pennsylvania
DEPARTMENT OF EDUCATION

PENNSYLVANIA
KEYSTONE EXAMS

TECHNICAL REPORT

ALGEBRA I, BIOLOGY, AND LITERATURE

2015

Provided by Data Recognition Corporation

TABLE OF CONTENTS

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Appendices

GLOSSARY OF COMMON TERMS

The following table contains some terms used in this technical report and their meanings. Some of these terms are used universally in the assessment community, and some of these terms are used commonly by psychometric professionals. A glossary of accommodation terms as applied to the Pennsylvania Keystone Exams is provided in Chapter Ten.

Term	Common Definition
Ability	In Rasch scaling, ability is a generic term indicating the level of an individual on the construct measured by an exam. As an example for the Keystone Exams, a student's literature ability is measured by how the student performed on the Literature Keystone Exam. A student who answered more items correctly has a higher ability than a student who answered fewer items correctly.
Adjacent Agreement	A score/rating difference of one (1) point in value usually assigned by two different raters under the same conditions (e.g., two independent raters give the same paper scores that differ by one point).
Alternate Forms	Two or more versions of a test that are considered exchangeable, i.e., they measure the same constructs in the same ways, are intended for the same purposes, and are administered using the same directions. More specific terminology applies depending on the degree of statistical similarity between the test forms (e.g., parallel forms, equivalent forms, and comparable forms) where parallel forms refers to the situation in which the test forms have the highest degree of similarity to each other.
Average	A measure of central tendency in a score distribution that usually refers to the arithmetic mean of a set of scores. In this case, it is determined by adding all the scores in a distribution and then dividing the obtained value by the total number of scores. Sometimes people use the word average to refer to other measures of central tendency such as the median (the score in the middle of a distribution) or mode (the score value with the greatest frequency).
Bias	In a statistical context, bias refers to any source of systematic error in the measurement of a test score. In discussing test fairness, bias may refer to construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers (e.g., gender, ethnicity, etc.). Attempts are made to reduce bias by conducting item fairness reviews and various differential item functioning (DIF) analyses, detecting potential areas of concern, and either removing or revising the flagged test items prior to the development of the final operational form of the test (see also Differential Item Functioning).
Constructed-Response Item	See Open-Ended Item.
Content Validity Evidence	Evidence regarding the extent to which a test provides an appropriate sampling of a content domain of interest (e.g., assessable portions of a state's Grade 6 mathematics curriculum in terms of the knowledge, skills, objectives, and processes sampled).

Glossary of Common Terms

Term	Common Definition
Criterion-Referenced Interpretation	When a score is interpreted as a measure of a student's performance with respect to an expected level of mastery, educational objective, or standard. The types of resulting score interpretations provide information about what a student knows or can do with respect to a given content area.
Cut Score	A specified point on a score scale such that scores at or above that point are interpreted or acted upon differently from scores below that point (e.g., a score designated as the minimum level of performance needed to pass a competency test). One or more cut scores can be set for a test that results in dividing the score range into various proficiency level ranges. Methods for establishing cut scores vary. For the Keystone Exams, three cut scores are used to place students into one of four performance levels (see also Performance Level Setting).
Decision Consistency	The extent to which classifications based on test scores would match the decisions based on scores from a second, parallel form of the same test. It is often expressed as the proportion of examinees who are classified the same way from the two test administrations.
Differential Item Functioning (DIF)	A statistical property of a test item in which different groups of test takers (who have the same total test score) have different average item scores. In other words, students with the same ability level but different group memberships do not have the same probability of answering the item correctly (see also Bias).
Distractor	An incorrect option in a multiple-choice item (also called a foil).
Equating	The strongest of several linking methods used to establish comparability between scores from multiple tests. Equated test scores should be considered exchangeable. Consequently, the criteria needed to refer to a linkage as equating are strong and somewhat complex (equal construct and precision, equity, and invariance). In practical terms, it is often stated that it should be a matter of indifference to a student if he/she takes any of the equated tests (see also Linking).
Exact Agreement	When identical scores/ratings are assigned by two different raters under the same conditions (e.g., two independent raters give a paper the same score).
Field-Test (FT) Items	The Keystone Exams use multiple test forms for each content-area test. Each form is composed of operational (OP) items and field-test (FT) items. An FT item is a newly developed item that is ready to be tried out to determine its statistical properties (see also *P*-value and Point-Biserial Correlation). Each test form includes a set of FT items. FT items are not part of any student scores.
Frequency	The number of times that a certain value or range of values (score interval) occurs in a distribution of scores.
Frequency Distribution	A tabulation of scores from low to high or high to low showing the number and/or percent of individuals who obtain each score or who fall within each score interval or category.

Glossary of Common Terms

Term	Common Definition
Infit/Outfit	Statistical indicators of the agreement of the data and the measurement model (see also Outfit/Infit).
Item Difficulty	For the Rasch model, the dichotomous item difficulty represents the point along the latent trait continuum where an examinee has a 0.50 probability of making a correct response. For a polytomous item, the difficulty is the average of the item's step difficulties (see also Step Difficulty).
Key	The correct response option or answer to a test item.
Linking	A generic term referring to one of a number of processes by which scores from one or more tests are made comparable to some degree. Linking includes several classes of transformations (equating, scale alignment, prediction, etc.). Equating is associated with the strongest degree of comparability (exchangeable scores). Other linkages may be very strong but fail to meet one or more of the strict criteria required of equating (see also Equating).
Logit	In Rasch scaling, logits are units used to express both examinee ability and item difficulty. When expressing examinee ability, a student who answers more items correctly has a higher logit than a student who answers fewer items correctly. Logits are transformed into scaled scores through a linear transformation. When expressing item difficulty, logits are transformed *p*-values (see also *P*-value). The logit difficulty scale is inversely related to *p*-values. A higher logit value would represent a relatively harder item, while a lower logit value would represent a relatively easier item.
Mean	Also referred to as the arithmetic mean of a set of scores, is found by adding all the score values in a distribution and dividing by the total number of scores. For example, the mean of the set {66, 76, 85, 97} is 81. The value of a mean can be influenced by extreme values in a score distribution.
Measure	In Rasch scaling, measure generally refers to a specific estimate of an examinee's ability (often expressed as logits) or an item's difficulty (again, often expressed as logits). As an example for the Keystone Exams, a student's literature measure might be equal to 0.525 logits. Or, a Literature Keystone Exam test item might have logit equal to -0.905.
Median	The middle point or score in a set of rank-ordered observations that divides the distribution into two equal parts such that each part contains 50 percent of the total data set. More simply put, half of the scores are below the median value and half of the scores are above the median value. As an example, the median for the following ranked set of scores {2, 3, 6, 8, 9} is 6.
Module	On score reports, a module often refers to a set of items on a test measuring the same contextual area (e.g., Operations and Linear Equations & Inequalities in Algebra I). Items developed to measure the same reporting category would be used to determine the module score (sometimes called "subscale" score).

Glossary of Common Terms

Term	Common Definition
Multiple-Choice Item	A type of item format that requires the test taker to select a response from a group of possible choices, one of which is the correct answer (or key) to the question posed (see also Open-Ended Item).
N-count	Sometimes designated as *N* or *n*, it is the number of observations (usually individuals or students) in a particular group. Some examples include the number of students tested, the number of students tested from a specific subpopulation (e.g., females), the number of students who attained a specific score, etc. In the following set {23, 32, 56, 65, 78, 87}, $n = 6$.
Open-Ended Item	An open-ended (OE) item—referred to by some as a constructed-response (CR) item—is an item format that requires examinees to create their own responses, which can be expressed in various forms (e.g., written paragraph, created table/graph, formulated calculation, etc.). Such items are frequently scored using more than two score categories, that is, polytomously (e.g., 0, 1, 2, and 3). This format is in contrast to when students make a choice from a supplied set of answer options (e.g., multiple-choice (MC) items which are typically dichotomously scored as right = 1 or wrong = 0). When interpreting item difficulty and discrimination indices it is important to consider whether an item is polytomously or dichotomously scored.
Operational Item	The Keystone Exams use multiple test forms for each content-area test. Each form is composed of operational (OP) items and field-test (FT) items. OP items are the same on all forms for any content-area test. Student total raw scores and scaled scores are based exclusively on the OP items.
Outfit/Infit	Statistical indicators of the agreement of the data and the measurement model. Infit and outfit are highly correlated, and both are highly correlated with the point-biserial correlation. Underfit can be caused when low-ability students correctly answer difficult items (perhaps by guessing or atypical experience) or high-ability students incorrectly answer easy items (perhaps because of carelessness or gaps in instruction). Any model expects some level of variability, so overfit can occur when nearly all low-ability students miss an item while nearly all high-ability students get the item correct.
Percent Correct	When referring to an individual item, the percent correct is the item's *p*-value expressed as a percent (instead of a proportion). When referring to a total test score, it is the percentage of the total number of points that a student received. The percent correct score is obtained by dividing the student's raw score by the total number of possible points and multiplying the result by 100. Percent correct scores are often used in criterion-referenced interpretations and are generally more helpful if the overall difficulty of a test is known. Sometimes percent correct scores are incorrectly interpreted as percentile ranks.

Glossary of Common Terms

Term	Common Definition
Percentile	The score or point in a score distribution at or below which a given percentage of scores fall. It should be emphasized that it is a value on the score scale, not the associated percentage (although sometimes in casual usage this misinterpretation is made). For example, if 72 percent of the students score at or below a scaled score of 1500 on a given test, then the scaled score of 1500 would be considered the 72nd percentile. As another example, the median is the 50th percentile.
Percentile Rank	The percentage of scores in a specified distribution falling at/below a certain point on a score distribution. Percentile ranks range in value from 1 to 99, and indicate the status or relative standing of an individual within a specified group by indicating the percent of individuals in that group who obtained equal or lower scores. An individual's percentile rank can vary depending on which group is used to determine the ranking. As suggested above, percentiles and percentile rank are sometimes used interchangeably; however, strictly speaking, a percentile is a value on the score scale.
Performance Level Descriptors	Descriptions of an individual's competency in a particular content area, usually defined as ordered categories on a continuum, often labeled from Below Basic to Advanced, that constitute broad ranges for classifying performance. The exact labeling of these categories, and narrative descriptions, may vary from one assessment or testing program to another.
Performance Level Setting	Also referred to as standard setting, a procedure used in the determination of the cut scores for a given assessment that is used to measure students' progress towards certain performance standards. Standard setting methods vary (e.g., modified Angoff, Bookmark Method, etc.), but most use a panel of educators and expert judgments to operationalize the level of achievement students must demonstrate in order to be categorized within each performance level.
Point-Biserial Correlation	In classical test theory this is an item discrimination index. It is the correlation between a dichotomously scored item and a continuous criterion, usually represented by the total test score (or the corrected total test score with the reference item removed). It reflects the extent to which an item differentiates between high-scoring and low-scoring examinees. This discrimination index ranges from –1.00 to +1.00. The higher the discrimination index (the closer to +1.00), the better the item is considered to be performing. For multiple-choice items scored as 0 or 1, it is rare for the value of this index to exceed 0.5.
P-value	An index indicating an item's difficulty for some specified group (perhaps grade). It is calculated as the proportion (sometimes percent) of students in the group who answer an item correctly. *P*-values range from 0.0 to 1.0 on the proportion scale. Lower values correspond to more difficult items and higher values correspond to easier items. *P*-values are usually provided for multiple-choice items or other items worth one point. For open-ended items or items worth more than one point, difficulty on a *p*-value-like scale can be estimated by dividing the item mean score by the maximum number of points possible for the item (see also Logit).

Glossary of Common Terms

Term	Common Definition
Raw Score	Sometimes abbreviated by RS—it is an unadjusted score usually determined by tallying the number of questions answered correctly, or by the sum of item scores (i.e., points). (Some rarer situations might include formula-scoring, the amount of time required to perform a task, the number of errors, application of basal/ceiling rules, etc.) Raw scores typically have little or no meaning by themselves and require additional information—like the number of items on the test, the difficulty of the test items, norm-referenced information, or criterion-referenced information.
Reliability	The expected degree to which test scores for a group of examinees are consistent over exchangeable replications of an assessment procedure, and therefore, are considered dependable and repeatable for an individual examinee. A test that produces highly consistent, stable results (i.e., relatively free from random error) is said to be highly reliable. The reliability of a test is typically expressed as a reliability coefficient or by the standard error of measurement derived by that coefficient.
Reliability Coefficient	A statistical index that reflects the degree to which scores are free from random measurement error. Theoretically, it expresses the consistency of test scores as the ratio of true score variance to total score variance (true score variance plus error variance). This statistic is often expressed as correlation coefficient (e.g., correlation between two forms of a test) or with an index that resembles a correlation coefficient (e.g., calculation of a test's internal consistency using coefficient alpha). Expressed this way, the reliability coefficient is a unitless index. The higher the value of the index (closer to 1.0), the greater the reliability of the test (see also Standard Error of Measurement).
Scaled Score	A mathematical transformation of a raw score developed through a process called scaling. Scaled scores are most useful when comparing test results over time. Several different methods of scaling exist, but each is intended to provide a continuous and meaningful score scale across different forms of a test.
Selected-Response Item	See Multiple-Choice Item.
Spiraling	A packaging process used when multiple forms of a test exist and it is desired that each form be tested in all classrooms (or other grouping unit (e.g., schools)) participating in the testing process. This process allows for the random distribution of test booklets to students. For example, if a package has four test forms labeled A, B, C, and D, the order of the test booklets in the package would be A, B, C, D, A, B, C, D, A, B, C, D, etc.
Standard Deviation (SD)	A statistic that measures the degree of spread or dispersion of a set of scores. The value of this statistic is always greater than or equal to zero. If all of the scores in a distribution are identical, the standard deviation is equal to zero. The further the scores are away from each other in value, the greater the standard deviation. This statistic is calculated using the information about the deviations (distances) between each score and the distribution's mean. It is equivalent to the square root of the variance statistic. The standard deviation is a commonly used method of examining a distribution's variability since the standard deviation is expressed in the same units as the data.

Glossary of Common Terms

Term	Common Definition
Standard Error of Measurement (SEM)	The amount an observed score is expected to fluctuate around the true score. As an example, across replications of a measurement procedure, the true score will not differ by more than plus or minus one standard error from the observed score about 68 percent of the time (assuming normally distributed errors). The SEM is frequently used to obtain an idea of the consistency of a person's score in actual score units or to set a confidence band around a score in terms of the error of measurement. Often a single SEM value is calculated for all test scores. On other occasions, however, the value of the SEM can vary along a score scale. Conditional standard errors of measurement (CSEMs) provide an SEM for each possible scaled score.
Step Difficulty	Step difficulty is a parameter estimate in Master's partial credit model (PCM) that represents the relative difficulty of each score step (e.g., going from a score of 1 to a score of 2). The higher the value of a particular step difficulty, the more difficult a particular step is relative to other score steps (e.g., is it harder to go from a 1 to a 2, or to go from a 2 to a 3).
Strand	On score reports, a strand often refers to a set of items on a test measuring the same contextual area (e.g., Number Sense in Mathematics). Items developed to measure the same reporting category would be used to determine the strand score (sometimes called "subscale" score).
Technical Advisory Committee (TAC)	A group of individuals, most often professionals in the field of testing, who are either appointed or selected to make recommendations for and to guide the technical development of a given testing program.
Validity	The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by the purposed uses of a test. There are various ways of gathering validity evidence.

Glossary of Common Terms

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 paper/pencil), collect, score, analyze, track, and report results of graduation competency assessments for 10 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 Exams served as the final course exams. Students’ Keystone Exam scores counted 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.

Preface: An Overview of the Assessments

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 Exam 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 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 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 2010–11 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)

Preface: An Overview of the Assessments

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 Three. 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)

Preface: An Overview of the Assessments

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 the 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 Exams 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 10 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 on 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 EXAMS 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 Exams 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 10 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 the fall of 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 the spring of 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 2015.

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	TBD
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 winter and summer administrations can be found in Chapter Three.

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

As shown in Table 2–1, 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 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/KeystoneDownloads.

HIGH-LEVEL TEST DESIGN CONSIDERATIONS

The Keystone Exams employs 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 an item 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 Exams, 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 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 Algebra I Keystone 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 content included in the 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 Keystone 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 Keystone 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 10 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 I. 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 I 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.pa.gov*

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 Biology Keystone 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 content included in the 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 Keystone 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 Keystone 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.pa.gov*

LITERATURE

The Literature Keystone 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 content included in the 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 Keystone 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 Keystone 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.pa.gov*

LITERATURE PASSAGES

One of the key requirements of the Literature Keystone 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 Assessment Anchors and Eligible Content for Literature 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 US 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.

MATCHING READER TO TEXT AND TASK

A number of factors are taken into consideration when deciding whether a passage will be placed in the pool for possible use on the Literature Keystone 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 Exams 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 passages to be included in the pool of passages for possible use on the Literature Keystone 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 Literature Keystone 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 Literature Keystone Exam are the Dale-Chall Formula, the Flesch Grade Level Formula, and the Fry Graph.

Chapter Two: Test Development Overview of the Keystone Exams

CHAPTER THREE: ITEM AND TEST DEVELOPMENT PROCESSES

GENERAL TEST DEVELOPMENT PROCESSES

The 2015 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. To accommodate the program hiatus, the 2015 Keystone Exam cores were made up of items that had appeared on the Spring 2012 and/or Winter 2013/2014 cores. The remainder of the operational 2015 exams were made up of items that were field tested on the Spring 2014 Keystone Exams embedded field test administration. Table 3–1 below 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 Occurring in Calendar Year							
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016***
2010–2011	Initial Item Development for Fall 2010 FT	**Fall 2010 Stand-alone FT** New Item Development for Spring 2011 FT	Data Review of Fall 2010 FT **Spring 2011 Operational & Embedded FT**	Data Review 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 Operational Administrations in 2011–2012 (Program in Hiatus)**							
2012–2013			New Item Development for Spring 2013 FT	**Winter 2012/13 Administration**	**Spring 2013 Operational & Embedded FT** Data Review of Spring 2013 FT **Summer 2013 Administration**		Biennial Core-to-Core Overlap (2013 core included as a portion of the 2015 core)	
2013–2014					**Winter 2013/14 Administration** New Item Development for Spring 2014 FT	**Spring 2014 Operational & Embedded FT** Data Review of Spring 2014 FT **Summer 2014 Administration**		
2014–2015						**Winter 2014/15 Administration** New Item Development for Spring 2015 FT	**Spring 2015 Operational & Embedded FT** Data Review of Spring 2015 FT **Summer 2015 Administration**	
2015–2016*							**Winter 2015/16 Administration** New Item Development for Spring 2016 FT	**Spring 2016 Operational & Embedded FT** Data Review of Spring 2016 FT **Summer 2016 Administration**

*Projected/scheduled tasks and activities

GENERAL TEST DEFINITION

The plan for the Keystone Exams 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 2015, 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 in 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 2015 Algebra I Keystone Exam was composed of 24 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. Tables 3–2 and 3–3 display the design for Algebra I for forms 1 through 24. 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 2015) per Operational Form

	Module	Core per Form		Field Test per Form		Total per Form	
		MC Items	CR Items	MC Items	CR Items	Core & FT MC Items	Core & FT CR Items
Algebra I	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 2015) per 24 Operational Forms

	Module	Core per 24 Forms		Field Test per 24 Forms		Total per 24 Forms	
		MC Items	CR Items	MC Items	CR Items	Core & FT MC Items	Core & FT CR Items
Algebra I	1	18	3	120	24	276	54
	2	18	3	120	24		
	Total	36	6	240	48		

The operational (core) portions of the Winter 2014/2015 and Summer 2015 administrations came from the same sources as the Spring 2013 core. Therefore 30% to 50% of the Winter 2014/2015 core overlaps with Spring 2013 and/or Winter 2013/2014 cores. The remaining core items that appeared on the Winter 2014/2015 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 (2015 Summer, Winter, and Breach) per Operational Form

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

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 shown in Table 3–5.

Table 3–5. Algebra I Core Points

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

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 Table 3–6.

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 2015 Biology Keystone Exam was composed of 24 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. Tables 3–7 and 3–8 display the design for Biology for forms 1 through 24. 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 2015) per Operational Form

	Module	Core per Form		Field Test per Form		Total per Form	
		MC Items	CR Items	MC Items	CR Items	Core & FT MC Items	Core & FT CR Items
Biology	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 2015) per 24 Operational Forms

	Module	Core per 24 Forms		Field Test per 24 Forms		Total per 24 Forms	
		MC Items	CR Items	MC Items	CR Items	Core & FT MC Items	Core & FT CR Items
Biology	1	24	3	192	24		
	2	24	3	192	24	432	54
	Total	48	6	384	48		

The operational (core) portions of the Winter 2014/2015 and Summer 2015 administrations came from the same sources as the Spring 2013 core. Therefore 30% to 50% of the Winter 2014/2015 core overlaps with Spring 2013 and/or Winter 2013/2014 cores. The remaining core items that appeared on the Winter 2014/2015 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 (2015 Summer, Winter, and Breach) per Operational Form

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

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 shown in Table 3–10.

Table 3–10. Biology Core Points

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

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

3. Cells and Cell Processes
4. Continuity and Unity of Life

The distribution of Biology items into these two categories is shown in Table 3–11.

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 2015 Literature Keystone Exam was composed of 24 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. Tables 3–12 and 3–13 display the design for Literature for forms 1 through 24. 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 2015) per Operational Form

Literature	Module	Core per Form			Field Test per Form			Total per Form		
		Passages	MC Items	CR Items	Passages	MC Items	CR Items	Core & FT Passages	Core & FT MC Items	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 2015) per 24 Operational Forms

Literature	Module	Core per 24 Forms			Field Test per 24 Forms			Total per 24 Forms		
		Passages	MC Items	CR Items	Passages	MC Items	CR Items	Core & FT Passages	Core & FT MC Items	Core & FT CR Items
	1	2	17	*3	12	144	24	28	322	54
	2	2	17	*3	12	144	24			
	Total	4	34	6	24	288	48			

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

The operational (core) portions of the Winter 2014/2015 and Summer 2015 administrations came from the same sources as the Spring 2013 core. Therefore 30% to 50% of the Winter 2014/2015 core overlaps with Spring 2013 and/or Winter 2013/2014 cores. The remaining core items that appeared on the Winter 2014/2015 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 items. Table 3–14 displays the design for the Literature Summer, Winter, and Breach operational forms.

Table 3–14. Literature Test Plan (2015 Summer, Winter, and Breach) per Operational Form

Literature	Module	Core per Form		Placeholders per Form		Total per Form		Number of Forms	
		MC Items	CR Items	MC Items	CR Items	Core & PH MC Items	Core & PH CR Items	Master Core	Scrambles
	1	17	*3	6	1	23	4	1	2
	2	17	*3	6	1	23	4		
	Total	34	6	12	2	46	8		

*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 shown in Table 3–15.

Table 3–15. Literature Core Points

Literature	Module 1			Module 2			Literature Total		
	50%			50%			100%		
	Passages	MC Items	CR Items	Passages	MC Items	CR Items	Passages	MC Items	CR Items
Core Items	2	17	3	2	17	3	4	34	6
Core Points		17	9		17	9		34	18
Total Points	26			26			52		

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 Table 3–16.

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 Exams 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

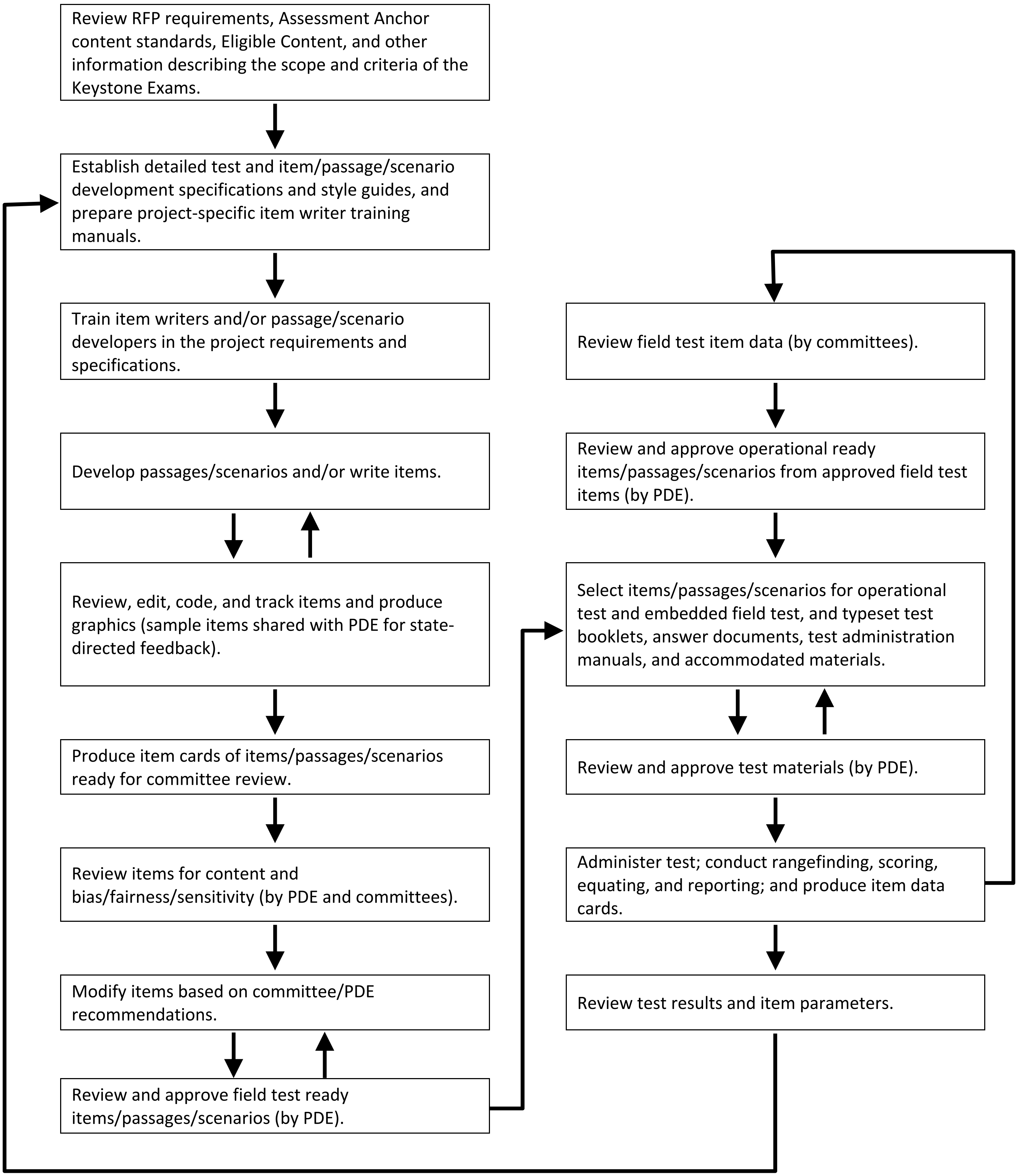
The item development process followed a logical cycle and timeline, which are outlined in Table 3–17 and Figure 3–1. On the front end of the schedule, tasks were generally completed with the goal of presenting field test candidate items to committees of Pennsylvania educators. On the back end of the 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 Exam field test event, not a stand-alone field test event or an accelerated development cycle.

The process flowchart, Figure 3–1, 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 Exam Administrations

Cycle	Steps in Development Life Cycle	Timeline	Approximate Window	
Primary	Development Planning	Summer/Fall	Month 1–4	Jul–Oct
	Literature Passage Selection	Summer/Fall	Month 1–6	Jul–Dec
	Item Writer Training	Fall	Month 5	Nov
	Initial Item Authoring	Fall/Winter	Month 5–9	Nov–Mar
	Internal Reviews and PDE Reviews	Fall to Spring	Month 6–12	Dec–Jun
	Bias, Fairness, and Sensitivity Review	Summer	Month 13	Jul
	New Item Content Committee Review (PA Educators)	Summer	Month 13	Jul
	Post-Review Resolution and Cleanup	Summer	Month 13–14	Jul–Aug
	Build Field Test Forms	Summer/Fall	Month 15–16	Sep–Oct
	Internal Form Reviews and PDE Reviews	Summer/Fall	Month 15–18	Sep–Dec
	Final Form and Printer Proof Approvals	Fall/Winter	Month 18–19	Dec–Jan
	Ancillary and Accommodated Form Development	Fall/Winter	Month 18–20	Dec–Feb
	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 19–22	Jan–Apr
	Field Test Administration	Spring	Month 23	May
	Material/Data Processing, Rangefinding, and Scoring	Spring/Summer	Month 23–26	May–Aug
	Field Test Item Data Review (PA Educators)	Summer	Month 27	Sept
	Select Operational Items	Summer/Fall	Month 27–28	Sep–Oct
	Build Operational Forms	Fall	Month 28–29	Oct–Nov
	Internal Form Reviews and PDE Reviews	Fall	Month 29–30	Nov–Dec
	Final Form and Printer Proof Approvals	Fall/Winter	Month 30–31	Dec–Jan
	Ancillary and Accommodated Form Development	Fall/Winter	Month 31–33	Jan–Mar
	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 31–33	Jan–Mar
	Operational Test Administration	Spring	Month 35	May
	Material/Data Processing and Scoring	Spring/Summer	Month 35–36	May–Jun
	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
	Build Operational Forms	Fall/Winter	Month 52–53	Oct–Nov
	Internal Form Reviews and PDE Reviews	Winter	Month 53–54	Nov–Dec
	Final Form and Printer Proof Approvals	Winter	Month 54–55	Dec–Jan
	Ancillary and Accommodated Form Development	Winter/Spring	Month 55–57	Jan–Mar
	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 56–58	Feb–Apr
	Second Operational Test Administration	Spring	Month 59	May
	Material/Data Processing and Scoring	Spring/Summer	Month 59–60	May–Jun
	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 WestEd—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 Exams 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 Exam Item Cards, which allow for preliminary sorting and reviewing. Although very similar, the Keystone Exam Item Cards for Multiple-Choice Items differ from the Keystone Exam Item Cards for Constructed-Response Items in that the former have 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, at WestEd, or at both companies. 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 Exam 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 2014

Prior to the 2015 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, or a flawed graphic in an item, or if an item requires specialized, non-content 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 July 28–August 1, 2014 for Biology and August 4–7, 2014, for Algebra I 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 JULY 2014

Prior to the 2015 field testing, all newly developed test items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. This review took place from July 23–25, 2014, 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, multi-ethnic 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, 2013). 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, except that the materials were locked up and stored at the DRC offices in Harrisburg. Table 3–18 shows the gender and race/ethnicity of the members of the bias committee who reviewed the Keystone Exam items and passages for bias, fairness, and sensitivity.

Table 3–18. Demographic Composition of the 2014 Keystone Exam Bias, Fairness, and Sensitivity Committee

Member #	July 2014	
	Gender	Race/Ethnicity
1.	Female	Asian American
2.	Male	Asian American
3.	Female	Hispanic American
4.	Male	Hispanic American
5.	Female	Caucasian American
6.	Male	Caucasian American
7.	Male	African American
8.	Female	African American
Totals	4 Male 4 Female	2 Asian Americans 2 African Americans 2 Caucasian Americans 2 Hispanic Americans

The results from the 2014 Bias, Fairness, and Sensitivity Committee reviews are summarized in Table 3–19.

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

Date	Algebra I			
	Total Reviewed	Accepted As Is	Accepted with Revision	Rejected
July 2014	332	320	12	0
Date	**Biology**			
	Total Reviewed	**Accepted As Is**	**Accepted with Revision**	**Rejected**
July 2014	22 Scenarios 606 Items	21 Scenarios 605 Items	1 Scenario 1 Item	0 Scenarios 0 Items
Date	**Literature**			
	Total Reviewed	**Accepted As Is**	**Accepted with Revision**	**Rejected**
July 2014	30 Passages 382 Items	30 Passages 382 Items	0 Passages 0 Items	0 Passages 0 Items
Date	**All Exams**			
	Total Reviewed	**Accepted As Is**	**Accepted with Revision**	**Rejected**
July 2014	22 Scenarios 30 Passages 1,320 Items	21 Scenarios 30 Passages 1307 Items	1 Scenario 0 Passages 13 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 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 Exams 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 use 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.

[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.

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 US 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 and WestEd 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 and WestEd 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).

To this end, DRC and WestEd embrace 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 and WestEd to limit item types that are difficult to format in Braille and that may become distorted when published in large print. DRC and WestEd are 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-language versions of the exams are discussed in Chapter Six.

CHAPTER FIVE: FIELD TEST LEADING TO THE SPRING 2015 CORE

FIELD TEST OVERVIEW

Approximately 50% of the core items appearing on the Spring 2015 Pennsylvania Keystone Exams came from the field test (nonlinking) items on the Spring 2014 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 2014 Keystone Exams were selected for use as common (core) items in the 2015 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

Field Test Administration	Initial Core Use			Core-to-Core Overlap Use		
	Spring	Summer	Winter	Spring	Summer	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 2014 KEYSTONE EXAMS EMBEDDED FIELD TEST

For 2014, the embedded field test (in spring) was designed to yield enough items to construct portions of the following operational forms: Spring 2015, Summer 2015, Winter 2015/2016, and a possible breach form. The next tables describe the embedded field test plans for the Keystone Exams in the spring of 2014.

SPRING 2014 ALGEBRA I KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2014 Algebra I Keystone Exam was composed of 24 forms. All of the forms contained core items that were identical for all students and sets of generally unique items. Tables 5–2 and 5–3 display the design for Algebra I for forms 1 through 24. 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 2014) per Operational Form

<table>
<tr><th rowspan="5">Algebra I</th><th rowspan="2">Module</th><th colspan="2">Core per Form</th><th colspan="2">Field Test per Form</th><th colspan="2">Total per Form</th></tr>
<tr><th>MC Items</th><th>CR Items</th><th>MC Items</th><th>CR Items</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>18</td><td>3</td><td>5</td><td>1</td><td>23</td><td>4</td></tr>
<tr><td>2</td><td>18</td><td>3</td><td>5</td><td>1</td><td>23</td><td>4</td></tr>
<tr><td>Total</td><td>36</td><td>6</td><td>10</td><td>2</td><td>46</td><td>8</td></tr>
</table>

Table 5–3. Algebra I Test Plan (Spring 2014) per 24 Operational Forms

<table>
<tr><th rowspan="5">Algebra I</th><th rowspan="2">Module</th><th colspan="2">Core per 24 Forms</th><th colspan="2">Field Test per 24 Forms</th><th colspan="2">Total per 24 Forms</th></tr>
<tr><th>MC Items</th><th>CR Items</th><th>MC Items</th><th>CR Items</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>18</td><td>3</td><td>100</td><td>20</td><td rowspan="3">236</td><td rowspan="3">46</td></tr>
<tr><td>2</td><td>18</td><td>3</td><td>100</td><td>20</td></tr>
<tr><td>Total</td><td>36</td><td>6</td><td>200</td><td>40</td></tr>
</table>

SPRING 2014 BIOLOGY KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2014 Biology Keystone Exam was composed of 24 forms. All of the forms contained core items that were identical for all students and sets of generally unique items. Tables 5–4 and 5–5 display the design for Biology for forms 1 through 24. 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 2014) per Operational Form

<table>
<tr><th rowspan="5">Biology</th><th rowspan="2">Module</th><th colspan="2">Core per Form</th><th colspan="2">Field Test per Form</th><th colspan="2">Total per Form</th></tr>
<tr><th>MC Items</th><th>CR Items</th><th>MC Items</th><th>CR Items</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>24</td><td>3</td><td>8</td><td>1</td><td>32</td><td>4</td></tr>
<tr><td>2</td><td>24</td><td>3</td><td>8</td><td>1</td><td>32</td><td>4</td></tr>
<tr><td>Total</td><td>48</td><td>6</td><td>16</td><td>2</td><td>64</td><td>8</td></tr>
</table>

Table 5–5. Biology Test Plan (Spring 2014) per 24 Operational Forms

<table>
<tr><th rowspan="5">Biology</th><th rowspan="2">Module</th><th colspan="2">Core per 24 Forms</th><th colspan="2">Field Test per 24 Forms</th><th colspan="2">Total per 24 Forms</th></tr>
<tr><th>MC Items</th><th>CR Items</th><th>MC Items</th><th>CR Items</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>24</td><td>3</td><td>160</td><td>20</td><td rowspan="3">368</td><td rowspan="3">46</td></tr>
<tr><td>2</td><td>24</td><td>3</td><td>160</td><td>20</td></tr>
<tr><td>Total</td><td>48</td><td>6</td><td>320</td><td>40</td></tr>
</table>

SPRING 2014 LITERATURE KEYSTONE EXAM EMBEDDED FIELD TEST PLAN

The Spring 2014 Literature Keystone Exam was composed of 24 forms. All of the forms contained common items that were identical for all students and sets of generally unique items. Tables 5–6 and 5–7 display the design for Literature for forms 1 through 24. 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 2014) per Operational Form

<table>
<tr><th rowspan="5">Literature</th><th rowspan="2">Module</th><th colspan="3">Core per Form</th><th colspan="3">Field Test per Form</th><th colspan="3">Total per Form</th></tr>
<tr><th>Passages</th><th>MC Items</th><th>CR Items</th><th>Passages</th><th>MC Items</th><th>CR Items</th><th>Core & FT Passages</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>2</td><td>17</td><td>3</td><td>1</td><td>6</td><td>1</td><td>3</td><td>23</td><td>4</td></tr>
<tr><td>2</td><td>2</td><td>17</td><td>3</td><td>1</td><td>6</td><td>1</td><td>3</td><td>23</td><td>4</td></tr>
<tr><td>Total</td><td>4</td><td>34</td><td>6</td><td>2</td><td>12</td><td>2</td><td>6</td><td>46</td><td>8</td></tr>
</table>

Table 5–7. Literature Test Plan (Spring 2014) per 24 Operational Forms

<table>
<tr><th rowspan="5">Literature</th><th rowspan="2">Module</th><th colspan="3">Core per 24 Forms</th><th colspan="3">Field Test per 24 Forms</th><th colspan="3">Total per 24 Forms</th></tr>
<tr><th>Passages</th><th>MC Items</th><th>CR Items</th><th>Passages</th><th>MC Items</th><th>CR Items</th><th>Core & FT Passages</th><th>Core & FT MC Items</th><th>Core & FT CR Items</th></tr>
<tr><td>1</td><td>2</td><td>17</td><td>3</td><td>10</td><td>120</td><td>20</td><td rowspan="3">24</td><td rowspan="3">274</td><td rowspan="3">46</td></tr>
<tr><td>2</td><td>2</td><td>17</td><td>3</td><td>10</td><td>120</td><td>20</td></tr>
<tr><td>Total</td><td>4</td><td>34</td><td>6</td><td>20</td><td>240</td><td>40</td></tr>
</table>

STATISTICAL ANALYSES AND RESULTS

All field test items were analyzed statistically following conventional item analysis methods. For MC items, the traditional or classical item statistics included the corrected item-total test correlation 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 non-scorable 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 item-total correlation coefficient. The item-total correlation 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 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, 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, if any item had a gender (male vs. female), ethnicity (white vs. black or Hispanic), and/or test administration mode (PPT vs. CBT) DIF code of C+ or C-, it was flagged.

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

1. Point-biserial correlation for the correct response of less than 0.25
2. Point-biserial correlation for any incorrect response greater than 0.0
3. Percent correct less than 0.3 or greater than 0.9
4. Percentage responding to any incorrect responses greater than the percent correct

For a CR item to be flagged, the following criterion was used in addition to the DIF criteria:

5. 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 2014 field test item analysis can be found in Chapter 9 of the *2014 Pennsylvania Keystone Exams Technical Report* (Pennsylvania Department of Education, 2014). Overall, the samples used to analyze the field test items embedded in 24 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*	Algebra I			Biology			Literature		
		Total *N*	*N*	%	Total *N*	*N*	%	Total *N*	*N*	%
MC	1	240	88	36.7	384	74	19.3	288	63	21.9
	2		70	29.2		78	20.3		39	13.5
	3		48	20.0		27	7.0		21	7.3
	4		55	22.9		47	12.2		29	10.1
CR	5	48	28	58.3	48	9	18.8	48	1	2.1

* See Section Criteria Used to Flag Items for what 1–5 stands for.

Table 5–9. DIF Summary — MC Items

Reference Group	Focal Group	Bias Code	Algebra I (Total *N*=240)		Biology (Total *N*=384)		Literature (Total *N*=288)	
			N	**%**	***N***	**%**	***N***	**%**
Male	Female	A-	128	53.3	188	49.0	158	54.9
		A+	105	43.8	194	50.5	114	39.6
		B-	3	1.3	2	0.5	12	4.2
		B+	2	0.8	0	0.0	0	0.0
		C-	2	0.8	0	0.0	4	1.4
		C+	0	0.0	0	0.0	0	0.0
White	Black	A-	156	65.0	240	62.5	202	70.1
		A+	79	32.9	136	35.4	48	16.7
		B-	3	1.3	7	1.8	32	11.1
		B+	0	0.0	0	0.0	1	0.3
		C-	2	0.8	1	0.3	5	1.7
		C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	150	62.5	263	68.5	202	70.1
		A+	84	35.0	113	29.4	52	18.1
		B-	5	2.1	8	2.1	28	9.7
		B+	0	0.0	0	0.0	0	0.0
		C-	1	0.4	0	0.0	6	2.1
		C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	89	37.1	171	44.5	73	25.3
		A+	150	62.5	210	54.7	202	70.1
		B-	0	0.0	0	0.0	1	0.3
		B+	1	0.4	3	0.8	11	3.8
		C-	0	0.0	0	0.0	0	0.0
		C+	0	0.0	0	0.0	1	0.3

Table 5–10. DIF Summary — CR Items

Reference Group	Focal Group	Bias Code	Algebra I (Total *N*=48)		Biology (Total *N*=48)		Literature (Total *N*=48)	
			N	%	*N*	%	*N*	%
Male	Female	A-	21	43.8	11	22.9	0	0.0
		A+	26	54.2	30	62.5	15	31.3
		B-	1	2.1	0	0.0	0	0.0
		B+	0	0.0	5	10.4	21	43.8
		C-	0	0.0	0	0.0	0	0.0
		C+	0	0.0	2	4.2	12	25.0
White	Black	A-	39	81.3	30	62.5	16	33.3
		A+	3	6.3	7	14.6	30	62.5
		B-	4	8.3	8	16.7	1	2.1
		B+	0	0.0	0	0.0	1	2.1
		C-	2	4.2	3	6.3	0	0.0
		C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	39	81.3	35	72.9	24	50.0
		A+	6	12.5	9	18.8	21	43.8
		B-	3	6.3	1	2.1	1	2.1
		B+	0	0.0	0	0.0	2	4.2
		C-	0	0.0	3	6.3	0	0.0
		C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	37	77.1	24	50.0	38	79.2
		A+	6	12.5	21	43.8	7	14.6
		B-	5	10.4	1	2.1	2	4.2
		B+	0	0.0	1	2.1	0	0.0
		C-	0	0.0	1	2.1	1	2.1
		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 2014 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 Pennsylvania educators (teachers and PDE staff) broken out into exam-based committees. The review took place on September 9 and 10, 2014. 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 field test items and made recommendations on each item. The results of the committee reviews are shown in Table 5–11. Further discussion on how this information was used is covered in Chapter Six.

Table 5–11. Spring 2014 Keystone Exam Data Review Results

Exam	Type	Number of Items in Spring 2014 EFT	Flagged Items in Spring 2014 EFT Examined at Sept 2014 Data Review Committee		Flagged Items in Spring 2014 EFT Rejected by Sept 2014 Data Review Committee		Items Classified as "Rejected" from Spring 2014 EFT (all sources)*	
			Number of Items	% of Field Test	Number of Items	% of Field Test	Number of Items	% of Field Test
Algebra I	MC	240	107	45%	34	14%	34	14%
	CR	48	29	60%	13	27%	13	27%
Biology	MC	384	108	28%	35	9%	35	9%
	CR	48	15	31%	3	6%	5	10%
Literature	MC	288	83	29%	15	5%	17	6%
	CR	48	14	29%	0	0%	0	0%
Subtotal	MC	912	298	33%	84	9%	86	9%
	CR	144	58	40%	16	11%	18	13%
	Total	1056	356	34%	100	9%	104	10%

*Data Review Committee, PDE, and DRC

CHAPTER SIX: OPERATIONAL FORMS CONSTRUCTION FOR 2015 ADMINISTRATIONS

FINAL SELECTION OF ITEMS AND FORMS CONSTRUCTION

Approximately 50% of the items that made up the Winter 2014/2015, Spring 2015, and Summer 2015 operational forms emerged from the Spring 2014 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) and WestEd 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 multi-ethnic 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 content-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 2015 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 Exams 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.

Chapter Six: Operational Forms Construction for 2015 Administrations

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 performs 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 replacing 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 Eight for more information about the steps following the final approval to print.

SPECIAL FORMS USED WITH THE OPERATIONAL 2015 KEYSTONE EXAMS

Four accommodated forms are available for the operational Keystone Exams: audio, Braille, large print, and Spanish translation. Audio versions and Spanish versions are available for Algebra I and Biology. Braille and large print are available for all operational exams. Braille, large print, and Spanish translation versions are available in the print mode of delivery. The audio version is only available in the online mode of delivery. The tables below summarize the usage for these four accommodated forms. Refer to the information about test accommodations in Chapter Ten for more information about this topic.

Table 6–1. Special Forms Usage Summary from Operational Spring 2015 Keystone Exams

Exam	Online Delivery	Print Form Delivery			All
	Audio	Braille	Large Print	Spanish	
Algebra I	701	9	92	1065	1867
Biology	957	7	75	643	1682
Literature		6	65		71
Total	1658	22	232	1708	3620
Total Administrations by Delivery Method	59,157	409,180			468,337
% of Total Administrations by Delivery Method	2.80%	0.005%	0.056%	0.417%	0.772%

Table 6–2. Special Forms Usage Summary from Operational Winter 2014/2015 Keystone Exams

Exam	Online Delivery	Print Form Delivery			All
	Audio	Braille	Large Print	Spanish	
Algebra I	308	5	26	467	806
Biology	565	4	22	256	847
Literature		4	18		22
Total	873	13	66	723	1675
Total Administrations by Delivery Method	26,736	145,296			172,032
% of Total Administrations by Delivery Method	3.03%	0.009%	0.045%	0.498%	0.974%

Table 6–3. Special Forms Usage Summary from Operational Summer 2015 Keystone Exams

Exam	Online Delivery	Print Form Delivery			All
	Audio	Braille	Large Print	Spanish	
Algebra I	8	0	1	0	9
Biology	2	0	0	0	2
Literature		0	0		0
Total	10	0	1	0	11
Total Administrations by Delivery Method	901	2,560			3,461
% of Total Administrations by Delivery Method	1.11%	0.000%	0.039%	0.000%	0.318%

On a yearly basis, the PDE examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. Accommodations manuals for Pennsylvania assessments titled *2014 Accommodations Guidelines* 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 AND LARGE PRINT

Students with visual impairments were able to respond to test materials that were available in either Braille or large print. For each operational exam, one form was selected for the creation of a Braille and a large print edition. School district personnel ordered Braille or large print assessment materials directly from DRC. They could also contact PaTTAN for technical assistance regarding students with visual impairments.

School personnel were directed to transcribe all student answers (MC 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.

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 translated Spanish text on facing pages. The translated Spanish 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 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.

Chapter Six: Operational Forms Construction for 2015 Administrations

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) and constructed-response (CR) items and to collect demographic information. The MC items and all stimulus text are placed within the test book. Table 7–1 identifies the exam material format for each 2014 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/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/pencil administrations in the 2014–2015 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 2014/2015 Operational Keystone Exam Testing Windows

Exam	Wave 1 Dates	Wave 2 Dates
Algebra I	December 3–17, 2014	January 7–21, 2015
Biology		
Literature		

Table 7–3. Spring 2015 Operational Keystone Exam Testing Windows

Exam	Dates
Algebra I	May 13–27, 2015
Biology	
Literature	

Table 7–4. Summer 2014 Operational Keystone Exam Testing Windows

Exam	Dates
Algebra I	July 27–July 31, 2015
Biology	
Literature	

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. Stimulus passages on the Literature exam are 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. 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. 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	20–30	–	1.5 [0.670]	10 [0.400]	0.400
Biology	20–30	–	1.25 [0.800]	8 [0.375]	0.458
Literature	20–30	60	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 that will be 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 by April 15, 2015.
- 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 by April 28, 2015.

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.

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 prior to the assessment being 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. The Certifications attested that all security measures were followed concerning the handling of secure materials.

SAMPLE MANUALS

Copies of the *Keystone Exams Test Coordinator Handbook, Keystone Exams User Guide*, and the *Directions for Administration Manual*s are available on the PA eDIRECT online system at https://pa.drcedirect.com. (Click on Documents under the General Information tab.)

TESTING WINDOW ASSESSMENT ACCOMMODATIONS

Three accommodations manuals, *Accommodations Guidelines for Students with IEPs and Students with 504 Plans, Accommodations for English Language Learners*, and *Accommodations Guidelines for All Students*, were developed for use with the Spring 2015 Keystone Exams. 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 on May 18, 2015, and concluded on June 10, 2015. 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.

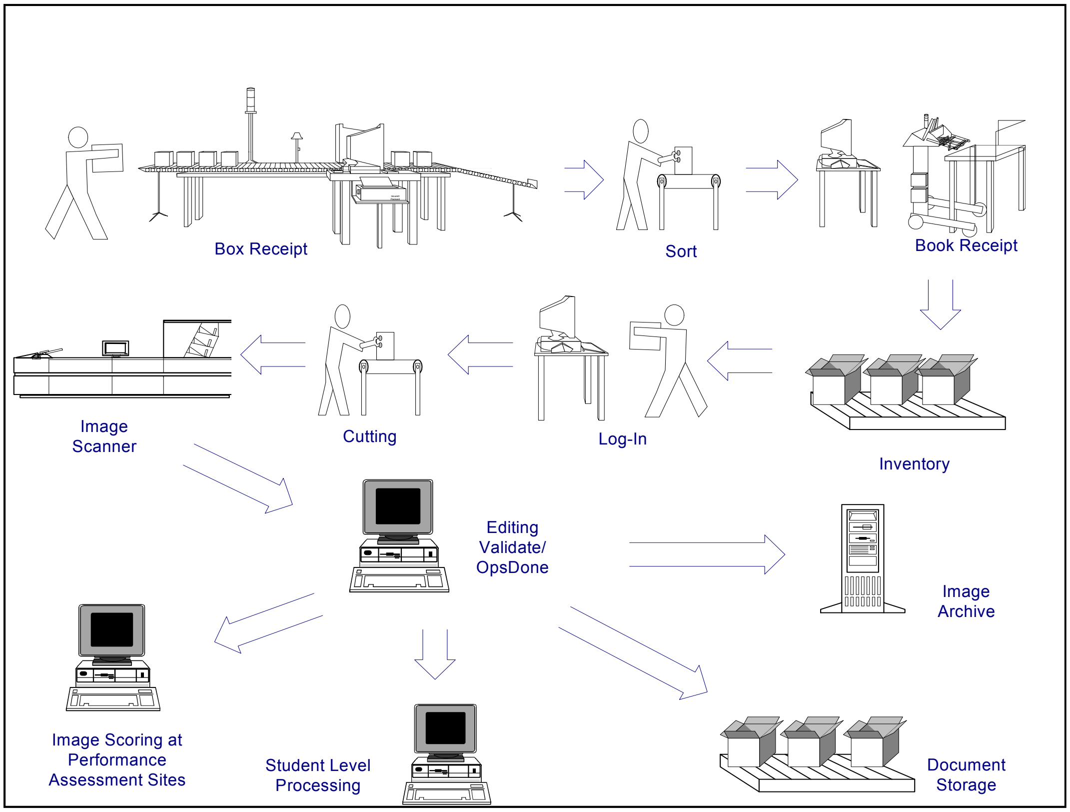
Table 8–1. Counts of 2015 Keystone Exams Materials Received: Algebra I, Biology, and Literature

Exam	Answer Books Received	Used Answer Books Received	Test Books Received	Total Books Received	Total Books Shipped
Algebra I	239,173	172,293	239,167	478,340	478,406
Biology	193,172	140,341	193,160	386,332	386,388
Literature	181,380	130,468	181,290	362,670	362,774

Chapter Eight: Processing and Scoring

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.

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 rangefinding meetings were held from July 13–16 at the Sheraton Harrisburg-Hershey, Harrisburg. The Literature and Biology rangefinding meetings were held from July 13–17, also at the Sheraton Harrisburg-Hershey.

Each rangefinding meeting began in a joint session with a review of the history of the assessment and then broke into groups by subject. 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 then 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.

SCORER RECRUITMENT AND QUALIFICATIONS

DRC retains a number of raters from year to year; the overall return rate in 2015 was 68%. This pool of experienced raters was drawn from to staff the scoring of the 2015 Keystones. To complete the rater staffing for this project, DRC placed advertisements in local newspapers and utilized a variety of web sites. Open houses 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. Thus, 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 AND QUALIFICATIONS

Scoring directors and team leaders were selected by content specialists 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, science, or writing and demonstrated organizational, leadership, and management skills. A majority of scoring directors and team leaders had at least five years of leadership experience working on large-scale assessments, including previous assessments for Pennsylvania. 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. This individual led all handscoring activities 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. All scoring directors were monitored by the project director, the project manager and the content specialists.

Team leaders assisted the scoring director with rater training by leading their teams in small group discussions and answering individual questions that raters may not have felt comfortable asking in a large group. Once raters were qualified, team leaders were responsible for 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. 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 2015 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 qualifying 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 each 8–10 raters ensured sufficient monitoring rates for team members.

The 2015 assessment included the opportunity for students to respond in Spanish to Algebra 1 and Biology items. The scoring director responsible for overseeing this is a Spanish language speaker who has a strong content background and has worked closely with the Keystone Exams in this capacity for five years. All Spanish raters were bilingual and hired specifically to score the Spanish portion of the assessment. They were required to meet the same training and scoring standards set for the raters of the English version of the assessment.

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 or team leaders led a thorough discussion of the responses, either in a large-group or small-group setting.

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 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.

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 DRC Scoring Centers in Sharonville and Columbus, Ohio; Plymouth and Woodbury, Minnesota; Pittsburgh, Pennsylvania; and Austin, Texas. Raters were seated at tables with two imaging stations at each table. 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 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. However, PDE does not receive students' responses or any other identifying information about the students. At no time in the alerts process do raters, or other DRC 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 2015 Keystone Exams, this scoring consistency was maintained, in part, through the validity process.

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

The validity responses were then implemented to test rater accuracy. The responses were scanned into the imaging system and dispersed intermittently to the raters. By the end of the project, raters had scored all 40 validity responses for any items they were qualified to score. Raters were unaware that 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. It is important to note that all raters who received validity responses had already successfully completed the training/qualifying process.

Next, the scores that the raters assigned to the validity responses 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 responses, 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 responses 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 papers 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, chosen by the content specialist, 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 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, content specialists, and project directors. The following reports and documents were used during the scoring of the open-ended items:

The Scoring Summary Report (includes two related reports)

1. 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.
2. The Score Point Distribution Report monitored the percentage of responses given each of the score points. For example, the Algebra 1 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). 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 8–2. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Winter 2015

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement %		% Validity Agreement	Percentage Awarded for Each Score Point					
					Exact	Adjacent		0	1	2	3	4	B/NS
Algebra 1	1	640123	A	0-1	99	1	97	34	60				6
			B	0-1	100	0	100	79	15				6
			C	0-1	100	0	98	74	19				6
			D	0-1	100	0	99	90	3				6
		640892		0-4	94	6	93	45	34	8	2	1	9
		612749	A	0-1	100	0	99	85	2				12
			B	0-1	100	0	99	52	36				12
			C	0-2	100	0	100	65	1	21			12
	2	640886	A	0-1	99	1	98	76	14				9
			B	0-1	99	1	98	84	6				9
			C	0-1	100	0	100	39	51				9
			D	0-1	99	1	98	80	10				9
		640898	A	0-1	100	0	99	68	21				10
			B	0-1	100	0	100	81	9				10
			C	0-1	100	0	99	84	5				10
			D	0-1	100	0	100	86	4				10
		622485	A	0-1	99	1	100	48	41				11
			B	0-1	100	0	100	84	5				11
			C	0-1	100	0	99	49	39				11
			D	0-1	100	0	100	54	35				11
Biology	1	629469		0-3	88	11	94	23	38	26	1		11
		644134		0-3	92	8	90	62	20	5	1		12
		641221		0-3	93	7	96	40	33	9	2		16
	2	641229		0-3	90	9	92	59	20	7	2		11
		641303		0-3	98	2	97	57	22	5	2		14
		611059		0-3	88	12	93	25	40	15	4		15
Literature	1	643730		0-3	82	18	92	18	42	24	5		11
		614675		0-3	84	16	96	16	30	39	4		11
		614674		0-3	86	14	82	13	36	25	7		19
	2	614016		0-3	80	20	90	16	37	27	6		13
		644768		0-3	85	15	92	6	36	40	6		12
		644767		0-3	86	14	90	8	34	35	6		16

Notes: B = blank; NS = non-scorable. Mathematics responses received a total of 0–4 points. For some mathematics items, readers simply applied a single score of 0, 1, 2, 3, or 4. Many of the mathematics 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, a mathematics 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 mathematics 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 8–3. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Summer 2015

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement %		% Validity Agreement	Percentage Awarded for Each Score Point					
					Exact	Adjacent		0	1	2	3	4	B/NS
Algebra 1	1	640376	A	0-2	99	1	98	64	14	19			3
			B	0-1	99	1	99	58	39				3
			C	0-1	100	0	100	42	54				3
		681303	A	0-1	99	1	99	27	69				3
			B	0-1	100	0	100	79	17				3
			C	0-1	100	0	99	96	1				3
			D	0-1	100	0	100	96	0				3
		674404		0-4	88	12	90	25	40	15	9	5	4
	2	633313	A	0-1	100	0	100	77	18				4
			B	0-1	93	7	97	57	38				4
			C	0-1	99	1	100	87	8				4
			D	0-1	99	1	100	84	11				4
			MU	0-1	100	0	100	95	1				4
		666557		0-4	95	5	91	51	36	5	2	1	5
		672275	A	0-1	100	0	99	75	19				6
			B	0-1	100	0	100	37	57				6
			C	0-1	100	0	100	79	16				6
			D	0-1	100	0	100	82	12				6
Biology	1	677887		0-3	93	7	96	29	35	21	7		7
		678932		0-3	94	6	94	18	23	37	15		7
		682984		0-3	88	12	88	28	26	36	1		7
	2	682669		0-3	87	13	88	9	43	30	7		7
		677890		0-3	92	8	82	25	54	9	9		7
		678996		0-3	94	6	95	41	20	26	9		12
Literature	1	643960		0-3	80	20	93	11	33	38	8		10
		616673		0-3	89	11	96	19	44	22	3		12
		616672		0-3	88	11	83	16	44	25	2		13
	2	613250		0-3	81	19	95	15	40	28	5		12
		683411		0-3	87	13	97	19	46	21	1		14
		683414		0-3	82	17	83	17	47	19	1		15

Notes: B = blank; NS = non-scorable. Mathematics responses received a total of 0–4 points. For some mathematics items, readers simply applied a single score of 0, 1, 2, 3, or 4. Many of the mathematics 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, a mathematics 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 mathematics 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 8–4. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Spring 2015

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement %		% Validity Agreement	Percentage Awarded for Each Score Point					
					Exact	Adjacent		0	1	2	3	4	B/NS
Algebra 1	1	672731	A	0-2	98	2	95	77	6	10			7
			B	0-1	99	1	98	59	34				7
			C	0-1	99	1	98	75	18				7
		641527	A	0-1	99	1	98	68	24				8
			B	0-1	98	2	100	66	26				8
			C	0-1	99	1	99	69	23				8
			D	0-1	100	0	100	83	9				8
		628222	A	0-2	98	2	98	57	13	21			9
			B	0-1	99	1	99	62	29				9
			C	0-1	97	3	100	35	56				9
			MU	0-1	100	0	100	91	0				9
	2	666526	A	0-1	99	1	100	37	57				6
			B	0-1	99	1	99	55	39				6
			C	0-1	100	0	100	88	6				6
			D	0-1	100	0	100	91	3				6
		628223	A	0-1	99	1	99	64	29				7
			B	0-1	98	2	99	51	42				7
			C	0-1	98	2	95	51	42				7
			D	0-1	98	3	98	61	32				7
			MU	0-1	100	0	100	93	0				7
		678819		0-4	93	7	94	23	30	17	12	8	9
Biology	1	678999		0-3	85	15	95	17	25	30	20		7
		679989		0-3	84	16	89	32	26	20	12		9
		678998		0-3	84	15	93	23	32	26	8		11
	2	684325		0-3	83	17	89	20	38	23	10		9
		682671		0-3	83	16	83	45	26	16	3		9
		644202		0-3	91	9	95	15	31	26	18		10
Literature	1	613661		0-3	77	23	90	17	30	37	8		8
		643181		0-3	77	23	81	23	27	30	11		9
		643182		0-3	78	22	85	10	25	39	15		10
	2	683334		0-3	76	24	83	7	28	45	11		9
		614551		0-3	77	23	83	9	21	40	21		9
		614552		0-3	80	20	81	12	29	34	15		10

Notes: B = blank; NS = non-scorable. Mathematics responses received a total of 0–4 points. For some mathematics items, readers simply applied a single score of 0, 1, 2, 3, or 4. Many of the mathematics 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, a mathematics 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 mathematics 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 2014/2015, Spring 2015, and Summer 2015 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

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 Complete Status = 02
- Module 2 Complete Status = 02

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 02/18/2015, 06/26/2015, and 08/25/2015 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	9,224	75,949	75,936
	Biology	5,543	54,247	54,236
	Literature	6,320	47,157	47,149
Spring	Algebra I	46,205	184,157	184,172
	Biology	33,428	154,596	154,585
	Literature	33,356	140,256	140,252
Summer	Algebra I	1,806	1,848	1,846
	Biology	1,080	1,116	1,116
	Literature	564	589	588

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

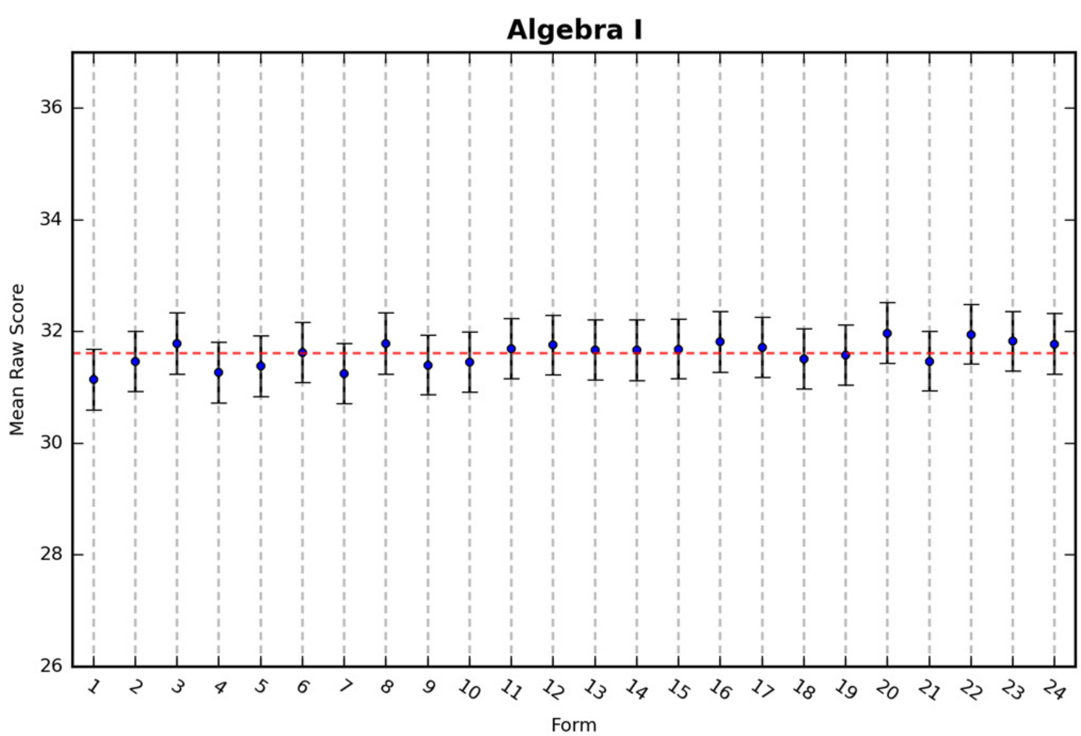
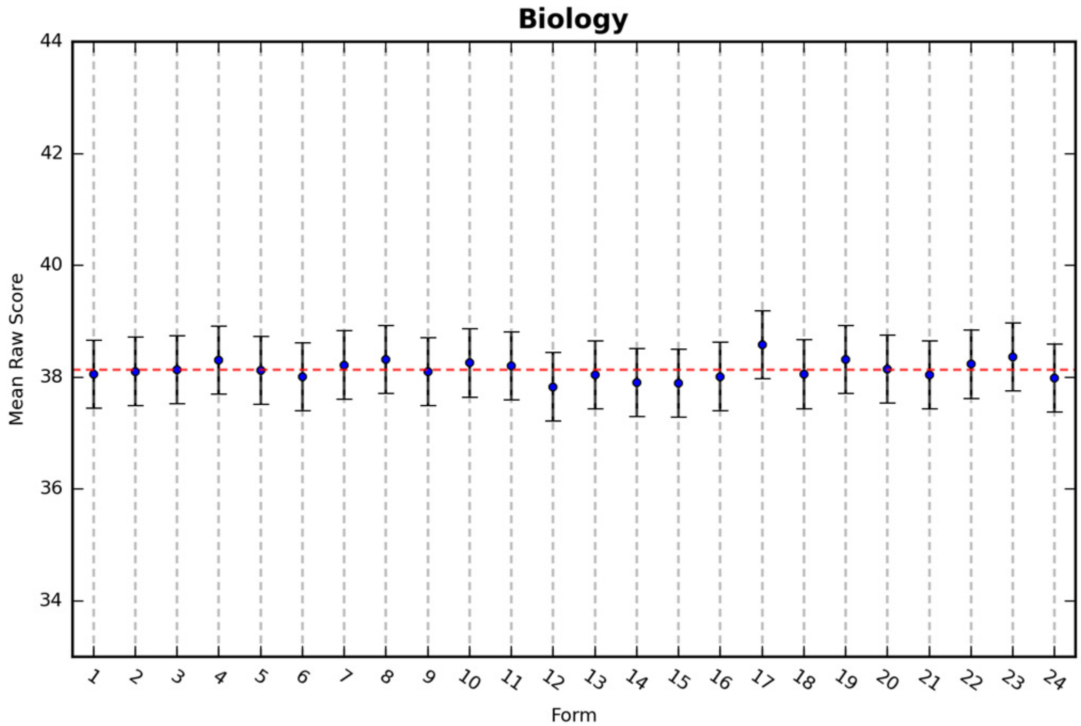
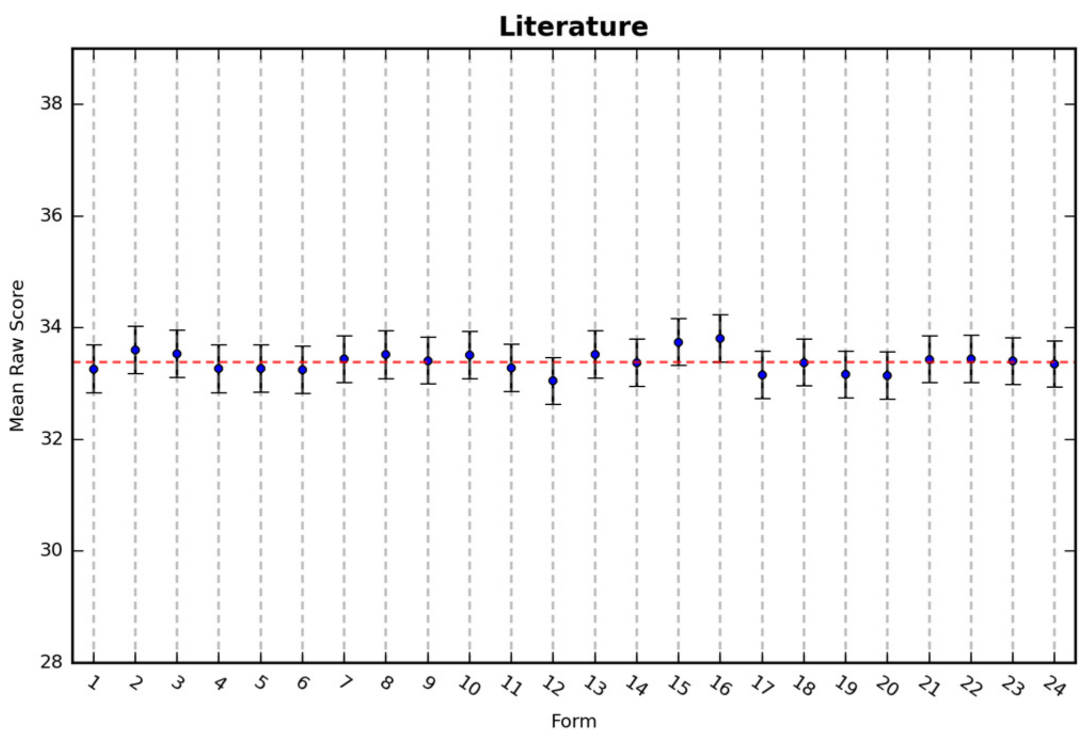


Figure 9–1 (continued). Form Mean Scores with +/- Three Standard Error (SE) Bands





CHAPTER TEN: SUMMARY DEMOGRAPHIC AND ACCOMMODATION DATA FOR SPRING 2015 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 2017, 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 2014–2015 school year. Students graduating before 2017 also participated in the assessment. 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 2015 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 2014/2015 and the other in the Summer 2015. 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 2015 (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 2015 Keystone Exam: Algebra I

	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	60	281	7,277	34,434	61,202	38,846	24,000	250	166,350
Total number of CBT processed	2	43	1,105	4,431	7,062	4,626	2,807	62	20,138
Total number of tests processed	62	324	8,382	38,865	68,264	43,472	26,807	312	186,488
Total number of tests processed with a score	57 91.9	323 99.7	8,350 99.6	38,657 99.5	66,294 97.1	41,405 95.2	24,634 91.9	290 92.9	180,010 96.5
Total number of tests processed without a score	5 8.1	1 0.3	32 0.4	208 0.5	1,970 2.9	2,067 4.8	2,173 8.1	22 7.1	6,478 3.5

*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 2015 Keystone Exam: Biology

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	29	209	44,602	64,349	25,700	301	135,190
Total number of CBT processed	0	5	6,896	11,667	3,658	69	22,295
Total number of tests processed	29	214	51,498	76,016	29,358	370	157,485
Total number of tests processed with a score	26 89.7	210 98.1	50,347 97.8	73,684 96.9	26,729 91.0	324 87.6	151,320 96.1
Total number of tests processed without a score	3 10.3	4 1.9	1,151 2.2	2,332 3.1	2,629 9.0	46 12.4	6,165 3.9

*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 2015 Keystone Exam: Literature

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	30	108	10,327	91,111	22,973	228	124,777
Total number of CBT processed	0	1	932	13,463	3,158	52	17,606
Total number of tests processed	30	109	11,259	104,574	26,131	280	142,383
Total number of tests processed with a score	26 86.7	109 100.0	10,618 94.3	102,120 97.7	23,889 91.4	245 87.5	137,007 96.2
Total number of tests processed without a score	4 13.3	0 0.0	641 5.7	2,454 2.3	2,242 8.6	35 12.5	5,376 3.8

*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 of Students without Scores on the Spring 2015 Keystone Exam: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	1 20.0	0 0.0	6 18.8	35 16.8	610 31.0	564 27.3	579 26.6	7 31.8	1,802 27.8
Absent without make-up	0 0.0	0 0.0	1 3.1	14 6.7	308 15.6	289 14.0	301 13.9	5 22.7	918 14.2
Non-attempt	3 60.0	0 0.0	5 15.6	40 19.2	698 35.4	804 38.9	645 29.7	5 22.7	2,200 34.0
Medical emergency	0 0.0	0 0.0	10 31.3	47 22.6	116 5.9	97 4.7	77 3.5	0 0.0	347 5.4
Parental request	0 0.0	1 100.0	6 18.8	19 9.1	56 2.8	50 2.4	122 5.6	0 0.0	254 3.9
Other reasons	1 20.0	0 0.0	4 12.5	53 25.5	182 9.2	263 12.7	449 20.7	5 22.7	957 14.8
Total not assessed	5	1	32	208	1,970	2,067	2,173	22	6,478

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

Table 10–2B. Counts of Students without Scores on the Spring 2015 Keystone Exam: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	1	0	379	667	802	18	1,867
	33.3	0.0	32.9	28.6	30.5	39.1	30.3
Absent without make-up	0	0	195	376	384	6	961
	0.0	0.0	16.9	16.1	14.6	13.0	15.6
Non-attempt	2	1	328	817	656	8	1,812
	66.7	25.0	28.5	35.0	25.0	17.4	29.4
Medical emergency	0	0	74	141	67	1	283
	0.0	0.0	6.4	6.0	2.5	2.2	4.6
Parental request	0	2	45	66	152	0	265
	0.0	50.0	3.9	2.8	5.8	0.0	4.3
Other reasons	0	1	130	265	568	13	977
	0.0	25.0	11.3	11.4	21.6	28.3	15.8
Total not assessed	3	4	1,151	2,332	2,629	46	6,165

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

Table 10–2L. Counts of Students without Scores on the Spring 2015 Keystone Exam: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	1	0	261	680	657	13	1,612
	25.0	0.0	40.7	27.7	29.3	37.1	30.0
Absent without make-up	0	0	98	356	354	5	813
	0.0	0.0	15.3	14.5	15.8	14.3	15.1
Non-attempt	3	0	182	893	636	7	1,721
	75.0	0.0	28.4	36.4	28.4	20.0	32.0
ELL in first year in U.S. schools	0	0	0	11	18	0	29
	0.0	0.0	0.0	0.4	0.8	0.0	0.5
Medical emergency	0	0	18	198	78	1	295
	0.0	0.0	2.8	8.1	3.5	2.9	5.5
Parental request	0	0	9	68	115	0	192
	0.0	0.0	1.4	2.8	5.1	0.0	3.6
Other reasons	0	0	73	248	384	9	714
	0.0	0.0	11.4	10.1	17.1	25.7	13.3
Total not assessed	4	0	641	2,454	2,242	35	5,376

*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 2015 administration, which was provided to the Pennsylvania Department of Education in July 2015. State summary statistics were based on students who received a total test score on the Spring 2015 administration with the exception of students who attended non-public schools or those who were home schooled. Also excluded were students who were not proficient.

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 2014/2015 and Summer 2015 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 2015 Keystone Exam: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender									
Female	13	112	3,787	20,175	32,734	19,977	11,839	113	88,750
	22.8	34.7	45.4	52.2	49.4	48.2	48.1	39.0	49.3
Male	23	211	4,563	18,466	33,547	21,412	12,781	177	91,180
	40.4	65.3	54.6	47.8	50.6	51.7	51.9	61.0	50.7
Race/Ethnicity									
American Indian/Alaskan Native (not Hispanic)	0	1	6	152	109	77	40	0	385
	0.0	0.3	0.1	0.4	0.2	0.2	0.2	0.0	0.2
Asian (not Hispanic)	8	108	829	1,739	1,900	809	614	11	6,018
	14.0	33.4	9.9	4.5	2.9	2.0	2.5	3.8	3.3
Black or African American (not Hispanic)	4	1	296	2,750	13,032	9,372	5,765	63	31,283
	7.0	0.3	3.5	7.1	19.7	22.6	23.4	21.7	17.4
Hispanic (any race)	1	2	239	1,788	7,685	5,744	3,126	74	18,659
	1.8	0.6	2.9	4.6	11.6	13.9	12.7	25.5	10.4
Multi-Racial (not Hispanic)	1	11	162	619	1,536	853	500	6	3,688
	1.8	3.4	1.9	1.6	2.3	2.1	2.0	2.1	2.0
White (not Hispanic)	22	200	6,812	31,555	41,956	24,498	14,556	136	119,735
	38.6	61.9	81.6	81.6	63.3	59.2	59.1	46.9	66.5
Native Hawaiian or Other Pacific Islander (not Hispanic)	0	0	6	38	61	35	17	0	157
	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.1
Educational Category and Other Demographic Groups									
IEP (not gifted)	3	10	191	1,352	9,672	10,092	7,083	91	28,494
	5.3	3.1	2.3	3.5	14.6	24.4	28.8	31.4	15.8
Student exited IEP in last 2 years	1	9	173	651	1,018	496	302	4	2,654
	1.8	2.8	2.1	1.7	1.5	1.2	1.2	1.4	1.5
Title I	5	20	760	5,889	18,700	12,890	7,699	145	46,108
	8.8	6.2	9.1	15.2	28.2	31.1	31.3	50.0	25.6
Title III served	0	0	5	129	2,082	2,008	1,311	22	5,557
	0.0	0.0	0.1	0.3	3.1	4.8	5.3	7.6	3.1
Title III not served	0	0	0	2	15	14	7	0	38
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Migrant student	0	0	0	2	34	35	29	0	100
	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.1
ELL (enrolled after 5/30/14)	0	0	2	25	491	337	277	8	1,140
	0.0	0.0	0.0	0.1	0.7	0.8	1.1	2.8	0.6
ELL (enrolled on or before 5/30/14)	0	0	6	130	1,775	1,833	1,148	29	4,921
	0.0	0.0	0.1	0.3	2.7	4.4	4.7	10.0	2.7

Table 10–3A (continued). Demographic Characteristics of Students taking the Spring 2015 Keystone Exam: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	6 0.1	88 0.2	238 0.4	182 0.4	121 0.5	1 0.3	636 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	1 0.3	8 0.1	83 0.2	206 0.3	124 0.3	110 0.4	0 0.0	532 0.3
Former ELL no longer monitored	0 0.0	6 1.9	122 1.5	829 2.1	1,681 2.5	789 1.9	335 1.4	2 0.7	3,764 2.1
Foreign exchange student	0 0.0	0 0.0	0 0.0	2 0.0	1 0.0	10 0.0	16 0.1	0 0.0	29 0.0
Economically disadvantaged	4 7.0	11 3.4	1,025 12.3	10,324 26.7	32,834 49.5	23,850 57.6	14,073 57.1	195 67.2	82,316 45.7
Historically Underperforming Subgroup	6 10.5	19 5.9	1,189 14.2	11,222 29.0	37,413 56.4	28,111 67.9	17,192 69.8	230 79.3	95,382 53.0
Enrollment in school of residence after 10/1/14	9 15.8	4 1.2	52 0.6	502 1.3	3,135 4.7	2,555 6.2	1,523 6.2	43 14.8	7,823 4.3
Enrollment in district of residence after 10/1/14	7 12.3	4 1.2	46 0.6	445 1.2	2,455 3.7	2,087 5.0	1,308 5.3	41 14.1	6,393 3.6
Enrollment as PA resident after 10/1/14	2 3.5	1 0.3	25 0.3	206 0.5	1,190 1.8	993 2.4	624 2.5	11 3.8	3,052 1.7
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	6 10.5	47 14.6	1,166 14.0	4,096 10.6	22,533 34.0	9,418 22.7	5,361 21.8	90 31.0	42,717 23.7
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	1 1.8	26 8.0	280 3.4	1,373 3.6	10,296 15.5	4,619 11.2	2,656 10.8	84 29.0	19,335 10.7
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	1 0.0	12 0.0	34 0.1	36 0.1	46 0.2	23 7.9	152 0.1
Number of assessed students	57	323	8,350	38,657	66,294	41,405	24,634	290	180,010

*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 2015 Keystone Exam: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	8 30.8	97 46.2	25,697 51.0	36,279 49.2	13,264 49.6	133 41.0	75,478 49.9
Male	4 15.4	113 53.8	24,639 48.9	37,391 50.7	13,463 50.4	190 58.6	75,800 50.1
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	63 0.1	111 0.2	36 0.1	0 0.0	210 0.1
Asian (not Hispanic)	1 3.8	7 3.3	2,437 4.8	1,988 2.7	741 2.8	10 3.1	5,184 3.4
Black or African American (not Hispanic)	4 15.4	53 25.2	6,147 12.2	11,723 15.9	5,998 22.4	78 24.1	24,003 15.9
Hispanic (any race)	0 0.0	2 1.0	3,202 6.4	7,506 10.2	3,157 11.8	83 25.6	13,950 9.2
Multi-Racial (not Hispanic)	0 0.0	5 2.4	1,034 2.1	1,376 1.9	542 2.0	8 2.5	2,965 2.0
White (not Hispanic)	7 26.9	143 68.1	37,406 74.3	50,892 69.1	16,232 60.7	142 43.8	104,822 69.3
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	45 0.1	71 0.1	20 0.1	1 0.3	137 0.1
Educational Category and Other Demographic Groups							
IEP (not gifted)	3 11.5	11 5.2	4,962 9.9	12,211 16.6	6,071 22.7	98 30.2	23,356 15.4
Student exited IEP in last 2 years	0 0.0	4 1.9	751 1.5	834 1.1	303 1.1	3 0.9	1,895 1.3
Title I	3 11.5	94 44.8	9,175 18.2	16,581 22.5	8,287 31.0	180 55.6	34,320 22.7
Title III served	0 0.0	0 0.0	686 1.4	2,129 2.9	1,200 4.5	9 2.8	4,024 2.7
Title III not served	0 0.0	0 0.0	6 0.0	18 0.0	11 0.0	0 0.0	35 0.0
Migrant student	0 0.0	0 0.0	6 0.0	45 0.1	30 0.1	0 0.0	81 0.1
ELL (enrolled after 5/30/14)	0 0.0	0 0.0	161 0.3	321 0.4	246 0.9	6 1.9	734 0.5
ELL (enrolled on or before 5/30/14)	0 0.0	0 0.0	635 1.3	2,005 2.7	1,076 4.0	17 5.2	3,733 2.5

Table 10–3B (continued). Demographic Characteristics of Students taking the Spring 2015 Keystone Exam: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	137 0.3	274 0.4	142 0.5	1 0.3	554 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	113 0.2	205 0.3	109 0.4	0 0.0	427 0.3
Former ELL no longer monitored	0 0.0	0 0.0	1,193 2.4	1,491 2.0	452 1.7	1 0.3	3,137 2.1
Foreign exchange student	0 0.0	0 0.0	5 0.0	13 0.0	18 0.1	1 0.3	37 0.0
Economically disadvantaged	3 11.5	70 33.3	17,480 34.7	33,452 45.4	14,810 55.4	218 67.3	66,033 43.6
Historically Underperforming Subgroup	5 19.2	73 34.8	19,967 39.7	39,016 53.0	17,453 65.3	247 76.2	76,761 50.7
Enrollment in school of residence after 10/1/14	8 30.8	2 1.0	1,440 2.9	2,820 3.8	1,553 5.8	46 14.2	5,869 3.9
Enrollment in district of residence after 10/1/14	7 26.9	2 1.0	1,221 2.4	2,325 3.2	1,319 4.9	41 12.7	4,915 3.2
Enrollment as PA resident after 10/1/14	2 7.7	1 0.5	599 1.2	1,040 1.4	601 2.2	9 2.8	2,252 1.5
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	1 3.8	15 7.1	13,066 26.0	12,381 16.8	5,915 22.1	89 27.5	31,467 20.8
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	0 0.0	15 7.1	5,439 10.8	5,782 7.8	2,705 10.1	89 27.5	14,030 9.3
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	13 0.0	35 0.0	44 0.2	27 8.3	119 0.1
Number of assessed students	26	210	50,347	73,684	26,729	324	151,320

*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 2015 Keystone Exam: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	5 19.2	52 47.7	5,300 49.9	50,362 49.3	10,125 42.4	76 31.0	65,920 48.1
Male	6 23.1	57 52.3	5,312 50.0	51,749 50.7	13,754 57.6	169 69.0	71,047 51.9
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	17 0.2	142 0.1	30 0.1	0 0.0	189 0.1
Asian (not Hispanic)	0 0.0	0 0.0	372 3.5	3,585 3.5	736 3.1	8 3.3	4,701 3.4
Black or African American (not Hispanic)	4 15.4	63 57.8	2,609 24.6	13,927 13.6	5,037 21.1	66 26.9	21,706 15.8
Hispanic (any race)	0 0.0	3 2.8	916 8.6	8,261 8.1	2,983 12.5	63 25.7	12,226 8.9
Multi-Racial (not Hispanic)	0 0.0	2 1.8	260 2.4	1,835 1.8	512 2.1	6 2.4	2,615 1.9
White (not Hispanic)	7 26.9	41 37.6	6,432 60.6	74,273 72.7	14,556 60.9	101 41.2	95,410 69.6
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	6 0.1	84 0.1	23 0.1	1 0.4	114 0.1
Educational Category and Other Demographic Groups							
IEP (not gifted)	4 15.4	11 10.1	1,479 13.9	14,127 13.8	7,034 29.4	95 38.8	22,750 16.6
Student exited IEP in last 2 years	0 0.0	0 0.0	119 1.1	1,203 1.2	283 1.2	3 1.2	1,608 1.2
Title I	4 15.4	104 95.4	3,541 33.3	19,026 18.6	7,255 30.4	150 61.2	30,080 22.0
Title III served	0 0.0	0 0.0	275 2.6	1,968 1.9	1,296 5.4	9 3.7	3,548 2.6
Title III not served	0 0.0	0 0.0	3 0.0	14 0.0	13 0.1	0 0.0	30 0.0
Migrant student	0 0.0	0 0.0	0 0.0	45 0.0	31 0.1	0 0.0	76 0.1
ELL (enrolled after 5/30/14)	0 0.0	0 0.0	23 0.2	166 0.2	168 0.7	2 0.8	359 0.3
ELL (enrolled on or before 5/30/14)	0 0.0	1 0.9	267 2.5	1,998 2.0	1,262 5.3	16 6.5	3,544 2.6

Table 10–3L (continued). Demographic Characteristics of Students taking the Spring 2015 Keystone Exam: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	18 0.2	332 0.3	134 0.6	1 0.4	485 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	31 0.3	250 0.2	140 0.6	0 0.0	421 0.3
Former ELL no longer monitored	0 0.0	0 0.0	191 1.8	2,076 2.0	351 1.5	1 0.4	2,619 1.9
Foreign exchange student	0 0.0	0 0.0	0 0.0	28 0.0	21 0.1	0 0.0	49 0.0
Economically disadvantaged	3 11.5	100 91.7	5,105 48.1	40,410 39.6	13,232 55.4	169 69.0	59,019 43.1
Historically Underperforming Subgroup	6 23.1	101 92.7	5,664 53.3	47,005 46.0	16,259 68.1	194 79.2	69,229 50.5
Enrollment in school of residence after 10/1/14	7 26.9	0 0.0	506 4.8	3,347 3.3	1,526 6.4	39 15.9	5,425 4.0
Enrollment in district of residence after 10/1/14	6 23.1	0 0.0	422 4.0	2,770 2.7	1,284 5.4	35 14.3	4,517 3.3
Enrollment as PA resident after 10/1/14	1 3.8	0 0.0	169 1.6	1,228 1.2	596 2.5	6 2.4	2,000 1.5
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	1 3.8	4 3.7	2,670 25.1	16,063 15.7	4,909 20.5	86 35.1	23,733 17.3
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	1 3.8	3 2.8	1,702 16.0	6,768 6.6	2,522 10.6	79 32.2	11,075 8.1
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	41 0.4	48 0.0	51 0.2	25 10.2	165 0.1
Number of assessed students	26	109	10,618	102,120	23,889	245	137,007

*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 (160,155), Biology (129,344), and Literature (119,681)
- CBT: Algebra I (19,855), Biology (21,976), and Literature (17,326)

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 12/17/2014), was updated for use with the 2015 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 2015 Keystone Exams varied by course exam and test administration mode as follows:

- PPT: Algebra I, 12; Biology, 12; and Literature, 7
- CBT: Algebra I, 10; Biology, 10; and Literature, 5

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 2015 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 2015 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 2015 Keystone Exam: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	9 0.0	N/A N/A	9 0.0
Large print format	92 0.1	N/A N/A	92 0.1
Computer Assistive Technology	4 0.0	N/A N/A	4 0.0
Some test items/questions read aloud	1,146 0.7	134 0.7	1,280 0.7
All test items/questions read aloud	688 0.4	209 1.1	897 0.5
Test items/questions signed	10 0.0	1 0.0	11 0.0
Test items/questions interpreted for ELL student	82 0.1	4 0.0	86 0.0
Amplification device	11 0.0	0 0.0	11 0.0
Magnification device	7 0.0	3 0.0	10 0.0
Color overlay	4 0.0	N/A N/A	4 0.0
Other (per Accommodations Guidelines)	37 0.0	35 0.2	72 0.0
Spanish version	1,065 0.7	N/A N/A	1,065 0.6
Online Accommodations Received			
Audio	N/A N/A	701 3.5	701 0.4
Color Chooser	N/A N/A	20 0.1	20 0.0
Contrasting Text Chooser	N/A N/A	5 0.0	5 0.0
Number of assessed students	160,155	19,855	180,010

Table 10–4B. Incidence of Presentation Accommodations Received on the Spring 2015 Keystone Exam: Biology

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	7	N/A	7
	0.0	N/A	0.0
Large print format	75	N/A	75
	0.1	N/A	0.0
Computer Assistive Technology	7	N/A	7
	0.0	N/A	0.0
Some test items/questions read aloud	615	118	733
	0.5	0.5	0.5
All test items/questions read aloud	784	266	1,050
	0.6	1.2	0.7
Test items/questions signed	17	3	20
	0.0	0.0	0.0
Test items/questions interpreted for ELL student	50	1	51
	0.0	0.0	0.0
Amplification device	7	3	10
	0.0	0.0	0.0
Magnification device	7	2	9
	0.0	0.0	0.0
Color overlay	2	N/A	2
	0.0	N/A	0.0
Other (per Accommodations Guidelines)	41	24	65
	0.0	0.1	0.0
Spanish version	643	N/A	643
	0.5	N/A	0.4
Online Accommodations Received			
Audio	N/A	957	957
	N/A	4.4	0.6
Color Chooser	N/A	21	21
	N/A	0.1	0.0
Contrasting Text Chooser	N/A	8	8
	N/A	0.0	0.0
Number of assessed students	129,344	21,976	151,320

Table 10–4L. Incidence of Presentation Accommodations Received on the Spring 2015 Keystone Exam: Literature

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	6 0.0	N/A N/A	6 0.0
Large print format	65 0.1	N/A N/A	65 0.0
Computer Assistive Technology	16 0.0	N/A N/A	16 0.0
Amplification device	4 0.0	2 0.0	6 0.0
Magnification device	6 0.0	1 0.0	7 0.0
Color overlay	3 0.0	N/A N/A	3 0.0
Other (per Accommodations Guidelines)	36 0.0	14 0.1	50 0.0
Online Accommodations Received			
Color Chooser	N/A N/A	10 0.1	10 0.0
Contrasting Text Chooser	N/A N/A	7 0.0	7 0.0
Number of assessed students	119,681	17,326	137,007

Table 10–5A. Incidence of Response Accommodations Received on the Spring 2015 Keystone Exam: Algebra I

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	87 0.1	2 0.0	89 0.0
Test administrator scribed open-ended responses at student's direction	98 0.1	2 0.0	100 0.1
Test administrator transcribed student responses	135 0.1	2 0.0	137 0.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	4 0.0	0 0.0	4 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	24 0.0	0 0.0	24 0.0
Keyboard, word processor, or computer	22 0.0	N/A N/A	22 0.0
Brailler/Notetaker	7 0.0	N/A N/A	7 0.0
Augmentative communication device	4 0.0	0 0.0	4 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	3 0.0	N/A N/A	3 0.0
Translation dictionary for ELL student	133 0.1	9 0.0	142 0.1
Other (per Accommodations Guidelines)	87 0.1	13 0.1	100 0.1
Number of assessed students	160,155	19,855	180,010

Table 10–5B. Incidence of Response Accommodations Received on the Spring 2015 Keystone Exam: Biology

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	53 0.0	0 0.0	53 0.0
Test administrator scribed open-ended responses at student's direction	85 0.1	1 0.0	86 0.1
Test administrator transcribed student responses	129 0.1	1 0.0	130 0.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	1 0.0	0 0.0	1 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	18 0.0	0 0.0	18 0.0
Keyboard, word processor, or computer	39 0.0	N/A N/A	39 0.0
Brailler/Notetaker	4 0.0	N/A N/A	4 0.0
Augmentative communication device	2 0.0	0 0.0	2 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	1 0.0	N/A N/A	1 0.0
Translation dictionary for ELL student	95 0.1	7 0.0	102 0.1
Other (per Accommodations Guidelines)	53 0.0	14 0.1	67 0.0
Number of assessed students	129,344	21,976	151,320

Table 10–5L. Incidence of Response Accommodations Received on the Spring 2015 Keystone Exam: Literature

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	74 0.1	0 0.0	74 0.1
Test administrator scribed open-ended responses at student's direction	103 0.1	0 0.0	103 0.1
Test administrator transcribed student responses	170 0.1	1 0.0	171 0.1
Keyboard, word processor, or computer	73 0.1	N/A N/A	73 0.1
Brailler/Notetaker	4 0.0	N/A N/A	4 0.0
Augmentative communication device	1 0.0	0 0.0	1 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	3 0.0	N/A N/A	3 0.0
Other (per Accommodations Guidelines)	60 0.1	3 0.0	63 0.0
Number of assessed students	119,681	17,326	137,007

Table 10–6A. Incidence of Setting Accommodations Received on the Spring 2015 Keystone Exam: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	34 0.0	0 0.0	34 0.0
One-on-one setting	365 0.2	7 0.0	372 0.2
Small group setting	12,413 7.8	1,293 6.5	13,706 7.6
Other (per Accommodations Guidelines)	153 0.1	8 0.0	161 0.1
Number of assessed students	160,155	19,855	180,010

Table 10–6B. Incidence of Setting Accommodations Received on the Spring 2015 Keystone Exam: Biology

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	29 0.0	0 0.0	29 0.0
One-on-one setting	290 0.2	6 0.0	296 0.2
Small group setting	10,090 7.8	1,302 5.9	11,392 7.5
Other (per Accommodations Guidelines)	107 0.1	9 0.0	116 0.1
Number of assessed students	129,344	21,976	151,320

Table 10–6L. Incidence of Setting Accommodations Received on the Spring 2015 Keystone Exam: Literature

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	26 0.0	0 0.0	26 0.0
One-on-one setting	223 0.2	5 0.0	228 0.2
Small group setting	9,994 8.4	1,036 6.0	11,030 8.1
Other (per Accommodations Guidelines)	98 0.1	8 0.0	106 0.1
Number of assessed students	119,681	17,326	137,007

Table 10–7A. Incidence of Timing Accommodations Received on the Spring 2015 Keystone Exam: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	16,707 10.4	1,120 5.6	17,827 9.9
Frequent breaks	771 0.5	318 1.6	1,089 0.6
Changed test schedule	204 0.1	4 0.0	208 0.1
Other (per Accommodations Guidelines)	23 0.0	1 0.0	24 0.0
Number of assessed students	160,155	19,855	180,010

Table 10–7B. Incidence of Timing Accommodations Received on the Spring 2015 Keystone Exam: Biology

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	5,503	1,098	6,601
	4.3	5.0	4.4
Frequent breaks	514	332	846
	0.4	1.5	0.6
Changed test schedule	123	9	132
	0.1	0.0	0.1
Other (per Accommodations Guidelines)	15	1	16
	0.0	0.0	0.0
Number of assessed students	129,344	21,976	151,320

Table 10–7L. Incidence of Timing Accommodations Received on the Spring 2015 Keystone Exam: Literature

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	13,119	1,016	14,135
	11.0	5.9	10.3
Frequent breaks	568	267	835
	0.5	1.5	0.6
Changed test schedule	153	10	163
	0.1	0.1	0.1
Other (per Accommodations Guidelines)	6	1	7
	0.0	0.0	0.0
Number of assessed students	119,681	17,326	137,007

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 or 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 10–8. Accommodations Rates for Non-IEP and IEP Students on the Spring 2015 Keystone Exams

Course Tested	Student Subgroup Tested	PPT	CBT	Total
		N/Pct	N/Pct	N/Pct
Algebra I	**Non-IEP Students**	134,958	16,558	151,516
	Non-Accommodated	120,071 89.0	15,996 96.6	136,067 89.8
	Accommodated	14,887 11.0	562 3.4	15,449 10.2
	IEP Students	25,197	3,297	28,494
	Non-Accommodated	12,868 51.1	1,720 52.2	14,588 51.2
	Accommodated	12,329 48.9	1,577 47.8	13,906 48.8
Biology	**Non-IEP Students**	109,473	18,491	127,964
	Non-Accommodated	104,659 95.6	17,980 97.2	122,639 95.8
	Accommodated	4,814 4.4	511 2.8	5,325 4.2
	IEP Students	19,871	3,485	23,356
	Non-Accommodated	10,034 50.5	1,744 50.0	11,778 50.4
	Accommodated	9,837 49.5	1,741 50.0	11,578 49.6
Literature	**Non-IEP Students**	99,819	14,438	114,257
	Non-Accommodated	88,759 88.9	13,888 96.2	102,647 89.8
	Accommodated	11,060 11.1	550 3.8	11,610 10.2
	IEP Students	19,862	2,888	22,750
	Non-Accommodated	9,710 48.9	1,775 61.5	11,485 50.5
	Accommodated	10,152 51.1	1,113 38.5	11,265 49.5

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 April 11, 2014. 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.

