TECHNICAL REPORT

pennsylvania
DEPARTMENT OF EDUCATION

2020–2021 CDT Technical Report

Provided by Data Recognition Corporation

This document has been formatted to be ADA compliant.

TABLE OF CONTENTS

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.

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 CDT, a student's reading ability is measured by how the student performed on the CDT Reading/Literature test.
Alternative Forms	*Alternative forms* are two or more versions of a test that are considered exchangeable; for example, 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, comparable forms), where parallel forms refers to the situation in which the test forms have the highest degree of similarity to each other.
Average	*Average* is 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).
Benchmark Activity	Also referred to as benchmarking, *benchmark activity* is a procedure used in the determination of the cut score(s) for a given assessment. It is used to measure students' progress towards certain performance standards. Methods vary (e.g., modified Angoff, Bookmark Method), 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.
Benchmark Cut	A *benchmark cut* marks a specified point on a score scale where scores at or above that point are interpreted differently from scores below that point (e.g., a score designated as the minimum level of performance needed to pass a competency test). A test can be divided into multiple proficiency levels by setting one or more cut scores. Methods for establishing cut scores vary. For the CDT, one benchmark cut was set that separates students into two categories: solidly ready for the next grade or course and not solidly ready for the next grade or course.
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). 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 including them in the final operational pools (see also *Differential Item Functioning*).
Computer Adaptive Test (CAT)	A *computer adaptive test* (*CAT*) is a computer-based test with an item selection routine that adjusts (adapts) to a student's performance during the test. For this reason, it has also been called a tailored test. Rather than all students taking the same set of items (fixed form), each student's test is individually tailored with items selected from a large item pool based on the student's performance.
Constructed-Response Item	A *constructed-response item*—referred to by some as an open-ended response item—is an item format that requires examinees to create their own responses, which can be expressed in various forms. This format is in contrast to multiple-choice items, which require students to make a choice from a supplied set of answer options. There are no constructed-response items on the CDT.
Content Validity Evidence	*Content validity evidence* shows the extent to which an exam 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).
Criterion-Referenced Interpretation	The *criterion-referenced interpretation* is a measure of a student's performance against 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 in a given content area.

Term	Common Definition
Decision Consistency	*Decision consistency* is the extent to which classifications based on test scores would match the decisions on students' proficiency levels 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.
Diagnostic Category	A *diagnostic category* is a grouping used for reporting results on the CDT. Each CDT test has four or five diagnostic categories which are based on the Pennsylvania Academic Standards (Mathematics, Reading, and Writing) or the Pennsylvania Academic Standards (Science).
Differential Item Functioning (DIF)	*Differential item functioning* (*DIF*) is 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	A *distractor* is an incorrect option in a multiple-choice item (also called a foil).
Equating	*The process that results in scores that can be used interchangeably across different test forms and/or test administrations. Equated test scores are considered exchangeable. Consequently, the requirements for equating are strong and somewhat complex (equal construct and precision, equity, and invariance). In practical terms, it is often stated that students should perceive no differences regardless of the test form administered (see also Scale Linking, Pre-equating, and Post-equating).*
Evidence-Based Selected-Response Item	A type of item that has two parts and requires the test taker to select a response from a group of possible answer choices in Part One, one of which is the correct answer (or key) to the question posed, and to then select one or two responses from a group of possible answer choices in Part Two, which provide evidence to support the correct answer in Part One.
Field-Test item	A *field-test item* is a newly developed item that is ready to be tried out to determine its statistical properties (e.g., see *p*-value and Point-Biserial Correlation). Items are field tested prior to operational administration. Items with acceptable statistical properties in field-test form the pool of CDT operational items.
Frequency	*Frequency* is the number of times that a certain value or range of values (score interval) occurs in a distribution of scores.
Frequency Distribution	*Frequency distribution* is a tabulation of scores from low to high or high to low with the number and/or percent of individuals who obtain each score or who fall within each score interval.
Infit/Outfit	*Infit* and *outfit* are statistical indicators of the agreement of the data and the measurement model. Infit and outfit are highly correlated, and they 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.
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.
Key	The *key* is the correct response option or answer to a test item.
Learning Progression	A *learning progression* shows the developmental sequences or building blocks of content/skills students need to master as they progress toward career and college readiness and is tied directly to the Assessment Anchors and Eligible Content as well as the Voluntary Model Curriculum Units and Lesson Plans.
Linking	A generic term referring to 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*). CDT scores are equated.
Logit	In Rasch scaling, *logits* are units used to express both examinee ability and item difficulty. When expressing examinee ability, if two students take the same set of items, a student who answers more items correctly has a higher logit than a student who answers fewer items correctly. Logits are transformed into scale scores through a linear transformation. When expressing item difficulty, logits are transformed *p*-value (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.

Term	Common Definition
Mean	*Mean* is also referred to as the arithmetic mean of a set of scores. It 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, and 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 CDT, a student's literature measure might be equal to 0.525 logit. Or, a CDT literature test item might have a logit equal to -0.905.
Median	The *median* is the middle point or score in a set of rank-ordered observations that divides the distribution into two equal parts; 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.
Multiple-Choice Item	A *multiple-choice item* is 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. All items on the CDT are multiple-choice items.
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), and the number of students who attained a specific score. In the following set {23, 32, 56, 65, 78, 87}, $n = 6$.
Operational Item	After initial item tryout (field test), all items with acceptable statistical properties form the pool of CDT *operational items*. Students' tests are selected from this pool.
Percent Correct	When referring to an individual item, the *percent correct* is the item's *p*-value from the field test administration expressed as a percent (instead of a proportion). Under a computer adaptive administration, percent correct scores are not appropriate for individual items or students.
Percentile	*Percentile* is 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 scale score of 1500 on a given test, then the scale score of 1500 would be considered the 72nd percentile. As another example, the median is the 50th percentile.
Percentile Rank	The *percentile rank* is the percentage of scores in a specified distribution that fall at/below a certain point on a score distribution. Percentile ranks range in value from 1 to 99. They 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 ranks are sometimes used interchangeably; however, strictly speaking, a percentile is a value on the score scale.
Point-Biserial Correlation	In classical test theory, *point-biserial correlation* 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.
Post-Equating	Post-equating refers to the method of utilizing data from the current administration for scale linking and equating. Post-equating relies heavily on collecting data from a representative sample, estimating new item parameters, linking the item parameters to the base sale, and estimating student ability based on the linked item parameters. In order to provide immediate results, CDT utilizes pre-equating. Post-equating is conducted for field-test analyses and updating item parameters.
Pre-Equating	Pre-equating refers to the method of utilizing previously estimated and linked item parameters for equating. Because item parameters have already been linked to the base scale, pre-equated solutions are available immediately after a CDT is completed.

Term	Common Definition
P-value	A *p-value* is 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*).
Raw Score	*Raw score* is an unadjusted score usually determined by tallying the number of questions answered correctly or by the sum of item scores (i.e., points). Raw scores typically have little or no meaning by themselves and require additional information like the number of items on the test and the difficulty of the test items. Under a computer adaptive administration, where each student takes a unique set of items, raw scores are not comparable across students.
Reliability	*Reliability* is the expected degree to which test scores for a group of examinees are consistent over exchangeable replications of an assessment procedure and, therefore, 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	*Reliability coefficient* is 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 a 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*).
Scale Linking	*The first step in any equating process in which independent item estimates are placed on the same scale of measurement (the logit scale). Scale linking results in item parameters that are on the same scale of measurement. Equating procedures can only be implemented once scale linking is achieved (see also Equating).*
Scale Score	*Scale score* is a mathematical transformation of a Rasch ability estimate developed through a process called scaling. Scale 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.
Standard Deviation	*Standard deviation* is 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 one another 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.
Standard Error of Measurement (SEM)	*Standard error of measurement* (*SEM*) is 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 error of measurement (CSEM) also indicates the degree of measurement error in scale score units but varies as a function of a student's unique set of items and actual scale 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).

Term	Common Definition
Technical Advisory Committee (TAC)	The *technical advisory committee* (*TAC*) is a group of individuals (most often professionals in the field of testing) that are either appointed or selected to make recommendations for and to guide the technical development of a given testing program.
Technology Enhanced (TE) Items	*Technology Enhanced (TE)* items are items that capitalize on computer-based interactions for collecting response data. Examples of TE items include drop-down menus, drag and drop functionality, text highlighting, and other interactions.
Validity	*Validity* is the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by the purpose of a test. There are various ways of gathering validity evidence.

PREFACE: AN OVERVIEW OF THE CDT

CLASSROOM DIAGNOSTIC TOOLS (CDT) OVERVIEW

The Pennsylvania Classroom Diagnostic Tools (CDT) is a set of online assessments, divided by content area, designed to provide diagnostic information in order to guide instruction and intervention. The CDT reporting system is fully integrated in Pennsylvania's Standards Aligned System (SAS). It assists educators in identifying student academic strengths and areas in need of improvement by providing links to classroom resources. The diagnostic reports feature easy-to-follow links to targeted curricular resources and materials, including units and lesson plans found within the SAS system. Students in grades 3 through high school at all Pennsylvania schools may take the CDT up to five times throughout the school year at no cost.

The purpose of the CDT is to provide information that will help guide instruction by providing support to students and teachers. The CDT reports are designed to provide a picture or snapshot of how students are performing in relation to the Pennsylvania Assessment Anchors and Eligible Content and Keystone Assessment Anchors and Eligible Content. The CDT goes beyond focusing only on **What** students should know and be able to do at a particular grade and/or course. It also provides a snapshot of **How** and **Why** students may still be struggling or extending beyond the grade and/or course Eligible Content. This valuable information is typically not identified through other types of assessments. Teachers, through the use of the CDT reports, may access additional information through the Learning Progression Map. The Learning Progression Map allows teachers to pinpoint where students are struggling or where they are extending beyond the learning continuum. The CDT helps identify and provides suggestions for next steps in student academic development.

The CDT consists of only multiple-choice questions and evidence-based selected response questions. The questions were developed to specifically align to the Pennsylvania Assessment Anchors and Eligible Content at kindergarten through high school and the Keystone Assessment Anchors and Eligible Content for end-of-course. The CDT is based on content assessed by the Pennsylvania System of School Assessments (PSSA) and the Keystone Exams. It includes interactive and dynamic reporting for various diagnostic reporting categories.

CDT Activities for the 2020–2021 School Year

Description	Date
Test Setup System Available	August 10, 2020
First Day of Testing	August 17, 2020
Last Day of Testing	August 6, 2021

CHAPTER ONE: BACKGROUND OF THE CLASSROOM DIAGNOSTIC TOOLS

This brief overview of the Pennsylvania Classroom Diagnostic Tools summarizes the program's intent and purpose, as well as key dates in the development process.

THE CLASSROOM DIAGNOSTIC TOOLS

The Classroom Diagnostic Tools (CDT) is a set of online assessments, divided by content area, designed to provide diagnostic information in order to a guide instruction and enrichment. The CDT reporting system is fully integrated in the Standards Aligned System (SAS). It assists educators in identifying student academic strengths and areas in need of improvement by providing links to classroom resources. The diagnostic reports feature easy-to-follow links to targeted curricular resources and materials, including units and lesson plans found within the SAS system. The CDT is available to districts at no cost.

The CDT is:

- Offered to students in grades 3 through high school
- Available for use in the classroom throughout the school year on a voluntary basis
- Based on content assessed by the Keystone Exams and the Pennsylvania System of School Assessment (PSSA)
- Comprised of multiple-choice items (all content areas), and evidence-based selected-response items (in Reading and Literature only)
- Delivered as an online Computer Adaptive Test (CAT), ensuring valid and reliable measures of a student's skills while minimizing testing time
- Designed to provide real-time results for students and teachers with links to Materials and Resources in SAS
- Available for Mathematics Lower Grades[1], Mathematics, Algebra I, Geometry, Algebra II, Reading Lower Grades, Reading/Literature, Science Lower Grades, Science, Biology, Chemistry, Writing Lower Grades, and Writing/English Composition
- Available as Full CDT, which covers multiple diagnostic categories, or as Diagnostic Category CDT, which covers a single category.

KEY DATES

The items for each course of the CDT were field tested online using fixed-form computer-based tests prior to their use in operational computer adaptive tests. Additional items were field tested as items embedded within the operational CDT to increase the pool of items aligned to the Pennsylvania Core Standards and to allow the extension of the CDT to students in grades 3 through 5. The timeline for implementation of the field tests and operational availability is shown in the following table.

[1] CDTs with the "Lower Grades" designation are for students in grades 3 through 5.

Table 1–1. Key Dates

Course	Field Test Dates	Operational Rollout Dates
Mathematics, Algebra I, Geometry, Algebra II	Spring 2010	Fall 2010
Reading/Literature	Fall 2010	Spring 2011
Science, Biology, Chemistry	Fall 2010	Spring 2011
Writing/English Composition	Spring 2011	Fall 2011
Mathematics, Reading/Literature, and Writing/English Composition aligned to the Pennsylvania Core Standards[2]	Spring 2013	Fall 2013
Mathematics Lower Grades, Reading Lower Grades, Science Lower Grades, and Writing Lower Grades	Fall 2013	Spring 2014
Mathematics, Algebra I, Algebra II, Reading, Literature, Writing, English Composition, Science, Biology, and Chemistry	Fall 2018	Fall 2019
Science and Biology	Spring 2019	Spring 2020

For more details on field-test events, see Chapter Six.[2]

[2] The alignment of Mathematics, Reading/Literature, and Writing/English Composition to the Pennsylvania Core Standards did not include field-test items for Writing/English Composition, as the Writing/English Composition pool did not require additional items to be fully aligned to the Pennsylvania Core Standards.

CHAPTER TWO: TEST DEVELOPMENT OVERVIEW OF THE PENNSYLVANIA CDT FRAMEWORK

The Pennsylvania Classroom Diagnostic Tools (CDT) is available for Mathematics Lower Grades, Mathematics, Algebra I, Geometry, Algebra II, Reading Lower Grades, Reading/Literature, Science Lower Grades, Science, Biology, Chemistry, Writing Lower Grades, and Writing/English Composition for students in grades 3 through high school. The assessments are administered online in a computer adaptive test (CAT) format.

The Pennsylvania CDT consists of multiple-choice, evidence-based selected-response, and technology-enhanced, questions that align to the Pennsylvania Assessment Anchors and Eligible Content at grades 3 through high school for mathematics, reading, writing, and science and the Keystone Assessment Anchors and Eligible Content for end-of-course for Algebra I, Algebra II, Geometry, Literature, English Composition, Biology, and Chemistry and evidence-based selected-response questions that align to the Pennsylvania Assessment Anchors and Eligible Content at grade 3 through 8 for reading. With the exception of grades 3, 5, 6, and 7 for Science, these Pennsylvania Assessment Anchors and Eligible Content were developed previously for the PSSA and Keystone Exams as described in the following sections. In addition, Learning Progressions were developed to show the pathways along which students travel as they progress towards mastery of the skills in each content area.

BACKGROUND FOR THE PSSA ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The PSSA Assessment Anchor Content Standards and Eligible Content in Mathematics, Reading, and Writing are based on the Pennsylvania Core Standards. The PSSA Assessment Anchor Content Standards and Eligible Content in Science are based on the Pennsylvania Academic Standards. Although the Pennsylvania Core Standards and the Pennsylvania Academic Standards indicate what students should know and be able to do, educator concerns regarding the number and breadth of Academic Standards led to an initiative by the Pennsylvania Department of Education (PDE) to develop Assessment Anchor Content Standards (Assessment Anchors) to indicate which parts of the Academic Standards (Instructional Standards) would be assessed on the PSSA. 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 across each grade span and focus the content of the standards into what is assessable on a large-scale test. The Assessment Anchor documents also serve to communicate Eligible Content, also called assessment limits, or the range of knowledge and skills from which the PSSA would be designed.

The Assessment Anchor's coding is read like an outline. The coding includes the content, grade level, Reporting Category, Assessment Anchor, descriptor (Sub-Assessment Anchor), and Eligible Content. Thus, S.4.A.1.3.1 would be Science, Grade 4, Reporting Category A, Assessment Anchor 1, descriptor (Sub-Assessment Anchor) 3, and Eligible Content 1.

Each of the Assessment Anchors has one or more descriptors (Sub-Assessment Anchors) and Eligible Content varying to reflect grade-level appropriateness. The Assessment Anchors form the basis of the test design for the grades undergoing new test development. In turn, this hierarchy is the basis for organizing the total content scores (based on the core [common] sections).

With Pennsylvania's decision to adopt the Pennsylvania Core Standards based on the Common Core State Standards, committees of Pennsylvania educators met in October 2011 to write, review, and approve the Assessment Anchors and Eligible Content statements. To provide initial focus, each content and grade span committee was presented with materials specific to the content and grade span in question, including a basic blueprint structure, the Pennsylvania Academic Standards, the Pennsylvania Assessment Anchors and Eligible Content aligned to the Pennsylvania Academic Standards, the Common Core State Standards, and draft Eligible Content statements. Committees then completed an iterative process of reviewing and revising the draft Eligible Content statements followed by discussions across grade-span committees to ensure vertical articulation across the grades. The results from the committee work were evaluated by national, state, and local subject matter experts, and, following revisions, they were ultimately validated by another committee of Pennsylvania educators. Following committee approval, the Pennsylvania Core Standards-aligned Assessment Anchors and Eligible Content for English Language Arts and Mathematics were approved by the State Board of Education in September 2013.

The complete set of Assessment Anchors and Eligible Content can be referenced at PDE's website: www.education.pa.gov.

- Roll over `Data and Reporting' in the bar across the top of the page.
- Select `Assessment and Accountability.' Click on the link that reads `PSSA - PA System of School Assessment'. Then click on Assessment Anchors/Eligible Content

For Science, Assessment Anchors and Eligible Content had only been previously developed at grades 4, 8, and 11 for the PSSA and for the Biology and Chemistry Keystone Exams. Therefore, to provide a vertical articulation of science content from grade to grade, a group of Pennsylvania educators were brought together to develop Assessment Anchors and Eligible Content for the off grades (those that do not assess Science on the PSSA). These educators, in collaboration with DRC Science Test Development staff, used the Assessment Anchors and Eligible Content for grades 4, 8, and 11 as the foundation to develop Assessment Anchors and Eligible Content for grades 3, 5, 6, and 7.

With the extension of the CDT to allow students in grades 3 through 5 to participate in the assessments, it was necessary to include items appropriate to assess skills and understandings that students should learn in kindergarten through grade 2. For Mathematics, Reading, and Writing, test questions were developed based to align to the Pennsylvania Core Standards for grades K through 2. For Science, a group of Pennsylvania educators was brought together in March 2013 to develop the Science Grades K-2 Assessment Anchors and Eligible Content, which are organized as a single grade band and contain foundational science concepts in order to promote flexibility in classroom instruction for these early grade levels.

BACKGROUND FOR THE KEYSTONE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The Keystone Test Blueprints—known as the Keystone Assessment Anchors and Eligible Content—are based on Pennsylvania Keystone Course Standards and the Common Core State 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 college and career ready.

Although the Keystone Course Standards indicate what students should know and be able to do, Assessment Anchors are designed to indicate which parts of the Keystone Course Standards (Instructional Standards) 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, or the range of knowledge and skills from which the Keystone Exams are designed.

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

- **Clear:** The Assessment Anchors are easy to read and are user-friendly; they clearly detail which standards are assessed on the Keystone Exams.
- **Focused:** The Assessment Anchors identify a core set of standards that could be reasonably assessed on a large-scale assessment, which will keep 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 first by Module (Reporting Category), then by Assessment Anchor, followed by Anchor Descriptor, and then finally, at the greatest level of detail, by an Eligible Content statement. The common format of this outline is followed across the Keystone Exams.

Here is a description of each level in the labeling system for the Keystone Exams.

- **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 provides further details that delineate 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 statements 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 on the Keystone Exams. This level is considered the assessment limit and 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 indicate certain clarifications about the scope of an Eligible Content.
- **Notes:** There are three types of notes included in the Assessment Anchor document:
  “e.g.” (“for example”)—sample approach, but not a limit to the Eligible Content
  “i.e.” (“that is”)—specific limit to the Eligible Content
  “Note”—content exclusions or definable range of the Eligible Content

The Assessment Anchor’s coding is read like an outline. The coding includes the Subject (Exam), Reporting Category/Module, Assessment Anchor, Anchor Descriptor, and Eligible Content. Each exam has two modules. Each Module has two or more Assessment Anchors. Each of the Assessment Anchors has one or more Anchor Descriptors, and each Anchor Descriptor has at least one Eligible Content statements (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.

Table 2–1. Sample Keystone Assessment Anchor Coding

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

The complete set of Assessment Anchors and Eligible Content can be referenced at *PDE's Standards Aligned System (SAS) website* at http://www.pdesas.org/Standard. Assessment Anchors and Eligible Content for Grades 3–8 can be found by selecting "Download PSSA and PASA Anchors and Eligible Content" while Assessment Anchors and Eligible Content for high school courses can be found by selecting "Download Keystone Anchors."

DIAGNOSTIC CATEGORIES FOR THE CLASSROOM DIAGNOSTIC TOOLS

The Classroom Diagnostic Tools provide information for teachers, students, and other stakeholders regarding student performance at the Overall Score level and also for each diagnostic category within the selected assessment. These diagnostic categories provide more detailed information about student strengths and areas of need for a related group of Eligible Content. A description of the diagnostic categories for each assessment follows.

MATHEMATICS LOWER GRADES AND MATHEMATICS

There are four diagnostic categories for the mathematics assessments. These are *Numbers & Operations*, *Algebraic Concepts*, Geometry, and *Measurement, Data, and Probability*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the table below.

Table 2–2. Number of Eligible Content per Diagnostic Category by Grade for Mathematics Lower Grades and Mathematics

Diagnostic Category	Kindergarten*	Grade 1*	Grade 2*	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Numbers & Operations	1	3	3	9	20	13	15	9	5	6
Algebraic Concepts	1	2	3	14	8	4	11	5	17	46
Geometry	2	2	2	3	3	3	6	8	8	29
Measurement, Data, and Probability	2	3	5	15	9	5	4	7	4	12

* Eligible Content for Kindergarten, Grade 1, and Grade 2 are not included in the Mathematics CDT.

ALGEBRA I

The Keystone Algebra I Assessment Anchors and Eligible Content has two reporting categories: Module 1, Operations and Linear Equations & Inequalities, and Module 2, Linear Functions and Data Organizations. These modules are each divided into two diagnostic categories. Module 1 is divided into *Operations with Real Numbers and Expressions* and *Linear Equations & Inequalities*. Module 2 is divided into *Functions & Coordinate Geometry* and *Data Analysis*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–3. Number of Eligible Content per Diagnostic Category by Grade for Algebra I

Diagnostic Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Module 1 – Operations with Real Numbers and Expressions	13	11	5	17	10	7	18
Module 1 – Linear Equations & Inequalities	0	0	0	3	3	8	16
Module 2 – Functions & Coordinate Geometry	0	3	1	4	1	10	21
Module 2 – Data Analysis	3	0	1	4	7	4	11

GEOMETRY

The Keystone Geometry Assessment Anchors and Eligible Content has two reporting categories: Module 1, Geometric Properties & Reasoning, and Module 2, Coordinate Geometry & Measurement. These modules are each divided into two diagnostic categories. Module 1 is divided into *Geometric Properties* and *Congruence, Similarity, & Proofs*. Module 2 is divided into *Coordinate Geometry & Right Triangles* and *Measurement*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–4. Number of Eligible Content per Diagnostic Category by Grade for Geometry

Diagnostic Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Module 1 – Geometric Properties	2	2	1	1	5	1	18
Module 1 – Congruence, Similarity, & Proofs	0	1	0	0	0	2	3
Module 2 – Coordinate Geometry & Right Triangles	0	0	1	3	1	7	5
Module 2 – Measurement	6	4	2	4	3	0	13

ALGEBRA II

The Keystone Algebra II Assessment Anchors and Eligible Content has two reporting categories: Module 1, Number Systems and Non-Linear Expressions & Equations, and Module 2, Functions and Data Analysis. These modules are each divided into two diagnostic categories. Module 1 is divided into *Operations with Complex Numbers* and *Non-Linear Expressions & Equations*. Module 2 is divided into *Functions* and *Data Analysis*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–5. Number of Eligible Content per Diagnostic Category by Grade for Algebra II

Diagnostic Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Module 1 – Operations with Complex Numbers	0	0	0	0	0	0	4
Module 1 – Non-Linear Expressions & Equations	0	1	1	16	9	8	30
Module 2 – Functions	0	3	0	1	0	5	20
Module 2 – Data Analysis	3	0	1	4	7	3	11

SCIENCE LOWER GRADES AND SCIENCE

There are four diagnostic categories for the science assessments. These are *The Nature of Science*, *Biological Sciences*, *Physical Sciences*, and *Earth/Space Sciences*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the table below.

Table 2–6. Number of Eligible Content per Diagnostic Category by Grade for Science Lower Grades and Science

Diagnostic Category	K–2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
The Nature of Science	7	9	20	8	10	19	31	27
Biological Sciences	7	14	18	11	7	21	21	38
Physical Sciences	1	10	9	12	12	12	12	46
Earth/Space Sciences	8	13	16	8	7	11	13	14

BIOLOGY

The Keystone Biology Exam has two reporting categories: Module 1[A], Cells and Cell Processes, and Module 2[B], Continuity and Unity of Life. These modules are each divided into two diagnostic categories. Module 1 is divided into *Basic Biological Principles/Chemical Basis for Life* and *Bioenergetics/Homeostasis & Transport*. Module 2 is divided into *Cell Growth & Reproduction/Genetics* and *Theory of Evolution/Ecology*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–7. Number of Eligible Content per Diagnostic Category by Grade for Biology

Diagnostic Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Module 1 – Basic Biological Principles/Chemical Basis for Life	5	5	3	3	5	5	9
Module 1 – Bioenergetics/Homeostasis & Transport	0	0	0	0	0	0	7
Module 2 – Cell Growth & Reproduction/Genetics	2	1	1	0	5	4	10
Module 2 – Theory of Evolution/Ecology	8	13	5	3	18	18	12

CHEMISTRY

The Keystone Chemistry Assessment Anchors and Eligible Content has two reporting categories: Module 1[A], Structure and Properties of Matter, and Module 2[B], The Mole Concept and Chemical Interactions. These modules are each divided into two diagnostic categories. Module 1 is divided into *Properties & Classification of Matter* and *Atomic Structure & the Periodic Table*. Module 2 is divided into *The Mole & Chemical Bonding* and *Chemical Relationships & Reactions*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–8. Number of Eligible Content per Diagnostic Category by Grade for Chemistry

Diagnostic Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Module 1 – Properties & Classification of Matter	7	4	7	7	3	3	10
Module 1 – Atomic Structure & The Periodic Table	0	0	0	0	1	0	8
Module 1 – The Mole & Chemical Bonding	0	0	0	0	1	1	9
Module 2 – Chemical Relationships & Reactions	0	0	1	0	1	1	7

READING LOWER GRADES AND READING/LITERATURE

The Reading Lower Grades and Reading/Literature Assessments use the same diagnostic categories across grades 3 through 8 and the high school Literature course. These diagnostic categories are not divided across the two Keystone Literature Modules (reporting categories) of Fiction and Non-fiction. The diagnostic categories for Reading Lower Grades and Reading/Literature are *Key Ideas and Details – Literature Text; Key Ideas and Details – Informational Text; Craft and Structure/Integration of Knowledge and Ideas – Literature Text; Craft and Structure/Integration of Knowledge and Ideas – Informational Text;* and *Vocabulary Acquisition and Use*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–9. Number of Eligible Content per Diagnostic Category by Grade for Reading Lower Grades and Reading/Literature

Diagnostic Category	Kindergarten*	Grade 1*	Grade 2*	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Key Ideas and Details—Literature Text	3	3	3	3	3	3	3	3	3	8
Key Ideas and Details—Informational Text	3	3	3	3	3	3	3	3	3	12
Craft and Structure/ Integration of Knowledge and Ideas—Literature Text	2	2	2	2	2	2	4	4	4	14
Craft and Structure/ Integration of Knowledge and Ideas—Informational Text	4	4	4	5	5	5	5	5	5	18
Vocabulary Acquisition and Use	2	2	2	4	4	4	4	4	4	6

* Eligible Content for Kindergarten, Grade 1, and Grade 2 are not included in the Reading/Literature CDT.

WRITING LOWER GRADES AND WRITING/ENGLISH COMPOSITION

The Writing Lower Grades and Writing/English Composition Assessments use the same diagnostic categories across grades 3 through 8 and the high school English Composition course. The diagnostic categories for Writing Lower Grades and Writing/English Composition are *Quality of Writing: Focus and Organization*, *Quality of Writing: Content and Style*, *Quality of Writing: Editing*, *Conventions: Punctuation, Capitalization, and Spelling*, and *Conventions: Grammar and Sentence Formation*. The number of Eligible Content from each grade that map to these diagnostic categories is shown in the following table.

Table 2–10. Number of Eligible Content per Diagnostic Category by Grade for Writing Lower Grades and Writing/English Composition

Diagnostic Category	Kindergarten*	Grade 1*	Grade 2*	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Quality of Writing: Focus and Organization	3	6	6	6	6	6	6	6	6	4
Quality of Writing: Content and Style	2	3	3	3	3	5	5	5	5	4
Quality of Writing: Editing	0	3	3	4	10	12	11	10	6	13
Conventions: Punctuation, Capitalization, and Spelling	1	3	2	6	4	5	3	3	3	5
Conventions: Grammar and Sentence Formation	2	3	2	10	9	9	9	7	5	2

* Eligible Content for Kindergarten, Grade 1, and Grade 2 are **not** included in the Writing/English Composition CDT.

CHAPTER THREE: GENERAL CLASSROOM DIAGNOSTIC TOOLS TEST DEVELOPMENT PROCESSES

The operational item pool for each Classroom Diagnostic Tool (CDT) subject is made up of multiple-choice items that were field tested in a stand-alone field test administration in addition to a smaller number of multiple choice, evidence-based selected-response (Reading only), and technology-enhanced (Science only) items embedded later in operational assessments. Due to the large number of items needed for each CDT Computer Adaptive Test (CAT) to provide reliable information about student strengths and areas of need, it was decided to stagger the content areas for both development and field testing. Appendix A shows a graphic representation of the basic process flow and overlap of the development cycles.

Mathematics (comprising Mathematics, Algebra I, Algebra II, and Geometry) was developed first. After initial development and internal reviews by DRC, the items were taken to be reviewed by Pennsylvania educators. Upon completion of the educator reviews, edits were incorporated, and items were placed into online field-test fixed-forms for a stand-alone, voluntary field test. For more information regarding the field test, see Chapter Six. After the field test, item statistics were reviewed, and those items that had questionable data were taken to an item data review with Pennsylvania educators. See Chapter Six for more information about this meeting. Following the item data review, all items administered during the field test were reviewed by a committee of Pennsylvania educators for alignment to the Learning Progression Maps. More information about this meeting is found later in this chapter. After the alignment review, committees of Pennsylvania educators participated in a benchmarking activity to determine the points on the scale at which students in each of grades 5 through high school could be considered solidly ready for the next course. For more information about the benchmarking process, see Chapter Ten. Following this set of meetings, the statuses of items were updated, and accepted items were included in the item pool for the operational administrations.

This same process was then repeated for Literature (comprising Reading and Literature) and for Science (comprising Science, Biology, and Chemistry), and then finally for Writing (comprising Writing and English Composition). See Appendix A for more information about the basic development cycles for these three subjects.

Additional items in Mathematics and Reading/Literature were developed for an embedded field test in spring 2013. The purpose of this development was to supplement the pool with additional items aligned to the Pennsylvania Core Standards in preparation for the transition to align all Mathematics and Literacy (Reading/Literature and Writing/English Composition) assessments with the Pennsylvania Core Standards. Following the field test, the items that had questionable data were taken to an item data review with Pennsylvania educators (more information about this meeting can be found in Chapter Six). Following the item data review, all items administered during the field test were reviewed by a committee of Pennsylvania educators for alignment to the Learning Progression Maps using the same procedure that was used for the initial development of each pool of items.

In fall 2013, a voluntary stand-alone field test was conducted for items aligned to the Mathematics and English Language Arts (Reading and Writing) Pennsylvania Core Standards in kindergarten through grade 2, the K–2 Science Assessment Anchors and Eligible Content, and the Mathematics, English Language Arts, and Science Assessment Anchors and Eligible Content for grades 3 and 4. These were administered to students in grades 3 through 5, as described in Chapter Six. At the same time, items developed to align to the Mathematics, English Language Arts, and Science Assessment Anchors and Eligible Content for grade 5 were administered as part of an embedded field test to students in grade 6 that completed an operational CDT administration. The purpose of these two field test administrations was to provide enough items to allow students in grades 3 through 5 to be included in the CDT assessments. The Mathematics Lower Grades, Reading Lower Grades, Science Lower Grades, and Writing Lower Grades assessments became available in spring 2014.

Additional items were developed in 2015 for an embedded field test in 2016. The purpose of this development was to supplement the pool with additional items including Evidence Based Selected Response (EBSR) items aligned to the Pennsylvania Core Standards for the reading/literature CDT. These EBSR items were developed to align to the English Language Arts Assessment Anchors and Eligible Content for grades 3 through 8 and were administered as part of an embedded field test to students that completed an operational CDT administration. Additional multiple-choice items were also field tested in mathematics and science.

An additional set of items in were developed in 2018 for an embedded field test in 2018. The purpose of this development was to supplement the pool with additional items in mathematics, English language arts and science. These items were aligned to the Mathematics and English Language Arts (Reading and Writing) Pennsylvania Core Standards in kindergarten through grade 2, the K–2 Science Assessment Anchors and Eligible Content, and the Mathematics, English Language Arts, and Science Assessment Anchors and Eligible Content. The additional items made for a more robust pool of items from which the Diagnostic Category assessments and the full CDT could draw.

An additional set of items were developed in 2018 and 2019 for science. These items were aligned to the Science Assessment Anchors and Eligible Content. All additional items were technology-enhanced items meant to increase the rigor of the science pool as well as provide alternative ways to assess various science concepts. The additional items made for a more robust pool of science items from which the Diagnostic Category assessments and the full CDT could draw.

ITEM DEVELOPMENT CONSIDERATIONS

Alignment to the PSSA and Keystone Assessment Anchors and Eligible Content, grade- or 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* (AERA, APA, NCME, 1999) and the *Principles of Universal Design* (Thompson, Johnstone, & Thurlow, 2002) guided the development process. In addition, DRC's *Bias, Fairness, and Sensitivity Guidelines* were 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 employs 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 for issues of bias, fairness, and sensitivity (as well as for 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 includes 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, regional/geographic, ethnic/cultural, socioeconomic/class, religious, experiential, 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.

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 Classroom Diagnostic Tools. 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.
- 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 found in Chapter Four.

DEPTH OF KNOWLEDGE (DOK) OVERVIEW

An important element in statewide assessments 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 depth of knowledge (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" (p. 7–8). The four levels of cognitive complexity (i.e., depths of knowledge) are as follows:

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

Depth-of-knowledge levels were incorporated in the item writing and review process, and items were coded with respect to the level each represented.

PASSAGE READABILITY OVERVIEW

Evaluating the readability of a passage is essentially a judgmental process by individuals familiar with the classroom context and what is linguistically appropriate. 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.

TEST ITEM READABILITY OVERVIEW

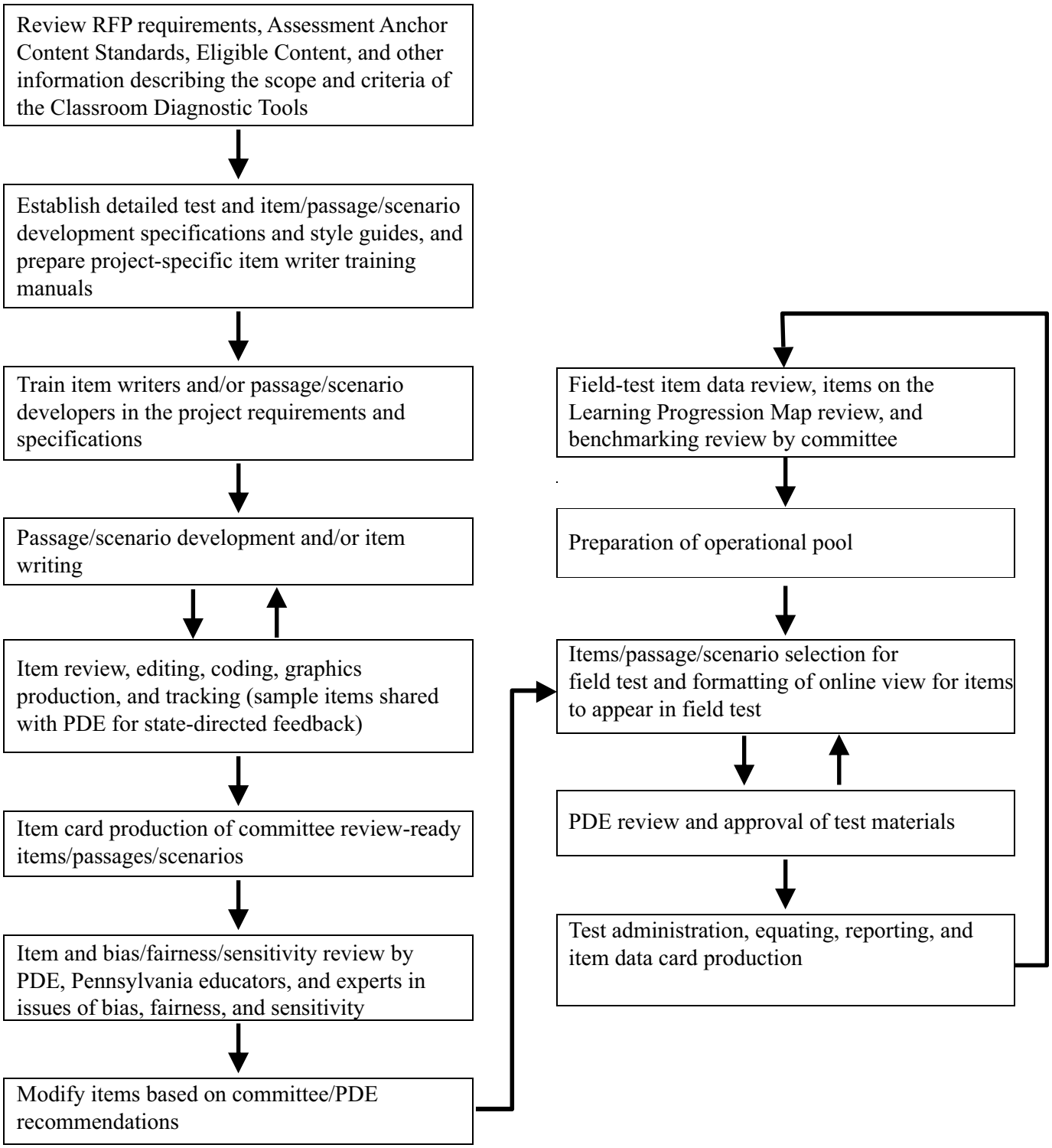
Careful attention was given to the readability of the items to make certain that the assessment focus of the item did not shift based on the difficulty of reading the item. Subject/course areas such as Mathematics, Algebra I, Science, or Biology contain many content-specific vocabulary terms. As a result, 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 grade or course level for non-Reading/Literature items. There was a conscious consideration made to ensure that each question was evaluating a student's ability to build toward mastery of the course standards versus 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 committees comprised 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, and they determine the validity of the vocabulary used and work to minimize the level of reading required.

ITEM AND TEST DEVELOPMENT CYCLE

The item development process for items followed a logical cycle and timeline, which is outlined in the figure on the following page. 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 lead to the field test data review and operational test construction. This presentation represents a typical life cycle for a field test event.

DRC Item and Test Development Primary Cycle



GENERAL ITEM AND TEST DEVELOPMENT PROCESS

The following describes the processes which lead up to an operational assessment. These processes were used to develop the entire pool of items that appeared within the field test administrations for potential inclusion in the operational item pool.

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, including the test blueprint, the field test plan (including development counts), procedures, timelines, etc.

ITEM WRITER TRAINING

Item writers were selected and trained for the subject areas of Mathematics, Algebra I, Algebra II, Geometry, Science, Biology, Chemistry, Reading, Literature, Writing, and English Composition. Qualified writers 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 multiple-choice items. Prior to developing items for the Classroom Diagnostic Tools, the cadre of item writers was trained with regard to the following:

- PSSA and Keystone Assessment Anchors and Eligible Content
- Webb's Levels of Cognitive Complexity, Depth of Knowledge
- Bias, Fairness, and Sensitivity Guidelines
- Principles of Universal Design
- Item Quality Technical Style Guidelines
- Reference Information
- Sample Items

LITERATURE PASSAGE DEVELOPMENT

The task of developing passages was conducted by DRC professionals with classroom experience in reading/English language arts. These professionals also underwent specialized training (provided by DRC) in the characteristics of acceptable passages. Guidelines for passage development 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 writers were given the task of writing a specified number of passages for each genre. Passages were commissioned by experienced authors.

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 able to stand the test of time.
- Passages are sufficiently rich to generate a variety of multiple-choice items.
- 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, those passages deemed potentially acceptable were reviewed by the Reading Content Committee and Bias, Fairness, and Sensitivity Committee for final approval.

ITEM AUTHORING AND TRACKING

Initially, items are generated with software-prepared Classroom Diagnostic Tools Item Cards, which allows for preliminary sorting and reviewing. 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, depth of knowledge (cognitive complexity), estimated difficulty, answer key, 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 within a test form. 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).

INTERNAL 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, knowledge of the content area curriculum, and 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 difficulties and cognitive complexities.

As part of the item construction process, each item was reviewed by content specialists and editors at DRC. Content specialists and editors evaluated each item to make sure that it measured the intended Eligible Content and Assessment Anchor. They also assessed each item to make certain that it was appropriate for the intended grade and that it provided only one correct answer. In addition, the difficulty level, depth of knowledge, graphics, language demand, and distractors were also evaluated. Other elements considered in this process include, but are not limited to, Universal Design, bias, source of challenge, grammar/punctuation, and Pennsylvania style. Following these reviews, the items were prepared for the content review meetings conducted with Pennsylvania educators.

ITEM CONTENT REVIEWS

Prior to the 2010, 2011, 2013, 2015, 2018 and 2019 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 or course appropriateness, estimated difficulty, depth of knowledge, and source of challenge. With source of challenge, items are identified where the cognitive demand is focused on an unintended content, concept, or skill (Webb, 2002). In addition, source of challenge may be attributed 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. They also suggested revisions and made recommendations for reclassification of items. The committee members also reviewed the items for adherence to the Principles of Universal Design, including language demand and issues of bias, fairness, and sensitivity.

The content review meetings were held in January 2010 for Mathematics, Algebra I, Algebra II, and Geometry, in May/June 2010 for Reading/Literature, Science, Biology, and Chemistry, and in January 2011 for Writing/English Composition. Additional content review meetings were held in November 2012 (for the additional items aligned to the Pennsylvania Core Standards) and in July 2013 (for the items to allow students in grades 3 through 5 to participate in the CDT). Content review meetings were again held in May of 2015 for Writing items and June of 2015 for Science, Reading, and Math (for additional items aligned to the Pennsylvania Core Standards and the Assessment Anchors and Eligible Content to supplement the pool). Another set of content review meetings took place in January of 2018 to supplement the item pool. The most recent content review meetings were held in January of 2019 for Science technology enhanced items. 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 content area, received training by working through and reviewing a group of items for quality and content, as well as for the following categories:

- Assessment Anchor Alignment
- Content Limits
- Grade-Level (Course-Level) Appropriateness
- Difficulty Level
- Depth of Knowledge
- Appropriate Source of Challenge
- Correct Answer
- Quality of Distractors
- Graphics in Regards to Appropriateness
- Appropriate Language Demand
- Freedom from Bias

The members then received a binder containing items to independently review and provided their recommendation for the status of each item: Approved, Accepted with Revision, or Rejected. All comments were reviewed and addressed by DRC content staff, and, when necessary, PDE staff were consulted.

Security was addressed by adhering to a strict set of procedures. 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

Prior to the 2010, 2011, 2013, 2015, 2018 and 2019 field testing, all newly developed test items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. These reviews took place prior to the Item Content Review for each content area. 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, 2003–2013). Members of the committee also had expertise with special-needs students and English Language Learners. 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 a consensus as to which issues 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 DRC content staff. This review followed the same security procedures as outlined above.

ITEMS ALIGNED TO LEARNING PROGRESSION MAPS

Following the field test of items, all items were brought before a committee of Pennsylvania educators for review of each item's alignment to the Learning Progression Map. DRC and PDE provided a general overview of the item and test development process for the Classroom Diagnostic Tools and provided information about the Learning Progression Maps and the purpose of the Classroom Diagnostic Tools. Then the committee reviewed the Learning Progression Map, which shows the vertical articulation of the Assessment Anchors and Eligible Content across grades within a given subject area. Once it was determined that the Learning Progression Map containing the Assessment Anchors and Eligible Content was an accurate representation of how the content progressed across grades, teachers worked in grade-span committees to review items for their alignment with the Assessment Anchor and Eligible Content. When reviewing the alignment to the Assessment Anchor and Eligible Content, educators considered whether the test item measured the content that it purported to measure, as well as the appropriateness of the difficulty and cognitive complexity of the item in relation to the Assessment Anchor and Eligible Content to which the item was aligned. Committees came to a consensus regarding the status of each item: Accepted, Accepted with Revised Alignment, or Rejected.

Security was addressed by adhering to a strict set of procedures. 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.

CHAPTER FOUR: UNIVERSAL DESIGN PROCEDURES APPLIED TO THE CLASSROOM DIAGNOSTIC TOOLS 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 were designed and developed using the elements of universally designed assessments developed by the National Center for 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 students with disabilities and English Language Learners. The benefits of universally designed assessments not only apply to these groups of students, but to all individuals with wide-ranging characteristics. Therefore, it is important that the development of all assessments, including voluntary assessments such as the Classroom Diagnostic Tools, be guided by the Principles of Universal Design.

DRC's test development team was trained in the elements of Universal Design as it relates 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 followed during all stages of the item development process for the Classroom Diagnostic Tools.

ELEMENTS OF UNIVERSALLY DESIGNED ASSESSMENTS

After a review of research relevant to the assessment development process and the Principles of Universal Design (Center for Universal Design, 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 Classroom Diagnostic Tools.

- **Inclusive Assessment Population**

  The target population includes students attending Commonwealth schools in grades 3 through 12 who will be participating in either the Pennsylvania System of School Assessment or the Keystone Exams.

- **Precisely Defined Constructs**

  An important function of well-designed assessments is that they actually measure what they are intended to measure. The Assessment Anchor Content Standards and Eligible Content for both PSSA and the Keystone Exams, as well as the Pennsylvania Academic Standards for Writing, provided clear descriptions of the constructs to be measured by the Classroom Diagnostic Tools assessments. Universally designed assessments must remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers.

- **Accessible, Non-biased Items**

  DRC conducted both internal and external reviews of items and test specifications to ensure that they did not create barriers because of lack of sensitivity to disability, culture, or other subgroups. Items and test specifications were developed by a team of individuals who understand 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.

- **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, text organization, and others. 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 Classroom Diagnostic Tools 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 how to direct 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 intended to assess. A style guide was developed and was utilized which 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 attended to.

1. **Items measure what they are intended to measure.** Item writing training included ensuring that writers and reviewers had a clear understanding of Pennsylvania's Core Standards, Pennsylvania's Academic Standards, and the PSSA and Keystone Assessment Anchors and Eligible Content. During all phases of test development, items were presented with content-standard information to ensure that each item reflected the intended Academic Standard (Mathematics, Reading, and Writing items aligned to Kindergarten, grade 1, or grade 2) or Eligible Content (all other grades and content areas). Careful consideration of the content standards was important in determining which skills involved in responding to an item were extraneous and which were relevant to what was being tested. In certain types of items an additional skill is 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 are presented to students must allow for maximum readability for all students. Appropriate fonts and point sizes were employed with minimal use of italics, which is far less legible and is 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 around paragraphs and between columns and staggered right margins were used.

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 where possible. Sufficient contrast between background and text, with minimal use of shading, increased readability for students with visual impairments. Color was not used to convey important information.

5. **Items have concise and readable text.** Linguistic demands of stimuli and items can interfere with a student's ability to demonstrate knowledge of the construct being assessed. During item writing and review, the following guidelines were used.

    - Simple, clear, commonly used words were used whenever possible.
    - Extraneous text was omitted.
    - Vocabulary and sentence complexity were appropriate for the grade level being assessed.
    - Technical terms and abbreviations were used only if they were related to the content being measured.
    - Definitions and examples were clear and understandable.
    - Idioms were avoided unless idiomatic speech was being assessed.
    - The questions to be answered were clearly identifiable.

6. **Items allow changes to format without changing meaning or difficulty.** An audio accommodation is available in Mathematics Lower Grades, Mathematics, Algebra I, Geometry, Algebra II, Science Lower Grades, Science, Biology, and Chemistry for any student with Individualized Education Program (IEP) requirements related to receiving audio assistance during testing. Additionally, a Magnifier tool that can be used to enlarge an area of the screen is available to all students. This tool can be used at the same time as other tools, such as the Highlighter or Line Guide.

7. **The test has an overall appearance that is clean and organized.** 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. Information was organized in a left-right, top-bottom format.

ITEM DEVELOPMENT

DRC works closely with the Pennsylvania Department of Education to help ensure that the Classroom Diagnostic Tools comply with nationally recognized Principles of Universal Design. In addition to the Principles of Universal Design as described in the Classroom Diagnostic Tools Technical Report, DRC applies 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 ensures that committee members at item and bias reviews are made aware of the Principles of Universal Design and of issues that may adversely affect students with disabilities with the goal of ensuring that Classroom Diagnostic Tools assessments are bias-free for all students.

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

ITEM FORMAT

For all Classroom Diagnostic Tools assessments, DRC formats the items to maximize accessibility for all students by using text that is in a size and font style that is easily readable. DRC limits shading, graphics, and charts. 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 where possible to convey pertinent information.

DRC ensures consistency across Classroom Diagnostic Tools assessments 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 where 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.

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 ensure 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. At this time, an audio accommodation is available in Mathematics Lower Grades, Mathematics, Algebra I, Geometry, Algebra II, Science Lower Grades, Science, Biology, and Chemistry for any student with Individualized Education Program (IEP) requirements related to receiving audio assistance during testing. A separate audio accommodation is available for all CDT assessments for students with visual impairments. Additionally, a color choices accommodation allows students who would benefit from a background other than white to select a background color from five available choices (in addition to the white background). A contrasting color allows students who would benefit from different text and background color combinations to select from seven options (in addition to black text on a white background).

CHAPTER FIVE: TEST ADMINISTRATION PROCEDURES

TEST SETUP

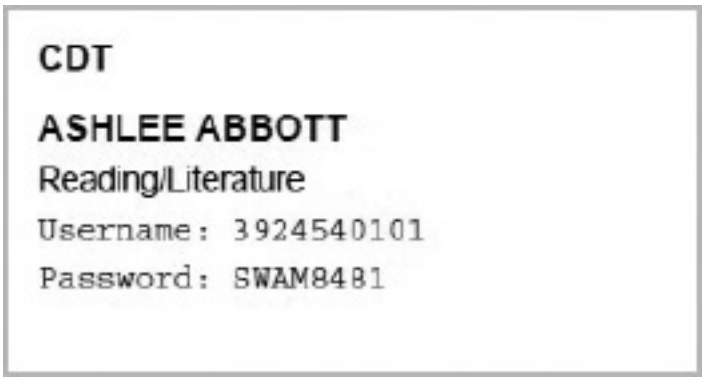
The process to set up students to take the Classroom Diagnostic Tools (CDT) is accomplished through an online interface located on the DRC INSIGHT Portal (https://www.drcedirect.com/all/eca-portal-ui/welcome/PA). The DRC INSIGHT Portal is a permission-based site that enables districts to assign users different roles and permissions depending on their role in the setup process. Each district can set up users with as much or as little permission as deemed necessary. A user's role and permission may be modified at any time.

The student and teacher information can be imported into the Portal at any time. Once the data is imported, users organize students into student groups and test sessions. Student groups and test sessions can be created by class, grade, school, or any other variation.

Each student group is assigned to a specific teacher. Students may belong to multiple student groups and multiple teachers can be assigned to the same student group. This allows districts/schools the ability to allow multiple users to view the data by class, grade, or even school. Student groups may be created and modified at any time during the administration window.

Test sessions are generated to create test tickets that are distributed to students prior to testing. A test ticket contains the student's full name, user name, password, and the assessment he/she will be taking. The test session, like the student group, may also be created by class, grade, and school. Each time an assessment is administered, a new test session must be created. Test sessions can be copied to simplify administering the CDT to the same students multiple times each year.

SAMPLE TEST SESSION TICKET


CDT

ASHLEE ABBOTT

Reading/Literature

Username: 3924540101

Password: SWAM8481

The CDT is untimed. Each full CDT should take the typical student 50 to 90 minutes to complete and is between 48 and 60 items in length. Each Diagnostic Category (DC) CDT should take the typical student 20–30 minutes to complete. The writing, science and math Diagnostic Category CDTs are between 15–18 items. The reading Diagnostic Category CDTs are between 35–45 items. The CDT may be administered in one sitting, but it is possible to administer the CDT over multiple days and recommended for the Grades 3–5 assessments.

Teachers have flexibility in using the different full and diagnostic category tests within a school year. For instance, some elementary teachers may choose to use the full mathematics CDT at the beginning of the year to understand where their students are starting, and follow-up with DC tests as they go through different units. High school teachers may choose a DC test first, based on the course or unit of study. Regardless of how the CDT is used in the classroom, there should be enough time between CDT administrations to allow for instructional impact to be reflected in the student's results. Though there are no restrictions on the time between CDTs, there is a restriction in the Test Setup system that only allows a student to be associated with a single CDT/DC CDT a maximum of five (5) times within a given school year.

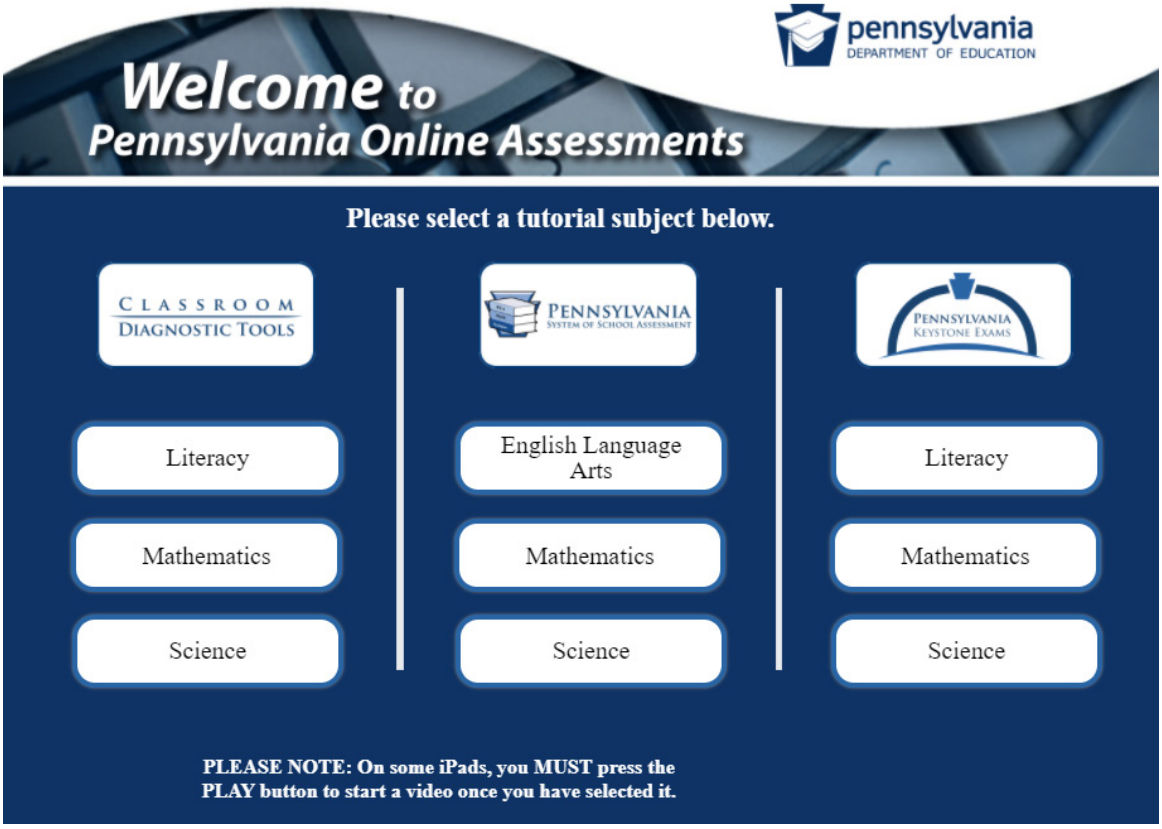
PA ONLINE ASSESSMENTS SOFTWARE

Prior to testing, each student computer needs to have the PA Online Assessments software installed. The testing software downloads are located on the DRC INSIGHT Portal. The installer is an MSI file that can be pushed out across a server to expedite the installation process. Once the software is installed, users also have access to the PA Assessment Online Student Tutorials and the PA Assessment Online Tools Training (OTT).

The PA Assessment Online Tools Training (OTT) is designed to provide an introductory experience using the online assessment software in preparation for taking the CDT. The purpose of the OTT is for students to observe and experiment with the features of the online assessment software prior to the actual assessment. The OTT is NOT designed to demonstrate complete coverage of the tested content, and it is NOT scored. Rather, sample items have been chosen to demonstrate online assessment features and uses.

Technology coordinators are encouraged to run the Online Tools Training prior to testing because it interacts with DRC servers exactly like an actual CDT assessment. Completion of the OTT will provide a good indication that the software installed correctly, and everything is configured properly on the network.

The web-based PA Online Assessment Student Tutorials are available for each operational assessment and are designed to be used by students at all grade levels. They use pictures, motion, and sound to present visual and verbal descriptions of the features and functionality of the PA Online Assessment system. It is recommended to allow a minimum of 20 minutes to view the tutorials. Tutorials may be reviewed as often as needed.



TRAINING AND CUSTOMER SERVICE SUPPORT

Prior to testing, training was provided to District Technology Coordinators and District Assessment Coordinators. All training was administered via web conference and lasted approximately 1½ hours. Test Coordinator Training goes over tasks that need to be completed prior to testing. A large portion of the training is dedicated to the setup of users and the creation of student groups and test sessions.

Technology Coordinator Training focuses on all technical aspects required for the setup of the CDT. Detailed installation instructions for the PA Online Assessments Software and Central Office Services – Service Device

(COS-SD) are provided. The COS-SD runs on a server within the local network and helps mitigate internet traffic by allowing student machines to retrieve items from the COS-SD rather than from DRC servers. The CDT requires an internet connection at all times.

Student Interface System Requirements

Windows / Linux Installer System Requirements

- 4 GB of RAM recommended
- Screen resolution of 1024 X 768 or higher
- Mouse; Keyboard
- 20 GB of available hard disk space or greater
- Dual-core i5 at 2 GHz or equivalent

Window/Linus Supported Operating Systems

- Windows Vista, SP2
- Windows 7, SP1
- Windows 8 (including 8.1)
- Windows 10 (version 1507, 1511, and Redstone 1)
- Windows Server 2008 (SP2, R2, SP1)
- Windows Server 2012 (R2)
- Ubuntu (12.04 and 12.04) LTS version, with 32- and 64-bit Gnome 3.4, Unity Shell

Macintosh Installer System Requirements

- 4 GB of RAM recommended
- Screen resolution of 1024 X 768 or higher
- Mouse; Keyboard
- 20 GB of available hard disk space or greater
- Dual-core i5 at 2 GHz or equivalent

Supported Operating Systems

- Apple® Mac OS X® 10.7
- Apple® Mac OS X® 10.8
- Apple® Mac OS X® 10.9
- Apple® Mac OS X® 10.10
- Apple® Mac OS X® 10.11

Chrome OS Installer System Requirements

- 4 GB of RAM or more
- Screen resolution 1024 x 768 or higher
- 2 GHz or faster processor

Chrome OS Supported Operating Systems

- Chrome OS recent stable channel

Apple iOS Installer System Requirements

- Screen resolution of 1024 x 768 or higher

Apple iOS Supported Operating System

- 9.3.x

Android Installer System Requirements

- Screen resolution of 1024 x 768 or higher

Android Supported Operating System

- Lollipop 5.x

Users are encouraged to call or email with any questions or error messages that cannot be resolved. If the problem cannot be resolved via a customer service representative, the issue is escalated to DRC developers. Ninety percent of the time, a solution is provided within twenty-four hours. If the issue requires more research, DRC will contact the caller daily to provide an update.

CHAPTER SIX: FIELD TEST

FIELD TEST OVERVIEW

All items appearing in the 2020–2021 Classroom Diagnostic Tools (CDT) operational item pools were field tested prior to their use on the operational CDT. The purpose of administering field-test items is to obtain statistics for them so they can be reviewed and approved before becoming operational. Based on this statistical review, many of the field-test items were selected for use in the 2020–2021 CDT operational item pools.

There were nine separate CDT field-test events that contributed items to the 2020–2021 operational item pools—four stand-alone field-test events and five embedded field-test events. Separate field-test events were needed because the operational CDT was rolled out in phases by content area and available grades.

There were three stand-alone field-test events to build the item pools for students in grade 6 and above. Items in mathematics were field tested in spring 2010. Items in reading and science were field tested in fall 2010. Items in writing were field tested in spring 2011. During these three field-test events, CDT items were field tested on stand-alone fixed forms. The forms were administered in computer-based format only. No paper/pencil versions were available. Field test administration mode was limited to computer-based to mirror the operational CDT, which is an adaptive test requiring computer administration. CDT stand-alone field tests were designed to build vertical scales across all grades and courses within a content area. In order to accomplish this, some field-test forms had items from one grade above or below in addition to on-grade level items. For example, some grade 7 mathematics forms contained items from grade 6 in addition to items from grade 7. Other grade 7 mathematics forms contained items from both grade 7 and grade 8. See Chapter Nine for more details.

There was one stand-alone field-test event to build the item pools for students in grades 3 through 5. Items in mathematics, reading, science, and writing were field tested in fall 2013. Again, CDT items were field tested on stand-alone fixed forms. The forms were administered in computer-based format only. No paper/pencil versions were available. In order to link to the existing operational scales, some operational grade-level items were included in the field-test forms. See Chapter Twelve for more details.

In addition to the four stand-alone field-test events that contributed items to the 2020–2021 operational item pools, there were five field-test events in which a small number of field-test items were included (embedded) within the operational CDT. In spring 2013, field-test items were included in mathematics and reading. The purpose of this embedded field test was to add items to the operational item pools that align to the Pennsylvania Core Standards. In fall 2013, field-test items were included in mathematics, reading, science, and writing. The purpose of this embedded field test was to field test additional items in grade 5 that could be used in the item pools for students in grades 3 through 5. In 2015–2016, seven of the thirteen CDTs included a small number of embedded field-test items. The purpose of this embedded field test was to supplement the existing item pools and to introduce the evidence-based selected-response (EBSR) item type in the reading content area. In 2018–2019, all CDTs included a small number of embedded field-test items. The purpose of this embedded field test was to supplement the existing item pools in all content areas and grades/courses. In 2019–2020, all CDTs in the science content area except Chemistry included a small number of embedded field-test items in addition to the operational items used to generate a student's score. The purpose of the embedded field test was to supplement the existing item pools and to introduce the technology-enhanced (TE) item type.

In the case of all five embedded field-test events, field-test items were included within the operational administration and students did not know which items were field-test items (items that do not count toward a student's score). Therefore, the embedded field-test items can be linked to the existing operational scales. See Chapter Twelve for details.

CDT STAND-ALONE FIELD TESTS

SPRING 2010—MATHEMATICS

The stand-alone field test administered in spring 2010 was designed to yield enough items to populate the item pool for CDT Mathematics. Items covering the Eligible Content in grades 3 through 8 and courses Algebra I, Geometry, and Algebra II were field tested. Items covering grade 11 Eligible Content that were NOT covered in Algebra I, Geometry, or Algebra II were also field tested.

Participation in the field test was voluntary. All schools that wanted to participate were allowed to field test. All students in volunteer schools were encouraged, but not required, to participate.

In order to encourage participation, field-test forms were limited in length. Forms for grades 3, 4, and 5 had 25 items. Forms for grades 6, 7, and 8 and Algebra I, Geometry, and Algebra II courses had 35 items. There were not separate grade 11 forms. Instead, grade 11 items were included on grade 8, Algebra I, Geometry, and Algebra II forms.

Since testing occurred in spring, students had nearly a full year of instruction. Therefore, grade-level forms were assigned to students in the corresponding grade (e.g., students in grade 7 took grade 7 forms). Course-level forms were assigned to students currently taking the course (e.g., students in a Geometry course took Geometry forms).

Each student was randomly assigned one of the appropriate grade- or course-level forms at the time of testing.

Table 6–1. Spring 2010 Mathematics Field-Test Form Details

Grade/Course	Number of Items	Number of Forms	Number of Vertical Linking Forms
Grade 3	86	8	4
Grade 4	86	10	8
Grade 5	85	10	8
Grade 6	259	16	8
Grade 7	258	16	8
Grade 8	257	18	12
Grade 11*	149	0	0
Algebra I	256	18	8
Geometry	257	16	4
Algebra II	256	16	4

* Grade 11 items were tested on grade 8, Algebra I, Geometry, and Algebra II forms.

FALL 2010—READING AND SCIENCE

The stand-alone field tests administered in fall 2010 were designed to yield enough items to populate the item pools for CDT Reading/Literature and CDT Science. Reading items covering the Eligible Content in grades 3 through 8 and Literature were field tested. Science items covering the Eligible Content in grades 3 through 8 and Biology and Chemistry courses were field tested. Items covering grade 11 science Eligible Content that were NOT covered in Biology or Chemistry were also field tested.

Participation in the field test was voluntary. All schools that wanted to participate were allowed to field test. All students in volunteer schools were encouraged, but not required, to participate. Schools were allowed to field test in both content areas.

In order to encourage participation, field-test forms were limited in length. Forms for grades 3, 4, and 5 had 25 items. Forms for grades 6, 7, and 8 and Literature, Biology, and Chemistry courses had 35 items. There were not separate grade 11 science forms. Instead, grade 11 science items were included on grade 8 science forms.

Since testing occurred in fall, students did NOT have a full year of instruction at their current grade level. Grade-level forms were therefore assigned one grade lower (e.g., students in grade 7 took grade 6 forms). Course-level forms were assigned to students who had completed the course during the prior school year.

Each student was randomly assigned one of the appropriate grade- or course-level forms at the time of testing.

Table 6–2. Fall 2010 Reading/Literature Field-Test Form Details

Grade/Course	Number of Items	Number of Forms	Number of Vertical Linking Forms
Grade 3	86	7	2
Grade 4	87	8	4
Grade 5	86	8	4
Grade 6	210	10	4
Grade 7	192	9	4
Grade 8	192	9	4
Literature	348	15	2

Table 6–3. Fall 2010 Science Field-Test Form Details

Grade/Course	Number of Items	Number of Forms	Number of Vertical Linking Forms
Grade 3	91	7	2
Grade 4	123	11	4
Grade 5	102	9	4
Grade 6	178	9	4
Grade 7	327	15	4
Grade 8	377	22	6
Grade 11*	115	0	0
Biology	390	16	2
Chemistry	335	14	2

* Grade 11 items were tested on grade 8 forms.

SPRING 2011—WRITING

The stand-alone field test administered in spring 2011 was designed to yield enough items to populate the item pool for CDT Writing/English Composition. Items covering the Pennsylvania Academic Standards for Writing in grades 3 through 8 and the Eligible Content for English Composition were field tested.

Participation in the field test was voluntary. All schools that wanted to participate were allowed to field test. All students in volunteer schools were encouraged, but not required, to participate.

In order to encourage participation, field-test forms were limited in length. Forms for grades 3, 4, and 5 had 25 items. Forms for grades 6, 7, and 8 and English Composition had 35 items.

Since testing occurred in spring, students had nearly a full year of instruction. Therefore, grade-level forms were assigned to students in the corresponding grade (e.g., students in grade 7 took grade 7 forms).

Each student was randomly assigned one of the appropriate grade- or course-level forms at the time of testing.

Table 6–4. Spring 2011 Writing/English Composition Field-Test Form Details

Grade/Course	Number of Items	Number of Forms	Number of Vertical Linking Forms
Grade 3	140	10	2
Grade 4	149	12	4
Grade 5	165	13	4
Grade 6	193	9	4
Grade 7	176	9	4
Grade 8	195	9	4
English Composition	365	15	2

FALL 2013—MATHEMATICS, READING, SCIENCE, AND WRITING

The stand-alone field tests administered in fall 2013 were designed to yield enough items to populate the item pools for each CDT for students in grades 3 through 5 in mathematics, reading, science, and writing. Items covering the Eligible Content in kindergarten through grade 4 were field tested[1]. In order to link to the existing operational scales, some operational grade-level items were included in the field-test forms.

Participation in the field test was voluntary. All schools that wanted to participate were allowed to field test. All students in volunteer schools were encouraged, but not required, to participate. Schools were allowed to field test in all content areas. In order to encourage participation, field-test forms were limited in length. All field-test forms had 25 items.

Since testing occurred in fall, students did NOT have a full year of instruction at their current grade level. Grade-level forms were therefore assigned one grade lower (e.g., students in grade 4 took forms containing grade 3 items). Each student was randomly assigned one of the appropriate grade-level forms at the time of testing.

[1] Items in grade 5 were part of the fall 2013 embedded field test.

Table 6–5. Fall 2013 Mathematics Field-Test Form Details

Student Grade	Item Grade(s)	Item Type	Number of Items	Number of Forms
Grade 3	K, 1, 2	Field Test	60, 90, 130	14
Grade 3	3	Link to Op Scale	15	14
Grade 4	3	Field Test	235	12
Grade 4	3	Link to Op Scale	15	12
Grade 5	4	Field Test	248	13
Grade 5	4	Link to Op Scale	15	13

Table 6–6. Fall 2013 Reading Field-Test Form Details

Student Grade	Item Grade(s)	Item Type	Number of Items	Number of Forms
Grade 3	K, 1, 2	Field Test	84, 98, 98	14
Grade 3	3	Link to Op Scale	15	14
Grade 4	3	Field Test	178	9
Grade 4	3	Link to Op Scale	15	9
Grade 5	4	Field Test	189	10
Grade 5	4	Link to Op Scale	15	10

Table 6–7. Fall 2013 Science Field-Test Form Details

Student Grade	Item Grade(s)	Item Type	Number of Items	Number of Forms
Grade 3	K–2 grade span	Field Test	280	14
Grade 3	3	Link to Op Scale	15	14
Grade 4	3	Field Test	155	8
Grade 4	3	Link to Op Scale	15	8
Grade 5	4	Field Test	213	11
Grade 5	4	Link to Op Scale	15	11

Table 6–8. Fall 2013 Writing Field-Test Form Details

Student Grade	Item Grade(s)	Item Type	Number of Items	Number of Forms
Grade 3	K, 1, 2	Field Test	44, 118, 117	14
Grade 3	3	Link to Op Scale	15	14
Grade 4	3	Field Test	60	3
Grade 4	3	Link to Op Scale	15	3
Grade 5	4	Field Test	60	3
Grade 5	4	Link to Op Scale	15	3

CDT EMBEDDED FIELD TESTS

SPRING 2013—MATHEMATICS AND READING

The embedded field test administered in spring 2013 was designed to augment the existing mathematics and reading/literature item pools. Items were aligned to the Pennsylvania Core Standards. Starting on February 14, 2013, all students testing CDT Mathematics took 5 field-test items. All students testing CDT Reading/Literature took 5–7 field-test items, depending on passage length. Students did not know which items were operational and which were field test. Field-test items did not count in calculation of total or diagnostic category scores. Since testing occurred in spring, students had received nearly a full year of instruction. Therefore, grade-level items were assigned to students in the corresponding grade wherever possible.

Table 6–9. Spring 2013 Embedded Field Test Details

Content Area	Grade/Course	Number of Items
Mathematics	Grade 3*	56
Mathematics	Grade 4*	67
Mathematics	Grade 5*	41
Mathematics	Grade 6	156
Mathematics	Grade 7	73
Mathematics	Grade 8	157
Reading	Grade 3*	58
Reading	Grade 4*	71
Reading	Grade 5*	60
Reading	Grade 6	56
Reading	Grade 7	58
Reading	Grade 8	57

*Items in grades 3 through 5 were initially field tested with students in grade 6 because CDT is available to students in grade 6 and above. However, this plan was revised after a few weeks of testing in favor of stand-alone field tests in fall 2013 with students in grades 3 through 5.

FALL 2013—MATHEMATICS, READING, SCIENCE, AND WRITING

The embedded field test administered in fall 2013 was designed to field test the grade 5 items needed to populate the item pools for each CDT for students in grades 3 through 5 in mathematics, reading, science, and writing. Starting on August 26, 2013, students in grade 6 testing CDT Mathematics, CDT Science, or CDT Writing/English Composition took 5 field-test items. Students in grade 6 testing CDT Reading/Literature took 5–7 field-test items, depending on passage length. Students did not know which items were operational and which were field test. Field-test items did not count in calculation of total or diagnostic category scores. Since testing occurred in fall, students had not received a full year of instruction. Therefore, grade 5 items were assigned to grade 6 students.

Table 6–10. Fall 2013 Embedded Field Test Details

CDT	Grade	Number of Items
Mathematics	Grade 5	221
Reading/Literature	Grade 5	134
Science	Grade 5	152
Writing/English Composition	Grade 5	71

FALL 2015—MATHEMATICS, READING, SCIENCE, AND WRITING

The embedded field test administered in fall 2015 was designed to field test new items to supplement the item pools in grades 6 and above in mathematics, reading, science, and writing as well as courses Algebra I and Biology. Additionally, the evidence-based selected-response item type was field tested in grades 3 through 8 reading.

Table 6–11. Fall 2015 Embedded Field Test Item Pools

Content Area	Item Grade/Course	Number of MC Items	Number of EBSR Items	Total Number of Items
Mathematics	6	122	0	122
Mathematics	7	177	0	177
Mathematics	8	151	0	151
Mathematics	Algebra I	150	0	150
Reading	3	0	22	22
Reading	4	0	22	22
Reading	5	0	22	22
Reading	6	105	21	126
Reading	7	105	21	126
Reading	8	105	21	126
Reading	Literature	126	0	126
Science	6	72	0	72
Science	7	159	0	159
Science	8	238	0	238
Science	Biology	136	0	136
Writing	6	93	0	93
Writing	7	93	0	93
Writing	8	110	0	110
Writing	English Composition	104	0	104

Starting on August 24, 2015, seven of the thirteen CDTs included embedded field-test items:

- Students using CDT Math Grades 6–HS, CDT Science Grades 6–HS, and CDT Writing/Eng Comp Grades 6–HS took 5 field-test items. Since testing occurred throughout the year, items were given to students whose grade matched the item's grade and to students one grade above the item's grade (e.g., grade 7 items were given to students in grades 7 and 8).
- Students using CDT Reading/Lit Grades 6–HS took one field-test passage and six associated items. Since testing occurred throughout the year, items were given to students whose grade matched the item's grade and to students one grade above the item's grade (e.g., grade 7 items were given to students in grades 7 and 8).
- Students using CDT Algebra I and CDT Biology took 5 field-test items from the relevant course.
- The only field-test items in grades 3 through 5 reading were EBSR items associated with existing operational passages. Students using CDT Reading Grades 3–5 were eligible to receive field-test EBSR items. However, operational passages that were not a good fit based on a student's performance were not administered just for the sake of field-test items. Instead, a field-test EBSR was administered only if the operational passage was selected for the student. The number of field-test EBSRs was limited to 3 per test.

In all cases, students did not know which items were operational and which were field test. Field test items did not count in the calculation of total or diagnostic category scores.

Table 6–12. Fall 2015 Embedded Field Test Design

Content Area	CDT	Item Grade/Course	Number of Items Embedded	Student Test Grade(s)
Mathematics	Math Grades 6–HS	6	5 MC	6,7
Mathematics	Math Grades 6–HS	7	5 MC	7, 8
Mathematics	Math Grades 6–HS	8	5 MC	8, 9+
Mathematics	Algebra I	Algebra I	5 MC	Algebra I
Reading	Reading Grades 3–5	3	0–3 EBSR	3,4,5
Reading	Reading Grades 3–5	4	0–3 EBSR	3,4,5
Reading	Reading Grades 3–5	5	0–3 EBSR	3,4,5
Reading	Reading/Lit Grades 6–HS	6	1 passage*	6,7
Reading	Reading/Lit Grades 6–HS	7	1 passage*	7, 8
Reading	Reading/Lit Grades 6–HS	8	1 passage*	8, 9+
Reading	Reading/Lit Grades 6–HS	Literature	1 passage*	9+
Science	Science Grades 6–HS	6	5 MC	6,7
Science	Science Grades 6–HS	7	5 MC	7, 8
Science	Science Grades 6–HS	8	5 MC	8, 9+
Science	Biology	Biology	5 MC	Biology
Writing	Writing/Eng Comp Gr 6–HS	6	5 MC	6,7
Writing	Writing/Eng Comp Gr 6–HS	7	5 MC	7, 8
Writing	Writing/Eng Comp Gr 6–HS	8	5 MC	8, 9+
Writing	Writing/Eng Comp Gr 6–HS	English Composition	5 MC	9+

* FT reading passages include six multiple-choice items OR five multiple-choice items and one evidence-based selected-response item.

FALL 2018—MATHEMATICS, READING, SCIENCE, AND WRITING

The embedded field test administered in fall 2018 was designed to field test new items to supplement the item pools in all content areas and grades/courses.

Table 6–13. Fall 2018 Embedded Field Test Item Pools

Content Area	Item Grade/Course	Number of MC Items	Number of EBSR Items	Total Number of Items
Mathematics	Kindergarten	20	0	20
Mathematics	1	20	0	20
Mathematics	2	20	0	20
Mathematics	3	178	0	178
Mathematics	4	179	0	179
Mathematics	5	180	0	180
Mathematics	6	96	0	96
Mathematics	7	103	0	103
Mathematics	8	99	0	99
Mathematics	Algebra I	299	0	299
Mathematics	Geometry	100	0	100
Mathematics	Algebra II	100	0	100
Reading	Kindergarten	32	0	32
Reading	1	20	0	20
Reading	2	32	0	32
Reading	3	135	27	162
Reading	4	135	27	162
Reading	5	135	27	162
Reading	6	102	21	123
Reading	7	102	21	123
Reading	8	99	21	120
Reading	Literature	249	0	249
Science	2	31	0	31
Science	3	89	0	89
Science	4	95	0	95
Science	5	90	0	90
Science	6	97	0	97
Science	7	99	0	99
Science	8	102	0	102
Science	Biology	290	0	290
Science	Chemistry	100	0	100

Table 6–13 (continued). Fall 2018 Embedded Field Test Item Pools

Content Area	Item Grade/Course	Number of MC Items	Number of EBSR Items	Total Number of Items
Writing	Kindergarten	10	0	10
Writing	1	10	0	10
Writing	2	12	0	12
Writing	3	99	0	99
Writing	4	90	0	90
Writing	5	90	0	90
Writing	6	93	0	93
Writing	7	111	0	111
Writing	8	93	0	93
Writing	English Composition	294	0	294

Starting on August 20, 2018, all CDTs included embedded field-test items:

- Students using grade level tests in content areas math, science, and writing took five field-test items. Since testing occurred throughout the year, items were given to students whose grade matched the item's grade and to students one grade above the item's grade (e.g., grade 7 items were given to students in grades 7 and 8).
- Students using CDTs in the reading content area took one field-test passage with four to six associated items. Since testing occurred throughout the year, items were given to students whose grade matched the item's grade and to students one grade above the item's grade (e.g., grade 7 items were given to students in grades 7 and 8).
- Students using CDT Algebra I, CDT Geometry, CDT Algebra II, CDT Biology, and CDT Chemistry took 5 field-test items from the relevant course.

In all cases, students did not know which items were operational and which were field test. Field test items did not count in calculation of total or diagnostic category scores.

Table 6–14. Fall 2018 Embedded Field Test Design

Content Area	CDT	Item Grade/Course	Number of Items Embedded	Student Test Grade(s)
Mathematics	Math Grades 3–5	Kindergarten	5 MC	3
Mathematics	Math Grades 3–5	1	5 MC	3
Mathematics	Math Grades 3–5	2	5 MC	3
Mathematics	Math Grades 3–5	3	5 MC	3,4
Mathematics	Math Grades 3–5	4	5 MC	4,5
Mathematics	Math Grades 3–5	5	5 MC	5,6
Mathematics	Math Grades 6–HS	6	5 MC	6,7
Mathematics	Math Grades 6–HS	7	5 MC	7, 8
Mathematics	Math Grades 6–HS	8	5 MC	8, 9+
Mathematics	Algebra I	Algebra I	5 MC	Algebra I
Mathematics	Geometry	Geometry	5 MC	Geometry
Mathematics	Algebra II	Algebra II	5 MC	Algebra II
Reading	Reading Grades 3–5	Kindergarten	1 passage*	3
Reading	Reading Grades 3–5	1	1 passage*	3
Reading	Reading Grades 3–5	2	1 passage*	3
Reading	Reading Grades 3–5	3	1 passage*	3,4
Reading	Reading Grades 3–5	4	1 passage*	4,5
Reading	Reading Grades 3–5	5	1 passage*	5,6
Reading	Reading/Lit Grades 6–HS	6	1 passage*	6,7
Reading	Reading/Lit Grades 6–HS	7	1 passage*	7, 8
Reading	Reading/Lit Grades 6–HS	8	1 passage*	8, 9+
Reading	Reading/Lit Grades 6–HS	Literature	1 passage*	9+
Science	Science Grades 3–5	2	5 MC	3
Science	Science Grades 3–5	3	5 MC	3,4
Science	Science Grades 3–5	4	5 MC	4,5
Science	Science Grades 3–5	5	5 MC	5,6
Science	Science Grades 6–HS	6	5 MC	6,7
Science	Science Grades 6–HS	7	5 MC	7, 8
Science	Science Grades 6–HS	8	5 MC	8, 9+
Science	Biology	Biology	5 MC	Biology
Science	Chemistry	Chemistry	5 MC	Chemistry

Table 6–14 (continued). Fall 2018 Embedded Field Test Design

Content Area	CDT	Item Grade/Course	Number of Items Embedded	Student Test Grade(s)
Writing	Writing Grades 3–5	Kindergarten	5 MC	3
Writing	Writing Grades 3–5	1	5 MC	3
Writing	Writing Grades 3–5	2	5 MC	3
Writing	Writing Grades 3–5	3	5 MC	3,4
Writing	Writing Grades 3–5	4	5 MC	4,5
Writing	Writing Grades 3–5	5	5 MC	5,6
Writing	Writing/Eng Comp Gr 6–HS	6	5 MC	6,7
Writing	Writing/Eng Comp Gr 6–HS	7	5 MC	7, 8
Writing	Writing/Eng Comp Gr 6–HS	8	5 MC	8, 9+
Writing	Writing/Eng Comp Gr 6–HS	English Composition	5 MC	9+

* FT reading passages include four to six items total with up to one evidence-based selected-response item.

FALL 2019—SCIENCE

The embedded field test administered in fall 2019 was designed to field test new technology-enhanced (TE) items in the science content area.

Table 6–15. Fall 2019 Embedded Field Test Item Pools

Content Area	Item Grade/Course	Number of MC Items	Number of TE Items	Total Number of Items
Science	K-2 grade span	0	0	0
Science	3	0	19	19
Science	4	0	22	22
Science	5	0	20	20
Science	6	0	18	18
Science	7	0	19	19
Science	8	0	20	20
Science	Biology	0	40	40
Science	Chemistry	0	0	0

Starting on August 19, 2019, CDTs in the science content area except Chemistry included embedded field-test items:

- Students using grade level tests took 2 field-test items. Since testing occurred throughout the year, items were given to students whose grade matched the item's grade and to students one grade above the item's grade (e.g., grade 7 items were given to students in grades 7 and 8).
- Students using CDT Biology took 2 field-test items from the Biology course.

In all cases, students did not know which items were operational and which were field test. Field test items did not count in calculation of total or diagnostic category scores.

Table 6–16. Fall 2019 Embedded Field Test Design

Content Area	CDT	Item Grade/Course	Number of Items Embedded	Student Test Grade(s)
Science	Science Grades 3–5	3	2	3, 4
Science	Science Grades 3–5	4	2	4, 5
Science	Science Grades 3–5	5	2	5, 6
Science	Science Grades 6–HS	6	2	6, 7
Science	Science Grades 6–HS	7	2	7, 8
Science	Science Grades 6–HS	8	2	8, 9+
Science	Biology	Biology	2	Biology

STATISTICAL ANALYSIS OF ITEM DATA

All field-tested items were analyzed statistically following conventional item analysis methods. For MC items, traditional or classical item statistics included the point-biserial correlation (Pt. Bis.) for the correct and incorrect responses (distractors), percent correct (p-value), and the percent selecting each incorrect response. For EBSR and TE items, the statistical indices included the item-test correlation, the point-biserial correlation for each score category, and the percent in each score category.

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 potential problem and the characteristics of the students affected. The primary way of detecting such conditions is through the point-biserial correlation coefficient for MC items and the item-test correlation for EBSR and TE items. In each case the statistic will be positive if the total-test mean score is higher for the students who respond correctly to MC items or attain a higher EBSR or TE score and negative when the reverse is true.

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 an MC item to be flagged, the criteria included any of the following:

- Point-biserial correlation for the correct response of less than 0.10
- Point-biserial correlation for any incorrect response greater than the point-biserial correlation for the correct response
- Differential item functioning (DIF) code of either C- or C+[2]

For an EBSR item to be flagged, the criteria included any of the following:

- Part One point-biserial correlation for the correct response of less than 0.10
- Part One point-biserial correlation for any incorrect response greater than the point-biserial correlation for the correct response
- Score proportion less than 0.05
- Differential item functioning (DIF) code of either C- or C+

[2] Items classified as C+ or C- have strong evidence of DIF. The plus sign indicates that the item favors the focal group (female or black or Hispanic) and a minus sign indicates that the item favors the reference group (male or white). For more details, see the section in this chapter on Differential Item Functioning.

For a TE item to be flagged, the criteria included any of the following:

- Item-test correlation less than 0.20
- Score proportion less than 0.05
- Differential item function (DIF) code of either C- or C+

These criteria differ slightly from the criteria used for end-of-year/course summative tests such as the Pennsylvania System of School Assessment (PSSA) or the Keystone Exams. For example, CDT items are not flagged for low and high *p*-values. While very easy and very difficult items may not be appropriate for summative tests, they are needed in diagnostic item pools so the computer adaptive item selection routine can find appropriate items for students at various levels.

Item analysis results for all items field tested prior to 2018-19 can be found in Appendix B of the 2017-2018 technical report. For field tests in 2018 or later, item analysis results are in Appendix B of the corresponding year's technical report.

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 field tests 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, or had extremely poor statistical quality and, consequently were regarded as unacceptable, were removed from the CDT item pools, 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 committees of Pennsylvania educators.

There were separate meetings to review items with data for each field-test event and content area. CDT mathematics items from the spring 2010 stand-alone field test were reviewed by fourteen Pennsylvania educators on August 9, 2010. CDT reading and science items from the fall 2010 stand-alone field test were reviewed by sixteen and fourteen Pennsylvania educators respectively on January 24, 2011. CDT writing items from the spring 2011 stand-alone field test were reviewed by fourteen Pennsylvania educators on August 1, 2011. CDT mathematics and reading items from the spring 2013 embedded field test were reviewed by twenty-two educators respectively on July 16–18, 2013. CDT mathematics, reading, science, and writing items from both the stand-alone and embedded field tests of fall 2013 were reviewed by seven, seven, seven, and eight Pennsylvania educators respectively on January 21–23, 2014. CDT mathematics, reading, science, and writing items from the embedded field tests of fall 2015 were reviewed by 10 Pennsylvania educators for each content group on June 9–10, 2016. CDT mathematics, reading, science, and writing items from the embedded field tests of fall 2018 were reviewed by 10 Pennsylvania educators for each content group on March 26–29, 2019. CDT science technology enhanced items from the embedded field tests in 2019 were reviewed by nine Pennsylvania educators on May 20–21, 2020.

At each of the item data review meetings committee members were first trained with regard to the statistical indices used in item evaluation. This was followed by a discussion with examples concerning reasons that 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 sensitivity/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 field-test items and made recommendations (i.e., accept or reject) for each item.

Table 6–17a. CDT Data Review Results for Mathematics in August 2010

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	86	4	4.7%	0	0.0%	0	0.0%
4	86	7	8.1%	0	0.0%	0	0.0%
5	85	0	0.0%	0	0.0%	0	0.0%
6	259	6	2.3%	0	0.0%	0	0.0%
7	258	19	7.4%	1	0.4%	1	0.4%
8	257	20	7.8%	1	0.4%	1	0.4%
11	149	13	8.7%	0	0.0%	0	0.0%
Algebra I	256	19	7.4%	6	2.3%	6	2.3%
Geometry	257	12	4.7%	3	1.2%	19	7.4%
Algebra II	256	15	5.9%	1	0.4%	2	0.8%

*Data Review Committee, PDE, and DRC

Table 6–17b. CDT Data Review Results for Reading in January 2011

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	86	0	0.0%	0	0.0%	0	0.0%
4	87	2	2.3%	0	0.0%	0	0.0%
5	86	3	3.5%	0	0.0%	0	0.0%
6	210	13	6.2%	1	0.5%	4	1.9%
7	192	8	4.2%	1	0.5%	2	1.0%
8	192	3	1.6%	0	0.0%	2	1.0%
Literature	348	16	4.6%	1	0.3%	8	2.3%

*Data Review Committee, PDE, and DRC

Table 6–17c. CDT Data Review Results for Science in January 2011

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	91	4	4.4%	1	1.1%	5	5.5%
4	123	6	4.9%	6	4.9%	9	7.3%
5	102	8	7.8%	3	2.9%	4	3.9%
6	178	13	7.3%	4	2.2%	10	5.6%
7	327	34	10.4%	28	8.6%	64	19.6%
8	377	43	11.4%	33	8.8%	56	14.9%
11	115	26	22.6%	9	7.8%	29	25.2%
Biology	390	43	11.0%	4	1.0%	61	15.6%
Chemistry	335	33	9.9%	8	2.4%	13	3.9%

*Data Review Committee, PDE, and DRC

Table 6–17d. CDT Data Review Results for Writing in August 2011

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	140	4	2.9%	1	0.7%	1	0.7%
4	149	10	6.7%	1	0.7%	1	0.7%
5	165	11	6.7%	4	2.4%	4	2.4%
6	193	13	6.7%	5	2.6%	5	2.6%
7	176	16	9.1%	5	2.8%	5	2.8%
8	195	21	10.8%	2	1.0%	2	1.0%
Eng. Comp	365	28	7.7%	10	2.7%	10	2.7%

*Data Review Committee, PDE, and DRC

Table 6–17e. CDT Data Review Results for Mathematics in July 2013

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
6	156	27	17.3%	7	4.5%	7	4.5%
7	73	15	20.5%	2	2.7%	2	2.7%
8	157	39	24.8%	4	2.5%	4	2.5%

*Data Review Committee, PDE, and DRC

Table 6–17f. CDT Data Review Results for Reading in July 2013

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
6	56	1	1.8%	1	1.8%	2	3.6%
7	58	4	6.9%	3	5.2%	4	6.9%
8	57	2	3.5%	1	1.8%	1	1.8%

*Data Review Committee, PDE, and DRC

Table 6–17g. CDT Data Review Results for Mathematics in January 2014

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K	60	14	23.3%	0	0.0%	1	1.7%
1	90	15	16.7%	0	0.0%	0	0.0%
2	130	11	8.5%	4	3.1%	5	3.8%
3	235	31	13.2%	3	1.3%	6	2.6%
4	248	20	8.1%	4	1.6%	11	4.4%
5	221	21	9.5%	4	1.8%	10	4.5%

*Data Review Committee, PDE, and DRC

Table 6–17h. CDT Data Review Results for Reading in January 2014

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K	84	11	13.1%	0	0.0%	0	0.0%
1	98	8	8.2%	3	3.1%	3	3.1%
2	98	1	1.0%	0	0.0%	0	0.0%
3	178	17	9.6%	2	1.1%	2	1.1%
4	189	11	5.8%	2	1.1%	2	1.1%
5	134	15	11.2%	0	0.0%	0	0.0%

*Data Review Committee, PDE, and DRC

Table 6–17i. CDT Data Review Results for Science in January 2014

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K–2	280	31	11.1%	5	1.8%	9	3.2%
3	155	9	5.8%	1	0.6%	4	2.6%
4	213	23	10.8%	4	1.9%	13	6.1%
5	152	44	28.9%	7	4.6%	10	6.6%

*Data Review Committee, PDE, and DRC

Table 6–17j. CDT Data Review Results for Writing in January 2014

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K	44	13	29.5%	2	4.5%	2	4.5%
1	118	18	15.3%	6	5.1%	6	5.1%
2	117	7	6.0%	3	2.6%	4	3.4%
3	60	4	6.7%	2	3.3%	2	3.3%
4	60	10	16.7%	3	5.0%	3	5.0%
5	71	15	21.1%	6	8.5%	6	8.5%

*Data Review Committee, PDE, and DRC

Table 6–17k. CDT Data Review Results for Mathematics in June 2016

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
6	122	17	13.9%	4	3.3%	4	3.3%
7	177	41	23.3%	10	5.7%	11	6.3%
8	151	31	20.4%	3	2.0%	4	2.6%
Algebra I	150	28	18.7%	1	0.7%	2	1.3%

*Data Review Committee, PDE, and DRC

Table 6–17l. CDT Data Review Results for Reading in June 2016

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	22	5	22.7%	0	0.0%	0	0.0%
4	22	6	27.3%	1	4.5%	1	4.5%
5	22	3	13.6%	0	0.0%	1	4.5%
6	126	10	7.9%	1	0.8%	4	3.2%
7	126	10	7.9%	1	0.8%	1	0.8%
8	126	12	9.5%	1	0.8%	3	2.4%
Literature	126	14	11.1%	1	0.8%	2	1.6%

*Data Review Committee, PDE, and DRC

Table 6–17m. CDT Data Review Results for Science in June 2016

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
6	72	12	16.7%	5	6.9%	6	8.3%
7	159	35	22.0%	6	3.8%	6	3.8%
8	238	65	27.3%	12	5.0%	12	5.0%
Biology	136	15	11.0%	1	0.7%	1	0.7%

*Data Review Committee, PDE, and DRC

Table 6–17n. CDT Data Review Results for Writing in June 2016

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
6	93	10	10.8%	3	3.2%	3	3.2%
7	93	9	9.7%	1	1.1%	1	1.1%
8	110	13	11.8%	3	2.7%	4	3.6%
Eng. Comp	104	12	11.5%	2	1.9%	2	1.9%

*Data Review Committee, PDE, and DRC

Table 6–17o. CDT Data Review Results for Mathematics in March 2019

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K-2	60	7	11.7%	0	0.0%	0	0.0%
3	178	21	11.8%	6	3.4%	6	3.4%
4	179	12	6.7%	1	0.6%	1	0.6%
5	180	8	4.4%	0	0.0%	0	0.0%
6	96	4	4.2%	1	1.0%	1	1.0%
7	103	10	9.7%	2	1.9%	2	1.9%
8	99	18	18.2%	4	4.0%	4	4.0%
Algebra I	299	64	21.4%	11	3.6%	11	3.6%
Geometry	100	22	22.0%	1	1.0%	1	1.0%
Algebra II	100	27	27.0%	2	2.0%	2	2.0%

*Data Review Committee, PDE, and DRC

Table 6–17p. CDT Data Review Results for Reading in March 2019

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K-2	84	10	11.9%	4	4.8%	4	4.8%
3	162	18	11.1%	4	2.5%	4	2.5%
4	162	16	10.5%	3	1.9%	4	2.5%
5	162	22	14.2%	5	3.1%	6	3.7%
6	123	10	8.1%	1	0.8%	1	0.8%
7	123	10	8.1%	4	3.3%	4	3.3%
8	120	14	11.7%	3	2.5%	3	2.5%
Literature	249	28	11.2%	2	0.8%	2	0.8%

*Data Review Committee, PDE, and DRC

Table 6–17q. CDT Data Review Results for Science in March 2019

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
2	31	4	12.9%	1	3.2%	1	3.2%
3	89	10	11.2%	2	2.2%	2	2.2%
4	95	14	14.7%	2	2.1%	2	2.1%
5	90	21	23.3%	2	2.2%	2	2.2%
6	97	25	25.8%	6	6.2%	6	6.2%
7	99	14	14.1%	1	1.0%	1	1.0%
8	102	24	23.5%	9	8.8%	9	8.8%
Biology	290	50	17.2%	11	3.8%	11	3.8%
Chemistry	100	49	49.0%	1	1.0%	1	1.0%

*Data Review Committee, PDE, and DRC

Table 6–17r. CDT Data Review Results for Writing in March 2019

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
K-2	32	4	12.5%	1	3.1%	3	9.4%
3	99	11	11.1%	1	1.0%	1	1.0%
4	90	7	7.8%	1	1.1%	2	2.2%
5	90	11	12.2%	1	1.1%	1	1.1%
6	93	9	9.7%	1	1.1%	2	2.2%
7	111	12	10.8%	1	0.9%	1	0.9%
8	93	10	10.8%	1	1.1%	1	1.1%
Eng. Comp	294	72	24.5%	18	6.1%	18	6.1%

*Data Review Committee, PDE, and DRC

Table 6–17s. CDT Data Review Results for Science in May 2020

Grade/Course	Number of Items Field Tested	Number Flagged and Examined at Data Review Committee	Percent Flagged and Examined at Data Review Committee	Number Rejected by Data Review Committee	Percent Rejected by Data Review Committee	Number Removed from CDT Item Pools (all sources)*	Percent Removed from CDT Item Pools (all sources)*
3	19	4	21.1%	2	10.53%	2	10.5%
4	22	9	40.9%	4	18.18%	5	22.7%
5	20	6	30.0%	1	5.00%	3	15.0%
6	18	4	22.2%	1	5.56%	1	5.6%
7	19	4	21.1%	3	15.79%	3	15.8%
8	20	7	35.0%	2	10.00%	2	10.0%
Biology	40	6	15.0%	3	7.50%	3	7.5%

*Data Review Committee, PDE, and DRC

DIFFERENTIAL ITEM FUNCTIONING

Differential item functioning (DIF) 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 sensitivity/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 sensitivity/bias is often easily recognized by 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.

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.

MANTEL-HAENSZEL PROCEDURE OF DIFFERENTIAL ITEM FUNCTIONING

For MC items, the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959) for detecting differential item functioning 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 and black or Hispanic[3] for ethnicity-based DIF; reference groups were male and white 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 population.

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 EBSR and TE items, a comparable statistic is computed based on the standardized mean difference (SMD) (Dorans, Schmitt, & Bleistein, 1992), which is 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 DIF statistic. 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 may be judged to be 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.

[3] Based on the population of CDT testers, ethnicity DIF on the white/Hispanic pairing was not run prior to 2018.

RESULTS AND OBSERVATIONS

Counts of the number of items field tested from each content area and grade/course that were assigned to each severity code are shown in Table 6–18. Some field-test items are classified as N/A (not applicable) because the number of students in either the reference or focal groups who took the item was insufficient for analysis. Where there are sufficient data to run DIF analyses, relatively few items had B or C DIF for the Male/Female, White/Black, or White/Hispanic reference and focal groups.

Table 6–18a. DIF Summary for Mathematics in August 2010

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	86	49	22	12	1	1	1	0	25	44	3	12	0	2	0
4	86	40	31	7	5	0	3	0	31	33	3	10	0	3	6
5	85	42	36	5	2	0	0	0	19	54	2	10	0	0	0
6	259	121	112	14	8	3	1	0	79	143	8	27	0	2	0
7	258	109	112	18	9	4	6	0	88	124	13	20	0	2	11
8	257	101	104	31	15	5	1	0	62	65	7	14	0	0	109
11	149	53	75	4	11	0	6	0	20	41	1	8	0	1	78
Algebra I	256	122	120	7	6	1	0	0	107	110	9	11	1	3	15
Geometry	257	115	123	7	8	1	3	0	93	109	6	15	1	2	31
Algebra II	256	124	115	6	9	0	2	0	58	89	4	14	2	4	85

N/A* Items with insufficient counts for DIF analysis
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18b. DIF Summary for Reading in January 2011

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	86	41	34	5	6	0	0	0	26	31	2	6	0	0	21
4	87	47	37	1	1	0	1	0	21	45	1	7	1	0	12
5	86	47	27	9	2	1	0	0	28	45	4	7	1	1	0
6	210	103	87	7	10	0	3	0	72	100	7	25	1	5	0
7	192	90	78	9	11	2	2	0	69	68	4	11	1	2	37
8	192	109	67	10	6	0	0	0	22	34	2	6	0	1	127
Literature	348	147	146	21	25	3	6	0	5	5	0	0	0	0	338

N/A* Items with insufficient counts for DIF analysis
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18c. DIF Summary for Science in January 2011

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	91	47	41	1	2	0	0	0	20	29	2	4	0	3	33
4	123	55	53	6	5	3	1	0	15	22	1	5	0	1	79
5	102	48	45	4	2	2	1	0	25	36	3	4	0	0	34
6	178	80	84	4	7	1	2	0	10	11	1	1	0	0	155
7	327	123	143	28	27	2	4	0	58	56	2	15	0	0	196
8	377	155	154	28	32	3	5	0	5	6	0	0	0	1	365
11	115	47	49	4	12	1	2	0	0	0	0	0	0	0	115
Biology	390	154	183	22	23	2	6	0	4	6	0	0	0	0	380
Chemistry	335	143	148	17	21	2	4	0	6	4	2	0	0	0	323

N/A* Items with insufficient counts for DIF analysis
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18d. DIF Summary for Writing in August 2011

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	140	71	59	4	4	1	1	0	24	44	3	4	0	0	65
4	149	69	67	7	5	1	0	0	15	26	3	2	0	0	103
5	165	78	62	15	7	3	0	0	12	14	1	2	0	1	135
6	193	94	82	8	7	1	1	0	53	67	4	12	0	4	53
7	176	73	81	16	3	3	0	0	11	20	1	3	0	0	141
8	195	95	81	10	3	3	3	0	4	3	0	2	0	1	185
Eng Comp	365	157	155	29	18	4	2	0	3	5	1	0	0	1	355

N/A* Items with insufficient counts for DIF analysis
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18e. DIF Summary for Mathematics in July 2013

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	56	0	0	0	0	0	0	56	0	0	0	0	0	0	56
4	67	0	0	0	0	0	0	67	0	0	0	0	0	0	67
5	41	0	0	0	0	0	0	41	0	0	0	0	0	0	41
6	156	67	65	9	14	1	0	0	2	1	0	2	0	0	151
7	73	37	32	2	1	0	1	0	13	16	1	4	0	0	39
8	157	72	63	8	12	2	0	0	2	5	0	1	0	0	149

N/A* Items with insufficient counts for DIF analysis or those that were re-field tested in fall 2013.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18f. DIF Summary for Reading in July 2013

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	58	0	0	0	0	0	0	58	0	0	0	0	0	0	58
4	71	0	0	0	0	0	0	71	0	0	0	0	0	0	71
5	60	0	0	0	0	0	0	60	0	0	0	0	0	0	60
6	56	29	21	4	2	0	0	0	4	6	0	2	0	0	44
7	58	29	21	4	3	1	0	0	11	34	1	3	0	0	9
8	57	34	20	2	1	0	0	0	13	38	0	5	0	1	0

N/A* Items with insufficient counts for DIF analysis or those that were re-field tested in fall 2013.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18g. DIF Summary for Mathematics in January 2014

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
K	60	31	19	6	3	1	0	0	6	14	1	5	0	2	32
1	90	40	38	8	4	0	0	0	18	25	0	5	0	0	42
2	130	47	56	7	16	1	3	0	24	32	3	4	0	1	66
3	235	101	101	11	15	4	3	0	28	41	2	5	1	1	157
4	248	105	110	16	14	2	1	0	37	44	7	11	0	2	147
5	221	108	84	13	12	2	2	0	31	41	3	8	0	1	137

N/A* Items with insufficient counts for DIF analysis or those that were re-field tested in fall 2013.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18h. DIF Summary for Reading in January 2014

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
K	84	50	21	9	2	2	0	0	9	10	0	3	0	0	62
1	98	57	31	6	3	1	0	0	7	11	0	0	0	0	80
2	98	47	43	3	4	0	1	0	5	13	0	2	0	0	78
3	178	81	75	8	10	3	1	0	54	69	5	11	0	1	38
4	189	93	78	12	6	0	0	0	40	54	2	7	0	2	84
5	134	75	49	6	2	0	2	0	23	53	1	6	0	2	49

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18i. DIF Summary for Science in January 2014

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
K–2	280	130	108	8	13	1	0	20	0	0	0	0	0	0	280
3	155	69	70	9	4	2	1	0	3	2	0	0	0	0	150
4	213	94	93	12	12	1	1	0	0	0	0	0	0	0	213
5	152	58	61	6	8	0	0	19	1	0	0	0	0	0	151

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18j. DIF Summary for Writing in January 2014

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
K	44	20	22	2	0	0	0	0	0	0	0	0	0	0	44
1	118	71	42	2	3	0	0	0	0	0	0	0	0	0	118
2	117	56	49	6	5	1	0	0	0	0	0	0	0	0	117
3	60	33	22	3	1	0	1	0	12	17	4	7	0	0	20
4	60	24	29	4	1	2	0	0	20	14	0	6	0	0	20
5	71	40	22	5	3	1	0	0	0	0	0	0	0	0	71

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white)

Table 6–18k. DIF Summary for Mathematics in June 2016

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
6	122	74	48	0	0	0	0	0	48	69	0	3	0	2	0
7	177	74	82	5	9	3	3	1	46	105	1	15	0	6	4
8	151	63	76	4	4	1	2	1	49	55	6	11	0	3	27
Algebra I	150	82	65	1	1	0	0	1	50	96	0	3	0	0	1

N/A* Items with insufficient counts for DIF analysis or those that were re-field tested in fall 2013.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18l. DIF Summary for Reading in June 2016

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
3	22	9	10	0	2	0	0	1	2	11	0	3	0	0	6
4	22	8	7	2	0	0	1	4	3	4	1	0	0	0	14
5	22	10	8	0	0	0	0	4	4	7	0	1	0	0	10
6	126	63	56	3	0	0	1	3	42	75	0	5	0	1	3
7	126	81	37	7	1	0	0	0	48	71	0	7	0	0	0
8	126	68	52	3	1	0	0	2	44	75	0	5	0	0	2
Literature	126	68	51	5	1	0	0	1	41	82	0	2	0	0	1

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18m. DIF Summary for Science in June 2016

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
6	72	37	31	2	2	0	0	0	19	30	2	6	0	0	15
7	159	75	67	5	10	0	2	0	31	54	2	13	0	1	58
8	238	106	106	11	8	4	3	0	36	69	4	17	0	0	112
Biology	136	64	70	1	1	0	0	0	34	101	0	1	0	0	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18n. DIF Summary for Writing in June 2016

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*	White/Black A+	White/Black A-	White/Black B+	White/Black B-	White/Black C+	White/Black C-	White/Black N/A*
6	93	53	34	2	4	0	0	0	26	27	2	11	0	0	27
7	93	48	38	2	3	2	0	0	6	13	1	3	0	0	70
8	110	66	38	3	1	1	1	0	3	6	0	2	0	0	99
Eng Comp	104	50	40	9	3	1	1	0	0	0	0	0	0	0	104

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female or black) and a minus sign indicates that the item favors the reference group (male or white).

Table 6–18o. Gender DIF Summary for Mathematics in March 2019

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*
K	20	10	7	2	0	1	0	0
1	20	11	6	2	0	1	0	0
2	20	9	10	0	1	0	0	0
3	178	81	79	8	5	3	2	0
4	179	83	82	7	6	1	0	0
5	180	96	71	7	3	3	0	0
6	96	51	40	2	3	0	0	0
7	103	42	53	4	4	0	0	0
8	99	52	41	1	4	0	1	0
Algebra I	299	157	131	2	8	0	1	0
Geometry	100	43	50	4	3	0	0	0
Algebra II	100	45	43	6	4	0	2	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female) and a minus sign indicates that the item favors the reference group (male).

Table 6–18p. Ethnicity DIF Summary for Mathematics in March 2019

Grade/ Course	Number of Field-test items	White/ Black A+	White/ Black A-	White/ Black B+	White/ Black B-	White/ Black C+	White/ Black C-	White/ Black N/A*	White/ Hispanic A+	White/ Hispanic A-	White/ Hispanic B+	White/ Hispanic B-	White/ Hispanic C+	White/ Hispanic C-	White/ Hispanic N/A*
K	20	2	11	0	3	0	0	4	2	11	0	6	0	1	0
1	20	8	8	0	3	0	0	1	6	11	0	3	0	0	0
2	20	6	7	0	2	0	2	3	4	11	0	4	0	1	0
3	178	47	93	2	22	0	4	10	46	102	5	16	0	8	1
4	179	54	92	3	20	0	2	8	62	93	2	15	0	6	1
5	180	61	97	1	19	0	1	1	68	101	1	8	0	1	1
6	96	28	59	0	9	0	0	0	31	64	1	0	0	0	0
7	103	35	57	0	10	0	1	0	29	62	1	9	0	2	0
8	99	43	47	1	5	0	0	3	42	54	0	2	0	0	1
Algebra I	299	114	169	2	11	0	3	0	112	172	2	12	0	1	0
Geometry	100	14	18	1	3	0	0	64	7	4	0	0	0	0	89
Algebra II	100	8	7	1	0	0	0	84	2	3	0	0	0	0	95

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (black or Hispanic) and a minus sign indicates that the item favors the reference group (white).

Table 6–18q. Gender DIF Summary for Reading in March 2019

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*
K	32	19	9	3	0	1	0	0
1	20	9	8	2	0	1	0	0
2	32	18	12	0	2	0	0	0
3	162	73	86	0	3	0	0	0
4	162	86	75	0	0	0	1	0
5	162	98	60	4	0	0	0	0
6	123	74	38	8	2	0	1	0
7	123	81	37	2	2	0	1	0
8	120	77	42	0	1	0	0	0
Literature	249	152	93	3	1	0	0	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female) and a minus sign indicates that the item favors the reference group (male).

Table 6–18r. Ethnicity DIF Summary for Reading in March 2019

Grade/ Course	Number of Field-test items	White/ Black A+	White/ Black A-	White/ Black B+	White/ Black B-	White/ Black C+	White/ Black C-	White/ Black N/A*	White/ Hispanic A+	White/ Hispanic A-	White/ Hispanic B+	White/ Hispanic B-	White/ Hispanic C+	White/ Hispanic C-	White/ Hispanic N/A*
K	32	4	12	0	0	0	0	16	10	15	0	3	0	0	32
1	20	1	5	1	0	0	1	12	9	10	0	1	0	0	20
2	32	10	10	0	3	0	1	8	12	17	1	2	0	0	32
3	162	58	82	2	10	2	2	6	51	95	4	11	0	1	162
4	162	62	78	2	10	1	3	6	68	79	0	13	1	1	162
5	162	64	76	3	7	0	0	12	54	100	2	4	0	2	162
6	123	47	70	2	4	0	0	0	47	69	3	4	0	0	123
7	123	43	72	3	5	0	0	0	53	64	3	3	0	0	123
8	120	41	72	0	6	0	1	0	40	76	0	4	0	0	120
Literature	249	103	131	1	11	0	3	0	97	140	2	9	0	1	249

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (black or Hispanic) and a minus sign indicates that the item favors the reference group (white).

Table 6–18s. Gender DIF Summary for Science in March 2019

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*
2	31	10	19	0	1	0	1	0
3	89	42	41	5	1	0	0	0
4	95	46	43	4	2	0	0	0
5	90	42	41	6	0	0	1	0
6	97	45	42	5	3	1	1	0
7	99	45	52	1	0	1	0	0
8	102	48	50	1	2	0	1	0
Biology	290	145	140	2	3	0	0	0
Chemistry	100	41	45	5	6	3	0	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female) and a minus sign indicates that the item favors the reference group (male).

Table 6–18t. Ethnicity DIF Summary for Science in March 2019

Grade/ Course	Number of Field-test items	White/ Black A+	White/ Black A-	White/ Black B+	White/ Black B-	White/ Black C+	White/ Black C-	White/ Black N/A*	White/ Hispanic A+	White/ Hispanic A-	White/ Hispanic B+	White/ Hispanic B-	White/ Hispanic C+	White/ Hispanic C-	White/ Hispanic N/A*
2	31	8	7	0	1	0	0	15	0	0	0	0	0	0	31
3	89	32	49	4	3	0	1	0	27	50	1	10	0	0	1
4	95	35	46	5	6	0	0	3	37	52	2	3	0	1	0
5	90	14	43	5	5	0	1	22	29	43	2	10	0	2	4
6	97	24	44	2	5	0	3	19	31	56	0	7	0	1	2
7	99	26	58	2	7	0	2	4	39	49	2	7	1	0	1
8	102	29	50	2	9	0	3	9	32	60	4	4	0	2	0
Biology	290	108	168	1	10	0	3	0	95	184	1	10	0	0	0
Chemistry	100	0	0	0	0	0	0	100	2	4	0	0	0	0	94

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (black or Hispanic) and a minus sign indicates that the item favors the reference group (white).

Table 6–18u. Gender DIF Summary for Writing in March 2019

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*
K	10	5	5	0	0	0	0	0
1	10	4	5	1	0	0	0	0
2	12	5	6	0	0	1	0	0
3	99	50	38	8	1	1	1	0
4	90	45	32	8	3	1	1	0
5	90	46	29	8	2	3	2	0
6	93	48	32	7	2	3	1	0
7	111	56	37	7	7	4	0	0
8	93	51	31	7	4	0	0	0
Eng Comp	294	64	43	5	1	0	1	180

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female) and a minus sign indicates that the item favors the reference group (male).

Table 6–18v. Ethnicity DIF Summary for Writing in March 2019

Grade/ Course	Number of Field-test items	White/ Black A+	White/ Black A-	White/ Black B+	White/ Black B-	White/ Black C+	White/ Black C-	White/ Black N/A*	White/ Hispanic A+	White/ Hispanic A-	White/ Hispanic B+	White/ Hispanic B-	White/ Hispanic C+	White/ Hispanic C-	White/ Hispanic N/A*
K	10	0	0	0	0	0	0	10	0	0	0	0	0	0	10
1	10	0	0	0	0	0	0	10	0	0	0	0	0	0	10
2	12	0	0	0	0	0	0	12	0	0	0	0	0	0	12
3	99	0	0	0	0	0	0	99	0	0	0	0	0	0	99
4	90	0	0	0	0	0	0	90	0	0	0	0	0	0	90
5	90	0	1	0	0	0	0	89	1	0	0	0	0	0	89
6	93	2	6	0	0	0	0	85	0	1	0	0	0	0	92
7	111	3	5	0	0	0	0	103	2	1	0	0	0	0	108
8	93	0	0	0	0	0	0	93	0	0	0	0	0	0	93
Eng Comp	294	0	0	0	0	0	0	294	0	0	0	0	0	0	294

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (black or Hispanic) and a minus sign indicates that the item favors the reference group (white).

Table 6–18w. Gender DIF Summary for Science in May 2020

Grade/ Course	Number of Field-test items	Male/ Female A+	Male/ Female A-	Male/ Female B+	Male/ Female B-	Male/ Female C+	Male/ Female C-	Male/ Female N/A*
3	19	7	11	1	0	0	0	0
4	22	9	11	0	2	0	0	0
5	20	5	14	1	0	0	0	0
6	18	6	10	0	2	0	0	0
7	19	12	6	0	1	0	0	0
8	20	8	11	0	1	0	0	0
Biology	40	18	21	0	1	0	0	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (female) and a minus sign indicates that the item favors the reference group (male).

Table 6–18x. Ethnicity DIF Summary for Science in May 2019

Grade/ Course	Number of Field-test items	White/ Black A+	White/ Black A-	White/ Black B+	White/ Black B-	White/ Black C+	White/ Black C-	White/ Black N/A*	White/ Hispanic A+	White/ Hispanic A-	White/ Hispanic B+	White/ Hispanic B-	White/ Hispanic C+	White/ Hispanic C-	White/ Hispanic N/A*
3	19	4	10	0	3	0	2	0	4	12	0	2	0	1	0
4	22	7	10	2	1	1	1	0	5	13	0	3	0	1	0
5	20	5	9	0	5	0	0	1	4	15	0	0	0	1	0
6	18	1	16	0	1	0	0	0	1	16	0	1	0	0	0
7	19	2	15	0	2	0	0	0	4	14	0	1	0	0	0
8	20	2	17	0	0	0	1	0	5	13	0	1	0	1	0
Biology	40	8	30	0	2	0	0	0	9	30	0	1	0	0	0

N/A* Items with insufficient counts for DIF analysis.
The plus sign indicates that the item favors the focal group (black or Hispanic) and a minus sign indicates that the item favors the reference group (white).

CHAPTER SEVEN: 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 to all items field tested in the stand-alone and embedded field-test events. Other statistics such as Rasch item statistics are discussed in Chapter Eight.

ITEM-LEVEL STATISTICS

Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–2019 or later can be found in Appendix B of the corresponding year's technical report. In all versions of appendix B, results are organized by content area, field-test event, and item type (multiple-choice, evidence-based selected-response, and technology-enhanced). 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 B include:

- Number of students taking the item (denoted as *N*)
- Indicators of item difficulty (denoted as PVal)
    - *p*-values for multiple-choice (MC) items
    - item mean divided by maximum possible item score for evidence-based selected-response (EBSR) and technology-enhanced (TE) items
- Proportions by response option or score point
    - proportions of students who chose each response option for MC items (denoted as P(A), P(B), P(C), P(D))
    - proportions of students who gained each score point for EBSR and TE items (denoted as P(0), P(1), P(2), and/or P(3))
    - Proportions of students who did not respond to an item (denoted as P(-))
- Indicators of item discrimination
    - item-total correlations (denoted as PtBis)
    - point-biserial correlation for each response option for MC items (denoted as PT(A), PT(B), PT(C), and PT(D))
    - point-biserial correlation for each score point for EBSR and TE items (denoted as PT(0), PT(1), PT(2), and PT(3))

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

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

For EBSR items, mean scores can range from the minimum possible score of zero to the maximum possible score of either two or three depending upon the item. Similarly, for TE items, mean scores can range from the minimum possible score of zero to the maximum possible score of either one or two depending upon the item. A *pseudo p*-value is provided for EBSR and TE items 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 Classroom Diagnostic Tools (CDT) because it is a computer adaptive test. The item selection routine selects items based on student performance during the test. While very easy or very difficult items may not be appropriate for many students, they are needed in the CDT item pools to ensure that the item selection routine can find appropriate items for students at various levels.

Utilizing the proportion of students who chose each MC 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 the key is not correct.

ITEM DISCRIMINATION

At the most general level, item discrimination[2] indicates an item's ability to differentiate between high and low achievers. It is expected that students with high ability (i.e., those who perform well on the CDT overall) would be more likely to answer any given CDT item correctly, while students with low ability (i.e., those who perform poorly on the CDT overall) would be more likely to answer the same item incorrectly. For the CDT, Pearson's product-moment correlation coefficient between item scores and test scores is used to indicate discrimination. 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), meaning 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 (i.e., those who perform well on the CDT overall) would be less likely to choose any distractors, while students with low ability (i.e., those who perform poorly on the CDT overall) would be more likely to choose a distractor. In other words, the item total correlations for the distractors are expected to be negative.

In summary, the correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly.[3] In other words, this indicates that 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., they have extreme *p*-values) can have reduced power to discriminate, and, thus, can have lower correlations.

Discrimination is an important consideration for the operational CDT because the use of more discriminating items on a test is associated with more precise score estimates (i.e., there will be smaller confidence intervals around the scores).

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

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

OBSERVATIONS AND INTERPRETATIONS

Table 7–1 provides the mean *p*-values and point-biserial correlations for the CDT item pools in each content area. The mean *p*-value ranged from 0.279 to 0.824. The mean point-biserial correlations ranged from 0.155 to 0.491.

It is difficult to make global conclusions about overall quality from these item statistics alone. With that caveat in mind, the results presented in this chapter indicate that the CDT item pools contain items within expected and acceptable ranges of item difficulty and discrimination.

Table 7–1. Mean *P*-value and Point-Biserial

Meeting Date	Content Area	Grade/Course	Number of Items Field Tested	Mean *P*-value	Mean Point-Biserial
Aug 2010	Mathematics	3	86	0.824	0.415
Aug 2010	Mathematics	4	86	0.737	0.414
Aug 2010	Mathematics	5	85	0.717	0.439
Aug 2010	Mathematics	6	259	0.684	0.413
Aug 2010	Mathematics	7	258	0.575	0.432
Aug 2010	Mathematics	8	257	0.497	0.361
Aug 2010	Mathematics	11	149	0.521	0.339
Aug 2010	Mathematics	Algebra I	256	0.411	0.317
Aug 2010	Mathematics	Geometry	257	0.439	0.349
Aug 2010	Mathematics	Algebra II	256	0.419	0.369
Jan 2011	Reading	3	86	0.595	0.437
Jan 2011	Reading	4	87	0.665	0.440
Jan 2011	Reading	5	86	0.666	0.433
Jan 2011	Reading	6	210	0.607	0.423
Jan 2011	Reading	7	192	0.679	0.395
Jan 2011	Reading	8	192	0.623	0.404
Jan 2011	Reading	Literature	348	0.568	0.408
Jan 2011	Science	3	91	0.637	0.371
Jan 2011	Science	4	123	0.602	0.348
Jan 2011	Science	5	102	0.482	0.335
Jan 2011	Science	6	178	0.503	0.322
Jan 2011	Science	7	327	0.486	0.322
Jan 2011	Science	8	377	0.504	0.335
Jan 2011	Science	11	115	0.381	0.238
Jan 2011	Science	Biology	390	0.420	0.294
Jan 2011	Science	Chemistry	335	0.355	0.255
Aug 2011	Writing	3	140	0.584	0.392
Aug 2011	Writing	4	149	0.566	0.372
Aug 2011	Writing	5	165	0.566	0.380
Aug 2011	Writing	6	193	0.556	0.369

Table 7–1 (continued). Mean *P*-value and Point-Biserial

Meeting Date	Content Area	Grade/Course	Number of Items Field Tested	Mean *P*-value	Mean Point-Biserial
Aug 2011	Writing	7	176	0.550	0.346
Aug 2011	Writing	8	195	0.538	0.332
Aug 2011	Writing	English Composition	365	0.514	0.357
July 2013	Mathematics	6	156	0.448	0.290
July 2013	Mathematics	7	73	0.431	0.257
July 2013	Mathematics	8	157	0.354	0.204
July 2013	Reading	6	56	0.585	0.351
July 2013	Reading	7	58	0.545	0.339
July 2013	Reading	8	57	0.577	0.358
Jan 2014	Mathematics	K	60	0.798	0.408
Jan 2014	Mathematics	1	90	0.801	0.426
Jan 2014	Mathematics	2	130	0.695	0.437
Jan 2014	Mathematics	3	235	0.596	0.413
Jan 2014	Mathematics	4	248	0.595	0.413
Jan 2014	Mathematics	5	221	0.508	0.326
Jan 2014	Reading	K	84	0.734	0.426
Jan 2014	Reading	1	98	0.575	0.415
Jan 2014	Reading	2	98	0.506	0.441
Jan 2014	Reading	3	178	0.546	0.398
Jan 2014	Reading	4	189	0.577	0.413
Jan 2014	Reading	5	134	0.566	0.364
Jan 2014	Science	K–2 span	280	0.619	0.404
Jan 2014	Science	3	155	0.641	0.391
Jan 2014	Science	4	213	0.570	0.362
Jan 2014	Science	5	152	0.424	0.240
Jan 2014	Writing	K	44	0.823	0.462
Jan 2014	Writing	1	118	0.729	0.444
Jan 2014	Writing	2	117	0.642	0.444
Jan 2014	Writing	3	60	0.626	0.415
Jan 2014	Writing	4	60	0.642	0.398
Jan 2014	Writing	5	71	0.550	0.326
June 2016	Mathematics	6	122	0.473	0.298
June 2016	Mathematics	7	177	0.456	0.286
June 2016	Mathematics	8	151	0.396	0.232
June 2016	Mathematics	Algebra I	150	0.414	0.228
June 2016	Reading	3	22	0.467	0.430

Table 7–1 (continued). Mean *P*-value and Point-Biserial

Meeting Date	Content Area	Grade/Course	Number of Items Field Tested	Mean *P*-value	Mean Point-Biserial
June 2016	Reading	4	22	0.568	0.421
June 2016	Reading	5	22	0.603	0.394
June 2016	Reading	6	126	0.535	0.360
June 2016	Reading	7	126	0.557	0.397
June 2016	Reading	8	126	0.577	0.398
June 2016	Reading	Literature	126	0.532	0.339
June 2016	Science	6	72	0.431	0.233
June 2016	Science	7	159	0.446	0.231
June 2016	Science	8	238	0.447	0.236
June 2016	Science	Biology	136	0.439	0.246
June 2016	Writing	6	93	0.531	0.327
June 2016	Writing	7	93	0.522	0.322
June 2016	Writing	8	110	0.504	0.308
June 2016	Writing	English Composition	104	0.485	0.298
March 2019	Mathematics	K	20	0.778	0.362
March 2019	Mathematics	1	20	0.758	0.389
March 2019	Mathematics	2	20	0.672	0.422
March 2019	Mathematics	3	178	0.602	0.379
March 2019	Mathematics	4	179	0.578	0.362
March 2019	Mathematics	5	180	0.569	0.350
March 2019	Mathematics	6	96	0.495	0.321
March 2019	Mathematics	7	103	0.476	0.328
March 2019	Mathematics	8	99	0.401	0.256
March 2019	Mathematics	Algebra I	299	0.401	0.246
March 2019	Mathematics	Geometry	100	0.378	0.228
March 2019	Mathematics	Algebra II	100	0.375	0.230
March 2019	Reading	K	32	0.527	0.368
March 2019	Reading	1	20	0.500	0.389
March 2019	Reading	2	32	0.459	0.343
March 2019	Reading	3	162	0.448	0.353
March 2019	Reading	4	162	0.484	0.357
March 2019	Reading	5	162	0.483	0.352
March 2019	Reading	6	123	0.508	0.371
March 2019	Reading	7	123	0.476	0.343
March 2019	Reading	8	120	0.503	0.356
March 2019	Reading	Literature	249	0.491	0.340

Table 7–1 (continued). Mean *P*-value and Point-Biserial

Meeting Date	Content Area	Grade/Course	Number of Items Field Tested	Mean *P*-value	Mean Point-Biserial
March 2019	Science	K-2 span	31	0.515	0.321
March 2019	Science	3	89	0.501	0.303
March 2019	Science	4	95	0.474	0.287
March 2019	Science	5	90	0.439	0.273
March 2019	Science	6	97	0.446	0.265
March 2019	Science	7	99	0.479	0.294
March 2019	Science	8	102	0.459	0.269
March 2019	Science	Biology	290	0.421	0.267
March 2019	Science	Chemistry	110	0.356	0.155
March 2019	Writing	K	10	0.713	0.491
March 2019	Writing	1	10	0.520	0.351
March 2019	Writing	2	12	0.445	0.281
March 2019	Writing	3	99	0.525	0.349
March 2019	Writing	4	90	0.589	0.364
March 2019	Writing	5	90	0.549	0.351
March 2019	Writing	6	93	0.517	0.329
March 2019	Writing	7	111	0.518	0.342
March 2019	Writing	8	93	0.514	0.333
March 2019	Writing	English Composition	294	0.475	0.285
May 2020	Science	3	19	0.458	0.336
May 2020	Science	4	22	0.300	0.282
May 2020	Science	5	20	0.293	0.307
May 2020	Science	6	18	0.284	0.275
May 2020	Science	7	19	0.312	0.283
May 2020	Science	8	20	0.283	0.294
May 2020	Science	Biology	40	0.279	0.319

CHAPTER EIGHT: RASCH ITEM CALIBRATION

The particular item response theory (IRT) model used for the Classroom Diagnostic Tools (CDT) 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 used to calibrate the Pennsylvania System of School Assessment (PSSA) items and Keystone Exam items. Consequently, this model was chosen to be used for the CDT. 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 CDT items. Generally, item calibration is the process of assigning a difficulty-parameter estimate to each item so that they are placed onto a common scale. This chapter briefly introduces the Rasch model and reports the results from evaluations of the adequacy of the Rasch assumptions. See Chapter Nine for a description of the common scale across grades and courses within a content area and for summaries of the Rasch item statistics for the CDT item pools.

DESCRIPTION OF THE RASCH MODEL

The Rasch partial credit model (RPCM) (Wright & Masters, 1982) was used to calibrate CDT items because the item pools contain multiple item types. The RPCM extends the Rasch model (Rasch, 1960) for dichotomous multiple-choice (0, 1) items so that it accommodates the polytomous evidence-based selected-response and technology-enhanced items. Under the RPCM, for a given item *i* with *mi* score categories, the probability of person *n* scoring *x* (*x* = 0, 1, 2,... *mi*) is given by:

$$P_{ni}(X = x) = \frac{\exp\sum_{j=0}^{x}(\theta_n - D_{ij})}{\sum_{k=0}^{m_i}\exp\sum_{j=0}^{k}(\theta_n - D_{ij})}, \; x = 0, 1, \ldots, m_i$$

where θ_n represents a student's proficiency (ability) level, and D_{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:

$$P_{ni}(X = 1) = \frac{\exp(\theta_n - D_{ij})}{1 + \exp(\theta_n - D_{ij})}.$$

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.

SOFTWARE AND ESTIMATION ALGORITHM

Item calibration was implemented via the WINSTEPS 3.71 computer program (Linacre, 2009). The unconditional, joint maximum likelihood (UCON) estimation procedure estimates the person parameters (i.e., ability) simultaneously with the item parameters (i.e., difficulty).

CHECKING RASCH ASSUMPTIONS

Because the Rasch model was the basis of all calibration, scoring, and scaling analyses associated with the CDT, 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.

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.

WINSTEPS provides three PCA residuals: raw, standardized, and logit. All three should yield similar results. The mixed residual setting was used for the PCA because previous research has demonstrated that raw residuals (PRCOMP=R) give a more realistic estimate of explained variance than do standardized residuals (PRCOMP=S), and standardized residuals are better for decomposing the unexplained variance into contrasts (Linacre, 2009).

Table 8–1 presents the PCA results for the CDT Mathematics item pool. The results include the total variance, variance explained by the model, unexplained total variance, and unexplained variance explained by the first factor (both eigenvalue units and percentage values are shown in the table). In addition, the modeled column provides variance components that would be explained if the data complied with the Rasch definition of unidimensionality.

As can been seen from Table 8–1, the primary dimension in the Rasch model explained between 21 and 63 percent of the total variances across the grades and courses. The empirical and model-based percentages were close, suggesting that the estimation of a primary Rasch dimension was successful. The unexplained variances were between 38 and 79 percent. This includes the Rasch-predicted randomness and any departures in the data from the Rasch model (e.g., departure from unidimensionality).

The most important variance for evaluating dimensionality is in the row named "unexplained variance explained by 1st factor." The eigenvalue of unexplained total variance equals the total number of items, since PCA was conducted with residuals. The eigenvalues of the first factor in the residual (again, this is the second dimension beyond the first Rasch model dimension in WINSTEPS PCA) were between 0.2 and 1.1 percent. Overall, WINSTEPS PCA suggests that there is one clearly dominant dimension for the CDT mathematics item pool.

Table 8–1. Results from PCA of Residuals in WINSTEPS for Mathematics

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Aug 2010	3	Total variance in observations	208.5	100.0%	100.0%
Aug 2010	3	Variance explained by model	122.5	58.7%	58.5%
Aug 2010	3	Unexplained variance (total)	86	41.3%	41.5%
Aug 2010	3	Unexplained variance explained by 1st factor	1.6	0.8%	
Aug 2010	4	Total variance in observations	167.8	100.0%	100.0%
Aug 2010	4	Variance explained by model	81.8	48.7%	48.1%
Aug 2010	4	Unexplained variance (total)	86	51.3%	51.9%
Aug 2010	4	Unexplained variance explained by 1st factor	1.5	0.9%	
Aug 2010	5	Total variance in observations	177.3	100.0%	100.0%
Aug 2010	5	Variance explained by model	92.3	52.1%	52.9%
Aug 2010	5	Unexplained variance (total)	85	47.9%	47.1%
Aug 2010	5	Unexplained variance explained by 1st factor	1.5	0.9%	
Aug 2010	6	Total variance in observations	606.2	100.0%	100.0%
Aug 2010	6	Variance explained by model	347.2	57.3%	58.0%
Aug 2010	6	Unexplained variance (total)	259	42.7%	42.0%
Aug 2010	6	Unexplained variance explained by 1st factor	2.0	0.3%	
Aug 2010	7	Total variance in observations	529.8	100.0%	100.0%
Aug 2010	7	Variance explained by model	271.8	51.3%	52.3%
Aug 2010	7	Unexplained variance (total)	258	48.7%	47.7%
Aug 2010	7	Unexplained variance explained by 1st factor	2.2	0.4%	
Aug 2010	8	Total variance in observations	476.9	100.0%	100.0%
Aug 2010	8	Variance explained by model	219.9	46.1%	47.3%
Aug 2010	8	Unexplained variance (total)	257	53.9%	52.7%
Aug 2010	8	Unexplained variance explained by 1st factor	2.1	0.4%	
Aug 2010	Algebra I*	Total variance in observations	365.4	100.0%	100.0%
Aug 2010	Algebra I*	Variance explained by model	109.4	29.9%	30.6%
Aug 2010	Algebra I*	Unexplained variance (total)	256	70.1%	69.4%
Aug 2010	Algebra I*	Unexplained variance explained by 1st factor	1.9	0.5%	
Aug 2010	Geometry*	Total variance in observations	408.9	100.0%	100.0%
Aug 2010	Geometry*	Variance explained by model	151.9	37.2%	38.3%
Aug 2010	Geometry*	Unexplained variance (total)	257	62.8%	61.7%
Aug 2010	Geometry*	Unexplained variance explained by 1st factor	1.9	0.5%	

*Grade 11 items were tested on grade 8, Algebra I, Geometry, and Algebra II forms.

Table 8–1 (continued). Results from PCA of Residuals in WINSTEPS for Mathematics

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Aug 2010	Algebra II*	Total variance in observations	464.8	100.0%	100.0%
Aug 2010	Algebra II*	Variance explained by model	208.8	44.9%	46.1%
Aug 2010	Algebra II*	Unexplained variance (total)	256	55.1%	53.9%
Aug 2010	Algebra II*	Unexplained variance explained by 1st factor	2.0	0.4%	
July 2013	6	Total variance in observations	323.3	100.0%	100.0%
July 2013	6	Variance explained by model	167.3	51.7%	48.4%
July 2013	6	Unexplained variance (total)	156	48.3%	51.6%
July 2013	6	Unexplained variance explained by 1st factor	1.3	0.4%	
July 2013	7	Total variance in observations	148.3	100.0%	100.0%
July 2013	7	Variance explained by model	75.3	50.8%	48.7%
July 2013	7	Unexplained variance (total)	73	49.2%	51.3%
July 2013	7	Unexplained variance explained by 1st factor	1.1	0.8%	
July 2013	8	Total variance in observations	243.3	100.0%	100.0%
July 2013	8	Variance explained by model	86.3	35.5%	33.0%
July 2013	8	Unexplained variance (total)	157	64.5%	67.0%
July 2013	8	Unexplained variance explained by 1st factor	1.3	0.6%	
Jan 2014	K–2**	Total variance in observations	728.0	100.0%	100.0%
Jan 2014	K–2**	Variance explained by model	448.0	61.5%	60.5%
Jan 2014	K–2**	Unexplained variance (total)	280	38.5%	39.5%
Jan 2014	K–2**	Unexplained variance explained by 1st factor	1.8	0.3%	
Jan 2014	3	Total variance in observations	564.0	100.0%	100.0%
Jan 2014	3	Variance explained by model	329.0	58.3%	59.4%
Jan 2014	3	Unexplained variance (total)	235	41.7%	40.6%
Jan 2014	3	Unexplained variance explained by 1st factor	1.9	0.3%	
Jan 2014	4	Total variance in observations	646.9	100.0%	100.0%
Jan 2014	4	Variance explained by model	398.9	61.7%	62.5%
Jan 2014	4	Unexplained variance (total)	248	38.3%	37.5%
Jan 2014	4	Unexplained variance explained by 1st factor	1.9	0.3%	
Jan 2014	5	Total variance in observations	417.9	100.0%	100.0%
Jan 2014	5	Variance explained by model	196.9	47.1%	43.1%
Jan 2014	5	Unexplained variance (total)	221	52.9%	56.9%
Jan 2014	5	Unexplained variance explained by 1st factor	1.2	0.3%	
June 2016	6	Total variance in observations	212.5	100.0%	100.0%
June 2016	6	Variance explained by model	94.5	44.5%	39.8%
June 2016	6	Unexplained variance (total)	118	55.5%	60.2%
June 2016	6	Unexplained variance explained by 1st factor	1.1	0.5%	

*Grade 11 items were tested on grade 8, Algebra I, Geometry, and Algebra II forms.
**Items in kindergarten through grade 2 were co-mingled on forms taken by students in grade 3.

Table 8–1 (continued). Results from PCA of Residuals in WINSTEPS for Mathematics

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2016	7	Total variance in observations	267.9	100.0%	100.0%
June 2016	7	Variance explained by model	101.9	38.0%	32.0%
June 2016	7	Unexplained variance (total)	166	62.0%	68.0%
June 2016	7	Unexplained variance explained by 1st factor	1.1	0.4%	
June 2016	8	Total variance in observations	197.5	100.0%	100.0%
June 2016	8	Variance explained by model	50.5	25.6%	20.9%
June 2016	8	Unexplained variance (total)	147	74.4%	79.1%
June 2016	8	Unexplained variance explained by 1st factor	1.1	0.6%	
June 2016	Algebra I	Total variance in observations	243.8	100.0%	100.0%
June 2016	Algebra I	Variance explained by model	95.8	39.3%	36.8%
June 2016	Algebra I	Unexplained variance (total)	148	60.7%	63.2%
June 2016	Algebra I	Unexplained variance explained by 1st factor	1.1	0.4%	
June 2019	K–2**	Total variance in observations	116.0	100.0%	100.0%
June 2019	K–2**	Variance explained by model	56.0	48.3%	35.6%
June 2019	K–2**	Unexplained variance (total)	60.0	51.7%	64.4%
June 2019	K–2**	Unexplained variance explained by 1st factor	1.2	1.1%	
June 2019	3	Total variance in observations	384.3	100.0%	100.0%
June 2019	3	Variance explained by model	206.3	53.7%	46.7%
June 2019	3	Unexplained variance (total)	178.0	46.3%	53.3%
June 2019	3	Unexplained variance explained by 1st factor	1.2	0.3%	
June 2019	4	Total variance in observations	338.4	100.0%	100.0%
June 2019	4	Variance explained by model	159.4	47.1%	38.0%
June 2019	4	Unexplained variance (total)	179.0	52.9%	62.0%
June 2019	4	Unexplained variance explained by 1st factor	1.1	0.3%	
June 2019	5	Total variance in observations	316.3	100.0%	100.0%
June 2019	5	Variance explained by model	136.3	43.1%	36.5%
June 2019	5	Unexplained variance (total)	180.0	56.9%	63.5%
June 2019	5	Unexplained variance explained by 1st factor	1.1	0.4%	
June 2019	6	Total variance in observations	156.0	100.0%	100.0%
June 2019	6	Variance explained by model	60.0	38.4%	31.1%
June 2019	6	Unexplained variance (total)	96.0	61.6%	68.9%
June 2019	6	Unexplained variance explained by 1st factor	1.1	0.7%	
June 2019	7	Total variance in observations	154.8	100.0%	100.0%
June 2019	7	Variance explained by model	51.8	33.5%	28.4%
June 2019	7	Unexplained variance (total)	103.0	66.5%	71.6%
June 2019	7	Unexplained variance explained by 1st factor	1.1	0.7%	

Table 8–1 (continued). Results from PCA of Residuals in WINSTEPS for Mathematics

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2019	8	Total variance in observations	147.9	100.0%	100.0%
June 2019	8	Variance explained by model	48.9	33.1%	27.6%
June 2019	8	Unexplained variance (total)	99.0	66.9%	72.4%
June 2019	8	Unexplained variance explained by 1st factor	1.1	0.8%	
June 2019	Algebra I	Total variance in observations	456.3	100.0%	100.0%
June 2019	Algebra I	Variance explained by model	157.3	34.5%	33.1%
June 2019	Algebra I	Unexplained variance (total)	299.0	65.5%	66.9%
June 2019	Algebra I	Unexplained variance explained by 1st factor	1.1	0.2%	
June 2019	Geometry	Total variance in observations	158.5	100.0%	100.0%
June 2019	Geometry	Variance explained by model	58.5	36.9%	35.5%
June 2019	Geometry	Unexplained variance (total)	100.0	63.1%	64.5%
June 2019	Geometry	Unexplained variance explained by 1st factor	1.2	0.7%	
June 2019	Algebra II	Total variance in observations	161.0	100.0%	100.0%
June 2019	Algebra II	Variance explained by model	61.0	37.9%	35.9%
June 2019	Algebra II	Unexplained variance (total)	100.0	62.1%	64.1%
June 2019	Algebra II	Unexplained variance explained by 1st factor	1.2	0.7%	

Table 8–2 presents the PCA results for the CDT reading item pool. The primary dimension in the Rasch model explained between 26 and 58 percent of the total variances across the grades and courses. The second dimension (the row named "unexplained variance explained by 1st factor") accounted for between 0.3 and 3.2 percent of the total variance in observations. These results suggest that the CDT reading item pool essentially measures a single dominant dimension.

Table 8–2. Results from PCA of Residuals in WINSTEPS for Reading

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Jan 2011	3	Total variance in observations	179.8	100.0%	100.0%
Jan 2011	3	Variance explained by model	93.8	52.2%	51.9%
Jan 2011	3	Unexplained variance (total)	86	47.8%	48.1%
Jan 2011	3	Unexplained variance explained by 1st factor	1.7	0.9%	
Jan 2011	4	Total variance in observations	157.4	100.0%	100.0%
Jan 2011	4	Variance explained by model	70.4	44.7%	43.9%
Jan 2011	4	Unexplained variance (total)	87	55.3%	56.1%
Jan 2011	4	Unexplained variance explained by 1st factor	1.6	1.0%	
Jan 2011	5	Total variance in observations	171.5	100.0%	100.0%
Jan 2011	5	Variance explained by model	85.5	49.8%	50.5%
Jan 2011	5	Unexplained variance (total)	86	50.2%	49.5%
Jan 2011	5	Unexplained variance explained by 1st factor	1.7	1.0%	
Jan 2011	6	Total variance in observations	442.8	100.0%	100.0%
Jan 2011	6	Variance explained by model	232.8	52.6%	53.5%
Jan 2011	6	Unexplained variance (total)	210	47.4%	46.5%
Jan 2011	6	Unexplained variance explained by 1st factor	2.3	0.5%	
Jan 2011	7	Total variance in observations	364.4	100.0%	100.0%
Jan 2011	7	Variance explained by model	172.4	47.3%	46.8%
Jan 2011	7	Unexplained variance (total)	192	52.7%	53.2%
Jan 2011	7	Unexplained variance explained by 1st factor	2.1	0.6%	
Jan 2011	8	Total variance in observations	345.5	100.0%	100.0%
Jan 2011	8	Variance explained by model	153.5	44.4%	44.5%
Jan 2011	8	Unexplained variance (total)	192	55.6%	55.5%
Jan 2011	8	Unexplained variance explained by 1st factor	2.0	0.6%	
Jan 2011	Literature	Total variance in observations	699.1	100.0%	100.0%
Jan 2011	Literature	Variance explained by model	351.1	50.2%	50.2%
Jan 2011	Literature	Unexplained variance (total)	348	49.8%	49.8%
Jan 2011	Literature	Unexplained variance explained by 1st factor	2.2	0.3%	
July 2013	6	Total variance in observations	111.7	100.0%	100.0%
July 2013	6	Variance explained by model	55.7	49.8%	47.3%
July 2013	6	Unexplained variance (total)	56	50.2%	52.7%
July 2013	6	Unexplained variance explained by 1st factor	1.5	1.3%	

Table 8–2 (continued). Results from PCA of Residuals in WINSTEPS for Reading

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
July 2013	7	Total variance in observations	103.4	100.0%	100.0%
July 2013	7	Variance explained by model	45.4	43.9%	42.2%
July 2013	7	Unexplained variance (total)	58	56.1%	57.8%
July 2013	7	Unexplained variance explained by 1st factor	1.4	1.4%	
July 2013	8	Total variance in observations	105.4	100.0%	100.0%
July 2013	8	Variance explained by model	48.4	45.9%	44.8%
July 2013	8	Unexplained variance (total)	57	54.1%	55.2%
July 2013	8	Unexplained variance explained by 1st factor	1.4	1.3%	
Jan 2014	K–2*	Total variance in observations	656.5	100.0%	100.0%
Jan 2014	K–2*	Variance explained by model	376.5	57.4%	57.6%
Jan 2014	K–2*	Unexplained variance (total)	280	42.6%	42.4%
Jan 2014	K–2*	Unexplained variance explained by 1st factor	1.9	0.3%	
Jan 2014	3	Total variance in observations	391.5	100.0%	100.0%
Jan 2014	3	Variance explained by model	213.5	54.5%	55.6%
Jan 2014	3	Unexplained variance (total)	178	45.5%	44.4%
Jan 2014	3	Unexplained variance explained by 1st factor	1.9	0.5%	
Jan 2014	4	Total variance in observations	434.7	100.0%	100.0%
Jan 2014	4	Variance explained by model	245.7	56.5%	57.1%
Jan 2014	4	Unexplained variance (total)	189	43.5%	42.9%
Jan 2014	4	Unexplained variance explained by 1st factor	1.7	0.4%	
Jan 2014	4	Total variance in observations	434.7	100.0%	100.0%
Jan 2014	4	Variance explained by model	245.7	56.5%	57.1%
Jan 2014	4	Unexplained variance (total)	189	43.5%	42.9%
Jan 2014	4	Unexplained variance explained by 1st factor	1.7	0.4%	
June 2016	3	Total variance in observations	53.5	100.0%	100.0%
June 2016	3	Variance explained by model	31.5	58.8%	41.7%
June 2016	3	Unexplained variance (total)	22	41.2%	58.3%
June 2016	3	Unexplained variance explained by 1st factor	1.1	2.1%	
June 2016	4	Total variance in observations	54.3	100.0%	100.0%
June 2016	4	Variance explained by model	33.3	61.4%	37.4%
June 2016	4	Unexplained variance (total)	21	38.6%	62.6%
June 2016	4	Unexplained variance explained by 1st factor	1.7	3.2%	
June 2016	5	Total variance in observations	57.5	100.0%	100.0%
June 2016	5	Variance explained by model	36.5	63.5%	43.5%
June 2016	5	Unexplained variance (total)	21	36.5%	56.6%
June 2016	5	Unexplained variance explained by 1st factor	1.2	2.1%	

Table 8–2 (continued). Results from PCA of Residuals in WINSTEPS for Reading

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2016	6	Total variance in observations	232.3	100.0%	100.0%
June 2016	6	Variance explained by model	110.3	47.5%	45.1%
June 2016	6	Unexplained variance (total)	122	52.5%	54.9%
June 2016	6	Unexplained variance explained by 1st factor	1.6	0.7%	
June 2016	7	Total variance in observations	245.8	100.0%	100.0%
June 2016	7	Variance explained by model	120.8	49.1%	47.2%
June 2016	7	Unexplained variance (total)	125	50.9%	52.8%
June 2016	7	Unexplained variance explained by 1st factor	1.6	0.6%	
June 2016	8	Variance explained by model	132.5	51.9%	49.8%
June 2016	8	Unexplained variance (total)	123	48.1%	50.2%
June 2016	8	Unexplained variance explained by 1st factor	1.7	0.7%	
June 2016	Literature	Total variance in observations	206.4	100.0%	100.0%
June 2016	Literature	Variance explained by model	82.4	39.9%	39.0%
June 2016	Literature	Unexplained variance (total)	124	60.1%	61.0%
June 2016	Literature	Unexplained variance explained by 1st factor	1.5	0.7%	
June 2019	K–2*	Total variance in observations	117.8	100.0%	100.0%
June 2019	K–2*	Variance explained by model	33.8	28.7%	26.0%
June 2019	K–2*	Unexplained variance (total)	84.0	71.3%	74.0%
June 2019	K–2*	Unexplained variance explained by 1st factor	1.5	1.3%	
June 2019	3	Total variance in observations	272.3	100.0%	100.0%
June 2019	3	Variance explained by model	110.3	40.5%	39.3%
June 2019	3	Unexplained variance (total)	162.0	59.5%	60.7%
June 2019	3	Unexplained variance explained by 1st factor	1.6	0.6%	
June 2019	4	Total variance in observations	288.0	100.0%	100.0%
June 2019	4	Variance explained by model	126.0	43.8%	42.7%
June 2019	4	Unexplained variance (total)	162.0	56.2%	57.3%
June 2019	4	Unexplained variance explained by 1st factor	1.7	0.6%	
June 2019	5	Total variance in observations	291.2	100.0%	100.0%
June 2019	5	Variance explained by model	129.2	44.4%	42.7%
June 2019	5	Unexplained variance (total)	162.0	55.6%	57.3%
June 2019	5	Unexplained variance explained by 1st factor	1.6	0.6%	
June 2019	6	Total variance in observations	216.3	100.0%	100.0%
June 2019	6	Variance explained by model	93.3	43.1%	42.2%
June 2019	6	Unexplained variance (total)	123.0	56.9%	57.8%
June 2019	6	Unexplained variance explained by 1st factor	1.6	0.7%	

Table 8–2 (continued). Results from PCA of Residuals in WINSTEPS for Reading

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2019	7	Variance explained by model	89.2	42.0%	41.2%
June 2019	7	Unexplained variance (total)	123.0	58.0%	58.8%
June 2019	7	Unexplained variance explained by 1st factor	1.6	0.8%	
June 2019	8	Total variance in observations	209.3	100.0%	100.0%
June 2019	8	Variance explained by model	89.3	42.7%	41.5%
June 2019	8	Unexplained variance (total)	120.0	57.3%	58.5%
June 2019	8	Unexplained variance explained by 1st factor	1.7	0.8%	
June 2019	Literature	Total variance in observations	396.2	100.0%	100.0%
June 2019	Literature	Variance explained by model	147.2	37.2%	36.4%
June 2019	Literature	Unexplained variance (total)	249.0	62.8%	63.6%
June 2019	Literature	Unexplained variance explained by 1st factor	1.5	0.4%	

*Items in kindergarten through grade 2 were co-mingled on forms taken by students in grade 3.

Table 8–3 presents the PCA results for the CDT science item pool. The primary dimension in the Rasch model explained between 20 and 68 percent of the total variances across the grades and courses. The second dimension (the row named "unexplained variance explained by 1st factor") accounted for between 0.3 and 4.6 percent of the total variance in observations. These results suggest that the CDT science item pool essentially measures a single dominant dimension.

