# Wisconsin Forward Exam

# Spring 2022
# Technical Report

WISCONSIN DEPARTMENT OF
**Public Instruction**

**Submitted to**
**Wisconsin Department of Public Instruction**
**October 2022**

DATA RECOGNITION
**DRC**
CORPORATION

# Copyright

# Foreword

The technical information herein is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

# Table of Contents

# Appendices

# List of Tables

# List of Figures

# Executive Summary

This report is a technical summary of the 2022 administration of the Wisconsin Forward Exam in English Language Arts (ELA) and Mathematics (administered in grades 3 through 8), Science (administered in grades 4 and 8), and Social Studies (administered in grades 4, 8, and 10).

The Wisconsin Forward Exam assessments are designed to measure students' knowledge of ELA, Mathematics, Science, and Social Studies, and they are aligned with Wisconsin Academic Standards. The ELA, Mathematics, and Science test forms administered in Spring 2022 were developed by Data Recognition Corporation (DRC) using DRC's college- and career-ready item bank. The Spring 2022 Social Studies assessments contained Wisconsin-owned items. Operational test forms originally developed for the Spring 2020 administration were administered for ELA, Mathematics, and Science. These forms were not administered in 2020 following test cancellations in Spring 2020 due to the COVID-19 pandemic. New Social Studies test forms, measuring new Wisconsin Standards for Social Studies, were developed for the Spring 2022 administration. All assessments except for Braille and accommodated paper-based forms were administered online.

## E.1 Overview of the Wisconsin Forward Exam

The Wisconsin Forward Exam is designed to measure Wisconsin Academic Standards, which define the knowledge and skills students need in each grade level to succeed in college, other postsecondary training, and careers.

The Wisconsin ELA and Mathematics grade-level tests have undergone multiple alignment changes since their first administration in the 2005–06 school year, with the latest changes in the 2015–16 administration, which was also the first administration year of the tests under the Wisconsin Forward Exam program. The current ELA and Mathematics assessments are aligned to the Wisconsin Academic Standards adopted in 2010. The reporting scales for the ELA and Mathematics tests were established after the Spring 2016 test administration, and the performance level cut scores were set in Summer 2016. The ELA and Mathematics 2015–16 results are considered the baseline for year-to-year student performance comparisons. The 2021–22 ELA and Mathematics assessments are statistically linked to the established scales, allowing for test score comparability between Spring 2021 and Spring 2022.

Science assessments (grades 4 and 8) have been on a different trajectory. A change to the Science test blueprint and design was made for the Spring 2019 operational test administration. New Science tests, aligned to the new Wisconsin Standards for Science (WSS) adopted in 2017 and the Next Generation Science Standards (NGSS), were developed and administered to Wisconsin students for the first time in Spring 2019. Due to the change of standards, new scales were developed for the new Science tests and new performance level cut scores were set after the Spring 2019 test administration. The 2021–22 Science assessments are statistically linked to the scales established in Spring 2019, allowing for test score comparability across the last three administrations.

A change to the Social Studies test blueprint and design was made for the Spring 2022 operational test administration. The Spring 2022 Social Studies assessments for grades 4, 8, and 10 are aligned to the new Wisconsin Standards for Social Studies adopted in 2018. The new reporting scales for the Social Studies tests were established after the Spring 2022 test administration, and new performance level cut scores were set for these assessments in Spring 2022. The Social Studies 2021–22 results are considered a new baseline for year-to-year student performance comparisons.

All Wisconsin assessments are administered online and contain various item types, including multiple-choice (MC), multi-select (MS), technology-enhanced (TE), evidence-based selected response (EBSR), and short-answer (SA). Braille, print-on-demand, and Spanish translation forms that contain the same items as regular online operational test forms are also available to students who need them.

## E.2 Administration

In Spring 2022, Wisconsin administered summative assessments in ELA and Mathematics to students in grades 3 through 8. Science assessments were administered to students in grades 4 and 8, and Social Studies assessments were administered in grades 4, 8, and 10. The Wisconsin Forward Exam was administered from March 21 to April 29, 2022. Test administration is discussed in Part 4 of this report.

A total of 455 public school districts, 383 choice schools, and 5 private schools had students who participated in at least one Wisconsin Forward Exam test in grades 3 through 8 or grade 10 (Social Studies only). Table E-1 shows test participation rates in Spring 2022. For the purposes of this report, participation rate is defined as the percentage of students who received a valid scale score given the total number of students expected to take the test. The "Enrolled" column shows the total number of students expected to take the test in Spring 2022. The "Number Tested" and "Percent Tested" columns show the number and percentage of students who participated in the test and received a valid scale score. The test participation rates for grades 3 through 7 ranged from approximately 95% to 96% across all content areas. The test participation rates for grade 8 were approximately 94% across all content areas. The participation rate for Social Studies grade 10 was approximately 88%. While the participation rates for grades 3 through 8 were similar to the pre-pandemic participation rates, the participation rate observed for grade 10 was lower compared to the rates observed in typical administration years (presented in Part 10 of the report). Further analysis of the Spring 2022 participation rates is provided in Part 4 of this report.

## E.3 Student Performance

This is the sixth year of the ELA and Mathematics scores being reported on the scales established in Spring 2016. Spring 2022 also marks the third year of the Science assessments measuring the new Wisconsin Standards for Science and the first year of the Social Studies assessments measuring the new Wisconsin Standards for Social Studies. Tables E-2 and E-3 present the percentages of students classified as *Proficient* or *Advanced* in 2016 through 2022 in ELA and Mathematics, respectively. Table E-4 shows the percentages of students classified as *Proficient* or *Advanced* in 2019, 2021, and 2022 in Science. Due to setting new scales and performance cut scores for Science after the Spring 2019 test administration, student results in Science are not directly comparable between the Spring 2019 administration and previous administrations, and the previous data are not reported in this table. Student results are comparable between the Spring 2019 and Spring 2022 administrations for Science. Table E-5 shows the percentages of students classified as *Proficient* or *Advanced* in 2022 in Social Studies. New performance level cut scores were established for Social Studies after the 2021–22 test administration. Therefore, student performance in Social Studies in the Spring 2022 administration cannot be directly compared with any of the previous administrations, and the past performance data are not presented in this report.

Caution should be used when making statewide data comparisons over time. Due to the COVID-19 pandemic and disruptions to student learning in the Spring 2020–21 school year, the participation rates for Spring 2021 were considerably lower than in previous years or the current year. In addition, the makeup of the Spring 2021 tested population was not representative of the Spring 2021 enrolled population. Therefore, the longitudinal data trend might have not accurately represented changes in state-level student performance between 2019 and 2021 and again between 2021 and 2022.

The percentages of students classified as *Proficient* or *Advanced* in ELA in 2022 ranged from approximately 33% for grade 8 to approximately 42% for grade 4. The percentages of students classified as *Proficient* or *Advanced* in Mathematics in 2022 ranged from 30% for grade 8 to 47% for grade 3. Approximately 51% of students were classified as *Proficient* or *Advanced* in Science grade 4, and approximately 49% of students were classified as *Proficient* or *Advanced* in Science grade 8. The percentages of students classified as *Proficient* or *Advanced* in Social Studies in 2022 were about 59% in grades 4 and 8 and approximately 48% in grade 10. More details on student performance are provided in Part 10 of this report.

It was observed that the percentages of students in the *Proficient* or *Advanced* performance categories were lower in Spring 2022 than in Spring 2019 for all ELA, Mathematics, and Science grades except for ELA grade 5 (where there was a negligible difference). This observed change in performance between Spring 2019 and Spring 2022 should be interpreted in the context of circumstances related to the long-lasting effects of the COVID-19 pandemic (including school closures, nonstandard instruction delivery modes in the 2020–21 school year, and potential diminished opportunity to learn for students) and slow educational recovery. The Spring 2021 and Spring 2022 results cannot be reliably compared due to differences in student populations in these administration years.

## E.4 Validity of Intended Interpretation of Test Scores

Most sections of this report are designed to provide validity evidence to support the use and intended interpretation of the Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies test scores. Test scores are used to identify strengths and areas for improvement in Wisconsin's student performance, to inform stakeholders (teachers, school administrators, district administrators, DPI staff members, parents, and the public) about the state's status with respect to its progress toward meeting the academic performance standards of the state, and to meet the requirements of the state's accountability program. Part 2 of this report provides the validity framework and a summary of the validity evidence for the Wisconsin Forward Exam.

Evidence of validity based on test content was supported by the test specifications, including the test design and test blueprint. Wisconsin assessments were developed in alignment with Wisconsin Academic Standards. A rigorous item review and test form development process were implemented to select items from DRC's college- and career-ready item pool for the ELA, Mathematics, and Science assessments and from Wisconsin-owned pool of items, written by DRC content specialists and reviewed by Wisconsin educators, for Social Studies. More details on test content and test development are provided in Part 3 of this report.

With the exceptions of Braille and a limited number of paper-based test forms, Wisconsin Forward Exam assessments were administered online in a standardized manner, further supporting validity of the intended score interpretation. Universal tools were available for all students to use. Designated supports and accommodations were available to students for whom such aids were deemed appropriate and/or indicated in their Individualized Education Programs. More details on test administration and use of universal tools, designated supports, and accommodations are provided in Part 4 of this report.

Scoring of technology-enhanced, short-answer, multi-select, evidence-based selected-response, and text-dependent analysis items followed predefined scoring criteria. The multiple-choice, multi-select, technology-enhanced, short-answer, and evidence-based selected-response items were autoscored. ELA text-dependent analysis items were scored via artificial intelligence engine supported by human reader

score verification. The inter-rater reliability statistics demonstrated that the text-dependent analysis items were scored reliably (refer to Parts 5 and 8 for details).

The test scaling and equating was conducted using item response theory (IRT) methodology. Students' scale scores were derived using item parameters estimated after the Spring 2022 test administration. The IRT models used for Wisconsin Forward Exam scaling were appropriate for the test data supporting the operational data analysis and ensuring that the test items, as well as the overall tests, were functioning appropriately. For details on test scaling and equating, refer to Part 6. The cut scores used to classify students into different performance levels and associated performance level descriptors were established during the Summer 2016 standard setting for ELA and Mathematics, the Spring 2019 standard setting for Science, and the Spring 2022 standard setting for Social Studies in a collaborative and participatory process, further supporting the validity and interpretation of the Wisconsin Forward Exam scores (refer to Part 7 for details).

Evidence of construct-related validity—supporting the intended interpretation of test scores and their use—was provided through studies of test reliability, evaluation of test fairness, evaluation of internal test structure, and evaluation of the relationship of test scores with external variables. The reliability analysis results indicated that the Wisconsin Forward Exam tests produce scores that would be relatively stable if the tests were administered repeatedly under similar conditions (refer to Part 8 of this report for details).

Test and item fairness were evaluated through differential item functioning analysis (refer to Part 9 of this report for details). The assumption that the content area Wisconsin Forward Exam tests were unidimensional (i.e., each grade-level test measured one primary dimension) was confirmed through principal component analysis. The evidence of the validity of the intended interpretation of the Wisconsin Forward Exam test scores based on the relationships with other variables was evaluated through the correlations computed between the ELA, Mathematics, Science, and Social Studies scale scores. The student scores were found to be highly, but not perfectly, related to each other, suggesting that while different constructs are being measured, the two assessments may also be tapping into a similar knowledge base or general underlying ability. When considering the Wisconsin Academic Standards and the percentages of students classified as *Proficient* or *Advanced* (based on the Wisconsin Forward Exam cut scores for ELA, Mathematics, and Science), the Wisconsin Forward Exam impact data are in alignment with the National Assessment of Educational Progress (NAEP) impact data. This provides evidence of the relationship between the state assessments and the national assessments in these content areas (see Part 9 of this report for details).

Finally, in Part 10 of this report, test results are presented in the context of score reports that aid the user in understanding the meaning of the test results. The current administration test results are presented for the total population and subgroups of students. The longitudinal test results are also presented for ELA, Mathematics, and Science. Monitoring group performance is possible if the test content and the construct measured by the test are comparable from year to year and if the scores are reported on the same scale used in previous years.

**Table E-1. Test Participation Rates in Spring 2022**

| Grade | Enrolled | English Language Arts | | Mathematics | | Science | | Social Studies | |
|---|---|---|---|---|---|---|---|---|---|
| | | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested |
| 3 | 60956 | 58275 | 95.60 | 58449 | 95.89 | | | | |
| 4 | 61465 | 58773 | 95.62 | 58931 | 95.88 | 58880 | 95.79 | 58833 | 95.72 |
| 5 | 61735 | 59044 | 95.64 | 59187 | 95.87 | | | | |
| 6 | 63048 | 60112 | 95.34 | 60234 | 95.54 | | | | |
| 7 | 65386 | 61871 | 94.62 | 61969 | 94.77 | | | | |
| 8 | 66789 | 62684 | 93.85 | 62762 | 93.97 | 62644 | 93.79 | 62606 | 93.74 |
| 10 | 67427 | | | | | | | 59391 | 88.08 |

**Table E-2. Percentage of Students Classified as *Proficient* or *Advanced* in 2016 through 2022, English Language Arts**

| Grade | English Language Arts | | | | | |
|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2021 | 2022 |
| 3 | 43.13 | 41.83 | 39.75 | 38.69 | 34.56 | 34.93 |
| 4 | 43.30 | 46.72 | 43.91 | 42.98 | 40.12 | 41.60 |
| 5 | 42.47 | 46.42 | 44.17 | 40.06 | 37.52 | 40.28 |
| 6 | 42.58 | 45.26 | 42.86 | 40.96 | 38.45 | 37.98 |
| 7 | 41.98 | 43.63 | 45.15 | 44.87 | 42.92 | 38.26 |
| 8 | 41.56 | 41.12 | 37.33 | 37.03 | 35.66 | 33.46 |

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 90%.

**Table E-3. Percentage of Students Classified as *Proficient* or *Advanced* in 2016 through 2022, Mathematics**

| Grade | Mathematics | | | | | |
|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2021 | 2022 |
| 3 | 48.00 | 48.03 | 49.83 | 49.44 | 44.99 | 47.20 |
| 4 | 44.20 | 43.50 | 44.46 | 45.05 | 41.07 | 43.73 |
| 5 | 44.08 | 44.46 | 45.95 | 46.58 | 41.59 | 44.80 |
| 6 | 42.84 | 43.61 | 43.96 | 42.49 | 35.57 | 38.76 |
| 7 | 39.26 | 39.29 | 38.97 | 38.83 | 34.84 | 33.73 |
| 8 | 33.86 | 34.62 | 36.61 | 35.85 | 30.00 | 30.31 |

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 90%.

**Table E-4. Percentage of Students Classified as *Proficient* or *Advanced* in 2019 through 2022, Science**

| Grade | Science | | |
|---|---|---|---|
| | 2019 | 2021 | 2022 |
| 4 | 52.78 | 51.16 | 50.85 |
| 8 | 53.95 | 51.47 | 48.90 |

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 90%.

**Table E-5. Percentage of Students Classified as *Proficient* or *Advanced* in 2022, Social Studies**

| Grade | Social Studies 2022 |
|-------|---------------------|
| 4     | 58.89               |
| 8     | 58.73               |
| 10    | 47.67               |

Note: Caution should be exercised when interpreting the Spring 2022 statewide data for grade 10 due to participation rates below 90%.

# Part 1: Overview

The *Wisconsin Forward Exam Spring 2022 Technical Report* documents the processes and procedures applied in test development, administration, and scoring, as well as the assessment results. This report also provides evidence in support of the validity and reliability of the testing program in adherence to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). This report demonstrates that the Spring 2022 Wisconsin Forward Exam adhered to the appropriate standards and practices of educational assessment. Ultimately, this report provides evidence that valid inferences about Wisconsin student performance can be derived from this assessment.

## 1.1 Historical Background

The Improving America's Schools Act of 1994 required that states establish challenging academic standards as well as aligned annual assessments. The Goals 2000: Educate America Act and the Elementary and Secondary Education Act (ESEA) spelled out additional requirements to ensure that citizens receive coherent information about whether and to what degree students are meeting rigorous academic standards. This Technical Report is an important part of meeting those requirements.

Wisconsin students in grades 4, 8, and 10 began taking the Wisconsin Knowledge and Concepts Examination (WKCE) norm-referenced assessments in the 1997 school year. At that time and in the following years, *TerraNova™* tests developed by CTB/McGraw-Hill (1997, 2000, 2009) were used. The selection of those tests was partly predicated on an awareness of the academic standards being developed. In January 1998, the Wisconsin Model Academic Standards (WMAS) were adopted. These new standards were the work of the Governor's Commission on Wisconsin Model Academic Standards, chaired by then Lieutenant Governor Scott McCallum and the Wisconsin Department of Public Instruction (DPI). The assessments aligned to WMAS would measure student performance in the same subjects as the *TerraNova* tests.

Beginning in the 2005–06 school year, the federal No Child Left Behind Act (NCLB) required all states to test all students in Reading and Mathematics in grades 3 through 8 and once in high school (in grade 10 under Wisconsin law § 118.30). Based on the NCLB legislation, student performance, reported in terms of proficiency categories, was used to determine the Adequate Yearly Progress (AYP) of students at the school, district, and state levels. Beginning with the 2007–08 school year, states were also required to administer Science assessments at least once in grades 3–5, once in grades 6–9, and once in grades 10–12.

It was within this policy context that the WKCE was constructed, as a criterion referenced test, for the Fall 2005 administration, replacing the previously existing norm referenced WKCE in Reading and Mathematics. The criterion-referenced WKCE was designed specifically for Wisconsin students to measure their performance on the WMAS. These assessments were designed to evaluate students' knowledge and to measure achievement in the basic skills taught in schools at grades 3–8 and 10. The Fall 2013 WKCE was the ninth administration of these assessments and the last administration of Reading, ELA, and Mathematics. The assessments in Science and Social Studies under the existing WKCE model continued to be administered until Fall 2014.

A major change in the Wisconsin assessments occurred for the 2014–15 test administration. First, the ELA and Mathematics assessments were moved from the Fall testing window to the Spring testing window. Second, the new ELA and Mathematics tests for grades 3–8 developed for the Spring 2015 administration consisted of new Smarter Balanced Assessment Consortium (SBAC) items aligned to the Common Core

State Standards (CCSS). Thus, the 2014–15 ELA and Mathematics assessments were not comparable content- and construct-wise to the assessments administered in prior years. Third, while the prior years' assessments included CTB's *TerraNova* items that yielded norm-referenced scores, the 2014–15 assessments did not include such items. Fourth, the regular versions of the 2014–15 assessments were administered as fixed forms in the online mode, in contrast to the previous assessments, which were all administered in the paper-and-pencil mode. Fifth, TE item types were introduced in the 2014–15 online test administration. Last, the student test scores for ELA and Mathematics were reported on SBAC scales and the students were classified into performance levels based on SBAC cut scores. Further details on the structure and reporting of the Spring 2015 ELA and Mathematics assessments (called the Wisconsin Badger Exam) can be found at https://dpi.wi.gov/assessment/historical/smarter.

The ELA and Mathematics assessments underwent yet another change in the 2015–16 administration year. The Wisconsin DPI partnered with DRC to develop new ELA and Mathematics assessments for grades 3–8 for the Spring 2016 administration. The items contained in these assessments were drawn from DRC's nationally field-tested college- and career-ready (CCR) item bank and aligned with Wisconsin Academic Standards for ELA and Mathematics. The new assessment program is called the Wisconsin Forward Exam, and the new ELA and Mathematics tests were administered online in Spring 2016. Since the new assessments did not contain any items from the 2014–15 Wisconsin Badger Exam tests, the new scales were not statistically linked to the previous scales. The new reporting scales for the ELA and Mathematics tests were developed after the Spring 2016 test administration, and the new performance level cut scores were set for these assessments in Summer 2016.

Science (grades 4 and 8) and Social Studies (grades 4, 8, and 10) assessments have been on a different trajectory, as they continued to be aligned with the WMAS. However, the test administration for these assessments was moved from the Fall window to the Spring window for the 2015–16 administration year. The items contained in the Science and Social Studies tests were mainly drawn from the pool of previously administered items, but new items were also included. Several of the previously administered items were edited to improve item quality and reflect test content changes over time. Despite the fact that many Science and Social Studies items in the Spring 2016 administration came from the previous item pool, statistically linking the Spring 2016 forms to the previous forms was not recommended due to the change of the testing window and the numerous changes to the items themselves. Instead, similar to what was done for the ELA and Mathematics assessments, new scales were developed for the Science and Social Studies tests under the new Wisconsin Forward Exam program. Following the new scale development, the new performance level cut scores were set for Science and Social Studies in Summer 2016.

Details regarding development, scaling, reporting, and standard setting for all Spring 2016 assessments are included in the *Wisconsin Forward Exam Spring 2016 Technical Report* available at https://dpi.wi.gov/assessment/forward/resources.

Spring 2022 was the sixth administration year for the Wisconsin Forward Exam in ELA and Mathematics. Spring 2022 was also the third administration year for the new Wisconsin Forward Exam in Science, aligned to the new WSS and the NGSS. The new Science assessments focus on content understanding linked to work with science and engineering practices and crosscutting concepts as detailed in the *National Research Council Framework for K–12 Science Education* (National Research Council, 2012). The Spring 2022 ELA and Mathematics assessments were statistically linked to their respective Spring 2016 scales, and Science assessments were statistically linked to the Spring 2019 Science scales, allowing for student score comparisons across administrations for these content areas.

Spring 2022 was the first administration year for the Wisconsin Forward Exam in Social Studies aligned to the new Wisconsin Standards for Social Studies. These assessments included all new items that were field-tested in Spring 2021. No items from the prior operational assessments were included in the Social Studies tests in Spring 2022. New scales were developed, and the new performance level cut scores were set for Social Studies in Spring 2022.

This Technical Report documents all aspects of the 2021–22 testing cycle. The structure of this report mirrors the testing cycle. A brief content summary of the report is provided later in this part of the report.

## 1.2 Uses of Test Scores

Validity is the overarching component of the Wisconsin Forward Exam program. The following excerpt is from the *Standards for Educational and Psychological Testing* (hereafter the Standards) (AERA, APA, & NCME, 2014):

> Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. Different components of validity evidence . . . include evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all test takers, as appropriate to the test interpretation in question. (p. 22)

As stated by the Standards, the validity of a testing program hinges on the use of the test scores. Validity evidence that supports the uses of the Wisconsin Forward Exam scores is provided in this Technical Report. This section examines some possible uses of the Wisconsin Forward Exam scores.

Part 2 of this Technical Report provides a summary of the evidence of the validity of intended score interpretation of the Wisconsin Forward Exam. Parts 3 through 10 of the report contain details of the validity evidence as well as technical support for some of the interpretations and uses of test scores. The information in Parts 3 through 10 also provides a firm foundation of evidence that the Wisconsin Forward Exam measures what it is intended to measure. However, this Technical Report cannot anticipate all possible interpretations and uses of the Wisconsin Forward Exam scores. It is recommended that policy and program evaluation studies, in accordance with the Standards, be conducted to support some of the uses of the Wisconsin Forward Exam scores.

The validity of a test score ultimately rests on how that test score is used. To understand whether a test score is being used properly, one must first understand the purpose of the test. The intended uses of the Wisconsin Forward Exam scores include the following:

- Identifying students' strengths and areas in need of improvement
- Communicating expectations for all students
- Evaluating school-, district-, and state-level programs
- Informing stakeholders (i.e., teachers, school administrators, district administrators, DPI staff members, parents, and the public) about the status of the progress toward meeting academic achievement standards of the state
- Meeting the requirements of the state's accountability program

This Technical Report refers to the use of the test-level scores (scale scores and performance levels) and standard-level (reporting category) scores (standard performance index [SPI] scores and performance levels).

## 1.2.1 Test-Level Scores

At the test level, an overall scale score that is based on student performance on the entire test is reported. In addition, an associated level of performance is reported. These scores indicate, in varying ways, a student's achievement in ELA, Mathematics, Science, or Social Studies. Test level scores are reported at four levels: state, school district, school, and student.

Two types of test-level scores are reported to indicate a student's achievement on the Wisconsin Forward Exam: (1) the scale score and (2) its associated level of performance.

### Scale Scores

A scale score indicating a student's performance is determined for each content area. The overall scale score for a content area quantifies the achievement being measured by the ELA, Mathematics, Science, or Social Studies test. In other words, the scale score represents the student's level of performance, where higher scale scores indicate higher levels of performance on the test and lower scale scores indicate lower levels of performance.

### Levels of Performance

A student's performance on the ELA, Mathematics, Science, or Social Studies Wisconsin Forward Exam is reported in one of four levels of performance: *Below Basic*, *Basic*, *Proficient*, or *Advanced*. The cut scores for the levels of performance for ELA and Mathematics were recommended by Wisconsin educators at the standard setting workshop in June 2016. The cut scores for Science were established during the standard setting workshop in May 2019 and the cut scores for Social Studies were established during the standard setting workshop in May 2022. The cut scores reflect the expectations of Wisconsin educators of what Wisconsin students should know and be able to do in ELA, Mathematics, Science, and Social Studies (see Part 7 of this report for a brief description of the Wisconsin Forward Exam standard setting).

### Use of Test-Level Scores

The Wisconsin Forward Exam scale scores and performance levels provide summary evidence of student achievement in ELA, Mathematics, Science, and Social Studies. Classroom teachers may use these scores as evidence of student achievement in these content areas. At the aggregate level, district and school administrators may use this information for activities such as curriculum planning. The results presented in this Technical Report provide evidence that the scale scores are valid and reliable indicators of student performance in ELA, Mathematics, Science, and Social Studies.

As stated earlier in this report, the Spring 2022 test results should be considered in the context of factors related to the potential long-term educational, social and emotional effects of the COVID-19 pandemic on students and their families.

### 1.2.2 Standard-Level Subscores and Performance Levels

The standard-level subscores (i.e., the SPI scores) indicate student performance on a content standard and can be interpreted as an estimate of the number of items a student would be expected to answer correctly if there had been 100 similar items for a given reporting category. The SPI scores are criterion-referenced scores, in that they estimate how much a student knows in a clearly defined skill domain (i.e., the criterion). The SPI scores are computed for content standards measured by at least four items.

Based on their SPI scores, students are classified in one of the four content category performance levels: *Below Basic*, *Basic*, *Proficient*, or *Advanced*. The SPI cut scores separating these performance levels are derived as expected percentages of possible score points for a given standard (content category) for students whose total test score is at the corresponding total test cut score (*Basic*, *Proficient*, or *Advanced*).

### Use of the Standard-Level Subscores

The purpose of reporting SPI scores on the Wisconsin Forward Exam is to show the relationship between the overall achievement being measured (represented by the test score) and the skills within each of the content standards associated with the content area. Teachers may use the SPI scores for individual students as indicators of strengths and needs, but the SPI scores are best corroborated by other evidence, such as homework, class participation, diagnostic test scores, or observation. Part 3 of this Technical Report provides evidence of content validity that supports the use of the standard-level subscores. Part 9 of this Technical Report provides evidence of construct validity that further supports the use of these subscores.

District and school administrators may compare their results by content standard and grade level with the state results to better understand students' strengths and needs within a particular content area and grade level. Caution should be exercised when comparing standard-level subscores across years because different items will contribute to these subscores and these items may vary in difficulty between test forms or test administrations.

## 1.3 Technical Report Structure

This Technical Report documents, in the subsequent parts, the major activities of the testing cycle. It provides comprehensive details that confirm that the processes and procedures applied in the Wisconsin Forward Exam adhere to appropriate professional standards and practices of educational assessment. Ultimately, this report provides evidence that valid inferences about Wisconsin student performance can be derived from the Wisconsin Forward Exam. An overview of the subsequent parts within this report is provided below.

### Part 2: Validity Framework

Part 2 of the Technical Report discusses the concept of validity evidence. This Technical Report is composed of evidence that supports the use of the Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies scores. In Part 2, some of the uses of the Wisconsin Forward Exam scores are discussed.

### Part 3: Test Content and Test Development

Part 3 of this report describes the test blueprint, test design, the item development and test form development process, and some aspects of the content-related validity of the Wisconsin Forward Exam. More specifically, it describes how DRC and DPI collaborated to ensure that the appropriate content was included in the Wisconsin Forward Exam and to ensure that the test items adequately sampled the domain of content knowledge necessary to make legitimate inferences about student performance. The Wisconsin

Academic Standards were the basis of the test blueprints and item specifications for their respective content areas. Wisconsin educators were involved in reviewing the items in all content areas to ensure the appropriateness of the test to the standards. The first item review for grades 3–8 in ELA and Mathematics occurred in December 2015. The first item review for new assessments for grades 4 and 8 in Science occurred in August 2017. In addition, the first item review for Social Studies assessments in grades 4, 8, and 10, measuring new Wisconsin Standards for Social Studies occurred in August 2019. Each year after that, with the exception of year 2020, new items were reviewed and added to the Wisconsin pool of items for future field-testing. The item reviews served to establish the accessibility of the items and reading passages. Simultaneously, DRC created the test specifications documents that were later approved by DPI and will continue to serve as a foundation for item and test development. Additional item reviews, supported by the item data, occurred after each field test administration and were conducted by DPI content experts. The purpose of these reviews was to refine the pool of items from which the subsequent operational test forms were selected.

Part 3 also presents the Wisconsin Forward Exam design and discusses features of the Spring 2022 Wisconsin Forward Exam forms. The Spring 2022 test forms adhered to the approved test blueprints, test designs, and psychometric specifications.

## Part 4: Test Administration

Part 4 describes test administration and accommodations. It also provides information on student participation in the ELA, Mathematics, Science, and Social Studies assessments in Spring 2022. In the 2021–22 school year, the Wisconsin Forward Exam was administered to Wisconsin students for the sixth time.

The Spring 2022 Wisconsin Forward Exam was an online assessment with a single print-on-demand form at each grade level. Student responses to the print-on-demand form were transcribed by a proctor into the online assessment system. Other variations of the forms included stacked Spanish translation forms, video sign language, and closed-captioning. These were provided in an online format at each grade level.

Test administration was conducted during a six-week window from March 21 to April 29, 2022. All testing was conducted online, administered via DRC's INSIGHT platform.

Part 4 of the Technical Report serves to describe the processes and activities implemented and information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students.

## Part 5: Scoring

Part 5 documents the scoring process for different item types: scanning of multiple choice (MC) items and multi-select (MS) items; autoscoring of technology-enhanced (TE) items, short-answer (SA) items, and evidence-based selected response (EBSR) items; and artificial intelligence (AI) scoring and handscoring of text-dependent analysis (TDA) items. The description of the handscoring process includes the development and review of the scoring rubrics, anchor (sample) paper selection, training of scoring personnel, ongoing quality assurance, and a systematic review of the resulting score distributions supporting reliable and valid reported test scores. The scoring rubric used in the handscoring of the TDA writing items is presented in detail.

## Part 6: Psychometric Analyses

The Spring 2022 administration year is the sixth administration year for the Wisconsin Forward Exam in all grades and content areas. Part 6 discusses characteristics of the sample of student data used for data analysis and describes the classical and item response model (IRT) procedure implemented to analyze the Wisconsin Forward Exam test data. The results of item analysis and test statistical properties are presented in this Part. The Wisconsin Forward Exam data were calibrated using two different item response theory (IRT) models, one for MC items and one for non-MC items. Evaluation of the sufficiency of the IRT model results included model-to-data fit and the standard error of measurement (SEM). The equating of Spring 2022 ELA and Mathematics test forms to the scales established after the Spring 2016 administration and the equating of Spring 2022 Science test forms to the scales established after the Spring 2019 administration were performed using the Stocking and Lord procedure. New scales were established for Social Studies assessments and these assessments were not equated to the previous scales. Item-pattern scoring was applied to the Spring 2022 Wisconsin Forward Exam. As discussed in Part 6, item-pattern scoring is generally recommended over number-correct scoring because it produces more accurate scores for individual students. Part 6 also explains how a student's scale score is derived from the raw score using item-pattern scoring.

## Part 7: Standard Setting

Part 7 provides a brief overview of the standard setting process, during which the performance level cut scores were set for the ELA and Mathematics tests in Summer 2016, for the Science tests in Spring 2019, and for the Social Studies tests in Spring 2022. The standard setting methodology and results, including short performance level descriptors and cut scores, are presented.

## Part 8: Studies of Reliability

Part 8 elaborates on the reliability of the test based on results presented in previous parts of the report. Standard error of measurement (SEM) was assessed for raw scores and scale scores. Internal consistency was evaluated for all tests for the total student population and for subgroups identified by gender, race/ethnicity, economic status, disability status, accommodation use, and English language proficiency. Classification consistency and accuracy were estimated for performance classification. In addition, inter-rater reliability was computed for TDA items on ELA tests that were scored using the AI scoring engine with human scorer verification.

## Part 9: Studies of Construct-Related Validity

Part 9 provides additional construct-related validity evidence supporting the Wisconsin Forward Exam. An analysis of differential item functioning is presented. Principal component analysis, correlations among content standards (reporting category scores), and a relationship between the Wisconsin Forward Exam scores and external variables are presented in the context of construct validity. Forensic analysis procedures, implemented to detect possible aberrant testing behavior, are also discussed.

## Part 10: Test Results

Part 10 includes short descriptions of reports provided to end users, including individual student reports and aggregate reports. It also contains information on the results of the Spring 2022 Wisconsin Forward Exam administration. Detailed summary statistics of the total scale scores and performance levels and the SPI scores are provided for the total population and for subgroups identified by gender, race/ethnicity, economic status, disability status, accommodation use, and English language proficiency. Longitudinal results are also presented for ELA, Mathematics, and Science.

## Part 11: Summary and Recommendations

Key findings of the Spring 2022 Wisconsin Forward Exam administration are presented in the body of the report. However, some issues of a more technical nature that stand out as key recommendations and summary statements that should be considered in subsequent administrations are presented in Part 11. Recommendations based on the Spring 2022 Wisconsin Forward Exam administration may cover different phases of the testing cycle: item development; scoring; and psychometric, or measurement-based, research and evaluation.

# Part 2: Validity Framework

Validity is the overarching component of the Wisconsin Forward Exam program. The following excerpt is from the Standards (AERA, APA, & NCME, 2014):

> Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. Different components of validity evidence . . . include evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all test takers, as appropriate to the test interpretation in question. (p. 22)

As stated by the Standards, the validity of a testing program hinges on the interpretation of the test scores. Validity evidence that supports the uses of the Wisconsin Forward Exam test scores is provided in this Technical Report.

The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for particular purposes or actions. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process. Every aspect of an assessment provides evidence in support of (or in challenge to) the validity of an intended interpretation of test scores, including design, content specifications, item development, psychometric quality, and inferences made from the results.

## 2.1 Sources of Validity Evidence

The sources of validity evidence described in the Standards (AERA et al. 2014, pp. 26–31) include evidence based on test content, evidence based on response processes, evidence based on internal test structure, evidence based on relationships with other variables, and evidence based on consequences of testing. These sources of validity evidence are briefly described below.

Validity evidence based on test content can be obtained from an analysis of the relationship between the content of a test and the construct it is intended to measure (AERA et al. 2014, p.14). It refers to traditional forms of content validity evidence and is supported by a correspondence between test content and a specification of the content domain. This type of evidence can be demonstrated through consistent adherence to test blueprints, through a high-quality test development process that includes review of items for accessibility to English language learners and students using testing accommodations, and through alignment studies.

Validity evidence based on response processes relies in large degree on the evaluation of the cognitive processes of examinees responding to various types of items and the relationship between these processes and the construct being measured. Direct evidence based on response processes typically comes from analyzing test takers' individual responses or from questioning test takers from various groups that make up the intended test-taking population about their performance or response strategies on specific items (AERA et al. 2014, p.15). Such evidence can be gathered through cognitive labs conducted as part of the field test data analysis. Validity evidence based on response process is also supported by a relationship between the item type, format, and content and the construct being measured. For example, if a test is intended to measure a certain set of skills, it is important to determine whether the items included in the test are, in fact, designed to measure these skills or knowledge. In addition, evaluation of student written responses (e.g., text-dependent analysis) further contributes to the validity evidence based on response

processes. In such cases, validity evidence includes the extent to which the processes of item response scoring, whether by a human reader or by an artificial intelligence engine, are consistent with the intended interpretation of scores. For example, scorers are expected to apply particular criteria in scoring students' responses and not be influenced by factors that are irrelevant to the intended interpretation of the scores (AERA et al., 2014, pp. 15–16). Recruitment and training of human scores as well as monitoring the artificial intelligence scoring processes and results, contribute to the validity evidence based on response processes.

Validity evidence based on internal test structure refers to the fact that "analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (AERA et al., 2014, p. 16). Such analyses may include statistical analyses of items and subscores conducted to investigate the dimensionality of an assessment. Procedures for gathering such evidence may include factor analysis for single assessments and evaluation of the continuity of the construct across grades for vertically scaled assessments. Internal test structure can also be evaluated using indices of measurement precision such as test reliability, decision accuracy and consistency, generalizability coefficients, and standard errors of measurement. Evaluation of the correlation coefficients that measure the relationship between the content standard (domain) scores and studies of whether test items may function differently for different subgroups of students are additional sources of validity evidence based on internal test structure.

Validity evidence based on relationships to other variables refers to "evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" (AERA et al., 2014, p. 16). In educational testing, such evidence is often gathered through studies of correlations between the test scores and measures of different or similar constructs. As stated in the Standards, relationships between test scores and other measures intended to assess the same or similar constructs provide convergent evidence, whereas relationships between test scores and measures of different constructs provide discriminant evidence (AERA et al., 2014, pp. 16–17).

Validity evidence based on the consequences of testing is ultimately determined by the stakeholders. Stakeholders decide the purpose and interpretation of scores within their system of reporting and accountability. DRC provides information about test content and technical quality but does not decide the use of test scores. As such, the validity evidence based on consequences of testing is not addressed in this report.

## 2.2 Summary of Validity Evidence for Wisconsin Forward Exam

In this Technical Report, validity evidence is presented in relation to test content, response processes, internal test structure, and relationship with other variables. Gathering validity evidence related to test consequences is beyond the scope of this Technical Report.

Parts 3 through 10 of this Technical Report provide evidence for the uses as well as technical support for some of the interpretations and uses of test scores. As the Technical Report progresses part by part, it moves through the phases of the testing cycle. Each part of the Technical Report details the procedures and processes applied in the Wisconsin Forward Exam program as well as the test results. Each part highlights the meaning and significance of the procedures, processes, and results in terms of validity evidence or a relationship to the Standards. A summary of Wisconsin Forward Exam validity evidence as documented in Parts 3 through 10 is presented here.

Part 3 of the Technical Report documents evidence of the content-related validity demonstrated through each Wisconsin Forward Exam assessment's consistent adherence to the assessment blueprints, which were constructed by DPI based on the Wisconsin Academic Standards. This part of the report also presents the test design and describes the key development tasks related to creating the Spring 2022 Wisconsin ELA, Mathematics, Science, and Social Studies operational test forms. This part documents the involvement of Wisconsin educators, DPI, and DRC in the item review and test development process. The test development process and the involvement of Wisconsin educators in that process forms an important part of the validity of the entire Wisconsin Forward Exam program. The knowledge, expertise, and professional judgment offered by Wisconsin educators ultimately ensures that the content of the Wisconsin Forward Exam forms an adequate and representative sample of appropriate content and that the content forms a legitimate basis upon which to derive valid conclusions about student achievement. The blueprint and design as well as the item and test development activities described in Part 3 explain how specific development processes provide evidence in support of the validity of an intended interpretation of test scores, primarily based on the test content and through the use of expert professional judgment from Wisconsin educators and from DRC test development specialists. The foundational documents—test blueprints and test designs—developed and approved during the initial phases of test development served as critical guides throughout the development and field-testing of items. These documents contribute to ensuring that each form of the test accurately measures the content in consistent and stable ways, thus providing evidence supporting using test scores as an indicator of student achievement of Wisconsin standards.

Part 3 provides evidence to support the validity of an intended interpretation of test scores based on test content of the Wisconsin Forward Exam and address AERA, APA, & NCME (2014) Standards 3.1, 3.2, 3.9, 4.0, 4.1, 4.7, and 4.12.

Part 4 of the Technical Report discusses the processes, procedures, and policies that guide the administration of the Wisconsin Forward Exam, including accommodations, security, and procedures provided to test administrators and school personnel. The following AERA, APA, & NCME (2014) Standards are addressed: 3.4, 3.5, 4.15, 4.16, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7. The process, procedures, and policies detailed in this section contribute to the validity of an intended interpretation of test scores by reducing the impact of construct-irrelevant variables (e.g., nonstandard administration methods, limitations associated with student disabilities, security breaches) on test performance.

Part 5 of the Technical Report demonstrates adherence to AERA, APA, & NCME (2014) Standards 4.18, 4.20, 6.8, and 6.9. It describes how MC, MS, EBSR, SA, and TE items were autoscored and how text-dependent analysis (TDA) items were scored by the artificial intelligence (AI) engine followed by a human reader verification. Training of the AI engine as well as the handscoring process, including training and selection of readers, and validation of scoring accuracy, are discussed and support the validity of an intended interpretation of test scores based on the response processes. The procedures described in this section contribute to the evidence of the validity of an intended interpretation of test scores of the Wisconsin Forward Exam by preventing hardware- or software-related errors in machine scoring and reducing construct-irrelevant score variance associated with variations in readers' interpretation and application of scoring rubrics for TDA items.

Part 6 describes the classical and IRT item and test analysis, including item calibration, test equating, and test scaling. The calibration, equating, and scaling methods as well as the processes and procedures for deriving scale scores from response patterns are also described in this part of the Technical Report. Some references to introductory and advanced discussions of IRT are provided. Several axes upon which to evaluate the calibration, equating, and scaling procedures, such as the models and data used, the software

applied, the vertical relationship across grades, the estimation of parameters, the fit, the SEM, and the IRT scoring method, are discussed. Part 6 of this report addresses AERA, APA, & NCME (2014) Standards 1.8, 2.13, 4.14, 5.2, 5.13, 5.15, and 7.2. These processes and procedures contribute to the validity of an intended interpretation of test scores of the ELA, Mathematics, Science, and Social Studies assessments by providing the opportunity to evaluate items contributing to the accurate and reliable measurement of the intended constructs and by ensuring the stability of the Wisconsin Forward assessments. The results of the psychometric analyses contribute to the validity evidence based on the internal test structure.

Part 7 of the Technical Report provides a summary of the Wisconsin Forward Exam standard setting for ELA and Mathematics, conducted in June 2016, for Science, conducted in May 2019, and for Social Studies, conducted in May 2022, during which the cut scores were set for the four content areas. The process of the standard setting adhered to AERA, APA, & NCME (2014) Standards 5.21 and 5.22, providing evidence of the procedural validity of the standard setting process, methodology, and outcomes.

Part 8 demonstrates adherence to the Standards (AERA, APA, & NCME, 2014) through analyses of the reliability of the Spring 2022 ELA, Mathematics, Science, and Social Studies assessments. It presents a reliability analysis using Cronbach's alpha, SEM, and CSEM results and a detailed analysis of classification consistency and classification accuracy for the total student population and by subgroup. The results of the inter-rater reliability for the ELA text dependent items are also discussed in this part of the report. These analyses address AERA, APA, & NCME (2014) Standards 2.0, 2.3, 2.7, 2.11, 2.13, 2.14, and 2.16. The results of the reliability studies indicate that the Wisconsin Forward Exam tests produce scores that would be stable if the test were administered repeatedly under similar conditions. Reliability is a prerequisite to score validity, and the analyses in this part contribute to the evidence of the validity of an intended interpretation of test scores based on the internal test structure by establishing the reliability of the ELA, Mathematics, Science, and Social Studies scores and proficiency classifications.

As presented in Part 9, additional metrics with which the validity of an intended interpretation of test scores of the ELA, Mathematics, Science, and Social Studies assessments was examined included evaluation of the performance of subgroups of students on the individual test items. As described in Part 9, the issue of item and test fairness is considered during the item development, item review, and test form construction processes and is formally assessed through an analysis of DIF. It is possible for items to function differently across different population groups, and it is also possible that results for an item do not reflect student ability but instead reflect irrelevant information influenced by demographic factors. The DIF analysis serves to determine whether that possibility occurred and, if so, to what degree, item by item, for each of the categories of gender and race/ethnicity as well as for students using testing accommodations. The evaluation of item and test fairness addresses AERA, APA, & NCME (2014) Standards 3.1, 3.2, 3.3, and 3.6.

Also included in Part 9 is additional evidence of the construct-related validity based on the internal test structure, gathered through the analysis of the relationships among test items and test components that conform to the test construct, which in turn provides a basis for test score interpretation. The assumption that the content area Wisconsin Forward Exam tests were unidimensional (that is, each grade-level test measured one primary dimension) was confirmed through principal component analysis. In addition, the relationship between the content area reporting category subscores was explored and validated through the measures of correlations between the reporting category scores within a content area. These analyses addressed AERA, APA, & NCME (2014) Standards 1.13 and 1.21.

The relationship between the Wisconsin Forward Exam scale scores and other variables was examined to provide evidence of the construct validity based on the relationships with other variables. These analyses included measures of cross-content correlations of the ELA, Mathematics, Science, and Social Studies scores for the total population and by subgroups and comparisons of student performance on the Wisconsin Forward Exam with student performance on the National Assessment of Educational Progress (NAEP). These analyses are in alignment with multiple best practices of the testing industry (AERA et al., 2014) and are also presented in Part 9 of the report.

Part 10 of the Technical Report contains descriptions of the score reports available to end users. It also provides information on the results of the Spring 2022 administration for all content areas and longitudinal data trends for ELA, Mathematics, and Science. AERA, APA, & NCME (2014) Standards 5.1, 6.10, 7.0, 7.1, and 12.18 are addressed in Part 10.

While the information in Parts 3 through 10 provides a firm foundation of evidence that the Wisconsin Forward Exam tests measure what they are intended to measure, this Technical Report cannot anticipate all possible interpretations and uses of the Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies scores. It is recommended that policy and program evaluation studies, in accordance with the Standards (AERA et al., 2014), be conducted to support some of the uses of the ELA, Mathematics, Science, and Social Studies scores.

# Part 3: Test Content and Test Development

The purpose of this section is to describe how DRC, DPI, and Wisconsin educators collaborated through a series of test development processes to ensure that appropriate content was included in the Wisconsin Forward Exam and to ensure that test items adequately sampled the domain of content knowledge necessary to make accurate inferences about student performance. Part 3 documents the test blueprints, test designs, item development process, review and field-testing of new items, and the test form development process for the Spring 2022 administration.

This part of the Technical Report is particularly relevant to AERA, APA, & NCME (2014) Standards 3.1, 3.2, 3.9, 4.0, 4.1, 4.7, and 4.12. Each of these Standards and the way each Standard is addressed will be presented in this section of the report. AERA, APA, & NCME (2014) Standard 4.0 states the following:

> Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population. (p. 85)

The test blueprint and item development activities described in this part explain how specific development processes provided evidence to support test validity, primarily content validity, through the use of expert professional judgment from Wisconsin DPI and from DRC test development specialists. The foundational documents—test blueprints and test designs—developed and approved during the initial phases of the project served as critical guides throughout development of the test forms. These documents contributed to ensuring that each test form accurately measured the content in consistent and stable ways, thus providing evidence supporting using the test as an indicator of student achievement of Wisconsin standards.

The Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies domains are generally defined as the knowledge and skills that are identified within the Wisconsin Academic Standards for these content areas. The framework of Wisconsin Academic Standards, in turn, is based on prior consensus among DPI, Wisconsin educators, and experienced subject-matter experts that the framework represents what is important for teachers to teach and students to learn.

Evidence of validity based on test content includes information about the test specifications, including the test design and test blueprint. Test development involves creating a design framework from the statement of the construct to be measured. The primary consideration in the development of the Wisconsin Forward Exam test specifications was the assessment alignment with the Wisconsin Academic Standards. Constraints of the assessment program and state policy decisions were also taken into consideration in development of the test specifications.

The Wisconsin Forward Exam test specifications consist of a test blueprint and a test design for each grade level and content area. In partnership with DRC, DPI created test blueprints and test designs. DRC and DPI content experts scrutinized each blueprint to ensure optimal content coverage and efficient use of time and resources.

## 3.1 Test Blueprints

AERA, APA, & NCME (2014) Standard 4.1 states the following:

> Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s). (p. 85)

The key structural aspect of the Wisconsin Forward Exam for ELA, Mathematics, Science, and Social Studies is the assessment blueprint that specifies the target score points for each grade and content strand or domain. These assessment blueprints were developed by Wisconsin DPI who made recommendations for the test content for each grade and content area, seeking to ensure optimal content coverage of the Wisconsin Forward Exam assessments. In general, each blueprint represents content sampling proportions that reflect the intended emphasis in instruction and mastery for each content area and at each grade level. Specifications for a range of items organized by standard and item type demonstrate the desired proportions within the summative assessment. In summary, the Wisconsin Forward Exam assessment blueprint for a given grade and content area provides guidance on how the standards are measured.

The test blueprints specify the number of item points for each reporting category and subskill as well as the allowable depth of knowledge (DOK) levels for the respective reporting categories. The process used for developing the blueprints for the Wisconsin Forward Exam was a collaborative effort between DRC and DPI. The DPI-approved blueprints can be found in Tables 3-1 through 3-4.

## 3.2 Test Design

The test design for the 2022 operational assessments included the use of items reviewed and approved by Wisconsin educators and DPI. Information concerning the item development process can be found in Section 3.4. Various item types were included in the Wisconsin Forward Exam in order to best assess students' understanding of the standards. A description of item types included in the Wisconsin Forward Exam is presented in Table 3-5. The following sections provide detailed information about the test design of the content areas assessed on the Spring 2022 Wisconsin Forward Exam assessments.

### 3.2.1 English Language Arts

Table 3-6 shows the ELA test design, including the number of passages, items, and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the sixteen test forms at each grade level. The unique items in each test form were field test items. Table 3-6 also identifies the various item types that appeared on the ELA forms, including the points for item scoring. Detailed descriptions of the item types are provided in Table 3-5 of this report.

The ELA section of the Forward Exam was divided into four sessions: text-dependent writing prompt, writing/language, listening, and reading. Students were able to take the sessions in any order. Recommended testing times for all sessions were included in the test design document as well as in the test administration manual.

### 3.2.2 Mathematics

Table 3-7 shows the Mathematics test design, including the number of items and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the sixteen test forms at each grade level. The unique items in each test form were field test items.

The Mathematics section of the exam was divided into two testing sessions, with students able to take the sessions in either order. In grades 3–5, no calculator was allowed for any of the Mathematics items. In grades 6–8, no calculator was allowed for the first session and the second session allowed students to use an embedded calculator. Recommended testing times for both sessions were included in the test design document as well as in the test administration manual.

### 3.2.3 Science

Table 3-8 shows the Science test design, including the number of items and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the twenty test forms at each grade level. The unique items in each test form were field test items.

The Science section of the exam was divided into three testing sessions, with students being allowed to take the sessions in any order. Recommended testing times for both sessions were included in the test design document as well as in the test administration manual.

### 3.2.4 Social Studies

Table 3-9 shows the Social Studies test design, including the number of items and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the six to nine test forms at each grade level. The unique items were field test items. The Social Studies exam included two test sessions that could be administered in either order. The Social Studies exams at grades 4, 8, and 10 included custom items developed specifically for the Wisconsin Forward Exam. Recommended testing times for both sessions were included in the test design document as well as in the test administration manual.

## 3.3 Universal Design

Assessments that are universally designed allow for the participation of the widest possible range of students, resulting in more valid inferences about student performance. Universally designed grade-level assessments may reduce the need for accommodations by reducing or eliminating access barriers associated with the tests themselves. Table 3-10 presents the elements of universal design that were implemented on the Wisconsin Forward Exam (Thompson & Thurlow, 2002).

These elements of universal design are relevant to both item development and form construction. This section addresses how the elements of universal design were addressed in the construction of the Spring 2022 test forms in compliance with AERA, APA, & NCME (2014) Standard 3.1, which states the following:

> Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (p. 63)

---

A goal of universal design is to measure the performance of students with a wide range of abilities and skills, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content. To accommodate the greatest number of students for the Wisconsin Forward Exam, the assessments include simple, clear, and intuitive instructions and procedures; maximum readability and comprehensibility; and maximum legibility. These design components are addressed primarily through the physical layout and formatting of the online test forms as well as the paper-based test forms used for accommodations. The page specifications define how directions and test items are placed on the pages, the location and appearance of headers and footers, the spacing between an item stem and the answer choices, and other page elements to ensure a consistent, legible appearance of online forms and paper-based test forms. Written instructions at the beginning of each test session are clearly and simply stated, and the wording of such instructions is standardized as much as possible across content areas and grade levels to ensure clarity and consistency.

> AERA, APA, & NCME (2014) Standard 3.9 states the following:
>
> Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs. (p. 67)

Students with disabilities or students who are English Learners may be provided with test administration accommodations based on their Individualized Education Programs (IEPs). Accommodation code definitions can be found in the Accessibility Guide available on the "Wisconsin Forward Exam Accommodations and Supports" page on DPI's website: https://dpi.wi.gov/assessment/forward/accommodations.

Braille test version was available for each grade and content area to enable students who are blind to participate in the Wisconsin Forward Exam testing. Braille forms for all grades and content areas were created by DRC test developers and consisted of the same items as those included in the regular operational online test forms. Specific recommendations on how to transcribe items into Braille were provided by an independent Braille expert who collaborated with the Braille publisher to produce the Braille version of the Wisconsin Forward Exam assessment and teacher's notes that accompany the Braille forms.

## 3.4 Item Development Process

As stated earlier in the report, the ELA, Mathematics, and Science test forms were originally developed for the Spring 2020 administration. Because of the Spring 2020 test cancellation, these test forms remained secure and were administered in Spring 2022. New test forms were developed for the Social Studies administration in 2022.

ELA, Mathematics, and Science test items included in the Spring 2022 Wisconsin Forward Exam were selected from DRC's College- and Career-Ready (CCR) item bank. DRC's CCR item bank contains nationally field-tested CCR items that support the next generation of standards and assessments. Items form the CCR bank are aligned to the College and Career Readiness standards in ELA and Mathematics grades 3–8. Science items are aligned to Wisconsin's Standards for Science and enhanced by the Next Generation Science Standards (NGSS) based on the National Research Council's Framework for K–12 Science Education. The item bank is designed to support states like Wisconsin that have adopted more rigorous content standards, curricula, and assessments that better prepare students for college and careers.

Alignment to standards, grade-level appropriateness, 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. DRC's item development process for the CCR item bank followed the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). DRC's item development work was and continues to be designed to produce reliable and instructionally valid tests that reflect the complete range of performance articulated in the AERA, APA, and NCME standards.

Furthermore, DRC's item development work adheres to the Principles of Universal Design (Thompson, Johnstone, & Thurlow, 2002) and reflects how items and tests must lend themselves to accessibility by diverse groups of students. Members of DRC's item development team have received direct training from the National Center on Educational Outcomes (NCEO). Therefore, DRC employs the Principles of Universal Design throughout all stages of both the item development process and the test development process.

All DRC's ELA, Mathematics, and Science items that appear on the Wisconsin Forward Exam were reviewed for content and for fairness not only by DRC's content experts but also by a panel of external experts and Wisconsin educators. The external reviewers have a broad range of experience in the educational field. All the reviewers have bachelor's-level, master's-level, or doctoral-level degrees and teaching experience in their specific area of expertise. Table 3-11 provides a high-level sequence of the activities that occurred in the development of the DRC CCR item bank.

Wisconsin-owned Social Studies items were developed by DRC internal item writers and contracted external consultant item writers. All item writers were required to hold a bachelor's degree or higher in the content area related to the subject for which they would be writing items, in curriculum and instruction, and/or in a related field. They also had to have three or more years of teaching experience in the content area for which they would be writing items and two or more years developing items to adhere to client specifications. DRC external item writers are trained annually at a series of content-focused writing training sessions. DRC item and test development staff prepared all materials for the item writing training session, including but not limited to Wisconsin Standards for Social Studies; guidelines for adhering to the Principles of Universal Design and other accessibility guidelines, including guidelines for English language learners; guidelines for freedom from issues of bias, fairness, and sensitivity; item specifications (including guidelines for writing items to cover a range of difficulty, a range of subject matter, and items focused on specific performance level descriptor alignments); item writing templates by item type, including scoring guidelines for those items scored using automatic scoring; and sample items. Social Studies item development was coordinated by DRC.

New Social Studies items underwent reviews by DRC content experts as well as DRC bias and sensitivity experts. All Social Studies items were also reviewed and approved by committees of Wisconsin educators. The efforts by DRC in developing items are in alignment with multiple best practices of the testing industry and, in particular, support the following AERA, APA, & NCME (2014) Standards:

> **Standard 3.1** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (p. 63)

**Standard 3.2** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (p. 64)

As stated earlier, Wisconsin licensed ELA, Mathematics, and Science items from DRC's CCR item bank. Due to the state-specific nature of the Social Studies standards, DPI owns the items for that content area. Details regarding the development of the items in the CCR bank created prior to their field-testing on the Forward Exam are provided in the *Wisconsin Forward Exam Spring 2016 Technical Report*, available on the DPI website at https://dpi.wi.gov.

### 3.4.1 Reading Passage and Item Reviews

The test items typically begin their life cycle two years prior to their operational administration. New ELA, Mathematics, Science, and Social Studies passages and/or items were first reviewed and approved for placement on the Wisconsin Forward Exam by both DPI and Wisconsin educators. For these reviews, educators from across the state convened in Madison, Wisconsin, to review items in an online format so that items could be evaluated in the same testing engine and style in which items are presented to students during the actual administration. ELA and Mathematics item reviews were held from July 31 to August 2, 2018; the Science item review was held from August 13 to 15, 2018. Social Studies item reviews were held from August 5 to 9, 2019. An example of the training PowerPoint presentation used at the reviews can be found in Appendix A of this report.

Items that were approved by the Wisconsin educators were then included in the next field test administration in Spring 2019 for ELA, Mathematics, and Science and in Spring 2021 for Social Studies. The purpose of the field tests was to build the pool of items eligible for inclusion in the subsequent operational test forms of the Forward Exam.

More information on the item and passage review for items later included in the Spring 2020 operational test forms for ELA, Mathematics and Science can be found in the *Wisconsin Forward Exam Spring 2020 Technical Report*, available on the DPI website at https://dpi.wi.gov. These forms were subsequently administered in Spring 2022.

## 3.5 Field-Testing

ELA, Mathematics, and Science items approved for the field test administration during the Summer 2018 item review were field-tested in Spring 2019 during the operational test administration. Field test items were fully embedded in the operational forms, and students were not able to distinguish between the operational and field test items. The field test items were embedded in several test forms administered in each grade and content area. Each test form contained the same operational test items and unique field test items. The test forms were spiraled at the student level within a grade and a content area. A total of 377 new items were field-tested for ELA. A total of 189 items were field-tested for Mathematics. A total of 240 items were field-tested for Science in Spring 2019. There was no field test administration in 2021 for these content areas.

Social Studies items measuring the new Wisconsin Standards for Social Studies were field-tested in Spring 2021. These items were embedded in the operational test forms. A total of 323 items were field-tested for Social Studies in Spring 2021.

The most recent field test of new items occurred in Spring 2022. A total of 811 items were field-tested for ELA, a total of 755 items were field-tested for Mathematics, a total of 187 items were field-tested for Science, and a total of 152 items were field-tested for Social Studies.

## 3.5.1 Statistical Analysis of Field Test Data

Following the field test data acquisition, the field test data analyses were conducted. The analyses included classical item analysis, differential item functioning (DIF) analysis, and item response theory (IRT). The classical item analysis included computation and evaluation of the following statistics: item $p$-values (difficulty), item-total test correlation, percentage of students selecting incorrect responses, point-biserial correlation for incorrect responses for the multiple choice (MC) items, score point distribution for items worth more than 1 point, and omit rates for all items. More details on classical item analysis methodology are provided in Part 6 of this report.

DIF was conducted for all field test items to examine potential item bias and to determine whether item performance differences between identifiable subgroups were due to factors other than student ability, making the items unfairly difficult for a particular subgroup in the student population. DIF analyses were conducted based on gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency (ELP) groups. More details on the DIF methodology are provided in Part 9 of this report.

As the last step of the field test data analysis, the field test items were calibrated and equated to operational test scales using the IRT methodology (explained in detail in Part 6 of this report). Note that ELA, Mathematics, and Science field test items were equated to their respective operational test scales in Spring 2019. All operational test items contained in the Spring 2019 operational test forms served as anchor items to place the field test items on the operational test scales using the Stocking and Lord (1983) equating procedure. Social Studies field test items were calibrated in Spring 2021.

The field test 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 that needed a closer review. For an item to be flagged for an additional review, the criteria included any of the following:

- $p$-value <0.20 or >0.90
- item-total test correlation (point biserial for MC items) <0.15
- positive point biserial on a distractor for an MC item
- omit rate >5%
- fewer than 5 students at any score point for multi-point items (EBSR and TDA item types in ELA)
- large differential item functioning (DIF) status

Items flagged for any of the above reasons were reviewed by the content area specialists prior to their review by DPI.

## 3.5.2 Item Data Review

In the preceding section, it was stated that test development content area specialists used certain statistics from item and DIF analyses of the 2019 field test for ELA, Mathematics, and Science and the 2021 field test for Social Studies to identify items for further review. Specific flagging criteria for this purpose were specified in the previous section. Items without statistical flags were regarded as statistically acceptable and were not included in the data review. Likewise, items of extremely poor statistical quality were regarded as unacceptable and needed no further review. Such items were excluded from the Wisconsin item pool prior to the data review with DPI. The remaining flagged items were regarded by DRC content area test development specialists and DRC psychometric specialists as needing further review. The intent was to capture all items that needed an additional review based on their statistical properties; thus, the criteria employed for item flagging tended to intentionally overidentify potential item issues.

The last review of the ELA, Mathematics, and Science field test items with data was conducted by DPI staff and DRC content specialists in July 2019 (for ELA and Science) and August 2019 (for Mathematics) in Madison, Wisconsin. A summary of the data review process and results, including the number of items that were field-tested, the number and percentage of items with statistical flags, and the number and percentage of items rejected by DPI during these reviews can be found in Table 3-13 in Part 3, "Test Content and Test Development," of the *Wisconsin Forward Exam Spring 2020 Technical Report*, available on the DPI website at https://dpi.wi.gov.

The review of the new Social Studies items with data was conducted in August 2021. This review was conducted online. As in past item review meetings, reviewers were first trained by a representative from DRC's staff 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 review process involved a brief exploration of possible reasons for the statistical profile of an item (e.g., possible bias, grade appropriateness, instructional issues) and a decision regarding acceptance. DRC content area test development specialists facilitated the review of the items. DPI reviewed the pool of field test items and made recommendations on each item and/or scenario/passage. The training presentation used at the Social Studies data review meeting may be found in Appendix B. A summary of the Social Studies data review results, including the number of items that were field-tested, the number and percentage of items with statistical flags, and the number and percentage of items rejected by DPI during the data review, is presented in Table 3-12. Social Studies items accepted for subsequent use in the Wisconsin Forward Exam were included in the pool of items for the Spring 2022 operational test form selection.

It is worth noting that the Social Studies field test was conducted in Spring 2021, and student performance on these items was likely negatively affected by the circumstances related to the COVID-19 pandemic and disrupted learning during the 2020–21 school year. Therefore, the statistical criteria for review and acceptance of these new items were more flexible than in a typical administration year. Items with statistics falling outside of the acceptable ranges could have been considered for inclusion in the item pool used for operational test form selection.

## 3.6 Form Development

The creation of test forms in a typical test development cycle involves the expertise of multiple DRC departments and DPI. The Wisconsin Forward Exam test development process complied with the following AERA, APA, & NCME (2014) Standards:

> **Standard 4.1** Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s). (p. 85)

> **Standard 4.7** The procedures used to develop, review, and try out items and to select items from the item pool should be documented. (p. 87)

> **Standard 4.12** Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications. (p. 89)

The DRC team works cooperatively with DPI content and assessment specialists to select passages and prompts with associated content-specific items for the online assessments. The DRC team constructs forms that comply with the approved test blueprints and form construction guidelines. DRC uses an integrated team approach to test development, which includes content area specialists, psychometricians, and scoring specialists working as a unit in collaboration with DPI content experts.

### 3.6.1 Wisconsin Forward Test Form Creation

As stated earlier in the report, the ELA, Mathematics, and Science operational test forms that were administered in Spring 2022 were originally developed for the Spring 2020 test administration. Except for the replacement of two items on the ELA grade 8 test between the original form selection and the Spring 2022 test administration, the operational portion of these assessments did not change. For details regarding the ELA, Mathematics, and Science form development and item selection process, refer to Part 3, "Test Content and Test Development," of the *Wisconsin Forward Exam Spring 2020 Technical Report*, available on the DPI website at https://dpi.wi.gov.

The items on the Spring 2022 Social Studies assessments for grades 4, 8, and 10 were selected from the pool of items measuring the Wisconsin Standards for Social Studies and field-tested in Spring 2021. The Social Studies assessments were not linked to the previous Social Studies scales, and no items from the previous operational Social Studies tests were included in the new assessments. The new Social Studies tests were not matched to the previous assessments in terms of statistical properties.

As a first step in building the new Social Studies assessments, the DRC team prepared all items that could be considered in the process in DRC's item banking system, which is called IDEAS. The form, format, extent, and organization of items in their respective test sessions were determined in consultation with DPI. Following the preparation of all necessary materials and resources, forms construction began. Construction of the test forms themselves was a collaborative effort within DRC's integrated development team of assessment specialists, psychometric services specialists, and scoring specialists.

Before test forms were created, passages, item/performance tasks, and artwork were carefully selected. The following process was used for item selection:

- DRC test development specialists first selected items to match the approved test blueprints.
- DRC test development specialists checked to see that each item clearly aligned with the standards and that each item, with available item statistics, met psychometric guidelines for inclusion in the test.
- DRC test development specialists verified that each item met technical quality standards for well-crafted items, including that each item
  - had one clearly correct answer (or answers if the item was multi-select);
  - used clear and concise wording;
  - was grammatically correct;
  - had an appropriate range of difficulty;
  - was free of any offensive, inappropriate, or biased content; and
  - met the Principles of Universal Design and maximum accessibility.

In addition to content requirements, the following statistical criteria were used in item selection:

- Test length and item types match the DPI-approved test design.
- Content coverage matches the DPI-approved test blueprint.
- Items that meet the following criteria are avoided whenever possible:
  - *p*-value $\leq 0.20$ or $\geq 0.90$
  - Item-total test correlation $<0.15$
  - Omit rates $\geq 3\%$
  - Poor item fit statistics (misfit flag)
  - Significant DIF statistics—If an item with DIF had to be included in the test to maintain blueprint coverage, the item was examined to determine whether any content reason exists for the DIF flag (sometimes items demonstrate statistical bias, but no content reason can be determined for the bias).

New field test items were added to each test form in each grade and content area. In constructing the final forms, the DRC content area test development specialists followed the guidelines provided below:

- Forms included adequate standards coverage, as required by test blueprints.
- No item in a form "clued" another item on that same form.
- Forms were diverse in terms of artwork and graphics.
- Forms included a wide range of topics and a variety of questions.
- Correct answer distributions were reasonable across MC items on the form.
- Forms did not contain any items that had been released to the public.
- DPI reviewed and gave final approval of all online test forms.

The test maps in Appendices C, D, E, and F provide details on the operational items on the Spring 2022 Wisconsin Forward Exam per grade and content area. The test maps include the session number, item sequence, item type, item usage, item maximum score, depth-of-knowledge level, standard code, and domain name. The ELA test maps are included in Appendix C, the Mathematics test maps are contained in Appendix D, the Science test maps are provided in Appendix E, and the Social Studies test maps are given in Appendix F.

## 3.6.2 Item and Form Quality Reviews

In all phases of the item and form development process, content area test development specialists and editorial specialists reviewed items and passages for technical quality; alignment with the standards; issues of bias, fairness, and sensitivity; depth of knowledge; estimated difficulty; and adherence to the Principles of Universal Design in all steps of the forms creation and forms review processes. The aim for this team approach was to conduct a multitiered internal review of all passages and items prior to submission for review by DPI and then, with approval by DPI prior to submission, for review by Wisconsin educators to ensure that all items align with Wisconsin's standards and adhere to DPI's standards for high-quality items.

DRC content and editorial teams reviewed all passages and items to ensure that they possessed the following traits:

- content alignment or congruence with the knowledge and skills specified in the standards;
- a range of estimated difficulty levels;
- appropriate grade-level vocabulary, subject matter, and assumed student knowledge;
- freedom from issues or concerns regarding bias, sensitivity, or fairness;
- accessibility, following the Principles of Universal Design; and
- correct grammar, usage, and structure/format.

As a part of DRC's internal review of the items and test forms, the test development team members and graphic specialists ensured that item art could be reproduced clearly and accurately when electronically displayed and when used in the print-on-demand forms.

Test specifications were reviewed to identify any potential display requirements that may present challenges in an electronic display environment. Display tolerances are impacted by line thickness, percentage of screening for shading, specialized fonts and symbols, photographs, and color. These are defined in the early stages of the item and test development process to help guide the delineation of style requirements and specifications.

Item art was produced using transparent vector graphics that allow for adjustments without the breakdown of image clarity, which is common with lower-quality formats, and that provide for the online accommodation of alternate background colors. The DRC multitiered quality assurance process made certain that converted item art was carefully compared to the original format throughout the test development and production process.

In reviewing forms in the online environment, multiple reviewers checked passages and items on the multiple electronic platforms on which students took the test to ensure a smooth testing experience.

## 3.7 DPI Approvals

DPI had the opportunity to review passages and items placed on the Spring 2022 Wisconsin Forward Exam during the following phases.

ELA, Mathematics, and Science

- prior to item content review in Summer 2018
- at item content review in Summer 2018
- during review of flagged field test data in Summer 2019
- during the form construction process in Fall 2019
- before the Spring 2022 test administration

Social Studies

- prior to item content review in Summer 2019
- at item content review in Summer 2019
- during review of flagged field test data in Summer 2021
- during the form construction process in Fall 2021
- before the Spring 2022 test administration

Prior to the opening of the testing window, all online forms were made accessible to DPI for review in DRC's secure INSIGHT testing engine.

## 3.8 Summary

In summary, the Spring 2022 Wisconsin Forward Exam assessment adhered to the Wisconsin test blueprints and test designs for each grade level and content area. The items included in the Spring 2022 Wisconsin Forward Exam were reviewed by DRC, DPI, and Wisconsin educators for issues regarding accessibility, bias, sensitivity, and content. During the reviews, experts identified (1) issues that could negatively affect a student's ability to access stimuli and items, (2) content in stimuli and items that could unfairly affect a student's response because of the student's background, (3) developmental appropriateness, and (4) the alignment of stimuli and items to the content specifications. Item content was checked for the accuracy of the content, answer keys, and scoring rules. Following Spring 2019 and Spring 2021 field-testing, items flagged for accessibility, bias and sensitivity, and/or other content concerns were further reviewed by DRC and DPI to determine whether these flagged items should be removed from the Wisconsin item pool prior to the form construction of the Wisconsin Forward Exam. The efforts and procedures used in the development of the Spring 2022 Wisconsin Forward Exam forms balanced the content and psychometric requirements for the form development. The psychometric properties of the ELA, Mathematics, and Science test forms were comparable to the psychometric properties of the Spring 2019 (target) forms. Such comparability was not required for Social Studies assessments, which measured the new Wisconsin Standards for Social Studies and for which new scales were developed after the Spring 2022 administration. Overall, the process implemented in the Spring 2022 operational form development was in alignment with multiple best practices of the testing industry.

**Table 3-1. Test Blueprints for English Language Arts Grades 3–8**

| Domain (Reporting Category) | Depth of Knowledge | Total Points by Grade | | | | | |
|---|---|---|---|---|---|---|---|
| | | 3 | 4 | 5 | 6 | 7 | 8 |
| **Reading** | | 22 | 24 | 24 | 24 | 24 | 24 |
| Key Ideas and Details | grade 3: 1–3 grades 4–8: 2–3 | 6–12 | 6–12 | 6–12 | 6–12 | 6–12 | 6–12 |
| Craft and Structure/Integration of Knowledge and Ideas | all grades: 2–3 | 4–10 | 4–10 | 4–10 | 4–10 | 4–10 | 4–10 |
| Vocabulary Use—Includes Language Standards 4 and 5 | grades 3–5: 1–3 grades 6–8: 2–3 | 4–6 | 4–6 | 4–6 | 4–6 | 4–6 | 4–6 |
| Literature | | about 60% | about 60% | about 60% | about 50% | about 50% | about 50% |
| Informational Text | | about 40% | about 40% | about 40% | about 50% | about 50% | about 50% |
| **Writing/Language** | | 24 | 24 | 24 | 24 | 24 | 24 |
| Text Types and Purposes/ Text-Dependent Analysis | all grades: 2–3 | 10–14 | 10–14 | 10–14 | 10–14 | 10–14 | 10–14 |
| Research | all grades: 2–3 | 6–8 | 6–8 | 6–8 | 6–8 | 6–8 | 6–8 |
| Language Conventions | all grades: 1–3 | 6–8 | 6–8 | 6–8 | 6–8 | 6–8 | 6–8 |
| **Listening** | all grades: 2–3 | 7 | 8 | 8 | 8 | 8 | 8 |
| **ELA Points Total** | | **53** | **56** | **56** | **56** | **56** | **56** |

**Table 3-2. Test Blueprints for Mathematics Grades 3–8**

| Domain (Reporting Category) | Depth of Knowledge | Total Points by Grade | | | | | |
|---|---|---|---|---|---|---|---|
| | | 3 | 4 | 5 | 6 | 7 | 8 |
| Operations and Algebraic Thinking | grade 3: 1–3 grades 4–5: 1–2 | 8–10 | 9–11 | 8–10 | | | |
| Number and Operations in Base Ten | grades 3–5: 1–3 | 7–9 | 8–10 | 8–10 | | | |
| Number and Operations—Fractions | grades 3–5: 1–3 | 7–9 | 9–11 | 8–10 | | | |
| Measurement and Data | grades 3–5: 1–3 | 9–11 | 9–11 | 9–11 | | | |
| Geometry | grades 3–4: 1–2 grades 5–8: 1–3 | 6–8 | 6–8 | 8–10 | 6–8 | 9–11 | 9–11 |
| Ratios and Proportional Relationships | grades 6–7: 1–3 | | | | 6–8 | 7–9 | |
| The Number System | grades 6–7: 1–3 grade 8: 1–2 | | | | 10–12 | 6–8 | 7–9 |
| Expressions and Equations | grades 6, 8: 1–3 grade 7: 1–2 | | | | 10–12 | 9–11 | 9–11 |
| Statistics and Probability | grade 6: 1–2 grades 7–8: 1–3 | | | | 9–11 | 10–12 | 7–9 |
| Functions | grade 8: 1–3 | | | | | | 9–11 |
| **Mathematics Points Total** | | **42** | **46** | **46** | **46** | **46** | **46** |

**Table 3-3. Test Blueprints for Science Grades 4 and 8**

| Domain (Reporting Category) | Depth of Knowledge | Total Points by Grade | |
|---|---|---|---|
| | | 4 | 8 |
| Practices and Crosscutting Concepts in Life Science | grades 4, 8: 2–3 | 8–12 | 8–12 |
| Practices and Crosscutting Concepts in Physical Science | grades 4, 8: 2–3 | 8–12 | 8–12 |
| Practices and Crosscutting Concepts in Earth and Space Science | grades 4, 8: 2–3 | 8–12 | 8–12 |
| Practices and Crosscutting Concepts in Engineering | grades 4, 8: 2–3 | 8–12 | 8–12 |
| **Science Total Points** | | **40** | **40** |

**Table 3-4. Test Blueprints for Social Studies Grades 4, 8, and 10**

| Domain (Reporting Category) | Depth of Knowledge | Total Points by Grade | | |
|---|---|---|---|---|
| | | 4 | 8 | 10 |
| Geography (Inquiry Practices and Processes) | grade 4: 1–3 grade 8: 2–3 grade 10: 2 | 8–12 | 8–12 | 8–10 |
| History (Inquiry Practices and Processes) | all grades: 2–3 | 8–12 | 8–12 | 8–10 |
| Political Science (Inquiry Practices and Processes) | grade 4: 2–3 grade 8: 1–3 grade 10: 2 | 6–8 | 6–8 | 8–10 |
| Economics (Inquiry Practices and Processes) | all grades: 2–3 | 6–8 | 6–8 | 6–8 |
| Behavioral Sciences (Inquiry Practices and Processes) | all grades: 2–3 | 6–8 | 6–8 | 6–8 |
| **Social Studies Total Points** | | **40** | **40** | **40** |

# Table 3-5. Item Type Descriptions for Items on the Wisconsin Forward Exam

| Item Type | Name | Description |
|---|---|---|
| **EBSR** | Evidence-Based Selected Response | Each evidence-based selected response item has two parts, and each two-part item is designed to elicit an evidence-based response from a student who has read a literature text passage, an informational text passage, or a writing concept. In part one, which is similar to a multiple-choice item, the student analyzes a passage or writing concept and chooses the best answer from four response options. In part two, the student uses evidence from the passage or writing concept to select one or more answers based on the response to part one. EBSR items can worth one or two points. |
| **MC** | Multiple Choice | Each multiple-choice item has four response options, only one of which is correct. Multiple-choice items are used to assess a variety of skill levels, from short-term recall of information to inference and problem-solving. Each of these items is worth one point. |
| **MS** | Multiple Select | Each multiple-select item requires a student to evaluate information presented and respond by choosing two or more correct responses. Multiple-select items can be used to assess multiple skills and concepts in a given content area. MS items can worth one or two points. |
| **SA** | Short Answer | Each short-answer item requires a student to enter a short numeric or algebraic response. These items are designed to assess a student's ability to formulate a solution to a pure or applied math problem without the assistance of response options. The short-answer items are scored on a 0–1-point scale using item-specific autoscoring rules. |
| **TDA** | Text- Dependent Analysis | Each text-dependent analysis item is a text-based analysis based on a passage or a multiple-passage set that each student has read during the assessment. Both literary and informational texts are addressed through this item type. Students must draw on basic writing skills while inferring and synthesizing information from the passage in order to develop a comprehensive, holistic essay response. The demands required of a student's reading and writing skills in response to a TDA item coincide with the similar demands required for a student to be college and career ready. The TDA prompts are scored using a holistic scoring guideline on a 1–4 point scale. A weight of 2 is applied to the item scores in the computation of the student total test raw scores and scale scores. That is, the TDA prompts contribute up to 8 raw score points toward the student total test raw score. This item type is supported by all Wisconsin ELA standards across all grades for both Reading Literature and Reading Informational Texts and by Writing standards 1, 2, 3, 4, and 9 across all grades. The TDA items are scored using artificial intelligence (AI) scoring, with an appropriate level of human scoring to validate the AI algorithms for all TDA items used in the Wisconsin ELA grades 3–8 assessments. |
| **TE** | Technology Enhanced | Each technology-enhanced item is designed to elicit evidence of a broad range of student understanding. A student interacts with the enhanced features of these computer-delivered, autoscorable test items to show understanding of skills and concepts. Item types such as drag-and-drop, hot-spot, number line and coordinate graphing, data displays, matching interaction, and drop-down menus are just some of the technology-enhanced items presented to a student. TE items can worth one or two points. |

## Table 3-6. Test Design for English Language Arts

| Test Design | | Grade | | | | | |
|---|---|---|---|---|---|---|---|
| | | **3** | **4** | **5** | **6** | **7** | **8** |
| **Number of Passage Sets** | Literature | 2 | 2 | 2 | 2 | 2 | 2 |
| | Informational | 1 | 2 | 3 | 2 | 2 | 2 |
| | Listening | 2 | 2 | 2 | 2 | 2 | 2 |
| **Number of Core (OP) Items** | Item Types: MC/TE (1 pt) | 27 | 28 | 30 | 24 | 22 | 28 |
| | Item Types: MS/TE/EBSR (2 pts) | 9 | 10 | 9 | 12 | 13 | 10 |
| | Item Type: TDA (4 pts x 2) | 1 | 1 | 1 | 1 | 1 | 1 |
| | **Total Core Items** | 37 | 39 | 40 | 37 | 36 | 39 |
| **Total Core Points** | | 53 | 53 | 56 | 56 | 56 | 56 |
| **Embedded Field Test (FT)** | **Number of Forms** | 16 | 16 | 16 | 16 | 16 | 16 |
| | **Passages (Reading + Listening)** | 2 | 2 | 2 | 2 | 2 | 2 |
| | **FT Items per Form** | 10 | 10 | 9 | 8 | 8 | 8 |
| | **Total Field Test Items** | 144 | 144 | 144 | 128 | 127 | 124 |
| **Total Items (Core + FT) per Form** | | 47 | 49 | 49 | 45 | 44 | 47 |
| **Total Estimated Testing Time (minutes)** | | 130 | 130 | 130 | 130 | 130 | 130 |

Note: TDA items are scored using a 1–4 point scoring rubric. A weight of 2 is applied to item scores in the computation of the student total test raw scores and scale scores.