SUBGROUP COMPARISONS FOR PPT ADMINISTRATION MODE

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.

SUBGROUP COMPARISONS FOR CBT ADMINISTRATION MODE

For the CBT administration the ELL/Non-IEP and IEP/ELL subgroup *n*-counts for Algebra I were only 269 and 80, respectively, and even lower for Biology and Literature. 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.

COMPARISONS BETWEEN PPT AND CBT

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 difference between PPT and CBT in percent of students receiving an extended time exceeded five percentage in Algebra I and Literature.
- 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 2015 Keystone Exam: Algebra I

Accommodation Received by Administration Mode		Classification of Students Regarding IEP and ELL: General Education (non-IEP or ELL) N/Pct	IEP and non-ELL N/Pct	ELL and non-IEP N/Pct	Both IEP and ELL N/Pct
PPT	Some test items/questions read aloud	70 0.1	904 3.7	138 2.8	34 4.0
	All test items/questions read aloud	23 0.0	648 2.7	5 0.1	12 1.4
	Small group setting	995 0.8	10,701 43.9	452 9.3	265 31.4
	Extended time	12,831 9.9	3,153 12.9	637 13.1	86 10.2
	Frequent breaks	56 0.0	676 2.8	17 0.3	22 2.6
	Number assessed	130,089	24,354	4,869	843
CBT	Some test items/questions read aloud	10 0.1	102 3.2	20 7.4	2 2.5
	All test items/questions read aloud	21 0.1	186 5.8	1 0.4	1 1.3
	Small group setting	65 0.4	1,197 37.2	7 2.6	24 30.0
	Extended time	446 2.7	616 19.1	40 14.9	18 22.5
	Frequent breaks	13 0.1	302 9.4	0 0.0	3 3.8
	Number assessed	16,289	3,217	269	80
Total	Some test items/questions read aloud	80 0.1	1,006 3.6	158 3.1	36 3.9
	All test items/questions read aloud	44 0.0	834 3.0	6 0.1	13 1.4
	Small group setting	1,060 0.7	11,898 43.2	459 8.9	289 31.3
	Extended time	13,277 9.1	3,769 13.7	677 13.2	104 11.3
	Frequent breaks	69 0.0	978 3.5	17 0.3	25 2.7
	Number assessed	146,378	27,571	5,138	923

Table 10-9B. Incidence of IEP and ELL Students Receiving Selected Accommodations on the Spring 2015 Keystone Exam: Biology

Accommodation Received by Administration Mode		Classification of Students Regarding IEP and ELL: General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Some test items/questions read aloud	22 0.0	510 2.7	68 1.9	15 2.3
	All test items/questions read aloud	32 0.0	724 3.8	7 0.2	21 3.3
	Small group setting	703 0.7	8,805 45.8	341 9.7	241 37.5
	Extended time	3,460 3.3	1,756 9.1	241 6.9	46 7.2
	Frequent breaks	30 0.0	456 2.4	13 0.4	15 2.3
	Number assessed	105,963	19,229	3,510	642
CBT	Some test items/questions read aloud	8 0.0	90 2.6	17 6.8	3 4.6
	All test items/questions read aloud	42 0.2	221 6.5	1 0.4	2 3.1
	Small group setting	52 0.3	1,224 35.8	7 2.8	19 29.2
	Extended time	354 1.9	679 19.9	49 19.6	16 24.6
	Frequent breaks	17 0.1	310 9.1	2 0.8	3 4.6
	Number assessed	18,241	3,420	250	65
Total	Some test items/questions read aloud	30 0.0	600 2.6	85 2.3	18 2.5
	All test items/questions read aloud	74 0.1	945 4.2	8 0.2	23 3.3
	Small group setting	755 0.6	10,029 44.3	348 9.3	260 36.8
	Extended time	3,814 3.1	2,435 10.8	290 7.7	62 8.8
	Frequent breaks	47 0.0	766 3.4	15 0.4	18 2.5
	Number assessed	124,204	22,649	3,760	707

Table 10–9L. Incidence of IEP and ELL Students Receiving Selected Accommodations on the Spring 2015 Keystone Exam: Literature

Accommodation Received by Administration Mode		Classification of Students Regarding IEP and ELL			
		General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Small group setting	729 0.8	8,745 45.5	313 10.5	207 32.6
	Extended time	9,638 10.0	2,936 15.3	464 15.6	81 12.8
	Frequent breaks	30 0.0	510 2.7	8 0.3	20 3.2
	Number assessed	96,836	19,228	2,983	634
CBT	Small group setting	36 0.3	987 34.9	3 1.3	10 17.9
	Extended time	466 3.3	500 17.7	45 19.6	5 8.9
	Frequent breaks	13 0.1	252 8.9	1 0.4	1 1.8
	Number assessed	14,208	2,832	230	56
Total	Small group setting	765 0.7	9,732 44.1	316 9.8	217 31.4
	Extended time	10,104 9.1	3,436 15.6	509 15.8	86 12.5
	Frequent breaks	43 0.0	762 3.5	9 0.3	21 3.0
	Number assessed	111,044	22,060	3,213	690

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, *2015 Accommodations Guidelines* (PDE, revised 12/17/2014, pages 23–40). 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 2015 PSSA and 2014–2015 Keystone Exams

Type of Testing Accommodation	Explanation
Student used the following 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.

Table 10–10 (continued). Glossary of Accommodation Terms as Applied in the 2015 PSSA and 2014–2015 Keystone Exams

Type of Testing Accommodation	Explanation
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.
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.
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.

Table 10–10 (continued). Glossary of Accommodation Terms as Applied in the 2015 PSSA and 2014–2015 Keystone Exams

Type of Testing Accommodation	Explanation
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.
Keyboard, word processor, or computer (per *Accommodations Guidelines*)	This is an allowable accommodation as a typing function only for students with the identified need. 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.
Audio recording of student responses (per *Accommodations Guidelines*)	An electronic recording device may be used to record responses, which must be transcribed into the answer booklet by the test administrator. (Students who are unable to use a pencil or have illegible handwriting may answer MC questions orally. Answers must be recorded in the answer booklet without alteration during the testing period.)
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.

Table 10–10 (continued). Glossary of Accommodation Terms as Applied in the 2015 PSSA and 2014–2015 Keystone Exams

Type of Testing Accommodation	Explanation
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.
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[2] 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.

---

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

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 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[3] 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.

[3] 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.

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.[4] 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.

SCATTER PLOTS OF ITEM DISCRIMINATION AND DIFFICULTY

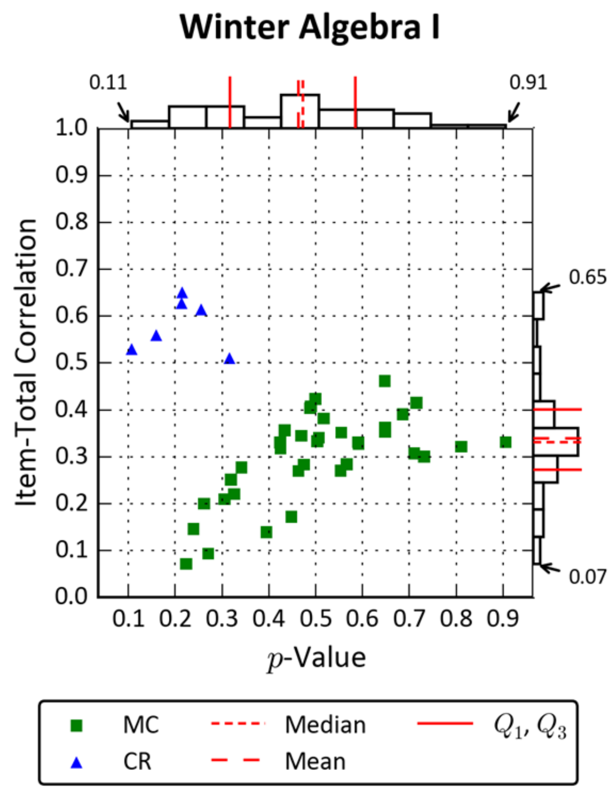
Figure 11–1 contains a series of scatter plots showing item discrimination (i.e., item-total correlation on *y*-axis) vs. item difficulty (i.e., *p*-value on *x*-axis) for the operational items in each content area exam by test administration. These plots provide information about item discrimination 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.

[4] 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.

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



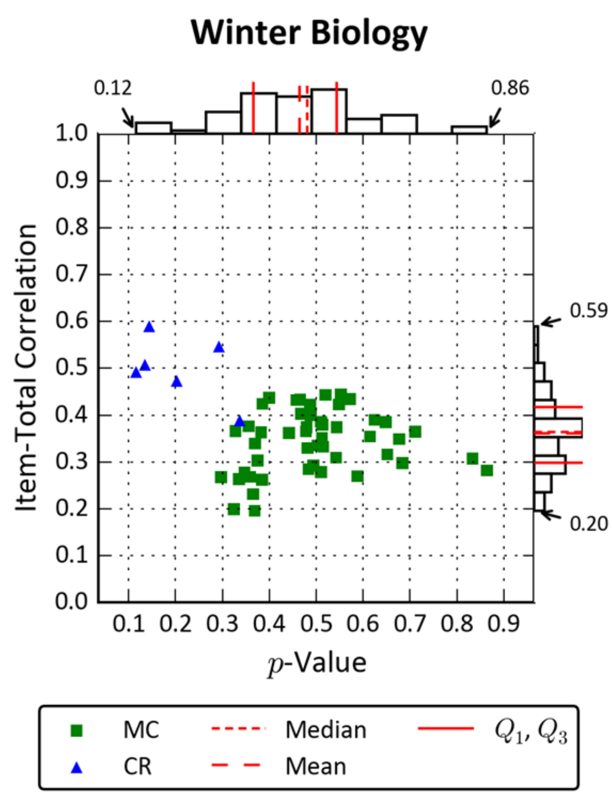
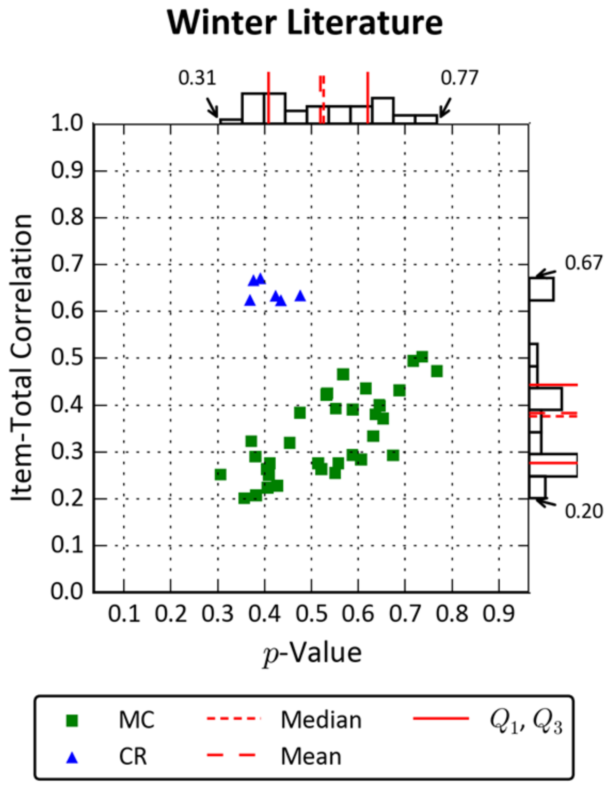


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



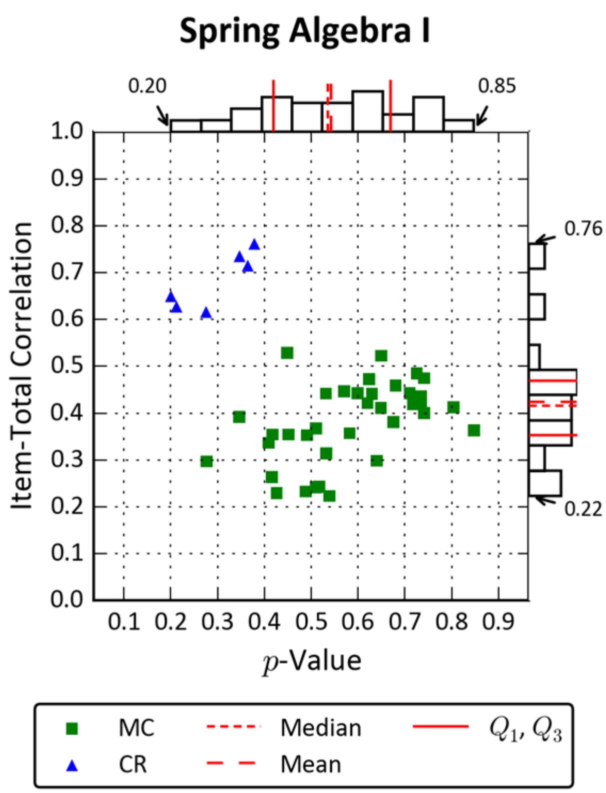
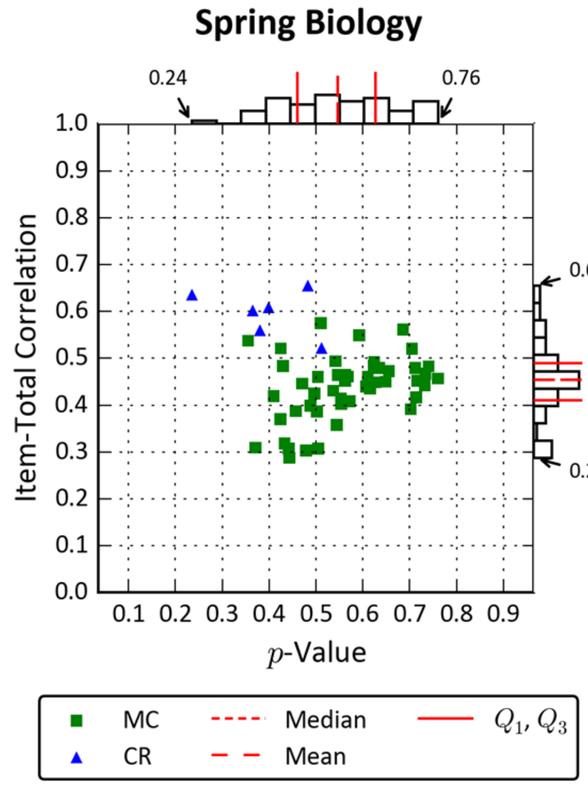


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



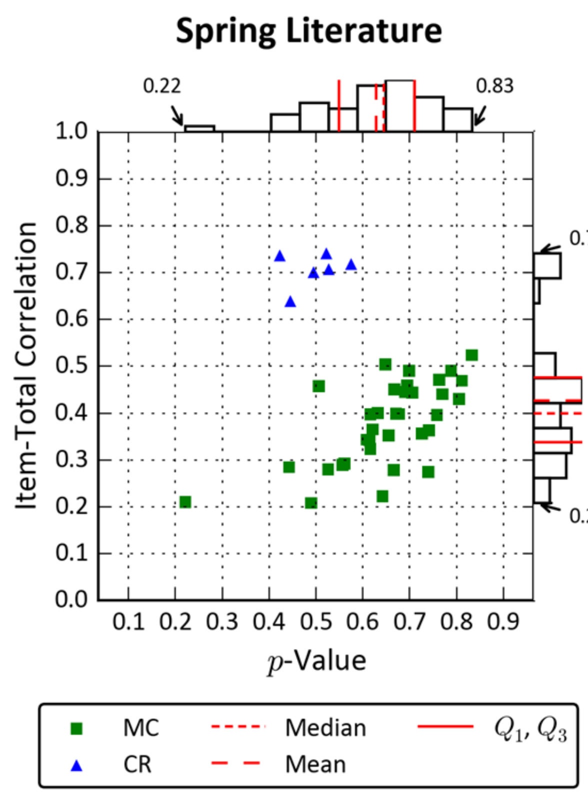
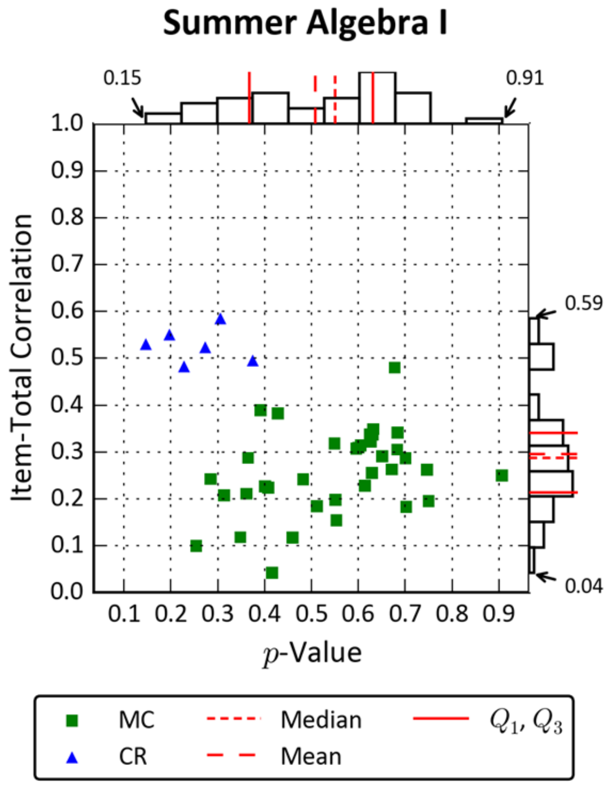


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



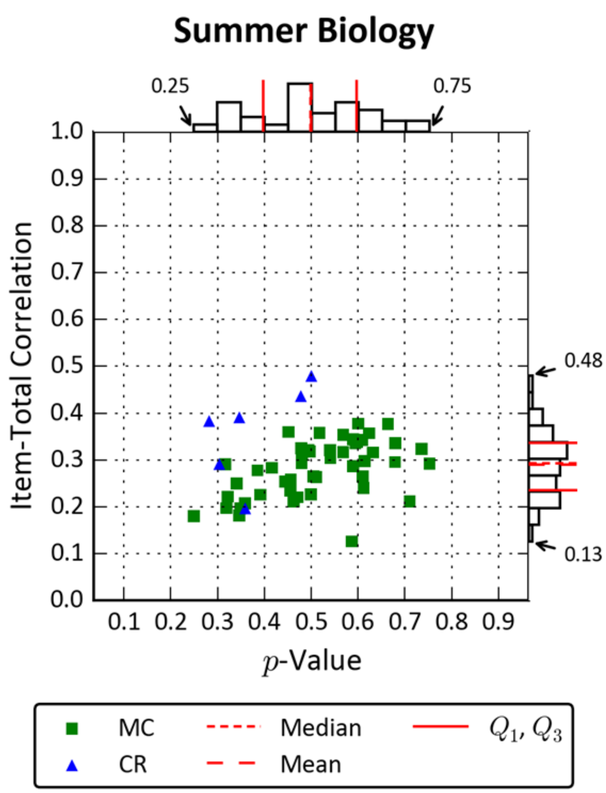
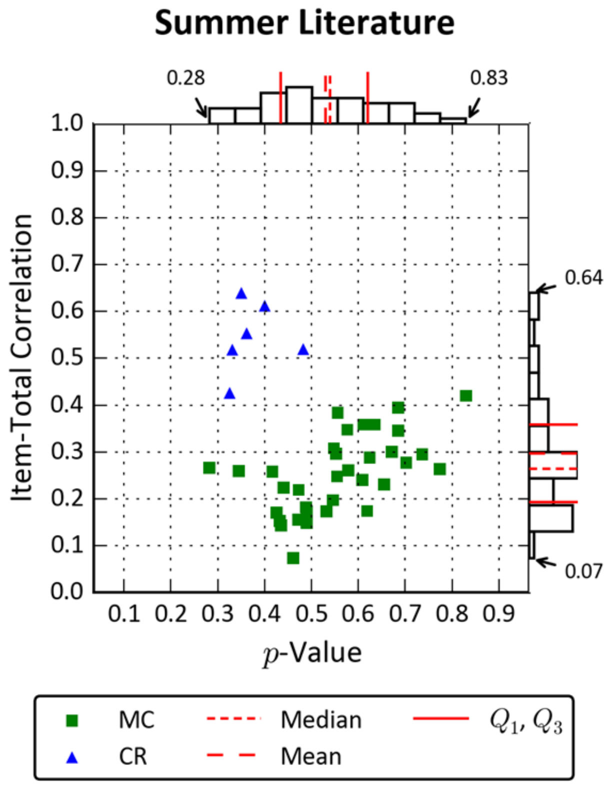


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty



OBSERVATIONS AND INTERPRETATIONS

Table 11–1 provides the mean and median *p*-values and median[5] 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 0.50 to 0.65 with the standard deviation (SD) ranging from 0.11 to 0.16, while the mean *p*-values for the CR items ranged from about 0.20 to 0.50 with the standard deviation ranging from 0.04 to 0.10. The median item-test correlations ranged from 0.26 to 0.45 and 0.39 to 0.71 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.

[5] 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.

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

Administration	Content Area	Multiple-Choice Items				Constructed-Response Items			
		Mean *p*-Value	SD *p*-Value	Median *p*-Value	Median I-T Corr.	Mean *p*-Value	SD *p*-Value	Median *p*-Value	Median I-T Corr.
Winter	Algebra I	0.51	0.16	0.49	0.32	0.21	0.07	0.21	0.59
	Biology	0.50	0.13	0.49	0.35	0.20	0.09	0.17	0.50
	Literature	0.54	0.12	0.55	0.32	0.41	0.04	0.41	0.63
Spring	Algebra I	0.58	0.14	0.59	0.40	0.30	0.08	0.31	0.68
	Biology	0.57	0.11	0.56	0.45	0.40	0.10	0.39	0.61
	Literature	0.65	0.12	0.67	0.40	0.50	0.06	0.51	0.71
Summer	Algebra I	0.55	0.15	0.60	0.26	0.25	0.08	0.25	0.53
	Biology	0.51	0.12	0.51	0.29	0.38	0.09	0.35	0.39
	Literature	0.56	0.12	0.55	0.26	0.38	0.06	0.36	0.54

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

Chapter Eleven: Classical Item Statistics

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,... 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 β_n represents a student's proficiency (ability) level, and δ_{ij} 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.)

Chapter Twelve: Rasch Item Calibration

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 19.7 to 50.9 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	33.5	44.4%
		2	1.8	2.4%
		3	1.4	1.8%
		4	1.3	1.7%
		5	1.2	1.6%
	Biology	1	29.3	35.2%
		2	1.6	1.9%
		3	1.4	1.7%
		4	1.3	1.6%
		5	1.2	1.4%
	Literature	1	16.9	29.7%
		2	2.3	4.0%
		3	1.6	2.7%
		4	1.3	2.3%
		5	1.3	2.2%
Spring	Algebra I	1	43.5	50.9%
		2	1.8	2.1%
		3	1.4	1.6%
		4	1.3	1.5%
		5	1.2	1.4%
	Biology	1	28.9	34.8%
		2	1.6	2.0%
		3	1.5	1.9%
		4	1.4	1.7%
		5	1.2	1.5%
	Literature	1	24	37.5%
		2	2.2	3.5%
		3	1.5	2.3%
		4	1.3	2.0%
		5	1.2	1.9%
Summer	Algebra I	1	30	41.6%
		2	1.7	2.4%
		3	1.6	2.2%
		4	1.4	2.0%
		5	1.4	1.9%
	Biology	1	13.3	19.7%
		2	1.7	2.5%
		3	1.6	2.3%
		4	13.3	19.7%
		5	1.7	2.5%
	Literature	1	13.7	25.5%
		2	2.6	4.8%
		3	1.7	3.1%
		4	1.6	3.0%
		5	1.4	2.7%

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 respect to the latent variable δ_n if, for all $x = (x_1, x_2, \ldots x_I)$ 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 covariances 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	Statistics										
		N	Mean	SD	Min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Max	>0.20
Winter	Algebra I	861	-0.02	0.03	-0.11	-0.06	-0.04	-0.02	-0.01	0.01	0.15	0
	Biology	1431	-0.02	0.02	-0.09	-0.04	-0.03	-0.02	-0.01	0.00	0.09	0
	Literature	780	-0.02	0.04	-0.13	-0.08	-0.05	-0.02	0.00	0.02	0.28	1
Spring	Algebra I	861	-0.02	0.03	-0.11	-0.05	-0.04	-0.02	0.00	0.01	0.08	0
	Biology	1431	-0.02	0.02	-0.11	-0.04	-0.03	-0.02	-0.01	0.01	0.21	1
	Literature	780	-0.02	0.04	-0.13	-0.07	-0.04	-0.02	0.00	0.01	0.19	0
Summer	Algebra I	861	-0.02	0.03	-0.15	-0.07	-0.05	-0.02	0.00	0.02	0.15	0
	Biology	1431	-0.02	0.04	-0.14	-0.06	-0.04	-0.02	0.00	0.03	0.17	0
	Literature	780	-0.02	0.06	-0.19	-0.10	-0.06	-0.03	0.01	0.05	0.38	2

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

Administration	Content Area	Infit Mean Square					Outfit Mean Square				
		Mean	SD	Min	Max	[0.7, 1.3]	Mean	SD	Min	Max	[0.7, 1.3]
Winter	Algebra I	1.01	0.15	0.68	1.39	39	1.04	0.25	0.49	1.83	35
	Biology	0.98	0.10	0.72	1.30	54	0.98	0.14	0.66	1.46	51
	Literature	1.05	0.13	0.74	1.33	38	1.08	0.20	0.73	1.60	35
Spring	Algebra I	1.01	0.13	0.78	1.24	42	1.02	0.19	0.71	1.33	37
	Biology	1.00	0.10	0.82	1.22	54	0.99	0.15	0.73	1.34	53
	Literature	1.01	0.16	0.66	1.47	38	1.04	0.26	0.62	1.80	32
Summer	Algebra I	1.00	0.12	0.69	1.28	41	1.01	0.16	0.60	1.40	38
	Biology	1.00	0.07	0.80	1.33	53	1.00	0.09	0.78	1.41	53
	Literature	1.03	0.14	0.64	1.44	37	1.04	0.16	0.64	1.51	37

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 difficulty ranged from -0.01 to 0.36. 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	0.28	0.01	0.99	-2.11	2.05
		MC	36	0.07	0.01	0.89	-2.11	1.85
		CR	6	1.57	0.00	0.38	0.97	2.05
	Biology	All	54	-0.01	0.01	0.74	-1.95	1.68
		MC	48	-0.17	0.01	0.61	-1.95	1.01
		CR	6	1.24	0.01	0.49	0.35	1.68
	Literature	All	40	0.36	0.01	0.75	-1.29	1.69
		MC	34	0.26	0.01	0.75	-1.29	1.54
		CR	6	0.95	0.01	0.48	0.26	1.69
Spring	Algebra I	All	42	0.17	0.01	0.90	-1.55	2.04
		MC	36	-0.04	0.01	0.77	-1.55	1.63
		CR	6	1.41	0.00	0.57	0.69	2.04
	Biology	All	54	0.11	0.01	0.60	-1.04	1.67
		MC	48	0.01	0.01	0.54	-1.04	1.33
		CR	6	0.90	0.00	0.49	0.38	1.67
	Literature	All	40	0.23	0.01	0.73	-1.07	2.14
		MC	34	0.11	0.01	0.72	-1.07	2.14
		CR	6	0.90	0.00	0.26	0.68	1.38
Summer	Algebra I	All	42	0.29	0.05	0.87	-1.74	2.43
		MC	36	0.09	0.05	0.76	-1.74	1.67
		CR	6	1.46	0.03	0.56	0.81	2.43
	Biology	All	54	0.17	0.06	0.57	-1.08	1.42
		MC	48	0.09	0.06	0.53	-1.08	1.03
		CR	6	0.81	0.04	0.44	0.14	1.42
	Literature	All	40	0.31	0.09	0.73	-1.52	1.66
		MC	34	0.15	0.09	0.67	-1.52	1.29
		CR	6	1.18	0.06	0.33	0.68	1.66

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.[6] 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).

[6] 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

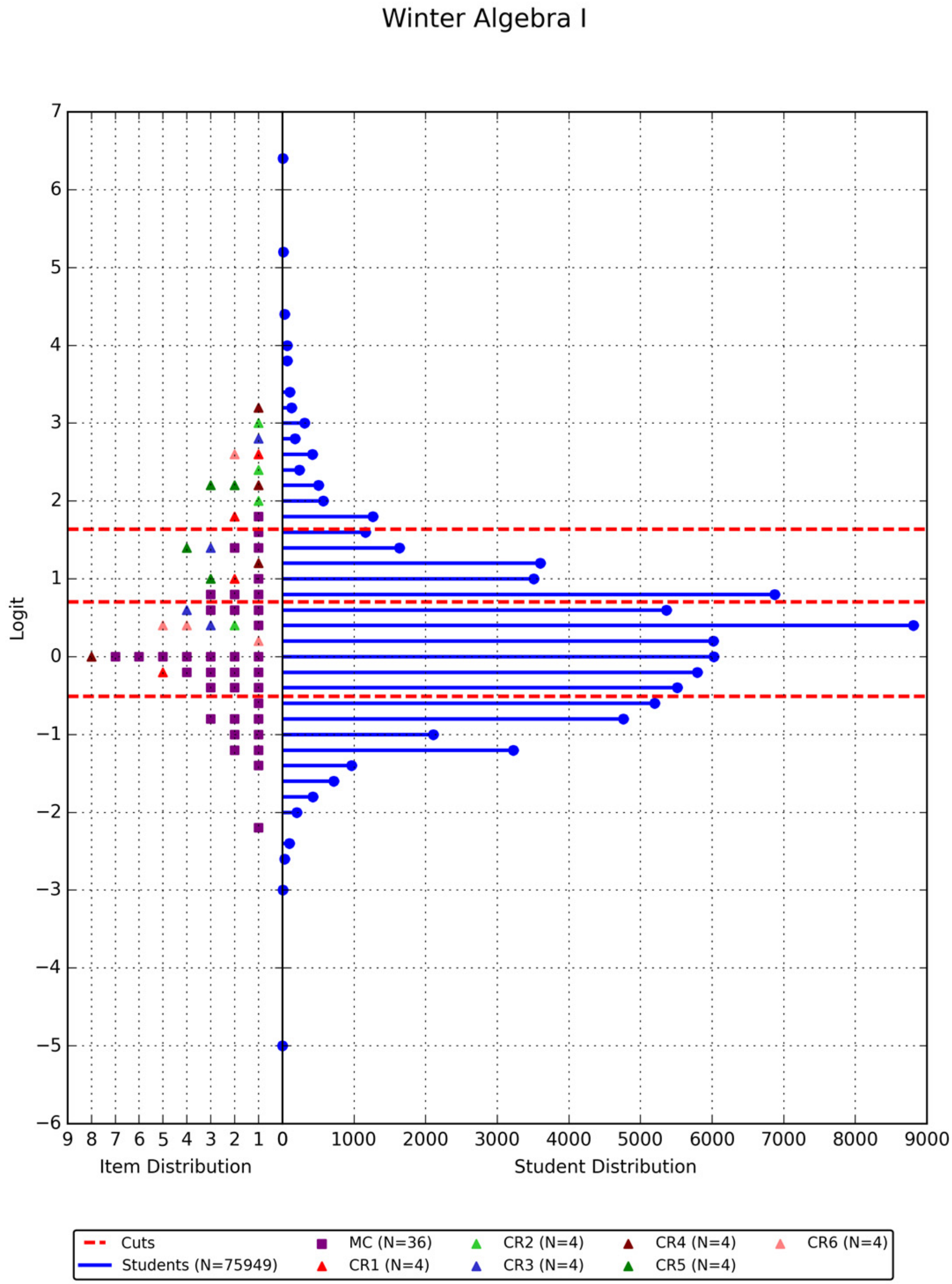


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

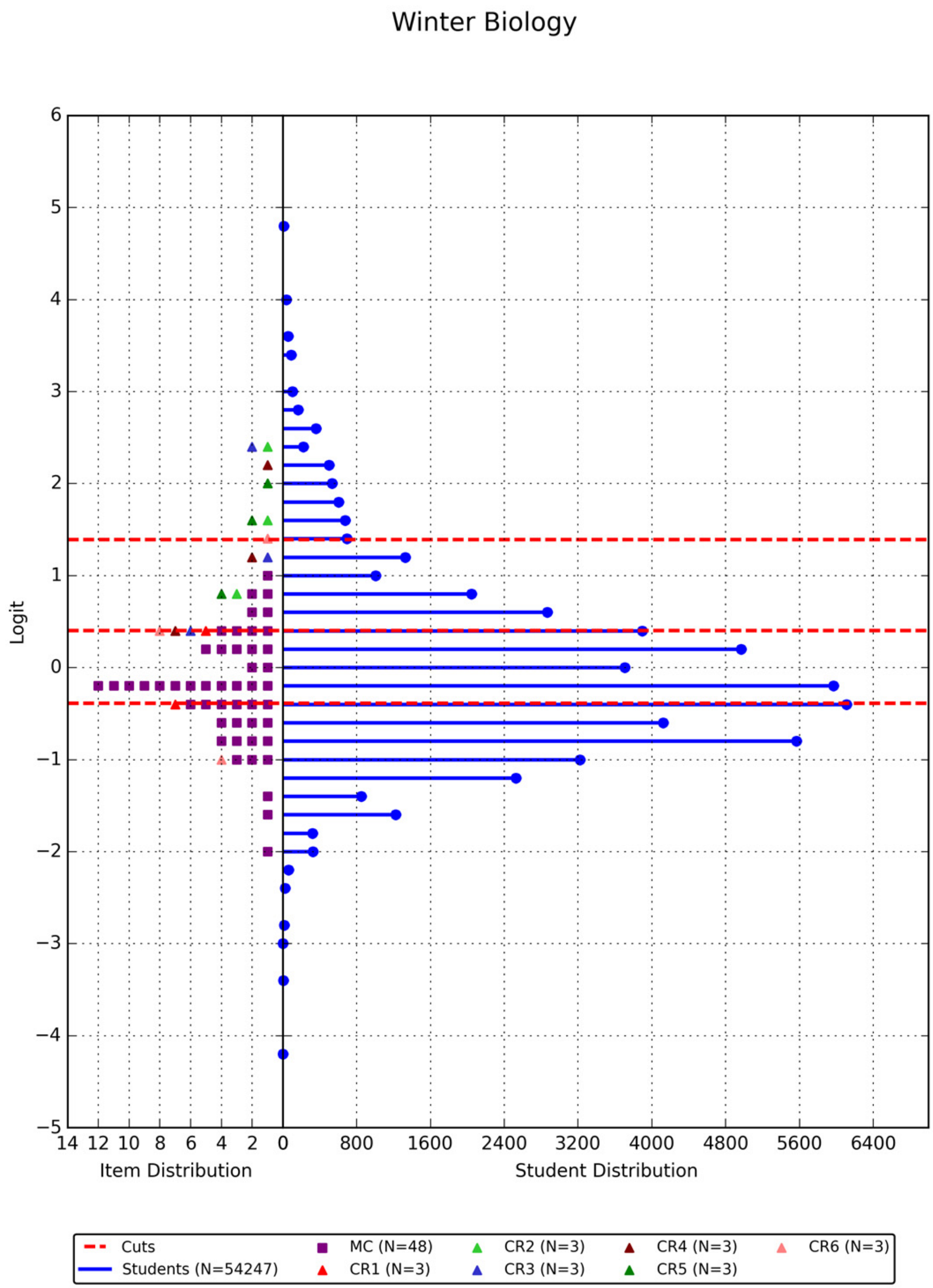


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

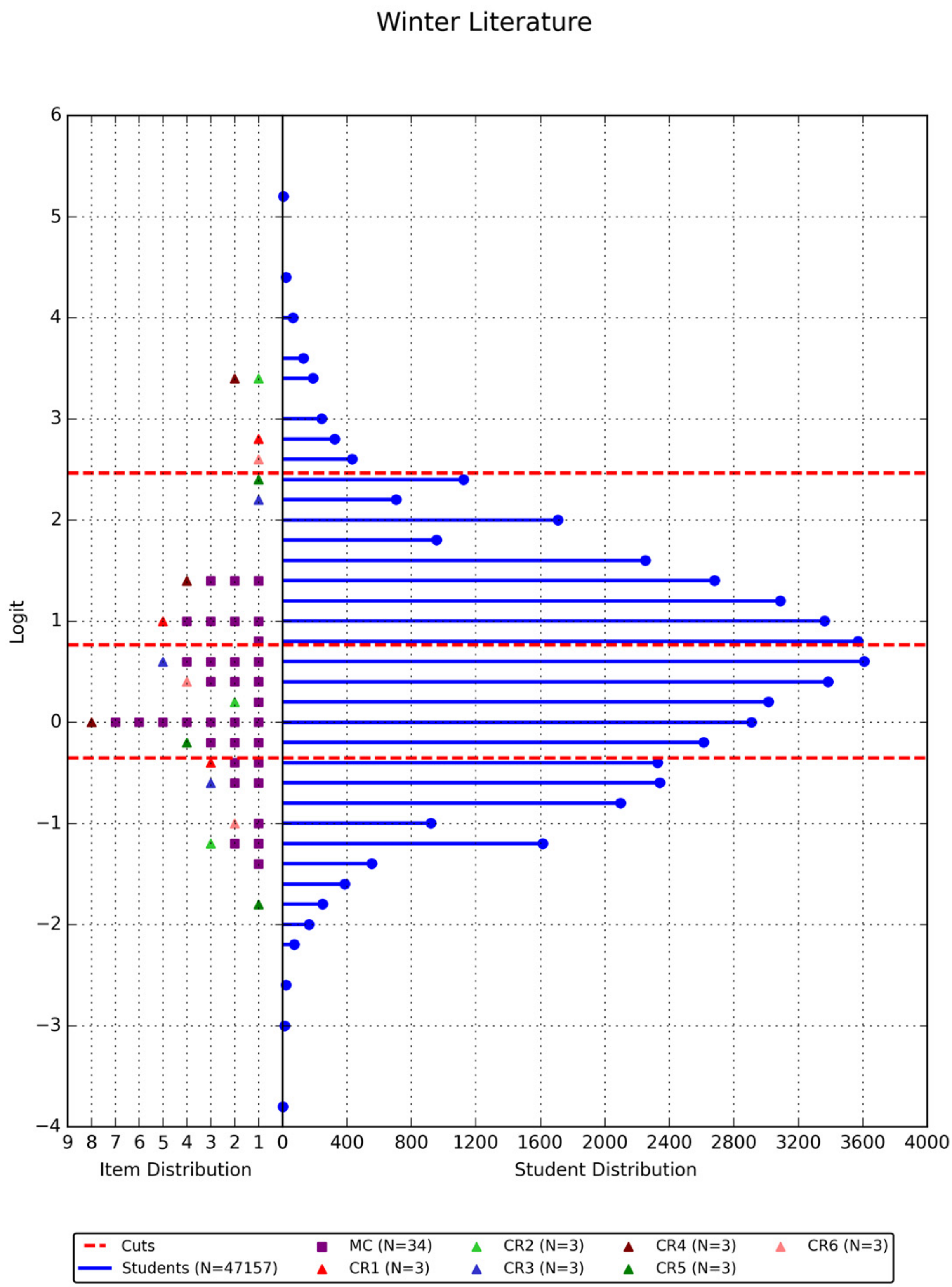


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

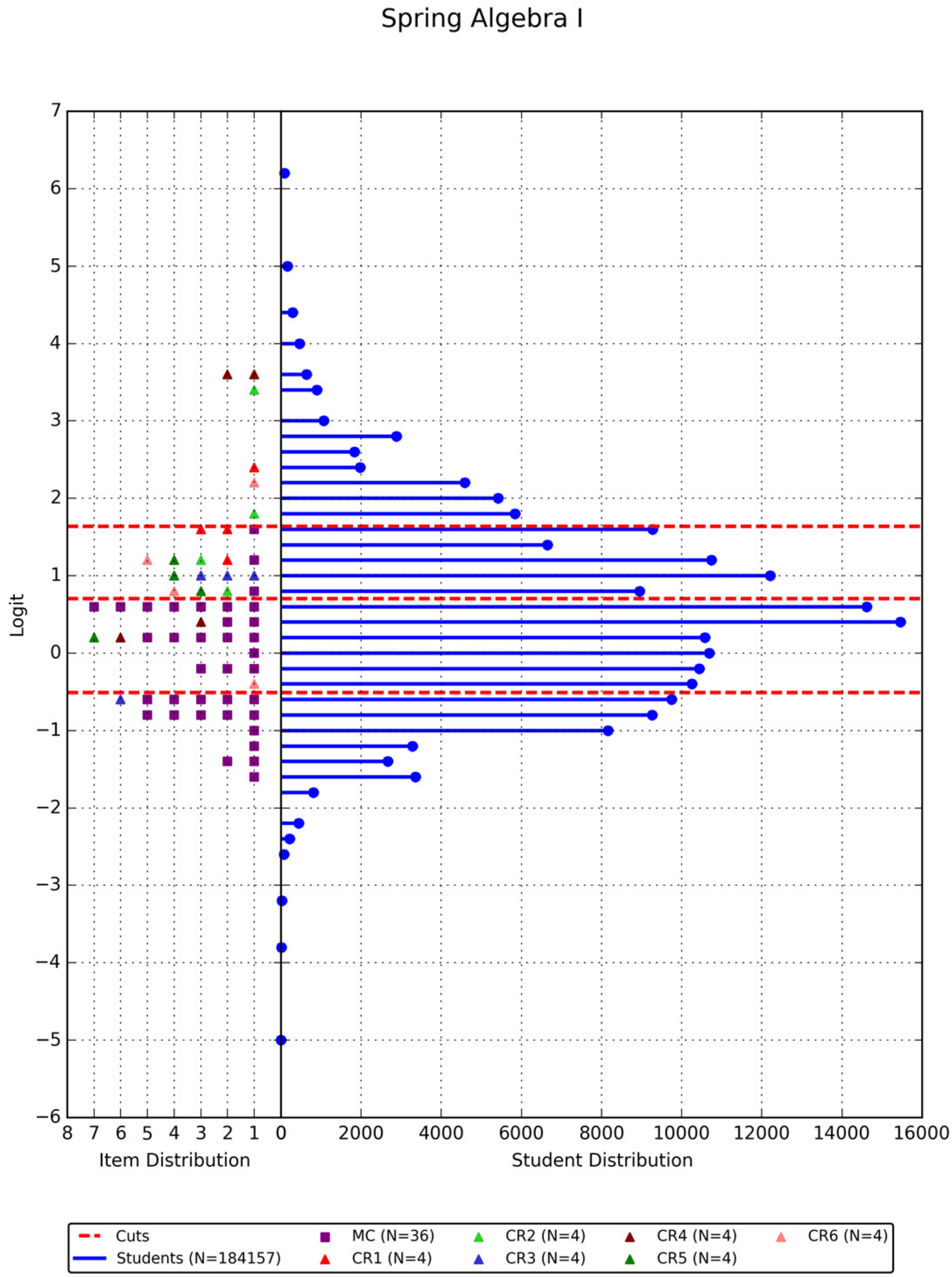


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

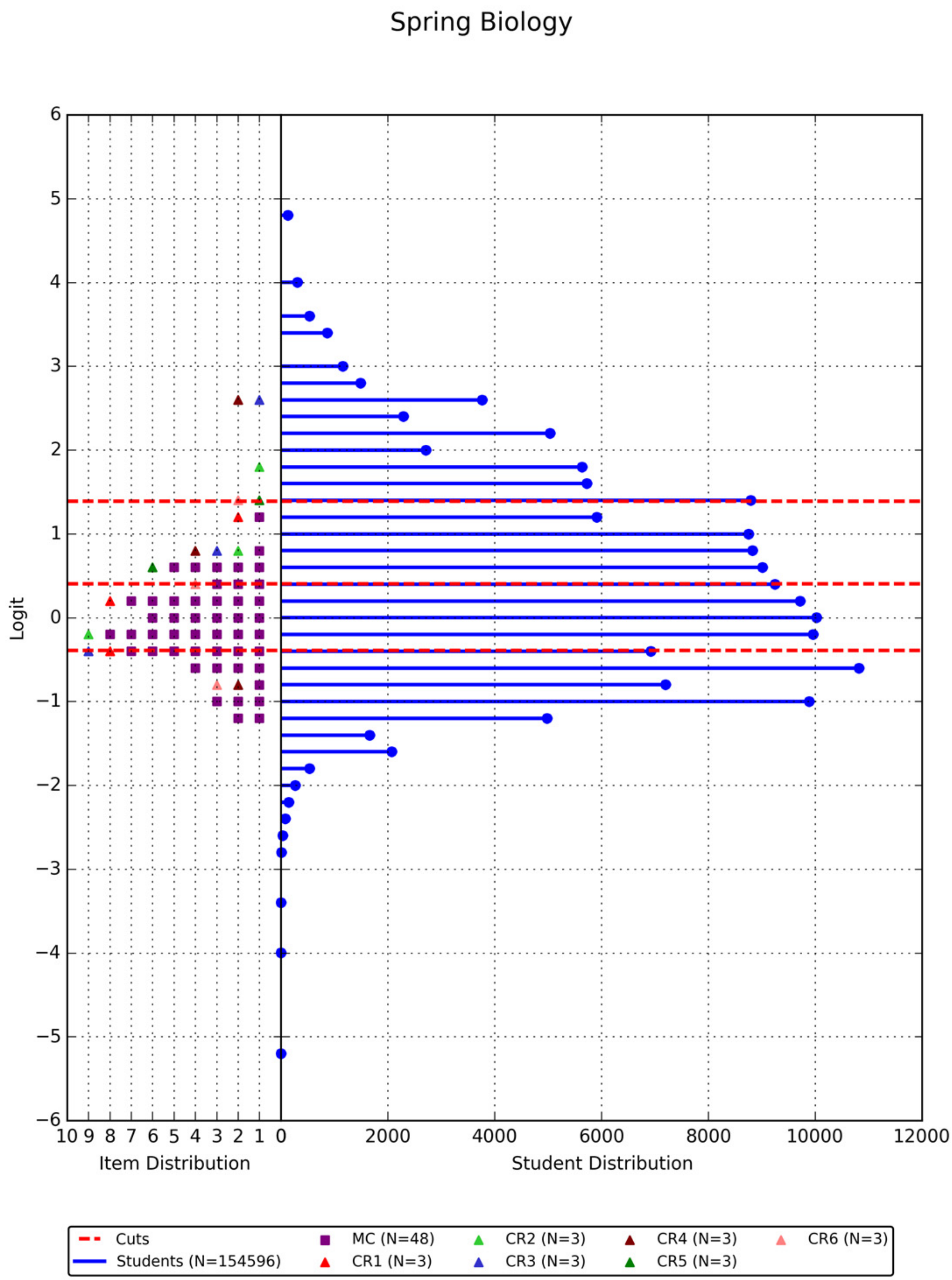


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

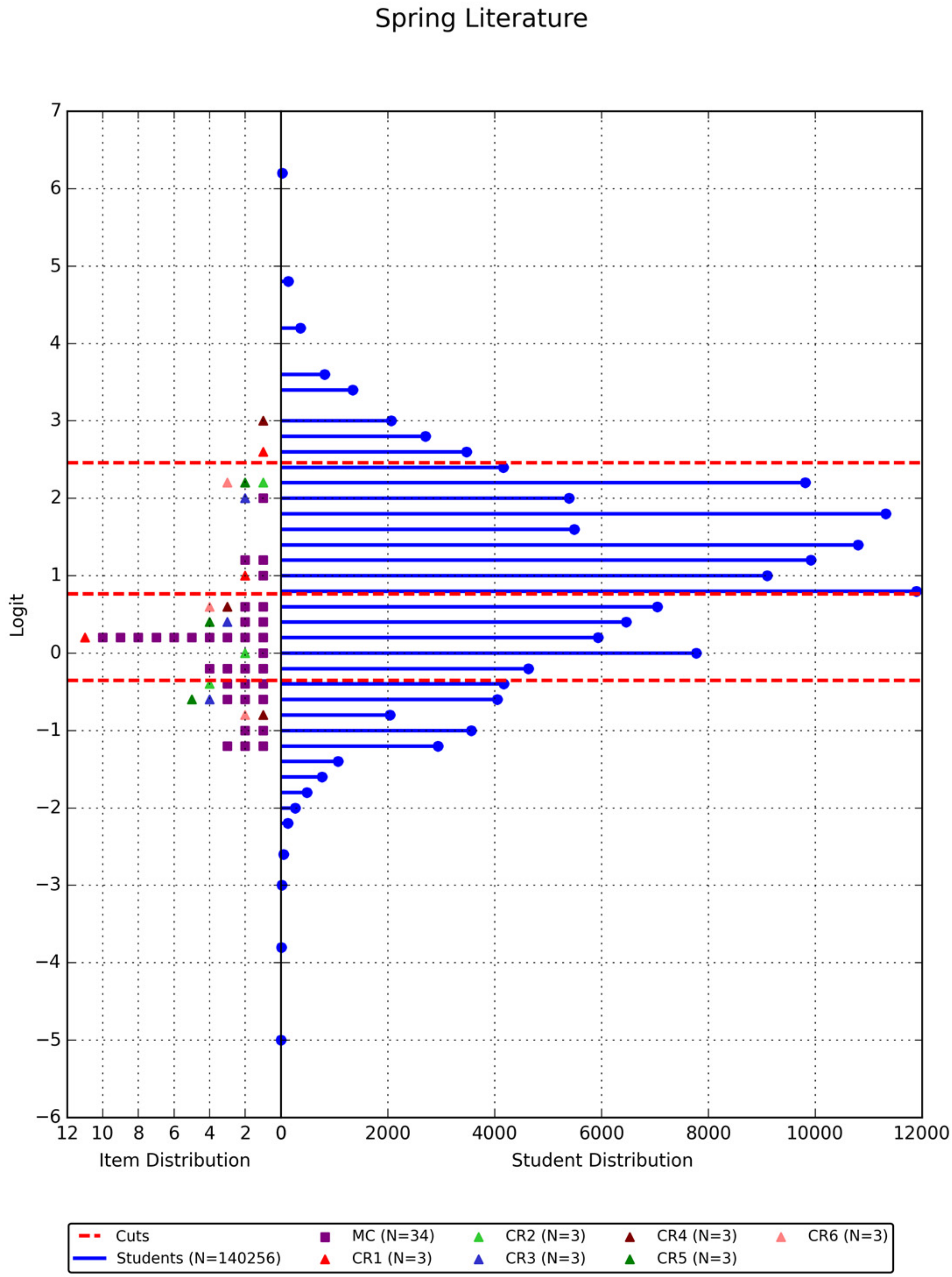


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

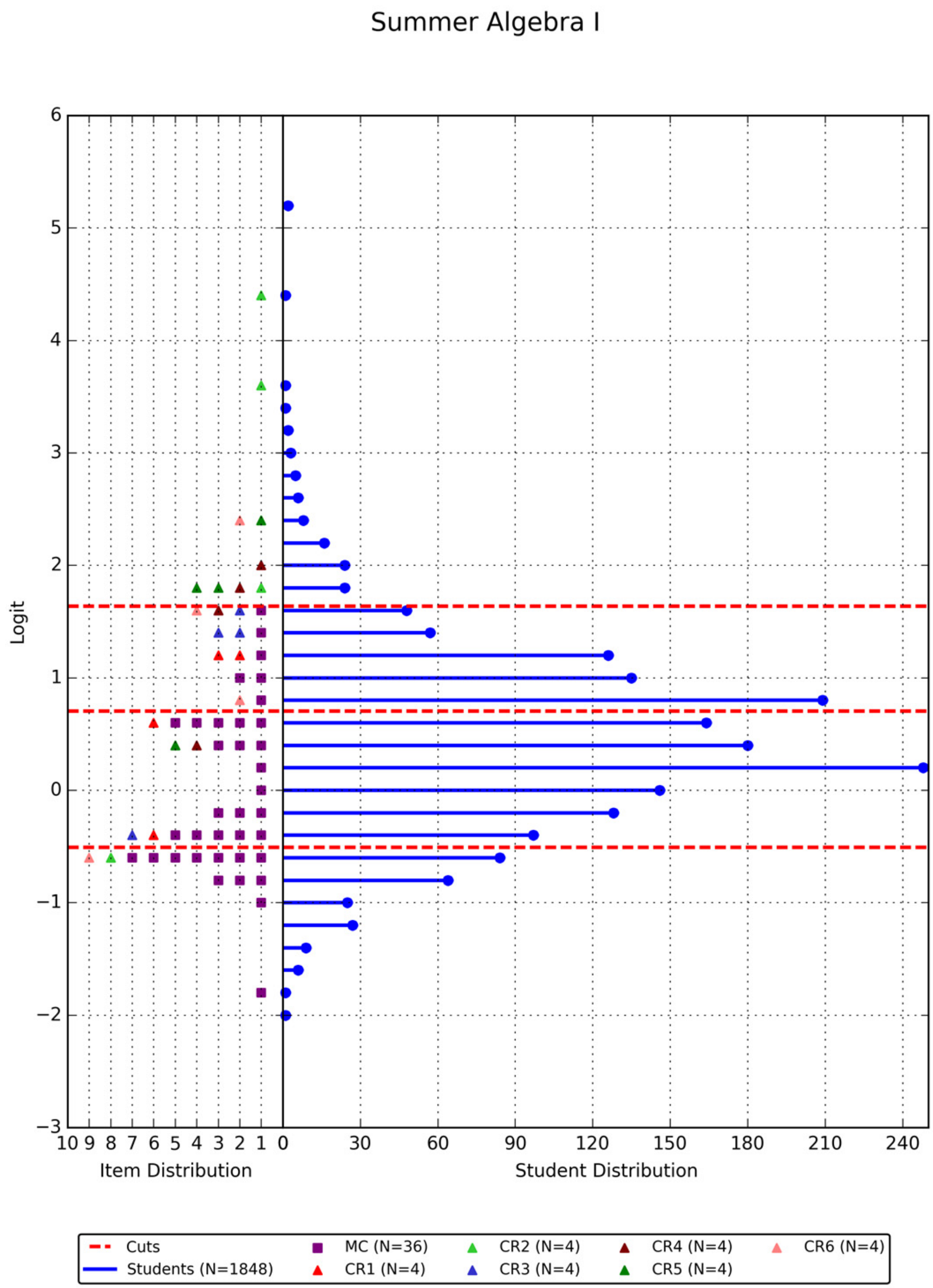


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

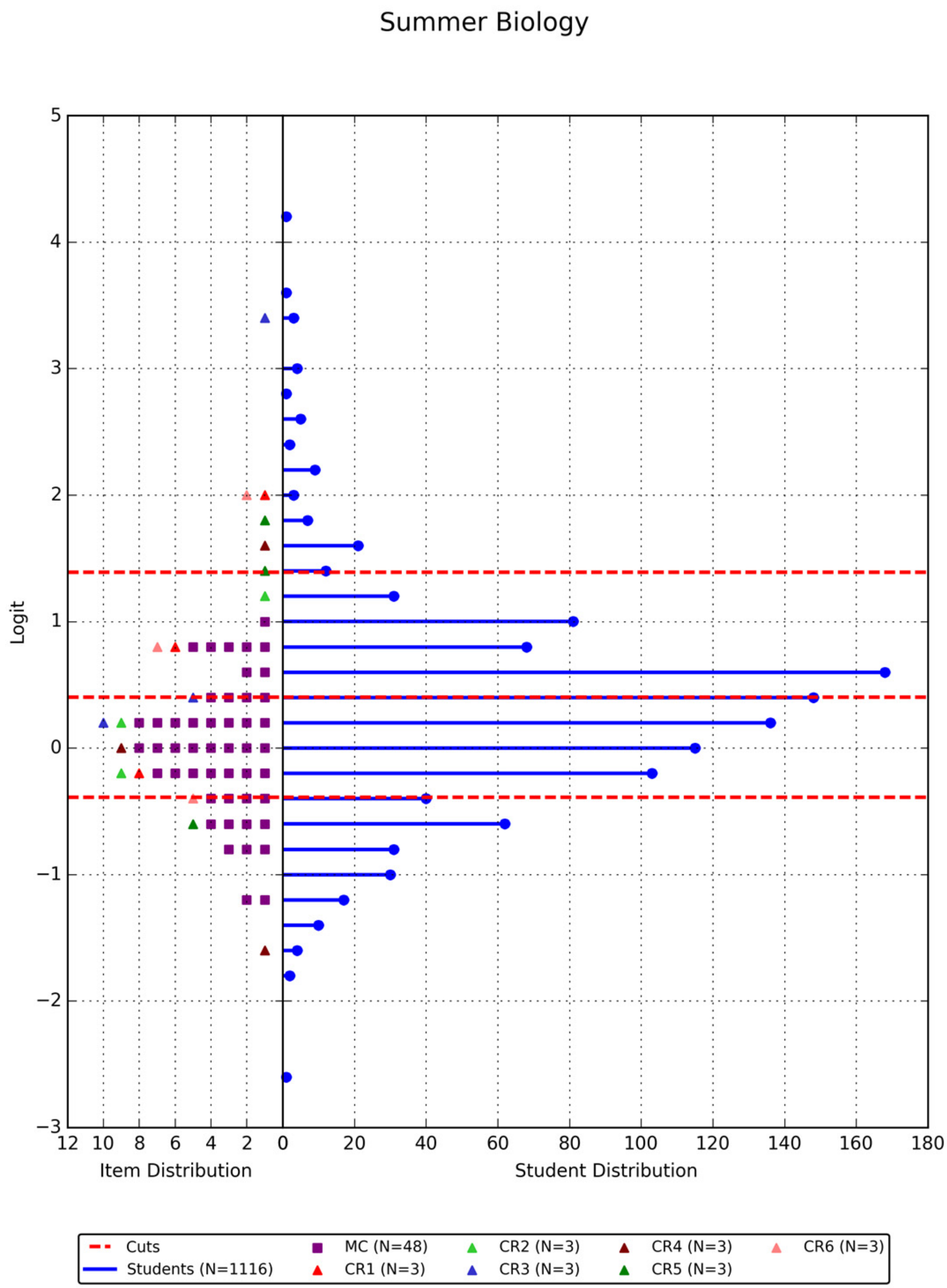
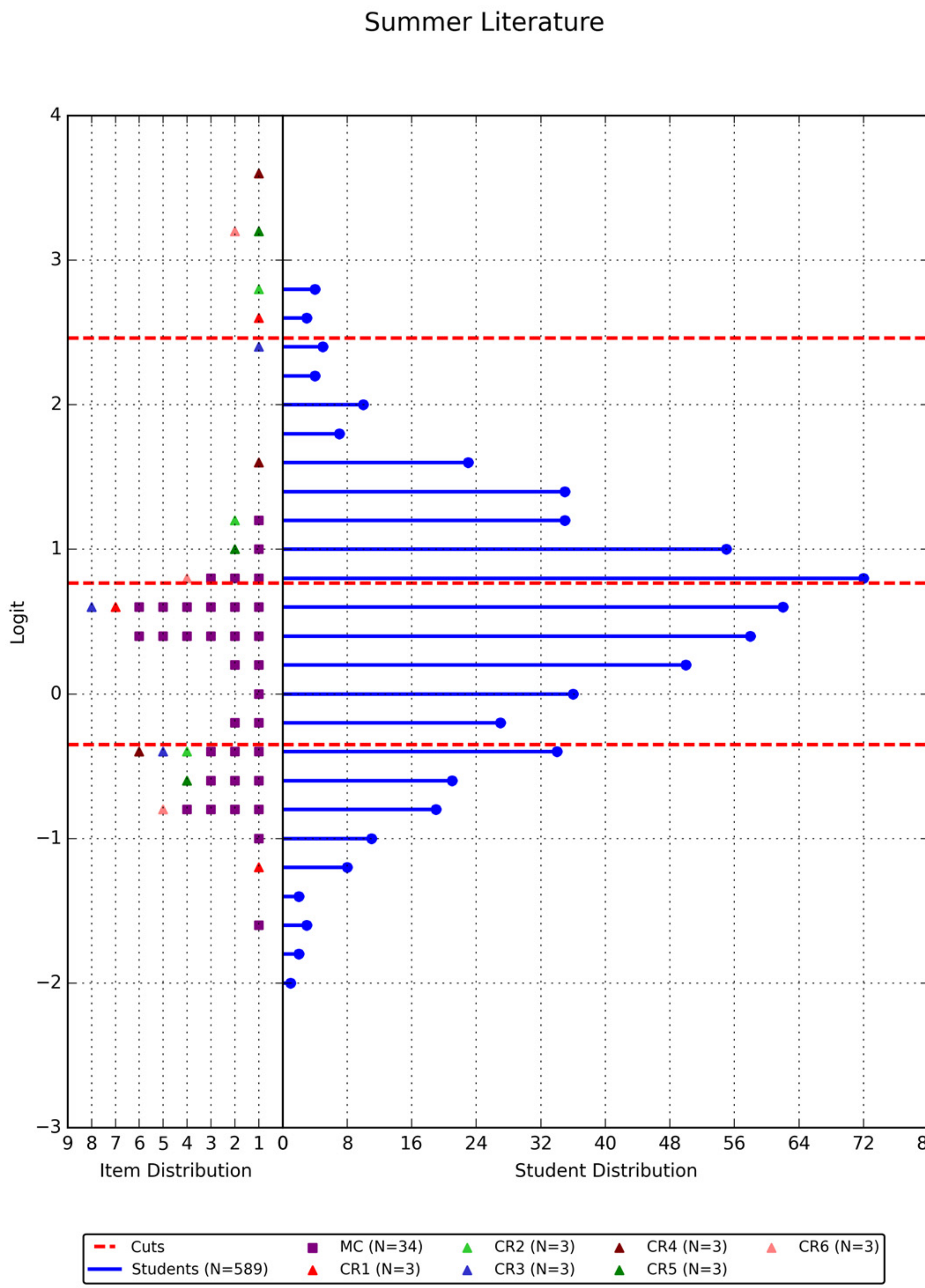


Figure 12–1 (continued). Item Difficulty-Student Ability Maps



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[7] 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

[7] 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.

- 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.

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:
    - 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?
    - 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?
    - 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?
    - 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?
    - 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?
    - Does any paragraph include information describing performance that might be too demanding or too easy? If so, what revisions need to be made?
4. The results of the discussion were summarized, and revisions to each general PLD paragraph were made. Committee consensus was reached.
5. Once consensus was reached, the paragraphs describing performance at each level were reviewed again by the committee to confirm the following:
    - The PLD paragraphs show a clear progression from one performance level to the next level.
    - The PLD paragraphs are consistent with the Pennsylvania Policy Definitions.
    - The PLD paragraphs are aligned to the Assessment Anchors and Eligible Content.

6. The results of the discussion were summarized, suggested revisions were made, and committee members' feedback was incorporated into the paragraphs. Committee consensus was reached.
7. 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 I		Biology		Literature	
		N	Percentage	*N*	Percentage	*N*	Percentage
Gender	Male	9	60.0%	5	38.5%	5	45.5%
	Female	6	40.0%	8	61.5%	6	54.5%
Ethnicity	Asian	1	6.7%	0	0.0%	0	0.0%
	American Indian	0	0.0%	0	0.0%	0	0.0%
	Black	1	6.7%	0	0.0%	1	9.1%
	Latino	0	0.0%	0	0.0%	0	0.0%
	Multi-Race	0	0.0%	0	0.0%	0	0.0%
	White	13	86.7%	13	100.0%	10	90.9%
Role	Classroom Teacher	8	53.3%	9	69.2%	4	36.4%
	Educator	3	20.0%	0	0.0%	0	0.0%
	Higher Education Educator	3	20.0%	1	7.7%	4	36.4%
	Other	1	6.7%	3	23.1%	3	27.3%
Special Education	Yes	7	46.7%	7	53.8%	4	36.4%
	No	2	13.3%	4	30.8%	3	27.3%
	N/A	6	40.0%	2	15.4%	4	36.4%
LEP	Yes	4	26.7%	5	38.5%	2	18.2%
	No	4	26.7%	6	46.2%	5	45.5%
	N/A	6	40.0%	2	15.4%	4	36.4%
	Missing	1	6.7%	0	0.0%	0	0.0%
Region	Urban	3	20.0%	2	15.4%	2	18.2%
	Suburban	7	46.7%	5	38.5%	5	45.5%
	Rural	4	26.7%	6	46.2%	3	27.3%
	Other	1	6.7%	0	0.0%	1	9.1%
Experience	Less than 10 years	0	0.0%	5	38.5%	0	0.0%
	10–20 years	4	26.7%	2	15.4%	4	36.4%
	20–30 years	8	53.3%	4	30.8%	3	27.3%
	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:

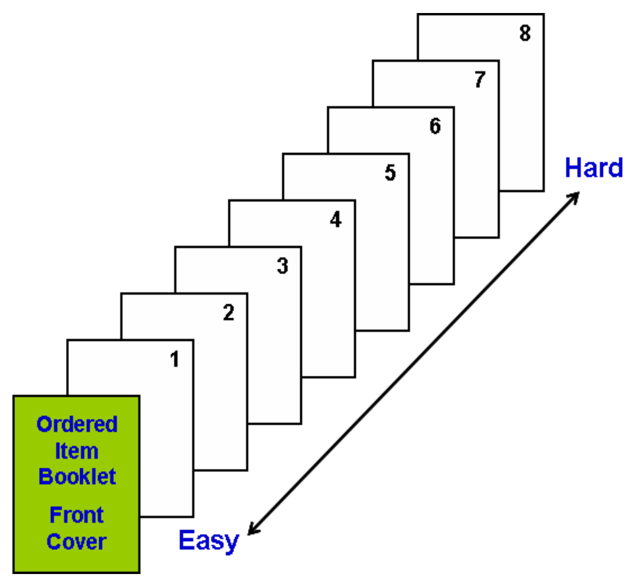
- Item Map
- Item Separation Map
- Ordered Item Booklet (OIB)
- Passages
- Scoring Rubrics
- 2011 Operational Test Form

- 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 Supplemented
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

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

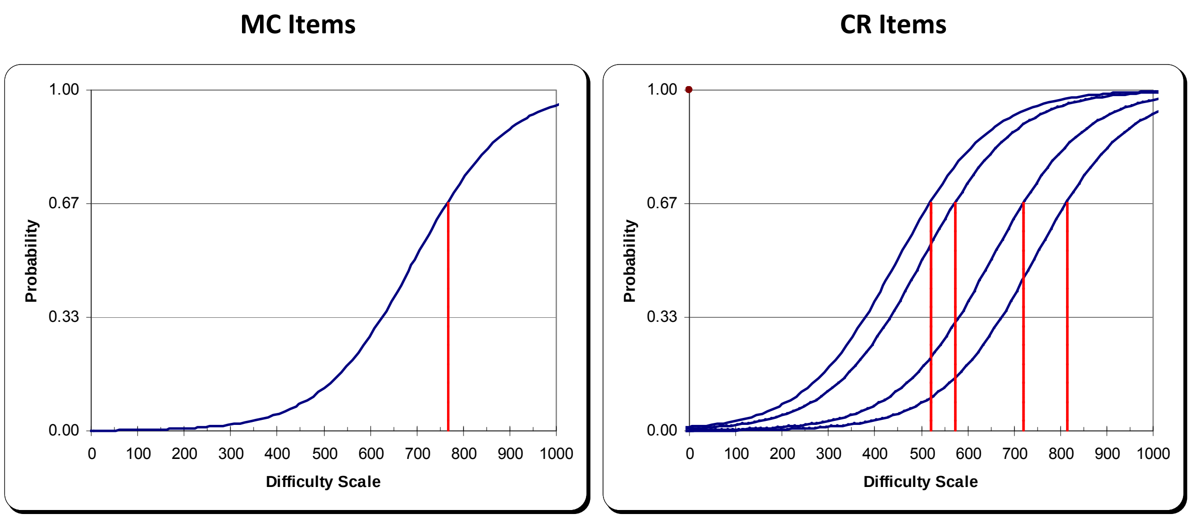
where m_i is the number of thresholds and, for notational convenience,

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

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.

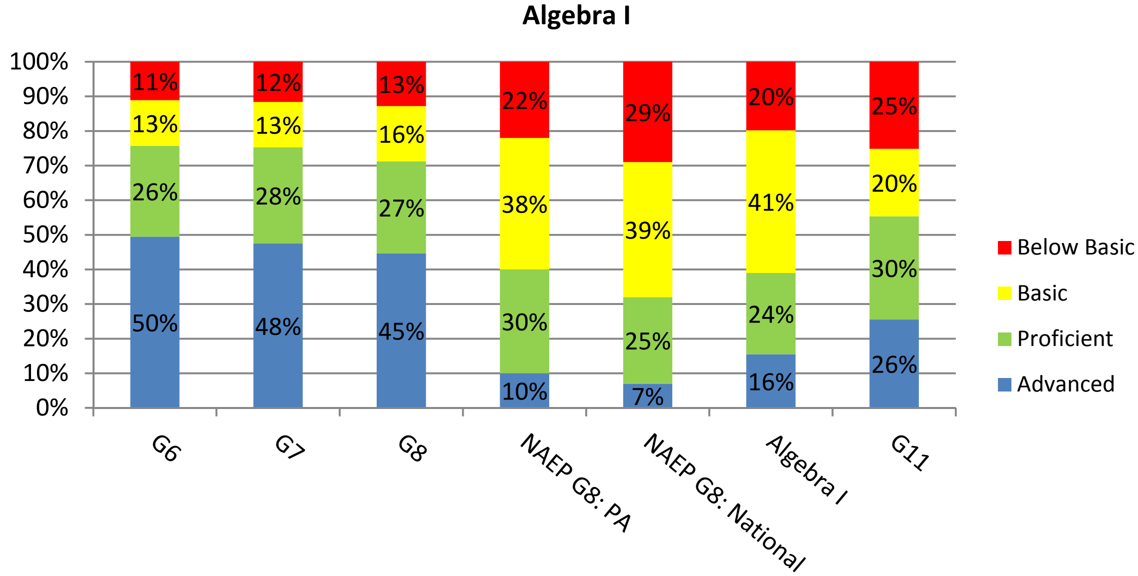
Round 3. Round 3 began with a discussion of Round 2 results. The process followed in Round 2 was used. More specifically, the individual panelist's Round 2 bookmark placements, the median bookmark placements from Round 2, and the percentage of students in each performance level were presented. Panelists 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.

Chapter Thirteen: Standard Setting

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:

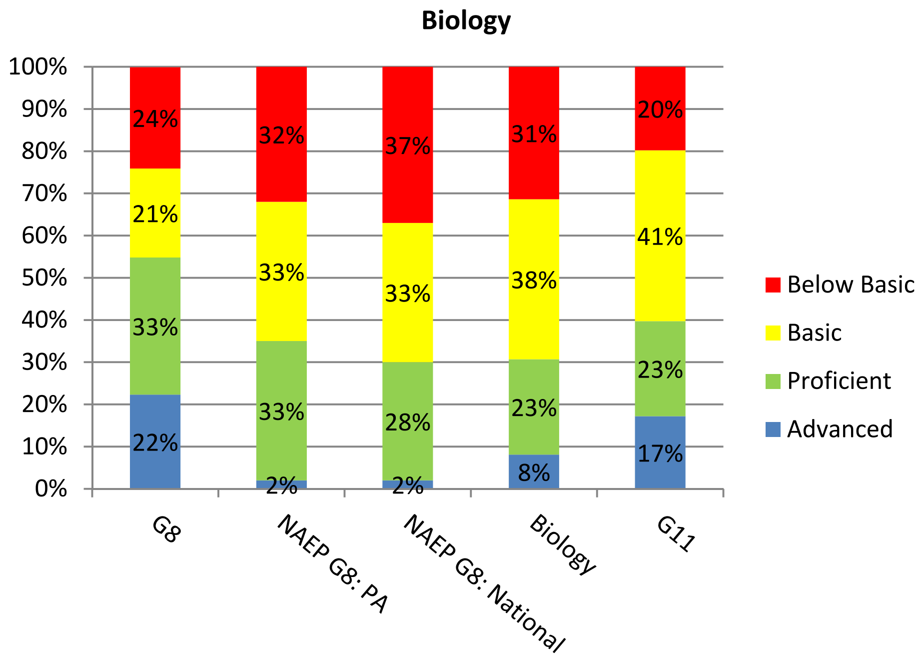
- The PSSA and NAEP results were based on students' performance in 2009. The PSSA results were from grades 6–8 and 11. The NAEP results were from grade 8.
- All students in grades 6–8 and 11 in Pennsylvania took the PSSA. A sample that represents the Pennsylvania grade 8 students took the NAEP tests.
- 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



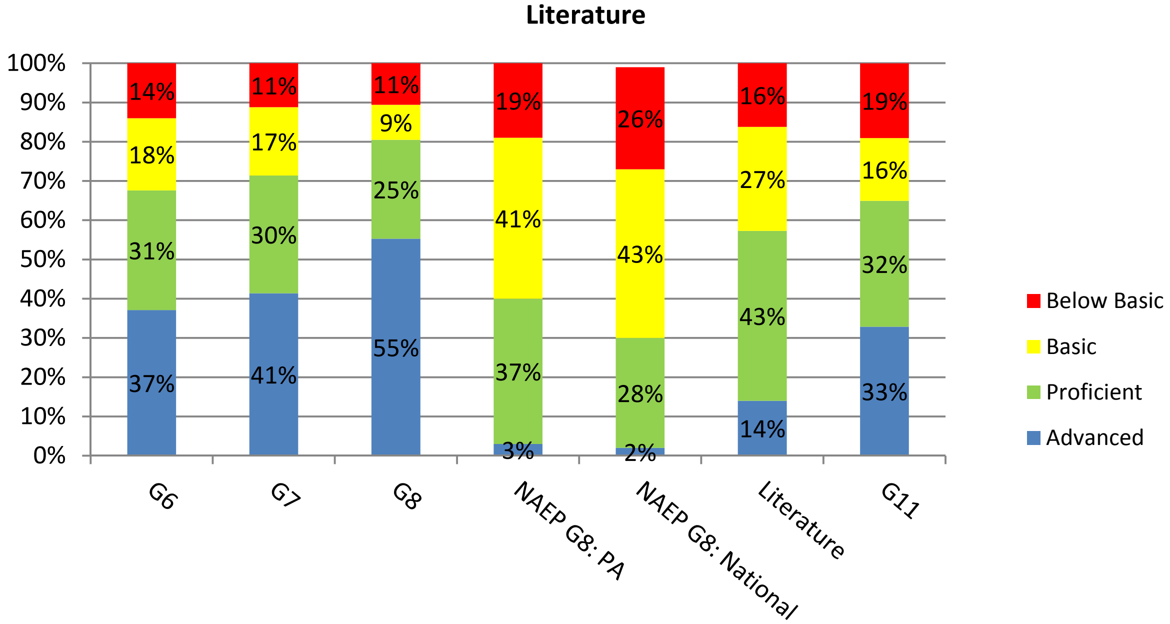
Performance Level	PSSA			NAEP Math		KE	PSSA	College Ready - Yes	SAT Math: College Ready - Yes	
	G6	G7	G8	NAEP G8: PA	NAEP G8: National	Algebra I	G11	Projected	PA	National
Below Basic	11.1%	11.6%	12.8%	22.0%	29.0%	19.8%	25.1%	0.9%		
Basic	13.2%	13.1%	16.0%	38.0%	39.0%	41.2%	19.5%	7.8%		
Proficient	26.2%	27.8%	26.6%	30.0%	25.0%	23.5%	29.8%	42.2%		
Advanced	49.5%	47.5%	44.7%	10.0%	7.0%	15.5%	25.5%	91.2%		
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%**		
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	38.1%		
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



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

Figure 13–3C. External Impact Data



Performance Level	PSSA			NAEP Reading		KE	PSSA	College Ready - Yes	SAT Critical Reading: College Ready - Yes	
	G6	G7	G8	NAEP G8: PA	NAEP G8: National	Literature	G11	Projected	PA	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%		
Proficient	30.5%	30.0%	25.2%	37.0%	28.0%	43.3%	32.1%	24.8%		
Advanced	37.1%	41.4%	55.3%	3.0%	2.0%	14.0%	32.9%	76.7%		
Below Basic + Basic	32.4%	28.6%	19.5%	60.0%	69.0%	42.7%	35.1%	2.7%		
Proficient + Advanced	**67.6%**	**71.4%**	**80.5%**	**40.0%**	**30.0%**	**57.3%**	**64.9%**	**51.1%**		
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	34.5%		
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

		Bookmark Page Number								
		Below Basic/Basic			Basic/Proficient			Proficient/Advanced		
Exam	Round	Median	Min.	Max.	Median	Min.	Max.	Median	Min.	Max.
Algebra I	1	11	6	19	28	14	42	45	33	56
	2	11	6	13	26	17	33	42	40	46
	3	11	10	12	26	18	30	46	41	46
Biology	1	9	4	15	26	20	30	56	43	62
	2	8	7	14	24	21	30	54	50	60
	3	8	7	12	22	20	30	54	50	60
Literature	1	8	5	14	27	12	34	47	38	52
	2	9	8	15	23	15	34	46	38	48
	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			Percentage in Performance Level (%)				
		BB/B	B/P	P/A	BB	B	P	A	P+A
Algebra I	Median-2SE	-0.7273	0.4291	1.3694	14.7	37.3	30.2	17.8	48.0
	Median-1SE	-0.6181	0.5659	1.5041	17.2	37.8	31.6	13.4	45.0
	Median	-0.5090	0.7027	1.6388	19.8	41.2	27.5	11.5	39.0
	Median+1SE	-0.3999	0.8395	1.7735	22.4	44.4	23.5	9.7	33.2
	Median+2SE	-0.2907	0.9763	1.9082	25.2	47.0	19.7	8.1	27.8
Biology	Median-2SE	-0.5977	0.3098	1.2500	22.5	38.9	27.9	10.7	38.6
	Median-1SE	-0.4933	0.3564	1.3205	28.3	33.1	29.3	9.3	38.6
	Median	-0.3888	0.4029	1.3910	31.4	32.7	27.8	8.1	35.9
	Median+1SE	-0.2843	0.4494	1.4615	34.5	32.2	25.2	8.1	33.3
	Median+2SE	-0.1799	0.4960	1.5320	40.7	26.0	26.4	6.9	33.3
Literature	Median-2SE	-0.6561	0.2338	1.7014	12.4	20.9	46.3	20.4	66.7
	Median-1SE	-0.5545	0.4116	1.9551	14.2	25.3	46.5	14.0	60.5
	Median	-0.4530	0.5894	2.2088	16.2	26.5	48.4	8.9	57.3
	Median+1SE	-0.3515	0.7672	2.4625	18.3	31.4	43.6	6.7	50.3
	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				Scaling Constants	
	Below Basic	Basic	Proficient	Advanced	Slope	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:

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

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[8] (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.

---

[8] 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).

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.

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	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 2014–2015.

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. For the Spring 2014 administration, the details of the number of students for each form and item type can be found in the *2014 Pennsylvania Keystone Exams Technical Report* (Pennsylvania Department of Education, 2014).

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 2014–2015 administrations, post-equating check analyses 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.

Figure 15–1. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Winter

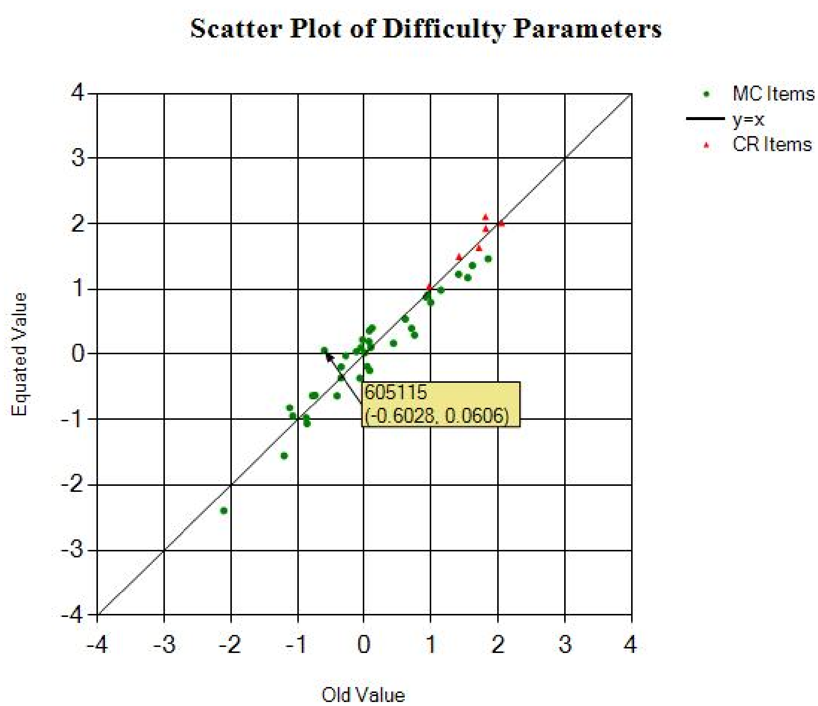


Figure 15–2. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Winter

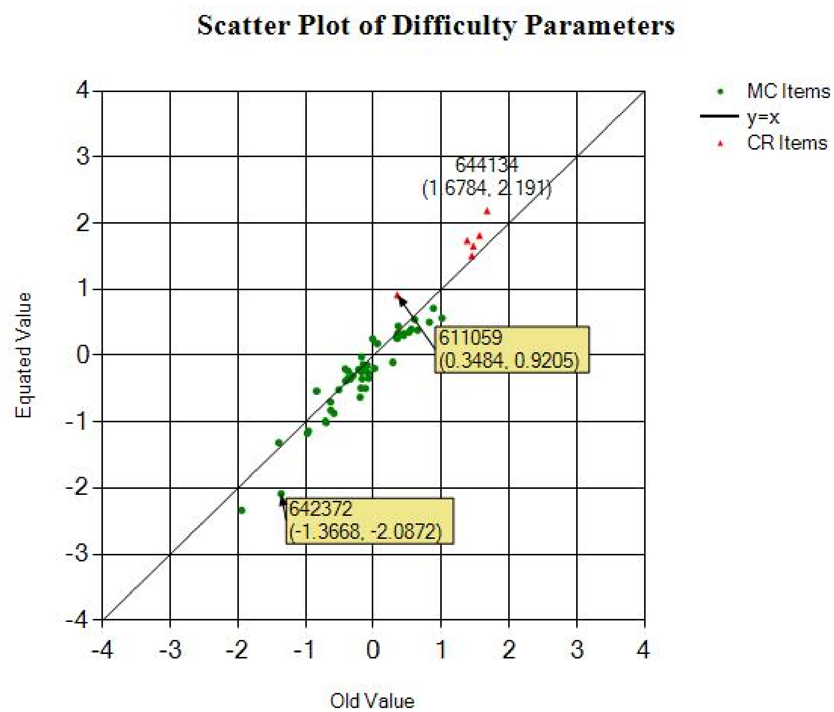


Figure 15–3. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Winter

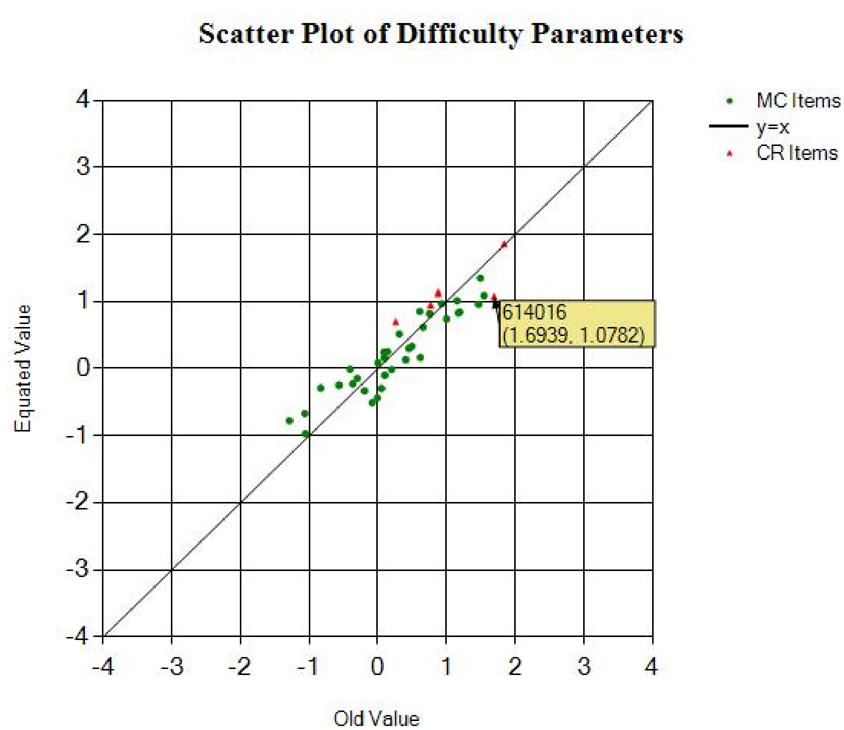


Figure 15–4. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Spring

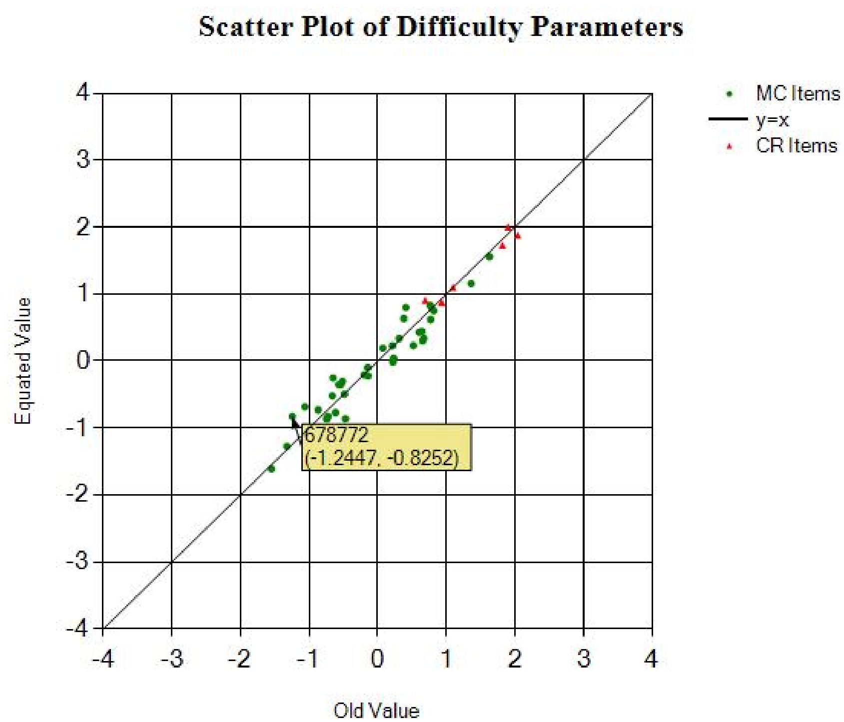


Figure 15–5. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Spring

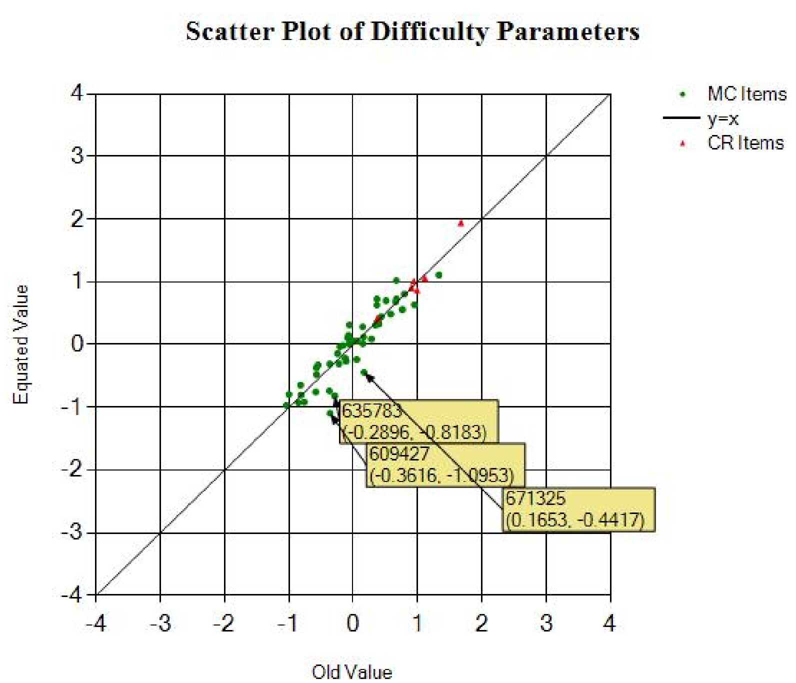


Figure 15–6. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Spring

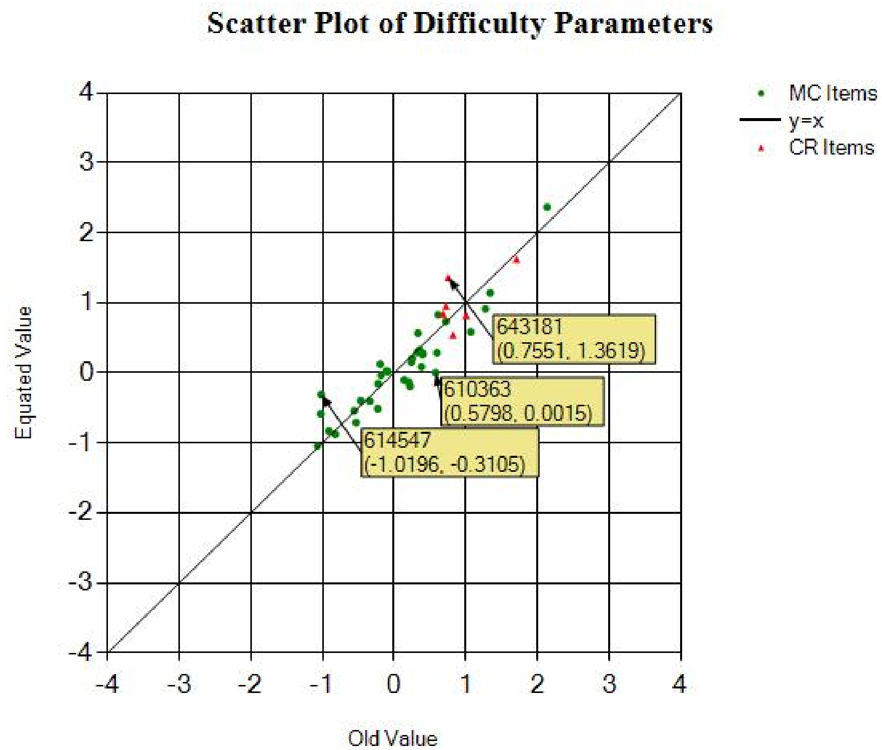


Figure 15–7. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Algebra I: Summer

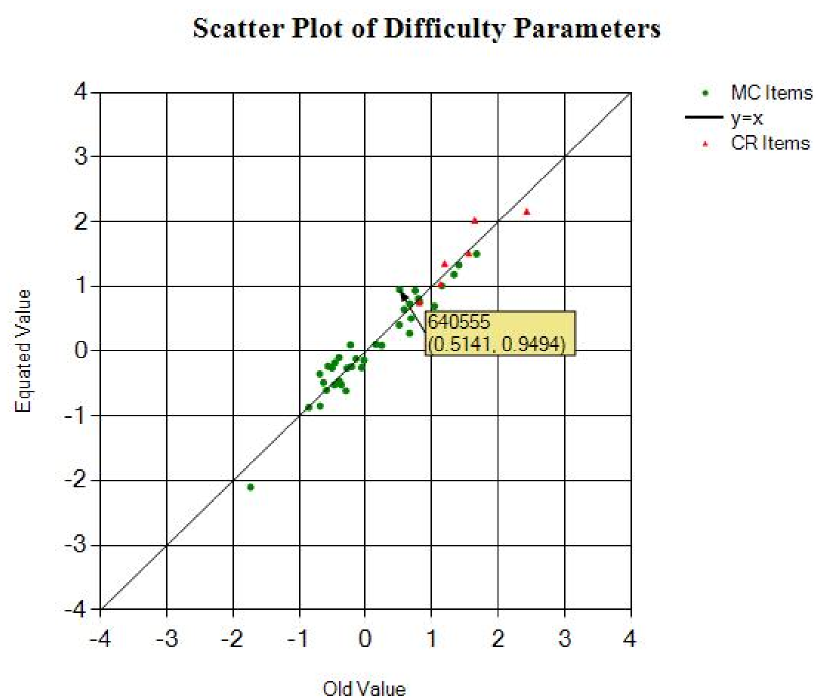


Figure 15–8. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Biology: Summer

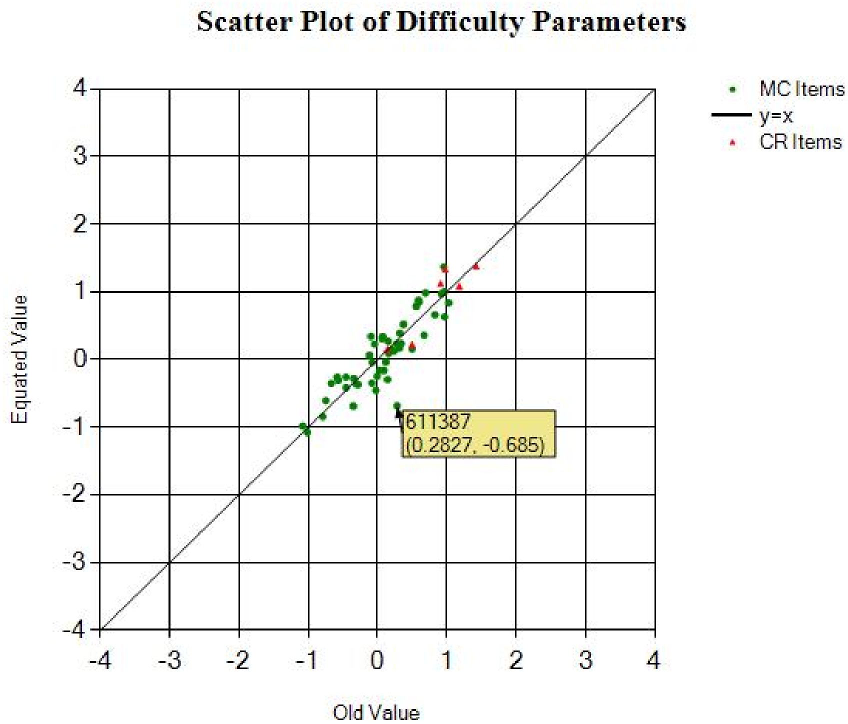


Figure 15–9. Scatter Plot of Old and Equated Item Parameter Difficulty Estimates for Literature: Summer

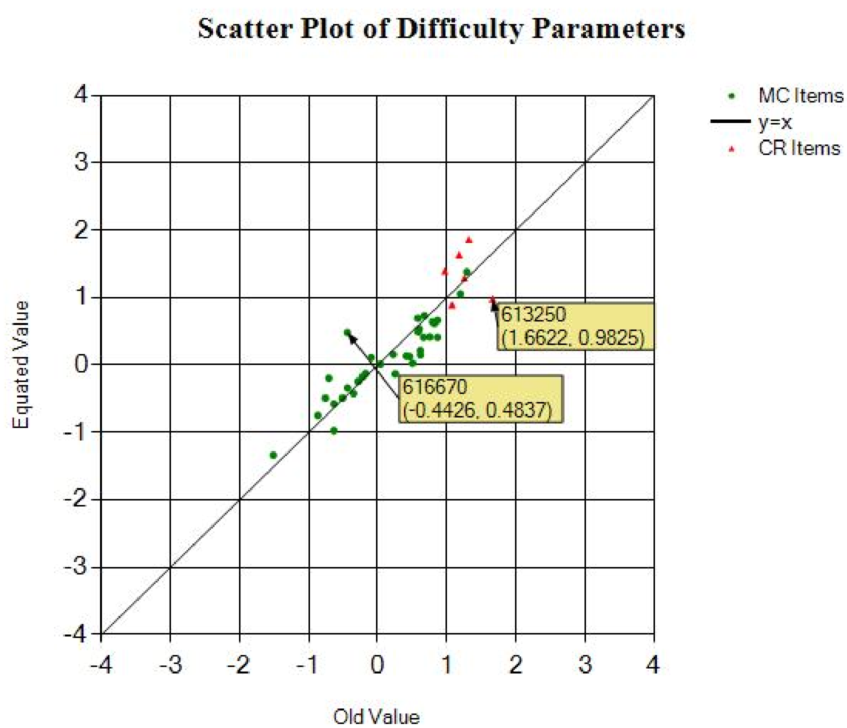


Table 15–1. Summary of Items Flagged by both the Scatter Plot and Displacement

Administration	Content Area	Item IDs
Winter	Algebra I	605115
	Biology	642372
	Literature	614016
Spring	Algebra I	NA
	Biology	635783, 609427, 671325
	Literature	610363, 614547, 643181
Summer	Algebra I	NA
	Biology	611387
	Literature	616670, 613250

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)	CR (3 point)	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.