Table 8–3. Results from PCA of Residuals in WINSTEPS for Science

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Jan 2011	3	Total variance in observations	229.1	100.0%	100.0%
Jan 2011	3	Variance explained by model	138.1	60.3%	60.3%
Jan 2011	3	Unexplained variance (total)	91	39.7%	39.7%
Jan 2011	3	Unexplained variance explained by 1st factor	1.7	0.7%	
Jan 2011	4	Total variance in observations	285.9	100.0%	100.0%
Jan 2011	4	Variance explained by model	162.9	57.0%	56.9%
Jan 2011	4	Unexplained variance (total)	123	43.0%	43.1%
Jan 2011	4	Unexplained variance explained by 1st factor	1.5	0.5%	
Jan 2011	5	Total variance in observations	161.9	100.0%	100.0%
Jan 2011	5	Variance explained by model	59.9	37.0%	37.4%
Jan 2011	5	Unexplained variance (total)	102	63.0%	62.6%
Jan 2011	5	Unexplained variance explained by 1st factor	1.5	0.9%	
Jan 2011	6	Total variance in observations	290.8	100.0%	100.0%
Jan 2011	6	Variance explained by model	112.8	38.8%	39.3%
Jan 2011	6	Unexplained variance (total)	178	61.2%	60.7%
Jan 2011	6	Unexplained variance explained by 1st factor	2.1	0.7%	
Jan 2011	7	Total variance in observations	487.1	100.0%	100.0%
Jan 2011	7	Variance explained by model	160.1	32.9%	33.3%
Jan 2011	7	Unexplained variance (total)	327	67.1%	66.7%
Jan 2011	7	Unexplained variance explained by 1st factor	2.2	0.4%	
Jan 2011	8*	Total variance in observations	658.8	100.0%	100.0%
Jan 2011	8*	Variance explained by model	281.8	42.8%	43.9%
Jan 2011	8*	Unexplained variance (total)	377	57.2%	56.1%
Jan 2011	8*	Unexplained variance explained by 1st factor	1.9	0.3%	
Jan 2011	Biology	Total variance in observations	545.2	100.0%	100.0%
Jan 2011	Biology	Variance explained by model	155.2	28.5%	29.7%
Jan 2011	Biology	Unexplained variance (total)	390	71.5%	70.3%
Jan 2011	Biology	Unexplained variance explained by 1st factor	2.0	0.4%	
Jan 2011	Chemistry	Total variance in observations	418.1	100.0%	100.0%
Jan 2011	Chemistry	Variance explained by model	83.1	19.9%	20.1%
Jan 2011	Chemistry	Unexplained variance (total)	335	80.1%	79.9%
Jan 2011	Chemistry	Unexplained variance explained by 1st factor	2.0	0.5%	
Jan 2014	K–2	Total variance in observations	652.2	100.0%	100.0%
Jan 2014	K–2	Variance explained by model	372.2	57.1%	57.4%

Table 8–3 (continued). Results from PCA of Residuals in WINSTEPS for Science

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Jan 2014	K–2	Unexplained variance (total)	280	42.9%	42.6%
Jan 2014	K–2	Unexplained variance explained by 1st factor	2.6	0.4%	
Jan 2014	3	Total variance in observations	369.9	100.0%	100.0%
Jan 2014	3	Variance explained by model	214.9	58.1%	57.8%
Jan 2014	3	Unexplained variance (total)	155	41.9%	42.2%
Jan 2014	3	Unexplained variance explained by 1st factor	2.0	0.5%	
Jan 2014	4	Total variance in observations	668.3	100.0%	100.0%
Jan 2014	4	Variance explained by model	455.3	68.1%	68.0%
Jan 2014	4	Unexplained variance (total)	213	31.9%	32.0%
Jan 2014	4	Unexplained variance explained by 1st factor	2.0	0.3%	
Jan 2014	5	Total variance in observations	235.5	100.0%	100.0%
Jan 2014	5	Variance explained by model	83.5	35.5%	34.5%
Jan 2014	5	Unexplained variance (total)	152	64.5%	65.5%
Jan 2014	5	Unexplained variance explained by 1st factor	1.3	0.6%	
June 2016	6	Total variance in observations	99.6	100.0%	100.0%
June 2016	6	Variance explained by model	33.6	33.7%	29.2%
June 2016	6	Unexplained variance (total)	66	66.3%	70.8%
June 2016	6	Unexplained variance explained by 1st factor	1.1	1.1%	
June 2016	7	Total variance in observations	218.9	100.0%	100.0%
June 2016	7	Variance explained by model	65.9	30.1%	24.9%
June 2016	7	Unexplained variance (total)	153	69.9%	75.1%
June 2016	7	Unexplained variance explained by 1st factor	1.1	0.5%	
June 2016	8	Total variance in observations	338.2	100.0%	100.0%
June 2016	8	Variance explained by model	112.2	33.2%	28.2%
June 2016	8	Unexplained variance (total)	226	66.8%	71.8%
June 2016	8	Unexplained variance explained by 1st factor	1.2	0.3%	
June 2016	Biology	Total variance in observations	205.4	100.0%	100.0%
June 2016	Biology	Variance explained by model	70.4	34.3%	32.0%
June 2016	Biology	Unexplained variance (total)	135	65.7%	68.0%
June 2016	Biology	Unexplained variance explained by 1st factor	1.1	0.5%	
June 2019	K–2**	Total variance in observations	49.6	100.0%	100.0%
June 2019	K–2**	Variance explained by model	18.6	37.5%	26.9%
June 2019	K–2**	Unexplained variance (total)	31.0	62.5%	73.1%
June 2019	K–2**	Unexplained variance explained by 1st factor	1.4	2.8%	
June 2019	3	Total variance in observations	154.7	100.0%	100.0%
June 2019	3	Variance explained by model	65.7	42.5%	36.0%
June 2019	3	Unexplained variance (total)	89.0	57.5%	64.0%
June 2019	3	Unexplained variance explained by 1st factor	1.1	0.7%	

Table 8–3 (continued). Results from PCA of Residuals in WINSTEPS for Science

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2019	4	Total variance in observations	140.1	100.0%	100.0%
June 2019	4	Variance explained by model	45.1	32.2%	27.3%
June 2019	4	Unexplained variance (total)	95.0	67.8%	72.7%
June 2019	4	Unexplained variance explained by 1st factor	1.1	0.8%	
June 2019	5	Total variance in observations	128.0	100.0%	100.0%
June 2019	5	Variance explained by model	38.0	29.7%	24.8%
June 2019	5	Unexplained variance (total)	90.0	70.3%	75.2%
June 2019	5	Unexplained variance explained by 1st factor	1.1	0.9%	
June 2019	6	Total variance in observations	136.2	100.0%	100.0%
June 2019	6	Variance explained by model	39.2	28.8%	24.2%
June 2019	6	Unexplained variance (total)	97.0	71.2%	75.8%
June 2019	6	Unexplained variance explained by 1st factor	1.1	0.8%	
June 2019	7	Total variance in observations	135.5	100.0%	100.0%
June 2019	7	Variance explained by model	36.5	26.9%	22.3%
June 2019	7	Unexplained variance (total)	99.0	73.1%	77.7%
June 2019	7	Unexplained variance explained by 1st factor	1.1	0.8%	
June 2019	8	Total variance in observations	152.6	100.0%	100.0%
June 2019	8	Variance explained by model	50.6	33.1%	27.6%
June 2019	8	Unexplained variance (total)	102.0	66.9%	72.4%
June 2019	8	Unexplained variance explained by 1st factor	1.1	0.7%	
June 2019	Biology	Total variance in observations	414.2	100.0%	100.0%
June 2019	Biology	Variance explained by model	124.2	30.0%	28.7%
June 2019	Biology	Unexplained variance (total)	290.0	70.0%	71.3%
June 2019	Biology	Unexplained variance explained by 1st factor	1.1	0.3%	
June 2019	Chemistry	Total variance in observations	142.9	100.0%	100.0%
June 2019	Chemistry	Variance explained by model	42.9	30.0%	27.5%
June 2019	Chemistry	Unexplained variance (total)	100.0	70.0%	72.5%
June 2019	Chemistry	Unexplained variance explained by 1st factor	1.3	0.9%	
May 2020	3	Total variance in observations	50.3	100.0%	100.0%
May 2020	3	Variance explained by model	31.3	62.2%	37.5%
May 2020	3	Unexplained variance (total)	19.0	37.8%	62.5%
May 2020	3	Unexplained variance explained by 1st factor	1.2	2.5%	
May 2020	4	Total variance in observations	50.1	100.0%	100.0%
May 2020	4	Variance explained by model	28.1	56.1%	34.2%
May 2020	4	Unexplained variance (total)	22.0	43.9%	65.8%
May 2020	4	Unexplained variance explained by 1st factor	1.3	2.5%	
May 2020	5	Total variance in observations	83.4	100.0%	100.0%
May 2020	5	Variance explained by model	63.4	76.0%	60.8%

Table 8–3 (continued). Results from PCA of Residuals in WINSTEPS for Science

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
May 2020	5	Unexplained variance (total)	20.0	24.0%	39.2%
May 2020	5	Unexplained variance explained by 1st factor	1.4	1.7%	
May 2020	6	Total variance in observations	27.8	100.0%	100.0%
May 2020	6	Variance explained by model	9.8	35.1%	19.8%
May 2020	6	Unexplained variance (total)	18.0	64.9%	80.2%
May 2020	6	Unexplained variance explained by 1st factor	1.3	4.6%	
May 2020	7	Total variance in observations	44.6	100.0%	100.0%
May 2020	7	Variance explained by model	25.6	57.4%	34.3%
May 2020	7	Unexplained variance (total)	19.0	42.6%	65.7%
May 2020	7	Unexplained variance explained by 1st factor	1.3	2.8%	
May 2020	8	Total variance in observations	39.3	100.0%	100.0%
May 2020	8	Variance explained by model	19.3	49.1%	29.4%
May 2020	8	Unexplained variance (total)	20.0	50.9%	70.6%
May 2020	8	Unexplained variance explained by 1st factor	1.3	3.2%	
May 2020	Biology	Total variance in observations	72.7	100.0%	100.0%
May 2020	Biology	Variance explained by model	32.7	45.0%	28.5%
May 2020	Biology	Unexplained variance (total)	40.0	55.0%	71.5%
May 2020	Biology	Unexplained variance explained by 1st factor	1.1	1.5%	

*Grade 11 items were tested on grade 8 forms.

Table 8–4 presents the PCA results for the CDT writing item pool. The primary dimension in the Rasch model explained between 22 and 55 percent of the total variances across the grades and courses. The second dimension (the row named "unexplained variance explained by 1st factor") accounted for between 0.3 and 2.2 percent of the total variance in observations. These results suggest that the CDT writing item pool essentially measures a single dominant dimension.

Table 8–4. Results from PCA of Residuals in WINSTEPS for Writing

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Aug 2011	3	Total variance in observations	297.7	100.0%	100.0%
Aug 2011	3	Variance explained by model	157.7	53.0%	55.0%
Aug 2011	3	Unexplained variance (total)	140	47.0%	45.0%
Aug 2011	3	Unexplained variance explained by 1st factor	1.7	0.6%	
Aug 2011	4	Total variance in observations	283.6	100.0%	100.0%
Aug 2011	4	Variance explained by model	134.6	47.5%	49.0%
Aug 2011	4	Unexplained variance (total)	149	52.5%	51.0%
Aug 2011	4	Unexplained variance explained by 1st factor	1.8	0.6%	
Aug 2011	5	Total variance in observations	280.7	100.0%	100.0%
Aug 2011	5	Variance explained by model	115.7	41.2%	42.2%
Aug 2011	5	Unexplained variance (total)	165	58.8%	57.8%
Aug 2011	5	Unexplained variance explained by 1st factor	1.8	0.6%	
Aug 2011	6	Total variance in observations	340.5	100.0%	100.0%
Aug 2011	6	Variance explained by model	147.5	43.3%	44.2%
Aug 2011	6	Unexplained variance (total)	193	56.7%	55.8%
Aug 2011	6	Unexplained variance explained by 1st factor	2.0	0.6%	
Aug 2011	7	Total variance in observations	317.9	100.0%	100.0%
Aug 2011	7	Variance explained by model	141.9	44.6%	45.5%
Aug 2011	7	Unexplained variance (total)	176	55.4%	54.5%
Aug 2011	7	Unexplained variance explained by 1st factor	2.1	0.6%	
Aug 2011	8	Total variance in observations	336.0	100.0%	100.0%
Aug 2011	8	Variance explained by model	141.0	42.0%	42.4%
Aug 2011	8	Unexplained variance (total)	195	58.0%	57.6%
Aug 2011	8	Unexplained variance explained by 1st factor	2.3	0.7%	
Aug 2011	English Composition	Total variance in observations	763.2	100.0%	100.0%
Aug 2011	English Composition	Variance explained by model	398.2	52.2%	53.4%
Aug 2011	English Composition	Unexplained variance (total)	365	47.8%	46.6%
Aug 2011	English Composition	Unexplained variance explained by 1st factor	2.3	0.3%	
Jan 2014	K–2*	Total variance in observations	93.2	100.0%	100.0%
Jan 2014	K–2*	Variance explained by model	49.2	52.8%	39.9%
Jan 2014	K–2*	Unexplained variance (total)	44	47.2%	60.1%
Jan 2014	K–2*	Unexplained variance explained by 1st factor	2.0	2.2%	
Jan 2014	3	Total variance in observations	132.5	100.0%	100.0%
Jan 2014	3	Variance explained by model	72.5	54.7%	54.6%

Table 8–4 (continued). Results from PCA of Residuals in WINSTEPS for Writing

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
Jan 2014	3	Unexplained variance (total)	60	45.3%	45.4%
Jan 2014	3	Unexplained variance explained by 1st factor	1.8	1.4%	
Jan 2014	4	Total variance in observations	132.4	100.0%	100.0%
Jan 2014	4	Variance explained by model	72.4	54.7%	55.4%
Jan 2014	4	Unexplained variance (total)	60	45.3%	44.6%
Jan 2014	4	Unexplained variance explained by 1st factor	1.7	1.3%	
Jan 2014	5	Total variance in observations	146.5	100.0%	100.0%
Jan 2014	5	Variance explained by model	75.5	51.5%	47.7%
Jan 2014	5	Unexplained variance (total)	71	48.5%	52.3%
Jan 2014	5	Unexplained variance explained by 1st factor	1.3	0.9%	
June 2016	6	Total variance in observations	154.7	100.0%	100.0%
June 2016	6	Variance explained by model	64.7	41.8%	38.2%
June 2016	6	Unexplained variance (total)	90	58.2%	61.8%
June 2016	6	Unexplained variance explained by 1st factor	1.2	0.8%	
June 2016	7	Total variance in observations	126.6	100.0%	100.0%
June 2016	7	Variance explained by model	34.6	27.3%	22.4%
June 2016	7	Unexplained variance (total)	92	72.7%	77.6%
June 2016	7	Unexplained variance explained by 1st factor	1.2	0.9%	
June 2016	8	Total variance in observations	150.7	100.0%	100.0%
June 2016	8	Variance explained by model	44.7	29.7%	25.2%
June 2016	8	Unexplained variance (total)	106	70.3%	74.8%
June 2016	8	Unexplained variance explained by 1st factor	1.2	0.8%	
June 2016	English Composition	Total variance in observations	149.5	100.0%	100.0%
June 2016	English Composition	Variance explained by model	47.5	31.8%	26.3%
June 2016	English Composition	Unexplained variance (total)	102	68.2%	73.7%
June 2016	English Composition	Unexplained variance explained by 1st factor	1.3	0.9%	
June 2019	K–2*	Total variance in observations	69.0	100.0%	100.0%
June 2019	K–2*	Variance explained by model	37.0	53.6%	40.9%
June 2019	K–2*	Unexplained variance (total)	32.0	46.4%	59.1%
June 2019	K–2*	Unexplained variance explained by 1st factor	1.5	2.2%	
June 2019	3	Total variance in observations	165.5	100.0%	100.0%
June 2019	3	Variance explained by model	66.5	40.2%	36.4%
June 2019	3	Unexplained variance (total)	99.0	59.8%	63.6%
June 2019	3	Unexplained variance explained by 1st factor	1.3	0.8%	
June 2019	4	Total variance in observations	163.6	100.0%	100.0%
June 2019	4	Variance explained by model	73.6	45.0%	37.8%
June 2019	4	Unexplained variance (total)	90.0	55.0%	62.2%
June 2019	4	Unexplained variance explained by 1st factor	1.3	0.8%	

Table 8–4 (continued). Results from PCA of Residuals in WINSTEPS for Writing

Date	Grade/Course	Statistic	Eigenvalue	Empirical	Modeled
June 2019	5	Total variance in observations	139.1	100.0%	100.0%
June 2019	5	Variance explained by model	49.1	35.3%	29.9%
June 2019	5	Unexplained variance (total)	90.0	64.7%	70.1%
June 2019	5	Unexplained variance explained by 1st factor	1.2	0.9%	
June 2019	6	Total variance in observations	136.5	100.0%	100.0%
June 2019	6	Variance explained by model	43.5	31.9%	26.3%
June 2019	6	Unexplained variance (total)	93.0	68.1%	73.7%
June 2019	6	Unexplained variance explained by 1st factor	1.2	0.9%	
June 2019	7	Total variance in observations	158.9	100.0%	100.0%
June 2019	7	Variance explained by model	47.9	30.1%	25.7%
June 2019	7	Unexplained variance (total)	111.0	69.9%	74.3%
June 2019	7	Unexplained variance explained by 1st factor	1.2	0.7%	
June 2019	8	Total variance in observations	131.9	100.0%	100.0%
June 2019	8	Variance explained by model	38.9	29.5%	24.5%
June 2019	8	Unexplained variance (total)	93.0	70.5%	75.5%
June 2019	8	Unexplained variance explained by 1st factor	1.3	1.0%	
June 2019	English Composition	Total variance in observations	523.2	100.0%	100.0%
June 2019	English Composition	Variance explained by model	229.2	43.8%	41.3%
June 2019	English Composition	Unexplained variance (total)	294.0	56.2%	58.7%
June 2019	English Composition	Unexplained variance explained by 1st factor	1.5	0.3%	

*Items in kindergarten through grade 2 were co-mingled on forms taken by students in grade 3.

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 comprised of items X_1, X_2,…X_n is locally independent with respect to the latent variable θ if, for all x = (x_1, x_2,…x_n) and θ,

$$P\left(\mathbf{X}=\mathbf{x}\,|\,\theta\right)=\prod_{i=1}^{I}P\left(X_i=x_i\,|\,\theta\right).$$

This formula essentially states that the probability of any pattern of responses across all items (*x*), after conditioning on the abilities 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, as 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 show below. (This is a "weaker" form because higher-order dependencies among items are allowed.) Based on the WLI, the following expression can be derived:

$$P\left(X_i=x_i,X_j=x_j\,|\,\theta\right)=P\left(X_i=x_i\,|\,\theta\right)P\left(X_j=x_j\,|\,\theta\right).$$

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 CDT 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, deviation (residual) 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.

As previously mentioned, three types of residual correlations are available in WINSTEPS: raw, standardized, and logit. Since the three residual correlations are very similar, the default "standardized residual correlation" in WINSTEPS was used for these analyses. Tables 8–5 through 8–8 show the summary statistics—mean, standard deviation (SD), minimum (Min), maximum (Max), and several percentiles (P10, P25, P50, P75, P90)—for all the residual correlations for each content area and grade/course. The total number of item pairs (N) and the number of pairs with the residual correlations greater than 0.20 are also reported in the tables.

Table 8–5. Summary of Item Residual Correlations for Mathematics

Date	Grade/ Course	N	Mean	SD	Min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Max	< -.20	>.20
Aug 2010	3	1,372	-0.03	0.03	-0.15	-0.06	-0.04	-0.03	-0.01	0.01	0.32	0	2
Aug 2010	4	1,122	-0.03	0.04	-0.18	-0.08	-0.06	-0.03	-0.01	0.01	0.28	0	2
Aug 2010	5	1,132	-0.03	0.04	-0.17	-0.07	-0.05	-0.03	-0.01	0.01	0.38	0	1
Aug 2010	6	5,410	-0.02	0.04	-0.15	-0.06	-0.04	-0.02	0.00	0.02	0.34	0	12
Aug 2010	7	5,409	-0.02	0.04	-0.24	-0.07	-0.05	-0.02	0.00	0.03	0.35	3	4
Aug 2010	8	4,935	-0.02	0.06	-0.36	-0.10	-0.06	-0.02	0.01	0.05	0.27	18	3
Aug 2010	Algebra I	5,024	-0.02	0.04	-0.19	-0.07	-0.05	-0.02	0.00	0.02	0.26	0	2
Aug 2010	Geometry	5,470	-0.02	0.04	-0.20	-0.07	-0.04	-0.02	0.00	0.02	0.27	0	1
Aug 2010	Algebra II	5,457	-0.02	0.04	-0.18	-0.07	-0.05	-0.02	0.00	0.02	0.22	0	2
July 2013	6	12,090	-0.01	0.01	-0.12	-0.02	-0.01	0.00	0.00	0.00	0.06	0	0
July 2013	7	2,628	-0.01	0.01	-0.05	-0.03	-0.02	-0.01	-0.01	0.00	0.01	0	0
July 2013	8	12,246	-0.01	0.01	-0.09	-0.02	-0.01	0.00	0.00	0.01	0.06	0	0
Jan 2014	K–2	2,660	-0.04	0.06	-0.23	-0.11	-0.08	-0.05	-0.01	0.02	0.35	4	4
Jan 2014	3	2,278	-0.05	0.06	-0.24	-0.12	-0.09	-0.05	-0.01	0.02	0.27	12	2
Jan 2014	4	2,462	-0.05	0.05	-0.24	-0.11	-0.08	-0.05	-0.01	0.02	0.46	2	2
Jan 2014	5	24,310	0.00	0.01	-0.05	-0.01	-0.01	0.00	0.00	0.00	0.02	0	0
June 2016	6	6,903	-0.01	0.00	-0.03	-0.01	-0.01	-0.01	-0.01	0.00	0.01	0	0
June 2016	7	13,695	-0.01	0.00	-0.03	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2016	8	10,731	-0.01	0.01	-0.03	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2016	Algebra I	10,878	-0.01	0.00	-0.02	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2019	K–2*	1,770	-0.02	0.01	-0.09	-0.03	-0.02	-0.02	-0.01	0.00	0.02	0	0
June 2019	3	15,753	-0.01	0.00	-0.05	-0.01	-0.01	-0.01	0.00	0.00	0.02	0	0
June 2019	4	15,931	-0.01	0.00	-0.04	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2019	5	16,110	-0.01	0.00	-0.03	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2019	6	4,560	-0.01	0.00	-0.03	-0.02	-0.01	-0.01	-0.01	0.00	0.00	0	0
June 2019	7	5,253	-0.01	0.00	-0.03	-0.01	-0.01	-0.01	-0.01	0.00	0.01	0	0
June 2019	8	4,851	-0.01	0.01	-0.04	-0.02	-0.01	-0.01	-0.01	0.00	0.01	0	0
June 2019	Algebra I	44,551	0.00	0.00	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.01	0	0
June 2019	Geometry	4,950	-0.01	0.01	-0.05	-0.02	-0.02	-0.01	0.00	0.00	0.02	0	0
June 2019	Algebra II	4,950	-0.01	0.01	-0.07	-0.02	-0.02	-0.01	0.00	0.00	0.02	0	0

Table 8–6. Summary of Item Residual Correlations for Reading

Date	Grade/ Course	N	Mean	SD	Min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Max	< -.20	>.20
Jan 2011	3	1,334	-0.02	0.04	-0.17	-0.07	-0.04	-0.02	-0.01	0.01	0.14	0	0
Jan 2011	4	1,272	-0.02	0.03	-0.18	-0.07	-0.04	-0.02	-0.01	0.01	0.27	0	2
Jan 2011	5	1,262	-0.02	0.03	-0.17	-0.06	-0.04	-0.02	-0.01	0.01	0.18	0	0
Jan 2011	6	4,245	-0.02	0.05	-0.24	-0.07	-0.04	-0.02	0.00	0.02	0.35	2	13
Jan 2011	7	3,782	-0.02	0.04	-0.23	-0.07	-0.04	-0.02	0.00	0.02	0.22	2	1
Jan 2011	8	3,782	-0.02	0.04	-0.26	-0.07	-0.04	-0.02	0.00	0.03	0.34	2	5
Jan 2011	Literature	7,517	-0.02	0.05	-0.28	-0.09	-0.04	-0.01	0.01	0.04	0.40	25	10
July 2013	6	1,540	-0.02	0.05	-0.43	-0.03	-0.01	0.00	0.00	0.00	0.05	42	0
July 2013	7	1,653	-0.02	0.05	-0.33	-0.04	-0.01	0.00	0.00	0.00	0.01	38	0
July 2013	8	1,596	-0.02	0.05	-0.32	-0.04	-0.01	0.00	0.00	0.00	0.02	39	0
Jan 2014	K–2	2,660	-0.05	0.06	-0.26	-0.12	-0.09	-0.05	-0.01	0.02	0.29	7	5
Jan 2014	3	1,709	-0.05	0.05	-0.23	-0.11	-0.08	-0.05	-0.02	0.02	0.20	2	0
Jan 2014	4	1,888	-0.05	0.05	-0.23	-0.10	-0.08	-0.05	-0.02	0.01	0.20	1	0
Jan 2014	5	8,911	-0.01	0.02	-0.26	-0.01	-0.01	0.00	0.00	0.00	0.03	33	0
June 2016	3	231	-0.04	0.02	-0.10	-0.08	-0.06	-0.04	-0.02	-0.01	0.00	0	0
June 2016	4	210	-0.04	0.06	-0.74	-0.08	-0.06	-0.03	-0.02	0.00	0.01	1	0
June 2016	5	210	-0.04	0.03	-0.13	-0.09	-0.06	-0.04	-0.02	-0.01	0.00	0	0
June 2016	6	7,381	-0.01	0.04	-0.36	0.00	0.00	0.00	0.00	0.00	0.00	117	0
June 2016	7	7,750	-0.01	0.04	-0.40	0.00	0.00	0.00	0.00	0.00	0.09	123	0
June 2016	8	7,503	-0.01	0.04	-0.38	0.00	0.00	0.00	0.00	0.00	0.04	115	0
June 2016	Literature	7,626	-0.01	0.04	-0.33	0.00	0.00	0.00	0.00	0.00	0.00	161	0
June 2019	K–2*	3,486	-0.01	0.06	-0.45	0.00	0.00	0.00	0.00	0.00	0.00	121	0
June 2019	3	13,041	-0.01	0.04	-0.38	0.00	0.00	0.00	0.00	0.00	0.05	150	0
June 2019	4	13,041	-0.01	0.04	-0.40	0.00	0.00	0.00	0.00	0.00	0.06	152	0
June 2019	5	13,041	-0.01	0.04	-0.38	0.00	0.00	0.00	0.00	0.00	0.06	156	0
June 2019	6	7,503	-0.01	0.04	-0.46	0.00	0.00	0.00	0.00	0.00	0.06	121	0
June 2019	7	7,503	-0.01	0.04	-0.44	0.00	0.00	0.00	0.00	0.00	0.09	109	0
June 2019	8	7,140	-0.01	0.04	-0.48	0.00	0.00	0.00	0.00	0.00	0.00	118	0
June 2019	Literature	30,876	0.00	0.03	-0.33	0.00	0.00	0.00	0.00	0.00	0.02	319	0

Table 8–7. Summary of Item Residual Correlations for Science

Date	Grade/ Course	N	Mean	SD	Min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Max	< -.20	>.20
Jan 2011	3	1,400	-0.03	0.03	-0.16	-0.07	-0.04	-0.02	-0.01	0.01	0.09	0	0
Jan 2011	4	1,950	-0.02	0.03	-0.19	-0.07	-0.04	-0.02	0.00	0.01	0.09	0	0
Jan 2011	5	1,530	-0.03	0.03	-0.17	-0.07	-0.04	-0.02	-0.01	0.01	0.08	0	0
Jan 2011	6	3,642	-0.02	0.04	-0.18	-0.07	-0.04	-0.02	0.00	0.02	0.19	0	0
Jan 2011	7	6,934	-0.02	0.04	-0.22	-0.08	-0.04	-0.01	0.00	0.03	0.24	7	2
Jan 2011	8	6,881	-0.02	0.05	-0.27	-0.09	-0.04	-0.01	0.00	0.02	0.24	30	2
Jan 2011	Biology	8,255	-0.02	0.05	-0.24	-0.09	-0.04	-0.01	0.00	0.03	0.26	17	1
Jan 2011	Chemistry	7,105	-0.02	0.05	-0.22	-0.08	-0.04	-0.01	0.01	0.03	0.24	8	2
Jan 2014	K–2	2,660	-0.05	0.10	-0.43	-0.17	-0.11	-0.05	0.01	0.08	0.68	152	28
Jan 2014	3	1,510	-0.05	0.06	-0.33	-0.12	-0.09	-0.05	-0.01	0.03	0.25	5	3
Jan 2014	4	2,069	-0.05	0.09	-0.31	-0.16	-0.11	-0.05	0.01	0.07	0.32	83	13
Jan 2014	5	11,476	-0.01	0.01	-0.08	-0.02	-0.01	-0.01	0.00	0.01	0.06	0	0
June 2016	6	2,145	-0.02	0.01	-0.05	-0.03	-0.02	-0.02	-0.01	0.00	0.02	0	0
June 2016	7	11,628	-0.01	0.01	-0.04	-0.01	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2016	8	25,425	0.00	0.01	-0.03	-0.01	-0.01	0.00	0.00	0.00	0.02	0	0
June 2016	Biology	9,045	-0.01	0.00	-0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0	0
June 2019	K–2*	465	-0.03	0.03	-0.15	-0.07	-0.05	-0.03	-0.01	0.00	0.04	0	0
June 2019	3	3,916	-0.01	0.01	-0.05	-0.02	-0.02	-0.01	-0.01	0.00	0.01	0	0
June 2019	4	4,465	-0.01	0.01	-0.04	-0.02	-0.02	-0.01	0.00	0.00	0.02	0	0
June 2019	5	4,005	-0.01	0.01	-0.06	-0.02	-0.02	-0.01	-0.01	0.00	0.01	0	0
June 2019	6	4,656	-0.01	0.01	-0.04	-0.02	-0.01	-0.01	-0.01	0.00	0.01	0	0
June 2019	7	4,851	-0.01	0.00	-0.03	-0.02	-0.01	-0.01	-0.01	0.00	0.00	0	0
June 2019	8	5,151	-0.01	0.01	-0.03	-0.02	-0.01	-0.01	-0.01	0.00	0.01	0	0
June 2019	Biology	41,905	0.00	0.00	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.01	0	0
June 2019	Chemistry	4,950	-0.01	0.01	-0.06	-0.03	-0.02	-0.01	0.00	0.00	0.04	0	0
May 2020	3	171	-0.05	0.02	-0.14	-0.08	-0.07	-0.05	-0.04	-0.02	0.00	0	0
May 2020	4	231	-0.05	0.03	-0.14	-0.08	-0.06	-0.04	-0.02	-0.01	0.00	0	0
May 2020	5	190	-0.05	0.04	-0.24	-0.10	-0.07	-0.04	-0.02	-0.01	0.01	1	0
May 2020	6	153	-0.06	0.05	-0.26	-0.12	-0.07	-0.04	-0.03	-0.02	-0.01	5	0
May 2020	7	171	-0.05	0.03	-0.22	-0.09	-0.06	-0.05	-0.04	-0.03	0.00	1	0
May 2020	8	190	-0.05	0.03	-0.19	-0.09	-0.06	-0.04	-0.03	-0.02	0.00	0	0
May 2020	Biology	780	-0.02	0.01	-0.08	-0.04	-0.03	-0.02	-0.01	-0.01	0.00	0	0

Table 8–8. Summary of Item Residual Correlations for Writing

Date	Grade/ Course	N	Mean	SD	Min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Max	< -.20	>.20
Aug 2011	3	2,205	-0.02	0.05	-0.26	-0.08	-0.04	-0.02	0.00	0.02	0.19	6	0
Aug 2011	4	2,315	-0.02	0.05	-0.24	-0.09	-0.04	-0.02	0.00	0.02	0.28	9	2
Aug 2011	5	2,580	-0.02	0.05	-0.25	-0.09	-0.04	-0.02	0.00	0.02	0.19	11	0
Aug 2011	6	3,795	-0.02	0.05	-0.25	-0.08	-0.04	-0.02	0.01	0.03	0.27	4	5
Aug 2011	7	3,544	-0.02	0.05	-0.24	-0.08	-0.04	-0.02	0.00	0.03	0.24	10	2
Aug 2011	8	3,815	-0.02	0.07	-0.29	-0.11	-0.05	-0.02	0.01	0.06	0.29	58	13
Aug 2011	Eng. Comp	7,705	-0.02	0.06	-0.30	-0.10	-0.04	-0.01	0.01	0.05	0.33	72	18
Jan 2014	K–2	2,641	-0.05	0.09	-0.39	-0.15	-0.11	-0.05	0.01	0.06	0.35	84	19
Jan 2014	3	570	-0.05	0.06	-0.20	-0.12	-0.08	-0.05	-0.02	0.02	0.23	1	1
Jan 2014	4	570	-0.05	0.04	-0.18	-0.10	-0.08	-0.05	-0.02	0.01	0.21	0	1
Jan 2014	5	2,485	-0.01	0.02	-0.13	-0.04	-0.02	-0.01	0.00	0.01	0.05	0	0
June 2016	6	4,005	-0.01	0.01	-0.05	-0.02	-0.02	-0.01	-0.01	0.00	0.02	0	0
June 2016	7	4,186	-0.01	0.01	-0.06	-0.02	-0.02	-0.01	0.00	0.00	0.01	0	0
June 2016	8	5,565	-0.01	0.01	-0.05	-0.02	-0.01	-0.01	0.00	0.00	0.01	0	0
June 2016	Eng. Comp	5,151	-0.01	0.01	-0.13	-0.03	-0.02	-0.01	0.00	0.00	0.03	0	0
June 2019	K–2*	496	-0.03	0.04	-0.39	-0.07	-0.05	-0.02	-0.01	0.00	0.04	3	0
June 2019	3	4,851	-0.01	0.01	-0.21	-0.03	-0.02	-0.01	0.00	0.00	0.05	1	0
June 2019	4	4,005	-0.01	0.01	-0.12	-0.03	-0.02	-0.01	0.00	0.00	0.04	0	0
June 2019	5	4,005	-0.01	0.01	-0.09	-0.03	-0.02	-0.01	0.00	0.00	0.06	0	0
June 2019	6	4,278	-0.01	0.01	-0.07	-0.02	-0.02	-0.01	0.00	0.00	0.02	0	0
June 2019	7	6,105	-0.01	0.01	-0.05	-0.02	-0.01	-0.01	0.00	0.00	0.02	0	0
June 2019	8	4,278	-0.01	0.01	-0.14	-0.03	-0.02	-0.01	0.00	0.00	0.04	0	0
June 2019	English Composition	43,071	0.00	0.01	-0.24	-0.02	-0.01	0.00	0.00	0.01	0.18	2	0

Across the content areas and grades/courses, the mean residual correlations were slightly negative and the values were close to zero. The vast majority of the correlations were very small, suggesting local item independence generally holds for the CDT mathematics, reading, science, and writing item pools.

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. MnSq values are presented in this chapter.

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 following results, values outside of 0.7 to 1.3 are given practical importance.

Table 8–9 presents the summary statistics of infit and outfit mean square statistics for the CDT item pools, including the mean, standard deviation, minimum, and maximum values. The number of items within the range of (0.7, 1.3) is also reported in Table 8–9. As can been seen, the mean values for both fit statistics were close to 1.00 for nearly all grades/courses. Nearly all items had infit values falling in the range of (0.7, 1.3). These results indicate that the Rasch model fits the CDT data well.

Table 8–9. Summary of Infit and Outfit Mean Square Statistics

Date	Content Area	Grade/Course	Number of Items	Infit Mean	Infit SD	Infit Min	Infit Max	Infit [0.7,1.3]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit [0.7,1.3]
Aug 2010	Mathematics	3	86	0.99	0.08	0.78	1.17	86/86	0.99	0.24	0.21	1.56	71/86
Aug 2010	Mathematics	4	86	0.99	0.08	0.81	1.20	86/86	0.98	0.18	0.50	1.65	78/86
Aug 2010	Mathematics	5	85	0.99	0.12	0.80	1.32	84/85	1.00	0.24	0.46	1.56	69/85
Aug 2010	Mathematics	6	259	0.99	0.11	0.80	1.38	256/259	1.00	0.31	0.40	3.92	217/259
Aug 2010	Mathematics	7	258	1.00	0.12	0.80	1.49	253/258	1.01	0.25	0.56	2.24	213/258
Aug 2010	Mathematics	8	257	1.00	0.11	0.75	1.37	254/257	1.03	0.22	0.48	2.40	226/257
Aug 2010	Mathematics	11	149	0.99	0.10	0.80	1.27	149/149	0.99	0.18	0.67	1.67	141/149
Aug 2010	Mathematics	Algebra I	256	1.00	0.09	0.79	1.28	256/256	1.02	0.14	0.65	1.61	249/256
Aug 2010	Mathematics	Geometry	257	1.00	0.10	0.81	1.31	256/257	1.02	0.17	0.66	1.78	239/257
Aug 2010	Mathematics	Algebra II	256	1.00	0.10	0.78	1.41	254/256	1.03	0.20	0.66	1.99	233/256
Jan 2011	Reading	3	86	0.99	0.12	0.74	1.30	86/86	0.97	0.24	0.40	1.53	66/86
Jan 2011	Reading	4	87	0.99	0.10	0.79	1.28	87/87	0.95	0.22	0.32	1.58	74/87
Jan 2011	Reading	5	86	0.96	0.09	0.78	1.22	86/86	0.91	0.20	0.44	1.64	72/86
Jan 2011	Reading	6	210	1.01	0.13	0.70	1.30	210/210	1.02	0.31	0.37	2.65	151/210
Jan 2011	Reading	7	192	1.00	0.10	0.76	1.30	192/192	0.96	0.23	0.21	2.00	162/192
Jan 2011	Reading	8	192	0.98	0.11	0.75	1.33	191/192	0.96	0.22	0.41	1.84	158/192
Jan 2011	Reading	Literature	348	1.01	0.13	0.75	1.31	347/348	1.01	0.25	0.38	2.00	282/348
Jan 2011	Science	3	91	1.01	0.09	0.83	1.20	91/91	1.00	0.21	0.45	1.48	80/91
Jan 2011	Science	4	123	1.01	0.08	0.85	1.23	123/123	1.00	0.18	0.52	1.81	112/123
Jan 2011	Science	5	102	1.00	0.08	0.84	1.21	102/102	1.02	0.16	0.74	1.85	98/102
Jan 2011	Science	6	178	1.00	0.09	0.80	1.22	178/178	1.02	0.17	0.61	1.82	165/178
Jan 2011	Science	7	327	0.99	0.09	0.78	1.22	327/327	1.01	0.17	0.54	1.83	300/327
Jan 2011	Science	8	377	1.02	0.12	0.77	1.37	372/377	1.06	0.24	0.57	2.12	307/377
Jan 2011	Science	11	115	1.08	0.10	0.81	1.30	115/115	1.19	0.26	0.73	2.19	82/115
Jan 2011	Science	Biology	390	1.00	0.08	0.84	1.28	390/390	1.03	0.14	0.73	1.63	372/390
Jan 2011	Science	Chemistry	335	1.00	0.06	0.85	1.26	335/335	1.02	0.09	0.79	1.48	333/335
Aug 2011	Writing	3	140	0.99	0.11	0.80	1.43	139/140	1.00	0.24	0.42	1.95	115/140
Aug 2011	Writing	4	149	0.99	0.10	0.79	1.26	149/149	1.00	0.24	0.52	1.74	123/149
Aug 2011	Writing	5	165	0.98	0.09	0.80	1.24	165/165	0.97	0.19	0.62	1.92	151/165
Aug 2011	Writing	6	193	0.99	0.10	0.78	1.23	193/193	0.98	0.20	0.53	1.76	170/193
Aug 2011	Writing	7	176	1.00	0.11	0.75	1.36	175/176	1.02	0.23	0.56	1.92	147/176
Aug 2011	Writing	8	195	0.99	0.11	0.77	1.31	194/195	0.99	0.21	0.45	1.68	166/195
Aug 2011	Writing	Eng. Comp.	365	1.00	0.12	0.77	1.38	362/365	1.03	0.25	0.38	2.16	304/365
July 2013	Mathematics	6	156	1.07	0.14	0.78	1.50	144/156	1.35	0.62	0.51	4.77	96/156
July 2013	Mathematics	7	73	1.11	0.13	0.82	1.40	69/73	1.52	0.68	0.76	4.74	33/73
July 2013	Mathematics	8	157	1.14	0.13	0.87	1.45	138/157	1.61	0.58	0.85	3.46	62/157
July 2013	Reading	6	56	1.03	0.13	0.78	1.31	55/56	1.13	0.37	0.58	2.48	35/56

Table 8–9 (continued). Summary of Infit and Outfit Mean Square Statistics

Date	Content Area	Grade/Course	Number of Items	Infit Mean	Infit SD	Infit Min	Infit Max	Infit [0.7,1.3]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit [0.7,1.3]
July 2013	Reading	7	58	1.05	0.14	0.82	1.42	55/58	1.17	0.38	0.65	2.91	41/58
July 2013	Reading	8	57	1.03	0.13	0.78	1.32	56/57	1.11	0.29	0.48	2.03	42/57
Jan 2014	Mathematics	K	60	0.98	0.12	0.77	1.34	58/60	0.90	0.30	0.40	1.53	37/60
Jan 2014	Mathematics	1	91	0.97	0.12	0.76	1.33	89/91	0.92	0.30	0.23	2.00	61/91
Jan 2014	Mathematics	2	130	0.99	0.10	0.77	1.29	130/130	0.98	0.27	0.36	1.95	99/130
Jan 2014	Mathematics	3	235	0.99	0.12	0.77	1.44	231/235	1.02	0.31	0.47	3.11	191/235
Jan 2014	Mathematics	4	248	1.00	0.12	0.75	1.31	247/248	1.03	0.27	0.45	2.21	199/248
Jan 2014	Mathematics	5	221	1.02	0.11	0.79	1.37	218/221	1.07	0.25	0.58	2.22	182/221
Jan 2014	Reading	K	84	0.97	0.11	0.77	1.36	83/84	0.91	0.24	0.39	1.51	61/84
Jan 2014	Reading	1	98	0.99	0.12	0.77	1.35	96/98	1.02	0.35	0.36	2.75	73/98
Jan 2014	Reading	2	98	0.98	0.11	0.76	1.24	98/98	1.02	0.25	0.44	1.80	77/98
Jan 2014	Reading	3	178	1.00	0.12	0.77	1.29	178/178	1.04	0.31	0.43	2.44	127/178
Jan 2014	Reading	4	189	1.00	0.11	0.78	1.35	188/189	1.01	0.28	0.40	2.70	149/189
Jan 2014	Reading	5	134	1.01	0.11	0.77	1.28	134/134	1.04	0.24	0.44	1.91	112/134
Jan 2014	Science	K-2 grade span	280	0.99	0.13	0.73	1.43	273/280	1.01	0.34	0.23	2.79	199/280
Jan 2014	Science	3	155	0.99	0.11	0.72	1.29	155/155	0.98	0.28	0.23	1.99	114/155
Jan 2014	Science	4	213	1.00	0.11	0.70	1.27	213/213	1.01	0.24	0.37	1.88	179/213
Jan 2014	Science	5	152	1.07	0.15	0.70	1.59	141/152	1.16	0.29	0.50	2.39	111/152
Jan 2014	Writing	K	44	0.90	0.11	0.73	1.20	44/44	0.72	0.26	0.33	1.38	20/44
Jan 2014	Writing	1	118	0.96	0.15	0.70	1.42	117/118	0.89	0.32	0.27	1.76	74/118
Jan 2014	Writing	2	117	0.98	0.13	0.70	1.46	115/117	0.99	0.26	0.32	1.65	93/117
Jan 2014	Writing	3	60	0.98	0.12	0.78	1.22	60/60	0.98	0.27	0.35	1.97	48/60
Jan 2014	Writing	4	60	1.00	0.11	0.83	1.34	59/60	1.02	0.29	0.60	2.41	51/60
Jan 2014	Writing	5	71	1.03	0.13	0.71	1.37	70/71	1.13	0.40	0.61	2.59	48/71
June 2016	Mathematics	6	122	1.08	0.13	0.87	1.49	113/122	1.31	0.36	0.72	2.38	70/122
June 2016	Mathematics	7	176	1.09	0.13	0.84	1.54	161/176	1.42	0.48	0.74	3.42	89/176
June 2016	Mathematics	8	150	1.13	0.12	0.85	1.61	139/150	1.61	0.50	0.82	3.32	51/150
June 2016	Mathematics	Algebra I	149	1.10	0.09	0.85	1.36	148/149	1.49	0.47	0.73	3.45	57/149
June 2016	Reading	3	22	1.13	0.17	0.85	1.49	18/22	1.15	0.19	0.82	1.54	16/22
June 2016	Reading	4	22	1.10	0.15	0.87	1.44	19/22	1.15	0.30	0.76	2.24	19/22
June 2016	Reading	5	21	1.10	0.13	0.96	1.40	20/21	1.14	0.20	0.91	1.67	18/21
June 2016	Reading	6	123	1.06	0.13	0.81	1.54	121/123	1.13	0.29	0.58	2.48	98/123
June 2016	Reading	7	126	1.04	0.15	0.79	1.51	122/126	1.12	0.37	0.40	2.91	90/126
June 2016	Reading	8	124	1.06	0.16	0.79	2.00	115/124	1.16	0.40	0.50	3.14	82/124
June 2016	Reading	Literature	125	1.07	0.12	0.75	1.36	122/125	1.24	0.38	0.60	2.53	83/125
June 2016	Science	6	72	1.08	0.10	0.87	1.30	72/72	1.27	0.35	0.73	2.36	45/72
June 2016	Science	7	159	1.08	0.09	0.82	1.34	158/159	1.29	0.32	0.64	2.28	98/159

Table 8–9 (continued). Summary of Infit and Outfit Mean Square Statistics

Date	Content Area	Grade/Course	Number of Items	Infit Mean	Infit SD	Infit Min	Infit Max	Infit [0.7,1.3]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit [0.7,1.3]
June 2016	Science	8	238	1.07	0.10	0.77	1.34	236/238	1.27	0.36	0.50	3.55	151/238
June 2016	Science	Biology	136	1.08	0.10	0.87	1.51	135/136	1.25	0.24	0.83	1.94	88/136
June 2016	Writing	6	93	1.06	0.12	0.83	1.34	91/93	1.24	0.47	0.70	4.66	62/93
June 2016	Writing	7	93	1.08	0.10	0.81	1.39	91/93	1.31	0.45	0.70	3.14	59/93
June 2016	Writing	8	110	1.09	0.11	0.88	1.37	106/110	1.37	0.48	0.76	3.93	63/110
June 2016	Writing	Eng. Comp.	104	1.08	0.11	0.75	1.34	103/104	1.46	0.84	0.58	8.30	51/104
June 2019	Mathematics	K	20	1.00	0.14	0.84	1.38	19/20	0.97	0.27	0.63	1.57	14/20
June 2019	Mathematics	1	20	1.00	0.11	0.84	1.25	20/20	0.98	0.27	0.53	1.47	15/20
June 2019	Mathematics	2	20	0.97	0.10	0.79	1.14	20/20	1.00	0.40	0.59	2.50	18/20
June 2019	Mathematics	3	178	1.02	0.11	0.81	1.38	174/178	1.13	0.41	0.40	3.97	142/178
June 2019	Mathematics	4	179	1.03	0.10	0.80	1.27	179/179	1.12	0.28	0.53	2.17	139/179
June 2019	Mathematics	5	180	1.05	0.10	0.85	1.32	179/180	1.14	0.27	0.66	2.12	136/180
June 2019	Mathematics	6	96	1.09	0.11	0.88	1.45	93/96	1.28	0.32	0.71	2.22	60/96
June 2019	Mathematics	7	103	1.09	0.12	0.86	1.45	101/103	1.35	0.45	0.73	3.07	60/103
June 2019	Mathematics	8	99	1.14	0.12	0.89	1.40	93/99	1.62	0.56	0.86	4.00	32/99
June 2019	Mathematics	Algebra I	299	1.14	0.11	0.86	1.45	270/299	1.58	0.56	0.73	5.27	110/299
June 2019	Mathematics	Geometry	100	1.18	0.15	0.88	1.51	82/100	1.84	0.75	0.80	5.11	25/100
June 2019	Mathematics	Algebra II	100	1.13	0.15	0.86	1.58	86/100	1.61	0.60	0.86	3.80	37/100
June 2019	Reading	K	32	1.01	0.15	0.82	1.37	30/32	1.03	0.22	0.71	1.60	28/32
June 2019	Reading	1	20	1.00	0.10	0.87	1.22	20/20	1.00	0.14	0.76	1.27	20/20
June 2019	Reading	2	32	1.05	0.15	0.82	1.43	30/32	1.08	0.24	0.72	1.55	25/32
June 2019	Reading	3	162	1.10	0.16	0.81	1.68	145/162	1.16	0.27	0.59	2.51	125/162
June 2019	Reading	4	162	1.09	0.17	0.75	1.67	139/162	1.19	0.35	0.52	2.78	109/162
June 2019	Reading	5	162	1.07	0.16	0.77	1.64	147/162	1.16	0.33	0.61	2.35	116/162
June 2019	Reading	6	123	1.06	0.16	0.76	1.60	114/123	1.14	0.35	0.61	2.93	87/123
June 2019	Reading	7	123	1.09	0.16	0.76	1.62	112/123	1.20	0.34	0.54	2.59	85/123
June 2019	Reading	8	120	1.09	0.15	0.79	1.67	111/120	1.19	0.32	0.63	2.34	81/120
June 2019	Reading	Literature	249	1.07	0.14	0.74	1.53	238/249	1.20	0.39	0.59	3.41	171/249
June 2019	Science	K-2 grade span	31	1.11	0.16	0.82	1.37	27/31	1.37	0.58	0.72	3.23	16/31
June 2019	Science	3	89	1.09	0.12	0.78	1.38	87/89	1.37	0.45	0.67	3.20	45/89
June 2019	Science	4	95	1.11	0.11	0.79	1.39	92/95	1.39	0.47	0.60	3.58	44/95
June 2019	Science	5	90	1.09	0.11	0.81	1.30	90/90	1.27	0.31	0.62	2.34	51/90
June 2019	Science	6	97	1.10	0.12	0.84	1.34	95/97	1.28	0.36	0.68	2.53	60/97
June 2019	Science	7	99	1.08	0.10	0.86	1.29	99/99	1.21	0.27	0.72	2.37	73/99
June 2019	Science	8	102	1.08	0.10	0.86	1.31	101/102	1.27	0.36	0.75	2.60	64/102
June 2019	Science	Biology	290	1.11	0.12	0.81	1.45	276/290	1.31	0.33	0.67	2.94	166/290
June 2019	Science	Chemistry	100	1.06	0.08	0.87	1.25	100/100	1.16	0.16	0.78	1.83	84/100

Table 8–9 (continued). Summary of Infit and Outfit Mean Square Statistics

Date	Content Area	Grade/Course	Number of Items	Infit Mean	Infit SD	Infit Min	Infit Max	Infit [0.7,1.3]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit [0.7,1.3]
June 2019	Writing	K	10	0.96	0.12	0.73	1.10	10/10	0.81	0.20	0.37	1.14	8/10
June 2019	Writing	1	10	1.07	0.19	0.71	1.25	10/10	1.47	0.82	0.51	2.90	5/10
June 2019	Writing	2	12	1.13	0.17	0.90	1.36	9/12	2.15	2.02	0.85	6.91	7/12
June 2019	Writing	3	99	1.10	0.14	0.78	1.44	91/99	1.39	0.82	0.39	8.12	54/99
June 2019	Writing	4	90	1.09	0.13	0.81	1.45	85/90	1.31	0.60	0.55	4.05	54/90
June 2019	Writing	5	90	1.09	0.14	0.79	1.47	83/90	1.24	0.42	0.49	2.80	54/90
June 2019	Writing	6	93	1.12	0.13	0.78	1.41	87/93	1.30	0.39	0.47	2.80	48/93
June 2019	Writing	7	111	1.11	0.13	0.76	1.42	101/111	1.32	0.41	0.63	2.77	70/111
June 2019	Writing	8	93	1.14	0.14	0.79	1.48	82/93	1.41	0.58	0.57	3.84	47/93
June 2019	Writing	Eng. Comp.	294	1.16	0.21	0.66	1.72	219/294	1.61	1.00	0.22	9.69	124/294
May 2020	Science	3	19	1.04	0.09	0.87	1.16	19/19	1.27	0.42	0.81	2.26	13/19
May 2020	Science	4	22	1.05	0.15	0.88	1.54	21/22	1.28	0.48	0.78	2.98	14/22
May 2020	Science	5	20	1.00	0.08	0.86	1.19	20/20	1.01	0.30	0.40	1.84	15/20
May 2020	Science	6	18	1.06	0.10	0.92	1.27	18/18	1.17	0.21	0.88	1.66	13/18
May 2020	Science	7	19	1.05	0.12	0.81	1.25	19/19	1.15	0.25	0.74	1.76	15/19
May 2020	Science	8	20	1.02	0.10	0.85	1.20	20/20	1.16	0.32	0.79	2.02	13/20
May 2020	Science	Biology	40	1.04	0.12	0.87	1.48	39/40	1.16	0.27	0.76	1.79	30/40

RASCH ITEM STATISTICS

As noted earlier, the Rasch model expresses item difficulty (and student ability) in units referred to as *logits*, rather than on 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 in any calibration at zero. For each CDT stand-alone field-test event and content area, all grades and courses were calibrated separately with the exception of grade 11 items in Mathematics and Science. As a result, items in each grade or course were centered at zero. See Chapter Nine for a description of how item parameters within a content area were re-scaled across grades and courses to build a single (vertical) scale.

For each CDT embedded field-test event and content area, field-test items were calibrated anchoring on operational items' parameters. As a result, the embedded field-test items were placed on operational vertical scale.

Rasch item difficulty measure on the vertical scale and associated standard error for all items field tested prior to 2018-2019 can be found in Appendix B of the 20178-2018 technical report. Statistics for items field tested in 2018–2019 or later can be found in Appendix B of the corresponding year's technical report.

CHAPTER NINE: VERTICAL LINKING

The Classroom Diagnostic Tools (CDT) is designed to enable educators to identify students' academic strengths and areas of need. As such, it is necessary for some students to take items out of grade or course level. In order to do this, all items within a content area must be on a common (vertical) scale.

As previously mentioned in Chapter Eight, items from the first stand-alone field-test event for each CDT content area and grade or course were calibrated separately and centered at zero. This chapter outlines the procedures used for vertically linking CDT items across grades and courses within a content area. The end results are four separate vertical scales—one for each content area.

Also mentioned in Chapter Eight, for each content area, the items from all embedded field-test events and the second stand-alone field-test event were calibrated anchoring on operational items' parameters. As a result, all field-test items after the first stand-alone field-test events were placed on the operational vertical scale.

VERTICAL LINKING DESIGN

The first CDT stand-alone field tests were designed to build vertical scales across all grades and courses within a content area. In order to accomplish this, some field-test forms had items from one grade above or below in addition to on-grade or course-level items.

Stand-alone field tests in each content area had two types of forms:

1. Vertical linking form
2. On-grade-only form

Students who received vertical linking forms took a set of on-grade items and a set of items either one grade above or one grade below. Students who received on-grade-only forms took just on-grade items.

All items in the pool were field tested on one or more forms. In Mathematics, on-grade items were chained across adjacent forms to provide a horizontal link across forms within a grade. There were eight to ten horizontal links across adjacent forms. In all other content areas, 10 on-grade items appeared on each form within a grade or course. These common items provide a horizontal link across forms within a grade.[1]

Items used in vertical linking were administered to students one grade above or one grade below in order to link the forms across grades. DRC test development specialists selected items to be administered off-grade level with the following guidelines:

- There are two types of linking sets.
    - Items administered one grade below (e.g., grade 7 items administered to grade 6 students).
    - Items administered one grade above (e.g., grade 7 items administered to grade 8 students).
- Linking sets span the diagnostic categories.
- Linking sets span the estimated difficulty range (item developers estimate easy, medium, or hard).
- Students have a reasonable chance of correctly answering a linking item based on the instruction received.
    - For items administered in the grade above, students should have received instruction the previous year.
    - For items administered in the grade below, they should be extensions of concepts the students have already covered, not something completely new.

[1] The change in horizontal linking design after the Mathematics field test was in response to lower-than-expected participation. Using the same horizontal links on all forms within a grade results in higher *n*-counts.

In Mathematics, each set of linking items appeared on two forms, once located at the beginning and once located at the end to counterbalance possible position effect. In all other content areas, vertical linking items were co-mingled throughout the form with on-grade items.[2]

See Tables 6–1 through 6–4 in Chapter Six for details on the stand-alone field tests including number of items, number of forms, and number of vertical linking forms.

VERTICAL LINKING — MATHEMATICS

Links were made between adjacent grades, grade 8 to Algebra I, Algebra I to Algebra II, and grade 8 to Geometry. Table 9–1 below shows the number of linking items from the lower grade and the upper grade for each link. There were two sets of linking items for each link and direction. For example, in linking grade 5 to grade 6, there were 30 grade 5 items (lower grade) and 20 grade 6 items (upper grade). The 30 grade 5 items were in two sets of 15, while the 20 grade 6 items were in two sets of 10. The number of linking items differs across grades because forms in grades 3, 4, and 5 had 25 items total while all of the others had 35. There was no overlap of linking items among the sets.

Table 9–1. Mathematics Linking Item Detail

Link	Lower Grade	Upper Grade	Total
Grade 3 to Grade 4	20	20	40
Grade 4 to Grade 5	20	20	40
Grade 5 to Grade 6	30	20	50
Grade 6 to Grade 7	30	30	60
Grade 8 to Grade 7	30	30	60
Algebra I to Grade 8	30	30	60
Algebra II to Algebra I	30	30	60
Geometry to Grade 8	30	30	60

A visual representation of the vertical linking design is provided in Table 9–2. Rows are item level and columns are forms. For example, looking at the second row, you can see grade 4 items were on grades 3, 4, and 5 forms. Grade 4 items on grade 4 forms were on-grade items. Grade 4 items on grade 3 and grade 5 forms were vertical linking items. These items also appeared on grade 4 forms and were used to calculate the vertical linking shift parameter.

In linking grades 4 and 5, look at the four cells in Table 9–2 where grade 4 and grade 5 rows and columns cross. There were 86 grade 4 items, and of those 86 items, 20 items were also given to grade 5 as linking items. Similarly, there were 85 grade 5 items, and 20 out of the 85 items were given to grade 4 students as linking items.

Items used to link to a lower grade were different from items used to link to an upper grade. For example, the 30 grade 7 items administered on grade 6 forms were not the same as the 30 grade 7 items administered on grade 8 forms.

[2] The change in vertical linking design after the Mathematics field test was in response to lower-than-expected participation.

Table 9–2. Mathematics Vertical Linking Design of Forms

Gr. 3 Forms	Gr. 4 Forms	Gr. 5 Forms	Gr. 6 Forms	Gr. 7 Forms	Gr. 8 Forms	Alg I Forms	Geo Forms	Alg II Forms
Gr. 3 Items (86)	Gr. 3 Items (20)							
Gr. 4 Items (20)	Gr. 4 Items (86)	Gr. 4 Items (20)						
	Gr. 5 Items (20)	Gr. 5 Items (85)	Gr. 5 Items (30)					
		Gr. 6 Items (20)	Gr. 6 Items (259)	Gr. 6 Items (30)				
			Gr. 7 Items (30)	Gr. 7 Items (258)	Gr. 7 Items (30)			
				Gr. 8 Items (30)	Gr. 8 Items (257)	Gr. 8 Items (30)	Gr. 8 Items (30)	
					Gr. 11 Items (30)	Gr. 11 Items (50)	Gr. 11 Items (50)	Gr. 11 Items (50)
					Alg I Items (15)	Alg I Items (256)		Alg I Items (30)
					Geo Items (15)		Geo Items (257)	
						Alg II Items (30)		Alg II Items (256)

See Appendix C for details related to vertical linking items such as *n*-counts, Eligible Content, and diagnostic categories.

VERTICAL LINKING — READING

Links were made between adjacent grades and grade 8 to Literature. Table 9–3 shows the number of linking items from the lower grade and the upper grade for each link. There were two sets of linking items for each link and direction. For example, in linking grade 5 to grade 6, there were 20 grade 5 items (lower grade) and 20 grade 6 items (upper grade). The number of linking items was the same across grades.

Table 9–3. Reading Linking Item Detail

Link	Lower Grade	Upper Grade	Total
Grade 3 to Grade 4	20	20	40
Grade 4 to Grade 5	20	20	40
Grade 5 to Grade 6	20	20	40
Grade 6 to Grade 7	20	20	40
Grade 8 to Grade 7	20	20	40
Literature to Grade 8	20	20	40

A visual representation of the vertical linking design is provided in Table 9–4.

Table 9–4. Reading Vertical Linking Design of Forms

Gr. 3 Forms	Gr. 4 Forms	Gr. 5 Forms	Gr. 6 Forms	Gr. 7 Forms	Gr. 8 Forms	Lit Forms
Gr. 3 Items (86)	Gr. 3 Items (20)					
Gr. 4 Items (20)	Gr. 4 Items (87)	Gr. 4 Items (20)				
	Gr. 5 Items (20)	Gr. 5 Items (86)	Gr. 5 Items (20)			
		Gr. 6 Items (20)	Gr. 6 Items (210)	Gr. 6 Items (20)		
			Gr. 7 Items (20)	Gr. 7 Items (192)	Gr. 7 Items (20)	
				Gr. 8 Items (20)	Gr. 8 Items (192)	Gr. 8 Items (20)
					Lit Items (20)	Lit Items (348)

See Appendix C for details related to vertical linking items such as *n*-counts, Eligible Content, and diagnostic categories.

VERTICAL LINKING — SCIENCE

Links were made between adjacent grades, grade 8 to Biology, and grade 8 to Chemistry. Table 9–5 below shows the number of linking items from the lower grade and the upper grade for each link. There were two sets of linking items for each link and direction. For example, in linking grade 5 to grade 6, there were 20 grade 5 items (lower grade) and 20 grade 6 items (upper grade). The number of linking items was the same across grades.

Table 9–5. Science Linking Item Detail

Link	Lower Grade	Upper Grade	Total
Grade 3 to Grade 4	20	20	40
Grade 4 to Grade 5	20	20	40
Grade 5 to Grade 6	20	20	40
Grade 6 to Grade 7	20	20	40
Grade 8 to Grade 7	20	20	40
Biology to Grade 8	20	20	40
Chemistry to Grade 8	20	20	40

A visual representation of the vertical linking design is provided in Table 9–6.

Table 9–6. Science Vertical Linking Design of Forms

Gr. 3 Forms	Gr. 4 Forms	Gr. 5 Forms	Gr. 6 Forms	Gr. 7 Forms	Gr. 8 Forms	Bio Forms	Chem Forms
Gr. 3 Items (91)	Gr. 3 Items (20)						
Gr. 4 Items (20)	Gr. 4 Items (123)	Gr. 4 Items (20)					
	Gr. 5 Items (20)	Gr. 5 Items (102)	Gr. 5 Items (20)				
		Gr. 6 Items (20)	Gr. 6 Items (178)	Gr. 6 Items (20)			
			Gr. 7 Items (20)	Gr. 7 Items (327)	Gr. 7 Items (20)		
				Gr. 8 Items (20)	Gr. 8 Items (377)	Gr. 8 Items (20)	Gr. 8 Items (20)
					Gr. 11 Items (115)		
					Bio Items (20)	Bio Items (390)	
					Chem Items (20)		Chem Items (335)

See Appendix C for details related to vertical linking items such as *n*-counts, Eligible Content, and diagnostic categories.

VERTICAL LINKING — WRITING

Links were made between adjacent grades and grade 8 to English Composition. Table 9–7 shows the number of linking items from the lower grade and the upper grade for each link. There were two sets of linking items for each link and direction. For example, in linking grade 5 to grade 6, there were 20 grade 5 items (lower grade) and 20 grade 6 items (upper grade). The number of linking items was the same across grades.

Table 9–7. Writing Linking Item Detail

Link	Lower Grade	Upper Grade	Total
Grade 3 to Grade 4	20	20	40
Grade 4 to Grade 5	20	20	40
Grade 5 to Grade 6	20	20	40
Grade 6 to Grade 7	20	20	40
Grade 8 to Grade 7	20	20	40
English Composition to Grade 8	20	20	40

A visual representation of the vertical linking design is provided in Table 9–8.

Table 9–8. Writing Vertical Linking Design of Forms

Gr. 3 Forms	Gr. 4 Forms	Gr. 5 Forms	Gr. 6 Forms	Gr. 7 Forms	Gr. 8 Forms	Eng Forms
Gr. 3 Items (140)	Gr. 3 Items (20)					
Gr. 4 Items (20)	Gr. 4 Items (149)	Gr. 4 Items (20)				
	Gr. 5 Items (20)	Gr. 5 Items (165)	Gr. 5 Items (20)			
		Gr. 6 Items (20)	Gr. 6 Items (193)	Gr. 6 Items (20)		
			Gr. 7 Items (20)	Gr. 7 Items (176)	Gr. 7 Items (20)	
				Gr. 8 Items (20)	Gr. 8 Items (195)	Gr. 8 Items (20)
					Eng Items (20)	Eng Items (365)

See Appendix C for details related to vertical linking items such as *n*-counts, Eligible Content, and diagnostic categories.

THE VERTICAL LINKING PROCEDURE

Each of the CDT content area vertical scales was centered at grade 7. Adjacent-grade shift parameters were calculated and applied such that all items were vertically linked to grade 7. For example, grade 4 science items were placed on the science vertical scale by applying three shift parameters:

- shift between grades 4 and 5 science
- shift between grades 5 and 6 science
- shift between grades 6 and 7 science

The steps used to calculate adjacent-grade shift parameters are described below. All item calibrations were done with WINSTEPS software version 3.71 (Linacre, 2009). The grade 4 to grade 5 link is provided as an example for the steps.

1. Calibrate all on-grade items.
   - Calibrate grade 4 items on grade 4 forms.
   - Calibrate grade 5 items on grade 5 forms.

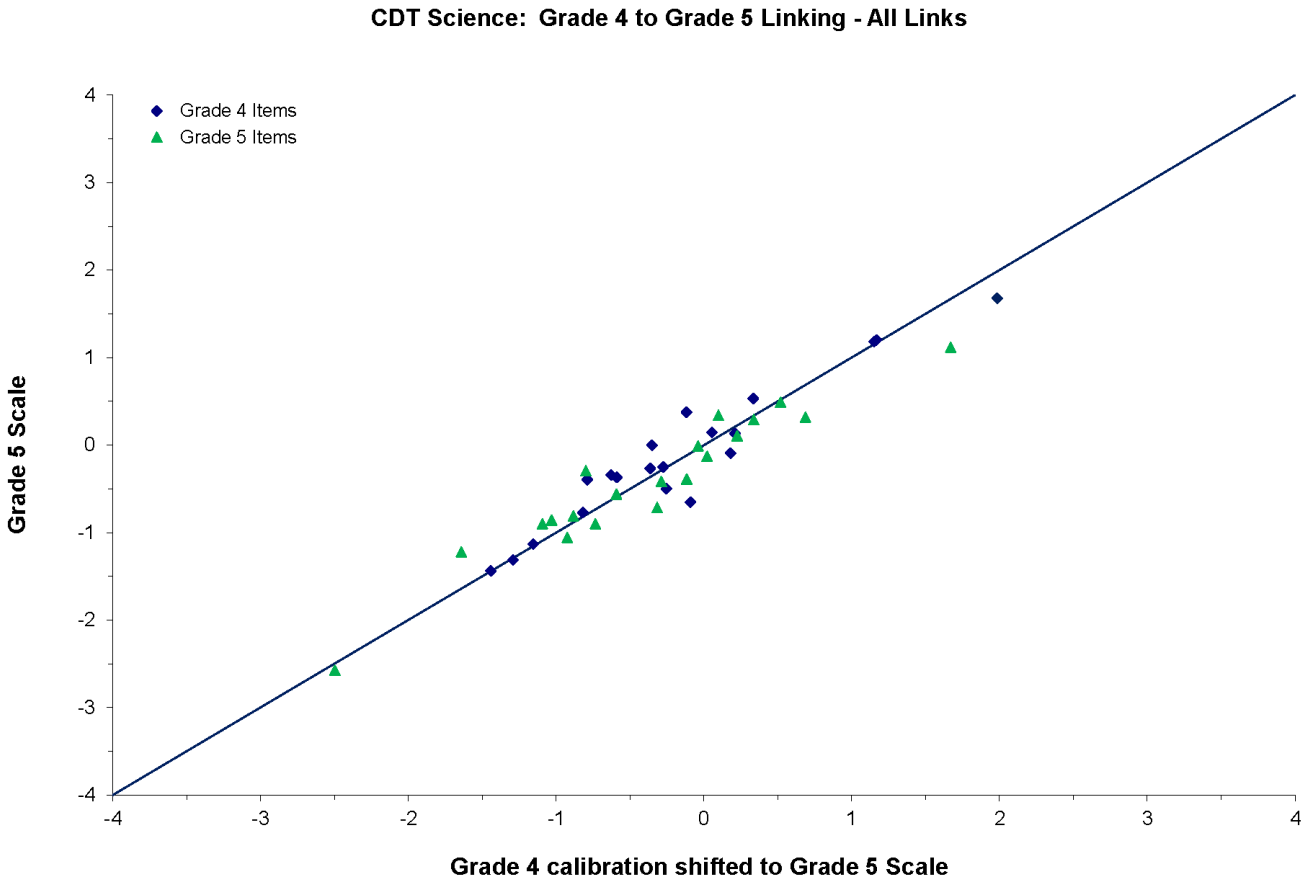
2. Calibrate off-grade items anchoring on the on-grade items. Anchor values come from step 1.
   - Calibrate grade 5 items on grade 4 forms anchoring on item parameters determined in grade 4 calibration in step 1.
   - Calibrate grade 4 items on grade 5 forms anchoring on item parameters determined in grade 5 calibration in step 1.

     Note: For the linking between grades 4 and 5, the calibration of off-grade items on grade 4 forms includes only grade 5 items. It does not include grade 3 items that appeared on grade 4 forms. That is, grade 3 and grade 5 items that appeared on grade 4 forms are not calibrated together.

   For each of the linking items, there are two estimates of item difficulty—one from each of the two calibrations. Correlation between these should be high. If not, vertical linking will be problematic.

3. Calculate the difference between the two estimates of item difficulty from step 2 for each linking item. The average of these differences is the adjacent grade shift parameter.
    - If grade is less than 7, determine the shift parameter needed to place items on upper grade scale.
    - If grade is greater than 7, determine the shift parameter needed to place items on lower grade scale.
    - Calculate the difference in item difficulty estimates between step 2, bullet 1 (grade 4 scale) and step 2, bullet 2 (grade 5 scale). An example of an Excel table used for calculations can be found in Appendix C.
4. Apply the adjacent grade shift parameter and plot the linking items along with a 45° line. Figure 9–1 below is an example. The 45° line is for visual reference only. Outliers are NOT identified by comparing to the line. See step 5 for details.

Figure 9–1. Sample of Linking Items Plot



Plots for all adjacent grade links can be found in Appendix C.

5. Determine if any items should be removed from the vertical linking process. Identify potential outliers using a combination of correlation, ratio of standard deviation, and robust Z. Discuss these items with test development specialists to determine if they should be removed. An item may be removed from the linking process and still remain in the item pool. In this case, the item is not removed from the on-grade calibrations. That is, do not re-run calibrations in step 1. Repeat steps 2 through 4.

6. Calculate the final shift parameter to the base grade (center of scale) by chaining together adjacent grade shift parameters
    - Grade 7 is the base grade. The final shift parameter for grade 4 items is the shift parameter between grades 4 and 5 plus the shift parameter between grades 5 and 6 plus the shift parameter between grades 6 and 7.

7. Apply the final shift parameters in step 6 to the item parameters calibrated in step 1.

VERTICAL LINKING RESULTS

Table 9–9 shows the number of links, correlation, and shift parameter for the both the initial and final vertical linking for each content area. Initial vertical linking includes all items. Final values were determined after some links were dropped after consultation with test development specialists.

Table 9–9. Vertical Linking Summary

Content Area	Link	Number of Links Initial	Number of Links Final	Correlation Initial	Correlation Final	Shift Parameter Initial	Shift Parameter Final
Mathematics	Grade 3 to Grade 4	40	39	0.960	0.964	-1.245	-1.212
Mathematics	Grade 4 to Grade 5	40	40	0.892	0.892	-0.622	-0.622
Mathematics	Grade 5 to Grade 6	50	49	0.914	0.910	-0.416	-0.395
Mathematics	Grade 6 to Grade 7	60	60	0.935	0.935	-0.782	-0.782
Mathematics	Grade 8 to Grade 7	60	60	0.887	0.887	0.301	0.301
Mathematics	Algebra I to Grade 8	60	58	0.933	0.941	0.766	0.808
Mathematics	Algebra II to Algebra I	60	59	0.880	0.905	0.516	0.544
Mathematics	Geometry to Grade 8	60	60	0.907	0.907	1.022	1.022
Reading	Grade 3 to Grade 4	40	40	0.956	0.956	-0.257	-0.257
Reading	Grade 4 to Grade 5	40	38	0.940	0.954	-0.410	-0.348
Reading	Grade 5 to Grade 6	40	39	0.948	0.965	-0.419	-0.389
Reading	Grade 6 to Grade 7	40	37	0.914	0.945	-0.066	-0.092
Reading	Grade 8 to Grade 7	40	40	0.934	0.934	0.352	0.352
Reading	Literature to Grade 8	40	40	0.929	0.929	0.383	0.383
Science	Grade 3 to Grade 4	40	40	0.952	0.952	-0.570	-0.570
Science	Grade 4 to Grade 5	40	40	0.956	0.956	-0.773	-0.773
Science	Grade 5 to Grade 6	40	40	0.968	0.968	-0.211	-0.211
Science	Grade 6 to Grade 7	40	39	0.938	0.945	-0.135	-0.111
Science	Grade 8 to Grade 7	40	40	0.973	0.973	0.140	0.140
Science	Biology to Grade 8	40	38	0.858	0.904	0.815	0.821
Science	Chemistry to Grade 8	40	37	0.882	0.932	1.172	1.136
Writing	Grade 3 to Grade 4	40	40	0.957	0.957	-0.597	-0.597
Writing	Grade 4 to Grade 5	40	40	0.954	0.954	-0.221	-0.221
Writing	Grade 5 to Grade 6	40	40	0.967	0.967	-0.305	-0.305
Writing	Grade 6 to Grade 7	40	40	0.950	0.950	-0.237	-0.237
Writing	Grade 8 to Grade 7	40	40	0.967	0.967	0.221	0.221
Writing	English Composition to Grade 8	40	40	0.961	0.961	0.176	0.176

Recall that for each content area the vertical scale is centered at grade 7. If the item's grade is less than 7, the shift parameter is the value that is added to place the item on the upper grade scale. For example, -1.212 is added to each grade 3 mathematics item's difficulty to place them on the grade 4 scale. The negative sign indicates that grade 3 items are less difficult than grade 4 items. If the item's grade is greater than 7, the shift parameter is the value added to place the item on the lower grade scale. For example, 0.301 is added to each grade 8 mathematics item's difficulty to place them on the grade 7 scale. The positive sign indicates that grade 8 items are more difficult than grade 7 items.

Items dropped from vertical linking are shown in Table 9–10. Linking plots in Appendix C show all linking items with dropped items in red.

Table 9–10. Items Dropped from Vertical Linking

Content Area	Link	Linking Items Removed
Mathematics	Grade 3 to Grade 4	603609 (gr. 4 item)
Mathematics	Grade 4 to Grade 5	None
Mathematics	Grade 5 to Grade 6	602104 (gr. 6 item)
Mathematics	Grade 6 to Grade 7	None
Mathematics	Grade 8 to Grade 7	None
Mathematics	Algebra I to Grade 8	601126 (gr. 8 item) and 602644 (gr. 11 item*)
Mathematics	Algebra II to Algebra I	603086 (Alg II item)
Mathematics	Geometry to Grade 8	None
Reading	Grade 3 to Grade 4	None
Reading	Grade 4 to Grade 5	611272 (gr. 5 item) and 611274 (gr. 5 item)
Reading	Grade 5 to Grade 6	610309 (gr. 6 item)
Reading	Grade 6 to Grade 7	610135 (gr. 6 item), 609022 (gr. 6 item), and 609023 (gr. 6 item)
Reading	Grade 8 to Grade 7	None
Reading	Literature to Grade 8	None
Science	Grade 3 to Grade 4	None
Science	Grade 4 to Grade 5	None
Science	Grade 5 to Grade 6	None
Science	Grade 6 to Grade 7	615238 (gr. 7 item)
Science	Grade 8 to Grade 7	None
Science	Biology to Grade 8	617395 (Bio item) and 617880 (Bio item)
Science	Chemistry to Grade 8	618699 (Chem item), 616511 (Chem item), and 616365 (Chem item)
Writing	Grade 3 to Grade 4	None
Writing	Grade 4 to Grade 5	None
Writing	Grade 5 to Grade 6	None
Writing	Grade 6 to Grade 7	None
Writing	Grade 8 to Grade 7	None
Writing	English Composition to Grade 8	None

*The grade 11 item was embedded on an Algebra I form

The final shift parameters were calculated by summing adjacent grade shift parameters. For example, grade 4 items were placed on the vertical scale by applying the grade 4 to grade 5 shift, the grade 5 to grade 6 shift, and the grade 6 to grade 7 shift. Similarly, Algebra I items were placed on the vertical scale by applying the Algebra I to grade 8 shift and the grade 8 to grade 7 shift. Table 9–11 shows the final shift parameters for each content area.

Table 9–11. Final Vertical Linking Shift Parameters

Content Area	Grade/Course	Shift
Mathematics	Grade 3	-3.011
Mathematics	Grade 4	-1.799
Mathematics	Grade 5	-1.177
Mathematics	Grade 6	-0.782
Mathematics	Grade 7	0.000
Mathematics	Grade 8	0.301
Mathematics	Algebra I	1.109
Mathematics	Geometry	1.323
Mathematics	Algebra II	1.653
Reading	Grade 3	-1.086
Reading	Grade 4	-0.829
Reading	Grade 5	-0.481
Reading	Grade 6	-0.092
Reading	Grade 7	0.000
Reading	Grade 8	0.352
Reading	Literature	0.735
Science	Grade 3	-1.665
Science	Grade 4	-1.095
Science	Grade 5	-0.322
Science	Grade 6	-0.111
Science	Grade 7	0.000
Science	Grade 8	0.140
Science	Biology	0.961
Science	Chemistry	1.276
Writing	Grade 3	-1.360
Writing	Grade 4	-0.763
Writing	Grade 5	-0.542
Writing	Grade 6	-0.237
Writing	Grade 7	0.000
Writing	Grade 8	0.221
Writing	English Composition	0.397

The final vertical linking shift parameters for grade 7 in each content area is zero because it is the base grade. The final vertical linking parameter applied to grade 11 items in mathematics and science is based on the grade or course where the items were field tested. For example, the Algebra I vertical linking constant is applied to grade 11 mathematics items which appeared on Algebra I forms.

BANKED ITEM PARAMETERS FROM STAND-ALONE FIELD TESTS

Table 9–12 provides summary information based on the first stand-alone field-test events which were used to establish the content area vertical scales. The table shows the mean, standard deviation, minimum, and maximum of the item parameter estimates for each grade or course level on the content area vertical scales.

Table 9–12. Summary Statistics for Vertically Scaled Item Parameters from Stand-alone Field Test

Content Area	Grade/Course	Mean	SD	Min	Max
Mathematics	Grade 3	-3.011	1.222	-6.641	0.052
Mathematics	Grade 4	-1.799	1.008	-4.388	0.781
Mathematics	Grade 5	-1.177	1.031	-4.367	1.172
Mathematics	Grade 6	-0.782	1.122	-3.821	2.748
Mathematics	Grade 7	0.000	0.979	-2.385	2.800
Mathematics	Grade 8	0.301	0.939	-2.743	2.985
Mathematics	Grade 11	0.939	1.014	-1.175	3.713
Mathematics	Algebra I	1.109	0.763	-0.888	3.099
Mathematics	Geometry	1.323	0.865	-1.125	3.482
Mathematics	Algebra II	1.653	0.955	-1.377	4.181
Reading	Grade 3	-1.086	1.045	-3.761	1.855
Reading	Grade 4	-0.829	0.944	-3.242	2.177
Reading	Grade 5	-0.481	1.039	-3.201	1.964
Reading	Grade 6	-0.092	1.060	-2.653	3.580
Reading	Grade 7	0.000	1.077	-3.744	3.259
Reading	Grade 8	0.352	1.039	-3.127	3.093
Reading	Literature	0.735	0.929	-2.115	3.313
Science	Grade 3	-1.665	1.302	-5.319	0.813
Science	Grade 4	-1.095	1.145	-4.453	1.663
Science	Grade 5	-0.322	0.948	-2.899	1.683
Science	Grade 6	-0.111	0.971	-2.347	2.546
Science	Grade 7	0.000	0.910	-2.531	2.532
Science	Grade 8	0.140	1.035	-2.654	3.309
Science	Grade 11	0.773	0.892	-2.216	2.377
Science	Biology	0.961	0.867	-1.331	3.731
Science	Chemistry	1.276	0.688	-1.101	3.064
Writing	Grade 3	-1.360	1.196	-4.536	2.958
Writing	Grade 4	-0.763	1.140	-3.608	1.899
Writing	Grade 5	-0.542	1.073	-3.780	2.462
Writing	Grade 6	-0.237	1.052	-2.724	4.390
Writing	Grade 7	0.000	1.132	-2.866	3.593
Writing	Grade 8	0.221	1.120	-3.234	2.883
Writing	English Composition	0.397	1.087	-2.531	3.617

Figures 9–2 through 9–5 show the banked item parameter estimates following the first stand-alone field-test events for each grade or course on the content area vertical scales.

Figure 9–2. Mathematics Item Parameters Estimates from Stand-alone Field Test

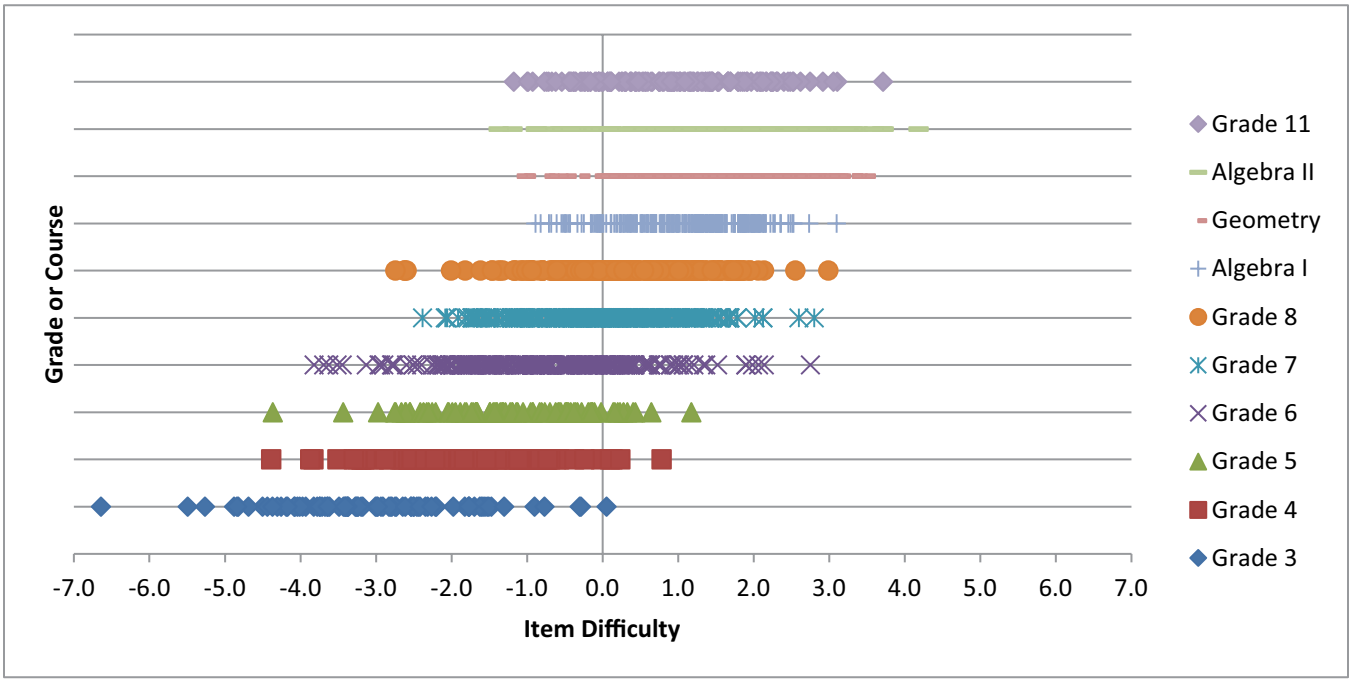


Figure 9–3. Reading Item Parameters Estimates from Stand-alone Field Test

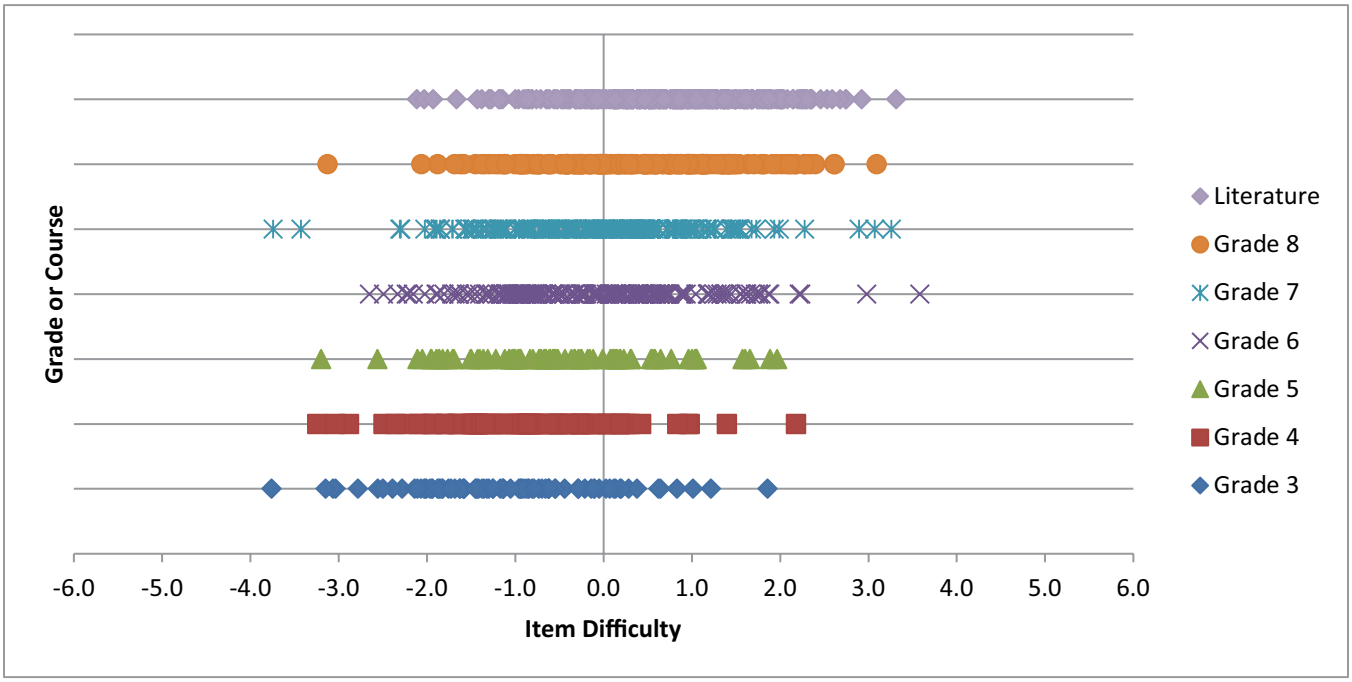


Figure 9–4. Science Item Parameters Estimates from Stand-alone Field Test

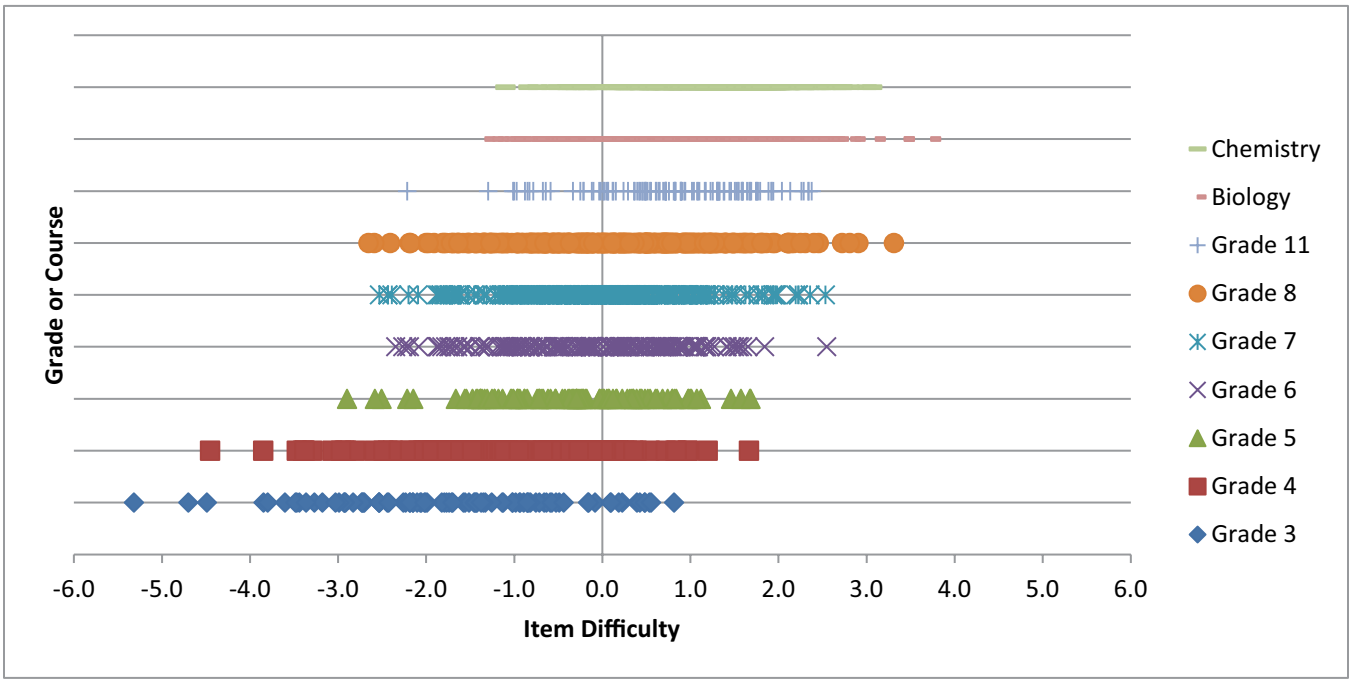
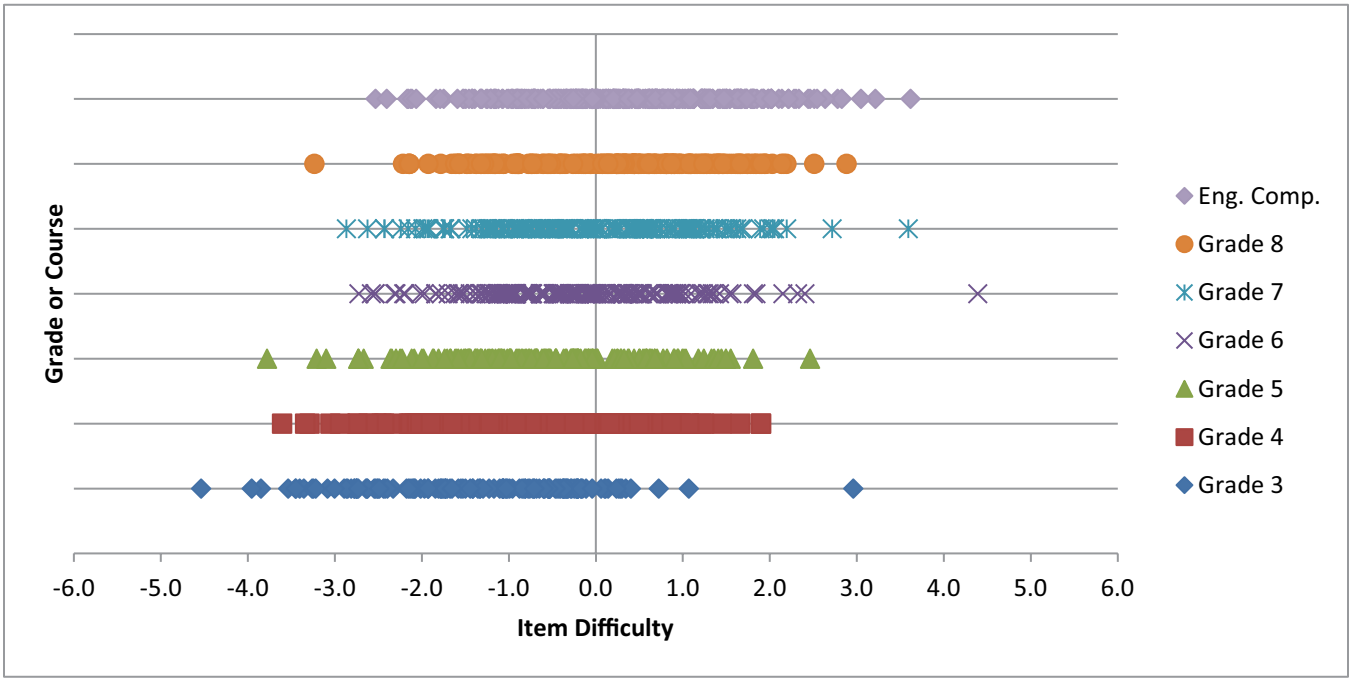


Figure 9–5. Writing Item Parameters Estimates from Stand-alone Field Test



Rasch item difficulty measure on the vertical scale and associated standard error for all items from the initial stand-alone field tests are presented in Appendix B of the 2017–2018 technical report.

BANKED ITEM PARAMETERS FOR THE 2020–2021 OPERATIONAL ITEM POOLS

A number of changes to the CDT item pools have occurred since the initial stand-alone field-test events and creation of the content area vertical scales. For example, there have been embedded field test events to augment the item pools as well as introduce items in kindergarten, grade 1, and grade 2. (See Chapter Six for details on the various field-test events.) Additionally, prior to the 2013–2014 school year CDT items in mathematics, reading, and writing were re-aligned to the new Pennsylvania Core Standards. Table 9–13 provides summary information based on the operational item pools for the 2020–2021 school year. The table shows the mean, standard deviation, minimum, and maximum of the item parameter estimates for each grade or course level on the content area vertical scales.

Table 9–13. Summary Statistics for Vertically Scaled Item Parameters for 2020–2021 School Year

Content Area	Grade/Course	Mean	SD	Min	Max
Mathematics	Kindergarten	-3.914	1.322	-6.433	-0.611
Mathematics	Grade 1	-3.732	1.069	-5.955	-0.610
Mathematics	Grade 2	-2.976	1.346	-5.987	0.402
Mathematics	Grade 3	-1.823	1.246	-5.632	2.158
Mathematics	Grade 4	-1.289	1.239	-6.641	2.748
Mathematics	Grade 5	-0.804	1.038	-3.831	2.139
Mathematics	Grade 6	-0.131	1.124	-3.821	3.389
Mathematics	Grade 7	0.278	0.933	-2.882	2.893
Mathematics	Grade 8	0.589	0.815	-1.662	3.651
Mathematics	Algebra I	0.870	0.794	-1.367	3.264
Mathematics	Geometry	1.193	0.904	-2.058	3.662
Mathematics	Algebra II	1.653	0.916	-1.377	4.181
Reading	Kindergarten	-2.239	1.037	-4.352	0.020
Reading	Grade 1	-1.613	0.995	-4.780	0.831
Reading	Grade 2	-1.148	0.816	-3.869	0.618
Reading	Grade 3	-0.701	0.959	-4.500	1.855
Reading	Grade 4	-0.285	0.975	-3.608	2.464
Reading	Grade 5	0.010	0.884	-3.201	2.101
Reading	Grade 6	0.126	0.917	-2.653	2.578
Reading	Grade 7	0.335	0.909	-3.744	3.259
Reading	Grade 8	0.551	0.916	-3.127	2.799
Reading	Literature	0.825	0.825	-2.115	2.859
Science	Grades K-2 span	-2.265	1.139	-5.446	1.864
Science	Grade 3	-1.691	1.229	-5.319	0.878
Science	Grade 4	-1.095	1.128	-7.111	1.689
Science	Grade 5	-0.512	0.848	-3.108	2.463
Science	Grade 6	-0.237	0.875	-2.723	2.071
Science	Grade 7	-0.094	0.841	-2.531	2.532
Science	Grade 8	0.004	0.921	-2.654	3.309

Table 9–13 (continued). Summary Statistics for Vertically Scaled Item Parameters for 2020–2021 School Year

Content Area	Grade/Course	Mean	SD	Min	Max
Science	Grade 11	0.672	0.944	-2.216	2.391
Science	Biology	0.728	0.805	-1.408	3.731
Science	Chemistry	1.192	0.690	-1.101	3.064
Writing	Kindergarten	-3.121	1.004	-5.685	0.047
Writing	Grade 1	-2.467	1.047	-5.107	0.693
Writing	Grade 2	-1.858	0.878	-4.436	-0.064
Writing	Grade 3	-1.114	1.224	-4.536	2.958
Writing	Grade 4	-0.820	1.177	-4.075	2.137
Writing	Grade 5	-0.663	1.027	-3.780	1.929
Writing	Grade 6	-0.318	0.934	-2.942	3.006
Writing	Grade 7	-0.086	0.862	-2.625	2.194
Writing	Grade 8	0.042	0.926	-3.234	2.192
Writing	English Composition	0.271	0.993	-3.507	3.214

Figures 9–6 through 9–9 show the banked item parameter estimates for the operational item pools for the 2020–2021 school year for each grade or course on the content area vertical scales.

Figure 9–6. Mathematics Item Parameters Estimates for 2020–2021 School Year

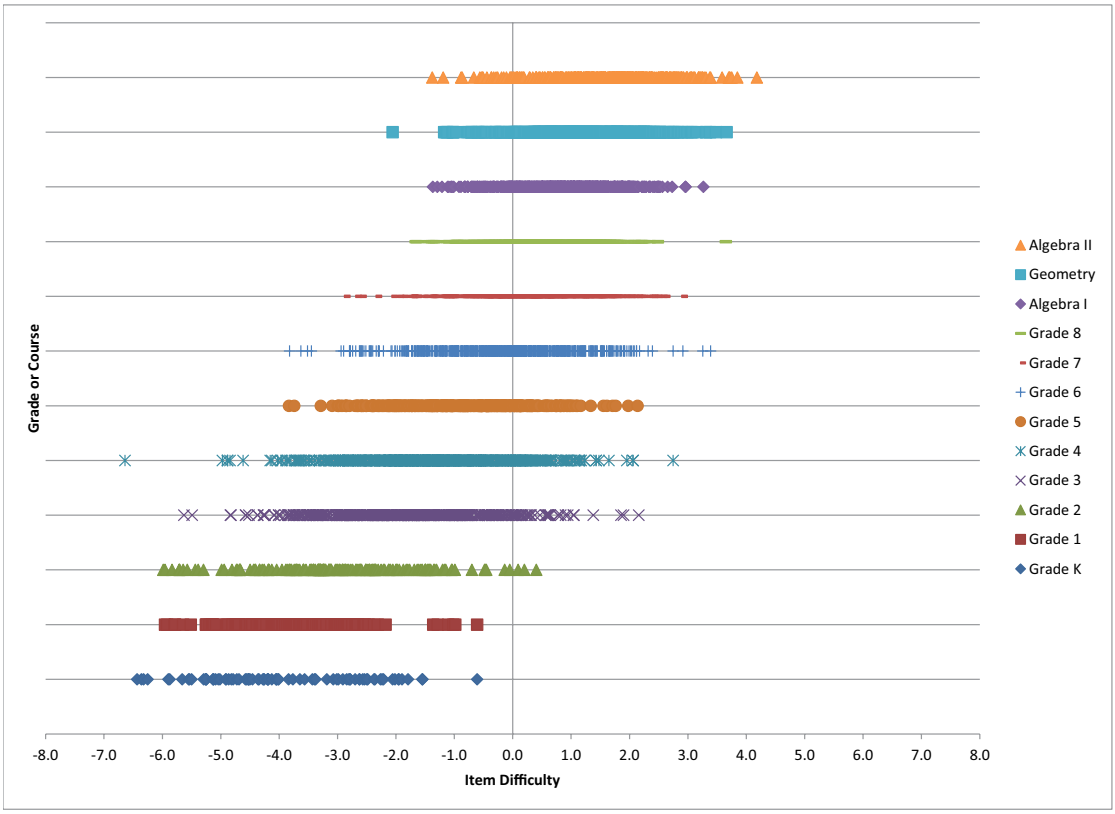


Figure 9–7. Reading Item Parameters Estimates for 2020–2021 School Year

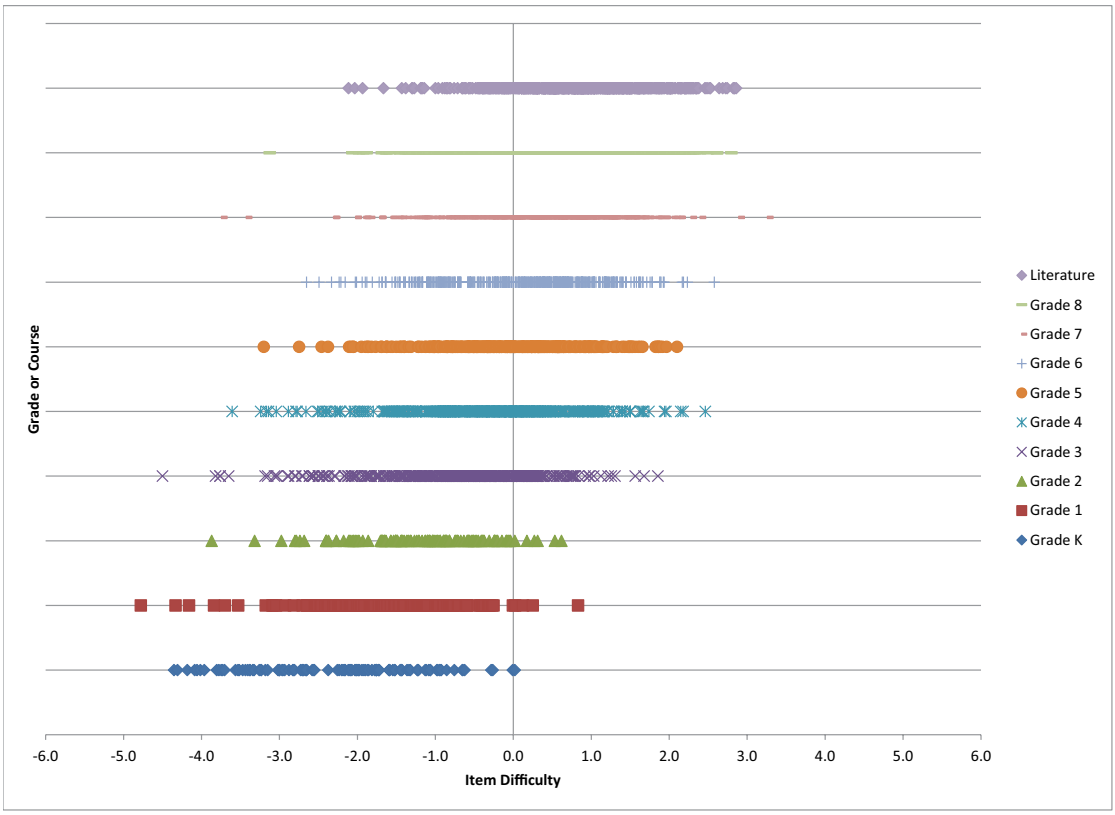


Figure 9–8. Science Item Parameters Estimates for 2020–2021 School Year

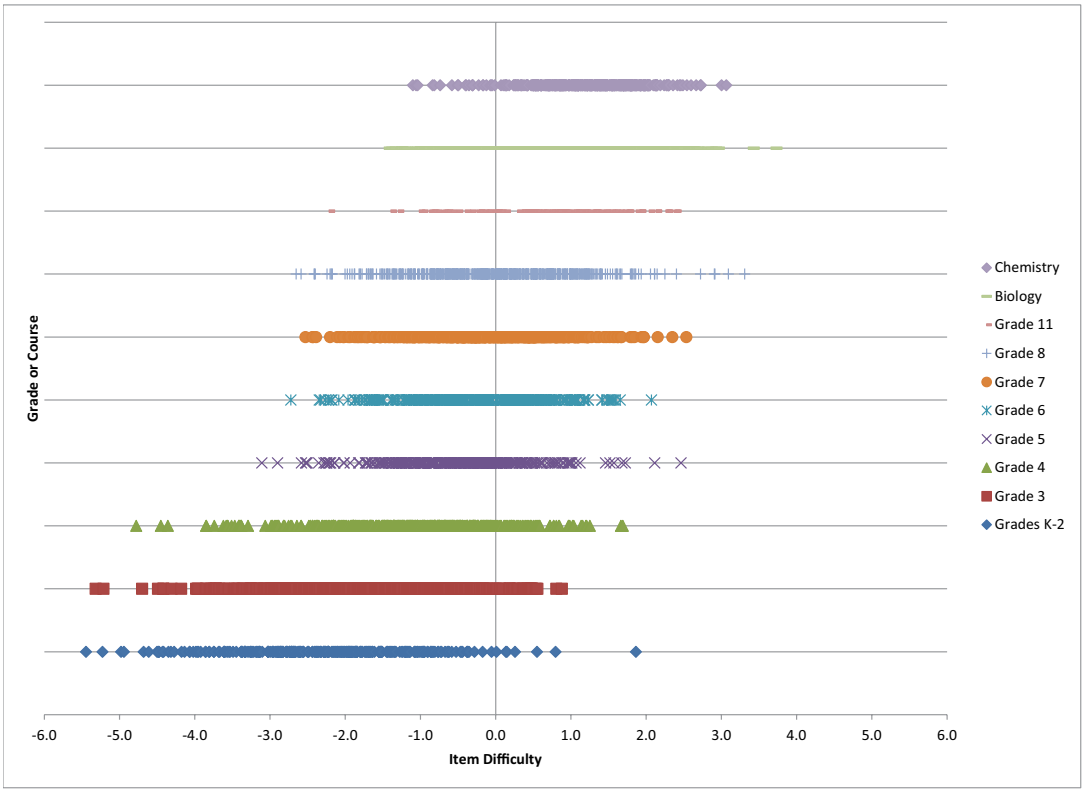
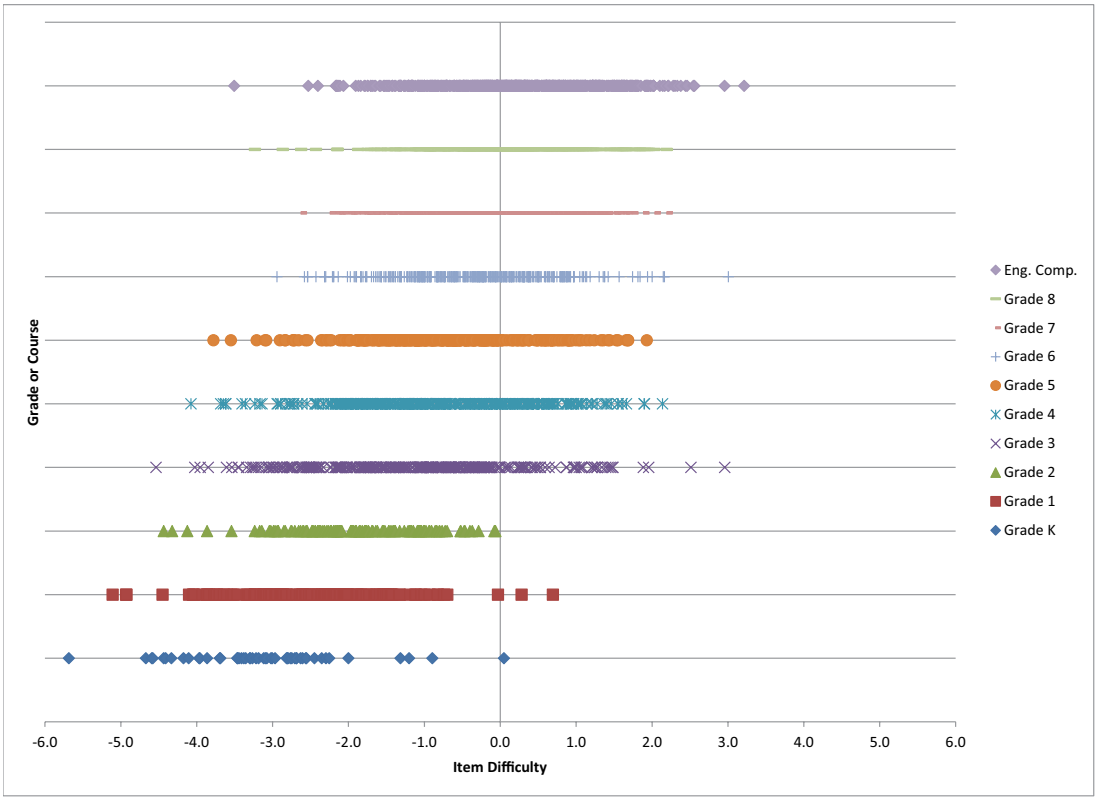


Figure 9–9. Writing Item Parameters Estimates for 2020–2021 School Year



Rasch item difficulty measure on the vertical scale and associated standard error for all operational items are presented in Appendix B of the 2017–2018, 2018–2019, and 2019–2020 technical reports.

CHAPTER TEN: BENCHMARKING

As described in Chapter Fourteen, CDT scores are placed along a continuum from "Areas of Need" to "Strengths to Build On." These are represented in the dynamic reporting suite with colors red, green, and blue. "Areas of Need" are depicted in the red range, while "Strengths to Build On" are depicted in the green and blue ranges. The center of the green range is the point that separates students into two categories: solidly ready for the next grade or course and not solidly ready for the next grade or course. In each content area, the center of the green range for grades 5 and above was established by panels of Pennsylvania educators during benchmarking activities[1].

BENCHMARKING ACTIVITIES

Table 10–1 below presents general information about the preliminary benchmarking activities for mathematics, reading, science, and writing. The cut points established are considered preliminary because they were set prior to the first operational administration of the CDT. This was necessary so teachers and students would have access to immediate scores and reports following operational administration. As operational data become available, preliminary cut points are reevaluated and possibly revised (see Chapter Nineteen for details including the benchmark cuts in place for the 2020–2021 school year).

Table 10–1. General Information about CDT Benchmarking Activities

Category	Information
Event Date	Mathematics: August 12–13, 2010
Event Date	Reading: January 27–28, 2011
Event Date	Science: January 27–28, 2011
Event Date	Writing: August 4–5, 2011
Grades/Courses	Mathematics: Grades 5–8, High School, Algebra I, Geometry, Algebra II
Grades/Courses	Reading: Grades 5–8, Literature
Grades/Courses	Science: Grades 5–8, High School, Biology, Chemistry
Grades/Courses	Writing: Grades 5–8, English Composition
Methodology	Randomly Ordered Item Booklet (ROIB) Angoff (Yes/No) Method
Categories	Not solidly ready for the next grade or course
Categories	Solidly ready for the next grade or course
Number of Panelists	Mathematics: 28
Number of Panelists	Reading: 23
Number of Panelists	Science: 20
Number of Panelists	Writing: 46
Rounds	Two

There were three separate CDT benchmarking events because the operational CDT was rolled out in phases by content area. Each benchmarking event followed the initial stand-alone field-test event for that content area.

When initially launched, the CDT was available to students in grades 6 and above. However, cut points were established for grades 5 and above. This is because CDT is available throughout the school year. Early in the school year it may be more appropriate to evaluate a student's scores based on the prior grade cut. For example, in October, a teacher may choose to evaluate a grade 6 student's scores relative to the grade 5 cut.

[1] The center of the green range for grades 2 through 4 was extrapolated from grades 5 and above prior to the launch of each CDT for students in grades 3 through 5 in spring of 2014. See Chapter Nineteen for details.

The Randomly Ordered Item Booklet (ROIB) Angoff (Yes/No) method was used to set CDT benchmark cut points. Panels of educators worked in grade/course groups to establish cut points for grades 5 through 8, high school, and content area courses Algebra I, Geometry, Algebra II, Literature, Biology, Chemistry, and English Composition. After a training session describing the process and definition of roles, a discussion was held in which panelists were asked to describe what "solidly ready for the next grade or course" means. Thereafter, panelists were asked to review approximately 40 test questions and make individual yes/no judgments as to whether a "solidly ready" student would be successful in answering each question. The judgments were made over two iterations or rounds with a sequence of Round 1 judgments, show and verification of Round 1 results, group discussion, and Round 2 judgments.

After cut points were set for each grade and course within a content area, the vertical articulation of cut points across grades and courses was reviewed. Given that each content area is vertically scaled, it was expected that cut points would increase as grade increased. For example, the grade 8 cut point would not be lower than the grade 7 cut point on the vertical scale. In some cases, post-smoothing was required to ensure increasing cut points across grades/courses and smooth transitions.

Complete descriptions of each benchmarking activity including post-smoothing are available in TAC documents:

- Classroom Diagnostic Tools—Results for Preliminary Benchmarking Activity—Mathematics
- Classroom Diagnostic Tools—Results for Preliminary Benchmarking Activity—Reading and Science
- Classroom Diagnostic Tools—Results for Preliminary Benchmarking Activity—Writing

BENCHMARKING RESULTS

Preliminary cut points in the logit metric for each content area are shown in Figures 10–1 through 10–4. In general, the difference between cut points is greater in the lower grades and then levels off.

Figure 10–1. Preliminary Benchmark Cut Points for Mathematics

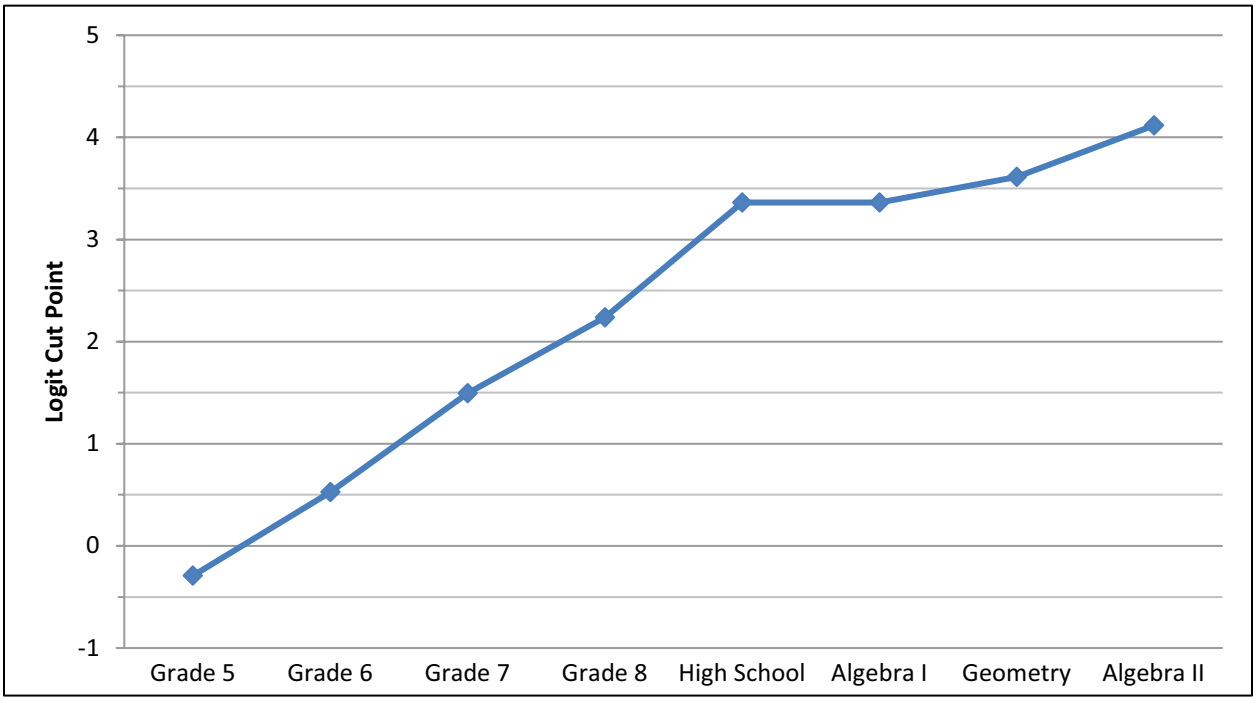


Figure 10–2. Preliminary Benchmark Cut Points for Reading

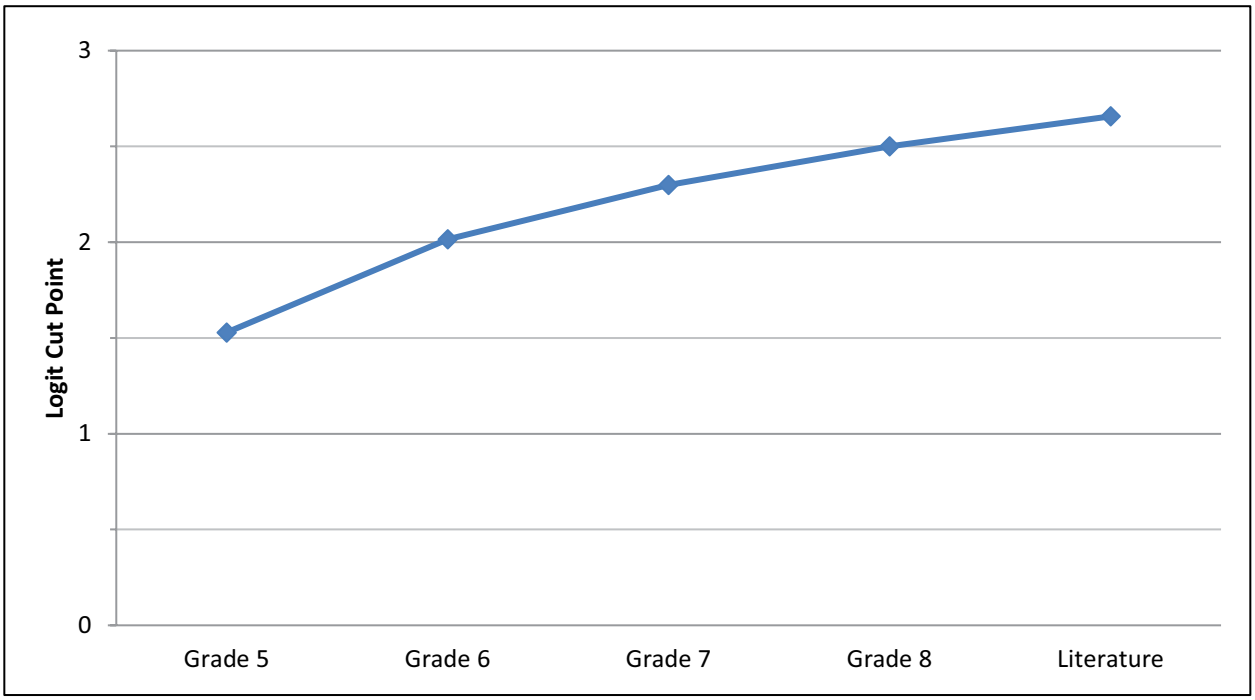


Figure 10–3. Preliminary Benchmark Cut Points for Science

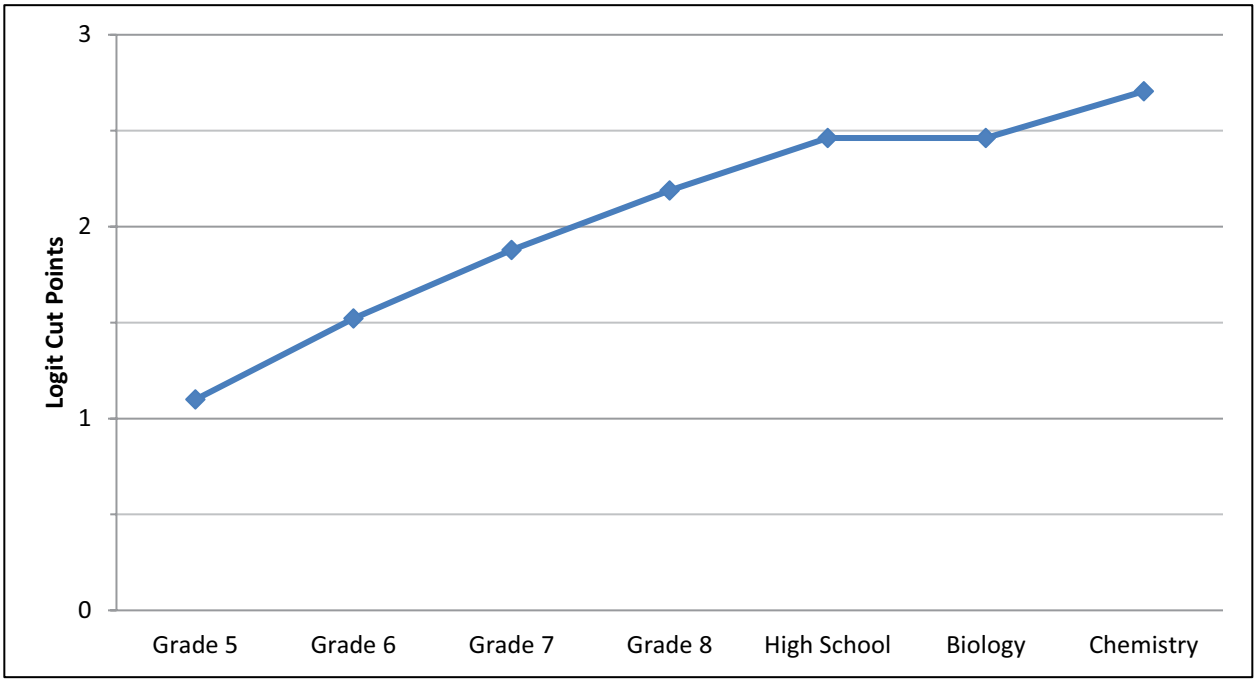


Figure 10–4. Preliminary Benchmark Cut Points for Writing

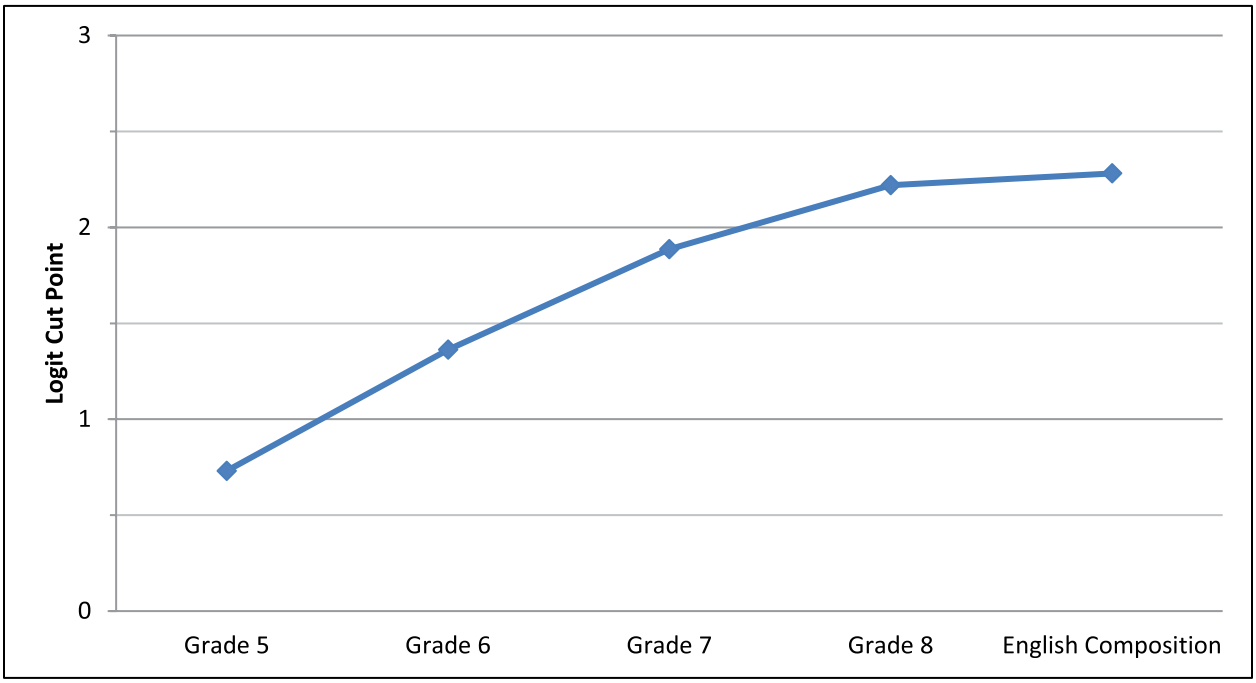


Table 10–2 shows the preliminary benchmark cuts in the logit metric for each content area. Also presented are the scale score ranges for each color on the CDT reports.

Table 10–2. Preliminary Benchmark Cuts and Scale Score Ranges

Content Area	Grade or Course	Logit Cut Point (Center of Green)	Red Scale Score Range	Green Scale Score Range	Blue Scale Score Range
Mathematics	Grade 5	-0.292	400 - 895	896 - 1058	1059 - 2000
Mathematics	Grade 6	0.526	400 - 997	998 - 1160	1161 - 2000
Mathematics	Grade 7	1.495	400 - 1118	1119 - 1281	1282 - 2000
Mathematics	Grade 8	2.238	400 - 1211	1212 - 1374	1375 - 2000
Mathematics	High School	3.363	400 - 1351	1352 - 1514	1515 - 2000
Mathematics	Algebra I	3.363	400 - 1351	1352 - 1514	1515 - 2000
Mathematics	Geometry	3.614	400 - 1383	1384 - 1546	1547 - 2000
Mathematics	Algebra II	4.117	400 - 1446	1447 - 1609	1610 - 2000
Reading	Grade 5	1.529	400 - 982	983 - 1197	1198 - 2000
Reading	Grade 6	2.015	400 - 1051	1052 - 1266	1267 - 2000
Reading	Grade 7	2.299	400 - 1092	1093 - 1307	1308 - 2000
Reading	Grade 8	2.500	400 - 1121	1122 - 1336	1337 - 2000
Reading	Literature	2.657	400 - 1143	1144 - 1358	1359 - 2000
Science	Grade 5	1.099	400 - 1009	1010 - 1182	1183 - 2000
Science	Grade 6	1.522	400 - 1066	1067 - 1239	1240 - 2000
Science	Grade 7	1.879	400 - 1113	1114 - 1286	1287 - 2000
Science	Grade 8	2.189	400 - 1154	1155 - 1327	1328 - 2000
Science	High School	2.462	400 - 1190	1191 - 1363	1364 - 2000
Science	Biology	2.462	400 - 1190	1191 - 1363	1364 - 2000
Science	Chemistry	2.706	400 - 1223	1224 - 1396	1397 - 2000
Writing	Grade 5	0.731	400 - 959	960 - 1132	1133 - 2000
Writing	Grade 6	1.363	400 - 1043	1044 - 1216	1217 - 2000
Writing	Grade 7	1.886	400 - 1113	1114 - 1286	1287 - 2000
Writing	Grade 8	2.219	400 - 1157	1158 - 1330	1331 - 2000
Writing	English Composition	2.281	400 - 1166	1167 - 1339	1340 - 2000

CHAPTER ELEVEN: SCALING

Scaling is used to transform test score values onto a scale that can be interpreted by users easily and correctly. Raw scores cannot be used to compare students' achievement on the CDT because they depend on the difficulty of the test items administered. Given the adaptive nature of the CDT, each student receives test items targeted at his or her level of achievement. Therefore, two students may have taken very different sets of items in terms of difficulty but have the same raw score. This makes use of raw scores for comparison across students, across administrations, or to a specific standard (cut point) meaningless. Rasch ability estimates in the logit metric do take into consideration the difficulty of the items administered. Therefore, they may be used to make comparisons. However, scale scores are introduced to report CDT results since scale scores may be easier to understand and interpret than logits.

Essentially, CDT scale scores are derived through a two-step process. First, there is a nonlinear transformation that converts an individual raw score on a unique set of items to Rasch ability (in logits). Second, a linear transformation is used to convert logits to scale scores. These and some additional considerations (e.g., rounding rules) are discussed in more detail below.

RAW SCORES TO RASCH ABILITY ESTIMATES

For each CDT test, the calibrated item difficulties associated with the unique set of items administered were used to obtain Rasch person ability estimates and asymptotic standard errors of measurement for the overall test, as well as each diagnostic category. Calibrated item difficulties were based on the field tests and vertical linking (further discussed in Chapter Eight and Chapter Nine).

Raw scores (total and diagnostic category) on the unique set of items that makes up an individual CDT test were mapped to Rasch ability estimates using unconditional, joint-maximum likelihood estimation. In the case of zero or perfect raw scores, a fractional raw score (a value less than one) was added to zero scores and subtracted from perfect scores to determine the corresponding logit values for these extreme scores. The Rasch ability estimates were then transformed to scale scores as discussed in the next section.

RASCH ABILITY ESTIMATES TO SCALE SCORES

Generally, scale 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 scale scores, all reported values can become positive integers, which makes more sense to teachers, parents, and students. Since Rasch ability estimates are comparative, the transformed scale scores have a common scale across administrations.

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

- Avoid scales that might be confused with scores for other types of assessment; for example:
    - Scale scores ranging from 0 to 100 (because this might be confused with percent correct scores or percentile ranks)
    - Scale scores ranging from 200 to 800 (because this might be confused with SAT scores)
    - Scale scores with similar ranges as the ones for the Pennsylvania System of School Assessment (PSSA) or Keystone Exams
- Avoid scales similar to raw scores.
- Avoid scales that might suggest the scores are more precise than they actually are (in other words, suggesting more precision than can be supported by the test scores).
- Avoid scales with negative numbers and decimals.

In terms of industry standard practice, a common perspective is that scale scores should facilitate score interpretation while at the same time minimize misinterpretation and unwarranted inferences. Often this is done by incorporating some kind of meaning to the scores[1] (Peterson, Kolen, and Hoover, 1989). The incorporation of content meaning is one way to facilitate score interpretation. This might be done in several different ways. For example, 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 the criterion-reference test like the PSSA.

For CDT, the scale must be sufficiently large to cover the entire vertical scale. As a result, an initial scale score range of 400 to 2000 was established for each content area. When CDT was expanded in spring of 2014 and made available to students in grades 3 through 5, the scale score range was expanded to 200 to 2000 for those students. Initially, the grade 7 benchmark logit cut point was mapped to a scale score of 1200 for all content areas. It is worth noting that, although careful consideration was given to the selection of these values, they are completely arbitrary. For example, the label of 1200 could have been called 100 or any other value 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 scale scores for the CDT for each content area 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 each CDT content area was derived by anchoring the grade 7 benchmark cut (i.e., Rasch ability estimate) to the scale score 1200 and a Rasch ability estimate of 7.9 to the scale score of 2000. The slopes of the scaling equations influence the variability of the scale scores. It is important that the slopes are sufficiently large to cover the full range of the vertical scale. The CDT scaling equations produce scale score distributions with standard deviations of approximately 150 scale score points and cover logit ranges of approximately -6.5 to 7.9. The final slopes and intercepts for deriving scale scores for the CDT are provided in Table 11–1.

Table 11–1. Scaling Constants by Content Area

Content Area	Slope	Intercept
Mathematics	124.90	1013.30
Reading	142.83	871.63
Science	132.87	950.34
Writing	133.02	949.12

ROUNDING

The linearly transformed scale scores are 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.

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

LOWEST OBTAINABLE SCALE SCORES

Each general content area CDT (mathematics, reading, science, and writing) has a lowest obtainable scale score (LOSS) of 200. Course specific CDTs (Algebra I, Geometry, Algebra II, Biology, and Chemistry) have a lowest obtainable scale score (LOSS) of 400. Any derived scale score less than LOSS is truncated to this minimum value. The selection of a LOSS is mainly based on two considerations:

1. Extremely low scale scores may have an impact on the average of the scale scores if CDT data is summarized at school, district, or state level.
2. Score truncation makes sense from a score precision perspective given measurement errors at the extremes are large.

HIGHEST OBTAINABLE SCALE SCORES

A highest obtainable scale score (HOSS), 2000, is set for the CDT for the same reasons as described for the LOSS value.

CHAPTER TWELVE: 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 the case of the CDT, the adaptive nature of the test means that each student takes a unique test form with items targeted at his or her level of achievement.

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, the model can be classified as either classical test theory based equating model or modern test theory (e.g., Rasch model or item response theory) based equating model. In terms of when the equating is conducted in the assessment cycle, the model can be classified as pre-equating or post-equating.

Given the requirements of adaptive testing and immediate score reporting, CDT is pre-equated. Also, it was based on the Rasch model. The following sections will focus on the discussion of pre-equating and the equating design for the CDT.

PRE-EQUATING VERSUS POST-EQUATING

Like other Pennsylvania assessment programs, the CDT uses the Rasch model to guide test design, calibration, scaling, and equating. The key element of equating test forms using the Rasch model is to place the item parameters on the same scale. Once this is done, raw scores can be converted to Rasch ability estimates and then to scale scores as described in Chapter Eleven. As a result, the scale scores can be compared across forms and administrations with different items.

A common practice in many K–12 large-scale assessment programs is to have all the items field tested before they are administered in an operational setting. 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 of them has its own advantages and disadvantages. The use of pre-equating can facilitate the operational process in terms of adaptive item selection, rapid or immediate score reporting, and more flexibility in the assessment. 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 stand-alone field tests, which may make the items appear harder in the field test than in the operational test (Eignor, 1985; Eignor and Stocking, 1986; Stocking and Eignor, 1986; Kolen and 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 because students cannot distinguish between operational and field-test items. Also, post-equating is sometimes considered to yield more accurate analysis results given the large number of students who take the operational tests. On the other hand, post-equating does not allow for adaptive item selection or immediate score reporting as required of the CDT.

EQUATING DESIGN FOR THE CDT

The CDT is an adaptive test, meaning that the test items selected are tailored to each student's achievement as the test progresses. This requires that all items in the pool be on the same scale and known at the time of testing. For CDT, this is accomplished by vertical linking the entire item pool within a content area based on the field-test events. See Chapter Eight and Chapter Nine for details. The known (pre-equated) item parameters are used in selecting items targeted for the student and to provide immediate scores to teachers and students.

In implementing the pre-equating model for the CDT, efforts were made to enhance the accuracy of pre-equating results. To address the concerns on students' motivation to take field tests, records were excluded from item calibrations if the student did not answer at least 5 questions. Also, records with high person outfit mean-squares values were excluded following the WINSTEPS suggestion that these may be the result of a few random responses by low performers. To address concerns of sample sizes, windows for field testing were scheduled so they did not overlap other testing in an attempt to increase volunteer participation. Also, field-test windows were extended in cases where schools were unable to complete testing in the allotted time. A small study of mathematics vertical linking items revealed no position effects. However, it should be noted that with adaptive tests students do not take the same items. Even if two students do take the same item, it will likely not be in the same test position.

EVALUATION OF ITEM PARAMETER STABILITY

After each school year, item parameter stability studies are conducted for each content area. If the differences between the newly estimated Rasch item difficulties and the estimates based on the field-test events are not statistically significant, the pre-equating results should be valid. See Chapter Eighteen for results of item parameter stability studies based on operational data from the 2020–2021 school year.

EQUATING ADDITIONAL FIELD-TEST ITEMS

Over time, additional items have been, and will continue to be, needed to replenish the CDT item pools. Plans to field test additional items must include an equating plan. Equating is needed to place the new items onto the existing vertical scale. In the case of stand-alone field-test events, common-item equating was used. That is, field-test forms included items from the current CDT item pool. In the case of embedded field-test events, field-test items were included within an operational administration such that students did not know which items were field test. With both stand-alone and embedded field test, 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. For each content area, the entire item pool, including field-test items, was calibrated using WINSTEPS with operational items anchored on the banked values.

CHAPTER THIRTEEN: OPERATIONAL TEST DESIGN AND CAT CONFIGURATIONS

The Pennsylvania Classroom Diagnostic Tools (CDT) was initially developed to support teachers and students in grades 6 through 12. In spring 2014, CDT was made available to students in grades 3 through 5 as well. The tools are fully integrated and aligned in the Standards Aligned System (SAS) and enable educators to identify students' academic strengths and areas of need as well as provide links to classroom resources. The assessment is voluntary and administered completely online using a computer adaptive test (CAT) model.

The CDT features a number of tests. Tests in Mathematics, Algebra I, Geometry, and Algebra II were introduced in October 2010 for students in grades 6 and above. Tests in Reading/Literature, Science, Biology, and Chemistry were first available in April 2011 for students in grades 6 and above. Tests in Writing /English Composition began in October 2011 for students in grades 6 and above. Tests in Mathematics, Reading, Science, and Writing for students in grades 3 through 5 started in April 2014.

This chapter details the operational CDT test design and configuration of the CAT algorithm. Test design elements include the number of diagnostic categories, the number of operational items to administer per diagnostic category, and the number of embedded field-test items. CAT algorithm elements include entry point, item selection criteria, test navigation, and termination.

OPERATIONAL TEST DESIGN

NUMBER OF DIAGNOSTIC CATEGORIES

The CDT tests include multiple-choice (MC) and evidence-based selected-response (EBSR) items. All items in the content areas of mathematics, reading, and writing are aligned to the Pennsylvania Core Standards. All items in the content area of science are aligned to the Pennsylvania Academic Standards. Each CDT is broken into four or five diagnostic categories and the items in the pool are grouped by these diagnostic categories based on the Assessment Anchors and Eligible Content. The diagnostic categories for each of the CDT tests are listed below.

Math Grades 3–5 and Math Grades 6–HS

- Numbers & Operations
- Algebraic Concepts
- Geometry
- Measurement, Data, and Probability

Algebra I

- Operations with Real Numbers and Expressions
- Linear Equations & Inequalities
- Functions & Coordinate Geometry
- Data Analysis

Geometry

- Geometric Properties
- Congruence, Similarity, & Proofs
- Coordinate Geometry & Right Triangles
- Measurement

Algebra II

- Operations with Complex Numbers
- Non-Linear Expressions & Equations
- Functions
- Data Analysis

Reading Grades 3–5 and Reading/Lit Grades 6–HS

- Key Ideas and Details—Literature Text
- Key Ideas and Details—Informational Text
- Craft and Structure/Integration of Knowledge and Ideas—Literature Text
- Craft and Structure/Integration of Knowledge and Ideas—Informational Text
- Vocabulary Acquisition and Use

Science Grades 3–5 and Science Grades 6–HS

- The Nature of Science
- Biological Sciences
- Physical Sciences
- Earth/Space Sciences

Biology

- Basic Biological Principles/Chemical Basis for Life
- Bioenergetics/Homeostasis & Transport
- Cell Growth & Reproduction/Genetics
- Theory of Evolution/Ecology

Chemistry

- Properties & Classification of Matter
- Atomic Structure & The Periodic Table
- The Mole & Chemical Bonding
- Chemical Relationships & Reactions

Writing Grades 3–5 and Writing/Eng Comp Grades 6–HS

- Quality of Writing: Focus and Organization
- Quality of Writing: Content and Style
- Quality of Writing: Editing
- Conventions: Punctuation, Capitalization, and Spelling
- Conventions: Grammar and Sentence Formation

NUMBER OF ITEMS PER DIAGNOSTIC CATEGORY

There were various factors considered when determining the number of operational items to administer per diagnostic category. The goal of the CDT is to provide diagnostic information. Therefore, the test must include a sufficient number of items to provide meaningful scores with low standard errors. However, testing time is limited and the item pools are finite. A very long test may produce lower standard errors, but if it is considered to be "too long" will teachers use it? Also, the longer the test, the more the items are exposed.

Prior to the launch of the first operational CDT in fall of 2010, simulations were run of various test lengths. Table 13–1 shows the average conditional standard error of measurement (CSEM) for total test and each diagnostic category[1] (DC) for five test lengths in simulations of CDT Mathematics. Also included is the theoretical minimum standard error that is possible for each test length. This is the standard error if the ability is known and there are sufficient items to administer where the item's difficulty is equal to the known ability and the test constraints are met.

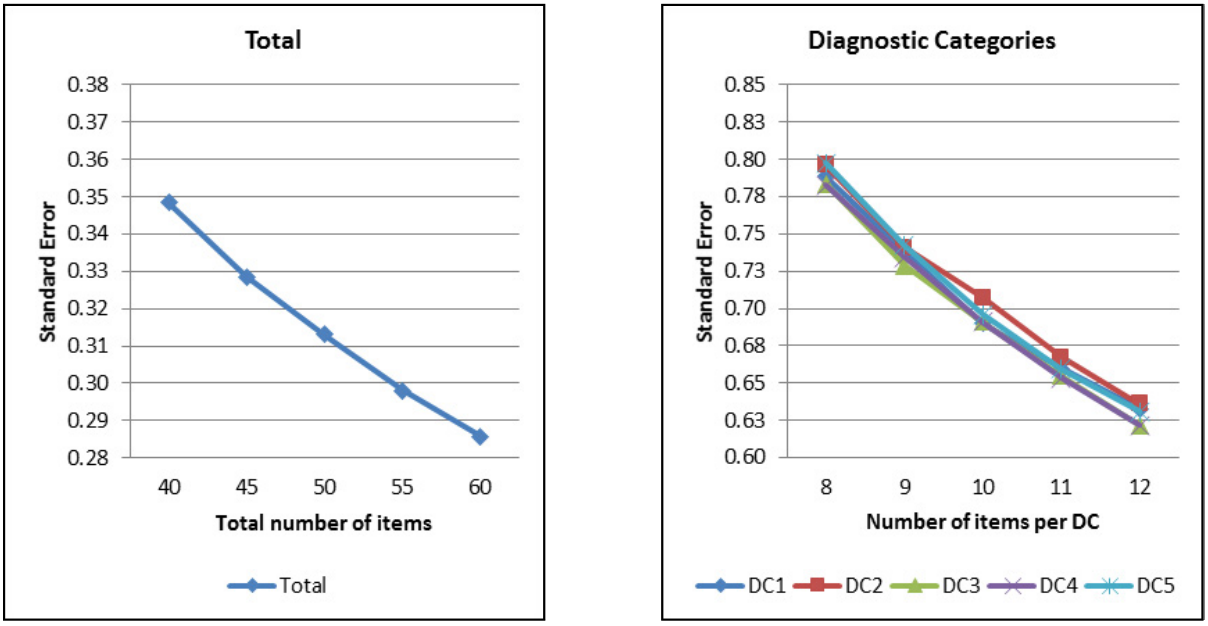
Table 13–1. Average Standard Errors for Various Test Lengths — Mathematics

Total Number of Points	Total Min Error	Total Avg Error	Diagnostic Categories Number of Points	Diagnostic Categories Min Error	Diagnostic Categories DC1 Avg Error	Diagnostic Categories DC2 Avg Error	Diagnostic Categories DC3 Avg Error	Diagnostic Categories DC4 Avg Error	Diagnostic Categories DC5 Avg Error
40	0.316	0.348	8	0.707	0.789	0.796	0.784	0.783	0.798
45	0.298	0.329	9	0.667	0.738	0.741	0.729	0.734	0.742
50	0.283	0.313	10	0.632	0.690	0.707	0.691	0.691	0.696
55	0.270	0.298	11	0.603	0.660	0.667	0.655	0.653	0.659
60	0.258	0.286	12	0.577	0.633	0.636	0.622	0.622	0.631

As expected, increasing the number of items decreases the standard error. Differences in standard errors at the diagnostic category level for the same number of items are a reflection of differences in the diagnostic category item pools.

Figure 13–1 shows average standard errors as a function of test length.

Figure 13–1. Average Standard Errors for Various Test Lengths – Mathematics



[1] At that time, there were five diagnostic categories in CDT Mathematics.

Considering test time factors and simulation results for various test lengths, it was determined that CDT tests with four diagnostic categories would have 12–15 items per category (48–60 items total) and CDT tests with five diagnostic categories would have 10–12 items per category (50–60 items total).

NUMBER OF EMBEDDED FIELD-TEST ITEMS

Over time, additional items will be needed to replenish the CDT item pools. Embedding field-test items within an operational CDT test is advantageous for two reasons. First, sufficient item level data can be gathered without the time and expense of a separate stand-alone administration. Second, it allows the new items to be placed on the existing operational scale. See Chapter Twelve for details.

As detailed in Chapter Six, there have been five embedded field-test events. Starting on February 14, 2013, field-test items were embedded within CDT Mathematics and Reading/Literature tests. Starting on August 26, 2013, items were embedded within CDT Mathematics, Reading/Literature, Science, and Writing/English Composition tests for students in grade 6. Starting on August 24, 2015, items were embedded within seven of the thirteen CDTs: Math Grades 6–HS, Algebra I, Reading Grades 3–5, Reading/Lit Grades 6–HS, Science Grades 6–HS, Biology, and Writing/Eng Comp Grades 6–HS. Starting on August 20, 2018, items were embedded within all thirteen of the CDTs. Starting on August 19, 2019, items were embedded within all CDTs in the science content area except Chemistry.

For each embedded field-test event, the factors considered when determining the number of field-test items to embed included the number of items to be field tested, the expected number of students testing, and the desired n-count per item for field-test analyses. In mathematics, science, and writing, field-test items were randomly assigned to fixed positions spread throughout the operational test. In reading, a field-test passage was randomly assigned near the middle of the test and students took all of the items associated with the passage. In all content areas, the positions of field-test items were unknown to students. Field-test items were not clustered at the end of the test in an effort to avoid any fatigue effect when placing the items on the operational scale.

CAT ALGORITHM

This section covers elements of the CAT algorithm including entry point, item selection criteria, test navigation, and termination.

ENTRY POINT

All CDT tests other than Reading Grades 3–5 and Reading/Lit Grades 6–HS begin with a small “locator” section in which one or two items per diagnostic category are administered. The order of the diagnostic categories is random. The two CDT tests in the reading content area are slightly different because they are passage-based. Those, too, have a small “locator” section, but they may not contain one or two items for each diagnostic category because not all passages have an item for each diagnostic category.

The CAT algorithm is designed to administer items targeted for the individual student based on performance. However, student performance in the current test setting is not known at the beginning of the test. With no prior information about a student, the starting point in each diagnostic category is an item of average difficulty. For CDT tests that are not course-specific (Math Grades 3–5, Math Grades 6–HS, Reading Grades 3–5, Reading/Lit Grades 6–HS, Science Grades 3–5, Science Grades 6–HS, Writing Grades 3–5, and Writing/Eng Comp Grades 6–HS), the student’s grade is considered in selecting an item of average difficulty. For example, a grade 7 student taking CDT Math Grades 6–HS will start with an item near the average difficulty of grade 7 items in the pool. For CDT tests that are course-specific (Algebra I, Geometry, Algebra II, Biology, and Chemistry), an average item will be selected regardless of the student’s grade. For example, a grade 7 student taking CDT Algebra I will start with an item near the average difficulty of Algebra I items in the pool.

If a student has previously taken the CDT, the prior CDT scores are used to give the CAT algorithm a “head start.” In this case, the first item in each diagnostic category is selected to match the characteristics of the prior information rather than an average item. For example, if a student previously took the CDT Math Grades 6–HS test and scored very high in “Measurement, Data, and Probability,” then the first item selected in that diagnostic category will be more difficult than the grade level average.

The CAT algorithm includes a randomization component when selecting items to control item exposure. That is, one item is selected from among a set of items that are near the targeted item difficulty. This is especially important at the beginning of the CDT when no prior information is available. Randomization of items and diagnostic categories ensure that students will not see the same set of items in the same order even when all of the students are assigned items of average difficulty.

ITEM SELECTION CRITERIA

Once the initial set of items has been administered, the CAT algorithm is designed to administer items targeted for the individual student based on performance. In targeting items, the CAT algorithm uses Rasch ability estimates from the current test session and considers a number of factors including test blueprint, response probability, item pool refinement, and passage-related concerns. Each of these is discussed in detail on the following pages.

RASCH ABILITY ESTIMATES

As described in Chapter Eight and Chapter Nine, CDT item pools are scaled using the Rasch partial credit model (Wright & Masters, 1982) and vertically linked across grades and courses. The CAT algorithm has access to all item parameters in the item pool. After each item response, Rasch ability estimates and standard errors are calculated via maximum likelihood estimation (MLE) for the total test and each diagnostic category. In the case of zero (all items incorrect) and perfect (all items correct) scores, a correction factor is applied before computing the relevant maximum likelihood estimates. A fractional value is added to a zero score and subtracted from a perfect score before estimation.

After the locator section of the CDT, but before a student has taken many items in each diagnostic category, the total Rasch ability estimate is used in item selection. This is because total and diagnostic category ability estimates tend to be highly correlated, and the total estimate does not change as dramatically as diagnostic category estimates given one additional item. Using the total estimate at this point prevents students from experiencing extreme fluctuations in the difficulty of items.

While use of the total Rasch ability estimate makes sense early in the test, the goal of the CDT is to be diagnostic, and some students exhibit clear strengths and areas of need in different diagnostic categories. Therefore, after four or five items have been administered in a diagnostic category, the corresponding Rasch ability estimate for that diagnostic category is used in item selection. This ensures, for example, that a student struggling in “Biological Sciences” while at the same time excelling in “Earth and Space Sciences” will be administered easier “Biological Sciences” items and more challenging “Earth and Space Sciences” items.

TEST BLUEPRINT

The CAT algorithm closely resembles a modified constrained CAT (MCCAT) design (Leung, Chang, & Hau, 2003). The general idea is that the CAT algorithm is configured with upper and lower bounds that specify the minimum and maximum numbers of items that will be administered to students for both total and diagnostic categories.

RESPONSE PROBABILITY

No matter which Rasch ability estimate is used in selecting an item, total or diagnostic category estimate, the CAT algorithm targets items where the student has response probability (RP) of answering correctly, based on the Rasch ability estimate and item’s difficulty. The most efficient way to run a CAT is to select items where RP is 0.5. That is, select items where the student has a 50% chance of getting the item correct. This response probability produces the smallest standard error for any given number of items.

Prior to the launch of the first operational CDT in fall of 2010, simulations were run for various response probabilities. Table 13–2 shows the average person standard errors for total test and each diagnostic category[2] for seven response probabilities in simulations of CDT Mathematics with 50 items. Figure 13–2 shows average standard errors as a function of response probability.