## Table 3-7. Test Design for Mathematics

| Test Design | | Grade | | | | | |
|---|---|---|---|---|---|---|---|
| | | **3** | **4** | **5** | **6** | **7** | **8** |
| **Number of Core (OP) Items** | Item Types: MC/SA(1 pt) | 35 | 40 | 40 | 38 | 39 | 37 |
| | Item Type: TE (1 pt) | 7 | 6 | 6 | 8 | 7 | 9 |
| | **Total Core Items** | 42 | 46 | 46 | 46 | 46 | 46 |
| **Total Core Points** | | 42 | 46 | 46 | 46 | 46 | 46 |
| **Embedded Field Test (FT)** | **Number of Forms** | 16 | 16 | 16 | 16 | 16 | 16 |
| | **FT Items per Form** | 8 | 8 | 8 | 8 | 8 | 8 |
| | **Total Field Test Items** | 128 | 128 | 127 | 116 | 128 | 128 |
| **Total Items (Core + FT) per Form** | | 50 | 54 | 54 | 54 | 54 | 54 |
| **Total Estimated Testing Time (minutes)** | | 90 | 90 | 90 | 105 | 105 | 115 |

## Table 3-8. Test Design for Science

| Test Design | | Grade | |
|---|---|---|---|
| | | **4** | **8** |
| **Number of Core (OP) Items** | Item Types: MC/MS/TE/EBSR (1 pt) | 40 | 40 |
| **Total Core Points** | | 40 | 40 |
| **Embedded Field Test (FT)** | Number of Forms | 20 | 20 |
| | Scenarios/Tasks | 10 | 10 |
| | FT Items per Form | 5 | 5 |
| | Total Field Test Items | 94 | 93 |
| **Total Items (Core + FT) per Form** | | 45 | 45 |
| **Total Estimated Testing Time (minutes)** | | 105 | 105 |

## Table 3-9. Test Design for Social Studies

| Test Design | | Grade | | |
|---|---|---|---|---|
| | | **4** | **8** | **10** |
| **Number of Core (OP) Items** | Item Types: MC/TE/MS/EBSR (1 pt) | 40 | 40 | 40 |
| **Total Core Points** | | 40 | 40 | 40 |
| **Embedded Field Test (FT)** | Number of Forms | 6 | 9 | 8 |
| | FT Items per Form | 8 | 8 | 10 |
| | Total Field Test Items | 37 | 62 | 53 |
| **Total Items (Core + FT) per Form** | | 48 | 48 | 50 |
| **Total Estimated Testing Time (minutes)** | | 70 | 70 | 70 |

## Table 3-10. Elements of Universal Design

| Element | Explanation |
|---|---|
| Inclusive Assessment Population | Tests designed for state, district, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field-testing procedures. |
| Precisely Defined Constructs | The specific constructs tested must be clearly defined so that all construct-irrelevant cognitive, sensory, emotional, and physical barriers can be removed. |
| Accessible, Unbiased Items | Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items. |
| Amenable to Accommodations | The test design facilitates the use of needed accommodations. |
| Simple, Clear, and Intuitive Instructions and Procedures | All instructions and procedures are simple, clear, and presented in understandable language. |
| Maximum Readability and Comprehensibility | Readability and plain language guidelines are followed (e.g., sentence length and number of difficult words are kept to a minimum) to produce readable and comprehensible text. |
| Maximum Legibility | Characteristics that ensure easy decipherability are applied to text, tables, figures, illustrations, and response formats. |

**Table 3-11. College- and Career-Ready Item Bank Development Activities**

| DRC College- and Career-Ready Item Bank Development Activities |
| --- |
| Establish item/passage development specifications and style guides and prepare item writing training manuals. |
| Determine item development plans. |
| Train item writers and/or passage developers in the project requirements and specifications. |
| Develop passages and write items. |
| Review, edit, code, and track items and produce graphics. |
| Produce review forms for content and bias/fairness/sensitivity reviews by external reviewers. |
| Modify items based on external reviewers' recommendations. |
| Review and approve field test–ready items and passages. |
| Develop field test forms and administer field test. |
| Internally review field test item data. |
| Approve items to be included in the item bank. |

**Table 3-12. Items Reviewed at Item and Data Reviews for Social Studies**

| Grade | Number of Items Reviewed at Item Review in Summer 2019 | Number of Items Field Tested in 2021 | Field Test Items Flagged for Poor Statistics or DIF | | Field Test Items Rejected at Data Review for Statistical or Content-Related Reasons | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | Number of Items | Percentage of All Field Test Items | Number of Items | Percentage of All Field Test Items |
| **4** | 149 | 120 | 25 | 21% | 18 | 15% |
| **8** | 139 | 102 | 32 | 31% | 11 | 11% |
| **10** | 149 | 101 | 23 | 23% | 17 | 17% |

# Part 4: Test Administration

In the Spring of 2022, Wisconsin administered assessments in ELA and Mathematics for grades 3–8. Science was administered in grades 4 and 8, and Social Studies was administered in grades 4, 8, and 10. The test administration window was March 21–April 29, 2022. Part 4 of the Technical Report provides information on student participation rates in the Spring 2022 assessments and describes a set of standardized procedures and policies applied to administer the Wisconsin Forward Exam. The issue of test security in test administration, which has important implications for the integrity of the results and, thus, the validity of Wisconsin Forward Exam scores, is also discussed. Documentation citing the written procedures provided to test administrators and school personnel to standardize the administration of the test is provided in this part as well. The following AERA, APA, & NCME (2014) Standards are addressed in Part 4: 3.4, 3.5, 4.15, 4.16, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7. Each standard will be explicated within the relevant section of this part of the report.

DPI is committed to the proposition that all schools and all students will be held accountable to a common set of high academic content standards, the Wisconsin Academic Standards. As an alternate assessment for students being instructed using alternate academic achievement standards, the Wisconsin Essential Elements, the Dynamic Learning Maps assessment measures the academic progress of students with the most significant cognitive disabilities in the subject areas of ELA and Mathematics at grades 3–11 and Science at grades 4 and 8–11. A teacher rater form is used to assess these students in Social Studies at grades 4, 8, and 10.

All other students are accountable to the grade-level knowledge and skills outlined in the Wisconsin Academic Standards. Those students who have an IEP, a 504 plan (under Section 504 of the Rehabilitation Act of 1973) or are identified as limited English proficient (LEP) or formerly LEP may be eligible to receive testing accommodations or supports. Accommodations and supports are practices and procedures that provide equitable access to grade-level content. They are intended to reduce or eliminate the effects of a student's disability or level of language acquisition; they do not reduce learning expectations. DPI guidance makes it clear that the accommodations or supports provided to a student must be consistent with classroom instruction, classroom assessments, and district and state assessments. It is important to note that while some accommodations or supports may be appropriate for instructional use, they may not be appropriate for use on a standardized assessment. AERA, APA, & NCME (2014) Standard 6.2 states the following:

> When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing. (p. 115)

An overview of the types of accommodations and supports available to students and the guidelines for test administration conditions are described below. Additionally, IEP teams were directed to the Wisconsin Forward Exam Accommodations and Supports page at http://dpi.wi.gov/assessment/forward/accommodations for guidance regarding all available accommodations and supports intended to provide equitable access to grade-level content and assessments.

District Assessment Coordinators (DACs) indicated which accommodations and supports were to be available for use by each student within the student learning profile in the DRC INSIGHT Portal. All student accommodations and supports are managed and can be monitored through the DRC INSIGHT Portal. This system is the interface to the administrative functions of the DRC INSIGHT Online Learning System, where students interface with their online assessments. As a function of this roles-based system,

the primary users of the DRC INSIGHT Portal were DACs and School Assessment Coordinators (SACs) who were assigned permissions accordingly for security purposes. The major functions are those of managing users and managing students. As such, the DRC INSIGHT Portal was used to manage and update student information, including demographic and accommodations/accessibilities information. All DRC INSIGHT Portal user roles and permission levels were approved by DPI.

## 4.1 Student Participation

For the purposes of this report, the test participation rate is defined as the percentage of students who received a valid scale score given the total number of students who were scheduled to take the test. The test participation rates for the students in grades 3 through 7 ranged from approximately 95% to 96% across content areas. The test participation rates for the students in grade 8 were approximately 94% across all content areas. These participation rates were comparable to the pre-pandemic participation rates, which were typically at least 95% or higher. The participation rate in Social Studies grade 10 was approximately 88%. The Spring 2022 test participation rate for Social Studies grade 10 was lower than participation rates observed in typical administration years before the COVID-19 pandemic. Typically, at least 94% of students take the Social Studies test in grade 10.

When test participation rates by demographic groups were considered, it was observed that students from ethnic minority groups (African American, Hispanic, and American Indian) and students from historically disadvantaged groups (limited English proficiency or economically disadvantaged students) participated in the assessment at slightly lower rates compared to White students, fully English proficient students, or students who were not economically disadvantaged. These differences ranged from less than 1% to about 5% across all grades, content areas, and groups. Larger differences in test participation rates were found for groups of students with and without disabilities, with less than 90% of students with disabilities participating in the assessments compared to over 95% of their peers without disabilities in grades 3 through 8. This discrepancy was larger in Social Studies grade 10, where 74% of students with disabilities participated in the assessment versus 90% of students without disabilities. The participation rates were about the same for male and female students and for White and Asian/Pacific Islander students across all grades and content areas. Detailed information on the test participation rates in Spring 2022 for all students and disaggregated by demographic characteristics is provided in Appendix G.

## 4.2 Standardized Test Administration

Unstandardized testing conditions can pose a serious threat to test validity by adding construct-irrelevant variance to the test scores. McCallin (2006) described a number of such threats to validity, including alterations in test administration requirements (e.g., changing time limits, modifying test instructions, giving hints to examinees), variability across test sites (e.g., differences in facilities/equipment, inadvertent posting of instructional aids in classrooms), interruptions during test sessions (e.g., power outages, relocation of students during testing, disturbances, other distractions), test administrator practices that may exacerbate test anxiety in particular students, practices that elicit test wiseness, and security breaches that may result in the exposure of test forms or items. Construct-irrelevant variance may exert a systematic effect on the scores of individual students or groups of students, resulting in an overestimation or underestimation of their true abilities.

Standardized test administration, extensive training of the test scorers and artificial intelligence (AI) engine, and rigorous scoring rules for autoscored items for the Wisconsin Forward Exam comply with AERA, APA, & NCME (2014) Standards 3.4 and 3.5.

**Standard 3.4** Test takers should receive comparable treatment during the test administration and scoring process. (p. 65)

**Standard 3.5** Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population. (p. 65)

The standardized Wisconsin Forward Exam test administration procedures described in this part of this report were designed to address these potential threats to validity through the use of comprehensive security measures and the provision of detailed Test Administration Manuals and other training materials for DACs, SACs, and TAs.

## 4.3 Accessibility Resources

Accommodations were allowed for eligible individual students participating in the Wisconsin Forward Exam. Accommodations provided to a student must be documented in a current IEP and used during routine instruction. IEP teams were directed to refer to the Wisconsin Forward Exam accommodations policy and guidance at https://dpi.wi.gov/assessment/forward/accommodations.

It is important to note that students were provided access to a range of supports that included universal tools (available to all students), designated supports, and accommodations, including the Braille version of the Wisconsin Forward Exam, based on students' needs. Those supports are defined as follows.

### 4.3.1 Universal Tools

Universal tools are accessibility features that are available to all students based on student preference and selection. These accessibility features of the assessment are either provided as digitally delivered components of the test administration system (embedded) or separate from it (non-embedded).

**Embedded Universal Tools (Online)**

- Calculators
- Click to Enlarge
- Cross-Off Tools
- Flag/Mark for Review
- Help/What's This?
- Highlighter
- Go to Question
- Keyboard Navigation
- Line Guide
- Magnifier Tool (Zoom)
- Measuring Tools
- Pause (Breaks)
- Review Page
- Sticky Notes (Digital Notepad)

- Test Directions
- Tool Tips
- Writer's Checklist (ELA TDA Session only)

Non-embedded Universal Tools (Standard)

- Graph Paper
- Scratch Paper
- Writer's Checklist (ELA TDA Session only)

## 4.3.2 Designated Supports

Designated supports are those features that are available for use by any student for whom the need has been indicated by an educator or team of educators (with parent/guardian and student input as appropriate) and are part of the student's classroom instruction. They are either provided as part of the online test administration system or separate from it (i.e., embedded or non embedded). All embedded and non-embedded designated supports must be entered into the DRC INSIGHT Portal prior to test administration. Embedded and non-embedded supports will appear on student test tickets.

### Embedded Designated Supports (Online)

- Color Choices (CC)
- Contrasting Color (CTC)
- Reverse Contrast (RC)
- Masking (MSK)
- Stacked Translations (Spanish)
- Text-to-Speech (TTS)

### Non-embedded Designated Supports (Standard)

- Amplification Device
- Word-to-Word Bilingual Dictionary
- Color Overlay
- Magnification
- Noise Buffers
- Read Aloud in English
- Read Aloud in Spanish
- Scribe
- Separate Setting
- Small Group Translation
- Translator/Interpreter

### 4.3.3 Accommodations

Accommodations are features that increase equitable access but do not compromise the grade-level standard or intended outcome of the assessment. They are available for students for whom there is a documented need in the IEP or 504 accommodation plan and who use a similar accommodation as part of their classroom instruction. Accommodations are either provided as part of the online test administration system or separate from it (i.e., embedded or non-embedded). All embedded and non-embedded accommodations must be entered into the DRC INSIGHT Portal prior to test administration. Embedded and non-embedded accommodations will appear on student test tickets.

**Embedded Accommodations (Online)**

- Video Sign Language (VSL)
- Closed-Captioning (C CAP)

**Non-embedded Accommodations (Standard)**

- Abacus
- Alternate Response Options
- Braille (Unified English Braille) (BRL)
- Calculator
- Listening Scripts (LS)
- Multiplication Table
- Print-on-Demand (POD)
- Read Aloud (Reading Passages)

### 4.3.4 Translation

For the Spring 2022 Wisconsin Forward Exam administration, the State of Wisconsin used an embedded stacked Spanish translation for Mathematics, Science, and Social Studies items. For ELA assessments, only the test directions are available in stacked translation. The stacked Spanish translation is a designated support for students who are native Spanish speakers and are identified as limited English proficient to demonstrate their knowledge on the Wisconsin Forward Exam. In addition to the embedded stacked translation, bilingual word lists and a translation of the test directions are allowable designated supports.

DPI recognizes that approximately five percent of the Wisconsin limited English proficient population speaks a language other than Spanish, and specific guidelines are provided for these students. Districts that serve students who speak languages other than Spanish may have used qualified translators to provide oral translation support to students. However, the use of translation support was restricted to Mathematics, Science, and Social Studies tests, given that the test constructs are not specific to the English language. DPI recommended that educators consult the list of allowable accommodations and supports (referenced above) to create the most appropriate testing situation for their students.

### 4.3.5 Additional Accessibility Resources

Additional accessibility resources and guidance included the following:

- **Multiplication Table:** This resource is a non-embedded accommodation available for students who have it in their IEP or 504 plan for grades 4–8 Mathematics.

- **Read Aloud Guidelines:** This document outlines the qualifications, guidelines, and procedures required for a test reader. The test reader must sign the Read Aloud Agreement to Maintain Security and Confidentiality prior to test administration. Completed agreement forms should be retained by the Site Assessment Coordinator.

- **Scribing Guidelines:** This document outlines the qualifications, guidelines, and procedures required when using a scribe.

- **Interpreter Guidelines:** This document outlines the qualifications, guidelines, and procedures required when using an interpreter.

Tables 4-1 through 4-7 provide the list of accommodations or designated supports made available for the Spring 2022 Wisconsin Forward Exam along with the number and percentage of students provided these accommodations or supports. The counts are based on the accommodations and designated supports selected via the DRC INSIGHT Portal. Scores of assessments taken with accommodations were included with the results for students who took these tests under standard conditions and presented at the school, district, and state levels.

## 4.4 Test Security

Maintaining the security of all test materials is crucial to preventing the possibility of random or systematic errors, such as an unauthorized exposure of test items that would affect the valid interpretation of test scores. Several test security measures have been implemented for the Wisconsin Forward Exam with compliance to the following AERA, APA, & NCME (2014) Standards:

> **Standard 6.6** Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means. (p. 116)

> **Standard 6.7** Test users have the responsibility of protecting the security of test materials at all times. (p. 117)

The primary goal of test security is to protect the integrity of the assessments and ensure that scores retain their interpretability. To ensure that trends in achievement results can be calculated across years and to provide longitudinal data, a certain number of test questions must be repeated from year to year. If any of these questions are made public, the validity of the test may be compromised. Because the Wisconsin Forward Exam is administered virtually 100 percent online, printed test materials are limited to the very few cases where a student requires a printed version of the test as provided in the IEP (i.e., Braille and Print-on-Demand), so the assessment exposure is limited to those educators who require access for those purposes. DPI and DRC ensured that all who had access to any materials associated with the Wisconsin Forward Exam understood the critical need for test security. They presented security requirements during the pre-test workshops and outlined the acceptable and unacceptable test preparation and administration practices. The Wisconsin Forward Exam was administered under secure testing conditions established by DPI.

Other security measures for Wisconsin Forward Exam test administrations are described below:

- The use of any unauthorized electronic device is prohibited during testing.
- Password-protected, role-based administrator access to all test setup, management, and reporting functions is required.
- Student Test Login Tickets provide secure student access to the test using a unique username and password.
- Test content is securely transferred using leading encryption technologies; content is decrypted when the student login is validated.
- Decrypted test content is purged from the system's memory upon completion of the test session.
- Device lockdown during testing prevents students from copying, pasting, printing, and accessing other applications.
- If the test is paused, content is removed from the screen to ensure security of test content. The system will time out and close the test after a defined period of inactivity.
- Extensive software quality assurance tests ensure that all data are scanned, captured, and accurately scored in the secure database and all associated reports contain accurate data.

The online systems provided by DRC that are associated with the administration of the Wisconsin Forward Exam have all been designed to provide the level of security required by DPI and described in the DPI Test Security Manual for its assessment programs. Student testing environments are designed to ensure the protection of responses as well as student data (as required under the federal Family Educational Rights and Privacy Act). DRC's information security policies and procedures are based on the National Institute of Standards and Technology (NIST) criteria (NIST Standard 800-53). This is a nationally recognized standard for information security practices.

## 4.4.1 Secure Student Access

Students are required to provide a valid username and password to access the online testing system. The Test Administrator (TA) provides each student with a Student Test Login Ticket, which contains the student's username and a unique, pre-generated password. A separate, unique password is generated for each assessment, ensuring that students can only access the content designated for that particular test. Passwords are generated randomly for each student to use. Test tickets are generated from within the DRC INSIGHT Portal secure administrative system, which is pre populated with student records. As an additional security measure, after a student logs in, a Student Verification Page prompts the student to verify their profile information, including any assigned accommodations, prior to initiating the test. The student's name is also displayed on the screen during the test, providing an additional verification check for the student and the TA.

Test tickets and rosters are considered secure materials. Therefore, it is recommended that test tickets be printed as close to the date of testing as possible, and sites are instructed to keep test tickets and rosters in a secure location until the session is scheduled to begin. Test tickets are distributed just prior to students logging in and are collected after all students have logged in and begun testing; directions also include a request to count the number of tickets that are distributed and collected after students log in to make sure the numbers of tickets are the same. After a testing session is complete, all test tickets are returned to the Site Assessment Coordinator for secure destruction or secure storage.

### 4.4.2 Test Security during Breaks

Test security must be maintained during all breaks within a testing session. To lessen the risk of a security breach occurring during these breaks, students requiring the use of restroom facilities must be escorted by either a proctor or a test examiner. In addition, students must not be allowed to use any form of wireless communication during these breaks.

## 4.5 Test Administration Training

DRC provided pre-recorded training for DACs, SACs, and TAs for the Spring 2022 administration of the Wisconsin Forward Exam. The webinars were recorded by DRC. The purpose of the webinars was to keep districts and schools informed about policies and procedures related to the Wisconsin Forward Exam administration. The information covered in the webinars included standardizing the administration of the Wisconsin Forward Exam, maintaining the security of the assessments, allowing access to the assessments for special populations by providing appropriate designated supports or accommodations, and providing guidance on appropriate interpretations of the test results. These communication efforts by DPI and the ancillary information developed by DRC are in alignment with multiple best practices of the testing industry and, in particular, support the following AERA, APA, & NCME (2014) Standards:

> **Standard 4.15** The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented. (p. 90)

> **Standard 4.16** The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions. (p. 90)

> **Standard 6.1** Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user. (p. 114)

> **Standard 6.2** When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing. (p. 115)

> **Standard 6.3** Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user. (p. 115)

> **Standard 6.4** The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance. (p. 116)

In order to ensure standardized testing administration for all students, a Guide for District and School Assessment Coordinators, included in the Test Administration Manual, was made available to all assessment coordinators. The guide included the following topics:

- Testing Roles and Responsibilities
- Test Security
- Resources and Training Materials
- Test Schedules
- DRC INSIGHT Portal
- Accessibility
- Student Transfers
- Prior to the Close of the Testing Window
- Data and Reporting

Test Administration Manuals, made available to all TAs, included the following topics:

- Key Dates
- TA Responsibilities
- Test Times
- Test Security
- Accessibility Information
- Prior to Testing Instructions
- Test Tickets
- During Testing Information
- Test Administration Script

These topics were also addressed in the recorded trainings that were posted for online access.

## Student Preparation for Online Testing

Prior to testing, schools and districts were encouraged to provide students with time to complete both a tutorial video series and an online tools training. Sample test items were also provided for each grade and content area.

## Student and Administrator Tutorial Videos

Student and administrator tutorial videos were available for students and TAs to become familiar with the online testing environment. Tutorials could be viewed as a class or at an individual student machine by launching INSIGHT and clicking on DRC INSIGHT Online Assessment Tutorials.

## Online Tools Training

The Online Tools Training (OTT) was provided for students to have a hands on opportunity to practice the types of items and tools available in the online testing system. The OTTs were available publicly for practice using a Chrome browser. Users (at home or school) could visit https://dpi.wi.gov/assessment/forward/sample-items to access the public OTTs. The OTTs could also be accessed on student testing devices once INSIGHT was installed. General OTTs were made available for each content area and grade level. Separate OTTs were available for students to practice using VSL with closed-captioning, TTS, Spanish translation, masking, and color choice tools. VSL and Spanish OTTs were available by grade band (3–5, 6–8, and 10). The OTTs were not scored and were not intended for content practice.

## Item Samplers

Item samplers were developed to be used by both educators and students to gain familiarity with the various item types and their varying functionalities. The format appears as a "guided practice test" in the online, PDF, and Braille versions of the tests.

Accommodated versions of the item samplers, reflecting the Wisconsin Forward Exam, were produced, including TTS, stacked Spanish translation (in Mathematics, Science, and Social Studies), and VSL with CC. All tools and supports available in the test engine were applied to this student online experience.

Access to the item samplers was granted through the OTT menu page. A username and password were displayed on the login screen. The "click to enlarge" item displayed the answer key and scoring guide for each item online. In addition, a paper answer key and scoring guide were provided as a document for posting.

## Administration Supports before and after Testing

With a few exceptions (i.e., accommodated student versions), the Wisconsin Forward Exam was administered fully online. Because DRC produced a variety of Wisconsin-specific manuals with process reviews by DRC program management staff, DRC editorial staff, and DPI staff, substantial consideration was given to the information required for successful online testing to occur. DPI provided final approval for each document prior to delivery and public posting.

Table 4-8 displays a list of electronic manual materials that DRC developed in conjunction with DPI. A final PDF of each deliverable was provided to DPI to post to the DPI informational website to allow districts to review and/or print.

For additional and more specific information related to test administration, refer to the Test Administration Manual that is available online at https://dpi.wi.gov/assessment/forward/resources#manuals.

## 4.6 Summary

This part of the report provides information on student participation rates and summarizes the processes and activities implemented and the information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. It describes how the test administration procedures implemented for the Wisconsin Forward Exam were in alignment with best practices of the testing industry.

**Table 4-1. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 3**

| Grade 3 | English Language Arts | | Mathematics | |
|---|---|---|---|---|
| Accommodation or Support | N Count | Percent | N Count | Percent |
| Used Braille [BRL] | 2 | 0.00 | 2 | 0.00 |
| Used Print on Demand [POD] | 3 | 0.01 | 3 | 0.01 |
| Used Bilingual Dictionary | | | 164 | 0.28 |
| Used Magnification | 48 | 0.08 | 48 | 0.08 |
| Used Noise Buffers | 513 | 0.88 | 514 | 0.88 |
| Used Read Aloud | 400 | 0.69 | 405 | 0.69 |
| Used Scribe | 505 | 0.87 | 450 | 0.77 |
| Used Separate Setting | 7650 | 13.13 | 7709 | 13.19 |
| Used Alternate Response Options | 10 | 0.02 | 9 | 0.02 |
| Used Read Aloud (Reading Passages) | 6 | 0.01 | | |
| Provided Color Choices [CC] | 53 | 0.09 | 53 | 0.09 |
| Used Contrasting Color [CTC] | 42 | 0.07 | 42 | 0.07 |
| Used Reverse Contrast [RC] | 42 | 0.07 | 39 | 0.07 |
| Used Masking [MSK] | 544 | 0.93 | 547 | 0.94 |
| Used Text-to-Speech [TTS] | 13318 | 22.85 | 13504 | 23.10 |
| Used Spanish Translation [ST] | 327 | 0.56 | 650 | 1.11 |
| Used Video Sign Language [VSL (ASL)] | 22 | 0.04 | 22 | 0.04 |
| Used Color Overlay | 19 | 0.03 | 19 | 0.03 |
| Amplification Device | 66 | 0.11 | 66 | 0.11 |
| Small Group Translation | 97 | 0.17 | 163 | 0.28 |
| Translator/Interpreter | 24 | 0.04 | 42 | 0.07 |
| Read Aloud in Spanish | 45 | 0.08 | 133 | 0.23 |
| Used Closed Captioning [C CAP] ELA | 53 | 0.09 | | |
| Used Listening Scripts [LS] ELA | 4 | 0.01 | | |
| Used Abacus Math | | | 28 | 0.05 |
| Used Non-embedded Calculator Math | | | 169 | 0.29 |

**Table 4-2. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 4**

| Grade 4<br>Accommodation or Support | English Language Arts N Count | English Language Arts Percent | Mathematics N Count | Mathematics Percent | Science N Count | Science Percent | Social Studies N Count | Social Studies Percent |
|---|---|---|---|---|---|---|---|---|
| Used Braille [BRL] | 3 | 0.01 | 2 | 0.00 | 3 | 0.01 | 3 | 0.01 |
| Used Print on Demand [POD] | 1 | 0.00 | 1 | 0.00 | 1 | 0.00 | 1 | 0.00 |
| Used Bilingual Dictionary | | | 143 | 0.24 | 151 | 0.26 | 151 | 0.26 |
| Used Magnification | 89 | 0.15 | 87 | 0.15 | 86 | 0.15 | 86 | 0.15 |
| Used Noise Buffers | 625 | 1.06 | 619 | 1.05 | 614 | 1.04 | 613 | 1.04 |
| Used Read Aloud | 401 | 0.68 | 410 | 0.70 | 401 | 0.68 | 394 | 0.67 |
| Used Scribe | 434 | 0.74 | 394 | 0.67 | 372 | 0.63 | 370 | 0.63 |
| Used Separate Setting | 8088 | 13.76 | 8146 | 13.82 | 8036 | 13.65 | 8019 | 13.63 |
| Used Alternate Response Options | 13 | 0.02 | 10 | 0.02 | 10 | 0.02 | 10 | 0.02 |
| Provided Color Choices [CC] | 87 | 0.15 | 87 | 0.15 | 86 | 0.15 | 86 | 0.15 |
| Used Contrasting Color [CTC] | 54 | 0.09 | 52 | 0.09 | 52 | 0.09 | 52 | 0.09 |
| Used Reverse Contrast [RC] | 26 | 0.04 | 25 | 0.04 | 25 | 0.04 | 25 | 0.04 |
| Used Masking [MSK] | 618 | 1.05 | 620 | 1.05 | 619 | 1.05 | 612 | 1.04 |
| Used Text-to-Speech [TTS] | 12617 | 21.47 | 12873 | 21.84 | 12766 | 21.68 | 12762 | 21.69 |
| Used Spanish Translation [ST] | 378 | 0.64 | 583 | 0.99 | 569 | 0.97 | 522 | 0.89 |
| Used Video Sign Language [VSL (ASL)] | 15 | 0.03 | 14 | 0.02 | 14 | 0.02 | 14 | 0.02 |
| Used Color Overlay | 18 | 0.03 | 17 | 0.03 | 17 | 0.03 | 19 | 0.03 |
| Amplification Device | 76 | 0.13 | 74 | 0.13 | 74 | 0.13 | 74 | 0.13 |
| Small Group Translation | 82 | 0.14 | 123 | 0.21 | 120 | 0.20 | 120 | 0.20 |
| Translator/Interpreter | 19 | 0.03 | 34 | 0.06 | 33 | 0.06 | 34 | 0.06 |
| Read Aloud in Spanish | 56 | 0.10 | 113 | 0.19 | 112 | 0.19 | 112 | 0.19 |
| Used Closed Captioning [C CAP] ELA | 62 | 0.11 | | | | | | |
| Used Listening Scripts [LS] ELA | 11 | 0.02 | | | | | | |
| Used Abacus Math | | | 33 | 0.06 | | | | |
| Used Non-embedded Calculator Math | | | 206 | 0.35 | | | | |
| Used Multiplication Table Math | | | 1808 | 3.07 | | | | |

**Table 4-3. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 5**

| Grade 5 | English Language Arts | | Mathematics | |
|---|---|---|---|---|
| Accommodation or Support | N Count | Percent | N Count | Percent |
| Used Braille [BRL] | 1 | 0.00 | 1 | 0.00 |
| Used Print on Demand [POD] | 1 | 0.00 | 1 | 0.00 |
| Used Bilingual Dictionary | | | 144 | 0.24 |
| Used Magnification | 50 | 0.08 | 51 | 0.09 |
| Used Noise Buffers | 581 | 0.98 | 581 | 0.98 |
| Used Read Aloud | 388 | 0.66 | 376 | 0.64 |
| Used Scribe | 379 | 0.64 | 356 | 0.60 |
| Used Separate Setting | 8155 | 13.81 | 8229 | 13.90 |
| Used Alternate Response Options | 9 | 0.02 | 8 | 0.01 |
| Used Read Aloud (Reading Passages) | 1 | 0.00 | | |
| Provided Color Choices [CC] | 85 | 0.14 | 85 | 0.14 |
| Used Contrasting Color [CTC] | 107 | 0.18 | 108 | 0.18 |
| Used Reverse Contrast [RC] | 74 | 0.13 | 74 | 0.13 |
| Used Masking [MSK] | 638 | 1.08 | 641 | 1.08 |
| Used Text-to-Speech [TTS] | 11488 | 19.46 | 11644 | 19.67 |
| Used Spanish Translation [ST] | 423 | 0.72 | 658 | 1.11 |
| Used Video Sign Language [VSL (ASL)] | 10 | 0.02 | 9 | 0.02 |
| Used Color Overlay | 23 | 0.04 | 23 | 0.04 |
| Amplification Device | 50 | 0.08 | 50 | 0.08 |
| Small Group Translation | 100 | 0.17 | 128 | 0.22 |
| Translator/Interpreter | 25 | 0.04 | 39 | 0.07 |
| Read Aloud in Spanish | 48 | 0.08 | 101 | 0.17 |
| Used Closed Captioning [C CAP] ELA | 66 | 0.11 | | |
| Used Listening Scripts [LS] ELA | 3 | 0.01 | | |
| Used Abacus Math | | | 27 | 0.05 |
| Used Non-embedded Calculator Math | | | 243 | 0.41 |
| Used Multiplication Table Math | | | 2227 | 3.76 |

**Table 4-4. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 6**

| Grade 6 | English Language Arts | | Mathematics | |
|---|---|---|---|---|
| Accommodation or Support | N Count | Percent | N Count | Percent |
| Used Braille [BRL] | 1 | 0.00 | 1 | 0.00 |
| Used Print on Demand [POD] | 2 | 0.00 | 2 | 0.00 |
| Used Bilingual Dictionary | | | 211 | 0.35 |
| Used Magnification | 87 | 0.14 | 87 | 0.14 |
| Used Noise Buffers | 410 | 0.68 | 410 | 0.68 |
| Used Read Aloud | 350 | 0.58 | 359 | 0.60 |
| Used Scribe | 253 | 0.42 | 226 | 0.38 |
| Used Separate Setting | 7419 | 12.34 | 7492 | 12.44 |
| Used Alternate Response Options | 12 | 0.02 | 7 | 0.01 |
| Used Read Aloud (Reading Passages) | 4 | 0.01 | | |
| Provided Color Choices [CC] | 102 | 0.17 | 100 | 0.17 |
| Used Contrasting Color [CTC] | 114 | 0.19 | 106 | 0.18 |
| Used Reverse Contrast [RC] | 64 | 0.11 | 63 | 0.10 |
| Used Masking [MSK] | 548 | 0.91 | 572 | 0.95 |
| Used Text-to-Speech [TTS] | 9259 | 15.40 | 9466 | 15.72 |
| Used Spanish Translation [ST] | 378 | 0.63 | 472 | 0.78 |
| Used Video Sign Language [VSL (ASL)] | 18 | 0.03 | 18 | 0.03 |
| Used Color Overlay | 22 | 0.04 | 23 | 0.04 |
| Amplification Device | 66 | 0.11 | 66 | 0.11 |
| Small Group Translation | 120 | 0.20 | 149 | 0.25 |
| Translator/Interpreter | 23 | 0.04 | 31 | 0.05 |
| Read Aloud in Spanish | 51 | 0.08 | 77 | 0.13 |
| Used Closed Captioning [C CAP] ELA | 74 | 0.12 | | |
| Used Listening Scripts [LS] ELA | 11 | 0.02 | | |
| Used Abacus Math | | | 8 | 0.01 |
| Used Non-embedded Calculator Math | | | 448 | 0.74 |
| Used Multiplication Table Math | | | 2590 | 4.30 |

**Table 4-5. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 7**

| Grade 7 | English Language Arts | | Mathematics | |
|---|---|---|---|---|
| Accommodation or Support | N Count | Percent | N Count | Percent |
| Used Braille [BRL] | 4 | 0.01 | 4 | 0.01 |
| Used Print on Demand [POD] | 1 | 0.00 | 1 | 0.00 |
| Used Bilingual Dictionary | | | 151 | 0.24 |
| Used Magnification | 84 | 0.14 | 85 | 0.14 |
| Used Noise Buffers | 312 | 0.50 | 313 | 0.51 |
| Used Read Aloud | 254 | 0.41 | 261 | 0.42 |
| Used Scribe | 194 | 0.31 | 175 | 0.28 |
| Used Separate Setting | 7471 | 12.08 | 7521 | 12.14 |
| Used Alternate Response Options | 9 | 0.01 | 8 | 0.01 |
| Used Read Aloud (Reading Passages) | 6 | 0.01 | | |
| Provided Color Choices [CC] | 111 | 0.18 | 91 | 0.15 |
| Used Contrasting Color [CTC] | 105 | 0.17 | 64 | 0.10 |
| Used Reverse Contrast [RC] | 63 | 0.10 | 30 | 0.05 |
| Used Masking [MSK] | 542 | 0.88 | 531 | 0.86 |
| Used Text-to-Speech [TTS] | 8926 | 14.43 | 9087 | 14.66 |
| Used Spanish Translation [ST] | 324 | 0.52 | 431 | 0.70 |
| Used Video Sign Language [VSL (ASL)] | 8 | 0.01 | 8 | 0.01 |
| Used Color Overlay | 13 | 0.02 | 14 | 0.02 |
| Amplification Device | 45 | 0.07 | 45 | 0.07 |
| Small Group Translation | 56 | 0.09 | 90 | 0.15 |
| Translator/Interpreter | 7 | 0.01 | 14 | 0.02 |
| Read Aloud in Spanish | 35 | 0.06 | 54 | 0.09 |
| Used Closed Captioning [C CAP] ELA | 72 | 0.12 | | |
| Used Listening Scripts [LS] ELA | 8 | 0.01 | | |
| Used Abacus Math | | | 4 | 0.01 |
| Used Non-embedded Calculator Math | | | 553 | 0.89 |
| Used Multiplication Table Math | | | 2693 | 4.35 |

**Table 4-6. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 8**

| Grade 8 | English Language Arts | | Mathematics | | Science | | Social Studies | |
|---|---|---|---|---|---|---|---|---|
| Accommodation or Support | N Count | Percent | N Count | Percent | N Count | Percent | N Count | Percent |
| Used Braille [BRL] | 5 | 0.01 | 5 | 0.01 | 5 | 0.01 | 5 | 0.01 |
| Used Print on Demand [POD] | 3 | 0.00 | 3 | 0.00 | 1 | 0.00 | 2 | 0.00 |
| Used Bilingual Dictionary | | | 167 | 0.27 | 166 | 0.26 | 166 | 0.27 |
| Used Magnification | 78 | 0.12 | 78 | 0.12 | 78 | 0.12 | 78 | 0.12 |
| Used Noise Buffers | 244 | 0.39 | 238 | 0.38 | 234 | 0.37 | 233 | 0.37 |
| Used Read Aloud | 270 | 0.43 | 258 | 0.41 | 265 | 0.42 | 264 | 0.42 |
| Used Scribe | 158 | 0.25 | 144 | 0.23 | 144 | 0.23 | 146 | 0.23 |
| Used Separate Setting | 7322 | 11.68 | 7385 | 11.77 | 7259 | 11.59 | 7237 | 11.56 |
| Used Alternate Response Options | 9 | 0.01 | 8 | 0.01 | 9 | 0.01 | 9 | 0.01 |
| Used Read Aloud (Reading Passages) | 7 | 0.01 | | | | | | |
| Provided Color Choices [CC] | 116 | 0.19 | 115 | 0.18 | 114 | 0.18 | 113 | 0.18 |
| Used Contrasting Color [CTC] | 114 | 0.18 | 112 | 0.18 | 111 | 0.18 | 111 | 0.18 |
| Used Reverse Contrast [RC] | 88 | 0.14 | 87 | 0.14 | 88 | 0.14 | 88 | 0.14 |
| Used Masking [MSK] | 481 | 0.77 | 492 | 0.78 | 490 | 0.78 | 488 | 0.78 |
| Used Text-to-Speech [TTS] | 8535 | 13.62 | 8614 | 13.72 | 8438 | 13.47 | 8419 | 13.45 |
| Used Spanish Translation [ST] | 325 | 0.52 | 455 | 0.72 | 456 | 0.73 | 459 | 0.73 |
| Used Video Sign Language [VSL (ASL)] | 14 | 0.02 | 15 | 0.02 | 16 | 0.03 | 14 | 0.02 |
| Used Color Overlay | 11 | 0.02 | 11 | 0.02 | 11 | 0.02 | 11 | 0.02 |
| Amplification Device | 52 | 0.08 | 52 | 0.08 | 52 | 0.08 | 52 | 0.08 |
| Small Group Translation | 58 | 0.09 | 97 | 0.15 | 97 | 0.15 | 97 | 0.15 |
| Translator/Interpreter | 15 | 0.02 | 18 | 0.03 | 15 | 0.02 | 15 | 0.02 |
| Read Aloud in Spanish | 30 | 0.05 | 51 | 0.08 | 50 | 0.08 | 50 | 0.08 |
| Used Closed Captioning [C CAP] ELA | 84 | 0.13 | | | | | | |
| Used Listening Scripts [LS] ELA | 8 | 0.01 | | | | | | |
| Used Abacus Math | | | 4 | 0.01 | | | | |
| Used Non-embedded Calculator Math | | | 620 | 0.99 | | | | |
| Used Multiplication Table Math | | | 2419 | 3.85 | | | | |

**Table 4-7. Number and Percentage of Students Using Accommodations or Designated Supports, Grade 10**

| Grade 10 | Social Studies | |
|---|---|---|
| Accommodation or Support | N Count | Percent |
| Used Braille [BRL] | 4 | 0.01 |
| Used Print on Demand [POD] | 1 | 0.00 |
| Used Bilingual Dictionary | 193 | 0.32 |
| Used Magnification | 63 | 0.11 |
| Used Noise Buffers | 56 | 0.09 |
| Used Read Aloud | 118 | 0.20 |
| Used Scribe | 70 | 0.12 |
| Used Separate Setting | 4465 | 7.52 |
| Used Alternate Response Options | 1 | 0.00 |
| Provided Color Choices [CC] | 10 | 0.02 |
| Used Contrasting Color [CTC] | 15 | 0.03 |
| Used Reverse Contrast [RC] | 12 | 0.02 |
| Used Masking [MSK] | 166 | 0.28 |
| Used Text-to-Speech [TTS] | 3276 | 5.52 |
| Used Spanish Translation [ST] | 255 | 0.43 |
| Used Video Sign Language [VSL (ASL)] | 15 | 0.03 |
| Used Color Overlay | 11 | 0.02 |
| Amplification Device | 18 | 0.03 |
| Small Group Translation | 87 | 0.15 |
| Translator/Interpreter | 9 | 0.02 |
| Read Aloud in Spanish | 20 | 0.03 |

**Table 4-8. Summary Table of Manual Materials**

| Material | Configuration |
|---|---|
| **DRC INSIGHT Portal Guide: Managing Users, Students, and Testing** | The DRC INSIGHT Portal Guide includes the following information:<br><br>• Managing user's own DRC INSIGHT Portal account<br>• Managing other DRC INSIGHT Portal users<br>• Adding and editing students and student demographics, accommodations, and testing codes<br>• Viewing, adding, and editing student test session information<br>• Printing and managing student test tickets<br>• Transferring students between schools and districts<br>• Entering Not-Tested or Invalidation Codes<br>• Unlocking or purging a student test<br>• Managing test sessions<br>• Monitoring testing status |
| **Accessibility Guide** | The Accessibility Guide outlines the various accessibility options available to students taking the Wisconsin Forward Exam. Guidelines for using the various accessibility features are also included. |
| **Student/Administrator Tutorials** | The Student Tutorial includes 11 videos intended for students in grades 4–10 and 7 videos for students in grade 3. It is designed to show students the interface of the online testing system and familiarize them with the tools and features available. It is intended to accompany the Online Tools Training (OTT).<br>The 2022 tutorial also includes 11 videos for test administrators to familiarize them with the administrative features and functionality of the DRC INSIGHT Portal as well as the accessibility features of the Wisconsin Forward Exam. |
| **Item Samplers** | The item samplers can be used by both educators and students to gain familiarity with the types of items and their functionalities. The format appears as a "guided practice test" in the online, PDF, and Braille versions. Accommodations, universal tools, and supports are available in the test engine for the item samplers.<br>Item samplers are accessible through the OTT menu page. The PDF versions include the answer key and scoring guide for each item. |
| **Online Tools Training (OTT)** | The OTT is a hands-on opportunity for students to become familiar with logging in, navigating, using tools, using accessibility features, reviewing, and submitting the test prior to signing in to an actual test. It is designed to be a second step after viewing the student tutorials. |

**Table 4-8 (continued). Summary Table of Manual Materials**

| Material | Configuration |
|---|---|
| **Test Administration Manual (TAM)** | TAM is a document intended for test administrators (TAs) and proctors. It includes the following information:<br><br>• Key dates<br>• TA and proctor responsibilities<br>• Test times<br>• Test security<br>• Accessibility information<br>• Procedures for before and during testing<br>• Test ticket management<br>• Test administration scripts<br><br>The TAM also includes a Guide for District and School Assessment Coordinators, which includes the following:<br><br>• Roles and responsibilities<br>• Test security<br>• Resources and training materials<br>• Test schedules<br>• DRC INSIGHT Portal and DRC INSIGHT secure browser<br>• Accessibility<br>• Student transfers<br>• Procedures to be completed prior to the close of the testing window<br>• Data and reporting |
| **Technology User Guide (TUG)** | The TUG is a document intended for Technology Coordinators. It is split into four volumes and includes detailed instructions on the installation and configuration of INSIGHT and the Central Office Services (COS) for all supported platforms. |
| **Interpretive Guide** | The Interpretive Guide is a document that includes the following information:<br><br>• Interpreting Wisconsin Forward Exam scores<br>• Accessing Individual Student Reports (ISRs) and interactive summary reports via the DRC INSIGHT Portal |
| **Technology Readiness Package** | The Technology Readiness Package is a suite of documents and tools for Technology Coordinators to prepare for the Wisconsin Forward Exam that includes the following:<br><br>• What is new and changing<br>• Assessing online testing readiness<br>• Capacity estimator<br>• COS decision guide<br>• Extended retries<br>• Headset guidance<br>• Keyboard settings<br>• Installation of COS and INSIGHT<br>• Installation of INSIGHT App<br>• Evaluation and troubleshooting<br>• System requirements<br>• Technology overview presentation<br>• Technology Coordinator Checklist<br>• Technology FAQ |

**Table 4-8 (continued). Summary Table of Manual Materials**

| Material | Configuration |
|---|---|
| **Technical Report** | The Technical Report is a manual that covers all grades and all psychometric details associated with administering the Wisconsin Forward Exam. The Technical Report provided by DRC presents thorough documentation to demonstrate the assessment validity. The document contains the following information:<br><br>• Description of the item pool used in the Wisconsin form-development process<br>• Description of the test administration process and test security<br>• Scoring of various types of items<br>• Summary information of student performance (including means and standard deviations of scale scores, percentage of examinees within each performance level for each content area and grade level, and scale score distribution tables)<br>• Item- and test-level analysis information for each content area and grade level, test scaling procedure, and student scoring process<br>• Evidence of test validity |
| **Data Forensic Report** | A separate Data Forensic Report includes analyses of the following:<br><br>• Evaluation of response changes<br>• Evaluation of student response time to items |

# Part 5: Scoring

The purpose of Part 5 is to demonstrate adherence to AERA, APA, & NCME (2014) Standard 4.18, which states the following:

> Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended response items such as performance tasks, portfolios, and essays. (p. 91)

Part 5 describes

- the scoring process of multiple-choice (MC) and multi-select (MS) items;
- the autoscoring process of technology-enhanced (TE), short-answer (SA), and evidence-based selected response (EBSR) items; and
- the scoring of text-dependent analysis (TDA) items, including
  - scoring rubrics,
  - artificial intelligence (AI) scoring process,
  - handscoring process,
  - scoring personnel selection,
  - anchor papers selection, and
  - TDA item scores distribution.

## 5.1 Multiple-Choice and Multi-Select Item Scoring Process

Responses to MC and MS items were captured during the online test administration. In the case of the Braille or paper-and-pencil form administrations, student responses to these items were transcribed into the online system by a TA. All MC and MS items had one and only one correct item response or a combination of responses.

## 5.2 Technology-Enhanced, Short-Answer, and Evidence-Based Selected Response Item Scoring Process

All TE, SA, and EBSR items were processed through DRC's autoscoring engine and scored according to the assigned scoring rules. DRC ensured that all rubrics and scoring rules were verified for accuracy before scoring any of these items. DRC established an adjudication process for these items and any gridded responses to verify that correct answers were identified. The quality control process for DRC's TE, SA, and EBSR item scoring included the following:

- A scoring rubric was created for each TE, SA, and EBSR item. It was similar to describing the one correct answer for dichotomously scored items (scored as either right or wrong). For ELA EBSR items worth 2 points, the rubric described in detail the type of response that could receive partial credit for 1 score point.

- The information from the scoring rubric was entered into the scoring system within the item banking system so that all information about the item resided in one place, along with the item image and other metadata. This scoring information designated specific information that varied by item type. For example, for a drag-and-drop item, the information included which objects are to be placed into which drop region to receive credit.

- The information was then verified by another autoscoring expert.

- After testing started, reports were generated that showed every response, how many students gave each response, and the score the scoring system provided.

- The scoring was then checked against the scoring rubric using two levels of verification.

- If any discrepancies were found, the scoring information was modified and verified again. Scoring was then rerun. This checking and modification process continued until no other issues were found.

- As a final check, a final report was run that showed all student responses, along with frequencies and received scores.

In the case of the Braille or paper-and-pencil form administrations, student responses to paper-and-pencil TE, SA, EBSR, or TE-equivalent items were transcribed (entered) into the online system by a TA.

## 5.3 Scoring of Text-Dependent Analysis Items

Sections 5.3 and 5.4 document the scoring processes used for TDA items. This documentation forms part of the validity evidence supporting the scoring process used for these items. Sections 5.3 and 5.4 describe the scoring rubrics, the scoring process, the selection of sample (anchor) papers used to train scoring personnel, the process of selecting personnel, and the distributions of scores for TDA items.

### 5.3.1 Description of Scoring Rubrics and Non-score Codes

In the 2022 test administration, the ELA forms in grades 3–8 contained one TDA item at each grade level. As stated in Part 3, Table 3-1, of this report, the TDA prompts are scored using a holistic scoring guideline on a 1–4 point scale. A weight of 2 is later applied to the item scores in order to compute the student total test raw scores and scale scores. That is, the TDA prompts will contribute up to 8 raw score points toward the student total test raw score.

The TDA responses were scored using an AI engine, and then validation scoring was performed by human scorers on approximately 10 percent of the AI scored responses. Table 5-1 presents the scoring rubric. In cases where student responses could not be scored, a non-score code was used. The non-score codes are presented in Table 5-2. All non-score codes were converted to a score of "0" in the derivation of student total test scores.

### 5.3.2 Artificial Intelligence Scoring

DRC partnered with Measurement Incorporated (MI) to score the TDA tasks. MI employed its essay scoring engine (called Project Essay Grade or PEG) to score all student responses. The AI model for scoring the Wisconsin student responses was built by first having DRC expert scorers score a representative sample of Wisconsin responses twice, independently, and resolve any scores that did not agree. While the engine only requires one score per response to build a model, the second score provides necessary information about how well two humans are able to agree on a score, which is then used as a benchmark for how well the engine's predictions should agree with the human scores. Approximately

3,000 student responses per grade from the Spring 2022 test administration were selected, handscored independently twice, and used in the AI model building.

The engine training sets consisting of scored sample responses and corresponding scores were delivered to the AI team at MI for model development. MI's linguistics experts, software developers, psychometricians, and human computer interaction specialists created task-specific algorithms that were then used to predict how humans would score these responses. The PEG team applied a standard stratified random sampling to all training sets, which is designed to produce two subsets of approximately 1,700 "training responses" and approximately 300 "validation responses," which approximated the score point distribution of the full training sets. The training responses were used to build the scoring model. The validation responses were used to verify the accuracy of AI scoring.

To build a scoring model, the engine analyzes the training set and calculates features that pertain to the content in question. The engine then sends the features to dozens of different algorithms that compete to see which ones can best associate the text features with the human assigned scores. These algorithms draw on many of the latest advances in the field of machine learning to generate both linear and nonlinear models. The strongest models are then automatically blended together to create a final model that retains the best elements from the various algorithms.

When the engine builds a model, it selects the model elements that maximize scoring accuracy for the data in question. Therefore, it is important to choose an agreement statistic on which the engine can optimize its models in such a way that the final model will exhibit reliable, accurate scoring. The inter-rater reliability of two human scorers is often measured via exact and adjacent agreement or the Pearson product-moment correlation coefficient (Pearson's $r$). It has been found that using quadratic weighted kappa, which has become the industry standard for AI scoring as the optimization and evaluation metric, leads to the most reliable and accurate scoring. Quadratic weighted kappa as a metric can detect changes in mean difference and variance between scorers and is, therefore, well suited for comparing the accuracy of AI scoring with that of human scoring and measuring the agreement of two independent human scorers.

MI's AI scoring software flagged student responses that could not be AI scored. The software has various triggers for identifying alert responses and responses in which it has low confidence. These responses lack proper development, lack enough content to be scored, are written in an unsupported language, contain inappropriate language, or represent a bad faith effort to complete the test (e.g., repeated text, off-topic text). These responses that could not be scored by AI were routed to DRC for human scoring with a condition code indicating why the response could not be AI scored. Of note for the 2022 Spring scoring is that the condition code used to indicate the threshold of "too few words" was changed. The previous threshold had been set to 45; the new threshold was set to 20. That is, all responses that had 20 or fewer words were sent to a human reader for review.

### 5.3.3 Handscoring Process

Human scoring of TDA items is referred to as "handscoring." The scoring personnel who score TDA items are referred to as scorers. The scorers were trained using customized training materials, such as the anchor papers described in Section 5.3.5. Once qualified, scorers were required to maintain accuracy standards throughout the project. These requirements were assessed primarily through each scorer's daily agreement rates with the AI scores (described below) and through targeted read-behinds with team leaders (described below). Reports were generated daily and monitored by the scoring director, team leaders, and project manager. Any scorers falling below the established quality standards for any item were retrained with the supervisors, who monitored scoring trends (such as difficulty with any particular score point). These

scorers also received additional reviews and read-behinds. Failure to recalibrate resulted in dismissal from the scoring assignment. This process was in place throughout the entire handscoring window.

## 5.3.4 Handscoring System

Scoreboard, DRC's handscoring system, was used to score TDA items as a validation method and to resolve cases where the AI engine returned a non-scorable condition code. Scoreboard presented images of rendered online responses to trained scorers who assigned scores for the TDA items. Images of each student's responses were automatically routed to designated groups of scorers who were trained and qualified to score these items.

## 5.3.5 Anchor Papers and Training Papers

DRC's project managers and scoring directors used the scoring guidelines, or rubrics, and Wisconsin student responses to select a representative sampling of student responses for each score point to create the necessary training materials for operational scoring. The sample reflects the various common response types produced for a specific item. The responses were then assembled into training/qualifying sets and shared with at least one other content expert for review. The scoring director for the specific grade took detailed notes, capturing scores and specific rationales for each score. Each grade and TDA item progressed in the same manner, using the same process. Once all sets were reviewed and scored, each grade-level scoring director had fully annotated a set of anchor papers, training papers, and qualifying papers. These anchor, training, and qualifying papers were then used to train a select group of scorers who scored the student responses that were used to train the AI engine in a process called model building. For this model-building activity, each student response was independently scored by two separate scorers. If there was any disagreement between the two readers, the scores were adjudicated to 100 percent agreement. Approximately 3,000 responses per grade were then delivered to the AI vendor to build the AI engine model. Once the model was built, the AI engine scored the remaining Wisconsin student responses. Upon completion of the AI scoring, a random sample consisting of approximately 10 percent of the student responses scored by the AI engine was sent to DRC for a human read. DRC then scored the 10 percent read-behind sample using a group of scorers trained to qualification standards to ensure consistency. The 10 percent read-behind with human scorers served as a validation check of the AI engine scoring data.

## 5.3.6 Scoring Personnel and Qualifications

AERA, APA, & NCME (2014) Standard 4.20 specifies the following:

The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring. (p. 92)

DRC recruited, trained, and managed personnel to complete all the handscoring operations within the timelines of the contract. The recruitment process and requirements of the scorers, team leaders, and scoring supervisors are described in the following sections.

**Scorers**—The DRC scorer pool included many retired and current educators, engineers, editors, published authors, and individuals with advanced degrees. The minimum qualification for all scorers was a bachelor's degree. Scorers were required to participate in training and successfully pass a qualification round. Once qualified, scorers could start scoring, but throughout the scoring process, scorer performance was assessed by a scoring director, a team leader, and the project manager through read-behinds and reviews of inter-rater reliability statistics as described in Sections 5.3.8 and 5.4.

**Team Leaders**—Team leaders were selected on the basis of their ability to maintain a high degree of scoring accuracy and consistency, often across multiple content areas and grades. Team leaders were also required to possess good interpersonal and leadership skills in order to be effective when training and counseling scorers. Each team leader was responsible for a small team of scorers. In addition to performing read-behinds on scorers, team leaders also coached scorers when needs were identified through data review or otherwise by supervisory staff.

**Scoring Directors**—Scoring directors comprised the core group at DRC who directed and organized the scoring process and trained team leaders and scorers. Scoring directors had extensive experience as team leaders prior to their qualification and selection, and all had previous scoring director experience. Scoring directors were content area experts. They oversaw all team leaders and scorers.

## 5.3.7 Scorer Training

AERA, APA, & NCME (2014) Standard 6.9 specifies the following:

Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected. (p. 118)

Qualification was a critical task in the training process and the final determinant of scorer readiness. All scorers, including team leaders, were required to achieve a certain level of scoring accuracy in the qualifying round that followed training. The standard to which they were held was the industry standard for TDA items: at least 70% exact agreement. Only those who were successfully validated were qualified as scorers to score tests.

## 5.3.8 Monitoring the Scoring Process

AERA, APA, & NCME (2014) Standard 6.8 states the following:

Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented. (p. 118)

The read-behind was used as a valuable monitoring technique. Each team leader was able to read a random selection of a scorer's scored responses. This reading could be targeted at the item and score-point level. The team leader scored the same item the scorer had scored. The scores (the scorer score and the team leader score) were compared, and if they agreed, the team leader was able to offer feedback, which enhanced the scorer's confidence and ability to score quickly and accurately. However, if a scorer strayed from the standards established in the training samples, the aberrant scoring was detected, and the team leader was able to offer guidance necessary to refocus the scorer's effort. Read-behinds by team leaders were more frequent for the scorers who had inconsistent scores, thus correcting any scoring variations. For

---

aberrant or inconsistent scoring, DRC has the capability to wholesale drop scores and have the responses rescored if deemed necessary.

### 5.3.9 Final Scores

All TDA responses were sent to the AI engine for scoring. The AI scores were the final scores (i.e., scores of record). In all cases where the AI engine returned a non-scorable condition code, the student responses were reviewed and scored by humans and a resolution was reached. If a human scorer was able to assign a score for a response that the AI engine was not able to score, then a score from a human scorer became the score of record.

## 5.4 Inter-rater Reliability

A random 10 percent of the AI-scored responses were sent to human scorers for second reads to validate the AI scores. The statistics for the inter-rater reliability were calculated for all TDA items. To determine the reliability of scoring, the score distribution and percentage of agreement of the two readers were examined. In this section, the distribution of TDA item scores is presented. Additional inter-rater reliability measures, including intra-class correlation and weighted kappa statistics, are presented in Part 8 of the Technical Report.

### 5.4.1 Distribution of TDA Item Scores

Table 5-3 shows the score and non-scorable code distributions for TDA items. The presented scores, on a 1–4 point scale, are from the AI engine supplemented by non-scorable responses resolved by human scorers. Students who did not attempt the TDA item session of the ELA assessments are not included in this table.

Table 5-4 shows the score and non-scorable code distributions for TDA items for responses selected for the second read (handscoring). Table 5-5 shows the associated percentages of scores and non-scorable codes for TDA items for responses selected for the second read. In both tables, Scorer 1 is the AI engine and Scorer 2 is a human scorer. It should be noted that all non-scorable responses returned by the AI engine were reviewed by the scoring directors and assigned either a specific condition code or a score. The data in the non-scorable code columns in Tables 5-4 and 5-5 show the numbers and percentages of the non-scorable responses from the AI engine and detailed condition codes for these responses assigned by the human scorers (scoring directors).

As shown in Tables 5-4 and 5-5, there was a generally acceptable degree of agreement between the AI engine and the human scorers, with the differences being approximately 1 percent or less for most score points. The exceptions were differences at score point 1 in grades 5, 7, and 8; score point 2 in grade 5; and score point 3 in grade 7. The differences between the AI engine and the human scorers were between approximately 2 and 3 percent in these cases. Greater differences between the AI engine and the human scorers were generally found at score points 1 and 2 compared to the differences at points 3 and 4 at all grade levels. These discrepancies ranged from less than half a percent to about 3 percent difference at score points 1 and 2, and from no practical difference to less than 2 percent difference at score points 3 and 4. It was observed that the percentages of scores 1 and 2 (combined) and scores 3 and 4 (combined) were comparable between the AI engine and the human scorers across all grades.

## 5.5 Items Suppressed from Scoring

There was a testing irregularity involving the ELA Listening Passage in Session 3 in grades 3, 4, and 6 when using an iPad with iOS version 15.4 and 15.4.1. Students who tested on the iPad with this version of iOS were not able to listen to one of the two operational listening passages due to the audio file size. There were three items associated with each of these passages. These items were considered "not presented" and were suppressed from scoring for the affected students. Scale scores, performance level scores, and SPI scores for these students were computed based on the remaining operational items on the test. Approximately between 300 and 400 students per grade were affected by this testing irregularity.

## 5.6 Summary

Taken together, the information presented in this part of the Technical Report summarizes the scoring procedures for different types of items and the steps taken by DRC to ensure accuracy in the TE item scoring, AI scoring, and handscoring processes. The score distribution statistics from the AI engine and the human scorer presented in Section 5.4 demonstrate that the items were scored reliably during the scoring process. These efforts by DRC follow multiple best practices of the testing industry and support AERA, APA, & NCME (2014) Standard 4.18, as presented in Part 5.

## Table 5-1. TDA Item Scoring Guidelines, Grades 3–8

| Score Value | Score Description | Scoring Rubrics |
|---|---|---|
| 4 | Demonstrates effective analysis of text and skillful writing | • Effective addressing of all parts of the task to demonstrate an in-depth understanding of the text(s)<br>• Strong organizational structure and focus on the task with logically grouped and related ideas, including an effective introduction, development, and conclusion<br>• Thorough analysis based on explicit and implicit meanings from the text(s) to support claims, opinions, and ideas<br>• Substantial, accurate, and direct reference to the text(s) using an effective combination of details, examples, quotes, and/or facts<br>• Substantial reference to the main ideas and relevant key details of the text(s)<br>• Skillful use of transitions to link ideas within categories of textual and supporting information<br>• Effective use of precise language and domain-specific vocabulary drawn from the text(s)<br>• Few errors, if any, in sentence formation, grammar, usage, spelling, capitalization, and punctuation that do not interfere with meaning |
| 3 | Demonstrates adequate analysis of text and appropriate writing | • Adequate addressing of all parts of the task to demonstrate a sufficient understanding of the text(s)<br>• Appropriate organizational structure and focus on the task with logically grouped and related ideas, including a clear introduction, development, and conclusion<br>• Clear analysis based on explicit and implicit meanings from the text(s) to support claims, opinions, and ideas<br>• Sufficient, accurate, and direct reference to the text(s) using an appropriate combination of details, examples, quotes, and/or facts<br>• Sufficient reference to the main ideas and relevant key details of the text(s)<br>• Appropriate use of transitions to link ideas within categories of textual and supporting information<br>• Appropriate use of precise language and domain-specific vocabulary drawn from the text(s)<br>• Some errors in sentence formation, grammar, usage, spelling, capitalization, and punctuation that seldom interfere with meaning |
| 2 | Demonstrates limited analysis of text and inconsistent writing | • Inconsistent addressing of some parts of the task to demonstrate a partial understanding of the text(s)<br>• Weak organizational structure and focus on the task with ineffectively grouped ideas, including a weak introduction, development, and/or conclusion<br>• Inconsistent analysis based on explicit and/or implicit meanings from the text(s) that ineffectively supports claims, opinions, and ideas<br>• Limited and/or vague reference to the text(s) using some details, examples, quotes, and/or facts<br>• Limited reference to the main ideas and relevant details of the text(s)<br>• Limited use of transitions to link ideas within categories of textual and supporting information<br>• Inconsistent use of precise language and domain-specific vocabulary drawn from the text(s)<br>• Errors in sentence formation, grammar, usage, spelling, capitalization, and punctuation that may interfere with meaning |

**Table 5-1 (continued). TDA Item Scoring Guidelines, Grades 3–8**

| Score Value | Score Description | Scoring Rubrics |
|---|---|---|
| 1 | Demonstrates minimal analysis of text and inadequate writing | • Minimal addressing of part(s) of the task to demonstrate an inadequate understanding of the text(s)<br>• Minimal evidence of an organizational structure and focus on the task with arbitrarily grouped ideas that may or may not include an introduction, development, and/or conclusion<br>• Minimal analysis based on the text(s) that may or may not support claims, opinions, and ideas<br>• Insufficient reference to the text(s) using few details, examples, quotes, and/or facts<br>• Minimal reference to the main ideas and relevant details of the text(s)<br>• Few, if any, transitions to link ideas<br>• Little or no use of precise language or domain-specific vocabulary drawn from the text(s)<br>• Many errors in sentence formation, grammar, usage, spelling, capitalization, and punctuation that often interfere with meaning |

## Table 5-2. TDA Item Non-scorable Codes, Grades 3–8

| Non-scorable Code | Definition/Example/Notes |
|---|---|
| B—Blank | A response that is completely blank. This includes responses that<br><br>• are completely erased (so that words are unreadable).<br>• are completely crossed out (so that words are unreadable).<br>• are online and consist solely of "white space" (e.g., spaces, tabs, returns). |
| R—Refusal | A response that indicates a refusal to attempt the task. This includes the following examples:<br><br>• *"I don't care"; "I'm not taking this test"; "This is stupid"; "I won't do it"; "you can't make me answer this question"*<br>• *"I don't know"; "IDK"; "we never learned this"; "X"; "NA"*<br>• *Unrelated song lyrics/rap lyrics/poetry (e.g., the lyrics to "Hotel California" in answer to a writing prompt asking whether backpacks should be allowed in class)*<br>• *Intentionally off-task response (e.g., a detailed description of what the student ate for breakfast that morning in answer to a question about Mozart's childhood)*<br><br>This also includes responses that consist solely of scribbles, random keystrokes ("*yyyyyyy*"; "*av:aeoiahvb*"; "*e, hhrrttuuvv*"), indecipherable writing/keystrokes ("*swensts mengetstets arawnstets*") emoticons, stray marks, doodles, drawings, circles, underlines, a couple of random letters (not a word), or other evidence that no attempt was made to address the task. |
| N—Non-scorable | This category includes<br><br>• responses written entirely in a language other than English.<br>• responses that are completely illegible due to poor handwriting.*<br>• online or typed responses that are incoherent due to consisting of incomprehensible strings of words that are not clearly a Refusal or Off Topic (e.g., "*best day school teacher inspired so I car*")<br>• responses too insufficient to be assessed by the criteria on the rubric.<br>• (for TDAs only) responses that address some part of the question but do not contain any logical/accurate/relevant reference to the passage(s) or any ideas contained in the passage(s).<br>• (for TDAs only) responses that consist solely, or almost solely, of text copied directly from the passage(s).<br><br>* If a response is difficult to read, every effort is made to read the response. Multiple people, including a team leader and/or a scoring director, will attempt to decipher the response, and the original answer document will be reviewed if necessary. If, ultimately, only a portion of the response is legible, that verbiage will be scored on its own merits. |
| T—Off Topic | A response makes no reference to the item or (if applicable) the passage provided but does not seem to constitute an intentional refusal.<br>If any part of the response relates to the item in any way, score the response. |
| C—Copied Item/ Directions | A response consists of text copied from the item and/or test directions. |

Note: Crossed out but legible/partially legible responses are scored according to the rubric based on whatever verbiage is legible.

**Table 5-3. TDA Item Score Distribution**

| Grade | Item Number | Total Count | Item Score 1 | 2 | 3 | 4 | Non-scorable Code B | C | N | R | T |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 58203 | 38323 | 13408 | 1231 | 39 | 358 | 112 | 3745 | 807 | 180 |
| 4 | 1 | 58712 | 36620 | 15753 | 3255 | 96 | 183 | 44 | 2199 | 502 | 60 |
| 5 | 1 | 58963 | 35273 | 17482 | 3963 | 86 | 152 | 31 | 1634 | 311 | 31 |
| 6 | 1 | 60026 | 30567 | 21188 | 5013 | 680 | 214 | 38 | 1715 | 561 | 50 |
| 7 | 1 | 61747 | 21313 | 27373 | 8960 | 1764 | 228 | 26 | 1623 | 419 | 41 |
| 8 | 1 | 62565 | 20112 | 27162 | 10585 | 2730 | 242 | 16 | 1029 | 646 | 43 |

**Table 5-4. TDA Item Score Distribution: AI Engine vs. Human Scorer**

| Grade | Scorer | Total Count | Score Count 1 | 2 | 3 | 4 | Non-scorable Code Count B | C | N | R | T |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Scorer 1 (AI Engine) | 15792 | 8927 | 1761 | 165 | 36 | | | 4903 | | |
| 3 | Scorer 2 (Human) | 15792 | 9040 | 1614 | 192 | 43 | 46 | 112 | 3748 | 816 | 181 |
| 4 | Scorer 1 (AI Engine) | 10926 | 5863 | 1787 | 410 | 21 | | | 2845 | | |
| 4 | Scorer 2 (Human) | 10926 | 5958 | 1753 | 339 | 31 | 26 | 44 | 2205 | 510 | 60 |
| 5 | Scorer 1 (AI Engine) | 10222 | 5738 | 1975 | 378 | 86 | | | 2045 | | |
| 5 | Scorer 2 (Human) | 10222 | 5555 | 2197 | 334 | 91 | 27 | 31 | 1643 | 313 | 31 |
| 6 | Scorer 1 (AI Engine) | 11511 | 5443 | 2838 | 699 | 118 | | | 2413 | | |
| 6 | Scorer 2 (Human) | 11511 | 5316 | 2915 | 751 | 116 | 23 | 38 | 1732 | 570 | 50 |
| 7 | Scorer 1 (AI Engine) | 15461 | 4390 | 6274 | 2202 | 428 | | | 2167 | | |
| 7 | Scorer 2 (Human) | 15461 | 4882 | 6130 | 1933 | 349 | 31 | 26 | 1644 | 425 | 41 |
| 8 | Scorer 1 (AI Engine) | 9228 | 3077 | 3036 | 1049 | 259 | | | 1807 | | |
| 8 | Scorer 2 (Human) | 9228 | 3242 | 2959 | 1000 | 220 | 34 | 17 | 1048 | 664 | 44 |

Note: TDA items are weighted x 2 in computation of student scores.

**Table 5-5. TDA Item Percentage Score Distribution: AI Engine vs. Human Scorer**

| Grade | Scorer | Total Count | Score Percentage | | | | Non-scorable Code Percentage | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | B | C | N | R | T |
| 3 | Scorer 1 (AI Engine) | 15792 | 56.53 | 11.15 | 1.04 | 0.23 | | | 31.05 | | |
| | Scorer 2 (Human) | 15792 | 57.24 | 10.22 | 1.22 | 0.27 | 0.29 | 0.71 | 23.73 | 5.17 | 1.15 |
| 4 | Scorer 1 (AI Engine) | 10926 | 53.66 | 16.36 | 3.75 | 0.19 | | | 26.04 | | |
| | Scorer 2 (Human) | 10926 | 54.53 | 16.04 | 3.10 | 0.28 | 0.24 | 0.40 | 20.18 | 4.67 | 0.55 |
| 5 | Scorer 1 (AI Engine) | 10222 | 56.13 | 19.32 | 3.70 | 0.84 | | | 20.01 | | |
| | Scorer 2 (Human) | 10222 | 54.34 | 21.49 | 3.27 | 0.89 | 0.26 | 0.30 | 16.07 | 3.06 | 0.30 |
| 6 | Scorer 1 (AI Engine) | 11511 | 47.29 | 24.65 | 6.07 | 1.03 | | | 20.96 | | |
| | Scorer 2 (Human) | 11511 | 46.18 | 25.32 | 6.52 | 1.01 | 0.20 | 0.33 | 15.05 | 4.95 | 0.43 |
| 7 | Scorer 1 (AI Engine) | 15461 | 28.39 | 40.58 | 14.24 | 2.77 | | | 14.02 | | |
| | Scorer 2 (Human) | 15461 | 31.58 | 39.65 | 12.50 | 2.26 | 0.20 | 0.17 | 10.63 | 2.75 | 0.27 |
| 8 | Scorer 1 (AI Engine) | 9228 | 33.34 | 32.90 | 11.37 | 2.81 | | | 19.58 | | |
| | Scorer 2 (Human) | 9228 | 35.13 | 32.07 | 10.84 | 2.38 | 0.37 | 0.18 | 11.36 | 7.20 | 0.48 |

Note: TDA items are weighted x 2 in computation of student scores.

# Part 6: Psychometric Analyses

This part of the Technical Report describes the analyses that were conducted with the ELA, Mathematics, Science, and Social Studies operational test data. These analyses included a classical item analysis and examination of the raw scores and an item response theory (IRT) analysis involving test calibration, scaling, and equating. These analyses were conducted using the calibration samples.

## 6.1 Overview of the Operational Test Data Analysis

This part of the Technical Report, the classical item statistics, including aggregate raw score statistics and individual item-level statistics are presented first. Next, the analyses involving test calibrating, equating, scaling, and student scoring that occurred for the Wisconsin Forward Exam after the 2022 test administration are described. The calibration samples are presented, followed by the data calibration results, including the model-data fit for the Wisconsin Forward Exam data. If the IRT models fit the empirical item response distributions for the population (i.e., Wisconsin students) for which generalizations are made, then the claim is strengthened that the scores are valid indicators of an underlying ability. The lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) for the Wisconsin Forward tests are also presented.

Part 6 demonstrates adherence in the Wisconsin Forward Exam program data analysis to AERA, APA, & NCME (2014) Standards 1.8, 4.14, 5.2, 5.13, 5.15, and 7.2. Each standard will be explicated within the appropriate section of this part. Standard 7.2 provides general guidance that is relevant to this part:

> The population for whom a test is intended and specifications for the test should be documented. If normative data are provided, the procedures used to gather the data should be explained; the norming population should be described in terms of relevant demographic variables; and the year(s) in which the data were collected should be reported. (p. 126)

## 6.2 Classical Item Analysis: Item Level Statistics

Three statistics are frequently used in item analysis: the proportion correct ($p$-value), the item-total correlation coefficient, and the omit rate for the item.

The $p$-value is an indication of the difficulty of an item. The $p$-value for an MC item or any item with a maximum score of 1 represents the proportion of students who answered the item correctly. If all students answered a given item correctly, its $p$-value would be 1.0. If only 30% of students answered the question correctly, the $p$-value would be 0.30. The lower the $p$-value is, the more difficult the item is. Item $p$-value is a good indication of difficulty, as it takes student performance into account and it makes comparing items in terms of a common statistic very simple. A test made up of items well distributed across the range of item difficulty levels is desirable because it supports the assessment of students at all ability levels.

The $p$-value for an item worth more than 1 point (e.g., EBSR item type) represents the mean proportion of possible raw score points that students actually obtained for the item. A $p$-value of 0.33 for an item with a maximum item score greater than 1 would indicate that, on average, students obtained one-third of the possible points for the item. If a $p$-value were 0.75, this would indicate a much easier item where, on average, students scored 75% of the maximum possible points for the item. Therefore, the $p$-value indicates difficulty for such items as well, with lower $p$-values indicating more difficult items.

The item-total correlation indicates the extent to which individual test items provide reliable measurement of the construct being measured by the total test, and it is an index of the item's ability to discriminate between high-ability and low-ability students. For dichotomously scored items, the item-total correlations are computed as point-biserial correlations between the score on the item and the score on the remaining items in the test. For multi-point items, the item total correlations are computed as Pearson product-moment correlations between the score on the item and the score on the remaining items in the test.[1] The item-total correlation coefficients can range from -1.0 to +1.0. A large positive value (such as 0.40) indicates a strong relationship between a score on an individual item and the total score, with students who earn high scores on the total test tending to score higher on the item than students with low scores on the total test. A low positive value (such as 0.10) indicates a weak relationship between scores on the item and the total score, while a negative value indicates that students who do well on the total test tend to score lower on the item than students who do poorly on the total test.

For MC items, the point-biserial correlation between each distractor and the total score was also calculated. In most cases, items will have negative correlations for each distractor and the total score. However, a weak positive correlation for a distractor does not necessarily mean that the item is defective, provided that the distractor correlation is substantially smaller than the item-total correlation for the correct response. In some cases, it may simply mean that the particular distractor is attractive to moderate-ability students and unattractive to low-ability students.

The omit rate is also computed for each item, reflecting the percentage of students who did not respond to the item. A high omit rate can indicate an especially difficult item or, if located near the end of the test, it can indicate what is referred to as a "speeded" test, where students have insufficient time to respond to all items.

The examination of omit rates complies with AERA, APA, & NCME (2014) Standard 4.14. This standard is concerned with the speededness of a test:

> For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure. (p. 90)

For the Spring 2022 Wisconsin Forward Exam, items were flagged for further investigation in the following situations:

- The *p*-value was less than 0.20. Such a *p*-value indicates a difficult item, where fewer than 20% of students obtained the correct answer.

- The item-total correlation was less than 0.15 for the correct answer. A low value may indicate that the item is not providing a high degree of discrimination between high-ability and low-ability students, and, in addition, it may be an indication that the correct answer is in question.

- A distractor had a positive correlation with the total test score.

- The omit rate was greater than 3%.

---

[1] For both the point-biserial and the Pearson correlations, the studied item is excluded from the computation of the total score so as to not artificially inflate the correlation statistic. This effect would be most noticeable for items worth several points.

Flagging an item for investigation is just one aspect of a complete evaluation of an item, and flagged items are not necessarily defective. It is desirable to include a small number of items with very high $p$-values (easy items) or very low $p$-values (difficult items) in order to provide more reliable measurement at the extreme high and low levels of ability and to fully represent the range of difficulty for particular content standards. In this case, the flagging of $p$-values is a useful way of verifying that the number of extremely easy or difficult items is relatively small and consistent with the purposes of the test. Thus, flagged items do not necessarily indicate a challenge to test validity, because items have been found to be appropriate during item reviews.

Omit rates may reflect a number of different properties, and an item that is omitted by more than 3 percent of students (the Wisconsin Forward Exam flagging criterion) is not necessarily problematic. Omit rates are often higher for non-MC items than for MC items because students who are fairly certain they do not know the answer may be inclined to simply skip the item altogether rather than taking the time to form a response. Items with high omit rates are referred to content specialists for further review to ensure there is no unintended ambiguity in the items. If these flagged items are judged to be clear and provide a valid measurement of the intended knowledge, skill, or ability, then they are retained on the test.

Items flagged for a low item-total correlation or for a positive distractor-total test correlation are more troublesome because these statistics show the relationship of each option to the construct being measured. In determining whether these items should be retained or removed from scoring, it is important to consider the relative magnitude of the correlation between the correct response and the total score and between the distractor and the total score. In most cases, removing an item with a modest item-total correlation and negative correlations for all of the distractors will actually lower the reliability of the total test, so it is generally preferable to retain these items. The same is true of an item with a small positive correlation for one of the distractors and a much larger positive correlation for the correct response. However, an item that exhibits a low correlation for the correct response in combination with a positive correlation for one or more distractors is likely to degrade the accuracy of the measurement and lower the reliability of the test. Such items should be removed from scoring.

Overall, 49 operational items across all Wisconsin assessments were flagged on the Spring 2022 operational tests as meeting one of the investigational criteria bulleted above. More flagged items were found in Mathematics grades than in ELA, Science, or Social Studies.

Table 6-1 shows the number of scored items in the Spring 2022 Wisconsin Forward Exam operational tests flagged for these conditions by grade and content area. Because some items were flagged for more than one condition, the number of flags may be greater than the number of flagged items.