Chapter Sixteen: Scores and Score Reports

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[9]; 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.

---

[9] 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.

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			Max
		BB/B	B/P	P/A	
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.

- **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

Chapter Sixteen: Scores and Score Reports

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

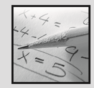
School: SAMPLE HS

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

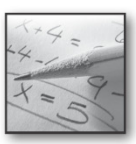
Student's total test scale score is indicated by the (⬆) . If this student were to test again under similar circumstances, his/her score would likely remain in the following range: **1520-1550.**

1200 – 1439	1439 – 1500	1500 – 1546	1546 – 1800
Below Basic	Basic	Proficient	Advanced

Highest Total Test Scale Score to Date - 1535

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.

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 17–1. Performance Level Percentages

Content Area	All Testers					First-Time Testers					Retesters				
	N	Percentage in Each Performance Level				*N*	Percentage in Each Performance Level				*N*	Percentage in Each Performance Level			
		Below Basic	Basic	Proficient	Advanced		Below Basic	Basic	Proficient	Advanced		Below Basic	Basic	Proficient	Advanced
Winter															
Algebra I	75,936	14.8	58.0	22.7	4.5	14,194	19.4	32.5	27.3	20.9	61,742	13.7	63.9	21.6	0.8
Biology	54,236	30.8	45.5	17.5	6.2	15,054	27.5	22.8	27.9	21.7	39,182	32.1	54.2	13.5	0.3
Literature	47,149	14.3	44.3	38.4	3.0	16,323	14.1	26.5	51.1	8.4	30,826	14.4	53.7	31.7	0.2
Spring															
Algebra I	184,172	15.0	45.8	25.0	14.2	121,255	15.0	34.9	29.1	21.0	62,917	15.0	66.9	17.0	1.2
Biology	154,585	23.3	29.9	25.7	21.1	115,936	19.9	22.1	30.1	27.9	38,649	33.4	53.6	12.6	0.5
Literature	140,252	11.2	30.1	50.9	7.8	114,387	9.9	22.8	57.7	9.5	25,865	16.9	62.4	20.6	0.1
Summer															
Algebra I	1,846	5.5	58.5	31.0	5.0	199	14.6	35.7	24.6	25.1	1,647	4.4	61.2	31.8	2.6
Biology	1,116	10.5	44.9	39.5	5.1	69	11.6	21.7	14.5	52.2	1,047	10.4	46.4	41.2	2.0
Literature	588	7.3	49.0	42.2	1.5	30	40.0	26.7	23.3	10.0	558	5.6	50.2	43.2	1.1

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	1479.1	40.1	1495.6	62.7	1475.4	31.6
	Biology	1480.1	40.8	1500.8	58.8	1472.1	27.4
	Literature	1492.1	45.9	1509.7	57.6	1482.8	34.9
Spring	Algebra I	1490.5	52.3	1499.6	58.1	1472.9	32.0
	Biology	1502.0	54.4	1512.1	57.3	1471.5	27.7
	Literature	1510.1	52.6	1517.8	53.1	1475.9	32.7
Summer	Algebra I	1491.9	33.4	1503.4	60.3	1490.6	28.2
	Biology	1495.9	33.1	1536.2	69.0	1493.2	27.3
	Literature	1494.3	34.1	1477.9	67.1	1495.2	31.2

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

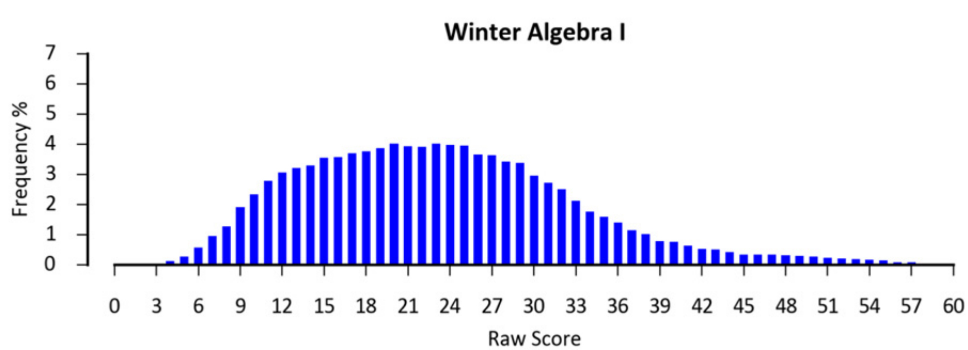
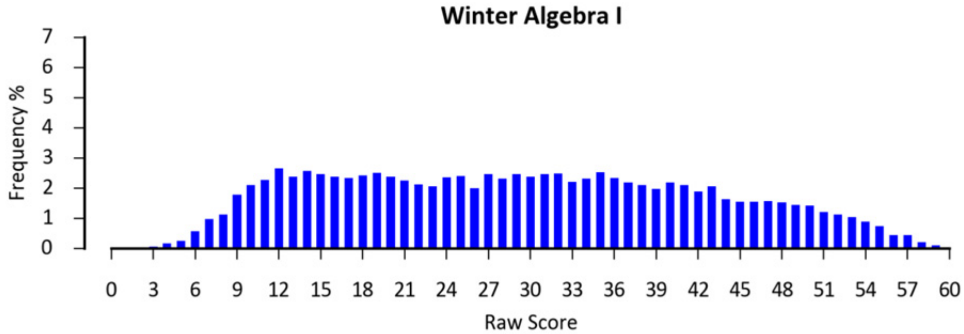
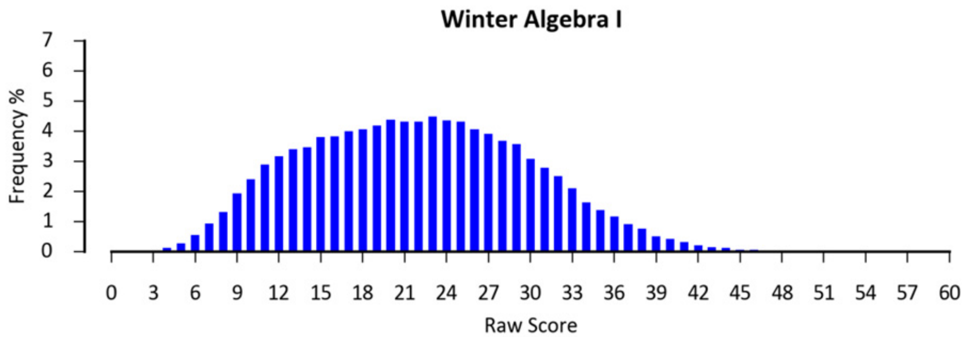


Figure 17–1 (continued). Raw Score Distributions

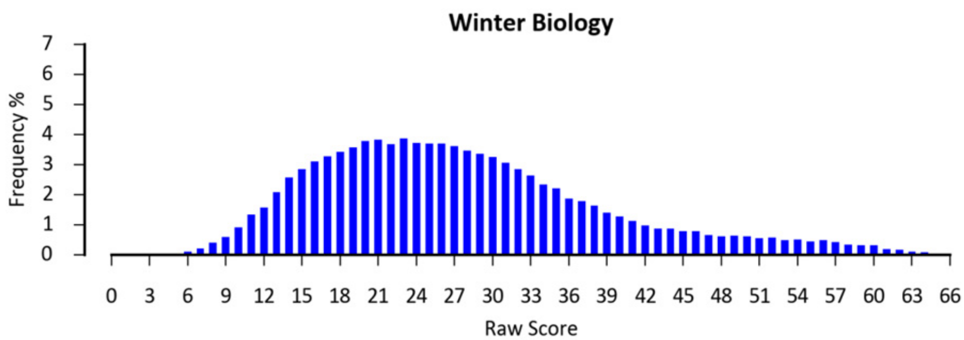
First-Time Testers



Retesters



All Testers



First-Time Testers

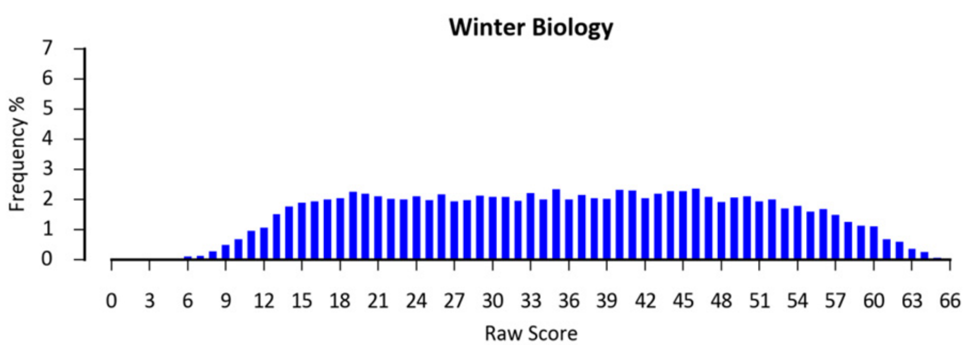
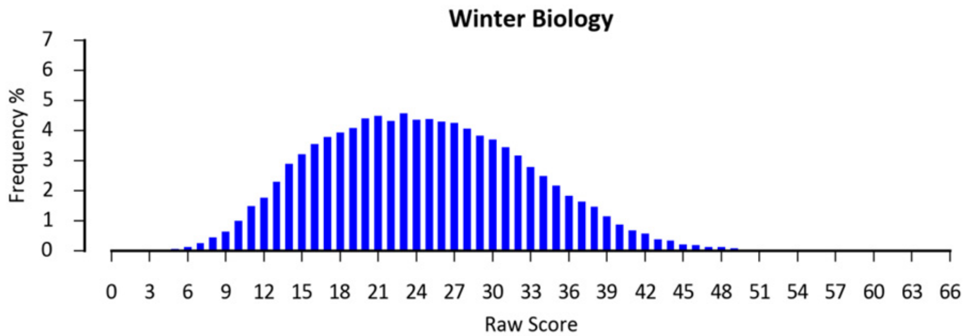
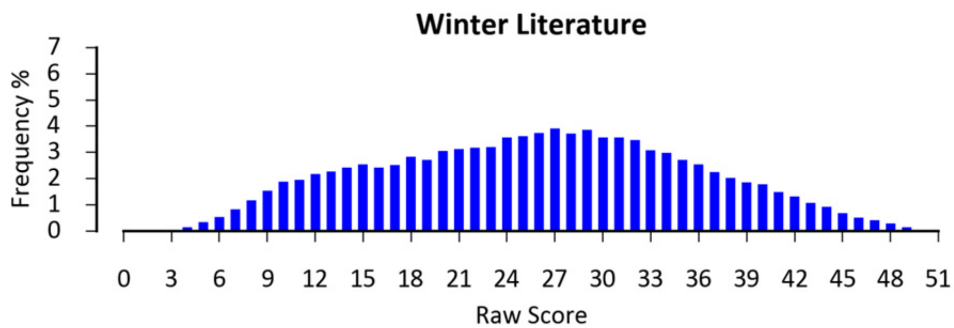


Figure 17–1 (continued). Raw Score Distributions

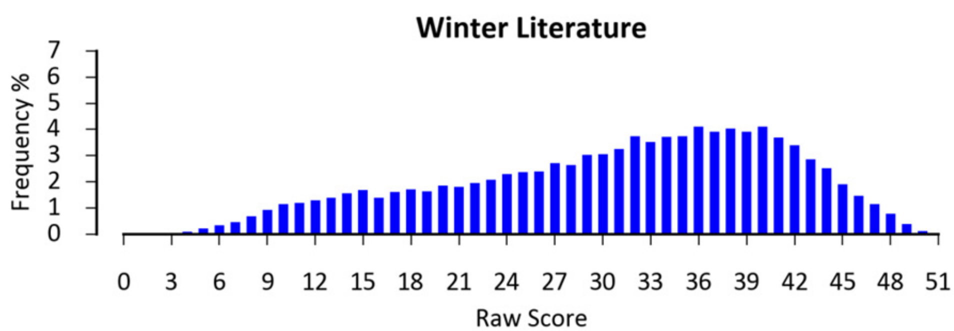
Retesters



All Testers



First-Time Testers



Retesters

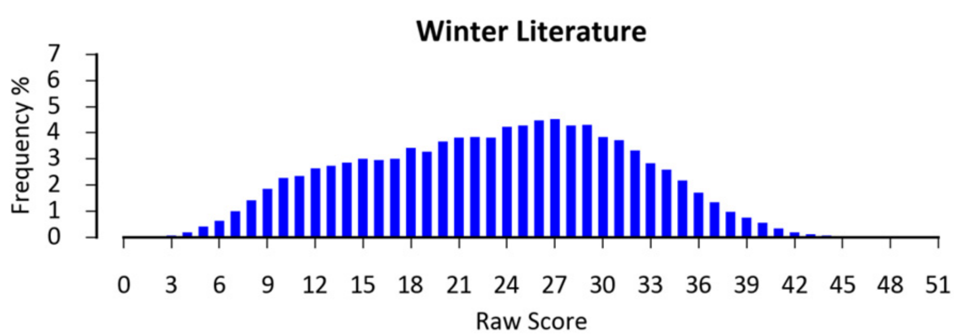
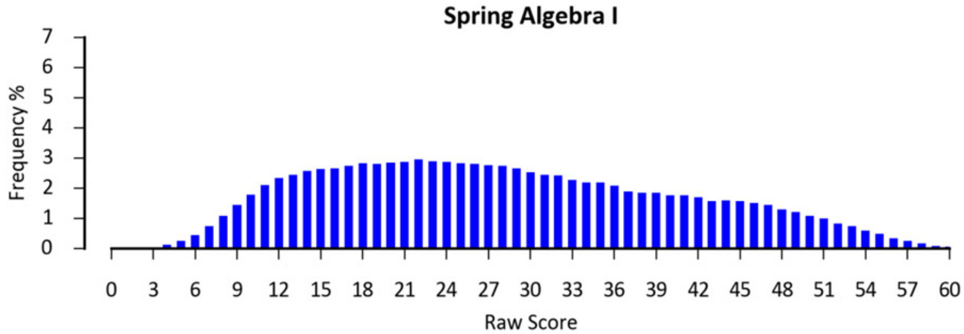
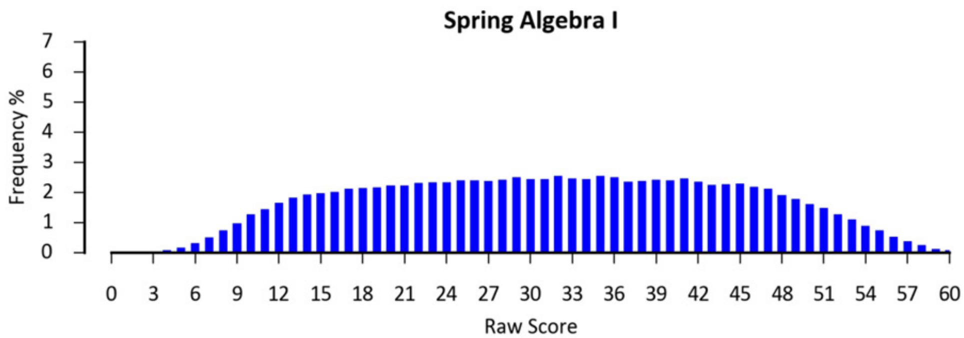


Figure 17–1 (continued). Raw Score Distributions

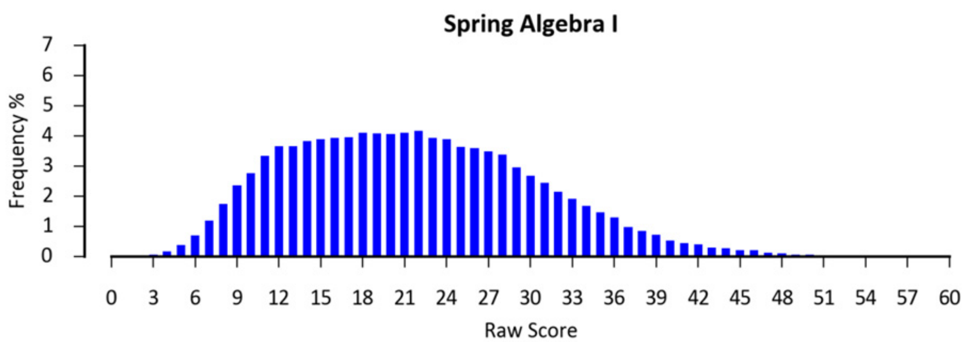
All Testers



First-Time Testers



Retesters



All Testers

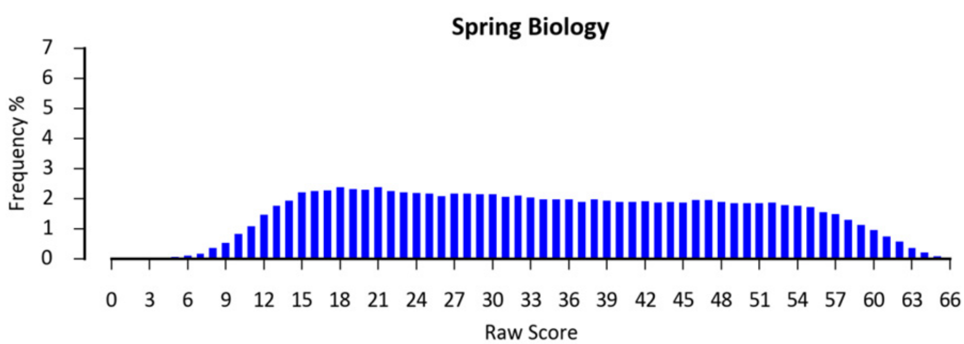
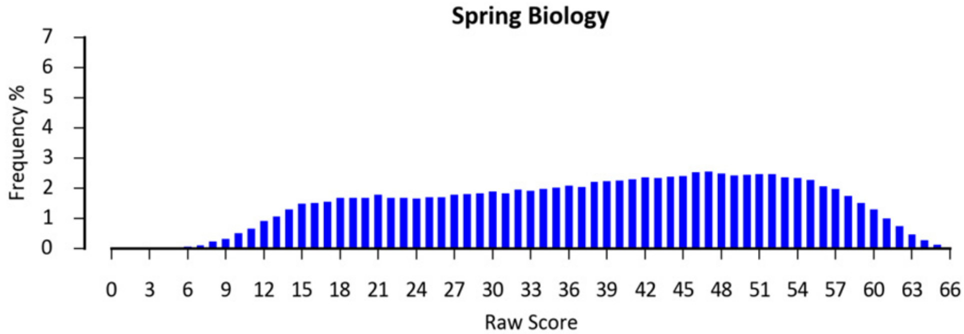
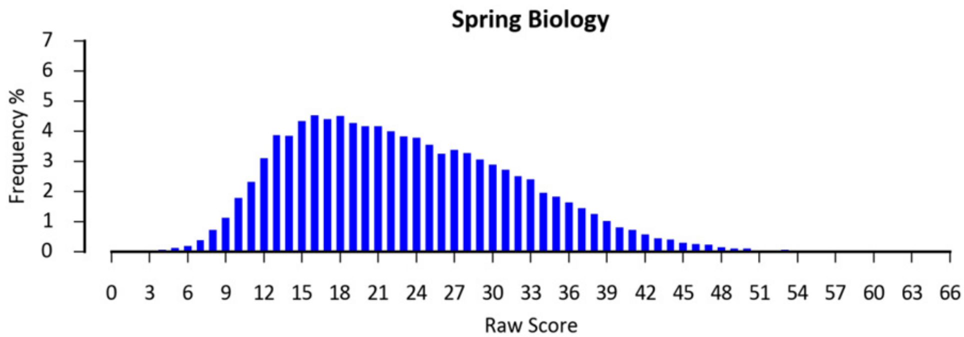


Figure 17–1 (continued). Raw Score Distributions

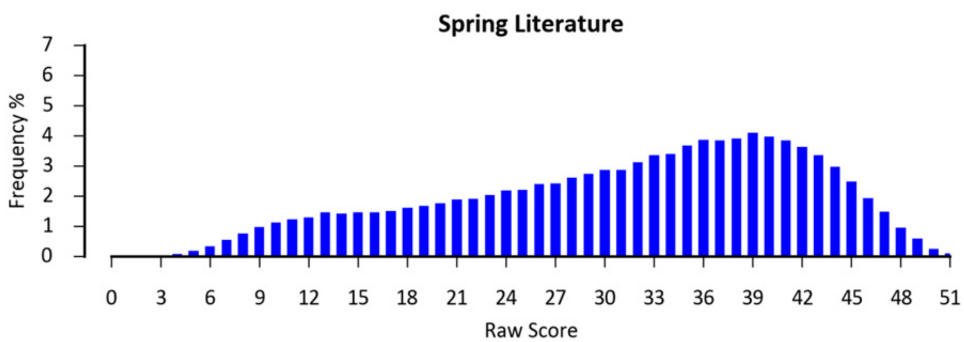
First-Time Testers



Retesters



All Testers



First-Time Testers

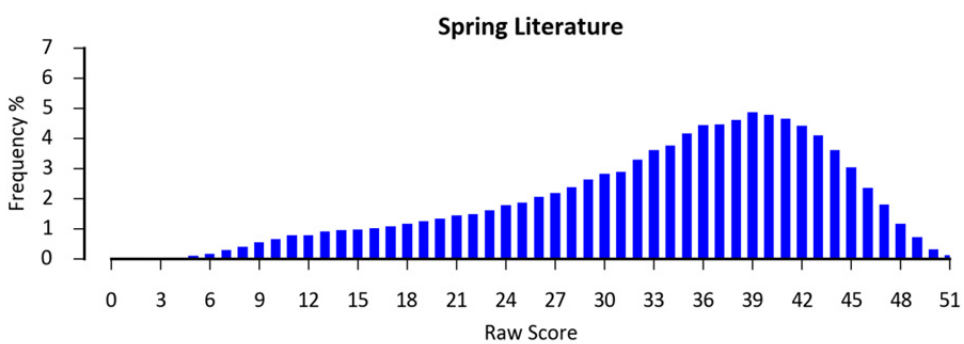
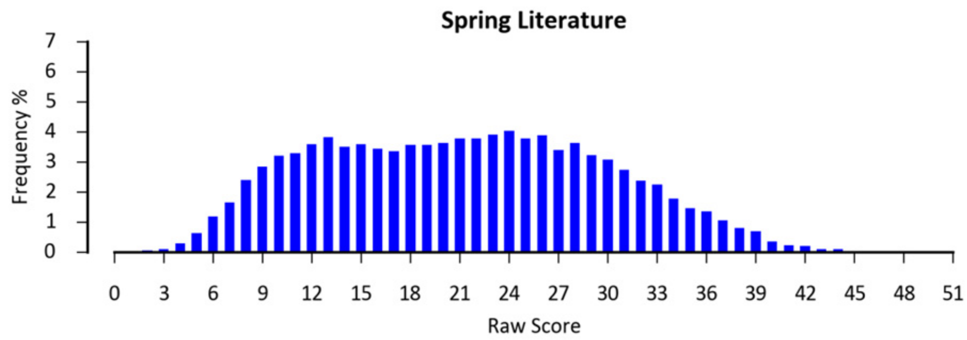
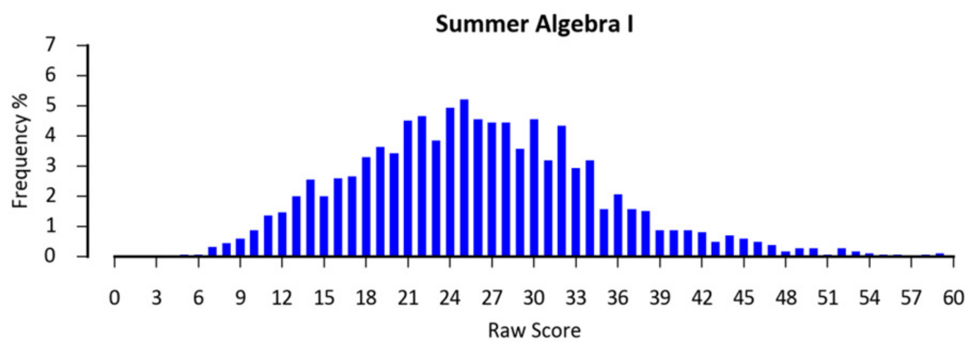


Figure 17–1 (continued). Raw Score Distributions

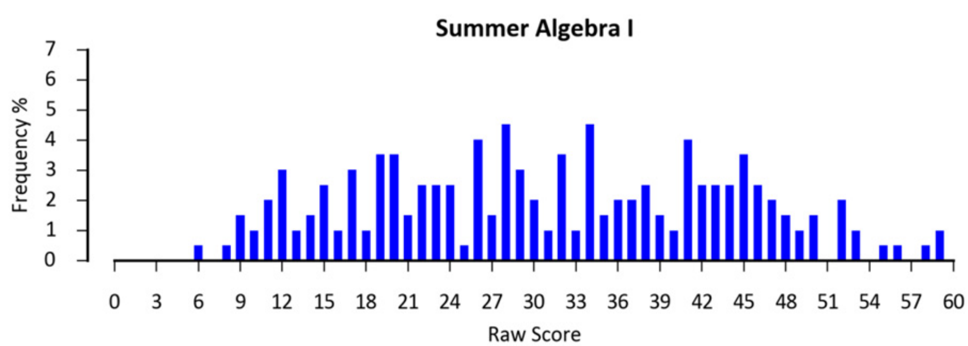
Retesters



All Testers



First-Time Testers



Retesters

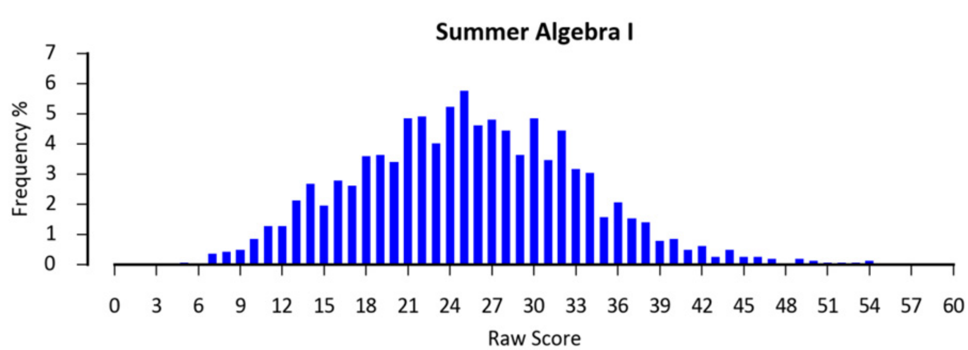
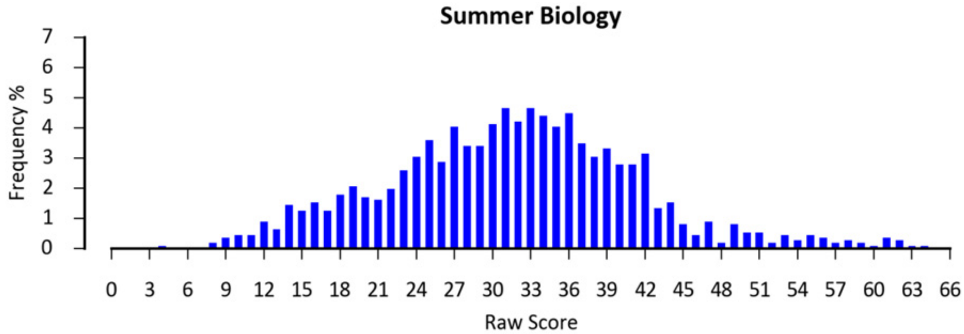
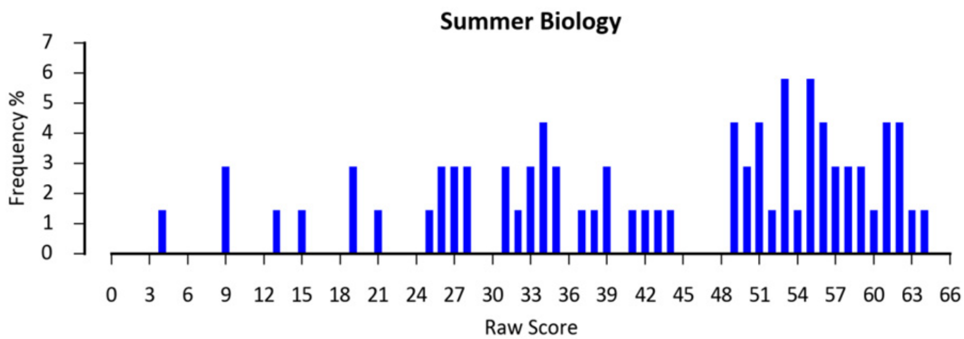


Figure 17–1 (continued). Raw Score Distributions

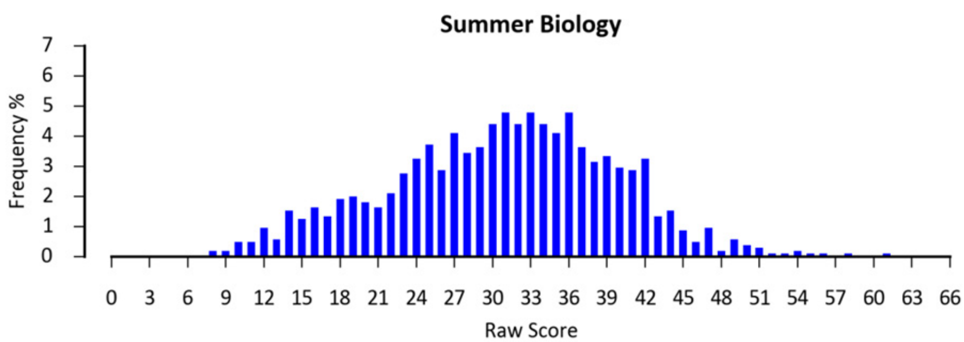
All Testers



First-Time Testers



Retesters



All Testers

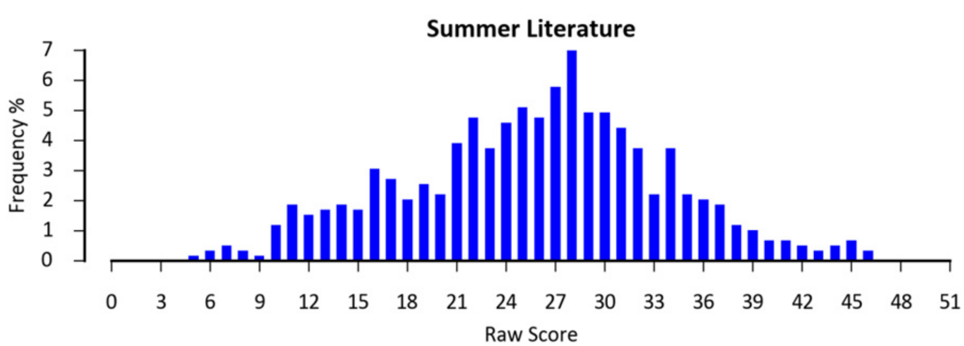
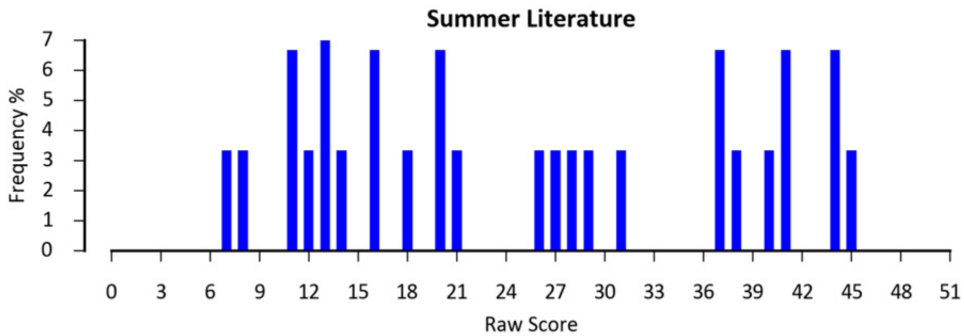
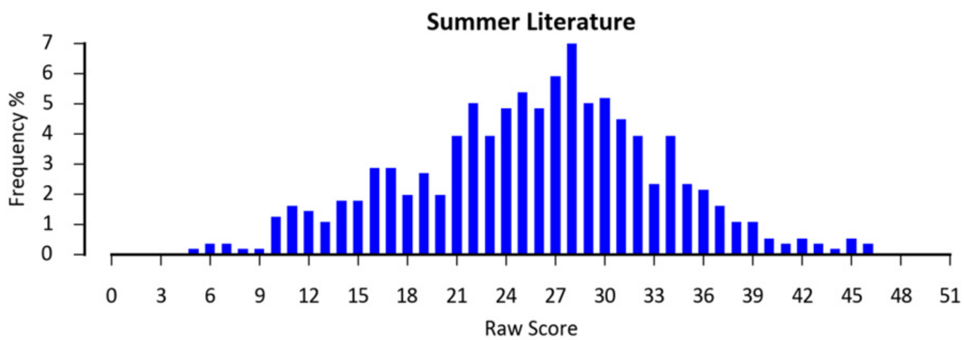


Figure 17–1 (continued). Raw Score Distributions

First-Time Testers



Retesters



Chapter Seventeen: Operational Test Statistics

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.[10]

$$\rho_X^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

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:

---

[10] 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 σ_{Yi}^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	Essentially-Tau Equivalent	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

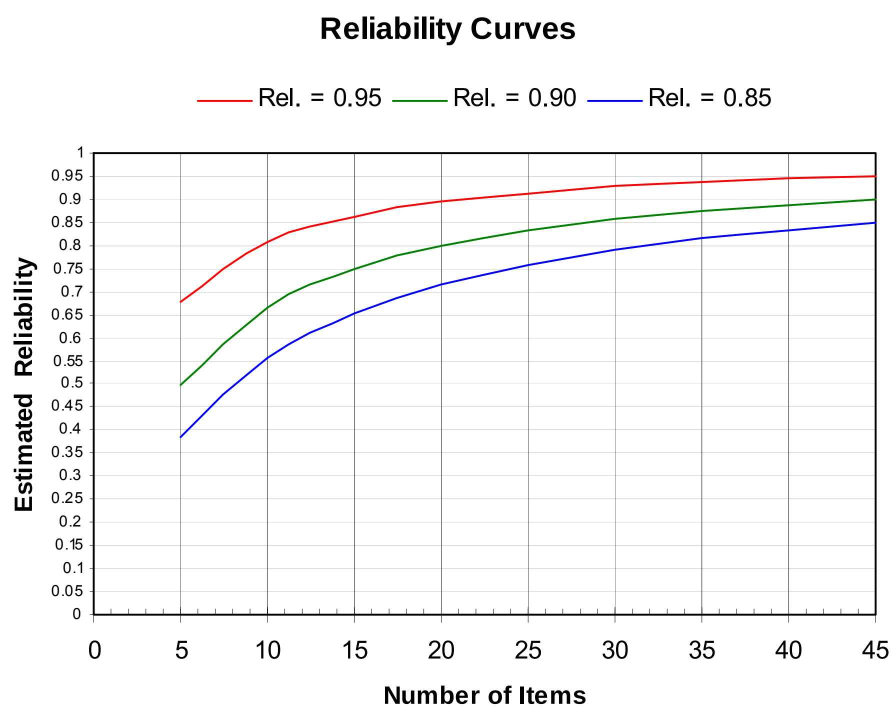
As emphasized in earlier sections, coefficient alpha only takes into account measurement error that arises from the selection of items used on a particular test form. There are other sources of random inaccuracy. One is due to the occasion of testing. Examples of other various random conditions that 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,[11] 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[12] 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 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.[13] 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.

---

[11] True score is the score the person would receive if the measurement process were perfect.
[12] 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.
[13] 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.

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.

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

$$\hat{\sigma}_T = \sqrt{\hat{\sigma}_x^2 - \hat{\sigma}_x^2(1 - \hat{\rho}_{xx})}\,.$$

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.

$$CSEM(\text{SS}) = CSEM(\hat{\beta}_n) * m$$

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

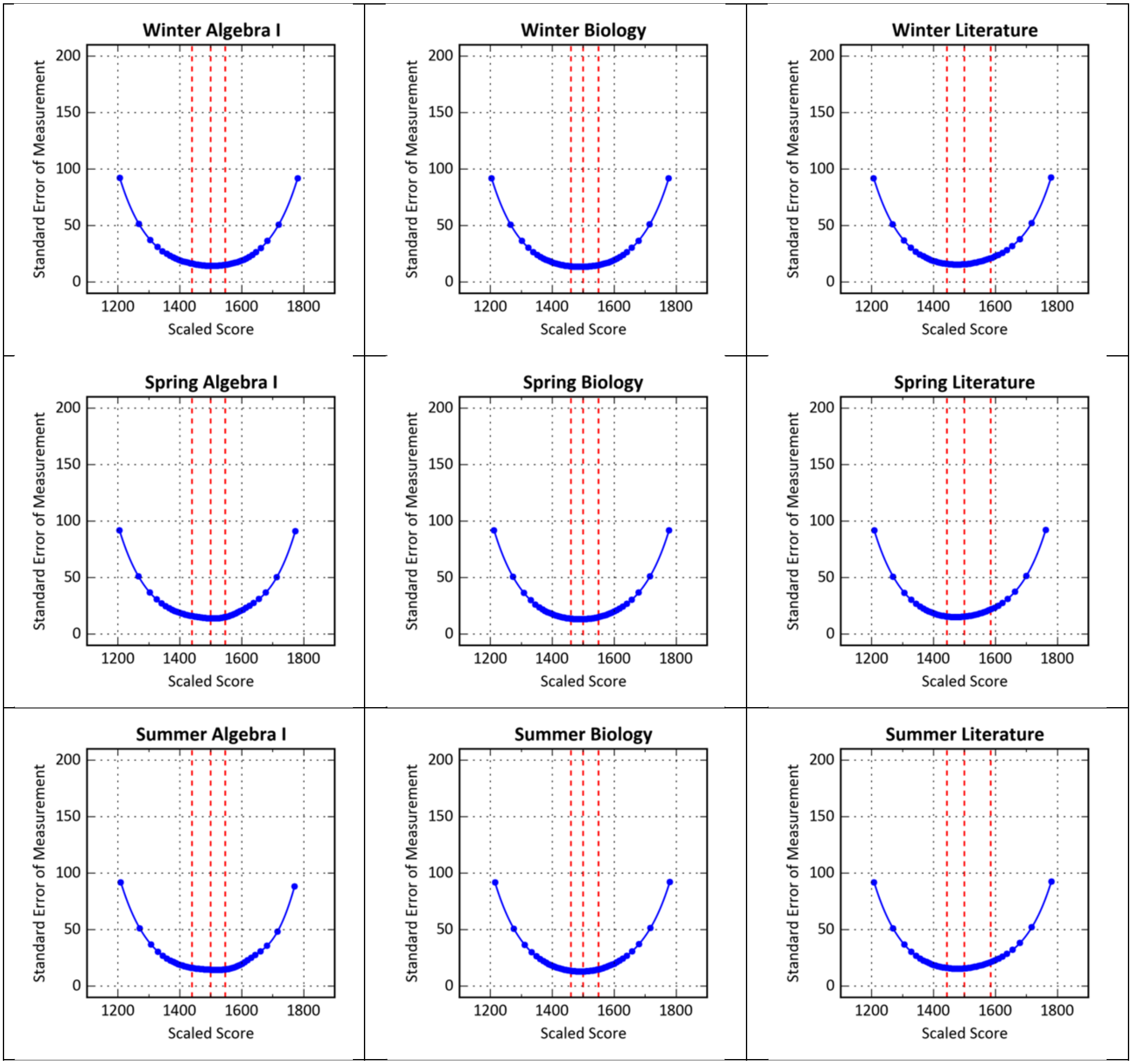
The CSEMs documented on the Keystone Exams score reports are the Rasch-based conditional standard errors of measurement described above. These are provided by the program WINSTEPS described in Chapter Twelve. As noted earlier, these CSEMs are based on the concept of statistical information. For the purpose 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.[14] 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.

[14] 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



Chapter Eighteen: Reliability

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 0.73 to 0.82 and the decision consistency ranged from 0.64 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 0.94 and the decision consistency ranged from 0.84 to 0.92. 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 84% 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	0.79	0.89	0.92	0.98
		Consistency	0.70	0.84	0.89	0.97
	Biology	Accuracy	0.81	0.90	0.94	0.98
		Consistency	0.74	0.86	0.92	0.97
	Literature	Accuracy	0.80	0.92	0.90	0.98
		Consistency	0.73	0.89	0.87	0.97
Spring	Algebra I	Accuracy	0.79	0.92	0.92	0.95
		Consistency	0.71	0.88	0.89	0.93
	Biology	Accuracy	0.82	0.93	0.94	0.95
		Consistency	0.75	0.91	0.91	0.93
	Literature	Accuracy	0.80	0.95	0.92	0.92
		Consistency	0.72	0.93	0.88	0.91
Summer	Algebra I	Accuracy	0.78	0.91	0.90	0.97
		Consistency	0.69	0.87	0.86	0.96
	Biology	Accuracy	0.76	0.91	0.89	0.97
		Consistency	0.67	0.87	0.84	0.96
	Literature	Accuracy	0.73	0.94	0.90	0.89
		Consistency	0.64	0.91	0.86	0.86

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 76 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 of CR Items: Winter

Content Area	Item	Inter-Rater Agreement %		% Exact + Adjacent Agreement	Percentage Awarded for Each Score Point					
		Exact	Adjacent		0	1	2	3	4	B/NS
Algebra I	1A	99	1	100	34	60	NA	NA	NA	6
	1B	100	0	100	79	15	NA	NA	NA	6
	1C	100	0	100	74	19	NA	NA	NA	6
	1D	100	0	100	90	3	NA	NA	NA	6
	2	94	6	100	45	34	8	2	1	9
	3A	100	0	100	85	2	NA	NA	NA	12
	3B	100	0	100	52	36	NA	NA	NA	12
	3C	100	0	100	65	1	21	NA	NA	12
	4A	99	1	100	76	14	NA	NA	NA	9
	4B	99	1	100	84	6	NA	NA	NA	9
	4C	100	0	100	39	51	NA	NA	NA	9
	4D	99	1	100	80	10	NA	NA	NA	9
	5A	100	0	100	68	21	NA	NA	NA	10
	5B	100	0	100	81	9	NA	NA	NA	10
	5C	100	0	100	84	5	NA	NA	NA	10
	5D	100	0	100	86	4	NA	NA	NA	10
	6A	99	1	100	48	41	NA	NA	NA	11
	6B	100	0	100	84	5	NA	NA	NA	11
	6C	100	0	100	49	39	NA	NA	NA	11
	6D	100	0	100	54	35	NA	NA	NA	11
Biology	1	88	11	99	23	38	26	1	NA	11
	2	92	8	100	62	20	5	1	NA	12
	3	93	7	100	40	33	9	2	NA	16
	4	90	9	99	59	20	7	2	NA	11
	5	98	2	100	57	22	5	2	NA	14
	6	88	12	100	25	40	15	4	NA	15
Literature	1	82	18	100	18	42	24	5	NA	11
	2	84	16	100	16	30	39	4	NA	11
	3	86	14	100	13	36	25	7	NA	19
	4	80	20	100	16	37	27	6	NA	13
	5	85	15	100	6	36	40	6	NA	12
	6	86	14	100	8	34	35	6	NA	16

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 + Adjacent Agreement	Percentage Awarded for Each Score Point					
		Exact	Adjacent		0	1	2	3	4	B/NS
Algebra I	1A	98	2	100	77	6	10	NA	NA	7
	1B	99	1	100	59	34	NA	NA	NA	7
	1C	99	1	100	75	18	NA	NA	NA	7
	2A	99	1	100	68	24	NA	NA	NA	8
	2B	98	2	100	66	26	NA	NA	NA	8
	2C	99	1	100	69	23	NA	NA	NA	8
	2D	100	0	100	83	9	NA	NA	NA	8
	3A	98	2	100	57	13	21	NA	NA	9
	3B	99	1	100	62	29	NA	NA	NA	9
	3C	97	3	100	35	56	NA	NA	NA	9
	3MU	100	0	100	91	0	NA	NA	NA	9
	4A	99	1	100	37	57	NA	NA	NA	6
	4B	99	1	100	55	39	NA	NA	NA	6
	4C	100	0	100	88	6	NA	NA	NA	6
	4D	100	0	100	91	3	NA	NA	NA	6
	5A	99	1	100	64	29	NA	NA	NA	7
	5B	98	2	100	51	42	NA	NA	NA	7
	5C	98	2	100	51	42	NA	NA	NA	7
	5D	98	3	100	61	32	NA	NA	NA	7
	5MU	100	0	100	93	0	NA	NA	NA	7
	6	93	7	100	23	30	17	12	8	9
Biology	1	85	15	100	17	25	30	20	NA	7
	2	84	16	100	32	26	20	12	NA	9
	3	84	15	99	23	32	26	8	NA	11
	4	83	17	100	20	38	23	10	NA	9
	5	83	16	99	45	26	16	3	NA	9
	6	91	9	100	15	31	26	18	NA	10
Literature	1	77	23	100	17	30	37	8	77	8
	2	77	23	100	23	27	30	11	77	9
	3	78	22	100	10	25	39	15	78	10
	4	76	24	100	7	28	45	11	76	9
	5	77	23	100	9	21	40	21	77	9
	6	80	20	100	12	29	34	15	80	10

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 + Adjacent Agreement	Percentage Awarded for Each Score Point					
		Exact	Adjacent		0	1	2	3	4	B/NS
Algebra I	1A	99	1	100	64	14	19	NA	NA	3
	1B	99	1	100	58	39	NA	NA	NA	3
	1C	100	0	100	42	54	NA	NA	NA	3
	2A	99	1	100	27	69	NA	NA	NA	3
	2B	100	0	100	79	17	NA	NA	NA	3
	2C	100	0	100	96	1	NA	NA	NA	3
	2D	100	0	100	96	0	NA	NA	NA	3
	3	88	12	100	25	40	15	9	5	4
	4A	100	0	100	77	18	NA	NA	NA	4
	4B	93	7	100	57	38	NA	NA	NA	4
	4C	99	1	100	87	8	NA	NA	NA	4
	4D	99	1	100	84	11	NA	NA	NA	4
	4MU	100	0	100	95	1	NA	NA	NA	4
	5	95	5	100	51	36	5	2	1	5
	6A	100	0	100	75	19	NA	NA	NA	6
	6B	100	0	100	37	57	NA	NA	NA	6
	6C	100	0	100	79	16	NA	NA	NA	6
	6D	100	0	100	82	12	NA	NA	NA	6
Biology	1	93	7	100	29	35	21	7	NA	7
	2	94	6	100	18	23	37	15	NA	7
	3	88	12	100	28	26	36	1	NA	7
	4	87	13	100	9	43	30	7	NA	7
	5	92	8	100	25	54	9	9	NA	7
	6	94	6	100	41	20	26	9	NA	12
Literature	1	80	20	100	11	33	38	8	NA	10
	2	89	11	100	19	44	22	3	NA	12
	3	88	11	99	16	44	25	2	NA	13
	4	81	19	100	15	40	28	5	NA	12
	5	87	13	100	19	46	21	1	NA	14
	6	82	17	99	17	47	19	1	NA	15

Chapter Eighteen: Reliability

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 2015 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.

Chapter Nineteen: Validity

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
		A/B → C	0	0	0
	Biology	C → A/B	0	0	0
		A/B → C	0	0	0
	Literature	C → A/B	1	0	0
		A/B → C	0	0	0
Spring	Algebra I	C → A/B	0	0	0
		A/B → C	0	0	0
	Biology	C → A/B	0	0	0
		A/B → C	0	0	0
	Literature	C → A/B	2	1	1
		A/B → C	0	0	0
Summer	Algebra I	C → A/B	0	0	0
		A/B → C	0	0	0
	Biology	C → A/B	0	0	0
		A/B → C	0	1	0
	Literature	C → A/B	2	0	0
		A/B → C	1	0	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 0.69 to 0.87. The intercorrelations between modules in different content areas ranged from 0.45 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		Biology		Literature	
			Module 1	Module 2	Module 1	Module 2	Module 1	Module 2
Winter	Algebra I	Module 1	-					
		Module 2	**0.77**	-				
	Biology	Module 1	0.52	0.50	-			
		Module 2	0.56	0.55	**0.80**	-		
	Literature	Module 1	0.52	0.50	0.59	0.63	-	
		Module 2	0.54	0.53	0.61	0.65	**0.78**	-
Spring	Algebra I	Module 1	-					
		Module 2	**0.84**	-				
	Biology	Module 1	0.64	0.66	-			
		Module 2	0.66	0.67	**0.87**	-		
	Literature	Module 1	0.53	0.57	0.71	0.72	-	
		Module 2	0.55	0.59	0.73	0.75	**0.83**	-
Summer	Algebra I	Module 1	-					
		Module 2	**0.70**	-				
	Biology	Module 1	0.45	0.57	-			
		Module 2	0.51	0.57	**0.71**	-		
	Literature	Module 1	0.50	0.51	0.52	0.55	-	
		Module 2	0.53	0.56	0.57	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	-					
		Module 2	**0.99**	-				
	Biology	Module 1	0.66	0.64	-			
		Module 2	0.69	0.68	**0.97**	-		
	Literature	Module 1	0.65	0.63	0.73	0.76	-	
		Module 2	0.68	0.67	0.74	0.78	**0.97**	-
Spring	Algebra I	Module 1	-					
		Module 2	**1.01**	-				
	Biology	Module 1	0.75	0.77	-			
		Module 2	0.77	0.78	**0.99**	-		
	Literature	Module 1	0.64	0.68	0.82	0.84	-	
		Module 2	0.65	0.71	0.85	0.87	**0.99**	-
Summer	Algebra I	Module 1	-					
		Module 2	**0.96**	-				
	Biology	Module 1	0.62	0.78	-			
		Module 2	0.69	0.76	**0.95**	-		
	Literature	Module 1	0.71	0.73	0.74	0.77	-	
		Module 2	0.71	0.75	0.77	0.90	**0.97**	-

The within-content-area correlations were high (e.g., above 0.95), 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 0.95, while across-content-area module correlations range from 0.62 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 64.72 to 74.11 percent of the total variance, while the second factor explains 11.63 to 12.69 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	4.01	**66.90**
	2	0.78	13.02
	3	0.56	9.29
	4	0.24	3.92
	5	0.21	3.57
	6	0.20	3.30
Spring	1	4.45	**74.11**
	2	0.74	12.29
	3	0.36	6.06
	4	0.17	2.81
	5	0.16	2.59
	6	0.13	2.14
Summer	1	3.88	**64.72**
	2	0.70	11.63
	3	0.57	9.44
	4	0.34	5.69
	5	0.28	4.70
	6	0.23	3.82

Figure 19–1. Scree Plot for Algebra I, Biology, and Literature Modules

Winter

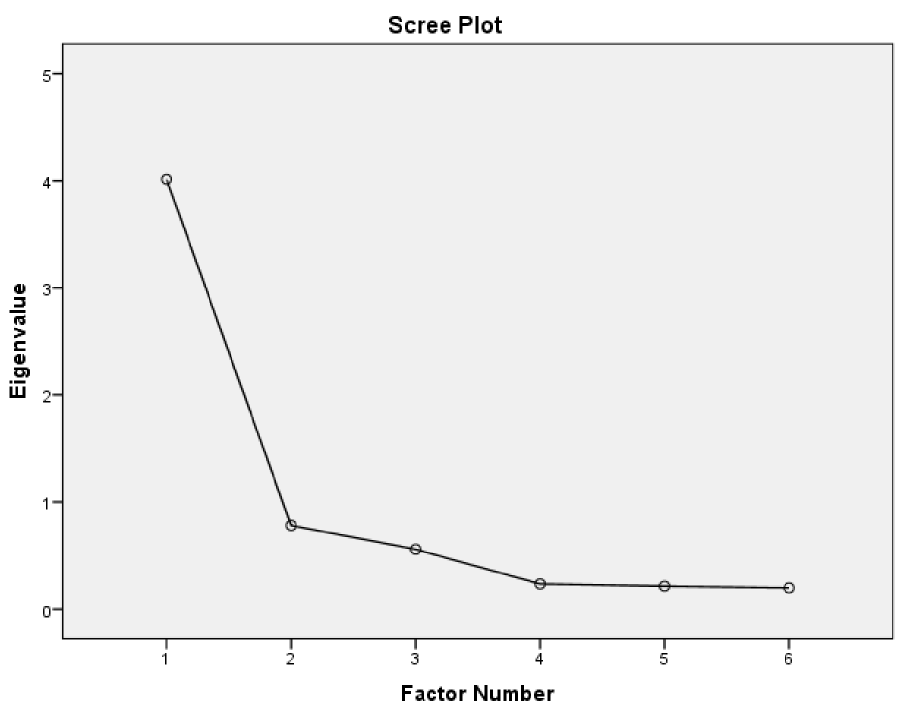
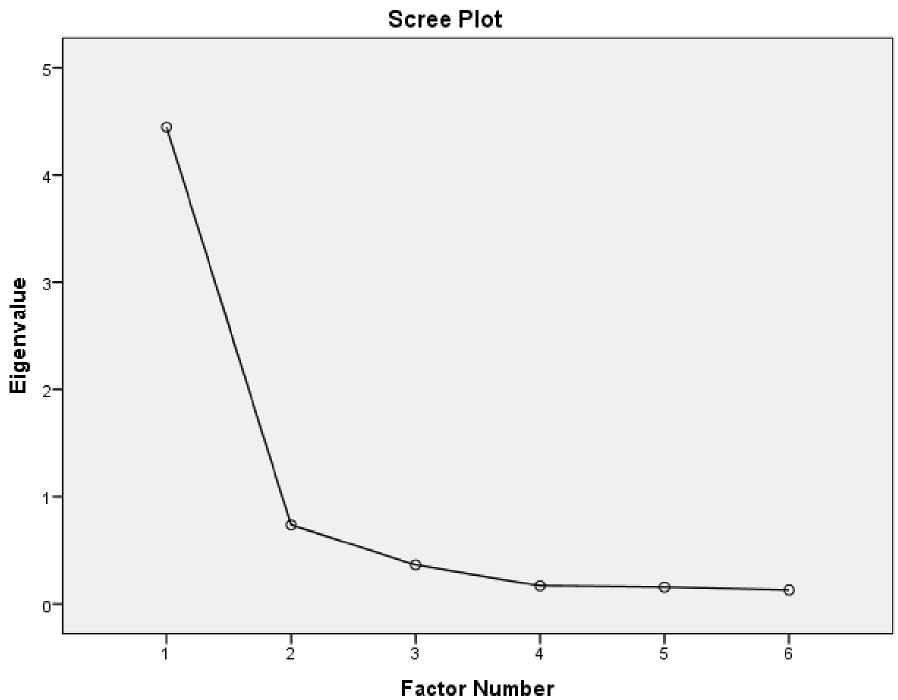
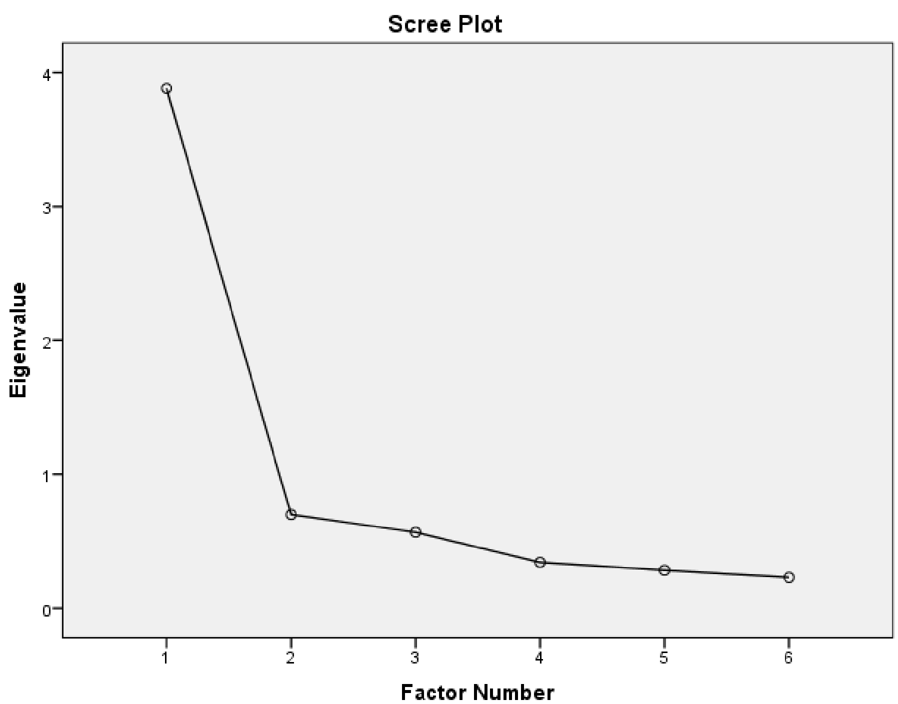


Figure 19–1 (continued). Scree Plot for Algebra I, Biology, and Literature Modules

Spring



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	**0.85**	0.02	0.02	Correlation (F1, F2) = 0.64
		Module 2	**0.86**	0.00	0.01	
	Biology	Module 1	0.00	0.02	**0.87**	Correlation (F1, F3) = 0.65
		Module 2	0.04	0.06	**0.84**	
	Literature	Module 1	-0.01	**0.86**	0.03	Correlation (F2, F3) = 0.65
		Module 2	0.04	**0.84**	0.03	
Spring	Algebra I	Module 1	-0.04	**0.90**	0.06	Correlation (F1, F2) = 0.63
		Module 2	0.07	**0.89**	-0.01	
	Biology	Module 1	0.09	0.04	**0.82**	Correlation (F1, F3) = 0.79
		Module 2	0.11	0.05	**0.81**	
	Literature	Module 1	**0.87**	0.01	0.03	Correlation (F2, F3) = 0.73
		Module 2	**0.86**	0.01	0.07	
Summer	Algebra I	Module 1	**0.83**	0.09	-0.10	Correlation (F1, F2) = 0.69
		Module 2	**0.79**	-0.05	0.15	
	Biology	Module 1	0.00	-0.04	**0.91**	Correlation (F1, F3) = 0.68
		Module 2	0.03	0.25	**0.61**	
	Literature	Module 1	0.12	**0.62**	0.06	Correlation (F2, F3) = 0.74
		Module 2	-0.02	**0.94**	0.01	

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	0.63 (0.71)	-
	Literature	0.61 (0.69)	0.70 (0.78)
Spring	Biology	0.74 (0.80)	-
	Literature	0.64 (0.70)	0.79 (0.86)
Summer	Biology	0.63 (0.74)	-
	Literature	0.63 (0.75)	0.69 (0.82)

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 0.61 to 0.79.

External evidence for the Keystone Exams is examined by using students’ scores on the 2015 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		Grade 8	
	N	Correlation	*N*	Correlation
Algebra I/Mathematics	8,274	0.81	38,092	0.84
Biology/Science	NA	NA	222	0.92
Literature/Reading	NA	NA	114	0.87

The correlations within the same content area ranged from 0.81 to 0.92. 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		Biology		Literature	
	N	%	*N*	%	*N*	%
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

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.

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.

References

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.

References

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.

References

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.

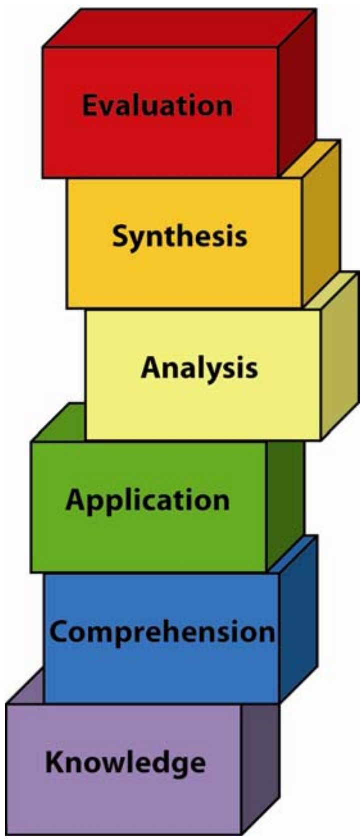
Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213.

Zwick, R., & Erickan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*, 55–66.

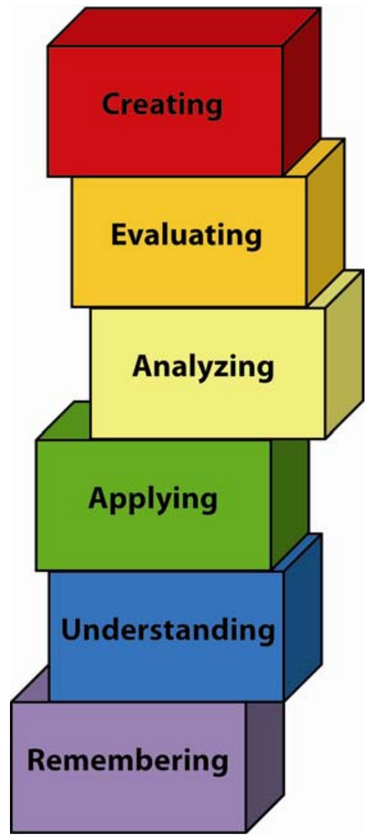
APPENDIX A: UNDERSTANDING DEPTH OF KNOWLEDGE AND COGNITIVE COMPLEXITY

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

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

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 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. 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

BIOLOGY MODULE 1

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

LITERATURE MODULE 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.

Table C–1 (continued). Item and Test Development Process for the Keystone Exams

Step	Description
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.
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.

Table C–1 (continued). Item and Test Development Process for the Keystone Exams

Step	Description
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.	

Table C–1 (continued). Item and Test Development Process for the Keystone Exams

Step	Description
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.
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.

Table C–1 (continued). Item and Test Development Process for the Keystone Exams

Step	Description
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).

1. [illegible]

A. [illegible]

B. [illegible]

C. [illegible]

D. [illegible]

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	

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

Data Recognition Corporation [illegible] 08/19/13 Page 4

APPENDIX E: ITEM RATING SHEET AND CRITERIA GUIDELINES

ITEM RATING SHEET

Reviewer Signature: ______________________

Item Rating Sheet

Content Area: ____________ **Module:** ____________

Summer 2014		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

ITEM REVIEW CRITERIA GUIDELINES

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.
- **Do NOT remove any items from the item binder at any time.**
- You must sign your item rating sheet.

APPENDIX F: KEYSTONE EXAMS TALLY SHEETS

ALGEBRA I-WINTER 2014/2015

Keystone Exam **Algebra I**

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
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	1	0	4
	1	1	1	Compare and/or order any real numbers.	0	0	0	0
	1	1	2	Simplify square roots.	1	0	1	0
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1	0	1	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	1	0
	1	4	1	Use estimation to solve problems.	0	0	0	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1	0	1	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1	0	1	0
	1	5	3	Simplify/reduce a rational algebraic expression.	1	0	1	0
	Total For Assessment Anchor A1.1.1				6	1	6	4
	2			**Linear Equations**	0	1	0	4
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	1	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1	0	1	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	2	0	2	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	Total For Assessment Anchor A1.1.2				6	1	6	4

	3			**Linear Inequalities**	0	1	0	4
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	1	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1	0	1	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	2	0	2	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1	0	1	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	Total For Assessment Anchor A1.1.3				6	1	6	4
Total For Reporting Category A1.1					18	3	18	12

Keystone Exam **Algebra I**

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
A1.2: Linear Functions and Data Organizations	1			**Functions**	0	1	0	4
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1	0	1	0
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1	0	1	0
	1	1	3	Identify the domain or range of a relation.	1	0	1	0
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	2	0	2	0
	1	2	2	Translate from one representation of a linear function to another.	1	0	1	0
	Total For Assessment Anchor A1.2.1				6	1	6	4
	2			**Coordinate Geometry**	0	2	0	4
	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	1	0
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1	0	1	0
	2	1	3	Write or identify a linear equation when given...	1	0	1	0
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	2	0	2	0
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1	0	1	0
	Total For Assessment Anchor A1.2.2				6	2	6	4
	3			**Data Analysis**	0	0	0	4

	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1	0	1	0
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	2	0	2	0
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1	0	1	0
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	1	0	1	0
	3	3	1	Find probabilities for compound events.	1	0	1	0
	Total For Assessment Anchor A1.2.3				6	0	6	4
Total For Reporting Category A1.2					18	3	18	12

ALGEBRA I-SPRING 2015

Keystone Exam **Algebra I**

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
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	1	0	4
	1	1	1	Compare and/or order any real numbers.	1	0	1	0
	1	1	2	Simplify square roots.	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	1	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	1	0
	1	4	1	Use estimation to solve problems.	0	0	0	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1	0	1	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1	0	1	0
	1	5	3	Simplify/reduce a rational algebraic expression.	1	0	1	0
	Total For Assessment Anchor A1.1.1				6	1	6	4
	2			**Linear Equations**	0	1	0	4
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	1	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1	0	1	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	2	0	2	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	Total For Assessment Anchor A1.1.2				6	1	6	4

	3			**Linear Inequalities**	0	1	0	4
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	1	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	2	0	2	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1	0	1	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	Total For Assessment Anchor A1.1.3				6	1	6	4
Total For Reporting Category A1.1					18	3	18	12

Keystone Exam **Algebra I**

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
A1.2: Linear Functions and Data Organizations	1			**Functions**	0	1	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	2	0
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1	0	1	0
	1	1	3	Identify the domain or range of a relation.	1	0	1	0
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	1	0	1	0
	1	2	2	Translate from one representation of a linear function to another.	1	0	1	0
	Total For Assessment Anchor A1.2.1				6	1	6	4
	2			**Coordinate Geometry**	0	1	0	4
	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	1	0
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1	0	1	0
	2	1	3	Write or identify a linear equation when given...	1	0	1	0
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	2	0	2	0
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1	0	1	0
	Total For Assessment Anchor A1.2.2				6	1	6	4

	3			**Data Analysis**	0	1	0	4
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	2	0	2	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	1	0
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1	0	1	0
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	2	0	2	0
	3	3	1	Find probabilities for compound events.	0	0	0	0
	Total For Assessment Anchor A1.2.3				6	1	6	4
Total For Reporting Category A1.2					18	3	18	12

ALGEBRA I-SUMMER 2015

Keystone Exam **Algebra I**

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
A1.1: Operations and Linear Equations & Inequalities	1			**Operations with Real Numbers and Expressions**	0	1	0	4
	1	1	1	Compare and/or order any real numbers.	0	0	0	0
	1	1	2	Simplify square roots.	1	0	1	0
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1	0	1	0
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1	0	1	0
	1	4	1	Use estimation to solve problems.	1	0	1	0
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1	0	1	0
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1	0	1	0
	1	5	3	Simplify/reduce a rational algebraic expression.	0	0	0	0
	Total For Assessment Anchor A1.1.1				6	1	6	4
	2			**Linear Equations**	0	1	0	4
	2	1	1	Write, solve, and/or apply a linear equation.	1	0	1	0
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	2	0	2	0
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1	0	1	0
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	Total For Assessment Anchor A1.1.2				6	1	6	4

	3			**Linear Inequalities**	0	1	0	4
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1	0	1	0
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1	0	1	0
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1	0	1	0
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1	0	1	0
	3	2	2	Interpret solutions to problems in the context of the problem situation.	2	0	2	0
	Total For Assessment Anchor A1.1.3				6	1	6	4
Total For Reporting Category A1.1					18	3	18	12

Keystone Exam **Algebra I**

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
A1.2: Linear Functions and Data Organizations	1			**Functions**	0	1	0	4
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1	0	1	0
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1	0	1	0
	1	1	3	Identify the domain or range of a relation.	1	0	1	0
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	2	0	2	0
	1	2	2	Translate from one representation of a linear function to another.	1	0	1	0
	Total For Assessment Anchor A1.2.1				6	1	6	4
	2			**Coordinate Geometry**	0	1	0	4
	2	1	1	Identify, describe, and/or use constant rates of change.	1	0	1	0
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1	0	1	0
	2	1	3	Write or identify a linear equation when given...	1	0	1	0
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1	0	1	0
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	2	0	2	0
	Total For Assessment Anchor A1.2.2				6	1	6	4

	3			**Data Analysis**	0	1	0	4
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1	0	1	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	1	0
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	2	0	2	0
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	2	0	2	0
	3	3	1	Find probabilities for compound events.	0	0	0	0
	Total For Assessment Anchor A1.2.3				6	1	6	4
Total For Reporting Category A1.2					18	3	18	12

BIOLOGY-WINTER 2014/2015

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.	2	0	2	0
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	1	1	1	3
	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.	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	1	1	3
	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				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.	2	0	2	0
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	2	0	2	0
	3	2	2	Describe the role of ATP in biochemical reactions.	1	1	1	3
	Total For Assessment Anchor BIO.A.3				5	1	5	3

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Number of Core Items MC	Number of Core Items CR	Core Points MC	Core Points CR
	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.	1	0	1	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.	2	0	2	0
	4	2	1	Explain how organisms maintain homeostasis.	2	0	2	0
	Total For Assessment Anchor BIO.A.4				7	0	7	0
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		Points: Core Points	
					MC	CR	MC	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.	1	0	1	0
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	1	0	1	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	1	1	3
	Total For Assessment Anchor BIO.B.1				4	1	4	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.	1	0	1	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1	0	1	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.	1	1	1	3
	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.	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	0	1	0
	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	0	1	0
	4	2	2	Describe biotic interactions in an ecosystem.	2	0	2	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.	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				9	0	9	0
Total For Assessment Anchor BIO.B					24	3	24	9

BIOLOGY-SPRING 2015

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.	0	1	0	3
	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				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.	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.	2	0	2	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.	2	0	2	0
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	2	0	2	0
	3	2	2	Describe the role of ATP in biochemical reactions.	2	0	2	0
	Total For Assessment Anchor BIO.A.3				6	0	6	0

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Number of Core Items MC	Number of Core Items CR	Core Points MC	Core Points CR
	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	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				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.	1	0	1	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				6	0	6	0
	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.	1	0	1	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1	0	1	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.	1	0	1	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	1	1	3
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				5	1	5	3
4			**Ecology**	0	0	0	0
4	1	1	Describe the levels of ecological organization.	1	1	1	3
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.	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.	1	0	1	0
Total For Assessment Anchor BIO.B.4				7	1	7	3
Total For Assessment Anchor BIO.B				24	3	24	9

BIOLOGY-SUMMER 2015

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.	1	0	1	0
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2	1	2	3
	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				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.	0	1	0	3
	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.	2	0	2	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.	2	0	2	0
	Total For Assessment Anchor BIO.A.2				8	1	8	3
	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.	2	0	2	0
	3	2	2	Describe the role of ATP in biochemical reactions.	2	0	2	0
	Total For Assessment Anchor BIO.A.3				6	0	6	0

	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	0	2	0
	4	1	3	Describe how membrane-bound cellular organelles.	1	0	1	0
	4	2	1	Explain how organisms maintain homeostasis.	0	1	0	3
	Total For Assessment Anchor BIO.A.4				5	1	5	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.	1	0	1	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				6	0	6	0
	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.	1	0	1	0
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1	0	1	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.	1	1	1	3
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				4	2	4	6
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	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.	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				8	0	8	0
Total For Assessment Anchor BIO.B				24	3	24	9

LITERATURE-WINTER 2014/2015

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.	1	0	1	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.	0	1	0	3
	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.	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 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	2	2	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:	2	0	2	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:	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	0	0	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.	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.	4	0	4	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				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
	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.	1	0	1	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.	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.	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.	0	1	0	3
	Total For Assessment Anchor L.N.1				4	1	4	3

L.N: Nonfiction	2			**Data Analysis**	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.	2	0	2	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	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:	1	0	1	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:	2	0	2	0
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	0	1	0	3
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	1	1	1	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.	1	0	1	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 2015

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.	1	0	1	0
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	0	1	0	3
	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.	2	0	2	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.	3	0	3	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.	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:	0	1	0	3
	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:	0	0	0	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:	0	0	0	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.	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.	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				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
	1			**Reading for Meaning—Nonfiction**	0	0	0	0
	1	1	1	Identify and/or analyze the author's intended purpose of a text.	1	0	1	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.	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.	2	0	2	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.	2	0	2	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.	0	0	0	0
	Total For Assessment Anchor L.N.1				7	0	7	0

L.N: Nonfiction	2			**Data Analysis**	0	0	0	0
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1	1	1	3
	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.	1	0	1	0
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:	1	1	1	3
	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:	1	0	1	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				10	3	10	9
Total For Reporting Category L.N					17	3	17	9

LITERATURE-SUMMER 2015

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.	1	0	1	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.	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.	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				5	1	5	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.	0	1	0	3
	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.	0	0	0	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.	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:	0	1	0	3
	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:	2	0	2	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.	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.	3	0	3	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				12	2	12	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
	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.	2	0	2	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.	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.	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.	1	0	1	0
	Total For Assessment Anchor L.N.1				6	0	6	0

L.N: Nonfiction	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.	0	0	0	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:	1	0	1	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:	1	0	1	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:	1	1	1	3
	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.	1	0	1	0
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.	1	0	1	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.	0	0	0	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.	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	1	0	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–1. Winter 2014/2015, Spring 2015, and Summer 2015 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
Algebra I	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–2. Winter 2014/2015, Spring 2015, and Summer 2015 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
Biology	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–3. Winter 2014/2015, Spring 2014, and Summer 2015 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
Literature	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	120096	42	60	0	60	31.6	32	12.7
1	5016	42	60	1	60	31.1	31	12.8
2	4964	42	60	3	60	31.5	31	12.6
3	4982	42	60	3	60	31.8	32	12.8
4	5046	42	60	3	60	31.3	31	12.9
5	5021	42	60	4	60	31.4	32	12.8
6	5021	42	60	4	60	31.6	32	12.7
7	5008	42	60	1	60	31.2	31	12.6
8	4876	42	60	2	60	31.8	32	12.8
9	5037	42	60	3	60	31.4	32	12.6
10	5027	42	60	3	60	31.5	31	12.7
11	4991	42	60	2	60	31.7	32	12.7
12	5003	42	60	2	60	31.8	32	12.5
13	5042	42	60	3	60	31.7	32	12.7
14	5026	42	60	3	60	31.7	32	12.8
15	5009	42	60	1	60	31.7	32	12.6
16	4993	42	60	1	60	31.8	32	12.7
17	5013	42	60	2	60	31.7	32	12.7
18	5023	42	60	3	60	31.5	31	12.8
19	5003	42	60	4	60	31.6	32	12.7
20	4978	42	60	2	60	32.0	33	12.8
21	5082	42	60	3	60	31.5	32	12.6
22	5000	42	60	4	60	32.0	32	12.5
23	5011	42	60	2	60	31.8	32	12.6
24	4924	42	60	0	60	31.8	32	12.7

BIOLOGY: SPRING

Table H–3. Biology Mean Raw Scores by Form Table

Form	*N*	L	Pts	Min	Max	Mean	Med	*SD*
All	114733	54	66	0	66	38.1	39	14.0
1	4828	54	66	2	66	38.1	39	14.1
2	4787	54	66	3	65	38.1	39	14.1
3	4730	54	66	5	65	38.1	40	13.9
4	4752	54	66	5	66	38.3	39	14.0
5	4756	54	66	5	65	38.1	40	13.9
6	4837	54	66	5	66	38.0	39	14.0
7	4788	54	66	3	66	38.2	40	14.1
8	4769	54	66	3	66	38.3	40	13.9
9	4730	54	66	5	66	38.1	39	13.9
10	4742	54	66	5	66	38.3	40	14.1
11	4797	54	66	2	66	38.2	40	14.1
12	4751	54	66	5	66	37.8	39	14.1
13	4804	54	66	5	66	38.0	39	14.1
14	4806	54	66	5	66	37.9	39	13.9
15	4776	54	66	5	66	37.9	39	14.0
16	4785	54	66	4	66	38.0	39	14.2
17	4773	54	66	5	66	38.6	40	13.9
18	4754	54	66	5	65	38.1	39	14.2
19	4823	54	66	0	66	38.3	40	14.1
20	4775	54	66	5	66	38.1	40	14.0
21	4826	54	66	4	66	38.0	40	14.1
22	4776	54	66	5	66	38.2	39	14.1
23	4809	54	66	6	66	38.4	40	14.1
24	4759	54	66	5	66	38.0	39	14.0

LITERATURE: SPRING

Table H–4. Literature Mean Raw Scores by Form Table

Form	*N*	L	Pts	Min	Max	Mean	Med	*SD*
All	114329	40	52	1	52	33.4	35	9.7
1	4772	40	52	3	52	33.3	35	9.9
2	4781	40	52	4	52	33.6	36	9.7
3	4801	40	52	3	52	33.5	36	9.8
4	4772	40	52	1	52	33.3	35	9.9
5	4755	40	52	4	52	33.3	35	9.8
6	4748	40	52	1	51	33.2	35	9.8
7	4799	40	52	4	52	33.4	35	9.7
8	4779	40	52	3	52	33.5	36	9.8
9	4754	40	52	3	52	33.4	35	9.6
10	4765	40	52	3	51	33.5	36	9.8
11	4802	40	52	2	51	33.3	35	9.7
12	4762	40	52	2	52	33.1	35	9.6
13	4782	40	52	3	52	33.5	36	9.7
14	4810	40	52	3	52	33.4	35	9.8
15	4773	40	52	3	52	33.7	36	9.7
16	4759	40	52	3	52	33.8	36	9.8
17	4757	40	52	1	52	33.2	35	9.7
18	4719	40	52	3	51	33.4	35	9.5
19	4747	40	52	1	52	33.2	35	9.7
20	4736	40	52	4	52	33.1	35	9.7
21	4735	40	52	5	52	33.4	35	9.6
22	4730	40	52	3	51	33.4	35	9.8
23	4735	40	52	3	52	33.4	35	9.7
24	4756	40	52	3	51	33.4	35	9.5

APPENDIX I: DEMOGRAPHIC AND ACCOMMODTION DATA

WINTER 2014/2015

Students Assessed on the Winter 2014/2015 Keystone: Algebra I

	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	42	3	13	465	10,294	29,218	26,751	824	67,610
Total number of CBT processed	0	1	2	217	1,976	3,902	4,001	204	10,303
Total number of tests processed	42	4	15	682	12,270	33,120	30,752	1,028	77,913
Total number of tests processed with a score	38 90.5	4 100.0	15 100.0	677 99.3	11,746 95.7	31,635 95.5	28,934 94.1	976 94.9	74,025 95.0
Total number of tests processed without a score	4 9.5	0 0.0	0 0.0	5 0.7	524 4.3	1,485 4.5	1,818 5.9	52 5.1	3,888 5.0

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

Students Assessed on the Winter 2014/2015 Keystone: Biology

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	19	58	2,158	18,601	25,094	664	46,594
Total number of CBT processed	0	0	327	3,602	4,870	232	9,031
Total number of tests processed	19	58	2,485	22,203	29,964	896	55,625
Total number of tests processed with a score	17 89.5	56 96.6	2,302 92.6	21,395 96.4	27,937 93.2	836 93.3	52,543 94.5
Total number of tests processed without a score	2 10.5	2 3.4	183 7.4	808 3.6	2,027 6.8	60 6.7	3,082 5.5

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

Students Assessed on the Winter 2014/2015 Keystone: Literature

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	15	52	780	12,139	26,506	552	40,044
Total number of CBT processed	0	0	82	2,641	5,227	134	8,084
Total number of tests processed	15	52	862	14,780	31,733	686	48,128
Total number of tests processed with a score	13 86.7	51 98.1	754 87.5	14,181 95.9	29,812 93.9	653 95.2	45,464 94.5
Total number of tests processed without a score	2 13.3	1 1.9	108 12.5	599 4.1	1,921 6.1	33 4.8	2,664 5.5

*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 2014/2015 Keystone: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	1	0	0	1	185	366	549	7	1,109
	25.0	0.0	0.0	20.0	35.3	24.6	30.2	13.5	28.5
Absent without make-up	0	0	0	0	58	219	208	5	490
	0.0	0.0	0.0	0.0	11.1	14.7	11.4	9.6	12.6
Non-attempt	2	0	0	1	210	669	665	14	1,561
	50.0	0.0	0.0	20.0	40.1	45.1	36.6	26.9	40.1
Medical emergency	0	0	0	1	12	49	57	2	121
	0.0	0.0	0.0	20.0	2.3	3.3	3.1	3.8	3.1
Parental request	0	0	0	2	5	11	81	1	100
	0.0	0.0	0.0	40.0	1.0	0.7	4.5	1.9	2.6
Other reasons	1	0	0	0	54	171	258	23	507
	25.0	0.0	0.0	0.0	10.3	11.5	14.2	44.2	13.0
Total not assessed	4	0	0	5	524	1,485	1,818	52	3,888

*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 2014/2015 Keystone: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	0	1	88	196	642	1	928
	0.0	50.0	48.1	24.3	31.7	1.7	30.1
Absent without make-up	1	0	20	138	245	1	405
	50.0	0.0	10.9	17.1	12.1	1.7	13.1
Non-attempt	0	0	50	343	660	14	1,067
	0.0	0.0	27.3	42.5	32.6	23.3	34.6
Medical emergency	0	1	4	30	68	1	104
	0.0	50.0	2.2	3.7	3.4	1.7	3.4
Parental request	0	0	0	7	115	1	123
	0.0	0.0	0.0	0.9	5.7	1.7	4.0
Other reasons	1	0	21	94	297	42	455
	50.0	0.0	11.5	11.6	14.7	70.0	14.8
Total not assessed	2	2	183	808	2,027	60	3,082

*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 2014/2015 Keystone: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	0	0	64	194	612	6	876
	0.0	0.0	59.3	32.4	31.9	18.2	32.9
Absent without make-up	0	0	11	92	240	2	345
	0.0	0.0	10.2	15.4	12.5	6.1	13.0
Non-attempt	1	0	15	227	670	7	920
	50.0	0.0	13.9	37.9	34.9	21.2	34.5
ELL in first year in U.S. schools	0	0	0	0	9	0	9
	0.0	0.0	0.0	0.0	0.5	0.0	0.3
Medical emergency	0	1	3	15	68	0	87
	0.0	100.0	2.8	2.5	3.5	0.0	3.3
Parental request	0	0	0	4	69	1	74
	0.0	0.0	0.0	0.7	3.6	3.0	2.8
Other reasons	1	0	15	67	253	17	353
	50.0	0.0	13.9	11.2	13.2	51.5	13.3
Total not assessed	2	1	108	599	1,921	33	2,664

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

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Algebra I

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender									
Female	11 28.9	0 0.0	8 53.3	317 46.8	5,808 49.4	15,277 48.3	13,814 47.7	410 42.0	35,645 48.2
Male	10 26.3	4 100.0	7 46.7	360 53.2	5,932 50.5	16,331 51.6	15,102 52.2	563 57.7	38,309 51.8
Race/Ethnicity									
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	1 6.7	1 0.1	18 0.2	52 0.2	45 0.2	0 0.0	117 0.2
Asian (not Hispanic)	1 2.6	2 50.0	10 66.7	30 4.4	370 3.2	756 2.4	695 2.4	14 1.4	1,878 2.5
Black or African American (not Hispanic)	9 23.7	0 0.0	0 0.0	80 11.8	1,517 12.9	7,072 22.4	6,456 22.3	212 21.7	15,346 20.7
Hispanic (any race)	3 7.9	0 0.0	0 0.0	28 4.1	1,113 9.5	3,844 12.2	3,294 11.4	102 10.5	8,384 11.3
Multi-Racial (not Hispanic)	0 0.0	0 0.0	0 0.0	5 0.7	258 2.2	619 2.0	547 1.9	21 2.2	1,450 2.0
White (not Hispanic)	5 13.2	2 50.0	4 26.7	533 78.7	8,451 71.9	19,239 60.8	17,860 61.7	621 63.6	46,715 63.1
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	0 0.0	0 0.0	11 0.1	20 0.1	17 0.1	3 0.3	51 0.1
Educational Category and Other Demographic Groups									
IEP (not gifted)	5 13.2	0 0.0	0 0.0	32 4.7	894 7.6	5,921 18.7	6,953 24.0	192 19.7	13,997 18.9
Student exited IEP in last 2 years	0 0.0	1 25.0	0 0.0	6 0.9	161 1.4	422 1.3	334 1.2	11 1.1	935 1.3
Title I	2 5.3	0 0.0	0 0.0	128 18.9	2,416 20.6	9,237 29.2	7,872 27.2	314 32.2	19,969 27.0
Title III served	1 2.6	0 0.0	0 0.0	1 0.1	218 1.9	1,308 4.1	1,228 4.2	18 1.8	2,774 3.7
Title III not served	0 0.0	0 0.0	0 0.0	0 0.0	3 0.0	15 0.0	15 0.1	0 0.0	33 0.0
Migrant student	0 0.0	0 0.0	0 0.0	1 0.1	0 0.0	16 0.1	14 0.0	0 0.0	31 0.0
ELL (enrolled after 1/22/14)	0 0.0	0 0.0	0 0.0	0 0.0	17 0.1	112 0.4	132 0.5	3 0.3	264 0.4
ELL (enrolled on or before 1/22/14)	1 2.6	0 0.0	0 0.0	1 0.1	219 1.9	1,263 4.0	1,161 4.0	19 1.9	2,664 3.6

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Algebra I (continued)

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	0 0.0	0 0.0	18 0.2	134 0.4	142 0.5	2 0.2	296 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	0 0.0	0 0.0	25 0.2	103 0.3	147 0.5	1 0.1	276 0.4
Former ELL no longer monitored	1 2.6	0 0.0	1 6.7	5 0.7	229 1.9	639 2.0	469 1.6	4 0.4	1,348 1.8
Foreign exchange student	0 0.0	0 0.0	0 0.0	0 0.0	2 0.0	6 0.0	17 0.1	3 0.3	28 0.0
Economically disadvantaged	4 10.5	0 0.0	0 0.0	183 27.0	4,741 40.4	16,661 52.7	15,328 53.0	544 55.7	37,461 50.6
Historically Underperforming Subgroup	8 21.1	0 0.0	0 0.0	201 29.7	5,155 43.9	19,466 61.5	18,517 64.0	623 63.8	43,970 59.4
Enrollment in school of residence after 10/1/14	7 18.4	0 0.0	0 0.0	3 0.4	85 0.7	299 0.9	378 1.3	11 1.1	783 1.1
Enrollment in district of residence after 10/1/14	6 15.8	0 0.0	0 0.0	1 0.1	50 0.4	205 0.6	263 0.9	11 1.1	536 0.7
Enrollment as PA resident after 10/1/14	2 5.3	0 0.0	0 0.0	1 0.1	21 0.2	91 0.3	102 0.4	2 0.2	219 0.3
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	3 7.9	0 0.0	3 20.0	25 3.7	3,423 29.1	8,068 25.5	6,735 23.3	112 11.5	18,369 24.8
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	3 7.9	0 0.0	3 20.0	21 3.1	1,429 12.2	2,814 8.9	2,827 9.8	94 9.6	7,191 9.7
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	0 0.0	0 0.0	3 0.0	9 0.0	16 0.1	7 0.7	35 0.0
Number of assessed students	38	4	15	677	11,746	31,635	28,934	976	74,025

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

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	5 29.4	20 35.7	1,088 47.3	10,701 50.0	13,752 49.2	339 40.6	25,905 49.3
Male	5 29.4	36 64.3	1,214 52.7	10,690 50.0	14,171 50.7	497 59.4	26,613 50.6
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	3 0.1	40 0.2	39 0.1	0 0.0	82 0.2
Asian (not Hispanic)	0 0.0	0 0.0	152 6.6	653 3.1	737 2.6	15 1.8	1,557 3.0
Black or African American (not Hispanic)	3 17.6	52 92.9	332 14.4	3,864 18.1	6,201 22.2	203 24.3	10,655 20.3
Hispanic (any race)	1 5.9	0 0.0	188 8.2	1,868 8.7	3,137 11.2	61 7.3	5,255 10.0
Multi-Racial (not Hispanic)	0 0.0	0 0.0	57 2.5	385 1.8	543 1.9	11 1.3	996 1.9
White (not Hispanic)	5 29.4	4 7.1	1,568 68.1	14,567 68.1	17,246 61.7	544 65.1	33,934 64.6
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	2 0.1	13 0.1	18 0.1	2 0.2	35 0.1
Educational Category and Other Demographic Groups							
IEP (not gifted)	2 11.8	8 14.3	216 9.4	3,726 17.4	6,109 21.9	161 19.3	10,222 19.5
Student exited IEP in last 2 years	0 0.0	0 0.0	30 1.3	252 1.2	289 1.0	7 0.8	578 1.1
Title I	1 5.9	52 92.9	509 22.1	4,461 20.9	7,879 28.2	283 33.9	13,185 25.1
Title III served	0 0.0	1 1.8	35 1.5	602 2.8	1,064 3.8	7 0.8	1,709 3.3
Title III not served	0 0.0	0 0.0	0 0.0	6 0.0	8 0.0	1 0.1	15 0.0
Migrant student	0 0.0	0 0.0	0 0.0	4 0.0	11 0.0	0 0.0	15 0.0
ELL (enrolled after 1/22/14)	0 0.0	0 0.0	1 0.0	41 0.2	110 0.4	2 0.2	154 0.3
ELL (enrolled on or before 1/22/14)	0 0.0	1 1.8	36 1.6	592 2.8	1,003 3.6	8 1.0	1,640 3.1

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Biology (continued)

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	5 0.2	70 0.3	130 0.5	2 0.2	207 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	6 0.3	60 0.3	120 0.4	1 0.1	187 0.4
Former ELL no longer monitored	0 0.0	0 0.0	67 2.9	447 2.1	602 2.2	6 0.7	1,122 2.1
Foreign exchange student	0 0.0	0 0.0	0 0.0	0 0.0	14 0.1	3 0.4	17 0.0
Economically disadvantaged	2 11.8	37 66.1	849 36.9	9,492 44.4	14,744 52.8	468 56.0	25,592 48.7
Historically Underperforming Subgroup	3 17.6	40 71.4	957 41.6	11,358 53.1	17,572 62.9	524 62.7	30,454 58.0
Enrollment in school of residence after 10/1/14	5 29.4	0 0.0	19 0.8	144 0.7	323 1.2	12 1.4	503 1.0
Enrollment in district of residence after 10/1/14	3 17.6	0 0.0	15 0.7	107 0.5	245 0.9	12 1.4	382 0.7
Enrollment as PA resident after 10/1/14	1 5.9	0 0.0	7 0.3	43 0.2	92 0.3	2 0.2	145 0.3
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	0 0.0	2 3.6	634 27.5	4,848 22.7	6,547 23.4	93 11.1	12,124 23.1
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	1 5.9	2 3.6	223 9.7	1,563 7.3	2,442 8.7	79 9.4	4,310 8.2
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	0 0.0	12 0.1	18 0.1	4 0.5	34 0.1
Number of assessed students	17	56	2,302	21,395	27,937	836	52,543

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

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	2 15.4	16 31.4	327 43.4	6,719 47.4	12,497 41.9	225 34.5	19,786 43.5
Male	3 23.1	35 68.6	426 56.5	7,452 52.5	17,295 58.0	426 65.2	25,637 56.4
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	0 0.0	15 0.1	42 0.1	0 0.0	57 0.1
Asian (not Hispanic)	0 0.0	0 0.0	17 2.3	369 2.6	753 2.5	11 1.7	1,150 2.5
Black or African American (not Hispanic)	3 23.1	51 100.0	182 24.1	2,332 16.4	5,950 20.0	152 23.3	8,670 19.1
Hispanic (any race)	0 0.0	0 0.0	127 16.8	1,688 11.9	3,469 11.6	51 7.8	5,335 11.7
Multi-Racial (not Hispanic)	0 0.0	0 0.0	33 4.4	205 1.4	548 1.8	12 1.8	798 1.8
White (not Hispanic)	1 7.7	0 0.0	393 52.1	9,554 67.4	19,006 63.8	424 64.9	29,378 64.6
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	1 0.1	7 0.0	21 0.1	1 0.2	30 0.1
Educational Category and Other Demographic Groups							
IEP (not gifted)	2 15.4	8 15.7	134 17.8	2,244 15.8	7,742 26.0	173 26.5	10,303 22.7
Student exited IEP in last 2 years	0 0.0	0 0.0	6 0.8	147 1.0	354 1.2	5 0.8	512 1.1
Title I	1 7.7	51 100.0	196 26.0	2,923 20.6	7,261 24.4	200 30.6	10,632 23.4
Title III served	0 0.0	1 2.0	34 4.5	292 2.1	1,197 4.0	11 1.7	1,535 3.4
Title III not served	0 0.0	0 0.0	0 0.0	4 0.0	11 0.0	1 0.2	16 0.0
Migrant student	0 0.0	0 0.0	0 0.0	9 0.1	13 0.0	0 0.0	22 0.0
ELL (enrolled after 1/22/14)	0 0.0	0 0.0	4 0.5	14 0.1	57 0.2	1 0.2	76 0.2
ELL (enrolled on or before 1/22/14)	0 0.0	1 2.0	34 4.5	302 2.1	1,196 4.0	13 2.0	1,546 3.4

Demographic Characteristics of Students taking the Winter 2014/2015 Keystone: Literature (continued)