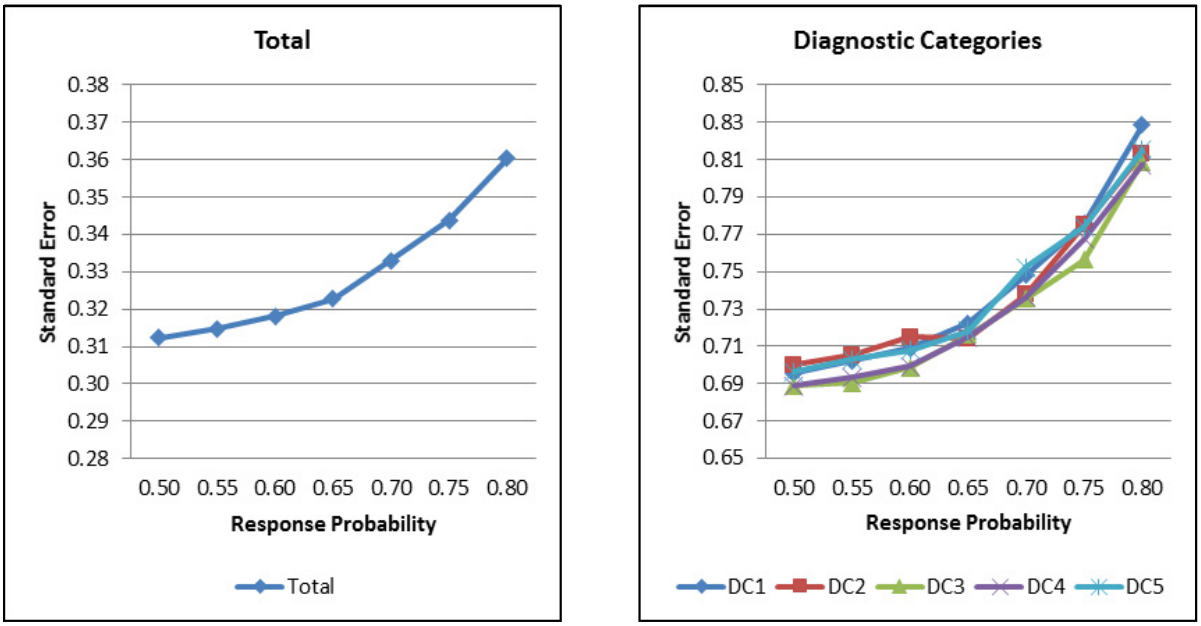
[2] At that time, there were five diagnostic categories in CDT Mathematics.

Table 13–2. Average Standard Errors for Various Response Probabilities — Mathematics

Number of Items	Response Probability	Total	DC 1	DC 2	DC 3	DC 4	DC 5
50 total (10 per DC)	0.50	0.312	0.696	0.700	0.689	0.689	0.696
50 total (10 per DC)	0.55	0.315	0.702	0.705	0.690	0.693	0.703
50 total (10 per DC)	0.60	0.318	0.709	0.715	0.699	0.699	0.708
50 total (10 per DC)	0.65	0.323	0.722	0.714	0.716	0.715	0.719
50 total (10 per DC)	0.70	0.333	0.748	0.738	0.735	0.736	0.752
50 total (10 per DC)	0.75	0.344	0.776	0.775	0.756	0.767	0.774
50 total (10 per DC)	0.80	0.360	0.829	0.813	0.809	0.807	0.815

As expected, increasing the response probability increases the standard error. Differences in standard errors at the diagnostic category level for the same response probability are a reflection of differences in the diagnostic category item pools.

Figure 13–2. Average Standard Errors for Various Response Probabilities — Mathematics



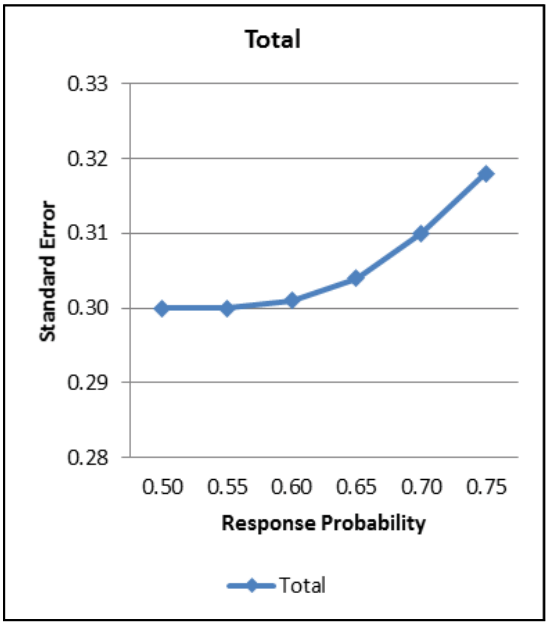
As can be seen in Figure 13–2, increasing response probability incrementally from 0.50 leads to increases in standard error. The increase in standard error is gradual at first and becomes more pronounced around 0.65.

Prior to the launch of the CDT for students in grades 3 through 5, the topic of response probability was revisited for each content area. Simulations for various response probabilities were run with fixed length tests equal to average test length. Results for each content area are presented in Tables 13–3 through 13–6 and Figures 13–3 through 13–6.

Table 13–3. Average Standard Errors for Various Response Probabilities — Mathematics

Number of Items	Response Probability	Total	DC 1	DC 2	DC 3	DC 4
52 total (13 per DC)	0.50	0.300	0.602	0.592	0.601	0.606
52 total (13 per DC)	0.55	0.300	0.602	0.594	0.602	0.607
52 total (13 per DC)	0.60	0.301	0.605	0.597	0.604	0.610
52 total (13 per DC)	0.65	0.304	0.613	0.608	0.613	0.619
52 total (13 per DC)	0.70	0.310	0.626	0.622	0.625	0.631
52 total (13 per DC)	0.75	0.318	0.646	0.644	0.645	0.651

Figure 13–3. Average Standard Errors for Various Response Probabilities — Mathematics



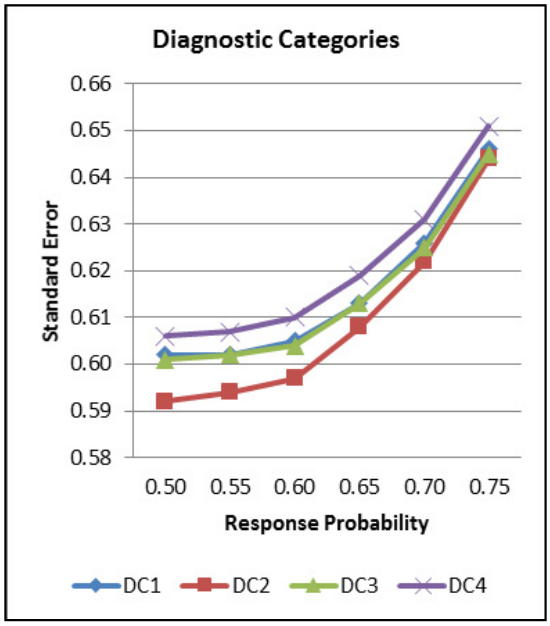
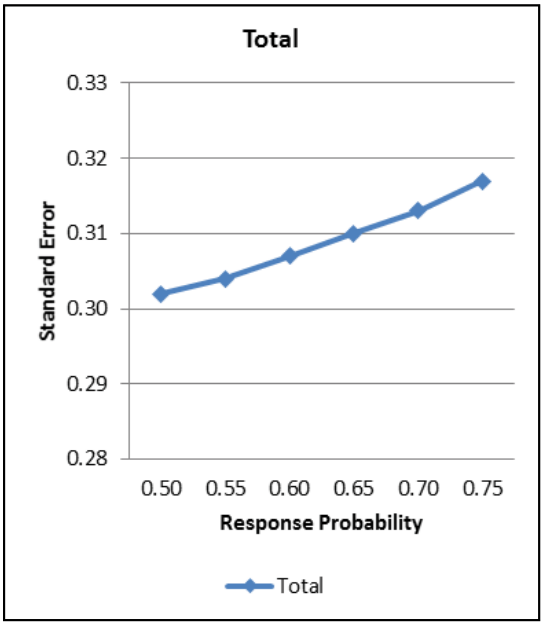


Table 13–4. Average Standard Errors for Various Response Probabilities — Reading

Number of Items	Response Probability	Total	DC 1	DC 2	DC 3	DC 4	DC 5
55 total (11 per DC)	0.50	0.302	0.738	0.739	0.723	0.743	0.743
55 total (11 per DC)	0.55	0.304	0.739	0.744	0.731	0.741	0.751
55 total (11 per DC)	0.60	0.307	0.742	0.744	0.733	0.756	0.771
55 total (11 per DC)	0.65	0.310	0.747	0.751	0.742	0.766	0.781
55 total (11 per DC)	0.70	0.313	0.755	0.756	0.751	0.772	0.800
55 total (11 per DC)	0.75	0.317	0.767	0.762	0.764	0.784	0.823

Figure 13–4. Average Standard Errors for Various Response Probabilities — Reading



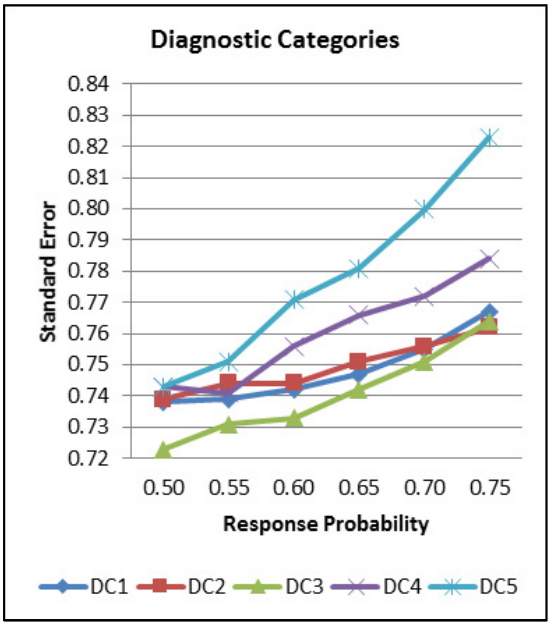


Table 13–5. Average Standard Errors for Various Response Probabilities — Science

Number of Items	Response Probability	Total	DC 1	DC 2	DC 3	DC 4
52 total (13 per DC)	0.50	0.300	0.601	0.599	0.602	0.599
52 total (13 per DC)	0.55	0.299	0.600	0.599	0.600	0.599
52 total (13 per DC)	0.60	0.300	0.602	0.601	0.603	0.604
52 total (13 per DC)	0.65	0.303	0.612	0.608	0.609	0.611
52 total (13 per DC)	0.70	0.308	0.624	0.622	0.619	0.626
52 total (13 per DC)	0.75	0.315	0.642	0.642	0.636	0.644

Figure 13–5. Average Standard Errors for Various Response Probabilities — Science

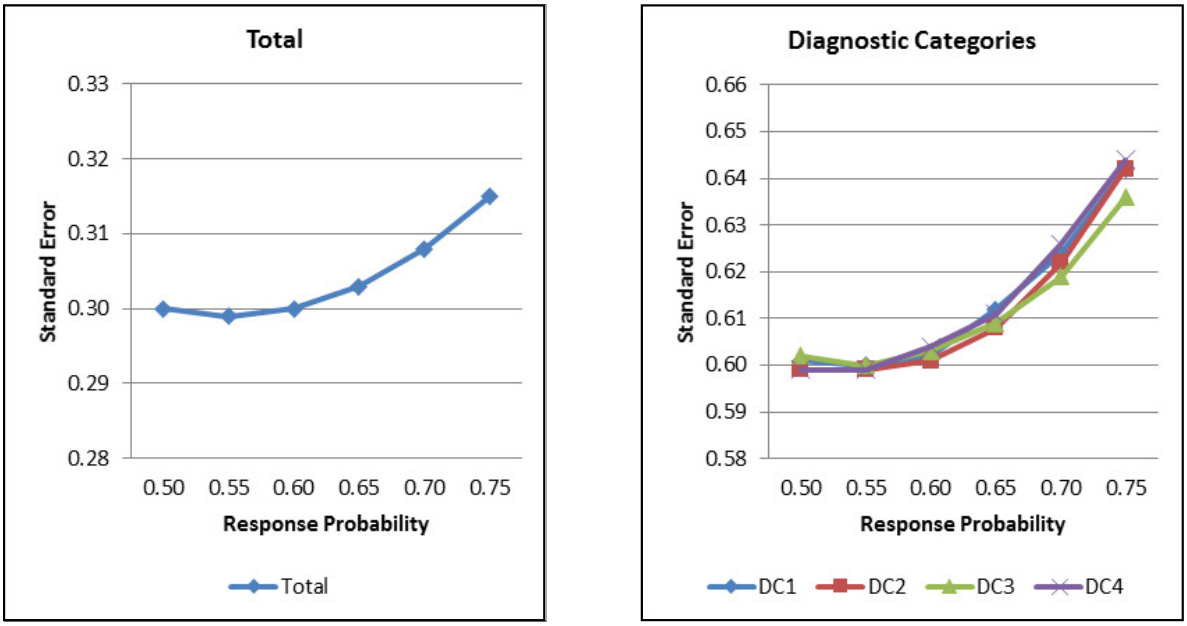
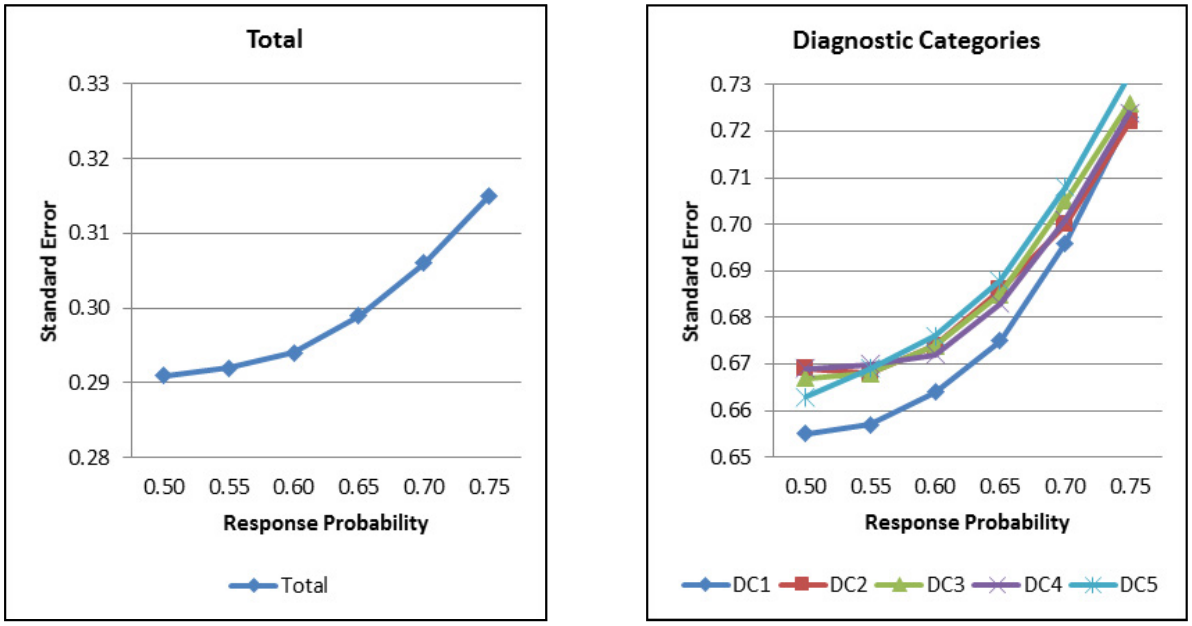


Table 13–6. Average Standard Errors for Various Response Probabilities — Writing

Number of Items	Response Probability	Total	DC 1	DC 2	DC 3	DC 4	DC 5
52 total (13 per DC)	0.50	0.291	0.655	0.669	0.667	0.669	0.663
52 total (13 per DC)	0.55	0.292	0.657	0.668	0.668	0.670	0.669
52 total (13 per DC)	0.60	0.294	0.664	0.674	0.674	0.672	0.676
52 total (13 per DC)	0.65	0.299	0.675	0.686	0.685	0.683	0.688
52 total (13 per DC)	0.70	0.306	0.696	0.700	0.705	0.701	0.708
52 total (13 per DC)	0.75	0.315	0.723	0.722	0.726	0.724	0.732

Figure 13–6. Average Standard Errors for Various Response Probabilities — Writing



Again, increasing response probability incrementally from 0.50 leads to increases in standard error. The increase in standard error is gradual at first and becomes more pronounced around 0.65.

For CDT tests designed for students in grade 6 and above, the response probability is set at 0.5. This is based on the desire for low standard errors at the diagnostic category level and the grade level of students testing. As part of the CDT training, students are told that the test is computer adaptive and designed to challenge them.

For CDT tests designed for students in grades 3 through 5, the response probability is set at 0.65. This response probability results in higher standard errors for the same number of items. However, there was concern that younger students may not have much experience with tests designed to be so challenging and could conceivably give up on a test that is perceived to be "too hard."

ITEM POOL REFINEMENT

The CAT algorithm has configurable elements that allow for refinement of the item pool used in item selection. The two configurable elements are:

- **Restrict pool**—The ability to restrict the available item pool by grade/course at various points in the test.

  For example, Chemistry items are not available for the first 20 items of CDT Science Grades 6–HS test.

- **Favor items**—The ability to favor items that are close to the student's grade when evaluating items near a student's estimated score.

  For example, if a student is in grade 8 and the item selection routine finds appropriate items (in terms of difficulty) in grades 4, 5, 6, 7, and 8, item selection can favor items at or close to grade 8. It is possible that no items near a student's grade are appropriate in terms of difficulty. In such a case, the CAT algorithm will select items further away from the student's grade but appropriate based on item difficulty.

The difference between restricting the pool and favoring items is that when the pool is restricted, some items may NOT be selected. With favoring, all non-restricted items are eligible for administration, but they are made more or less LIKELY to be selected based on closeness to student grade.

PASSAGE RELATED CONCERNS

As previously mentioned, the CDT tests in the reading content area are passage-based. CDT passages have between one and seven associated items. The CAT algorithm does not require that all items associated with a passage be administered. Instead, it evaluates all possible combinations of items within a passage. Item sequencing within a passage is preserved when items are presented to the student. For example, if a six-item passage is selected and items 1 and 4 are NOT administered, then the items administered in order will be 2, 3, 5, and 6.

The configurable elements of passage-based CAT include:

- **Passage minimum percent**—Define the minimum percentage of the items associated with a passage to be used.

  For example, if the passage minimum percent is set at 80, then the selection routine will consider combinations such as 1 of 1 (100%), 4 of 5 (80%), 5 of 6 (83%), and 6 of 6 (100%). It will not consider combinations such as 1 of 2 (50%), 3 of 4 (75%), 3 of 5 (60%), etc. Near the end of a test, the passage minimum percent constraint may need to be loosened in order to meet content constraints such as number of items per diagnostic category.

- **Passage evaluation criteria**—Multiple factors are considered when evaluating and ranking each passage combination to determine the best combination to administer to a student. They include:
  - Percent of items associated with the passage used; the higher the percent, the higher the combination is ranked
  - Number of items associated with the passage used; the higher the number, the higher the combination is ranked
  - Distance between items' difficulties and student's estimated score; the smaller the distance, the higher the combination is ranked
  - Distance between the items' grade levels and the student's grade level; the smaller the distance, the higher the combination is ranked

  Different weights may be assigned to each of the factors. For example, if all of the weight is put on number of items used, then the algorithm will select the passages with the most associated items and administer all of them until the maximum number of items is reached.

TEST NAVIGATION

Many versions of computer adaptive tests do not allow students to skip items in the test or back up to previously answered items and change answers due to some complicating factors.

If students are allowed to skip items, the CAT algorithm would need to select additional items without any additional information (no change to Rasch ability estimates). Taken to the extreme, a student with no prior CDT scores who skipped every item starting with the first would receive an entire test of average items. It would not be adaptive at all.

If students are allowed to back up and change answers, Rasch ability estimates are re-calculated when answered are changed. This additional information can be used to select additional items but would not change previously selected items. For example, suppose a student is on item twenty-five and goes back to change the answer to item eleven from wrong to right. The total and corresponding diagnostic category Rasch ability estimates would go up. That additional information can be used in selection of items twenty-six and beyond. However, items twelve through twenty-five are not reselected even though different items may have been selected if item eleven was initially answered correctly. When it comes to items twelve through twenty-five, "the train has left the station."

Also, if students are allowed to back up in the test, additional considerations must be put in place to ensure that the answer to one item does not cue another.

Currently all CDT tests except Reading Grades 3–5 and Reading/Lit Grades 6–HS do not allow skipping items or backing up and changing answers. On CDT tests in the reading content area, students are allowed to skip items within a passage. For example, when presented with a passage and five associated items, the student does not have to answer questions one through five in that order without skipping. If a student tries to navigate to the next passage without answering all of the items associated with a passage, the test engine will prompt the student to answer all items and will not move on to the next passage until all are answered.

TERMINATION

The CAT algorithm allows for both a fixed- or variable-length test.

With fixed length, the test ends when a student has taken a predefined number of items total and in each diagnostic category.

With variable length, the algorithm stops administering items from a diagnostic category when one of two conditions is satisfied:

- A student has taken at least a predefined minimum number of items in that diagnostic category and the standard error is below a predefined threshold

  OR

- A student has taken a predefined maximum number of items in that diagnostic category

The test ends when one of the two conditions above is satisfied for each of the diagnostic categories.

Note that with both fixed- and variable-length tests, there is no requirement that the predefined number of items in diagnostic categories be equal.

CAT CONFIGURATION – MATH GRADES 6–HS

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 7 student will start with an item near the average difficulty of grade 7 items. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to restrict the pool and to favor items close to a student's grade. The pool restrictions are:

- no Algebra I items will be administered in the first 5 items,
- no Geometry items will be administered in the first 10 items, and
- no Algebra II items will be administered in the first 20 items.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.59 to 0.62.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Math Grades 6–HS. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – ALGEBRA I

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to restrict the pool and to favor items close to Algebra I. The pool restriction is that no Algebra II items will be administered in the first 16 items.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.60 to 0.62.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Algebra I. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – GEOMETRY

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to favor items close to Geometry. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.60 to 0.61.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Geometry. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – ALGEBRA II

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to favor items close to Algebra II. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.60 to 0.66.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Algebra II. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – MATH GRADES 3–5

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 4 student will start with an item near the average difficulty of grade 4 items. Items are selected where the response probability is 0.65, meaning a student has a 65% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.62, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to restrict the pool and to favor items close to a student's grade. The pool restrictions are:

- no grade 7 items will be administered in the first 5 items,
- no grade 8 items will be administered in the first 10 items,
- no Algebra I items will be administered in the first 20 items, and
- no Geometry or Algebra II items will be administered.

Simulations were run with this configuration. On average:

- a total of 52 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.31, and
- standard errors for the diagnostic categories are in the range of 0.61 to 0.62.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Math Grades 3–5. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – READING/LIT GRADES 6–HS

The test has five diagnostic categories. Each student will take between 10 and 12 operational items per diagnostic category for a total test of 50 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 7 student will start with an item near the average difficulty of grade 7 items. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 10 operational items in that diagnostic category and the standard error is below 0.75, or
- a student has taken 12 operational items in that diagnostic category.

Functionality is used to run CAT with passages and favor items close to student's grade. There are no pool restrictions.

Passage minimum percent is set at 66%. That is, whenever possible, only passage combinations that use 66% or more of the associated items are used. (Near the end of a test, the passage minimum percent constraint may need to be loosened in order to meet content constraints.) Many simulations were run to arrive at this percent. On the one hand, testing time and reading load should be minimized. Therefore, students should not have to read long passages for only one or two items. On the other hand, using all items associated with a passage may not be desirable since some items are far from a student's estimated score. Given a limited number of items, those that are either too easy or too hard should not be used.

In evaluating and ranking passages, percent of items associated with the passage is not used. Simulation results indicate that if it is factored into evaluations, students take many short passages because 1 of 1 (100%) and 2 of 2 (100%) are ranked higher than 5 of 6 (83%) and 4 of 5 (80%), for example.

Simulations were run with this configuration. On average:

- a total of 56 operational items are administered—about 11 per diagnostic category,
- a total of 14 passages are administered,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.72 to 0.80.

DIAGNOSTIC CATEGORY TESTS

Diagnostic category tests in the reading content area are different than the other content areas because items are passage-based. Testing a single diagnostic category would result in students reading full passages for only one or two items. Instead, diagnostic category tests associated with Reading/Literature Grades 6–HS are separated by text type – literature text or informational text. Each of the two tests have three diagnostic categories[3]. Students take between 10 and 12 operational items per diagnostic category for a total test of 30 to 36 operational items. Diagnostic category tests were first available on January 28, 2019.

CAT CONFIGURATION – READING GRADES 3–5

The test has five diagnostic categories. Each student will take between 10 and 12 operational items per diagnostic category for a total test of 50 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 4 student will start with an item near the average difficulty of grade 4 items. Items are selected where the response probability is 0.65, meaning a student has a 65% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 10 operational items in that diagnostic category and the standard error is below 0.77, or
- a student has taken 12 operational items in that diagnostic category.

Functionality is used to run CAT with passages and favor items close to student's grade. The pool is restricted so that students will not receive passages associated with a grade that is more than four grades above the student's grade.

[3] Key Ideas and Details, Craft and Structure/Integration of Knowledge and Ideas, Vocabulary

Passage minimum percent is set at 66%. That is, whenever possible, only passage combinations that use 66% or more of the associated items are used. (Near the end of a test, the passage minimum percent constraint may need to be loosened in order to meet content constraints.) Many simulations were run to arrive at this percent. On the one hand, testing time and reading load should be minimized. Therefore, students should not have to read long passages for only one or two items. On the other hand, using all items associated with a passage may not be desirable since some items are far from a student's estimated score. Given a limited number of items, those that are either too easy or too hard should not be used.

In evaluating and ranking passages, percent of items associated with the passage is not used. Simulation results indicate that if it is factored into evaluations, students take many short passages because 1 of 1 (100%) and 2 of 2 (100%) are ranked higher than 5 of 6 (83%) and 4 of 5 (80%), for example.

Simulations were run with this configuration. On average:

- a total of 56 operational items are administered—about 11 per diagnostic category,
- a total of 14 passages are administered,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.73 to 0.78.

Note that the standard error is higher for in reading than the other content areas. This is because Reading/Lit Grades 6–HS and Reading Grades 3–5 are passage-based. Rather than selecting one targeted item at a time, the item selection routine evaluates and selects multiple items associated with a given passage. In general, items selected in this manner are not as close to the targeted response probability as stand-alone items selected one by one.

DIAGNOSTIC CATEGORY TESTS

Diagnostic category tests in the reading content area are different than the other content areas because items are passage-based. Testing a single diagnostic category would result in students reading full passages for only one or two items. Instead, diagnostic category tests associated with Reading/Literature Grades 6–HS are separated by text type – literature text or informational text. Each of the two tests have three diagnostic categories[4]. Students take between 10 and 12 operational items per diagnostic category for a total test of 30 to 36 operational items. Diagnostic category tests were first available on January 28, 2019.

CAT CONFIGURATION – SCIENCE GRADES 6–HS

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 7 student will start with an item near the average difficulty of grade 7 items. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to restrict the pool and to favor items close to a student's grade. The pool restrictions are:

- no grade 11 items will be administered in the first 20 items UNLESS the student is in grade 11 or 12,
- no Biology or Chemistry items will be administered in the first 20 items.

[4] Key Ideas and Details, Craft and Structure/Integration of Knowledge and Ideas, Vocabulary

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.61 to 0.63.

CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Science Grades 6–HS. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – BIOLOGY

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to favor items close to Biology. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.61 to 0.63.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Biology. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – CHEMISTRY

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.60, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to favor items close to Chemistry. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 53 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.31, and
- standard errors for the diagnostic categories are in the range of 0.60 to 0.65.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Chemistry. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – SCIENCE GRADES 3–5

The test has four diagnostic categories. Each student will take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 4 student will start with an item near the average difficulty of grade 4 items. Items are selected where the response probability is 0.65, meaning a student has a 65% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 12 operational items in that diagnostic category and the standard error is below 0.62, or
- a student has taken 15 operational items in that diagnostic category.

Functionality is used to restrict the pool and to favor items close to a student's grade. The pool restrictions are:

- no grade 11 items will be administered in the first 40 items, and
- no Biology or Chemistry items will be administered.

Simulations were run with this configuration. On average:

- a total of 52 operational items are administered—about 13 per diagnostic category,
- standard error for the total score is 0.31, and
- standard errors for the diagnostic categories are in the range of 0.61 to 0.62.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the four diagnostic categories associated with Science Grades 3–5. Given that the content is limited to a single diagnostic category, the number of items is increased from 12 to 15 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – WRITING/ENG COMP GRADES 6–HS

The test has five diagnostic categories. Each student will take between 10 and 12 operational items per diagnostic category for a total test of 50 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 7 student will start with an item near the average difficulty of grade 7 items. Items are selected where the response probability is 0.5, meaning a student has a 50% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 10 operational items in that diagnostic category and the standard error is below 0.65, or
- a student has taken 12 operational items in that diagnostic category.

Functionality is used to favor items close to the student's grade. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 56 operational items are administered—about 11 per diagnostic category,
- standard error for the total score is 0.29, and
- standard errors for the diagnostic categories are in the range of 0.66 to 0.70.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the five diagnostic categories associated with Writing/English Composition Grades 6–HS. Given that the content is limited to a single diagnostic category, the number of items is increased from 10 to 12 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

CAT CONFIGURATION – WRITING GRADES 3–5

The test has five diagnostic categories. Each student will take between 10 and 12 operational items per diagnostic category for a total test of 50 to 60 operational items. With no prior information about a student, the starting point in each diagnostic category will be an item of average difficulty by grade level. For example, a grade 4 student will start with an item near the average difficulty of grade 4 items. Items are selected where the response probability is 0.65, meaning a student has a 65% chance of answering correctly. The CAT algorithm will stop administering items in a diagnostic category when one of two conditions is satisfied:

- a student has taken at least 10 operational items in that diagnostic category and the standard error is below 0.67, or
- a student has taken 12 operational items in that diagnostic category.

Functionality is used to favor items close to the student's grade. There are no pool restrictions.

Simulations were run with this configuration. On average:

- a total of 55 operational items are administered—about 11 per diagnostic category,
- standard error for the total score is 0.30, and
- standard errors for the diagnostic categories are in the range of 0.68 to 0.70.

DIAGNOSTIC CATEGORY TESTS

Starting on January 28, 2019, CDTs were available that allowed students to take a single one of the five diagnostic categories associated with Writing Grades 3–5. Given that the content is limited to a single diagnostic category, the number of items is increased from 10 to 12 per diagnostic category to 15 to 18. This allows for more precise estimates (lower standard error) than the full test in which all diagnostic categories are tested.

Tables 13–7 through 13–12 summarize CAT configurations by content area.

Table 13–7. CAT Configuration Summary – Mathematics

	Math Grades 3–5	Math Grades 6–HS
Number of DCs	4	4
Number of OP Items per DC	12–15	12–15
Number of OP Items Total	48–60	48–60
Number of FT Items Total	0	0
Entry Point: No Prior CDT	average item by grade	average item by grade
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.65	0.50
Item Selection: Favor Items	close to student grade	close to student grade
Item Selection: Pool Restriction	Items 1–5: no Grade 7	Items 1–5: no Algebra I
Item Selection: Pool Restriction	Items 1–10: no Grade 8	Items 1–10: no Geometry
Item Selection: Pool Restriction	Items 1–20: no Algebra I	Items 1–20: no Algebra II
Item Selection: Pool Restriction	No Geometry	
Item Selection: Pool Restriction	No Algebra II	
Navigation	no skip; no backtrack	no skip; no backtrack
Termination	12 items per DC, SE < 0.62 OR 15 items per DC	12 items per DC, SE < 0.60 OR 15 items per DC

DC = Diagnostic Category

Table 13–8. CAT Configuration Summary – Algebra I, Geometry, and Algebra II

	Algebra I	Geometry	Algebra II
Number of DCs	4	4	4
Number of OP Items per DC	12–15	12–15	12–15
Number of OP Items Total	48–60	48–60	48–60
Number of FT Items Total	0	0	0
Entry Point: No Prior CDT	average item	average item	average item
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.50	0.50	0.50
Item Selection: Favor Items	close to Algebra I	close to Geometry	close to Algebra II
Item Selection: Pool Restriction	Items 1–16: no Algebra II	None	None
Navigation	no skip; no backtrack	no skip; no backtrack	no skip; no backtrack
Termination	12 items per DC, SE < 0.60 OR 15 items per DC	12 items per DC, SE < 0.60 OR 15 items per DC	12 items per DC, SE < 0.60 OR 15 items per DC

DC = Diagnostic Category

Table 13–9. CAT Configuration Summary – Reading

	Reading Grades 3–5	Reading/Lit Grades 6–HS
Number of DCs	5	5
Number of OP Items per DC	10–12	10–12
Number of OP Items Total	50–60	50–60
Number of FT Items Total	0	0
Entry Point: No Prior CDT	average item by grade	average item by grade
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.65	0.50
Item Selection: Favor Items	close to student grade	close to student grade
Item Selection: Pool Restriction	No items from grades more than four above student grade	None
Passage Min %	66	66
Navigation	skip items within passage	skip items within passage
Termination	10 items per DC, SE < 0.77 OR 12 items per DC	10 items per DC, SE < 0.75 OR 12 items per DC

DC = Diagnostic Category

Table 13–10. CAT Configuration Summary – Science

	Science Grades 3–5	Science Grades 6–HS
Number of DCs	4	4
Number of OP Items per DC	12–15	12–15
Number of OP Items Total	48–60	48–60
Number of FT Items Total	0	0
Entry Point: No Prior CDT	average item by grade	average item by grade
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.65	0.50
Item Selection: Favor Items	close to student grade	close to student grade
Item Selection: Pool Restriction	Items 1–40: no grade 11	Students in grades 6–10 Items 1–20: no grade 11, Biology, or Chemistry
Item Selection: Pool Restriction	No Biology	Students in grades 11–12 Items 1–20: no Biology, or Chemistry
Item Selection: Pool Restriction	No Chemistry	
Navigation	no skip; no backtrack	no skip; no backtrack
Termination	12 items per DC, SE < 0.62 OR 15 items per DC	12 items per DC, SE < 0.60 OR 15 items per DC

DC = Diagnostic Category

Table 13–11. CAT Configuration Summary – Biology and Chemistry

	Biology	Chemistry
Number of DCs	4	4
Number of OP Items per DC	12–15	12–15
Number of OP Items Total	48–60	48–60
Number of FT Items Total	0	0
Entry Point: No Prior CDT	average item	average item
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.50	0.50
Item Selection: Favor Items	close to Biology	close to Chemistry
Item Selection: Pool Restriction	None	None
Navigation	no skip; no backtrack	no skip; no backtrack
Termination	12 items per DC, SE < 0.60 OR 15 items per DC	12 items per DC, SE < 0.60 OR 15 items per DC

DC = Diagnostic Category

Table 13–12. CAT Configuration Summary – Writing

	Writing Grades 3–5	Writing/Eng Comp Gr 6–HS
Number of DCs	5	5
Number of OP Items per DC	10–12	10–12
Number of OP Items Total	50–60	50–60
Number of FT Items Total	0	0
Entry Point: No Prior CDT	average item by grade	average item by grade
Entry Point: Prior CDT	prior diagnostic scores	prior diagnostic scores
Item Selection: Rasch Ability Estimates	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate	After locator, use total estimate until the fifth item in a DC; then switch to DC estimate
Item Selection: Response Probability	0.65	0.50
Item Selection: Favor Items	close to student grade	close to student grade
Item Selection: Pool Restriction	None	None
Navigation	no skip; no backtrack	no skip; no backtrack
Termination	10 items per DC, SE < 0.67 OR 12 items per DC	10 items per DC, SE < 0.65 OR 12 items per DC

DC = Diagnostic Category

CHAPTER FOURTEEN: SCORES AND SCORE REPORTS

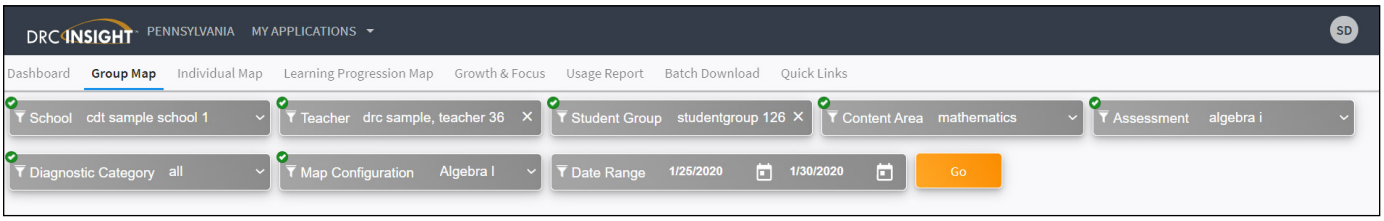
Teachers will receive immediate and usable data to be used for targeting instruction to meet the needs of individual students. The CDT Interactive Reports provide direct links to resources in SAS, including specific lesson plans, interventions, and other resources. The reports can also show the progress of students across test administrations. This overview summarizes the steps in accessing the interactive reports, as well as the types of information available for each type of report.

ACCESSING INTERACTIVE REPORTS

Any user with the role of District, School, or Teacher has the ability to view CDT Interactive Reports accessed through the DRC INSIGHT Portal. Once the user is logged in, Report Delivery can be selected under MY APPLICATIONS, at the top of the screen. Next, the user selects CDT Interactive Reports. The user is presented general information on the Dashboard with separate tabs for each report. Once a report is selected, the user will begin to make selections within the available pre-filters to generate the report to be displayed.

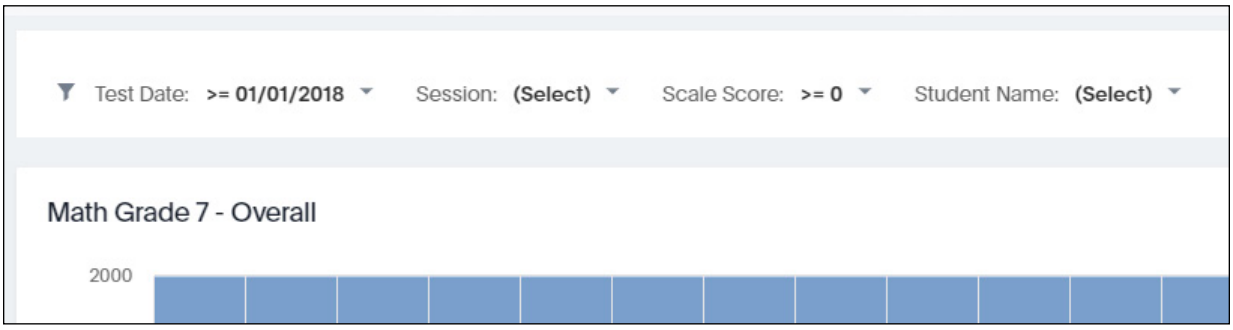
The pre-filters include: District, School, Teacher, Student Group, Content Area, Assessment, Diagnostic Category, Map Configuration, Date Range. The filters are smart filters. This means the filter will pre-populate based on the data the user has access to or based on the previous filter selections made.

Figure 14–1. Pre-Filter Screen



A secondary set of filters is available within each report to further refine the data reported on the page. Each reporting table and map has its own filters and selections to sort the data in a way that maximizes the ability for teachers to evaluate performance for a group or sub-set of students. The secondary filters enable teachers to view a subset of the data displayed. In the example below, filters include test date, test session selection(s), scale score range, and student name selection(s).

Figure 14–2. Secondary-Filter Screen

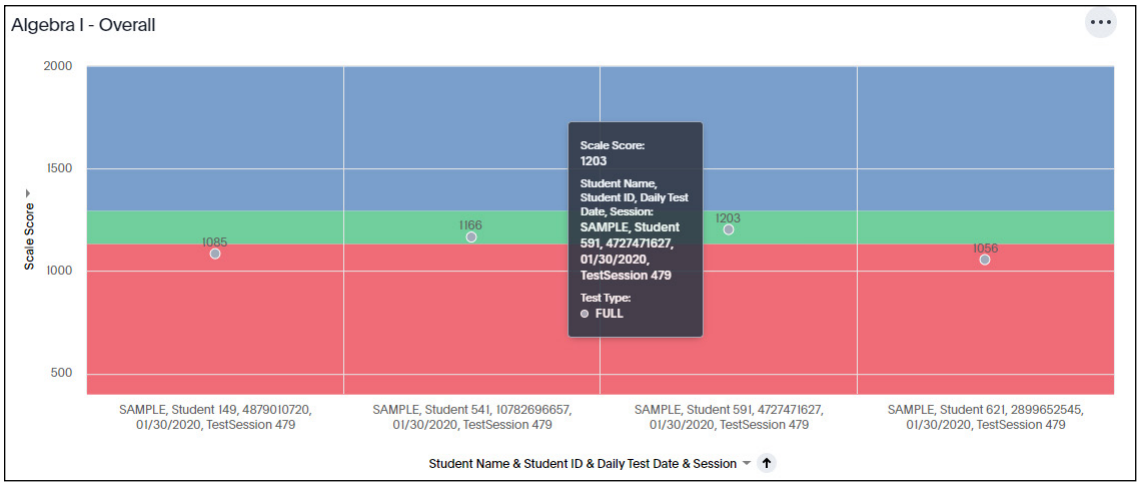


There are four types of interactive reports for the CDT: Group Map, Individual Map, Learning Progression Map, and Growth and Focus Map.

GROUP MAP

The group-level reports provide teachers insightful information and data about classroom performance, including students' strengths to build on and areas of need. The group maps allow users to view overall classroom performance on a given assessment; to view eligible content associated with student scores; and to sort the data in various ways to make smaller student groups for targeted instruction. The group map is made up of several different data displays, which are discussed below.

Figure 14–3. Sample Overall Group Map



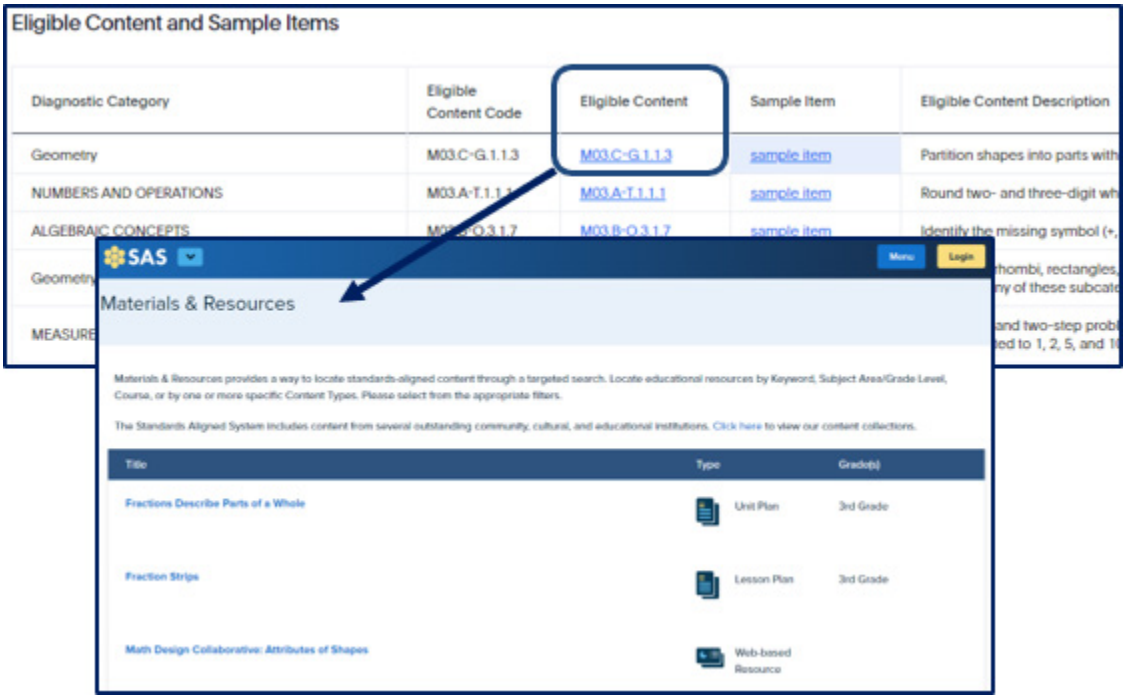
The Interactive Reports use colors to indicate relative **Strengths to Build On** and **Areas of Need**. Each descriptor correlates with a color range on the scale: Green/Blue = Strengths to Build On; Red = Areas of Need.

- Each gray dot on the Group Map represents a single student score.
- Additional information displays when the user hovers over the dot: student name, test date, and score.
- Only students within the Student Group with scores will have a gray dot appear on the map.
- All dots represent the assessment score(s) during the administration window, identified using the Date Range filter.
- The Group Map is intended to provide general assessment information based on a group of student scores within a full CDT assessment and/or Diagnostic Category CDT.
- The Diagnostic Category maps, found below the Group Map, provide all scores associated with the Diagnostic Categories tested within the full assessment, as well as for all individual Diagnostic Category CDT assessments completed. The scores are represented with yellow plotted dots.
- The data is also displayed in a grid that provides a complete list of the students within the selected student group with accompanying score information. The data from the grid can be exported as a CSV file.

Initially, the Group Map shows the entire vertical scale (representing scores from 200 to 2000 for Lower Grades Mathematics, Lower Grades Reading, Lower Grades Science, and Lower Grades Writing; representing scores from 400 to 2000 for Mathematics, Algebra I, Algebra II, Geometry, Reading/Literature, Science, Biology, Chemistry, and Writing/English Composition). The **Scale Score** filter provides the user the ability to narrow the reported set of students down to those falling in similar ranges.

If a user chooses one diagnostic category from the prefilters then additional detail is displayed at an eligible content level, including a description of the eligible content, links to a sample item, and links to instructional resources found on the SAS website.

Figure 14–4. Eligible Content and Sample Items

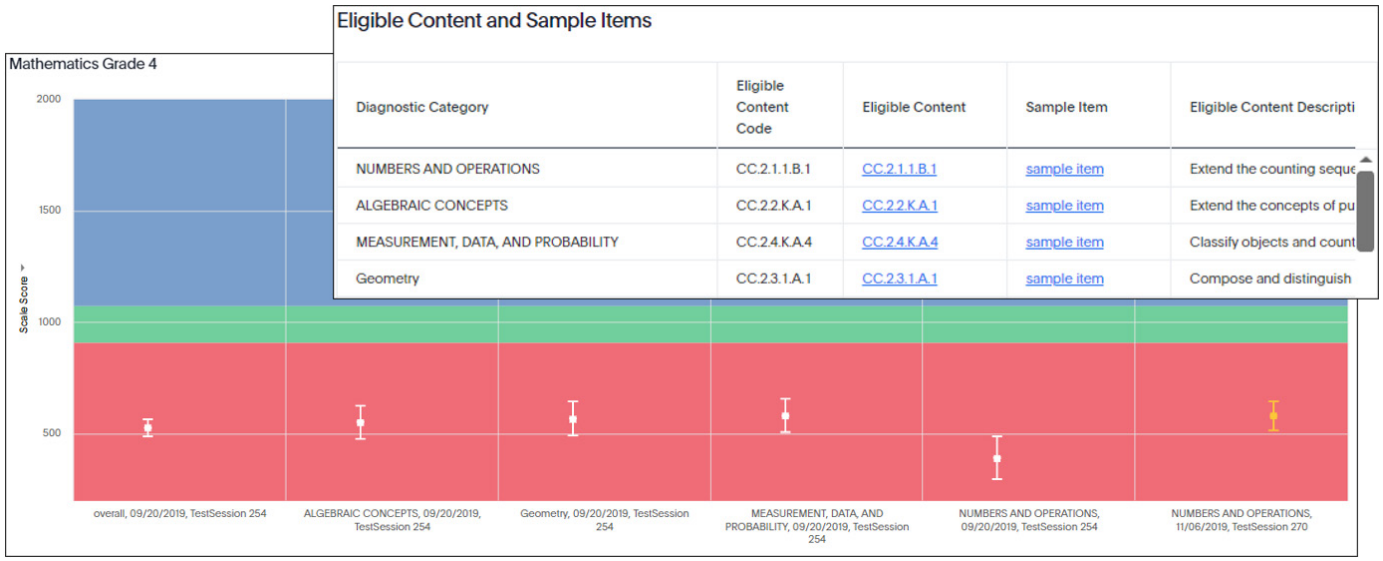


INDIVIDUAL MAP

The CDT Individual Map shows how an individual student performed on a given assessment, with scores plotted on the CDT scale. The columns in the Individual Map represent the individual tests taken by the student. In adherence reporting guidelines outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), a standard error band is displayed for each score. This interval represents the range where the student would likely score if tested again without additional instruction. The use of error bands supports more-accurate interpretation of scores (i.e., not over-interpreting scores) since error bands that overlap indicate that scores are not significantly different.

Similar to the Group Map, the Individual Map provides Eligible Content and Sample Items at the student level. This display contains sample items, eligible content descriptions, and links to materials and resources on SAS.

Figure 14–5. Sample Individual Map and Eligible Content Associated with a Student's Score



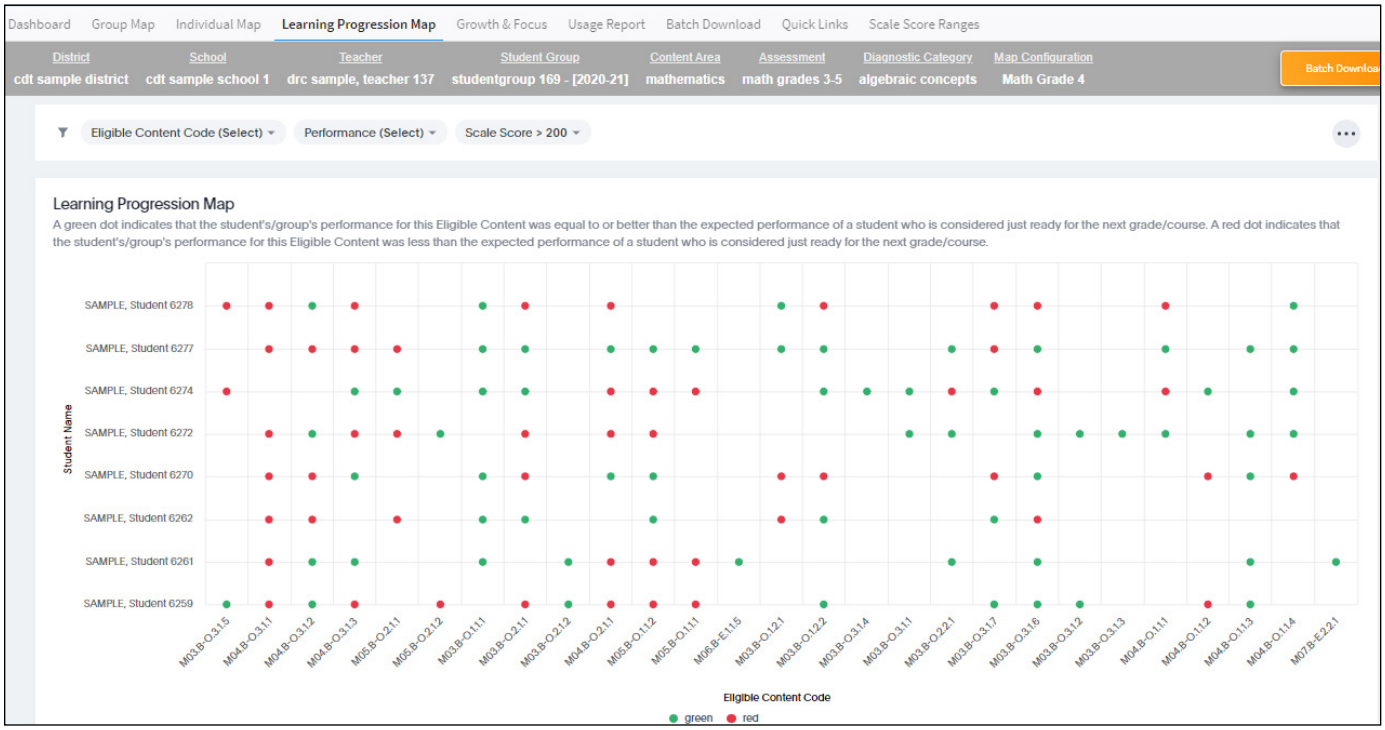
The Individual Map has the ability to show the all assessments that apply to the preliminary filter selections for an individual student. The Individual Map is intended to provide general Instructional Enrichment (a set of Eligible Content) based on a student's score within a Diagnostic Category. Additional data displays on the Individual Map include hover overs and a grid view.

GROUP AND INDIVIDUAL LEARNING PROGRESSION MAP

The Group and Individual Learning Progression Map is a graphical representation about how learning may typically move toward increased understanding over time based on Eligible Content. Each column represents the Eligible Content in a subject's domain and subdomain and for a specific grade level or course. Each row represents student performance on the eligible content.

- A ***green*** dot indicates that the student was presented with at least one test item for the Eligible Content and performed as well or better than the expected performance of a student who is considered just ready for the next grade/course.
- A ***red*** dot indicates that the student was presented with at least one test item from the Eligible Content and the student's performance was less than the expected performance of a student who is considered just ready for the next grade/course.
- An empty box represents Eligible Content that is available, but the student was not presented with any test items from that Eligible Content.

Figure 14–6. Sample Learning Progression Map



Additional data displays within the Learning Progression map include a summary by eligible content code, a grade-level summary, and information in a grid format.

GROWTH AND FOCUS REPORT

The Growth and Focus report is designed to aid teachers in goal-setting with students by identifying students that fall in the “all” group or a “focus” group.

Students within the “all” group have an overall score higher or equal to the score at the bottom of the green area of the Group Map for the previous grade level. Students within the “focus” group are students who have an overall score that is less than the bottom of green of the previous grade level. These are students who could benefit from individual or small-group interventions.

The table is designed to allow educators to view one test event or compare two test events to determine if a student had significant growth between test sessions. The calculations that generate this report use the standard error information found in the Individual Maps to determine if there was significant growth.

Figure 14–7. Growth and Focus Report

Growth & Focus - Math Grade 4

Student Name ↑	Test Session 1	Scale Score 1	Test Session 2	Scale Score 2	Change in Score	SEM	Significant Growth	Group
SAMPLE, Student 6259	TestSession 192	773	TestSession 489	1022	249	72	yes	all
SAMPLE, Student 6261	TestSession 192	940	TestSession 489	1115	175	74	yes	all
SAMPLE, Student 6270	TestSession 192	762	TestSession 489	709	-53	72	no	focus
SAMPLE, Student 6272	TestSession 192	985	TestSession 489	1063	78	81	no	all
SAMPLE, Student 6274	TestSession 192	772	TestSession 489	813	41	77	no	focus
SAMPLE, Student 6277	TestSession 192	967	TestSession 489	885	-82	76	no	all
SAMPLE, Student 6278	TestSession 192	583	TestSession 489	756	173	74	yes	focus

OTHER CDT REPORTING COMPONENTS

STUDENT CONFERENCING REPORT: Data gives educators a comprehensive student-level report that compares recent test events for the same content area tested. This can include full CDT events, as well as individual Diagnostic Category CDT results. Teachers frequently use this report during one-on-one conferences with students and during conversations with parents because it provides a clear picture of student performance that can be easily printed or distributed via email.

DISTRICT STUDENT DATA FILE: District-level data is easily accessible using the District Data File download feature. This file is updated nightly and can be downloaded at any time throughout the CDT testing window. It includes student-level data for all schools within the district that have completed test events.

USAGE REPORT: DRC provides CDT usage reports in a variety of user-friendly formats (pie charts, bar graphs, CSV export files) that will allow administrators at SDP to easily view a summary of CDT usage by school. Users can filter the report content to best match their intended use.

CHAPTER FIFTEEN: OPERATIONAL ADMINISTRATION 2020–2021

This chapter contains summary information about the operational administration of the Classroom Diagnostic Tools (CDT) during the 2020–2021 school year. Two types of CDTs were available—full CDTs and diagnostic category CDTs. Full CDTs test four or five diagnostic categories in one test session. Diagnostic category CDTs focus on a single diagnostic category in math, science and writing, or a single text type with three diagnostic categories in reading. Results in this chapter focus on full CDTs except where specifically noted.

FREQUENCIES

Tables 15–1 through 15–3 present information related to the number of students who were administered one or more CDT tests in the 2020–2021 school year. Tables 15–1a and 15–1b show the number of students who have taken each CDT. Some of these students have taken the same CDT test multiple times or have taken multiple CDT tests. Tables 15–1a and 15–1b count only the first administration of each CDT test. Data about multiple administrations of the same test and multiple CDT tests are presented in Tables 15–2 and 15–3, respectively.

Table 15–1a. Number of Students Taking the First Administration of a Full CDT by Grade Level

CDT	3	4	5	6	7	8	9	10	11	12	TOTAL
Math Grades 3–5	13,635	13,913	16,333	-	-	-	-	-	-	-	43,881
Math Grades 6–HS	-	-	-	19,519	20,056	16,419	197	102	49	17	56,359
Algebra I	-	-	-	37	848	4,656	14,026	4,873	1,906	581	26,927
Geometry	-	-	-	3	20	219	942	1,434	633	75	3,326
Algebra II	-	-	-	2	7	151	985	1,668	1,031	278	4,122
Reading Grades 3–5	11,401	11,693	13,379	-	-	-	-	-	-	-	36,473
Reading/Lit Grades 6–HS	-	-	-	15,106	16,271	15,879	13,420	20,408	4,584	606	86,274
Science Grades 3–5	882	7,393	2,923	-	-	-	-	-	-	-	11,198
Science Grades 6–HS	-	-	-	5,413	8,847	15,827	246	97	51	23	30,504
Biology	-	-	-	2	3	63	12,357	13,139	2,650	399	28,613
Chemistry	-	-	-	0	0	2	398	923	569	51	1,943
Writing Grades 3–5	1,369	1,426	2,005	-	-	-	-	-	-	-	4,800
Writing/Eng Comp Grades 6–HS				3,049	3,855	3,928	1,451	1,481	382	128	14,274

Table 15–1b. Number of Students Taking the First Administration of a Diagnostic Category CDT by Grade Level

CDT	Diagnostic Category	3	4	5	6	7	8	9	10	11	12	TOTAL
Math Grades 3–5	Numbers and Operations	6,031	6,225	7,319	-	-	-	-	-	-	-	19,575
Math Grades 3–5	Algebraic Concepts	3,023	2,931	2,461	-	-	-	-	-	-	-	8,415
Math Grades 3–5	Geometry	1,413	1,313	1,436	-	-	-	-	-	-	-	4,162
Math Grades 3–5	Measurement, Data, and Probability	1,431	1,499	1,308	-	-	-	-	-	-	-	4,238
Math Grades 6–HS	Numbers and Operations	-	-	-	5,871	4,497	2,974	194	17	8	4	13,565
Math Grades 6–HS	Algebraic Concepts	-	-	-	3,270	3,757	3,103	34	5	6	3	10,178
Math Grades 6–HS	Geometry	-	-	-	1,543	1,279	897	17	1	2	0	3,739
Math Grades 6–HS	Measurement, Data, and Probability	-	-	-	961	713	1,166	16	2	4	0	2,862
Algebra I	Operations with Real Numbers and Expressions	-	-	-	0	11	414	3,098	1,148	317	97	5,085
Algebra I	Linear Equations & Inequalities	-	-	-	8	418	736	3,350	1,345	435	173	6,465
Algebra I	Functions & Coordinate Geometry	-	-	-	8	141	592	2,083	933	293	93	4,143
Algebra I	Data Analysis	-	-	-	0	0	94	1,133	616	166	70	2,079
Geometry	Geometric Properties	-	-	-	0	0	0	22	164	137	77	400
Geometry	Congruence, Similarity, and Proofs	-	-	-	0	0	0	14	81	17	8	120
Geometry	Coordinate Geometry and Right Triangles	-	-	-	0	0	69	25	129	140	24	387
Geometry	Measurement	-	-	-	0	0	4	88	178	59	11	340
Algebra II	Operations with Complex Numbers	-	-	-	0	0	7	97	81	20	17	222
Algebra II	Non-Linear Expressions & Equations	-	-	-	0	0	0	1	103	60	45	209

Table 15–1b (continued). Number of Students Taking the First Administration of a Diagnostic Category CDT by Grade Level

CDT	Diagnostic Category	3	4	5	6	7	8	9	10	11	12	TOTAL
Algebra II	Functions	-	-	-	0	0	7	133	254	203	37	634
Algebra II	Data Analysis	-	-	-	0	0	0	0	3	14	15	32
Reading Grades 3–5	Informational Text	3,785	4,508	5,229	-	-	-	-	-	-	-	13,522
Reading Grades 3–5	Literature Text	3,633	3,953	4,953	-	-	-	-	-	-	-	12,539
Reading/Lit Grades 6–HS	Informational Text	-	-	-	4,965	3,125	3,503	2,352	3,235	766	233	18,179
Reading/Lit Grades 6–HS	Literature Text	-	-	-	5,487	4,565	4,141	1,949	4,621	681	144	21,588
Science Grades 3–5	The Nature of Science	0	886	291	-	-	-	-	-	-	-	1,177
Science Grades 3–5	Biological Sciences	37	521	149	-	-	-	-	-	-	-	707
Science Grades 3–5	Physical Sciences	36	856	1	-	-	-	-	-	-	-	893
Science Grades 3–5	Earth and Space Sciences	52	616	0	-	-	-	-	-	-	-	668
Science Grades 6–HS	The Nature of Science	-	-	-	2,361	3,065	4,629	112	11	13	0	10,191
Science Grades 6–HS	Biological Sciences	-	-	-	360	3,481	1,155	76	9	11	0	5,092
Science Grades 6–HS	Physical Sciences	-	-	-	378	732	2,512	22	14	0	4	3,662
Science Grades 6–HS	Earth and Space Sciences	-	-	-	1,447	1,522	1,946	10	9	15	0	4,949
Biology	Basic Biological Principles/ Chemical Basis for Life	-	-	-	2	1	8	3,599	3,788	550	99	8,047
Biology	Bioenergetics/ Homeostasis and Transport	-	-	-	1	1	2	1,929	2,830	426	52	5,241
Biology	Cell Growth and Reproduction/ Genetics	-	-	-	0	0	2	1,071	2,227	289	27	3,616
Biology	Theory of Evolution/Ecology	-	-	-	0	0	4	993	2,117	214	36	3,364

Table 15–1b (continued). Number of Students Taking the First Administration of a Diagnostic Category CDT by Grade Level

CDT	Diagnostic Category	3	4	5	6	7	8	9	10	11	12	TOTAL
Chemistry	Properties and Classification of Matter	-	-	-	0	0	0	4	72	148	19	243
Chemistry	Atomic Structure and The Periodic Table	-	-	-	0	66	1	17	208	394	52	738
Chemistry	The Mole and Chemical Bonding	-	-	-	0	0	0	0	40	114	13	167
Chemistry	Chemical Relationships and Reactions	-	-	-	0	0	0	0	2	20	9	31
Writing Grades 3–5	Quality of Writing: Focus and Organization	405	498	319	-	-	-	-	-	-	-	1,222
Writing Grades 3–5	Quality of Writing: Content and Style	39	126	105	-	-	-	-	-	-	-	270
Writing Grades 3–5	Quality of Writing: Editing	67	127	34	-	-	-	-	-	-	-	228
Writing Grades 3–5	Conventions: Punctuation, Capitalization, and Spelling	694	627	691	-	-	-	-	-	-	-	2,012
Writing Grades 3–5	Conventions: Grammar and Sentence Formation	830	943	1,233	-	-	-	-	-	-	-	3,006
Writing/Eng Comp Grades 6–HS	Quality of Writing: Focus and Organization	-	-	-	305	351	301	127	92	82	12	1,270
Writing/Eng Comp Grades 6–HS	Quality of Writing: Content and Style	-	-	-	0	224	391	18	96	84	9	822
Writing/Eng Comp Grades 6–HS	Quality of Writing: Editing	-	-	-	0	96	237	5	15	18	9	380
Writing/Eng Comp Grades 6–HS	Conventions: Punctuation, Capitalization, and Spelling	-	-	-	855	746	1,281	87	296	410	54	3,729
Writing/Eng Comp Grades 6–HS	Conventions: Grammar and Sentence Formation	-	-	-	1,109	975	1,297	22	161	57	8	3,629

Table 15–2a. Multiple Administrations of the Same Full CDT Test

CDT	Students with 1 Administration	Students with 2 Administrations	Students with 3 Administrations	Students with 4 Administrations	Students with 5 Administrations
Math Grades 3–5	43,881	23,597	7,705	833	3
Math Grades 6–HS	56,359	29,985	10,530	682	1
Algebra I	26,927	12,161	3,523	114	0
Geometry	3,326	1,581	222	2	0
Algebra II	4,122	1,728	173	0	0
Reading Grades 3–5	36,473	20,663	7,332	382	3
Reading/Lit Grades 6–HS	86,274	40,934	11,376	470	1
Science Grades 3–5	11,198	6,699	2,670	111	0
Science Grades 6–HS	30,504	15,598	5,765	121	1
Biology	28,613	12,584	4,638	639	3
Chemistry	1,943	849	278	0	0
Writing Grades 3–5	4,800	2,823	851	4	0
Writing/Eng Comp Gr 6–HS	14,274	5,924	1,349	1	0

Table 15–2b. Multiple Administrations of the Same Diagnostic Category CDT Test

CDT	Diagnostic Category	Students with 1 Administration	Students with 2 Administrations	Students with 3 Administrations	Students with 4 Administrations	Students with 5 Administrations
Math Grades 3–5	Numbers and Operations	19,575	5,712	1,802	86	0
Math Grades 3–5	Algebraic Concepts	8,415	1,599	142	68	0
Math Grades 3–5	Geometry	4,162	1,265	123	27	0
Math Grades 3–5	Measurement, Data, and Probability	4,238	1,384	116	23	0
Math Grades 6–HS	Numbers and Operations	13,565	4,658	1,410	56	0
Math Grades 6–HS	Algebraic Concepts	10,178	3,417	632	0	0
Math Grades 6–HS	Geometry	3,739	1,221	74	0	0
Math Grades 6–HS	Measurement, Data, and Probability	2,862	1,132	30	0	0
Algebra I	Operations with Real Numbers and Expressions	5,085	1,418	221	0	0
Algebra I	Linear Equations & Inequalities	6,465	2,321	618	21	0
Algebra I	Functions & Coordinate Geometry	4,143	1,020	110	4	0
Algebra I	Data Analysis	2,079	519	20	0	0
Geometry	Geometric Properties	400	177	24	0	0
Geometry	Congruence, Similarity, and Proofs	120	37	0	0	0
Geometry	Coordinate Geometry and Right Triangles	387	157	0	0	0
Geometry	Measurement	340	133	10	0	0
Algebra II	Operations with Complex Numbers	222	57	0	0	0
Algebra II	Non-Linear Expressions & Equations	209	104	0	0	0
Algebra II	Functions	634	238	6	0	0
Algebra II	Data Analysis	32	13	0	0	0
Reading Grades 3–5	Informational Text	13,522	5,239	250	79	0
Reading Grades 3–5	Literature Text	12,539	5,154	148	79	5

Table 15–2b (continued). Multiple Administrations of the Same Diagnostic Category CDT Test

CDT	Diagnostic Category	Students with 1 Administration	Students with 2 Administrations	Students with 3 Administrations	Students with 4 Administrations	Students with 5 Administrations
Reading/Lit Grades 6–HS	Informational Text	18,179	5,943	829	48	0
Reading/Lit Grades 6–HS	Literature Text	21,588	8,885	775	11	1
Science Grades 3–5	The Nature of Science	1,177	528	0	0	0
Science Grades 3–5	Biological Sciences	707	304	0	0	0
Science Grades 3–5	Physical Sciences	893	267	0	0	0
Science Grades 3–5	Earth and Space Sciences	668	251	0	0	0
Science Grades 6–HS	The Nature of Science	10,191	3,752	190	0	0
Science Grades 6–HS	Biological Sciences	5,092	1,953	302	1	0
Science Grades 6–HS	Physical Sciences	3,662	1,464	374	0	0
Science Grades 6–HS	Earth and Space Sciences	4,949	2,130	508	0	0
Biology	Basic Biological Principles/ Chemical Basis for Life	8,047	1,625	139	0	0
Biology	Bioenergetics/ Homeostasis and Transport	5,241	948	63	0	0
Biology	Cell Growth and Reproduction/ Genetics	3,616	608	78	0	0
Biology	Theory of Evolution/Ecology	3,364	851	1	1	0
Chemistry	Properties and Classification of Matter	243	19	0	0	0
Chemistry	Atomic Structure and The Periodic Table	738	324	34	1	0
Chemistry	The Mole and Chemical Bonding	167	21	0	0	0
Chemistry	Chemical Relationships and Reactions	31	10	0	0	0

Table 15–2b (continued). Multiple Administrations of the Same Diagnostic Category CDT Test

CDT	Diagnostic Category	Students with 1 Administration	Students with 2 Administrations	Students with 3 Administrations	Students with 4 Administrations	Students with 5 Administrations
Writing Grades 3–5	Quality of Writing: Focus and Organization	1,222	283	97	24	0
Writing Grades 3–5	Quality of Writing: Content and Style	270	122	115	83	0
Writing Grades 3–5	Quality of Writing: Editing	228	120	99	25	0
Writing Grades 3–5	Conventions: Punctuation, Capitalization, and Spelling	2,012	1,241	639	84	0
Writing Grades 3–5	Conventions: Grammar and Sentence Formation	3,006	1,446	665	84	0
Writing/Eng Comp Gr 6–HS	Quality of Writing: Focus and Organization	1,270	264	44	0	0
Writing/Eng Comp Gr 6–HS	Quality of Writing: Content and Style	822	240	0	0	0
Writing/Eng Comp Gr 6–HS	Quality of Writing: Editing	380	93	0	0	0
Writing/Eng Comp Gr 6–HS	Conventions: Punctuation, Capitalization, and Spelling	3,729	1,752	761	0	0
Writing/Eng Comp Gr 6–HS	Conventions: Grammar and Sentence Formation	3,629	1,487	712	0	0

Table 15–3a. Number of Students in Grades 3 through 5 Taking Multiple Full CDT Tests

Grades 3 through 5	Math	Reading	Science	Writing
Math Grades 3–5	-	-	-	-
Reading Grades 3–5	34,077	-	-	-
Science Grades 3–5	6,963	6,235	-	-
Writing Grades 3–5	4,531	4,236	1,824	-

Table 15–3b. Number of Students in Grades 6 and above Taking Multiple Full CDT Tests

Grades 6 and above	Math	Algebra I	Geometry	Algebra II	Reading/ Literature	Science	Biology	Chemistry	Writing/ English Comp.
Math Grades 6–HS	-	-	-	-	-	-	-	-	-
Algebra I	1,543	-	-	-	-	-	-	-	-
Geometry	42	135	-	-	-	-	-	-	-
Algebra II	5	70	242	-	-	-	-	-	-
Reading/Lit Grades 6–HS	38,580	13,092	1,512	1,915	-	-	-	-	-
Science Grades 6–HS	18,331	2,765	130	146	17,294	-	-	-	-
Biology	121	7,656	694	909	12,764	103	-	-	-
Chemistry	7	67	649	331	662	0	76	-	-
Writing/Eng Comp Grades 6–HS	8,273	2,303	394	748	10,961	4,827	796	317	-

Further demographic information about students tested with the CDT is found in the next section.

DEMOGRAPHIC CHARACTERISTICS

COMPOSITION OF SAMPLE USED IN SUBSEQUENT TABLES

To avoid double counting of students, the following demographic tables are based on students' first administration for a given CDT test. Students taking only diagnostic category tests are counted with the parent test[1]. For example, a student taking Math Grades 3–5 Numbers and Operations is counted under Math Grades 3–5. Students who took the same test multiple times are counted only once. Students who took different tests are counted for each test. For example, if a student took CDT Algebra I twice, he or she is counted only once in the Algebra I counts; if a student took Algebra I once and Biology once, he or she is counted in both Algebra I and Biology counts.

COLLECTION OF STUDENT DEMOGRAPHIC INFORMATION

Data for analyses of demographic characteristics were obtained primarily from information supplied by school district personnel through the Pennsylvania Information Management System (PIMS) and subsequently transmitted to DRC. However, teachers may assign CDT tests to students who do not have data in PIMS at the time of testing. This may result in CDT records with incomplete demographic information.

DEMOGRAPHIC CHARACTERISTICS

Frequency data for various demographic categories are presented in Tables 15–4 through 15–16. Shown at the bottom of the appropriate table is the number of students with a total test score on which the column percentages are based. Percentages in some categories may sum to a quantity below 100 percent due to missing data.

Analyses are broken out by grade level. However, in the case of course-specific CDT tests (Algebra I, Geometry, Algebra II, Biology, and Chemistry), students across multiple grades may be enrolled in the course.

Caution should be used in interpreting CDT demographic data, since participation is voluntary and complete demographic data via PIMS is not required for testing. This is especially true for rows in the lower half of the tables (e.g. IEP, Migrant, and Economically Disadvantaged) because these typically have more than ninety-five percent blank responses.

[1] Approximately 16% of students take only diagnostic category tests.

Table 15–4. Demographic Characteristics of Students Taking CDT Math Grades 3–5

Demographic or Educational Characteristic	Gr. 3	Gr. 4	Gr. 5	Total
Female (N)	8,113	8,126	9,699	25,938
Female (Pct)	49.84%	48.65%	48.99%	49.14%
Male (N)	8,166	8,578	10,100	26,844
Male (Pct)	50.16%	51.35%	51.01%	50.86%
American Indian or Alaskan Native (N)	34	41	44	119
American Indian or Alaskan Native (Pct)	0.21%	0.25%	0.22%	0.23%
Black/African American non-Hispanic (N)	1,382	1,455	1,745	4,582
Black/African American non-Hispanic (Pct)	8.49%	8.71%	8.81%	8.68%
Hispanic (N)	1,803	1,852	2,545	6,200
Hispanic (Pct)	11.08%	11.09%	12.85%	11.75%
White/Caucasian non-Hispanic (N)	11,470	11,837	13,557	36,864
White/Caucasian non-Hispanic (Pct)	70.46%	70.86%	68.47%	69.84%
Multi-Racial non-Hispanic (N)	1,082	1,015	1,265	3,362
Multi-Racial non-Hispanic (Pct)	6.65%	6.08%	6.39%	6.37%
Asian non-Hispanic (N)	493	492	626	1,611
Asian non-Hispanic (Pct)	3.03%	2.95%	3.16%	3.05%
Native Hawaiian or Pacific Islander (N)	15	12	17	44
Native Hawaiian or Pacific Islander (Pct)	0.09%	0.07%	0.09%	0.08%
IEP (N)	56	57	49	162
IEP (Pct)	0.34%	0.34%	0.25%	0.31%
Migrant student (N)	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	25	25	55	105
Economically disadvantaged (Pct)	0.15%	0.15%	0.28%	0.20%
Number of students	16,279	16,704	19,799	52,782

Table 15–5. Demographic Characteristics of Students Taking CDT Math Grades 6–HS

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	11,119	12,061	9,452	165	53	13	6	32,869
Female (Pct)	49.60%	49.83%	48.36%	41.98%	44.54%	22.81%	28.57%	49.24%
Male (N)	11,300	12,142	10,092	228	66	44	15	33,887
Male (Pct)	50.40%	50.17%	51.64%	58.02%	55.46%	77.19%	71.43%	50.76%
American Indian or Alaskan Native (N)	50	59	47	1	1	0	0	158
American Indian or Alaskan Native (Pct)	0.22%	0.24%	0.24%	0.25%	0.84%	0.00%	0.00%	0.24%
Black/African American non-Hispanic (N)	2,135	2,377	2,108	48	22	9	2	6,701
Black/African American non-Hispanic (Pct)	9.52%	9.82%	10.79%	12.21%	18.49%	15.79%	9.52%	10.04%
Hispanic (N)	2,581	2,620	2,465	49	11	7	4	7,737
Hispanic (Pct)	11.51%	10.83%	12.61%	12.47%	9.24%	12.28%	19.05%	11.59%
White/Caucasian non-Hispanic (N)	15,729	17,317	13,596	267	74	38	14	47,035
White/Caucasian non-Hispanic (Pct)	70.16%	71.55%	69.57%	67.94%	62.18%	66.67%	66.67%	70.46%
Multi-Racial non-Hispanic (N)	1,328	1,336	908	24	10	3	1	3,610
Multi-Racial non-Hispanic (Pct)	5.92%	5.52%	4.65%	6.11%	8.40%	5.26%	4.76%	5.41%
Asian non-Hispanic (N)	574	460	365	4	1	0	0	1,404
Asian non-Hispanic (Pct)	2.56%	1.90%	1.87%	1.02%	0.84%	0.00%	0.00%	2.10%
Native Hawaiian or Pacific Islander (N)	22	34	55	0	0	0	0	111
Native Hawaiian or Pacific Islander (Pct)	0.10%	0.14%	0.28%	0.00%	0.00%	0.00%	0.00%	0.17%
IEP (N)	78	123	99	4	4	2	1	311
IEP (Pct)	0.35%	0.51%	0.51%	1.02%	3.36%	3.51%	4.76%	0.47%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	55	64	57	4	0	0	1	181
Economically disadvantaged (Pct)	0.25%	0.26%	0.29%	1.02%	0.00%	0.00%	4.76%	0.27%
Number of students	22,419	24,203	19,544	393	119	57	21	66,756

Table 15–6. Demographic Characteristics of Students Taking CDT Algebra I

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	19	557	2,906	8,859	3,048	1,098	355	16,842
Female (Pct)	42.22%	44.85%	52.12%	49.06%	44.87%	43.45%	45.57%	48.09%
Male (N)	26	685	2,670	9,200	3,745	1,429	424	18,179
Male (Pct)	57.78%	55.15%	47.88%	50.94%	55.13%	56.55%	54.43%	51.91%
American Indian or Alaskan Native (N)	1	1	8	79	42	11	0	142
American Indian or Alaskan Native (Pct)	2.22%	0.08%	0.14%	0.44%	0.62%	0.44%	0.00%	0.41%
Black/African American non-Hispanic (N)	3	21	143	2,447	1,026	552	212	4,404
Black/African American non-Hispanic (Pct)	6.67%	1.69%	2.56%	13.55%	15.10%	21.84%	27.21%	12.58%
Hispanic (N)	0	30	223	2,510	1,088	461	175	4,487
Hispanic (Pct)	0.00%	2.42%	4.00%	13.90%	16.02%	18.24%	22.46%	12.81%
White/Caucasian non-Hispanic (N)	32	1,019	4,731	11,808	4,105	1,336	356	23,387
White/Caucasian non-Hispanic (Pct)	71.11%	82.05%	84.85%	65.39%	60.43%	52.87%	45.70%	66.78%
Multi-Racial non-Hispanic (N)	0	64	273	766	407	121	18	1,649
Multi-Racial non-Hispanic (Pct)	0.00%	5.15%	4.90%	4.24%	5.99%	4.79%	2.31%	4.71%
Asian non-Hispanic (N)	9	106	183	433	124	45	18	918
Asian non-Hispanic (Pct)	20.00%	8.53%	3.28%	2.40%	1.83%	1.78%	2.31%	2.62%
Native Hawaiian or Pacific Islander (N)	0	1	15	16	1	1	0	34
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.08%	0.27%	0.09%	0.01%	0.04%	0.00%	0.10%
IEP (N)	0	0	3	54	46	33	17	153
IEP (Pct)	0.00%	0.00%	0.05%	0.30%	0.68%	1.31%	2.18%	0.44%
Migrant student (N)	0	0	0	0	0	0	1	1
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.13%	0.00%
Economically disadvantaged (N)	0	1	4	57	37	30	24	153
Economically disadvantaged (Pct)	0.00%	0.08%	0.07%	0.32%	0.54%	1.19%	3.08%	0.44%
Number of students	45	1,242	5,576	18,059	6,793	2,527	779	35,021

Table 15–7. Demographic Characteristics of Students Taking CDT Geometry

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	0	6	130	528	966	477	79	2,186
Female (Pct)	0.00%	30.00%	44.52%	49.53%	52.39%	50.32%	46.20%	50.32%
Male (N)	3	14	162	538	878	471	92	2,158
Male (Pct)	100.00%	70.00%	55.48%	50.47%	47.61%	49.68%	53.80%	49.68%
American Indian or Alaskan Native (N)	0	0	1	3	12	1	1	18
American Indian or Alaskan Native (Pct)	0.00%	0.00%	0.34%	0.28%	0.65%	0.11%	0.58%	0.41%
Black/African American non-Hispanic (N)	0	0	11	79	298	130	41	559
Black/African American non-Hispanic (Pct)	0.00%	0.00%	3.77%	7.41%	16.16%	13.71%	23.98%	12.87%
Hispanic (N)	0	1	7	50	160	143	32	393
Hispanic (Pct)	0.00%	5.00%	2.40%	4.69%	8.68%	15.08%	18.71%	9.05%
White/Caucasian non-Hispanic (N)	2	11	210	850	1,263	616	83	3,035
White/Caucasian non-Hispanic (Pct)	66.67%	55.00%	71.92%	79.74%	68.49%	64.98%	48.54%	69.87%
Multi-Racial non-Hispanic (N)	0	1	44	32	38	30	10	155
Multi-Racial non-Hispanic (Pct)	0.00%	5.00%	15.07%	3.00%	2.06%	3.16%	5.85%	3.57%
Asian non-Hispanic (N)	1	7	19	50	70	26	4	177
Asian non-Hispanic (Pct)	33.33%	35.00%	6.51%	4.69%	3.80%	2.74%	2.34%	4.07%
Native Hawaiian or Pacific Islander (N)	0	0	0	2	3	2	0	7
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.00%	0.00%	0.19%	0.16%	0.21%	0.00%	0.16%
IEP (N)	0	0	0	4	9	24	12	49
IEP (Pct)	0.00%	0.00%	0.00%	0.38%	0.49%	2.53%	7.02%	1.13%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	0	0	0	8	8	16	10	42
Economically disadvantaged (Pct)	0.00%	0.00%	0.00%	0.75%	0.43%	1.69%	5.85%	0.97%
Number of students	3	20	292	1,066	1,844	948	171	4,344

Table 15–8. Demographic Characteristics of Students Taking CDT Algebra II

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	0	3	74	577	1,050	685	185	2,574
Female (Pct)	0.00%	42.86%	46.54%	51.33%	51.88%	52.81%	52.56%	51.84%
Male (N)	2	4	85	547	974	612	167	2,391
Male (Pct)	100.00%	57.14%	53.46%	48.67%	48.12%	47.19%	47.44%	48.16%
American Indian or Alaskan Native (N)	0	0	0	7	7	2	1	17
American Indian or Alaskan Native (Pct)	0.00%	0.00%	0.00%	0.62%	0.35%	0.15%	0.28%	0.34%
Black/African American non-Hispanic (N)	0	0	1	33	121	73	33	261
Black/African American non-Hispanic (Pct)	0.00%	0.00%	0.63%	2.94%	5.98%	5.63%	9.38%	5.26%
Hispanic (N)	0	0	6	24	140	100	33	303
Hispanic (Pct)	0.00%	0.00%	3.77%	2.14%	6.92%	7.71%	9.38%	6.10%
White/Caucasian non-Hispanic (N)	1	6	141	936	1,553	1,004	264	3,905
White/Caucasian non-Hispanic (Pct)	50.00%	85.71%	88.68%	83.27%	76.73%	77.41%	75.00%	78.65%
Multi-Racial non-Hispanic (N)	0	0	3	12	42	45	5	107
Multi-Racial non-Hispanic (Pct)	0.00%	0.00%	1.89%	1.07%	2.08%	3.47%	1.42%	2.16%
Asian non-Hispanic (N)	1	1	8	111	158	70	16	365
Asian non-Hispanic (Pct)	50.00%	14.29%	5.03%	9.88%	7.81%	5.40%	4.55%	7.35%
Native Hawaiian or Pacific Islander (N)	0	0	0	1	3	3	0	7
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.00%	0.00%	0.09%	0.15%	0.23%	0.00%	0.14%
IEP (N)	0	0	0	0	5	4	5	14
IEP (Pct)	0.00%	0.00%	0.00%	0.00%	0.25%	0.31%	1.42%	0.28%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	0	0	0	0	7	1	1	9
Economically disadvantaged (Pct)	0.00%	0.00%	0.00%	0.00%	0.35%	0.08%	0.28%	0.18%
Number of students	2	7	159	1,124	2,024	1,297	352	4,965

Table 15–9. Demographic Characteristics of Students Taking CDT Reading Grades 3–5

Demographic or Educational Characteristic	Gr. 3	Gr. 4	Gr. 5	Total
Female (N)	7,000	7,305	8,661	22,966
Female (Pct)	49.69%	48.65%	49.26%	49.19%
Male (N)	7,086	7,711	8,921	23,718
Male (Pct)	50.31%	51.35%	50.74%	50.81%
American Indian or Alaskan Native (N)	30	39	43	112
American Indian or Alaskan Native (Pct)	0.21%	0.26%	0.24%	0.24%
Black/African American non-Hispanic (N)	1,223	1,347	1,603	4,173
Black/African American non-Hispanic (Pct)	8.68%	8.97%	9.12%	8.94%
Hispanic (N)	1,622	1,743	2,071	5,436
Hispanic (Pct)	11.51%	11.61%	11.78%	11.64%
White/Caucasian non-Hispanic (N)	9,849	10,561	12,180	32,590
White/Caucasian non-Hispanic (Pct)	69.92%	70.33%	69.28%	69.81%
Multi-Racial non-Hispanic (N)	885	817	1,034	2,736
Multi-Racial non-Hispanic (Pct)	6.28%	5.44%	5.88%	5.86%
Asian non-Hispanic (N)	464	495	639	1,598
Asian non-Hispanic (Pct)	3.29%	3.30%	3.63%	3.42%
Native Hawaiian or Pacific Islander (N)	13	14	12	39
Native Hawaiian or Pacific Islander (Pct)	0.09%	0.09%	0.07%	0.08%
IEP (N)	54	52	46	152
IEP (Pct)	0.38%	0.35%	0.26%	0.33%
Migrant student (N)	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	24	23	57	104
Economically disadvantaged (Pct)	0.17%	0.15%	0.32%	0.22%
Number of students	14,086	15,016	17,582	46,684

Table 15–10. Demographic Characteristics of Students Taking CDT Reading/Lit Grades 6–HS

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	9,606	9,868	9,410	8,148	12,415	2,678	408	52,533
Female (Pct)	49.75%	49.77%	48.63%	49.32%	48.89%	46.71%	44.16%	49.07%
Male (N)	9,701	9,961	9,940	8,374	12,981	3,055	516	54,528
Male (Pct)	50.25%	50.23%	51.37%	50.68%	51.11%	53.29%	55.84%	50.93%
American Indian or Alaskan Native (N)	38	40	50	55	116	24	1	324
American Indian or Alaskan Native (Pct)	0.20%	0.20%	0.26%	0.33%	0.46%	0.42%	0.11%	0.30%
Black/African American non-Hispanic (N)	1,733	1,723	1,675	1,505	2,297	605	149	9,687
Black/African American non-Hispanic (Pct)	8.98%	8.69%	8.66%	9.11%	9.04%	10.55%	16.13%	9.05%
Hispanic (N)	2,074	2,100	2,030	1,722	2,149	571	195	10,841
Hispanic (Pct)	10.74%	10.59%	10.49%	10.42%	8.46%	9.96%	21.10%	10.13%
White/Caucasian non-Hispanic (N)	13,784	14,412	14,110	11,745	18,934	4,002	538	77,525
White/Caucasian non-Hispanic (Pct)	71.39%	72.68%	72.92%	71.09%	74.56%	69.81%	58.23%	72.41%
Multi-Racial non-Hispanic (N)	1,142	1,123	997	924	1,115	366	32	5,699
Multi-Racial non-Hispanic (Pct)	5.91%	5.66%	5.15%	5.59%	4.39%	6.38%	3.46%	5.32%
Asian non-Hispanic (N)	525	409	435	554	774	161	9	2,867
Asian non-Hispanic (Pct)	2.72%	2.06%	2.25%	3.35%	3.05%	2.81%	0.97%	2.68%
Native Hawaiian or Pacific Islander (N)	11	22	53	17	11	4	0	118
Native Hawaiian or Pacific Islander (Pct)	0.06%	0.11%	0.27%	0.10%	0.04%	0.07%	0.00%	0.11%
IEP (N)	88	112	92	57	75	54	27	505
IEP (Pct)	0.46%	0.56%	0.48%	0.34%	0.30%	0.94%	2.92%	0.47%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	55	46	48	63	87	58	27	384
Economically disadvantaged (Pct)	0.28%	0.23%	0.25%	0.38%	0.34%	1.01%	2.92%	0.36%
Number of students	19,307	19,829	19,350	16,522	25,396	5,733	924	107,061

Table 15–11. Demographic Characteristics of Students Taking CDT Science Grades 3–5

Demographic or Educational Characteristic	Gr. 3	Gr. 4	Gr. 5	Total
Female (N)	446	3,876	1,644	5,966
Female (Pct)	45.93%	49.29%	49.19%	49.00%
Male (N)	525	3,987	1,698	6,210
Male (Pct)	54.07%	50.71%	50.81%	51.00%
American Indian or Alaskan Native (N)	1	11	3	15
American Indian or Alaskan Native (Pct)	0.10%	0.14%	0.09%	0.12%
Black/African American non-Hispanic (N)	184	1,098	381	1,663
Black/African American non-Hispanic (Pct)	18.95%	13.96%	11.40%	13.66%
Hispanic (N)	71	1,397	1,017	2,485
Hispanic (Pct)	7.31%	17.77%	30.43%	20.41%
White/Caucasian non-Hispanic (N)	612	4,755	1,719	7,086
White/Caucasian non-Hispanic (Pct)	63.03%	60.47%	51.44%	58.20%
Multi-Racial non-Hispanic (N)	49	459	137	645
Multi-Racial non-Hispanic (Pct)	5.05%	5.84%	4.10%	5.30%
Asian non-Hispanic (N)	54	137	84	275
Asian non-Hispanic (Pct)	5.56%	1.74%	2.51%	2.26%
Native Hawaiian or Pacific Islander (N)	0	6	1	7
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.08%	0.03%	0.06%
IEP (N)	0	15	2	17
IEP (Pct)	0.00%	0.19%	0.06%	0.14%
Migrant student (N)	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	2	17	12	31
Economically disadvantaged (Pct)	0.21%	0.22%	0.36%	0.25%
Number of students	971	7,863	3,342	12,176

Table 15–12. Demographic Characteristics of Students Taking CDT Science Grades 6–HS

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	4,239	6,781	9,846	176	51	46	10	21,149
Female (Pct)	49.16%	49.48%	48.49%	41.81%	46.36%	51.11%	43.48%	48.87%
Male (N)	4,383	6,923	10,461	245	59	44	13	22,128
Male (Pct)	50.84%	50.52%	51.51%	58.19%	53.64%	48.89%	56.52%	51.13%
American Indian or Alaskan Native (N)	15	35	68	2	0	0	0	120
American Indian or Alaskan Native (Pct)	0.17%	0.26%	0.33%	0.48%	0.00%	0.00%	0.00%	0.28%
Black/African American non-Hispanic (N)	804	1,645	2,210	71	10	16	2	4,758
Black/African American non-Hispanic (Pct)	9.32%	12.00%	10.88%	16.86%	9.09%	17.78%	8.70%	10.99%
Hispanic (N)	1,210	1,680	2,280	56	20	21	10	5,277
Hispanic (Pct)	14.03%	12.26%	11.23%	13.30%	18.18%	23.33%	43.48%	12.19%
White/Caucasian non-Hispanic (N)	5,874	9,305	14,147	260	76	45	11	29,718
White/Caucasian non-Hispanic (Pct)	68.13%	67.90%	69.67%	61.76%	69.09%	50.00%	47.83%	68.67%
Multi-Racial non-Hispanic (N)	435	719	1,044	21	0	4	0	2,223
Multi-Racial non-Hispanic (Pct)	5.05%	5.25%	5.14%	4.99%	0.00%	4.44%	0.00%	5.14%
Asian non-Hispanic (N)	278	300	513	11	4	4	0	1,110
Asian non-Hispanic (Pct)	3.22%	2.19%	2.53%	2.61%	3.64%	4.44%	0.00%	2.56%
Native Hawaiian or Pacific Islander (N)	6	20	45	0	0	0	0	71
Native Hawaiian or Pacific Islander (Pct)	0.07%	0.15%	0.22%	0.00%	0.00%	0.00%	0.00%	0.16%
IEP (N)	36	85	72	2	1	2	4	202
IEP (Pct)	0.42%	0.62%	0.35%	0.48%	0.91%	2.22%	17.39%	0.47%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	13	17	40	2	0	3	7	82
Economically disadvantaged (Pct)	0.15%	0.12%	0.20%	0.48%	0.00%	3.33%	30.43%	0.19%
Number of students	8,622	13,704	20,307	421	110	90	23	43,277

Table 15–13. Demographic Characteristics of Students Taking CDT Biology

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	1	2	38	7,907	8,118	1,394	248	17,708
Female (Pct)	33.33%	50.00%	52.78%	51.26%	47.30%	46.48%	47.69%	48.93%
Male (N)	2	2	34	7,519	9,045	1,605	272	18,479
Male (Pct)	66.67%	50.00%	47.22%	48.74%	52.70%	53.52%	52.31%	51.07%
American Indian or Alaskan Native (N)	0	0	0	82	114	70	36	302
American Indian or Alaskan Native (Pct)	0.00%	0.00%	0.00%	0.53%	0.66%	2.33%	6.92%	0.83%
Black/African American non-Hispanic (N)	1	0	1	1,287	2,086	411	81	3,867
Black/African American non-Hispanic (Pct)	33.33%	0.00%	1.39%	8.34%	12.15%	13.70%	15.58%	10.69%
Hispanic (N)	0	0	2	1,642	2,172	567	119	4,502
Hispanic (Pct)	0.00%	0.00%	2.78%	10.64%	12.66%	18.91%	22.88%	12.44%
White/Caucasian non-Hispanic (N)	2	4	58	11,302	11,625	1,589	262	24,842
White/Caucasian non-Hispanic (Pct)	66.67%	100.00%	80.56%	73.27%	67.73%	52.98%	50.38%	68.65%
Multi-Racial non-Hispanic (N)	0	0	7	541	752	260	12	1,572
Multi-Racial non-Hispanic (Pct)	0.00%	0.00%	9.72%	3.51%	4.38%	8.67%	2.31%	4.34%
Asian non-Hispanic (N)	0	0	4	561	403	99	9	1,076
Asian non-Hispanic (Pct)	0.00%	0.00%	5.56%	3.64%	2.35%	3.30%	1.73%	2.97%
Native Hawaiian or Pacific Islander (N)	0	0	0	11	11	3	1	26
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.00%	0.00%	0.07%	0.06%	0.10%	0.19%	0.07%
IEP (N)	0	0	0	19	44	8	0	71
IEP (Pct)	0.00%	0.00%	0.00%	0.12%	0.26%	0.27%	0.00%	0.20%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	0	0	0	24	49	5	0	78
Economically disadvantaged (Pct)	0.00%	0.00%	0.00%	0.16%	0.29%	0.17%	0.00%	0.22%
Number of students	3	4	72	15,426	17,163	2,999	520	36,187

Table 15–14. Demographic Characteristics of Students Taking CDT Chemistry

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	0	33	1	203	641	503	56	1,437
Female (Pct)	N/A	50.00%	33.33%	48.80%	55.16%	51.07%	53.85%	52.52%
Male (N)	0	33	2	213	521	482	48	1,299
Male (Pct)	N/A	50.00%	66.67%	51.20%	44.84%	48.93%	46.15%	47.48%
American Indian or Alaskan Native (N)	0	0	0	0	0	3	0	3
American Indian or Alaskan Native (Pct)	N/A	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%	0.11%
Black/African American non-Hispanic (N)	0	0	0	5	87	124	28	244
Black/African American non-Hispanic (Pct)	N/A	0.00%	0.00%	1.20%	7.49%	12.59%	26.92%	8.92%
Hispanic (N)	0	2	0	12	152	228	36	430
Hispanic (Pct)	N/A	3.03%	0.00%	2.88%	13.08%	23.15%	34.62%	15.72%
White/Caucasian non-Hispanic (N)	0	64	3	362	806	542	35	1,812
White/Caucasian non-Hispanic (Pct)	N/A	96.97%	100.00%	87.02%	69.36%	55.03%	33.65%	66.23%
Multi-Racial non-Hispanic (N)	0	0	0	16	45	48	1	110
Multi-Racial non-Hispanic (Pct)	N/A	0.00%	0.00%	3.85%	3.87%	4.87%	0.96%	4.02%
Asian non-Hispanic (N)	0	0	0	20	71	40	4	135
Asian non-Hispanic (Pct)	N/A	0.00%	0.00%	4.81%	6.11%	4.06%	3.85%	4.93%
Native Hawaiian or Pacific Islander (N)	0	0	0	1	1	0	0	2
Native Hawaiian or Pacific Islander (Pct)	N/A	0.00%	0.00%	0.24%	0.09%	0.00%	0.00%	0.07%
IEP (N)	0	0	0	0	0	10	4	14
IEP (Pct)	N/A	0.00%	0.00%	0.00%	0.00%	1.02%	3.85%	0.51%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	N/A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	0	1	1	0	0	7	1	10
Economically disadvantaged (Pct)	N/A	1.52%	33.33%	0.00%	0.00%	0.71%	0.96%	0.37%
Number of students	0	66	3	416	1,162	985	104	2,736

Table 15–15. Demographic Characteristics of Students Taking CDT Writing Grades 3–5

Demographic or Educational Characteristic	Gr. 3	Gr. 4	Gr. 5	Total
Female (N)	1,189	1,232	1,661	4,082
Female (Pct)	49.98%	50.33%	49.95%	50.07%
Male (N)	1,190	1,216	1,664	4,070
Male (Pct)	50.02%	49.67%	50.05%	49.93%
American Indian or Alaskan Native (N)	3	8	6	17
American Indian or Alaskan Native (Pct)	0.13%	0.33%	0.18%	0.21%
Black/African American non-Hispanic (N)	349	341	356	1,046
Black/African American non-Hispanic (Pct)	14.67%	13.93%	10.71%	12.83%
Hispanic (N)	367	366	401	1,134
Hispanic (Pct)	15.43%	14.95%	12.06%	13.91%
White/Caucasian non-Hispanic (N)	1,484	1,543	2,279	5,306
White/Caucasian non-Hispanic (Pct)	62.38%	63.03%	68.54%	65.09%
Multi-Racial non-Hispanic (N)	142	143	163	448
Multi-Racial non-Hispanic (Pct)	5.97%	5.84%	4.90%	5.50%
Asian non-Hispanic (N)	33	44	120	197
Asian non-Hispanic (Pct)	1.39%	1.80%	3.61%	2.42%
Native Hawaiian or Pacific Islander (N)	1	3	0	4
Native Hawaiian or Pacific Islander (Pct)	0.04%	0.12%	0.00%	0.05%
IEP (N)	5	10	6	21
IEP (Pct)	0.21%	0.41%	0.18%	0.26%
Migrant student (N)	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	10	5	12	27
Economically disadvantaged (Pct)	0.42%	0.20%	0.36%	0.33%
Number of students	2,379	2,448	3,325	8,152

Table 15–16. Demographic Characteristics of Students Taking CDT Writing/Eng Comp Grades 6–HS

Demographic or Educational Characteristic	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10	Gr. 11	Gr. 12	Total
Female (N)	2,121	2,614	2,831	862	900	425	106	9,859
Female (Pct)	49.87%	50.25%	48.60%	53.08%	49.21%	48.35%	55.50%	49.79%
Male (N)	2,132	2,588	2,994	762	929	454	85	9,944
Male (Pct)	50.13%	49.75%	51.40%	46.92%	50.79%	51.65%	44.50%	50.21%
American Indian or Alaskan Native (N)	7	6	9	1	5	3	0	31
American Indian or Alaskan Native (Pct)	0.16%	0.12%	0.15%	0.06%	0.27%	0.34%	0.00%	0.16%
Black/African American non-Hispanic (N)	431	474	509	188	233	213	82	2,130
Black/African American non-Hispanic (Pct)	10.13%	9.11%	8.74%	11.58%	12.74%	24.23%	42.93%	10.76%
Hispanic (N)	604	483	587	298	178	177	33	2,360
Hispanic (Pct)	14.20%	9.28%	10.08%	18.35%	9.73%	20.14%	17.28%	11.92%
White/Caucasian non-Hispanic (N)	2,912	3,852	4,334	1,010	1,305	423	70	13,906
White/Caucasian non-Hispanic (Pct)	68.47%	74.05%	74.40%	62.19%	71.35%	48.12%	36.65%	70.22%
Multi-Racial non-Hispanic (N)	202	236	230	71	44	32	5	820
Multi-Racial non-Hispanic (Pct)	4.75%	4.54%	3.95%	4.37%	2.41%	3.64%	2.62%	4.14%
Asian non-Hispanic (N)	97	144	116	56	62	30	1	506
Asian non-Hispanic (Pct)	2.28%	2.77%	1.99%	3.45%	3.39%	3.41%	0.52%	2.56%
Native Hawaiian or Pacific Islander (N)	0	7	40	0	2	1	0	50
Native Hawaiian or Pacific Islander (Pct)	0.00%	0.13%	0.69%	0.00%	0.11%	0.11%	0.00%	0.25%
IEP (N)	25	34	31	9	7	6	7	119
IEP (Pct)	0.59%	0.65%	0.53%	0.55%	0.38%	0.68%	3.66%	0.60%
Migrant student (N)	0	0	0	0	0	0	0	0
Migrant student (Pct)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Economically disadvantaged (N)	15	4	15	13	4	9	8	68
Economically disadvantaged (Pct)	0.35%	0.08%	0.26%	0.80%	0.22%	1.02%	4.19%	0.34%
Number of students	4,253	5,202	5,825	1,624	1,829	879	191	19,803

SUMMARY STATISTICS—TEST LENGTH

The analyses from here until the section titled "Multiple Administrations of the Same CDT Test" include all records in the full CDT operational assessments. When a student took CDT Math Grades 6–HS twice, for example, both records were used in the analyses.

As noted in Chapter Thirteen, full CDT tests have either four or five diagnostic categories. On tests with five diagnostic categories (Reading Grades 3–5, Reading/Lit Grades 6–HS, Writing Grades 3–5, and Writing/Eng Comp Grades 6–HS), students take between 10 and 12 operational items per diagnostic category for a total test of 50 to 60 operational items. On tests with four diagnostic categories (Math Grades 3–5, Math Grades 6–HS, Algebra I, Geometry, Algebra II, Science Grades 3–5, Science Grades 6–HS, Biology, and Chemistry), students take between 12 and 15 operational items per diagnostic category for a total test of 48 to 60 operational items.

Tables 15–17a and 15–17b show the summary statistics for the test length for each assessment. Summary statistics are based on the number of items presented to the student and include minimum, maximum, quartiles 1 and 3, mean, and median.

Table 15–17a. Summary Statistics for Full CDT Test Length (Number of Operational Items Administered)

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math Grades 3–5	76,019	48	50	51	51.63	53	60
Math Grades 6–HS	97,557	48	50	52	52.13	54	60
Algebra I	42,725	48	50	52	52.74	55	60
Geometry	5,131	48	50	52	52.31	54	60
Algebra II	6,023	48	51	52	52.71	54	60
Reading Grades 3–5	64,853	50	54	55	55.15	57	60
Reading/Lit Grades 6–HS	139,055	50	54	55	55.42	57	60
Science Grades 3–5	20,678	48	50	51	51.44	53	60
Science Grades 6–HS	51,989	48	50	51	52.21	54	60
Biology	46,477	48	50	52	52.31	54	60
Chemistry	3,070	48	50	52	52.55	55	60
Writing Grades 3–5	8,478	50	53	55	54.76	56	60
Writing/Eng Comp Gr 6–HS	21,548	50	54	55	55.61	57	60

The minimum number of items was quite similar, ranging from 48 to 50. The mean and median were higher for tests in the reading and writing content areas, which have five diagnostic categories. The maximum number of items administered was fixed at 60 for all CDT tests.

Table 15–17b. Summary Statistics for Diagnostic Category CDT Test Length (Number of Operational Items Administered)

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math Grades 3–5	Numbers and Operations	27,175	15	15	16	16.29	17	18
Math Grades 3–5	Algebraic Concepts	10,224	15	15	16	16.33	18	18
Math Grades 3–5	Geometry	5,577	15	15	16	16.31	17	18
Math Grades 3–5	Measurement, Data, and Probability	5,761	15	15	16	16.29	17	18
Math Grades 6–HS	Numbers and Operations	19,689	15	15	16	16.34	17	18
Math Grades 6–HS	Algebraic Concepts	14,227	15	15	16	16.44	18	18
Math Grades 6–HS	Geometry	5,034	15	15	16	16.38	17	18
Math Grades 6–HS	Measurement, Data, and Probability	4,024	15	16	17	16.66	18	18
Algebra I	Operations with Real Numbers and Expressions	6,724	15	16	16	16.57	18	18
Algebra I	Linear Equations & Inequalities	9,425	15	15	16	16.37	18	18
Algebra I	Functions & Coordinate Geometry	5,277	15	15	16	16.40	18	18
Algebra I	Data Analysis	2,618	15	16	16	16.58	18	18
Geometry	Geometric Properties	601	15	15	16	16.33	17	18
Geometry	Congruence, Similarity, and Proofs	157	15	16	18	17.13	18	18
Geometry	Coordinate Geometry and Right Triangles	544	15	16	16	16.57	18	18
Geometry	Measurement	483	15	16	16	16.54	18	18
Algebra II	Operations with Complex Numbers	279	15	16	17	16.92	18	18
Algebra II	Non-Linear Expressions & Equations	313	15	16	16	16.56	18	18
Algebra II	Functions	878	15	16	16	16.56	18	18
Algebra II	Data Analysis	45	15	16	17	16.91	18	18

Table 15–17b (continued). Summary Statistics for Diagnostic Category CDT Test Length (Number of Operational Items Administered)

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Reading Grades 3–5	Informational Text	19,090	30	32	33	32.89	34	36
Reading Grades 3–5	Literature Text	17,925	30	32	33	32.77	34	36
Reading/Lit Grades 6–HS	Informational Text	24,999	30	32	33	32.82	34	36
Reading/Lit Grades 6–HS	Literature Text	31,260	30	32	33	32.98	34	36
Science Grades 3–5	The Nature of Science	1,705	15	15	16	16.33	18	18
Science Grades 3–5	Biological Sciences	1,011	15	15	16	16.37	18	18
Science Grades 3–5	Physical Sciences	1,160	15	15	16	16.40	18	18
Science Grades 3–5	Earth and Space Sciences	919	15	15	16	16.31	18	18
Science Grades 6–HS	The Nature of Science	14,133	15	15	16	16.19	17	18
Science Grades 6–HS	Biological Sciences	7,348	15	15	16	16.30	17	18
Science Grades 6–HS	Physical Sciences	5,500	15	15	16	16.45	18	18
Science Grades 6–HS	Earth and Space Sciences	7,587	15	15	16	16.30	17	18
Biology	Basic Biological Principles/ Chemical Basis for Life	9,811	15	15	16	16.37	17	18
Biology	Bioenergetics/ Homeostasis and Transport	6,252	15	15	16	16.34	17	18
Biology	Cell Growth and Reproduction/ Genetics	4,302	15	15	16	16.41	18	18
Biology	Theory of Evolution/ Ecology	4,217	15	15	16	16.36	18	18
Chemistry	Properties and Classification of Matter	262	15	15	16	16.46	18	18
Chemistry	Atomic Structure and The Periodic Table	1,097	15	16	16	16.62	18	18

Table 15–17b (continued). Summary Statistics for Diagnostic Category CDT Test Length (Number of Operational Items Administered)

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Chemistry	The Mole and Chemical Bonding	188	15	15	16	16.31	18	18
Chemistry	Chemical Relationships and Reactions	41	15	15	16	16.29	18	18
Writing Grades 3–5	Quality of Writing: Focus and Organization	1,626	15	15	16	16.38	18	18
Writing Grades 3–5	Quality of Writing: Content and Style	590	15	15	16	16.25	17	18
Writing Grades 3–5	Quality of Writing: Editing	472	15	15	16	16.21	17	18
Writing Grades 3–5	Conventions: Punctuation, Capitalization, and Spelling	3,976	15	15	16	16.32	18	18
Writing Grades 3–5	Conventions: Grammar and Sentence Formation	5,201	15	15	16	16.38	18	18
Writing/Eng Comp Gr 6–HS	Quality of Writing: Focus and Organization	1,578	15	16	16	16.53	18	18
Writing/Eng Comp Gr 6–HS	Quality of Writing: Content and Style	1,062	15	15	16	16.53	18	18
Writing/Eng Comp Gr 6–HS	Quality of Writing: Editing	473	15	15	16	16.45	18	18
Writing/Eng Comp Gr 6–HS	Conventions: Punctuation, Capitalization, and Spelling	6,242	15	16	16	16.55	18	18
Writing/Eng Comp Gr 6–HS	Conventions: Grammar and Sentence Formation	5,828	15	16	16	16.52	18	18

All diagnostic category CDTs in the math, science and writing content areas focus on a single diagnostic category. Tests range from 15 to 18 items. Diagnostic category CDTs in the reading content area focus on a single text type with three diagnostic categories. Tests range from 30 to 36 items.

SUMMARY STATISTICS—SCALE SCORES AND CONDITIONAL STANDARD ERRORS

Tables 15–18a and 15–18b show the summary statistics for the scale scores. Tests with multiple benchmark cuts are broken down to match the grade level of the cuts. Tests that are course-specific are not broken down.

Table 15–18a. Summary Statistics for Scale Score Based on Full CDT

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math – G3	23,191	208	592	712	705.47	817	1414
Math – G4	24,069	287	694	800	791.62	895	1446
Math – G5	28,759	308	755	860	848.86	952	1582
Math – G6	34,688	400	813	924	914.65	1022	1666
Math – G7	34,947	383	845	957	943.92	1049	1714
Math – G8	27,468	423	855	980	963.03	1078	1675
Math – HS	454	519	766	887	874.82	979	1308
Algebra I	42,725	474	891	1015	994.29	1106	1642
Geometry	5,131	497	992	1095	1077.90	1174	1583
Algebra II	6,023	597	1088	1178	1174.28	1274	2000
Reading – G3	19,838	318	594	696	715.19	824	1276
Reading – G4	20,691	352	653	783	785.37	909	1337
Reading – G5	24,324	408	729	868	855.99	981	1368
Reading – G6	25,753	424	772	902	890.49	1005	1442
Reading – G7	26,355	366	786	929	912.24	1033	1403
Reading – G8	24,777	368	799	944	928.82	1053	1425
Literature	62,170	457	875	1017	991.15	1113	1528
Science – G3	1,557	200	591	726	695.14	822	1205
Science – G4	13,182	200	633	761	732.93	850	1275
Science – G5	5,939	277	642	782	753.33	877	1241
Science – G6	10,228	400	689	825	807.07	925	1300
Science – G7	15,341	410	704	847	826.56	948	1382
Science – G8	25,927	422	754	895	867.65	985	1352
Science – HS	493	492	685	819	810.34	923	1189
Biology	46,477	430	836	955	938.77	1042	1567
Chemistry	3,070	591	907	988	982.74	1056	1388
Writing – G3	2,677	247	577	735	711.92	847	1182
Writing – G4	2,177	273	647	803	770.89	898	1245
Writing – G5	3,624	329	716	858	824.11	950	1241
Writing – G6	4,985	413	749	895	867.14	992	1345
Writing – G7	6,262	398	783	931	898.68	1028	1369
Writing – G8	5,666	393	753	922	890.28	1030	1393
English Composition	4,635	467	829	982	943.37	1070	1391

Table 15–18b. Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math – G3	Numbers and Operations	8,236	200	594	712	709.70	822	1437
Math – G3	Algebraic Concepts	3,597	200	579	720	712.98	838	1293
Math – G3	Geometry	1,888	200	558	669	661.05	759	1249
Math – G3	Measurement, Data, and Probability	1,988	200	527	678	670.28	805	1285
Math – G4	Numbers and Operations	8,482	200	704	817	808.91	916	1633
Math – G4	Algebraic Concepts	3,616	200	698	813	801.16	924	1297
Math – G4	Geometry	1,755	200	644	731	729.89	808	1241
Math – G4	Measurement, Data, and Probability	1,965	203	613	754	741.37	870	1365
Math – G5	Numbers and Operations	10,457	200	741	864	852.29	973	1657
Math – G5	Algebraic Concepts	3,011	217	728	841	827.27	950	1390
Math – G5	Geometry	1,934	229	688	783	793.66	905	1723
Math – G5	Measurement, Data, and Probability	1,808	200	673	796	787.47	908	1394
Math – G6	Numbers and Operations	8,320	200	789	907	901.85	1032	1692
Math – G6	Algebraic Concepts	4,197	247	808	939	916.78	1048	1629
Math – G6	Geometry	2,049	236	760	884	868.75	976	1430
Math – G6	Measurement, Data, and Probability	1,291	200	687	810	811.59	926	1309
Math – G7	Numbers and Operations	6,793	208	825	950	934.82	1065	1727
Math – G7	Algebraic Concepts	4,848	252	849	978	952.29	1074	1771
Math – G7	Geometry	1,629	302	825	935	930.08	1048	1493
Math – G7	Measurement, Data, and Probability	1,003	216	721	831	844.58	967	1426
Math – G8	Numbers and Operations	4,266	200	797	934	915.34	1064	1441
Math – G8	Algebraic Concepts	5,102	200	859	1003	981.82	1099	1683

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math – G8	Geometry	1,326	444	827	937	942.75	1058	1530
Math – G8	Measurement, Data, and Probability	1,699	243	747	878	876.43	1011	1538
Math – HS	Numbers and Operations	310	306	793	915	896.00	1032	1365
Math – HS	Algebraic Concepts	80	367	770	923	906.08	1055	1191
Math – HS	Geometry	30	618	766	877	867.33	952	1096
Math – HS	Measurement, Data, and Probability	31	520	721	811	818.65	884	1194
Algebra I	Operations with Real Numbers and Expressions	6,724	400	903	1033	1012.56	1139	1826
Algebra I	Linear Equations & Inequalities	9,425	412	919	1042	1027.75	1135	1808
Algebra I	Functions & Coordinate Geometry	5,277	400	931	1041	1024.55	1130	1778
Algebra I	Data Analysis	2,618	400	868	1000	975.86	1086	1842
Geometry	Geometric Properties	601	400	927	1044	1019.14	1122	1395
Geometry	Congruence, Similarity, and Proofs	157	400	887	1025	1043.05	1219	1791
Geometry	Coordinate Geometry and Right Triangles	544	523	987	1113	1103.48	1240	1788
Geometry	Measurement	483	434	957	1071	1050.05	1161	1659
Algebra II	Operations with Complex Numbers	279	823	1072	1173	1244.27	1439	1857
Algebra II	Non-Linear Expressions & Equations	313	433	1023	1129	1114.27	1237	1509
Algebra II	Functions	878	600	998	1103	1097.28	1198	1718
Algebra II	Data Analysis	45	423	954	1059	1053.44	1189	1725
Reading – G3	Informational Text	5,230	246	599	708	722.97	841	1233
Reading – G3	Literature Text	5,098	279	609	736	736.27	859	1392

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Reading – G4	Informational Text	6,409	308	660	799	792.76	921	1473
Reading – G4	Literature Text	5,675	200	663	806	794.69	922	1333
Reading – G5	Informational Text	7,451	242	692	852	837.76	978	1391
Reading – G5	Literature Text	7,152	325	745	889	870.42	1000	1496
Reading – G6	Informational Text	6,924	366	757	916	894.31	1024	1347
Reading – G6	Literature Text	7,784	259	793	923	905.38	1021	1490
Reading – G7	Informational Text	4,045	385	727	890	875.31	1016	1417
Reading – G7	Literature Text	6,498	393	779	923	906.92	1029	1370
Reading – G8	Informational Text	4,436	319	759	925	906.30	1049	1403
Reading – G8	Literature Text	5,962	382	802	952	935.31	1065	1417
Literature	Informational Text	9,594	377	845	1000	971.37	1102	1599
Literature	Literature Text	11,016	364	862	1015	986.88	1117	1562
Science – G3	The Nature of Science	0	0	0	0	0.00	0	0
Science – G3	Biological Sciences	65	511	691	785	794.63	901	1089
Science – G3	Physical Sciences	60	200	623	783	741.85	869	959
Science – G3	Earth and Space Sciences	52	304	605	679	687.42	786	1003
Science – G4	The Nature of Science	1,220	200	518	688	667.54	812	1258
Science – G4	Biological Sciences	796	200	482	659	647.03	800	1154
Science – G4	Physical Sciences	1,099	200	541	726	688.78	842	1343
Science – G4	Earth and Space Sciences	867	200	529	660	665.43	801	1162
Science – G5	The Nature of Science	485	319	788	875	865.28	965	1389
Science – G5	Biological Sciences	150	255	805	905	893.82	975	1456
Science – G5	Physical Sciences	1	1164	1164	1164	1164.00	1164	1164
Science – G5	Earth and Space Sciences	0	0	0	0	0.00	0	0

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Science – G6	The Nature of Science	3,375	305	806	904	885.29	982	1336
Science – G6	Biological Sciences	737	227	857	946	934.21	1035	1379
Science – G6	Physical Sciences	969	256	736	838	831.19	927	1216
Science – G6	Earth and Space Sciences	2,400	302	792	885	870.92	962	1324
Science – G7	The Nature of Science	4,231	281	804	914	884.68	992	1387
Science – G7	Biological Sciences	5,000	200	760	891	866.44	982	1562
Science – G7	Physical Sciences	1,092	200	701	820	811.30	927	1263
Science – G7	Earth and Space Sciences	2,636	347	731	853	841.02	956	1309
Science – G8	The Nature of Science	6,362	200	782	917	885.60	1006	1488
Science – G8	Biological Sciences	1,441	255	658	811	806.59	947	1320
Science – G8	Physical Sciences	3,390	200	760	890	869.06	990	1350
Science – G8	Earth and Space Sciences	2,506	313	724	856	838.21	957	1248
Science – HS	The Nature of Science	165	200	667	796	780.52	912	1119
Science – HS	Biological Sciences	170	200	655	790	786.76	927	1189
Science – HS	Physical Sciences	49	286	609	784	784.86	932	1234
Science – HS	Earth and Space Sciences	45	200	613	847	793.20	961	1086
Biology	Basic Biological Principles/ Chemical Basis for Life	9,811	400	878	984	976.84	1082	1780
Biology	Bioenergetics/ Homeostasis and Transport	6,252	400	888	994	993.42	1096	1724
Biology	Cell Growth and Reproduction/ Genetics	4,302	400	878	998	985.99	1094	1600
Biology	Theory of Evolution/ Ecology	4,217	400	828	973	947.31	1076	1490

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Chemistry	Properties and Classification of Matter	262	400	863	980	945.98	1064	1263
Chemistry	Atomic Structure and The Periodic Table	1,097	503	922	994	997.31	1081	1580
Chemistry	The Mole and Chemical Bonding	188	612	928	1035	1024.00	1121	1407
Chemistry	Chemical Relationships and Reactions	41	691	922	986	985.29	1082	1231
Writing – G3	Quality of Writing: Focus and Organization	561	200	505	690	666.06	829	1243
Writing – G3	Quality of Writing: Content and Style	137	207	569	726	678.57	808	1022
Writing – G3	Quality of Writing: Editing	169	200	568	705	700.21	854	1231
Writing – G3	Conventions: Punctuation, Capitalization, and Spelling	1,344	200	507	615	645.41	779	1229
Writing – G3	Conventions: Grammar and Sentence Formation	1,415	200	474	653	647.68	810	1179
Writing – G4	Quality of Writing: Focus and Organization	658	200	580	746	723.10	882	1174
Writing – G4	Quality of Writing: Content and Style	272	273	675	804	769.99	900	1157
Writing – G4	Quality of Writing: Editing	222	200	688	803	780.70	911	1154
Writing – G4	Conventions: Punctuation, Capitalization, and Spelling	1,310	200	557	684	695.85	826	1212
Writing – G4	Conventions: Grammar and Sentence Formation	1,649	200	581	729	710.00	865	1282

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Writing – G5	Quality of Writing: Focus and Organization	407	200	588	771	757.40	918	1234
Writing – G5	Quality of Writing: Content and Style	181	277	691	837	805.35	932	1227
Writing – G5	Quality of Writing: Editing	81	280	625	789	748.79	892	1067
Writing – G5	Conventions: Punctuation, Capitalization, and Spelling	1,322	200	591	699	720.04	840	1236
Writing – G5	Conventions: Grammar and Sentence Formation	2,137	200	695	837	809.02	954	1691
Writing – G6	Quality of Writing: Focus and Organization	322	200	669	847	819.10	962	1191
Writing – G6	Quality of Writing: Content and Style	0	0	0	0	0.00	0	0
Writing – G6	Quality of Writing: Editing	0	0	0	0	0.00	0	0
Writing – G6	Conventions: Punctuation, Capitalization, and Spelling	1,575	200	656	784	784.03	911	1346
Writing – G6	Conventions: Grammar and Sentence Formation	1,867	200	674	813	791.76	924	1243
Writing – G7	Quality of Writing: Focus and Organization	459	309	730	890	857.89	992	1416
Writing – G7	Quality of Writing: Content and Style	224	383	857	957	947.81	1066	1360
Writing – G7	Quality of Writing: Editing	96	402	736	885	867.28	991	1174
Writing – G7	Conventions: Punctuation, Capitalization, and Spelling	1,276	200	668	806	808.68	940	1377

Table 15–18b (continued). Summary Statistics for Scale Score Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Writing – G7	Conventions: Grammar and Sentence Formation	1,558	200	690	836	811.99	945	1229
Writing – G8	Quality of Writing: Focus and Organization	419	205	685	833	825.72	967	1217
Writing – G8	Quality of Writing: Content and Style	562	216	807	957	929.43	1066	1489
Writing – G8	Quality of Writing: Editing	317	338	730	893	862.56	1000	1202
Writing – G8	Conventions: Punctuation, Capitalization, and Spelling	2,127	200	730	890	873.23	1008	1469
Writing – G8	Conventions: Grammar and Sentence Formation	2,066	200	742	903	868.23	1009	1562
English Composition	Quality of Writing: Focus and Organization	378	200	701	896	867.91	1046	1364
English Composition	Quality of Writing: Content and Style	276	206	666	784	801.02	936	1262
English Composition	Quality of Writing: Editing	60	291	704	826	832.17	992	1210
English Composition	Conventions: Punctuation, Capitalization, and Spelling	1,264	200	803	926	912.75	1040	1382
English Composition	Conventions: Grammar and Sentence Formation	337	200	728	869	839.12	968	1248

Tables 15–19a and 15–19b show the summary statistics for the conditional standard errors of measurement (CSEMs) in the scale score metric. The final column in the table shows the theoretical minimum CSEM that is possible for a test length equal to the mean number of points. This is the standard error if the student's ability is known and there are sufficient items in the operational pool to administer where the item's difficulty is equal to the known ability and the test constraints are met.

Table 15–19a. Summary Statistics for Conditional Standard Errors Based on Full CDT

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Math – G3	23,191	35	37	38	38.18	39	56	36.31
Math – G4	24,069	35	37	38	38.15	39	53	36.31
Math – G5	28,759	34	37	38	38.10	39	75	36.67
Math – G6	34,688	34	37	37	37.27	38	90	34.64
Math – G7	34,947	34	37	37	37.25	38	91	34.64
Math – G8	27,468	34	37	37	37.39	38	75	34.64
Math – HS	454	35	37	37	37.78	38	48	33.99
Algebra I	42,725	34	37	37	37.73	38	66	34.31
Geometry	5,131	35	37	37	37.47	38	55	34.64
Algebra II	6,023	34	37	37	37.49	38	229	34.31
Reading – G3	19,838	38	41	42	43.09	44	69	39.32
Reading – G4	20,691	38	41	42	42.64	44	76	39.32
Reading – G5	24,324	37	41	42	42.38	43	76	39.66
Reading – G6	25,753	38	40	41	41.91	43	87	37.84
Reading – G7	26,355	38	40	41	42.12	43	78	37.84
Reading – G8	24,777	38	41	42	42.44	44	87	37.84
Literature	62,170	37	41	42	42.72	44	104	37.84
Science – G3	1,557	38	40	40	40.54	41	52	38.63
Science – G4	13,182	37	40	40	40.46	41	61	39.01
Science – G5	5,939	38	40	40	40.43	41	55	38.63
Science – G6	10,228	37	39	39	39.60	40	59	36.85
Science – G7	15,341	37	39	39	39.64	40	58	36.85
Science – G8	25,927	37	39	39	39.61	40	58	36.85
Science – HS	493	38	39	39	40.07	40	49	36.16
Biology	46,477	36	39	39	39.77	40	71	36.85
Chemistry	3,070	37	39	39	40.00	40	62	36.50
Writing – G3	2,677	36	39	39	39.45	40	55	37.60
Writing – G4	2,177	36	39	39	39.39	40	59	37.60
Writing – G5	3,624	36	39	39	39.29	40	53	37.60
Writing – G6	4,985	36	38	38	38.45	39	55	35.55
Writing – G7	6,262	36	38	38	38.42	39	56	35.55
Writing – G8	5,666	36	38	38	38.61	39	64	35.55
English Composition	4,635	36	38	38	38.47	39	53	35.55

Table 15–19b. Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Math – G3	Numbers and Operations	8,236	64	66	67	68.17	68	232	65.47
Math – G3	Algebraic Concepts	3,597	65	66	68	68.41	68	231	65.47
Math – G3	Geometry	1,888	65	67	68	68.39	68	134	65.47
Math – G3	Measurement, Data, and Probability	1,988	65	67	68	68.82	68	133	65.47
Math – G4	Numbers and Operations	8,482	65	66	67	68.14	68	232	65.47
Math – G4	Algebraic Concepts	3,616	65	66	68	68.15	68	130	65.47
Math – G4	Geometry	1,755	65	66	67	67.99	68	141	65.47
Math – G4	Measurement, Data, and Probability	1,965	65	66	68	68.38	68	133	65.47
Math – G5	Numbers and Operations	10,457	65	66	67	68.05	68	231	65.47
Math – G5	Algebraic Concepts	3,011	65	66	67	67.98	68	135	65.47
Math – G5	Geometry	1,934	65	66	67	67.83	68	230	65.47
Math – G5	Measurement, Data, and Probability	1,808	65	66	67	68.21	68	231	65.47
Math – G6	Numbers and Operations	8,320	63	65	65	66.03	66	231	62.45
Math – G6	Algebraic Concepts	4,197	63	65	65	66.46	66	132	62.45
Math – G6	Geometry	2,049	63	65	65	66.27	66	231	62.45
Math – G6	Measurement, Data, and Probability	1,291	63	65	66	67.69	66	231	60.59
Math – G7	Numbers and Operations	6,793	63	65	65	66.26	66	232	62.45
Math – G7	Algebraic Concepts	4,848	63	65	65	66.75	66	232	62.45
Math – G7	Geometry	1,629	63	65	65	66.05	66	133	62.45
Math – G7	Measurement, Data, and Probability	1,003	63	65	65	67.72	66	231	60.59

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Math – G8	Numbers and Operations	4,266	63	65	65	66.47	66	232	62.45
Math – G8	Algebraic Concepts	5,102	63	65	65	66.66	66	233	62.45
Math – G8	Geometry	1,326	63	65	65	66.29	66	132	62.45
Math – G8	Measurement, Data, and Probability	1,699	63	65	66	67.89	67	232	60.59
Math – HS	Numbers and Operations	310	63	65	65	67.27	66	132	60.59
Math – HS	Algebraic Concepts	80	63	65	65	68.07	68	133	60.59
Math – HS	Geometry	30	63	65	66	65.66	66	69	62.45
Math – HS	Measurement, Data, and Probability	31	64	65	66	66.16	66	82	62.45
Algebra I	Operations with Real Numbers and Expressions	6,724	63	65	65	66.90	66	231	60.59
Algebra I	Linear Equations & Inequalities	9,425	63	65	65	67.29	66	231	62.45
Algebra I	Functions & Coordinate Geometry	5,277	63	65	65	66.82	66	234	62.45
Algebra I	Data Analysis	2,618	63	65	65	67.51	66	233	60.59
Geometry	Geometric Properties	601	63	65	65	66.28	66	149	62.45
Geometry	Congruence, Similarity, and Proofs	157	63	65	66	70.96	69	230	60.59
Geometry	Coordinate Geometry and Right Triangles	544	63	65	65	67.71	66	230	60.59
Geometry	Measurement	483	63	65	65	66.70	66	131	60.59
Algebra II	Operations with Complex Numbers	279	63	65	66	80.46	74	231	60.59
Algebra II	Non-Linear Expressions & Equations	313	63	65	65	66.73	66	134	60.59
Algebra II	Functions	878	63	65	65	67.12	66	130	60.59
Algebra II	Data Analysis	45	64	65	66	71.37	68	134	60.59

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Reading – G3	Informational Text	5,230	48	52	54	55.39	57	108	50.62
Reading – G3	Literature Text	5,098	48	53	54	55.51	57	146	49.91
Reading – G4	Informational Text	6,409	48	52	54	55.00	57	146	50.62
Reading – G4	Literature Text	5,675	48	52	54	55.06	57	146	49.91
Reading – G5	Informational Text	7,451	47	52	54	55.22	57	146	50.62
Reading – G5	Literature Text	7,152	47	52	54	55.02	57	147	50.62
Reading – G6	Informational Text	6,924	48	52	53	54.35	56	91	48.29
Reading – G6	Literature Text	7,784	46	51	53	53.97	55	148	48.29
Reading – G7	Informational Text	4,045	47	52	54	55.35	57	92	48.29
Reading – G7	Literature Text	6,498	47	51	53	54.21	56	90	48.29
Reading – G8	Informational Text	4,436	48	53	54	55.71	57	109	48.99
Reading – G8	Literature Text	5,962	47	52	54	55.05	57	110	48.29
Literature	Informational Text	9,594	48	53	55	56.15	58	146	49.73
Literature	Literature Text	11,016	47	53	54	55.91	58	147	48.99
Science – G3	The Nature of Science	0	0	0	0	0.00	0	0	N/A
Science – G3	Biological Sciences	65	69	70	72	71.65	72	77	69.64
Science – G3	Physical Sciences	60	69	71	72	72.34	73	104	69.64
Science – G3	Earth and Space Sciences	52	69	71	71	71.74	72	81	69.64
Science – G4	The Nature of Science	1,220	66	71	72	73.73	73	168	69.64
Science – G4	Biological Sciences	796	68	71	72	74.10	73	247	69.64
Science – G4	Physical Sciences	1,099	67	71	72	73.34	73	246	69.64
Science – G4	Earth and Space Sciences	867	64	71	72	73.03	73	144	69.64

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Science – G5	The Nature of Science	485	64	71	72	72.37	73	139	69.64
Science – G5	Biological Sciences	150	69	71	72	73.39	73	140	69.64
Science – G5	Physical Sciences	1	74	74	74	73.61	74	74	65.66
Science – G5	Earth and Space Sciences	0	0	0	0	0.00	0	0	N/A
Science – G6	The Nature of Science	3,375	64	69	69	69.69	70	138	66.44
Science – G6	Biological Sciences	737	64	69	69	69.74	70	144	66.44
Science – G6	Physical Sciences	969	66	69	69	70.31	70	141	66.44
Science – G6	Earth and Space Sciences	2,400	66	69	69	69.92	70	143	66.44
Science – G7	The Nature of Science	4,231	66	69	69	70.12	70	138	66.44
Science – G7	Biological Sciences	5,000	64	69	69	70.79	70	253	66.44
Science – G7	Physical Sciences	1,092	66	69	70	72.87	71	248	64.45
Science – G7	Earth and Space Sciences	2,636	66	69	69	70.94	70	139	66.44
Science – G8	The Nature of Science	6,362	65	69	69	70.48	70	246	66.44
Science – G8	Biological Sciences	1,441	66	69	70	72.02	71	150	64.45
Science – G8	Physical Sciences	3,390	67	69	69	71.25	70	248	66.44
Science – G8	Earth and Space Sciences	2,506	66	69	69	71.08	70	143	66.44

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Science – HS	The Nature of Science	165	67	69	70	72.54	71	246	64.45
Science – HS	Biological Sciences	170	66	69	70	73.62	71	246	64.45
Science – HS	Physical Sciences	49	68	69	71	76.09	79	114	64.45
Science – HS	Earth and Space Sciences	45	67	69	70	77.20	77	247	64.45
Biology	Basic Biological Principles/ Chemical Basis for Life	9,811	64	69	69	70.75	70	248	66.44
Biology	Bioenergetics/ Homeostasis and Transport	6,252	65	69	69	70.86	70	246	66.44
Biology	Cell Growth and Reproduction/ Genetics	4,302	65	69	69	71.76	70	246	66.44
Biology	Theory of Evolution/ Ecology	4,217	65	69	69	71.20	70	246	66.44
Chemistry	Properties and Classification of Matter	262	67	69	70	71.93	70	247	66.44
Chemistry	Atomic Structure and The Periodic Table	1,097	67	69	70	72.51	71	245	64.45
Chemistry	The Mole and Chemical Bonding	188	67	69	69	71.30	70	138	66.44
Chemistry	Chemical Relationships and Reactions	41	67	69	69	71.38	71	103	66.44

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Writing – G3	Quality of Writing: Focus and Organization	561	69	71	72	73.66	73	139	69.72
Writing – G3	Quality of Writing: Content and Style	137	69	71	72	73.20	73	102	69.72
Writing – G3	Quality of Writing: Editing	169	69	71	72	73.36	73	138	69.72
Writing – G3	Conventions: Punctuation, Capitalization, and Spelling	1,344	69	71	72	75.52	73	245	69.72
Writing – G3	Conventions: Grammar and Sentence Formation	1,415	69	71	72	73.86	73	246	69.72
Writing – G4	Quality of Writing: Focus and Organization	658	69	71	72	73.66	73	140	69.72
Writing – G4	Quality of Writing: Content and Style	272	69	71	72	73.27	73	105	69.72
Writing – G4	Quality of Writing: Editing	222	69	71	72	73.12	72	246	69.72
Writing – G4	Conventions: Punctuation, Capitalization, and Spelling	1,310	69	71	72	74.28	73	245	69.72
Writing – G4	Conventions: Grammar and Sentence Formation	1,649	69	71	72	73.58	73	246	69.72

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Writing – G5	Quality of Writing: Focus and Organization	407	69	71	72	73.56	73	141	69.72
Writing – G5	Quality of Writing: Content and Style	181	69	71	72	72.96	73	107	69.72
Writing – G5	Quality of Writing: Editing	81	69	71	72	72.82	73	101	69.72
Writing – G5	Conventions: Punctuation, Capitalization, and Spelling	1,322	69	71	72	73.93	73	246	69.72
Writing – G5	Conventions: Grammar and Sentence Formation	2,137	69	71	72	72.88	73	247	69.72
Writing – G6	Quality of Writing: Focus and Organization	322	67	69	70	72.12	71	246	66.51
Writing – G6	Quality of Writing: Content and Style	0	0	0	0	0.00	0	0	N/A
Writing – G6	Quality of Writing: Editing	0	0	0	0	0.00	0	0	N/A
Writing – G6	Conventions: Punctuation, Capitalization, and Spelling	1,575	67	69	70	72.87	72	245	64.52
Writing – G6	Conventions: Grammar and Sentence Formation	1,867	67	69	70	71.77	71	247	64.52

Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Writing – G7	Quality of Writing: Focus and Organization	459	67	69	70	71.92	71	139	66.51
Writing – G7	Quality of Writing: Content and Style	224	68	69	70	71.07	70	139	66.51
Writing – G7	Quality of Writing: Editing	96	68	69	70	70.36	70	101	66.51
Writing – G7	Conventions: Punctuation, Capitalization, and Spelling	1,276	67	69	70	72.92	71	245	64.52
Writing – G7	Conventions: Grammar and Sentence Formation	1,558	67	69	70	71.86	71	248	64.52
Writing – G8	Quality of Writing: Focus and Organization	419	67	69	70	71.68	71	246	64.52
Writing – G8	Quality of Writing: Content and Style	562	67	69	70	71.67	70	245	66.51
Writing – G8	Quality of Writing: Editing	317	67	69	69	71.74	70	139	66.51
Writing – G8	Conventions: Punctuation, Capitalization, and Spelling	2,127	67	69	70	71.97	70	246	66.51
Writing – G8	Conventions: Grammar and Sentence Formation	2,066	67	69	70	71.68	70	247	66.51

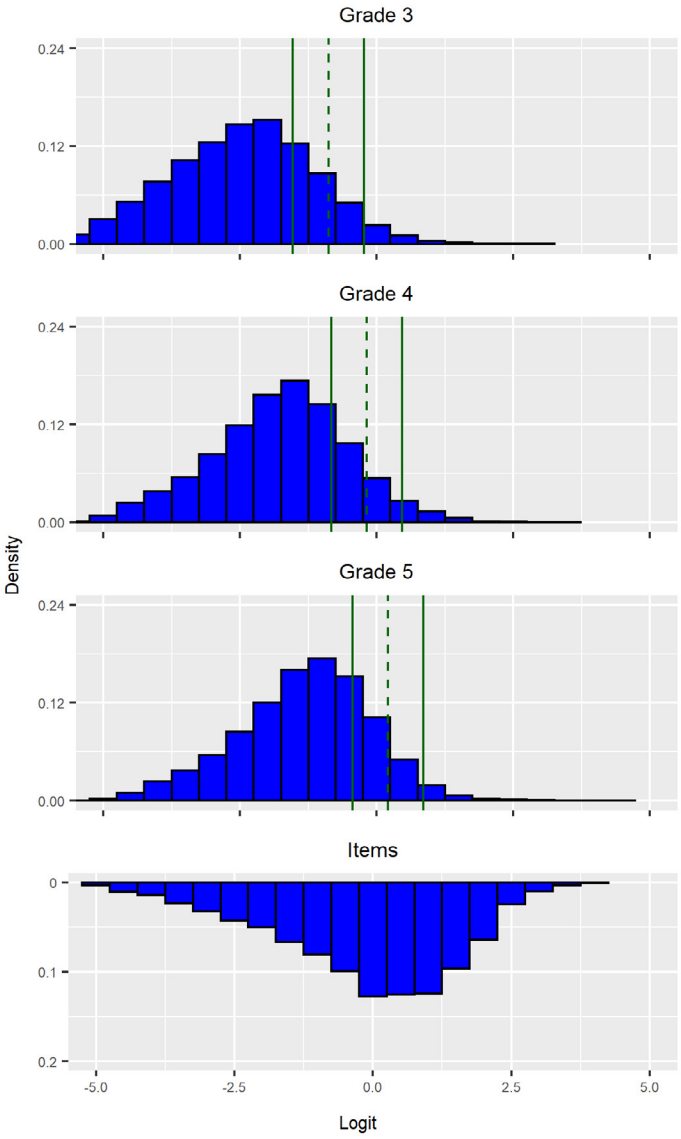
Table 15–19b (continued). Summary Statistics for Conditional Standard Errors Based on Diagnostic Category CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum	Theoretical Minimum
Eng. Comp.	Quality of Writing: Focus and Organization	378	67	69	70	73.60	72	247	64.52
Eng. Comp.	Quality of Writing: Content and Style	276	67	69	70	74.97	76	245	64.52
Eng. Comp.	Quality of Writing: Editing	60	67	69	70	73.80	71	140	64.52
Eng. Comp.	Conventions: Punctuation, Capitalization, and Spelling	1,264	67	69	70	71.51	70	247	66.51
Eng. Comp.	Conventions: Grammar and Sentence Formation	337	67	69	70	72.42	71	248	64.52

Values in the “Minimum” column that are less than the “Theoretical Minimum” are due to students taking more than the mean number of points. Recall that calculation of “Theoretical Minimum” is based on the mean number of points.

Figures 15–1 through 15–8 show the scale score distributions for the total test for the content areas mathematics, reading, science, and writing. Tests with multiple benchmark cuts are broken down to match the grade level of the cuts while tests that are course-specific are not broken down. The benchmark cuts in place during the 2020–2021 school year are shown in green[2]. The bottom plot in each figure represents the distribution of items in the content area pools.

Figure 15–1. Scale Score Distribution – Math Grades 3–5 Total Scores



[2] For details on benchmark cuts, see Chapter Ten and Chapter Nineteen.

Figure 15–2. Scale Score Distribution – Math Total Scores

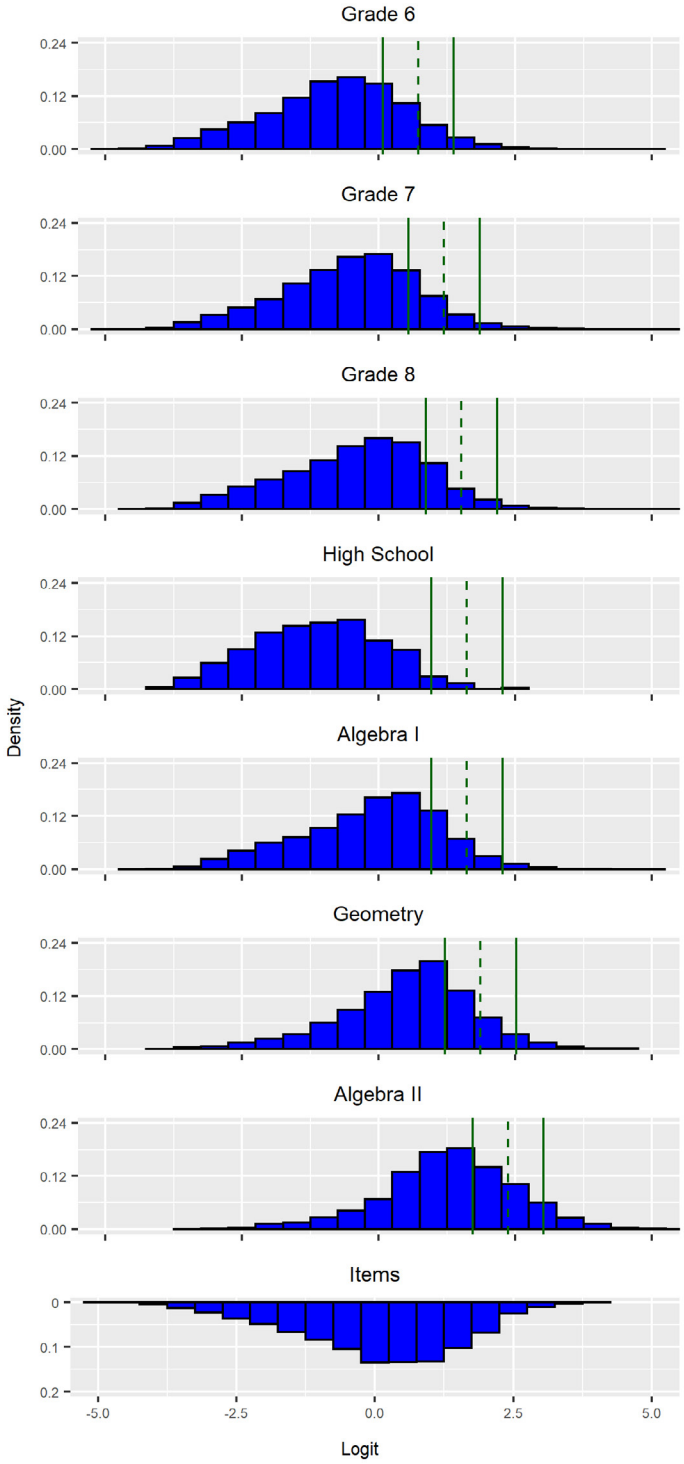


Figure 15–3. Scale Score Distribution – Reading Grades 3–5 Total Scores

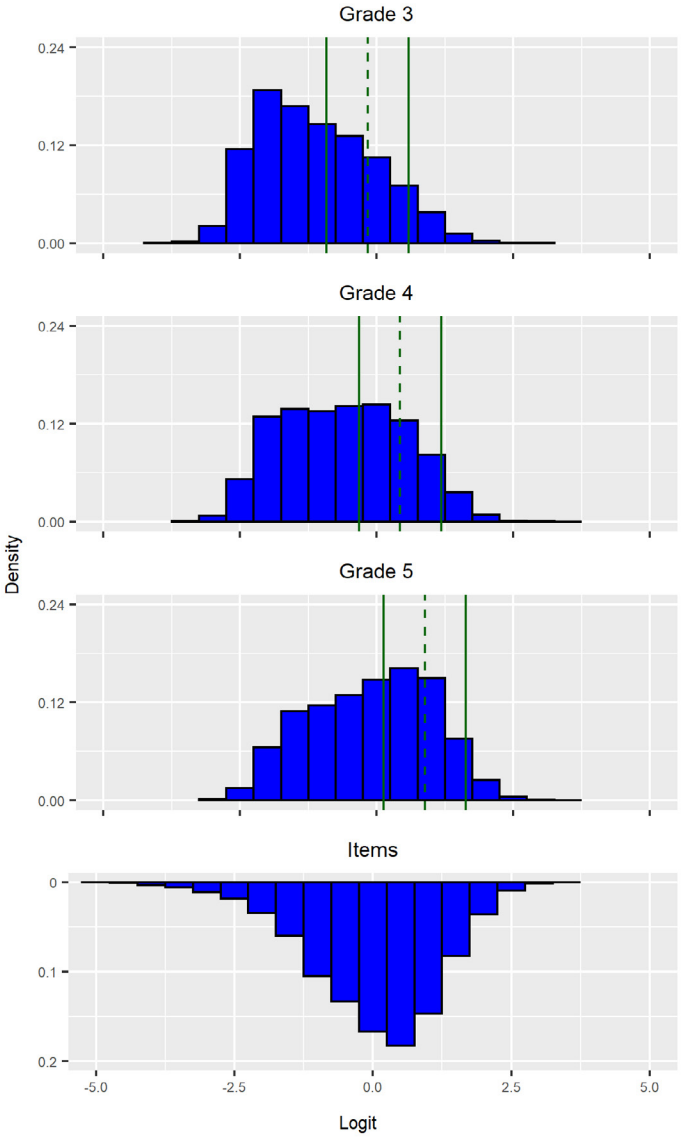


Figure 15–4. Scale Score Distribution – Reading/Literature Total Scores

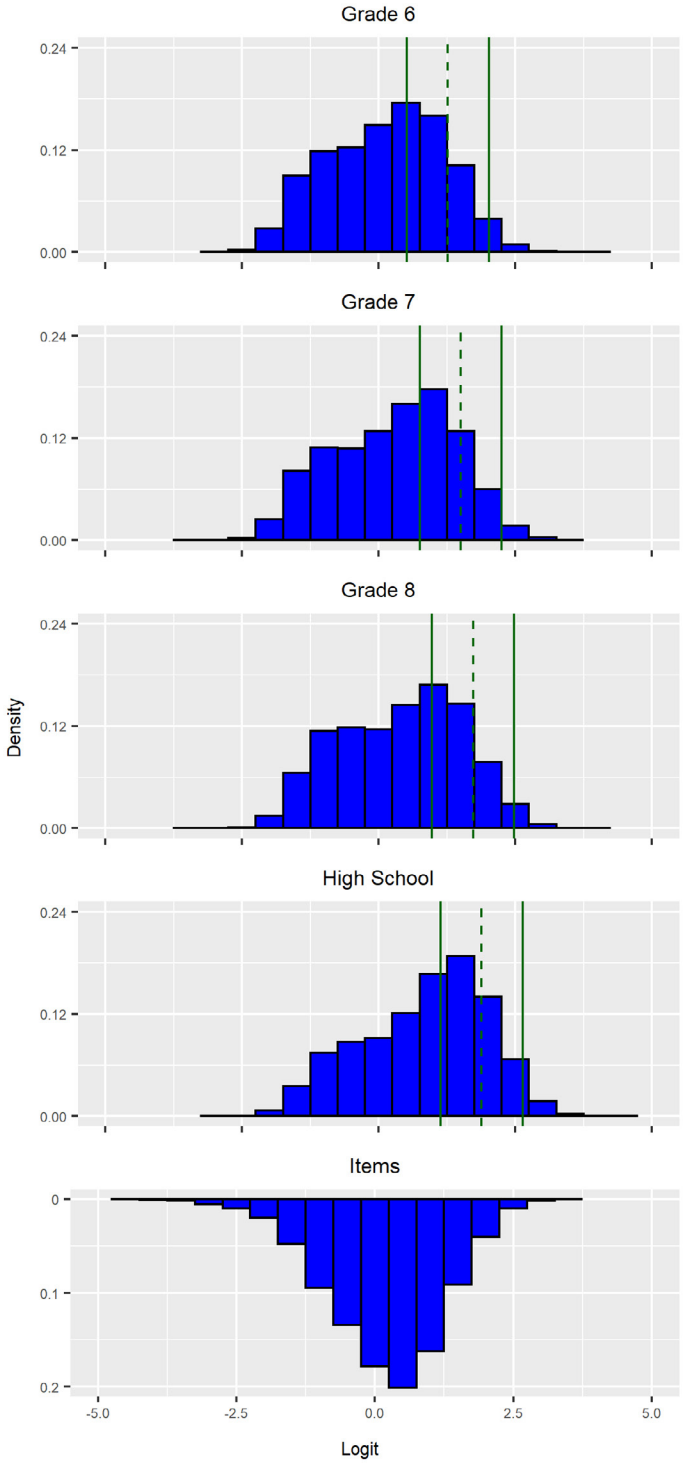


Figure 15–5. Scale Score Distribution – Science Grades 3–5 Total Scores

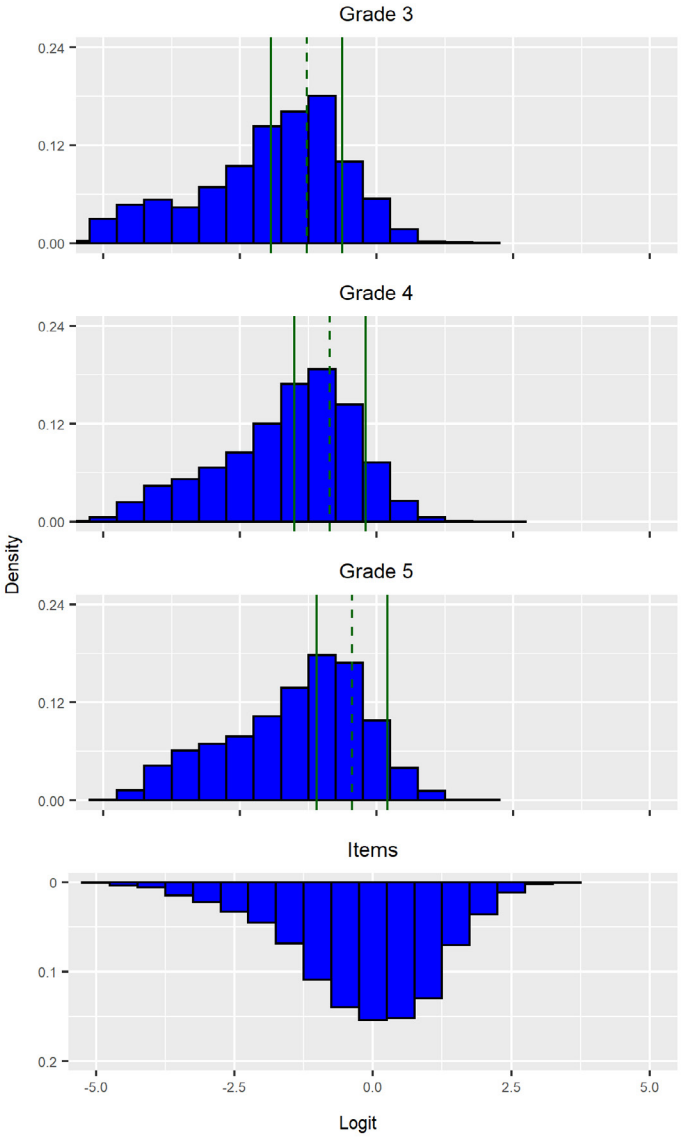


Figure 15–6. Scale Score Distribution – Science Total Scores

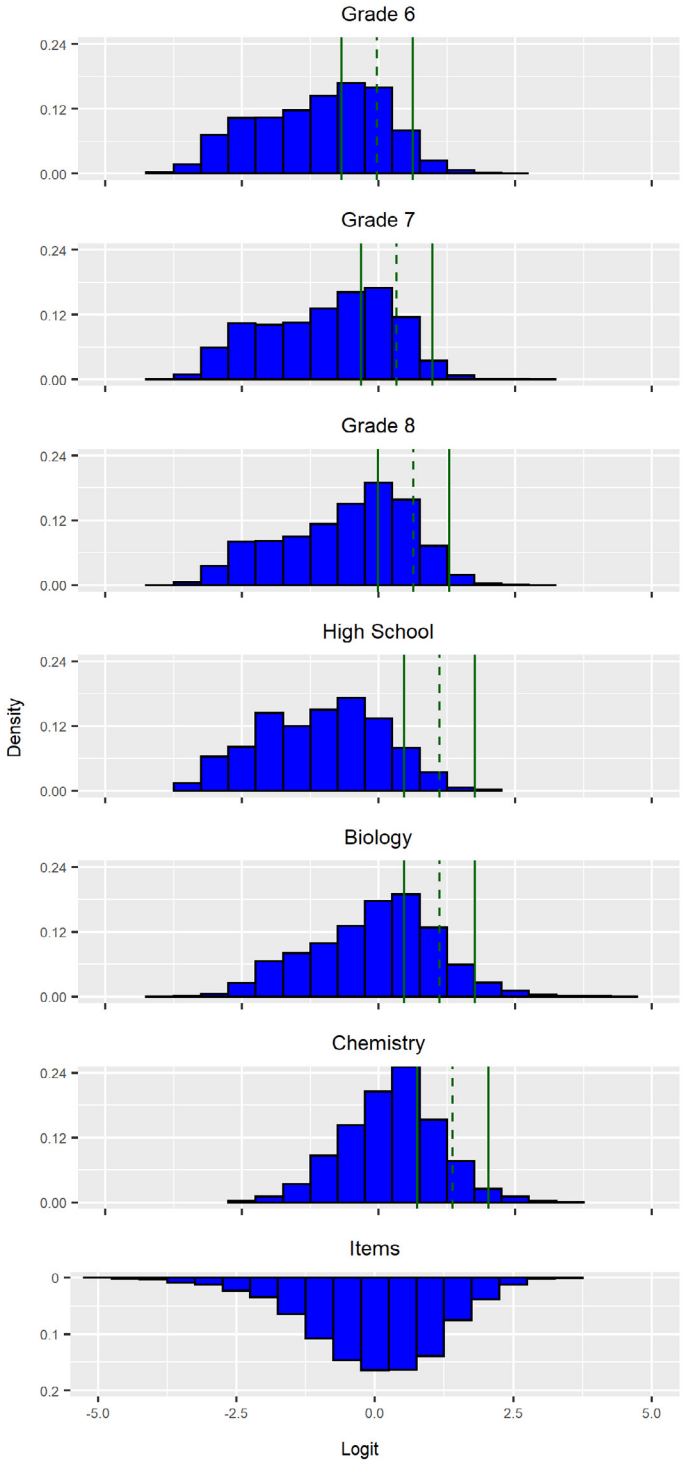


Figure 15–7. Scale Score Distribution – Writing Grades 3–5 Total Scores

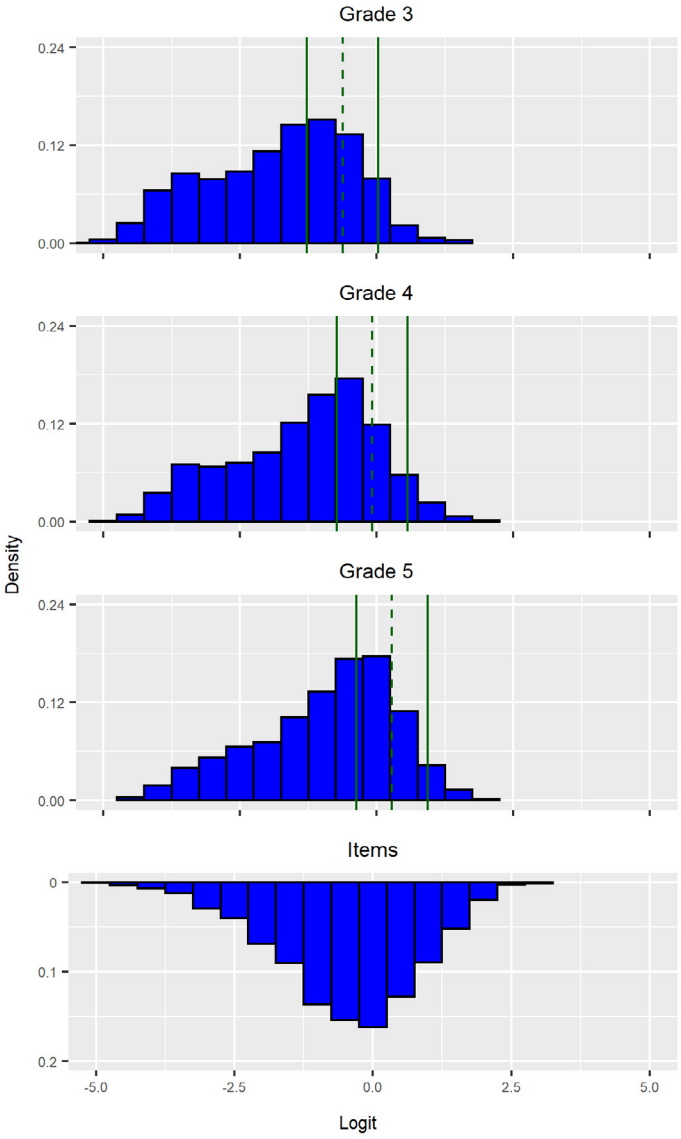
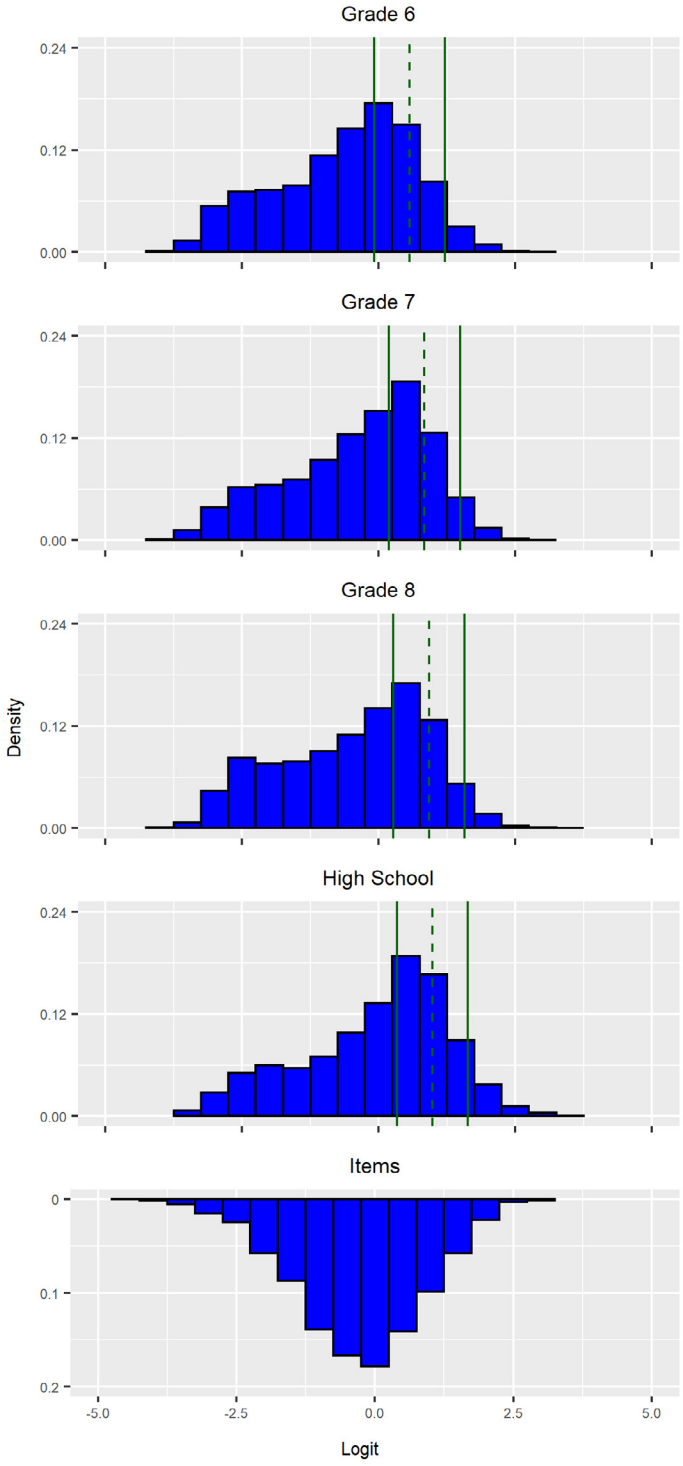


Figure 15–8. Scale Score Distribution – Writing/English Composition Total Scores



SUMMARY STATISTICS – SCALE SCORES AND CONDITIONAL STANDARD ERRORS FOR DIAGNOSTIC CATEGORY SUB-SCORES FROM FULL CDT

Earlier in this chapter, tables 15–18b and 15–19b show summary statistics for the diagnostic category scale scores and conditional standard errors from diagnostic category CDT tests. In this section, tables Table 15–20 and Table 15–21 show summary statistics for diagnostic categories from full CDT tests. Diagnostic category sub-scores from full CDTs are presented here because N-counts are significantly higher. For example, there are only 45 tests of Algebra II Data Analysis while there are 6,023 tests of Algebra II which includes the sub-score Data Analysis. To be consistent with previous tables, tests with multiple benchmark cuts are broken down to match the grade level of the cuts, while tests that are course-specific are not broken down.

