-0.92	0.06	-6.57	0.81	-6.34	0.61
189	800169	16	5846	0.66	0.11	0.15	0.08	0.66	0.00	0.00	0.44	-0.17	-0.28	-0.20	0.44	B-	A-	A-	A+	-0.06	0.05	-1.47	0.97	-1.97	0.92
190	818762	16	5846	0.50	0.50	0.16	0.26	0.07	0.00	0.00	0.32	0.32	-0.18	-0.13	-0.14	A+	A-	A+	A-	0.76	0.05	7.15	1.16	5.42	1.18
191	818262	16	5846	0.74	0.06	0.74	0.12	0.08	0.00	0.00	0.44	-0.27	0.44	-0.18	-0.25	A-	B-	A-	A-	-0.58	0.06	-2.68	0.93	-0.75	0.96
192	819634	16	5846	0.56	0.56	0.14	0.18	0.12	0.00	0.00	0.52	0.52	-0.27	-0.31	-0.14	A+	A-	A-	A+	0.43	0.05	-4.81	0.90	-4.16	0.86
193	817736	16	5846	0.53	0.17	0.17	0.12	0.53	0.00	0.00	0.43	-0.16	-0.25	-0.19	0.43	A+	A-	A-	A-	0.60	0.05	1.91	1.04	2.50	1.08
194	800479	16	5846	0.61	0.21	0.61	0.12	0.06	0.00	0.00	0.47	-0.29	0.47	-0.21	-0.19	A-	A-	A-	A+	0.19	0.05	-2.87	0.94	-2.83	0.89
195	724133	16	5846	0.59	0.15	0.12	0.14	0.59	0.00	0.00	0.54	-0.22	-0.31	-0.25	0.54	A+	A-	A-	A-	0.29	0.05	-5.99	0.88	-4.88	0.83
196	820458	16	5846	0.37	0.11	0.21	0.37	0.31	0.00	0.00	0.31	-0.08	-0.18	0.31	-0.11	A+	A+	A+	A+	1.47	0.05	7.09	1.18	9.09	1.39
197	696816	17	5876	0.11	0.11	0.49	0.13	0.27	0.00	0.00	0.05	0.05	0.14	-0.09	-0.13	A-	A-	A+	A-	3.38	0.08	2.01	1.10	8.82	2.21
198	700844	17	5876	0.70	0.18	0.70	0.09	0.04	0.00	0.00	0.46	-0.25	0.46	-0.25	-0.21	A-	B-	B-	A+	-0.25	0.05	-2.39	0.94	-0.80	0.96
199	819628	17	5876	0.41	0.16	0.28	0.41	0.15	0.00	0.00	0.28	-0.14	-0.06	0.28	-0.16	A+	A-	A-	A+	1.26	0.05	8.88	1.21	9.80	1.36
200	819210	17	5876	0.56	0.09	0.23	0.11	0.56	0.00	0.00	0.49	-0.26	-0.25	-0.18	0.49	A-	B-	A-	A+	0.50	0.05	-2.99	0.94	-1.88	0.94
201	817702	17	5876	0.74	0.74	0.05	0.16	0.04	0.00	0.00	0.48	0.48	-0.22	-0.30	-0.25	A+	A-	A-	A-	-0.53	0.06	-3.96	0.90	-2.80	0.84
202	820455	17	5876	0.52	0.10	0.52	0.20	0.17	0.00	0.00	0.41	-0.10	0.41	-0.19	-0.25	A+	A+	A+	A+	0.72	0.05	3.26	1.07	5.17	1.18
203	816446	17	5876	0.70	0.70	0.14	0.08	0.08	0.00	0.00	0.46	0.46	-0.30	-0.26	-0.14	A-	A-	A-	A+	-0.29	0.06	-3.86	0.91	-2.28	0.88
204	818790	17	5876	0.59	0.12	0.19	0.10	0.59	0.00	0.00	0.35	-0.15	-0.16	-0.20	0.35	A+	A-	A-	A-	0.33	0.05	4.28	1.09	3.58	1.14
205	817160	17	5876	0.58	0.58	0.12	0.12	0.18	0.00	0.00	0.29	0.29	-0.19	-0.22	-0.03	A+	A-	A-	A+	0.37	0.05	7.62	1.17	4.72	1.18
206	817157	17	5876	0.53	0.15	0.53	0.24	0.07	0.00	0.00	0.39	-0.15	0.39	-0.23	-0.15	A-	A-	A+	A+	0.63	0.05	2.61	1.06	2.18	1.07
207	820042	18	5841	0.38	0.14	0.38	0.35	0.13	0.00	0.00	0.18	-0.13	0.18	-0.07	-0.03	A-	A-	A-	A-	1.45	0.05	9.90	1.31	9.90	1.52
208	820044	18	5841	0.64	0.17	0.14	0.64	0.04	0.00	0.00	0.46	-0.39	-0.11	0.46	-0.17	A-	B-	A-	A+	0.05	0.05	-2.63	0.94	-0.46	0.98
209	819632	18	5841	0.50	0.15	0.25	0.50	0.09	0.00	0.00	0.38	-0.16	-0.15	0.38	-0.22	A+	A-	A+	A+	0.78	0.05	5.24	1.11	4.32	1.14

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
210	736781	18	5841	0.21	0.14	0.27	0.38	0.21	0.00	0.00	0.17	-0.15	0.04	-0.07	0.17	A-	A+	A-	A-	2.48	0.06	6.91	1.25	8.88	1.67
211	818274	18	5841	0.54	0.08	0.28	0.09	0.54	0.00	0.00	0.38	-0.24	-0.16	-0.18	0.38	A+	A-	A+	A+	0.57	0.05	1.41	1.03	0.65	1.02
212	818294	18	5841	0.47	0.16	0.18	0.47	0.19	0.00	0.00	0.35	-0.19	-0.08	0.35	-0.19	A+	A-	A-	A+	0.93	0.05	5.71	1.12	4.73	1.16
213	819215	18	5841	0.64	0.64	0.13	0.18	0.05	0.00	0.00	0.51	0.51	-0.29	-0.26	-0.22	A+	A-	A-	A-	0.04	0.05	-6.06	0.87	-5.38	0.79
214	818291	18	5841	0.81	0.04	0.81	0.06	0.09	0.00	0.00	0.36	-0.16	0.36	-0.26	-0.16	A-	A-	A-	A+	-1.02	0.06	-0.54	0.98	0.83	1.06
215	819883	18	5841	0.64	0.20	0.09	0.64	0.07	0.00	0.00	0.51	-0.28	-0.27	0.51	-0.22	A+	A+	A-	A-	0.07	0.05	-6.46	0.86	-5.50	0.79
216	736730	18	5841	0.27	0.27	0.20	0.33	0.20	0.00	0.00	0.17	0.17	0.01	0.02	-0.21	A-	A-	A+	A-	2.09	0.06	8.62	1.27	9.90	1.71
217	816623	19	5775	0.64	0.08	0.24	0.64	0.04	0.00	0.00	0.52	-0.19	-0.41	0.52	-0.12	A+	A-	A-	A-	0.08	0.05	-5.89	0.88	-4.79	0.81
218	819082	19	5775	0.20	0.29	0.20	0.42	0.09	0.00	0.00	0.00	-0.10	0.00	0.13	-0.07	A+	A-	A-	A-	2.60	0.06	8.83	1.35	9.90	2.34
219	816620	19	5775	0.46	0.46	0.06	0.13	0.35	0.00	0.00	0.47	0.47	-0.19	-0.28	-0.20	A+	A-	A-	A-	1.00	0.05	-0.92	0.98	-0.34	0.99
220	818265	19	5775	0.54	0.54	0.13	0.26	0.06	0.00	0.00	0.36	0.36	-0.24	-0.12	-0.19	A+	A-	A-	A+	0.58	0.05	4.87	1.11	4.86	1.18
221	817155	19	5775	0.42	0.38	0.42	0.09	0.10	0.00	0.00	0.26	-0.04	0.26	-0.19	-0.17	A-	A-	A+	A-	1.22	0.05	9.49	1.22	9.13	1.34
222	724119	19	5775	0.63	0.08	0.63	0.22	0.07	0.00	0.00	0.44	-0.28	0.44	-0.21	-0.20	A+	A-	A+	A+	0.12	0.05	-2.26	0.95	-1.74	0.93
223	819224	19	5775	0.61	0.13	0.12	0.13	0.61	0.00	0.00	0.49	-0.22	-0.21	-0.28	0.49	A+	A-	A-	A+	0.23	0.05	-3.90	0.92	-3.18	0.88
224	816631	19	5775	0.58	0.10	0.14	0.58	0.18	0.00	0.00	0.33	-0.17	-0.19	0.33	-0.11	A+	A-	A-	A-	0.41	0.05	5.38	1.12	6.04	1.24
225	818389	19	5775	0.72	0.13	0.07	0.72	0.08	0.00	0.00	0.49	-0.26	-0.26	0.49	-0.24	A+	A+	A+	A+	-0.37	0.06	-5.17	0.87	-4.76	0.76
226	818775	19	5775	0.55	0.11	0.55	0.15	0.19	0.00	0.00	0.38	-0.23	0.38	-0.22	-0.10	A+	A-	A-	A+	0.56	0.05	2.96	1.06	2.57	1.09
227	820448	20	5845	0.25	0.08	0.36	0.25	0.30	0.00	0.00	0.15	-0.14	-0.08	0.15	0.03	A+	A+	A+	A-	2.20	0.06	7.19	1.23	9.90	1.73
228	820453	20	5845	0.29	0.20	0.29	0.18	0.32	0.00	0.00	0.17	-0.18	0.17	-0.18	0.15	A-	A+	A+	A-	1.89	0.06	9.90	1.29	9.90	1.60
229	820449	20	5845	0.43	0.43	0.32	0.15	0.09	0.00	0.00	0.26	0.26	-0.05	-0.15	-0.18	A+	A-	A+	A-	1.12	0.05	9.90	1.23	9.90	1.40
230	700828	20	5845	0.35	0.12	0.35	0.25	0.28	0.00	0.00	0.26	-0.14	0.26	-0.18	0.01	A-	A+	A+	A-	1.58	0.05	8.73	1.22	9.26	1.38
231	818382	20	5845	0.55	0.55	0.26	0.13	0.06	0.00	0.00	0.51	0.51	-0.23	-0.31	-0.20	A+	A-	A-	A-	0.54	0.05	-4.76	0.91	-3.30	0.89
232	819081	20	5845	0.38	0.14	0.29	0.19	0.38	0.00	0.00	0.32	-0.24	0.08	-0.27	0.32	A-	A+	A+	A-	1.41	0.05	6.02	1.14	6.22	1.23
233	819101	20	5845	0.44	0.20	0.44	0.11	0.24	0.00	0.00	0.56	-0.18	0.56	-0.30	-0.25	A-	A-	A-	A-	1.08	0.05	-6.30	0.87	-4.02	0.88
234	800481	20	5845	0.41	0.41	0.16	0.31	0.12	0.00	0.00	0.48	0.48	-0.21	-0.26	-0.11	A-	A-	A-	A+	1.22	0.05	-2.41	0.95	-1.04	0.97
235	818293	20	5845	0.33	0.16	0.33	0.31	0.20	0.00	0.00	0.24	-0.14	0.24	-0.20	0.08	A-	A+	A+	A-	1.71	0.05	7.08	1.19	8.91	1.39
236	819226	20	5845	0.55	0.07	0.30	0.55	0.07	0.00	0.00	0.29	-0.19	-0.14	0.29	-0.13	A-	A-	A-	A+	0.50	0.05	7.00	1.15	5.66	1.20

Table J–6. Biology Multiple-Choice Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	677970	0	142262	0.69	0.12	0.11	0.69	0.08	0.00	0.00	0.39	-0.19	-0.25	0.39	-0.15					-0.35	0.01	-9.90	0.91	-9.90	0.84
2	674135	0	142262	0.68	0.14	0.04	0.68	0.14	0.00	0.00	0.40	-0.28	-0.19	0.40	-0.15					-0.64	0.01	5.02	1.03	-1.60	0.98
3	642329	0	142262	0.43	0.05	0.10	0.41	0.43	0.00	0.00	0.37	-0.22	-0.28	-0.10	0.37					0.60	0.01	9.90	1.04	9.90	1.07
4	674080	0	142262	0.67	0.67	0.11	0.13	0.09	0.00	0.00	0.45	0.45	-0.29	-0.23	-0.15					0.05	0.01	-9.90	0.86	-9.90	0.79
5	740934	0	142262	0.55	0.05	0.55	0.31	0.09	0.00	0.00	0.41	-0.23	0.41	-0.21	-0.20					0.24	0.01	-5.12	0.98	-7.15	0.95
6	674139	0	142262	0.72	0.14	0.72	0.07	0.07	0.00	0.00	0.45	-0.22	0.45	-0.26	-0.22					-0.35	0.01	-9.90	0.80	-9.90	0.72
7	735097	0	142262	0.55	0.07	0.13	0.55	0.25	0.00	0.00	0.44	-0.30	-0.25	0.44	-0.13					-0.07	0.01	2.69	1.01	-3.11	0.98
8	739696	0	142262	0.72	0.72	0.13	0.10	0.05	0.00	0.00	0.43	0.43	-0.29	-0.19	-0.16					-0.52	0.01	-9.90	0.86	-9.90	0.79
9	735494	0	142262	0.64	0.17	0.09	0.09	0.64	0.00	0.00	0.42	-0.13	-0.29	-0.24	0.42					-0.19	0.01	-9.90	0.93	-9.90	0.90
10	736847	0	142262	0.61	0.12	0.61	0.14	0.14	0.00	0.00	0.35	-0.08	0.35	-0.21	-0.20					0.15	0.01	7.34	1.03	9.90	1.10
11	679666	0	142262	0.69	0.69	0.10	0.14	0.07	0.00	0.00	0.49	0.49	-0.24	-0.30	-0.19					-0.67	0.01	-9.90	0.91	-9.90	0.79
12	679665	0	142262	0.61	0.13	0.11	0.61	0.15	0.00	0.00	0.32	-0.21	-0.19	0.32	-0.07					-0.01	0.01	9.90	1.08	9.90	1.08
13	735301	0	142262	0.50	0.13	0.50	0.20	0.17	0.01	0.00	0.31	-0.14	0.31	-0.20	-0.06					0.49	0.01	9.90	1.08	9.90	1.08
14	739676	0	142262	0.69	0.14	0.09	0.69	0.07	0.01	0.00	0.37	-0.10	-0.26	0.37	-0.21					-0.59	0.01	2.80	1.02	6.25	1.07
15	678965	0	142262	0.41	0.41	0.22	0.12	0.24	0.01	0.00	0.42	0.42	-0.15	-0.28	-0.10					0.79	0.01	-7.06	0.97	-5.13	0.97
16	714620	0	142262	0.54	0.22	0.19	0.54	0.05	0.01	0.00	0.27	-0.01	-0.21	0.27	-0.18					0.12	0.01	9.90	1.16	9.90	1.19
17	739671	0	142262	0.53	0.12	0.53	0.14	0.21	0.01	0.00	0.43	-0.26	0.43	-0.26	-0.09					-0.03	0.01	7.87	1.04	5.04	1.04
18	714626	0	142262	0.59	0.59	0.07	0.12	0.21	0.01	0.00	0.33	0.33	-0.23	-0.19	-0.08					0.22	0.01	9.90	1.05	6.42	1.04
19	678419	0	142262	0.45	0.21	0.13	0.45	0.20	0.01	0.00	0.44	-0.14	-0.23	0.44	-0.18					0.74	0.01	-9.90	0.94	-9.90	0.93
20	702101	0	142262	0.53	0.17	0.18	0.11	0.53	0.01	0.00	0.35	-0.17	-0.15	-0.15	0.35					0.08	0.01	9.90	1.12	9.90	1.16
21	741369	0	142262	0.64	0.12	0.64	0.14	0.09	0.01	0.00	0.47	-0.17	0.47	-0.32	-0.17					-0.44	0.01	-9.90	0.93	-5.80	0.94
22	736841	0	142262	0.51	0.14	0.13	0.51	0.21	0.01	0.00	0.35	-0.16	-0.15	0.35	-0.14					0.42	0.01	9.90	1.07	9.90	1.10
23	740016	0	142262	0.62	0.62	0.11	0.16	0.10	0.01	0.00	0.53	0.53	-0.28	-0.26	-0.21					-0.41	0.01	-9.90	0.89	-9.90	0.79
24	740018	0	142262	0.53	0.53	0.14	0.24	0.08	0.01	0.00	0.39	0.39	-0.13	-0.19	-0.22					0.11	0.01	9.90	1.05	5.65	1.04
25	643400	0	142262	0.45	0.19	0.13	0.45	0.23	0.00	0.00	0.34	-0.22	-0.27	0.34	0.02					0.95	0.01	9.90	1.08	9.90	1.12
26	739971	0	142262	0.46	0.14	0.27	0.46	0.12	0.00	0.00	0.36	-0.11	-0.15	0.36	-0.21					0.49	0.01	9.90	1.05	7.99	1.05
27	741663	0	142262	0.57	0.10	0.22	0.11	0.57	0.00	0.00	0.27	-0.13	-0.05	-0.23	0.27					0.20	0.01	9.90	1.12	9.90	1.15
28	678303	0	142262	0.43	0.43	0.25	0.19	0.13	0.00	0.00	0.40	0.40	-0.32	-0.08	-0.08					1.03	0.01	4.03	1.02	8.58	1.06
29	741014	0	142262	0.76	0.04	0.13	0.08	0.76	0.00	0.00	0.40	-0.20	-0.19	-0.26	0.40					-1.10	0.01	-2.67	0.98	-2.22	0.97

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	641211	0	142262	0.41	0.41	0.16	0.29	0.14	0.00	0.00	0.25	0.25	-0.09	-0.10	-0.13					0.97	0.01	9.90	1.16	9.90	1.23
31	703161	0	142262	0.61	0.13	0.15	0.61	0.12	0.00	0.00	0.40	-0.20	-0.17	0.40	-0.22					-0.08	0.01	-2.74	0.99	2.50	1.02
32	683582	0	142262	0.71	0.71	0.11	0.12	0.05	0.00	0.00	0.54	0.54	-0.29	-0.31	-0.23					-0.75	0.01	-9.90	0.79	-9.90	0.65
33	683555	0	142262	0.54	0.54	0.10	0.07	0.29	0.00	0.00	0.42	0.42	-0.20	-0.23	-0.20					-0.07	0.01	6.61	1.03	4.75	1.04
34	703159	0	142262	0.72	0.18	0.72	0.05	0.05	0.00	0.00	0.39	-0.26	0.39	-0.21	-0.13					-0.54	0.01	-9.90	0.91	-9.90	0.89
35	741291	0	142262	0.55	0.55	0.18	0.15	0.12	0.00	0.00	0.49	0.49	-0.17	-0.30	-0.20					0.10	0.01	-9.90	0.91	-9.90	0.86
36	736565	0	142262	0.45	0.11	0.45	0.24	0.20	0.00	0.00	0.31	-0.20	0.31	-0.11	-0.11					0.62	0.01	9.90	1.10	9.90	1.12
37	610870	0	142262	0.57	0.12	0.57	0.12	0.18	0.00	0.00	0.33	-0.25	0.33	-0.18	-0.04					0.09	0.01	9.90	1.09	9.90	1.13
38	677886	0	142262	0.53	0.53	0.18	0.12	0.16	0.00	0.00	0.42	0.42	-0.18	-0.25	-0.14					0.17	0.01	-2.84	0.99	-5.00	0.97
39	607737	0	142262	0.52	0.12	0.52	0.19	0.17	0.00	0.00	0.35	-0.10	0.35	-0.18	-0.18					0.24	0.01	9.90	1.07	9.60	1.07
40	742296	0	142262	0.53	0.11	0.14	0.22	0.53	0.00	0.00	0.37	-0.20	-0.11	-0.19	0.37					0.18	0.01	9.90	1.07	9.90	1.09
41	642901	0	142262	0.59	0.19	0.59	0.12	0.09	0.00	0.00	0.42	-0.23	0.42	-0.21	-0.14					0.20	0.01	-9.90	0.94	-9.90	0.92
42	683552	0	142262	0.65	0.65	0.14	0.08	0.12	0.00	0.00	0.52	0.52	-0.25	-0.25	-0.26					-0.34	0.01	-9.90	0.83	-9.90	0.78
43	678986	0	142262	0.69	0.06	0.16	0.08	0.69	0.00	0.00	0.54	-0.27	-0.27	-0.30	0.54					-0.82	0.01	-9.90	0.86	-9.90	0.73
44	739770	0	142262	0.63	0.05	0.20	0.63	0.11	0.00	0.00	0.34	-0.14	-0.25	0.34	-0.09					-0.41	0.01	9.90	1.11	9.90	1.14
45	678894	0	142262	0.44	0.18	0.44	0.14	0.24	0.00	0.00	0.37	-0.24	0.37	-0.25	0.00					0.76	0.01	8.03	1.03	8.87	1.05
46	642871	0	142262	0.45	0.23	0.45	0.18	0.13	0.00	0.00	0.34	-0.03	0.34	-0.25	-0.16					0.66	0.01	9.90	1.07	9.90	1.09
47	742284	0	142262	0.71	0.17	0.06	0.71	0.05	0.00	0.00	0.40	-0.15	-0.27	0.40	-0.24					-0.70	0.01	-8.26	0.95	3.07	1.04
48	742323	0	142262	0.63	0.63	0.13	0.11	0.12	0.00	0.00	0.55	0.55	-0.29	-0.28	-0.22					-0.34	0.01	-9.90	0.84	-9.90	0.75
49	809556	1	5791	0.50	0.11	0.50	0.19	0.20	0.00	0.00	0.33	-0.17	0.33	-0.09	-0.19	A+	A-	A-	A-	0.65	0.05	3.92	1.07	3.94	1.10
50	798792	1	5791	0.57	0.16	0.16	0.57	0.11	0.00	0.00	0.36	-0.15	-0.09	0.36	-0.28	A+	A-	A-	A+	0.32	0.05	1.31	1.02	1.03	1.03
51	816619	1	5791	0.82	0.04	0.09	0.06	0.82	0.00	0.00	0.37	-0.27	-0.13	-0.24	0.37	A+	A-	A-	A+	-1.10	0.06	-2.18	0.93	-2.49	0.85
52	808703	1	5791	0.59	0.59	0.15	0.11	0.14	0.00	0.00	0.41	0.41	-0.19	-0.21	-0.19	A+	A-	A-	A+	0.22	0.05	-2.47	0.95	-3.07	0.91
53	741377	1	5791	0.30	0.36	0.30	0.20	0.14	0.00	0.00	0.09	0.15	0.09	-0.12	-0.17	A-	A+	A+	A-	1.68	0.05	9.74	1.25	9.90	1.43
54	714647	1	5791	0.23	0.23	0.19	0.21	0.37	0.00	0.00	0.18	0.18	-0.06	-0.16	0.03	A-	A-	A+	A+	2.11	0.06	4.12	1.13	8.26	1.48
55	809877	1	5791	0.29	0.29	0.09	0.21	0.41	0.00	0.00	0.04	0.04	-0.21	-0.03	0.11	A-	A+	A+	A-	1.74	0.05	9.90	1.34	9.90	1.68
56	809696	1	5791	0.27	0.40	0.27	0.13	0.19	0.00	0.00	0.13	0.04	0.13	-0.20	-0.02	A-	A+	A-	A-	1.85	0.05	4.21	1.11	8.80	1.43
57	799652	1	5791	0.29	0.19	0.29	0.26	0.25	0.00	0.00	0.21	-0.05	0.21	-0.19	0.02	A-	A+	A+	A+	1.74	0.05	4.36	1.11	7.22	1.32
58	799645	1	5791	0.53	0.17	0.53	0.11	0.18	0.00	0.00	0.41	-0.20	0.41	-0.29	-0.09	A-	A-	A+	A+	0.50	0.05	-1.78	0.97	-1.40	0.96
59	740866	1	5791	0.55	0.22	0.13	0.55	0.10	0.00	0.00	0.39	-0.10	-0.27	0.39	-0.20	A-	A-	A+	A-	0.40	0.05	-0.05	1.00	-0.10	1.00

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
60	741474	1	5791	0.34	0.10	0.35	0.21	0.34	0.00	0.00	0.36	-0.14	-0.07	-0.22	0.36	A+	A+	A+	A+	1.46	0.05	-0.28	0.99	2.54	1.09
61	810528	1	5791	0.31	0.31	0.31	0.13	0.24	0.00	0.00	0.19	0.00	0.19	-0.20	-0.04	A+	A-	A-	A+	1.63	0.05	7.16	1.18	9.90	1.43
62	808547	1	5791	0.40	0.39	0.40	0.13	0.08	0.00	0.00	0.23	0.08	0.23	-0.25	-0.23	A-	A-	A-	A+	1.16	0.05	7.26	1.15	7.93	1.24
63	809199	1	5791	0.72	0.08	0.04	0.72	0.16	0.00	0.00	0.28	-0.16	-0.25	0.28	-0.08	A+	A-	A-	A-	-0.48	0.05	4.42	1.11	5.68	1.27
64	734725	1	5791	0.64	0.05	0.08	0.64	0.22	0.00	0.00	0.51	-0.21	-0.26	0.51	-0.30	A+	A-	A-	A+	-0.05	0.05	-7.08	0.86	-6.30	0.80
65	809558	2	5681	0.36	0.17	0.36	0.17	0.29	0.00	0.00	0.16	-0.11	0.16	-0.08	-0.01	A-	A-	A-	A-	1.34	0.05	8.68	1.20	9.90	1.37
66	798794	2	5681	0.67	0.67	0.13	0.12	0.07	0.00	0.00	0.48	0.48	-0.27	-0.24	-0.20	A-	A-	A-	A-	-0.23	0.05	-4.82	0.89	-4.56	0.83
67	821212	2	5681	0.43	0.34	0.13	0.43	0.10	0.00	0.00	0.44	-0.23	-0.10	0.44	-0.25	A-	A-	A+	A-	1.01	0.05	-3.76	0.93	-2.37	0.93
68	808359	2	5681	0.47	0.47	0.15	0.08	0.30	0.00	0.00	0.34	0.34	-0.18	-0.28	-0.05	A-	A+	A-	A+	0.80	0.05	3.26	1.06	3.15	1.09
69	809466	2	5681	0.63	0.08	0.13	0.63	0.16	0.00	0.00	0.41	-0.16	-0.24	0.41	-0.21	A+	A-	A-	A-	0.00	0.05	-1.13	0.98	0.43	1.01
70	809872	2	5681	0.67	0.67	0.10	0.12	0.10	0.00	0.00	0.47	0.47	-0.24	-0.28	-0.18	A-	A-	A-	A+	-0.25	0.05	-3.37	0.92	-4.05	0.84
71	734731	2	5681	0.28	0.28	0.22	0.24	0.26	0.00	0.00	0.16	0.16	-0.09	-0.06	-0.02	A-	A-	A+	A+	1.84	0.06	7.62	1.22	9.37	1.47
72	809699	2	5681	0.71	0.08	0.71	0.11	0.09	0.00	0.00	0.47	-0.24	0.47	-0.25	-0.23	A+	A-	A-	A+	-0.45	0.05	-3.91	0.91	-2.66	0.88
73	809510	2	5681	0.48	0.24	0.48	0.18	0.10	0.00	0.00	0.21	-0.01	0.21	-0.03	-0.29	A+	A-	A-	A-	0.75	0.05	9.90	1.23	9.90	1.32
74	809509	2	5681	0.68	0.12	0.09	0.68	0.11	0.00	0.00	0.46	-0.27	-0.19	0.46	-0.23	A+	A-	A-	A+	-0.28	0.05	-3.42	0.92	-2.23	0.91
75	808872	2	5681	0.22	0.22	0.09	0.07	0.61	0.00	0.00	0.29	0.29	-0.24	-0.29	0.06	A-	A+	A-	A-	2.19	0.06	-1.30	0.96	3.79	1.22
76	700896	2	5681	0.41	0.22	0.41	0.17	0.20	0.00	0.00	0.50	-0.33	0.50	-0.19	-0.10	A+	A+	A+	A+	1.10	0.05	-6.83	0.87	-5.37	0.85
77	741048	2	5681	0.77	0.14	0.77	0.04	0.04	0.00	0.00	0.30	-0.11	0.30	-0.25	-0.18	A+	A+	A-	A+	-0.82	0.06	-0.18	0.99	2.72	1.16
78	730014	2	5681	0.52	0.14	0.10	0.52	0.23	0.00	0.00	0.44	-0.28	-0.20	0.44	-0.14	A-	B-	A-	A-	0.52	0.05	-1.49	0.97	-1.11	0.97
79	739782	2	5681	0.45	0.45	0.25	0.18	0.11	0.00	0.00	0.38	0.38	-0.07	-0.28	-0.15	A+	A-	A+	A-	0.90	0.05	1.20	1.02	1.17	1.03
80	737649	2	5681	0.52	0.22	0.10	0.17	0.52	0.00	0.00	0.39	-0.12	-0.24	-0.20	0.39	A+	A-	A-	A-	0.56	0.05	0.09	1.00	0.37	1.01
81	800306	3	5602	0.27	0.04	0.53	0.15	0.27	0.00	0.00	0.14	-0.21	0.12	-0.22	0.14	A-	A-	A-	A+	1.84	0.06	6.80	1.20	8.28	1.42
82	800304	3	5602	0.30	0.30	0.14	0.23	0.32	0.00	0.00	0.19	0.19	-0.11	-0.06	-0.04	A-	A+	A-	A+	1.66	0.05	6.77	1.18	8.25	1.37
83	808537	3	5602	0.54	0.25	0.54	0.15	0.06	0.00	0.00	0.30	-0.19	0.30	-0.07	-0.18	A+	A+	A+	A-	0.46	0.05	5.88	1.12	4.96	1.15
84	808544	3	5602	0.50	0.32	0.50	0.07	0.11	0.00	0.00	0.23	-0.10	0.23	-0.12	-0.13	A-	A-	A-	A+	0.62	0.05	7.47	1.15	6.55	1.19
85	808548	3	5602	0.64	0.18	0.06	0.12	0.64	0.00	0.00	0.42	-0.19	-0.19	-0.26	0.42	A+	A-	A-	A-	-0.08	0.05	-3.12	0.93	-2.72	0.90
86	809884	3	5602	0.38	0.38	0.24	0.06	0.32	0.00	0.00	0.44	0.44	-0.19	-0.29	-0.13	A+	A-	A-	A+	1.23	0.05	-2.80	0.94	0.96	1.03
87	809467	3	5602	0.78	0.04	0.10	0.78	0.07	0.00	0.00	0.41	-0.22	-0.21	0.41	-0.23	A+	A-	A-	A-	-0.90	0.06	-1.85	0.94	-1.04	0.94
88	810611	3	5602	0.76	0.06	0.07	0.76	0.11	0.00	0.00	0.52	-0.27	-0.27	0.52	-0.28	A+	A-	A-	A+	-0.76	0.06	-5.47	0.85	-5.28	0.74
89	809508	3	5602	0.44	0.44	0.14	0.26	0.17	0.00	0.00	0.32	0.32	-0.18	-0.13	-0.10	A-	A+	A-	A-	0.93	0.05	2.96	1.06	2.90	1.08

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
90	809511	3	5602	0.46	0.08	0.15	0.30	0.46	0.00	0.00	0.43	-0.25	-0.21	-0.15	0.43	A-	A-	A-	A+	0.84	0.05	-0.73	0.99	0.71	1.02
91	808874	3	5602	0.66	0.10	0.66	0.06	0.18	0.00	0.00	0.38	-0.17	0.38	-0.25	-0.18	A-	A-	A-	A+	-0.18	0.05	-0.62	0.99	2.15	1.09
92	679972	3	5602	0.71	0.71	0.17	0.06	0.06	0.00	0.00	0.43	0.43	-0.24	-0.20	-0.24	A+	A-	A-	A+	-0.44	0.06	-2.93	0.93	-1.70	0.92
93	813415	3	5602	0.56	0.10	0.15	0.56	0.19	0.00	0.00	0.30	-0.19	-0.10	0.30	-0.14	A+	A-	A-	A+	0.35	0.05	5.92	1.12	4.06	1.12
94	811170	3	5602	0.70	0.07	0.08	0.15	0.70	0.00	0.00	0.47	-0.28	-0.30	-0.18	0.47	A+	A-	A-	A+	-0.38	0.05	-3.63	0.91	-2.47	0.90
95	816149	3	5602	0.46	0.23	0.46	0.13	0.17	0.00	0.00	0.35	-0.17	0.35	-0.24	-0.05	A-	A+	A+	A-	0.84	0.05	2.12	1.04	2.45	1.07
96	808357	3	5602	0.32	0.05	0.11	0.51	0.32	0.00	0.00	0.24	-0.18	-0.21	-0.02	0.24	A+	A+	A+	A+	1.55	0.05	5.28	1.13	7.08	1.29
97	800301	4	5692	0.45	0.14	0.32	0.08	0.45	0.00	0.00	0.47	-0.20	-0.23	-0.21	0.47	A-	A-	A-	A+	0.88	0.05	-4.81	0.91	-3.61	0.90
98	800302	4	5692	0.59	0.59	0.17	0.11	0.12	0.00	0.00	0.47	0.47	-0.21	-0.26	-0.20	A+	A-	A-	A-	0.19	0.05	-3.94	0.92	-3.58	0.89
99	809867	4	5692	0.46	0.11	0.46	0.09	0.35	0.00	0.00	0.21	-0.25	0.21	-0.12	0.02	A-	A-	A-	A+	0.85	0.05	9.90	1.26	9.90	1.33
100	816511	4	5692	0.71	0.08	0.09	0.71	0.12	0.00	0.00	0.50	-0.23	-0.28	0.50	-0.25	A-	A-	A-	A+	-0.47	0.05	-5.49	0.87	-5.30	0.77
101	810613	4	5692	0.55	0.11	0.55	0.15	0.18	0.00	0.00	0.50	-0.25	0.50	-0.20	-0.24	A+	A-	A-	A-	0.37	0.05	-2.88	0.94	-1.73	0.95
102	808017	4	5692	0.46	0.46	0.18	0.13	0.22	0.00	0.00	0.42	0.42	-0.14	-0.32	-0.12	A-	A-	A-	A+	0.82	0.05	-2.43	0.95	-2.01	0.95
103	816143	4	5692	0.67	0.08	0.67	0.12	0.13	0.00	0.00	0.36	-0.20	0.36	-0.25	-0.10	A+	A-	A+	A+	-0.22	0.05	-0.96	0.98	-0.16	0.99
104	811933	4	5692	0.45	0.21	0.45	0.21	0.12	0.00	0.00	0.39	-0.15	0.39	-0.15	-0.21	A-	A+	A-	A-	0.86	0.05	1.15	1.02	1.84	1.05
105	798801	4	5692	0.56	0.08	0.14	0.22	0.56	0.00	0.00	0.34	-0.24	-0.14	-0.12	0.34	A+	A-	A-	A-	0.33	0.05	3.47	1.07	3.20	1.10
106	809557	4	5692	0.53	0.08	0.21	0.53	0.18	0.00	0.00	0.44	-0.26	-0.25	0.44	-0.12	A-	A-	A-	A-	0.48	0.05	-0.74	0.99	-1.11	0.97
107	813412	4	5692	0.55	0.11	0.18	0.55	0.15	0.00	0.00	0.39	-0.23	-0.19	0.39	-0.12	A+	A+	A-	A-	0.41	0.05	1.39	1.03	0.28	1.01
108	741037	4	5692	0.75	0.75	0.06	0.10	0.09	0.00	0.00	0.48	0.48	-0.29	-0.25	-0.21	A+	A-	A-	A-	-0.72	0.06	-3.86	0.89	-3.95	0.80
109	808875	4	5692	0.43	0.28	0.08	0.43	0.21	0.00	0.00	0.22	-0.03	-0.26	0.22	-0.05	A-	A-	A-	A-	1.00	0.05	9.46	1.20	9.44	1.29
110	816426	4	5692	0.53	0.08	0.29	0.53	0.10	0.00	0.00	0.39	-0.27	-0.13	0.39	-0.19	A-	A-	A-	A+	0.47	0.05	1.97	1.04	2.17	1.06
111	721622	4	5692	0.67	0.67	0.12	0.09	0.12	0.00	0.00	0.53	0.53	-0.25	-0.29	-0.25	A+	A+	A-	A-	-0.25	0.05	-6.35	0.86	-5.30	0.80
112	742280	4	5692	0.41	0.12	0.32	0.41	0.15	0.00	0.00	0.18	-0.17	0.07	0.18	-0.19	A-	A-	A-	A+	1.07	0.05	9.90	1.25	9.90	1.34
113	813701	5	5675	0.40	0.24	0.19	0.40	0.17	0.00	0.00	0.25	-0.13	-0.07	0.25	-0.10	A-	A-	A-	A+	1.14	0.05	5.92	1.13	6.38	1.20
114	813702	5	5675	0.48	0.48	0.16	0.14	0.22	0.00	0.00	0.42	0.42	-0.22	-0.23	-0.11	A-	A-	A-	A-	0.76	0.05	-2.14	0.96	-2.29	0.94
115	809461	5	5675	0.63	0.12	0.17	0.63	0.09	0.00	0.00	0.37	-0.28	-0.12	0.37	-0.16	A-	A-	A-	A+	0.01	0.05	1.17	1.02	0.59	1.02
116	809871	5	5675	0.49	0.28	0.13	0.49	0.09	0.00	0.00	0.28	0.04	-0.26	0.28	-0.23	A+	A-	A-	A+	0.67	0.05	7.87	1.16	8.69	1.25
117	816603	5	5675	0.23	0.44	0.23	0.26	0.07	0.00	0.00	0.02	-0.11	0.02	0.16	-0.10	A-	A+	A+	A-	2.10	0.06	7.35	1.24	9.90	1.69
118	809875	5	5675	0.57	0.57	0.15	0.09	0.20	0.00	0.00	0.48	0.48	-0.13	-0.29	-0.27	A-	A-	A-	A+	0.30	0.05	-4.42	0.92	-3.01	0.91
119	809688	5	5675	0.40	0.12	0.36	0.40	0.11	0.00	0.00	0.23	-0.14	-0.04	0.23	-0.15	A+	A+	A-	A-	1.16	0.05	6.32	1.13	6.21	1.20

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
120	810620	5	5675	0.51	0.18	0.51	0.15	0.15	0.00	0.00	0.44	-0.16	0.44	-0.21	-0.23	A+	A-	A-	A+	0.58	0.05	-2.34	0.96	-2.20	0.94
121	798804	5	5675	0.47	0.47	0.14	0.11	0.28	0.00	0.00	0.43	0.43	-0.16	-0.28	-0.15	A-	A+	A-	A-	0.79	0.05	-2.62	0.95	-1.37	0.96
122	798803	5	5675	0.63	0.08	0.08	0.63	0.21	0.00	0.00	0.41	-0.27	-0.30	0.41	-0.10	A-	A-	A-	A-	0.01	0.05	-0.15	1.00	-0.21	0.99
123	810326	5	5675	0.45	0.45	0.34	0.15	0.05	0.00	0.00	0.30	0.30	-0.08	-0.19	-0.17	A-	A-	A-	A-	0.87	0.05	4.99	1.10	4.22	1.12
124	811179	5	5675	0.59	0.05	0.13	0.59	0.22	0.00	0.00	0.43	-0.27	-0.24	0.43	-0.16	A+	A-	A-	A+	0.20	0.05	-2.28	0.96	-0.40	0.99
125	809289	5	5675	0.68	0.19	0.04	0.68	0.08	0.00	0.00	0.45	-0.22	-0.26	0.45	-0.25	A+	A-	A-	A+	-0.27	0.05	-3.68	0.92	-3.96	0.85
126	810636	5	5675	0.51	0.11	0.51	0.32	0.05	0.00	0.00	0.14	-0.06	0.14	-0.06	-0.10	A+	A+	A-	A+	0.57	0.05	9.90	1.27	9.90	1.37
127	679966	5	5675	0.72	0.17	0.72	0.09	0.03	0.00	0.00	0.42	-0.24	0.42	-0.26	-0.14	A+	A-	A-	A-	-0.48	0.06	-0.48	0.99	-1.49	0.93
128	739776	5	5675	0.40	0.40	0.16	0.24	0.20	0.00	0.00	0.48	0.48	-0.09	-0.19	-0.28	A-	A-	A-	A+	1.15	0.05	-6.61	0.87	-4.34	0.88
129	813703	6	5729	0.49	0.49	0.12	0.25	0.14	0.00	0.00	0.29	0.29	-0.12	-0.13	-0.15	A+	A-	A-	A-	0.66	0.05	5.89	1.12	5.20	1.15
130	813706	6	5729	0.45	0.14	0.20	0.20	0.45	0.00	0.00	0.43	-0.15	-0.24	-0.16	0.43	A+	A-	A-	A-	0.83	0.05	-0.86	0.98	-0.18	0.99
131	816604	6	5729	0.70	0.10	0.70	0.06	0.14	0.00	0.00	0.52	-0.25	0.52	-0.23	-0.32	A-	A-	A-	A-	-0.44	0.05	-6.04	0.86	-5.23	0.79
132	810029	6	5729	0.32	0.21	0.13	0.32	0.34	0.00	0.00	0.06	-0.05	-0.25	0.06	0.16	A+	A-	A+	A-	1.52	0.05	9.90	1.37	9.90	1.63
133	739674	6	5729	0.47	0.47	0.08	0.36	0.09	0.00	0.00	0.53	0.53	-0.18	-0.38	-0.11	A+	A-	A-	A-	0.76	0.05	-7.18	0.87	-5.58	0.85
134	809155	6	5729	0.51	0.17	0.13	0.51	0.19	0.00	0.00	0.49	-0.24	-0.26	0.49	-0.17	A+	A-	A-	A+	0.57	0.05	-3.24	0.94	-2.98	0.92
135	739673	6	5729	0.60	0.60	0.18	0.06	0.16	0.00	0.00	0.14	0.14	0.04	-0.14	-0.14	A-	A-	A+	A+	0.11	0.05	9.90	1.24	9.90	1.43
136	809460	6	5729	0.66	0.10	0.14	0.66	0.09	0.00	0.00	0.46	-0.21	-0.27	0.46	-0.20	A+	A-	A-	A+	-0.21	0.05	-3.26	0.93	-3.78	0.86
137	798802	6	5729	0.24	0.33	0.31	0.12	0.24	0.00	0.00	0.22	0.11	-0.16	-0.21	0.22	A+	A+	A-	A-	2.00	0.06	5.11	1.16	7.22	1.41
138	798796	6	5729	0.67	0.67	0.07	0.09	0.17	0.00	0.00	0.33	0.33	-0.27	-0.21	-0.07	A+	A-	A-	A+	-0.25	0.05	0.21	1.00	0.93	1.04
139	816148	6	5729	0.43	0.15	0.17	0.43	0.24	0.00	0.00	0.33	-0.29	-0.30	0.33	0.12	A+	A-	A-	A+	0.93	0.05	3.49	1.07	4.02	1.12
140	808877	6	5729	0.53	0.13	0.25	0.53	0.08	0.00	0.00	0.34	-0.15	-0.14	0.34	-0.21	A+	A+	A+	A+	0.45	0.05	4.65	1.09	3.61	1.10
141	713515	6	5729	0.23	0.23	0.32	0.32	0.12	0.00	0.00	0.16	0.16	0.00	0.02	-0.23	A+	A-	A+	A-	2.08	0.06	5.42	1.18	9.90	1.69
142	810571	6	5729	0.47	0.47	0.12	0.25	0.16	0.00	0.00	0.41	0.41	-0.22	-0.17	-0.15	A-	A-	A-	A+	0.75	0.05	0.79	1.02	1.24	1.03
143	810527	6	5729	0.57	0.22	0.03	0.57	0.17	0.00	0.00	0.38	-0.18	-0.19	0.38	-0.20	A+	A-	A-	A-	0.26	0.05	1.42	1.03	0.59	1.02
144	742309	6	5729	0.66	0.09	0.66	0.11	0.14	0.00	0.00	0.49	-0.21	0.49	-0.30	-0.22	A+	A-	A-	A+	-0.19	0.05	-4.74	0.90	-4.92	0.83
145	799626	7	5658	0.45	0.37	0.45	0.06	0.12	0.00	0.00	0.10	0.06	0.10	-0.16	-0.12	A-	A-	A-	A-	0.92	0.05	9.90	1.34	9.90	1.47
146	799630	7	5658	0.48	0.18	0.25	0.48	0.10	0.00	0.00	0.23	-0.15	0.01	0.23	-0.21	A+	A-	A-	A+	0.77	0.05	9.20	1.18	9.63	1.28
147	812930	7	5658	0.62	0.19	0.09	0.62	0.10	0.00	0.00	0.43	-0.17	-0.23	0.43	-0.25	A+	A-	A-	A+	0.09	0.05	-0.26	0.99	0.59	1.02
148	809690	7	5658	0.59	0.59	0.10	0.18	0.13	0.00	0.00	0.47	0.47	-0.18	-0.33	-0.15	A+	A-	A-	A-	0.21	0.05	-4.31	0.92	-4.95	0.85
149	809870	7	5658	0.64	0.17	0.08	0.10	0.64	0.00	0.00	0.42	-0.13	-0.30	-0.23	0.42	A+	A-	A-	A-	-0.06	0.05	-0.90	0.98	-1.30	0.95