The flagged items were referred to DRC's content specialists for further review to ensure that the items were unambiguous and the answer keys were correct. As part of this review, DRC's content experts also evaluated each flagged item against the Wisconsin Forward Exam depth-of-knowledge criteria to ensure that the cognitive demands of the item reflected the skills and knowledge that the item was designed to measure. Tables 6-2, 6-3, and 6-4 provide more information about the flagged items.

### 6.2.1 Flagging for a Positive Distractor Correlation

In Tables 6-2 through 6-4, the distractor correlation coefficients are provided for items that were flagged because of positive distractor correlations. The distractor correlations tend to be small and are generally much smaller than the item-total correlations for the correct answer. The majority of items flagged for a positive distractor-total test correlation had a distractor total test correlation close to 0 and an acceptable item-total test correlation for the correct answer. More items were flagged for positive distractor correlation in Mathematics than in other content areas. All flagged items were judged to be acceptable based on their content and other statistics and were retained in order to meet the Wisconsin Forward Exam test blueprints.

### 6.2.2 Flagging for the Item-Total Correlation

One item per grade was flagged for item-total test correlation <0.15 in Mathematics grades 4 and 5 and Science grades 4 and 8. The item-total test correlations for the flagged items ranged from 0.04 to 0.14.

### 6.2.3 Flagging for *p*-Value

One item per grade in Mathematics grades 4 and 5 and two items per grade in Mathematics grades 6 through 8 were flagged for *p*-values <0.20. The flagged items had *p*-values between 0.14 and 0.19. While these statistics indicate items that were difficult, the number of items flagged for difficulty was found to be reasonable, given that learning disruptions related to the COVID-19 pandemic likely occurred in the 2020–21 school year and the post-pandemic recovery. No operational items were flagged for difficulty in ELA, Science, or Social Studies.

### 6.2.4 Flagging for Omit Rate

No operational items on the Wisconsin Forward Exam were flagged for an omit rate of higher than 3%. Most of the items had omit rates of less than 1%.

### 6.2.5 Speededness

The degree to which a test is speeded can be evaluated by examining the percentage of students who fail to respond to the final items on a test or the last items in a timed section. One criterion of test speededness currently in use in the testing industry is a rule introduced by Educational Testing Services, which stipulates that at least 80% of test takers should be able to answer all of the items and all test takers should be able to answer at least 75% of the items (Swineford, 1956). However, a more stringent requirement is often applied, considering tests to be non-speeded only if at least 95% of examinees attempt the final item. As shown in Table 6-5, the Wisconsin Forward Exam satisfies this more stringent requirement, with approximately 99% of the examinees attempting the final item in each of the four content areas.

### 6.2.6 Supplemental Tables on Classical Item Analysis

Tables 6-6 through 6-22 present more comprehensive results from the classical item analysis for all the items retained in each grade and content area. In those tables, the item-total test correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the total test score, the omit rate is flagged when it is above 3 percent, and the *p*-value is flagged when it is below 0.20.

Tables 6-6 through 6-22 also show the item numbers, which can be used to understand the location of test items as students actually encountered them on the test. The item analysis tables also indicate item type (e.g., MC, EBSR).

The numbers of flagged items across grade and content areas are summarized in Table 6-1. As indicated above, relatively few items were flagged. The item analysis indicated that the p values of the items in the operational tests were well distributed throughout the range of difficulty levels, with reasonably high point-biserial correlations for most items. Detailed item analysis results including distractor statistics for MC items and score point distributions for non multiple choice items are included in Appendix H.

## 6.3 Test-Level Statistics

Test-level statistics, including test reliability, were computed for the Spring 2022 Wisconsin Forward Exam data for students with complete operational test data. These statistics are presented in Table 6-23. To facilitate interpretation of the test-level statistics, Table 6-23 provides the maximum possible score, the number of students, a measure of test difficulty, the standard deviation (SD) of raw scores, the skewness of the raw score distribution, the kurtosis, the minimum obtained score, the maximum obtained score, the reliability (Cronbach's alpha), and the standard error of measurement (SEM) for raw scores. These measurements are further explained below. Readers can refer to Tables 3-6 through 3-9 for a count of the number of items in the test and the number of score points corresponding to each test.

The mean raw score varies by grade and content area and, specifically, in the context of the maximum possible score points. In ELA, for example, the maximum possible raw score is 53 in grade 3 and 56 in grades 4 through 8. In Mathematics, the maximum possible raw score is 42 in grade 3 and 46 in grades 4 through 8. The maximum possible raw score is 40 in both Science grades and in all Social Studies grades.

Test difficulty is computed as the mean raw score divided by the maximum possible score points. Test difficulty ranges from 0 to 1.0. A larger test difficulty value indicates a mean raw score that is closer to the maximum possible score and, therefore, indicates an easier test. A smaller test difficulty value indicates a mean raw score that is further from the maximum possible score and, therefore, indicates a more difficult test. Consider an example: A test difficulty statistic would be 0.90 if a mean score of 45 were obtained on a test with a maximum possible score of 50. This would be considered an easier test. On the other hand, test difficulty would be 0.50 if a mean raw score of 25 were obtained on the same test. This would then be considered a more difficult test. For example, the Mathematics grade 3 test mean raw score is 22.75 and the maximum possible score is 42, resulting in the test mean *p*-value of approximately 0.54. Evaluation of the mean *p*-values indicates that with the exception of Mathematics grade 3, Mathematics tests were more difficult for students than ELA, Science, and Social Studies tests.

Table 6-23 also shows the skewness and kurtosis statistics for each distribution of raw scores. Skewness and kurtosis describe the shape of a distribution. When a distribution is perfectly normal, skewness is zero. A negative skew has a long tail on the left side of the distribution because of the presence of some low scores, and, because the mean is sensitive to extreme scores, it indicates that most student scores are clustered on the high end of the scale. A positive skew indicates a distribution with some very high scores and a larger number of scores below the mean. Kurtosis describes a distribution in terms of its shape relative to a perfectly normal distribution. When a distribution is perfectly normal, kurtosis is zero. A negative kurtosis statistic indicates a distribution that is flatter than a perfectly normal curve, and a positive kurtosis statistic indicates a distribution that has more scores in the center of the score distribution (making it peaked) than a perfectly normal curve. Table 6-23 reveals that, in most cases, Wisconsin Forward Exam students are not normally distributed along the test scale in each grade and content area. Although this has implications for practitioners who wish to use Wisconsin Forward Exam raw scores in statistical analyses (normality of the data cannot be assumed), from a criterion-referenced testing standpoint, it indicates that students on the whole are mastering the Wisconsin Standards for ELA, Science, and Social Studies. The

Mathematics assessments in grades 4 through 8 tended to be more difficult, however, showing most of the scores clustered below the mean (as indicated by positively skewed score distributions).

In addition, Table 6-23 shows that the minimum obtained score in all grades and content areas, except for ELA grade 6, was zero, meaning that at least one student failed all items for each of those tests. With the exception of ELA grades 3, 4, and 7, the maximum obtained scores were equal to the maximum number of points possible on the test in all grades, meaning that at least one student obtained the full score for all items on each of those tests. For example, as displayed in Table 6-23, in Mathematics grade 3, there was at least one student who failed all items and at least one student who obtained the maximum raw score of 42.

A reliable test is one with high reliability as represented by statistics such as Cronbach's alpha and a low SEM. When interpreting reliability statistics, readers should note that test length (number of items and score points) is one of the important factors that influences reliability statistics and SEM. These concepts are described further in Part 8. For present purposes, the reader should note that measurement error is associated with every test score. A student's true score is the hypothetical average score that would result if the test could be administered repeatedly without the effects of practice or fatigue. Obtained scores should not be regarded as absolute but as one point within a range that, with a certain degree of probability, includes a student's true score.

The test-level statistics for each content area are summarized and discussed below using the measurements described above.

## English Language Arts

- Test difficulty ranged from 0.54 to 0.57.
- Reliability coefficient Alpha was relatively high in every grade (0.88 to 0.90).
- SEM ranged from 3.28 to 3.50.

## Mathematics

- Test difficulty ranged from 0.43 to 0.54, with generally lower difficulty in lower grades and higher difficulty in higher grades.
- Reliability coefficient Alpha was high in every grade (0.92 to 0.93).
- SEM ranged from 2.68 to 2.90.

## Science

- Test difficulty was 0.54 in grade 4 and 0.52 in grade 8.
- Reliability coefficient Alpha was 0.89 in grade 4 and 0.88 in grade 8.
- SEM was 2.72 and 2.80 for grades 4 and 8, respectively.

## Social Studies

- Test difficulty ranged from 0.53 to 0.59.
- Reliability coefficient Alpha was 0.91 in grades 4 and 8 and 0.90 in grade 10.
- SEM ranged from 2.68 to 2.77.

## 6.4 Item Response Theory Methodology

This section of the report outlines the item response theory (IRT) methodology including item calibration, test equating and test scaling, as well as the methodology of computation of the scale scores based on Wisconsin Forward Exam test data. Readers should note that calibration, equating, and scoring using IRT are mathematically complex and computationally intensive processes. A full understanding of these topics requires a background in psychometrics. However, in order to make these processes more accessible and transparent to a wider range of audiences, a brief, nontechnical explanation of the IRT process and how scale scores are derived from student raw responses is provided. Additional references are also suggested to interested readers.

### 6.4.1 Item Calibration

This section of the report outlines the calibration procedures and results for the Spring 2022 Wisconsin Forward Exam. Student responses on the Wisconsin Forward Exam are inputted into complex mathematical algorithms designed to model the relationship between a student's ability in a content area and a test item. The group of algorithms is collectively known as item response theory (IRT). Wisconsin Forward Exam scores are established through the processes of calibration, scaling, and item-pattern scoring.

Calibration is the mathematical process of estimating characteristics of individual items. These characteristics are termed "item parameters." Section 6.4.1 serves to explain this process, beginning with a description of the calibration methods that were applied to the Spring 2022 Wisconsin Forward Exam, followed by a presentation of a calibration sample, and a discussion of the calibration models and the software used. The results of the calibration process, using model-to-data fit statistics, and the outcomes of test scaling are also discussed in this section.

### 6.4.1.1 Calibration Models

The three-parameter logistic (3PL) model and the two-parameter partial credit (2PPC) IRT model (Bock & Aitkin, 1981; Thissen, 1982) were used to estimate parameters for multiple choice (MC) items and constructed-response (CR) items, respectively. All non MC items, including technology-enhanced (TE) items, evidence-based selected response (EBSR) items, short-answer (SA) items, and text-dependent analysis (TDA) items, were treated as CR items in calibrations. Item parameters for items contained in all Wisconsin assessments were estimated using a marginal maximum-likelihood procedure.

Under the 3PL model, the probability that a student with a trait or scale score $\theta$ will respond correctly to MC item $j$ is

$$P_j(\theta) = c_j + (1 - c_j)/[1 + exp(-1.7a_j(\theta - b_j))]$$

In the equation, $a_j$ is the item discrimination, $b_j$ is the item difficulty, and $c_j$ is the probability of a correct response by a very low-ability student. Under the 2PPC model, the probability that a student with a trait or scale score $\theta$ will respond in category $k$ to partial-credit item $j$ is

$$P_{jk}(\theta) = exp(z_{jk})/\sum_{i=1}^{m_j} exp(z_{ji})$$

where $z_{jk} = (k-1)f_j - \sum\limits_{i=0}^{k-1} g_{ji}$ and $g_{j0} = 0$ for all $j$.

The summary output of the 3PL and 2PPC models is in two different metrics. The discrimination and location parameters for the MC items are in the traditional 3PL metric and are labeled $a$ and $b$, respectively. In the 2PPC model, $f$ (alpha) and $g$ (gamma) are analogous to $a$ and $b$, where alpha is the discrimination parameter and gamma over alpha ($g/f$) is the location in which adjacent trace lines cross on the ability scale. Because of the different metrics used, the 3PL parameters $a$ and $b$ are not directly comparable to the 2PPC parameters $g$ and $f$; however, they can be converted to a common metric. The two metrics are related by $a = f/1.7$ and $b = g/f$ (Burket, 2002). As a result of this procedure, the MC and CR items are placed on the same scale. Note that for the 2PPC model, there are $m_j - 1$ (where $m_j$ is a score level $j$) independent $g$'s and one $f$, for a total of $m_j$ independent parameters estimated for each item, while there is one $a$ and one $b$ per item in the 3PL model.

Using the 3PL/2PPC models for estimation of ELA, Mathematics, Science, and Social Studies, item parameters were consistent with the past methodology (except for the 2014–15 administration for ELA and Mathematics) implemented for Wisconsin assessments. Item parameters estimated after the 2021–22 test administration were used to score the responses of Wisconsin students who took these tests.

### 6.4.1.2 Calibration Sample

The calibration of the Wisconsin Forward Exam occurred after the Spring 2022 test administration and was based on the student data acquired during the entire testing window. This section provides information on the comparability of the calibration sample to the census data in terms of demographic characteristics in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) Standards:

> The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics. (p. 25)

The calibration samples consisted of the student data acquired during the entire testing window and included records with complete operational test data. That is, to be included in the sample, a student had to respond to all test questions and not have any missing responses. Students from public, choice, and private schools were included in the calibration data. Students who had non-tested codes indicated in the data were excluded from the calibration samples. Students in grades 3, 4, and 6 who experienced iPad/iOS issues that affected certain listening passages were also excluded from the ELA calibration samples and subsequent data analyses. The characteristics of the calibration samples are presented in Tables 6-24 through 6-27 for ELA, Mathematics, Science, and Social Studies, respectively. The grade 3 through 8 calibration samples consisted of more than 99 percent of students who later received valid test scores in these grades across all content areas. The grade 10 calibration sample consisted of approximately 98% of students who later received valid test scores. In addition, calibration samples for grades 3 through 7 consisted of approximately 94% to 95% of all students scheduled to participate in the assessment. The grade 8 calibration samples consisted of approximately 93% of all students scheduled to take the test across all content areas. The grade 10 calibration sample for Social Studies included approximately 86% of all students scheduled to participate in the assessment. These percentages are reported in Tables 6-24 through 6-27 in the top row ("All Students") for each grade.

When the calibration sample characteristics were compared with the characteristics of students who later received valid test scores, it was found that the differences across subgroups were less than a quarter of a

percent for any subgroup across all grades and content areas. When the calibration sample characteristics were compared to the characteristics of all students scheduled to test, it was found that, with few exceptions, the differences across subgroups were less than one percent across all content areas in grades 3 through 8. The exceptions were differences in the percentages of students with disabilities included in the calibration samples compared to the percentages of students with disabilities scheduled to test. These differences were just over 1% in grades 6 through 8 across all content areas. Larger differences between the calibration sample and the population of students scheduled to test were found for the grade 10 Social Studies. White students and students with disabilities were over-represented by over 3% in the calibration sample compared to the population of students scheduled to test. African American and economically disadvantaged students were under-represented by over 2% in the calibration sample compared to the population of students scheduled to test. The difference between the calibration sample and the population of students scheduled to test was less than 1% for all other subgroups for Social Studies grade 10. No adjustment to the calibration sample was made for grade 10.

### 6.4.1.3 Calibration Procedure

The calibrations were conducted separately for each grade level and content area using the marginal maximum-likelihood procedures implemented with the expected maximum algorithm (Bock & Aitkin, 1981; Thissen, 1982). In a process of item calibration, the number of estimation cycles was set to 99 with the convergence criterion of 0.001 for all content areas. The maximum value of $a$-parameter was set to 5.0, and the range for $b$-parameter was set between -7.5 and 7.5. For all items, the estimated $a$- and $b$-parameters were within the prescribed parameter ranges. The $c$-parameters for anchor items were fixed to their Spring 2019 values for ELA, Mathematics, and Science. It should be noted that there was a small number of items with the default value for the $c$-parameter on all tests. When the algorithm, which is implemented to calibrate the items, encounters difficulty estimating the $c$-parameter, it assigns a default $c$-parameter value of 0.20.

### 6.4.1.4 Calibration Software

Calibration of the Wisconsin Forward Exam data was performed using PARDUX software (Burket, 2002; Shu 2020). PARDUX is designed to produce a single scale by jointly analyzing data resulting from students' responses to both MC items and CR items for assessments that include both item types. In PARDUX, items are calibrated based on IRT, using the 3PL model (Lord & Novick, 1968) for MC items and the 2PPC model (Yen, 1993) for CR items.

PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. Extensive simulation studies and comparisons between PARDUX and MULTILOG (Thissen, 1990), PARSCALE (Muraki & Bock, 1991), and BIGSTEPS (Wright & Linacre, 1992) have shown that PARDUX provides precise parameter and ability estimates and performs as well or more efficiently than these programs (Fitzpatrick, 1991; Fitzpatrick and Julian, 1996). Extensive research with simulation data has also shown that the IRT procedures used for calibration and scaling of Wisconsin assessments produce accurate vertical scaling (Yen & Burket, 1997). PARDUX software has been regularly updated by adding advanced estimation algorithms and other features, keeping it up-to-date to face the challenges of large-scale assessments (Shu, 2020).

### 6.4.1.5 Calibration Results

This section describes the calibration results in terms of the estimation of item parameters and model-to-data fit for all content areas and grades.

---

## IRT Item Parameters

During calibration, items may not converge, meaning the characteristics of the items will not be determined. When this occurs, items may be suppressed from student scoring and future assessments. In Spring 2022, no non-convergence issues occurred for any item on the operational tests.

## IRT Item Fit

The calibration process produces ability and item parameter estimates that can be used to predict student response patterns to each item. For example, based on the item parameter estimates for item difficulty and item discrimination, low-ability students are expected to be less likely to answer a difficult and highly discriminating item correctly than higher-ability students. After parameters are produced, the predicted scoring patterns can be compared to the observed scoring patterns in what are referred to as item-to-model fit comparisons. Where there is little difference between the predicted scoring patterns and the observed scoring patterns, the model can be said to "fit" the data.

A procedure developed by Yen (1981) was used to assess model-to-data fit for all test items. In this procedure, students are rank ordered on the basis of their $\hat{\theta}$ values and sorted into ten cells, with 10 percent of the sample in each cell. Each item $j$ in each decile $i$ has a response from $N_{ij}$ examinees. The fitted IRT models are used to calculate an expected proportion $E_{ijk}$ of examinees who respond to item $j$ in category $k$. The observed proportion $O_{ijk}$ is also tabulated for each decile. The fit index for item $i$ is

$$Q_{1j} = \sum_{i=1}^{10} \sum_{k=1}^{m_j} \frac{N_{ij}(O_{ijk} - E_{ijk})^2}{E_{ijk}}.$$

$Q_{1j}$ should be approximately chi-square distributed with degrees of freedom ($DF$) equal to the number of "independent" cells, $10(m_j - 1)$, minus the number of estimated parameters. For the 3PL model, $m_j = 2$, so $DF = 10(2 - 1) - 3 = 7$. For the 2PPC model, $DF = 10(m_j - 1) - m_j = 9m_j - 10$.

DRC evaluated item-to-model fit in a two-step process. First, item-to-model fit information was obtained for each item using a $Z$-statistic. The $Z$-statistic is an index of the degree to which obtained proportions of students with each item score match the proportions predicted by the estimated student ability and item parameters. When the difference between the obtained proportions of students with each item score and the proportions predicted by the estimated student ability and item parameters reached a certain threshold, the item was flagged for "misfit."

The $Z$-statistic is a transformation of the chi-square ($Q_1$) statistic that takes into account differing numbers of score levels as well as sample size using the equation

$$Z_j = \frac{(Q_{1j} - DF_j)}{\sqrt{2DF_j}},$$

where $Q_{1j}$ is the item chi-square statistic, $j$ is an item, and $DF$ is the degrees of freedom for a given item $j$.

Because the value of $Z$ increases as the sample size increases, the critical values for $Z$ were established using the following equation (Yen & Candell, 1991):

$$Z_{crit,j} = \frac{4N_j}{1500},$$

where $Z$ crit, $j$ is the critical value of $Z$ for item $j$ and $N_j$ is the number of students who responded to item $j$. These values and the associated chi-squares ($Q_1$) are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984).

Table 6-28 presents items that were flagged for less than optimal fit when the obtained $Z$ statistic exceeded the critical $Z$-statistic value. This table specifies the content area, grade level, item number in the calibration, item type (MC or CR), $N$ size (i.e., the number of students who took this item), $Z$, and critical $Z$ as described previously. Twenty-three items were flagged for poor fit for ELA, six items were flagged for Mathematics, four items were flagged for Science, and five items were flagged for Social Studies. Most of the flagged items were CR items (TE and EBSR). For example, item #14 for ELA grade 3 was flagged because the observed $Z$ of 201.16 was larger than the critical $Z$ value of 154.29 based on a sample size of 57,860. For many of the flagged items, the observed $Z$ and the critical $Z$ were not very far apart, indicating small or moderate misfit; however, it was observed that for some items, the misfit was larger (e.g., item #28 in ELA grade 4, items #18 and #23 in ELA grade 7, and item #38 in ELA grade 8).

In order to evaluate item-to-model fit further, DRC inspected the observed-to-predicted item characteristic curve (ICC) for each flagged item. These ICCs simultaneously plot the characteristics of an item (e.g., item difficulty, item discrimination, level of guessing) using IRT model predictions and the observed student responses. The ICCs show exactly where along the ability continuum the misfit occurs and the extent of the misfit.

All cases of MC items flagged for misfit had empirical (observed) information that differed from the model in the lower-ability range, where there are fewer students to provide information at the tail end of the distribution. Similarly, for CR items, there were, in general, fewer students at the lower score levels, which provides less information at the tail ends of the student distribution. Items that only show misfit at the tail ends of the distribution provide stable information about the majority of the students—those in the middle range of the distribution. However, if the misfit happens around the middle of the ability range, where there are many students, this may be a concern and may lead to the item being dropped from the item pool.

In a large-scale assessment, such as the Wisconsin Forward Exam, with 17 combinations of grades and content areas, it is expected that some items will be flagged for misfit. As noted, the difference between the obtained Z-statistic and the critical Z-statistic was often small or moderate. Items flagged for misfit were reported to the DRC Test Development team for additional review. Such items are flagged in the Wisconsin Forward Exam item bank and are avoided during the form selection process unless there is a compelling reason that they should be included, such as meeting the test blueprint.

## 6.4.2 Test Equating

Test equating is the statistical process of placing scores from two or more parallel assessments onto a common scale, resulting in direct comparability of scores from two different test forms. A common-item design was used to link the assessments from 2022 to the established ELA, Mathematics, and Science scales for the Wisconsin Forward Exam. Sets of items that were administered to Wisconsin students in previous operational test administrations and that were included in the Spring 2022 assessments served as the anchor sets in each ELA, Mathematics, and Science grade. All anchor items were selected from Spring 2018 or 2019 operational test assessments for ELA and Mathematics. All Science anchor items were selected from Spring 2019 operational tests. The anchor sets constituted at least one-third of the Spring 2022 assessments and were representative of the Spring 2022 test content. After the item calibration, item parameters were linked to the Wisconsin Forward Exam scales using the Stocking & Lord (1983) equating procedure.

Standard 5.13 of the AERA, APA, & NCME (2014) Standards states the following:

When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions were established and on the accuracy of the equating functions. (p. 105)

The Stocking & Lord procedure minimizes the mean squared difference between the two test characteristic curves (TCCs), one based on estimates from the previous calibration and the other based on transformed estimates from the current calibration. Let $\Psi_j$ be the TCC based on estimates from a previous calibration and $\Psi_j^*$ be the TCC based on transformed estimates from the current calibration:

$$\widehat{\Psi}_j = \widehat{\Psi}(\theta_j) = \sum_{i=1}^{n} P_i\left(\theta_j;\ a_i, b_i, c_i\right)$$

$$\widehat{\Psi}_j^* = \widehat{\Psi}(\theta_j) = \sum_{i=1}^{n} P_i\left(\theta_j;\frac{a_i}{A}, Ab_i + B, c_i\right).$$

The TCC method determines the equating constants ($A$ and $B$) by minimizing the following quadratic loss function ($F$):

$$F = \frac{1}{N}\sum_{a=1}^{N}\left(\widehat{\Psi}_j - \widehat{\Psi}_j^*\right)^2.$$

The Stocking & Lord equating procedure is commonly used in large-scale assessments. The standard error of the equating (SEE) is difficult and cumbersome to estimate for IRT equating procedures like the Stocking & Lord procedure (Kolen & Brennan, 1995; Michaelides & Haertel, 2004). The estimation of the SEE is beyond the scope of this report.

### 6.4.2.1 Evaluation of Anchor Items

AERA, APA, & NCME (2014) Standard 5.15 requires information about the anchors, stating the following:

In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented. (p. 105)

Two statistical methods were used to evaluate anchor items: (1) iterative linking (Candell & Drasgow, 1988) using Stocking & Lord's (1983) TCC method and (2) differences between the item-ability regression curves.

### Test Characteristic Curve Method

The Stocking & Lord (1983) procedure, also called the TCC method, for which the mathematical equation was provided in a previous section of this document, minimizes the mean squared difference between the two TCCs, one based on estimates from the previous calibration and the other based on transformed estimates from the current calibration.

Differential item functioning was evaluated by examining previous (input) and transformed (estimated) item parameters. Items with an absolute difference of parameters greater than two times the root mean square deviation were flagged for review. These differences were monitored by plotting input and estimated item parameters.

### Item Response Theory Item-Ability Regression Curves

Differences between the item-ability regression curves of the anchor items in the Spring 2022 Wisconsin Forward Exam administration were also compared to previous calibrations from Spring 2019. The differences between the item curves were evaluated using the following statistics:

- UnWtd Mean = Average signed difference in estimated probability

- UnWtd Mean Abs = Average absolute (unsigned) difference in estimated probability

- UnWtd RMSD = Root mean squared difference

- Wtd Mean = Weighted average signed difference in estimated probability

- Wtd Mean Abs = Weighted average absolute (unsigned) difference in estimated probability

- Wtd RMSD = Weighted root mean squared difference

Both unweighted and weighted versions of these statistics were calculated. Unweighted differences give equal weight to differences across the ability spectrum. Weighted differences assign weights according to the number of test takers that are impacted (that is, the frequency distribution of estimated student abilities during the calibration).

For the six statistics listed above, differences greater than +/-0.10 are considered large and differences between +/-0.07 and +/-0.10 are considered moderate.

Additionally, the maximum absolute difference (Max Abs) was identified. For Max Abs, large differences are those greater than +/-0.15 and moderate differences are all differences between +/-0.125 and +/-0.15.

## 6.4.2.2 Removal of Anchor Items

One of the key requirements of anchor items in deriving valid and reliable linking results is that the anchor items form a miniature of the test in terms of content coverage, or test blueprint. While dropping a flagged anchor item based solely on statistical criteria has its simplicity, this option may change the content coverage and invalidate results. Before an anchor item is dropped from an anchor set, the item characteristics, adequacy of the content coverage, and impact on the size of the anchor set must be evaluated.

An item may be removed from the anchor set only if it adversely affects the quality of scaling, not the desirability of the results. Therefore, DRC does not consider how the removal of an item affects the overall mean scale score or the impact data (i.e., percentage of students in each achievement level) when recommending items for removal.

Items removed from the anchor set are still scored as part of the whole test. DRC recommends that the anchor items be considered for exclusion from the Wisconsin Forward Exam equating sets under the following conditions:

1.  An item may be a candidate for removal if it is flagged for moderate or large differences on at least four of the seven statistics (listed in Section 6.4.2.1) considered when examining the differences between the IRT item-ability regression curves.

2.  Removal of the item will only be considered after alternative explanations have been considered that may explain shifts in performance. For example, performance on the anchor item may improve because of a statewide initiative emphasizing instruction on a particular set of skills. In this case, improved performance on the item represents true growth in that area. Removing the anchor item may artificially lower test scores.

3.  Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of the anchor items across the content standards should remain within 10 percent of the Wisconsin Forward Exam test blueprint.

4.  The number of remaining items will remain at an acceptable level of anchor set reliability. Operationally, this means the anchor set will still be representative of the total test blueprint and the anchor set may not be less than 20 percent of the total test length.

Flagged items are reviewed by DRC test development experts to verify that no changes to item content or format occurred between the administration in which the anchor items were used and the current administration. In addition, for the flagged non-MC anchor items, verification that no changes to scoring rubrics occurred between the two administrations is performed.

## 6.4.2.3 Evaluation of Equating Results

Table 6-29 provides equating results for the TCC method for ELA, Mathematics, and Science. This table summarizes the following information for each grade and content area: number of anchors, number of iterations, quadratic loss function ($F$), correlation between the $a$-parameter input and estimates, correlation between the $b$-parameter input and estimates, number of $a$- and $b$-parameter outliers as indicated by the root mean square deviation method, and equating constants ($A$ and $B$). Note that two sets of equating

results are included for Science grade 8 due to the exclusion of one anchor item from equating in this grade.

The overall alignment of the anchor TCCs was very good for all grades in the three content areas. Figures 6-1 through 6-3 show the TCC alignment of the anchor set before and after equating for all grades of ELA, Mathematics, and Science. In these figures, the input anchor set TCC (before equating) is indicated by the dashed red line and the new anchor estimate TCC (after equating) is indicated by the solid blue line. The correlations between the *a*-parameter input and estimates were at least 0.91 or higher, and the correlations between the *b*-parameter input and estimates were 0.97 or higher for all grades and content areas. One anchor item was flagged as an *a*-parameter outlier in ELA grades 3 and 4 and in Mathematics grades 3, 5, and 8. Two anchor items were flagged as *a*-parameter outliers in ELA grade 8 and Mathematics grade 7. One anchor item was flagged as a *b*-parameter outlier in ELA grades 3, 4, 6, and 8 and in Mathematics grades 3 through 8. Overall, the number of anchor items flagged using the TCC method was small.

Table 6-30 presents the item-ability regression statistics for the Mathematics grade 8 anchor item (anchor position 44; test question 37 in Session 2 of the test) flagged using the item-ability regression curve criteria described in an earlier section of this report. This item was flagged by four of the statistics used to examine ICC differences using the IRT item-ability regression curve method. The magnitude of the flags was moderate. Figure 6-4 show the ICCs before and after equating for the flagged Mathematics anchor item. In these figures, the dashed red line is the ICC before equating (based on input parameters) and the solid blue line is the ICC after equating (based on new parameter estimates). Examination of statistical properties of the flagged anchor items revealed that students performed better on this item during the Spring 2022 test administration than in the Spring 2019 test administration. No other anchor items in any other grades or content areas were flagged using the IRT item-ability regression curve method.

The flagged Mathematics grade 8 anchor item was reviewed by DRC test development experts who verified that no changes to item content, format, or position on the test occurred between the Spring 2019 and Spring 2022 administrations. No plausible explanation was found for differential item performance between the two administration years, and this anchor item was not excluded from equating.

One Science grade 8 anchor item (anchor position 25; test question 10 in Session 2 of the test) was excluded from equating due to poor statistics in Spring 2022. While this item did not meet the Wisconsin Forward Exam statistical criteria to be considered for removal from the anchor set, an exception was made based on the item's poor discrimination statistics in 2022. The item-total test correlation decreased from 0.12 in Spring 2019 to 0.04 in Spring 2022. While no plausible explanation was found for differential item discrimination between the two administration years for this item, poor item discrimination warranted its removal from the anchor set. Exclusion of the Science anchor item from the anchor set did not significantly affect the anchor set content coverage or the equating results for these grades.

### 6.4.3 Test Scaling and Scale Evaluation

The purpose of scaling a test is to enhance its validity by increasing the comparability of test takers' scores. This section explicates the way in which the Wisconsin Forward Exam scales are produced to comply with Standard 5.2 of the AERA, APA, & NCME (2014) Standards, which states the following:

> The procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly. (p. 102)

---

The Wisconsin Forward Exam scales were established for ELA and Mathematics after the Spring 2016 test administration. New scales were established for Science assessments after the Spring 2019 test administration. New scales were established for Social Studies assessments after the Spring 2022 test administration. In this section, the results of the test scaling of the Wisconsin Forward Exam are described and evaluated.

Following the test equating for ELA, Mathematics, and Science, the equated item parameter estimates in the theta metric were transformed into the scale score metric for the purpose of the evaluation of the scale properties. For Social Studies, the item parameters estimated during the calibration process were transformed into the scale score metric. The scale evaluation included

- evaluation of the TCCs,
- evaluation of the standard error (SE) curves, and
- examination of the growth at quartiles.

The scaling constants, *M1* and *M2*, used to transform equated item parameters and ability estimates in the theta metric into the scale score metric are the same as the scaling constants used in the Spring 2016 scale development for ELA and Mathematics. The *M1* and *M2* scaling constants for Science were established after the Spring 2019 test administration. These scaling constants are presented in Table 6-31.

As stated in the previous section, new scales were established for Social Studies grades 4, 8, and 10 after the Spring 2022 test administration. The mean and standard deviation of ability estimates for each grade were estimated and used to identify transformation constants that allowed the transformation of ability estimates in a theta metric into a scale score metric and to produce a student scale score distribution with a target mean and standard deviation for each grade. In order to differentiate the new Social Studies scales from the previous ones, a scale score mean of 600 and a standard deviation of 50 were set for grade 4, a scale score mean of 700 and a standard deviation of 50 were set for grade 8, and a scale score mean of 800 and a standard deviation of 50 were set for grade 10. The resulting grade level scale score means show "vertical relationships" (increasing scale score means across grades), but they are not true vertical scales.

The following formulae were used to compute transformation constants for the transformation of the Science grade 4 and grade 8 ability estimates from the theta metric to the scale score metric:

$$M1 = \frac{SD_{ss,G}}{SD_{\theta,G}} \text{ and}$$

$$M2 = \bar{X}_G - (\bar{\theta}_G * M1),$$

where

- *M1* and *M2* are the transformation constants,
- $SD_{ss,G}$ is the target standard deviation in the scale score metric for the given grade,
- $SD_{\theta,G}$ is the estimated standard deviation in the theta metric for the given grade,
- $\bar{\theta}_G$ is the estimated population mean in the theta metric for the given grade, and
- $\bar{X}_G$ is the target mean in the scale score metric for the given grade.

Future Social Studies test forms will be equated to the Spring 2022 baseline scale using the same equating methodology as described for ELA, Mathematics, and Science in Section 6.4.2. The scaling constants for Social Studies are also presented in Table 6-31.

## ELA Scale

***Test Characteristic Curves***—Figure 6-5 shows the TCCs for ELA tests. As shown in Figure 6-5, the ELA TCCs are ordinal at most ability levels, indicating that the difficulty of these assessments increases as the grade level increases. The grade 4 TCC overlaps with the grade 5 TCC at the upper end of the ability scale, indicating that the ELA grade 4 and grade 5 assessments have comparable difficulties for higher ability students. The grade 7 and grade 8 TCCs cross at the lower ability level, indicating that the grade 7 assessments may be more difficult for lower ability students than the grade 8 assessments.

It should be noted that while TCC ordinality is a desirable property of a vertical scale, the lack of it does not necessarily affect student scores or grade-to-grade growth interpretation. As demonstrated by the pattern of scale scores at quartiles (see the "Growth at Quartiles" paragraph below) for grades 3–8, student ability on ELA assessments increases as grade level increases at all grade levels, indicating grade-to-grade growth.

***Standard Error Curves***—The SE curves for ELA presented in Figure 6-6 are U-shaped, indicating smaller errors around ability estimates that are roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring very high- and very low-achieving students are found. Overall, the SEs around the scale score were found to be reasonable for ELA assessments (for more details, see Section 8.1.1 of this report).

***Scale Scores and Growth at Quartiles***—The estimated scale scores for the ELA calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-7. It can be observed that the scale scores increase as the percentile increases within each grade. Consistent with the properties of a vertical scale, the scale scores also increase at the same percentile across grade levels, indicating growth on the ELA ability scale as students move from one grade to the next (except for the grade 8 scale score at the 25th percentile, where the grade 8 scale score is about the same as the grade 7 scale score, indicating that the lower-ability grade 7 students scored about the same on the ELA test for grade 7 as the lower-ability grade 8 students scored on the ELA test for grade 8).

## Mathematics Scale

***Test Characteristic Curves***—Figure 6-8 shows the TCCs for Mathematics assessments, which are on a vertical scale. As observed in Figure 6-8, the TCCs for Mathematics are ordinal, indicating that the difficulty of the assessments increases as the grade level increases.

***Standard Error Curves***—The SE curves for Mathematics presented in Figure 6-9 are U shaped (as expected), indicating smaller errors around ability estimates that are roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Mathematics assessments (for more details, see Section 8.1.1 of this report).

***Scale Scores and Growth at Quartiles***—The estimated scale scores for the calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-10. It can be observed that the scale scores increase as the percentile increases within each grade level. With the exception of Mathematics grades 5 and 6, which show the same scale score at the 25th percentile, the scale scores increase at the

same percentile across grade levels, indicating growth on the Mathematics ability scale as students move from one grade to the next.

## Science Scale

*Test Characteristic Curves*—Although the Science assessments are not vertically scaled, the TCCs for grades 4 and 8 are presented together in Figure 6-11 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a student's ability increases. The grade 4 and grade 8 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

*Standard Error Curves*—Figure 6-12 shows the SE curves for Science grades 4 and 8. The SE curves are U-shaped, indicating smaller errors around ability estimates that are approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Science assessments (for more details, see Section 8.1.1 of this report).

*Scale Scores at Quartiles*—The estimated scale scores for the Science calibration sample at the 25th, 50th, and 75th percentiles for both grade levels are presented in Figure 6-13. The data pattern presented in this figure indicates that the scale scores increase as the percentile increases within each grade level. Because the Science assessments are not on a vertical scale, it is not appropriate to compare scale scores between grades.

## Social Studies Scale

*Test Characteristic Curves*—Although the Social Studies assessments are not vertically scaled, the TCCs for grades 4, 8, and 10 are presented together in Figure 6-14 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a student's ability increases. The grade 4 and grade 8 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

*Standard Error Curves*—Figure 6-15 shows Social Studies SE curves for grades 4, 8, and 10. The SE curves are U-shaped, indicating smaller errors around ability estimates that are approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Social Studies assessments (for more details, see Section 8.1.1 of this report).

*Scale Scores at Quartiles*—The estimated scale scores for the Social Studies calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-16. The data pattern presented in this figure indicates that the scale scores increase as the percentile increases within each grade level. Because the Social Studies assessments are not on a vertical scale, it is not appropriate to compare scale scores between grades.

### 6.4.4 Derivation of Scale Scores

A scale score can be interpreted as a highly probable estimate of a student's ability in a given content area. Scale scores are based on the student's responses to all items on a given test and account for the characteristics of the items that are on the test (such as item difficulty). Item parameters estimated after the Spring 2022 test administration were used to derive student scale scores.

Scale scores in the Wisconsin Forward Exam are based on the theoretical models of the item response process described above and elaborated upon below. The essential idea behind these models is that the probability of a correct response to a given item is a function of examinee ability and the characteristics of the item, such as the difficulty of the item. It is expected that as examinee ability increases, the probability of a correct response to a given item also increases, given certain conditions and assumptions. This description applies specifically to MC items; non-MC items are treated as CR items and are handled slightly differently, but they follow a logic that is essentially the same.

Whether looking at an individual item or at a group of items that make up a complete test, IRT uses probability models to describe the relationship between a student's ability and that student's observed scores. As described above, the 3PL model is used to estimate the probability of a correct response for each of the MC items. The model is provided here because its components are reviewed in the following paragraphs.

$$P(u_i = 1|\theta) = c_i + \frac{1 - c_i}{1 + e^{-1.7a_i(\theta - b_i)}} \quad (1)$$

In this model, $\theta$ denotes a measured ability (e.g., ELA ability) and $u_i$ represents an observed score on a particular item. For MC items, the observed score $u_i$ is either 0 or 1, indicating either an incorrect or a correct response, respectively. For an MC item, the probability model can be denoted as $P(u_i = 1|\theta)$. That is, $P$ is an estimation of the probability that a student with an ability value $\theta$ would answer item i correctly.

The terms on the right side of the equation above ($a_i$, $b_i$, $c_i$) represent the parameters in the model: discrimination, difficulty (or location), and a pseudo-guessing factor. Discrimination refers to how well an item sorts students by ability level, difficulty represents the difficulty of the item or its location on an ability continuum, and the pseudo-guessing factor represents the probability of a low-ability student guessing the correct response.

Given any particular response pattern ($u_1 u_2 \cdots u_n$) on a test with some number of items ($n$ items), the "likelihood function," or the probability that a student with a given ability value ($\theta$) would produce this particular response pattern, is given by

$$P(u_1 u_2 \cdots u_n|\theta) = \prod_{i=1}^{n} P(u_i|\theta). \quad (2)$$

The formula indicates that the "estimated maximum likelihood" IRT item-pattern scoring method searches for the ability estimate ($\theta_0$) that maximizes the probability function in the equation shown above (2) and assigns an ability estimate ($\theta_0$) as the test score for the student with the response pattern ($u_1 u_2 \cdots u_n$). In other words, the scale score is the most likely, or most probable, estimate of student ability, produced in a context in which item parameters are known and based on all the items in a given test.

As indicated, the item-pattern scoring method takes into account not only a student's total raw score but also the psychometric characteristics of all items the student responded to, including the items the student responded to incorrectly.

Consider the following example. Suppose six examinees in grade 4 take an ELA test with 30 MC items. Suppose further that the properties, or parameters, of the items on that test are as follows (see Table 6-A).

## Table 6-A. Example of Item Parameters for a Test

| Item | Discrimination (*a*) | Location (*b*) | Guessing (*c*) | Item | Discrimination (*a*) | Location (*b*) | Guessing (*c*) |
|------|---------------------|----------------|----------------|------|---------------------|----------------|----------------|
| 1 | 0.0341 | 318.75 | 0.16 | 16 | 0.0398 | 286.13 | 0.13 |
| 2 | 0.0342 | 244.62 | 0.20 | 17 | 0.0523 | 290.65 | 0.26 |
| 3 | 0.0234 | 257.56 | 0.20 | 18 | 0.0387 | 280.23 | 0.14 |
| 4 | 0.0306 | 235.00 | 0.20 | 19 | 0.0329 | 315.71 | 0.21 |
| 5 | 0.0125 | 342.39 | 0.17 | 20 | 0.0370 | 287.88 | 0.25 |
| 6 | 0.0305 | 261.51 | 0.16 | 21 | 0.0387 | 280.25 | 0.18 |
| 7 | 0.0316 | 296.93 | 0.19 | 22 | 0.0321 | 285.86 | 0.17 |
| 8 | 0.0228 | 252.70 | 0.20 | 23 | 0.0219 | 302.52 | 0.13 |
| 9 | 0.0383 | 266.28 | 0.20 | 24 | 0.0551 | 301.11 | 0.26 |
| 10 | 0.0229 | 308.84 | 0.11 | 25 | 0.0165 | 324.24 | 0.19 |
| 11 | 0.0536 | 259.00 | 0.21 | 26 | 0.0279 | 297.19 | 0.11 |
| 12 | 0.0478 | 245.19 | 0.20 | 27 | 0.0423 | 296.06 | 0.28 |
| 13 | 0.0418 | 276.25 | 0.28 | 28 | 0.0658 | 324.76 | 0.21 |
| 14 | 0.0377 | 287.60 | 0.23 | 29 | 0.0488 | 281.56 | 0.32 |
| 15 | 0.0177 | 316.08 | 0.24 | 30 | 0.0237 | 345.32 | 0.37 |

Now suppose that the student response patterns for these six examinees are as follows, where 0 represents an incorrect response and 1 represents a correct response (see Table 6-B).

## Table 6-B. Example of Item Response Pattern

| Student | Response Pattern ($u_1 u_2 \cdots u_n$) | Raw Score | Item-Pattern Score |
|---------|------------------------------------------|-----------|--------------------|
| Pam | 100001100101000000000000000101 | 7 | 140 |
| Craig | 101010101010101010101010101010 | 15 | 246 |
| Vicki | 010101010101010101010101010101 | 15 | 266 |
| Tom | 001100110011001100110011001101 | 15 | 259 |
| Evan | 110011001100110011001100110010 | 15 | 265 |
| Dan | 111111111111111111111111011111 | 29 | 379 |

The first student, Pam, answered 7 of the items correctly and obtained a scale score of 140, which is equal to the lowest point on the scale score range, called the lowest obtainable scale score, or LOSS. The next four students each answered 15 out of 30 items correctly, but the response pattern of each of these students is different. The raw score of each of these students is 15. However, the maximum likelihood item-pattern scoring method produced a different scale score for each examinee. Scale scores were 246 for Craig, 266 for Vicki, 259 for Tom, and 265 for Evan. These scores can be accounted for by considering the pattern of the student responses on the test in conjunction with the properties (or parameters) of the items as shown in Table 6-A. By referring to Table 6-A, the reader can observe that Vicki and Evan answered some difficult and highly discriminating items correctly, whereas Craig and Tom did not. The remaining student, Dan, scored 29 out of the 30 items correctly and obtained a scale score of 379, which is near the upper limit of the scale score range, called the highest obtainable scale score, or HOSS.

Figure 6-A shows the probability of each ability estimate (or scale score) for the six examinees. The total scale score range for the test is plotted on the horizontal axis. As indicated by the two vertical lines in the plot, the lower and upper limits of the scale score range are 140 and 420, respectively. The likelihood, or probability, of all possible ability estimates for each examinee is plotted on the vertical axis and ranges

from 0 to 1.0. The higher the likelihood, the more probable it is that the ability estimate accurately reflects the examinee's ability level.

As indicated above, scale scores are the most likely, or the maximum likelihood, estimates of examinee ability. As can be observed for Vicki, Tom, and Evan, scores that are plus or minus only a few scale score points are markedly less likely estimates of the students' abilities. The same is true for Craig and Dan, though to a slightly lesser extent. In the case of Pam, a few scores were almost as likely as the maximum likelihood estimate reported. Those scores that appear to be more likely than the reported score are outside of the scale score range of the test (below the LOSS).

There are two IRT-based scoring methods generally used for large-scale assessments: number-correct scoring and item-pattern scoring. Item-pattern scoring may be recommended over number-correct scoring for several reasons. Two reasons, accuracy and reliability, are pertinent for the present purposes.

First, item-pattern scoring generally produces more accurate scores for individual students. Specifically, it produces a smaller conditional standard error of measurement (CSEM) across the scale score range for a given test compared to number-correct scoring. The smaller the CSEM, the more confident one can be in the accuracy of the test results. The increase in accuracy provided by item-pattern scoring is equivalent, on average, to an increase by approximately 15 to 20 percent in test length (Yen, 1984; Yen & Candell, 1991).

Second, reliability tends to be higher using item-pattern scoring, which means (a) fewer items are needed to achieve a given level of reliability and (b) a given test with a given number of items will have higher reliability than it would when using number-correct scoring. Yen (1984) has demonstrated that an equivalent level of reliability for a 20-item test scored by the number correct scoring method could be obtained with a 16- or 17-item test scored by the item pattern scoring method.

Because of the nature of item-pattern scoring, a scoring table showing a simple, direct conversion of raw score to scale score cannot be generated for the Spring 2022 Wisconsin Forward Exam. However, scoring tables showing an approximate raw score-to-scale score relationship and the associated CSEM can be produced, and these are provided in Tables 6-32 through 6-48. These tables are provided to illustrate the approximate raw score-to-scale score relationship for each unique raw score and do not include all combinations of raw score-to-scale score associations.

Several supplements to this simplified outline of IRT are available. Introductory discussions of IRT can be found in *Educational Measurement* (Linn, 1989) or Chapter 11 in *Introduction to Measurement Theory* (Allen & Yen, 1979). More advanced discussions of partial-credit models may be found in Muraki (1990, 1992), Yen (1993), and van der Linden & Hambleton (1997). For additional information on the technical details of item-pattern scoring, readers can also refer to Yen & Candell (1991).

**Figure 6-A. Examples of Likelihood Functions or the Probability of Each Ability Level Estimate (or Scale Score)**

a. Pam
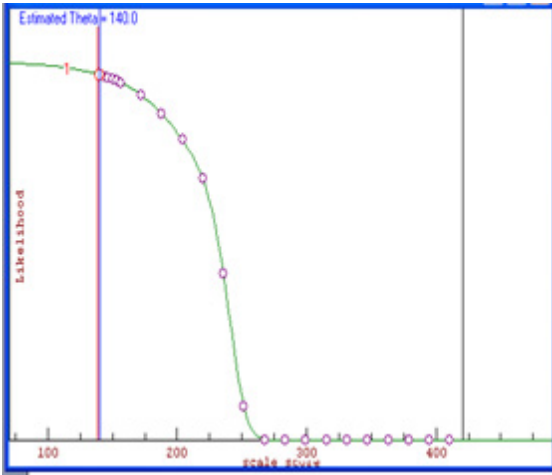
b. Craig
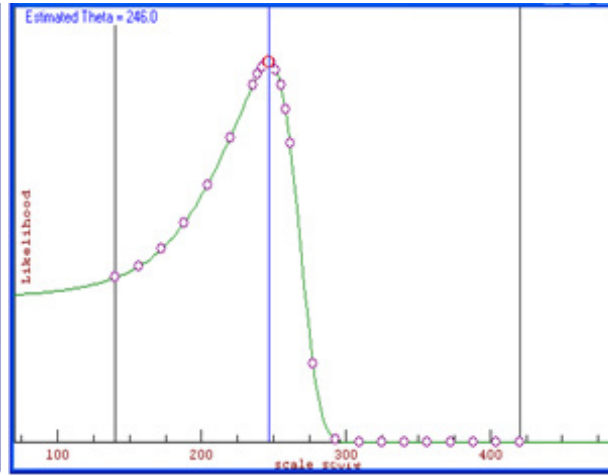
c. Vicki
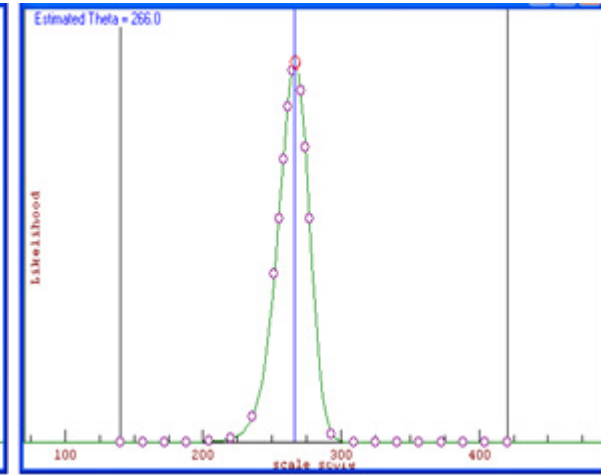
d. Tom
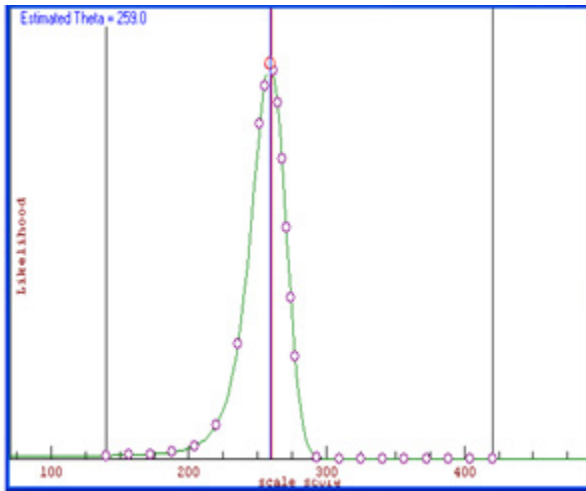
e. Evan
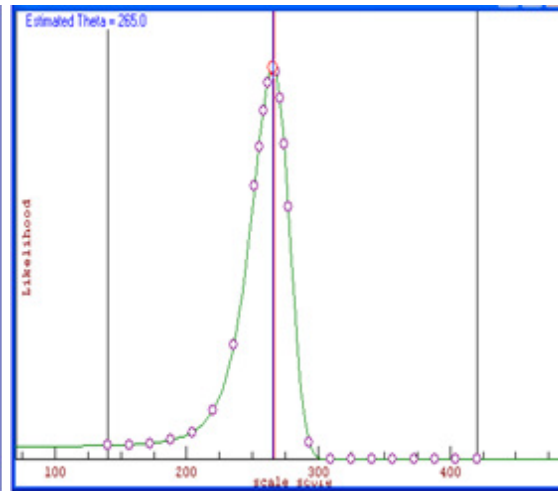
f. Dan
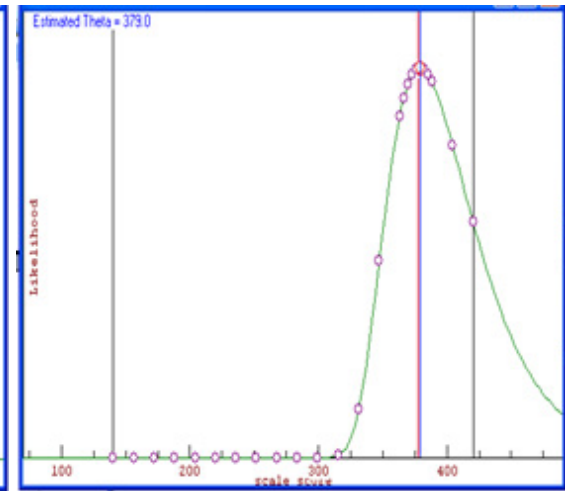


Note: The circular dots in the likelihood functions indicate that the software program used is searching for a maximum likelihood estimate (scale score) for the student.

### 6.4.5 Lowest and Highest Obtainable Scale Scores

As previously established, a scale score is a maximum likelihood ability estimate. The maximum-likelihood procedure cannot produce scale score estimates for students with perfect scores or scores below the scoring level expected by guessing. Although maximum likelihood estimates are available for students with extreme scores other than zero or a perfect score, these estimates generally have large SEMs. Therefore, scores are established for these extreme highs and lows based on a rational, but necessarily non-maximum, likelihood procedure. These values are set separately by grade and are called the LOSS and the HOSS. The LOSS and HOSS values for ELA and Mathematics were established after the Spring 2016 test administration and remained unchanged through the Spring 2022 test administration. New LOSS and HOSS values were established after the Spring 2019 test administration for Science and after the Spring 2022 test administration for Social Studies.

Table 6-49 shows the number and percentage of students at the LOSS and the HOSS. In general, there should not be many students clustered at the LOSS or HOSS. A high proportion of students at the LOSS or HOSS may indicate a floor or ceiling effect.

It should be noted that for ELA and Mathematics, the LOSS and HOSS values were set in such a way during the Spring 2016 scale development that they increase as the grade level increases. Setting increasing LOSS values as the grade level increases is an important property of a vertical scale and constrains student ability in each grade in such a way that the lowest-ability students in a given grade will always have a higher scale score than the lowest-ability students in a grade below and a lower scale score than the lowest-ability students in a grade above. Conversely, setting increasing HOSS values as the grade level increases constrains student ability in each grade in such a way that the highest-ability students in a given grade will always have a higher scale score than the highest-ability students in a grade below and a lower scale score than the highest-ability students in a grade above.

Less than one-tenth of one percent of students received the lowest obtainable scale scores in all ELA and Science grades. In addition, about half a percent of students scored at the LOSS in Social Studies grade 4. The percentages of students scoring at the LOSS was higher for all Mathematics grades and ranged from over 1% for grade 3 to over 3% for grade 8. Approximately 2% of students scored at the LOSS in Social Studies grades 8 and 10. These percentages of students at the LOSS in Mathematics and Social Studies grades 8 and 10 were investigated. It was found that students who scored this way typically correctly answered very few MC items and no non-MC items, which resulted in their LOSS values. For these students to receive a scale score above the LOSS, they would need to correctly answer more items, including some non-MC items. Non-MC items do not assume guessing, so the correct responses tend to represent student ability more accurately.

No students scored at the HOSS in ELA grades 3, 4, and 7 and only between 1 and 4 students received the highest obtainable score in ELA grades 5, 6, and 8. Less than half of one percent of students in all grades of Mathematics, Science, and Social Studies scored at the HOSS.

## 6.5 Summary

In summary, the overall purpose of the test psychometric data analysis, including scaling and equating, is to ensure that the test items, as well as the overall test, are functioning appropriately. It also helps maintain the test scale so that test results may be appropriately compared across years. The data analyses undertaken by DRC are in alignment with multiple best practices of the testing industry and, in particular, support the following AERA, APA, & NCME (2014) Standards: 1.8, 4.14, 5.2, 5.13, 5.15, and 7.2.

**Table 6-1. Summary of Flagged Operational Items on the Wisconsin Forward Exam**

| Content | Grade | # of Items Flagged | Number of Flags | | | |
|---|---|---|---|---|---|---|
| | | | Correlation <0.15 | Distractor Correlation >0 | Omit >3% | *p*-Value <0.20 |
| ELA | 3 | 4 | 0 | 4 | 0 | 0 |
| | 4 | 0 | 0 | 0 | 0 | 0 |
| | 5 | 1 | 0 | 1 | 0 | 0 |
| | 6 | 2 | 0 | 2 | 0 | 0 |
| | 7 | 2 | 0 | 2 | 0 | 0 |
| | 8 | 1 | 0 | 1 | 0 | 0 |
| Mathematics | 3 | 1 | 0 | 1 | 0 | 0 |
| | 4 | 6 | 1 | 5 | 0 | 1 |
| | 5 | 6 | 1 | 5 | 0 | 1 |
| | 6 | 6 | 0 | 4 | 0 | 2 |
| | 7 | 5 | 0 | 3 | 0 | 2 |
| | 8 | 7 | 0 | 5 | 0 | 2 |
| Science | 4 | 1 | 1 | 0 | 0 | 0 |
| | 8 | 3 | 1 | 2 | 0 | 0 |
| Social Studies | 4 | 1 | 0 | 1 | 0 | 0 |
| | 8 | 1 | 0 | 1 | 0 | 0 |
| | 10 | 2 | 0 | 2 | 0 | 0 |
| **Total** | | **49** | **4** | **39** | **0** | **8** |

Note: The number of flags may be greater than the number of flagged items.

**Table 6-2. Items Flagged for Classical Item Analysis Statistics, English Language Arts**

| Grade | Content | Item | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
| | | | | | | | Corr | Distractor | Omit | *p*-Value |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | ELA | 5 | MC | 0.67 | 0.20 | 0.19% | | + | 0.04 | |
| 3 | ELA | 17 | MC | 0.44 | 0.27 | 0.21% | | + | 0.03 | |
| 3 | ELA | 24 | MC | 0.50 | 0.21 | 0.36% | | + | 0.06 | |
| 3 | ELA | 28 | MC | 0.50 | 0.22 | 0.28% | | + | 0.11 | |
| 5 | ELA | 3 | MC | 0.29 | 0.22 | 0.11% | | + | 0.02 | |
| 6 | ELA | 5 | MC | 0.34 | 0.19 | 0.24% | | + | 0.03 | |
| 6 | ELA | 17 | MC | 0.42 | 0.27 | 0.18% | | + | 0.01 | |
| 7 | ELA | 22 | MC | 0.41 | 0.22 | 0.42% | | + | 0.04 | |
| 7 | ELA | 24 | MC | 0.28 | 0.18 | 0.28% | | + | 0.12 | |
| 8 | ELA | 35 | MC | 0.39 | 0.20 | 0.43% | | + | 0.07 | |

**Table 6-3. Items Flagged for Classical Item Analysis Statistics, Mathematics**

| Grade | Content | Item | Item Type | p-Value | Corr | Percent Omit | Flags | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Corr | Distractor | | Omit | p-Value |
| 3 | Math | 16 | MC | 0.32 | 0.40 | 0.31% | | + | 0.04 | | |
| 4 | Math | 14 | MC | 0.41 | 0.29 | 0.46% | | + | 0.01 | | |
| 4 | Math | 15 | MC | 0.40 | 0.29 | 0.20% | | + | 0.03 | | |
| 4 | Math | 17 | MC | 0.26 | 0.06 | 0.22% | + | + | 0.28 | | |
| 4 | Math | 22 | MC | 0.30 | 0.17 | 0.49% | | + | 0.13 | | |
| 4 | Math | 28 | MC | 0.34 | 0.19 | 0.15% | | + | 0.10 | | |
| 4 | Math | 35 | SA | 0.18 | 0.54 | 0.27% | | | | | + |
| 5 | Math | 11 | MC | 0.33 | 0.21 | 0.28% | | + | 0.04 | | |
| 5 | Math | 20 | MC | 0.38 | 0.26 | 0.26% | | + | 0.01 | | |
| 5 | Math | 21 | SA | 0.19 | 0.43 | 0.33% | | | | | + |
| 5 | Math | 23 | MC | 0.34 | 0.43 | 0.55% | | + | 0.01 | | |
| 5 | Math | 33 | MC | 0.42 | 0.30 | 0.17% | | + | 0.11 | | |
| 5 | Math | 45 | MC | 0.24 | 0.14 | 0.29% | + | + | 0.13 | | |
| 6 | Math | 1 | MC | 0.36 | 0.38 | 0.18% | | + | 0.04 | | |
| 6 | Math | 10 | SA | 0.15 | 0.48 | 0.24% | | | | | + |
| 6 | Math | 25 | MC | 0.30 | 0.31 | 0.51% | | + | 0.05 | | |
| 6 | Math | 36 | SA | 0.17 | 0.56 | 0.74% | | | | | + |
| 6 | Math | 38 | MC | 0.63 | 0.31 | 0.45% | | + | 0.04 | | |
| 6 | Math | 43 | MC | 0.35 | 0.23 | 0.42% | | + | 0.02 | | |
| 7 | Math | 8 | TE | 0.17 | 0.42 | 0.17% | | | | | + |
| 7 | Math | 9 | MC | 0.33 | 0.28 | 0.18% | | + | 0.07 | | |
| 7 | Math | 20 | MC | 0.37 | 0.25 | 0.43% | | + | 0.02 | | |
| 7 | Math | 25 | SA | 0.17 | 0.57 | 0.87% | | | | | + |
| 7 | Math | 32 | MC | 0.30 | 0.19 | 0.58% | | + | 0.03 | | |
| 8 | Math | 3 | MC | 0.36 | 0.31 | 0.16% | | + | 0.05 | | |
| 8 | Math | 4 | MC | 0.44 | 0.50 | 0.14% | | + | 0.04 | | |
| 8 | Math | 5 | SA | 0.14 | 0.50 | 0.34% | | | | | + |
| 8 | Math | 9 | TE | 0.18 | 0.54 | 0.26% | | | | | + |
| 8 | Math | 21 | MC | 0.39 | 0.24 | 0.36% | | + | 0.01 | | |
| 8 | Math | 35 | MC | 0.43 | 0.26 | 0.65% | | + | 0.05 | | |
| 8 | Math | 41 | MC | 0.37 | 0.23 | 0.70% | | + | 0.02 | | |

**Table 6-4. Items Flagged for Classical Item Analysis Statistics, Science and Social Studies**

| Grade | Content | Item | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Corr | Distractor | Omit | *p*-Value |
| 4 | Science | 13 | MC | 0.38 | 0.12 | 0.25% | + | | | |
| 8 | Science | 2 | MC | 0.47 | 0.22 | 0.15% | | + | 0.04 | |
| 8 | Science | 10 | MC | 0.41 | 0.16 | 0.16% | | + | 0.07 | |
| 8 | Science | 25 | TE | 0.46 | 0.04 | 0.20% | + | | | |
| 4 | Social Studies | 8 | MC | 0.51 | 0.29 | 0.17% | | + | 0.05 | |
| 8 | Social Studies | 27 | MC | 0.55 | 0.28 | 0.13% | | + | 0.01 | |
| 10 | Social Studies | 5 | MC | 0.41 | 0.28 | 0.24% | | + | 0.01 | |
| 10 | Social Studies | 17 | MC | 0.33 | 0.27 | 0.71% | | + | 0.04 | |

**Table 6-5. Percentage of Students Attempting Last Operational Item in Test**

| Content | Grade | | | | | | |
|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 | 10 |
| English Language Arts | 98.97 | 99.45 | 99.45 | 99.62 | 99.53 | 99.42 | |
| Mathematics | 98.72 | 99.56 | 99.62 | 99.47 | 99.29 | 99.36 | |
| Science | | 99.84 | | | | 99.72 | |
| Social Studies | | 99.66 | | | | 99.39 | 98.89 |

**Table 6-6. Item Analysis, English Language Arts Grade 3**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 1 | TDA | 0.33 | 0.38 | 0.62 | | | | |
| 2 | MC | 0.79 | 0.42 | 0.35 | | | | |
| 3 | MC | 0.63 | 0.23 | 0.20 | | | | |
| 4 | TE | 0.70 | 0.53 | 0.21 | | | | |
| 5 | MC | 0.67 | 0.20 | 0.19 | | + | | |
| 6 | TE | 0.55 | 0.45 | 0.39 | | | | |
| 7 | MC | 0.72 | 0.39 | 0.21 | | | | |
| 8 | MC | 0.73 | 0.48 | 0.24 | | | | |
| 9 | MC | 0.60 | 0.38 | 0.21 | | | | |
| 10 | MC | 0.44 | 0.28 | 0.23 | | | | |
| 11 | TE | 0.45 | 0.46 | 0.28 | | | | |
| 12 | MC | 0.42 | 0.36 | 0.24 | | | | |
| 13 | MC | 0.51 | 0.30 | 0.25 | | | | |
| 14 | TE | 0.40 | 0.22 | 1.12 | | | | |
| 15 | EBSR | 0.53 | 0.34 | 0.07 | | | | |
| 16 | MC | 0.57 | 0.35 | 0.28 | | | | |
| 17 | MC | 0.44 | 0.27 | 0.21 | | + | | |
| 18 | MC | 0.53 | 0.37 | 0.23 | | | | |
| 19 | MS | 0.68 | 0.53 | 0.21 | | | | |
| 20 | MC | 0.52 | 0.20 | 0.17 | | | | |
| 21 | MC | 0.63 | 0.41 | 0.33 | | | | |
| 22 | MC | 0.42 | 0.29 | 0.34 | | | | |
| 23 | EBSR | 0.54 | 0.51 | 0.15 | | | | |
| 24 | MC | 0.50 | 0.21 | 0.36 | | + | | |
| 25 | TE | 0.66 | 0.65 | 0.40 | | | | |
| 26 | MC | 0.55 | 0.44 | 1.04 | | | | |
| 27 | MC | 0.77 | 0.47 | 0.95 | | | | |
| 28 | MC | 0.50 | 0.22 | 0.28 | | + | | |
| 29 | EBSR | 0.49 | 0.55 | 0.19 | | | | |
| 30 | MC | 0.70 | 0.50 | 0.31 | | | | |
| 31 | MC | 0.54 | 0.35 | 0.34 | | | | |
| 32 | MC | 0.42 | 0.33 | 0.50 | | | | |
| 33 | MC | 0.43 | 0.37 | 0.46 | | | | |
| 34 | MC | 0.57 | 0.53 | 0.40 | | | | |
| 35 | EBSR | 0.43 | 0.54 | 0.30 | | | | |
| 36 | MC | 0.48 | 0.30 | 0.70 | | | | |
| 37 | MC | 0.42 | 0.31 | 1.03 | | | | |

**Table 6-7. Item Analysis, English Language Arts Grade 4**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 1 | TDA | 0.35 | 0.41 | 0.31 | | | | |
| 2 | MC | 0.47 | 0.35 | 0.08 | | | | |
| 3 | TE | 0.60 | 0.42 | 0.18 | | | | |
| 4 | MC | 0.83 | 0.44 | 0.11 | | | | |
| 5 | TE | 0.51 | 0.47 | 0.30 | | | | |
| 6 | TE | 0.48 | 0.42 | 0.21 | | | | |
| 7 | MC | 0.52 | 0.44 | 0.21 | | | | |
| 8 | MC | 0.56 | 0.38 | 0.18 | | | | |
| 9 | MC | 0.56 | 0.41 | 0.15 | | | | |
| 10 | MC | 0.42 | 0.25 | 0.16 | | | | |
| 11 | TE | 0.68 | 0.41 | 0.68 | | | | |
| 12 | MC | 0.65 | 0.50 | 0.18 | | | | |
| 13 | TE | 0.63 | 0.29 | 0.18 | | | | |
| 14 | MC | 0.42 | 0.23 | 0.18 | | | | |
| 15 | MC | 0.56 | 0.28 | 0.06 | | | | |
| 16 | MC | 0.74 | 0.37 | 0.15 | | | | |
| 17 | EBSR | 0.32 | 0.34 | 0.05 | | | | |
| 18 | MC | 0.49 | 0.22 | 0.13 | | | | |
| 19 | MC | 0.53 | 0.33 | 0.16 | | | | |
| 20 | EBSR | 0.58 | 0.46 | 0.08 | | | | |
| 21 | MC | 0.72 | 0.46 | 0.09 | | | | |
| 22 | TE | 0.56 | 0.43 | 0.45 | | | | |
| 23 | MC | 0.49 | 0.41 | 0.23 | | | | |
| 24 | MC | 0.59 | 0.54 | 0.29 | | | | |
| 25 | MS | 0.46 | 0.46 | 0.21 | | | | |
| 26 | MC | 0.87 | 0.46 | 0.29 | | | | |
| 27 | MC | 0.78 | 0.45 | 0.20 | | | | |
| 28 | EBSR | 0.41 | 0.46 | 0.21 | | | | |
| 29 | MC | 0.69 | 0.42 | 0.41 | | | | |
| 30 | MC | 0.63 | 0.41 | 0.24 | | | | |
| 31 | MC | 0.49 | 0.44 | 0.77 | | | | |
| 32 | EBSR | 0.27 | 0.42 | 0.18 | | | | |
| 33 | MC | 0.52 | 0.40 | 0.29 | | | | |
| 34 | MC | 0.64 | 0.51 | 0.29 | | | | |
| 35 | MC | 0.53 | 0.40 | 0.34 | | | | |
| 36 | MC | 0.64 | 0.40 | 0.30 | | | | |
| 37 | MC | 0.56 | 0.43 | 0.27 | | | | |
| 38 | TE | 0.46 | 0.45 | 0.31 | | | | |
| 39 | MC | 0.59 | 0.44 | 0.55 | | | | |

**Table 6-8. Item Analysis, English Language Arts Grade 5**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | TDA | 0.36 | 0.48 | 0.26 | | | | |
| 2 | MC | 0.78 | 0.42 | 0.06 | | | | |
| 3 | MC | 0.29 | 0.22 | 0.11 | | + | | |
| 4 | TE | 0.58 | 0.47 | 0.10 | | | | |
| 5 | MC | 0.41 | 0.28 | 0.12 | | | | |
| 6 | MC | 0.62 | 0.44 | 0.12 | | | | |
| 7 | TE | 0.74 | 0.36 | 0.11 | | | | |
| 8 | TE | 0.70 | 0.53 | 0.29 | | | | |
| 9 | MC | 0.46 | 0.29 | 0.10 | | | | |
| 10 | MC | 0.52 | 0.35 | 0.14 | | | | |
| 11 | TE | 0.72 | 0.57 | 0.12 | | | | |
| 12 | MC | 0.76 | 0.46 | 0.10 | | | | |
| 13 | TE | 0.59 | 0.39 | 0.28 | | | | |
| 14 | MC | 0.89 | 0.38 | 0.06 | | | | |
| 15 | MC | 0.86 | 0.41 | 0.22 | | | | |
| 16 | EBSR | 0.60 | 0.34 | 0.04 | | | | |
| 17 | MC | 0.61 | 0.42 | 0.09 | | | | |
| 18 | MC | 0.58 | 0.38 | 0.17 | | | | |
| 19 | EBSR | 0.56 | 0.53 | 0.07 | | | | |
| 20 | MC | 0.79 | 0.38 | 0.09 | | | | |
| 21 | MC | 0.55 | 0.30 | 0.24 | | | | |
| 22 | EBSR | 0.37 | 0.42 | 0.06 | | | | |
| 23 | MC | 0.61 | 0.40 | 0.26 | | | | |
| 24 | MC | 0.55 | 0.43 | 0.33 | | | | |
| 25 | MC | 0.48 | 0.38 | 0.24 | | | | |
| 26 | MC | 0.34 | 0.21 | 0.52 | | | | |
| 27 | MC | 0.50 | 0.25 | 0.41 | | | | |
| 28 | MC | 0.52 | 0.29 | 0.30 | | | | |
| 29 | MS | 0.56 | 0.53 | 0.23 | | | | |
| 30 | MC | 0.60 | 0.40 | 0.54 | | | | |
| 31 | MC | 0.54 | 0.50 | 0.52 | | | | |
| 32 | MC | 0.46 | 0.45 | 0.62 | | | | |
| 33 | MC | 0.49 | 0.39 | 0.36 | | | | |
| 34 | MC | 0.62 | 0.52 | 0.35 | | | | |
| 35 | MC | 0.53 | 0.35 | 0.38 | | | | |
| 36 | MC | 0.49 | 0.38 | 0.31 | | | | |
| 37 | TE | 0.70 | 0.27 | 0.70 | | | | |
| 38 | MC | 0.43 | 0.45 | 0.32 | | | | |
| 39 | MC | 0.47 | 0.33 | 0.34 | | | | |
| 40 | MC | 0.56 | 0.45 | 0.55 | | | | |

**Table 6-9. Item Analysis, English Language Arts Grade 6**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | TDA | 0.39 | 0.49 | 0.35 | | | | |
| 2 | MC | 0.67 | 0.23 | 0.08 | | | | |
| 3 | MC | 0.42 | 0.35 | 0.13 | | | | |
| 4 | MC | 0.69 | 0.44 | 0.21 | | | | |
| 5 | MC | 0.34 | 0.19 | 0.24 | | + | | |
| 6 | TE | 0.27 | 0.27 | 0.25 | | | | |
| 7 | MC | 0.36 | 0.26 | 0.18 | | | | |
| 8 | MC | 0.41 | 0.31 | 0.16 | | | | |
| 9 | TE | 0.63 | 0.45 | 0.14 | | | | |
| 10 | TE | 0.64 | 0.35 | 0.40 | | | | |
| 11 | TE | 0.60 | 0.47 | 0.21 | | | | |
| 12 | MC | 0.54 | 0.34 | 0.21 | | | | |
| 13 | MC | 0.43 | 0.32 | 0.19 | | | | |
| 14 | MC | 0.70 | 0.42 | 0.08 | | | | |
| 15 | MC | 0.54 | 0.34 | 0.21 | | | | |
| 16 | TE | 0.60 | 0.43 | 0.16 | | | | |
| 17 | MC | 0.42 | 0.27 | 0.18 | | + | | |
| 18 | MC | 0.78 | 0.48 | 0.21 | | | | |
| 19 | EBSR | 0.65 | 0.54 | 0.10 | | | | |
| 20 | MC | 0.64 | 0.45 | 0.12 | | | | |
| 21 | MC | 0.71 | 0.36 | 0.28 | | | | |
| 22 | MC | 0.68 | 0.47 | 0.37 | | | | |
| 23 | MS | 0.64 | 0.50 | 0.34 | | | | |
| 24 | MC | 0.55 | 0.35 | 0.40 | | | | |
| 25 | MC | 0.79 | 0.53 | 0.38 | | | | |
| 26 | TE | 0.55 | 0.42 | 0.43 | | | | |
| 27 | MC | 0.62 | 0.44 | 0.66 | | | | |
| 28 | TE | 0.64 | 0.45 | 0.50 | | | | |
| 29 | MC | 0.74 | 0.53 | 0.40 | | | | |
| 30 | MC | 0.48 | 0.37 | 0.37 | | | | |
| 31 | EBSR | 0.57 | 0.48 | 0.28 | | | | |
| 32 | TE | 0.67 | 0.54 | 0.43 | | | | |
| 33 | MC | 0.67 | 0.47 | 0.42 | | | | |
| 34 | EBSR | 0.50 | 0.62 | 0.28 | | | | |
| 35 | MC | 0.42 | 0.39 | 0.40 | | | | |
| 36 | MC | 0.57 | 0.42 | 0.34 | | | | |
| 37 | MC | 0.64 | 0.55 | 0.38 | | | | |

**Table 6-10. Item Analysis, English Language Arts Grade 7**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 1 | TDA | 0.46 | 0.52 | 0.36 | | | | |
| 2 | MC | 0.66 | 0.42 | 0.10 | | | | |
| 3 | TE | 0.49 | 0.40 | 0.14 | | | | |
| 4 | EBSR | 0.54 | 0.46 | 0.08 | | | | |
| 5 | MC | 0.79 | 0.36 | 0.12 | | | | |
| 6 | TE | 0.62 | 0.30 | 0.37 | | | | |
| 7 | MC | 0.53 | 0.26 | 0.17 | | | | |
| 8 | MC | 0.58 | 0.48 | 0.16 | | | | |
| 9 | MC | 0.68 | 0.37 | 0.18 | | | | |
| 10 | TE | 0.45 | 0.24 | 0.24 | | | | |
| 11 | MC | 0.42 | 0.29 | 0.24 | | | | |
| 12 | TE | 0.83 | 0.31 | 0.13 | | | | |
| 13 | MC | 0.59 | 0.44 | 0.21 | | | | |
| 14 | EBSR | 0.73 | 0.45 | 0.05 | | | | |
| 15 | MC | 0.56 | 0.36 | 0.19 | | | | |
| 16 | MS | 0.61 | 0.52 | 0.14 | | | | |
| 17 | MC | 0.78 | 0.45 | 0.19 | | | | |
| 18 | EBSR | 0.62 | 0.54 | 0.10 | | | | |
| 19 | MC | 0.47 | 0.41 | 0.14 | | | | |
| 20 | MS | 0.36 | 0.25 | 0.28 | | | | |
| 21 | MC | 0.74 | 0.41 | 0.28 | | | | |
| 22 | MC | 0.41 | 0.22 | 0.42 | | + | | |
| 23 | EBSR | 0.50 | 0.54 | 0.15 | | | | |
| 24 | MC | 0.28 | 0.18 | 0.28 | | + | | |
| 25 | TE | 0.71 | 0.42 | 1.13 | | | | |
| 26 | MC | 0.56 | 0.53 | 0.57 | | | | |
| 27 | EBSR | 0.48 | 0.42 | 0.17 | | | | |
| 28 | MC | 0.48 | 0.37 | 0.32 | | | | |
| 29 | TE | 0.30 | 0.42 | 1.19 | | | | |
| 30 | MC | 0.67 | 0.57 | 0.51 | | | | |
| 31 | EBSR | 0.42 | 0.44 | 0.28 | | | | |
| 32 | MC | 0.54 | 0.40 | 0.42 | | | | |
| 33 | MC | 0.46 | 0.46 | 0.47 | | | | |
| 34 | MS | 0.55 | 0.51 | 0.38 | | | | |
| 35 | MS | 0.71 | 0.50 | 0.35 | | | | |
| 36 | MC | 0.53 | 0.43 | 0.47 | | | | |