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	0 0.0	45 0.3	143 0.5	2 0.3	190 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	2 0.3	40 0.3	183 0.6	4 0.6	229 0.5
Former ELL no longer monitored	0 0.0	0 0.0	12 1.6	332 2.3	589 2.0	1 0.2	934 2.1
Foreign exchange student	0 0.0	0 0.0	0 0.0	4 0.0	15 0.1	4 0.6	23 0.1
Economically disadvantaged	1 7.7	36 70.6	444 58.9	5,901 41.6	15,671 52.6	375 57.4	22,428 49.3
Historically Underperforming Subgroup	2 15.4	39 76.5	485 64.3	6,908 48.7	19,201 64.4	436 66.8	27,071 59.5
Enrollment in school of residence after 10/1/14	2 15.4	0 0.0	14 1.9	148 1.0	342 1.1	9 1.4	515 1.1
Enrollment in district of residence after 10/1/14	1 7.7	0 0.0	4 0.5	117 0.8	262 0.9	9 1.4	393 0.9
Enrollment as PA resident after 10/1/14	0 0.0	0 0.0	2 0.3	47 0.3	96 0.3	1 0.2	146 0.3
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	0 0.0	0 0.0	197 26.1	2,602 18.3	6,668 22.4	82 12.6	9,549 21.0
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	1 7.7	0 0.0	128 17.0	959 6.8	2,790 9.4	70 10.7	3,948 8.7
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	2 0.3	15 0.1	18 0.1	3 0.5	38 0.1
Number of assessed students	13	51	754	14,181	29,812	653	45,464

*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 2014/2015 Keystone: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	5	N/A	5
	0.0	N/A	0.0
Large print format	26	N/A	26
	0.0	N/A	0.0
Computer Assistive Technology	0	N/A	0
	0.0	N/A	0.0
Some test items/questions read aloud	482	97	579
	0.8	1.0	0.8
All test items/questions read aloud	275	61	336
	0.4	0.6	0.5
Test items/questions signed	3	0	3
	0.0	0.0	0.0
Test items/questions interpreted for ELL student	4	0	4
	0.0	0.0	0.0
Amplification device	7	1	8
	0.0	0.0	0.0
Magnification device	1	1	2
	0.0	0.0	0.0
Color overlay	1	N/A	1
	0.0	N/A	0.0
Other (per Accommodations Guidelines)	2	12	14
	0.0	0.1	0.0
Spanish version	467	N/A	467
	0.7	N/A	0.6
Online Accommodations Received			
Audio	N/A	308	308
	N/A	3.1	0.4
Color Chooser	N/A	2	2
	N/A	0.0	0.0
Contrasting Text Chooser	N/A	1	1
	N/A	0.0	0.0
Number of assessed students	63,978	10,047	74,025

Incidence of Presentation Accommodations Received on the Winter 2014/2015 Keystone: Biology

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	4	N/A	4
	0.0	N/A	0.0
Large print format	22	N/A	22
	0.1	N/A	0.0
Computer Assistive Technology	1	N/A	1
	0.0	N/A	0.0
Some test items/questions read aloud	224	85	309
	0.5	1.0	0.6
All test items/questions read aloud	327	107	434
	0.7	1.2	0.8
Test items/questions signed	3	2	5
	0.0	0.0	0.0
Test items/questions interpreted for ELL student	16	0	16
	0.0	0.0	0.0
Amplification device	1	0	1
	0.0	0.0	0.0
Magnification device	1	0	1
	0.0	0.0	0.0
Color overlay	1	N/A	1
	0.0	N/A	0.0
Other (per Accommodations Guidelines)	9	11	20
	0.0	0.1	0.0
Spanish version	256	N/A	256
	0.6	N/A	0.5
Online Accommodations Received			
Audio	N/A	565	565
	N/A	6.4	1.1
Color Chooser	N/A	4	4
	N/A	0.0	0.0
Contrasting Text Chooser	N/A	4	4
	N/A	0.0	0.0
Number of assessed students	43,750	8,793	52,543

Incidence of Presentation Accommodations Received on the Winter 2014/2015 Keystone: Literature

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	4 0.0	N/A N/A	4 0.0
Large print format	18 0.0	N/A N/A	18 0.0
Computer Assistive Technology	1 0.0	N/A N/A	1 0.0
Amplification device	2 0.0	0 0.0	2 0.0
Magnification device	0 0.0	1 0.0	1 0.0
Color overlay	0 0.0	N/A N/A	0 0.0
Other (per Accommodations Guidelines)	13 0.0	13 0.2	26 0.1
Online Accommodations Received			
Color Chooser	N/A N/A	0 0.0	0 0.0
Contrasting Text Chooser	N/A N/A	1 0.0	1 0.0
Number of assessed students	37,568	7,896	45,464

Incidence of Response Accommodations Received on the Winter 2014/2015 Keystone: Algebra I

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	22 0.0	0 0.0	22 0.0
Test administrator scribed open-ended responses at student's direction	28 0.0	1 0.0	29 0.0
Test administrator transcribed student responses	31 0.0	0 0.0	31 0.0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	3 0.0	0 0.0	3 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	8 0.0	0 0.0	8 0.0
Keyboard, word processor, or computer	5 0.0	N/A N/A	5 0.0
Brailler/Notetaker	5 0.0	N/A N/A	5 0.0
Augmentative communication device	1 0.0	0 0.0	1 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	1 0.0	N/A N/A	1 0.0
Translation dictionary for ELL student	62 0.1	12 0.1	74 0.1
Other (per Accommodations Guidelines)	2 0.0	11 0.1	13 0.0
Number of assessed students	63,978	10,047	74,025

Incidence of Response Accommodations Received on the Winter 2014/2015 Keystone: Biology

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	20 0.0	0 0.0	20 0.0
Test administrator scribed open-ended responses at student's direction	26 0.1	1 0.0	27 0.1
Test administrator transcribed student responses	48 0.1	1 0.0	49 0.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	4 0.0	0 0.0	4 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	7 0.0	0 0.0	7 0.0
Keyboard, word processor, or computer	16 0.0	N/A N/A	16 0.0
Brailler/Notetaker	5 0.0	N/A N/A	5 0.0
Augmentative communication device	0 0.0	0 0.0	0 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Translation dictionary for ELL student	50 0.1	1 0.0	51 0.1
Other (per Accommodations Guidelines)	14 0.0	11 0.1	25 0.0
Number of assessed students	43,750	8,793	52,543

Incidence of Response Accommodations Received on the Winter 2014/2015 Keystone: Literature

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	19 0.1	0 0.0	19 0.0
Test administrator scribed open-ended responses at student's direction	32 0.1	0 0.0	32 0.1
Test administrator transcribed student responses	33 0.1	0 0.0	33 0.1
Keyboard, word processor, or computer	17 0.0	N/A N/A	17 0.0
Brailler/Notetaker	1 0.0	N/A N/A	1 0.0
Augmentative communication device	0 0.0	0 0.0	0 0.0
Audio recording of student responses	1 0.0	0 0.0	1 0.0
Computer Assistive Technology	1 0.0	N/A N/A	1 0.0
Other (per Accommodations Guidelines)	4 0.0	8 0.1	12 0.0
Number of assessed students	37,568	7,896	45,464

Incidence of Setting Accommodations Received on the Winter 2014/2015 Keystone: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	9 0.0	0 0.0	9 0.0
One-on-one setting	123 0.2	1 0.0	124 0.2
Small group setting	5,380 8.4	749 7.5	6,129 8.3
Other (per Accommodations Guidelines)	68 0.1	39 0.4	107 0.1
Number of assessed students	63,978	10,047	74,025

Incidence of Setting Accommodations Received on the Winter 2014/2015 Keystone: Biology

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	8 0.0	0 0.0	8 0.0
One-on-one setting	125 0.3	2 0.0	127 0.2
Small group setting	3,959 9.0	730 8.3	4,689 8.9
Other (per Accommodations Guidelines)	43 0.1	37 0.4	80 0.2
Number of assessed students	43,750	8,793	52,543

Incidence of Setting Accommodations Received on the Winter 2014/2015 Keystone: Literature

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	8	1	9
	0.0	0.0	0.0
One-on-one setting	96	2	98
	0.3	0.0	0.2
Small group setting	3,968	722	4,690
	10.6	9.1	10.3
Other (per Accommodations Guidelines)	65	34	99
	0.2	0.4	0.2
Number of assessed students	37,568	7,896	45,464

Incidence of Timing Accommodations Received on the Winter 2014/2015 Keystone: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	5,280	842	6,122
	8.3	8.4	8.3
Frequent breaks	292	128	420
	0.5	1.3	0.6
Changed test schedule	103	1	104
	0.2	0.0	0.1
Other (per Accommodations Guidelines)	6	1	7
	0.0	0.0	0.0
Number of assessed students	63,978	10,047	74,025

Incidence of Timing Accommodations Received on the Winter 2014/2015 Keystone: Biology

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	1,825	541	2,366
	4.2	6.2	4.5
Frequent breaks	250	138	388
	0.6	1.6	0.7
Changed test schedule	49	0	49
	0.1	0.0	0.1
Other (per Accommodations Guidelines)	1	1	2
	0.0	0.0	0.0
Number of assessed students	43,750	8,793	52,543

Incidence of Timing Accommodations Received on the Winter 2014/2015 Keystone: Literature

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	3,077 8.2	583 7.4	3,660 8.1
Frequent breaks	179 0.5	124 1.6	303 0.7
Changed test schedule	51 0.1	2 0.0	53 0.1
Other (per Accommodations Guidelines)	3 0.0	1 0.0	4 0.0
Number of assessed students	37,568	7,896	45,464

Accommodation Rate for Non-IEP and IEP Students on the Winter 2014/2015 Keystone

Course Tested	Student Subgroup Tested	PPT	CBT	Total
		N/Pct	N/Pct	N/Pct
Algebra I	**Non-IEP Students**	51,991	8,037	60,028
	Non-Accommodated	47,638 91.6	7,595 94.5	55,233 92.0
	Accommodated	4,353 8.4	442 5.5	4,795 8.0
	IEP Students	11,987	2,010	13,997
	Non-Accommodated	6,640 55.4	1,099 54.7	7,739 55.3
	Accommodated	5,347 44.6	911 45.3	6,258 44.7
Biology	**Non-IEP Students**	35,356	6,965	42,321
	Non-Accommodated	33,977 96.1	6,812 97.8	40,789 96.4
	Accommodated	1,379 3.9	153 2.2	1,532 3.6
	IEP Students	8,394	1,828	10,222
	Non-Accommodated	4,529 54.0	880 48.1	5,409 52.9
	Accommodated	3,865 46.0	948 51.9	4,813 47.1
Literature	**Non-IEP Students**	29,088	6,073	35,161
	Non-Accommodated	26,740 91.9	5,897 97.1	32,637 92.8
	Accommodated	2,348 8.1	176 2.9	2,524 7.2
	IEP Students	8,480	1,823	10,303
	Non-Accommodated	4,486 52.9	1,056 57.9	5,542 53.8
	Accommodated	3,994 47.1	767 42.1	4,761 46.2

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2014/2015 Keystone: Algebra I

Classification of Students Regarding IEP and ELL					
Accommodation Received by Administration Mode		General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Some test items/questions read aloud	12 0.0	452 3.9	11 0.5	7 1.7
	All test items/questions read aloud	7 0.0	256 2.2	5 0.2	7 1.7
	Small group setting	345 0.7	4,760 41.1	149 6.3	126 31.0
	Extended time	3,655 7.4	1,362 11.8	232 9.8	31 7.6
	Frequent breaks	23 0.0	255 2.2	8 0.3	6 1.5
	Number assessed	49,629	11,580	2,362	407
CBT	Some test items/questions read aloud	8 0.1	73 3.7	13 10.2	3 9.7
	All test items/questions read aloud	0 0.0	58 2.9	3 2.3	0 0.0
	Small group setting	25 0.3	706 35.7	11 8.6	7 22.6
	Extended time	380 4.8	415 21.0	43 33.6	4 12.9
	Frequent breaks	7 0.1	121 6.1	0 0.0	0 0.0
	Number assessed	7,909	1,979	128	31
Total	Some test items/questions read aloud	20 0.0	525 3.9	24 1.0	10 2.3
	All test items/questions read aloud	7 0.0	314 2.3	8 0.3	7 1.6
	Small group setting	370 0.6	5,466 40.3	160 6.4	133 30.4
	Extended time	4,035 7.0	1,777 13.1	275 11.0	35 8.0
	Frequent breaks	30 0.1	376 2.8	8 0.3	6 1.4
	Number assessed	57,538	13,559	2,490	438

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2014/2015 Keystone: Biology

Accommodation Received by Administration Mode		Classification of Students Regarding IEP and ELL			
		General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Some test items/questions read aloud	3 0.0	194 2.4	21 1.4	6 2.7
	All test items/questions read aloud	6 0.0	310 3.8	4 0.3	7 3.2
	Small group setting	220 0.6	3,534 43.2	125 8.6	80 36.2
	Extended time	938 2.8	797 9.8	80 5.5	10 4.5
	Frequent breaks	24 0.1	212 2.6	10 0.7	4 1.8
	Number assessed	33,899	8,173	1,457	221
CBT	Some test items/questions read aloud	5 0.1	74 4.1	5 5.8	1 3.3
	All test items/questions read aloud	0 0.0	105 5.8	1 1.2	1 3.3
	Small group setting	15 0.2	696 38.7	3 3.5	16 53.3
	Extended time	118 1.7	390 21.7	24 27.9	9 30.0
	Frequent breaks	6 0.1	131 7.3	0 0.0	1 3.3
	Number assessed	6,879	1,798	86	30
Total	Some test items/questions read aloud	8 0.0	268 2.7	26 1.7	7 2.8
	All test items/questions read aloud	6 0.0	415 4.2	5 0.3	8 3.2
	Small group setting	235 0.6	4,230 42.4	128 8.3	96 38.2
	Extended time	1,056 2.6	1,187 11.9	104 6.7	19 7.6
	Frequent breaks	30 0.1	343 3.4	10 0.6	5 2.0
	Number assessed	40,778	9,971	1,543	251

Incidence of IEP and ELL Students Receiving Accommodations on the Winter 2014/2015 Keystone: Literature

Classification of Students Regarding IEP and ELL					
Accommodation Received by Administration Mode		General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Small group setting	199 0.7	3,580 43.5	95 7.6	94 37.2
	Extended time	1,980 7.1	969 11.8	106 8.4	22 8.7
	Frequent breaks	13 0.0	153 1.9	7 0.6	6 2.4
	Number assessed	27,833	8,227	1,255	253
CBT	Small group setting	20 0.3	690 38.5	2 2.4	10 34.5
	Extended time	135 2.3	415 23.1	26 30.6	7 24.1
	Frequent breaks	7 0.1	115 6.4	1 1.2	1 3.4
	Number assessed	5,988	1,794	85	29
Total	Small group setting	219 0.6	4,270 42.6	97 7.2	104 36.9
	Extended time	2,115 6.3	1,384 13.8	132 9.9	29 10.3
	Frequent breaks	20 0.1	268 2.7	8 0.6	7 2.5
	Number assessed	33,821	10,021	1,340	282

SUMMER 2015

Students Assessed on the Summer 2015 Keystone: Algebra I

	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	11	0	10	139	598	447	133	5	1,343
Total number of CBT processed	0	0	17	120	163	184	73	0	557
Total number of tests processed	11	0	27	259	761	631	206	5	1,900
Total number of tests processed with a score	11 100.0	0 0.0	15 55.6	245 94.6	733 96.3	610 96.7	198 96.1	5 100.0	1,817 95.6
Total number of tests processed without a score	0 0.0	0 0.0	12 44.4	14 5.4	28 3.7	21 3.3	8 3.9	0 0.0	83 4.4

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

Students Assessed on the Summer 2015 Keystone: Biology

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	4	0	272	503	89	4	872
Total number of CBT processed	0	0	28	181	32	0	241
Total number of tests processed	4	0	300	684	121	4	1,113
Total number of tests processed with a score	4 100.0	0 0.0	298 99.3	663 96.9	114 94.2	4 100.0	1,083 97.3
Total number of tests processed without a score	0 0.0	0 0.0	2 0.7	21 3.1	7 5.8	0 0.0	30 2.7

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

Students Assessed on the Summer 2015 Keystone: Literature

	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Total number of PPT processed	5	0	48	260	98	1	412
Total number of CBT processed	0	1	7	94	79	3	184
Total number of tests processed	5	1	55	354	177	4	596
Total number of tests processed with a score	5 100.0	1 100.0	50 90.9	340 96.0	164 92.7	1 25.0	561 94.1
Total number of tests processed without a score	0 0.0	0 0.0	5 9.1	14 4.0	13 7.3	3 75.0	35 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 Summer 2015 Keystone: Algebra I

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	3 14.3	0 0.0	0 0.0	3 3.6
Absent without make-up	0 0.0	0 0.0	6 50.0	7 50.0	6 21.4	3 14.3	0 0.0	0 0.0	22 26.5
Non-attempt	0 0.0	0 0.0	6 50.0	6 42.9	21 75.0	11 52.4	6 75.0	0 0.0	50 60.2
Medical emergency	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	1 4.8	0 0.0	0 0.0	1 1.2
Parental request	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Other reasons	0 0.0	0 0.0	0 0.0	1 7.1	1 3.6	3 14.3	2 25.0	0 0.0	7 8.4
Total not assessed	0	0	12	14	28	21	8	0	83

*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 2015 Keystone: Biology

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	0 0.0	0 0.0	0 0.0	1 4.8	1 14.3	0 0.0	2 6.7
Absent without make-up	0 0.0	0 0.0	0 0.0	8 38.1	0 0.0	0 0.0	8 26.7
Non-attempt	0 0.0	0 0.0	1 50.0	10 47.6	3 42.9	0 0.0	14 46.7
Medical emergency	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Parental request	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Other reasons	0 0.0	0 0.0	1 50.0	2 9.5	3 42.9	0 0.0	6 20.0
Total not assessed	0	0	2	21	7	0	30

*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 2015 Keystone: Literature

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Extended absence from school	0	0	0	2	0	0	2
	0.0	0.0	0.0	14.3	0.0	0.0	5.7
Absent without make-up	0	0	0	2	0	0	2
	0.0	0.0	0.0	14.3	0.0	0.0	5.7
Non-attempt	0	0	3	8	11	3	25
	0.0	0.0	60.0	57.1	84.6	100.0	71.4
ELL in first year in U.S. schools	0	0	0	0	0	0	0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Medical emergency	0	0	0	0	0	0	0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Parental request	0	0	0	0	0	0	0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Other reasons	0	0	2	2	2	0	6
	0.0	0.0	40.0	14.3	15.4	0.0	17.1
Total not assessed	0	0	5	14	13	3	35

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

Demographic Characteristics of Students taking the Summer 2015 Keystone: Algebra I

Demographic or Educational Characteristic	Other* N/Pct	Gr.6 N/Pct	Gr.7 N/Pct	Gr.8 N/Pct	Gr.9 N/Pct	Gr.10 N/Pct	Gr.11 N/Pct	Gr.12 N/Pct	Total N/Pct
Gender									
Female	5 45.5	0 0.0	8 53.3	129 52.7	347 47.3	316 51.8	87 43.9	0 0.0	892 49.1
Male	5 45.5	0 0.0	7 46.7	116 47.3	358 48.8	286 46.9	110 55.6	4 80.0	886 48.8
Race/Ethnicity									
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0	0 0.0	1 0.1
Asian (not Hispanic)	1 9.1	0 0.0	2 13.3	8 3.3	19 2.6	13 2.1	2 1.0	0 0.0	45 2.5
Black or African American (not Hispanic)	2 18.2	0 0.0	0 0.0	10 4.1	59 8.0	162 26.6	30 15.2	0 0.0	263 14.5
Hispanic (any race)	0 0.0	0 0.0	1 6.7	12 4.9	33 4.5	27 4.4	3 1.5	0 0.0	76 4.2
Multi-Racial (not Hispanic)	0 0.0	0 0.0	0 0.0	3 1.2	16 2.2	13 2.1	6 3.0	0 0.0	38 2.1
White (not Hispanic)	5 45.5	0 0.0	12 80.0	209 85.3	563 76.8	383 62.8	155 78.3	4 80.0	1,331 73.3
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	0 0.0	1 0.4	0 0.0	1 0.2	0 0.0	0 0.0	2 0.1
Educational Category and Other Demographic Groups									
IEP (not gifted)	0 0.0	0 0.0	1 6.7	10 4.1	98 13.4	119 19.5	59 29.8	3 60.0	290 16.0
Student exited IEP in last 2 years	0 0.0	0 0.0	0 0.0	1 0.4	12 1.6	6 1.0	1 0.5	0 0.0	20 1.1
Title I	0 0.0	0 0.0	0 0.0	12 4.9	17 2.3	161 26.4	50 25.3	0 0.0	240 13.2
Title III served	0 0.0	0 0.0	0 0.0	1 0.4	6 0.8	8 1.3	0 0.0	0 0.0	15 0.8
Title III not served	0 0.0	0 0.0	0 0.0	18 7.3	27 3.7	28 4.6	2 1.0	0 0.0	75 4.1
Migrant student	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
ELL (enrolled after 8/1/14)	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	3 0.5	0 0.0	0 0.0	3 0.2
ELL (enrolled on or before 8/1/14)	0 0.0	0 0.0	0 0.0	1 0.4	5 0.7	5 0.8	0 0.0	0 0.0	11 0.6

Demographic Characteristics of Students taking the Summer 2015 Keystone: Algebra I (continued)

Demographic or Educational Characteristic	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	0 0.0	1 0.4	1 0.1	0 0.0	0 0.0	0 0.0	2 0.1
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	0 0.0	0 0.0	1 0.1	0 0.0	0 0.0	0 0.0	1 0.1
Former ELL no longer monitored	0 0.0	0 0.0	0 0.0	5 2.0	11 1.5	4 0.7	0 0.0	0 0.0	20 1.1
Foreign exchange student	0 0.0	0 0.0	0 0.0	0 0.0	1 0.1	0 0.0	0 0.0	0 0.0	1 0.1
Economically disadvantaged	0 0.0	0 0.0	2 13.3	34 13.9	168 22.9	225 36.9	63 31.8	1 20.0	493 27.1
Historically Underperforming Subgroup	0 0.0	0 0.0	3 20.0	42 17.1	244 33.3	297 48.7	104 52.5	3 60.0	693 38.1
Enrollment in school of residence after 10/1/14	2 18.2	0 0.0	0 0.0	15 6.1	29 4.0	26 4.3	6 3.0	0 0.0	78 4.3
Enrollment in district of residence after 10/1/14	2 18.2	0 0.0	0 0.0	8 3.3	21 2.9	24 3.9	4 2.0	0 0.0	59 3.2
Enrollment as PA resident after 10/1/14	2 18.2	0 0.0	0 0.0	5 2.0	4 0.5	16 2.6	3 1.5	0 0.0	30 1.7
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	1 9.1	0 0.0	2 13.3	7 2.9	99 13.5	52 8.5	28 14.1	0 0.0	189 10.4
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	0 0.0	0 0.0	2 13.3	5 2.0	40 5.5	48 7.9	27 13.6	0 0.0	122 6.7
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Number of assessed students	11	0	15	245	733	610	198	5	1,817

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

Demographic Characteristics of Students taking the Summer 2015 Keystone: Biology

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	0 0.0	0 0.0	158 53.0	334 50.4	64 56.1	3 75.0	559 51.6
Male	3 75.0	0 0.0	130 43.6	317 47.8	50 43.9	1 25.0	501 46.3
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	0 0.0	3 0.5	0 0.0	0 0.0	3 0.3
Asian (not Hispanic)	0 0.0	0 0.0	22 7.4	21 3.2	4 3.5	0 0.0	47 4.3
Black or African American (not Hispanic)	1 25.0	0 0.0	45 15.1	59 8.9	25 21.9	0 0.0	130 12.0
Hispanic (any race)	0 0.0	0 0.0	15 5.0	55 8.3	2 1.8	0 0.0	72 6.6
Multi-Racial (not Hispanic)	0 0.0	0 0.0	9 3.0	13 2.0	4 3.5	0 0.0	26 2.4
White (not Hispanic)	2 50.0	0 0.0	196 65.8	499 75.3	78 68.4	4 100.0	779 71.9
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Educational Category and Other Demographic Groups							
IEP (not gifted)	0 0.0	0 0.0	51 17.1	130 19.6	26 22.8	1 25.0	208 19.2
Student exited IEP in last 2 years	0 0.0	0 0.0	4 1.3	13 2.0	0 0.0	1 25.0	18 1.7
Title I	0 0.0	0 0.0	12 4.0	45 6.8	24 21.1	0 0.0	81 7.5
Title III served	0 0.0	0 0.0	3 1.0	14 2.1	2 1.8	0 0.0	19 1.8
Title III not served	0 0.0	0 0.0	2 0.7	20 3.0	0 0.0	0 0.0	22 2.0
Migrant student	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
ELL (enrolled after 8/1/14)	0 0.0	0 0.0	1 0.3	4 0.6	1 0.9	0 0.0	6 0.6
ELL (enrolled on or before 8/1/14)	0 0.0	0 0.0	2 0.7	10 1.5	1 0.9	0 0.0	13 1.2

Demographic Characteristics of Students taking the Summer 2015 Keystone: Biology (continued)

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	0 0.0	4 0.6	0 0.0	0 0.0	4 0.4
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0	0 0.0	1 0.1
Former ELL no longer monitored	0 0.0	0 0.0	4 1.3	12 1.8	0 0.0	0 0.0	16 1.5
Foreign exchange student	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Economically disadvantaged	0 0.0	0 0.0	86 28.9	179 27.0	35 30.7	0 0.0	300 27.7
Historically Underperforming Subgroup	0 0.0	0 0.0	118 39.6	257 38.8	55 48.2	1 25.0	431 39.8
Enrollment in school of residence after 10/1/14	1 25.0	0 0.0	9 3.0	29 4.4	2 1.8	0 0.0	41 3.8
Enrollment in district of residence after 10/1/14	1 25.0	0 0.0	7 2.3	20 3.0	1 0.9	0 0.0	29 2.7
Enrollment as PA resident after 10/1/14	1 25.0	0 0.0	2 0.7	10 1.5	1 0.9	0 0.0	14 1.3
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	1 25.0	0 0.0	42 14.1	37 5.6	30 26.3	1 25.0	111 10.2
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	0 0.0	0 0.0	21 7.0	34 5.1	27 23.7	0 0.0	82 7.6
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Number of assessed students	4	0	298	663	114	4	1,083

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

Demographic Characteristics of Students taking the Summer 2015 Keystone: Literature

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Gender							
Female	2 40.0	0 0.0	18 36.0	146 42.9	73 44.5	1 100.0	240 42.8
Male	3 60.0	1 100.0	32 64.0	187 55.0	91 55.5	0 0.0	314 56.0
Race/Ethnicity							
American Indian/Alaskan Native (not Hispanic)	0 0.0	0 0.0	0 0.0	2 0.6	0 0.0	0 0.0	2 0.4
Asian (not Hispanic)	0 0.0	0 0.0	1 2.0	7 2.1	4 2.4	0 0.0	12 2.1
Black or African American (not Hispanic)	2 40.0	0 0.0	3 6.0	43 12.6	24 14.6	0 0.0	72 12.8
Hispanic (any race)	0 0.0	1 100.0	0 0.0	25 7.4	6 3.7	1 100.0	33 5.9
Multi-Racial (not Hispanic)	0 0.0	0 0.0	0 0.0	16 4.7	7 4.3	0 0.0	23 4.1
White (not Hispanic)	3 60.0	0 0.0	46 92.0	239 70.3	120 73.2	0 0.0	408 72.7
Native Hawaiian or Other Pacific Islander (not Hispanic)	0 0.0	0 0.0	0 0.0	0 0.0	2 1.2	0 0.0	2 0.4
Educational Category and Other Demographic Groups							
IEP (not gifted)	0 0.0	0 0.0	8 16.0	73 21.5	50 30.5	0 0.0	131 23.4
Student exited IEP in last 2 years	0 0.0	0 0.0	0 0.0	7 2.1	0 0.0	0 0.0	7 1.2
Title I	0 0.0	0 0.0	0 0.0	16 4.7	41 25.0	1 100.0	58 10.3
Title III served	0 0.0	0 0.0	2 4.0	6 1.8	5 3.0	0 0.0	13 2.3
Title III not served	0 0.0	0 0.0	0 0.0	24 7.1	7 4.3	1 100.0	32 5.7
Migrant student	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
ELL (enrolled after 8/1/14)	0 0.0	0 0.0	0 0.0	1 0.3	0 0.0	0 0.0	1 0.2
ELL (enrolled on or before 8/1/14)	0 0.0	0 0.0	0 0.0	4 1.2	5 3.0	0 0.0	9 1.6

Demographic Characteristics of Students taking the Summer 2015 Keystone: Literature (continued)

Demographic or Educational Characteristic	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct	N/Pct
Exited ESL/bilingual program and in first year of monitoring	0 0.0	0 0.0	1 2.0	4 1.2	0 0.0	0 0.0	5 0.9
Exited ESL/bilingual program and in 2nd year of monitoring	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Former ELL no longer monitored	0 0.0	0 0.0	0 0.0	2 0.6	0 0.0	1 100.0	3 0.5
Foreign exchange student	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Economically disadvantaged	0 0.0	1 100.0	5 10.0	116 34.1	57 34.8	1 100.0	180 32.1
Historically Underperforming Subgroup	0 0.0	1 100.0	11 22.0	161 47.4	91 55.5	1 100.0	265 47.2
Enrollment in school of residence after 10/1/14	1 20.0	0 0.0	1 2.0	15 4.4	7 4.3	0 0.0	24 4.3
Enrollment in district of residence after 10/1/14	1 20.0	0 0.0	1 2.0	14 4.1	3 1.8	0 0.0	19 3.4
Enrollment as PA resident after 10/1/14	1 20.0	0 0.0	0 0.0	5 1.5	3 1.8	0 0.0	9 1.6
Enrollment in school of residence after 10/1/13 but on/before 10/1/14	1 20.0	0 0.0	12 24.0	31 9.1	19 11.6	0 0.0	63 11.2
Enrollment in district of residence after 10/1/13 but on/before 10/1/14	0 0.0	0 0.0	4 8.0	24 7.1	19 11.6	0 0.0	47 8.4
Home schooled	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Court/agency placed	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
Number of assessed students	5	1	50	340	164	1	561

*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 2015 Keystone: Algebra I

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	0 0.0	N/A N/A	0 0.0
Large print format	1 0.1	N/A N/A	1 0.1
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Some test items/questions read aloud	1 0.1	0 0.0	1 0.1
All test items/questions read aloud	5 0.4	3 0.6	8 0.4
Test items/questions signed	1 0.1	0 0.0	1 0.1
Test items/questions interpreted for ELL student	0 0.0	0 0.0	0 0.0
Amplification device	0 0.0	0 0.0	0 0.0
Magnification device	0 0.0	0 0.0	0 0.0
Color overlay	0 0.0	N/A N/A	0 0.0
Other (per Accommodations Guidelines)	0 0.0	0 0.0	0 0.0
Spanish version	0 0.0	N/A N/A	0 0.0
Online Accommodations Received			
Audio	N/A N/A	8 1.6	8 0.4
Color Chooser	N/A N/A	1 0.2	1 0.1
Contrasting Text Chooser	N/A N/A	0 0.0	0 0.0
Number of assessed students	1,312	505	1,817

Incidence of Presentation Accommodations Received on the Summer 2015 Keystone: Biology

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	0	N/A	0
	0.0	N/A	0.0
Large print format	0	N/A	0
	0.0	N/A	0.0
Computer Assistive Technology	0	N/A	0
	0.0	N/A	0.0
Some test items/questions read aloud	5	0	5
	0.6	0.0	0.5
All test items/questions read aloud	8	1	9
	0.9	0.4	0.8
Test items/questions signed	0	0	0
	0.0	0.0	0.0
Test items/questions interpreted for ELL student	0	0	0
	0.0	0.0	0.0
Amplification device	0	0	0
	0.0	0.0	0.0
Magnification device	0	0	0
	0.0	0.0	0.0
Color overlay	0	N/A	0
	0.0	N/A	0.0
Other (per Accommodations Guidelines)	0	1	1
	0.0	0.4	0.1
Spanish version	0	N/A	0
	0.0	N/A	0.0
Online Accommodations Received			
Audio	N/A	2	2
	N/A	0.9	0.2
Color Chooser	N/A	0	0
	N/A	0.0	0.0
Contrasting Text Chooser	N/A	0	0
	N/A	0.0	0.0
Number of assessed students	855	228	1,083

Incidence of Presentation Accommodations Received on the Summer 2015 Keystone: Literature

Type of Presentation Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Braille format	0 0.0	N/A N/A	0 0.0
Large print format	0 0.0	N/A N/A	0 0.0
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Amplification device	0 0.0	0 0.0	0 0.0
Magnification device	0 0.0	0 0.0	0 0.0
Color overlay	0 0.0	N/A N/A	0 0.0
Other (per Accommodations Guidelines)	0 0.0	1 0.6	1 0.2
Online Accommodations Received			
Color Chooser	N/A N/A	0 0.0	0 0.0
Contrasting Text Chooser	N/A N/A	0 0.0	0 0.0
Number of assessed students	393	168	561

Incidence of Response Accommodations Received on the Summer 2015 Keystone: Algebra I

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	1 0.1	0 0.0	1 0.1
Test administrator scribed open-ended responses at student's direction	0 0.0	0 0.0	0 0.0
Test administrator transcribed student responses	4 0.3	0 0.0	4 0.2
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	0 0.0	0 0.0	0 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	0 0.0	0 0.0	0 0.0
Keyboard, word processor, or computer	0 0.0	N/A N/A	0 0.0
Brailler/Notetaker	0 0.0	N/A N/A	0 0.0
Augmentative communication device	0 0.0	0 0.0	0 0.0
Audio recording of student responses	0 0.0	0 0.0	0 0.0
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Translation dictionary for ELL student	0 0.0	0 0.0	0 0.0
Other (per Accommodations Guidelines)	1 0.1	0 0.0	1 0.1
Number of assessed students	1,312	505	1,817

Incidence of Response Accommodations Received on the Summer 2015 Keystone: Biology

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	1 0.1	0 0.0	1 0.1
Test administrator scribed open-ended responses at student's direction	2 0.2	0 0.0	2 0.2
Test administrator transcribed student responses	1 0.1	0 0.0	1 0.1
Qualified interpreter translated, transcribed, and/or scribed student's signed responses	0 0.0	0 0.0	0 0.0
Qualified interpreter translated, transcribed, and/or scribed ELL student responses	0 0.0	0 0.0	0 0.0
Keyboard, word processor, or computer	0 0.0	N/A N/A	0 0.0
Brailler/Notetaker	0 0.0	N/A N/A	0 0.0
Augmentative communication device	0 0.0	0 0.0	0 0.0
Audio recording of student responses	0 0.0	0 0.0	0 0.0
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Translation dictionary for ELL student	0 0.0	0 0.0	0 0.0
Other (per Accommodations Guidelines)	1 0.1	0 0.0	1 0.1
Number of assessed students	855	228	1,083

Incidence of Response Accommodations Received on the Summer 2015 Keystone: Literature

Type of Response Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Test administrator marked multiple-choice responses at student's direction	0 0.0	0 0.0	0 0.0
Test administrator scribed open-ended responses at student's direction	1 0.3	0 0.0	1 0.2
Test administrator transcribed student responses	0 0.0	0 0.0	0 0.0
Keyboard, word processor, or computer	2 0.5	N/A N/A	2 0.4
Brailler/Notetaker	0 0.0	N/A N/A	0 0.0
Augmentative communication device	0 0.0	0 0.0	0 0.0
Audio recording of student responses	0 0.0	0 0.0	0 0.0
Computer Assistive Technology	0 0.0	N/A N/A	0 0.0
Other (per Accommodations Guidelines)	0 0.0	0 0.0	0 0.0
Number of assessed students	393	168	561

Incidence of Setting Accommodations Received on the Summer 2015 Keystone: Algebra I

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	0	0	0
	0.0	0.0	0.0
One-on-one setting	1	0	1
	0.1	0.0	0.1
Small group setting	39	15	54
	3.0	3.0	3.0
Other (per Accommodations Guidelines)	0	0	0
	0.0	0.0	0.0
Number of assessed students	1,312	505	1,817

Incidence of Setting Accommodations Received on the Summer 2015 Keystone: Biology

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	0	0	0
	0.0	0.0	0.0
One-on-one setting	7	0	7
	0.8	0.0	0.6
Small group setting	35	15	50
	4.1	6.6	4.6
Other (per Accommodations Guidelines)	0	0	0
	0.0	0.0	0.0
Number of assessed students	855	228	1,083

Incidence of Setting Accommodations Received on the Summer 2015 Keystone: Literature

Type of Setting Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Hospital/home setting	0 0.0	0 0.0	0 0.0
One-on-one setting	2 0.5	0 0.0	2 0.4
Small group setting	25 6.4	8 4.8	33 5.9
Other (per Accommodations Guidelines)	0 0.0	1 0.6	1 0.2
Number of assessed students	393	168	561

Incidence of Timing Accommodations Received on the Summer 2015 Keystone: Algebra I

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	64 4.9	10 2.0	74 4.1
Frequent breaks	4 0.3	3 0.6	7 0.4
Changed test schedule	0 0.0	0 0.0	0 0.0
Other (per Accommodations Guidelines)	1 0.1	0 0.0	1 0.1
Number of assessed students	1,312	505	1,817

Incidence of Timing Accommodations Received on the Summer 2015 Keystone: Biology

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	9 1.1	4 1.8	13 1.2
Frequent breaks	3 0.4	1 0.4	4 0.4
Changed test schedule	1 0.1	0 0.0	1 0.1
Other (per Accommodations Guidelines)	1 0.1	0 0.0	1 0.1
Number of assessed students	855	228	1,083

Incidence of Timing Accommodations Received on the Summer 2015 Keystone: Literature

Type of Timing Accommodation	PPT	CBT	Total
	N/Pct	N/Pct	N/Pct
Extended time	0 0.0	4 2.4	4 0.7
Frequent breaks	0 0.0	2 1.2	2 0.4
Changed test schedule	0 0.0	0 0.0	0 0.0
Other (per Accommodations Guidelines)	1 0.3	0 0.0	1 0.2
Number of assessed students	393	168	561

Accommodation Rate for Non-IEP and IEP Students on the Summer 2015 Keystone Exams

Course Tested	Student Subgroup Tested	PPT N/Pct	CBT N/Pct	Total N/Pct
Algebra I	**Non-IEP Students**	1,083	444	1,527
	Non-Accommodated	1,030 95.1	444 100.0	1,474 96.5
	Accommodated	53 4.9	0 0.0	53 3.5
	IEP Students	229	61	290
	Non-Accommodated	174 76.0	43 70.5	217 74.8
	Accommodated	55 24.0	18 29.5	73 25.2
Biology	**Non-IEP Students**	679	196	875
	Non-Accommodated	667 98.2	196 100.0	863 98.6
	Accommodated	12 1.8	0 0.0	12 1.4
	IEP Students	176	32	208
	Non-Accommodated	136 77.3	16 50.0	152 73.1
	Accommodated	40 22.7	16 50.0	56 26.9
Literature	**Non-IEP Students**	297	133	430
	Non-Accommodated	296 99.7	132 99.2	428 99.5
	Accommodated	1 0.3	1 0.8	2 0.5
	IEP Students	96	35	131
	Non-Accommodated	71 74.0	27 77.1	98 74.8
	Accommodated	25 26.0	8 22.9	33 25.2

Incidence of IEP and ELL Students Receiving Selected Accommodations on the Summer 2015 Keystone: Algebra I

Accommodation Received by Administration Mode		Classification of Students Regarding IEP and ELL			
		General Education (non-IEP or ELL)	IEP and non-ELL	ELL and non-IEP	Both IEP and ELL
		N/Pct	N/Pct	N/Pct	N/Pct
PPT	Some test items/questions read aloud	1 0.1	0 0.0	0 0.0	0 0.0
	All test items/questions read aloud	1 0.1	4 1.7	0 0.0	0 0.0
	Small group setting	6 0.6	33 14.4	0 0.0	0 0.0
	Extended time	44 4.1	19 8.3	1 7.7	0 0.0
	Frequent breaks	0 0.0	4 1.7	0 0.0	0 0.0
	Number assessed	1,070	229	13	0
CBT	Some test items/questions read aloud	0 0.0	0 0.0	0 0.0	0 0.0
	All test items/questions read aloud	0 0.0	3 4.9	0 0.0	0 0.0
	Small group setting	0 0.0	15 24.6	0 0.0	0 0.0
	Extended time	0 0.0	10 16.4	0 0.0	0 0.0
	Frequent breaks	0 0.0	3 4.9	0 0.0	0 0.0
	Number assessed	443	61	1	0
Total	Some test items/questions read aloud	1 0.1	0 0.0	0 0.0	0 0.0
	All test items/questions read aloud	1 0.1	7 2.4	0 0.0	0 0.0
	Small group setting	6 0.4	48 16.6	0 0.0	0 0.0
	Extended time	44 2.9	29 10.0	1 7.1	0 0.0
	Frequent breaks	0 0.0	7 2.4	0 0.0	0 0.0
	Number assessed	1,513	290	14	0

Incidence of IEP and ELL Students Receiving Selected Accommodations on the Summer 2015 Keystone: Biology

Accommodation Received by Administration Mode		General Education (non-IEP or ELL) N/Pct	IEP and non-ELL N/Pct	ELL and non-IEP N/Pct	Both IEP and ELL N/Pct
		Classification of Students Regarding IEP and ELL			
PPT	Some test items/questions read aloud	0 0.0	5 2.9	0 0.0	0 0.0
	All test items/questions read aloud	0 0.0	8 4.6	0 0.0	0 0.0
	Small group setting	7 1.1	28 16.2	0 0.0	0 0.0
	Extended time	5 0.8	4 2.3	0 0.0	0 0.0
	Frequent breaks	1 0.2	2 1.2	0 0.0	0 0.0
	Number assessed	666	173	13	3
CBT	Some test items/questions read aloud	0 0.0	0 0.0	0 0.0	0 0.0
	All test items/questions read aloud	0 0.0	1 3.2	0 0.0	0 0.0
	Small group setting	0 0.0	15 48.4	0 0.0	0 0.0
	Extended time	0 0.0	4 12.9	0 0.0	0 0.0
	Frequent breaks	0 0.0	1 3.2	0 0.0	0 0.0
	Number assessed	194	31	2	1
Total	Some test items/questions read aloud	0 0.0	5 2.5	0 0.0	0 0.0
	All test items/questions read aloud	0 0.0	9 4.4	0 0.0	0 0.0
	Small group setting	7 0.8	43 21.1	0 0.0	0 0.0
	Extended time	5 0.6	8 3.9	0 0.0	0 0.0
	Frequent breaks	1 0.1	3 1.5	0 0.0	0 0.0
	Number assessed	860	204	15	4

Incidence of IEP and ELL Students Receiving Selected Accommodations on the Summer 2015 Keystone: Literature

Accommodation Received by Administration Mode		General Education (non-IEP or ELL) N/Pct	IEP and non-ELL N/Pct	ELL and non-IEP N/Pct	Both IEP and ELL N/Pct
PPT	Small group setting	1 0.3	24 25.3	0 0.0	0 0.0
	Extended time	0 0.0	0 0.0	0 0.0	0 0.0
	Frequent breaks	0 0.0	0 0.0	0 0.0	0 0.0
	Number assessed	289	95	8	1
CBT	Small group setting	0 0.0	8 22.9	0 0.0	0 0.0
	Extended time	1 0.8	3 8.6	0 0.0	0 0.0
	Frequent breaks	0 0.0	2 5.7	0 0.0	0 0.0
	Number assessed	132	35	1	0
Total	Small group setting	1 0.2	32 24.6	0 0.0	0 0.0
	Extended time	1 0.2	3 2.3	0 0.0	0 0.0
	Frequent breaks	0 0.0	2 1.5	0 0.0	0 0.0
	Number assessed	421	130	9	1

APPENDIX J: ITEM STATISTICS

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	*z* fit statistic
MS	Mean square fit 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

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	640597	0	75936	0.65	0.13	0.65	0.16	0.06	0.00	0.00	0.35	-0.20	0.35	-0.20	-0.12	-0.78	0.01	1.25	1.00	-1.25	0.99
2	605261	0	75936	0.57	0.28	0.11	0.57	0.04	0.00	0.00	0.28	-0.09	-0.22	0.28	-0.15	0.08	0.01	9.90	1.04	9.90	1.04
3	640591	0	75936	0.73	0.73	0.12	0.06	0.09	0.00	0.00	0.30	0.30	-0.16	-0.16	-0.15	-0.86	0.01	-9.90	0.91	-9.90	0.91
4	640578	0	75936	0.43	0.17	0.31	0.43	0.09	0.00	0.00	0.32	-0.15	-0.17	0.32	-0.07	0.70	0.01	9.90	1.07	9.90	1.11
5	605195	0	75936	0.55	0.55	0.24	0.14	0.06	0.00	0.00	0.27	0.27	-0.13	-0.15	-0.11	0.04	0.01	9.90	1.05	9.90	1.05
6	605114	0	75936	0.59	0.23	0.09	0.59	0.09	0.00	0.00	0.33	-0.16	-0.20	0.33	-0.13	-0.36	0.01	-2.77	0.99	-4.88	0.98
7	640568	0	75936	0.24	0.24	0.18	0.27	0.31	0.00	0.00	0.15	0.15	-0.04	-0.11	0.00	1.62	0.01	9.90	1.22	9.90	1.53
8	605255	0	75936	0.72	0.04	0.12	0.12	0.72	0.00	0.00	0.42	-0.20	-0.22	-0.22	0.42	-0.87	0.01	-9.90	0.85	-9.90	0.79
9	605262	0	75936	0.32	0.24	0.14	0.30	0.32	0.00	0.00	0.25	-0.03	-0.17	-0.09	0.25	0.95	0.01	9.90	1.05	9.90	1.12
10	640599	0	75936	0.39	0.22	0.19	0.39	0.19	0.01	0.00	0.14	-0.09	-0.10	0.14	0.04	0.61	0.01	9.90	1.17	9.90	1.25
11	640629	0	75936	0.47	0.47	0.13	0.26	0.13	0.01	0.00	0.34	0.34	-0.20	-0.09	-0.17	0.07	0.01	-5.26	0.99	0.56	1.00
12	640572	0	75936	0.31	0.29	0.25	0.31	0.15	0.01	0.00	0.21	-0.11	-0.07	0.21	-0.03	1.15	0.01	9.90	1.13	9.90	1.26
13	640543	0	75936	0.71	0.07	0.15	0.71	0.07	0.01	0.00	0.31	-0.20	-0.08	0.31	-0.21	-1.07	0.01	3.96	1.02	9.90	1.07
14	605057	0	75936	0.59	0.07	0.18	0.16	0.59	0.01	0.00	0.33	-0.17	-0.18	-0.11	0.33	-0.07	0.01	-7.25	0.98	-6.34	0.97
15	640579	0	75936	0.26	0.26	0.19	0.29	0.25	0.01	0.00	0.20	0.20	-0.10	-0.07	-0.03	1.41	0.01	9.90	1.14	9.90	1.32
16	640529	0	75936	0.46	0.15	0.46	0.24	0.14	0.01	0.00	0.27	-0.10	0.27	-0.10	-0.14	-0.03	0.01	9.90	1.07	9.90	1.09
17	605254	0	75936	0.65	0.16	0.10	0.09	0.65	0.01	0.00	0.46	-0.21	-0.24	-0.23	0.46	-0.74	0.01	-9.90	0.89	-9.90	0.83
18	605170	0	75936	0.69	0.08	0.69	0.11	0.12	0.01	0.00	0.39	-0.22	0.39	-0.18	-0.17	-1.12	0.01	8.64	1.04	0.54	1.00
19	605180	0	75936	0.81	0.05	0.81	0.11	0.03	0.00	0.00	0.32	-0.18	0.32	-0.19	-0.16	-1.20	0.01	-9.90	0.79	-9.90	0.73
20	605273	0	75936	0.91	0.04	0.04	0.91	0.02	0.00	0.00	0.33	-0.21	-0.20	0.33	-0.12	-2.11	0.01	-9.90	0.68	-9.90	0.49
21	605280	0	75936	0.49	0.14	0.24	0.13	0.49	0.00	0.00	0.40	-0.16	-0.21	-0.16	0.40	0.10	0.01	-9.90	0.94	-9.90	0.92
22	641441	0	75936	0.45	0.19	0.11	0.26	0.45	0.00	0.00	0.17	-0.11	-0.11	-0.02	0.17	0.75	0.01	9.90	1.25	9.90	1.35
23	641539	0	75936	0.27	0.33	0.27	0.24	0.16	0.00	0.00	0.09	-0.04	0.09	-0.01	-0.05	1.55	0.01	9.90	1.35	9.90	1.68
24	641466	0	75936	0.55	0.55	0.21	0.13	0.10	0.00	0.00	0.35	0.35	-0.18	-0.19	-0.12	-0.35	0.01	0.22	1.00	1.02	1.00
25	641524	0	75936	0.33	0.35	0.33	0.23	0.09	0.00	0.00	0.22	-0.18	0.22	-0.01	-0.05	0.92	0.01	9.90	1.08	9.90	1.18
26	605283	0	75936	0.65	0.12	0.15	0.65	0.08	0.00	0.00	0.36	-0.12	-0.23	0.36	-0.19	-0.42	0.01	-9.90	0.92	-9.90	0.89
27	605206	0	75936	0.51	0.08	0.14	0.51	0.27	0.00	0.00	0.34	-0.22	-0.19	0.34	-0.10	0.01	0.01	-2.18	0.99	-3.86	0.99
28	641523	0	75936	0.49	0.11	0.10	0.30	0.49	0.00	0.00	0.41	-0.12	-0.22	-0.21	0.41	-0.05	0.01	-9.90	0.94	-9.90	0.92
29	641460	0	75936	0.43	0.11	0.31	0.43	0.14	0.01	0.00	0.36	-0.19	-0.13	0.36	-0.15	0.07	0.01	-3.45	0.99	-5.98	0.98

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
30	605115	0	75936	0.50	0.18	0.50	0.22	0.10	0.00	0.00	0.42	-0.16	0.42	-0.24	-0.15	-0.60	0.01	9.90	1.11	9.90	1.10
31	641519	0	75936	0.22	0.24	0.39	0.22	0.15	0.01	0.00	0.07	-0.21	0.19	0.07	-0.08	1.85	0.01	9.90	1.39	9.90	1.83
32	641478	0	75936	0.42	0.26	0.15	0.16	0.42	0.01	0.00	0.33	-0.14	-0.14	-0.13	0.33	0.12	0.01	3.65	1.01	-0.81	1.00
33	641451	0	75936	0.52	0.52	0.14	0.24	0.10	0.01	0.00	0.38	0.38	-0.17	-0.17	-0.18	-0.28	0.01	-2.38	0.99	-4.80	0.98
34	605230	0	75936	0.50	0.17	0.50	0.19	0.14	0.01	0.00	0.33	-0.11	0.33	-0.17	-0.16	-0.12	0.01	3.85	1.01	5.70	1.02
35	605096	0	75936	0.48	0.17	0.48	0.22	0.13	0.01	0.00	0.28	-0.13	0.28	-0.12	-0.11	0.43	0.01	9.90	1.07	9.90	1.11
36	605168	0	75936	0.34	0.34	0.21	0.32	0.12	0.01	0.00	0.28	0.28	-0.10	-0.10	-0.11	1.00	0.01	9.90	1.09	9.90	1.17

Table J–3. Biology Multiple-Choice Item Statistics: Winter

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	643772	0	54236	0.65	0.08	0.12	0.15	0.65	0.00	0.00	0.32	-0.19	-0.19	-0.11	0.32	-0.70	0.01	-9.90	0.95	-9.79	0.94
2	643380	0	54236	0.65	0.16	0.65	0.13	0.06	0.00	0.00	0.39	-0.22	0.39	-0.22	-0.11	-0.72	0.01	-9.90	0.90	-9.90	0.86
3	635760	0	54236	0.61	0.25	0.05	0.61	0.08	0.00	0.00	0.35	-0.22	-0.20	0.35	-0.11	-0.63	0.01	-9.90	0.95	-9.90	0.91
4	641198	0	54236	0.51	0.24	0.12	0.51	0.13	0.00	0.00	0.39	-0.15	-0.21	0.39	-0.18	-0.37	0.01	-9.90	0.96	-9.90	0.93
5	610456	0	54236	0.68	0.04	0.68	0.08	0.19	0.00	0.00	0.30	-0.15	0.30	-0.20	-0.13	-0.98	0.01	-9.90	0.96	-9.90	0.93
6	607797	0	54236	0.59	0.29	0.06	0.59	0.06	0.00	0.00	0.27	-0.08	-0.23	0.27	-0.19	-0.64	0.01	9.90	1.03	9.90	1.08
7	642866	0	54236	0.36	0.36	0.19	0.21	0.24	0.00	0.00	0.27	0.27	-0.09	-0.22	-0.01	0.65	0.01	9.90	1.13	9.90	1.20
8	642850	0	54236	0.51	0.07	0.51	0.14	0.28	0.00	0.00	0.28	-0.17	0.28	-0.19	-0.07	-0.07	0.01	9.90	1.07	9.90	1.06
9	607704	0	54236	0.48	0.27	0.08	0.17	0.48	0.00	0.00	0.37	-0.11	-0.20	-0.21	0.37	0.01	0.01	-3.40	0.99	-4.35	0.98
10	611033	0	54236	0.30	0.30	0.15	0.36	0.20	0.00	0.00	0.27	0.27	-0.17	-0.04	-0.11	0.89	0.01	9.90	1.09	9.90	1.21
11	642841	0	54236	0.51	0.51	0.24	0.12	0.12	0.00	0.00	0.33	0.33	-0.11	-0.23	-0.14	-0.38	0.01	1.54	1.00	-3.49	0.98
12	640665	0	54236	0.34	0.10	0.45	0.34	0.11	0.00	0.00	0.26	-0.20	-0.07	0.26	-0.09	0.83	0.01	9.90	1.16	9.90	1.28
13	644729	0	54236	0.49	0.49	0.15	0.20	0.16	0.00	0.00	0.40	0.40	-0.21	-0.18	-0.14	-0.18	0.01	-9.90	0.95	-9.90	0.92
14	644733	0	54236	0.33	0.26	0.20	0.21	0.33	0.00	0.00	0.37	-0.08	-0.16	-0.18	0.37	0.61	0.01	-5.82	0.97	-0.39	1.00
15	607732	0	54236	0.62	0.62	0.14	0.18	0.05	0.01	0.00	0.39	0.39	-0.18	-0.21	-0.17	-0.59	0.01	-9.90	0.91	-9.90	0.86
16	642843	0	54236	0.54	0.16	0.54	0.13	0.16	0.01	0.00	0.37	-0.16	0.37	-0.25	-0.09	-0.12	0.01	-6.39	0.98	-8.56	0.96
17	642335	0	54236	0.38	0.14	0.28	0.38	0.20	0.01	0.00	0.30	-0.22	-0.08	0.30	-0.06	0.44	0.01	9.90	1.05	9.90	1.08
18	643387	0	54236	0.38	0.21	0.21	0.38	0.18	0.01	0.00	0.26	-0.08	-0.09	0.26	-0.13	0.35	0.01	9.90	1.08	9.90	1.11
19	611445	0	54236	0.57	0.12	0.19	0.10	0.57	0.01	0.00	0.43	-0.17	-0.19	-0.24	0.43	-0.20	0.01	-9.90	0.92	-9.90	0.88
20	611154	0	54236	0.48	0.48	0.14	0.18	0.18	0.01	0.00	0.33	0.33	-0.21	-0.11	-0.11	-0.22	0.01	0.50	1.00	2.24	1.01
21	642864	0	54236	0.48	0.20	0.19	0.48	0.12	0.01	0.00	0.37	-0.17	-0.15	0.37	-0.15	-0.42	0.01	-4.20	0.99	-3.72	0.98
22	642837	0	54236	0.38	0.16	0.30	0.15	0.38	0.01	0.00	0.36	-0.20	-0.07	-0.18	0.36	0.34	0.01	-4.25	0.98	-2.21	0.99
23	643371	0	54236	0.40	0.16	0.25	0.18	0.40	0.01	0.00	0.44	-0.17	-0.20	-0.15	0.44	0.05	0.01	-9.90	0.90	-9.90	0.88
24	643411	0	54236	0.37	0.26	0.37	0.24	0.13	0.01	0.00	0.23	-0.03	0.23	-0.10	-0.14	0.52	0.01	9.90	1.13	9.90	1.18
25	607756	0	54236	0.86	0.86	0.06	0.05	0.03	0.00	0.00	0.28	0.28	-0.14	-0.17	-0.16	-1.95	0.01	-9.90	0.76	-9.90	0.70
26	611614	0	54236	0.47	0.38	0.10	0.05	0.47	0.00	0.00	0.40	-0.18	-0.25	-0.17	0.40	-0.09	0.01	-9.90	0.94	-9.90	0.92
27	643390	0	54236	0.36	0.23	0.16	0.25	0.36	0.00	0.00	0.38	-0.19	-0.15	-0.10	0.38	0.56	0.01	-0.68	1.00	0.55	1.00
28	611055	0	54236	0.68	0.05	0.68	0.12	0.15	0.00	0.00	0.35	-0.17	0.35	-0.17	-0.20	-0.96	0.01	-9.90	0.92	-9.90	0.89
29	642372	0	54236	0.83	0.06	0.06	0.05	0.83	0.00	0.00	0.31	-0.15	-0.17	-0.18	0.31	-1.37	0.01	-9.90	0.72	-9.90	0.66
30	608250	0	54236	0.51	0.51	0.22	0.15	0.11	0.00	0.00	0.38	0.38	-0.18	-0.23	-0.10	-0.34	0.01	-9.90	0.96	-8.84	0.96
31	641193	0	54236	0.54	0.54	0.14	0.19	0.13	0.00	0.00	0.31	0.31	-0.19	-0.23	0.01	-0.19	0.01	6.86	1.02	9.44	1.04

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	641202	0	54236	0.55	0.12	0.22	0.55	0.10	0.00	0.00	0.44	-0.21	-0.25	0.44	-0.15	-0.84	0.01	-6.26	0.98	-9.90	0.93
33	609392	0	54236	0.71	0.10	0.08	0.71	0.11	0.00	0.00	0.36	-0.16	-0.21	0.36	-0.18	-1.40	0.01	-4.74	0.98	-4.41	0.96
34	608284	0	54236	0.55	0.55	0.13	0.21	0.11	0.00	0.00	0.42	0.42	-0.20	-0.22	-0.17	-0.51	0.01	-9.90	0.92	-9.90	0.91
35	642387	0	54236	0.37	0.15	0.37	0.31	0.17	0.00	0.00	0.20	-0.14	0.20	-0.04	-0.06	0.35	0.01	9.90	1.12	9.90	1.16
36	642385	0	54236	0.49	0.49	0.22	0.16	0.13	0.00	0.00	0.42	0.42	-0.17	-0.19	-0.21	-0.37	0.01	-9.90	0.94	-9.90	0.91
37	644616	0	54236	0.51	0.51	0.17	0.19	0.13	0.00	0.00	0.35	0.35	-0.19	-0.17	-0.11	-0.17	0.01	-4.37	0.99	-3.61	0.98
38	608261	0	54236	0.37	0.15	0.23	0.25	0.37	0.00	0.00	0.34	-0.06	-0.17	-0.16	0.34	0.44	0.01	3.69	1.01	7.09	1.04
39	657189	0	54236	0.49	0.15	0.20	0.49	0.14	0.01	0.00	0.29	-0.13	-0.16	0.29	-0.07	-0.08	0.01	9.90	1.05	9.90	1.05
40	641282	0	54236	0.35	0.27	0.16	0.21	0.35	0.01	0.00	0.28	-0.08	-0.15	-0.08	0.28	0.36	0.01	7.17	1.03	9.50	1.05
41	642339	0	54236	0.32	0.28	0.17	0.32	0.22	0.01	0.00	0.20	-0.06	-0.15	0.20	0.00	1.01	0.01	9.90	1.30	9.90	1.46
42	611469	0	54236	0.46	0.46	0.23	0.17	0.13	0.01	0.00	0.43	0.43	-0.19	-0.22	-0.12	0.28	0.01	-4.50	0.98	-5.77	0.97
43	642859	0	54236	0.39	0.20	0.15	0.26	0.39	0.01	0.00	0.42	-0.18	-0.20	-0.12	0.42	-0.01	0.01	-9.90	0.91	-9.90	0.89
44	642371	0	54236	0.44	0.21	0.22	0.44	0.12	0.01	0.00	0.36	-0.08	-0.18	0.36	-0.20	-0.18	0.01	-7.83	0.98	-4.17	0.98
45	642378	0	54236	0.47	0.21	0.09	0.22	0.47	0.01	0.00	0.43	-0.18	-0.19	-0.19	0.43	-0.15	0.01	-9.90	0.91	-9.90	0.89
46	607597	0	54236	0.48	0.27	0.48	0.12	0.12	0.01	0.00	0.28	-0.03	0.28	-0.22	-0.16	-0.12	0.01	9.90	1.04	9.90	1.06
47	609822	0	54236	0.50	0.18	0.50	0.16	0.15	0.01	0.00	0.33	-0.15	0.33	-0.23	-0.05	-0.30	0.01	-0.45	1.00	-0.75	1.00
48	611419	0	54236	0.52	0.21	0.15	0.52	0.12	0.01	0.00	0.44	-0.19	-0.23	0.44	-0.17	-0.41	0.01	-9.90	0.91	-9.90	0.88

Table J–4. Literature Multiple-Choice Item Statistics: Winter

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	641600	0	47149	0.31	0.17	0.21	0.31	0.31	0.00	0.00	0.25	-0.19	-0.04	-0.06	0.25	1.49	0.01	9.90	1.13	9.90	1.26
2	641594	0	47149	0.56	0.15	0.56	0.14	0.15	0.00	0.00	0.28	-0.03	0.28	-0.22	-0.14	0.40	0.01	9.90	1.12	9.90	1.15
3	641612	0	47149	0.65	0.65	0.16	0.13	0.05	0.00	0.00	0.37	0.37	-0.18	-0.20	-0.20	-0.20	0.01	-5.45	0.98	-7.30	0.95
4	641613	0	47149	0.59	0.30	0.04	0.59	0.07	0.00	0.00	0.29	-0.07	-0.22	0.29	-0.26	-0.41	0.01	9.90	1.20	9.90	1.28
5	641598	0	47149	0.43	0.15	0.43	0.37	0.05	0.00	0.00	0.23	-0.16	0.23	-0.04	-0.15	1.00	0.01	9.90	1.19	9.90	1.33
6	641595	0	47149	0.55	0.55	0.13	0.10	0.22	0.00	0.00	0.26	0.26	-0.21	-0.21	0.02	0.62	0.01	9.90	1.17	9.90	1.25
7	641599	0	47149	0.51	0.19	0.20	0.51	0.08	0.00	0.00	0.28	-0.01	-0.18	0.28	-0.21	0.49	0.01	9.90	1.12	9.90	1.17
8	641592	0	47149	0.38	0.10	0.31	0.20	0.38	0.00	0.00	0.21	-0.16	-0.03	-0.09	0.21	1.47	0.01	9.90	1.33	9.90	1.60
9	614671	0	47149	0.64	0.12	0.64	0.14	0.10	0.01	0.00	0.40	-0.20	0.40	-0.22	-0.15	-0.84	0.01	9.90	1.20	9.90	1.22
10	614662	0	47149	0.55	0.08	0.23	0.14	0.55	0.01	0.00	0.39	-0.22	-0.10	-0.25	0.39	0.09	0.01	0.17	1.00	-0.30	1.00
11	614663	0	47149	0.62	0.10	0.19	0.62	0.08	0.01	0.00	0.44	-0.28	-0.20	0.44	-0.15	-0.30	0.01	-5.20	0.98	-7.29	0.95
12	614667	0	47149	0.40	0.12	0.40	0.28	0.19	0.01	0.00	0.26	-0.21	0.26	-0.03	-0.11	0.61	0.01	9.90	1.11	9.90	1.16
13	614668	0	47149	0.64	0.64	0.10	0.14	0.12	0.01	0.00	0.38	0.38	-0.19	-0.15	-0.20	-0.57	0.01	9.90	1.10	8.76	1.08
14	614664	0	47149	0.63	0.63	0.10	0.05	0.21	0.01	0.00	0.33	0.33	-0.27	-0.26	-0.04	-0.37	0.01	9.90	1.06	9.90	1.16
15	614665	0	47149	0.74	0.05	0.10	0.11	0.74	0.01	0.00	0.50	-0.27	-0.30	-0.22	0.50	-1.29	0.01	9.90	1.09	-1.95	0.97
16	614661	0	47149	0.41	0.29	0.12	0.41	0.17	0.01	0.00	0.28	-0.12	-0.20	0.28	-0.03	0.75	0.01	9.90	1.09	9.90	1.16
17	614669	0	47149	0.72	0.09	0.10	0.08	0.72	0.01	0.00	0.49	-0.24	-0.23	-0.28	0.49	-1.07	0.01	3.80	1.02	-6.41	0.93
18	612408	0	47149	0.77	0.10	0.06	0.77	0.06	0.00	0.00	0.47	-0.26	-0.27	0.47	-0.21	-1.06	0.01	-9.90	0.88	-9.90	0.76
19	612400	0	47149	0.59	0.14	0.59	0.16	0.11	0.00	0.00	0.39	-0.23	0.39	-0.20	-0.12	0.20	0.01	-4.82	0.98	-4.29	0.98
20	612403	0	47149	0.67	0.13	0.12	0.67	0.07	0.00	0.00	0.29	-0.02	-0.25	0.29	-0.17	-0.01	0.01	5.29	1.02	9.90	1.12
21	612402	0	47149	0.41	0.17	0.28	0.13	0.41	0.00	0.00	0.25	-0.09	-0.05	-0.18	0.25	1.17	0.01	9.90	1.19	9.90	1.35
22	612406	0	47149	0.69	0.08	0.10	0.13	0.69	0.00	0.00	0.43	-0.24	-0.19	-0.22	0.43	-0.08	0.01	-9.90	0.88	-9.90	0.86
23	612405	0	47149	0.65	0.65	0.10	0.07	0.18	0.00	0.00	0.40	0.40	-0.28	-0.28	-0.09	0.05	0.01	-9.90	0.94	-9.89	0.94
24	612446	0	47149	0.37	0.37	0.36	0.18	0.09	0.00	0.00	0.32	0.32	-0.08	-0.18	-0.16	1.15	0.01	9.90	1.06	9.90	1.17
25	640207	0	47149	0.61	0.16	0.16	0.06	0.61	0.00	0.00	0.28	-0.04	-0.16	-0.26	0.28	0.10	0.01	9.90	1.07	9.90	1.19
26	612404	0	47149	0.41	0.26	0.41	0.23	0.10	0.00	0.00	0.22	-0.04	0.22	-0.06	-0.21	1.19	0.01	9.90	1.23	9.90	1.40
27	642702	0	47149	0.45	0.13	0.45	0.33	0.09	0.01	0.00	0.32	-0.16	0.32	-0.10	-0.19	0.66	0.01	9.90	1.06	9.90	1.11
28	642704	0	47149	0.53	0.13	0.53	0.15	0.18	0.01	0.00	0.42	-0.23	0.42	-0.23	-0.11	0.14	0.01	-5.98	0.98	-5.21	0.97
29	642703	0	47149	0.52	0.10	0.15	0.52	0.23	0.01	0.00	0.26	-0.15	-0.09	0.26	-0.12	0.45	0.01	9.90	1.13	9.90	1.16
30	642707	0	47149	0.48	0.11	0.26	0.48	0.15	0.01	0.00	0.38	-0.21	-0.09	0.38	-0.22	0.31	0.01	3.95	1.02	2.59	1.01
31	642706	0	47149	0.36	0.17	0.25	0.22	0.36	0.01	0.00	0.20	-0.09	-0.07	-0.06	0.20	1.54	0.01	9.90	1.32	9.90	1.56

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	642724	0	47149	0.38	0.23	0.38	0.20	0.18	0.01	0.00	0.29	-0.15	0.29	-0.12	-0.05	0.93	0.01	9.90	1.06	9.90	1.15
33	642705	0	47149	0.53	0.09	0.22	0.15	0.53	0.01	0.00	0.43	-0.16	-0.20	-0.22	0.43	0.09	0.01	-5.13	0.98	-5.02	0.97
34	642723	0	47149	0.57	0.19	0.10	0.14	0.57	0.01	0.00	0.47	-0.17	-0.26	-0.23	0.47	0.00	0.01	-9.90	0.93	-9.90	0.90

Table J–5. Algebra I Multiple-Choice Item Statistics: Spring

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	674448	0	184172	0.80	0.08	0.80	0.06	0.05	0.00	0.00	0.41	-0.22	0.41	-0.24	-0.19					-1.32	0.01	-9.90	0.85	-9.90	0.72
2	605137	0	184172	0.62	0.06	0.08	0.23	0.62	0.00	0.00	0.47	-0.26	-0.23	-0.24	0.47					-0.14	0.01	-9.90	0.87	-9.90	0.82
3	605189	0	184172	0.49	0.18	0.49	0.23	0.09	0.00	0.00	0.23	-0.14	0.23	-0.09	-0.08					0.64	0.01	9.90	1.19	9.90	1.24
4	633250	0	184172	0.68	0.68	0.22	0.07	0.04	0.00	0.00	0.38	0.38	-0.21	-0.22	-0.18					-0.48	0.01	-5.71	0.97	-5.87	0.94
5	674148	0	184172	0.49	0.28	0.15	0.49	0.07	0.00	0.00	0.35	-0.14	-0.21	0.35	-0.14					0.61	0.01	8.27	1.03	8.25	1.05
6	632148	0	184172	0.74	0.74	0.15	0.07	0.04	0.00	0.00	0.47	0.47	-0.30	-0.25	-0.19					-0.74	0.01	-9.90	0.75	-9.90	0.63
7	674491	0	184172	0.52	0.13	0.10	0.25	0.52	0.00	0.00	0.24	-0.07	-0.11	-0.15	0.24					0.65	0.01	9.90	1.17	9.90	1.22
8	674455	0	184172	0.51	0.10	0.22	0.51	0.17	0.00	0.00	0.24	-0.08	-0.13	0.24	-0.11					0.67	0.01	9.90	1.17	9.90	1.23
9	631599	0	184172	0.73	0.10	0.73	0.10	0.06	0.00	0.00	0.44	-0.24	0.44	-0.25	-0.18					-0.72	0.01	-9.90	0.83	-9.90	0.71
10	640609	0	184172	0.41	0.27	0.41	0.17	0.14	0.01	0.00	0.34	-0.06	0.34	-0.20	-0.16					0.76	0.01	9.90	1.04	9.90	1.07
11	657749	0	184172	0.65	0.65	0.10	0.18	0.07	0.01	0.00	0.52	0.52	-0.24	-0.31	-0.22					-0.57	0.01	-9.90	0.88	-9.90	0.79
12	605108	0	184172	0.60	0.13	0.10	0.60	0.17	0.01	0.00	0.44	-0.16	-0.23	0.44	-0.24					-0.15	0.01	-9.90	0.94	-9.90	0.91
13	640576	0	184172	0.45	0.26	0.17	0.12	0.45	0.01	0.00	0.53	-0.26	-0.17	-0.25	0.53					0.38	0.01	-9.90	0.87	-9.90	0.84
14	674511	0	184172	0.65	0.10	0.65	0.09	0.15	0.01	0.00	0.41	-0.18	0.41	-0.27	-0.16					-0.54	0.01	2.60	1.01	3.25	1.03
15	674419	0	184172	0.53	0.16	0.16	0.53	0.14	0.01	0.00	0.31	-0.04	-0.19	0.31	-0.19					0.21	0.01	9.90	1.12	9.90	1.17
16	605172	0	184172	0.42	0.42	0.30	0.13	0.14	0.01	0.00	0.35	0.35	-0.17	-0.16	-0.10					0.78	0.01	9.61	1.04	9.90	1.06
17	605286	0	184172	0.35	0.16	0.34	0.15	0.35	0.01	0.00	0.39	-0.14	-0.14	-0.17	0.39					1.36	0.01	1.95	1.01	8.73	1.06
18	674384	0	184172	0.51	0.24	0.12	0.51	0.12	0.01	0.00	0.37	-0.13	-0.26	0.37	-0.12					0.31	0.01	9.90	1.05	9.90	1.07
19	678779	0	184172	0.85	0.04	0.85	0.09	0.02	0.00	0.00	0.36	-0.14	0.36	-0.28	-0.16					-1.55	0.01	-9.90	0.80	-9.90	0.70
20	682132	0	184172	0.43	0.43	0.20	0.18	0.20	0.00	0.00	0.23	0.23	-0.17	-0.10	-0.01					0.82	0.01	9.90	1.19	9.90	1.26
21	678752	0	184172	0.68	0.13	0.10	0.09	0.68	0.00	0.00	0.46	-0.24	-0.22	-0.23	0.46					-0.66	0.01	-9.90	0.92	-9.90	0.85
22	678731	0	184172	0.42	0.42	0.19	0.31	0.08	0.00	0.00	0.26	0.26	-0.16	-0.06	-0.13					0.41	0.01	9.90	1.21	9.90	1.30
23	605110	0	184172	0.73	0.06	0.12	0.09	0.73	0.00	0.00	0.49	-0.25	-0.27	-0.24	0.49					-0.62	0.01	-9.90	0.76	-9.90	0.65
24	678806	0	184172	0.45	0.10	0.45	0.25	0.20	0.00	0.00	0.35	-0.09	0.35	-0.13	-0.23					0.77	0.01	9.28	1.04	9.90	1.06
25	632358	0	184172	0.74	0.74	0.10	0.08	0.09	0.00	0.00	0.40	0.40	-0.23	-0.19	-0.20					-0.47	0.01	-9.90	0.82	-9.90	0.76
26	633251	0	184172	0.64	0.09	0.14	0.64	0.13	0.00	0.00	0.30	-0.20	-0.07	0.30	-0.18					-0.51	0.01	9.90	1.19	9.90	1.32
27	605301	0	184172	0.72	0.07	0.13	0.72	0.09	0.00	0.00	0.42	-0.22	-0.27	0.42	-0.15					-0.87	0.01	-8.52	0.95	-5.23	0.94
28	678772	0	184172	0.73	0.73	0.16	0.06	0.04	0.00	0.00	0.42	0.42	-0.25	-0.22	-0.18					-1.24	0.01	9.90	1.12	9.34	1.15
29	605245	0	184172	0.58	0.17	0.58	0.14	0.10	0.00	0.00	0.36	-0.18	0.36	-0.18	-0.13					0.22	0.01	5.39	1.02	6.14	1.04
30	641484	0	184172	0.28	0.40	0.18	0.14	0.28	0.00	0.00	0.30	0.03	-0.18	-0.22	0.30					1.63	0.01	9.63	1.05	9.90	1.16
31	633531	0	184172	0.57	0.09	0.57	0.23	0.10	0.00	0.00	0.45	-0.19	0.45	-0.25	-0.18					0.23	0.01	-9.90	0.91	-9.90	0.86

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	678814	0	184172	0.54	0.10	0.27	0.09	0.54	0.00	0.00	0.22	-0.15	0.00	-0.22	0.22					0.08	0.01	9.90	1.27	9.90	1.35
33	605101	0	184172	0.71	0.71	0.17	0.06	0.05	0.00	0.00	0.44	0.44	-0.26	-0.20	-0.22					-1.06	0.01	6.14	1.04	0.85	1.01
34	673951	0	184172	0.63	0.08	0.22	0.63	0.06	0.00	0.00	0.44	-0.23	-0.23	0.44	-0.21					-0.65	0.01	9.90	1.07	-1.41	0.98
35	678774	0	184172	0.53	0.19	0.53	0.15	0.13	0.00	0.00	0.44	-0.23	0.44	-0.22	-0.15					0.52	0.01	-9.90	0.94	-9.90	0.92
36	678785	0	184172	0.62	0.06	0.14	0.62	0.18	0.00	0.00	0.42	-0.24	-0.23	0.42	-0.17					-0.20	0.01	-9.36	0.96	-9.48	0.92
37	700883	1	5016	0.80	0.80	0.04	0.12	0.03	0.00	0.00	0.48	0.48	-0.19	-0.35	-0.20	A-	C-	B-	A+	-1.06	0.06	-4.63	0.85	-4.46	0.72
38	700847	1	5016	0.32	0.13	0.20	0.34	0.32	0.00	0.00	0.24	-0.12	0.00	-0.15	0.24	A+	A-	A+	A-	1.61	0.05	5.80	1.15	8.27	1.39
39	700881	1	5016	0.46	0.13	0.23	0.18	0.46	0.00	0.00	0.44	-0.10	-0.29	-0.17	0.44	A-	A-	A-	A+	0.86	0.05	-1.06	0.98	0.45	1.01
40	681998	1	5016	0.41	0.41	0.24	0.09	0.25	0.00	0.00	0.38	0.38	-0.11	-0.16	-0.21	A+	A+	A-	A-	1.10	0.05	2.91	1.06	5.34	1.19
41	674394	1	5016	0.51	0.18	0.07	0.23	0.51	0.00	0.00	0.40	-0.16	-0.18	-0.22	0.40	A-	A-	B-	A+	0.59	0.05	1.87	1.04	1.16	1.04
42	702567	1	5016	0.75	0.07	0.75	0.13	0.05	0.00	0.00	0.46	-0.28	0.46	-0.24	-0.22	B+	A-	A+	A+	-0.71	0.06	-5.61	0.85	-4.49	0.76
43	704022	1	5016	0.78	0.78	0.08	0.08	0.05	0.00	0.00	0.53	0.53	-0.29	-0.28	-0.27	A+	A-	A+	A+	-0.93	0.06	-7.44	0.78	-7.27	0.60
44	703354	1	5016	0.41	0.41	0.21	0.27	0.11	0.00	0.00	0.21	0.21	-0.24	0.07	-0.10	A-	A-	A-	A+	1.14	0.05	9.90	1.24	9.90	1.41
45	702491	1	5016	0.68	0.05	0.20	0.68	0.06	0.00	0.00	0.46	-0.25	-0.27	0.46	-0.21	A+	A+	A+	A+	-0.29	0.05	-2.90	0.93	-1.91	0.91
46	703979	1	5016	0.70	0.70	0.10	0.14	0.06	0.00	0.00	0.48	0.48	-0.29	-0.22	-0.22	A-	A-	A-	A+	-0.39	0.05	-4.13	0.90	-3.61	0.83
47	704000	2	4964	0.40	0.22	0.20	0.18	0.40	0.00	0.00	0.21	-0.12	-0.01	-0.13	0.21	A-	A-	A-	A+	1.21	0.05	8.95	1.20	8.45	1.30
48	700870	2	4964	0.81	0.06	0.06	0.81	0.06	0.00	0.00	0.42	-0.23	-0.20	0.42	-0.23	A+	A-	A-	A+	-1.08	0.06	-2.34	0.92	-0.87	0.94
49	700819	2	4964	0.65	0.11	0.17	0.65	0.07	0.00	0.00	0.30	-0.16	-0.17	0.30	-0.11	A+	A+	A+	A+	-0.07	0.05	5.19	1.12	4.67	1.19
50	674514	2	4964	0.67	0.67	0.12	0.15	0.06	0.00	0.00	0.51	0.51	-0.27	-0.29	-0.20	B-	A-	A-	A+	-0.20	0.05	-4.98	0.89	-4.74	0.82
51	666549	2	4964	0.29	0.22	0.32	0.17	0.29	0.00	0.00	0.16	-0.06	-0.01	-0.11	0.16	A+	A+	A+	A-	1.83	0.05	5.40	1.15	8.23	1.42
52	702477	2	4964	0.33	0.22	0.25	0.20	0.33	0.00	0.00	0.35	-0.11	-0.15	-0.13	0.35	A+	A-	A-	A+	1.55	0.05	3.61	1.09	7.69	1.33
53	702522	2	4964	0.32	0.31	0.25	0.32	0.13	0.00	0.00	0.19	-0.09	-0.05	0.19	-0.07	B-	A-	A-	A+	1.66	0.05	6.05	1.15	7.86	1.36
54	702539	2	4964	0.53	0.22	0.53	0.17	0.08	0.00	0.00	0.44	-0.29	0.44	-0.16	-0.14	A-	A+	A+	A+	0.54	0.05	-1.60	0.97	-0.86	0.97
55	702511	2	4964	0.70	0.07	0.70	0.16	0.07	0.00	0.00	0.41	-0.29	0.41	-0.16	-0.20	A+	A-	A-	A-	-0.33	0.05	-2.00	0.95	-1.25	0.95
56	702531	2	4964	0.36	0.25	0.16	0.36	0.23	0.00	0.00	0.22	-0.03	-0.19	0.22	-0.05	A+	A-	A-	A+	1.43	0.05	8.18	1.19	9.82	1.40
57	703999	3	4982	0.43	0.43	0.13	0.32	0.12	0.00	0.00	0.43	0.43	-0.22	-0.11	-0.28	A-	A-	A-	A+	1.09	0.05	-1.46	0.97	0.22	1.01
58	703973	3	4982	0.13	0.36	0.23	0.27	0.13	0.00	0.00	0.22	-0.12	-0.10	0.07	0.22	B-	A-	A+	A+	3.06	0.07	1.51	1.07	6.27	1.69
59	700810	3	4982	0.33	0.24	0.28	0.14	0.33	0.00	0.00	0.48	-0.26	-0.18	-0.10	0.48	A+	A-	A-	A-	1.59	0.05	-3.71	0.92	-2.13	0.92
60	696821	3	4982	0.35	0.35	0.30	0.17	0.18	0.00	0.00	0.21	0.21	-0.02	-0.15	-0.07	A-	A+	A-	A-	1.51	0.05	9.90	1.25	9.90	1.54
61	700800	3	4982	0.16	0.16	0.33	0.26	0.25	0.01	0.00	0.06	0.06	0.23	-0.19	-0.10	A-	A-	A+	A-	2.78	0.07	4.16	1.18	9.90	2.15
62	702554	3	4982	0.45	0.34	0.45	0.17	0.04	0.00	0.00	0.26	-0.04	0.26	-0.18	-0.22	A-	A-	A-	A+	0.98	0.05	8.04	1.17	7.15	1.23
63	678805	3	4982	0.35	0.27	0.22	0.35	0.16	0.00	0.00	0.24	0.06	-0.20	0.24	-0.15	A-	A+	A-	A+	1.53	0.05	7.02	1.17	8.64	1.36

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
64	702546	3	4982	0.34	0.34	0.44	0.16	0.07	0.00	0.00	0.23	0.23	0.14	-0.33	-0.23	A-	A-	A-	A+	1.58	0.05	6.66	1.16	8.50	1.37
65	696806	3	4982	0.73	0.12	0.73	0.09	0.06	0.00	0.00	0.46	-0.26	0.46	-0.27	-0.17	A+	A-	A-	A+	-0.49	0.06	-3.32	0.91	-2.72	0.87
66	702482	3	4982	0.65	0.06	0.20	0.65	0.09	0.00	0.00	0.51	-0.28	-0.32	0.51	-0.16	A+	A+	A+	A+	-0.05	0.05	-5.21	0.88	-3.15	0.88
67	700857	4	5046	0.33	0.33	0.22	0.24	0.20	0.00	0.00	-0.03	-0.03	0.05	-0.06	0.06	A+	A+	A+	A+	1.57	0.05	9.90	1.49	9.90	2.02
68	700774	4	5046	0.16	0.29	0.20	0.16	0.34	0.00	0.00	-0.11	-0.06	0.02	-0.11	0.14	A-	A+	A-	A+	2.68	0.07	9.90	1.44	9.90	2.76
69	700820	4	5046	0.80	0.14	0.04	0.80	0.03	0.00	0.00	0.46	-0.32	-0.25	0.46	-0.16	A+	A-	A-	A-	-1.02	0.06	-3.78	0.88	-4.30	0.74
70	700879	4	5046	0.40	0.40	0.24	0.14	0.21	0.00	0.00	0.41	0.41	-0.10	-0.20	-0.22	A+	A-	A+	A+	1.16	0.05	1.40	1.03	2.29	1.07
71	700801	4	5046	0.30	0.15	0.30	0.42	0.12	0.00	0.00	0.28	0.02	0.28	-0.14	-0.19	A-	A-	A-	A+	1.74	0.05	3.86	1.10	8.09	1.39
72	702467	4	5046	0.90	0.04	0.90	0.02	0.03	0.00	0.00	0.37	-0.23	0.37	-0.18	-0.19	A-	C-	B-	A+	-1.93	0.08	-2.29	0.88	-3.69	0.65
73	703172	4	5046	0.35	0.09	0.35	0.21	0.35	0.00	0.00	0.38	-0.03	0.38	-0.11	-0.27	A+	A-	A+	A+	1.47	0.05	1.30	1.03	4.23	1.16
74	704023	4	5046	0.75	0.11	0.05	0.75	0.08	0.00	0.00	0.47	-0.29	-0.27	0.47	-0.19	A+	A-	A+	A+	-0.70	0.06	-3.75	0.90	-3.19	0.84
75	702563	4	5046	0.53	0.06	0.14	0.27	0.53	0.00	0.00	0.47	-0.26	-0.24	-0.20	0.47	A-	A+	A-	A-	0.50	0.05	-1.25	0.98	-1.14	0.97
76	702493	4	5046	0.49	0.19	0.15	0.17	0.49	0.00	0.00	0.40	-0.15	-0.15	-0.22	0.40	A+	A-	A-	A+	0.72	0.05	0.88	1.02	0.68	1.02
77	704018	5	5021	0.69	0.17	0.69	0.10	0.04	0.00	0.00	0.41	-0.23	0.41	-0.22	-0.19	A-	A-	A-	A+	-0.35	0.05	-0.60	0.98	-0.72	0.97
78	703177	5	5021	0.40	0.12	0.40	0.19	0.28	0.00	0.00	0.39	-0.15	0.39	-0.20	-0.13	A+	A-	A-	A+	1.20	0.05	1.64	1.04	4.14	1.14
79	703346	5	5021	0.59	0.08	0.08	0.24	0.59	0.00	0.00	0.41	-0.23	-0.26	-0.15	0.41	A+	A-	A-	A+	0.21	0.05	0.77	1.02	0.82	1.03
80	700791	5	5021	0.51	0.24	0.16	0.51	0.09	0.00	0.00	0.41	-0.14	-0.22	0.41	-0.23	A-	A-	A+	A+	0.62	0.05	2.35	1.05	1.38	1.04
81	700882	5	5021	0.26	0.26	0.26	0.22	0.26	0.00	0.00	0.24	-0.23	0.04	-0.05	0.24	A-	A-	A-	A-	2.02	0.06	3.23	1.09	6.11	1.34
82	702568	5	5021	0.20	0.09	0.07	0.64	0.20	0.00	0.00	0.35	-0.09	-0.14	-0.17	0.35	A-	A+	A+	A-	2.41	0.06	-0.26	0.99	5.09	1.36
83	702474	5	5021	0.43	0.09	0.09	0.43	0.38	0.00	0.00	0.31	-0.27	-0.27	0.31	0.01	A-	B-	A-	A+	1.05	0.05	6.93	1.15	6.18	1.20
84	702532	5	5021	0.26	0.24	0.31	0.19	0.26	0.00	0.00	0.28	-0.18	-0.10	0.00	0.28	A-	A+	A+	A+	1.98	0.06	3.78	1.11	7.30	1.40
85	704024	5	5021	0.55	0.06	0.55	0.23	0.15	0.00	0.00	0.54	-0.17	0.54	-0.31	-0.27	A+	A-	A+	A+	0.42	0.05	-6.04	0.88	-5.42	0.84
86	702502	5	5021	0.38	0.30	0.16	0.38	0.16	0.00	0.00	0.47	-0.18	-0.19	0.47	-0.19	A-	A-	A+	A+	1.33	0.05	-2.69	0.94	0.62	1.02
87	700829	6	5021	0.23	0.21	0.23	0.32	0.23	0.01	0.00	0.17	-0.11	0.10	-0.13	0.17	A-	A+	A-	A-	2.22	0.06	3.82	1.12	9.32	1.60
88	700818	6	5021	0.44	0.16	0.44	0.22	0.18	0.00	0.00	0.29	-0.20	0.29	0.00	-0.19	A-	A-	A-	A+	1.00	0.05	6.18	1.13	6.05	1.18
89	674417	6	5021	0.61	0.16	0.14	0.61	0.09	0.00	0.00	0.38	-0.28	-0.13	0.38	-0.13	A+	A-	A-	A-	0.15	0.05	2.54	1.05	1.76	1.06
90	700848	6	5021	0.34	0.20	0.20	0.26	0.34	0.00	0.00	0.18	-0.08	-0.10	-0.02	0.18	A-	A-	A-	A-	1.56	0.05	9.90	1.25	9.90	1.49
91	700779	6	5021	0.75	0.75	0.08	0.07	0.10	0.00	0.00	0.45	0.45	-0.22	-0.26	-0.22	A-	A-	A-	A+	-0.64	0.06	-3.69	0.90	-2.73	0.86
92	703345	6	5021	0.37	0.12	0.37	0.04	0.47	0.00	0.00	0.40	-0.12	0.40	-0.21	-0.22	A-	A-	A+	A+	1.37	0.05	-1.36	0.97	1.24	1.04
93	702550	6	5021	0.57	0.13	0.22	0.57	0.08	0.00	0.00	0.45	-0.15	-0.27	0.45	-0.23	A-	A+	A-	A+	0.34	0.05	-2.51	0.95	-2.55	0.92
94	703353	6	5021	0.63	0.26	0.08	0.63	0.02	0.00	0.00	0.33	-0.15	-0.23	0.33	-0.19	A-	A-	A-	A+	0.05	0.05	4.18	1.09	4.02	1.15
95	703169	6	5021	0.38	0.35	0.38	0.06	0.21	0.00	0.00	0.24	0.00	0.24	-0.19	-0.16	A-	A+	A+	A-	1.33	0.05	6.72	1.15	7.94	1.28

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
96	702462	6	5021	0.38	0.13	0.16	0.32	0.38	0.01	0.00	0.41	-0.16	-0.11	-0.21	0.41	A+	A+	A-	A-	1.31	0.05	-0.73	0.98	2.23	1.07
97	700822	7	5008	0.52	0.26	0.52	0.16	0.06	0.00	0.00	0.33	-0.20	0.33	-0.15	-0.10	B-	A-	A-	A+	0.60	0.05	3.92	1.08	3.52	1.10
98	674503	7	5008	0.36	0.19	0.36	0.18	0.27	0.00	0.00	0.23	0.08	0.23	-0.12	-0.20	A+	A+	A-	A+	1.38	0.05	6.56	1.15	6.69	1.25
99	700812	7	5008	0.55	0.55	0.22	0.13	0.09	0.00	0.00	0.34	0.34	-0.22	-0.15	-0.10	A-	A+	A-	A+	0.41	0.05	3.33	1.07	2.79	1.08
100	700863	7	5008	0.36	0.23	0.17	0.23	0.36	0.00	0.00	0.26	-0.07	-0.13	-0.10	0.26	A+	A+	A+	A+	1.39	0.05	6.74	1.15	7.47	1.28
101	700777	7	5008	0.32	0.38	0.32	0.19	0.10	0.00	0.00	0.14	0.17	0.14	-0.22	-0.20	A+	A-	A+	A+	1.60	0.05	9.90	1.28	9.90	1.56
102	702498	7	5008	0.74	0.74	0.10	0.12	0.03	0.00	0.00	0.34	0.34	-0.26	-0.10	-0.19	A-	A-	A-	A-	-0.63	0.06	1.10	1.03	4.42	1.24
103	702564	7	5008	0.33	0.24	0.33	0.18	0.25	0.00	0.00	0.13	-0.14	0.13	-0.12	0.11	A-	A-	A+	A-	1.55	0.05	9.72	1.24	9.90	1.43
104	702527	7	5008	0.34	0.05	0.34	0.37	0.24	0.00	0.00	0.40	-0.19	0.40	-0.19	-0.12	A-	A+	A-	A-	1.50	0.05	-1.87	0.96	3.01	1.11
105	702470	7	5008	0.59	0.59	0.14	0.13	0.15	0.00	0.00	0.47	0.47	-0.22	-0.22	-0.23	A-	A+	A+	A-	0.25	0.05	-3.69	0.93	-3.43	0.90
106	702534	7	5008	0.26	0.15	0.08	0.26	0.50	0.00	0.00	0.27	-0.35	-0.07	0.27	0.06	A-	A+	A-	A+	1.94	0.06	0.85	1.02	5.02	1.26
107	681816	8	4876	0.31	0.23	0.31	0.33	0.12	0.00	0.00	0.23	-0.06	0.23	-0.06	-0.15	A-	A+	A-	A+	1.70	0.05	5.36	1.13	9.10	1.41
108	700845	8	4876	0.41	0.41	0.31	0.17	0.11	0.00	0.00	0.23	0.23	0.02	-0.20	-0.14	A-	A-	A+	A+	1.17	0.05	8.34	1.18	8.30	1.28
109	700854	8	4876	0.37	0.33	0.37	0.14	0.15	0.00	0.00	0.05	0.27	0.05	-0.21	-0.21	A+	A+	A-	A+	1.36	0.05	9.90	1.41	9.90	1.66
110	700853	8	4876	0.29	0.29	0.25	0.20	0.25	0.00	0.00	0.29	0.29	-0.15	-0.08	-0.08	A+	A-	A+	A+	1.81	0.05	3.00	1.08	7.57	1.36
111	696817	8	4876	0.39	0.09	0.28	0.39	0.24	0.00	0.00	0.22	-0.21	-0.13	0.22	0.03	A-	A+	A+	A+	1.29	0.05	8.59	1.19	9.60	1.35
112	696818	8	4876	0.40	0.16	0.40	0.17	0.27	0.00	0.00	0.31	-0.09	0.31	-0.19	-0.09	A+	A-	A+	A+	1.24	0.05	3.54	1.07	4.47	1.15
113	702481	8	4876	0.61	0.11	0.13	0.15	0.61	0.00	0.00	0.25	-0.12	-0.04	-0.19	0.25	A-	A-	A-	A+	0.17	0.05	7.63	1.17	5.71	1.21
114	702552	8	4876	0.51	0.18	0.16	0.51	0.15	0.00	0.00	0.18	-0.15	-0.17	0.18	0.08	A+	A+	A-	A+	0.69	0.05	9.90	1.24	9.90	1.35
115	702488	8	4876	0.10	0.70	0.11	0.09	0.10	0.00	0.00	0.13	0.07	-0.13	-0.10	0.13	A-	A-	A-	A+	3.35	0.08	-1.02	0.94	6.58	1.85
116	702551	8	4876	0.60	0.60	0.14	0.15	0.10	0.00	0.00	0.46	0.46	-0.24	-0.22	-0.20	A-	A-	A-	A+	0.20	0.05	-3.49	0.93	-3.41	0.89
117	700831	9	5037	0.35	0.26	0.26	0.35	0.12	0.00	0.00	0.22	-0.03	-0.10	0.22	-0.14	A-	A-	A-	A-	1.44	0.05	6.32	1.14	7.74	1.28
118	700789	9	5037	0.40	0.12	0.24	0.24	0.40	0.00	0.00	0.24	-0.07	-0.13	-0.09	0.24	A-	A-	A-	A-	1.19	0.05	5.97	1.12	6.12	1.19
119	673945	9	5037	0.42	0.37	0.10	0.42	0.10	0.00	0.00	0.04	0.20	-0.19	0.04	-0.19	A-	A+	A+	A+	1.09	0.05	9.90	1.35	9.90	1.48
120	700793	9	5037	0.31	0.20	0.31	0.40	0.10	0.00	0.00	0.20	-0.06	0.20	-0.03	-0.16	A-	A-	A-	A+	1.68	0.05	7.88	1.20	9.90	1.46
121	700871	9	5037	0.43	0.43	0.24	0.23	0.10	0.00	0.00	0.05	0.05	-0.14	0.11	-0.04	A+	A-	A-	A+	1.04	0.05	9.90	1.38	9.90	1.52
122	702557	9	5037	0.31	0.14	0.27	0.28	0.31	0.00	0.00	0.18	-0.17	-0.01	-0.04	0.18	A-	A-	A-	A-	1.66	0.05	4.41	1.11	5.61	1.23
123	702553	9	5037	0.35	0.15	0.35	0.35	0.14	0.00	0.00	0.15	-0.13	0.04	0.15	-0.12	A+	A+	A+	A+	1.42	0.05	9.10	1.21	9.70	1.35
124	666555	9	5037	0.40	0.18	0.40	0.14	0.28	0.00	0.00	0.09	-0.07	0.09	-0.07	0.02	A+	A-	A-	A-	1.19	0.05	9.90	1.26	9.90	1.40
125	704026	9	5037	0.48	0.35	0.48	0.10	0.06	0.00	0.00	0.34	-0.09	0.34	-0.20	-0.24	A-	A-	A-	A-	0.76	0.05	3.32	1.06	3.06	1.08
126	702468	9	5037	0.49	0.49	0.18	0.19	0.14	0.01	0.00	0.33	0.33	-0.15	-0.14	-0.15	A-	A+	A-	A+	0.76	0.05	3.00	1.05	2.98	1.08
127	703974	10	5027	0.58	0.11	0.23	0.08	0.58	0.00	0.00	0.52	-0.18	-0.30	-0.27	0.52	A+	A-	A-	A+	0.30	0.05	-5.64	0.89	-4.97	0.86