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
150	810023	7	5658	0.74	0.14	0.07	0.74	0.05	0.00	0.00	0.36	-0.16	-0.22	0.36	-0.21	A-	A-	B-	A+	-0.59	0.06	0.41	1.01	2.24	1.12
151	809874	7	5658	0.47	0.08	0.15	0.30	0.47	0.00	0.00	0.29	-0.15	-0.11	-0.14	0.29	A+	A-	A-	A+	0.80	0.05	6.41	1.13	5.21	1.14
152	714651	7	5658	0.52	0.21	0.14	0.12	0.52	0.00	0.00	0.45	-0.11	-0.24	-0.29	0.45	A+	A-	A-	A+	0.55	0.05	-3.31	0.94	-2.58	0.93
153	799959	7	5658	0.63	0.23	0.08	0.06	0.63	0.00	0.00	0.20	-0.03	-0.07	-0.26	0.20	A+	A-	A+	A-	0.00	0.05	9.18	1.21	9.90	1.44
154	809553	7	5658	0.56	0.22	0.10	0.56	0.12	0.00	0.00	0.45	-0.19	-0.29	0.45	-0.17	A-	A+	A-	A-	0.36	0.05	-4.00	0.92	-4.48	0.88
155	741475	7	5658	0.47	0.47	0.33	0.17	0.04	0.00	0.00	0.35	0.35	-0.15	-0.17	-0.22	A-	A+	A-	A+	0.82	0.05	3.49	1.07	2.67	1.07
156	741019	7	5658	0.51	0.20	0.51	0.14	0.15	0.00	0.00	0.41	-0.19	0.41	-0.27	-0.09	A+	A+	A-	A-	0.59	0.05	0.55	1.01	0.48	1.01
157	809056	7	5658	0.35	0.19	0.20	0.35	0.25	0.00	0.00	0.18	-0.11	0.01	0.18	-0.09	A+	A+	A-	A-	1.41	0.05	8.49	1.20	9.10	1.33
158	737648	7	5658	0.56	0.56	0.13	0.11	0.20	0.00	0.00	0.22	0.22	-0.15	-0.22	0.03	A-	A-	A-	A+	0.37	0.05	9.62	1.20	9.35	1.29
159	810564	7	5658	0.58	0.14	0.15	0.58	0.13	0.00	0.00	0.36	-0.19	-0.22	0.36	-0.09	A+	A+	A+	A+	0.28	0.05	1.70	1.03	-0.27	0.99
160	810559	7	5658	0.64	0.07	0.12	0.17	0.64	0.00	0.00	0.51	-0.26	-0.29	-0.21	0.51	A-	A-	A-	A+	-0.03	0.05	-6.15	0.87	-5.69	0.81
161	799632	8	5734	0.59	0.13	0.13	0.59	0.14	0.00	0.00	0.37	-0.13	-0.26	0.37	-0.13	A-	A-	A+	A-	0.18	0.05	2.52	1.05	1.86	1.06
162	799628	8	5734	0.49	0.25	0.49	0.11	0.15	0.00	0.00	0.34	-0.08	0.34	-0.23	-0.18	A-	A-	A+	A-	0.68	0.05	3.93	1.08	3.48	1.10
163	809695	8	5734	0.67	0.67	0.17	0.13	0.04	0.00	0.00	0.23	0.23	-0.15	-0.06	-0.16	A+	A-	A-	A+	-0.21	0.05	5.55	1.13	3.56	1.14
164	809686	8	5734	0.55	0.55	0.08	0.07	0.29	0.00	0.00	0.41	0.41	-0.24	-0.28	-0.14	A-	A-	A+	A+	0.38	0.05	1.63	1.03	2.44	1.07
165	741112	8	5734	0.78	0.07	0.08	0.78	0.07	0.00	0.00	0.45	-0.22	-0.27	0.45	-0.22	A+	A-	A-	A-	-0.89	0.06	-4.48	0.87	-4.39	0.76
166	816602	8	5734	0.63	0.12	0.15	0.11	0.63	0.00	0.00	0.47	-0.18	-0.28	-0.23	0.47	A+	A-	A+	A+	0.01	0.05	-2.75	0.94	-2.61	0.91
167	810024	8	5734	0.55	0.15	0.12	0.55	0.18	0.00	0.00	0.46	-0.25	-0.19	0.46	-0.20	A+	A-	A-	A+	0.41	0.05	-3.02	0.94	-3.25	0.91
168	736840	8	5734	0.48	0.17	0.20	0.14	0.48	0.00	0.00	0.42	-0.14	-0.19	-0.22	0.42	A-	A-	A-	A-	0.75	0.05	-2.69	0.95	-2.11	0.94
169	799958	8	5734	0.73	0.13	0.05	0.08	0.73	0.00	0.00	0.49	-0.21	-0.28	-0.29	0.49	A+	A-	A-	A+	-0.55	0.06	-5.54	0.86	-4.87	0.78
170	799955	8	5734	0.37	0.37	0.14	0.19	0.30	0.00	0.00	0.34	0.34	-0.17	-0.20	-0.05	A-	A+	A+	A-	1.29	0.05	1.39	1.03	2.80	1.09
171	808879	8	5734	0.71	0.71	0.07	0.17	0.04	0.00	0.00	0.51	0.51	-0.29	-0.28	-0.24	A+	A-	A-	A-	-0.45	0.05	-5.76	0.86	-5.60	0.77
172	816419	8	5734	0.32	0.39	0.32	0.20	0.08	0.00	0.00	0.18	0.02	0.18	-0.07	-0.22	A+	A+	A+	A-	1.55	0.05	7.52	1.19	9.90	1.41
173	809206	8	5734	0.65	0.11	0.08	0.65	0.16	0.00	0.00	0.43	-0.18	-0.24	0.43	-0.23	A+	A-	A-	A+	-0.14	0.05	-2.25	0.95	-1.71	0.94
174	810328	8	5734	0.45	0.15	0.45	0.16	0.24	0.00	0.00	0.31	-0.10	0.31	-0.25	-0.06	A+	A-	A-	A-	0.91	0.05	4.02	1.08	4.74	1.13
175	811181	8	5734	0.28	0.22	0.33	0.28	0.17	0.00	0.00	0.10	-0.15	0.10	0.10	-0.08	A-	A-	A-	A+	1.82	0.06	8.23	1.24	9.90	1.50
176	704207	8	5734	0.59	0.15	0.07	0.19	0.59	0.00	0.00	0.33	-0.14	-0.26	-0.12	0.33	A+	A-	A-	A-	0.18	0.05	3.50	1.07	4.76	1.16
177	812169	9	5717	0.49	0.22	0.49	0.11	0.18	0.00	0.00	0.28	0.07	0.28	-0.22	-0.26	A-	A+	A-	A-	0.65	0.05	6.29	1.12	6.30	1.18
178	812170	9	5717	0.29	0.20	0.16	0.35	0.29	0.00	0.00	0.18	-0.11	-0.19	0.06	0.18	A+	A-	A-	A-	1.69	0.05	4.45	1.12	6.11	1.27
179	810030	9	5717	0.48	0.14	0.24	0.14	0.48	0.00	0.00	0.49	-0.24	-0.24	-0.16	0.49	A-	A-	A+	A+	0.72	0.05	-6.55	0.88	-5.14	0.87

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
180	812548	9	5717	0.51	0.22	0.11	0.51	0.16	0.00	0.00	0.30	-0.12	-0.15	0.30	-0.14	A+	A-	A+	A+	0.54	0.05	6.68	1.13	4.88	1.14
181	812936	9	5717	0.57	0.11	0.57	0.14	0.18	0.00	0.00	0.40	-0.21	0.40	-0.22	-0.16	A-	A-	A+	A-	0.28	0.05	0.26	1.00	-0.66	0.98
182	809878	9	5717	0.51	0.09	0.21	0.18	0.51	0.00	0.00	0.39	-0.25	-0.16	-0.14	0.39	A+	A-	A-	A+	0.56	0.05	0.33	1.01	0.98	1.03
183	812929	9	5717	0.54	0.12	0.13	0.21	0.54	0.00	0.00	0.40	-0.13	-0.28	-0.15	0.40	A+	A-	A-	A+	0.41	0.05	-0.36	0.99	-0.14	1.00
184	813788	9	5717	0.39	0.30	0.18	0.39	0.12	0.00	0.00	0.35	-0.12	-0.08	0.35	-0.25	A+	A+	A-	A-	1.15	0.05	0.54	1.01	0.79	1.02
185	800387	9	5717	0.42	0.12	0.19	0.26	0.42	0.00	0.00	0.22	-0.15	-0.07	-0.07	0.22	A+	A-	A-	A+	1.00	0.05	8.71	1.18	7.64	1.23
186	800384	9	5717	0.52	0.12	0.52	0.14	0.21	0.00	0.00	0.40	-0.26	0.40	-0.25	-0.07	A+	A+	A-	A-	0.49	0.05	0.81	1.02	0.75	1.02
187	809057	9	5717	0.51	0.51	0.05	0.07	0.36	0.00	0.00	0.40	0.40	-0.27	-0.25	-0.16	A-	A+	A+	A-	0.54	0.05	-0.40	0.99	0.77	1.02
188	742291	9	5717	0.76	0.11	0.05	0.08	0.76	0.00	0.00	0.37	-0.12	-0.22	-0.26	0.37	A+	A-	A-	A+	-0.79	0.06	-1.56	0.96	-1.18	0.94
189	810557	9	5717	0.65	0.08	0.65	0.07	0.20	0.00	0.00	0.22	-0.25	0.22	-0.25	0.08	A+	A+	A+	A-	-0.14	0.05	6.32	1.14	7.37	1.30
190	810637	9	5717	0.43	0.37	0.43	0.11	0.09	0.00	0.00	0.36	-0.10	0.36	-0.31	-0.12	A+	A-	A+	A-	0.95	0.05	1.52	1.03	2.07	1.06
191	809288	9	5717	0.74	0.12	0.08	0.74	0.06	0.00	0.00	0.46	-0.23	-0.29	0.46	-0.21	A+	A-	A-	A+	-0.64	0.06	-4.70	0.88	-3.54	0.83
192	809286	9	5717	0.48	0.48	0.26	0.15	0.10	0.00	0.00	0.40	0.40	-0.19	-0.16	-0.19	A-	A-	A-	A-	0.71	0.05	-0.81	0.98	-0.46	0.99
193	812167	10	5682	0.41	0.21	0.24	0.41	0.15	0.00	0.00	0.36	-0.04	-0.17	0.36	-0.24	A-	A+	A-	A+	1.12	0.05	0.03	1.00	1.38	1.04
194	812171	10	5682	0.43	0.20	0.26	0.43	0.11	0.00	0.00	0.32	-0.15	-0.14	0.32	-0.12	A+	A-	A-	A-	1.02	0.05	2.67	1.06	3.57	1.10
195	740017	10	5682	0.44	0.44	0.13	0.27	0.16	0.00	0.00	0.28	0.28	-0.15	-0.05	-0.18	A-	A-	A-	A+	0.95	0.05	5.46	1.11	5.63	1.16
196	809869	10	5682	0.45	0.21	0.45	0.10	0.24	0.00	0.00	0.35	-0.13	0.35	-0.27	-0.09	A-	A-	A-	A+	0.89	0.05	1.11	1.02	1.48	1.04
197	809689	10	5682	0.55	0.14	0.55	0.16	0.15	0.00	0.00	0.34	-0.22	0.34	-0.14	-0.11	A-	A-	A-	A+	0.40	0.05	3.30	1.07	2.12	1.06
198	819086	10	5682	0.69	0.07	0.13	0.11	0.69	0.00	0.00	0.54	-0.25	-0.21	-0.37	0.54	A+	A-	A-	A-	-0.33	0.05	-7.08	0.84	-5.73	0.78
199	810610	10	5682	0.66	0.66	0.09	0.13	0.11	0.00	0.00	0.45	0.45	-0.26	-0.26	-0.16	A-	A+	A-	A+	-0.19	0.05	-1.50	0.97	-1.76	0.93
200	821171	10	5682	0.54	0.20	0.54	0.18	0.07	0.00	0.00	0.38	-0.06	0.38	-0.23	-0.29	A+	A-	A-	A+	0.46	0.05	0.92	1.02	0.73	1.02
201	800386	10	5682	0.33	0.13	0.25	0.28	0.33	0.00	0.00	0.22	-0.07	-0.13	-0.05	0.22	A-	A-	A+	A+	1.54	0.05	7.79	1.20	9.20	1.37
202	800385	10	5682	0.73	0.11	0.08	0.73	0.08	0.00	0.00	0.48	-0.23	-0.29	0.48	-0.22	A+	A-	A-	A+	-0.54	0.06	-4.10	0.90	-3.45	0.84
203	810292	10	5682	0.54	0.19	0.54	0.20	0.08	0.00	0.00	0.40	-0.18	0.40	-0.20	-0.17	A+	A-	A-	A-	0.47	0.05	1.23	1.02	0.69	1.02
204	810633	10	5682	0.45	0.18	0.30	0.07	0.45	0.00	0.00	0.46	-0.25	-0.12	-0.30	0.46	A+	A+	A-	A-	0.90	0.05	-3.64	0.93	-2.77	0.93
205	678899	10	5682	0.70	0.06	0.70	0.15	0.09	0.00	0.00	0.43	-0.22	0.43	-0.26	-0.16	A+	A-	A+	A+	-0.36	0.05	-2.96	0.93	-3.09	0.87
206	678969	10	5682	0.49	0.15	0.14	0.21	0.49	0.00	0.00	0.33	-0.19	-0.16	-0.10	0.33	A-	A-	A-	A-	0.69	0.05	3.61	1.07	4.29	1.12
207	734727	10	5682	0.16	0.16	0.55	0.15	0.14	0.00	0.00	0.11	0.11	-0.01	-0.06	-0.03	A-	A-	A+	A-	2.67	0.07	2.55	1.11	8.65	1.76
208	809195	10	5682	0.67	0.07	0.16	0.67	0.09	0.00	0.00	0.52	-0.29	-0.24	0.52	-0.28	A+	A-	A-	A+	-0.23	0.05	-7.48	0.84	-5.85	0.78
209	818023	11	5697	0.41	0.40	0.12	0.07	0.41	0.00	0.00	0.41	-0.14	-0.19	-0.26	0.41	A-	A-	A-	A-	1.14	0.05	-0.18	1.00	0.88	1.03

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
210	818022	11	5697	0.42	0.11	0.42	0.27	0.20	0.00	0.00	0.35	-0.16	0.35	-0.16	-0.13	A-	A-	A-	A+	1.05	0.05	2.80	1.06	4.15	1.12
211	740940	11	5697	0.65	0.65	0.13	0.12	0.09	0.00	0.00	0.48	0.48	-0.24	-0.28	-0.19	A+	A-	A+	A+	-0.09	0.05	-3.67	0.92	-4.12	0.86
212	816601	11	5697	0.49	0.14	0.49	0.24	0.14	0.00	0.00	0.23	-0.22	0.23	-0.05	-0.04	A-	A-	A-	A+	0.73	0.05	8.65	1.17	9.30	1.27
213	808539	11	5697	0.59	0.20	0.59	0.12	0.09	0.00	0.00	0.41	-0.27	0.41	-0.14	-0.17	A-	A-	A-	A+	0.24	0.05	-0.34	0.99	-0.82	0.97
214	810612	11	5697	0.74	0.74	0.06	0.13	0.07	0.00	0.00	0.49	0.49	-0.27	-0.24	-0.26	A+	A-	A-	A+	-0.58	0.06	-4.85	0.87	-3.75	0.82
215	821263	11	5697	0.47	0.09	0.47	0.19	0.25	0.00	0.00	0.34	-0.22	0.34	-0.14	-0.11	A-	A-	A-	A+	0.81	0.05	2.03	1.04	1.61	1.04
216	808350	11	5697	0.50	0.11	0.22	0.50	0.16	0.00	0.00	0.49	-0.22	-0.22	0.49	-0.23	A-	A-	A-	A+	0.67	0.05	-4.56	0.92	-3.51	0.91
217	800588	11	5697	0.76	0.11	0.76	0.08	0.05	0.00	0.00	0.49	-0.24	0.49	-0.28	-0.25	A+	A-	A-	A-	-0.69	0.06	-3.98	0.89	-4.75	0.77
218	800586	11	5697	0.73	0.10	0.73	0.12	0.06	0.00	0.00	0.46	-0.25	0.46	-0.23	-0.22	A-	A-	B-	A-	-0.51	0.06	-2.73	0.93	-1.62	0.92
219	741666	11	5697	0.57	0.57	0.16	0.16	0.10	0.00	0.00	0.42	0.42	-0.17	-0.23	-0.19	A-	A-	A+	A-	0.31	0.05	-1.64	0.97	-0.58	0.98
220	809204	11	5697	0.51	0.21	0.51	0.16	0.12	0.00	0.00	0.40	-0.12	0.40	-0.22	-0.21	A+	A-	A-	A-	0.60	0.05	-0.41	0.99	-0.86	0.98
221	734735	11	5697	0.36	0.36	0.10	0.42	0.12	0.00	0.00	0.23	0.23	-0.17	0.00	-0.17	A+	A-	A+	A+	1.39	0.05	6.66	1.15	6.93	1.24
222	739772	11	5697	0.55	0.12	0.11	0.22	0.55	0.00	0.00	0.44	-0.22	-0.18	-0.21	0.44	A+	A-	A-	A+	0.41	0.05	-1.89	0.97	-1.50	0.96
223	809060	11	5697	0.53	0.16	0.14	0.18	0.53	0.00	0.00	0.29	-0.13	-0.15	-0.11	0.29	A+	A-	A-	A+	0.55	0.05	6.14	1.12	5.17	1.15
224	809205	11	5697	0.47	0.47	0.19	0.22	0.12	0.00	0.00	0.22	0.22	-0.12	-0.06	-0.10	A-	A+	A+	A+	0.82	0.05	9.74	1.19	8.38	1.24
225	818020	12	5723	0.61	0.61	0.13	0.18	0.08	0.00	0.00	0.46	0.46	-0.29	-0.19	-0.19	A-	A-	A-	A-	0.08	0.05	-1.76	0.96	-2.41	0.92
226	818021	12	5723	0.26	0.06	0.26	0.24	0.44	0.00	0.00	0.08	-0.24	0.08	-0.12	0.15	A+	A+	A+	A-	1.91	0.06	9.56	1.29	9.90	1.67
227	809880	12	5723	0.78	0.09	0.08	0.05	0.78	0.00	0.00	0.46	-0.28	-0.19	-0.26	0.46	A+	A-	A-	A-	-0.87	0.06	-2.94	0.91	-3.38	0.81
228	809682	12	5723	0.66	0.66	0.18	0.07	0.09	0.00	0.00	0.44	0.44	-0.24	-0.25	-0.18	A+	A-	A+	A-	-0.19	0.05	-2.37	0.95	-3.28	0.88
229	810323	12	5723	0.69	0.09	0.12	0.10	0.69	0.00	0.00	0.49	-0.26	-0.27	-0.22	0.49	A+	B-	A-	A+	-0.34	0.05	-6.29	0.86	-4.70	0.81
230	809468	12	5723	0.64	0.17	0.11	0.64	0.09	0.00	0.00	0.38	-0.09	-0.18	0.38	-0.32	A-	A-	A-	A+	-0.05	0.05	0.44	1.01	-0.43	0.98
231	810618	12	5723	0.65	0.13	0.15	0.07	0.65	0.00	0.00	0.52	-0.19	-0.32	-0.28	0.52	A+	A-	A-	A+	-0.12	0.05	-6.47	0.87	-6.17	0.79
232	809885	12	5723	0.64	0.10	0.09	0.64	0.17	0.00	0.00	0.46	-0.26	-0.27	0.46	-0.16	A+	A-	A-	A+	-0.08	0.05	-2.01	0.96	-0.96	0.97
233	800592	12	5723	0.71	0.71	0.13	0.09	0.08	0.00	0.00	0.53	0.53	-0.33	-0.26	-0.21	A+	A-	A-	A-	-0.44	0.05	-7.11	0.84	-6.41	0.74
234	800591	12	5723	0.53	0.05	0.53	0.17	0.25	0.00	0.00	0.52	-0.26	0.52	-0.20	-0.29	A+	A-	A-	A-	0.50	0.05	-4.92	0.91	-4.17	0.89
235	816150	12	5723	0.48	0.18	0.48	0.20	0.14	0.00	0.00	0.42	-0.18	0.42	-0.19	-0.17	A+	A+	A+	A-	0.74	0.05	0.13	1.00	0.63	1.02
236	810529	12	5723	0.49	0.21	0.08	0.49	0.22	0.00	0.00	0.39	-0.17	-0.24	0.39	-0.13	A+	A-	A+	A-	0.69	0.05	1.49	1.03	1.54	1.04
237	741704	12	5723	0.57	0.57	0.14	0.10	0.19	0.00	0.00	0.37	0.37	-0.20	-0.26	-0.08	A+	A+	A-	A-	0.29	0.05	3.66	1.07	5.30	1.17
238	816424	12	5723	0.42	0.14	0.14	0.30	0.42	0.00	0.00	0.37	-0.15	-0.23	-0.10	0.37	A+	A-	A+	A-	1.04	0.05	1.65	1.03	1.73	1.05
239	811180	12	5723	0.24	0.24	0.07	0.28	0.41	0.00	0.00	0.22	0.22	-0.19	0.03	-0.12	A-	A-	A-	A+	2.04	0.06	2.84	1.09	7.95	1.46

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
240	816422	12	5723	0.56	0.21	0.13	0.56	0.10	0.00	0.00	0.37	-0.08	-0.26	0.37	-0.20	A+	A+	A-	A+	0.37	0.05	0.69	1.01	-0.10	1.00
241	799966	13	5635	0.53	0.53	0.17	0.05	0.24	0.00	0.00	0.16	0.16	-0.22	-0.21	0.13	A+	A+	A+	A-	0.50	0.05	9.90	1.26	9.90	1.35
242	799967	13	5635	0.52	0.25	0.08	0.52	0.15	0.00	0.00	0.26	-0.05	-0.16	0.26	-0.17	A-	A-	A-	A-	0.53	0.05	8.95	1.18	7.55	1.23
243	816144	13	5635	0.67	0.17	0.67	0.06	0.10	0.00	0.00	0.40	-0.31	0.40	-0.24	-0.04	A-	A-	A-	A-	-0.22	0.05	-1.42	0.97	2.25	1.09
244	809876	13	5635	0.36	0.36	0.16	0.09	0.39	0.00	0.00	0.37	0.37	-0.21	-0.30	-0.03	A+	A-	A-	A+	1.35	0.05	0.33	1.01	1.98	1.07
245	816142	13	5635	0.66	0.07	0.22	0.66	0.05	0.00	0.00	0.51	-0.23	-0.30	0.51	-0.23	A+	A-	A-	A+	-0.15	0.05	-5.62	0.88	-4.81	0.82
246	810616	13	5635	0.74	0.06	0.07	0.12	0.74	0.00	0.00	0.49	-0.28	-0.24	-0.26	0.49	B+	A-	A-	A+	-0.65	0.06	-5.23	0.86	-4.81	0.76
247	809879	13	5635	0.37	0.13	0.31	0.19	0.37	0.00	0.00	0.39	-0.20	-0.17	-0.11	0.39	A-	A-	A+	A+	1.30	0.05	-0.36	0.99	2.16	1.07
248	810022	13	5635	0.70	0.14	0.70	0.09	0.07	0.00	0.00	0.40	-0.18	0.40	-0.22	-0.22	A+	A-	A-	A+	-0.39	0.05	-0.25	0.99	1.16	1.05
249	800693	13	5635	0.66	0.13	0.06	0.66	0.16	0.00	0.00	0.35	-0.16	-0.24	0.35	-0.16	A-	A-	A+	A-	-0.18	0.05	3.88	1.09	2.74	1.11
250	800695	13	5635	0.54	0.16	0.18	0.54	0.11	0.00	0.00	0.42	-0.19	-0.19	0.42	-0.20	A+	A+	A-	A+	0.44	0.05	-0.99	0.98	-1.12	0.97
251	811175	13	5635	0.48	0.48	0.12	0.21	0.19	0.00	0.00	0.46	0.46	-0.25	-0.21	-0.15	A-	A-	A-	A-	0.75	0.05	-3.67	0.93	-2.82	0.92
252	810574	13	5635	0.70	0.08	0.70	0.15	0.07	0.00	0.00	0.43	-0.22	0.43	-0.17	-0.30	A+	A-	A-	A+	-0.39	0.05	-1.62	0.96	-1.29	0.94
253	810290	13	5635	0.53	0.15	0.53	0.15	0.18	0.00	0.00	0.23	-0.05	0.23	-0.11	-0.15	A-	A-	A-	A+	0.50	0.05	9.90	1.23	8.67	1.27
254	741743	13	5635	0.75	0.75	0.05	0.16	0.04	0.00	0.00	0.49	0.49	-0.25	-0.29	-0.25	A-	A-	B-	A-	-0.67	0.06	-3.19	0.91	-3.85	0.81
255	811182	13	5635	0.60	0.07	0.10	0.60	0.23	0.00	0.00	0.36	-0.30	-0.19	0.36	-0.10	A+	A-	A-	A+	0.14	0.05	1.84	1.04	1.58	1.05
256	813414	13	5635	0.67	0.14	0.67	0.10	0.08	0.00	0.00	0.43	-0.20	0.43	-0.24	-0.21	A+	A-	A-	A+	-0.25	0.05	-1.42	0.97	-2.30	0.91
257	799960	14	5630	0.13	0.16	0.29	0.13	0.43	0.00	0.00	-0.07	-0.18	0.13	-0.07	0.07	A+	A+	A+	A+	2.96	0.07	5.47	1.28	9.90	2.42
258	799963	14	5630	0.40	0.42	0.12	0.40	0.06	0.00	0.00	0.28	-0.08	-0.24	0.28	-0.07	A-	A+	A+	A-	1.15	0.05	4.76	1.10	5.84	1.18
259	741115	14	5630	0.54	0.15	0.54	0.10	0.20	0.00	0.00	0.40	-0.16	0.40	-0.22	-0.18	A+	A-	A-	A+	0.42	0.05	0.11	1.00	0.08	1.00
260	739698	14	5630	0.67	0.08	0.13	0.67	0.12	0.00	0.00	0.45	-0.31	-0.16	0.45	-0.22	A-	A-	A-	A+	-0.20	0.05	-2.39	0.95	-3.23	0.88
261	816652	14	5630	0.32	0.20	0.24	0.24	0.32	0.00	0.00	0.24	-0.13	0.09	-0.22	0.24	A-	A-	A+	A-	1.58	0.05	5.47	1.13	6.93	1.28
262	809886	14	5630	0.41	0.41	0.21	0.29	0.10	0.00	0.00	0.50	0.50	-0.09	-0.31	-0.22	A-	A-	A-	A-	1.12	0.05	-7.51	0.86	-6.09	0.83
263	714619	14	5630	0.64	0.07	0.17	0.12	0.64	0.00	0.00	0.48	-0.27	-0.24	-0.21	0.48	A+	A-	A-	A-	-0.08	0.05	-3.86	0.92	-4.22	0.85
264	714645	14	5630	0.81	0.81	0.09	0.06	0.04	0.00	0.00	0.46	0.46	-0.31	-0.22	-0.20	A+	A-	A-	A+	-1.09	0.06	-4.90	0.84	-5.58	0.67
265	800694	14	5630	0.51	0.04	0.22	0.23	0.51	0.00	0.00	0.48	-0.20	-0.30	-0.18	0.48	A-	A-	A-	A-	0.61	0.05	-4.54	0.92	-4.09	0.89
266	800689	14	5630	0.50	0.07	0.28	0.15	0.50	0.00	0.00	0.54	-0.27	-0.22	-0.29	0.54	A+	A-	A-	A-	0.65	0.05	-8.91	0.84	-7.73	0.81
267	742279	14	5630	0.71	0.71	0.13	0.07	0.09	0.00	0.00	0.39	0.39	-0.19	-0.20	-0.21	A-	A-	A-	A-	-0.43	0.05	-1.32	0.97	-0.45	0.98
268	816421	14	5630	0.26	0.14	0.26	0.18	0.41	0.00	0.00	0.06	-0.08	0.06	-0.12	0.10	A-	A+	A+	A+	1.92	0.06	9.15	1.27	9.90	1.80
269	811185	14	5630	0.81	0.10	0.05	0.04	0.81	0.00	0.00	0.41	-0.17	-0.24	-0.28	0.41	A-	B-	A-	A+	-1.07	0.06	-3.31	0.89	-2.69	0.83

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
270	810635	14	5630	0.55	0.17	0.55	0.17	0.11	0.00	0.00	0.34	-0.17	0.34	-0.22	-0.07	A-	A-	A-	A-	0.38	0.05	4.01	1.08	2.00	1.06
271	816423	14	5630	0.54	0.54	0.16	0.21	0.09	0.00	0.00	0.45	0.45	-0.16	-0.27	-0.18	A+	A-	A-	A-	0.45	0.05	-2.05	0.96	-2.05	0.94
272	813416	14	5630	0.54	0.18	0.11	0.17	0.54	0.00	0.00	0.45	-0.11	-0.19	-0.31	0.45	A+	A+	A+	A-	0.46	0.05	-1.86	0.96	-2.44	0.93
273	812315	15	5686	0.51	0.19	0.18	0.12	0.51	0.00	0.00	0.40	-0.26	-0.09	-0.19	0.40	A+	A-	A-	A-	0.61	0.05	-2.08	0.96	-1.33	0.97
274	812316	15	5686	0.39	0.11	0.25	0.39	0.24	0.00	0.00	0.34	-0.06	-0.24	0.34	-0.09	A-	A-	A-	A-	1.16	0.05	0.82	1.02	2.75	1.08
275	808358	15	5686	0.23	0.25	0.19	0.33	0.23	0.00	0.00	0.06	0.11	-0.19	0.01	0.06	A-	A+	A-	A-	2.11	0.06	7.77	1.26	9.90	1.61
276	736850	15	5686	0.63	0.63	0.14	0.14	0.08	0.00	0.00	0.44	0.44	-0.16	-0.26	-0.24	A-	A-	A-	A+	-0.02	0.05	-3.02	0.94	-3.09	0.90
277	808342	15	5686	0.42	0.42	0.18	0.14	0.26	0.00	0.00	0.31	0.31	-0.21	-0.24	0.02	A+	A+	A+	A-	1.04	0.05	4.28	1.08	5.74	1.16
278	810621	15	5686	0.48	0.09	0.22	0.20	0.48	0.00	0.00	0.39	-0.23	-0.17	-0.13	0.39	A-	A-	A-	A-	0.73	0.05	0.08	1.00	0.51	1.01
279	812546	15	5686	0.45	0.13	0.45	0.30	0.12	0.00	0.00	0.18	-0.16	0.18	0.06	-0.19	A-	A-	A-	A-	0.91	0.05	9.90	1.21	9.83	1.28
280	818998	15	5686	0.41	0.41	0.19	0.08	0.32	0.00	0.00	0.15	0.15	-0.11	-0.23	0.07	A+	A+	A-	A+	1.10	0.05	9.90	1.22	8.74	1.26
281	816272	15	5686	0.46	0.09	0.35	0.46	0.10	0.00	0.00	0.28	-0.26	0.00	0.28	-0.22	A+	A+	A-	A-	0.84	0.05	6.12	1.12	6.78	1.18
282	816270	15	5686	0.64	0.14	0.64	0.12	0.10	0.00	0.00	0.46	-0.25	0.46	-0.27	-0.15	A+	A-	A-	A+	-0.03	0.05	-2.32	0.95	-2.01	0.94
283	810289	15	5686	0.60	0.13	0.21	0.60	0.07	0.00	0.00	0.16	-0.14	0.04	0.16	-0.20	A-	A+	A-	A+	0.14	0.05	9.04	1.19	8.11	1.26
284	813651	15	5686	0.56	0.10	0.18	0.17	0.56	0.00	0.00	0.35	-0.29	-0.09	-0.14	0.35	A+	A-	A-	A-	0.35	0.05	2.30	1.04	1.55	1.04
285	809059	15	5686	0.72	0.08	0.72	0.06	0.13	0.00	0.00	0.53	-0.27	0.53	-0.31	-0.26	A+	A-	A-	A+	-0.52	0.06	-7.32	0.82	-7.29	0.71
286	741479	15	5686	0.37	0.29	0.18	0.37	0.16	0.00	0.00	0.36	-0.13	-0.18	0.36	-0.13	A+	A-	A-	A-	1.30	0.05	0.41	1.01	4.01	1.13
287	642876	15	5686	0.44	0.16	0.24	0.44	0.15	0.00	0.00	0.26	-0.13	-0.02	0.26	-0.20	A+	A-	A-	A-	0.92	0.05	7.20	1.14	7.01	1.19
288	800388	15	5686	0.40	0.17	0.26	0.17	0.40	0.00	0.00	0.35	-0.19	-0.09	-0.17	0.35	A-	A-	A+	A+	1.15	0.05	0.93	1.02	1.26	1.04
289	812313	16	5660	0.30	0.35	0.19	0.30	0.16	0.00	0.00	0.23	0.06	-0.25	0.23	-0.11	A+	A-	A-	A+	1.67	0.05	3.73	1.09	7.18	1.31
290	812314	16	5660	0.44	0.22	0.44	0.16	0.17	0.00	0.00	0.33	-0.15	0.33	-0.19	-0.08	A-	A-	A+	A-	0.94	0.05	3.14	1.06	3.83	1.11
291	702060	16	5660	0.67	0.05	0.67	0.12	0.16	0.00	0.00	0.46	-0.23	0.46	-0.27	-0.21	A-	A-	A-	A-	-0.20	0.05	-2.78	0.94	-2.83	0.89
292	810617	16	5660	0.37	0.17	0.32	0.37	0.14	0.00	0.00	0.30	-0.24	-0.08	0.30	-0.04	A-	A-	A-	A-	1.32	0.05	4.08	1.09	6.05	1.21
293	714634	16	5660	0.52	0.52	0.14	0.19	0.16	0.00	0.00	0.41	0.41	-0.20	-0.24	-0.12	A+	A-	A-	A+	0.56	0.05	-1.87	0.96	-2.21	0.94
294	809698	16	5660	0.36	0.14	0.36	0.28	0.22	0.00	0.00	0.25	-0.09	0.25	-0.04	-0.17	A-	A-	A-	A+	1.35	0.05	5.94	1.13	7.89	1.28
295	809868	16	5660	0.76	0.12	0.06	0.06	0.76	0.00	0.00	0.60	-0.36	-0.29	-0.29	0.60	A-	A-	A-	A+	-0.74	0.06	-9.90	0.74	-9.84	0.56
296	809684	16	5660	0.58	0.08	0.58	0.14	0.19	0.00	0.00	0.34	-0.18	0.34	-0.23	-0.09	A+	A+	A-	A+	0.25	0.05	1.60	1.03	0.67	1.02
297	816268	16	5660	0.57	0.27	0.11	0.57	0.05	0.00	0.00	0.25	-0.07	-0.17	0.25	-0.18	A-	A-	A-	A+	0.31	0.05	7.07	1.14	6.96	1.22
298	816269	16	5660	0.44	0.11	0.44	0.39	0.06	0.00	0.00	0.30	-0.21	0.30	-0.08	-0.19	A+	A-	A-	A+	0.92	0.05	5.53	1.11	5.63	1.16
299	808873	16	5660	0.39	0.39	0.21	0.17	0.23	0.00	0.00	0.40	0.40	-0.23	-0.22	-0.03	A+	A+	A-	A+	1.19	0.05	-2.23	0.95	0.60	1.02

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
300	810291	16	5660	0.57	0.57	0.10	0.15	0.18	0.00	0.00	0.51	0.51	-0.30	-0.23	-0.20	A-	A-	A-	A-	0.33	0.05	-7.18	0.87	-6.48	0.82
301	811178	16	5660	0.49	0.05	0.19	0.49	0.27	0.00	0.00	0.36	-0.23	-0.23	0.36	-0.09	A-	A-	A-	A+	0.69	0.05	2.43	1.05	3.26	1.09
302	810638	16	5660	0.47	0.09	0.47	0.24	0.19	0.00	0.00	0.25	-0.14	0.25	-0.05	-0.16	A-	A+	A-	A+	0.79	0.05	8.91	1.18	8.24	1.24
303	808880	16	5660	0.80	0.80	0.11	0.06	0.04	0.00	0.00	0.37	0.37	-0.24	-0.16	-0.20	A+	A-	A-	A-	-0.98	0.06	1.21	1.04	1.76	1.11
304	816420	16	5660	0.37	0.26	0.37	0.18	0.19	0.00	0.00	0.18	0.21	0.18	-0.27	-0.19	A-	A-	A+	A+	1.30	0.05	9.90	1.25	9.90	1.37
305	818070	17	5714	0.25	0.46	0.17	0.25	0.12	0.00	0.00	0.02	0.17	-0.03	0.02	-0.25	A+	A-	A+	A+	1.93	0.06	9.90	1.31	9.90	1.77
306	818069	17	5714	0.55	0.09	0.19	0.55	0.17	0.00	0.00	0.33	-0.21	-0.17	0.33	-0.09	A+	A-	A+	A+	0.38	0.05	3.43	1.06	4.56	1.13
307	721614	17	5714	0.77	0.77	0.06	0.08	0.09	0.00	0.00	0.44	0.44	-0.22	-0.30	-0.18	A-	C-	A-	A+	-0.80	0.06	-3.46	0.90	-4.38	0.78
308	821184	17	5714	0.50	0.50	0.09	0.20	0.20	0.00	0.00	0.34	0.34	-0.27	-0.22	0.01	A-	A+	A-	A-	0.61	0.05	2.12	1.04	1.72	1.04
309	813193	17	5714	0.28	0.30	0.31	0.11	0.28	0.00	0.00	0.30	0.11	-0.22	-0.28	0.30	A-	A+	A+	A-	1.76	0.05	1.07	1.03	2.46	1.10
310	809873	17	5714	0.76	0.10	0.07	0.76	0.08	0.00	0.00	0.45	-0.19	-0.25	0.45	-0.27	A+	A-	A-	A+	-0.77	0.06	-2.56	0.93	-2.93	0.85
311	714654	17	5714	0.61	0.29	0.05	0.61	0.04	0.00	0.00	0.36	-0.16	-0.25	0.36	-0.22	A-	A-	A-	A+	0.07	0.05	2.58	1.05	2.97	1.09
312	739664	17	5714	0.41	0.13	0.14	0.31	0.41	0.00	0.00	0.40	-0.23	-0.18	-0.12	0.40	A-	A+	A-	A-	1.05	0.05	-2.93	0.94	-1.63	0.96
313	821019	17	5714	0.40	0.40	0.24	0.19	0.17	0.00	0.00	0.34	0.34	-0.07	-0.23	-0.12	A-	A+	A+	A-	1.10	0.05	0.42	1.01	1.66	1.05
314	821020	17	5714	0.29	0.19	0.30	0.29	0.22	0.00	0.00	0.19	-0.13	-0.05	0.19	-0.02	A-	A+	A+	A+	1.68	0.05	5.83	1.15	9.06	1.40
315	739779	17	5714	0.31	0.31	0.36	0.11	0.21	0.00	0.00	0.19	0.19	-0.16	-0.15	0.08	A+	A+	A+	A-	1.57	0.05	5.11	1.13	6.84	1.27
316	742283	17	5714	0.53	0.32	0.04	0.11	0.53	0.00	0.00	0.22	0.03	-0.26	-0.22	0.22	A+	A-	A-	A-	0.46	0.05	8.93	1.17	7.28	1.20
317	808353	17	5714	0.62	0.15	0.62	0.17	0.06	0.00	0.00	0.45	-0.21	0.45	-0.24	-0.21	A-	A+	A+	A+	0.04	0.05	-5.03	0.90	-5.21	0.85
318	810526	17	5714	0.28	0.20	0.18	0.33	0.28	0.00	0.00	0.14	-0.11	-0.18	0.10	0.14	A+	A-	A-	A-	1.74	0.05	6.21	1.17	8.22	1.37
319	811171	17	5714	0.63	0.63	0.08	0.07	0.22	0.00	0.00	0.56	0.56	-0.28	-0.31	-0.28	A+	A+	A-	A+	-0.01	0.05	-8.88	0.83	-8.21	0.76
320	809287	17	5714	0.42	0.42	0.42	0.10	0.06	0.00	0.00	0.29	0.29	0.03	-0.30	-0.29	A-	A-	A-	A-	0.99	0.05	2.88	1.06	3.67	1.10
321	818065	18	5693	0.32	0.33	0.16	0.32	0.19	0.00	0.00	0.24	-0.09	-0.17	0.24	0.00	A-	A-	A-	A-	1.54	0.05	6.36	1.16	9.55	1.39
322	818067	18	5693	0.39	0.39	0.29	0.07	0.25	0.00	0.00	0.17	0.17	-0.01	-0.29	-0.01	A-	A-	A+	A+	1.16	0.05	9.90	1.24	9.90	1.35
323	809156	18	5693	0.60	0.15	0.09	0.14	0.60	0.00	0.00	0.48	-0.26	-0.27	-0.16	0.48	A+	A-	A-	A-	0.11	0.05	-4.46	0.91	-4.44	0.87
324	809881	18	5693	0.42	0.20	0.19	0.42	0.19	0.00	0.00	0.31	-0.16	-0.10	0.31	-0.13	A-	A+	A-	A-	1.01	0.05	4.13	1.08	5.32	1.15
325	809883	18	5693	0.41	0.16	0.15	0.28	0.41	0.00	0.00	0.36	-0.06	-0.19	-0.20	0.36	A-	A-	A-	A+	1.07	0.05	2.35	1.05	3.94	1.12
326	809693	18	5693	0.54	0.13	0.17	0.54	0.16	0.00	0.00	0.47	-0.29	-0.27	0.47	-0.10	A-	A-	A-	A-	0.41	0.05	-3.12	0.94	-2.13	0.94
327	809882	18	5693	0.60	0.60	0.08	0.11	0.21	0.00	0.00	0.43	0.43	-0.17	-0.28	-0.19	A-	A+	A+	A-	0.14	0.05	-1.72	0.97	-2.67	0.92
328	808545	18	5693	0.60	0.13	0.11	0.15	0.60	0.00	0.00	0.49	-0.29	-0.22	-0.19	0.49	A+	A-	A-	A-	0.11	0.05	-5.32	0.90	-5.16	0.85
329	821024	18	5693	0.20	0.20	0.15	0.19	0.46	0.00	0.00	-0.09	-0.09	-0.27	-0.13	0.37	A+	A+	A+	A-	2.32	0.06	9.05	1.35	9.90	1.94