**Table 6-11. Item Analysis, English Language Arts Grade 8**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | TDA | 0.48 | 0.48 | 0.37 | | | | |
| 2 | MC | 0.47 | 0.21 | 0.06 | | | | |
| 3 | MC | 0.59 | 0.41 | 0.18 | | | | |
| 4 | MC | 0.61 | 0.40 | 0.19 | | | | |
| 5 | MC | 0.56 | 0.36 | 0.20 | | | | |
| 6 | MC | 0.80 | 0.40 | 0.20 | | | | |
| 7 | MC | 0.57 | 0.35 | 0.19 | | | | |
| 8 | MC | 0.65 | 0.37 | 0.18 | | | | |
| 9 | EBSR | 0.49 | 0.54 | 0.11 | | | | |
| 10 | TE | 0.45 | 0.29 | 0.35 | | | | |
| 11 | MC | 0.39 | 0.29 | 0.19 | | | | |
| 12 | TE | 0.85 | 0.39 | 0.20 | | | | |
| 13 | MC | 0.51 | 0.25 | 0.26 | | | | |
| 14 | MC | 0.48 | 0.38 | 0.20 | | | | |
| 15 | MS | 0.63 | 0.46 | 0.06 | | | | |
| 16 | EBSR | 0.66 | 0.50 | 0.05 | | | | |
| 17 | MC | 0.67 | 0.48 | 0.16 | | | | |
| 18 | MC | 0.47 | 0.38 | 0.20 | | | | |
| 19 | EBSR | 0.42 | 0.41 | 0.10 | | | | |
| 20 | MC | 0.36 | 0.34 | 0.14 | | | | |
| 21 | MC | 0.69 | 0.45 | 0.31 | | | | |
| 22 | MC | 0.53 | 0.33 | 0.31 | | | | |
| 23 | MS | 0.59 | 0.48 | 0.35 | | | | |
| 24 | MC | 0.38 | 0.47 | 0.32 | | | | |
| 25 | MC | 0.59 | 0.49 | 0.42 | | | | |
| 26 | MC | 0.51 | 0.28 | 0.27 | | | | |
| 27 | TE | 0.67 | 0.56 | 0.59 | | | | |
| 28 | MC | 0.71 | 0.55 | 0.49 | | | | |
| 29 | MC | 0.44 | 0.20 | 0.34 | | | | |
| 30 | MC | 0.67 | 0.43 | 0.44 | | | | |
| 31 | MC | 0.51 | 0.28 | 0.50 | | | | |
| 32 | MC | 0.53 | 0.34 | 0.54 | | | | |
| 33 | EBSR | 0.56 | 0.64 | 0.31 | | | | |
| 34 | MC | 0.73 | 0.54 | 0.44 | | | | |
| 35 | MC | 0.39 | 0.20 | 0.43 | | + | | |
| 36 | MC | 0.50 | 0.30 | 0.49 | | | | |
| 37 | MC | 0.55 | 0.44 | 0.44 | | | | |
| 38 | EBSR | 0.23 | 0.16 | 0.33 | | | | |
| 39 | MC | 0.61 | 0.43 | 0.58 | | | | |

**Table 6-12. Item Analysis, Mathematics Grade 3**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags Corr | Flags Distractor | Flags Omit | Flags *p*-Value |
|---|---|---|---|---|---|---|---|---|
| 1 | TE | 0.70 | 0.55 | 0.19 | | | | |
| 2 | SA | 0.53 | 0.57 | 0.17 | | | | |
| 3 | MC | 0.76 | 0.47 | 0.16 | | | | |
| 4 | TE | 0.42 | 0.62 | 0.14 | | | | |
| 5 | MC | 0.49 | 0.46 | 0.21 | | | | |
| 6 | SA | 0.45 | 0.59 | 1.10 | | | | |
| 7 | MC | 0.35 | 0.42 | 0.85 | | | | |
| 8 | SA | 0.36 | 0.55 | 0.29 | | | | |
| 9 | TE | 0.31 | 0.49 | 1.32 | | | | |
| 10 | MC | 0.49 | 0.42 | 0.26 | | | | |
| 11 | MC | 0.41 | 0.35 | 0.27 | | | | |
| 12 | MC | 0.44 | 0.32 | 0.31 | | | | |
| 13 | MC | 0.58 | 0.44 | 0.58 | | | | |
| 14 | SA | 0.36 | 0.34 | 0.67 | | | | |
| 15 | MC | 0.44 | 0.41 | 0.33 | | | | |
| 16 | MC | 0.32 | 0.40 | 0.31 | | + | | |
| 17 | SA | 0.56 | 0.56 | 0.37 | | | | |
| 18 | TE | 0.83 | 0.44 | 1.84 | | | | |
| 19 | MC | 0.72 | 0.49 | 0.29 | | | | |
| 20 | MC | 0.58 | 0.44 | 0.32 | | | | |
| 21 | SA | 0.70 | 0.59 | 0.29 | | | | |
| 22 | MC | 0.45 | 0.39 | 0.18 | | | | |
| 23 | TE | 0.81 | 0.42 | 0.18 | | | | |
| 24 | SA | 0.70 | 0.62 | 0.21 | | | | |
| 25 | TE | 0.58 | 0.36 | 0.22 | | | | |
| 26 | SA | 0.40 | 0.57 | 0.24 | | | | |
| 27 | SA | 0.62 | 0.59 | 0.71 | | | | |
| 28 | MC | 0.40 | 0.37 | 0.80 | | | | |
| 29 | MC | 0.92 | 0.32 | 0.20 | | | | |
| 30 | MC | 0.44 | 0.27 | 0.20 | | | | |
| 31 | MC | 0.39 | 0.43 | 0.21 | | | | |
| 32 | SA | 0.52 | 0.54 | 0.37 | | | | |
| 33 | SA | 0.52 | 0.61 | 0.30 | | | | |
| 34 | MC | 0.68 | 0.49 | 0.56 | | | | |
| 35 | SA | 0.48 | 0.57 | 0.71 | | | | |
| 36 | MC | 0.64 | 0.46 | 0.25 | | | | |
| 37 | MC | 0.69 | 0.59 | 0.22 | | | | |
| 38 | MC | 0.61 | 0.48 | 0.24 | | | | |
| 39 | MC | 0.72 | 0.49 | 0.23 | | | | |
| 40 | MC | 0.66 | 0.51 | 0.21 | | | | |
| 41 | SA | 0.38 | 0.41 | 0.27 | | | | |
| 42 | TE | 0.44 | 0.52 | 1.28 | | | | |

**Table 6-13. Item Analysis, Mathematics Grade 4**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags Corr | Flags Distractor | Flags Omit | Flags p-Value |
|---|---|---|---|---|---|---|---|---|
| 1 | MC | 0.71 | 0.43 | 0.14 | | | | |
| 2 | SA | 0.55 | 0.59 | 0.21 | | | | |
| 3 | TE | 0.69 | 0.40 | 0.11 | | | | |
| 4 | MC | 0.48 | 0.31 | 0.18 | | | | |
| 5 | MC | 0.61 | 0.45 | 0.09 | | | | |
| 6 | MC | 0.63 | 0.48 | 0.41 | | | | |
| 7 | MC | 0.61 | 0.41 | 0.38 | | | | |
| 8 | SA | 0.35 | 0.63 | 0.24 | | | | |
| 9 | MC | 0.70 | 0.52 | 0.16 | | | | |
| 10 | TE | 0.55 | 0.62 | 0.16 | | | | |
| 11 | MC | 0.54 | 0.44 | 0.17 | | | | |
| 12 | MC | 0.58 | 0.42 | 0.16 | | | | |
| 13 | MC | 0.50 | 0.60 | 0.34 | | | | |
| 14 | MC | 0.41 | 0.29 | 0.46 | | + | | |
| 15 | MC | 0.40 | 0.29 | 0.20 | | + | | |
| 16 | SA | 0.29 | 0.50 | 0.25 | | | | |
| 17 | MC | 0.26 | 0.06 | 0.22 | + | + | | |
| 18 | MC | 0.40 | 0.49 | 0.24 | | | | |
| 19 | TE | 0.32 | 0.60 | 0.20 | | | | |
| 20 | SA | 0.41 | 0.50 | 0.20 | | | | |
| 21 | MC | 0.50 | 0.50 | 0.19 | | | | |
| 22 | MC | 0.30 | 0.17 | 0.49 | | + | | |
| 23 | MC | 0.51 | 0.50 | 0.43 | | | | |
| 24 | TE | 0.69 | 0.39 | 0.48 | | | | |
| 25 | MC | 0.82 | 0.39 | 0.15 | | | | |
| 26 | MC | 0.37 | 0.45 | 0.17 | | | | |
| 27 | MC | 0.33 | 0.58 | 0.20 | | | | |
| 28 | MC | 0.34 | 0.19 | 0.15 | | + | | |
| 29 | MC | 0.54 | 0.49 | 0.35 | | | | |
| 30 | MC | 0.40 | 0.33 | 0.51 | | | | |
| 31 | SA | 0.42 | 0.57 | 0.21 | | | | |
| 32 | TE | 0.20 | 0.50 | 0.73 | | | | |
| 33 | MC | 0.49 | 0.19 | 0.18 | | | | |
| 34 | MC | 0.39 | 0.42 | 0.20 | | | | |
| 35 | SA | 0.18 | 0.54 | 0.27 | | | | + |
| 36 | MC | 0.52 | 0.43 | 0.41 | | | | |
| 37 | MC | 0.84 | 0.37 | 0.37 | | | | |
| 38 | MC | 0.35 | 0.55 | 0.22 | | | | |
| 39 | TE | 0.67 | 0.45 | 0.21 | | | | |
| 40 | SA | 0.61 | 0.50 | 0.21 | | | | |
| 41 | MC | 0.64 | 0.53 | 0.19 | | | | |
| 42 | MC | 0.68 | 0.45 | 0.19 | | | | |
| 43 | SA | 0.45 | 0.48 | 0.27 | | | | |

**Table 6-13 (continued). Item Analysis, Mathematics Grade 4**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|-------|-----------|-----------|------|--------------|-------|-----------|------|-----------|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 44 | MC | 0.56 | 0.37 | 0.20 | | | | |
| 45 | SA | 0.51 | 0.60 | 0.38 | | | | |
| 46 | MC | 0.58 | 0.43 | 0.44 | | | | |

**Table 6-14. Item Analysis, Mathematics Grade 5**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | SA | 0.30 | 0.59 | 0.23 | | | | |
| 2 | MC | 0.34 | 0.20 | 0.22 | | | | |
| 3 | TE | 0.52 | 0.55 | 0.18 | | | | |
| 4 | MC | 0.52 | 0.46 | 0.17 | | | | |
| 5 | MC | 0.39 | 0.45 | 0.20 | | | | |
| 6 | SA | 0.35 | 0.44 | 0.49 | | | | |
| 7 | TE | 0.20 | 0.53 | 0.45 | | | | |
| 8 | MC | 0.66 | 0.52 | 0.18 | | | | |
| 9 | MC | 0.75 | 0.41 | 0.21 | | | | |
| 10 | SA | 0.48 | 0.57 | 0.24 | | | | |
| 11 | MC | 0.33 | 0.21 | 0.28 | | + | | |
| 12 | MC | 0.53 | 0.31 | 0.20 | | | | |
| 13 | SA | 0.48 | 0.55 | 0.44 | | | | |
| 14 | MC | 0.65 | 0.35 | 0.44 | | | | |
| 15 | MC | 0.38 | 0.44 | 0.26 | | | | |
| 16 | SA | 0.30 | 0.35 | 0.26 | | | | |
| 17 | MC | 0.49 | 0.32 | 0.25 | | | | |
| 18 | SA | 0.43 | 0.47 | 0.30 | | | | |
| 19 | MC | 0.63 | 0.54 | 0.21 | | | | |
| 20 | MC | 0.38 | 0.26 | 0.26 | | + | | |
| 21 | SA | 0.19 | 0.43 | 0.33 | | | | + |
| 22 | TE | 0.52 | 0.50 | 0.68 | | | | |
| 23 | MC | 0.34 | 0.43 | 0.55 | | + | | |
| 24 | MC | 0.70 | 0.38 | 0.10 | | | | |
| 25 | TE | 0.21 | 0.58 | 0.34 | | | | |
| 26 | MC | 0.38 | 0.37 | 0.15 | | | | |
| 27 | SA | 0.43 | 0.60 | 0.28 | | | | |
| 28 | MC | 0.47 | 0.42 | 0.15 | | | | |
| 29 | MC | 0.76 | 0.38 | 0.22 | | | | |
| 30 | TE | 0.29 | 0.58 | 2.55 | | | | |
| 31 | TE | 0.53 | 0.59 | 0.24 | | | | |
| 32 | MC | 0.71 | 0.39 | 0.23 | | | | |
| 33 | MC | 0.42 | 0.30 | 0.17 | | + | | |
| 34 | SA | 0.26 | 0.57 | 0.35 | | | | |
| 35 | MC | 0.61 | 0.57 | 0.17 | | | | |
| 36 | MC | 0.32 | 0.46 | 0.25 | | | | |
| 37 | SA | 0.47 | 0.55 | 0.67 | | | | |
| 38 | MC | 0.48 | 0.50 | 0.22 | | | | |
| 39 | SA | 0.24 | 0.55 | 0.36 | | | | |
| 40 | MC | 0.47 | 0.33 | 0.15 | | | | |
| 41 | SA | 0.68 | 0.36 | 0.22 | | | | |
| 42 | MC | 0.83 | 0.45 | 0.19 | | | | |
| 43 | MC | 0.22 | 0.25 | 0.25 | | | | |

**Table 6-14 (continued). Item Analysis, Mathematics Grade 5**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 44 | SA | 0.61 | 0.52 | 0.26 | | | | |
| 45 | MC | 0.24 | 0.14 | 0.29 | + | + | | |
| 46 | MC | 0.55 | 0.47 | 0.38 | | | | |

**Table 6-15. Item Analysis, Mathematics Grade 6**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | MC | 0.36 | 0.38 | 0.18 | | + | | |
| 2 | MC | 0.36 | 0.48 | 0.10 | | | | |
| 3 | TE | 0.51 | 0.56 | 0.96 | | | | |
| 4 | MC | 0.51 | 0.38 | 0.25 | | | | |
| 5 | MC | 0.51 | 0.47 | 0.21 | | | | |
| 6 | SA | 0.20 | 0.56 | 0.47 | | | | |
| 7 | SA | 0.41 | 0.55 | 0.64 | | | | |
| 8 | MC | 0.79 | 0.39 | 0.24 | | | | |
| 9 | MC | 0.41 | 0.44 | 0.24 | | | | |
| 10 | SA | 0.15 | 0.48 | 0.24 | | | | + |
| 11 | MC | 0.34 | 0.35 | 0.23 | | | | |
| 12 | SA | 0.46 | 0.60 | 0.43 | | | | |
| 13 | MC | 0.75 | 0.39 | 0.23 | | | | |
| 14 | TE | 0.35 | 0.66 | 0.50 | | | | |
| 15 | MC | 0.48 | 0.49 | 0.26 | | | | |
| 16 | TE | 0.32 | 0.57 | 1.00 | | | | |
| 17 | MC | 0.35 | 0.49 | 0.18 | | | | |
| 18 | MC | 0.53 | 0.52 | 0.30 | | | | |
| 19 | MC | 0.69 | 0.39 | 0.28 | | | | |
| 20 | SA | 0.58 | 0.55 | 0.29 | | | | |
| 21 | TE | 0.53 | 0.55 | 0.40 | | | | |
| 22 | MC | 0.49 | 0.42 | 0.47 | | | | |
| 23 | TE | 0.68 | 0.34 | 2.74 | | | | |
| 24 | MC | 0.62 | 0.40 | 0.52 | | | | |
| 25 | MC | 0.30 | 0.31 | 0.51 | | + | | |
| 26 | MC | 0.43 | 0.37 | 0.31 | | | | |
| 27 | SA | 0.36 | 0.23 | 0.60 | | | | |
| 28 | SA | 0.33 | 0.54 | 0.61 | | | | |
| 29 | SA | 0.29 | 0.41 | 0.76 | | | | |
| 30 | MC | 0.57 | 0.34 | 0.32 | | | | |
| 31 | TE | 0.44 | 0.60 | 0.59 | | | | |
| 32 | MC | 0.28 | 0.39 | 0.54 | | | | |
| 33 | MC | 0.64 | 0.47 | 0.48 | | | | |
| 34 | MC | 0.29 | 0.46 | 0.57 | | | | |
| 35 | MC | 0.42 | 0.47 | 0.60 | | | | |
| 36 | SA | 0.17 | 0.56 | 0.74 | | | | + |
| 37 | MC | 0.59 | 0.37 | 0.45 | | | | |
| 38 | MC | 0.63 | 0.31 | 0.45 | | + | | |
| 39 | TE | 0.50 | 0.55 | 1.70 | | | | |
| 40 | MC | 0.29 | 0.48 | 0.46 | | | | |
| 41 | MC | 0.51 | 0.48 | 0.41 | | | | |
| 42 | MC | 0.38 | 0.26 | 0.41 | | | | |
| 43 | MC | 0.35 | 0.23 | 0.42 | | + | | |

**Table 6-15 (continued). Item Analysis, Grade 6**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 44 | MC | 0.45 | 0.41 | 0.53 | | | | |
| 45 | TE | 0.36 | 0.58 | 0.65 | | | | |
| 46 | MC | 0.26 | 0.49 | 0.53 | | | | |

**Table 6-16. Item Analysis, Mathematics Grade 7**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | SA | 0.42 | 0.65 | 0.30 | | | | |
| 2 | SA | 0.43 | 0.61 | 0.29 | | | | |
| 3 | MC | 0.47 | 0.33 | 0.18 | | | | |
| 4 | SA | 0.22 | 0.48 | 0.45 | | | | |
| 5 | MC | 0.62 | 0.49 | 0.18 | | | | |
| 6 | MC | 0.42 | 0.44 | 0.20 | | | | |
| 7 | MC | 0.49 | 0.43 | 0.17 | | | | |
| 8 | TE | 0.17 | 0.42 | 0.17 | | | | + |
| 9 | MC | 0.33 | 0.28 | 0.18 | | + | | |
| 10 | MC | 0.47 | 0.41 | 0.19 | | | | |
| 11 | MC | 0.53 | 0.34 | 0.15 | | | | |
| 12 | MC | 0.72 | 0.26 | 0.17 | | | | |
| 13 | MC | 0.39 | 0.38 | 0.22 | | | | |
| 14 | MC | 0.43 | 0.41 | 0.20 | | | | |
| 15 | SA | 0.28 | 0.42 | 0.68 | | | | |
| 16 | TE | 0.72 | 0.22 | 0.33 | | | | |
| 17 | MC | 0.47 | 0.55 | 0.51 | | | | |
| 18 | MC | 0.40 | 0.21 | 0.50 | | | | |
| 19 | TE | 0.32 | 0.63 | 1.30 | | | | |
| 20 | MC | 0.37 | 0.25 | 0.43 | | + | | |
| 21 | TE | 0.58 | 0.48 | 0.46 | | | | |
| 22 | MC | 0.34 | 0.38 | 0.51 | | | | |
| 23 | MC | 0.28 | 0.20 | 0.45 | | | | |
| 24 | SA | 0.35 | 0.55 | 0.71 | | | | |
| 25 | SA | 0.17 | 0.57 | 0.87 | | | | + |
| 26 | MC | 0.49 | 0.48 | 0.44 | | | | |
| 27 | MC | 0.43 | 0.47 | 0.43 | | | | |
| 28 | MC | 0.52 | 0.16 | 0.48 | | | | |
| 29 | SA | 0.51 | 0.56 | 1.02 | | | | |
| 30 | MC | 0.36 | 0.35 | 0.60 | | | | |
| 31 | TE | 0.63 | 0.24 | 1.58 | | | | |
| 32 | MC | 0.30 | 0.19 | 0.58 | | + | | |
| 33 | MC | 0.47 | 0.52 | 0.58 | | | | |
| 34 | MC | 0.54 | 0.33 | 0.51 | | | | |
| 35 | SA | 0.36 | 0.68 | 0.92 | | | | |
| 36 | TE | 0.64 | 0.51 | 0.63 | | | | |
| 37 | MC | 0.26 | 0.48 | 0.56 | | | | |
| 38 | SA | 0.29 | 0.57 | 0.90 | | | | |
| 39 | MC | 0.48 | 0.59 | 0.60 | | | | |
| 40 | SA | 0.27 | 0.53 | 0.96 | | | | |
| 41 | MC | 0.42 | 0.33 | 0.68 | | | | |
| 42 | TE | 0.44 | 0.25 | 1.34 | | | | |
| 43 | MC | 0.44 | 0.52 | 0.71 | | | | |

**Table 6-16 (continued). Item Analysis, Grade 7**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 44 | MC | 0.24 | 0.18 | 0.76 | | | | |
| 45 | SA | 0.62 | 0.60 | 0.94 | | | | |
| 46 | MC | 0.59 | 0.48 | 0.71 | | | | |

**Table 6-17. Item Analysis, Mathematics Grade 8**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | MC | 0.59 | 0.49 | 0.12 | | | | |
| 2 | MC | 0.51 | 0.39 | 0.15 | | | | |
| 3 | MC | 0.36 | 0.31 | 0.16 | | + | | |
| 4 | MC | 0.44 | 0.50 | 0.14 | | + | | |
| 5 | SA | 0.14 | 0.50 | 0.34 | | | | + |
| 6 | MC | 0.31 | 0.44 | 0.17 | | | | |
| 7 | MC | 0.38 | 0.34 | 0.15 | | | | |
| 8 | SA | 0.39 | 0.67 | 0.44 | | | | |
| 9 | TE | 0.18 | 0.54 | 0.26 | | | | + |
| 10 | MC | 0.51 | 0.39 | 0.21 | | | | |
| 11 | SA | 0.42 | 0.65 | 0.36 | | | | |
| 12 | MC | 0.46 | 0.29 | 0.20 | | | | |
| 13 | MC | 0.45 | 0.34 | 0.16 | | | | |
| 14 | MC | 0.74 | 0.33 | 0.19 | | | | |
| 15 | TE | 0.43 | 0.52 | 0.25 | | | | |
| 16 | MC | 0.74 | 0.43 | 0.23 | | | | |
| 17 | SA | 0.33 | 0.54 | 0.81 | | | | |
| 18 | MC | 0.53 | 0.41 | 0.32 | | | | |
| 19 | MC | 0.44 | 0.34 | 0.47 | | | | |
| 20 | MC | 0.57 | 0.36 | 0.48 | | | | |
| 21 | MC | 0.39 | 0.24 | 0.36 | | + | | |
| 22 | MC | 0.68 | 0.42 | 0.38 | | | | |
| 23 | TE | 0.50 | 0.52 | 1.78 | | | | |
| 24 | MC | 0.54 | 0.37 | 0.34 | | | | |
| 25 | TE | 0.20 | 0.49 | 0.75 | | | | |
| 26 | MC | 0.27 | 0.19 | 0.42 | | | | |
| 27 | SA | 0.30 | 0.67 | 0.90 | | | | |
| 28 | MC | 0.53 | 0.57 | 0.40 | | | | |
| 29 | MC | 0.30 | 0.39 | 0.61 | | | | |
| 30 | MC | 0.53 | 0.51 | 0.49 | | | | |
| 31 | SA | 0.25 | 0.61 | 2.01 | | | | |
| 32 | TE | 0.32 | 0.51 | 0.82 | | | | |
| 33 | MC | 0.52 | 0.38 | 0.64 | | | | |
| 34 | TE | 0.22 | 0.66 | 0.71 | | | | |
| 35 | MC | 0.43 | 0.26 | 0.65 | | + | | |
| 36 | MC | 0.43 | 0.43 | 0.54 | | | | |
| 37 | TE | 0.28 | 0.25 | 1.44 | | | | |
| 38 | TE | 0.44 | 0.38 | 1.16 | | | | |
| 39 | MC | 0.38 | 0.37 | 0.61 | | | | |
| 40 | MC | 0.45 | 0.30 | 0.55 | | | | |
| 41 | MC | 0.37 | 0.23 | 0.70 | | + | | |
| 42 | TE | 0.23 | 0.50 | 2.54 | | | | |
| 43 | MC | 0.66 | 0.42 | 0.66 | | | | |

**Table 6-17 (continued). Item Analysis, Grade 8**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 44 | MC | 0.60 | 0.39 | 0.66 | | | | |
| 45 | MC | 0.65 | 0.43 | 0.65 | | | | |
| 46 | MC | 0.37 | 0.35 | 0.64 | | | | |

**Table 6-18. Item Analysis, Science Grade 4**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | TE | 0.77 | 0.36 | 0.21 | | | | |
| 2 | TE | 0.73 | 0.52 | 0.11 | | | | |
| 3 | MC | 0.50 | 0.32 | 0.13 | | | | |
| 4 | TE | 0.32 | 0.32 | 0.13 | | | | |
| 5 | TE | 0.44 | 0.44 | 0.15 | | | | |
| 6 | TE | 0.52 | 0.39 | 0.35 | | | | |
| 7 | TE | 0.60 | 0.52 | 0.18 | | | | |
| 8 | MC | 0.35 | 0.30 | 0.14 | | | | |
| 9 | TE | 0.59 | 0.43 | 0.17 | | | | |
| 10 | EBSR | 0.41 | 0.45 | 0.09 | | | | |
| 11 | MC | 0.65 | 0.38 | 0.11 | | | | |
| 12 | MC | 0.66 | 0.49 | 0.12 | | | | |
| 13 | MC | 0.38 | 0.12 | 0.25 | + | | | |
| 14 | TE | 0.42 | 0.57 | 0.26 | | | | |
| 15 | TE | 0.42 | 0.37 | 0.15 | | | | |
| 16 | TE | 0.82 | 0.41 | 0.04 | | | | |
| 17 | TE | 0.68 | 0.43 | 0.17 | | | | |
| 18 | MC | 0.59 | 0.34 | 0.13 | | | | |
| 19 | TE | 0.50 | 0.24 | 0.34 | | | | |
| 20 | TE | 0.87 | 0.32 | 0.15 | | | | |
| 21 | TE | 0.51 | 0.42 | 0.31 | | | | |
| 22 | MC | 0.50 | 0.55 | 0.17 | | | | |
| 23 | MC | 0.57 | 0.43 | 0.17 | | | | |
| 24 | TE | 0.23 | 0.40 | 0.10 | | | | |
| 25 | MC | 0.42 | 0.37 | 0.12 | | | | |
| 26 | MC | 0.64 | 0.35 | 0.08 | | | | |
| 27 | MC | 0.83 | 0.33 | 0.09 | | | | |
| 28 | MC | 0.38 | 0.35 | 0.16 | | | | |
| 29 | TE | 0.39 | 0.34 | 0.58 | | | | |
| 30 | TE | 0.36 | 0.16 | 0.16 | | | | |
| 31 | TE | 0.45 | 0.61 | 0.22 | | | | |
| 32 | TE | 0.34 | 0.36 | 0.31 | | | | |
| 33 | MC | 0.66 | 0.44 | 0.13 | | | | |
| 34 | TE | 0.71 | 0.41 | 0.11 | | | | |
| 35 | TE | 0.89 | 0.37 | 0.10 | | | | |
| 36 | TE | 0.55 | 0.34 | 0.18 | | | | |
| 37 | TE | 0.35 | 0.30 | 0.17 | | | | |
| 38 | MC | 0.44 | 0.42 | 0.17 | | | | |
| 39 | TE | 0.53 | 0.38 | 0.12 | | | | |
| 40 | MC | 0.51 | 0.52 | 0.16 | | | | |

**Table 6-19. Item Analysis, Science Grade 8**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | p-Value |
| 1 | MS | 0.43 | 0.44 | 0.08 | | | | |
| 2 | MC | 0.47 | 0.22 | 0.15 | | + | | |
| 3 | TE | 0.73 | 0.31 | 0.24 | | | | |
| 4 | TE | 0.66 | 0.34 | 0.76 | | | | |
| 5 | TE | 0.43 | 0.40 | 0.60 | | | | |
| 6 | TE | 0.22 | 0.24 | 0.08 | | | | |
| 7 | TE | 0.47 | 0.39 | 0.29 | | | | |
| 8 | MC | 0.61 | 0.39 | 0.28 | | | | |
| 9 | TE | 0.78 | 0.32 | 0.23 | | | | |
| 10 | MC | 0.41 | 0.16 | 0.16 | | + | | |
| 11 | MC | 0.54 | 0.41 | 0.19 | | | | |
| 12 | MC | 0.44 | 0.23 | 0.30 | | | | |
| 13 | TE | 0.40 | 0.49 | 0.15 | | | | |
| 14 | TE | 0.68 | 0.36 | 0.19 | | | | |
| 15 | TE | 0.33 | 0.41 | 0.18 | | | | |
| 16 | TE | 0.60 | 0.37 | 0.06 | | | | |
| 17 | MC | 0.49 | 0.24 | 0.23 | | | | |
| 18 | TE | 0.49 | 0.31 | 0.15 | | | | |
| 19 | MC | 0.53 | 0.42 | 0.31 | | | | |
| 20 | MC | 0.48 | 0.36 | 0.22 | | | | |
| 21 | TE | 0.79 | 0.38 | 0.18 | | | | |
| 22 | TE | 0.53 | 0.43 | 0.18 | | | | |
| 23 | TE | 0.69 | 0.43 | 0.26 | | | | |
| 24 | TE | 0.61 | 0.45 | 0.22 | | | | |
| 25 | TE | 0.46 | 0.04 | 0.20 | + | | | |
| 26 | TE | 0.48 | 0.55 | 0.10 | | | | |
| 27 | TE | 0.57 | 0.17 | 0.24 | | | | |
| 28 | TE | 0.32 | 0.51 | 0.08 | | | | |
| 29 | TE | 0.75 | 0.33 | 0.14 | | | | |
| 30 | TE | 0.75 | 0.37 | 0.17 | | | | |
| 31 | TE | 0.40 | 0.37 | 0.58 | | | | |
| 32 | MC | 0.50 | 0.43 | 0.30 | | | | |
| 33 | TE | 0.60 | 0.42 | 0.23 | | | | |
| 34 | EBSR | 0.42 | 0.48 | 0.15 | | | | |
| 35 | TE | 0.41 | 0.50 | 0.16 | | | | |
| 36 | MC | 0.55 | 0.47 | 0.22 | | | | |
| 37 | TE | 0.39 | 0.40 | 0.27 | | | | |
| 38 | MC | 0.61 | 0.55 | 0.26 | | | | |
| 39 | TE | 0.27 | 0.36 | 0.23 | | | | |
| 40 | TE | 0.64 | 0.38 | 0.28 | | | | |

**Table 6-20. Item Analysis, Social Studies Grade 4**

| Grade | Item Type | p-Value | Corr | Percent Omit | Flags Corr | Flags Distractor | Flags Omit | Flags p-Value |
|---|---|---|---|---|---|---|---|---|
| 1 | TE | 0.54 | 0.47 | 0.07 | | | | |
| 2 | MC | 0.72 | 0.41 | 0.12 | | | | |
| 3 | TE | 0.25 | 0.20 | 0.16 | | | | |
| 4 | MC | 0.72 | 0.48 | 0.16 | | | | |
| 5 | MC | 0.72 | 0.44 | 0.10 | | | | |
| 6 | MS | 0.36 | 0.50 | 0.11 | | | | |
| 7 | MC | 0.73 | 0.49 | 0.16 | | | | |
| 8 | MC | 0.51 | 0.29 | 0.17 | | + | | |
| 9 | TE | 0.81 | 0.50 | 0.15 | | | | |
| 10 | MC | 0.35 | 0.23 | 0.18 | | | | |
| 11 | MC | 0.53 | 0.34 | 0.16 | | | | |
| 12 | MS | 0.37 | 0.50 | 0.61 | | | | |
| 13 | MC | 0.65 | 0.47 | 0.24 | | | | |
| 14 | TE | 0.61 | 0.45 | 0.15 | | | | |
| 15 | TE | 0.64 | 0.46 | 0.15 | | | | |
| 16 | MC | 0.76 | 0.49 | 0.19 | | | | |
| 17 | TE | 0.76 | 0.51 | 0.35 | | | | |
| 18 | MC | 0.59 | 0.50 | 0.17 | | | | |
| 19 | MC | 0.40 | 0.41 | 0.17 | | | | |
| 20 | MC | 0.53 | 0.40 | 0.46 | | | | |
| 21 | MC | 0.44 | 0.31 | 0.05 | | | | |
| 22 | MC | 0.71 | 0.53 | 0.13 | | | | |
| 23 | TE | 0.53 | 0.52 | 0.10 | | | | |
| 24 | MC | 0.60 | 0.49 | 0.19 | | | | |
| 25 | TE | 0.68 | 0.61 | 0.13 | | | | |
| 26 | TE | 0.50 | 0.37 | 0.25 | | | | |
| 27 | TE | 0.51 | 0.63 | 0.20 | | | | |
| 28 | MC | 0.46 | 0.30 | 0.34 | | | | |
| 29 | TE | 0.53 | 0.36 | 0.15 | | | | |
| 30 | TE | 0.28 | 0.32 | 0.14 | | | | |
| 31 | MC | 0.71 | 0.38 | 0.14 | | | | |
| 32 | MC | 0.71 | 0.45 | 0.18 | | | | |
| 33 | TE | 0.43 | 0.43 | 0.15 | | | | |
| 34 | MC | 0.74 | 0.49 | 0.26 | | | | |
| 35 | MC | 0.51 | 0.45 | 0.27 | | | | |
| 36 | MC | 0.79 | 0.49 | 0.21 | | | | |
| 37 | MC | 0.70 | 0.43 | 0.16 | | | | |
| 38 | TE | 0.38 | 0.30 | 0.21 | | | | |
| 39 | MC | 0.57 | 0.36 | 0.15 | | | | |
| 40 | MC | 0.55 | 0.27 | 0.34 | | | | |

**Table 6-21. Item Analysis, Social Studies Grade 8**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 1 | MC | 0.75 | 0.51 | 0.05 | | | | |
| 2 | MC | 0.41 | 0.32 | 0.13 | | | | |
| 3 | MC | 0.56 | 0.56 | 0.17 | | | | |
| 4 | MS | 0.25 | 0.26 | 0.16 | | | | |
| 5 | MC | 0.64 | 0.49 | 0.16 | | | | |
| 6 | MC | 0.78 | 0.51 | 0.23 | | | | |
| 7 | MC | 0.52 | 0.40 | 0.21 | | | | |
| 8 | MC | 0.39 | 0.38 | 0.28 | | | | |
| 9 | MC | 0.67 | 0.54 | 0.14 | | | | |
| 10 | MC | 0.80 | 0.41 | 0.18 | | | | |
| 11 | MC | 0.53 | 0.42 | 0.21 | | | | |
| 12 | MC | 0.52 | 0.52 | 0.25 | | | | |
| 13 | MC | 0.71 | 0.54 | 0.21 | | | | |
| 14 | MC | 0.75 | 0.47 | 0.33 | | | | |
| 15 | MC | 0.72 | 0.40 | 0.28 | | | | |
| 16 | MC | 0.75 | 0.39 | 0.31 | | | | |
| 17 | MC | 0.46 | 0.44 | 0.31 | | | | |
| 18 | MC | 0.72 | 0.38 | 0.27 | | | | |
| 19 | TE | 0.50 | 0.60 | 0.30 | | | | |
| 20 | MC | 0.50 | 0.43 | 0.36 | | | | |
| 21 | MC | 0.79 | 0.54 | 0.06 | | | | |
| 22 | MC | 0.83 | 0.45 | 0.16 | | | | |
| 23 | MC | 0.36 | 0.41 | 0.18 | | | | |
| 24 | MC | 0.55 | 0.35 | 0.15 | | | | |
| 25 | TE | 0.77 | 0.46 | 0.13 | | | | |
| 26 | MC | 0.65 | 0.56 | 0.23 | | | | |
| 27 | MC | 0.55 | 0.28 | 0.13 | | + | | |
| 28 | MC | 0.63 | 0.49 | 0.22 | | | | |
| 29 | MC | 0.48 | 0.39 | 0.27 | | | | |
| 30 | MC | 0.63 | 0.44 | 0.35 | | | | |
| 31 | MC | 0.44 | 0.27 | 0.25 | | | | |
| 32 | TE | 0.48 | 0.45 | 0.30 | | | | |
| 33 | TE | 0.63 | 0.47 | 0.33 | | | | |
| 34 | TE | 0.57 | 0.42 | 0.28 | | | | |
| 35 | MC | 0.48 | 0.37 | 0.31 | | | | |
| 36 | MC | 0.69 | 0.45 | 0.29 | | | | |
| 37 | MC | 0.48 | 0.33 | 0.24 | | | | |
| 38 | MC | 0.57 | 0.43 | 0.31 | | | | |
| 39 | MC | 0.59 | 0.45 | 0.30 | | | | |
| 40 | TE | 0.46 | 0.37 | 0.61 | | | | |

**Table 6-22. Item Analysis, Social Studies Grade 10**

| Grade | Item Type | *p*-Value | Corr | Percent Omit | Flags | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Corr | Distractor | Omit | *p*-Value |
| 1 | TE | 0.70 | 0.42 | 0.35 | | | | |
| 2 | MC | 0.63 | 0.44 | 0.14 | | | | |
| 3 | MC | 0.77 | 0.32 | 0.15 | | | | |
| 4 | MC | 0.63 | 0.39 | 0.23 | | | | |
| 5 | MC | 0.41 | 0.28 | 0.24 | | + | | |
| 6 | MC | 0.70 | 0.35 | 0.19 | | | | |
| 7 | MC | 0.62 | 0.41 | 0.25 | | | | |
| 8 | MC | 0.48 | 0.44 | 0.41 | | | | |
| 9 | MC | 0.70 | 0.46 | 0.42 | | | | |
| 10 | MC | 0.76 | 0.27 | 0.42 | | | | |
| 11 | MC | 0.72 | 0.41 | 0.42 | | | | |
| 12 | MC | 0.54 | 0.55 | 0.44 | | | | |
| 13 | MC | 0.37 | 0.34 | 0.51 | | | | |
| 14 | MC | 0.57 | 0.38 | 0.63 | | | | |
| 15 | MC | 0.72 | 0.54 | 0.62 | | | | |
| 16 | MC | 0.48 | 0.48 | 0.77 | | | | |
| 17 | MC | 0.33 | 0.27 | 0.71 | | + | | |
| 18 | MC | 0.66 | 0.43 | 0.62 | | | | |
| 19 | MC | 0.52 | 0.29 | 0.66 | | | | |
| 20 | MS | 0.24 | 0.27 | 0.64 | | | | |
| 21 | MC | 0.59 | 0.39 | 0.11 | | | | |
| 22 | MS | 0.42 | 0.50 | 0.26 | | | | |
| 23 | MC | 0.69 | 0.53 | 0.28 | | | | |
| 24 | MC | 0.62 | 0.38 | 0.32 | | | | |
| 25 | ESR | 0.23 | 0.37 | 0.24 | | | | |
| 26 | TE | 0.37 | 0.50 | 0.62 | | | | |
| 27 | MC | 0.55 | 0.42 | 0.58 | | | | |
| 28 | EBSR | 0.35 | 0.50 | 0.41 | | | | |
| 29 | MC | 0.42 | 0.35 | 0.60 | | | | |
| 30 | MC | 0.55 | 0.38 | 0.64 | | | | |
| 31 | TE | 0.32 | 0.26 | 0.79 | | | | |
| 32 | MC | 0.58 | 0.57 | 0.85 | | | | |
| 33 | MC | 0.49 | 0.43 | 0.89 | | | | |
| 34 | TE | 0.39 | 0.44 | 0.83 | | | | |
| 35 | MC | 0.53 | 0.49 | 0.80 | | | | |
| 36 | TE | 0.37 | 0.47 | 0.90 | | | | |
| 37 | MC | 0.68 | 0.36 | 0.83 | | | | |
| 38 | MC | 0.65 | 0.28 | 0.81 | | | | |
| 39 | MC | 0.47 | 0.31 | 0.85 | | | | |
| 40 | TE | 0.56 | 0.43 | 1.11 | | | | |

**Table 6-23. Test-Level Descriptive Statistics**

| Content | Grade | N Count | Mean Raw Score | Test Difficulty | Raw Score SD | Skewness | Kurtosis | Min Obtained | Max Obtained | Max Possible | Alpha | SEM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| English Language Arts | 3 | 57800 | 27.20 | 0.54 | 9.70 | -0.03 | -0.90 | 0 | 52 | 53 | 0.88 | 3.33 |
| | 4 | 58244 | 28.87 | 0.55 | 10.06 | -0.01 | -0.93 | 0 | 55 | 56 | 0.89 | 3.29 |
| | 5 | 58969 | 30.58 | 0.56 | 10.01 | 0.00 | -0.89 | 0 | 56 | 56 | 0.89 | 3.28 |
| | 6 | 59678 | 30.64 | 0.57 | 10.37 | -0.20 | -0.88 | 1 | 56 | 56 | 0.90 | 3.33 |
| | 7 | 61722 | 30.18 | 0.56 | 10.49 | -0.08 | -0.89 | 0 | 55 | 56 | 0.89 | 3.50 |
| | 8 | 62550 | 30.24 | 0.55 | 10.37 | -0.05 | -0.90 | 0 | 56 | 56 | 0.89 | 3.48 |
| Mathematics | 3 | 58400 | 22.75 | 0.54 | 10.19 | -0.08 | -1.07 | 0 | 42 | 42 | 0.93 | 2.68 |
| | 4 | 58872 | 22.83 | 0.50 | 10.49 | 0.20 | -1.07 | 0 | 46 | 46 | 0.93 | 2.87 |
| | 5 | 59137 | 20.99 | 0.46 | 10.26 | 0.30 | -0.88 | 0 | 46 | 46 | 0.92 | 2.84 |
| | 6 | 60143 | 20.09 | 0.44 | 10.59 | 0.47 | -0.79 | 0 | 46 | 46 | 0.93 | 2.84 |
| | 7 | 61867 | 19.55 | 0.43 | 10.09 | 0.52 | -0.71 | 0 | 46 | 46 | 0.92 | 2.90 |
| | 8 | 62658 | 19.64 | 0.43 | 10.04 | 0.59 | -0.62 | 0 | 46 | 46 | 0.92 | 2.88 |
| Science | 4 | 58792 | 21.46 | 0.54 | 8.25 | -0.07 | -0.96 | 0 | 40 | 40 | 0.89 | 2.72 |
| | 8 | 62527 | 20.86 | 0.52 | 8.11 | 0.13 | -0.90 | 0 | 40 | 40 | 0.88 | 2.80 |
| Social Studies | 4 | 58789 | 22.82 | 0.57 | 8.82 | -0.21 | -0.97 | 0 | 40 | 40 | 0.91 | 2.68 |
| | 8 | 62531 | 23.47 | 0.59 | 9.00 | -0.20 | -1.00 | 0 | 40 | 40 | 0.91 | 2.68 |
| | 10 | 59157 | 21.11 | 0.53 | 8.69 | 0.06 | -1.00 | 0 | 40 | 40 | 0.90 | 2.77 |

**Table 6-24A. Calibration Sample Demographics Compared to Population, English Language Arts, Grades 3 and 4**

| Grade 3 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 57861 | | 58275 | | 60956 | | 99.29 | 94.92 |
| **Gender** | | | | | | | | |
| Male | 29379 | 50.78 | 29613 | 50.82 | 31115 | 51.05 | 0.04 | 0.27 |
| Female | 28482 | 49.22 | 28662 | 49.18 | 29841 | 48.95 | -0.04 | -0.27 |
| **Race/Ethnicity** | | | | | | | | |
| White | 37874 | 65.46 | 38091 | 65.36 | 39550 | 64.88 | -0.09 | -0.57 |
| African-American | 5599 | 9.68 | 5679 | 9.75 | 6053 | 9.93 | 0.07 | 0.25 |
| Hispanic | 7943 | 13.73 | 7998 | 13.72 | 8521 | 13.98 | 0.00 | 0.25 |
| Asian/Pacific Islander | 2718 | 4.70 | 2744 | 4.71 | 2860 | 4.69 | 0.01 | -0.01 |
| American Indian | 592 | 1.02 | 599 | 1.03 | 646 | 1.06 | 0.00 | 0.04 |
| Other | 3135 | 5.42 | 3164 | 5.43 | 3326 | 5.46 | 0.01 | 0.04 |
| **LEP** | | | | | | | | |
| No | 52807 | 91.27 | 53189 | 91.27 | 55533 | 91.10 | 0.01 | -0.16 |
| Yes | 5054 | 8.73 | 5086 | 8.73 | 5423 | 8.90 | -0.01 | 0.16 |
| **Disability** | | | | | | | | |
| No | 50029 | 86.46 | 50372 | 86.44 | 52153 | 85.56 | -0.03 | -0.91 |
| Yes | 7832 | 13.54 | 7903 | 13.56 | 8803 | 14.44 | 0.03 | 0.91 |
| **SES Disadvantaged** | | | | | | | | |
| No | 33570 | 58.02 | 33821 | 58.04 | 35209 | 57.76 | 0.02 | -0.26 |
| Yes | 24291 | 41.98 | 24454 | 41.96 | 25747 | 42.24 | -0.02 | 0.26 |
| **Grade 4** | N | % | N | % | N | % | % | % |
| All Students | 58296 | | 58773 | | 61465 | | 99.19 | 94.84 |
| **Gender** | | | | | | | | |
| Male | 29847 | 51.20 | 30084 | 51.19 | 31558 | 51.34 | -0.01 | 0.14 |
| Female | 28449 | 48.80 | 28689 | 48.81 | 29907 | 48.66 | 0.01 | -0.14 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38172 | 65.48 | 38419 | 65.37 | 39960 | 65.01 | -0.11 | -0.47 |
| African-American | 5532 | 9.49 | 5604 | 9.53 | 6001 | 9.76 | 0.05 | 0.27 |
| Hispanic | 8210 | 14.08 | 8305 | 14.13 | 8792 | 14.30 | 0.05 | 0.22 |
| Asian/Pacific Islander | 2669 | 4.58 | 2706 | 4.60 | 2819 | 4.59 | 0.03 | 0.01 |
| American Indian | 588 | 1.01 | 593 | 1.01 | 625 | 1.02 | 0.00 | 0.01 |
| Other | 3125 | 5.36 | 3146 | 5.35 | 3268 | 5.32 | -0.01 | -0.04 |
| **LEP** | | | | | | | | |
| No | 52983 | 90.89 | 53390 | 90.84 | 55701 | 90.62 | -0.05 | -0.26 |
| Yes | 5313 | 9.11 | 5383 | 9.16 | 5764 | 9.38 | 0.05 | 0.26 |
| **Disability** | | | | | | | | |
| No | 50507 | 86.64 | 50901 | 86.61 | 52699 | 85.74 | -0.03 | -0.90 |
| Yes | 7789 | 13.36 | 7872 | 13.39 | 8766 | 14.26 | 0.03 | 0.90 |
| **SES Disadvantaged** | | | | | | | | |
| No | 33830 | 58.03 | 34131 | 58.07 | 35488 | 57.74 | 0.04 | -0.29 |
| Yes | 24466 | 41.97 | 24642 | 41.93 | 25977 | 42.26 | -0.04 | 0.29 |

**Table 6-24B. Calibration Sample Demographics Compared to Population, English Language Arts, Grades 5 and 6**

| Grade 5 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 58842 | | 59044 | | 61735 | | 99.66 | 95.31 |
| **Gender** | | | | | | | | |
| Male | 29981 | 50.95 | 30093 | 50.97 | 31576 | 51.15 | 0.02 | 0.20 |
| Female | 28861 | 49.05 | 28951 | 49.03 | 30159 | 48.85 | -0.02 | -0.20 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38737 | 65.83 | 38782 | 65.68 | 40292 | 65.27 | -0.15 | -0.57 |
| African-American | 5687 | 9.66 | 5774 | 9.78 | 6170 | 9.99 | 0.11 | 0.33 |
| Hispanic | 8270 | 14.05 | 8320 | 14.09 | 8824 | 14.29 | 0.04 | 0.24 |
| Asian/Pacific Islander | 2585 | 4.39 | 2590 | 4.39 | 2686 | 4.35 | -0.01 | -0.04 |
| American Indian | 631 | 1.07 | 634 | 1.07 | 676 | 1.10 | 0.00 | 0.02 |
| Other | 2932 | 4.98 | 2944 | 4.99 | 3087 | 5.00 | 0.00 | 0.02 |
| **LEP** | | | | | | | | |
| No | 54080 | 91.91 | 54257 | 91.89 | 56611 | 91.70 | -0.01 | -0.21 |
| Yes | 4762 | 8.09 | 4787 | 8.11 | 5124 | 8.30 | 0.01 | 0.21 |
| **Disability** | | | | | | | | |
| No | 51131 | 86.90 | 51294 | 86.87 | 53057 | 85.94 | -0.02 | -0.95 |
| Yes | 7711 | 13.10 | 7750 | 13.13 | 8678 | 14.06 | 0.02 | 0.95 |
| **SES Disadvantaged** | | | | | | | | |
| No | 34672 | 58.92 | 34743 | 58.84 | 36093 | 58.46 | -0.08 | -0.46 |
| Yes | 24170 | 41.08 | 24301 | 41.16 | 25642 | 41.54 | 0.08 | 0.46 |
| **Grade 6** | N | % | N | % | N | % | % | % |
| All Students | 59590 | | 60112 | | 63048 | | 99.13 | 94.52 |
| **Gender** | | | | | | | | |
| Male | 30379 | 50.98 | 30649 | 50.99 | 32243 | 51.14 | 0.01 | 0.16 |
| Female | 29211 | 49.02 | 29463 | 49.01 | 30805 | 48.86 | -0.01 | -0.16 |
| **Race/Ethnicity** | | | | | | | | |
| White | 39662 | 66.56 | 39875 | 66.33 | 41457 | 65.75 | -0.22 | -0.80 |
| African-American | 5723 | 9.60 | 5873 | 9.77 | 6393 | 10.14 | 0.17 | 0.54 |
| Hispanic | 8342 | 14.00 | 8431 | 14.03 | 8975 | 14.24 | 0.03 | 0.24 |
| Asian/Pacific Islander | 2441 | 4.10 | 2470 | 4.11 | 2569 | 4.07 | 0.01 | -0.02 |
| American Indian | 609 | 1.02 | 620 | 1.03 | 655 | 1.04 | 0.01 | 0.02 |
| Other | 2813 | 4.72 | 2843 | 4.73 | 2999 | 4.76 | 0.01 | 0.04 |
| **LEP** | | | | | | | | |
| No | 55634 | 93.36 | 56089 | 93.31 | 58705 | 93.11 | -0.05 | -0.25 |
| Yes | 3956 | 6.64 | 4023 | 6.69 | 4343 | 6.89 | 0.05 | 0.25 |
| **Disability** | | | | | | | | |
| No | 52167 | 87.54 | 52585 | 87.48 | 54476 | 86.40 | -0.06 | -1.14 |
| Yes | 7423 | 12.46 | 7527 | 12.52 | 8572 | 13.60 | 0.06 | 1.14 |
| **SES Disadvantaged** | | | | | | | | |
| No | 35377 | 59.37 | 35632 | 59.28 | 37070 | 58.80 | -0.09 | -0.57 |
| Yes | 24213 | 40.63 | 24480 | 40.72 | 25978 | 41.20 | 0.09 | 0.57 |

**Table 6-24C. Calibration Sample Demographics Compared to Population, English Language Arts, Grades 7 and 8**

| Grade 7 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 61538 | | 61871 | | 65386 | | 99.46 | 94.11 |
| **Gender** | | | | | | | | |
| Male | 31523 | 51.23 | 31682 | 51.21 | 33554 | 51.32 | -0.02 | 0.09 |
| Female | 30015 | 48.77 | 30189 | 48.79 | 31832 | 48.68 | 0.02 | -0.09 |
| **Race/Ethnicity** | | | | | | | | |
| White | 40859 | 66.40 | 40934 | 66.16 | 42791 | 65.44 | -0.24 | -0.95 |
| African-American | 5910 | 9.60 | 6072 | 9.81 | 6717 | 10.27 | 0.21 | 0.67 |
| Hispanic | 8638 | 14.04 | 8702 | 14.06 | 9326 | 14.26 | 0.03 | 0.23 |
| Asian/Pacific Islander | 2663 | 4.33 | 2671 | 4.32 | 2782 | 4.25 | -0.01 | -0.07 |
| American Indian | 616 | 1.00 | 623 | 1.01 | 681 | 1.04 | 0.01 | 0.04 |
| Other | 2852 | 4.63 | 2869 | 4.64 | 3089 | 4.72 | 0.00 | 0.09 |
| **LEP** | | | | | | | | |
| No | 57680 | 93.73 | 57982 | 93.71 | 61152 | 93.52 | -0.02 | -0.21 |
| Yes | 3858 | 6.27 | 3889 | 6.29 | 4234 | 6.48 | 0.02 | 0.21 |
| **Disability** | | | | | | | | |
| No | 54023 | 87.79 | 54280 | 87.73 | 56616 | 86.59 | -0.06 | -1.20 |
| Yes | 7515 | 12.21 | 7591 | 12.27 | 8770 | 13.41 | 0.06 | 1.20 |
| **SES Disadvantaged** | | | | | | | | |
| No | 36951 | 60.05 | 37042 | 59.87 | 38738 | 59.25 | -0.18 | -0.80 |
| Yes | 24587 | 39.95 | 24829 | 40.13 | 26648 | 40.75 | 0.18 | 0.80 |
| **Grade 8** | N | % | N | % | N | % | % | % |
| All Students | 62315 | | 62684 | | 66789 | | 99.41 | 93.30 |
| **Gender** | | | | | | | | |
| Male | 31950 | 51.27 | 32133 | 51.26 | 34164 | 51.15 | -0.01 | -0.12 |
| Female | 30365 | 48.73 | 30551 | 48.74 | 32625 | 48.85 | 0.01 | 0.12 |
| **Race/Ethnicity** | | | | | | | | |
| White | 41677 | 66.88 | 41778 | 66.65 | 44041 | 65.94 | -0.23 | -0.94 |
| African-American | 5981 | 9.60 | 6135 | 9.79 | 6848 | 10.25 | 0.19 | 0.66 |
| Hispanic | 8780 | 14.09 | 8846 | 14.11 | 9588 | 14.36 | 0.02 | 0.27 |
| Asian/Pacific Islander | 2530 | 4.06 | 2540 | 4.05 | 2628 | 3.93 | -0.01 | -0.13 |
| American Indian | 581 | 0.93 | 590 | 0.94 | 647 | 0.97 | 0.01 | 0.04 |
| Other | 2765 | 4.44 | 2795 | 4.46 | 3037 | 4.55 | 0.02 | 0.11 |
| **LEP** | | | | | | | | |
| No | 58458 | 93.81 | 58797 | 93.80 | 62556 | 93.66 | -0.01 | -0.15 |
| Yes | 3856 | 6.19 | 3887 | 6.20 | 4233 | 6.34 | 0.01 | 0.15 |
| **Disability** | | | | | | | | |
| No | 55248 | 88.66 | 55531 | 88.59 | 58317 | 87.32 | -0.07 | -1.35 |
| Yes | 7066 | 11.34 | 7153 | 11.41 | 8472 | 12.68 | 0.07 | 1.35 |
| **SES Disadvantaged** | | | | | | | | |
| No | 37735 | 60.56 | 37833 | 60.36 | 39852 | 59.67 | -0.20 | -0.89 |
| Yes | 24579 | 39.44 | 24851 | 39.64 | 26937 | 40.33 | 0.20 | 0.89 |

**Table 6-25A. Calibration Sample Demographics Compared to Population, Mathematics, Grades 3 and 4**

| Grade 3 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 58345 | | 58449 | | 60956 | | 99.82 | 95.72 |
| **Gender** | | | | | | | | |
| Male | 29650 | 50.82 | 29706 | 50.82 | 31115 | 51.05 | 0.01 | 0.23 |
| Female | 28695 | 49.18 | 28743 | 49.18 | 29841 | 48.95 | -0.01 | -0.23 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38065 | 65.24 | 38088 | 65.16 | 39550 | 64.88 | -0.08 | -0.36 |
| African-American | 5632 | 9.65 | 5675 | 9.71 | 6053 | 9.93 | 0.06 | 0.28 |
| Hispanic | 8134 | 13.94 | 8152 | 13.95 | 8521 | 13.98 | 0.01 | 0.04 |
| Asian/Pacific Islander | 2763 | 4.74 | 2770 | 4.74 | 2860 | 4.69 | 0.00 | -0.04 |
| American Indian | 595 | 1.02 | 597 | 1.02 | 646 | 1.06 | 0.00 | 0.04 |
| Other | 3156 | 5.41 | 3167 | 5.42 | 3326 | 5.46 | 0.01 | 0.05 |
| **LEP** | | | | | | | | |
| No | 53104 | 91.02 | 53198 | 91.02 | 55533 | 91.10 | 0.00 | 0.09 |
| Yes | 5241 | 8.98 | 5251 | 8.98 | 5423 | 8.90 | 0.00 | -0.09 |
| **Disability** | | | | | | | | |
| No | 50476 | 86.51 | 50553 | 86.49 | 52153 | 85.56 | -0.02 | -0.95 |
| Yes | 7869 | 13.49 | 7896 | 13.51 | 8803 | 14.44 | 0.02 | 0.95 |
| **SES Disadvantaged** | | | | | | | | |
| No | 33921 | 58.14 | 33939 | 58.07 | 35209 | 57.76 | -0.07 | -0.38 |
| Yes | 24424 | 41.86 | 24510 | 41.93 | 25747 | 42.24 | 0.07 | 0.38 |
| **Grade 4** | N | % | N | % | N | % | % | % |
| All Students | 58822 | | 58931 | | 61465 | | 99.82 | 95.70 |
| **Gender** | | | | | | | | |
| Male | 30090 | 51.15 | 30161 | 51.18 | 31558 | 51.34 | 0.03 | 0.19 |
| Female | 28732 | 48.85 | 28770 | 48.82 | 29907 | 48.66 | -0.03 | -0.19 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38385 | 65.26 | 38425 | 65.20 | 39960 | 65.01 | -0.05 | -0.24 |
| African-American | 5566 | 9.46 | 5606 | 9.51 | 6001 | 9.76 | 0.05 | 0.30 |
| Hispanic | 8416 | 14.31 | 8432 | 14.31 | 8792 | 14.30 | 0.00 | 0.00 |
| Asian/Pacific Islander | 2723 | 4.63 | 2725 | 4.62 | 2819 | 4.59 | -0.01 | -0.04 |
| American Indian | 589 | 1.00 | 591 | 1.00 | 625 | 1.02 | 0.00 | 0.02 |
| Other | 3143 | 5.34 | 3152 | 5.35 | 3268 | 5.32 | 0.01 | -0.03 |
| **LEP** | | | | | | | | |
| No | 53298 | 90.61 | 53396 | 90.61 | 55701 | 90.62 | 0.00 | 0.01 |
| Yes | 5524 | 9.39 | 5535 | 9.39 | 5764 | 9.38 | 0.00 | -0.01 |
| **Disability** | | | | | | | | |
| No | 50974 | 86.66 | 51067 | 86.66 | 52699 | 85.74 | 0.00 | -0.92 |
| Yes | 7848 | 13.34 | 7864 | 13.34 | 8766 | 14.26 | 0.00 | 0.92 |
| **SES Disadvantaged** | | | | | | | | |
| No | 34191 | 58.13 | 34226 | 58.08 | 35488 | 57.74 | -0.05 | -0.39 |
| Yes | 24631 | 41.87 | 24705 | 41.92 | 25977 | 42.26 | 0.05 | 0.39 |

**Table 6-25B. Calibration Sample Demographics Compared to Population, Mathematics, Grades 5 and 6**

| Grade 5 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 59081 | | 59187 | | 61735 | | 99.82 | 95.70 |
| **Gender** | | | | | | | | |
| Male | 30124 | 50.99 | 30171 | 50.98 | 31576 | 51.15 | -0.01 | 0.16 |
| Female | 28957 | 49.01 | 29016 | 49.02 | 30159 | 48.85 | 0.01 | -0.16 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38747 | 65.58 | 38776 | 65.51 | 40292 | 65.27 | -0.07 | -0.32 |
| African-American | 5736 | 9.71 | 5776 | 9.76 | 6170 | 9.99 | 0.05 | 0.29 |
| Hispanic | 8424 | 14.26 | 8446 | 14.27 | 8824 | 14.29 | 0.01 | 0.03 |
| Asian/Pacific Islander | 2606 | 4.41 | 2612 | 4.41 | 2686 | 4.35 | 0.00 | -0.06 |
| American Indian | 632 | 1.07 | 634 | 1.07 | 676 | 1.10 | 0.00 | 0.03 |
| Other | 2936 | 4.97 | 2943 | 4.97 | 3087 | 5.00 | 0.00 | 0.03 |
| **LEP** | | | | | | | | |
| No | 54164 | 91.68 | 54257 | 91.67 | 56611 | 91.70 | -0.01 | 0.02 |
| Yes | 4917 | 8.32 | 4930 | 8.33 | 5124 | 8.30 | 0.01 | -0.02 |
| **Disability** | | | | | | | | |
| No | 51363 | 86.94 | 51441 | 86.91 | 53057 | 85.94 | -0.02 | -0.99 |
| Yes | 7718 | 13.06 | 7746 | 13.09 | 8678 | 14.06 | 0.02 | 0.99 |
| **SES Disadvantaged** | | | | | | | | |
| No | 34812 | 58.92 | 34842 | 58.87 | 36093 | 58.46 | -0.05 | -0.46 |
| Yes | 24269 | 41.08 | 24345 | 41.13 | 25642 | 41.54 | 0.05 | 0.46 |
| **Grade 6** | N | % | N | % | N | % | % | % |
| All Students | 60059 | | 60234 | | 63048 | | 99.71 | 95.26 |
| **Gender** | | | | | | | | |
| Male | 30627 | 50.99 | 30732 | 51.02 | 32243 | 51.14 | 0.03 | 0.15 |
| Female | 29432 | 49.01 | 29502 | 48.98 | 30805 | 48.86 | -0.03 | -0.15 |
| **Race/Ethnicity** | | | | | | | | |
| White | 39817 | 66.30 | 39861 | 66.18 | 41457 | 65.75 | -0.12 | -0.54 |
| African-American | 5793 | 9.65 | 5874 | 9.75 | 6393 | 10.14 | 0.11 | 0.49 |
| Hispanic | 8518 | 14.18 | 8545 | 14.19 | 8975 | 14.24 | 0.00 | 0.05 |
| Asian/Pacific Islander | 2483 | 4.13 | 2489 | 4.13 | 2569 | 4.07 | 0.00 | -0.06 |
| American Indian | 615 | 1.02 | 618 | 1.03 | 655 | 1.04 | 0.00 | 0.01 |
| Other | 2833 | 4.72 | 2847 | 4.73 | 2999 | 4.76 | 0.01 | 0.04 |
| **LEP** | | | | | | | | |
| No | 55924 | 93.12 | 56081 | 93.11 | 58705 | 93.11 | -0.01 | 0.00 |
| Yes | 4135 | 6.88 | 4153 | 6.89 | 4343 | 6.89 | 0.01 | 0.00 |
| **Disability** | | | | | | | | |
| No | 52587 | 87.56 | 52710 | 87.51 | 54476 | 86.40 | -0.05 | -1.15 |
| Yes | 7472 | 12.44 | 7524 | 12.49 | 8572 | 13.60 | 0.05 | 1.15 |
| **SES Disadvantaged** | | | | | | | | |
| No | 35657 | 59.37 | 35702 | 59.27 | 37070 | 58.80 | -0.10 | -0.57 |
| Yes | 24402 | 40.63 | 24532 | 40.73 | 25978 | 41.20 | 0.10 | 0.57 |

**Table 6-25C. Calibration Sample Demographics Compared to Population, Mathematics, Grades 7 and 8**

| Grade 7 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) − (1) | Difference (3) − (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 61773 | | 61969 | | 65386 | | 99.68 | 94.47 |
| **Gender** | | | | | | | | |
| Male | 31654 | 51.24 | 31740 | 51.22 | 33554 | 51.32 | -0.02 | 0.07 |
| Female | 30119 | 48.76 | 30229 | 48.78 | 31832 | 48.68 | 0.02 | -0.07 |
| **Race/Ethnicity** | | | | | | | | |
| White | 40860 | 66.15 | 40900 | 66.00 | 42791 | 65.44 | -0.14 | -0.70 |
| African-American | 5962 | 9.65 | 6067 | 9.79 | 6717 | 10.27 | 0.14 | 0.62 |
| Hispanic | 8800 | 14.25 | 8832 | 14.25 | 9326 | 14.26 | 0.01 | 0.02 |
| Asian/Pacific Islander | 2685 | 4.35 | 2687 | 4.34 | 2782 | 4.25 | -0.01 | -0.09 |
| American Indian | 617 | 1.00 | 620 | 1.00 | 681 | 1.04 | 0.00 | 0.04 |
| Other | 2849 | 4.61 | 2863 | 4.62 | 3089 | 4.72 | 0.01 | 0.11 |
| **LEP** | | | | | | | | |
| No | 57759 | 93.50 | 57942 | 93.50 | 61152 | 93.52 | 0.00 | 0.02 |
| Yes | 4014 | 6.50 | 4027 | 6.50 | 4234 | 6.48 | 0.00 | -0.02 |
| **Disability** | | | | | | | | |
| No | 54237 | 87.80 | 54389 | 87.77 | 56616 | 86.59 | -0.03 | -1.21 |
| Yes | 7536 | 12.20 | 7580 | 12.23 | 8770 | 13.41 | 0.03 | 1.21 |
| **SES Disadvantaged** | | | | | | | | |
| No | 37072 | 60.01 | 37112 | 59.89 | 38738 | 59.25 | -0.13 | -0.77 |
| Yes | 24701 | 39.99 | 24857 | 40.11 | 26648 | 40.75 | 0.13 | 0.77 |
| **Grade 8** | N | % | N | % | N | % | % | % |
| All Students | 62555 | | 62762 | | 66789 | | 99.67 | 93.66 |
| **Gender** | | | | | | | | |
| Male | 32069 | 51.27 | 32172 | 51.26 | 34164 | 51.15 | 0.00 | -0.11 |
| Female | 30486 | 48.73 | 30590 | 48.74 | 32625 | 48.85 | 0.00 | 0.11 |
| **Race/Ethnicity** | | | | | | | | |
| White | 41691 | 66.65 | 41754 | 66.53 | 44041 | 65.94 | -0.12 | -0.71 |
| African-American | 6054 | 9.68 | 6139 | 9.78 | 6848 | 10.25 | 0.10 | 0.58 |
| Hispanic | 8926 | 14.27 | 8960 | 14.28 | 9588 | 14.36 | 0.01 | 0.09 |
| Asian/Pacific Islander | 2537 | 4.06 | 2539 | 4.05 | 2628 | 3.93 | -0.01 | -0.12 |
| American Indian | 573 | 0.92 | 584 | 0.93 | 647 | 0.97 | 0.01 | 0.05 |
| Other | 2773 | 4.43 | 2786 | 4.44 | 3037 | 4.55 | 0.01 | 0.11 |
| **LEP** | | | | | | | | |
| No | 58584 | 93.65 | 58777 | 93.65 | 62556 | 93.66 | 0.00 | 0.01 |
| Yes | 3970 | 6.35 | 3985 | 6.35 | 4233 | 6.34 | 0.00 | -0.01 |
| **Disability** | | | | | | | | |
| No | 55467 | 88.67 | 55616 | 88.61 | 58317 | 87.32 | -0.06 | -1.36 |
| Yes | 7087 | 11.33 | 7146 | 11.39 | 8472 | 12.68 | 0.06 | 1.36 |
| **SES Disadvantaged** | | | | | | | | |
| No | 37831 | 60.48 | 37881 | 60.36 | 39852 | 59.67 | -0.12 | -0.81 |
| Yes | 24723 | 39.52 | 24881 | 39.64 | 26937 | 40.33 | 0.12 | 0.81 |

**Table 6-26. Calibration Sample Demographics Compared to Population, Science**

| Grade 4 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 58676 | | 58880 | | 61465 | | 99.65 | 95.46 |
| **Gender** | | | | | | | | |
| Male | 30030 | 51.18 | 30129 | 51.17 | 31558 | 51.34 | -0.01 | 0.16 |
| Female | 28646 | 48.82 | 28751 | 48.83 | 29907 | 48.66 | 0.01 | -0.16 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38363 | 65.38 | 38422 | 65.25 | 39960 | 65.01 | -0.13 | -0.37 |
| African-American | 5494 | 9.36 | 5577 | 9.47 | 6001 | 9.76 | 0.11 | 0.40 |
| Hispanic | 8377 | 14.28 | 8422 | 14.30 | 8792 | 14.30 | 0.03 | 0.03 |
| Asian/Pacific Islander | 2717 | 4.63 | 2721 | 4.62 | 2819 | 4.59 | -0.01 | -0.04 |
| American Indian | 588 | 1.00 | 590 | 1.00 | 625 | 1.02 | 0.00 | 0.01 |
| Other | 3137 | 5.35 | 3148 | 5.35 | 3268 | 5.32 | 0.00 | -0.03 |
| **LEP** | | | | | | | | |
| No | 53176 | 90.63 | 53359 | 90.62 | 55701 | 90.62 | 0.00 | 0.00 |
| Yes | 5500 | 9.37 | 5521 | 9.38 | 5764 | 9.38 | 0.00 | 0.00 |
| **Disability** | | | | | | | | |
| No | 50847 | 86.66 | 51016 | 86.64 | 52699 | 85.74 | -0.01 | -0.92 |
| Yes | 7829 | 13.34 | 7864 | 13.36 | 8766 | 14.26 | 0.01 | 0.92 |
| **SES Disadvantaged** | | | | | | | | |
| No | 34145 | 58.19 | 34215 | 58.11 | 35488 | 57.74 | -0.08 | -0.46 |
| Yes | 24531 | 41.81 | 24665 | 41.89 | 25977 | 42.26 | 0.08 | 0.46 |
| **Grade 8** | N | % | N | % | N | % | % | % |
| All Students | 62287 | | 62644 | | 66789 | | 99.43 | 93.26 |
| **Gender** | | | | | | | | |
| Male | 31903 | 51.22 | 32104 | 51.25 | 34164 | 51.15 | 0.03 | -0.07 |
| Female | 30384 | 48.78 | 30540 | 48.75 | 32625 | 48.85 | -0.03 | 0.07 |
| **Race/Ethnicity** | | | | | | | | |
| White | 41628 | 66.83 | 41734 | 66.62 | 44041 | 65.94 | -0.21 | -0.89 |
| African-American | 5928 | 9.52 | 6068 | 9.69 | 6848 | 10.25 | 0.17 | 0.74 |
| Hispanic | 8866 | 14.23 | 8942 | 14.27 | 9588 | 14.36 | 0.04 | 0.12 |
| Asian/Pacific Islander | 2534 | 4.07 | 2540 | 4.05 | 2628 | 3.93 | -0.01 | -0.13 |
| American Indian | 570 | 0.92 | 580 | 0.93 | 647 | 0.97 | 0.01 | 0.05 |
| Other | 2761 | 4.43 | 2780 | 4.44 | 3037 | 4.55 | 0.01 | 0.11 |
| **LEP** | | | | | | | | |
| No | 58344 | 93.67 | 58668 | 93.65 | 62556 | 93.66 | -0.02 | -0.01 |
| Yes | 3943 | 6.33 | 3976 | 6.35 | 4233 | 6.34 | 0.02 | 0.01 |
| **Disability** | | | | | | | | |
| No | 55249 | 88.70 | 55524 | 88.63 | 58317 | 87.32 | -0.07 | -1.39 |
| Yes | 7038 | 11.30 | 7120 | 11.37 | 8472 | 12.68 | 0.07 | 1.39 |
| **SES Disadvantaged** | | | | | | | | |
| No | 37749 | 60.60 | 37845 | 60.41 | 39852 | 59.67 | -0.19 | -0.94 |
| Yes | 24538 | 39.40 | 24799 | 39.59 | 26937 | 40.33 | 0.19 | 0.94 |

**Table 6-27A. Calibration Sample Demographics Compared to Population, Social Studies, Grades 4 and 8**

| Grade 4 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 58715 | | 58833 | | 61465 | | 99.80 | 95.53 |
| **Gender** | | | | | | | | |
| Male | 30037 | 51.16 | 30104 | 51.17 | 31558 | 51.34 | 0.01 | 0.19 |
| Female | 28678 | 48.84 | 28729 | 48.83 | 29907 | 48.66 | -0.01 | -0.19 |
| **Race/Ethnicity** | | | | | | | | |
| White | 38372 | 65.35 | 38403 | 65.27 | 39960 | 65.01 | -0.08 | -0.34 |
| African-American | 5518 | 9.40 | 5566 | 9.46 | 6001 | 9.76 | 0.06 | 0.37 |
| Hispanic | 8375 | 14.26 | 8407 | 14.29 | 8792 | 14.30 | 0.03 | 0.04 |
| Asian/Pacific Islander | 2721 | 4.63 | 2721 | 4.62 | 2819 | 4.59 | -0.01 | -0.05 |
| American Indian | 592 | 1.01 | 593 | 1.01 | 625 | 1.02 | 0.00 | 0.01 |
| Other | 3137 | 5.34 | 3143 | 5.34 | 3268 | 5.32 | 0.00 | -0.03 |
| **LEP** | | | | | | | | |
| No | 53211 | 90.63 | 53317 | 90.62 | 55701 | 90.62 | 0.00 | 0.00 |
| Yes | 5504 | 9.37 | 5516 | 9.38 | 5764 | 9.38 | 0.00 | 0.00 |
| **Disability** | | | | | | | | |
| No | 50895 | 86.68 | 50983 | 86.66 | 52699 | 85.74 | -0.02 | -0.94 |
| Yes | 7820 | 13.32 | 7850 | 13.34 | 8766 | 14.26 | 0.02 | 0.94 |
| **SES Disadvantaged** | | | | | | | | |
| No | 34162 | 58.18 | 34201 | 58.13 | 35488 | 57.74 | -0.05 | -0.45 |
| Yes | 24553 | 41.82 | 24632 | 41.87 | 25977 | 42.26 | 0.05 | 0.45 |
| **Grade 8** | N | % | N | % | N | % | % | % |
| All Students | 62348 | | 62606 | | 66789 | | 99.59 | 93.35 |
| **Gender** | | | | | | | | |
| Male | 31942 | 51.23 | 32088 | 51.25 | 34164 | 51.15 | 0.02 | -0.08 |
| Female | 30406 | 48.77 | 30518 | 48.75 | 32625 | 48.85 | -0.02 | 0.08 |
| **Race/Ethnicity** | | | | | | | | |
| White | 41680 | 66.85 | 41752 | 66.69 | 44041 | 65.94 | -0.16 | -0.91 |
| African-American | 5925 | 9.50 | 6027 | 9.63 | 6848 | 10.25 | 0.12 | 0.75 |
| Hispanic | 8866 | 14.22 | 8930 | 14.26 | 9588 | 14.36 | 0.04 | 0.14 |
| Asian/Pacific Islander | 2536 | 4.07 | 2539 | 4.06 | 2628 | 3.93 | -0.01 | -0.13 |
| American Indian | 569 | 0.91 | 577 | 0.92 | 647 | 0.97 | 0.01 | 0.06 |
| Other | 2772 | 4.45 | 2781 | 4.44 | 3037 | 4.55 | 0.00 | 0.10 |
| **LEP** | | | | | | | | |
| No | 58401 | 93.67 | 58629 | 93.65 | 62556 | 93.66 | -0.02 | -0.01 |
| Yes | 3947 | 6.33 | 3977 | 6.35 | 4233 | 6.34 | 0.02 | 0.01 |
| **Disability** | | | | | | | | |
| No | 55296 | 88.69 | 55493 | 88.64 | 58317 | 87.32 | -0.05 | -1.37 |
| Yes | 7052 | 11.31 | 7113 | 11.36 | 8472 | 12.68 | 0.05 | 1.37 |
| **SES Disadvantaged** | | | | | | | | |
| No | 37783 | 60.60 | 37851 | 60.46 | 39852 | 59.67 | -0.14 | -0.93 |
| Yes | 24565 | 39.40 | 24755 | 39.54 | 26937 | 40.33 | 0.14 | 0.93 |

**Table 6-27B. Calibration Sample Demographics Compared to Population, Social Studies, Grade 10**

| Grade 10 | Calibration Sample (1) | | Students with Valid Scores (2) | | Scheduled to Test (3) | | Difference (2) – (1) | Difference (3) – (1) |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | % | % |
| All Students | 58120 | | 59391 | | 67427 | | 97.86 | 86.20 |
| **Gender** | | | | | | | | |
| Male | 29891 | 51.43 | 30546 | 51.43 | 34566 | 51.26 | 0.00 | -0.17 |
| Female | 28229 | 48.57 | 28845 | 48.57 | 32861 | 48.74 | 0.00 | 0.17 |
| **Race/Ethnicity** | | | | | | | | |
| White | 41541 | 71.47 | 41976 | 70.68 | 45824 | 67.96 | -0.80 | -3.51 |
| African-American | 3764 | 6.48 | 4203 | 7.08 | 6174 | 9.16 | 0.60 | 2.68 |
| Hispanic | 7663 | 13.18 | 7932 | 13.36 | 9344 | 13.86 | 0.17 | 0.67 |
| Asian/Pacific Islander | 2412 | 4.15 | 2458 | 4.14 | 2641 | 3.92 | -0.01 | -0.23 |
| American Indian | 555 | 0.95 | 570 | 0.96 | 740 | 1.10 | 0.00 | 0.14 |
| Other | 2185 | 3.76 | 2252 | 3.79 | 2704 | 4.01 | 0.03 | 0.25 |
| **LEP** | | | | | | | | |
| No | 55425 | 95.36 | 56567 | 95.25 | 64034 | 94.97 | -0.12 | -0.40 |
| Yes | 2695 | 4.64 | 2824 | 4.75 | 3393 | 5.03 | 0.12 | 0.40 |
| **Disability** | | | | | | | | |
| No | 52177 | 89.77 | 53228 | 89.62 | 59075 | 87.61 | -0.15 | -2.16 |
| Yes | 5943 | 10.23 | 6163 | 10.38 | 8352 | 12.39 | 0.15 | 2.16 |
| **SES Disadvantaged** | | | | | | | | |
| No | 38578 | 66.38 | 39119 | 65.87 | 42633 | 63.23 | -0.51 | -3.15 |
| Yes | 19542 | 33.62 | 20272 | 34.13 | 24794 | 36.77 | 0.51 | 3.15 |

**Table 6-28 Items Flagged Based on Yen's Q1**

| Content | Grade | Item Number in Calibration | Type | N | Z | Critical Z |
|---|---|---|---|---|---|---|
| English Language Arts | 3 | 14 | CR | 57860 | 201.16 | 154.29 |
| | 3 | 28 | MC | 57860 | 159.19 | 154.29 |
| | 3 | 29 | CR | 57860 | 216.88 | 154.29 |
| | 3 | 35 | CR | 57860 | 285.67 | 154.29 |
| | 4 | 20 | CR | 58295 | 305.89 | 155.45 |
| | 4 | 22* | CR | 58295 | 157.53 | 155.45 |
| | 4 | 28 | CR | 58295 | 447.94 | 155.45 |
| | 4 | 30 | MC | 58295 | 282.29 | 155.45 |
| | 4 | 32 | CR | 58295 | 307.77 | 155.45 |
| | 4 | 38 | CR | 58295 | 197.32 | 155.45 |
| | 5 | 16 | CR | 58840 | 255.03 | 156.91 |
| | 5 | 22 | CR | 58840 | 200.85 | 156.91 |
| | 6 | 17 | MC | 59589 | 213.27 | 158.90 |
| | 6 | 19 | CR | 59589 | 622.63 | 158.90 |
| | 7 | 14* | CR | 61538 | 174.41 | 164.10 |
| | 7 | 18 | CR | 61538 | 472.83 | 164.10 |
| | 7 | 23 | CR | 61538 | 818.63 | 164.10 |
| | 7 | 25 | CR | 61538 | 165.37 | 164.10 |
| | 7 | 27 | CR | 61538 | 329.66 | 164.10 |
| | 7 | 31 | CR | 61538 | 183.4 | 164.10 |
| | 8 | 19* | CR | 62310 | 200.31 | 166.16 |
| | 8 | 33 | CR | 62310 | 177.82 | 166.16 |
| | 8 | 38 | CR | 62310 | 800.16 | 166.16 |
| Mathematics | 3 | 41 | CR | 58096 | 193.78 | 154.92 |
| | 4 | 17 | MC | 58758 | 268.38 | 156.69 |
| | 6 | 40* | MC | 59944 | 186.38 | 159.85 |
| | 6 | 45 | CR | 59944 | 295.67 | 159.85 |
| | 7 | 42* | CR | 61715 | 353.07 | 164.57 |
| | 8 | 26* | MC | 62487 | 258.78 | 166.63 |
| Science | 4 | 21* | CR | 58661 | 196.37 | 156.43 |
| | 8 | 4 | CR | 62243 | 208.65 | 165.98 |
| | 8 | 27* | CR | 62243 | 224.86 | 165.98 |
| | 8 | 39 | CR | 62243 | 379.09 | 165.98 |
| Social Studies | 4 | 3 | CR | 58612 | 161 | 156.30 |
| | 8 | 32 | CR | 62144 | 225.54 | 165.72 |
| | 8 | 40 | CR | 62144 | 188.37 | 165.72 |
| | 10 | 10 | MC | 58053 | 181.45 | 154.81 |
| | 10 | 34 | CR | 58053 | 166.13 | 154.81 |

Note: An asterisk (*) indicates an anchor item.