Table 15–20. Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math – G3	Numbers and Operations	23,191	200	575	701	694.24	810	1644
Math – G3	Algebraic Concepts	23,191	200	588	732	723.32	856	1600
Math – G3	Geometry	23,191	200	582	691	685.31	788	1426
Math – G3	Measurement, Data, and Probability	23,191	200	572	715	709.08	840	1378
Math – G4	Numbers and Operations	24,069	200	675	796	786.81	899	1648
Math – G4	Algebraic Concepts	24,069	200	698	816	808.98	937	1572
Math – G4	Geometry	24,069	200	667	755	762.98	856	1475
Math – G4	Measurement, Data, and Probability	24,069	200	680	809	796.00	919	1460
Math – G5	Numbers and Operations	28,759	200	747	870	859.46	983	1688
Math – G5	Algebraic Concepts	28,759	200	745	861	845.46	960	1724
Math – G5	Geometry	28,759	200	727	825	829.70	944	1745
Math – G5	Measurement, Data, and Probability	28,759	200	759	862	850.37	953	1608
Math – G6	Numbers and Operations	34,688	200	806	935	926.76	1063	1718
Math – G6	Algebraic Concepts	34,688	200	795	925	907.55	1034	1776
Math – G6	Geometry	34,688	205	815	926	916.67	1026	1765
Math – G6	Measurement, Data, and Probability	34,688	206	801	910	910.50	1029	1624
Math – G7	Numbers and Operations	34,947	200	840	968	953.90	1087	1719
Math – G7	Algebraic Concepts	34,947	205	843	970	949.43	1071	1799
Math – G7	Geometry	34,947	200	853	955	947.76	1055	1824
Math – G7	Measurement, Data, and Probability	34,947	222	810	931	927.71	1052	1635
Math – G8	Numbers and Operations	27,468	200	836	982	958.20	1109	1759
Math – G8	Algebraic Concepts	27,468	236	854	994	973.29	1093	1826

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math – G8	Geometry	27,468	270	857	975	967.05	1083	1843
Math – G8	Measurement, Data, and Probability	27,468	211	834	971	957.30	1089	1808
Math – HS	Numbers and Operations	454	342	740	888	872.47	1023	1443
Math – HS	Algebraic Concepts	454	266	765	885	882.02	1009	1378
Math – HS	Geometry	454	400	779	874	884.29	998	1337
Math – HS	Measurement, Data, and Probability	454	234	734	869	864.27	1007	1236
Algebra I	Operations with Real Numbers and Expressions	42,725	400	863	1020	983.94	1128	1801
Algebra I	Linear Equations & Inequalities	42,725	420	891	1021	1009.86	1123	1786
Algebra I	Functions & Coordinate Geometry	42,725	400	899	1022	1007.27	1119	1809
Algebra I	Data Analysis	42,725	400	868	1006	981.35	1109	1677
Geometry	Geometric Properties	5,131	400	984	1099	1076.25	1185	1779
Geometry	Congruence, Similarity, and Proofs	5,131	400	998	1094	1080.87	1188	1637
Geometry	Coordinate Geometry and Right Triangles	5,131	400	990	1106	1085.84	1195	1795
Geometry	Measurement	5,131	400	962	1094	1070.99	1186	1679
Algebra II	Operations with Complex Numbers	6,023	570	1087	1207	1250.60	1404	1842
Algebra II	Non-Linear Expressions & Equations	6,023	400	1073	1182	1169.96	1285	1895
Algebra II	Functions	6,023	471	1072	1173	1163.18	1261	1867
Algebra II	Data Analysis	6,023	400	1039	1140	1124.19	1224	1850
Reading – G3	Key Ideas – Lit text	19,838	200	583	703	712.02	842	1463
Reading – G3	Key Ideas – Info text	19,838	200	574	698	704.69	840	1508
Reading – G3	Craft & Structure – Lit text	19,838	200	615	719	726.15	838	1518
Reading – G3	Craft & Structure – Info text	19,838	200	594	704	711.59	826	1508
Reading – G3	Vocabulary	19,838	200	566	707	706.35	843	1479
Reading – G4	Key Ideas – Lit text	20,691	200	638	781	781.15	924	1524
Reading – G4	Key Ideas – Info text	20,691	200	640	779	775.61	913	1559

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Reading – G4	Craft & Structure – Lit text	20,691	200	679	799	801.95	928	1522
Reading – G4	Craft & Structure – Info text	20,691	200	649	780	780.83	919	1548
Reading – G4	Vocabulary	20,691	200	637	792	778.08	917	1515
Reading – G5	Key Ideas – Lit text	24,324	200	719	872	856.63	997	1567
Reading – G5	Key Ideas – Info text	24,324	200	724	858	849.34	979	1564
Reading – G5	Craft & Structure – Lit text	24,324	200	750	873	870.06	997	1613
Reading – G5	Craft & Structure – Info text	24,324	200	701	863	844.70	995	1570
Reading – G5	Vocabulary	24,324	200	715	871	851.69	988	1545
Reading – G6	Key Ideas – Lit text	25,753	200	771	905	896.15	1022	1592
Reading – G6	Key Ideas – Info text	25,753	200	768	903	891.76	1020	1600
Reading – G6	Craft & Structure – Lit text	25,753	243	775	901	891.61	1011	1594
Reading – G6	Craft & Structure – Info text	25,753	200	750	903	882.12	1021	1595
Reading – G6	Vocabulary	25,753	200	750	904	882.71	1017	1580
Reading – G7	Key Ideas – Lit text	26,355	200	782	924	912.83	1046	1642
Reading – G7	Key Ideas – Info text	26,355	200	788	935	918.65	1054	1599
Reading – G7	Craft & Structure – Lit text	26,355	219	792	925	913.43	1040	1620
Reading – G7	Craft & Structure – Info text	26,355	200	763	929	901.81	1047	1664
Reading – G7	Vocabulary	26,355	200	772	933	907.47	1051	1605
Reading – G8	Key Ideas – Lit text	24,777	211	786	936	925.27	1064	1647
Reading – G8	Key Ideas – Info text	24,777	200	795	950	929.31	1071	1627
Reading – G8	Craft & Structure – Lit text	24,777	255	816	944	935.61	1061	1637
Reading – G8	Craft & Structure – Info text	24,777	235	786	940	920.13	1061	1687
Reading – G8	Vocabulary	24,777	200	782	952	927.00	1078	1641
Literature	Key Ideas – Lit text	62,170	203	849	1000	980.87	1117	1654
Literature	Key Ideas – Info text	62,170	200	874	1017	993.88	1125	1646
Literature	Craft & Structure – Lit text	62,170	292	869	1007	990.36	1116	1672
Literature	Craft & Structure – Info text	62,170	220	867	1013	991.45	1124	1654
Literature	Vocabulary	62,170	200	875	1030	1007.85	1147	1644

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Science – G3	The Nature of Science	1,557	200	556	713	686.32	823	1294
Science – G3	Biological Sciences	1,557	200	573	719	693.92	832	1193
Science – G3	Physical Sciences	1,557	200	581	732	698.54	835	1199
Science – G3	Earth and Space Sciences	1,557	200	587	715	694.04	829	1198
Science – G4	The Nature of Science	13,182	200	607	747	723.76	852	1362
Science – G4	Biological Sciences	13,182	200	617	750	728.75	858	1324
Science – G4	Physical Sciences	13,182	200	635	766	736.51	859	1353
Science – G4	Earth and Space Sciences	13,182	200	616	754	734.74	855	1404
Science – G5	The Nature of Science	5,939	200	615	765	741.45	877	1304
Science – G5	Biological Sciences	5,939	200	618	769	746.58	891	1342
Science – G5	Physical Sciences	5,939	200	649	784	760.98	883	1327
Science – G5	Earth and Space Sciences	5,939	200	637	780	756.29	882	1304
Science – G6	The Nature of Science	10,228	200	662	822	800.53	942	1475
Science – G6	Biological Sciences	10,228	200	671	823	803.07	934	1504
Science – G6	Physical Sciences	10,228	200	699	823	812.69	930	1425
Science – G6	Earth and Space Sciences	10,228	200	701	828	813.83	939	1379
Science – G7	The Nature of Science	15,341	200	675	844	817.83	963	1559
Science – G7	Biological Sciences	15,341	200	677	842	821.10	961	1503
Science – G7	Physical Sciences	15,341	200	723	852	840.24	960	1386
Science – G7	Earth and Space Sciences	15,341	200	712	844	829.30	950	1348
Science – G8	The Nature of Science	25,927	200	732	896	864.37	1001	1475
Science – G8	Biological Sciences	25,927	200	737	892	866.62	1003	1494
Science – G8	Physical Sciences	25,927	200	765	896	878.30	1000	1397
Science – G8	Earth and Space Sciences	25,927	200	752	887	862.91	983	1369
Science – HS	The Nature of Science	493	365	653	826	805.33	944	1254
Science – HS	Biological Sciences	493	208	679	819	809.40	936	1245
Science – HS	Physical Sciences	493	273	703	840	828.59	948	1201

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Science – HS	Earth and Space Sciences	493	293	663	818	800.72	937	1262
Biology	Basic Biological Principles/Chemical Basis for Life	46,477	400	829	952	936.85	1057	1789
Biology	Bioenergetics/ Homeostasis and Transport	46,477	400	846	949	948.76	1048	1727
Biology	Cell Growth and Reproduction/ Genetics	46,477	400	846	960	950.24	1053	1763
Biology	Theory of Evolution/ Ecology	46,477	400	797	954	921.89	1058	1733
Chemistry	Properties and Classification of Matter	3,070	400	853	986	951.42	1070	1550
Chemistry	Atomic Structure and The Periodic Table	3,070	513	922	1004	1007.50	1090	1546
Chemistry	The Mole and Chemical Bonding	3,070	440	909	996	993.05	1079	1590
Chemistry	Chemical Relationships and Reactions	3,070	421	901	985	985.02	1073	1552
Writing – G3	Quality of Writing: Focus and Organization	2,677	200	558	722	704.39	858	1219
Writing – G3	Quality of Writing: Content and Style	2,677	200	578	733	704.76	846	1307
Writing – G3	Quality of Writing: Editing	2,677	200	578	727	711.20	843	1313
Writing – G3	Conventions: Punctuation, Capitalization, and Spelling	2,677	200	585	724	717.31	848	1290
Writing – G3	Conventions: Grammar and Sentence Formation	2,677	200	575	725	712.78	864	1356
Writing – G4	Quality of Writing: Focus and Organization	2,177	200	612	785	761.15	915	1537
Writing – G4	Quality of Writing: Content and Style	2,177	200	633	793	768.81	903	1360
Writing – G4	Quality of Writing: Editing	2,177	200	623	786	761.63	903	1307

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Writing – G4	Conventions: Punctuation, Capitalization, and Spelling	2,177	200	646	785	775.10	907	1390
Writing – G4	Conventions: Grammar and Sentence Formation	2,177	200	661	813	777.86	915	1419
Writing – G5	Quality of Writing: Focus and Organization	3,624	200	670	838	809.14	957	1373
Writing – G5	Quality of Writing: Content and Style	3,624	200	703	844	822.58	954	1527
Writing – G5	Quality of Writing: Editing	3,624	200	703	843	817.04	954	1571
Writing – G5	Conventions: Punctuation, Capitalization, and Spelling	3,624	200	704	848	832.80	964	1444
Writing – G5	Conventions: Grammar and Sentence Formation	3,624	200	724	860	828.27	957	1431
Writing – G6	Quality of Writing: Focus and Organization	4,985	200	713	888	856.92	1009	1529
Writing – G6	Quality of Writing: Content and Style	4,985	233	733	887	873.83	1011	1558
Writing – G6	Quality of Writing: Editing	4,985	200	738	886	859.72	1001	1445
Writing – G6	Conventions: Punctuation, Capitalization, and Spelling	4,985	225	755	907	883.35	1009	1610
Writing – G6	Conventions: Grammar and Sentence Formation	4,985	200	748	891	857.62	990	1444
Writing – G7	Quality of Writing: Focus and Organization	6,262	200	751	929	889.49	1046	1568
Writing – G7	Quality of Writing: Content and Style	6,262	230	753	927	902.46	1048	1608
Writing – G7	Quality of Writing: Editing	6,262	200	773	923	893.07	1038	1456

Table 15–20 (continued). Summary Statistics for Diagnostic Category Scale Score Based on Full CDT

CDT	Diagnostic Category	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Writing – G7	Conventions: Punctuation, Capitalization, and Spelling	6,262	246	794	940	916.60	1055	1717
Writing – G7	Conventions: Grammar and Sentence Formation	6,262	200	780	923	887.60	1022	1652
Writing – G8	Quality of Writing: Focus and Organization	5,666	200	720	913	877.96	1050	1429
Writing – G8	Quality of Writing: Content and Style	5,666	240	734	912	890.00	1044	1622
Writing – G8	Quality of Writing: Editing	5,666	215	742	913	883.16	1035	1444
Writing – G8	Conventions: Punctuation, Capitalization, and Spelling	5,666	201	760	929	904.61	1053	1709
Writing – G8	Conventions: Grammar and Sentence Formation	5,666	200	765	925	887.54	1032	1679
English Composition	Quality of Writing: Focus and Organization	4,635	200	815	981	939.28	1091	1618
English Composition	Quality of Writing: Content and Style	4,635	259	804	971	937.35	1083	1624
English Composition	Quality of Writing: Editing	4,635	211	806	966	932.75	1071	1669
English Composition	Conventions: Punctuation, Capitalization, and Spelling	4,635	200	832	989	962.63	1106	1705
English Composition	Conventions: Grammar and Sentence Formation	4,635	200	838	975	939.90	1072	1676

Table 15–21 shows the summary statistics for the conditional standard errors of measurement (CSEMs) for diagnostic categories in the scale score metric based on full CDT. The final column in the table shows the theoretical minimum CSEM that is possible for a test length equal to the mean number of points. Minimum values in the table that are less than the theoretical minimum are due to students taking more than the mean number of points.

Table 15–21. Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Math – G3	Numbers and Operations	23,191	71	74	76	76.29	77	233	72.63
Math – G3	Algebraic Concepts	23,191	72	74	76	76.09	77	234	72.63
Math – G3	Geometry	23,191	72	75	76	75.98	77	234	72.63
Math – G3	Measurement, Data, and Probability	23,191	71	74	76	76.20	77	234	72.63
Math – G4	Numbers and Operations	24,069	72	74	76	75.93	77	232	72.63
Math – G4	Algebraic Concepts	24,069	72	74	76	75.82	77	232	72.63
Math – G4	Geometry	24,069	72	74	76	75.75	77	233	72.63
Math – G4	Measurement, Data, and Probability	24,069	71	74	76	75.97	77	233	72.63
Math – G5	Numbers and Operations	28,759	72	74	76	76.11	77	231	72.63
Math – G5	Algebraic Concepts	28,759	72	74	76	75.74	77	232	72.63
Math – G5	Geometry	28,759	72	74	76	75.72	77	241	72.63
Math – G5	Measurement, Data, and Probability	28,759	72	74	76	75.83	77	234	72.63
Math – G6	Numbers and Operations	34,688	69	73	74	74.22	74	232	69.28
Math – G6	Algebraic Concepts	34,688	69	73	74	74.40	74	237	69.28
Math – G6	Geometry	34,688	69	73	74	73.96	74	231	69.28
Math – G6	Measurement, Data, and Probability	34,688	69	73	74	74.26	74	232	69.28
Math – G7	Numbers and Operations	34,947	69	73	74	74.34	74	233	69.28
Math – G7	Algebraic Concepts	34,947	69	73	74	74.43	74	234	69.28
Math – G7	Geometry	34,947	69	73	74	73.85	74	235	69.28
Math – G7	Measurement, Data, and Probability	34,947	69	73	74	74.45	74	233	69.28
Math – G8	Numbers and Operations	27,468	69	73	74	74.39	74	233	69.28
Math – G8	Algebraic Concepts	27,468	69	73	74	74.92	74	238	69.28
Math – G8	Geometry	27,468	69	73	74	74.33	74	233	69.28
Math – G8	Measurement, Data, and Probability	27,468	69	73	74	74.79	74	235	69.28
Math – HS	Numbers and Operations	454	70	73	74	75.24	75	135	69.28
Math – HS	Algebraic Concepts	454	70	73	74	76.36	75	234	66.76
Math – HS	Geometry	454	70	73	74	75.24	75	147	69.28
Math – HS	Measurement, Data, and Probability	454	70	73	74	76.12	75	233	69.28

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Algebra I	Operations with Real Numbers and Expressions	42,725	69	73	74	74.91	75	232	69.28
Algebra I	Linear Equations & Inequalities	42,725	69	73	74	75.73	75	232	69.28
Algebra I	Functions & Coordinate Geometry	42,725	69	73	74	75.24	74	238	69.28
Algebra I	Data Analysis	42,725	69	73	74	75.35	75	236	69.28
Geometry	Geometric Properties	5,131	70	73	74	74.14	74	232	69.28
Geometry	Congruence, Similarity, and Proofs	5,131	70	73	74	74.88	75	239	69.28
Geometry	Coordinate Geometry and Right Triangles	5,131	69	73	74	74.96	74	233	69.28
Geometry	Measurement	5,131	70	73	74	74.25	74	234	69.28
Algebra II	Operations with Complex Numbers	6,023	69	73	74	85.40	81	232	66.76
Algebra II	Non-Linear Expressions & Equations	6,023	70	73	74	74.78	74	234	69.28
Algebra II	Functions	6,023	69	73	74	74.68	74	231	69.28
Algebra II	Data Analysis	6,023	70	73	74	74.31	74	233	69.28
Reading – G3	Key Ideas – Lit text	19,838	73	89	95	98.99	101	275	86.44
Reading – G3	Key Ideas – Info text	19,838	76	91	97	101.34	103	275	90.29
Reading – G3	Craft & Structure – Lit text	19,838	76	93	99	103.76	105	279	90.29
Reading – G3	Craft & Structure – Info text	19,838	72	90	96	100.83	102	276	86.44
Reading – G3	Vocabulary	19,838	77	93	98	103.06	103	276	90.29
Reading – G4	Key Ideas – Lit text	20,691	74	89	95	98.51	101	275	86.44
Reading – G4	Key Ideas – Info text	20,691	75	90	96	99.56	101	277	90.29
Reading – G4	Craft & Structure – Lit text	20,691	73	92	97	101.15	103	274	86.44
Reading – G4	Craft & Structure – Info text	20,691	72	89	94	99.11	101	278	86.44
Reading – G4	Vocabulary	20,691	76	93	98	102.41	103	273	90.29
Reading – G5	Key Ideas – Lit text	24,324	72	89	95	98.62	101	280	86.44
Reading – G5	Key Ideas – Info text	24,324	72	90	95	97.87	100	280	86.44

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Reading – G5	Craft & Structure – Lit text	24,324	72	89	95	97.74	101	271	86.44
Reading – G5	Craft & Structure – Info text	24,324	71	90	95	98.87	101	277	86.44
Reading – G5	Vocabulary	24,324	75	94	98	102.90	103	277	90.29
Reading – G6	Key Ideas – Lit text	25,753	72	88	93	97.62	100	273	82.46
Reading – G6	Key Ideas – Info text	25,753	73	89	94	97.77	100	286	82.46
Reading – G6	Craft & Structure – Lit text	25,753	72	89	94	97.25	100	272	86.13
Reading – G6	Craft & Structure – Info text	25,753	73	88	93	97.50	100	275	82.46
Reading – G6	Vocabulary	25,753	78	94	98	102.37	103	280	86.13
Reading – G7	Key Ideas – Lit text	26,355	72	87	92	96.91	100	276	82.46
Reading – G7	Key Ideas – Info text	26,355	75	89	94	98.14	100	270	82.46
Reading – G7	Craft & Structure – Lit text	26,355	74	89	94	97.64	100	273	86.13
Reading – G7	Craft & Structure – Info text	26,355	75	89	94	98.43	101	271	86.13
Reading – G7	Vocabulary	26,355	81	94	98	103.39	103	285	86.13
Reading – G8	Key Ideas – Lit text	24,777	73	88	94	98.69	100	276	82.46
Reading – G8	Key Ideas – Info text	24,777	75	89	94	98.84	101	272	86.13
Reading – G8	Craft & Structure – Lit text	24,777	75	89	94	98.59	100	276	86.13
Reading – G8	Craft & Structure – Info text	24,777	74	89	93	98.55	100	273	86.13
Reading – G8	Vocabulary	24,777	82	96	100	104.73	104	280	86.13
Literature	Key Ideas – Lit text	62,170	73	89	95	100.22	101	275	82.46
Literature	Key Ideas – Info text	62,170	75	89	94	99.44	101	280	86.13
Literature	Craft & Structure – Lit text	62,170	81	90	94	98.38	100	274	86.13
Literature	Craft & Structure – Info text	62,170	78	89	93	98.81	100	273	86.13
Literature	Vocabulary	62,170	83	96	101	109.92	107	285	86.13
Science – G3	The Nature of Science	1,557	76	79	80	81.05	82	259	77.26
Science – G3	Biological Sciences	1,557	76	79	81	80.75	82	139	77.26
Science – G3	Physical Sciences	1,557	76	79	80	81.00	82	247	77.26
Science – G3	Earth and Space Sciences	1,557	71	79	80	80.88	82	246	77.26

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Science – G4	The Nature of Science	13,182	71	79	80	81.10	82	263	77.26
Science – G4	Biological Sciences	13,182	74	79	80	80.98	82	247	77.26
Science – G4	Physical Sciences	13,182	73	79	80	80.81	82	248	77.26
Science – G4	Earth and Space Sciences	13,182	73	79	80	80.81	82	249	77.26
Science – G5	The Nature of Science	5,939	72	79	80	81.81	82	273	77.26
Science – G5	Biological Sciences	5,939	73	79	80	81.11	82	247	77.26
Science – G5	Physical Sciences	5,939	74	79	80	81.07	82	251	77.26
Science – G5	Earth and Space Sciences	5,939	74	79	80	80.87	82	248	77.26
Science – G6	The Nature of Science	10,228	70	77	78	79.59	79	276	73.70
Science – G6	Biological Sciences	10,228	72	77	78	79.96	79	246	73.70
Science – G6	Physical Sciences	10,228	71	77	78	80.23	79	248	73.70
Science – G6	Earth and Space Sciences	10,228	73	77	78	79.56	79	248	73.70
Science – G7	The Nature of Science	15,341	74	77	78	79.74	79	246	73.70
Science – G7	Biological Sciences	15,341	71	77	78	80.09	79	246	73.70
Science – G7	Physical Sciences	15,341	73	77	78	79.74	79	248	73.70
Science – G7	Earth and Space Sciences	15,341	71	77	78	79.67	79	246	73.70
Science – G8	The Nature of Science	25,927	73	77	78	79.50	79	246	73.70
Science – G8	Biological Sciences	25,927	71	77	78	79.67	79	246	73.70
Science – G8	Physical Sciences	25,927	74	77	78	79.63	79	249	73.70
Science – G8	Earth and Space Sciences	25,927	73	77	78	79.52	79	251	73.70
Science – HS	The Nature of Science	493	74	78	79	79.97	79	139	73.70
Science – HS	Biological Sciences	493	74	78	79	81.43	80	246	73.70
Science – HS	Physical Sciences	493	74	78	79	80.84	79	245	73.70
Science – HS	Earth and Space Sciences	493	74	77	79	81.20	80	145	73.70

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Biology	Basic Biological Principles/Chemical Basis for Life	46,477	69	77	78	79.86	79	256	73.70
Biology	Bioenergetics/ Homeostasis and Transport	46,477	71	77	78	79.97	79	246	73.70
Biology	Cell Growth and Reproduction/ Genetics	46,477	72	77	78	80.37	79	247	73.70
Biology	Theory of Evolution/ Ecology	46,477	71	77	78	80.12	79	247	73.70
Chemistry	Properties and Classification of Matter	3,070	74	77	78	79.55	79	251	73.70
Chemistry	Atomic Structure and The Periodic Table	3,070	74	77	79	81.00	79	246	73.70
Chemistry	The Mole and Chemical Bonding	3,070	74	77	79	80.03	79	246	73.70
Chemistry	Chemical Relationships and Reactions	3,070	74	77	79	80.54	79	246	73.70
Writing – G3	Quality of Writing: Focus and Organization	2,677	82	86	87	89.25	89	249	84.09
Writing – G3	Quality of Writing: Content and Style	2,677	82	86	87	89.84	89	255	84.09
Writing – G3	Quality of Writing: Editing	2,677	82	86	87	89.25	89	247	84.09
Writing – G3	Conventions: Punctuation, Capitalization, and Spelling	2,677	82	86	87	89.98	89	247	84.09
Writing – G3	Conventions: Grammar and Sentence Formation	2,677	82	86	87	89.11	89	248	84.09
Writing – G4	Quality of Writing: Focus and Organization	2,177	82	86	87	89.39	89	250	84.09
Writing – G4	Quality of Writing: Content and Style	2,177	82	86	87	89.34	89	249	84.09
Writing – G4	Quality of Writing: Editing	2,177	82	86	87	88.80	89	247	84.09

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Writing – G4	Conventions: Punctuation, Capitalization, and Spelling	2,177	82	86	87	89.30	89	247	84.09
Writing – G4	Conventions: Grammar and Sentence Formation	2,177	82	86	87	88.56	89	248	84.09
Writing – G5	Quality of Writing: Focus and Organization	3,624	82	86	87	88.75	89	250	84.09
Writing – G5	Quality of Writing: Content and Style	3,624	82	86	87	89.13	89	252	84.09
Writing – G5	Quality of Writing: Editing	3,624	82	86	87	88.29	88	248	84.09
Writing – G5	Conventions: Punctuation, Capitalization, and Spelling	3,624	82	86	87	88.41	89	247	84.09
Writing – G5	Conventions: Grammar and Sentence Formation	3,624	82	86	87	88.00	89	247	84.09
Writing – G6	Quality of Writing: Focus and Organization	4,985	81	84	85	88.19	86	248	80.21
Writing – G6	Quality of Writing: Content and Style	4,985	81	84	85	89.06	86	247	80.21
Writing – G6	Quality of Writing: Editing	4,985	81	83	85	86.79	86	248	80.21
Writing – G6	Conventions: Punctuation, Capitalization, and Spelling	4,985	81	84	85	87.27	86	248	80.21
Writing – G6	Conventions: Grammar and Sentence Formation	4,985	80	83	85	86.36	86	251	80.21
Writing – G7	Quality of Writing: Focus and Organization	6,262	81	84	85	88.06	86	250	80.21
Writing – G7	Quality of Writing: Content and Style	6,262	81	84	85	88.32	86	247	80.21
Writing – G7	Quality of Writing: Editing	6,262	81	83	85	86.38	86	250	80.21

Table 15–21 (continued). Summary Statistics for Diagnostic Category Conditional Standard Errors Based on Full CDT

CDT	Diagnostic Category	*N*	Min	Q1	Median	Mean	Q3	Max	Theoretical Minimum
Writing – G7	Conventions: Punctuation, Capitalization, and Spelling	6,262	81	84	85	87.20	86	247	80.21
Writing – G7	Conventions: Grammar and Sentence Formation	6,262	81	83	85	86.64	86	251	80.21
Writing – G8	Quality of Writing: Focus and Organization	5,666	81	84	85	88.61	86	249	80.21
Writing – G8	Quality of Writing: Content and Style	5,666	81	84	85	88.65	86	248	80.21
Writing – G8	Quality of Writing: Editing	5,666	81	84	85	87.50	86	247	80.21
Writing – G8	Conventions: Punctuation, Capitalization, and Spelling	5,666	80	84	85	88.13	86	249	80.21
Writing – G8	Conventions: Grammar and Sentence Formation	5,666	81	84	85	87.36	86	255	80.21
Eng. Comp.	Quality of Writing: Focus and Organization	4,635	81	83	85	88.05	86	250	80.21
Eng. Comp.	Quality of Writing: Content and Style	4,635	81	84	85	88.06	86	248	80.21
Eng. Comp.	Quality of Writing: Editing	4,635	81	84	85	87.02	86	248	80.21
Eng. Comp.	Conventions: Punctuation, Capitalization, and Spelling	4,635	81	84	85	87.68	86	248	80.21
Eng. Comp.	Conventions: Grammar and Sentence Formation	4,635	81	83	85	86.66	86	251	80.21

DIAGNOSTIC CATEGORY SCORE DIFFERENCES

As described in Chapter Fourteen, the CDT reports that are available to teachers display scale scores and probable score ranges for each diagnostic category. The probable score range is the scale score ± one standard error. Probable score range differences—ranges that do not overlap—may indicate to teachers a meaningful difference between two diagnostic category scores. Tables 15–22a through 15–34a show the number of students with score range differences (non-overlapping probable score ranges) between pairs of diagnostic categories for each full[3] CDT test. For example, according to Table 15–22a, 16,646 students who took the Math Grades 3–5 assessment had score range differences between diagnostic categories 1 and 2 while 59,373 students did not. Tables 15–22b through 15–34b show the total number of score range differences. For example, 14,279 students had two pairs of diagnostic categories with score range differences, which was 18.8% of the total students who took Math Grades 3–5.

Table 15–22a. Diagnostic Category Score Range Differences – Math Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	16,646	59,373	21.9%	78.1%
DC1	DC3	20,877	55,142	27.5%	72.5%
DC1	DC4	16,666	59,353	21.9%	78.1%
DC2	DC3	21,539	54,480	28.3%	71.7%
DC2	DC4	15,990	60,029	21.0%	79.0%
DC3	DC4	20,265	55,754	26.7%	73.3%

Table 15–22b. Total Number of Diagnostic Category Score Range Differences – Math Grades 3–5

Number of Score Range Differences	Number of Students	Percent of Students
0	27,143	35.7%
1	13,483	17.7%
2	14,279	18.8%
3	15,442	20.3%
4	4,756	6.3%
5	904	1.2%
6	12	0.0%

Table 15–23a. Diagnostic Category Score Range Differences – Math Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	24,212	73,345	24.8%	75.2%
DC1	DC3	26,289	71,268	26.9%	73.1%
DC1	DC4	25,864	71,693	26.5%	73.5%
DC2	DC3	24,808	72,749	25.4%	74.6%
DC2	DC4	25,532	72,025	26.2%	73.8%
DC3	DC4	25,438	72,119	26.1%	73.9%

[3] Score differences between diagnostic categories are based on full CDTs because scores are based on the same test event. Comparisons are not made based on diagnostic category CDTs which may be taken at very different times.

Table 15–23b. Total Number of Diagnostic Category Score Range Differences – Math Grades 6–HS

Number of Score Range Differences	Number of Students	Percent of Students
0	33,561	34.4%
1	16,506	16.9%
2	17,660	18.1%
3	20,766	21.3%
4	7,336	7.5%
5	1,693	1.7%
6	35	0.0%

Table 15–24a. Diagnostic Category Score Range Differences – Algebra I

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	14,486	28,239	33.9%	66.1%
DC1	DC3	14,265	28,460	33.4%	66.6%
DC1	DC4	14,033	28,692	32.8%	67.2%
DC2	DC3	10,337	32,388	24.2%	75.8%
DC2	DC4	12,176	30,549	28.5%	71.5%
DC3	DC4	11,457	31,268	26.8%	73.2%

Table 15–24b. Total Number of Diagnostic Category Score Range Differences – Algebra I

Number of Score Range Differences	Number of Students	Percent of Students
0	12,727	29.8%
1	6,464	15.1%
2	7,504	17.6%
3	10,283	24.1%
4	4,331	10.1%
5	1,387	3.2%
6	29	0.1%

Table 15–25a. Diagnostic Category Score Range Differences – Geometry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,346	3,785	26.2%	73.8%
DC1	DC3	1,332	3,799	26.0%	74.0%
DC1	DC4	1,363	3,768	26.6%	73.4%
DC2	DC3	1,272	3,859	24.8%	75.2%
DC2	DC4	1,402	3,729	27.3%	72.7%
DC3	DC4	1,320	3,811	25.7%	74.3%

Table 15–25b. Total Number of Diagnostic Category Score Range Differences – Geometry

Number of Score Range Differences	Number of Students	Percent of Students
0	1,814	35.4%
1	826	16.1%
2	935	18.2%
3	1,017	19.8%
4	413	8.0%
5	120	2.3%
6	6	0.1%

Table 15–26a. Diagnostic Category Score Range Differences – Algebra II

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	2,428	3,595	40.3%	59.7%
DC1	DC3	2,544	3,479	42.2%	57.8%
DC1	DC4	2,972	3,051	49.3%	50.7%
DC2	DC3	1,617	4,406	26.8%	73.2%
DC2	DC4	1,895	4,128	31.5%	68.5%
DC3	DC4	1,681	4,342	27.9%	72.1%

Table 15–26b. Total Number of Diagnostic Category Score Range Differences – Algebra II

Number of Score Range Differences	Number of Students	Percent of Students
0	1,416	23.5%
1	758	12.6%
2	903	15.0%
3	1,688	28.0%
4	804	13.3%
5	431	7.2%
6	23	0.4%

Table 15–27a. Diagnostic Category Score Range Differences – Reading Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	12,767	52,086	19.7%	80.3%
DC1	DC3	11,774	53,079	18.2%	81.8%
DC1	DC4	13,233	51,620	20.4%	79.6%
DC1	DC5	12,109	52,744	18.7%	81.3%
DC2	DC3	12,761	52,092	19.7%	80.3%
DC2	DC4	12,431	52,422	19.2%	80.8%
DC2	DC5	12,409	52,444	19.1%	80.9%
DC3	DC4	12,981	51,872	20.0%	80.0%
DC3	DC5	12,830	52,023	19.8%	80.2%
DC4	DC5	12,328	52,525	19.0%	81.0%

Table 15–27b. Total Number of Diagnostic Category Score Range Differences – Reading Grades 3–5

Number of Score Range Differences	Number of Students	Percent of Students
0	20,751	32.0%
1	9,757	15.0%
2	10,630	16.4%
3	8,568	13.2%
4	9,600	14.8%
5	3,112	4.8%
6	2,139	3.3%
7	260	0.4%
8	36	0.1%
9	0	0.0%
10	0	0.0%

Table 15–28a. Diagnostic Category Score Range Differences – Reading/Lit Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	27,719	111,336	19.9%	80.1%
DC1	DC3	25,785	113,270	18.5%	81.5%
DC1	DC4	28,094	110,961	20.2%	79.8%
DC1	DC5	29,098	109,957	20.9%	79.1%
DC2	DC3	26,820	112,235	19.3%	80.7%
DC2	DC4	25,490	113,565	18.3%	81.7%
DC2	DC5	27,927	111,128	20.1%	79.9%
DC3	DC4	27,371	111,684	19.7%	80.3%
DC3	DC5	28,591	110,464	20.6%	79.4%
DC4	DC5	27,774	111,281	20.0%	80.0%

Table 15–28b. Total Number of Diagnostic Category Score Range Differences – Reading/Lit Grades 6–HS

Number of Score Range Differences	Number of Students	Percent of Students
0	43,653	31.4%
1	21,051	15.1%
2	22,550	16.2%
3	18,279	13.1%
4	20,739	14.9%
5	6,890	5.0%
6	5,087	3.7%
7	695	0.5%
8	111	0.1%
9	0	0.0%
10	0	0.0%

Table 15–29a. Diagnostic Category Score Range Differences – Science Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	3,799	16,879	18.4%	81.6%
DC1	DC3	3,932	16,746	19.0%	81.0%
DC1	DC4	3,878	16,800	18.8%	81.2%
DC2	DC3	3,973	16,705	19.2%	80.8%
DC2	DC4	4,002	16,676	19.4%	80.6%
DC3	DC4	4,012	16,666	19.4%	80.6%

Table 15–29b. Total Number of Diagnostic Category Score Range Differences – Science Grades 3–5

Number of Score Range Differences	Number of Students	Percent of Students
0	9,517	46.0%
1	3,722	18.0%
2	3,447	16.7%
3	3,108	15.0%
4	764	3.7%
5	120	0.6%
6	0	0.0%

Table 15–30a. Diagnostic Category Score Range Differences – Science Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	11,468	40,521	22.1%	77.9%
DC1	DC3	11,933	40,056	23.0%	77.0%
DC1	DC4	11,908	40,081	22.9%	77.1%
DC2	DC3	11,787	40,202	22.7%	77.3%
DC2	DC4	11,806	40,183	22.7%	77.3%
DC3	DC4	11,316	40,673	21.8%	78.2%

Table 15–30b. Total Number of Diagnostic Category Score Range Differences – Science Grades 6–HS

Number of Score Range Differences	Number of Students	Percent of Students
0	20,804	40.0%
1	9,142	17.6%
2	9,154	17.6%
3	9,345	18.0%
4	3,000	5.8%
5	531	1.0%
6	13	0.0%

Table 15–31a. Diagnostic Category Score Range Differences – Biology

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	10,817	35,660	23.3%	76.7%
DC1	DC3	11,044	35,433	23.8%	76.2%
DC1	DC4	11,317	35,160	24.3%	75.7%
DC2	DC3	10,222	36,255	22.0%	78.0%
DC2	DC4	12,657	33,820	27.2%	72.8%
DC3	DC4	11,824	34,653	25.4%	74.6%

Table 15–31b. Total Number of Diagnostic Category Score Range Differences – Biology

Number of Score Range Differences	Number of Students	Percent of Students
0	17,183	37.0%
1	8,048	17.3%
2	8,350	18.0%
3	9,057	19.5%
4	3,240	7.0%
5	592	1.3%
6	7	0.0%

Table 15–32a. Diagnostic Category Score Range Differences – Chemistry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	905	2,165	29.5%	70.5%
DC1	DC3	778	2,292	25.3%	74.7%
DC1	DC4	890	2,180	29.0%	71.0%
DC2	DC3	590	2,480	19.2%	80.8%
DC2	DC4	651	2,419	21.2%	78.8%
DC3	DC4	632	2,438	20.6%	79.4%

Table 15–32b. Total Number of Diagnostic Category Score Range Differences – Chemistry

Number of Score Range Differences	Number of Students	Percent of Students
0	1,146	37.3%
1	558	18.2%
2	505	16.4%
3	612	19.9%
4	203	6.6%
5	46	1.5%
6	0	0.0%

Table 15–33a. Diagnostic Category Score Range Differences – Writing Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,572	6,906	18.5%	81.5%
DC1	DC3	1,647	6,831	19.4%	80.6%
DC1	DC4	1,915	6,563	22.6%	77.4%
DC1	DC5	1,822	6,656	21.5%	78.5%
DC2	DC3	1,617	6,861	19.1%	80.9%
DC2	DC4	1,856	6,622	21.9%	78.1%
DC2	DC5	1,681	6,797	19.8%	80.2%
DC3	DC4	1,647	6,831	19.4%	80.6%
DC3	DC5	1,501	6,977	17.7%	82.3%
DC4	DC5	1,762	6,716	20.8%	79.2%

Table 15–33b. Total Number of Diagnostic Category Score Range Differences – Writing Grades 3–5

Number of Score Range Differences	Number of Students	Percent of Students
0	2,706	31.9%
1	1,204	14.2%
2	1,300	15.3%
3	1,153	13.6%
4	1,299	15.3%
5	421	5.0%
6	325	3.8%
7	54	0.6%
8	16	0.2%
9	0	0.0%
10	0	0.0%

Table 15–34a. Diagnostic Category Score Range Differences – Writing/Eng Comp Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	4,759	16,789	22.1%	77.9%
DC1	DC3	4,696	16,852	21.8%	78.2%
DC1	DC4	5,416	16,132	25.1%	74.9%
DC1	DC5	5,112	16,436	23.7%	76.3%
DC2	DC3	4,824	16,724	22.4%	77.6%
DC2	DC4	5,308	16,240	24.6%	75.4%
DC2	DC5	5,163	16,385	24.0%	76.0%
DC3	DC4	5,036	16,512	23.4%	76.6%
DC3	DC5	4,730	16,818	22.0%	78.0%
DC4	DC5	5,069	16,479	23.5%	76.5%

Table 15–34b. Total Number of Diagnostic Category Score Range Differences – Writing/Eng Comp Grades 6–HS

Number of Score Range Differences	Number of Students	Percent of Students
0	5,675	26.3%
1	2,936	13.6%
2	3,372	15.6%
3	2,800	13.0%
4	3,751	17.4%
5	1,443	6.7%
6	1,246	5.8%
7	263	1.2%
8	61	0.3%
9	1	0.0%
10	0	0.0%

Significant differences among diagnostic categories were tested based on t-test. Using the diagnostic category scale scores and the conditional standard errors for each student, the differences between pairs of diagnostic category scores were examined based on t-test for each student. A Bonferroni correction for multiple comparisons was performed to keep the family wise Type I error rate at 0.32. This results in the number of significant differences being smaller than the number of score range differences (non-overlapping probable score ranges) presented above. Tables 15–35a through 15–47a show the number of students who had significant differences between pairs of diagnostic categories for each assessment. Tables 15–35b through 15–47b show the total number of significant differences.

Table 15–35a. Diagnostic Category Significant Differences – Math Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,555	74,464	2.0%	98.0%
DC1	DC3	2,798	73,221	3.7%	96.3%
DC1	DC4	1,592	74,427	2.1%	97.9%
DC2	DC3	2,978	73,041	3.9%	96.1%
DC2	DC4	1,556	74,463	2.0%	98.0%
DC3	DC4	2,469	73,550	3.2%	96.8%

Note: Z value is 1.94

Table 15–35b. Total Number of Diagnostic Category Significant Differences – Math Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	66,963	88.1%
1	6,020	7.9%
2	2,230	2.9%
3	759	1.0%
4	44	0.1%
5	3	0.0%
6	0	0.0%

Table 15–36a. Diagnostic Category Significant Differences – Math Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	3,280	94,277	3.4%	96.6%
DC1	DC3	3,869	93,688	4.0%	96.0%
DC1	DC4	3,607	93,950	3.7%	96.3%
DC2	DC3	3,398	94,159	3.5%	96.5%
DC2	DC4	3,538	94,019	3.6%	96.4%
DC3	DC4	3,543	94,014	3.6%	96.4%

Note: Z value is 1.94

Table 15–36b. Total Number of Diagnostic Category Significant Differences – Math Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	83,716	85.8%
1	8,368	8.6%
2	3,730	3.8%
3	1,568	1.6%
4	172	0.2%
5	3	0.0%
6	0	0.0%

Table 15–37a. Diagnostic Category Significant Differences – Algebra I

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	3,162	39,563	7.4%	92.6%
DC1	DC3	2,943	39,782	6.9%	93.1%
DC1	DC4	3,210	39,515	7.5%	92.5%
DC2	DC3	1,059	41,666	2.5%	97.5%
DC2	DC4	1,813	40,912	4.2%	95.8%
DC3	DC4	1,647	41,078	3.9%	96.1%

Note: Z value is 1.94

Table 15–37b. Total Number of Diagnostic Category Significant Differences – Algebra I

Number of Significant Differences	Number of Students	Percent of Students
0	34,094	79.8%
1	4,831	11.3%
2	2,547	6.0%
3	1,105	2.6%
4	146	0.3%
5	2	0.0%
6	0	0.0%

Table 15–38a. Diagnostic Category Significant Differences – Geometry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	175	4,956	3.4%	96.6%
DC1	DC3	214	4,917	4.2%	95.8%
DC1	DC4	220	4,911	4.3%	95.7%
DC2	DC3	191	4,940	3.7%	96.3%
DC2	DC4	239	4,892	4.7%	95.3%
DC3	DC4	226	4,905	4.4%	95.6%

Note: Z value is 1.94

Table 15–38b. Total Number of Diagnostic Category Significant Differences – Geometry

Number of Significant Differences	Number of Students	Percent of Students
0	4,356	84.9%
1	439	8.6%
2	212	4.1%
3	94	1.8%
4	30	0.6%
5	0	0.0%
6	0	0.0%

Table 15–39a. Diagnostic Category Significant Differences – Algebra II

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	581	5,442	9.6%	90.4%
DC1	DC3	706	5,317	11.7%	88.3%
DC1	DC4	1,143	4,880	19.0%	81.0%
DC2	DC3	246	5,777	4.1%	95.9%
DC2	DC4	419	5,604	7.0%	93.0%
DC3	DC4	219	5,804	3.6%	96.4%

Note: Z value is 1.94

Table 15–39b. Total Number of Diagnostic Category Significant Differences – Algebra II

Number of Significant Differences	Number of Students	Percent of Students
0	4,164	69.1%
1	884	14.7%
2	568	9.4%
3	339	5.6%
4	63	1.0%
5	5	0.1%
6	0	0.0%

Table 15–40a. Diagnostic Category Significant Differences – Reading Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	150	64,703	0.2%	99.8%
DC1	DC3	132	64,721	0.2%	99.8%
DC1	DC4	191	64,662	0.3%	99.7%
DC1	DC5	158	64,695	0.2%	99.8%
DC2	DC3	194	64,659	0.3%	99.7%
DC2	DC4	161	64,692	0.2%	99.8%
DC2	DC5	155	64,698	0.2%	99.8%
DC3	DC4	162	64,691	0.2%	99.8%
DC3	DC5	244	64,609	0.4%	99.6%
DC4	DC5	194	64,659	0.3%	99.7%

Note: Z value is 2.15

Table 15–40b. Total Number of Diagnostic Category Significant Differences – Reading Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	63,442	97.8%
1	1,144	1.8%
2	215	0.3%
3	41	0.1%
4	11	0.0%
5	0	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table 15–41a. Diagnostic Category Significant Differences – Reading/Lit Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	405	138,650	0.3%	99.7%
DC1	DC3	308	138,747	0.2%	99.8%
DC1	DC4	367	138,688	0.3%	99.7%
DC1	DC5	584	138,471	0.4%	99.6%
DC2	DC3	298	138,757	0.2%	99.8%
DC2	DC4	293	138,762	0.2%	99.8%
DC2	DC5	593	138,462	0.4%	99.6%
DC3	DC4	336	138,719	0.2%	99.8%
DC3	DC5	702	138,353	0.5%	99.5%
DC4	DC5	504	138,551	0.4%	99.6%

Note: Z value is 2.15

Table 15–41b. Total Number of Diagnostic Category Significant Differences – Reading/Lit Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	135,572	97.5%
1	2,764	2.0%
2	564	0.4%
3	127	0.1%
4	24	0.0%
5	3	0.0%
6	1	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table 15–42a. Diagnostic Category Significant Differences – Science Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	267	20,411	1.3%	98.7%
DC1	DC3	331	20,347	1.6%	98.4%
DC1	DC4	334	20,344	1.6%	98.4%
DC2	DC3	346	20,332	1.7%	98.3%
DC2	DC4	334	20,344	1.6%	98.4%
DC3	DC4	348	20,330	1.7%	98.3%

Note: Z value is 1.94

Table 15–42b. Total Number of Diagnostic Category Significant Differences – Science Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	19,278	93.2%
1	967	4.7%
2	309	1.5%
3	121	0.6%
4	3	0.0%
5	0	0.0%
6	0	0.0%

Table 15–43a. Diagnostic Category Significant Differences – Science Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,325	50,664	2.5%	97.5%
DC1	DC3	1,291	50,698	2.5%	97.5%
DC1	DC4	1,196	50,793	2.3%	97.7%
DC2	DC3	1,324	50,665	2.5%	97.5%
DC2	DC4	1,203	50,786	2.3%	97.7%
DC3	DC4	1,120	50,869	2.2%	97.8%

Note: Z value is 1.94

Table 15–43b. Total Number of Diagnostic Category Significant Differences – Science Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	47,011	90.4%
1	3,120	6.0%
2	1,287	2.5%
3	519	1.0%
4	52	0.1%
5	0	0.0%
6	0	0.0%

Table 15–44a. Diagnostic Category Significant Differences – Biology

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,203	45,274	2.6%	97.4%
DC1	DC3	1,229	45,248	2.6%	97.4%
DC1	DC4	1,344	45,133	2.9%	97.1%
DC2	DC3	852	45,625	1.8%	98.2%
DC2	DC4	1,514	44,963	3.3%	96.7%
DC3	DC4	1,360	45,117	2.9%	97.1%

Note: Z value is 1.94

Table 15–44b. Total Number of Diagnostic Category Significant Differences – Biology

Number of Significant Differences	Number of Students	Percent of Students
0	41,215	88.7%
1	3,489	7.5%
2	1,347	2.9%
3	387	0.8%
4	37	0.1%
5	2	0.0%
6	0	0.0%

Table 15–45a. Diagnostic Category Significant Differences – Chemistry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	201	2,869	6.5%	93.5%
DC1	DC3	133	2,937	4.3%	95.7%
DC1	DC4	154	2,916	5.0%	95.0%
DC2	DC3	32	3,038	1.0%	99.0%
DC2	DC4	42	3,028	1.4%	98.6%
DC3	DC4	42	3,028	1.4%	98.6%

Note: Z value is 1.94

Table 15–45b. Total Number of Diagnostic Category Significant Differences – Chemistry

Number of Significant Differences	Number of Students	Percent of Students
0	2,679	87.3%
1	231	7.5%
2	111	3.6%
3	45	1.5%
4	4	0.1%
5	0	0.0%
6	0	0.0%

Table 15–46a. Diagnostic Category Significant Differences – Writing Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	70	8,408	0.8%	99.2%
DC1	DC3	58	8,420	0.7%	99.3%
DC1	DC4	75	8,403	0.9%	99.1%
DC1	DC5	94	8,384	1.1%	98.9%
DC2	DC3	45	8,433	0.5%	99.5%
DC2	DC4	67	8,411	0.8%	99.2%
DC2	DC5	60	8,418	0.7%	99.3%
DC3	DC4	54	8,424	0.6%	99.4%
DC3	DC5	60	8,418	0.7%	99.3%
DC4	DC5	69	8,409	0.8%	99.2%

Note: Z value is 2.15

Table 15–46b. Total Number of Diagnostic Category Significant Differences – Writing Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	8,048	94.9%
1	282	3.3%
2	92	1.1%
3	38	0.4%
4	18	0.2%
5	0	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table 15–47a. Diagnostic Category Significant Differences – Writing/Eng Comp Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	191	21,357	0.9%	99.1%
DC1	DC3	212	21,336	1.0%	99.0%
DC1	DC4	278	21,270	1.3%	98.7%
DC1	DC5	289	21,259	1.3%	98.7%
DC2	DC3	181	21,367	0.8%	99.2%
DC2	DC4	281	21,267	1.3%	98.7%
DC2	DC5	280	21,268	1.3%	98.7%
DC3	DC4	261	21,287	1.2%	98.8%
DC3	DC5	227	21,321	1.1%	98.9%
DC4	DC5	273	21,275	1.3%	98.7%

Note: Z value is 2.15

Table 15–47b. Total Number of Diagnostic Category Significant Differences – Writing/Eng Comp Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	19,949	92.6%
1	1,016	4.7%
2	363	1.7%
3	154	0.7%
4	62	0.3%
5	3	0.0%
6	1	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Low numbers of significant differences across diagnostic categories, along with the high disattenuated correlations between categories and exploratory factor analyses discussed in Chapter Seventeen, suggest that some diagnostic categories might be measuring essentially the same construct. While this may be the case in general, when looking at group summary information, diagnostic category scores for individual students can provide useful information to teachers. For example, while 79.8% of students showed no significant differences between Algebra I diagnostic categories, 20.2% of students did. CDT diagnostic category scores for these students along with links to instructional resources are a valuable tool for teachers.

The tables in Appendix D show the significant differences with the familywise Type I error rate at 0.10.

DISTRIBUTION OF BENCHMARK RANGES

As described in Chapter Ten, committees of Pennsylvania educators established preliminary CDT cut scores for grade 5 and above prior to the first operational use. Following the 2010–2011 school year, the preliminary cut scores were revised for the mathematics content-area tests. See Chapter Nineteen of the 2010–2011 technical report for details. Following the 2011–2012 school year, the preliminary cut scores were revised for the reading, science, and writing content-area tests. See Chapter Nineteen of the 2011–2012 technical report for details. Cut points for grades 2 through 4 were interpolated from existing cuts in grade 5 and above prior to the first operational use of CDT tests for grades 3 through 5. See Chapter Nineteen of the 2013–2014 technical report for details. Following the 2014–2015 school year, the cut scores were revised for the mathematics, reading, and writing content-area tests based on the revised PSSA tests. See Chapter Nineteen of the 2015–2016 technical report for details.

The benchmark cuts in place during the 2020–2021 school year determine the color ranges (red/green/blue) in the CDT dynamic reporting suite. The cut scores and standard errors (SE)[4] were used to define ranges as follows: The green range is defined as the scale score cut ± one SE. The red range is defined as the scale minimum (200 for all CDTs except Algebra I, Geometry, Algebra II, Biology, and Chemistry which are 400) to the lower bound of the green range. The blue range is defined as the upper bound of the green range to the scale maximum (2000).

Table 15–48 shows the number and percentage of students in each benchmark range for each full CDT test. Tests with multiple benchmark cuts are broken down to match the grade level of the cuts. Tests that are course-specific are not broken down. All results are based on the cut points in place for the 2020–2021 school year.

[4] The standard error was estimated based on simulations using the operational configuration of the CAT in terms of the content constraints and stopping rules.

Table 15–48. Number and Percent of Students in Each CDT Score Range

CDT	Red *N*	Red Percent	Green *N*	Green Percent	Blue *N*	Blue Percent
Math – G3	17,567	75.7%	4,732	20.4%	892	3.8%
Math – G4	18,852	78.3%	4,461	18.5%	756	3.1%
Math – G5	21,937	76.3%	6,145	21.4%	677	2.4%
Math – G6	26,042	75.1%	7,429	21.4%	1,217	3.5%
Math – G7	28,766	82.3%	5,543	15.9%	638	1.8%
Math – G8	23,195	84.4%	3,843	14.0%	430	1.6%
Math – HS	443	97.6%	10	2.2%	1	0.2%
Algebra I	34,961	81.8%	7,057	16.5%	707	1.7%
Geometry	3,717	72.4%	1,229	24.0%	185	3.6%
Algebra II	3,876	64.4%	1,744	29.0%	403	6.7%
Reading – G3	11,743	59.2%	6,674	33.6%	1,421	7.2%
Reading – G4	12,082	58.4%	7,457	36.0%	1,152	5.6%
Reading – G5	13,262	54.5%	10,050	41.3%	1,012	4.2%
Reading – G6	15,439	60.0%	9,711	37.7%	603	2.3%
Reading – G7	16,144	61.3%	9,653	36.6%	558	2.1%
Reading – G8	16,029	64.7%	8,346	33.7%	402	1.6%
Literature	33,449	53.8%	26,836	43.2%	1,885	3.0%
Science – G3	663	42.6%	655	42.1%	239	15.4%
Science – G4	6,243	47.4%	5,688	43.1%	1,251	9.5%
Science – G5	3,290	55.4%	2,291	38.6%	358	6.0%
Science – G6	5,984	58.5%	3,774	36.9%	470	4.6%
Science – G7	9,888	64.5%	5,133	33.5%	320	2.1%
Science – G8	16,651	64.2%	8,738	33.7%	538	2.1%
Science – HS	453	91.9%	39	7.9%	1	0.2%
Biology	30,944	66.6%	13,649	29.4%	1,884	4.1%
Chemistry	2,192	71.4%	804	26.2%	74	2.4%
Writing – G3	1,593	59.5%	923	34.5%	161	6.0%
Writing – G4	1,347	61.9%	725	33.3%	105	4.8%
Writing – G5	2,204	60.8%	1,291	35.6%	129	3.6%
Writing – G6	3,028	60.7%	1,730	34.7%	227	4.6%
Writing – G7	3,750	59.9%	2,275	36.3%	237	3.8%
Writing – G8	3,596	63.5%	1,863	32.9%	207	3.7%
English Composition	2,452	52.9%	1,853	40.0%	330	7.1%

MULTIPLE ADMINISTRATIONS OF THE SAME CDT TEST

As previously indicated, there are a number of students who took the same full CDT test multiple times. This section focuses on the number of days between administrations and both changes in scale score and benchmark range across a student's first and last administrations.

Table 15–49 shows the summary statistics for the number of days from the first to last administration.

Table 15–49. Summary Statistics for Number of Days between Administrations

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math Grades 3–5	23,597	1	118	158	161.22	209	272
Math Grades 6–HS	29,985	0	115	148	162.16	218	321
Algebra I	12,161	0	119	159	156.53	190	279
Geometry	1,581	6	112	166	165.90	218	267
Algebra II	1,728	2	132	194	171.81	212	274
Reading Grades 3–5	20,663	0	115	140	152.74	188	270
Reading/Lit Grades 6–HS	40,934	0	107	140	146.72	185	277
Science Grades 3–5	6,699	0	112	137	142.46	162	268
Science Grades 6–HS	15,598	0	118	144	150.82	176	269
Biology	12,584	0	105	154	151.12	196	266
Chemistry	849	0	71	133	126.16	147	274
Writing Grades 3–5	2,823	0	116	133	143.68	176	249
Writing/Eng Comp Gr 6–HS	5,924	0	104	139	141.30	173	266

Table 15–50 shows the summary statistics for the change in total scale score from the first to last administration.

Table 15–50. Summary Statistics for Change in Total Scale Score between Administrations

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math Grades 3–5	23,597	-845	6	68	65.87	130	621
Math Grades 6–HS	29,985	-626	-28	35	29.76	94	606
Algebra I	12,161	-542	-45	29	22.48	97	584
Geometry	1,581	-458	-16	44	34.66	102	468
Algebra II	1,728	-546	-8	62	55.76	125	734
Reading Grades 3–5	20,663	-670	-33	28	27.06	90	688
Reading/Lit Grades 6–HS	40,934	-631	-60	1	-2.05	60	544
Science Grades 3–5	6,699	-805	-39	28	21.99	89	556
Science Grades 6–HS	15,598	-500	-53	10	5.64	68	606
Biology	12,584	-481	-26	43	39.52	108	651
Chemistry	849	-308	-35	28	23.53	85	340
Writing Grades 3–5	2,823	-549	-34	32	29.50	96	526
Writing/Eng Comp Gr 6–HS	5,924	-510	-53	10	3.61	67	583

Tables 15–51a through 15–51m show the changes in benchmark range from the first to last administration. For example, 4,320 students who scored in the red range on the first administration of the Math Grades 3–5 test scored in the green range on the last administration.

Table 15–51a. Change in Benchmark Range between First and Last Administrations – Math Grades 3–5

	Red–last test	Green – last test	Blue – last test
Red–first test	15,158	4,320	307
Green–first test	600	2,145	679
Blue–first test	65	83	240

Table 15–51b. Change in Benchmark Range between First and Last Administrations – Math Grades 6–HS

	Red – last test	Green – last test	Blue – last test
Red–first test	21,780	3,684	133
Green–first test	756	2,586	735
Blue–first test	16	48	247

Table 15–51c. Change in Benchmark Range between First and Last Administrations – Algebra I

	Red – last test	Green – last test	Blue – last test
Red–first test	9,087	1,760	54
Green–first test	276	765	177
Blue–first test	3	9	30

Table 15–51d. Change in Benchmark Range between First and Last Administrations – Geometry

	Red – last test	Green – last test	Blue – last test
Red–first test	961	335	10
Green–first test	35	174	56
Blue–first test	0	3	7

Table 15–51e. Change in Benchmark Range between First and Last Administrations – Algebra II

	Red – last test	Green – last test	Blue – last test
Red–first test	818	426	30
Green–first test	81	250	77
Blue–first test	0	12	34

Table 15–51f. Change in Benchmark Range between First and Last Administrations – Reading Grades 3–5

	Red – last test	Green – last test	Blue – last test
Red–first test	9,506	2,724	69
Green–first test	1,158	5,403	958
Blue–first test	48	308	489

Table 15–51g. Change in Benchmark Range between First and Last Administrations – Reading/Lit Grades 6–HS

	Red – last test	Green – last test	Blue – last test
Red–first test	20,185	3,638	13
Green–first test	3,309	12,149	766
Blue–first test	8	485	381

Table 15–51h. Change in Benchmark Range between First and Last Administrations – Science Grades 3–5

	Red – last test	Green – last test	Blue – last test
Red–first test	2,564	936	43
Green–first test	468	1,794	418
Blue–first test	45	152	279

Table 15–51i. Change in Benchmark Range between First and Last Administrations – Science Grades 6–HS

	Red – last test	Green – last test	Blue – last test
Red–first test	8,655	1,778	9
Green–first test	1,054	3,487	369
Blue–first test	4	86	156

Table 15–51j. Change in Benchmark Range between First and Last Administrations – Biology

	Red – last test	Green – last test	Blue – last test
Red–first test	6,901	2,522	150
Green–first test	422	1,843	634
Blue–first test	0	16	96

Table 15–51k. Change in Benchmark Range between First and Last Administrations – Chemistry

	Red – last test	Green – last test	Blue – last test
Red–first test	554	161	4
Green–first test	33	74	19
Blue–first test	0	0	4

Table 15–51l. Change in Benchmark Range between First and Last Administrations – Writing Grades 3–5

	Red – last test	Green – last test	Blue – last test
Red–first test	1,390	436	14
Green–first test	157	625	90
Blue–first test	9	43	59

Table 15–51m. Change in Benchmark Range between First and Last Administrations – Writing/Eng Comp Grades 6–HS

	Red – last test	Green – last test	Blue – last test
Red–first test	3,051	585	4
Green–first test	418	1,535	157
Blue–first test	2	53	119

CHAPTER SIXTEEN: RELIABILITY

This chapter addresses the reliability of Classroom Diagnostic Tools (CDT) test scores. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the general notion of reliability/precision refers to

> the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported (p.33).

Frisbie (2005) highlighted several elements of reliability. 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 *Standards'* definition yet are still important for test users to understand. While freedom from measurement error is very important, 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 Seventeen.

Another noteworthy issue is that multiple sources of error exist (e.g., the day of testing, the items used). However, most widely used reliability indices only reflect a single type of error. Consequently, it is important for test users to understand what specific type of error is being considered in a reliability study, and equally, if not more importantly, what types are not.

Understanding the distinction between relative error and absolute error is also important, as 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. Understanding examinee rank-order stability is important; however, such stability might be well achieved even when the specific score values are considerably different. 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.

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 (SEMs and CSEMs)
- Decision consistency

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: 1) variance in true scores and 2) variance due to the imperfections in the measurement process. Put differently, total variance equals true score variance plus error variance.[1]

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

[1] A covariance term is not required, as true scores and error are assumed to be uncorrelated in classical test theory.

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 (i.e., all measurement error). If the index had a value of 1.0, scores would be perfectly consistent (i.e., 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 (student health, testing environment, etc.).

COEFFICIENT ALPHA

Although a number of reliability indices exist, one of the most frequently reported for achievement tests is coefficient alpha. For example, both PSSA and Keystone programs report alpha.

FORMULA FOR ALPHA

Consider the following data matrix representing the scores of persons (rows) on items (columns):

Table 16–1. Person × Item Score (*Xpi*) Infinite (Population-Universe) Matrix

Person	Item 1	Item 2	. . .	Item *i*	. . .	Item *k*
Person 1	*Y*11	*Y*12	. . .	*Y*1*i*	. . .	*X*1*k*
Person 2	*Y*21	*Y*22	. . .	*Y*2*i*	. . .	*X*2*k*
. . . .	. . .	. . .	. . .	. . .	. . .	. . .
Person *p*	*Yp*1	*Yp*2	. . .	*Ypi*	. . .	*Xpk*
. . . .	. . .	. . .	. . .	. . .	. . .	. . .
Person *N*	*YN*1	*YN*2	. . .	*YNi*	. . .	*XNk*

Note: Adapted from Cronbach and Shavelson (2004).

Then, a general computational formula for alpha is as follows:

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma_{Yi}^2}{\sigma_X^2}\right),$$

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*.

Examination of the formula for alpha indicates why the coefficient is not appropriate for CDT. In the case of CDT, tests are adaptive. Each student takes a unique set of test items rather than the same fixed form. A person item score matrix for CDT analogous to Table 16–1 would include all items in the available item pool (over 5,000 in some cases). Each student takes only a small subset of items (48–60) from the available pool. Summing the variance of more than 5,000 item scores and dividing by the variance of test scores based on 48–60 items is not appropriate. Therefore, a measure of reliability other than alpha must be used for CDT.

SPLIT-HALF RELIABILITY

Like alpha, split-half 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 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:

Table 16–2. 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 split-half reliability index for those test scores was high, this would suggest that Test Two would provide a very similar rank ordering of the students if they had taken it instead. If split-half reliability was low, dissimilar rank orderings would likely be observed—again, relative-error variance is reflected.

CALCULATION OF SPLIT-HALF RELIABILITY

To determine split-half reliability for a given CDT test, such as Biology, each administration of the test was split into two halves. Each item's difficulty was considered in the split so the halves represent approximately equivalent alternative forms. Rasch ability estimates were then calculated for each of the two halves. Then, Pearson correlation was computed between the Rasch ability estimates from the two halves. Finally, the Pearson correlation was adjusted for test length using the Spearman-Brown prediction formula as described below.

$$\text{Split-Half reliability} = \frac{2r}{1+r} \qquad \text{where } r = \text{Pearson correlation}$$

Split-half reliability is related to coefficient alpha in that alpha is often interpreted as the mean of all possible split-half coefficients.

FURTHER INTERPRETATIONS

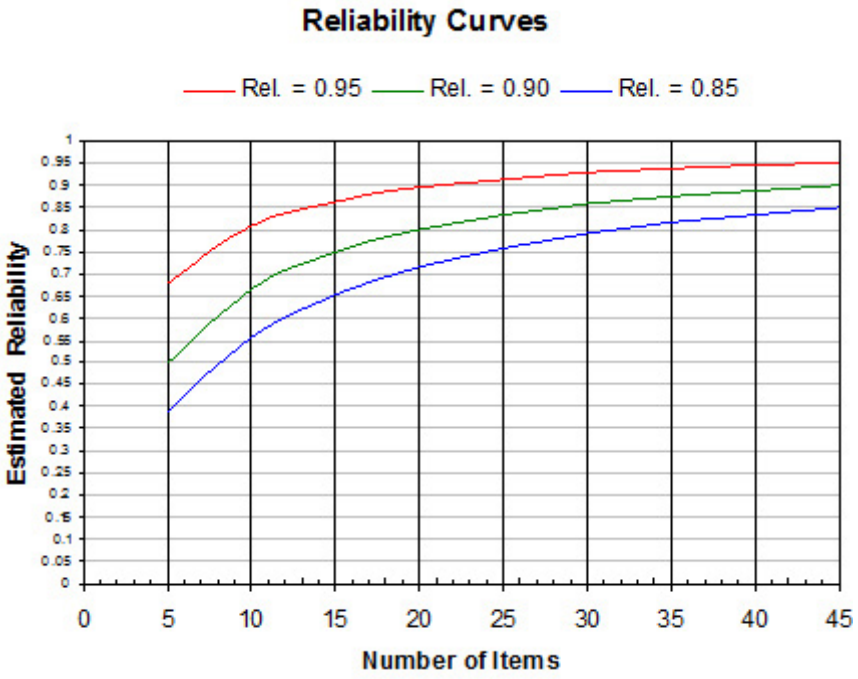
What reliability value is considered high enough? What values are considered too low? Although frequently asked for, any rules of thumb for interpreting the magnitude of reliability indices are mostly arbitrary. One approach is to research the reliabilities from similar testing instruments to see what values are commonly observed. For 2021 PSSA tests in Mathematics, English Language Arts (ELA), and Science, reliability coefficients ranged from 0.82 to 0.92. For spring 2021 Keystone exams in Algebra I, Literature, and Biology, reliability coefficients were 0.92, 0.92, and 0.93, respectively. For many other state assessment programs, reliabilities in the low 0.90s are usually the highest 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.

DIAGNOSTIC CATEGORY 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 16–1 illustrates this relationship for a hypothetical 45-item 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 diagnostic category score would be expected to have a score reliability of just over 0.65. The use of the Spearman-Brown prediction formula assumes all items are exchangeable, which, in practice, they may not be. While such a chart may not perfectly model actual diagnostic category reliability, the intent is to illustrate the substantial impact that limited numbers of items can have on diagnostic category score reliability.

Figure 16–1. Example of the Relationship between Test Length and Reliability



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 is somewhat unwieldy. A beginning point for understanding the concept is as follows. If everyone being tested had the same true score,[2] there would still be some variation in observed scores due to imperfections in the measurement process, such as random differences in attention during instruction or concentration during testing, the sampling of test items, etc. The standard error is defined as the standard deviation[3] of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents very important information for test score users.

[2] True score is the score the person would receive if the measurement process were perfect.
[3] The standard deviation of a distribution is a measure of the dispersion of the observations. For the normal distribution, about 16 percent of the observations are more than one standard deviation above the mean.

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.

TRADITIONAL SEM CONFIDENCE INTERVALS

The SEM is an index of the random variability in test scores in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the precision of individual tests scores. SEMs help place reasonable limits (Gulliksen, 1950) around observed scores through construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, X, and adding and subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between ± 1 SEM about two-thirds of the time.[4] For ±2 SEM confidence intervals, the percentage increases to about 95 percent.

FURTHER INTERPRETATIONS

ONE SEM FOR ALL TEST SCORES

The SEM approach described above only provides a single numerical estimate for constructing the confidence intervals for examinees regardless of their score levels. In reality, however, such confidence intervals vary according to one's score. Consequently, care should be taken when using the SEM for students with extreme scores. An alternate approach is described in the next section that conditions the SEM on a student's score estimate.

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.

SCALE SCORE METRIC

The SEM approach is calculated using scale scores, and as such, the resulting confidence interval bands are in the scale score metric.

TYPE OF ERROR REFLECTED

The interpretation of the SEM should be driven by the type of score reliability that underpins it. So, the CDT 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 CDT. 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.

[4] Some prefer the following interpretation: if a student were tested an infinite number of times, the ± 1 SEM confidence intervals constructed for each score would capture the student's true score 68 percent of the time.

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 1078 and the SEM band was 1038 to 1118, then a student would be likely to receive a score somewhere between 1038 and 1118 if he or she took a different version of the test without additional instruction.

RESULTS AND OBSERVATIONS

Split-half reliability coefficients and associated (traditional) SEMs for CDT tests are presented in Table 16–3. Values were derived using the operational data from the 2020–2021 school year. The results are presented for total scores and each diagnostic category score. The statistics reported include number of students tested (*N*), mean scale score, standard deviation of scale score, split-half reliability, and traditional standard error of measurement (SEM) in the scale score metric.

Table 16–3. CDT Reliabilities

CDT	Score	Average Number of Points	*N*	Scale Score Mean	Scale Score SD	Split-Half Reliability	SEM in Scale Score Metric
Math Grades 3–5	**Total**	51.6	76,019	786.994	167.205	0.943	40.0
Math Grades 3–5	Numbers and Operations	12.9	76,019	786.051	194.501	0.824	81.5
Math Grades 3–5	Algebraic Concepts	12.9	76,019	796.652	186.104	0.807	81.8
Math Grades 3–5	Geometry	12.9	76,019	764.526	172.000	0.772	82.1
Math Grades 3–5	Measurement, Data, and Probability	12.9	76,019	790.053	185.755	0.807	81.7
Math Grades 6–HS	**Total**	52.1	97,557	938.570	158.616	0.941	38.6
Math Grades 6–HS	Numbers and Operations	13.0	97,557	945.080	193.028	0.838	77.7
Math Grades 6–HS	Algebraic Concepts	13.1	97,557	940.942	180.987	0.814	78.1
Math Grades 6–HS	Geometry	13.0	97,557	941.843	162.717	0.775	77.3
Math Grades 6–HS	Measurement, Data, and Probability	13.1	97,557	929.624	180.207	0.810	78.5
Algebra I	**Total**	52.7	42,725	994.288	158.091	0.937	39.7
Algebra I	Operations with Real Numbers and Expressions	13.3	42,725	983.937	210.739	0.861	78.6
Algebra I	Linear Equations & Inequalities	13.2	42,725	1009.862	163.502	0.763	79.7
Algebra I	Functions & Coordinate Geometry	13.1	42,725	1007.272	166.906	0.778	78.7
Algebra I	Data Analysis	13.2	42,725	981.354	184.719	0.813	79.8
Geometry	**Total**	52.3	5,131	1077.897	151.388	0.935	38.7
Geometry	Geometric Properties	13.0	5,131	1076.254	165.193	0.785	76.6
Geometry	Congruence, Similarity, & Proofs	13.1	5,131	1080.871	170.282	0.783	79.2
Geometry	Coordinate Geometry and Right Triangles	13.1	5,131	1085.838	180.662	0.813	78.1
Geometry	Measurement	13.1	5,131	1070.991	180.724	0.811	78.6

Table 16–3 (continued). CDT Reliabilities

CDT	Score	Average Number of Points	*N*	Scale Score Mean	Scale Score SD	Split-Half Reliability	SEM in Scale Score Metric
Algebra II	**Total**	52.7	6,023	1174.279	151.936	0.933	39.5
Algebra II	Operations with Complex Numbers	13.6	6,023	1250.599	232.325	0.876	81.9
Algebra II	Non-linear Expressions & Equations	13.0	6,023	1169.962	184.659	0.825	77.2
Algebra II	Functions	13.0	6,023	1163.182	164.153	0.770	78.7
Algebra II	Data Analysis	13.1	6,023	1124.191	165.497	0.779	77.7
Reading Grades 3–5	**Total**	57.6	64,853	790.389	166.594	0.927	45.0
Reading Grades 3–5	Key Ideas and Details-Literature Text	11.8	64,853	788.316	203.705	0.735	104.9
Reading Grades 3–5	Key Ideas and Details-Informational Text	11.5	64,853	781.572	198.714	0.718	105.6
Reading Grades 3–5	Craft and Structure-Literature Text	11.6	64,853	804.311	186.336	0.665	107.9
Reading Grades 3–5	Craft and Structure-Informational Text	11.6	64,853	783.605	200.529	0.720	106.1
Reading Grades 3–5	Vocabulary Acquisition and Use	11.1	64,853	783.749	212.160	0.750	106.1
Reading/Lit Grades 6–HS	**Total**	56.8	139,055	946.449	164.155	0.927	44.2
Reading/Lit Grades 6–HS	Key Ideas and Details-Literature Text	11.8	139,055	942.376	196.819	0.723	103.6
Reading/Lit Grades 6–HS	Key Ideas and Details-Informational Text	11.5	139,055	949.204	196.319	0.720	103.9
Reading/Lit Grades 6–HS	Craft and Structure-Literature Text	11.3	139,055	947.735	181.772	0.672	104.0
Reading/Lit Grades 6–HS	Craft and Structure-Informational Text	11.4	139,055	941.506	203.409	0.740	103.8
Reading/Lit Grades 6–HS	Vocabulary Acquisition and Use	10.8	139,055	951.243	223.952	0.762	109.2
Science Grades 3–5	**Total**	51.5	20,678	735.942	164.841	0.935	41.9
Science Grades 3–5	The Nature of Science	12.9	20,678	726.022	185.918	0.777	87.8
Science Grades 3–5	Biological Sciences	12.9	20,678	731.249	185.319	0.788	85.4
Science Grades 3–5	Physical Sciences	12.9	20,678	740.681	179.100	0.771	85.8
Science Grades 3–5	Earth and Space Sciences	12.8	20,678	737.865	174.935	0.755	86.5
Science Grades 6–HS	**Total**	52.2	51,989	843.061	154.249	0.930	40.8
Science Grades 6–HS	The Nature of Science	13.0	51,989	837.517	187.992	0.804	83.2
Science Grades 6–HS	Biological Sciences	13.1	51,989	840.144	185.063	0.799	83.0
Science Grades 6–HS	Physical Sciences	13.1	51,989	853.691	166.545	0.745	84.0
Science Grades 6–HS	Earth and Space Sciences	13.0	51,989	842.745	164.421	0.741	83.7

Table 16–3 (continued). CDT Reliabilities

CDT	Score	Average Number of Points	*N*	Scale Score Mean	Scale Score SD	Split-Half Reliability	SEM in Scale Score Metric
Biology	**Total**	52.4	46,477	938.774	150.397	0.925	41.2
Biology	Basic Biological Principles/ Chemical Basis for Life	13.1	46,477	936.847	185.324	0.795	83.9
Biology	Bioenergetics/Homeostasis and Transport	13.1	46,477	948.764	154.223	0.706	83.6
Biology	Cell Growth and Reproduction/ Genetics	13.1	46,477	950.236	162.647	0.735	83.7
Biology	Theory of Evolution/Ecology	13.1	46,477	921.893	189.251	0.805	83.5
Chemistry	**Total**	52.5	3,070	982.741	115.685	0.873	41.2
Chemistry	Properties and Classification of Matter	13.1	3,070	951.416	177.368	0.775	84.2
Chemistry	Atomic Structure and the Periodic Table	13.2	3,070	1007.505	125.021	0.536	85.1
Chemistry	The Mole and Chemical Bonding	13.1	3,070	993.050	130.807	0.597	83.1
Chemistry	Chemical Relationships and Reactions	13.2	3,070	985.019	137.478	0.610	85.9
Writing Grades 3–5	**Total**	54.8	8,478	775.020	178.741	0.950	40.2
Writing Grades 3–5	Quality of Writing: Focus and Organization	11.0	8,478	763.741	209.908	0.799	94.1
Writing Grades 3–5	Quality of Writing: Content and Style	11.0	8,478	771.569	202.087	0.786	93.5
Writing Grades 3–5	Quality of Writing: Editing	10.9	8,478	769.392	198.384	0.780	93.0
Writing Grades 3–5	Conventions: Punctuation, Capitalization, and Spelling	10.9	8,478	781.514	187.886	0.756	92.8
Writing Grades 3–5	Conventions: Grammar and Sentence Formation	11.0	8,478	778.857	202.511	0.790	92.7
Writing/Eng Comp Gr 6–HS	**Total**	55.6	21,548	898.789	174.647	0.947	40.0
Writing/Eng Comp Gr 6–HS	Quality of Writing: Focus and Organization	11.1	21,548	889.633	209.993	0.804	93.0
Writing/Eng Comp Gr 6–HS	Quality of Writing: Content and Style	11.1	21,548	900.067	201.848	0.790	92.6
Writing/Eng Comp Gr 6–HS	Quality of Writing: Editing	11.1	21,548	891.286	198.416	0.793	90.3
Writing/Eng Comp Gr 6–HS	Conventions: Punctuation, Capitalization, and Spelling	11.1	21,548	915.657	202.599	0.798	91.1
Writing/Eng Comp Gr 6–HS	Conventions: Grammar and Sentence Formation	11.1	21,548	891.896	196.167	0.787	90.6

The overall test score reliability values are high and similar to those reported for PSSA and Keystone Exams. The reliabilities at the diagnostic category level are lower due to the fact that each diagnostic category contains fewer items.

RASCH CONDITIONAL STANDARD ERRORS OF MEASUREMENT

The CSEM also indicates the degree of measurement error in scale score units, but varies as a function of a student's actual scale 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 the conditional standard error of measurement and $I(\hat{\beta}_n)$ is the 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 ($\hat{\beta}_n$) metric. The conditional standard error on the scale 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 scale scores:

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

Chapter Eleven provides the linear transformation formulas for each of the CDT content areas.

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 scale 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. On fixed form tests, the magnitude of the CSEM values is often "U" shaped, with larger CSEM values associated with lower and higher scores. With a fixed set of items, there is less information for students scoring at the extremes, and CSEM is inversely related to the information function (the more information, the lower the CSEM). Given that CDT tests are adaptive, this "U" shape tends to be less pronounced as students are presented with items targeted at their level. While there is some "U" shape at the extreme ends of the vertical scale, there is a much larger area on the scale where CSEMs are relatively flat compared to fixed form tests. The adaptive tests allow for greater information and, therefore, lower CSEMs across a wide range of the vertical scale.

GROUP SPECIFIC

Assuming reasonable model-data fit—as explored in Chapter Eight—the Rasch based CSEMs (conditioned on score level) should not vary across groups.

SCALE SCORE METRIC

The CSEM and associated confidence interval bands are in the scale score metric.

TYPE OF ERROR REFLECTED

The CSEMs reported in the dynamic reporting suite are the Rasch-based conditional standard errors of measurement described above. Score report content is considered in greater detail in Chapter Fourteen.

RESULTS AND OBSERVATIONS

Figures 16–2 through 16–14 show the average Rasch CSEMs associated with various scale score ranges based on operational data from the 2020–2021 school year. The values are fairly consistent across a large range of scores on the vertical scale. The values increase at the low and high ends of the scale score range.

Figure 16–2. Average Conditional Standard Errors for Math Grades 3–5

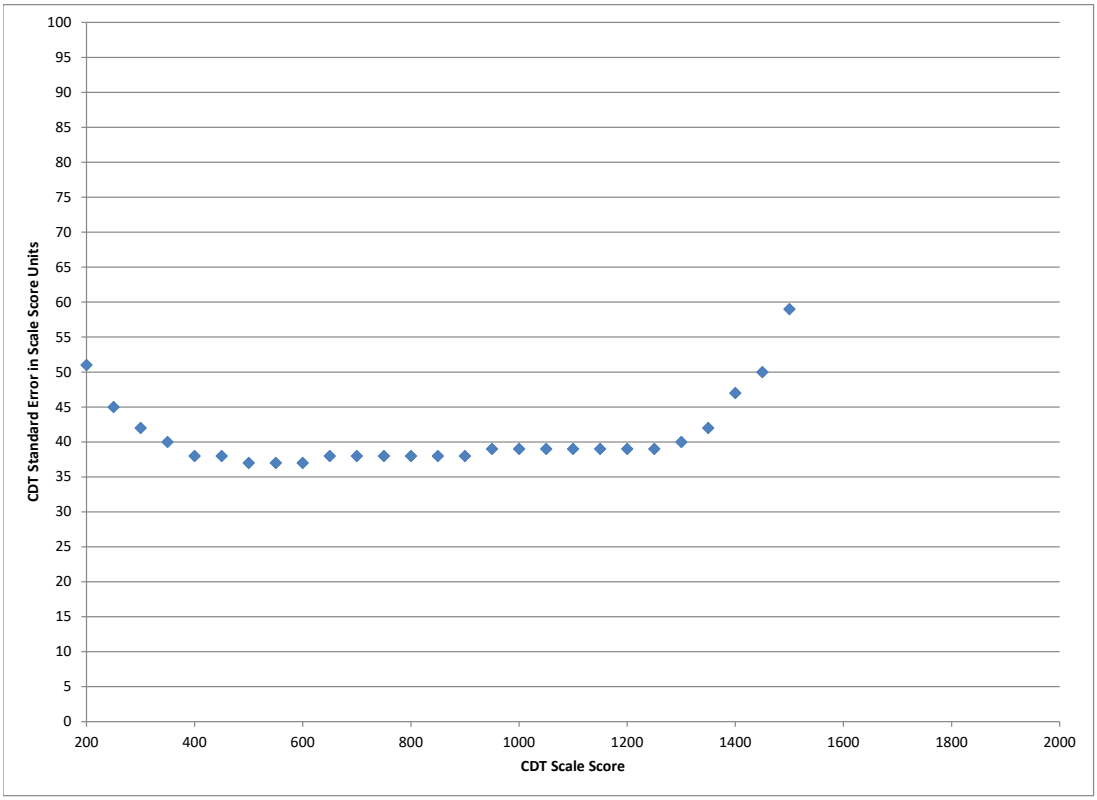


Figure 16–3. Average Conditional Standard Errors for Math Grades 6–HS

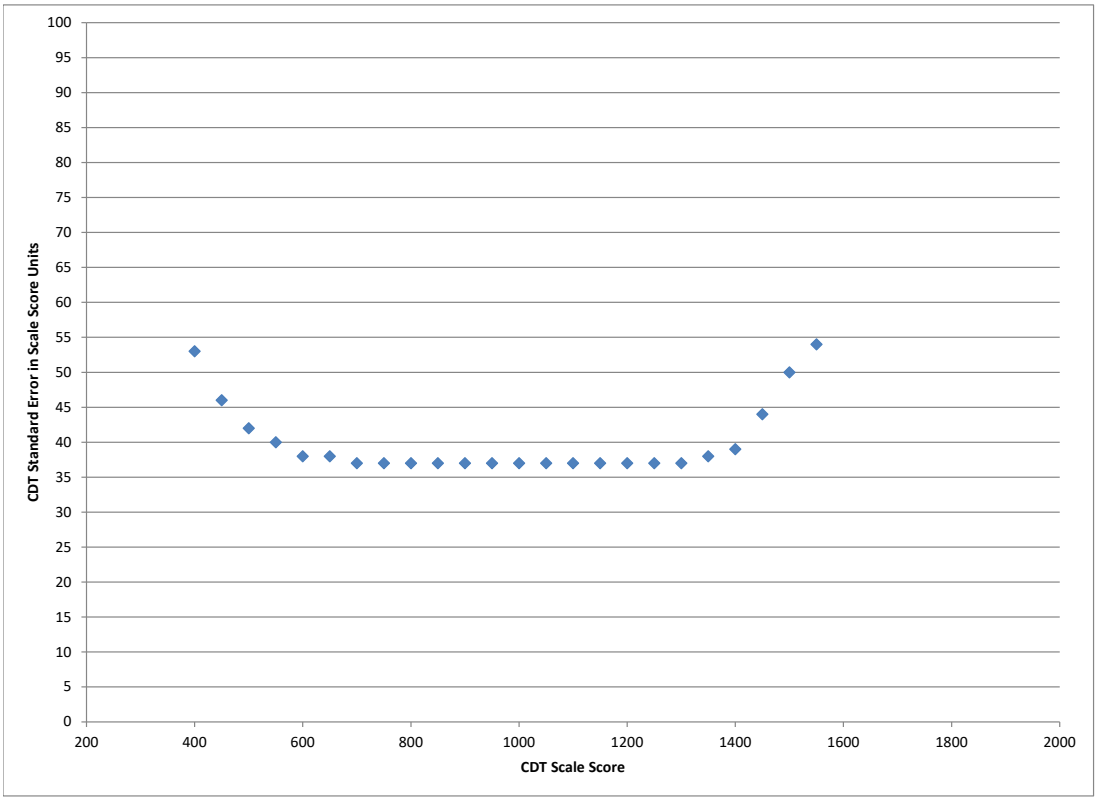


Figure 16–4. Average Conditional Standard Errors for Algebra I

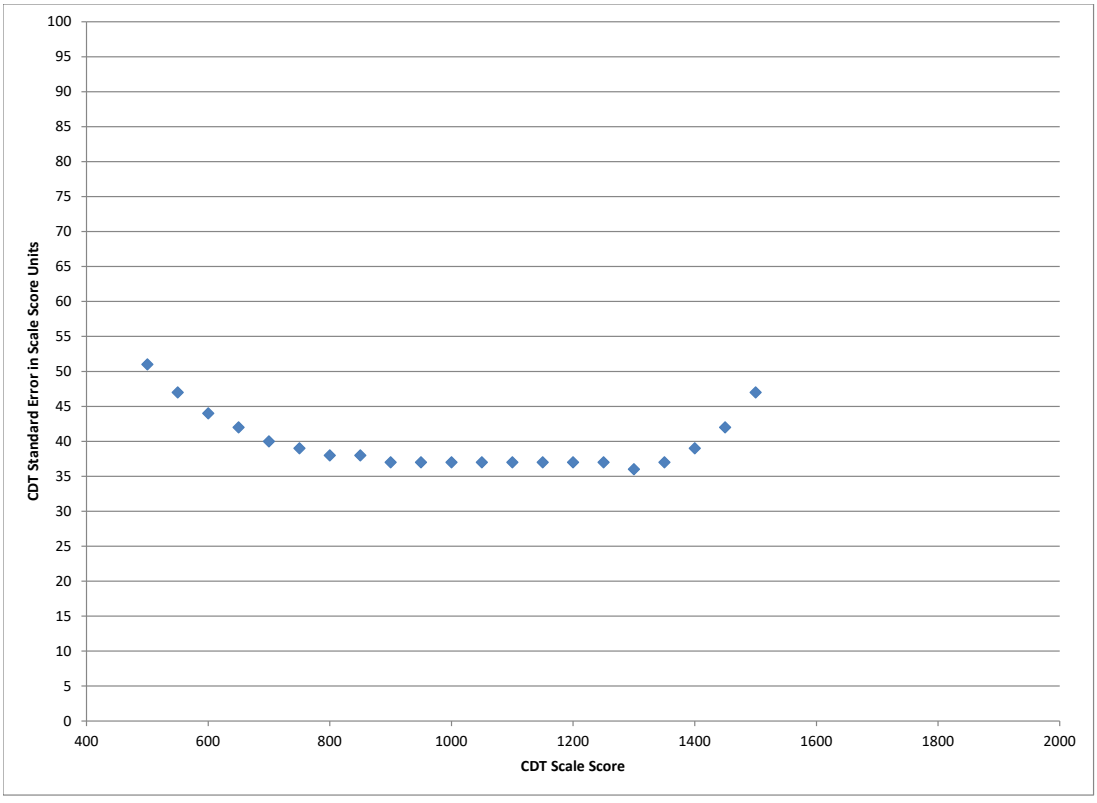


Figure 16–5. Average Conditional Standard Errors for Geometry

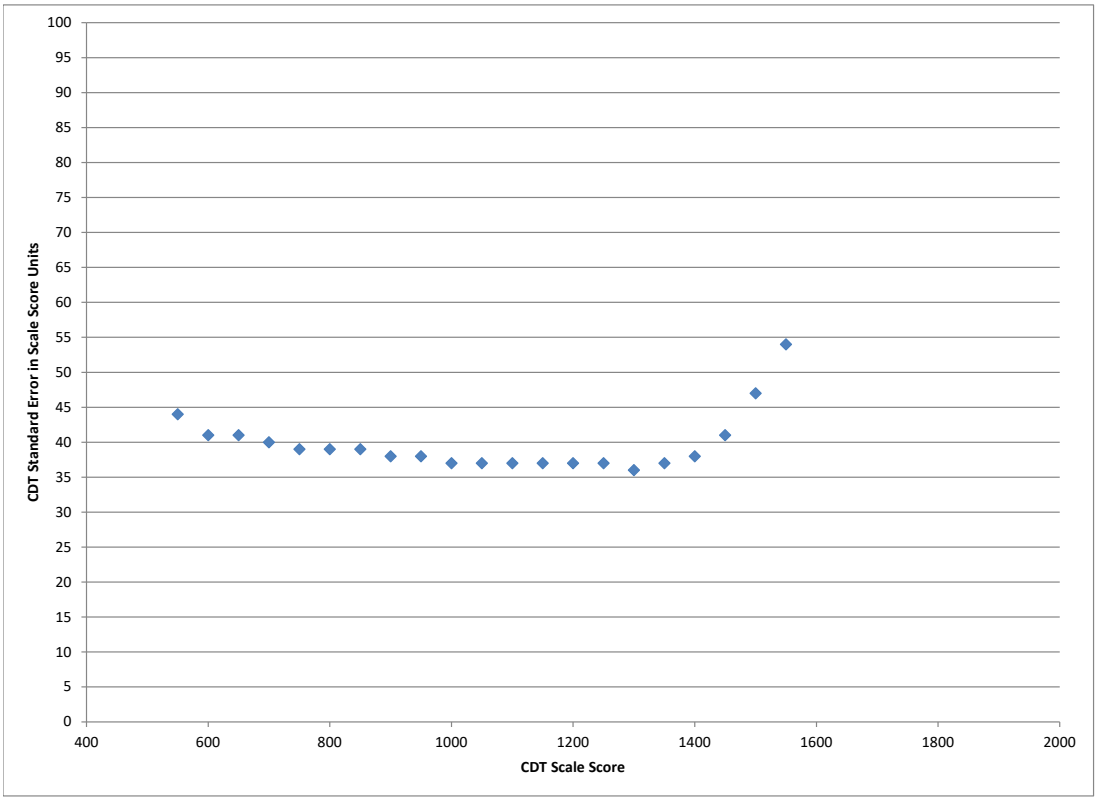


Figure 16–6. Average Conditional Standard Errors for Algebra II

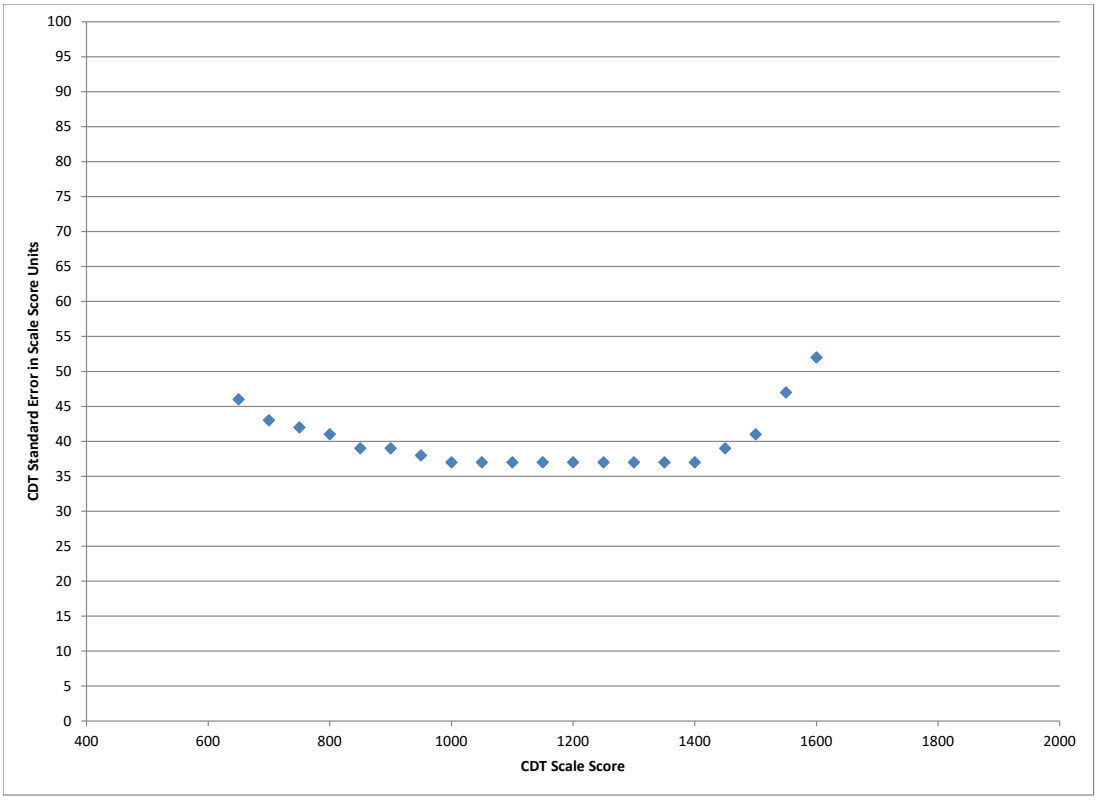


Figure 16–7. Average Conditional Standard Errors for Reading Grades 3–5

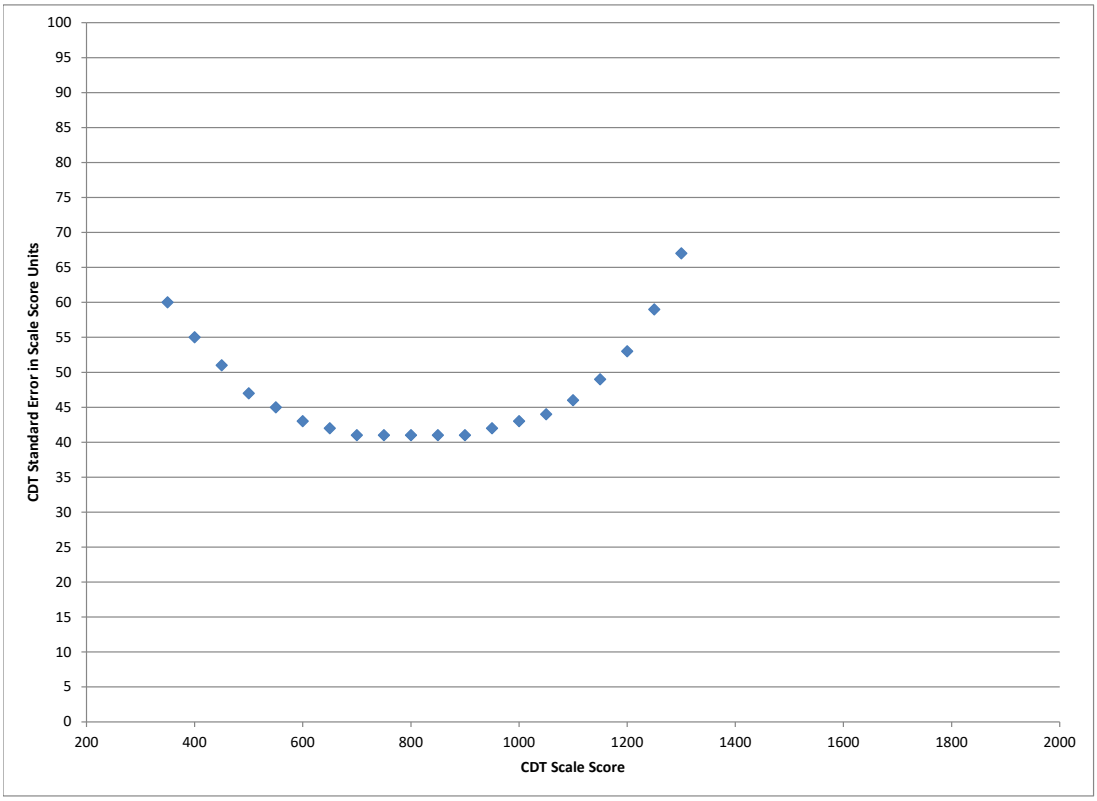
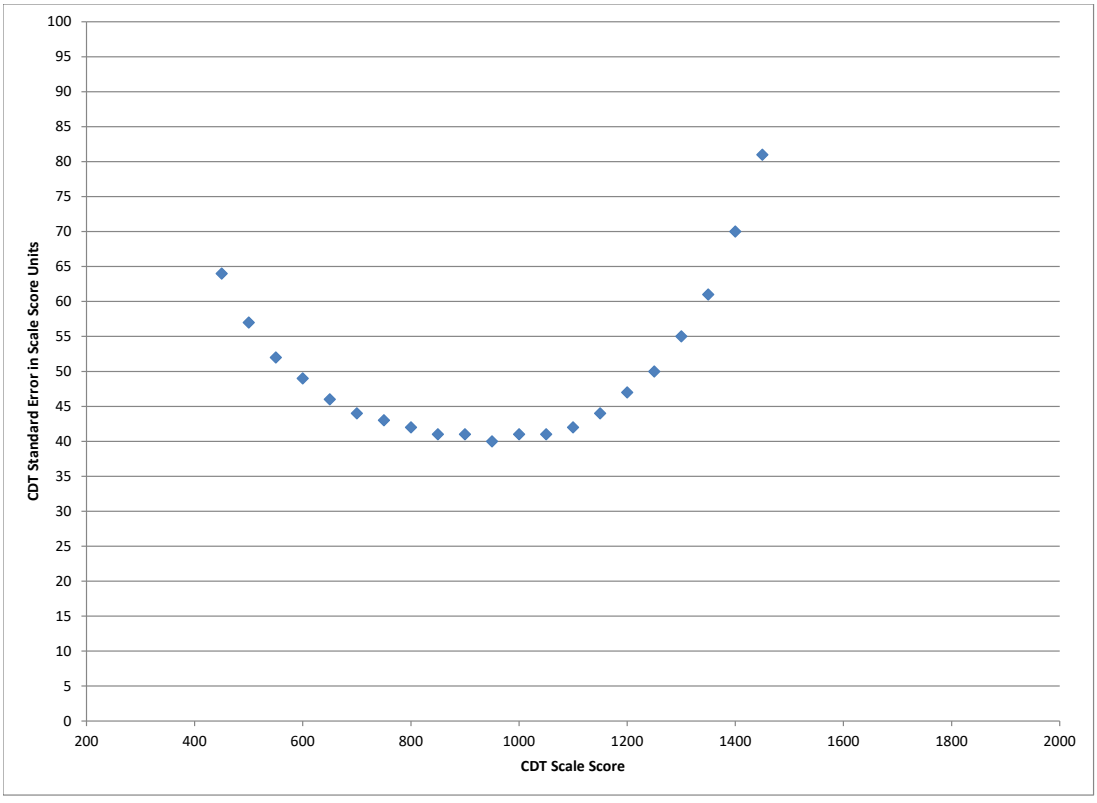


Figure 16–8. Average Conditional Standard Errors for Reading/Lit Grades 6–HS



CSEMs tend to be higher in the reading content area. This is due to the fact that CDT Reading Grades 3–5 and CDT Reading/Lit Grades 6–HS are passage-based. The items from a selected passage may not be as closely targeted to the student's level as when individual items are selected one at a time. For more information on adaptive selection of passages, see Chapter Thirteen.

Figure 16–9. Average Conditional Standard Errors for Science Grades 3–5

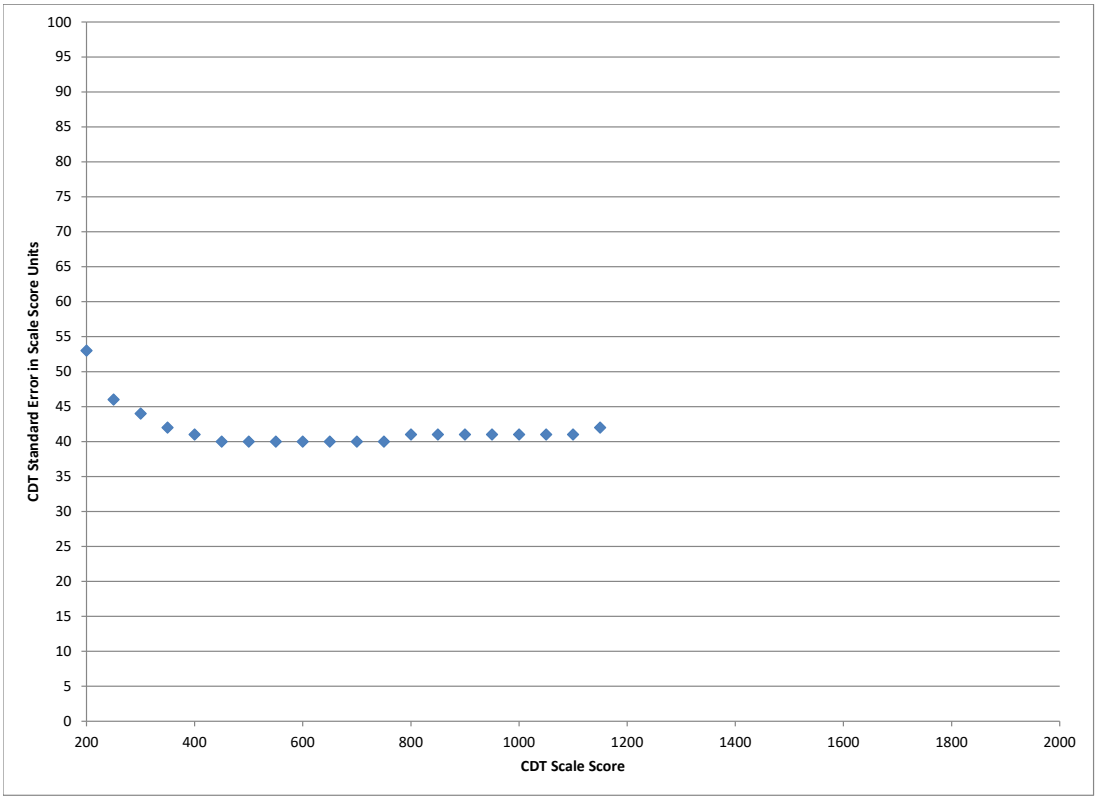


Figure 16–10. Average Conditional Standard Errors for Science Grades 6–HS

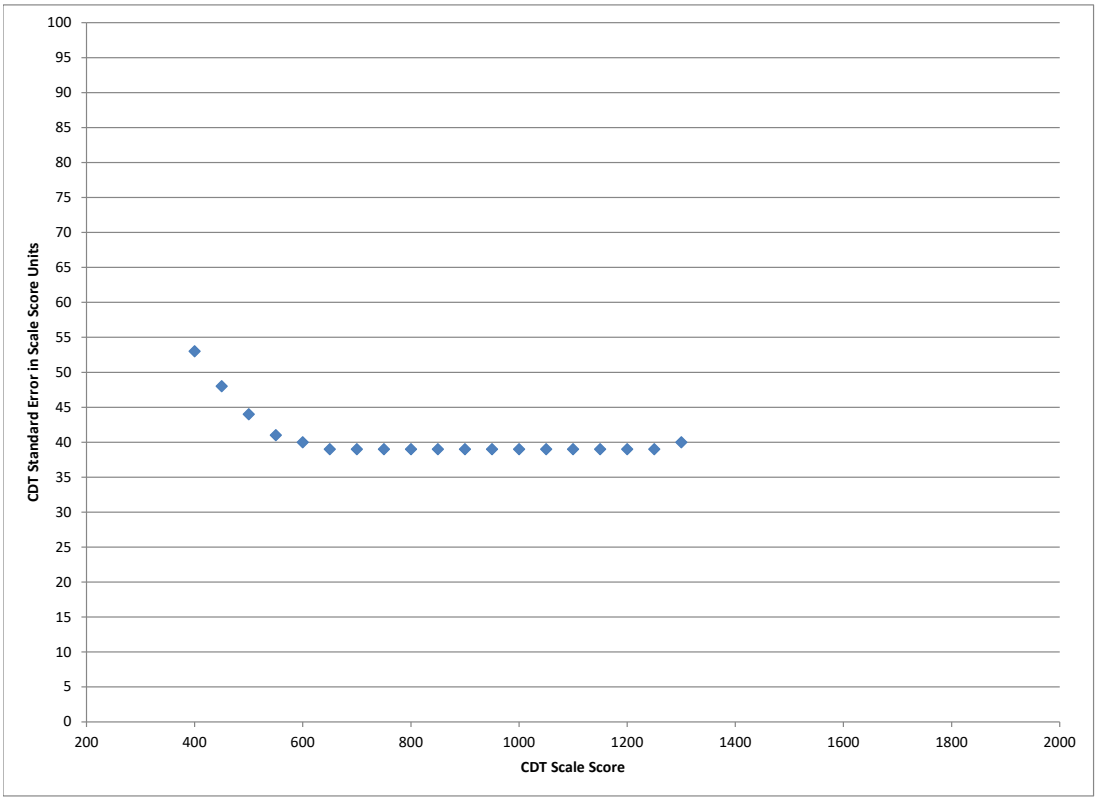


Figure 16–11. Average Conditional Standard Errors for Biology

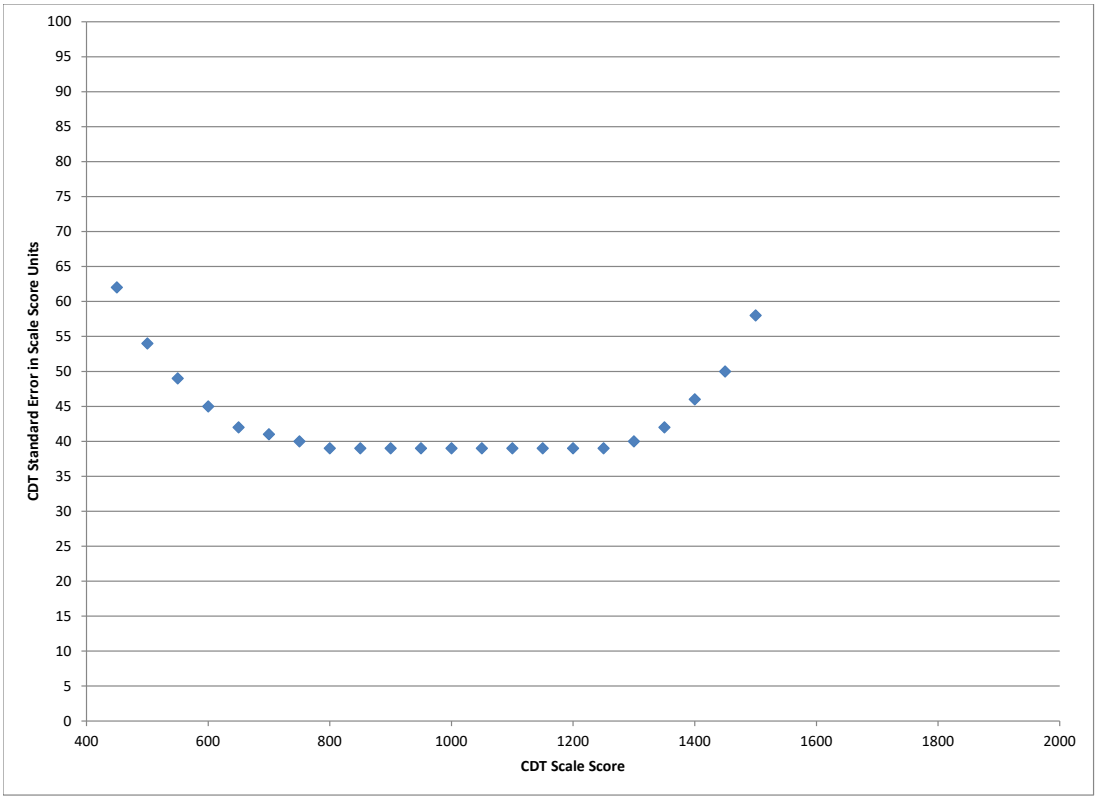


Figure 16–12. Average Conditional Standard Errors for Chemistry

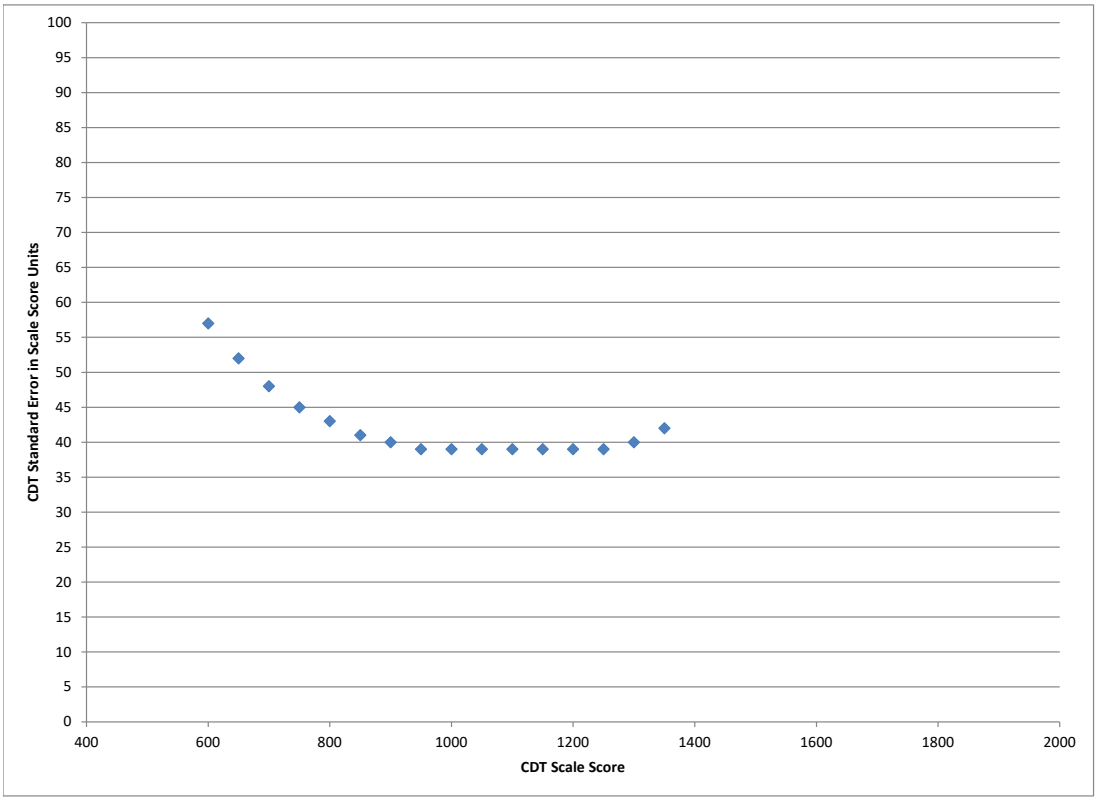


Figure 16–13. Average Conditional Standard Errors for Writing Grades 3–5

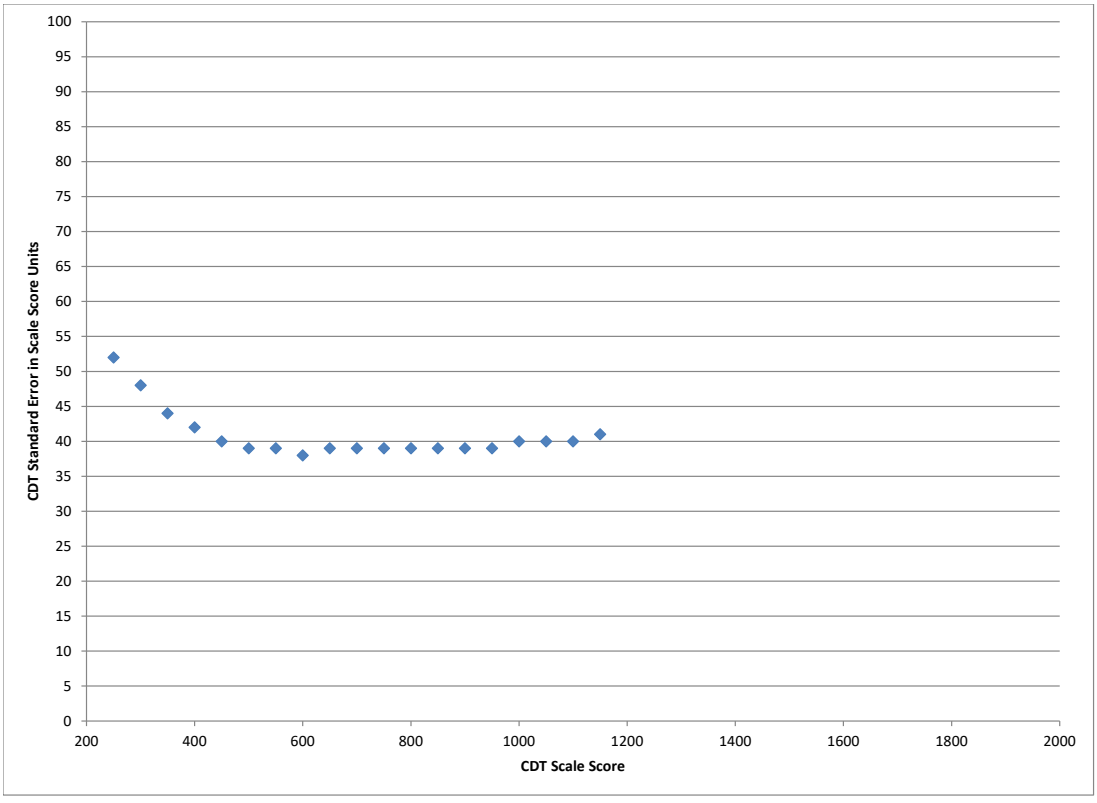
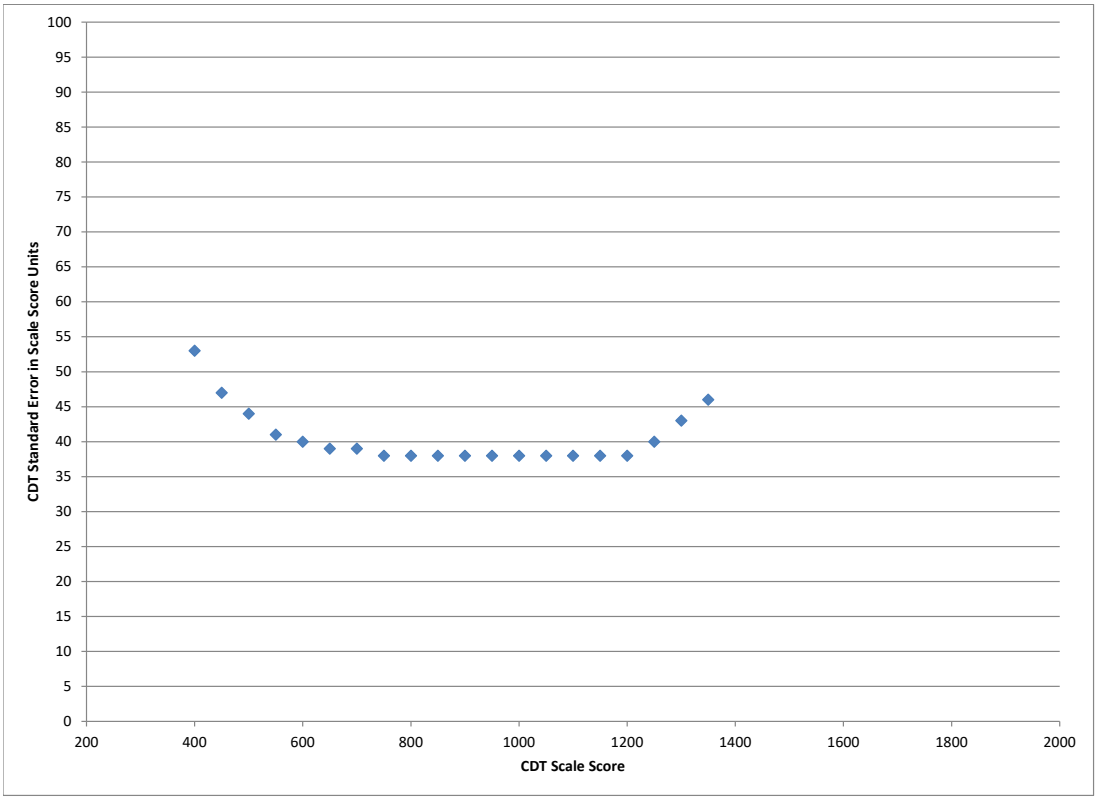


Figure 16–14. Average Conditional Standard Errors for Writing/Eng Comp Grades 6–HS



DECISION CONSISTENCY

Classification decision consistency refers to the degree to which the achievement level for each student can be replicated upon retesting using an equivalent form (Huynh, 1976). While CDT is designed to be administered multiple times in the school year to gauge progress following instruction, retesting in the context of decision consistency refers to retesting shortly after testing without additional instruction.

In a standards-based testing program, there should be great interest in knowing how accurately students are classified into performance categories. In contrast to reliability, which is concerned with the relative rank-ordering of students, it is the absolute values of student scores that are important in decision consistency.

Decision consistency answers the question "What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?" If two parallel forms of the test were given to the same students (without additional instruction), the consistency of the measure would be reflected by the extent to which the classification decisions made based on the first set of test scores matched the decisions based on the second set of test scores. Consider Table 16–4:

Table 16–4. Pseudo-Decision Table for Three Hypothetical Categories

Test Level	Test One – Level I	Test One – Level II	Test One – Level III	Test One – Marginal
Test Two – Level I	φ_{11}	φ_{12}	φ_{13}	$\varphi_{1\bullet}$
Test Two – Level II	φ_{21}	φ_{22}	φ_{23}	$\varphi_{2\bullet}$
Test Two – Level III	φ_{31}	φ_{32}	φ_{33}	$\varphi_{3\bullet}$
Test Two – Marginal	$\varphi_{\bullet 1}$	$\varphi_{\bullet 2}$	$\varphi_{\bullet 3}$	1

If a student is classified as in one category based on Test One's score, how probable would it be that the student would be reclassified in the same category if he or she took Test Two (a non-overlapping, equally difficult form of the test)?

The proportions of correct decisions, φ, for three categories is computed as:

$$\tilde{O} = \varphi 11 + \varphi 22 + \varphi 33$$

It is the sum of the diagonal entries—that is, the proportion of students classified by the two forms into exactly the same level—that would signify the overall consistency.

Since it is not feasible to repeat CDT tests one right after the other with no additional instruction 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). Two well-known methods were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995) utilizing specific true score models. While both measures are reported for PSSA and Keystone Exams, the statistical models imposed on the data depend upon a beta binomial distribution of raw scores. Given that the CDT is adaptive (i.e., raw scores using a response probability of 0.5 are generally equal to one-half of test length), these measures are not reported for CDT. Instead, decision consistency measures in this section are a Rasch-based index that relies on conditional standard errors (CSEMs). Also reported are results based on simulations and kappa.

The decision consistency measures reported in the section are based on the Rasch model and conditional standard errors (Stearns and Smith, 2007). Each person's scale score has an associated conditional standard error. Each of the performance levels on the test has an established benchmark cut in the scale score metric. Given these three pieces of information, the assumption of a normal distribution of measurement error allows one to calculate the probability that a student would receive the same classification on retesting. Using the statistic:

$$z = \frac{SS_n - SSBC}{SE_{SS_n}}$$

where *SSn* is the scale score estimate for person *n*, *SSBC* is the scale score benchmark cut, and SE_{ss_n} is the asymptotic standard error of the person scale score estimate. Using cumulative normal probabilities, the probability that a retest would produce the same performance level classification and the probability of a different performance level classification were calculated. The process was repeated for each cut score which results in a probability of classification in each of the performance levels. The total classification rate for the entire sample is the average of the probabilities of the same classification on retesting.

Table 16–5 provides an example based on CDT Algebra I operational data from the 2020–2021 school year. Recall that in the dynamic reporting suite, scores are classified into one of three color ranges—red, green, or blue. The benchmark cut points used for the analyses are the cut points in place during the 2020–2021 school year.

Table 16–5. Retest Classification Probability – Algebra I

	Red – retest	Green – retest	Blue – retest
Red – test	0.954	0.046	0.000
Green – test	0.157	0.809	0.034
Blue – test	0.000	0.179	0.821

Consider students with scores in the green range: The probability of scoring in the red range if retested is 0.157. The probability of scoring in the green range again is 0.809. The probability of scoring in the blue range is 0.034.

The total classification rate is determined by taking the weighted average of the diagonal probabilities where the weights are the number of students in the corresponding range. There are 42,725 students in the sample: 34,961 with total scores in the red range, 7,057 in the green range, and 707 in the blue range. The total classification rate is [(0.954)*(34,961)+(0809)*(7,057)+(0.821)*(707)]/42,725 = 0.927.

In addition to the exact agreement rate, Cohen's kappa[5] was also calculated as 0.767.

In cases with multiple categories, an alternative to kappa, which treats every misclassification as equally important, is a weighted kappa that considers differences that are non-adjacent as more "off." While relevant, given there are three categories, weighted kappa is the same as kappa in this case because both the red/blue and blue/red cells in Table 16–5 are zero.

3 X 3 retest classification probability tables for all CDT tests and benchmark cuts comparable to Table 16–5 are presented in Appendix E.

Stearns and Smith (2007) point out that one advantage of this method is that each student can understand how likely it is that he or she would be classified in the same range if the student took the test over without additional instruction. In addition, each student can learn the probability with which he or she would be reclassified in any of the ranges. A student scoring right at the cut score will have a lower rate of consistent classification than a student scoring in the middle of a performance level band. This can be seen in Table 16–6, which is based on the same Algebra I data set and cut points and shows for various scale scores the percent chance of scoring in each color range if retested.

[5] Kappa, κ, takes into account the agreement occurring by chance.

Table 16–6. Retest Classification Percent for Various Scale Score Ranges – Algebra I

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	2	>99.9%	0.0%	0.0%	>99.9%
500 to 549	22	>99.9%	0.0%	0.0%	>99.9%
550 to 599	206	>99.9%	0.0%	0.0%	>99.9%
600 to 649	615	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,157	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,632	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,160	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,469	>99.9%	0.0%	0.0%	>99.9%
850 to 899	2,997	>99.9%	0.0%	0.0%	>99.9%
900 to 949	3,770	>99.9%	0.0%	0.0%	>99.9%
950 to 999	4,710	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	5,661	99.7%	0.3%	0.0%	99.7%
1050 to 1099	5,965	93.0%	7.0%	0.0%	93.0%
1100 to 1149 (Red/Green cut = 1134)	4,983	59.8%	40.2%	0.0%	64.8%
1150 to 1199	3,182	16.2%	83.7%	0.1%	83.7%
1200 to 1249	1,711	1.2%	96.0%	2.9%	96.0%
1250 to 1299 (Green/Blue cut = 1297)	811	0.0%	74.4%	25.6%	74.4%
1300 to 1349	389	0.0%	26.7%	73.3%	73.3%
1350 to 1399	173	0.0%	3.3%	96.7%	96.7%
1400 to 1449	69	0.0%	0.1%	99.9%	99.9%
1450 to 1499	28	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	8	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	3	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	2	0.0%	0.0%	>99.9%	>99.9%
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	42,725				

* Retest assuming no additional instruction

Tables for all CDT tests and benchmark cuts comparable to Table 16–6 are presented in Appendix E.

As previously mentioned, it is not feasible to repeat CDT tests one right after the other with no additional instruction in order to estimate decision consistency. However, simulations were run as a validation of the results based on the Stearns and Smith method. The reported Algebra I scores from 2020–2021 were used as true scores in order to simulate retest results. Table 16–7 repeats the Algebra I results from Table 16–5, shows the simulation results, and displays the differences.

Table 16–7. Compare Stearns and Smith Results to Simulation Retest Classification Probability – Algebra I

	Red – retest	Green – retest	Blue – retest
Red – Stearns & Smith	0.954	0.046	0.000
Green – Stearns & Smith	0.157	0.809	0.034
Blue – Stearns & Smith	0.000	0.180	0.820

Exact Agreement Rate = 0.927
Kappa = 0.767

	Red – retest	Green – retest	Blue – retest
Red – Simulated test	0.953	0.047	0.000
Green – Simulated test	0.158	0.805	0.037
Blue – Simulated test	0.000	0.181	0.819

Exact Agreement Rate = 0.927
Kappa = 0.764

	Red – retest	Green – retest	Blue – retest
Red – Difference	0.001	-0.001	0.000
Green – Difference	-0.001	0.004	-0.003
Blue – Difference	0.000	-0.001	0.001

Exact Agreement Rate = 0.000
Kappa = 0.003

Based on results of the simulation validation, Stearns and Smith methodology was applied to all CDT tests and benchmark cut points using data from the 2020–2021 school year. Results are presented in Table 16–8.

Table 16–8. Decision Consistency for All CDT Tests

CDT	Benchmark Cut	N-count	Exact Agreement Rate	Kappa
Mathematics Grades 3–5	Grade 3	23,191	0.922	0.798
Mathematics Grades 3–5	Grade 4	24,069	0.923	0.784
Mathematics Grades 3–5	Grade 5	28,759	0.915	0.774
Mathematics Grades 6–HS	Grade 6	34,688	0.918	0.793
Mathematics Grades 6–HS	Grade 7	34,947	0.932	0.776
Mathematics Grades 6–HS	Grade 8	27,468	0.936	0.767
Mathematics Grades 6–HS	High School	454	0.974	0.614
Algebra I	Algebra I	42,725	0.927	0.767
Geometry	Geometry	5,131	0.903	0.772
Algebra II	Algebra II	6,023	0.896	0.794
Reading Grades 3–5	Grade 3	19,838	0.903	0.818
Reading Grades 3–5	Grade 4	20,691	0.903	0.817
Reading Grades 3–5	Grade 5	24,324	0.902	0.816
Reading/Lit Grades 6–HS	Grade 6	25,753	0.903	0.805
Reading/Lit Grades 6–HS	Grade 7	26,355	0.906	0.809
Reading/Lit Grades 6–HS	Grade 8	24,777	0.910	0.807
Reading/Lit Grades 6–HS	Literature	62,170	0.894	0.798
Science Grades 3–5	Grade 3	1,557	0.871	0.792
Science Grades 3–5	Grade 4	13,182	0.870	0.777
Science Grades 3–5	Grade 5	5,939	0.884	0.786
Science Grades 6–HS	Grade 6	10,228	0.893	0.795
Science Grades 6–HS	Grade 7	15,341	0.901	0.792
Science Grades 6–HS	Grade 8	25,927	0.895	0.779
Science Grades 6–HS	High School	493	0.957	0.733
Biology	Biology	46,477	0.894	0.775
Chemistry	Chemistry	3,070	0.887	0.736
Writing Grades 3–5	Grade 3	2,677	0.895	0.800
Writing Grades 3–5	Grade 4	2,177	0.896	0.793
Writing Grades 3–5	Grade 5	3,624	0.892	0.785
Writing/Eng Comp Gr 6–HS	Grade 6	4,985	0.895	0.795
Writing/Eng Comp Gr 6–HS	Grade 7	6,262	0.899	0.801
Writing/Eng Comp Gr 6–HS	Grade 8	5,666	0.906	0.808
Writing/Eng Comp Gr 6–HS	English Composition	4,635	0.890	0.802

See Appendix E for the 3 X 3 retest classification probability tables.

CHAPTER SEVENTEEN: VALIDITY

As defined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), validity refers to "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). The *Standards* provides a framework for describing the sources of evidence that should be considered when evaluating validity. These sources include evidence based on 1) test content, 2) response processes, 3) the internal structure of the test, 4) the relationships between test scores and other variables, and 5) the consequences of testing. In addition, when Item Response Theory (IRT) 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 Classroom Diagnostic Tools (CDT) in support of its 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 is used to summarize and synthesize the evidence based on the framework of the *Standards*. The purposes and intended use of the CDT is reviewed first, and then each type of validity evidence is addressed in turn.

PURPOSES AND INTENDED USES OF THE CDT

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 CDT will be reviewed first. The CDT was developed to support teachers and students in grades 3 through 12. These tools, available at no cost to districts, are fully integrated and aligned in the Standards Aligned System (SAS) and enable educators to identify students' academic strengths and areas of need, as well as provide links to classroom resources. The assessment is administered completely online using a computer adaptive test (CAT) model, and participation is voluntary. CDT scores are available immediately after testing in the dynamic reporting suite. In addition to the scores, this suite includes links to instructional resources. The CDT may be used multiple times throughout the school year.

EVIDENCE BASED ON TEST CONTENT

Test content validity evidence for the CDT 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 Assessment Anchors and Eligible Content. The CDT is intended to measure the knowledge and skills described in the Assessment Anchors and Eligible Content for grades 3 through 8 and high school in mathematics, reading, science, and writing, and courses Algebra I, Geometry, Algebra II, Literature, Biology, Chemistry, and English Composition.

Lane (1999) suggests taking the following steps to support the content validity of an assessment. In the case of the operational CDT, one should:

- evaluate the degree to which the test specifications represent and align with the knowledge and skills described in the corresponding Assessment Anchors and Eligible Content.
- evaluate the alignment between the CDT 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 CDT 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, practice/training with online tools and tests, and time limits for the assessments.
- submit operational tests to third-party independent reviews.

Chapters Two through Five 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 Assessment Anchors and Eligible Content. After development and prior to field testing, items were reviewed for content and bias issues. After being field tested, items were reviewed with respect to their statistical properties and alignment with the learning progressions. Items selected for inclusion in the operational pools 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 CDT 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. Appropriateness for the intended students was also considered, as well as depth of knowledge, 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
    - Depth of knowledge
    - 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 prior hurdles were tried out in a stand-alone or embedded field-test event. Several statistical analyses were conducted on the field-test data including classical item analyses, distractor analyses, and differential item functioning (DIF) analyses. 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.
- Following field testing, the items were submitted to content committees (composed of Pennsylvania educators) for review and alignment with the learning progressions.
- The CDT was administered according to standardized procedures with allowable accommodations. Students were given ample time to complete the tests (i.e., there were no speediness issues).

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.

For the operational 2020–2021 CDT, no cognitive lab studies were conducted to collect the response process evidence.

EVIDENCE BASED ON INTERNAL STRUCTURE

As described in the *Standards* (2014), internal-structure evidence refers to the degree to which the relationships among test items and test components conform to the construct on which the proposed test interpretations are based. For each CDT, one total test score as well as diagnostic category scores were reported (see Chapter Fourteen for more information about CDT 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 discussed in Chapter Seven and provided in Appendix B of the 2017–2018, 2018–2019, and 2019-2020 technical reports. All items in the final operational pools had values that were positive and of acceptable magnitude.

DIMENSIONALITY

Dimensionality analyses were conducted for the CDT using WINSTEPS's principal components analyses on response residuals for each content area. Results are shown in Chapter Eight. The principal component analysis results provided evidence that each CDT test was essentially unidimensional, supporting the validity of using the total scores to estimate a student's overall ability.

DIAGNOSTIC CATEGORY CORRELATIONS

Correlations and disattenuated correlations among diagnostic category scores for the CDT are presented below. Values were derived from the CDT operational data from the 2020–2021 school year. This data can also provide information on score dimensionality that is part of internal-structure evidence. Each CDT has either four or five diagnostic categories. Full diagnostic category names can be found in Chapter Thirteen.

Table 17–1. Correlations among Diagnostic Categories — Math Grades 3–5

Diagnostic Category	Numbers.	Alg. Con	Geo.	Meas.
Numbers.	-	-	-	-
Alg. Con.	0.781	-	-	-
Geo.	0.712	0.697	-	-
Meas.	0.777	0.774	0.711	-

Table 17–2. Correlations among Diagnostic Categories — Math Grades 6–HS

Diagnostic Category	Numbers.	Alg. Con	Geo.	Meas.
Numbers.	-	-	-	-
Alg. Con.	0.739	-	-	-
Geo.	0.698	0.688	-	-
Meas.	0.726	0.706	0.683	-

Table 17–3. Correlations among Diagnostic Categories — Algebra I

Diagnostic Category	Operations.	Linear.	Functions.	Data.
Operations.	-	-	-	-
Linear.	0.644	-	-	-
Functions.	0.656	0.669	-	-
Data.	0.656	0.656	0.680	-

Table 17–4. Correlations among Diagnostic Categories — Geometry

Diagnostic Category	Properties.	Congruence.	Coordinate.	Measure.
Properties.	-	-	-	-
Congruence.	0.663	-	-	-
Coordinate.	0.667	0.678	-	-
Measure.	0.671	0.654	0.683	-

Table 17–5. Correlations among Diagnostic Categories — Algebra II

Diagnostic Category	Complex.	Non-Linear.	Functions.	Data.
Complex.	-	-	-	-
Non-Linear.	0.607	-	-	-
Functions.	0.526	0.674	-	-
Data.	0.416	0.618	0.652	-

Table 17–6. Correlations among Diagnostic Categories — Reading Grades 3–5

Diagnostic Category	Key – Lit.	Key – Info.	Craft – Lit.	Craft – Info.	Vocab.
Key – Lit.	-	-	-	-	-
Key – Info.	0.681	-	-	-	-
Craft – Lit.	0.685	0.652	-	-	-
Craft – Info.	0.674	0.678	0.652	-	-
Vocab.	0.706	0.690	0.665	0.692	-

Table 17–7. Correlations among Diagnostic Categories — Reading/Lit Grades 6–HS

Diagnostic Category	Key – Lit.	Key – Info.	Craft – Lit.	Craft – Info.	Vocab.
Key – Lit.	-	-	-	-	-
Key – Info.	0.660	-	-	-	-
Craft – Lit.	0.660	0.645	-	-	-
Craft – Info.	0.669	0.692	0.657	-	-
Vocab.	0.673	0.681	0.662	0.695	-

Table 17–8. Correlations among Diagnostic Categories — Science Grades 3–5

Diagnostic Category	Nature.	Bio.	Phys.	Earth/Space.
Nature.	-	-	-	-
Bio.	0.772	-	-	-
Phys.	0.759	0.754	-	-
Earth/Space.	0.753	0.751	0.736	-

Table 17–9. Correlations among Diagnostic Categories — Science Grades 6–HS

Diagnostic Category	Nature.	Bio.	Phys.	Earth/Space.
Nature.	-	-	-	-
Bio.	0.732	-	-	-
Phys.	0.703	0.697	-	-
Earth/Space.	0.705	0.697	0.669	-

Table 17–10. Correlations among Diagnostic Categories — Biology

Diagnostic Category	Basic.	Bioenerg.	Cell Growth.	Evol./Ecol.
Basic.	-	-	-	-
Bioenerg.	0.672	-	-	-
Cell Growth.	0.679	0.635	-	-
Evol./Ecol.	0.711	0.645	0.679	-

Table 17–11. Correlations among Diagnostic Categories — Chemistry

Diagnostic Category	Matter.	Atomic.	Mole.	Chem.
Matter.	-	-	-	-
Atomic.	0.488	-	-	-
Mole.	0.578	0.521	-	-
Chem.	0.522	0.495	0.524	-

Table 17–12. Correlations among Diagnostic Categories — Writing Grades 3–5

Diagnostic Category	Focus.	Content.	Edit.	Punct.	Gram.
Focus.	-	-	-	-	-
Content.	0.768	-	-	-	-
Edit.	0.763	0.757	-	-	-
Punct.	0.721	0.708	0.737	-	-
Gram.	0.743	0.747	0.769	0.725	-

Table 17–13. Correlations among Diagnostic Categories — Writing/Eng Comp Grades 6–HS

Diagnostic Category	Focus.	Content.	Edit.	Punct.	Gram.
Focus.	-	-	-	-	-
Content.	0.729	-	-	-	-
Edit.	0.733	0.719	-	-	-
Punct.	0.707	0.694	0.718	-	-
Gram.	0.707	0.690	0.717	0.710	-

The correlations in Tables 17–1 through 17–13 are based on the observed diagnostic category scores. These observed-score correlations are weakened by existing measurement error contained within each diagnostic category. As a result, disattenuated correlations could provide an estimate of the relationships among diagnostic categories 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, 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 diagnostic categories 1 and 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 diagnostic categories are measuring slightly different aspects of the same construct. Values markedly less than 1.00 suggest the diagnostic categories reflect different constructs.

Tables 17–14 through 17–26 show the corresponding disattenuated correlations. Given that none of these diagnostic categories had perfect reliabilities (see Chapter Sixteen), the disattenuated correlations are higher than their observed score counterparts.

Table 17–14. Disattenuated Correlations among Diagnostic Categories — Math Grades 3–5

Diagnostic Category	Numbers.	Alg. Con	Geo.	Meas.
Numbers.	-	-	-	-
Alg. Con.	0.957	-	-	-
Geo.	0.893	0.883	-	-
Meas.	0.953	0.960	0.901	-

Table 17–15. Disattenuated Correlations among Diagnostic Categories — Math Grades 6–HS

Diagnostic Category	Numbers.	Alg. Con	Geo.	Meas.
Numbers.	-	-	-	-
Alg. Con.	0.894	-	-	-
Geo.	0.867	0.866	-	-
Meas.	0.881	0.870	0.862	-

Table 17–16. Disattenuated Correlations among Diagnostic Categories — Algebra I

Diagnostic Category	Operations.	Linear.	Functions.	Data.
Operations.	-	-	-	-
Linear.	0.794	-	-	-
Functions.	0.802	0.869	-	-
Data.	0.784	0.833	0.854	-

Table 17–17. Disattenuated Correlations among Diagnostic Categories — Geometry

Diagnostic Category	Properties.	Congruence.	Coordinate.	Measure.
Properties.	-	-	-	-
Congruence.	0.846	-	-	-
Coordinate.	0.834	0.849	-	-
Measure.	0.840	0.821	0.840	-

Table 17–18. Disattenuated Correlations among Diagnostic Categories — Algebra II

Diagnostic Category	Complex.	Non-Linear.	Functions.	Data.
Complex.	-	-	-	-
Non-Linear.	0.714	-	-	-
Functions.	0.640	0.845	-	-
Data.	0.504	0.771	0.841	-

Table 17–19. Disattenuated Correlations among Diagnostic Categories — Reading Grades 3–5

Diagnostic Category	Key – Lit.	Key – Info.	Craft – Lit.	Craft – Info.	Vocab.
Key – Lit.	-	-	-	-	-
Key – Info.	0.938	-	-	-	-
Craft – Lit.	0.980	0.943	-	-	-
Craft – Info.	0.926	0.943	0.942	-	-
Vocab.	0.951	0.940	0.942	0.942	-

Table 17–20. Disattenuated Correlations among Diagnostic Categories — Reading/Lit Grades 6–HS

Diagnostic Category	Key – Lit.	Key – Info.	Craft – Lit.	Craft – Info.	Vocab.
Key – Lit.	-	-	-	-	-
Key – Info.	0.915	-	-	-	-
Craft – Lit.	0.946	0.928	-	-	-
Craft – Info.	0.915	0.948	0.931	-	-
Vocab.	0.906	0.919	0.924	0.925	-

Table 17–21. Disattenuated Correlations among Diagnostic Categories — Science Grades 3–5

Diagnostic Category	Nature.	Bio.	Phys.	Earth/Space.
Nature.	-	-	-	-
Bio.	0.988	-	-	-
Phys.	0.981	0.967	-	-
Earth/Space.	0.983	0.973	0.965	-

Table 17–22. Disattenuated Correlations among Diagnostic Categories — Science Grades 6–HS

Diagnostic Category	Nature.	Bio.	Phys.	Earth/Space.
Nature.	-	-	-	-
Bio.	0.913	-	-	-
Phys.	0.908	0.903	-	-
Earth/Space.	0.913	0.906	0.900	-

Table 17–23. Disattenuated Correlations among Diagnostic Categories — Biology

Diagnostic Category	Basic.	Bioenerg.	Cell Growth.	Evol./Ecol.
Basic.	-	-	-	-
Bioenerg.	0.897	-	-	-
Cell Growth.	0.888	0.881	-	-
Evol./Ecol.	0.888	0.856	0.882	-

Table 17–24. Disattenuated Correlations among Diagnostic Categories — Chemistry

Diagnostic Category	Matter.	Atomic.	Mole.	Chem.
Matter.	-	-	-	-
Atomic.	0.758	-	-	-
Mole.	0.850	0.920	-	-
Chem.	0.759	0.866	0.869	-

Table 17–25. Disattenuated Correlations among Diagnostic Categories — Writing Grades 3–5

Diagnostic Category	Focus.	Content.	Edit.	Punct.	Gram.
Focus.	-	-	-	-	-
Content.	0.969	-	-	-	-
Edit.	0.966	0.967	-	-	-
Punct.	0.928	0.919	0.959	-	-
Gram.	0.935	0.948	0.979	0.938	-

Table 17–26. Disattenuated Correlations among Diagnostic Categories — Writing/Eng Comp Grades 6–HS

Diagnostic Category	Focus.	Content.	Edit.	Punct.	Gram.
Focus.	-	-	-	-	-
Content.	0.915	-	-	-	-
Edit.	0.918	0.909	-	-	-
Punct.	0.882	0.875	0.903	-	-
Gram.	0.889	0.876	0.908	0.897	-

In reviewing the differences between the simple correlations and the disattenuated ones, it is clear that the impact of the "less than perfect" reliabilities on the disattenuated correlations is large for most of the tests. For example, Science Grades 3–5 found virtually no differences between any pair of disattenuated correlations. This indicates that, for the majority of students, the diagnostic category scores are merely shorter versions of what the total scores are measuring. Note that, while the theoretical maximum for observed correlations is 1.00, disattenuated correlations can exceed this value when high observed correlations are combined with low reliabilities. The other tests' disattenuated correlations are somewhat lower, generally in the range of .82 to .95. The test with the lowest disattenuated correlations is Algebra II, with Complex Numbers showing the most uniqueness.