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
330	821025	18	5693	0.61	0.12	0.61	0.12	0.15	0.00	0.00	0.45	-0.28	0.45	-0.21	-0.16	A-	A-	A-	A-	0.10	0.05	-1.46	0.97	-1.84	0.94
331	740989	18	5693	0.41	0.41	0.16	0.25	0.18	0.00	0.00	0.27	0.27	-0.18	-0.15	0.00	A+	A-	A+	A+	1.07	0.05	7.31	1.15	7.95	1.24
332	809192	18	5693	0.41	0.22	0.19	0.19	0.41	0.00	0.00	0.29	-0.15	-0.10	-0.10	0.29	A+	A-	A+	A-	1.09	0.05	4.81	1.10	5.28	1.16
333	809290	18	5693	0.73	0.03	0.73	0.03	0.21	0.00	0.00	0.37	-0.24	0.37	-0.23	-0.20	A-	A-	A+	A-	-0.60	0.06	-0.50	0.99	-0.26	0.99
334	810288	18	5693	0.48	0.48	0.17	0.19	0.16	0.00	0.00	0.44	0.44	-0.22	-0.15	-0.20	A+	A-	A+	A-	0.70	0.05	-1.84	0.97	-1.57	0.96
335	810530	18	5693	0.56	0.06	0.10	0.28	0.56	0.00	0.00	0.33	-0.28	-0.14	-0.12	0.33	A-	A+	A+	A+	0.31	0.05	4.12	1.08	3.54	1.10
336	813411	18	5693	0.55	0.24	0.11	0.09	0.55	0.00	0.00	0.48	-0.12	-0.30	-0.31	0.48	A+	A-	A-	A+	0.37	0.05	-5.37	0.90	-4.09	0.89
337	819332	19	5696	0.26	0.17	0.43	0.14	0.26	0.00	0.00	0.22	-0.02	-0.03	-0.22	0.22	A+	A-	A-	A-	1.94	0.06	3.64	1.11	7.33	1.40
338	819336	19	5696	0.25	0.30	0.25	0.29	0.16	0.00	0.00	0.25	0.06	0.25	-0.10	-0.23	A-	A-	A-	A+	2.00	0.06	2.07	1.06	8.48	1.49
339	713982	19	5696	0.62	0.62	0.08	0.20	0.10	0.00	0.00	0.38	0.38	-0.25	-0.10	-0.24	A+	A-	A-	A+	0.04	0.05	-0.19	1.00	-0.34	0.99
340	810032	19	5696	0.35	0.08	0.39	0.18	0.35	0.00	0.00	0.36	-0.20	-0.09	-0.19	0.36	A-	A-	A-	A-	1.41	0.05	-0.20	1.00	2.37	1.09
341	812549	19	5696	0.63	0.10	0.14	0.63	0.13	0.00	0.00	0.41	-0.27	-0.21	0.41	-0.14	A+	A-	A-	A-	-0.02	0.05	0.14	1.00	-0.82	0.97
342	812550	19	5696	0.42	0.42	0.09	0.17	0.33	0.00	0.00	0.43	0.43	-0.28	-0.19	-0.12	A+	A+	A-	A-	1.05	0.05	-2.56	0.95	-0.57	0.98
343	816147	19	5696	0.20	0.12	0.20	0.40	0.27	0.00	0.00	0.29	-0.23	0.29	-0.05	-0.04	A+	A-	A-	A-	2.32	0.06	0.04	1.00	6.51	1.45
344	741598	19	5696	0.64	0.16	0.10	0.10	0.64	0.00	0.00	0.49	-0.25	-0.23	-0.24	0.49	A+	B-	B-	A+	-0.09	0.05	-5.42	0.89	-5.33	0.82
345	816746	19	5696	0.36	0.36	0.21	0.13	0.30	0.00	0.00	0.24	0.24	-0.24	-0.17	0.09	A+	A+	A-	A-	1.33	0.05	6.75	1.16	7.75	1.29
346	816748	19	5696	0.38	0.21	0.16	0.25	0.38	0.00	0.00	0.37	-0.11	-0.21	-0.13	0.37	A+	A-	A+	A+	1.23	0.05	1.54	1.03	3.13	1.10
347	809062	19	5696	0.68	0.13	0.07	0.12	0.68	0.00	0.00	0.52	-0.24	-0.30	-0.25	0.52	A+	B-	A-	A+	-0.29	0.05	-7.32	0.84	-6.69	0.75
348	809198	19	5696	0.49	0.22	0.14	0.49	0.15	0.00	0.00	0.17	0.06	-0.23	0.17	-0.07	A+	A+	A-	A+	0.69	0.05	9.90	1.27	9.90	1.41
349	816425	19	5696	0.59	0.17	0.59	0.11	0.13	0.00	0.00	0.39	-0.05	0.39	-0.25	-0.27	A+	A-	A+	A-	0.20	0.05	0.89	1.02	2.15	1.07
350	741038	19	5696	0.76	0.05	0.76	0.05	0.15	0.00	0.00	0.51	-0.22	0.51	-0.25	-0.33	A-	A-	A-	A+	-0.73	0.06	-5.90	0.84	-6.29	0.70
351	810287	19	5696	0.51	0.51	0.16	0.23	0.11	0.00	0.00	0.48	0.48	-0.27	-0.19	-0.21	A+	A-	A+	A-	0.58	0.05	-5.23	0.90	-4.91	0.87
352	809201	19	5696	0.79	0.05	0.79	0.06	0.10	0.00	0.00	0.55	-0.28	0.55	-0.30	-0.30	A+	B-	A-	A+	-0.94	0.06	-8.33	0.76	-7.48	0.61
353	819337	20	5743	0.31	0.33	0.31	0.08	0.28	0.00	0.00	0.04	0.21	0.04	-0.22	-0.12	A-	A-	A+	A+	1.60	0.05	9.90	1.32	9.90	1.51
354	819334	20	5743	0.59	0.12	0.59	0.21	0.08	0.00	0.00	0.31	-0.12	0.31	-0.10	-0.26	A+	A+	A+	A-	0.18	0.05	5.16	1.10	4.87	1.15
355	809469	20	5743	0.52	0.52	0.41	0.03	0.03	0.00	0.00	0.10	0.10	0.04	-0.20	-0.21	A-	A-	A-	A-	0.49	0.05	9.90	1.26	9.90	1.34
356	808346	20	5743	0.46	0.20	0.24	0.11	0.46	0.00	0.00	0.47	-0.23	-0.20	-0.19	0.47	A-	A+	A-	A-	0.82	0.05	-5.42	0.90	-3.80	0.91
357	714633	20	5743	0.36	0.19	0.36	0.23	0.22	0.00	0.00	0.29	-0.01	0.29	-0.22	-0.11	A-	A+	A-	A+	1.29	0.05	3.09	1.07	5.30	1.17
358	809463	20	5743	0.31	0.16	0.30	0.31	0.23	0.00	0.00	0.23	-0.15	-0.05	0.23	-0.06	A-	A-	A+	A-	1.59	0.05	2.41	1.06	4.62	1.18
359	808348	20	5743	0.77	0.77	0.07	0.06	0.10	0.00	0.00	0.47	0.47	-0.29	-0.26	-0.20	A+	A-	A-	A+	-0.83	0.06	-5.59	0.84	-5.09	0.75

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
360	816146	20	5743	0.58	0.13	0.17	0.58	0.12	0.00	0.00	0.35	-0.16	-0.09	0.35	-0.25	A-	A+	A+	A+	0.20	0.05	0.89	1.02	1.02	1.03
361	816745	20	5743	0.44	0.24	0.12	0.20	0.44	0.00	0.00	0.39	-0.11	-0.27	-0.14	0.39	A-	A-	A-	A-	0.88	0.05	-0.76	0.99	-0.49	0.99
362	816744	20	5743	0.56	0.23	0.56	0.12	0.08	0.00	0.00	0.34	-0.12	0.34	-0.13	-0.25	A+	A+	A-	A+	0.32	0.05	2.88	1.05	3.65	1.10
363	811176	20	5743	0.45	0.08	0.27	0.45	0.19	0.00	0.00	0.17	-0.16	-0.01	0.17	-0.09	A+	A-	A-	A-	0.83	0.05	9.14	1.18	9.60	1.26
364	810640	20	5743	0.66	0.10	0.16	0.66	0.08	0.00	0.00	0.49	-0.17	-0.38	0.49	-0.16	A+	A-	A-	A+	-0.20	0.05	-6.02	0.88	-5.55	0.81
365	809202	20	5743	0.37	0.37	0.17	0.34	0.12	0.00	0.00	0.12	0.12	-0.14	0.08	-0.12	A+	A-	A-	A+	1.23	0.05	9.90	1.25	9.90	1.39
366	811183	20	5743	0.42	0.42	0.14	0.17	0.27	0.00	0.00	0.22	0.22	-0.03	-0.27	0.01	A-	A-	A-	A+	0.98	0.05	7.59	1.15	7.36	1.21
367	816418	20	5743	0.36	0.23	0.14	0.36	0.26	0.00	0.00	0.44	-0.20	-0.29	0.44	-0.06	A+	A-	A-	A-	1.29	0.05	-4.05	0.92	-1.28	0.96
368	702732	20	5743	0.70	0.08	0.70	0.04	0.17	0.00	0.00	0.43	-0.28	0.43	-0.27	-0.17	A+	A-	A-	A+	-0.39	0.05	-3.20	0.93	-2.62	0.90

Table J–7. Literature Multiple-Choice Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	734612	0	130328	0.61	0.19	0.10	0.10	0.61	0.00	0.00	0.44	-0.24	-0.20	-0.19	0.44					0.44	0.01	-5.15	0.97	-3.30	0.97
2	734591	0	130328	0.57	0.12	0.57	0.12	0.18	0.00	0.00	0.36	-0.15	0.36	-0.18	-0.18					0.61	0.01	9.90	1.06	9.54	1.07
3	734573	0	130328	0.55	0.20	0.16	0.55	0.08	0.00	0.00	0.34	-0.10	-0.19	0.34	-0.21					0.72	0.01	9.90	1.08	9.90	1.11
4	734569	0	130328	0.60	0.14	0.09	0.16	0.60	0.00	0.00	0.43	-0.20	-0.24	-0.19	0.43					0.66	0.01	-4.51	0.98	-4.38	0.97
5	734576	0	130328	0.71	0.06	0.13	0.10	0.71	0.00	0.00	0.49	-0.25	-0.24	-0.26	0.49					-0.06	0.01	-9.90	0.87	-9.90	0.80
6	734614	0	130328	0.53	0.27	0.15	0.53	0.04	0.00	0.00	0.25	-0.01	-0.19	0.25	-0.24					0.91	0.01	9.90	1.19	9.90	1.30
7	734615	0	130328	0.63	0.15	0.63	0.14	0.07	0.00	0.00	0.33	-0.18	0.33	-0.11	-0.20					0.29	0.01	9.90	1.11	9.90	1.18
8	734611	0	130328	0.27	0.18	0.27	0.18	0.36	0.00	0.00	0.04	-0.13	0.04	-0.06	0.12					1.38	0.01	9.90	1.29	9.90	1.49
9	734575	0	130328	0.50	0.50	0.14	0.22	0.14	0.00	0.00	0.30	0.30	-0.18	-0.09	-0.14					1.28	0.01	9.90	1.15	9.90	1.26
10	640841	0	130328	0.75	0.12	0.08	0.75	0.05	0.00	0.00	0.45	-0.20	-0.30	0.45	-0.21					-0.56	0.01	-5.27	0.96	-3.44	0.95
11	640840	0	130328	0.64	0.64	0.08	0.14	0.14	0.00	0.00	0.37	0.37	-0.27	-0.12	-0.16					0.33	0.01	9.12	1.05	9.47	1.08
12	640857	0	130328	0.80	0.08	0.05	0.07	0.80	0.00	0.00	0.43	-0.19	-0.24	-0.26	0.43					-0.91	0.01	-1.33	0.99	-2.06	0.97
13	640839	0	130328	0.76	0.76	0.08	0.10	0.06	0.00	0.00	0.43	0.43	-0.24	-0.20	-0.23					-1.03	0.02	9.90	1.23	9.90	1.27
14	640858	0	130328	0.61	0.21	0.61	0.09	0.09	0.00	0.00	0.40	-0.21	0.40	-0.20	-0.15					0.39	0.01	8.72	1.05	8.78	1.07
15	640842	0	130328	0.66	0.09	0.14	0.11	0.66	0.00	0.00	0.50	-0.25	-0.22	-0.26	0.50					-0.19	0.01	-0.73	1.00	-6.00	0.94
16	640859	0	130328	0.69	0.09	0.14	0.69	0.08	0.00	0.00	0.46	-0.19	-0.27	0.46	-0.22					-0.09	0.01	-9.90	0.94	-9.90	0.87
17	640836	0	130328	0.66	0.09	0.07	0.66	0.18	0.01	0.00	0.36	-0.22	-0.28	0.36	-0.08					0.38	0.01	9.90	1.05	9.90	1.11
18	683335	0	130328	0.84	0.84	0.09	0.04	0.03	0.00	0.00	0.52	0.52	-0.34	-0.26	-0.24					-1.07	0.02	-9.90	0.73	-9.90	0.52
19	683343	0	130328	0.51	0.51	0.13	0.25	0.11	0.00	0.00	0.27	0.27	-0.10	-0.15	-0.12					0.72	0.01	9.90	1.19	9.90	1.25
20	683461	0	130328	0.61	0.61	0.10	0.23	0.06	0.00	0.00	0.34	0.34	-0.23	-0.17	-0.12					0.35	0.01	9.90	1.08	9.90	1.08
21	683338	0	130328	0.63	0.20	0.05	0.12	0.63	0.00	0.00	0.40	-0.17	-0.26	-0.21	0.40					0.25	0.01	5.41	1.03	1.18	1.01
22	683342	0	130328	0.71	0.10	0.71	0.11	0.08	0.00	0.00	0.49	-0.25	0.49	-0.28	-0.23					-0.22	0.01	-9.90	0.89	-9.90	0.77
23	683336	0	130328	0.81	0.03	0.81	0.12	0.04	0.00	0.00	0.46	-0.24	0.46	-0.31	-0.19					-0.82	0.01	-9.90	0.83	-9.90	0.75
24	683340	0	130328	0.56	0.20	0.56	0.17	0.07	0.00	0.00	0.30	-0.07	0.30	-0.13	-0.28					1.07	0.01	9.90	1.14	9.90	1.21
25	683460	0	130328	0.45	0.45	0.21	0.07	0.26	0.00	0.00	0.30	0.30	-0.14	-0.23	-0.06					1.34	0.01	9.90	1.10	9.90	1.22
26	742875	0	130328	0.68	0.09	0.15	0.08	0.68	0.00	0.00	0.40	-0.20	-0.12	-0.30	0.40					0.02	0.01	4.35	1.03	1.22	1.01
27	742869	0	130328	0.61	0.61	0.05	0.09	0.24	0.00	0.00	0.42	0.42	-0.30	-0.31	-0.10					0.38	0.01	3.73	1.02	5.78	1.05
28	742873	0	130328	0.88	0.04	0.88	0.04	0.04	0.00	0.00	0.49	-0.27	0.49	-0.29	-0.23					-1.90	0.02	-5.57	0.92	-9.90	0.61
29	742874	0	130328	0.44	0.21	0.30	0.44	0.04	0.00	0.00	0.35	-0.23	-0.05	0.35	-0.27					1.28	0.01	9.90	1.04	9.90	1.14

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	742878	0	130328	0.37	0.14	0.32	0.37	0.15	0.00	0.00	0.24	-0.24	0.07	0.24	-0.16					1.37	0.01	9.90	1.13	9.90	1.28
31	742880	0	130328	0.68	0.08	0.11	0.68	0.12	0.00	0.00	0.43	-0.26	-0.27	0.43	-0.12					0.02	0.01	-4.51	0.97	-5.60	0.95
32	742870	0	130328	0.32	0.32	0.36	0.25	0.06	0.00	0.00	0.19	0.19	0.13	-0.19	-0.25					1.63	0.01	9.90	1.12	9.90	1.30
33	742872	0	130328	0.72	0.13	0.72	0.08	0.06	0.00	0.00	0.42	-0.16	0.42	-0.26	-0.24					-0.48	0.01	9.90	1.09	7.79	1.10
34	742879	0	130328	0.74	0.11	0.74	0.09	0.06	0.00	0.00	0.52	-0.25	0.52	-0.29	-0.26					-0.56	0.01	-9.90	0.89	-9.90	0.78
35	820933	1	5554	0.85	0.06	0.85	0.07	0.02	0.00	0.00	0.43	-0.31	0.43	-0.21	-0.19	A-	A-	B-	A-	-1.06	0.07	-0.30	0.99	-1.66	0.88
36	820938	1	5554	0.46	0.08	0.13	0.34	0.46	0.00	0.00	0.30	-0.22	-0.05	-0.15	0.30	A-	A-	A-	A-	1.27	0.05	6.16	1.12	7.45	1.24
37	820944	1	5554	0.77	0.77	0.05	0.12	0.07	0.00	0.00	0.48	0.48	-0.23	-0.28	-0.24	A-	A-	A-	A-	-0.43	0.06	-2.76	0.92	-3.97	0.80
38	820937	1	5554	0.69	0.05	0.69	0.06	0.19	0.00	0.00	0.33	-0.24	0.33	-0.17	-0.13	A-	A+	A-	A-	0.04	0.05	0.87	1.02	0.94	1.04
39	820935	1	5554	0.75	0.07	0.75	0.14	0.04	0.00	0.00	0.42	-0.21	0.42	-0.23	-0.24	A+	A+	A-	A-	-0.29	0.06	-1.21	0.96	-1.45	0.93
40	820945	1	5554	0.64	0.64	0.15	0.18	0.03	0.00	0.00	0.21	0.21	-0.18	0.02	-0.25	A-	A-	A-	A-	0.35	0.05	9.32	1.23	9.35	1.35
41	812980	1	5554	0.70	0.21	0.06	0.70	0.03	0.00	0.00	0.45	-0.27	-0.23	0.45	-0.23	A-	A-	A-	A-	0.03	0.05	-0.48	0.99	-2.31	0.91
42	812982	1	5554	0.45	0.45	0.08	0.30	0.17	0.00	0.00	0.15	0.15	-0.22	0.09	-0.14	A-	A-	A+	A+	1.32	0.05	9.90	1.25	9.90	1.44
43	812985	1	5554	0.76	0.76	0.10	0.11	0.03	0.00	0.00	0.49	0.49	-0.28	-0.25	-0.26	A-	B-	B-	A+	-0.36	0.06	-3.54	0.90	-3.97	0.81
44	812986	1	5554	0.64	0.10	0.17	0.64	0.09	0.00	0.00	0.38	-0.09	-0.30	0.38	-0.15	A+	A+	A-	A+	0.33	0.05	2.32	1.05	1.53	1.05
45	812981	1	5554	0.52	0.11	0.24	0.13	0.52	0.00	0.00	0.21	-0.30	0.01	-0.04	0.21	B+	A+	A-	A-	0.96	0.05	9.90	1.21	8.86	1.27
46	812988	1	5554	0.78	0.05	0.03	0.78	0.13	0.00	0.00	0.35	-0.09	-0.17	0.35	-0.27	A-	A-	A-	A+	-0.52	0.06	0.33	1.01	0.64	1.04
47	820934	2	5563	0.60	0.17	0.60	0.17	0.06	0.00	0.00	0.29	-0.17	0.29	-0.11	-0.13	A-	A+	A-	A+	0.60	0.05	5.85	1.13	4.49	1.15
48	820940	2	5563	0.75	0.08	0.08	0.08	0.75	0.00	0.00	0.43	-0.20	-0.20	-0.27	0.43	A-	A-	A-	A-	-0.28	0.06	-2.23	0.93	-2.24	0.89
49	820942	2	5563	0.69	0.11	0.10	0.11	0.69	0.00	0.00	0.53	-0.26	-0.30	-0.25	0.53	A-	A+	A-	A-	0.10	0.05	-4.47	0.89	-4.74	0.82
50	820936	2	5563	0.82	0.05	0.10	0.82	0.02	0.00	0.00	0.55	-0.33	-0.31	0.55	-0.25	A-	A-	A-	A-	-0.73	0.06	-4.81	0.83	-6.00	0.65
51	820941	2	5563	0.62	0.19	0.12	0.62	0.07	0.00	0.00	0.36	-0.08	-0.25	0.36	-0.23	A+	A+	A+	A-	0.50	0.05	2.93	1.07	2.67	1.09
52	820939	2	5563	0.76	0.76	0.06	0.13	0.05	0.00	0.00	0.38	0.38	-0.25	-0.13	-0.26	A+	A-	A-	A-	-0.30	0.06	-0.02	1.00	-0.23	0.99
53	812977	2	5563	0.44	0.27	0.05	0.44	0.23	0.00	0.00	0.35	-0.13	-0.24	0.35	-0.14	B-	A-	A-	A+	1.38	0.05	2.46	1.05	4.41	1.14
54	812978	2	5563	0.34	0.32	0.34	0.09	0.24	0.00	0.00	0.12	0.12	0.12	-0.29	-0.06	A-	A-	A-	A+	1.90	0.05	9.90	1.23	9.90	1.70
55	812987	2	5563	0.50	0.03	0.05	0.42	0.50	0.00	0.00	0.22	-0.24	-0.28	0.00	0.22	A+	A+	A-	A+	1.11	0.05	9.90	1.22	9.90	1.31
56	812979	2	5563	0.14	0.04	0.14	0.39	0.43	0.00	0.00	0.01	-0.31	0.01	0.11	0.02	A-	A+	A-	A-	3.20	0.07	2.61	1.12	9.90	2.35
57	812984	2	5563	0.64	0.10	0.11	0.15	0.64	0.00	0.00	0.47	-0.27	-0.15	-0.27	0.47	A+	A-	A-	A+	0.36	0.05	-2.54	0.94	-1.66	0.94
58	812983	2	5563	0.22	0.22	0.39	0.22	0.17	0.00	0.00	0.04	0.04	0.03	-0.04	-0.04	A-	A+	A-	A-	2.58	0.06	5.97	1.19	9.90	1.92
59	798613	3	5537	0.61	0.61	0.13	0.09	0.17	0.00	0.00	0.46	0.46	-0.31	-0.14	-0.20	A-	A-	B-	A+	0.52	0.05	-1.60	0.96	-0.42	0.99

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
60	798602	3	5537	0.67	0.15	0.08	0.67	0.10	0.00	0.00	0.44	-0.21	-0.22	0.44	-0.23	A+	B-	B-	A+	0.19	0.05	-1.45	0.96	-1.54	0.94
61	798612	3	5537	0.78	0.11	0.07	0.04	0.78	0.00	0.00	0.45	-0.21	-0.28	-0.27	0.45	A+	A-	A-	A+	-0.51	0.06	-1.48	0.95	-1.19	0.93
62	798604	3	5537	0.85	0.85	0.02	0.08	0.05	0.00	0.00	0.44	0.44	-0.25	-0.26	-0.21	A+	A-	A-	A+	-1.07	0.07	-3.16	0.87	-4.04	0.71
63	798607	3	5537	0.69	0.11	0.10	0.10	0.69	0.00	0.00	0.48	-0.25	-0.23	-0.25	0.48	A+	A-	A-	A+	0.08	0.06	-2.43	0.94	-2.88	0.88
64	798603	3	5537	0.42	0.05	0.42	0.40	0.14	0.00	0.00	0.21	-0.19	0.21	-0.06	-0.09	A+	A-	A-	A+	1.49	0.05	9.90	1.27	9.90	1.54
65	826272	3	5537	0.81	0.09	0.06	0.04	0.81	0.00	0.00	0.51	-0.23	-0.33	-0.26	0.51	A-	B-	B-	A+	-0.73	0.06	-4.43	0.85	-3.69	0.77
66	826277	3	5537	0.87	0.04	0.87	0.04	0.05	0.00	0.00	0.52	-0.27	0.52	-0.29	-0.29	A+	B-	B-	B+	-1.30	0.07	-6.18	0.74	-7.20	0.47
67	826275	3	5537	0.58	0.58	0.16	0.18	0.08	0.00	0.00	0.40	0.40	-0.20	-0.15	-0.23	A-	A-	A-	A-	0.69	0.05	0.55	1.01	0.65	1.02
68	826270	3	5537	0.78	0.07	0.09	0.05	0.78	0.00	0.00	0.53	-0.29	-0.22	-0.35	0.53	A+	A-	B-	A+	-0.51	0.06	-4.64	0.85	-4.30	0.77
69	826280	3	5537	0.73	0.09	0.13	0.73	0.05	0.00	0.00	0.52	-0.30	-0.24	0.52	-0.28	A+	B-	B-	A-	-0.16	0.06	-3.36	0.91	-4.30	0.81
70	826269	3	5537	0.68	0.22	0.68	0.06	0.04	0.00	0.00	0.24	-0.04	0.24	-0.21	-0.22	A-	B-	A-	A-	0.11	0.06	7.20	1.20	8.70	1.40
71	798608	4	5526	0.83	0.83	0.07	0.06	0.03	0.00	0.00	0.45	0.45	-0.28	-0.25	-0.19	A-	A-	B-	A+	-0.89	0.07	-1.27	0.95	-3.20	0.78
72	798605	4	5526	0.60	0.12	0.09	0.60	0.19	0.00	0.00	0.29	-0.15	-0.17	0.29	-0.11	A-	A-	A-	A+	0.60	0.05	6.86	1.15	6.49	1.21
73	798606	4	5526	0.65	0.15	0.65	0.11	0.10	0.00	0.00	0.34	-0.15	0.34	-0.14	-0.22	A-	A-	A-	A+	0.31	0.05	2.14	1.05	0.87	1.03
74	798611	4	5526	0.69	0.07	0.16	0.08	0.69	0.00	0.00	0.39	-0.23	-0.11	-0.30	0.39	A-	B-	B-	A+	0.09	0.05	-0.74	0.98	-0.16	0.99
75	798610	4	5526	0.28	0.51	0.28	0.11	0.10	0.00	0.00	0.04	0.07	0.04	-0.15	-0.02	A+	B+	A+	A-	2.22	0.06	8.64	1.23	9.90	1.67
76	798609	4	5526	0.76	0.08	0.09	0.76	0.07	0.00	0.00	0.34	-0.15	-0.23	0.34	-0.16	A+	A-	A-	A+	-0.34	0.06	0.85	1.03	3.22	1.18
77	826276	4	5526	0.43	0.43	0.18	0.32	0.06	0.00	0.00	0.19	0.19	-0.20	0.01	-0.10	A-	A+	A+	A-	1.41	0.05	8.11	1.16	8.69	1.28
78	826271	4	5526	0.67	0.07	0.18	0.67	0.08	0.00	0.00	0.43	-0.27	-0.20	0.43	-0.19	B-	B-	A-	A+	0.23	0.05	-1.80	0.96	-2.58	0.91
79	826279	4	5526	0.55	0.14	0.55	0.15	0.16	0.00	0.00	0.29	-0.17	0.29	-0.20	-0.03	A-	A-	A-	A-	0.83	0.05	4.61	1.09	4.97	1.15
80	826274	4	5526	0.56	0.30	0.05	0.56	0.09	0.00	0.00	0.41	-0.15	-0.26	0.41	-0.27	A-	A-	A-	A-	0.76	0.05	-0.63	0.99	-0.16	1.00
81	826281	4	5526	0.76	0.03	0.10	0.11	0.76	0.00	0.00	0.40	-0.26	-0.14	-0.26	0.40	A+	A-	A-	A-	-0.31	0.06	0.94	1.03	0.66	1.03
82	826273	4	5526	0.71	0.71	0.07	0.16	0.06	0.00	0.00	0.35	0.35	-0.32	-0.11	-0.15	A+	A-	A+	A-	-0.03	0.06	0.32	1.01	0.54	1.02
83	820431	5	5609	0.09	0.62	0.05	0.09	0.23	0.00	0.00	-0.05	0.22	-0.18	-0.05	-0.12	A-	A-	A-	A+	3.69	0.08	3.27	1.20	9.90	3.44
84	820432	5	5609	0.38	0.37	0.09	0.17	0.38	0.00	0.00	0.29	0.04	-0.25	-0.23	0.29	A-	A+	A+	A-	1.68	0.05	1.84	1.04	3.27	1.11
85	820433	5	5609	0.60	0.21	0.06	0.60	0.13	0.00	0.00	0.28	-0.01	-0.25	0.28	-0.21	A+	A-	A-	A-	0.58	0.05	6.86	1.16	7.82	1.26
86	820423	5	5609	0.52	0.52	0.17	0.22	0.08	0.00	0.00	0.31	0.31	-0.17	-0.09	-0.18	A+	A-	A-	A+	0.96	0.05	4.30	1.09	5.71	1.16
87	820427	5	5609	0.59	0.11	0.59	0.14	0.15	0.00	0.00	0.46	-0.16	0.46	-0.20	-0.28	A+	A-	A-	A-	0.63	0.05	-2.69	0.94	-1.54	0.95
88	820434	5	5609	0.48	0.26	0.13	0.48	0.13	0.00	0.00	0.28	-0.02	-0.24	0.28	-0.14	A+	A-	A-	A+	1.16	0.05	6.34	1.13	6.82	1.20
89	818084	5	5609	0.77	0.09	0.08	0.05	0.77	0.00	0.00	0.51	-0.33	-0.23	-0.23	0.51	A+	A-	A-	A-	-0.41	0.06	-4.34	0.87	-4.62	0.77

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
90	818073	5	5609	0.56	0.25	0.56	0.15	0.04	0.00	0.00	0.42	-0.20	0.42	-0.21	-0.22	A-	A-	A-	A-	0.80	0.05	-2.39	0.95	-2.40	0.93
91	818083	5	5609	0.53	0.16	0.53	0.19	0.12	0.00	0.00	0.42	-0.09	0.42	-0.23	-0.24	A-	B-	A-	A+	0.94	0.05	-0.29	0.99	0.49	1.01
92	818075	5	5609	0.71	0.07	0.08	0.71	0.14	0.00	0.00	0.53	-0.28	-0.37	0.53	-0.19	A+	A-	A-	A-	-0.05	0.06	-5.03	0.87	-5.79	0.77
93	818081	5	5609	0.66	0.66	0.15	0.11	0.08	0.00	0.00	0.36	0.36	-0.21	-0.09	-0.22	A-	A-	A-	A+	0.28	0.05	3.16	1.08	3.15	1.12
94	818072	5	5609	0.40	0.40	0.20	0.05	0.35	0.00	0.00	-0.02	-0.02	0.12	-0.24	0.04	A+	A+	A+	A-	1.56	0.05	9.90	1.43	9.90	1.62
95	820424	6	5543	0.51	0.51	0.19	0.15	0.15	0.00	0.00	0.33	0.33	-0.20	-0.09	-0.15	A-	A-	A-	A+	1.00	0.05	4.79	1.10	4.75	1.15
96	820426	6	5543	0.80	0.80	0.04	0.10	0.06	0.00	0.00	0.44	0.44	-0.24	-0.21	-0.27	A+	A-	B-	A-	-0.61	0.06	-1.53	0.95	-2.24	0.87
97	820425	6	5543	0.73	0.08	0.73	0.07	0.12	0.00	0.00	0.45	-0.18	0.45	-0.23	-0.27	A+	A-	A-	B-	-0.19	0.06	1.02	1.03	0.21	1.01
98	820430	6	5543	0.74	0.18	0.74	0.04	0.04	0.00	0.00	0.31	-0.10	0.31	-0.23	-0.24	A+	A+	A+	A-	-0.22	0.06	4.64	1.14	4.37	1.23
99	820429	6	5543	0.63	0.05	0.14	0.18	0.63	0.00	0.00	0.47	-0.28	-0.21	-0.24	0.47	A+	A-	A-	A-	0.43	0.05	-2.51	0.94	-2.62	0.91
100	820428	6	5543	0.68	0.08	0.04	0.20	0.68	0.00	0.00	0.36	-0.20	-0.24	-0.16	0.36	A+	A-	A-	A+	0.16	0.05	0.89	1.02	2.24	1.09
101	818078	6	5543	0.50	0.50	0.14	0.07	0.29	0.00	0.00	0.26	0.26	-0.14	-0.25	-0.04	A-	A-	A-	A-	1.05	0.05	7.95	1.17	8.06	1.26
102	818076	6	5543	0.59	0.13	0.21	0.07	0.59	0.00	0.00	0.47	-0.26	-0.19	-0.25	0.47	A+	A-	A-	A+	0.62	0.05	-1.88	0.96	-2.20	0.93
103	818079	6	5543	0.48	0.43	0.48	0.04	0.04	0.00	0.00	0.22	0.01	0.22	-0.28	-0.28	A-	A-	A-	A-	1.14	0.05	9.90	1.25	9.90	1.39
104	818077	6	5543	0.50	0.16	0.19	0.15	0.50	0.00	0.00	0.38	-0.09	-0.14	-0.27	0.38	A+	A-	A-	A-	1.08	0.05	0.55	1.01	1.61	1.05
105	818085	6	5543	0.71	0.05	0.17	0.71	0.06	0.00	0.00	0.32	-0.23	-0.12	0.32	-0.18	A+	A-	A-	A+	-0.05	0.06	4.16	1.11	5.78	1.28
106	818080	6	5543	0.77	0.07	0.08	0.77	0.08	0.00	0.00	0.43	-0.20	-0.22	0.43	-0.25	A+	B-	B-	A+	-0.45	0.06	-3.73	0.89	-2.26	0.88
107	812303	7	5549	0.58	0.25	0.11	0.58	0.06	0.00	0.00	0.31	-0.12	-0.26	0.31	-0.08	A+	A-	A-	A-	0.68	0.05	6.51	1.14	6.00	1.20
108	812307	7	5549	0.66	0.13	0.66	0.14	0.07	0.00	0.00	0.37	-0.27	0.37	-0.19	-0.05	A+	A-	A-	A-	0.27	0.05	2.83	1.07	3.08	1.12
109	812304	7	5549	0.71	0.07	0.71	0.03	0.18	0.00	0.00	0.39	-0.23	0.39	-0.25	-0.18	A+	A-	A-	A+	-0.04	0.06	-0.19	0.99	-0.59	0.97
110	812301	7	5549	0.60	0.07	0.27	0.60	0.06	0.00	0.00	0.27	-0.21	-0.07	0.27	-0.19	A-	A-	A-	A+	0.58	0.05	6.36	1.15	4.99	1.17
111	812312	7	5549	0.84	0.84	0.06	0.05	0.06	0.00	0.00	0.50	0.50	-0.30	-0.26	-0.25	A+	A-	A-	A-	-0.95	0.07	-5.25	0.80	-4.96	0.67
112	812308	7	5549	0.90	0.90	0.03	0.03	0.05	0.00	0.00	0.53	0.53	-0.27	-0.26	-0.35	B+	A-	A-	A+	-1.54	0.08	-5.35	0.74	-7.64	0.39
113	810298	7	5549	0.65	0.07	0.17	0.10	0.65	0.00	0.00	0.44	-0.31	-0.20	-0.17	0.44	A-	A-	A-	A-	0.29	0.05	0.14	1.00	1.94	1.07
114	810297	7	5549	0.54	0.41	0.03	0.54	0.02	0.00	0.00	0.37	-0.24	-0.22	0.37	-0.16	C-	C-	B-	A+	0.89	0.05	1.53	1.03	2.09	1.06
115	810300	7	5549	0.67	0.12	0.67	0.10	0.10	0.00	0.00	0.36	-0.11	0.36	-0.23	-0.19	A+	A+	A-	A+	0.19	0.05	1.97	1.05	1.67	1.07
116	810306	7	5549	0.41	0.41	0.18	0.18	0.23	0.00	0.00	0.33	0.33	-0.21	-0.14	-0.06	A-	A-	A-	A+	1.53	0.05	1.96	1.04	4.87	1.17
117	810301	7	5549	0.59	0.10	0.12	0.19	0.59	0.00	0.00	0.44	-0.22	-0.22	-0.19	0.44	A-	A-	A-	A+	0.62	0.05	-1.95	0.96	-1.82	0.94
118	810296	7	5549	0.62	0.07	0.19	0.62	0.11	0.00	0.00	0.36	-0.26	-0.07	0.36	-0.25	A-	A-	A-	A+	0.44	0.05	3.90	1.09	3.02	1.11
119	812302	8	5610	0.77	0.06	0.77	0.09	0.09	0.00	0.00	0.46	-0.20	0.46	-0.28	-0.24	A+	A-	A-	A-	-0.44	0.06	-2.14	0.93	-2.52	0.86