**Table 6-29. Equating Evaluation Results, Stocking and Lord Method**

| Content Area | Grade | Number of Anchors | TCC Results | | a-Parameter | | b-Parameter | | Equating Constants | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | # of Iterations | F Value | Corr | # of RMSD Outliers | Corr | # of RMSD Outliers | A | B |
| English Language Arts | 3 | 15 | 5 | 0.11633 | 0.98 | 1 | 0.98 | 1 | 0.956 | -1.325 |
| | 4 | 17 | 3 | 0.12119 | 0.98 | 1 | 0.99 | 1 | 1.061 | -0.669 |
| | 5 | 14 | 5 | 0.18427 | 0.98 | 0 | 0.99 | 0 | 1.075 | -0.374 |
| | 6 | 16 | 4 | 0.09229 | 0.96 | 0 | 0.99 | 1 | 1.040 | -0.126 |
| | 7 | 15 | 5 | 0.14713 | 0.99 | 0 | 0.97 | 0 | 1.145 | 0.231 |
| | 8 | 18 | 3 | 0.04797 | 0.97 | 2 | 0.98 | 1 | 1.237 | 0.333 |
| Mathematics | 3 | 18 | 3 | 0.02910 | 0.97 | 1 | 0.98 | 1 | 1.027 | -1.263 |
| | 4 | 19 | 9 | 0.13690 | 0.97 | 0 | 0.99 | 1 | 1.028 | -0.760 |
| | 5 | 20 | 13 | 0.20649 | 0.91 | 1 | 0.99 | 1 | 0.932 | -0.204 |
| | 6 | 22 | 14 | 0.07918 | 0.99 | 0 | 0.99 | 1 | 1.051 | -0.022 |
| | 7 | 26 | 25 | 0.06275 | 0.94 | 2 | 0.98 | 1 | 1.095 | 0.210 |
| | 8 | 24 | 22 | 0.04953 | 0.97 | 1 | 0.97 | 1 | 1.047 | 0.626 |
| Science | 4 | 15 | 9 | 0.16064 | 0.99 | 0 | 1.00 | 0 | 1.028 | -0.032 |
| | 8 | 15 | 7 | 0.07286 | 1.00 | 0 | 1.00 | 0 | 0.984 | -0.123 |
| | 8* | 14 | 5 | 0.06049 | 1.00 | 0 | 1.00 | 0 | 0.999 | -0.111 |

*Equating results obtained in test equating with a reduced anchor set (final)

**Table 6-30. Statistics Comparing IRT Item-Ability Regression Curves for Flagged Anchor Items**

| Content Area | Grade | Anchor Item Position | UnWtd RMSD | UnWtd Mean Abs | Max Abs | UnWtd Mean | Wtd RMSD | Wtd Mean Abs | Wtd Mean |
|---|---|---|---|---|---|---|---|---|---|
| Mathematics | 8 | 44 | 0.0631 | 0.0410 | **0.1442** | 0.0410 | **0.0937** | **0.0798** | **0.0798** |

Note: Item-Ability Regression statistics meeting the flagging criteria are indicated in bold print.

**Table 6-31. Scale Transformation Constants**

| Content Area | Grade | Scale Transformation Constants | |
|---|---|---|---|
| | | M1 | M2 |
| ELA | 3–8 | 43.7445 | 610.4987 |
| Mathematics | 3–8 | 46.4684 | 612.0818 |
| Science | 4 | 45.0450 | 500.4505 |
| | 8 | 45.0450 | 699.5496 |
| Social Studies | 4 | 42.9923 | 503.8693 |
| | 8 | 40.2577 | 708.0515 |
| | 10 | 42.0521 | 803.7847 |

## Table 6-32. Scoring Table for English Language Arts Grade 3

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 330 | 85 |
| 1 | 330 | 85 |
| 2 | 330 | 85 |
| 3 | 330 | 85 |
| 4 | 330 | 85 |
| 5 | 330 | 85 |
| 6 | 379 | 49 |
| 7 | 410 | 36 |
| 8 | 430 | 29 |
| 9 | 444 | 26 |
| 10 | 456 | 23 |
| 11 | 466 | 21 |
| 12 | 474 | 20 |
| 13 | 482 | 18 |
| 14 | 489 | 17 |
| 15 | 495 | 16 |
| 16 | 501 | 15 |
| 17 | 506 | 15 |
| 18 | 511 | 14 |
| 19 | 516 | 14 |
| 20 | 520 | 13 |
| **21** | **524** | **13** |
| 22 | 529 | 13 |
| 23 | 533 | 13 |
| 24 | 537 | 13 |
| 25 | 541 | 13 |
| 26 | 545 | 13 |
| 27 | 549 | 13 |
| 28 | 553 | 13 |
| 29 | 557 | 13 |
| 30 | 561 | 13 |
| 31 | 565 | 13 |
| 32 | 569 | 13 |
| **33** | **574** | **13** |
| 34 | 578 | 13 |
| 35 | 583 | 13 |
| 36 | 587 | 14 |
| 37 | 592 | 14 |
| 38 | 597 | 14 |
| 39 | 603 | 15 |
| 40 | 609 | 15 |
| 41 | 615 | 16 |
| 42 | 621 | 16 |
| **43** | **629** | **17** |

**Table 6-32 (continued). Scoring Table for English Language Arts Grade 3**

| Raw Score | Scale Score | SEM |
|-----------|-------------|-----|
| 44 | 637 | 19 |
| 45 | 647 | 20 |
| 46 | 657 | 22 |
| 47 | 670 | 24 |
| 48 | 685 | 26 |
| 49 | 702 | 28 |
| 50 | 723 | 31 |
| 51 | 749 | 35 |
| 52 | 787 | 46 |
| 53 | 900 | 138 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-33. Scoring Table for English Language Arts Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 340 | 61 |
| 1 | 340 | 61 |
| 2 | 340 | 61 |
| 3 | 340 | 61 |
| 4 | 340 | 61 |
| 5 | 340 | 61 |
| 6 | 370 | 50 |
| 7 | 404 | 41 |
| 8 | 428 | 36 |
| 9 | 447 | 32 |
| 10 | 461 | 28 |
| 11 | 474 | 25 |
| 12 | 484 | 23 |
| 13 | 493 | 21 |
| 14 | 501 | 19 |
| 15 | 509 | 18 |
| 16 | 515 | 17 |
| 17 | 522 | 16 |
| 18 | 527 | 16 |
| 19 | 533 | 16 |
| 20 | 538 | 15 |
| 21 | 543 | 15 |
| **22** | **548** | **15** |
| 23 | 553 | 15 |
| 24 | 558 | 14 |
| 25 | 562 | 14 |
| 26 | 567 | 14 |
| 27 | 571 | 14 |
| 28 | 575 | 14 |
| 29 | 580 | 14 |
| 30 | 584 | 13 |
| 31 | 588 | 13 |
| **32** | **593** | **13** |
| 33 | 597 | 14 |
| 34 | 601 | 14 |
| 35 | 606 | 14 |
| 36 | 610 | 14 |
| 37 | 615 | 14 |
| 38 | 620 | 15 |
| 39 | 625 | 15 |
| 40 | 631 | 16 |
| 41 | 636 | 16 |
| 42 | 642 | 17 |
| 43 | 648 | 17 |

**Table 6-33 (continued). Scoring Table for English Language Arts Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| **44** | **655** | **18** |
| 45 | 662 | 19 |
| 46 | 670 | 20 |
| 47 | 679 | 21 |
| 48 | 689 | 22 |
| 49 | 699 | 24 |
| 50 | 712 | 26 |
| 51 | 727 | 29 |
| 52 | 746 | 33 |
| 53 | 769 | 37 |
| 54 | 801 | 43 |
| 55 | 848 | 56 |
| 56 | 930 | 109 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-34. Scoring Table for English Language Arts Grade 5**

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 0 | 350 | 63 |
| 1 | 350 | 63 |
| 2 | 350 | 63 |
| 3 | 350 | 63 |
| 4 | 350 | 63 |
| 5 | 350 | 63 |
| 6 | 350 | 63 |
| 7 | 359 | 59 |
| 8 | 402 | 43 |
| 9 | 429 | 36 |
| 10 | 449 | 32 |
| 11 | 466 | 29 |
| 12 | 479 | 26 |
| 13 | 491 | 24 |
| 14 | 501 | 22 |
| 15 | 510 | 20 |
| 16 | 518 | 19 |
| 17 | 525 | 18 |
| 18 | 532 | 18 |
| 19 | 538 | 17 |
| 20 | 544 | 17 |
| 21 | 550 | 16 |
| 22 | 556 | 16 |
| 23 | 561 | 16 |
| **24** | **566** | **15** |
| 25 | 571 | 15 |
| 26 | 576 | 15 |
| 27 | 581 | 14 |
| 28 | 585 | 14 |
| 29 | 590 | 14 |
| 30 | 594 | 14 |
| 31 | 598 | 13 |
| 32 | 602 | 13 |
| 33 | 607 | 13 |
| **34** | **611** | **13** |
| 35 | 615 | 13 |
| 36 | 619 | 12 |
| 37 | 623 | 13 |
| 38 | 628 | 13 |
| 39 | 632 | 13 |
| 40 | 637 | 13 |
| 41 | 642 | 13 |
| 42 | 647 | 14 |
| 43 | 652 | 14 |

**Table 6-34 (continued). Scoring Table for English Language Arts Grade 5**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 658 | 15 |
| 45 | 664 | 15 |
| **46** | **670** | **16** |
| 47 | 678 | 17 |
| 48 | 686 | 18 |
| 49 | 695 | 20 |
| 50 | 706 | 22 |
| 51 | 720 | 25 |
| 52 | 737 | 29 |
| 53 | 761 | 35 |
| 54 | 794 | 41 |
| 55 | 841 | 50 |
| 56 | 940 | 120 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-35. Scoring Table for English Language Arts Grade 6**

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 0 | 360 | 65 |
| 1 | 360 | 65 |
| 2 | 360 | 65 |
| 3 | 360 | 65 |
| 4 | 360 | 65 |
| 5 | 360 | 65 |
| 6 | 381 | 54 |
| 7 | 420 | 39 |
| 8 | 445 | 33 |
| 9 | 463 | 30 |
| 10 | 478 | 27 |
| 11 | 491 | 26 |
| 12 | 502 | 24 |
| 13 | 512 | 22 |
| 14 | 521 | 21 |
| 15 | 528 | 20 |
| 16 | 535 | 19 |
| 17 | 542 | 17 |
| 18 | 547 | 17 |
| 19 | 553 | 16 |
| 20 | 558 | 15 |
| 21 | 563 | 14 |
| 22 | 567 | 14 |
| **23** | **572** | **14** |
| 24 | 576 | 13 |
| 25 | 580 | 13 |
| 26 | 584 | 13 |
| 27 | 588 | 13 |
| 28 | 592 | 13 |
| 29 | 596 | 13 |
| 30 | 600 | 13 |
| 31 | 604 | 13 |
| 32 | 608 | 13 |
| 33 | 612 | 13 |
| 34 | 616 | 13 |
| 35 | 621 | 13 |
| **36** | **625** | **13** |
| 37 | 629 | 13 |
| 38 | 634 | 13 |
| 39 | 638 | 14 |
| 40 | 643 | 14 |
| 41 | 648 | 14 |
| 42 | 654 | 14 |
| 43 | 659 | 15 |

**Table 6-35 (continued). Scoring Table for English Language Arts Grade 6**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 665 | 15 |
| **45** | **671** | **16** |
| 46 | 678 | 16 |
| 47 | 685 | 17 |
| 48 | 693 | 18 |
| 49 | 702 | 19 |
| 50 | 712 | 21 |
| 51 | 723 | 23 |
| 52 | 737 | 25 |
| 53 | 754 | 29 |
| 54 | 778 | 36 |
| 55 | 820 | 55 |
| 56 | 950 | 170 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

# Table 6-36. Scoring Table for English Language Arts Grade 7

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 370 | 63 |
| 1 | 370 | 63 |
| 2 | 370 | 63 |
| 3 | 370 | 63 |
| 4 | 370 | 63 |
| 5 | 397 | 51 |
| 6 | 430 | 40 |
| 7 | 452 | 35 |
| 8 | 470 | 31 |
| 9 | 484 | 28 |
| 10 | 497 | 26 |
| 11 | 508 | 25 |
| 12 | 517 | 23 |
| 13 | 526 | 22 |
| 14 | 534 | 21 |
| 15 | 541 | 20 |
| 16 | 548 | 19 |
| 17 | 554 | 19 |
| 18 | 560 | 18 |
| 19 | 566 | 18 |
| 20 | 571 | 17 |
| 21 | 577 | 17 |
| 22 | 582 | 16 |
| **23** | **586** | **16** |
| 24 | 591 | 16 |
| 25 | 596 | 15 |
| 26 | 600 | 15 |
| 27 | 605 | 15 |
| 28 | 609 | 15 |
| 29 | 614 | 15 |
| 30 | 618 | 15 |
| 31 | 622 | 15 |
| 32 | 627 | 15 |
| 33 | 631 | 15 |
| 34 | 636 | 15 |
| **35** | **640** | **15** |
| 36 | 645 | 15 |
| 37 | 650 | 15 |
| 38 | 655 | 15 |
| 39 | 660 | 16 |
| 40 | 666 | 16 |
| 41 | 671 | 16 |
| 42 | 677 | 17 |
| 43 | 683 | 17 |

**Table 6-36 (continued). Scoring Table for English Language Arts Grade 7**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 690 | 18 |
| **45** | **697** | **18** |
| 46 | 704 | 19 |
| 47 | 713 | 20 |
| 48 | 722 | 21 |
| 49 | 732 | 23 |
| 50 | 744 | 25 |
| 51 | 758 | 28 |
| 52 | 775 | 33 |
| 53 | 799 | 40 |
| 54 | 833 | 52 |
| 55 | 899 | 86 |
| 56 | 960 | 128 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-37. Scoring Table for English Language Arts Grade 8**

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 0 | 380 | 60 |
| 1 | 380 | 60 |
| 2 | 380 | 60 |
| 3 | 380 | 60 |
| 4 | 380 | 60 |
| 5 | 380 | 60 |
| 6 | 380 | 60 |
| 7 | 380 | 60 |
| 8 | 419 | 47 |
| 9 | 446 | 40 |
| 10 | 467 | 36 |
| 11 | 485 | 32 |
| 12 | 499 | 30 |
| 13 | 512 | 28 |
| 14 | 523 | 26 |
| 15 | 533 | 25 |
| 16 | 542 | 23 |
| 17 | 550 | 22 |
| 18 | 558 | 21 |
| 19 | 565 | 20 |
| 20 | 571 | 19 |
| 21 | 577 | 18 |
| 22 | 583 | 17 |
| 23 | 589 | 17 |
| **24** | **594** | **16** |
| 25 | 599 | 16 |
| 26 | 604 | 16 |
| 27 | 609 | 15 |
| 28 | 614 | 15 |
| 29 | 618 | 15 |
| 30 | 623 | 15 |
| 31 | 628 | 15 |
| 32 | 633 | 15 |
| 33 | 637 | 15 |
| 34 | 642 | 15 |
| 35 | 647 | 15 |
| **36** | **652** | **15** |
| 37 | 657 | 16 |
| 38 | 662 | 16 |
| 39 | 668 | 16 |
| 40 | 673 | 16 |
| 41 | 679 | 17 |
| 42 | 685 | 17 |
| 43 | 691 | 18 |

**Table 6-37 (continued). Scoring Table for English Language Arts Grade 8**

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 44 | 698 | 18 |
| 45 | 706 | 19 |
| **46** | **714** | **20** |
| 47 | 723 | 22 |
| 48 | 732 | 23 |
| 49 | 744 | 26 |
| 50 | 757 | 28 |
| 51 | 772 | 32 |
| 52 | 792 | 37 |
| 53 | 818 | 45 |
| 54 | 857 | 60 |
| 55 | 933 | 103 |
| 56 | 970 | 131 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

## Table 6-38. Scoring Table for Mathematics Grade 3

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 0 | 360 | 101 |
| 1 | 360 | 101 |
| 2 | 360 | 101 |
| 3 | 360 | 101 |
| 4 | 361 | 100 |
| 5 | 433 | 36 |
| 6 | 456 | 26 |
| 7 | 471 | 21 |
| 8 | 481 | 18 |
| 9 | 490 | 16 |
| 10 | 497 | 15 |
| 11 | 504 | 14 |
| 12 | 509 | 13 |
| 13 | 515 | 12 |
| **14** | **519** | **12** |
| 15 | 524 | 12 |
| 16 | 528 | 11 |
| 17 | 532 | 11 |
| 18 | 536 | 11 |
| 19 | 540 | 11 |
| 20 | 544 | 11 |
| 21 | 548 | 11 |
| 22 | 552 | 11 |
| 23 | 555 | 11 |
| 24 | 559 | 11 |
| **25** | **563** | **11** |
| 26 | 567 | 11 |
| 27 | 570 | 11 |
| 28 | 574 | 11 |
| 29 | 578 | 11 |
| 30 | 583 | 11 |
| 31 | 587 | 11 |
| 32 | 591 | 12 |
| 33 | 596 | 12 |
| 34 | 601 | 12 |
| 35 | 607 | 13 |
| **36** | **614** | **14** |
| 37 | 621 | 15 |
| 38 | 629 | 17 |
| 39 | 640 | 19 |
| 40 | 656 | 24 |
| 41 | 682 | 36 |
| 42 | 760 | 102 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-39. Scoring Table for Mathematics Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 405 | 96 |
| 1 | 405 | 96 |
| 2 | 405 | 96 |
| 3 | 405 | 96 |
| 4 | 405 | 96 |
| 5 | 405 | 96 |
| 6 | 405 | 96 |
| 7 | 456 | 49 |
| 8 | 485 | 32 |
| 9 | 501 | 25 |
| 10 | 513 | 21 |
| 11 | 523 | 19 |
| 12 | 530 | 17 |
| **13** | **537** | **16** |
| 14 | 543 | 15 |
| 15 | 549 | 14 |
| 16 | 554 | 13 |
| 17 | 559 | 12 |
| 18 | 563 | 12 |
| 19 | 567 | 12 |
| 20 | 571 | 11 |
| 21 | 575 | 11 |
| 22 | 579 | 11 |
| 23 | 582 | 10 |
| 24 | 586 | 10 |
| **25** | **590** | **10** |
| 26 | 593 | 10 |
| 27 | 597 | 10 |
| 28 | 600 | 10 |
| 29 | 604 | 10 |
| 30 | 607 | 10 |
| 31 | 611 | 10 |
| 32 | 614 | 10 |
| 33 | 618 | 10 |
| 34 | 622 | 10 |
| 35 | 626 | 11 |
| 36 | 631 | 11 |
| **37** | **635** | **11** |
| 38 | 640 | 12 |
| 39 | 646 | 12 |
| 40 | 652 | 13 |
| 41 | 659 | 13 |
| 42 | 667 | 14 |
| 43 | 677 | 17 |

**Table 6-39 (continued). Scoring Table for Scoring Table for Mathematics Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 692 | 23 |
| 45 | 719 | 37 |
| 46 | 800 | 106 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-40. Scoring Table for Mathematics Grade 5**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 430 | 108 |
| 1 | 430 | 108 |
| 2 | 430 | 108 |
| 3 | 430 | 108 |
| 4 | 430 | 108 |
| 5 | 430 | 108 |
| 6 | 498 | 42 |
| 7 | 522 | 27 |
| 8 | 537 | 21 |
| 9 | 547 | 18 |
| 10 | 556 | 16 |
| 11 | 563 | 15 |
| 12 | 569 | 14 |
| **13** | **575** | **13** |
| 14 | 580 | 12 |
| 15 | 585 | 12 |
| 16 | 589 | 11 |
| 17 | 593 | 11 |
| 18 | 597 | 11 |
| 19 | 601 | 10 |
| 20 | 605 | 10 |
| 21 | 608 | 10 |
| **22** | **612** | **10** |
| 23 | 615 | 10 |
| 24 | 618 | 9 |
| 25 | 622 | 9 |
| 26 | 625 | 9 |
| 27 | 628 | 9 |
| 28 | 631 | 9 |
| 29 | 635 | 9 |
| 30 | 638 | 9 |
| 31 | 641 | 9 |
| 32 | 645 | 9 |
| 33 | 648 | 10 |
| 34 | 652 | 10 |
| 35 | 656 | 10 |
| **36** | **660** | **10** |
| 37 | 664 | 11 |
| 38 | 669 | 11 |
| 39 | 674 | 12 |
| 40 | 680 | 13 |
| 41 | 686 | 13 |
| 42 | 694 | 15 |
| 43 | 704 | 17 |

**Table 6-40 (continued). Scoring Table for Scoring Table for Mathematics Grade 5**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 717 | 21 |
| 45 | 740 | 30 |
| 46 | 830 | 110 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-41. Scoring Table for Mathematics Grade 6**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 440 | 96 |
| 1 | 440 | 96 |
| 2 | 440 | 96 |
| 3 | 440 | 96 |
| 4 | 440 | 96 |
| 5 | 440 | 96 |
| 6 | 495 | 46 |
| 7 | 523 | 32 |
| 8 | 540 | 26 |
| 9 | 553 | 22 |
| 10 | 563 | 20 |
| 11 | 571 | 18 |
| 12 | 578 | 16 |
| **13** | **585** | **15** |
| 14 | 590 | 14 |
| 15 | 595 | 13 |
| 16 | 600 | 13 |
| 17 | 605 | 12 |
| 18 | 609 | 12 |
| 19 | 613 | 11 |
| 20 | 617 | 11 |
| 21 | 621 | 11 |
| 22 | 624 | 10 |
| **23** | **628** | **10** |
| 24 | 631 | 10 |
| 25 | 635 | 10 |
| 26 | 638 | 10 |
| 27 | 641 | 10 |
| 28 | 644 | 9 |
| 29 | 648 | 9 |
| 30 | 651 | 9 |
| 31 | 654 | 9 |
| 32 | 658 | 9 |
| 33 | 661 | 10 |
| 34 | 665 | 10 |
| 35 | 668 | 10 |
| 36 | 672 | 10 |
| 37 | 677 | 11 |
| 38 | 681 | 11 |
| 39 | 686 | 12 |
| **40** | **692** | **13** |
| 41 | 699 | 14 |
| 42 | 707 | 16 |
| 43 | 718 | 19 |

**Table 6-41 (continued). Scoring Table for Scoring Table for Mathematics Grade 6**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 733 | 24 |
| 45 | 763 | 40 |
| 46 | 870 | 139 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-42. Scoring Table for Mathematics Grade 7**

| Raw Score | Scale Score | SEM |
|:---:|:---:|:---:|
| 0 | 450 | 105 |
| 1 | 450 | 105 |
| 2 | 450 | 105 |
| 3 | 450 | 105 |
| 4 | 450 | 105 |
| 5 | 450 | 105 |
| 6 | 450 | 105 |
| 7 | 506 | 59 |
| 8 | 542 | 35 |
| 9 | 562 | 26 |
| 10 | 575 | 21 |
| 11 | 585 | 18 |
| 12 | 593 | 16 |
| 13 | 600 | 15 |
| **14** | **606** | **14** |
| 15 | 611 | 13 |
| 16 | 616 | 13 |
| 17 | 621 | 12 |
| 18 | 625 | 12 |
| 19 | 630 | 11 |
| 20 | 634 | 11 |
| 21 | 637 | 11 |
| 22 | 641 | 11 |
| 23 | 645 | 10 |
| **24** | **649** | **10** |
| 25 | 652 | 10 |
| 26 | 656 | 10 |
| 27 | 659 | 10 |
| 28 | 663 | 10 |
| 29 | 667 | 10 |
| 30 | 670 | 10 |
| 31 | 674 | 11 |
| 32 | 678 | 11 |
| 33 | 682 | 11 |
| 34 | 687 | 11 |
| 35 | 691 | 12 |
| 36 | 696 | 12 |
| 37 | 701 | 12 |
| 38 | 706 | 13 |
| **39** | **713** | **14** |
| 40 | 719 | 15 |
| 41 | 727 | 16 |
| 42 | 737 | 18 |
| 43 | 749 | 21 |

**Table 6-42 (continued). Scoring Table for Scoring Table for Mathematics Grade 7**

| Raw Score | Scale Score | SEM |
|-----------|-------------|-----|
| 44 | 767 | 27 |
| 45 | 798 | 41 |
| 46 | 880 | 111 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-43. Scoring Table for Mathematics Grade 8**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 470 | 115 |
| 1 | 470 | 115 |
| 2 | 470 | 115 |
| 3 | 470 | 115 |
| 4 | 470 | 115 |
| 5 | 470 | 115 |
| 6 | 470 | 115 |
| 7 | 518 | 67 |
| 8 | 555 | 38 |
| 9 | 575 | 28 |
| 10 | 589 | 24 |
| 11 | 599 | 21 |
| 12 | 608 | 18 |
| 13 | 616 | 17 |
| **14** | **622** | **15** |
| 15 | 628 | 14 |
| 16 | 633 | 13 |
| 17 | 638 | 12 |
| 18 | 643 | 12 |
| 19 | 647 | 11 |
| 20 | 651 | 11 |
| 21 | 655 | 10 |
| 22 | 658 | 10 |
| 23 | 662 | 10 |
| 24 | 665 | 10 |
| **25** | **669** | **10** |
| 26 | 672 | 10 |
| 27 | 675 | 10 |
| 28 | 679 | 10 |
| 29 | 682 | 10 |
| 30 | 686 | 10 |
| 31 | 689 | 10 |
| 32 | 693 | 10 |
| 33 | 697 | 10 |
| 34 | 701 | 10 |
| 35 | 705 | 11 |
| 36 | 709 | 11 |
| 37 | 714 | 12 |
| **38** | **719** | **12** |
| 39 | 724 | 13 |
| 40 | 731 | 14 |
| 41 | 738 | 15 |
| 42 | 747 | 17 |
| 43 | 758 | 20 |

**Table 6-43 (continued). Scoring Table for Scoring Table for Mathematics Grade 8**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 44 | 774 | 25 |
| 45 | 804 | 40 |
| 46 | 890 | 116 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-44. Scoring Table for Science Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 300 | 62 |
| 1 | 300 | 62 |
| 2 | 300 | 62 |
| 3 | 312 | 54 |
| 4 | 349 | 36 |
| 5 | 371 | 28 |
| 6 | 386 | 25 |
| 7 | 399 | 22 |
| 8 | 409 | 21 |
| 9 | 418 | 19 |
| 10 | 427 | 18 |
| 11 | 434 | 18 |
| 12 | 442 | 17 |
| **13** | **448** | **17** |
| 14 | 455 | 16 |
| 15 | 461 | 16 |
| 16 | 466 | 15 |
| 17 | 472 | 15 |
| 18 | 477 | 14 |
| 19 | 483 | 14 |
| 20 | 488 | 14 |
| 21 | 493 | 14 |
| **22** | **498** | **14** |
| 23 | 503 | 14 |
| 24 | 508 | 14 |
| 25 | 513 | 14 |
| 26 | 519 | 14 |
| 27 | 524 | 14 |
| 28 | 530 | 15 |
| 29 | 536 | 15 |
| **30** | **543** | **16** |
| 31 | 550 | 17 |
| 32 | 558 | 18 |
| 33 | 567 | 20 |
| 34 | 578 | 22 |
| 35 | 590 | 24 |
| 36 | 605 | 28 |
| 37 | 624 | 33 |
| 38 | 652 | 43 |
| 39 | 703 | 67 |
| 40 | 725 | 80 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-45. Scoring Table for Science Grade 8**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 480 | 77 |
| 1 | 480 | 77 |
| 2 | 480 | 77 |
| 3 | 480 | 77 |
| 4 | 524 | 49 |
| 5 | 559 | 35 |
| 6 | 580 | 29 |
| 7 | 596 | 25 |
| 8 | 608 | 23 |
| 9 | 619 | 21 |
| 10 | 628 | 20 |
| 11 | 636 | 19 |
| 12 | 644 | 18 |
| 13 | 651 | 17 |
| **14** | **657** | **16** |
| 15 | 664 | 16 |
| 16 | 669 | 15 |
| 17 | 675 | 15 |
| 18 | 680 | 15 |
| 19 | 686 | 14 |
| 20 | 691 | 14 |
| **21** | **696** | **14** |
| 22 | 701 | 14 |
| 23 | 706 | 14 |
| 24 | 711 | 14 |
| 25 | 716 | 14 |
| 26 | 721 | 14 |
| 27 | 726 | 14 |
| 28 | 732 | 14 |
| **29** | **737** | **15** |
| 30 | 744 | 15 |
| 31 | 750 | 16 |
| 32 | 757 | 17 |
| 33 | 765 | 18 |
| 34 | 775 | 19 |
| 35 | 785 | 21 |
| 36 | 799 | 25 |
| 37 | 816 | 29 |
| 38 | 844 | 40 |
| 39 | 905 | 77 |
| 40 | 945 | 114 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-46. Scoring Table for Social Studies Grade 4**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 330 | 83 |
| 1 | 330 | 83 |
| 2 | 330 | 83 |
| 3 | 330 | 83 |
| 4 | 330 | 83 |
| 5 | 360 | 54 |
| 6 | 391 | 32 |
| 7 | 408 | 24 |
| 8 | 419 | 21 |
| 9 | 429 | 18 |
| 10 | 436 | 16 |
| 11 | 443 | 15 |
| 12 | 449 | 14 |
| 13 | 454 | 13 |
| 14 | 459 | 13 |
| **15** | **464** | **12** |
| 16 | 468 | 12 |
| 17 | 473 | 12 |
| 18 | 477 | 12 |
| 19 | 481 | 12 |
| 20 | 486 | 12 |
| 21 | 490 | 11 |
| **22** | **494** | **12** |
| 23 | 498 | 12 |
| 24 | 503 | 12 |
| 25 | 507 | 12 |
| 26 | 512 | 12 |
| 27 | 516 | 12 |
| 28 | 521 | 12 |
| 29 | 526 | 13 |
| 30 | 532 | 13 |
| **31** | **538** | **14** |
| 32 | 544 | 14 |
| 33 | 551 | 15 |
| 34 | 559 | 17 |
| 35 | 569 | 19 |
| 36 | 581 | 22 |
| 37 | 597 | 26 |
| 38 | 619 | 34 |
| 39 | 660 | 53 |
| 40 | 700 | 81 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-47. Scoring Table for Social Studies Grade 8**

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 540 | 94 |
| 1 | 540 | 94 |
| 2 | 540 | 94 |
| 3 | 540 | 94 |
| 4 | 540 | 94 |
| 5 | 540 | 94 |
| 6 | 540 | 94 |
| 7 | 578 | 56 |
| 8 | 607 | 34 |
| 9 | 623 | 25 |
| 10 | 634 | 20 |
| 11 | 642 | 18 |
| 12 | 649 | 16 |
| 13 | 655 | 14 |
| 14 | 660 | 13 |
| **15** | **665** | **13** |
| 16 | 670 | 12 |
| 17 | 674 | 12 |
| 18 | 678 | 11 |
| 19 | 682 | 11 |
| 20 | 686 | 11 |
| 21 | 690 | 11 |
| **22** | **694** | **11** |
| 23 | 698 | 10 |
| 24 | 702 | 10 |
| 25 | 706 | 11 |
| 26 | 710 | 11 |
| 27 | 714 | 11 |
| 28 | 718 | 11 |
| 29 | 722 | 11 |
| 30 | 727 | 11 |
| 31 | 732 | 12 |
| **32** | **737** | **12** |
| 33 | 742 | 13 |
| 34 | 749 | 14 |
| 35 | 756 | 15 |
| 36 | 765 | 17 |
| 37 | 777 | 20 |
| 38 | 793 | 25 |
| 39 | 823 | 39 |
| 40 | 860 | 65 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

## Table 6-48. Scoring Table for Social Studies Grade 10

| Raw Score | Scale Score | SEM |
|---|---|---|
| 0 | 645 | 87 |
| 1 | 645 | 87 |
| 2 | 645 | 87 |
| 3 | 645 | 87 |
| 4 | 645 | 87 |
| 5 | 645 | 87 |
| 6 | 645 | 87 |
| 7 | 685 | 50 |
| 8 | 712 | 34 |
| 9 | 728 | 27 |
| 10 | 740 | 23 |
| 11 | 749 | 20 |
| 12 | 757 | 18 |
| 13 | 764 | 16 |
| **14** | **770** | **15** |
| 15 | 775 | 14 |
| 16 | 780 | 13 |
| 17 | 785 | 13 |
| 18 | 790 | 12 |
| 19 | 794 | 12 |
| 20 | 798 | 12 |
| 21 | 802 | 11 |
| **22** | **806** | **11** |
| 23 | 810 | 11 |
| 24 | 815 | 11 |
| 25 | 819 | 11 |
| 26 | 823 | 11 |
| 27 | 827 | 11 |
| 28 | 831 | 11 |
| 29 | 836 | 12 |
| **30** | **841** | **12** |
| 31 | 846 | 12 |
| 32 | 851 | 13 |
| 33 | 857 | 13 |
| 34 | 864 | 14 |
| 35 | 872 | 16 |
| 36 | 882 | 18 |
| 37 | 894 | 21 |
| 38 | 912 | 28 |
| 39 | 945 | 43 |
| 40 | 980 | 67 |

Note: **Bold** represents CSEM around cut score (or the next higher scale score if the cut score value is not in the table).

**Table 6-49. Numbers and Percentages of Students at LOSS and HOSS**

| Content | Grade | LOSS | N | Percentage | HOSS | N | Percentage |
|---|---|---|---|---|---|---|---|
| ELA | 3 | 330 | 35 | 0.06 | 900 | 0 | 0.00 |
| | 4 | 340 | 27 | 0.05 | 930 | 0 | 0.00 |
| | 5 | 350 | 31 | 0.05 | 940 | 2 | 0.00 |
| | 6 | 360 | 11 | 0.02 | 950 | 1 | 0.00 |
| | 7 | 370 | 10 | 0.02 | 960 | 0 | 0.00 |
| | 8 | 380 | 52 | 0.08 | 970 | 4 | 0.01 |
| Mathematics | 3 | 360 | 752 | 1.29 | 760 | 244 | 0.42 |
| | 4 | 405 | 1634 | 2.77 | 800 | 63 | 0.11 |
| | 5 | 430 | 1500 | 2.53 | 830 | 59 | 0.10 |
| | 6 | 440 | 1519 | 2.52 | 870 | 107 | 0.18 |
| | 7 | 450 | 1611 | 2.60 | 880 | 48 | 0.08 |
| | 8 | 470 | 2050 | 3.27 | 890 | 62 | 0.10 |
| Science | 4 | 300 | 19 | 0.03 | 725 | 16 | 0.03 |
| | 8 | 480 | 34 | 0.05 | 945 | 69 | 0.11 |
| Social Studies | 4 | 330 | 312 | 0.53 | 700 | 102 | 0.17 |
| | 8 | 540 | 1394 | 2.23 | 860 | 201 | 0.32 |
| | 10 | 645 | 1003 | 1.69 | 980 | 65 | 0.11 |

**Figure 6-1. Anchor Set Test Characteristic Curves, English Language Arts Grades 3 through 8**
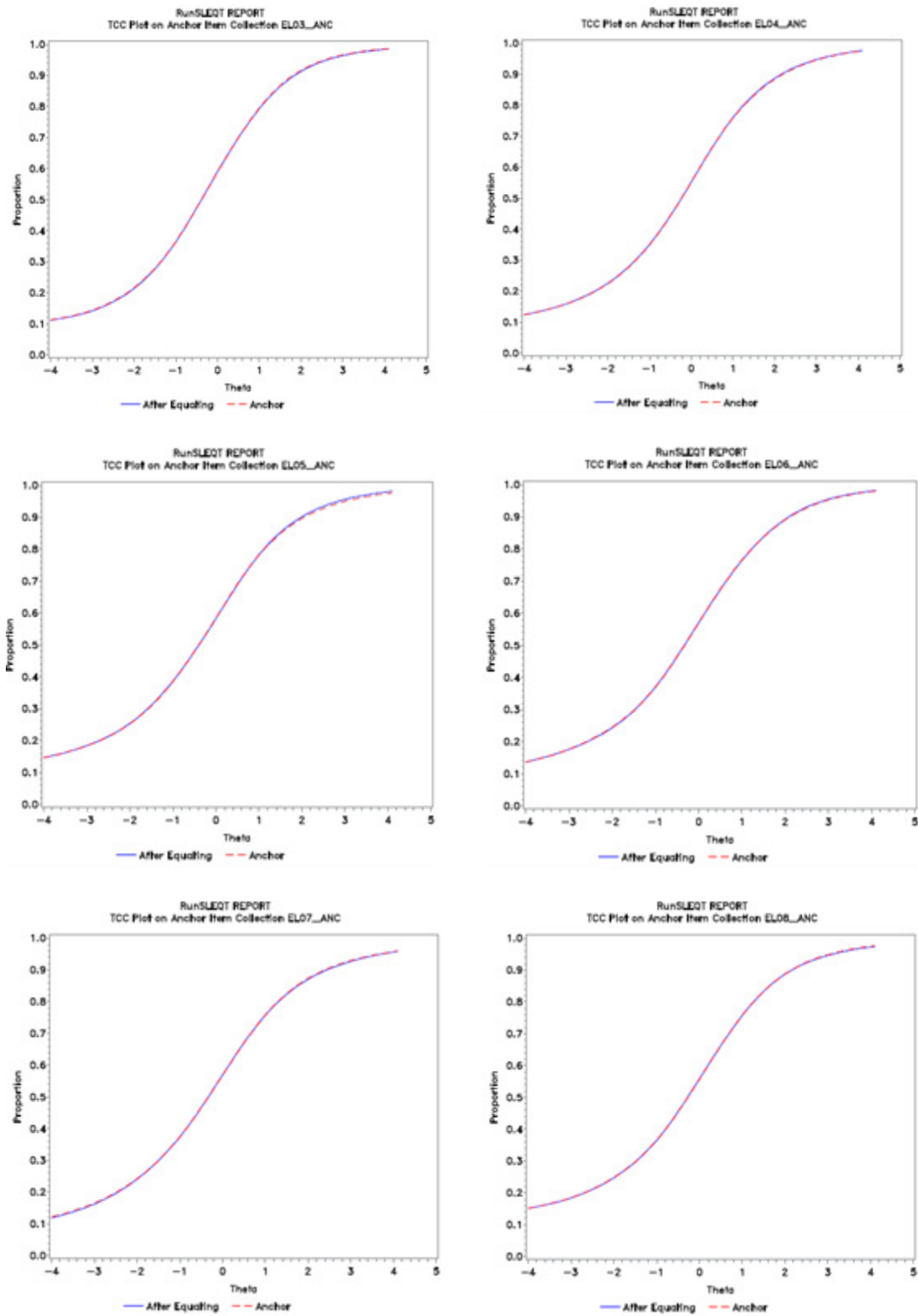
# Figure 6-2. Anchor Set Test Characteristic Curves, Mathematics Grades 3 through 8
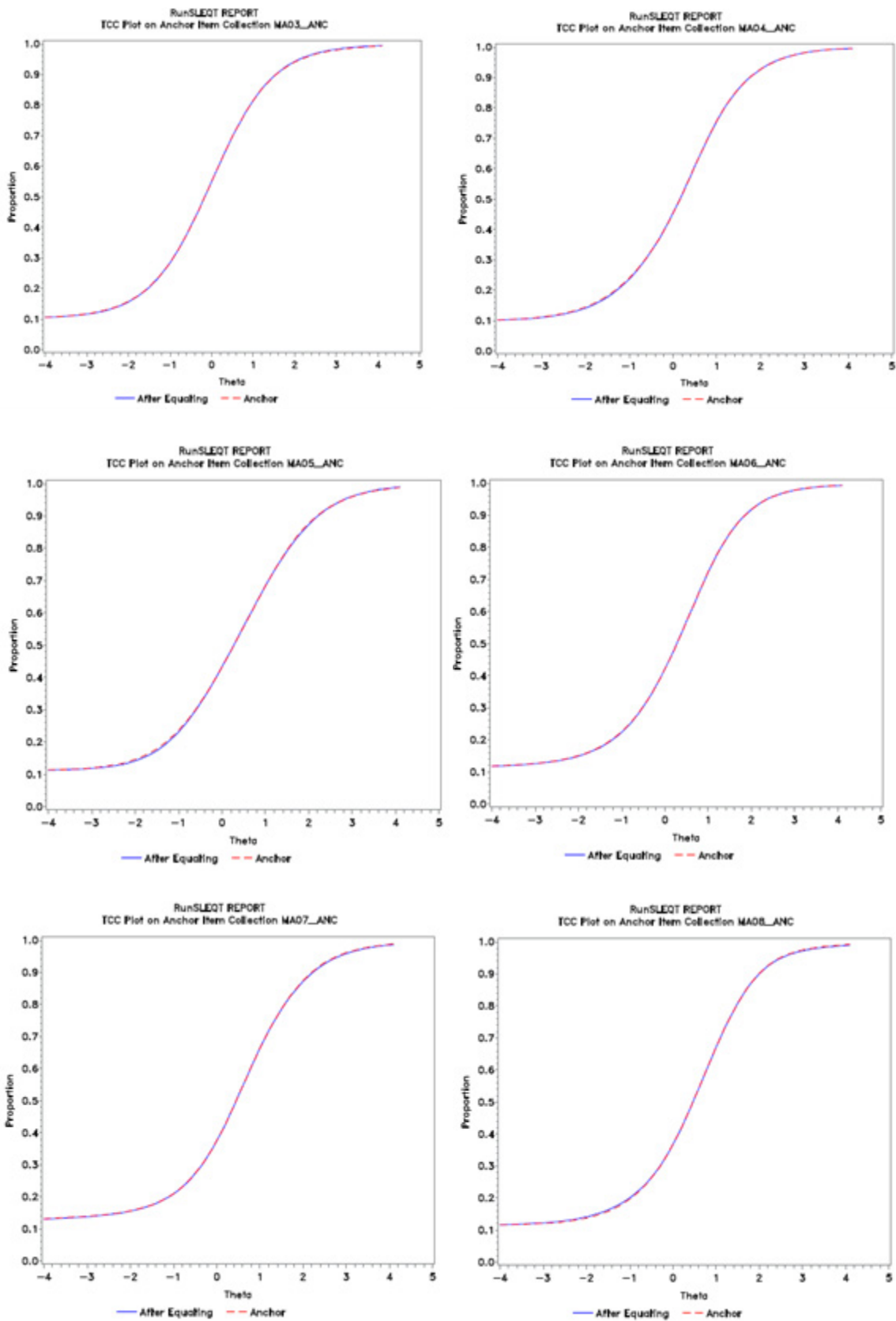
**Figure 6-3. Anchor Set Test Characteristic Curves, Science Grades 4 and 8**
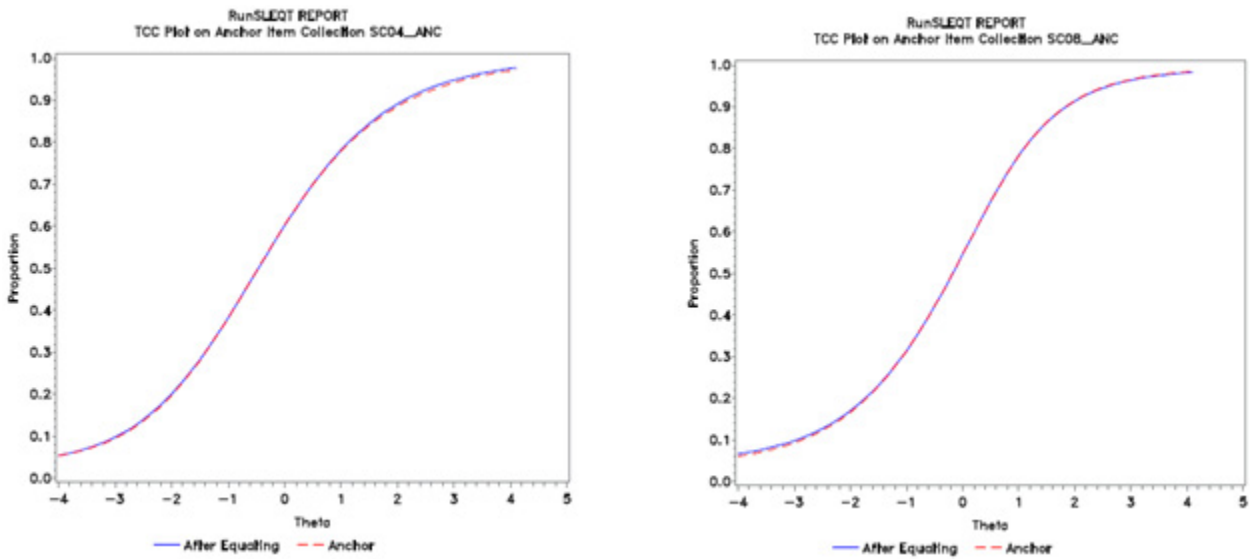


**Figure 6-4. Item Characteristic Curves for the Flagged Mathematics Grade 8, Anchor #44**
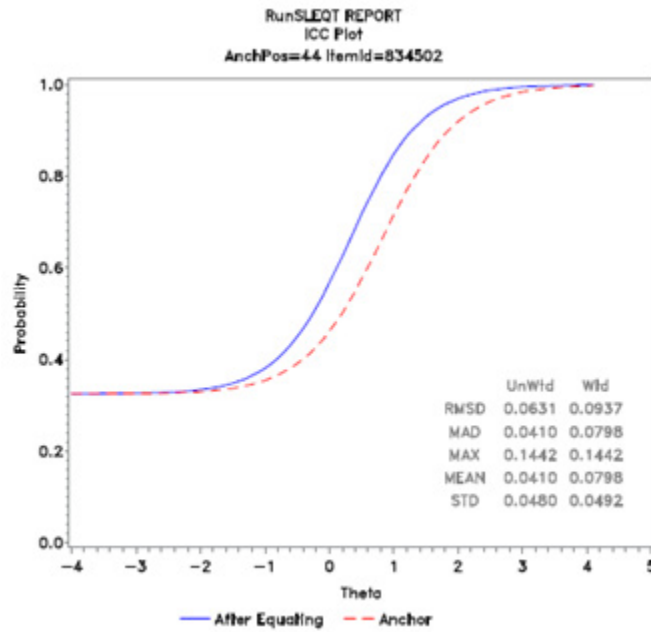
**Figure 6-5. Test Characteristic Curves, English Language Arts**
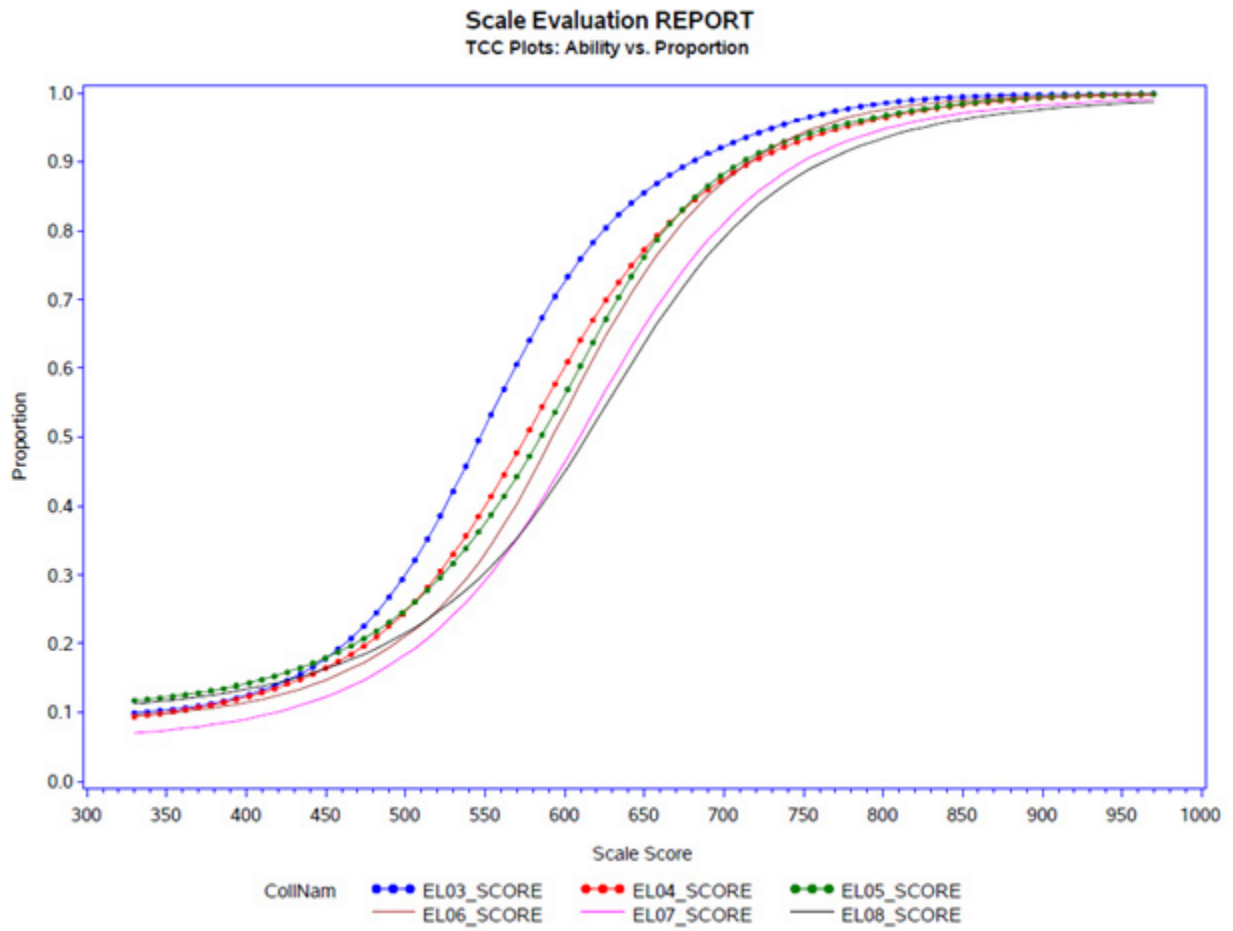
**Figure 6-6. Standard Error Curves, English Language Arts**



Scale Evaluation REPORT REPORT
Constrained IP SEM Plots: Ability vs. SEM
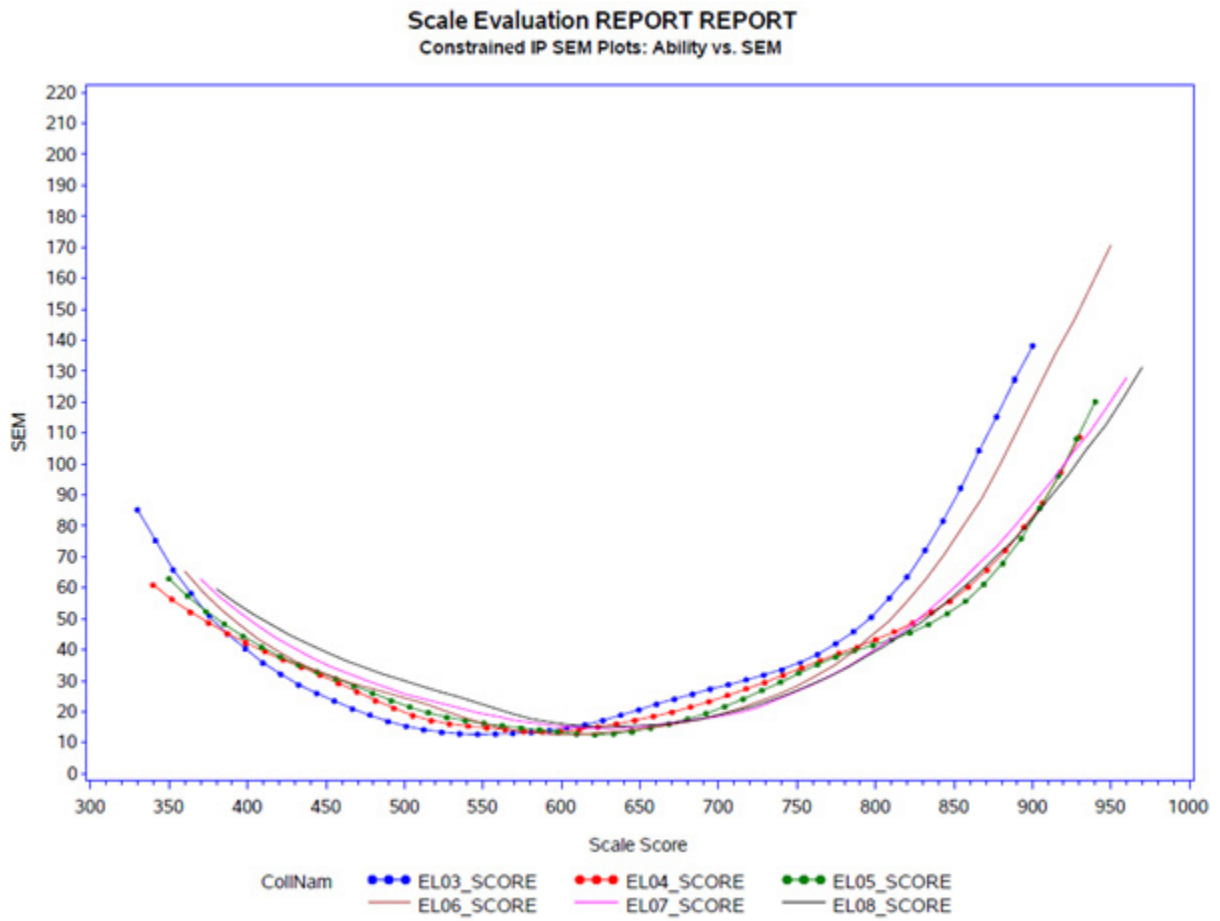
**Figure 6-7. Scale Scores and Growth at Quartiles, English Language Arts**



Scale Evaluation REPORT
Growth Curve at Quantiles
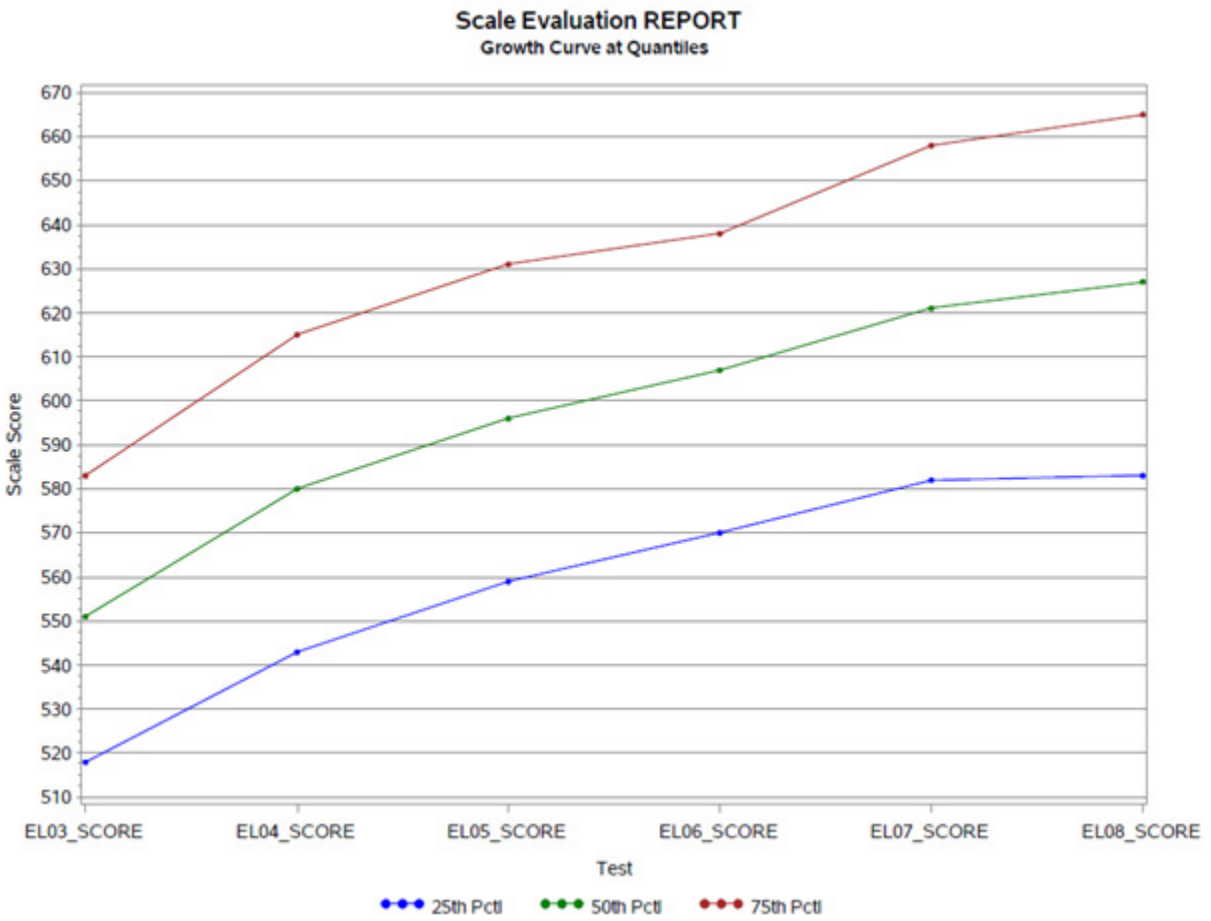
**Figure 6-8. Test Characteristic Curves, Mathematics**
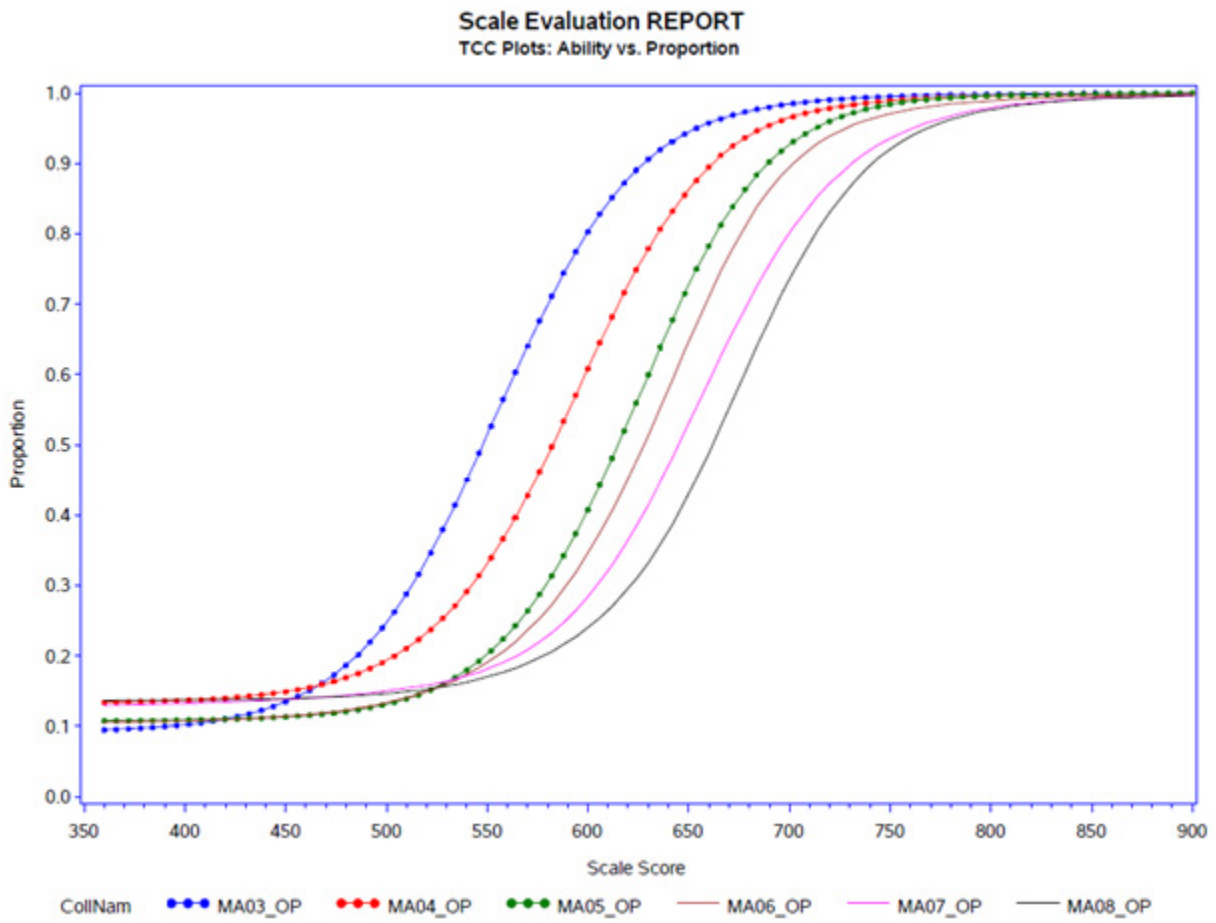
**Figure 6-9. Standard Error Curves, Mathematics**



**Scale Evaluation REPORT REPORT**
Constrained IP SEM Plots: Ability vs. SEM
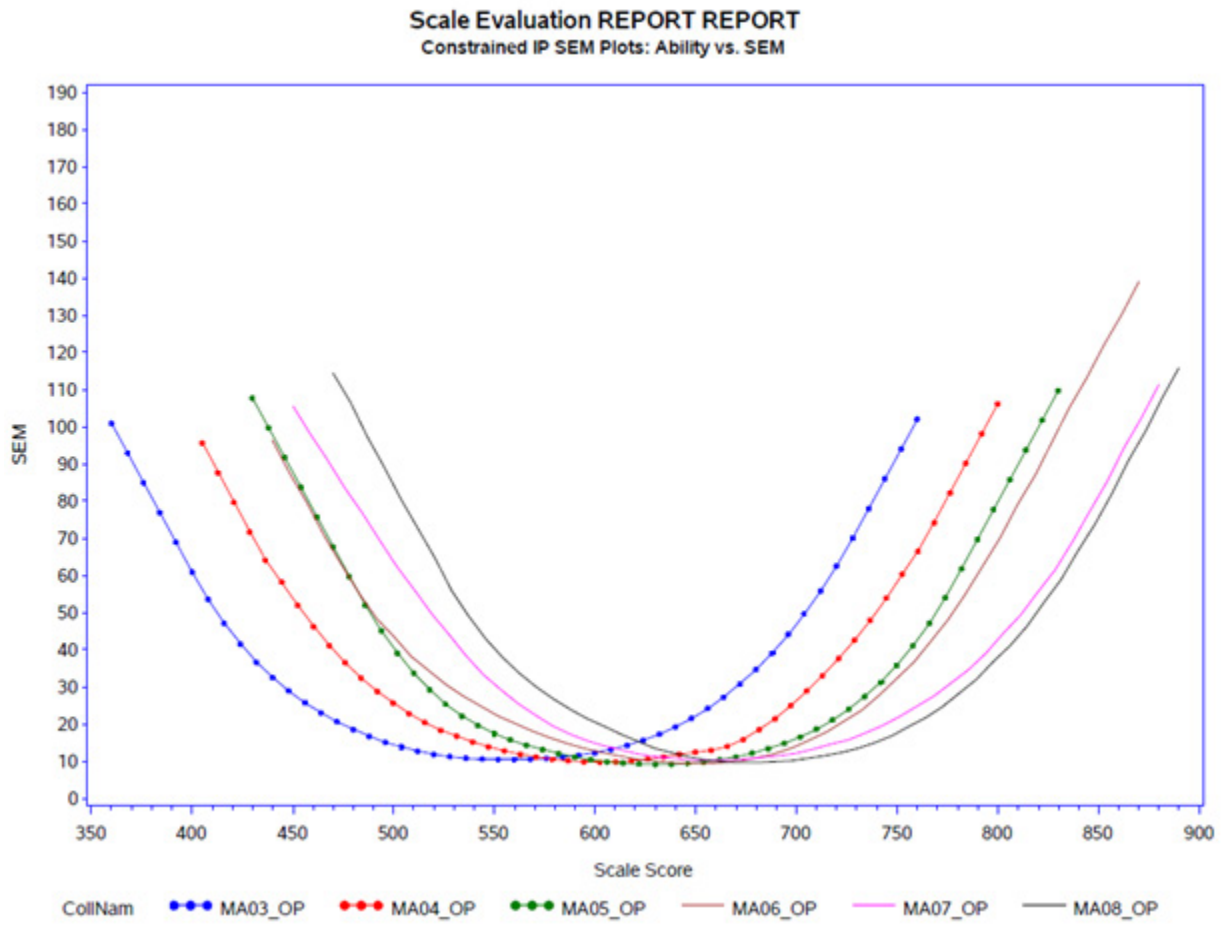
**Figure 6-10. Scale Scores and Growth at Quartiles, Mathematics**
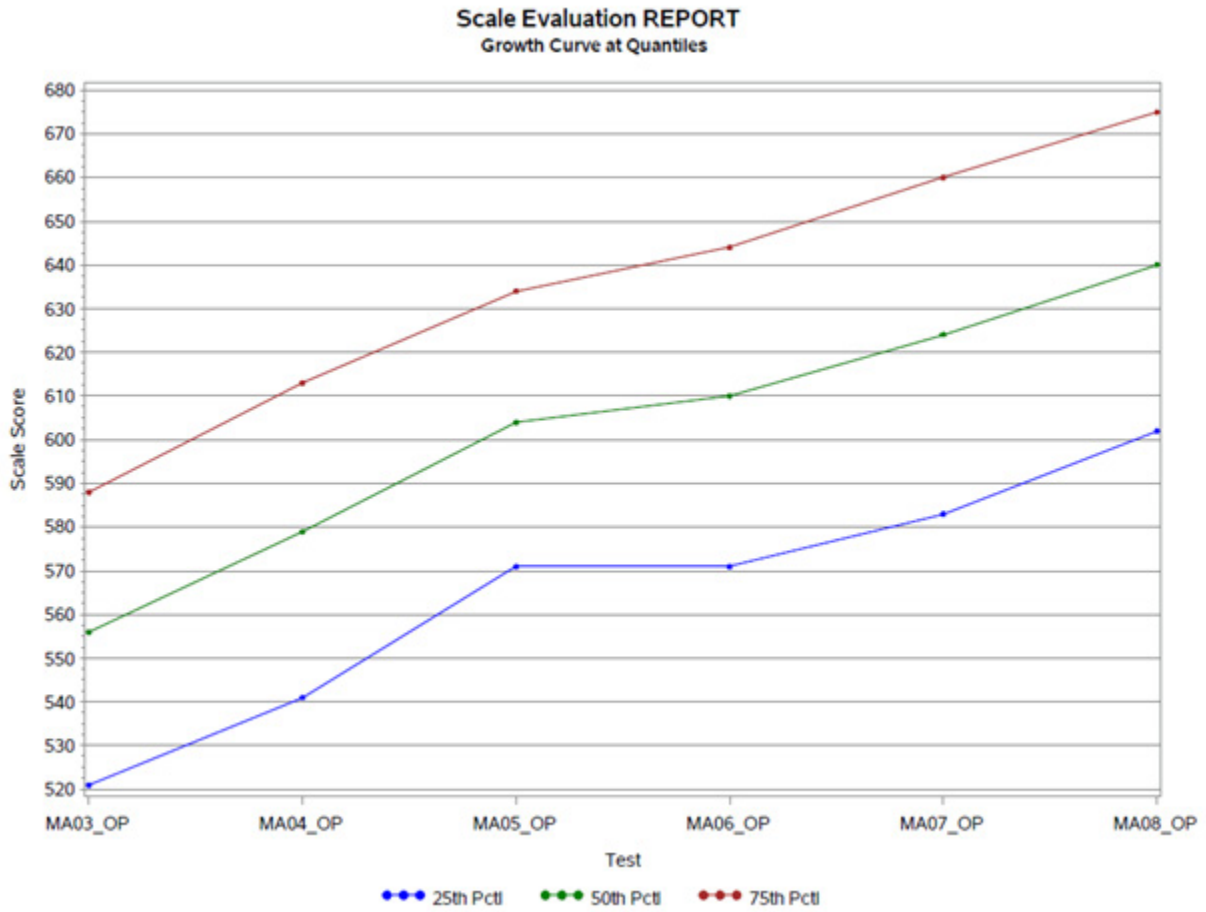
**Figure 6-11. Test Characteristic Curves, Science**
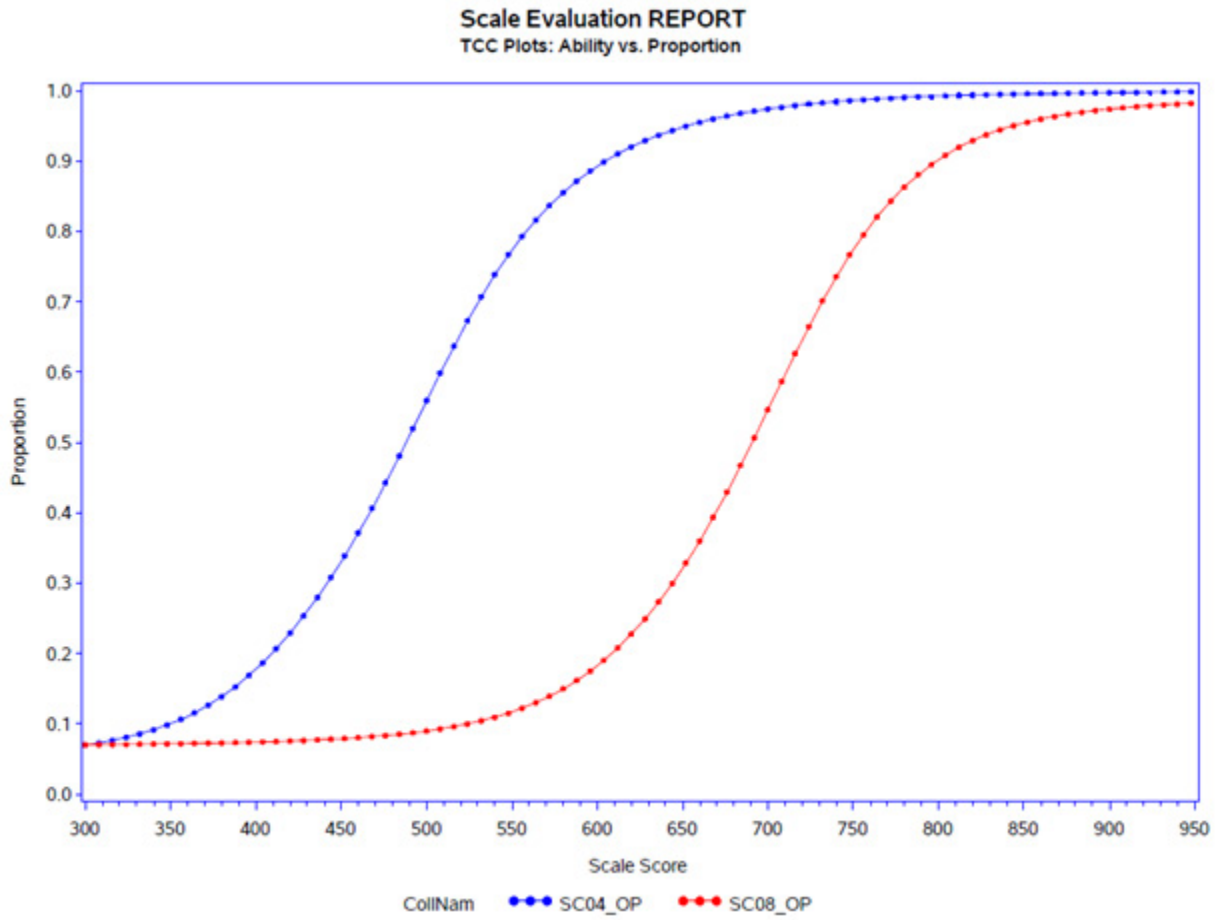
**Figure 6-12. Standard Error Curves, Science**
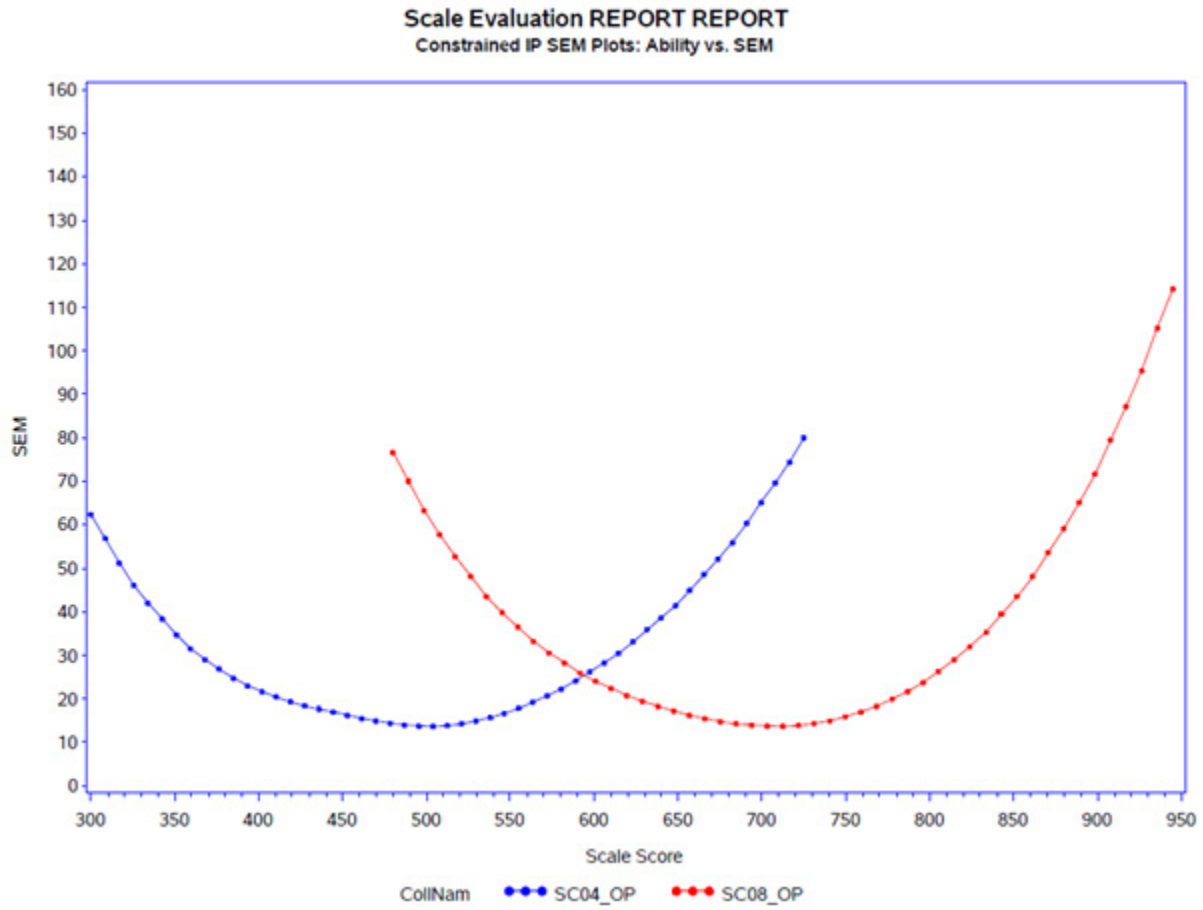
**Figure 6-13. Scale Scores at Quartiles, Science**



Scale Evaluation REPORT
Growth Curve at Quantiles
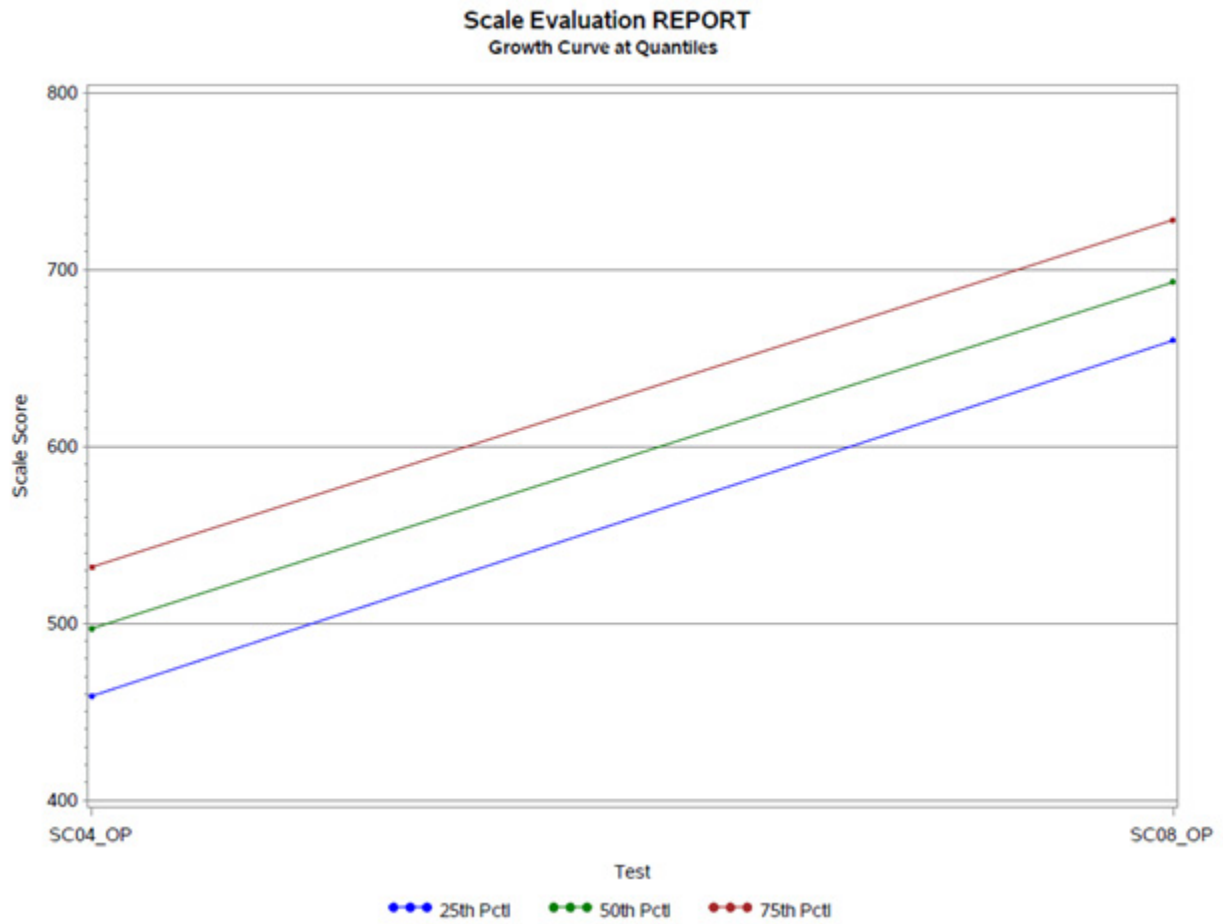
**Figure 6-14. Test Characteristic Curves, Social Studies**



Scale Evaluation REPORT
TCC Plots: Ability vs. Proportion
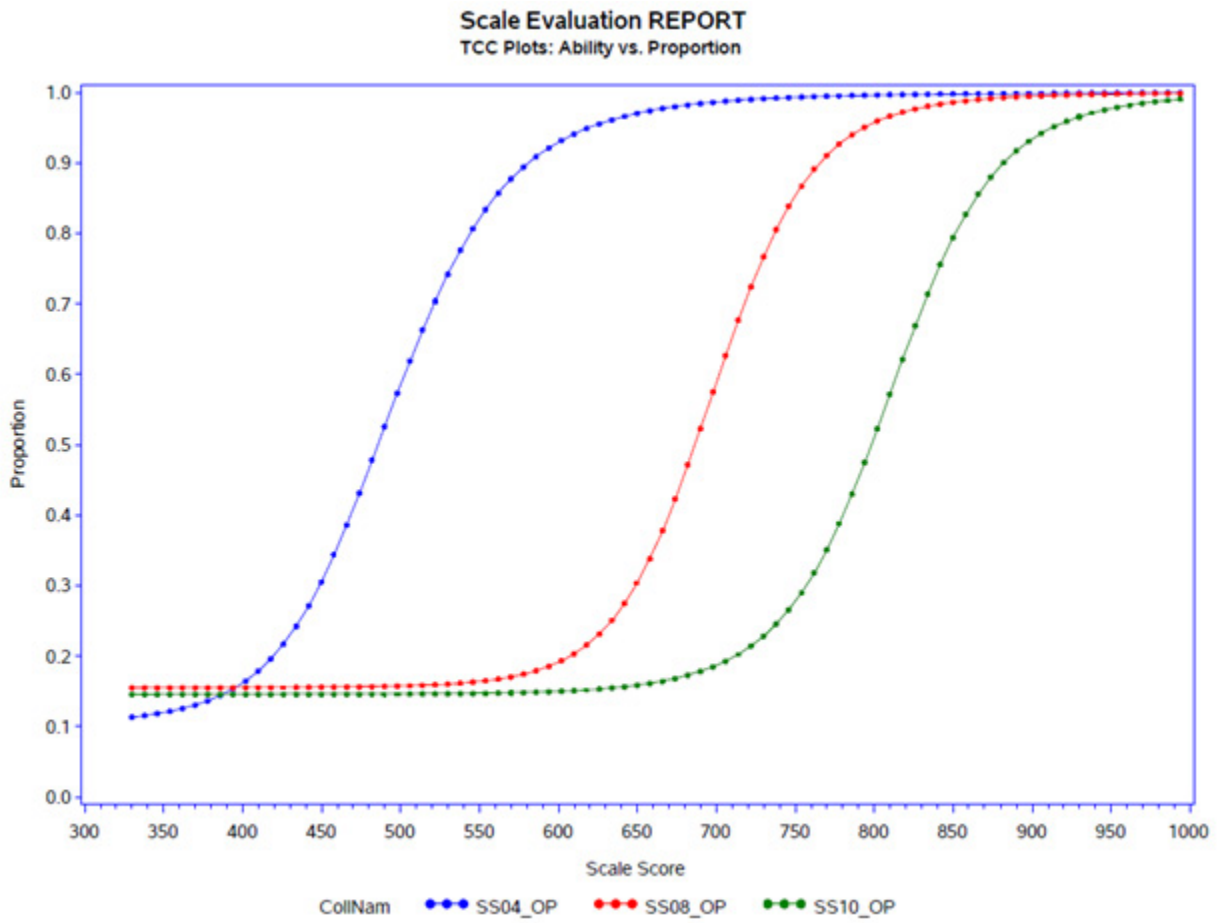
**Figure 6-15. Standard Error Curves, Social Studies**



Scale Evaluation REPORT REPORT
Constrained IP SEM Plots: Ability vs. SEM
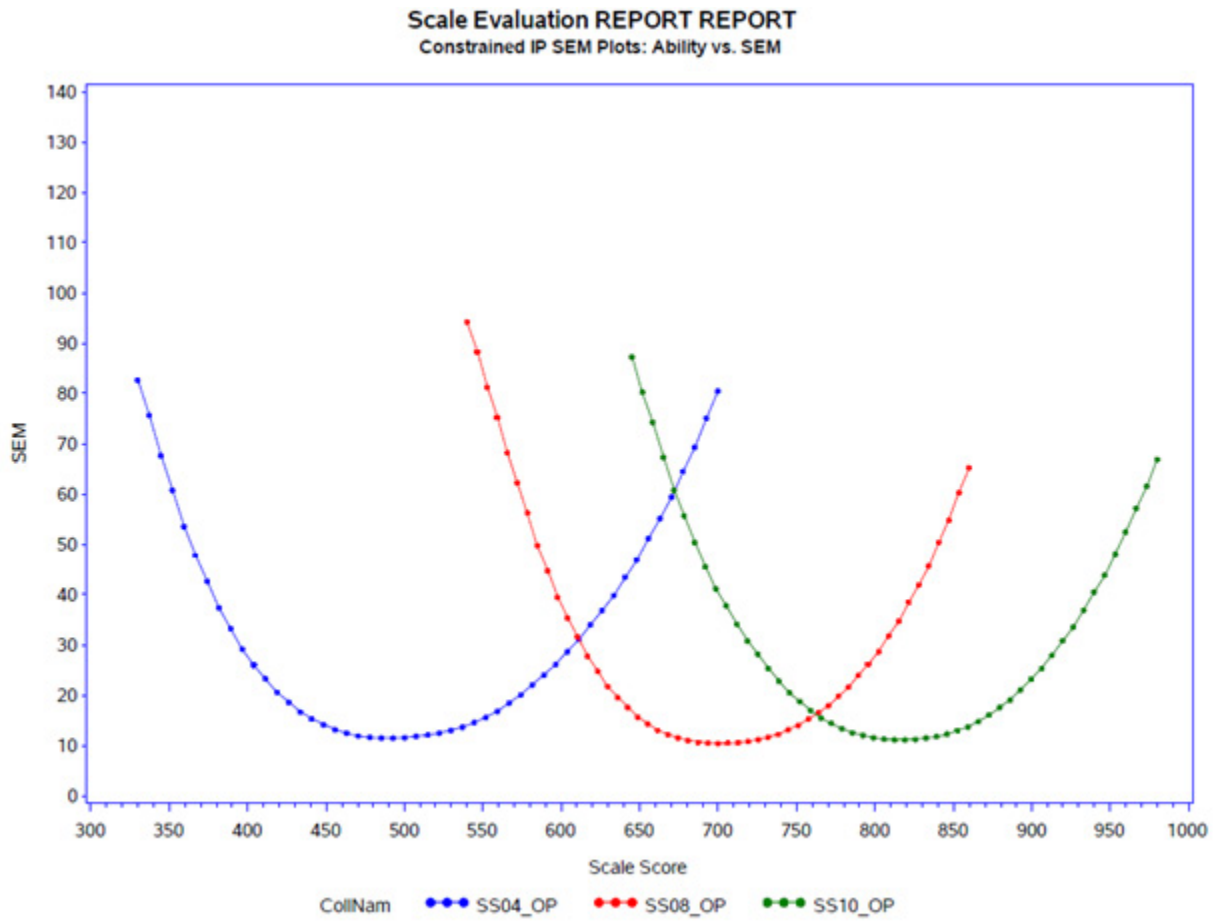
**Figure 6-16. Scale Scores at Quartiles, Social Studies**



Scale Evaluation REPORT
Growth Curve at Quantiles
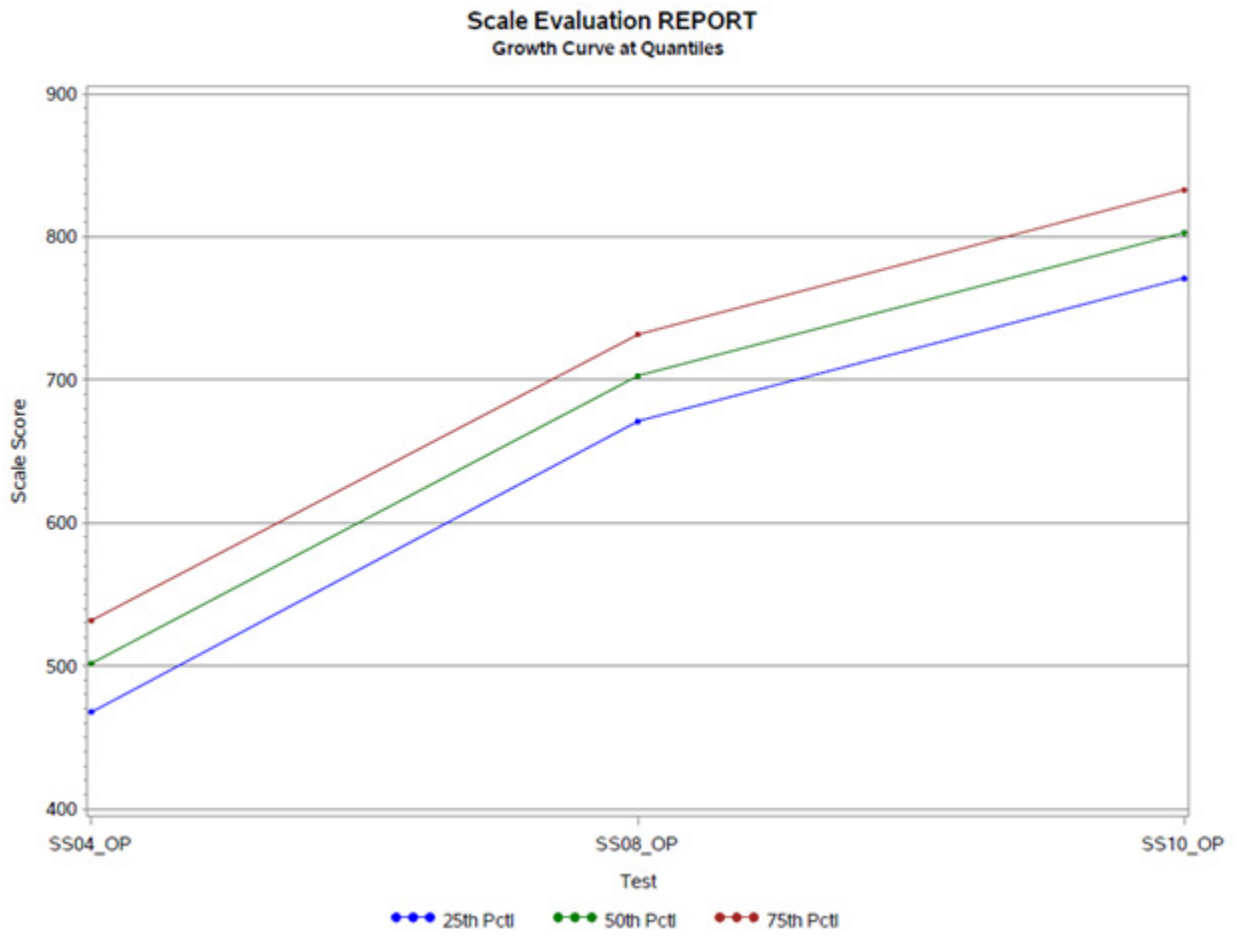
# Part 7: Standard Setting

In this part of the report, the standard settings that were conducted for Wisconsin Forward Exams are briefly described. The first standard setting for ELA, Mathematics, Science, and Social Studies occurred in Spring 2016. After the implementation of new Wisconsin Standards for Science, a new standard setting was conducted for Science in Spring 2019. Following the implementation of new Wisconsin Standards for Social Studies, a new standard setting for Social Studies was held in Spring 2022.