As a practical consideration, and despite these results, diagnostic category scores for individual students may still provide useful information to the teacher. For example, a student may still have statistically significant differences between pairs of diagnostic scores ("areas of needs" versus "strengths to build on") with large observed scale score differences. The diagnostic reporting suite shows these differences in a graphic that includes the level of precision for each scale score in the form of an "error band." The error band is the scale score ± one conditional standard error. Any two pairs of scores can be interpreted as statistically different if their respective error bands do not overlap. More details about the use and interpretation of error bands may be found in Chapter Fourteen. Additionally, Chapter Fifteen provides summary information about conditional standard errors for each diagnostic category and tables that indicate the incidence of non-overlapping error bands in the 2020–2021 operational testing population.

EXPLORATORY FACTOR ANALYSIS

In order to further explore the internal structure of each CDT, an exploratory factor analysis (EFA) of the diagnostic category scores was conducted. Operational data from the 2020–2021 school year was used to create the observed correlation matrices shown in Tables 17–1 through 17–13. These, in turn, were used in the EFA. In the *Statistical Package for the Social Sciences (SPSS)*, Principal Axis Factor extraction was utilized with an oblique rotation (Promax) of the initial factor solution to improve interpretability. Oblique rotations allow for correlated factors.

Tables 17–27 through 17–39 present the eigenvalues and the explained variance for the extracted factors. Figures 17–1 through 17–13 are scree plot graphs of the eigenvalues against the factor number. In general, the first factor accounts for approximately 76% of the total variance for all CDT tests except Chemistry, while the second factor accounts for approximately 8% of the total variance. For Chemistry, the first factor accounts for 64% of the total variance, while the second factor accounts for 13%. For each CDT, only the first factor had an eigenvalue greater than 1.0, typically suggesting a one-factor solution using the Kaiser criterion.

Table 17–27. Eigenvalues and Explained Variance for Math Grades 3–5 Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.23	80.68
2	0.33	8.21
3	0.23	5.65
4	0.22	5.45

Figure 17–1. Scree Plot for Math Grades 3–5 Diagnostic Categories

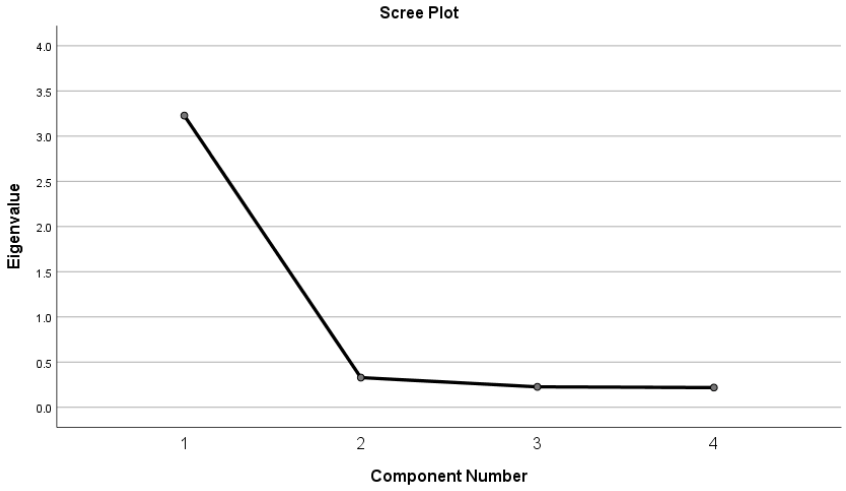


Table 17–28. Eigenvalues and Explained Variance for Math Grades 6–HS Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.12	78.01
2	0.33	8.18
3	0.30	7.39
4	0.26	6.42

Figure 17–2. Scree Plot for Math Grades 6–HS Diagnostic Categories

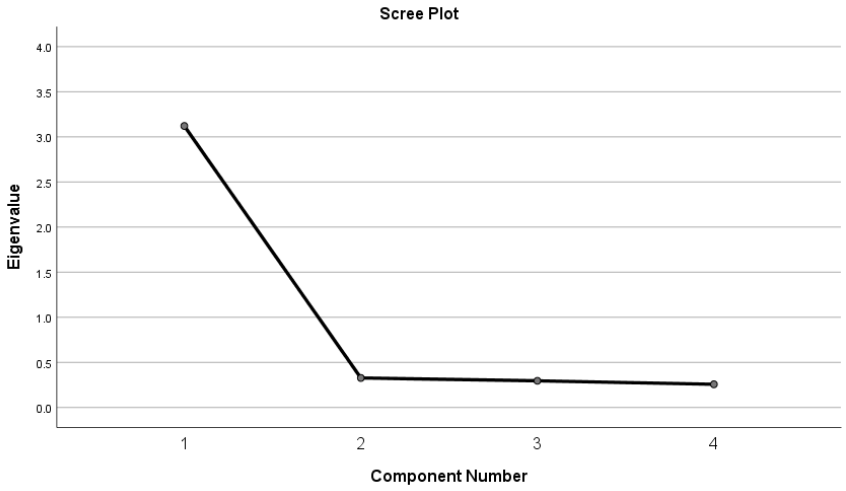


Table 17–29. Eigenvalues and Explained Variance for Algebra I Diagnostic Categories

Factor	Eigenvalue	Percent
1	2.98	74.52
2	0.36	8.95
3	0.34	8.59
4	0.32	7.94

Figure 17–3. Scree Plot for Algebra I Diagnostic Categories

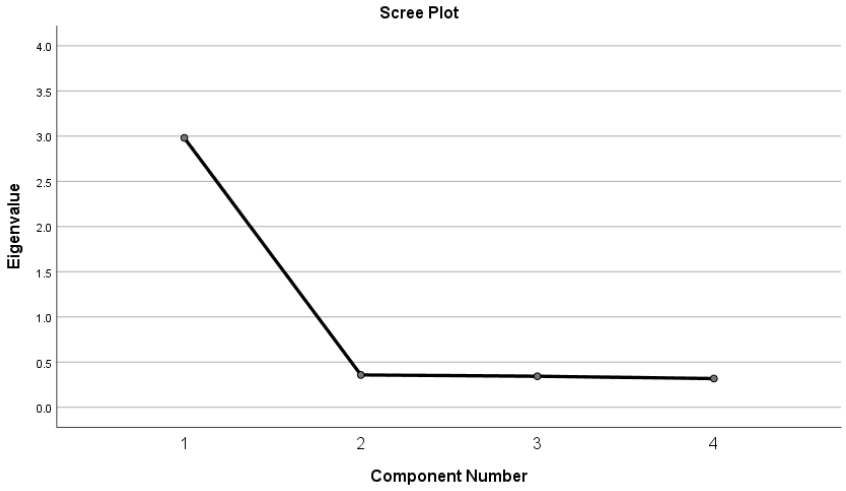


Table 17–30. Eigenvalues and Explained Variance for Geometry Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.01	75.20
2	0.35	8.68
3	0.34	8.40
4	0.31	7.72

Figure 17–4. Scree Plot for Geometry Diagnostic Categories

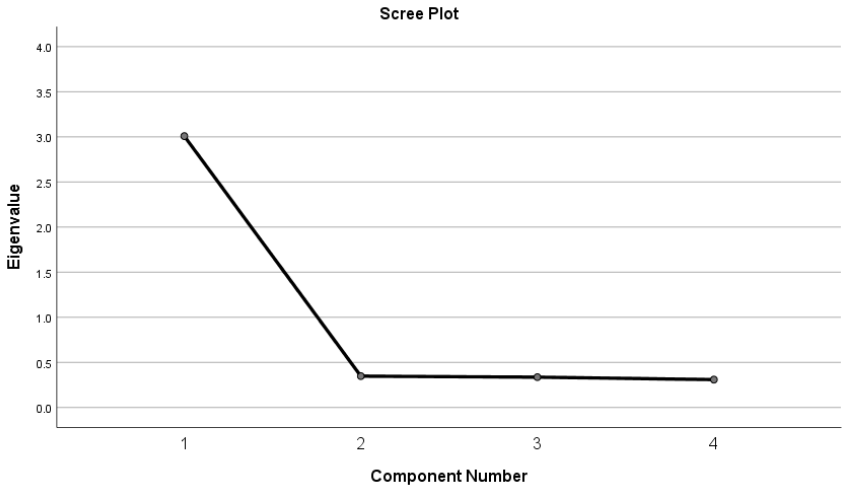


Table 17–31. Eigenvalues and Explained Variance for Algebra II Diagnostic Categories

Factor	Eigenvalue	Percent
1	2.75	68.87
2	0.60	15.10
3	0.33	8.27
4	0.31	7.76

Figure 17–5. Scree Plot for Algebra II Diagnostic Categories

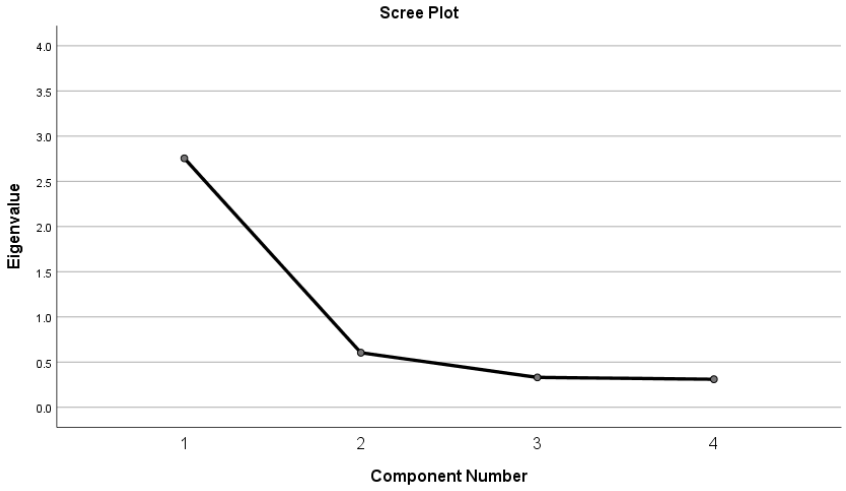


Table 17–32. Eigenvalues and Explained Variance for Reading Grades 3–5 Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.71	74.21
2	0.36	7.24
3	0.32	6.47
4	0.32	6.33
5	0.29	5.76

Figure 17–6. Scree Plot for Reading Grades 3–5 Diagnostic Categories

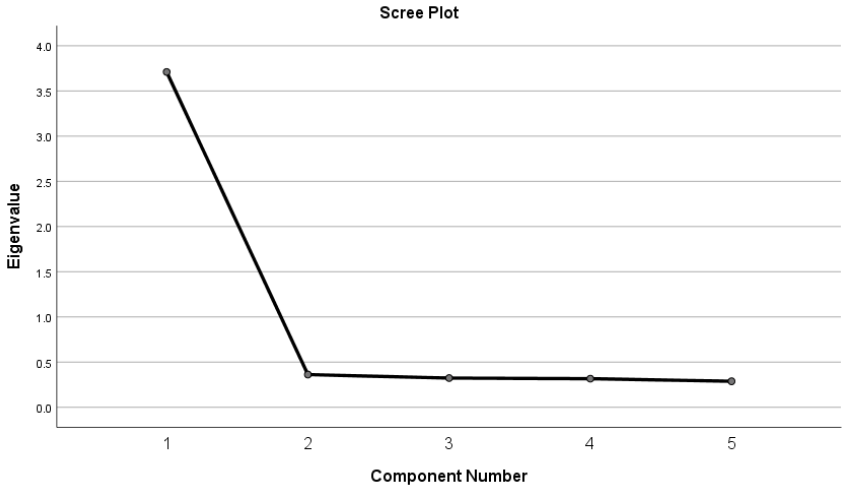


Table 17–33. Eigenvalues and Explained Variance for Reading/Lit Grades 6–HS Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.68	73.56
2	0.37	7.34
3	0.34	6.73
4	0.32	6.33
5	0.30	6.04

Figure 17–7. Scree Plot for Reading/Lit Grades 6–HS Diagnostic Categories

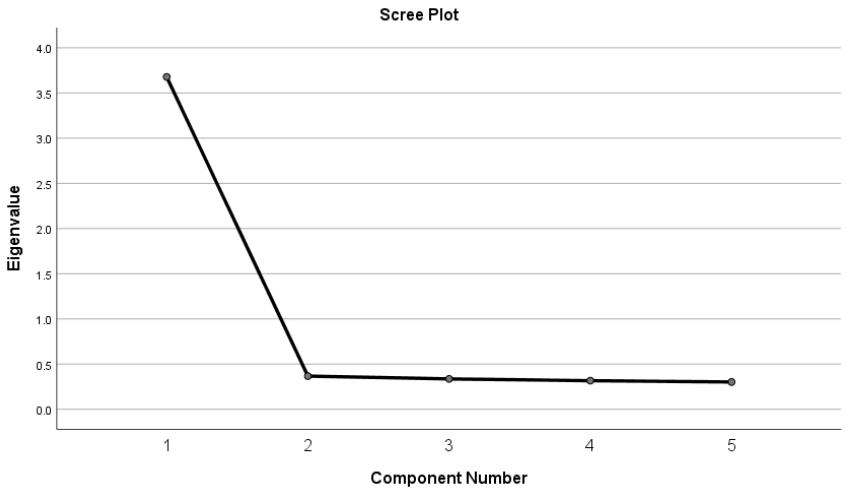


Table 17–34. Eigenvalues and Explained Variance for Science Grades 3–5 Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.26	81.57
2	0.26	6.61
3	0.25	6.13
4	0.23	5.69

Figure 17–8. Scree Plot for Science Grades 3–5 Diagnostic Categories

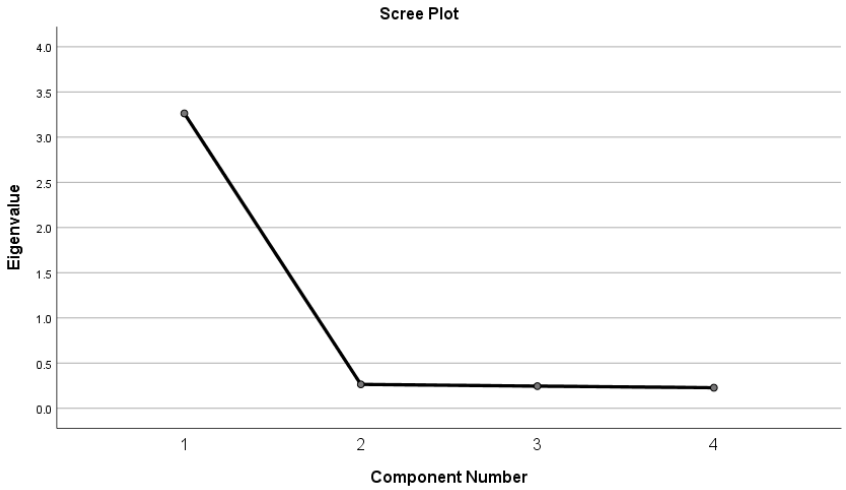


Table 17–35. Eigenvalues and Explained Variance for Science Grades 6–HS Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.10	77.55
2	0.33	8.28
3	0.30	7.50
4	0.27	6.68

Figure 17–9. Scree Plot for Science Grades 6–HS Diagnostic Categories

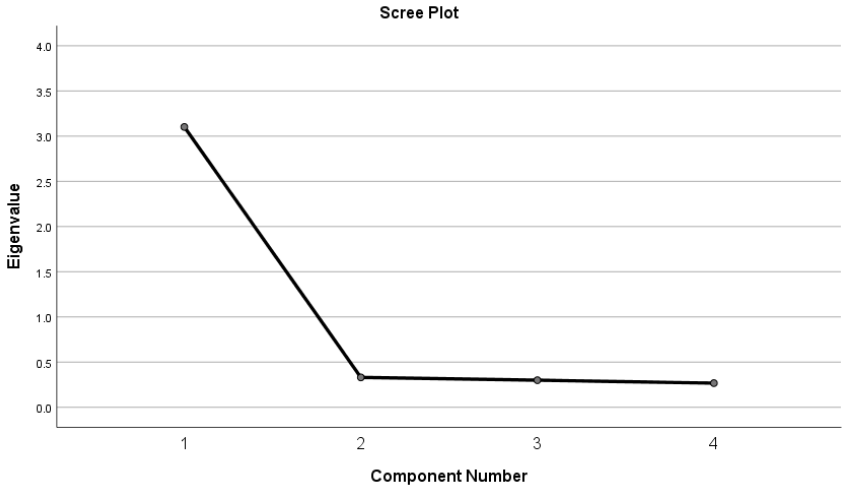


Table 17–36. Eigenvalues and Explained Variance for Biology Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.01	75.28
2	0.37	9.34
3	0.33	8.24
4	0.29	7.14

Figure 17–10. Scree Plot for Biology Diagnostic Categories

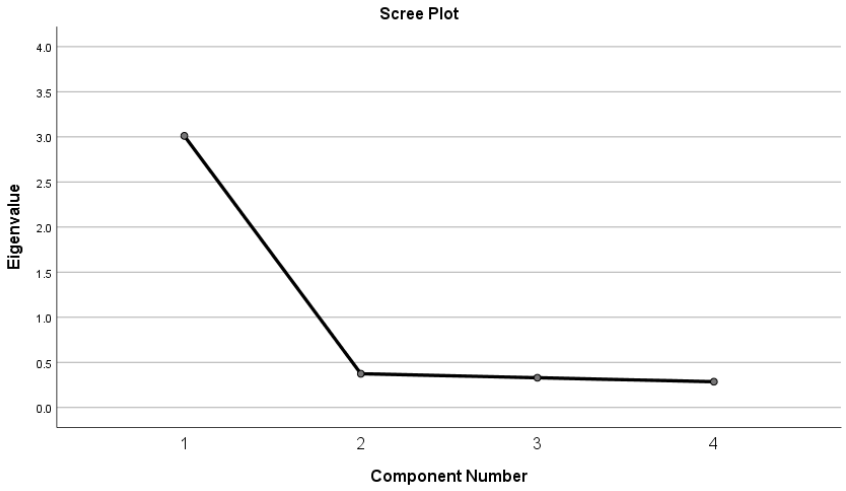


Table 17–37. Eigenvalues and Explained Variance for Chemistry Diagnostic Categories

Factor	Eigenvalue	Percent
1	2.56	64.12
2	0.52	13.08
3	0.49	12.34
4	0.42	10.45

Figure 17–11. Scree Plot for Chemistry Diagnostic Categories

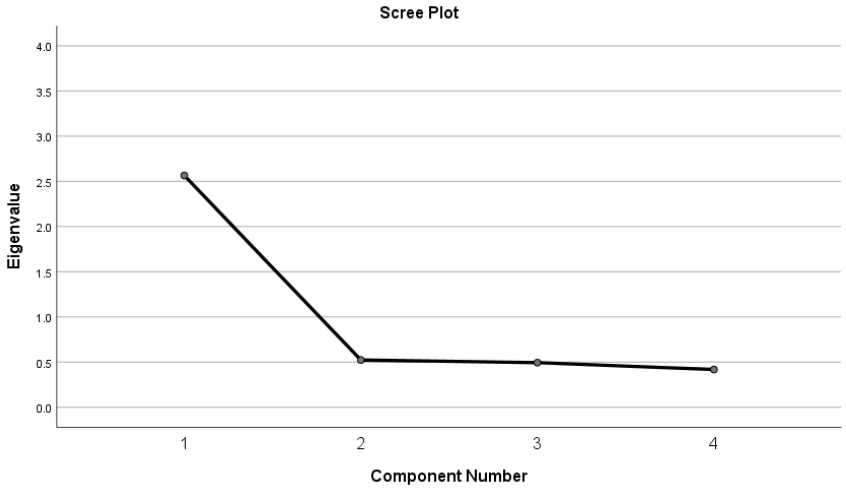


Table 17–38. Eigenvalues and Explained Variance for Writing Grades 3–5 Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.98	79.52
2	0.30	6.08
3	0.26	5.25
4	0.23	4.67
5	0.22	4.49

Figure 17–12. Scree Plot for Writing Grades 3–5 Diagnostic Categories

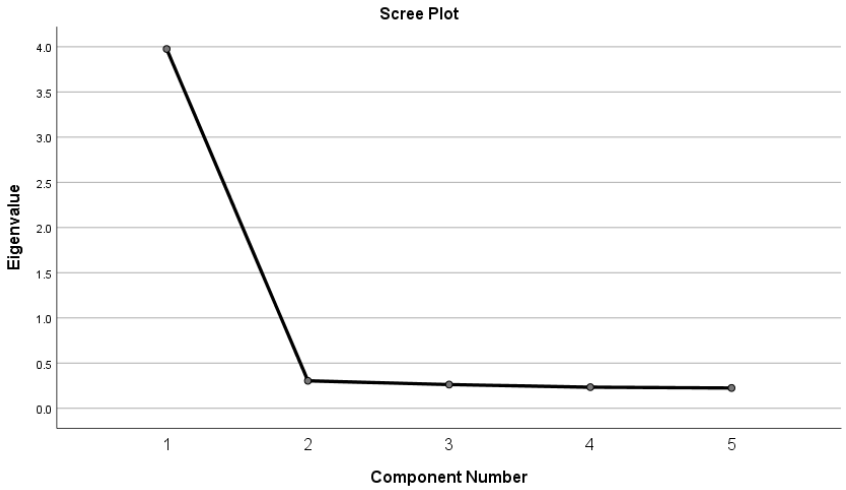
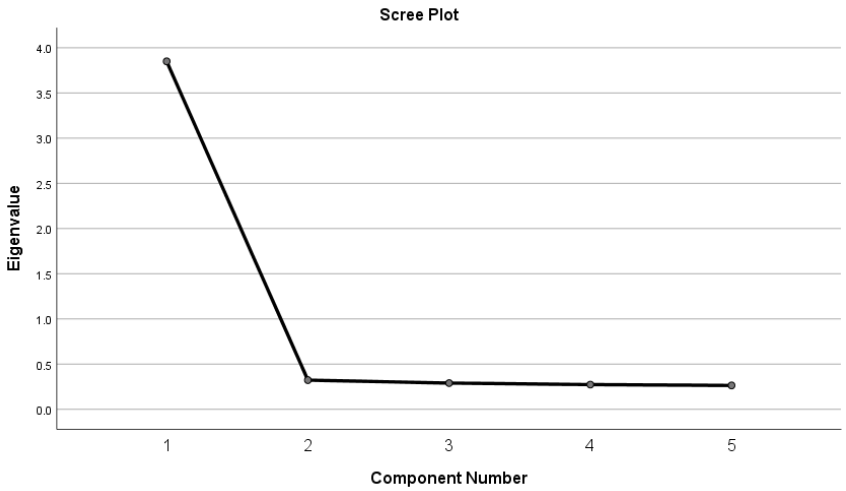


Table 17–39. Eigenvalues and Explained Variance for Writing/Eng Comp Grades 6–HS Diagnostic Categories

Factor	Eigenvalue	Percent
1	3.85	77.00
2	0.32	6.46
3	0.29	5.80
4	0.27	5.46
5	0.26	5.28

Figure 17–13. Scree Plot for Writing/Eng Comp Grades 6–HS Diagnostic Categories



Taken as a whole, the internal structure evidence presented generally indicates that related elements of each of the CDT tests are correlated in the intended manner. This further supports using a total score to report students' performances in the different content areas.

The diagnostic category scores present more of a mixed message. Since the diagnostic categories in each of the CDT tests were designed to measure distinct components, it is reasonable to expect that the diagnostic category correlations should be positive and strong but, ideally, not extremely high. However, the disattenuated correlations imply that some diagnostic categories are essentially measuring the same constructs. While there is content rationale underlying the creation of the diagnostic category scores, the empirical correlations illustrate that caution is required when using these scores when identifying an individual student's areas of need and strengths to build on.

EVIDENCE BASED ON RELATIONSHIPS WITH OTHER VARIABLES

As described in the *Standards* (AERA, APA, & NCME, 2014), ". . . Evidence based on relationships with other variables provides evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" (p. 16). 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).

Correlations and disattenuated correlations among students' test scores across different CDT content areas provide some discriminant validity evidence. These are provided in Tables 17–40 and 17–41.

Table 17–40a. Correlations among CDT Grades 3–5 Tests

CDT	Math Grades 3–5	Reading Grades 3–5	Science Grades 3–5	Writing Grades 3–5
Math Grades 3–5	-	-	-	-
Reading Grades 3–5	0.760	-	-	-
Science Grades 3–5	0.741	0.749	-	-
Writing Grades 3–5	0.728	0.797	0.757	-

Table 17–40b. Correlations among CDT Tests

CDT	Math Gr 6–HS	Algebra I	Geometry	Algebra II	Read/Lit Gr 6–HS	Science Gr 6–HS	Biology	Chemistry	Writing/ Eng Comp Gr 6–HS
Math Gr 6–HS	-	-	-	-	-	-	-	-	-
Algebra I	0.806	-	-	-	-	-	-	-	-
Geometry	0.367	0.820	-	-	-	-	-	-	-
Algebra II	-	0.531	0.728	-	-	-	-	-	-
Read/Lit Gr 6–HS	0.735	0.677	0.672	0.622	-	-	-	-	-
Science Gr 6–HS	0.730	0.663	0.678	0.586	0.759	-	-	-	-
Biology	0.792	0.640	0.728	0.604	0.736	0.824	-	-	-
Chemistry	-	0.447	0.710	0.714	0.644	-	0.582	-	-
Writing Gr 6–HS	0.732	0.684	0.655	0.637	0.807	0.744	0.725	0.671	-

Table 17–41a. Disattenuated Correlations among CDT Grades 3–5 Tests

CDT	Math Grades 3–5	Reading Grades 3–5	Science Grades 3–5	Writing Grades 3–5
Math Grades 3–5	-	-	-	-
Reading Grades 3–5	0.813	-	-	-
Science Grades 3–5	0.789	0.804	-	-
Writing Grades 3–5	0.770	0.850	0.803	-

Table 17–41b. Disattenuated Correlations among CDT Tests

CDT	Math Gr 6–HS	Algebra I	Geometry	Algebra II	Read/Lit Gr 6–HS	Science Gr 6–HS	Biology	Chemistry	Writing/ Eng Comp Gr 6–HS
Math Gr 6–HS	-	-	-	-	-	-	-	-	-
Algebra I	0.858	-	-	-	-	-	-	-	-
Geometry	0.392	0.876	-	-	-	-	-	-	-
Algebra II	-	0.568	0.780	-	-	-	-	-	-
Read/Lit Gr 6–HS	0.787	0.727	0.722	0.668	-	-	-	-	-
Science Gr 6–HS	0.781	0.711	0.727	0.629	0.817	-	-	-	-
Biology	0.849	0.687	0.783	0.650	0.794	0.888	-	-	-
Chemistry	-	0.494	0.786	0.791	0.715	-	0.647	-	-
Writing Gr 6–HS	0.775	0.726	0.696	0.678	0.861	0.793	0.775	0.738	-

Each CDT test measures a different construct, so the correlations among them were not expected to be extremely high. The values in the tables are consistent with this expectation. Correlations among the CDT tests ranged from 0.367 to 0.824. Correlations across tests within a content area tend to be more highly correlated than across content areas. For example, the correlation between Algebra I and Geometry is 0.820, whereas the correlation between Algebra I and Biology is 0.640.

External evidence for the CDT is examined by using students' scores on the 2021 Pennsylvania System of School Assessment (PSSA) and/or 2021 Keystone Exams as external criteria. For each content area, CDT results from the 2020–2021 school year were matched to spring 2021 PSSA in the corresponding content area using the PA secure ID. Similarly, CDT tests in Algebra I, Biology, and Reading/Literature were matched to corresponding spring 2021 Keystone Exams. The correlations between students' total scale scores on the CDT and PSSA or Keystone are calculated as one piece of external evidence. Table 17–42 summarizes the sample sizes and correlations.

Table 17–42. Correlation between CDT and PSSA or Keystone Exams Scores

Student Grade	CDT	PSSA or Keystone Test	*N*	Correlation of Total Scale Scores
3	Math Grades 3–5	PSSA Math Grade 3	11,958	0.745
4	Math Grades 3–5	PSSA Math Grade 4	12,011	0.759
5	Math Grades 3–5	PSSA Math Grade 5	13,622	0.772
6	Math Grades 6–HS	PSSA Math Grade 6	15,622	0.805
7	Math Grades 6–HS	PSSA Math Grade 7	15,784	0.779
8	Math Grades 6–HS	PSSA Math Grade 8	12,068	0.752
3	Reading Grades 3–5	PSSA ELA Grade 3	10,050	0.750
4	Reading Grades 3–5	PSSA ELA Grade 4	10,244	0.794
5	Reading Grades 3–5	PSSA ELA Grade 5	11,353	0.774
6	Reading/Lit Grades 6–HS	PSSA ELA Grade 6	12,537	0.785
7	Reading/Lit Grades 6–HS	PSSA ELA Grade 7	13,237	0.750
8	Reading/Lit Grades 6–HS	PSSA ELA Grade 8	12,515	0.741
4	Science Grades 3–5	PSSA Science Grade 4	5,393	0.719
8	Science Grades 6–HS	PSSA Science Grade 8	11,547	0.759
3	Writing Grades 3–5	PSSA ELA Grade 3	1,247	0.761
4	Writing Grades 3–5	PSSA ELA Grade 4	1,272	0.738
5	Writing Grades 3–5	PSSA ELA Grade 5	1,809	0.738
6	Writing/Eng Comp Gr 6–HS	PSSA ELA Grade 6	2,517	0.768
7	Writing/Eng Comp Gr 6–HS	PSSA ELA Grade 7	3,145	0.746
8	Writing/Eng Comp Gr 6–HS	PSSA ELA Grade 8	3,133	0.721
6–12	Algebra I	Keystone Algebra I	16,714	0.741
6–12	Biology	Keystone Biology	20,504	0.788
6–12	Reading/Literature	Keystone Literature	19,357	0.732

These results provide external evidence in support of CDT as a valid measure of students' achievement.

The collection of external evidence related to the CDT is an ongoing process. As more CDT data become available, other criterion-related evidence will be evaluated. In addition to examining the relationship between CDT and PSSA or Keystone Exams, other criterion variables such as Scholastic Aptitude Test (SAT) scores, American College Test (ACT) scores, or student grade point average (GPA) may be considered.

EVIDENCE BASED ON CONSEQUENCES OF TESTS

According to the *Standards* (AERA, APA, & NCME, 2014), 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 changes in scale scores for students who took the same CDT test multiple times. Table 17–43 below summarizes scale score changes between the first and last administrations of the CDT.

Table 17–43. Summary of Scale Score Changes between CDT Administrations

CDT	*N*	Minimum	Q1	Median	Mean	Q3	Maximum
Math Grades 3–5	23,597	-845	6	68	65.87	130	621
Math Grades 6–HS	29,985	-626	-28	35	29.76	94	606
Algebra I	12,161	-542	-45	29	22.48	97	584
Geometry	1,581	-458	-16	44	34.66	102	468
Algebra II	1,728	-546	-8	62	55.76	125	734
Reading Grades 3–5	20,663	-670	-33	28	27.06	90	688
Reading/Lit Grades 6–HS	40,934	-631	-60	1	-2.05	60	544
Science Grades 3–5	6,699	-805	-39	28	21.99	89	556
Science Grades 6–HS	15,598	-500	-53	10	5.64	68	606
Biology	12,584	-481	-26	43	39.52	108	651
Chemistry	849	-308	-35	28	23.53	85	340
Writing Grades 3–5	2,823	-549	-34	32	29.50	96	526
Writing/Eng Comp Gr 6–HS	5,924	-510	-53	10	3.61	67	583

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 Fourteen, CDT offers a dynamic suite of reports. The extent to which various groups of users (e.g., students and teachers) interpret these reports appropriately affects the validity of subsequent uses of these results. As noted in Chapter Fourteen, there are report training scenarios for each content area. The intent is that the scenarios will help users avoid unintended uses and interpretations of the CDT results.

EVIDENCE RELATED TO USE OF THE RASCH MODEL

Since the Rasch model is the basis of all calibration, scaling, and equating analyses associated with the CDT, 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 Eight, the underlying assumptions of Rasch models were essentially met for all the CDT data, indicating the appropriateness of using Rasch models to analyze the CDT 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 CDT item and its associated Eligible Content. Detailed information regarding educator reviews are presented in Chapter Six.

Diagnostic category score intercorrelations were also presented in this chapter. They provide some favorable evidence regarding the internal 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 Sixteen) were very good, with many in the low 0.90s.

Reported in Chapter Six, differential item functioning (DIF) with respect to gender and ethnicity 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 teacher committees are eligible for operational use.

CHAPTER EIGHTEEN: PARAMETER STABILITY

The Classroom Diagnostic Tools (CDT) features a number of tests. Tests in Mathematics, Algebra I, Geometry, and Algebra II have been available since October 2010 for students in grades 6 and above. Tests in Reading/ Literature, Science, Biology, and Chemistry have been available since April 2011 for students in grades 6 and above. Tests in Writing /English Composition have been available since October 2011 for students in grades 6 and above. Tests in Mathematics, Reading, Science, and Writing have been available since April 2014 for students in grades 3 through 5. During the 2020–2021 school year, CAT item selection and Rasch ability estimates were based on initial item parameters estimated from the stand-alone and embedded field-test events and vertical linking (see Chapter Eight and Chapter Nine for details). The only exceptions were 113 items in the mathematics content area that had parameters re-estimated following the 2010–2011 school year and 74 items in the science content area that had parameters re-estimated following the 2011–2012 school year. Following the 2020–2021 school year, item parameter stability was checked for all items in the banks.

METHODOLOGY

In the first two years of CDT, four separate methods were investigated to evaluate the stability of the item parameters in the CDT operational administration

1. Calibrate the entire bank within a content area in a single concurrent calibration. Do not anchor item parameters on banked values. Compare new parameter estimates to the banked values.
2. Calibrate the entire bank within a content area in a single concurrent calibration. Anchor item parameters on banked values. Examine displacements.
3. Calibrate each grade/course level item with students in that grade/course. Do not anchor item parameters on banked values. Compare new parameter estimates to the banked values.
4. Calibrate each grade/course level item with students in that grade/course. Anchor item parameters on banked values. Examine displacements.

As noted in Chapter Twelve, CDT tests are pre-equated. Immediate score reports are based on banked item parameters. Therefore, this chapter focuses on anchored calibrations and examination of displacement values to evaluate item parameter stability[1].

ANCHORED CONCURRENT CALIBRATION WITHIN CONTENT AREA ACROSS GRADES/ COURSES

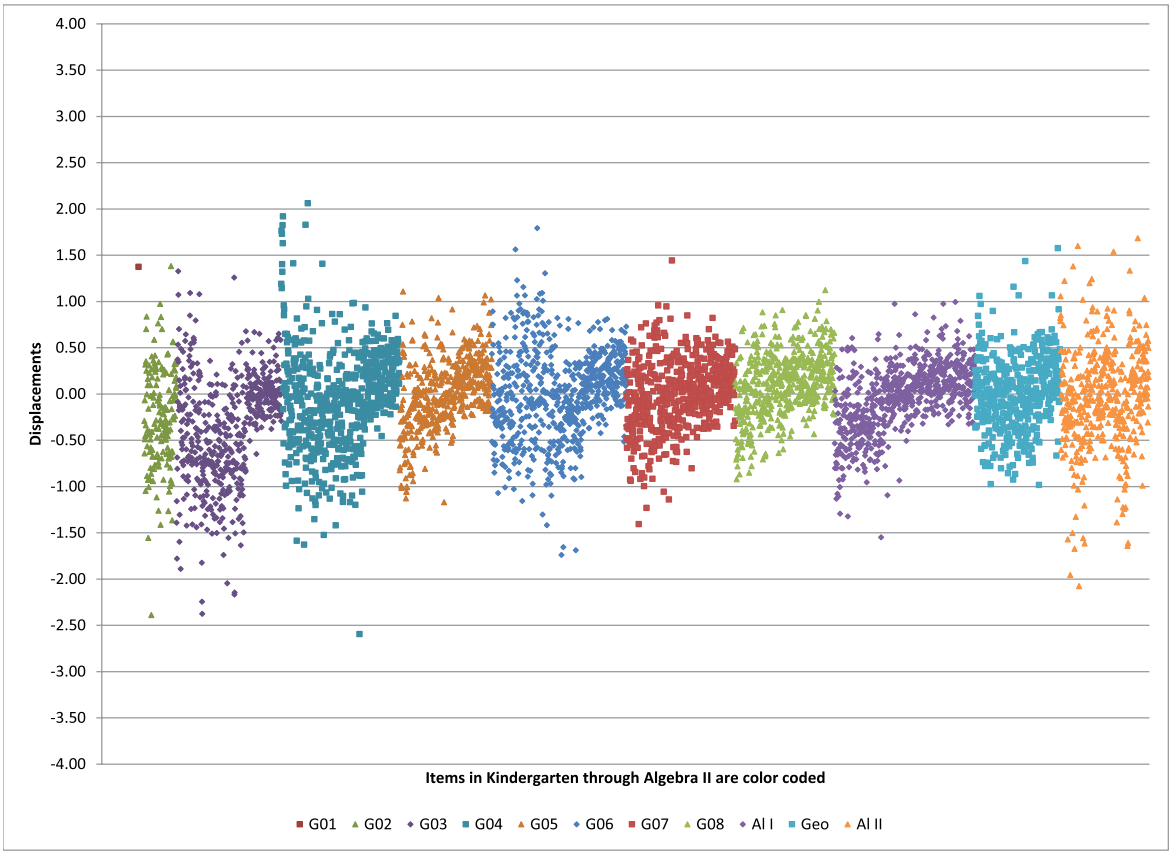
One method used to evaluate the stability of the item parameters in the operational administration was to calibrate the entire bank within a content area anchoring on the banked item parameters and examine the displacements. For each item, the displacement value is the size of the change in the parameter estimate that would be estimated if the parameter for the item was unanchored and all other parameters were anchored at their current value. Given that the banked values were developed into a single, vertical scale, all items within a content area were calibrated in a single concurrent calibration using WINSTEPS software version 3.71 (Linacre, 2009).

MATHEMATICS

Figure 18–1 shows the displacements from a concurrent anchored calibration of all mathematics items using the operational data set. Items are color-coded by grade/course.

[1] For results of all four methods for the 2011–2012 school year, see Chapter Eighteen of the 2011–2012 technical report.

Figure 18–1. Anchored Calibration Displacements — All Items



Note: Many kindergarten and grade 1 items were not estimated by WINSTEPS software due to insufficient counts.

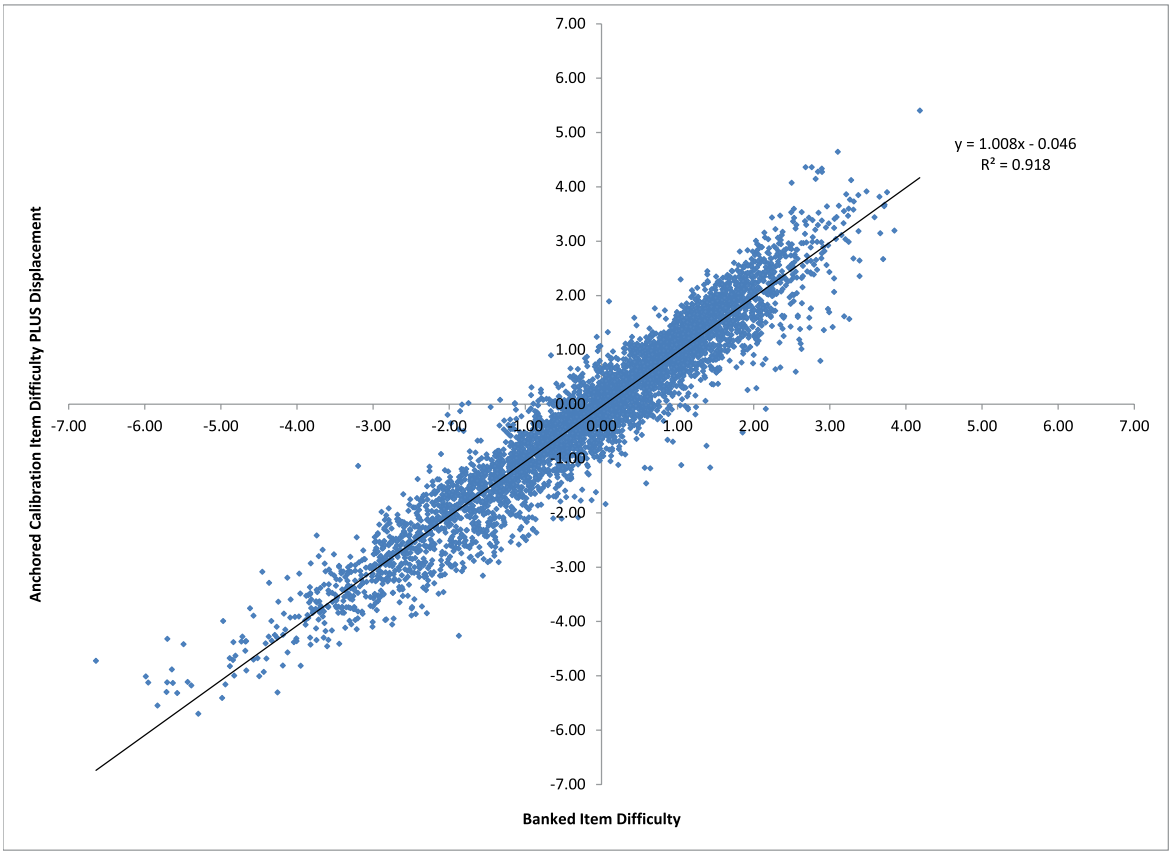
Table 18–1 summarizes the data in Figure 18–1. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Seventy-five percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–1).

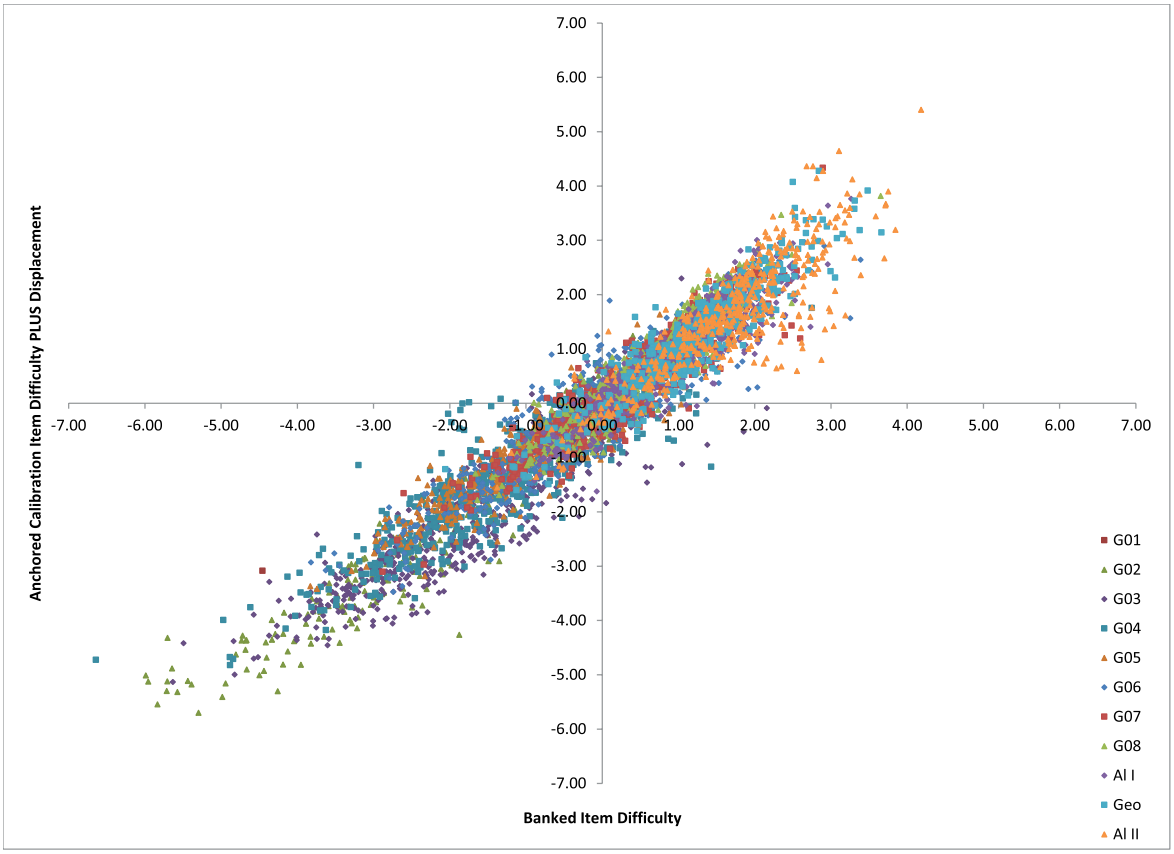
Table 18–1. Number of Mathematics Items by Grade/Course and Displacement Interval

Interval	K	G01	G02	G03	G04	G05	G06	G07	G08	ALI	GEO	ALII	Total
Disp. ≤ -1.0	0	0	10	72	24	7	11	4	0	8	0	19	155
-1.0 < Disp. ≤ -0.9	0	0	7	18	12	3	7	6	1	4	3	6	67
-0.9 < Disp. ≤ -0.8	0	0	3	23	8	5	15	5	2	8	3	6	78
-0.8 < Disp. ≤ -0.7	0	0	6	38	20	6	17	5	4	12	8	7	123
-0.7 < Disp. ≤ -0.6	0	0	18	27	23	6	20	12	10	18	12	18	164
-0.6 < Disp. ≤ -0.5	0	0	4	24	38	10	38	15	10	30	22	14	205
-0.5 < Disp. ≤ -0.4	0	0	15	17	23	19	18	24	12	23	15	26	192
-0.4 < Disp. ≤ -0.3	0	0	14	28	35	22	40	41	17	47	28	27	299
-0.3 < Disp. ≤ -0.2	0	0	9	35	27	39	43	38	29	48	36	34	338
-0.2 < Disp. ≤ -0.1	0	0	11	37	45	45	50	51	40	54	34	38	405
-0.1 < Disp. ≤ 0.0	0	0	13	52	50	52	64	70	50	79	42	36	508
0.0 < Disp. ≤ 0.1	0	0	8	45	51	45	67	55	57	106	48	36	518
0.1 < Disp. ≤ 0.2	0	0	5	32	46	52	55	59	58	86	40	40	473
0.2 < Disp. ≤ 0.3	0	0	14	24	43	48	55	49	54	69	42	26	424
0.3 < Disp. ≤ 0.4	0	0	10	12	38	39	47	37	59	48	28	25	343
0.4 < Disp. ≤ 0.5	0	0	4	8	35	20	41	30	24	24	22	28	236
0.5 < Disp. ≤ 0.6	0	0	3	11	28	22	23	16	25	10	18	13	169
0.6 < Disp. ≤ 0.7	0	0	0	7	10	6	14	13	19	8	13	9	99
0.7 < Disp. ≤ 0.8	0	0	3	2	7	7	17	5	12	3	1	11	68
0.8 < Disp. ≤ 0.9	0	0	2	1	5	2	16	3	5	2	2	7	45
0.9 < Disp. ≤ 1.0	0	0	1	0	7	2	4	2	3	3	2	4	28
1.0 < Disp.	0	1	1	5	14	4	11	1	1	0	6	11	55
TOTAL	0	1	161	518	589	461	673	541	492	690	425	441	4992

Figure 18–2 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored concurrent calibration of operational data for the mathematics item bank. A line of best fit is included in the upper plot. If item difficulties from the operational calibration are close to the banked values, the line will approach an intercept of zero and a slope of one. The lower plot displays the same data as the upper, but color codes items by grade/course in an attempt to lend insight into the possible causes for the deviations.

Figure 18–2. Mathematics Banked Item Parameters vs. Anchored Calibration — All Items





Based on Figure 18–2, one can see that there are a number of items with operational estimates that differ from their banked values. Some of these are in kindergarten through grade 2. Recall that the operational CDT is available to students in grade 3 and above. While items were developed to sample content in kindergarten through grade 2 to provide better diagnostic information for lower-performing students, the data from the operational administration did not include students below grade 3. To investigate whether this had an impact on the stability of the item parameter estimates, a concurrent anchored calibration of all items in grade 3 and above was run.

Figure 18–3 and Table 18–2 summarize the displacements from a concurrent anchored calibration of all items in grade 3 and above. Seventy-five percent of the items in the calibration have displacement less than 0.5 in magnitude (gray shaded in Table 18–2). Figure 18–4 shows banked item difficulties plotted against the item difficulties plus displacement. Again, a line of best fit is included in the upper plot.

Figure 18–3. Mathematics Anchored Calibration Displacements — All Items in Grade 3 and Above

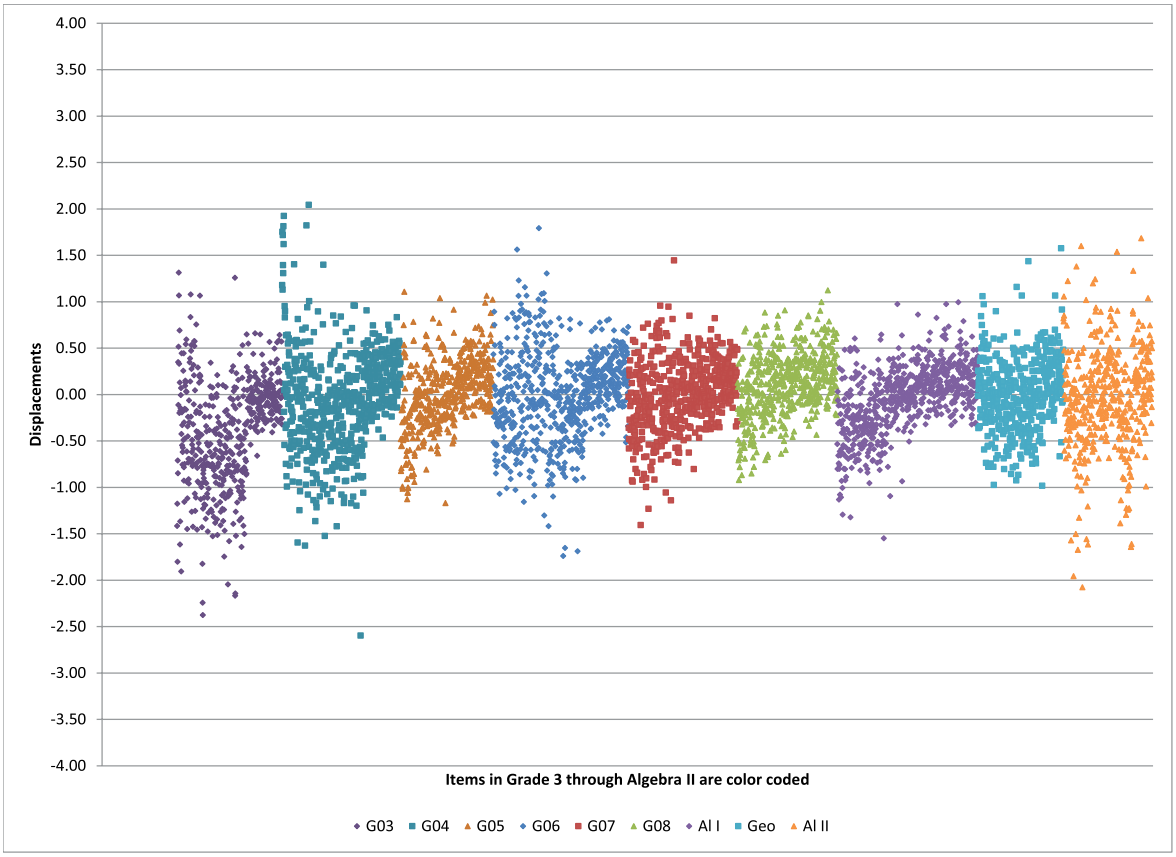
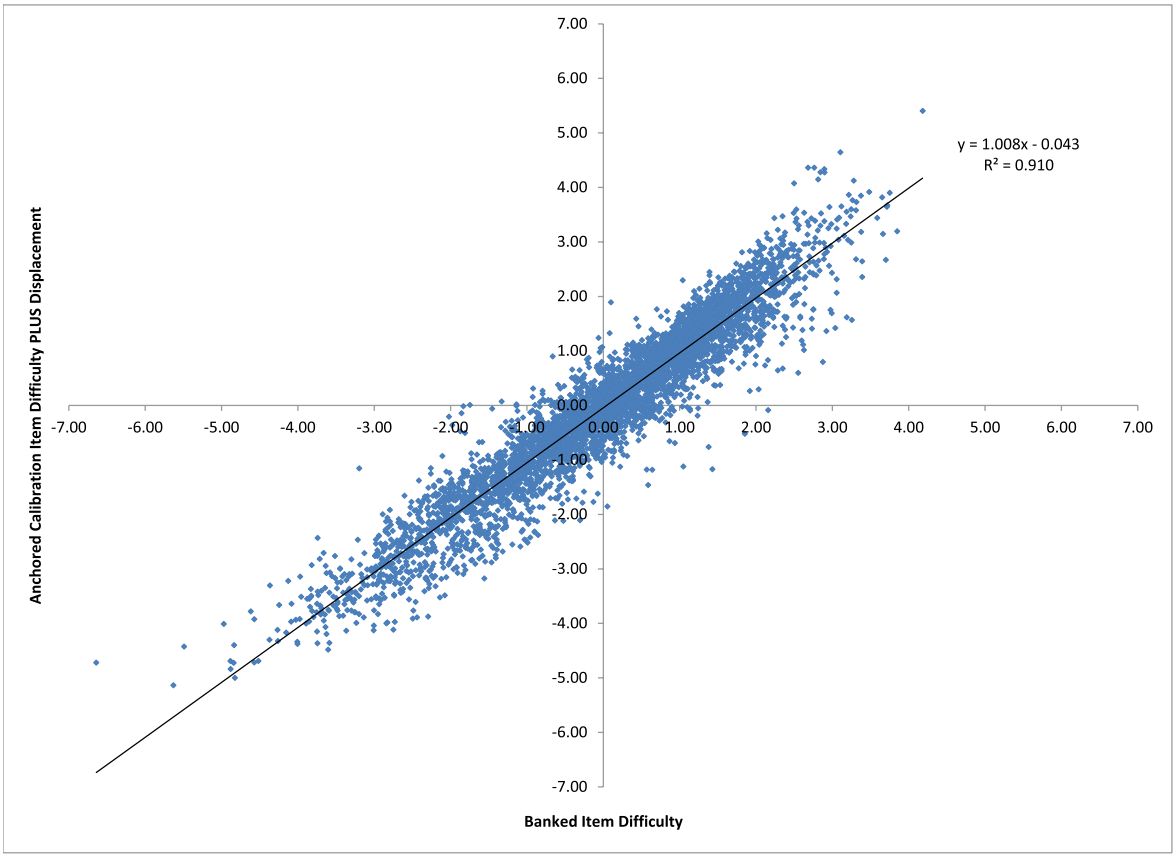
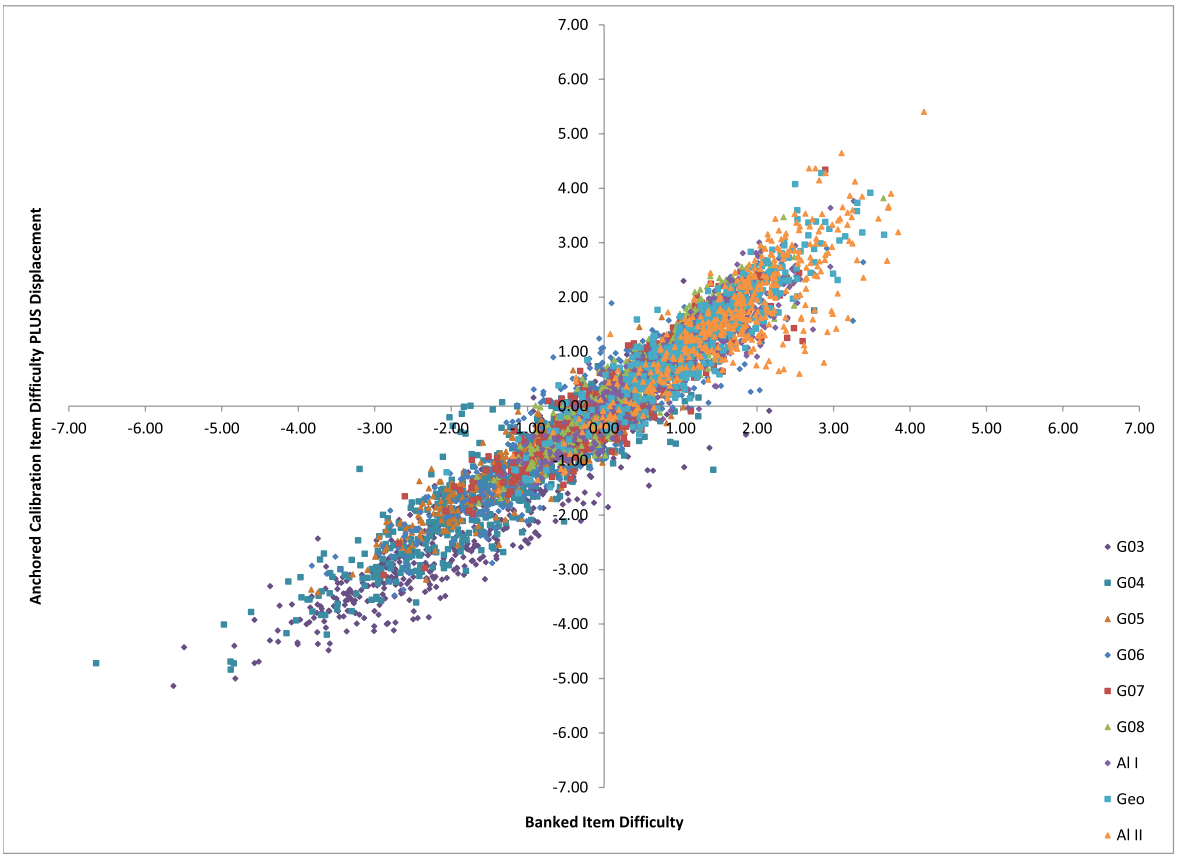


Table 18–2. Number of Mathematics Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	ALI	GEO	ALII	Total
Disp. ≤ -1.0	73	24	7	11	4	0	8	0	19	146
-1.0 < Disp. ≤ -0.9	21	12	3	7	6	1	4	3	6	63
-0.9 < Disp. ≤ -0.8	23	9	5	15	5	2	8	3	6	76
-0.8 < Disp. ≤ -0.7	38	21	6	17	5	4	12	8	7	118
-0.7 < Disp. ≤ -0.6	29	25	6	20	12	10	18	12	18	150
-0.6 < Disp. ≤ -0.5	19	35	10	38	15	10	30	22	14	193
-0.5 < Disp. ≤ -0.4	23	23	19	18	24	12	23	15	26	183
-0.4 < Disp. ≤ -0.3	30	35	22	40	41	17	47	28	27	287
-0.3 < Disp. ≤ -0.2	31	29	39	43	38	29	48	36	34	327
-0.2 < Disp. ≤ -0.1	47	49	45	50	51	40	54	34	38	408
-0.1 < Disp. ≤ 0.0	48	50	52	64	70	50	79	42	36	491
0.0 < Disp. ≤ 0.1	42	52	45	67	55	57	106	48	36	508
0.1 < Disp. ≤ 0.2	30	42	52	55	59	58	86	40	40	462
0.2 < Disp. ≤ 0.3	21	43	48	55	49	54	69	42	26	407
0.3 < Disp. ≤ 0.4	11	35	39	47	37	59	48	28	25	329
0.4 < Disp. ≤ 0.5	7	36	20	41	30	24	24	22	28	232
0.5 < Disp. ≤ 0.6	13	27	22	23	16	25	10	18	13	167
0.6 < Disp. ≤ 0.7	5	9	6	14	13	19	8	13	9	96
0.7 < Disp. ≤ 0.8	1	7	7	17	5	12	3	1	11	64
0.8 < Disp. ≤ 0.9	1	7	2	16	3	5	2	2	7	45
0.9 < Disp. ≤ 1.0	0	5	2	4	2	3	3	2	4	25
1.0 < Disp.	5	14	4	11	1	1	0	6	11	53
TOTAL	518	589	461	673	541	492	690	425	441	4830

Figure 18–4. Mathematics Banked Item Parameters vs. Anchored Calibration — All Items in Grade 3 and Above





It is evident from this series of plots that the item parameter estimates are reasonably stable for the items in grade 3 and above.

For both of the anchored calibrations described in this section, banked item parameters were compared to the banked item parameters plus the displacements by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–3 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in each of the calibrations.

Table 18–3. Summary of Robust Z across Anchored Calibrations in Mathematics

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
Kindergarten	0	0	N/A	0	0	N/A
Grade 1	1	1	100%	0	0	N/A
Grade 2	161	36	22%	0	0	N/A
Grade 3	518	160	31%	518	170	33%
Grade 4	589	101	17%	589	106	18%
Grade 5	461	36	8%	461	36	8%
Grade 6	673	102	15%	673	102	15%
Grade 7	541	35	6%	541	35	6%
Grade 8	492	30	6%	492	34	7%
Algebra I	690	42	6%	690	43	6%
Geometry	425	26	6%	425	30	7%
Algebra II	441	74	17%	441	80	18%
Total	4992	643	13%	4830	636	13%
	Correlation = 0.958			Correlation = 0.954		
	Additive Constant = -0.046			Additive Constant = -0.043		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 2, all items with absolute displacement greater than 0.681 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.674 to 0.681, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.674 have absolute value of robust Z greater than 1.645.

READING/LITERATURE

Figure 18–5 shows the displacements from a concurrent anchored calibration of all reading items using the operational data set. Items are color-coded by grade/course.

Figure 18–5. Reading Anchored Calibration Displacements — All Items

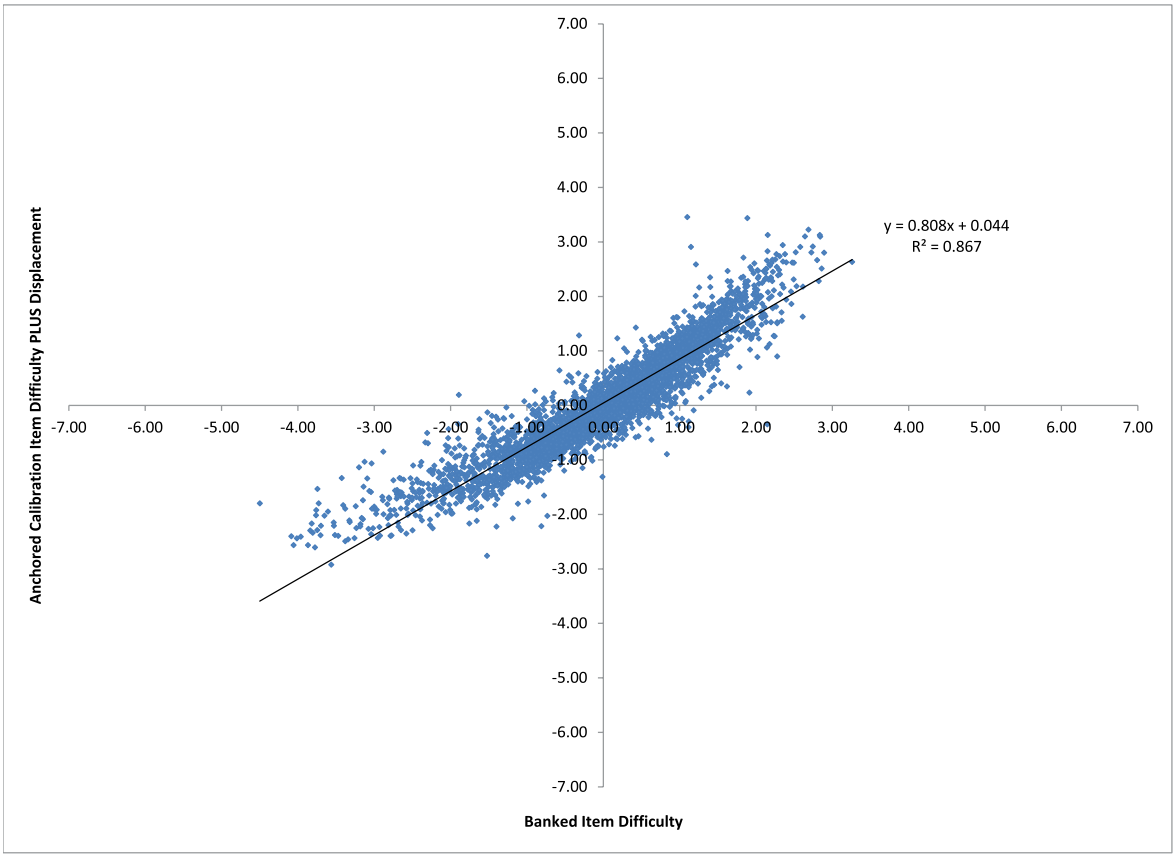
Table 18–4 summarizes the data in Figure 18–5. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-two percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–4).

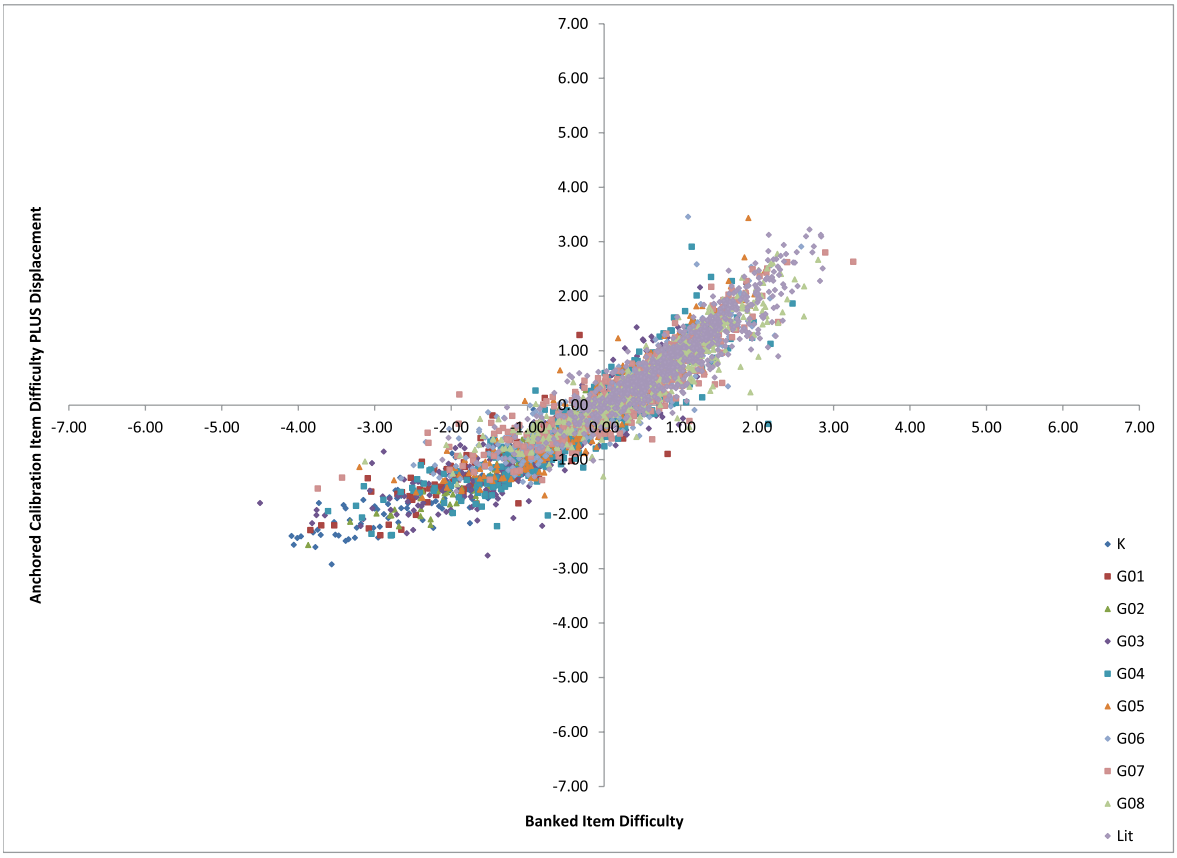
Table 18–4. Number of Reading Items by Grade/Course and Displacement Interval

Interval	K	G01	G02	G03	G04	G05	G06	G07	G08	LIT	Total
Disp. ≤ -1.0	0	1	0	5	5	0	3	4	9	4	31
-1.0 < Disp. ≤ -0.9	0	0	0	1	1	0	1	0	1	2	6
-0.9 < Disp. ≤ -0.8	0	1	0	5	5	1	4	3	1	5	25
-0.8 < Disp. ≤ -0.7	0	0	0	5	6	1	6	7	8	8	41
-0.7 < Disp. ≤ -0.6	0	2	2	6	11	5	4	11	14	19	74
-0.6 < Disp. ≤ -0.5	0	0	0	3	13	8	9	15	13	25	86
-0.5 < Disp. ≤ -0.4	1	2	1	15	14	12	20	18	17	31	131
-0.4 < Disp. ≤ -0.3	0	0	1	20	25	20	28	24	39	42	199
-0.3 < Disp. ≤ -0.2	3	2	8	48	32	37	41	48	45	51	315
-0.2 < Disp. ≤ -0.1	5	4	11	50	55	65	58	50	55	76	429
-0.1 < Disp. ≤ 0.0	5	6	15	40	59	60	75	60	84	103	507
0.0 < Disp. ≤ 0.1	5	6	21	54	59	52	79	73	54	113	516
0.1 < Disp. ≤ 0.2	4	7	21	53	44	36	41	52	49	93	400
0.2 < Disp. ≤ 0.3	10	8	20	25	33	35	41	32	32	46	282
0.3 < Disp. ≤ 0.4	7	12	9	14	21	18	19	20	11	35	166
0.4 < Disp. ≤ 0.5	8	9	7	14	11	9	12	16	13	21	120
0.5 < Disp. ≤ 0.6	9	10	5	17	15	8	17	10	9	14	114
0.6 < Disp. ≤ 0.7	6	9	1	12	8	5	12	5	2	7	67
0.7 < Disp. ≤ 0.8	7	5	3	8	3	6	5	7	3	4	51
0.8 < Disp. ≤ 0.9	7	8	1	5	4	3	4	6	4	2	44
0.9 < Disp. ≤ 1.0	5	1	1	6	4	0	4	2	3	4	30
1.0 < Disp.	26	14	2	20	12	7	13	9	5	6	114
TOTAL	108	107	129	426	440	388	496	472	471	711	3748

Figure 18–6 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored concurrent calibration of operational data for the reading item bank. A line of best fit is included in the upper plot. The lower plot displays the same data as the upper, but color codes items by grade/course in an attempt to lend insight into the possible causes for the deviations.

Figure 18–6. Reading Banked Item Parameters vs. Anchored Calibration — All Items





Based on Figure 18–6, one can see that there are a number of items with operational estimates that differ from their banked values. Some of these are in kindergarten through grade 2. Recall that the operational CDT is available to students in grade 3 and above. While items were developed to sample content in kindergarten through grade 2 to provide better diagnostic information for lower performing students, the data from the operational administration did not include students below grade 3. To investigate whether this had an impact on the stability of the item parameter estimates, a concurrent anchored calibration of all items in grade 3 and above was run.

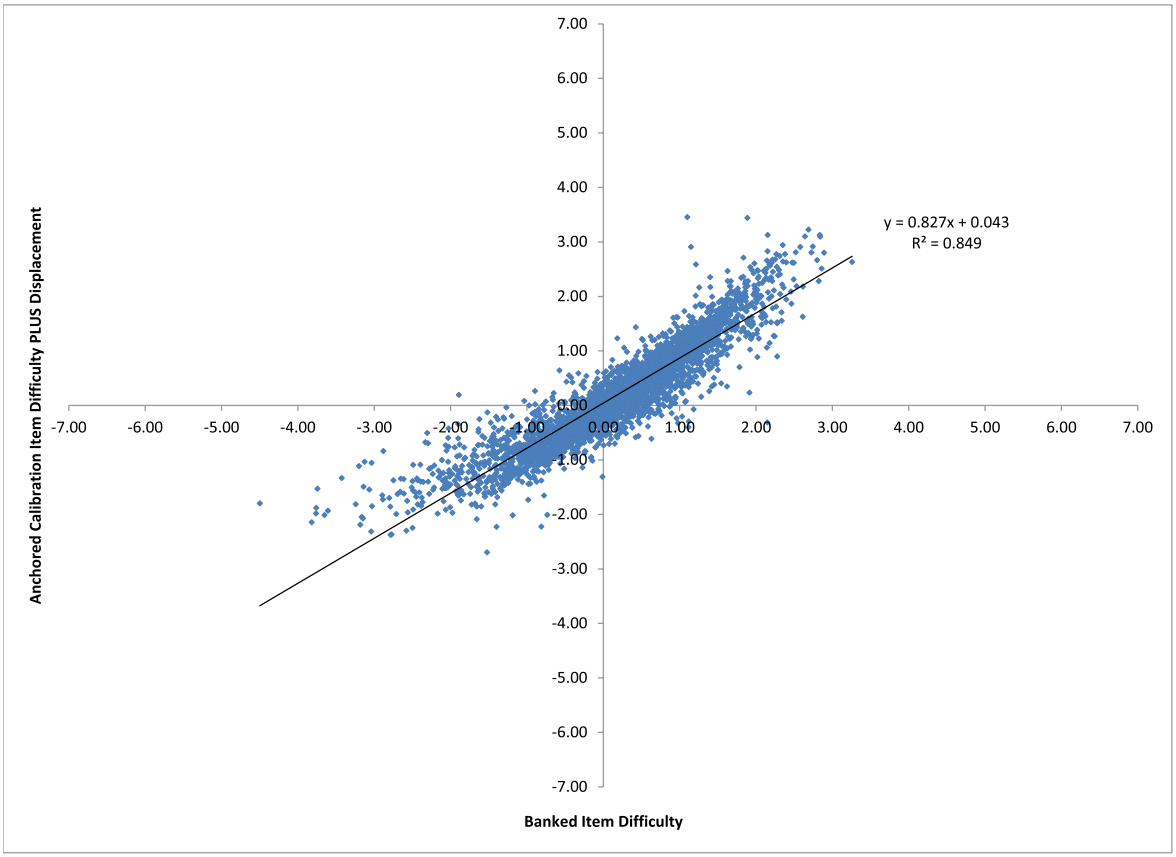
Figure 18–7 and Table 18–5 summarize the displacements from a concurrent anchored calibration of all items in grade 3 and above. Eighty-four percent of the items in the calibration have displacement less than 0.5 in magnitude (gray shaded in Table 18–5). Figure 18–8 shows banked item difficulties plotted against the item difficulties plus displacement. Again, a line of best fit is included in the upper plot.

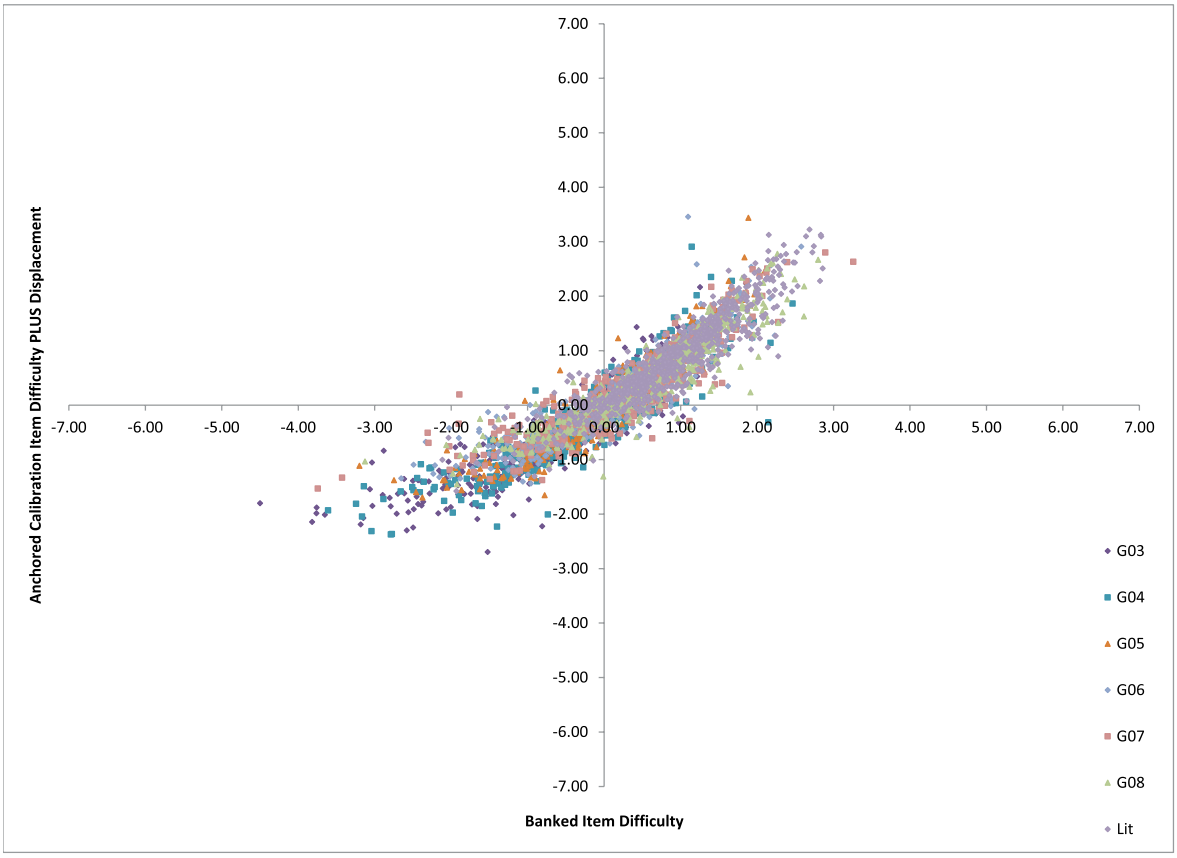
Figure 18–7. Reading Anchored Calibration Displacements — All Items in Grade 3 and Above

Table 18–5. Number of Reading Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	LIT	Total
Disp. ≤ -1.0	4	5	0	3	4	9	4	29
-1.0 < Disp. ≤ -0.9	2	1	0	1	0	1	2	7
-0.9 < Disp. ≤ -0.8	5	4	1	4	3	1	5	23
-0.8 < Disp. ≤ -0.7	4	6	1	6	6	8	8	39
-0.7 < Disp. ≤ -0.6	6	9	5	3	12	14	18	67
-0.6 < Disp. ≤ -0.5	2	16	6	9	15	13	26	87
-0.5 < Disp. ≤ -0.4	13	13	13	21	18	17	30	125
-0.4 < Disp. ≤ -0.3	16	17	20	28	24	39	43	187
-0.3 < Disp. ≤ -0.2	35	37	34	41	48	45	51	291
-0.2 < Disp. ≤ -0.1	59	47	67	54	49	55	76	407
-0.1 < Disp. ≤ 0.0	46	61	61	78	61	84	103	494
0.0 < Disp. ≤ 0.1	44	65	49	78	73	54	113	476
0.1 < Disp. ≤ 0.2	57	44	39	42	52	49	93	376
0.2 < Disp. ≤ 0.3	30	32	34	42	31	31	46	246
0.3 < Disp. ≤ 0.4	21	24	19	19	20	12	35	150
0.4 < Disp. ≤ 0.5	14	13	10	12	17	13	21	100
0.5 < Disp. ≤ 0.6	14	12	8	17	10	9	14	84
0.6 < Disp. ≤ 0.7	12	9	5	11	5	2	7	51
0.7 < Disp. ≤ 0.8	10	4	5	6	7	3	4	39
0.8 < Disp. ≤ 0.9	5	5	4	4	5	4	2	29
0.9 < Disp. ≤ 1.0	7	3	0	4	3	3	4	24
1.0 < Disp.	20	13	7	13	9	5	6	73
TOTAL	426	440	388	496	472	471	711	3404

Figure 18–8. Reading Banked Item Parameters vs. Anchored Calibration — All Items in Grade 3 and Above





It is evident from this series of plots that the item parameter estimates are reasonably stable for the items in grade 3 and above.

For both of the anchored calibrations described in this section, banked item parameters were compared to the banked item parameters plus the displacements by calculating a robust Z statistic for each item pairing. Table 18–6 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in each of the calibrations.

Table 18–6. Summary of Robust Z across Anchored Calibrations in Reading

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
Kindergarten	108	60	56%	0	0	N/A
Grade 1	107	53	50%	0	0	N/A
Grade 2	129	16	12%	0	0	N/A
Grade 3	426	95	22%	426	102	24%
Grade 4	440	89	20%	440	93	21%
Grade 5	388	45	12%	388	48	12%
Grade 6	496	84	17%	496	86	17%
Grade 7	472	80	17%	472	85	18%
Grade 8	471	76	16%	471	81	17%
Literature	711	105	15%	711	116	16%
Total	3748	703	19%	3404	611	18%
	Correlation = 0.931			Correlation = 0.921		
	Additive Constant = 0.039			Additive Constant = 0.010		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.493 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.482 to 0.493, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.482 have absolute value of robust Z greater than 1.645.

SCIENCE

Figure 18–9 shows the displacements from a concurrent anchored calibration of all science items using the operational data set. Items are color-coded by grade/course.

Figure 18–9. Science Anchored Calibration Displacements — All Items

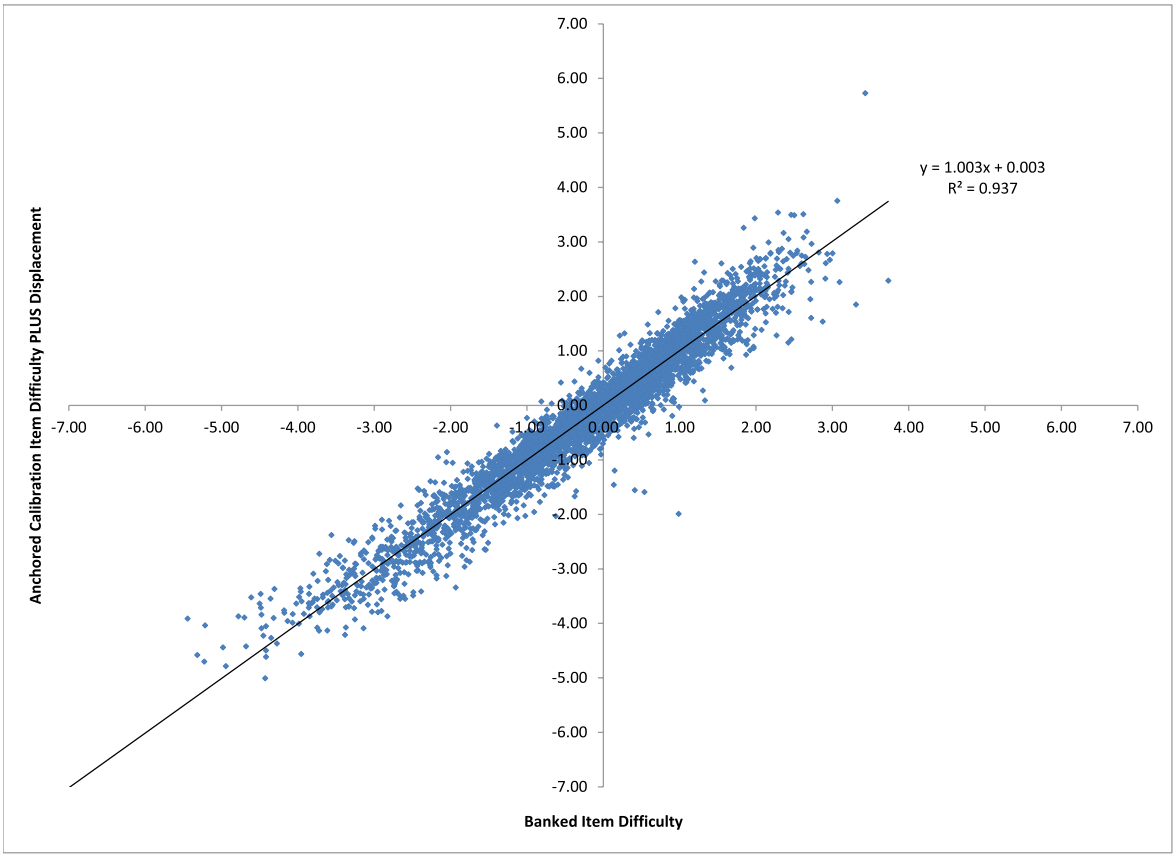
Table 18–7 summarizes the data in Figure 18–9. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-seven percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–7).

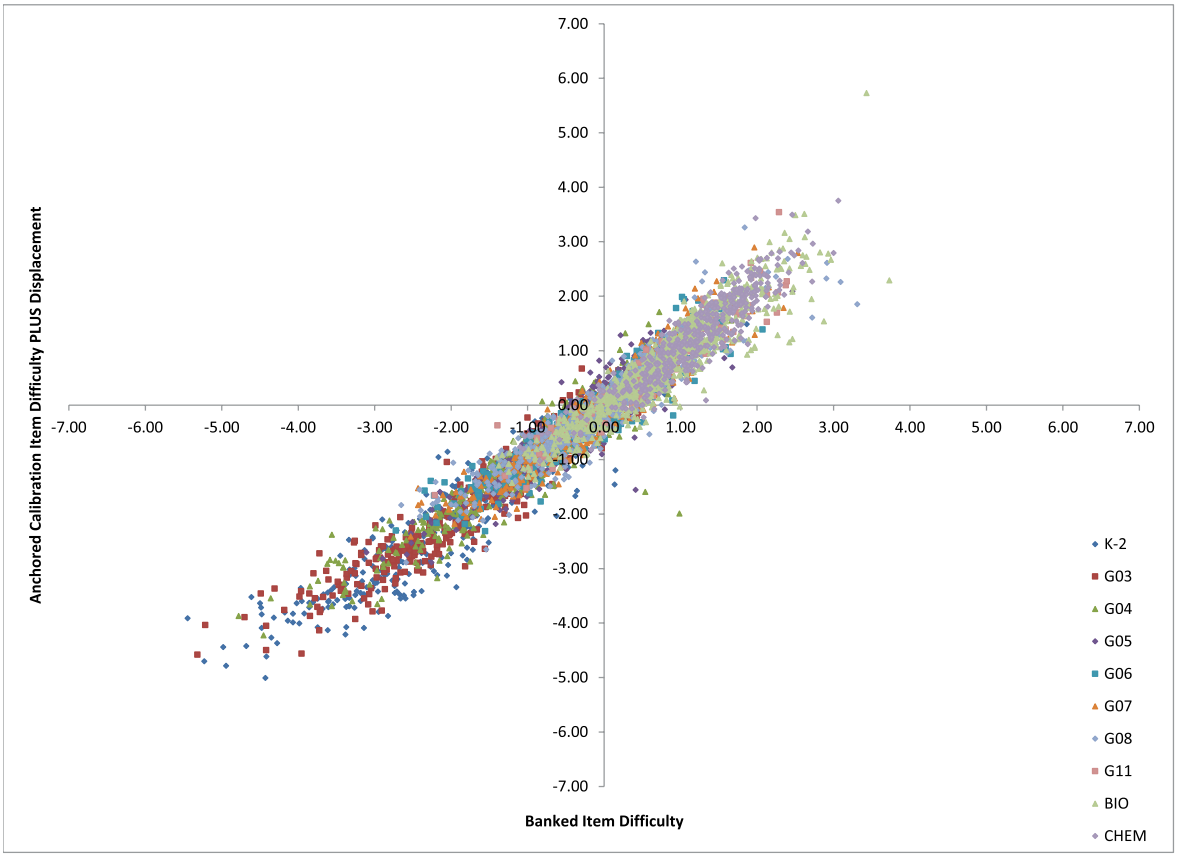
Table 18–7. Number of Science Items by Grade/Course and Displacement Interval

Interval	K–2	G03	G04	G05	G06	G07	G08	G11	BIO	CHEM	Total
Disp. ≤ -1.0	12	3	3	1	1	0	4	0	6	1	31
-1.0 < Disp. ≤ -0.9	7	1	1	2	1	0	0	0	4	0	16
-0.9 < Disp. ≤ -0.8	9	2	5	2	2	2	4	0	4	0	30
-0.8 < Disp. ≤ -0.7	8	5	6	2	3	1	7	0	4	3	39
-0.7 < Disp. ≤ -0.6	4	9	9	9	6	9	7	1	8	5	67
-0.6 < Disp. ≤ -0.5	21	17	8	6	8	18	17	2	22	9	128
-0.5 < Disp. ≤ -0.4	14	15	15	13	11	27	17	3	35	15	165
-0.4 < Disp. ≤ -0.3	22	28	21	19	21	32	41	7	41	24	256
-0.3 < Disp. ≤ -0.2	28	30	34	24	30	45	55	12	69	33	360
-0.2 < Disp. ≤ -0.1	20	33	42	25	32	52	64	15	70	51	404
-0.1 < Disp. ≤ 0.0	25	37	60	43	41	79	88	11	118	35	537
0.0 < Disp. ≤ 0.1	32	45	45	41	45	75	90	15	118	53	559
0.1 < Disp. ≤ 0.2	22	33	54	46	46	92	96	13	104	46	552
0.2 < Disp. ≤ 0.3	27	26	45	37	53	47	68	7	90	49	449
0.3 < Disp. ≤ 0.4	15	22	27	34	15	22	56	5	55	43	294
0.4 < Disp. ≤ 0.5	8	14	16	19	12	16	34	5	29	26	179
0.5 < Disp. ≤ 0.6	11	7	16	10	4	8	12	2	16	15	101
0.6 < Disp. ≤ 0.7	6	3	4	5	5	6	2	2	9	9	51
0.7 < Disp. ≤ 0.8	3	6	4	3	2	2	3	0	3	2	28
0.8 < Disp. ≤ 0.9	3	1	4	2	3	1	2	0	4	2	22
0.9 < Disp. ≤ 1.0	0	3	3	1	1	3	2	0	1	0	14
1.0 < Disp.	4	3	3	1	0	0	3	2	2	2	20
TOTAL	301	343	425	345	342	537	672	102	812	423	4302

Figure 18–10 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored concurrent calibration of operational data for the science item bank. A line of best fit is included in the upper plot. If item difficulties from the operational calibration are close to the banked values, the line will approach an intercept of zero and a slope of one. The lower plot displays the same data as the upper, but color codes items by grade/course in an attempt to lend insight into the possible causes for the deviations.

Figure 18–10. Science Banked Item Parameters vs. Anchored Calibration — All Items





Based on Figure 18–10, one can see that there are a number of items with operational estimates that differ from their banked values. Some of these are in the K–2 span. Recall that the operational CDT is available to students in grade 3 and above. While items were developed to sample content in the K–2 span to provide better diagnostic information for lower performing students, the data from the operational administration did not include students below grade 3. To investigate whether this had an impact on the stability of the item parameter estimates, a concurrent anchored calibration of all items in grade 3 and above was run.

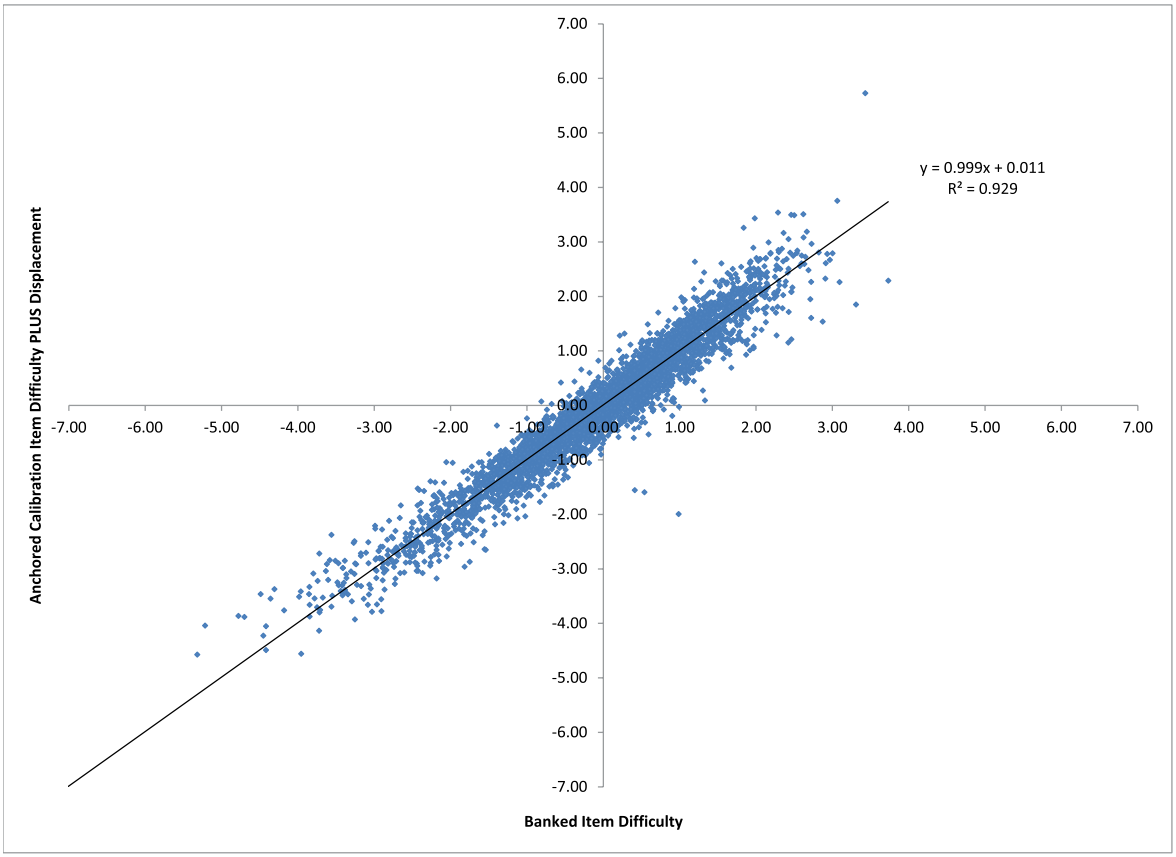
Figure 18–11 and Table 18–8 summarize the displacements from a concurrent anchored calibration of all items in grade 3 and above. Eighty-nine percent of the items in the calibration have displacement less than 0.5 in magnitude (gray shaded in Table 18–8). Figure 18–12 shows banked item difficulties plotted against the item difficulties plus displacement. Again, a line of best fit is included in the upper plot.

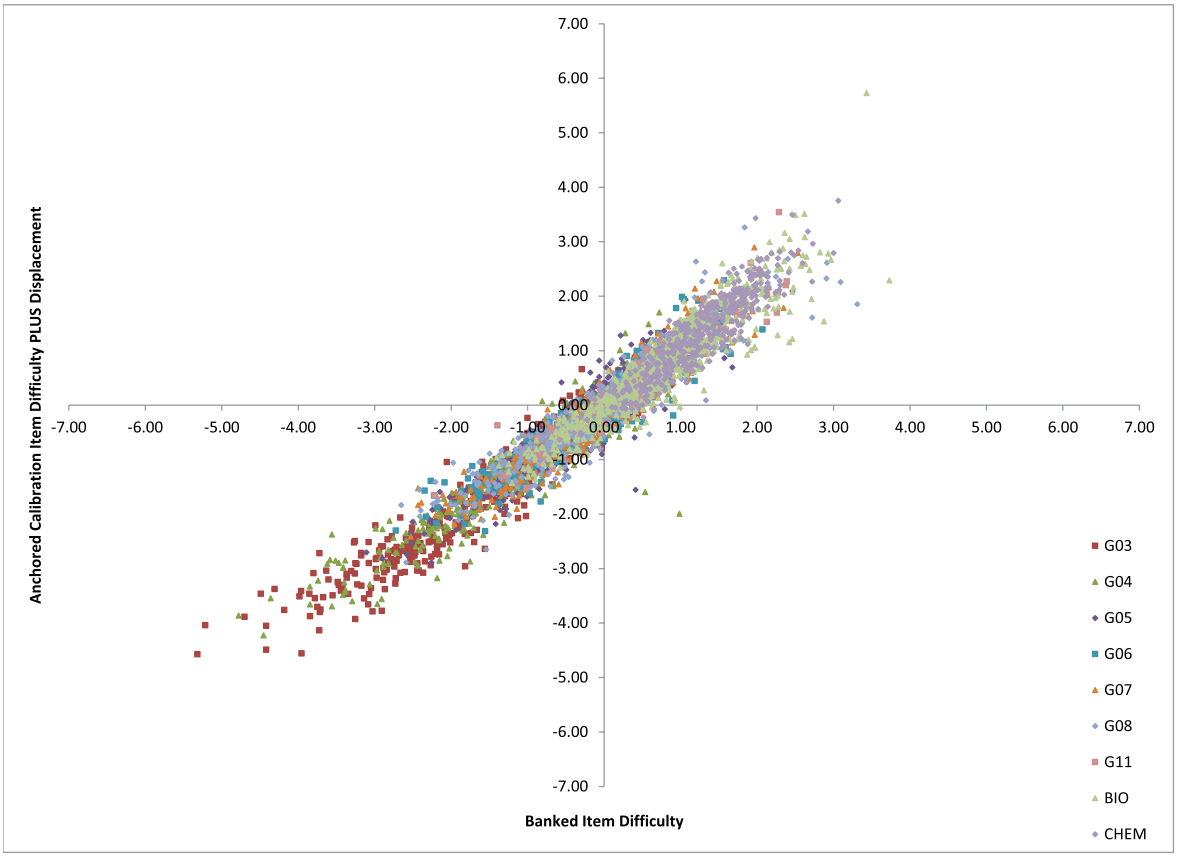
Figure 18–11. Science Anchored Calibration Displacements — All Items in Grade 3 and Above

Table 18–8. Number of Science Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	G11	BIO	CHEM	Total
Disp. ≤ -1.0	3	3	1	1	0	4	0	6	1	19
-1.0 < Disp. ≤ -0.9	1	1	2	1	0	0	0	4	0	9
-0.9 < Disp. ≤ -0.8	2	5	2	2	2	4	0	4	0	21
-0.8 < Disp. ≤ -0.7	5	6	2	3	1	7	0	4	3	31
-0.7 < Disp. ≤ -0.6	9	9	9	6	9	7	1	8	5	63
-0.6 < Disp. ≤ -0.5	17	8	6	8	18	17	2	22	9	107
-0.5 < Disp. ≤ -0.4	16	15	13	11	27	17	3	35	15	152
-0.4 < Disp. ≤ -0.3	28	21	19	21	32	41	7	41	24	234
-0.3 < Disp. ≤ -0.2	29	35	24	30	45	55	12	69	33	332
-0.2 < Disp. ≤ -0.1	35	41	25	32	52	64	15	70	51	385
-0.1 < Disp. ≤ 0.0	36	63	43	41	79	88	11	118	35	514
0.0 < Disp. ≤ 0.1	44	44	41	45	75	90	15	118	53	525
0.1 < Disp. ≤ 0.2	34	54	46	46	92	96	13	104	46	531
0.2 < Disp. ≤ 0.3	27	44	38	53	47	68	7	90	49	423
0.3 < Disp. ≤ 0.4	20	26	33	15	22	56	5	55	43	275
0.4 < Disp. ≤ 0.5	14	16	19	12	16	34	5	29	26	171
0.5 < Disp. ≤ 0.6	7	16	10	4	8	12	2	16	15	90
0.6 < Disp. ≤ 0.7	3	4	5	5	6	2	2	9	9	45
0.7 < Disp. ≤ 0.8	6	4	3	2	2	3	0	3	2	25
0.8 < Disp. ≤ 0.9	1	4	2	2	1	2	0	4	2	18
0.9 < Disp. ≤ 1.0	2	3	1	2	3	2	0	1	0	14
1.0 < Disp.	4	3	1	0	0	3	2	2	2	17
TOTAL	343	425	345	342	537	672	102	812	423	4001

Figure 18–12. Science Banked Item Parameters vs. Anchored Calibration — All Items in Grade 3 and Above





It is evident from this series of plots that the item parameter estimates are reasonably stable for the items in grade 3 and above.

For both of the anchored calibrations described in this section, banked item parameters were compared to the banked item parameters plus the displacements by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–9 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in each of the calibrations.

Table 18–9. Summary of Robust Z across Anchored Calibrations in Science

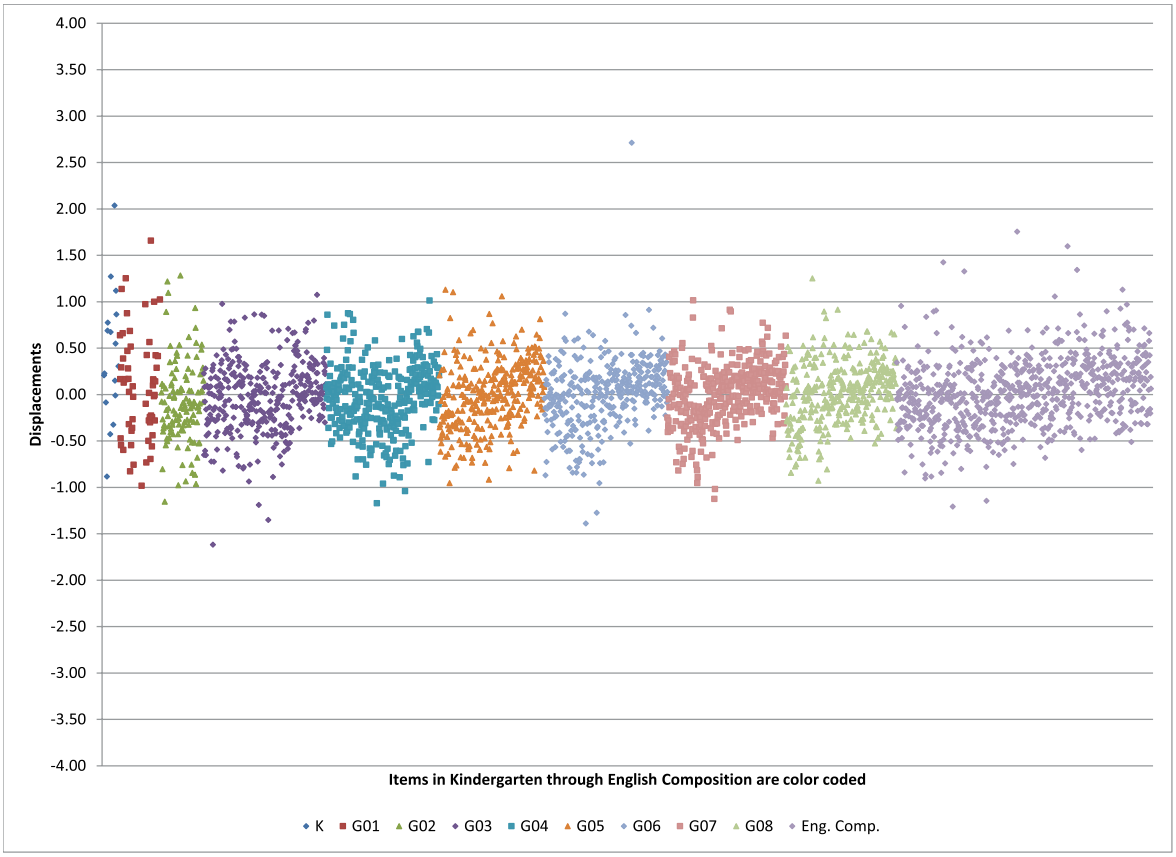
Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
K–2 span	301	87	29%	0	0	N/A
Grade 3	343	58	17%	343	60	17%
Grade 4	425	59	14%	425	66	16%
Grade 5	345	42	12%	345	44	13%
Grade 6	342	38	11%	342	39	11%
Grade 7	537	53	10%	537	59	11%
Grade 8	672	59	9%	672	61	9%
Grade 11	102	9	9%	102	11	11%
Biology	812	85	10%	812	92	11%
Chemistry	423	44	10%	423	47	11%
Total	4302	534	12%	4001	479	12%
	Correlation = 0.968			Correlation = 0.964		
	Additive Constant = 0.003			Additive Constant = 0.011		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.537 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.485 to 0.537, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.485 have absolute value of robust Z greater than 1.645.

WRITING/ENGLISH COMPOSITION

Figure 18–13 shows the displacements from a concurrent anchored calibration of all writing items using the operational data set. Items are color-coded by grade/course.

Figure 18–13. Writing Anchored Calibration Displacements — All Items



Note: Many kindergarten items were not estimated by WINSTEPS software due to insufficient counts.

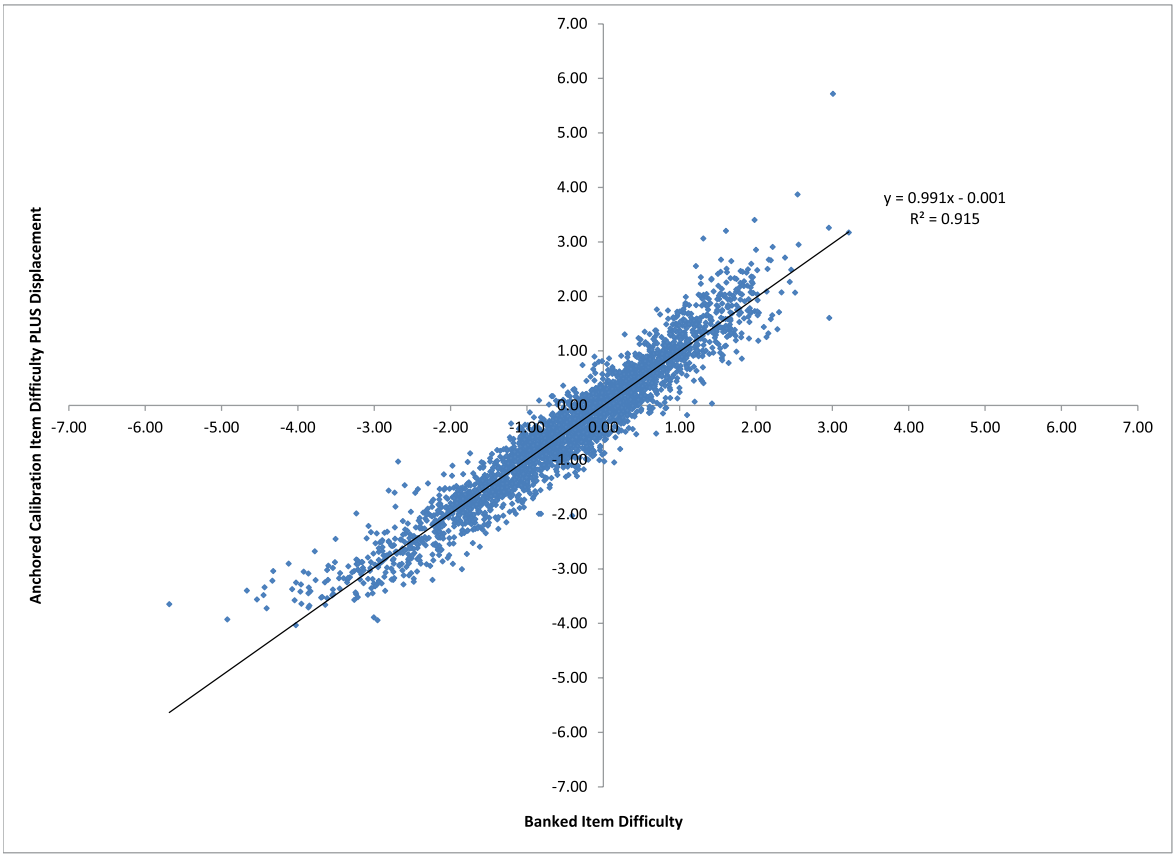
Table 18–10 summarizes the data in Figure 18–13. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-five percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–10).

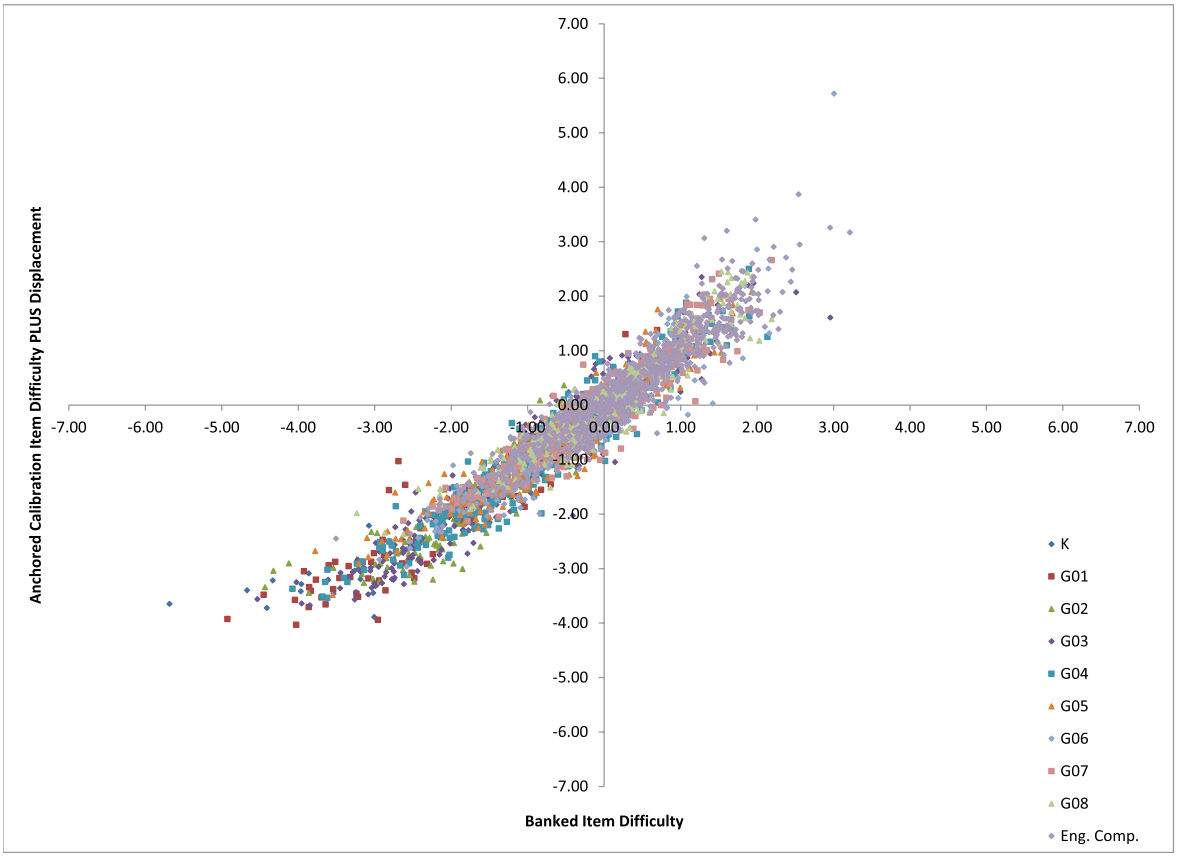
Table 18–10. Number of Writing Items by Grade/Course and Displacement Interval

Interval	K	G01	G02	G03	G04	G05	G06	G07	G08	COMP	Total
Disp. ≤ -1.0	0	0	1	3	2	0	2	2	0	2	12
-1.0 < Disp. ≤ -0.9	0	1	3	1	1	2	1	1	1	1	12
-0.9 < Disp. ≤ -0.8	1	1	3	2	3	1	4	3	2	5	25
-0.8 < Disp. ≤ -0.7	0	2	2	8	8	5	8	3	2	4	42
-0.7 < Disp. ≤ -0.6	0	1	1	7	8	8	8	7	7	9	56
-0.6 < Disp. ≤ -0.5	0	5	8	10	12	13	8	3	8	21	88
-0.5 < Disp. ≤ -0.4	1	2	6	18	15	13	13	14	10	32	124
-0.4 < Disp. ≤ -0.3	1	4	13	24	24	24	16	23	25	60	214
-0.3 < Disp. ≤ -0.2	0	6	14	36	31	22	25	24	23	72	253
-0.2 < Disp. ≤ -0.1	0	1	13	39	36	28	31	43	33	65	289
-0.1 < Disp. ≤ 0.0	2	3	14	40	32	36	45	32	38	78	320
0.0 < Disp. ≤ 0.1	0	4	14	47	44	30	49	50	46	87	371
0.1 < Disp. ≤ 0.2	1	7	6	37	37	37	58	53	36	77	349
0.2 < Disp. ≤ 0.3	2	3	7	25	33	34	30	40	34	64	272
0.3 < Disp. ≤ 0.4	1	1	2	25	14	17	26	16	22	69	193
0.4 < Disp. ≤ 0.5	0	4	5	15	11	19	20	20	13	27	134
0.5 < Disp. ≤ 0.6	1	2	4	6	6	7	5	6	13	20	70
0.6 < Disp. ≤ 0.7	2	3	1	6	6	6	6	2	8	14	54
0.7 < Disp. ≤ 0.8	1	0	1	6	3	1	2	3	0	11	28
0.8 < Disp. ≤ 0.9	1	1	1	4	4	3	2	2	2	4	24
0.9 < Disp. ≤ 1.0	0	2	1	1	0	0	1	1	1	4	11
1.0 < Disp.	3	4	3	1	1	3	1	1	1	7	25
TOTAL	17	57	123	361	331	309	361	349	325	733	2966

Figure 18–14 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored concurrent calibration of operational data for the writing item bank. A line of best fit is included in the upper plot. If item difficulties from the operational calibration are close to the banked values, the line will approach an intercept of zero and a slope of one. The lower plot displays the same data as the upper, but color codes items by grade/course in an attempt to lend insight into the possible causes for the deviations.

Figure 18–14. Writing Banked Item Parameters vs. Anchored Calibration — All Items





Based on Figure 18–14, one can see that there are a number of items with operational estimates that differ from their banked values. Some of these are in kindergarten through grade 2. Recall that the operational CDT is available to students in grade 3 and above. While items were developed to sample content in kindergarten through grade 2 to provide better diagnostic information for lower performing students, the data from the operational administration did not include students below grade 3. To investigate whether this had an impact on the stability of the item parameter estimates, a concurrent anchored calibration of all items in grade 3 and above was run.

Figure 18–15 and Table 18–11 summarize the displacements from a concurrent anchored calibration of all items in grade 3 and above. Eighty-six percent of the items in the calibration have displacement less than 0.5 in magnitude (gray shaded in Table 18–11). Figure 18–16 shows banked item difficulties plotted against the item difficulties plus displacement. Again, a line of best fit is included in the upper plot.

Figure 18–15. Writing Anchored Calibration Displacements — All Items in Grade 3 and Above

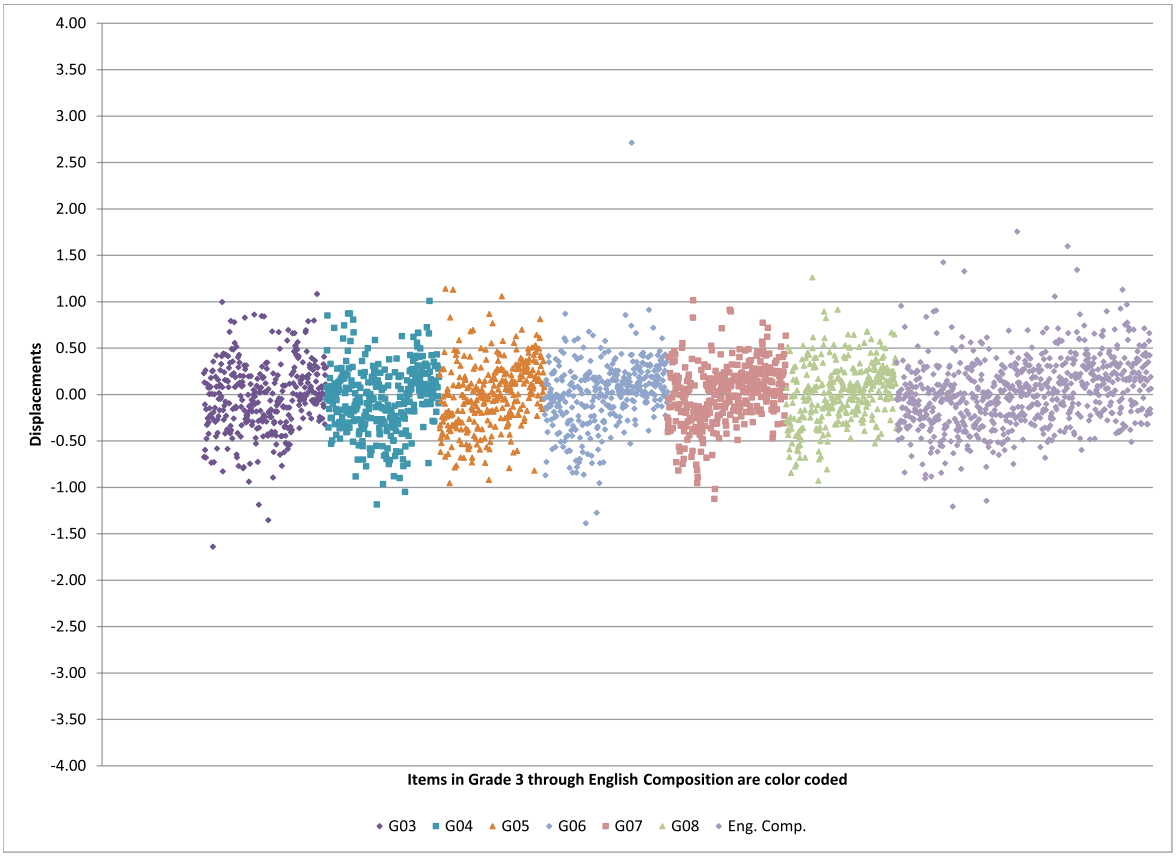
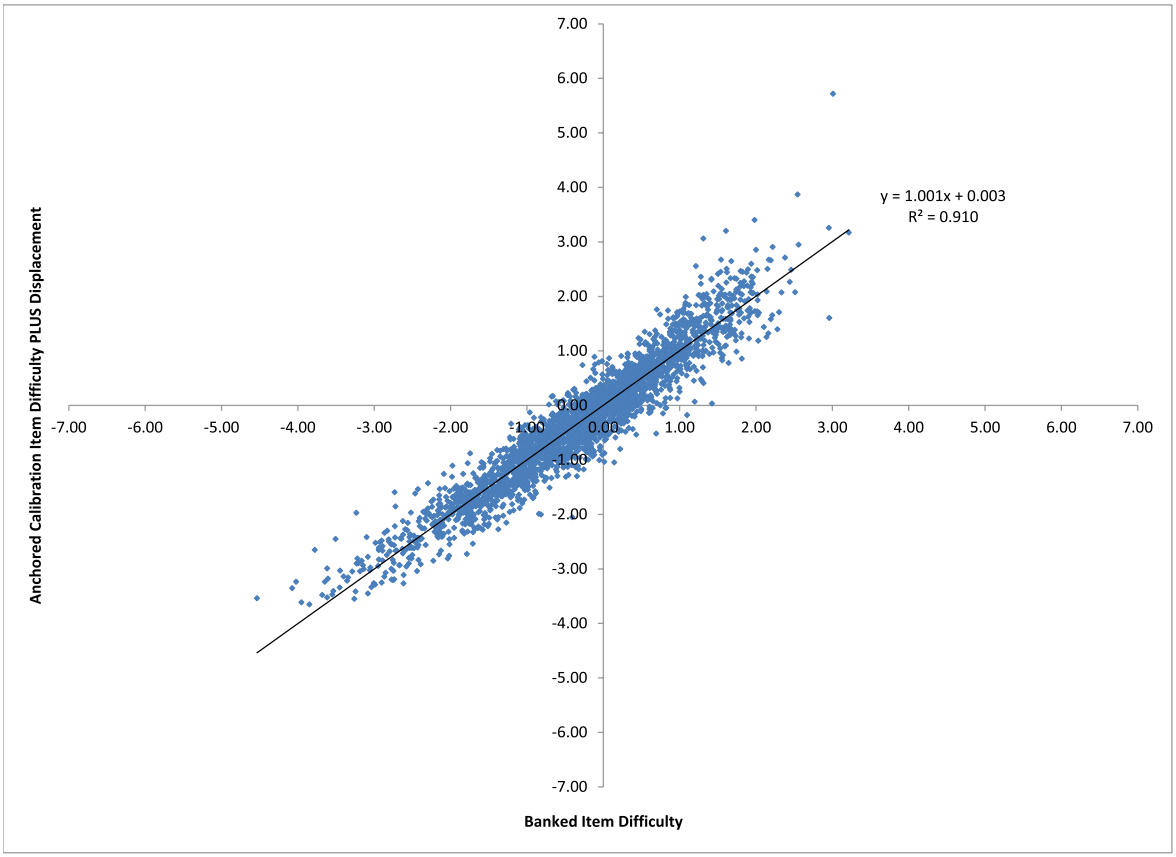
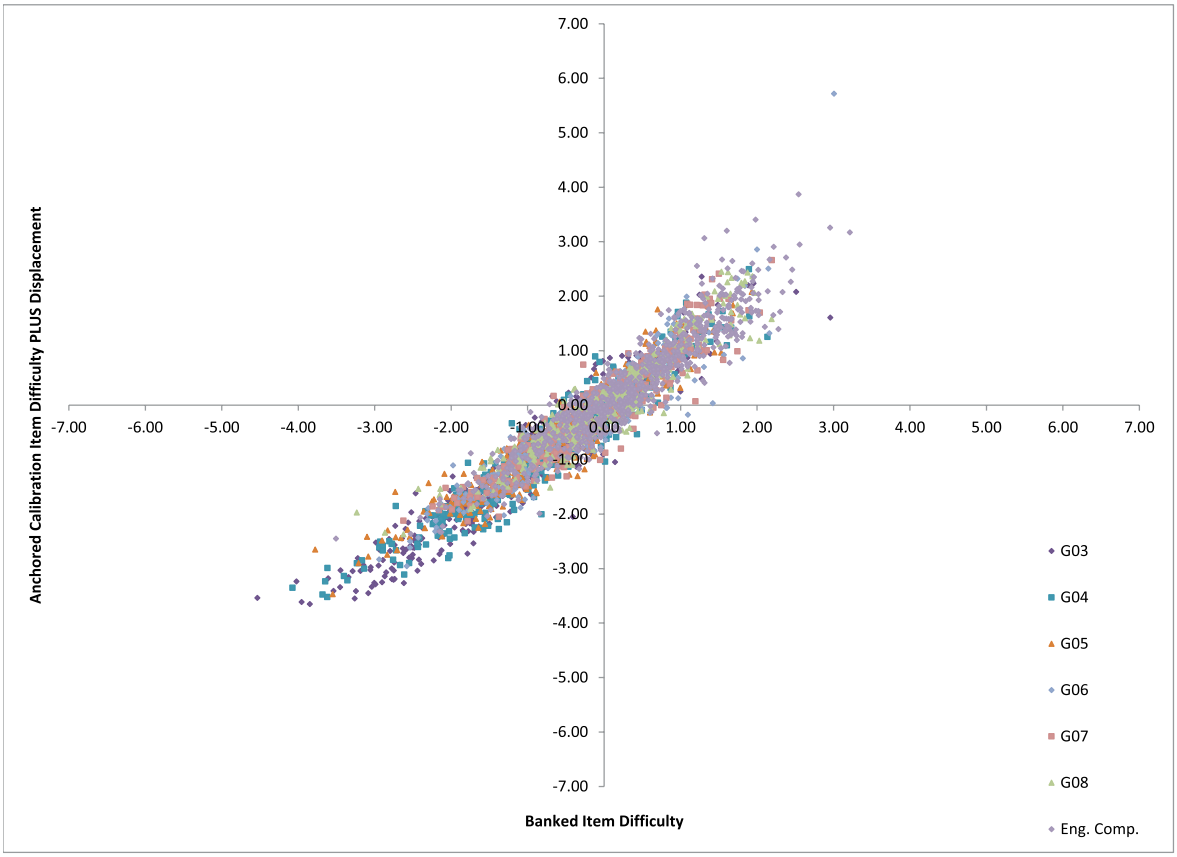


Table 18–11. Number of Writing Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	COMP	Total
Disp. ≤ -1.0	3	2	0	2	2	0	2	11
-1.0 < Disp. ≤ -0.9	1	2	2	1	1	1	1	9
-0.9 < Disp. ≤ -0.8	2	2	1	4	3	2	5	19
-0.8 < Disp. ≤ -0.7	8	9	5	8	3	2	4	39
-0.7 < Disp. ≤ -0.6	7	8	8	8	7	7	9	54
-0.6 < Disp. ≤ -0.5	10	11	13	8	3	8	21	74
-0.5 < Disp. ≤ -0.4	22	14	13	13	14	10	32	118
-0.4 < Disp. ≤ -0.3	19	26	22	16	23	25	60	191
-0.3 < Disp. ≤ -0.2	39	30	25	25	24	23	72	238
-0.2 < Disp. ≤ -0.1	38	38	26	32	43	33	65	275
-0.1 < Disp. ≤ 0.0	42	30	36	44	32	38	78	300
0.0 < Disp. ≤ 0.1	45	48	31	49	50	46	87	356
0.1 < Disp. ≤ 0.2	37	34	36	57	53	36	77	330
0.2 < Disp. ≤ 0.3	30	30	32	31	40	34	64	261
0.3 < Disp. ≤ 0.4	17	17	19	26	16	22	69	186
0.4 < Disp. ≤ 0.5	17	11	20	20	20	13	27	128
0.5 < Disp. ≤ 0.6	6	4	7	5	6	13	20	61
0.6 < Disp. ≤ 0.7	8	7	6	6	2	8	14	51
0.7 < Disp. ≤ 0.8	4	3	1	2	3	0	11	24
0.8 < Disp. ≤ 0.9	4	4	3	2	2	2	4	21
0.9 < Disp. ≤ 1.0	1	0	0	1	1	1	4	8
1.0 < Disp.	1	1	3	1	1	1	7	15
TOTAL	361	331	309	361	349	325	733	2769

Figure 18–16. Writing Banked Item Parameters vs. Anchored Calibration — All Items in Grade 3 and Above





It is evident from this series of plots that the item parameter estimates are reasonably stable for the items in grade 3 and above.

For both of the anchored calibrations described in this section, banked item parameters were compared to the banked item parameters plus the displacements by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–12 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in each of the calibrations.

Table 18–12. Summary of Robust Z across Anchored Calibrations in Writing

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
Kindergarten	17	8	47%	0	0	N/A
Grade 1	57	17	30%	0	0	N/A
Grade 2	123	20	16%	0	0	N/A
Grade 3	361	44	12%	361	49	14%
Grade 4	331	45	14%	331	47	14%
Grade 5	309	38	12%	309	41	13%
Grade 6	361	43	12%	361	43	12%
Grade 7	349	27	8%	349	28	8%
Grade 8	325	28	9%	325	30	9%
English Comp	733	78	11%	733	82	11%
Total	2966	348	12%	2769	320	12%
	Correlation = 0.957			Correlation = 0.954		
	Additive Constant = 0.003			Additive Constant = 0.002		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.573 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.546 to 0.573, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.546 have absolute value of robust Z greater than 1.645.

ANCHORED GRADE LEVEL CALIBRATIONS

While the CDT content area item banks are vertically scaled with items from Kindergarten through high school courses, the assessments themselves are first made available in grade 3. Also, while the items are selected adaptively, most students take a large number of items at grade level. Given these conditions, item parameters were also evaluated by running anchored grade level item calibrations—grade 3 items calibrated with grade 3 students, and so on. This is similar to how field-test items were calibrated. Table 18–13 shows the number of students in each grade level calibration.

Table 18–13. Number of Students in Grade Level Calibrations

Content Area	Grade/Course	Number of Students
Mathematics	Grade 3	38,900
Mathematics	Grade 4	39,887
Mathematics	Grade 5	45,969
Mathematics	Grade 6	50,545
Mathematics	Grade 7	49,220
Mathematics	Grade 8	39,861
Mathematics	Algebra I	66,769
Mathematics	Geometry	6,916
Mathematics	Algebra II	7,538
Reading	Grade 3	30,166
Reading	Grade 4	32,775
Reading	Grade 5	38,927
Reading	Grade 6	40,461
Reading	Grade 7	36,898
Reading	Grade 8	35,175
Reading	Reading/Literature	82,780
Science	Grade 3	1,734
Science	Grade 4	17,164
Science	Grade 5	6,575
Science	Grade 6	17,709
Science	Grade 7	28,300
Science	Grade 8	39,626
Science	High School	922
Science	Biology	71,059
Science	Chemistry	4,658
Writing	Grade 3	6,303
Writing	Grade 4	6,288
Writing	Grade 5	7,752
Writing	Grade 6	8,749
Writing	Grade 7	9,875
Writing	Grade 8	11,157
Writing	Writing/English Composition	6,950

MATHEMATICS

Figure 18–17 shows the displacements from the anchored grade level calibrations of operational data for the mathematics item bank. Items are color-coded by grade/course.

Figure 18–17. Mathematics Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above

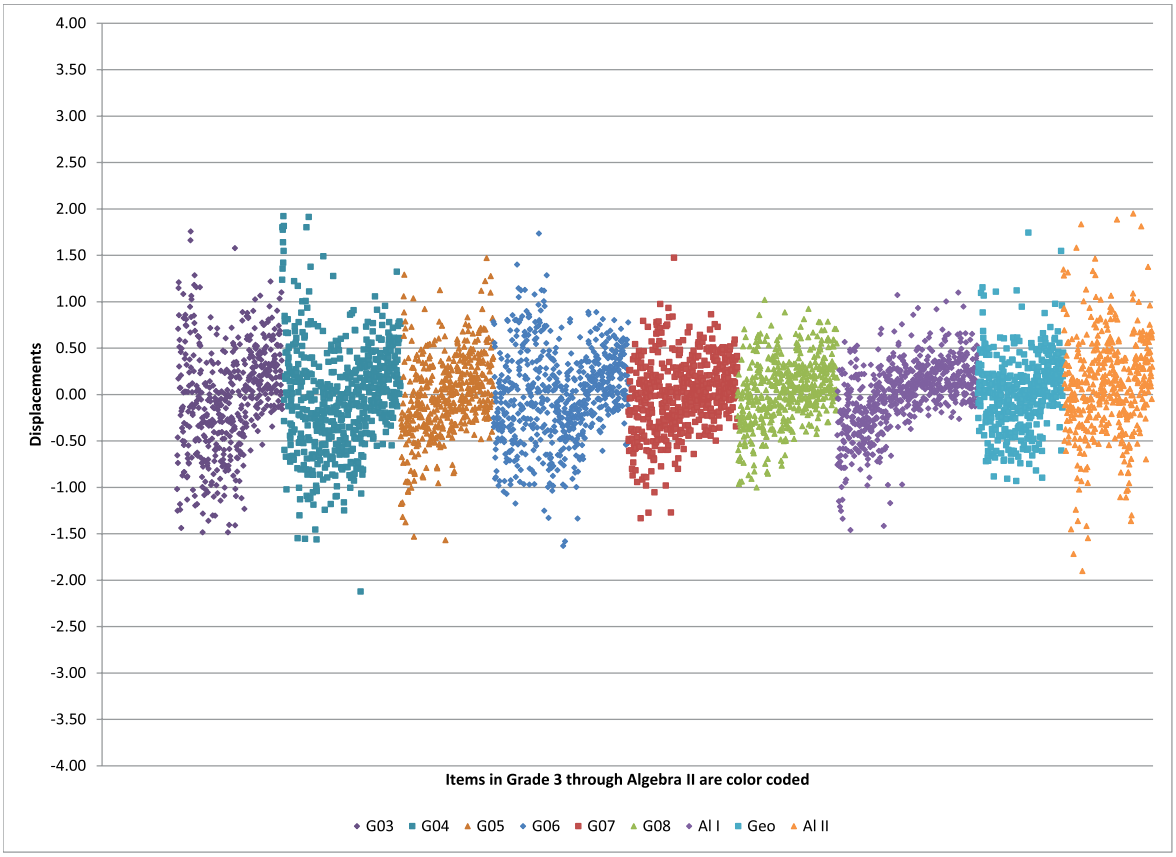


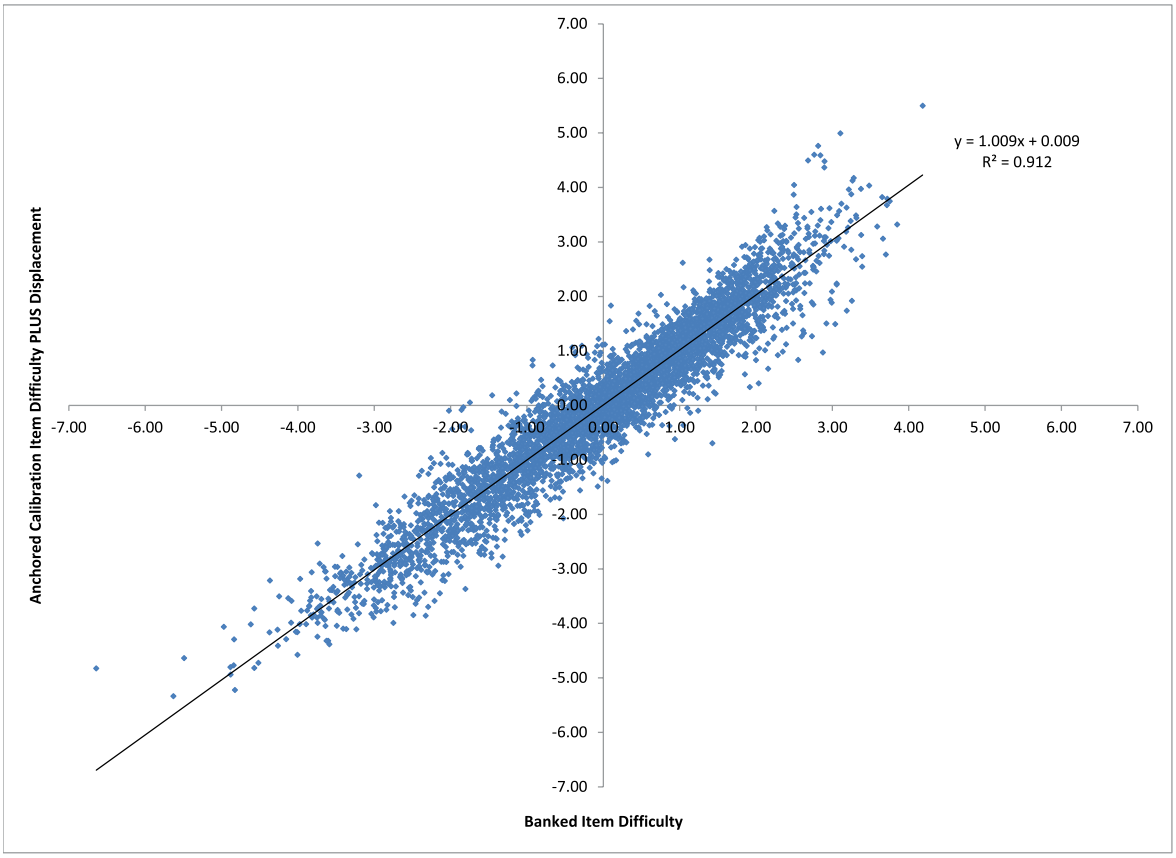
Table 18–14 summarizes the data in Figure 18–17. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Seventy-five percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–14).

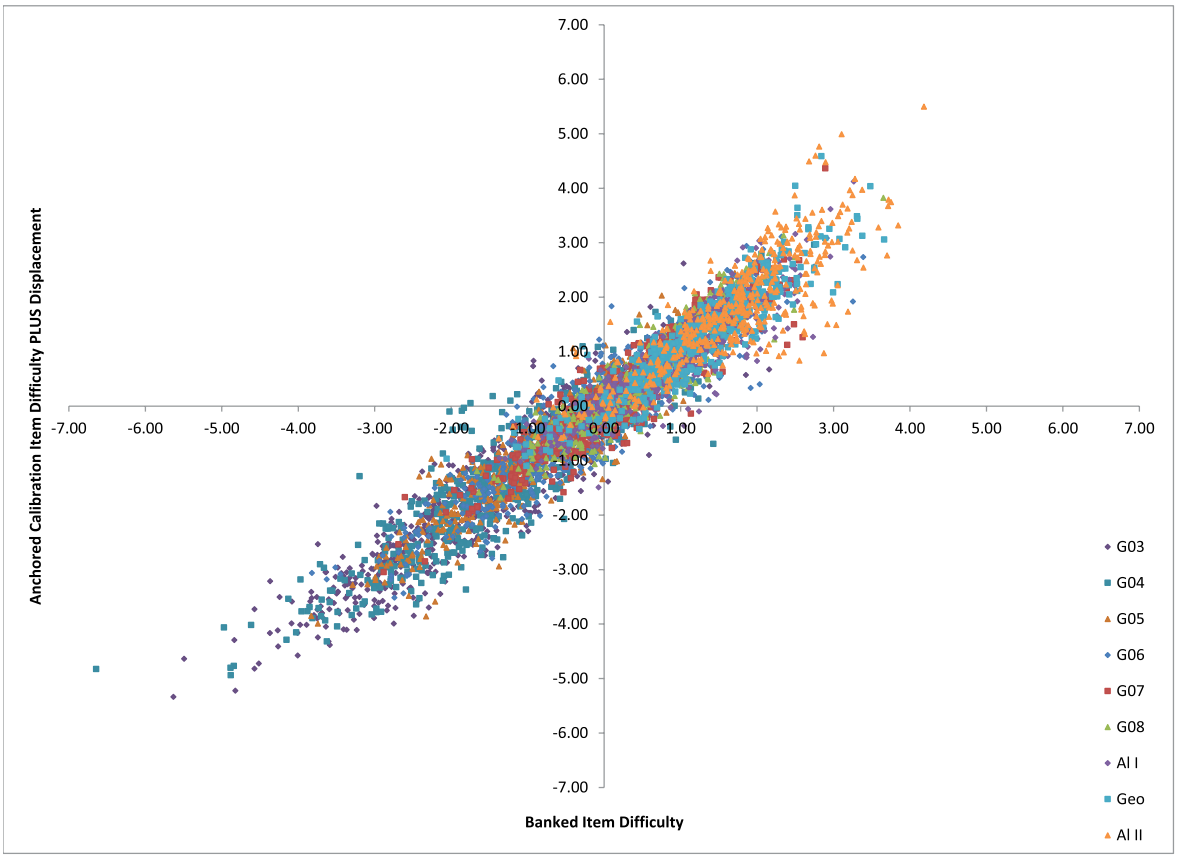
Table 18–14. Number of Mathematics Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	ALI	GEO	ALII	Total
Disp. ≤ -1.0	30	24	10	9	4	1	8	0	14	100
-1.0 < Disp. ≤ -0.9	8	4	6	17	4	4	6	2	4	55
-0.9 < Disp. ≤ -0.8	13	17	7	11	7	4	2	3	4	68
-0.8 < Disp. ≤ -0.7	21	20	7	14	5	8	12	7	5	99
-0.7 < Disp. ≤ -0.6	20	31	5	26	10	5	18	16	7	138
-0.6 < Disp. ≤ -0.5	20	23	14	23	23	11	21	13	13	161
-0.5 < Disp. ≤ -0.4	22	30	23	28	26	13	19	21	20	202
-0.4 < Disp. ≤ -0.3	26	28	35	48	31	40	37	23	21	289
-0.3 < Disp. ≤ -0.2	31	39	39	45	38	30	38	26	28	314
-0.2 < Disp. ≤ -0.1	36	44	40	58	48	45	60	40	26	397
-0.1 < Disp. ≤ 0.0	28	42	48	53	69	58	74	42	43	457
0.0 < Disp. ≤ 0.1	43	49	42	70	65	68	84	52	31	504
0.1 < Disp. ≤ 0.2	32	42	42	51	53	47	100	52	45	464
0.2 < Disp. ≤ 0.3	39	47	29	51	44	52	90	35	33	420
0.3 < Disp. ≤ 0.4	33	33	38	40	31	33	54	26	37	325
0.4 < Disp. ≤ 0.5	27	26	24	28	35	20	31	27	22	240
0.5 < Disp. ≤ 0.6	22	22	18	29	22	25	13	17	23	191
0.6 < Disp. ≤ 0.7	17	19	9	21	9	15	14	10	22	136
0.7 < Disp. ≤ 0.8	15	17	11	20	11	8	2	1	8	93
0.8 < Disp. ≤ 0.9	14	7	3	17	3	3	1	2	9	59
0.9 < Disp. ≤ 1.0	4	4	2	3	2	1	3	3	7	29
1.0 < Disp.	17	21	9	11	1	1	3	7	19	89
TOTAL	518	589	461	673	541	492	690	425	441	4830

Figure 18–18 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored grade level calibrations of all items using the operational data set. Again, a line of best fit is included in the upper plot.

Figure 18–18. Mathematics Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above





For the anchored grade level calibrations described above, banked item parameters were compared to the newly calibrated values by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–15 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in the calibrations.

Table 18–15. Summary of Robust Z across Anchored Grade Level Calibrations in Mathematics

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645
Kindergarten	0	0	N/A
Grade 1	0	0	N/A
Grade 2	0	0	N/A
Grade 3	518	136	26%
Grade 4	589	136	23%
Grade 5	461	57	12%
Grade 6	673	122	18%
Grade 7	541	41	8%
Grade 8	492	33	7%
Algebra I	690	47	7%
Geometry	425	30	7%
Algebra II	441	74	17%
Total	4830	676	14%
	Correlation = 0.955		
	Additive Constant = 0.009		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, all items with absolute displacement greater than 0.699 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.642 to 0.699, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.642 have absolute value of robust Z greater than 1.645.

READING/LITERATURE

Figure 18–19 shows the displacements from the anchored grade level calibrations of operational data for the reading item bank. Items are color-coded by grade/course.

Figure 18–19. Reading Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above

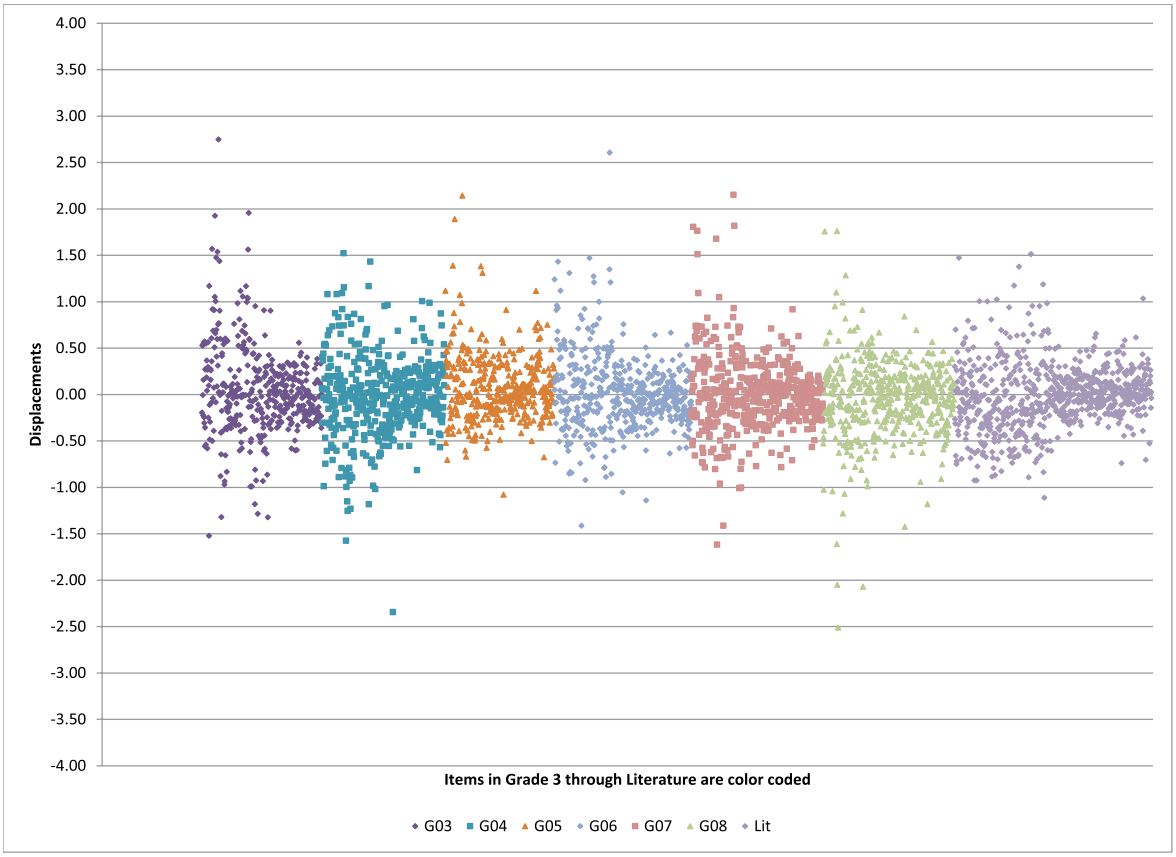


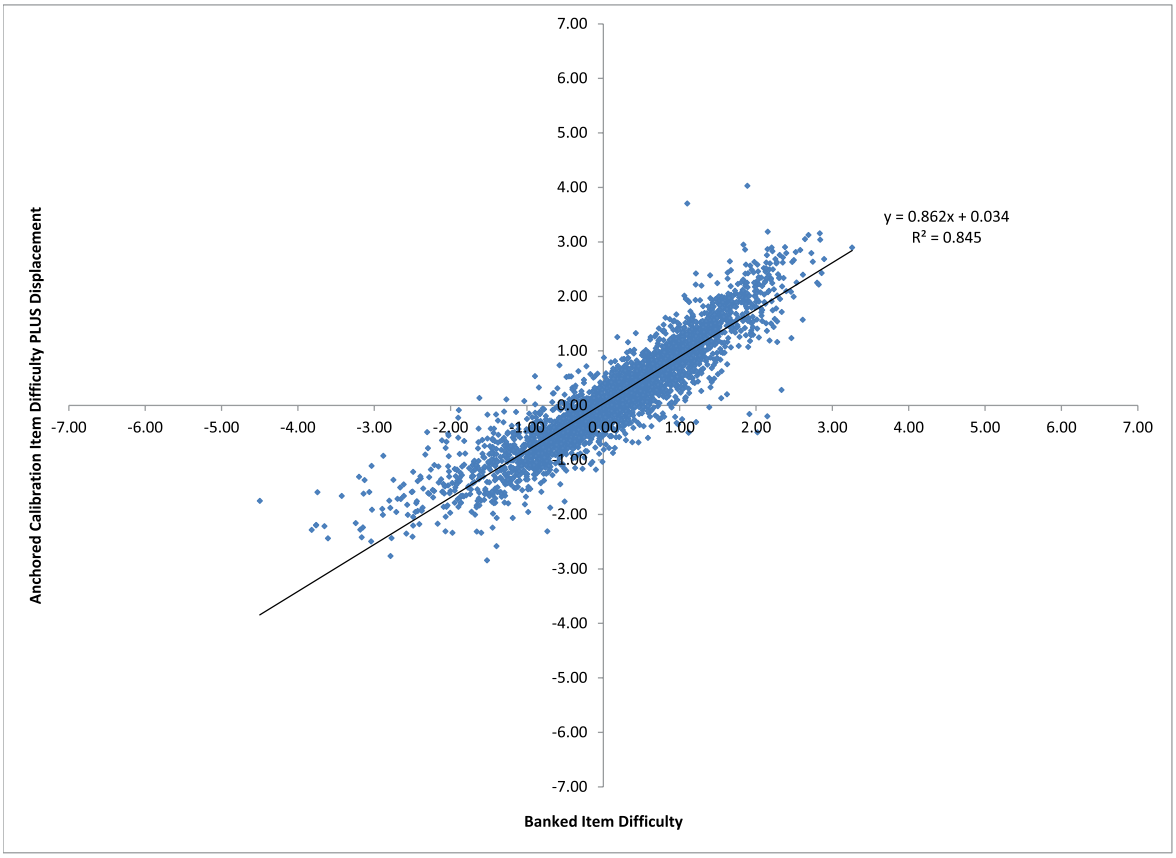
Table 18–16 summarizes the data in Figure 18–19. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-four percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–16).

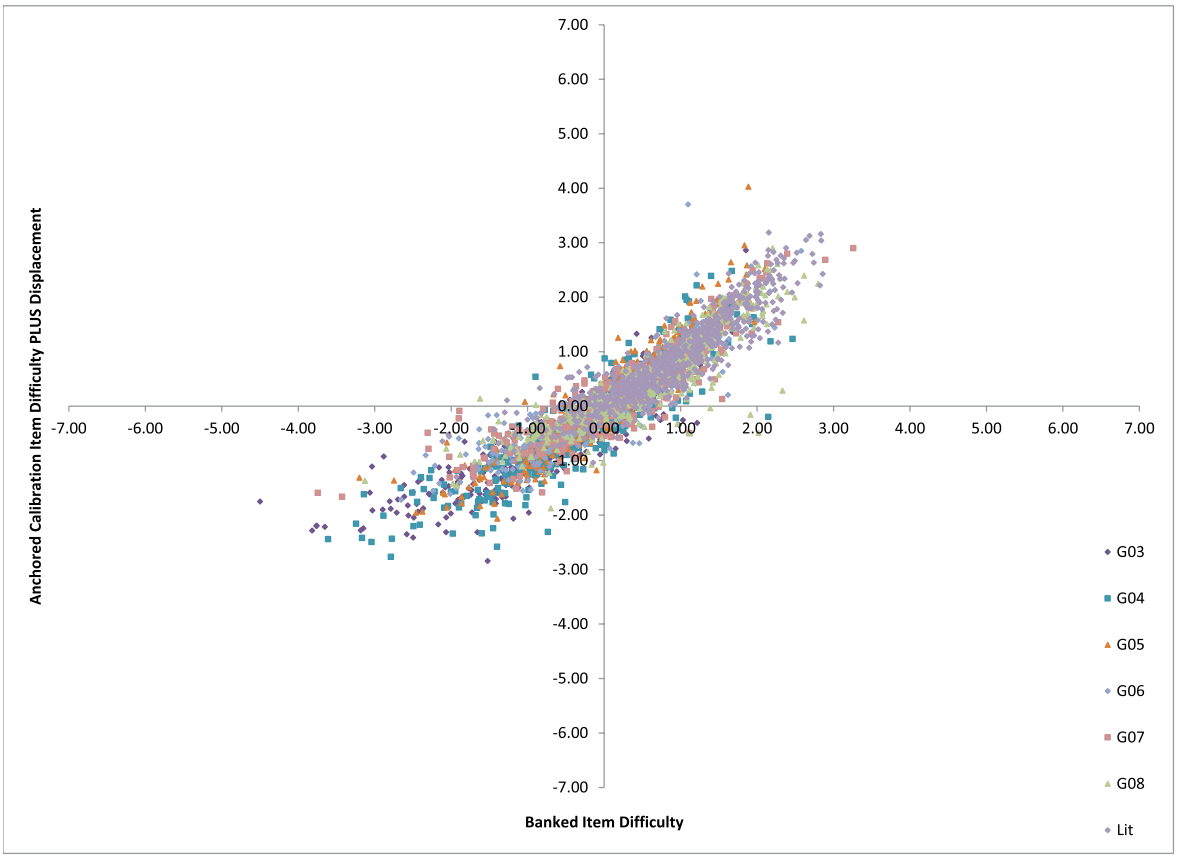
Table 18–16. Number of Reading Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	LIT	Total
Disp. ≤ -1.0	5	7	1	3	4	10	1	31
-1.0 < Disp. ≤ -0.9	6	4	0	1	1	5	2	19
-0.9 < Disp. ≤ -0.8	4	6	0	5	2	1	5	23
-0.8 < Disp. ≤ -0.7	0	7	1	4	7	3	8	30
-0.7 < Disp. ≤ -0.6	7	10	3	6	12	11	26	75
-0.6 < Disp. ≤ -0.5	10	10	3	11	10	18	16	78
-0.5 < Disp. ≤ -0.4	10	20	12	18	10	24	27	121
-0.4 < Disp. ≤ -0.3	32	27	19	34	36	21	37	206
-0.3 < Disp. ≤ -0.2	42	35	38	47	49	46	57	314
-0.2 < Disp. ≤ -0.1	50	46	38	59	54	51	83	381
-0.1 < Disp. ≤ 0.0	48	53	66	77	54	52	99	449
0.0 < Disp. ≤ 0.1	47	49	49	65	82	66	105	463
0.1 < Disp. ≤ 0.2	39	47	37	48	44	59	88	362
0.2 < Disp. ≤ 0.3	28	39	30	32	22	31	56	238
0.3 < Disp. ≤ 0.4	22	20	32	21	33	30	32	190
0.4 < Disp. ≤ 0.5	16	14	22	19	11	18	25	125
0.5 < Disp. ≤ 0.6	19	12	8	17	11	8	10	85
0.6 < Disp. ≤ 0.7	10	7	8	5	10	6	15	61
0.7 < Disp. ≤ 0.8	4	7	7	6	8	1	4	37
0.8 < Disp. ≤ 0.9	1	7	1	3	2	2	2	18
0.9 < Disp. ≤ 1.0	9	4	2	5	2	3	4	29
1.0 < Disp.	16	8	8	10	8	4	9	63
TOTAL	425	439	385	496	472	470	711	3398

Figure 18–20 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored grade level calibrations of all items using the operational data set. Again, a line of best fit is included in the upper plot.