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
120	812309	8	5610	0.67	0.09	0.05	0.20	0.67	0.00	0.00	0.50	-0.36	-0.21	-0.21	0.50	A+	A-	A-	A-	0.21	0.05	-4.89	0.88	-4.51	0.83
121	812310	8	5610	0.83	0.83	0.08	0.06	0.03	0.00	0.00	0.45	0.45	-0.27	-0.25	-0.19	A-	A-	B-	A-	-0.89	0.07	-1.02	0.96	-2.52	0.82
122	812305	8	5610	0.56	0.18	0.18	0.56	0.08	0.00	0.00	0.30	-0.20	-0.10	0.30	-0.12	A+	A-	A-	A-	0.74	0.05	4.23	1.09	3.97	1.13
123	812306	8	5610	0.64	0.09	0.13	0.14	0.64	0.00	0.00	0.33	-0.20	-0.13	-0.15	0.33	A-	A-	A-	A-	0.35	0.05	4.40	1.11	5.18	1.20
124	820817	8	5610	0.73	0.05	0.02	0.20	0.73	0.00	0.00	0.38	-0.26	-0.27	-0.17	0.38	A+	A-	A+	A-	-0.16	0.06	1.08	1.03	1.86	1.09
125	810302	8	5610	0.52	0.52	0.09	0.26	0.12	0.00	0.00	0.16	0.16	-0.29	0.09	-0.09	A-	A-	A-	A-	0.95	0.05	9.90	1.28	9.90	1.44
126	810304	8	5610	0.32	0.18	0.32	0.36	0.13	0.00	0.00	0.23	-0.10	0.23	-0.02	-0.16	A+	A-	A-	A-	1.97	0.05	3.65	1.08	7.94	1.37
127	810295	8	5610	0.85	0.05	0.03	0.85	0.06	0.00	0.00	0.47	-0.25	-0.27	0.47	-0.25	A-	B-	B-	A+	-1.10	0.07	-4.73	0.81	-4.32	0.68
128	810303	8	5610	0.53	0.53	0.10	0.24	0.12	0.00	0.00	0.34	0.34	-0.21	-0.08	-0.20	A-	A-	A-	A-	0.91	0.05	4.23	1.09	3.77	1.11
129	810305	8	5610	0.83	0.05	0.83	0.09	0.03	0.00	0.00	0.49	-0.29	0.49	-0.27	-0.24	A+	A-	A-	A-	-0.85	0.07	-4.40	0.84	-4.73	0.70
130	810299	8	5610	0.74	0.09	0.07	0.10	0.74	0.00	0.00	0.47	-0.26	-0.24	-0.22	0.47	A-	B-	B-	A-	-0.23	0.06	-2.80	0.92	-2.87	0.86
131	815898	9	5553	0.60	0.60	0.09	0.10	0.21	0.00	0.00	0.36	0.36	-0.24	-0.18	-0.12	A-	A-	B-	A+	0.57	0.05	3.90	1.09	3.30	1.11
132	815893	9	5553	0.73	0.73	0.07	0.12	0.08	0.00	0.00	0.55	0.55	-0.25	-0.36	-0.22	A+	A-	B-	A-	-0.20	0.06	-5.10	0.86	-5.62	0.76
133	815901	9	5553	0.39	0.14	0.39	0.43	0.04	0.00	0.00	0.05	-0.18	0.05	0.18	-0.24	A-	A-	A-	A-	1.65	0.05	9.90	1.36	9.90	1.67
134	815896	9	5553	0.55	0.19	0.13	0.13	0.55	0.00	0.00	0.43	-0.16	-0.21	-0.22	0.43	A-	A-	B-	A-	0.82	0.05	-1.62	0.97	-1.37	0.96
135	815899	9	5553	0.80	0.09	0.04	0.07	0.80	0.00	0.00	0.54	-0.35	-0.29	-0.23	0.54	A+	A-	A-	A-	-0.66	0.06	-4.77	0.84	-5.30	0.70
136	815902	9	5553	0.38	0.03	0.49	0.10	0.38	0.00	0.00	0.25	-0.28	0.03	-0.26	0.25	A+	B-	A-	A+	1.69	0.05	7.74	1.16	9.62	1.40
137	799210	9	5553	0.90	0.03	0.03	0.90	0.03	0.00	0.00	0.42	-0.22	-0.26	0.42	-0.21	A-	A-	A-	B+	-1.58	0.08	-2.30	0.88	-3.33	0.69
138	799209	9	5553	0.39	0.39	0.52	0.04	0.05	0.00	0.00	0.22	0.22	0.00	-0.25	-0.26	B+	A-	A+	A-	1.61	0.05	5.77	1.12	7.81	1.30
139	799212	9	5553	0.83	0.08	0.04	0.05	0.83	0.00	0.00	0.51	-0.28	-0.25	-0.30	0.51	C+	A-	A-	A-	-0.89	0.07	-3.05	0.88	-3.93	0.74
140	799205	9	5553	0.76	0.06	0.76	0.09	0.09	0.00	0.00	0.34	-0.17	0.34	-0.20	-0.15	A+	A-	A+	A-	-0.36	0.06	2.81	1.09	3.70	1.21
141	799204	9	5553	0.80	0.80	0.05	0.08	0.07	0.00	0.00	0.52	0.52	-0.24	-0.31	-0.26	A+	A-	A-	A-	-0.62	0.06	-3.41	0.89	-3.98	0.77
142	799208	9	5553	0.68	0.14	0.68	0.08	0.10	0.00	0.00	0.45	-0.15	0.45	-0.32	-0.24	A+	A-	A-	A-	0.16	0.05	-1.22	0.97	-1.65	0.94
143	815894	10	5517	0.68	0.68	0.06	0.24	0.02	0.00	0.00	0.46	0.46	-0.27	-0.31	-0.13	A+	A-	A-	A-	0.16	0.05	-1.76	0.96	-1.93	0.92
144	815895	10	5517	0.75	0.06	0.12	0.75	0.08	0.00	0.00	0.43	-0.27	-0.22	0.43	-0.20	A+	A-	A-	A+	-0.25	0.06	0.21	1.01	-0.16	0.99
145	815897	10	5517	0.68	0.14	0.14	0.68	0.03	0.00	0.00	0.44	-0.25	-0.22	0.44	-0.23	B+	A+	A-	A+	0.15	0.05	-1.60	0.96	-1.18	0.95
146	815891	10	5517	0.88	0.05	0.88	0.02	0.05	0.00	0.00	0.46	-0.27	0.46	-0.22	-0.26	A+	A-	A-	A-	-1.30	0.07	-2.36	0.89	-4.61	0.62
147	815892	10	5517	0.79	0.11	0.06	0.79	0.04	0.00	0.00	0.51	-0.28	-0.29	0.51	-0.26	A+	A-	A-	A-	-0.55	0.06	-2.83	0.90	-3.26	0.81
148	815900	10	5517	0.78	0.04	0.78	0.11	0.07	0.00	0.00	0.42	-0.25	0.42	-0.13	-0.32	B+	A-	A-	A-	-0.49	0.06	0.02	1.00	1.52	1.09
149	799211	10	5517	0.84	0.05	0.05	0.84	0.05	0.00	0.00	0.46	-0.28	-0.29	0.46	-0.18	B+	A-	A-	A-	-0.98	0.07	-3.38	0.86	-3.40	0.75

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
150	799203	10	5517	0.74	0.08	0.74	0.07	0.10	0.00	0.00	0.47	-0.23	0.47	-0.28	-0.24	A-	A-	A-	A+	-0.23	0.06	-1.77	0.95	-2.70	0.87
151	799202	10	5517	0.91	0.91	0.03	0.03	0.02	0.00	0.00	0.47	0.47	-0.26	-0.27	-0.26	B+	A-	A-	A-	-1.76	0.09	-3.47	0.80	-3.99	0.59
152	799207	10	5517	0.44	0.04	0.24	0.27	0.44	0.00	0.00	0.39	-0.24	-0.10	-0.23	0.39	A+	A-	A-	A+	1.42	0.05	-0.18	1.00	1.85	1.06
153	799206	10	5517	0.82	0.03	0.05	0.82	0.10	0.00	0.00	0.45	-0.25	-0.24	0.45	-0.27	A+	A+	A+	A-	-0.74	0.06	-1.56	0.94	-2.02	0.87
154	799201	10	5517	0.61	0.10	0.14	0.15	0.61	0.00	0.00	0.39	-0.21	-0.16	-0.20	0.39	A+	A-	A-	A+	0.55	0.05	0.77	1.02	0.40	1.01
155	819127	11	5550	0.80	0.80	0.09	0.04	0.07	0.00	0.00	0.38	0.38	-0.16	-0.29	-0.19	A-	A-	A-	A+	-0.63	0.06	-1.61	0.94	-2.27	0.86
156	819122	11	5550	0.83	0.05	0.83	0.05	0.08	0.00	0.00	0.48	-0.28	0.48	-0.30	-0.20	A-	A-	A-	A+	-0.83	0.07	-4.30	0.84	-4.71	0.70
157	819121	11	5550	0.60	0.23	0.60	0.03	0.13	0.00	0.00	0.39	-0.18	0.39	-0.18	-0.23	A-	A+	A-	A+	0.59	0.05	1.33	1.03	1.28	1.04
158	819126	11	5550	0.91	0.03	0.04	0.91	0.02	0.00	0.00	0.51	-0.28	-0.32	0.51	-0.24	A+	A-	C-	A-	-1.76	0.08	-5.60	0.70	-7.54	0.35
159	819130	11	5550	0.70	0.06	0.15	0.10	0.70	0.00	0.00	0.48	-0.21	-0.30	-0.21	0.48	A+	A-	A-	A-	0.07	0.05	-2.97	0.92	-3.90	0.84
160	819123	11	5550	0.54	0.54	0.11	0.16	0.19	0.00	0.00	0.40	0.40	-0.26	-0.06	-0.24	A+	A-	A-	A-	0.90	0.05	-0.76	0.98	0.19	1.01
161	824931	11	5550	0.29	0.45	0.21	0.29	0.05	0.00	0.00	0.26	-0.02	-0.12	0.26	-0.24	A+	A+	A+	A+	2.22	0.06	2.39	1.06	7.55	1.42
162	824934	11	5550	0.75	0.13	0.75	0.08	0.04	0.00	0.00	0.50	-0.26	0.50	-0.29	-0.24	B+	B-	A-	A-	-0.28	0.06	-2.03	0.94	-3.83	0.81
163	824925	11	5550	0.76	0.04	0.16	0.05	0.76	0.00	0.00	0.47	-0.20	-0.29	-0.28	0.47	B+	A+	A-	A-	-0.31	0.06	-2.12	0.94	-2.54	0.87
164	824929	11	5550	0.65	0.10	0.07	0.18	0.65	0.00	0.00	0.27	-0.06	-0.17	-0.17	0.27	B+	A-	A-	A+	0.35	0.05	5.06	1.12	5.56	1.21
165	824928	11	5550	0.74	0.07	0.11	0.74	0.07	0.00	0.00	0.52	-0.25	-0.27	0.52	-0.29	A+	A-	A-	A+	-0.21	0.06	-4.15	0.88	-5.34	0.76
166	824924	11	5550	0.33	0.06	0.59	0.33	0.03	0.00	0.00	0.09	-0.30	0.14	0.09	-0.24	A-	A-	A-	A+	2.00	0.05	9.90	1.28	9.90	1.69
167	819120	12	5522	0.77	0.12	0.04	0.77	0.06	0.00	0.00	0.46	-0.26	-0.24	0.46	-0.23	A+	A-	A-	A+	-0.42	0.06	-2.65	0.92	-2.39	0.88
168	819124	12	5522	0.80	0.10	0.04	0.05	0.80	0.00	0.00	0.33	-0.17	-0.15	-0.21	0.33	A+	A-	A-	A+	-0.64	0.06	0.45	1.02	0.88	1.05
169	819131	12	5522	0.42	0.42	0.33	0.07	0.18	0.00	0.00	0.05	0.05	-0.14	-0.13	0.20	A-	A-	A-	A+	1.46	0.05	9.90	1.38	9.90	1.54
170	819125	12	5522	0.47	0.16	0.47	0.22	0.14	0.00	0.00	0.25	-0.17	0.25	-0.13	-0.02	A-	A-	A-	A-	1.20	0.05	6.73	1.13	6.86	1.21
171	819128	12	5522	0.53	0.19	0.13	0.53	0.15	0.00	0.00	0.36	-0.22	0.01	0.36	-0.26	A+	A+	A-	A-	0.93	0.05	2.19	1.04	3.42	1.10
172	819129	12	5522	0.31	0.21	0.14	0.34	0.31	0.00	0.00	0.15	0.02	-0.24	0.02	0.15	A+	A-	A+	A+	2.05	0.05	9.48	1.23	9.90	1.57
173	824932	12	5522	0.57	0.25	0.57	0.10	0.08	0.00	0.00	0.43	-0.21	0.43	-0.22	-0.20	A+	A+	A-	A-	0.70	0.05	-1.02	0.98	-1.21	0.96
174	824926	12	5522	0.76	0.76	0.06	0.09	0.09	0.00	0.00	0.40	0.40	-0.29	-0.24	-0.10	A+	A-	A-	A+	-0.35	0.06	-1.46	0.96	-0.79	0.96
175	824927	12	5522	0.39	0.03	0.21	0.39	0.36	0.00	0.00	0.29	-0.25	-0.15	0.29	-0.07	A+	A-	A-	A-	1.61	0.05	2.29	1.05	5.42	1.19
176	824923	12	5522	0.69	0.69	0.22	0.05	0.04	0.00	0.00	0.39	0.39	-0.14	-0.29	-0.28	A+	A-	A-	A+	0.06	0.05	1.12	1.03	0.94	1.04
177	824933	12	5522	0.68	0.68	0.06	0.22	0.04	0.00	0.00	0.48	0.48	-0.31	-0.26	-0.21	B+	A-	A+	A-	0.14	0.05	-3.46	0.92	-4.38	0.84
178	824930	12	5522	0.18	0.19	0.18	0.19	0.43	0.00	0.00	-0.05	-0.03	-0.05	0.00	0.07	A+	A-	A-	A-	2.85	0.06	6.26	1.24	9.90	2.48
179	822108	13	5484	0.75	0.15	0.06	0.03	0.75	0.00	0.00	0.48	-0.31	-0.25	-0.20	0.48	A-	B-	B-	A+	-0.31	0.06	-2.59	0.92	-4.14	0.80

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
180	822112	13	5484	0.81	0.12	0.04	0.81	0.03	0.00	0.00	0.52	-0.36	-0.24	0.52	-0.20	B+	A-	A-	A-	-0.71	0.06	-4.32	0.85	-5.03	0.70
181	822104	13	5484	0.57	0.57	0.28	0.03	0.12	0.00	0.00	0.35	0.35	-0.14	-0.29	-0.17	A+	A-	A-	A-	0.74	0.05	5.08	1.11	5.14	1.17
182	822110	13	5484	0.39	0.15	0.11	0.34	0.39	0.00	0.00	0.24	-0.14	-0.18	-0.01	0.24	A+	A-	A-	A+	1.65	0.05	7.03	1.15	8.25	1.33
183	822115	13	5484	0.72	0.72	0.14	0.08	0.06	0.00	0.00	0.42	0.42	-0.18	-0.24	-0.25	A-	A-	A+	A-	-0.09	0.06	0.66	1.02	-0.32	0.98
184	822106	13	5484	0.34	0.04	0.19	0.34	0.42	0.00	0.00	0.32	-0.27	-0.08	0.32	-0.13	A-	A+	A-	A-	1.89	0.05	-0.59	0.99	4.02	1.18
185	805341	13	5484	0.52	0.07	0.36	0.52	0.05	0.00	0.00	0.36	-0.15	-0.22	0.36	-0.17	A-	A-	A-	A-	0.99	0.05	2.91	1.06	3.72	1.11
186	805335	13	5484	0.88	0.88	0.02	0.05	0.05	0.00	0.00	0.39	0.39	-0.24	-0.17	-0.25	A+	B-	C-	A-	-1.37	0.08	-1.82	0.91	-0.68	0.93
187	805340	13	5484	0.86	0.06	0.86	0.03	0.04	0.00	0.00	0.49	-0.28	0.49	-0.26	-0.26	A+	C-	C-	A+	-1.19	0.07	-5.07	0.79	-5.06	0.62
188	805342	13	5484	0.81	0.14	0.03	0.81	0.02	0.00	0.00	0.49	-0.31	-0.29	0.49	-0.24	A+	A-	A-	A-	-0.74	0.06	-3.71	0.87	-3.41	0.79
189	805344	13	5484	0.79	0.06	0.79	0.12	0.03	0.00	0.00	0.52	-0.29	0.52	-0.29	-0.27	B+	A-	A-	A-	-0.55	0.06	-4.92	0.84	-3.73	0.79
190	805338	13	5484	0.85	0.04	0.04	0.06	0.85	0.00	0.00	0.52	-0.25	-0.31	-0.28	0.52	B+	B-	B-	A-	-1.01	0.07	-3.74	0.85	-4.70	0.67
191	822114	14	5526	0.66	0.10	0.66	0.10	0.13	0.00	0.00	0.30	-0.14	0.30	-0.22	-0.10	B-	A-	A-	A-	0.22	0.05	5.16	1.13	5.11	1.21
192	822111	14	5526	0.70	0.07	0.17	0.05	0.70	0.00	0.00	0.41	-0.31	-0.13	-0.25	0.41	A+	A-	A-	A-	0.01	0.06	-0.92	0.98	-1.83	0.92
193	822105	14	5526	0.45	0.45	0.42	0.07	0.06	0.00	0.00	0.40	0.40	-0.14	-0.30	-0.24	A-	A-	A-	A-	1.32	0.05	-2.11	0.96	0.24	1.01
194	822107	14	5526	0.70	0.09	0.70	0.16	0.06	0.00	0.00	0.39	-0.27	0.39	-0.18	-0.17	A+	A-	A-	A+	0.04	0.05	-1.01	0.97	-1.24	0.95
195	822113	14	5526	0.47	0.09	0.47	0.38	0.06	0.00	0.00	0.46	-0.19	0.46	-0.24	-0.23	A+	A-	A+	A-	1.22	0.05	-6.30	0.88	-2.07	0.94
196	822109	14	5526	0.39	0.11	0.29	0.39	0.21	0.00	0.00	0.26	-0.23	0.07	0.26	-0.20	A-	A-	A-	A-	1.61	0.05	5.14	1.11	8.15	1.31
197	805333	14	5526	0.65	0.65	0.17	0.05	0.13	0.00	0.00	0.41	0.41	-0.15	-0.25	-0.24	A+	A-	A-	A+	0.30	0.05	0.22	1.01	-0.96	0.96
198	805336	14	5526	0.80	0.07	0.09	0.80	0.04	0.00	0.00	0.51	-0.27	-0.32	0.51	-0.21	A+	A-	A-	A-	-0.63	0.06	-3.90	0.87	-5.06	0.71
199	805334	14	5526	0.74	0.08	0.74	0.08	0.11	0.00	0.00	0.38	-0.21	0.38	-0.26	-0.13	A-	A-	A-	A-	-0.20	0.06	1.37	1.04	1.22	1.06
200	805337	14	5526	0.74	0.09	0.09	0.08	0.74	0.00	0.00	0.49	-0.22	-0.26	-0.30	0.49	A+	A-	A-	A-	-0.21	0.06	-4.56	0.87	-4.85	0.78
201	805339	14	5526	0.28	0.28	0.24	0.22	0.26	0.00	0.00	0.10	0.10	0.02	0.02	-0.13	A-	A+	A-	A-	2.24	0.06	8.84	1.25	9.90	1.65
202	805343	14	5526	0.64	0.12	0.18	0.05	0.64	0.00	0.00	0.33	-0.12	-0.17	-0.24	0.33	A+	A-	A-	A-	0.34	0.05	4.09	1.10	3.41	1.13
203	804726	15	5526	0.88	0.88	0.06	0.04	0.02	0.00	0.00	0.40	0.40	-0.26	-0.21	-0.19	A+	A-	A-	A-	-1.32	0.07	-0.66	0.97	0.69	1.06
204	804724	15	5526	0.88	0.88	0.06	0.04	0.02	0.00	0.00	0.40	0.40	-0.22	-0.23	-0.21	A-	A-	B-	A+	-1.35	0.07	-1.99	0.91	-2.44	0.78
205	804731	15	5526	0.69	0.05	0.22	0.04	0.69	0.00	0.00	0.46	-0.28	-0.23	-0.29	0.46	A+	A-	A-	A-	0.08	0.06	-1.57	0.96	-2.48	0.90
206	804730	15	5526	0.78	0.13	0.04	0.78	0.05	0.00	0.00	0.47	-0.25	-0.29	0.47	-0.24	A-	B-	B-	A-	-0.47	0.06	-1.71	0.95	-2.86	0.84
207	804733	15	5526	0.71	0.04	0.71	0.20	0.04	0.00	0.00	0.44	-0.28	0.44	-0.23	-0.22	A+	A+	A-	B-	-0.03	0.06	-0.44	0.99	0.09	1.00
208	804734	15	5526	0.71	0.12	0.11	0.05	0.71	0.00	0.00	0.45	-0.22	-0.23	-0.26	0.45	A-	B-	A-	A-	-0.03	0.06	-2.53	0.93	-2.91	0.87
209	811022	15	5526	0.71	0.14	0.08	0.07	0.71	0.00	0.00	0.46	-0.22	-0.22	-0.27	0.46	A-	A-	B-	A+	-0.01	0.06	-2.41	0.94	-2.44	0.89

Ref	ID	Form	N	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
210	811016	15	5526	0.82	0.02	0.82	0.10	0.05	0.00	0.00	0.43	-0.26	0.43	-0.19	-0.30	A+	A-	A-	A-	-0.83	0.06	-1.47	0.95	-1.64	0.89
211	811017	15	5526	0.74	0.74	0.18	0.05	0.03	0.00	0.00	0.34	0.34	-0.16	-0.26	-0.15	A-	A-	A-	A-	-0.23	0.06	3.25	1.10	2.71	1.14
212	811015	15	5526	0.81	0.81	0.09	0.06	0.04	0.00	0.00	0.39	0.39	-0.16	-0.28	-0.20	A-	A-	A-	A+	-0.73	0.06	0.38	1.01	0.28	1.02
213	811021	15	5526	0.78	0.11	0.78	0.06	0.05	0.00	0.00	0.50	-0.26	0.50	-0.31	-0.23	A-	A-	A-	A-	-0.49	0.06	-3.08	0.90	-3.01	0.83
214	811024	15	5526	0.74	0.09	0.08	0.74	0.09	0.00	0.00	0.39	-0.08	-0.22	0.39	-0.30	A+	A-	A-	A-	-0.24	0.06	-0.04	1.00	1.45	1.07
215	804727	16	5565	0.57	0.23	0.57	0.16	0.04	0.00	0.00	0.26	-0.04	0.26	-0.20	-0.20	A+	A-	A+	A-	0.73	0.05	9.90	1.24	9.90	1.43
216	804728	16	5565	0.87	0.03	0.05	0.87	0.06	0.00	0.00	0.43	-0.28	-0.22	0.43	-0.23	A+	B-	B-	A-	-1.25	0.07	-4.26	0.82	-4.02	0.68
217	804722	16	5565	0.91	0.02	0.04	0.03	0.91	0.00	0.00	0.44	-0.21	-0.27	-0.24	0.44	A+	B-	B-	A-	-1.72	0.08	-3.26	0.82	-3.62	0.63
218	804732	16	5565	0.81	0.03	0.11	0.81	0.05	0.00	0.00	0.51	-0.26	-0.33	0.51	-0.23	A+	A-	A-	A-	-0.76	0.06	-5.23	0.82	-5.44	0.67
219	804729	16	5565	0.75	0.15	0.75	0.07	0.03	0.00	0.00	0.54	-0.33	0.54	-0.29	-0.24	A+	A-	B-	A-	-0.30	0.06	-5.29	0.85	-6.26	0.71
220	804725	16	5565	0.81	0.81	0.06	0.06	0.07	0.00	0.00	0.42	0.42	-0.24	-0.29	-0.15	A+	A-	A-	A-	-0.73	0.06	-1.21	0.96	-0.47	0.97
221	811014	16	5565	0.57	0.04	0.57	0.30	0.09	0.00	0.00	0.30	-0.24	0.30	-0.07	-0.22	A-	B-	A-	A-	0.75	0.05	7.85	1.17	8.76	1.30
222	811019	16	5565	0.73	0.11	0.12	0.73	0.03	0.00	0.00	0.45	-0.22	-0.27	0.45	-0.22	A-	A-	A-	A-	-0.18	0.06	-1.41	0.96	-1.84	0.91
223	811025	16	5565	0.77	0.04	0.08	0.10	0.77	0.00	0.00	0.44	-0.26	-0.22	-0.24	0.44	B-	B-	B-	A-	-0.46	0.06	-1.49	0.95	-1.18	0.93
224	811018	16	5565	0.81	0.81	0.03	0.12	0.03	0.00	0.00	0.51	0.51	-0.26	-0.32	-0.27	A-	B-	C-	B-	-0.75	0.06	-2.64	0.91	-3.49	0.78
225	811023	16	5565	0.60	0.06	0.15	0.60	0.19	0.00	0.00	0.38	-0.27	-0.22	0.38	-0.11	A+	A+	A+	B-	0.60	0.05	2.25	1.05	2.92	1.10
226	811020	16	5565	0.75	0.08	0.08	0.09	0.75	0.00	0.00	0.54	-0.28	-0.29	-0.27	0.54	A+	B-	B-	A-	-0.28	0.06	-5.14	0.85	-5.59	0.74
227	811532	17	5572	0.78	0.78	0.06	0.13	0.02	0.00	0.00	0.36	0.36	-0.22	-0.21	-0.14	A-	A-	A-	A-	-0.45	0.06	1.60	1.05	1.43	1.08
228	811531	17	5572	0.71	0.11	0.12	0.06	0.71	0.00	0.00	0.36	-0.14	-0.22	-0.21	0.36	A+	A-	A+	A+	0.02	0.06	1.07	1.03	0.98	1.04
229	811529	17	5572	0.79	0.02	0.79	0.09	0.10	0.00	0.00	0.39	-0.22	0.39	-0.24	-0.19	A+	A-	A-	A-	-0.49	0.06	-0.27	0.99	0.13	1.01
230	811533	17	5572	0.70	0.70	0.09	0.15	0.06	0.00	0.00	0.38	0.38	-0.14	-0.19	-0.27	A+	A-	A+	A+	0.04	0.06	-0.77	0.98	-1.72	0.93
231	811534	17	5572	0.83	0.04	0.06	0.07	0.83	0.00	0.00	0.53	-0.26	-0.27	-0.33	0.53	B+	A-	A-	A-	-0.83	0.07	-4.80	0.82	-5.24	0.66
232	811526	17	5572	0.55	0.34	0.55	0.07	0.04	0.00	0.00	0.38	-0.18	0.38	-0.23	-0.21	A+	A+	A-	A-	0.89	0.05	1.11	1.02	0.87	1.03
233	808360	17	5572	0.91	0.04	0.91	0.02	0.02	0.00	0.00	0.45	-0.29	0.45	-0.26	-0.17	A-	C-	A-	A+	-1.72	0.09	-2.96	0.83	-3.26	0.66
234	808364	17	5572	0.28	0.52	0.09	0.28	0.11	0.00	0.00	0.07	0.06	-0.11	0.07	-0.09	A-	A-	A-	A-	2.27	0.06	8.73	1.24	9.90	1.91
235	808367	17	5572	0.88	0.04	0.88	0.05	0.03	0.00	0.00	0.53	-0.28	0.53	-0.32	-0.28	A+	A-	B-	A+	-1.34	0.08	-4.39	0.80	-6.80	0.47
236	808371	17	5572	0.46	0.15	0.19	0.20	0.46	0.00	0.00	0.26	-0.21	0.01	-0.13	0.26	A-	A-	A-	A-	1.32	0.05	7.05	1.14	8.56	1.29
237	808368	17	5572	0.44	0.44	0.22	0.17	0.16	0.00	0.00	0.35	0.35	-0.16	-0.16	-0.12	A-	A-	A-	A-	1.40	0.05	1.77	1.03	3.73	1.12
238	808369	17	5572	0.30	0.25	0.04	0.30	0.41	0.00	0.00	0.00	0.12	-0.25	0.00	0.01	A-	A-	A+	A-	2.13	0.05	9.90	1.40	9.90	2.11
239	811524	18	5487	0.61	0.10	0.17	0.61	0.11	0.00	0.00	0.41	-0.22	-0.16	0.41	-0.22	A+	B-	B-	A+	0.55	0.05	1.35	1.03	0.64	1.02

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
240	811525	18	5487	0.69	0.07	0.15	0.09	0.69	0.00	0.00	0.23	-0.18	-0.07	-0.11	0.23	A+	A-	A-	A+	0.14	0.05	7.17	1.20	6.37	1.28
241	811528	18	5487	0.88	0.88	0.05	0.05	0.03	0.00	0.00	0.39	0.39	-0.14	-0.29	-0.20	A-	B-	C-	A+	-1.28	0.07	-3.49	0.84	-3.09	0.74
242	811535	18	5487	0.54	0.31	0.54	0.13	0.02	0.00	0.00	0.39	-0.24	0.39	-0.14	-0.21	A-	A-	A-	A-	0.89	0.05	2.63	1.05	2.93	1.09
243	811530	18	5487	0.88	0.02	0.07	0.88	0.03	0.00	0.00	0.36	-0.20	-0.21	0.36	-0.19	A+	B-	A-	A+	-1.36	0.08	-2.16	0.90	-0.62	0.94
244	811527	18	5487	0.72	0.17	0.05	0.72	0.05	0.00	0.00	0.43	-0.23	-0.18	0.43	-0.28	A+	A-	B-	A+	-0.09	0.06	-1.40	0.96	0.27	1.01
245	808362	18	5487	0.87	0.06	0.05	0.03	0.87	0.00	0.00	0.54	-0.34	-0.28	-0.26	0.54	A-	B-	B-	A-	-1.17	0.07	-5.08	0.78	-6.91	0.51
246	808363	18	5487	0.76	0.76	0.07	0.08	0.08	0.00	0.00	0.39	0.39	-0.24	-0.27	-0.11	A+	A-	A-	A-	-0.35	0.06	1.16	1.04	1.32	1.07
247	808366	18	5487	0.56	0.10	0.19	0.14	0.56	0.00	0.00	0.35	-0.12	-0.27	-0.09	0.35	A+	A-	A-	A+	0.79	0.05	3.83	1.08	4.92	1.16
248	808365	18	5487	0.24	0.02	0.24	0.53	0.21	0.00	0.00	0.02	-0.27	0.02	0.13	-0.07	A+	A+	A+	A-	2.52	0.06	8.05	1.25	9.90	2.16
249	808361	18	5487	0.81	0.81	0.10	0.05	0.03	0.00	0.00	0.38	0.38	-0.22	-0.16	-0.24	A+	A-	A-	A+	-0.69	0.06	-0.53	0.98	1.58	1.11
250	808370	18	5487	0.84	0.04	0.07	0.84	0.05	0.00	0.00	0.44	-0.24	-0.23	0.44	-0.25	A+	C-	B-	A+	-0.92	0.07	-3.07	0.88	-2.59	0.82
251	824176	19	5561	0.75	0.10	0.10	0.05	0.75	0.00	0.00	0.27	-0.12	-0.21	-0.06	0.27	A+	A+	A-	A-	-0.24	0.06	4.17	1.13	5.02	1.28
252	824173	19	5561	0.72	0.05	0.21	0.02	0.72	0.00	0.00	0.38	-0.23	-0.22	-0.23	0.38	A+	A+	A-	A+	-0.07	0.06	2.14	1.06	0.91	1.04
253	824172	19	5561	0.47	0.16	0.47	0.30	0.07	0.00	0.00	0.33	-0.20	0.33	-0.11	-0.15	A-	A-	A-	A-	1.27	0.05	5.41	1.11	6.97	1.23
254	824219	19	5561	0.40	0.20	0.40	0.22	0.18	0.00	0.00	0.12	0.01	0.12	-0.14	-0.01	A-	A-	A-	A-	1.59	0.05	9.90	1.32	9.90	1.62
255	824169	19	5561	0.68	0.68	0.05	0.15	0.12	0.00	0.00	0.51	0.51	-0.30	-0.33	-0.16	A+	A-	A-	A+	0.17	0.05	-3.64	0.91	-3.83	0.85
256	824178	19	5561	0.35	0.33	0.08	0.35	0.24	0.00	0.00	0.02	0.20	-0.28	0.02	-0.06	A+	A-	A+	A+	1.88	0.05	9.90	1.39	9.90	1.77
257	809476	19	5561	0.79	0.02	0.13	0.79	0.06	0.00	0.00	0.35	-0.28	-0.16	0.35	-0.20	B+	A-	A-	A-	-0.51	0.06	-0.62	0.98	1.77	1.11
258	809475	19	5561	0.55	0.13	0.20	0.12	0.55	0.00	0.00	0.40	-0.17	-0.20	-0.20	0.40	A+	A-	A-	A-	0.86	0.05	2.31	1.05	2.47	1.08
259	809470	19	5561	0.81	0.03	0.81	0.11	0.05	0.00	0.00	0.45	-0.30	0.45	-0.26	-0.19	B-	C-	C-	A+	-0.64	0.06	-3.67	0.87	-2.69	0.83
260	809481	19	5561	0.78	0.78	0.08	0.10	0.05	0.00	0.00	0.48	0.48	-0.29	-0.19	-0.29	A+	B-	B-	A-	-0.42	0.06	-4.64	0.86	-4.47	0.76
261	809477	19	5561	0.73	0.73	0.13	0.05	0.08	0.00	0.00	0.51	0.51	-0.22	-0.33	-0.28	A+	A-	A-	A-	-0.12	0.06	-5.12	0.86	-4.72	0.79
262	809471	19	5561	0.85	0.07	0.85	0.03	0.05	0.00	0.00	0.46	-0.27	0.46	-0.29	-0.20	A+	A-	A-	A+	-1.04	0.07	-4.11	0.83	-2.64	0.80
263	824174	20	5550	0.95	0.95	0.01	0.01	0.02	0.00	0.00	0.33	0.33	-0.16	-0.20	-0.20	B+	A-	C-	A-	-2.51	0.11	-0.79	0.93	-3.31	0.53
264	824170	20	5550	0.68	0.68	0.15	0.10	0.07	0.00	0.00	0.39	0.39	-0.13	-0.29	-0.19	A+	A+	A+	A-	0.19	0.05	1.87	1.05	1.13	1.05
265	824168	20	5550	0.65	0.10	0.65	0.20	0.05	0.00	0.00	0.43	-0.29	0.43	-0.14	-0.28	A+	A-	A-	A-	0.34	0.05	-0.02	1.00	-0.01	1.00
266	824175	20	5550	0.88	0.05	0.03	0.88	0.04	0.00	0.00	0.49	-0.25	-0.28	0.49	-0.27	A+	A-	A-	B-	-1.40	0.08	-2.53	0.88	-3.83	0.66
267	824177	20	5550	0.80	0.04	0.13	0.80	0.03	0.00	0.00	0.46	-0.29	-0.27	0.46	-0.22	A+	A-	A-	A-	-0.63	0.06	-3.01	0.90	-3.91	0.76
268	824171	20	5550	0.54	0.02	0.38	0.06	0.54	0.00	0.00	0.32	-0.20	-0.10	-0.33	0.32	A+	A-	A-	A-	0.91	0.05	5.89	1.12	5.29	1.17
269	809472	20	5550	0.50	0.22	0.12	0.50	0.16	0.00	0.00	0.33	-0.23	-0.25	0.33	0.04	A+	A-	A-	A-	1.12	0.05	6.31	1.13	6.61	1.22

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
270	809480	20	5550	0.74	0.74	0.08	0.05	0.12	0.00	0.00	0.51	0.51	-0.28	-0.30	-0.24	A+	B-	B-	A-	-0.21	0.06	-4.15	0.88	-4.59	0.78
271	809474	20	5550	0.61	0.22	0.13	0.04	0.61	0.00	0.00	0.46	-0.22	-0.28	-0.20	0.46	A+	A-	A-	A+	0.54	0.05	-2.12	0.95	-1.87	0.94
272	809478	20	5550	0.66	0.08	0.13	0.13	0.66	0.00	0.00	0.47	-0.26	-0.16	-0.29	0.47	A+	A-	A-	A+	0.26	0.05	-4.23	0.90	-3.87	0.85
273	809479	20	5550	0.73	0.08	0.08	0.73	0.12	0.00	0.00	0.48	-0.26	-0.27	0.48	-0.22	A+	A-	A-	A-	-0.11	0.06	-3.33	0.91	-3.76	0.83
274	809473	20	5550	0.72	0.11	0.72	0.13	0.04	0.00	0.00	0.41	-0.23	0.41	-0.17	-0.29	A-	A-	A-	A+	-0.08	0.06	-1.26	0.96	-1.33	0.94

Table J–8. Algebra I Multiple-Choice Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	674427	0	1454	0.85	0.05	0.02	0.85	0.08	0.00	0.00	0.31	-0.14	-0.16	0.31	-0.22	-1.15	0.07	-6.25	0.77	-6.46	0.67
2	674487	0	1454	0.36	0.16	0.24	0.24	0.36	0.00	0.00	0.36	-0.10	-0.21	-0.10	0.36	1.04	0.06	-3.72	0.92	-3.44	0.91
3	674474	0	1454	0.42	0.14	0.42	0.16	0.28	0.00	0.00	0.15	-0.02	0.15	-0.07	-0.10	1.25	0.06	8.98	1.23	9.18	1.32
4	674385	0	1454	0.49	0.49	0.16	0.15	0.20	0.00	0.00	0.24	0.24	-0.07	-0.18	-0.07	0.46	0.06	1.05	1.02	1.64	1.03
5	640555	0	1454	0.40	0.08	0.11	0.40	0.41	0.00	0.00	0.25	-0.10	-0.06	0.25	-0.16	0.51	0.06	0.24	1.00	0.40	1.01
6	674458	0	1454	0.48	0.18	0.24	0.10	0.48	0.00	0.00	0.37	-0.17	-0.19	-0.12	0.37	0.79	0.06	-1.76	0.97	-2.35	0.95
7	681817	0	1454	0.50	0.28	0.50	0.11	0.11	0.00	0.00	0.33	-0.20	0.33	-0.06	-0.18	0.57	0.06	-2.29	0.96	-2.17	0.95
8	674490	0	1454	0.63	0.10	0.11	0.63	0.16	0.00	0.00	0.27	-0.15	-0.17	0.27	-0.09	-0.27	0.06	0.42	1.01	0.23	1.01
9	674495	0	1454	0.70	0.70	0.08	0.16	0.05	0.00	0.00	0.29	0.29	-0.17	-0.14	-0.15	-0.22	0.06	-5.03	0.90	-4.44	0.88
10	674503	0	1454	0.35	0.18	0.35	0.21	0.26	0.00	0.00	0.15	0.06	0.15	-0.08	-0.14	1.38	0.06	5.95	1.17	6.42	1.25
11	640576	0	1454	0.41	0.33	0.16	0.10	0.41	0.00	0.00	0.42	-0.23	-0.12	-0.17	0.42	0.38	0.06	-5.82	0.91	-5.60	0.89
12	713534	0	1454	0.56	0.15	0.14	0.15	0.56	0.00	0.00	0.20	-0.10	-0.05	-0.13	0.20	-0.16	0.06	5.53	1.11	4.38	1.13
13	700839	0	1454	0.68	0.13	0.10	0.68	0.08	0.00	0.00	0.28	-0.08	-0.23	0.28	-0.12	-0.33	0.06	-2.33	0.95	-2.15	0.93
14	713394	0	1454	0.23	0.23	0.07	0.66	0.03	0.00	0.00	0.30	0.30	-0.22	-0.11	-0.09	1.47	0.06	-5.28	0.86	-4.99	0.82
15	657740	0	1454	0.46	0.32	0.46	0.11	0.11	0.00	0.00	0.30	-0.14	0.30	-0.13	-0.13	0.64	0.06	-1.37	0.98	-0.66	0.99
16	700790	0	1454	0.30	0.26	0.27	0.16	0.30	0.00	0.00	0.12	-0.01	-0.06	-0.06	0.12	1.27	0.06	2.12	1.05	3.86	1.13
17	674419	0	1454	0.54	0.16	0.16	0.54	0.13	0.00	0.00	0.16	0.01	-0.11	0.16	-0.13	0.21	0.06	4.26	1.07	4.26	1.10
18	713762	0	1454	0.60	0.60	0.16	0.19	0.05	0.00	0.00	0.28	0.28	-0.07	-0.18	-0.18	-0.05	0.06	-0.72	0.99	-1.03	0.97
19	666556	0	1454	0.73	0.73	0.09	0.11	0.07	0.00	0.00	0.29	0.29	-0.15	-0.15	-0.15	-0.47	0.06	-4.54	0.89	-2.76	0.90
20	678765	0	1454	0.58	0.22	0.09	0.58	0.11	0.00	0.00	0.12	-0.03	-0.07	0.12	-0.07	0.24	0.06	5.04	1.08	5.25	1.12
21	678772	0	1454	0.83	0.83	0.13	0.03	0.02	0.00	0.00	0.21	0.21	-0.18	-0.04	-0.10	-1.24	0.07	-0.92	0.96	-0.41	0.97
22	702462	0	1454	0.26	0.11	0.19	0.43	0.26	0.01	0.00	0.23	-0.08	-0.07	-0.09	0.23	1.31	0.06	-3.43	0.92	-2.82	0.91
23	641506	0	1454	0.75	0.02	0.03	0.19	0.75	0.00	0.00	0.25	-0.12	-0.15	-0.15	0.25	-1.02	0.07	2.51	1.09	1.77	1.09
24	702548	0	1454	0.52	0.12	0.52	0.30	0.06	0.00	0.00	0.26	-0.13	0.26	-0.12	-0.13	-0.51	0.06	9.90	1.36	9.90	1.42
25	641450	0	1454	0.54	0.20	0.13	0.54	0.12	0.00	0.00	0.22	-0.14	-0.10	0.22	-0.06	0.81	0.06	6.60	1.12	6.38	1.16
26	678806	0	1454	0.43	0.10	0.43	0.25	0.22	0.00	0.00	0.24	-0.06	0.24	-0.09	-0.14	0.77	0.06	0.81	1.01	1.54	1.04
27	678779	0	1454	0.90	0.03	0.90	0.05	0.01	0.00	0.00	0.20	-0.06	0.20	-0.16	-0.12	-1.55	0.08	-5.63	0.73	-4.70	0.69
28	678726	0	1454	0.31	0.21	0.24	0.23	0.31	0.00	0.00	0.22	-0.11	-0.04	-0.09	0.22	1.26	0.06	0.56	1.01	0.30	1.01
29	724198	0	1454	0.35	0.21	0.35	0.05	0.40	0.00	0.00	0.19	-0.18	0.19	-0.19	0.05	1.11	0.06	1.86	1.04	2.12	1.06

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	712399	0	1454	0.24	0.09	0.24	0.32	0.36	0.00	0.00	0.13	-0.12	0.13	-0.06	0.02	1.50	0.06	-0.42	0.99	1.07	1.04
31	724155	0	1454	0.41	0.27	0.14	0.41	0.18	0.00	0.00	0.15	0.01	-0.13	0.15	-0.08	0.99	0.06	5.17	1.11	5.16	1.14
32	703172	0	1454	0.30	0.11	0.30	0.27	0.32	0.00	0.00	0.26	0.05	0.26	-0.07	-0.20	1.47	0.06	0.85	1.02	1.64	1.06
33	702473	0	1454	0.34	0.14	0.08	0.34	0.45	0.00	0.00	0.25	-0.09	-0.09	0.25	-0.12	0.79	0.06	-2.30	0.96	-1.90	0.96
34	678731	0	1454	0.61	0.61	0.15	0.19	0.05	0.00	0.00	0.21	0.21	-0.13	-0.07	-0.12	0.41	0.06	2.18	1.03	2.83	1.06
35	678740	0	1454	0.59	0.59	0.17	0.15	0.08	0.00	0.00	0.24	0.24	-0.18	-0.08	-0.06	0.51	0.06	1.69	1.03	1.47	1.03
36	678809	0	1454	0.50	0.18	0.14	0.18	0.50	0.00	0.00	0.25	-0.10	-0.14	-0.10	0.25	0.12	0.06	2.51	1.04	2.22	1.05