The cut scores established during these workshops and the performance level descriptors derived from the standard setting are also presented in this section of the report. The information on the ELA, Mathematics, Science, and Social Studies Spring 2016 standard setting comes from the *Wisconsin Standard Setting 2016 Final Technical Report*, and the information on the Science Spring 2019 standard setting comes from the *Wisconsin Standard Setting 2019 Final Technical Report*. Both reports are available at http://dpi.wi.gov/assessment/forward/resources.

## 7.1 Background Information

Several changes were made to Wisconsin's statewide tests in recent years. In the 2014–15 school year, the Wisconsin Badger Exam measured students' abilities in ELA and Mathematics using assessments developed by the Smarter Balanced Assessment Consortium (SBAC). Cut scores for the Wisconsin Badger Exam were taken from the national SBAC standard setting, conducted in 2014. For Science and Social Studies, the Wisconsin Knowledge and Concepts Examination (WKCE) was administered. Cut scores for the WKCE were established in 2005.

In the 2015–16 school year, DPI consolidated the Wisconsin Badger Exam and the WKCE into a unified program, the Wisconsin Forward Exam. At the inception of the Wisconsin Forward Exam, DPI indicated that they would no longer use SBAC items or test scales for ELA and Mathematics and that new test scales would be established for the Wisconsin Forward Exam. New test scales and performance levels were established for all four content areas using data from the Spring 2016 administration of the Wisconsin Forward Exam.

Changes to Wisconsin Science standards, test blueprint, and test design were implemented for the Spring 2019 Science operational test administration. New scales were developed, and new performance level cut scores were set for Science tests in Spring 2019.

New test blueprints and test designs were implemented for Spring 2022 Social Studies after adopting new Wisconsin Standards for Social Studies. New scales were developed and new performance level cut scores were set for Social Studies tests in Spring 2022.

## 7.2 Standard Setting Methodology and Process

Prior to the standard setting workshops in Spring 2016, 2019, and 2022, DPI worked in collaboration with DRC and its other technical advisors to select the methodology to be used at the standard setting. In recognition of its use in Wisconsin and widespread use across the country, the Bookmark Standard Setting Procedure (BSSP) for the Wisconsin Forward Exam was selected for use by DPI. The BSSP was well suited for standard setting for these assessments because (a) the tests are composed of both MC and non-MC items, (b) the items are scaled and can be mapped using item mapping techniques, and (c) the BSSP allows participants to focus on the knowledge, skills, and abilities expected of students in each

performance level. The BSSP has been well documented in standard setting literature. Developed in 1996, the BSSP has been implemented in over half of the states in the United States and abroad by DRC and by other major testing firms, making it the most widely used standard setting procedure in K–12 education (Karantonis & Sireci, 2006; Cizek & Bunch, 2007).

### 7.2.1 Spring 2016 Standard Setting for All Content Areas

On June 14–17, 2016, DPI and DRC conducted the Wisconsin Forward Exam standard setting for grades 3–8 in ELA and Mathematics, grades 4 and 8 in Science, and grades 4, 8, and 10 in Social Studies. The purpose of the standard setting was to develop performance standards for the Wisconsin Forward Exam, including the development of cut scores that divide students into four performance levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. During this benchmarked standard setting, DPI developed cut scores on the Wisconsin Forward Exam that reflected these content-based expectations on the tests, as informed by test data from well respected measures of student achievement.

A total of 59 Wisconsin educators and stakeholders worked individually and in committees to recommend performance standards associated with the four performance levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. This process yielded performance standards for the 17 tests of the Wisconsin Forward Exam program. The performance standards were approved by the Superintendent of Public Instruction in July 2016. For more information on the ELA, Mathematics, Science, and Social Studies standard setting, refer to *Wisconsin Standard Setting 2016 Final Technical Report*, available at http://dpi.wi.gov/assessment/forward/resources.

### 7.2.2 Spring 2019 Standard Setting for Science

Because the Science test blueprint and design changed for the Spring 2019 administration and new Science reporting scales were developed, a new performance level setting was needed for this content area. On May 29 and 30, 2019, DPI and DRC conducted the Wisconsin Forward Exam standard setting for grades 4 and 8 in Science. The purpose of the standard setting was to develop new performance standards for the Science tests, including the development of cut scores that divided students into the four performance levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. During the standard setting, Wisconsin educators made recommendations for cut scores on the Wisconsin Forward Exam that reflected the content-based expectations on the tests, as informed by test data from other measures of student Science achievement.

A total of 27 Wisconsin educators, 13 for grade 4 and 14 for grade 8, working individually and in grade-specific committees, recommended performance standards associated with the four performance levels for the two Science assessments: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. Participants engaged in three rounds of discussions and judgments to make their cut score recommendations. The cut scores recommended by the committee were approved by the State Superintendent of Public Instruction on June 5, 2019. *Wisconsin Standard Setting 2019 Final Technical Report*. Both reports are available at http://dpi.wi.gov/assessment/forward/resources.

### 7.2.3 Spring 2022 Standard Setting for Social Studies

On May 24–26, 2022, a committee of 50 Wisconsin educators participated in a standard setting workshop for the Wisconsin Forward Exam for Social Studies in grades 4, 8, and 10. At the workshop, participating educators recommended cut scores to divide students into four performance levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. The standard setting was required because of recent changes to the tests: the Spring 2022 Social Studies tests were the first to measure operationally the Wisconsin Standards for Social Studies, published in 2018, and new cut scores were needed to align with the new content standards.

During the standard setting workshop, participants were divided into three groups, one per tested grade. To make recommendations, participants engaged in three rounds of discussions and judgments. Specifically, the committee performed the following tasks:

1. Participants reviewed the Wisconsin Standards for Social Studies and the performance level descriptors (PLDs) for the test. The PLDs described the content-based expectations for students in each performance level.

2. Participants discussed the expectations for students at the point-of-entry for each performance level, that is, the *threshold students*.

3. Participants studied a collection of items from the pool assessing the new content standards, comprising operational test items from spring 2022 and field test items from spring 2021 and 2022. Items were ordered by difficulty based on student performance.

4. For each item, participants considered the knowledge and skills needed to answer the question correctly. Participants considered which items measured skills that each threshold student was expected to show command of.

5. Participants individually engaged in three rounds of cut score recommendations, termed bookmark placement at the workshop. After Round 1, the committee was shown benchmarks based on the performance of students on the 2019 administration of the social studies tests. These benchmarks were shown for participants' reference.

6. After Round 1, participants in each grade-group discussed their bookmark placements in two small groups, and they shared why they placed their bookmarks where they did. Participants then worked individually to revise their judgments.

7. After Round 2, participants were shown the percentage of students who would be classified in each performance level if the committee's Round 2 recommendations were applied to students (i.e., the *impact data*). Participants discussed their bookmarks in grade-groups. Participants then worked individually to revise their bookmark placements.

8. Participants then examined their Round 3 recommendations and impact data. The committee examined the across-grade consistency (articulation) of the performance standards, and they assented to sending their recommendations forward to DPI.

After the workshop, the participants' recommendations were reviewed by DPI and the Superintendent, who have the responsibility of implementing cut scores for the assessment such that the final cut scores (a) reflect the knowledge and skills expected of students in each performance level, and (b) reflect the policy-based expectations for educators and stakeholders across the Wisconsin Forward Exam program. Accordingly, adjustments to the cut scores recommended by the participants may be considered to reflect the policy-based expectations for the program. After the review of the recommended cut scores, the *Advanced* cut scores for grades 4 and 8 were adjusted by one-half of one conditional standard error of measurement value (+0.5 CSEM) in order to promote consistency with the grade 10 percentages of students classified in the highest performance level.

The process of all three standard settings adhered to AERA, APA, & NCME (2014) Standards 5.21 and 5.22, which state the following:

> **Standard 5.21** When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly. (p. 107)

**Standard 5.22** When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way. (p. 108)

## 7.3 Performance Level Descriptors

In terms of the validity of the Wisconsin Forward Exam scores, it is essential to understand that descriptors and cut scores are established in a collaborative and participatory process. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, particularly cut scores. Performance level descriptors (PLDs) summarize the knowledge, skills, and abilities expected of students in each performance level. DPI provided policy PLDs for the Wisconsin Forward Exam during the Spring 2016, Spring 2019, and Spring 2022 standard settings. The brief policy performance level descriptors, shown in Table 7-1, describe DPI's vision for each performance level. In addition, the standards-based PLDs for the Wisconsin Forward Exam in Science were provided to the standard setting participants in Spring 2019 and Spring 2022. (For detailed standards-based PLDs, refer to the *Wisconsin Standard Setting 2019 Final Technical Report* and the *Wisconsin Forward Exam Social Studies Standard Setting 2022 Final Technical Report*.) At the most recent standard settings for Science and Social Studies, Wisconsin educators used the policy PLDs in conjunction with standards-based PLDs to consider the content-based expectations for students in each performance level on each Science test in the Wisconsin Forward Exam program.

## 7.4 Cut Scores

Table 7-2 shows the cut scores for all grades and content areas. The cut scores reflect the content-based expectations for students and policy-based decisions (i.e., the impact of the cut scores on Wisconsin students as shown through the impact data). The cut scores for ELA and Mathematics, established in Spring 2016, remained unchanged for the 2022 assessments. New cut scores were established for Science after the Spring 2019 test administration, and these cut scores were used for student classification into performance levels in Spring 2019 through 2022. New cut scores reflecting Wisconsin student performance on the new Social Studies assessments were established for Social Studies after the Spring 2022 test administration.

## 7.5 Summary

Part 7 presented a brief overview of the standard setting process used to establish the Wisconsin Forward Exam cut scores for all content areas after the Spring 2016 test administration, for Science after the Spring 2019 test administration, and for Social Studies after the Spring 2022 administration. All standard setting workshops are described in detail in their respective technical reports: *Wisconsin Standard Setting 2016 Final Technical Report*, *Wisconsin Standard Setting 2019 Final Technical Report*, and *Wisconsin Forward Exam Social Studies Standard Setting 2022 Final Technical Report*. The standard settings undertaken by DPI and facilitated by DRC support Standards 5.21 and 5.22 from the Standards (AERA, APA, & NCME, 2014).

**Table 7-1. Policy Performance Level Descriptors for the Wisconsin Forward Exam**

| Level | Performance Level Descriptor |
|---|---|
| *Below Basic* | Student demonstrates **minimal** understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness. |
| *Basic* | Student demonstrates **partial** understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness. |
| *Proficient* | Student demonstrates understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness. |
| *Advanced* | Student demonstrates **exemplary** understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness. |

**Table 7-2. Wisconsin Forward Exam Cut Scores**

| Content | Grade | *Basic* | *Proficient* | *Advanced* |
|---|---|---|---|---|
| English Language Arts | 3 | 522 | 570 | 624 |
| | 4 | 546 | 592 | 650 |
| | 5 | 564 | 610 | 670 |
| | 6 | 572 | 622 | 671 |
| | 7 | 585 | 638 | 697 |
| | 8 | 592 | 652 | 708 |
| Mathematics | 3 | 517 | 560 | 611 |
| | 4 | 536 | 588 | 633 |
| | 5 | 574 | 611 | 658 |
| | 6 | 582 | 626 | 688 |
| | 7 | 606 | 647 | 712 |
| | 8 | 620 | 667 | 718 |
| Science | 4 | 447 | 496 | 543 |
| | 8 | 653 | 695 | 737 |
| Social Studies | 4 | 461 | 491 | 537 |
| | 8 | 662 | 693 | 734 |
| | 10 | 770 | 805 | 837 |

# Part 8: Studies of Reliability

Part 8 of the Technical Report builds upon existing analyses of the summary results by providing additional estimates of the reliability of those results. Reliability can be defined as the consistency of an assessment when the testing procedure is repeated with the same testing target group. A reliable assessment is one that would produce stable scores if the same group of students were to take the same test repeatedly, without any fatigue or memory of the test. As detailed below, the reliability of the Spring 2022 Wisconsin Forward Exam was estimated in three ways:

- Internal consistency was assessed for all items using Cronbach's alpha (1951).

- Standard error of measurement (SEM) was calculated for raw score and scale score.

- Classification consistency and classification accuracy were estimated for the performance level classifications.

This part of the report addresses AERA, APA, & NCME (2014) Standards 2.0, 2.3, 2.11, 2.13, 2.14, and 2.16, which are cited below:

> **Standard 2.0** Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use. (p. 42)

> **Standard 2.3** For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported. (p. 43)

> **Standard 2.11** Test publishers should provide estimates of reliability/precision as soon as feasible for each relevant subgroup for which the test is recommended. (p. 45)

> **Standard 2.13** The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score. (p. 45)

> **Standard 2.14** When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score. (p. 46)

> **Standard 2.16** When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure. (p. 46)

Standard 2.3 advises providing reliability estimates and the SEM for all total scores and subscores reported, Standard 2.13 advises reporting SEM in both raw score and scale score units, and Standard 2.11 advises assessing reliability and SEM for all population subgroups. This part of the report presents raw score reliability coefficients and SEMs for the four Wisconsin Forward Exam content areas, for each reported content standard for the total group of examinees, and for the subgroups identified by gender, race/ethnicity, economic status, disability status, and English language proficiency. The scale score CSEMs are provided in Section 8.1.1.

Standard 2.16 advises that when testing measures are used to make categorical decisions, the reliability of those decisions should be estimated. In the present context, Standard 2.16 applies specifically to

performance level determinations, such as *Proficient* or *Advanced*. As described below, the Spring 2022 Wisconsin Forward Exam adhered to this standard by applying a detailed analysis of classification consistency and classification accuracy—two related measures used to evaluate the reliability of the performance level classifications used in the test program. This analysis also addresses Standard 2.14 by providing a CSEM for the cut scores that separate the performance levels.

Combined, Cronbach's alpha, SEM, classification consistency, and classification accuracy provide several forms of evidence related to the reliability of the Wisconsin Forward Exam. Cronbach's alpha and the SEM operate at the content level. For example, they provide estimates of reliability for student scores in ELA or Mathematics. Classification consistency and classification accuracy operate on the associated performance level classifications. These are of particular interest in the context of the Elementary and Secondary Education Act (ESEA) and the associated accountability requirements. In addition, the Cronbach's alpha statistics and the SEM were computed for content standards and domains, providing evidence of the reliability and precision of measurement of the Wisconsin Forward Exam subscores. Altogether, the evidence provided in this part of the Technical Report, which is targeted at each intended use of the Wisconsin Forward Exam scores, addresses Standard 2.0.

## 8.1 Measures of Internal Consistency and Standard Error of Measurement

Cronbach's alpha is a frequently used measure of internal consistency for tests consisting of MC and CR items. Cronbach's alpha ($\alpha$) is computed as

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_X^2}\right),$$

where $k$ = number of items, $\sigma_x^2$ = the total score variance, and $\sigma_i^2$ = the variance of item $i$ (Crocker & Algina, 1986). SEM is defined as

$$\text{SEM} = SD\sqrt{1 - R_{xx}}$$

where $SD$ represents the standard deviation of the raw score distribution and $R_{xx}$ represents Cronbach's alpha.

Cronbach's alpha and the SEM are shown in Tables 8-1 and 8-2, respectively. These tables include information for all students and for the subgroup categories of gender, race/ethnicity, English language proficiency, disability status, economic status, and accommodation use.

As shown in the "Total" column of Table 8-1, reliability ranges from 0.88 to 0.90 across grades for ELA and from 0.92 to 0.93 across grades for Mathematics. The reliability coefficient was 0.89 for Science grade 4 and 0.88 for Science grade 8. The reliability coefficient was 0.91 for Social Studies in grades 4 and 8 and 0.90 for Social Studies grade 10. All reliability coefficients would ideally be 0.90 or above. However, for relatively short tests that are designed to measure a fairly broad range of content, this is not always a realistic expectation. If 0.90 is considered a conservative criterion for an acceptable level of reliability as measured by Cronbach's alpha, then the ELA assessments for grades 3, 4, 5, 7, and 8 and the Science assessments in both grades would not meet this criterion. The reliability coefficients for these tests may be affected by the number of items (and score points) and the diversity of the content being assessed. Applying the Spearman-Brown prophecy formula to these results indicates that to achieve the 0.90 reliability threshold, the current ELA assessments for grades 3, 4, 5, 7, and 8 would need to be increased by

approximately 11, 4, 5, 7, and 9 points, respectively. For the current Science assessments in grades 4 and 8, the increase would need to be 4 score points and at least 9 score points, respectively.

Table 8-1 shows that many of the subgroup reliability coefficients were similar to, albeit slightly lower than, the total reliability coefficients. Reliability coefficients are particularly sensitive to score distribution and variance, so this result is consistent with the general variability among many of these subgroups.

The test reliability coefficients were comparable within 0.02 for male and female students across all grades and content areas. Most differences among the five racial/ethnic groups were small and within 0.05 of one another for all grades in ELA, Science, and Social Studies. In Mathematics, higher test reliabilities were observed for White and Asian students and the lowest reliability was observed for African American students in all grades.

The differences between test reliability coefficients for economically disadvantaged and not economically disadvantaged students were 0.02 for all grades and content areas, except for Mathematics grades 6 through 8, where the differences ranged from 0.03 to 0.04. The test reliability was higher for students who were not economically disadvantaged.

The differences in reliability of test scores between disabled and not disabled students were within 0.03 for most grades and content areas, except for ELA grade 7, Mathematics grades 6 through 8, and Social Studies grade 10, where the differences were between 0.04 to 0.08. The test reliability was found to be higher for students without disabilities.

The greatest differences in test reliability were between fully English proficient and limited English proficient students and between students using and not using testing accommodations, with consistently lower reliability among limited English proficient students and students using testing accommodations. The test reliability coefficients for limited English proficient students ranged from 0.81 to 0.84 for ELA, from 0.78 to 0.91 for Mathematics, from 0.75 to 0.83 for Science, and from 0.77 to 0.84 for Social Studies.

With the exception of grade 3, the reliability coefficients for students using testing accommodations in Mathematics were the lowest of all subgroups and ranged from 0.59 to 0.77 in grades 4 through 8. The reliability coefficients for students using testing accommodations in ELA should be interpreted with caution because of the low number of students using the accommodations. The reliability coefficients were not computed for students using testing accommodations in Science or Social Studies, because the number of students using accommodations in these subject areas was less than 50.

The reliability coefficient is affected by, among other factors, the variability of students' scores. The higher the variability of scores, the higher the reliability coefficient will tend to be. While the overall patterns of the differences in test reliability for different subgroups are similar to the patterns observed in Spring 2021, it was observed that the test reliability increased for the subgroups of African American students, students with disabilities, economically disadvantaged students, and particularly for students of limited English proficiency in Spring 2022 compared to Spring 2021.

Table 8-2 presents the raw score SEM for the total population and for the subgroups described above. These values provide important information for raw score interpretation since an individual's obtained score can be expected to fall within two SEMs of the individual's true score approximately 95% of the time. Although there were some observable differences in SEM for the different subgroups, all differences were generally within one-half of a score point. The SEMs for ELA assessments were slightly larger than those for the other content areas. Because these SEMs are in the raw score metric, this result is consistent with the fact that ELA tests have more raw score points and relatively larger raw score SDs than the other

content areas. For every grade and content area, the CSEM for individual scale scores is provided in the scoring tables previously discussed in Part 6 (Tables 6-32 through 6-48).

Reliability, as measured by Cronbach's alpha, was also computed for content standards (or reporting categories) within each content area as well as for each language domain in ELA. These data are presented in Table 8-3. The last column presents the reliability for the total test per grade for each content area (with all content standards or domains) for all examinees. It is clear that the reliability per content standard or domain is lower than the reliability for the total test per content area. The number of items or score points has a close relationship with reliability, and a smaller number of items or score points is generally associated with lower reliability. The number of score points for ELA per domain was 7 or 8 in Listening, 22 or 24 in Reading, and 24 in Writing/Language. The number of score points ranged from 4 to 12 per content standard (or reporting category) for ELA, from 7 to 11 per standard for Mathematics, from 8 to 13 per standard for Science, and from 6 to 11 per standard for Social Studies. A lower level of reliability per content standard or domain is therefore expected compared to each full content area assessment. The lower level of reliability per standard or domain is one of the reasons why the information based on the content standards or domains should be used for low-stakes purposes only. (This issue is also discussed in the context of standard performance index scores in Part 10.)

As shown in Table 8-3, the reliability ranges by content standard/domain were as follows:

- For ELA, reliability indices by content standard ranged from 0.28 for the Reading—Vocabulary Use standard in grade 5 to 0.71 for the Reading—Key Ideas and Details standard in grades 3 and 6. When the ELA domains were considered, the highest reliability, ranging from 0.80 to 0.85 across grades, was found for the Reading domain, followed by the Writing domain, with reliability coefficients ranging from 0.69 to 0.76 across grades, and followed by the Listening domain, which included a smaller number of items, resulting in the reliability coefficients ranging from 0.48 to 0.63 across grades.

- For Mathematics, reliability indices by content standard ranged from 0.56 (for the Geometry standard in grade 7) to 0.81 (for the Number and Operations—Fractions standard in grade 4).

- For Science, reliability indices by content standard ranged from 0.58 (for the Earth and Space Science standard in grade 4) to 0.74 (for the Engineering standard also in grade 4).

- For Social Studies, reliability indices by content standard ranged from 0.47 (for the Behavioral Sciences standard in grade 10) to 0.79 (for the Geography standard in grade 4).

The SEM associated with each content standard or domain is presented in Table 8-4 by content area and grade level. Some differences in SEM by content standard can be observed. As indicated by the discussion above, these SEMs were smaller than those for the total test and were generally consistent with the number of items within each content standard.

In summary, the reliability indices, as measured by Cronbach's alpha at the test level, are in a reasonable range given the number of items in each test. As described above, readers should also note that, because reliability is influenced by the number of items, lower reliability for the content standards with fewer items is to be expected.

### 8.1.1 Conditional Standard Error of Measurement

In contrast to the SEM, the CSEM expresses the degree of measurement error in scale score units and is conditioned on the ability of the student. The CSEM is defined as the reciprocal of the square root of the test information function and can be estimated across all points of the ability continuum (Hambleton & Swaminathan, 1985):

$$\text{CSEM}(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}},$$

where $I(\theta_i)$ is the test information function, computed as a sum of item information functions, obtained as

$$I(\theta_i) = \sum_j \frac{p'_{ij}(\theta_i)^2}{p_{ij}(\theta_i)q_{ij}(\theta_i)},$$

where $p'_{ij}(\theta_i)$ is the derivative of $p_{ij}(\theta_i)$ and $q_{ij}(\theta_i) = 1 - p_{ij}(\theta_i)$.

The CSEM can be used to obtain the range within which a student's true score is likely to fall (that is, with a certain degree of probability). It is expected that a student's score obtained from a single testing will fall within one CSEM of that student's true score 68 percent of the time and that the obtained score will fall within two CSEMs of the true score 95 percent of the time.

Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scale scores) and are smaller in the middle of the score distribution and larger at the tails. This pattern is seen for all Wisconsin Forward Exam CSEMs and is to be expected when IRT methods are used. In compliance with Standards 2.13 and 2.14, the CSEM of each cut score was presented in the raw score-to-scale score tables (Tables 6-32 through 6-48) for all grades and content areas in Part 6 of this report. In addition, graphical representation of the CSEM with the cut scores is presented in Figures I-1 through I-17 of Appendix I for all grades and content areas. As shown in Appendix I, the estimates of CSEM tend to be higher at the low and high ends of the scale score range. The CSEM increases when there are few observations at a particular ability level. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. Figures I-1 through I-17 demonstrate that the CSEM is minimized at the cut scores and in the middle of the scale range, where most students are located.

## 8.2 Classification Consistency and Accuracy

One of the primary goals of education policy is to improve the performance of all students, with a specific goal of having all students become *Proficient*. Because of this heavy emphasis on moving all students to levels of academic performance at or above each state's self defined *Proficient* category, the consistency and accuracy of the classification of students into these performance levels are of particular interest. The following section describes how the consistency and accuracy of these classifications were evaluated and provides evidence that supports the validity of these classifications.

Conceptually, classification consistency is defined as the extent to which two classifications of a single student agree, based either on two independent administrations of the same test or on one administration of two parallel test forms. However, it is difficult to obtain data from repeated administrations of the same form because of the cost, time, and student memory from prior administrations. It is also difficult

---

to construct two psychometrically parallel forms. For these reasons, the common practice is to estimate classification consistency from a single administration.

A contingency table representing the probability of particular classification outcomes under specific scenarios is a convenient way to measure classification consistency. The table below is a contingency table of $(H + 1) \times (H + 1)$, where H is the number of cut scores. Three cut scores yield a $4 \times 4$ contingency table, as can be seen below in Table 8-A.

It is common to report two indices of classification consistency: the classification agreement "P" and the coefficient kappa. Hambleton and Novick (1973) proposed P as a measure of classification consistency, where P is defined as the sum of diagonal values of the contingency table:

$$P = P_{11} + P_{22} + P_{33} + P_{44}.$$

**Table 8-A. Example Contingency Table with Three Cut Scores**

|  | Level 1 | Level 2 | Level 3 | Level 4 | Sum |
|---|---|---|---|---|---|
| Level 1 | $P_{11}$ | $P_{21}$ | $P_{31}$ | $P_{41}$ | $P_{.1}$ |
| Level 2 | $P_{12}$ | $P_{22}$ | $P_{32}$ | $P_{42}$ | $P_{.2}$ |
| Level 3 | $P_{13}$ | $P_{23}$ | $P_{33}$ | $P_{43}$ | $P_{.3}$ |
| Level 4 | $P_{14}$ | $P_{24}$ | $P_{34}$ | $P_{44}$ | $P_{.4}$ |
| Sum | $P_{1.}$ | $P_{2.}$ | $P_{3.}$ | $P_{4.}$ | 1.0 |

To reflect statistical chance agreement, Swaminathan, Hambleton, and Algina (1974) suggest using Cohen's kappa (1960) as

$$\text{kappa} = \frac{P - P_c}{1 - P_c},$$

where $P_c$ is the chance probability of a consistent classification under two completely random assignments. Probability $P_c$ is the sum of the probabilities obtained by multiplying the marginal probability of the first administration and the corresponding marginal probability of the second administration as

$$P_c = (P_{1.} \times P_{.1}) + (P_{2.} \times P_{.2}) + (P_{3.} \times P_{.3}) + (P_{4.} \times P_{.4}).$$

To aid in the interpretation of the kappa statistic, the following cutoffs have been suggested (Landis & Koch, 1977; Altman, 1991):

| Kappa Value | Strength of Agreement |
|---|---|
| 0 | None |
| <0.20 | Poor |
| 0.21–0.40 | Fair |
| 0.41–0.60 | Moderate |
| 0.61–0.80 | Good |
| 0.81–1.00 | Very Good |

While classification consistency refers to the agreement between two observed scores, classification accuracy refers to the agreement between the observed score and the true score. Classification accuracy is defined as the extent to which the actual classifications of test takers agree with the classifications that would be made on the basis of their true scores (Livingston & Lewis, 1995). It is common to estimate classification accuracy by assuming the psychometric model to find true scores that correspond to observed scores. For the Wisconsin Forward Exam, the method used to estimate classification accuracy and consistency is the Kolen and Kim method (2004), which is described in the next section of this report (see also Kim, Choi, Um, & Kim, 2006; Kim, Barton, & Kim, 2007).

### 8.2.1 Kolen and Kim's Method for Pattern Scoring

As stated in Part 6, when IRT is applied to score examinees' responses, two types of scoring are available: number correct scoring and item pattern scoring. The Wisconsin Forward Exam uses item-pattern scoring. Many methods of estimating the consistency and accuracy of classification based on number-correct scoring have been suggested in psychometric literature. However, there have been relatively few studies dealing with item-pattern scoring based on IRT. Kolen and Kim (2004) suggest a simple procedure for pattern scoring (KKM) based on IRT and simulated item responses. The procedure is described below and was implemented with KKCLASS software (Kim, 2005):

Step 1: Obtain item parameters (I) and the ability distribution weight ($\hat{g}(\theta)$) at each quadrature point.

Step 2: Compute two ability estimates at each quadrature point. At a given quadrature point, $\theta_j$, generate two sets of item responses using the item parameters from a test form, assuming that the same test form was administered twice to an examinee with the true ability $\theta_j$.

$$\theta_j \longrightarrow (1,1,0,0, \ldots: \text{Item response from the first administration, or Form 1}) \longrightarrow \hat{\theta}_{j1}$$

$$\theta_j \longrightarrow (0,1,1,0, \ldots: \text{Item response from the second administration, or Form 2}) \longrightarrow \hat{\theta}_{j2}$$

If two parallel (or alternative) forms (e.g., Form 1 and Form 2) are available, the two response patterns can be generated based on the item parameters from the two forms.

Step 3: Construct a classification matrix at each quadrature point. Determine the joint event for the cells in Table 8-B by using the two ability estimates obtained in Step 2. Note that this table is constructed for each quadrature point and replication. One, and only one, cell will have a value of one and zeros elsewhere.

**Table 8-B. Example Classification Table for One Cut Point (C1)**

First Administration, or Form 1

| | $\hat{\theta}_{j1} \geq C_1$ | $\hat{\theta}_{j1} < C_1$ | |
|---|---|---|---|
| $\hat{\theta}_{j2} \geq C_1$ | | | Second Administration, |
| $\hat{\theta}_{j2} < C_1$ | | | or Form 2 |

Step 4: Repeat Steps 2 and 3 $R$ times and get average values over $R$ replications. $R$ should be a large number (e.g., 500) to obtain stable results.

Step 5: Multiply distribution weight ($\hat{g}(\theta)$) by the average values in Step 4 for each quadrature point and sum across all quadrature points. From this, a final contingency table and classification consistency indices, such as kappa, can be computed.

Because the examinees' abilities are estimated at each quadrature point, these quadrature points can be considered the true scores. Therefore, classification accuracy is computed using the examinees' estimated abilities (observed scores) and quadrature points (true scores). Just as 0.90 is generally considered the criterion for acceptable test score reliability, the criterion value of 0.90 is considered to be an acceptably high level of classification accuracy.

In Tables 8-5 through 8-21, there are two tables for each grade and content area. The first table is a contingency table with all three cut scores, which was prepared based on the KKM procedure. The rows represent the first administration of an assessment, and the columns represent the second administration of the same assessment to the same students. As mentioned above, in the KKM procedure, the score distributions for the first administration and the second administration are estimated using a simulation. Therefore, the value in each cell represents the probability of belonging to a particular pair of performance levels in the first administration and the second administration. For example, when considering the first column of data in the ELA grade 3 table, 0.23 represents the probability of belonging in *Below Basic* in both the first and second administrations. The 0.06 value represents the probability of belonging in *Basic* in the first administration and *Below Basic* in the second administration. The probability of belonging in *Proficient* or *Advanced* in the first administration and *Below Basic* in the second administration is 0.00. "Sum" is obtained simply by adding the four row values or the four column values. Because the values displayed have been rounded to two decimal places, this sum is not always identical to the sum of the values shown in the table.

The second table shows indices for classification consistency and classification accuracy. Because there are four performance levels for the Wisconsin Forward Exam, there are three cut scores. The values in "All Cuts" were obtained by applying all three cuts together. In Table 8-5 for ELA grade 3, when all three cuts were used for the computation, classification consistency (P) is 0.74, probability of chance is 0.30, kappa ($k$) is 0.62, and classification accuracy is 0.81. The values for "Cut 1" were obtained by applying only the first cut score. There are two levels whenever only one cut is applied (i.e., performance levels above and below the cut). It is clear that the values for P, $k$, and classification accuracy with all three cuts are smaller than those for any single cut point. The probability of assigning students to the incorrect performance level will increase with the number of cut scores.

Because the *Proficient* cut score is a criterion for accountability reports, the reliability values for this second cut need to be considered carefully. In Table 8-5, for example, the P for the second cut, which establishes the *Proficient* performance level, was 0.89, kappa was 0.77, and classification accuracy was 0.92. The interpretation of the values illustrated for Table 8-5 is the same for Tables 8-6 through 8-21.

As shown in Tables 8-5 through 8-21, when only the *Proficient* cut score was applied, the classification consistency (P) was greater than or equal to 0.87 and the classification accuracy was greater than or equal to 0.91 for all tests. The kappa value was greater than or equal to 0.68 for all tests. According to criteria for kappa ($k$) values (presented earlier in this section of the report in the discussion of classification consistency), all tests showed good or very good agreement based on the cut for the *Proficient* performance level.

In addition, the indices for classification consistency and classification accuracy were computed for the subgroups of students. These data are presented in Appendix J. As seen in Tables J-1 through J-17, when the *Proficient* cut is considered, classification consistency, accuracy coefficients, and kappa values were good or very good for all subgroups, grades, and content areas. Specifically, for ELA, the classification consistency was greater than or equal to 0.88 and the classification accuracy was greater than or equal to 0.91 for all subgroups across all grades. For Mathematics, the classification consistency was greater than or equal to 0.89 and the classification accuracy was greater than or equal to 0.92 for all subgroups across all grades. For Science, the classification consistency was greater than or equal to 0.86 and the classification accuracy was greater than or equal to 0.90 for all subgroups across both grades. For Social Studies, the classification consistency was greater than or equal to 0.87 and the classification accuracy was greater than or equal to 0.91 for all subgroups across all grades.

The kappa values indicated good or very good agreement based on the cut for the *Proficient* performance level for all subgroups across all grades and content areas. The kappa values were greater than or equal to 0.65 for all subgroups in ELA, greater than or equal to 0.62 for all subgroups in Mathematics, greater than or equal to 0.62 for all subgroups in Science, and greater than or equal to 0.65 for all subgroups in Social Studies. Somewhat lower kappa values were observed for the limited English proficiency subgroups in all content areas and for students using testing accommodations in ELA and Mathematics compared to other subgroups. This is consistent with the trend of the test reliability coefficients, which were found to be lower for these groups of students compared to other subgroups. Because the number of students using testing accommodations in Science and Social Studies was less than 50 per grade, the indices for classification consistency and classification accuracy were not computed for students using testing accommodations in these subject areas. The indices for classification consistency and classification accuracy for students using testing accommodations in ELA should be interpreted with caution because of the low number of students using accommodations in this content area.

## 8.3 Inter-rater Reliability for TDA Items

The reliability of scoring of TDA items was measured in two ways: (1) tabulations of exact and adjacent agreement of two scorers and (2) reliability coefficients. Reliability for TDA items was examined by calculating indices of inter-rater agreement, which is the degree of reliability with which the AI engine and a human scorer assign scores to a given student response. Two indices for inter-rater reliability, intraclass correlation and weighted kappa, are presented here.

Notation: To assess reliability, it is necessary to replicate the scoring process for a subset of papers. This is usually done with "blind double-reads." Suppose that there are $N$ responses, each of which is scored twice. The two scores of response $n$ are denoted by $X_{n1}$ and $X_{n2}$, where $n = 1, 2, \ldots N$. The resulting data may be presented in two ways: enumeration by response and cross tabulation. Table 8-C shows the enumeration by response data structure, where each row represents a single student response.

**Table 8-C. Data Structure 1: Enumeration by Response**

| Response # | Score 1 | Score 2 | Mean Score |
|---|---|---|---|
| 1 | $X_{11}$ | $X_{12}$ | $\bar{X}_{1.}$ |
| 2 | $X_{21}$ | $X_{22}$ | $\bar{X}_{2.}$ |
| . | . | . | . |
| . | . | . | . |
| $N$ | $X_{N1}$ | $X_{N2}$ | $\bar{X}_{N.}$ |
| Column Mean | $\bar{X}_{.1}$ | $\bar{X}_{.2}$ | $\bar{X}_{..}$ |

where

$$\bar{X}_{1.} = (X_{11} + X_{12})/2$$

is the mean score for Response 1 (similarly for responses 2, 3, … $N$),

$$\bar{X}_{.1} = \frac{1}{N}\sum_{n=1}^{N} X_{.1} = (X_{11} + X_{21} + \ldots + X_{N1})/N$$

is the mean of Score 1 over all responses (similarly for Score 2), and

$$\bar{X}_{..} = \frac{1}{N}\sum_{n=1}^{N} (X_{n1} + X_{n2})/2$$

is the overall mean score across both scores of all responses.

As an alternative, a square table of counts may be created for each Score 1 by Score 2 (i.e., $X_{n1} \times X_{n2}$) combination. An example of this data structure is presented in Table 8-D.

**Table 8-D. Data Structure 2: Cross-Tabulation of Score 1 and Score 2**

| | | Score 2 | | | | Row Total |
|---|---|---|---|---|---|---|
| | | **0** | **1** | **...** | **m** | |
| | 0 | $n_{00}$ | $n_{01}$ | ... | $n_{0m}$ | $n_{0+}$ |
| | 1 | $n_{10}$ | $n_{11}$ | ... | $n_{1m}$ | $n_{1+}$ |
| Score 1 | . | . | . | ... | . | . |
| | . | . | . | ... | . | . |
| | m | $n_{m0}$ | $n_{m1}$ | ... | $n_{mm}$ | $n_{m+}$ |
| Column Total | | $n_{+0}$ | $n_{+1}$ | ... | $n_{+m}$ | $n_{++}$ |

where $m$ is the maximum score (for a rubric including zero) obtainable for an item, $n_{ij}$ is the number of responses for which Score 1 = $i$ and Score 2 = $j$, $n_{i+}$ is the number of responses for which Score 1 = $i$, and $n_{+j}$ is the number of responses for which Score 2 = $j$.

Formulas for the two reliability coefficients of interest are then given:

1. Intraclass correlation, $\rho_{IC}$, describes the percentage of overall score variance accounted for by the variance of mean response scores:

$$\rho_{IC} = \frac{Var_n(\bar{X}_n.)}{Var_n(X_{n1}, X_{n2})} = \frac{\frac{1}{N-1}\sum_{n=1}^{N}(\bar{X}_n. - \bar{X}..)^2}{\frac{1}{2(N-1)}\sum_{n=1}^{N}[(X_{n1} - \bar{X}..)^2 + (X_{n2} - \bar{X}..)^2]}$$

If agreement is perfect, $\rho_{IC} = 1$. The following is always true: $0 \leq \rho_{IC} \leq 1$.

2. Weighted kappa, $k$, is used in many contexts as a measure of association in square contingency tables:

$$k = \frac{\sum_{i=0}^{m}\sum_{j=0}^{m}w_{ij}\frac{n_{ij}}{n_{++}} - \sum_{i=0}^{m}\sum_{j=0}^{m}w_{ij}\frac{n_{i+}n_{+j}}{n^2_{++}}}{1 - \sum_{i=0}^{m}\sum_{j=0}^{m}w_{ij}\frac{n_{i+}n_{+j}}{n^2_{++}}}, \text{ where } w_{ij} = 1 - \frac{(i-j)^2}{M^2}.$$

If agreement is perfect, $k = 1$. If agreement is what would be expected by chance, $k = 0$. The following is always true: $0 \leq k \leq 1$.

Ordinal rating scales (e.g., 1, 2, 3, 4) used in scoring TDA items contain a certain level of chance agreement that is expected. Although the intraclass correlation is reported in this report, it does not take into account the possibility of chance agreement between the two raters. Cohen's kappa does take this into consideration. In general, $k$ will have values equal to or less than the intraclass correlation. If agreement is perfect, the value of $k$ is 1.0. If agreement is at chance levels, the value of $k$ is 0. As noted in Section 8.2, values of $k$ greater than 0.81 indicate "very good agreement," values between 0.61 and 0.80 represent "good agreement," values between 0.41 and 0.60 represent "moderate agreement," values between 0.21 and 0.40 represent "fair agreement" beyond chance, and values below 0.20 denote "poor agreement." Specific criteria for intraclass correlation or weighted $k$ are not established.

Table 8-22 presents the rater agreement statistics for TDA items. The evidence supporting inter-rater reliability is presented in terms of the percentage of agreement between raters (the AI engine and a human rater), two indices of inter-rater reliability, and the distributions of scores across score levels. In the table, "Exact" agreement is defined as scores that are exactly the same. "Adjacent" agreement is defined as scores differing by 1 point. "Discrepant" cases are those cases in which the scores of the two raters differed by more than one raw score point. For example, as shown in Table 8-22, for the grade 3 TDA item, the exact agreement, adjacent agreement, and discrepant agreement rates are 91.71%, 8.14%, and 0.16%, respectively. "Mean" reflects the item mean score from the second reads, which are done by human scorers. "No. of Second Reads" is the number of student responses selected for the purpose of the second read and computing inter-rater reliability. The "Score Frequency" columns represent the scoring outcomes for the student responses based on the raw scores given by the human scorers. The column for "All Codes" reflects the number of students who received the condition codes B, C, N, R, or T (described in detail in Part 5, Table 5-2 of this report).

Overall, the exact rater agreement percentages were high for all TDA items and ranged from 84.72% in grade 6 to 91.71% in grade 3. The combined exact and adjacent agreement percentages were over 99% in all grades. The intraclass correlation coefficients ranged from 0.90 in grade 3 to 0.94 in grade 8. The weighted kappa ranged from 0.80 in grade 3 to 0.88 in grade 8, indicating very good rater agreement for all TDA items.

## 8.4 Summary

Overall, the analyses discussed in this section of the report indicated acceptable levels of reliability for the Wisconsin Forward Exam. The internal consistency reliability estimates, as measured by Cronbach's alpha coefficient, were reasonable given the number of items in each test. The analyses of classification consistency and accuracy indicated acceptable levels of consistency and accuracy of student proficiency level classifications, and the SEM around the *Proficient* cut score was low in every grade and content area. The levels of rater agreement were high, and the discrepancy rates were low, with acceptably high values for the weighted kappa and intraclass correlations. The results of the inter-rater reliability analyses indicated an acceptable degree of reliability for scores on the ELA TDA items in the Wisconsin Forward Exam.

**Table 8-1. Cronbach's Alpha Reliability Coefficients for Total Group and Subgroups**

| Category | English Language Arts | | | | | | Mathematics | | | | | | Science | | Social Studies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GR 3 | GR 4 | GR 5 | GR 6 | GR 7 | GR 8 | GR 3 | GR 4 | GR 5 | GR 6 | GR 7 | GR 8 | GR 4 | GR 8 | GR 4 | GR 8 | GR 10 |
| **Total** | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.89 | 0.93 | 0.93 | 0.92 | 0.93 | 0.92 | 0.92 | 0.89 | 0.88 | 0.91 | 0.91 | 0.90 |
| **Gender** | | | | | | | | | | | | | | | | | |
| Female | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.88 | 0.93 | 0.92 | 0.92 | 0.92 | 0.91 | 0.91 | 0.89 | 0.87 | 0.90 | 0.90 | 0.89 |
| Male | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.89 | 0.93 | 0.93 | 0.93 | 0.93 | 0.92 | 0.92 | 0.90 | 0.89 | 0.91 | 0.92 | 0.91 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | | | |
| White | 0.87 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.92 | 0.92 | 0.91 | 0.92 | 0.91 | 0.91 | 0.87 | 0.87 | 0.89 | 0.90 | 0.90 |
| African American | 0.83 | 0.85 | 0.83 | 0.87 | 0.86 | 0.85 | 0.89 | 0.84 | 0.85 | 0.84 | 0.82 | 0.81 | 0.83 | 0.78 | 0.87 | 0.87 | 0.86 |
| Hispanic | 0.85 | 0.87 | 0.87 | 0.88 | 0.87 | 0.87 | 0.92 | 0.90 | 0.90 | 0.89 | 0.88 | 0.88 | 0.87 | 0.85 | 0.89 | 0.90 | 0.87 |
| Asian | 0.88 | 0.90 | 0.90 | 0.89 | 0.89 | 0.89 | 0.94 | 0.93 | 0.93 | 0.94 | 0.93 | 0.93 | 0.88 | 0.88 | 0.91 | 0.91 | 0.89 |
| American Indian | 0.84 | 0.86 | 0.87 | 0.87 | 0.86 | 0.86 | 0.90 | 0.88 | 0.90 | 0.88 | 0.86 | 0.86 | 0.85 | 0.84 | 0.88 | 0.88 | 0.86 |
| Two or More | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.88 | 0.92 | 0.92 | 0.92 | 0.92 | 0.91 | 0.91 | 0.89 | 0.88 | 0.91 | 0.91 | 0.90 |
| **ELP** | | | | | | | | | | | | | | | | | |
| Fully English Proficient | 0.88 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.93 | 0.93 | 0.92 | 0.93 | 0.92 | 0.92 | 0.89 | 0.88 | 0.91 | 0.91 | 0.90 |
| Limited English Proficient | 0.83 | 0.84 | 0.82 | 0.84 | 0.81 | 0.82 | 0.91 | 0.88 | 0.85 | 0.83 | 0.80 | 0.78 | 0.83 | 0.75 | 0.87 | 0.84 | 0.77 |
| **Disability** | | | | | | | | | | | | | | | | | |
| Disabled | 0.86 | 0.87 | 0.85 | 0.86 | 0.84 | 0.84 | 0.93 | 0.90 | 0.90 | 0.88 | 0.85 | 0.84 | 0.88 | 0.85 | 0.90 | 0.89 | 0.85 |
| Not Disabled | 0.88 | 0.89 | 0.89 | 0.89 | 0.88 | 0.88 | 0.93 | 0.92 | 0.92 | 0.93 | 0.91 | 0.92 | 0.89 | 0.88 | 0.90 | 0.90 | 0.89 |
| **Economic Status** | | | | | | | | | | | | | | | | | |
| Economically Disadvantaged | 0.86 | 0.88 | 0.87 | 0.88 | 0.87 | 0.87 | 0.92 | 0.90 | 0.90 | 0.90 | 0.88 | 0.88 | 0.88 | 0.86 | 0.89 | 0.90 | 0.88 |
| Not Economically Disadvantaged | 0.87 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.92 | 0.92 | 0.92 | 0.93 | 0.92 | 0.92 | 0.88 | 0.87 | 0.89 | 0.90 | 0.90 |
| **Accommodations** | | | | | | | | | | | | | | | | | |
| Students with Accommodations | 0.87 | 0.91 | 0.82 | 0.88 | 0.89 | 0.89 | 0.86 | 0.76 | 0.77 | 0.67 | 0.63 | 0.59 | - | - | - | - | - |
| Students without Accommodations | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.89 | 0.93 | 0.92 | 0.92 | 0.93 | 0.92 | 0.92 | 0.89 | 0.88 | 0.91 | 0.91 | 0.90 |

Note: GR = Grade. The reliability coefficients were not computed for students using testing accommodations in Science or Social Studies because the number of students using testing accommodations in these subject areas was less than 50 per grade.

**Table 8-2. Standard Error of Measurement for Total Group and Subgroups**

| Category | English Language Arts | | | | | | Mathematics | | | | | | Science | | Social Studies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GR 3 | GR 4 | GR 5 | GR 6 | GR 7 | GR 8 | GR 3 | GR 4 | GR 5 | GR 6 | GR 7 | GR 8 | GR 4 | GR 8 | GR 4 | GR 8 | GR 10 |
| **Total** | 3.33 | 3.29 | 3.28 | 3.33 | 3.50 | 3.48 | 2.68 | 2.87 | 2.84 | 2.84 | 2.90 | 2.88 | 2.72 | 2.80 | 2.68 | 2.68 | 2.77 |
| **Gender** | | | | | | | | | | | | | | | | | |
| Female | 3.31 | 3.29 | 3.26 | 3.31 | 3.50 | 3.46 | 2.70 | 2.88 | 2.85 | 2.85 | 2.90 | 2.89 | 2.74 | 2.82 | 2.69 | 2.69 | 2.78 |
| Male | 3.34 | 3.28 | 3.29 | 3.34 | 3.49 | 3.48 | 2.65 | 2.85 | 2.83 | 2.83 | 2.90 | 2.87 | 2.71 | 2.78 | 2.67 | 2.66 | 2.75 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | | | |
| White | 3.28 | 3.28 | 3.24 | 3.29 | 3.47 | 3.43 | 2.67 | 2.87 | 2.87 | 2.87 | 2.94 | 2.91 | 2.71 | 2.78 | 2.64 | 2.63 | 2.73 |
| African American | 3.32 | 3.23 | 3.35 | 3.39 | 3.52 | 3.49 | 2.62 | 2.78 | 2.65 | 2.66 | 2.69 | 2.73 | 2.69 | 2.80 | 2.74 | 2.81 | 2.78 |
| Hispanic | 3.37 | 3.29 | 3.34 | 3.38 | 3.53 | 3.51 | 2.72 | 2.87 | 2.80 | 2.79 | 2.84 | 2.83 | 2.75 | 2.82 | 2.77 | 2.77 | 2.82 |
| Asian | 3.35 | 3.29 | 3.27 | 3.32 | 3.46 | 3.44 | 2.66 | 2.84 | 2.82 | 2.81 | 2.87 | 2.84 | 2.75 | 2.76 | 2.70 | 2.64 | 2.76 |
| American Indian | 3.40 | 3.31 | 3.31 | 3.38 | 3.51 | 3.46 | 2.74 | 2.90 | 2.81 | 2.73 | 2.79 | 2.85 | 2.76 | 2.85 | 2.79 | 2.80 | 2.80 |
| Two or More | 3.34 | 3.29 | 3.29 | 3.35 | 3.51 | 3.53 | 2.71 | 2.88 | 2.83 | 2.83 | 2.86 | 2.86 | 2.73 | 2.81 | 2.70 | 2.70 | 2.76 |
| **ELP** | | | | | | | | | | | | | | | | | |
| Fully English Proficient | 3.32 | 3.29 | 3.28 | 3.33 | 3.50 | 3.47 | 2.68 | 2.87 | 2.85 | 2.85 | 2.91 | 2.89 | 2.72 | 2.80 | 2.67 | 2.67 | 2.76 |
| Limited English Proficient | 3.37 | 3.28 | 3.34 | 3.38 | 3.51 | 3.48 | 2.72 | 2.87 | 2.77 | 2.71 | 2.73 | 2.77 | 2.75 | 2.80 | 2.79 | 2.83 | 2.80 |
| **Disability** | | | | | | | | | | | | | | | | | |
| Disabled | 3.35 | 3.25 | 3.32 | 3.34 | 3.42 | 3.40 | 2.66 | 2.84 | 2.71 | 2.68 | 2.71 | 2.73 | 2.72 | 2.77 | 2.74 | 2.76 | 2.75 |
| Not Disabled | 3.31 | 3.28 | 3.26 | 3.31 | 3.49 | 3.46 | 2.68 | 2.87 | 2.86 | 2.86 | 2.92 | 2.90 | 2.72 | 2.80 | 2.67 | 2.67 | 2.77 |
| **Economic Status** | | | | | | | | | | | | | | | | | |
| Economically Disadvantaged | 3.37 | 3.28 | 3.33 | 3.38 | 3.53 | 3.51 | 2.72 | 2.88 | 2.81 | 2.80 | 2.84 | 2.84 | 2.75 | 2.83 | 2.76 | 2.77 | 2.81 |
| Not Economically Disadvantaged | 3.27 | 3.28 | 3.23 | 3.27 | 3.46 | 3.42 | 2.65 | 2.85 | 2.86 | 2.87 | 2.93 | 2.90 | 2.70 | 2.78 | 2.62 | 2.61 | 2.73 |
| **Accommodations** | | | | | | | | | | | | | | | | | |
| Students with Accommodations | 3.30 | 3.23 | 3.44 | 3.47 | 3.42 | 3.50 | 2.53 | 2.76 | 2.58 | 2.56 | 2.59 | 2.66 | - | - | - | - | - |
| Students without Accommodations | 3.33 | 3.29 | 3.28 | 3.33 | 3.50 | 3.48 | 2.68 | 2.87 | 2.85 | 2.85 | 2.91 | 2.89 | 2.72 | 2.80 | 2.68 | 2.68 | 2.77 |

Note: GR = Grade. The SEMs were not computed for students using testing accommodations in Science or Social Studies because the number of students using testing accommodations in these subject areas was less than 50 per grade.

# Table 8-3. Cronbach's Alpha Reliability Coefficients for Content Standards and Domains

## English Language Arts

| Grade | Alpha per Content Standard and Domain | | | | | | | | | |
|-------|------|------|------|------|------|------|-----------|---------|---------|------------|
|       | A | B | C | D | E | F | Listening | Reading | Writing | Total Test |
| 3 | 0.71 | 0.47 | 0.47 | 0.42 | 0.54 | 0.48 | 0.49 | 0.80 | 0.73 | 0.88 |
| 4 | 0.66 | 0.59 | 0.65 | 0.45 | 0.58 | 0.49 | 0.48 | 0.83 | 0.75 | 0.89 |
| 5 | 0.70 | 0.61 | 0.28 | 0.40 | 0.66 | 0.56 | 0.58 | 0.81 | 0.76 | 0.89 |
| 6 | 0.71 | 0.67 | 0.52 | 0.45 | 0.35 | 0.46 | 0.60 | 0.85 | 0.69 | 0.90 |
| 7 | 0.64 | 0.55 | 0.60 | 0.46 | 0.48 | 0.43 | 0.63 | 0.81 | 0.71 | 0.89 |
| 8 | 0.61 | 0.57 | 0.62 | 0.48 | 0.49 | 0.37 | 0.59 | 0.81 | 0.72 | 0.89 |

ELA standards: A=Reading—Key Ideas and Details; B=Reading—Craft & Structure/Integration of Knowledge & Ideas; C=Reading—Vocabulary Use; D=Writing/Language—Text Types and Purposes; E=Writing/Language—Research; F=Writing/Language—Language Conventions

## Mathematics

| Grade | Alpha per Content Standard | | | | | | | | | | Total Test |
|-------|------|------|------|------|------|------|------|------|------|------|------------|
|       | A | B | C | D | E | F | G | H | I | J | |
| 3 | 0.76 | 0.77 | 0.76 | 0.74 | 0.66 | | | | | | 0.93 |
| 4 | 0.73 | 0.73 | 0.81 | 0.66 | 0.68 | | | | | | 0.93 |
| 5 | 0.70 | 0.74 | 0.72 | 0.72 | 0.72 | | | | | | 0.92 |
| 6 | | | | | 0.71 | 0.68 | 0.75 | 0.80 | 0.70 | | 0.93 |
| 7 | | | | | 0.56 | 0.76 | 0.74 | 0.64 | 0.76 | | 0.92 |
| 8 | | | | | 0.70 | | 0.78 | 0.68 | 0.63 | 0.75 | 0.92 |

Mathematics standards: A=Operations and Algebraic Thinking; B=Number and Operations in Base Ten; C=Number and Operations—Fractions; D=Measurement and Data; E=Geometry; F=Ratios and Proportional Relationships; G=The Number System; H=Expressions and Equations; I=Statistics and Probability; J=Functions

## Science

| Grade | Alpha per Content Standard | | | | Total Test |
|-------|------|------|------|------|------------|
|       | A | B | C | D | |
| 4 | 0.64 | 0.73 | 0.58 | 0.74 | 0.89 |
| 8 | 0.66 | 0.68 | 0.59 | 0.68 | 0.88 |

Science standards: A=Life Science; B=Physical Science; C=Earth and Space Science; D=Engineering.

## Social Studies

| Grade | Alpha per Content Standard | | | | | Total Test |
|-------|------|------|------|------|------|------------|
|       | A | B | C | D | E | |
| 4 | 0.79 | 0.65 | 0.53 | 0.55 | 0.71 | 0.91 |
| 8 | 0.71 | 0.64 | 0.74 | 0.62 | 0.64 | 0.91 |
| 10 | 0.69 | 0.67 | 0.70 | 0.60 | 0.47 | 0.90 |

Social Studies standards: A=Geography; B=History; C=Political Science; D=Economics; E= Behavioral Sciences

# Table 8-4. Standard Error of Measurement per Content Standards and Domains

## English Language Arts

| Grade | SEM per Content Standard and Domain | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | Listening | Reading | Writing | Total Test |
| 3 | 1.69 | 1.23 | 0.86 | 1.05 | 1.03 | 1.11 | 1.24 | 2.29 | 1.85 | 3.33 |
| 4 | 1.54 | 1.20 | 0.97 | 0.99 | 1.04 | 1.01 | 1.42 | 2.19 | 1.76 | 3.29 |
| 5 | 1.62 | 1.28 | 0.87 | 1.03 | 0.97 | 1.02 | 1.27 | 2.24 | 1.78 | 3.28 |
| 6 | 1.45 | 1.38 | 0.81 | 1.02 | 1.11 | 1.14 | 1.30 | 2.16 | 1.89 | 3.33 |
| 7 | 1.54 | 1.47 | 0.88 | 1.08 | 1.27 | 0.99 | 1.32 | 2.30 | 1.95 | 3.50 |
| 8 | 1.52 | 1.27 | 0.94 | 1.19 | 1.26 | 1.04 | 1.31 | 2.22 | 2.02 | 3.48 |

ELA standards: A=Reading—Key Ideas and Details; B=Reading—Craft & Structure/Integration of Knowledge & Ideas; C=Reading—Vocabulary Use; D=Writing/Language—Text Types and Purposes; E=Writing/Language—Research; F=Writing/Language—Language Conventions

## Mathematics

| Grade | SEM per Content Standard | | | | | | | | | | Total Test |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J | |
| 3 | 1.21 | 1.14 | 1.14 | 1.34 | 1.09 | | | | | | 2.68 |
| 4 | 1.33 | 1.25 | 1.28 | 1.37 | 1.08 | | | | | | 2.87 |
| 5 | 1.25 | 1.24 | 1.24 | 1.29 | 1.28 | | | | | | 2.84 |
| 6 | | | | | 1.08 | 1.05 | 1.38 | 1.37 | 1.37 | | 2.84 |
| 7 | | | | | 1.40 | 1.14 | 1.15 | 1.33 | 1.40 | | 2.90 |
| 8 | | | | | 1.35 | | 1.12 | 1.38 | 1.20 | 1.30 | 2.88 |

Mathematics standards: A=Operations and Algebraic Thinking; B=Number and Operations in Base Ten; C=Number and Operations—Fractions; D=Measurement and Data; E=Geometry; F=Ratios and Proportional Relationships; G=The Number System; H=Expressions and Equations; I=Statistics and Probability; J=Functions

## Science

| Grade | SEM per Content Standard | | | | Total Test |
|---|---|---|---|---|---|
| | A | B | C | D | |
| 4 | 1.38 | 1.48 | 1.36 | 1.19 | 2.72 |
| 8 | 1.40 | 1.51 | 1.47 | 1.19 | 2.80 |

Science standards: A=Life Science; B=Physical Science; C=Earth and Space Science; D=Engineering

## Social Studies

| Grade | SEM per Content Standard | | | | | Total Test |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | |
| 4 | 1.36 | 1.21 | 1.15 | 1.18 | 1.05 | 2.68 |
| 8 | 1.27 | 1.30 | 1.15 | 1.22 | 1.01 | 2.68 |
| 10 | 1.37 | 1.29 | 1.30 | 1.08 | 1.10 | 2.77 |

Social Studies standards: A=Geography; B=History; C=Political Science; D=Economics; E= Behavioral Sciences

**Table 8-5. Classification Consistency and Classification Accuracy for English Language Arts Grade 3**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.23 | 0.05 | 0.00 | 0.00 | 0.28 |
| *Basic* | 0.06 | 0.26 | 0.05 | 0.00 | 0.37 |
| *Proficient* | 0.00 | 0.05 | 0.20 | 0.02 | 0.28 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.04 | 0.07 |
| **Sum** | 0.28 | 0.37 | 0.28 | 0.07 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.89 | 0.89 | 0.95 | 0.74 |
| **Probability of Chance** | 0.60 | 0.55 | 0.87 | 0.30 |
| **Kappa ($k$)** | 0.74 | 0.77 | 0.59 | 0.62 |
| **Classification Accuracy** | 0.93 | 0.92 | 0.96 | 0.81 |

**Table 8-6. Classification Consistency and Classification Accuracy for English Language Arts Grade 4**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.22 | 0.05 | 0.00 | 0.00 | 0.27 |
| *Basic* | 0.05 | 0.21 | 0.05 | 0.00 | 0.31 |
| *Proficient* | 0.00 | 0.06 | 0.23 | 0.03 | 0.32 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.06 | 0.10 |
| **Sum** | 0.28 | 0.31 | 0.32 | 0.09 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.90 | 0.89 | 0.94 | 0.73 |
| **Probability of Chance** | 0.60 | 0.51 | 0.83 | 0.28 |
| **Kappa ($k$)** | 0.74 | 0.77 | 0.63 | 0.62 |
| **Classification Accuracy** | 0.93 | 0.92 | 0.95 | 0.80 |

**Table 8-7. Classification Consistency and Classification Accuracy for English Language Arts Grade 5**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.23 | 0.05 | 0.00 | 0.00 | 0.28 |
| *Basic* | 0.05 | 0.20 | 0.05 | 0.00 | 0.30 |
| *Proficient* | 0.00 | 0.05 | 0.26 | 0.03 | 0.34 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.05 | 0.08 |
| **Sum** | 0.28 | 0.31 | 0.34 | 0.07 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.90 | 0.89 | 0.95 | 0.74 |
| **Probability of Chance** | 0.59 | 0.52 | 0.86 | 0.29 |
| **Kappa ($k$)** | 0.75 | 0.78 | 0.61 | 0.63 |
| **Classification Accuracy** | 0.93 | 0.93 | 0.97 | 0.82 |

**Table 8-8. Classification Consistency and Classification Accuracy for English Language Arts Grade 6**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.22 | 0.05 | 0.00 | 0.00 | 0.26 |
| *Basic* | 0.04 | 0.25 | 0.06 | 0.00 | 0.35 |
| *Proficient* | 0.00 | 0.06 | 0.20 | 0.03 | 0.29 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.06 | 0.09 |
| **Sum** | 0.26 | 0.36 | 0.29 | 0.09 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.91 | 0.89 | 0.94 | 0.74 |
| **Probability of Chance** | 0.61 | 0.53 | 0.84 | 0.29 |
| **Kappa ($k$)** | 0.77 | 0.76 | 0.62 | 0.63 |
| **Classification Accuracy** | 0.94 | 0.92 | 0.95 | 0.81 |

**Table 8-9. Classification Consistency and Classification Accuracy for English Language Arts Grade 7**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.22 | 0.05 | 0.00 | 0.00 | 0.27 |
| *Basic* | 0.05 | 0.23 | 0.06 | 0.00 | 0.34 |
| *Proficient* | 0.00 | 0.06 | 0.21 | 0.03 | 0.30 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.06 | 0.09 |
| Sum | 0.27 | 0.34 | 0.30 | 0.09 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.91 | 0.88 | 0.94 | 0.73 |
| Probability of Chance | 0.61 | 0.52 | 0.83 | 0.29 |
| Kappa (*k*) | 0.76 | 0.76 | 0.61 | 0.62 |
| Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.80 |

**Table 8-10. Classification Consistency and Classification Accuracy for English Language Arts Grade 8**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.25 | 0.05 | 0.00 | 0.00 | 0.30 |
| *Basic* | 0.05 | 0.26 | 0.05 | 0.00 | 0.36 |
| *Proficient* | 0.00 | 0.05 | 0.18 | 0.03 | 0.26 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.05 | 0.08 |
| Sum | 0.29 | 0.37 | 0.26 | 0.08 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.91 | 0.89 | 0.94 | 0.74 |
| Probability of Chance | 0.58 | 0.55 | 0.85 | 0.29 |
| Kappa (*k*) | 0.77 | 0.76 | 0.61 | 0.63 |
| Classification Accuracy | 0.94 | 0.92 | 0.96 | 0.82 |

**Table 8-11. Classification Consistency and Classification Accuracy for Mathematics Grade 3**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.20 | 0.04 | 0.00 | 0.00 | 0.23 |
| *Basic* | 0.03 | 0.22 | 0.05 | 0.00 | 0.30 |
| *Proficient* | 0.00 | 0.05 | 0.25 | 0.03 | 0.33 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.10 | 0.13 |
| **Sum** | 0.23 | 0.30 | 0.33 | 0.14 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.93 | 0.90 | 0.93 | 0.76 |
| **Probability of Chance** | 0.64 | 0.50 | 0.77 | 0.27 |
| **Kappa (*k*)** | 0.81 | 0.80 | 0.72 | 0.68 |
| **Classification Accuracy** | 0.95 | 0.93 | 0.95 | 0.83 |

**Table 8-12. Classification Consistency and Classification Accuracy for Mathematics Grade 4**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.19 | 0.04 | 0.00 | 0.00 | 0.23 |
| *Basic* | 0.04 | 0.25 | 0.04 | 0.00 | 0.33 |
| *Proficient* | 0.00 | 0.04 | 0.23 | 0.03 | 0.30 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.11 | 0.14 |
| **Sum** | 0.22 | 0.34 | 0.30 | 0.14 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.92 | 0.91 | 0.94 | 0.77 |
| **Probability of Chance** | 0.65 | 0.51 | 0.76 | 0.27 |
| **Kappa (*k*)** | 0.77 | 0.82 | 0.75 | 0.69 |
| **Classification Accuracy** | 0.94 | 0.94 | 0.96 | 0.84 |

**Table 8-13. Classification Consistency and Classification Accuracy for Mathematics Grade 5**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.23 | 0.05 | 0.00 | 0.00 | 0.28 |
| *Basic* | 0.04 | 0.19 | 0.04 | 0.00 | 0.27 |
| *Proficient* | 0.00 | 0.04 | 0.27 | 0.03 | 0.34 |
| *Advanced* | 0.00 | 0.00 | 0.03 | 0.09 | 0.11 |
| **Sum** | 0.27 | 0.28 | 0.33 | 0.12 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.91 | 0.92 | 0.95 | 0.77 |
| Probability of Chance | 0.60 | 0.51 | 0.80 | 0.28 |
| Kappa (*k*) | 0.77 | 0.83 | 0.73 | 0.69 |
| Classification Accuracy | 0.94 | 0.94 | 0.96 | 0.84 |

**Table 8-14. Classification Consistency and Classification Accuracy for Mathematics Grade 6**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.26 | 0.05 | 0.00 | 0.00 | 0.31 |
| *Basic* | 0.05 | 0.20 | 0.04 | 0.00 | 0.30 |
| *Proficient* | 0.00 | 0.04 | 0.27 | 0.02 | 0.33 |
| *Advanced* | 0.00 | 0.00 | 0.02 | 0.04 | 0.06 |
| **Sum** | 0.31 | 0.30 | 0.33 | 0.06 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.90 | 0.91 | 0.97 | 0.78 |
| Probability of Chance | 0.57 | 0.52 | 0.89 | 0.30 |
| Kappa (*k*) | 0.76 | 0.81 | 0.72 | 0.69 |
| Classification Accuracy | 0.93 | 0.93 | 0.98 | 0.84 |

**Table 8-15. Classification Consistency and Classification Accuracy for Mathematics Grade 7**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.33 | 0.05 | 0.00 | 0.00 | 0.38 |
| *Basic* | 0.05 | 0.19 | 0.04 | 0.00 | 0.28 |
| *Proficient* | 0.00 | 0.04 | 0.24 | 0.01 | 0.29 |
| *Advanced* | 0.00 | 0.00 | 0.01 | 0.03 | 0.05 |
| **Sum** | 0.38 | 0.28 | 0.29 | 0.05 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.90 | 0.92 | 0.97 | 0.79 |
| **Probability of Chance** | 0.53 | 0.55 | 0.91 | 0.31 |
| **Kappa ($k$)** | 0.78 | 0.81 | 0.71 | 0.69 |
| **Classification Accuracy** | 0.92 | 0.94 | 0.98 | 0.85 |

**Table 8-16. Classification Consistency and Classification Accuracy for Mathematics Grade 8**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.31 | 0.06 | 0.00 | 0.00 | 0.37 |
| *Basic* | 0.05 | 0.23 | 0.04 | 0.00 | 0.32 |
| *Proficient* | 0.00 | 0.04 | 0.19 | 0.02 | 0.24 |
| *Advanced* | 0.00 | 0.00 | 0.02 | 0.05 | 0.07 |
| **Sum** | 0.36 | 0.33 | 0.24 | 0.07 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.89 | 0.93 | 0.97 | 0.78 |
| **Probability of Chance** | 0.54 | 0.57 | 0.87 | 0.30 |
| **Kappa ($k$)** | 0.76 | 0.83 | 0.73 | 0.69 |
| **Classification Accuracy** | 0.92 | 0.95 | 0.98 | 0.84 |

**Table 8-17. Classification Consistency and Classification Accuracy for Science Grade 4**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.14 | 0.05 | 0.00 | 0.00 | 0.19 |
| *Basic* | 0.04 | 0.20 | 0.05 | 0.00 | 0.30 |
| *Proficient* | 0.00 | 0.06 | 0.20 | 0.05 | 0.31 |
| *Advanced* | 0.00 | 0.00 | 0.05 | 0.15 | 0.20 |
| **Sum** | 0.19 | 0.31 | 0.30 | 0.20 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.91 | 0.89 | 0.90 | 0.70 |
| **Probability of Chance** | 0.69 | 0.50 | 0.68 | 0.26 |
| **Kappa ($k$)** | 0.70 | 0.77 | 0.68 | 0.59 |
| **Classification Accuracy** | 0.94 | 0.92 | 0.93 | 0.78 |

**Table 8-18. Classification Consistency and Classification Accuracy for Science Grade 8**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.16 | 0.05 | 0.00 | 0.00 | 0.22 |
| *Basic* | 0.05 | 0.17 | 0.06 | 0.00 | 0.29 |
| *Proficient* | 0.00 | 0.06 | 0.18 | 0.05 | 0.29 |
| *Advanced* | 0.00 | 0.00 | 0.04 | 0.16 | 0.21 |
| **Sum** | 0.22 | 0.28 | 0.29 | 0.21 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| **Classification Consistency (P)** | 0.89 | 0.87 | 0.91 | 0.68 |
| **Probability of Chance** | 0.66 | 0.50 | 0.67 | 0.26 |
| **Kappa ($k$)** | 0.68 | 0.74 | 0.72 | 0.57 |
| **Classification Accuracy** | 0.92 | 0.91 | 0.93 | 0.77 |

**Table 8-19. Classification Consistency and Classification Accuracy for Social Studies Grade 4**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.18 | 0.04 | 0.00 | 0.00 | 0.22 |
| *Basic* | 0.04 | 0.12 | 0.04 | 0.00 | 0.20 |
| *Proficient* | 0.00 | 0.04 | 0.25 | 0.05 | 0.35 |
| *Advanced* | 0.00 | 0.00 | 0.05 | 0.18 | 0.24 |
| Sum | 0.22 | 0.20 | 0.35 | 0.23 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.92 | 0.91 | 0.90 | 0.73 |
| Probability of Chance | 0.66 | 0.51 | 0.64 | 0.26 |
| Kappa (*k*) | 0.77 | 0.80 | 0.71 | 0.63 |
| Classification Accuracy | 0.94 | 0.93 | 0.93 | 0.80 |

**Table 8-20. Classification Consistency and Classification Accuracy for Social Studies Grade 8**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.16 | 0.04 | 0.00 | 0.00 | 0.20 |
| *Basic* | 0.04 | 0.13 | 0.05 | 0.00 | 0.22 |
| *Proficient* | 0.00 | 0.05 | 0.23 | 0.05 | 0.33 |
| *Advanced* | 0.00 | 0.00 | 0.05 | 0.20 | 0.25 |
| Sum | 0.20 | 0.22 | 0.33 | 0.25 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.93 | 0.90 | 0.90 | 0.73 |
| Probability of Chance | 0.68 | 0.51 | 0.63 | 0.26 |
| Kappa (*k*) | 0.77 | 0.79 | 0.73 | 0.63 |
| Classification Accuracy | 0.94 | 0.93 | 0.93 | 0.81 |

**Table 8-21. Classification Consistency and Classification Accuracy for Social Studies Grade 10**

**Contingency Table with All Cut Scores**

| Performance Level | *Below Basic* | *Basic* | *Proficient* | *Advanced* | Sum |
|---|---|---|---|---|---|
| *Below Basic* | 0.20 | 0.05 | 0.00 | 0.00 | 0.25 |
| *Basic* | 0.05 | 0.15 | 0.05 | 0.00 | 0.26 |
| *Proficient* | 0.00 | 0.05 | 0.16 | 0.04 | 0.26 |
| *Advanced* | 0.00 | 0.00 | 0.04 | 0.19 | 0.23 |
| Sum | 0.26 | 0.25 | 0.26 | 0.23 | 1.00 |

**Indices for Classification Consistency and Classification Accuracy**

| Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|
| Classification Consistency (P) | 0.89 | 0.89 | 0.91 | 0.70 |
| Probability of Chance | 0.62 | 0.50 | 0.64 | 0.25 |
| Kappa ($k$) | 0.72 | 0.78 | 0.76 | 0.60 |
| Classification Accuracy | 0.93 | 0.92 | 0.94 | 0.79 |

**Table 8-22. Inter-Rater Reliability, English Language Arts**

| Grade | Item No. | Max | Percentage of Agreement | | | Intra. Corr. | Weighted Kappa | Mean | Score Frequency | | | | | |
| | | | Exact | Adjacent | Discrepant | | | | No. of Second Reads | 1 | 2 | 3 | 4 | All Codes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 4 | 91.71 | 8.14 | 0.16 | 0.90 | 0.80 | 1.20 | 15792 | 9040 | 1614 | 192 | 43 | 4903 |
| 4 | 1 | 4 | 89.46 | 10.39 | 0.15 | 0.92 | 0.83 | 1.33 | 10926 | 5958 | 1753 | 339 | 31 | 2845 |
| 5 | 1 | 4 | 88.93 | 11.02 | 0.05 | 0.93 | 0.85 | 1.37 | 10222 | 5555 | 2197 | 334 | 91 | 2045 |
| 6 | 1 | 4 | 84.72 | 15.11 | 0.16 | 0.92 | 0.84 | 1.50 | 11511 | 5316 | 2915 | 751 | 116 | 2413 |
| 7 | 1 | 4 | 85.27 | 14.38 | 0.35 | 0.93 | 0.87 | 1.90 | 15461 | 4882 | 6130 | 1933 | 349 | 2167 |
| 8 | 1 | 4 | 85.88 | 13.88 | 0.24 | 0.94 | 0.88 | 1.80 | 9228 | 3242 | 2959 | 1000 | 220 | 1807 |

Note: The sum of the modes of agreement and codes may not equal exactly 100% due to rounding.
Note: TDA item scores presented in this table reflect a 1–4 point scoring rubric (before application of a weight of 2).