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
128	700797	10	5027	0.47	0.16	0.20	0.47	0.17	0.00	0.00	0.11	0.01	-0.06	0.11	-0.09	A+	A-	A+	A+	0.83	0.05	9.90	1.34	9.90	1.49
129	700824	10	5027	0.44	0.44	0.27	0.18	0.11	0.00	0.00	0.39	0.39	-0.20	-0.11	-0.20	A-	A-	A-	A+	0.98	0.05	0.38	1.01	1.64	1.05
130	674476	10	5027	0.53	0.07	0.09	0.31	0.53	0.00	0.00	0.09	-0.07	-0.07	-0.02	0.09	A+	A-	A-	A+	0.56	0.05	9.90	1.38	9.90	1.56
131	700782	10	5027	0.36	0.19	0.36	0.36	0.09	0.00	0.00	0.23	-0.17	0.23	0.00	-0.15	A-	A-	A-	A-	1.42	0.05	7.89	1.18	9.87	1.39
132	678783	10	5027	0.32	0.36	0.09	0.23	0.32	0.00	0.00	0.30	-0.21	-0.12	-0.02	0.30	A-	A-	A-	A-	1.65	0.05	2.81	1.07	4.94	1.21
133	702461	10	5027	0.43	0.10	0.28	0.43	0.19	0.00	0.00	0.34	-0.12	-0.16	0.34	-0.15	A+	A-	A-	A+	1.03	0.05	4.87	1.10	5.80	1.18
134	674152	10	5027	0.77	0.10	0.77	0.05	0.08	0.00	0.00	0.47	-0.28	0.47	-0.16	-0.28	A+	A+	A-	A+	-0.79	0.06	-5.30	0.85	-4.90	0.75
135	696810	10	5027	0.63	0.14	0.63	0.10	0.13	0.00	0.00	0.50	-0.20	0.50	-0.22	-0.31	A+	A+	A-	A+	0.04	0.05	-5.68	0.88	-5.19	0.83
136	702544	10	5027	0.27	0.27	0.64	0.05	0.03	0.00	0.00	0.35	0.35	-0.15	-0.20	-0.20	A-	A-	A-	A+	1.93	0.06	0.52	1.01	5.35	1.28
137	703176	11	4991	0.34	0.34	0.25	0.30	0.11	0.00	0.00	0.16	0.16	-0.12	-0.01	-0.06	A-	A-	A-	A-	1.53	0.05	7.69	1.18	9.90	1.40
138	674480	11	4991	0.46	0.12	0.30	0.46	0.12	0.00	0.00	0.24	-0.21	0.02	0.24	-0.19	A+	A-	A-	A+	0.93	0.05	8.57	1.17	8.02	1.24
139	703350	11	4991	0.44	0.16	0.18	0.44	0.23	0.00	0.00	0.36	-0.20	-0.16	0.36	-0.11	A-	A-	A-	B+	1.02	0.05	3.28	1.06	4.14	1.12
140	700840	11	4991	0.35	0.35	0.23	0.28	0.14	0.00	0.00	0.24	0.24	-0.15	-0.01	-0.12	A-	A+	A-	A+	1.47	0.05	5.48	1.12	8.48	1.32
141	700830	11	4991	0.25	0.45	0.14	0.25	0.15	0.00	0.00	0.22	-0.04	-0.06	0.22	-0.14	A-	A-	A-	A-	2.06	0.06	3.24	1.09	8.88	1.50
142	704002	11	4991	0.50	0.21	0.50	0.22	0.07	0.00	0.00	0.35	-0.09	0.35	-0.24	-0.16	A-	A+	A+	A-	0.69	0.05	2.00	1.04	1.49	1.04
143	696809	11	4991	0.40	0.14	0.34	0.40	0.11	0.00	0.00	0.22	-0.11	-0.06	0.22	-0.13	A+	A+	A-	A+	1.19	0.05	7.62	1.15	8.71	1.28
144	657752	11	4991	0.72	0.07	0.72	0.07	0.13	0.00	0.00	0.45	-0.13	0.45	-0.24	-0.30	A+	A-	A-	A+	-0.47	0.05	-4.04	0.90	-3.08	0.86
145	696820	11	4991	0.46	0.11	0.22	0.21	0.46	0.00	0.00	0.29	-0.13	-0.05	-0.21	0.29	A-	A+	A-	A+	0.89	0.05	6.17	1.12	5.94	1.17
146	702503	11	4991	0.33	0.24	0.33	0.28	0.15	0.00	0.00	0.09	-0.16	0.09	0.10	-0.04	A-	A-	A-	A+	1.61	0.05	9.90	1.30	9.90	1.53
147	703971	12	5003	0.60	0.29	0.04	0.60	0.08	0.00	0.00	0.38	-0.19	-0.18	0.38	-0.24	A+	A-	A-	A+	0.23	0.05	2.53	1.05	1.46	1.05
148	703998	12	5003	0.31	0.31	0.13	0.17	0.40	0.00	0.00	0.45	0.45	-0.15	-0.28	-0.10	B-	A-	A-	A+	1.72	0.05	-3.27	0.92	0.90	1.04
149	700814	12	5003	0.40	0.20	0.40	0.18	0.22	0.00	0.00	0.20	-0.13	0.20	-0.05	-0.06	A-	A-	A-	A-	1.22	0.05	9.90	1.25	9.90	1.41
150	703975	12	5003	0.39	0.39	0.24	0.21	0.16	0.00	0.00	0.29	0.29	-0.01	-0.23	-0.11	A-	A-	A-	A+	1.26	0.05	5.51	1.12	6.97	1.24
151	696807	12	5003	0.31	0.17	0.37	0.14	0.31	0.00	0.00	0.35	-0.23	-0.06	-0.12	0.35	A-	A+	A+	A-	1.71	0.05	1.28	1.03	4.70	1.21
152	702476	12	5003	0.45	0.13	0.26	0.15	0.45	0.00	0.00	0.41	-0.20	-0.15	-0.19	0.41	A+	A+	A+	A+	0.95	0.05	-0.49	0.99	-0.23	0.99
153	702562	12	5003	0.32	0.32	0.33	0.18	0.17	0.00	0.00	0.30	0.30	-0.10	-0.15	-0.10	A-	A-	A-	A-	1.68	0.05	3.29	1.08	6.22	1.27
154	702547	12	5003	0.36	0.36	0.15	0.24	0.25	0.00	0.00	0.31	0.31	-0.16	0.01	-0.21	A+	A-	A+	A-	1.44	0.05	3.87	1.09	5.32	1.20
155	702528	12	5003	0.67	0.11	0.11	0.67	0.11	0.00	0.00	0.38	-0.30	-0.21	0.38	-0.06	A-	A-	A-	A+	-0.16	0.05	-0.33	0.99	0.33	1.01
156	702535	12	5003	0.68	0.10	0.13	0.68	0.09	0.00	0.00	0.43	-0.25	-0.20	0.43	-0.19	A+	A+	A+	A+	-0.18	0.05	-1.77	0.96	-1.79	0.93
157	700839	13	5042	0.70	0.12	0.08	0.70	0.10	0.00	0.00	0.41	-0.21	-0.27	0.41	-0.16	A+	A-	A+	A+	-0.33	0.05	-0.42	0.99	-0.97	0.96
158	674484	13	5042	0.14	0.25	0.48	0.13	0.14	0.00	0.00	0.11	-0.03	0.03	-0.12	0.11	A+	A+	A+	A-	2.93	0.07	1.91	1.09	8.77	1.91
159	674517	13	5042	0.34	0.24	0.34	0.29	0.13	0.00	0.00	0.26	-0.06	0.26	-0.08	-0.18	A-	A-	A-	A+	1.55	0.05	4.88	1.12	6.63	1.26

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
160	700846	13	5042	0.27	0.24	0.19	0.29	0.27	0.00	0.00	0.21	0.00	-0.14	-0.08	0.21	A-	A+	A-	A-	1.92	0.06	4.95	1.14	8.82	1.46
161	703175	13	5042	0.58	0.18	0.58	0.08	0.16	0.00	0.00	0.35	-0.15	0.35	-0.21	-0.15	A-	A-	A-	A+	0.32	0.05	2.91	1.06	1.42	1.04
162	702566	13	5042	0.19	0.41	0.19	0.23	0.17	0.00	0.00	-0.01	0.12	-0.01	-0.14	0.01	A-	A-	A+	A+	2.49	0.06	7.30	1.28	9.90	2.18
163	696815	13	5042	0.50	0.13	0.50	0.22	0.14	0.00	0.00	0.38	-0.12	0.38	-0.23	-0.13	B-	A-	A-	A-	0.70	0.05	2.97	1.06	1.68	1.05
164	702536	13	5042	0.43	0.43	0.17	0.23	0.16	0.00	0.00	0.23	0.23	-0.10	-0.02	-0.17	A-	A+	A+	A+	1.06	0.05	8.76	1.18	8.48	1.27
165	703977	13	5042	0.64	0.10	0.09	0.17	0.64	0.00	0.00	0.47	-0.19	-0.26	-0.25	0.47	A+	A-	A-	A-	0.01	0.05	-2.50	0.95	-2.72	0.91
166	704004	13	5042	0.58	0.58	0.14	0.12	0.16	0.01	0.00	0.48	0.48	-0.19	-0.23	-0.26	A+	A+	A+	A-	0.33	0.05	-2.95	0.94	-2.58	0.92
167	704017	14	5026	0.86	0.09	0.86	0.03	0.02	0.00	0.00	0.40	-0.27	0.40	-0.19	-0.19	A-	A-	A-	A+	-1.50	0.07	-3.06	0.87	-4.15	0.68
168	700786	14	5026	0.48	0.48	0.10	0.30	0.12	0.00	0.00	0.40	0.40	-0.22	-0.14	-0.22	A-	A-	A+	A-	0.83	0.05	0.46	1.01	0.43	1.01
169	704019	14	5026	0.59	0.17	0.09	0.14	0.59	0.00	0.00	0.44	-0.24	-0.14	-0.22	0.44	A-	A-	A-	A+	0.27	0.05	-0.31	0.99	-0.04	1.00
170	696822	14	5026	0.63	0.63	0.19	0.12	0.06	0.00	0.00	0.43	0.43	-0.24	-0.21	-0.18	A+	A+	A-	A-	0.04	0.05	-1.49	0.97	-0.21	0.99
171	673946	14	5026	0.60	0.14	0.17	0.60	0.09	0.00	0.00	0.51	-0.33	-0.21	0.51	-0.19	A-	A-	A-	A+	0.20	0.05	-4.83	0.90	-4.35	0.86
172	702529	14	5026	0.39	0.26	0.14	0.39	0.21	0.00	0.00	0.30	-0.12	-0.13	0.30	-0.12	A-	A-	A-	A-	1.27	0.05	6.03	1.13	7.17	1.25
173	703978	14	5026	0.50	0.24	0.13	0.13	0.50	0.00	0.00	0.38	-0.09	-0.27	-0.20	0.38	A+	A+	A+	A-	0.69	0.05	1.18	1.02	1.19	1.03
174	678719	14	5026	0.47	0.47	0.28	0.12	0.13	0.00	0.00	0.33	0.33	-0.22	-0.17	-0.03	A+	A+	A+	A+	0.84	0.05	5.32	1.11	5.32	1.16
175	702524	14	5026	0.42	0.24	0.42	0.18	0.15	0.00	0.00	0.22	-0.12	0.22	-0.17	0.03	A-	A+	A+	A-	1.13	0.05	9.89	1.21	9.62	1.32
176	702540	14	5026	0.32	0.32	0.08	0.17	0.43	0.00	0.00	0.24	0.24	-0.21	-0.19	0.04	A-	A-	A+	B-	1.68	0.05	4.93	1.12	7.50	1.33
177	700865	15	5009	0.33	0.15	0.19	0.33	0.33	0.00	0.00	0.11	-0.05	-0.06	-0.01	0.11	A-	A-	A-	A-	1.59	0.05	9.90	1.31	9.90	1.63
178	696813	15	5009	0.46	0.30	0.46	0.16	0.07	0.00	0.00	0.37	-0.26	0.37	-0.11	-0.10	A-	A-	A+	A+	0.93	0.05	2.00	1.04	1.85	1.06
179	703349	15	5009	0.63	0.17	0.08	0.11	0.63	0.00	0.00	0.50	-0.27	-0.14	-0.31	0.50	A-	A-	A+	A+	0.05	0.05	-4.44	0.91	-3.82	0.87
180	700785	15	5009	0.28	0.28	0.19	0.20	0.32	0.00	0.00	0.27	0.27	-0.13	-0.20	0.02	A-	A-	A+	A+	1.86	0.06	3.61	1.10	6.56	1.33
181	674450	15	5009	0.67	0.06	0.13	0.15	0.67	0.00	0.00	0.57	-0.18	-0.36	-0.29	0.57	A+	A+	A+	A+	-0.14	0.05	-9.17	0.81	-7.51	0.73
182	702510	15	5009	0.31	0.23	0.41	0.31	0.05	0.00	0.00	0.18	-0.10	-0.02	0.18	-0.13	A-	A+	A-	A-	1.73	0.05	7.54	1.20	8.83	1.41
183	702480	15	5009	0.23	0.23	0.22	0.31	0.23	0.00	0.00	0.21	-0.04	-0.12	-0.05	0.21	A+	A-	A+	A-	2.18	0.06	3.77	1.12	9.08	1.58
184	702500	15	5009	0.42	0.42	0.10	0.29	0.19	0.00	0.00	0.36	0.36	-0.20	-0.14	-0.13	A+	A+	A+	A-	1.12	0.05	2.47	1.05	2.97	1.10
185	696805	15	5009	0.29	0.29	0.35	0.15	0.21	0.00	0.00	0.27	0.27	-0.11	-0.14	-0.05	A-	A-	A-	A+	1.81	0.05	6.10	1.16	9.90	1.50
186	702471	15	5009	0.44	0.18	0.44	0.18	0.20	0.00	0.00	0.41	-0.21	0.41	-0.12	-0.19	A-	A+	A+	A+	1.01	0.05	0.70	1.01	1.04	1.03
187	700859	16	4993	0.30	0.19	0.30	0.38	0.12	0.00	0.00	0.23	-0.16	0.23	0.01	-0.15	A+	A+	A-	A+	1.76	0.05	5.68	1.14	7.27	1.34
188	657753	16	4993	0.83	0.83	0.07	0.07	0.04	0.00	0.00	0.34	0.34	-0.24	-0.16	-0.15	A+	A-	A-	A+	-1.18	0.06	-1.26	0.95	1.37	1.10
189	700813	16	4993	0.30	0.07	0.22	0.42	0.30	0.00	0.00	0.27	-0.07	-0.22	-0.02	0.27	A+	A-	A-	A-	1.80	0.05	1.87	1.05	5.62	1.26
190	700868	16	4993	0.49	0.49	0.18	0.14	0.19	0.00	0.00	0.42	0.42	-0.22	-0.21	-0.13	A+	A-	A+	A+	0.77	0.05	-0.25	0.99	0.67	1.02
191	700780	16	4993	0.46	0.09	0.19	0.25	0.46	0.00	0.00	0.43	-0.23	-0.14	-0.21	0.43	A-	A-	A-	A+	0.94	0.05	-0.95	0.98	-0.41	0.99

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
192	702565	16	4993	0.22	0.49	0.18	0.11	0.22	0.00	0.00	0.01	0.19	-0.05	-0.23	0.01	A-	A-	A-	B+	2.29	0.06	9.90	1.34	9.90	1.99
193	702475	16	4993	0.53	0.12	0.22	0.53	0.13	0.00	0.00	0.26	-0.05	-0.13	0.26	-0.17	A+	A-	A-	A+	0.58	0.05	9.13	1.19	7.17	1.22
194	696825	16	4993	0.68	0.17	0.68	0.10	0.04	0.00	0.00	0.51	-0.38	0.51	-0.17	-0.20	A+	A+	A+	A+	-0.22	0.05	-5.37	0.87	-4.59	0.82
195	702465	16	4993	0.67	0.67	0.24	0.04	0.05	0.00	0.00	0.35	0.35	-0.26	-0.12	-0.13	A-	C-	B-	A+	-0.18	0.05	1.46	1.04	1.78	1.07
196	702496	16	4993	0.54	0.07	0.11	0.54	0.28	0.00	0.00	0.34	-0.29	-0.29	0.34	-0.01	A-	A-	A+	A+	0.53	0.05	5.56	1.11	5.70	1.17
197	700838	17	5013	0.69	0.15	0.12	0.69	0.03	0.00	0.00	0.52	-0.22	-0.38	0.52	-0.20	B-	B-	A-	A+	-0.31	0.05	-4.45	0.89	-4.47	0.83
198	700877	17	5013	0.87	0.06	0.04	0.04	0.87	0.00	0.00	0.42	-0.25	-0.23	-0.22	0.42	A-	A-	A-	A+	-1.60	0.07	-4.09	0.82	-5.18	0.61
199	704020	17	5013	0.80	0.04	0.08	0.08	0.80	0.00	0.00	0.47	-0.24	-0.28	-0.23	0.47	A-	A-	A-	A+	-1.00	0.06	-4.78	0.85	-3.28	0.81
200	700804	17	5013	0.52	0.08	0.52	0.28	0.11	0.00	0.00	0.34	-0.12	0.34	-0.18	-0.16	A-	A-	A-	A-	0.62	0.05	4.05	1.08	3.19	1.09
201	700880	17	5013	0.40	0.33	0.40	0.15	0.11	0.00	0.00	0.23	-0.14	0.23	-0.14	0.03	A+	A+	A-	A+	1.20	0.05	9.90	1.22	9.90	1.38
202	702549	17	5013	0.44	0.44	0.11	0.39	0.06	0.00	0.00	0.28	0.28	-0.22	-0.11	-0.06	A+	A-	A-	A+	1.03	0.05	4.90	1.10	6.50	1.20
203	702460	17	5013	0.55	0.13	0.55	0.14	0.18	0.00	0.00	0.46	-0.17	0.46	-0.20	-0.26	B-	A-	A-	A+	0.45	0.05	-1.82	0.96	-1.10	0.97
204	702570	17	5013	0.46	0.17	0.25	0.46	0.12	0.00	0.00	0.40	-0.26	-0.15	0.40	-0.10	A-	A-	A-	A-	0.94	0.05	-0.78	0.99	1.10	1.03
205	702558	17	5013	0.38	0.22	0.38	0.28	0.12	0.00	0.00	-0.12	-0.04	-0.12	0.26	-0.11	A+	A-	A+	A+	1.34	0.05	9.90	1.56	9.90	1.89
206	702478	17	5013	0.63	0.63	0.05	0.08	0.23	0.00	0.00	0.48	0.48	-0.19	-0.20	-0.31	A-	A-	A-	A+	0.06	0.05	-5.16	0.89	-5.02	0.84
207	700792	18	5023	0.51	0.03	0.26	0.51	0.20	0.00	0.00	0.38	-0.19	-0.19	0.38	-0.18	A-	A-	A-	A+	0.64	0.05	3.18	1.07	2.95	1.10
208	700798	18	5023	0.26	0.35	0.27	0.11	0.26	0.00	0.00	0.23	-0.16	0.09	-0.19	0.23	A+	A-	A+	A+	2.00	0.06	5.75	1.17	7.39	1.43
209	700806	18	5023	0.48	0.48	0.20	0.21	0.10	0.00	0.00	0.35	0.35	-0.15	-0.14	-0.18	A-	A-	A-	A+	0.79	0.05	3.22	1.07	2.92	1.09
210	700842	18	5023	0.77	0.77	0.05	0.12	0.06	0.00	0.00	0.53	0.53	-0.26	-0.31	-0.25	A-	A-	A-	A+	-0.80	0.06	-5.94	0.83	-4.83	0.73
211	700788	18	5023	0.38	0.18	0.38	0.31	0.13	0.00	0.00	0.41	-0.14	0.41	-0.20	-0.15	A-	A-	A-	A+	1.33	0.05	1.75	1.04	3.72	1.14
212	696819	18	5023	0.24	0.23	0.24	0.20	0.33	0.00	0.00	0.11	-0.12	0.11	-0.10	0.10	A-	A-	A-	A-	2.14	0.06	6.93	1.22	9.90	1.88
213	702543	18	5023	0.79	0.04	0.05	0.11	0.79	0.00	0.00	0.47	-0.23	-0.23	-0.29	0.47	A-	A-	A+	A-	-0.97	0.06	-4.30	0.87	-3.72	0.77
214	702458	18	5023	0.44	0.44	0.27	0.14	0.15	0.00	0.00	0.51	0.51	-0.16	-0.19	-0.32	A-	A+	A+	A-	1.02	0.05	-5.37	0.89	-2.30	0.93
215	702571	18	5023	0.32	0.32	0.25	0.20	0.23	0.00	0.00	0.26	0.26	-0.09	-0.14	-0.05	A-	A-	A-	A+	1.65	0.05	5.66	1.14	8.19	1.38
216	702530	18	5023	0.46	0.24	0.46	0.21	0.09	0.00	0.00	0.45	-0.27	0.45	-0.15	-0.15	A+	A+	A-	A+	0.91	0.05	-2.72	0.95	-1.31	0.96
217	704021	19	5003	0.82	0.04	0.07	0.82	0.07	0.00	0.00	0.42	-0.20	-0.26	0.42	-0.22	A+	A-	A-	A+	-1.17	0.06	-2.80	0.90	-3.31	0.78
218	700841	19	5003	0.73	0.05	0.10	0.12	0.73	0.00	0.00	0.48	-0.27	-0.24	-0.24	0.48	A+	A-	A-	A+	-0.52	0.06	-4.53	0.88	-4.20	0.81
219	700852	19	5003	0.33	0.25	0.33	0.22	0.20	0.00	0.00	0.12	-0.02	0.12	-0.10	0.00	A-	A-	A-	A-	1.61	0.05	9.90	1.27	9.90	1.60
220	703997	19	5003	0.31	0.39	0.15	0.15	0.31	0.00	0.00	0.33	-0.03	-0.10	-0.28	0.33	A+	A-	A+	A-	1.71	0.05	2.83	1.07	5.45	1.24
221	700776	19	5003	0.47	0.40	0.47	0.09	0.04	0.00	0.00	0.42	-0.15	0.42	-0.31	-0.21	A-	A-	A+	A+	0.87	0.05	0.89	1.02	2.12	1.06
222	704005	19	5003	0.60	0.15	0.60	0.17	0.09	0.00	0.00	0.43	-0.28	0.43	-0.17	-0.16	A+	A-	A-	A+	0.22	0.05	0.20	1.00	-0.34	0.99
223	702556	19	5003	0.30	0.13	0.37	0.20	0.30	0.00	0.00	0.09	0.03	-0.08	-0.03	0.09	A-	A-	A-	A-	1.77	0.05	9.90	1.31	9.90	1.59

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
224	702521	19	5003	0.20	0.17	0.20	0.14	0.49	0.00	0.00	0.17	-0.28	0.17	-0.03	0.10	A-	A+	A-	A-	2.44	0.06	3.86	1.14	9.90	1.78
225	703976	19	5003	0.60	0.10	0.11	0.60	0.18	0.00	0.00	0.43	-0.19	-0.23	0.43	-0.21	A+	B-	A-	A+	0.19	0.05	-2.65	0.95	-2.85	0.91
226	703171	19	5003	0.63	0.63	0.07	0.25	0.05	0.00	0.00	0.35	0.35	-0.20	-0.15	-0.25	A-	A-	A+	A+	0.04	0.05	2.59	1.06	2.27	1.08
227	700851	20	4978	0.27	0.21	0.14	0.39	0.27	0.00	0.00	0.10	-0.10	-0.10	0.06	0.10	A-	A-	A+	A-	2.00	0.06	9.77	1.29	9.90	1.68
228	700861	20	4978	0.63	0.07	0.17	0.13	0.63	0.00	0.00	0.47	-0.21	-0.20	-0.30	0.47	A+	A-	A-	A-	0.05	0.05	-3.28	0.93	-3.60	0.88
229	700784	20	4978	0.39	0.19	0.39	0.24	0.18	0.00	0.00	0.12	-0.06	0.12	0.02	-0.10	A-	A-	A-	A+	1.28	0.05	9.90	1.36	9.90	1.54
230	700855	20	4978	0.66	0.66	0.13	0.13	0.08	0.00	0.00	0.42	0.42	-0.23	-0.20	-0.19	A+	A-	A-	A-	-0.09	0.05	-0.64	0.99	0.27	1.01
231	700803	20	4978	0.70	0.04	0.12	0.70	0.13	0.00	0.00	0.41	-0.21	-0.23	0.41	-0.20	A-	A-	A-	A+	-0.33	0.05	0.01	1.00	-0.35	0.98
232	702555	20	4978	0.57	0.57	0.16	0.18	0.09	0.00	0.00	0.40	0.40	-0.15	-0.25	-0.15	A-	A+	A+	A-	0.41	0.05	0.90	1.02	-0.30	0.99
233	702499	20	4978	0.42	0.24	0.23	0.42	0.11	0.00	0.00	0.17	-0.14	0.01	0.17	-0.08	A+	A+	A-	A-	1.14	0.05	9.90	1.22	9.90	1.35
234	702492	20	4978	0.58	0.58	0.16	0.10	0.16	0.00	0.00	0.41	0.41	-0.28	-0.21	-0.10	A+	A-	A+	A-	0.34	0.05	-0.02	1.00	-0.27	0.99
235	702490	20	4978	0.59	0.11	0.17	0.59	0.13	0.00	0.00	0.44	-0.27	-0.18	0.44	-0.20	A+	A+	A+	A+	0.26	0.05	-1.72	0.96	-1.76	0.94
236	703178	20	4978	0.47	0.12	0.47	0.16	0.25	0.00	0.00	0.35	-0.15	0.35	-0.16	-0.15	A-	A-	A-	A+	0.89	0.05	3.15	1.06	3.12	1.09
237	700878	21	5082	0.33	0.23	0.25	0.19	0.33	0.00	0.00	0.12	-0.21	0.05	0.03	0.12	A-	A+	A+	A-	1.59	0.05	9.90	1.27	9.90	1.54
238	700783	21	5082	0.34	0.14	0.28	0.34	0.24	0.00	0.00	0.22	-0.12	-0.11	0.22	-0.04	A-	A-	A-	A+	1.51	0.05	6.46	1.15	8.19	1.32
239	700815	21	5082	0.35	0.28	0.30	0.35	0.06	0.00	0.00	0.23	-0.07	-0.05	0.23	-0.22	A+	A-	A-	A+	1.43	0.05	8.19	1.18	9.79	1.37
240	700823	21	5082	0.43	0.12	0.43	0.21	0.24	0.00	0.00	0.30	-0.18	0.30	-0.15	-0.07	A-	A-	A-	A+	1.06	0.05	6.69	1.13	8.48	1.26
241	703174	21	5082	0.22	0.27	0.18	0.22	0.33	0.00	0.00	0.17	-0.22	-0.22	0.17	0.24	A+	A+	A+	A-	2.24	0.06	4.24	1.14	7.04	1.44
242	702560	21	5082	0.22	0.37	0.22	0.20	0.20	0.00	0.00	-0.16	0.38	-0.16	-0.18	-0.11	A+	A+	A-	A+	2.22	0.06	9.90	1.47	9.90	2.31
243	702537	21	5082	0.69	0.16	0.69	0.05	0.11	0.00	0.00	0.51	-0.30	0.51	-0.22	-0.26	A-	A-	A-	A-	-0.27	0.05	-5.53	0.87	-5.18	0.81
244	702509	21	5082	0.17	0.30	0.35	0.17	0.17	0.00	0.00	0.07	-0.01	0.14	-0.24	0.07	A-	A-	A-	A-	2.58	0.06	3.78	1.15	8.80	1.72
245	704003	21	5082	0.77	0.77	0.07	0.12	0.04	0.00	0.00	0.43	0.43	-0.22	-0.25	-0.22	A+	A-	A-	A+	-0.79	0.06	-2.29	0.93	-1.59	0.91
246	702484	21	5082	0.43	0.43	0.20	0.18	0.19	0.00	0.00	0.40	0.40	-0.18	-0.16	-0.17	A+	A+	A+	A-	1.02	0.05	0.54	1.01	1.18	1.03
247	700832	22	5000	0.32	0.26	0.32	0.17	0.25	0.00	0.00	0.28	0.05	0.28	-0.29	-0.09	A-	A+	A-	A+	1.69	0.05	3.32	1.08	5.32	1.22
248	674443	22	5000	0.83	0.83	0.03	0.10	0.03	0.00	0.00	0.50	0.50	-0.22	-0.35	-0.22	A+	A-	B-	A+	-1.18	0.06	-7.32	0.74	-7.86	0.53
249	700790	22	5000	0.39	0.23	0.21	0.17	0.39	0.00	0.00	0.30	-0.10	-0.21	-0.04	0.30	A-	A-	A-	A+	1.27	0.05	3.46	1.07	5.92	1.20
250	700799	22	5000	0.21	0.45	0.18	0.21	0.16	0.00	0.00	0.19	-0.16	-0.01	0.19	0.01	A-	A-	A+	A+	2.35	0.06	2.47	1.08	6.51	1.42
251	700802	22	5000	0.66	0.66	0.12	0.15	0.07	0.00	0.00	0.50	0.50	-0.27	-0.22	-0.27	A+	A-	A-	A+	-0.11	0.05	-5.50	0.88	-4.47	0.83
252	703980	22	5000	0.45	0.46	0.45	0.06	0.03	0.00	0.00	0.09	0.04	0.09	-0.17	-0.16	A-	A-	A-	A+	0.97	0.05	9.90	1.36	9.90	1.53
253	704025	22	5000	0.61	0.05	0.61	0.21	0.14	0.00	0.00	0.44	-0.15	0.44	-0.30	-0.17	A+	A-	A-	A-	0.19	0.05	-1.40	0.97	-1.84	0.94
254	702559	22	5000	0.26	0.28	0.27	0.20	0.26	0.00	0.00	0.06	0.04	-0.03	-0.08	0.06	A-	A-	A-	A-	2.04	0.06	8.91	1.27	9.90	1.67
255	704001	22	5000	0.29	0.12	0.09	0.29	0.50	0.00	0.00	0.33	-0.11	-0.24	0.33	-0.09	A+	A+	A-	A+	1.86	0.05	1.04	1.03	4.44	1.20

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
256	702459	22	5000	0.48	0.22	0.12	0.48	0.18	0.00	0.00	0.48	-0.22	-0.22	0.48	-0.20	B-	A-	A-	A+	0.82	0.05	-4.47	0.92	-2.62	0.93
257	696808	23	5011	0.43	0.31	0.07	0.43	0.18	0.00	0.00	0.17	0.05	-0.20	0.17	-0.14	A+	A-	A-	A-	1.06	0.05	9.90	1.23	9.55	1.31
258	703972	23	5011	0.33	0.28	0.33	0.19	0.20	0.00	0.00	0.26	-0.05	0.26	-0.18	-0.07	A+	A-	A+	A+	1.61	0.05	5.27	1.13	7.39	1.30
259	703348	23	5011	0.32	0.32	0.22	0.21	0.24	0.00	0.00	0.09	0.09	0.04	-0.04	-0.08	A+	A+	A-	A+	1.64	0.05	9.90	1.32	9.90	1.65
260	700860	23	5011	0.21	0.21	0.32	0.25	0.22	0.00	0.00	0.08	0.08	0.08	-0.11	-0.05	A+	A-	A-	A+	2.33	0.06	7.28	1.26	9.90	1.74
261	700869	23	5011	0.38	0.13	0.42	0.38	0.06	0.01	0.00	0.36	-0.12	-0.21	0.36	-0.12	A-	A-	A+	A+	1.31	0.05	1.30	1.03	3.39	1.11
262	702464	23	5011	0.22	0.18	0.18	0.42	0.22	0.00	0.00	0.19	-0.13	-0.21	0.10	0.19	A-	A-	A-	A+	2.24	0.06	3.11	1.10	8.03	1.51
263	702501	23	5011	0.32	0.32	0.22	0.12	0.34	0.00	0.00	0.22	0.22	-0.20	-0.23	0.11	A+	A+	A+	A-	1.63	0.05	6.96	1.17	8.40	1.35
264	702469	23	5011	0.30	0.08	0.36	0.30	0.26	0.00	0.00	0.27	-0.14	-0.06	0.27	-0.12	A-	A+	A-	A-	1.78	0.05	2.14	1.05	6.11	1.28
265	678786	23	5011	0.26	0.19	0.41	0.14	0.26	0.00	0.00	0.18	-0.17	0.00	-0.04	0.18	A-	A-	A+	A+	2.02	0.06	6.59	1.20	8.49	1.46
266	702541	23	5011	0.56	0.56	0.18	0.17	0.08	0.00	0.00	0.51	0.51	-0.23	-0.23	-0.28	A-	A-	A-	A+	0.41	0.05	-6.33	0.88	-5.15	0.85
267	700864	24	4924	0.56	0.56	0.12	0.22	0.10	0.00	0.00	0.45	0.45	-0.17	-0.24	-0.23	A-	A-	A-	A+	0.40	0.05	-1.72	0.97	-1.45	0.96
268	696824	24	4924	0.17	0.43	0.33	0.07	0.17	0.00	0.00	0.25	0.04	-0.20	-0.08	0.25	B-	A+	A-	A-	2.65	0.06	-1.41	0.95	5.32	1.42
269	681818	24	4924	0.51	0.12	0.18	0.51	0.19	0.00	0.00	0.32	-0.18	-0.19	0.32	-0.07	A+	A-	A-	A+	0.67	0.05	5.44	1.11	4.93	1.14
270	700849	24	4924	0.29	0.42	0.23	0.29	0.06	0.00	0.00	0.34	-0.22	-0.06	0.34	-0.08	A-	A-	A-	A+	1.85	0.05	-0.43	0.99	1.59	1.07
271	700816	24	4924	0.63	0.06	0.16	0.14	0.63	0.00	0.00	0.43	-0.24	-0.24	-0.17	0.43	A+	A+	A-	A+	0.06	0.05	-0.80	0.98	-0.24	0.99
272	702479	24	4924	0.33	0.27	0.09	0.33	0.32	0.00	0.00	-0.12	0.28	-0.09	-0.12	-0.08	A-	A-	A-	A+	1.61	0.05	9.90	1.57	9.90	2.05
273	702520	24	4924	0.35	0.35	0.20	0.21	0.23	0.00	0.00	0.25	0.25	-0.18	-0.16	0.04	A-	A+	A+	A-	1.48	0.05	6.30	1.15	7.42	1.28
274	703351	24	4924	0.73	0.06	0.08	0.13	0.73	0.00	0.00	0.46	-0.23	-0.27	-0.22	0.46	A+	A-	A-	A+	-0.48	0.05	-4.46	0.89	-3.32	0.85
275	702508	24	4924	0.33	0.12	0.33	0.25	0.29	0.00	0.00	0.21	-0.23	0.21	-0.28	0.23	A+	A+	A-	A+	1.61	0.05	6.74	1.17	8.85	1.38
276	702494	24	4924	0.47	0.23	0.47	0.15	0.15	0.00	0.00	0.27	-0.17	0.27	-0.14	-0.04	A+	A-	A+	A-	0.89	0.05	7.53	1.15	7.08	1.21

Table J–6. Biology Multiple-Choice Item Statistics: Spring

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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***	**MS**	***z***	**MS**
1	607777	0	154585	0.59	0.11	0.08	0.22	0.59	0.00	0.00	0.55	-0.25	-0.21	-0.32	0.55					-0.24	0.01	-9.90	0.84	-9.90	0.77
2	674080	0	154585	0.61	0.61	0.14	0.16	0.09	0.00	0.00	0.46	0.46	-0.29	-0.24	-0.13					0.05	0.01	-9.90	0.89	-9.90	0.83
3	611477	0	154585	0.74	0.07	0.06	0.74	0.12	0.00	0.00	0.48	-0.30	-0.21	0.48	-0.25					-1.04	0.01	-9.90	0.86	-9.90	0.67
4	680195	0	154585	0.63	0.06	0.06	0.25	0.63	0.00	0.00	0.45	-0.22	-0.26	-0.23	0.45					-0.55	0.01	5.77	1.03	1.14	1.01
5	635783	0	154585	0.72	0.72	0.08	0.08	0.12	0.00	0.00	0.45	0.45	-0.24	-0.23	-0.23					-0.29	0.01	-9.90	0.80	-9.90	0.68
6	678545	0	154585	0.71	0.08	0.09	0.71	0.12	0.00	0.00	0.42	-0.20	-0.24	0.42	-0.20					-0.81	0.01	-4.80	0.97	-6.05	0.93
7	680242	0	154585	0.62	0.09	0.62	0.10	0.19	0.00	0.00	0.44	-0.27	0.44	-0.30	-0.11					-0.11	0.01	-9.90	0.95	-9.90	0.90
8	609427	0	154585	0.76	0.76	0.05	0.07	0.12	0.00	0.00	0.46	0.46	-0.26	-0.23	-0.25					-0.36	0.01	-9.90	0.74	-9.90	0.64
9	635766	0	154585	0.48	0.14	0.17	0.21	0.48	0.00	0.00	0.30	-0.26	-0.16	-0.01	0.30					0.43	0.01	9.90	1.15	9.90	1.20
10	610882	0	154585	0.61	0.19	0.61	0.12	0.09	0.00	0.00	0.44	-0.22	0.44	-0.24	-0.18					-0.12	0.01	-9.90	0.95	-9.90	0.90
11	644752	0	154585	0.51	0.51	0.33	0.10	0.06	0.00	0.00	0.31	0.31	-0.07	-0.24	-0.21					0.34	0.01	9.90	1.16	9.90	1.24
12	644755	0	154585	0.55	0.16	0.10	0.55	0.20	0.00	0.00	0.46	-0.09	-0.30	0.46	-0.28					0.28	0.01	-9.90	0.94	-9.90	0.94
13	678875	0	154585	0.73	0.08	0.12	0.73	0.07	0.00	0.00	0.47	-0.28	-0.25	0.47	-0.19					-0.85	0.01	-9.90	0.83	-9.90	0.72
14	635786	0	154585	0.50	0.18	0.22	0.50	0.10	0.00	0.00	0.39	-0.19	-0.14	0.39	-0.20					0.40	0.01	6.93	1.03	3.22	1.02
15	677989	0	154585	0.51	0.11	0.26	0.51	0.11	0.00	0.00	0.58	-0.25	-0.35	0.58	-0.17					0.15	0.01	-9.90	0.81	-9.90	0.74
16	671325	0	154585	0.65	0.11	0.65	0.15	0.09	0.00	0.00	0.45	-0.18	0.45	-0.27	-0.19					0.17	0.01	-9.90	0.91	-9.90	0.92
17	678965	0	154585	0.41	0.41	0.22	0.14	0.23	0.00	0.00	0.42	0.42	-0.15	-0.28	-0.10					0.79	0.01	-5.22	0.98	-3.14	0.98
18	610880	0	154585	0.57	0.18	0.09	0.15	0.57	0.00	0.00	0.46	-0.19	-0.26	-0.21	0.46					-0.16	0.01	-3.67	0.98	-8.62	0.93
19	642317	0	154585	0.36	0.43	0.13	0.09	0.36	0.00	0.00	0.54	-0.20	-0.26	-0.24	0.54					1.33	0.01	-9.90	0.87	-9.90	0.88
20	635799	0	154585	0.71	0.71	0.10	0.13	0.06	0.00	0.00	0.52	0.52	-0.30	-0.23	-0.27					-0.58	0.01	-9.90	0.76	-9.90	0.67
21	635597	0	154585	0.63	0.63	0.18	0.11	0.07	0.00	0.00	0.48	0.48	-0.20	-0.27	-0.25					-0.57	0.01	-5.85	0.97	-9.90	0.87
22	607712	0	154585	0.50	0.17	0.23	0.50	0.10	0.00	0.00	0.42	-0.27	-0.08	0.42	-0.23					0.37	0.01	-2.36	0.99	-4.21	0.97
23	681052	0	154585	0.43	0.11	0.43	0.24	0.21	0.00	0.00	0.48	-0.21	0.48	-0.30	-0.10					0.51	0.01	-9.90	0.92	-9.90	0.90
24	677992	0	154585	0.47	0.09	0.47	0.25	0.19	0.00	0.00	0.45	-0.22	0.45	-0.25	-0.12					0.58	0.01	-9.90	0.95	-8.91	0.95
25	643400	0	154585	0.44	0.19	0.15	0.44	0.22	0.00	0.00	0.31	-0.19	-0.27	0.31	0.05					0.95	0.01	9.90	1.13	9.90	1.18
26	681523	0	154585	0.56	0.14	0.56	0.18	0.12	0.00	0.00	0.46	-0.24	0.46	-0.24	-0.17					-0.06	0.01	-9.90	0.95	-9.90	0.89
27	608245	0	154585	0.73	0.17	0.06	0.04	0.73	0.00	0.00	0.44	-0.27	-0.23	-0.20	0.44					-0.76	0.01	-9.90	0.86	-9.90	0.82
28	607589	0	154585	0.50	0.50	0.20	0.25	0.04	0.00	0.00	0.46	0.46	-0.35	-0.13	-0.17					-0.06	0.01	9.90	1.05	4.46	1.03
29	611072	0	154585	0.70	0.11	0.10	0.09	0.70	0.00	0.00	0.39	-0.15	-0.22	-0.22	0.39					-0.37	0.01	-9.90	0.91	-3.29	0.97
30	611066	0	154585	0.57	0.11	0.12	0.57	0.20	0.00	0.00	0.41	-0.23	-0.17	0.41	-0.18					-0.21	0.01	9.90	1.07	9.90	1.10
31	611046	0	154585	0.56	0.13	0.12	0.20	0.56	0.00	0.00	0.45	-0.19	-0.24	-0.21	0.45					0.15	0.01	-9.90	0.96	-5.71	0.96

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	608286	0	154585	0.44	0.44	0.17	0.23	0.15	0.00	0.00	0.29	0.29	-0.24	-0.07	-0.07					0.36	0.01	9.90	1.20	9.90	1.29
33	607578	0	154585	0.37	0.30	0.37	0.24	0.08	0.00	0.00	0.31	-0.11	0.31	-0.16	-0.11					0.67	0.01	9.90	1.08	9.90	1.10
34	683555	0	154585	0.54	0.54	0.10	0.07	0.29	0.00	0.00	0.43	0.43	-0.19	-0.24	-0.21					-0.07	0.01	8.56	1.04	8.18	1.06
35	684511	0	154585	0.66	0.08	0.18	0.66	0.09	0.00	0.00	0.47	-0.20	-0.27	0.47	-0.24					-0.57	0.01	-9.90	0.92	-9.90	0.84
36	682405	0	154585	0.49	0.20	0.11	0.49	0.20	0.00	0.00	0.40	-0.23	-0.22	0.40	-0.10					0.39	0.01	2.16	1.01	-0.76	1.00
37	611476	0	154585	0.55	0.55	0.23	0.07	0.14	0.00	0.00	0.40	0.40	-0.23	-0.25	-0.08					0.00	0.01	7.48	1.03	4.45	1.03
38	612588	0	154585	0.42	0.25	0.21	0.12	0.42	0.00	0.00	0.52	-0.13	-0.32	-0.20	0.52					0.37	0.01	-9.90	0.91	-9.90	0.89
39	612675	0	154585	0.62	0.16	0.62	0.11	0.10	0.00	0.00	0.48	-0.15	0.48	-0.32	-0.23					-0.22	0.01	-9.90	0.89	-9.90	0.85
40	610909	0	154585	0.55	0.55	0.17	0.18	0.09	0.00	0.00	0.41	0.41	-0.09	-0.23	-0.28					0.05	0.01	0.59	1.00	-1.33	0.99
41	608019	0	154585	0.54	0.54	0.11	0.20	0.14	0.00	0.00	0.49	0.49	-0.27	-0.23	-0.17					0.16	0.01	-9.90	0.91	-9.90	0.90
42	608382	0	154585	0.71	0.08	0.08	0.13	0.71	0.00	0.00	0.48	-0.26	-0.26	-0.22	0.48					-1.00	0.01	-7.32	0.95	-9.90	0.87
43	642871	0	154585	0.43	0.25	0.43	0.18	0.13	0.00	0.00	0.32	-0.02	0.32	-0.22	-0.16					0.66	0.01	9.90	1.11	9.90	1.14
44	608279	0	154585	0.54	0.09	0.11	0.54	0.25	0.00	0.00	0.36	-0.18	-0.15	0.36	-0.17					-0.09	0.01	9.90	1.14	9.90	1.20
45	641293	0	154585	0.62	0.16	0.13	0.62	0.08	0.00	0.00	0.49	-0.22	-0.27	0.49	-0.23					-0.37	0.01	-9.90	0.91	-9.90	0.87
46	678986	0	154585	0.69	0.07	0.16	0.08	0.69	0.00	0.00	0.56	-0.28	-0.28	-0.31	0.56					-0.82	0.01	-9.90	0.82	-9.90	0.69
47	607750	0	154585	0.42	0.19	0.21	0.17	0.42	0.00	0.00	0.37	-0.14	-0.07	-0.25	0.37					0.67	0.01	9.90	1.06	9.90	1.08
48	678894	0	154585	0.46	0.19	0.46	0.15	0.20	0.00	0.00	0.39	-0.22	0.39	-0.25	-0.03					0.76	0.01	9.90	1.04	9.90	1.06
49	703253	1	4828	0.50	0.21	0.17	0.50	0.12	0.00	0.00	0.23	-0.10	-0.13	0.23	-0.08	A-	A-	A-	A-	0.63	0.05	9.56	1.20	9.70	1.32
50	703251	1	4828	0.39	0.21	0.26	0.14	0.39	0.00	0.00	0.46	-0.27	-0.12	-0.17	0.46	A-	A+	A-	A-	1.22	0.05	-3.07	0.94	-0.94	0.97
51	703261	1	4828	0.53	0.53	0.19	0.15	0.13	0.00	0.00	0.40	0.40	-0.15	-0.24	-0.15	A-	A-	A+	A-	0.50	0.05	0.95	1.02	1.04	1.03
52	702628	1	4828	0.66	0.07	0.09	0.66	0.17	0.00	0.00	0.43	-0.24	-0.28	0.43	-0.15	A+	A-	A-	A-	-0.21	0.05	-1.33	0.97	-1.87	0.92
53	702262	1	4828	0.64	0.19	0.64	0.11	0.06	0.00	0.00	0.44	-0.29	0.44	-0.10	-0.26	A+	A-	A-	A+	-0.06	0.05	-1.40	0.97	-0.97	0.96
54	702103	1	4828	0.26	0.34	0.07	0.26	0.33	0.00	0.00	0.28	-0.11	-0.31	0.28	0.02	A-	A-	A-	A-	1.94	0.06	0.95	1.03	6.33	1.34
55	703487	1	4828	0.58	0.12	0.58	0.11	0.19	0.00	0.00	0.34	-0.15	0.34	-0.23	-0.11	A+	A-	A+	A-	0.24	0.05	4.79	1.10	2.71	1.09
56	641257	1	4828	0.53	0.13	0.53	0.20	0.14	0.00	0.00	0.37	-0.11	0.37	-0.17	-0.23	A+	A-	A-	A+	0.49	0.05	3.95	1.08	2.59	1.08
57	702618	1	4828	0.80	0.09	0.80	0.08	0.04	0.00	0.00	0.47	-0.24	0.47	-0.28	-0.23	A+	A-	A-	A-	-1.04	0.06	-4.11	0.87	-3.62	0.77
58	702616	1	4828	0.43	0.27	0.43	0.20	0.09	0.00	0.00	0.27	-0.08	0.27	-0.10	-0.19	A+	A-	A-	A+	0.99	0.05	6.68	1.14	7.17	1.23
59	674136	1	4828	0.74	0.74	0.10	0.05	0.10	0.00	0.00	0.50	0.50	-0.28	-0.27	-0.23	A+	A-	A+	A-	-0.64	0.06	-3.70	0.90	-4.15	0.79
60	704782	1	4828	0.40	0.27	0.16	0.16	0.40	0.00	0.00	0.39	-0.04	-0.20	-0.27	0.39	A-	A-	A+	A-	1.14	0.05	0.58	1.01	1.53	1.05
61	702149	1	4828	0.31	0.13	0.40	0.31	0.15	0.00	0.00	0.30	-0.02	-0.20	0.30	-0.08	A-	A-	A+	A-	1.65	0.05	2.36	1.06	6.28	1.28
62	702731	1	4828	0.75	0.05	0.12	0.07	0.75	0.00	0.00	0.52	-0.27	-0.28	-0.28	0.52	A+	A-	A-	A-	-0.74	0.06	-6.41	0.83	-5.10	0.73
63	705231	1	4828	0.43	0.23	0.43	0.19	0.15	0.00	0.00	0.29	0.05	0.29	-0.23	-0.19	A+	A-	A+	A-	1.01	0.05	6.83	1.14	5.92	1.19

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
64	673877	1	4828	0.58	0.58	0.08	0.18	0.16	0.00	0.00	0.26	0.26	-0.20	-0.08	-0.12	A-	A-	A-	A-	0.22	0.05	9.90	1.23	6.40	1.23
65	703250	2	4787	0.31	0.31	0.21	0.25	0.22	0.00	0.00	0.15	0.15	-0.05	-0.06	-0.05	A-	A-	A-	A+	1.63	0.05	9.10	1.23	9.72	1.43
66	703249	2	4787	0.24	0.17	0.26	0.24	0.33	0.00	0.00	0.08	-0.14	0.00	0.08	0.04	A-	A-	A-	A-	2.05	0.06	7.21	1.22	9.90	1.64
67	702630	2	4787	0.63	0.12	0.13	0.13	0.63	0.00	0.00	0.54	-0.28	-0.30	-0.22	0.54	A-	A-	A-	A+	-0.01	0.05	-6.02	0.87	-5.48	0.82
68	703488	2	4787	0.61	0.61	0.06	0.20	0.13	0.00	0.00	0.56	0.56	-0.25	-0.32	-0.26	A+	A+	A+	A+	0.09	0.05	-9.05	0.82	-8.13	0.75
69	702104	2	4787	0.61	0.16	0.61	0.13	0.10	0.00	0.00	0.50	-0.28	0.50	-0.23	-0.21	A-	A-	A-	A-	0.11	0.05	-3.60	0.93	-4.20	0.87
70	702263	2	4787	0.47	0.47	0.16	0.20	0.17	0.00	0.00	0.50	0.50	-0.29	-0.20	-0.17	A-	A-	A-	A-	0.78	0.05	-4.87	0.91	-2.89	0.92
71	703263	2	4787	0.39	0.39	0.08	0.40	0.14	0.00	0.00	0.29	0.29	-0.21	-0.12	-0.09	A+	A+	A-	A-	1.22	0.05	5.71	1.13	7.06	1.24
72	683613	2	4787	0.56	0.20	0.11	0.13	0.56	0.00	0.00	0.36	-0.12	-0.28	-0.12	0.36	A+	A+	A+	A-	0.34	0.05	4.71	1.10	3.95	1.12
73	702621	2	4787	0.72	0.10	0.72	0.08	0.10	0.00	0.00	0.49	-0.26	0.49	-0.28	-0.21	A+	A-	A+	A-	-0.52	0.06	-3.88	0.90	-4.73	0.79
74	702620	2	4787	0.45	0.45	0.13	0.27	0.15	0.00	0.00	0.34	0.34	-0.27	-0.02	-0.20	A+	A-	A-	A-	0.89	0.05	3.07	1.06	3.73	1.11
75	644033	2	4787	0.56	0.07	0.07	0.56	0.30	0.00	0.00	0.35	-0.29	-0.27	0.35	-0.07	A+	A-	A-	A-	0.34	0.05	4.28	1.09	2.74	1.08
76	678910	2	4787	0.34	0.34	0.30	0.23	0.13	0.00	0.00	0.09	0.09	0.08	-0.10	-0.10	A+	A-	A+	A-	1.47	0.05	9.90	1.33	9.90	1.58
77	703166	2	4787	0.81	0.03	0.11	0.04	0.81	0.00	0.00	0.48	-0.25	-0.28	-0.26	0.48	A-	A-	A-	A-	-1.15	0.06	-3.90	0.87	-4.25	0.73
78	704783	2	4787	0.48	0.16	0.23	0.13	0.48	0.00	0.00	0.39	-0.15	-0.18	-0.18	0.39	A+	A-	A+	A+	0.77	0.05	1.62	1.03	1.70	1.05
79	642887	2	4787	0.44	0.15	0.44	0.13	0.28	0.00	0.00	0.27	-0.11	0.27	-0.22	-0.04	A-	A-	A-	A+	0.94	0.05	7.42	1.16	7.04	1.21
80	702150	2	4787	0.70	0.70	0.04	0.14	0.12	0.00	0.00	0.51	0.51	-0.25	-0.37	-0.16	A+	A+	A-	A-	-0.39	0.05	-5.83	0.86	-4.90	0.80
81	704374	3	4730	0.40	0.12	0.42	0.06	0.40	0.00	0.00	0.32	-0.11	-0.13	-0.24	0.32	A-	A-	A-	A-	1.15	0.05	3.45	1.07	4.26	1.13
82	704383	3	4730	0.35	0.23	0.19	0.35	0.23	0.00	0.00	0.20	-0.08	-0.12	0.20	-0.03	A-	A+	A-	A-	1.41	0.05	7.24	1.17	8.81	1.33
83	703485	3	4730	0.64	0.16	0.10	0.09	0.64	0.00	0.00	0.44	-0.23	-0.16	-0.26	0.44	A+	A-	A-	A-	-0.08	0.05	-0.62	0.99	-1.41	0.95
84	703989	3	4730	0.42	0.42	0.20	0.20	0.18	0.00	0.00	0.21	0.21	-0.11	-0.19	0.04	A+	A-	A-	A+	1.03	0.05	9.62	1.20	9.09	1.28
85	703489	3	4730	0.33	0.28	0.18	0.21	0.33	0.00	0.00	0.37	-0.10	-0.11	-0.21	0.37	A-	A-	A+	A+	1.51	0.05	-0.81	0.98	2.77	1.10
86	703237	3	4730	0.32	0.19	0.31	0.17	0.32	0.00	0.00	0.31	-0.24	-0.03	-0.10	0.31	A+	A-	A+	A+	1.55	0.05	2.27	1.05	6.30	1.25
87	702201	3	4730	0.32	0.09	0.38	0.21	0.32	0.00	0.00	0.14	-0.14	0.05	-0.12	0.14	A-	A-	A-	A+	1.60	0.05	7.47	1.18	8.16	1.35
88	702073	3	4730	0.72	0.72	0.09	0.08	0.10	0.00	0.00	0.45	0.45	-0.28	-0.20	-0.21	A+	A-	A-	A+	-0.52	0.06	-2.98	0.92	-2.80	0.87
89	702617	3	4730	0.60	0.15	0.12	0.60	0.13	0.00	0.00	0.30	-0.11	-0.18	0.30	-0.15	A-	A-	A+	A+	0.13	0.05	4.84	1.10	3.68	1.12
90	702619	3	4730	0.70	0.70	0.07	0.16	0.07	0.00	0.00	0.45	0.45	-0.28	-0.20	-0.24	A-	A-	A-	A-	-0.41	0.05	-2.97	0.93	-2.92	0.88
91	681051	3	4730	0.61	0.15	0.61	0.19	0.05	0.00	0.00	0.44	-0.27	0.44	-0.23	-0.13	A+	A-	A+	A+	0.09	0.05	-2.47	0.95	-2.61	0.92
92	680251	3	4730	0.59	0.12	0.14	0.15	0.59	0.00	0.00	0.42	-0.32	-0.09	-0.20	0.42	A+	A+	A+	A-	0.18	0.05	-0.19	1.00	-0.39	0.99
93	702170	3	4730	0.76	0.06	0.76	0.14	0.04	0.00	0.00	0.43	-0.31	0.43	-0.17	-0.23	A+	A-	A+	A-	-0.75	0.06	-0.55	0.98	-0.71	0.96
94	642858	3	4730	0.33	0.10	0.20	0.37	0.33	0.00	0.00	0.11	-0.18	-0.03	0.03	0.11	A+	A+	A+	A-	1.53	0.05	9.90	1.28	9.90	1.48
95	677881	3	4730	0.59	0.16	0.11	0.59	0.14	0.00	0.00	0.44	-0.25	-0.26	0.44	-0.13	A-	A-	A-	A+	0.20	0.05	-0.49	0.99	0.24	1.01

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
96	704201	3	4730	0.48	0.40	0.07	0.48	0.05	0.00	0.00	0.31	-0.08	-0.24	0.31	-0.26	A-	A-	A-	A-	0.75	0.05	4.22	1.08	4.84	1.14
97	704372	4	4752	0.24	0.54	0.10	0.13	0.24	0.00	0.00	0.27	0.00	-0.27	-0.10	0.27	A-	A-	A-	A+	2.13	0.06	2.58	1.08	5.20	1.30
98	704373	4	4752	0.42	0.42	0.08	0.15	0.35	0.00	0.00	0.45	0.45	-0.20	-0.34	-0.10	A-	A-	A-	A+	1.07	0.05	-2.82	0.94	-1.69	0.95
99	702631	4	4752	0.39	0.14	0.39	0.30	0.17	0.00	0.00	0.24	-0.01	0.24	-0.03	-0.27	A-	A-	A-	A-	1.25	0.05	8.86	1.20	8.49	1.30
100	678872	4	4752	0.64	0.64	0.15	0.08	0.14	0.00	0.00	0.34	0.34	-0.17	-0.20	-0.15	A-	A-	A-	A+	-0.05	0.05	3.04	1.07	2.05	1.08
101	703356	4	4752	0.49	0.35	0.09	0.07	0.49	0.00	0.00	0.36	-0.09	-0.28	-0.24	0.36	A-	A+	A-	A-	0.69	0.05	4.47	1.09	3.25	1.09
102	702058	4	4752	0.49	0.32	0.49	0.13	0.06	0.00	0.00	0.25	0.03	0.25	-0.26	-0.22	A+	A+	A+	A-	0.69	0.05	9.70	1.21	8.60	1.26
103	703697	4	4752	0.43	0.16	0.43	0.23	0.18	0.00	0.00	0.39	-0.20	0.39	-0.24	-0.04	A-	A-	A-	A+	1.03	0.05	1.43	1.03	3.46	1.10
104	702261	4	4752	0.39	0.14	0.18	0.39	0.28	0.00	0.00	0.38	-0.10	-0.28	0.38	-0.09	A-	A-	A+	A-	1.23	0.05	-0.26	0.99	1.21	1.04
105	703149	4	4752	0.75	0.06	0.75	0.08	0.11	0.00	0.00	0.44	-0.22	0.44	-0.29	-0.18	A+	A-	A+	C-	-0.69	0.06	0.14	1.00	-0.41	0.98
106	703151	4	4752	0.64	0.06	0.20	0.64	0.10	0.00	0.00	0.54	-0.28	-0.28	0.54	-0.26	A+	A-	A-	A+	-0.06	0.05	-6.86	0.86	-6.66	0.77
107	641287	4	4752	0.64	0.12	0.12	0.64	0.12	0.00	0.00	0.37	-0.21	-0.23	0.37	-0.11	A+	A-	A+	A+	-0.07	0.05	3.16	1.07	1.09	1.04
108	703156	4	4752	0.46	0.46	0.21	0.17	0.15	0.00	0.00	0.43	0.43	-0.16	-0.20	-0.21	A+	A-	A-	A-	0.86	0.05	-1.60	0.97	-0.32	0.99
109	679971	4	4752	0.28	0.28	0.25	0.22	0.24	0.00	0.00	0.33	0.33	-0.15	-0.02	-0.16	A-	A+	A+	A-	1.83	0.06	0.49	1.01	4.18	1.20
110	679208	4	4752	0.72	0.09	0.03	0.72	0.16	0.00	0.00	0.38	-0.13	-0.22	0.38	-0.26	A+	A-	A-	A-	-0.50	0.06	2.18	1.06	0.47	1.02
111	702733	4	4752	0.68	0.68	0.09	0.19	0.04	0.00	0.00	0.45	0.45	-0.23	-0.24	-0.25	A-	A-	A-	A-	-0.27	0.05	-1.92	0.96	-2.07	0.91
112	679974	4	4752	0.42	0.13	0.22	0.42	0.23	0.00	0.00	0.29	-0.18	-0.22	0.29	0.01	A-	A-	A-	A+	1.07	0.05	6.73	1.15	8.06	1.26
113	701042	5	4756	0.41	0.41	0.27	0.15	0.17	0.00	0.00	0.36	0.36	-0.16	-0.21	-0.08	A+	A+	A-	A-	1.10	0.05	1.68	1.03	2.02	1.06
114	701039	5	4756	0.44	0.13	0.33	0.44	0.10	0.00	0.00	0.22	-0.17	-0.02	0.22	-0.14	A+	A+	A+	A+	0.92	0.05	9.90	1.20	9.37	1.27
115	702633	5	4756	0.56	0.20	0.56	0.12	0.12	0.00	0.00	0.39	-0.14	0.39	-0.27	-0.14	A+	A+	A-	A-	0.37	0.05	-0.54	0.99	-1.39	0.96
116	703698	5	4756	0.53	0.53	0.29	0.10	0.07	0.00	0.00	0.26	0.26	-0.01	-0.23	-0.22	A+	A-	A+	A-	0.50	0.05	6.64	1.13	4.88	1.13
117	702222	5	4756	0.37	0.37	0.32	0.13	0.18	0.00	0.00	0.24	0.24	0.01	-0.30	-0.04	A-	A+	A+	A+	1.29	0.05	6.99	1.15	7.79	1.27
118	700891	5	4756	0.55	0.17	0.11	0.55	0.17	0.00	0.00	0.25	-0.12	-0.13	0.25	-0.10	A+	A+	A+	A-	0.39	0.05	8.42	1.17	7.12	1.20
119	702057	5	4756	0.48	0.17	0.25	0.48	0.10	0.00	0.00	0.29	-0.15	-0.20	0.29	0.00	A-	A-	A-	A+	0.77	0.05	7.10	1.14	7.14	1.20
120	703357	5	4756	0.38	0.35	0.06	0.20	0.38	0.00	0.00	0.11	0.00	-0.08	-0.08	0.11	A+	A-	A-	A+	1.22	0.05	9.90	1.31	9.90	1.42
121	703147	5	4756	0.67	0.15	0.10	0.08	0.67	0.00	0.00	0.48	-0.25	-0.21	-0.26	0.48	A+	A+	A+	A+	-0.23	0.05	-3.96	0.91	-3.84	0.86
122	703148	5	4756	0.77	0.77	0.08	0.10	0.05	0.00	0.00	0.49	0.49	-0.27	-0.26	-0.25	A-	A-	A-	A-	-0.81	0.06	-3.34	0.90	-3.80	0.81
123	643788	5	4756	0.42	0.25	0.42	0.11	0.22	0.00	0.00	0.24	-0.06	0.24	-0.15	-0.11	A+	A+	A+	A-	1.04	0.05	7.27	1.15	8.04	1.24
124	702734	5	4756	0.70	0.12	0.09	0.70	0.10	0.00	0.00	0.40	-0.19	-0.24	0.40	-0.18	A-	A-	A+	A+	-0.40	0.05	-0.58	0.99	-1.52	0.94
125	703157	5	4756	0.53	0.18	0.15	0.14	0.53	0.00	0.00	0.54	-0.27	-0.28	-0.18	0.54	A+	A-	A+	A-	0.51	0.05	-6.91	0.87	-6.51	0.84
126	702157	5	4756	0.27	0.13	0.27	0.39	0.22	0.00	0.00	0.00	-0.12	0.00	0.07	0.01	A+	A+	A-	A+	1.86	0.06	9.90	1.37	9.90	1.80
127	677869	5	4756	0.31	0.14	0.31	0.25	0.30	0.00	0.00	0.20	-0.31	0.20	0.01	0.02	A-	A-	A-	A-	1.61	0.05	6.04	1.15	8.39	1.35

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
128	679253	5	4756	0.69	0.06	0.08	0.69	0.17	0.00	0.00	0.40	-0.19	-0.26	0.40	-0.18	A-	A-	A-	A-	-0.32	0.05	-0.79	0.98	0.55	1.02
129	701040	6	4837	0.36	0.24	0.22	0.36	0.18	0.00	0.00	0.20	-0.01	-0.21	0.20	0.00	A-	A+	A+	A-	1.36	0.05	8.07	1.18	9.63	1.36
130	701044	6	4837	0.51	0.16	0.15	0.18	0.51	0.00	0.00	0.29	-0.14	-0.11	-0.14	0.29	A+	A-	A-	A-	0.57	0.05	7.26	1.15	6.62	1.20
131	703358	6	4837	0.59	0.59	0.16	0.15	0.09	0.00	0.00	0.51	0.51	-0.21	-0.25	-0.28	A-	A-	A-	A+	0.16	0.05	-4.06	0.92	-4.69	0.86
132	702062	6	4837	0.39	0.15	0.20	0.26	0.39	0.00	0.00	0.44	-0.23	-0.33	0.00	0.44	A-	A-	A+	A-	1.20	0.05	-2.77	0.94	-0.21	0.99
133	702056	6	4837	0.63	0.05	0.14	0.63	0.19	0.00	0.00	0.34	-0.20	-0.12	0.34	-0.20	A-	A-	A-	A+	0.00	0.05	4.35	1.10	2.74	1.10
134	702093	6	4837	0.81	0.07	0.81	0.04	0.08	0.00	0.00	0.50	-0.29	0.50	-0.24	-0.27	A+	A-	A-	A+	-1.13	0.06	-6.02	0.80	-5.72	0.65
135	703991	6	4837	0.39	0.15	0.39	0.36	0.11	0.00	0.00	0.37	-0.13	0.37	-0.15	-0.20	A-	A+	A+	A-	1.21	0.05	2.52	1.05	5.13	1.17
136	703262	6	4837	0.59	0.12	0.59	0.14	0.15	0.00	0.00	0.37	-0.19	0.37	-0.26	-0.08	A-	A-	A-	A+	0.17	0.05	2.95	1.06	1.41	1.05
137	703146	6	4837	0.33	0.43	0.11	0.33	0.14	0.00	0.00	0.16	0.09	-0.09	0.16	-0.25	A-	A+	A-	A+	1.54	0.05	9.06	1.22	9.90	1.44
138	703150	6	4837	0.77	0.10	0.07	0.06	0.77	0.00	0.00	0.48	-0.27	-0.24	-0.24	0.48	A+	A-	A+	A+	-0.86	0.06	-5.89	0.83	-5.23	0.72
139	641291	6	4837	0.27	0.27	0.15	0.36	0.22	0.00	0.00	0.15	0.15	-0.25	-0.06	0.12	A-	A+	A+	A+	1.87	0.06	7.62	1.21	9.80	1.51
140	702158	6	4837	0.38	0.12	0.37	0.12	0.38	0.00	0.00	0.37	-0.20	-0.04	-0.29	0.37	A+	A+	A+	A-	1.24	0.05	0.88	1.02	3.09	1.10
141	677870	6	4837	0.37	0.45	0.37	0.05	0.12	0.00	0.00	0.26	-0.08	0.26	-0.19	-0.12	A+	A-	A+	A+	1.28	0.05	4.20	1.09	6.21	1.21
142	678980	6	4837	0.65	0.12	0.07	0.16	0.65	0.00	0.00	0.44	-0.17	-0.28	-0.22	0.44	A+	A-	A-	A+	-0.13	0.05	-1.44	0.97	-1.39	0.95
143	703152	6	4837	0.50	0.50	0.11	0.17	0.22	0.00	0.00	0.49	0.49	-0.25	-0.29	-0.13	A-	A+	A+	A+	0.63	0.05	-4.90	0.91	-4.32	0.89
144	702735	6	4837	0.61	0.07	0.61	0.19	0.14	0.00	0.00	0.31	-0.23	0.31	-0.11	-0.15	A-	A-	A-	A-	0.08	0.05	6.33	1.14	6.46	1.23
145	702195	7	4788	0.33	0.16	0.34	0.17	0.33	0.00	0.00	0.34	-0.19	0.00	-0.23	0.34	A-	A+	A-	A-	1.53	0.05	1.71	1.04	4.60	1.18
146	702197	7	4788	0.48	0.11	0.13	0.48	0.28	0.00	0.00	0.30	0.07	-0.20	0.30	-0.23	A-	A-	A-	A+	0.77	0.05	6.88	1.14	6.99	1.21
147	702094	7	4788	0.69	0.13	0.10	0.69	0.08	0.00	0.00	0.42	-0.17	-0.28	0.42	-0.19	A-	A-	A-	A+	-0.34	0.05	-0.61	0.99	-1.81	0.92
148	703359	7	4788	0.45	0.25	0.45	0.17	0.13	0.00	0.00	0.32	-0.14	0.32	-0.11	-0.16	A-	A-	A-	A+	0.92	0.05	4.47	1.09	4.07	1.12
149	683619	7	4788	0.46	0.11	0.22	0.21	0.46	0.00	0.00	0.25	-0.15	-0.10	-0.09	0.25	A+	A-	A-	A-	0.84	0.05	9.51	1.20	9.90	1.33
150	679984	7	4788	0.42	0.42	0.28	0.20	0.10	0.00	0.00	0.20	0.20	-0.02	-0.04	-0.25	A+	A-	A+	A+	1.06	0.05	9.90	1.23	9.90	1.35
151	700898	7	4788	0.58	0.06	0.16	0.20	0.58	0.00	0.00	0.40	-0.13	-0.13	-0.29	0.40	A+	A-	A-	A+	0.25	0.05	1.92	1.04	1.28	1.04
152	678414	7	4788	0.43	0.21	0.25	0.43	0.12	0.00	0.00	0.37	-0.22	-0.05	0.37	-0.23	A-	A+	A+	A-	1.03	0.05	3.13	1.06	3.96	1.12
153	703505	7	4788	0.63	0.17	0.63	0.09	0.11	0.00	0.00	0.41	-0.19	0.41	-0.26	-0.17	A+	A-	A-	A-	-0.01	0.05	-0.62	0.99	-0.08	1.00
154	703506	7	4788	0.53	0.13	0.21	0.53	0.13	0.00	0.00	0.44	-0.18	-0.24	0.44	-0.18	A+	A-	A-	A-	0.51	0.05	-1.52	0.97	-1.81	0.95
155	678294	7	4788	0.53	0.19	0.53	0.12	0.17	0.00	0.00	0.47	-0.21	0.47	-0.17	-0.26	A+	A+	A-	A+	0.51	0.05	-4.45	0.92	-4.19	0.88
156	703153	7	4788	0.49	0.16	0.49	0.18	0.17	0.00	0.00	0.35	-0.18	0.35	-0.26	-0.02	A+	A-	A+	A+	0.71	0.05	3.29	1.07	3.04	1.09
157	702736	7	4788	0.15	0.15	0.29	0.25	0.31	0.00	0.00	0.02	0.02	0.06	-0.09	0.01	A-	A-	A-	A-	2.74	0.07	4.02	1.17	9.90	2.15
158	683547	7	4788	0.82	0.04	0.06	0.82	0.08	0.00	0.00	0.35	-0.26	-0.25	0.35	-0.09	B+	A-	A-	B+	-1.21	0.06	-1.60	0.94	1.27	1.09
159	642349	7	4788	0.28	0.32	0.15	0.28	0.25	0.00	0.00	0.29	-0.08	-0.18	0.29	-0.06	A-	A+	A-	A+	1.84	0.05	1.75	1.05	6.19	1.30

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
160	681246	7	4788	0.67	0.09	0.67	0.13	0.11	0.00	0.00	0.54	-0.29	0.54	-0.27	-0.25	A+	A-	A-	A+	-0.21	0.05	-5.93	0.87	-6.16	0.77
161	702200	8	4769	0.24	0.23	0.24	0.46	0.07	0.00	0.00	-0.08	-0.10	-0.08	0.25	-0.20	A-	A-	A+	A+	2.09	0.06	9.90	1.39	9.90	2.05
162	702196	8	4769	0.27	0.22	0.27	0.20	0.31	0.00	0.00	0.26	-0.03	0.26	-0.13	-0.11	A-	A+	A+	A-	1.89	0.06	1.68	1.04	6.49	1.32
163	702063	8	4769	0.62	0.62	0.26	0.08	0.04	0.00	0.00	0.34	0.34	-0.17	-0.23	-0.14	A-	A-	A-	A-	0.05	0.05	4.42	1.10	2.97	1.10
164	678865	8	4769	0.61	0.10	0.61	0.15	0.14	0.00	0.00	0.46	-0.17	0.46	-0.29	-0.19	A-	B-	A-	A-	0.12	0.05	-1.29	0.97	-2.50	0.92
165	703493	8	4769	0.75	0.75	0.13	0.06	0.05	0.00	0.00	0.57	0.57	-0.36	-0.29	-0.24	A+	A+	A-	A+	-0.71	0.06	-8.23	0.78	-8.10	0.62
166	680246	8	4769	0.47	0.12	0.15	0.47	0.26	0.00	0.00	0.30	-0.13	-0.16	0.30	-0.12	A-	A+	A-	A-	0.80	0.05	6.45	1.13	5.75	1.16
167	673866	8	4769	0.53	0.10	0.24	0.53	0.12	0.00	0.00	0.45	-0.17	-0.26	0.45	-0.19	A-	A-	A-	A+	0.51	0.05	-1.57	0.97	-1.46	0.96
168	678861	8	4769	0.32	0.32	0.13	0.35	0.21	0.00	0.00	0.26	0.26	-0.18	-0.01	-0.15	A-	A+	A-	A-	1.62	0.05	2.81	1.07	6.54	1.27
169	703502	8	4769	0.50	0.08	0.13	0.29	0.50	0.00	0.00	0.41	-0.19	-0.27	-0.15	0.41	A-	A-	A-	A+	0.64	0.05	-0.08	1.00	0.02	1.00
170	703504	8	4769	0.49	0.27	0.11	0.13	0.49	0.00	0.00	0.32	0.03	-0.29	-0.25	0.32	A+	A-	A-	A-	0.73	0.05	5.29	1.11	5.63	1.16
171	703608	8	4769	0.53	0.53	0.26	0.16	0.04	0.00	0.00	0.33	0.33	-0.35	0.08	-0.19	A-	A-	A-	A-	0.49	0.05	4.72	1.10	3.87	1.11
172	642388	8	4769	0.38	0.43	0.07	0.38	0.11	0.00	0.00	0.11	0.03	-0.20	0.11	-0.05	A-	A-	A+	A+	1.25	0.05	9.90	1.29	9.90	1.43
173	703154	8	4769	0.58	0.07	0.31	0.58	0.04	0.00	0.00	0.41	-0.23	-0.21	0.41	-0.22	A+	A+	A-	A-	0.28	0.05	-0.68	0.99	-0.93	0.97
174	702164	8	4769	0.41	0.08	0.41	0.35	0.16	0.00	0.00	0.43	-0.24	0.43	-0.04	-0.33	A-	A-	A-	A-	1.13	0.05	-0.83	0.98	0.92	1.03
175	703165	8	4769	0.72	0.07	0.15	0.05	0.72	0.00	0.00	0.39	-0.17	-0.21	-0.24	0.39	A+	A-	A-	A+	-0.52	0.06	-0.37	0.99	-0.82	0.96
176	677862	8	4769	0.62	0.13	0.09	0.62	0.16	0.00	0.00	0.40	-0.27	-0.15	0.40	-0.16	A+	A-	A-	A+	0.08	0.05	0.25	1.01	-1.25	0.96
177	703402	9	4730	0.76	0.76	0.12	0.08	0.04	0.00	0.00	0.43	0.43	-0.29	-0.20	-0.20	A-	A-	A-	A-	-0.73	0.06	-2.45	0.93	-1.83	0.91
178	703403	9	4730	0.44	0.17	0.16	0.23	0.44	0.00	0.00	0.29	-0.03	-0.18	-0.16	0.29	A+	A+	A-	A+	0.96	0.05	5.62	1.11	4.52	1.12
179	678883	9	4730	0.76	0.07	0.76	0.10	0.07	0.00	0.00	0.42	-0.22	0.42	-0.21	-0.24	A+	A+	A-	A+	-0.76	0.06	-2.05	0.94	-3.17	0.84
180	703264	9	4730	0.39	0.17	0.21	0.23	0.39	0.00	0.00	0.25	-0.10	-0.02	-0.18	0.25	A-	A-	A-	A-	1.17	0.05	7.20	1.15	7.19	1.22
181	702078	9	4730	0.40	0.37	0.14	0.09	0.40	0.00	0.00	0.28	-0.14	-0.08	-0.15	0.28	A+	A-	A-	A+	1.12	0.05	5.79	1.12	5.64	1.17
182	703492	9	4730	0.56	0.12	0.12	0.56	0.20	0.00	0.00	0.36	-0.15	-0.12	0.36	-0.23	A+	A-	A-	A-	0.35	0.05	2.42	1.05	1.09	1.03
183	702220	9	4730	0.34	0.10	0.14	0.42	0.34	0.00	0.00	0.33	-0.22	-0.26	0.00	0.33	A+	A+	A+	A-	1.47	0.05	1.67	1.04	5.46	1.20
184	702198	9	4730	0.53	0.14	0.23	0.09	0.53	0.00	0.00	0.28	-0.21	-0.05	-0.14	0.28	A+	A+	A-	A+	0.49	0.05	7.71	1.16	7.07	1.20
185	703189	9	4730	0.53	0.53	0.20	0.13	0.13	0.00	0.00	0.40	0.40	-0.23	-0.21	-0.11	A+	A+	A-	A-	0.47	0.05	0.49	1.01	0.13	1.00
186	701056	9	4730	0.45	0.38	0.09	0.45	0.07	0.00	0.00	0.19	0.01	-0.15	0.19	-0.20	A+	A-	A+	A+	0.88	0.05	9.90	1.26	9.90	1.37
187	703160	9	4730	0.47	0.47	0.19	0.16	0.19	0.00	0.00	0.35	0.35	-0.13	-0.13	-0.18	A+	A-	A-	A+	0.80	0.05	0.73	1.01	0.73	1.02
188	700900	9	4730	0.60	0.08	0.15	0.60	0.17	0.00	0.00	0.22	-0.21	-0.19	0.22	0.05	A+	A+	A-	A-	0.13	0.05	9.12	1.20	9.90	1.38
189	702165	9	4730	0.32	0.16	0.32	0.13	0.39	0.00	0.00	0.11	-0.04	0.11	-0.01	-0.06	A+	A+	A+	A-	1.56	0.05	8.40	1.20	9.90	1.41
190	702730	9	4730	0.23	0.23	0.34	0.19	0.23	0.00	0.00	0.10	-0.15	0.08	-0.04	0.10	A-	A+	A+	A+	2.10	0.06	6.78	1.21	9.90	1.69
191	700897	9	4730	0.66	0.09	0.13	0.66	0.12	0.00	0.00	0.51	-0.27	-0.23	0.51	-0.26	A+	A-	A-	A+	-0.19	0.05	-5.11	0.89	-5.76	0.80

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
192	704206	9	4730	0.54	0.54	0.19	0.07	0.21	0.00	0.00	0.40	0.40	-0.13	-0.29	-0.19	A-	A-	A-	A+	0.46	0.05	-1.45	0.97	-1.42	0.96
193	702690	10	4742	0.64	0.07	0.64	0.09	0.20	0.00	0.00	0.40	-0.22	0.40	-0.27	-0.15	A+	B-	A-	A-	-0.05	0.05	0.59	1.01	0.82	1.03
194	702691	10	4742	0.53	0.20	0.11	0.53	0.16	0.00	0.00	0.33	-0.14	-0.19	0.33	-0.13	A+	A+	A-	A-	0.49	0.05	3.94	1.08	2.84	1.08
195	677998	10	4742	0.49	0.27	0.49	0.14	0.09	0.00	0.00	0.27	-0.12	0.27	-0.12	-0.14	A-	A-	A+	A+	0.69	0.05	7.46	1.15	7.07	1.21
196	702096	10	4742	0.64	0.09	0.07	0.64	0.20	0.00	0.00	0.28	-0.14	-0.20	0.28	-0.10	A+	A-	A-	A+	-0.05	0.05	6.50	1.15	3.88	1.15
197	702086	10	4742	0.46	0.13	0.19	0.46	0.22	0.00	0.00	0.38	-0.21	-0.34	0.38	0.04	A-	A-	A-	A+	0.86	0.05	0.90	1.02	1.43	1.04
198	673867	10	4742	0.68	0.68	0.08	0.06	0.18	0.00	0.00	0.55	0.55	-0.25	-0.27	-0.32	A-	A-	A-	A-	-0.28	0.05	-6.18	0.86	-6.11	0.76
199	702084	10	4742	0.54	0.30	0.08	0.54	0.08	0.00	0.00	0.48	-0.23	-0.27	0.48	-0.22	A-	A-	A+	A-	0.44	0.05	-2.60	0.95	-2.56	0.93
200	702203	10	4742	0.33	0.51	0.33	0.07	0.09	0.00	0.00	0.30	-0.02	0.30	-0.25	-0.25	A-	A-	A+	A-	1.54	0.05	0.69	1.02	4.04	1.15
201	701055	10	4742	0.49	0.08	0.49	0.09	0.34	0.00	0.00	0.27	-0.21	0.27	-0.17	-0.05	A-	A-	A-	A-	0.73	0.05	9.06	1.19	8.06	1.24
202	701047	10	4742	0.46	0.46	0.30	0.11	0.13	0.00	0.00	0.25	0.25	0.01	-0.23	-0.17	A-	A-	A-	A-	0.88	0.05	8.77	1.18	8.52	1.25
203	703163	10	4742	0.48	0.18	0.48	0.14	0.19	0.00	0.00	0.34	-0.10	0.34	-0.29	-0.07	A+	A+	A-	A-	0.74	0.05	3.52	1.07	3.96	1.11
204	704202	10	4742	0.70	0.09	0.14	0.70	0.07	0.00	0.00	0.49	-0.24	-0.28	0.49	-0.23	A-	A-	A-	A-	-0.41	0.05	-4.42	0.89	-4.57	0.81
205	700899	10	4742	0.33	0.15	0.19	0.33	0.33	0.00	0.00	0.29	-0.12	-0.21	-0.01	0.29	A-	A-	A+	A-	1.58	0.05	2.85	1.07	4.84	1.19
206	679256	10	4742	0.21	0.21	0.24	0.26	0.29	0.00	0.00	0.18	0.18	0.10	-0.17	-0.08	A-	A+	A+	A-	2.28	0.06	2.82	1.09	8.63	1.57
207	703158	10	4742	0.56	0.56	0.23	0.13	0.09	0.00	0.00	0.43	0.43	-0.20	-0.23	-0.19	A+	A-	A-	A+	0.36	0.05	-1.26	0.97	-2.22	0.94
208	704211	10	4742	0.72	0.72	0.18	0.06	0.03	0.00	0.00	0.57	0.57	-0.43	-0.19	-0.23	A-	A-	A-	A+	-0.52	0.06	-9.02	0.78	-8.25	0.65
209	701955	11	4797	0.30	0.16	0.10	0.30	0.43	0.00	0.00	0.04	-0.20	-0.14	0.04	0.21	A+	A+	A+	A-	1.70	0.05	9.90	1.27	9.90	1.65
210	701950	11	4797	0.47	0.18	0.47	0.10	0.25	0.00	0.00	0.30	-0.05	0.30	-0.28	-0.10	A-	A+	A+	A-	0.82	0.05	5.33	1.11	4.48	1.13
211	702087	11	4797	0.56	0.16	0.11	0.56	0.16	0.00	0.00	0.45	-0.24	-0.28	0.45	-0.11	A-	A+	A-	A+	0.34	0.05	-2.09	0.96	-2.23	0.94
212	703484	11	4797	0.53	0.53	0.14	0.22	0.11	0.00	0.00	0.43	0.43	-0.30	-0.14	-0.16	A-	A+	A-	A-	0.52	0.05	-0.34	0.99	-1.14	0.97
213	643360	11	4797	0.39	0.07	0.49	0.39	0.05	0.00	0.00	0.08	-0.21	0.13	0.08	-0.21	A+	A+	A-	A+	1.22	0.05	9.90	1.35	9.90	1.54
214	702066	11	4797	0.36	0.21	0.08	0.36	0.36	0.00	0.00	0.26	-0.23	-0.17	0.26	0.03	A-	A-	A-	A+	1.36	0.05	3.52	1.08	4.92	1.17
215	703494	11	4797	0.38	0.38	0.18	0.09	0.35	0.00	0.00	0.15	0.15	-0.32	-0.13	0.18	A-	A-	A-	A+	1.28	0.05	9.90	1.29	9.90	1.41
216	703239	11	4797	0.49	0.21	0.17	0.13	0.49	0.00	0.00	0.39	-0.20	-0.17	-0.15	0.39	A+	A-	A-	A+	0.68	0.05	0.90	1.02	0.68	1.02
217	701046	11	4797	0.29	0.30	0.21	0.29	0.20	0.00	0.00	0.11	0.18	-0.13	0.11	-0.19	A+	A-	A+	A+	1.73	0.05	8.20	1.21	9.90	1.51
218	701054	11	4797	0.56	0.13	0.56	0.26	0.06	0.00	0.00	0.32	-0.22	0.32	-0.09	-0.18	A+	A-	A-	A-	0.37	0.05	6.37	1.13	4.83	1.15
219	704203	11	4797	0.73	0.04	0.16	0.07	0.73	0.00	0.00	0.39	-0.24	-0.19	-0.22	0.39	A-	C-	B-	A+	-0.58	0.06	-0.49	0.99	0.46	1.02
220	704791	11	4797	0.79	0.79	0.11	0.05	0.05	0.00	0.00	0.48	0.48	-0.27	-0.27	-0.22	A-	B-	A-	A-	-0.94	0.06	-4.58	0.86	-4.45	0.75
221	703167	11	4797	0.55	0.10	0.11	0.55	0.24	0.00	0.00	0.41	-0.22	-0.31	0.41	-0.09	A-	A+	A+	A+	0.40	0.05	1.09	1.02	-0.23	0.99
222	702738	11	4797	0.62	0.62	0.20	0.06	0.11	0.00	0.00	0.21	0.21	-0.06	-0.22	-0.06	A+	A-	A-	A+	0.02	0.05	8.84	1.20	9.14	1.35
223	679978	11	4797	0.73	0.05	0.73	0.18	0.04	0.00	0.00	0.26	-0.14	0.26	-0.11	-0.19	A-	A-	A-	A-	-0.55	0.06	6.32	1.18	5.66	1.30

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
224	678895	11	4797	0.27	0.28	0.32	0.13	0.27	0.00	0.00	0.25	0.05	-0.09	-0.27	0.25	A+	A-	A+	A-	1.88	0.05	2.39	1.06	5.72	1.28
225	701954	12	4751	0.61	0.61	0.15	0.17	0.07	0.00	0.00	0.54	0.54	-0.25	-0.30	-0.25	A-	A-	A-	A-	0.04	0.05	-5.73	0.88	-5.99	0.81
226	701952	12	4751	0.44	0.10	0.44	0.17	0.28	0.00	0.00	0.34	-0.37	0.34	-0.08	-0.05	A-	A+	A-	A+	0.93	0.05	3.17	1.06	3.13	1.09
227	702067	12	4751	0.49	0.10	0.20	0.20	0.49	0.00	0.00	0.38	-0.16	-0.30	-0.06	0.38	A-	A-	A-	A-	0.66	0.05	0.80	1.02	0.99	1.03
228	703483	12	4751	0.59	0.59	0.19	0.11	0.11	0.00	0.00	0.46	0.46	-0.16	-0.26	-0.26	A-	A-	A+	A+	0.17	0.05	-1.58	0.97	-2.19	0.93
229	680194	12	4751	0.72	0.05	0.06	0.72	0.17	0.00	0.00	0.55	-0.26	-0.27	0.55	-0.33	A-	A-	A-	A-	-0.54	0.05	-7.55	0.82	-7.91	0.67
230	703240	12	4751	0.52	0.52	0.19	0.21	0.09	0.00	0.00	0.39	0.39	-0.18	-0.22	-0.13	A-	A+	A+	A+	0.52	0.05	1.37	1.03	0.84	1.02
231	703495	12	4751	0.79	0.06	0.07	0.09	0.79	0.00	0.00	0.41	-0.24	-0.26	-0.16	0.41	A+	A-	A-	A+	-0.97	0.06	-2.36	0.93	-1.45	0.91
232	703700	12	4751	0.22	0.12	0.22	0.41	0.25	0.00	0.00	0.15	-0.27	0.15	0.12	-0.06	A-	A+	A+	A-	2.21	0.06	3.46	1.11	9.90	1.73
233	701419	12	4751	0.47	0.33	0.08	0.12	0.47	0.00	0.00	0.48	-0.27	-0.19	-0.18	0.48	A-	A-	A+	A+	0.79	0.05	-3.48	0.93	-1.67	0.95
234	701414	12	4751	0.66	0.13	0.11	0.10	0.66	0.00	0.00	0.44	-0.23	-0.22	-0.21	0.44	A+	A-	A-	A-	-0.18	0.05	-0.81	0.98	0.46	1.02
235	704789	12	4751	0.63	0.07	0.18	0.11	0.63	0.00	0.00	0.47	-0.27	-0.27	-0.17	0.47	A+	A-	A-	A-	-0.04	0.05	-2.55	0.95	-1.74	0.94
236	704204	12	4751	0.62	0.11	0.16	0.62	0.11	0.00	0.00	0.47	-0.22	-0.25	0.47	-0.21	A-	A-	A-	A+	-0.01	0.05	-3.49	0.93	-3.91	0.87
237	678886	12	4751	0.48	0.48	0.22	0.17	0.14	0.00	0.00	0.31	0.31	-0.17	-0.20	-0.03	A+	A-	A-	A-	0.75	0.05	6.05	1.12	5.30	1.15
238	704780	12	4751	0.35	0.20	0.35	0.36	0.09	0.00	0.00	0.23	0.05	0.23	-0.11	-0.26	A+	A-	A-	A+	1.40	0.05	6.27	1.14	8.48	1.32
239	641194	12	4751	0.23	0.44	0.23	0.15	0.18	0.00	0.00	0.21	0.02	0.21	-0.16	-0.10	A-	A-	A-	A-	2.14	0.06	2.99	1.09	9.90	1.65
240	678990	12	4751	0.22	0.22	0.16	0.47	0.15	0.00	0.00	-0.02	-0.02	-0.19	0.23	-0.10	A-	A-	A-	A+	2.22	0.06	9.90	1.36	9.90	1.89
241	701953	13	4804	0.40	0.25	0.20	0.40	0.15	0.00	0.00	0.33	-0.12	-0.23	0.33	-0.05	A-	A-	A+	A-	1.15	0.05	3.92	1.08	5.29	1.17
242	701951	13	4804	0.68	0.68	0.09	0.13	0.10	0.00	0.00	0.43	0.43	-0.23	-0.21	-0.20	A+	A-	A-	A+	-0.30	0.05	-0.88	0.98	-1.42	0.94
243	702081	13	4804	0.76	0.09	0.07	0.76	0.07	0.00	0.00	0.45	-0.30	-0.22	0.45	-0.18	A-	A-	A-	A+	-0.82	0.06	-3.75	0.89	-2.92	0.84
244	702068	13	4804	0.47	0.16	0.47	0.20	0.18	0.00	0.00	0.41	-0.33	0.41	-0.15	-0.07	A+	A+	A-	A-	0.80	0.05	0.17	1.00	1.16	1.03
245	703701	13	4804	0.48	0.48	0.13	0.19	0.19	0.00	0.00	0.41	0.41	-0.32	-0.07	-0.17	A-	A+	A-	A-	0.74	0.05	1.09	1.02	0.49	1.01
246	702204	13	4804	0.59	0.12	0.59	0.18	0.11	0.00	0.00	0.48	-0.19	0.48	-0.25	-0.26	A-	A-	A-	A+	0.18	0.05	-4.02	0.92	-4.44	0.86
247	674098	13	4804	0.71	0.11	0.71	0.08	0.10	0.00	0.00	0.49	-0.23	0.49	-0.27	-0.25	A+	A+	A-	A+	-0.49	0.05	-3.34	0.92	-3.54	0.84
248	703241	13	4804	0.52	0.13	0.52	0.08	0.27	0.00	0.00	0.45	-0.26	0.45	-0.22	-0.17	A-	A-	A-	A+	0.54	0.05	-0.52	0.99	-0.88	0.97
249	701418	13	4804	0.24	0.24	0.31	0.19	0.25	0.00	0.00	0.11	0.11	-0.08	-0.17	0.14	A-	A-	B-	A-	2.08	0.06	7.19	1.22	9.90	1.69
250	702637	13	4804	0.44	0.24	0.44	0.20	0.12	0.00	0.00	0.29	-0.07	0.29	-0.14	-0.17	A-	A-	A-	A-	0.96	0.05	6.93	1.14	7.09	1.22
251	678300	13	4804	0.60	0.07	0.17	0.16	0.60	0.00	0.00	0.51	-0.23	-0.19	-0.33	0.51	A+	A+	A-	A-	0.13	0.05	-4.34	0.91	-3.79	0.88
252	705226	13	4804	0.29	0.29	0.27	0.08	0.35	0.00	0.00	0.26	0.26	-0.01	-0.19	-0.12	A-	B-	A-	A-	1.77	0.05	4.35	1.11	7.21	1.34
253	704205	13	4804	0.62	0.12	0.62	0.17	0.08	0.00	0.00	0.53	-0.21	0.53	-0.32	-0.23	A+	A-	A-	A+	0.01	0.05	-5.75	0.88	-5.53	0.82
254	704256	13	4804	0.60	0.60	0.14	0.16	0.10	0.00	0.00	0.43	0.43	-0.22	-0.18	-0.23	A+	A-	A-	A-	0.13	0.05	0.50	1.01	-0.37	0.99
255	642870	13	4804	0.53	0.20	0.12	0.15	0.53	0.00	0.00	0.45	-0.33	-0.15	-0.12	0.45	A+	A-	A+	A+	0.48	0.05	-0.47	0.99	-0.84	0.98