Table J–9. Biology Multiple-Choice Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	671335	0	960	0.62	0.62	0.25	0.07	0.06	0.00	0.00	0.28	0.28	-0.17	-0.16	-0.09	0.16	0.07	-1.05	0.98	-1.35	0.95
2	741368	0	960	0.64	0.08	0.14	0.64	0.15	0.00	0.00	0.19	-0.13	-0.12	0.19	-0.04	-0.25	0.07	0.93	1.02	0.23	1.01
3	735497	0	960	0.79	0.06	0.10	0.05	0.79	0.00	0.00	0.33	-0.11	-0.24	-0.16	0.33	-0.80	0.08	-6.06	0.80	-4.41	0.74
4	674117	0	960	0.74	0.07	0.16	0.74	0.04	0.00	0.00	0.26	-0.16	-0.14	0.26	-0.14	-0.79	0.08	-1.26	0.96	-0.64	0.96
5	739663	0	960	0.80	0.01	0.03	0.80	0.16	0.00	0.00	0.27	0.03	-0.21	0.27	-0.20	-1.03	0.08	-3.23	0.87	-1.52	0.89
6	740928	0	960	0.44	0.16	0.44	0.22	0.18	0.00	0.00	0.33	-0.08	0.33	-0.21	-0.13	0.57	0.07	-1.74	0.96	-1.39	0.95
7	702100	0	960	0.54	0.31	0.09	0.07	0.54	0.00	0.00	0.19	-0.10	-0.06	-0.11	0.19	0.09	0.07	3.14	1.06	1.79	1.07
8	739691	0	960	0.54	0.54	0.06	0.20	0.20	0.00	0.00	0.19	0.19	-0.01	-0.11	-0.12	0.13	0.07	2.81	1.05	1.95	1.07
9	713983	0	960	0.49	0.49	0.24	0.12	0.15	0.00	0.00	0.31	0.31	-0.16	-0.11	-0.13	0.23	0.07	-1.29	0.98	-0.98	0.97
10	678418	0	960	0.51	0.51	0.09	0.35	0.05	0.00	0.00	0.29	0.29	-0.14	-0.17	-0.11	0.50	0.07	0.49	1.01	-0.05	1.00
11	734747	0	960	0.64	0.09	0.15	0.64	0.12	0.00	0.00	0.22	-0.14	-0.12	0.22	-0.08	-0.56	0.07	3.10	1.09	1.95	1.11
12	734755	0	960	0.44	0.23	0.44	0.14	0.19	0.00	0.00	0.21	-0.12	0.21	-0.06	-0.08	0.79	0.07	3.51	1.09	3.48	1.13
13	677989	0	960	0.50	0.10	0.31	0.50	0.08	0.00	0.00	0.41	-0.18	-0.25	0.41	-0.14	0.15	0.07	-4.92	0.91	-3.57	0.88
14	714627	0	960	0.65	0.03	0.65	0.24	0.08	0.00	0.00	0.25	-0.15	0.25	-0.13	-0.14	-0.39	0.07	-0.12	1.00	0.49	1.02
15	678539	0	960	0.40	0.40	0.18	0.29	0.13	0.00	0.00	0.23	0.23	-0.10	-0.09	-0.10	0.69	0.07	1.12	1.03	1.09	1.04
16	741370	0	960	0.44	0.18	0.18	0.44	0.21	0.00	0.00	0.25	-0.05	-0.10	0.25	-0.16	0.49	0.07	0.28	1.01	0.15	1.00
17	740931	0	960	0.72	0.72	0.11	0.08	0.09	0.00	0.00	0.36	0.36	-0.25	-0.19	-0.09	-0.98	0.08	0.67	1.03	-0.67	0.95
18	739695	0	960	0.73	0.73	0.20	0.03	0.03	0.00	0.00	0.17	0.17	-0.07	-0.09	-0.18	-0.28	0.07	-2.46	0.94	-1.46	0.93
19	714652	0	960	0.65	0.08	0.08	0.20	0.65	0.00	0.00	0.29	-0.05	-0.18	-0.20	0.29	-0.21	0.07	-2.70	0.94	-2.05	0.91
20	735307	0	960	0.50	0.20	0.50	0.19	0.11	0.00	0.00	0.35	-0.10	0.35	-0.24	-0.14	-0.20	0.07	1.95	1.04	0.38	1.02
21	713514	0	960	0.36	0.11	0.34	0.36	0.18	0.00	0.00	0.27	-0.13	-0.08	0.27	-0.12	0.65	0.07	-1.62	0.96	-1.42	0.95
22	702079	0	960	0.41	0.05	0.41	0.39	0.15	0.00	0.00	0.30	-0.14	0.30	-0.17	-0.10	0.36	0.07	-1.26	0.98	0.48	1.02
23	681052	0	960	0.55	0.10	0.55	0.16	0.19	0.00	0.00	0.30	-0.14	0.30	-0.17	-0.12	0.51	0.07	1.69	1.03	0.59	1.02
24	677984	0	960	0.28	0.23	0.28	0.24	0.25	0.00	0.00	0.21	-0.13	0.21	-0.15	0.06	0.96	0.07	-2.54	0.93	-0.68	0.97
25	607761	0	960	0.57	0.10	0.26	0.06	0.57	0.00	0.00	0.39	-0.18	-0.26	-0.11	0.39	-0.01	0.07	-4.75	0.91	-3.54	0.87
26	678972	0	960	0.55	0.13	0.55	0.17	0.16	0.00	0.00	0.26	-0.13	0.26	-0.23	0.01	0.00	0.07	0.81	1.02	0.41	1.02
27	678297	0	960	0.73	0.04	0.04	0.19	0.73	0.00	0.00	0.20	-0.09	-0.13	-0.11	0.20	-0.53	0.07	-2.41	0.93	0.45	1.02
28	681050	0	960	0.44	0.44	0.12	0.28	0.16	0.00	0.00	0.27	0.27	-0.11	-0.16	-0.07	0.56	0.07	0.64	1.01	0.65	1.02
29	742278	0	960	0.46	0.13	0.11	0.46	0.29	0.00	0.00	0.33	-0.19	-0.17	0.33	-0.10	0.25	0.07	-1.77	0.97	-1.61	0.95

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	737713	0	960	0.55	0.24	0.55	0.16	0.06	0.00	0.00	0.27	-0.05	0.27	-0.19	-0.19	0.16	0.07	-0.63	0.99	-1.10	0.96
31	741380	0	960	0.73	0.02	0.23	0.03	0.73	0.00	0.00	0.22	-0.11	-0.14	-0.15	0.22	-0.64	0.07	-1.48	0.96	-1.20	0.93
32	677882	0	960	0.49	0.49	0.23	0.09	0.18	0.00	0.00	0.35	0.35	-0.12	-0.22	-0.16	0.32	0.07	-2.81	0.95	-2.07	0.93
33	741599	0	960	0.50	0.34	0.07	0.50	0.09	0.00	0.00	0.25	-0.04	-0.23	0.25	-0.16	0.19	0.07	1.17	1.02	1.13	1.04
34	704200	0	960	0.54	0.08	0.25	0.13	0.54	0.00	0.00	0.32	-0.16	-0.16	-0.14	0.32	0.10	0.07	-1.94	0.96	-1.66	0.94
35	682196	0	960	0.65	0.65	0.12	0.16	0.07	0.00	0.00	0.34	0.34	-0.24	-0.14	-0.13	-0.34	0.07	-3.19	0.92	-2.47	0.89
36	682198	0	960	0.76	0.08	0.76	0.07	0.08	0.00	0.00	0.30	-0.13	0.30	-0.20	-0.15	-1.08	0.08	-0.49	0.98	0.07	1.00
37	741036	0	960	0.59	0.13	0.59	0.18	0.11	0.00	0.00	0.33	-0.12	0.33	-0.17	-0.19	0.11	0.07	-3.20	0.94	-2.09	0.93
38	734738	0	960	0.50	0.14	0.50	0.15	0.20	0.00	0.00	0.30	-0.05	0.30	-0.22	-0.13	0.14	0.07	-0.30	0.99	-0.76	0.97
39	702737	0	960	0.49	0.11	0.32	0.49	0.08	0.00	0.00	0.24	-0.20	-0.06	0.24	-0.11	0.25	0.07	1.06	1.02	0.57	1.02
40	643346	0	960	0.55	0.07	0.20	0.18	0.55	0.00	0.00	0.30	-0.12	-0.07	-0.22	0.30	0.31	0.07	-0.36	0.99	-0.87	0.97
41	702144	0	960	0.52	0.11	0.23	0.52	0.14	0.00	0.00	0.21	-0.12	-0.08	0.21	-0.10	0.25	0.07	2.14	1.04	3.26	1.12
42	681522	0	960	0.45	0.45	0.18	0.15	0.22	0.00	0.00	0.26	0.26	-0.13	-0.17	-0.04	0.83	0.07	3.70	1.10	3.16	1.12
43	742317	0	960	0.43	0.13	0.25	0.18	0.43	0.00	0.00	0.31	-0.13	-0.14	-0.12	0.31	0.14	0.07	0.32	1.01	-0.29	0.99
44	679258	0	960	0.54	0.22	0.54	0.14	0.09	0.00	0.00	0.23	-0.06	0.23	-0.15	-0.13	-0.10	0.07	2.37	1.05	3.80	1.16
45	737650	0	960	0.56	0.56	0.18	0.14	0.12	0.00	0.00	0.28	0.28	-0.12	-0.20	-0.07	-0.22	0.07	1.25	1.03	1.88	1.08
46	741017	0	960	0.63	0.16	0.11	0.63	0.10	0.00	0.00	0.32	-0.11	-0.21	0.32	-0.15	-0.46	0.07	-0.26	0.99	-0.86	0.96
47	643389	0	960	0.45	0.45	0.23	0.18	0.14	0.00	0.00	0.27	0.27	-0.14	-0.13	-0.08	0.10	0.07	1.90	1.04	1.16	1.04
48	742316	0	960	0.42	0.28	0.42	0.16	0.14	0.00	0.00	0.22	-0.04	0.22	-0.15	-0.09	0.41	0.07	1.32	1.03	0.77	1.03

Table J–10. Literature Multiple-Choice Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	641566	0	446	0.53	0.05	0.21	0.20	0.53	0.00	0.00	0.17	-0.16	-0.15	0.04	0.17	0.61	0.10	2.42	1.07	2.41	1.11
2	641544	0	446	0.46	0.46	0.37	0.10	0.08	0.00	0.00	0.14	0.14	0.00	-0.20	-0.02	0.79	0.10	3.17	1.10	2.65	1.12
3	641564	0	446	0.72	0.72	0.06	0.06	0.15	0.00	0.00	0.24	0.24	-0.11	-0.15	-0.14	-0.44	0.11	0.66	1.03	0.34	1.02
4	641561	0	446	0.53	0.08	0.53	0.29	0.10	0.00	0.00	0.13	-0.10	0.13	0.03	-0.17	0.86	0.10	4.07	1.13	3.84	1.18
5	641547	0	446	0.80	0.10	0.06	0.04	0.80	0.00	0.00	0.26	-0.15	-0.16	-0.09	0.26	-0.64	0.12	-1.59	0.90	-1.65	0.86
6	641562	0	446	0.46	0.46	0.04	0.02	0.48	0.00	0.00	0.13	0.13	-0.22	-0.24	0.02	0.86	0.10	3.32	1.10	3.17	1.14
7	641542	0	446	0.66	0.10	0.66	0.11	0.13	0.00	0.00	0.31	-0.18	0.31	-0.20	-0.09	-0.35	0.11	1.52	1.08	0.93	1.07
8	641543	0	446	0.59	0.18	0.03	0.59	0.19	0.01	0.00	0.21	-0.05	-0.22	0.21	-0.12	-0.18	0.11	4.01	1.19	3.11	1.21
9	641565	0	446	0.49	0.49	0.18	0.16	0.17	0.00	0.00	0.19	0.19	0.02	-0.10	-0.17	0.66	0.10	2.05	1.06	1.76	1.08
10	701003	0	446	0.46	0.19	0.25	0.10	0.46	0.00	0.00	0.12	-0.06	-0.05	-0.04	0.12	0.91	0.10	3.80	1.12	4.40	1.21
11	701000	0	446	0.86	0.03	0.86	0.08	0.03	0.00	0.00	0.20	-0.18	0.20	-0.14	-0.03	-0.99	0.13	-2.70	0.81	-1.16	0.87
12	700982	0	446	0.63	0.06	0.17	0.63	0.13	0.00	0.00	0.17	-0.18	-0.07	0.17	-0.02	0.35	0.10	1.33	1.04	1.10	1.05
13	701004	0	446	0.55	0.15	0.14	0.15	0.55	0.00	0.00	0.29	-0.18	-0.09	-0.13	0.29	0.42	0.10	-0.39	0.99	0.07	1.00
14	701006	0	446	0.55	0.15	0.55	0.09	0.21	0.00	0.00	0.26	-0.13	0.26	-0.17	-0.09	0.45	0.10	0.21	1.01	0.80	1.04
15	700981	0	446	0.50	0.20	0.50	0.08	0.22	0.00	0.00	0.22	-0.06	0.22	-0.16	-0.10	0.70	0.10	1.16	1.03	2.77	1.12
16	700999	0	446	0.34	0.12	0.28	0.26	0.34	0.00	0.00	0.16	-0.18	0.03	-0.07	0.16	1.26	0.10	0.85	1.03	2.30	1.13
17	701005	0	446	0.62	0.62	0.05	0.13	0.19	0.00	0.00	0.38	0.38	-0.30	-0.24	-0.09	-0.34	0.11	2.03	1.10	1.12	1.08
18	683429	0	446	0.63	0.08	0.63	0.27	0.02	0.00	0.00	0.34	-0.20	0.34	-0.17	-0.25	0.18	0.10	-2.20	0.92	-1.91	0.90
19	683426	0	446	0.51	0.05	0.30	0.51	0.14	0.00	0.00	0.32	-0.23	-0.14	0.32	-0.13	0.38	0.10	-0.20	0.99	-0.19	0.99
20	683425	0	446	0.48	0.35	0.48	0.13	0.04	0.00	0.00	0.20	-0.07	0.20	-0.16	-0.05	0.69	0.10	1.84	1.06	1.36	1.06
21	683430	0	446	0.74	0.04	0.13	0.08	0.74	0.00	0.00	0.40	-0.12	-0.26	-0.23	0.40	-0.42	0.11	-2.72	0.86	-2.71	0.80
22	683432	0	446	0.45	0.22	0.20	0.45	0.13	0.00	0.00	0.24	-0.08	-0.18	0.24	-0.03	0.92	0.10	0.75	1.02	0.62	1.03
23	683424	0	446	0.36	0.16	0.04	0.44	0.36	0.00	0.00	0.08	-0.08	-0.16	0.05	0.08	1.98	0.12	7.05	1.49	6.35	1.68
24	683422	0	446	0.71	0.71	0.09	0.13	0.07	0.00	0.00	0.39	0.39	-0.25	-0.12	-0.22	-0.01	0.11	-4.37	0.83	-3.36	0.82
25	683421	0	446	0.43	0.07	0.39	0.43	0.11	0.00	0.00	0.21	-0.16	-0.01	0.21	-0.18	1.28	0.10	2.73	1.10	2.83	1.16
26	673032	0	446	0.75	0.75	0.07	0.14	0.04	0.00	0.00	0.30	0.30	-0.22	-0.13	-0.14	-0.63	0.12	-0.18	0.99	-0.23	0.98
27	673031	0	446	0.49	0.49	0.15	0.22	0.13	0.00	0.00	0.22	0.22	-0.11	-0.08	-0.08	0.58	0.10	1.55	1.05	1.29	1.06
28	683404	0	446	0.85	0.06	0.85	0.04	0.05	0.00	0.00	0.38	-0.13	0.38	-0.22	-0.28	-1.52	0.15	0.44	1.04	-0.89	0.86
29	683419	0	446	0.59	0.21	0.59	0.16	0.04	0.00	0.00	0.29	-0.17	0.29	-0.08	-0.22	-0.10	0.11	2.49	1.11	1.35	1.08

Ref	ID	Form	*N*	PVal	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	673030	0	446	0.62	0.09	0.02	0.62	0.26	0.00	0.00	0.30	-0.20	-0.16	0.30	-0.14	0.04	0.10	-0.04	1.00	-0.22	0.99
31	673023	0	446	0.37	0.37	0.26	0.16	0.20	0.00	0.00	0.28	0.28	-0.02	-0.18	-0.15	1.20	0.10	-0.59	0.98	-0.54	0.97
32	673024	0	446	0.72	0.07	0.18	0.72	0.04	0.00	0.00	0.33	-0.20	-0.13	0.33	-0.25	-0.76	0.12	2.20	1.15	1.33	1.13
33	673025	0	446	0.62	0.11	0.62	0.12	0.15	0.00	0.00	0.20	-0.02	0.20	-0.20	-0.06	0.50	0.10	1.13	1.03	1.61	1.07
34	673026	0	446	0.67	0.09	0.07	0.67	0.17	0.00	0.00	0.30	-0.25	-0.22	0.30	-0.04	-0.29	0.11	0.79	1.04	0.14	1.01

Table J–11. Algebra I Constructed-Response Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	616748	0	55500	0.33	0.40	0.13	0.23	0.05	0.15	0.04	0.61	-0.55	0.04	0.14	0.18	0.45	0.93	0.00	9.90	1.42	9.90	1.43
2	700872	0	55500	0.17	0.37	0.42	0.06	0.03	0.01	0.09	0.60	-0.49	0.22	0.28	0.28	0.24	1.75	0.01	-9.90	0.72	-9.90	0.71
3	672277	0	55500	0.30	0.31	0.27	0.14	0.10	0.08	0.08	0.66	-0.53	0.00	0.14	0.28	0.46	0.68	0.00	-9.90	0.84	-9.90	0.79
4	701634	0	55500	0.46	0.07	0.28	0.37	0.18	0.06	0.03	0.56	-0.33	-0.21	-0.02	0.30	0.36	0.39	0.00	-9.90	0.93	-8.15	0.95
5	644984	0	55500	0.19	0.36	0.44	0.13	0.02	0.00	0.04	0.50	-0.39	0.10	0.30	0.25	0.15	2.16	0.01	0.67	1.00	-3.80	0.98
6	704112	0	55500	0.19	0.47	0.25	0.12	0.05	0.03	0.07	0.68	-0.58	0.15	0.30	0.32	0.36	1.28	0.01	-9.90	0.78	-9.90	0.71

Table J–12. Biology Constructed-Response Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	703365	0	41916	0.39	0.23	0.35	0.27	0.09	0.04	0.54	-0.41	-0.08	0.27	0.36	0.77	0.01	9.90	1.20	9.90	1.21
2	703537	0	41916	0.16	0.63	0.15	0.09	0.05	0.06	0.67	-0.61	0.21	0.39	0.41	1.07	0.01	-9.90	0.70	-9.90	0.58
3	703528	0	41916	0.25	0.47	0.22	0.16	0.07	0.07	0.62	-0.55	0.11	0.35	0.40	1.29	0.01	9.90	1.15	7.13	1.07
4	607815	0	41916	0.37	0.24	0.37	0.21	0.11	0.05	0.59	-0.41	-0.11	0.27	0.44	0.48	0.01	-3.75	0.98	-4.84	0.97
5	703605	0	41916	0.25	0.41	0.30	0.16	0.04	0.07	0.61	-0.53	0.15	0.37	0.32	1.27	0.01	2.93	1.02	-5.09	0.96
6	705264	0	41916	0.39	0.21	0.43	0.22	0.10	0.04	0.60	-0.38	-0.16	0.26	0.46	0.26	0.01	-9.90	0.88	-9.90	0.88

Table J–13. Literature Constructed-Response Item Statistics: Winter

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	643730	0	36551	0.36	0.25	0.36	0.28	0.05	0.04	0.64	-0.50	-0.09	0.47	0.30	1.23	0.01	-9.90	0.83	-9.90	0.82
2	643178	0	36551	0.43	0.13	0.34	0.40	0.05	0.06	0.73	-0.56	-0.20	0.52	0.32	0.68	0.01	-9.90	0.67	-9.90	0.66
3	643179	0	36551	0.40	0.15	0.37	0.29	0.09	0.09	0.72	-0.55	-0.13	0.41	0.41	0.87	0.01	-9.90	0.71	-9.90	0.70
4	644041	0	36551	0.51	0.16	0.21	0.40	0.17	0.05	0.74	-0.57	-0.22	0.28	0.50	0.56	0.01	-9.90	0.66	-9.90	0.64
5	644768	0	36551	0.51	0.05	0.33	0.47	0.09	0.05	0.66	-0.42	-0.35	0.39	0.36	0.26	0.01	-9.90	0.80	-9.90	0.79
6	644767	0	36551	0.42	0.10	0.38	0.35	0.06	0.09	0.67	-0.49	-0.19	0.45	0.34	0.77	0.01	-9.90	0.77	-9.90	0.76

Table J–14. Algebra I Constructed-Response Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	724679	0	167097	0.48	0.13	0.19	0.26	0.37	0.03	0.02	0.65	-0.39	-0.31	-0.05	0.49	0.29					1.11	0.01	-7.50	0.95	-4.80	0.96
2	730202	0	167097	0.19	0.54	0.20	0.10	0.06	0.04	0.04	0.71	-0.64	0.15	0.31	0.37	0.36					2.05	0.01	-9.90	0.88	-9.90	0.76
3	674404	0	167097	0.38	0.25	0.29	0.13	0.10	0.16	0.06	0.74	-0.56	-0.14	0.11	0.27	0.55					1.13	0.01	9.90	1.10	6.19	1.05
4	730208	0	167097	0.48	0.12	0.27	0.25	0.24	0.11	0.02	0.73	-0.42	-0.33	-0.01	0.33	0.51					0.95	0.01	0.11	1.00	-0.88	0.99
5	666526	0	167097	0.28	0.34	0.23	0.32	0.05	0.02	0.03	0.68	-0.55	-0.05	0.40	0.28	0.28					2.04	0.01	-0.04	1.00	-3.51	0.97
6	739459	0	167097	0.36	0.17	0.21	0.44	0.11	0.01	0.05	0.70	-0.52	-0.22	0.32	0.42	0.18					1.40	0.01	-9.90	0.78	-9.90	0.78
7	821557	1	2055	0.13	0.55	0.34	0.04	0.03	0.00	0.03	0.64	-0.62	0.42	0.27	0.29	0.06	A+	A-	A-	A+	3.65	0.04	-6.08	0.78	-8.38	0.68
8	730209	1	2055	0.44	0.19	0.22	0.27	0.20	0.10	0.02	0.74	-0.54	-0.21	0.07	0.39	0.41	A+	A-	A-	A+	1.15	0.03	-5.96	0.83	-6.08	0.82
9	821370	2	1983	0.19	0.43	0.39	0.11	0.03	0.01	0.02	0.59	-0.54	0.26	0.27	0.26	0.16	A+	A-	A-	A-	2.99	0.04	-0.11	1.00	-1.00	0.96
10	818199	2	1983	0.42	0.12	0.13	0.68	0.05	0.01	0.01	0.56	-0.40	-0.24	0.28	0.33	0.13	A-	C-	B-	A+	1.94	0.04	-2.59	0.91	-2.06	0.90
11	821619	3	1984	0.37	0.27	0.27	0.13	0.24	0.05	0.03	0.72	-0.59	-0.09	0.12	0.46	0.33	B+	A-	A-	A-	1.60	0.03	-1.90	0.94	-2.41	0.91
12	818209	3	1984	0.37	0.22	0.32	0.26	0.13	0.06	0.01	0.74	-0.54	-0.22	0.29	0.38	0.34	A-	A-	A-	A+	1.53	0.03	-6.21	0.82	-7.04	0.80
13	724693	4	1979	0.48	0.14	0.13	0.37	0.33	0.01	0.01	0.75	-0.52	-0.29	-0.02	0.60	0.14	A+	B-	A-	A+	1.78	0.03	-9.90	0.69	-9.78	0.70
14	795660	4	1979	0.25	0.36	0.37	0.13	0.10	0.02	0.02	0.71	-0.60	0.07	0.29	0.43	0.22	A+	A-	A-	A+	2.42	0.03	-5.37	0.83	-6.45	0.79
15	821011	5	1990	0.21	0.43	0.36	0.08	0.05	0.04	0.03	0.68	-0.61	0.18	0.31	0.28	0.33	A+	A-	A-	A-	2.23	0.03	-5.62	0.80	-6.98	0.74
16	817476	5	1990	0.23	0.40	0.32	0.18	0.07	0.01	0.01	0.67	-0.56	0.06	0.34	0.37	0.21	A-	A-	A-	A-	2.44	0.03	-5.27	0.84	-5.33	0.82
17	821540	6	1974	0.04	0.86	0.07	0.03	0.01	0.00	0.03	0.42	-0.43	0.29	0.24	0.16	0.12	A-	A+	A+	A-	4.00	0.05	-0.93	0.93	-2.40	0.68
18	795636	6	1974	0.28	0.29	0.39	0.18	0.08	0.03	0.02	0.66	-0.52	-0.04	0.33	0.33	0.27	A+	A-	A-	A-	2.14	0.03	-1.53	0.95	-2.34	0.92
19	817473	7	1972	0.28	0.16	0.58	0.12	0.07	0.02	0.04	0.55	-0.28	-0.21	0.32	0.35	0.19	A+	A-	A+	B-	2.24	0.03	1.37	1.05	0.21	1.01
20	821013	7	1972	0.17	0.54	0.28	0.10	0.04	0.02	0.02	0.60	-0.51	0.14	0.31	0.29	0.27	A-	A-	A-	A-	2.63	0.03	-1.27	0.95	-2.07	0.91
21	821569	8	1983	0.31	0.23	0.43	0.19	0.11	0.03	0.02	0.75	-0.59	-0.12	0.36	0.45	0.21	A+	A-	A-	B-	2.00	0.03	-8.02	0.76	-9.24	0.73
22	821029	8	1983	0.24	0.33	0.44	0.11	0.08	0.01	0.02	0.72	-0.64	0.17	0.32	0.39	0.18	A+	A-	A-	A+	2.64	0.03	-6.70	0.79	-7.83	0.76
23	704006	9	1994	0.19	0.55	0.24	0.13	0.06	0.02	0.01	0.66	-0.61	0.19	0.35	0.33	0.21	A-	A-	A-	A+	2.63	0.03	-1.57	0.94	-3.15	0.85
24	734693	9	1994	0.25	0.24	0.55	0.13	0.04	0.01	0.02	0.65	-0.52	0.05	0.37	0.31	0.17	A+	A-	A-	A-	2.53	0.04	-4.94	0.84	-5.91	0.82
25	795489	10	1978	0.12	0.61	0.27	0.07	0.02	0.00	0.04	0.54	-0.50	0.29	0.30	0.22	0.11	A-	A-	A+	B-	3.52	0.04	-0.27	0.99	-1.60	0.92
26	818665	10	1978	0.22	0.46	0.27	0.12	0.11	0.01	0.02	0.74	-0.69	0.18	0.32	0.45	0.16	A-	A-	A-	A-	2.62	0.03	-6.89	0.78	-8.50	0.68
27	724685	11	1985	0.29	0.32	0.32	0.19	0.10	0.04	0.02	0.73	-0.60	-0.05	0.33	0.39	0.29	A+	A-	A-	A-	1.95	0.03	-5.85	0.82	-7.37	0.77
28	739460	11	1985	0.35	0.22	0.34	0.23	0.15	0.04	0.01	0.74	-0.57	-0.19	0.29	0.42	0.28	A-	B-	A-	A+	1.67	0.03	-5.91	0.83	-6.57	0.81

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
29	818113	12	1985	0.63	0.14	0.12	0.10	0.27	0.35	0.02	0.70	-0.52	-0.24	-0.12	0.09	0.56	A-	A-	A-	A-	0.27	0.02	-3.38	0.89	0.08	1.00
30	821623	12	1985	0.37	0.21	0.19	0.45	0.11	0.02	0.02	0.63	-0.49	-0.15	0.24	0.37	0.21	B-	C-	C-	A+	1.93	0.03	-0.03	1.00	0.22	1.01
31	817475	13	1988	0.23	0.46	0.24	0.11	0.13	0.02	0.03	0.73	-0.66	0.13	0.29	0.47	0.22	A+	A-	A-	A-	2.52	0.03	-5.93	0.80	-7.42	0.69
32	817339	13	1988	0.38	0.24	0.24	0.27	0.24	0.01	0.01	0.73	-0.58	-0.18	0.20	0.53	0.14	A-	B-	B-	A-	2.11	0.03	-5.38	0.85	-4.74	0.85
33	821004	14	1984	0.20	0.45	0.33	0.14	0.06	0.01	0.02	0.56	-0.48	0.10	0.31	0.31	0.14	A+	A-	B-	B-	2.75	0.03	2.25	1.08	3.16	1.13
34	820175	14	1984	0.44	0.21	0.22	0.20	0.21	0.13	0.03	0.77	-0.63	-0.17	0.09	0.39	0.42	A-	B-	B-	A+	1.10	0.02	-6.69	0.81	-6.18	0.80
35	795492	15	1987	0.50	0.14	0.21	0.21	0.25	0.15	0.02	0.75	-0.54	-0.31	0.05	0.35	0.43	A+	B-	A-	A-	0.82	0.03	-4.97	0.85	-4.17	0.87
36	820277	15	1987	0.22	0.27	0.58	0.11	0.02	0.00	0.01	0.57	-0.50	0.19	0.32	0.21	0.02	A+	A-	A-	A-	3.48	0.04	-1.57	0.95	-2.09	0.93
37	818106	16	1970	0.11	0.58	0.36	0.02	0.01	0.00	0.02	0.62	-0.61	0.51	0.22	0.17	0.03	A-	A-	A-	A+	4.15	0.05	-5.77	0.80	-7.34	0.73
38	820313	16	1970	0.29	0.19	0.52	0.18	0.07	0.02	0.02	0.68	-0.48	-0.15	0.37	0.36	0.24	A-	B-	B-	A-	2.03	0.03	-5.35	0.83	-6.09	0.81
39	821026	17	1974	0.12	0.62	0.25	0.10	0.01	0.00	0.02	0.66	-0.64	0.36	0.43	0.21	0.07	A+	A-	A-	A-	3.85	0.04	-7.03	0.75	-8.59	0.61
40	817477	17	1974	0.13	0.55	0.33	0.07	0.01	0.01	0.03	0.51	-0.44	0.21	0.34	0.14	0.20	B-	B-	A-	A-	3.12	0.04	0.06	1.00	-0.05	1.00
41	821550	18	1980	0.50	0.09	0.23	0.34	0.23	0.09	0.01	0.77	-0.42	-0.42	-0.01	0.45	0.41	A+	B-	A-	A-	0.87	0.03	-9.90	0.71	-9.90	0.72
42	734698	18	1980	0.28	0.37	0.19	0.32	0.08	0.02	0.03	0.74	-0.67	0.04	0.39	0.39	0.22	A-	B-	A-	A+	2.29	0.03	-7.08	0.79	-7.57	0.73
43	714218	19	1975	0.10	0.71	0.15	0.07	0.03	0.00	0.03	0.59	-0.61	0.37	0.30	0.28	0.12	A-	A-	A-	A+	3.53	0.04	-3.12	0.86	-5.03	0.65
44	821620	19	1975	0.29	0.49	0.16	0.10	0.11	0.12	0.02	0.72	-0.61	-0.02	0.17	0.31	0.51	A-	A-	A+	A-	1.71	0.02	-2.38	0.91	-1.81	0.90
45	795499	20	1992	0.25	0.41	0.26	0.18	0.09	0.03	0.03	0.69	-0.59	0.03	0.34	0.38	0.26	A-	A-	A-	A-	2.27	0.03	-3.62	0.88	-4.25	0.84
46	714437	20	1992	0.42	0.23	0.25	0.19	0.20	0.12	0.02	0.69	-0.52	-0.17	0.08	0.35	0.41	A+	A-	A-	A-	1.20	0.02	1.04	1.03	0.05	1.00

Table J–15. Biology Constructed-Response Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	736837	0	142262	0.38	0.43	0.12	0.22	0.20	0.03	0.63	-0.56	0.02	0.17	0.51					1.05	0.01	9.90	1.12	9.90	1.15
2	679989	0	142262	0.38	0.31	0.26	0.23	0.15	0.04	0.60	-0.49	-0.05	0.27	0.41					0.94	0.01	9.90	1.18	9.90	1.16
3	735487	0	142262	0.40	0.24	0.31	0.29	0.10	0.05	0.60	-0.50	-0.06	0.35	0.34					0.92	0.01	9.90	1.07	9.27	1.06
4	736555	0	142262	0.41	0.31	0.22	0.28	0.15	0.03	0.75	-0.61	-0.12	0.35	0.51					0.87	0.01	-9.90	0.79	-9.90	0.77
5	736552	0	142262	0.47	0.18	0.30	0.35	0.13	0.03	0.62	-0.48	-0.16	0.29	0.39					0.50	0.01	-9.90	0.90	-9.84	0.93
6	736553	0	142262	0.38	0.30	0.22	0.19	0.17	0.09	0.70	-0.61	-0.01	0.28	0.51					0.88	0.01	-9.90	0.88	-9.90	0.85
7	808341	1	2025	0.66	0.17	0.13	0.21	0.48	0.02	0.68	-0.56	-0.18	-0.04	0.59	A+	A-	A-	A-	0.07	0.03	-4.75	0.86	-4.64	0.79
8	809283	1	2025	0.37	0.35	0.24	0.25	0.12	0.03	0.64	-0.60	0.05	0.36	0.34	A+	A-	A-	A+	1.28	0.03	-2.45	0.93	-3.61	0.88
9	810319	2	1961	0.41	0.32	0.20	0.22	0.20	0.05	0.72	-0.66	0.03	0.25	0.52	A-	A-	A-	A-	1.05	0.03	-7.02	0.80	-6.67	0.76
10	810324	2	1961	0.38	0.36	0.21	0.20	0.18	0.05	0.67	-0.59	0.00	0.27	0.48	A+	A-	A-	A-	1.18	0.03	-2.74	0.92	-3.13	0.88
11	808704	3	1923	0.32	0.27	0.54	0.06	0.10	0.02	0.60	-0.51	0.12	0.22	0.40	A+	A-	A-	A-	1.41	0.03	-3.14	0.89	-3.21	0.89
12	810325	3	1923	0.35	0.40	0.21	0.21	0.14	0.03	0.73	-0.68	0.07	0.36	0.46	A+	A+	A+	A-	1.33	0.03	-9.52	0.73	-9.21	0.66
13	808340	4	1949	0.33	0.43	0.22	0.18	0.14	0.03	0.66	-0.61	0.08	0.32	0.42	A+	A-	A-	A+	1.37	0.03	-2.51	0.92	-3.81	0.84
14	809284	4	1949	0.44	0.22	0.36	0.26	0.15	0.02	0.54	-0.36	-0.18	0.23	0.41	A+	A-	A-	A+	0.95	0.03	3.38	1.10	2.54	1.08
15	742501	5	1964	0.11	0.71	0.19	0.06	0.01	0.03	0.48	-0.48	0.34	0.28	0.14	A+	A-	A-	B+	3.07	0.04	-2.66	0.88	-3.82	0.77
16	809860	5	1964	0.63	0.10	0.21	0.32	0.35	0.02	0.61	-0.44	-0.27	0.08	0.46	A-	C-	C-	A+	0.13	0.03	-1.12	0.97	0.36	1.01
17	742503	6	1947	0.24	0.52	0.22	0.15	0.07	0.03	0.69	-0.67	0.23	0.40	0.37	A+	A-	A-	A+	1.83	0.03	-8.60	0.73	-9.04	0.62
18	819535	6	1947	0.52	0.21	0.23	0.26	0.27	0.03	0.62	-0.53	-0.11	0.18	0.44	A+	B-	B-	A+	0.51	0.03	1.88	1.06	2.98	1.11
19	812684	7	1950	0.29	0.49	0.23	0.13	0.12	0.02	0.65	-0.59	0.12	0.30	0.44	A+	A-	A-	A-	1.60	0.03	-5.09	0.84	-5.08	0.78
20	809203	7	1950	0.39	0.38	0.20	0.19	0.19	0.03	0.72	-0.68	0.07	0.30	0.48	A+	A-	A+	A-	1.17	0.03	-7.16	0.79	-6.68	0.74
21	808339	8	1956	0.17	0.60	0.23	0.11	0.02	0.03	0.54	-0.52	0.26	0.34	0.23	A+	A-	A-	A+	2.49	0.04	-2.19	0.92	-3.66	0.82
22	813408	8	1956	0.35	0.36	0.29	0.19	0.13	0.02	0.61	-0.55	0.06	0.30	0.38	B+	A-	A-	A+	1.32	0.03	-0.46	0.99	-1.76	0.94
23	810321	9	1945	0.19	0.51	0.34	0.09	0.01	0.03	0.51	-0.48	0.27	0.29	0.18	A+	A-	A-	A-	2.55	0.04	-0.73	0.97	-1.63	0.94
24	813410	9	1945	0.28	0.40	0.29	0.19	0.06	0.05	0.66	-0.63	0.20	0.39	0.31	A+	A+	A-	A-	1.76	0.03	-5.69	0.83	-6.60	0.77
25	812687	10	1938	0.29	0.46	0.24	0.17	0.10	0.03	0.64	-0.58	0.11	0.34	0.40	A-	A-	A-	A-	1.67	0.03	-4.81	0.85	-4.22	0.83
26	736550	10	1938	0.24	0.48	0.31	0.13	0.05	0.03	0.57	-0.54	0.22	0.32	0.28	A+	A-	B-	A+	2.02	0.03	-0.96	0.97	-2.61	0.90
27	813191	11	1949	0.52	0.22	0.23	0.26	0.27	0.01	0.63	-0.45	-0.22	0.13	0.51	A+	B-	A-	A-	0.60	0.03	-1.10	0.97	-0.83	0.97
28	819528	11	1949	0.39	0.32	0.26	0.25	0.13	0.02	0.60	-0.54	0.02	0.30	0.35	A-	A-	B-	A+	1.22	0.03	0.64	1.02	0.00	1.00
29	812926	12	1970	0.38	0.34	0.26	0.21	0.15	0.03	0.69	-0.62	0.03	0.33	0.43	A+	A-	A-	A+	1.20	0.03	-4.99	0.86	-5.71	0.80

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	736556	12	1970	0.23	0.45	0.31	0.15	0.03	0.04	0.43	-0.40	0.17	0.26	0.19	A+	B-	A-	B+	2.19	0.03	6.83	1.24	5.28	1.22
31	812927	13	1935	0.41	0.32	0.21	0.23	0.18	0.05	0.70	-0.63	-0.01	0.30	0.47	A+	B-	A-	A+	1.09	0.03	-4.96	0.86	-5.32	0.81
32	813409	13	1935	0.61	0.11	0.18	0.41	0.27	0.01	0.63	-0.45	-0.18	-0.02	0.53	A-	C-	C-	A+	0.20	0.03	-1.93	0.94	-1.58	0.95
33	809153	14	1947	0.31	0.40	0.27	0.20	0.09	0.03	0.65	-0.61	0.11	0.39	0.35	A-	A-	A-	A+	1.59	0.03	-4.27	0.87	-5.81	0.79
34	810563	14	1947	0.41	0.32	0.24	0.23	0.17	0.03	0.66	-0.62	0.05	0.25	0.44	A+	A-	A-	A-	1.11	0.03	-3.04	0.91	-3.68	0.87
35	809154	15	1945	0.31	0.43	0.25	0.18	0.10	0.03	0.62	-0.53	0.04	0.32	0.40	A-	A-	A-	A-	1.52	0.03	-4.28	0.87	-3.84	0.85
36	741657	15	1945	0.23	0.50	0.26	0.11	0.07	0.04	0.54	-0.50	0.17	0.29	0.32	A+	A-	A-	A+	1.87	0.03	-1.04	0.96	-1.90	0.91
37	808349	16	1945	0.19	0.63	0.16	0.12	0.05	0.03	0.50	-0.50	0.22	0.30	0.25	A+	A-	A-	A+	2.18	0.03	1.63	1.06	0.90	1.06
38	821017	16	1945	0.37	0.33	0.26	0.21	0.14	0.04	0.68	-0.61	0.03	0.34	0.42	A+	A-	A-	A+	1.27	0.03	-5.58	0.84	-6.01	0.79
39	808536	17	1946	0.51	0.16	0.21	0.49	0.12	0.01	0.68	-0.54	-0.20	0.31	0.41	A+	C-	A-	A-	0.75	0.03	-8.61	0.77	-7.69	0.78
40	737708	17	1946	0.33	0.44	0.23	0.18	0.13	0.02	0.41	-0.32	-0.03	0.16	0.33	A-	C-	A+	A+	1.37	0.03	9.90	1.44	9.90	1.50
41	812685	18	1964	0.35	0.33	0.25	0.34	0.05	0.03	0.65	-0.62	0.06	0.46	0.24	A-	A-	A-	A-	1.61	0.03	-5.18	0.86	-5.35	0.83
42	810565	18	1964	0.33	0.35	0.26	0.21	0.11	0.05	0.64	-0.62	0.12	0.36	0.34	A+	A-	A-	A+	1.41	0.03	-2.79	0.92	-4.07	0.85
43	810320	19	1933	0.22	0.55	0.20	0.16	0.05	0.03	0.66	-0.66	0.27	0.42	0.29	A-	A+	A-	A-	2.09	0.03	-6.28	0.79	-6.94	0.67
44	736549	19	1933	0.41	0.23	0.35	0.26	0.12	0.03	0.66	-0.56	-0.06	0.36	0.37	A+	A-	A-	A-	1.08	0.03	-3.09	0.91	-3.71	0.89
45	810318	20	1951	0.23	0.45	0.40	0.08	0.04	0.02	0.53	-0.44	0.15	0.33	0.27	A+	B-	A-	A+	2.02	0.04	-2.24	0.92	-3.29	0.88
46	820999	20	1951	0.35	0.35	0.30	0.21	0.11	0.02	0.54	-0.48	0.05	0.23	0.36	A+	A-	A-	A+	1.28	0.03	2.68	1.08	1.56	1.05