Figure 18–20. Reading Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above





An examination of the items with larger differences between banked values and operational estimates revealed that a number of these have low *n*-counts in the operational calibration. To investigate whether this had an impact on the stability of the item parameter estimates, anchored grade level calibrations of all items in grade 3 and above with larger *n*-counts were run. Figure 18–21 shows the displacements from these calibrations. Items are color-coded by grade/course.

Figure 18–21. Reading Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above with *N*>100

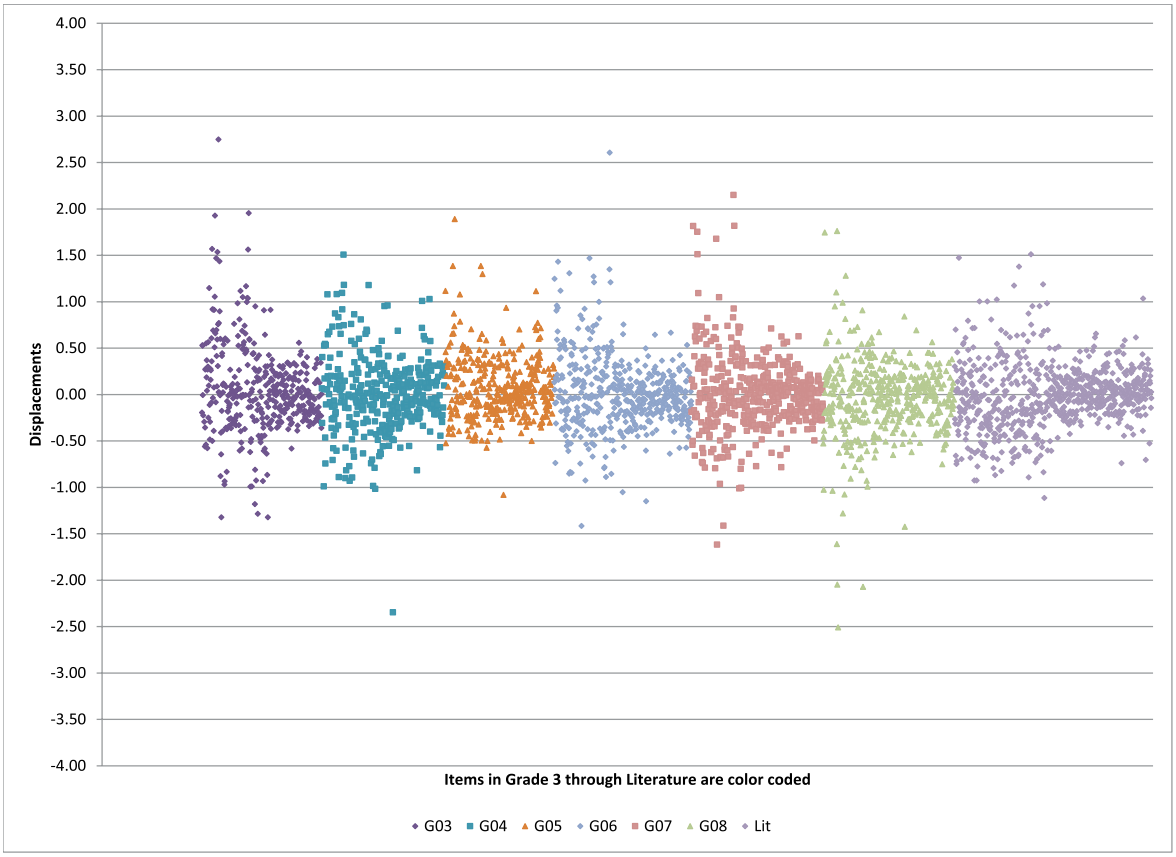


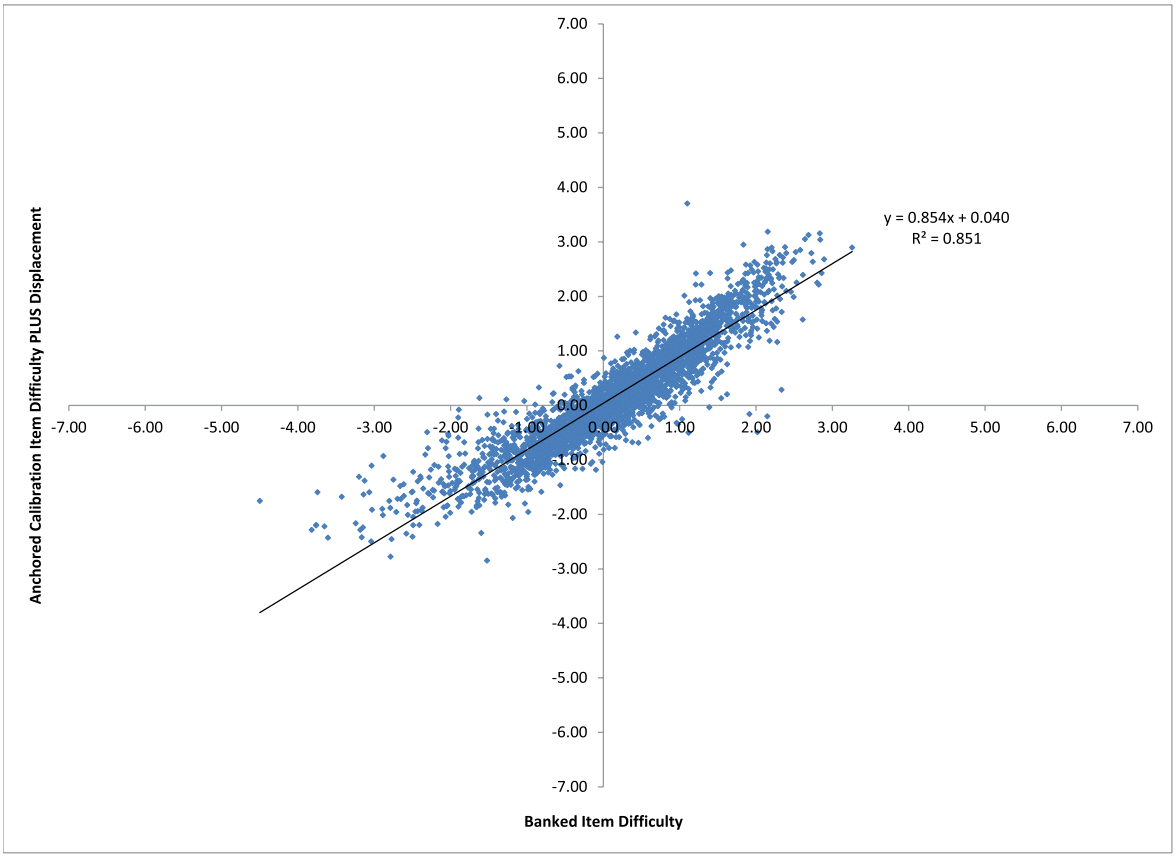
Table 18–17 summarizes the data in Figure 18–21. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-five percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–17).

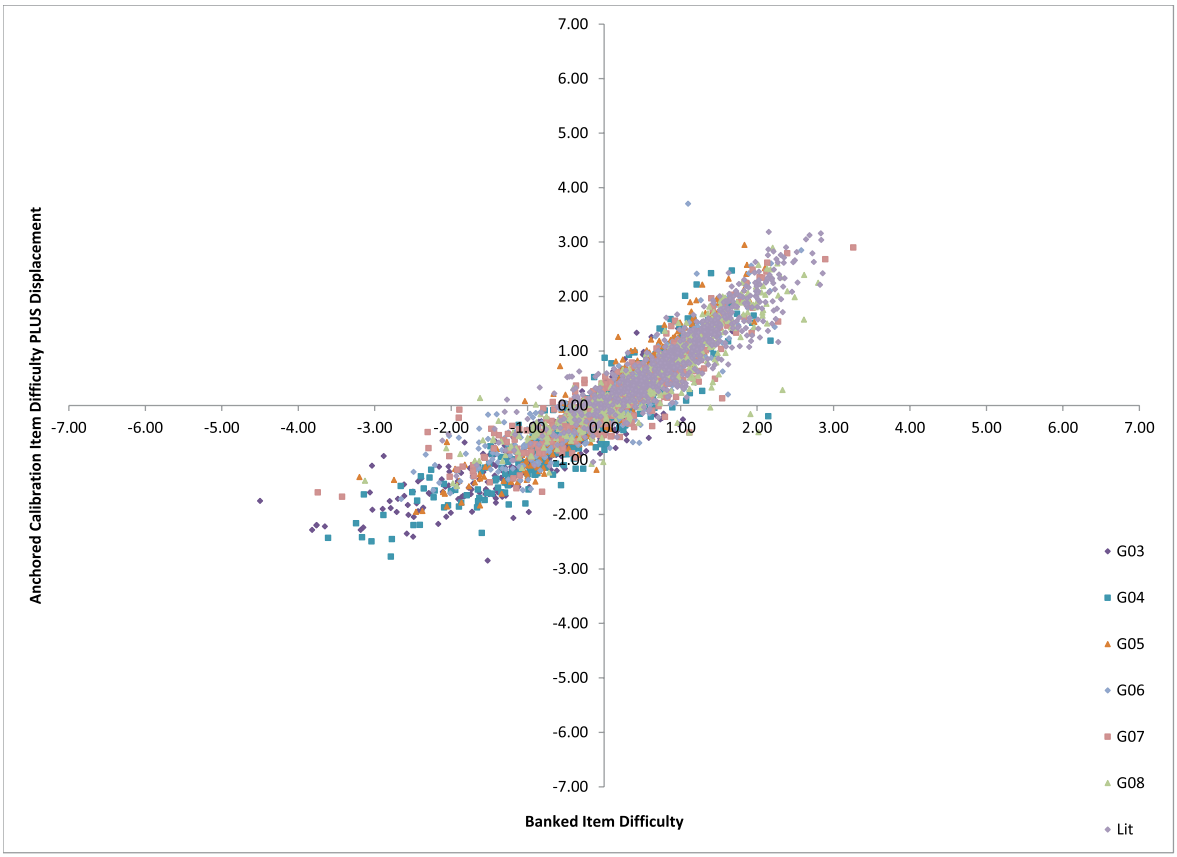
Table 18–17. Number of Reading Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	LIT	Total
Disp. ≤ -1.0	4	2	1	3	4	9	1	24
-1.0 < Disp. ≤ -0.9	6	4	0	1	1	3	2	17
-0.9 < Disp. ≤ -0.8	4	4	0	5	1	1	4	19
-0.8 < Disp. ≤ -0.7	0	6	0	4	7	3	8	28
-0.7 < Disp. ≤ -0.6	5	7	0	6	12	10	24	64
-0.6 < Disp. ≤ -0.5	9	7	3	10	9	18	14	70
-0.5 < Disp. ≤ -0.4	9	16	12	17	9	20	26	109
-0.4 < Disp. ≤ -0.3	30	23	16	32	35	20	37	193
-0.3 < Disp. ≤ -0.2	43	32	37	46	47	44	57	306
-0.2 < Disp. ≤ -0.1	49	42	38	58	54	49	82	372
-0.1 < Disp. ≤ 0.0	47	48	61	76	54	50	98	434
0.0 < Disp. ≤ 0.1	48	44	48	65	80	62	104	451
0.1 < Disp. ≤ 0.2	36	40	32	48	43	56	86	341
0.2 < Disp. ≤ 0.3	27	36	31	32	20	30	56	232
0.3 < Disp. ≤ 0.4	22	17	32	20	34	26	31	182
0.4 < Disp. ≤ 0.5	16	10	17	18	10	15	24	110
0.5 < Disp. ≤ 0.6	17	9	9	16	11	8	10	80
0.6 < Disp. ≤ 0.7	9	8	7	5	10	6	14	59
0.7 < Disp. ≤ 0.8	4	5	7	6	8	1	4	35
0.8 < Disp. ≤ 0.9	2	4	1	3	2	2	2	16
0.9 < Disp. ≤ 1.0	8	3	1	5	1	3	3	24
1.0 < Disp.	15	8	7	10	8	4	9	61
TOTAL	410	375	360	486	460	440	696	3227

Figure 18–22 mirrors Figure 18–20, except the calibrations exclude items with fewer than 100 administrations. Again, a line of best fit is included in the upper plot.

Figure 18–22. Reading Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above with N>100





For the two sets of anchored grade level calibrations described above, banked item parameters were compared to the newly calibrated values by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–18 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in the calibrations.

Table 18–18. Summary of Robust Z across Two Sets of Anchored Grade Level Calibrations in Reading

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
Kindergarten	0	0	N/A	0	0	N/A
Grade 1	0	0	N/A	0	0	N/A
Grade 2	0	0	N/A	0	0	N/A
Grade 3	425	92	22%	410	86	21%
Grade 4	439	88	20%	375	68	18%
Grade 5	385	42	11%	360	38	11%
Grade 6	496	78	16%	486	79	16%
Grade 7	472	78	17%	460	75	16%
Grade 8	470	71	15%	440	68	15%
Literature	711	103	14%	696	98	14%
Total	3398	552	16%	3227	512	16%
	Correlation = 0.919			Correlation = 0.922		
	Additive Constant = 0.008			Additive Constant = 0.012		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.504 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.495 to 0.504, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.495 have absolute value of robust Z greater than 1.645.

SCIENCE

Figure 18–23 shows the displacements from the anchored grade level calibrations of operational data for the science item bank. Items are color-coded by grade/course.

Figure 18–23. Science Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above

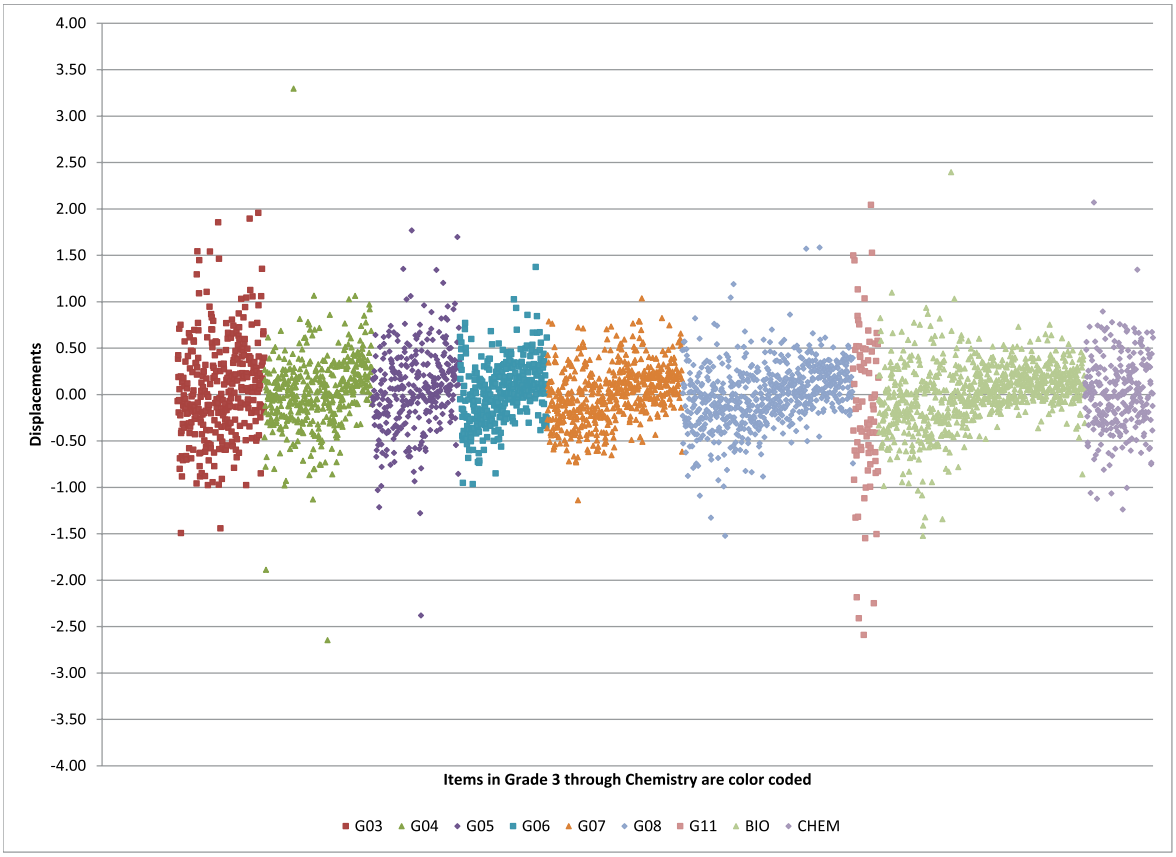


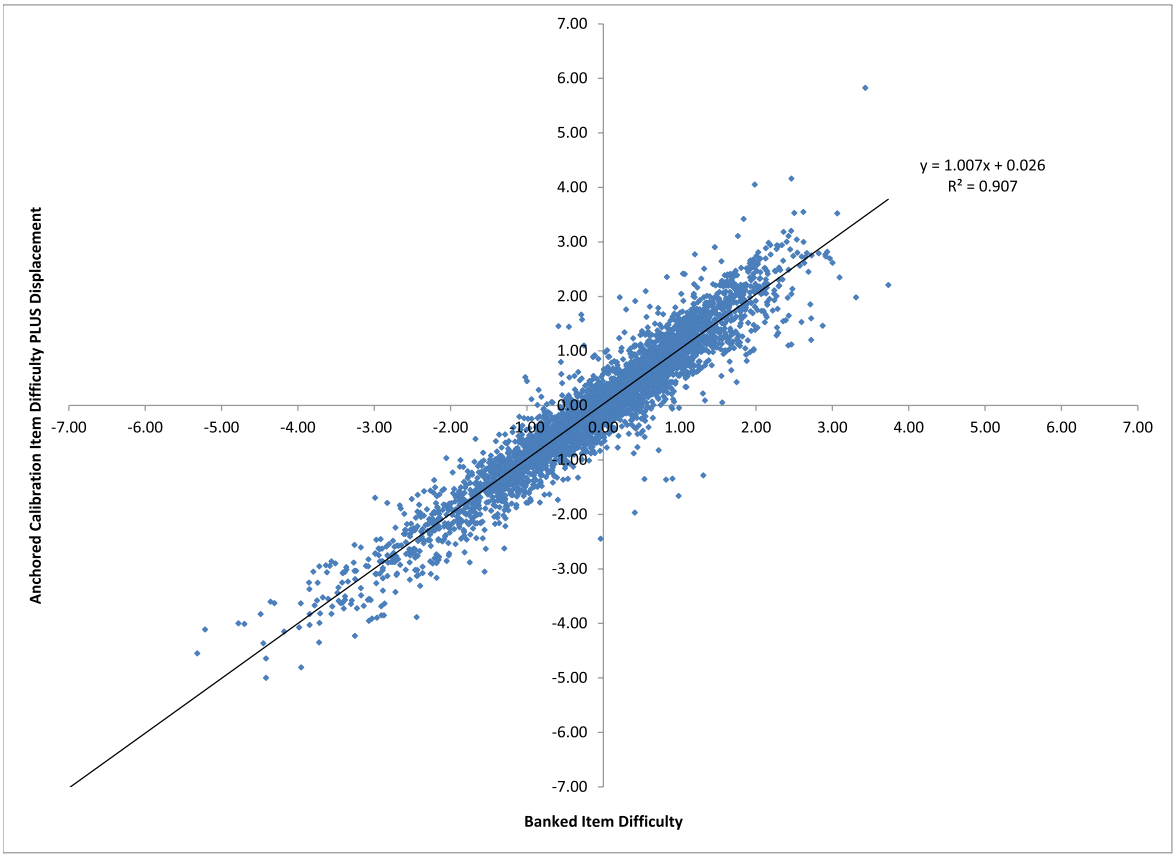
Table 18–19 summarizes the data in Figure 18–23. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-four percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–19).

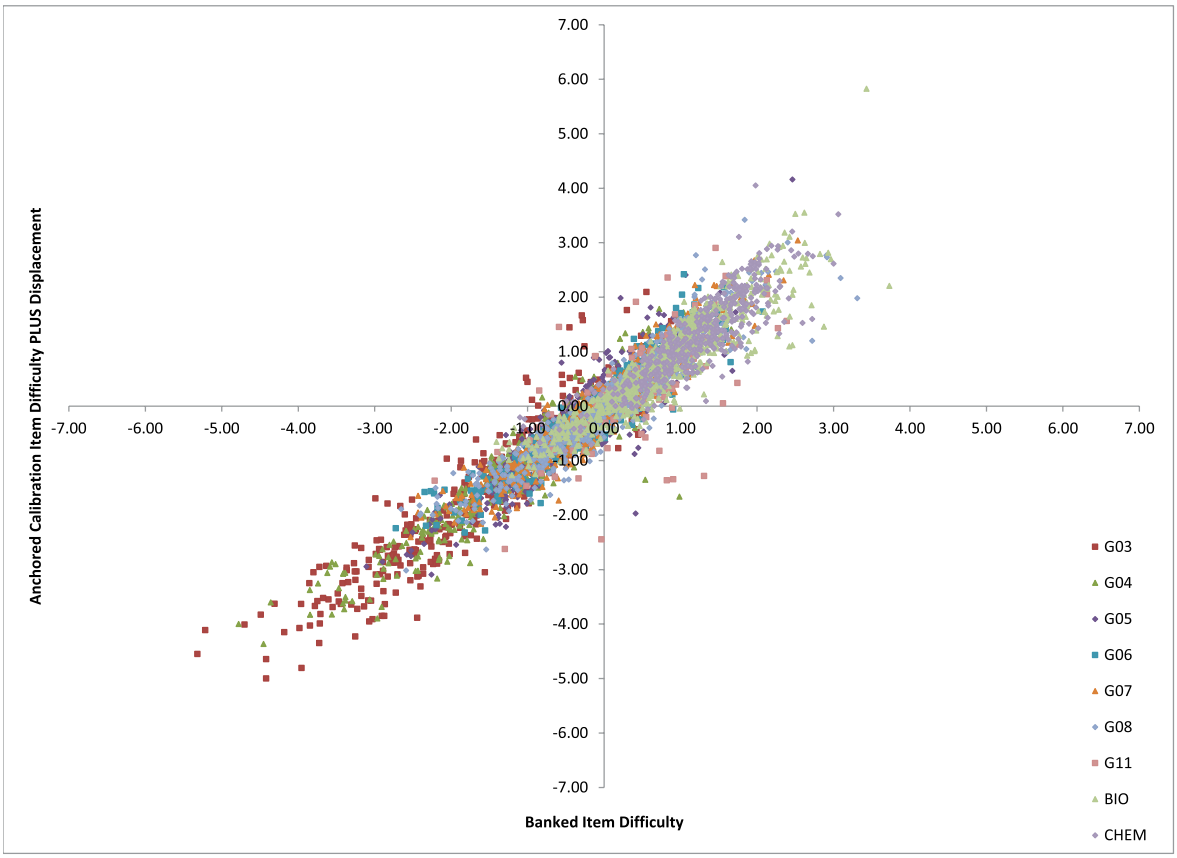
Table 18–19. Number of Science Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	G11	BIO	CHEM	Total
Disp. ≤ -1.0	2	3	4	0	1	3	10	6	5	34
-1.0 < Disp. ≤ -0.9	6	2	2	2	0	2	2	5	1	22
-0.9 < Disp. ≤ -0.8	6	5	2	1	0	5	3	4	1	27
-0.8 < Disp. ≤ -0.7	5	6	4	3	3	6	2	2	3	34
-0.7 < Disp. ≤ -0.6	13	4	10	2	9	7	4	9	8	66
-0.6 < Disp. ≤ -0.5	16	8	11	7	13	16	5	24	12	112
-0.5 < Disp. ≤ -0.4	16	15	10	16	31	20	4	31	13	156
-0.4 < Disp. ≤ -0.3	21	25	18	24	20	46	7	32	26	219
-0.3 < Disp. ≤ -0.2	16	32	22	34	46	51	2	65	28	296
-0.2 < Disp. ≤ -0.1	42	34	25	24	58	85	4	52	49	373
-0.1 < Disp. ≤ 0.0	23	54	42	39	58	89	5	131	41	482
0.0 < Disp. ≤ 0.1	25	52	40	47	80	91	0	116	42	493
0.1 < Disp. ≤ 0.2	26	56	31	51	80	90	2	107	47	490
0.2 < Disp. ≤ 0.3	25	37	33	29	54	57	6	98	43	382
0.3 < Disp. ≤ 0.4	23	39	27	25	33	49	3	64	29	292
0.4 < Disp. ≤ 0.5	16	17	17	16	17	23	6	26	23	161
0.5 < Disp. ≤ 0.6	14	10	13	8	17	19	6	14	19	120
0.6 < Disp. ≤ 0.7	14	11	14	7	5	4	2	14	18	89
0.7 < Disp. ≤ 0.8	10	6	5	2	10	3	2	5	11	54
0.8 < Disp. ≤ 0.9	5	3	4	2	1	2	1	3	2	23
0.9 < Disp. ≤ 1.0	3	2	4	1	0	0	0	1	0	11
1.0 < Disp.	16	4	7	2	1	4	6	3	2	45
TOTAL	343	425	345	342	537	672	82	812	423	3981

Figure 18–24 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored grade level calibrations of all items using the operational data set. Again, a line of best fit is included in the upper plot.

Figure 18–24. Science Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above





An examination of the items with larger differences between banked values and operational estimates revealed that a number of these have low *n*-counts in the operational calibration. To investigate whether this had an impact on the stability of the item parameter estimates, anchored grade level calibrations of all items in grade 3 and above with larger *n*-counts were run. Figure 18–25 shows the displacements from these calibrations. Items are color-coded by grade/course.

Figure 18–25. Science Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above with N>100

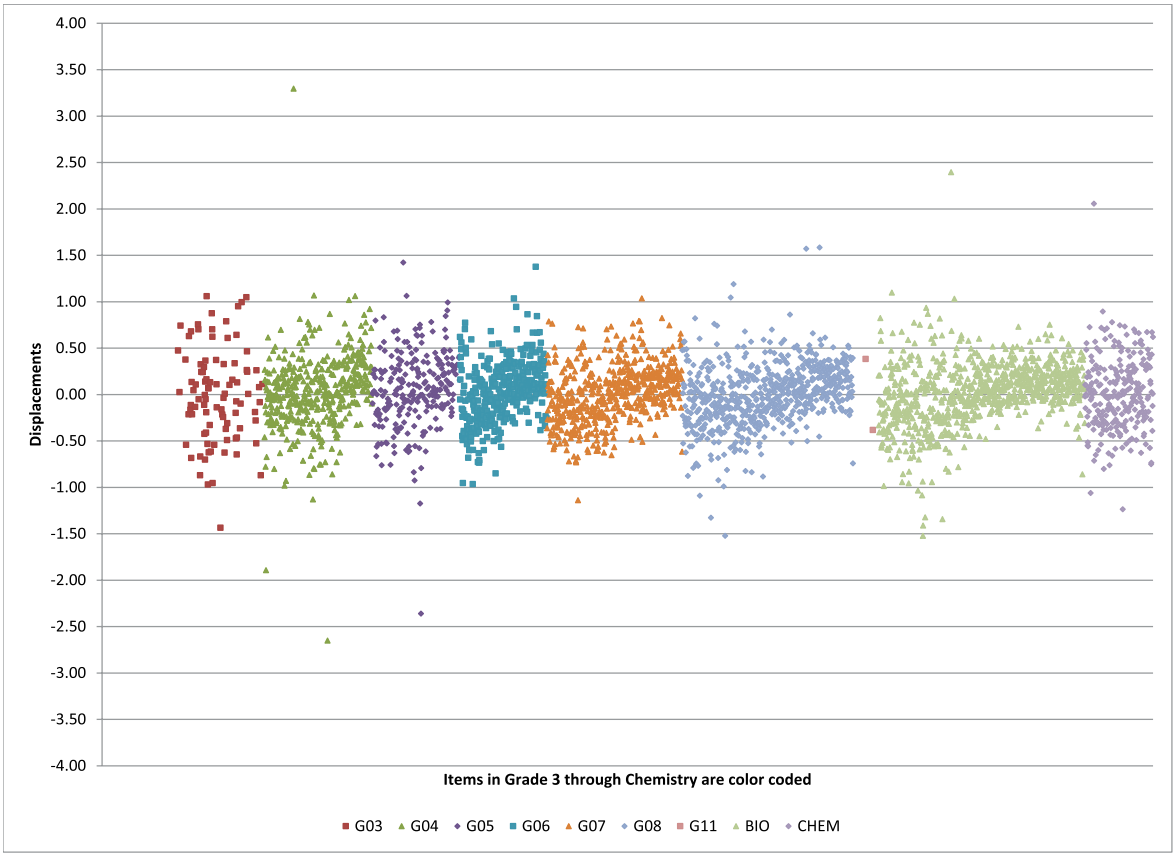


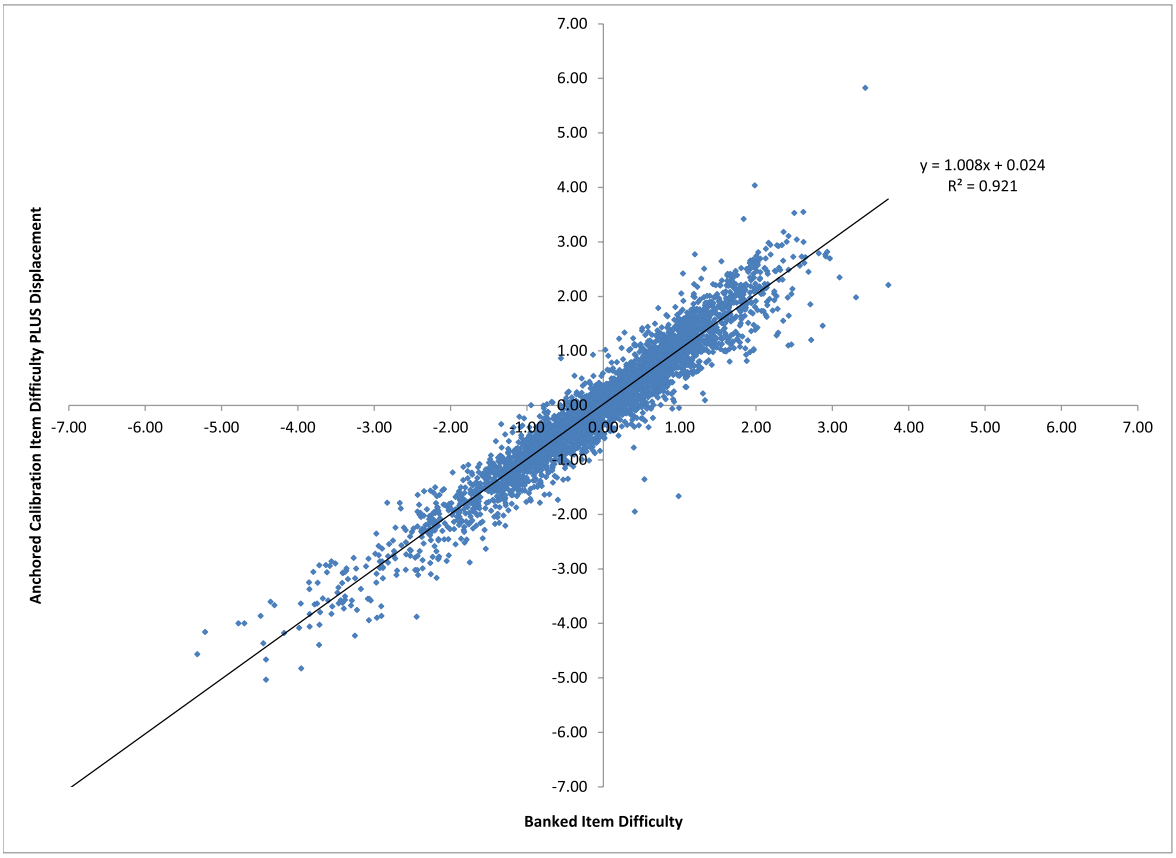
Table 18–20 summarizes the data in Figure 18–25. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-six percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–20).

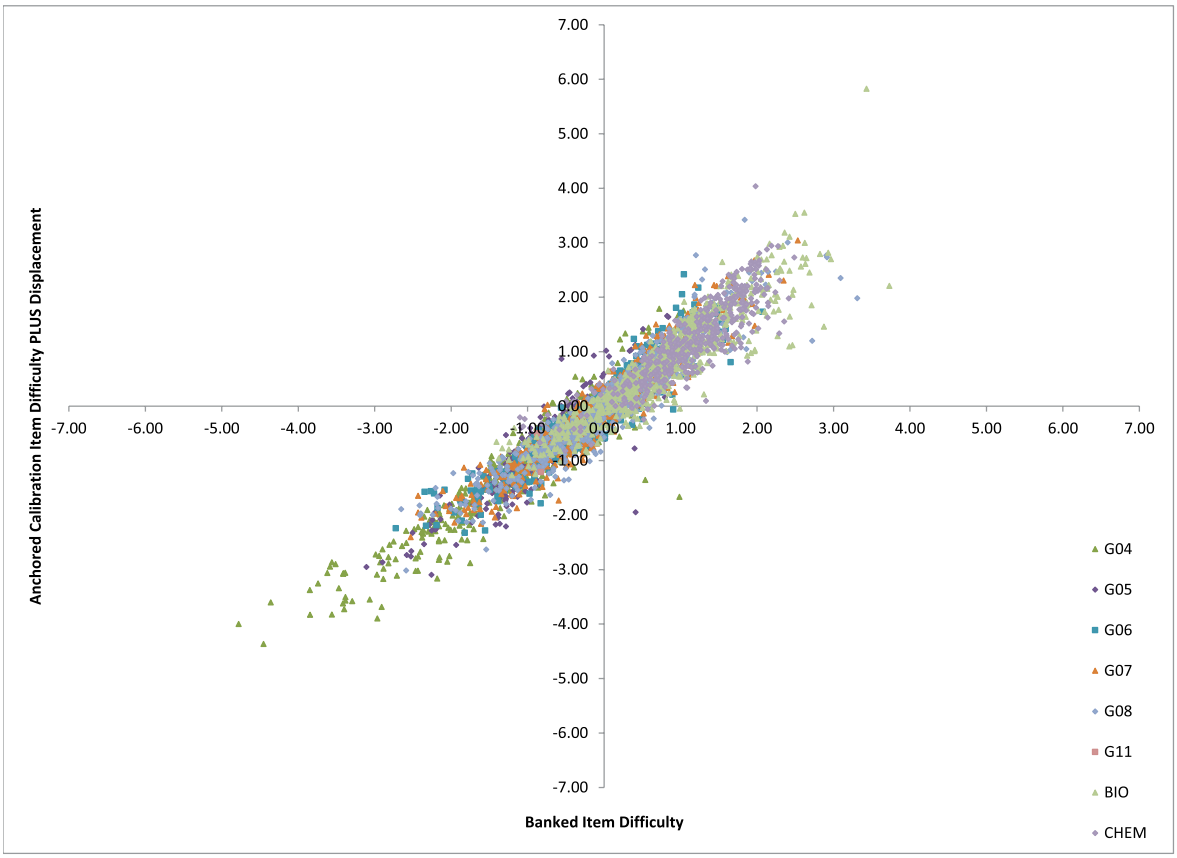
Table 18–20. Number of Science Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	G11	BIO	CHEM	Total
Disp. ≤ -1.0	1	3	2	0	1	3	0	6	2	18
-1.0 < Disp. ≤ -0.9	2	2	1	2	0	2	0	5	1	15
-0.9 < Disp. ≤ -0.8	2	4	1	1	0	5	0	4	1	18
-0.8 < Disp. ≤ -0.7	1	7	3	3	3	6	0	2	4	29
-0.7 < Disp. ≤ -0.6	6	4	6	3	9	7	0	9	7	51
-0.6 < Disp. ≤ -0.5	4	8	10	7	13	16	0	24	11	93
-0.5 < Disp. ≤ -0.4	5	14	6	16	31	20	0	31	13	136
-0.4 < Disp. ≤ -0.3	4	24	13	21	20	46	1	31	23	183
-0.3 < Disp. ≤ -0.2	5	32	10	34	46	51	0	66	27	271
-0.2 < Disp. ≤ -0.1	9	33	24	23	58	85	0	52	47	331
-0.1 < Disp. ≤ 0.0	7	53	30	38	58	89	0	131	40	446
0.0 < Disp. ≤ 0.1	7	52	28	48	80	91	0	116	41	463
0.1 < Disp. ≤ 0.2	10	56	22	51	79	90	0	107	45	460
0.2 < Disp. ≤ 0.3	6	36	28	28	55	57	0	98	42	350
0.3 < Disp. ≤ 0.4	6	38	21	25	33	48	1	63	29	264
0.4 < Disp. ≤ 0.5	2	17	13	17	17	23	0	26	20	135
0.5 < Disp. ≤ 0.6	0	10	7	8	16	19	0	14	21	95
0.6 < Disp. ≤ 0.7	5	10	7	5	5	4	0	14	15	65
0.7 < Disp. ≤ 0.8	5	7	6	3	10	3	0	5	10	49
0.8 < Disp. ≤ 0.9	1	3	2	2	1	2	0	3	2	16
0.9 < Disp. ≤ 1.0	2	1	2	1	0	0	0	1	0	7
1.0 < Disp.	2	4	2	2	1	4	0	3	1	19
TOTAL	92	418	244	338	536	671	2	811	402	3514

Figure 18–26 mirrors Figure 18–24, except the calibrations exclude items with fewer than 100 administrations. Again, a line of best fit is included in the upper plot.

Figure 18–26. Science Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above with N>100





For the two sets of anchored grade level calibrations described above, banked item parameters were compared to the newly calibrated values by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–21 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in the calibrations.

Table 18–21. Summary of Robust Z across Two Sets of Anchored Grade Level Calibrations in Science

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
K–2 span	0	0	N/A	0	0	N/A
Grade 3	343	102	30%	92	32	35%
Grade 4	425	55	13%	418	61	15%
Grade 5	345	71	21%	244	47	19%
Grade 6	342	30	9%	338	40	12%
Grade 7	537	46	9%	536	57	11%
Grade 8	672	59	9%	671	65	10%
Grade 11	82	39	48%	2	0	0%
Biology	812	82	10%	811	89	11%
Chemistry	423	69	16%	402	69	17%
Total	3981	553	14%	3514	460	13%
	Correlation = 0.953			Correlation = 0.960		
	Additive Constant = 0.026			Additive Constant = 0.024		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.564 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.500 to 0.564, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.500 have absolute value of robust Z greater than 1.645.

WRITING/ENGLISH COMPOSITION

Figure 18–27 shows the displacements from the anchored grade level calibrations of operational data for the writing item bank. Items are color-coded by grade/course.

Figure 18–27. Writing Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above

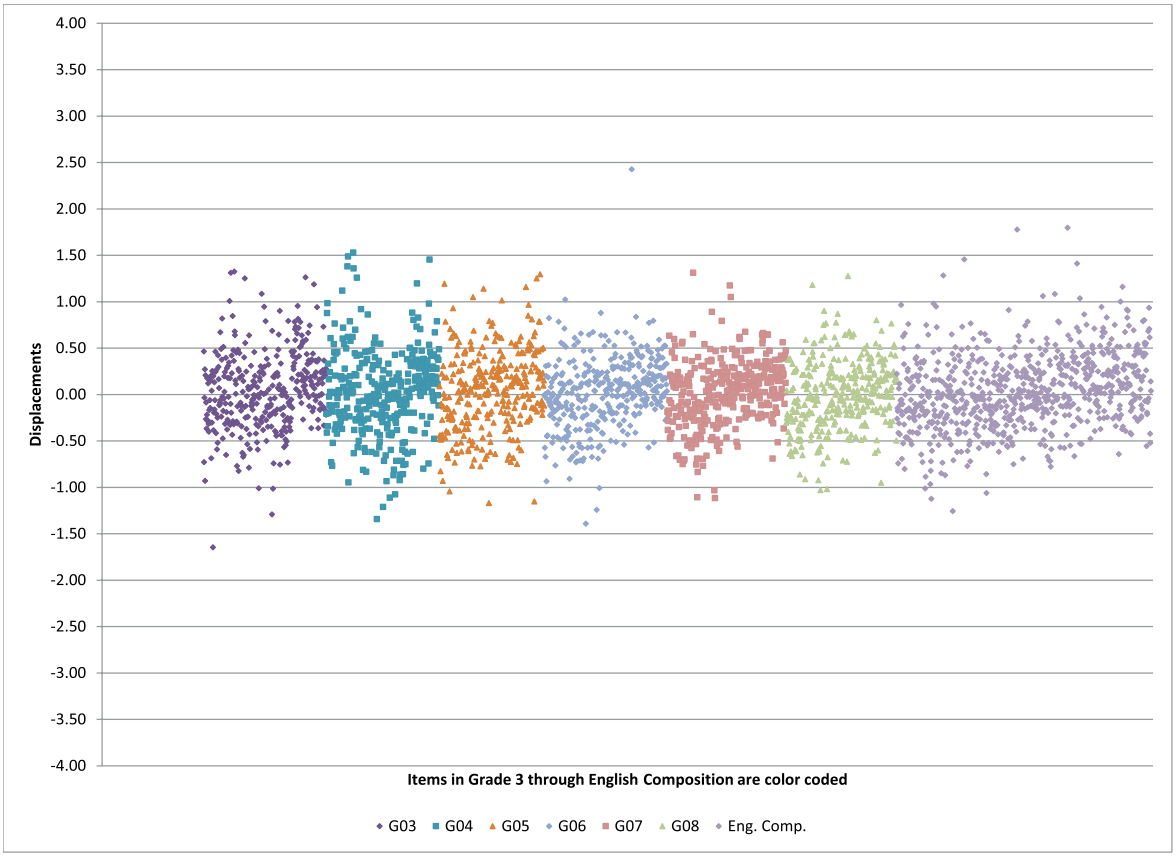


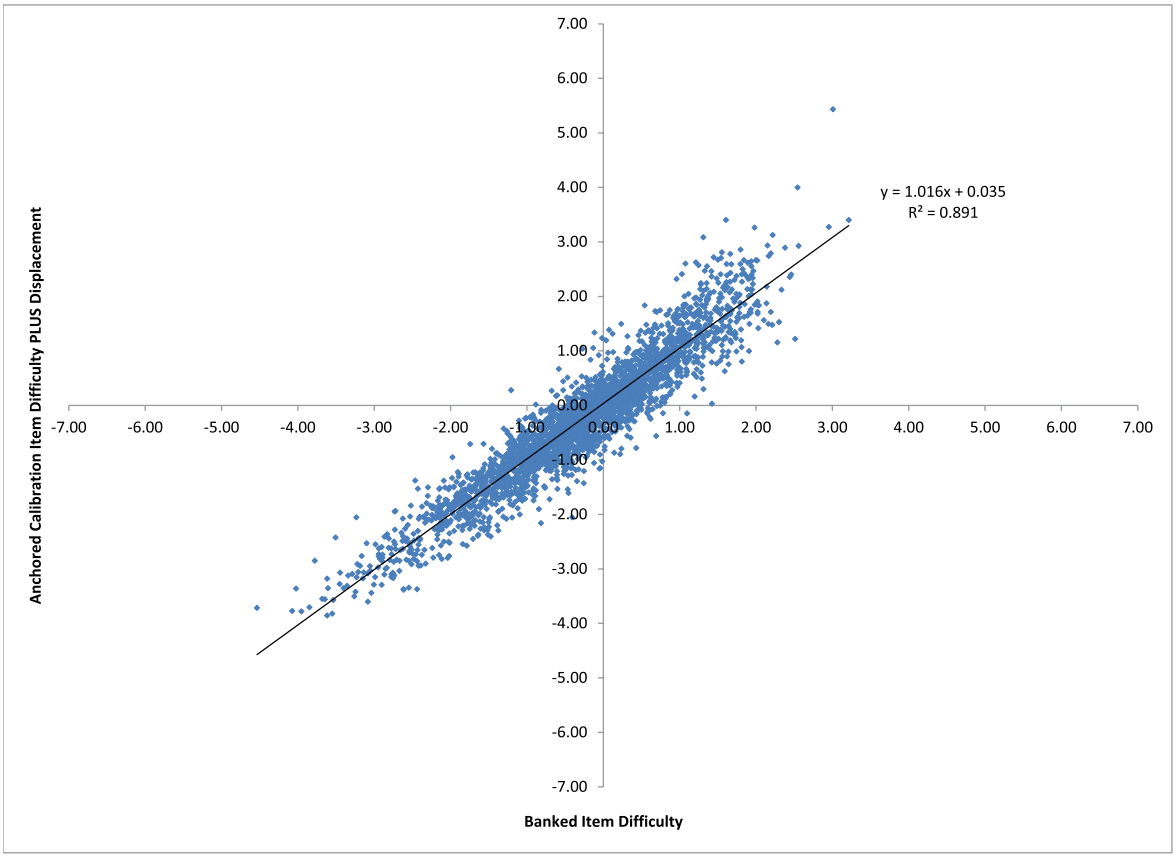
Table 18–22 summarizes the data in Figure 18–27. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-one percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–22).

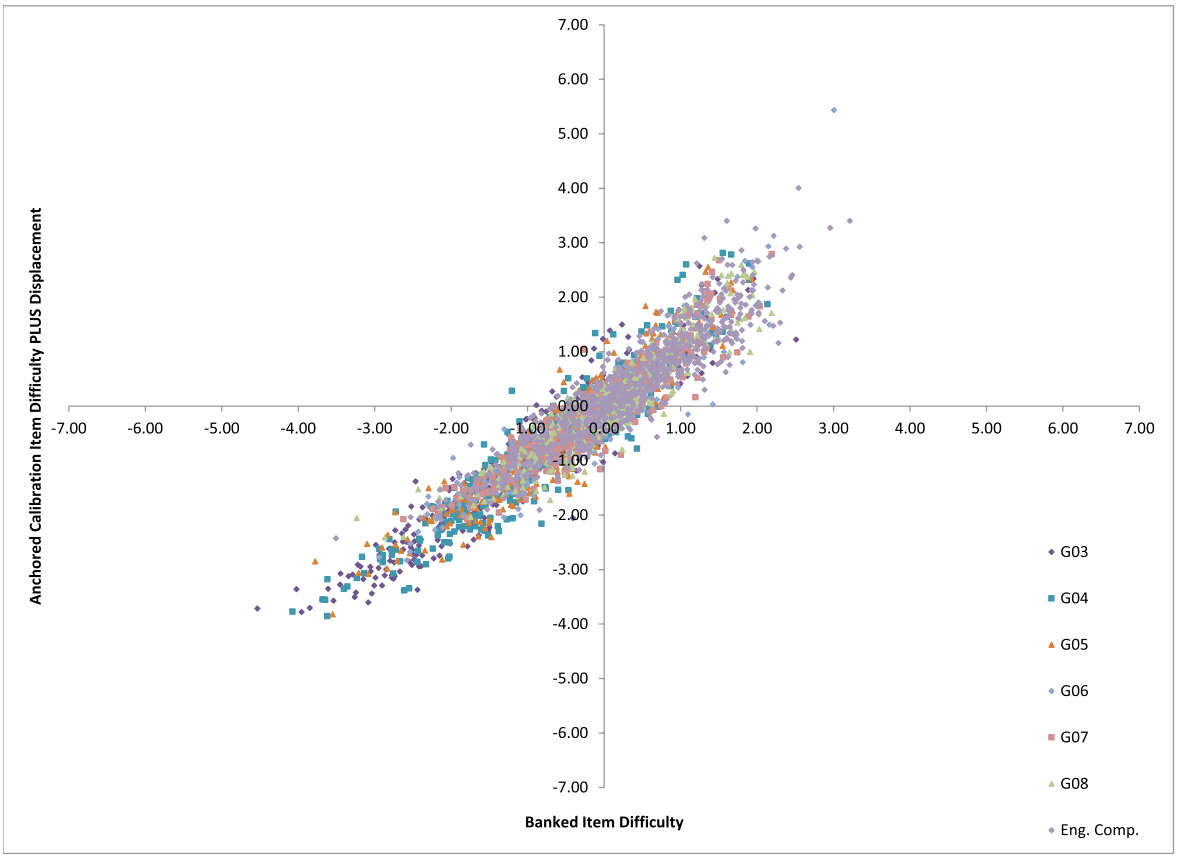
Table 18–22. Number of Writing Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	COMP	Total
Disp. ≤ -1.0	4	4	3	3	3	2	4	23
-1.0 < Disp. ≤ -0.9	1	3	1	2	0	3	1	11
-0.9 < Disp. ≤ -0.8	1	6	1	0	1	1	6	16
-0.8 < Disp. ≤ -0.7	6	9	6	6	6	3	11	47
-0.7 < Disp. ≤ -0.6	5	8	12	7	7	10	11	60
-0.6 < Disp. ≤ -0.5	9	10	11	15	9	5	29	88
-0.5 < Disp. ≤ -0.4	12	11	14	15	14	21	34	121
-0.4 < Disp. ≤ -0.3	28	19	15	15	15	19	42	153
-0.3 < Disp. ≤ -0.2	30	26	24	24	33	30	57	224
-0.2 < Disp. ≤ -0.1	39	25	24	36	28	34	85	271
-0.1 < Disp. ≤ 0.0	37	32	22	37	40	35	79	282
0.0 < Disp. ≤ 0.1	29	37	34	40	36	34	68	278
0.1 < Disp. ≤ 0.2	37	39	35	55	58	33	69	326
0.2 < Disp. ≤ 0.3	36	19	27	29	38	32	64	245
0.3 < Disp. ≤ 0.4	25	26	15	30	22	24	47	189
0.4 < Disp. ≤ 0.5	20	19	22	15	17	10	43	146
0.5 < Disp. ≤ 0.6	9	9	13	14	9	11	27	92
0.6 < Disp. ≤ 0.7	9	7	12	9	8	9	17	71
0.7 < Disp. ≤ 0.8	9	6	7	4	1	4	17	48
0.8 < Disp. ≤ 0.9	3	5	2	3	1	2	4	20
0.9 < Disp. ≤ 1.0	4	3	2	0	0	1	8	18
1.0 < Disp.	7	8	7	2	3	2	10	39
TOTAL	360	331	309	361	349	325	733	2768

Figure 18–28 shows banked item difficulties plotted against the item difficulties plus displacement from the anchored grade level calibrations of all items using the operational data set. Again, a line of best fit is included in the upper plot.

Figure 18–28. Writing Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above





An examination of the items with larger differences between banked values and operational estimates revealed that a number of these have low *n*-counts in the operational calibration. To investigate whether this had an impact on the stability of the item parameter estimates, anchored grade level calibrations of all items in grade 3 and above with larger *n*-counts were run. Figure 18–29 shows the displacements from these calibrations. Items are color-coded by grade/course.

Figure 18–29. Writing Anchored Grade Level Calibrations Displacements — All Items in Grade 3 and Above with *N*>100

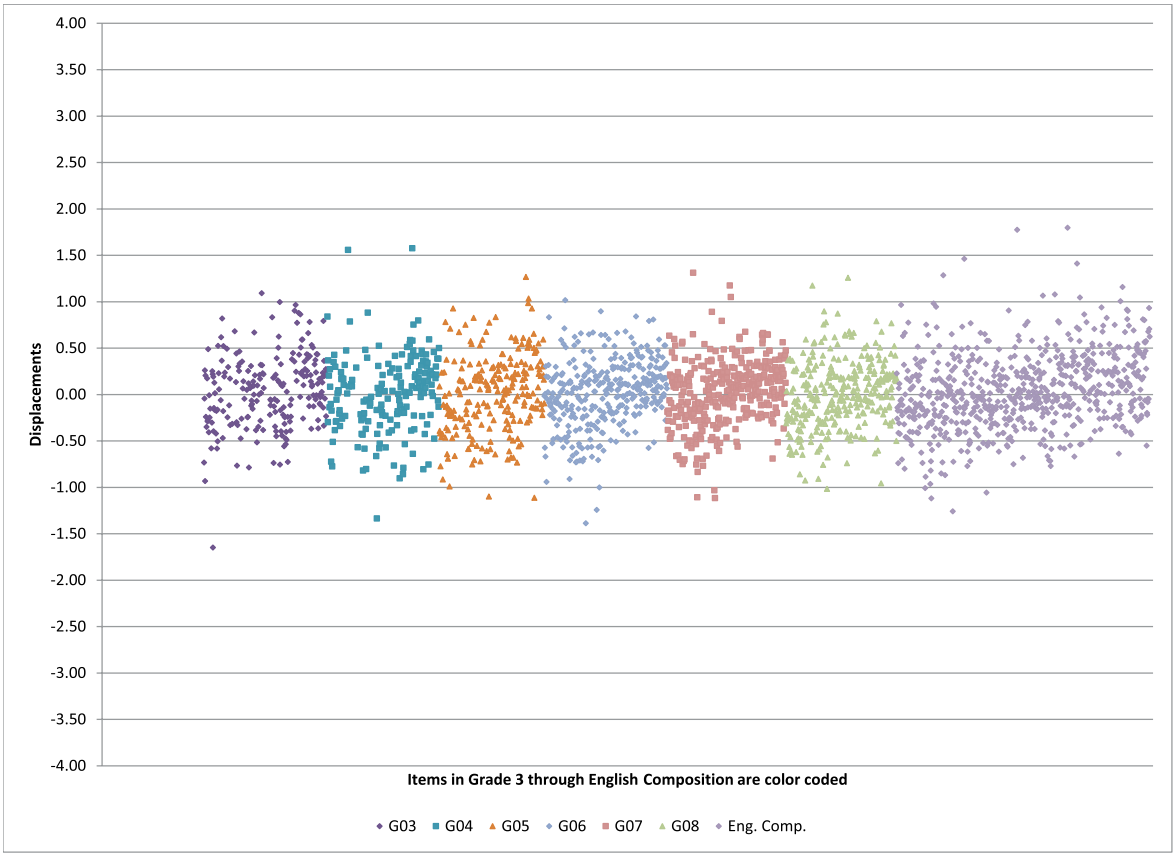


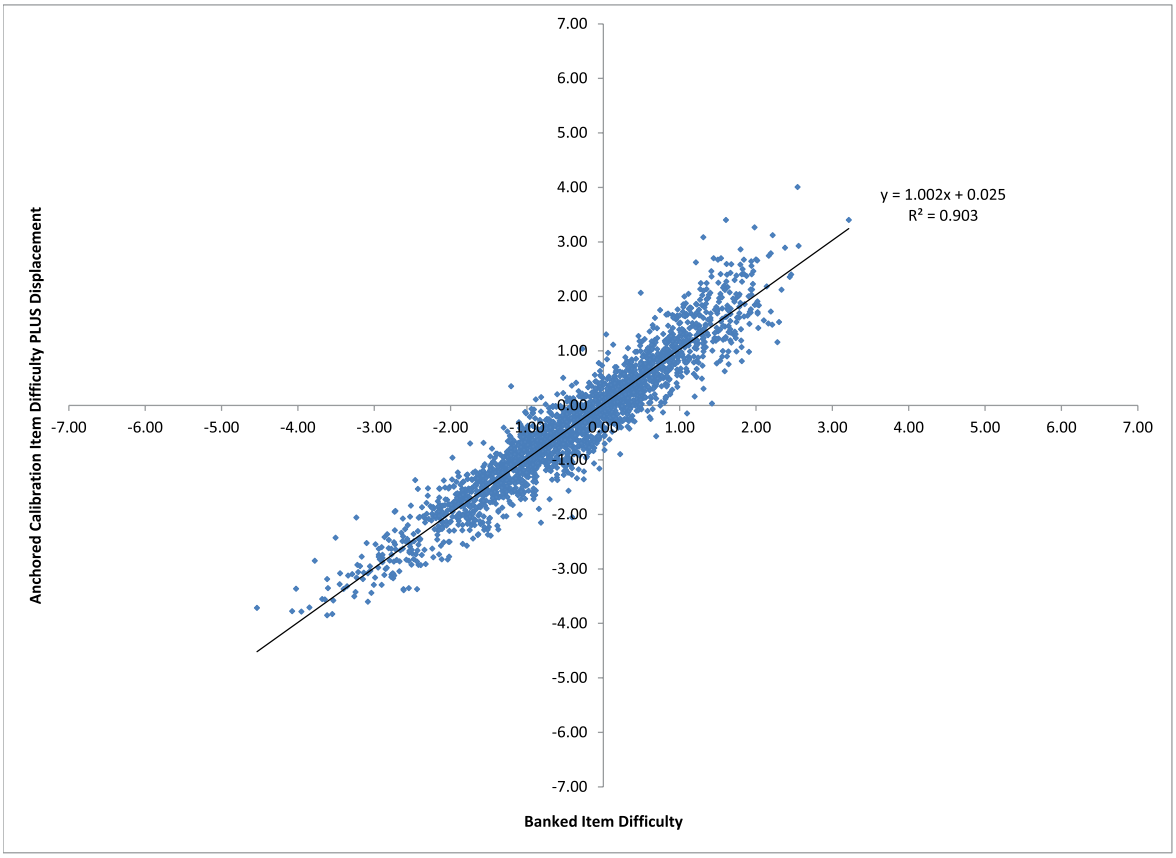
Table 18–23 summarizes the data in Figure 18–29. It contains item counts by grade/course and displacements in intervals of 0.1 logits. According to the WINSTEPS manual, in an anchored calibration, half of the displacements are expected to be negative and half positive. Displacements less than 0.5 in magnitude are considered small (unlikely to have much impact). Eighty-two percent of the items in the bank have a displacement less than 0.5 in magnitude (gray shaded in Table 18–23).

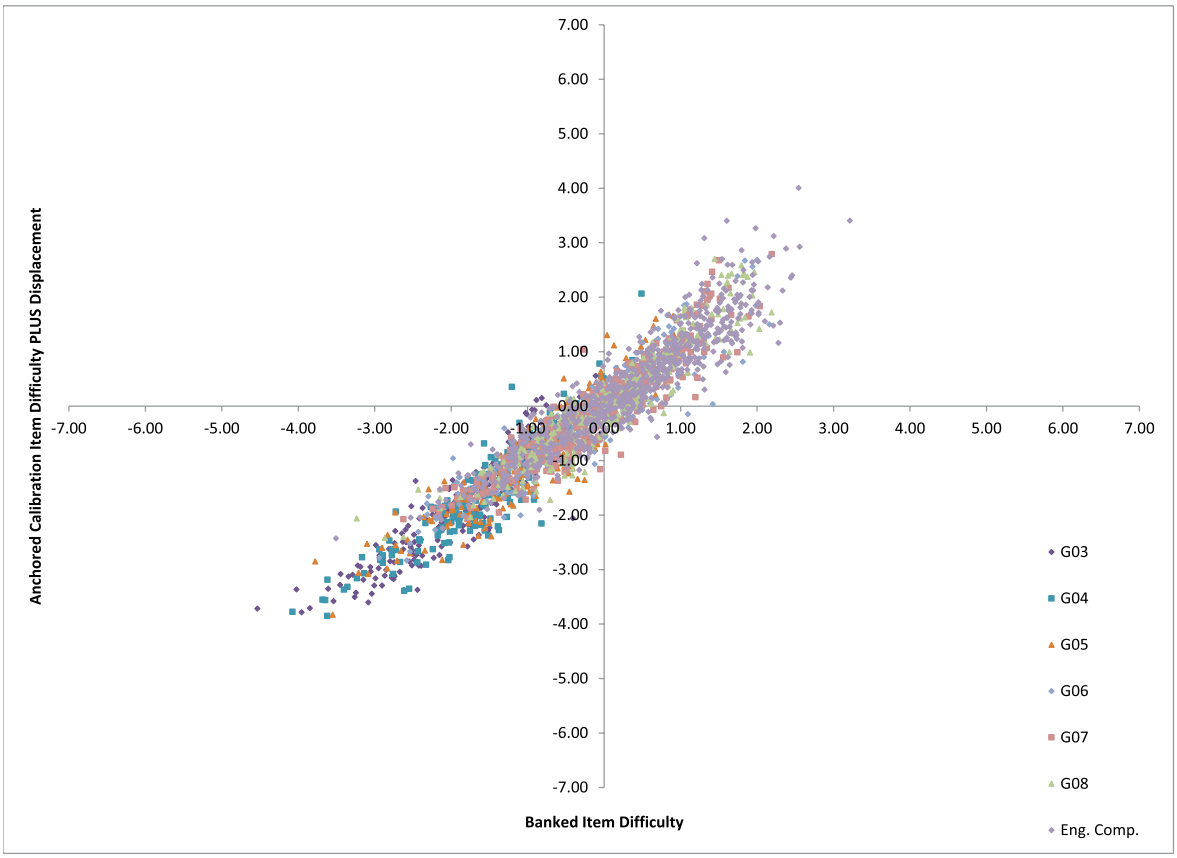
Table 18–23. Number of Writing Items by Grade/Course and Displacement Interval

Interval	G03	G04	G05	G06	G07	G08	COMP	Total
Disp. ≤ -1.0	1	1	2	3	3	1	4	15
-1.0 < Disp. ≤ -0.9	1	1	2	2	0	3	1	10
-0.9 < Disp. ≤ -0.8	0	4	0	0	1	1	4	10
-0.8 < Disp. ≤ -0.7	6	5	6	6	6	2	13	44
-0.7 < Disp. ≤ -0.6	0	2	8	6	7	9	8	40
-0.6 < Disp. ≤ -0.5	6	7	9	15	9	7	24	77
-0.5 < Disp. ≤ -0.4	9	6	11	12	14	19	33	104
-0.4 < Disp. ≤ -0.3	18	9	13	13	15	16	37	121
-0.3 < Disp. ≤ -0.2	19	18	20	23	33	27	54	194
-0.2 < Disp. ≤ -0.1	20	16	22	33	28	30	83	232
-0.1 < Disp. ≤ 0.0	23	12	13	37	40	32	74	231
0.0 < Disp. ≤ 0.1	15	25	25	39	36	29	61	230
0.1 < Disp. ≤ 0.2	24	19	23	56	58	38	63	281
0.2 < Disp. ≤ 0.3	27	17	23	30	38	27	56	218
0.3 < Disp. ≤ 0.4	16	16	17	24	22	20	45	160
0.4 < Disp. ≤ 0.5	11	12	6	13	17	12	33	104
0.5 < Disp. ≤ 0.6	8	7	13	14	9	9	27	87
0.6 < Disp. ≤ 0.7	5	0	6	9	8	9	14	51
0.7 < Disp. ≤ 0.8	3	3	4	2	1	4	17	34
0.8 < Disp. ≤ 0.9	4	2	2	4	1	2	5	20
0.9 < Disp. ≤ 1.0	3	0	3	0	0	0	8	14
1.0 < Disp.	1	2	2	1	3	2	10	21
TOTAL	220	184	230	342	349	299	674	2298

Figure 18–30 mirrors Figure 18–28, except the calibrations exclude items with fewer than 100 administrations. Again, a line of best fit is included in the upper plot.

Figure 18–30. Writing Banked Item Parameters vs. Anchored Grade Level Calibrations — All Items in Grade 3 and Above with *N*>100





For the two sets of anchored grade level calibrations described above, banked item parameters were compared to the newly calibrated values by calculating a robust Z statistic for each item pairing. If item difficulties from the operational calibration are close to the banked values, the correlation will be high and the additive constant near zero. Table 18–24 shows the number of items in each grade/course and the number and percent of items with absolute value of robust Z greater than 1.645 in the calibrations.

Table 18–24. Summary of Robust Z across Two Sets of Anchored Grade Level Calibrations in Writing

Grade/ Course	Cal 1: Number of Items	Cal 1: Number of Items with ABS(Z) > 1.645	Cal 1: Percent of Items with ABS(Z) > 1.645	Cal 2: Number of Items	Cal 2: Number of Items with ABS(Z) > 1.645	Cal 2: Percent of Items with ABS(Z) > 1.645
Kindergarten	0	0	N/A	0	0	N/A
Grade 1	0	0	N/A	0	0	N/A
Grade 2	0	0	N/A	0	0	N/A
Grade 3	360	53	15%	220	27	12%
Grade 4	331	59	18%	184	24	13%
Grade 5	309	54	17%	230	41	18%
Grade 6	361	42	12%	342	43	13%
Grade 7	349	30	9%	349	34	10%
Grade 8	325	37	11%	299	37	12%
English Comp	733	91	12%	674	95	14%
Total	2768	366	13%	2298	301	13%
	Correlation = 0.944			Correlation = 0.95		
	Additive Constant = 0.031			Additive Constant = 0.025		

For the most part, whether high absolute displacement values or robust Z was used to identify items with operational estimates that differ from banked values, the same items were identified. For example, in calibration 1, all items with absolute displacement greater than 0.624 have an absolute value of robust Z greater than 1.645. In the displacement range of 0.563 to 0.624, some items have absolute value of robust Z greater than 1.645 while others do not. No items with absolute displacement less than 0.563 have absolute value of robust Z greater than 1.645.

For each of the content areas, it is evident from this series of plots that the item parameter estimates are reasonably stable for the items in grade 3 and above.

CHAPTER NINETEEN: REVISION OF BENCHMARK CUTS

As described in Chapter Fourteen, CDT scores are placed along a continuum from "Areas of Need" to "Strengths to Build On." These are represented in the dynamic reporting suite with colors red, green, and blue. "Areas of Need" are depicted in the red range, while "Strengths to Build On" are depicted in the green and blue ranges. The center of the green range for grades 5 and above was established by panels of Pennsylvania educators during preliminary benchmarking activities (see Chapter Ten for details). The center of the green range for grades 2 through 4 was extrapolated from grades 5 and above prior to the launch of the CDT tests for students in grades 3 through 5 in spring of 2014.

The preliminary benchmarking activities took place prior to the first operational administration in each content area so that, once operational, immediate score reports would be available to students and teachers. Given that the preliminary benchmark cuts were set prior to the operational administration and based on field-test data, it was planned at that time to revisit the location of the cut scores after enough operational data had been collected. The preliminary benchmark cut points in the mathematics content area were analyzed and revised based on operational data following the 2010–2011 school year. The preliminary benchmark cut points in the reading, science, and writing content areas were analyzed and revised based on operational data following the 2011–2012 school year.

The introduction of CDT tests for students in grades 3 through 5 in spring 2014 required benchmark cuts for grades 2 through 4. For each content area, the benchmark cuts in place for the 2013–2014 school year in grades 5 and above were used to extrapolate cuts in grades 2 through 4.

Prior to the start of the 2015–2016 school year, the benchmark cut points in mathematics, reading, and writing were revised based on the revised Pennsylvania System of School Assessments (PSSA) tests and cut points established in spring 2015.

This chapter summarizes changes to the benchmark cuts.

FIRST REVISION OF BENCHMARK CUTS BASED ON OPERATIONAL DATA

In each content area, the benchmark cut points set during preliminary benchmarking activities were analyzed based on matched data sets – operational CDT with PSSA and Keystone Exams (Keystone). CDT benchmark cuts were not revised to exactly match PSSA and Keystone cuts or be predictive. However, CDT, PSSA, and Keystone are based on the same eligible content. As such, it is reasonable to expect that students who do well on CDT will do well on PSSA/Keystone and vice versa. In looking at CDT results matched to PSSA and Keystone results it was determined that many students who scored in the CDT red range scored Proficient or Advanced on PSSA or Keystone suggesting that CDT benchmark cuts were set too high. Therefore, CDT benchmark cuts were lowered to make CDT red/green/blue classifications more consistent with PSSA and Keystone results. See Chapter Nineteen of the 2010–2011 and 2011–2012 technical reports for details. Table 19–1 provides a summary of the first revisions to the benchmark cut points.

Table 19–1. Summary of First Revision to Benchmark Cuts

Content Area	Course /Grade	Benchmarking Logit Cut Point	First Revision to Logit Cut Point	Difference in Logit Cut Point	Difference in Scale Score
Mathematics	Grade 5	-0.292	-0.792	-0.500	-63
Mathematics	Grade 6	0.526	0.026	-0.500	-62
Mathematics	Grade 7	1.495	0.495	-1.000	-125
Mathematics	Grade 8	2.238	0.838	-1.400	-175
Mathematics	High School	3.363	1.613	-1.750	-218
Mathematics	Algebra I	3.363	1.613	-1.750	-218
Mathematics	Geometry	3.614	1.864	-1.750	-219
Mathematics	Algebra II	4.117	2.367	-1.750	-219
Reading	Grade 5	1.529	0.529	-1.000	-143
Reading	Grade 6	2.015	1.015	-1.000	-142
Reading	Grade 7	2.299	1.299	-1.000	-143
Reading	Grade 8	2.500	1.500	-1.000	-143
Reading	Literature	2.657	1.657	-1.000	-143
Science	Grade 5	1.099	-0.451	-1.550	-206
Science	Grade 6	1.522	-0.028	-1.550	-206
Science	Grade 7	1.879	0.329	-1.550	-206
Science	Grade 8	2.189	0.639	-1.550	-206
Science	High School	2.462	1.112	-1.350	-179
Science	Biology	2.462	1.112	-1.350	-179
Science	Chemistry	2.706	1.356	-1.350	-179
Writing	Grade 5	0.731	-0.569	-1.300	-173
Writing	Grade 6	1.363	0.063	-1.300	-172
Writing	Grade 7	1.886	0.586	-1.300	-173
Writing	Grade 8	2.219	0.919	-1.300	-173
Writing	English Composition	2.281	0.981	-1.300	-173

EXTRAPOLATION OF BENCHMARK CUTS FOR GRADES 2 THROUGH 4

The introduction of CDT tests for students in grades 3 through 5 in spring 2014 required benchmark cuts for grades 2 through 4[1]. For each content area, the benchmark cuts in place for the 2013–2014 school year in grades 5 and above were used to extrapolate cuts in grades 2 through 4. See Chapter Nineteen of the 2013–2014 technical report for details.

[1] It is not expected that students in grade 2 will use the CDT. However, teachers may want to use a grade 2 benchmark when looking at reports for students in grade 3, especially early in the school year.

REVISION OF BENCHMARK CUTS BASED ON CHANGES TO PSSA

In spring 2015, changes were made to PSSA test designs and cut points in mathematics and English language arts. In light of these changes, CDT benchmark cuts were analyzed again using matched data sets - operational CDT with PSSA and Keystone. The new PSSA cut points approved in July 2015 represented higher, more rigorous, standards. Therefore, CDT benchmark cuts in mathematics, reading, and writing were raised to make CDT red/green/blue classifications more consistent with PSSA. See Chapter Nineteen of the 2015–2016 technical report for details. Table 19–2 provides a summary of the revisions to the benchmark cut points based on changes to PSSA.

Table 19–2. Summary of Second Revision to Benchmark Cuts

CDT	Course /Grade	2014–2015 Logit Cut Point	2015–2016 Logit Cut Point	Difference in Logit Cut Point	Difference in Scale Score
Math Grades 3–5	Grade 2	-2.828	-1.628	1.200	150
Math Grades 3–5	Grade 3	-2.083	-0.883	1.200	150
Math Grades 3–5	Grade 4	-1.380	-0.180	1.200	150
Math Grades 3–5	Grade 5	-0.792	0.208	1.000	125
Math Gr 6–HS	Grade 6	0.026	0.726	0.700	87
Math Gr 6–HS	Grade 7	0.495	1.195	0.700	88
Math Gr 6–HS	Grade 8	0.838	1.513	0.675	84
Math Gr 6–HS	High School	1.613	1.613	0.000	0
Algebra I	Algebra I	1.613	1.613	0.000	0
Geometry	Geometry	1.864	1.864	0.000	0
Algebra II	Algebra II	2.367	2.367	0.000	0
Reading Grades 3–5	Grade 2	-1.136	-0.936	0.200	29
Reading Grades 3–5	Grade 3	-0.367	-0.167	0.200	29
Reading Grades 3–5	Grade 4	0.179	0.429	0.250	36
Reading Grades 3–5	Grade 5	0.529	0.879	0.350	50
Read/Lit Grades 6–HS	Grade 6	1.015	1.265	0.250	35
Read/Lit Grades 6–HS	Grade 7	1.299	1.499	0.200	29
Read/Lit Grades 6–HS	Grade 8	1.500	1.725	0.225	32
Read/Lit Grades 6–HS	Literature	1.657	1.882	0.225	32
Writing Grades 3–5	Grade 2	-2.989	-1.739	1.250	166
Writing Grades 3–5	Grade 3	-1.874	-0.624	1.250	166
Writing Grades 3–5	Grade 4	-1.084	-0.084	1.000	133
Writing Grades 3–5	Grade 5	-0.569	0.281	0.850	113
Writing/Eng Comp Gr 6–HS	Grade 6	0.063	0.563	0.500	66
Writing/Eng Comp Gr 6–HS	Grade 7	0.586	0.836	0.250	33
Writing/Eng Comp Gr 6–HS	Grade 8	0.919	0.919	0.000	0
Writing/Eng Comp Gr 6–HS	English Composition	0.981	0.981	0.000	0

BENCHMARK CUTS FOR ALL GRADES AND COURSES FOR THE 2020–2021 SCHOOL YEAR

Table 19–3 shows the benchmark cuts used for student reporting during the 2020–2021 school year in the logit metric for each content area. Also presented are the scale score ranges for each color on the CDT reports.

Table 19–3. Benchmark Cuts and Scale Score Ranges for the 2020–2021 School Year

CDT	Course/Grade	Logit Cut Point (Center of Green)	Red Scale Score Range	Green Scale Score Range	Blue Scale Score Range
Math Grades 3–5	Grade 2	-1.628	200 - 728	729 - 891	892 - 2000
Math Grades 3–5	Grade 3	-0.883	200 - 821	822 - 984	985 - 2000
Math Grades 3–5	Grade 4	-0.180	200 - 909	910 - 1072	1073 - 2000
Math Grades 3–5	Grade 5	0.208	200 - 957	958 - 1120	1121 - 2000
Math Gr 6–HS	Grade 6	0.726	200 - 1022	1023 - 1185	1186 - 2000
Math Gr 6–HS	Grade 7	1.195	200 - 1081	1082 - 1244	1245 - 2000
Math Gr 6–HS	Grade 8	1.513	200 - 1120	1121 - 1283	1284 - 2000
Math Gr 6–HS	High School	1.613	400 - 1133	1134 - 1296	1297 - 2000
Algebra I	Algebra I	1.613	400 - 1133	1134 - 1296	1297 - 2000
Geometry	Geometry	1.864	400 - 1164	1165 - 1327	1328 - 2000
Algebra II	Algebra II	2.367	400 - 1227	1228 - 1390	1391 - 2000
Reading Grades 3–5	Grade 2	-0.936	200 - 630	631 - 845	846 - 2000
Reading Grades 3–5	Grade 3	-0.167	200 - 740	741 - 955	956 - 2000
Reading Grades 3–5	Grade 4	0.429	200 - 825	826 - 1040	1041 - 2000
Reading Grades 3–5	Grade 5	0.879	200 - 889	890 - 1104	1105 - 2000
Read/Lit Grades 6–HS	Grade 6	1.265	200 - 944	945 - 1159	1160 - 2000
Read/Lit Grades 6–HS	Grade 7	1.499	200 - 978	979 - 1193	1194 - 2000
Read/Lit Grades 6–HS	Grade 8	1.725	200 - 1010	1011 - 1225	1226 - 2000
Read/Lit Grades 6–HS	Literature	1.882	200 - 1032	1033 - 1247	1248 - 2000
Science Grades 3–5	Grade 2	-1.723	200 - 634	635 - 807	808 - 2000
Science Grades 3–5	Grade 3	-1.282	200 - 693	694 - 866	867 - 2000
Science Grades 3–5	Grade 4	-0.855	200 - 750	751 - 923	924 - 2000
Science Grades 3–5	Grade 5	-0.451	200 - 803	804 - 976	977 - 2000
Science Gr 6–HS	Grade 6	-0.028	200 - 860	861 - 1033	1034 - 2000
Science Gr 6–HS	Grade 7	0.329	200 - 907	908 - 1080	1081 - 2000
Science Gr 6–HS	Grade 8	0.639	200 - 948	949 - 1121	1122 - 2000
Science Gr 6–HS	High School	1.112	400 - 1011	1012 - 1184	1185 - 2000
Biology	Biology	1.112	400 - 1011	1012 - 1184	1185 - 2000
Chemistry	Chemistry	1.356	400 - 1044	1045 - 1217	1218 - 2000
Writing Grades 3–5	Grade 2	-1.739	200 - 631	632 - 804	805 - 2000
Writing Grades 3–5	Grade 3	-0.624	200 - 779	780 - 952	953 - 2000
Writing Grades 3–5	Grade 4	-0.084	200 - 851	852 - 1024	1025 - 2000
Writing Grades 3–5	Grade 5	0.281	200 - 899	900 - 1072	1073 - 2000
Writing/Eng Comp Gr 6–HS	Grade 6	0.563	200 - 937	938 - 1110	1111 - 2000

Table 19–3 (continued). Benchmark Cuts and Scale Score Ranges for the 2018–2019 School Year

CDT	Course/Grade	Logit Cut Point (Center of Green)	Red Scale Score Range	Green Scale Score Range	Blue Scale Score Range
Writing/Eng Comp Gr 6–HS	Grade 7	0.836	200 - 973	974 - 1146	1147 - 2000
Writing/Eng Comp Gr 6–HS	Grade 8	0.919	200 - 984	985 - 1157	1158 - 2000
Writing/Eng Comp Gr 6–HS	English Composition	0.981	200 - 993	994 - 1166	1167 - 2000

APPENDIX A: GENERAL DEVELOPMENT AND FIELD TEST CYCLE FOR THE CLASSROOM DIAGNOSTIC TOOLS

Table A–1. General Development and Field Test Cycle for the Classroom Diagnostic Tools

	Mathematics	Reading/Literature	Science	Writing/English Composition
Summer/Fall 2009	Item Development and Internal Reviews			
Winter 2009/2010	Item Review by Pennsylvania Educators	Item Development and Internal Reviews	Item Development and Internal Reviews	
Spring 2010	Stand-alone Field Test	Item Development and Internal Reviews	Item Development and Internal Reviews	
Summer 2010	Data Review, Items Aligned to the Learning Progression Map, and Benchmarking	Item Review by Pennsylvania Educators	Item Review by Pennsylvania Educators	Item Development and Internal Reviews
Fall 2010	Operational Assessments Available	Stand-alone Field Test	Stand-alone Field Test	Item Development and Internal Reviews
Winter 2010/2011	Operational Assessments Available	Data Review, Items Aligned to the Learning Progression Map, and Benchmarking	Data Review, Items Aligned to the Learning Progression Map, and Benchmarking	Item Review by Pennsylvania Educators
Spring 2011	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available	Stand-alone Field Test
Summer 2011				Data Review, Items Aligned to the Learning Progression Map, and Benchmarking
Fall 2011	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available
Winter 2011/2012	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available
Spring 2012	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available	Operational Assessments Available
Summer 2012	Item Development and Internal Reviews of Items Aligned to Pennsylvania Core Standards Begins	Item Development and Internal Reviews of Items Aligned to Pennsylvania Core Standards Begins		
Fall 2012	Operational Assessments Available and Completion of Item Development and Internal Reviews of Items Aligned to Pennsylvania Core Standards	Operational Assessments Available and Completion of Item Development and Internal Reviews of Items Aligned to Pennsylvania Core Standards	Operational Assessments Available	Operational Assessments Available

Table A–1 (continued). General Development and Field Test Cycle for the Classroom Diagnostic Tools

	Mathematics	Reading/Literature	Science	Writing/English Composition
Winter 2012/2013	Operational Assessments Available and Item Review by Pennsylvania Educators for Items Aligned to Pennsylvania Core Standards	Operational Assessments Available and Item Review by Pennsylvania Educators for Items Aligned to Pennsylvania Core Standards	Operational Assessments Available	Operational Assessments Available
Spring 2013	Operational Assessments with Embedded Field Test Items Aligned to the Pennsylvania Core Standards Available and Item Development and Internal Reviews of Items for Lower Grades CDT	Operational Assessments with Embedded Field Test Items Aligned to the Pennsylvania Core Standards Available and Item Development and Internal Reviews of Items Lower Grades CDT	Operational Assessments Available and Item Development and Internal Reviews of Items for Lower Grades CDT	Operational Assessments Available and Item Development and Internal Reviews of Items for Lower Grades CDT
Summer 2013	Data Review and Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators for Items for Lower Grades	Data Review and Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators for Items for Lower Grades	Item Review by Pennsylvania Educators for Items for Lower Grades	Item Review by Pennsylvania Educators for Items for Lower Grades
Fall 2013	Operational Assessments Aligned to PCS Including Embedded Field Test Items at Grade 6 Available and Stand-alone Field Test for Lower Grades	Operational Assessments Aligned to PCS Including Embedded Field Test Items at Grade 6 Available and Stand-alone Field Test for Lower Grades	Operational Assessments Aligned to PCS Including Embedded Field Test Items at Grade 6 Available and Stand-alone Field Test for Lower Grades	Operational Assessments Aligned to PCS Including Embedded Field Test Items at Grade 6 Available and Stand-alone Field Test for Lower Grades
Winter 2013/2014	Operational Assessments Aligned to PCS Available and Data Review and Items Aligned to the Learning Progression Map for Items for Lower Grades CDT	Operational Assessments Aligned to PCS Available and Data Review and Items Aligned to the Learning Progression Map for Items for Lower Grades CDT	Operational Assessments Aligned to PCS Available and Data Review and Items Aligned to the Learning Progression Map for Items for Lower Grades CDT	Operational Assessments Aligned to PCS Available and Data Review and Items Aligned to the Learning Progression Map for Items for Lower Grades CDT
Spring 2014	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available
Winter 2014/2015	Item Development and Internal Reviews of Replenishment Items for Grades 6–HS CDT	Item Development and Internal Reviews of Replenishment Items for Grades 6–HS and EBSR items for all grade levels CDT	Item Development and Internal Reviews of Replenishment Items for Grades 6–HS CDT	Item Development and Internal Reviews of Replenishment Items for Grades 6–HS CDT
Spring 2015	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available

Table A–1 (continued). General Development and Field Test Cycle for the Classroom Diagnostic Tools

	Mathematics	Reading/Literature	Science	Writing/English Composition
Spring 2016	Data Review of Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators and Operational Assessments, including Lower Grades, Available	Data Review of Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators and Operational Assessments, including Lower Grades, Available	Data Review of Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators and Operational Assessments, including Lower Grades, Available	Data Review of Items Aligned to the Learning Progression Map for Items Aligned to the Pennsylvania Core Standards and Item Review by Pennsylvania Educators and Operational Assessments, including Lower Grades, Available
Spring 2017	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available	Operational Assessments, including Lower Grades, Available
Winter 2017/2018	Item Development and Internal Reviews of Replenishment Items for Grades K–HS. Item Review by Pennsylvania Educators.	Item Development and Internal Reviews of Replenishment Items for Grades K–HS. Item Review by Pennsylvania Educators.	Item Development and Internal Reviews of Replenishment Items for Grades K–HS. Item Review by Pennsylvania Educators.	Item Development and Internal Reviews of Replenishment Items for Grades K–HS. Item Review by Pennsylvania Educators.
Fall 2018	Operational Assessments Aligned to PCS Including Embedded Field Test Items	Operational Assessments Aligned to PCS Including Embedded Field Test Items	Operational Assessments Aligned to PCS Including Embedded Field Test Items. Item Development and Internal Reviews of Technology Enhanced Items.	Operational Assessments Aligned to PCS Including Embedded Field Test Items
Spring 2019	Data Review by Pennsylvania Educators.	Data Review by Pennsylvania Educators.	Data Review by Pennsylvania Educators.	Data Review by Pennsylvania Educators.
Summer 2019	Operational Assessments of full CDT assessments and Diagnostic Category Assessments Aligned to PCS	Operational Assessments of full CDT assessments and Diagnostic Category Assessments Aligned to PCS	Operational Assessments of full CDT assessments and Diagnostic Category Assessments Aligned to PCS	Operational Assessments of full CDT assessments and Diagnostic Category Assessments Aligned to PCS
Spring 2020			Data Review of Technology Enhanced Items by Pennsylvania Educators.	
Summer 2020			Operational Assessments of full CDT assessments and Diagnostic Category Assessments Aligned to PCS	

APPENDIX B: FIELD TEST ITEM STATISTICS

There were no items field tested during the 2020-2021 school year. Location of classical item statistics for previously field tested items are detailed below.

MATHEMATICS MULTIPLE-CHOICE ITEMS

There were no multiple-choice items in the mathematics content area field tested during the 2020–2021 school year. Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–19 can be found in Appendix B of the 2018–2019 technical report.

READING/LITERATURE MULTIPLE-CHOICE ITEMS

There were no multiple-choice items in the reading content area field tested during the 2020–2021 school year. Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–19 can be found in Appendix B of the 2018–2019 technical report.

SCIENCE MULTIPLE-CHOICE ITEMS

There were no multiple-choice items in the science content area field tested during the 2020–2021 school year. Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–2019 can be found in Appendix B of the 2018–2019 technical report.

WRITING/ENGLISH COMPOSITION MULTIPLE-CHOICE ITEMS

There were no multiple-choice items in the writing content area field tested during the 2020–2021 school year. Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–2019 can be found in Appendix B of the 2018–2019 technical report.

READING/LITERATURE EVIDENCE-BASED SELECTED-RESPONSE ITEMS

There were no evidence-based selected-response items in the reading content area field tested during the 2020–2021 school year. Classical item statistics for all items field tested prior to 2018–2019 can be found in Appendix B of the 2017–2018 technical report. Classical item statistics for items field tested in 2018–2019 can be found in Appendix B of the 2018–2019 technical report.

SCIENCE TECHNOLOGY-ENHANCED ITEMS

There were no technology-enhanced items in the science content area field tested during the 2020–2021 school year. Classical item statistics for items field tested in 2019–2020 can be found in Appendix B of the 2019–2020 technical report.

APPENDIX C: VERTICAL LINKING ITEM DETAILS

This appendix provides details on the items used to build the vertical scales in each content area. Information such as grade, n-count, eligible content code, and diagnostic category is provided for each of the vertical linking items. This information is based on the academic standards in place at the time each of the content area vertical scale was established[1]. Summary tables indicate the number of linking items in each diagnostic category. A sample of the vertical linking Excel file is provided as well as plots of the vertical linking items.

MATHEMATICS

Tables C–1 through C–8 show n-counts, eligible content code, and diagnostic category for each of the vertical linking items.

Each item was administered in two grades so there are two n-counts: one for the lower grade and one for the upper grade. For example, item 600869 is a grade 3 item used to link grades 3 and 4. It was administered 1,280 times on the lower grade forms (grade 3) and 964 times on the upper grade forms (grade 4).

Diagnostic categories for Algebra I, Geometry, and Algebra II are different than diagnostic categories for grades 3 through 8 and 11 Mathematics. Items may fall into both a Mathematics diagnostic category and an Algebra I, Geometry, or Algebra II diagnostic category. This is shown in Tables C–6, C–7, and C–8. For example, item 601329 is in the Mathematics diagnostic category "Geometry" and the Geometry diagnostic category "Coordinate Geometry and Right Triangles".

The Mathematics diagnostic categories are[2]:

- Numbers and Operations
- Measurement
- Geometry
- Algebraic Concepts
- Data Analysis and Probability

The Algebra I diagnostic categories are:

- Operations with Real Numbers and Expressions
- Linear Equations & Inequalities
- Functions & Coordinate Geometry
- Data Analysis

The Geometry diagnostic categories are:

- Geometric Properties
- Congruence, Similarity, & Proofs
- Coordinate Geometry and Right Triangles
- Measurement

The Algebra II diagnostic categories are:

- Operations with Complex Numbers
- Non-linear Expressions & Equations
- Functions
- Data Analysis

[1] Before the 2013–2014 school year items in mathematics, reading, and writing were re-aligned to the new Pennsylvania Core Standards.

[2] Mathematics diagnostic categories changed at the start of the 2013–2014 school year due to re-alignment to the Pennsylvania Core Standards. See Chapter Thirteen for a list of the current diagnostic categories.

Table C–1. Mathematics Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
600869	3	Grade 3 to Grade 4	1280	964	M3.B.1.1.1	Measure.
600871	3	Grade 3 to Grade 4	1275	964	M3.B.2.2.1	Measure.
601980	3	Grade 3 to Grade 4	1280	964	M3.B.1.2.1	Measure.
604352	3	Grade 3 to Grade 4	1281	964	M3.D.2.1.1	Alg. Con.
600442	3	Grade 3 to Grade 4	1280	964	M3.C.2.1.1	Geo.
600431	3	Grade 3 to Grade 4	1274	964	M3.A.1.1.1	Numbers & Op.
601975	3	Grade 3 to Grade 4	1281	964	M3.A.2.1.1	Numbers & Op.
600865	3	Grade 3 to Grade 4	1279	964	M3.A.1.3.1	Numbers & Op.
601985	3	Grade 3 to Grade 4	1285	963	M3.E.1.1.1	Data & Prob.
601897	3	Grade 3 to Grade 4	1282	964	M3.A.1.2.1	Numbers & Op.
601437	3	Grade 3 to Grade 4	1274	963	M3.A.1.1.4	Numbers & Op.
600438	3	Grade 3 to Grade 4	1277	963	M3.A.1.2.2	Numbers & Op.
600427	3	Grade 3 to Grade 4	1282	963	M3.C.1.1.1	Geo.
600877	3	Grade 3 to Grade 4	1283	963	M3.E.1.2.1	Data & Prob.
601587	3	Grade 3 to Grade 4	1276	963	M3.A.2.1.3	Numbers & Op.
600440	3	Grade 3 to Grade 4	639	963	M3.B.2.1.1	Measure.
600921	3	Grade 3 to Grade 4	1271	963	M3.A.1.3.2	Numbers & Op.
601589	3	Grade 3 to Grade 4	639	962	M3.D.1.1.1	Alg. Con.
601440	3	Grade 3 to Grade 4	1272	962	M3.B.1.1.3	Measure.
601984	3	Grade 3 to Grade 4	1278	962	M3.D.2.1.2	Alg. Con.
604193	4	Grade 3 to Grade 4	1283	959	M4.D.1.1.2	Alg. Con.
602015	4	Grade 3 to Grade 4	1284	481	M4.E.1.2.1	Data & Prob.
601993	4	Grade 3 to Grade 4	1282	1447	M4.C.1.1.1	Geo.
603609	4	Grade 3 to Grade 4	1284	959	M4.B.2.1.1	Measure.
604189	4	Grade 3 to Grade 4	1280	962	M4.B.1.1.3	Measure.
602010	4	Grade 3 to Grade 4	1285	961	M4.C.1.1.2	Geo.
601646	4	Grade 3 to Grade 4	1283	960	M4.D.2.2.2	Alg. Con.
604186	4	Grade 3 to Grade 4	1279	965	M4.A.3.1.1	Numbers & Op.
601958	4	Grade 3 to Grade 4	1281	961	M4.A.1.1.2	Numbers & Op.
604488	4	Grade 3 to Grade 4	1279	958	M4.A.1.2.2	Numbers & Op.
603744	4	Grade 3 to Grade 4	1279	481	M4.B.2.2.1	Measure.
602009	4	Grade 3 to Grade 4	1279	963	M4.C.1.1.2	Geo.
604514	4	Grade 3 to Grade 4	1280	481	M4.C.2.1.1	Geo.
604492	4	Grade 3 to Grade 4	1278	961	M4.A.3.1.2	Numbers & Op.
601972	4	Grade 3 to Grade 4	1281	965	M4.E.1.2.2	Data & Prob.
601962	4	Grade 3 to Grade 4	1278	962	M4.A.1.3.2	Numbers & Op.
601987	4	Grade 3 to Grade 4	1278	961	M4.A.1.1.4	Numbers & Op.

Table C–1 (continued). Mathematics Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
604195	4	Grade 3 to Grade 4	1279	481	M4.D.2.1.1	Alg. Con.
604501	4	Grade 3 to Grade 4	1279	959	M4.E.1.1.1	Data & Prob.
604493	4	Grade 3 to Grade 4	1279	1443	M4.B.1.1.4	Measure.

Table C–2. Mathematics Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
601646	4	Grade 4 to Grade 5	960	1187	M4.D.2.2.2	Alg. Con.
601987	4	Grade 4 to Grade 5	961	1186	M4.A.1.1.4	Numbers & Op.
604493	4	Grade 4 to Grade 5	1443	1183	M4.B.1.1.4	Measure.
601961	4	Grade 4 to Grade 5	965	1184	M4.A.1.3.2	Numbers & Op.
604499	4	Grade 4 to Grade 5	962	1188	M4.E.1.1.1	Data & Prob.
602889	4	Grade 4 to Grade 5	962	1187	M4.E.1.2.2	Data & Prob.
602885	4	Grade 4 to Grade 5	965	1186	M4.B.2.2.1	Measure.
602887	4	Grade 4 to Grade 5	962	1187	M4.C.3.1.1	Geo.
601639	4	Grade 4 to Grade 5	960	1184	M4.A.3.1.3	Numbers & Op.
604969	4	Grade 4 to Grade 5	480	1184	M4.C.1.2.2	Geo.
601994	4	Grade 4 to Grade 5	479	1185	M4.D.1.2.2	Alg. Con.
601998	4	Grade 4 to Grade 5	960	1191	M4.E.3.1.1	Data & Prob.
602000	4	Grade 4 to Grade 5	959	1190	M4.C.1.1.1	Geo.
601991	4	Grade 4 to Grade 5	959	1189	M4.A.2.1.2	Numbers & Op.
604879	4	Grade 4 to Grade 5	1441	1188	M4.D.1.1.3	Alg. Con.
601964	4	Grade 4 to Grade 5	961	1188	M4.A.3.2.2	Numbers & Op.
602971	4	Grade 4 to Grade 5	480	1187	M4.B.2.1.1	Measure.
604486	4	Grade 4 to Grade 5	481	1186	M4.E.1.2.1	Data & Prob.
604967	4	Grade 4 to Grade 5	962	1187	M4.A.1.2.2	Numbers & Op.
602973	4	Grade 4 to Grade 5	964	1186	M4.C.2.1.1	Geo.
600853	5	Grade 4 to Grade 5	964	1790	M5.B.2.1.1	Measure.
604790	5	Grade 4 to Grade 5	964	586	M5.C.2.1.2	Geo.
604956	5	Grade 4 to Grade 5	959	1175	M5.A.2.1.1	Numbers & Op.
604862	5	Grade 4 to Grade 5	960	1182	M5.D.1.2.1	Alg. Con.
604783	5	Grade 4 to Grade 5	961	1179	M5.A.1.2.1	Numbers & Op.
606159	5	Grade 4 to Grade 5	960	1190	M5.A.1.5.1	Numbers & Op.
604848	5	Grade 4 to Grade 5	961	1784	M5.E.3.1.1	Data & Prob.
604843	5	Grade 4 to Grade 5	959	1186	M5.C.1.1.2	Geo.
604966	5	Grade 4 to Grade 5	961	596	M5.E.1.1.1	Data & Prob.
606163	5	Grade 4 to Grade 5	961	1188	M5.B.1.1.1	Measure.
601532	5	Grade 4 to Grade 5	956	2369	M5.A.1.1.1	Numbers & Op.
606160	5	Grade 4 to Grade 5	958	1190	M5.A.3.1.1	Numbers & Op.
604960	5	Grade 4 to Grade 5	957	594	M5.B.2.2.3	Measure.
600852	5	Grade 4 to Grade 5	958	1178	M5.D.1.1.1	Alg. Con.
604834	5	Grade 4 to Grade 5	954	1189	M5.A.1.3.1	Numbers & Op.
604959	5	Grade 4 to Grade 5	956	1183	M5.B.1.2.2	Measure.
604961	5	Grade 4 to Grade 5	956	1193	M5.C.1.2.1	Geo.

Table C–2 (continued). Mathematics Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
606278	5	Grade 4 to Grade 5	954	1177	M5.D.2.1.2	Alg. Con.
604965	5	Grade 4 to Grade 5	957	1190	M5.E.1.1.1	Data & Prob.
604865	5	Grade 4 to Grade 5	956	1192	M5.A.1.6.2	Numbers & Op.

Table C–3. Mathematics Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
606277	5	Grade 5 to Grade 6	1175	1225	M5.D.2.1.2	Alg. Con.
606153	5	Grade 5 to Grade 6	590	1225	M5.A.1.4.2	Numbers & Op.
604796	5	Grade 5 to Grade 6	1194	1224	M5.B.1.3.2	Measure.
606154	5	Grade 5 to Grade 6	1195	1223	M5.A.2.1.3	Numbers & Op.
604962	5	Grade 5 to Grade 6	1192	1222	M5.C.1.2.1	Geo.
606826	5	Grade 5 to Grade 6	593	1221	M5.A.1.3.2	Numbers & Op.
604859	5	Grade 5 to Grade 6	1766	1223	M5.C.1.1.1	Geo.
604860	5	Grade 5 to Grade 6	1184	1215	M5.D.1.2.1	Alg. Con.
606167	5	Grade 5 to Grade 6	1181	1216	M5.E.3.1.1	Data & Prob.
604836	5	Grade 5 to Grade 6	1176	1216	M5.A.1.6.1	Numbers & Op.
606162	5	Grade 5 to Grade 6	593	1216	M5.B.1.1.1	Measure.
604841	5	Grade 5 to Grade 6	594	1215	M5.B.2.2.1	Measure.
606155	5	Grade 5 to Grade 6	1193	1215	M5.C.2.1.2	Geo.
601592	5	Grade 5 to Grade 6	595	1214	M5.E.2.1.1	Data & Prob.
601590	5	Grade 5 to Grade 6	2372	1214	M5.A.1.1.1	Numbers & Op.
604953	5	Grade 5 to Grade 6	1171	1226	M5.A.1.3.3	Numbers & Op.
604853	5	Grade 5 to Grade 6	1175	1227	M5.A.1.5.1	Numbers & Op.
604784	5	Grade 5 to Grade 6	1178	1227	M5.A.1.2.1	Numbers & Op.
604868	5	Grade 5 to Grade 6	1176	1225	M5.B.1.2.1	Measure.
604964	5	Grade 5 to Grade 6	1190	1226	M5.E.1.1.1	Data & Prob.
601542	5	Grade 5 to Grade 6	1189	1225	M5.B.2.1.1	Measure.
606276	5	Grade 5 to Grade 6	590	1223	M5.C.2.1.1	Geo.
604856	5	Grade 5 to Grade 6	1180	1219	M5.A.3.1.1	Numbers & Op.
606166	5	Grade 5 to Grade 6	1181	1220	M5.D.2.1.1	Alg. Con.
604958	5	Grade 5 to Grade 6	1176	1219	M5.A.2.1.1	Numbers & Op.
604842	5	Grade 5 to Grade 6	1182	1219	M5.C.1.1.2	Geo.
606157	5	Grade 5 to Grade 6	1188	1219	M5.D.1.1.2	Alg. Con.
604794	5	Grade 5 to Grade 6	1177	1217	M5.E.2.1.2	Data & Prob.
604869	5	Grade 5 to Grade 6	1191	1216	M5.B.2.2.2	Measure.
606279	5	Grade 5 to Grade 6	1196	1219	M5.E.3.1.2	Data & Prob.
601040	6	Grade 5 to Grade 6	1190	609	M6.E.3.1.1	Data & Prob.
602096	6	Grade 5 to Grade 6	1190	1213	M6.B.2.1.1	Measure.
601730	6	Grade 5 to Grade 6	1191	1223	M6.B.2.2.1	Measure.
602081	6	Grade 5 to Grade 6	1188	1199	M6.E.1.1.3	Data & Prob.
599668	6	Grade 5 to Grade 6	1186	608	M6.A.1.3.1	Numbers & Op.
600989	6	Grade 5 to Grade 6	1184	1223	M6.D.1.1.1	Alg. Con.
602070	6	Grade 5 to Grade 6	1184	614	M6.E.1.1.1	Data & Prob.

Table C–3 (continued). Mathematics Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
601689	6	Grade 5 to Grade 6	1185	609	M6.C.1.2.2	Geo.
601031	6	Grade 5 to Grade 6	1185	1206	M6.D.2.1.2	Alg. Con.
602174	6	Grade 5 to Grade 6	1181	1210	M6.A.3.2.1	Numbers & Op.
601249	6	Grade 5 to Grade 6	1186	600	M6.C.3.1.1	Geo.
599670	6	Grade 5 to Grade 6	1181	1199	M6.A.1.3.2	Numbers & Op.
600978	6	Grade 5 to Grade 6	1184	615	M6.D.2.2.1	Alg. Con.
601706	6	Grade 5 to Grade 6	1186	1209	M6.E.2.1.1	Data & Prob.
601024	6	Grade 5 to Grade 6	1183	608	M6.D.1.2.1	Alg. Con.
602176	6	Grade 5 to Grade 6	1183	1213	M6.B.1.1.1	Measure.
602071	6	Grade 5 to Grade 6	1184	1210	M6.E.1.1.2	Data & Prob.
602104	6	Grade 5 to Grade 6	1179	607	M6.B.2.1.2	Measure.
599667	6	Grade 5 to Grade 6	1181	1226	M6.A.1.2.1	Numbers & Op.
601260	6	Grade 5 to Grade 6	1181	610	M6.C.1.1.1	Geo.

Table C–4. Mathematics Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
599606	6	Grade 6 to Grade 7	1224	792	M6.A.1.2.1	Numbers & Op.
601257	6	Grade 6 to Grade 7	1214	792	M6.C.3.1.1	Geo.
601026	6	Grade 6 to Grade 7	614	790	M6.D.1.2.1	Alg. Con.
601705	6	Grade 6 to Grade 7	1221	786	M6.E.1.1.1	Data & Prob.
601811	6	Grade 6 to Grade 7	1220	785	M6.A.2.1.1	Numbers & Op.
601714	6	Grade 6 to Grade 7	1203	786	M6.C.1.2.1	Geo.
601032	6	Grade 6 to Grade 7	1210	783	M6.D.2.1.2	Alg. Con.
599590	6	Grade 6 to Grade 7	2447	783	M6.A.1.1.1	Numbers & Op.
602095	6	Grade 6 to Grade 7	606	784	M6.B.2.1.3	Measure.
601700	6	Grade 6 to Grade 7	1230	785	M6.C.1.1.3	Geo.
601277	6	Grade 6 to Grade 7	1223	785	M6.E.3.1.1	Data & Prob.
602073	6	Grade 6 to Grade 7	603	784	M6.E.1.1.3	Data & Prob.
599643	6	Grade 6 to Grade 7	1217	778	M6.A.1.3.2	Numbers & Op.
602177	6	Grade 6 to Grade 7	1217	778	M6.B.1.1.1	Measure.
601220	6	Grade 6 to Grade 7	1205	778	M6.B.2.3.1	Measure.
601030	6	Grade 6 to Grade 7	1217	789	M6.D.2.1.1	Alg. Con.
601275	6	Grade 6 to Grade 7	592	786	M6.E.2.1.1	Data & Prob.
601678	6	Grade 6 to Grade 7	1220	785	M6.D.1.1.1	Alg. Con.
601301	6	Grade 6 to Grade 7	1220	785	M6.E.1.1.2	Data & Prob.
601245	6	Grade 6 to Grade 7	1225	783	M6.E.3.1.2	Data & Prob.
599593	6	Grade 6 to Grade 7	1221	784	M6.A.1.1.2	Numbers & Op.
601664	6	Grade 6 to Grade 7	600	780	M6.C.1.1.4	Geo.
599609	6	Grade 6 to Grade 7	1207	776	M6.A.1.3.1	Numbers & Op.
601799	6	Grade 6 to Grade 7	1211	778	M6.A.1.4.1	Numbers & Op.
602101	6	Grade 6 to Grade 7	612	775	M6.B.2.1.1	Measure.
602175	6	Grade 6 to Grade 7	614	773	M6.A.3.2.1	Numbers & Op.
601044	6	Grade 6 to Grade 7	1210	773	M6.D.2.2.1	Alg. Con.
601694	6	Grade 6 to Grade 7	1211	773	M6.C.1.1.2	Geo.
602088	6	Grade 6 to Grade 7	1226	772	M6.B.2.2.1	Measure.
601702	6	Grade 6 to Grade 7	605	771	M6.C.1.2.2	Geo.
601287	7	Grade 6 to Grade 7	1222	395	M7.D.2.1.1	Alg. Con.
601050	7	Grade 6 to Grade 7	1223	399	M7.E.2.1.1	Data & Prob.
601772	7	Grade 6 to Grade 7	1222	793	M7.D.1.1.1	Alg. Con.
602215	7	Grade 6 to Grade 7	1222	765	M7.B.2.1.3	Measure.
601132	7	Grade 6 to Grade 7	1221	764	M7.E.4.1.1	Data & Prob.
599720	7	Grade 6 to Grade 7	1221	757	M7.A.2.1.1	Numbers & Op.
602190	7	Grade 6 to Grade 7	1219	788	M7.B.1.1.1	Measure.

Table C–4 (continued). Mathematics Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
601273	7	Grade 6 to Grade 7	1215	762	M7.D.2.2.1	Alg. Con.
599734	7	Grade 6 to Grade 7	1215	792	M7.A.1.2.1	Numbers & Op.
601784	7	Grade 6 to Grade 7	1216	373	M7.C.1.1.2	Geo.
601278	7	Grade 6 to Grade 7	1213	401	M7.D.3.1.1	Alg. Con.
601704	7	Grade 6 to Grade 7	1214	788	M7.C.3.1.1	Geo.
602189	7	Grade 6 to Grade 7	1212	780	M7.A.3.2.2	Numbers & Op.
601123	7	Grade 6 to Grade 7	1209	385	M7.E.3.1.1	Data & Prob.
599633	7	Grade 6 to Grade 7	1209	797	M7.A.2.2.4	Numbers & Op.
601099	7	Grade 6 to Grade 7	1218	777	M7.E.1.1.1	Data & Prob.
599685	7	Grade 6 to Grade 7	1214	400	M7.A.2.2.2	Numbers & Op.
601124	7	Grade 6 to Grade 7	1216	785	M7.E.3.1.2	Data & Prob.
602193	7	Grade 6 to Grade 7	1214	792	M7.B.2.1.1	Measure.
601827	7	Grade 6 to Grade 7	1211	772	M7.C.1.1.3	Geo.
601067	7	Grade 6 to Grade 7	1208	781	M7.D.2.1.1	Alg. Con.
601379	7	Grade 6 to Grade 7	1212	793	M7.E.2.1.2	Data & Prob.
599708	7	Grade 6 to Grade 7	1206	563	M7.A.1.1.1	Numbers & Op.
601771	7	Grade 6 to Grade 7	1202	767	M7.D.1.1.1	Alg. Con.
601271	7	Grade 6 to Grade 7	1206	761	M7.D.2.2.1	Alg. Con.
599715	7	Grade 6 to Grade 7	1206	781	M7.A.1.2.2	Numbers & Op.
599650	7	Grade 6 to Grade 7	1193	798	M7.A.3.2.1	Numbers & Op.
602180	7	Grade 6 to Grade 7	1199	789	M7.B.1.1.1	Measure.
601355	7	Grade 6 to Grade 7	1190	399	M7.D.3.1.1	Alg. Con.
602202	7	Grade 6 to Grade 7	1194	795	M7.C.1.1.1	Geo.

Table C–5. Mathematics Items Used to Link Grade 8 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
601054	7	Grade 8 to Grade 7	745	312	M7.E.3.1.1	Data & Prob.
601365	7	Grade 8 to Grade 7	746	312	M7.D.3.1.1	Alg. Con.
601117	7	Grade 8 to Grade 7	747	311	M7.E.1.1.1	Data & Prob.
601835	7	Grade 8 to Grade 7	748	310	M7.C.1.1.3	Geo.
601677	7	Grade 8 to Grade 7	749	312	M7.C.1.2.2	Geo.
602155	7	Grade 8 to Grade 7	750	312	M7.A.3.2.2	Numbers & Op.
602142	7	Grade 8 to Grade 7	751	312	M7.B.2.1.3	Measure.
601300	7	Grade 8 to Grade 7	752	312	M7.D.2.1.2	Alg. Con.
601130	7	Grade 8 to Grade 7	753	312	M7.E.3.1.3	Data & Prob.
599682	7	Grade 8 to Grade 7	754	311	M7.A.2.2.1	Numbers & Op.
602144	7	Grade 8 to Grade 7	755	309	M7.B.2.2.2	Measure.
599732	7	Grade 8 to Grade 7	756	309	M7.A.2.2.6	Numbers & Op.
599727	7	Grade 8 to Grade 7	757	309	M7.A.1.2.1	Numbers & Op.
599686	7	Grade 8 to Grade 7	758	309	M7.A.2.2.3	Numbers & Op.
601687	7	Grade 8 to Grade 7	759	307	M7.C.3.1.2	Geo.
601218	7	Grade 8 to Grade 7	760	315	M7.C.3.1.1	Geo.
599722	7	Grade 8 to Grade 7	761	314	M7.A.2.1.1	Numbers & Op.
599684	7	Grade 8 to Grade 7	762	313	M7.A.2.2.2	Numbers & Op.
602141	7	Grade 8 to Grade 7	763	311	M7.B.2.1.2	Measure.
601051	7	Grade 8 to Grade 7	764	314	M7.E.2.1.2	Data & Prob.
599712	7	Grade 8 to Grade 7	765	314	M7.A.3.2.1	Numbers & Op.
602234	7	Grade 8 to Grade 7	766	314	M7.C.1.1.1	Geo.
602146	7	Grade 8 to Grade 7	767	314	M7.C.1.2.1	Geo.
601773	7	Grade 8 to Grade 7	768	313	M7.D.2.1.1	Alg. Con.
599711	7	Grade 8 to Grade 7	769	313	M7.A.2.2.5	Numbers & Op.
602143	7	Grade 8 to Grade 7	770	313	M7.B.2.2.1	Measure.
601110	7	Grade 8 to Grade 7	771	313	M7.E.3.1.2	Data & Prob.
601272	7	Grade 8 to Grade 7	772	312	M7.D.2.2.1	Alg. Con.
601357	7	Grade 8 to Grade 7	773	313	M7.D.3.1.2	Alg. Con.
601086	7	Grade 8 to Grade 7	774	313	M7.E.4.1.1	Data & Prob.
601263	8	Grade 8 to Grade 7	775	309	M8.C.3.1.1	Geo.
601757	8	Grade 8 to Grade 7	776	158	M8.D.1.1.2	Alg. Con.
601069	8	Grade 8 to Grade 7	777	308	M8.E.4.1.2	Data & Prob.
599651	8	Grade 8 to Grade 7	778	318	M8.A.3.1.2	Numbers & Op.
601073	8	Grade 8 to Grade 7	779	314	M8.D.2.1.3	Alg. Con.
601801	8	Grade 8 to Grade 7	780	154	M8.B.1.1.1	Measure.
599610	8	Grade 8 to Grade 7	781	160	M8.A.2.1.1	Numbers & Op.

Table C–5 (continued). Mathematics Items Used to Link Grade 8 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category
601097	8	Grade 8 to Grade 7	782	159	M8.E.1.1.1	Data & Prob.
601725	8	Grade 8 to Grade 7	783	316	M8.B.1.1.3	Measure.
601744	8	Grade 8 to Grade 7	784	157	M8.B.2.2.3	Measure.
601288	8	Grade 8 to Grade 7	785	157	M8.D.2.1.1	Alg. Con.
601247	8	Grade 8 to Grade 7	786	312	M8.D.2.2.2	Alg. Con.
599698	8	Grade 8 to Grade 7	787	156	M8.A.2.2.2	Numbers & Op.
601763	8	Grade 8 to Grade 7	788	306	M8.D.4.1.2	Alg. Con.
601090	8	Grade 8 to Grade 7	789	154	M8.E.1.1.3	Data & Prob.
601804	8	Grade 8 to Grade 7	790	318	M8.B.1.1.4	Measure.
599640	8	Grade 8 to Grade 7	791	311	M8.A.3.1.1	Numbers & Op.
602158	8	Grade 8 to Grade 7	792	310	M8.B.1.1.2	Measure.
602072	8	Grade 8 to Grade 7	793	315	M8.D.1.1.1	Alg. Con.
601707	8	Grade 8 to Grade 7	794	317	M8.D.1.1.3	Alg. Con.
601332	8	Grade 8 to Grade 7	795	312	M8.D.2.1.2	Alg. Con.
599613	8	Grade 8 to Grade 7	796	317	M8.A.2.2.1	Numbers & Op.
601675	8	Grade 8 to Grade 7	797	317	M8.D.4.1.3	Alg. Con.
601100	8	Grade 8 to Grade 7	798	157	M8.E.3.1.1	Data & Prob.
599583	8	Grade 8 to Grade 7	799	636	M8.A.1.1.1	Numbers & Op.
601340	8	Grade 8 to Grade 7	800	156	M8.D.2.2.1	Alg. Con.
601344	8	Grade 8 to Grade 7	801	321	M8.D.4.1.1	Alg. Con.
600990	8	Grade 8 to Grade 7	802	306	M8.E.1.1.2	Data & Prob.
599645	8	Grade 8 to Grade 7	803	160	M8.A.3.3.1	Numbers & Op.
602058	8	Grade 8 to Grade 7	804	307	M8.C.1.1.1	Geo.

Table C–6. Mathematics Items Used to Link Algebra I to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category	Algebra I Diagnostic Category
601121	8	Algebra I to Grade 8	316	1400	M8.A.3.3.1	Numbers & Op.	Op. with Real Num.
601102	8	Algebra I to Grade 8	310	1406	M8.E.3.1.1	Data & Prob.	Data Anal.
601360	8	Algebra I to Grade 8	155	1403	M8.D.4.1.1	Alg. Con.	Functions & Geo.
601764	8	Algebra I to Grade 8	316	1396	M8.D.4.1.3	Alg. Con.	Functions & Geo.
602052	8	Algebra I to Grade 8	318	1396	M8.D.1.1.3	Alg. Con.	Functions & Geo.
599639	8	Algebra I to Grade 8	154	1391	M8.A.3.1.1	Numbers & Op.	Op. with Real Num.
602065	8	Algebra I to Grade 8	156	1376	M8.D.1.1.1	Alg. Con.	Functions & Geo.
601346	8	Algebra I to Grade 8	306	1390	M8.D.2.2.2	Alg. Con.	Linear Eq.
599582	8	Algebra I to Grade 8	625	1387	M8.A.1.1.1	Numbers & Op.	Op. with Real Num.
599697	8	Algebra I to Grade 8	314	1377	M8.A.2.2.1	Numbers & Op.	Op. with Real Num.
600980	8	Algebra I to Grade 8	318	1376	M8.D.2.1.3	Alg. Con.	Linear Eq.
601127	8	Algebra I to Grade 8	158	1376	M8.E.4.1.1	Data & Prob.	Data Anal.
601776	8	Algebra I to Grade 8	311	1370	M8.D.4.1.2	Alg. Con.	Functions & Geo.
601092	8	Algebra I to Grade 8	306	1362	M8.E.1.1.2	Data & Prob.	Data Anal.
601232	8	Algebra I to Grade 8	151	1359	M8.D.2.1.1	Alg. Con.	Linear Eq.
601348	8	Algebra I to Grade 8	311	1402	M8.D.2.2.1	Alg. Con.	Linear Eq.
601777	8	Algebra I to Grade 8	307	1401	M8.D.4.1.3	Alg. Con.	Functions & Geo.
599619	8	Algebra I to Grade 8	314	1388	M8.A.2.2.2	Numbers & Op.	Op. with Real Num.
601222	8	Algebra I to Grade 8	311	1389	M8.C.3.1.1	Geo.	None
601384	8	Algebra I to Grade 8	317	1388	M8.D.4.1.1	Alg. Con.	Functions & Geo.
601091	8	Algebra I to Grade 8	314	1390	M8.E.1.1.3	Data & Prob.	Data Anal.
599585	8	Algebra I to Grade 8	310	1377	M8.A.2.1.1	Numbers & Op.	Op. with Real Num.
599637	8	Algebra I to Grade 8	308	1380	M8.A.3.1.2	Numbers & Op.	Op. with Real Num.
601231	8	Algebra I to Grade 8	313	1374	M8.D.2.1.1	Alg. Con.	Linear Eq.
601663	8	Algebra I to Grade 8	155	1368	M8.D.1.1.2	Alg. Con.	Functions & Geo.
601126	8	Algebra I to Grade 8	308	1370	M8.E.4.1.2	Data & Prob.	Data Anal.
601089	8	Algebra I to Grade 8	151	1357	M8.E.1.1.2	Data & Prob.	Data Anal.
601234	8	Algebra I to Grade 8	303	1356	M8.D.2.1.2	Alg. Con.	Linear Eq.
601775	8	Algebra I to Grade 8	312	1349	M8.D.4.1.2	Alg. Con.	Functions & Geo.
601103	8	Algebra I to Grade 8	319	1344	M8.E.3.2.1	Data & Prob.	Data Anal.
602259	11	Algebra I to Grade 8	312	714	M11.E.2.1.3	Data & Prob.	Data Anal.
604952	11	Algebra I to Grade 8	312	710	M11.E.4.1.2	Data & Prob.	Data Anal.
601837	A1	Algebra I to Grade 8	312	700	A1.2.2.1.1	Alg. Con.	Functions & Geo.
602184	A1	Algebra I to Grade 8	313	1421	A1.2.1.1.1	Alg. Con.	Functions & Geo.
601554	11	Algebra I to Grade 8	313	711	M11.E.2.1.3	Data & Prob.	Data Anal.
602171	A1	Algebra I to Grade 8	309	1382	A1.2.1.2.2	Alg. Con.	Functions & Geo.
601841	A1	Algebra I to Grade 8	313	1383	A1.2.2.1.2	Alg. Con.	Functions & Geo.

Table C–6 (continued). Mathematics Items Used to Link Algebra I to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category	Algebra I Diagnostic Category
604806	11	Algebra I to Grade 8	312	710	M11.E.4.1.2	Data & Prob.	Data Anal.
600839	11	Algebra I to Grade 8	313	713	M11.E.1.1.1	Data & Prob.	Data Anal.
601461	11	Algebra I to Grade 8	313	711	M11.E.1.1.1	Data & Prob.	Data Anal.
604804	11	Algebra I to Grade 8	313	705	M11.E.2.1.3	Data & Prob.	Data Anal.
602241	A1	Algebra I to Grade 8	312	1420	A1.2.1.2.1	Alg. Con.	Functions & Geo.
601793	A1	Algebra I to Grade 8	313	1425	A1.2.2.1.4	Alg. Con.	Functions & Geo.
602159	A1	Algebra I to Grade 8	312	1416	A1.2.2.2.1	Alg. Con.	Functions & Geo.
602274	11	Algebra I to Grade 8	312	713	M11.E.4.1.2	Data & Prob.	Data Anal.
601135	A1	Algebra I to Grade 8	315	1418	A1.2.3.3.1	Data & Prob.	Data Anal.
601144	A1	Algebra I to Grade 8	317	1415	A1.1.2.1.3	Alg. Con.	Linear Eq.
600842	11	Algebra I to Grade 8	316	717	M11.A.2.1.3	Numbers & Op.	Op. with Real Num.
601370	A1	Algebra I to Grade 8	314	1364	A1.1.3.1.3	Alg. Con.	Linear Eq.
600646	11	Algebra I to Grade 8	315	710	M11.A.3.1.1	Numbers & Op.	Op. with Real Num.
601630	11	Algebra I to Grade 8	314	718	M11.A.3.1.1	Numbers & Op.	Op. with Real Num.
601138	A1	Algebra I to Grade 8	313	1378	A1.2.3.2.1	Data & Prob.	Data Anal.
601139	A1	Algebra I to Grade 8	310	1413	A1.2.3.2.2	Data & Prob.	Data Anal.
600826	11	Algebra I to Grade 8	311	716	M11.A.3.1.1	Numbers & Op.	Op. with Real Num.
601140	A1	Algebra I to Grade 8	310	1408	A1.2.3.2.3	Data & Prob.	Data Anal.
600930	A1	Algebra I to Grade 8	311	707	A1.1.1.4.1	Numbers & Op.	Op. with Real Num.
602260	11	Algebra I to Grade 8	312	717	M11.A.2.1.1	Numbers & Op.	Op. with Real Num.
600931	A1	Algebra I to Grade 8	310	1375	A1.1.1.5.1	Alg. Con.	Op. with Real Num.
602644	11	Algebra I to Grade 8	311	714	M11.A.2.1.1	Numbers & Op.	Op. with Real Num.
604162	11	Algebra I to Grade 8	310	714	M11.A.2.1.2	Numbers & Op.	Op. with Real Num.

Table C–7. Mathematics Items Used to Link Geometry to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category	Geometry Diagnostic Category
601740	8	Geometry to Grade 8	306	1052	M8.B.2.1.3	Measure.	Measure.
602118	8	Geometry to Grade 8	319	1049	M8.B.2.2.1	Measure.	Measure.
602056	8	Geometry to Grade 8	306	1052	M8.C.1.1.2	Geo.	Geo. Prop.
602059	8	Geometry to Grade 8	156	1052	M8.C.1.1.2	Geo.	Geo. Prop.
601733	8	Geometry to Grade 8	151	1039	M8.B.2.1.1	Measure.	Measure.
602133	8	Geometry to Grade 8	320	1049	M8.C.1.1.3	Geo.	Geo. Prop.
602117	8	Geometry to Grade 8	151	1046	M8.B.2.2.2	Measure.	Measure.
602128	8	Geometry to Grade 8	312	1047	M8.C.1.1.1	Geo.	Geo. Prop.
601802	8	Geometry to Grade 8	319	1047	M8.B.1.1.3	Measure.	None
602205	8	Geometry to Grade 8	318	1047	M8.C.1.1.1	Geo.	Geo. Prop.
601723	8	Geometry to Grade 8	306	1037	M8.B.1.1.1	Measure.	None
602208	8	Geometry to Grade 8	317	1043	M8.C.1.1.3	Geo.	Geo. Prop.
601326	8	Geometry to Grade 8	317	1038	M8.C.1.2.1	Geo.	Coor. Geo.
601338	8	Geometry to Grade 8	311	1038	M8.C.3.1.1	Geo.	Coor. Geo.
601371	8	Geometry to Grade 8	316	1031	M8.C.3.1.1	Geo.	Coor. Geo.
601736	8	Geometry to Grade 8	316	1048	M8.B.2.1.2	Measure.	Measure.
602136	8	Geometry to Grade 8	316	1034	M8.C.1.2.1	Geo.	Coor. Geo.
601755	8	Geometry to Grade 8	306	1039	M8.C.1.2.1	Geo.	Coor. Geo.
601372	8	Geometry to Grade 8	316	1037	M8.C.3.1.1	Geo.	Coor. Geo.
601782	8	Geometry to Grade 8	156	1028	M8.B.1.1.4	Measure.	None
602204	8	Geometry to Grade 8	308	1039	M8.C.1.1.1	Geo.	Geo. Prop.
602131	8	Geometry to Grade 8	317	1037	M8.C.1.1.2	Geo.	Geo. Prop.
602061	8	Geometry to Grade 8	314	1035	M8.C.1.1.2	Geo.	Geo. Prop.
602115	8	Geometry to Grade 8	317	1029	M8.B.2.2.2	Measure.	Measure.
602087	8	Geometry to Grade 8	312	1034	M8.C.1.1.3	Geo.	Geo. Prop.
602212	8	Geometry to Grade 8	319	1030	M8.C.1.1.3	Geo.	Geo. Prop.
601724	8	Geometry to Grade 8	310	1023	M8.B.1.1.2	Measure.	None
602113	8	Geometry to Grade 8	315	1023	M8.B.2.2.1	Measure.	Measure.
601329	8	Geometry to Grade 8	302	1031	M8.C.3.1.1	Geo.	Coor. Geo.
601743	8	Geometry to Grade 8	305	1029	M8.B.2.2.3	Measure.	Measure.
602661	11	Geometry to Grade 8	316	531	M11.B.2.1.1	Measure.	Measure.
604163	11	Geometry to Grade 8	317	531	M11.B.2.2.2	Measure.	Measure.
604671	GE	Geometry to Grade 8	311	1963	G.1.1.1.1	Geo.	Geo. Prop.
604400	GE	Geometry to Grade 8	316	992	G.1.3.1.1	Geo.	Congruence
604389	GE	Geometry to Grade 8	316	1001	G.2.1.1.1	Geo.	Coor. Geo.
604799	11	Geometry to Grade 8	316	528	M11.B.2.3.1	Measure.	Measure.
604418	GE	Geometry to Grade 8	312	478	G.1.2.1.4	Geo.	Geo. Prop.

Table C–7 (continued). Mathematics Items Used to Link Geometry to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Mathematics Diagnostic Category	Geometry Diagnostic Category
600651	11	Geometry to Grade 8	315	531	M11.B.2.2.4	Measure.	Measure.
604707	GE	Geometry to Grade 8	314	1053	G.1.2.1.5	Geo.	Geo. Prop.
604180	11	Geometry to Grade 8	316	528	M11.B.2.2.3	Measure.	Measure.
604378	GE	Geometry to Grade 8	316	1048	G.2.2.1.1	Geo.	Measure.
601544	11	Geometry to Grade 8	316	532	M11.B.2.1.1	Measure.	Measure.
600749	11	Geometry to Grade 8	314	531	M11.B.2.2.4	Measure.	Measure.
604392	GE	Geometry to Grade 8	315	1053	G.1.1.1.4	Geo.	Geo. Prop.
604395	GE	Geometry to Grade 8	314	1024	G.1.3.1.2	Geo.	Congruence
604178	11	Geometry to Grade 8	315	531	M11.C.1.3.1	Geo.	Congruence
600785	11	Geometry to Grade 8	315	530	M11.C.1.2.2	Geo.	Geo. Prop.
604522	11	Geometry to Grade 8	313	533	M11.C.1.4.1	Geo.	Coor. Geo.
604763	GE	Geometry to Grade 8	308	503	G.2.2.2.1	Geo.	Measure.
602650	11	Geometry to Grade 8	313	530	M11.C.1.3.1	Geo.	Congruence
604474	GE	Geometry to Grade 8	313	988	G.2.2.1.2	Geo.	Measure.
604600	GE	Geometry to Grade 8	310	1053	G.2.2.2.4	Geo.	Measure.
604361	GE	Geometry to Grade 8	312	525	G.2.3.2.1	Geo.	Measure.
601550	11	Geometry to Grade 8	311	530	M11.C.1.2.3	Geo.	Geo. Prop.
604360	GE	Geometry to Grade 8	309	1042	G.2.3.1.3	Geo.	Measure.
604170	11	Geometry to Grade 8	309	528	M11.C.1.4.1	Geo.	Coor. Geo.
604354	GE	Geometry to Grade 8	306	1007	G.2.2.3.1	Geo.	Measure.
601549	11	Geometry to Grade 8	306	530	M11.C.1.2.3	Geo.	Geo. Prop.
602268	11	Geometry to Grade 8	305	527	M11.C.1.3.1	Geo.	Congruence
604453	GE	Geometry to Grade 8	304	955	G.2.2.2.2	Geo.	Measure.

Table C–8. Mathematics Items Used to Link Algebra II to Algebra I

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Algebra I Diagnostic Category	Algebra II Diagnostic Category
602167	A1	Algebra II to Algebra I	701	949	A1.1.3.2.1	Linear Eq.	Non-linear
601423	A1	Algebra II to Algebra I	709	951	A1.1.2.1.3	Linear Eq.	Non-linear
602188	A1	Algebra II to Algebra I	708	943	A1.2.2.1.4	Functions & Geo.	Functions
600971	A1	Algebra II to Algebra I	1407	944	A1.1.1.5.1	Op. with Real Num.	Non-linear
601180	A1	Algebra II to Algebra I	1372	948	A1.1.2.1.1	Linear Eq.	Non-linear
601854	A1	Algebra II to Algebra I	670	937	A1.1.2.2.2	Linear Eq.	Non-linear
602253	A1	Algebra II to Algebra I	705	939	A1.2.2.1.2	Functions & Geo.	Functions
601419	A1	Algebra II to Algebra I	693	941	A1.1.3.1.2	Linear Eq.	Non-linear
602251	A1	Algebra II to Algebra I	1371	942	A1.2.1.2.2	Functions & Geo.	Functions
601176	A1	Algebra II to Algebra I	676	941	A1.2.3.2.3	Data Anal.	Data Anal.
600928	A1	Algebra II to Algebra I	1405	935	A1.1.1.2.1	Op. with Real Num.	Non-linear
600926	A1	Algebra II to Algebra I	2816	940	A1.1.1.1.1	Op. with Real Num.	Non-linear
602237	A1	Algebra II to Algebra I	662	931	A1.2.1.1.1	Functions & Geo.	Functions
601394	A1	Algebra II to Algebra I	697	931	A1.2.1.1.3	Functions & Geo.	Functions
600973	A1	Algebra II to Algebra I	682	925	A1.1.1.5.3	Op. with Real Num.	Non-linear
601397	A1	Algebra II to Algebra I	1378	943	A1.1.3.1.1	Linear Eq.	Non-linear
601368	A1	Algebra II to Algebra I	1374	948	A1.1.3.1.3	Linear Eq.	Non-linear
601136	A1	Algebra II to Algebra I	709	942	A1.1.2.1.2	Linear Eq.	Non-linear
601836	A1	Algebra II to Algebra I	713	946	A1.2.2.1.1	Functions & Geo.	Functions
601148	A1	Algebra II to Algebra I	1395	942	A1.2.3.3.1	Data Anal.	Data Anal.
602160	A1	Algebra II to Algebra I	1397	947	A1.2.2.2.1	Functions & Geo.	Functions
601813	A1	Algebra II to Algebra I	1424	941	A1.2.1.2.1	Functions & Geo.	Functions
601805	A1	Algebra II to Algebra I	1348	920	A1.2.2.1.3	Functions & Geo.	Functions
600953	A1	Algebra II to Algebra I	659	940	A1.1.1.1.2	Op. with Real Num.	Non-linear
600932	A1	Algebra II to Algebra I	1411	941	A1.1.1.5.2	Op. with Real Num.	Non-linear
601398	A1	Algebra II to Algebra I	1410	931	A1.1.2.2.1	Linear Eq.	Non-linear
600948	A1	Algebra II to Algebra I	1387	920	A1.2.3.1.1	Data Anal.	Data Anal.
600966	A1	Algebra II to Algebra I	1395	912	A1.1.1.3.1	Op. with Real Num.	Non-linear
602154	A1	Algebra II to Algebra I	1387	918	A1.1.3.2.2	Linear Eq.	Non-linear
601380	A1	Algebra II to Algebra I	1392	915	A1.2.1.1.2	Functions & Geo.	Functions
604700	A2	Algebra II to Algebra I	1406	927	A2.2.1.1.1	Functions & Geo.	Functions
603013	A2	Algebra II to Algebra I	1406	957	A2.1.3.1.4	Linear Eq.	Non-linear
604570	A2	Algebra II to Algebra I	1386	462	A2.2.2.1.3	Functions & Geo.	Functions
603086	A2	Algebra II to Algebra I	1400	914	A2.1.2.1.4	Op. with Real Num.	Non-linear
604625	A2	Algebra II to Algebra I	1380	948	A2.2.1.1.3	Functions & Geo.	Functions
604530	A2	Algebra II to Algebra I	1380	935	A2.1.3.2.2	Linear Eq.	Non-linear
604686	A2	Algebra II to Algebra I	1379	446	A2.2.2.2.1	Functions & Geo.	Functions

Table C–8 (continued). Mathematics Items Used to Link Algebra II to Algebra I

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Algebra I Diagnostic Category	Algebra II Diagnostic Category
603043	A2	Algebra II to Algebra I	1383	932	A2.1.2.1.2	Op. with Real Num.	Non-linear
603037	A2	Algebra II to Algebra I	1366	950	A2.2.1.1.4	Functions & Geo.	Functions
604572	A2	Algebra II to Algebra I	1377	453	A2.2.2.1.4	Functions & Geo.	Functions
603000	A2	Algebra II to Algebra I	1372	471	A2.1.2.2.2	Op. with Real Num.	Non-linear
604537	A2	Algebra II to Algebra I	1373	908	A2.2.1.1.2	Functions & Geo.	Functions
604634	A2	Algebra II to Algebra I	1369	472	A2.2.3.2.3	Data Anal.	Data Anal.
603106	A2	Algebra II to Algebra I	1360	898	A2.2.3.1.2	Data Anal.	Data Anal.
603057	A2	Algebra II to Algebra I	1351	456	A2.2.3.2.1	Data Anal.	Data Anal.
603055	A2	Algebra II to Algebra I	1397	919	A2.2.3.1.1	Data Anal.	Data Anal.
603018	A2	Algebra II to Algebra I	1408	937	A2.1.2.2.1	Op. with Real Num.	Non-linear
604685	A2	Algebra II to Algebra I	1404	476	A2.2.2.2.1	Functions & Geo.	Functions
603126	A2	Algebra II to Algebra I	1396	474	A2.2.3.2.3	Data Anal.	Data Anal.
604539	A2	Algebra II to Algebra I	1395	941	A2.1.3.2.1	Linear Eq.	Non-linear
604540	A2	Algebra II to Algebra I	1382	889	A2.1.3.2.2	Linear Eq.	Non-linear
604703	A2	Algebra II to Algebra I	1397	479	A2.2.1.1.1	Functions & Geo.	Functions
604629	A2	Algebra II to Algebra I	1387	902	A2.2.2.1.1	Functions & Geo.	Functions
603056	A2	Algebra II to Algebra I	1390	928	A2.2.3.2.1	Data Anal.	Data Anal.
603003	A2	Algebra II to Algebra I	1376	473	A2.1.3.1.2	Linear Eq.	Non-linear
604550	A2	Algebra II to Algebra I	1369	939	A2.2.2.1.4	Functions & Geo.	Functions
603098	A2	Algebra II to Algebra I	1374	944	A2.1.2.1.3	Op. with Real Num.	Non-linear
604544	A2	Algebra II to Algebra I	1370	461	A2.2.1.1.2	Functions & Geo.	Functions
604627	A2	Algebra II to Algebra I	1363	953	A2.2.1.1.3	Functions & Geo.	Functions
603042	A2	Algebra II to Algebra I	1368	936	A2.1.2.1.1	Op. with Real Num.	Non-linear

Tables C–9 through C–16 summarize the number of linking items by diagnostic category. Items coded in a Mathematics diagnostic category and an Algebra I, Geometry, or Algebra II diagnostic category are noted.

Table C–9. Number of Items Linking Grade 3 to Grade 4 by Diagnostic Category

Diagnostic Category	Grade 3 Items	Grade 4 Items	Total
Numbers & Operations	8	6	14
Measurement	5	4	9
Geometry	2	4	6
Algebraic Concepts	3	3	6
Data Analysis & Probability	2	3	5
TOTAL	20	20	40

Table C–10. Number of Items Linking Grade 4 to Grade 5 by Diagnostic Category

Diagnostic Category	Grade 4 Items	Grade 5 Items	Total
Numbers & Operations	6	7	13
Measurement	3	4	7
Geometry	4	3	7
Algebraic Concepts	3	3	6
Data Analysis & Probability	4	3	7
TOTAL	20	20	40

Table C–11. Number of Items Linking Grade 5 to Grade 6 by Diagnostic Category

Diagnostic Category	Grade 5 Items	Grade 6 Items	Total
Numbers & Operations	10	4	14
Measurement	6	4	10
Geometry	5	3	8
Algebraic Concepts	4	4	8
Data Analysis & Probability	5	5	10
TOTAL	30	20	50

Table C–12. Number of Items Linking Grade 6 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 6 Items	Grade 7 Items	Total
Numbers & Operations	8	8	16
Measurement	5	4	9
Geometry	6	4	10
Algebraic Concepts	5	8	13
Data Analysis & Probability	6	6	12
TOTAL	30	30	60

Table C–13. Number of Items Linking Grade 8 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 7 Items	Grade 8 Items	Total
Numbers & Operations	9	7	16
Measurement	4	5	9
Geometry	6	2	8
Algebraic Concepts	5	11	16
Data Analysis & Probability	6	5	11
TOTAL	30	30	60

Table C–14a. Number of Items Linking Algebra I to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Algebra I Items	Total
Numbers & Operations	7	8	15
Measurement	0	0	0
Geometry	1	0	1
Algebraic Concepts	15	10	25
Data Analysis & Probability	7	12	19
No Grade 8 DC	0	0	0
TOTAL	30	30	60

Table C–14b. Number of Items Linking Algebra I to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Algebra I Items	Total
Operations with Real Numbers	7	9	16
Linear Equations	6	2	8
Functions	9	7	16
Data Analysis	7	12	19
No Algebra I DC	1	0	1
TOTAL	30	30	60

Table C–15a. Number of Items Linking Geometry to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Geometry Items	Total
Numbers & Operations	0	0	0
Measurement	12	0	12
Geometry	18	30	48
Algebraic Concepts	0	0	0
Data Analysis & Probability	0	0	0
No Grade 8 DC	0	0	0
TOTAL	30	30	60

Table C–15b. Number of Items Linking Geometry to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Geometry Items	Total
Geometric Properties	11	8	19
Congruence	0	4	4
Coordinate	7	2	9
Measurement	8	16	24
No Geometry DC	4	0	4
TOTAL	30	30	60

Table C–16a. Number of Items Linking Algebra II to Algebra I by Diagnostic Category

Diagnostic Category	Algebra I Items	Algebra II Items	Total
Operations with Real Numbers	7	6	13
Linear Equations	10	5	15
Functions	10	13	23
Data Analysis	3	6	9
No Algebra I DC	0	0	0
TOTAL	30	30	60

Table C–16b. Number of Items Linking Algebra II to Algebra I by Diagnostic Category

Diagnostic Category	Algebra I Items	Algebra II Items	Total
Op. with Complex Numbers	0	0	0
Non-linear	17	11	28
Functions	10	13	23
Data Analysis	3	6	9
No Algebra II DC	0	0	0
TOTAL	30	30	60

Table C–17. Mathematics Example of Vertical Linking Workbook

		Grade 4 Calibration			Grade 5 Calibration				Grade 4 on		
Item ID	Item Grade	Difficulty	Fit	Displace	Difficulty	Fit	Displace	Discrepancy	Grade 5 Scale	Robust Z	Flag
601646	4	-1.028	1.020	-0.006	-1.880	1.000	-0.004	-0.852	-1.650	-0.458	
601987	4	0.195	0.970	0.001	-0.384	0.930	0.000	-0.579	-0.427	0.205	
604493	4	0.784	1.030	0.000	0.204	1.010	0.000	-0.580	0.162	0.203	
601961	4	0.684	1.000	0.002	-0.469	0.910	0.000	-1.153	0.062	-1.189	
604499	4	-0.488	0.900	0.001	-0.492	0.910	0.000	-0.004	-1.110	1.601	
602889	4	-0.160	0.920	-0.002	-1.157	0.840	0.000	-0.997	-0.782	-0.810	
602885	4	0.112	1.200	0.003	0.051	1.220	0.000	-0.061	-0.510	1.463	
602887	4	-0.493	1.070	-0.002	-1.063	1.030	0.000	-0.570	-1.115	0.227	
601639	4	0.397	1.070	0.001	0.149	1.090	0.000	-0.248	-0.225	1.009	
604969	4	1.559	1.060	0.000	1.469	1.080	0.000	-0.090	0.937	1.393	
601994	4	0.257	0.950	0.000	0.100	1.090	0.000	-0.157	-0.365	1.230	
601998	4	-0.551	1.120	-0.001	-1.376	1.140	-0.004	-0.825	-1.173	-0.392	
602000	4	2.034	1.070	-0.006	1.248	1.060	-0.003	-0.786	1.412	-0.297	
601991	4	1.106	0.900	0.001	0.095	0.860	-0.003	-1.011	0.484	-0.844	
604879	4	-0.099	1.020	0.000	-1.101	0.870	-0.003	-1.002	-0.721	-0.822	
601964	4	1.069	1.020	0.001	0.154	1.010	-0.003	-0.915	0.447	-0.611	
602971	4	-0.355	1.000	0.000	-0.858	1.070	-0.003	-0.503	-0.977	0.390	
604486	4	-0.420	0.940	0.000	-0.749	0.970	-0.003	-0.329	-1.042	0.812	
604967	4	-1.495	0.900	0.001	-1.254	0.960	-0.003	0.241	-2.117	2.196	high robust Z
602973	4	-0.035	0.940	0.003	0.362	1.220	-0.003	0.397	-0.657	2.575	high robust Z
600853	5	0.883	1.100	0.004	-0.047	1.100	-0.003	-0.930	0.261	-0.647	
604790	5	-0.495	1.010	0.004	-1.082	0.970	0.000	-0.587	-1.117	0.186	
604956	5	1.299	0.870	0.004	0.590	0.820	-0.003	-0.709	0.677	-0.110	
604862	5	1.405	0.920	0.004	0.368	0.850	-0.003	-1.037	0.783	-0.907	
604783	5	0.764	0.970	0.004	-0.814	0.890	0.001	-1.578	0.142	-2.221	high robust Z
606159	5	0.793	1.090	0.004	-0.157	0.990	-0.003	-0.950	0.171	-0.696	
604848	5	0.301	0.910	0.004	-0.707	1.020	0.001	-1.008	-0.321	-0.837	
604843	5	1.481	1.050	0.004	0.819	0.940	0.001	-0.662	0.859	0.004	
604966	5	-1.974	0.920	0.004	-3.190	0.870	-0.005	-1.216	-2.596	-1.342	
606163	5	0.780	1.130	0.004	0.478	1.200	0.002	-0.302	0.158	0.878	
601532	5	-0.368	0.950	0.000	-1.033	0.920	-0.001	-0.665	-0.990	-0.004	
606160	5	0.382	1.070	0.000	-0.313	0.940	-0.005	-0.695	-0.240	-0.076	
604960	5	0.618	0.910	0.000	0.223	1.050	0.000	-0.395	-0.004	0.652	
600852	5	0.753	1.100	0.000	0.050	1.020	0.002	-0.703	0.131	-0.096	
604834	5	-0.673	0.980	0.000	-1.151	0.980	-0.004	-0.478	-1.295	0.450	
604959	5	0.012	0.880	0.000	-0.871	0.840	-0.001	-0.883	-0.610	-0.533	
604961	5	0.141	1.000	0.000	-0.319	1.010	0.002	-0.460	-0.481	0.494	
606278	5	1.197	1.000	0.000	0.700	0.960	0.001	-0.497	0.575	0.404	
604965	5	-1.454	0.890	0.000	-1.565	0.900	-0.005	-0.111	-2.076	1.342	
604865	5	0.454	0.930	0.000	-0.537	0.910	-0.001	-0.991	-0.168	-0.795	
	Mean	0.234			-0.388			-0.622	-0.388	0.101	
	SD	0.887			0.893			0.413	0.887	1.002	
	SD Ratio	0.993									
	Correlation	0.892									
	Add. Constant	-0.622									
	Median							-0.664			
	Q							0.557			

Figures C–1 through C–8 are the adjacent grade linking plots. Items removed from final linking procedure are colored red.