# Part 9: Studies of Construct-Related Validity

As stated in Part 2 of this Technical Report, validity is the overarching component of the Wisconsin Forward Exam program. The following excerpt is from the Standards (AERA, APA, & NCME, 2014):

Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. Different components of validity evidence . . . include evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all test takers, as appropriate to the test interpretation in question. (p. 22)

Part 9 addresses four additional issues related to the evidence of the validity of an intended interpretation of test scores: test fairness, evidence of validity based on the internal structure of the test, evidence of validity based on the relationship between test scores and other variables, and t3est integrity. In the subsequent pages, Part 9 will, as stated, present additional metrics with which to evaluate the validity of an intended interpretation of test scores of the Wisconsin Forward Exam program.

As described below, the Wisconsin Forward Exam program formally assessed the issue of test fairness through an analysis of differential item functioning (DIF). It is possible for items to function differently across different population groups, and it is also possible that results for an item do not reflect student ability but instead reflect irrelevant information influenced by demographic factors. The DIF analysis provided below serves to determine whether that possibility occurred and, if so, to what degree, item by item, for each of the categories of gender, race/ethnicity, economic status, disability status, accommodation use, and English language proficiency.

This part is particularly relevant to AERA, APA, & NCME (2014) Standards 3.1, 3.2, 3.3, and 3.6. Each of these standards and the way in which the standard is addressed will be presented in this part.

> **Standard 3.6** Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (p. 65)

There is no particular research on the Wisconsin Forward Exam showing that the test scores of examinee subgroups differ in meaning; however, this is an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, DRC follows multiple best practices of the testing industry in item development and selection, as is explained in Part 3. These practices adhere to AERA, APA, & NCME (2014) Standards 3.1, 3.2, and 3.3:

> **Standard 3.1** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (p. 63)

> **Standard 3.2** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (p. 64)

**Standard 3.3** Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test. (p. 64)

DRC conducted DIF studies following the operational administration of the Wisconsin Forward Exam. Items are first evaluated for possible DIF in the field test phase of test development and again after their operational administration. Items flagged for DIF are further examined for possible bias. All items on the Spring 2022 test forms, except for ELA TDA items, were field-tested in Wisconsin, and DIF analyses were already performed on these items after the field test administration. The DIF analyses were repeated in Spring 2022. Items flagged for DIF were again evaluated by DRC content experts for potential bias. Section 9.1 of this part of the Technical Report explains the steps taken to evaluate the Wisconsin Forward Exam items through the use of DIF.

Section 9.2 of the report provides the evidence of the validity of an intended interpretation of test scores related to test construct. Two measures of the test internal structure are provided: correlations between content area reporting category (standard) scores and principal component analysis. Both of these measures are provided to demonstrate the existence of a single, underlying trait or ability for each content area, such as ELA ability or Mathematics ability. The presence of a single, underlying trait is a fundamental issue when scaling and analyzing results through IRT models. Therefore, these analyses are essential elements in assessing the validity of the Wisconsin Forward Exam.

In Section 9.3, the relationship between the Wisconsin Forward Exam scores and other variables is explored in order to support the evidence of the validity of an intended interpretation of test scores. These measures include evaluation of the correlations of the content area scores with other content area scores for the total population and by subgroups. They also include comparisons of student performance on the Wisconsin Forward Exam with performance on the NAEP.

In addition, Section 9.4 provides an overview of the forensic analysis procedures that were employed to ensure the integrity of test scores by identifying schools and individual students that might have engaged in inappropriate behaviors during testing.

## 9.1 Differential Item Functioning

An empirical DIF approach was used to examine potential item bias and to determine whether item performance differences between identifiable subgroups were due to extraneous or construct-irrelevant information, making the items unfairly difficult for a particular subgroup in the student population. An item was flagged for DIF when there was a significant difference in the scores between a focal group of students and a reference group of students, with both groups at the same overall ability level. Thus, an item flagged for DIF is more difficult for a particular group of students than would be expected based on their total test scores (Camilli & Shepard, 1994; Green, 1975).

DIF analyses were conducted based on gender, race/ethnicity, economic status, disability status, English language proficiency, and accommodation use groups. The reference and focal groups are as follows:

- **Gender**—reference group: male students; focal group: female students

- **Race/Ethnicity**—reference group: White students; focal groups: African American, Asian, Hispanic, American Indian students

- **Economic status**—reference group: not economically disadvantaged students; focal group: economically disadvantaged students

- **Disability status**—reference group: students without disabilities; focal group: students with disabilities
- **English language proficiency**—reference group: fully English proficient students; focal group: students of limited English proficiency
- **Accommodation use**—reference group: students not using testing accommodations; focal group: students using testing accommodations

Two DIF statistics that are commonly used for this purpose are the Mantel-Haenszel (MH) statistic (1959) and the Standardized Mean Difference (SMD) between the reference and focal groups, proposed by Dorans and Schmitt (1991).

The MH statistic is computed as follows (Zwick, Donoghue, & Grima, 1993):

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

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

In addition to the MH chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. The delta statistic was developed by Educational Testing Service (Holland & Thayer, 1985, 1986). To compute delta, alpha (the odds ratio) is first computed:

$$\alpha MH = \frac{\sum_{k=1}^{K} Nr1kNf0k/Nk}{\sum_{k=1}^{K} Nf1kNr0k/Nk},$$

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

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

For selected response items, the MH ($\chi^2_{MH}$) statistic was used to evaluate potential DIF items. In the MH procedure, subgroups are matched by their raw total test score using a contingency table with $k$ ability levels. When applying the MH procedure, the log-odds ratio $\alpha$ is assumed to be constant across the $k$ matched levels. Then the $\chi^2_{MH}$ estimates a pooled common odds ratio. Taking the natural logarithm of the common-odds ratio and its confidence limits and multiplying these with the constant $-2.35$, the resulting values may then be placed on the MH delta metric ($\Delta_{MH}$) for interpretive purposes. Items were flagged for DIF using the following criteria:

- Moderate DIF: Significant MH chi-square statistic ($p < 0.05$) and $1.0 \leq |\text{MH D-DIF}| < 1.5$
- Large DIF: Significant MH chi-square statistic ($p < 0.05$) and $|\text{MH D-DIF}| \geq 1.5$

For or non-MC items, an effect size (ES) statistic based on the MH chi-square was used. The ES is obtained by dividing the SMD statistics by the standard deviation (SD) of the item. The SMD is an ES index of DIF, which is relatively easy to interpret (Zwick et al., 1993). The SMD compares the means of the reference and focal groups, adjusting for the distribution of the reference and focal group members on the conditioning variable (Zwick et al., 1993), which for these analyses is the Wisconsin Forward Exam raw score. SMD is computed as follows (Zwick et al., 1993):

$$SMD = pFk\left(\sum_k mFk - \sum_k mRk\right),$$

where $pFk$ = the proportion of the focal group members at the $k$th level of the matching variable, $mFk = 1/N_{F1k}$, and $mRk = 1/N_{R1k}$. Items are flagged using the same rules that are used in the NAEP:

- Moderate DIF: If the MH statistic is significant ($p < 0.05$) and |ES| is between 0.17 and 0.25

- Large DIF: If the MH statistic is significant ($p < 0.05$) and |ES| $\geq 0.25$

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group. Tables 9-1 through 9-8 show the DIF results for all subgroups of students.

A negative SMD value implies that the focal group has a lower mean item score than the reference group, whereas a positive value implies that the focal group has a higher mean item score than the reference group, conditioned on the matching test score.

The minimum case count for the focal group was set at 200, and the minimum case count for the reference group was set at 400. The DIF analyses were not performed for subgroups of fewer than 200 students. In these cases, the statistical procedures do not have sufficient power to detect differences should they exist.

Tables 9-1 through 9-8 show items that were flagged based on the criteria described above. The B flag represents a lower threshold for DIF. Only items that were flagged with a B or C flag were included in Tables 9-1 through 9-8.

The DIF results for gender are presented in Table 9-1, results for race/ethnicity are presented in Tables 9-2 through 9-5, results for English language proficiency are presented in Table 9-6, results for disability status are presented in Table 9-7, and results for accommodation use are presented in Table 9-8. No items were flagged for DIF by student economic status in the Spring 2022 administration.

Each DIF table references the grade and content area of the items flagged for DIF, the item number on the test, and the item type. The tables present the MH SMD statistics and the Mantel-Haenszel statistics ($\Delta_{MH}$). After specifying these statistics for each item, the final column provides a flag status. The flag is based on SMD statistics and on MH ($\Delta_{MH}$) statistics.

In Table 9-1, looking at all items and all grades and content areas, eight items were flagged for moderate (B flag) gender DIF in the ELA tests (grades 5, 6, 7, and 8). Of these items, five were flagged in favor of the focal group (females) and three were flagged against the focal group. Six items were flagged for moderate DIF, and one item was flagged for large DIF in the Mathematics tests (grades 3, 5, 6, and 7). For the six moderate DIF items, two items were flagged in favor of female students and four items were flagged against female students; the item with large DIF was flagged against female students. Three items were flagged for moderate DIF against female students in the Science grade 8 test. In addition, a total of

six items were flagged for moderate DIF in the Social Studies tests (five items in grade 8 and one item in grade 10). A total of four items were flagged against female students and two items were flagged in favor of female students in Social Studies. Overall, nine items were flagged in favor of the focal group (females) and fifteen items were flagged against the focal group across all grades and content areas. Of all items flagged for gender DIF, one displayed large DIF (against female students) and twenty-three items displayed moderate DIF (either in favor of or against female students).

The other DIF results in Tables 9-2 through 9-8 can be understood in the same fashion. Note that a single item can be flagged for multiple subgroup categories, such as for ethnicity and language proficiency.

When looking at DIF results by item type, it was observed that most of the flagged items were MC and TE items across all content areas and subgroups. In addition, ELA TDA items were flagged for DIF by disability status (against students with disabilities in all grades), by ethnicity status (in favor of Asian students in grade 3, against American Indian students in grades 5 through 8, and against African American students in grade 8), and by gender (in favor of female students in grades 5 through 8).

Combined, the DIF statistical analyses discussed above and the expert reviews provide an appropriate set of tools with which to minimize the extraneous or construct-irrelevant information associated with item bias or DIF in the Wisconsin Forward Exam. It should be noted that in large-scale assessments, such as the Wisconsin Forward Exam, it is expected that some items will show DIF. All items flagged for DIF are annotated as such in the item pool so that content experts would be able to reevaluate these items in future item selection activities. Items with DIF (particularly items flagged for large DIF) are to be avoided in future selections.

## 9.2 Validity Evidence Based on Internal Test Structure

Construct-related evidence of the validity of an intended interpretation of test scores can be defined as the extent to which tests measure the skills or constructs they intend to measure and is the central concept underlying the Spring 2022 Wisconsin Forward Exam validation process. Evidence for construct-related validity is comprehensive and integrates evidence from both content- and criterion-related validity. The Wisconsin Forward Exam development process included specifications, item writing, review, and test construction.

Threats to construct-related validity include the unintended measurement of variables unrelated to the desired constructs and multidimensionality of the tests. To ensure that the test items are focused on the desired constructs, standardized procedures are employed to select items with sound statistical properties, to align the items to content standards, and to ensure that each test form meets the Wisconsin Forward Exam blueprint. A test can be said to be unidimensional when all of the items in the test measure the same underlying ability or trait. For example, Mathematics items should measure Mathematics ability and not Reading skills. Standard 1.13 of the Standards (AERA, APA, & NCME, 2014) states the following:

If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided. (pp. 26–27)

## 9.2.1 Correlations between Content Standards

Analyses of the internal structure of a test can indicate the extent to which the relationships between test items and components conform to the construct the test purports to measure. For educational assessments that are designed to measure a single construct or content domain, the correlations between content standards within a test can be expected to be relatively high. Table 9-9 shows the correlations between the main test domains for ELA, and Tables 9-10 through 9-13 show the correlations between content standards for each Wisconsin Forward Exam content area. The correlation coefficients here reflect the degree of linear relationship and direction between any two given content standards. The correlation can range from +1 to −1. A correlation of +1 indicates a perfect positive linear relationship between two content standards, and a correlation of −1 indicates a perfect negative linear relationship between two content standards. A correlation of zero means there is no linear relationship. In general, the size of the correlation coefficient is influenced by the number of items or score points and by the score variance. Readers are cautioned not to confuse correlation with causation. The presence of a high correlation between two content standards should not be taken as an indication that there is a causal relationship between them.

As may be observed in Table 9-9, the correlations between the ELA main test domains of Reading, Writing, and Listening are moderate to high and range from 0.56 to 0.74 across all grades. Lower correlations (at or lower than 0.67) were observed between the Listening and Writing domains and the Listening and Reading domains, while higher correlations (at least 0.72 or higher) were observed between the Reading and Writing domains. The correlations between ELA content standards (see Table 9-10) are typically moderate for all grades and all standard pairs and range from 0.33 to 0.71. It should be noted that the number of items associated with each content standard was smaller than the number of items associated with each ELA domain, resulting in lower correlations at the standard level compared to the correlations at the ELA domain level.

As indicated in Table 9-11, the correlations between Mathematics content standards are moderate to high and range from 0.58 to 0.77. The correlations between Science content standards range from 0.60 to 0.71 (see Table 9-12), and the correlations between Social Studies content standards range from 0.53 to 0.74 (as shown in Table 9-13). Overall, the correlations for all content areas are within the moderate to high range.

Although it may be tempting to try to interpret the differences in magnitude within and across content areas, it is important to note that these correlations are highly dependent upon the numbers of items and the score variance for the different standards. The important finding is that within each content area, the correlations between content standards are low enough to indicate that the standards are, as intended, somewhat distinct from one another but high enough to indicate that the individual standards are measuring related components of a single content area.

## 9.2.2 Principal Component Analysis

Wisconsin Forward Exam items are calibrated using unidimensional IRT models, which suggests that the test items are measuring an essentially unidimensional construct. To assess the dimensionality of the Wisconsin Forward Exam, a principal components analysis was conducted for each content area and grade. A principal components analysis is a statistical technique commonly used to evaluate dimensionality by detecting patterns of relationships among items. This method is useful in determining whether the observed scores on a test can be explained largely or entirely in terms of a much smaller number of components. For example, if answering the Mathematics items in a Mathematics test required a high level of reading ability, the Mathematics test would be measuring not only mathematics ability but also reading ability. Such a test would be said to be multidimensional rather than essentially unidimensional. One way of evaluating the dimensions detected in the analysis is by examining the eigenvectors and eigenvalues. In

a principal components analysis, the eigenvectors correspond to factors, and the eigenvalues correspond to the variance explained by these factors. The sum of the eigenvalues is equal to the number of items in the test. The eigenvalues can be ordered from first to last in terms of the amount of common variance that each explains. Data are generally considered to be unidimensional if the second eigenvalue is less than or equal to 1.0. Previous research shows that an examination of the ratio of the first two (i.e., the two largest) eigenvalues can be useful in determining the existence of dominant factors. Specifically, where large ratios exist between the first and second eigenvalues, a single dominant factor can be said to exist. Although the definition of "large" in the present context is subjective, the results in Table 9-14 show that the eigenvalue of the first factor is more than five times as large as the eigenvalue of the second factor.

As can be seen in Table 9-14, the ratios of the first two eigenvalues range from 5.72 to 7.49. The eigenvalues are proportional to the amount of common variance explained by each component, indicating that the variance explained by the first component alone is between six and eight times greater than the variance explained by the second component. The eigenvalue ratios range from 5.94 to 6.91 in ELA, from 6.93 to 7.49 in Mathematics, from 5.72 to 6.69 in Science, and from 6.71 to 7.19 in Social Studies. These ratios suggest that the unidimensionality of each of the Wisconsin Forward Exam content assessments is sufficient to meet the requirements of a unidimensional IRT calibration model.

Overall, these results provide support for the construct validity of the Wisconsin Forward Exam assessments. The correlations between content standards and the presence of a single dominant factor for each test confirm that the content standards are sufficiently unidimensional to be combined into a single score.

## 9.3 Validity Evidence Based on Relationship with Other Variables

The relationship between the Wisconsin Forward Exam scores and other variables was examined to further support the validity of the intended score interpretation. This was done using two measures: evaluation of correlations between the Wisconsin Forward Exam content area scores and comparisons of the percentages of students classified in different performance levels (impact data) on the State assessment and on the NAEP assessment.

### 9.3.1 Correlations between Content Area Test Scores

The test score relationship with other variables can be assessed by the extent to which measures of constructs that theoretically should not be related to each other are, in fact, observed as not related to each other. Typically, correlation coefficients between measures of unrelated or distantly related constructs are examined in support of the validity evidence based on the relationship of the test scores with other variables.

To assess the relationships between the Wisconsin Forward Exam content area scores, the correlations between the ELA, Mathematics, Science, and Social Studies scale scores for students who took more than one subject area test in 2022 were computed and examined for the total student population and for each subgroup. Table 9-15 shows the correlations between the content area scores for the total population of Wisconsin students. The correlations between ELA and Mathematics scores ranged from 0.74 to 0.77 across all grades. The correlations between ELA and Science scores and between ELA and Social Studies scores ranged from 0.78 to 0.83 for grades 4 and 8. The correlations between Mathematics and Science scores and between Mathematics and Social Studies scores ranged from 0.72 to 0.76 for grades 4 and 8. Lastly, the correlations between Science and Social Studies scale scores were 0.84 for grade 4 and 0.80 for grade 8.

Tables 9-16 through 9-20 show correlation coefficients between the content area scores by gender, ethnicity, English language proficiency status, economic status, and disability status, respectively. As seen in Table 9-16, the correlations between the content area scores for male or female groups ranged from 0.71 to 0.84 and were comparable for the two gender groups for each pair of correlated scores. The correlations between the content area scores for different ethnic groups ranged from 0.56 to 0.88 (see Table 9-17). The highest correlations by ethnic group were observed for Asian students. Correlations between the content area scores for the African American student subgroup were generally lower than the correlations for other subgroups. As shown in Table 9-18, the correlations between the content area scores by English proficiency status ranged from 0.54 to 0.78 for limited English proficiency students and from 0.72 to 0.84 for fully English proficient students across all grade levels and all pairs of correlated scores. The correlations between the content area scores by student economic status are presented in Table 9-19. These correlations ranged from 0.71 to 0.82 for students who were not economically disadvantaged and from 0.65 to 0.82 for economically disadvantaged students across all grades and pairs of correlated scores. The correlations between the content area scores by student disability status are shown in Table 9-20. These correlations ranged from 0.71 to 0.83 for students without disabilities and from 0.57 to 0.81 for students with disabilities across all grades and pairs of correlated scores. In all grade levels, the correlations between each pair of scores were, in most cases, lower for the groups of students classified as English language learners, economically disadvantaged, or disabled compared to the groups of students classified as fully English proficient, not economically disadvantaged, or not disabled. In all grade levels and for each subgroup, the correlations between Mathematics test scores and other content area scores tended to be lower than the correlations between ELA scores and other content area scores. The correlation coefficients between the content area scores were not computed by accommodation use, because the accommodation use status is not consistent across content areas for the same students (e.g., students who used accommodations in one content area did not necessarily use accommodations in another content area).

Overall, the correlations between the content area scores for the total population of students were found to be highly related. As mentioned, correlations between Mathematics test scores and other content area scores were found to be consistently lower than the correlations between ELA scores and other content area scores, suggesting larger differences between the Mathematics constructs and constructs measured by other content areas. The correlations between the content area scores for the subgroups of students were found to be moderately to highly related. Despite relatively high correlations, the tests are not perfectly related to one another, suggesting that different constructs are being tapped; however, if the test scores are highly related to one another, they may be tapping into a similar knowledge base or general underlying ability. This outcome is not unexpected for the new generations of large-scale assessments based on new CCR standards, such as the Wisconsin State Standards that emphasize teaching and learning various content skills across content areas. All assessments are intended to be aligned to performance expectations that are more reflective of the current knowledge and skill demands of postsecondary education and careers. Given the cross-content complexity and cognitive demands of the assessments, the relationship between student scores from different tests has necessarily become strengthened, and larger correlations between the content areas are expected.

## Partial Correlations

In addition to the simple correlations between the content area scores, partial correlations, which are measures of the strength of the relationship between the content area scores while controlling for the student demographic characteristics (gender, ethnicity, English proficiency status, disability status, and economic status), were also computed. Partial correlations allow for the evaluation of the relationship between two content area scores with the effect of the student demographic characteristics removed (or held constant). The partial correlations between the ELA, Mathematics, Science, and Social Studies test

scores for the total population of students and at each grade level are presented in Table 9-21. The partial correlations between ELA and Mathematics scores ranged from 0.66 to 0.69 across all grades. The partial correlations between ELA and Science scores and ELA and Social Studies scores ranged from 0.73 to 0.78. The partial correlations between Mathematics and Science scores and Mathematics and Social Studies scores ranged from 0.62 to 0.68. And the partial correlations between Science and Social Studies scores were 0.79 for grade 4 and 0.75 for grade 8.

Although the magnitude of these correlations is considered to be strong, the partial correlations between the content area scores were lower than the corresponding simple correlations, indicating that the student demographic characteristics did contribute to the strength of the relationship between the content area test scores. The differences between the simple correlation coefficients and corresponding partial correlation coefficients suggested that there may be some effect of the student demographic characteristics on the relationships between the ELA, Mathematics, Science, and Social Studies test scores.

## 9.3.2 Comparison of the Wisconsin Forward Exam and Wisconsin NAEP Impact Data

The NAEP is the largest nationally representative and continuing assessment of what America's students know and can do in various content areas. Assessments in several content areas, including Reading, Mathematics, and Science, are administered to students in grades 4, 8, and 12 and conducted periodically. Representative samples of students from different states, including Wisconsin, participated in the latest NAEP assessment, which occurred in Spring 2019.

The main NAEP assessments are constructed using detailed frameworks that result from a comprehensive national process in which teachers, curriculum experts, policymakers, and members of the general public work to create a unified vision of how a particular subject ought to be assessed. This vision is based on current educational research on achievement and its measurement as well as good educational practices. The frameworks are devised through a development process that ensures they meet current educational requirements. (For details, refer to https://nces.ed.gov/nationsreportcard/assessments/frameworks.aspx.)

The NAEP results are reported for all assessed content areas and for all participating grades at the national level. At the state level, the results for Reading, Mathematics, Science, and Writing are reported for grades 4 and 8. The results may also be reported at the district level (within a state) for these four content areas. No results are reported at the student level.

Wisconsin students participated in the latest Reading, Mathematics, and Science NAEP assessments in Spring 2019. The percentages of Wisconsin students classified in different performance levels on the Wisconsin Forward Exam in Spring 2022 and the corresponding NAEP assessments for Reading and Mathematics from Spring 2019 are presented in Table 9-22. Because the Spring 2019 Science results by state are not available at the time of preparation of this report, the Spring 2015 Science NAEP results for Wisconsin students are presented (also in Table 9-22). With three exceptions, the percentages of students classified in the different performance levels on the NAEP assessments and on the Wisconsin Forward Exam were comparable and within 10% of each other for every performance level across both grades and all three content areas. The exceptions were higher percentages of students classified in the *Below Basic* performance level on the Wisconsin Forward Mathematics test in grade 8 and higher percentages of students classified in the *Advanced* performance level on the Wisconsin Forward Science test for grades 4 and 8 compared to the percentages of students classified in the respective performance levels on the corresponding NAEP assessments.

Looking at the percentages of students classified at or above *Proficient*, higher proportions of students were classified in this combined category of performance levels on the Wisconsin Forward Exam in ELA grade 4 and Science grades 4 and 8 in Spring 2022 compared to the corresponding NAEP Reading and Science assessments in Spring 2019 (Reading) or Spring 2015 (Science). Higher proportions of students were classified at or above *Proficient* on the NAEP assessment in Reading grade 8 and both Mathematics grades in Spring 2019 compared to the corresponding Wisconsin Forward Exam ELA and Mathematics assessments in Spring 2022. As stated earlier in this report, such comparisons should be made with caution because student performance on the Wisconsin Forward Exam in Spring 2022 was likely affected by slow recovery from disrupted learning in the 2020–21 academic year and was found to be lower than student performance in typical pre-pandemic administration years in almost all grades across ELA, Mathematics, and Science.

It should be noted that the Spring 2015 Reading and Mathematics Wisconsin NAEP impact data were used as benchmarks during the Wisconsin Forward Exam standard setting for ELA and Mathematics after the Spring 2016 test administration. The Spring 2015 Science Wisconsin NAEP impact data were also shown to the participants for reference and guidance in performance level setting during the Spring 2019 standard setting. While the standard setting participants were free to deviate from the NAEP impact data while placing their bookmarks in the ordered item booklets in consideration of the Wisconsin performance level descriptors (PLDs), the final Wisconsin impact data achieved after the standard setting were generally aligned with the Wisconsin state-level NAEP data. When considering the Wisconsin content standards and impact data articulation across grades, the Wisconsin Forward Exam cut scores for ELA, Mathematics, and Science remained in most cases aligned with the *Proficient* benchmarks, further supporting the evidence of the relationship between the state and the national assessments in these content areas.

## 9.4 Test Integrity: Data Forensic Analyses

With the high-stakes nature of large-scale statewide assessment programs, there can be situations in which student responses, and hence their scores, may not be a true representation of student ability. Various activities may take place, such as a student copying from another student's paper, a student receiving inappropriate assistance before or during testing, or a student's responses being altered during or after testing. To maintain the integrity of the Wisconsin Forward Exam and the validity of the results, it is important that any such instances be discovered.

Two studies were conducted to evaluate the Wisconsin Forward Exam student data for any indicators of possible inappropriate testing behavior. The first study examines incorrect student responses to MC items on the Spring 2022 Wisconsin Forward Exam in ELA, Mathematics, Science, and Social Studies that were changed to correct responses. These answer changes are referred to as wrong-to-right answer changes. Inordinate numbers of wrong-to-right answer changes in a specifically identifiable testing administration group may indicate inappropriate student behavior or intervention by an educator during the testing session.

The second study evaluates the time spent on the test and individual test items by students. These analyses serve to inform of any events in which students (within one school) spent a very short or very long time on the test or specific items. Inordinate numbers of unusual test or item response times may indicate inappropriate pre-knowledge of the items or other interventions during the testing session.

The results of the two studies are provided to DPI for evaluation. We emphasize that the results from these studies may be used in conjunction with other information to investigate whether inappropriate interventions may have taken place. The statistical results by themselves may simply be coincidental and do not necessarily indicate inappropriate behavior.

## 9.5 Summary

In summary, the overall purpose of Part 9 was to provide additional evidence of the validity of an intended interpretation of test scores related to test construct. Through the measures of correlations between content area reporting category scores and principal components analysis, the existence of a single underlying trait or ability for each content area was demonstrated. Next, the relationship between the Wisconsin Forward Exam scores and other variables was explored and validated through the evaluation of correlations between content area scores for the total population and by subgroups. In addition, student performance on the Wisconsin Forward Exam was compared with student performance on the NAEP assessment. The forensic analysis procedures that were employed to ensure the integrity of test scores by identifying schools and individual students that might have engaged in inappropriate behaviors during testing were also described in this part of the report.

**Table 9-1. Items Flagged for DIF by Gender, Focal Group: Female**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| English Language Arts | 5 | 1 | TDA | 0.13 | | B |
| | 5 | 34 | MC | -0.07 | -1.05 | B- |
| | 6 | 1 | TDA | 0.19 | | B |
| | 6 | 14 | MC | -0.07 | -1.02 | B- |
| | 7 | 1 | TDA | 0.21 | | B |
| | 8 | 1 | TDA | 0.20 | | B |
| | 8 | 5 | MC | -0.09 | -1.07 | B- |
| | 8 | 39 | MC | 0.08 | 1.04 | B |
| Mathematics | 3 | 3 | MC | -0.06 | -1.04 | B- |
| | 5 | 7 | TE | -0.08 | -1.74 | B- |
| | 6 | 13 | MC | -0.06 | -1.04 | B- |
| | 6 | 15 | MC | -0.09 | -1.15 | B- |
| | 7 | 17 | MC | -0.12 | -1.77 | C- |
| | 7 | 36 | TE | 0.08 | 1.29 | B |
| | 7 | 46 | MC | 0.08 | 1.07 | B |
| Science | 8 | 5 | TE | -0.12 | -1.43 | B- |
| | 8 | 22 | TE | -0.09 | -1.12 | B- |
| | 8 | 39 | TE | -0.09 | -1.33 | B- |
| Social Studies | 8 | 5 | MC | -0.08 | -1.10 | B- |
| | 8 | 8 | MC | -0.10 | -1.21 | B- |
| | 8 | 28 | MC | 0.10 | 1.36 | B |
| | 8 | 34 | TE | -0.09 | -1.09 | B- |
| | 8 | 40 | TE | 0.12 | 1.42 | B |
| | 10 | 27 | MC | -0.10 | -1.14 | B- |

**Table 9-2. Items Flagged for DIF by Race/Ethnicity, Focal Group: African American**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| English Language Arts | 8 | 1 | TDA | -0.21 | | B- |
| Mathematics | 3 | 24 | SA | -0.07 | -1.33 | B- |
| | 4 | 40 | SA | -0.09 | -1.20 | B- |
| | 6 | 23 | TE | -0.09 | -0.95 | B- |
| | 8 | 15 | TE | -0.11 | -2.21 | B- |
| Social Studies | 4 | 14 | TE | -0.14 | -1.86 | C- |
| | 8 | 1 | MC | -0.09 | -1.18 | B- |
| | 8 | 5 | MC | -0.10 | -1.29 | B- |
| | 8 | 19 | TE | 0.09 | 1.62 | B |
| | 8 | 34 | TE | -0.18 | -2.19 | C- |

**Table 9-3. Items Flagged for DIF by Race/Ethnicity, Focal Group: Hispanic**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| Social Studies | 8 | 34 | TE | -0.09 | -0.98 | B- |

**Table 9-4. Items Flagged for DIF by Race/Ethnicity, Focal Group: Asian**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| English Language Arts | 3 | 1 | TDA | 0.12 | | B |
| | 5 | 15 | MC | -0.05 | -1.10 | B- |
| | 5 | 35 | MC | 0.09 | 1.11 | B |
| | 7 | 12 | TE | 0.06 | 1.50 | B |
| Mathematics | 3 | 32 | SA | 0.09 | 1.25 | B |
| | 4 | 17 | MC | -0.07 | -1.06 | B- |
| | 5 | 22 | TE | 0.10 | 1.36 | B |
| | 6 | 35 | MC | -0.09 | -1.25 | B- |
| | 8 | 15 | TE | -0.09 | -1.22 | B- |
| Science | 8 | 5 | TE | -0.09 | -1.12 | B- |
| Social Studies | 4 | 1 | TE | 0.09 | 1.12 | B |
| | 4 | 15 | TE | -0.09 | -1.14 | B- |
| | 8 | 1 | MC | -0.09 | -1.58 | C- |
| | 8 | 9 | MC | -0.07 | -1.11 | B- |
| | 8 | 13 | MC | -0.07 | -1.24 | B- |
| | 8 | 16 | MC | 0.08 | 1.63 | C |
| | 8 | 18 | MC | 0.07 | 1.14 | B |
| | 8 | 21 | MC | 0.05 | 1.32 | B |
| | 8 | 25 | TE | -0.09 | -1.53 | B- |
| | 8 | 28 | MC | 0.10 | 1.49 | B |
| | 8 | 34 | TE | -0.08 | -1.04 | B- |
| | 10 | 1 | TE | -0.14 | -1.86 | C- |
| | 10 | 2 | MC | -0.12 | -1.58 | C- |
| | 10 | 12 | MC | 0.07 | 1.02 | B |
| | 10 | 28 | EBSR | 0.08 | 1.16 | B |

**Table 9-5. Items Flagged for DIF by Race/Ethnicity, Focal Group: American Indian**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| English Language Arts | 5 | 1 | TDA | -0.14 | | B- |
| | 6 | 1 | TDA | -0.18 | | B- |
| | 7 | 1 | TDA | -0.23 | | C- |
| | 7 | 12 | TE | 0.06 | 0.99 | B |
| | 8 | 1 | TDA | -0.26 | | C- |
| Mathematics | 8 | 14 | MC | -0.09 | -1.06 | B- |
| | 8 | 16 | MC | -0.08 | -1.05 | B- |
| Social Studies | 4 | 9 | TE | -0.06 | -1.00 | B- |
| | 10 | 32 | MC | 0.14 | 1.99 | C |

**Table 9-6. Items Flagged for DIF by English Language Proficiency, Focal Group: Students Not English Language Proficient**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| Social Studies | 8 | 25 | TE | -0.11 | -1.36 | C- |
| | 8 | 34 | TE | -0.11 | -1.36 | B- |
| | 10 | 1 | TE | -0.12 | -1.30 | C- |
| | 10 | 2 | MC | -0.10 | -1.16 | B- |

**Table 9-7. Items Flagged for DIF by Disability Status, Focal Group: Students with One or More Disabilities**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| English Language Arts | 3 | 1 | TDA | -0.15 | | B- |
| | 4 | 1 | TDA | -0.19 | | C- |
| | 4 | 4 | MC | -0.1 | -1.47 | B- |
| | 5 | 1 | TDA | -0.16 | | B- |
| | 5 | 2 | MC | -0.09 | -1.09 | B- |
| | 5 | 12 | MC | -0.08 | -1.01 | B- |
| | 6 | 1 | TDA | -0.21 | | C- |
| | 6 | 10 | TE | -0.11 | | B- |
| | 7 | 1 | TDA | -0.27 | | C- |
| | 7 | 12 | TE | -0.12 | -1.43 | C- |
| | 8 | 1 | TDA | -0.31 | | C- |
| Mathematics | 8 | 14 | MC | -0.10 | -1.14 | B- |
| | 8 | 16 | MC | -0.10 | -1.17 | B- |
| Social Studies | 4 | 9 | TE | -0.07 | -1.15 | B- |
| | 10 | 1 | TE | -0.12 | -1.40 | C- |

**Table 9-8. Items Flagged for DIF by Accommodation Use, Focal Group: Students Using Testing Accommodations**

| Content | Grade | Item Number | Item Type | MH SMD Statistic | MH Delta Statistic | DIF Flag |
|---|---|---|---|---|---|---|
| Mathematics | 3 | 16 | MC | 0.06 | 1.01 | B |
| | 3 | 18 | TE | -0.11 | -1.55 | C- |
| | 3 | 23 | TE | -0.07 | -0.88 | B- |
| | 4 | 3 | TE | -0.09 | -1.03 | B- |
| | 4 | 17 | MC | 0.11 | 1.19 | B |
| | 5 | 42 | MC | -0.10 | -1.17 | B- |
| | 6 | 13 | MC | -0.12 | -1.30 | B- |
| | 6 | 20 | SA | -0.09 | -1.73 | B- |
| | 7 | 37 | MC | 0.05 | 1.31 | B |
| | 7 | 45 | SA | -0.09 | -2.07 | B- |
| | 8 | 16 | MC | -0.11 | -1.18 | B- |

Note: DIF analysis by accommodation use was not performed on ELA, Science, and Social Studies data due to an insufficient number of students using testing accommodations in these content areas.

**Table 9-9. Correlations between English Language Arts Test Domains**

| Grade | ELA Domain | Listening | Reading |
|-------|-----------|-----------|---------|
| 3 | Reading | 0.61 | |
| 3 | Writing | 0.57 | 0.72 |
| 4 | Reading | 0.59 | |
| 4 | Writing | 0.56 | 0.74 |
| 5 | Reading | 0.63 | |
| 5 | Writing | 0.61 | 0.73 |
| 6 | Reading | 0.67 | |
| 6 | Writing | 0.60 | 0.72 |
| 7 | Reading | 0.66 | |
| 7 | Writing | 0.64 | 0.74 |
| 8 | Reading | 0.65 | |
| 8 | Writing | 0.61 | 0.73 |

**Table 9-10. Correlations between Content Standards, English Language Arts**

| Grade | Standard Code | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|
| 3 | B | 0.60 | | | | | |
| | C | 0.59 | 0.48 | | | | |
| | D | 0.54 | 0.46 | 0.45 | | | |
| | E | 0.58 | 0.48 | 0.49 | 0.51 | | |
| | F | 0.52 | 0.44 | 0.46 | 0.45 | 0.48 | |
| | G | 0.58 | 0.46 | 0.47 | 0.45 | 0.51 | 0.43 |
| 4 | B | 0.65 | | | | | |
| | C | 0.64 | 0.62 | | | | |
| | D | 0.50 | 0.49 | 0.50 | | | |
| | E | 0.57 | 0.55 | 0.55 | 0.49 | | |
| | F | 0.51 | 0.50 | 0.52 | 0.48 | 0.48 | |
| | G | 0.52 | 0.49 | 0.52 | 0.43 | 0.49 | 0.44 |
| 5 | B | 0.68 | | | | | |
| | C | 0.44 | 0.40 | | | | |
| | D | 0.54 | 0.50 | 0.33 | | | |
| | E | 0.60 | 0.55 | 0.38 | 0.48 | | |
| | F | 0.52 | 0.48 | 0.34 | 0.46 | 0.49 | |
| | G | 0.59 | 0.54 | 0.38 | 0.46 | 0.56 | 0.47 |
| 6 | B | 0.71 | | | | | |
| | C | 0.63 | 0.61 | | | | |
| | D | 0.56 | 0.55 | 0.48 | | | |
| | E | 0.48 | 0.47 | 0.41 | 0.41 | | |
| | F | 0.53 | 0.50 | 0.45 | 0.47 | 0.41 | |
| | G | 0.62 | 0.60 | 0.54 | 0.50 | 0.44 | 0.46 |
| 7 | B | 0.62 | | | | | |
| | C | 0.59 | 0.53 | | | | |
| | D | 0.55 | 0.51 | 0.51 | | | |
| | E | 0.55 | 0.49 | 0.49 | 0.48 | | |
| | F | 0.48 | 0.43 | 0.47 | 0.45 | 0.44 | |
| | G | 0.59 | 0.53 | 0.58 | 0.53 | 0.52 | 0.47 |
| 8 | B | 0.60 | | | | | |
| | C | 0.62 | 0.61 | | | | |
| | D | 0.52 | 0.52 | 0.53 | | | |
| | E | 0.54 | 0.53 | 0.56 | 0.50 | | |
| | F | 0.45 | 0.43 | 0.46 | 0.44 | 0.44 | |
| | G | 0.56 | 0.54 | 0.58 | 0.49 | 0.53 | 0.43 |

ELA standards: A=Reading—Key Ideas and Details; B=Reading—Craft & Structure/Integration of Knowledge & Ideas; C=Reading—Vocabulary Use; D=Writing/Language—Text Types and Purposes; E=Writing/Language—Research; F=Writing/Language—Language Conventions; G = Listening

**Table 9-11. Correlations between Content Standards, Mathematics**

| Grade | Standard Code | A | B | C | D | E | F | G | H | I |
|-------|---------------|------|------|------|------|------|------|------|------|------|
| 3 | B | 0.77 | | | | | | | | |
| | C | 0.67 | 0.67 | | | | | | | |
| | D | 0.74 | 0.74 | 0.67 | | | | | | |
| | E | 0.67 | 0.67 | 0.66 | 0.68 | | | | | |
| 4 | B | 0.73 | | | | | | | | |
| | C | 0.73 | 0.71 | | | | | | | |
| | D | 0.67 | 0.65 | 0.68 | | | | | | |
| | E | 0.61 | 0.60 | 0.63 | 0.61 | | | | | |
| 5 | B | 0.70 | | | | | | | | |
| | C | 0.69 | 0.71 | | | | | | | |
| | D | 0.63 | 0.65 | 0.65 | | | | | | |
| | E | 0.68 | 0.69 | 0.67 | 0.64 | | | | | |
| 6 | F | | | | | 0.64 | | | | |
| | G | | | | | 0.70 | 0.70 | | | |
| | H | | | | | 0.71 | 0.70 | 0.76 | | |
| | I | | | | | 0.63 | 0.63 | 0.68 | 0.67 | |
| 7 | F | | | | | 0.64 | | | | |
| | G | | | | | 0.61 | 0.73 | | | |
| | H | | | | | 0.59 | 0.69 | 0.67 | | |
| | I | | | | | 0.64 | 0.75 | 0.70 | 0.68 | |
| 8 | G | | | | | 0.62 | | | | |
| | H | | | | | 0.63 | | 0.66 | | |
| | I | | | | | 0.61 | | 0.58 | 0.61 | |
| | J | | | | | 0.68 | | 0.65 | 0.69 | 0.69 |

Note: Standard Codes are as follows: A = Operations and Algebraic Thinking; B = Number and Operations in Base Ten;
C = Number and Operations—Fractions; D = Measurement and Data; E = Geometry; F = Ratios and Proportional Relationships;
G = The Number System; H = Expressions and Equations; I = Statistics and Probability; J = Functions

**Table 9-12. Correlations between Content Standards, Science**

| Grade | Standard Code | A | B | C |
|---|---|---|---|---|
| 4 | B | 0.68 | | |
| | C | 0.60 | 0.64 | |
| | D | 0.67 | 0.71 | 0.65 |
| 8 | B | 0.66 | | |
| | C | 0.64 | 0.62 | |
| | D | 0.66 | 0.67 | 0.63 |

Note: Standard Codes are as follows: A = Life Science; B = Physical Science; C = Earth and Space Science; D = Engineering

**Table 9-13. Correlations between Content Standards, Social Studies**

| Grade | Standard Code | A | B | C | D |
|---|---|---|---|---|---|
| 4 | B | 0.69 | | | |
| | C | 0.64 | 0.58 | | |
| | D | 0.64 | 0.59 | 0.53 | |
| | E | 0.74 | 0.66 | 0.60 | 0.61 |
| 8 | B | 0.68 | | | |
| | C | 0.72 | 0.70 | | |
| | D | 0.67 | 0.64 | 0.69 | |
| | E | 0.67 | 0.64 | 0.70 | 0.63 |
| 10 | B | 0.69 | | | |
| | C | 0.71 | 0.70 | | |
| | D | 0.64 | 0.63 | 0.64 | |
| | E | 0.59 | 0.59 | 0.60 | 0.54 |

Note: Standard Codes are as follows: A = Geography; B = History; C = Political Science; D = Economics; E = Behavioral Sciences

**Table 9-14. Principal Components Analysis**

| Content Area | Grade | First Eigenvalue | Second Eigenvalue | Ratio of First Two Eigenvalues |
|---|---|---|---|---|
| English Language Arts | 3 | 7.457 | 1.239 | 6.021 |
| | 4 | 8.255 | 1.232 | 6.698 |
| | 5 | 8.182 | 1.378 | 5.937 |
| | 6 | 8.271 | 1.229 | 6.728 |
| | 7 | 7.869 | 1.168 | 6.738 |
| | 8 | 8.024 | 1.161 | 6.912 |
| Mathematics | 3 | 11.306 | 1.524 | 7.417 |
| | 4 | 11.252 | 1.502 | 7.493 |
| | 5 | 10.958 | 1.582 | 6.925 |
| | 6 | 11.341 | 1.574 | 7.205 |
| | 7 | 10.680 | 1.437 | 7.430 |
| | 8 | 10.549 | 1.515 | 6.964 |
| Science | 4 | 8.025 | 1.199 | 6.693 |
| | 8 | 7.571 | 1.324 | 5.720 |
| Social Studies | 4 | 9.140 | 1.272 | 7.186 |
| | 8 | 9.331 | 1.381 | 6.754 |
| | 10 | 8.349 | 1.245 | 6.707 |

**Table 9-15. Correlations between Content Area Scale Scores**

| Grade | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|
| 3 | 0.76 | | | | | |
| 4 | 0.76 | 0.83 | 0.83 | 0.76 | 0.75 | 0.84 |
| 5 | 0.76 | | | | | |
| 6 | 0.77 | | | | | |
| 7 | 0.76 | | | | | |
| 8 | 0.74 | 0.78 | 0.80 | 0.75 | 0.72 | 0.80 |

**Table 9-16. Correlations between Content Area Scale Scores by Gender**

| Grade | Demographic Group | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|---|
| 3 | Female | 0.77 | | | | | |
| | Male | 0.77 | | | | | |
| 4 | Female | 0.77 | 0.83 | 0.84 | 0.77 | 0.75 | 0.84 |
| | Male | 0.77 | 0.83 | 0.83 | 0.76 | 0.75 | 0.84 |
| 5 | Female | 0.77 | | | | | |
| | Male | 0.77 | | | | | |
| 6 | Female | 0.78 | | | | | |
| | Male | 0.78 | | | | | |
| 7 | Female | 0.77 | | | | | |
| | Male | 0.77 | | | | | |
| 8 | Female | 0.74 | 0.80 | 0.81 | 0.75 | 0.72 | 0.81 |
| | Male | 0.74 | 0.79 | 0.80 | 0.75 | 0.71 | 0.80 |

**Table 9-17. Correlations between Content Area Scale Scores by Ethnicity/Race**

| Grade | Demographic Group | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|---|
| 3 | White | 0.72 | | | | | |
| | African American | 0.65 | | | | | |
| | Hispanic | 0.72 | | | | | |
| | Asian | 0.78 | | | | | |
| | American Indian | 0.67 | | | | | |
| | Two or More | 0.75 | | | | | |
| 4 | White | 0.73 | 0.80 | 0.81 | 0.73 | 0.71 | 0.81 |
| | African American | 0.64 | 0.73 | 0.75 | 0.62 | 0.63 | 0.75 |
| | Hispanic | 0.70 | 0.79 | 0.79 | 0.71 | 0.70 | 0.81 |
| | Asian | 0.78 | 0.84 | 0.85 | 0.77 | 0.77 | 0.84 |
| | American Indian | 0.65 | 0.77 | 0.82 | 0.67 | 0.66 | 0.80 |
| | Two or More | 0.76 | 0.82 | 0.83 | 0.75 | 0.74 | 0.84 |
| 5 | White | 0.74 | | | | | |
| | African American | 0.61 | | | | | |
| | Hispanic | 0.70 | | | | | |
| | Asian | 0.78 | | | | | |
| | American Indian | 0.70 | | | | | |
| | Two or More | 0.75 | | | | | |
| 6 | White | 0.75 | | | | | |
| | African American | 0.65 | | | | | |
| | Hispanic | 0.72 | | | | | |
| | Asian | 0.78 | | | | | |
| | American Indian | 0.71 | | | | | |
| | Two or More | 0.77 | | | | | |
| 7 | White | 0.74 | | | | | |
| | African American | 0.64 | | | | | |
| | Hispanic | 0.71 | | | | | |
| | Asian | 0.78 | | | | | |
| | American Indian | 0.70 | | | | | |
| | Two or More | 0.76 | | | | | |
| 8 | White | 0.73 | 0.77 | 0.79 | 0.73 | 0.70 | 0.79 |
| | African American | 0.60 | 0.69 | 0.69 | 0.59 | 0.56 | 0.69 |
| | Hispanic | 0.68 | 0.75 | 0.77 | 0.68 | 0.66 | 0.76 |
| | Asian | 0.77 | 0.80 | 0.83 | 0.78 | 0.76 | 0.82 |
| | American Indian | 0.60 | 0.74 | 0.74 | 0.63 | 0.61 | 0.74 |
| | Two or More | 0.73 | 0.77 | 0.78 | 0.74 | 0.70 | 0.79 |

**Table 9-18. Correlations between Content Area Scale Scores by English Proficiency Status**

| Grade | Demographic Group | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|---|
| 3 | Fully English Proficient | 0.76 | | | | | |
| | Limited English Proficient | 0.70 | | | | | |
| 4 | Fully English Proficient | 0.76 | 0.82 | 0.83 | 0.76 | 0.75 | 0.84 |
| | Limited English Proficient | 0.67 | 0.76 | 0.76 | 0.67 | 0.66 | 0.78 |
| 5 | Fully English Proficient | 0.76 | | | | | |
| | Limited English Proficient | 0.61 | | | | | |
| 6 | Fully English Proficient | 0.77 | | | | | |
| | Limited English Proficient | 0.62 | | | | | |
| 7 | Fully English Proficient | 0.76 | | | | | |
| | Limited English Proficient | 0.59 | | | | | |
| 8 | Fully English Proficient | 0.74 | 0.78 | 0.80 | 0.75 | 0.72 | 0.80 |
| | Limited English Proficient | 0.57 | 0.66 | 0.67 | 0.56 | 0.54 | 0.65 |

**Table 9-19. Correlations between Content Area Scale Scores by Economic Status**

| Grade | Demographic Group | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|---|
| 3 | Not Economically Disadvantaged | 0.73 | | | | | |
|  | Economically Disadvantaged | 0.72 | | | | | |
| 4 | Not Economically Disadvantaged | 0.73 | 0.80 | 0.81 | 0.74 | 0.72 | 0.82 |
|  | Economically Disadvantaged | 0.72 | 0.80 | 0.81 | 0.72 | 0.72 | 0.82 |
| 5 | Not Economically Disadvantaged | 0.75 | | | | | |
|  | Economically Disadvantaged | 0.71 | | | | | |
| 6 | Not Economically Disadvantaged | 0.75 | | | | | |
|  | Economically Disadvantaged | 0.72 | | | | | |
| 7 | Not Economically Disadvantaged | 0.75 | | | | | |
|  | Economically Disadvantaged | 0.71 | | | | | |
| 8 | Not Economically Disadvantaged | 0.73 | 0.77 | 0.79 | 0.74 | 0.71 | 0.80 |
|  | Economically Disadvantaged | 0.68 | 0.75 | 0.76 | 0.69 | 0.65 | 0.77 |

**Table 9-20. Correlations between Content Area Scale Scores by Disability Status**

| Grade | Demographic Group | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|---|
| 3 | Not Disabled | 0.75 | | | | | |
|   | Disabled | 0.73 | | | | | |
| 4 | Not Disabled | 0.75 | 0.82 | 0.83 | 0.76 | 0.74 | 0.83 |
|   | Disabled | 0.71 | 0.79 | 0.78 | 0.71 | 0.71 | 0.81 |
| 5 | Not Disabled | 0.75 | | | | | |
|   | Disabled | 0.67 | | | | | |
| 6 | Not Disabled | 0.75 | | | | | |
|   | Disabled | 0.66 | | | | | |
| 7 | Not Disabled | 0.75 | | | | | |
|   | Disabled | 0.65 | | | | | |
| 8 | Not Disabled | 0.72 | 0.77 | 0.79 | 0.74 | 0.71 | 0.80 |
|   | Disabled | 0.60 | 0.71 | 0.70 | 0.62 | 0.57 | 0.72 |

**Table 9-21. Partial Correlations between Content Area Scale Scores**

| Grade | ELA & Mathematics | ELA & Science | ELA & Social Studies | Mathematics & Science | Mathematics & Social Studies | Science & Social Studies |
|---|---|---|---|---|---|---|
| 3 | 0.69 | | | | | |
| 4 | 0.68 | 0.77 | 0.78 | 0.68 | 0.66 | 0.79 |
| 5 | 0.68 | | | | | |
| 6 | 0.68 | | | | | |
| 7 | 0.68 | | | | | |
| 8 | 0.66 | 0.73 | 0.74 | 0.67 | 0.62 | 0.75 |

**Table 9-22. Comparison of Most Recent Wisconsin NAEP and Spring 2022 Wisconsin Forward Exam Impact Data**

| Content | Grade | NAEP Year | Below Basic | Basic | Proficient | Advanced | At or Above Proficient | At or Above Basic | Below Basic | Basic | Proficient | Advanced | At or Above Proficient | At or Above Basic |
|---------|-------|-----------|-------------|-------|------------|----------|------------------------|-------------------|-------------|-------|------------|----------|------------------------|-------------------|
| Reading/ELA | 4 | 2019 | 34 | 30 | 26 | 9 | 36 | 66 | 26.66 | 31.74 | 33.63 | 7.96 | 41.60 | 73.34 |
| Reading/ELA | 8 | 2019 | 24 | 38 | 34 | 4 | 39 | 76 | 29.81 | 36.72 | 26.45 | 7.02 | 33.46 | 70.19 |
| Math | 4 | 2019 | 20 | 35 | 34 | 11 | 45 | 80 | 22.24 | 34.03 | 30.95 | 12.79 | 43.73 | 77.76 |
| Math | 8 | 2019 | 24 | 34 | 29 | 12 | 41 | 76 | 35.96 | 33.73 | 24.05 | 6.26 | 30.31 | 64.04 |
| Science | 4 | 2015 | 21 | 38 | 40 | 1 | 41 | 79 | 18.20 | 30.95 | 32.32 | 18.53 | 50.85 | 81.80 |
| Science | 8 | 2015 | 25 | 35 | 38 | 2 | 40 | 75 | 20.77 | 30.33 | 28.99 | 19.91 | 48.90 | 79.23 |

Note: The NAEP assessed student knowledge and skills in Reading, while the Wisconsin Forward Exam assessed student knowledge and skills in ELA, which included Reading, Listening, and Writing.

Note: NAEP data are from https://nces.ed.gov/nationsreportcard.

# Part 10: Test Results

Part 10 of the Technical Report provides short descriptions of the Wisconsin Forward Exam score reports and interpretive guide. It also presents a summary of student test results for the Spring 2022 Wisconsin Forward Exam administration. The summary results are presented for all Wisconsin students and cover four types of reported scores: total test scale scores; total test performance levels; scores based on each of the content standards within each content area, which are called standard performance index (SPI) scores; and performance levels based on SPI scores. The four types of scores offer the reader several points from which to understand and evaluate the performance of Wisconsin students on the Wisconsin Forward Exam. In addition, the longitudinal test participation rates and test results are presented in this part of the report. The AERA, APA, & NCME (2014) Standards addressed in Part 8 include 5.1, 6.10, 7.0, 7.1, and 12.18.

## 10.1 Types of Reports

Score reports are the primary means of communicating test scores to relevant district personnel (e.g., district assessment coordinators, superintendents), teachers, and parents. AERA, APA, & NCME (2014) Standard 6.10 states the following:

> When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used. (p. 119)

> **Standard 5.1** is related in that it states the following:

> Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations. (p. 102)

Interpretations related to the test scores are disseminated in two ways: (1) the individual score report and (2) the *User's Guide to Interpreting Reports* (DRC, 2022).

In addition to providing an explanation of the intended interpretation of test scores, a testing program must also ensure that the information related to the test scores is understandable by the target audience. Standards 7.0 and 7.1 of the Standards (AERA, APA, & NCME, 2014) state the following:

> **Standard 7.0** Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores. (p. 125)

> **Standard 7.1** The rationale for a test, recommended uses of the test, support for such uses, and information that assists in score interpretation should be documented. When particular misuses of a test can be reasonably anticipated, cautions against such misuses should be specified. (p. 125)

In support of Standards 7.0 and 7.1, the User's Guide to Interpreting Reports is accessible to parents, teachers, and the general public at https://dpi.wi.gov/assessment/forward/data#resources.

In the 2021–22 administration year, DPI reported the Wisconsin Forward Exam results in WISEdash Public and WISEdash for Districts. These dashboards provide comprehensive data analysis for statewide assessments, attendance, graduation, coursework and other data of interest to district, school and other

statewide data users. DRC reported the Wisconsin Forward Exam results through the Wisconsin Forward Exam Reporting System, which is a browser-based system designed to deliver online interactive reporting to authorized users at the state and district levels for Wisconsin schools.

## 10.1.1 Description of Each Type of Report

In this section, descriptions of the following reports are provided: Individual Student Report (ISR), Student Roster, Summary by Subject, and Summary by Reporting Category. In compliance with AERA, APA, & NCME (2014) Standard 12.18, the Wisconsin Forward Exam score reports provide clear information about the achievements of individual students and groups of students. Standard 12.18 states the following:

> In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports. (p. 200)

### Individual Student Report

The Individual Student Report (ISR) is one of the types of reports available through the Wisconsin Forward Exam Reporting System. The ISR is the primary means for sharing student test results with parents and guardians. It is a stand-alone document, giving parents or guardians relevant information that enables them to understand their child's test score. The ISRs are provided to schools to be sent home to parents or guardians. The ISR consists of three to four pages (depending on how many tests a student took). On the upper-left side of the first page, the student's identifying information is provided. Underneath the student information, there is a short description of the Wisconsin Forward Exam and the purpose of the report. On the upper-right side of the first page there is a table with information on the student's performance level in each content area as well as the student's percentile rank in that content. Under the table summarizing the student's proficiency classification, a short description of the Wisconsin Forward Exam performance levels is presented.

The second page of the report includes the presentation of the total test scale scores and the performance levels for ELA and Mathematics. This information is presented in the form of a bar graph, and the student's scale score for a given content area is shown, along with the performance level associated with that scale score. The total test results are followed by the reporting category results presented for that content area. These results include number of points obtained, number of points possible, SPI score, and the SPI proficiency classification for each reporting category. The third page of the report includes the total test and reporting category results for Science and Social Studies for students in grades 4, 8, and 10 (the grades at which these assessments are administered).

The last page of the ISR includes information on accommodations and designated support use, Wisconsin Academic Standards, and a short explanation of the reported scores. An example of an ISR can be found in the *User's Guide to Interpreting Reports*.

### Roster

Another report available from the Wisconsin Forward Exam Reporting System is an online interactive Roster that displays a list of students based on the specific report filter options selected, such as test administration, grade, school, district, gender, race/ethnicity, disability status, and English proficiency status. Total test scale scores and performance level indicators, as well as the reporting category performance levels, are displayed in a table format for the content area chosen.

## Subject Summary Report

The Subject Summary online interactive report contains performance level information for the school, district, and state. It includes a mean scale score along with numeric and graphic representations of the performance level summary for the subject and grade. This report also includes the number and percentage of students in each of the performance levels.

## Reporting Category Summary Report

The online interactive Reporting Category Summary Report includes performance level information based on the individual reporting category selected for the school, district, and state. By selecting any one of the available reporting categories from the Reporting Category filter, the table and chart data will be based on that category for the subject and grade chosen. Along with numeric and graphic representations of the performance level by reporting category, subject, and grade, the report includes the number and percentage of students in each of the performance levels for the reporting category selected.

## Demographic Summary Report

The online interactive Demographic Summary Report provides at-a-glance comparisons of performance between various demographic subgroups. The percentages of students in each performance level, means scale scores, as well as the test participation rates are presented graphically and numerically for each subgroup within each demographic category. The summary table shows the number of students tested, the mean scale scores, and the percentage of students in each performance level for each demographic subgroup combination.

Examples of Roster, Subject Summary, Reporting Category Summary, and Demographic Summary reports are presented in the *User's Guide to Interpreting Reports* available on DPI's website.

## 10.1.2 Interpreting Test Results

A student's correct responses to the assessment questions are used to derive that student's Wisconsin Forward Exam scale score. The scale score describes performance on a continuum that spans the complete range of grades 3–8 for ELA and Mathematics. These scores range in value from 330 to 970 for ELA and from 360 to 890 for Mathematics. Because ELA and Mathematics assessments are on vertical scales, scores from adjacent grades may be compared within a content area. For example, it is appropriate to compare a student's grade 5 Mathematics scale score with the student's grade 6 Mathematics scale score in a subsequent administration year. ELA and Mathematics scale scores can also be compared within a content area across the administrations from Spring 2016 to Spring 2022.

Science scale scores range from 300 to 725 for grade 4 and from 480 to 945 for grade 8. Science scores can be compared within a grade level, but since Science assessments are not on a vertical scale, the scale scores cannot be compared across grades. Because new reporting scales were developed for the Science assessments in Spring 2019, the Science scale scores from the current administration can only be compared with the scores from the Spring 2019 and 2021 administrations.

Social Studies scale scores range from 330 to 700 for grade 4, from 540 to 860 for grade 8, and from 645 to 980 for grade 10. Social Studies scores can be compared within a grade level, but since the assessments are not on a vertical scale, the scale scores cannot be compared across grades. New Social Studies scales were established in Spring 2022, and this year's scores are considered a new baseline. As such, the Spring 2022 scale scores cannot be compared with the Social Studies scores from previous administrations.

Scale scores cannot be compared across content areas. For example, it is not appropriate to compare a student's Mathematics and ELA scores as they do not represent comparable achievement.

The Wisconsin Forward Exam scale scores determine a student's performance level. Student performance is reported in terms of four performance levels that describe a pathway to proficiency and college and career readiness. Each performance level represents standards of performance for each assessed content area. Performance level scores provide a description of what students can do in terms of the content and skills assessed, as described in the Wisconsin Academic Standards.

In addition to the total test score, students receive scores in each reporting category of the test taken. The reporting category scores are SPI scores and performance levels. The SPI is an estimate of the number of questions that a student could be expected to answer correctly if there had been 100 such questions measuring that content standard on the test in a given administration year. More information on the SPI scores is provided in Section 10.4 of this report.

Last but not least, state percentile ranks are computed for each student based on the student's total test scale score. The state percentile ranks, ranging from 1 to 99, provide information that compares the student's achievement with that of a larger reference group, the state. The percentile rank tables for the most recent test administration can be found on DPI's website at https://dpi.wi.gov/assessment/forward/data.

Information on score interpretation is included in the *User's Guide to Interpreting Reports*, which was written for Wisconsin teachers and administrators who received score reports from the 2021–22 administration of the Wisconsin Forward Exam. The *Guide* was developed collaboratively by DRC and DPI staff.

## 10.2 Scale Scores Summary Statistics

The primary scores reported in Wisconsin Forward Exam program reports are scale scores. The scale score of a student in a given content area represents the student's level of performance in that content area. Higher scale scores indicate higher levels of performance, and lower scale scores indicate lower levels of performance. Scale scores are based on the entire set of scored operational items per grade and content area.

Summary descriptive statistics based on the scale score results are described below.

Table 10-1 is the summary scale score table based on the Spring 2022 census data. The table shows the following: mean scale score, standard deviation of the scale scores, skewness and kurtosis, minimum and maximum obtained scale scores, and lowest and highest obtainable scale scores (LOSS and HOSS, respectively) for all content areas and grades based on the census data (all students with valid test scores). The LOSS and HOSS, as discussed in Part 6, identify the lower and upper limits of the scale score range. These values were established when the current scales were developed and do not change from one administration to another.

### English Language Arts

- Mean scale score increased as grade level increased, ranging from 549.75 for grade 3 to 622.78 for grade 8. This mean scale score pattern supports the ELA vertical scale properties.

- Standard deviations ranged from 47.41 to 61.00 scale score points, across grades.

- Student scores spanned the full-scale score range from the LOSS to the HOSS in grades 5, 6, and 8. No student reached the HOSS in grades 3, 4, or 7.

## Mathematics

- Mean scale score increased as grade level increased, ranging from 552.25 for grade 3 to 634.95 for grade 8. This mean scale score pattern supports the Mathematics vertical scale properties.

- Standard deviations ranged from 52.67 to 62.13 scale score points, across grades.

- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

## Science

- Mean scale scores were 495.81 and 693.86 for grades 4 and 8, respectively.

- Standard deviations were 52.11 and 51.11 scale score points for grades 4 and 8, respectively.

- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

## Social Studies

- Mean scale scores were 499.85, 699.24, and 798.49 for grades 4, 8, and 10, respectively.

- Standard deviations ranged from 50.27 to 50.80 scale score points.

- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

## 10.2.1 Subgroup Performance Patterns in Scale Score Results

The scale score results, like the raw score results, showed some consistent performance patterns in terms of subgroups. The results for gender, race/ethnicity, economic status, disability status, English language proficiency, and accommodation use are presented in Tables 10-2 through 10-10. The scale score statistics were computed based on the census data.

## Gender

- In ELA, male students as a group showed lower mean scale scores than female students as a group in each grade level. The difference ranged from about 8 scale score points in grade 3 to 16 scale score points in grade 8 or between one-fifth and one-quarter of a standard deviation.

- In Mathematics, male students as a group showed higher mean scale scores in grades 3 through 7 (differences between 1 and 5 scale score points) and a lower mean scale score in grade 8 (difference of approximately 3 scale score points) than female students. All differences were less than one-tenth of a standard deviation.

- In Science, the mean scale scores were only slightly lower for female students, with a difference of about 2 scale score points in both grades. These differences were less than one-twentieth of a standard deviation.

- In Social Studies, there was no practical difference between female and male student mean scale scores in grade 4. Female students performed better than male students in grades 8 and 10, with a difference between mean scale score of approximately 2 points in grade 8 (less than one-twentieth of a standard deviation) and about 5 scale score points in grade 10 (approximately one-tenth of a standard deviation).

## Race/Ethnicity

- The scale score results showed some consistent performance differences by ethnicity.

- With two exceptions, White students as a group had the highest mean scale scores, followed by Asian students, Hispanic students, American Indian students, and African American students. The exceptions were ELA grade 8 and Mathematics grade 8, where White and Asian students had comparable scale score means.

- The mean scale scores of African American students were typically more than one standard deviation lower than the mean scale scores of White students. The mean scale scores of Hispanic and American Indian students were approximately two-thirds to three-quarters of a standard deviation lower than the mean scale scores of White students in all grades and content areas. The mean scale scores of Asian students in grades 3 through 5 were approximately one-quarter to one-third of the standard deviation lower than the mean scale score of White students, regardless of the content area. The mean scores of Asian students in grades 6 through 8 and 10 were comparable with the mean scores of White students within about one-tenth of a standard deviation across all content areas.

## Economic Status

- Economically disadvantaged students as a group scored lower than students who were not economically disadvantaged as a group across all grades and content areas. Differences ranged from over 30 scale score points in all ELA, Science, and Social Studies grades to over 40 scale score points in Mathematics grades 3, 4, 6, 7 and 8.

- For every grade and content area, the mean scale scores of students who were economically disadvantaged were typically two-thirds to three-quarters of a standard deviation lower than the mean scale scores of students who were not economically disadvantaged.

## Disability Status

- Students with disabilities and students without disabilities showed consistent and large differences in mean scale scores by group. These differences tended to increase as grade level increased. For example, the mean scale score difference by disability status ranged from approximately 33 scale score points in ELA grade 3 to about 64 scale score points in ELA grade 8 and from about 42 scale score points in Mathematics grade 3 to about 60 points in Mathematics grades 7 and 8.

- For every grade and content area, the mean scale scores of students with disabilities were lower than the mean scale scores of students without disabilities by about three-quarters of a standard deviation (in lower grades) to over one standard deviation (in higher grades).

## English Language Proficiency

- Students who were fully English proficient and students of limited English proficiency showed consistent and large differences in mean scale scores by group. These differences were larger in grades 6 through 8 compared to lower grades. For example, the mean scale score difference by English language proficiency status ranged from approximately 29 scale score points in ELA grade 3 to about 48 scale score points in ELA grades 7 and 8, and from about 32 to 37 scale score points in Mathematics grade 3 through 5 to approximately 46 to 51 points in Mathematics grades 6 through 8.

- For every grade and content area, the mean scale scores of students with limited English proficiency were lower than the mean scale scores of fully English proficient students by two-thirds of a standard deviation to approximately one standard deviation.

## Accommodation Use

- Students using testing accommodations (listed in Part 4 this report) performed less well on the tests compared to their peers not using testing accommodations. The differences ranged from approximately 22 to 36 scale score points for ELA and from 66 to 77 scale score points for Mathematics. Note that the comparison of the mean scale scores for students using and not using testing accommodations is less reliable for ELA than for Mathematics due to the fact that only 82 to 113 students per grade used ELA testing accommodations.

- The mean scale scores of students using testing accommodations were lower than the mean scale scores of students not using testing accommodations by about one-third to three-quarters of a standard deviation for ELA grades and by more than one standard deviation in each Mathematics grade.

- Science and Social Studies student performance for students using testing accommodations were not compared with the performance of their peers not using testing accommodations. Fewer than 50 students per grade used testing accommodations in these content areas.

## 10.3 Performance Level Classifications

Student performance on the Wisconsin Forward Exam is reported in terms of four performance categories: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. These performance categories are established through cut scores.

Standard 5.21 of the Standards (AERA, APA, & NCME, 2014) indicates that "when proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly" (p. 107).

In terms of the validity of the Wisconsin Forward Exam, it is essential to understand that cut scores and PLDs are established in a collaborative and participatory process. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, particularly cut scores. PLDs summarize the knowledge, skills, and abilities expected of students in each achievement level. As stated in Part 7, DPI provided policy PLDs for the Wisconsin Forward Exam assessments. At the standard setting, DPI used the policy PLDs in conjunction with the content standards to consider the content-based expectations for students in each achievement level on each test in the Wisconsin Forward Exam program.

Tables 10-11 through 10-14 provide the scale score ranges that define performance levels together with the percentage of students in each performance level. The results for each content area and grade are summarized below.

### English Language Arts

- Between approximately 33% (grade 8) and 42% (grade 4) of students were either *Proficient* or *Advanced*.

- Between about 5% and 8% of students were classified as *Advanced*, depending on the grade level.

- Across all grade levels, more than 55% of students were below *Proficient*. These percentages ranged from approximately 58% below *Proficient* in grade 4 to approximately 67% below *Proficient* in grade 8.

## Mathematics

- Between approximately 30% (grade 8) and 47% (grade 3) of students were either *Proficient* or *Advanced*.

- The percentage of students who were *Advanced* ranged from approximately 4% (grade 7) to over 12% (grades 3 and 4).

- Across all grade levels, the percentages of students below *Proficient* ranged from approximately 53% in grade 3 to 70% in grade 8.

## Science

- Approximately 51% of students in grade 4 and about 49% of students in grade 8 were either *Proficient* or *Advanced*.

- The percentage of students classified as *Advanced* was approximately 19% in grade 4 and about 20% in grade 8.

- The percentage of students classified as below *Proficient* was approximately 49% in grade 4 and about 51% in grade 8.

## Social Studies

- Approximately 59% of students in grades 4 and 8 and about 48% of students in grade 10 were classified as *Proficient* or *Advanced*.

- Between 22% and 24% of students were classified as *Advanced* across the three grades.

- The percentage of students classified as below *Proficient* ranged from approximately 41% in grade 4 to 52% in grade 10.

### 10.3.1 Subgroup Patterns in Performance Level Results

The performance level results varied by subgroup: gender, race/ethnicity, English language proficiency, disability status, economic status, and accommodation use. The main subgroup performance patterns are described below. These comparisons are based on Tables 10-15 through 10-18.

In terms of gender, higher percentages of female students were classified as *Proficient* or above in all ELA grades. The differences in the percentages of male and female students in the *Proficient* or above categories ranged from approximately 6% (grades 3, 4, and 7) to about 9% (grades 6 and 8) for ELA. A reversed trend was observed for Mathematics. Between approximately 2.5% (grade 6) and 6% (grade 4) more male students than female students were classified as *Proficient* or above in grades 3 through 7. The difference in Mathematics grade 8 was about half a percent. In Science, more male students than female students were classified as *Proficient* or above in both grades (with a difference of approximately 1%). In Social Studies, the differences in percentages of male and female students classified as *Proficient* or above were about 2% in grade 8 and close to 5% in grade 10, with more female students being classified in the two highest performance categories in these grades. There was no practical difference in the percentages of male and female students classified as *Proficient* or above in grade 4.

There were some consistent patterns in performance by ethnicity across grades and content areas. This pattern followed the pattern of the mean scale score differences. In terms of the *Proficient* or above categories, the prevailing tendency was that there were higher percentages of White students as a group, followed by Asian students, Hispanic students, American Indian students, and African American students. The inverse sequence was found at the *Below Basic* performance level.

Performance level results showed that there were higher percentages of fully English proficient students who were classified as *Proficient* or above compared to students who were of limited English proficiency in every grade and content area. These differences ranged from approximately 25% in ELA grade 3 to over 41% in Social Studies grade 10.

Performance level results showed a similar pattern in comparisons for students without disabilities who were classified as *Proficient* or above compared to students with disabilities, with differences ranging from approximately 24% (ELA grade 3) to 40% (Science grade 8) depending on grade level and content area.

There were consistent differences in performance between economically disadvantaged students and not economically disadvantaged students. In every grade and content area, between approximately 24% (ELA grade 8) and over 30% (Mathematics grades 3 through 6, Science grade 4, and Social Studies grade 4) more students who were not economically disadvantaged were classified as *Proficient* or above compared to their economically disadvantaged peers.

Performance level results showed that there were higher percentages of students not using testing accommodations who were classified as *Proficient* or above compared to students using testing accommodations. These differences ranged from approximately 12% to 27% across ELA grades and from 31% to 42% across Mathematics grades. The differences in the percentages of students in different performance levels between groups of students using and not using testing accommodations should be interpreted with caution for ELA due to the fact that only 113 students or fewer per grade used ELA testing accommodations. Similarly, the comparisons of percentages of students using and not using testing accommodations for Science and Social Studies are less reliable due to that fact that fewer than 50 students per grade used testing accommodations in these content areas.

## 10.4 Standard Performance Index for Content Standards

In addition to raw scores and scale scores, teachers and educational decision makers frequently need diagnostic information to inform instructional strategies. Diagnostic information also helps to identify individual student strengths and needs. This kind of information can be derived from scores on subsets of test items that estimate how much a student knows in a clearly defined skill domain. These skill domains are called content standards or standards and they reflect Wisconsin Forward Exam reporting categories. Scores on subsets of test items at the content standard level are called standard performance index (SPI) scores. The purpose of reporting SPI scores on the Wisconsin Forward Exam is to show the relationship between the overall achievement being measured (represented by the test score) and the skills within each of the content standards associated with the overall content area. Teachers may use the SPI scores for individual students as indicators of strengths and weaknesses, but the SPI scores are best corroborated by other evidence, such as homework, class participation, diagnostic test scores, or observation. District and school administrators may compare their results by content standard and grade level with the state mean percentage to better understand their strengths and weaknesses within a particular content area and grade level.

An SPI score can be interpreted as an estimate of the number of items a student would be expected to answer correctly if there had been 100 similar items for a given reporting category. For example, an SPI score of 77 for a given reporting category means that, if the student were given 100 similar items, the student would be expected to answer 77 of them correctly. This is a criterion-referenced score because it estimates how much a student knows in a clearly defined skill domain (i.e., the criterion). Technical readers can refer to Appendix K of this report for more details.

This approach—identifying student proficiency on each content standard—relates to the ELA and Mathematics Wisconsin Academic Standards and Wisconsin Standards for Science and Social Studies. SPI scores provide a more reliable estimate of student achievement on each content standard than is possible by simply reporting the percentage correct. However, *SPI scores should be used for low-stakes purposes because these scores cannot be considered stable for any content standard with a small number of items.*

Readers should note that the average difficulty of items will vary across content standards and grades. Content standards vary in their complexity, level of abstraction, and cognitive demand. Some standards may be intrinsically more difficult than others, and the difficulty of individual items is determined, in part, by the difficulty of the content domain being measured. The current test blueprints do not specify the average difficulty level of items for each content standard within grades or across grades. If the difficulty of the items varies across years, grades, or content standards, the mean SPI scores will be affected by differences in item difficulty as well as differences in student ability. *Thus, differences in SPI scores across years, grades, or content standards should not be seen as reliable indicators of differences in student ability since these differences may be explained in whole or in part by differences in the difficulty of the items themselves.*

In general, comparisons across years, grades, or content standards are appropriate for assessing the relative difficulty of the items. Comparisons of individual student scores or group mean scores can provide useful information about the *relative* strengths and needs of individual students or groups on these standards in a given administration year.