Table J–16. Literature Constructed-Response Item Statistics: Spring

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	735336	0	130328	0.49	0.12	0.31	0.43	0.10	0.03	0.69	-0.53	-0.24	0.42	0.34					1.00	0.01	-9.90	0.79	-9.90	0.79
2	643181	0	130328	0.44	0.26	0.21	0.33	0.15	0.04	0.75	-0.68	-0.06	0.40	0.42					0.76	0.01	9.90	1.08	4.15	1.03
3	643182	0	130328	0.52	0.10	0.23	0.45	0.14	0.06	0.74	-0.60	-0.23	0.38	0.39					0.68	0.01	-9.90	0.67	-9.90	0.68
4	683334	0	130328	0.54	0.09	0.26	0.47	0.14	0.03	0.70	-0.54	-0.29	0.36	0.36					1.01	0.01	-9.90	0.84	-9.90	0.84
5	742911	0	130328	0.50	0.09	0.26	0.52	0.07	0.05	0.77	-0.59	-0.31	0.53	0.31					0.77	0.01	-9.90	0.58	-9.90	0.57
6	742912	0	130328	0.47	0.11	0.32	0.39	0.11	0.06	0.76	-0.62	-0.20	0.45	0.37					1.32	0.01	-9.90	0.67	-9.90	0.67
7	821032	1	1935	0.48	0.20	0.26	0.34	0.17	0.03	0.70	-0.60	-0.12	0.31	0.43	A+	A-	A-	A-	1.15	0.03	-8.42	0.77	-8.98	0.74
8	824314	1	1935	0.53	0.09	0.32	0.41	0.15	0.02	0.69	-0.54	-0.26	0.33	0.38	B+	A+	A-	A-	0.85	0.03	-7.64	0.79	-7.76	0.79
9	821033	2	1929	0.48	0.12	0.37	0.36	0.11	0.02	0.70	-0.57	-0.19	0.39	0.35	A+	A-	B-	A-	1.23	0.03	-9.62	0.74	-9.78	0.74
10	824309	2	1929	0.58	0.07	0.26	0.43	0.20	0.02	0.71	-0.52	-0.32	0.25	0.44	B+	A-	A+	A-	0.70	0.03	-9.01	0.75	-9.08	0.75
11	800286	3	1921	0.48	0.12	0.36	0.39	0.11	0.02	0.70	-0.58	-0.20	0.42	0.32	C+	A-	A-	A-	1.18	0.03	-8.58	0.76	-8.70	0.76
12	826290	3	1921	0.56	0.12	0.22	0.42	0.20	0.03	0.72	-0.60	-0.22	0.28	0.43	A+	A-	A-	A-	0.81	0.03	-8.10	0.77	-8.20	0.76
13	800287	4	1921	0.53	0.10	0.31	0.41	0.15	0.02	0.72	-0.55	-0.29	0.37	0.39	B+	A-	B-	A-	0.96	0.03	-9.90	0.73	-9.90	0.73
14	826291	4	1921	0.43	0.22	0.34	0.28	0.13	0.03	0.68	-0.60	-0.05	0.36	0.36	A+	A-	B-	A-	1.44	0.03	-8.06	0.78	-8.50	0.76
15	820826	5	1940	0.50	0.08	0.35	0.46	0.08	0.02	0.68	-0.53	-0.28	0.44	0.29	A+	A-	A-	B-	1.14	0.04	-8.79	0.75	-8.66	0.75
16	826235	5	1940	0.44	0.21	0.31	0.31	0.13	0.03	0.73	-0.65	-0.06	0.40	0.38	B+	A+	A-	B-	1.41	0.03	-9.90	0.70	-9.90	0.69
17	820824	6	1934	0.55	0.08	0.25	0.50	0.13	0.02	0.75	-0.57	-0.32	0.38	0.39	A+	A-	A-	A-	0.86	0.03	-9.90	0.67	-9.90	0.67
18	826234	6	1934	0.44	0.18	0.32	0.33	0.11	0.05	0.71	-0.63	-0.07	0.40	0.36	B+	A-	A-	C-	1.41	0.03	-9.29	0.75	-9.69	0.73
19	821002	7	1939	0.49	0.12	0.31	0.45	0.09	0.02	0.70	-0.56	-0.24	0.45	0.31	C+	A-	A+	A-	1.24	0.03	-8.33	0.77	-8.10	0.77
20	824974	7	1939	0.45	0.16	0.30	0.44	0.06	0.03	0.71	-0.62	-0.13	0.49	0.26	A+	A-	A+	A-	1.54	0.03	-8.99	0.75	-8.85	0.74
21	821000	8	1930	0.55	0.08	0.28	0.51	0.12	0.02	0.68	-0.51	-0.32	0.38	0.33	B+	A-	A-	A-	0.88	0.03	-7.84	0.77	-7.65	0.77
22	824975	8	1930	0.48	0.13	0.32	0.43	0.08	0.02	0.68	-0.54	-0.21	0.43	0.30	A+	A+	A+	B-	1.32	0.03	-7.58	0.79	-7.72	0.78
23	820948	9	1948	0.48	0.12	0.36	0.38	0.11	0.02	0.69	-0.60	-0.16	0.40	0.32	C+	A+	A-	B-	1.18	0.03	-8.62	0.76	-8.51	0.77
24	800319	9	1948	0.48	0.12	0.31	0.45	0.08	0.03	0.73	-0.62	-0.19	0.48	0.29	B+	A-	A-	A-	1.31	0.03	-9.90	0.71	-9.90	0.71
25	821001	10	1936	0.42	0.20	0.35	0.35	0.07	0.02	0.67	-0.58	-0.06	0.41	0.29	A+	A-	A+	B-	1.63	0.03	-7.12	0.80	-7.26	0.79
26	800339	10	1936	0.51	0.10	0.31	0.48	0.09	0.02	0.69	-0.54	-0.26	0.42	0.31	B+	A+	A+	A-	1.19	0.03	-7.85	0.77	-7.94	0.77
27	821030	11	1947	0.51	0.12	0.31	0.40	0.14	0.02	0.62	-0.48	-0.22	0.29	0.36	A+	A+	A+	C-	1.03	0.03	-2.97	0.91	-3.17	0.91
28	824952	11	1947	0.49	0.17	0.28	0.35	0.16	0.03	0.72	-0.61	-0.15	0.35	0.40	B+	A-	A+	A-	1.18	0.03	-9.88	0.74	-9.90	0.72
29	821031	12	1947	0.54	0.12	0.28	0.40	0.18	0.02	0.70	-0.54	-0.25	0.29	0.42	B+	A-	A-	A-	0.92	0.03	-8.51	0.77	-8.74	0.76

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
30	824953	12	1947	0.52	0.12	0.28	0.42	0.15	0.02	0.72	-0.59	-0.21	0.34	0.39	B+	A-	A-	A+	1.03	0.03	-9.90	0.72	-9.90	0.72
31	824873	13	1939	0.51	0.13	0.26	0.43	0.14	0.03	0.74	-0.63	-0.20	0.39	0.38	A+	A-	A+	B-	1.11	0.03	-9.90	0.70	-9.90	0.69
32	826239	13	1939	0.52	0.11	0.29	0.43	0.13	0.02	0.71	-0.58	-0.23	0.38	0.37	A+	A+	A-	A-	1.05	0.03	-8.76	0.76	-8.90	0.75
33	824876	14	1941	0.39	0.33	0.22	0.28	0.13	0.02	0.69	-0.63	0.01	0.36	0.41	A+	B-	B-	A-	1.61	0.03	-9.90	0.71	-9.90	0.66
34	826244	14	1941	0.51	0.08	0.37	0.39	0.13	0.02	0.69	-0.54	-0.27	0.39	0.34	B+	A-	A-	A-	0.95	0.03	-7.14	0.80	-7.32	0.80
35	824307	15	1937	0.49	0.13	0.35	0.34	0.14	0.03	0.73	-0.61	-0.17	0.35	0.41	B+	A-	A-	A-	1.12	0.03	-9.90	0.71	-9.90	0.70
36	824937	15	1937	0.53	0.06	0.35	0.45	0.12	0.02	0.68	-0.51	-0.32	0.37	0.34	B+	A+	A+	A-	0.82	0.04	-7.49	0.79	-7.42	0.79
37	824306	16	1959	0.47	0.13	0.35	0.41	0.08	0.02	0.72	-0.59	-0.20	0.44	0.33	B+	A-	A-	A-	1.29	0.03	-9.90	0.72	-9.90	0.71
38	824938	16	1959	0.51	0.09	0.35	0.37	0.15	0.03	0.69	-0.54	-0.26	0.36	0.37	B+	A-	A-	A-	0.96	0.03	-7.56	0.79	-7.92	0.78
39	821616	17	1938	0.53	0.09	0.27	0.51	0.09	0.02	0.69	-0.51	-0.31	0.42	0.32	B+	A-	A-	A-	1.11	0.03	-7.20	0.79	-6.97	0.79
40	824989	17	1938	0.46	0.13	0.36	0.39	0.08	0.03	0.71	-0.61	-0.15	0.44	0.30	A+	A-	A-	A-	1.45	0.03	-9.90	0.73	-9.90	0.72
41	821613	18	1932	0.57	0.06	0.25	0.53	0.14	0.02	0.69	-0.49	-0.35	0.33	0.36	B+	A+	A-	A-	0.72	0.04	-7.62	0.77	-7.44	0.78
42	824988	18	1932	0.53	0.08	0.29	0.48	0.11	0.02	0.70	-0.53	-0.29	0.38	0.35	A+	A-	A-	A-	0.97	0.03	-8.40	0.76	-8.24	0.76
43	824936	19	1934	0.53	0.09	0.32	0.43	0.13	0.02	0.72	-0.54	-0.31	0.38	0.38	B+	A-	A-	A-	0.95	0.03	-9.90	0.73	-9.90	0.73
44	826264	19	1934	0.55	0.10	0.22	0.52	0.13	0.02	0.75	-0.60	-0.29	0.40	0.36	A+	A-	A-	A-	0.93	0.03	-9.90	0.69	-9.90	0.69
45	824935	20	1951	0.56	0.06	0.30	0.46	0.15	0.02	0.71	-0.54	-0.33	0.35	0.37	B+	A-	A-	A-	0.76	0.03	-9.29	0.74	-9.28	0.74
46	826283	20	1951	0.57	0.06	0.24	0.53	0.13	0.03	0.73	-0.55	-0.36	0.41	0.35	B+	A-	B-	A+	0.81	0.04	-9.90	0.70	-9.90	0.70

Table J–17. Algebra I Constructed-Response Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	724699	0	1454	0.37	0.15	0.45	0.18	0.18	0.03	0.00	0.51	-0.30	-0.22	0.11	0.34	0.24	1.20	0.03	-1.32	0.95	-1.53	0.94
2	681303	0	1454	0.25	0.14	0.70	0.15	0.00	0.00	0.01	0.41	-0.25	-0.09	0.31	0.15	0.15	2.43	0.04	-8.09	0.73	-8.40	0.73
3	701631	0	1454	0.17	0.44	0.43	0.10	0.01	0.00	0.02	0.55	-0.46	0.19	0.32	0.23	0.16	1.76	0.03	-9.75	0.65	-9.90	0.65
4	734692	0	1454	0.42	0.11	0.35	0.27	0.26	0.00	0.01	0.48	-0.30	-0.21	0.04	0.38	0.14	1.30	0.03	-3.88	0.88	-3.51	0.89
5	714265	0	1454	0.25	0.35	0.36	0.20	0.06	0.01	0.01	0.56	-0.43	-0.01	0.28	0.31	0.20	1.53	0.03	-8.29	0.73	-8.51	0.70
6	678819	0	1454	0.31	0.21	0.44	0.22	0.09	0.02	0.02	0.55	-0.39	-0.08	0.19	0.31	0.26	1.10	0.03	-9.09	0.71	-8.80	0.71

Table J–18. Biology Constructed-Response Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	736839	0	960	0.35	0.26	0.47	0.23	0.04	0.00	0.50	-0.32	-0.12	0.32	0.34	1.35	0.04	-5.40	0.80	-4.79	0.80
2	742581	0	960	0.32	0.30	0.43	0.21	0.04	0.01	0.32	-0.10	-0.19	0.16	0.40	0.73	0.04	1.05	1.05	0.45	1.02
3	741445	0	960	0.46	0.24	0.30	0.27	0.17	0.01	0.51	-0.36	-0.13	0.13	0.41	0.33	0.04	-3.60	0.87	-3.05	0.87
4	737711	0	960	0.30	0.36	0.39	0.20	0.04	0.01	0.42	-0.29	-0.03	0.22	0.34	1.22	0.04	-1.78	0.92	-1.37	0.93
5	703003	0	960	0.48	0.20	0.32	0.28	0.19	0.01	0.41	-0.30	-0.14	0.18	0.27	0.64	0.04	3.26	1.14	4.75	1.24
6	677890	0	960	0.29	0.34	0.45	0.12	0.05	0.02	0.49	-0.34	0.00	0.25	0.36	0.98	0.05	-0.19	0.99	-0.71	0.97

Table J–19. Literature Constructed-Response Item Statistics: Summer

Ref	ID	Form	*N*	PVal	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	*z*-Infit	MS-Infit	*z*-Outfit	MS-Outfit
1	643960	0	446	0.54	0.07	0.34	0.47	0.12	0.01	0.50	-0.40	-0.21	0.26	0.24	0.68	0.07	-1.62	0.90	-1.64	0.90
2	703918	0	446	0.40	0.14	0.52	0.31	0.02	0.01	0.57	-0.46	-0.08	0.36	0.26	1.44	0.07	-6.83	0.64	-6.89	0.64
3	703919	0	446	0.37	0.14	0.55	0.26	0.02	0.03	0.50	-0.35	-0.09	0.33	0.27	1.18	0.07	-4.69	0.74	-4.69	0.74
4	683634	0	446	0.44	0.13	0.41	0.41	0.03	0.01	0.55	-0.41	-0.19	0.42	0.18	1.19	0.06	-6.32	0.67	-6.43	0.66
5	683411	0	446	0.39	0.14	0.52	0.31	0.01	0.00	0.57	-0.43	-0.14	0.46	0.15	1.32	0.07	-8.40	0.57	-8.34	0.57
6	683414	0	446	0.42	0.08	0.54	0.31	0.03	0.02	0.52	-0.42	-0.11	0.32	0.23	1.18	0.07	-6.00	0.67	-5.90	0.68

APPENDIX K: RAW-TO-SCALE SCORE CONVERSION TABLES

Table K–1. Raw-to-Scaled Score Conversion Tables

Column Heading	Definition
Raw	Raw score
SS	Scaled score
CSEM	Conditional standard error of measurement
LCI	Lower confidence interval
UCI	Upper confidence interval

WINTER

Table K–2. Algebra I Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1208	92	1200	1300
1	1269	51	1218	1320
2	1305	36	1269	1341
3	1327	30	1297	1357
4	1343	27	1316	1370
5	1356	24	1332	1380
6	1366	22	1344	1388
7	1376	21	1355	1397
8	1384	20	1364	1404
9	1391	19	1372	1410
10	1398	18	1380	1416
11	1405	18	1387	1423
12	1411	17	1394	1428
13	1416	17	1399	1433
14	1422	16	1406	1438
15	1427	16	1411	1443
16	1432	16	1416	1448
17	1437	15	1422	1452
18	1442	15	1427	1457
19	1446	15	1431	1461
20	1450	15	1435	1465
21	1455	15	1440	1470
22	1459	14	1445	1473
23	1463	14	1449	1477
24	1467	14	1453	1481
25	1471	14	1457	1485
26	1475	14	1461	1489
27	1479	14	1465	1493
28	1483	14	1469	1497
29	1486	14	1472	1500
30	1490	14	1476	1504
31	1494	14	1480	1508
32	1497	14	1483	1511
33	1501	14	1487	1515
34	1505	14	1491	1519
35	1509	14	1495	1523
36	1512	14	1498	1526

Raw	SS	CSEM	LCI	UCI
37	1516	14	1502	1530
38	1520	14	1506	1534
39	1524	14	1510	1538
40	1528	14	1514	1542
41	1532	14	1518	1546
42	1536	14	1522	1550
43	1540	15	1525	1555
44	1545	15	1530	1560
45	1549	15	1534	1564
46	1554	16	1538	1570
47	1559	16	1543	1575
48	1564	17	1547	1581
49	1570	17	1553	1587
50	1576	18	1558	1594
51	1583	19	1564	1602
52	1591	20	1571	1611
53	1599	21	1578	1620
54	1609	23	1586	1632
55	1621	25	1596	1646
56	1635	28	1607	1663
57	1654	33	1621	1687
58	1681	41	1640	1722
59	1729	58	1671	1787
60	1800	95	1705	1800

Table K–3. Biology Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1207	92	1200	1299
1	1268	51	1217	1319
2	1304	36	1268	1340
3	1325	30	1295	1355
4	1341	26	1315	1367
5	1353	24	1329	1377
6	1363	22	1341	1385
7	1372	20	1352	1392
8	1380	19	1361	1399
9	1387	18	1369	1405
10	1394	18	1376	1412
11	1400	17	1383	1417
12	1405	16	1389	1421
13	1411	16	1395	1427
14	1416	16	1400	1432
15	1420	15	1405	1435
16	1425	15	1410	1440
17	1429	15	1414	1444
18	1434	14	1420	1448
19	1438	14	1424	1452
20	1442	14	1428	1456
21	1445	14	1431	1459
22	1449	14	1435	1463
23	1453	14	1439	1467
24	1457	13	1444	1470
25	1460	13	1447	1473
26	1464	13	1451	1477
27	1467	13	1454	1480
28	1470	13	1457	1483
29	1474	13	1461	1487
30	1477	13	1464	1490
31	1481	13	1468	1494
32	1484	13	1471	1497
33	1487	13	1474	1500
34	1490	13	1477	1503
35	1494	13	1481	1507
36	1497	13	1484	1510
37	1500	13	1487	1513

Raw	SS	CSEM	LCI	UCI
38	1504	13	1491	1517
39	1507	13	1494	1520
40	1510	13	1497	1523
41	1514	13	1501	1527
42	1517	13	1504	1530
43	1521	13	1508	1534
44	1524	13	1511	1537
45	1528	14	1514	1542
46	1532	14	1518	1546
47	1536	14	1522	1550
48	1540	14	1526	1554
49	1544	14	1530	1558
50	1548	15	1533	1563
51	1553	15	1538	1568
52	1557	15	1542	1572
53	1562	16	1546	1578
54	1567	16	1551	1583
55	1573	17	1556	1590
56	1579	18	1561	1597
57	1585	18	1567	1603
58	1592	19	1573	1611
59	1600	20	1580	1620
60	1609	22	1587	1631
61	1620	24	1596	1644
62	1632	26	1606	1658
63	1648	30	1618	1678
64	1670	36	1634	1706
65	1706	51	1655	1757
66	1767	92	1675	1800

Table K–4. Literature Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1209	92	1200	1301
1	1270	51	1219	1321
2	1307	37	1270	1344
3	1329	30	1299	1359
4	1345	27	1318	1372
5	1358	24	1334	1382
6	1368	22	1346	1390
7	1378	21	1357	1399
8	1386	20	1366	1406
9	1394	19	1375	1413
10	1401	18	1383	1419
11	1408	18	1390	1426
12	1414	17	1397	1431
13	1420	17	1403	1437
14	1425	16	1409	1441
15	1430	16	1414	1446
16	1436	16	1420	1452
17	1441	16	1425	1457
18	1445	15	1430	1460
19	1450	15	1435	1465
20	1455	15	1440	1470
21	1459	15	1444	1474
22	1464	15	1449	1479
23	1468	15	1453	1483
24	1473	15	1458	1488
25	1477	15	1462	1492
26	1482	15	1467	1497
27	1486	15	1471	1501
28	1491	15	1476	1506
29	1495	15	1480	1510
30	1500	15	1485	1515
31	1505	15	1490	1520
32	1509	16	1493	1525
33	1514	16	1498	1530
34	1519	16	1503	1535
35	1524	16	1508	1540
36	1530	16	1514	1546
37	1535	17	1518	1552

Raw	SS	CSEM	LCI	UCI
38	1541	17	1524	1558
39	1547	18	1529	1565
40	1553	18	1535	1571
41	1560	19	1541	1579
42	1567	19	1548	1586
43	1574	20	1554	1594
44	1583	21	1562	1604
45	1592	22	1570	1614
46	1602	23	1579	1625
47	1614	25	1589	1639
48	1627	27	1600	1654
49	1644	31	1613	1675
50	1667	37	1630	1704
51	1705	51	1654	1756
52	1767	92	1675	1800

SPRING

Table K–5. Algebra I Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1218	92	1200	1310
1	1279	51	1228	1330
2	1315	37	1278	1352
3	1337	30	1307	1367
4	1353	27	1326	1380
5	1366	24	1342	1390
6	1377	22	1355	1399
7	1386	21	1365	1407
8	1395	20	1375	1415
9	1402	19	1383	1421
10	1409	18	1391	1427
11	1416	18	1398	1434
12	1422	17	1405	1439
13	1428	17	1411	1445
14	1433	16	1417	1449
15	1438	16	1422	1454
16	1443	16	1427	1459
17	1448	15	1433	1463
18	1453	15	1438	1468
19	1458	15	1443	1473
20	1462	15	1447	1477
21	1466	15	1451	1481
22	1471	15	1456	1486
23	1475	15	1460	1490
24	1479	15	1464	1494
25	1484	14	1470	1498
26	1488	14	1474	1502
27	1492	14	1478	1506
28	1496	14	1482	1510
29	1500	14	1486	1514
30	1504	14	1490	1518
31	1508	14	1494	1522
32	1513	14	1499	1527
33	1517	14	1503	1531
34	1521	14	1507	1535
35	1525	15	1510	1540
36	1529	15	1514	1544

Raw	SS	CSEM	LCI	UCI
37	1534	15	1519	1549
38	1538	15	1523	1553
39	1542	15	1527	1557
40	1547	15	1532	1562
41	1552	15	1537	1567
42	1556	16	1540	1572
43	1561	16	1545	1577
44	1567	16	1551	1583
45	1572	17	1555	1589
46	1578	17	1561	1595
47	1584	18	1566	1602
48	1590	18	1572	1608
49	1597	19	1578	1616
50	1605	20	1585	1625
51	1613	21	1592	1634
52	1622	22	1600	1644
53	1632	23	1609	1655
54	1642	24	1618	1666
55	1655	26	1629	1681
56	1669	28	1641	1697
57	1687	32	1655	1719
58	1710	38	1672	1748
59	1748	52	1696	1800
60	1800	83	1719	1800

Table K–6. Biology Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1216	92	1200	1308
1	1277	51	1226	1328
2	1312	36	1276	1348
3	1334	30	1304	1364
4	1349	26	1323	1375
5	1362	24	1338	1386
6	1372	22	1350	1394
7	1381	20	1361	1401
8	1388	19	1369	1407
9	1395	18	1377	1413
10	1402	18	1384	1420
11	1408	17	1391	1425
12	1413	16	1397	1429
13	1419	16	1403	1435
14	1424	15	1409	1439
15	1428	15	1413	1443
16	1433	15	1418	1448
17	1437	14	1423	1451
18	1441	14	1427	1455
19	1445	14	1431	1459
20	1449	14	1435	1463
21	1453	14	1439	1467
22	1456	13	1443	1469
23	1460	13	1447	1473
24	1463	13	1450	1476
25	1467	13	1454	1480
26	1470	13	1457	1483
27	1474	13	1461	1487
28	1477	13	1464	1490
29	1480	13	1467	1493
30	1483	13	1470	1496
31	1487	13	1474	1500
32	1490	13	1477	1503
33	1493	13	1480	1506
34	1496	13	1483	1509
35	1499	13	1486	1512
36	1502	13	1489	1515
37	1505	13	1492	1518

Raw	SS	CSEM	LCI	UCI
38	1509	13	1496	1522
39	1512	13	1499	1525
40	1515	13	1502	1528
41	1518	13	1505	1531
42	1522	13	1509	1535
43	1525	13	1512	1538
44	1528	13	1515	1541
45	1532	13	1519	1545
46	1535	14	1521	1549
47	1539	14	1525	1553
48	1543	14	1529	1557
49	1547	14	1533	1561
50	1551	15	1536	1566
51	1555	15	1540	1570
52	1560	15	1545	1575
53	1565	16	1549	1581
54	1570	16	1554	1586
55	1575	17	1558	1592
56	1581	17	1564	1598
57	1588	18	1570	1606
58	1595	19	1576	1614
59	1602	20	1582	1622
60	1611	22	1589	1633
61	1622	24	1598	1646
62	1634	26	1608	1660
63	1650	30	1620	1680
64	1671	36	1635	1707
65	1707	51	1656	1758
66	1768	92	1676	1800

Table K–7. Literature Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1207	92	1200	1299
1	1268	51	1217	1319
2	1305	37	1268	1342
3	1327	30	1297	1357
4	1343	27	1316	1370
5	1356	24	1332	1380
6	1367	23	1344	1390
7	1377	21	1356	1398
8	1385	20	1365	1405
9	1393	19	1374	1412
10	1400	19	1381	1419
11	1407	18	1389	1425
12	1413	17	1396	1430
13	1419	17	1402	1436
14	1425	17	1408	1442
15	1430	16	1414	1446
16	1435	16	1419	1451
17	1440	16	1424	1456
18	1445	16	1429	1461
19	1450	15	1435	1465
20	1455	15	1440	1470
21	1460	15	1445	1475
22	1464	15	1449	1479
23	1469	15	1454	1484
24	1473	15	1458	1488
25	1478	15	1463	1493
26	1482	15	1467	1497
27	1487	15	1472	1502
28	1492	15	1477	1507
29	1496	15	1481	1511
30	1501	15	1486	1516
31	1506	16	1490	1522
32	1511	16	1495	1527
33	1516	16	1500	1532
34	1521	16	1505	1537
35	1526	16	1510	1542
36	1531	17	1514	1548
37	1537	17	1520	1554

Raw	SS	CSEM	LCI	UCI
38	1543	17	1526	1560
39	1549	18	1531	1567
40	1556	18	1538	1574
41	1562	19	1543	1581
42	1570	19	1551	1589
43	1578	20	1558	1598
44	1586	21	1565	1607
45	1595	22	1573	1617
46	1606	24	1582	1630
47	1618	25	1593	1643
48	1632	28	1604	1660
49	1649	31	1618	1680
50	1672	37	1635	1709
51	1710	51	1659	1761
52	1772	92	1680	1800

SUMMER

Table K–8. Algebra Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1215	92	1200	1307
1	1277	51	1226	1328
2	1313	37	1276	1350
3	1336	31	1305	1367
4	1352	27	1325	1379
5	1365	25	1340	1390
6	1377	23	1354	1400
7	1386	21	1365	1407
8	1395	20	1375	1415
9	1403	19	1384	1422
10	1410	19	1391	1429
11	1417	18	1399	1435
12	1423	18	1405	1441
13	1429	17	1412	1446
14	1435	17	1418	1452
15	1441	16	1425	1457
16	1446	16	1430	1462
17	1451	16	1435	1467
18	1456	15	1441	1471
19	1460	15	1445	1475
20	1465	15	1450	1480
21	1470	15	1455	1485
22	1474	15	1459	1489
23	1478	15	1463	1493
24	1482	14	1468	1496
25	1487	14	1473	1501
26	1491	14	1477	1505
27	1495	14	1481	1509
28	1499	14	1485	1513
29	1503	14	1489	1517
30	1507	14	1493	1521
31	1511	14	1497	1525
32	1514	14	1500	1528
33	1518	14	1504	1532
34	1522	14	1508	1536
35	1526	14	1512	1540
36	1531	14	1517	1545

Raw	SS	CSEM	LCI	UCI
37	1535	14	1521	1549
38	1539	15	1524	1554
39	1543	15	1528	1558
40	1548	15	1533	1563
41	1552	15	1537	1567
42	1557	15	1542	1572
43	1562	16	1546	1578
44	1567	16	1551	1583
45	1572	17	1555	1589
46	1578	17	1561	1595
47	1584	17	1567	1601
48	1590	18	1572	1608
49	1597	19	1578	1616
50	1604	19	1585	1623
51	1612	20	1592	1632
52	1621	21	1600	1642
53	1630	23	1607	1653
54	1641	24	1617	1665
55	1653	26	1627	1679
56	1667	28	1639	1695
57	1684	31	1653	1715
58	1707	36	1671	1743
59	1743	51	1692	1794
60	1800	88	1712	1800

Table K–9. Biology Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1215	92	1200	1307
1	1276	51	1225	1327
2	1311	36	1275	1347
3	1333	30	1303	1363
4	1348	26	1322	1374
5	1361	24	1337	1385
6	1371	22	1349	1393
7	1380	20	1360	1400
8	1388	19	1369	1407
9	1395	18	1377	1413
10	1401	18	1383	1419
11	1407	17	1390	1424
12	1413	16	1397	1429
13	1418	16	1402	1434
14	1423	16	1407	1439
15	1428	15	1413	1443
16	1432	15	1417	1447
17	1437	15	1422	1452
18	1441	14	1427	1455
19	1445	14	1431	1459
20	1449	14	1435	1463
21	1452	14	1438	1466
22	1456	14	1442	1470
23	1460	13	1447	1473
24	1463	13	1450	1476
25	1467	13	1454	1480
26	1470	13	1457	1483
27	1474	13	1461	1487
28	1477	13	1464	1490
29	1480	13	1467	1493
30	1484	13	1471	1497
31	1487	13	1474	1500
32	1490	13	1477	1503
33	1493	13	1480	1506
34	1496	13	1483	1509
35	1500	13	1487	1513
36	1503	13	1490	1516
37	1506	13	1493	1519

Raw	SS	CSEM	LCI	UCI
38	1509	13	1496	1522
39	1513	13	1500	1526
40	1516	13	1503	1529
41	1519	13	1506	1532
42	1523	13	1510	1536
43	1526	13	1513	1539
44	1530	13	1517	1543
45	1533	13	1520	1546
46	1537	14	1523	1551
47	1541	14	1527	1555
48	1545	14	1531	1559
49	1549	14	1535	1563
50	1553	15	1538	1568
51	1557	15	1542	1572
52	1562	15	1547	1577
53	1566	16	1550	1582
54	1571	16	1555	1587
55	1577	17	1560	1594
56	1583	17	1566	1600
57	1589	18	1571	1607
58	1596	19	1577	1615
59	1604	20	1584	1624
60	1612	22	1590	1634
61	1622	23	1599	1645
62	1635	26	1609	1661
63	1650	30	1620	1680
64	1671	36	1635	1707
65	1707	51	1656	1758
66	1768	92	1676	1800

Table K–10. Literature Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1210	92	1200	1302
1	1271	51	1220	1322
2	1308	37	1271	1345
3	1330	30	1300	1360
4	1346	27	1319	1373
5	1359	24	1335	1383
6	1370	23	1347	1393
7	1379	21	1358	1400
8	1388	20	1368	1408
9	1396	19	1377	1415
10	1403	19	1384	1422
11	1410	18	1392	1428
12	1416	18	1398	1434
13	1422	17	1405	1439
14	1428	17	1411	1445
15	1433	16	1417	1449
16	1438	16	1422	1454
17	1443	16	1427	1459
18	1449	16	1433	1465
19	1453	16	1437	1469
20	1458	16	1442	1474
21	1463	15	1448	1478
22	1468	15	1453	1483
23	1472	15	1457	1487
24	1477	15	1462	1492
25	1482	15	1467	1497
26	1487	15	1472	1502
27	1491	15	1476	1506
28	1496	15	1481	1511
29	1501	16	1485	1517
30	1506	16	1490	1522
31	1511	16	1495	1527
32	1516	16	1500	1532
33	1521	16	1505	1537
34	1526	16	1510	1542
35	1531	17	1514	1548
36	1537	17	1520	1554
37	1543	17	1526	1560

Raw	SS	CSEM	LCI	UCI
38	1549	18	1531	1567
39	1555	18	1537	1573
40	1562	19	1543	1581
41	1569	19	1550	1588
42	1577	20	1557	1597
43	1585	21	1564	1606
44	1594	22	1572	1616
45	1604	23	1581	1627
46	1615	24	1591	1639
47	1628	26	1602	1654
48	1643	28	1615	1671
49	1661	32	1629	1693
50	1685	38	1647	1723
51	1723	52	1671	1775
52	1786	92	1694	1800

APPENDIX L: POST-EQUATING CHECK ANALYSES RESULTS

ITEM LEVEL

Table L–1. Evaluation of Algebra I Item Difficulty Stability: Winter

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
640543	MC	-9	9,293	55,510	-1.07	0.06	-1.06	0.01	-1.01	0.01
640528	MC	-11	6,438	55,510	0.55	0.05	0.31	0.01	0.36	-0.19
703971	MC	-7	5,003	55,510	0.23	0.05	0.27	0.01	0.32	0.09
674501	MC	-6	6,155	55,510	0.68	0.05	0.72	0.01	0.76	0.09
605092	MC	-7	4,207	55,510	0.78	0.05	0.63	0.01	0.67	-0.10
640530	MC	-6	10,082	55,510	-0.89	0.06	-0.96	0.01	-0.92	-0.09
640563	MC	-3	9,186	55,510	0.40	0.05	0.29	0.01	0.33	-0.07
704018	MC	-2	5,021	55,510	-0.35	0.05	-0.36	0.01	-0.32	0.00
700861	MC	-2	4,978	55,510	0.05	0.05	0.09	0.01	0.13	0.07
700819	MC	5	4,964	55,510	-0.07	0.05	-0.44	0.01	-0.40	-0.36
700855	MC	5	4,978	55,510	-0.09	0.05	0.04	0.01	0.08	0.16
640597	MC	9	6,471	55,510	-0.78	0.06	-0.62	0.01	-0.58	0.17
640540	MC	9	6,493	55,510	0.19	0.05	-0.05	0.01	-0.01	-0.21
700818	MC	10	5,021	55,510	1.00	0.05	0.69	0.01	0.73	-0.25
640579	MC	9	9,211	55,510	1.41	0.05	1.14	0.01	1.18	-0.20
704019	MC	11	5,026	55,510	0.27	0.05	0.40	0.01	0.44	0.17
674514	MC	11	4,964	55,510	-0.20	0.05	0.26	0.01	0.31	0.50
700870	MC	14	4,964	55,510	-1.08	0.06	-0.84	0.01	-0.80	0.24
704022	MC	-10	5,016	55,510	-0.93	0.06	-0.74	0.01	-0.70	0.19
704003	MC	-11	5,082	55,510	-0.79	0.06	-0.92	0.01	-0.88	-0.13
641466	MC	-8	6,494	55,510	-0.35	0.05	-0.12	0.01	-0.08	0.26
641477	MC	-7	9,194	55,510	0.67	0.05	0.36	0.01	0.40	-0.26
605102	MC	-9	6,472	55,510	0.20	0.05	-0.09	0.01	-0.05	-0.26
696810	MC	-7	5,027	55,510	0.04	0.05	-0.10	0.01	-0.05	-0.11
702529	MC	-3	5,026	55,510	1.27	0.05	0.86	0.01	0.90	-0.34
702528	MC	-5	5,003	55,510	-0.16	0.05	-0.31	0.01	-0.27	-0.14
641437	MC	-5	6,540	55,510	0.00	0.05	-0.31	0.01	-0.27	-0.29
641502	MC	7	9,226	55,510	0.66	0.05	0.50	0.01	0.55	-0.10
703979	MC	4	5,016	55,510	-0.39	0.05	-0.25	0.01	-0.21	0.16
657752	MC	7	4,991	55,510	-0.47	0.05	-0.46	0.01	-0.42	0.02
702537	MC	9	5,082	55,510	-0.27	0.05	-0.09	0.01	-0.05	0.21
678791	MC	9	6,152	55,510	-0.12	0.06	-0.23	0.01	-0.19	-0.09
703976	MC	9	5,003	55,510	0.19	0.05	0.10	0.01	0.14	-0.05
641523	MC	11	6,488	55,510	-0.05	0.05	-0.26	0.01	-0.22	-0.19

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
703345	MC	14	5,021	55,510	1.37	0.05	1.39	0.01	1.43	0.08
641481	MC	15	9,271	55,510	-0.59	0.05	0.18	0.01	0.22	0.80
616748	CR	-1	1,939	55,510	0.93	0.02	0.55	0.00	0.59	-0.34
700872	CR	10	2,010	55,510	1.75	0.03	1.81	0.01	1.85	0.05
672277	CR	11	1,675	55,510	0.68	0.02	0.76	0.00	0.80	0.10
701634	CR	-1	1,999	55,510	0.39	0.03	0.21	0.01	0.25	-0.17
644984	CR	10	2,003	55,510	2.16	0.03	2.15	0.01	2.19	0.41
704112	CR	11	1,993	55,510	1.28	0.03	1.41	0.01	1.46	0.32

Table L–2. Evaluation of Biology Item Difficulty Stability: Winter

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
643379	MC	-21	8,058	41,924	-1.33	0.06	-1.31	0.01	-1.42	-0.13
703485	MC	-13	4,730	41,924	-0.08	0.05	0.35	0.01	0.23	0.31
703515	MC	-15	4,773	41,924	0.65	0.05	0.51	0.01	0.39	-0.26
703358	MC	-11	4,837	41,924	0.16	0.05	0.33	0.01	0.21	0.05
680559	MC	-14	4,039	41,924	-0.11	0.06	-0.23	0.01	-0.35	-0.26
642841	MC	-13	5,347	41,924	-0.38	0.05	-0.27	0.01	-0.39	-0.02
703516	MC	-13	4,754	41,924	0.97	0.05	0.67	0.01	0.55	-0.41
678540	MC	-10	5,809	41,924	0.17	0.05	0.00	0.01	-0.12	-0.30
642320	MC	-11	8,109	41,924	-0.27	0.05	-0.28	0.01	-0.40	-0.14
643772	MC	-12	5,353	41,924	-0.70	0.05	-0.66	0.01	-0.78	-0.10
702614	MC	-2	4,754	41,924	-0.11	0.05	-0.38	0.01	-0.50	-0.41
703268	MC	-2	4,754	41,924	0.16	0.05	0.38	0.01	0.26	0.10
641270	MC	0	5,347	41,924	-0.23	0.05	-0.10	0.01	-0.22	0.00
680194	MC	-1	4,751	41,924	-0.54	0.05	-0.18	0.01	-0.30	0.24
703483	MC	1	4,751	41,924	0.17	0.05	0.02	0.01	-0.10	-0.28
702072	MC	2	4,775	41,924	0.51	0.05	0.80	0.01	0.68	0.18
703493	MC	2	4,769	41,924	-0.71	0.06	-0.31	0.01	-0.43	0.27
642861	MC	1	5,341	41,924	-0.09	0.05	0.11	0.01	-0.01	0.08
642314	MC	-1	5,364	41,924	-0.34	0.05	-0.13	0.01	-0.25	0.08
643411	MC	1	8,062	41,924	0.52	0.05	0.50	0.01	0.38	-0.14
642850	MC	8	5,378	41,924	-0.07	0.05	-0.12	0.01	-0.24	-0.18
702208	MC	5	4,809	41,924	-0.23	0.05	-0.27	0.01	-0.39	-0.18
673866	MC	6	4,769	41,924	0.51	0.05	0.53	0.01	0.41	-0.10
702070	MC	6	4,776	41,924	0.53	0.05	0.70	0.01	0.58	0.06
674147	MC	-25	4,823	41,924	-0.28	0.05	-0.13	0.01	-0.25	0.02
705228	MC	-26	4,826	41,924	-0.25	0.05	-0.11	0.01	-0.23	0.01
679253	MC	-26	4,756	41,924	-0.32	0.05	-0.51	0.01	-0.63	-0.34
641191	MC	-24	4,039	41,924	-0.10	0.06	0.11	0.01	-0.01	0.09
703153	MC	-20	4,788	41,924	0.71	0.05	0.88	0.01	0.77	0.06
678892	MC	-23	3,974	41,924	-1.31	0.07	-1.19	0.01	-1.31	-0.04
642371	MC	-22	5,360	41,924	-0.18	0.05	0.02	0.01	-0.10	0.07
641265	MC	-17	8,149	41,924	-0.09	0.05	-0.01	0.01	-0.13	-0.04
704198	MC	-15	4,754	41,924	-0.34	0.05	-0.28	0.01	-0.40	-0.08
678968	MC	-15	5,869	41,924	-0.76	0.06	-0.56	0.01	-0.68	0.07
643343	MC	-12	8,149	41,924	-1.16	0.06	-1.05	0.01	-1.17	-0.04
610863	MC	-11	8,901	41,924	0.17	0.05	0.19	0.01	0.08	-0.10
704202	MC	-10	4,742	41,924	-0.41	0.05	-0.19	0.01	-0.31	0.09

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
641202	MC	-8	5,378	41,924	-0.84	0.06	-0.29	0.01	-0.41	0.43
679206	MC	-7	4,759	41,924	-0.09	0.05	0.09	0.01	-0.03	0.06
642860	MC	-11	8,065	41,924	0.08	0.05	0.18	0.01	0.06	-0.03
679964	MC	-10	4,826	41,924	0.29	0.05	0.27	0.01	0.15	-0.13
704786	MC	-6	4,776	41,924	-0.72	0.06	-0.25	0.01	-0.37	0.34
703154	MC	-5	4,769	41,924	0.28	0.05	0.48	0.01	0.36	0.09
608250	MC	-3	5,347	41,924	-0.34	0.05	0.19	0.01	0.07	0.42
700897	MC	-5	4,730	41,924	-0.19	0.05	0.13	0.01	0.01	0.20
643401	MC	-1	4,823	41,924	0.74	0.05	0.72	0.01	0.60	-0.13
644022	MC	-6	8,129	41,924	0.10	0.05	0.45	0.01	0.33	0.24
704256	MC	-1	4,804	41,924	0.13	0.05	0.21	0.01	0.09	-0.04
703365	CR	-9	1,988	41,924	0.77	0.03	0.59	0.01	-0.74	-0.32
703537	CR	-8	1,996	41,924	1.07	0.03	1.63	0.01	1.37	0.51
703528	CR	5	1,993	41,924	1.29	0.03	1.19	0.01	0.61	-0.17
607815	CR	-14	1,970	41,924	0.48	0.03	0.57	0.01	-0.71	-0.05
703605	CR	-17	1,979	41,924	1.27	0.03	1.34	0.01	0.15	0.03
705264	CR	-4	2,011	41,924	0.26	0.03	0.53	0.01	-1.08	0.13