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
256	683556	13	4804	0.76	0.06	0.08	0.09	0.76	0.00	0.00	0.56	-0.28	-0.32	-0.29	0.56	A-	A-	A-	A+	-0.81	0.06	-8.73	0.76	-7.21	0.64
257	704371	14	4806	0.31	0.34	0.31	0.13	0.22	0.00	0.00	0.29	-0.05	0.29	-0.08	-0.20	A-	A-	A-	A-	1.60	0.05	4.09	1.10	7.44	1.31
258	704381	14	4806	0.41	0.41	0.41	0.09	0.09	0.00	0.00	0.04	0.04	0.22	-0.28	-0.15	A-	A+	A-	A+	1.08	0.05	9.90	1.38	9.90	1.53
259	684083	14	4806	0.54	0.54	0.21	0.21	0.05	0.00	0.00	0.24	0.24	-0.16	-0.06	-0.13	A+	A+	A+	A+	0.44	0.05	9.00	1.19	8.08	1.24
260	678902	14	4806	0.51	0.13	0.51	0.13	0.22	0.00	0.00	0.31	-0.09	0.31	-0.16	-0.16	A-	A+	A+	A+	0.58	0.05	5.17	1.10	4.15	1.12
261	702061	14	4806	0.55	0.16	0.17	0.12	0.55	0.00	0.00	0.46	-0.18	-0.19	-0.27	0.46	A-	A-	A-	A+	0.36	0.05	-2.53	0.95	-2.90	0.92
262	702199	14	4806	0.42	0.42	0.24	0.26	0.09	0.00	0.00	0.24	0.24	-0.27	0.11	-0.17	A+	A+	A-	A-	1.04	0.05	7.78	1.16	7.08	1.21
263	702205	14	4806	0.59	0.16	0.16	0.59	0.10	0.00	0.00	0.43	-0.17	-0.29	0.43	-0.16	A-	A-	A-	A+	0.19	0.05	-0.47	0.99	-2.01	0.94
264	703702	14	4806	0.42	0.42	0.23	0.11	0.24	0.00	0.00	0.31	0.31	-0.09	-0.20	-0.13	A-	A+	A+	A-	1.00	0.05	6.08	1.13	5.81	1.17
265	701417	14	4806	0.52	0.52	0.07	0.20	0.20	0.00	0.00	0.37	0.37	-0.24	-0.12	-0.17	A-	A-	A+	A-	0.51	0.05	1.82	1.04	0.66	1.02
266	701416	14	4806	0.62	0.13	0.20	0.05	0.62	0.00	0.00	0.40	-0.22	-0.17	-0.25	0.40	A+	A+	A+	A-	0.03	0.05	1.38	1.03	2.06	1.07
267	704209	14	4806	0.63	0.63	0.03	0.07	0.28	0.00	0.00	0.19	0.19	-0.15	-0.17	-0.05	A+	A-	A-	A+	-0.01	0.05	8.70	1.20	9.73	1.37
268	702725	14	4806	0.59	0.26	0.04	0.11	0.59	0.00	0.00	0.38	-0.17	-0.23	-0.21	0.38	B-	A-	A-	A-	0.16	0.05	2.63	1.05	1.88	1.06
269	674123	14	4806	0.51	0.08	0.51	0.23	0.18	0.00	0.00	0.38	-0.16	0.38	-0.23	-0.13	A-	A-	A-	A-	0.55	0.05	1.38	1.03	0.98	1.03
270	678982	14	4806	0.61	0.06	0.24	0.09	0.61	0.00	0.00	0.42	-0.27	-0.19	-0.20	0.42	A+	A+	A+	A+	0.08	0.05	-0.01	1.00	1.60	1.05
271	702154	14	4806	0.61	0.12	0.17	0.09	0.61	0.00	0.00	0.44	-0.17	-0.23	-0.24	0.44	A+	A-	A+	A+	0.07	0.05	-3.42	0.93	-2.62	0.92
272	642363	14	4806	0.49	0.09	0.49	0.16	0.26	0.00	0.00	0.22	-0.06	0.22	-0.12	-0.11	A-	A-	A+	A+	0.69	0.05	9.64	1.20	8.23	1.23
273	704377	15	4776	0.76	0.10	0.09	0.05	0.76	0.00	0.00	0.47	-0.30	-0.28	-0.15	0.47	A+	A-	A-	A-	-0.79	0.06	-3.05	0.91	-3.43	0.81
274	704379	15	4776	0.43	0.26	0.07	0.43	0.24	0.00	0.00	0.29	-0.13	-0.21	0.29	-0.07	A-	A+	A-	A-	0.97	0.05	8.19	1.17	8.83	1.28
275	702083	15	4776	0.52	0.15	0.52	0.16	0.17	0.00	0.00	0.39	-0.18	0.39	-0.16	-0.19	A+	A-	A+	A-	0.52	0.05	2.28	1.05	2.63	1.08
276	678939	15	4776	0.57	0.15	0.09	0.20	0.57	0.00	0.00	0.50	-0.30	-0.28	-0.16	0.50	A+	A-	A-	A+	0.27	0.05	-4.47	0.91	-3.18	0.90
277	703360	15	4776	0.50	0.10	0.25	0.50	0.14	0.00	0.00	0.50	-0.17	-0.21	0.50	-0.31	A-	A+	A+	A+	0.61	0.05	-6.08	0.88	-4.97	0.87
278	702099	15	4776	0.66	0.66	0.05	0.18	0.10	0.00	0.00	0.40	0.40	-0.26	-0.19	-0.19	A+	A-	A-	A+	-0.21	0.05	1.15	1.03	1.48	1.06
279	703708	15	4776	0.46	0.14	0.26	0.14	0.46	0.00	0.00	0.46	-0.24	-0.18	-0.18	0.46	A+	A+	A+	A+	0.84	0.05	-2.69	0.95	-1.92	0.95
280	702070	15	4776	0.52	0.52	0.17	0.19	0.12	0.00	0.00	0.43	0.43	-0.16	-0.16	-0.28	A+	A+	A+	A-	0.53	0.05	-1.41	0.97	-0.96	0.97
281	702615	15	4776	0.50	0.50	0.08	0.07	0.34	0.00	0.00	0.18	0.18	-0.24	-0.28	0.11	A+	A-	A-	A-	0.63	0.05	9.90	1.30	9.90	1.44
282	702604	15	4776	0.36	0.42	0.15	0.07	0.36	0.00	0.00	0.14	0.16	-0.22	-0.25	0.14	A-	A-	A-	A-	1.37	0.05	9.90	1.28	9.90	1.43
283	704210	15	4776	0.66	0.66	0.17	0.07	0.09	0.00	0.00	0.44	0.44	-0.19	-0.27	-0.24	A-	A-	A-	A+	-0.21	0.05	-1.26	0.97	-1.41	0.94
284	642907	15	4776	0.60	0.11	0.11	0.60	0.17	0.00	0.00	0.36	-0.23	-0.20	0.36	-0.11	A+	A-	A-	A-	0.10	0.05	2.88	1.06	1.90	1.07
285	704786	15	4776	0.75	0.08	0.09	0.08	0.75	0.00	0.00	0.54	-0.28	-0.27	-0.30	0.54	A+	A-	A-	A-	-0.72	0.06	-4.81	0.87	-4.82	0.76
286	702156	15	4776	0.50	0.07	0.19	0.25	0.50	0.00	0.00	0.41	-0.22	-0.09	-0.26	0.41	A+	A-	A-	A+	0.64	0.05	1.24	1.02	0.50	1.01
287	679975	15	4776	0.65	0.07	0.07	0.65	0.21	0.00	0.00	0.54	-0.24	-0.25	0.54	-0.32	A-	B-	A-	B+	-0.12	0.05	-7.42	0.84	-6.96	0.75

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
288	702726	15	4776	0.59	0.10	0.59	0.09	0.22	0.00	0.00	0.40	-0.24	0.40	-0.29	-0.11	A-	A-	A-	A-	0.16	0.05	1.22	1.03	0.91	1.03
289	703672	16	4785	0.30	0.26	0.27	0.30	0.17	0.00	0.00	0.14	0.00	-0.10	0.14	-0.05	A+	A+	A-	A+	1.70	0.05	5.93	1.15	8.74	1.39
290	703669	16	4785	0.67	0.15	0.07	0.11	0.67	0.00	0.00	0.39	-0.24	-0.20	-0.14	0.39	A+	A-	A-	A+	-0.22	0.05	1.53	1.04	0.33	1.01
291	678543	16	4785	0.65	0.10	0.08	0.65	0.17	0.00	0.00	0.44	-0.23	-0.29	0.44	-0.15	A-	A+	A-	A-	-0.13	0.05	-1.90	0.96	-2.82	0.90
292	702088	16	4785	0.54	0.16	0.21	0.54	0.09	0.00	0.00	0.13	-0.26	0.13	0.13	-0.06	A-	A+	A+	A+	0.43	0.05	9.90	1.31	9.90	1.40
293	702098	16	4785	0.85	0.05	0.05	0.04	0.85	0.00	0.00	0.46	-0.25	-0.25	-0.25	0.46	A+	A-	B-	A+	-1.48	0.07	-4.24	0.83	-5.01	0.64
294	702211	16	4785	0.31	0.31	0.28	0.28	0.13	0.00	0.00	0.28	0.28	-0.15	0.01	-0.19	A-	A+	A+	A-	1.63	0.05	2.82	1.07	5.84	1.24
295	702257	16	4785	0.59	0.07	0.07	0.59	0.28	0.00	0.00	0.40	-0.22	-0.21	0.40	-0.19	A-	A-	A-	A-	0.20	0.05	0.74	1.01	-0.09	1.00
296	679205	16	4785	0.72	0.72	0.07	0.13	0.08	0.00	0.00	0.46	0.46	-0.27	-0.21	-0.24	A+	A+	A+	A-	-0.50	0.05	-2.74	0.93	-3.33	0.85
297	702606	16	4785	0.29	0.21	0.10	0.39	0.29	0.00	0.00	0.16	0.02	-0.18	-0.05	0.16	A+	A+	A-	A-	1.75	0.05	5.77	1.15	7.15	1.33
298	702612	16	4785	0.41	0.08	0.38	0.41	0.12	0.00	0.00	0.25	-0.21	0.02	0.25	-0.23	A-	A-	A+	A+	1.07	0.05	7.58	1.16	6.77	1.21
299	704787	16	4785	0.71	0.08	0.13	0.71	0.08	0.00	0.00	0.37	-0.22	-0.20	0.37	-0.15	A-	A+	A+	A-	-0.44	0.05	1.30	1.03	0.23	1.01
300	702162	16	4785	0.27	0.27	0.23	0.21	0.28	0.00	0.00	0.08	0.08	-0.08	0.03	-0.02	A-	A-	A+	A-	1.85	0.05	8.91	1.25	9.90	1.70
301	642890	16	4785	0.30	0.30	0.17	0.22	0.31	0.00	0.00	0.28	0.28	-0.11	-0.24	0.02	A-	A-	A-	A-	1.66	0.05	3.65	1.09	6.35	1.27
302	703162	16	4785	0.50	0.11	0.22	0.16	0.50	0.00	0.00	0.57	-0.21	-0.25	-0.30	0.57	A+	A-	A-	A+	0.62	0.05	-8.35	0.85	-7.11	0.82
303	643406	16	4785	0.36	0.36	0.36	0.15	0.13	0.00	0.00	0.06	0.13	0.06	-0.18	-0.08	A+	A+	A+	A-	1.35	0.05	9.90	1.35	9.90	1.52
304	702727	16	4785	0.71	0.71	0.08	0.10	0.11	0.00	0.00	0.47	0.47	-0.25	-0.15	-0.31	A-	A-	A-	A+	-0.46	0.05	-4.68	0.89	0.06	1.00
305	703671	17	4773	0.77	0.07	0.77	0.13	0.03	0.00	0.00	0.36	-0.28	0.36	-0.15	-0.17	A-	A-	A-	A-	-0.80	0.06	1.10	1.03	1.92	1.11
306	703668	17	4773	0.70	0.70	0.14	0.09	0.07	0.00	0.00	0.45	0.45	-0.16	-0.31	-0.24	A-	A-	A+	A-	-0.35	0.05	-1.36	0.97	-1.65	0.93
307	678863	17	4773	0.52	0.19	0.52	0.23	0.06	0.00	0.00	0.48	-0.22	0.48	-0.25	-0.18	A-	A+	A-	A+	0.61	0.05	-3.58	0.93	-3.32	0.91
308	700894	17	4773	0.40	0.40	0.28	0.08	0.24	0.00	0.00	0.12	0.12	-0.03	-0.12	-0.03	A-	A-	A+	A+	1.20	0.05	9.90	1.32	9.90	1.48
309	702258	17	4773	0.71	0.12	0.05	0.71	0.12	0.00	0.00	0.46	-0.23	-0.20	0.46	-0.26	A-	A+	A-	A-	-0.39	0.05	-1.65	0.96	-2.07	0.91
310	703515	17	4773	0.51	0.10	0.51	0.08	0.31	0.00	0.00	0.42	-0.18	0.42	-0.26	-0.19	A-	A+	A-	A-	0.65	0.05	0.89	1.02	1.39	1.04
311	702625	17	4773	0.36	0.36	0.30	0.12	0.22	0.00	0.00	0.18	0.18	0.12	-0.20	-0.19	A-	A+	A-	A+	1.39	0.05	9.90	1.23	9.90	1.40
312	703362	17	4773	0.69	0.11	0.12	0.69	0.07	0.00	0.00	0.57	-0.28	-0.33	0.57	-0.25	A-	A-	A-	A-	-0.31	0.05	-5.95	0.86	-6.39	0.75
313	702605	17	4773	0.13	0.17	0.49	0.21	0.13	0.00	0.00	-0.07	-0.10	0.09	0.04	-0.07	A-	A-	A-	A+	2.99	0.07	4.35	1.21	9.90	2.67
314	702613	17	4773	0.49	0.32	0.49	0.04	0.15	0.00	0.00	0.30	-0.06	0.30	-0.24	-0.19	A-	A-	A-	A+	0.76	0.05	5.99	1.12	5.41	1.15
315	702161	17	4773	0.78	0.03	0.06	0.13	0.78	0.00	0.00	0.35	-0.24	-0.20	-0.16	0.35	A+	A-	A-	A+	-0.88	0.06	-0.07	1.00	2.21	1.14
316	700895	17	4773	0.38	0.16	0.12	0.33	0.38	0.00	0.00	0.34	-0.25	-0.24	0.02	0.34	A+	A-	A-	A+	1.28	0.05	3.09	1.06	4.42	1.15
317	702728	17	4773	0.58	0.06	0.18	0.17	0.58	0.00	0.00	0.38	-0.23	-0.19	-0.15	0.38	A-	A-	A-	A+	0.27	0.05	0.94	1.02	0.76	1.02
318	680877	17	4773	0.67	0.67	0.19	0.07	0.07	0.00	0.00	0.47	0.47	-0.21	-0.30	-0.25	A-	A-	A-	A+	-0.20	0.05	-3.95	0.91	-2.76	0.89
319	677866	17	4773	0.33	0.17	0.18	0.33	0.32	0.00	0.00	0.21	-0.16	-0.18	0.21	0.06	A-	A+	A+	A-	1.58	0.05	5.24	1.12	7.65	1.31

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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***	**MS**	***z***	**MS**
320	674095	17	4773	0.61	0.61	0.23	0.09	0.07	0.00	0.00	0.50	0.50	-0.22	-0.29	-0.26	A-	A-	A-	A-	0.15	0.05	-2.95	0.94	-3.42	0.89
321	702614	18	4754	0.65	0.65	0.14	0.08	0.13	0.00	0.00	0.32	0.32	-0.15	-0.21	-0.12	A+	A+	A-	A-	-0.11	0.05	5.24	1.12	3.03	1.12
322	703268	18	4754	0.59	0.14	0.10	0.17	0.59	0.00	0.00	0.47	-0.19	-0.21	-0.27	0.47	A+	A+	A+	A+	0.16	0.05	-2.72	0.94	-1.71	0.94
323	702259	18	4754	0.60	0.17	0.60	0.15	0.07	0.00	0.00	0.43	-0.17	0.43	-0.30	-0.16	A-	A-	A-	A-	0.12	0.05	0.20	1.00	-1.06	0.96
324	703363	18	4754	0.54	0.13	0.20	0.54	0.13	0.00	0.00	0.46	-0.07	-0.30	0.46	-0.25	A-	A+	A-	A-	0.45	0.05	-1.29	0.97	-2.06	0.94
325	702626	18	4754	0.48	0.12	0.17	0.24	0.48	0.00	0.00	0.44	-0.24	-0.21	-0.15	0.44	A-	A-	A-	A-	0.77	0.05	-1.85	0.96	-1.52	0.96
326	703521	18	4754	0.69	0.13	0.11	0.07	0.69	0.00	0.00	0.37	-0.15	-0.27	-0.14	0.37	A+	A-	A+	A+	-0.36	0.05	1.96	1.05	1.56	1.07
327	678542	18	4754	0.65	0.19	0.08	0.65	0.08	0.00	0.00	0.42	-0.25	-0.26	0.42	-0.11	A-	A-	A-	A-	-0.12	0.05	1.88	1.04	1.23	1.05
328	703516	18	4754	0.44	0.20	0.22	0.44	0.14	0.00	0.00	0.37	-0.14	-0.16	0.37	-0.18	A-	A-	A-	A+	0.97	0.05	3.09	1.07	4.68	1.15
329	703477	18	4754	0.56	0.12	0.20	0.12	0.56	0.00	0.00	0.36	-0.19	-0.14	-0.18	0.36	A+	A-	A+	A+	0.33	0.05	4.07	1.09	2.96	1.10
330	703478	18	4754	0.71	0.71	0.06	0.12	0.10	0.00	0.00	0.36	0.36	-0.21	-0.20	-0.14	A-	A-	A-	A-	-0.48	0.05	3.15	1.08	2.72	1.14
331	704198	18	4754	0.69	0.69	0.08	0.06	0.17	0.00	0.00	0.40	0.40	-0.24	-0.29	-0.12	A-	A-	A-	A+	-0.34	0.05	0.07	1.00	0.32	1.01
332	702167	18	4754	0.42	0.19	0.20	0.19	0.42	0.00	0.00	0.34	-0.16	-0.04	-0.23	0.34	A-	A-	A+	A-	1.06	0.05	4.93	1.11	4.88	1.16
333	645053	18	4754	0.50	0.28	0.50	0.14	0.07	0.00	0.00	0.33	-0.04	0.33	-0.24	-0.23	A-	A+	A+	A-	0.63	0.05	6.01	1.13	4.91	1.15
334	679970	18	4754	0.62	0.13	0.62	0.15	0.09	0.00	0.00	0.53	-0.25	0.53	-0.26	-0.26	A-	A-	A-	A-	0.04	0.05	-5.45	0.89	-4.67	0.84
335	674162	18	4754	0.52	0.13	0.08	0.52	0.26	0.00	0.00	0.51	-0.24	-0.30	0.51	-0.21	A-	A-	B-	A+	0.54	0.05	-5.25	0.90	-4.60	0.87
336	677860	18	4754	0.52	0.09	0.24	0.15	0.52	0.00	0.00	0.37	-0.20	-0.16	-0.16	0.37	A+	A-	A+	A+	0.55	0.05	3.94	1.08	3.56	1.11
337	702607	19	4823	0.39	0.15	0.19	0.26	0.39	0.00	0.00	0.35	-0.19	-0.15	-0.10	0.35	A+	A+	A+	A-	1.23	0.05	1.32	1.03	2.08	1.07
338	702610	19	4823	0.20	0.20	0.53	0.12	0.15	0.00	0.00	-0.04	-0.04	0.28	-0.20	-0.15	A-	A-	A-	A+	2.38	0.06	8.93	1.33	9.90	2.20
339	703244	19	4823	0.75	0.07	0.05	0.75	0.13	0.00	0.00	0.42	-0.28	-0.28	0.42	-0.14	A+	A-	A-	A+	-0.68	0.06	-2.65	0.93	-0.44	0.98
340	702090	19	4823	0.55	0.55	0.06	0.04	0.35	0.00	0.00	0.30	0.30	-0.24	-0.27	-0.07	A-	A+	A-	A-	0.42	0.05	5.84	1.12	3.39	1.10
341	703517	19	4823	0.54	0.08	0.54	0.14	0.25	0.00	0.00	0.28	-0.13	0.28	-0.31	0.01	A+	A-	A-	A+	0.47	0.05	8.16	1.17	7.06	1.21
342	660917	19	4823	0.50	0.26	0.10	0.50	0.14	0.00	0.00	0.37	-0.15	-0.30	0.37	-0.08	A-	A-	A-	A+	0.69	0.05	2.63	1.05	2.49	1.07
343	702265	19	4823	0.80	0.04	0.08	0.08	0.80	0.00	0.00	0.50	-0.27	-0.28	-0.26	0.50	A+	B-	A-	A+	-1.03	0.06	-4.97	0.84	-5.45	0.69
344	703523	19	4823	0.62	0.27	0.62	0.06	0.05	0.00	0.00	0.29	-0.07	0.29	-0.25	-0.21	A+	A-	A-	A+	0.07	0.05	6.19	1.14	4.57	1.16
345	703486	19	4823	0.65	0.25	0.65	0.06	0.05	0.00	0.00	0.43	-0.26	0.43	-0.18	-0.22	A-	A-	A-	A+	-0.08	0.05	-1.47	0.97	-1.90	0.93
346	703476	19	4823	0.51	0.19	0.51	0.14	0.16	0.00	0.00	0.34	-0.03	0.34	-0.31	-0.14	A-	A-	A-	A+	0.64	0.05	5.28	1.11	4.51	1.13
347	704784	19	4823	0.45	0.26	0.45	0.16	0.12	0.00	0.00	0.33	-0.03	0.33	-0.21	-0.21	A-	A-	A-	A-	0.90	0.05	2.89	1.06	3.65	1.11
348	643401	19	4823	0.48	0.48	0.32	0.11	0.08	0.00	0.00	0.37	0.37	-0.09	-0.27	-0.20	A+	A-	A-	A-	0.74	0.05	2.12	1.04	1.17	1.03
349	674147	19	4823	0.68	0.13	0.06	0.68	0.12	0.00	0.00	0.48	-0.22	-0.28	0.48	-0.23	A-	A-	A-	A+	-0.28	0.05	-3.94	0.91	-3.82	0.85
350	681824	19	4823	0.18	0.18	0.29	0.10	0.42	0.00	0.00	0.05	0.05	-0.06	-0.30	0.21	A-	A+	A+	A-	2.48	0.06	5.48	1.20	9.90	1.92
351	704199	19	4823	0.57	0.16	0.19	0.57	0.08	0.00	0.00	0.46	-0.24	-0.16	0.46	-0.27	A-	A-	A-	A+	0.30	0.05	-2.39	0.95	-1.11	0.97

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
352	702168	19	4823	0.66	0.18	0.07	0.10	0.66	0.00	0.00	0.42	-0.24	-0.24	-0.15	0.42	A+	A-	A-	A-	-0.13	0.05	-0.89	0.98	-0.54	0.98
353	702611	20	4775	0.44	0.16	0.19	0.44	0.22	0.00	0.00	0.26	-0.14	-0.18	0.26	-0.02	A+	A-	A+	A-	0.95	0.05	6.82	1.14	6.48	1.19
354	702608	20	4775	0.46	0.27	0.46	0.15	0.12	0.00	0.00	0.23	-0.11	0.23	-0.08	-0.13	A+	A+	A-	A-	0.85	0.05	9.09	1.19	8.06	1.23
355	703248	20	4775	0.29	0.09	0.31	0.32	0.29	0.00	0.00	0.30	-0.14	0.02	-0.23	0.30	A-	A-	A-	A+	1.77	0.05	2.91	1.07	5.38	1.24
356	702072	20	4775	0.53	0.12	0.53	0.17	0.19	0.00	0.00	0.38	-0.17	0.38	-0.33	-0.04	A-	A-	A-	A-	0.51	0.05	1.72	1.03	2.54	1.07
357	703524	20	4775	0.73	0.06	0.73	0.13	0.09	0.00	0.00	0.49	-0.29	0.49	-0.24	-0.25	A-	A-	A-	A-	-0.55	0.06	-3.01	0.92	-3.39	0.85
358	702091	20	4775	0.54	0.16	0.54	0.13	0.18	0.00	0.00	0.37	-0.18	0.37	-0.21	-0.13	A-	A-	A-	A+	0.44	0.05	4.15	1.08	4.28	1.12
359	700892	20	4775	0.50	0.18	0.11	0.22	0.50	0.00	0.00	0.47	-0.19	-0.28	-0.18	0.47	A+	A+	A-	A+	0.66	0.05	-3.10	0.94	-2.01	0.95
360	702266	20	4775	0.68	0.15	0.68	0.09	0.08	0.00	0.00	0.43	-0.13	0.43	-0.30	-0.25	A+	A-	A-	A+	-0.29	0.05	-1.67	0.96	-1.15	0.95
361	701423	20	4775	0.24	0.18	0.29	0.29	0.24	0.00	0.00	0.13	-0.12	0.05	-0.08	0.13	A-	A-	A+	A-	2.07	0.06	5.92	1.18	9.90	1.69
362	701421	20	4775	0.39	0.39	0.10	0.22	0.28	0.00	0.00	0.23	0.23	-0.20	-0.17	0.04	A-	A-	A-	A-	1.18	0.05	8.41	1.18	8.33	1.27
363	704785	20	4775	0.59	0.08	0.59	0.06	0.26	0.00	0.00	0.40	-0.28	0.40	-0.31	-0.10	A-	A-	A+	A-	0.19	0.05	3.58	1.07	2.69	1.08
364	705225	20	4775	0.44	0.15	0.31	0.09	0.44	0.00	0.00	0.31	-0.23	-0.02	-0.22	0.31	A-	A-	A+	A-	0.93	0.05	6.42	1.13	6.88	1.20
365	702163	20	4775	0.43	0.20	0.30	0.07	0.43	0.00	0.00	0.41	-0.31	-0.07	-0.18	0.41	A+	A-	A-	A+	0.98	0.05	-0.60	0.99	0.73	1.02
366	678992	20	4775	0.55	0.14	0.55	0.16	0.15	0.00	0.00	0.40	-0.15	0.40	-0.21	-0.20	A-	A-	A-	A+	0.38	0.05	1.21	1.02	1.09	1.03
367	679252	20	4775	0.72	0.15	0.72	0.08	0.06	0.00	0.00	0.36	-0.17	0.36	-0.16	-0.25	A+	A-	A-	A+	-0.51	0.05	0.15	1.00	0.85	1.04
368	681053	20	4775	0.58	0.16	0.13	0.58	0.13	0.00	0.00	0.35	-0.11	-0.16	0.35	-0.23	A+	A-	A-	A+	0.27	0.05	3.90	1.08	3.73	1.11
369	703663	21	4826	0.39	0.39	0.12	0.36	0.12	0.00	0.00	0.31	0.31	-0.29	-0.06	-0.08	A-	A-	A-	A-	1.19	0.05	5.48	1.12	6.85	1.23
370	703662	21	4826	0.36	0.13	0.36	0.31	0.20	0.00	0.00	0.29	-0.20	0.29	0.09	-0.28	A-	A+	A+	A-	1.39	0.05	4.49	1.10	6.22	1.23
371	702074	21	4826	0.79	0.09	0.79	0.07	0.05	0.00	0.00	0.41	-0.24	0.41	-0.23	-0.18	A-	A-	A-	A+	-0.99	0.06	-1.20	0.96	-1.10	0.93
372	642345	21	4826	0.33	0.33	0.37	0.11	0.19	0.00	0.00	0.39	0.39	-0.05	-0.32	-0.15	B-	A-	A-	A+	1.56	0.05	-1.70	0.96	3.15	1.12
373	702206	21	4826	0.66	0.07	0.14	0.12	0.66	0.00	0.00	0.54	-0.29	-0.19	-0.35	0.54	A-	A-	A-	B-	-0.19	0.05	-4.44	0.90	-4.34	0.83
374	702213	21	4826	0.38	0.38	0.14	0.36	0.13	0.00	0.00	0.24	0.24	-0.27	0.04	-0.12	A-	A+	A-	A-	1.29	0.05	8.29	1.19	9.43	1.34
375	702059	21	4826	0.69	0.07	0.06	0.17	0.69	0.00	0.00	0.57	-0.32	-0.16	-0.37	0.57	A-	A-	A-	A+	-0.39	0.05	-7.92	0.81	-6.69	0.72
376	703526	21	4826	0.37	0.37	0.27	0.23	0.12	0.00	0.00	0.27	0.27	-0.02	-0.09	-0.25	A-	A-	A-	A-	1.30	0.05	7.41	1.17	7.09	1.25
377	701422	21	4826	0.33	0.06	0.25	0.36	0.33	0.00	0.00	0.27	-0.29	-0.14	0.01	0.27	A-	A-	A-	A-	1.53	0.05	5.70	1.14	6.25	1.25
378	701425	21	4826	0.61	0.61	0.13	0.13	0.13	0.00	0.00	0.42	0.42	-0.24	-0.21	-0.16	A-	A-	A-	A-	0.06	0.05	0.88	1.02	2.35	1.09
379	683553	21	4826	0.51	0.16	0.14	0.18	0.51	0.00	0.00	0.50	-0.26	-0.31	-0.12	0.50	A-	A-	B-	A+	0.57	0.05	-5.35	0.90	-4.84	0.87
380	702145	21	4826	0.31	0.31	0.21	0.09	0.38	0.00	0.00	0.29	0.29	-0.30	-0.11	0.04	A+	A+	A+	A-	1.63	0.05	0.54	1.01	5.42	1.23
381	678991	21	4826	0.42	0.34	0.42	0.14	0.10	0.00	0.00	0.26	0.05	0.26	-0.14	-0.32	B-	A-	A-	A+	1.06	0.05	9.43	1.21	8.80	1.29
382	702159	21	4826	0.57	0.09	0.14	0.57	0.20	0.00	0.00	0.44	-0.19	-0.39	0.44	-0.08	A-	A-	A-	A-	0.27	0.05	-0.61	0.99	-1.31	0.96
383	705228	21	4826	0.67	0.67	0.14	0.12	0.08	0.00	0.00	0.53	0.53	-0.31	-0.25	-0.23	A+	A+	A+	A-	-0.25	0.05	-5.86	0.87	-5.46	0.79

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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	MS	z	MS
384	679964	21	4826	0.57	0.12	0.57	0.21	0.10	0.00	0.00	0.42	-0.25	0.42	-0.14	-0.22	A-	A-	A-	A+	0.29	0.05	0.12	1.00	-0.66	0.98
385	703660	22	4776	0.58	0.17	0.11	0.14	0.58	0.00	0.00	0.46	-0.21	-0.18	-0.26	0.46	A+	B-	A-	A-	0.28	0.05	-0.56	0.99	-0.38	0.99
386	703658	22	4776	0.39	0.39	0.20	0.34	0.07	0.00	0.00	0.29	0.29	-0.18	-0.06	-0.16	A+	A-	A+	A-	1.24	0.05	5.59	1.12	5.86	1.20
387	673871	22	4776	0.58	0.14	0.11	0.16	0.58	0.00	0.00	0.38	-0.20	-0.21	-0.14	0.38	A+	A-	A+	A+	0.24	0.05	2.85	1.06	0.69	1.02
388	702207	22	4776	0.52	0.52	0.17	0.26	0.06	0.00	0.00	0.40	0.40	-0.28	-0.13	-0.16	A-	A-	A+	A-	0.58	0.05	2.00	1.04	1.10	1.03
389	703871	22	4776	0.77	0.05	0.77	0.11	0.07	0.00	0.00	0.52	-0.27	0.52	-0.31	-0.24	A-	C-	C-	A+	-0.81	0.06	-5.38	0.85	-4.57	0.75
390	673849	22	4776	0.56	0.13	0.56	0.10	0.21	0.00	0.00	0.29	-0.30	0.29	-0.27	0.09	A+	A+	A+	A+	0.36	0.05	7.91	1.17	7.87	1.26
391	702077	22	4776	0.68	0.68	0.07	0.13	0.12	0.00	0.00	0.48	0.48	-0.26	-0.25	-0.23	A+	A-	A+	A+	-0.30	0.05	-4.32	0.90	-3.65	0.85
392	678864	22	4776	0.35	0.10	0.21	0.34	0.35	0.00	0.00	0.38	-0.23	-0.25	-0.01	0.38	B-	A-	A+	A-	1.45	0.05	-0.23	0.99	3.49	1.13
393	702141	22	4776	0.68	0.11	0.14	0.68	0.06	0.00	0.00	0.35	-0.21	-0.16	0.35	-0.17	A+	A-	A-	A+	-0.30	0.05	1.75	1.04	1.62	1.07
394	702138	22	4776	0.73	0.09	0.10	0.08	0.73	0.00	0.00	0.52	-0.25	-0.30	-0.25	0.52	A-	A-	A-	A+	-0.59	0.06	-5.06	0.87	-4.84	0.77
395	702146	22	4776	0.29	0.36	0.29	0.13	0.22	0.00	0.00	0.19	0.14	0.19	-0.25	-0.17	A-	A-	A-	A-	1.77	0.05	6.30	1.17	8.73	1.43
396	702147	22	4776	0.77	0.02	0.17	0.04	0.77	0.00	0.00	0.34	-0.21	-0.16	-0.27	0.34	A+	A-	B-	A-	-0.84	0.06	1.51	1.05	3.19	1.20
397	679965	22	4776	0.52	0.14	0.11	0.23	0.52	0.00	0.00	0.52	-0.26	-0.30	-0.17	0.52	A+	A-	A-	A+	0.58	0.05	-5.55	0.89	-4.05	0.89
398	705227	22	4776	0.46	0.18	0.23	0.46	0.13	0.00	0.00	0.36	-0.20	-0.13	0.36	-0.13	A+	A-	A+	A-	0.87	0.05	1.13	1.02	1.60	1.05
399	702160	22	4776	0.53	0.11	0.15	0.20	0.53	0.00	0.00	0.41	-0.22	-0.16	-0.18	0.41	A+	A-	A-	A+	0.52	0.05	1.35	1.03	1.76	1.05
400	683845	22	4776	0.79	0.05	0.79	0.12	0.05	0.00	0.00	0.40	-0.20	0.40	-0.23	-0.22	A+	A-	A-	A+	-0.96	0.06	-1.88	0.94	0.85	1.05
401	703873	23	4809	0.50	0.50	0.14	0.18	0.17	0.00	0.00	0.31	0.31	-0.14	-0.17	-0.11	A-	A-	A-	A+	0.65	0.05	7.39	1.15	7.24	1.21
402	702216	23	4809	0.36	0.06	0.40	0.36	0.18	0.00	0.00	0.30	-0.12	-0.10	0.30	-0.17	A+	A+	A+	A-	1.38	0.05	3.06	1.07	5.13	1.18
403	702219	23	4809	0.64	0.17	0.08	0.64	0.11	0.00	0.00	0.26	-0.17	-0.18	0.26	-0.04	A-	A-	A-	A-	-0.08	0.05	6.19	1.14	6.43	1.26
404	674118	23	4809	0.52	0.22	0.12	0.13	0.52	0.00	0.00	0.35	-0.16	-0.22	-0.10	0.35	A+	A-	A-	A-	0.55	0.05	3.96	1.08	3.92	1.11
405	703874	23	4809	0.87	0.04	0.04	0.87	0.05	0.00	0.00	0.47	-0.28	-0.26	0.47	-0.24	A-	B-	A-	A+	-1.60	0.07	-3.11	0.87	-5.57	0.58
406	703242	23	4809	0.36	0.47	0.12	0.36	0.04	0.00	0.00	0.36	-0.05	-0.33	0.36	-0.21	A-	A-	A-	A+	1.38	0.05	1.20	1.03	5.13	1.18
407	702208	23	4809	0.67	0.67	0.15	0.15	0.02	0.00	0.00	0.30	0.30	-0.24	-0.08	-0.20	A-	A-	A+	A-	-0.23	0.05	4.94	1.12	3.69	1.16
408	703259	23	4809	0.50	0.04	0.50	0.18	0.28	0.00	0.00	0.46	-0.15	0.46	-0.33	-0.16	A-	A+	A-	A-	0.65	0.05	-3.40	0.93	-3.16	0.91
409	702142	23	4809	0.71	0.71	0.12	0.06	0.11	0.00	0.00	0.49	0.49	-0.24	-0.27	-0.25	A+	B-	A-	A+	-0.43	0.05	-4.45	0.89	-4.54	0.81
410	702143	23	4809	0.74	0.74	0.13	0.09	0.04	0.00	0.00	0.48	0.48	-0.25	-0.28	-0.24	A-	A-	A-	B-	-0.60	0.06	-2.08	0.94	-2.89	0.86
411	674169	23	4809	0.49	0.49	0.12	0.30	0.09	0.00	0.00	0.41	0.41	-0.21	-0.20	-0.16	A-	A-	A-	A+	0.74	0.05	1.02	1.02	1.27	1.04
412	702151	23	4809	0.41	0.23	0.22	0.13	0.41	0.00	0.00	0.29	-0.11	-0.13	-0.12	0.29	A+	A-	A-	A-	1.14	0.05	7.40	1.16	7.79	1.25
413	704788	23	4809	0.48	0.23	0.21	0.48	0.08	0.00	0.00	0.26	0.01	-0.18	0.26	-0.21	A-	A+	A+	A+	0.78	0.05	9.90	1.23	9.90	1.30
414	679683	23	4809	0.58	0.58	0.10	0.13	0.19	0.00	0.00	0.44	0.44	-0.19	-0.26	-0.19	A-	A-	A-	A+	0.26	0.05	-1.17	0.98	-0.66	0.98
415	682253	23	4809	0.67	0.07	0.23	0.67	0.04	0.00	0.00	0.47	-0.23	-0.28	0.47	-0.23	A-	A-	A+	A-	-0.19	0.05	-3.01	0.93	-3.93	0.85

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
416	674092	23	4809	0.73	0.07	0.12	0.73	0.08	0.00	0.00	0.53	-0.22	-0.30	0.53	-0.30	A-	B-	C-	A+	-0.55	0.06	-6.94	0.83	-6.42	0.72
417	702632	24	4759	0.38	0.27	0.21	0.38	0.14	0.00	0.00	0.33	-0.02	-0.23	0.33	-0.15	A-	A+	A-	A+	1.26	0.05	2.67	1.06	5.31	1.18
418	705267	24	4759	0.69	0.69	0.09	0.15	0.07	0.00	0.00	0.39	0.39	-0.17	-0.26	-0.16	A-	B-	A-	A+	-0.36	0.05	-0.17	1.00	-1.18	0.95
419	700890	24	4759	0.36	0.36	0.23	0.18	0.24	0.00	0.00	0.45	0.45	-0.19	-0.10	-0.23	A-	A+	B-	A+	1.37	0.05	-3.89	0.92	-0.99	0.97
420	702209	24	4759	0.31	0.16	0.20	0.31	0.32	0.00	0.00	0.23	-0.15	-0.13	0.23	0.00	A-	A-	A-	A-	1.62	0.05	4.00	1.10	6.98	1.30
421	703875	24	4759	0.59	0.21	0.59	0.10	0.10	0.00	0.00	0.41	-0.23	0.41	-0.25	-0.10	A-	A-	A-	A-	0.20	0.05	-0.30	0.99	-1.64	0.95
422	702627	24	4759	0.46	0.18	0.22	0.46	0.13	0.00	0.00	0.35	-0.14	-0.11	0.35	-0.22	A-	A-	A-	A-	0.81	0.05	2.21	1.04	2.54	1.07
423	703243	24	4759	0.50	0.20	0.50	0.12	0.19	0.00	0.00	0.49	-0.22	0.49	-0.24	-0.20	A-	A-	A-	A+	0.64	0.05	-4.70	0.91	-4.12	0.89
424	703260	24	4759	0.47	0.16	0.23	0.13	0.47	0.00	0.00	0.43	-0.21	-0.23	-0.12	0.43	A-	A+	A-	A-	0.77	0.05	-0.14	1.00	0.48	1.01
425	702139	24	4759	0.29	0.15	0.43	0.13	0.29	0.00	0.00	0.24	-0.15	0.01	-0.17	0.24	A-	A-	A-	A-	1.74	0.05	4.36	1.12	6.48	1.30
426	702140	24	4759	0.63	0.21	0.08	0.08	0.63	0.00	0.00	0.49	-0.24	-0.19	-0.32	0.49	A-	A-	A-	A-	-0.01	0.05	-2.73	0.94	-1.48	0.95
427	679206	24	4759	0.64	0.06	0.64	0.09	0.21	0.00	0.00	0.48	-0.25	0.48	-0.25	-0.25	A+	A-	A-	A+	-0.09	0.05	-4.04	0.91	-4.29	0.85
428	705229	24	4759	0.40	0.31	0.40	0.17	0.12	0.00	0.00	0.26	0.00	0.26	-0.24	-0.10	A-	A+	A+	A+	1.14	0.05	8.02	1.18	8.87	1.29
429	704781	24	4759	0.39	0.18	0.14	0.28	0.39	0.00	0.00	0.23	-0.14	-0.13	-0.02	0.23	A+	A-	A+	A+	1.17	0.05	8.77	1.19	9.53	1.32
430	642905	24	4759	0.64	0.14	0.09	0.12	0.64	0.00	0.00	0.41	-0.17	-0.24	-0.22	0.41	A-	B-	B-	A-	-0.10	0.05	-0.84	0.98	-0.43	0.98
431	674165	24	4759	0.68	0.07	0.68	0.04	0.20	0.00	0.00	0.36	-0.28	0.36	-0.27	-0.10	A+	A-	A-	A+	-0.31	0.05	0.87	1.02	1.05	1.04
432	680556	24	4759	0.48	0.48	0.17	0.06	0.28	0.00	0.00	0.24	0.24	-0.05	-0.25	-0.08	A+	A+	A+	A+	0.73	0.05	9.59	1.20	8.75	1.26

Table J–7. Literature Multiple-Choice Item Statistics: Spring

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	610362	0	140252	0.76	0.76	0.07	0.05	0.12	0.00	0.00	0.40	0.40	-0.25	-0.29	-0.13					-0.23	0.01	-9.90	0.91	-9.90	0.89
2	610359	0	140252	0.74	0.12	0.08	0.74	0.06	0.00	0.00	0.36	-0.11	-0.25	0.36	-0.24					-0.34	0.01	4.62	1.03	9.90	1.12
3	610357	0	140252	0.69	0.08	0.69	0.10	0.12	0.00	0.00	0.46	-0.20	0.46	-0.25	-0.25					0.20	0.01	-9.90	0.85	-9.90	0.78
4	610364	0	140252	0.64	0.04	0.10	0.21	0.64	0.00	0.00	0.22	-0.25	-0.23	0.03	0.22					0.24	0.01	9.90	1.25	9.90	1.45
5	610366	0	140252	0.62	0.62	0.13	0.06	0.20	0.00	0.00	0.32	0.32	-0.25	-0.30	0.00					0.60	0.01	9.90	1.06	9.90	1.09
6	610363	0	140252	0.67	0.19	0.07	0.67	0.06	0.00	0.00	0.40	-0.17	-0.26	0.40	-0.21					0.58	0.01	-9.90	0.94	-9.90	0.91
7	610361	0	140252	0.67	0.14	0.12	0.67	0.07	0.00	0.00	0.28	-0.17	-0.07	0.28	-0.18					-0.11	0.01	9.90	1.25	9.90	1.44
8	610365	0	140252	0.71	0.14	0.71	0.05	0.10	0.00	0.00	0.44	-0.24	0.44	-0.27	-0.19					0.22	0.01	-9.90	0.87	-9.90	0.82
9	610358	0	140252	0.49	0.49	0.18	0.11	0.22	0.00	0.00	0.21	0.21	-0.05	-0.20	-0.05					1.27	0.01	9.90	1.22	9.90	1.35
10	640841	0	140252	0.76	0.10	0.08	0.76	0.05	0.00	0.00	0.47	-0.22	-0.30	0.47	-0.21					-0.56	0.01	-9.90	0.88	-9.90	0.82
11	640840	0	140252	0.56	0.56	0.09	0.18	0.17	0.00	0.00	0.29	0.29	-0.24	-0.06	-0.12					0.33	0.01	9.90	1.24	9.90	1.35
12	640857	0	140252	0.81	0.08	0.05	0.07	0.81	0.00	0.00	0.43	-0.18	-0.24	-0.26	0.43					-0.91	0.01	-7.18	0.94	-6.37	0.91
13	640839	0	140252	0.77	0.77	0.08	0.09	0.06	0.00	0.00	0.44	0.44	-0.25	-0.21	-0.23					-1.03	0.01	9.90	1.17	8.03	1.13
14	640858	0	140252	0.62	0.20	0.62	0.09	0.09	0.00	0.00	0.40	-0.21	0.40	-0.20	-0.16					0.39	0.01	3.55	1.02	4.13	1.03
15	640842	0	140252	0.65	0.09	0.14	0.12	0.65	0.00	0.00	0.50	-0.25	-0.23	-0.26	0.50					-0.19	0.01	2.66	1.02	-5.90	0.94
16	640859	0	140252	0.67	0.09	0.15	0.67	0.09	0.00	0.00	0.45	-0.19	-0.27	0.45	-0.20					-0.09	0.01	-2.16	0.99	-7.86	0.93
17	640836	0	140252	0.65	0.10	0.06	0.65	0.18	0.00	0.00	0.35	-0.22	-0.27	0.35	-0.08					0.38	0.01	5.96	1.03	9.90	1.08
18	683335	0	140252	0.83	0.83	0.09	0.05	0.03	0.00	0.00	0.52	0.52	-0.34	-0.26	-0.25					-1.07	0.01	-9.90	0.74	-9.90	0.53
19	683343	0	140252	0.53	0.53	0.12	0.24	0.11	0.00	0.00	0.28	0.28	-0.11	-0.15	-0.13					0.72	0.01	9.90	1.13	9.90	1.17
20	683461	0	140252	0.61	0.61	0.10	0.24	0.05	0.00	0.00	0.34	0.34	-0.23	-0.17	-0.12					0.35	0.01	9.90	1.07	9.90	1.08
21	683338	0	140252	0.63	0.20	0.05	0.12	0.63	0.00	0.00	0.40	-0.16	-0.26	-0.21	0.40					0.25	0.01	4.75	1.02	1.04	1.01
22	683342	0	140252	0.70	0.10	0.70	0.11	0.08	0.00	0.00	0.49	-0.25	0.49	-0.28	-0.22					-0.22	0.01	-9.90	0.91	-9.90	0.80
23	683336	0	140252	0.81	0.03	0.81	0.12	0.04	0.00	0.00	0.47	-0.24	0.47	-0.32	-0.20					-0.82	0.01	-9.90	0.80	-9.90	0.70
24	683340	0	140252	0.56	0.20	0.56	0.17	0.07	0.00	0.00	0.29	-0.06	0.29	-0.13	-0.28					1.07	0.01	9.90	1.10	9.90	1.16
25	683460	0	140252	0.44	0.44	0.21	0.07	0.27	0.00	0.00	0.28	0.28	-0.13	-0.25	-0.05					1.34	0.01	9.90	1.07	9.90	1.15
26	614543	0	140252	0.74	0.08	0.07	0.11	0.74	0.00	0.00	0.27	-0.16	-0.27	-0.01	0.27					-0.47	0.01	9.90	1.20	9.90	1.61
27	614547	0	140252	0.73	0.73	0.05	0.14	0.07	0.00	0.00	0.36	0.36	-0.24	-0.15	-0.17					-1.02	0.01	9.90	1.57	9.90	1.90
28	614541	0	140252	0.68	0.06	0.68	0.04	0.23	0.00	0.00	0.40	-0.31	0.40	-0.26	-0.15					-0.18	0.01	9.03	1.06	8.71	1.09
29	614542	0	140252	0.51	0.09	0.35	0.51	0.05	0.00	0.00	0.46	-0.28	-0.19	0.46	-0.24					0.61	0.01	-9.90	0.95	-9.90	0.92
30	614544	0	140252	0.79	0.79	0.07	0.09	0.04	0.00	0.00	0.49	0.49	-0.33	-0.21	-0.24					-0.53	0.01	-9.90	0.79	-9.90	0.69
31	614545	0	140252	0.22	0.22	0.15	0.46	0.16	0.00	0.00	0.21	0.21	-0.11	0.01	-0.13					2.14	0.01	-9.90	0.93	9.90	1.11

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	614550	0	140252	0.69	0.11	0.69	0.15	0.05	0.00	0.00	0.45	-0.19	0.45	-0.24	-0.26					0.14	0.01	-9.90	0.90	-9.90	0.87
33	614546	0	140252	0.62	0.11	0.62	0.15	0.12	0.00	0.00	0.37	-0.15	0.37	-0.20	-0.16					0.40	0.01	3.22	1.02	1.36	1.01
34	614539	0	140252	0.61	0.30	0.04	0.61	0.03	0.00	0.00	0.34	-0.11	-0.29	0.34	-0.27					0.33	0.01	9.90	1.10	9.90	1.14
35	701988	1	4772	0.95	0.02	0.03	0.95	0.01	0.00	0.00	0.38	-0.23	-0.23	0.38	-0.17	A+	A-	A-	A+	-2.25	0.11	-1.42	0.88	-3.51	0.55
36	701992	1	4772	0.80	0.04	0.07	0.09	0.80	0.00	0.00	0.47	-0.20	-0.27	-0.27	0.47	B+	A-	A-	A+	-0.57	0.06	-3.48	0.88	-3.81	0.78
37	701989	1	4772	0.91	0.04	0.04	0.01	0.91	0.00	0.00	0.44	-0.29	-0.22	-0.22	0.44	A+	A-	A-	A-	-1.62	0.08	-1.79	0.89	-4.38	0.58
38	701984	1	4772	0.74	0.14	0.06	0.06	0.74	0.00	0.00	0.43	-0.19	-0.25	-0.25	0.43	A+	A+	A+	A+	-0.13	0.06	-1.89	0.95	-1.92	0.91
39	701990	1	4772	0.72	0.06	0.72	0.18	0.04	0.00	0.00	0.53	-0.24	0.53	-0.35	-0.22	A-	B-	A-	A-	-0.02	0.06	-4.90	0.87	-5.79	0.76
40	701985	1	4772	0.71	0.71	0.10	0.14	0.05	0.00	0.00	0.39	0.39	-0.27	-0.15	-0.20	A-	A-	A-	A-	0.03	0.06	0.33	1.01	-0.39	0.98
41	704152	1	4772	0.29	0.34	0.29	0.23	0.14	0.00	0.00	0.02	0.00	0.02	0.02	-0.04	A+	A-	A+	A-	2.20	0.05	9.79	1.25	9.90	1.80
42	704150	1	4772	0.66	0.09	0.18	0.07	0.66	0.00	0.00	0.43	-0.23	-0.16	-0.29	0.43	A+	A-	A-	A-	0.34	0.05	-0.58	0.99	-0.55	0.98
43	704157	1	4772	0.85	0.05	0.04	0.06	0.85	0.00	0.00	0.49	-0.28	-0.28	-0.24	0.49	A+	A-	B-	A+	-0.96	0.07	-4.21	0.83	-5.68	0.62
44	704158	1	4772	0.55	0.55	0.08	0.22	0.15	0.00	0.00	0.39	0.39	-0.31	-0.17	-0.11	A-	A-	A-	A+	0.91	0.05	1.93	1.04	2.00	1.06
45	704149	1	4772	0.80	0.80	0.07	0.05	0.07	0.00	0.00	0.48	0.48	-0.32	-0.24	-0.21	A-	A-	A-	A+	-0.60	0.06	-3.03	0.89	-3.37	0.80
46	704151	1	4772	0.31	0.17	0.31	0.25	0.26	0.00	0.00	0.07	-0.03	0.07	0.00	-0.04	A+	A+	A+	A+	2.11	0.05	9.90	1.31	9.90	1.77
47	701981	2	4781	0.44	0.44	0.04	0.28	0.24	0.00	0.00	0.29	0.29	-0.22	0.11	-0.35	A-	A-	A-	A+	1.47	0.05	3.19	1.06	5.15	1.16
48	701982	2	4781	0.96	0.96	0.01	0.02	0.01	0.00	0.00	0.38	0.38	-0.16	-0.27	-0.20	A+	B-	B-	A+	-2.40	0.11	-1.63	0.86	-3.08	0.57
49	701991	2	4781	0.51	0.17	0.12	0.51	0.19	0.00	0.00	0.31	-0.20	-0.09	0.31	-0.13	A+	A-	A-	A+	1.10	0.05	3.47	1.06	3.69	1.10
50	701983	2	4781	0.67	0.03	0.03	0.67	0.26	0.00	0.00	0.27	-0.25	-0.33	0.27	-0.04	A-	A+	A-	A-	0.31	0.05	5.11	1.13	6.04	1.24
51	701986	2	4781	0.80	0.01	0.80	0.02	0.17	0.00	0.00	0.37	-0.19	0.37	-0.17	-0.27	A-	A-	A-	A+	-0.54	0.06	-0.63	0.98	-0.73	0.95
52	701987	2	4781	0.75	0.07	0.75	0.16	0.02	0.00	0.00	0.39	-0.28	0.39	-0.20	-0.17	A+	A-	A+	A-	-0.18	0.06	0.39	1.01	0.45	1.02
53	704153	2	4781	0.64	0.21	0.03	0.11	0.64	0.00	0.00	0.40	-0.20	-0.25	-0.18	0.40	A-	A-	A-	A-	0.49	0.05	1.05	1.02	0.79	1.03
54	704155	2	4781	0.72	0.04	0.72	0.06	0.18	0.00	0.00	0.39	-0.27	0.39	-0.31	-0.12	A-	A-	A-	A+	0.04	0.05	-1.73	0.95	-1.95	0.92
55	704156	2	4781	0.63	0.21	0.10	0.63	0.05	0.00	0.00	0.35	-0.16	-0.21	0.35	-0.16	A-	A-	A-	A+	0.50	0.05	-0.55	0.99	-1.65	0.95
56	704147	2	4781	0.64	0.02	0.05	0.64	0.28	0.00	0.00	0.44	-0.21	-0.25	0.44	-0.27	A+	A-	A-	A+	0.44	0.05	-1.93	0.96	-1.83	0.94
57	704154	2	4781	0.53	0.53	0.33	0.09	0.05	0.00	0.00	0.34	0.34	-0.11	-0.25	-0.23	A-	A-	A-	A+	1.01	0.05	1.91	1.04	2.05	1.06
58	704159	2	4781	0.63	0.03	0.63	0.15	0.18	0.00	0.00	0.39	-0.21	0.39	-0.20	-0.20	B-	B-	A-	A+	0.50	0.05	-0.12	1.00	-0.77	0.97
59	702136	3	4801	0.66	0.11	0.17	0.06	0.66	0.00	0.00	0.46	-0.20	-0.25	-0.25	0.46	A-	B-	B-	A+	0.33	0.05	-2.51	0.94	-2.32	0.92
60	702135	3	4801	0.71	0.07	0.15	0.71	0.06	0.00	0.00	0.40	-0.26	-0.18	0.40	-0.19	A-	A-	B-	A-	0.04	0.05	0.05	1.00	-0.83	0.96
61	702127	3	4801	0.79	0.05	0.79	0.12	0.05	0.00	0.00	0.45	-0.24	0.45	-0.27	-0.22	A+	A-	A-	A+	-0.43	0.06	-3.58	0.88	-4.02	0.79
62	702133	3	4801	0.68	0.07	0.68	0.23	0.03	0.00	0.00	0.20	-0.13	0.20	-0.06	-0.21	A+	A+	A+	A+	0.24	0.05	5.98	1.15	4.79	1.19
63	702132	3	4801	0.83	0.06	0.83	0.05	0.06	0.00	0.00	0.51	-0.31	0.51	-0.28	-0.24	A+	C-	C-	A+	-0.78	0.07	-5.12	0.81	-5.21	0.68

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
64	702130	3	4801	0.70	0.70	0.19	0.07	0.05	0.00	0.00	0.35	0.35	-0.14	-0.24	-0.20	A+	A-	A-	A-	0.15	0.05	2.53	1.07	1.83	1.07
65	703570	3	4801	0.71	0.05	0.09	0.71	0.16	0.00	0.00	0.33	-0.26	-0.15	0.33	-0.14	A+	A-	A-	A+	0.08	0.05	2.26	1.06	3.31	1.14
66	703573	3	4801	0.79	0.79	0.05	0.09	0.06	0.00	0.00	0.49	0.49	-0.34	-0.25	-0.19	A-	B-	B-	A+	-0.45	0.06	-3.89	0.87	-4.62	0.75
67	703574	3	4801	0.77	0.08	0.77	0.07	0.08	0.00	0.00	0.50	-0.28	0.50	-0.28	-0.22	C-	B-	B-	A-	-0.34	0.06	-2.18	0.93	-4.30	0.78
68	703575	3	4801	0.71	0.04	0.21	0.71	0.05	0.00	0.00	0.52	-0.29	-0.30	0.52	-0.27	A-	A-	A-	A+	0.08	0.05	-4.34	0.89	-4.90	0.81
69	703571	3	4801	0.44	0.43	0.44	0.06	0.06	0.00	0.00	0.21	0.06	0.21	-0.27	-0.27	A-	A-	A-	A+	1.46	0.05	8.23	1.16	8.56	1.27
70	703565	3	4801	0.63	0.63	0.16	0.09	0.12	0.00	0.00	0.35	0.35	-0.14	-0.26	-0.13	A-	B-	A-	A+	0.52	0.05	1.57	1.03	0.92	1.03
71	702124	4	4772	0.64	0.27	0.05	0.04	0.64	0.00	0.00	0.21	-0.01	-0.21	-0.26	0.21	A+	A+	A-	A-	0.42	0.05	8.42	1.20	9.69	1.35
72	702131	4	4772	0.49	0.14	0.17	0.20	0.49	0.00	0.00	0.29	-0.09	-0.16	-0.13	0.29	A+	A-	A-	B+	1.19	0.05	5.43	1.10	5.77	1.17
73	702126	4	4772	0.81	0.81	0.13	0.03	0.03	0.00	0.00	0.37	0.37	-0.18	-0.26	-0.22	A+	B-	B-	A+	-0.65	0.06	-1.33	0.95	-1.65	0.90
74	702137	4	4772	0.61	0.61	0.20	0.03	0.16	0.00	0.00	0.21	0.21	-0.13	-0.17	-0.04	A-	A-	A+	A+	0.60	0.05	8.65	1.19	7.82	1.26
75	702125	4	4772	0.80	0.04	0.11	0.80	0.04	0.00	0.00	0.43	-0.23	-0.27	0.43	-0.18	A+	A-	A-	A+	-0.58	0.06	-3.02	0.90	-3.13	0.82
76	702128	4	4772	0.75	0.06	0.08	0.75	0.11	0.00	0.00	0.46	-0.22	-0.27	0.46	-0.22	A-	A-	A-	A-	-0.24	0.06	-2.20	0.93	-2.46	0.88
77	703576	4	4772	0.39	0.18	0.10	0.33	0.39	0.00	0.00	0.37	-0.18	-0.18	-0.11	0.37	A-	B-	C-	A-	1.70	0.05	-0.62	0.99	1.59	1.05
78	703572	4	4772	0.47	0.07	0.14	0.31	0.47	0.00	0.00	0.22	-0.21	-0.12	-0.02	0.22	A-	A-	A-	A+	1.26	0.05	9.34	1.18	9.50	1.29
79	703569	4	4772	0.66	0.66	0.23	0.08	0.03	0.00	0.00	0.48	0.48	-0.27	-0.26	-0.25	A-	A-	A-	A+	0.35	0.05	-4.71	0.89	-5.48	0.82
80	703567	4	4772	0.90	0.04	0.02	0.03	0.90	0.00	0.00	0.52	-0.30	-0.27	-0.30	0.52	A+	B-	B-	B+	-1.49	0.08	-5.51	0.73	-7.18	0.44
81	703566	4	4772	0.62	0.07	0.07	0.62	0.23	0.00	0.00	0.11	-0.22	-0.16	0.11	0.11	A+	A-	A-	A-	0.53	0.05	9.90	1.33	9.90	1.52
82	703577	4	4772	0.85	0.06	0.85	0.06	0.02	0.00	0.00	0.48	-0.32	0.48	-0.23	-0.26	A+	B-	A-	A+	-0.99	0.07	-4.26	0.83	-4.15	0.71
83	702122	5	4755	0.73	0.07	0.73	0.10	0.09	0.00	0.00	0.46	-0.26	0.46	-0.28	-0.18	A-	A+	A-	A+	-0.10	0.06	-3.28	0.91	-3.85	0.84
84	702119	5	4755	0.73	0.73	0.03	0.04	0.20	0.00	0.00	0.37	0.37	-0.27	-0.26	-0.16	A+	A-	A-	A-	-0.05	0.06	-0.50	0.99	-0.25	0.99
85	702113	5	4755	0.68	0.11	0.12	0.09	0.68	0.00	0.00	0.45	-0.19	-0.24	-0.25	0.45	A+	A-	A-	A+	0.20	0.05	-2.40	0.94	-2.05	0.92
86	702123	5	4755	0.61	0.05	0.08	0.61	0.26	0.00	0.00	0.19	-0.16	-0.15	0.19	-0.03	A-	A-	A-	A+	0.61	0.05	9.90	1.25	9.90	1.36
87	702110	5	4755	0.76	0.76	0.07	0.06	0.10	0.00	0.00	0.44	0.44	-0.25	-0.27	-0.17	A+	A-	A-	A+	-0.30	0.06	-2.98	0.91	-3.86	0.81
88	702115	5	4755	0.62	0.06	0.24	0.07	0.62	0.00	0.00	0.43	-0.27	-0.19	-0.22	0.43	A+	A-	A-	A-	0.52	0.05	-2.58	0.94	-2.28	0.93
89	703653	5	4755	0.59	0.59	0.14	0.22	0.05	0.00	0.00	0.25	0.25	-0.22	0.02	-0.24	A-	B-	A-	A-	0.71	0.05	7.17	1.15	7.37	1.23
90	703656	5	4755	0.70	0.70	0.17	0.09	0.04	0.00	0.00	0.43	0.43	-0.18	-0.30	-0.22	A+	A-	A-	A+	0.12	0.05	-2.63	0.93	-1.84	0.93
91	703655	5	4755	0.53	0.12	0.20	0.15	0.53	0.00	0.00	0.52	-0.14	-0.26	-0.31	0.52	A+	A-	A-	A-	1.00	0.05	-9.23	0.84	-6.96	0.83
92	703651	5	4755	0.71	0.07	0.15	0.07	0.71	0.00	0.00	0.38	-0.29	-0.13	-0.18	0.38	A+	A-	A-	A-	0.06	0.05	1.78	1.05	1.00	1.04
93	703650	5	4755	0.72	0.72	0.11	0.10	0.07	0.00	0.00	0.43	0.43	-0.15	-0.34	-0.16	A+	A-	A-	A+	-0.05	0.06	-1.24	0.97	0.25	1.01
94	703647	5	4755	0.38	0.15	0.38	0.12	0.35	0.00	0.00	0.18	-0.05	0.18	-0.17	-0.02	A+	A-	A-	A-	1.72	0.05	8.41	1.17	9.90	1.43
95	702120	6	4748	0.58	0.58	0.06	0.08	0.28	0.00	0.00	0.22	0.22	-0.21	-0.29	0.06	A-	A-	A-	A+	0.73	0.05	8.49	1.18	7.40	1.22

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
96	702112	6	4748	0.74	0.12	0.06	0.74	0.08	0.00	0.00	0.34	-0.15	-0.20	0.34	-0.19	A+	A-	A-	A-	-0.16	0.06	1.69	1.05	1.36	1.06
97	702111	6	4748	0.54	0.29	0.54	0.13	0.05	0.00	0.00	0.20	-0.07	0.20	-0.07	-0.21	A+	A-	A-	A+	0.94	0.05	9.90	1.20	9.90	1.29
98	702117	6	4748	0.43	0.25	0.43	0.11	0.21	0.00	0.00	0.23	-0.12	0.23	-0.12	-0.05	A-	A-	A-	A-	1.46	0.05	6.60	1.12	7.84	1.24
99	702118	6	4748	0.26	0.14	0.41	0.19	0.26	0.00	0.00	0.15	-0.20	0.08	-0.08	0.15	A+	A+	A-	A-	2.38	0.05	2.00	1.05	5.13	1.26
100	702114	6	4748	0.27	0.52	0.09	0.27	0.11	0.00	0.00	0.21	0.06	-0.23	0.21	-0.18	A-	A+	A-	A+	2.31	0.05	1.84	1.05	6.97	1.35
101	703649	6	4748	0.67	0.08	0.16	0.67	0.09	0.00	0.00	0.40	-0.24	-0.16	0.40	-0.22	A-	A-	A-	A+	0.29	0.05	-1.67	0.96	-1.61	0.94
102	703648	6	4748	0.78	0.06	0.78	0.10	0.06	0.00	0.00	0.48	-0.25	0.48	-0.29	-0.22	A+	B-	A-	A+	-0.40	0.06	-3.52	0.89	-4.11	0.80
103	703654	6	4748	0.56	0.13	0.15	0.56	0.16	0.00	0.00	0.36	-0.19	-0.18	0.36	-0.13	A-	B-	A-	A+	0.82	0.05	0.83	1.02	0.62	1.02
104	703646	6	4748	0.46	0.24	0.20	0.10	0.46	0.00	0.00	0.42	-0.07	-0.22	-0.30	0.42	A+	A+	A-	A-	1.33	0.05	-4.22	0.93	-1.99	0.95
105	703652	6	4748	0.47	0.09	0.47	0.10	0.34	0.00	0.00	0.32	-0.16	0.32	-0.18	-0.12	A-	B-	A-	A+	1.26	0.05	0.98	1.02	2.52	1.07
106	703645	6	4748	0.63	0.16	0.17	0.63	0.04	0.00	0.00	0.41	-0.10	-0.28	0.41	-0.26	A+	A-	A-	A+	0.48	0.05	-0.88	0.98	-0.82	0.97
107	701999	7	4799	0.73	0.13	0.04	0.73	0.09	0.00	0.00	0.41	-0.27	-0.20	0.41	-0.16	A+	A+	A-	A-	-0.07	0.06	-0.57	0.98	-0.04	1.00
108	702005	7	4799	0.86	0.01	0.05	0.86	0.08	0.00	0.00	0.31	-0.19	-0.17	0.31	-0.17	A-	A+	B-	A+	-1.04	0.07	-0.33	0.98	0.19	1.01
109	701998	7	4799	0.58	0.13	0.18	0.58	0.10	0.00	0.00	0.37	-0.21	-0.22	0.37	-0.08	A+	A+	A-	A+	0.74	0.05	1.77	1.04	1.45	1.04
110	702004	7	4799	0.84	0.02	0.84	0.06	0.08	0.00	0.00	0.38	-0.20	0.38	-0.20	-0.23	A-	A-	A-	A-	-0.86	0.07	-0.89	0.96	-1.34	0.90
111	701995	7	4799	0.84	0.84	0.02	0.06	0.07	0.00	0.00	0.37	0.37	-0.20	-0.23	-0.17	A+	A-	A-	A-	-0.88	0.07	-1.33	0.95	-0.75	0.94
112	701993	7	4799	0.47	0.12	0.47	0.21	0.20	0.00	0.00	0.32	-0.07	0.32	-0.22	-0.11	A-	A-	A+	A-	1.30	0.05	3.74	1.07	4.83	1.14
113	704228	7	4799	0.51	0.51	0.18	0.26	0.06	0.00	0.00	0.37	0.37	-0.13	-0.18	-0.25	A-	B-	A-	A+	1.11	0.05	0.30	1.01	2.72	1.08
114	704223	7	4799	0.52	0.19	0.13	0.15	0.52	0.00	0.00	0.26	-0.06	-0.11	-0.18	0.26	A-	A-	A-	A+	1.04	0.05	7.47	1.15	6.49	1.19
115	704230	7	4799	0.62	0.04	0.62	0.14	0.20	0.00	0.00	0.36	-0.30	0.36	-0.23	-0.09	A-	A-	A+	A+	0.55	0.05	1.11	1.02	0.68	1.02
116	704232	7	4799	0.77	0.04	0.02	0.16	0.77	0.00	0.00	0.39	-0.22	-0.26	-0.22	0.39	A-	A-	A-	A+	-0.34	0.06	-0.62	0.98	-0.56	0.97
117	704234	7	4799	0.86	0.07	0.05	0.86	0.02	0.00	0.00	0.48	-0.30	-0.29	0.48	-0.19	A+	A-	A-	B+	-1.06	0.07	-5.83	0.76	-6.56	0.55
118	704225	7	4799	0.77	0.10	0.77	0.08	0.06	0.00	0.00	0.48	-0.19	0.48	-0.27	-0.31	A+	A+	A-	A-	-0.29	0.06	-4.03	0.88	-3.82	0.81
119	701994	8	4779	0.79	0.06	0.11	0.04	0.79	0.00	0.00	0.49	-0.26	-0.29	-0.24	0.49	A+	A-	B-	A-	-0.46	0.06	-3.29	0.89	-3.60	0.80
120	702003	8	4779	0.50	0.50	0.06	0.24	0.20	0.00	0.00	0.33	0.33	-0.17	-0.21	-0.09	C-	A-	A-	B-	1.18	0.05	3.49	1.07	3.23	1.10
121	702002	8	4779	0.70	0.02	0.70	0.12	0.15	0.00	0.00	0.37	-0.25	0.37	-0.31	-0.08	A+	A-	A-	A+	0.13	0.05	1.41	1.04	2.17	1.09
122	702000	8	4779	0.69	0.18	0.07	0.06	0.69	0.00	0.00	0.37	-0.28	-0.14	-0.12	0.37	A+	C-	B-	A-	0.18	0.05	0.45	1.01	0.89	1.04
123	702001	8	4779	0.62	0.62	0.28	0.08	0.03	0.00	0.00	0.25	0.25	-0.10	-0.15	-0.20	A-	A+	A+	A+	0.59	0.05	7.98	1.19	6.63	1.24
124	701996	8	4779	0.78	0.15	0.02	0.04	0.78	0.00	0.00	0.38	-0.26	-0.21	-0.16	0.38	A+	A-	B-	A-	-0.42	0.06	-0.76	0.97	-0.34	0.98
125	704231	8	4779	0.71	0.14	0.08	0.71	0.07	0.00	0.00	0.46	-0.23	-0.27	0.46	-0.22	A-	A-	A-	A+	0.08	0.05	-2.60	0.93	-2.50	0.89
126	704221	8	4779	0.85	0.07	0.85	0.06	0.03	0.00	0.00	0.55	-0.33	0.55	-0.32	-0.24	A-	B-	B-	A+	-0.96	0.07	-7.48	0.71	-8.23	0.47
127	704233	8	4779	0.89	0.89	0.02	0.05	0.04	0.00	0.00	0.46	0.46	-0.27	-0.29	-0.22	A+	A-	B-	A+	-1.34	0.08	-2.79	0.86	-4.36	0.63

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
128	704227	8	4779	0.77	0.03	0.05	0.15	0.77	0.00	0.00	0.47	-0.24	-0.31	-0.25	0.47	A+	A-	B-	A-	-0.36	0.06	-1.71	0.95	-1.91	0.89
129	704222	8	4779	0.84	0.05	0.06	0.84	0.05	0.00	0.00	0.50	-0.24	-0.29	0.50	-0.28	A+	A-	A-	A+	-0.88	0.07	-3.67	0.86	-4.95	0.66
130	704224	8	4779	0.82	0.82	0.09	0.03	0.05	0.00	0.00	0.46	0.46	-0.28	-0.28	-0.21	B+	A-	B-	A+	-0.73	0.06	-3.14	0.88	-3.66	0.76
131	700959	9	4754	0.51	0.18	0.51	0.09	0.22	0.00	0.00	0.22	-0.09	0.22	-0.17	-0.05	A-	A+	A+	A+	1.11	0.05	8.10	1.15	8.87	1.24
132	700967	9	4754	0.83	0.07	0.06	0.03	0.83	0.00	0.00	0.49	-0.27	-0.29	-0.24	0.49	B-	C-	C-	A+	-0.78	0.06	-5.83	0.79	-6.62	0.62
133	700970	9	4754	0.68	0.18	0.07	0.68	0.07	0.00	0.00	0.37	-0.21	-0.23	0.37	-0.12	A-	A-	A-	A+	0.25	0.05	-1.42	0.97	-2.07	0.93
134	700962	9	4754	0.39	0.39	0.14	0.05	0.42	0.00	0.00	0.18	0.18	-0.24	-0.28	0.12	A-	A-	A-	A+	1.68	0.05	9.23	1.18	9.62	1.32
135	700965	9	4754	0.77	0.06	0.08	0.77	0.10	0.00	0.00	0.40	-0.24	-0.19	0.40	-0.20	B-	A-	A-	A+	-0.28	0.06	-0.90	0.97	-1.36	0.93
136	700969	9	4754	0.35	0.22	0.35	0.20	0.23	0.00	0.00	0.19	-0.18	0.19	-0.07	0.03	A-	A+	A-	A-	1.87	0.05	5.31	1.11	8.46	1.32
137	697076	9	4754	0.56	0.56	0.17	0.19	0.08	0.00	0.00	0.18	0.18	-0.11	-0.01	-0.15	A+	A-	A+	A+	0.84	0.05	9.90	1.21	9.90	1.30
138	697078	9	4754	0.57	0.57	0.19	0.16	0.07	0.00	0.00	0.30	0.30	-0.18	-0.07	-0.19	A-	A-	A-	A-	0.80	0.05	3.08	1.06	3.19	1.09
139	697095	9	4754	0.30	0.26	0.29	0.30	0.16	0.00	0.00	0.01	0.09	0.01	0.01	-0.12	A-	A+	A+	A-	2.14	0.05	9.90	1.28	9.90	1.66
140	697081	9	4754	0.71	0.12	0.08	0.09	0.71	0.00	0.00	0.50	-0.24	-0.28	-0.25	0.50	A+	A-	A-	A+	0.05	0.05	-4.42	0.89	-4.52	0.83
141	697082	9	4754	0.56	0.19	0.17	0.56	0.07	0.00	0.00	0.29	-0.10	-0.07	0.29	-0.27	A+	A-	A-	A+	0.85	0.05	5.88	1.11	6.15	1.17
142	697092	9	4754	0.57	0.12	0.08	0.23	0.57	0.00	0.00	0.35	-0.27	-0.26	-0.03	0.35	A+	A-	A-	A-	0.82	0.05	2.07	1.04	3.09	1.08
143	700971	10	4765	0.66	0.09	0.17	0.66	0.08	0.00	0.00	0.37	-0.27	-0.13	0.37	-0.18	A-	A-	A-	A-	0.36	0.05	-0.08	1.00	-0.38	0.99
144	700963	10	4765	0.40	0.40	0.12	0.15	0.33	0.00	0.00	0.20	0.20	-0.07	-0.18	-0.02	A+	A-	A-	A-	1.61	0.05	7.40	1.14	8.72	1.27
145	700960	10	4765	0.21	0.18	0.15	0.46	0.21	0.00	0.00	-0.07	-0.08	-0.23	0.29	-0.07	A+	A+	A+	A-	2.71	0.06	9.31	1.32	9.90	2.32
146	700964	10	4765	0.38	0.41	0.08	0.14	0.38	0.00	0.00	0.35	0.00	-0.28	-0.27	0.35	B-	A-	A-	A-	1.75	0.05	-2.08	0.96	1.38	1.04
147	700966	10	4765	0.71	0.11	0.71	0.09	0.10	0.00	0.00	0.45	-0.18	0.45	-0.28	-0.23	A-	B-	A-	A+	0.09	0.05	-3.37	0.92	-4.27	0.84
148	700961	10	4765	0.46	0.46	0.38	0.09	0.07	0.00	0.00	0.19	0.19	-0.05	-0.07	-0.19	A-	A-	A-	A+	1.36	0.05	9.83	1.18	9.87	1.28
149	697077	10	4765	0.67	0.15	0.67	0.05	0.12	0.00	0.00	0.33	-0.19	0.33	-0.22	-0.11	A+	A-	A-	A+	0.28	0.05	3.22	1.08	3.68	1.13
150	697094	10	4765	0.27	0.36	0.27	0.26	0.10	0.00	0.00	0.06	0.16	0.06	-0.16	-0.10	A-	A+	A+	A+	2.30	0.05	7.93	1.21	9.90	1.68
151	697084	10	4765	0.74	0.05	0.09	0.11	0.74	0.00	0.00	0.36	-0.19	-0.15	-0.22	0.36	A+	A-	A+	A-	-0.13	0.06	1.20	1.03	1.95	1.09
152	697080	10	4765	0.29	0.58	0.29	0.08	0.05	0.00	0.00	0.19	-0.02	0.19	-0.07	-0.25	A-	A+	A+	A+	2.23	0.05	2.77	1.07	5.18	1.23
153	697083	10	4765	0.67	0.67	0.14	0.06	0.12	0.00	0.00	0.35	0.35	-0.20	-0.28	-0.07	A-	B-	A-	A+	0.28	0.05	0.66	1.02	2.65	1.09
154	697093	10	4765	0.51	0.08	0.07	0.51	0.34	0.00	0.00	0.26	-0.20	-0.24	0.26	-0.02	A+	A-	A-	A-	1.13	0.05	5.03	1.09	4.88	1.13
155	701439	11	4802	0.86	0.86	0.07	0.04	0.03	0.00	0.00	0.39	0.39	-0.22	-0.26	-0.15	A-	A-	A-	A+	-1.08	0.07	-0.53	0.98	-1.22	0.90
156	701434	11	4802	0.84	0.84	0.04	0.09	0.04	0.00	0.00	0.37	0.37	-0.27	-0.19	-0.17	A+	B-	B-	B+	-0.85	0.07	-0.21	0.99	-0.25	0.98
157	701442	11	4802	0.25	0.26	0.26	0.23	0.25	0.00	0.00	0.06	0.18	-0.07	-0.18	0.06	A-	A-	A-	A-	2.44	0.06	7.82	1.22	9.90	1.82
158	701435	11	4802	0.81	0.08	0.81	0.05	0.06	0.00	0.00	0.50	-0.32	0.50	-0.19	-0.29	A+	A-	A-	A-	-0.63	0.06	-3.92	0.86	-4.57	0.74
159	701436	11	4802	0.85	0.04	0.04	0.85	0.06	0.00	0.00	0.46	-0.26	-0.32	0.46	-0.19	A+	C-	C-	A+	-0.94	0.07	-5.10	0.80	-6.16	0.61

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
160	701443	11	4802	0.76	0.05	0.76	0.14	0.05	0.00	0.00	0.52	-0.27	0.52	-0.32	-0.22	A+	B-	A-	A-	-0.24	0.06	-5.00	0.85	-5.55	0.75
161	704056	11	4802	0.72	0.72	0.17	0.08	0.03	0.00	0.00	0.37	0.37	-0.27	-0.10	-0.20	B-	B-	B-	A+	0.00	0.06	1.67	1.05	-0.02	1.00
162	704063	11	4802	0.78	0.08	0.78	0.05	0.09	0.00	0.00	0.50	-0.26	0.50	-0.30	-0.24	A-	B-	C-	A+	-0.40	0.06	-4.54	0.86	-5.88	0.71
163	704059	11	4802	0.83	0.05	0.06	0.07	0.83	0.00	0.00	0.57	-0.32	-0.27	-0.33	0.57	A+	A-	B-	C-	-0.76	0.06	-3.65	0.87	-4.85	0.71
164	704061	11	4802	0.72	0.19	0.72	0.05	0.04	0.00	0.00	0.44	-0.19	0.44	-0.30	-0.28	A-	B-	A-	A+	-0.04	0.06	-1.42	0.96	-2.88	0.88
165	704066	11	4802	0.74	0.74	0.09	0.13	0.03	0.00	0.00	0.42	0.42	-0.26	-0.20	-0.22	A+	A-	A+	A+	-0.17	0.06	-1.53	0.95	-2.29	0.89
166	704062	11	4802	0.33	0.03	0.16	0.33	0.48	0.00	0.00	-0.12	-0.23	0.03	-0.12	0.16	A+	A-	A-	A-	1.98	0.05	9.90	1.50	9.90	2.16
167	701444	12	4762	0.80	0.11	0.07	0.80	0.02	0.00	0.00	0.48	-0.29	-0.28	0.48	-0.19	A-	A-	B-	A+	-0.61	0.06	-4.16	0.86	-4.34	0.76
168	701432	12	4762	0.83	0.83	0.09	0.05	0.03	0.00	0.00	0.23	0.23	-0.12	-0.16	-0.09	A+	C+	A+	A-	-0.83	0.07	4.43	1.19	4.47	1.33
169	701438	12	4762	0.44	0.12	0.10	0.33	0.44	0.00	0.00	0.31	-0.14	-0.22	-0.09	0.31	A+	A-	A-	A-	1.42	0.05	2.78	1.05	5.63	1.17
170	715522	12	4762	0.85	0.06	0.85	0.05	0.04	0.00	0.00	0.46	-0.29	0.46	-0.20	-0.26	A+	B-	B-	A-	-1.00	0.07	-2.11	0.91	-1.87	0.86
171	701437	12	4762	0.77	0.07	0.13	0.77	0.03	0.00	0.00	0.49	-0.26	-0.28	0.49	-0.24	A-	A-	A-	A+	-0.39	0.06	-3.47	0.89	-4.23	0.79
172	701441	12	4762	0.71	0.16	0.71	0.07	0.06	0.00	0.00	0.43	-0.24	0.43	-0.20	-0.24	A-	A-	A-	A+	0.02	0.05	-1.23	0.97	-1.03	0.96
173	704068	12	4762	0.79	0.06	0.04	0.11	0.79	0.00	0.00	0.35	-0.16	-0.23	-0.20	0.35	B-	A-	A-	A+	-0.52	0.06	0.15	1.00	0.30	1.02
174	704067	12	4762	0.88	0.08	0.03	0.88	0.02	0.00	0.00	0.50	-0.31	-0.31	0.50	-0.23	B-	B-	B-	A-	-1.24	0.07	-2.69	0.87	-4.09	0.69
175	704057	12	4762	0.64	0.07	0.25	0.64	0.04	0.00	0.00	0.33	-0.28	-0.10	0.33	-0.21	A-	A-	B-	A+	0.40	0.05	3.24	1.07	2.83	1.09
176	704065	12	4762	0.60	0.08	0.60	0.26	0.06	0.00	0.00	0.22	-0.24	0.22	0.04	-0.25	A-	A+	A-	A+	0.60	0.05	8.71	1.19	8.77	1.28
177	704064	12	4762	0.85	0.05	0.05	0.05	0.85	0.00	0.00	0.54	-0.29	-0.31	-0.28	0.54	A+	B-	B-	A-	-1.00	0.07	-3.90	0.84	-4.51	0.70
178	704060	12	4762	0.67	0.67	0.15	0.13	0.05	0.00	0.00	0.35	0.35	-0.25	-0.08	-0.21	A-	A-	A-	A-	0.24	0.05	2.70	1.07	2.50	1.09
179	701010	13	4782	0.80	0.01	0.03	0.15	0.80	0.00	0.00	0.51	-0.25	-0.29	-0.34	0.51	A-	B-	A-	A+	-0.52	0.06	-5.26	0.82	-6.90	0.65
180	701019	13	4782	0.81	0.05	0.03	0.81	0.12	0.00	0.00	0.31	-0.21	-0.24	0.31	-0.11	A+	A-	A-	A+	-0.55	0.06	0.64	1.02	1.62	1.10
181	701011	13	4782	0.72	0.72	0.11	0.04	0.13	0.00	0.00	0.38	0.38	-0.14	-0.27	-0.22	A+	A-	A+	A-	0.01	0.06	-0.22	0.99	-0.39	0.98
182	701008	13	4782	0.80	0.03	0.10	0.06	0.80	0.00	0.00	0.43	-0.24	-0.20	-0.28	0.43	A+	A-	A+	A-	-0.53	0.06	-3.26	0.89	-3.64	0.80
183	701015	13	4782	0.54	0.12	0.54	0.07	0.27	0.00	0.00	0.26	-0.16	0.26	-0.23	-0.04	A+	A-	A-	A-	0.95	0.05	6.52	1.13	6.99	1.19
184	701014	13	4782	0.44	0.06	0.48	0.44	0.03	0.00	0.00	0.33	-0.24	-0.14	0.33	-0.21	A+	A-	A-	A+	1.46	0.05	1.32	1.02	4.49	1.13
185	700989	13	4782	0.45	0.10	0.12	0.32	0.45	0.00	0.00	0.26	-0.30	-0.16	0.04	0.26	A+	A+	A+	A-	1.39	0.05	6.47	1.12	6.84	1.19
186	700985	13	4782	0.51	0.31	0.10	0.08	0.51	0.00	0.00	0.29	-0.03	-0.19	-0.26	0.29	A-	A-	A-	A+	1.11	0.05	4.48	1.08	5.40	1.14
187	700987	13	4782	0.49	0.04	0.31	0.49	0.16	0.00	0.00	0.12	-0.26	-0.09	0.12	0.08	A+	A+	A+	A-	1.19	0.05	9.90	1.25	9.90	1.37
188	700992	13	4782	0.37	0.37	0.08	0.25	0.30	0.00	0.00	0.11	0.11	-0.23	0.00	0.02	A-	A+	A-	A-	1.79	0.05	9.90	1.21	9.70	1.34
189	700986	13	4782	0.34	0.42	0.14	0.34	0.10	0.00	0.00	0.03	0.27	-0.19	0.03	-0.27	A-	A+	A-	A+	1.94	0.05	9.90	1.33	9.90	1.59
190	700990	13	4782	0.52	0.07	0.52	0.18	0.23	0.00	0.00	0.29	-0.22	0.29	-0.13	-0.09	A-	A-	A-	A-	1.04	0.05	3.85	1.07	3.51	1.09
191	701007	14	4810	0.27	0.33	0.32	0.27	0.08	0.00	0.00	0.12	-0.03	-0.01	0.12	-0.12	A+	A+	A-	A-	2.34	0.05	4.88	1.13	8.84	1.46

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
192	701009	14	4810	0.76	0.07	0.03	0.13	0.76	0.00	0.00	0.42	-0.23	-0.24	-0.22	0.42	A+	A-	A-	A+	-0.27	0.06	-1.75	0.95	-0.96	0.95
193	701012	14	4810	0.87	0.87	0.06	0.05	0.02	0.00	0.00	0.44	0.44	-0.24	-0.29	-0.20	A+	A+	A+	A-	-1.08	0.07	-2.58	0.89	-4.79	0.66
194	701017	14	4810	0.57	0.08	0.57	0.14	0.22	0.00	0.00	0.37	-0.20	0.37	-0.24	-0.11	A-	A-	A-	A-	0.82	0.05	0.87	1.02	1.90	1.05
195	701018	14	4810	0.84	0.84	0.07	0.04	0.04	0.00	0.00	0.48	0.48	-0.26	-0.32	-0.21	A+	A-	A-	A-	-0.86	0.07	-3.29	0.87	-4.75	0.70
196	701016	14	4810	0.40	0.11	0.37	0.40	0.11	0.00	0.00	0.08	-0.12	0.14	0.08	-0.21	A+	A+	A+	A-	1.62	0.05	9.90	1.33	9.90	1.49
197	700993	14	4810	0.57	0.07	0.09	0.27	0.57	0.00	0.00	0.50	-0.29	-0.31	-0.19	0.50	A-	A-	A-	A+	0.82	0.05	-6.80	0.87	-6.10	0.85
198	700994	14	4810	0.38	0.38	0.18	0.27	0.17	0.00	0.00	0.17	0.17	-0.06	-0.14	0.00	A-	A-	A+	A+	1.71	0.05	9.77	1.20	9.90	1.41
199	700988	14	4810	0.56	0.21	0.56	0.17	0.05	0.00	0.00	0.29	-0.06	0.29	-0.17	-0.23	A-	A-	A-	A+	0.85	0.05	5.74	1.11	6.11	1.17
200	700996	14	4810	0.40	0.40	0.07	0.16	0.37	0.00	0.00	0.26	0.26	-0.25	-0.04	-0.09	A-	A-	A+	A-	1.64	0.05	3.56	1.07	5.01	1.16
201	700991	14	4810	0.42	0.32	0.20	0.42	0.05	0.00	0.00	0.11	0.16	-0.19	0.11	-0.21	A+	A-	A-	A-	1.51	0.05	9.90	1.28	9.90	1.43
202	700995	14	4810	0.15	0.20	0.15	0.53	0.12	0.00	0.00	-0.03	0.01	-0.03	0.10	-0.13	A-	A-	A-	A-	3.16	0.07	3.57	1.15	9.90	2.25
203	701000	15	4773	0.86	0.03	0.86	0.05	0.05	0.00	0.00	0.27	-0.20	0.27	-0.07	-0.18	A-	B-	A-	A-	-0.99	0.07	1.14	1.05	2.17	1.18
204	701002	15	4773	0.64	0.64	0.12	0.11	0.12	0.00	0.00	0.27	0.27	-0.09	-0.13	-0.17	A+	A-	A-	A+	0.49	0.05	5.17	1.12	3.85	1.13
205	701001	15	4773	0.42	0.42	0.24	0.14	0.19	0.00	0.00	0.06	0.06	-0.05	-0.03	0.01	B-	A-	A-	A+	1.57	0.05	9.90	1.30	9.90	1.54
206	700982	15	4773	0.66	0.04	0.15	0.66	0.14	0.00	0.00	0.27	-0.22	-0.17	0.27	-0.06	A-	A-	A-	A+	0.35	0.05	4.38	1.11	4.08	1.15
207	701004	15	4773	0.65	0.11	0.10	0.14	0.65	0.00	0.00	0.36	-0.20	-0.17	-0.17	0.36	A+	A-	A-	A-	0.42	0.05	-0.29	0.99	-0.62	0.98
208	700981	15	4773	0.60	0.16	0.60	0.08	0.16	0.00	0.00	0.33	-0.05	0.33	-0.20	-0.24	A+	A-	A-	A+	0.70	0.05	2.07	1.04	1.77	1.05
209	702957	15	4773	0.63	0.21	0.63	0.13	0.03	0.00	0.00	0.34	-0.15	0.34	-0.22	-0.17	A+	A+	A+	A+	0.53	0.05	3.25	1.07	3.50	1.12
210	702952	15	4773	0.71	0.20	0.06	0.71	0.02	0.00	0.00	0.30	-0.10	-0.24	0.30	-0.24	B-	A-	B-	B-	0.11	0.05	5.01	1.14	5.09	1.22
211	702955	15	4773	0.83	0.83	0.07	0.03	0.07	0.00	0.00	0.41	0.41	-0.20	-0.26	-0.23	A-	B-	C-	A+	-0.71	0.06	-2.12	0.92	-2.43	0.84
212	702947	15	4773	0.55	0.55	0.04	0.04	0.37	0.00	0.00	0.20	0.20	-0.27	-0.23	0.00	B-	B-	A-	A+	0.94	0.05	9.90	1.21	9.55	1.28
213	702950	15	4773	0.70	0.08	0.70	0.12	0.10	0.00	0.00	0.41	-0.24	0.41	-0.21	-0.19	A+	A-	A-	A+	0.15	0.05	-2.02	0.95	-1.45	0.94
214	702951	15	4773	0.80	0.80	0.06	0.05	0.09	0.00	0.00	0.55	0.55	-0.31	-0.33	-0.26	A-	B-	A-	A+	-0.46	0.06	-5.30	0.83	-6.87	0.65
215	701003	16	4759	0.56	0.13	0.22	0.08	0.56	0.00	0.00	0.24	-0.06	-0.16	-0.11	0.24	A+	B-	A-	A+	0.91	0.05	9.20	1.19	7.51	1.23
216	700997	16	4759	0.51	0.11	0.17	0.51	0.21	0.00	0.00	0.25	-0.27	-0.11	0.25	0.00	A-	A-	A-	A-	1.15	0.05	7.84	1.15	8.51	1.25
217	700998	16	4759	0.63	0.21	0.10	0.07	0.63	0.00	0.00	0.33	-0.03	-0.28	-0.25	0.33	A+	A-	A-	A+	0.56	0.05	0.72	1.02	2.32	1.08
218	701006	16	4759	0.65	0.13	0.65	0.07	0.16	0.00	0.00	0.32	-0.22	0.32	-0.22	-0.06	A-	A-	A-	A+	0.45	0.05	4.51	1.11	3.95	1.14
219	700999	16	4759	0.49	0.09	0.16	0.26	0.49	0.00	0.00	0.36	-0.16	-0.16	-0.16	0.36	A+	A-	A-	A+	1.26	0.05	0.74	1.01	2.16	1.06
220	701005	16	4759	0.78	0.78	0.03	0.08	0.11	0.00	0.00	0.40	0.40	-0.24	-0.22	-0.19	A-	A-	A-	A+	-0.34	0.06	-0.82	0.97	-0.97	0.95
221	702956	16	4759	0.85	0.04	0.85	0.07	0.04	0.00	0.00	0.42	-0.25	0.42	-0.22	-0.22	A-	B-	C-	A+	-0.89	0.07	-2.33	0.90	-2.18	0.84
222	702958	16	4759	0.82	0.02	0.09	0.07	0.82	0.00	0.00	0.49	-0.26	-0.28	-0.29	0.49	A-	B-	B-	A+	-0.65	0.06	-4.06	0.85	-5.59	0.67
223	702953	16	4759	0.75	0.05	0.13	0.75	0.07	0.00	0.00	0.45	-0.32	-0.18	0.45	-0.26	A-	B-	B-	A+	-0.17	0.06	-3.72	0.89	-4.43	0.79