Figure C–1. CDT Mathematics: Grade 3 to Grade 4 Linking – All Links

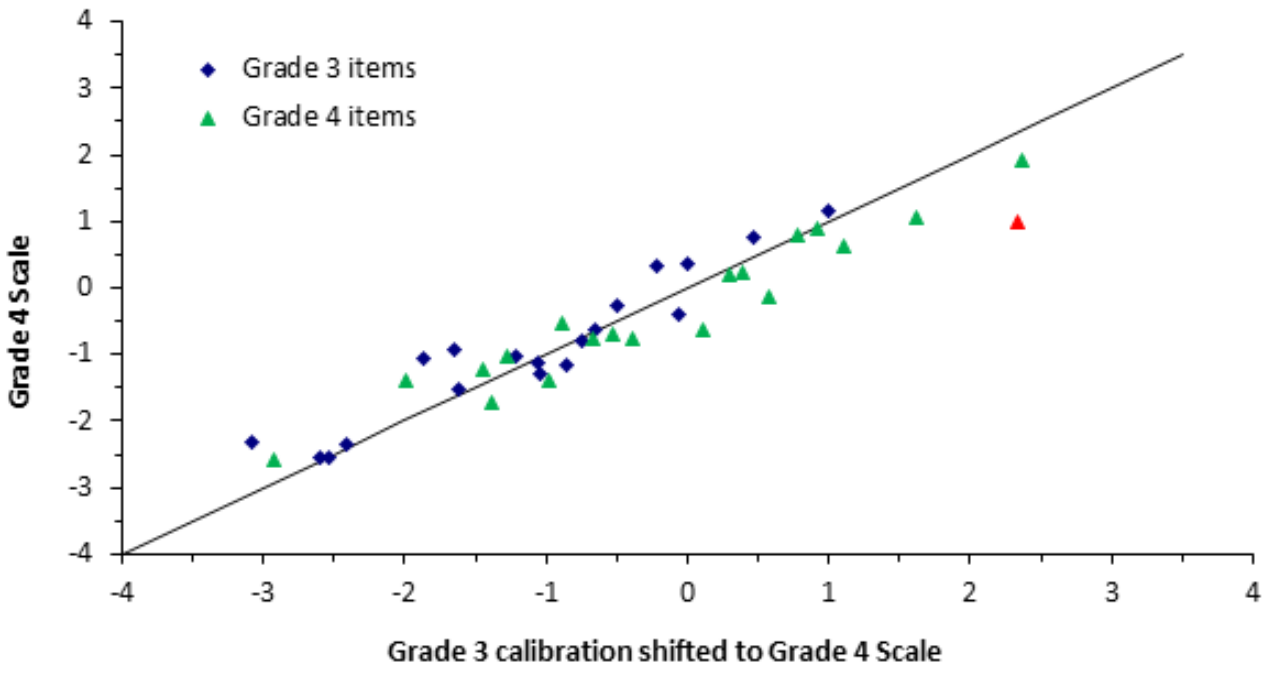


Figure C–2. CDT Mathematics: Grade 4 to Grade 5 Linking – All Links

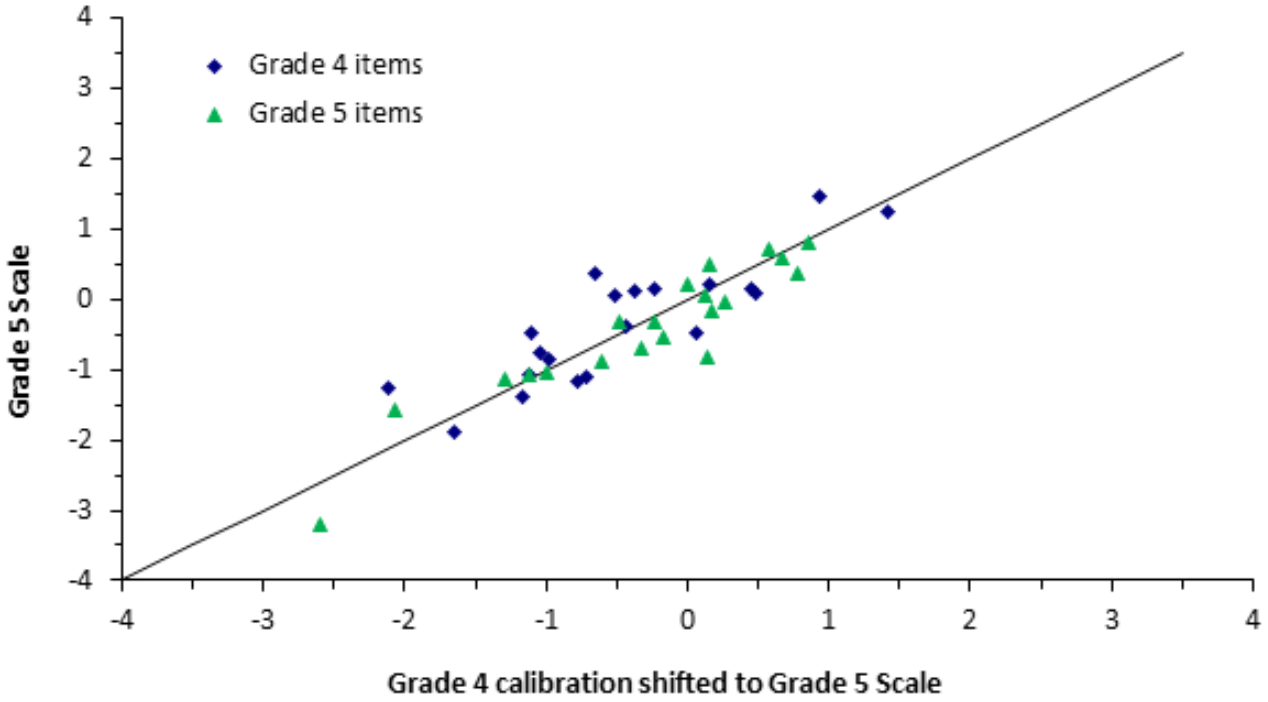


Figure C–3. CDT Mathematics: Grade 5 to Grade 6 Linking – All Links

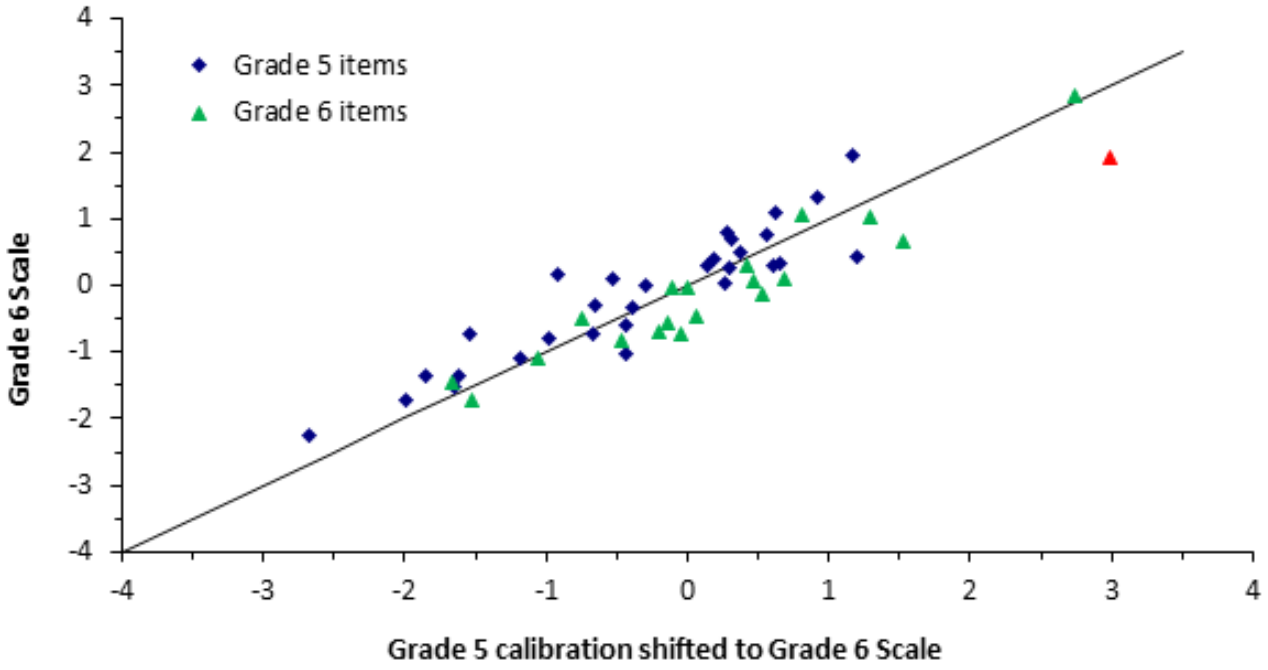


Figure C–4. CDT Mathematics: Grade 6 to Grade 7 Linking – All Links

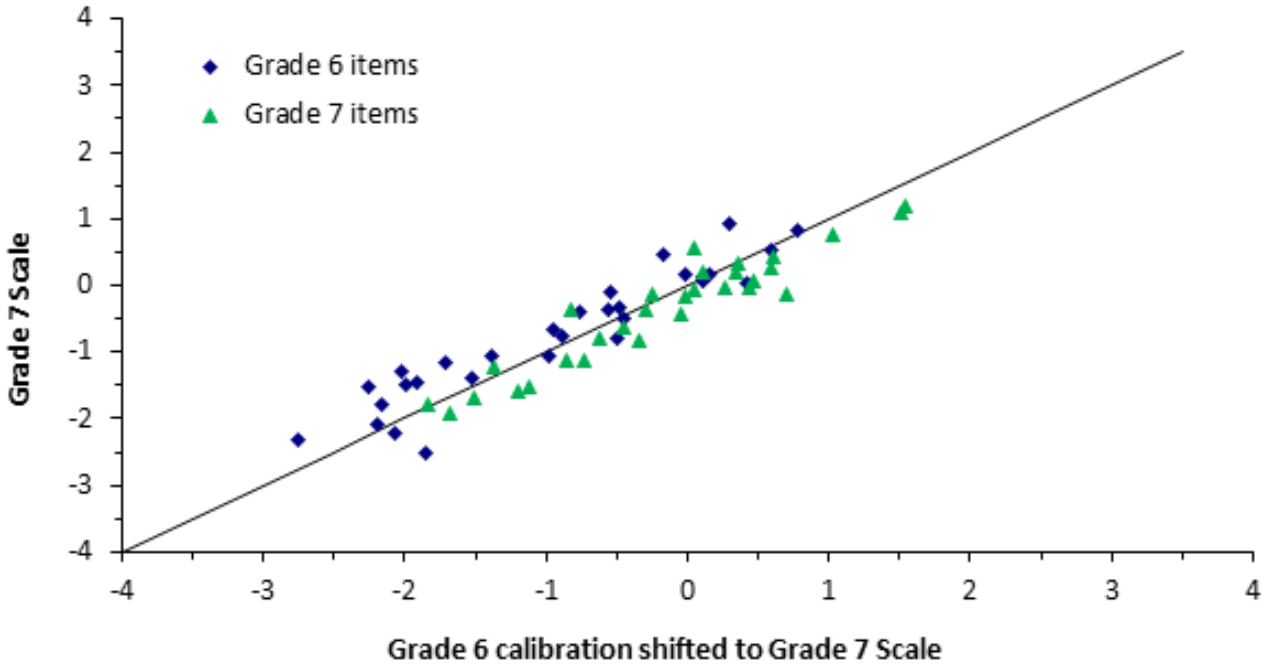


Figure C–5. CDT Mathematics: Grade 8 to Grade 7 Linking – All Links

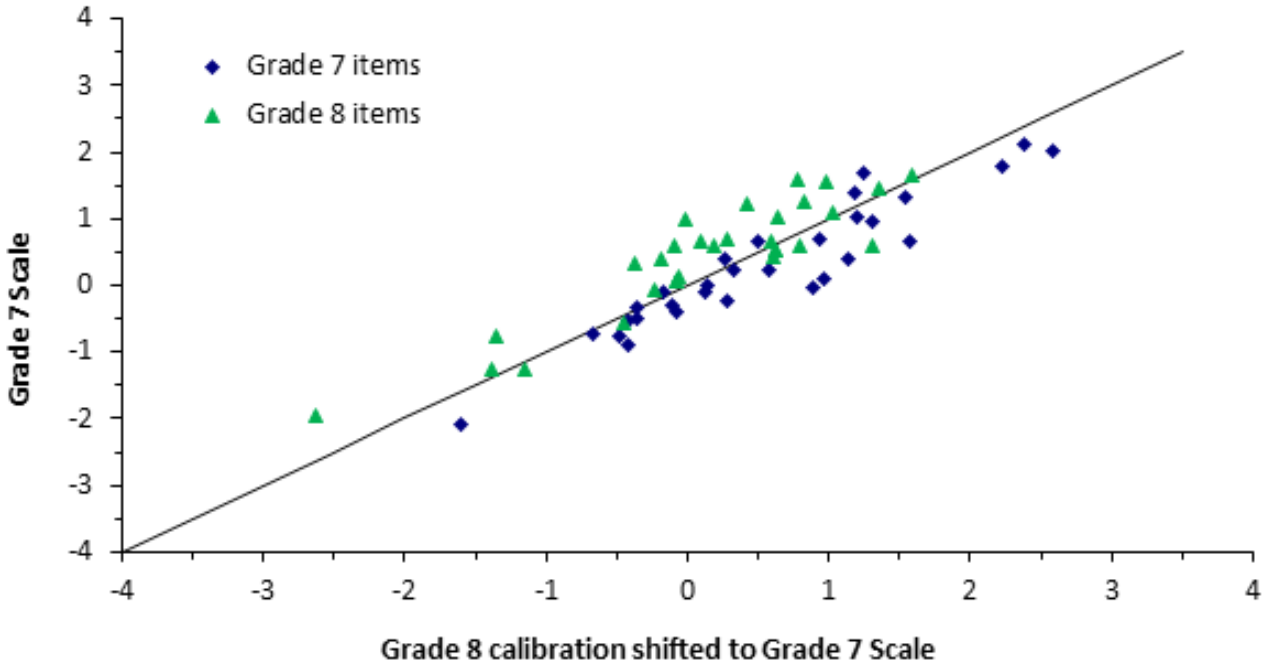


Figure C–6. CDT Mathematics: Algebra I to Grade 8 Linking – All Links

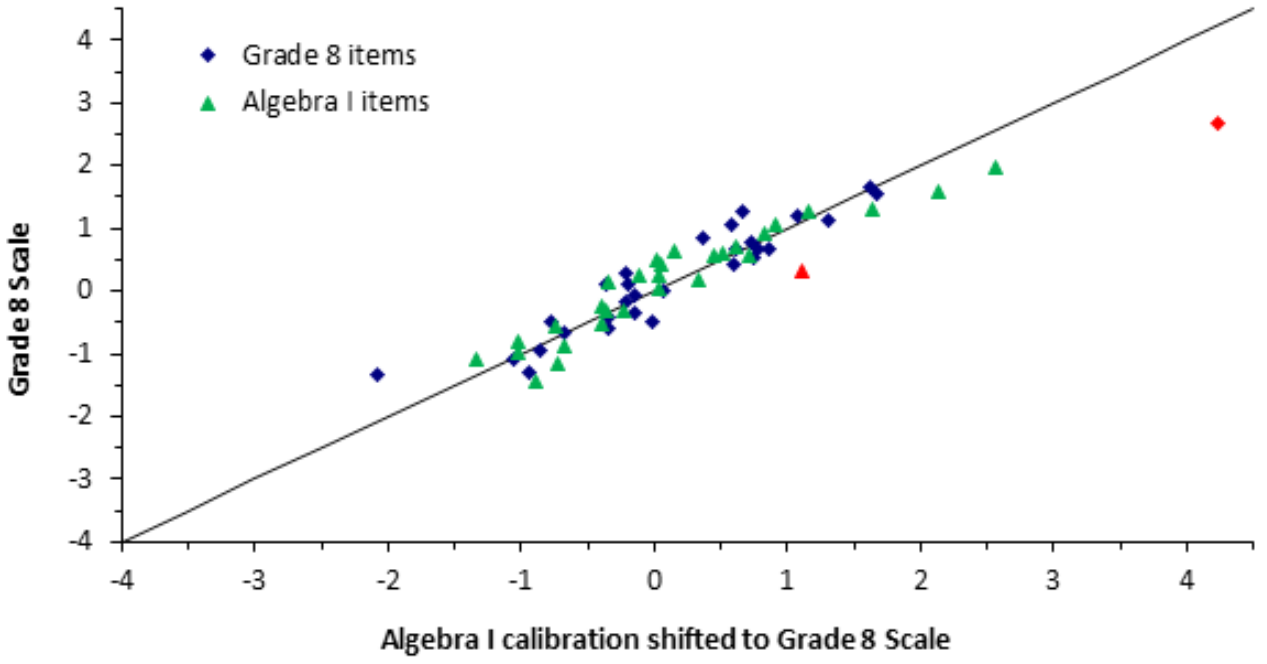


Figure C–7. CDT Mathematics: Geometry to Grade 8 Linking – All Links

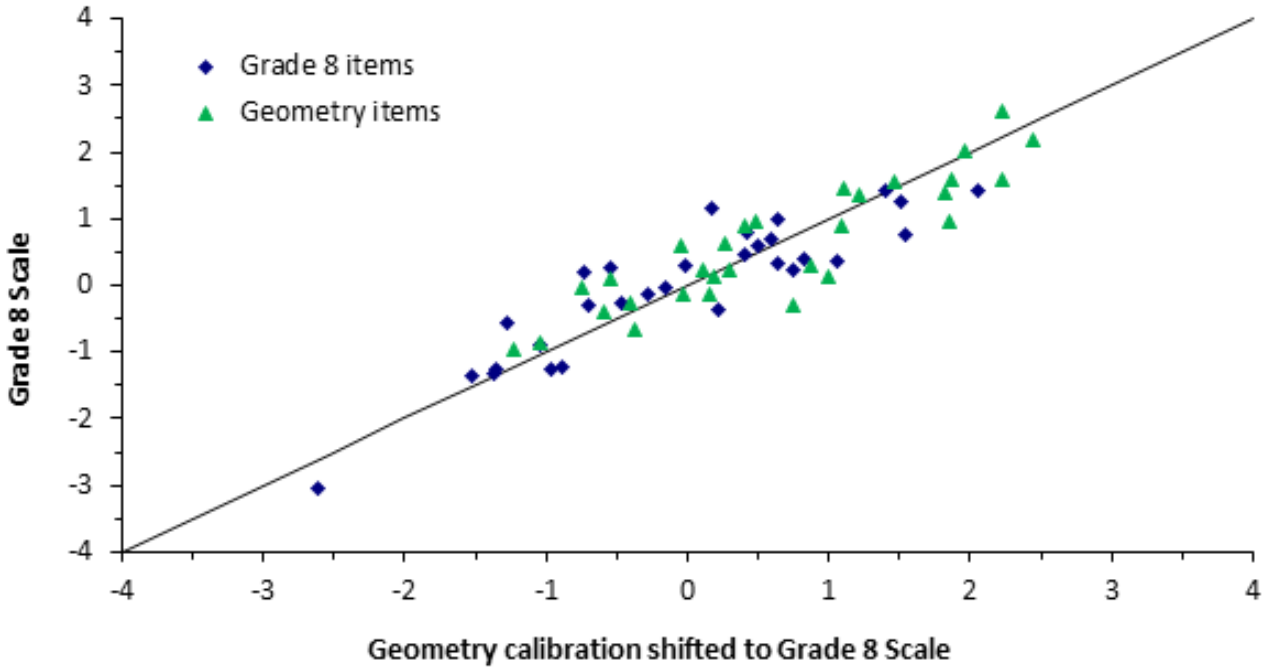
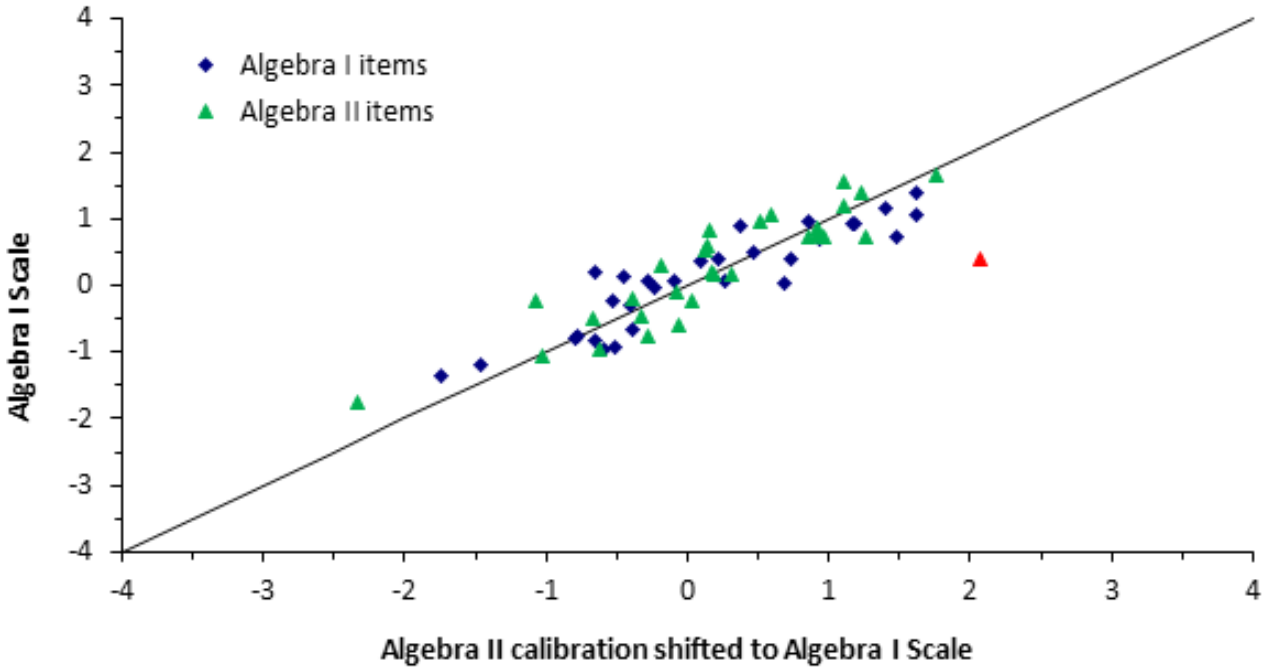


Figure C–8. CDT Mathematics: Algebra II to Algebra I Linking – All Links



READING/LITERATURE

Tables C–18 through C–23 show n-counts, eligible content code, and diagnostic category for each of the vertical linking items.

Each item was administered in two grades so there are two n-counts: one for the lower grade and one for the upper grade. For example, item 613607 is a grade 3 item used to link grades 3 and 4. It was administered 761 times on the lower grade form (grade 3) and 826 times on the upper grade form (grade 4). In some cases, a linking item was also a common item. This results in n-count that is much higher in one of the two grades. For example, item 613400 is a grade 4 item used to link grades 3 and 4. It was also a common grade 4 item (meaning it appeared on all grade 4 forms). The n-counts reflect this: Grade 3 n-count is 754 while grade 4 n-count is 6,574.

The diagnostic categories are[3]

- Comprehension
- Vocabulary
- Interpretation/Analysis Literary Elements & Devices
- Interpretation/Analysis Persuasive Techniques
- Interpretation/Analysis Organizational Skills

[3] Reading diagnostic categories changed at the start of the 2013–2014 school year due to re-alignment to the Pennsylvania Core Standards. See Chapter Thirteen for a list of the current diagnostic categories.

Table C–18. Reading/Literature Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
613605	3	Grade 3 to Grade 4	5272	823	R3A.1.1.2	Vocabulary
613613	3	Grade 3 to Grade 4	5270	822	R3A.2.2.1	Vocabulary
613614	3	Grade 3 to Grade 4	5275	822	R3A.2.1.1	Vocabulary
613592	3	Grade 3 to Grade 4	5262	822	R3A.2.3.1	Comprehension
613593	3	Grade 3 to Grade 4	5263	822	R3A.2.4.1	Comprehension
613460	3	Grade 3 to Grade 4	5251	823	R3A.1.2.2	Vocabulary
613459	3	Grade 3 to Grade 4	5245	822	R3A.1.1.1	Vocabulary
613461	3	Grade 3 to Grade 4	5242	823	R3A.1.4.1	Comprehension
613463	3	Grade 3 to Grade 4	5246	823	R3B.2.1.1	I/A Literary
613462	3	Grade 3 to Grade 4	5241	823	R3A.1.5.1	Comprehension
613607	3	Grade 3 to Grade 4	761	826	R3A.1.2.1	Vocabulary
613446	3	Grade 3 to Grade 4	752	825	R3A.1.1.1	Vocabulary
613444	3	Grade 3 to Grade 4	752	824	R3B.1.1.1	I/A Literary
613445	3	Grade 3 to Grade 4	751	823	R3A.1.5.1	Comprehension
613440	3	Grade 3 to Grade 4	744	823	R3A.1.2.2	Vocabulary
613439	3	Grade 3 to Grade 4	740	823	R3A.1.1.1	Vocabulary
613438	3	Grade 3 to Grade 4	739	822	R3B.1.1.1	I/A Literary
613443	3	Grade 3 to Grade 4	739	823	R3A.1.6.1	Comprehension
613442	3	Grade 3 to Grade 4	735	822	R3A.1.5.1	Comprehension
613441	3	Grade 3 to Grade 4	733	821	R3A.1.3.1	Comprehension
613220	4	Grade 3 to Grade 4	755	6576	R4B.2.1.3	I/A Literary
613219	4	Grade 3 to Grade 4	754	6573	R4B.2.1.2	I/A Literary
613399	4	Grade 3 to Grade 4	757	6569	R4A.2.2.1	Vocabulary
613400	4	Grade 3 to Grade 4	754	6574	R4A.2.3.1	Comprehension
613402	4	Grade 3 to Grade 4	756	6568	R4B.3.2.1	I/A Persuasive
613403	4	Grade 3 to Grade 4	759	6566	R4B.3.2.1	I/A Persuasive
613401	4	Grade 3 to Grade 4	756	6570	R4A.2.6.1	Comprehension
613288	4	Grade 3 to Grade 4	757	6569	R4A.1.1.2	Vocabulary
613291	4	Grade 3 to Grade 4	756	6567	R4A.1.1.1	Vocabulary
613295	4	Grade 3 to Grade 4	757	6563	R4A.2.2.1	Vocabulary
613289	4	Grade 3 to Grade 4	756	804	R4A.1.2.1	Vocabulary
613292	4	Grade 3 to Grade 4	756	805	R4A.1.2.2	Vocabulary
613215	4	Grade 3 to Grade 4	755	805	R4A.1.2.2	Vocabulary
613213	4	Grade 3 to Grade 4	751	803	R4B.2.1.1	I/A Literary
613214	4	Grade 3 to Grade 4	752	804	R4A.1.4.1	Comprehension
613388	4	Grade 3 to Grade 4	749	827	R4A.2.3.1	Comprehension
613389	4	Grade 3 to Grade 4	750	827	R4A.2.4.1	Comprehension

Table C–18 (continued). Reading/Literature Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
613391	4	Grade 3 to Grade 4	748	827	R4B.3.3.2	I/A Organizational
613392	4	Grade 3 to Grade 4	746	826	R4B.3.3.3	I/A Organizational
613390	4	Grade 3 to Grade 4	746	826	R4A.2.5.1	Comprehension

Table C–19. Reading/Literature Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
613220	4	Grade 4 to Grade 5	6576	955	R4B.2.1.3	I/A Literary
613219	4	Grade 4 to Grade 5	6573	957	R4B.2.1.2	I/A Literary
613399	4	Grade 4 to Grade 5	6569	958	R4A.2.2.1	Vocabulary
613400	4	Grade 4 to Grade 5	6574	958	R4A.2.3.1	Comprehension
613402	4	Grade 4 to Grade 5	6568	957	R4B.3.2.1	I/A Persuasive
613403	4	Grade 4 to Grade 5	6566	957	R4B.3.2.1	I/A Persuasive
613401	4	Grade 4 to Grade 5	6570	958	R4A.2.6.1	Comprehension
613288	4	Grade 4 to Grade 5	6569	958	R4A.1.1.2	Vocabulary
613291	4	Grade 4 to Grade 5	6567	958	R4A.1.1.1	Vocabulary
613295	4	Grade 4 to Grade 5	6563	958	R4A.2.2.1	Vocabulary
613293	4	Grade 4 to Grade 5	830	931	R4A.2.1.2	Vocabulary
613297	4	Grade 4 to Grade 5	829	930	R4A.2.2.2	Vocabulary
613212	4	Grade 4 to Grade 5	829	930	R4A.1.1.2	Vocabulary
613211	4	Grade 4 to Grade 5	830	926	R4A.1.5.1	Comprehension
613210	4	Grade 4 to Grade 5	829	925	R4A.1.6.1	Comprehension
613369	4	Grade 4 to Grade 5	815	920	R4A.2.2.1	Vocabulary
613370	4	Grade 4 to Grade 5	813	920	R4A.2.4.1	Comprehension
613372	4	Grade 4 to Grade 5	813	919	R4B.3.1.1	I/A Persuasive
613371	4	Grade 4 to Grade 5	813	917	R4A.2.5.1	Comprehension
613373	4	Grade 4 to Grade 5	812	915	R4B.3.3.1	I/A Organizational
611554	5	Grade 4 to Grade 5	812	7546	R5A.2.1.1	Vocabulary
613007	5	Grade 4 to Grade 5	813	7530	R5B.2.1.4	I/A Literary
613005	5	Grade 4 to Grade 5	810	7528	R5B.1.1.1	I/A Literary
613006	5	Grade 4 to Grade 5	812	7526	R5A.1.6.2	Comprehension
611354	5	Grade 4 to Grade 5	811	7530	R5A.2.1.2	Vocabulary
611377	5	Grade 4 to Grade 5	808	7524	R5B.3.3.2	I/A Organizational
611376	5	Grade 4 to Grade 5	812	7526	R5B.3.1.1	I/A Persuasive
611390	5	Grade 4 to Grade 5	810	7517	R5B.3.3.3	I/A Organizational
611374	5	Grade 4 to Grade 5	807	7510	R5A.2.5.1	Comprehension
611375	5	Grade 4 to Grade 5	808	7509	R5A.2.6.2	Comprehension
611550	5	Grade 4 to Grade 5	826	931	R5A.2.1.2	Vocabulary
611245	5	Grade 4 to Grade 5	826	924	R5B.2.1.1	I/A Literary
611246	5	Grade 4 to Grade 5	826	924	R5B.2.2.1	I/A Literary
611244	5	Grade 4 to Grade 5	826	921	R5A.1.4.1	Comprehension
611269	5	Grade 4 to Grade 5	826	935	R5A.2.1.1	Vocabulary
611272	5	Grade 4 to Grade 5	824	935	R5B.3.1.1	I/A Persuasive
611270	5	Grade 4 to Grade 5	823	935	R5A.2.3.1	Comprehension

Table C–19 (continued). Reading/Literature Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
611274	5	Grade 4 to Grade 5	824	935	R5B.3.3.2	I/A Organizational
611271	5	Grade 4 to Grade 5	824	934	R5A.2.6.1	Comprehension
611273	5	Grade 4 to Grade 5	824	933	R5B.3.3.1	I/A Organizational

Table C–20. Reading/Literature Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
611554	5	Grade 5 to Grade 6	7546	716	R5A.2.1.1	Vocabulary
613007	5	Grade 5 to Grade 6	7530	719	R5B.2.1.4	I/A Literary
613005	5	Grade 5 to Grade 6	7528	721	R5B.1.1.1	I/A Literary
613006	5	Grade 5 to Grade 6	7526	720	R5A.1.6.2	Comprehension
611354	5	Grade 5 to Grade 6	7530	719	R5A.2.1.2	Vocabulary
611377	5	Grade 5 to Grade 6	7524	717	R5B.3.3.2	I/A Organizational
611376	5	Grade 5 to Grade 6	7526	719	R5B.3.1.1	I/A Persuasive
611390	5	Grade 5 to Grade 6	7517	718	R5B.3.3.3	I/A Organizational
611374	5	Grade 5 to Grade 6	7510	717	R5A.2.5.1	Comprehension
611375	5	Grade 5 to Grade 6	7509	717	R5A.2.6.2	Comprehension
611247	5	Grade 5 to Grade 6	928	697	R5A.1.1.1	Vocabulary
611251	5	Grade 5 to Grade 6	928	698	R5B.2.1.4	I/A Literary
611250	5	Grade 5 to Grade 6	926	697	R5B.2.1.3	I/A Literary
611249	5	Grade 5 to Grade 6	926	696	R5A.1.3.2	Comprehension
611248	5	Grade 5 to Grade 6	926	694	R5A.1.3.1	Comprehension
611309	5	Grade 5 to Grade 6	925	688	R5B.3.3.3	I/A Organizational
611278	5	Grade 5 to Grade 6	924	687	R5A.2.3.2	Comprehension
611291	5	Grade 5 to Grade 6	921	685	R5B.3.3.1	I/A Organizational
611545	5	Grade 5 to Grade 6	942	682	R5A.1.1.2	Vocabulary
611553	5	Grade 5 to Grade 6	945	680	R5A.2.1.1	Vocabulary
610132	6	Grade 5 to Grade 6	936	7111	R6A.1.2.1	Vocabulary
610135	6	Grade 5 to Grade 6	937	7105	R6B.2.1.2	I/A Literary
610133	6	Grade 5 to Grade 6	935	7086	R6A.1.4.1	Comprehension
610355	6	Grade 5 to Grade 6	935	7075	R6A.1.3.2	Comprehension
610136	6	Grade 5 to Grade 6	935	7066	R6B.2.2.2	I/A Literary
610134	6	Grade 5 to Grade 6	936	7069	R6A.1.6.1	Comprehension
612249	6	Grade 5 to Grade 6	937	7035	R6B.3.3.4	I/A Organizational
612248	6	Grade 5 to Grade 6	936	7026	R6A.2.6.2	Comprehension
607918	6	Grade 5 to Grade 6	937	7150	R6A.2.1.1	Vocabulary
607921	6	Grade 5 to Grade 6	937	7142	R6A.2.1.2	Vocabulary
607927	6	Grade 5 to Grade 6	941	713	R6A.2.2.1	Vocabulary
607917	6	Grade 5 to Grade 6	941	716	R6A.2.1.1	Vocabulary
610141	6	Grade 5 to Grade 6	938	703	R6A.1.1.1	Vocabulary
610144	6	Grade 5 to Grade 6	937	701	R6B.2.1.1	I/A Literary
610305	6	Grade 5 to Grade 6	933	700	R6A.1.3.1	Comprehension
610145	6	Grade 5 to Grade 6	932	695	R6B.2.2.2	I/A Literary
610142	6	Grade 5 to Grade 6	927	695	R6A.1.5.1	Comprehension

Table C–20 (continued). Reading/Literature Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
610143	6	Grade 5 to Grade 6	925	694	R6A.1.6.1	Comprehension
610310	6	Grade 5 to Grade 6	917	726	R6B.3.2.2	I/A Persuasive
610309	6	Grade 5 to Grade 6	917	726	R6A.2.6.1	Comprehension

Table C–21. Reading/Literature Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
610132	6	Grade 6 to Grade 7	7111	549	R6A.1.2.1	Vocabulary
610135	6	Grade 6 to Grade 7	7105	550	R6B.2.1.2	I/A Literary
610133	6	Grade 6 to Grade 7	7086	551	R6A.1.4.1	Comprehension
610355	6	Grade 6 to Grade 7	7075	551	R6A.1.3.2	Comprehension
610136	6	Grade 6 to Grade 7	7066	551	R6B.2.2.2	I/A Literary
610134	6	Grade 6 to Grade 7	7069	551	R6A.1.6.1	Comprehension
607921	6	Grade 6 to Grade 7	7142	550	R6A.2.1.2	Vocabulary
610327	6	Grade 6 to Grade 7	685	550	R6A.1.2.2	Vocabulary
610328	6	Grade 6 to Grade 7	682	549	R6B.2.1.4	I/A Literary
610329	6	Grade 6 to Grade 7	679	548	R6B.2.2.1	I/A Literary
610065	6	Grade 6 to Grade 7	696	551	R6A.1.1.1	Vocabulary
610071	6	Grade 6 to Grade 7	692	550	R6A.1.3.1	Comprehension
610066	6	Grade 6 to Grade 7	691	550	R6B.2.1.4	I/A Literary
610070	6	Grade 6 to Grade 7	689	551	R6A.1.3.2	Comprehension
610078	6	Grade 6 to Grade 7	687	551	R6B.2.1.3	I/A Literary
609022	6	Grade 6 to Grade 7	1433	551	R6A.1.1.2	Vocabulary
609025	6	Grade 6 to Grade 7	1431	550	R6B.2.1.1	I/A Literary
609026	6	Grade 6 to Grade 7	1431	550	R6B.2.1.4	I/A Literary
609023	6	Grade 6 to Grade 7	1431	549	R6A.1.3.1	Comprehension
609024	6	Grade 6 to Grade 7	1432	548	R6A.1.6.2	Comprehension
609658	7	Grade 6 to Grade 7	722	4978	R7A.1.1.1	Vocabulary
609663	7	Grade 6 to Grade 7	725	4976	R7B.2.2.1	I/A Literary
609661	7	Grade 6 to Grade 7	723	4971	R7A.1.5.1	Comprehension
610324	7	Grade 6 to Grade 7	724	4974	R7A.2.2.1	Vocabulary
610325	7	Grade 6 to Grade 7	723	4968	R7A.2.3.2	Comprehension
610146	7	Grade 6 to Grade 7	722	563	R7A.1.1.1	Vocabulary
610149	7	Grade 6 to Grade 7	723	565	R7B.2.1.1	I/A Literary
610147	7	Grade 6 to Grade 7	722	564	R7A.1.3.1	Comprehension
610338	7	Grade 6 to Grade 7	721	563	R7B.1.1.1	I/A Literary
610148	7	Grade 6 to Grade 7	721	564	R7A.1.6.1	Comprehension
607933	7	Grade 6 to Grade 7	705	545	R7A.1.1.2	Vocabulary
607936	7	Grade 6 to Grade 7	703	545	R7A.1.2.1	Vocabulary
609243	7	Grade 6 to Grade 7	701	544	R7B.2.1.2	I/A Literary
609053	7	Grade 6 to Grade 7	700	544	R7A.1.3.2	Comprehension
609219	7	Grade 6 to Grade 7	700	544	R7A.1.6.2	Comprehension
609037	7	Grade 6 to Grade 7	695	553	R7A.2.2.2	Vocabulary
609038	7	Grade 6 to Grade 7	692	552	R7A.2.4.1	Comprehension

Table C–21 (continued). Reading/Literature Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
609039	7	Grade 6 to Grade 7	684	551	R7A.2.6.2	Comprehension
609040	7	Grade 6 to Grade 7	680	553	R7B.3.1.1	I/A Persuasive
609041	7	Grade 6 to Grade 7	678	552	R7B.3.3.1	I/A Organizational

Table C–22. Reading/Literature Items Used to Link Grade 7 to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
609658	7	Grade 8 to Grade 7	4978	518	R7A.1.1.1	Vocabulary
609663	7	Grade 8 to Grade 7	4976	518	R7B.2.2.1	I/A Literary
609661	7	Grade 8 to Grade 7	4971	517	R7A.1.5.1	Comprehension
610324	7	Grade 8 to Grade 7	4974	516	R7A.2.2.1	Vocabulary
610325	7	Grade 8 to Grade 7	4968	515	R7A.2.3.2	Comprehension
610146	7	Grade 8 to Grade 7	563	491	R7A.1.1.1	Vocabulary
610149	7	Grade 8 to Grade 7	565	491	R7B.2.1.1	I/A Literary
610147	7	Grade 8 to Grade 7	564	490	R7A.1.3.1	Comprehension
610338	7	Grade 8 to Grade 7	563	488	R7B.1.1.1	I/A Literary
610148	7	Grade 8 to Grade 7	564	485	R7A.1.6.1	Comprehension
614855	7	Grade 8 to Grade 7	559	516	R7A.1.1.2	Vocabulary
614859	7	Grade 8 to Grade 7	558	516	R7B.2.2.1	I/A Literary
614858	7	Grade 8 to Grade 7	559	515	R7B.2.1.2	I/A Literary
614856	7	Grade 8 to Grade 7	559	515	R7A.1.3.2	Comprehension
614857	7	Grade 8 to Grade 7	558	514	R7A.1.6.1	Comprehension
609152	7	Grade 8 to Grade 7	550	504	R7B.3.1.1	I/A Persuasive
609072	7	Grade 8 to Grade 7	551	502	R7A.2.5.1	Comprehension
609209	7	Grade 8 to Grade 7	548	500	R7B.1.1.1	I/A Literary
609210	7	Grade 8 to Grade 7	548	496	R7B.2.1.1	I/A Literary
609208	7	Grade 8 to Grade 7	548	495	R7A.1.3.1	Comprehension
609060	8	Grade 8 to Grade 7	550	4645	R8B.3.1.1	I/A Persuasive
609059	8	Grade 8 to Grade 7	550	4647	R8A.2.5.1	Comprehension
608017	8	Grade 8 to Grade 7	550	4637	R8A.1.1.2	Vocabulary
608016	8	Grade 8 to Grade 7	551	4629	R8B.2.1.2	I/A Literary
607999	8	Grade 8 to Grade 7	550	4622	R8A.1.6.2	Comprehension
610087	8	Grade 8 to Grade 7	550	510	R8B.3.3.4	I/A Organizational
610260	8	Grade 8 to Grade 7	550	509	R8B.3.3.2	I/A Organizational
610090	8	Grade 8 to Grade 7	550	511	R8B.3.3.4	I/A Organizational
610089	8	Grade 8 to Grade 7	550	511	R8B.3.3.4	I/A Organizational
610088	8	Grade 8 to Grade 7	550	510	R8B.3.3.4	I/A Organizational
609135	8	Grade 8 to Grade 7	540	531	R8B.3.2.1	I/A Persuasive
609131	8	Grade 8 to Grade 7	540	532	R8B.3.2.1	I/A Persuasive
609120	8	Grade 8 to Grade 7	539	532	R8B.3.3.2	I/A Organizational
609143	8	Grade 8 to Grade 7	539	531	R8A.2.3.2	Comprehension
609140	8	Grade 8 to Grade 7	539	532	R8A.2.6.2	Comprehension
609264	8	Grade 8 to Grade 7	539	513	R8A.1.1.2	Vocabulary
609267	8	Grade 8 to Grade 7	539	513	R8B.2.1.2	I/A Literary

Table C–22 (continued). Reading/Literature Items Used to Link Grade 7 to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
609265	8	Grade 8 to Grade 7	539	514	R8A.1.3.2	Comprehension
609269	8	Grade 8 to Grade 7	539	514	R8B.2.2.1	I/A Literary
609266	8	Grade 8 to Grade 7	539	515	R8A.1.6.1	Comprehension

Table C–23. Reading/Literature Items Used to Link Literature to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
608017	8	Literature to Grade 8	4637	255	R8A.1.1.2	Vocabulary
608016	8	Literature to Grade 8	4629	253	R8B.2.1.2	I/A Literary
607999	8	Literature to Grade 8	4622	252	R8A.1.6.2	Comprehension
610087	8	Literature to Grade 8	510	256	R8B.3.3.4	I/A Organizational
610260	8	Literature to Grade 8	509	256	R8B.3.3.2	I/A Organizational
610090	8	Literature to Grade 8	511	255	R8B.3.3.4	I/A Organizational
610089	8	Literature to Grade 8	511	255	R8B.3.3.4	I/A Organizational
610088	8	Literature to Grade 8	510	255	R8B.3.3.4	I/A Organizational
607957	8	Literature to Grade 8	502	254	R8A.1.1.2	Vocabulary
607963	8	Literature to Grade 8	501	254	R8A.1.1.1	Vocabulary
607958	8	Literature to Grade 8	516	258	R8A.1.2.1	Vocabulary
607962	8	Literature to Grade 8	516	258	R8A.1.1.1	Vocabulary
612324	8	Literature to Grade 8	516	257	R8B.3.3.4	I/A Organizational
612280	8	Literature to Grade 8	517	257	R8B.3.3.4	I/A Organizational
612279	8	Literature to Grade 8	517	257	R8A.2.6.1	Comprehension
609244	8	Literature to Grade 8	523	257	R8A.1.1.1	Vocabulary
609254	8	Literature to Grade 8	523	256	R8B.2.1.1	I/A Literary
609279	8	Literature to Grade 8	522	256	R8B.1.1.1	I/A Literary
609245	8	Literature to Grade 8	523	256	R8A.1.3.1	Comprehension
609252	8	Literature to Grade 8	523	256	R8A.1.6.1	Comprehension
608136	Lit	Literature to Grade 8	515	258	L.F.1.3.1	Comprehension
608138	Lit	Literature to Grade 8	515	258	L.F.2.3.4	I/A Literary
608137	Lit	Literature to Grade 8	512	257	L.F.2.2.1	Comprehension
614029	Lit	Literature to Grade 8	515	271	L.F.1.2.4	Vocabulary
614032	Lit	Literature to Grade 8	515	271	L.F.2.3.1	I/A Literary
614030	Lit	Literature to Grade 8	515	271	L.F.2.1.1	Comprehension
614031	Lit	Literature to Grade 8	515	271	L.F.2.2.2	Comprehension
614033	Lit	Literature to Grade 8	515	271	L.F.2.3.2	I/A Literary
614034	Lit	Literature to Grade 8	510	271	L.F.2.5.1	I/A Literary
608118	Lit	Literature to Grade 8	514	265	L.F.1.2.4	Vocabulary
610352	Lit	Literature to Grade 8	516	261	L.F.2.5.2	I/A Literary
610092	Lit	Literature to Grade 8	511	261	L.F.2.2.1	Comprehension
610094	Lit	Literature to Grade 8	509	260	L.F.2.3.6	I/A Literary
610095	Lit	Literature to Grade 8	510	259	L.F.2.4.1	I/A Literary
610093	Lit	Literature to Grade 8	509	260	L.F.2.3.4	I/A Literary
610091	Lit	Literature to Grade 8	507	260	L.F.1.1.1	Comprehension
612547	Lit	Literature to Grade 8	504	258	L.F.1.2.2	Vocabulary

Table C–23 (continued). Reading/Literature Items Used to Link Literature to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Reading/Literature Diagnostic Category
612498	Lit	Literature to Grade 8	502	258	L.F.2.2.2	Comprehension
612548	Lit	Literature to Grade 8	499	258	L.F.1.3.2	Comprehension
612496	Lit	Literature to Grade 8	497	258	L.F.1.1.1	Comprehension

Tables C–24 through C–29 summarize the number of linking items by diagnostic category.

Vertical linking items are not distributed evenly across the diagnostic categories. This is due to the fact that Reading and Literature items are passage based. The three passage types (literary, persuasive, and organizational) may each have associated comprehension and vocabulary items, as well as interpretation/analysis items.

Table C–24. Number of Items Linking Grade 3 to Grade 4 by Diagnostic Category

Diagnostic Category	Grade 3 Items	Grade 4 Items	Total
Comprehension	8	6	14
Vocabulary	9	7	16
I/A Literary	3	3	6
I/A Persuasive	0	2	2
I/A Organizational	0	2	2
TOTAL	20	20	40

Table C–25. Number of Items Linking Grade 4 to Grade 5 by Diagnostic Category

Diagnostic Category	Grade 4 Items	Grade 5 Items	Total
Comprehension	6	6	12
Vocabulary	8	4	12
I/A Literary	2	4	6
I/A Persuasive	3	2	5
I/A Organizational	1	4	5
TOTAL	20	20	40

Table C–26. Number of Items Linking Grade 5 to Grade 6 by Diagnostic Category

Diagnostic Category	Grade 5 Items	Grade 6 Items	Total
Comprehension	6	8	14
Vocabulary	5	6	11
I/A Literary	4	4	8
I/A Persuasive	1	1	2
I/A Organizational	4	1	5
TOTAL	20	20	40

Table C–27. Number of Items Linking Grade 6 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 6 Items	Grade 7 Items	Total
Comprehension	7	8	15
Vocabulary	5	6	11
I/A Literary	8	4	12
I/A Persuasive	0	1	1
I/A Organizational	0	1	1
TOTAL	20	20	40

Table C–28. Number of Items Linking Grade 8 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 7 Items	Grade 8 Items	Total
Comprehension	8	6	14
Vocabulary	4	2	6
I/A Literary	7	3	10
I/A Persuasive	1	3	4
I/A Organizational	0	6	6
TOTAL	20	20	40

Table C–29. Number of Items Linking Literature to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Literature Items	Total
Comprehension	4	9	13
Vocabulary	6	3	9
I/A Literary	3	8	11
I/A Persuasive	0	0	0
I/A Organizational	7	0	7
TOTAL	20	20	40

Table C–30. Reading/Literature Example of Vertical Linking Workbook

		Grade 4 Calibration			Grade 5 Calibration				Grade 4 on		
Item ID	Item Grade	Difficulty	Fit	Displace	Difficulty	Fit	Displace	Discrepancy	Grade 5 Scale	Robust Z	Flag
613220	4	0.700	1.090	0.000	0.258	1.040	-0.003	-0.442	0.290	-0.271	
613219	4	-0.063	0.980	0.000	-0.495	0.960	-0.003	-0.432	-0.473	-0.235	
613399	4	0.557	1.040	0.000	0.056	0.980	-0.003	-0.501	0.147	-0.486	
613400	4	0.589	1.020	0.000	0.131	1.000	-0.003	-0.458	0.179	-0.329	
613402	4	0.316	1.070	0.000	0.014	0.930	-0.003	-0.302	-0.094	0.238	
613403	4	0.295	0.970	0.000	-0.446	0.890	-0.003	-0.741	-0.115	-1.360	
613401	4	-0.657	0.810	0.000	-1.307	0.810	-0.003	-0.650	-1.067	-1.028	
613288	4	-0.608	0.960	0.000	-1.044	0.950	-0.003	-0.436	-1.018	-0.249	
613291	4	0.927	1.200	0.000	0.628	1.170	-0.003	-0.299	0.517	0.249	
613295	4	-1.117	0.880	0.000	-1.712	0.900	-0.003	-0.595	-1.527	-0.828	
613293	4	0.173	0.930	0.002	-0.113	0.880	0.000	-0.286	-0.237	0.297	
613297	4	0.807	1.070	0.002	0.424	0.990	0.000	-0.383	0.397	-0.056	
613212	4	1.664	1.210	0.003	1.491	1.220	0.000	-0.173	1.254	0.708	
613211	4	0.245	0.930	0.002	0.082	0.890	0.000	-0.163	-0.165	0.744	
613210	4	0.203	1.000	0.002	-0.273	0.910	0.000	-0.476	-0.207	-0.395	
613369	4	-0.556	0.900	0.004	-0.791	0.920	0.000	-0.235	-0.966	0.482	
613370	4	0.433	0.930	0.004	0.151	0.950	0.000	-0.282	0.023	0.311	
613372	4	-0.305	0.860	0.004	-0.698	0.870	0.000	-0.393	-0.715	-0.093	
613371	4	-0.513	0.910	0.004	-0.670	0.960	0.000	-0.157	-0.923	0.766	
613373	4	1.012	1.060	0.004	1.002	1.040	0.000	-0.010	0.602	1.301	
611554	5	1.180	1.170	0.003	1.126	1.050	0.000	-0.054	0.770	1.141	
613007	5	-0.124	0.900	0.003	-0.476	0.960	-0.001	-0.352	-0.534	0.056	
613005	5	2.069	1.250	0.003	2.138	1.220	0.000	0.069	1.659	1.589	
613006	5	2.275	1.240	0.003	2.367	1.120	0.000	0.092	1.865	1.673	
611354	5	0.669	1.020	0.003	0.576	1.020	-0.001	-0.093	0.259	0.999	
611377	5	0.336	1.060	0.003	0.559	1.010	-0.001	0.223	-0.074	2.149	high robust Z
611376	5	-0.804	0.840	0.003	-0.946	0.850	-0.001	-0.142	-1.214	0.821	
611390	5	1.351	1.110	0.003	1.443	1.040	0.000	0.092	0.941	1.673	
611374	5	0.109	0.930	0.003	-0.065	0.920	-0.001	-0.174	-0.301	0.704	
611375	5	0.581	1.160	0.003	0.605	1.120	-0.001	0.024	0.171	1.425	
611550	5	0.355	1.000	0.001	-0.586	0.900	0.000	-0.941	-0.055	-2.088	high robust Z
611245	5	1.298	1.070	0.001	0.635	1.030	0.000	-0.663	0.888	-1.076	
611246	5	-0.051	0.860	0.001	-0.532	0.850	0.000	-0.481	-0.461	-0.413	
611244	5	-0.152	0.910	0.001	-0.226	0.940	0.000	-0.074	-0.562	1.068	
611269	5	-0.287	0.900	0.001	-1.341	0.960	-0.006	-1.054	-0.697	-2.499	high robust Z
611272	5	-0.860	0.840	0.001	-2.081	0.930	-0.006	-1.221	-1.270	-3.107	high robust Z
611270	5	-0.274	0.900	0.001	-1.286	0.960	-0.006	-1.012	-0.684	-2.346	high robust Z
611274	5	-0.784	0.760	0.001	-2.720	0.870	-0.006	-1.936	-1.194	-5.709	high robust Z
611271	5	0.972	0.910	0.001	0.157	0.900	-0.005	-0.815	0.562	-1.629	
611273	5	2.533	1.250	0.001	2.056	1.040	-0.004	-0.477	2.123	-0.399	
	Mean	0.362			-0.048			-0.410	-0.048	-0.155	
	SD	0.868			1.107			0.415	0.868	1.511	
	SD Ratio	0.784									
	Correlation	0.940									
	Add. Constant	-0.410									
	Median							-0.368			
	Q							0.371			

Figures C–9 through C–14 are the adjacent grade linking plots. Items removed from final linking procedure are colored red.

Figure C–9. CDT Reading/Literature: Grade 3 to Grade 4 Linking – All Links

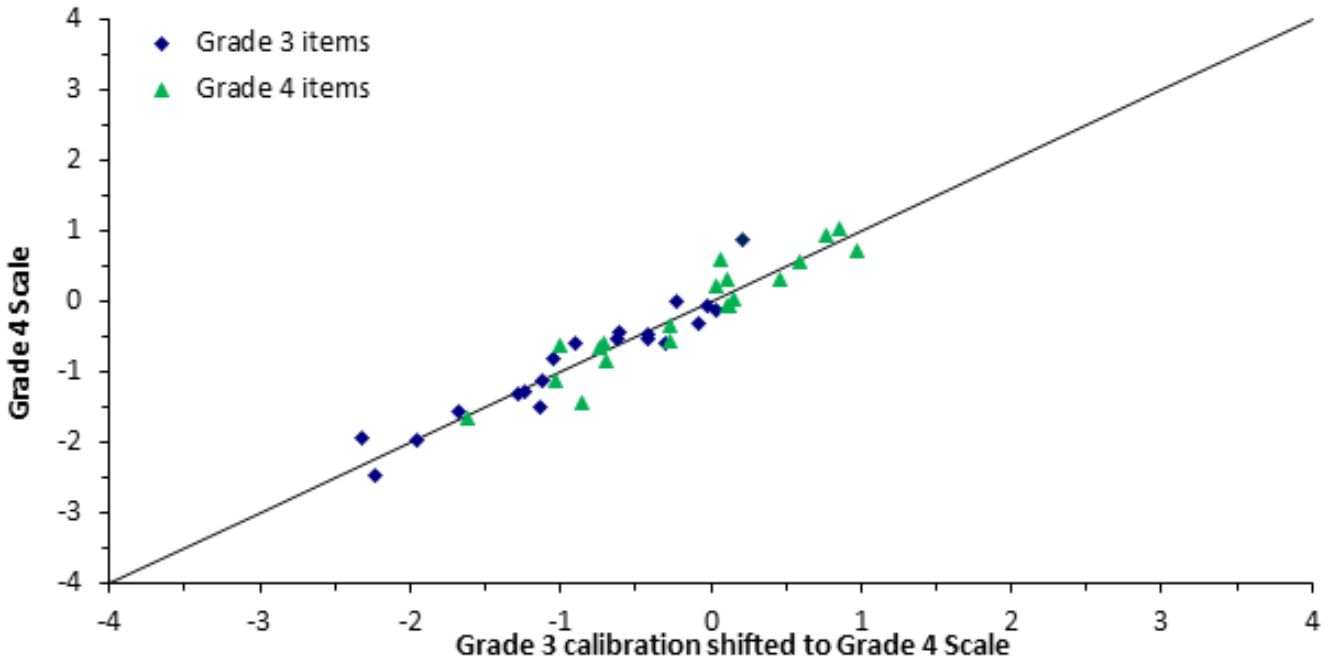


Figure C–10. CDT Reading/Literature: Grade 4 to Grade 5 Linking – All Links

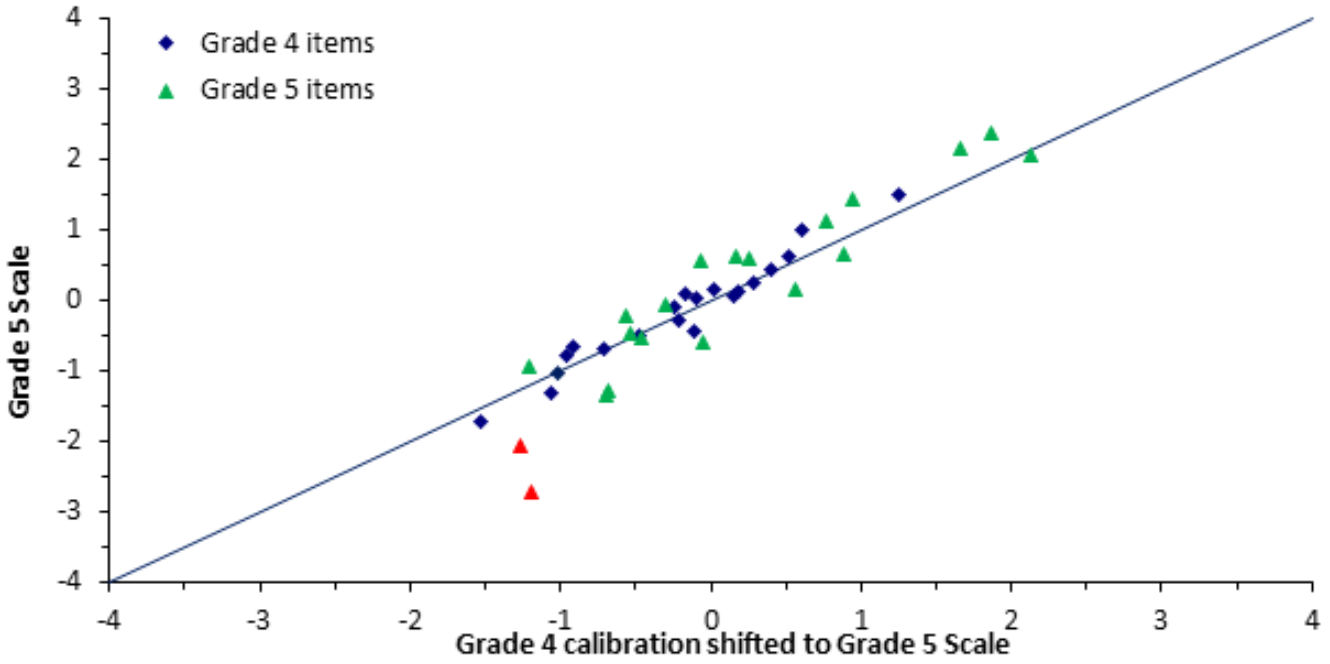


Figure C–11. CDT Reading/Literature: Grade 5 to Grade 6 Linking – All Links

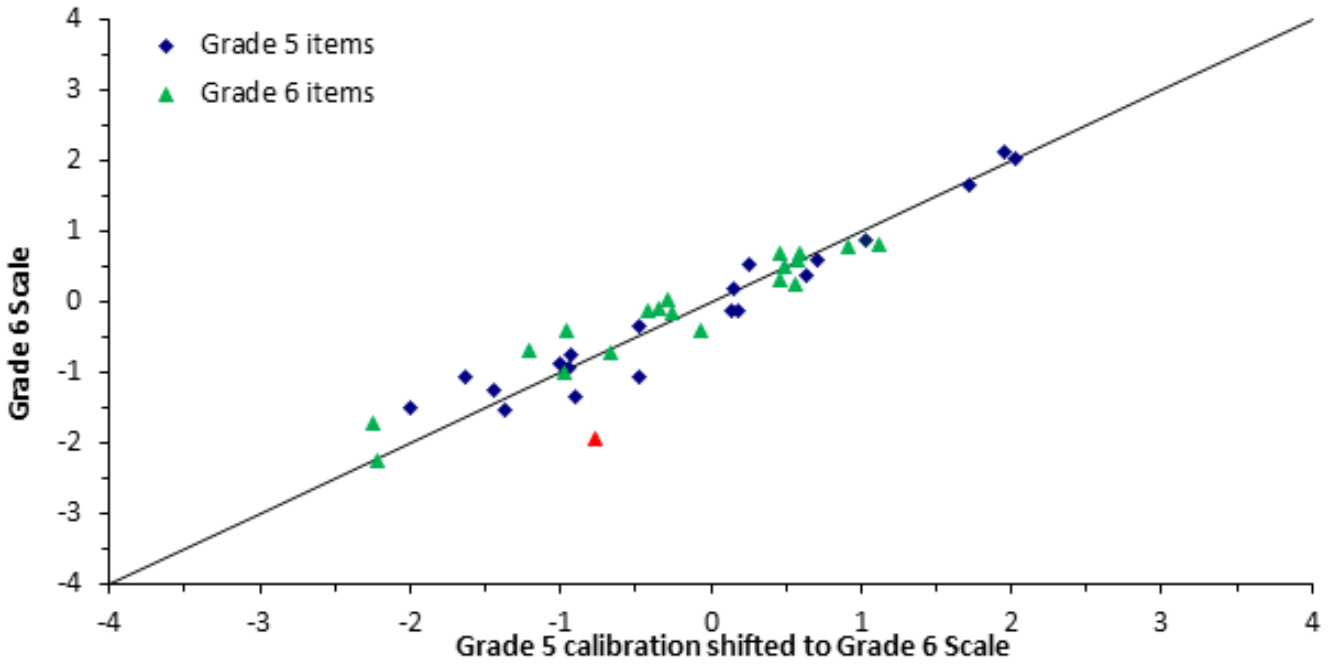


Figure C–12. CDT Reading/Literature: Grade 6 to Grade 7 Linking – All Links

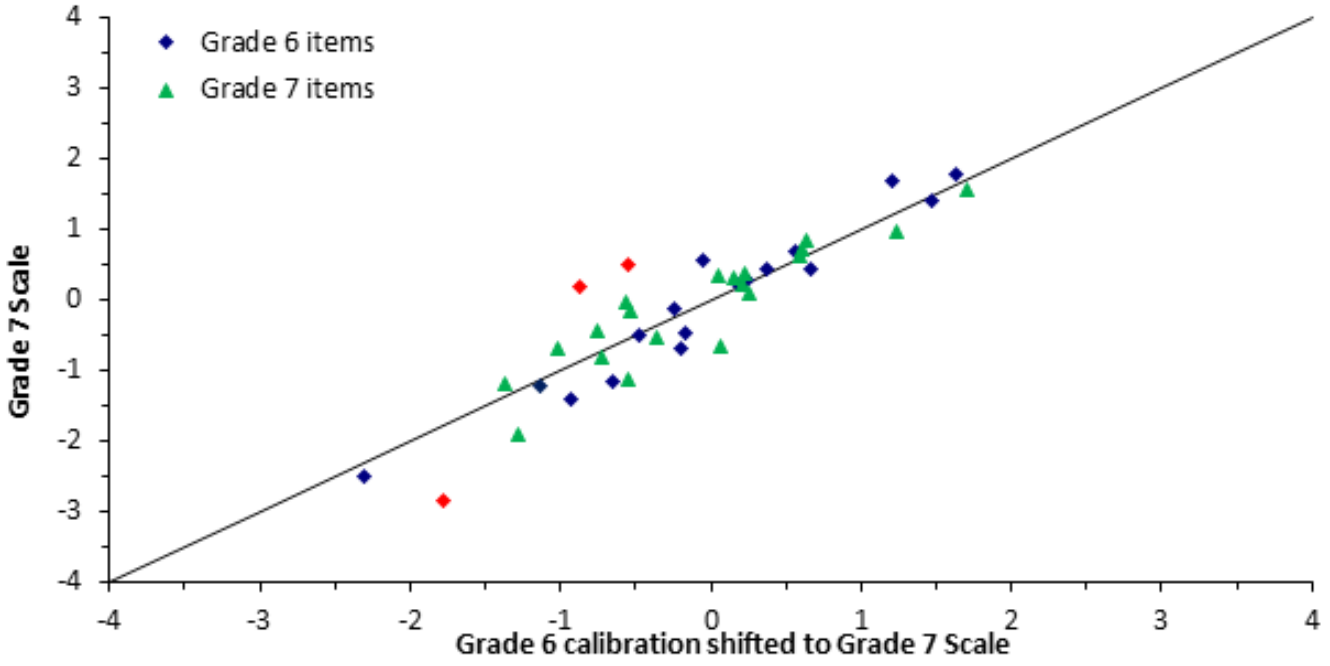


Figure C–13. CDT Reading/Literature: Grade 8 to Grade 7 Linking – All Links

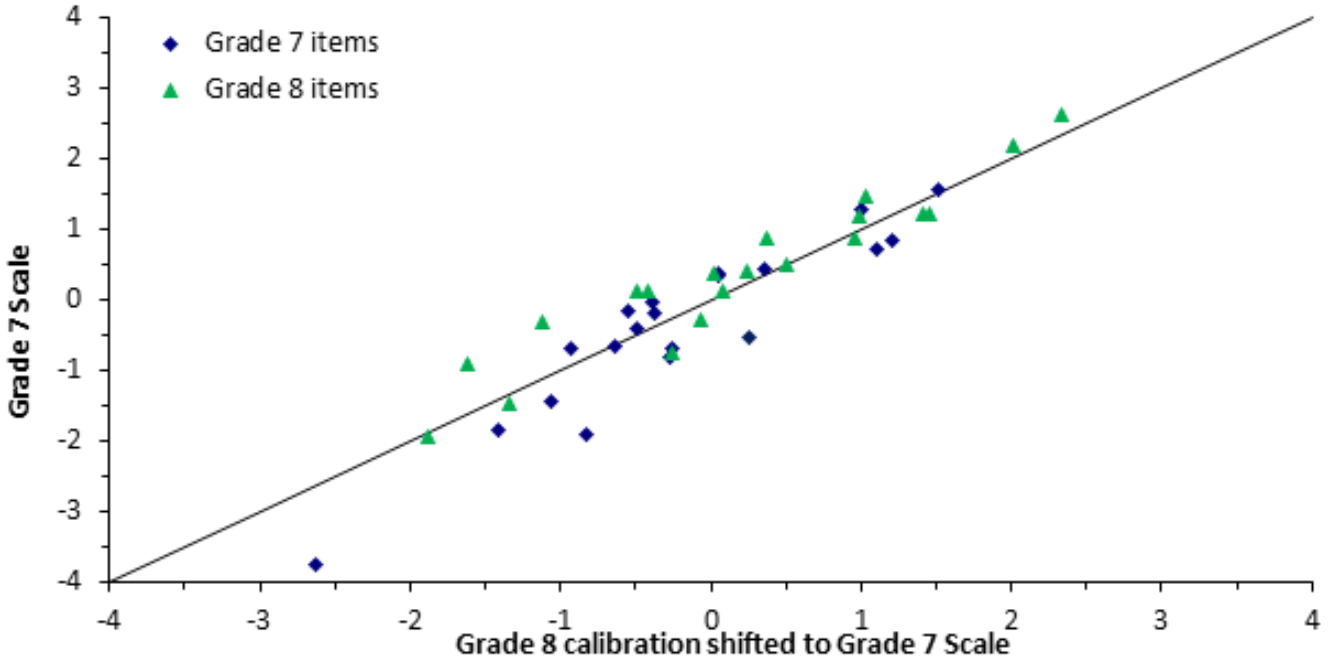
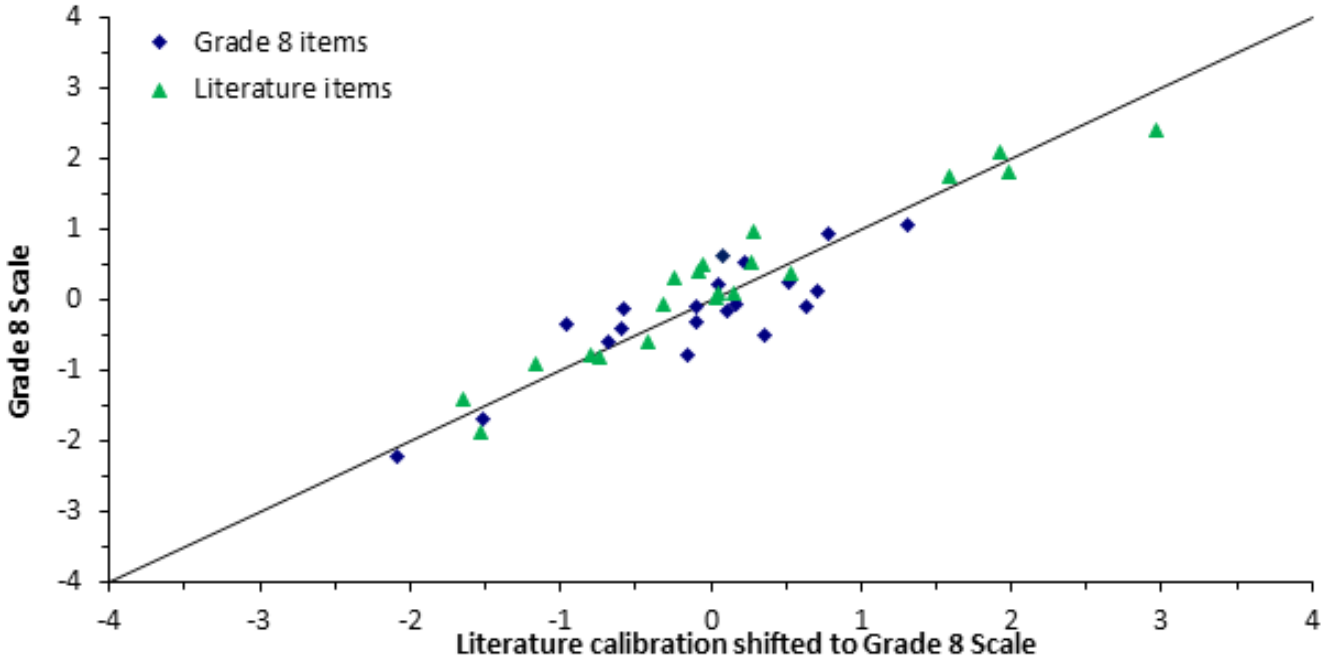


Figure C–14. CDT Reading/Literature: Literature to Grade 8 Linking – All Links



SCIENCE

Tables C–31 through C–37 show n-counts, eligible content code, and diagnostic category for each of the vertical linking items.

Each item was administered in two grades so there are two n-counts: one for the lower grade and one for the upper grade. For example, item 615315 is a grade 3 item used to link grades 3 and 4. It was administered 789 times on the lower grade form (grade 3) and 530 times on the upper grade form (grade 4). In some cases, a linking item was also a common item. This results in n-count that is much higher in one of the two grades. For example, item 617401 is a Biology item used to link Biology and grade 8. It was also a common Biology item (meaning it appeared on all Biology forms). The n-counts reflect this: Grade 8 n-count is 256 while Biology n-count is 4,874.

Diagnostic categories for Biology and Chemistry are different than diagnostic categories for grades 3 through 8 and 11 Science. Items may fall into both a Science diagnostic category and a Biology or Chemistry diagnostic category. This is shown in Tables C–36 and C–37. For example, item 615777 is in the Science diagnostic category “Biological Sciences” and the Biology diagnostic category “Basic Biological Principles”.

The Science diagnostic categories are:

- The Nature of Science
- Biological Science
- Physical Sciences
- Earth and Space Sciences

The Biology diagnostic categories are:

- Basic Biological Principles/Chemical Basis for Life
- Bioenergetics/Homeostasis and Transport
- Cell Growth and Reproduction/Genetics
- Theory of Evolution/Ecology

The Chemistry diagnostic categories are:

- Properties and Classification of Matter
- Atomic Structure and the Periodic Table
- The Mole and Chemical Bonding
- Chemical Relationships and Reactions

Table C–31. Science Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
615315	3	Grade 3 to Grade 4	789	530	S3.A.2.1.3	Nature of Science
615379	3	Grade 3 to Grade 4	790	530	S3.D.1.2.1	Earth and Space Sci.
615333	3	Grade 3 to Grade 4	770	530	S3.B.2.1.1	Biological Sci.
615395	3	Grade 3 to Grade 4	797	530	S3.D.1.3.3	Earth and Space Sci.
615363	3	Grade 3 to Grade 4	1559	530	S3.C.1.1.4	Physical Sci.
615368	3	Grade 3 to Grade 4	773	530	S3.C.2.1.2	Physical Sci.
615314	3	Grade 3 to Grade 4	796	530	S3.A.2.1.2	Nature of Science
615331	3	Grade 3 to Grade 4	782	529	S3.B.1.1.4	Biological Sci.
615324	3	Grade 3 to Grade 4	786	529	S3.A.2.1.3	Nature of Science
615347	3	Grade 3 to Grade 4	796	528	S3.B.3.1.2	Biological Sci.
615385	3	Grade 3 to Grade 4	771	525	S3.D.1.2.1	Earth and Space Sci.
615319	3	Grade 3 to Grade 4	790	524	S3.A.3.1.1	Nature of Science
615339	3	Grade 3 to Grade 4	785	524	S3.B.2.2.1	Biological Sci.
617274	3	Grade 3 to Grade 4	796	525	S3.A.1.1.1	Nature of Science
615400	3	Grade 3 to Grade 4	771	524	S3.D.3.1.1	Earth and Space Sci.
615322	3	Grade 3 to Grade 4	1572	523	S3.A.3.2.1	Nature of Science
615325	3	Grade 3 to Grade 4	773	523	S3.B.1.1.1	Biological Sci.
615376	3	Grade 3 to Grade 4	785	521	S3.D.1.1.1	Earth and Space Sci.
615327	3	Grade 3 to Grade 4	787	521	S3.B.1.1.2	Biological Sci.
615334	3	Grade 3 to Grade 4	794	521	S3.B.2.1.2	Biological Sci.
617229	4	Grade 3 to Grade 4	792	538	S4.C.1.1.2	Physical Sci.
617061	4	Grade 3 to Grade 4	793	1086	S4.A.2.1.4	Nature of Science
617244	4	Grade 3 to Grade 4	789	558	S4.D.1.1.1	Earth and Space Sci.
617095	4	Grade 3 to Grade 4	792	1097	S4.B.2.1.2	Biological Sci.
615621	4	Grade 3 to Grade 4	793	1065	S4.A.1.1.1	Nature of Science
617239	4	Grade 3 to Grade 4	793	1073	S4.C.3.1.1	Physical Sci.
617099	4	Grade 3 to Grade 4	793	539	S4.B.2.2.1	Biological Sci.
617249	4	Grade 3 to Grade 4	792	539	S4.D.1.1.3	Earth and Space Sci.
617084	4	Grade 3 to Grade 4	790	536	S4.B.1.1.1	Biological Sci.
615625	4	Grade 3 to Grade 4	791	539	S4.A.1.3.1	Nature of Science
617233	4	Grade 3 to Grade 4	780	535	S4.C.2.1.2	Physical Sci.
615632	4	Grade 3 to Grade 4	782	534	S4.A.1.3.5	Nature of Science
617245	4	Grade 3 to Grade 4	780	536	S4.D.1.1.1	Earth and Space Sci.
617096	4	Grade 3 to Grade 4	780	1092	S4.B.2.1.2	Biological Sci.
615627	4	Grade 3 to Grade 4	781	528	S4.A.1.3.2	Nature of Science
617255	4	Grade 3 to Grade 4	779	538	S4.D.1.2.3	Earth and Space Sci.
617101	4	Grade 3 to Grade 4	778	540	S4.B.3.1.1	Biological Sci.

Table C–31 (continued). Science Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617253	4	Grade 3 to Grade 4	779	559	S4.D.1.2.2	Earth and Space Sci.
617071	4	Grade 3 to Grade 4	779	531	S4.A.3.1.4	Nature of Science
617091	4	Grade 3 to Grade 4	779	529	S4.B.1.1.5	Biological Sci.

Table C–32. Science Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617231	4	Grade 4 to Grade 5	1099	608	S4.C.2.1.1	Physical Sci.
617060	4	Grade 4 to Grade 5	527	606	S4.A.2.1.3	Nature of Science
617092	4	Grade 4 to Grade 5	524	607	S4.B.1.1.5	Biological Sci.
617074	4	Grade 4 to Grade 5	528	608	S4.A.3.2.2	Nature of Science
617246	4	Grade 4 to Grade 5	537	606	S4.D.1.1.2	Earth and Space Sci.
617237	4	Grade 4 to Grade 5	538	607	S4.C.2.1.4	Physical Sci.
617068	4	Grade 4 to Grade 5	536	607	S4.A.3.1.3	Nature of Science
617102	4	Grade 4 to Grade 5	534	604	S4.B.3.1.2	Biological Sci.
617075	4	Grade 4 to Grade 5	557	606	S4.A.3.2.2	Nature of Science
617259	4	Grade 4 to Grade 5	523	604	S4.D.1.3.3	Earth and Space Sci.
617072	4	Grade 4 to Grade 5	539	599	S4.A.3.2.1	Nature of Science
617240	4	Grade 4 to Grade 5	540	600	S4.C.3.1.2	Physical Sci.
617112	4	Grade 4 to Grade 5	533	600	S4.B.3.3.3	Biological Sci.
617080	4	Grade 4 to Grade 5	533	601	S4.A.3.3.1	Nature of Science
617257	4	Grade 4 to Grade 5	538	600	S4.D.1.3.1	Earth and Space Sci.
617271	4	Grade 4 to Grade 5	533	600	S4.D.3.1.3	Earth and Space Sci.
617089	4	Grade 4 to Grade 5	534	600	S4.B.1.1.4	Biological Sci.
617234	4	Grade 4 to Grade 5	527	600	S4.C.2.1.3	Physical Sci.
617070	4	Grade 4 to Grade 5	537	599	S4.A.3.1.4	Nature of Science
617260	4	Grade 4 to Grade 5	531	599	S4.D.1.3.3	Earth and Space Sci.
617311	5	Grade 4 to Grade 5	532	604	S5.B.1.1.2	Biological Sci.
616317	5	Grade 4 to Grade 5	533	609	S5.A.1.1.2	Nature of Science
615950	5	Grade 4 to Grade 5	532	616	S5.B.2.1.1	Biological Sci.
617328	5	Grade 4 to Grade 5	532	610	S5.C.3.2.1	Physical Sci.
617304	5	Grade 4 to Grade 5	533	598	S5.A.2.1.2	Nature of Science
615962	5	Grade 4 to Grade 5	533	606	S5.D.3.1.1	Earth and Space Sci.
615936	5	Grade 4 to Grade 5	533	633	S5.A.1.1.2	Nature of Science
617330	5	Grade 4 to Grade 5	532	636	S5.D.1.1.1	Earth and Space Sci.
615958	5	Grade 4 to Grade 5	532	629	S5.C.1.2.1	Physical Sci.
617307	5	Grade 4 to Grade 5	528	635	S5.A.2.2.1	Nature of Science
617338	5	Grade 4 to Grade 5	540	617	S5.D.1.2.2	Earth and Space Sci.
615939	5	Grade 4 to Grade 5	538	610	S5.A.2.1.1	Nature of Science
617504	5	Grade 4 to Grade 5	541	630	S5.B.3.2.2	Biological Sci.
616969	5	Grade 4 to Grade 5	541	637	S5.C.2.1.1	Physical Sci.
615943	5	Grade 4 to Grade 5	538	627	S5.B.1.1.1	Biological Sci.
617502	5	Grade 4 to Grade 5	539	616	S5.B.2.1.3	Biological Sci.
617499	5	Grade 4 to Grade 5	540	614	S5.A.1.1.3	Nature of Science

Table C–32 (continued). Science Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
615965	5	Grade 4 to Grade 5	540	608	S5.D.1.1.1	Earth and Space Sci.
615942	5	Grade 4 to Grade 5	539	608	S5.A.3.1.1	Nature of Science
617507	5	Grade 4 to Grade 5	539	607	S5.C.2.1.2	Physical Sci.

Table C–33. Science Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617334	5	Grade 5 to Grade 6	605	621	S5.C.2.1.4	Physical Sci.
615949	5	Grade 5 to Grade 6	629	622	S5.B.1.1.3	Biological Sci.
615938	5	Grade 5 to Grade 6	608	622	S5.A.2.1.1	Nature of Science
615963	5	Grade 5 to Grade 6	617	623	S5.D.3.1.2	Earth and Space Sci.
615946	5	Grade 5 to Grade 6	617	621	S5.B.1.1.3	Biological Sci.
616968	5	Grade 5 to Grade 6	608	620	S5.C.1.2.2	Physical Sci.
617725	5	Grade 5 to Grade 6	602	620	S5.A.2.2.2	Nature of Science
616319	5	Grade 5 to Grade 6	637	618	S5.C.1.1.2	Physical Sci.
617318	5	Grade 5 to Grade 6	629	618	S5.B.3.1.2	Biological Sci.
616970	5	Grade 5 to Grade 6	637	617	S5.C.2.1.1	Physical Sci.
617339	5	Grade 5 to Grade 6	602	624	S5.D.1.2.1	Earth and Space Sci.
617729	5	Grade 5 to Grade 6	1215	623	S5.B.2.1.4	Biological Sci.
617501	5	Grade 5 to Grade 6	606	625	S5.A.1.1.3	Nature of Science
617342	5	Grade 5 to Grade 6	616	627	S5.D.2.1.2	Earth and Space Sci.
617310	5	Grade 5 to Grade 6	628	626	S5.A.3.2.1	Nature of Science
617326	5	Grade 5 to Grade 6	636	625	S5.C.2.1.4	Physical Sci.
617305	5	Grade 5 to Grade 6	617	625	S5.A.2.1.2	Nature of Science
617323	5	Grade 5 to Grade 6	1219	626	S5.C.1.1.1	Physical Sci.
617312	5	Grade 5 to Grade 6	634	618	S5.B.1.1.2	Biological Sci.
617327	5	Grade 5 to Grade 6	629	609	S5.C.2.1.4	Physical Sci.
615560	6	Grade 5 to Grade 6	614	623	S6.C.1.2.2	Physical Sci.
615518	6	Grade 5 to Grade 6	614	625	S6.A.2.2.1	Nature of Science
617741	6	Grade 5 to Grade 6	614	616	S6.B.2.1.2	Biological Sci.
615520	6	Grade 5 to Grade 6	614	619	S6.A.2.1.1	Nature of Science
615594	6	Grade 5 to Grade 6	614	624	S6.D.2.1.1	Earth and Space Sci.
619132	6	Grade 5 to Grade 6	614	617	S6.C.2.1.3	Physical Sci.
615554	6	Grade 5 to Grade 6	613	625	S6.B.3.2.1	Biological Sci.
615557	6	Grade 5 to Grade 6	613	620	S6.C.1.2.1	Physical Sci.
615514	6	Grade 5 to Grade 6	614	624	S6.A.1.1.3	Nature of Science
615603	6	Grade 5 to Grade 6	612	616	S6.D.3.1.2	Earth and Space Sci.
615574	6	Grade 5 to Grade 6	613	620	S6.C.2.1.3	Physical Sci.
618591	6	Grade 5 to Grade 6	612	625	S6.A.1.2.2	Nature of Science
615532	6	Grade 5 to Grade 6	612	621	S6.B.2.1.2	Biological Sci.
619296	6	Grade 5 to Grade 6	611	625	S6.A.2.1.1	Nature of Science
615601	6	Grade 5 to Grade 6	610	616	S6.D.3.1.1	Earth and Space Sci.
617512	6	Grade 5 to Grade 6	610	625	S6.C.2.1.1	Physical Sci.
615540	6	Grade 5 to Grade 6	610	624	S6.B.3.1.1	Biological Sci.

Table C–33 (continued). Science Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617508	6	Grade 5 to Grade 6	608	619	S6.B.1.1.1	Biological Sci.
615526	6	Grade 5 to Grade 6	608	620	S6.A.3.2.1	Nature of Science
619365	6	Grade 5 to Grade 6	608	618	S6.D.2.1.1	Earth and Space Sci.

Table C–34. Science Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
615535	6	Grade 6 to Grade 7	1248	428	S6.A.3.2.1	Nature of Science
615562	6	Grade 6 to Grade 7	620	428	S6.C.1.2.2	Physical Sci.
615530	6	Grade 6 to Grade 7	1234	428	S6.B.2.1.1	Biological Sci.
619141	6	Grade 6 to Grade 7	616	426	S6.D.2.1.3	Earth and Space Sci.
615510	6	Grade 6 to Grade 7	1253	425	S6.A.1.1.2	Nature of Science
618609	6	Grade 6 to Grade 7	625	426	S6.C.3.1.2	Physical Sci.
618590	6	Grade 6 to Grade 7	1243	425	S6.A.1.2.1	Nature of Science
615576	6	Grade 6 to Grade 7	621	424	S6.C.2.1.3	Physical Sci.
615551	6	Grade 6 to Grade 7	621	424	S6.C.1.2.1	Physical Sci.
615512	6	Grade 6 to Grade 7	1233	423	S6.A.1.1.3	Nature of Science
615577	6	Grade 6 to Grade 7	619	428	S6.C.3.1.1	Physical Sci.
618791	6	Grade 6 to Grade 7	1235	428	S6.A.1.2.1	Nature of Science
615531	6	Grade 6 to Grade 7	1225	428	S6.B.2.1.1	Biological Sci.
619624	6	Grade 6 to Grade 7	627	428	S6.D.3.1.2	Earth and Space Sci.
616332	6	Grade 6 to Grade 7	1228	426	S6.A.1.1.3	Nature of Science
619149	6	Grade 6 to Grade 7	618	425	S6.C.3.2.1	Physical Sci.
617533	6	Grade 6 to Grade 7	1249	427	S6.B.2.1.1	Biological Sci.
618794	6	Grade 6 to Grade 7	624	426	S6.C.3.2.1	Physical Sci.
615517	6	Grade 6 to Grade 7	1245	426	S6.A.1.2.2	Nature of Science
615567	6	Grade 6 to Grade 7	616	425	S6.C.2.1.1	Physical Sci.
616616	7	Grade 6 to Grade 7	619	428	S7.D.1.1.2	Earth and Space Sci.
615235	7	Grade 6 to Grade 7	619	430	S7.B.1.1.2	Biological Sci.
617184	7	Grade 6 to Grade 7	616	424	S7.A.1.1.1	Nature of Science
618806	7	Grade 6 to Grade 7	618	427	S7.D.2.1.1	Earth and Space Sci.
615974	7	Grade 6 to Grade 7	618	443	S7.A.1.2.1	Nature of Science
618603	7	Grade 6 to Grade 7	617	439	S7.C.2.1.3	Physical Sci.
615973	7	Grade 6 to Grade 7	617	424	S7.A.1.1.4	Nature of Science
615275	7	Grade 6 to Grade 7	614	870	S7.B.3.3.2	Biological Sci.
615238	7	Grade 6 to Grade 7	609	427	S7.B.1.1.3	Biological Sci.
618802	7	Grade 6 to Grade 7	606	430	S7.C.2.1.1	Physical Sci.
617531	7	Grade 6 to Grade 7	624	424	S7.D.1.1.2	Earth and Space Sci.
616339	7	Grade 6 to Grade 7	626	431	S7.A.2.2.3	Nature of Science
615970	7	Grade 6 to Grade 7	625	429	S7.A.1.1.2	Nature of Science
616626	7	Grade 6 to Grade 7	625	443	S7.D.3.1.1	Earth and Space Sci.
617195	7	Grade 6 to Grade 7	626	444	S7.A.1.3.1	Nature of Science
617526	7	Grade 6 to Grade 7	624	422	S7.C.1.2.2	Physical Sci.
619627	7	Grade 6 to Grade 7	625	428	S7.A.1.1.4	Nature of Science

Table C–34 (continued). Science Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
615252	7	Grade 6 to Grade 7	624	444	S7.B.2.1.3	Biological Sci.
615234	7	Grade 6 to Grade 7	620	427	S7.B.1.1.1	Biological Sci.
616039	7	Grade 6 to Grade 7	618	424	S7.C.2.1.3	Physical Sci.

Table C–35. Science Items Used to Link Grade 8 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617198	7	Grade 8 to Grade 7	431	256	S7.A.1.3.2	Nature of Science
616619	7	Grade 8 to Grade 7	426	256	S7.D.1.2.2	Earth and Space Sci.
615969	7	Grade 8 to Grade 7	427	255	S7.A.1.1.1	Nature of Science
616038	7	Grade 8 to Grade 7	424	256	S7.C.2.1.2	Physical Sci.
616622	7	Grade 8 to Grade 7	427	254	S7.D.2.1.1	Earth and Space Sci.
615971	7	Grade 8 to Grade 7	429	254	S7.A.1.1.3	Nature of Science
615249	7	Grade 8 to Grade 7	425	255	S7.B.2.1.2	Biological Sci.
618803	7	Grade 8 to Grade 7	432	254	S7.D.2.1.1	Earth and Space Sci.
618801	7	Grade 8 to Grade 7	427	252	S7.C.2.1.3	Physical Sci.
615999	7	Grade 8 to Grade 7	423	251	S7.B.1.1.3	Biological Sci.
615308	7	Grade 8 to Grade 7	422	253	S7.C.3.1.3	Physical Sci.
618855	7	Grade 8 to Grade 7	430	254	S7.A.2.1.1	Nature of Science
618853	7	Grade 8 to Grade 7	425	254	S7.A.1.3.1	Nature of Science
616348	7	Grade 8 to Grade 7	438	254	S7.B.2.2.2	Biological Sci.
616621	7	Grade 8 to Grade 7	426	254	S7.D.1.2.3	Earth and Space Sci.
617000	7	Grade 8 to Grade 7	441	254	S7.D.3.1.3	Earth and Space Sci.
616014	7	Grade 8 to Grade 7	419	254	S7.B.3.1.1	Biological Sci.
617196	7	Grade 8 to Grade 7	441	252	S7.A.1.3.1	Nature of Science
616313	7	Grade 8 to Grade 7	430	251	S7.C.3.1.1	Physical Sci.
616007	7	Grade 8 to Grade 7	429	252	S7.B.2.1.2	Biological Sci.
615771	8	Grade 8 to Grade 7	445	262	S8.A.3.3.2	Nature of Science
617489	8	Grade 8 to Grade 7	445	257	S8.C.3.1.1	Physical Sci.
615784	8	Grade 8 to Grade 7	444	262	S8.B.2.1.1	Biological Sci.
620362	8	Grade 8 to Grade 7	444	271	S8.D.1.2.1	Earth and Space Sci.
618535	8	Grade 8 to Grade 7	444	267	S8.A.3.2.2	Nature of Science
617484	8	Grade 8 to Grade 7	444	258	S8.D.1.1.2	Earth and Space Sci.
618896	8	Grade 8 to Grade 7	443	272	S8.D.1.3.2	Earth and Space Sci.
615776	8	Grade 8 to Grade 7	443	255	S8.B.1.1.2	Biological Sci.
618543	8	Grade 8 to Grade 7	442	264	S8.C.2.2.2	Physical Sci.
617735	8	Grade 8 to Grade 7	441	287	S8.A.2.1.2	Nature of Science
617294	8	Grade 8 to Grade 7	432	262	S8.D.2.1.3	Earth and Space Sci.
617289	8	Grade 8 to Grade 7	432	255	S8.B.2.2.1	Biological Sci.
618544	8	Grade 8 to Grade 7	432	260	S8.C.2.2.2	Physical Sci.
620027	8	Grade 8 to Grade 7	432	289	S8.A.3.1.5	Nature of Science
617962	8	Grade 8 to Grade 7	432	259	S8.A.1.3.4	Nature of Science
615810	8	Grade 8 to Grade 7	432	267	S8.C.2.1.1	Physical Sci.
617279	8	Grade 8 to Grade 7	432	258	S8.B.1.1.1	Biological Sci.

Table C–35 (continued). Science Items Used to Link Grade 8 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category
617293	8	Grade 8 to Grade 7	430	286	S8.D.2.1.3	Earth and Space Sci.
620020	8	Grade 8 to Grade 7	430	256	S8.A.1.1.2	Nature of Science
620400	8	Grade 8 to Grade 7	430	255	S8.B.3.2.3	Biological Sci.

Table C–36. Science Items Used to Link Biology to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category	Biology Diagnostic Category
615777	8	Biology to Grade 8	261	306	S8.B.1.1.3	Biological Sci.	Basic Bio. Princ.
615790	8	Biology to Grade 8	259	306	S8.B.2.1.3	Biological Sci.	Cell Growth
615817	8	Biology to Grade 8	519	306	S8.C.2.1.3	Physical Sci.	No Biology DC
620364	8	Biology to Grade 8	256	305	S8.D.1.3.1	Earth and Space Sci.	Theory of Evolution
617739	8	Biology to Grade 8	288	304	S8.A.2.1.4	Nature of Science	No Biology DC
615789	8	Biology to Grade 8	257	303	S8.B.2.1.2	Biological Sci.	Theory of Evolution
618786	8	Biology to Grade 8	257	305	S8.A.3.2.3	Nature of Science	No Biology DC
617059	8	Biology to Grade 8	266	306	S8.B.1.1.1	Biological Sci.	Basic Bio. Princ.
615791	8	Biology to Grade 8	529	305	S8.B.2.1.3	Biological Sci.	Cell Growth
617284	8	Biology to Grade 8	259	305	S8.B.2.1.3	Biological Sci.	Cell Growth
620015	8	Biology to Grade 8	254	298	S8.A.1.1.1	Nature of Science	No Biology DC
620396	8	Biology to Grade 8	256	298	S8.B.3.2.2	Biological Sci.	Theory of Evolution
617737	8	Biology to Grade 8	252	298	S8.A.2.1.3	Nature of Science	No Biology DC
617292	8	Biology to Grade 8	255	297	S8.B.2.2.2	Biological Sci.	Cell Growth
615822	8	Biology to Grade 8	542	298	S8.C.2.2.3	Physical Sci.	Theory of Evolution
620637	8	Biology to Grade 8	262	298	S8.B.3.1.3	Biological Sci.	Theory of Evolution
618540	8	Biology to Grade 8	259	298	S8.A.3.3.1	Nature of Science	No Biology DC
618548	8	Biology to Grade 8	260	298	S8.D.1.3.4	Earth and Space Sci.	Theory of Evolution
620029	8	Biology to Grade 8	522	298	S8.A.3.2.3	Nature of Science	No Biology DC
620401	8	Biology to Grade 8	259	298	S8.B.3.2.3	Biological Sci.	Theory of Evolution
617377	Bio	Biology to Grade 8	257	305	BIO.A.4.2.1	Biological Sci.	Bioenergetics
617565	Bio	Biology to Grade 8	256	311	BIO.B.4.2.5	Biological Sci.	Theory of Evolution
616111	Bio	Biology to Grade 8	256	303	BIO.A.1.2.1	Biological Sci.	Basic Bio. Princ.
617401	Bio	Biology to Grade 8	256	4874	BIO.B.2.1.1	Biological Sci.	Cell Growth
617430	Bio	Biology to Grade 8	256	309	BIO.B.3.1.1	Biological Sci.	Theory of Evolution
617395	Bio	Biology to Grade 8	256	310	BIO.B.1.2.2	Biological Sci.	Cell Growth
617013	Bio	Biology to Grade 8	257	311	BIO.A.2.2.3	Biological Sci.	Basic Bio. Princ.
617444	Bio	Biology to Grade 8	257	311	BIO.B.3.2.1	Biological Sci.	Theory of Evolution
617458	Bio	Biology to Grade 8	256	295	BIO.B.4.1.2	Biological Sci.	Theory of Evolution
617449	Bio	Biology to Grade 8	256	311	BIO.B.3.3.1	Biological Sci.	Theory of Evolution
617839	Bio	Biology to Grade 8	263	300	BIO.A.4.2.1	Biological Sci.	Bioenergetics
617462	Bio	Biology to Grade 8	263	297	BIO.B.3.3.1	Biological Sci.	Theory of Evolution
616112	Bio	Biology to Grade 8	263	305	BIO.A.1.2.1	Biological Sci.	Basic Bio. Princ.
617457	Bio	Biology to Grade 8	263	4863	BIO.B.4.1.2	Biological Sci.	Theory of Evolution
617394	Bio	Biology to Grade 8	262	296	BIO.B.1.2.2	Biological Sci.	Cell Growth
617454	Bio	Biology to Grade 8	263	310	BIO.B.4.1.1	Biological Sci.	Theory of Evolution
617349	Bio	Biology to Grade 8	263	309	BIO.A.3.1.1	Biological Sci.	Bioenergetics

Table C–36 (continued). Science Items Used to Link Biology to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category	Biology Diagnostic Category
617414	Bio	Biology to Grade 8	263	300	BIO.B.2.2.2	Biological Sci.	Cell Growth
617880	Bio	Biology to Grade 8	263	305	BIO.B.2.2.2	Biological Sci.	Cell Growth
617451	Bio	Biology to Grade 8	263	298	BIO.B.3.3.1	Biological Sci.	Theory of Evolution

Table C–37. Science Items Used to Link Chemistry to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category	Chemistry Diagnostic Category
615817	8	Chemistry to Grade 8	519	305	S8.C.2.1.3	Physical Sci.	Properties of Matter
615822	8	Chemistry to Grade 8	542	304	S8.C.2.2.3	Physical Sci.	No Chemistry DC
620029	8	Chemistry to Grade 8	522	307	S8.A.3.2.3	Nature of Science	No Chemistry DC
620025	8	Chemistry to Grade 8	258	308	S8.A.2.1.1	Nature of Science	No Chemistry DC
615819	8	Chemistry to Grade 8	261	308	S8.C.2.2.1	Physical Sci.	No Chemistry DC
620021	8	Chemistry to Grade 8	262	308	S8.A.1.1.3	Nature of Science	No Chemistry DC
615833	8	Chemistry to Grade 8	265	306	S8.D.1.1.2	Earth and Space Sci.	No Chemistry DC
615749	8	Chemistry to Grade 8	259	307	S8.A.2.2.3	Nature of Science	No Chemistry DC
620426	8	Chemistry to Grade 8	253	306	S8.B.3.3.4	Biological Sci.	No Chemistry DC
615723	8	Chemistry to Grade 8	270	305	S8.A.1.3.3	Nature of Science	No Chemistry DC
615809	8	Chemistry to Grade 8	511	307	S8.C.1.1.3	Physical Sci.	Chem. Relation.
615884	8	Chemistry to Grade 8	253	306	S8.A.2.1.1	Nature of Science	No Chemistry DC
615919	8	Chemistry to Grade 8	260	306	S8.C.1.1.1	Physical Sci.	Mole
620030	8	Chemistry to Grade 8	258	307	S8.A.3.2.3	Nature of Science	No Chemistry DC
620427	8	Chemistry to Grade 8	287	304	S8.B.3.3.4	Biological Sci.	No Chemistry DC
615927	8	Chemistry to Grade 8	266	305	S8.A.1.3.1	Nature of Science	No Chemistry DC
615826	8	Chemistry to Grade 8	262	306	S8.C.3.1.2	Physical Sci.	No Chemistry DC
620023	8	Chemistry to Grade 8	262	305	S8.A.1.3.2	Nature of Science	No Chemistry DC
615857	8	Chemistry to Grade 8	267	304	S8.D.2.1.1	Earth and Space Sci.	No Chemistry DC
615804	8	Chemistry to Grade 8	259	306	S8.C.1.1.1	Physical Sci.	Mole
616406	Chem	Chemistry to Grade 8	258	305	CHEM.A.2.1.2	Physical Sci.	Atomic Structure

Table C–37 (continued). Science Items Used to Link Chemistry to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Science Diagnostic Category	Chemistry Diagnostic Category
618699	Chem	Chemistry to Grade 8	259	302	CHEM.B.2.1.5	Physical Sci.	Chem. Relation.
616511	Chem	Chemistry to Grade 8	259	299	CHEM.B.1.4.1	Physical Sci.	Mole
616362	Chem	Chemistry to Grade 8	258	303	CHEM.A.1.1.2	Physical Sci.	Properties of Matter
618734	Chem	Chemistry to Grade 8	259	307	CHEM.B.2.1.4	Physical Sci.	Chem. Relation.
616367	Chem	Chemistry to Grade 8	259	615	CHEM.A.1.2.2	Physical Sci.	Properties of Matter
616559	Chem	Chemistry to Grade 8	259	305	CHEM.A.1.1.5	Physical Sci.	Properties of Matter
619910	Chem	Chemistry to Grade 8	259	306	CHEM.B.1.4.2	Physical Sci.	Mole
616494	Chem	Chemistry to Grade 8	259	305	CHEM.A.1.2.3	Physical Sci.	Properties of Matter
616518	Chem	Chemistry to Grade 8	259	304	CHEM.B.2.1.5	Physical Sci.	Chem. Relation.
616427	Chem	Chemistry to Grade 8	260	306	CHEM.A.1.1.1	Physical Sci.	Properties of Matter
618726	Chem	Chemistry to Grade 8	260	309	CHEM.B.1.3.1	Physical Sci.	Mole
616365	Chem	Chemistry to Grade 8	260	301	CHEM.A.1.1.5	Physical Sci.	Properties of Matter
616516	Chem	Chemistry to Grade 8	260	306	CHEM.B.2.1.3	Physical Sci.	Chem. Relation.
618733	Chem	Chemistry to Grade 8	260	307	CHEM.B.2.1.3	Physical Sci.	Chem. Relation.
620468	Chem	Chemistry to Grade 8	260	315	CHEM.B.2.1.1	Physical Sci.	Chem. Relation.
616561	Chem	Chemistry to Grade 8	260	307	CHEM.A.1.2.2	Physical Sci.	Properties of Matter
616376	Chem	Chemistry to Grade 8	259	304	CHEM.A.2.3.1	Physical Sci.	Atomic Structure
616533	Chem	Chemistry to Grade 8	259	306	CHEM.A.2.2.2	Physical Sci.	Atomic Structure
618698	Chem	Chemistry to Grade 8	259	302	CHEM.B.2.1.4	Physical Sci.	Chem. Relation.

Tables C–38 through C–44 summarize the number of linking items by diagnostic category. Items coded in a Science diagnostic category and a Biology or Chemistry diagnostic category are noted.

Table C–38. Number of Items Linking Grade 3 to Grade 4 by Diagnostic Category

Diagnostic Category	Grade 3 Items	Grade 4 Items	Total
Nature of Science	6	6	12
Biological Sciences	7	6	13
Physical Sciences	2	3	5
Earth and Space Sciences	5	5	10
TOTAL	20	20	40

Table C–39. Number of Items Linking Grade 4 to Grade 5 by Diagnostic Category

Diagnostic Category	Grade 4 Items	Grade 5 Items	Total
Nature of Science	7	7	14
Biological Sciences	4	5	9
Physical Sciences	4	4	8
Earth and Space Sciences	5	4	9
TOTAL	20	20	40

Table C–40. Number of Items Linking Grade 5 to Grade 6 by Diagnostic Category

Diagnostic Category	Grade 5 Items	Grade 6 Items	Total
Nature of Science	5	6	11
Biological Sciences	5	5	10
Physical Sciences	7	5	12
Earth and Space Sciences	3	4	7
TOTAL	20	20	40

Table C–41. Number of Items Linking Grade 6 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 6 Items	Grade 7 Items	Total
Nature of Science	7	7	14
Biological Sciences	3	5	8
Physical Sciences	8	4	12
Earth and Space Sciences	2	4	6
TOTAL	20	20	40

Table C–42. Number of Items Linking Grade 8 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 7 Items	Grade 8 Items	Total
Nature of Science	6	6	12
Biological Sciences	5	5	10
Physical Sciences	4	4	8
Earth and Space Sciences	5	5	10
TOTAL	20	20	40

Table C–43a. Number of Items Linking Biology to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Biology Items	Total
Nature of Science	6	0	6
Biological Sciences	10	20	30
Physical Sciences	2	0	2
Earth and Space Sciences	2	0	2
No Grade 8 DC	0	0	0
TOTAL	20	20	40

Table C–43b. Number of Items Linking Biology to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Biology Items	Total
Basic Biological Principles	2	3	5
Bioenergetics	0	3	3
Cell Growth	4	5	9
Theory of Evolution	7	9	16
No Biology DC	7	0	7
TOTAL	20	20	40

Table C–44a. Number of Items Linking Chemistry to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Chemistry Items	Total
Nature of Science	9	0	9
Biological Sciences	2	0	2
Physical Sciences	7	20	27
Earth and Space Sciences	2	0	2
No Grade 8 DC	0	0	0
TOTAL	20	20	40

Table C–44b. Number of Items Linking Chemistry to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Chemistry Items	Total
Properties of Matter	1	7	8
Atomic Structure	0	3	3
The Mole	2	3	5
Chemical Relationships	1	7	8
No Chemistry DC	16	0	16
TOTAL	20	20	40

Table C–45. Science Example of Vertical Linking Workbook

		Grade 4 Calibration			Grade 5 Calibration				Grade 4 on		
Item ID	Item Grade	Difficulty	Fit	Displace	Difficulty	Fit	Displace	Discrepancy	Grade 5 Scale	Robust Z	Flag
617231	4	-0.669	0.980	0.001	-1.440	1.040	-0.004	-0.771	-1.442	-0.097	
617060	4	0.409	1.030	-0.002	-0.267	1.050	-0.003	-0.676	-0.364	0.312	
617092	4	-0.519	1.040	-0.002	-1.314	0.930	-0.004	-0.795	-1.292	-0.200	
617074	4	-0.048	0.950	-0.002	-0.773	1.000	-0.003	-0.725	-0.821	0.101	
617246	4	0.952	0.930	0.000	-0.093	0.900	-0.003	-1.045	0.179	-1.275	
617237	4	0.497	0.970	0.000	-0.250	0.950	-0.003	-0.747	-0.276	0.006	
617068	4	-0.016	1.030	0.002	-0.396	0.980	-0.003	-0.380	-0.789	1.585	
617102	4	2.758	1.090	-0.006	1.678	1.100	-0.003	-1.080	1.985	-1.426	
617075	4	0.654	1.030	-0.001	0.375	1.010	-0.003	-0.279	-0.119	2.019	high robust Z
617259	4	1.107	1.120	-0.001	0.532	1.070	-0.003	-0.575	0.334	0.746	
617072	4	0.683	0.970	0.004	-0.653	0.950	-0.002	-1.336	-0.090	-2.526	high robust Z
617240	4	0.983	1.080	0.004	0.131	1.100	-0.002	-0.852	0.210	-0.445	
617112	4	0.827	0.970	-0.001	0.145	0.930	-0.002	-0.682	0.054	0.286	
617080	4	1.924	1.230	-0.001	1.183	1.110	-0.002	-0.741	1.151	0.032	
617257	4	0.184	0.950	0.004	-0.368	0.960	-0.002	-0.552	-0.589	0.845	
617271	4	0.518	0.980	0.002	-0.502	0.920	-0.002	-1.020	-0.255	-1.168	
617089	4	0.146	1.140	-0.006	-0.345	1.080	-0.002	-0.491	-0.627	1.107	
617234	4	0.420	0.990	-0.002	0.000	1.060	-0.002	-0.420	-0.353	1.413	
617070	4	-0.383	0.940	0.000	-1.133	0.920	-0.002	-0.750	-1.156	-0.006	
617260	4	1.940	1.120	0.003	1.201	1.140	-0.002	-0.739	1.167	0.041	
617311	5	-0.320	1.000	0.002	-0.902	0.970	-0.001	-0.582	-1.093	0.716	
616317	5	-0.027	1.040	0.002	-0.296	1.080	0.001	-0.269	-0.800	2.062	high robust Z
615950	5	0.038	0.970	0.002	-0.902	0.920	0.001	-0.940	-0.735	-0.823	
617328	5	-0.257	0.960	0.002	-0.859	0.860	0.001	-0.602	-1.030	0.630	
617304	5	1.292	1.120	0.002	0.486	1.020	0.001	-0.806	0.519	-0.247	
615962	5	-0.868	0.940	0.002	-1.223	0.930	0.001	-0.355	-1.641	1.692	
615936	5	-0.152	0.990	0.002	-1.059	0.890	0.003	-0.907	-0.925	-0.682	
617330	5	0.732	0.940	0.002	-0.012	0.840	-0.002	-0.744	-0.041	0.019	
615958	5	0.180	1.070	0.002	-0.560	1.010	0.003	-0.740	-0.593	0.037	
617307	5	1.109	0.950	0.002	0.289	0.970	-0.002	-0.820	0.336	-0.307	
617338	5	0.456	0.940	0.005	-0.715	0.920	0.001	-1.171	-0.317	-1.817	
615939	5	0.484	0.980	0.005	-0.418	0.850	-0.002	-0.902	-0.289	-0.660	
617504	5	2.443	0.990	0.005	1.115	1.020	0.004	-1.328	1.670	-2.492	high robust Z
616969	5	-0.111	1.080	0.005	-0.812	1.030	-0.002	-0.701	-0.884	0.204	
615943	5	0.657	1.070	0.005	-0.391	0.940	0.003	-1.048	-0.116	-1.288	
617502	5	0.997	0.980	0.005	0.107	0.970	0.001	-0.890	0.224	-0.608	
617499	5	0.794	1.030	0.005	-0.130	1.000	-0.002	-0.924	0.021	-0.755	
615965	5	1.460	0.920	0.005	0.316	0.920	-0.001	-1.144	0.687	-1.701	
615942	5	-1.725	0.940	0.005	-2.577	0.940	-0.001	-0.852	-2.498	-0.445	
617507	5	0.870	1.230	0.005	0.340	1.130	0.001	-0.530	0.097	0.940	
	Mean	0.510			-0.262			-0.773	-0.262	-0.104	
	SD	0.875			0.810			0.259	0.875	1.114	
	SD Ratio	1.080									
	Correlation	0.956									
	Add. Constant	-0.773									
	Median							-0.749			
	Q							0.314			

Figures C–15 through C–21 are the adjacent grade linking plots. Items removed from final linking procedure are colored red.

Figure C–15. CDT Science: Grade 3 to Grade 4 Linking – All Links

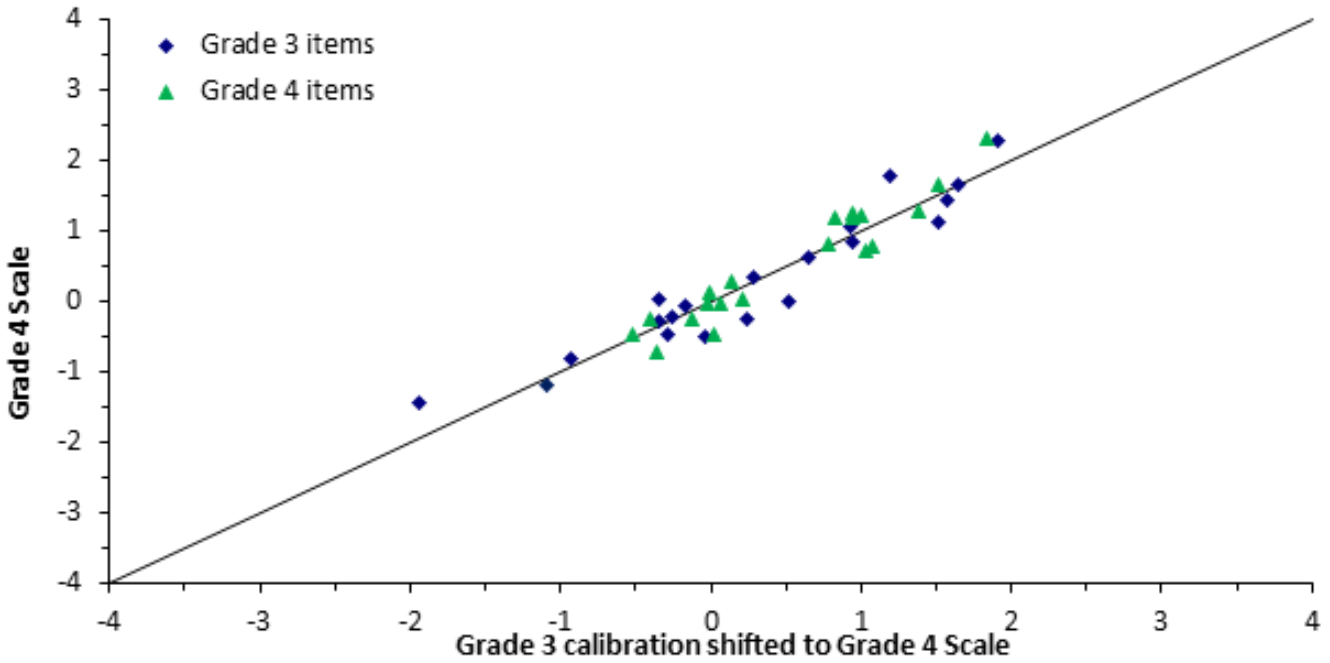


Figure C–16. CDT Science: Grade 4 to Grade 5 Linking – All Links

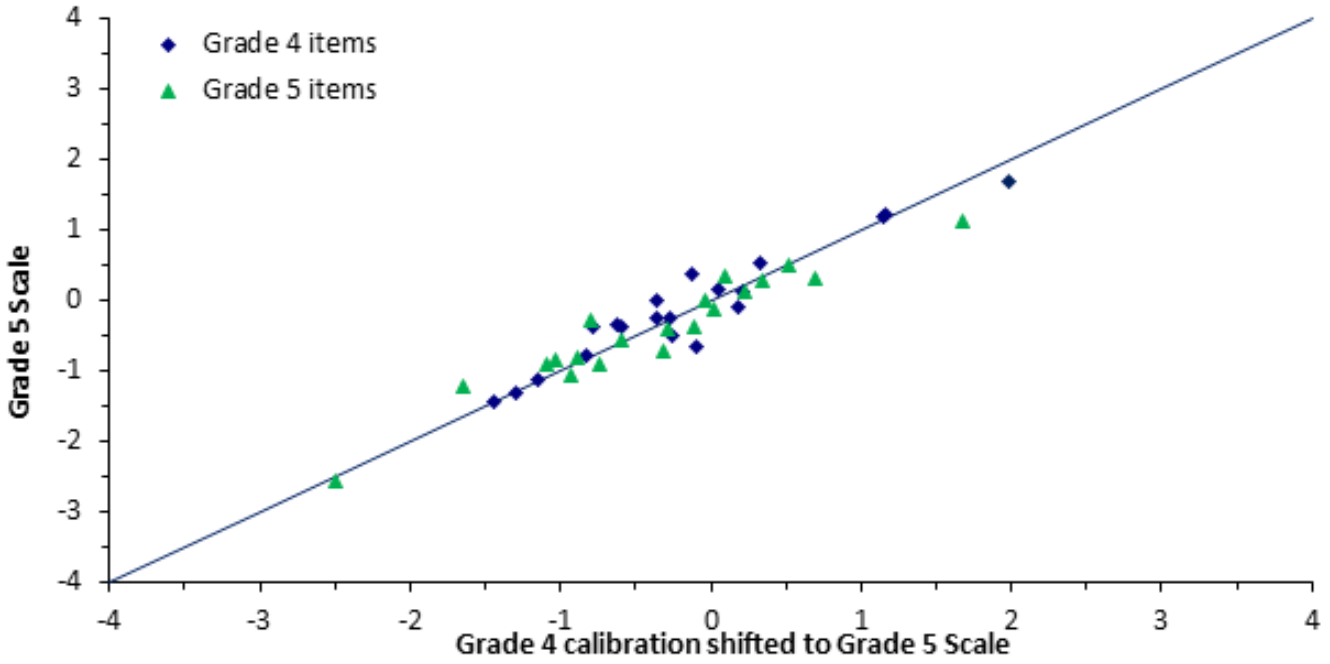


Figure C–17. CDT Science: Grade 5 to Grade 6 Linking – All Links

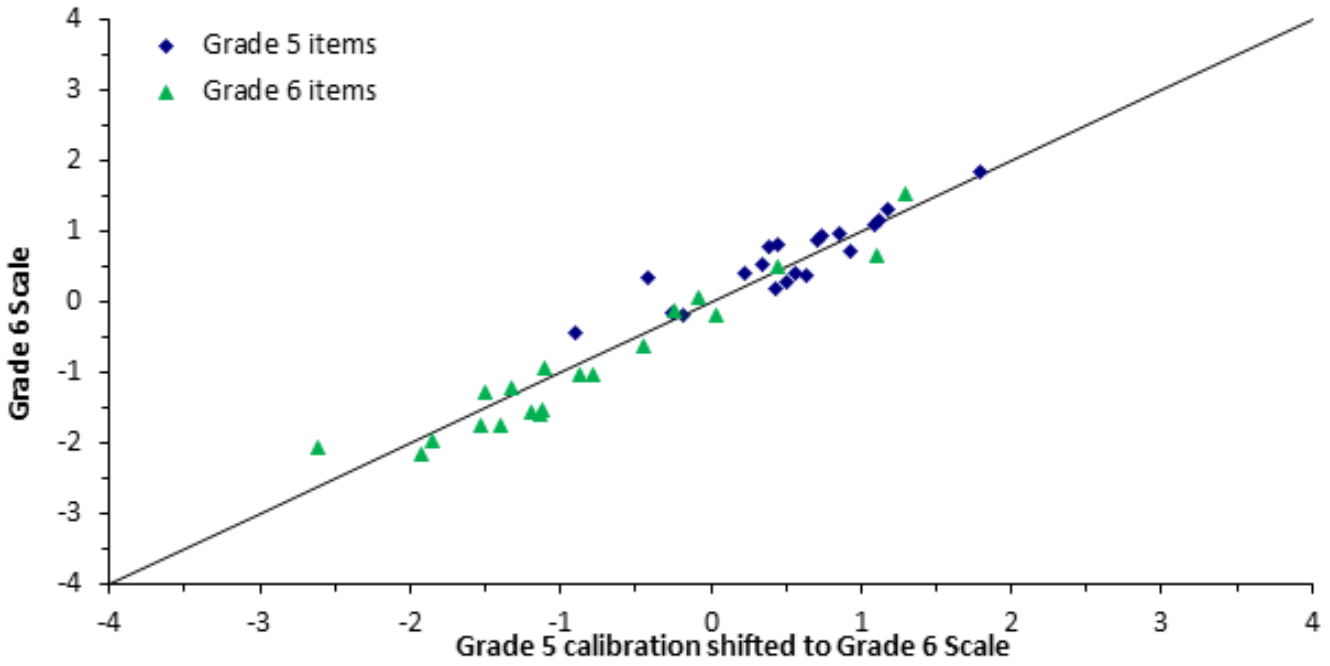


Figure C–18. CDT Science: Grade 6 to Grade 7 Linking – All Links

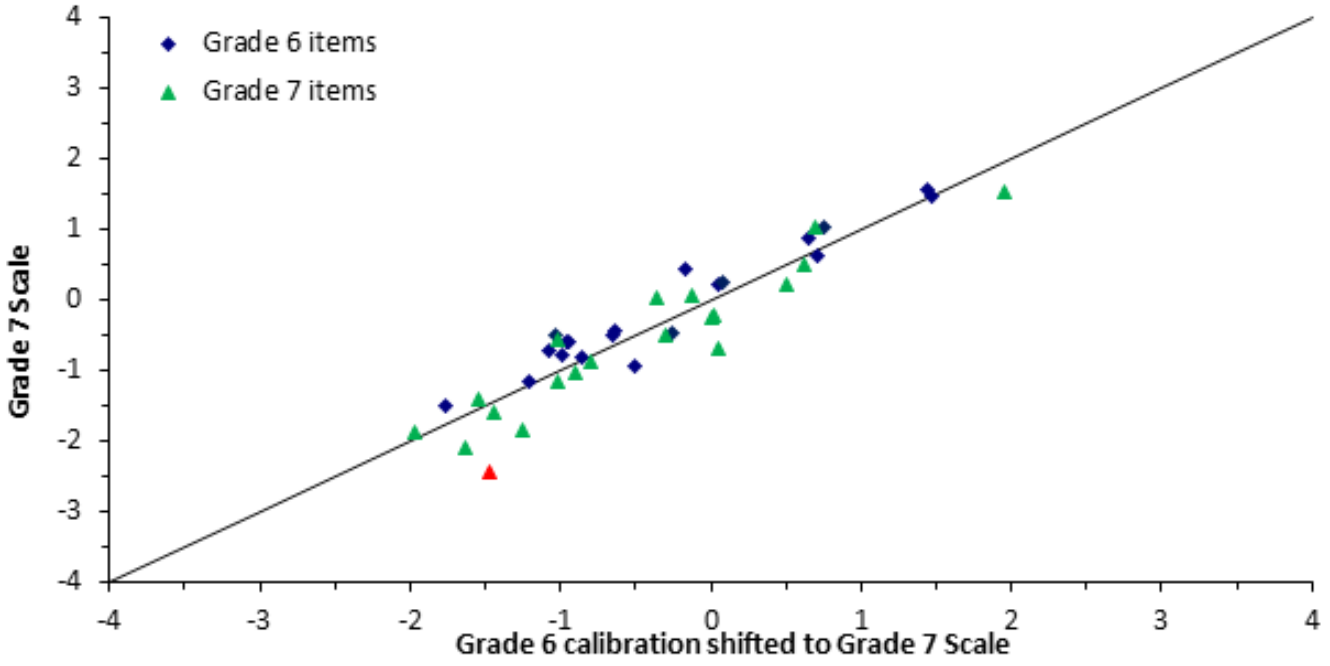


Figure C–19. CDT Science: Grade 8 to Grade 7 Linking – All Links

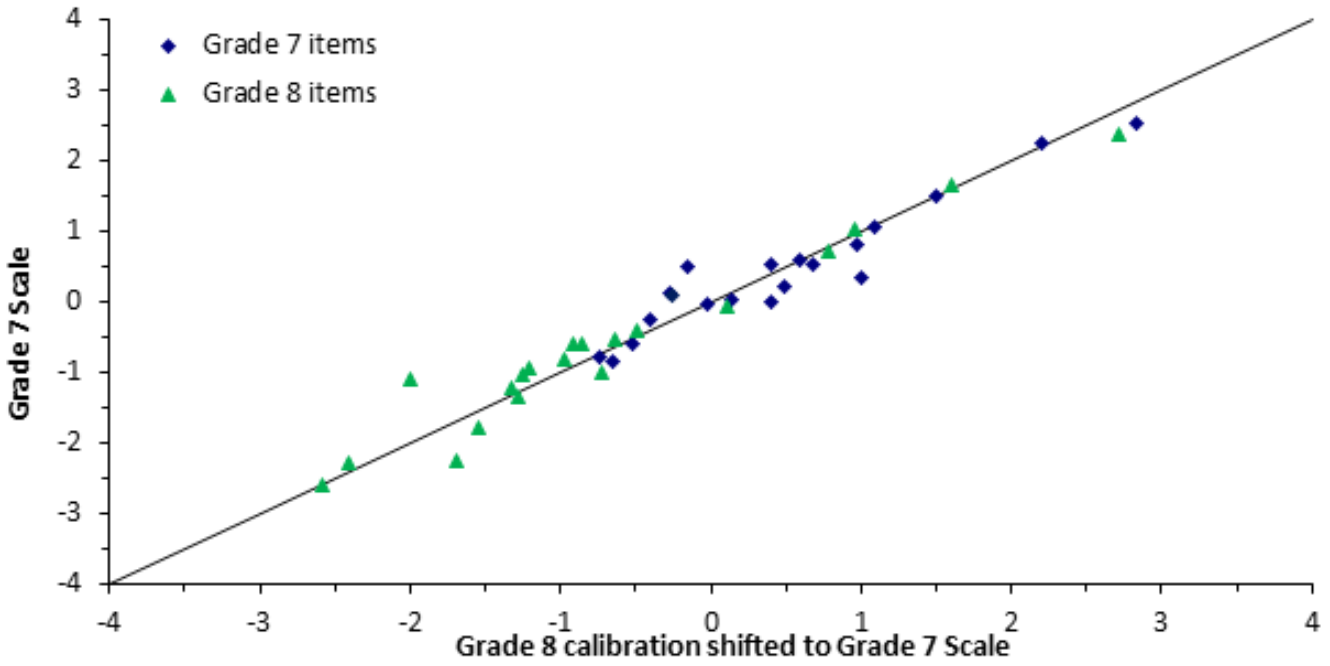


Figure C–20. CDT Science: Biology to Grade 8 Linking – All Links

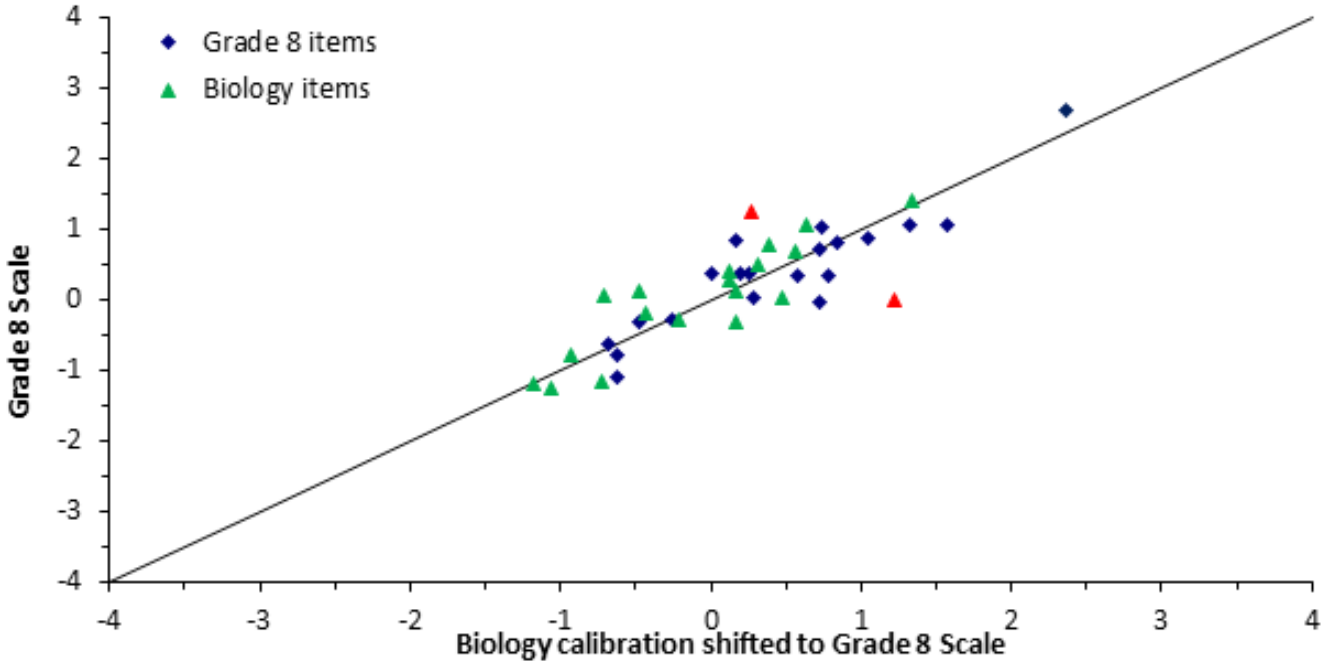
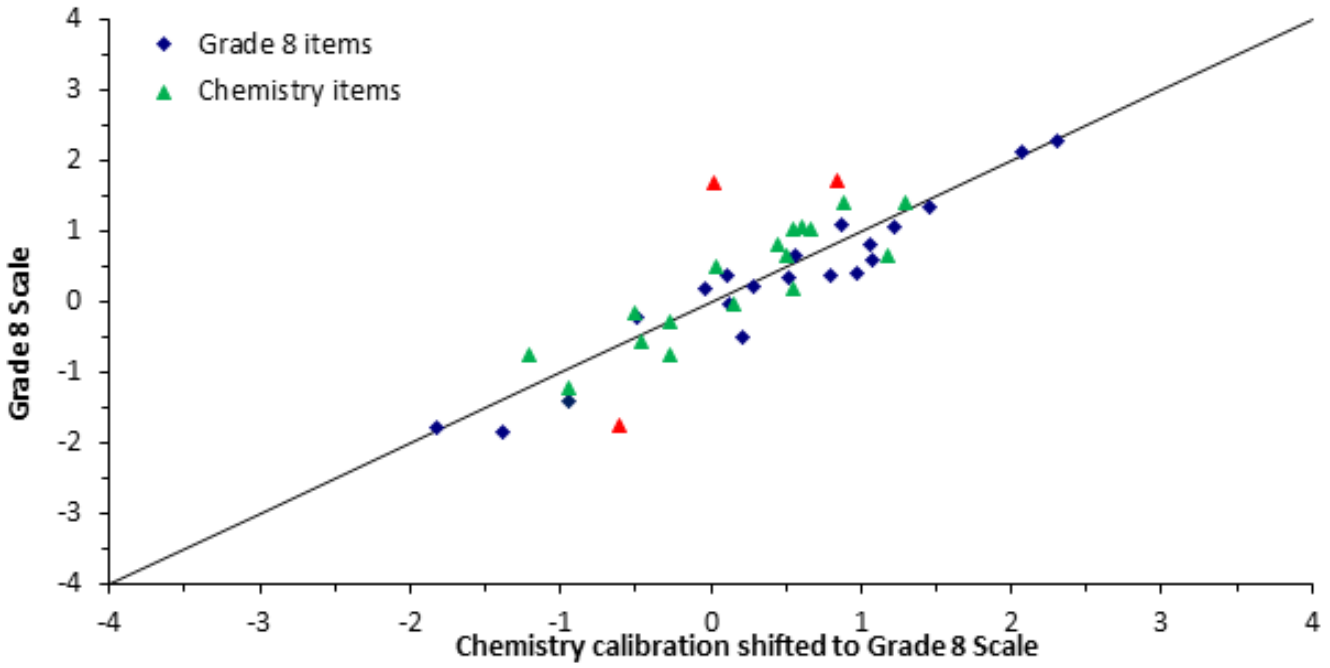


Figure C–21. CDT Science: Chemistry to Grade 8 Linking – All Links



WRITING/ENGLISH COMPOSITION

Tables C–46 through C–51 show n-counts, eligible content code, and diagnostic category for each of the vertical linking items.

Each item was administered in two grades so there are two n-counts: one for the lower grade and one for the upper grade. For example, item 626547 is a grade 3 item used to link grades 3 and 4. It was administered 274 times on the lower grade form (grade 3) and 234 times on the upper grade form (grade 4).

The diagnostic categories are[4]:

- Quality of Writing: Focus and Content
- Quality of Writing: Organization and Style
- Quality of Writing: Editing
- Conventions: Spelling, Capitalization, and Punctuation
- Conventions: Grammar and Sentence Formation

[4] Writing diagnostic categories changed at the start of the 2013–2014 school year due to re-alignment to the Pennsylvania Core Standards. See Chapter Thirteen for a list of the current diagnostic categories.

Table C–46. Writing/English Composition Items Used to Link Grade 3 to Grade 4

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
626547	3	Grade 3 to Grade 4	274	234	W.1.5.3.F.b	Spell., Cap., Punct.
621012	3	Grade 3 to Grade 4	276	234	W.1.5.3.F.d	Gram. and Sent.
634030	3	Grade 3 to Grade 4	277	234	W.1.5.3.F.a	Spell., Cap., Punct.
634160	3	Grade 3 to Grade 4	275	234	W.1.5.3.D	Org and Style
623056	3	Grade 3 to Grade 4	275	234	W.1.5.3.C	Org and Style
621006	3	Grade 3 to Grade 4	277	234	W.1.5.3.F.d	Gram. and Sent.
624801	3	Grade 3 to Grade 4	276	234	W.1.5.3.A	Focus and Content
623023	3	Grade 3 to Grade 4	274	234	W.1.5.3.F.d	Gram. and Sent.
622985	3	Grade 3 to Grade 4	274	234	W.1.5.3.B	Focus and Content
624847	3	Grade 3 to Grade 4	277	234	W.1.5.3.F.c	Spell., Cap., Punct.
624849	3	Grade 3 to Grade 4	276	232	W.1.5.3.F.b	Spell., Cap., Punct.
622465	3	Grade 3 to Grade 4	277	232	W.1.5.3.F.d	Gram. and Sent.
634029	3	Grade 3 to Grade 4	275	232	W.1.5.3.F.a	Spell., Cap., Punct.
634162	3	Grade 3 to Grade 4	275	232	W.1.5.3.D	Org and Style
626574	3	Grade 3 to Grade 4	277	232	W.1.5.3.C	Org and Style
636550	3	Grade 3 to Grade 4	276	232	W.1.5.3.F.d	Gram. and Sent.
622979	3	Grade 3 to Grade 4	274	232	W.1.5.3.A	Focus and Content
621008	3	Grade 3 to Grade 4	274	232	W.1.5.3.F.d	Gram. and Sent.
623107	3	Grade 3 to Grade 4	276	232	W.1.5.3.B	Focus and Content
625516	3	Grade 3 to Grade 4	275	232	W.1.5.3.F.c	Spell., Cap., Punct.
623113	4	Grade 3 to Grade 4	274	233	W.1.5.4.C	Org and Style
637175	4	Grade 3 to Grade 4	274	232	W.1.5.4.D	Org and Style
633445	4	Grade 3 to Grade 4	274	235	W.1.5.4.F.a	Spell., Cap., Punct.
635414	4	Grade 3 to Grade 4	274	233	W.1.5.4.A	Focus and Content
639852	4	Grade 3 to Grade 4	274	234	W.1.5.4.F.c	Spell., Cap., Punct.
623033	4	Grade 3 to Grade 4	274	232	W.1.5.4.F.b	Spell., Cap., Punct.
623013	4	Grade 3 to Grade 4	274	233	W.1.5.4.B	Focus and Content
633852	4	Grade 3 to Grade 4	274	233	W.1.5.4.C	Org and Style
624765	4	Grade 3 to Grade 4	274	233	W.1.5.4.F.d	Gram. and Sent.
625527	4	Grade 3 to Grade 4	274	232	W.1.5.4.E	Editing
627004	4	Grade 3 to Grade 4	275	232	W.1.5.4.E	Editing
637177	4	Grade 3 to Grade 4	275	235	W.1.5.4.D	Org and Style
633432	4	Grade 3 to Grade 4	275	233	W.1.5.4.F.a	Spell., Cap., Punct.
633464	4	Grade 3 to Grade 4	275	234	W.1.5.4.A	Focus and Content
639854	4	Grade 3 to Grade 4	275	232	W.1.5.4.F.c	Spell., Cap., Punct.
623136	4	Grade 3 to Grade 4	275	233	W.1.5.4.F.b	Spell., Cap., Punct.
635900	4	Grade 3 to Grade 4	275	233	W.1.5.4.B	Focus and Content
635412	4	Grade 3 to Grade 4	275	233	W.1.5.4.C	Org and Style
630419	4	Grade 3 to Grade 4	275	232	W.1.5.4.F.d	Gram. and Sent.
630295	4	Grade 3 to Grade 4	275	235	W.1.5.4.E	Editing

Table C–47. Writing/English Composition Items Used to Link Grade 4 to Grade 5

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
623017	4	Grade 4 to Grade 5	235	221	W.1.5.4.E	Editing
625455	4	Grade 4 to Grade 5	233	221	W.1.5.4.A	Focus and Content
622453	4	Grade 4 to Grade 5	234	221	W.1.5.4.E	Editing
623135	4	Grade 4 to Grade 5	232	221	W.1.5.4.F.b	Spell., Cap., Punct.
632573	4	Grade 4 to Grade 5	233	221	W.1.5.4.F.d	Gram. and Sent.
623020	4	Grade 4 to Grade 5	233	221	W.1.5.4.C	Org and Style
633435	4	Grade 4 to Grade 5	233	221	W.1.5.4.F.a	Spell., Cap., Punct.
623108	4	Grade 4 to Grade 5	232	221	W.1.5.4.B	Focus and Content
633468	4	Grade 4 to Grade 5	235	221	W.1.5.4.C	Org and Style
627696	4	Grade 4 to Grade 5	233	221	W.1.5.4.F.c	Spell., Cap., Punct.
623115	4	Grade 4 to Grade 5	233	221	W.1.5.4.E	Editing
622983	4	Grade 4 to Grade 5	234	221	W.1.5.4.A	Focus and Content
622454	4	Grade 4 to Grade 5	232	221	W.1.5.4.E	Editing
621395	4	Grade 4 to Grade 5	233	221	W.1.5.4.F.b	Spell., Cap., Punct.
632587	4	Grade 4 to Grade 5	233	221	W.1.5.4.F.d	Gram. and Sent.
623019	4	Grade 4 to Grade 5	233	221	W.1.5.4.C	Org and Style
634025	4	Grade 4 to Grade 5	232	221	W.1.5.4.F.a	Spell., Cap., Punct.
626922	4	Grade 4 to Grade 5	235	221	W.1.5.4.B	Focus and Content
633469	4	Grade 4 to Grade 5	233	221	W.1.5.4.C	Org and Style
628471	4	Grade 4 to Grade 5	234	221	W.1.5.4.F.c	Spell., Cap., Punct.
637149	5	Grade 4 to Grade 5	233	218	W.1.5.5.F.d	Gram. and Sent.
633440	5	Grade 4 to Grade 5	233	221	W.1.5.5.F.a	Spell., Cap., Punct.
635884	5	Grade 4 to Grade 5	233	221	W.1.5.5.E	Editing
637062	5	Grade 4 to Grade 5	233	218	W.1.5.5.F.d	Gram. and Sent.
623027	5	Grade 4 to Grade 5	233	220	W.1.5.5.F.d	Gram. and Sent.
622469	5	Grade 4 to Grade 5	233	221	W.1.5.5.F.b	Spell., Cap., Punct.
639843	5	Grade 4 to Grade 5	233	222	W.1.5.5.F.c	Spell., Cap., Punct.
635417	5	Grade 4 to Grade 5	233	221	W.1.5.5.C	Org and Style
620819	5	Grade 4 to Grade 5	233	220	W.1.5.5.C	Org and Style
635605	5	Grade 4 to Grade 5	233	221	W.1.5.5.C	Org and Style
637148	5	Grade 4 to Grade 5	232	221	W.1.5.5.C	Org and Style
633439	5	Grade 4 to Grade 5	232	221	W.1.5.5.F.a	Spell., Cap., Punct.
620820	5	Grade 4 to Grade 5	232	218	W.1.5.5.E	Editing
626566	5	Grade 4 to Grade 5	232	220	W.1.5.5.F.d	Gram. and Sent.
623129	5	Grade 4 to Grade 5	232	221	W.1.5.5.F.d	Gram. and Sent.
629858	5	Grade 4 to Grade 5	232	222	W.1.5.5.F.b	Spell., Cap., Punct.
639864	5	Grade 4 to Grade 5	232	221	W.1.5.5.F.c	Spell., Cap., Punct.
627291	5	Grade 4 to Grade 5	232	220	W.1.5.5.C	Org and Style
639349	5	Grade 4 to Grade 5	232	218	W.1.5.5.C	Org and Style
626818	5	Grade 4 to Grade 5	232	221	W.1.5.5.C	Org and Style

Table C–48. Writing/English Composition Items Used to Link Grade 5 to Grade 6

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
623105	5	Grade 5 to Grade 6	221	303	W.1.5.5.A	Focus and Content
626927	5	Grade 5 to Grade 6	218	303	W.1.5.5.F.d	Gram. and Sent.
632608	5	Grade 5 to Grade 6	220	303	W.1.5.5.E	Editing
625460	5	Grade 5 to Grade 6	221	303	W.1.5.5.C	Org and Style
626923	5	Grade 5 to Grade 6	222	303	W.1.5.5.E	Editing
628065	5	Grade 5 to Grade 6	221	303	W.1.5.5.F.b	Spell., Cap., Punct.
633443	5	Grade 5 to Grade 6	220	303	W.1.5.5.F.a	Spell., Cap., Punct.
621390	5	Grade 5 to Grade 6	218	303	W.1.5.5.F.c	Spell., Cap., Punct.
626820	5	Grade 5 to Grade 6	221	303	W.1.5.5.E	Editing
624842	5	Grade 5 to Grade 6	218	303	W.1.5.5.F.d	Gram. and Sent.
624800	5	Grade 5 to Grade 6	218	304	W.1.5.5.A	Focus and Content
627413	5	Grade 5 to Grade 6	220	304	W.1.5.5.F.d	Gram. and Sent.
630403	5	Grade 5 to Grade 6	221	304	W.1.5.5.E	Editing
624804	5	Grade 5 to Grade 6	222	304	W.1.5.5.C	Org and Style
626570	5	Grade 5 to Grade 6	221	304	W.1.5.5.E	Editing
624773	5	Grade 5 to Grade 6	220	304	W.1.5.5.F.b	Spell., Cap., Punct.
633442	5	Grade 5 to Grade 6	218	304	W.1.5.5.F.a	Spell., Cap., Punct.
629854	5	Grade 5 to Grade 6	221	304	W.1.5.5.F.c	Spell., Cap., Punct.
623060	5	Grade 5 to Grade 6	221	304	W.1.5.5.E	Editing
627488	5	Grade 5 to Grade 6	220	304	W.1.5.5.F.d	Gram. and Sent.
624292	6	Grade 5 to Grade 6	221	304	W.1.5.6.E	Editing
626934	6	Grade 5 to Grade 6	221	303	W.1.5.6.A	Focus and Content
627013	6	Grade 5 to Grade 6	221	304	W.1.5.6.F.b	Spell., Cap., Punct.
632646	6	Grade 5 to Grade 6	221	305	W.1.5.6.F.d	Gram. and Sent.
624829	6	Grade 5 to Grade 6	221	304	W.1.5.6.F.d	Gram. and Sent.
630378	6	Grade 5 to Grade 6	221	304	W.1.5.6.B	Focus and Content
624297	6	Grade 5 to Grade 6	221	303	W.1.5.6.C	Org and Style
635654	6	Grade 5 to Grade 6	221	304	W.1.5.6.F.c	Spell., Cap., Punct.
639363	6	Grade 5 to Grade 6	221	305	W.1.5.6.C	Org and Style
633448	6	Grade 5 to Grade 6	221	304	W.1.5.6.F.a	Spell., Cap., Punct.
623114	6	Grade 5 to Grade 6	222	303	W.1.5.6.E	Editing
626932	6	Grade 5 to Grade 6	222	304	W.1.5.6.A	Focus and Content
635660	6	Grade 5 to Grade 6	222	305	W.1.5.6.F.b	Spell., Cap., Punct.
626822	6	Grade 5 to Grade 6	222	304	W.1.5.6.F.d	Gram. and Sent.
625478	6	Grade 5 to Grade 6	222	304	W.1.5.6.F.d	Gram. and Sent.
626776	6	Grade 5 to Grade 6	222	303	W.1.5.6.B	Focus and Content
624296	6	Grade 5 to Grade 6	222	304	W.1.5.6.C	Org and Style
628055	6	Grade 5 to Grade 6	222	305	W.1.5.6.F.c	Spell., Cap., Punct.
627289	6	Grade 5 to Grade 6	222	304	W.1.5.6.C	Org and Style
633444	6	Grade 5 to Grade 6	222	304	W.1.5.6.F.a	Spell., Cap., Punct.

Table C–49. Writing/English Composition Items Used to Link Grade 6 to Grade 7

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
633446	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.a	Spell., Cap., Punct.
635619	6	Grade 6 to Grade 7	305	279	W.1.5.6.D	Org and Style
635662	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.b	Spell., Cap., Punct.
623111	6	Grade 6 to Grade 7	304	279	W.1.5.6.E	Editing
624754	6	Grade 6 to Grade 7	303	279	W.1.5.6.F.d	Gram. and Sent.
628060	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.c	Spell., Cap., Punct.
627415	6	Grade 6 to Grade 7	305	279	W.1.5.6.F.d	Gram. and Sent.
624287	6	Grade 6 to Grade 7	304	279	W.1.5.6.E	Editing
624763	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.d	Gram. and Sent.
627960	6	Grade 6 to Grade 7	303	279	W.1.5.6.A	Focus and Content
633447	6	Grade 6 to Grade 7	305	279	W.1.5.6.F.a	Spell., Cap., Punct.
639392	6	Grade 6 to Grade 7	304	279	W.1.5.6.D	Org and Style
635661	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.b	Spell., Cap., Punct.
624289	6	Grade 6 to Grade 7	303	279	W.1.5.6.E	Editing
624756	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.d	Gram. and Sent.
628061	6	Grade 6 to Grade 7	305	279	W.1.5.6.F.c	Spell., Cap., Punct.
628112	6	Grade 6 to Grade 7	304	279	W.1.5.6.F.d	Gram. and Sent.
626567	6	Grade 6 to Grade 7	304	279	W.1.5.6.E	Editing
624840	6	Grade 6 to Grade 7	303	279	W.1.5.6.F.d	Gram. and Sent.
627030	6	Grade 6 to Grade 7	304	279	W.1.5.6.A	Focus and Content
627052	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.d	Gram. and Sent.
639447	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.d	Gram. and Sent.
627058	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.d	Gram. and Sent.
639380	7	Grade 6 to Grade 7	303	279	W.1.5.7.A	Focus and Content
624286	7	Grade 6 to Grade 7	303	280	W.1.5.7.B	Focus and Content
624822	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.b	Spell., Cap., Punct.
636003	7	Grade 6 to Grade 7	303	280	W.1.5.7.C	Org and Style
633454	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.a	Spell., Cap., Punct.
635909	7	Grade 6 to Grade 7	303	279	W.1.5.7.D	Org and Style
634300	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.c	Spell., Cap., Punct.
626992	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.d	Gram. and Sent.
639438	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.d	Gram. and Sent.
628116	7	Grade 6 to Grade 7	303	279	W.1.5.7.F.d	Gram. and Sent.
626764	7	Grade 6 to Grade 7	303	280	W.1.5.7.A	Focus and Content
639394	7	Grade 6 to Grade 7	303	280	W.1.5.7.B	Focus and Content
628476	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.b	Spell., Cap., Punct.
636008	7	Grade 6 to Grade 7	303	280	W.1.5.7.C	Org and Style
633455	7	Grade 6 to Grade 7	303	279	W.1.5.7.F.a	Spell., Cap., Punct.
639420	7	Grade 6 to Grade 7	303	280	W.1.5.7.D	Org and Style
634299	7	Grade 6 to Grade 7	303	280	W.1.5.7.F.c	Spell., Cap., Punct.

Table C–50. Writing/English Composition Items Used to Link Grade 7 to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
627684	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.d	Gram. and Sent.
625487	7	Grade 8 to Grade 7	279	145	W.1.5.7.F.d	Gram. and Sent.
627464	7	Grade 8 to Grade 7	280	145	W.1.5.7.A	Focus and Content
639375	7	Grade 8 to Grade 7	280	145	W.1.5.7.C	Org and Style
633458	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.a	Spell., Cap., Punct.
626996	7	Grade 8 to Grade 7	280	145	W.1.5.7.E	Editing
628098	7	Grade 8 to Grade 7	279	145	W.1.5.7.F.b	Spell., Cap., Punct.
639358	7	Grade 8 to Grade 7	280	145	W.1.5.7.B	Focus and Content
635665	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.c	Spell., Cap., Punct.
627361	7	Grade 8 to Grade 7	280	145	W.1.5.7.C	Org and Style
627056	7	Grade 8 to Grade 7	279	145	W.1.5.7.F.d	Gram. and Sent.
639407	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.d	Gram. and Sent.
626943	7	Grade 8 to Grade 7	280	145	W.1.5.7.A	Focus and Content
639364	7	Grade 8 to Grade 7	280	145	W.1.5.7.C	Org and Style
633457	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.a	Spell., Cap., Punct.
626997	7	Grade 8 to Grade 7	279	145	W.1.5.7.F.d	Gram. and Sent.
630429	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.b	Spell., Cap., Punct.
625506	7	Grade 8 to Grade 7	280	145	W.1.5.7.B	Focus and Content
635668	7	Grade 8 to Grade 7	280	145	W.1.5.7.F.c	Spell., Cap., Punct.
627362	7	Grade 8 to Grade 7	280	145	W.1.5.7.C	Org and Style
633498	8	Grade 8 to Grade 7	279	144	W.1.5.8.F.a	Spell., Cap., Punct.
639580	8	Grade 8 to Grade 7	279	145	W.1.5.8.C	Org and Style
624848	8	Grade 8 to Grade 7	279	143	W.1.5.8.F.b	Spell., Cap., Punct.
639612	8	Grade 8 to Grade 7	279	144	W.1.5.8.B	Focus and Content
628115	8	Grade 8 to Grade 7	279	144	W.1.5.8.F.d	Gram. and Sent.
627963	8	Grade 8 to Grade 7	279	144	W.1.5.8.A	Focus and Content
628311	8	Grade 8 to Grade 7	279	145	W.1.5.8.F.d	Gram. and Sent.
628242	8	Grade 8 to Grade 7	279	143	W.1.5.8.B	Focus and Content
639857	8	Grade 8 to Grade 7	279	144	W.1.5.8.F.c	Spell., Cap., Punct.
639441	8	Grade 8 to Grade 7	279	144	W.1.5.8.F.d	Gram. and Sent.
633497	8	Grade 8 to Grade 7	280	145	W.1.5.8.F.a	Spell., Cap., Punct.
639588	8	Grade 8 to Grade 7	280	143	W.1.5.8.C	Org and Style
625522	8	Grade 8 to Grade 7	280	144	W.1.5.8.F.b	Spell., Cap., Punct.
639610	8	Grade 8 to Grade 7	280	144	W.1.5.8.B	Focus and Content
624828	8	Grade 8 to Grade 7	280	144	W.1.5.8.F.d	Gram. and Sent.
625520	8	Grade 8 to Grade 7	280	145	W.1.5.8.A	Focus and Content
625508	8	Grade 8 to Grade 7	280	143	W.1.5.8.F.d	Gram. and Sent.
626775	8	Grade 8 to Grade 7	280	144	W.1.5.8.B	Focus and Content
639856	8	Grade 8 to Grade 7	280	144	W.1.5.8.F.c	Spell., Cap., Punct.
639439	8	Grade 8 to Grade 7	280	144	W.1.5.8.F.d	Gram. and Sent.

Table C–51. Writing/English Composition Items Used to Link English Composition to Grade 8

Item ID	Item Grade	Link	N Count Lower Grade	N Count Upper Grade	Eligible Content	Writing/Composition Diagnostic Category
636213	8	English Comp to Grade 8	143	173	W.1.5.8.F.d	Gram. and Sent.
639599	8	English Comp to Grade 8	144	173	W.1.5.8.C	Org and Style
633503	8	English Comp to Grade 8	144	173	W.1.5.8.F.a	Spell., Cap., Punct.
629857	8	English Comp to Grade 8	144	173	W.1.5.8.F.b	Spell., Cap., Punct.
634156	8	English Comp to Grade 8	145	173	W.1.5.8.F.c	Spell., Cap., Punct.
639577	8	English Comp to Grade 8	143	173	W.1.5.8.E	Editing
635385	8	English Comp to Grade 8	144	173	W.1.5.8.F.d	Gram. and Sent.
635351	8	English Comp to Grade 8	144	173	W.1.5.8.F.d	Gram. and Sent.
627964	8	English Comp to Grade 8	144	173	W.1.5.8.A	Focus and Content
626786	8	English Comp to Grade 8	145	173	W.1.5.8.C	Org and Style
636212	8	English Comp to Grade 8	144	171	W.1.5.8.F.d	Gram. and Sent.
639597	8	English Comp to Grade 8	144	171	W.1.5.8.C	Org and Style
633502	8	English Comp to Grade 8	144	171	W.1.5.8.F.a	Spell., Cap., Punct.
629860	8	English Comp to Grade 8	145	171	W.1.5.8.F.b	Spell., Cap., Punct.
634157	8	English Comp to Grade 8	143	171	W.1.5.8.F.c	Spell., Cap., Punct.
639608	8	English Comp to Grade 8	144	171	W.1.5.8.E	Editing
635386	8	English Comp to Grade 8	144	171	W.1.5.8.F.d	Gram. and Sent.
635350	8	English Comp to Grade 8	144	171	W.1.5.8.F.d	Gram. and Sent.
628143	8	English Comp to Grade 8	145	171	W.1.5.8.A	Focus and Content
626785	8	English Comp to Grade 8	143	171	W.1.5.8.C	Org and Style
622816	EC	English Comp to Grade 8	143	173	C.E.1.1.1	Focus and Content
639932	EC	English Comp to Grade 8	143	173	C.E.3.1.5	Gram. and Sent.
639920	EC	English Comp to Grade 8	143	171	C.E.3.1.4	Gram. and Sent.
634313	EC	English Comp to Grade 8	143	173	C.E.3.1.2	Spell., Cap., Punct.
633540	EC	English Comp to Grade 8	143	172	C.E.3.1.1	Spell., Cap., Punct.
622613	EC	English Comp to Grade 8	143	173	C.E.1.1.3	Org and Style
623126	EC	English Comp to Grade 8	143	173	C.E.3.1.4	Gram. and Sent.
639971	EC	English Comp to Grade 8	143	174	C.E.1.1.2	Focus and Content
629853	EC	English Comp to Grade 8	143	174	C.E.3.1.3	Spell., Cap., Punct.
630391	EC	English Comp to Grade 8	143	173	C.E.1.1.3	Org and Style
622815	EC	English Comp to Grade 8	145	174	C.P.1.1.1	Focus and Content
639933	EC	English Comp to Grade 8	145	173	C.E.3.1.5	Gram. and Sent.
639919	EC	English Comp to Grade 8	145	173	C.E.3.1.4	Gram. and Sent.
634349	EC	English Comp to Grade 8	145	174	C.E.3.1.2	Spell., Cap., Punct.
633536	EC	English Comp to Grade 8	145	174	C.E.3.1.1	Spell., Cap., Punct.
622611	EC	English Comp to Grade 8	145	174	C.E.1.1.3	Org and Style
621166	EC	English Comp to Grade 8	145	173	C.E.3.1.4	Gram. and Sent.
630659	EC	English Comp to Grade 8	145	173	C.E.1.1.2	Focus and Content
629822	EC	English Comp to Grade 8	145	173	C.E.3.1.3	Spell., Cap., Punct.
630392	EC	English Comp to Grade 8	145	171	C.E.1.1.3	Org and Style

Tables C–52 through C–57 summarize the number of linking items by diagnostic category.

Table C–52. Number of Items Linking Grade 3 to Grade 4 by Diagnostic Category

Diagnostic Category	Grade 3 Items	Grade 4 Items	Total
Focus and Content	4	4	8
Org and Style	4	5	9
Editing	0	3	3
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	6	2	8
TOTAL	20	20	40

Table C–53. Number of Items Linking Grade 4 to Grade 5 by Diagnostic Category

Diagnostic Category	Grade 4 Items	Grade 5 Items	Total
Focus and Content	4	0	4
Org and Style	4	7	11
Editing	4	2	6
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	2	5	7
TOTAL	20	20	40

Table C–54. Number of Items Linking Grade 5 to Grade 6 by Diagnostic Category

Diagnostic Category	Grade 5 Items	Grade 6 Items	Total
Focus and Content	2	4	6
Org and Style	2	4	6
Editing	6	2	8
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	4	4	8
TOTAL	20	20	40

Table C–55. Number of Items Linking Grade 6 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 6 Items	Grade 7 Items	Total
Focus and Content	2	4	6
Org and Style	2	4	6
Editing	4	0	4
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	6	6	12
TOTAL	20	20	40

Table C–56. Number of Items Linking Grade 8 to Grade 7 by Diagnostic Category

Diagnostic Category	Grade 7 Items	Grade 8 Items	Total
Focus and Content	4	6	10
Org and Style	4	2	6
Editing	1	0	1
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	5	6	11
TOTAL	20	20	40

Table C–57. Number of Items Linking English Composition to Grade 8 by Diagnostic Category

Diagnostic Category	Grade 8 Items	Eng Comp Items	Total
Focus and Content	2	4	6
Org and Style	4	4	8
Editing	2	0	2
Spell., Cap., Punct.	6	6	12
Gram. and Sent.	6	6	12
TOTAL	20	20	40

Table C–58. Writing/English Composition Example of Vertical Linking Workbook

		Grade 4 Calibration			Grade 5 Calibration				Grade 4 on		
Item ID	Item Grade	Difficulty	Fit	Displace	Difficulty	Fit	Displace	Discrepancy	Grade 5 Scale	Robust Z	Flag
623017	4	-0.784	0.910	-0.006	-0.927	0.910	0.000	-0.143	-1.005	0.233	
625455	4	-0.205	1.030	-0.001	0.132	1.010	0.001	0.337	-0.426	1.437	
622453	4	-0.955	0.910	0.003	-1.526	0.860	0.000	-0.571	-1.176	-0.840	
623135	4	1.520	1.200	0.005	1.516	1.110	0.001	-0.004	1.299	0.582	
632573	4	0.527	1.250	-0.002	0.872	1.190	0.001	0.345	0.306	1.457	
623020	4	-1.254	0.890	-0.001	-1.487	0.900	0.000	-0.233	-1.475	0.008	
633435	4	-0.452	1.020	-0.003	-0.441	0.910	0.000	0.011	-0.673	0.620	
623108	4	-0.152	0.830	0.000	0.025	0.920	0.000	0.177	-0.373	1.036	
633468	4	-0.857	0.900	-0.006	-0.475	0.860	0.000	0.382	-1.078	1.550	
627696	4	1.837	1.210	-0.001	1.968	1.140	0.001	0.131	1.616	0.921	
623115	4	-0.678	0.960	-0.001	-1.072	0.890	-0.003	-0.394	-0.899	-0.396	
622983	4	-0.797	1.020	0.003	-1.360	0.980	-0.003	-0.563	-1.018	-0.820	
622454	4	0.922	1.070	0.005	0.483	1.000	-0.002	-0.439	0.701	-0.509	
621395	4	1.634	1.080	-0.002	0.998	1.090	-0.002	-0.636	1.413	-1.003	
632587	4	0.650	0.830	-0.001	0.149	0.980	-0.002	-0.501	0.429	-0.665	
623019	4	-1.134	0.990	-0.003	-1.611	1.020	-0.003	-0.477	-1.355	-0.605	
634025	4	-0.885	0.960	0.000	-1.496	0.920	-0.003	-0.611	-1.106	-0.941	
626922	4	0.516	1.000	-0.006	0.159	0.970	-0.002	-0.357	0.295	-0.304	
633469	4	-0.151	0.880	-0.001	-0.121	0.900	-0.002	0.030	-0.372	0.667	
628471	4	2.662	1.140	0.003	2.119	1.130	-0.001	-0.543	2.441	-0.770	
637149	5	-2.406	0.960	0.003	-2.126	0.960	0.005	0.280	-2.627	1.294	
633440	5	-0.302	1.040	0.003	-0.227	0.960	0.001	0.075	-0.523	0.780	
635884	5	-1.607	0.840	0.003	-1.708	0.870	-0.001	-0.101	-1.828	0.339	
637062	5	0.739	1.110	0.004	0.794	1.170	0.000	0.055	0.518	0.730	
623027	5	-0.341	0.780	0.003	-0.917	0.800	-0.004	-0.576	-0.562	-0.853	
622469	5	1.057	1.110	0.004	0.730	1.000	0.000	-0.327	0.836	-0.228	
639843	5	-0.548	0.910	0.003	-1.127	0.990	-0.002	-0.579	-0.769	-0.860	
635417	5	0.499	1.050	0.004	0.561	1.050	-0.005	0.062	0.278	0.747	
620819	5	0.739	0.970	0.004	0.337	0.950	-0.005	-0.402	0.518	-0.416	
635605	5	1.417	1.220	0.004	1.437	1.080	0.001	0.020	1.196	0.642	
637148	5	-0.606	0.950	0.002	-1.440	0.920	0.001	-0.834	-0.827	-1.500	
633439	5	0.404	1.100	0.002	0.544	1.050	-0.001	0.140	0.183	0.943	
620820	5	0.287	0.950	0.002	0.089	0.960	0.000	-0.198	0.066	0.095	
626566	5	-0.764	0.860	0.002	-1.003	0.860	-0.004	-0.239	-0.985	-0.008	
623129	5	-1.331	0.800	0.002	-1.323	0.820	0.000	0.008	-1.552	0.612	
629858	5	1.124	1.020	0.003	0.983	1.020	-0.002	-0.141	0.903	0.238	
639864	5	-0.729	0.950	0.002	-1.075	0.900	-0.005	-0.346	-0.950	-0.276	
627291	5	0.515	0.880	0.002	0.008	0.970	-0.005	-0.507	0.294	-0.680	
639349	5	0.658	1.040	0.002	0.285	0.890	0.005	-0.373	0.437	-0.344	
626818	5	1.722	0.970	0.003	0.913	0.990	-0.001	-0.809	1.501	-1.437	
	Mean	0.062			-0.159			-0.221	-0.159	0.037	
	SD	1.088			1.095			0.330	1.088	0.828	
	SD Ratio	0.993									
	Correlation	0.954									
	Add. Constant	-0.221									
	Median							-0.236			
	Q							0.539			

Figures C–22 through C–27 are the adjacent grade linking plots. No items were removed from final linking procedure so there are no red items in these plots.