Table L–3. Evaluation of Literature Item Difficulty Stability: Winter

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
641600	MC	-11	7,017	36,552	1.49	0.05	1.27	0.01	1.46	0.06
641594	MC	-11	7,017	36,552	0.40	0.05	-0.05	0.01	0.14	-0.23
641611	MC	-9	7,040	36,552	0.47	0.05	0.09	0.01	0.28	-0.15
641612	MC	-10	7,017	36,552	-0.20	0.06	-0.47	0.01	-0.27	-0.07
641598	MC	-9	7,040	36,552	1.00	0.05	0.69	0.01	0.89	-0.05
641595	MC	-9	7,040	36,552	0.62	0.05	-0.02	0.01	0.18	-0.40
641599	MC	-8	7,017	36,552	0.49	0.05	0.03	0.01	0.23	-0.23
641592	MC	-8	7,040	36,552	1.47	0.05	0.71	0.01	0.91	-0.47
640877	MC	5	7,026	36,552	-0.43	0.06	-0.65	0.01	-0.45	-0.03
640876	MC	6	6,972	36,552	0.03	0.06	0.00	0.01	0.20	0.20
640879	MC	8	7,026	36,552	-0.49	0.06	-0.73	0.01	-0.53	-0.05
640863	MC	7	7,026	36,552	-1.48	0.08	-1.25	0.01	-1.06	0.40
640865	MC	7	6,972	36,552	0.86	0.05	0.53	0.01	0.72	-0.08
640881	MC	9	6,972	36,552	0.71	0.05	0.28	0.01	0.47	-0.18
640864	MC	8	7,026	36,552	-0.53	0.06	-0.30	0.01	-0.11	0.45
640878	MC	10	7,026	36,552	0.38	0.05	0.15	0.01	0.35	0.01
640860	MC	10	6,972	36,552	0.69	0.05	-0.08	0.01	0.12	-0.54
641571	MC	-9	7,040	36,552	-1.25	0.07	-1.38	0.01	-1.18	0.03
641570	MC	-8	7,017	36,552	0.49	0.05	0.27	0.01	0.46	0.02
641586	MC	-9	7,040	36,552	0.56	0.05	0.26	0.01	0.46	-0.05
641590	MC	-8	7,017	36,552	-0.91	0.07	-1.08	0.01	-0.88	0.00
641588	MC	-8	7,040	36,552	0.74	0.05	0.40	0.01	0.60	-0.08
641589	MC	-7	7,017	36,552	-0.10	0.06	-0.14	0.01	0.06	0.19
641569	MC	-7	7,017	36,552	0.49	0.05	-0.13	0.01	0.06	-0.40
641591	MC	-7	7,017	36,552	-0.46	0.06	-0.65	0.01	-0.46	0.00
641567	MC	-6	7,040	36,552	1.22	0.05	-0.06	0.01	0.14	-1.02
642702	MC	8	4,690	36,552	0.66	0.05	0.41	0.01	0.60	0.00
642704	MC	8	4,742	36,552	0.14	0.06	0.04	0.01	0.23	0.13
642703	MC	9	4,690	36,552	0.45	0.05	0.03	0.01	0.23	-0.19
642707	MC	9	4,742	36,552	0.31	0.06	0.34	0.01	0.53	0.28
642706	MC	9	4,742	36,552	1.54	0.05	0.92	0.01	1.12	-0.34
642724	MC	10	4,690	36,552	0.93	0.05	0.72	0.01	0.92	0.06
642705	MC	10	4,690	36,552	0.09	0.06	-0.01	0.01	0.19	0.13
642723	MC	11	4,742	36,552	0.00	0.06	-0.16	0.01	0.03	0.06
643730	CR	-8	1,872	36,552	1.23	0.03	1.21	0.01	1.41	0.13
643178	CR	9	1,836	36,552	0.68	0.03	0.91	0.01	1.10	0.32
643179	CR	10	1,794	36,552	0.87	0.03	0.84	0.01	1.04	0.20

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
644041	CR	-6	1,872	36,552	0.56	0.03	0.30	0.01	0.50	-0.07
644768	CR	10	1,794	36,552	0.26	0.04	0.31	0.01	0.50	0.25
644767	CR	11	1,808	36,552	0.77	0.03	0.87	0.01	1.07	0.23

Table L–4. Evaluation of Algebra I Item Difficulty Stability: Spring

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
674150	MC	-10	4,152	167,097	-1.10	0.07	-1.54	0.01	-1.06	-0.07
674468	MC	-10	5,320	167,097	1.28	0.05	0.97	0.01	1.45	0.21
700864	MC	-7	4,924	167,097	0.40	0.05	-0.19	0.01	0.29	-0.14
640618	MC	-7	6,494	167,097	0.51	0.05	-0.09	0.01	0.39	-0.13
674453	MC	-5	6,147	167,097	1.25	0.06	0.50	0.01	0.98	-0.25
640603	MC	-7	6,510	167,097	0.24	0.05	-0.26	0.01	0.22	-0.04
674391	MC	-6	4,210	167,097	-0.22	0.06	-0.71	0.01	-0.23	-0.07
700863	MC	-5	5,008	167,097	1.39	0.05	0.94	0.01	1.42	0.07
674148	MC	-4	6,155	167,097	0.61	0.05	-0.03	0.01	0.45	-0.17
674382	MC	4	6,182	167,097	0.72	0.05	0.20	0.01	0.68	-0.04
666551	MC	4	4,114	167,097	-0.86	0.07	-1.08	0.01	-0.60	0.19
674448	MC	8	6,136	167,097	-1.32	0.07	-1.24	0.01	-0.76	0.48
736783	MC	9	4,860	167,097	0.16	0.06	-0.22	0.01	0.25	0.07
700823	MC	8	5,082	167,097	1.06	0.05	0.43	0.01	0.91	-0.13
674433	MC	9	6,148	167,097	1.68	0.06	1.17	0.01	1.65	0.02
681810	MC	12	4,155	167,097	1.04	0.06	0.30	0.01	0.77	-0.26
696817	MC	10	4,876	167,097	1.29	0.05	0.75	0.01	1.23	-0.03
674446	MC	12	4,146	167,097	1.43	0.06	0.95	0.01	1.43	0.04
682039	MC	-10	4,210	167,097	-0.30	0.06	-1.11	0.01	-0.63	-0.43
678771	MC	-11	6,146	167,097	-0.89	0.06	-1.22	0.01	-0.74	0.08
702571	MC	-10	5,023	167,097	1.65	0.05	1.04	0.01	1.51	-0.08
724197	MC	-8	4,958	167,097	1.57	0.06	1.18	0.01	1.66	0.14
724194	MC	-6	4,949	167,097	0.55	0.06	-0.06	0.01	0.41	-0.15
641531	MC	-5	9,186	167,097	-0.89	0.06	-1.12	0.01	-0.64	0.17
678752	MC	-3	4,114	167,097	-0.66	0.06	-0.75	0.01	-0.27	0.35
682132	MC	-2	6,126	167,097	0.82	0.05	0.35	0.01	0.83	0.02
712563	MC	-1	4,932	167,097	1.32	0.06	1.07	0.01	1.55	0.28
724114	MC	5	4,860	167,097	1.05	0.06	0.44	0.01	0.92	-0.11
678782	MC	4	4,210	167,097	0.92	0.05	0.48	0.01	0.95	0.05
678717	MC	7	6,214	167,097	-0.49	0.06	-0.92	0.01	-0.45	-0.03
724143	MC	8	4,988	167,097	-0.30	0.06	-0.48	0.01	0.00	0.26
641436	MC	8	6,488	167,097	0.77	0.05	0.30	0.01	0.78	0.02
641453	MC	9	9,226	167,097	0.11	0.05	-0.34	0.01	0.14	0.00
724120	MC	9	4,886	167,097	0.03	0.06	-0.30	0.01	0.17	0.12
605156	MC	11	4,165	167,097	0.87	0.05	0.57	0.01	1.05	0.20
674156	MC	11	4,114	167,097	0.49	0.06	0.04	0.01	0.52	0.02
724679	CR	-1	1,474	167,097	1.11	0.03	0.56	0.00	1.04	-0.10

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
730202	CR	10	1,457	167,097	2.05	0.03	1.60	0.00	2.08	0.25
674404	CR	11	1,678	167,097	1.13	0.03	0.48	0.00	0.95	-0.15
730208	CR	-1	1,459	167,097	0.95	0.03	0.17	0.00	0.65	-0.30
666526	CR	10	1,688	167,097	2.04	0.03	1.45	0.00	1.92	0.01
739459	CR	11	1,423	167,097	1.40	0.04	1.30	0.00	1.78	0.22

Table L–5. Evaluation of Biology Item Difficulty Stability: Spring

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
677970	MC	-15	4,028	142,245	-0.35	0.06	-0.59	0.01	-0.57	-0.25
674135	MC	-15	5,843	142,245	-0.64	0.06	-0.52	0.01	-0.50	0.12
642329	MC	-13	8,058	142,245	0.60	0.05	0.72	0.01	0.73	0.15
674080	MC	-10	5,790	142,245	0.05	0.06	-0.48	0.01	-0.46	-0.54
740934	MC	-10	4,782	142,245	0.24	0.06	0.16	0.01	0.18	-0.06
674139	MC	-14	3,986	142,245	-0.35	0.06	-0.74	0.01	-0.72	-0.41
735097	MC	-9	4,799	142,245	-0.07	0.06	0.16	0.01	0.17	0.25
739696	MC	-11	4,820	142,245	-0.52	0.06	-0.74	0.01	-0.72	-0.22
735494	MC	-10	4,824	142,245	-0.19	0.06	-0.29	0.01	-0.28	-0.10
736847	MC	-9	4,822	142,245	0.15	0.06	-0.14	0.01	-0.12	-0.28
679666	MC	-2	5,872	142,245	-0.67	0.06	-0.57	0.01	-0.56	0.10
679665	MC	-1	4,039	142,245	-0.01	0.05	-0.14	0.01	-0.12	-0.12
735301	MC	6	4,820	142,245	0.49	0.06	0.41	0.01	0.42	-0.06
739676	MC	8	4,816	142,245	-0.59	0.06	-0.57	0.01	-0.55	0.02
678965	MC	10	5,845	142,245	0.79	0.06	0.82	0.01	0.84	0.06
714620	MC	11	4,830	142,245	0.12	0.06	0.21	0.01	0.23	0.11
739671	MC	11	4,830	142,245	-0.03	0.06	0.26	0.01	0.27	0.30
714626	MC	13	4,807	142,245	0.22	0.06	-0.07	0.01	-0.06	-0.29
678419	MC	11	4,752	142,245	0.74	0.06	0.66	0.01	0.68	-0.06
702101	MC	15	4,767	142,245	0.08	0.06	0.22	0.01	0.24	0.16
741369	MC	11	4,767	142,245	-0.44	0.06	-0.33	0.01	-0.32	0.11
736841	MC	15	4,782	142,245	0.42	0.06	0.35	0.01	0.36	-0.06
740016	MC	13	4,745	142,245	-0.41	0.06	-0.22	0.01	-0.21	0.20
740018	MC	16	4,782	142,245	0.11	0.06	0.24	0.01	0.26	0.16
643400	MC	-19	4,028	142,245	0.95	0.05	0.63	0.01	0.65	-0.29
739971	MC	-14	4,756	142,245	0.49	0.06	0.57	0.01	0.58	0.11
741663	MC	-14	4,791	142,245	0.20	0.06	0.05	0.01	0.06	-0.15
678303	MC	-15	3,986	142,245	1.03	0.06	0.75	0.01	0.76	-0.25
741014	MC	-12	4,822	142,245	-1.10	0.07	-0.98	0.01	-0.96	0.11
641211	MC	-12	5,872	142,245	0.97	0.05	0.83	0.01	0.84	-0.11
703161	MC	-8	4,779	142,245	-0.08	0.06	-0.16	0.01	-0.15	-0.08
683582	MC	-6	5,786	142,245	-0.75	0.06	-0.67	0.01	-0.66	0.07
683555	MC	-9	5,842	142,245	-0.07	0.06	0.18	0.01	0.19	0.27
703159	MC	-5	4,745	142,245	-0.54	0.06	-0.74	0.01	-0.72	-0.21
741291	MC	-2	4,816	142,245	0.10	0.06	0.16	0.01	0.17	0.08
736565	MC	-2	4,816	142,245	0.62	0.06	0.64	0.01	0.66	0.05
610870	MC	8	4,807	142,245	0.09	0.06	0.03	0.01	0.04	-0.05

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
677886	MC	4	4,834	142,245	0.17	0.06	0.23	0.01	0.24	0.08
607737	MC	8	4,811	142,245	0.24	0.06	0.32	0.01	0.33	0.09
742296	MC	8	4,804	142,245	0.18	0.06	0.27	0.01	0.28	0.10
642901	MC	10	4,782	142,245	0.20	0.06	-0.07	0.01	-0.05	-0.26
683552	MC	9	4,816	142,245	-0.34	0.06	-0.35	0.01	-0.34	-0.01
678986	MC	10	4,025	142,245	-0.82	0.06	-0.59	0.01	-0.58	0.23
739770	MC	15	4,816	142,245	-0.41	0.06	-0.26	0.01	-0.25	0.15
678894	MC	11	5,795	142,245	0.76	0.06	0.67	0.01	0.68	-0.07
642871	MC	14	4,025	142,245	0.66	0.05	0.63	0.01	0.65	0.00
742284	MC	13	4,782	142,245	-0.70	0.06	-0.71	0.01	-0.70	-0.02
742323	MC	17	4,779	142,245	-0.34	0.06	-0.23	0.01	-0.22	0.12
736837	CR	-1	1,645	142,245	1.05	0.03	0.89	0.00	0.91	-0.13
679989	CR	13	1,507	142,245	0.94	0.03	0.94	0.00	0.96	0.05
735487	CR	14	1,631	142,245	0.92	0.03	1.01	0.00	1.02	0.11
736555	CR	-1	1,641	142,245	0.87	0.03	0.84	0.00	0.86	0.01
736552	CR	13	1,642	142,245	0.50	0.03	0.65	0.00	0.66	0.12
736553	CR	14	1,645	142,245	0.88	0.03	0.93	0.00	0.95	0.07

Table L–6. Evaluation of Literature Item Difficulty Stability: Spring

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
734612	MC	-9	4,703	130,310	0.44	0.06	0.10	0.01	0.29	-0.12
734591	MC	-8	4,719	130,310	0.61	0.06	0.29	0.01	0.48	-0.09
734573	MC	-8	4,703	130,310	0.72	0.06	0.37	0.01	0.57	-0.11
734569	MC	-8	4,703	130,310	0.66	0.06	0.14	0.01	0.33	-0.30
734576	MC	-8	4,703	130,310	-0.06	0.06	-0.43	0.01	-0.24	-0.19
734614	MC	-7	4,719	130,310	0.91	0.06	0.47	0.01	0.66	-0.19
734615	MC	-7	4,703	130,310	0.29	0.06	-0.03	0.01	0.17	-0.10
734611	MC	-6	4,719	130,310	1.38	0.06	1.81	0.01	2.00	0.73
734575	MC	-6	4,719	130,310	1.28	0.06	0.66	0.01	0.86	-0.35
640841	MC	7	4,742	130,310	-0.56	0.06	-0.68	0.01	-0.49	0.05
640840	MC	7	4,690	130,310	0.33	0.05	-0.05	0.01	0.15	-0.16
640857	MC	8	4,742	130,310	-0.91	0.07	-0.98	0.01	-0.79	0.08
640839	MC	8	4,690	130,310	-1.03	0.07	-0.73	0.01	-0.54	0.47
640858	MC	9	4,742	130,310	0.39	0.05	0.12	0.01	0.31	-0.06
640842	MC	9	4,742	130,310	-0.19	0.06	-0.17	0.01	0.02	0.23
640859	MC	10	4,690	130,310	-0.09	0.06	-0.35	0.01	-0.15	-0.06
640836	MC	10	4,742	130,310	0.38	0.05	-0.16	0.01	0.03	-0.34
683335	MC	-10	3,854	130,310	-1.07	0.07	-1.30	0.01	-1.11	-0.10
683343	MC	-9	3,891	130,310	0.72	0.05	0.61	0.01	0.80	0.13
683461	MC	-9	3,854	130,310	0.35	0.06	0.10	0.01	0.30	-0.03
683338	MC	-9	3,891	130,310	0.25	0.06	-0.02	0.01	0.18	-0.05
683342	MC	-9	3,854	130,310	-0.22	0.06	-0.45	0.01	-0.26	-0.04
683336	MC	-9	3,891	130,310	-0.82	0.07	-1.10	0.01	-0.91	-0.13
683340	MC	-9	3,891	130,310	1.07	0.05	0.33	0.01	0.52	-0.50
683460	MC	-8	3,854	130,310	1.34	0.05	0.89	0.01	1.08	-0.18
742875	MC	6	4,648	130,310	0.02	0.06	-0.25	0.01	-0.06	-0.08
742869	MC	6	4,648	130,310	0.38	0.06	0.10	0.01	0.29	-0.06
742873	MC	6	4,684	130,310	-1.90	0.10	-1.68	0.01	-1.49	0.33
742874	MC	7	4,648	130,310	1.28	0.06	0.95	0.01	1.14	-0.06
742878	MC	7	4,684	130,310	1.37	0.06	1.27	0.01	1.46	0.17
742880	MC	8	4,648	130,310	0.02	0.06	-0.29	0.01	-0.10	-0.12
742870	MC	8	4,684	130,310	1.63	0.06	1.56	0.01	1.75	0.23
742872	MC	8	4,648	130,310	-0.48	0.07	-0.47	0.01	-0.28	0.19
742879	MC	9	4,684	130,310	-0.56	0.07	-0.62	0.01	-0.43	0.12
735336	CR	-6	1,588	130,310	1.00	0.04	0.83	0.00	1.02	-0.03
643181	CR	9	1,794	130,310	0.76	0.03	1.05	0.00	1.24	0.55
643182	CR	10	1,808	130,310	0.68	0.03	0.67	0.00	0.86	0.12

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
683334	CR	-8	1,670	130,310	1.01	0.04	0.53	0.00	0.73	-0.28
742911	CR	9	1,589	130,310	0.77	0.04	0.89	0.00	1.08	0.22
742912	CR	10	1,597	130,310	1.32	0.03	0.89	0.00	1.08	-0.23

Table L–7. Evaluation of Algebra I Item Difficulty Stability: Summer

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
674427	MC	-13	6,148	1,448	-1.15	0.07	-1.80	0.07	-1.45	-0.32
674487	MC	-12	4,205	1,448	1.04	0.06	0.68	0.06	1.03	0.03
674474	MC	-7	6,214	1,448	1.25	0.06	0.39	0.06	0.74	-0.46
674385	MC	-8	4,153	1,448	0.46	0.06	0.07	0.06	0.43	0.00
640555	MC	-9	6,540	1,448	0.51	0.05	0.47	0.06	0.83	0.35
674458	MC	-5	4,146	1,448	0.79	0.06	0.14	0.06	0.49	-0.27
681817	MC	-7	4,152	1,448	0.57	0.06	0.04	0.06	0.39	-0.15
674490	MC	-2	6,182	1,448	-0.27	0.06	-0.55	0.06	-0.20	0.08
674495	MC	-2	4,205	1,448	-0.22	0.06	-0.88	0.06	-0.53	-0.30
674503	MC	6	5,008	1,448	1.38	0.05	0.73	0.06	1.08	-0.25
640576	MC	7	6,399	1,448	0.38	0.05	0.43	0.06	0.79	0.45
713534	MC	7	4,921	1,448	-0.16	0.06	-0.22	0.06	0.13	0.32
700839	MC	10	5,042	1,448	-0.33	0.05	-0.78	0.06	-0.43	-0.09
713394	MC	7	4,922	1,448	1.47	0.06	1.35	0.07	1.70	0.29
657740	MC	8	6,155	1,448	0.64	0.05	0.20	0.06	0.55	-0.05
700790	MC	11	5,000	1,448	1.27	0.05	0.95	0.06	1.30	0.08
674419	MC	13	6,148	1,448	0.21	0.06	-0.13	0.06	0.23	0.04
713762	MC	13	4,944	1,448	-0.05	0.06	-0.39	0.06	-0.04	0.04
666556	MC	-13	4,205	1,448	-0.47	0.06	-1.04	0.06	-0.68	-0.21
678765	MC	-11	6,182	1,448	0.24	0.06	-0.32	0.06	0.04	-0.18
678772	MC	-7	4,146	1,448	-1.24	0.07	-1.63	0.07	-1.28	-0.04
702462	MC	-10	5,021	1,448	1.31	0.05	1.19	0.06	1.55	0.29
641506	MC	-6	10,082	1,448	-1.02	0.06	-1.16	0.06	-0.81	0.21
702548	MC	-5	4,971	1,448	-0.51	0.07	-0.03	0.06	0.32	0.87
641450	MC	-3	10,082	1,448	0.81	0.05	-0.14	0.06	0.22	-0.56
678806	MC	-6	4,152	1,448	0.77	0.06	0.37	0.06	0.72	-0.01
678779	MC	-2	6,126	1,448	-1.55	0.08	-2.30	0.09	-1.95	-0.44
678726	MC	3	6,146	1,448	1.26	0.06	0.88	0.06	1.24	0.03
724198	MC	6	4,971	1,448	1.11	0.06	0.72	0.06	1.08	0.01
712399	MC	9	4,938	1,448	1.50	0.06	1.31	0.06	1.66	0.21
724155	MC	10	4,957	1,448	0.99	0.06	0.44	0.06	0.79	-0.15
703172	MC	10	5,046	1,448	1.47	0.05	0.97	0.06	1.33	-0.08
702473	MC	11	4,946	1,448	0.79	0.06	0.77	0.06	1.12	0.38
678731	MC	10	4,146	1,448	0.41	0.06	-0.46	0.06	-0.10	-0.49
678740	MC	12	6,136	1,448	0.51	0.05	-0.33	0.06	0.02	-0.46
678809	MC	13	4,155	1,448	0.12	0.06	0.06	0.06	0.41	0.33
724699	CR	-1	1,423	1,448	1.20	0.03	0.65	0.03	1.00	-0.17

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
681303	CR	10	1,725	1,448	2.43	0.04	1.67	0.05	2.03	-0.27
701631	CR	11	1,462	1,448	1.76	0.03	1.88	0.04	2.23	0.36
734692	CR	-1	1,430	1,448	1.30	0.03	1.05	0.03	1.41	-0.15
714265	CR	10	1,457	1,448	1.53	0.03	1.37	0.03	1.72	0.08
678819	CR	11	1,685	1,448	1.10	0.03	0.96	0.03	1.32	0.17

Table L–8. Evaluation of Biology Item Difficulty Stability: Summer

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
671335	MC	-18	3,986	942	0.16	0.06	-0.29	0.07	-0.29	-0.42
741368	MC	-18	4,807	942	-0.25	0.06	-0.38	0.07	-0.38	-0.10
735497	MC	-16	4,787	942	-0.80	0.06	-1.17	0.08	-1.17	-0.36
674117	MC	-14	4,019	942	-0.79	0.06	-0.87	0.08	-0.86	-0.05
739663	MC	-13	4,782	942	-1.03	0.07	-1.23	0.08	-1.23	-0.18
740928	MC	-11	4,816	942	0.57	0.06	0.50	0.07	0.50	-0.04
702100	MC	-9	4,782	942	0.09	0.06	0.07	0.07	0.07	0.01
739691	MC	-12	4,776	942	0.13	0.06	0.08	0.07	0.08	-0.01
713983	MC	-6	4,791	942	0.23	0.06	0.31	0.07	0.31	0.11
678418	MC	-7	5,872	942	0.50	0.05	0.19	0.07	0.19	-0.28
734747	MC	-2	4,787	942	-0.56	0.06	-0.39	0.07	-0.39	0.21
734755	MC	-1	4,756	942	0.79	0.06	0.53	0.07	0.53	-0.22
677989	MC	8	5,790	942	0.15	0.06	0.22	0.07	0.22	0.11
714627	MC	8	4,804	942	-0.39	0.06	-0.45	0.07	-0.45	-0.03
678539	MC	6	5,798	942	0.69	0.06	0.70	0.07	0.71	0.05
741370	MC	6	4,811	942	0.49	0.06	0.55	0.07	0.55	0.10
740931	MC	12	4,823	942	-0.98	0.07	-0.80	0.08	-0.80	0.21
739695	MC	9	4,804	942	-0.28	0.06	-0.83	0.08	-0.83	-0.54
714652	MC	9	4,820	942	-0.21	0.06	-0.43	0.07	-0.43	-0.19
735307	MC	12	4,791	942	-0.20	0.06	0.24	0.07	0.24	0.48
713514	MC	16	4,804	942	0.65	0.06	0.88	0.07	0.88	0.26
702079	MC	16	4,811	942	0.36	0.06	0.64	0.07	0.64	0.32
681052	MC	19	3,990	942	0.51	0.05	0.00	0.07	0.01	-0.48
677984	MC	17	5,275	942	0.96	0.05	1.34	0.08	1.34	0.42
607761	MC	-7	46,979	942	-0.01	0.01	-0.08	0.07	-0.08	-0.04
678972	MC	-15	4,018	942	0.00	0.06	0.04	0.07	0.04	0.08
678297	MC	-16	4,807	942	-0.53	0.06	-0.83	0.08	-0.83	-0.28
681050	MC	-15	4,022	942	0.56	0.05	0.51	0.07	0.51	-0.02
742278	MC	-15	4,820	942	0.25	0.06	0.40	0.07	0.40	0.18
737713	MC	-13	4,782	942	0.16	0.06	0.04	0.07	0.04	-0.09
741380	MC	-11	4,782	942	-0.64	0.06	-0.84	0.08	-0.84	-0.18
677882	MC	-11	5,843	942	0.32	0.05	0.26	0.07	0.26	-0.02
741599	MC	-8	4,752	942	0.19	0.06	0.23	0.07	0.23	0.07
704200	MC	-10	4,811	942	0.10	0.06	0.09	0.07	0.09	0.02
682196	MC	-3	4,025	942	-0.34	0.06	-0.45	0.07	-0.45	-0.08
682198	MC	-1	4,022	942	-1.08	0.06	-1.02	0.08	-1.02	0.08
741036	MC	4	4,756	942	0.11	0.06	-0.16	0.07	-0.16	-0.25

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
734738	MC	6	4,804	942	0.14	0.06	0.22	0.07	0.22	0.12
702737	MC	5	4,776	942	0.25	0.06	0.29	0.07	0.29	0.08
643346	MC	7	5,842	942	0.31	0.05	0.03	0.07	0.03	-0.25
702144	MC	11	4,737	942	0.25	0.06	0.16	0.07	0.16	-0.06
681522	MC	8	3,990	942	0.83	0.05	0.46	0.07	0.46	-0.33
742317	MC	11	4,799	942	0.14	0.06	0.54	0.07	0.54	0.43
679258	MC	15	5,786	942	-0.10	0.05	0.07	0.07	0.07	0.20
737650	MC	11	4,737	942	-0.22	0.06	-0.04	0.07	-0.04	0.21
741017	MC	16	4,782	942	-0.46	0.06	-0.35	0.07	-0.35	0.14
643389	MC	18	4,811	942	0.10	0.06	0.43	0.07	0.43	0.37
742316	MC	18	4,816	942	0.41	0.06	0.61	0.07	0.61	0.24
736839	CR	-1	1,630	942	1.35	0.03	1.19	0.05	1.19	-0.26
742581	CR	0	1,645	942	0.73	0.03	1.25	0.05	1.25	0.32
741445	CR	13	1,656	942	0.33	0.03	0.40	0.04	0.40	0.10
737711	CR	-1	1,643	942	1.22	0.03	1.35	0.05	1.36	-0.07
703003	CR	0	1,649	942	0.64	0.03	0.30	0.04	0.30	-0.35
677890	CR	13	1,744	942	0.98	0.03	1.22	0.05	1.22	0.34

Table L–9. Evaluation of Literature Item Difficulty Stability: Summer

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
641566	MC	-10	4,744	444	0.61	0.05	0.25	0.10	0.44	-0.10
641544	MC	-9	4,714	444	0.79	0.05	0.59	0.10	0.79	0.08
641564	MC	-10	4,744	444	-0.44	0.06	-0.64	0.11	-0.44	0.04
641561	MC	-9	4,714	444	0.86	0.05	0.26	0.10	0.45	-0.33
641547	MC	-9	4,714	444	-0.64	0.06	-1.07	0.12	-0.87	-0.22
641562	MC	-8	4,744	444	0.86	0.05	0.57	0.10	0.77	-0.01
641542	MC	-8	4,714	444	-0.35	0.06	-0.33	0.11	-0.13	0.27
641543	MC	-8	4,744	444	-0.18	0.06	-0.03	0.10	0.16	0.41
641565	MC	-7	4,714	444	0.66	0.05	0.44	0.10	0.64	0.06
701003	MC	7	4,759	444	0.91	0.05	0.55	0.10	0.75	-0.07
701000	MC	8	4,773	444	-0.99	0.07	-1.58	0.14	-1.38	-0.39
700982	MC	6	4,773	444	0.35	0.05	-0.20	0.10	0.00	-0.30
701004	MC	6	4,773	444	0.42	0.05	0.17	0.10	0.36	0.02
701006	MC	8	4,759	444	0.45	0.05	0.17	0.10	0.36	-0.02
700981	MC	7	4,773	444	0.70	0.05	0.38	0.10	0.58	-0.04
700999	MC	9	4,759	444	1.26	0.05	1.11	0.11	1.31	0.14
701005	MC	9	4,759	444	-0.34	0.06	-0.14	0.10	0.06	0.46
683429	MC	-10	4,730	444	0.18	0.05	-0.21	0.10	-0.01	-0.13
683426	MC	-8	4,735	444	0.38	0.05	0.37	0.10	0.56	0.25
683425	MC	-9	4,735	444	0.69	0.05	0.46	0.10	0.66	0.05
683430	MC	-8	4,730	444	-0.42	0.06	-0.76	0.11	-0.56	-0.11
683432	MC	-8	4,735	444	0.92	0.05	0.63	0.10	0.83	-0.01
683424	MC	-7	4,730	444	1.98	0.05	1.05	0.11	1.25	-0.62
683422	MC	-7	4,735	444	-0.01	0.06	-0.59	0.11	-0.39	-0.34
683421	MC	-7	4,735	444	1.28	0.05	0.72	0.10	0.92	-0.27
673032	MC	6	5,623	444	-0.63	0.07	-0.81	0.12	-0.61	0.06
673031	MC	6	5,623	444	0.58	0.05	0.44	0.10	0.64	0.14
683404	MC	6	5,623	444	-1.52	0.08	-1.50	0.14	-1.30	0.22
683419	MC	7	5,545	444	-0.10	0.06	0.01	0.10	0.20	0.36
673030	MC	7	5,623	444	0.04	0.06	-0.15	0.10	0.05	0.07
673023	MC	7	5,545	444	1.20	0.05	0.97	0.10	1.17	0.06
673024	MC	8	5,623	444	-0.76	0.07	-0.63	0.11	-0.43	0.38
673025	MC	8	5,623	444	0.50	0.05	-0.15	0.10	0.05	-0.39
673026	MC	9	5,545	444	-0.29	0.06	-0.36	0.11	-0.16	0.17
643960	CR	-7	1,855	444	0.68	0.03	0.22	0.07	0.42	-0.24
703918	CR	9	1,992	444	1.44	0.03	1.25	0.08	1.45	-0.10
703919	CR	10	2,001	444	1.18	0.03	1.43	0.08	1.63	0.37

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
683634	CR	-7	1,995	444	1.19	0.03	1.00	0.07	1.20	-0.12
683411	CR	9	1,705	444	1.32	0.03	1.67	0.08	1.87	0.04
683414	CR	10	1,644	444	1.18	0.04	0.97	0.08	1.17	0.07

FORM LEVEL

Table L–10. Raw-to-Scaled Score Comparison for Algebra I: Winter

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1208	92	1210	92	1211	92
1	1269	51	1271	51	1272	51
2	1305	36	1307	36	1308	36
3	1327	30	1329	30	1330	30
4	1343	27	1344	26	1346	26
5	1356	24	1357	24	1358	24
6	1366	22	1368	22	1369	22
7	1376	21	1377	21	1378	21
8	1384	20	1385	20	1386	20
9	1391	19	1393	19	1394	19
10	1398	18	1400	18	1401	18
11	1405	18	1406	18	1407	18
12	1411	17	1412	17	1413	17
13	1416	17	1418	17	1419	17
14	1422	16	1423	16	1424	16
15	1427	16	1428	16	1429	16
16	1432	16	1433	16	1434	16
17	1437	15	1438	15	1439	15
18	1442	15	1442	15	1443	15
19	1446	15	1447	15	1448	15
20	1450	15	1451	15	1452	15
21	1455	15	1455	14	1456	14
22	1459	14	1459	14	1461	14
23	1463	14	1463	14	1465	14
24	1467	14	1467	14	1468	14
25	1471	14	1471	14	1472	14
26	1475	14	1475	14	1476	14
27	1479	14	1479	14	1480	14
28	1483	14	1483	14	1484	14
29	1486	14	1486	14	1487	14
30	1490	14	1490	14	1491	14
31	1494	14	1494	14	1495	14
32	1497	14	1497	14	1498	14
33	1501	14	1501	14	1502	14
34	1505	14	1505	14	1506	14

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
35	1509	14	1508	14	1509	14
36	1512	14	1512	14	1513	14
37	1516	14	1516	14	1517	14
38	1520	14	1520	14	1521	14
39	1524	14	1524	14	1525	14
40	1528	14	1528	14	1529	14
41	1532	14	1532	15	1533	15
42	1536	14	1536	15	1537	15
43	1540	15	1541	15	1542	15
44	1545	15	1545	15	1546	15
45	1549	15	1550	16	1551	16
46	1554	16	1555	16	1556	16
47	1559	16	1560	16	1561	16
48	1564	17	1566	17	1567	17
49	1570	17	1572	17	1573	17
50	1576	18	1578	18	1579	18
51	1583	19	1585	19	1586	19
52	1591	20	1592	20	1593	20
53	1599	21	1601	21	1602	21
54	1609	23	1610	22	1611	22
55	1621	25	1621	24	1622	24
56	1635	28	1634	27	1635	27
57	1654	33	1651	31	1652	31
58	1681	41	1674	38	1675	38
59	1729	58	1713	52	1714	52
60	1800	97	1776	93	1777	93

Table L–11. Raw-to-Scaled Score Comparison for Biology: Winter

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1207	92	1207	92	1208	92
1	1268	51	1268	51	1269	51
2	1304	36	1304	36	1305	36
3	1325	30	1325	30	1326	30
4	1341	26	1341	26	1342	26
5	1353	24	1353	24	1354	24
6	1363	22	1364	22	1365	22
7	1372	20	1372	20	1374	20
8	1380	19	1380	19	1381	19
9	1387	18	1387	18	1389	18
10	1394	18	1394	18	1395	18
11	1400	17	1400	17	1401	17
12	1405	16	1406	16	1407	16
13	1411	16	1411	16	1412	16
14	1416	16	1416	16	1417	16
15	1420	15	1421	15	1422	15
16	1425	15	1425	15	1426	15
17	1429	15	1430	15	1431	15
18	1434	14	1434	14	1435	14
19	1438	14	1438	14	1439	14
20	1442	14	1442	14	1443	14
21	1445	14	1446	14	1447	14
22	1449	14	1449	14	1450	14
23	1453	14	1453	14	1454	14
24	1457	13	1457	13	1458	13
25	1460	13	1460	13	1461	13
26	1464	13	1464	13	1465	13
27	1467	13	1467	13	1468	13
28	1470	13	1471	13	1472	13
29	1474	13	1474	13	1475	13
30	1477	13	1477	13	1478	13
31	1481	13	1481	13	1482	13
32	1484	13	1484	13	1485	13
33	1487	13	1487	13	1488	13
34	1490	13	1491	13	1492	13
35	1494	13	1494	13	1495	13
36	1497	13	1497	13	1498	13

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1500	13	1501	13	1502	13
38	1504	13	1504	13	1505	13
39	1507	13	1507	13	1508	13
40	1510	13	1511	13	1512	13
41	1514	13	1514	13	1515	13
42	1517	13	1518	13	1519	13
43	1521	13	1521	13	1522	13
44	1524	13	1525	13	1526	13
45	1528	14	1529	14	1530	14
46	1532	14	1532	14	1533	14
47	1536	14	1536	14	1537	14
48	1540	14	1540	14	1541	14
49	1544	14	1544	14	1545	14
50	1548	15	1548	15	1549	15
51	1553	15	1553	15	1554	15
52	1557	15	1557	15	1558	15
53	1562	16	1562	16	1563	16
54	1567	16	1567	16	1568	16
55	1573	17	1573	17	1574	17
56	1579	18	1579	17	1580	17
57	1585	18	1585	18	1586	18
58	1592	19	1592	19	1593	19
59	1600	20	1600	20	1601	20
60	1609	22	1608	22	1609	22
61	1620	24	1619	24	1620	24
62	1632	26	1631	26	1632	26
63	1648	30	1646	30	1648	30
64	1670	36	1668	36	1669	36
65	1706	51	1704	51	1705	51
66	1767	92	1765	92	1766	92

Table L–12. Raw-to-Scaled Score Comparison for Literature: Winter

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1209	92	1210	92	1209	92
1	1270	51	1271	51	1270	51
2	1307	37	1307	36	1306	36
3	1329	30	1329	30	1328	30
4	1345	27	1345	26	1344	26
5	1358	24	1357	24	1356	24
6	1368	22	1368	22	1367	22
7	1378	21	1377	21	1376	21
8	1386	20	1386	20	1385	20
9	1394	19	1393	19	1392	19
10	1401	18	1400	18	1399	18
11	1408	18	1406	18	1405	18
12	1414	17	1412	17	1411	17
13	1420	17	1418	17	1417	17
14	1425	16	1423	16	1422	16
15	1430	16	1429	16	1428	16
16	1436	16	1434	16	1433	16
17	1441	16	1439	16	1437	16
18	1445	15	1443	15	1442	15
19	1450	15	1448	15	1447	15
20	1455	15	1452	15	1451	15
21	1459	15	1457	15	1456	15
22	1464	15	1461	15	1460	15
23	1468	15	1466	15	1465	15
24	1473	15	1470	15	1469	15
25	1477	15	1475	15	1474	15
26	1482	15	1479	15	1478	15
27	1486	15	1484	15	1483	15
28	1491	15	1488	15	1487	15
29	1495	15	1493	15	1492	15
30	1500	15	1497	15	1496	15
31	1505	15	1502	15	1501	15
32	1509	16	1507	16	1506	16
33	1514	16	1512	16	1511	16
34	1519	16	1517	16	1516	16
35	1524	16	1522	16	1521	16
36	1530	16	1528	17	1527	17

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1535	17	1533	17	1532	17
38	1541	17	1539	18	1538	18
39	1547	18	1546	18	1545	18
40	1553	18	1552	19	1551	19
41	1560	19	1560	19	1559	19
42	1567	19	1567	20	1566	20
43	1574	20	1576	21	1575	21
44	1583	21	1585	22	1584	22
45	1592	22	1595	23	1594	23
46	1602	23	1606	25	1605	25
47	1614	25	1619	26	1618	26
48	1627	27	1634	29	1633	29
49	1644	31	1653	32	1652	32
50	1667	37	1678	38	1677	38
51	1705	51	1717	52	1716	52
52	1767	92	1781	93	1780	93

Table L–13. Raw-to-Scaled Score Comparison for Algebra I: Spring

RS	Preequating SS	Preequating SEM	Post-equating SS	Post-equating SEM
0	1218	92	1224	92
1	1279	51	1285	51
2	1315	37	1321	36
3	1337	30	1343	30
4	1353	27	1358	26
5	1366	24	1371	24
6	1377	22	1382	22
7	1386	21	1391	21
8	1395	20	1399	20
9	1402	19	1406	19
10	1409	18	1413	18
11	1416	18	1419	17
12	1422	17	1425	17
13	1428	17	1431	16
14	1433	16	1436	16
15	1438	16	1441	16
16	1443	16	1446	15
17	1448	15	1450	15
18	1453	15	1455	15
19	1458	15	1459	15
20	1462	15	1463	15
21	1466	15	1468	14
22	1471	15	1472	14
23	1475	15	1476	14
24	1479	15	1480	14
25	1484	14	1484	14
26	1488	14	1488	14
27	1492	14	1492	14
28	1496	14	1496	14
29	1500	14	1500	14
30	1504	14	1504	14
31	1508	14	1508	14
32	1513	14	1511	14
33	1517	14	1515	14
34	1521	14	1519	14
35	1525	15	1523	14
36	1529	15	1528	14

RS	Preequating SS	Preequating SEM	Post-equating SS	Post-equating SEM
37	1534	15	1532	14
38	1538	15	1536	15
39	1542	15	1540	15
40	1547	15	1545	15
41	1552	15	1549	15
42	1556	16	1554	15
43	1561	16	1559	16
44	1567	16	1564	16
45	1572	17	1569	17
46	1578	17	1575	17
47	1584	18	1581	17
48	1590	18	1587	18
49	1597	19	1594	19
50	1605	20	1601	19
51	1613	21	1609	20
52	1622	22	1618	21
53	1632	23	1627	23
54	1642	24	1638	24
55	1655	26	1650	26
56	1669	28	1665	29
57	1687	32	1684	32
58	1710	38	1708	39
59	1748	52	1749	53
60	1800	83	1800	93

Table L–14. Raw-to-Scaled Score Comparison for Biology: Spring

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1216	92	1216	92	1216	92
1	1277	51	1277	51	1277	51
2	1312	36	1313	36	1312	36
3	1334	30	1334	30	1334	30
4	1349	26	1350	26	1349	26
5	1362	24	1362	24	1362	24
6	1372	22	1372	22	1372	22
7	1381	20	1381	20	1381	20
8	1388	19	1389	19	1388	19
9	1395	18	1396	18	1395	18
10	1402	18	1402	18	1402	18
11	1408	17	1408	17	1408	17
12	1413	16	1414	16	1413	16
13	1419	16	1419	16	1419	16
14	1424	15	1424	15	1423	15
15	1428	15	1429	15	1428	15
16	1433	15	1433	15	1433	15
17	1437	14	1437	14	1437	14
18	1441	14	1441	14	1441	14
19	1445	14	1445	14	1445	14
20	1449	14	1449	14	1449	14
21	1453	14	1453	14	1453	14
22	1456	13	1457	13	1456	13
23	1460	13	1460	13	1460	13
24	1463	13	1464	13	1463	13
25	1467	13	1467	13	1467	13
26	1470	13	1470	13	1470	13
27	1474	13	1474	13	1473	13
28	1477	13	1477	13	1477	13
29	1480	13	1480	13	1480	13
30	1483	13	1483	13	1483	13
31	1487	13	1487	13	1486	13
32	1490	13	1490	13	1489	13
33	1493	13	1493	12	1492	12
34	1496	13	1496	12	1496	12
35	1499	13	1499	12	1499	12
36	1502	13	1502	12	1502	12

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1505	13	1505	13	1505	13
38	1509	13	1508	13	1508	13
39	1512	13	1512	13	1511	13
40	1515	13	1515	13	1514	13
41	1518	13	1518	13	1518	13
42	1522	13	1521	13	1521	13
43	1525	13	1525	13	1524	13
44	1528	13	1528	13	1528	13
45	1532	13	1532	13	1531	13
46	1535	14	1535	13	1535	13
47	1539	14	1539	14	1538	14
48	1543	14	1543	14	1542	14
49	1547	14	1547	14	1546	14
50	1551	15	1551	14	1550	14
51	1555	15	1555	15	1555	15
52	1560	15	1559	15	1559	15
53	1565	16	1564	16	1564	16
54	1570	16	1569	16	1569	16
55	1575	17	1575	17	1574	17
56	1581	17	1581	17	1580	17
57	1588	18	1587	18	1587	18
58	1595	19	1594	19	1594	19
59	1602	20	1602	20	1601	20
60	1611	22	1611	22	1610	22
61	1622	24	1621	24	1621	24
62	1634	26	1634	26	1633	26
63	1650	30	1649	30	1649	30
64	1671	36	1671	36	1671	36
65	1707	51	1707	51	1707	51
66	1768	92	1768	92	1768	92