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
224	702954	16	4759	0.64	0.19	0.13	0.04	0.64	0.00	0.00	0.41	-0.11	-0.29	-0.29	0.41	A-	B-	A-	B+	0.50	0.05	-0.81	0.98	0.01	1.00
225	702948	16	4759	0.81	0.06	0.10	0.03	0.81	0.00	0.00	0.54	-0.35	-0.28	-0.26	0.54	A-	C-	C-	A+	-0.59	0.06	-5.56	0.81	-6.62	0.63
226	702949	16	4759	0.72	0.10	0.14	0.72	0.04	0.00	0.00	0.38	-0.29	-0.11	0.38	-0.22	A+	B-	A-	A+	0.06	0.05	-1.04	0.97	-0.66	0.97
227	701065	17	4757	0.82	0.82	0.06	0.03	0.09	0.00	0.00	0.37	0.37	-0.20	-0.24	-0.18	A-	C-	B-	A-	-0.68	0.06	-0.16	0.99	-1.04	0.94
228	701073	17	4757	0.59	0.31	0.06	0.59	0.03	0.00	0.00	0.37	-0.16	-0.31	0.37	-0.17	B-	B-	A-	A-	0.66	0.05	-0.93	0.98	-0.63	0.98
229	701074	17	4757	0.67	0.07	0.15	0.12	0.67	0.00	0.00	0.49	-0.25	-0.23	-0.27	0.49	A+	A-	A-	A-	0.27	0.05	-4.63	0.89	-4.85	0.85
230	701068	17	4757	0.23	0.04	0.69	0.23	0.04	0.00	0.00	-0.01	-0.26	0.23	-0.01	-0.28	A+	A+	A+	A-	2.50	0.06	7.78	1.23	9.90	1.78
231	701076	17	4757	0.81	0.05	0.81	0.04	0.10	0.00	0.00	0.49	-0.27	0.49	-0.25	-0.28	A-	B-	C-	A+	-0.62	0.06	-4.89	0.83	-6.32	0.68
232	701075	17	4757	0.59	0.06	0.15	0.59	0.20	0.00	0.00	0.27	-0.16	-0.14	0.27	-0.10	A+	A-	A+	A-	0.67	0.05	6.60	1.14	6.55	1.19
233	700980	17	4757	0.45	0.17	0.31	0.07	0.45	0.00	0.00	0.25	-0.17	-0.06	-0.13	0.25	A+	A-	A-	A+	1.33	0.05	6.20	1.11	6.09	1.17
234	700978	17	4757	0.28	0.20	0.28	0.36	0.15	0.00	0.00	0.08	-0.11	0.08	0.08	-0.08	A+	A-	A+	A+	2.20	0.05	8.96	1.22	9.90	1.64
235	700977	17	4757	0.48	0.11	0.27	0.14	0.48	0.00	0.00	0.36	-0.19	-0.14	-0.15	0.36	A-	A+	A-	A-	1.18	0.05	1.37	1.02	1.99	1.05
236	700956	17	4757	0.82	0.10	0.82	0.04	0.03	0.00	0.00	0.44	-0.24	0.44	-0.21	-0.28	B+	A-	A-	A+	-0.71	0.06	-2.14	0.92	-2.07	0.88
237	700974	17	4757	0.16	0.12	0.18	0.16	0.53	0.00	0.00	-0.18	-0.19	-0.17	-0.18	0.39	A-	A+	A-	A-	3.02	0.06	5.70	1.24	9.90	2.72
238	700955	17	4757	0.33	0.33	0.14	0.18	0.35	0.00	0.00	0.08	0.08	-0.20	-0.13	0.18	A-	A-	A-	A+	1.92	0.05	9.90	1.22	9.90	1.47
239	701067	18	4719	0.87	0.87	0.04	0.07	0.02	0.00	0.00	0.36	0.36	-0.21	-0.23	-0.15	A+	B-	B-	A+	-1.13	0.07	-0.82	0.96	-2.00	0.84
240	701066	18	4719	0.90	0.04	0.90	0.03	0.03	0.00	0.00	0.38	-0.18	0.38	-0.23	-0.22	A+	C-	C-	A+	-1.48	0.08	-3.04	0.84	-2.83	0.74
241	701077	18	4719	0.63	0.10	0.11	0.16	0.63	0.00	0.00	0.37	-0.17	-0.21	-0.17	0.37	A-	B-	A-	A-	0.48	0.05	1.93	1.04	1.94	1.06
242	701070	18	4719	0.86	0.86	0.05	0.04	0.05	0.00	0.00	0.41	0.41	-0.25	-0.22	-0.20	A+	B-	C-	A-	-1.04	0.07	-1.91	0.92	-2.61	0.81
243	701071	18	4719	0.74	0.09	0.09	0.08	0.74	0.00	0.00	0.42	-0.19	-0.19	-0.27	0.42	A+	A-	A-	A+	-0.11	0.06	-2.56	0.93	-1.85	0.92
244	701072	18	4719	0.69	0.06	0.69	0.17	0.07	0.00	0.00	0.36	-0.18	0.36	-0.22	-0.14	A-	A-	A-	A+	0.16	0.05	1.32	1.03	1.89	1.07
245	700976	18	4719	0.68	0.12	0.16	0.68	0.05	0.00	0.00	0.40	-0.23	-0.20	0.40	-0.19	A-	A-	A-	A+	0.22	0.05	-1.00	0.98	-0.73	0.97
246	700979	18	4719	0.58	0.58	0.12	0.22	0.08	0.00	0.00	0.33	0.33	-0.18	-0.13	-0.18	A-	A-	A+	A-	0.74	0.05	3.49	1.07	3.42	1.10
247	700972	18	4719	0.56	0.08	0.28	0.07	0.56	0.00	0.00	0.38	-0.29	-0.09	-0.25	0.38	A-	B-	B-	A+	0.84	0.05	-0.35	0.99	0.56	1.01
248	700975	18	4719	0.12	0.12	0.13	0.12	0.62	0.00	0.00	0.11	-0.16	-0.06	0.11	0.08	A-	A-	A-	A-	3.42	0.07	0.81	1.04	9.90	2.23
249	700958	18	4719	0.37	0.20	0.37	0.08	0.34	0.00	0.00	0.15	-0.02	0.15	-0.32	0.05	A-	A-	A+	A+	1.76	0.05	9.83	1.20	9.90	1.39
250	700973	18	4719	0.67	0.67	0.08	0.14	0.11	0.00	0.00	0.44	0.44	-0.23	-0.25	-0.19	A-	A-	A-	A-	0.27	0.05	-1.96	0.95	-2.39	0.92
251	701191	19	4747	0.50	0.33	0.05	0.12	0.50	0.00	0.00	0.08	0.10	-0.28	-0.09	0.08	A-	A-	A-	A-	1.10	0.05	9.90	1.30	9.90	1.39
252	701198	19	4747	0.66	0.06	0.06	0.22	0.66	0.00	0.00	0.52	-0.29	-0.28	-0.26	0.52	A-	A-	B-	A+	0.30	0.05	-5.46	0.87	-5.12	0.83
253	701196	19	4747	0.74	0.10	0.12	0.74	0.04	0.00	0.00	0.41	-0.19	-0.24	0.41	-0.20	A+	A-	A-	A+	-0.12	0.06	-1.45	0.96	-1.56	0.93
254	701200	19	4747	0.79	0.08	0.79	0.03	0.10	0.00	0.00	0.46	-0.24	0.46	-0.29	-0.23	A+	A-	B-	A-	-0.50	0.06	-1.35	0.95	-2.36	0.87
255	701194	19	4747	0.59	0.59	0.14	0.23	0.03	0.00	0.00	0.23	0.23	-0.13	-0.06	-0.22	A-	A-	A-	A+	0.67	0.05	7.45	1.16	8.43	1.27

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
256	701190	19	4747	0.73	0.05	0.11	0.73	0.11	0.00	0.00	0.37	-0.30	-0.19	0.37	-0.13	A+	A-	A-	A-	-0.10	0.06	0.28	1.01	-0.19	0.99
257	702780	19	4747	0.79	0.79	0.07	0.03	0.11	0.00	0.00	0.49	0.49	-0.32	-0.29	-0.20	A-	A-	A-	A-	-0.47	0.06	-4.27	0.86	-5.19	0.74
258	702791	19	4747	0.65	0.03	0.65	0.26	0.06	0.00	0.00	0.29	-0.18	0.29	-0.13	-0.19	A-	A-	A-	A+	0.36	0.05	2.91	1.07	2.53	1.09
259	702782	19	4747	0.58	0.11	0.16	0.58	0.15	0.00	0.00	0.33	-0.18	-0.11	0.33	-0.18	A-	A-	A-	A+	0.72	0.05	2.71	1.05	2.95	1.09
260	702786	19	4747	0.74	0.74	0.06	0.13	0.06	0.00	0.00	0.46	0.46	-0.29	-0.22	-0.22	A-	B-	A-	A-	-0.17	0.06	-1.94	0.94	-3.17	0.86
261	702784	19	4747	0.90	0.03	0.90	0.03	0.03	0.00	0.00	0.53	-0.27	0.53	-0.30	-0.29	A-	C-	B-	B+	-1.53	0.08	-5.48	0.72	-8.01	0.38
262	702788	19	4747	0.58	0.04	0.31	0.06	0.58	0.00	0.00	0.16	-0.27	0.12	-0.30	0.16	A-	A+	A+	A-	0.73	0.05	9.90	1.26	9.90	1.42
263	701199	20	4736	0.33	0.33	0.13	0.15	0.39	0.00	0.00	0.24	0.24	-0.24	-0.14	0.04	B-	A-	A-	A+	1.98	0.05	4.07	1.09	6.63	1.27
264	701192	20	4736	0.65	0.65	0.15	0.05	0.15	0.00	0.00	0.40	0.40	-0.16	-0.25	-0.21	A-	A-	A-	A-	0.38	0.05	-0.53	0.99	-1.16	0.96
265	701201	20	4736	0.62	0.16	0.08	0.62	0.14	0.00	0.00	0.37	-0.14	-0.24	0.37	-0.17	A-	A-	A-	A+	0.52	0.05	1.44	1.03	0.22	1.01
266	701193	20	4736	0.39	0.10	0.34	0.17	0.39	0.00	0.00	0.32	-0.23	-0.06	-0.15	0.32	A-	A-	A+	A+	1.66	0.05	1.06	1.02	3.28	1.11
267	701197	20	4736	0.68	0.12	0.68	0.14	0.06	0.00	0.00	0.44	-0.24	0.44	-0.19	-0.25	A+	A-	A-	A-	0.21	0.05	-2.15	0.95	-2.84	0.90
268	701195	20	4736	0.56	0.21	0.56	0.07	0.15	0.00	0.00	0.18	-0.06	0.18	-0.21	-0.01	A-	A+	A+	A-	0.80	0.05	9.90	1.21	9.90	1.34
269	702792	20	4736	0.94	0.94	0.02	0.02	0.02	0.00	0.00	0.38	0.38	-0.24	-0.20	-0.19	A-	B-	B-	A-	-2.07	0.10	-1.56	0.88	-3.69	0.58
270	702783	20	4736	0.48	0.11	0.07	0.35	0.48	0.00	0.00	0.41	-0.22	-0.18	-0.19	0.41	A-	A-	A-	A-	1.22	0.05	-1.80	0.97	-0.82	0.98
271	702787	20	4736	0.28	0.08	0.12	0.52	0.28	0.00	0.00	0.23	-0.14	-0.30	0.08	0.23	A-	A-	A-	A+	2.27	0.05	2.53	1.06	6.76	1.33
272	702781	20	4736	0.70	0.10	0.14	0.70	0.06	0.00	0.00	0.45	-0.21	-0.24	0.45	-0.25	B-	B-	A-	A+	0.07	0.05	-3.38	0.92	-3.82	0.85
273	702785	20	4736	0.63	0.10	0.63	0.11	0.16	0.00	0.00	0.37	-0.16	0.37	-0.20	-0.17	A-	B-	A-	A-	0.48	0.05	0.62	1.01	1.29	1.04
274	702789	20	4736	0.43	0.22	0.28	0.43	0.07	0.00	0.00	0.26	-0.19	0.01	0.26	-0.20	A-	B-	A-	A-	1.45	0.05	2.99	1.06	5.63	1.17
275	697086	21	4735	0.31	0.31	0.17	0.06	0.46	0.00	0.00	-0.01	-0.01	-0.18	-0.06	0.17	A+	A+	A+	A-	2.10	0.05	9.90	1.32	9.90	1.77
276	697091	21	4735	0.44	0.24	0.12	0.20	0.44	0.00	0.00	0.31	-0.10	-0.21	-0.10	0.31	A-	A+	A-	A-	1.42	0.05	2.23	1.04	2.78	1.08
277	697249	21	4735	0.44	0.06	0.44	0.45	0.05	0.00	0.00	0.19	-0.21	0.19	-0.01	-0.19	A-	A-	A+	A-	1.44	0.05	8.26	1.15	8.37	1.24
278	697250	21	4735	0.59	0.11	0.25	0.59	0.06	0.00	0.00	0.37	-0.22	-0.18	0.37	-0.14	A-	A-	A+	A+	0.70	0.05	2.61	1.05	2.52	1.07
279	697089	21	4735	0.50	0.15	0.06	0.50	0.29	0.00	0.00	0.08	-0.13	-0.23	0.08	0.14	A+	A+	A+	A-	1.14	0.05	9.90	1.30	9.90	1.42
280	697085	21	4735	0.49	0.16	0.49	0.07	0.29	0.00	0.00	0.30	-0.14	0.30	-0.22	-0.09	A+	A-	A-	A-	1.21	0.05	4.49	1.08	4.94	1.13
281	683426	21	4735	0.65	0.03	0.22	0.65	0.10	0.00	0.00	0.40	-0.27	-0.25	0.40	-0.13	A-	B-	A-	A-	0.38	0.05	-0.90	0.98	-1.49	0.95
282	683428	21	4735	0.68	0.03	0.68	0.19	0.10	0.00	0.00	0.42	-0.19	0.42	-0.31	-0.14	C-	A-	B-	A-	0.23	0.05	-0.70	0.98	-2.09	0.93
283	683425	21	4735	0.59	0.25	0.59	0.11	0.04	0.00	0.00	0.42	-0.18	0.42	-0.27	-0.21	A-	A-	A-	A-	0.69	0.05	-1.20	0.98	-0.90	0.97
284	683432	21	4735	0.55	0.16	0.15	0.55	0.14	0.00	0.00	0.33	-0.14	-0.23	0.33	-0.10	A-	A-	A-	A-	0.92	0.05	2.63	1.05	2.07	1.05
285	683422	21	4735	0.72	0.72	0.09	0.11	0.08	0.00	0.00	0.37	0.37	-0.20	-0.11	-0.26	A-	A-	A-	A+	-0.01	0.06	-1.38	0.96	1.23	1.05
286	683421	21	4735	0.47	0.09	0.33	0.47	0.11	0.00	0.00	0.31	-0.20	-0.10	0.31	-0.17	A-	A-	A-	A-	1.28	0.05	0.81	1.01	2.25	1.06
287	697086	22	4730	0.30	0.30	0.15	0.06	0.49	0.00	0.00	0.01	0.01	-0.23	-0.10	0.21	A+	A-	A+	A+	2.10	0.05	9.90	1.32	9.90	1.77

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
288	697252	22	4730	0.58	0.08	0.58	0.30	0.03	0.00	0.00	0.26	-0.10	0.26	-0.13	-0.22	A-	A+	A+	A+	0.77	0.05	6.37	1.13	5.60	1.16
289	697088	22	4730	0.37	0.26	0.15	0.37	0.22	0.00	0.00	0.30	-0.02	-0.18	0.30	-0.16	A-	A+	A-	A+	1.78	0.05	1.73	1.03	5.61	1.19
290	697090	22	4730	0.43	0.47	0.03	0.07	0.43	0.00	0.00	0.34	-0.16	-0.26	-0.15	0.34	A+	A+	A+	A-	1.49	0.05	-0.02	1.00	1.27	1.04
291	697251	22	4730	0.69	0.69	0.08	0.19	0.04	0.00	0.00	0.38	0.38	-0.24	-0.13	-0.28	A+	A-	A-	A+	0.19	0.05	-0.78	0.98	-0.18	0.99
292	697248	22	4730	0.65	0.65	0.11	0.15	0.09	0.00	0.00	0.37	0.37	-0.26	-0.14	-0.15	A-	A-	A-	A+	0.40	0.05	-2.16	0.95	-1.20	0.96
293	683427	22	4730	0.73	0.12	0.10	0.06	0.73	0.00	0.00	0.54	-0.29	-0.26	-0.28	0.54	A-	C-	B-	A-	-0.03	0.06	-6.18	0.84	-6.62	0.74
294	683429	22	4730	0.69	0.06	0.69	0.23	0.02	0.00	0.00	0.38	-0.26	0.38	-0.19	-0.22	A-	A-	B-	A+	0.18	0.05	0.08	1.00	0.72	1.03
295	683430	22	4730	0.79	0.06	0.09	0.06	0.79	0.00	0.00	0.55	-0.24	-0.35	-0.28	0.55	A-	A-	A-	A+	-0.42	0.06	-5.98	0.81	-6.33	0.69
296	683424	22	4730	0.33	0.13	0.05	0.48	0.33	0.00	0.00	0.19	-0.11	-0.19	-0.01	0.19	A-	A-	A-	A+	1.98	0.05	5.59	1.12	9.08	1.37
297	683431	22	4730	0.56	0.56	0.07	0.19	0.17	0.00	0.00	0.22	0.22	-0.30	-0.06	-0.01	A-	A-	A-	A+	0.84	0.05	8.14	1.16	7.84	1.22
298	683423	22	4730	0.23	0.23	0.34	0.23	0.19	0.00	0.00	-0.03	-0.03	0.09	0.03	-0.11	A-	A-	A-	A+	2.56	0.06	8.48	1.26	9.90	1.88
299	701627	23	4735	0.89	0.04	0.89	0.02	0.05	0.00	0.00	0.43	-0.23	0.43	-0.20	-0.28	A-	C-	B-	C+	-1.35	0.08	-4.88	0.76	-5.70	0.55
300	701626	23	4735	0.66	0.16	0.10	0.08	0.66	0.00	0.00	0.39	-0.22	-0.19	-0.17	0.39	A-	A-	A-	A-	0.33	0.05	-0.35	0.99	0.03	1.00
301	701628	23	4735	0.80	0.80	0.13	0.06	0.02	0.00	0.00	0.50	0.50	-0.31	-0.30	-0.19	A-	A-	A-	A+	-0.50	0.06	-3.44	0.88	-5.13	0.73
302	701621	23	4735	0.75	0.05	0.75	0.05	0.15	0.00	0.00	0.49	-0.28	0.49	-0.21	-0.28	A-	A-	A+	B-	-0.16	0.06	-3.22	0.91	-4.63	0.79
303	701618	23	4735	0.30	0.52	0.11	0.30	0.06	0.00	0.00	0.09	0.10	-0.08	0.09	-0.27	A-	A+	A-	A-	2.15	0.05	8.46	1.20	9.90	1.56
304	701625	23	4735	0.85	0.05	0.03	0.85	0.07	0.00	0.00	0.47	-0.21	-0.24	0.47	-0.30	A-	C-	A-	A+	-0.93	0.07	-3.62	0.85	-3.76	0.74
305	701021	23	4735	0.66	0.66	0.07	0.15	0.11	0.00	0.00	0.32	0.32	-0.26	-0.07	-0.17	B-	A-	A-	A-	0.32	0.05	3.97	1.10	3.68	1.14
306	701027	23	4735	0.51	0.22	0.22	0.51	0.05	0.00	0.00	0.30	-0.08	-0.17	0.30	-0.20	A-	A+	A-	A+	1.08	0.05	3.28	1.06	4.63	1.13
307	701023	23	4735	0.45	0.20	0.45	0.30	0.05	0.00	0.00	0.26	-0.15	0.26	-0.05	-0.18	A-	A+	A-	A+	1.41	0.05	5.07	1.09	7.95	1.24
308	701024	23	4735	0.79	0.79	0.05	0.08	0.08	0.00	0.00	0.39	0.39	-0.21	-0.15	-0.27	A+	A-	B-	A+	-0.45	0.06	-1.41	0.95	0.65	1.04
309	701025	23	4735	0.54	0.15	0.07	0.23	0.54	0.00	0.00	0.40	-0.23	-0.22	-0.13	0.40	A-	A-	A-	A+	0.93	0.05	-1.77	0.97	-0.72	0.98
310	701020	23	4735	0.10	0.34	0.10	0.15	0.41	0.00	0.00	-0.12	0.18	-0.12	-0.20	0.06	A-	A+	A-	A-	3.71	0.08	1.51	1.08	9.90	3.77
311	701616	24	4756	0.51	0.51	0.44	0.02	0.03	0.00	0.00	0.14	0.14	-0.03	-0.21	-0.14	A-	A-	B-	A-	1.07	0.05	9.90	1.21	9.90	1.29
312	703953	24	4756	0.83	0.83	0.08	0.06	0.03	0.00	0.00	0.45	0.45	-0.23	-0.28	-0.24	A-	B-	A-	A-	-0.78	0.06	-2.99	0.89	-4.64	0.73
313	701617	24	4756	0.51	0.02	0.23	0.51	0.24	0.00	0.00	0.25	-0.22	-0.09	0.25	-0.13	A-	A-	A-	A+	1.09	0.05	5.12	1.09	5.76	1.14
314	701623	24	4756	0.59	0.04	0.30	0.07	0.59	0.00	0.00	0.22	-0.25	0.01	-0.23	0.22	A-	A-	A-	A+	0.69	0.05	6.34	1.12	7.04	1.19
315	701619	24	4756	0.84	0.05	0.05	0.06	0.84	0.00	0.00	0.48	-0.24	-0.29	-0.25	0.48	A-	A-	B-	A-	-0.81	0.07	-5.19	0.80	-5.91	0.66
316	701624	24	4756	0.64	0.15	0.64	0.08	0.13	0.00	0.00	0.26	-0.01	0.26	-0.27	-0.12	A+	A-	A-	A+	0.45	0.05	5.41	1.12	5.68	1.18
317	700984	24	4756	0.44	0.21	0.21	0.44	0.14	0.00	0.00	0.06	-0.03	0.00	0.06	-0.04	A-	A-	A-	A-	1.43	0.05	9.90	1.27	9.90	1.42
318	701029	24	4756	0.46	0.16	0.10	0.28	0.46	0.00	0.00	0.17	-0.12	-0.15	0.01	0.17	A+	A-	A-	A+	1.32	0.05	9.03	1.16	8.62	1.23
319	701026	24	4756	0.28	0.19	0.44	0.09	0.28	0.00	0.00	0.14	-0.20	0.13	-0.16	0.14	A-	A-	A-	A+	2.21	0.05	5.71	1.14	8.15	1.38

Item Information			Classical													DIF				Rasch		Infit		Outfit	
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***	**MS**	***z***	**MS**
320	701028	24	4756	0.26	0.39	0.26	0.29	0.07	0.00	0.00	0.00	0.24	0.00	-0.11	-0.26	A-	A+	B+	A-	2.36	0.05	7.05	1.20	9.90	1.57
321	700983	24	4756	0.19	0.19	0.12	0.24	0.45	0.00	0.00	0.04	0.04	-0.19	-0.03	0.12	A-	A-	A-	A+	2.77	0.06	4.32	1.15	9.90	1.79
322	701022	24	4756	0.57	0.05	0.29	0.57	0.08	0.00	0.00	0.39	-0.28	-0.17	0.39	-0.18	A-	A-	A-	A-	0.79	0.05	-0.46	0.99	0.05	1.00

Table J–8. Algebra I Multiple-Choice Item Statistics: Summer

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	605225	0	1846	0.91	0.06	0.91	0.02	0.01	0.00	0.00	0.25	-0.17	0.25	-0.14	-0.12	-1.74	0.07	-7.08	0.70	-6.53	0.60
2	674458	0	1846	0.39	0.19	0.30	0.12	0.39	0.00	0.00	0.39	-0.12	-0.24	-0.10	0.39	0.79	0.05	-4.71	0.92	-4.63	0.89
3	640555	0	1846	0.36	0.10	0.10	0.36	0.44	0.00	0.00	0.21	-0.08	-0.04	0.21	-0.13	0.51	0.05	2.03	1.03	1.67	1.04
4	631597	0	1846	0.65	0.22	0.11	0.65	0.02	0.00	0.00	0.29	-0.09	-0.25	0.29	-0.15	-0.69	0.05	4.22	1.10	2.75	1.11
5	681810	0	1846	0.42	0.42	0.29	0.16	0.13	0.00	0.00	0.04	0.04	-0.07	0.07	-0.04	1.04	0.05	9.90	1.28	9.90	1.40
6	605216	0	1846	0.55	0.11	0.21	0.55	0.13	0.00	0.00	0.20	-0.08	-0.15	0.20	-0.03	-0.23	0.05	7.12	1.13	5.73	1.17
7	632351	0	1846	0.55	0.20	0.55	0.10	0.14	0.00	0.00	0.32	-0.12	0.32	-0.21	-0.13	0.15	0.05	-1.90	0.97	-1.97	0.96
8	605205	0	1846	0.31	0.14	0.31	0.24	0.31	0.00	0.00	0.21	-0.10	-0.03	-0.11	0.21	1.33	0.05	3.78	1.10	4.71	1.17
9	674495	0	1846	0.63	0.63	0.11	0.18	0.09	0.00	0.00	0.32	0.32	-0.20	-0.14	-0.15	-0.22	0.05	-2.84	0.95	-2.18	0.94
10	605264	0	1846	0.61	0.09	0.61	0.14	0.15	0.00	0.00	0.23	-0.10	0.23	-0.18	-0.05	-0.46	0.05	5.39	1.12	4.22	1.14
11	674454	0	1846	0.35	0.35	0.16	0.33	0.15	0.01	0.00	0.12	0.12	-0.13	-0.01	-0.01	1.15	0.05	6.98	1.16	7.58	1.25
12	674516	0	1846	0.60	0.11	0.20	0.60	0.10	0.00	0.00	0.31	-0.15	-0.13	0.31	-0.19	-0.15	0.05	-1.65	0.97	2.32	1.06
13	605203	0	1846	0.48	0.31	0.48	0.16	0.05	0.00	0.00	0.24	-0.12	0.24	-0.12	-0.09	0.51	0.05	2.71	1.04	3.10	1.07
14	681809	0	1846	0.46	0.46	0.25	0.12	0.17	0.00	0.00	0.12	0.12	-0.06	-0.11	0.02	0.68	0.05	8.98	1.16	9.15	1.23
15	674493	0	1846	0.28	0.24	0.22	0.25	0.28	0.00	0.00	0.24	-0.02	-0.16	-0.07	0.24	1.40	0.06	1.43	1.04	2.40	1.09
16	674456	0	1846	0.63	0.05	0.26	0.63	0.06	0.00	0.00	0.34	-0.14	-0.21	0.34	-0.16	-0.51	0.05	1.37	1.03	-0.15	0.99
17	640534	0	1846	0.40	0.40	0.18	0.30	0.12	0.00	0.00	0.23	0.23	-0.07	-0.10	-0.11	0.81	0.05	3.02	1.05	3.14	1.08
18	674509	0	1846	0.68	0.08	0.09	0.15	0.68	0.00	0.00	0.48	-0.21	-0.29	-0.24	0.48	-0.64	0.05	-5.76	0.87	-5.61	0.81
19	605175	0	1846	0.75	0.20	0.02	0.04	0.75	0.00	0.00	0.26	-0.15	-0.17	-0.16	0.26	-0.69	0.05	-3.88	0.91	-2.28	0.92
20	678811	0	1846	0.61	0.09	0.61	0.24	0.07	0.00	0.00	0.31	-0.18	0.31	-0.16	-0.13	-0.03	0.05	-2.86	0.95	-2.52	0.94
21	682039	0	1846	0.70	0.70	0.11	0.06	0.13	0.00	0.00	0.18	0.18	-0.12	-0.13	-0.04	-0.30	0.05	-0.41	0.99	-0.76	0.98
22	632154	0	1846	0.70	0.07	0.70	0.08	0.15	0.00	0.00	0.29	-0.13	0.29	-0.19	-0.13	-0.59	0.05	-1.57	0.96	-1.41	0.95
23	605246	0	1846	0.75	0.07	0.12	0.75	0.06	0.00	0.00	0.19	-0.09	-0.09	0.19	-0.12	-0.86	0.06	0.29	1.01	0.77	1.03
24	633004	0	1846	0.51	0.21	0.15	0.12	0.51	0.00	0.00	0.18	-0.10	-0.03	-0.12	0.18	0.66	0.05	7.95	1.14	7.49	1.18
25	678802	0	1846	0.63	0.09	0.21	0.63	0.07	0.00	0.00	0.26	-0.10	-0.18	0.26	-0.08	-0.06	0.05	-0.60	0.99	-0.68	0.98
26	605306	0	1846	0.62	0.06	0.62	0.14	0.18	0.00	0.00	0.34	-0.17	0.34	-0.14	-0.19	-0.57	0.05	2.17	1.05	1.71	1.06
27	666556	0	1846	0.68	0.68	0.12	0.10	0.10	0.00	0.00	0.34	0.34	-0.17	-0.15	-0.20	-0.47	0.05	-3.97	0.92	-3.92	0.88
28	678777	0	1846	0.68	0.21	0.68	0.09	0.02	0.00	0.00	0.31	-0.20	0.31	-0.15	-0.12	-0.37	0.05	-3.97	0.92	-2.94	0.91
29	678765	0	1846	0.55	0.23	0.11	0.55	0.10	0.00	0.00	0.15	-0.05	-0.09	0.15	-0.07	0.24	0.05	5.88	1.09	7.94	1.18
30	674153	0	1846	0.63	0.63	0.07	0.18	0.12	0.00	0.00	0.35	0.35	-0.20	-0.16	-0.17	-0.29	0.05	-3.28	0.94	-2.72	0.92
31	633411	0	1846	0.67	0.14	0.12	0.07	0.67	0.00	0.00	0.26	-0.12	-0.14	-0.13	0.26	-0.40	0.05	-1.14	0.98	-0.39	0.99

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	681814	0	1846	0.36	0.36	0.25	0.23	0.16	0.00	0.00	0.29	0.29	-0.18	-0.05	-0.10	0.75	0.05	-1.90	0.97	-0.98	0.98
33	632156	0	1846	0.43	0.09	0.29	0.19	0.43	0.00	0.00	0.38	-0.11	-0.18	-0.18	0.38	0.58	0.05	-4.92	0.92	-4.02	0.91
34	641533	0	1846	0.25	0.34	0.25	0.33	0.07	0.00	0.00	0.10	-0.10	0.10	0.07	-0.11	1.67	0.06	5.65	1.18	7.83	1.38
35	678766	0	1846	0.41	0.09	0.14	0.36	0.41	0.00	0.00	0.22	-0.13	-0.09	-0.08	0.22	0.66	0.05	2.29	1.04	2.37	1.06
36	678736	0	1846	0.60	0.08	0.19	0.60	0.13	0.00	0.00	0.31	-0.11	-0.18	0.31	-0.14	-0.40	0.05	3.11	1.06	1.46	1.05

Table J–9. Biology Multiple-Choice Item Statistics: Summer

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	607724	0	1116	0.46	0.22	0.10	0.46	0.22	0.00	0.00	0.21	-0.14	-0.10	0.21	-0.04	0.07	0.06	3.09	1.06	1.72	1.05
2	671335	0	1116	0.51	0.51	0.32	0.10	0.07	0.00	0.00	0.26	0.26	-0.15	-0.16	-0.05	0.16	0.06	0.27	1.00	-0.42	0.99
3	674117	0	1116	0.71	0.09	0.15	0.71	0.05	0.00	0.00	0.21	-0.14	-0.10	0.21	-0.09	-0.79	0.07	0.32	1.01	0.44	1.02
4	642329	0	1116	0.35	0.05	0.12	0.48	0.35	0.00	0.00	0.18	-0.16	-0.19	0.02	0.18	0.60	0.06	0.59	1.01	0.91	1.03
5	678418	0	1116	0.50	0.50	0.08	0.34	0.09	0.00	0.00	0.32	0.32	-0.17	-0.11	-0.22	0.50	0.06	0.59	1.01	0.39	1.01
6	642837	0	1116	0.48	0.10	0.30	0.12	0.48	0.00	0.00	0.33	-0.21	-0.10	-0.16	0.33	0.34	0.06	-1.84	0.97	-1.64	0.96
7	611014	0	1116	0.47	0.13	0.28	0.12	0.47	0.00	0.00	0.22	-0.11	-0.07	-0.12	0.22	0.15	0.06	2.02	1.04	1.02	1.03
8	635759	0	1116	0.50	0.50	0.24	0.17	0.09	0.00	0.00	0.23	0.23	-0.05	-0.10	-0.20	0.21	0.06	1.72	1.03	1.52	1.04
9	674139	0	1116	0.68	0.14	0.68	0.11	0.07	0.00	0.00	0.34	-0.13	0.34	-0.23	-0.15	-0.35	0.06	-5.49	0.89	-4.54	0.86
10	610887	0	1116	0.48	0.23	0.12	0.17	0.48	0.00	0.00	0.29	-0.07	-0.24	-0.10	0.29	0.28	0.06	-0.89	0.98	-1.05	0.97
11	679666	0	1116	0.61	0.61	0.12	0.16	0.11	0.00	0.00	0.27	0.27	-0.11	-0.19	-0.08	-0.67	0.07	4.06	1.11	2.35	1.10
12	679665	0	1116	0.59	0.12	0.11	0.59	0.18	0.00	0.00	0.13	-0.13	-0.05	0.13	-0.01	-0.01	0.06	4.47	1.08	4.20	1.12
13	678539	0	1116	0.32	0.32	0.18	0.34	0.15	0.00	0.00	0.20	0.20	-0.12	0.00	-0.12	0.69	0.06	-1.04	0.98	0.28	1.01
14	607702	0	1116	0.59	0.59	0.11	0.16	0.14	0.00	0.00	0.29	0.29	-0.16	-0.06	-0.19	-0.45	0.07	1.48	1.03	1.42	1.05
15	607107	0	1116	0.54	0.10	0.54	0.18	0.18	0.00	0.00	0.30	-0.25	0.30	-0.13	-0.06	-0.08	0.06	-1.06	0.98	-1.29	0.96
16	607735	0	1116	0.42	0.23	0.15	0.20	0.42	0.00	0.00	0.28	-0.01	-0.21	-0.14	0.28	0.37	0.06	-1.26	0.98	-1.32	0.97
17	671328	0	1116	0.59	0.10	0.22	0.09	0.59	0.00	0.00	0.34	-0.16	-0.14	-0.21	0.34	-0.58	0.07	1.78	1.05	0.89	1.03
18	677984	0	1116	0.25	0.24	0.25	0.30	0.21	0.00	0.00	0.18	0.00	0.18	-0.20	0.04	0.96	0.07	-3.10	0.92	-1.18	0.96
19	611387	0	1116	0.68	0.15	0.05	0.68	0.12	0.00	0.00	0.30	-0.17	-0.18	0.30	-0.11	0.28	0.06	2.64	1.05	2.14	1.06
20	609821	0	1116	0.61	0.11	0.17	0.61	0.10	0.00	0.00	0.30	-0.14	-0.17	0.30	-0.12	-0.28	0.06	-1.71	0.97	-1.63	0.95
21	611384	0	1116	0.52	0.09	0.52	0.22	0.16	0.00	0.00	0.36	-0.15	0.36	-0.21	-0.11	-0.12	0.06	-2.00	0.96	-2.07	0.94
22	635764	0	1116	0.48	0.18	0.48	0.12	0.22	0.00	0.00	0.31	-0.12	0.31	-0.08	-0.19	-0.04	0.06	0.07	1.00	-0.92	0.97
23	610459	0	1116	0.57	0.10	0.21	0.11	0.57	0.00	0.00	0.35	-0.10	-0.19	-0.19	0.35	0.03	0.06	-3.92	0.93	-3.53	0.91
24	607571	0	1116	0.32	0.39	0.16	0.13	0.32	0.00	0.00	0.29	-0.01	-0.14	-0.22	0.29	0.96	0.07	-1.61	0.96	-0.66	0.98
25	611398	0	1116	0.34	0.10	0.20	0.36	0.34	0.00	0.00	0.25	-0.19	-0.14	-0.01	0.25	0.59	0.06	-1.81	0.96	-1.39	0.96
26	679982	0	1116	0.36	0.14	0.30	0.36	0.20	0.00	0.00	0.21	-0.14	-0.06	0.21	-0.06	0.56	0.06	0.08	1.00	0.34	1.01
27	678303	0	1116	0.35	0.35	0.31	0.19	0.15	0.00	0.00	0.20	0.20	-0.18	-0.03	0.00	1.03	0.07	3.15	1.09	4.57	1.19
28	611926	0	1116	0.46	0.23	0.46	0.21	0.11	0.00	0.00	0.26	-0.06	0.26	-0.16	-0.12	0.07	0.06	1.19	1.02	0.63	1.02
29	612587	0	1116	0.61	0.15	0.08	0.61	0.16	0.00	0.00	0.34	-0.21	-0.23	0.34	-0.08	-0.09	0.06	-4.08	0.93	-3.72	0.90
30	682252	0	1116	0.54	0.08	0.26	0.54	0.12	0.00	0.00	0.32	-0.20	-0.17	0.32	-0.09	0.12	0.06	-2.17	0.96	-2.00	0.95
31	641211	0	1116	0.39	0.39	0.22	0.25	0.14	0.00	0.00	0.23	0.23	-0.13	-0.02	-0.15	0.97	0.07	4.90	1.14	4.47	1.17

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	677882	0	1116	0.44	0.44	0.22	0.10	0.23	0.00	0.00	0.25	0.25	-0.03	-0.22	-0.11	0.32	0.06	0.28	1.00	0.36	1.01
33	611507	0	1116	0.50	0.10	0.12	0.50	0.27	0.00	0.00	0.26	-0.19	-0.12	0.26	-0.08	0.23	0.06	0.32	1.01	0.69	1.02
34	683582	0	1116	0.66	0.66	0.14	0.15	0.05	0.00	0.00	0.38	0.38	-0.23	-0.17	-0.18	-0.75	0.07	-0.82	0.98	-1.70	0.93
35	682196	0	1116	0.59	0.59	0.12	0.19	0.09	0.00	0.00	0.34	0.34	-0.20	-0.17	-0.11	-0.34	0.06	-1.58	0.97	-0.86	0.97
36	682198	0	1116	0.74	0.10	0.74	0.08	0.08	0.00	0.00	0.32	-0.15	0.32	-0.23	-0.12	-1.08	0.07	-0.29	0.99	-0.37	0.98
37	611921	0	1116	0.75	0.09	0.75	0.13	0.03	0.00	0.00	0.29	-0.19	0.29	-0.12	-0.15	-1.02	0.07	-1.70	0.94	-1.95	0.90
38	643346	0	1116	0.49	0.09	0.20	0.21	0.49	0.00	0.00	0.32	-0.18	-0.03	-0.23	0.32	0.31	0.06	-1.55	0.97	-1.27	0.97
39	611012	0	1116	0.57	0.13	0.20	0.57	0.10	0.00	0.00	0.32	-0.21	-0.03	0.32	-0.23	0.09	0.06	-2.27	0.96	-2.15	0.95
40	611925	0	1116	0.61	0.19	0.11	0.09	0.61	0.00	0.00	0.24	-0.14	-0.01	-0.19	0.24	-0.31	0.06	0.11	1.00	0.89	1.03
41	681522	0	1116	0.39	0.39	0.18	0.19	0.24	0.00	0.00	0.28	0.28	-0.11	-0.12	-0.10	0.83	0.07	1.08	1.03	1.25	1.04
42	611465	0	1116	0.32	0.08	0.25	0.34	0.32	0.00	0.00	0.22	-0.16	-0.08	-0.05	0.22	0.92	0.07	0.18	1.00	0.51	1.02
43	608357	0	1116	0.63	0.04	0.63	0.26	0.07	0.00	0.00	0.32	-0.19	0.32	-0.14	-0.20	-0.03	0.06	-3.33	0.94	-3.13	0.92
44	607754	0	1116	0.45	0.45	0.18	0.15	0.21	0.00	0.00	0.36	0.36	-0.19	-0.26	-0.03	0.67	0.06	-0.13	1.00	-0.23	0.99
45	642902	0	1116	0.60	0.08	0.14	0.18	0.60	0.00	0.00	0.38	-0.21	-0.20	-0.15	0.38	-0.57	0.07	-0.19	1.00	-0.01	1.00
46	611053	0	1116	0.62	0.11	0.62	0.16	0.10	0.00	0.00	0.36	-0.16	0.36	-0.15	-0.22	-0.45	0.07	-2.29	0.95	-2.24	0.92
47	679258	0	1116	0.46	0.24	0.46	0.18	0.12	0.00	0.00	0.23	-0.08	0.23	-0.13	-0.09	-0.10	0.06	3.98	1.08	5.05	1.15
48	608358	0	1116	0.60	0.60	0.07	0.17	0.16	0.00	0.00	0.34	0.34	-0.15	-0.14	-0.20	0.14	0.06	-2.76	0.95	-2.77	0.93

Table J–10. Literature Multiple-Choice Item Statistics: Summer

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	641566	0	588	0.53	0.06	0.20	0.21	0.53	0.00	0.00	0.17	-0.20	-0.11	0.01	0.17	0.61	0.09	4.23	1.12	3.50	1.14
2	641544	0	588	0.44	0.44	0.44	0.07	0.06	0.00	0.00	0.14	0.14	-0.04	-0.12	-0.08	0.79	0.09	3.81	1.12	4.77	1.21
3	641564	0	588	0.65	0.65	0.05	0.11	0.18	0.00	0.00	0.23	0.23	-0.09	-0.21	-0.06	-0.44	0.09	2.05	1.08	1.74	1.10
4	641561	0	588	0.49	0.07	0.49	0.31	0.13	0.00	0.00	0.15	-0.11	0.15	0.02	-0.16	0.86	0.09	5.61	1.18	5.28	1.24
5	641547	0	588	0.77	0.09	0.09	0.05	0.77	0.00	0.00	0.26	-0.12	-0.19	-0.10	0.26	-0.64	0.10	-2.64	0.88	-1.62	0.90
6	641562	0	588	0.43	0.43	0.05	0.02	0.49	0.00	0.00	0.15	0.15	-0.21	-0.21	0.00	0.86	0.09	3.84	1.12	4.45	1.20
7	641542	0	588	0.67	0.10	0.67	0.09	0.13	0.00	0.00	0.30	-0.24	0.30	-0.17	-0.05	-0.35	0.09	-0.72	0.97	-0.78	0.96
8	641543	0	588	0.61	0.20	0.04	0.61	0.15	0.00	0.00	0.24	-0.15	-0.15	0.24	-0.09	-0.18	0.09	1.65	1.06	1.73	1.08
9	641565	0	588	0.49	0.49	0.20	0.14	0.17	0.00	0.00	0.17	0.17	-0.02	-0.11	-0.11	0.66	0.09	3.82	1.11	3.51	1.14
10	616676	0	588	0.69	0.13	0.69	0.04	0.14	0.00	0.00	0.35	-0.06	0.35	-0.15	-0.31	-0.51	0.10	-0.94	0.96	-0.54	0.97
11	616683	0	588	0.74	0.07	0.12	0.07	0.74	0.00	0.00	0.29	-0.23	-0.07	-0.19	0.29	-0.87	0.10	0.58	1.03	1.03	1.08
12	616677	0	588	0.55	0.12	0.19	0.15	0.55	0.00	0.00	0.20	-0.13	-0.14	-0.01	0.20	0.22	0.09	2.50	1.07	2.29	1.09
13	616671	0	588	0.46	0.05	0.46	0.40	0.08	0.00	0.00	0.07	-0.04	0.07	-0.06	0.01	0.60	0.09	5.78	1.17	5.56	1.22
14	616678	0	588	0.62	0.20	0.11	0.62	0.06	0.00	0.00	0.29	-0.10	-0.19	0.29	-0.16	-0.71	0.10	5.02	1.26	3.46	1.26
15	616675	0	588	0.62	0.12	0.19	0.62	0.07	0.00	0.00	0.17	-0.05	-0.09	0.17	-0.13	-0.22	0.09	2.91	1.10	3.41	1.18
16	616670	0	588	0.47	0.47	0.10	0.39	0.04	0.00	0.00	0.22	0.22	-0.19	-0.02	-0.21	-0.44	0.09	9.76	1.44	7.79	1.51
17	616681	0	588	0.47	0.17	0.47	0.28	0.07	0.00	0.00	0.16	-0.10	0.16	0.03	-0.20	0.58	0.09	3.70	1.10	4.08	1.16
18	608943	0	588	0.55	0.18	0.17	0.10	0.55	0.00	0.00	0.30	-0.12	-0.14	-0.14	0.30	0.41	0.09	-0.17	1.00	0.33	1.01
19	608939	0	588	0.61	0.17	0.07	0.61	0.15	0.00	0.00	0.36	-0.23	-0.20	0.36	-0.10	0.25	0.09	-2.51	0.93	-2.47	0.91
20	616846	0	588	0.42	0.10	0.30	0.42	0.18	0.00	0.00	0.26	-0.18	-0.04	0.26	-0.13	0.68	0.09	0.27	1.01	0.51	1.02
21	608941	0	588	0.55	0.55	0.19	0.06	0.19	0.00	0.00	0.31	0.31	-0.18	-0.26	-0.02	0.62	0.09	0.80	1.02	0.85	1.03
22	640221	0	588	0.44	0.25	0.44	0.19	0.12	0.01	0.00	0.22	-0.13	0.22	-0.06	-0.07	0.82	0.09	2.03	1.06	2.79	1.12
23	640205	0	588	0.49	0.22	0.49	0.21	0.08	0.01	0.00	0.18	-0.08	0.18	-0.02	-0.15	0.76	0.09	4.02	1.12	3.87	1.16
24	608948	0	588	0.55	0.55	0.07	0.29	0.08	0.01	0.00	0.25	0.25	-0.16	-0.08	-0.13	0.46	0.09	1.46	1.04	1.81	1.07
25	608942	0	588	0.28	0.30	0.28	0.26	0.16	0.00	0.00	0.27	0.01	0.27	-0.21	-0.07	1.29	0.09	-1.86	0.93	-0.38	0.98
26	673032	0	588	0.70	0.70	0.07	0.20	0.02	0.00	0.00	0.28	0.28	-0.27	-0.09	-0.15	-0.63	0.10	0.54	1.02	1.42	1.09
27	673031	0	588	0.43	0.43	0.21	0.21	0.16	0.00	0.00	0.17	0.17	-0.07	-0.10	-0.04	0.58	0.09	2.88	1.08	2.92	1.11
28	683404	0	588	0.83	0.08	0.83	0.05	0.04	0.00	0.00	0.42	-0.24	0.42	-0.23	-0.21	-1.52	0.12	-0.33	0.97	-1.71	0.81
29	683419	0	588	0.56	0.18	0.56	0.18	0.08	0.00	0.00	0.38	-0.06	0.38	-0.22	-0.32	-0.10	0.09	-0.59	0.98	-0.72	0.97
30	673030	0	588	0.58	0.11	0.07	0.58	0.24	0.00	0.00	0.26	-0.10	-0.28	0.26	-0.06	0.04	0.09	0.95	1.03	0.41	1.02
31	673023	0	588	0.35	0.35	0.27	0.17	0.21	0.00	0.00	0.26	0.26	-0.09	-0.14	-0.07	1.20	0.09	0.53	1.02	1.46	1.08

Item Information			Classical													Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	673024	0	588	0.69	0.11	0.14	0.69	0.07	0.00	0.00	0.40	-0.21	-0.17	0.40	-0.23	-0.76	0.10	0.81	1.04	-0.16	0.99
33	673025	0	588	0.58	0.14	0.58	0.13	0.15	0.00	0.00	0.35	-0.07	0.35	-0.24	-0.18	0.50	0.09	-0.90	0.98	-0.23	0.99
34	673026	0	588	0.63	0.12	0.10	0.63	0.14	0.00	0.00	0.36	-0.23	-0.21	0.36	-0.10	-0.29	0.09	-0.86	0.97	-1.22	0.94

Table J–11. Algebra I Constructed-Response Item Statistics: Winter

Item Information			Classical														Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	640123	0	75936	0.26	0.28	0.45	0.19	0.05	0.01	0.02	0.61	-0.44	-0.02	0.28	0.35	0.25	1.45	0.00	-9.90	0.74	-9.90	0.76
2	640892	0	75936	0.16	0.47	0.36	0.08	0.03	0.01	0.05	0.56	-0.45	0.18	0.27	0.29	0.22	2.05	0.01	-9.90	0.93	-9.90	0.89
3	612749	0	75936	0.21	0.47	0.22	0.08	0.13	0.02	0.08	0.63	-0.53	0.10	0.14	0.43	0.32	1.42	0.00	-9.90	0.86	-9.90	0.79
4	640886	0	75936	0.21	0.37	0.38	0.14	0.05	0.01	0.05	0.65	-0.54	0.13	0.29	0.33	0.26	1.71	0.00	-9.90	0.81	-9.90	0.77
5	640898	0	75936	0.11	0.66	0.18	0.07	0.02	0.01	0.05	0.53	-0.44	0.17	0.28	0.26	0.26	1.81	0.01	-9.90	0.74	-9.90	0.65
6	622485	0	75936	0.32	0.37	0.16	0.14	0.23	0.03	0.06	0.51	-0.47	0.09	0.09	0.25	0.37	0.97	0.00	9.90	1.15	9.90	1.16

Table J–12. Biology Constructed-Response Item Statistics: Winter

Item Information			Classical												Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	629469	0	54236	0.34	0.24	0.41	0.28	0.01	0.06	0.39	-0.35	0.08	0.20	0.25	0.99	0.01	6.36	1.03	9.90	1.07
2	644134	0	54236	0.12	0.66	0.21	0.05	0.01	0.06	0.49	-0.45	0.27	0.31	0.21	1.68	0.01	-9.90	0.82	-9.90	0.75
3	641221	0	54236	0.20	0.42	0.35	0.09	0.02	0.11	0.47	-0.29	-0.04	0.37	0.35	1.45	0.01	-5.05	0.96	4.70	1.04
4	641299	0	54236	0.14	0.63	0.22	0.07	0.02	0.05	0.59	-0.52	0.25	0.36	0.31	1.38	0.01	-9.90	0.75	-9.90	0.67
5	641303	0	54236	0.13	0.60	0.23	0.05	0.02	0.09	0.51	-0.39	0.13	0.34	0.33	1.57	0.01	-9.90	0.82	-9.90	0.83
6	611059	0	54236	0.29	0.27	0.42	0.17	0.04	0.09	0.55	-0.39	-0.01	0.33	0.36	0.35	0.01	-9.90	0.86	-9.90	0.86

Table J–13. Literature Constructed-Response Item Statistics: Winter

Item Information			Classical												Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	643730	0	47149	0.37	0.19	0.45	0.25	0.05	0.05	0.62	-0.47	-0.09	0.41	0.32	1.23	0.01	-9.90	0.74	-9.90	0.74
2	614675	0	47149	0.42	0.17	0.32	0.41	0.04	0.05	0.63	-0.52	-0.14	0.51	0.19	0.88	0.01	-0.72	1.00	-4.73	0.97
3	614674	0	47149	0.38	0.14	0.38	0.27	0.07	0.14	0.67	-0.51	-0.09	0.40	0.38	0.87	0.01	-9.90	0.74	-9.90	0.73
4	614016	0	47149	0.39	0.17	0.40	0.29	0.07	0.07	0.67	-0.52	-0.10	0.41	0.36	1.69	0.01	6.71	1.04	4.09	1.03
5	644768	0	47149	0.48	0.07	0.38	0.43	0.06	0.06	0.63	-0.43	-0.28	0.41	0.32	0.26	0.01	-9.90	0.88	-9.90	0.88
6	644767	0	47149	0.43	0.09	0.37	0.38	0.06	0.10	0.62	-0.47	-0.18	0.41	0.31	0.77	0.01	-9.90	0.81	-9.90	0.80

Table J–14. Algebra I Constructed-Response Item Statistics: Spring

Item Information			Classical														DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	672731	0	184172	0.20	0.58	0.15	0.12	0.05	0.06	0.03	0.65	-0.61	0.14	0.31	0.26	0.37					1.82	0.01	9.90	1.11	0.29	1.00
2	641527	0	184172	0.21	0.47	0.23	0.15	0.08	0.02	0.05	0.63	-0.56	0.10	0.30	0.34	0.24					1.90	0.01	-9.90	0.93	-9.90	0.90
3	628222	0	184172	0.36	0.26	0.30	0.15	0.10	0.14	0.05	0.71	-0.50	-0.16	0.12	0.26	0.53					0.69	0.00	-9.90	0.93	-9.90	0.91
4	666526	0	184172	0.28	0.34	0.26	0.33	0.04	0.01	0.02	0.62	-0.52	-0.02	0.38	0.25	0.23					2.04	0.01	2.37	1.01	0.22	1.00
5	628223	0	184172	0.38	0.31	0.22	0.15	0.13	0.15	0.03	0.76	-0.59	-0.12	0.09	0.30	0.55					0.93	0.00	-9.90	0.78	-9.90	0.75
6	678819	0	184172	0.35	0.24	0.31	0.18	0.13	0.08	0.06	0.73	-0.54	-0.14	0.18	0.37	0.44					1.10	0.01	-9.90	0.86	-9.90	0.84
7	703741	1	2000	0.20	0.33	0.55	0.09	0.03	0.00	0.01	0.66	-0.58	0.27	0.34	0.27	0.08	A-	A-	B-	A-	3.06	0.04	-7.00	0.77	-7.83	0.76
8	703737	2	1987	0.34	0.16	0.42	0.22	0.13	0.03	0.03	0.68	-0.49	-0.20	0.27	0.41	0.24	A+	B-	B-	A-	1.55	0.03	-5.21	0.84	-6.06	0.82
9	703739	2	1987	0.32	0.38	0.19	0.18	0.19	0.04	0.02	0.74	-0.64	-0.04	0.24	0.48	0.27	A-	A-	A-	A-	1.71	0.03	-7.84	0.77	-8.04	0.70
10	700825	3	1987	0.17	0.47	0.38	0.08	0.03	0.01	0.03	0.62	-0.58	0.30	0.33	0.23	0.16	A-	A-	A-	A-	2.75	0.03	-5.05	0.82	-6.49	0.76
11	701635	3	1987	0.29	0.27	0.37	0.24	0.08	0.02	0.02	0.68	-0.56	-0.04	0.35	0.34	0.21	A+	A-	A-	A+	1.97	0.03	-4.83	0.85	-5.38	0.84
12	700826	4	1981	0.12	0.69	0.12	0.11	0.03	0.02	0.03	0.59	-0.58	0.23	0.37	0.26	0.21	A-	A-	A-	A-	2.66	0.03	-3.82	0.84	-5.60	0.59
13	696812	4	1981	0.43	0.21	0.29	0.16	0.18	0.15	0.02	0.70	-0.53	-0.20	0.14	0.28	0.44	A-	A-	A-	A-	0.97	0.02	-0.84	0.97	-1.00	0.97
14	700873	5	1980	0.13	0.62	0.24	0.10	0.03	0.00	0.01	0.55	-0.54	0.30	0.31	0.21	0.11	A-	A-	A-	A-	3.12	0.04	0.50	1.02	-2.51	0.87
15	704031	5	1980	0.45	0.17	0.24	0.30	0.19	0.09	0.01	0.77	-0.53	-0.29	0.12	0.38	0.42	A+	B-	A-	A-	1.02	0.03	-9.41	0.74	-9.42	0.74
16	704030	6	1987	0.25	0.46	0.14	0.30	0.05	0.03	0.02	0.67	-0.60	-0.03	0.47	0.24	0.25	A+	A-	A+	A+	2.03	0.03	-3.54	0.89	-3.87	0.83
17	702519	6	1987	0.28	0.32	0.34	0.19	0.11	0.02	0.01	0.63	-0.53	0.01	0.27	0.36	0.21	A+	A-	A+	A+	1.95	0.03	-1.17	0.96	-2.07	0.93
18	696811	7	2005	0.10	0.73	0.15	0.06	0.03	0.01	0.02	0.55	-0.54	0.28	0.29	0.28	0.16	A+	A-	A-	A-	2.93	0.04	-4.79	0.78	-3.84	0.72
19	702514	7	2005	0.19	0.49	0.30	0.13	0.05	0.01	0.01	0.54	-0.47	0.14	0.29	0.27	0.16	A-	A-	A-	A-	2.50	0.03	1.14	1.04	0.16	1.01
20	703742	8	1980	0.24	0.35	0.40	0.13	0.07	0.02	0.03	0.66	-0.55	0.06	0.35	0.33	0.22	A+	A-	A-	A-	2.14	0.03	-5.33	0.82	-6.44	0.79
21	632538	8	1980	0.28	0.43	0.19	0.16	0.13	0.06	0.03	0.67	-0.63	0.07	0.25	0.35	0.32	A+	A-	A+	A-	1.77	0.03	-1.49	0.95	-3.36	0.85
22	704111	9	1997	0.32	0.20	0.47	0.16	0.11	0.04	0.02	0.63	-0.42	-0.18	0.27	0.38	0.25	A+	A-	A-	A-	1.58	0.03	-3.46	0.89	-4.12	0.87
23	702516	9	1997	0.21	0.59	0.09	0.17	0.13	0.00	0.02	0.63	-0.60	0.11	0.26	0.47	0.09	A+	A-	A-	A-	2.68	0.03	-3.72	0.88	-3.58	0.77
24	701641	10	1995	0.37	0.27	0.25	0.19	0.23	0.04	0.02	0.75	-0.64	-0.09	0.20	0.49	0.25	A-	A-	B-	A-	1.50	0.03	-8.04	0.77	-8.18	0.74
25	678825	10	1995	0.27	0.29	0.39	0.22	0.07	0.01	0.01	0.60	-0.52	0.03	0.33	0.27	0.17	A-	A-	B-	A-	2.13	0.03	-1.39	0.96	-1.83	0.94
26	703743	11	1999	0.66	0.08	0.09	0.18	0.38	0.26	0.01	0.69	-0.45	-0.32	-0.16	0.17	0.46	A-	C-	B-	A-	0.10	0.03	-6.20	0.81	-5.16	0.83
27	701638	11	1999	0.41	0.08	0.45	0.24	0.14	0.08	0.01	0.63	-0.30	-0.36	0.12	0.32	0.39	A-	A-	A-	A-	0.97	0.03	-3.30	0.90	-3.33	0.89
28	704028	12	1983	0.17	0.43	0.44	0.09	0.01	0.00	0.03	0.65	-0.59	0.31	0.41	0.16	0.07	A+	A-	A-	A-	3.30	0.04	-6.63	0.80	-8.04	0.76
29	701636	12	1983	0.23	0.43	0.30	0.16	0.06	0.03	0.02	0.69	-0.63	0.13	0.35	0.31	0.28	A-	A-	A-	A+	2.09	0.03	-7.28	0.77	-7.38	0.73
30	701632	13	1995	0.39	0.35	0.09	0.23	0.22	0.09	0.02	0.70	-0.65	-0.01	0.13	0.40	0.34	A-	A-	A-	A-	1.32	0.02	-2.49	0.92	-2.18	0.91
31	701637	13	1995	0.42	0.15	0.28	0.30	0.25	0.02	0.01	0.65	-0.43	-0.28	0.14	0.45	0.21	B-	A-	A-	B-	1.52	0.03	-2.64	0.92	-2.45	0.93

Item Information			Classical														DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	700872	14	2010	0.29	0.31	0.36	0.14	0.11	0.05	0.03	0.68	-0.59	0.00	0.32	0.34	0.29	A-	A-	A-	A+	1.75	0.03	-5.17	0.84	-6.08	0.80
33	704033	14	2010	0.39	0.34	0.19	0.13	0.20	0.13	0.01	0.74	-0.67	-0.02	0.13	0.32	0.46	A-	A-	A-	A-	1.23	0.02	-5.60	0.83	-5.72	0.78
34	700876	15	1992	0.14	0.52	0.37	0.07	0.02	0.00	0.03	0.61	-0.57	0.36	0.33	0.21	0.09	A-	A-	A-	A+	3.58	0.04	-5.02	0.83	-6.65	0.76
35	703738	16	1994	0.11	0.68	0.14	0.11	0.02	0.00	0.05	0.54	-0.53	0.25	0.36	0.22	0.08	A+	A+	A-	A-	3.25	0.04	-1.96	0.92	-4.52	0.68
36	703735	16	1994	0.26	0.37	0.29	0.24	0.08	0.01	0.01	0.69	-0.65	0.11	0.39	0.32	0.14	A-	A-	A-	A-	2.33	0.03	-6.08	0.82	-6.85	0.77
37	701639	17	1995	0.22	0.40	0.35	0.19	0.04	0.01	0.02	0.57	-0.55	0.17	0.38	0.18	0.11	A+	A+	A-	A-	2.36	0.03	-0.01	1.00	-1.06	0.96
38	703744	17	1995	0.26	0.41	0.29	0.17	0.08	0.04	0.02	0.59	-0.51	0.04	0.26	0.32	0.25	A+	A-	A-	A+	1.90	0.03	-0.53	0.98	0.37	1.01
39	704029	18	1993	0.55	0.16	0.13	0.19	0.26	0.23	0.02	0.79	-0.59	-0.25	-0.07	0.26	0.55	A+	B-	B-	A-	0.52	0.02	-9.90	0.70	-8.81	0.71
40	704112	18	1993	0.38	0.26	0.29	0.21	0.13	0.10	0.01	0.78	-0.61	-0.15	0.21	0.36	0.44	A-	A-	A-	A-	1.28	0.03	-9.90	0.70	-9.90	0.69
41	704110	19	1989	0.11	0.64	0.24	0.08	0.01	0.00	0.03	0.55	-0.52	0.28	0.36	0.19	0.09	A+	A-	A+	A-	3.26	0.04	-3.87	0.85	-3.97	0.79
42	704032	19	1989	0.18	0.47	0.36	0.10	0.05	0.01	0.02	0.63	-0.58	0.25	0.32	0.31	0.14	A+	A-	A-	A-	2.55	0.03	-4.08	0.85	-5.89	0.78
43	700805	20	1994	0.17	0.51	0.33	0.10	0.03	0.01	0.02	0.53	-0.45	0.15	0.32	0.24	0.17	A-	A-	A-	A-	2.66	0.03	-0.80	0.97	-0.81	0.96
44	702518	20	1994	0.09	0.74	0.15	0.06	0.02	0.01	0.02	0.55	-0.55	0.33	0.30	0.23	0.17	A-	A+	A+	A-	2.99	0.04	-4.30	0.80	-5.63	0.61
45	700850	21	2002	0.24	0.26	0.55	0.09	0.05	0.01	0.03	0.63	-0.50	0.07	0.34	0.32	0.19	A-	A-	A-	A+	2.17	0.03	-6.13	0.78	-7.84	0.74
46	673354	21	2002	0.43	0.22	0.31	0.09	0.25	0.12	0.01	0.63	-0.45	-0.20	0.10	0.30	0.40	A-	A-	A-	A-	1.05	0.02	3.34	1.11	2.66	1.10
47	678817	22	1996	0.09	0.73	0.16	0.04	0.03	0.01	0.04	0.43	-0.39	0.20	0.20	0.23	0.18	A-	A-	A-	B-	2.97	0.04	0.81	1.04	0.07	1.00
48	702523	22	1996	0.31	0.32	0.29	0.18	0.17	0.02	0.02	0.72	-0.61	-0.03	0.24	0.48	0.22	A+	A-	A-	A+	1.91	0.03	-8.16	0.77	-8.31	0.73
49	703613	23	1999	0.08	0.68	0.27	0.02	0.00	0.00	0.03	0.57	-0.56	0.49	0.23	0.07	0.05	A+	A-	A-	A+	4.47	0.05	-5.68	0.80	-7.33	0.68
50	701634	23	1999	0.56	0.06	0.21	0.31	0.26	0.16	0.02	0.65	-0.37	-0.30	-0.14	0.27	0.44	A+	B-	A-	B-	0.39	0.03	-1.72	0.95	-1.25	0.96
51	700827	24	1989	0.14	0.61	0.24	0.08	0.02	0.02	0.02	0.54	-0.49	0.22	0.27	0.22	0.25	A-	A-	A-	A-	2.53	0.03	-2.37	0.90	-2.42	0.87
52	702515	24	1989	0.29	0.32	0.30	0.28	0.09	0.01	0.01	0.68	-0.59	0.03	0.33	0.37	0.16	C-	A-	A-	A-	2.21	0.03	-5.74	0.84	-6.09	0.81

Table J–15. Biology Constructed-Response Item Statistics: Spring

Item Information			Classical												DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	678999	0	154585	0.51	0.17	0.27	0.32	0.21	0.03	0.52	-0.40	-0.16	0.16	0.38					0.39	0.01	9.90	1.20	9.90	1.30
2	679989	0	154585	0.37	0.33	0.27	0.21	0.13	0.05	0.60	-0.50	-0.03	0.29	0.41					0.94	0.01	9.90	1.19	9.90	1.17
3	678998	0	154585	0.38	0.23	0.34	0.27	0.09	0.07	0.56	-0.46	-0.03	0.29	0.34					1.11	0.01	9.90	1.13	9.90	1.12
4	684325	0	154585	0.40	0.21	0.40	0.24	0.10	0.04	0.61	-0.51	-0.02	0.29	0.37					0.90	0.01	9.90	1.07	9.90	1.07
5	682671	0	154585	0.24	0.48	0.27	0.16	0.04	0.05	0.64	-0.60	0.22	0.42	0.25					1.67	0.01	-9.90	0.90	-9.90	0.84
6	644202	0	154585	0.48	0.15	0.32	0.28	0.19	0.05	0.66	-0.46	-0.22	0.22	0.49					0.38	0.01	-9.90	0.94	-9.69	0.94
7	703246	1	4828	0.09	0.24	0.09	0.05	0.02	0.02	0.36	-0.34	0.17	0.21	0.21	A-	A+	A+	A-	2.14	0.03	-4.10	0.86	-4.96	0.75
8	703007	1	4828	0.13	0.18	0.11	0.08	0.04	0.01	0.37	-0.33	0.10	0.22	0.24	A-	B-	A-	A+	1.65	0.03	-4.80	0.86	-6.01	0.77
9	684968	2	4787	0.10	0.21	0.10	0.06	0.02	0.02	0.36	-0.35	0.16	0.24	0.19	A+	A-	A-	A-	2.00	0.03	-5.27	0.83	-6.74	0.72
10	703002	2	4787	0.27	0.06	0.06	0.11	0.17	0.01	0.27	-0.18	-0.13	0.03	0.29	A+	C-	A-	A+	0.08	0.03	-1.12	0.96	0.81	1.03
11	644151	3	4730	0.18	0.09	0.13	0.12	0.06	0.01	0.31	-0.24	-0.02	0.20	0.23	A-	A-	C-	A+	1.02	0.03	-3.58	0.90	-3.76	0.89
12	678420	3	4730	0.25	0.06	0.08	0.14	0.13	0.01	0.28	-0.19	-0.10	0.06	0.30	A-	C-	B-	A+	0.27	0.03	-3.41	0.90	-3.01	0.90
13	702635	4	4752	0.14	0.13	0.15	0.08	0.04	0.02	0.32	-0.23	0.00	0.23	0.22	A+	A-	A-	A+	1.52	0.03	-1.30	0.96	-2.06	0.93
14	679990	4	4752	0.19	0.10	0.10	0.12	0.08	0.01	0.31	-0.24	-0.05	0.18	0.25	A+	B-	A-	A+	0.93	0.03	0.69	1.02	0.14	1.00
15	702636	5	4756	0.12	0.18	0.11	0.10	0.02	0.01	0.35	-0.32	0.12	0.27	0.15	A+	A-	A-	A-	1.96	0.03	-7.23	0.80	-7.72	0.74
16	700902	5	4756	0.08	0.24	0.09	0.04	0.03	0.02	0.34	-0.33	0.18	0.19	0.20	A+	A-	A-	A+	2.04	0.03	-5.35	0.81	-6.75	0.67
17	703364	6	4837	0.14	0.10	0.22	0.08	0.01	0.01	0.23	-0.18	0.06	0.17	0.14	A-	A-	A+	A+	1.86	0.04	2.65	1.08	2.58	1.08
18	702739	6	4837	0.13	0.12	0.18	0.10	0.01	0.01	0.32	-0.27	0.08	0.28	0.08	A+	A-	A-	A-	2.21	0.04	-4.78	0.87	-4.97	0.86
19	703365	7	4788	0.20	0.07	0.13	0.14	0.07	0.01	0.28	-0.18	-0.09	0.16	0.25	A+	A-	A-	A+	0.77	0.03	-1.12	0.97	-1.82	0.95
20	702742	7	4788	0.14	0.14	0.14	0.09	0.03	0.01	0.32	-0.27	0.06	0.21	0.21	A+	A-	A-	A+	1.56	0.03	-2.03	0.94	-3.07	0.90
21	703496	8	4769	0.21	0.09	0.11	0.09	0.11	0.01	0.34	-0.26	-0.06	0.13	0.32	A+	A-	A-	A-	0.71	0.03	-5.44	0.85	-5.44	0.82
22	702744	8	4769	0.26	0.06	0.08	0.14	0.14	0.01	0.29	-0.20	-0.12	0.08	0.29	A+	B-	B-	A+	0.22	0.03	-2.99	0.91	-0.66	0.98
23	703499	9	4730	0.17	0.16	0.09	0.07	0.09	0.01	0.38	-0.35	0.07	0.17	0.30	A-	A-	A-	A-	1.14	0.03	-4.92	0.86	-5.46	0.78
24	702745	9	4730	0.19	0.12	0.10	0.11	0.09	0.01	0.27	-0.22	-0.01	0.13	0.23	A+	B-	A-	C+	0.90	0.03	8.27	1.26	8.00	1.29
25	703528	10	4742	0.16	0.14	0.11	0.10	0.05	0.01	0.32	-0.26	0.01	0.20	0.24	A+	B-	A-	A-	1.29	0.03	1.99	1.06	1.67	1.06
26	702998	10	4742	0.23	0.01	0.20	0.13	0.08	0.01	0.26	-0.09	-0.22	0.17	0.28	A+	A-	B-	A+	0.16	0.03	-4.39	0.87	-4.41	0.86
27	703529	11	4797	0.22	0.07	0.10	0.13	0.11	0.01	0.32	-0.23	-0.09	0.12	0.30	A-	A-	A-	A+	0.55	0.03	-3.76	0.89	-3.37	0.89
28	702999	11	4797	0.17	0.13	0.11	0.08	0.08	0.02	0.38	-0.32	0.04	0.18	0.31	A+	A-	A-	A+	1.10	0.03	-7.07	0.80	-8.18	0.73
29	703531	12	4751	0.07	0.25	0.10	0.05	0.01	0.02	0.25	-0.24	0.15	0.17	0.10	A+	A-	A-	A-	2.59	0.04	4.23	1.16	3.21	1.18
30	703000	12	4751	0.17	0.09	0.17	0.12	0.03	0.01	0.29	-0.22	-0.01	0.23	0.17	A-	C-	C-	A+	1.27	0.03	-3.01	0.91	-3.49	0.90
31	703532	13	4804	0.32	0.04	0.03	0.11	0.23	0.01	0.18	-0.14	-0.05	-0.02	0.20	A-	C-	C-	A+	-0.38	0.03	6.43	1.24	9.07	1.52

Item Information			Classical												DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	703001	13	4804	0.29	0.04	0.06	0.11	0.20	0.01	0.25	-0.16	-0.16	0.01	0.28	A+	A-	B-	A+	-0.19	0.03	-3.32	0.90	0.23	1.01
33	703536	14	4806	0.14	0.15	0.12	0.09	0.04	0.02	0.35	-0.29	0.04	0.23	0.24	A-	A-	A-	A+	1.46	0.03	-2.37	0.93	-3.41	0.88
34	703004	14	4806	0.11	0.21	0.09	0.06	0.03	0.01	0.35	-0.32	0.11	0.23	0.21	A+	A-	B-	B+	1.84	0.03	-2.06	0.93	-3.13	0.86
35	703535	15	4776	0.15	0.16	0.11	0.08	0.06	0.01	0.37	-0.31	0.05	0.20	0.28	A+	A-	A-	A+	1.39	0.03	-3.78	0.89	-4.60	0.84
36	703006	15	4776	0.13	0.20	0.06	0.04	0.08	0.04	0.37	-0.35	0.10	0.14	0.31	A+	A-	A+	A-	1.47	0.03	-0.06	1.00	-0.90	0.94
37	703534	16	4785	0.17	0.14	0.10	0.11	0.06	0.01	0.37	-0.32	0.01	0.25	0.26	A-	A-	A-	A-	1.17	0.03	-8.64	0.77	-8.22	0.74
38	703481	16	4785	0.10	0.21	0.11	0.05	0.02	0.02	0.31	-0.27	0.10	0.20	0.19	A+	A-	A-	A-	2.02	0.03	0.07	1.00	0.02	1.00
39	703539	17	4773	0.15	0.11	0.17	0.08	0.04	0.01	0.33	-0.27	0.06	0.21	0.22	A-	A-	A+	A-	1.38	0.03	-4.99	0.86	-5.76	0.83
40	703605	17	4773	0.17	0.10	0.15	0.12	0.04	0.01	0.29	-0.22	0.00	0.19	0.20	A+	B-	A-	A+	1.27	0.03	-0.15	1.00	-0.78	0.98
41	703537	18	4754	0.18	0.14	0.08	0.10	0.08	0.01	0.37	-0.33	0.03	0.21	0.27	A-	A-	A-	A-	1.07	0.03	-5.22	0.85	-5.25	0.81
42	703666	18	4754	0.11	0.21	0.09	0.08	0.02	0.02	0.35	-0.32	0.11	0.25	0.18	A+	A-	A-	A+	2.01	0.03	-1.92	0.94	-3.28	0.85
43	703540	19	4823	0.08	0.27	0.06	0.04	0.04	0.02	0.32	-0.33	0.17	0.18	0.21	A+	A+	A+	A-	2.04	0.03	1.38	1.06	-1.91	0.86
44	703878	19	4823	0.20	0.10	0.09	0.13	0.08	0.01	0.26	-0.20	-0.03	0.08	0.26	A+	C-	B-	B+	0.82	0.03	6.91	1.22	6.97	1.25
45	703541	20	4775	0.12	0.21	0.06	0.07	0.05	0.02	0.34	-0.33	0.13	0.17	0.24	A+	A+	A-	A-	1.56	0.03	0.02	1.00	-1.92	0.89
46	704768	20	4775	0.10	0.23	0.08	0.06	0.03	0.02	0.35	-0.32	0.12	0.20	0.23	B+	A-	A-	A+	1.89	0.03	-4.14	0.86	-4.83	0.75
47	703703	21	4826	0.13	0.14	0.16	0.09	0.02	0.01	0.28	-0.20	0.00	0.25	0.16	A-	A-	A+	A+	1.91	0.03	1.16	1.04	0.78	1.02
48	704769	21	4826	0.27	0.04	0.05	0.19	0.12	0.01	0.26	-0.18	-0.13	0.03	0.31	A+	C-	B-	A+	0.07	0.03	-5.84	0.83	-4.66	0.85
49	703705	22	4776	0.13	0.18	0.11	0.08	0.04	0.01	0.35	-0.31	0.07	0.23	0.23	A+	A-	A-	A+	1.65	0.03	-4.49	0.86	-4.36	0.83
50	704770	22	4776	0.17	0.14	0.09	0.09	0.09	0.02	0.35	-0.30	0.02	0.15	0.30	A-	B-	A-	A+	1.08	0.03	-2.23	0.93	-2.73	0.89
51	703877	23	4809	0.10	0.19	0.16	0.06	0.02	0.00	0.23	-0.17	0.05	0.16	0.15	A-	A-	A-	A-	2.12	0.03	5.97	1.21	6.80	1.26
52	705264	23	4809	0.24	0.05	0.13	0.13	0.11	0.00	0.27	-0.14	-0.17	0.09	0.30	A-	B-	B-	A+	0.26	0.03	-3.84	0.89	-3.66	0.89
53	705274	24	4759	0.14	0.16	0.12	0.08	0.05	0.01	0.36	-0.31	0.08	0.20	0.25	A+	A-	A-	A+	1.44	0.03	-6.10	0.82	-6.37	0.78
54	705265	24	4759	0.17	0.13	0.12	0.08	0.07	0.01	0.29	-0.23	0.00	0.14	0.25	A-	A-	A-	A+	1.07	0.03	3.86	1.12	3.11	1.11

Table J–16. Literature Constructed-Response Item Statistics: Spring

Item Information			Classical												DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	613661	0	140252	0.45	0.18	0.31	0.39	0.08	0.04	0.64	-0.54	-0.11	0.40	0.29					1.38	0.01	-9.90	0.77	-9.90	0.79
2	643181	0	140252	0.42	0.23	0.28	0.32	0.12	0.05	0.74	-0.66	-0.04	0.43	0.37					0.76	0.01	9.90	1.07	5.97	1.04
3	643182	0	140252	0.52	0.10	0.26	0.41	0.16	0.06	0.74	-0.60	-0.23	0.35	0.41					0.68	0.01	-9.90	0.67	-9.90	0.67
4	683334	0	140252	0.53	0.08	0.29	0.47	0.12	0.04	0.71	-0.52	-0.32	0.41	0.35					1.01	0.01	-9.90	0.70	-9.90	0.70
5	614551	0	140252	0.58	0.09	0.22	0.42	0.22	0.04	0.72	-0.56	-0.28	0.25	0.45					0.82	0.01	-9.90	0.76	-9.90	0.75
6	614552	0	140252	0.49	0.12	0.30	0.36	0.15	0.06	0.70	-0.57	-0.20	0.34	0.40					0.73	0.01	-9.90	0.84	-9.90	0.84
7	703951	1	1957	0.58	0.07	0.27	0.49	0.16	0.01	0.65	-0.46	-0.34	0.28	0.36	B+	A+	A+	B-	0.71	0.03	-6.47	0.81	-6.49	0.81
8	705116	1	1957	0.47	0.09	0.44	0.37	0.08	0.02	0.68	-0.51	-0.26	0.43	0.31	B+	A-	A-	A-	1.33	0.03	-9.83	0.73	-9.83	0.73
9	703952	2	1990	0.57	0.07	0.24	0.55	0.13	0.01	0.68	-0.52	-0.33	0.36	0.32	C+	B+	A-	B-	0.86	0.03	-7.72	0.77	-7.49	0.77
10	705115	2	1990	0.42	0.15	0.45	0.28	0.08	0.03	0.71	-0.64	-0.05	0.41	0.31	B+	B-	A-	A+	1.58	0.03	-9.90	0.68	-9.90	0.68
11	704330	3	1995	0.44	0.13	0.43	0.35	0.07	0.03	0.65	-0.53	-0.15	0.44	0.26	B+	A+	A-	A-	1.57	0.03	-7.41	0.79	-7.49	0.79
12	704771	3	1995	0.54	0.10	0.28	0.44	0.16	0.02	0.70	-0.60	-0.20	0.30	0.39	B+	A-	A-	A+	0.99	0.03	-8.97	0.76	-8.94	0.75
13	704331	4	1989	0.47	0.14	0.38	0.37	0.09	0.02	0.70	-0.57	-0.19	0.42	0.33	B+	A-	A-	A-	1.36	0.03	-9.90	0.71	-9.90	0.71
14	704772	4	1989	0.50	0.11	0.31	0.45	0.10	0.03	0.71	-0.58	-0.21	0.41	0.34	B+	A+	A-	A-	1.23	0.03	-9.90	0.72	-9.90	0.72
15	704396	5	1998	0.56	0.11	0.24	0.42	0.20	0.03	0.73	-0.59	-0.28	0.33	0.41	B+	B-	A-	A-	0.91	0.03	-9.90	0.70	-9.90	0.69
16	705114	5	1998	0.45	0.18	0.33	0.36	0.10	0.02	0.62	-0.53	-0.10	0.36	0.30	C+	A-	A-	A+	1.49	0.03	-4.54	0.87	-4.74	0.86
17	704397	6	1996	0.45	0.19	0.31	0.37	0.10	0.03	0.69	-0.59	-0.12	0.42	0.32	B+	A-	A-	A-	1.49	0.03	-9.90	0.73	-9.90	0.72
18	705113	6	1996	0.46	0.21	0.26	0.40	0.11	0.03	0.66	-0.57	-0.11	0.40	0.32	B+	A+	A+	A-	1.47	0.03	-7.05	0.81	-7.02	0.79
19	704326	7	1992	0.67	0.02	0.20	0.49	0.28	0.01	0.69	-0.40	-0.47	0.11	0.46	C+	A+	A-	A-	0.06	0.03	-9.25	0.74	-9.41	0.74
20	705145	7	1992	0.60	0.05	0.27	0.46	0.20	0.02	0.67	-0.48	-0.36	0.26	0.39	B+	A-	A-	A-	0.60	0.03	-7.36	0.80	-7.37	0.80
21	704327	8	2000	0.57	0.04	0.32	0.49	0.14	0.02	0.69	-0.45	-0.42	0.36	0.36	B+	A+	A+	A-	0.70	0.04	-8.92	0.75	-8.78	0.76
22	705144	8	2000	0.60	0.06	0.23	0.50	0.19	0.02	0.71	-0.50	-0.35	0.23	0.43	B+	A+	A+	A-	0.64	0.03	-9.20	0.74	-9.28	0.74
23	703982	9	1995	0.40	0.22	0.36	0.30	0.09	0.03	0.67	-0.59	-0.03	0.40	0.32	A+	A-	A-	A+	1.67	0.03	-9.85	0.74	-9.90	0.73
24	698382	9	1995	0.51	0.06	0.40	0.42	0.09	0.03	0.68	-0.51	-0.29	0.39	0.33	B+	A-	A-	A+	1.08	0.03	-9.63	0.74	-9.53	0.74
25	703981	10	2002	0.46	0.17	0.32	0.40	0.09	0.03	0.72	-0.62	-0.13	0.44	0.32	A+	A-	A+	A-	1.49	0.03	-9.90	0.69	-9.90	0.69
26	698357	10	2002	0.50	0.09	0.38	0.38	0.12	0.03	0.69	-0.56	-0.23	0.39	0.33	C+	A-	A+	A-	1.16	0.03	-9.90	0.74	-9.90	0.74
27	703940	11	1999	0.47	0.16	0.35	0.35	0.13	0.02	0.67	-0.56	-0.15	0.35	0.36	A+	A-	A-	B-	1.30	0.03	-8.54	0.77	-8.95	0.76
28	705103	11	1999	0.60	0.07	0.21	0.51	0.19	0.02	0.74	-0.56	-0.36	0.29	0.41	B+	A+	A-	A-	0.66	0.03	-9.90	0.68	-9.90	0.69
29	703937	12	1991	0.54	0.13	0.22	0.47	0.15	0.03	0.76	-0.66	-0.20	0.38	0.38	B+	A-	A-	A-	1.02	0.03	-9.90	0.64	-9.90	0.64
30	705093	12	1991	0.57	0.05	0.27	0.54	0.12	0.02	0.69	-0.54	-0.33	0.36	0.33	C+	A+	A-	A-	0.75	0.04	-9.45	0.73	-9.36	0.73
31	704322	13	1991	0.61	0.05	0.25	0.46	0.22	0.02	0.70	-0.48	-0.40	0.23	0.42	C+	A+	A+	A-	0.55	0.03	-9.90	0.73	-9.90	0.73

Item Information			Classical												DIF				Rasch		Infit		Outfit	
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*	MS	*z*	MS
32	704672	13	1991	0.54	0.06	0.32	0.51	0.09	0.02	0.68	-0.50	-0.34	0.41	0.32	C+	A+	A-	A-	1.05	0.04	-9.51	0.73	-9.40	0.73
33	704320	14	1995	0.57	0.10	0.21	0.50	0.17	0.02	0.72	-0.56	-0.30	0.29	0.41	B+	A+	A+	A-	0.87	0.03	-9.90	0.69	-9.90	0.69
34	704674	14	1995	0.44	0.16	0.38	0.35	0.08	0.03	0.69	-0.58	-0.14	0.45	0.30	B+	A-	A-	A-	1.55	0.03	-9.90	0.72	-9.90	0.72
35	703918	15	1992	0.47	0.13	0.38	0.39	0.08	0.02	0.63	-0.48	-0.23	0.42	0.28	A+	A-	A-	A-	1.44	0.03	-6.55	0.82	-6.76	0.81
36	704763	15	1992	0.52	0.13	0.25	0.48	0.12	0.02	0.70	-0.59	-0.21	0.39	0.33	A+	B-	A-	A-	1.21	0.03	-9.20	0.75	-8.92	0.74
37	703919	16	2001	0.52	0.09	0.32	0.45	0.10	0.02	0.64	-0.46	-0.29	0.37	0.32	B+	A+	A-	B-	1.18	0.03	-5.81	0.83	-5.94	0.83
38	704764	16	2001	0.58	0.06	0.22	0.59	0.12	0.02	0.70	-0.53	-0.35	0.37	0.32	A+	B-	B-	A-	0.82	0.04	-9.11	0.73	-8.83	0.72
39	701069	17	1993	0.37	0.22	0.45	0.26	0.05	0.03	0.61	-0.57	0.06	0.38	0.23	A+	B-	B-	A-	1.90	0.03	-6.98	0.81	-7.51	0.79
40	704755	17	1993	0.47	0.15	0.29	0.45	0.07	0.03	0.69	-0.59	-0.16	0.46	0.29	B+	A+	A+	A-	1.46	0.03	-9.45	0.74	-9.13	0.74
41	703916	18	1975	0.45	0.20	0.31	0.36	0.11	0.03	0.70	-0.62	-0.08	0.42	0.32	B+	A-	A-	A-	1.49	0.03	-9.90	0.72	-9.90	0.71
42	704757	18	1975	0.47	0.13	0.33	0.44	0.07	0.03	0.71	-0.61	-0.16	0.48	0.26	A+	A-	B-	A+	1.48	0.03	-9.90	0.70	-9.90	0.70
43	703935	19	1975	0.44	0.09	0.48	0.35	0.05	0.03	0.66	-0.56	-0.15	0.43	0.25	A+	A-	A-	A-	1.53	0.04	-8.82	0.75	-8.88	0.75
44	705161	19	1975	0.41	0.08	0.57	0.27	0.04	0.03	0.62	-0.47	-0.18	0.43	0.23	A+	A-	A-	A+	1.61	0.04	-8.16	0.76	-8.25	0.75
45	703936	20	1982	0.51	0.04	0.44	0.38	0.11	0.02	0.67	-0.45	-0.37	0.39	0.36	A+	A-	A-	A-	0.93	0.03	-9.75	0.74	-9.69	0.73
46	705174	20	1982	0.49	0.08	0.42	0.38	0.09	0.03	0.69	-0.50	-0.30	0.43	0.34	A-	B-	A-	A-	1.20	0.03	-9.90	0.71	-9.90	0.72
47	698361	21	1995	0.51	0.08	0.37	0.45	0.09	0.01	0.65	-0.45	-0.33	0.41	0.30	C+	A+	A+	B-	1.12	0.03	-7.49	0.79	-7.48	0.79
48	683634	21	1995	0.51	0.11	0.31	0.42	0.13	0.03	0.72	-0.59	-0.22	0.39	0.36	A+	A-	A-	A-	1.19	0.03	-9.90	0.70	-9.90	0.70
49	698373	22	1996	0.41	0.21	0.32	0.37	0.06	0.04	0.68	-0.62	-0.02	0.46	0.25	B+	A+	A-	A-	1.82	0.03	-9.90	0.73	-9.90	0.72
50	683631	22	1996	0.50	0.14	0.28	0.45	0.11	0.03	0.71	-0.58	-0.20	0.40	0.34	A+	A+	A-	B-	1.27	0.03	-9.90	0.72	-9.90	0.72
51	703955	23	1990	0.51	0.15	0.23	0.46	0.13	0.03	0.73	-0.65	-0.17	0.41	0.34	A+	B-	A+	A-	1.23	0.03	-9.90	0.70	-9.90	0.69
52	705132	23	1990	0.51	0.08	0.37	0.40	0.12	0.03	0.66	-0.52	-0.26	0.39	0.32	B+	A-	A-	A-	1.13	0.03	-7.57	0.79	-7.69	0.79
53	703954	24	1983	0.51	0.12	0.29	0.48	0.10	0.02	0.69	-0.56	-0.25	0.43	0.30	B+	A-	A-	B-	1.18	0.03	-9.90	0.72	-9.90	0.71
54	705133	24	1983	0.53	0.06	0.39	0.36	0.17	0.03	0.70	-0.51	-0.33	0.30	0.42	A+	B-	B-	A-	0.82	0.03	-9.90	0.70	-9.90	0.70

Table J–17. Algebra I Constructed-Response Item Statistics: Summer

Item Information			Classical														Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	640376	0	1846	0.37	0.23	0.32	0.24	0.11	0.09	0.01	0.50	-0.34	-0.14	0.12	0.21	0.32	0.81	0.02	0.63	1.02	0.89	1.03
2	681303	0	1846	0.23	0.23	0.63	0.13	0.00	0.00	0.01	0.48	-0.34	0.03	0.33	0.17	0.18	2.43	0.04	-7.86	0.77	-8.22	0.77
3	674404	0	1846	0.31	0.26	0.42	0.16	0.09	0.06	0.02	0.59	-0.41	-0.08	0.17	0.27	0.37	1.13	0.03	-1.41	0.95	-2.21	0.92
4	633313	0	1846	0.20	0.50	0.29	0.11	0.05	0.03	0.02	0.55	-0.42	0.06	0.21	0.24	0.36	1.56	0.03	-0.39	0.98	-1.77	0.92
5	666557	0	1846	0.15	0.52	0.38	0.05	0.02	0.01	0.02	0.53	-0.42	0.20	0.26	0.25	0.23	1.65	0.03	-8.71	0.69	-9.90	0.64
6	672275	0	1846	0.27	0.27	0.42	0.19	0.07	0.02	0.03	0.52	-0.35	-0.06	0.18	0.28	0.32	1.19	0.03	-3.73	0.88	-4.04	0.87

Table J–18. Biology Constructed-Response Item Statistics: Summer

Item Information			Classical												Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	677887	0	1116	0.35	0.31	0.38	0.22	0.07	0.01	0.39	-0.30	-0.01	0.19	0.26	0.91	0.04	1.14	1.04	0.96	1.04
2	678932	0	1116	0.50	0.19	0.25	0.39	0.16	0.01	0.48	-0.30	-0.20	0.14	0.39	0.50	0.03	-0.86	0.97	-0.84	0.97
3	682984	0	1116	0.36	0.31	0.28	0.38	0.02	0.02	0.20	-0.13	-0.06	0.14	0.16	1.42	0.04	8.82	1.33	9.45	1.41
4	682669	0	1116	0.48	0.10	0.46	0.32	0.11	0.01	0.44	-0.31	-0.16	0.19	0.28	0.14	0.04	-1.59	0.94	-1.41	0.95
5	677890	0	1116	0.28	0.27	0.58	0.10	0.02	0.03	0.38	-0.27	0.04	0.23	0.22	0.98	0.04	-5.01	0.80	-5.38	0.78
6	678996	0	1116	0.30	0.43	0.21	0.28	0.05	0.03	0.29	-0.26	0.03	0.18	0.17	0.91	0.03	2.88	1.10	3.54	1.16

Table J–19. Literature Constructed-Response Item Statistics: Summer

Item Information			Classical												Rasch		Infit		Outfit	
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*	MS	*z*	MS
1	643960	0	588	0.48	0.13	0.37	0.40	0.09	0.01	0.52	-0.39	-0.17	0.27	0.30	0.68	0.06	-0.87	0.95	-0.98	0.95
2	616673	0	588	0.35	0.20	0.49	0.23	0.03	0.04	0.64	-0.51	-0.02	0.43	0.28	1.26	0.06	-7.64	0.64	-7.56	0.64
3	616672	0	588	0.36	0.19	0.47	0.28	0.02	0.05	0.55	-0.47	0.00	0.36	0.22	0.97	0.05	-7.53	0.66	-7.32	0.66
4	613250	0	588	0.40	0.17	0.44	0.30	0.05	0.04	0.61	-0.50	-0.07	0.37	0.30	1.66	0.06	2.09	1.12	1.65	1.09
5	683411	0	588	0.33	0.21	0.50	0.24	0.01	0.05	0.52	-0.42	0.01	0.40	0.12	1.32	0.06	-5.62	0.73	-5.58	0.73
6	683414	0	588	0.33	0.19	0.52	0.21	0.01	0.07	0.43	-0.36	0.05	0.28	0.16	1.18	0.06	-1.42	0.93	-1.57	0.92

APPENDIX K: RAW-TO-SCALED 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	1206	92	1200	1298
1	1268	51	1217	1319
2	1305	37	1268	1342
3	1327	31	1296	1358
4	1344	27	1317	1371
5	1357	25	1332	1382
6	1369	23	1346	1392
7	1379	22	1357	1401
8	1387	20	1367	1407
9	1395	20	1375	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	1439	16	1423	1455
17	1444	16	1428	1460
18	1449	16	1433	1465
19	1454	15	1439	1469
20	1458	15	1443	1473
21	1463	15	1448	1478
22	1467	15	1452	1482
23	1471	15	1456	1486
24	1476	14	1462	1490
25	1480	14	1466	1494
26	1484	14	1470	1498
27	1488	14	1474	1502
28	1492	14	1478	1506
29	1496	14	1482	1510

Raw	SS	CSEM	LCI	UCI
30	1500	14	1486	1514
31	1504	14	1490	1518
32	1507	14	1493	1521
33	1511	14	1497	1525
34	1515	14	1501	1529
35	1519	14	1505	1533
36	1523	14	1509	1537
37	1527	14	1513	1541
38	1532	14	1518	1546
39	1536	15	1521	1551
40	1540	15	1525	1555
41	1544	15	1529	1559
42	1549	15	1534	1564
43	1554	15	1539	1569
44	1558	16	1542	1574
45	1563	16	1547	1579
46	1569	16	1553	1585
47	1574	17	1557	1591
48	1580	17	1563	1597
49	1586	18	1568	1604
50	1592	18	1574	1610
51	1599	19	1580	1618
52	1606	20	1586	1626
53	1614	21	1593	1635
54	1623	22	1601	1645
55	1634	24	1610	1658
56	1646	26	1620	1672
57	1662	30	1632	1692
58	1683	36	1647	1719
59	1719	51	1668	1770
60	1780	92	1688	1800

Table K–3. Biology Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1204	92	1200	1296
1	1265	51	1214	1316
2	1301	36	1265	1337
3	1323	30	1293	1353
4	1339	26	1313	1365
5	1351	24	1327	1375
6	1362	22	1340	1384
7	1371	21	1350	1392
8	1379	20	1359	1399
9	1386	19	1367	1405
10	1393	18	1375	1411

Raw	SS	CSEM	LCI	UCI
11	1399	17	1382	1416
12	1405	17	1388	1422
13	1410	16	1394	1426
14	1415	16	1399	1431
15	1420	16	1404	1436
16	1425	15	1410	1440
17	1430	15	1415	1445
18	1434	15	1419	1449
19	1438	14	1424	1452
20	1442	14	1428	1456
21	1446	14	1432	1460
22	1450	14	1436	1464
23	1454	14	1440	1468
24	1458	14	1444	1472
25	1462	14	1448	1476
26	1465	14	1451	1479
27	1469	13	1456	1482
28	1473	13	1460	1486
29	1476	13	1463	1489
30	1480	13	1467	1493
31	1483	13	1470	1496
32	1487	13	1474	1500
33	1490	13	1477	1503
34	1494	13	1481	1507
35	1497	13	1484	1510
36	1501	13	1488	1514
37	1505	13	1492	1518
38	1508	13	1495	1521
39	1512	13	1499	1525
40	1515	14	1501	1529
41	1519	14	1505	1533
42	1523	14	1509	1537
43	1527	14	1513	1541
44	1530	14	1516	1544
45	1534	14	1520	1548
46	1538	14	1524	1552
47	1543	14	1529	1557
48	1547	15	1532	1562
49	1551	15	1536	1566
50	1556	15	1541	1571
51	1560	15	1545	1575
52	1565	16	1549	1581
53	1570	16	1554	1586
54	1576	17	1559	1593
55	1581	17	1564	1598
56	1587	18	1569	1605
57	1594	19	1575	1613

Raw	SS	CSEM	LCI	UCI
58	1601	19	1582	1620
59	1609	21	1588	1630
60	1618	22	1596	1640
61	1628	24	1604	1652
62	1641	26	1615	1667
63	1657	30	1627	1687
64	1678	36	1642	1714
65	1715	51	1664	1766
66	1776	92	1684	1800

Table K–4. Literature Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1207	92	1200	1299
1	1268	51	1217	1319
2	1304	37	1267	1341
3	1327	30	1297	1357
4	1343	27	1316	1370
5	1356	24	1332	1380
6	1367	23	1344	1390
7	1376	21	1355	1397
8	1385	20	1365	1405
9	1392	19	1373	1411
10	1400	19	1381	1419
11	1406	18	1388	1424
12	1413	18	1395	1431
13	1418	17	1401	1435
14	1424	17	1407	1441
15	1430	16	1414	1446
16	1435	16	1419	1451
17	1440	16	1424	1456
18	1445	16	1429	1461
19	1450	16	1434	1466
20	1455	16	1439	1471
21	1460	15	1445	1475
22	1465	15	1450	1480
23	1469	15	1454	1484
24	1474	15	1459	1489
25	1479	15	1464	1494
26	1483	15	1468	1498
27	1488	15	1473	1503
28	1493	15	1478	1508
29	1497	15	1482	1512
30	1502	16	1486	1518
31	1507	16	1491	1523
32	1512	16	1496	1528

Raw	SS	CSEM	LCI	UCI
33	1517	16	1501	1533
34	1522	16	1506	1538
35	1528	17	1511	1545
36	1533	17	1516	1550
37	1539	17	1522	1556
38	1545	18	1527	1563
39	1551	18	1533	1569
40	1558	18	1540	1576
41	1565	19	1546	1584
42	1573	20	1553	1593
43	1581	20	1561	1601
44	1589	21	1568	1610
45	1599	23	1576	1622
46	1610	24	1586	1634
47	1622	26	1596	1648
48	1636	28	1608	1664
49	1654	32	1622	1686
50	1678	38	1640	1716
51	1716	52	1664	1768
52	1779	92	1687	1800

SPRING

Table K–5. Algebra I Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1205	92	1200	1297
1	1266	51	1215	1317
2	1303	37	1266	1340
3	1325	30	1295	1355
4	1341	27	1314	1368
5	1354	24	1330	1378
6	1365	23	1342	1388
7	1375	21	1354	1396
8	1383	20	1363	1403
9	1391	19	1372	1410
10	1398	19	1379	1417
11	1405	18	1387	1423
12	1412	18	1394	1430
13	1417	17	1400	1434
14	1423	17	1406	1440
15	1429	16	1413	1445
16	1434	16	1418	1450
17	1439	16	1423	1455
18	1444	16	1428	1460
19	1449	15	1434	1464
20	1453	15	1438	1468
21	1458	15	1443	1473
22	1462	15	1447	1477
23	1466	15	1451	1481
24	1471	14	1457	1485
25	1475	14	1461	1489
26	1479	14	1465	1493
27	1483	14	1469	1497
28	1487	14	1473	1501
29	1491	14	1477	1505
30	1494	14	1480	1508
31	1498	14	1484	1512
32	1502	14	1488	1516
33	1506	14	1492	1520
34	1509	14	1495	1523
35	1513	14	1499	1527
36	1517	14	1503	1531
37	1520	14	1506	1534
38	1524	14	1510	1538
39	1528	14	1514	1542
40	1532	14	1518	1546
41	1536	14	1522	1550
42	1540	14	1526	1554