Figure C–22. CDT Writing/English Composition: Grade 3 to Grade 4 Linking – All Links

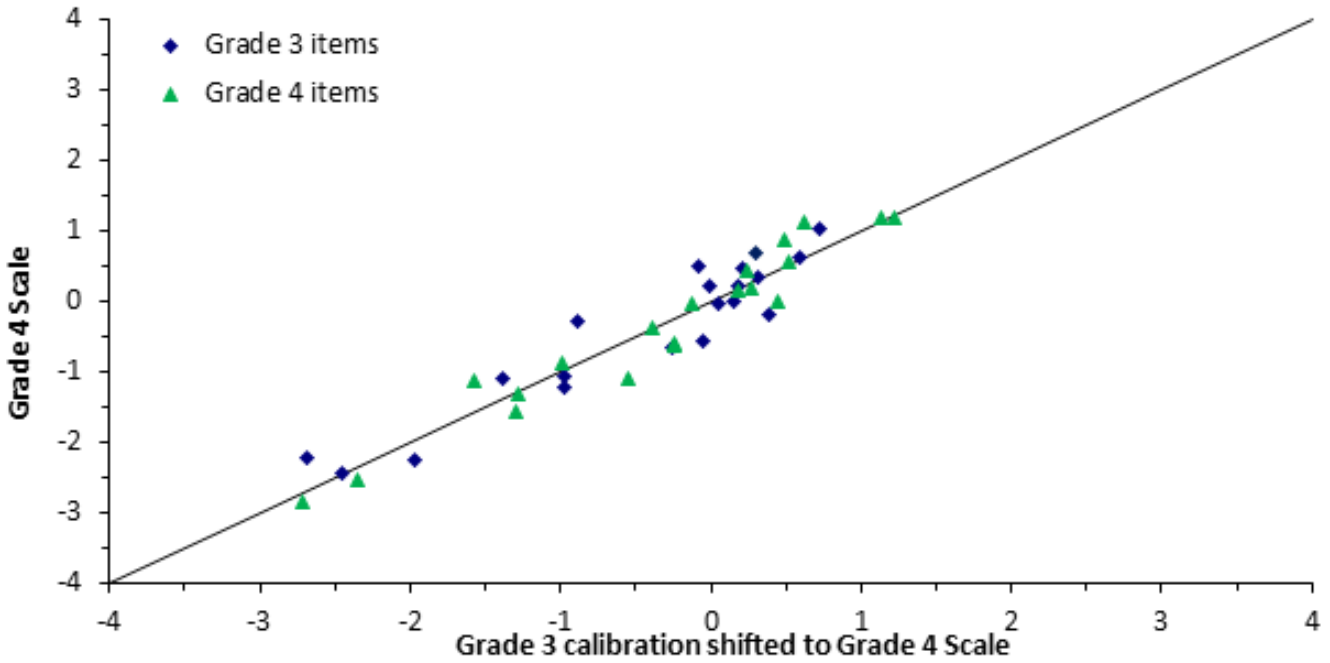


Figure C–23. CDT Writing/English Composition: Grade 4 to Grade 5 Linking – All Links

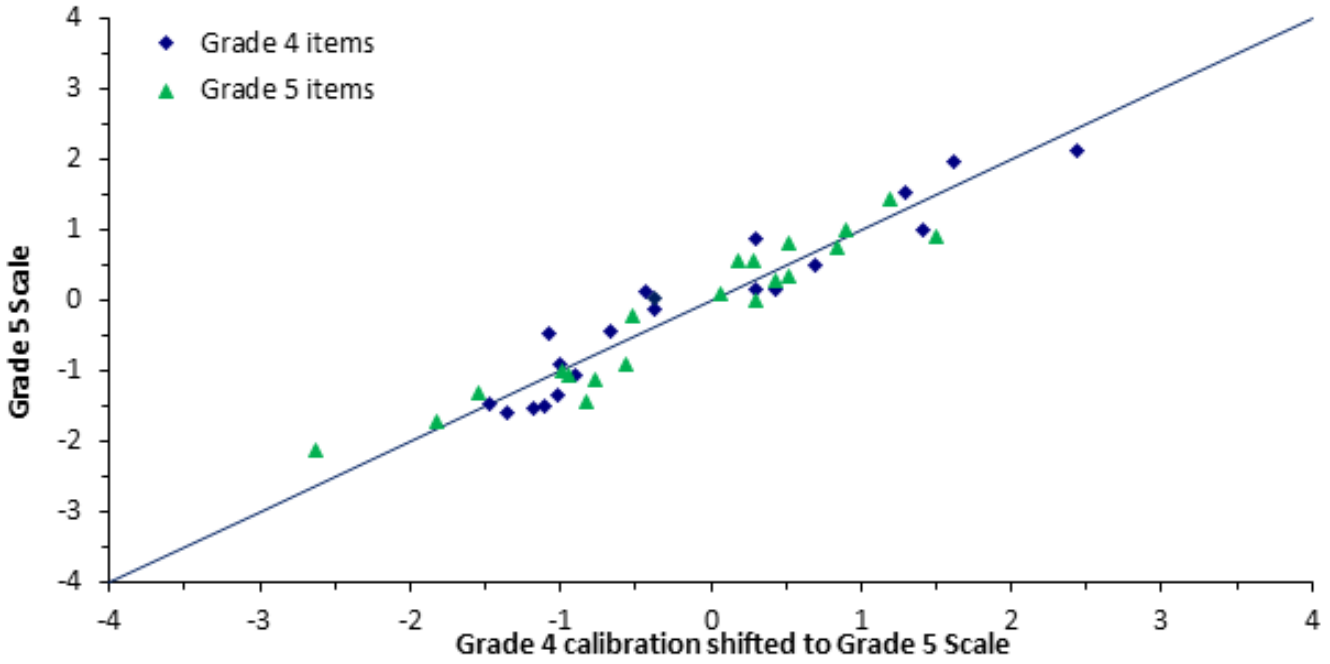


Figure C–24. CDT Writing/English Composition: Grade 5 to Grade 6 Linking – All Links

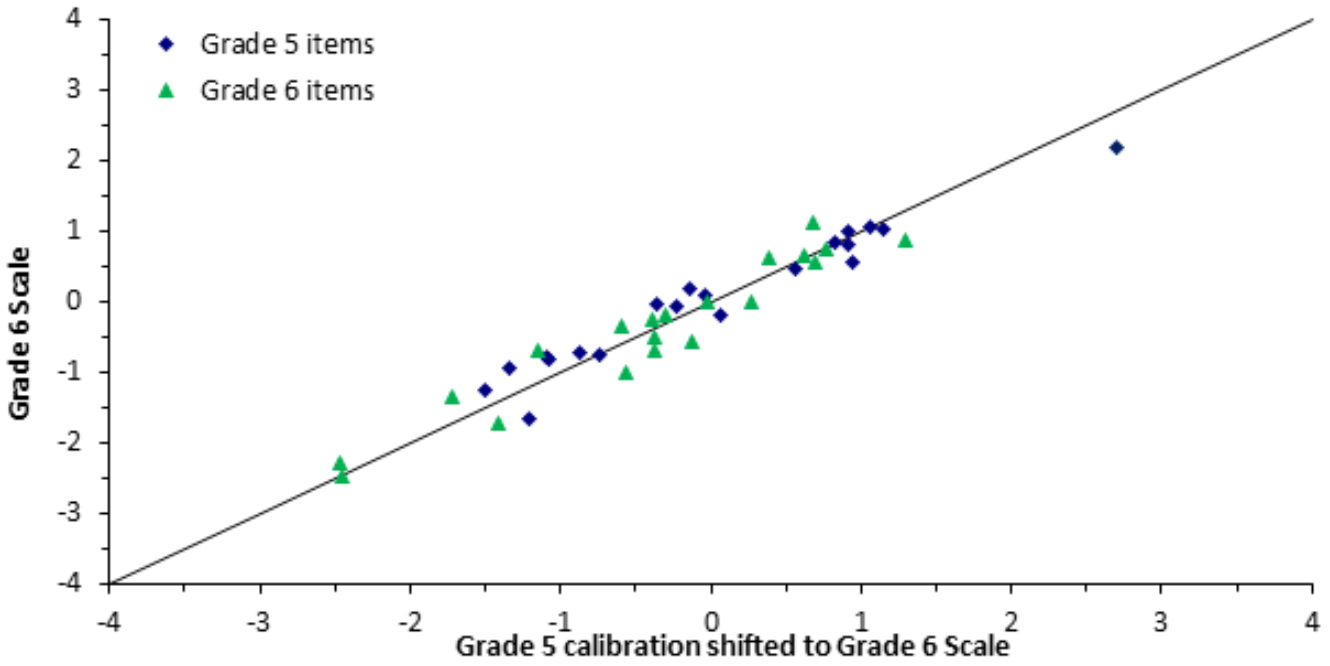


Figure C–25. CDT Writing/English Composition: Grade 6 to Grade 7 Linking – All Links

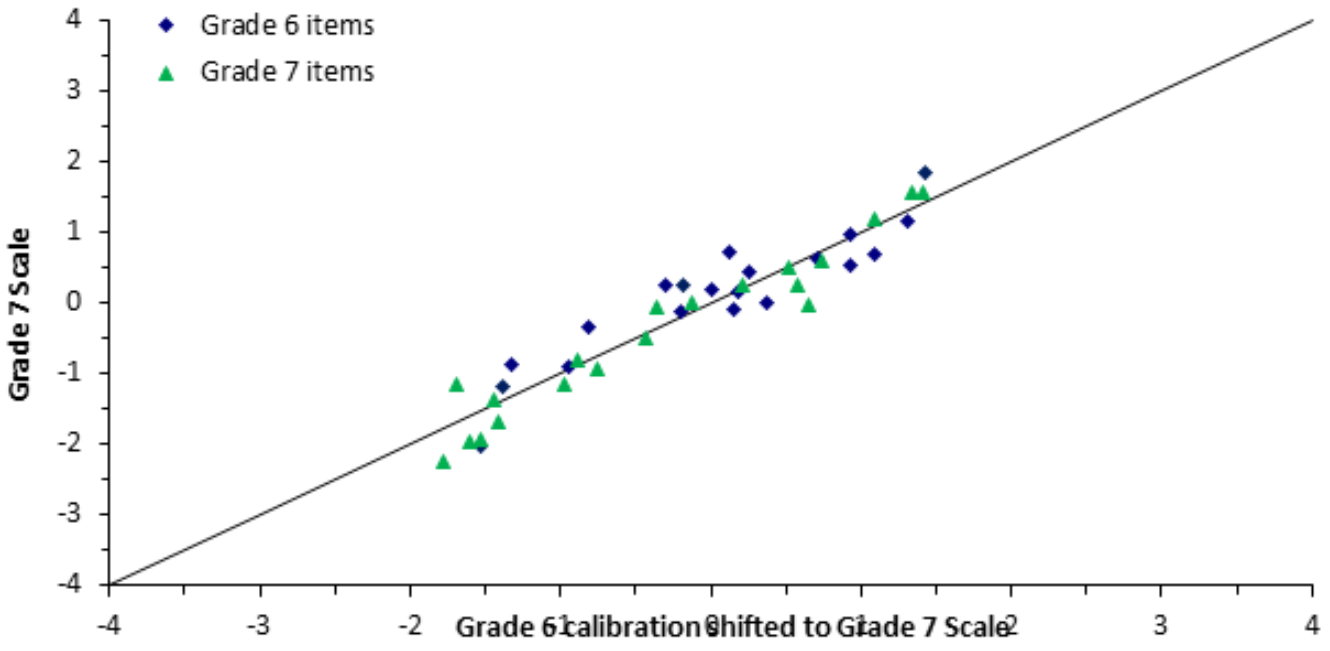


Figure C–26. CDT Writing/English Composition: Grade 8 to Grade 7 Linking – All Links

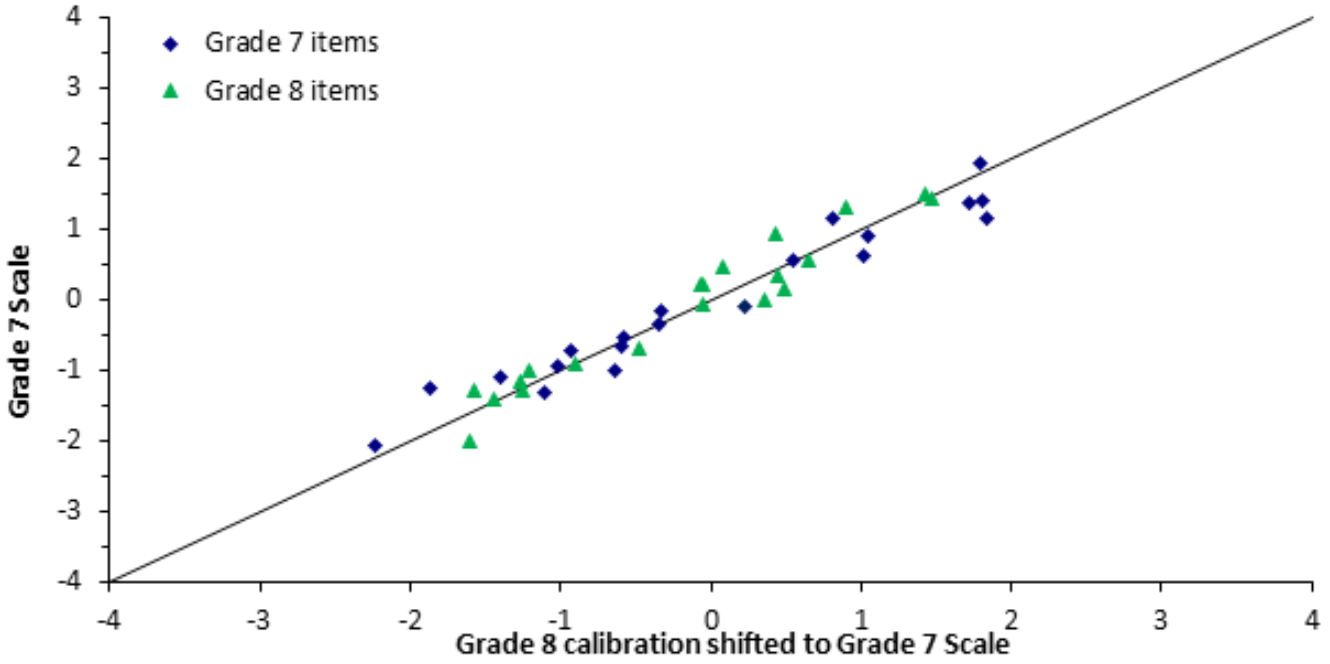
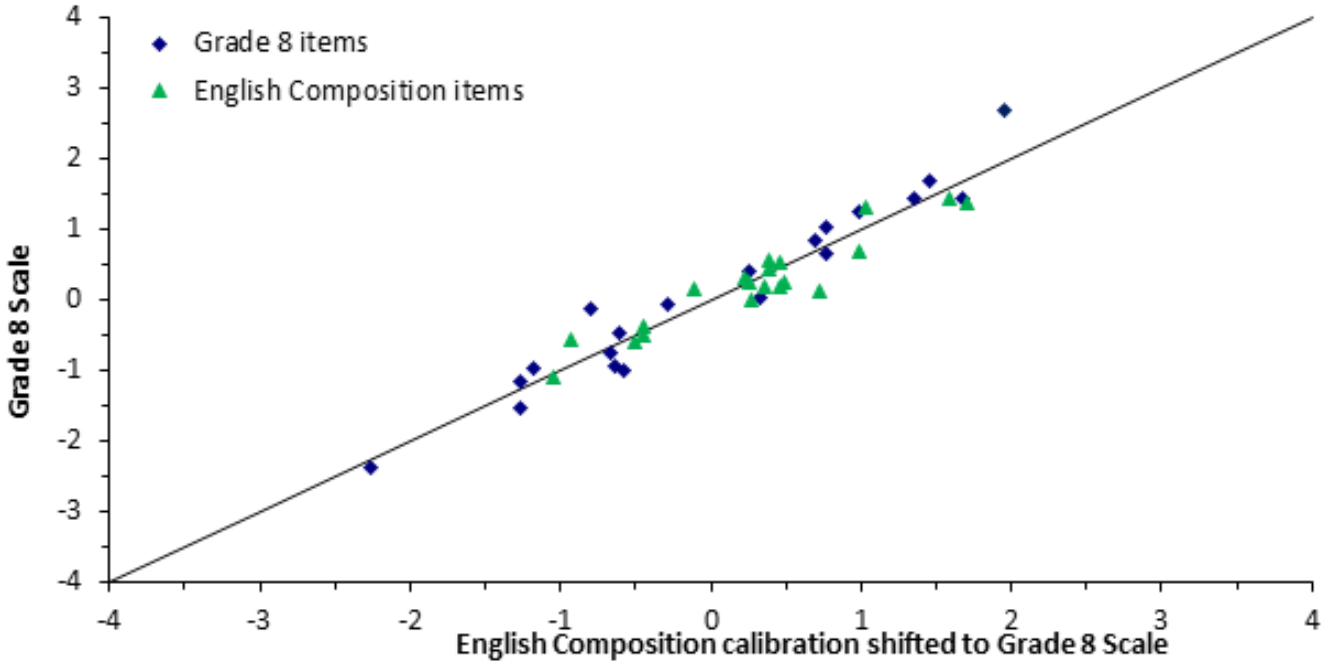


Figure C–27. CDT Writing/English Composition: Literature to Grade 8 Linking – All Links



APPENDIX D: SIGNIFICANT DIFFERENCES AMONG DIAGNOSTIC CATEGORIES

In Chapter Fifteen (Operational Administration 2020–2021), significant differences among diagnostic categories were tested with a t-test using a Bonferroni correction for multiple comparisons to keep the familywise Type I error rate at 0.32. The tables in this appendix show the significant differences with the familywise Type I error rate at 0.10.

DIAGNOSTIC CATEGORY SIGNIFICANT DIFFERENCES

Table D–1a. Diagnostic Category Significant Differences – Math Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	405	75,614	0.5%	99.5%
DC1	DC3	849	75,170	1.1%	98.9%
DC1	DC4	423	75,596	0.6%	99.4%
DC2	DC3	877	75,142	1.2%	98.8%
DC2	DC4	420	75,599	0.6%	99.4%
DC3	DC4	728	75,291	1.0%	99.0%

Note: Z value is 2.39

Table D–1b. Total Number of Diagnostic Category Significant Differences – Math Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	73,197	96.3%
1	2,107	2.8%
2	553	0.7%
3	159	0.2%
4	3	0.0%
5	0	0.0%
6	0	0.0%

Table D–2a. Diagnostic Category Significant Differences – Math Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,089	96,468	1.1%	98.9%
DC1	DC3	1,390	96,167	1.4%	98.6%
DC1	DC4	1,171	96,386	1.2%	98.8%
DC2	DC3	1,111	96,446	1.1%	98.9%
DC2	DC4	1,122	96,435	1.2%	98.8%
DC3	DC4	1,122	96,435	1.2%	98.8%

Note: Z value is 2.39

Table D–2b. Total Number of Diagnostic Category Significant Differences – Math Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	92,514	94.8%
1	3,494	3.6%
2	1,160	1.2%
3	365	0.4%
4	24	0.0%
5	0	0.0%
6	0	0.0%

Table D–3a. Diagnostic Category Significant Differences – Algebra I

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	1,191	41,534	2.8%	97.2%
DC1	DC3	1,154	41,571	2.7%	97.3%
DC1	DC4	1,358	41,367	3.2%	96.8%
DC2	DC3	257	42,468	0.6%	99.4%
DC2	DC4	516	42,209	1.2%	98.8%
DC3	DC4	529	42,196	1.2%	98.8%

Note: Z value is 2.39

Table D–3b. Total Number of Diagnostic Category Significant Differences – Algebra I

Number of Significant Differences	Number of Students	Percent of Students
0	39,147	91.6%
1	2,413	5.6%
2	916	2.1%
3	236	0.6%
4	13	0.0%
5	0	0.0%
6	0	0.0%

Table D–4a. Diagnostic Category Significant Differences – Geometry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	60	5,071	1.2%	98.8%
DC1	DC3	81	5,050	1.6%	98.4%
DC1	DC4	80	5,051	1.6%	98.4%
DC2	DC3	82	5,049	1.6%	98.4%
DC2	DC4	97	5,034	1.9%	98.1%
DC3	DC4	96	5,035	1.9%	98.1%

Note: Z value is 2.39

Table D–4b. Total Number of Diagnostic Category Significant Differences – Geometry

Number of Significant Differences	Number of Students	Percent of Students
0	4,812	93.8%
1	192	3.7%
2	88	1.7%
3	28	0.5%
4	11	0.2%
5	0	0.0%
6	0	0.0%

Table D–5a. Diagnostic Category Significant Differences – Algebra II

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	233	5,790	3.9%	96.1%
DC1	DC3	300	5,723	5.0%	95.0%
DC1	DC4	555	5,468	9.2%	90.8%
DC2	DC3	94	5,929	1.6%	98.4%
DC2	DC4	175	5,848	2.9%	97.1%
DC3	DC4	67	5,956	1.1%	98.9%

Note: Z value is 2.39

Table D–5b. Total Number of Diagnostic Category Significant Differences – Algebra II

Number of Significant Differences	Number of Students	Percent of Students
0	5,124	85.1%
1	514	8.5%
2	259	4.3%
3	112	1.9%
4	14	0.2%
5	0	0.0%
6	0	0.0%

Table D–6a. Diagnostic Category Significant Differences – Reading Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	9	64,844	0.0%	100.0%
DC1	DC3	15	64,838	0.0%	100.0%
DC1	DC4	19	64,834	0.0%	100.0%
DC1	DC5	11	64,842	0.0%	100.0%
DC2	DC3	18	64,835	0.0%	100.0%
DC2	DC4	15	64,838	0.0%	100.0%
DC2	DC5	7	64,846	0.0%	100.0%
DC3	DC4	17	64,836	0.0%	100.0%
DC3	DC5	34	64,819	0.1%	99.9%
DC4	DC5	15	64,838	0.0%	100.0%

Note: Z value is 2.58

Table D–6b. Total Number of Diagnostic Category Significant Differences – Reading Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	64,721	99.8%
1	110	0.2%
2	17	0.0%
3	4	0.0%
4	1	0.0%
5	0	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table D–7a. Diagnostic Category Significant Differences – Reading/Lit Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	21	139,034	0.0%	100.0%
DC1	DC3	17	139,038	0.0%	100.0%
DC1	DC4	13	139,042	0.0%	100.0%
DC1	DC5	54	139,001	0.0%	100.0%
DC2	DC3	25	139,030	0.0%	100.0%
DC2	DC4	16	139,039	0.0%	100.0%
DC2	DC5	76	138,979	0.1%	99.9%
DC3	DC4	23	139,032	0.0%	100.0%
DC3	DC5	94	138,961	0.1%	99.9%
DC4	DC5	66	138,989	0.0%	100.0%

Note: Z value is 2.58

Table D–7b. Total Number of Diagnostic Category Significant Differences – Reading/Lit Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	138,695	99.7%
1	320	0.2%
2	35	0.0%
3	5	0.0%
4	0	0.0%
5	0	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table D–8a. Diagnostic Category Significant Differences – Science Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	70	20,608	0.3%	99.7%
DC1	DC3	84	20,594	0.4%	99.6%
DC1	DC4	97	20,581	0.5%	99.5%
DC2	DC3	87	20,591	0.4%	99.6%
DC2	DC4	76	20,602	0.4%	99.6%
DC3	DC4	89	20,589	0.4%	99.6%

Note: Z value is 2.39

Table D–8b. Total Number of Diagnostic Category Significant Differences – Science Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	20,286	98.1%
1	301	1.5%
2	71	0.3%
3	20	0.1%
4	0	0.0%
5	0	0.0%
6	0	0.0%

Table D–9a. Diagnostic Category Significant Differences – Science Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	388	51,601	0.7%	99.3%
DC1	DC3	398	51,591	0.8%	99.2%
DC1	DC4	331	51,658	0.6%	99.4%
DC2	DC3	388	51,601	0.7%	99.3%
DC2	DC4	368	51,621	0.7%	99.3%
DC3	DC4	341	51,648	0.7%	99.3%

Note: Z value is 2.39

Table D–9b. Total Number of Diagnostic Category Significant Differences – Science Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	50,396	96.9%
1	1,099	2.1%
2	373	0.7%
3	115	0.2%
4	6	0.0%
5	0	0.0%
6	0	0.0%

Table D–10a. Diagnostic Category Significant Differences – Biology

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	275	46,202	0.6%	99.4%
DC1	DC3	313	46,164	0.7%	99.3%
DC1	DC4	397	46,080	0.9%	99.1%
DC2	DC3	163	46,314	0.4%	99.6%
DC2	DC4	391	46,086	0.8%	99.2%
DC3	DC4	337	46,140	0.7%	99.3%

Note: Z value is 2.39

Table D–10b. Total Number of Diagnostic Category Significant Differences – Biology

Number of Significant Differences	Number of Students	Percent of Students
0	44,997	96.8%
1	1,133	2.4%
2	300	0.6%
3	45	0.1%
4	2	0.0%
5	0	0.0%
6	0	0.0%

Table D–11a. Diagnostic Category Significant Differences – Chemistry

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	86	2,984	2.8%	97.2%
DC1	DC3	47	3,023	1.5%	98.5%
DC1	DC4	52	3,018	1.7%	98.3%
DC2	DC3	5	3,065	0.2%	99.8%
DC2	DC4	7	3,063	0.2%	99.8%
DC3	DC4	3	3,067	0.1%	99.9%

Note: Z value is 2.39

Table D–11b. Total Number of Diagnostic Category Significant Differences – Chemistry

Number of Significant Differences	Number of Students	Percent of Students
0	2,923	95.2%
1	103	3.4%
2	35	1.1%
3	9	0.3%
4	0	0.0%
5	0	0.0%
6	0	0.0%

Table D–12a. Diagnostic Category Significant Differences – Writing Grades 3–5

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	19	8,459	0.2%	99.8%
DC1	DC3	14	8,464	0.2%	99.8%
DC1	DC4	16	8,462	0.2%	99.8%
DC1	DC5	22	8,456	0.3%	99.7%
DC2	DC3	10	8,468	0.1%	99.9%
DC2	DC4	18	8,460	0.2%	99.8%
DC2	DC5	18	8,460	0.2%	99.8%
DC3	DC4	12	8,466	0.1%	99.9%
DC3	DC5	14	8,464	0.2%	99.8%
DC4	DC5	19	8,459	0.2%	99.8%

Note: Z value is 2.58

Table D–12b. Total Number of Diagnostic Category Significant Differences – Writing Grades 3–5

Number of Significant Differences	Number of Students	Percent of Students
0	8,367	98.7%
1	76	0.9%
2	21	0.2%
3	12	0.1%
4	2	0.0%
5	0	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

Table D–13a. Diagnostic Category Significant Differences – Writing/Eng Comp Grades 6–HS

Group 1	Group 2	Yes	No	% Yes	% No
DC1	DC2	33	21,515	0.2%	99.8%
DC1	DC3	49	21,499	0.2%	99.8%
DC1	DC4	74	21,474	0.3%	99.7%
DC1	DC5	79	21,469	0.4%	99.6%
DC2	DC3	38	21,510	0.2%	99.8%
DC2	DC4	66	21,482	0.3%	99.7%
DC2	DC5	50	21,498	0.2%	99.8%
DC3	DC4	71	21,477	0.3%	99.7%
DC3	DC5	58	21,490	0.3%	99.7%
DC4	DC5	69	21,479	0.3%	99.7%

Note: Z value is 2.58

Table D–13b. Total Number of Diagnostic Category Significant Differences – Writing/Eng Comp Grades 6–HS

Number of Significant Differences	Number of Students	Percent of Students
0	21,134	98.1%
1	288	1.3%
2	88	0.4%
3	30	0.1%
4	7	0.0%
5	1	0.0%
6	0	0.0%
7	0	0.0%
8	0	0.0%
9	0	0.0%
10	0	0.0%

APPENDIX E: DECISION CONSISTENCY

In Chapter Sixteen (Reliability), decision consistency for each CDT test and benchmark cut is reported with two values: exact agreement rate and kappa. However, as noted in the chapter, retest classification probability varies at different points along the scale. For example, the retest probability of green is higher for scores near the red/green cut than for scores very low in the red range. This appendix provides a more detailed examination of the differences in retest probability across the score range. 3 X 3 retest classification probability tables and retest classification percent tables by score range are presented for all CDT tests and benchmark cuts.

3 X 3 RETEST CLASSIFICATION PROBABILITY

Table E–1. Retest Classification Probability – Mathematics Grade 3

	Red - retest	Green - retest	Blue - retest
Red – test	0.955	0.045	0.000
Green – test	0.133	0.813	0.054
Blue – test	0.000	0.158	0.842

Exact Agreement Rate = 0.922
Kappa = 0.798
N-count = 23,191

Table E–2. Retest Classification Probability – Mathematics Grade 4

	Red - retest	Green - retest	Blue - retest
Red – test	0.953	0.047	0.000
Green – test	0.142	0.810	0.048
Blue – test	0.000	0.162	0.838

Exact Agreement Rate = 0.923
Kappa = 0.784
N-count = 24,069

Table E–3. Retest Classification Probability – Mathematics Grade 5

	Red - retest	Green - retest	Blue - retest
Red – test	0.948	0.052	0.000
Green – test	0.147	0.809	0.044
Blue – test	0.000	0.182	0.818

Exact Agreement Rate = 0.915
Kappa = 0.774
N-count = 28,759

Table E–4. Retest Classification Probability – Mathematics Grade 6

	Red - retest	Green - retest	Blue - retest
Red – test	0.950	0.050	0.000
Green – test	0.137	0.818	0.044
Blue – test	0.000	0.154	0.846

Exact Agreement Rate = 0.918
Kappa = 0.793
N-count = 34,688

Table E–5. Retest Classification Probability – Mathematics

Grade 7

	Red - retest	Green - retest	Blue - retest
Red – test	0.957	0.043	0.000
Green – test	0.156	0.809	0.035
Blue – test	0.000	0.162	0.838

Exact Agreement Rate = 0.932
Kappa = 0.776
N-count = 34,947

Table E–6. Retest Classification Probability – Mathematics Grade 8

	Red - retest	Green - retest	Blue - retest
Red – test	0.960	0.040	0.000
Green – test	0.162	0.804	0.035
Blue – test	0.000	0.168	0.832

Exact Agreement Rate = 0.936
Kappa = 0.767
N-count = 27,468

Table E–7. Retest Classification Probability – Mathematics High School

	Red - retest	Green - retest	Blue - retest
Red – test	0.976	0.024	0.000
Green – test	0.108	0.885	0.007
Blue – test	0.000	0.258	0.742

Exact Agreement Rate = 0.974
Kappa = 0.614
N-count = 454

Table E–8. Retest Classification Probability – Algebra I

	Red - retest	Green - retest	Blue - retest
Red – test	0.954	0.046	0.000
Green – test	0.157	0.809	0.034
Blue – test	0.000	0.179	0.821

Exact Agreement Rate = 0.927
Kappa = 0.767
N-count = 42,725

Table E–9. Retest Classification Probability – Geometry

	Red - retest	Green - retest	Blue - retest
Red – test	0.936	0.064	0.000
Green – test	0.143	0.813	0.044
Blue – test	0.000	0.159	0.841

Exact Agreement Rate = 0.903

Kappa = 0.772

N-count = 5,131

Table E–10. Retest Classification Probability – Algebra II

	Red - retest	Green - retest	Blue - retest
Red – test	0.933	0.067	0.000
Green – test	0.114	0.830	0.057
Blue – test	0.000	0.164	0.836

Exact Agreement Rate = 0.896
Kappa = 0.794
N-count = 6,023

Table E–11. Retest Classification Probability – Reading Grade 3

	Red - retest	Green - retest	Blue - retest
Red – test	0.941	0.059	0.000
Green – test	0.093	0.853	0.055
Blue – test	0.000	0.176	0.824

Exact Agreement Rate = 0.903
Kappa = 0.818
N-count = 19,838

Table E–12. Retest Classification Probability – Reading Grade 4

	Red - retest	Green - retest	Blue - retest
Red – test	0.944	0.056	0.000
Green – test	0.090	0.857	0.053
Blue – test	0.000	0.225	0.775

Exact Agreement Rate = 0.903
Kappa = 0.817
N-count = 20,691

Table E–13. Retest Classification Probability – Reading Grade 5

	Red - retest	Green - retest	Blue - retest
Red – test	0.939	0.061	0.000
Green – test	0.084	0.867	0.049
Blue – test	0.000	0.243	0.757

Exact Agreement Rate = 0.902
Kappa = 0.816
N-count = 24,324

Table E–14. Retest Classification Probability – Reading Grade 6

	Red - retest	Green - retest	Blue - retest
Red – test	0.936	0.064	0.000
Green – test	0.105	0.859	0.036
Blue – test	0.000	0.255	0.745

Exact Agreement Rate = 0.903
Kappa = 0.805
N-count = 25,753

Table E–15. Retest Classification Probability – Reading Grade 7

	Red - retest	Green - retest	Blue - retest
Red – test	0.939	0.061	0.000
Green – test	0.107	0.861	0.033
Blue – test	0.000	0.255	0.745

Exact Agreement Rate = 0.906
Kappa = 0.809
N-count = 26,355

Table E–16. Retest Classification Probability – Reading Grade 8

	Red - retest	Green - retest	Blue - retest
Red – test	0.943	0.057	0.000
Green – test	0.113	0.854	0.033
Blue – test	0.000	0.286	0.714

Exact Agreement Rate = 0.910
Kappa = 0.807
N-count = 24,777

Table E–17. Retest Classification Probability – Literature

	Red - retest	Green - retest	Blue - retest
Red – test	0.931	0.069	0.000
Green – test	0.098	0.858	0.043
Blue – test	0.000	0.265	0.735

Exact Agreement Rate = 0.894
Kappa = 0.798
N-count = 62,170

Table E–18. Retest Classification Probability – Science Grade 3

	Red - retest	Green - retest	Blue - retest
Red – test	0.922	0.078	0.000
Green – test	0.088	0.832	0.080
Blue – test	0.000	0.161	0.838

Exact Agreement Rate = 0.871
Kappa = 0.792
N-count = 1,557

Table E–19. Retest Classification Probability – Science Grade 4

	Red - retest	Green - retest	Blue - retest
Red – test	0.922	0.078	0.000
Green – test	0.100	0.829	0.071
Blue – test	0.000	0.204	0.796

Exact Agreement Rate = 0.870
Kappa = 0.777
N-count = 13,182

Table E–20. Retest Classification Probability – Science Grade 5

	Red - retest	Green - retest	Blue - retest
Red – test	0.930	0.070	0.000
Green – test	0.111	0.830	0.059
Blue – test	0.000	0.199	0.801

Exact Agreement Rate = 0.884
Kappa = 0.786
N-count = 5,939

Table E–21. Retest Classification Probability – Science Grade 6

	Red - retest	Green - retest	Blue - retest
Red – test	0.936	0.064	0.000
Green – test	0.105	0.839	0.056
Blue – test	0.000	0.227	0.773

Exact Agreement Rate = 0.893
Kappa = 0.795
N-count = 10,228

Table E–22. Retest Classification Probability – Science Grade 7

	Red - retest	Green - retest	Blue - retest
Red – test	0.940	0.060	0.000
Green – test	0.121	0.835	0.044
Blue – test	0.000	0.225	0.775

Exact Agreement Rate = 0.901
Kappa = 0.792
N-count = 15,341

Table E–23. Retest Classification Probability – Science Grade 8

	Red - retest	Green - retest	Blue - retest
Red – test	0.934	0.066	0.000
Green – test	0.132	0.830	0.039
Blue – test	0.000	0.245	0.755

Exact Agreement Rate = 0.895
Kappa = 0.779
N-count = 25,927

Table E–24. Retest Classification Probability – Science High School

	Red - retest	Green - retest	Blue - retest
Red – test	0.970	0.030	0.000
Green – test	0.188	0.792	0.020
Blue – test	0.000	0.458	0.542

Exact Agreement Rate = 0.957

Kappa = 0.733

N-count = 493

Table E–25. Retest Classification Probability – Biology

	Red - retest	Green - retest	Blue - retest
Red – test	0.930	0.070	0.000
Green – test	0.139	0.819	0.042
Blue – test	0.000	0.161	0.839

Exact Agreement Rate = 0.894

Kappa = 0.775

N-count = 46,477

Table E–26. Retest Classification Probability – Chemistry

	Red - retest	Green - retest	Blue - retest
Red – test	0.917	0.083	0.000
Green – test	0.158	0.811	0.031
Blue – test	0.000	0.184	0.816

Exact Agreement Rate = 0.887
Kappa = 0.736
N-count = 3,070

Table E–27. Retest Classification Probability – Writing Grade 3

	Red - retest	Green - retest	Blue - retest
Red – test	0.942	0.058	0.000
Green – test	0.106	0.829	0.065
Blue – test	0.000	0.204	0.796

Exact Agreement Rate = 0.895
Kappa = 0.800
N-count = 2,677

Table E–28. Retest Classification Probability – Writing Grade 4

	Red - retest	Green - retest	Blue - retest
Red – test	0.940	0.060	0.000
Green – test	0.125	0.826	0.050
Blue – test	0.000	0.192	0.808

Exact Agreement Rate = 0.896
Kappa = 0.793
N-count = 2,177

Table E–29. Retest Classification Probability – Writing Grade 5

	Red - retest	Green - retest	Blue - retest
Red – test	0.936	0.064	0.000
Green – test	0.124	0.828	0.047
Blue – test	0.000	0.229	0.771

Exact Agreement Rate = 0.892
Kappa = 0.785
N-count = 3,624

Table E–30. Retest Classification Probability – Writing Grade 6

	Red - retest	Green - retest	Blue - retest
Red – test	0.937	0.063	0.000
Green – test	0.113	0.836	0.052
Blue – test	0.000	0.200	0.800

Exact Agreement Rate = 0.895
Kappa = 0.795
N-count = 4,985

Table E–31. Retest Classification Probability – Writing Grade 7

	Red - retest	Green - retest	Blue - retest
Red – test	0.942	0.058	0.000
Green – test	0.113	0.841	0.046
Blue – test	0.000	0.221	0.779

Exact Agreement Rate = 0.899
Kappa = 0.801
N-count = 6,262

Table E–32. Retest Classification Probability – Writing Grade 8

	Red - retest	Green - retest	Blue - retest
Red – test	0.946	0.054	0.000
Green – test	0.112	0.844	0.044
Blue – test	0.000	0.224	0.776

Exact Agreement Rate = 0.906
Kappa = 0.808
N-count = 5,666

Table E–33. Retest Classification Probability – English Composition

	Red - retest	Green - retest	Blue - retest
Red – test	0.935	0.065	0.000
Green – test	0.105	0.845	0.050
Blue – test	0.000	0.196	0.804

Exact Agreement Rate = 0.890
Kappa = 0.805
N-count = 4,635

Retest Classification Percent for Various Scale Score Ranges

Tables E–34 through E–66 show the percent chance of scoring in each color range if retested without additional instruction for various scale scores ranges.

Table E–34. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 3

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	12	>99.9%	0.0%	0.0%	>99.9%
250 to 299	56	>99.9%	0.0%	0.0%	>99.9%
300 to 349	230	>99.9%	0.0%	0.0%	>99.9%
350 to 399	437	>99.9%	0.0%	0.0%	>99.9%
400 to 449	825	>99.9%	0.0%	0.0%	>99.9%
450 to 499	1,132	>99.9%	0.0%	0.0%	>99.9%
500 to 549	1,511	>99.9%	0.0%	0.0%	>99.9%
550 to 599	1,868	>99.9%	0.0%	0.0%	>99.9%
600 to 649	2,254	>99.9%	0.0%	0.0%	>99.9%
650 to 699	2,603	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,802	99.1%	0.9%	0.0%	99.1%
750 to 799	2,771	87.2%	12.8%	0.0%	87.2%
800 to 849 (Red/Green cut = 822)	2,273	47.8%	52.2%	0.0%	62.7%
850 to 899	1,793	10.4%	89.3%	0.3%	89.3%
900 to 949	1,196	0.7%	92.6%	6.7%	92.6%
950 to 999 (Green/Blue cut = 985)	689	0.0%	63.0%	37.0%	66.0%
1000 to 1049	375	0.0%	19.0%	81.0%	81.0%
1050 to 1099	192	0.0%	1.6%	98.4%	98.4%
1100 to 1149	92	0.0%	0.1%	99.9%	99.9%
1150 to 1199	42	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	22	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	7	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	2	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	5	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	2	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	23,191				

* Retest assuming no additional instruction

Table E–35. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 4

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	2	>99.9%	0.0%	0.0%	>99.9%
300 to 349	20	>99.9%	0.0%	0.0%	>99.9%
350 to 399	102	>99.9%	0.0%	0.0%	>99.9%
400 to 449	316	>99.9%	0.0%	0.0%	>99.9%
450 to 499	595	>99.9%	0.0%	0.0%	>99.9%
500 to 549	765	>99.9%	0.0%	0.0%	>99.9%
550 to 599	1,072	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,428	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,968	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,621	>99.9%	0.0%	0.0%	>99.9%
750 to 799	3,082	>99.9%	0.0%	0.0%	>99.9%
800 to 849	3,344	98.1%	1.9%	0.0%	98.1%
850 to 899	2,997	80.7%	19.3%	0.0%	80.7%
900 to 949 (Red/Green cut = 910)	2,272	36.7%	63.3%	0.0%	65.5%
950 to 999	1,532	6.0%	93.2%	0.8%	93.2%
1000 to 1049	926	0.3%	88.5%	11.2%	88.5%
1050 to 1099 (Green/Blue cut = 1073)	507	0.0%	51.5%	48.5%	62.4%
1100 to 1149	272	0.0%	11.4%	88.6%	88.6%
1150 to 1199	158	0.0%	0.9%	99.1%	99.1%
1200 to 1249	57	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	20	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	6	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	3	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	4	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	24,069				

* Retest assuming no additional instruction

Table E–36. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 5

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	4	>99.9%	0.0%	0.0%	>99.9%
350 to 399	20	>99.9%	0.0%	0.0%	>99.9%
400 to 449	116	>99.9%	0.0%	0.0%	>99.9%
450 to 499	332	>99.9%	0.0%	0.0%	>99.9%
500 to 549	586	>99.9%	0.0%	0.0%	>99.9%
550 to 599	849	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,129	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,664	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,218	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,938	>99.9%	0.0%	0.0%	>99.9%
800 to 849	3,665	99.9%	0.1%	0.0%	99.9%
850 to 899	4,018	97.8%	2.2%	0.0%	97.8%
900 to 949	3,864	78.9%	21.1%	0.0%	78.9%
950 to 999 (Red/Green cut = 958)	3,128	34.3%	65.7%	0.0%	66.9%
1000 to 1049	2,083	5.6%	93.6%	0.8%	93.6%
1050 to 1099	1,165	0.2%	88.0%	11.7%	88.0%
1100 to 1149 (Green/Blue cut = 1121)	564	0.0%	50.8%	49.2%	61.7%
1150 to 1199	202	0.0%	11.3%	88.7%	88.7%
1200 to 1249	103	0.0%	0.7%	99.3%	99.3%
1250 to 1299	45	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	31	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	13	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	9	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	6	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	5	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	2	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	28,759				

* Retest assuming no additional instruction

Table E–37. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 6

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	8	>99.9%	0.0%	0.0%	>99.9%
450 to 499	64	>99.9%	0.0%	0.0%	>99.9%
500 to 549	250	>99.9%	0.0%	0.0%	>99.9%
550 to 599	674	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,127	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,466	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,890	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,369	>99.9%	0.0%	0.0%	>99.9%
800 to 849	3,230	>99.9%	0.0%	0.0%	>99.9%
850 to 899	4,031	>99.9%	0.0%	0.0%	>99.9%
900 to 949	4,567	99.3%	0.7%	0.0%	99.3%
950 to 999	4,408	88.5%	11.5%	0.0%	88.5%
1000 to 1049 (Red/Green cut = 1023)	3,976	48.6%	51.4%	0.0%	62.9%
1050 to 1099	2,898	9.9%	89.9%	0.2%	89.9%
1100 to 1149	1,767	0.6%	94.0%	5.5%	94.0%
1150 to 1199 (Green/Blue cut = 1186)	945	0.0%	64.0%	36.0%	66.9%
1200 to 1249	537	0.0%	18.3%	81.7%	81.7%
1250 to 1299	258	0.0%	1.4%	98.6%	98.6%
1300 to 1349	123	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	56	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	21	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	15	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	6	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	1	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	1	0.0%	0.0%	>99.9%	>99.9%
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	34,688				

* Retest assuming no additional instruction

Table E–38. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 7

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	1	>99.9%	0.0%	0.0%	>99.9%
400 to 449	1	>99.9%	0.0%	0.0%	>99.9%
450 to 499	20	>99.9%	0.0%	0.0%	>99.9%
500 to 549	135	>99.9%	0.0%	0.0%	>99.9%
550 to 599	435	>99.9%	0.0%	0.0%	>99.9%
600 to 649	823	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,143	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,533	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,039	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,898	>99.9%	0.0%	0.0%	>99.9%
850 to 899	3,503	>99.9%	0.0%	0.0%	>99.9%
900 to 949	4,247	>99.9%	0.0%	0.0%	>99.9%
950 to 999	4,862	99.6%	0.4%	0.0%	99.6%
1000 to 1049	4,597	92.3%	7.7%	0.0%	92.3%
1050 to 1099 (Red/Green cut = 1082)	3,722	58.0%	42.0%	0.0%	64.3%
1100 to 1149	2,438	15.1%	84.8%	0.1%	84.8%
1150 to 1199	1,306	1.1%	95.8%	3.2%	95.8%
1200 to 1249 (Green/Blue cut = 1245)	635	0.0%	73.7%	26.3%	73.8%
1250 to 1299	338	0.0%	25.4%	74.6%	74.6%
1300 to 1349	160	0.0%	2.2%	97.8%	97.8%
1350 to 1399	58	0.0%	0.1%	99.9%	99.9%
1400 to 1449	30	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	14	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	4	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	4	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	1	0.0%	0.0%	>99.9%	>99.9%
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	34,947				

* Retest assuming no additional instruction

Table E–39. Retest Classification Percent for Various Scale Score Ranges – Mathematics Grade 8

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	3	>99.9%	0.0%	0.0%	>99.9%
450 to 499	14	>99.9%	0.0%	0.0%	>99.9%
500 to 549	68	>99.9%	0.0%	0.0%	>99.9%
550 to 599	281	>99.9%	0.0%	0.0%	>99.9%
600 to 649	621	>99.9%	0.0%	0.0%	>99.9%
650 to 699	979	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,288	>99.9%	0.0%	0.0%	>99.9%
750 to 799	1,513	>99.9%	0.0%	0.0%	>99.9%
800 to 849	1,875	>99.9%	0.0%	0.0%	>99.9%
850 to 899	2,233	>99.9%	0.0%	0.0%	>99.9%
900 to 949	2,857	>99.9%	0.0%	0.0%	>99.9%
950 to 999	3,400	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	3,477	99.2%	0.8%	0.0%	99.2%
1050 to 1099	3,320	87.7%	12.3%	0.0%	87.7%
1100 to 1149 (Red/Green cut = 1121)	2,603	47.8%	52.2%	0.0%	62.7%
1150 to 1199	1,471	9.7%	90.0%	0.2%	90.0%
1200 to 1249	771	0.5%	93.5%	6.0%	93.5%
1250 to 1299 (Green/Blue cut = 1284)	354	0.0%	62.0%	38.0%	65.8%
1300 to 1349	196	0.0%	16.7%	83.3%	83.3%
1350 to 1399	73	0.0%	1.4%	98.6%	98.6%
1400 to 1449	41	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	18	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	6	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	4	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	2	0.0%	0.0%	>99.9%	>99.9%
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	27,468				

* Retest assuming no additional instruction

Table E–40. Retest Classification Percent for Various Scale Score Ranges – Mathematics High School

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	0	N/A	N/A	N/A	N/A
500 to 549	3	>99.9%	0.0%	0.0%	>99.9%
3550 to 3599	9	>99.9%	0.0%	0.0%	>99.9%
600 to 649	20	>99.9%	0.0%	0.0%	>99.9%
650 to 699	27	>99.9%	0.0%	0.0%	>99.9%
700 to 749	36	>99.9%	0.0%	0.0%	>99.9%
750 to 799	55	>99.9%	0.0%	0.0%	>99.9%
800 to 849	52	>99.9%	0.0%	0.0%	>99.9%
850 to 899	36	>99.9%	0.0%	0.0%	>99.9%
900 to 949	62	>99.9%	0.0%	0.0%	>99.9%
950 to 999	57	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	39	99.4%	0.6%	0.0%	99.4%
1050 to 1099	33	86.3%	13.7%	0.0%	86.3%
1100 to 1149 (Red/Green cut = 1134)	16	54.0%	46.0%	0.0%	63.3%
1150 to 1199	5	8.4%	91.3%	0.3%	91.3%
1200 to 1249	3	1.2%	97.1%	1.7%	97.1%
1250 to 1299 (Green/Blue cut = 1297)	0	N/A	N/A	N/A	N/A
1300 to 1349	1	0.0%	25.8%	74.2%	74.2%
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	454				

* Retest assuming no additional instruction

Table E–41. Retest Classification Percent for Various Scale Score Ranges – Algebra I

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	2	>99.9%	0.0%	0.0%	>99.9%
500 to 549	22	>99.9%	0.0%	0.0%	>99.9%
550 to 599	206	>99.9%	0.0%	0.0%	>99.9%
600 to 649	615	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,157	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,632	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,160	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,469	>99.9%	0.0%	0.0%	>99.9%
850 to 899	2,997	>99.9%	0.0%	0.0%	>99.9%
900 to 949	3,770	>99.9%	0.0%	0.0%	>99.9%
950 to 999	4,710	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	5,661	99.7%	0.3%	0.0%	99.7%
1050 to 1099	5,965	93.0%	7.0%	0.0%	93.0%
1100 to 1149 (Red/Green cut = 1134)	4,983	59.8%	40.2%	0.0%	64.8%
1150 to 1199	3,182	16.2%	83.7%	0.1%	83.7%
1200 to 1249	1,711	1.2%	96.0%	2.9%	96.0%
1250 to 1299 (Green/Blue cut = 1297)	811	0.0%	74.4%	25.6%	74.4%
1300 to 1349	389	0.0%	26.7%	73.3%	73.3%
1350 to 1399	173	0.0%	3.3%	96.7%	96.7%
1400 to 1449	69	0.0%	0.1%	99.9%	99.9%
1450 to 1499	28	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	8	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	3	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	2	0.0%	0.0%	>99.9%	>99.9%
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	42,725				

* Retest assuming no additional instruction

Table E–42. Retest Classification Percent for Various Scale Score Ranges – Geometry

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	1	>99.9%	0.0%	0.0%	>99.9%
500 to 549	3	>99.9%	0.0%	0.0%	>99.9%
550 to 599	12	>99.9%	0.0%	0.0%	>99.9%
600 to 649	20	>99.9%	0.0%	0.0%	>99.9%
650 to 699	47	>99.9%	0.0%	0.0%	>99.9%
700 to 749	74	>99.9%	0.0%	0.0%	>99.9%
750 to 799	106	>99.9%	0.0%	0.0%	>99.9%
800 to 849	141	>99.9%	0.0%	0.0%	>99.9%
850 to 899	221	>99.9%	0.0%	0.0%	>99.9%
900 to 949	305	>99.9%	0.0%	0.0%	>99.9%
950 to 999	418	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	581	>99.9%	0.0%	0.0%	>99.9%
1050 to 1099	711	98.7%	1.3%	0.0%	98.7%
1100 to 1149	857	84.3%	15.7%	0.0%	84.3%
1150 to 1199 (Red/Green cut = 1165)	669	40.6%	59.4%	0.0%	64.2%
1200 to 1249	422	7.0%	92.6%	0.4%	92.6%
1250 to 1299	262	0.3%	92.6%	7.2%	92.6%
1300 to 1349 (Green/Blue cut = 1328)	144	0.0%	56.8%	43.2%	63.4%
1350 to 1399	71	0.0%	13.7%	86.3%	86.3%
1400 to 1449	36	0.0%	1.3%	98.7%	98.7%
1450 to 1499	19	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	6	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	5	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	5,131				

* Retest assuming no additional instruction

Table E–43. Retest Classification Percent for Various Scale Score Ranges – Algebra II

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	0	N/A	N/A	N/A	N/A
500 to 549	0	N/A	N/A	N/A	N/A
550 to 599	1	>99.9%	0.0%	0.0%	>99.9%
600 to 649	3	>99.9%	0.0%	0.0%	>99.9%
650 to 699	10	>99.9%	0.0%	0.0%	>99.9%
700 to 749	34	>99.9%	0.0%	0.0%	>99.9%
750 to 799	59	>99.9%	0.0%	0.0%	>99.9%
800 to 849	70	>99.9%	0.0%	0.0%	>99.9%
850 to 899	118	>99.9%	0.0%	0.0%	>99.9%
900 to 949	160	>99.9%	0.0%	0.0%	>99.9%
950 to 999	241	>99.9%	0.0%	0.0%	>99.9%
1000 to 1049	370	>99.9%	0.0%	0.0%	>99.9%
1050 to 1099	601	>99.9%	0.0%	0.0%	>99.9%
1100 to 1149	821	99.5%	0.5%	0.0%	99.5%
1150 to 1199	907	90.7%	9.3%	0.0%	90.7%
1200 to 1249 (Red/Green cut = 1228)	789	53.8%	46.2%	0.0%	63.4%
1250 to 1299	664	11.8%	88.0%	0.1%	88.0%
1300 to 1349	490	0.8%	95.2%	4.1%	95.2%
1350 to 1399 (Green/Blue cut = 1391)	333	0.0%	68.3%	31.7%	69.7%
1400 to 1449	182	0.0%	22.3%	77.7%	77.7%
1450 to 1499	92	0.0%	2.2%	97.8%	97.8%
1500 to 1549	44	0.0%	0.1%	99.9%	99.9%
1550 to 1599	20	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	6	0.0%	0.0%	>99.9%	>99.9%
1650 to 1699	4	0.0%	0.0%	>99.9%	>99.9%
1700 to 1749	1	0.0%	0.0%	>99.9%	>99.9%
1750 to 1799	2	0.0%	0.0%	>99.9%	>99.9%
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	1	0.0%	0.4%	99.6%	99.6%
TOTAL	6,023				

* Retest assuming no additional instruction

Table E–44. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 3

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	1	>99.9%	0.0%	0.0%	>99.9%
350 to 399	15	>99.9%	0.0%	0.0%	>99.9%
400 to 449	157	>99.9%	0.0%	0.0%	>99.9%
450 to 499	668	>99.9%	0.0%	0.0%	>99.9%
500 to 549	1,850	>99.9%	0.0%	0.0%	>99.9%
550 to 599	2,566	>99.9%	0.0%	0.0%	>99.9%
600 to 649	2,580	99.5%	0.5%	0.0%	99.5%
650 to 699	2,227	93.3%	6.7%	0.0%	93.3%
700 to 749 (Red/Green cut = 741)	2,023	64.2%	35.8%	0.0%	65.9%
750 to 799	1,933	21.5%	78.5%	0.0%	78.5%
800 to 849	1,715	2.7%	97.2%	0.1%	97.2%
850 to 899	1,434	0.1%	96.7%	3.2%	96.7%
900 to 949	1,133	0.0%	76.7%	23.3%	76.7%
950 to 999 (Green/Blue cut = 956)	763	0.0%	35.7%	64.3%	65.4%
1000 to 1049	460	0.0%	8.1%	91.9%	91.9%
1050 to 1099	191	0.0%	0.9%	99.1%	99.1%
1100 to 1149	86	0.0%	0.1%	99.9%	99.9%
1150 to 1199	23	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	8	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	5	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	19,838				

* Retest assuming no additional instruction

Table E–45. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 4

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	7	>99.9%	0.0%	0.0%	>99.9%
400 to 449	43	>99.9%	0.0%	0.0%	>99.9%
450 to 499	248	>99.9%	0.0%	0.0%	>99.9%
500 to 549	921	>99.9%	0.0%	0.0%	>99.9%
550 to 599	1,766	>99.9%	0.0%	0.0%	>99.9%
600 to 649	2,043	>99.9%	0.0%	0.0%	>99.9%
650 to 699	2,001	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,903	99.0%	1.0%	0.0%	99.0%
750 to 799	2,058	88.2%	11.8%	0.0%	88.2%
800 to 849 (Red/Green cut = 826)	2,077	50.6%	49.4%	0.0%	61.9%
850 to 899	2,071	13.1%	86.9%	0.0%	86.9%
900 to 949	1,896	1.2%	98.4%	0.4%	98.4%
950 to 999	1,572	0.0%	93.5%	6.4%	93.5%
1000 to 1049 (Green/Blue cut = 1041)	1,106	0.0%	65.7%	34.3%	66.8%
1050 to 1099	634	0.0%	26.1%	73.9%	73.9%
1100 to 1149	234	0.0%	5.4%	94.6%	94.6%
1150 to 1199	80	0.0%	0.7%	99.3%	99.3%
1200 to 1249	20	0.0%	0.1%	99.9%	99.9%
1250 to 1299	7	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	4	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	20,691				

* Retest assuming no additional instruction

Table E–46. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 5

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	14	>99.9%	0.0%	0.0%	>99.9%
450 to 499	67	>99.9%	0.0%	0.0%	>99.9%
500 to 549	314	>99.9%	0.0%	0.0%	>99.9%
550 to 599	878	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,715	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,900	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,983	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,048	99.6%	0.4%	0.0%	99.6%
800 to 849	2,364	93.2%	6.8%	0.0%	93.2%
850 to 899 (Red/Green cut = 890)	2,477	63.0%	37.0%	0.0%	65.3%
900 to 949	2,701	20.5%	79.5%	0.0%	79.5%
950 to 999	2,800	2.5%	97.3%	0.2%	97.3%
1000 to 1049	2,465	0.1%	96.5%	3.4%	96.5%
1050 to 1099	1,482	0.0%	76.1%	23.9%	76.1%
1100 to 1149 (Green/Blue cut = 1105)	751	0.0%	36.6%	63.4%	64.2%
1150 to 1199	268	0.0%	9.3%	90.7%	90.7%
1200 to 1249	76	0.0%	1.7%	98.3%	98.3%
1250 to 1299	16	0.0%	0.2%	99.8%	99.8%
1300 to 1349	4	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	1	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	24,324				

* Retest assuming no additional instruction

Table E–47. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 6

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	3	>99.9%	0.0%	0.0%	>99.9%
450 to 499	11	>99.9%	0.0%	0.0%	>99.9%
500 to 549	65	>99.9%	0.0%	0.0%	>99.9%
550 to 599	357	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,035	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,892	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,147	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,150	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,346	99.7%	0.3%	0.0%	99.7%
850 to 899	2,714	94.7%	5.3%	0.0%	94.7%
900 to 949 (Red/Green cut = 945)	3,021	67.0%	33.0%	0.0%	67.6%
950 to 999	3,235	24.1%	75.9%	0.0%	75.9%
1000 to 1049	2,814	3.2%	96.8%	0.1%	96.8%
1050 to 1099	2,033	0.2%	97.5%	2.3%	97.5%
1100 to 1149	1,158	0.0%	80.6%	19.4%	80.6%
1150 to 1199 (Green/Blue cut = 1160)	538	0.0%	41.7%	58.3%	61.3%
1200 to 1249	164	0.0%	12.6%	87.4%	87.4%
1250 to 1299	51	0.0%	2.4%	97.6%	97.6%
1300 to 1349	14	0.0%	0.6%	99.4%	99.4%
1350 to 1399	3	0.0%	0.1%	99.9%	99.9%
1400 to 1449	2	0.0%	0.1%	99.9%	99.9%
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	25,753				

* Retest assuming no additional instruction

Table E–48. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 7

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	1	>99.9%	0.0%	0.0%	>99.9%
400 to 449	1	>99.9%	0.0%	0.0%	>99.9%
450 to 499	7	>99.9%	0.0%	0.0%	>99.9%
500 to 549	61	>99.9%	0.0%	0.0%	>99.9%
550 to 599	306	>99.9%	0.0%	0.0%	>99.9%
600 to 649	971	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,781	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,997	>99.9%	0.0%	0.0%	>99.9%
750 to 799	1,975	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,059	>99.9%	0.0%	0.0%	>99.9%
850 to 899	2,370	99.2%	0.8%	0.0%	99.2%
900 to 949	2,816	89.2%	10.8%	0.0%	89.2%
950 to 999 (Red/Green cut = 979)	3,137	53.0%	47.0%	0.0%	62.1%
1000 to 1049	3,307	14.1%	85.9%	0.0%	85.9%
1050 to 1099	2,574	1.4%	98.4%	0.3%	98.4%
1100 to 1149	1,682	0.1%	94.5%	5.4%	94.5%
1150 to 1199 (Green/Blue cut = 1194)	820	0.0%	69.3%	30.7%	69.6%
1200 to 1249	339	0.0%	30.0%	70.0%	70.0%
1250 to 1299	100	0.0%	7.3%	92.7%	92.7%
1300 to 1349	42	0.0%	1.5%	98.5%	98.5%
1350 to 1399	8	0.0%	0.3%	99.7%	99.7%
1400 to 1449	1	0.0%	0.1%	99.9%	99.9%
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	26,355				

* Retest assuming no additional instruction

Table E–49. Retest Classification Percent for Various Scale Score Ranges – Reading Grade 8

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	1	>99.9%	0.0%	0.0%	>99.9%
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	4	>99.9%	0.0%	0.0%	>99.9%
500 to 549	22	>99.9%	0.0%	0.0%	>99.9%
550 to 599	171	>99.9%	0.0%	0.0%	>99.9%
600 to 649	598	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,422	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,016	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,000	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,066	>99.9%	0.0%	0.0%	>99.9%
850 to 899	1,986	99.9%	0.1%	0.0%	99.9%
900 to 949	2,426	97.5%	2.5%	0.0%	97.5%
950 to 999	2,663	79.2%	20.8%	0.0%	79.2%
1000 to 1049 (Red/Green cut = 1011)	2,997	36.8%	63.2%	0.0%	65.3%
1050 to 1099	2,674	7.0%	93.0%	0.0%	93.0%
1100 to 1149	1,900	0.5%	98.5%	1.0%	98.5%
1150 to 1199	1,102	0.0%	88.1%	11.8%	88.1%
1200 to 1249 (Green/Blue cut = 1226)	524	0.0%	54.3%	45.7%	61.0%
1250 to 1299	156	0.0%	20.5%	79.5%	79.5%
1300 to 1349	44	0.0%	4.9%	95.1%	95.1%
1350 to 1399	4	0.0%	1.4%	98.6%	98.6%
1400 to 1449	1	0.0%	0.2%	99.8%	99.8%
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	24,777				

* Retest assuming no additional instruction

Table E–50. Retest Classification Percent for Various Scale Score Ranges – Literature

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	11	>99.9%	0.0%	0.0%	>99.9%
500 to 549	26	>99.9%	0.0%	0.0%	>99.9%
550 to 599	188	>99.9%	0.0%	0.0%	>99.9%
600 to 649	740	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,971	>99.9%	0.0%	0.0%	>99.9%
700 to 749	3,257	>99.9%	0.0%	0.0%	>99.9%
750 to 799	3,768	>99.9%	0.0%	0.0%	>99.9%
800 to 849	3,717	>99.9%	0.0%	0.0%	>99.9%
850 to 899	4,019	>99.9%	0.0%	0.0%	>99.9%
900 to 949	4,855	99.3%	0.7%	0.0%	99.3%
950 to 999	6,118	90.5%	9.5%	0.0%	90.5%
1000 to 1049 (Red/Green cut = 1033)	7,503	56.1%	43.9%	0.0%	62.4%
1050 to 1099	8,311	16.5%	83.5%	0.0%	83.5%
1100 to 1149	7,439	1.9%	97.8%	0.3%	97.8%
1150 to 1199	5,330	0.1%	94.7%	5.2%	94.7%
1200 to 1249 (Green/Blue cut = 1248)	3,108	0.0%	70.7%	29.3%	70.8%
1250 to 1299	1,241	0.0%	33.5%	66.5%	66.5%
1300 to 1349	422	0.0%	10.1%	89.9%	89.9%
1350 to 1399	115	0.0%	2.5%	97.5%	97.5%
1400 to 1449	22	0.0%	0.8%	99.2%	99.2%
1450 to 1499	8	0.0%	0.3%	99.7%	99.7%
1500 to 1549	1	0.0%	0.4%	99.6%	99.6%
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	62,170				

* Retest assuming no additional instruction

Table E–51. Retest Classification Percent for Various Scale Score Ranges – Science Grade 3

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	4	>99.9%	0.0%	0.0%	>99.9%
250 to 299	27	>99.9%	0.0%	0.0%	>99.9%
300 to 349	43	>99.9%	0.0%	0.0%	>99.9%
350 to 399	63	>99.9%	0.0%	0.0%	>99.9%
400 to 449	67	>99.9%	0.0%	0.0%	>99.9%
450 to 499	54	>99.9%	0.0%	0.0%	>99.9%
500 to 549	57	>99.9%	0.0%	0.0%	>99.9%
550 to 599	93	99.7%	0.3%	0.0%	99.7%
600 to 649	114	94.2%	5.8%	0.0%	94.2%
650 to 699 (Red/Green cut = 694)	164	65.3%	34.7%	0.0%	66.1%
700 to 749	179	22.7%	77.3%	0.0%	77.3%
750 to 799	204	3.1%	95.3%	1.6%	95.3%
800 to 849	209	0.1%	84.4%	15.5%	84.4%
850 to 899 (Green/Blue cut = 867)	123	0.0%	43.0%	57.0%	62.6%
900 to 949	89	0.0%	10.0%	90.0%	90.0%
950 to 999	42	0.0%	0.6%	99.4%	99.4%
1000 to 1049	19	0.0%	0.0%	>99.9%	>99.9%
1050 to 1099	2	0.0%	0.0%	>99.9%	>99.9%
1100 to 1149	3	0.0%	0.0%	>99.9%	>99.9%
1150 to 1199	0	N/A	N/A	N/A	N/A
1200 to 1249	1	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	1,557				

* Retest assuming no additional instruction

Table E–52. Retest Classification Percent for Various Scale Score Ranges – Science Grade 4

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	4	>99.9%	0.0%	0.0%	>99.9%
250 to 299	34	>99.9%	0.0%	0.0%	>99.9%
300 to 349	138	>99.9%	0.0%	0.0%	>99.9%
350 to 399	312	>99.9%	0.0%	0.0%	>99.9%
400 to 449	453	>99.9%	0.0%	0.0%	>99.9%
450 to 499	521	>99.9%	0.0%	0.0%	>99.9%
500 to 549	592	>99.9%	0.0%	0.0%	>99.9%
550 to 599	678	>99.9%	0.0%	0.0%	>99.9%
600 to 649	871	99.8%	0.2%	0.0%	99.8%
650 to 699	1,111	96.0%	4.0%	0.0%	96.0%
700 to 749	1,495	71.6%	28.4%	0.0%	71.6%
750 to 799 (Red/Green cut = 751)	1,773	28.3%	71.7%	0.0%	71.7%
800 to 849	1,882	4.3%	94.7%	1.0%	94.7%
850 to 899	1,520	0.2%	87.9%	11.9%	87.9%
900 to 949 (Green/Blue cut = 924)	1,034	0.0%	51.4%	48.6%	62.0%
950 to 999	475	0.0%	13.6%	86.4%	86.4%
1000 to 1049	203	0.0%	1.3%	98.7%	98.7%
1050 to 1099	66	0.0%	0.0%	>99.9%	>99.9%
1100 to 1149	10	0.0%	0.0%	>99.9%	>99.9%
1150 to 1199	8	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	1	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	1	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	13,182				

* Retest assuming no additional instruction

Table E–53. Retest Classification Percent for Various Scale Score Ranges – Science Grade 5

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	2	>99.9%	0.0%	0.0%	>99.9%
300 to 349	14	>99.9%	0.0%	0.0%	>99.9%
350 to 399	99	>99.9%	0.0%	0.0%	>99.9%
400 to 449	193	>99.9%	0.0%	0.0%	>99.9%
450 to 499	266	>99.9%	0.0%	0.0%	>99.9%
500 to 549	295	>99.9%	0.0%	0.0%	>99.9%
550 to 599	330	>99.9%	0.0%	0.0%	>99.9%
600 to 649	349	>99.9%	0.0%	0.0%	>99.9%
650 to 699	440	99.9%	0.1%	0.0%	99.9%
700 to 749	550	96.6%	3.4%	0.0%	96.6%
750 to 799	691	74.0%	26.0%	0.0%	74.0%
800 to 849 (Red/Green cut = 804)	803	30.7%	69.3%	0.0%	69.5%
850 to 899	793	4.8%	94.3%	0.8%	94.3%
900 to 949	570	0.3%	89.1%	10.7%	89.1%
950 to 999 (Green/Blue cut = 977)	302	0.0%	55.2%	44.8%	62.0%
1000 to 1049	166	0.0%	13.7%	86.3%	86.3%
1050 to 1099	56	0.0%	1.4%	98.6%	98.6%
1100 to 1149	15	0.0%	0.1%	99.9%	99.9%
1150 to 1199	3	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	2	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	5,939				

* Retest assuming no additional instruction

Table E–54. Retest Classification Percent for Various Scale Score Ranges – Science Grade 6

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	15	>99.9%	0.0%	0.0%	>99.9%
450 to 499	90	>99.9%	0.0%	0.0%	>99.9%
500 to 549	347	>99.9%	0.0%	0.0%	>99.9%
550 to 599	702	>99.9%	0.0%	0.0%	>99.9%
600 to 649	807	>99.9%	0.0%	0.0%	>99.9%
650 to 699	762	>99.9%	0.0%	0.0%	>99.9%
700 to 749	878	99.9%	0.1%	0.0%	99.9%
750 to 799	958	97.9%	2.1%	0.0%	97.9%
800 to 849	1,140	79.6%	20.4%	0.0%	79.6%
850 to 899 (Red/Green cut = 861)	1,284	36.4%	63.6%	0.0%	65.8%
900 to 949	1,347	6.2%	93.4%	0.4%	93.4%
950 to 999	1,006	0.4%	92.8%	6.9%	92.8%
1000 to 1049 (Green/Blue cut = 1034)	561	0.0%	61.7%	38.3%	65.2%
1050 to 1099	228	0.0%	20.2%	79.8%	79.8%
1100 to 1149	70	0.0%	1.8%	98.2%	98.2%
1150 to 1199	22	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	9	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	1	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	1	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	10,228				

* Retest assuming no additional instruction

Table E–55. Retest Classification Percent for Various Scale Score Ranges – Science Grade 7

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	7	>99.9%	0.0%	0.0%	>99.9%
450 to 499	65	>99.9%	0.0%	0.0%	>99.9%
500 to 549	364	>99.9%	0.0%	0.0%	>99.9%
550 to 599	933	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,219	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,140	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,200	>99.9%	0.0%	0.0%	>99.9%
750 to 799	1,286	99.9%	0.1%	0.0%	99.9%
800 to 849	1,544	97.4%	2.6%	0.0%	97.4%
850 to 899	1,810	77.6%	22.4%	0.0%	77.6%
900 to 949 (Red/Green cut = 908)	2,015	34.0%	66.0%	0.0%	67.1%
950 to 999	1,856	5.7%	93.8%	0.5%	93.8%
1000 to 1049	1,218	0.3%	91.9%	7.8%	91.9%
1050 to 1099 (Green/Blue cut = 1081)	480	0.0%	60.8%	39.2%	64.9%
1100 to 1149	141	0.0%	16.5%	83.5%	83.5%
1150 to 1199	38	0.0%	1.7%	98.3%	98.3%
1200 to 1249	13	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	5	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	6	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	1	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	15,341				

* Retest assuming no additional instruction

Table E–56. Retest Classification Percent for Various Scale Score Ranges – Science Grade 8

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	2	>99.9%	0.0%	0.0%	>99.9%
450 to 499	51	>99.9%	0.0%	0.0%	>99.9%
500 to 549	312	>99.9%	0.0%	0.0%	>99.9%
550 to 599	1,064	>99.9%	0.0%	0.0%	>99.9%
600 to 649	1,650	>99.9%	0.0%	0.0%	>99.9%
650 to 699	1,616	>99.9%	0.0%	0.0%	>99.9%
700 to 749	1,635	>99.9%	0.0%	0.0%	>99.9%
750 to 799	1,863	>99.9%	0.0%	0.0%	>99.9%
800 to 849	2,237	99.8%	0.2%	0.0%	99.8%
850 to 899	2,831	95.8%	4.2%	0.0%	95.8%
900 to 949 (Red/Green cut = 949)	3,462	70.7%	29.3%	0.0%	70.7%
950 to 999	3,793	26.5%	73.5%	0.0%	73.5%
1000 to 1049	2,899	3.7%	95.4%	0.9%	95.4%
1050 to 1099	1,621	0.1%	88.7%	11.2%	88.7%
1100 to 1149 (Green/Blue cut = 1122)	628	0.0%	52.1%	47.9%	62.1%
1150 to 1199	190	0.0%	13.1%	86.9%	86.9%
1200 to 1249	55	0.0%	0.9%	99.1%	99.1%
1250 to 1299	16	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	1	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	1	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	25,927				

* Retest assuming no additional instruction

Table E–57. Retest Classification Percent for Various Scale Score Ranges – Science High School

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	2	>99.9%	0.0%	0.0%	>99.9%
500 to 549	16	>99.9%	0.0%	0.0%	>99.9%
550 to 599	27	>99.9%	0.0%	0.0%	>99.9%
600 to 649	33	>99.9%	0.0%	0.0%	>99.9%
650 to 699	56	>99.9%	0.0%	0.0%	>99.9%
700 to 749	40	>99.9%	0.0%	0.0%	>99.9%
750 to 799	52	>99.9%	0.0%	0.0%	>99.9%
800 to 849	54	>99.9%	0.0%	0.0%	>99.9%
850 to 899	61	99.9%	0.1%	0.0%	99.9%
900 to 949	56	98.2%	1.8%	0.0%	98.2%
950 to 999	47	82.2%	17.8%	0.0%	82.2%
1000 to 1049 (Red/Green cut = 1012)	27	40.3%	59.6%	0.0%	63.1%
1050 to 1099	16	8.8%	91.0%	0.2%	91.0%
1100 to 1149	3	0.6%	94.9%	4.6%	94.9%
1150 to 1199 (Green/Blue cut = 1185)	3	0.0%	61.6%	38.4%	64.4%
1200 to 1249	0	N/A	N/A	N/A	N/A
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	493				

* Retest assuming no additional instruction

Table E–58. Retest Classification Percent for Various Scale Score Ranges – Biology

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	2	>99.9%	0.0%	0.0%	>99.9%
450 to 499	6	>99.9%	0.0%	0.0%	>99.9%
500 to 549	59	>99.9%	0.0%	0.0%	>99.9%
550 to 599	298	>99.9%	0.0%	0.0%	>99.9%
600 to 649	969	>99.9%	0.0%	0.0%	>99.9%
650 to 699	2,162	>99.9%	0.0%	0.0%	>99.9%
700 to 749	2,680	>99.9%	0.0%	0.0%	>99.9%
750 to 799	2,940	>99.9%	0.0%	0.0%	>99.9%
800 to 849	3,552	>99.9%	0.0%	0.0%	>99.9%
850 to 899	4,365	99.9%	0.1%	0.0%	99.9%
900 to 949	5,500	97.9%	2.1%	0.0%	97.9%
950 to 999	6,739	80.6%	19.4%	0.0%	80.6%
1000 to 1049 (Red/Green cut = 1012)	6,536	38.0%	62.0%	0.0%	64.8%
1050 to 1099	4,828	6.9%	92.7%	0.4%	92.7%
1100 to 1149	2,814	0.4%	92.9%	6.7%	92.9%
1150 to 1199 (Green/Blue cut = 1185)	1,482	0.0%	62.4%	37.6%	65.6%
1200 to 1249	801	0.0%	18.6%	81.4%	81.4%
1250 to 1299	427	0.0%	1.8%	98.2%	98.2%
1300 to 1349	195	0.0%	0.1%	99.9%	99.9%
1350 to 1399	76	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	30	0.0%	0.0%	>99.9%	>99.9%
1450 to 1499	9	0.0%	0.0%	>99.9%	>99.9%
1500 to 1549	5	0.0%	0.0%	>99.9%	>99.9%
1550 to 1599	2	0.0%	0.0%	>99.9%	>99.9%
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	46,477				

* Retest assuming no additional instruction

Table E–59. Retest Classification Percent for Various Scale Score Ranges – Chemistry

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 400	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	0	N/A	N/A	N/A	N/A
500 to 549	0	N/A	N/A	N/A	N/A
550 to 599	2	>99.9%	0.0%	0.0%	>99.9%
600 to 649	5	>99.9%	0.0%	0.0%	>99.9%
650 to 699	15	>99.9%	0.0%	0.0%	>99.9%
700 to 749	49	>99.9%	0.0%	0.0%	>99.9%
750 to 799	117	>99.9%	0.0%	0.0%	>99.9%
800 to 849	217	>99.9%	0.0%	0.0%	>99.9%
850 to 899	292	>99.9%	0.0%	0.0%	>99.9%
900 to 949	437	99.8%	0.2%	0.0%	99.8%
950 to 999	541	94.9%	5.1%	0.0%	94.9%
1000 to 1049 (Red/Green cut = 1045)	569	68.1%	31.9%	0.0%	68.6%
1050 to 1099	382	25.0%	74.9%	0.0%	74.9%
1100 to 1149	243	2.8%	96.1%	1.1%	96.1%
1150 to 1199	105	0.1%	87.1%	12.8%	87.1%
1200 to 1249 (Green/Blue cut = 1218)	54	0.0%	45.0%	55.0%	63.1%
1250 to 1299	27	0.0%	9.0%	91.0%	91.0%
1300 to 1349	9	0.0%	0.9%	99.1%	99.1%
1350 to 1399	6	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	3,070				

* Retest assuming no additional instruction

Table E–60. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 3

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	1	>99.9%	0.0%	0.0%	>99.9%
250 to 299	4	>99.9%	0.0%	0.0%	>99.9%
300 to 349	28	>99.9%	0.0%	0.0%	>99.9%
350 to 399	78	>99.9%	0.0%	0.0%	>99.9%
400 to 449	136	>99.9%	0.0%	0.0%	>99.9%
450 to 499	171	>99.9%	0.0%	0.0%	>99.9%
500 to 549	164	>99.9%	0.0%	0.0%	>99.9%
550 to 599	159	>99.9%	0.0%	0.0%	>99.9%
600 to 649	185	>99.9%	0.0%	0.0%	>99.9%
650 to 699	218	99.4%	0.6%	0.0%	99.4%
700 to 749	274	90.6%	9.4%	0.0%	90.6%
750 to 799 (Red/Green cut = 780)	297	53.8%	46.2%	0.0%	61.4%
800 to 849	308	14.4%	85.5%	0.1%	85.5%
850 to 899	276	1.2%	95.9%	2.9%	95.9%
900 to 949	211	0.0%	76.8%	23.1%	76.8%
950 to 999 (Green/Blue cut = 953)	102	0.0%	33.2%	66.8%	67.0%
1000 to 1049	38	0.0%	5.3%	94.7%	94.7%
1050 to 1099	15	0.0%	0.3%	99.7%	99.7%
1100 to 1149	10	0.0%	0.0%	>99.9%	>99.9%
1150 to 1199	2	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	0	N/A	N/A	N/A	N/A
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	2,677				

* Retest assuming no additional instruction

Table E–61. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 4

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	2	>99.9%	0.0%	0.0%	>99.9%
300 to 349	4	>99.9%	0.0%	0.0%	>99.9%
350 to 399	26	>99.9%	0.0%	0.0%	>99.9%
400 to 449	64	>99.9%	0.0%	0.0%	>99.9%
450 to 499	120	>99.9%	0.0%	0.0%	>99.9%
500 to 549	105	>99.9%	0.0%	0.0%	>99.9%
550 to 599	105	>99.9%	0.0%	0.0%	>99.9%
600 to 649	127	>99.9%	0.0%	0.0%	>99.9%
650 to 699	132	>99.9%	0.0%	0.0%	>99.9%
700 to 749	183	99.9%	0.1%	0.0%	99.9%
750 to 799	207	96.6%	3.4%	0.0%	96.6%
800 to 849	266	73.2%	26.8%	0.0%	73.2%
850 to 899 (Red/Green cut = 852)	301	27.8%	72.2%	0.0%	72.2%
900 to 949	219	4.3%	95.0%	0.7%	95.0%
950 to 999	163	0.2%	89.4%	10.4%	89.4%
1000 to 1049 (Green/Blue cut = 1025)	85	0.0%	52.6%	47.4%	62.6%
1050 to 1099	42	0.0%	13.5%	86.5%	86.5%
1100 to 1149	20	0.0%	1.1%	98.9%	98.9%
1150 to 1199	4	0.0%	0.0%	>99.9%	>99.9%
1200 to 1249	2	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	2,177				

* Retest assuming no additional instruction

Table E–62. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 5

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	2	>99.9%	0.0%	0.0%	>99.9%
350 to 399	14	>99.9%	0.0%	0.0%	>99.9%
400 to 449	61	>99.9%	0.0%	0.0%	>99.9%
450 to 499	104	>99.9%	0.0%	0.0%	>99.9%
500 to 549	143	>99.9%	0.0%	0.0%	>99.9%
550 to 599	136	>99.9%	0.0%	0.0%	>99.9%
600 to 649	187	>99.9%	0.0%	0.0%	>99.9%
650 to 699	181	>99.9%	0.0%	0.0%	>99.9%
700 to 749	265	>99.9%	0.0%	0.0%	>99.9%
750 to 799	263	99.9%	0.1%	0.0%	99.9%
800 to 849	401	96.2%	3.8%	0.0%	96.2%
850 to 899	447	71.9%	28.1%	0.0%	71.9%
900 to 949 (Red/Green cut = 900)	514	28.0%	71.9%	0.0%	71.9%
950 to 999	436	3.7%	95.4%	0.9%	95.4%
1000 to 1049	267	0.2%	88.6%	11.2%	88.6%
1050 to 1099 (Green/Blue cut = 1073)	132	0.0%	52.9%	47.1%	61.9%
1100 to 1149	49	0.0%	12.9%	87.1%	87.1%
1150 to 1199	20	0.0%	1.2%	98.8%	98.8%
1200 to 1249	2	0.0%	0.0%	>99.9%	>99.9%
1250 to 1299	0	N/A	N/A	N/A	N/A
1300 to 1349	0	N/A	N/A	N/A	N/A
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	3,624				

* Retest assuming no additional instruction

Table E–63. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 6

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	8	>99.9%	0.0%	0.0%	>99.9%
450 to 499	36	>99.9%	0.0%	0.0%	>99.9%
500 to 549	137	>99.9%	0.0%	0.0%	>99.9%
550 to 599	242	>99.9%	0.0%	0.0%	>99.9%
600 to 649	277	>99.9%	0.0%	0.0%	>99.9%
650 to 699	270	>99.9%	0.0%	0.0%	>99.9%
700 to 749	279	>99.9%	0.0%	0.0%	>99.9%
750 to 799	328	>99.9%	0.0%	0.0%	>99.9%
800 to 849	446	99.7%	0.3%	0.0%	99.7%
850 to 899	526	93.5%	6.5%	0.0%	93.5%
900 to 949 (Red/Green cut = 938)	644	61.5%	38.5%	0.0%	64.8%
950 to 999	623	18.3%	81.7%	0.0%	81.7%
1000 to 1049	555	1.6%	96.9%	1.4%	96.9%
1050 to 1099	332	0.0%	82.6%	17.4%	82.6%
1100 to 1149 (Green/Blue cut = 1111)	173	0.0%	41.1%	58.9%	63.1%
1150 to 1199	74	0.0%	7.0%	93.0%	93.0%
1200 to 1249	27	0.0%	0.5%	99.5%	99.5%
1250 to 1299	6	0.0%	0.0%	>99.9%	>99.9%
1300 to 1349	2	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	0	N/A	N/A	N/A	N/A
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	4,985				

* Retest assuming no additional instruction

Table E–64. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 7

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	1	>99.9%	0.0%	0.0%	>99.9%
400 to 449	3	>99.9%	0.0%	0.0%	>99.9%
450 to 499	35	>99.9%	0.0%	0.0%	>99.9%
500 to 549	129	>99.9%	0.0%	0.0%	>99.9%
550 to 599	239	>99.9%	0.0%	0.0%	>99.9%
600 to 649	302	>99.9%	0.0%	0.0%	>99.9%
650 to 699	300	>99.9%	0.0%	0.0%	>99.9%
700 to 749	316	>99.9%	0.0%	0.0%	>99.9%
750 to 799	366	>99.9%	0.0%	0.0%	>99.9%
800 to 849	463	>99.9%	0.0%	0.0%	>99.9%
850 to 899	558	99.2%	0.8%	0.0%	99.2%
900 to 949	683	88.1%	11.9%	0.0%	88.1%
950 to 999 (Red/Green cut = 974)	821	47.3%	52.7%	0.0%	63.4%
1000 to 1049	842	10.3%	89.6%	0.1%	89.6%
1050 to 1099	625	0.7%	96.1%	3.2%	96.1%
1100 to 1149 (Green/Blue cut = 1147)	357	0.0%	74.5%	25.5%	74.6%
1150 to 1199	158	0.0%	27.3%	72.7%	72.7%
1200 to 1249	48	0.0%	3.9%	96.1%	96.1%
1250 to 1299	11	0.0%	0.2%	99.8%	99.8%
1300 to 1349	4	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	1	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	6,262				

* Retest assuming no additional instruction

Table E–65. Retest Classification Percent for Various Scale Score Ranges – Writing Grade 8

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	1	>99.9%	0.0%	0.0%	>99.9%
400 to 449	4	>99.9%	0.0%	0.0%	>99.9%
450 to 499	16	>99.9%	0.0%	0.0%	>99.9%
500 to 549	110	>99.9%	0.0%	0.0%	>99.9%
550 to 599	273	>99.9%	0.0%	0.0%	>99.9%
600 to 649	355	>99.9%	0.0%	0.0%	>99.9%
650 to 699	317	>99.9%	0.0%	0.0%	>99.9%
700 to 749	321	>99.9%	0.0%	0.0%	>99.9%
750 to 799	372	>99.9%	0.0%	0.0%	>99.9%
800 to 849	378	>99.9%	0.0%	0.0%	>99.9%
850 to 899	463	99.6%	0.4%	0.0%	99.6%
900 to 949	523	92.6%	7.4%	0.0%	92.6%
950 to 999 (Red/Green cut = 985)	661	58.9%	41.1%	0.0%	64.2%
1000 to 1049	743	15.9%	84.0%	0.0%	84.0%
1050 to 1099	592	1.4%	96.8%	1.8%	96.8%
1100 to 1149	300	0.0%	80.9%	19.1%	80.9%
1150 to 1199 (Green/Blue cut = 1158)	155	0.0%	38.2%	61.8%	64.0%
1200 to 1249	61	0.0%	6.1%	93.9%	93.9%
1250 to 1299	14	0.0%	0.4%	99.6%	99.6%
1300 to 1349	4	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	3	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	5,666				

* Retest assuming no additional instruction

Table E–66. Retest Classification Percent for Various Scale Score Ranges – English Composition

Scale Score Range	Number of Students	Red (% Chance in Category if Retested*)	Green (% Chance in Category if Retested*)	Blue (% Chance in Category if Retested*)	% Chance in Same Category if Retested*
< 200	0	N/A	N/A	N/A	N/A
200 to 249	0	N/A	N/A	N/A	N/A
250 to 299	0	N/A	N/A	N/A	N/A
300 to 349	0	N/A	N/A	N/A	N/A
350 to 399	0	N/A	N/A	N/A	N/A
400 to 449	0	N/A	N/A	N/A	N/A
450 to 499	13	>99.9%	0.0%	0.0%	>99.9%
500 to 549	63	>99.9%	0.0%	0.0%	>99.9%
550 to 599	131	>99.9%	0.0%	0.0%	>99.9%
600 to 649	185	>99.9%	0.0%	0.0%	>99.9%
650 to 699	210	>99.9%	0.0%	0.0%	>99.9%
700 to 749	197	>99.9%	0.0%	0.0%	>99.9%
750 to 799	208	>99.9%	0.0%	0.0%	>99.9%
800 to 849	252	>99.9%	0.0%	0.0%	>99.9%
850 to 899	316	99.8%	0.2%	0.0%	99.8%
900 to 949	417	95.3%	4.7%	0.0%	95.3%
950 to 999 (Red/Green cut = 994)	536	66.5%	33.5%	0.0%	67.5%
1000 to 1049	679	21.4%	78.5%	0.0%	78.5%
1050 to 1099	615	2.4%	96.6%	1.0%	96.6%
1100 to 1149	400	0.1%	87.0%	13.0%	87.0%
1150 to 1199 (Green/Blue cut = 1167)	235	0.0%	44.0%	56.0%	62.5%
1200 to 1249	104	0.0%	9.9%	90.1%	90.1%
1250 to 1299	44	0.0%	0.6%	99.4%	99.4%
1300 to 1349	18	0.0%	0.0%	>99.9%	>99.9%
1350 to 1399	12	0.0%	0.0%	>99.9%	>99.9%
1400 to 1449	0	N/A	N/A	N/A	N/A
1450 to 1499	0	N/A	N/A	N/A	N/A
1500 to 1549	0	N/A	N/A	N/A	N/A
1550 to 1599	0	N/A	N/A	N/A	N/A
1600 to 1649	0	N/A	N/A	N/A	N/A
1650 to 1699	0	N/A	N/A	N/A	N/A
1700 to 1749	0	N/A	N/A	N/A	N/A
1750 to 1799	0	N/A	N/A	N/A	N/A
1800 to 1849	0	N/A	N/A	N/A	N/A
1850 to 1899	0	N/A	N/A	N/A	N/A
1900 to 1949	0	N/A	N/A	N/A	N/A
1950 to 1999	0	N/A	N/A	N/A	N/A
>= 2000	0	N/A	N/A	N/A	N/A
TOTAL	4,635				

* Retest assuming no additional instruction

APPENDIX F: CDT LEARNING PROGRESSIONS

The CDT learning progressions were developed by the Pennsylvania Department of Education (PDE) and its curriculum consultants, including staff from Data Recognition Corporation (DRC), to show the developmental sequences or building blocks of content/skills students need to master as they progress toward career and college readiness. The progressions were developed for each content area (i.e., English language arts, mathematics, and science.) They served and continue to serve as roadmaps or the pathways (K-12) that students travel as they progress toward mastery of the skills needed for career and college readiness. As such, each learning progression was developed to provide teachers with the opportunity to determine whether students have navigated successfully through the building blocks and are able to move forward along the road to career and college readiness for a given content area. Each progression also provides teachers with the opportunity to identify students who may need additional instruction in a given content area, as well as to identify students who have navigated successfully beyond the building blocks or mileposts for each grade and/or course and are in need of accelerated curriculum. The learning progressions are directly aligned and based upon the Pennsylvania Academic Standards, the Assessment Anchors, and the Eligible Contents and as such provide evidence of the linkage between the CDT and the Pennsylvania PSSA and Keystone assessments addressing career and college readiness success with interpretations.

The learning progressions were first developed in 2009. Upon the initial development of the learning progression, the progressions were reviewed by Pennsylvania educators to confirm alignment to the Pennsylvania Standards and to confirm that the progressions, do, in fact, serve to show the development sequences of content/skills students need to master as they progress toward career and college readiness. At this meeting with educators, PDE and DRC provided information about the development of the learning progressions, the purpose of the progressions, and the actual progressions for each content area. The committees of Pennsylvania educators reviewed the progressions, which serve to show the vertical articulation of the Pennsylvania Academic Standards, Assessment Anchors and Eligible Content across grades within a given subject area (e.g., reading, mathematics). Pennsylvania educators were asked to confirm that the progressions were an accurate representation of how the content/skills included in the Pennsylvania Academic Standards progressed across grades and provided a broad description of the essential content and general sequencing for student learning and skill development as each student progresses toward college and career readiness.

Beginning 2010, the learning progressions have continued to be used during item reviews for the CDT, as well as for the PSSA and the Keystone assessments. For example, during each subsequent review of items for potential use on these assessments, including the CDT, Pennsylvania educators, in addition to reviewing items for alignment to the standards, cognitive complexity, technical quality, etc. also review items for alignment to the learning progressions. The learning progressions are included in this evidence to demonstrate the content/skills linkage between the CDT to address career and college readiness success.

APPENDIX G: DEVELOPMENT OF THE PENNSYLVANIA ACADEMIC STANDARDS, ASSESSMENT ANCHOR CONTENT STANDARDS, AND ELIGIBLE CONTENT

The Assessment Anchor Content Standards and Eligible Content statements are based on the Pennsylvania Academic Standards in English language arts and mathematics and the Pennsylvania Academic Standards in science. Although the Academic Standards indicated in broad terms what students should know and be able to do, educator concerns regarding the number and breadth of the Academic Standards led to an initiative by the Pennsylvania Department of Education (PDE) to develop Assessment Anchor Content Standards to indicate which parts of the Academic Standards (Instructional Content Standards) would be assessed on the summative assessments. Based on recommendations from Pennsylvania educators, the Assessment Anchor Content Standards were designed to improve the articulation of curricular, instructional, and assessment practices. The anchors clarify what is expected across each grade and content area and focus the content of the standards into what is assessable on a large-scale test. The Assessment Anchor Content Standards also serve to communicate Eligible Content or assessment limits. The Eligible Content statements also provide for the range of knowledge and skills from which the summative assessments and the CDT is designed.

The Assessment Anchor Content Standards' structure includes the content, grade level, Reporting Category, Assessment Anchor, descriptor (Sub-Assessment Anchor), and Eligible Content. Each of the Assessment Anchor Content Standards has one or more descriptors (Sub-Assessment Anchors) and Eligible Content to reflect grade-level appropriateness. The Assessment Anchor Content Standards form the basis of the test design. In turn, this hierarchy is the basis for organizing the total content scores (based on the core [common] sections). The Assessment Anchor Content Standards, therefore, are the general descriptions of what students should know and be able to do. The Eligible Content statements are the more specific statements of the knowledge and/or skills that students are expected to demonstrate in a given grade and content area. The Eligible Content statements are considered the granular level to which items are written. As such, they serve to define at a more granular level what students should know and be able to do. They also serve as the checkpoints that monitor progress toward meeting the board Pennsylvania Academic Standards. In other states' structures of content standards, the Assessment Anchor Content Standards are often labeled Benchmarks, and the Eligible Content statements are often labeled grade-level expectations.

The complete set of Assessment Anchor Content Standards and Eligible Content statements aligned to the board Pennsylvania Academic Standards can be found at the PDE's website: www.education.pa.gov.

DEVELOPMENT OF THE ASSESSMENT ANCHOR CONTENT STANDARDS AND THE ELIGIBLE CONTENT STATEMENTS

With Pennsylvania's decision to adopt the Pennsylvania Academic Standards in July 2010, committees of Pennsylvania educators then met in October 3–6, 2011 to write and review the Assessment Anchor Content Standards and Eligible Content statements aligned to the new Pennsylvania Academic Standards. Members of the committees included representatives from the PDE curriculum and instruction, the PDE assessment, Pennsylvania educators, and a team of expert consultants appointed by the PDE. The consultants were Pennsylvania known and nationally known experts representing specific areas of expertise. These appointed consultants were members of the Pennsylvania Quality Review Team, and their function was to oversee the process, ensuring quality throughout.

Prior to the beginning of the development of the Assessment Anchor Content Standards and the Eligible Content statements, the PDE-selected Quality Review Team consultants and the PDE assessment and curriculum staff analyzed pertinent national career- and college-ready standards and curriculum framework documents including frameworks from the National Assessment of Educational Progress (NAEP). Once the analysis was completed, members of the PDE-selected Quality Review Team met with the testing vendor, Data Recognition Corporation (DRC) to provide recommendations as to what materials and documents would be needed to facilitate the committees of Pennsylvania educators in the development and review of the Assessment Anchor Content Standards and the Eligible Content statements. In addition, the purpose of this meeting with the Quality Review Team was to come to agreement on the Assessment Anchor Content Standards and Eligible Content development process, including the role of the Pennsylvania educators, the PDE assessment staff, the PDE curriculum staff, the Quality Review Team members, and the testing vendor, DRC.

To provide initial focus at the October 2011 meetings, each content and grade committee of Pennsylvania educators was presented with materials specific to the content and grade to which the anchors and Eligible Content statements were to be developed, including a basic blueprint structure of the summative assessment and the CDT. The Pennsylvania Academic Standards, the 2005 version of the Pennsylvania Assessment Anchor Content Standards and Eligible Content aligned to the previous Pennsylvania Academic Standards, other career- and college-ready state standards, and draft Eligible Content statements aligned to the newly revised Pennsylvania Academic Standards were also provided. Committees then completed an iterative process of developing, reviewing, and revising the Assessment Anchor Content Standards and Eligible Content statements followed by discussions across grade-level committees to ensure vertical articulation across the grades. The results from the committee work were recorded and eventually evaluated by national, state, and local subject experts as noted in the sections below.

To begin the process, a general training session was held for all meeting participants. The training included welcome remarks, setting of the context for the task by the PDE staff and the PDE Quality Review Team member staff, and a presentation of the procedural training and meeting logistics by the testing vendor, DRC. Each meeting began with an introduction to Pennsylvania's Standards Aligned System and an overview of the assessment program. The PDE staff and the PDE Quality Review Team members articulated Pennsylvania's vision for the content standards, including the role that the Assessment Anchor Content Standards and Eligible Content statements would play in defining what students should know and be able to do. The opening presentation also included providing educators with the definition, structure, and purpose of the content standards, including definitions of Assessment Anchor Content Standards and Eligible Content statements. Training was also provided concerning writing, reviewing, and revising the Assessment Anchor Content Standards and Eligible Content statements. The focus of the training was to follow the design parameters to include clear, focused, rigorous, manageable, and subject-area statements.

The following materials were provided at the meeting:

- Pennsylvania Curriculum Framework: The Curriculum Framework specifies what is to be taught for each subject in the curriculum. In Pennsylvania, Curriculum Frameworks include Big Ideas, Concepts, Competencies, and Essential Questions aligned to standards. They are defined as follows:

  Big Ideas: The big ideas are the declarative statements that describe concepts that transcend grade levels. Big Ideas are essential to provide focus on specific content for all students.

  Concepts: The concepts are what students should know (key knowledge) as a result of this instruction specific to grade level.

  Competencies: The competences are what students should be able to do (key skills) as a result of this instruction, specific to grade level.

  Essential Questions: The essential questions are connected to the Standards Aligned System (SAS) framework and are specifically linked to the big ideas. They frame student inquiry, promote critical thinking, and assist in learning transfer.

- Pennsylvania Academic Standards

- Other documents as relevant, including hard copy working documents with adequate white space

After the training, committee members were instructed to begin the development process. Committee members were provided with hard copy working documents. Using their background knowledge and the materials they were provided during the meeting (e.g., documents from the Standards Aligned System, curriculum framework, Pennsylvania's Academic Standards), Pennsylvania educators created their own short list of the critical concepts that Pennsylvania students must know and be able to do for each grade and content area. Beginning with one concept at a time, concepts or Eligible Content statements were recorded on the master list; Assessment Anchor Content Standards were then developed and reviewed. As the Assessment Anchor Content Standards and Eligible Content statements were developed, they were displayed using a laptop and projector. A scribe from the testing vendor, DRC, served to record the committee members' work as well as other comments. The scribe also recorded changes or additions to the anchors and/or statements as directed from the consensus of the group.

Next, the entire group reviewed and discussed the recommendations for the anchors and the Eligible Content statements. Consensus was reached. The committee of Pennsylvania educators proceeded in this manner until all Assessment Anchor Content Standards and Eligible Content statements for each grade and content area were developed, reviewed, and discussed. DRC's facilitator took notes verbatim regarding the intent and direction of the committee. The notes were prepared for use in subsequent meetings.

FOLLOW-UP MEETINGS WITH THE QUALITY REVIEW TEAM AND PDE

A series of follow-up meetings took place with the PDE-appointed team of consultants, PDE assessment staff, and PDE-appointed Quality Review Team members. Prior to the follow-up meetings, a draft of the Assessment Anchor Content Standards and Eligible Content statements for each grade and content area were prepared for review, including all notes from the meeting with Pennsylvania educators. During the follow-up meetings, the Assessment Anchor Content Standards and Eligible Content statements were reviewed, and revisions were suggested. After the follow-up meetings, the Assessment Anchor Content Standards and Eligible Content statements were revised by the PDE and the PDE Quality Review Team per agreed-upon feedback. This revised draft was then posted on the Pennsylvania System of Aligned Standards (SAS) website for public review and opinion. All additional feedback from the public review was reviewed again by the PDE and the PDE-appointed Quality Review Team and agreed upon revisions to the Assessment Anchor Content Standards and Eligible Content statements were made. The Assessment Anchor Content Standards and Eligible Content statements were then finalized and prepared for the Pennsylvania Board of Education for approval as the official Pennsylvania Academic Content Standards.

PENNSYLVANIA BOARD OF EDUCATION APPROVAL

The Assessment Anchor Content Standards and Eligible Content statements were presented to the State Board of Education in September 2013. They were subsequently approved by the State Board at the September 2013 State Board meeting as Pennsylvania Content Standards.

APPENDIX H: CDT PASSAGE DEVELOPMENT PROCESS

The task of writing passages or securing passages and or other stimuli for the CDT is conducted by Data Recognition Corporation (DRC) professionals with classroom experience in reading/language arts as well as experience writing the various types of passages and/or stimuli required by the CDT and the Pennsylvania Academic Standards, Assessment Anchors, and Eligible Content. Guidelines provided to writers for passage/stimulus writing for the CDT include appropriate length, text structure, density, and vocabulary for the grade level as reviewed and approved by the Pennsylvania Department of Education (PDE) and as aligned to the Pennsylvania Academic Standards, Assessment Anchors, and Eligible Content. Passage/stimulus writers are given a specified number of passages/stimuli to write for each genre/standard per grade. Passage/stimulus training includes training writers to develop passages/stimuli to meet the following requirements:

- Grade appropriateness
- Appropriate readability for the assigned grade
- Interest value for students
- Freedom from bias, fairness, and sensitivity issues
- Representation of different cultures
- Ability to generate a variety of item types
- Avoidance of dated subject matter, unless a relevant historical context is provided
- No need for extensive background knowledge in a certain discipline or subject area

While DRC does train passage writers to be knowledgeable of each passage's readability, for the CDT we also statistically analyze readability of each passage, using Lexile, Flesch-Kincaid, Powers, and Spache measurements. The process that DRC's item and test development team uses to determine text complexity involves (1) the quantitative evaluation of the text, and (2) the qualitative evaluation of the text. This analysis is documented on a passage placemat. (See example passage placemat at the end of this section.) A third component, matching reader to text and task, is also taken into consideration during passage evaluation and internal reviews.

QUANTITATIVE EVALUATION

Evaluating the complexity of a passage is a judgment process conducted by DRC passage writers and internal reviewers who are familiar with the classroom context and what is developmentally and linguistically appropriate for students at a given grade level. DRC uses common readability formulas along with the qualitative information when selecting passages during development.

QUALITATIVE EVALUATION

For programs such as the CDT, DRC also implements qualitative measures to help determine placement and appropriateness of passages. These measures include rubric-based qualitative evaluations and external reviewers to provide expert opinions on grade-level appropriateness, as part of considerations for matching the reader to text and task. Rubrics provide the qualitative measures for literary and informational passages. As indicated on the placemats, the quantitative rubrics suggest the appropriate grade band of the passage, while the qualitative rubrics help to further clarify the specific grade level of the passage. These rubrics provide a powerful and comprehensive way of evaluating a range of stimulus materials that cover the literary and informational scope outlined in the client state's standards.

TEXT COMPLEXITY: QUALITATIVE-MEASURES RUBRIC—LITERARY TEXTS

The English Language Arts State Collaborative on Assessment and Student Standards (SCASS) developed the following qualitative-measures rubric for determining the text complexity of literary passages. The rubric examines criteria judged as central to students' successful comprehension of text meaning, text structure, language features, and knowledge demands. Each of these categories is ranked based on descriptors associated with the following levels: slightly complex, moderately complex, very complex, and exceedingly complex.

Qualitative-Measures Rubric—Literary Passages

Features	Exceedingly Complex	Very Complex	Moderately Complex	Slightly Complex
Meaning	Several levels and competing elements of meaning that are difficult to identify, separate, and interpret; theme is implicit or subtle, often ambiguous and revealed over the entirety of the text	Several levels of meaning that may be difficult to identify or separate; theme is implicit or subtle and may be revealed over the entirety of the text	More than one level of meaning with levels clearly distinguished from each other; theme is clear but may be conveyed with some subtlety	One level of meaning; theme is obvious and revealed early in the text
Organization	Organization is intricate with regard to elements such as narrative viewpoint, time shifts, multiple characters, storylines, and detail	Organization may include subplots, time shifts, and more complex characters	Organization may have two or more storylines and is occasionally difficult to predict	Organization of text is clear, chronological, or easy to predict
Use of images	If used, minimal illustrations that support the text	If used, a few illustrations that support the text	If used, a range of illustrations that support selected parts of the text	If used, extensive illustrations that directly support and assist in interpreting the written text
Language Features	**Conventionality** Dense and complex; contains abstract, ironic, and/or figurative language	**Conventionality** Complex; contains some abstract, ironic, and/or figurative language	**Conventionality** Largely explicit and easy to understand, with some occasions for more complex meaning	**Conventionality** Explicit, literal, straightforward, easy to understand
	Vocabulary Generally unfamiliar, archaic, subject- specific, or overly academic language; may be ambiguous or purposefully misleading	**Vocabulary** Somewhat complex language that is sometimes unfamiliar, archaic, subject- specific, or overly academic	**Vocabulary** Mostly contemporary, familiar, conversational; rarely unfamiliar or overly academic	**Vocabulary** Contemporary, familiar, conversational language
	Sentence Structure Mainly complex sentences, often containing multiple concepts	**Sentence Structure** Many complex sentences with several subordinate phrases or clauses and transition words	**Sentence Structure** Simple and compound sentences, with some more complex constructions	**Sentence Structure** Mainly simple sentences
Knowledge Demands	**Life Experiences** Explores complex, sophisticated themes; experiences are distinctly different from those of the common reader	**Life Experiences** Explores themes of varying levels of complexity; experiences portrayed are uncommon to most readers	**Life Experiences** Explores a single theme; experiences portrayed are common to many readers	**Life Experiences** Explores a single theme; experiences portrayed are everyday and common to most readers
	Intertextuality and Cultural Knowledge Many references or allusions to other texts or cultural elements	**Intertextuality and Cultural Knowledge** Some references or allusions to other texts or cultural elements	**Intertextuality and Cultural Knowledge** A few references or allusions to other texts or cultural elements	**Intertextuality and Cultural Knowledge** No references or allusions to other texts or cultural elements

Qualitative-Measures Rubric—Informational Texts

Features	Exceedingly Complex	Very Complex	Moderately Complex	Slightly Complex
Purpose	**Purpose** Subtle, implied, difficult to determine; intricate, theoretical elements	**Purpose** Implied but fairly easy to infer; more theoretical than concrete	**Purpose** Implied but easy to identify based upon context or source	**Purpose** Explicitly stated; clear, concrete with a narrow focus
Text Structure	**Organization of Main Ideas** Connections between an extensive range of ideas, processes, or events are deep and often implicit or subtle; organization of the text is intricate or specialized for a particular discipline	**Organization of Main Ideas** Connections between an expanded range of ideas, processes, or events are deeper and often implicit or subtle; organization may contain multiple pathways and may exhibit traits common to a specific discipline	**Organization of Main Ideas** Connections between some ideas or events are implicit or subtle; organization is evident and generally sequential	**Organization of Main Ideas** Connections between ideas, processes, or events are explicit and clear; organization of text is clear or chronological or easy to predict
	Text Features If used, are essential in understanding content	**Text Features** If used, greatly enhance the reader's understanding of content	**Text Features** If used, enhance the reader's understanding of content	**Text Features** If used, help the reader navigate and understand content but are not essential
	Use of Images If used, extensive, intricate, essential integrated images, tables, charts, etc., necessary to make meaning of text; also may provide information not otherwise conveyed in the text	**Use of Images** If used, essential integrated images, tables, charts, etc., may occasionally be essential to understanding the text	**Use of Images** If used, images such as indexes and glossaries are mostly supplementary to understanding of the text; graphs, pictures, tables, and charts directly support the text	**Use of Images** If used, images are simple and unnecessary to understanding the text but directly support and assist in interpreting the written text
Language Features	**Conventionality** Dense and complex; contains abstract, ironic, and/or figurative language	**Conventionality** Complex; contains some abstract, ironic, and/or figurative language	**Conventionality** Largely explicit and easy to understand, with some occasions for more complex meaning	**Conventionality** Explicit, literal, straightforward, easy to understand
	Vocabulary Generally unfamiliar, archaic, subject- specific, or overly academic language; may be ambiguous or purposefully misleading	**Vocabulary** Somewhat complex language that is sometimes unfamiliar, archaic, subject-specific, or overly academic	**Vocabulary** Mostly contemporary, familiar, conversational; rarely unfamiliar or overly academic	**Vocabulary** Contemporary, familiar, conversational language
	Sentence Structure Mainly complex sentences, often containing multiple concepts	**Sentence Structure** Many complex sentences with several subordinate phrases or clauses and transition words	**Sentence Structure** Simple and compound sentences, with some more complex constructions	**Sentence Structure** Mainly simple sentences
Knowledge Demands	**Subject Matter Knowledge** Extensive, perhaps specialized or even theoretical discipline-specific content knowledge; range of challenging abstract and theoretical concepts	**Subject Matter Knowledge** Moderate levels of discipline-specific content knowledge; some theoretical knowledge may enhance understanding; range of recognizable ideas and challenging abstract concepts	**Subject Matter Knowledge** Everyday practical knowledge and some discipline-specific content knowledge; both simple and more complicated, abstract ideas	**Subject Matter Knowledge** Everyday, practical knowledge; simple, concrete ideas
	Intertextuality Many references or allusions to other texts or outside ideas, theories, etc.	**Intertextuality** Some references or allusions to other texts or outside ideas, theories, etc.	**Intertextuality** A few references or allusions to other texts or outside ideas, theories, etc.	**Intertextuality** No references or allusions to other texts, or outside ideas, theories, etc.

Adapted from Smarter Balanced and © 2012 by the ELA SCASS

Passage Placemat

Below is an example of a passage placemat for item writer use.

Worksheet: Text Complexity Analysis		
Title	**Author**	**Text Description**

Recommended Placement for Assessment: Grade X

Qualitative Measures	Quantitative Measures
PURPOSE : **TEXT STRUCTURE** **Organization of Main Ideas:** : **Text Features:** N/A **Use of Images:** N/A **LANGUAGE FEATURES** **Conventionality:** : **Vocabulary:** : **Sentence Structure:** : **KNOWLEDGE DEMANDS** **Subject Matter Knowledge:** : **Intertextuality:** :	**Common Core State Standards Appendix A Complexity Band Level** (if applicable): **Lexile or Other Quantitative Measure of the Text:** Lexile: Flesch-Kincaid: **Considerations for Passage Selection** Passage selection should be based on the ELA Content Specifications targets and the cognitive demands of the assessment tasks. **Potential Challenges This Text May Pose (check all that apply):** ☐ Accessibility ☐ Sentence and text structures ☐ Archaic language, slang, idioms, or other language challenges ☐ Background knowledge ☐ Bias and sensitivity issues ☐ Word count

Adapted from Smarter Balanced and the 2012 ELA SCASS work

APPENDIX I: DEPTH OF KNOWLEDGE

Diagnostic Categories-Mathematics

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
3	M03.A-F.1.1.1	Numbers and Operations	15
3	M03.A-F.1.1.2	Numbers and Operations	6
3	M03.A-F.1.1.3	Numbers and Operations	8
3	M03.A-F.1.1.4	Numbers and Operations	6
3	M03.A-F.1.1.5	Numbers and Operations	6
3	M03.A-T.1.1.1	Numbers and Operations	8
3	M03.A-T.1.1.2	Numbers and Operations	10
3	M03.A-T.1.1.3	Numbers and Operations	6
3	M03.A-T.1.1.4	Numbers and Operations	10
3	M03.B-O.1.1.1	Algebraic Concepts	7
3	M03.B-O.1.1.2	Algebraic Concepts	5
3	M03.B-O.1.2.1	Algebraic Concepts	10
3	M03.B-O.1.2.2	Algebraic Concepts	9
3	M03.B-O.2.1.1	Algebraic Concepts	8
3	M03.B-O.2.1.2	Algebraic Concepts	4
3	M03.B-O.2.2.1	Algebraic Concepts	6
3	M03.B-O.3.1.1	Algebraic Concepts	8
3	M03.B-O.3.1.2	Algebraic Concepts	6
3	M03.B-O.3.1.3	Algebraic Concepts	5
3	M03.B-O.3.1.4	Algebraic Concepts	5
3	M03.B-O.3.1.5	Algebraic Concepts	4
3	M03.B-O.3.1.6	Algebraic Concepts	15
3	M03.B-O.3.1.7	Algebraic Concepts	8
3	M03.C-G.1.1.1	Geometry	12
3	M03.C-G.1.1.2	Geometry	12
3	M03.C-G.1.1.3	Geometry	11
3	M03.D-M.1.1.1	Measurement Data and Probability	9
3	M03.D-M.1.1.2	Measurement Data and Probability	17
3	M03.D-M.1.2.1	Measurement Data and Probability	6
3	M03.D-M.1.2.2	Measurement Data and Probability	5
3	M03.D-M.1.2.3	Measurement Data and Probability	17
3	M03.D-M.1.3.1	Measurement Data and Probability	9
3	M03.D-M.1.3.2	Measurement Data and Probability	7
3	M03.D-M.1.3.3	Measurement Data and Probability	7
3	M03.D-M.2.1.1	Measurement Data and Probability	5
3	M03.D-M.2.1.2	Measurement Data and Probability	15

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
3	M03.D-M.2.1.3	Measurement Data and Probability	8
3	M03.D-M.2.1.4	Measurement Data and Probability	9
3	M03.D-M.3.1.1	Measurement Data and Probability	5
3	M03.D-M.3.1.2	Measurement Data and Probability	3
3	M03.D-M.4.1.1	Measurement Data and Probability	14
4	M04.A-F.1.1.1	Numbers and Operations	6
4	M04.A-F.1.1.2	Numbers and Operations	7
4	M04.A-F.2.1.1	Numbers and Operations	6
4	M04.A-F.2.1.2	Numbers and Operations	13
4	M04.A-F.2.1.3	Numbers and Operations	6
4	M04.A-F.2.1.4	Numbers and Operations	8
4	M04.A-F.2.1.5	Numbers and Operations	5
4	M04.A-F.2.1.6	Numbers and Operations	11
4	M04.A-F.2.1.7	Numbers and Operations	5
4	M04.A-F.3.1.1	Numbers and Operations	1
4	M04.A-F.3.1.2	Numbers and Operations	6
4	M04.A-F.3.1.3	Numbers and Operations	5
4	M04.A-T.1.1.1	Numbers and Operations	4
4	M04.A-T.1.1.2	Numbers and Operations	9
4	M04.A-T.1.1.3	Numbers and Operations	4
4	M04.A-T.1.1.4	Numbers and Operations	11
4	M04.A-T.2.1.1	Numbers and Operations	7
4	M04.A-T.2.1.2	Numbers and Operations	7
4	M04.A-T.2.1.3	Numbers and Operations	6
4	M04.A-T.2.1.4	Numbers and Operations	8
4	M04.B-O.1.1.1	Algebraic Concepts	5
4	M04.B-O.1.1.2	Algebraic Concepts	5
4	M04.B-O.1.1.3	Algebraic Concepts	11
4	M04.B-O.1.1.4	Algebraic Concepts	9
4	M04.B-O.2.1.1	Algebraic Concepts	17
4	M04.B-O.3.1.1	Algebraic Concepts	17
4	M04.B-O.3.1.2	Algebraic Concepts	36
4	M04.B-O.3.1.3	Algebraic Concepts	30
4	M04.C-G.1.1.1	Geometry	38
4	M04.C-G.1.1.2	Geometry	15
4	M04.C-G.1.1.3	Geometry	21
4	M04.D-M.1.1.1	Measurement Data and Probability	3
4	M04.D-M.1.1.2	Measurement Data and Probability	5

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
4	M04.D-M.1.1.3	Measurement Data and Probability	8
4	M04.D-M.1.1.4	Measurement Data and Probability	13
4	M04.D-M.2.1.1	Measurement Data and Probability	14
4	M04.D-M.2.1.2	Measurement Data and Probability	12
4	M04.D-M.2.1.3	Measurement Data and Probability	2
4	M04.D-M.3.1.1	Measurement Data and Probability	14
4	M04.D-M.3.1.2	Measurement Data and Probability	1
5	M05.A-F.1.1.1	Numbers and Operations	10
5	M05.A-F.2.1.1	Numbers and Operations	7
5	M05.A-F.2.1.2	Numbers and Operations	8
5	M05.A-F.2.1.3	Numbers and Operations	7
5	M05.A-F.2.1.4	Numbers and Operations	9
5	M05.A-T.1.1.1	Numbers and Operations	11
5	M05.A-T.1.1.2	Numbers and Operations	8
5	M05.A-T.1.1.3	Numbers and Operations	11
5	M05.A-T.1.1.4	Numbers and Operations	15
5	M05.A-T.1.1.5	Numbers and Operations	7
5	M05.A-T.2.1.1	Numbers and Operations	7
5	M05.A-T.2.1.2	Numbers and Operations	7
5	M05.A-T.2.1.3	Numbers and Operations	9
5	M05.B-O.1.1.1	Algebraic Concepts	10
5	M05.B-O.1.1.2	Algebraic Concepts	10
5	M05.B-O.2.1.1	Algebraic Concepts	10
5	M05.B-O.2.1.2	Algebraic Concepts	8
5	M05.C-G.1.1.1	Geometry	21
5	M05.C-G.1.1.2	Geometry	13
5	M05.C-G.2.1.1	Geometry	20
5	M05.D-M.1.1.1	Measurement Data and Probability	22
5	M05.D-M.2.1.1	Measurement Data and Probability	8
5	M05.D-M.2.1.2	Measurement Data and Probability	23
5	M05.D-M.3.1.1	Measurement Data and Probability	11
5	M05.D-M.3.1.2	Measurement Data and Probability	9
6	M06.A-N.1.1.1	Numbers and Operations	7
6	M06.A-N.2.1.1	Numbers and Operations	16
6	M06.A-N.2.2.1	Numbers and Operations	19
6	M06.A-N.2.2.2	Numbers and Operations	16
6	M06.A-N.3.1.1	Numbers and Operations	7
6	M06.A-N.3.1.2	Numbers and Operations	8

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
6	M06.A-N.3.1.3	Numbers and Operations	34
6	M06.A-N.3.2.1	Numbers and Operations	9
6	M06.A-N.3.2.2	Numbers and Operations	8
6	M06.A-N.3.2.3	Numbers and Operations	7
6	M06.A-R.1.1.1	Numbers and Operations	9
6	M06.A-R.1.1.2	Numbers and Operations	11
6	M06.A-R.1.1.3	Numbers and Operations	7
6	M06.A-R.1.1.4	Numbers and Operations	11
6	M06.A-R.1.1.5	Numbers and Operations	19
6	M06.B-E.1.1.1	Algebraic Concepts	16
6	M06.B-E.1.1.2	Algebraic Concepts	15
6	M06.B-E.1.1.3	Algebraic Concepts	14
6	M06.B-E.1.1.4	Algebraic Concepts	15
6	M06.B-E.1.1.5	Algebraic Concepts	15
6	M06.B-E.2.1.1	Algebraic Concepts	26
6	M06.B-E.2.1.2	Algebraic Concepts	22
6	M06.B-E.2.1.3	Algebraic Concepts	31
6	M06.B-E.2.1.4	Algebraic Concepts	15
6	M06.B-E.3.1.1	Algebraic Concepts	15
6	M06.B-E.3.1.2	Algebraic Concepts	26
6	M06.C-G.1.1.1	Geometry	11
6	M06.C-G.1.1.2	Geometry	13
6	M06.C-G.1.1.3	Geometry	11
6	M06.C-G.1.1.4	Geometry	10
6	M06.C-G.1.1.5	Geometry	12
6	M06.C-G.1.1.6	Geometry	10
6	M06.D-S.1.1.1	Measurement Data and Probability	19
6	M06.D-S.1.1.2	Measurement Data and Probability	33
6	M06.D-S.1.1.3	Measurement Data and Probability	42
6	M06.D-S.1.1.4	Measurement Data and Probability	19
7	M07.A-N.1.1.1	Numbers and Operations	9
7	M07.A-N.1.1.2	Numbers and Operations	9
7	M07.A-N.1.1.3	Numbers and Operations	10
7	M07.A-R.1.1.1	Numbers and Operations	9
7	M07.A-R.1.1.2	Numbers and Operations	10
7	M07.A-R.1.1.3	Numbers and Operations	13
7	M07.A-R.1.1.4	Numbers and Operations	10
7	M07.A-R.1.1.5	Numbers and Operations	9

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
7	M07.A-R.1.1.6	Numbers and Operations	20
7	M07.B-E.1.1.1	Algebraic Concepts	13
7	M07.B-E.2.1.1	Algebraic Concepts	29
7	M07.B-E.2.2.1	Algebraic Concepts	11
7	M07.B-E.2.2.2	Algebraic Concepts	9
7	M07.B-E.2.3.1	Algebraic Concepts	9
7	M07.C-G.1.1.1	Geometry	23
7	M07.C-G.1.1.2	Geometry	14
7	M07.C-G.1.1.3	Geometry	12
7	M07.C-G.1.1.4	Geometry	13
7	M07.C-G.2.1.1	Geometry	13
7	M07.C-G.2.1.2	Geometry	12
7	M07.C-G.2.2.1	Geometry	13
7	M07.C-G.2.2.2	Geometry	12
7	M07.D-S.1.1.1	Measurement Data and Probability	18
7	M07.D-S.1.1.2	Measurement Data and Probability	19
7	M07.D-S.2.1.1	Measurement Data and Probability	18
7	M07.D-S.3.1.1	Measurement Data and Probability	22
7	M07.D-S.3.2.1	Measurement Data and Probability	18
7	M07.D-S.3.2.2	Measurement Data and Probability	30
7	M07.D-S.3.2.3	Measurement Data and Probability	33
8	M08.A-N.1.1.1	Numbers and Operations	10
8	M08.A-N.1.1.2	Numbers and Operations	15
8	M08.A-N.1.1.3	Numbers and Operations	12
8	M08.A-N.1.1.4	Numbers and Operations	7
8	M08.A-N.1.1.5	Numbers and Operations	10
8	M08.B-E.1.1.1	Algebraic Concepts	9
8	M08.B-E.1.1.2	Algebraic Concepts	10
8	M08.B-E.1.1.3	Algebraic Concepts	7
8	M08.B-E.1.1.4	Algebraic Concepts	20
8	M08.B-E.2.1.1	Algebraic Concepts	9
8	M08.B-E.2.1.2	Algebraic Concepts	10
8	M08.B-E.2.1.3	Algebraic Concepts	12
8	M08.B-E.3.1.1	Algebraic Concepts	9
8	M08.B-E.3.1.2	Algebraic Concepts	10
8	M08.B-E.3.1.3	Algebraic Concepts	9
8	M08.B-E.3.1.4	Algebraic Concepts	10
8	M08.B-E.3.1.5	Algebraic Concepts	9

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
8	M08.B-F.1.1.1	Algebraic Concepts	9
8	M08.B-F.1.1.2	Algebraic Concepts	10
8	M08.B-F.1.1.3	Algebraic Concepts	10
8	M08.B-F.2.1.1	Algebraic Concepts	10
8	M08.B-F.2.1.2	Algebraic Concepts	17
8	M08.C-G.1.1.1	Geometry	25
8	M08.C-G.1.1.2	Geometry	7
8	M08.C-G.1.1.3	Geometry	10
8	M08.C-G.1.1.4	Geometry	11
8	M08.C-G.2.1.1	Geometry	12
8	M08.C-G.2.1.2	Geometry	19
8	M08.C-G.2.1.3	Geometry	12
8	M08.C-G.3.1.1	Geometry	13
8	M08.D-S.1.1.1	Measurement Data and Probability	14
8	M08.D-S.1.1.2	Measurement Data and Probability	15
8	M08.D-S.1.1.3	Measurement Data and Probability	12
8	M08.D-S.1.2.1	Measurement Data and Probability	13
A1	A1.1.1.1.1	Numbers and Operations	11
A1	A1.1.1.1.2	Numbers and Operations	11
A1	A1.1.1.2.1	Numbers and Operations	11
A1	A1.1.1.3.1	Algebraic Concepts	11
A1	A1.1.1.4.1	Numbers and Operations	11
A1	A1.1.1.5.1	Algebraic Concepts	12
A1	A1.1.1.5.2	Algebraic Concepts	10
A1	A1.1.1.5.3	Algebraic Concepts	11
A1	A1.1.2.1.1	Algebraic Concepts	13
A1	A1.1.2.1.2	Algebraic Concepts	12
A1	A1.1.2.1.3	Algebraic Concepts	12
A1	A1.1.2.2.1	Algebraic Concepts	12
A1	A1.1.2.2.2	Algebraic Concepts	12
A1	A1.1.3.1.1	Algebraic Concepts	13
A1	A1.1.3.1.2	Algebraic Concepts	11
A1	A1.1.3.1.3	Algebraic Concepts	12
A1	A1.1.3.2.1	Algebraic Concepts	12
A1	A1.1.3.2.2	Algebraic Concepts	13
A1	A1.2.1.1.1	Algebraic Concepts	14
A1	A1.2.1.1.2	Algebraic Concepts	14
A1	A1.2.1.1.3	Algebraic Concepts	16

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
A1	A1.2.1.2.1	Algebraic Concepts	14
A1	A1.2.1.2.2	Algebraic Concepts	14
A1	A1.2.2.1.1	Algebraic Concepts	12
A1	A1.2.2.1.2	Algebraic Concepts	13
A1	A1.2.2.1.3	Algebraic Concepts	14
A1	A1.2.2.1.4	Algebraic Concepts	14
A1	A1.2.2.2.1	Measurement Data and Probability	14
A1	A1.2.3.1.1	Measurement Data and Probability	11
A1	A1.2.3.2.1	Measurement Data and Probability	10
A1	A1.2.3.2.2	Measurement Data and Probability	11
A1	A1.2.3.2.3	Measurement Data and Probability	10
A1	A1.2.3.3.1	Measurement Data and Probability	11
A2	A2.1.1.1.1	Numbers and Operations	26
A2	A2.1.1.1.2	Numbers and Operations	24
A2	A2.1.1.2.1	Algebraic Concepts	26
A2	A2.1.1.2.2	Algebraic Concepts	24
A2	A2.1.2.1.1	Algebraic Concepts	6
A2	A2.1.2.1.2	Algebraic Concepts	6
A2	A2.1.2.1.3	Algebraic Concepts	6
A2	A2.1.2.1.4	Algebraic Concepts	6
A2	A2.1.2.2.1	Algebraic Concepts	6
A2	A2.1.2.2.2	Algebraic Concepts	6
A2	A2.1.3.1.1	Algebraic Concepts	7
A2	A2.1.3.1.2	Algebraic Concepts	7
A2	A2.1.3.1.3	Algebraic Concepts	7
A2	A2.1.3.1.4	Algebraic Concepts	7
A2	A2.1.3.2.1	Algebraic Concepts	9
A2	A2.1.3.2.2	Algebraic Concepts	7
A2	A2.2.1.1.1	Algebraic Concepts	13
A2	A2.2.1.1.2	Algebraic Concepts	13
A2	A2.2.1.1.3	Algebraic Concepts	13
A2	A2.2.1.1.4	Algebraic Concepts	13
A2	A2.2.2.1.1	Algebraic Concepts	13
A2	A2.2.2.1.2	Algebraic Concepts	12
A2	A2.2.2.1.3	Algebraic Concepts	13
A2	A2.2.2.1.4	Algebraic Concepts	13
A2	A2.2.2.2.1	Algebraic Concepts	13
A2	A2.2.3.1.1	Measurement Data and Probability	8

Diagnostic Categories-Mathematics (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
A2	A2.2.3.1.2	Measurement Data and Probability	6
A2	A2.2.3.2.1	Measurement Data and Probability	11
A2	A2.2.3.2.2	Measurement Data and Probability	10
A2	A2.2.3.2.3	Measurement Data and Probability	12
G	G.1.1.1.1	Geometry	22
G	G.1.1.1.2	Geometry	14
G	G.1.1.1.3	Geometry	10
G	G.1.1.1.4	Geometry	10
G	G.1.2.1.1	Geometry	15
G	G.1.2.1.2	Geometry	15
G	G.1.2.1.3	Geometry	9
G	G.1.2.1.4	Geometry	10
G	G.1.2.1.5	Geometry	15
G	G.1.3.1.1	Geometry	23
G	G.1.3.1.2	Geometry	24
G	G.1.3.2.1	Geometry	24
G	G.2.1.1.1	Geometry	9
G	G.2.1.1.2	Geometry	9
G	G.2.1.2.1	Geometry	9
G	G.2.1.2.2	Geometry	14
G	G.2.1.2.3	Geometry	9
G	G.2.2.1.1	Geometry	6
G	G.2.2.1.2	Geometry	6
G	G.2.2.2.1	Geometry	6
G	G.2.2.2.2	Geometry	6
G	G.2.2.2.3	Geometry	6
G	G.2.2.2.4	Geometry	6
G	G.2.2.2.5	Geometry	6
G	G.2.2.3.1	Geometry	6
G	G.2.2.4.1	Measurement Data and Probability	6
G	G.2.3.1.1	Geometry	9
G	G.2.3.1.2	Geometry	6
G	G.2.3.1.3	Geometry	6
G	G.2.3.2.1	Geometry	10

Diagnostic Categories-Reading

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
K	CC.1.2.K.A	Key Ideas and Details-Informational Text	6
K	CC.1.2.K.B	Key Ideas and Details-Informational Text	7
K	CC.1.2.K.C	Key Ideas and Details-Informational Text	8
K	CC.1.2.K.E	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6
K	CC.1.2.K.G	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	5
K	CC.1.2.K.H	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6
K	CC.1.2.K.I	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	4
K	CC.1.2.K.K	Vocabulary Acquisition and Use	6
K	CC.1.3.K.A	Key Ideas and Details-Literature Text	6
K	CC.1.3.K.B	Key Ideas and Details-Literature Text	6
K	CC.1.3.K.C	Key Ideas and Details-Literature Text	6
K	CC.1.3.K.D	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	6
K	CC.1.3.K.H	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	6
K	CC.1.3.K.I	Vocabulary Acquisition and Use	6
1	CC.1.2.1.A	Key Ideas and Details-Informational Text	7
1	CC.1.2.1.B	Key Ideas and Details-Informational Text	7
1	CC.1.2.1.C	Key Ideas and Details-Informational Text	7
1	CC.1.2.1.E	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
1	CC.1.2.1.G	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
1	CC.1.2.1.H	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
1	CC.1.2.1.I	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6
1	CC.1.2.1.K	Vocabulary Acquisition and Use	7
1	CC.1.3.1.A	Key Ideas and Details-Literature Text	6
1	CC.1.3.1.B	Key Ideas and Details-Literature Text	7
1	CC.1.3.1.C	Key Ideas and Details-Literature Text	7
1	CC.1.3.1.D	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
1	CC.1.3.1.H	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
1	CC.1.3.1.I	Vocabulary Acquisition and Use	6
2	CC.1.2.2.A	Key Ideas and Details-Informational Text	7
2	CC.1.2.2.B	Key Ideas and Details-Informational Text	7

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
2	CC.1.2.2.C	Key Ideas and Details-Informational Text	7
2	CC.1.2.2.E	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
2	CC.1.2.2.G	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
2	CC.1.2.2.H	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
2	CC.1.2.2.I	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
2	CC.1.2.2.K	Vocabulary Acquisition and Use	7
2	CC.1.3.2.A	Key Ideas and Details-Literature Text	7
2	CC.1.3.2.B	Key Ideas and Details-Literature Text	7
2	CC.1.3.2.C	Key Ideas and Details-Literature Text	7
2	CC.1.3.2.D	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
2	CC.1.3.2.H	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
2	CC.1.3.2.I	Vocabulary Acquisition and Use	7
3	E03.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	15
3	E03.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	14
3	E03.A-K.1.1.1	Key Ideas and Details-Literature Text	14
3	E03.A-K.1.1.2	Key Ideas and Details-Literature Text	14
3	E03.A-K.1.1.3	Key Ideas and Details-Literature Text	13
3	E03.A-V.4.1.1	Vocabulary Acquisition and Use	24
3	E03.A-V.4.1.2	Vocabulary Acquisition and Use	22
3	E03.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	13
3	E03.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	14
3	E03.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	13
3	E03.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	10
3	E03.B-C.3.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	12
3	E03.B-K.1.1.1	Key Ideas and Details-Informational Text	15
3	E03.B-K.1.1.2	Key Ideas and Details-Informational Text	14
3	E03.B-K.1.1.3	Key Ideas and Details-Informational Text	14
3	E03.B-V.4.1.1	Vocabulary Acquisition and Use	29
3	E03.B-V.4.1.2	Vocabulary Acquisition and Use	22

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
4	E04.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	13
4	E04.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	16
4	E04.A-K.1.1.1	Key Ideas and Details-Literature Text	17
4	E04.A-K.1.1.2	Key Ideas and Details-Literature Text	14
4	E04.A-K.1.1.3	Key Ideas and Details-Literature Text	14
4	E04.A-V.4.1.1	Vocabulary Acquisition and Use	23
4	E04.A-V.4.1.2	Vocabulary Acquisition and Use	28
4	E04.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	14
4	E04.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	11
4	E04.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	13
4	E04.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	9
4	E04.B-C.3.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	14
4	E04.B-K.1.1.1	Key Ideas and Details-Informational Text	20
4	E04.B-K.1.1.2	Key Ideas and Details-Informational Text	16
4	E04.B-K.1.1.3	Key Ideas and Details-Informational Text	14
4	E04.B-V.4.1.1	Vocabulary Acquisition and Use	27
4	E04.B-V.4.1.2	Vocabulary Acquisition and Use	22
5	E05.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	16
5	E05.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	12
5	E05.A-K.1.1.1	Key Ideas and Details-Literature Text	18
5	E05.A-K.1.1.2	Key Ideas and Details-Literature Text	11
5	E05.A-K.1.1.3	Key Ideas and Details-Literature Text	7
5	E05.A-V.4.1.1	Vocabulary Acquisition and Use	17
5	E05.A-V.4.1.2	Vocabulary Acquisition and Use	25
5	E05.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
5	E05.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6
5	E05.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	15
5	E05.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	4

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
5	E05.B-C.3.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	13
5	E05.B-K.1.1.1	Key Ideas and Details-Informational Text	20
5	E05.B-K.1.1.2	Key Ideas and Details-Informational Text	9
5	E05.B-K.1.1.3	Key Ideas and Details-Informational Text	13
5	E05.B-V.4.1.1	Vocabulary Acquisition and Use	23
5	E05.B-V.4.1.2	Vocabulary Acquisition and Use	16
6	E06.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	25
6	E06.A-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	11
6	E06.A-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	12
6	E06.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
6	E06.A-K.1.1.1	Key Ideas and Details-Literature Text	25
6	E06.A-K.1.1.2	Key Ideas and Details-Literature Text	18
6	E06.A-K.1.1.3	Key Ideas and Details-Literature Text	13
6	E06.A-V.4.1.1	Vocabulary Acquisition and Use	27
6	E06.A-V.4.1.2	Vocabulary Acquisition and Use	41
6	E06.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	10
6	E06.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	24
6	E06.B-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	16
6	E06.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	14
6	E06.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
6	E06.B-K.1.1.1	Key Ideas and Details-Informational Text	38
6	E06.B-K.1.1.2	Key Ideas and Details-Informational Text	20
6	E06.B-K.1.1.3	Key Ideas and Details-Informational Text	12
6	E06.B-V.4.1.1	Vocabulary Acquisition and Use	32
6	E06.B-V.4.1.2	Vocabulary Acquisition and Use	20
7	E07.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	14
7	E07.A-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	12
7	E07.A-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	18

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
7	E07.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
7	E07.A-K.1.1.1	Key Ideas and Details-Literature Text	38
7	E07.A-K.1.1.2	Key Ideas and Details-Literature Text	17
7	E07.A-K.1.1.3	Key Ideas and Details-Literature Text	13
7	E07.A-V.4.1.1	Vocabulary Acquisition and Use	26
7	E07.A-V.4.1.2	Vocabulary Acquisition and Use	23
7	E07.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	13
7	E07.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	17
7	E07.B-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	12
7	E07.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	14
7	E07.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6
7	E07.B-K.1.1.1	Key Ideas and Details-Informational Text	44
7	E07.B-K.1.1.2	Key Ideas and Details-Informational Text	16
7	E07.B-K.1.1.3	Key Ideas and Details-Informational Text	13
7	E07.B-V.4.1.1	Vocabulary Acquisition and Use	33
7	E07.B-V.4.1.2	Vocabulary Acquisition and Use	17
8	E08.A-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	13
8	E08.A-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
8	E08.A-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	14
8	E08.A-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	11
8	E08.A-K.1.1.1	Key Ideas and Details-Literature Text	33
8	E08.A-K.1.1.2	Key Ideas and Details-Literature Text	17
8	E08.A-K.1.1.3	Key Ideas and Details-Literature Text	13
8	E08.A-V.4.1.1	Vocabulary Acquisition and Use	28
8	E08.A-V.4.1.2	Vocabulary Acquisition and Use	30
8	E08.B-C.2.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	18
8	E08.B-C.2.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	23
8	E08.B-C.2.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	12

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
8	E08.B-C.3.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	10
8	E08.B-C.3.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
8	E08.B-K.1.1.1	Key Ideas and Details-Informational Text	30
8	E08.B-K.1.1.2	Key Ideas and Details-Informational Text	21
8	E08.B-K.1.1.3	Key Ideas and Details-Informational Text	12
8	E08.B-V.4.1.1	Vocabulary Acquisition and Use	34
8	E08.B-V.4.1.2	Vocabulary Acquisition and Use	21
Lit	L.F.1.1.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.1.1.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.1.1.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.1.2.1	Vocabulary Acquisition and Use	7
Lit	L.F.1.2.2	Vocabulary Acquisition and Use	7
Lit	L.F.1.2.3	Vocabulary Acquisition and Use	7
Lit	L.F.1.2.4	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
Lit	L.F.1.3.1	Key Ideas and Details-Literature Text	10
Lit	L.F.1.3.2	Key Ideas and Details-Literature Text	8
Lit	L.F.2.1.1	Key Ideas and Details-Literature Text	8
Lit	L.F.2.1.2	Key Ideas and Details-Literature Text	8
Lit	L.F.2.2.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7
Lit	L.F.2.2.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.2.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.2.4	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.3.1	Key Ideas and Details-Literature Text	8
Lit	L.F.2.3.2	Key Ideas and Details-Literature Text	8
Lit	L.F.2.3.3	Key Ideas and Details-Literature Text	10
Lit	L.F.2.3.4	Key Ideas and Details-Literature Text	8
Lit	L.F.2.3.5	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.3.6	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.4.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
Lit	L.F.2.5.1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.F.2.5.2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	9
Lit	L.F.2.5.3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8
Lit	L.N.1.1.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
Lit	L.N.1.1.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
Lit	L.N.1.1.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.1.1.4	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.1.2.1	Vocabulary Acquisition and Use	7
Lit	L.N.1.2.2	Vocabulary Acquisition and Use	7
Lit	L.N.1.2.3	Vocabulary Acquisition and Use	7
Lit	L.N.1.2.4	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
Lit	L.N.1.3.1	Key Ideas and Details-Informational Text	8
Lit	L.N.1.3.2	Key Ideas and Details-Informational Text	8
Lit	L.N.1.3.3	Key Ideas and Details-Informational Text	8
Lit	L.N.2.1.1	Key Ideas and Details-Informational Text	7
Lit	L.N.2.1.2	Key Ideas and Details-Informational Text	8
Lit	L.N.2.2.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.2.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	10
Lit	L.N.2.2.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.3.1	Key Ideas and Details-Informational Text	8
Lit	L.N.2.3.2	Key Ideas and Details-Informational Text	8
Lit	L.N.2.3.3	Key Ideas and Details-Informational Text	9
Lit	L.N.2.3.4	Key Ideas and Details-Informational Text	9
Lit	L.N.2.3.5	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	10
Lit	L.N.2.3.6	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	9
Lit	L.N.2.4.1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	7
Lit	L.N.2.4.2	Key Ideas and Details-Informational Text	8

Diagnostic Categories-Reading (continued)

Item Grade/Course	Eligible Content	Diagnostic Category	Number of Items
Lit	L.N.2.4.3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.4.4	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	9
Lit	L.N.2.4.5	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.5.1	Key Ideas and Details-Informational Text	7
Lit	L.N.2.5.2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.5.3	Key Ideas and Details-Informational Text	8
Lit	L.N.2.5.4	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	9
Lit	L.N.2.5.5	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8
Lit	L.N.2.5.6	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	8

Depth of Knowledge-Mathematics

Grade	Diagnostic Category	DOK 1	DOK 2	DOK 3
3	Numbers and Operations	27	48	0
3	Algebraic Concepts	30	70	0
3	Geometry	17	18	0
3	Measurement Data and Probability	35	101	0
4	Numbers and Operations	37	98	0
4	Algebraic Concepts	26	104	0
4	Geometry	33	41	0
4	Measurement Data and Probability	21	51	0
5	Numbers and Operations	32	84	0
5	Algebraic Concepts	7	31	0
5	Geometry	17	37	0
5	Measurement Data and Probability	7	66	0
6	Numbers and Operations	120	68	0
6	Algebraic Concepts	108	102	0
6	Geometry	15	51	1
6	Measurement Data and Probability	31	82	0
7	Numbers and Operations	21	78	0
7	Algebraic Concepts	26	45	0
7	Geometry	26	86	0
7	Measurement Data and Probability	6	150	2
8	Numbers and Operations	31	23	0
8	Algebraic Concepts	44	135	1
8	Geometry	15	92	2
8	Measurement Data and Probability	9	44	1

Depth of Knowledge-Reading

Grade	Diagnostic Category	DOK 1	DOK 2	DOK 3
K	Key Ideas and Details-Informational Text	7	7	7
K	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	6	10	5
K	Vocabulary Acquisition and Use	0	6	0
		1	11	6
K	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	6	0	6
K	Vocabulary Acquisition and Use	0	6	0
1	Key Ideas and Details-Informational Text	6	8	7
1	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	0	19	8
1	Vocabulary Acquisition and Use	0	7	0
1	Key Ideas and Details-Literature Text	7	8	5
1	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	2	6	6
1	Vocabulary Acquisition and Use	0	6	0
2	Key Ideas and Details-Informational Text	3	14	4
2	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	0	22	6
2	Vocabulary Acquisition and Use	1	6	0
2	Key Ideas and Details-Literature Text	5	10	6
2	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	0	2	12
2	Vocabulary Acquisition and Use	0	7	0
3	Key Ideas and Details-Informational Text	0	8	21
3	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	2	23	16
3	Vocabulary Acquisition and Use	0	44	2
3	Key Ideas and Details-Literature Text	0	40	22
3	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	7	25	11
3	Vocabulary Acquisition and Use	1	48	2
4	Key Ideas and Details-Informational Text	0	2	26
4	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	3	27	15
4	Vocabulary Acquisition and Use	7	41	3
4	Key Ideas and Details-Literature Text	0	25	36

Depth of Knowledge-Reading (continued)

Grade	Diagnostic Category	DOK 1	DOK 2	DOK 3
4	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	8	26	16
4	Vocabulary Acquisition and Use	5	42	2
5	Key Ideas and Details-Informational Text	0	8	20
5	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	1	21	13
5	Vocabulary Acquisition and Use	5	36	1
5	Key Ideas and Details-Literature Text	0	18	27
5	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	2	20	20
5	Vocabulary Acquisition and Use	4	32	3
6	Key Ideas and Details-Informational Text	0	23	33
6	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	1	38	17
6	Vocabulary Acquisition and Use	7	60	1
6	Key Ideas and Details-Literature Text	0	47	25
6	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	4	38	28
6	Vocabulary Acquisition and Use	10	39	3
7	Key Ideas and Details-Informational Text	0	24	28
7	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	1	34	33
7	Vocabulary Acquisition and Use	7	41	1
7	Key Ideas and Details-Literature Text	0	37	25
7	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	4	44	25
7	Vocabulary Acquisition and Use	5	45	0
8	Key Ideas and Details-Informational Text	0	16	30
8	Craft and Structure and Integration of Knowledge and Ideas –Informational Text	3	40	20
8	Vocabulary Acquisition and Use	10	47	1
8	Key Ideas and Details-Literature Text	0	39	32
8	Craft and Structure and Integration of Knowledge and Ideas –Literature Text	3	44	16
8	Vocabulary Acquisition and Use	4	50	1

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 edition). Louisville, KY: American Printing House for the Blind. Available from http://www.aph.org.

Alonzo, A.C. & Gearhart, M. (2006). Considering learning progressions from a classroom assessment perspective. Measurement: Interdisciplinary Research and Perspectives. Vol. 4(1&2) Mahwah, NJ: Lawrence Erlbaum. 99-108.

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton NJ: Educational Testing Service. [Reprint of chapter in R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 508–600). Washington, DC: American Council on Education, 1971.]

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Black, P. & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (ed.), Towards Coherence between Classroom Assessment and Accountability. 103RF Yearbook of the National Society for the Study of Education, Part 2. Chicago, IL: National Society for the Study of Education. 20-50.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., p. 443–507). Washington, DC: American Council on Education.

Cronbach, L., & Shavelson R. L. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391–418.

Data Recognition Corporation. (2003–2010). *Fairness in Testing: Training Manual for Issues of Bias, Fairness, and Sensitivity*. Maple Grove, MN: DRC.

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 preequating the SAT verbal and mathematical sections. (Research Report 85–10). Princeton, NJ: Educational Testing Service.

Eignor, D. R., & Stocking, M. L. (1986). An investigation of the possible causes for the inadequacy of IRT preequating. (Research Report 86–14). Princeton, NJ: Educational Testing Service.

Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24(3) 21–28.

Gong, B. (2008). Developing Learning Progressions to inform Formative Assessment: Five areas to develop. Presentation at the CCSSO FAST SCASS Meeting, February 6, 2008, Atlanta, GA. Center for Assessment

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.

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. (2008). Developing and Using Learning Progressions as a Schema for Measuring Progress National Center for the Improvement of Educational Assessment, Dover, New Hampshire

Hess, K. (2008). Tools and Strategies for Developing and Using Learning Progressions. Five areas to develop. Presentation at the CCSSO FAST SCASS Meeting, February 6, 2008, Atlanta, GA. Center for Assessment

Heritage, M. (2008). Learning Progressions: Supporting Instruction and Formative Assessment. National Center for Research on Evaluation, Standards, and Student Tests (CRESST) paper prepared for the Formative Assessment for Teachers and Students (FAST) State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSS)

Huynh, H. (1976). On the reliability of decisions in domain referenced testing. *Journal of Educational Measurement*, 13, 253–264.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking. New York, NY: 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, No. 1 (Spring), pp. 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.

Leung, C. K., Chang, H. H., & Hau, K. T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. Journal of Technology, Learning, and Assessment, 2(5).

Linacre, J. M. (2009). *A user's guide to WINSTEPS MININSTEP Rasch-model computer programs*. Chicago, IL: Winsteps.

Linacre, J. M. (2009). *WINSTEPS 3.71: 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.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.

Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(3), 200–215.

McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of general theory. *Multivariate Behavioral Research*, 14, 21–38.

Mctighe, J. and Wiggins, G.P. (2005) Understanding by Design. Alexandria, VA: Association for Supervision and Curriculum and Development.

Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational Measurement* (3rd ed., pp. 3–104). New York: American Council on Education.

Pennsylvania Department of Education. (2010). Classroom Diagnostic Tools *Results for Preliminary Benchmarking Activity – Mathematics*. Harrisburg, PA: PDE.

Pennsylvania Department of Education. (2011). Classroom Diagnostic Tools *Results for Preliminary Benchmarking Activity – Reading and Science*. Harrisburg, PA: PDE.

Pennsylvania Department of Education. (2011). Classroom Diagnostic Tools *Results for Preliminary Benchmarking Activity – Writing*. Harrisburg, PA: PDE.

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: 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.

Stearns, M., & Smith, R. M. (2007). *Estimation of classification consistency indices for complex assessments: Model based approaches*. Paper presented at the 2007 Annual Convention of the American Educational Research Association. Chicago, IL.

Stocking, M. L., & Eignor, D. R. (1986). The impact of different ability distributions on IRT preequating. (Research Report, 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.

Webb, N. L. (1999). *Research Monograph No. 18: Alignment of Science and Mathematics Standards and Assessments in Four States*. Madison, WI: National Institute for Science Education.

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.

Wright, B., & Masters, G. (1982). *Rating scale analysis*. Chicago: MESA Press.