Table L–15. Raw-to-Scaled Score Comparison for Literature: Spring

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1207	92	1210	92	1210	92
1	1268	51	1271	51	1272	51
2	1305	37	1307	36	1308	36
3	1327	30	1329	30	1330	30
4	1343	27	1345	27	1346	27
5	1356	24	1358	24	1358	24
6	1367	23	1368	22	1369	22
7	1377	21	1378	21	1378	21
8	1385	20	1386	20	1387	20
9	1393	19	1394	19	1394	19
10	1400	19	1400	18	1401	18
11	1407	18	1407	18	1408	18
12	1413	17	1413	17	1414	17
13	1419	17	1419	17	1419	17
14	1425	17	1424	16	1425	16
15	1430	16	1429	16	1430	16
16	1435	16	1434	16	1435	16
17	1440	16	1439	16	1440	16
18	1445	16	1444	15	1445	15
19	1450	15	1449	15	1449	15
20	1455	15	1453	15	1454	15
21	1460	15	1458	15	1459	15
22	1464	15	1462	15	1463	15
23	1469	15	1467	15	1468	15
24	1473	15	1471	15	1472	15
25	1478	15	1476	15	1476	15
26	1482	15	1480	15	1481	15
27	1487	15	1485	15	1485	15
28	1492	15	1489	15	1490	15
29	1496	15	1494	15	1495	15
30	1501	15	1499	15	1499	15
31	1506	16	1503	16	1504	16
32	1511	16	1508	16	1509	16
33	1516	16	1513	16	1514	16
34	1521	16	1518	16	1519	16
35	1526	16	1524	16	1524	16
36	1531	17	1529	17	1530	17

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1537	17	1535	17	1536	17
38	1543	17	1541	18	1542	18
39	1549	18	1548	18	1548	18
40	1556	18	1554	19	1555	19
41	1562	19	1562	19	1562	19
42	1570	19	1569	20	1570	20
43	1578	20	1578	21	1578	21
44	1586	21	1587	22	1587	22
45	1595	22	1597	23	1597	23
46	1606	24	1608	24	1609	24
47	1618	25	1621	26	1621	26
48	1632	28	1636	29	1636	29
49	1649	31	1654	32	1655	32
50	1672	37	1678	38	1679	38
51	1710	51	1717	52	1718	52
52	1772	92	1780	92	1781	92

Table L–16. Raw-to-Scaled Score Comparison for Algebra I: Summer

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1215	92	1207	92	1207	92
1	1277	51	1269	51	1269	51
2	1313	37	1306	37	1306	37
3	1336	31	1328	31	1329	31
4	1352	27	1345	27	1346	27
5	1365	25	1359	25	1359	25
6	1377	23	1370	23	1371	23
7	1386	21	1380	22	1381	22
8	1395	20	1389	21	1390	21
9	1403	19	1397	20	1398	20
10	1410	19	1405	19	1405	19
11	1417	18	1412	18	1412	18
12	1423	18	1418	18	1419	18
13	1429	17	1425	17	1425	17
14	1435	17	1431	17	1431	17
15	1441	16	1436	17	1437	17
16	1446	16	1442	16	1442	16
17	1451	16	1447	16	1448	16
18	1456	15	1452	16	1453	16
19	1460	15	1457	16	1458	16
20	1465	15	1462	15	1462	15
21	1470	15	1466	15	1467	15
22	1474	15	1471	15	1472	15
23	1478	15	1476	15	1476	15
24	1482	14	1480	15	1481	15
25	1487	14	1484	15	1485	15
26	1491	14	1489	15	1489	15
27	1495	14	1493	15	1494	15
28	1499	14	1497	15	1498	15
29	1503	14	1502	15	1502	15
30	1507	14	1506	15	1507	15
31	1511	14	1510	15	1511	15
32	1514	14	1514	15	1515	15
33	1518	14	1519	15	1519	15
34	1522	14	1523	15	1524	15
35	1526	14	1528	15	1528	15
36	1531	14	1532	15	1533	15

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1535	14	1537	15	1537	15
38	1539	15	1541	15	1542	15
39	1543	15	1546	16	1547	16
40	1548	15	1551	16	1552	16
41	1552	15	1556	16	1557	16
42	1557	15	1561	16	1562	16
43	1562	16	1566	17	1567	17
44	1567	16	1572	17	1573	17
45	1572	17	1578	17	1579	17
46	1578	17	1584	18	1585	18
47	1584	17	1590	18	1591	18
48	1590	18	1597	18	1598	18
49	1597	19	1604	19	1605	19
50	1604	19	1611	19	1612	19
51	1612	20	1619	20	1620	20
52	1621	21	1627	21	1628	21
53	1630	23	1636	22	1637	22
54	1641	24	1646	23	1647	23
55	1653	26	1657	25	1658	25
56	1667	28	1671	27	1671	27
57	1684	31	1688	31	1688	31
58	1707	36	1712	39	1712	39
59	1743	51	1754	55	1755	55
60	1800	92	1800	96	1800	96

Table L–17. Raw-to-Scaled Score Comparison for Biology: Summer

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1215	92	1211	92	1211	92
1	1276	51	1272	51	1272	51
2	1311	36	1308	36	1307	36
3	1333	30	1329	30	1329	30
4	1348	26	1345	26	1344	26
5	1361	24	1357	24	1357	24
6	1371	22	1368	22	1367	22
7	1380	20	1376	20	1376	20
8	1388	19	1384	19	1384	19
9	1395	18	1391	18	1391	18
10	1401	18	1398	18	1398	18
11	1407	17	1404	17	1404	17
12	1413	16	1410	16	1409	16
13	1418	16	1415	16	1414	16
14	1423	16	1420	16	1419	16
15	1428	15	1425	15	1424	15
16	1432	15	1429	15	1429	15
17	1437	15	1433	15	1433	15
18	1441	14	1438	14	1437	14
19	1445	14	1442	14	1441	14
20	1449	14	1446	14	1445	14
21	1452	14	1450	14	1449	14
22	1456	14	1453	14	1453	14
23	1460	13	1457	13	1457	13
24	1463	13	1461	13	1460	13
25	1467	13	1464	13	1464	13
26	1470	13	1468	13	1467	13
27	1474	13	1471	13	1471	13
28	1477	13	1475	13	1474	13
29	1480	13	1478	13	1477	13
30	1484	13	1481	13	1481	13
31	1487	13	1485	13	1484	13
32	1490	13	1488	13	1488	13
33	1493	13	1491	13	1491	13
34	1496	13	1495	13	1494	13
35	1500	13	1498	13	1498	13
36	1503	13	1501	13	1501	13

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1506	13	1505	13	1504	13
38	1509	13	1508	13	1508	13
39	1513	13	1512	13	1511	13
40	1516	13	1515	13	1515	13
41	1519	13	1519	13	1518	13
42	1523	13	1522	14	1522	14
43	1526	13	1526	14	1526	14
44	1530	13	1530	14	1529	14
45	1533	13	1534	14	1533	14
46	1537	14	1538	14	1537	14
47	1541	14	1542	14	1541	14
48	1545	14	1546	15	1545	15
49	1549	14	1550	15	1550	15
50	1553	15	1555	15	1554	15
51	1557	15	1559	16	1559	16
52	1562	15	1564	16	1564	16
53	1566	16	1570	16	1569	16
54	1571	16	1575	17	1575	17
55	1577	17	1581	17	1581	17
56	1583	17	1587	18	1587	18
57	1589	18	1594	19	1594	19
58	1596	19	1602	20	1601	20
59	1604	20	1610	21	1610	21
60	1612	22	1620	22	1619	22
61	1622	23	1631	24	1630	24
62	1635	26	1644	27	1643	27
63	1650	30	1660	31	1660	31
64	1671	36	1682	37	1682	37
65	1707	51	1719	51	1719	51
66	1768	92	1781	92	1781	92

Table L–18. Raw-to-Scaled Score Comparison for Literature: Summer

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
0	1210	92	1203	92	1202	92
1	1271	51	1264	51	1263	51
2	1308	37	1300	37	1300	37
3	1330	30	1323	30	1322	30
4	1346	27	1339	27	1338	27
5	1359	24	1352	24	1351	24
6	1370	23	1363	23	1362	23
7	1379	21	1373	21	1372	21
8	1388	20	1381	20	1381	20
9	1396	19	1389	19	1388	19
10	1403	19	1396	19	1396	19
11	1410	18	1403	18	1402	18
12	1416	18	1410	18	1409	18
13	1422	17	1416	17	1415	17
14	1428	17	1422	17	1421	17
15	1433	16	1427	17	1427	17
16	1438	16	1433	16	1432	16
17	1443	16	1438	16	1437	16
18	1449	16	1443	16	1443	16
19	1453	16	1448	16	1448	16
20	1458	16	1453	16	1453	16
21	1463	15	1458	16	1458	16
22	1468	15	1463	16	1462	16
23	1472	15	1468	16	1467	16
24	1477	15	1473	16	1472	16
25	1482	15	1478	16	1477	16
26	1487	15	1483	16	1482	16
27	1491	15	1488	16	1487	16
28	1496	15	1492	16	1492	16
29	1501	16	1497	16	1497	16
30	1506	16	1503	16	1502	16
31	1511	16	1508	16	1507	16
32	1516	16	1513	16	1512	16
33	1521	16	1518	17	1518	17
34	1526	16	1524	17	1523	17
35	1531	17	1530	17	1529	17
36	1537	17	1536	18	1535	18

RS	Preequating SS	Preequating SEM	Post-equating All Items SS	Post-equating All Items SEM	Post-equating Outlier Removed SS	Post-equating Outlier Removed SEM
37	1543	17	1542	18	1542	18
38	1549	18	1549	19	1548	19
39	1555	18	1556	19	1555	19
40	1562	19	1564	20	1563	20
41	1569	19	1572	21	1571	21
42	1577	20	1581	22	1580	22
43	1585	21	1590	23	1590	23
44	1594	22	1601	24	1600	24
45	1604	23	1613	25	1613	25
46	1615	24	1627	27	1626	27
47	1628	26	1643	29	1642	29
48	1643	28	1662	32	1661	32
49	1661	32	1684	36	1684	36
50	1685	38	1714	42	1713	42
51	1723	52	1758	55	1758	55
52	1786	92	1800	94	1800	94

APPENDIX M: RELIABILITIES

Table M–1. Reliabilities

Column Heading	Definition
Level	Total test or module level
Group	Student group: all students or subgroup
Pts.	Max points possible
Len.	Test length
N	Number of students
Mean	Mean of raw score
SD	Standard deviation of raw score
r	Reliability coefficient: Cronbach's alpha
SEM	Standard error of measurement

Note: "DNR" in the tables below represents "Do Not Report". This happened only when the *N* count was small.

Table M–2. Winter: Algebra I Reliabilities

Table M–2. Winter: Algebra I Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	55500	24.46	10.71	0.88	3.64
	Module 1	All	30	21	55500	12.02	5.93	0.79	2.70
	Module 2	All	30	21	55500	12.44	5.39	0.80	2.44
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	27439	25.08	10.48	0.88	3.65
		Male	60	42	28015	23.86	10.89	0.89	3.62
	Module 1	Female	30	21	27439	12.34	5.84	0.78	2.71
		Male	30	21	28015	11.71	5.99	0.80	2.67
	Module 2	Female	30	21	27439	12.74	5.27	0.79	2.44
		Male	30	21	28015	12.15	5.50	0.80	2.44
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	10868	19.58	9.00	0.85	3.44
		American Indian	60	42	108	24.23	10.41	0.88	3.66
		Asian	60	42	1350	29.73	13.44	0.92	3.81
		Hispanic	60	42	7823	20.94	9.81	0.87	3.51
		Multi-racial	60	42	1400	23.65	10.36	0.88	3.62
		Native Hawaiian/ Pacific Islander	60	42	42	23.43	9.50	0.86	3.56
		White	60	42	33857	26.67	10.53	0.88	3.68
	Module 1	African American	30	21	10868	9.51	4.99	0.75	2.52
		American Indian	30	21	108	12.01	6.04	0.80	2.69
		Asian	30	21	1350	14.84	7.39	0.85	2.82
		Hispanic	30	21	7823	10.10	5.47	0.78	2.57
		Multi-racial	30	21	1400	11.63	5.75	0.78	2.68
		Native Hawaiian/ Pacific Islander	30	21	42	11.43	5.79	0.79	2.65
		White	30	21	33857	13.17	5.87	0.78	2.74
	Module 2	African American	30	21	10868	10.06	4.69	0.75	2.34
		American Indian	30	21	108	12.22	5.18	0.77	2.47
		Asian	30	21	1350	14.89	6.50	0.85	2.56
		Hispanic	30	21	7823	10.84	5.02	0.77	2.39
		Multi-racial	30	21	1400	12.03	5.21	0.78	2.43
		Native Hawaiian/ Pacific Islander	30	21	42	12.00	4.51	0.73	2.36
		White	30	21	33857	13.50	5.30	0.78	2.46
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	2524	15.54	7.73	0.82	3.25
	Module 1	All	30	21	2524	7.26	4.36	0.73	2.27
	Module 2	All	30	21	2524	8.29	4.08	0.68	2.32
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	10194	18.06	8.64	0.85	3.35
	Module 1	All	30	21	10194	8.61	4.71	0.74	2.41
	Module 2	All	30	21	10194	9.46	4.61	0.74	2.33
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	28939	21.72	9.68	0.87	3.53
	Module 1	All	30	21	28939	10.59	5.38	0.77	2.60
	Module 2	All	30	21	28939	11.12	4.96	0.77	2.39

Table M–3. Winter: Biology Reliabilities

Table M–3. Winter: Biology Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	41916	29.05	12.94	0.92	3.71
	Module 1	All	30	21	41916	14.07	6.68	0.85	2.63
	Module 2	All	30	21	41916	14.98	6.86	0.85	2.61
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	20489	29.49	12.60	0.91	3.72
		Male	60	42	21389	28.65	13.23	0.92	3.70
	Module 1	Female	30	21	20489	14.22	6.46	0.83	2.63
		Male	30	21	21389	13.93	6.88	0.85	2.63
	Module 2	Female	30	21	20489	15.26	6.76	0.85	2.62
		Male	30	21	21389	14.72	6.95	0.86	2.60
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	7390	22.57	9.97	0.87	3.60
		American Indian	60	42	60	26.97	12.42	0.91	3.66
		Asian	60	42	1073	35.79	16.40	0.95	3.64
		Hispanic	60	42	5339	24.50	11.03	0.89	3.65
		Multi-racial	60	42	1014	27.69	11.81	0.90	3.71
		Native Hawaiian/ Pacific Islander	60	42	34	27.21	10.97	0.89	3.63
		White	60	42	26963	31.53	12.98	0.92	3.73
	Module 1	African American	30	21	7390	11.07	5.15	0.76	2.52
		American Indian	30	21	60	13.13	6.86	0.86	2.56
		Asian	30	21	1073	17.66	8.45	0.91	2.58
		Hispanic	30	21	5339	11.89	5.72	0.80	2.57
		Multi-racial	30	21	1014	13.42	6.17	0.82	2.62
		Native Hawaiian/ Pacific Islander	30	21	34	13.06	5.60	0.79	2.56
		White	30	21	26963	15.22	6.77	0.85	2.65
	Module 2	African American	30	21	7390	11.50	5.56	0.79	2.56
		American Indian	30	21	60	13.83	6.63	0.85	2.56
		Asian	30	21	1073	18.13	8.38	0.91	2.57
		Hispanic	30	21	5339	12.61	5.99	0.81	2.58
		Multi-racial	30	21	1014	14.27	6.30	0.83	2.62
		Native Hawaiian/ Pacific Islander	30	21	34	14.15	6.22	0.83	2.55
		White	30	21	26963	16.31	6.82	0.85	2.61
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	1675	18.04	6.97	0.76	3.42
	Module 1	All	30	21	1675	8.84	3.88	0.62	2.41
	Module 2	All	30	21	1675	9.20	4.00	0.63	2.43
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	8243	22.06	9.46	0.86	3.59
	Module 1	All	30	21	8243	10.75	5.01	0.74	2.53
	Module 2	All	30	21	8243	11.31	5.22	0.76	2.54
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	20972	25.31	11.07	0.89	3.67
	Module 1	All	30	21	20972	12.30	5.74	0.80	2.59
	Module 2	All	30	21	20972	13.00	6.03	0.81	2.60

Table M–4. Winter: Literature Reliabilities

Table M–4. Winter: Literature Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	36551	26.37	10.66	0.91	3.19
	Module 1	All	30	21	36551	12.78	5.52	0.83	2.25
	Module 2	All	30	21	36551	13.58	5.66	0.84	2.25
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	16306	28.51	10.46	0.91	3.17
		Male	60	42	20218	24.64	10.50	0.91	3.17
	Module 1	Female	30	21	16306	13.90	5.47	0.83	2.25
		Male	30	21	20218	11.89	5.39	0.83	2.23
	Module 2	Female	30	21	16306	14.61	5.52	0.84	2.22
		Male	30	21	20218	12.76	5.64	0.84	2.25
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	6221	21.54	9.22	0.88	3.25
		American Indian	60	42	64	25.83	9.59	0.89	3.21
		Asian	60	42	972	29.90	11.57	0.93	3.14
		Hispanic	60	42	4854	23.10	9.90	0.89	3.23
		Multi-racial	60	42	807	25.38	10.31	0.90	3.19
		Native Hawaiian/ Pacific Islander	60	42	29	28.52	10.14	0.90	3.25
		White	60	42	23567	28.21	10.56	0.91	3.14
	Module 1	African American	30	21	6221	10.33	4.83	0.78	2.27
		American Indian	30	21	64	12.33	5.08	0.81	2.24
		Asian	30	21	972	14.27	5.99	0.86	2.25
		Hispanic	30	21	4854	11.09	5.14	0.81	2.26
		Multi-racial	30	21	807	12.21	5.38	0.83	2.24
		Native Hawaiian/ Pacific Islander	30	21	29	13.31	5.36	0.81	2.32
		White	30	21	23567	13.74	5.46	0.83	2.23
	Module 2	African American	30	21	6221	11.21	5.06	0.79	2.31
		American Indian	30	21	64	13.50	5.16	0.80	2.28
		Asian	30	21	972	15.63	6.02	0.87	2.18
		Hispanic	30	21	4854	12.01	5.34	0.82	2.29
		Multi-racial	30	21	807	13.17	5.53	0.83	2.26
		Native Hawaiian/ Pacific Islander	30	21	29	15.21	5.18	0.80	2.29
		White	30	21	23567	14.47	5.60	0.84	2.21
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	1575	16.60	7.04	0.80	3.15
	Module 1	All	30	21	1575	7.69	3.64	0.64	2.18
	Module 2	All	30	21	1575	8.92	4.16	0.70	2.27
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	8111	19.46	8.16	0.85	3.18
	Module 1	All	30	21	8111	9.40	4.37	0.74	2.22
	Module 2	All	30	21	8111	10.06	4.49	0.75	2.26
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	18090	23.10	9.66	0.89	3.23
	Module 1	All	30	21	18090	11.17	5.06	0.80	2.26
	Module 2	All	30	21	18090	11.93	5.22	0.81	2.29

Table M–5. Spring: Algebra Reliabilities

Table M–5. Spring: Algebra Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	167097	26.82	12.28	0.91	3.61
	Module 1	All	30	21	167097	13.08	6.32	0.82	2.66
	Module 2	All	30	21	167097	13.73	6.44	0.86	2.44
Gender	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	Female	60	42	82069	27.37	11.97	0.91	3.60
		Male	60	42	84839	26.30	12.55	0.92	3.61
	Module 1	Female	30	21	82069	13.37	6.16	0.81	2.66
		Male	30	21	84839	12.81	6.45	0.83	2.66
	Module 2	Female	30	21	82069	14.00	6.29	0.85	2.43
		Male	30	21	84839	13.49	6.56	0.86	2.45
Ethnicity	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	African American	60	42	26976	19.12	8.96	0.85	3.42
		American Indian	60	42	272	25.67	11.55	0.90	3.64
		Asian	60	42	6270	35.55	13.63	0.93	3.62
		Hispanic	60	42	19203	20.61	10.11	0.88	3.48
		Multi-racial	60	42	4358	25.44	11.90	0.91	3.60
		Native Hawaiian/ Pacific Islander	60	42	142	29.49	11.25	0.90	3.62
		White	60	42	109657	29.36	11.95	0.91	3.59
	Module 1	African American	30	21	26976	9.34	4.61	0.72	2.45
		American Indian	30	21	272	12.54	5.97	0.80	2.69
		Asian	30	21	6270	17.58	7.18	0.86	2.72
		Hispanic	30	21	19203	10.08	5.15	0.76	2.51
		Multi-racial	30	21	4358	12.44	6.12	0.81	2.65
		Native Hawaiian/ Pacific Islander	30	21	142	14.46	5.72	0.78	2.69
		White	30	21	109657	14.30	6.21	0.81	2.67
	Module 2	African American	30	21	26976	9.78	4.95	0.77	2.39
		American Indian	30	21	272	13.13	6.11	0.84	2.46
		Asian	30	21	6270	17.97	6.86	0.88	2.40
		Hispanic	30	21	19203	10.54	5.51	0.81	2.41
		Multi-racial	30	21	4358	13.00	6.23	0.85	2.44
		Native Hawaiian/ Pacific Islander	30	21	142	15.03	6.07	0.84	2.42
		White	30	21	109657	15.06	6.24	0.85	2.41
ELL	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	6486	15.78	7.94	0.83	3.24
	Module 1	All	30	21	6486	7.91	4.29	0.70	2.34
	Module 2	All	30	21	6486	7.87	4.26	0.72	2.25
IEP	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	26205	17.76	8.79	0.85	3.36
	Module 1	All	30	21	26205	8.76	4.54	0.72	2.40
	Module 2	All	30	21	26205	9.00	4.84	0.76	2.35
ED	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	76757	22.23	10.49	0.89	3.52
	Module 1	All	30	21	76757	10.84	5.37	0.77	2.55
	Module 2	All	30	21	76757	11.39	5.66	0.82	2.43

Table M–6. Spring: Biology Reliabilities

Table M–6. Spring: Biology Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	142262	34.97	14.33	0.93	3.83
	Module 1	All	30	21	142262	17.59	7.34	0.86	2.73
	Module 2	All	30	21	142262	17.38	7.51	0.87	2.69
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	69967	35.48	13.89	0.92	3.85
		Male	60	42	72161	34.51	14.72	0.93	3.80
	Module 1	Female	30	21	69967	17.82	7.09	0.85	2.75
		Male	30	21	72161	17.38	7.57	0.87	2.71
	Module 2	Female	30	21	69967	17.66	7.34	0.86	2.70
		Male	30	21	72161	17.12	7.65	0.88	2.66
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	21084	25.91	11.73	0.90	3.78
		American Indian	60	42	235	33.25	13.24	0.92	3.84
		Asian	60	42	5212	43.15	15.26	0.94	3.66
		Hispanic	60	42	15076	27.15	12.27	0.90	3.80
		Multi-racial	60	42	3376	32.48	13.79	0.92	3.84
		Native Hawaiian/ Pacific Islander	60	42	122	37.34	13.41	0.92	3.86
		White	60	42	96997	37.83	13.72	0.92	3.81
	Module 1	African American	30	21	21084	13.24	6.19	0.81	2.69
		American Indian	30	21	235	16.77	6.67	0.83	2.75
		Asian	30	21	5212	21.79	7.77	0.89	2.60
		Hispanic	30	21	15076	13.77	6.40	0.82	2.71
		Multi-racial	30	21	3376	16.36	7.16	0.85	2.74
		Native Hawaiian/ Pacific Islander	30	21	122	18.89	6.85	0.84	2.75
		White	30	21	96997	18.96	7.06	0.85	2.72
	Module 2	African American	30	21	21084	12.66	6.18	0.82	2.65
		American Indian	30	21	235	16.48	7.10	0.86	2.68
		Asian	30	21	5212	21.36	7.92	0.89	2.58
		Hispanic	30	21	15076	13.38	6.48	0.83	2.66
		Multi-racial	30	21	3376	16.12	7.18	0.86	2.70
		Native Hawaiian/ Pacific Islander	30	21	122	18.45	7.11	0.85	2.71
		White	30	21	96997	18.87	7.20	0.86	2.66
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	4868	20.14	8.76	0.83	3.61
	Module 1	All	30	21	4868	10.41	5.02	0.74	2.57
	Module 2	All	30	21	4868	9.74	4.52	0.69	2.53
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	22784	24.07	11.01	0.89	3.72
	Module 1	All	30	21	22784	12.33	5.80	0.79	2.66
	Module 2	All	30	21	22784	11.74	5.88	0.81	2.60
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	62883	29.46	12.81	0.91	3.83
	Module 1	All	30	21	62883	14.93	6.64	0.83	2.73
	Module 2	All	30	21	62883	14.53	6.77	0.84	2.68

Table M–7. Spring: Literature Reliabilities

Table M–7. Spring: Literature Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	130328	30.01	10.75	0.92	3.13
	Module 1	All	30	21	130328	14.90	5.71	0.84	2.28
	Module 2	All	30	21	130328	15.11	5.49	0.85	2.15
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	62697	31.92	10.16	0.91	3.09
		Male	60	42	67506	28.25	10.97	0.92	3.13
	Module 1	Female	30	21	62697	15.93	5.43	0.83	2.25
		Male	30	21	67506	13.95	5.79	0.85	2.27
	Module 2	Female	30	21	62697	15.98	5.19	0.83	2.11
		Male	30	21	67506	14.30	5.63	0.85	2.15
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	20025	23.74	10.09	0.90	3.25
		American Indian	60	42	194	27.32	10.25	0.90	3.18
		Asian	60	42	4847	34.65	10.50	0.92	2.98
		Hispanic	60	42	13051	24.67	10.60	0.91	3.23
		Multi-racial	60	42	3071	28.98	10.64	0.91	3.15
		Native Hawaiian/ Pacific Islander	60	42	105	31.86	9.79	0.90	3.14
		White	60	42	88900	32.00	10.06	0.91	3.05
	Module 1	African American	30	21	20025	11.67	5.44	0.81	2.35
		American Indian	30	21	194	13.36	5.50	0.82	2.32
		Asian	30	21	4847	17.27	5.59	0.85	2.19
		Hispanic	30	21	13051	12.16	5.68	0.83	2.33
		Multi-racial	30	21	3071	14.44	5.62	0.83	2.29
		Native Hawaiian/ Pacific Islander	30	21	105	16.09	5.37	0.82	2.29
		White	30	21	88900	15.92	5.35	0.83	2.23
	Module 2	African American	30	21	20025	12.06	5.21	0.81	2.24
		American Indian	30	21	194	13.96	5.33	0.84	2.16
		Asian	30	21	4847	17.38	5.30	0.85	2.02
		Hispanic	30	21	13051	12.51	5.42	0.83	2.24
		Multi-racial	30	21	3071	14.54	5.51	0.85	2.16
		Native Hawaiian/ Pacific Islander	30	21	105	15.77	4.88	0.80	2.16
		White	30	21	88900	16.08	5.16	0.84	2.09
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	3962	16.58	7.27	0.81	3.14
	Module 1	All	30	21	3962	7.93	3.94	0.69	2.20
	Module 2	All	30	21	3962	8.65	4.01	0.69	2.23
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	21755	20.13	9.03	0.88	3.17
	Module 1	All	30	21	21755	9.81	4.84	0.78	2.25
	Module 2	All	30	21	21755	10.32	4.77	0.78	2.22
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	56713	26.02	10.50	0.91	3.21
	Module 1	All	30	21	56713	12.85	5.61	0.83	2.32
	Module 2	All	30	21	56713	13.17	5.39	0.83	2.21

Table M–8. Summer: Algebra Reliabilities

Table M–8. Summer: Algebra Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	1454	25.24	8.14	0.83	3.39
	Module 1	All	30	21	1454	12.14	4.46	0.73	2.33
	Module 2	All	30	21	1454	13.09	4.39	0.69	2.46
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	731	25.45	7.84	0.81	3.39
		Male	60	42	714	25.04	8.48	0.84	3.38
	Module 1	Female	30	21	731	12.38	4.36	0.72	2.32
		Male	30	21	714	11.91	4.57	0.74	2.32
	Module 2	Female	30	21	731	13.07	4.19	0.65	2.47
		Male	30	21	714	13.12	4.60	0.72	2.45
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	218	22.29	8.32	0.84	3.36
		American Indian	60	42	1	DNR	DNR	DNR	DNR
		Asian	60	42	51	31.78	9.08	0.86	3.42
		Hispanic	60	42	115	25.95	7.57	0.80	3.39
		Multi-racial	60	42	32	23.97	8.55	0.84	3.47
		Native Hawaiian/ Pacific Islander	60	42	3	DNR	DNR	DNR	DNR
		White	60	42	1018	25.58	7.88	0.82	3.38
	Module 1	African American	30	21	218	10.84	4.52	0.75	2.27
		American Indian	30	21	1	DNR	DNR	DNR	DNR
		Asian	30	21	51	15.31	4.93	0.77	2.36
		Hispanic	30	21	115	12.12	3.91	0.66	2.29
		Multi-racial	30	21	32	11.44	4.71	0.74	2.41
		Native Hawaiian/ Pacific Islander	30	21	3	DNR	DNR	DNR	DNR
		White	30	21	1018	12.33	4.39	0.72	2.33
	Module 2	African American	30	21	218	11.45	4.54	0.70	2.47
		American Indian	30	21	1	DNR	DNR	DNR	DNR
		Asian	30	21	51	16.47	4.87	0.74	2.47
		Hispanic	30	21	115	13.83	4.50	0.69	2.49
		Multi-racial	30	21	32	12.53	4.50	0.69	2.50
		Native Hawaiian/ Pacific Islander	30	21	3	DNR	DNR	DNR	DNR
		White	30	21	1018	13.24	4.18	0.66	2.44
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	15	25.47	8.17	0.83	3.36
	Module 1	All	30	21	15	11.67	4.72	0.78	2.22
	Module 2	All	30	21	15	13.80	4.21	0.65	2.50
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	209	21.49	8.13	0.84	3.30
	Module 1	All	30	21	209	10.14	4.35	0.73	2.27
	Module 2	All	30	21	209	11.35	4.50	0.72	2.39
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	353	23.25	8.30	0.84	3.36
	Module 1	All	30	21	353	11.15	4.55	0.74	2.31
	Module 2	All	30	21	353	12.10	4.46	0.70	2.44

Table M–9. Summer: Biology Reliabilities

Table M–9. Summer: Biology Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	960	33.07	10.16	0.86	3.80
	Module 1	All	30	21	960	16.75	5.39	0.76	2.66
	Module 2	All	30	21	960	16.32	5.61	0.77	2.70
Gender	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	Female	60	42	471	33.78	9.49	0.84	3.81
		Male	60	42	461	32.00	10.58	0.87	3.78
	Module 1	Female	30	21	471	16.82	5.20	0.74	2.66
		Male	30	21	461	16.47	5.53	0.77	2.66
	Module 2	Female	30	21	471	16.96	5.13	0.72	2.72
		Male	30	21	461	15.53	5.88	0.79	2.67
Ethnicity	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	African American	60	42	94	30.28	10.25	0.87	3.72
		American Indian	60	42	2	DNR	DNR	DNR	DNR
		Asian	60	42	51	43.96	12.74	0.92	3.65
		Hispanic	60	42	110	31.84	8.26	0.79	3.81
		Multi-racial	60	42	19	34.42	11.42	0.89	3.76
		Native Hawaiian/ Pacific Islander	60	42	4	DNR	DNR	DNR	DNR
		White	60	42	642	32.58	9.54	0.84	3.81
	Module 1	African American	30	21	94	15.76	5.57	0.78	2.60
		American Indian	30	21	2	DNR	DNR	DNR	DNR
		Asian	30	21	51	22.39	6.64	0.86	2.51
		Hispanic	30	21	110	15.94	4.61	0.67	2.67
		Multi-racial	30	21	19	16.53	6.37	0.83	2.61
		Native Hawaiian/ Pacific Islander	30	21	4	DNR	DNR	DNR	DNR
		White	30	21	642	16.46	5.07	0.72	2.67
	Module 2	African American	30	21	94	14.52	5.70	0.79	2.63
		American Indian	30	21	2	DNR	DNR	DNR	DNR
		Asian	30	21	51	21.57	6.80	0.85	2.63
		Hispanic	30	21	110	15.90	4.65	0.66	2.70
		Multi-racial	30	21	19	17.89	5.80	0.78	2.69
		Native Hawaiian/ Pacific Islander	30	21	4	DNR	DNR	DNR	DNR
		White	30	21	642	16.12	5.33	0.74	2.70
ELL	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	12	28.42	11.04	0.89	3.59
	Module 1	All	30	21	12	15.67	6.41	0.84	2.58
	Module 2	All	30	21	12	12.75	4.85	0.73	2.52
IEP	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	141	29.40	9.20	0.83	3.81
	Module 1	All	30	21	141	15.36	4.77	0.68	2.69
	Module 2	All	30	21	141	14.04	5.42	0.76	2.67
ED	**Level**	**Group**	**Pts.**	**Len.**	***N***	**Mean**	**SD**	**r**	***SEM***
	Total	All	60	42	218	30.95	9.40	0.84	3.77
	Module 1	All	30	21	218	16.10	5.11	0.73	2.64
	Module 2	All	30	21	218	14.85	5.28	0.74	2.67

Table M–10. Summer: Literature Reliabilities

Table M–10. Summer: Literature Reliabilities

	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
Overall	Total	All	60	42	446	27.47	7.48	0.83	3.11
	Module 1	All	30	21	446	13.72	3.83	0.66	2.24
	Module 2	All	30	21	446	13.76	4.30	0.75	2.15
Gender	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	Female	60	42	171	28.11	7.31	0.82	3.11
		Male	60	42	275	27.08	7.57	0.83	3.10
	Module 1	Female	30	21	171	14.16	3.68	0.63	2.25
		Male	30	21	275	13.44	3.91	0.68	2.22
	Module 2	Female	30	21	171	13.95	4.13	0.73	2.15
		Male	30	21	275	13.64	4.40	0.76	2.15
Ethnicity	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	African American	60	42	41	26.80	6.52	0.77	3.15
		American Indian	60	42	1	DNR	DNR	DNR	DNR
		Asian	60	42	27	29.67	7.01	0.80	3.16
		Hispanic	60	42	36	24.72	6.80	0.79	3.11
		Multi-racial	60	42	6	20.33	7.39	0.84	2.99
		Native Hawaiian/ Pacific Islander	60	42	0	DNR	DNR	DNR	DNR
		White	60	42	334	27.86	7.55	0.83	3.08
	Module 1	African American	30	21	41	13.12	2.98	0.43	2.25
		American Indian	30	21	1	DNR	DNR	DNR	DNR
		Asian	30	21	27	14.07	3.94	0.67	2.25
		Hispanic	30	21	36	12.14	3.54	0.58	2.30
		Multi-racial	30	21	6	10.00	3.85	0.70	2.12
		Native Hawaiian/ Pacific Islander	30	21	0	DNR	DNR	DNR	DNR
		White	30	21	334	14.01	3.88	0.67	2.22
	Module 2	African American	30	21	41	13.68	4.26	0.74	2.19
		American Indian	30	21	1	DNR	DNR	DNR	DNR
		Asian	30	21	27	15.59	3.59	0.61	2.23
		Hispanic	30	21	36	12.58	4.04	0.73	2.09
		Multi-racial	30	21	6	10.33	4.18	0.75	2.10
		Native Hawaiian/ Pacific Islander	30	21	0	DNR	DNR	DNR	DNR
		White	30	21	334	13.85	4.31	0.76	2.13
ELL	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	15	21.80	7.06	0.78	3.28
	Module 1	All	30	21	15	10.33	3.44	0.54	2.32
	Module 2	All	30	21	15	11.47	4.31	0.71	2.32
IEP	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	91	24.24	7.76	0.83	3.17
	Module 1	All	30	21	91	12.03	3.85	0.65	2.29
	Module 2	All	30	21	91	12.21	4.42	0.75	2.20
ED	Level	Group	Pts.	Len.	*N*	Mean	SD	r	*SEM*
	Total	All	60	42	104	24.97	6.69	0.77	3.21
	Module 1	All	30	21	104	12.39	3.49	0.56	2.31
	Module 2	All	30	21	104	12.58	3.96	0.68	2.23

REFERENCES

Allman, C. (2004). *Test access: Making tests accessible for students with visual impairments–A guide for test publishers, test developers, and state assessment personnel* (2nd ed.). Louisville, KY: American Printing House for the Blind. Retrieved from www.aph.org.

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

Anderson, L.W. and Krathwohl, D.R. (Eds.) (2001). A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives: Complete Edition. New York: Longman.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 508–600). Washington, DC: American Council on Education.

Bloom, B.S. (1956). Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain. New York: Longman.

Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice, 17*(1), 5–9.

Brennan, R. L. (2004). BB-Class (Version 1.0) [Computer Software]. Retrieved from http://www.education.uiowa.edu/casma.

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289.

Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., et al. (1997). The principles of universal design. Raleigh: North Carolina University, College of Design.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Cronbach, L. J., & Shavelson, R. L. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418.

Data Recognition Corporation. (2010). *Fairness in testing: Training manual for issues of bias, fairness, and sensitivity*. Maple Grove, MN.

Dorans, N., Schmitt, A., & Bleistein, C. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, *29*, 309–319.

Eignor, D. R. (1985). *An investigation of the feasibility and practical outcomes of pre-equating the SAT verbal and mathematical sections* (Research Report 85-10). Princeton, NJ: Educational Testing Service.

Eignor, D. R., & Stocking, M. L. (1986). *An investigation of the possible causes for the inadequacy of IRT preequating* (Research Report 86-14). Princeton, NJ: Educational Testing Service.

Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, *24*(3), 21–28.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.

Hambleton, R., & Novick, M. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, *10*, 159–170.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score theory models. *Journal of Educational Measurement*, *27*(4), 345–359.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practices*, *10*(2), 33–41.

Hess, K. (2004). *Applying Webb's depth-of-knowledge levels in reading*. [online] available: www.nciea.org.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking: methods and practices (2nd ed.). New York: Springer.

Kolen, M. J., & Harris, D. J. (1990). Comparison of item preequating and random groups equating using IRT and equipercentile methods. *Journal of Educational Measurement*, *27* (1), 27–39.

Lane, S. (1999). *Validity evidence for assessments*. Paper presented at the 1999 Edward F. Reidy, Interactive Lecture Series, Providence, RI.

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, *21*(1), 23–30.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A bookmark approach*. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Linacre, J. M. (2009). *A user's guide to WINSTEPS MINNISTEP Rasch-model computer programs*. Chicago: Winsteps.

Linacre, J. M. (2013). *A user's guide to WINSTEPS MINNISTEP Rasch-model computer programs*. Chicago: Winsteps.

Linacre, J. M., & Wright, B. D. (2003). *WINSTEPS 3.54: Multiple-choice, rating scale, and partial credit Rasch analysis* [computer software]. Chicago: MESA Press.

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

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

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719–748.

Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200–215.

McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of a general theory. *Multivariate Behavioral Research*, *14*, 21–38.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 3–104). Washington, DC: American Council on Education.

Pennsylvania Department of Education. (2010). *Psychometric analysis report for the fall 2010 Keystone field tests*. Harrisburg, PA: PDE.

Pennsylvania Department of Education. (2011). *Algebra I Keystone Feb 2011 item and scoring sampler*. Retrieved November 11, 2011, from www.pdesas.org/Assessment/Keystone.

Pennsylvania Department of Education. (2011). *Biology Keystone Feb 2011 item and scoring sampler*. Retrieved November 11, 2011, from www.pdesas.org/Assessment/Keystone.

Pennsylvania Department of Education. (2011). *Keystone Exams score report focus group findings*. Harrisburg, PA: PDE.

Pennsylvania Department of Education. (2011). *Keystone standard setting technical report: Algebra I, Biology, and Literature*. Harrisburg, PA: PDE.

Pennsylvania Department of Education. (2011). *Literature Keystone Feb 2011 Item and Scoring Sampler*. Retrieved November 11, 2011, from www.pdesas.org/Assessment/Keystone.

Pennsylvania Department of Education. (2011). *PSSA, PSSA-M, Keystone (paper/pencil) accommodations guidelines for students with IEPs and students with 504 plans, revised 1-12-2011*. Harrisburg, PA: PDE. Retrieved February 25, 2011, from *http://www.pde.state.pa.us.*

Pennsylvania Department of Education. (2012). *Accommodations Guidelines: Keystone Exams and PSSA* (Revised 10/31/2012). Retrieved January 22, 2013, http://www.education.state.pa.us.

Pennsylvania Department of Education. (2013). *2012–2013 PSSA: Handbook for Assessment Coordinators*. Retrieved January 22, 2013, from http://www.education.state.pa.us.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 221–262). Washington, DC: American Council on Education.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*, 271–295.

Stocking, M. L., & Eignor, D. R. (1986). The impact of different ability distributions on IRT preequating (Research Report No. 86–14). Princeton, NJ: Educational Testing Service.

Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

U.S. Department of Education, Office of Elementary and Secondary Education. (2004, April 28). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author.

Valencia, S.W. and Wixson, K.K. (2000). *Policy-oriented research on literary standards and assessment*. In M.L. Kamil, P.B. Mosenthat, P.D. Pearson, and R. Barr (Eds.), Handbook of Reading Research: Vol. III. Mahwah, NJ: Lawrence Erlbaum.

Webb, N. L. (1997). *Criteria for alignment of expectations and tests in mathematics and science education*. Research Monograph No. 6). Madison: University of Wisconsin–Madison, National Institute for Science Education. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1997; 2006). *Research monograph number 6: Criteria for alignment of expectations and assessments on mathematics and science education*. Washington, D.C.: CCSSO.

Webb, N. L. (1999). *18: Alignment of science and mathematics standards and assessments in four states*. Research Monograph No. Madison, WI: National Institute for Science Education.

Webb, N. L. (1999). Research monograph No. 18: Alignment of science and mathematics standards and assessments in four states. Washington, D.C.: CCSSO.

Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and tests for four states: State collaborative on test and state standards (SCASS)*. Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.

Webb, N. L. (November, 2005). *Depth-of-Knowledge levels for four content areas*. Presentation to the Florida Education Research Association, 50th Annual Meeting, Miami, Florida.

Webb, N. L. (2006). *Web alignment tool* [Computer Software]. Madison: Wisconsin Center of Educational Research. University of Wisconsin-Madison.

Webb Alignment Tool (WAT) Training Manual retrieved from http://www.wcer.wisc.edu/WAT/index.aspx.

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

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213.

Zwick, R., & Erickan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*, 55–66.