Raw	SS	CSEM	LCI	UCI
43	1544	15	1529	1559
44	1549	15	1534	1564
45	1553	15	1538	1568
46	1558	16	1542	1574
47	1563	16	1547	1579
48	1569	17	1552	1586
49	1574	17	1557	1591
50	1581	18	1563	1599
51	1587	19	1568	1606
52	1595	20	1575	1615
53	1604	21	1583	1625
54	1613	23	1590	1636
55	1625	25	1600	1650
56	1638	27	1611	1665
57	1655	31	1624	1686
58	1677	36	1641	1713
59	1713	50	1663	1763
60	1773	91	1682	1800

Table K–6. Biology Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1212	92	1200	1304
1	1273	51	1222	1324
2	1309	36	1273	1345
3	1330	30	1300	1360
4	1345	26	1319	1371
5	1358	24	1334	1382
6	1368	22	1346	1390
7	1377	20	1357	1397
8	1385	19	1366	1404
9	1392	18	1374	1410
10	1398	18	1380	1416
11	1404	17	1387	1421
12	1410	16	1394	1426
13	1415	16	1399	1431
14	1420	16	1404	1436
15	1425	15	1410	1440
16	1429	15	1414	1444
17	1433	15	1418	1448
18	1438	14	1424	1452
19	1442	14	1428	1456
20	1446	14	1432	1460
21	1449	14	1435	1463
22	1453	14	1439	1467
23	1457	13	1444	1470

Raw	SS	CSEM	LCI	UCI
24	1460	13	1447	1473
25	1464	13	1451	1477
26	1467	13	1454	1480
27	1471	13	1458	1484
28	1474	13	1461	1487
29	1478	13	1465	1491
30	1481	13	1468	1494
31	1484	13	1471	1497
32	1488	13	1475	1501
33	1491	13	1478	1504
34	1494	13	1481	1507
35	1498	13	1485	1511
36	1501	13	1488	1514
37	1504	13	1491	1517
38	1508	13	1495	1521
39	1511	13	1498	1524
40	1515	13	1502	1528
41	1518	13	1505	1531
42	1522	13	1509	1535
43	1525	14	1511	1539
44	1529	14	1515	1543
45	1533	14	1519	1547
46	1537	14	1523	1551
47	1541	14	1527	1555
48	1545	15	1530	1560
49	1549	15	1534	1564
50	1554	15	1539	1569
51	1558	15	1543	1573
52	1563	16	1547	1579
53	1568	16	1552	1584
54	1574	17	1557	1591
55	1580	17	1563	1597
56	1586	18	1568	1604
57	1593	19	1574	1612
58	1600	20	1580	1620
59	1608	21	1587	1629
60	1617	22	1595	1639
61	1628	24	1604	1652
62	1641	27	1614	1668
63	1657	30	1627	1687
64	1679	37	1642	1716
65	1716	51	1665	1767
66	1777	92	1685	1800

Table K–7. Literature Raw-to-Scaled Score Conversion

Raw	SS	CSEM	LCI	UCI
0	1208	92	1200	1300
1	1269	51	1218	1320
2	1306	36	1270	1342
3	1327	30	1297	1357
4	1343	26	1317	1369
5	1356	24	1332	1380
6	1367	22	1345	1389
7	1376	21	1355	1397
8	1384	20	1364	1404
9	1392	19	1373	1411
10	1399	18	1381	1417
11	1405	18	1387	1423
12	1411	17	1394	1428
13	1417	17	1400	1434
14	1422	16	1406	1438
15	1428	16	1412	1444
16	1433	16	1417	1449
17	1438	16	1422	1454
18	1442	15	1427	1457
19	1447	15	1432	1462
20	1452	15	1437	1467
21	1456	15	1441	1471
22	1461	15	1446	1476
23	1465	15	1450	1480
24	1469	15	1454	1484
25	1474	15	1459	1489
26	1478	15	1463	1493
27	1483	15	1468	1498
28	1487	15	1472	1502
29	1491	15	1476	1506
30	1496	15	1481	1511
31	1501	15	1486	1516
32	1505	15	1490	1520
33	1510	16	1494	1526
34	1515	16	1499	1531
35	1520	16	1504	1536
36	1525	16	1509	1541
37	1531	17	1514	1548
38	1536	17	1519	1553
39	1542	17	1525	1559
40	1549	18	1531	1567
41	1555	18	1537	1573
42	1562	19	1543	1581
43	1570	20	1550	1590
44	1578	21	1557	1599

Raw	SS	CSEM	LCI	UCI
45	1587	22	1565	1609
46	1597	23	1574	1620
47	1609	25	1584	1634
48	1623	27	1596	1650
49	1640	31	1609	1671
50	1663	37	1626	1700
51	1700	51	1649	1751
52	1762	92	1670	1800

SUMMER

Table K–8. Algebra Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1209	92	1200	1301
1	1271	51	1220	1322
2	1307	37	1270	1344
3	1329	30	1299	1359
4	1345	27	1318	1372
5	1358	24	1334	1382
6	1369	22	1347	1391
7	1379	21	1358	1400
8	1387	20	1367	1407
9	1395	19	1376	1414
10	1402	19	1383	1421
11	1409	18	1391	1427
12	1415	17	1398	1432
13	1421	17	1404	1438
14	1426	17	1409	1443
15	1432	16	1416	1448
16	1437	16	1421	1453
17	1442	16	1426	1458
18	1447	16	1431	1463
19	1452	15	1437	1467
20	1456	15	1441	1471
21	1461	15	1446	1476
22	1465	15	1450	1480
23	1470	15	1455	1485
24	1474	15	1459	1489
25	1478	15	1463	1493
26	1483	15	1468	1498
27	1487	14	1473	1501
28	1491	14	1477	1505
29	1495	14	1481	1509
30	1499	14	1485	1513
31	1503	14	1489	1517

Raw	SS	CSEM	LCI	UCI
32	1507	14	1493	1521
33	1511	14	1497	1525
34	1515	14	1501	1529
35	1519	14	1505	1533
36	1523	14	1509	1537
37	1527	14	1513	1541
38	1531	14	1517	1545
39	1535	14	1521	1549
40	1540	14	1526	1554
41	1544	14	1530	1558
42	1548	15	1533	1563
43	1552	15	1537	1567
44	1556	15	1541	1571
45	1561	15	1546	1576
46	1566	15	1551	1581
47	1571	16	1555	1587
48	1576	16	1560	1592
49	1581	17	1564	1598
50	1587	18	1569	1605
51	1594	19	1575	1613
52	1601	20	1581	1621
53	1609	21	1588	1630
54	1619	23	1596	1642
55	1630	25	1605	1655
56	1644	27	1617	1671
57	1660	31	1629	1691
58	1682	36	1646	1718
59	1715	48	1667	1763
60	1770	88	1682	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

Raw	SS	CSEM	LCI	UCI
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	13	1443	1469
23	1460	13	1447	1473
24	1463	13	1450	1476
25	1467	13	1454	1480
26	1470	13	1457	1483
27	1473	13	1460	1486
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	1500	13	1487	1513
36	1503	13	1490	1516
37	1506	13	1493	1519
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	1527	13	1514	1540
44	1530	14	1516	1544
45	1534	14	1520	1548
46	1538	14	1524	1552
47	1542	14	1528	1556
48	1546	14	1532	1560
49	1550	15	1535	1565
50	1554	15	1539	1569
51	1559	15	1544	1574
52	1564	16	1548	1580
53	1569	16	1553	1585
54	1574	17	1557	1591
55	1580	17	1563	1597
56	1586	18	1568	1604
57	1593	19	1574	1612
58	1600	20	1580	1620
59	1608	21	1587	1629

Raw	SS	CSEM	LCI	UCI
60	1617	22	1595	1639
61	1628	24	1604	1652
62	1641	27	1614	1668
63	1657	31	1626	1688
64	1680	37	1643	1717
65	1717	51	1666	1768
66	1779	92	1687	1800

Table K–10. Literature Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1208	92	1200	1300
1	1269	51	1218	1320
2	1306	37	1269	1343
3	1328	30	1298	1358
4	1344	27	1317	1371
5	1357	24	1333	1381
6	1368	22	1346	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	1436	16	1420	1452
17	1441	16	1425	1457
18	1446	16	1430	1462
19	1451	16	1435	1467
20	1455	15	1440	1470
21	1460	15	1445	1475
22	1465	15	1450	1480
23	1470	15	1455	1485
24	1474	15	1459	1489
25	1479	15	1464	1494
26	1483	15	1468	1498
27	1488	15	1473	1503
28	1493	15	1478	1508
29	1498	15	1483	1513
30	1502	16	1486	1518
31	1507	16	1491	1523
32	1512	16	1496	1528
33	1517	16	1501	1533
34	1522	16	1506	1538

Raw	SS	CSEM	LCI	UCI
35	1528	17	1511	1545
36	1533	17	1516	1550
37	1539	17	1522	1556
38	1545	18	1527	1563
39	1551	18	1533	1569
40	1558	18	1540	1576
41	1565	19	1546	1584
42	1573	20	1553	1593
43	1581	21	1560	1602
44	1590	21	1569	1611
45	1599	23	1576	1622
46	1610	24	1586	1634
47	1622	26	1596	1648
48	1637	28	1609	1665
49	1655	32	1623	1687
50	1679	38	1641	1717
51	1717	52	1665	1769
52	1780	92	1688	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
640597	MC	-9	6471	75949	-0.78	0.06	-0.65	0.01	-0.63	0.16
605261	MC	-12	4179	75949	0.08	0.05	-0.27	0.01	-0.25	-0.31
640591	MC	-11	9137	75949	-0.86	0.06	-1.08	0.01	-1.06	-0.22
640578	MC	-9	9169	75949	0.70	0.05	0.38	0.01	0.40	-0.27
605195	MC	3	94697	75949	0.04	0.01	-0.20	0.01	-0.18	-0.21
605114	MC	-2	94697	75949	-0.36	0.01	-0.38	0.01	-0.36	0.01
640568	MC	-6	6538	75949	1.62	0.05	1.34	0.01	1.36	-0.19
605255	MC	-2	4349	75949	-0.87	0.05	-1.00	0.01	-0.97	-0.11
605262	MC	4	94697	75949	0.95	0.01	0.90	0.01	0.92	0.02
640599	MC	4	6494	75949	0.61	0.05	0.52	0.01	0.54	-0.03
640629	MC	4	6472	75949	0.07	0.05	0.18	0.01	0.20	0.16
640572	MC	7	9186	75949	1.15	0.05	0.97	0.01	0.99	-0.11
640543	MC	10	9293	75949	-1.07	0.06	-0.97	0.01	-0.95	0.12
605057	MC	-3	94697	75949	-0.07	0.01	-0.38	0.01	-0.36	-0.29
640579	MC	9	9211	75949	1.41	0.05	1.21	0.01	1.23	-0.12
640529	MC	9	6488	75949	-0.03	0.05	0.20	0.01	0.22	0.28
605254	MC	-1	94697	75949	-0.74	0.01	-0.65	0.01	-0.63	0.12
605170	MC	24	94697	75949	-1.12	0.01	-0.84	0.01	-0.82	0.30
605180	MC	-21	94697	75949	-1.20	0.01	-1.57	0.01	-1.55	-0.39
605273	MC	-8	9137	75949	-2.11	0.08	-2.42	0.01	-2.40	-0.36
605280	MC	-5	94697	75949	0.10	0.01	0.09	0.01	0.11	0.04
641441	MC	-7	9271	75949	0.75	0.05	0.27	0.01	0.29	-0.42
641539	MC	-8	9186	75949	1.55	0.05	1.16	0.01	1.18	-0.31
641466	MC	-5	6494	75949	-0.35	0.05	-0.21	0.01	-0.19	0.18
641524	MC	-5	6538	75949	0.92	0.05	0.86	0.01	0.88	0.01
605283	MC	-15	94697	75949	-0.42	0.01	-0.65	0.01	-0.63	-0.22
605206	MC	3	94697	75949	0.01	0.01	0.00	0.01	0.03	0.04
641523	MC	5	6488	75949	-0.05	0.05	0.08	0.01	0.10	0.18
641460	MC	4	6538	75949	0.07	0.05	0.34	0.01	0.36	0.32
605115	MC	5	4187	75949	-0.60	0.05	0.04	0.01	0.06	0.69
641519	MC	9	6540	75949	1.85	0.06	1.45	0.01	1.47	-0.31
641478	MC	8	10082	75949	0.12	0.05	0.38	0.01	0.40	0.32
641451	MC	9	9169	75949	-0.28	0.05	-0.04	0.01	-0.02	0.28
605230	MC	-1	94697	75949	-0.12	0.01	0.02	0.01	0.04	0.19
605096	MC	11	4176	75949	0.43	0.05	0.15	0.01	0.17	-0.23
605168	MC	16	94697	75949	1.00	0.01	0.78	0.01	0.80	-0.15
640123	CR	-1	1972	75949	1.45	0.03	1.49	0.01	1.51	-0.11

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
640892	CR	10	1995	75949	2.05	0.03	1.99	0.01	2.01	-0.09
612749	CR	11	1945	75949	1.42	0.02	1.48	0.00	1.50	-0.01
640886	CR	-1	2029	75949	1.71	0.03	1.62	0.00	1.64	0.01
640898	CR	10	1995	75949	1.81	0.03	2.09	0.01	2.12	0.27
622485	CR	11	2008	75949	0.97	0.02	1.02	0.00	1.04	0.02

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
643772	MC	-21	5353	54247	-0.70	0.05	-0.72	0.01	-1.01	-0.24
643380	MC	-20	8129	54247	-0.72	0.05	-0.71	0.01	-0.99	-0.22
635760	MC	-12	1713	54247	-0.63	0.05	-0.54	0.01	-0.83	-0.13
641198	MC	-11	8065	54247	-0.37	0.05	-0.06	0.01	-0.34	0.10
610456	MC	-9	1821	54247	-0.98	0.05	-0.89	0.01	-1.17	-0.13
607797	MC	-14	1735	54247	-0.64	0.05	-0.42	0.01	-0.70	0.00
642866	MC	-13	5191	54247	0.65	0.05	0.67	0.01	0.38	-0.20
642850	MC	-7	5378	54247	-0.07	0.05	-0.06	0.01	-0.34	-0.20
607704	MC	2	46979	54247	0.01	0.01	0.09	0.01	-0.19	-0.14
611033	MC	-2	46979	54247	0.89	0.01	1.00	0.01	0.72	-0.10
642841	MC	-8	5347	54247	-0.38	0.05	-0.08	0.01	-0.36	0.09
640665	MC	-3	8149	54247	0.83	0.05	0.79	0.01	0.51	-0.25
644729	MC	-1	8072	54247	-0.18	0.05	0.04	0.01	-0.24	0.01
644733	MC	0	8062	54247	0.61	0.05	0.83	0.01	0.55	0.01
607732	MC	-2	46979	54247	-0.59	0.01	-0.59	0.01	-0.88	-0.22
642843	MC	5	5356	54247	-0.12	0.05	-0.21	0.01	-0.50	-0.31
642335	MC	7	8060	54247	0.44	0.05	0.59	0.01	0.30	-0.07
643387	MC	7	5378	54247	0.35	0.05	0.54	0.01	0.26	-0.02
611445	MC	13	1735	54247	-0.20	0.05	-0.35	0.01	-0.63	-0.37
611154	MC	6	46979	54247	-0.22	0.01	0.08	0.01	-0.21	0.08
642864	MC	11	8065	54247	-0.42	0.05	0.09	0.01	-0.20	0.29
642837	MC	16	5341	54247	0.34	0.05	0.55	0.01	0.27	0.00
643371	MC	13	5341	54247	0.05	0.05	0.47	0.01	0.18	0.20
643411	MC	13	8062	54247	0.52	0.05	0.64	0.01	0.35	-0.10
607756	MC	-2	46979	54247	-1.95	0.01	-2.05	0.01	-2.34	-0.35
611614	MC	-5	46979	54247	-0.09	0.01	0.14	0.01	-0.14	0.02
643390	MC	-18	8065	54247	0.56	0.05	0.69	0.01	0.40	-0.09
611055	MC	-7	46979	54247	-0.96	0.01	-0.85	0.01	-1.14	-0.12
642372	MC	-15	8149	54247	-1.37	0.06	-1.80	0.01	-2.09	-0.69
608250	MC	-10	5347	54247	-0.34	0.05	-0.07	0.01	-0.35	0.06
641193	MC	-9	8058	54247	-0.19	0.05	-0.21	0.01	-0.49	-0.24
641202	MC	-7	5378	54247	-0.84	0.06	-0.25	0.01	-0.54	0.38
609392	MC	8	46979	54247	-1.40	0.01	-1.03	0.01	-1.32	0.15
608284	MC	-2	46979	54247	-0.51	0.01	-0.23	0.01	-0.52	0.07

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
642387	MC	-8	8062	54247	0.35	0.05	0.62	0.01	0.34	0.05
642385	MC	-3	5356	54247	-0.37	0.05	0.05	0.01	-0.23	0.21
644616	MC	0	8065	54247	-0.17	0.05	-0.07	0.01	-0.35	-0.11
608261	MC	0	8065	54247	0.44	0.05	0.62	0.01	0.33	-0.04
657189	MC	4	5341	54247	-0.08	0.05	0.02	0.01	-0.26	-0.12
641282	MC	7	8901	54247	0.36	0.05	0.73	0.01	0.45	0.15
642339	MC	8	8109	54247	1.01	0.05	0.85	0.01	0.57	-0.37
611469	MC	-1	46979	54247	0.28	0.01	0.18	0.01	-0.10	-0.31
642859	MC	9	8062	54247	-0.01	0.05	0.54	0.01	0.25	0.34
642371	MC	10	5360	54247	-0.18	0.05	0.26	0.01	-0.02	0.23
642378	MC	12	5191	54247	-0.15	0.05	0.15	0.01	-0.13	0.09
607597	MC	17	8058	54247	-0.12	0.05	0.07	0.01	-0.22	-0.03
609822	MC	11	1757	54247	-0.30	0.05	-0.02	0.01	-0.31	0.07
611419	MC	9	46979	54247	-0.41	0.01	-0.10	0.01	-0.38	0.10
629469	CR	5	1726	54247	0.99	0.03	1.44	0.01	1.15	-0.15
644134	CR	17	1870	54247	1.68	0.03	2.48	0.01	2.19	0.29
641221	CR	18	1976	54247	1.45	0.03	1.80	0.01	1.51	-0.15
641299	CR	5	1926	54247	1.38	0.03	2.02	0.01	1.74	0.37
641303	CR	17	1964	54247	1.57	0.03	2.10	0.01	1.81	0.14
611059	CR	18	1820	54247	0.35	0.03	1.21	0.01	0.92	0.46

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	7017	47157	1.49	0.05	1.14	0.01	1.35	-0.06
641594	MC	-11	7017	47157	0.40	0.05	-0.08	0.01	0.13	-0.26
641612	MC	-11	7017	47157	-0.20	0.06	-0.54	0.01	-0.33	-0.15
641613	MC	-9	7040	47157	-0.41	0.06	-0.22	0.01	-0.01	0.40
641598	MC	-9	7040	47157	1.00	0.05	0.53	0.01	0.74	-0.21
641595	MC	-9	7040	47157	0.62	0.05	-0.05	0.01	0.16	-0.44
641599	MC	-8	7017	47157	0.49	0.05	0.12	0.01	0.33	-0.14
641592	MC	-8	7040	47157	1.47	0.05	0.75	0.01	0.96	-0.44
614671	MC	7	1859	47157	-0.84	0.06	-0.50	0.01	-0.29	0.54
614662	MC	6	1860	47157	0.09	0.05	-0.05	0.01	0.16	0.08
614663	MC	6	1859	47157	-0.30	0.06	-0.36	0.01	-0.15	0.15
614667	MC	8	1859	47157	0.61	0.05	0.64	0.01	0.85	0.29
614668	MC	11	1860	47157	-0.57	0.06	-0.46	0.01	-0.25	0.31
614664	MC	8	1860	47157	-0.37	0.06	-0.44	0.01	-0.23	0.13
614665	MC	12	1859	47157	-1.29	0.07	-0.99	0.01	-0.78	0.47
614661	MC	10	1859	47157	0.75	0.05	0.60	0.01	0.82	0.11
614669	MC	10	1859	47157	-1.07	0.07	-0.88	0.01	-0.67	0.36
612408	MC	-11	1859	47157	-1.06	0.07	-1.18	0.01	-0.97	0.04
612400	MC	-11	1859	47157	0.20	0.05	-0.23	0.01	-0.01	-0.21

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
612403	MC	-12	1859	47157	-0.01	0.05	-0.65	0.01	-0.44	-0.46
612402	MC	-9	1860	47157	1.17	0.05	0.61	0.01	0.83	-0.29
612406	MC	-10	1860	47157	-0.08	0.05	-0.72	0.01	-0.51	-0.46
612405	MC	-10	1860	47157	0.05	0.05	-0.51	0.01	-0.30	-0.36
612446	MC	-10	1859	47157	1.15	0.05	0.80	0.01	1.01	-0.08
640207	MC	-6	1860	47157	0.10	0.05	-0.31	0.01	-0.10	-0.20
612404	MC	-8	1860	47157	1.19	0.05	0.63	0.01	0.84	-0.29
642702	MC	8	4690	47157	0.66	0.05	0.40	0.01	0.62	-0.01
642704	MC	8	4742	47157	0.14	0.06	0.04	0.01	0.25	0.13
642703	MC	9	4690	47157	0.45	0.05	0.09	0.01	0.30	-0.13
642707	MC	9	4742	47157	0.31	0.06	0.30	0.01	0.51	0.23
642706	MC	9	4742	47157	1.54	0.05	0.87	0.01	1.09	-0.39
642724	MC	10	4690	47157	0.93	0.05	0.76	0.01	0.97	0.10
642705	MC	10	4690	47157	0.09	0.06	0.03	0.01	0.24	0.17
642723	MC	11	4742	47157	0.00	0.06	-0.13	0.01	0.09	0.09
643730	CR	-8	1872	47157	1.23	0.03	1.01	0.01	1.22	0.01
614675	CR	10	1860	47157	0.88	0.04	0.93	0.01	1.14	0.27
614674	CR	11	1859	47157	0.87	0.03	0.91	0.01	1.12	0.25
614016	CR	-8	1858	47157	1.69	0.03	0.87	0.01	1.08	-0.55
644768	CR	10	1794	47157	0.26	0.04	0.49	0.01	0.70	0.45
644767	CR	11	1808	47157	0.77	0.03	0.73	0.01	0.95	0.10

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
674448	MC	-10	6136	184157	-1.32	0.07	-1.23	0.01	-1.27	-0.01
605137	MC	-10	4207	184157	-0.14	0.05	-0.18	0.01	-0.22	-0.11
605189	MC	-9	4195	184157	0.64	0.05	0.49	0.01	0.44	-0.20
633250	MC	-7	3682	184157	-0.48	0.05	-0.45	0.01	-0.49	-0.04
674148	MC	-8	6155	184157	0.61	0.05	0.47	0.01	0.43	-0.17
632148	MC	-8	4213	184157	-0.74	0.05	-0.82	0.01	-0.86	-0.18
674491	MC	-7	6135	184157	0.65	0.05	0.35	0.01	0.30	-0.35
674455	MC	-3	6155	184157	0.67	0.05	0.38	0.01	0.34	-0.33
631599	MC	-3	3664	184157	-0.72	0.06	-0.78	0.01	-0.83	-0.15
640609	MC	4	6471	184157	0.76	0.05	0.88	0.01	0.84	0.09
657749	MC	6	4114	184157	-0.57	0.06	-0.31	0.01	-0.36	0.19
605108	MC	7	4193	184157	-0.15	0.05	-0.05	0.01	-0.10	0.03
640576	MC	9	6399	184157	0.38	0.05	0.68	0.01	0.64	0.27
674511	MC	8	4153	184157	-0.54	0.06	-0.31	0.01	-0.35	0.16
674419	MC	11	6148	184157	0.21	0.06	0.27	0.01	0.23	0.01
605172	MC	9	6438	184157	0.78	0.05	0.84	0.01	0.79	0.02
605286	MC	10	4213	184157	1.36	0.05	1.20	0.01	1.16	-0.17
674384	MC	12	4152	184157	0.31	0.06	0.38	0.01	0.34	0.02

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
678779	MC	-10	6126	184157	-1.55	0.08	-1.57	0.01	-1.61	-0.14
682132	MC	-8	6126	184157	0.82	0.05	0.79	0.01	0.75	-0.06
678752	MC	-7	4114	184157	-0.66	0.06	-0.47	0.01	-0.52	0.11
678731	MC	-9	4146	184157	0.41	0.06	0.84	0.01	0.80	0.40
605110	MC	-6	3698	184157	-0.62	0.05	-0.73	0.01	-0.77	-0.21
678806	MC	-8	4152	184157	0.77	0.06	0.66	0.01	0.62	-0.14
632358	MC	-3	3671	184157	-0.47	0.05	-0.82	0.01	-0.86	-0.45
633251	MC	-3	4195	184157	-0.51	0.05	-0.26	0.01	-0.30	0.18
605301	MC	-4	3698	184157	-0.87	0.06	-0.69	0.01	-0.73	0.10
678772	MC	7	4146	184157	-1.24	0.07	-0.78	0.01	-0.83	0.38
605245	MC	7	3709	184157	0.22	0.05	0.03	0.01	-0.01	-0.25
641484	MC	7	9211	184157	1.63	0.05	1.60	0.01	1.56	-0.03
633531	MC	9	4213	184157	0.23	0.05	0.08	0.01	0.04	-0.20
678814	MC	9	4146	184157	0.08	0.06	0.24	0.01	0.19	0.12
605101	MC	12	4207	184157	-1.06	0.06	-0.64	0.01	-0.69	0.34
673951	MC	12	6146	184157	-0.65	0.06	-0.21	0.01	-0.25	0.38
678774	MC	14	6148	184157	0.52	0.05	0.28	0.01	0.23	-0.29
678785	MC	11	4153	184157	-0.20	0.06	-0.16	0.01	-0.21	-0.03
672731	CR	-1	1640	184157	1.82	0.03	1.77	0.00	1.73	-0.08
641527	CR	10	1963	184157	1.90	0.03	2.05	0.00	2.00	0.10
628222	CR	11	2002	184157	0.69	0.02	0.96	0.00	0.91	0.23
666526	CR	-1	1688	184157	2.04	0.03	1.93	0.00	1.88	-0.07
628223	CR	10	2002	184157	0.93	0.02	0.92	0.00	0.88	-0.07
678819	CR	11	1685	184157	1.10	0.03	1.14	0.00	1.10	0.04

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
607777	MC	-12	1613	154596	-0.24	0.05	-0.13	0.01	-0.14	0.10
674080	MC	-12	5790	154596	0.05	0.06	-0.23	0.01	-0.24	-0.30
611477	MC	-16	1734	154596	-1.04	0.06	-0.96	0.01	-0.97	0.07
680195	MC	-10	5869	154596	-0.55	0.06	-0.32	0.01	-0.33	0.23
635783	MC	-16	1744	154596	-0.29	0.05	-0.81	0.01	-0.82	-0.55
678545	MC	-14	4025	154596	-0.81	0.06	-0.79	0.01	-0.80	0.01
680242	MC	-12	5842	154596	-0.11	0.06	-0.25	0.01	-0.26	-0.15
609427	MC	-14	1620	154596	-0.36	0.05	-1.09	0.01	-1.10	-0.77
635766	MC	-12	1592	154596	0.43	0.05	0.46	0.01	0.45	0.01
610882	MC	-11	1613	154596	-0.12	0.05	-0.21	0.01	-0.22	-0.09
644752	MC	-2	8060	154596	0.34	0.05	0.32	0.01	0.31	-0.03
644755	MC	-2	8065	154596	0.28	0.05	0.10	0.01	0.09	-0.19
678875	MC	3	5845	154596	-0.85	0.06	-0.92	0.01	-0.93	-0.08
635786	MC	2	1748	154596	0.40	0.05	0.33	0.01	0.32	-0.07
677989	MC	10	5790	154596	0.15	0.06	0.29	0.01	0.28	0.14

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
671325	MC	11	4025	154596	0.17	0.05	-0.43	0.01	-0.44	-0.62
678965	MC	12	5845	154596	0.79	0.06	0.82	0.01	0.81	0.01
610880	MC	6	1632	154596	-0.16	0.05	-0.01	0.01	-0.02	0.15
642317	MC	10	8072	154596	1.33	0.05	1.12	0.01	1.11	-0.22
635799	MC	8	1734	154596	-0.58	0.05	-0.75	0.01	-0.76	-0.18
635597	MC	12	1761	154596	-0.57	0.05	-0.35	0.01	-0.36	0.21
607712	MC	18	1735	154596	0.37	0.05	0.36	0.01	0.35	-0.01
681052	MC	19	3990	154596	0.51	0.05	0.71	0.01	0.70	0.19
677992	MC	17	4018	154596	0.58	0.05	0.50	0.01	0.49	-0.09
643400	MC	-19	4028	154596	0.95	0.05	0.65	0.01	0.64	-0.31
681523	MC	-14	5797	154596	-0.06	0.05	0.02	0.01	0.01	0.08
608245	MC	-13	1616	154596	-0.76	0.06	-0.91	0.01	-0.92	-0.16
607589	MC	-12	1744	154596	-0.06	0.05	0.32	0.01	0.31	0.38
611072	MC	-11	1744	154596	-0.37	0.05	-0.73	0.01	-0.74	-0.38
611066	MC	-16	1713	154596	-0.21	0.05	-0.02	0.01	-0.03	0.18
611046	MC	-8	1713	154596	0.15	0.05	0.03	0.01	0.02	-0.13
608286	MC	-7	1619	154596	0.36	0.05	0.64	0.01	0.63	0.27
607578	MC	-5	1620	154596	0.67	0.06	1.03	0.01	1.02	0.35
683555	MC	-8	5842	154596	-0.07	0.06	0.16	0.01	0.14	0.22
684511	MC	-3	5869	154596	-0.57	0.06	-0.47	0.01	-0.48	0.10
682405	MC	-2	5845	154596	0.39	0.06	0.40	0.01	0.39	0.00
611476	MC	10	1595	154596	0.00	0.05	0.07	0.01	0.06	0.06
612588	MC	2	1737	154596	0.37	0.05	0.74	0.01	0.73	0.36
612675	MC	11	1609	154596	-0.22	0.05	-0.30	0.01	-0.31	-0.09
610909	MC	13	1592	154596	0.05	0.05	0.07	0.01	0.06	0.01
608019	MC	14	1748	154596	0.16	0.05	0.13	0.01	0.12	-0.04
608382	MC	13	1598	154596	-1.00	0.06	-0.78	0.01	-0.79	0.21
642871	MC	11	4025	154596	0.66	0.05	0.69	0.01	0.68	0.02
608279	MC	14	1619	154596	-0.09	0.05	0.12	0.01	0.11	0.20
641293	MC	12	5797	154596	-0.37	0.06	-0.30	0.01	-0.31	0.06
678986	MC	13	4025	154596	-0.82	0.06	-0.64	0.01	-0.65	0.18
607750	MC	13	1611	154596	0.67	0.06	0.74	0.01	0.73	0.06
678894	MC	14	5795	154596	0.76	0.06	0.57	0.01	0.56	-0.20
678999	CR	-1	1700	154596	0.39	0.03	0.31	0.00	0.30	-0.10
679989	CR	13	1507	154596	0.94	0.03	1.03	0.00	1.02	0.09
678998	CR	14	1745	154596	1.11	0.03	1.07	0.00	1.06	-0.05
684325	CR	-1	1777	154596	0.90	0.03	0.92	0.00	0.91	0.07
682671	CR	13	1702	154596	1.67	0.03	1.96	0.00	1.95	0.19
644202	CR	18	1874	154596	0.38	0.03	0.43	0.00	0.42	0.04

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
610362	2	-12	1881	140256	-0.23	0.06	-0.58	0.01	-0.51	-0.31
610359	2	-10	1860	140256	-0.34	0.06	-0.48	0.01	-0.41	-0.08
610357	2	-7	1881	140256	0.20	0.05	-0.20	0.01	-0.13	-0.33
610364	2	-10	1860	140256	0.24	0.05	0.08	0.01	0.15	-0.07
610366	2	-10	1881	140256	0.60	0.05	0.22	0.01	0.29	-0.28
610363	2	-7	1860	140256	0.58	0.05	-0.07	0.01	0.00	-0.56
610361	2	-7	1881	140256	-0.11	0.05	-0.04	0.01	0.03	0.15
610365	2	-7	1860	140256	0.22	0.05	-0.26	0.01	-0.19	-0.42
610358	2	-2	1860	140256	1.27	0.05	0.84	0.01	0.91	-0.30
640841	2	7	4742	140256	-0.56	0.06	-0.61	0.01	-0.54	-0.01
640840	2	7	4690	140256	0.33	0.05	0.50	0.01	0.56	0.27
640857	2	8	4742	140256	-0.91	0.07	-0.90	0.01	-0.83	0.05
640839	2	8	4690	140256	-1.03	0.07	-0.65	0.01	-0.58	0.43
640858	2	9	4742	140256	0.39	0.05	0.22	0.01	0.29	-0.08
640842	2	9	4742	140256	-0.19	0.06	0.05	0.01	0.12	0.33
640859	2	10	4690	140256	-0.09	0.06	-0.05	0.01	0.02	0.12
640836	2	10	4742	140256	0.38	0.05	0.02	0.01	0.09	-0.28
683335	2	-10	3854	140256	-1.07	0.07	-1.11	0.01	-1.04	-0.02
683343	2	-9	3891	140256	0.72	0.05	0.66	0.01	0.73	0.06
683461	2	-9	3854	140256	0.35	0.06	0.26	0.01	0.33	0.00
683338	2	-9	3891	140256	0.25	0.06	0.13	0.01	0.20	-0.03
683342	2	-9	3854	140256	-0.22	0.06	-0.22	0.01	-0.15	0.07
683336	2	-9	3891	140256	-0.82	0.07	-0.95	0.01	-0.88	-0.09
683340	2	-9	3891	140256	1.07	0.05	0.51	0.01	0.58	-0.44
683460	2	-8	3854	140256	1.34	0.05	1.07	0.01	1.14	-0.13
614543	2	2	1871	140256	-0.47	0.06	-0.47	0.01	-0.40	0.06
614547	2	5	1871	140256	-1.02	0.06	-0.38	0.01	-0.31	0.71
614541	2	7	1867	140256	-0.18	0.06	-0.10	0.01	-0.03	0.16
614542	2	7	1867	140256	0.61	0.05	0.76	0.01	0.83	0.27
614544	2	7	1867	140256	-0.53	0.06	-0.78	0.01	-0.71	-0.21
614545	2	5	1867	140256	2.14	0.06	2.30	0.01	2.37	0.32
614550	2	9	1871	140256	0.14	0.05	-0.17	0.01	-0.10	-0.24
614546	2	9	1867	140256	0.40	0.05	0.19	0.01	0.26	-0.11
614539	2	8	1871	140256	0.33	0.05	0.23	0.01	0.30	0.00
613661	4	-6	1860	140256	1.38	0.03	1.27	0.00	1.34	-0.14
643181	4	9	1794	140256	0.76	0.03	1.29	0.00	1.36	0.66
643182	4	10	1808	140256	0.68	0.03	0.77	0.00	0.84	0.13
683334	4	-8	1670	140256	1.01	0.04	0.75	0.00	0.82	-0.18
614551	4	8	1865	140256	0.82	0.03	0.47	0.00	0.54	-0.28
614552	4	9	1868	140256	0.73	0.03	0.89	0.00	0.96	0.22

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
605225	2	-10	4207	1848	-1.74	0.07	-2.18	0.08	-2.10	-0.46
674458	2	-9	4146	1848	0.79	0.06	0.73	0.05	0.81	0.02
640555	2	-11	6540	1848	0.51	0.05	0.87	0.05	0.95	0.45
631597	2	-6	3682	1848	-0.69	0.05	-0.43	0.05	-0.35	0.32
681810	2	-6	4155	1848	1.04	0.06	0.62	0.05	0.70	-0.33
605216	2	-5	3671	1848	-0.23	0.05	0.02	0.05	0.10	0.32
632351	2	-6	3652	1848	0.15	0.05	0.03	0.05	0.11	-0.06
605205	2	-5	4195	1848	1.33	0.05	1.11	0.05	1.19	-0.12
674495	2	-2	4205	1848	-0.22	0.06	-0.32	0.05	-0.24	-0.05
605264	2	7	3671	1848	-0.46	0.05	-0.26	0.05	-0.18	0.26
674454	2	5	6146	1848	1.15	0.05	0.93	0.05	1.01	-0.12
674516	2	7	6147	1848	-0.15	0.06	-0.20	0.05	-0.12	0.01
605203	2	8	3698	1848	0.51	0.05	0.32	0.05	0.40	-0.11
681809	2	10	6165	1848	0.68	0.05	0.43	0.05	0.51	-0.18
674493	2	9	4114	1848	1.40	0.06	1.25	0.05	1.33	-0.05
674456	2	9	4146	1848	-0.51	0.06	-0.34	0.05	-0.26	0.22
640534	2	12	9194	1848	0.81	0.05	0.68	0.05	0.76	-0.04
674509	2	11	4153	1848	-0.64	0.06	-0.56	0.05	-0.48	0.12
605175	2	-9	3682	1848	-0.69	0.05	-0.93	0.06	-0.85	-0.21
678811	2	-9	6165	1848	-0.03	0.06	-0.22	0.05	-0.14	-0.14
682039	2	-8	4210	1848	-0.30	0.06	-0.69	0.05	-0.61	-0.35
632154	2	-8	4207	1848	-0.59	0.05	-0.68	0.05	-0.60	-0.05
605246	2	-8	3664	1848	-0.86	0.06	-0.95	0.06	-0.87	-0.05
633004	2	-7	4179	1848	0.66	0.05	0.19	0.05	0.28	-0.39
678802	2	-5	4131	1848	-0.06	0.06	-0.34	0.05	-0.25	-0.22
605306	2	-2	4207	1848	-0.57	0.05	-0.31	0.05	-0.23	0.31
666556	2	-5	4205	1848	-0.47	0.06	-0.60	0.05	-0.52	-0.08
678777	2	7	6182	1848	-0.37	0.06	-0.60	0.05	-0.52	-0.18
678765	2	5	6182	1848	0.24	0.06	0.01	0.05	0.09	-0.16
674153	2	6	6148	1848	-0.29	0.06	-0.35	0.05	-0.26	0.00
633411	2	7	4213	1848	-0.40	0.05	-0.54	0.05	-0.46	-0.09
681814	2	10	6214	1848	0.75	0.05	0.85	0.05	0.93	0.20
632156	2	10	3697	1848	0.58	0.05	0.57	0.05	0.65	0.07
641533	2	9	6438	1848	1.67	0.05	1.42	0.06	1.50	-0.14
678766	2	12	6146	1848	0.66	0.05	0.65	0.05	0.73	0.07
678736	2	12	4114	1848	-0.40	0.06	-0.18	0.05	-0.10	0.28
640376	5	-1	1929	1848	0.81	0.02	0.68	0.02	0.76	-0.07
681303	5	10	1725	1848	2.43	0.04	2.09	0.04	2.17	-0.12
674404	5	11	1678	1848	1.13	0.03	0.96	0.03	1.04	-0.14
633313	5	-1	1929	1848	1.56	0.03	1.44	0.03	1.52	-0.04
666557	5	10	1675	1848	1.65	0.03	1.95	0.04	2.03	0.37
672275	5	11	1672	1848	1.19	0.03	1.28	0.03	1.36	0.24

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
607724	2	-12	1624	1116	0.07	0.05	0.23	0.06	0.31	0.27
671335	2	-17	3986	1116	0.16	0.06	0.02	0.06	0.09	-0.04
674117	2	-15	4019	1116	-0.79	0.06	-0.92	0.07	-0.85	-0.03
642329	2	-12	8058	1116	0.60	0.05	0.77	0.07	0.85	0.27
678418	2	-12	5872	1116	0.50	0.05	0.08	0.06	0.15	-0.32
642837	2	-10	5341	1116	0.34	0.05	0.16	0.06	0.24	-0.08
611014	2	-7	1713	1116	0.15	0.05	0.19	0.06	0.27	0.14
635759	2	-7	1632	1116	0.21	0.05	0.07	0.06	0.15	-0.04
674139	2	-11	3986	1116	-0.35	0.06	-0.77	0.07	-0.69	-0.32
610887	2	-4	1748	1116	0.28	0.05	0.16	0.06	0.23	-0.02
679666	2	-2	5872	1116	-0.67	0.06	-0.43	0.06	-0.35	0.34
679665	2	-1	4039	1116	-0.01	0.05	-0.32	0.06	-0.25	-0.21
678539	2	4	5798	1116	0.69	0.06	0.91	0.07	0.98	0.32
607702	2	2	1613	1116	-0.45	0.05	-0.34	0.06	-0.26	0.22
607107	2	10	1624	1116	-0.08	0.05	-0.12	0.06	-0.04	0.07
607735	2	11	1611	1116	0.37	0.05	0.44	0.06	0.52	0.17
671328	2	12	5872	1116	-0.58	0.06	-0.34	0.06	-0.26	0.35
677984	2	11	5275	1116	0.96	0.05	1.29	0.07	1.37	0.43
611387	2	14	1753	1116	0.28	0.05	-0.76	0.07	-0.69	-0.95
609821	2	17	1753	1116	-0.28	0.05	-0.45	0.06	-0.37	-0.07
611384	2	17	1734	1116	-0.12	0.05	-0.01	0.06	0.06	0.20
635764	2	16	1717	1116	-0.04	0.05	0.15	0.06	0.23	0.30
610459	2	13	1734	1116	0.03	0.05	-0.24	0.06	-0.16	-0.16
607571	2	13	1619	1116	0.96	0.06	0.92	0.07	1.00	0.06
611398	2	-12	1619	1116	0.59	0.05	0.80	0.07	0.88	0.31
679982	2	-14	5842	1116	0.56	0.05	0.71	0.07	0.79	0.25
678303	2	-16	3986	1116	1.03	0.06	0.76	0.07	0.84	-0.16
611926	2	-11	1613	1116	0.07	0.05	0.26	0.06	0.34	0.29
612587	2	-17	1761	1116	-0.09	0.05	-0.43	0.06	-0.35	-0.24
682252	2	-9	4022	1116	0.12	0.05	-0.11	0.06	-0.04	-0.13
641211	2	-11	5872	1116	0.97	0.05	0.55	0.06	0.63	-0.31
677882	2	-11	5843	1116	0.32	0.05	0.31	0.06	0.39	0.09
611507	2	-13	1598	1116	0.23	0.05	0.05	0.06	0.12	-0.09
683582	2	-4	5786	1116	-0.75	0.06	-0.69	0.07	-0.61	0.16
682196	2	-3	4025	1116	-0.34	0.06	-0.36	0.06	-0.28	0.09
682198	2	-1	4022	1116	-1.08	0.06	-1.06	0.07	-0.98	0.12
611921	2	4	1602	1116	-1.02	0.06	-1.16	0.07	-1.08	-0.04
643346	2	5	5842	1116	0.31	0.05	0.09	0.06	0.17	-0.12
611012	2	-1	46979	1116	0.09	0.01	-0.24	0.06	-0.16	-0.23
611925	2	13	1744	1116	-0.31	0.05	-0.44	0.06	-0.36	-0.03
681522	2	7	3990	1116	0.83	0.05	0.58	0.07	0.66	-0.14

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
611465	2	15	1624	1116	0.92	0.06	0.89	0.07	0.97	0.08
608357	2	14	1717	1116	-0.03	0.05	-0.53	0.07	-0.45	-0.41
607754	2	17	1735	1116	0.67	0.05	0.28	0.06	0.36	-0.29
642902	2	12	5341	1116	-0.57	0.05	-0.39	0.06	-0.31	0.29
611053	2	11	1598	1116	-0.45	0.06	-0.49	0.07	-0.42	0.06
679258	2	18	5786	1116	-0.10	0.05	0.26	0.06	0.34	0.46
608358	2	21	1632	1116	0.14	0.05	-0.37	0.06	-0.30	-0.42
677887	4	-1	1763	1116	0.91	0.03	0.76	0.04	0.83	-0.03
678932	4	13	1726	1116	0.50	0.03	0.15	0.04	0.23	-0.31
682984	4	14	1725	1116	1.42	0.03	1.31	0.04	1.39	-0.28
682669	4	-1	1761	1116	0.14	0.03	0.08	0.04	0.16	0.15
677890	4	13	1744	1116	0.98	0.03	1.26	0.05	1.34	0.26
678996	4	14	1772	1116	0.91	0.03	1.05	0.04	1.13	0.17

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	2	-10	4744	589	0.61	0.05	0.12	0.09	0.22	-0.34
641544	2	-9	4714	589	0.79	0.05	0.55	0.09	0.64	-0.08
641564	2	-10	4744	589	-0.44	0.06	-0.43	0.09	-0.34	0.13
641561	2	-9	4714	589	0.86	0.05	0.31	0.09	0.41	-0.38
641547	2	-9	4714	589	-0.64	0.06	-1.07	0.10	-0.97	-0.34
641562	2	-8	4744	589	0.86	0.05	0.57	0.09	0.66	-0.12
641542	2	-8	4714	589	-0.35	0.06	-0.52	0.09	-0.42	-0.04
641543	2	-8	4744	589	-0.18	0.06	-0.22	0.09	-0.13	0.10
641565	2	-7	4714	589	0.66	0.05	0.31	0.09	0.41	-0.19
616676	2	8	1696	589	-0.51	0.06	-0.58	0.09	-0.49	0.05
616683	2	9	1721	589	-0.87	0.06	-0.84	0.10	-0.75	0.14
616677	2	6	1721	589	0.22	0.05	0.06	0.09	0.16	-0.01
616671	2	10	1696	589	0.60	0.05	0.44	0.09	0.54	0.01
616678	2	7	1721	589	-0.71	0.06	-0.29	0.09	-0.20	0.56
616675	2	11	1721	589	-0.22	0.06	-0.27	0.09	-0.17	0.09
616670	2	11	1696	589	-0.44	0.06	0.39	0.09	0.48	0.99
616681	2	10	1696	589	0.58	0.05	0.40	0.09	0.49	-0.02
608943	2	-11	1696	589	0.41	0.05	0.04	0.09	0.13	-0.22
608939	2	-12	1696	589	0.25	0.06	-0.23	0.09	-0.13	-0.34
616846	2	-10	1696	589	0.68	0.05	0.64	0.09	0.73	0.13
608941	2	-11	1696	589	0.62	0.05	0.05	0.09	0.15	-0.41
640221	2	-12	1696	589	0.82	0.05	0.52	0.09	0.62	-0.13
640205	2	-8	1721	589	0.76	0.05	0.32	0.09	0.42	-0.27
608948	2	-9	1696	589	0.46	0.05	0.03	0.09	0.13	-0.28
608942	2	-9	1721	589	1.29	0.06	1.29	0.10	1.38	0.19
673032	2	6	5623	589	-0.63	0.07	-0.67	0.10	-0.58	0.08

Item ID	Item Type	Sequence Change	Old *N*	New *N*	Old Logit	Old SEM	New Logit	New SEM	Equated Logit	Displacement
673031	2	6	5623	589	0.58	0.05	0.60	0.09	0.69	0.19
683404	2	6	5623	589	-1.52	0.08	-1.43	0.11	-1.34	0.17
683419	2	7	5545	589	-0.10	0.06	0.02	0.09	0.11	0.26
673030	2	7	5623	589	0.04	0.06	-0.08	0.09	0.01	0.03
673023	2	7	5545	589	1.20	0.05	0.96	0.09	1.06	-0.05
673024	2	8	5623	589	-0.76	0.07	-0.58	0.09	-0.49	0.30
673025	2	8	5623	589	0.50	0.05	-0.07	0.09	0.03	-0.42
673026	2	9	5545	589	-0.29	0.06	-0.34	0.09	-0.24	0.08
643960	4	-7	1855	589	0.68	0.03	0.44	0.06	0.53	-0.13
616673	4	10	1720	589	1.26	0.03	1.20	0.06	1.29	-0.01
616672	4	11	1696	589	0.97	0.03	1.31	0.06	1.40	0.21
613250	4	-9	1696	589	1.66	0.04	0.89	0.06	0.98	-0.55
683411	4	9	1705	589	1.32	0.03	1.77	0.06	1.87	0.17
683414	4	10	1644	589	1.18	0.04	1.55	0.06	1.64	0.44

FORM LEVEL

Table L–10. Raw-to-Scaled Score Comparison for Algebra I: Winter

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1206	92	1203	92	1203	92
1	1268	51	1265	51	1265	51
2	1305	37	1302	37	1303	37
3	1327	31	1325	31	1326	31
4	1344	27	1342	27	1343	27
5	1357	25	1356	25	1356	25
6	1369	23	1367	23	1368	23
7	1379	22	1377	22	1378	22
8	1387	20	1386	21	1387	21
9	1395	20	1394	20	1395	20
10	1403	19	1401	19	1402	19
11	1410	18	1408	18	1409	18
12	1416	18	1415	18	1415	18
13	1422	17	1421	17	1421	17
14	1428	17	1427	17	1427	17
15	1433	16	1432	16	1433	16
16	1439	16	1437	16	1438	16
17	1444	16	1443	16	1443	16
18	1449	16	1447	15	1448	15
19	1454	15	1452	15	1453	15
20	1458	15	1457	15	1457	15
21	1463	15	1461	15	1462	15
22	1467	15	1466	15	1466	15
23	1471	15	1470	14	1470	14
24	1476	14	1474	14	1474	14
25	1480	14	1478	14	1479	14
26	1484	14	1482	14	1483	14
27	1488	14	1486	14	1487	14
28	1492	14	1490	14	1491	14
29	1496	14	1494	14	1495	14
30	1500	14	1498	14	1499	14
31	1504	14	1502	14	1502	14
32	1507	14	1506	14	1506	14
33	1511	14	1510	14	1510	14
34	1515	14	1514	14	1514	14
35	1519	14	1518	14	1519	14
36	1523	14	1522	14	1523	14

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
37	1527	14	1526	15	1527	15
38	1532	14	1531	15	1531	15
39	1536	15	1535	15	1536	15
40	1540	15	1540	15	1540	15
41	1544	15	1544	15	1545	15
42	1549	15	1549	16	1550	16
43	1554	15	1554	16	1554	16
44	1558	16	1559	16	1559	16
45	1563	16	1564	16	1565	16
46	1569	16	1570	17	1570	17
47	1574	17	1575	17	1576	17
48	1580	17	1581	17	1582	17
49	1586	18	1587	18	1588	18
50	1592	18	1594	19	1595	19
51	1599	19	1601	19	1602	19
52	1606	20	1609	20	1610	20
53	1614	21	1618	21	1618	21
54	1623	22	1627	23	1628	23
55	1634	24	1639	25	1639	25
56	1646	26	1652	27	1652	27
57	1662	30	1668	31	1669	31
58	1683	36	1691	37	1691	37
59	1719	51	1728	51	1728	51
60	1780	92	1790	92	1790	92

Table L–11. Raw-to-Scaled Score Comparison for Biology: Winter

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1204	92	1200	92	1200	92
1	1265	51	1257	51	1256	51
2	1301	36	1293	37	1293	37
3	1323	30	1315	30	1315	30
4	1339	26	1331	27	1331	27
5	1351	24	1344	24	1343	24
6	1362	22	1355	22	1354	22
7	1371	21	1364	21	1363	21
8	1379	20	1372	20	1372	20
9	1386	19	1380	19	1379	19
10	1393	18	1387	18	1386	18
11	1399	17	1393	18	1392	18

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
12	1405	17	1399	17	1398	17
13	1410	16	1404	16	1404	16
14	1415	16	1410	16	1409	16
15	1420	16	1415	16	1414	16
16	1425	15	1420	15	1419	15
17	1430	15	1424	15	1424	15
18	1434	15	1429	15	1428	15
19	1438	14	1433	15	1433	15
20	1442	14	1437	15	1437	15
21	1446	14	1442	14	1441	14
22	1450	14	1446	14	1445	14
23	1454	14	1450	14	1449	14
24	1458	14	1454	14	1453	14
25	1462	14	1457	14	1457	14
26	1465	14	1461	14	1461	14
27	1469	13	1465	14	1465	14
28	1473	13	1469	14	1468	14
29	1476	13	1473	14	1472	14
30	1480	13	1476	14	1476	14
31	1483	13	1480	14	1480	14
32	1487	13	1484	14	1483	14
33	1490	13	1488	14	1487	14
34	1494	13	1491	14	1491	14
35	1497	13	1495	14	1495	14
36	1501	13	1499	14	1498	14
37	1505	13	1503	14	1502	14
38	1508	13	1507	14	1506	14
39	1512	13	1511	14	1510	14
40	1515	14	1514	14	1514	14
41	1519	14	1519	14	1518	14
42	1523	14	1523	14	1522	14
43	1527	14	1527	15	1526	15
44	1530	14	1531	15	1531	15
45	1534	14	1535	15	1535	15
46	1538	14	1540	15	1539	15
47	1543	14	1545	15	1544	15
48	1547	15	1549	16	1549	16
49	1551	15	1554	16	1554	16
50	1556	15	1559	16	1559	16
51	1560	15	1565	16	1564	16
52	1565	16	1570	17	1570	17

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
53	1570	16	1576	17	1575	17
54	1576	17	1582	18	1581	18
55	1581	17	1589	18	1588	18
56	1587	18	1596	19	1595	19
57	1594	19	1603	20	1602	20
58	1601	19	1611	21	1611	21
59	1609	21	1620	22	1620	22
60	1618	22	1631	23	1630	23
61	1628	24	1642	25	1642	25
62	1641	26	1656	28	1656	28
63	1657	30	1674	32	1673	32
64	1678	36	1698	38	1697	38
65	1715	51	1737	52	1736	52
66	1776	92	1800	93	1800	93

Table L–12. Raw-to-Scaled Score Comparison for Literature: Winter

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1207	92	1211	92	1209	92
1	1268	51	1272	51	1270	51
2	1304	37	1308	36	1306	36
3	1327	30	1330	30	1328	30
4	1343	27	1346	26	1344	26
5	1356	24	1359	24	1357	24
6	1367	23	1369	22	1367	22
7	1376	21	1378	21	1377	21
8	1385	20	1387	20	1385	20
9	1392	19	1394	19	1392	19
10	1400	19	1401	18	1399	18
11	1406	18	1408	18	1406	18
12	1413	18	1414	17	1412	17
13	1418	17	1419	17	1418	17
14	1424	17	1425	16	1423	16
15	1430	16	1430	16	1428	16
16	1435	16	1435	16	1433	16
17	1440	16	1440	16	1438	16
18	1445	16	1445	15	1443	15
19	1450	16	1450	15	1448	15
20	1455	16	1455	15	1453	15
21	1460	15	1459	15	1457	15

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
22	1465	15	1464	15	1462	15
23	1469	15	1468	15	1466	15
24	1474	15	1473	15	1471	15
25	1479	15	1477	15	1475	15
26	1483	15	1482	15	1480	15
27	1488	15	1486	15	1485	15
28	1493	15	1491	15	1489	15
29	1497	15	1496	15	1494	15
30	1502	16	1500	15	1499	15
31	1507	16	1505	16	1503	16
32	1512	16	1510	16	1508	16
33	1517	16	1515	16	1513	16
34	1522	16	1520	16	1518	16
35	1528	17	1526	17	1524	17
36	1533	17	1531	17	1529	17
37	1539	17	1537	17	1535	17
38	1545	18	1543	18	1541	18
39	1551	18	1549	18	1548	18
40	1558	18	1556	19	1554	19
41	1565	19	1563	19	1561	19
42	1573	20	1571	20	1569	20
43	1581	20	1579	21	1577	21
44	1589	21	1588	22	1586	22
45	1599	23	1598	23	1596	23
46	1610	24	1609	24	1607	24
47	1622	26	1622	26	1620	26
48	1636	28	1637	29	1635	29
49	1654	32	1655	32	1653	32
50	1678	38	1679	38	1677	38
51	1716	52	1718	52	1716	52
52	1779	92	1781	92	1779	92

Table L–13. Raw-to-Scaled Score Comparison for Algebra I: Spring

RS	Preequating		Post-equating	
	SS	SEM	SS	SEM
0	1205	92	1207	92
1	1266	51	1268	51
2	1303	37	1305	37
3	1325	30	1327	30
4	1341	27	1343	27
5	1354	24	1356	24
6	1365	23	1367	23
7	1375	21	1377	21
8	1383	20	1385	20
9	1391	19	1393	19
10	1398	19	1400	19
11	1405	18	1407	18
12	1412	18	1413	17
13	1417	17	1419	17
14	1423	17	1424	17
15	1429	16	1430	16
16	1434	16	1435	16
17	1439	16	1440	16
18	1444	16	1445	15
19	1449	15	1449	15
20	1453	15	1454	15
21	1458	15	1458	15
22	1462	15	1463	15
23	1466	15	1467	14
24	1471	14	1471	14
25	1475	14	1475	14
26	1479	14	1479	14
27	1483	14	1483	14
28	1487	14	1487	14
29	1491	14	1491	14
30	1494	14	1494	14
31	1498	14	1498	14
32	1502	14	1502	14
33	1506	14	1506	14
34	1509	14	1509	14
35	1513	14	1513	14
36	1517	14	1517	14
37	1520	14	1520	14
38	1524	14	1524	14
39	1528	14	1528	14
40	1532	14	1532	14

RS	Preequating		Post-equating	
	SS	SEM	SS	SEM
41	1536	14	1536	14
42	1540	14	1540	14
43	1544	15	1544	15
44	1549	15	1548	15
45	1553	15	1553	15
46	1558	16	1558	16
47	1563	16	1563	16
48	1569	17	1568	16
49	1574	17	1573	17
50	1581	18	1580	18
51	1587	19	1586	19
52	1595	20	1594	20
53	1604	21	1602	21
54	1613	23	1611	22
55	1625	25	1622	24
56	1638	27	1635	27
57	1655	31	1651	30
58	1677	36	1673	36
59	1713	50	1708	50
60	1773	91	1768	91

Table L–14. Raw-to-Scaled Score Comparison for Biology: Spring

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1212	92	1210	92	1209	92
1	1273	51	1271	51	1270	51
2	1309	36	1307	36	1305	36
3	1330	30	1328	30	1327	30
4	1345	26	1344	26	1342	26
5	1358	24	1356	24	1355	24
6	1368	22	1366	22	1365	22
7	1377	20	1375	20	1374	20
8	1385	19	1383	19	1382	19
9	1392	18	1390	18	1389	18
10	1398	18	1397	18	1395	18
11	1404	17	1403	17	1401	17
12	1410	16	1408	16	1407	16
13	1415	16	1414	16	1412	16
14	1420	16	1419	16	1417	16
15	1425	15	1424	15	1422	15
16	1429	15	1428	15	1427	15

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
17	1433	15	1432	15	1431	15
18	1438	14	1437	14	1435	14
19	1442	14	1441	14	1439	14
20	1446	14	1445	14	1443	14
21	1449	14	1449	14	1447	14
22	1453	14	1452	14	1451	14
23	1457	13	1456	14	1455	14
24	1460	13	1460	13	1458	13
25	1464	13	1463	13	1462	13
26	1467	13	1467	13	1465	13
27	1471	13	1470	13	1469	13
28	1474	13	1474	13	1472	13
29	1478	13	1477	13	1476	13
30	1481	13	1481	13	1479	13
31	1484	13	1484	13	1483	13
32	1488	13	1488	13	1486	13
33	1491	13	1491	13	1489	13
34	1494	13	1494	13	1493	13
35	1498	13	1498	13	1496	13
36	1501	13	1501	13	1500	13
37	1504	13	1505	13	1503	13
38	1508	13	1508	13	1507	13
39	1511	13	1511	13	1510	13
40	1515	13	1515	13	1514	13
41	1518	13	1519	13	1517	13
42	1522	13	1522	14	1521	14
43	1525	14	1526	14	1524	14
44	1529	14	1530	14	1528	14
45	1533	14	1534	14	1532	14
46	1537	14	1537	14	1536	14
47	1541	14	1542	14	1540	14
48	1545	15	1546	15	1544	15
49	1549	15	1550	15	1549	15
50	1554	15	1555	15	1553	15
51	1558	15	1559	15	1558	15
52	1563	16	1564	16	1563	16
53	1568	16	1569	16	1568	16
54	1574	17	1575	17	1573	17
55	1580	17	1581	17	1579	17
56	1586	18	1587	18	1585	18
57	1593	19	1594	19	1592	19

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
58	1600	20	1601	20	1600	20
59	1608	21	1610	21	1608	21
60	1617	22	1619	22	1617	22
61	1628	24	1630	24	1628	24
62	1641	27	1643	27	1641	27
63	1657	30	1659	31	1658	31
64	1679	37	1682	37	1680	37
65	1716	51	1719	51	1718	51
66	1777	92	1781	92	1780	92

Table L–15. Raw-to-Scaled Score Comparison for Literature: Spring

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1208	92	1209	92	1211	92
1	1269	51	1270	51	1272	51
2	1306	36	1306	36	1308	36
3	1327	30	1328	30	1330	30
4	1343	26	1343	26	1345	26
5	1356	24	1356	24	1358	24
6	1367	22	1366	22	1369	22
7	1376	21	1376	21	1378	21
8	1384	20	1384	20	1386	20
9	1392	19	1391	19	1393	19
10	1399	18	1398	18	1400	18
11	1405	18	1404	18	1406	18
12	1411	17	1410	17	1412	17
13	1417	17	1416	17	1418	17
14	1422	16	1421	16	1423	16
15	1428	16	1426	16	1429	16
16	1433	16	1431	16	1434	16
17	1438	16	1436	15	1438	15
18	1442	15	1441	15	1443	15
19	1447	15	1446	15	1448	15
20	1452	15	1450	15	1452	15
21	1456	15	1455	15	1457	15
22	1461	15	1459	15	1461	15
23	1465	15	1463	15	1465	15
24	1469	15	1468	15	1470	15
25	1474	15	1472	15	1474	15
26	1478	15	1477	15	1479	15

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
27	1483	15	1481	15	1483	15
28	1487	15	1485	15	1487	15
29	1491	15	1490	15	1492	15
30	1496	15	1494	15	1497	15
31	1501	15	1499	15	1501	15
32	1505	15	1504	15	1506	15
33	1510	16	1509	16	1511	16
34	1515	16	1514	16	1516	16
35	1520	16	1519	16	1521	16
36	1525	16	1524	17	1526	17
37	1531	17	1530	17	1532	17
38	1536	17	1536	17	1538	17
39	1542	17	1542	18	1544	18
40	1549	18	1548	18	1550	18
41	1555	18	1555	19	1557	19
42	1562	19	1563	20	1565	20
43	1570	20	1571	20	1573	20
44	1578	21	1580	21	1582	21
45	1587	22	1589	23	1591	23
46	1597	23	1600	24	1602	24
47	1609	25	1612	26	1615	26
48	1623	27	1627	28	1629	28
49	1640	31	1645	32	1647	32
50	1663	37	1669	38	1671	38
51	1700	51	1707	52	1709	52
52	1762	92	1770	92	1772	92

Table L–16. Raw-to-Scaled Score Comparison for Algebra I: Summer

RS	Preequating		Post-equating	
	SS	SEM	SS	SEM
0	1209	92	1207	92
1	1271	51	1269	51
2	1307	37	1306	37
3	1329	30	1328	31
4	1345	27	1344	27
5	1358	24	1358	24
6	1369	22	1369	23
7	1379	21	1378	21
8	1387	20	1387	20
9	1395	19	1395	19
10	1402	19	1402	19

RS	Preequating		Post-equating	
	SS	SEM	SS	SEM
11	1409	18	1408	18
12	1415	17	1415	17
13	1421	17	1421	17
14	1426	17	1426	17
15	1432	16	1432	16
16	1437	16	1437	16
17	1442	16	1442	16
18	1447	16	1447	16
19	1452	15	1452	15
20	1456	15	1457	15
21	1461	15	1461	15
22	1465	15	1466	15
23	1470	15	1470	15
24	1474	15	1475	15
25	1478	15	1479	15
26	1483	15	1483	15
27	1487	14	1487	14
28	1491	14	1492	14
29	1495	14	1496	14
30	1499	14	1500	14
31	1503	14	1504	14
32	1507	14	1508	14
33	1511	14	1512	14
34	1515	14	1516	14
35	1519	14	1520	14
36	1523	14	1524	14
37	1527	14	1528	14
38	1531	14	1532	14
39	1535	14	1536	14
40	1540	14	1541	14
41	1544	14	1545	15
42	1548	15	1549	15
43	1552	15	1553	15
44	1556	15	1558	15
45	1561	15	1563	15
46	1566	15	1568	16
47	1571	16	1573	16
48	1576	16	1578	17
49	1581	17	1584	17
50	1587	18	1590	18
51	1594	19	1596	18
52	1601	20	1603	19

RS	Preequating		Post-equating	
	SS	SEM	SS	SEM
53	1609	21	1611	20
54	1619	23	1620	22
55	1630	25	1630	23
56	1644	27	1641	25
57	1660	31	1656	28
58	1682	36	1675	34
59	1715	48	1705	47
60	1770	88	1760	88

Table L–17. Raw-to-Scaled Score Comparison for Biology: Summer

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1215	92	1212	92	1211	92
1	1276	51	1273	51	1273	51
2	1311	36	1309	36	1308	36
3	1333	30	1331	30	1330	30
4	1348	26	1346	26	1345	26
5	1361	24	1358	24	1358	24
6	1371	22	1369	22	1368	22
7	1380	20	1378	20	1377	20
8	1388	19	1386	19	1385	19
9	1395	18	1393	18	1392	18
10	1401	18	1399	18	1398	18
11	1407	17	1405	17	1404	17
12	1413	16	1411	16	1410	16
13	1418	16	1416	16	1415	16
14	1423	16	1421	16	1420	16
15	1428	15	1426	15	1425	15
16	1432	15	1430	15	1429	15
17	1437	15	1434	15	1434	15
18	1441	14	1439	14	1438	14
19	1445	14	1443	14	1442	14
20	1449	14	1446	14	1446	14
21	1452	14	1450	14	1450	14
22	1456	13	1454	14	1453	14
23	1460	13	1458	13	1457	13
24	1463	13	1461	13	1460	13
25	1467	13	1465	13	1464	13
26	1470	13	1468	13	1467	13
27	1473	13	1472	13	1471	13
28	1477	13	1475	13	1474	13

RS	Preequating		Post-equating			
			All Items		Outliers Removed	
	SS	SEM	SS	SEM	SS	SEM
29	1480	13	1478	13	1478	13
30	1483	13	1482	13	1481	13
31	1487	13	1485	13	1484	13
32	1490	13	1488	13	1488	13
33	1493	13	1492	13	1491	13
34	1496	13	1495	13	1494	13
35	1500	13	1498	13	1498	13
36	1503	13	1502	13	1501	13
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	1523	14	1522	14
43	1527	13	1526	14	1526	14
44	1530	14	1530	14	1529	14
45	1534	14	1534	14	1533	14
46	1538	14	1538	14	1537	14
47	1542	14	1542	14	1541	14
48	1546	14	1546	15	1546	15
49	1550	15	1551	15	1550	15
50	1554	15	1556	15	1555	15
51	1559	15	1560	16	1560	16
52	1564	16	1565	16	1565	16
53	1569	16	1571	17	1570	17
54	1574	17	1576	17	1576	17
55	1580	17	1583	18	1582	18
56	1586	18	1589	18	1588	18
57	1593	19	1596	19	1595	19
58	1600	20	1604	20	1603	20
59	1608	21	1612	21	1612	21
60	1617	22	1622	23	1621	23
61	1628	24	1634	25	1633	25
62	1641	27	1647	28	1646	28
63	1657	31	1664	31	1664	31
64	1680	37	1688	38	1687	38
65	1717	51	1727	53	1727	53
66	1779	92	1792	93	1791	93

Table L–18. Raw-to-Scaled Score Comparison for Literature: Summer

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
0	1208	92	1207	92	1206	92
1	1269	51	1268	51	1267	51
2	1306	37	1304	36	1303	36
3	1328	30	1326	30	1325	30
4	1344	27	1342	27	1341	27
5	1357	24	1355	24	1354	24
6	1368	22	1366	22	1365	22
7	1377	21	1375	21	1374	21
8	1385	20	1383	20	1382	20
9	1393	19	1391	19	1390	19
10	1400	19	1398	18	1397	18
11	1407	18	1405	18	1403	18
12	1413	17	1411	17	1410	17
13	1419	17	1417	17	1415	17
14	1425	17	1422	17	1421	17
15	1430	16	1428	16	1426	16
16	1436	16	1433	16	1432	16
17	1441	16	1438	16	1437	16
18	1446	16	1443	16	1442	16
19	1451	16	1448	15	1446	15
20	1455	15	1452	15	1451	15
21	1460	15	1457	15	1456	15
22	1465	15	1462	15	1461	15
23	1470	15	1466	15	1465	15
24	1474	15	1471	15	1470	15
25	1479	15	1476	15	1474	15
26	1483	15	1480	15	1479	15
27	1488	15	1485	15	1484	15
28	1493	15	1490	15	1489	15
29	1498	15	1494	16	1493	16
30	1502	16	1499	16	1498	16
31	1507	16	1504	16	1503	16
32	1512	16	1509	16	1508	16
33	1517	16	1515	16	1513	16
34	1522	16	1520	17	1519	17
35	1528	17	1526	17	1524	17
36	1533	17	1531	17	1530	17
37	1539	17	1537	18	1536	18
38	1545	18	1544	18	1543	18

RS	Preequating		Post-equating			
			All Items		Outlier Removed	
	SS	SEM	SS	SEM	SS	SEM
39	1551	18	1550	19	1549	19
40	1558	18	1557	19	1556	19
41	1565	19	1565	20	1564	20
42	1573	20	1573	21	1572	21
43	1581	21	1582	22	1581	22
44	1590	21	1592	23	1591	23
45	1599	23	1603	24	1602	24
46	1610	24	1616	26	1615	26
47	1622	26	1630	28	1629	28
48	1637	28	1647	31	1646	31
49	1655	32	1668	34	1667	34
50	1679	38	1696	40	1695	40
51	1717	52	1739	54	1737	54
52	1780	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 less than 5.

Table M–2. Winter: Algebra I Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	75936	23.25	9.66	0.87	3.47
	Module 1	All	30	21	75936	11.83	5.10	0.78	2.41
	Module 2	All	30	21	75936	11.42	5.18	0.77	2.50
Gender	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	Female	60	42	36393	23.62	9.39	0.86	3.47
		Male	60	42	39385	22.92	9.88	0.88	3.47
	Module 1	Female	30	21	36393	12.05	5.01	0.77	2.42
		Male	30	21	39385	11.64	5.17	0.78	2.40
	Module 2	Female	30	21	36393	11.57	5.00	0.75	2.49
		Male	30	21	39385	11.28	5.33	0.78	2.50
Ethnicity	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	African American	60	42	16205	19.05	7.96	0.83	3.31
		American Indian	60	42	117	24.31	11.10	0.90	3.51
		Asian	60	42	1913	27.53	12.60	0.91	3.68
		Hispanic	60	42	8717	19.93	8.54	0.85	3.34
		Multi-racial	60	42	1482	22.64	9.14	0.86	3.45
		Native Hawaiian/ Pacific Islander	60	42	51	25.80	9.08	0.85	3.48
		White	60	42	47268	25.16	9.60	0.87	3.52
	Module 1	African American	30	21	16205	9.67	4.33	0.72	2.30
		American Indian	30	21	117	12.67	5.98	0.84	2.43
		Asian	30	21	1913	13.98	6.64	0.85	2.57
		Hispanic	30	21	8717	10.20	4.61	0.75	2.31
		Multi-racial	30	21	1482	11.49	4.88	0.76	2.39
		Native Hawaiian/ Pacific Islander	30	21	51	12.96	4.54	0.71	2.46
		White	30	21	47268	12.80	5.04	0.77	2.44
	Module 2	African American	30	21	16205	9.38	4.34	0.70	2.38
		American Indian	30	21	117	11.64	5.67	0.80	2.53
		Asian	30	21	1913	13.55	6.44	0.83	2.63
		Hispanic	30	21	8717	9.74	4.58	0.72	2.40
		Multi-racial	30	21	1482	11.15	4.92	0.74	2.49
		Native Hawaiian/ Pacific Islander	30	21	51	12.84	5.01	0.76	2.48
		White	30	21	47268	12.35	5.20	0.76	2.53
ELL	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	2987	16.62	7.77	0.83	3.21
	Module 1	All	30	21	2987	8.54	4.40	0.73	2.27
	Module 2	All	30	21	2987	8.09	4.02	0.68	2.27
IEP	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	14549	17.60	7.69	0.82	3.24
	Module 1	All	30	21	14549	9.05	4.22	0.71	2.26
	Module 2	All	30	21	14549	8.55	4.17	0.69	2.32
ED	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	38604	20.72	8.50	0.84	3.37
	Module 1	All	30	21	38604	10.57	4.57	0.74	2.34
	Module 2	All	30	21	38604	10.15	4.61	0.72	2.43

Table M–3. Winter: Biology Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	54236	27.52	11.27	0.90	3.52
	Module 1	All	33	27	54236	13.58	5.77	0.81	2.51
	Module 2	All	33	27	54236	13.94	6.11	0.84	2.46
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	66	54	26559	27.81	10.89	0.90	3.53
		Male	66	54	27562	27.27	11.61	0.91	3.51
	Module 1	Female	33	27	26559	13.65	5.58	0.80	2.50
		Male	33	27	27562	13.53	5.94	0.82	2.51
	Module 2	Female	33	27	26559	14.16	5.95	0.83	2.48
		Male	33	27	27562	13.74	6.27	0.85	2.45
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	66	54	11430	22.25	8.59	0.84	3.45
		American Indian	66	54	83	28.27	11.44	0.90	3.57
		Asian	66	54	1574	30.25	13.79	0.93	3.52
		Hispanic	66	54	5533	23.64	9.60	0.87	3.48
		Multi-racial	66	54	1032	26.99	10.88	0.90	3.52
		Native Hawaiian/ Pacific Islander	66	54	35	29.23	10.56	0.89	3.57
		White	66	54	34420	29.80	11.42	0.90	3.54
	Module 1	African American	33	27	11430	11.15	4.51	0.70	2.48
		American Indian	33	27	83	13.90	6.11	0.83	2.53
		Asian	33	27	1574	15.07	7.00	0.87	2.48
		Hispanic	33	27	5533	11.75	4.97	0.75	2.49
		Multi-racial	33	27	1032	13.44	5.58	0.80	2.51
		Native Hawaiian/ Pacific Islander	33	27	35	14.49	5.45	0.79	2.51
		White	33	27	34420	14.63	5.88	0.82	2.51
	Module 2	African American	33	27	11430	11.10	4.84	0.76	2.39
		American Indian	33	27	83	14.36	5.89	0.82	2.52
		Asian	33	27	1574	15.17	7.29	0.88	2.49
		Hispanic	33	27	5533	11.89	5.32	0.79	2.43
		Multi-racial	33	27	1032	13.55	5.97	0.83	2.46
		Native Hawaiian/ Pacific Islander	33	27	35	14.74	5.74	0.81	2.52
		White	33	27	34420	15.17	6.15	0.84	2.48
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	1829	18.33	7.01	0.77	3.36
	Module 1	All	33	27	1829	9.45	3.92	0.62	2.42
	Module 2	All	33	27	1829	8.88	3.93	0.65	2.33
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	10724	21.69	8.51	0.84	3.45
	Module 1	All	33	27	10724	10.86	4.47	0.69	2.48
	Module 2	All	33	27	10724	10.83	4.80	0.75	2.40
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	26601	24.13	9.30	0.86	3.49
	Module 1	All	33	27	26601	11.98	4.83	0.73	2.50
	Module 2	All	33	27	26601	12.15	5.20	0.78	2.43

Table M–4. Winter: Literature Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	47149	25.72	9.77	0.89	3.19
	Module 1	All	26	20	47149	12.83	5.14	0.81	2.27
	Module 2	All	26	20	47149	12.88	5.20	0.81	2.24
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	52	40	20445	27.47	9.55	0.89	3.18
		Male	52	40	26574	24.39	9.72	0.89	3.17
	Module 1	Female	26	20	20445	13.73	5.03	0.80	2.27
		Male	26	20	26574	12.16	5.12	0.81	2.25
	Module 2	Female	26	20	20445	13.74	5.10	0.81	2.23
		Male	26	20	26574	12.23	5.18	0.82	2.22
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	52	40	9449	21.70	8.74	0.86	3.23
		American Indian	52	40	58	25.53	11.32	0.93	3.07
		Asian	52	40	1164	27.25	10.54	0.91	3.22
		Hispanic	52	40	5639	22.69	9.14	0.88	3.22
		Multi-racial	52	40	832	25.08	9.57	0.89	3.21
		Native Hawaiian/ Pacific Islander	52	40	30	25.20	9.70	0.89	3.18
		White	52	40	29837	27.55	9.63	0.89	3.15
	Module 1	African American	26	20	9449	10.95	4.71	0.76	2.29
		American Indian	26	20	58	12.98	5.81	0.86	2.16
		Asian	26	20	1164	13.35	5.54	0.83	2.29
		Hispanic	26	20	5639	11.37	4.86	0.78	2.29
		Multi-racial	26	20	832	12.67	5.12	0.80	2.28
		Native Hawaiian/ Pacific Islander	26	20	30	12.30	5.46	0.83	2.22
		White	26	20	29837	13.71	5.08	0.81	2.24
	Module 2	African American	26	20	9449	10.75	4.73	0.77	2.27
		American Indian	26	20	58	12.55	6.01	0.87	2.17
		Asian	26	20	1164	13.90	5.50	0.83	2.25
		Hispanic	26	20	5639	11.31	4.90	0.79	2.26
		Multi-racial	26	20	832	12.41	5.06	0.80	2.25
		Native Hawaiian/ Pacific Islander	26	20	30	12.90	4.69	0.76	2.27
		White	26	20	29837	13.84	5.11	0.81	2.21
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	1665	16.51	6.71	0.78	3.16
	Module 1	All	26	20	1665	8.21	3.66	0.63	2.23
	Module 2	All	26	20	1665	8.30	3.79	0.65	2.24
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	10813	19.53	7.89	0.84	3.15
	Module 1	All	26	20	10813	9.73	4.32	0.73	2.23
	Module 2	All	26	20	10813	9.81	4.26	0.73	2.22
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	23453	22.96	8.91	0.87	3.21
	Module 1	All	26	20	23453	11.51	4.77	0.77	2.28
	Module 2	All	26	20	23453	11.45	4.79	0.78	2.25

Table M–5. Spring: Algebra Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	184172	28.10	12.45	0.91	3.73
	Module 1	All	30	21	184172	13.26	6.49	0.83	2.65
	Module 2	All	30	21	184172	14.84	6.49	0.84	2.63
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	60	42	90355	28.70	12.27	0.91	3.76
		Male	60	42	93679	27.53	12.59	0.91	3.70
	Module 1	Female	30	21	90355	13.57	6.46	0.83	2.68
		Male	30	21	93679	12.96	6.49	0.84	2.62
	Module 2	Female	30	21	90355	15.13	6.35	0.83	2.65
		Male	30	21	93679	14.58	6.62	0.84	2.61
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	60	42	32745	20.73	9.90	0.88	3.44
		American Indian	60	42	389	29.94	12.93	0.91	3.81
		Asian	60	42	6076	36.08	14.01	0.92	3.88
		Hispanic	60	42	19110	22.08	10.54	0.89	3.50
		Multi-racial	60	42	3841	26.38	11.97	0.90	3.69
		Native Hawaiian/ Pacific Islander	60	42	160	30.10	12.57	0.91	3.80
		White	60	42	121695	30.68	12.05	0.90	3.76
	Module 1	African American	30	21	32745	9.61	5.08	0.78	2.38
		American Indian	30	21	389	14.24	6.83	0.84	2.73
		Asian	30	21	6076	17.41	7.52	0.86	2.82
		Hispanic	30	21	19110	10.24	5.39	0.80	2.44
		Multi-racial	30	21	3841	12.45	6.19	0.82	2.60
		Native Hawaiian/ Pacific Islander	30	21	160	14.43	6.86	0.84	2.74
		White	30	21	121695	14.53	6.35	0.82	2.69
	Module 2	African American	30	21	32745	11.12	5.45	0.80	2.47
		American Indian	30	21	389	15.70	6.67	0.84	2.66
		Asian	30	21	6076	18.67	6.95	0.85	2.67
		Hispanic	30	21	19110	11.84	5.75	0.81	2.51
		Multi-racial	30	21	3841	13.92	6.33	0.83	2.63
		Native Hawaiian/ Pacific Islander	30	21	160	15.68	6.20	0.82	2.65
		White	30	21	121695	16.16	6.25	0.82	2.63
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	60	42	6159	17.31	8.99	0.87	3.29
	Module 1	All	30	21	6159	8.07	4.70	0.76	2.30
	Module 2	All	30	21	6159	9.23	4.97	0.78	2.34
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	60	42	30114	18.80	9.31	0.87	3.33
	Module 1	All	30	21	30114	8.76	4.66	0.75	2.31
	Module 2	All	30	21	30114	10.04	5.27	0.79	2.40
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	60	42	84749	23.32	10.76	0.89	3.55
	Module 1	All	30	21	84749	10.86	5.52	0.80	2.48
	Module 2	All	30	21	84749	12.46	5.83	0.81	2.53

Table M–6. Spring: Biology Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	154585	34.25	14.49	0.94	3.66
	Module 1	All	33	27	154585	17.75	7.60	0.88	2.60
	Module 2	All	33	27	154585	16.50	7.38	0.88	2.57
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	66	54	76732	34.66	14.22	0.93	3.66
		Male	66	54	77758	33.86	14.74	0.94	3.64
	Module 1	Female	33	27	76732	17.88	7.41	0.88	2.60
		Male	33	27	77758	17.63	7.78	0.89	2.59
	Module 2	Female	33	27	76732	16.78	7.31	0.88	2.57
		Male	33	27	77758	16.23	7.44	0.88	2.56
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	66	54	25239	24.72	11.55	0.90	3.67
		American Indian	66	54	216	31.28	13.40	0.92	3.69
		Asian	66	54	5215	39.86	15.96	0.95	3.52
		Hispanic	66	54	14350	26.68	12.58	0.91	3.68
		Multi-racial	66	54	3086	31.92	14.06	0.93	3.68
		Native Hawaiian/ Pacific Islander	66	54	139	36.68	14.59	0.94	3.60
		White	66	54	106233	37.34	13.88	0.93	3.63
	Module 1	African American	33	27	25239	13.03	6.33	0.83	2.63
		American Indian	33	27	216	16.26	7.14	0.86	2.63
		Asian	33	27	5215	20.82	8.26	0.91	2.48
		Hispanic	33	27	14350	13.94	6.78	0.85	2.63
		Multi-racial	33	27	3086	16.58	7.48	0.88	2.62
		Native Hawaiian/ Pacific Islander	33	27	139	19.12	7.80	0.90	2.51
		White	33	27	106233	19.28	7.29	0.88	2.57
	Module 2	African American	33	27	25239	11.68	5.84	0.81	2.55
		American Indian	33	27	216	15.01	6.88	0.86	2.58
		Asian	33	27	5215	19.04	8.12	0.91	2.50
		Hispanic	33	27	14350	12.74	6.38	0.84	2.56
		Multi-racial	33	27	3086	15.33	7.14	0.87	2.57
		Native Hawaiian/ Pacific Islander	33	27	139	17.56	7.32	0.88	2.58
		White	33	27	106233	18.07	7.11	0.87	2.55
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	4558	19.12	8.85	0.84	3.52
	Module 1	All	33	27	4558	10.04	5.18	0.76	2.52
	Module 2	All	33	27	4558	9.08	4.41	0.69	2.45
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	24776	23.29	10.76	0.89	3.63
	Module 1	All	33	27	24776	12.27	5.90	0.80	2.61
	Module 2	All	33	27	24776	11.02	5.51	0.79	2.51
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	68160	28.12	12.72	0.92	3.69
	Module 1	All	33	27	68160	14.68	6.82	0.85	2.64
	Module 2	All	33	27	68160	13.44	6.48	0.84	2.57

Table M–7. Spring: Literature Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	140252	31.13	10.65	0.91	3.15
	Module 1	All	26	20	140252	15.64	5.61	0.84	2.24
	Module 2	All	26	20	140252	15.49	5.52	0.84	2.21
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	52	40	67195	32.89	10.08	0.90	3.11
		Male	52	40	72966	29.52	10.90	0.92	3.14
	Module 1	Female	26	20	67195	16.65	5.32	0.83	2.22
		Male	26	20	72966	14.72	5.71	0.85	2.23
	Module 2	Female	26	20	67195	16.24	5.25	0.83	2.18
		Male	26	20	72966	14.79	5.66	0.85	2.20
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	52	40	22928	25.50	10.22	0.90	3.29
		American Indian	52	40	194	29.73	9.88	0.89	3.23
		Asian	52	40	4731	34.54	10.71	0.92	3.05
		Hispanic	52	40	12624	25.84	10.52	0.90	3.26
		Multi-racial	52	40	2735	29.89	10.40	0.91	3.20
		Native Hawaiian/ Pacific Islander	52	40	117	32.01	9.53	0.89	3.13
		White	52	40	96821	33.03	10.03	0.91	3.07
	Module 1	African American	26	20	22928	12.98	5.54	0.82	2.32
		American Indian	26	20	194	15.01	5.40	0.82	2.30
		Asian	26	20	4731	17.46	5.51	0.84	2.18
		Hispanic	26	20	12624	13.04	5.62	0.83	2.30
		Multi-racial	26	20	2735	15.09	5.51	0.83	2.27
		Native Hawaiian/ Pacific Islander	26	20	117	16.31	5.13	0.81	2.25
		White	26	20	96821	16.54	5.31	0.83	2.20
	Module 2	African American	26	20	22928	12.52	5.26	0.80	2.33
		American Indian	26	20	194	14.73	5.01	0.79	2.27
		Asian	26	20	4731	17.07	5.62	0.86	2.14
		Hispanic	26	20	12624	12.80	5.42	0.82	2.30
		Multi-racial	26	20	2735	14.80	5.41	0.83	2.25
		Native Hawaiian/ Pacific Islander	26	20	117	15.70	5.07	0.82	2.15
		White	26	20	96821	16.49	5.20	0.83	2.14
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	4003	18.50	7.73	0.83	3.22
	Module 1	All	26	20	4003	9.34	4.28	0.72	2.26
	Module 2	All	26	20	4003	9.16	4.13	0.69	2.29
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	24168	21.35	9.17	0.88	3.20
	Module 1	All	26	20	24168	10.73	5.02	0.80	2.25
	Module 2	All	26	20	24168	10.62	4.77	0.77	2.27
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	52	40	61073	26.95	10.40	0.90	3.24
	Module 1	All	26	20	61073	13.62	5.58	0.83	2.29
	Module 2	All	26	20	61073	13.33	5.35	0.82	2.28

Table M–8. Summer: Algebra Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	1846	25.91	8.75	0.84	3.51
	Module 1	All	30	21	1846	12.90	4.75	0.71	2.54
	Module 2	All	30	21	1846	13.01	4.74	0.74	2.43
Gender	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	Female	60	42	897	26.63	8.60	0.83	3.55
		Male	60	42	910	25.33	8.92	0.85	3.48
	Module 1	Female	30	21	897	13.31	4.66	0.69	2.58
		Male	30	21	910	12.55	4.83	0.73	2.49
	Module 2	Female	30	21	897	13.32	4.67	0.73	2.43
		Male	30	21	910	12.78	4.83	0.75	2.41
Ethnicity	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	African American	60	42	278	23.73	9.97	0.87	3.55
		American Indian	60	42	DNR	DNR	DNR	DNR	DNR
		Asian	60	42	45	30.22	10.85	0.88	3.75
		Hispanic	60	42	83	26.08	7.92	0.82	3.39
		Multi-racial	60	42	38	23.45	8.72	0.84	3.52
		Native Hawaiian/ Pacific Islander	60	42	DNR	DNR	DNR	DNR	DNR
		White	60	42	1338	26.32	8.41	0.83	3.50
	Module 1	African American	30	21	278	11.55	5.06	0.75	2.52
		American Indian	30	21	DNR	DNR	DNR	DNR	DNR
		Asian	30	21	45	14.98	6.13	0.80	2.71
		Hispanic	30	21	83	12.64	4.24	0.68	2.41
		Multi-racial	30	21	38	11.45	4.61	0.70	2.51
		Native Hawaiian/ Pacific Islander	30	21	DNR	DNR	DNR	DNR	DNR
		White	30	21	1338	13.18	4.62	0.70	2.54
	Module 2	African American	30	21	278	12.18	5.52	0.79	2.50
		American Indian	30	21	DNR	DNR	DNR	DNR	DNR
		Asian	30	21	45	15.24	5.30	0.76	2.60
		Hispanic	30	21	83	13.45	4.71	0.75	2.36
		Multi-racial	30	21	38	12.00	4.95	0.75	2.45
		Native Hawaiian/ Pacific Islander	30	21	DNR	DNR	DNR	DNR	DNR
		White	30	21	1338	13.14	4.54	0.72	2.40
ELL	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	15	20.80	8.60	0.84	3.42
	Module 1	All	30	21	15	10.07	5.09	0.78	2.41
	Module 2	All	30	21	15	10.73	4.68	0.74	2.37
IEP	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	304	22.42	8.47	0.84	3.39
	Module 1	All	30	21	304	11.24	4.68	0.72	2.46
	Module 2	All	30	21	304	11.18	4.51	0.73	2.33
ED	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	60	42	519	24.13	9.12	0.85	3.50
	Module 1	All	30	21	519	11.91	4.79	0.73	2.48
	Module 2	All	30	21	519	12.21	5.08	0.77	2.45

Table M–9. Summer: Biology Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	1116	31.50	9.90	0.85	3.80
	Module 1	All	33	27	1116	15.83	5.30	0.73	2.73
	Module 2	All	33	27	1116	15.67	5.39	0.76	2.63
Gender	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	Female	66	54	565	32.38	9.11	0.83	3.81
		Male	66	54	528	30.72	10.67	0.88	3.77
	Module 1	Female	33	27	565	16.18	4.93	0.69	2.74
		Male	33	27	528	15.54	5.70	0.77	2.72
	Module 2	Female	33	27	565	16.21	5.04	0.72	2.64
		Male	33	27	528	15.18	5.72	0.79	2.60
Ethnicity	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	African American	66	54	145	27.30	9.57	0.85	3.75
		American Indian	66	54	DNR	DNR	DNR	DNR	DNR
		Asian	66	54	47	42.06	12.18	0.91	3.56
		Hispanic	66	54	79	30.05	8.54	0.79	3.88
		Multi-racial	66	54	26	34.12	10.70	0.87	3.86
		Native Hawaiian/ Pacific Islander	66	54	DNR	DNR	DNR	DNR	DNR
		White	66	54	790	31.79	9.43	0.84	3.80
	Module 1	African American	33	27	145	13.91	5.17	0.73	2.69
		American Indian	33	27	DNR	DNR	DNR	DNR	DNR
		Asian	33	27	47	21.40	6.68	0.86	2.48
		Hispanic	33	27	79	15.54	4.75	0.65	2.82
		Multi-racial	33	27	26	16.65	5.86	0.78	2.76
		Native Hawaiian/ Pacific Islander	33	27	DNR	DNR	DNR	DNR	DNR
		White	33	27	790	15.90	5.06	0.71	2.74
	Module 2	African American	33	27	145	13.39	5.24	0.75	2.60
		American Indian	33	27	3	17.67	2.52	-0.33	2.90
		Asian	33	27	47	20.66	6.25	0.84	2.53
		Hispanic	33	27	79	14.51	4.56	0.66	2.67
		Multi-racial	33	27	26	17.46	6.06	0.81	2.65
		Native Hawaiian/ Pacific Islander	33	27	DNR	DNR	DNR	DNR	DNR
		White	33	27	790	15.89	5.19	0.74	2.63
ELL	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	20	30.00	8.25	0.77	3.94
	Module 1	All	33	27	20	15.70	5.19	0.70	2.84
	Module 2	All	33	27	20	14.30	4.37	0.62	2.69
IEP	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	224	27.40	9.57	0.84	3.78
	Module 1	All	33	27	224	13.72	5.11	0.71	2.75
	Module 2	All	33	27	224	13.68	5.29	0.76	2.59
ED	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
	Total	All	66	54	328	29.17	9.44	0.84	3.81
	Module 1	All	33	27	328	14.80	5.00	0.70	2.76
	Module 2	All	33	27	328	14.37	5.35	0.76	2.61

Table M–10. Summer: Literature Reliabilities

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	588	25.68	7.88	0.84	3.20
	Module 1	All	26	20	588	13.29	4.06	0.68	2.30
	Module 2	All	26	20	588	12.40	4.50	0.76	2.21
Gender	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	Female	52	40	241	27.64	7.30	0.81	3.21
		Male	52	40	339	24.28	7.99	0.84	3.16
	Module 1	Female	26	20	241	14.28	3.81	0.64	2.30
		Male	26	20	339	12.58	4.10	0.69	2.29
	Module 2	Female	26	20	241	13.36	4.15	0.71	2.24
		Male	26	20	339	11.70	4.62	0.78	2.17
Ethnicity	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	African American	52	40	84	20.55	8.74	0.86	3.23
		American Indian	52	40	DNR	DNR	DNR	DNR	DNR
		Asian	52	40	12	32.08	5.84	0.72	3.11
		Hispanic	52	40	39	23.82	7.41	0.81	3.24
		Multi-racial	52	40	23	24.43	7.39	0.80	3.33
		Native Hawaiian/ Pacific Islander	52	40	DNR	DNR	DNR	DNR	DNR
		White	52	40	416	26.72	7.35	0.81	3.17
	Module 1	African American	26	20	84	10.95	4.24	0.69	2.35
		American Indian	26	20	DNR	DNR	DNR	DNR	DNR
		Asian	26	20	12	16.17	3.01	0.42	2.29
		Hispanic	26	20	39	12.13	4.21	0.70	2.30
		Multi-racial	26	20	23	12.13	3.95	0.64	2.37
		Native Hawaiian/ Pacific Islander	26	20	DNR	DNR	DNR	DNR	DNR
		White	26	20	416	13.82	3.84	0.65	2.29
	Module 2	African American	26	20	84	9.60	5.12	0.81	2.21
		American Indian	26	20	DNR	DNR	DNR	DNR	DNR
		Asian	26	20	12	15.92	3.60	0.66	2.10
		Hispanic	26	20	39	11.69	4.12	0.70	2.26
		Multi-racial	26	20	23	12.30	3.90	0.63	2.36
		Native Hawaiian/ Pacific Islander	26	20	DNR	DNR	DNR	DNR	DNR
		White	26	20	416	12.90	4.24	0.73	2.18
ELL	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	52	40	11	22.27	5.83	0.68	3.30
	Module 1	All	26	20	11	11.09	3.67	0.61	2.30
	Module 2	All	26	20	11	11.18	3.03	0.39	2.37
IEP	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	52	40	145	21.60	7.72	0.83	3.18
	Module 1	All	26	20	145	11.43	4.13	0.69	2.31
	Module 2	All	26	20	145	10.17	4.24	0.73	2.19
ED	**Level**	**Group**	**Pts.**	**Len.**	**N**	**Mean**	**SD**	**r**	**SEM**
	Total	All	52	40	201	23.97	7.84	0.83	3.25
	Module 1	All	26	20	201	12.61	4.10	0.67	2.34
	Module 2	All	26	20	201	11.36	4.45	0.75	2.25