Tables 10-19 through 10-23 identify the content standards or domains, the number of MC and non-MC items within each standard or domain, the total number of possible points per standard or domain, the mean raw score, the mean *p*-value, the standard deviation of the raw scores, the mean SPI score, and the standard deviation of SPI scores for all content areas across grades. The results from Tables 10-19 through 10-23 are summarized below. Tables 10-24 through 10-27 identify the SPI cut scores for each content area reporting category and grade level.

## English Language Arts

Tables 10-19 and 10-20 present mean *p*-values and SPI scores for ELA across content standards and across domains, respectively. Results show that the mean ELA SPI scores across grades ranged from approximately 40 to 69 for content standards (including Listening) and from about 47 to 61 for domains (Reading or Writing), indicating that the items were, on average, moderately difficult for examinees. Some patterns of standard difficulty were found across ELA grades and standards. Content standard C (Reading—Vocabulary Use) was found to be easier than other standards in grades 4, 6, 7, and 8 with a mean SPI of at least 60 or higher. On the other hand, content standard D (Writing/Language—Text Types and Purposes) tended to be more difficult than other content standards in grades 3 through 7, with a mean SPI for this standard ranging from 40 to 48 in these grades. No particular pattern of difficulty was detected across other standards.

The Writing domain was more difficult than the Reading and Listening domains for students in grades 3 through 6 (with mean SPI scores ranging from 47 to 52 for the Writing domain), and the Reading domain was more difficult than the other two domains in grades 7 and 8 (with mean SPI scores of approximately 50 and 53, respectively). The Listening domain was the easiest of the three domains in grades 3 and 5 through 8 (with mean SPI scores ranging from 56 to 65).

## Mathematics

Table 10-21 presents mean $p$-values and SPI scores for Mathematics across grades and content standards. Results show that the mean $p$-values and SPI scores varied across standards in all grades. Mean SPI scores across all content standards ranged from approximately 39 for content standard E (Geometry) and content standard G (The Number System) in grade 8 to about 58 for content standard A (Operations and Algebraic Thinking) in grade 3. The Mathematics items were more challenging in higher grades than lower grades. No particular pattern of difficulty was detected across Mathematics standards within a grade.

## Science

Table 10-22 presents mean $p$-values and SPI scores for Science across grades and content standards. The mean Science SPI scores across both grades and all content standards ranged from approximately 42 to 62, indicating that the test items were of medium difficulty, on average. Mean SPI scores indicated that content standard C (Earth and Space Science) was the most difficult in grade 4 and that content standards B (Physical Science) and C (Earth and Space Science) were the most difficult in grade 8.

## Social Studies

Table 10-23 presents mean $p$-values and SPI scores for Social Studies across grades and content standards. The mean Social Studies SPI scores across all grades and content standards ranged from approximately 47 to 62, indicating that the test items were, on average, of medium difficulty. The mean SPI scores indicated that the most difficult content standard varied between the three Social Studies grades. The most difficult standard in each grade was content standard D (Economics) in grade 4, content standard B (History) in grade 8, and content standard E (Behavioral Sciences) in grade 10.

## Summary of Student Performance Indicator Results

Overall, the mean SPI scores across grades and content standards ranged in difficulty from approximately 39 (standard G, The Number System and standard E, Geometry in Mathematics grade 8) to 69 (standard C, Reading—Vocabulary Use in ELA grade 7).

It is important to note that some variation in the difficulty of the items across content standards within and across grades and test forms is inevitable and that some of that variation is independent of any intrinsic differences in the difficulty of the standards themselves (e.g., variations in the difficulty of the particular items that were selected for the test forms). For this reason, SPI scores should be interpreted with caution and should not be used to make comparisons of student performance across testing years or grade levels.

# 10.5 Longitudinal Comparisons of Test Scores

It is often desirable to examine the scores of students across time and to monitor group performance. This is possible if the test content and the construct measured by the test are comparable from year to year and if the scores are reported on the same scale in multiple years.

Six years of test scores on the same reporting scales are available for the Wisconsin Forward Exam assessments in ELA and Mathematics. The state-level mean scale scores and standard deviations for the 2016, 2017, 2018, 2019, 2021, and 2022 administrations are presented for ELA and Mathematics in Tables 10-28 and 10-29, respectively. New scales were established for the Science assessments after the Spring 2019 test administration. Because the new Science assessments were not linked to the previous scales, the Spring 2022 scale scores are comparable only with the Spring 2019 and 2021 scale scores and not with the scores prior to the Spring 2019 administration. Therefore, only three years of scale score data are presented for Science in Table 10-30. New scales were established for the Social Studies assessments after the Spring 2022 test administration. Because the new Social Studies assessments were not linked to the previous scales, the Spring 2022 scale scores are not comparable with the previous administration scores. Therefore, only one year of scale score data is presented for Social Studies in Table 10-31. The Spring 2022 student performance in Social Studies is a new baseline for longitudinal comparisons. The historical Social Studies data can be found in the *Wisconsin Forward Exam Spring 2021 Technical Report* available at DPI's website: https://dpi.wi.gov/assessment/forward/resources#documentation.

The statistics presented in Tables 10-28 through 10-31 are based on the total population of Wisconsin students, including students attending public, choice, and private schools. The test participation rates for each year are also included in Tables 10-28 through 10-31. The participation rates are computed as the percentage of students who received a valid scale score given the number of students expected to take the test in each grade and content area. The "Enrolled" column shows the total number of students expected to take the test each year. The "Number Tested" and "Percent Tested" columns show the number and percentage of students who participated in the assessment and received a valid scale score.

It should be noted that students who participated in the Spring 2021 test administration were not fully representative of the Wisconsin student population. As such, the interpretation of the state-level performance trend changes between Spring 2019 and Spring 2021 and between Spring 2021 and Spring 2022 should be done with caution and in the context of disrupted learning due to the COVID-19 pandemic in the 2020–21 academic year. Therefore, in this report we focus on the differences in student performance between the Spring 2019 and Spring 2022 administrations, rather than a comparison of student performance in the last two administrations.

It was observed that the mean scale scores for ELA were consistently lower for all grades in Spring 2022 compared to Spring 2019. Smaller differences between administrations, ranging from approximately 2 to less than 5 scale score points, were observed in grades 3 through 6 compared to differences of approximately 6 scale score points in grade 8 and close to 9 scale score points in grade 7 (see Table 10-28).

Similar to ELA, the Mathematics mean scale scores decreased for all grades between Spring 2019 and Spring 2022. The decreases in mean scale scores ranged from approximately 2 to 2.5 scale score points for grades 3 through 5 and from about 6 to 9.5 scale score points for grades 6 through 8 (see Table 10-29).

As with ELA and Mathematics, student performance declined for both Science grades between the last two administrations. The mean scale score decrease was about 4 points in grade 4 and close to 6 points in grade 8 (refer to Table 10-30).

Tables 10-32 through 10-35 show the percentages of students in each achievement level in the Spring 2016 through 2022 test administrations for ELA and Mathematics, in the Spring 2019 through 2022 test administrations for Science, and in the Spring 2022 test administration for Social Studies. The results presented in these tables are based on all Wisconsin students who participated in the assessments in a given year, including students attending public, choice, and private schools. The pattern of student performance

classification change between Spring 2019 and Spring 2022 was consistent with the pattern of the mean scale score change between the two test administrations in all grades of ELA, Mathematics, and Science.

For ELA, a decrease in the percentage of students at or above the *Proficient* cut was observed for all grades except for grade 5. The decrease ranged from approximately 1% in grade 4 to over 6% in grade 7. The change in the percentage of students at or above the *Proficient* cut score was negligible for grade 5 (see Table 10-32). For Mathematics, the decrease in the percentage of students classified as *Proficient* or above ranged from approximately 1% in grade 4 to over 5% in grades 7 and 8 (see Table 10-33).

For Science, approximately 2% fewer students were classified as *Proficient* or above in Spring 2022 compared to Spring 2019 in grade 5. About 5% fewer students were classified in the two highest performance levels in grade 8 between the two administrations (see Table 10-34). As stated earlier in the report, new performance level cut scores were established for Science after the Spring 2019 test administration, which is considered to be a new baseline for longitudinal comparisons. Therefore, no impact data from administrations prior to 2019 are presented in this report. The historical Science data can be found in the *Wisconsin Forward Exam Spring 2018 Technical Report* available at DPI's website: https://dpi.wi.gov/assessment/forward/resources#documentation.

As stated earlier in the report, new performance level cut scores were established for Social Studies after the Spring 2022 test administration. Therefore, the Spring 2022 Social Studies impact data should not be directly compared with the previous years' impact data. The Spring 2022 Social Studies impact data are a new baseline for longitudinal comparisons. The historical Social Studies data can be found in the *Wisconsin Forward Exam Spring 2021 Technical Report* available at DPI's website: https://dpi.wi.gov/assessment/forward/resources#documentation.

Overall, the percentages of students classified in the *Proficient* or above categories were found to be lower in Spring 2022 compared to Spring 2019 for each grade and content area, except for ELA grade 5 (where there was no difference). These results are not unexpected and likely reflect a slow recovery from the effects of the COVID-19 pandemic on student learning in the last two years.

## 10.6 Summary

In the Wisconsin Forward Exam, the purpose of the ELA, Mathematics, Science, and Social Studies assessments is to demonstrate student achievement through test scores in the respective content areas. The results presented in Part 10, together with the reliability and validity evidence presented in Parts 8 and 9, indicate that the scale scores and performance levels reported in the Wisconsin Forward Exam program are valid and reliable evidence of student achievement in the tested content areas and grades in Spring 2022. However, due to the circumstances related to the COVID-19 pandemic and lower participation in the assessment than in a typical year, we recommend that the results of the Spring 2021 test administration be treated and interpreted with caution. More reliable comparisons of student performance can be made between Spring 2019 and Spring 2022.

Classroom teachers may use the Spring 2022 scores as evidence of student achievement for students who participated in the assessment. District and school administrators may use this information for activities such as planning teaching activities in the next school year.

**Table 10-1. Scale Score Descriptive Statistics for Total Population**

| Content | Grade | N Count | Mean | SD | Skewness | Kurtosis | Min | Max | LOSS | HOSS |
|---|---|---|---|---|---|---|---|---|---|---|
| English Language Arts | 3 | 58275 | 549.75 | 47.41 | -0.13 | 0.29 | 330 | 810 | 330 | 900 |
| | 4 | 58773 | 578.32 | 52.20 | -0.16 | 0.21 | 340 | 862 | 340 | 930 |
| | 5 | 59044 | 593.86 | 52.07 | -0.22 | 0.43 | 350 | 940 | 350 | 940 |
| | 6 | 60112 | 602.26 | 51.02 | -0.38 | 0.35 | 360 | 950 | 360 | 950 |
| | 7 | 61871 | 618.80 | 56.30 | -0.10 | 0.21 | 370 | 923 | 370 | 960 |
| | 8 | 62684 | 622.78 | 61.00 | -0.20 | 0.30 | 380 | 970 | 380 | 970 |
| Mathematics | 3 | 58449 | 552.25 | 56.04 | -0.41 | 1.56 | 360 | 760 | 360 | 760 |
| | 4 | 58931 | 573.54 | 57.00 | -0.61 | 1.00 | 405 | 800 | 405 | 800 |
| | 5 | 59187 | 599.38 | 52.67 | -0.70 | 1.61 | 430 | 830 | 430 | 830 |
| | 6 | 60234 | 604.64 | 59.36 | -0.39 | 1.00 | 440 | 870 | 440 | 870 |
| | 7 | 61969 | 617.55 | 62.13 | -0.50 | 0.62 | 450 | 880 | 450 | 880 |
| | 8 | 62762 | 634.95 | 60.04 | -0.46 | 0.87 | 470 | 890 | 470 | 890 |
| Science | 4 | 58880 | 495.81 | 52.11 | 0.06 | 0.05 | 300 | 725 | 300 | 725 |
| | 8 | 62644 | 693.86 | 51.11 | 0.14 | 0.72 | 480 | 945 | 480 | 945 |
| Social Studies | 4 | 58833 | 499.85 | 50.27 | -0.06 | 0.81 | 330 | 700 | 330 | 700 |
| | 8 | 62606 | 699.24 | 50.80 | -0.50 | 1.27 | 540 | 860 | 540 | 860 |
| | 10 | 59391 | 798.49 | 50.78 | -0.40 | 0.79 | 645 | 980 | 645 | 980 |

**Table 10-2. Scale Score Descriptive Statistics by Gender**

| Content | Grade | Male | | | | | Female | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N Count | Mean | SD | Min | Max | N Count | Mean | SD | Min | Max |
| English Language Arts | 3 | 29613 | 545.70 | 47.55 | 330 | 768 | 28662 | 553.94 | 46.91 | 330 | 810 |
| | 4 | 30084 | 574.09 | 52.31 | 340 | 793 | 28689 | 582.76 | 51.72 | 340 | 862 |
| | 5 | 30093 | 588.88 | 52.41 | 350 | 853 | 28951 | 599.04 | 51.20 | 350 | 940 |
| | 6 | 30649 | 596.27 | 51.47 | 360 | 792 | 29463 | 608.49 | 49.79 | 372 | 950 |
| | 7 | 31682 | 613.35 | 56.80 | 370 | 923 | 30189 | 624.51 | 55.20 | 370 | 923 |
| | 8 | 32133 | 614.98 | 61.34 | 380 | 970 | 30551 | 630.98 | 59.55 | 380 | 970 |
| Mathematics | 3 | 29706 | 554.91 | 58.48 | 360 | 760 | 28743 | 549.49 | 53.27 | 360 | 760 |
| | 4 | 30161 | 575.99 | 58.75 | 405 | 800 | 28770 | 570.97 | 54.98 | 405 | 800 |
| | 5 | 30171 | 601.60 | 54.51 | 430 | 830 | 29016 | 597.06 | 50.58 | 430 | 830 |
| | 6 | 30732 | 605.03 | 61.19 | 440 | 870 | 29502 | 604.23 | 57.40 | 440 | 870 |
| | 7 | 31740 | 618.76 | 64.52 | 450 | 880 | 30229 | 616.29 | 59.50 | 450 | 880 |
| | 8 | 32172 | 633.49 | 62.66 | 470 | 890 | 30590 | 636.47 | 57.13 | 470 | 890 |
| Science | 4 | 30129 | 496.68 | 52.93 | 300 | 725 | 28751 | 494.90 | 51.21 | 300 | 725 |
| | 8 | 32104 | 694.85 | 54.09 | 480 | 945 | 30540 | 692.82 | 47.76 | 480 | 945 |
| Social Studies | 4 | 30104 | 499.77 | 51.68 | 330 | 700 | 28729 | 499.93 | 48.75 | 330 | 700 |
| | 8 | 32088 | 698.33 | 53.78 | 540 | 860 | 30518 | 700.20 | 47.46 | 540 | 860 |
| | 10 | 30546 | 795.98 | 53.28 | 645 | 980 | 28845 | 801.15 | 47.85 | 645 | 980 |

**Table 10-3. Scale Score Descriptive Statistics by Race/Ethnicity for English Language Arts**

| Race/Ethnicity | Grade | N Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| White | 3 | 38091 | 561.27 | 43.33 | 330 | 810 |
| | 4 | 38419 | 590.64 | 47.66 | 340 | 862 |
| | 5 | 38782 | 605.79 | 47.66 | 350 | 940 |
| | 6 | 39875 | 613.18 | 46.55 | 360 | 950 |
| | 7 | 40934 | 630.60 | 52.03 | 381 | 923 |
| | 8 | 41778 | 634.12 | 57.28 | 380 | 970 |
| African American | 3 | 5679 | 506.24 | 43.73 | 330 | 695 |
| | 4 | 5604 | 532.70 | 48.88 | 340 | 744 |
| | 5 | 5774 | 546.89 | 49.00 | 350 | 731 |
| | 6 | 5873 | 559.49 | 50.93 | 360 | 728 |
| | 7 | 6072 | 573.17 | 54.30 | 370 | 847 |
| | 8 | 6135 | 577.39 | 57.95 | 380 | 792 |
| Hispanic | 3 | 7998 | 530.23 | 43.78 | 330 | 725 |
| | 4 | 8305 | 555.67 | 49.31 | 340 | 747 |
| | 5 | 8320 | 574.39 | 49.05 | 350 | 786 |
| | 6 | 8431 | 582.97 | 49.12 | 360 | 763 |
| | 7 | 8702 | 597.03 | 53.54 | 378 | 862 |
| | 8 | 8846 | 601.74 | 59.19 | 380 | 843 |
| Asian | 3 | 2744 | 546.66 | 47.41 | 330 | 709 |
| | 4 | 2706 | 578.40 | 52.66 | 404 | 814 |
| | 5 | 2590 | 593.56 | 52.34 | 386 | 786 |
| | 6 | 2470 | 606.77 | 49.19 | 412 | 773 |
| | 7 | 2671 | 626.77 | 55.23 | 430 | 886 |
| | 8 | 2540 | 634.19 | 59.47 | 431 | 868 |
| American Indian | 3 | 599 | 526.24 | 42.62 | 330 | 666 |
| | 4 | 593 | 555.60 | 45.62 | 386 | 688 |
| | 5 | 634 | 570.71 | 50.43 | 407 | 877 |
| | 6 | 620 | 573.42 | 49.46 | 360 | 712 |
| | 7 | 623 | 591.85 | 51.82 | 381 | 746 |
| | 8 | 590 | 590.45 | 57.48 | 380 | 775 |
| Two or More | 3 | 3164 | 545.71 | 46.29 | 330 | 708 |
| | 4 | 3146 | 573.15 | 51.83 | 340 | 769 |
| | 5 | 2944 | 589.11 | 50.83 | 350 | 755 |
| | 6 | 2843 | 596.96 | 52.09 | 360 | 758 |
| | 7 | 2869 | 611.44 | 57.38 | 376 | 917 |
| | 8 | 2795 | 615.94 | 61.64 | 380 | 842 |

**Table 10-4. Scale Score Descriptive Statistics by Race/Ethnicity for Mathematics**

| Race/Ethnicity | Grade | N Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| White | 3 | 38088 | 566.59 | 48.97 | 360 | 760 |
| | 4 | 38425 | 589.12 | 48.71 | 405 | 800 |
| | 5 | 38776 | 612.92 | 45.37 | 430 | 830 |
| | 6 | 39861 | 619.71 | 52.33 | 440 | 870 |
| | 7 | 40900 | 632.88 | 54.99 | 450 | 880 |
| | 8 | 41754 | 648.12 | 54.28 | 470 | 890 |
| African American | 3 | 5675 | 498.29 | 55.49 | 360 | 760 |
| | 4 | 5606 | 515.97 | 56.83 | 405 | 702 |
| | 5 | 5776 | 546.54 | 54.01 | 430 | 719 |
| | 6 | 5874 | 547.32 | 57.25 | 440 | 870 |
| | 7 | 6067 | 559.22 | 59.84 | 450 | 743 |
| | 8 | 6139 | 583.17 | 57.50 | 470 | 890 |
| Hispanic | 3 | 8152 | 527.71 | 54.12 | 360 | 760 |
| | 4 | 8432 | 545.66 | 55.45 | 405 | 800 |
| | 5 | 8446 | 577.48 | 51.15 | 430 | 830 |
| | 6 | 8545 | 578.06 | 55.42 | 440 | 870 |
| | 7 | 8832 | 590.56 | 59.16 | 450 | 880 |
| | 8 | 8960 | 610.18 | 57.61 | 470 | 890 |
| Asian | 3 | 2770 | 552.93 | 60.82 | 360 | 760 |
| | 4 | 2725 | 575.58 | 57.28 | 405 | 800 |
| | 5 | 2612 | 600.99 | 54.13 | 430 | 830 |
| | 6 | 2489 | 610.95 | 62.59 | 440 | 870 |
| | 7 | 2687 | 625.89 | 63.68 | 450 | 880 |
| | 8 | 2539 | 647.84 | 64.35 | 470 | 890 |
| American Indian | 3 | 597 | 522.73 | 51.21 | 360 | 760 |
| | 4 | 591 | 544.57 | 52.88 | 405 | 705 |
| | 5 | 634 | 575.91 | 51.44 | 430 | 726 |
| | 6 | 618 | 564.75 | 57.06 | 440 | 720 |
| | 7 | 620 | 580.42 | 59.74 | 450 | 796 |
| | 8 | 584 | 603.38 | 59.21 | 470 | 890 |
| Two or More | 3 | 3167 | 544.51 | 53.60 | 360 | 760 |
| | 4 | 3152 | 564.32 | 57.52 | 405 | 800 |
| | 5 | 2943 | 591.10 | 52.95 | 430 | 830 |
| | 6 | 2847 | 594.81 | 59.59 | 440 | 870 |
| | 7 | 2863 | 605.72 | 63.90 | 450 | 880 |
| | 8 | 2786 | 626.13 | 60.69 | 470 | 890 |

**Table 10-5. Scale Score Descriptive Statistics by Race/Ethnicity for Science**

| Race/Ethnicity | Grade | N Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| White | 4 | 38422 | 509.83 | 47.54 | 300 | 725 |
| | 8 | 41734 | 705.20 | 48.22 | 480 | 945 |
| African American | 4 | 5577 | 446.89 | 42.78 | 300 | 638 |
| | 8 | 6068 | 650.39 | 42.32 | 480 | 885 |
| Hispanic | 4 | 8422 | 470.24 | 47.96 | 300 | 725 |
| | 8 | 8942 | 672.74 | 46.83 | 480 | 945 |
| Asian | 4 | 2721 | 488.52 | 51.17 | 300 | 709 |
| | 8 | 2540 | 698.75 | 50.31 | 480 | 945 |
| American Indian | 4 | 590 | 472.84 | 44.03 | 342 | 617 |
| | 8 | 580 | 670.61 | 46.02 | 521 | 825 |
| Two or More | 4 | 3148 | 490.43 | 51.17 | 324 | 673 |
| | 8 | 2780 | 686.75 | 50.75 | 480 | 945 |

**Table 10-6. Scale Score Descriptive Statistics by Race/Ethnicity for Social Studies**

| Race/Ethnicity | Grade | N Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| White | 4 | 38403 | 512.14 | 45.77 | 330 | 700 |
| | 8 | 41752 | 709.43 | 46.60 | 540 | 860 |
| | 10 | 41976 | 806.78 | 47.73 | 645 | 980 |
| African American | 4 | 5566 | 455.01 | 47.18 | 330 | 685 |
| | 8 | 6027 | 658.90 | 51.52 | 540 | 860 |
| | 10 | 4203 | 757.38 | 51.39 | 645 | 980 |
| Hispanic | 4 | 8407 | 477.84 | 47.06 | 330 | 700 |
| | 8 | 8930 | 680.42 | 49.73 | 540 | 860 |
| | 10 | 7932 | 778.07 | 49.19 | 645 | 963 |
| Asian | 4 | 2721 | 495.66 | 50.07 | 330 | 700 |
| | 8 | 2539 | 704.91 | 50.30 | 540 | 860 |
| | 10 | 2458 | 803.62 | 48.74 | 645 | 980 |
| American Indian | 4 | 593 | 479.96 | 43.25 | 330 | 621 |
| | 8 | 577 | 676.05 | 47.19 | 540 | 860 |
| | 10 | 570 | 771.96 | 48.70 | 645 | 920 |
| Two or More | 4 | 3143 | 495.34 | 50.04 | 330 | 700 |
| | 8 | 2781 | 693.91 | 51.46 | 540 | 860 |
| | 10 | 2252 | 793.84 | 52.66 | 645 | 980 |

**Table 10-7. Scale Score Descriptive Statistics by Economic Status**

| Content | Grade | Economically Disadvantaged | | | | | Not Economically Disadvantaged | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N Count | Mean | SD | Min | Max | N Count | Mean | SD | Min | Max |
| English Language Arts | 3 | 24454 | 530.16 | 44.65 | 330 | 725 | 33821 | 563.92 | 44.20 | 330 | 810 |
| | 4 | 24642 | 556.82 | 49.73 | 340 | 761 | 34131 | 593.84 | 48.30 | 340 | 862 |
| | 5 | 24301 | 571.79 | 49.25 | 350 | 853 | 34743 | 609.29 | 48.29 | 350 | 940 |
| | 6 | 24480 | 580.95 | 49.84 | 360 | 780 | 35632 | 616.90 | 46.45 | 360 | 950 |
| | 7 | 24829 | 595.79 | 53.98 | 370 | 917 | 37042 | 634.22 | 52.44 | 370 | 923 |
| | 8 | 24851 | 598.99 | 58.65 | 380 | 929 | 37833 | 638.40 | 57.35 | 380 | 970 |
| Mathematics | 3 | 24510 | 528.53 | 55.12 | 360 | 760 | 33939 | 569.38 | 50.15 | 360 | 760 |
| | 4 | 24705 | 548.47 | 56.97 | 405 | 800 | 34226 | 591.64 | 49.70 | 405 | 800 |
| | 5 | 24345 | 576.45 | 52.70 | 430 | 830 | 34842 | 615.40 | 46.34 | 430 | 830 |
| | 6 | 24532 | 578.28 | 57.84 | 440 | 870 | 35702 | 622.75 | 53.31 | 440 | 870 |
| | 7 | 24857 | 590.53 | 60.46 | 450 | 880 | 37112 | 635.66 | 56.40 | 450 | 880 |
| | 8 | 24881 | 609.83 | 58.49 | 470 | 890 | 37881 | 651.44 | 55.14 | 470 | 890 |
| Science | 4 | 24665 | 474.35 | 49.03 | 300 | 725 | 34215 | 511.28 | 48.67 | 300 | 725 |
| | 8 | 24799 | 673.76 | 47.66 | 480 | 945 | 37845 | 707.03 | 48.97 | 480 | 945 |
| Social Studies | 4 | 24632 | 479.16 | 47.87 | 330 | 700 | 34201 | 514.75 | 46.54 | 330 | 700 |
| | 8 | 24755 | 679.69 | 50.31 | 540 | 860 | 37851 | 712.03 | 46.90 | 540 | 860 |
| | 10 | 20272 | 778.26 | 50.14 | 645 | 980 | 39119 | 808.98 | 47.86 | 645 | 980 |

**Table 10-8. Scale Score Descriptive Statistics by Disability**

| Content | Grade | Disabled | | | | | Not Disabled | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N Count | Mean | SD | Min | Max | N Count | Mean | SD | Min | Max |
| English Language Arts | 3 | 7903 | 520.90 | 45.08 | 330 | 693 | 50372 | 554.28 | 46.16 | 330 | 810 |
| | 4 | 7872 | 542.39 | 50.95 | 340 | 766 | 50901 | 583.88 | 50.15 | 340 | 862 |
| | 5 | 7750 | 551.26 | 50.68 | 350 | 785 | 51294 | 600.30 | 49.17 | 350 | 940 |
| | 6 | 7527 | 555.54 | 49.30 | 360 | 745 | 52585 | 608.95 | 47.65 | 360 | 950 |
| | 7 | 7591 | 567.02 | 50.61 | 370 | 764 | 54280 | 626.04 | 53.18 | 370 | 923 |
| | 8 | 7153 | 566.24 | 56.11 | 380 | 829 | 55531 | 630.06 | 57.70 | 380 | 970 |
| Mathematics | 3 | 7896 | 515.96 | 62.78 | 360 | 760 | 50553 | 557.91 | 52.71 | 360 | 760 |
| | 4 | 7864 | 534.69 | 62.30 | 405 | 800 | 51067 | 579.52 | 53.69 | 405 | 800 |
| | 5 | 7746 | 558.71 | 57.81 | 430 | 754 | 51441 | 605.50 | 49.01 | 430 | 830 |
| | 6 | 7524 | 553.78 | 59.68 | 440 | 870 | 52710 | 611.90 | 55.65 | 440 | 870 |
| | 7 | 7580 | 563.71 | 61.53 | 450 | 880 | 54389 | 625.06 | 58.40 | 450 | 880 |
| | 8 | 7146 | 582.25 | 59.04 | 470 | 890 | 55616 | 641.72 | 56.72 | 470 | 890 |
| Science | 4 | 7864 | 466.10 | 50.01 | 300 | 725 | 51016 | 500.39 | 50.90 | 300 | 725 |
| | 8 | 7120 | 656.40 | 47.69 | 480 | 945 | 55524 | 698.66 | 49.53 | 480 | 945 |
| Social Studies | 4 | 7850 | 468.13 | 51.49 | 330 | 700 | 50983 | 504.73 | 48.26 | 330 | 700 |
| | 8 | 7113 | 655.89 | 54.44 | 540 | 860 | 55493 | 704.80 | 47.54 | 540 | 860 |
| | 10 | 6163 | 753.66 | 51.16 | 645 | 958 | 53228 | 803.68 | 48.11 | 645 | 980 |

**Table 10-9. Scale Score Descriptive Statistics by English Language Proficiency**

| Content | Grade | Limited English Proficient | | | | | Fully English Proficient | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N Count | Mean | SD | Min | Max | N Count | Mean | SD | Min | Max |
| English Language Arts | 3 | 5086 | 523.15 | 40.29 | 330 | 683 | 53189 | 552.29 | 47.26 | 330 | 810 |
| | 4 | 5383 | 547.19 | 44.39 | 340 | 709 | 53390 | 581.46 | 51.90 | 340 | 862 |
| | 5 | 4787 | 559.44 | 44.10 | 350 | 877 | 54257 | 596.90 | 51.63 | 350 | 940 |
| | 6 | 4023 | 563.70 | 44.84 | 373 | 704 | 56089 | 605.02 | 50.31 | 360 | 950 |
| | 7 | 3889 | 573.63 | 45.90 | 378 | 737 | 57982 | 621.83 | 55.64 | 370 | 923 |
| | 8 | 3887 | 577.38 | 52.01 | 380 | 803 | 58797 | 625.78 | 60.35 | 380 | 970 |
| Mathematics | 3 | 5251 | 522.43 | 52.70 | 360 | 760 | 53198 | 555.19 | 55.50 | 360 | 760 |
| | 4 | 5535 | 539.89 | 54.23 | 405 | 800 | 53396 | 577.03 | 56.13 | 405 | 800 |
| | 5 | 4930 | 566.58 | 48.39 | 430 | 830 | 54257 | 602.36 | 52.03 | 430 | 830 |
| | 6 | 4153 | 559.66 | 52.96 | 440 | 778 | 56081 | 607.97 | 58.45 | 440 | 870 |
| | 7 | 4027 | 570.05 | 54.69 | 450 | 767 | 57942 | 620.86 | 61.26 | 450 | 880 |
| | 8 | 3985 | 591.57 | 52.72 | 470 | 890 | 58777 | 637.89 | 59.37 | 470 | 890 |
| Science | 4 | 5521 | 460.77 | 42.65 | 300 | 631 | 53359 | 499.44 | 51.65 | 300 | 725 |
| | 8 | 3976 | 654.40 | 38.92 | 480 | 842 | 58668 | 696.53 | 50.74 | 480 | 945 |
| Social Studies | 4 | 5516 | 469.10 | 43.17 | 330 | 644 | 53317 | 503.03 | 49.88 | 330 | 700 |
| | 8 | 3977 | 661.17 | 45.95 | 540 | 841 | 58629 | 701.83 | 50.08 | 540 | 860 |
| | 10 | 2824 | 751.16 | 45.70 | 645 | 980 | 56567 | 800.86 | 49.86 | 645 | 980 |

**Table 10-10. Scale Score Descriptive Statistics by Accommodation Use**

| Content | Grade | Students Using Testing Accommodations | | | | | Students Not Using Accommodations | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N Count | Mean | SD | Min | Max | N Count | Mean | SD | Min | Max |
| English Language Arts | 3 | 84 | 513.36 | 47.81 | 406 | 611 | 58191 | 549.80 | 47.39 | 330 | 810 |
| | 4 | 86 | 550.12 | 61.03 | 402 | 698 | 58687 | 578.36 | 52.18 | 340 | 862 |
| | 5 | 82 | 558.94 | 44.86 | 433 | 655 | 58962 | 593.91 | 52.06 | 350 | 940 |
| | 6 | 100 | 576.59 | 53.43 | 434 | 700 | 60012 | 602.30 | 51.00 | 360 | 950 |
| | 7 | 93 | 582.66 | 56.28 | 466 | 708 | 61778 | 618.85 | 56.29 | 370 | 923 |
| | 8 | 113 | 601.10 | 63.62 | 438 | 829 | 62571 | 622.82 | 60.98 | 380 | 970 |
| Mathematics | 3 | 217 | 475.29 | 57.89 | 360 | 628 | 58232 | 552.53 | 55.84 | 360 | 760 |
| | 4 | 1865 | 507.69 | 53.66 | 405 | 665 | 57066 | 575.69 | 55.81 | 405 | 800 |
| | 5 | 2281 | 535.59 | 51.24 | 430 | 687 | 56906 | 601.93 | 51.09 | 430 | 830 |
| | 6 | 2718 | 532.06 | 49.72 | 440 | 713 | 57516 | 608.07 | 57.56 | 440 | 870 |
| | 7 | 2848 | 544.40 | 52.73 | 450 | 721 | 59121 | 621.08 | 60.35 | 450 | 880 |
| | 8 | 2646 | 565.79 | 52.04 | 470 | 718 | 60116 | 637.99 | 58.52 | 470 | 890 |
| Science | 4 | 28 | 444.21 | 38.78 | 375 | 533 | 58852 | 495.84 | 52.10 | 300 | 725 |
| | 8 | 30 | 668.63 | 53.89 | 536 | 775 | 62614 | 693.87 | 51.11 | 480 | 945 |
| Social Studies | 4 | 28 | 447.29 | 48.88 | 330 | 547 | 58805 | 499.87 | 50.26 | 330 | 700 |
| | 8 | 29 | 677.90 | 65.41 | 540 | 791 | 62577 | 699.25 | 50.79 | 540 | 860 |
| | 10 | 21 | 760.52 | 61.27 | 645 | 857 | 59370 | 798.51 | 50.77 | 645 | 980 |

**Table 10-11. Score Ranges and Associated Impact Data, English Language Arts**

| Grade | Score Range | | | | Impact Data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Proficient + Advanced* |
| 3 | 330–521 | 522–569 | 570–623 | 624–900 | 27.55 | 37.52 | 29.61 | 5.32 | 34.93 |
| 4 | 340–545 | 546–591 | 592–649 | 650–930 | 26.66 | 31.74 | 33.63 | 7.96 | 41.60 |
| 5 | 350–563 | 564–609 | 610–669 | 670–940 | 27.99 | 31.73 | 34.21 | 6.06 | 40.28 |
| 6 | 360–571 | 572–621 | 622–670 | 671–950 | 26.19 | 35.83 | 30.71 | 7.27 | 37.98 |
| 7 | 370–584 | 585–637 | 638–696 | 697–960 | 26.71 | 35.03 | 30.70 | 7.55 | 38.26 |
| 8 | 380–591 | 592–651 | 652–707 | 708–970 | 29.81 | 36.72 | 26.45 | 7.02 | 33.46 |

**Table 10-12. Score Ranges and Associated Impact Data, Mathematics**

| Grade | Score Range | | | | Impact Data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Proficient + Advanced* |
| 3 | 360–516 | 517–559 | 560–610 | 611–760 | 22.76 | 30.04 | 35.02 | 12.18 | 47.20 |
| 4 | 405–535 | 536–587 | 588–632 | 633–800 | 22.24 | 34.03 | 30.95 | 12.79 | 43.73 |
| 5 | 430–573 | 574–610 | 611–657 | 658–830 | 26.82 | 28.38 | 34.28 | 10.52 | 44.80 |
| 6 | 440–581 | 582–625 | 626–687 | 688–870 | 30.96 | 30.28 | 33.27 | 5.49 | 38.76 |
| 7 | 450–605 | 606–646 | 647–711 | 712–880 | 38.11 | 28.16 | 29.58 | 4.15 | 33.73 |
| 8 | 470–619 | 620–666 | 667–717 | 718–890 | 35.96 | 33.73 | 24.05 | 6.26 | 30.31 |

**Table 10-13. Score Ranges and Associated Impact Data, Science**

| Grade | Score Range | | | | Impact Data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Proficient + Advanced* |
| 4 | 300–446 | 447–495 | 496–542 | 543–725 | 18.20 | 30.95 | 32.32 | 18.53 | 50.85 |
| 8 | 480–652 | 653–694 | 695–736 | 737–945 | 20.77 | 30.33 | 28.99 | 19.91 | 48.90 |

**Table 10-14. Score Ranges and Associated Impact Data, Social Studies**

| Grade | Score Range | | | | Impact Data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Proficient + Advanced* |
| 4 | 330–460 | 461-490 | 491–536 | 537–700 | 21.02 | 20.09 | 36.79 | 22.10 | 58.89 |
| 8 | 540–661 | 662–692 | 693–733 | 734–860 | 19.53 | 21.74 | 35.08 | 23.64 | 58.73 |
| 10 | 645–769 | 770–804 | 805–836 | 837–980 | 25.49 | 26.84 | 25.73 | 21.94 | 47.67 |

**Table 10-15A. Percentage of Students in Each Performance Level by Subgroup, English Language Arts Grades 3–5**

| Demographic Group | Grade 3 | | | | | Grade 4 | | | | | Grade 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | | BB | B | P | A | |
| **All Students** | 16056 | 21865 | 17256 | 3098 | 58275 | 15669 | 18655 | 19768 | 4681 | 58773 | 16528 | 18736 | 20199 | 3581 | 59044 |
| **Percentage All** | 27.6 | 37.5 | 29.6 | 5.3 | 100 | 26.7 | 31.7 | 33.6 | 8.0 | 100 | 28.0 | 31.7 | 34.2 | 6.1 | 100 |
| **Gender** | | | | | | | | | | | | | | | |
| Female | 24.5 | 37.3 | 32.1 | 6.1 | 28662 | 24.0 | 31.1 | 35.5 | 9.4 | 28689 | 24.3 | 32.0 | 36.4 | 7.4 | 28951 |
| Male | 30.5 | 37.8 | 27.2 | 4.5 | 29613 | 29.2 | 32.3 | 31.9 | 6.6 | 30084 | 31.6 | 31.5 | 32.1 | 4.8 | 30093 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | |
| White | 18.0 | 38.5 | 36.6 | 6.9 | 38091 | 17.3 | 32.1 | 40.5 | 10.2 | 38419 | 19.0 | 32.0 | 41.2 | 7.8 | 38782 |
| African American | 66.2 | 25.7 | 7.5 | 0.6 | 5679 | 62.7 | 25.5 | 10.8 | 1.0 | 5604 | 64.9 | 25.2 | 9.2 | 0.6 | 5774 |
| Hispanic | 42.1 | 39.9 | 16.3 | 1.6 | 7998 | 42.8 | 34.1 | 20.3 | 2.9 | 8305 | 41.4 | 34.3 | 22.2 | 2.1 | 8320 |
| Asian | 29.8 | 39.8 | 24.8 | 5.6 | 2744 | 27.5 | 32.6 | 30.7 | 9.2 | 2706 | 29.8 | 31.6 | 31.5 | 7.1 | 2590 |
| American Indian | 45.2 | 40.1 | 13.4 | 1.3 | 599 | 44.2 | 34.1 | 19.6 | 2.2 | 593 | 43.5 | 36.4 | 18.6 | 1.4 | 634 |
| Two or More | 30.7 | 38.8 | 25.8 | 4.7 | 3164 | 30.6 | 31.7 | 31.0 | 6.6 | 3146 | 31.7 | 32.5 | 30.3 | 5.5 | 2944 |
| **English Language Proficiency** | | | | | | | | | | | | | | | |
| Fully Proficient | 25.6 | 37.3 | 31.4 | 5.8 | 53189 | 24.4 | 31.4 | 35.5 | 8.7 | 53390 | 25.7 | 31.6 | 36.2 | 6.6 | 54257 |
| Limited Proficiency | 48.5 | 39.5 | 11.3 | 0.7 | 5086 | 49.4 | 34.8 | 15.0 | 0.9 | 5383 | 54.3 | 33.4 | 11.9 | 0.3 | 4787 |
| **Disability Status** | | | | | | | | | | | | | | | |
| Yes | 52.4 | 33.7 | 12.3 | 1.5 | 7903 | 55.0 | 27.8 | 15.2 | 2.0 | 7872 | 62.1 | 25.1 | 11.7 | 1.1 | 7750 |
| No | 23.6 | 38.1 | 32.3 | 5.9 | 50372 | 22.3 | 32.3 | 36.5 | 8.9 | 50901 | 22.8 | 32.7 | 37.6 | 6.8 | 51294 |
| **Economically Disadvantaged** | | | | | | | | | | | | | | | |
| Yes | 42.8 | 38.0 | 17.5 | 1.7 | 24454 | 41.5 | 33.9 | 21.8 | 2.9 | 24642 | 43.6 | 33.5 | 21.1 | 1.8 | 24301 |
| No | 16.5 | 37.2 | 38.4 | 7.9 | 33821 | 15.9 | 30.2 | 42.2 | 11.6 | 34131 | 17.1 | 30.5 | 43.3 | 9.1 | 34743 |
| **Accommodation Use** | | | | | | | | | | | | | | | |
| Yes | 56.0 | 32.1 | 11.9 | 0.0 | 84 | 50.0 | 24.4 | 19.8 | 5.8 | 86 | 57.3 | 29.3 | 13.4 | 0.0 | 82 |
| No | 27.5 | 37.5 | 29.6 | 5.3 | 58191 | 26.6 | 31.8 | 33.7 | 8.0 | 58687 | 28.0 | 31.7 | 34.2 | 6.1 | 58962 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-15B. Percentage of Students in Each Performance Level by Subgroup, English Language Arts Grades 6–8**

| Demographic Group | Grade 6 | | | | | Grade 7 | | | | | Grade 8 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | | BB | B | P | A | |
| **All Students** | 15744 | 21539 | 18458 | 4371 | 60112 | 16526 | 21676 | 18995 | 4674 | 61871 | 18688 | 23020 | 16578 | 4398 | 62684 |
| **Percentage All** | 26.2 | 35.8 | 30.7 | 7.3 | 100 | 26.7 | 35.0 | 30.7 | 7.6 | 100 | 29.8 | 36.7 | 26.4 | 7.0 | 100 |
| **Gender** | | | | | | | | | | | | | | | |
| Female | 22.3 | 35.2 | 33.4 | 9.1 | 29463 | 22.9 | 35.8 | 32.2 | 9.1 | 30189 | 25.1 | 36.9 | 29.2 | 8.8 | 30551 |
| Male | 29.9 | 36.5 | 28.1 | 5.6 | 30649 | 30.3 | 34.3 | 29.2 | 6.1 | 31682 | 34.3 | 36.5 | 23.8 | 5.3 | 32133 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | |
| White | 18.1 | 36.3 | 36.3 | 9.3 | 39875 | 18.5 | 35.6 | 36.5 | 9.4 | 40934 | 22.3 | 37.9 | 31.1 | 8.7 | 41778 |
| African American | 59.4 | 29.5 | 10.2 | 1.0 | 5873 | 59.3 | 28.4 | 10.9 | 1.4 | 6072 | 61.0 | 28.2 | 9.7 | 1.0 | 6135 |
| Hispanic | 39.6 | 37.8 | 20.3 | 2.4 | 8431 | 41.1 | 36.3 | 19.5 | 3.0 | 8702 | 43.2 | 36.5 | 17.2 | 3.1 | 8846 |
| Asian | 23.0 | 37.4 | 30.9 | 8.6 | 2470 | 22.5 | 36.5 | 30.9 | 10.1 | 2671 | 23.2 | 38.2 | 27.8 | 10.9 | 2540 |
| American Indian | 45.0 | 39.0 | 14.4 | 1.6 | 620 | 43.7 | 38.4 | 15.2 | 2.7 | 623 | 50.3 | 36.1 | 12.4 | 1.2 | 590 |
| Two or More | 30.6 | 34.6 | 28.4 | 6.5 | 2843 | 32.0 | 34.7 | 27.2 | 6.2 | 2869 | 33.1 | 37.4 | 23.8 | 5.7 | 2795 |
| **English Language Proficiency** | | | | | | | | | | | | | | | |
| Fully Proficient | 24.1 | 35.9 | 32.3 | 7.8 | 56089 | 24.6 | 35.1 | 32.3 | 8.0 | 57982 | 27.7 | 37.0 | 27.8 | 7.4 | 58797 |
| Limited Proficiency | 56.0 | 34.6 | 8.9 | 0.6 | 4023 | 58.8 | 33.6 | 7.0 | 0.5 | 3889 | 61.0 | 32.0 | 6.2 | 0.7 | 3887 |
| **Disability Status** | | | | | | | | | | | | | | | |
| Yes | 64.4 | 26.2 | 8.2 | 1.1 | 7527 | 65.4 | 25.9 | 7.8 | 0.9 | 7591 | 69.3 | 23.5 | 6.3 | 0.8 | 7153 |
| No | 20.7 | 37.2 | 33.9 | 8.2 | 52585 | 21.3 | 36.3 | 33.9 | 8.5 | 54280 | 24.7 | 38.4 | 29.0 | 7.8 | 55531 |
| **Economically Disadvantaged** | | | | | | | | | | | | | | | |
| Yes | 41.3 | 37.1 | 19.2 | 2.5 | 24480 | 41.7 | 36.1 | 19.5 | 2.8 | 24829 | 44.6 | 36.3 | 16.7 | 2.4 | 24851 |
| No | 15.8 | 35.0 | 38.6 | 10.6 | 35632 | 16.7 | 34.3 | 38.2 | 10.7 | 37042 | 20.1 | 37.0 | 32.8 | 10.0 | 37833 |
| **Accommodation Use** | | | | | | | | | | | | | | | |
| Yes | 45.0 | 37.0 | 15.0 | 3.0 | 100 | 49.5 | 33.3 | 15.1 | 2.2 | 93 | 47.8 | 31.0 | 16.8 | 4.4 | 113 |
| No | 26.2 | 35.8 | 30.7 | 7.3 | 60012 | 26.7 | 35.0 | 30.7 | 7.6 | 61778 | 29.8 | 36.7 | 26.5 | 7.0 | 62571 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-16A. Percentage of Students in Each Performance Level by Subgroup, Mathematics Grades 3–5**

| Demographic Group | Grade 3 | | | | | Grade 4 | | | | | Grade 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | | BB | B | P | A | |
| All Students | 13302 | 17560 | 20468 | 7119 | 58449 | 13106 | 20052 | 18237 | 7536 | 58931 | 15873 | 16799 | 20291 | 6224 | 59187 |
| Percentage All | 22.8 | 30.0 | 35.0 | 12.2 | 100 | 22.2 | 34.0 | 30.9 | 12.8 | 100 | 26.8 | 28.4 | 34.3 | 10.5 | 100 |
| **Gender** | | | | | | | | | | | | | | | |
| Female | 24.1 | 31.1 | 34.5 | 10.3 | 28743 | 22.9 | 36.4 | 29.8 | 11.0 | 28770 | 27.5 | 30.4 | 33.5 | 8.6 | 29016 |
| Male | 21.5 | 29.0 | 35.5 | 14.0 | 29706 | 21.6 | 31.8 | 32.1 | 14.5 | 30161 | 26.2 | 26.4 | 35.1 | 12.3 | 30171 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | |
| White | 13.0 | 29.1 | 42.1 | 15.9 | 38088 | 12.1 | 33.2 | 38.0 | 16.7 | 38425 | 16.4 | 28.1 | 41.8 | 13.7 | 38776 |
| African American | 61.5 | 27.2 | 10.1 | 1.2 | 5675 | 61.6 | 30.0 | 7.3 | 1.1 | 5606 | 68.1 | 22.9 | 8.3 | 0.7 | 5776 |
| Hispanic | 38.2 | 34.0 | 23.7 | 4.1 | 8152 | 39.7 | 37.9 | 18.5 | 4.0 | 8432 | 43.0 | 32.3 | 21.2 | 3.5 | 8446 |
| Asian | 24.4 | 30.6 | 31.1 | 13.9 | 2770 | 22.3 | 36.0 | 27.1 | 14.6 | 2725 | 26.9 | 28.6 | 31.9 | 12.6 | 2612 |
| American Indian | 40.0 | 38.2 | 19.3 | 2.5 | 597 | 38.6 | 42.6 | 16.6 | 2.2 | 591 | 44.2 | 32.8 | 18.9 | 4.1 | 634 |
| Two or More | 26.8 | 33.8 | 30.6 | 8.8 | 3167 | 26.3 | 37.9 | 25.9 | 9.9 | 3152 | 32.8 | 30.5 | 29.1 | 7.6 | 2943 |
| **English Language Proficiency** | | | | | | | | | | | | | | | |
| Fully Proficient | 20.9 | 29.6 | 36.5 | 13.1 | 53198 | 20.1 | 33.5 | 32.6 | 13.9 | 53396 | 24.5 | 28.0 | 36.1 | 11.4 | 54257 |
| Limited Proficiency | 41.8 | 35.0 | 20.2 | 3.0 | 5251 | 43.1 | 39.2 | 15.4 | 2.2 | 5535 | 51.8 | 33.1 | 14.1 | 1.1 | 4930 |
| **Disability Status** | | | | | | | | | | | | | | | |
| Yes | 47.2 | 29.0 | 18.6 | 5.1 | 7896 | 48.5 | 31.6 | 15.6 | 4.3 | 7864 | 59.1 | 23.7 | 14.1 | 3.1 | 7746 |
| No | 18.9 | 30.2 | 37.6 | 13.3 | 50553 | 18.2 | 34.4 | 33.3 | 14.1 | 51067 | 22.0 | 29.1 | 37.3 | 11.6 | 51441 |
| **Economically Disadvantaged** | | | | | | | | | | | | | | | |
| Yes | 37.4 | 33.8 | 24.1 | 4.7 | 24510 | 37.2 | 38.0 | 19.9 | 4.9 | 24705 | 43.3 | 31.1 | 22.1 | 3.5 | 24345 |
| No | 12.2 | 27.3 | 42.9 | 17.6 | 33939 | 11.5 | 31.1 | 38.9 | 18.5 | 34226 | 15.3 | 26.5 | 42.8 | 15.4 | 34842 |
| **Accommodation Use** | | | | | | | | | | | | | | | |
| Yes | 77.0 | 17.1 | 5.5 | 0.5 | 217 | 68.5 | 26.6 | 4.4 | 0.4 | 1865 | 78.7 | 16.8 | 4.1 | 0.4 | 2281 |
| No | 22.6 | 30.1 | 35.1 | 12.2 | 58232 | 20.7 | 34.3 | 31.8 | 13.2 | 57066 | 24.7 | 28.8 | 35.5 | 10.9 | 56906 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-16B. Percentage of Students in Each Performance Level by Subgroup, Mathematics Grades 6–8**

| Demographic Group | Grade 6 | | | | | Grade 7 | | | | | Grade 8 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | | BB | B | P | A | |
| **All Students** | 18647 | 18241 | 20040 | 3306 | 60234 | 23615 | 17449 | 18332 | 2573 | 61969 | 22570 | 21168 | 15095 | 3929 | 62762 |
| **Percentage All** | 31.0 | 30.3 | 33.3 | 5.5 | 100 | 38.1 | 28.2 | 29.6 | 4.2 | 100 | 36.0 | 33.7 | 24.1 | 6.3 | 100 |
| **Gender** | | | | | | | | | | | | | | | |
| Female | 30.9 | 31.6 | 32.7 | 4.8 | 29502 | 39.0 | 29.8 | 27.7 | 3.6 | 30229 | 33.9 | 36.1 | 24.4 | 5.6 | 30590 |
| Male | 31.0 | 29.1 | 33.8 | 6.1 | 30732 | 37.3 | 26.6 | 31.4 | 4.7 | 31740 | 37.9 | 31.5 | 23.7 | 6.9 | 32172 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | |
| White | 20.1 | 31.5 | 41.3 | 7.1 | 39861 | 26.8 | 30.8 | 37.2 | 5.2 | 40900 | 25.9 | 36.4 | 29.9 | 7.7 | 41754 |
| African American | 72.1 | 20.4 | 7.2 | 0.2 | 5874 | 78.6 | 15.4 | 5.8 | 0.2 | 6067 | 74.2 | 20.2 | 5.1 | 0.4 | 6139 |
| Hispanic | 49.4 | 31.1 | 18.2 | 1.3 | 8545 | 58.3 | 26.0 | 14.6 | 1.1 | 8832 | 54.6 | 31.1 | 12.1 | 2.2 | 8960 |
| Asian | 29.8 | 30.7 | 29.7 | 9.8 | 2489 | 36.2 | 26.5 | 29.1 | 8.2 | 2687 | 30.8 | 31.9 | 24.3 | 13.0 | 2539 |
| American Indian | 57.6 | 27.8 | 14.2 | 0.3 | 618 | 64.2 | 25.6 | 9.7 | 0.5 | 620 | 57.0 | 32.7 | 8.9 | 1.4 | 584 |
| Two or More | 37.7 | 31.2 | 26.7 | 4.5 | 2847 | 47.2 | 26.6 | 22.8 | 3.5 | 2863 | 42.3 | 33.3 | 19.8 | 4.7 | 2786 |
| **English Language Proficiency** | | | | | | | | | | | | | | | |
| Fully Proficient | 28.5 | 30.6 | 35.1 | 5.9 | 56081 | 35.5 | 28.8 | 31.2 | 4.4 | 57942 | 33.7 | 34.3 | 25.4 | 6.7 | 58777 |
| Limited Proficiency | 64.6 | 26.1 | 9.0 | 0.2 | 4153 | 75.3 | 18.7 | 5.7 | 0.3 | 4027 | 69.9 | 25.3 | 4.4 | 0.4 | 3985 |
| **Disability Status** | | | | | | | | | | | | | | | |
| Yes | 68.2 | 20.8 | 9.9 | 1.1 | 7524 | 76.1 | 16.0 | 7.2 | 0.7 | 7580 | 74.8 | 18.8 | 5.3 | 1.0 | 7146 |
| No | 25.6 | 31.6 | 36.6 | 6.1 | 52710 | 32.8 | 29.9 | 32.7 | 4.6 | 54389 | 31.0 | 35.6 | 26.5 | 6.9 | 55616 |
| **Economically Disadvantaged** | | | | | | | | | | | | | | | |
| Yes | 48.9 | 30.5 | 19.1 | 1.6 | 24532 | 57.5 | 25.7 | 15.7 | 1.1 | 24857 | 54.0 | 31.3 | 12.8 | 2.0 | 24881 |
| No | 18.6 | 30.2 | 43.0 | 8.2 | 35702 | 25.1 | 29.8 | 38.9 | 6.2 | 37112 | 24.1 | 35.4 | 31.5 | 9.1 | 37881 |
| **Accommodation Use** | | | | | | | | | | | | | | | |
| Yes | 85.4 | 12.6 | 2.0 | 0.0 | 2718 | 90.0 | 8.5 | 1.5 | 0.0 | 2848 | 86.7 | 12.3 | 1.0 | 0.0 | 2646 |
| No | 28.4 | 31.1 | 34.7 | 5.7 | 57516 | 35.6 | 29.1 | 30.9 | 4.4 | 59121 | 33.7 | 34.7 | 25.1 | 6.5 | 60116 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-17. Percentage of Students in Each Performance Level by Subgroup, Science**

| Demographic Group | Grade 4 | | | | | Grade 8 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | |
| **All Students** | 10716 | 18224 | 19029 | 10911 | 58880 | 13009 | 18999 | 18161 | 12475 | 62644 |
| **Percentage All** | 18.2 | 31.0 | 32.3 | 18.5 | 100 | 20.8 | 30.3 | 29.0 | 19.9 | 100 |
| **Gender** | | | | | | | | | | |
| Female | 18.1 | 31.7 | 32.8 | 17.5 | 28751 | 19.4 | 32.3 | 30.7 | 17.7 | 30540 |
| Male | 18.3 | 30.3 | 31.9 | 19.5 | 30129 | 22.1 | 28.5 | 27.4 | 22.1 | 32104 |
| **Race/Ethnicity** | | | | | | | | | | |
| White | 9.3 | 28.3 | 38.2 | 24.2 | 38422 | 13.0 | 28.5 | 33.3 | 25.2 | 41734 |
| African American | 53.3 | 33.2 | 11.4 | 2.0 | 5577 | 53.5 | 32.5 | 11.4 | 2.6 | 6068 |
| Hispanic | 32.9 | 38.3 | 21.9 | 6.9 | 8422 | 33.8 | 35.5 | 22.2 | 8.6 | 8942 |
| Asian | 22.1 | 34.6 | 28.7 | 14.6 | 2721 | 16.9 | 32.4 | 29.1 | 21.7 | 2540 |
| American Indian | 28.0 | 42.4 | 23.2 | 6.4 | 590 | 34.0 | 37.1 | 21.9 | 7.1 | 580 |
| Two or More | 20.3 | 34.0 | 30.1 | 15.6 | 3148 | 25.0 | 33.2 | 25.3 | 16.5 | 2780 |
| **English Language Proficiency** | | | | | | | | | | |
| Fully Proficient | 16.1 | 30.0 | 33.9 | 20.1 | 53359 | 19.0 | 29.7 | 30.2 | 21.1 | 58668 |
| Limited Proficiency | 38.7 | 40.5 | 17.5 | 3.4 | 5521 | 47.3 | 39.4 | 11.3 | 2.0 | 3976 |
| **Disability Status** | | | | | | | | | | |
| Yes | 38.6 | 34.3 | 19.6 | 7.4 | 7864 | 50.5 | 31.1 | 12.4 | 6.0 | 7120 |
| No | 15.0 | 30.4 | 34.3 | 20.2 | 51016 | 17.0 | 30.2 | 31.1 | 21.7 | 55524 |
| **Economically Disadvantaged** | | | | | | | | | | |
| Yes | 30.2 | 37.0 | 24.0 | 8.8 | 24665 | 33.4 | 34.9 | 22.1 | 9.6 | 24799 |
| No | 9.5 | 26.6 | 38.3 | 25.6 | 34215 | 12.5 | 27.3 | 33.5 | 26.7 | 37845 |
| **Accommodation Use** | | | | | | | | | | |
| Yes | 60.7 | 28.6 | 10.7 | 0.0 | 28 | 43.3 | 23.3 | 30.0 | 3.3 | 30 |
| No | 18.2 | 31.0 | 32.3 | 18.5 | 58852 | 20.8 | 30.3 | 29.0 | 19.9 | 62614 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-18. Percentage of Students in Each Performance Level by Subgroup, Social Studies**

| Demographic Group | Grade 4 | | | | | Grade 8 | | | | | Grade 10 | | | | |
| | Performance Level | | | | Total Count | Performance Level | | | | Total Count | Performance Level | | | | Total Count |
| | BB | B | P | A | | BB | B | P | A | | BB | B | P | A | |
| **All Students** | 12366 | 11818 | 21644 | 13005 | 58833 | 12230 | 13610 | 21964 | 14802 | 62606 | 15140 | 15939 | 15281 | 13031 | 59391 |
| **Percentage All** | 21.0 | 20.1 | 36.8 | 22.1 | 100 | 19.5 | 21.7 | 35.1 | 23.6 | 100 | 25.5 | 26.8 | 25.7 | 21.9 | 100 |
| **Gender** | | | | | | | | | | | | | | | |
| Female | 20.3 | 20.7 | 37.5 | 21.5 | 28729 | 17.2 | 22.8 | 37.7 | 22.3 | 30518 | 22.3 | 27.7 | 28.1 | 22.0 | 28845 |
| Male | 21.7 | 19.5 | 36.1 | 22.7 | 30104 | 21.8 | 20.7 | 32.6 | 24.9 | 32088 | 28.5 | 26.0 | 23.5 | 21.9 | 30546 |
| **Race/Ethnicity** | | | | | | | | | | | | | | | |
| White | 12.1 | 18.4 | 41.5 | 28.1 | 38403 | 12.8 | 19.6 | 38.4 | 29.1 | 41752 | 19.2 | 26.3 | 28.3 | 26.1 | 41976 |
| African American | 56.6 | 21.2 | 17.9 | 4.2 | 5566 | 47.8 | 27.3 | 20.0 | 4.9 | 6027 | 59.5 | 23.2 | 12.2 | 5.1 | 4203 |
| Hispanic | 35.2 | 26.0 | 29.1 | 9.7 | 8407 | 31.2 | 26.2 | 30.7 | 11.9 | 8930 | 39.3 | 30.2 | 20.6 | 9.8 | 7932 |
| Asian | 23.6 | 21.9 | 35.3 | 19.2 | 2721 | 15.7 | 23.2 | 34.6 | 26.5 | 2539 | 21.1 | 29.3 | 26.3 | 23.3 | 2458 |
| American Indian | 32.4 | 29.0 | 29.5 | 9.1 | 593 | 33.1 | 31.0 | 27.4 | 8.5 | 577 | 45.6 | 28.9 | 19.6 | 5.8 | 570 |
| Two or More | 24.7 | 20.0 | 36.1 | 19.1 | 3143 | 22.5 | 23.5 | 33.4 | 20.5 | 2781 | 30.1 | 28.0 | 21.4 | 20.5 | 2252 |
| **English Language Proficiency** | | | | | | | | | | | | | | | |
| Fully Proficient | 18.9 | 19.3 | 37.8 | 23.9 | 53317 | 17.7 | 21.1 | 36.1 | 25.0 | 58629 | 23.6 | 26.8 | 26.7 | 23.0 | 56567 |
| Limited Proficiency | 41.2 | 27.7 | 26.6 | 4.6 | 5516 | 45.9 | 31.3 | 19.7 | 3.1 | 3977 | 64.0 | 27.4 | 7.1 | 1.5 | 2824 |
| **Disability Status** | | | | | | | | | | | | | | | |
| Yes | 45.8 | 22.3 | 22.8 | 9.2 | 7850 | 52.6 | 24.6 | 16.1 | 6.7 | 7113 | 62.4 | 23.2 | 9.7 | 4.8 | 6163 |
| No | 17.2 | 19.8 | 38.9 | 24.1 | 50983 | 15.3 | 21.4 | 37.5 | 25.8 | 55493 | 21.2 | 27.3 | 27.6 | 23.9 | 53228 |
| **Economically Disadvantaged** | | | | | | | | | | | | | | | |
| Yes | 34.7 | 24.5 | 30.3 | 10.6 | 24632 | 31.3 | 26.9 | 29.9 | 11.9 | 24755 | 40.0 | 29.5 | 19.9 | 10.7 | 20272 |
| No | 11.2 | 16.9 | 41.5 | 30.4 | 34201 | 11.8 | 18.4 | 38.5 | 31.3 | 37851 | 18.0 | 25.5 | 28.8 | 27.8 | 39119 |
| **Accommodation Use** | | | | | | | | | | | | | | | |
| Yes | 64.3 | 25.0 | 3.6 | 7.1 | 28 | 31.0 | 20.7 | 24.1 | 24.1 | 29 | 57.1 | 14.3 | 19.0 | 9.5 | 21 |
| No | 21.0 | 20.1 | 36.8 | 22.1 | 58805 | 19.5 | 21.7 | 35.1 | 23.6 | 62577 | 25.5 | 26.8 | 25.7 | 21.9 | 59370 |

Note: The abbreviation "BB" is for the *Below Basic* performance level, "B" is for the *Basic* performance level, "P" is for the *Proficient* performance level, and "A" is for the *Advanced* performance level.

**Table 10-19. Summary Statistics for Content Standards Raw and SPI Scores, English Language Arts**

| Grade | Content Standard | Standard | No. of Items | | Total Score Points | Mean | Mean p-Value | SD | SPI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MC | CR | | | | | Mean | SD |
| **3** | A | Reading—Key Ideas and Details | 3 | 4 | 11 | 5.76 | 0.52 | 3.13 | 52.42 | 25.57 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 7 | 0 | 7 | 3.51 | 0.50 | 1.68 | 50.51 | 18.46 |
| | C | Reading—Vocabulary Use | 4 | 0 | 4 | 2.35 | 0.59 | 1.19 | 58.82 | 22.56 |
| | D | Writing/Language—Text Types and Purposes | 1 | 4 | 12 | 4.82 | 0.52 | 1.86 | 40.31 | 12.33 |
| | E | Writing/Language—Research | 2 | 2 | 6 | 3.51 | 0.63 | 1.52 | 58.22 | 20.44 |
| | F | Writing/Language—Language Conventions | 6 | 0 | 6 | 3.30 | 0.55 | 1.54 | 55.18 | 18.68 |
| | G | Listening | 3 | 2 | 7 | 3.94 | 0.55 | 1.74 | 56.02 | 19.29 |
| **4** | A | Reading—Key Ideas and Details | 3 | 4 | 11 | 5.55 | 0.53 | 2.65 | 50.63 | 20.80 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 5 | 1 | 7 | 3.41 | 0.53 | 1.87 | 49.08 | 22.04 |
| | C | Reading—Vocabulary Use | 6 | 0 | 6 | 4.02 | 0.67 | 1.65 | 66.54 | 23.86 |
| | D | Writing/Language—Text Types and Purposes | 1 | 4 | 12 | 4.99 | 0.51 | 1.87 | 41.92 | 12.14 |
| | E | Writing/Language—Research | 3 | 2 | 6 | 3.11 | 0.53 | 1.61 | 52.15 | 21.13 |
| | F | Writing/Language—Language Conventions | 4 | 1 | 6 | 3.66 | 0.60 | 1.41 | 60.90 | 17.71 |
| | G | Listening | 4 | 2 | 8 | 4.11 | 0.54 | 1.97 | 51.40 | 18.64 |
| **5** | A | Reading—Key Ideas and Details | 8 | 2 | 12 | 6.26 | 0.54 | 2.95 | 52.42 | 22.04 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 8 | 0 | 8 | 4.09 | 0.51 | 2.04 | 51.60 | 21.35 |
| | C | Reading—Vocabulary Use | 2 | 1 | 4 | 2.38 | 0.57 | 1.03 | 59.48 | 15.62 |
| | D | Writing/Language—Text Types and Purposes | 2 | 3 | 12 | 4.97 | 0.45 | 1.88 | 41.76 | 12.19 |
| | E | Writing/Language—Research | 1 | 3 | 6 | 3.91 | 0.66 | 1.66 | 64.79 | 22.78 |
| | F | Writing/Language—Language Conventions | 4 | 1 | 6 | 3.69 | 0.62 | 1.53 | 61.23 | 19.07 |
| | G | Listening | 4 | 2 | 8 | 5.25 | 0.68 | 1.97 | 65.19 | 20.05 |

**Table 10-19 (continued). Summary Statistics for Content Standards Raw and SPI Scores, English Language Arts**

| Grade | Content Standard | Standard | No. of Items | | Total Score Points | Mean | Mean *p*-Value | SD | SPI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MC | CR | | | | | Mean | SD |
| **6** | A | Reading—Key Ideas and Details | 3 | 4 | 11 | 6.56 | 0.61 | 2.68 | 59.56 | 22.06 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 5 | 2 | 9 | 5.42 | 0.60 | 2.43 | 60.04 | 23.79 |
| | C | Reading—Vocabulary Use | 4 | 0 | 4 | 2.58 | 0.65 | 1.18 | 64.18 | 22.87 |
| | D | Writing/Language—Text Types and Purposes | 3 | 2 | 11 | 4.79 | 0.54 | 2.01 | 43.97 | 14.65 |
| | E | Writing/Language—Research | 2 | 2 | 6 | 2.60 | 0.44 | 1.38 | 43.75 | 15.77 |
| | F | Writing/Language—Language Conventions | 3 | 2 | 7 | 3.72 | 0.49 | 1.56 | 53.47 | 16.55 |
| | G | Listening | 4 | 2 | 8 | 4.94 | 0.62 | 2.07 | 61.51 | 20.92 |
| **7** | A | Reading—Key Ideas and Details | 4 | 4 | 11 | 4.86 | 0.44 | 2.57 | 44.58 | 20.60 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 4 | 2 | 8 | 3.71 | 0.47 | 2.20 | 46.88 | 22.43 |
| | C | Reading—Vocabulary Use | 2 | 2 | 5 | 3.50 | 0.71 | 1.40 | 69.30 | 23.43 |
| | D | Writing/Language—Text Types and Purposes | 1 | 3 | 11 | 5.19 | 0.54 | 2.19 | 47.55 | 16.29 |
| | E | Writing/Language—Research | 3 | 2 | 7 | 3.68 | 0.54 | 1.76 | 52.57 | 19.39 |
| | F | Writing/Language—Language Conventions | 3 | 2 | 6 | 3.95 | 0.67 | 1.31 | 65.63 | 16.10 |
| | G | Listening | 2 | 3 | 8 | 5.25 | 0.66 | 2.18 | 64.84 | 23.34 |
| **8** | A | Reading—Key Ideas and Details | 7 | 2 | 11 | 5.44 | 0.52 | 2.45 | 49.69 | 18.97 |
| | B | Reading—Craft & Structure/ Integration of Knowledge & Ideas | 8 | 0 | 8 | 4.27 | 0.54 | 1.93 | 53.60 | 19.70 |
| | C | Reading—Vocabulary Use | 1 | 2 | 5 | 3.03 | 0.61 | 1.54 | 60.25 | 26.48 |
| | D | Writing/Language—Text Types and Purposes | 4 | 2 | 12 | 6.16 | 0.58 | 2.40 | 51.64 | 16.18 |
| | E | Writing/Language—Research | 4 | 1 | 6 | 3.14 | 0.53 | 1.77 | 52.39 | 23.25 |
| | F | Writing/Language—Language Conventions | 2 | 2 | 6 | 3.63 | 0.59 | 1.31 | 60.33 | 15.51 |
| | G | Listening | 2 | 3 | 8 | 4.55 | 0.57 | 2.06 | 56.91 | 21.01 |

**Table 10-20. Summary Statistics for Domain Raw and SPI Scores, English Language Arts**

| Grade | Standard | No. of Items | | Total Score Points | Mean | Mean p-Value | SD | SPI | |
| | | MC | CR | | | | | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| 3 | Listening | 3 | 2 | 7 | 3.92 | 0.55 | 1.75 | 56.02 | 19.29 |
| | Reading | 14 | 4 | 22 | 11.63 | 0.53 | 5.17 | 52.98 | 22.40 |
| | Writing | 9 | 5 | 24 | 11.64 | 0.57 | 3.98 | 48.54 | 15.39 |
| 4 | Listening | 4 | 2 | 8 | 4.09 | 0.54 | 1.97 | 51.41 | 18.63 |
| | Reading | 14 | 5 | 24 | 13.00 | 0.57 | 5.39 | 54.18 | 21.58 |
| | Writing | 8 | 6 | 24 | 11.77 | 0.55 | 3.98 | 49.21 | 15.33 |
| 5 | Listening | 4 | 2 | 8 | 5.25 | 0.68 | 1.97 | 65.19 | 20.05 |
| | Reading | 18 | 3 | 24 | 12.73 | 0.53 | 5.15 | 53.24 | 20.37 |
| | Writing | 7 | 6 | 24 | 12.57 | 0.58 | 4.08 | 52.46 | 15.87 |
| 6 | Listening | 4 | 2 | 8 | 4.93 | 0.62 | 2.08 | 61.51 | 20.93 |
| | Reading | 12 | 6 | 24 | 14.55 | 0.62 | 5.59 | 60.61 | 22.65 |
| | Writing | 8 | 5 | 24 | 11.11 | 0.49 | 3.92 | 46.58 | 14.95 |
| 7 | Listening | 2 | 3 | 8 | 5.25 | 0.66 | 2.18 | 64.84 | 23.34 |
| | Reading | 10 | 8 | 24 | 12.07 | 0.51 | 5.28 | 50.44 | 21.13 |
| | Writing | 7 | 6 | 24 | 12.82 | 0.59 | 4.23 | 53.55 | 16.34 |
| 8 | Listening | 2 | 3 | 8 | 4.55 | 0.57 | 2.06 | 56.91 | 21.01 |
| | Reading | 16 | 4 | 24 | 12.74 | 0.54 | 5.10 | 53.15 | 20.36 |
| | Writing | 10 | 4 | 24 | 12.93 | 0.57 | 4.44 | 53.98 | 17.09 |

**Table 10-21. Summary Statistics for Content Standards Raw and SPI Scores, Mathematics**

| Grade | Content Standard | Standard | No. of Items MC | No. of Items CR | Total Score Points | Mean | Mean p-Value | SD | SPI Mean | SPI SD |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | A | Operations and Algebraic Thinking | 4 | 5 | 9 | 5.21 | 0.58 | 2.49 | 57.82 | 25.49 |
| | B | Number and Operations in Base Ten | 4 | 4 | 8 | 4.25 | 0.53 | 2.40 | 53.24 | 27.26 |
| | C | Number and Operations—Fractions | 5 | 3 | 8 | 4.46 | 0.56 | 2.31 | 55.78 | 25.02 |
| | D | Measurement and Data | 6 | 4 | 10 | 4.88 | 0.49 | 2.63 | 49.11 | 23.87 |
| | E | Geometry | 3 | 4 | 7 | 3.94 | 0.57 | 1.86 | 56.12 | 22.76 |
| 4 | A | Operations and Algebraic Thinking | 7 | 3 | 10 | 5.10 | 0.51 | 2.58 | 51.10 | 23.26 |
| | B | Number and Operations in Base Ten | 5 | 4 | 9 | 4.77 | 0.53 | 2.42 | 52.92 | 24.14 |
| | C | Number and Operations—Fractions | 7 | 3 | 10 | 4.95 | 0.50 | 2.97 | 49.56 | 27.23 |
| | D | Measurement and Data | 8 | 2 | 10 | 4.37 | 0.44 | 2.35 | 43.92 | 20.63 |
| | E | Geometry | 4 | 3 | 7 | 3.63 | 0.52 | 1.91 | 51.55 | 22.28 |
| 5 | A | Operations and Algebraic Thinking | 6 | 3 | 9 | 4.02 | 0.45 | 2.29 | 44.79 | 22.41 |
| | B | Number and Operations in Base Ten | 4 | 5 | 9 | 4.49 | 0.50 | 2.44 | 49.90 | 24.43 |
| | C | Number and Operations—Fractions | 6 | 3 | 9 | 3.70 | 0.41 | 2.34 | 41.40 | 23.14 |
| | D | Measurement and Data | 6 | 4 | 10 | 4.55 | 0.46 | 2.45 | 45.35 | 20.90 |
| | E | Geometry | 5 | 4 | 9 | 4.23 | 0.47 | 2.43 | 46.90 | 23.65 |
| 6 | E | Geometry | 4 | 3 | 7 | 2.91 | 0.42 | 2.00 | 41.68 | 24.65 |
| | F | Ratios and Proportional Relationships | 4 | 3 | 7 | 2.94 | 0.42 | 1.88 | 42.09 | 23.09 |
| | G | The Number System | 6 | 5 | 11 | 4.98 | 0.46 | 2.79 | 45.14 | 23.26 |
| | H | Expressions and Equations | 7 | 4 | 11 | 4.61 | 0.42 | 3.05 | 42.09 | 25.52 |
| | I | Statistics and Probability | 8 | 2 | 10 | 4.64 | 0.47 | 2.50 | 46.27 | 21.50 |

**Table 10-21 (continued). Summary Statistics for Content Standards Raw and SPI Scores, Mathematics**

| Grade | Content Standard | Standard | No. of Items | | Total Score Points | Mean | Mean p-Value | SD | SPI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MC | CR | | | | | Mean | SD |
| 7 | E | Geometry | 8 | 2 | 10 | 4.26 | 0.43 | 2.10 | 42.64 | 17.12 |
| | F | Ratios and Proportional Relationships | 4 | 4 | 8 | 3.33 | 0.42 | 2.32 | 41.72 | 26.36 |
| | G | The Number System | 4 | 4 | 8 | 3.27 | 0.41 | 2.24 | 41.04 | 24.78 |
| | H | Expressions and Equations | 6 | 3 | 9 | 4.03 | 0.45 | 2.22 | 45.05 | 20.72 |
| | I | Statistics and Probability | 6 | 5 | 11 | 4.64 | 0.43 | 2.85 | 42.13 | 23.57 |
| 8 | E | Geometry | 7 | 3 | 10 | 3.95 | 0.40 | 2.46 | 39.37 | 21.54 |
| | G | The Number System | 5 | 3 | 8 | 3.13 | 0.39 | 2.39 | 39.44 | 26.10 |
| | H | Expressions and Equations | 7 | 3 | 10 | 4.14 | 0.42 | 2.43 | 41.74 | 21.11 |
| | I | Statistics and Probability | 6 | 2 | 8 | 3.90 | 0.49 | 1.99 | 48.39 | 21.05 |
| | J | Functions | 6 | 4 | 10 | 4.51 | 0.45 | 2.60 | 44.92 | 23.72 |

**Table 10-22. Summary Statistics for Content Standards Raw and SPI Scores, Science**

| Grade | Content Standard | Standard | No. of Items MC | No. of Items CR | Total Score Points | Mean | Mean *p*-Value | SD | SPI Mean | SPI SD |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | A | Life Science | 3 | 7 | 10 | 5.56 | 0.56 | 2.30 | 55.51 | 20.00 |
| | B | Physical Science | 6 | 7 | 13 | 8.08 | 0.62 | 2.84 | 62.01 | 20.00 |
| | C | Earth and Space Science | 4 | 5 | 9 | 3.75 | 0.42 | 2.08 | 42.13 | 19.36 |
| | D | Engineering | 2 | 6 | 8 | 4.06 | 0.51 | 2.33 | 50.87 | 26.25 |
| 8 | A | Life Science | 2 | 8 | 10 | 5.23 | 0.53 | 2.41 | 52.43 | 21.36 |
| | B | Physical Science | 4 | 7 | 11 | 5.34 | 0.49 | 2.65 | 48.76 | 21.29 |
| | C | Earth and Space Science | 3 | 8 | 11 | 5.33 | 0.49 | 2.31 | 48.54 | 17.83 |
| | D | Engineering | 2 | 6 | 8 | 4.94 | 0.62 | 2.10 | 61.47 | 23.38 |

**Table 10-23. Summary Statistics for Content Standards Raw and SPI Scores, Social Studies**

| Grade | Content Standard | Standard | No. of Items MC | No. of Items CR | Total Score Points | Mean | Mean *p*-Value | SD | SPI Mean | SPI SD |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | A | Geography | 6 | 5 | 11 | 6.75 | 0.62 | 2.98 | 61.38 | 25.15 |
| | B | History | 7 | 1 | 8 | 4.49 | 0.56 | 2.05 | 56.37 | 21.76 |
| | C | Political Science | 2 | 5 | 7 | 3.71 | 0.53 | 1.68 | 53.15 | 19.22 |
| | D | Economics | 6 | 1 | 7 | 3.66 | 0.52 | 1.76 | 52.66 | 20.42 |
| | E | Behavioral Sciences | 3 | 4 | 7 | 4.20 | 0.60 | 1.95 | 59.99 | 24.73 |
| 8 | A | Geography | 7 | 2 | 9 | 5.31 | 0.59 | 2.35 | 59.11 | 23.13 |
| | B | History | 8 | 1 | 9 | 4.90 | 0.55 | 2.16 | 54.86 | 20.49 |
| | C | Political Science | 6 | 2 | 8 | 4.84 | 0.61 | 2.26 | 60.49 | 25.85 |
| | D | Economics | 7 | 1 | 8 | 4.69 | 0.59 | 1.99 | 58.81 | 21.16 |
| | E | Behavioral Sciences | 5 | 1 | 6 | 3.73 | 0.62 | 1.69 | 61.93 | 23.99 |
| 10 | A | Geography | 8 | 2 | 10 | 5.69 | 0.58 | 2.45 | 56.90 | 21.81 |
| | B | History | 7 | 2 | 9 | 4.75 | 0.54 | 2.24 | 52.90 | 22.02 |
| | C | Political Science | 6 | 3 | 9 | 4.84 | 0.55 | 2.37 | 53.90 | 23.59 |
| | D | Economics | 5 | 1 | 6 | 2.97 | 0.50 | 1.70 | 49.68 | 23.12 |
| | E | Behavioral Sciences | 4 | 2 | 6 | 2.81 | 0.48 | 1.51 | 47.46 | 19.02 |

**Table 10-24A. SPI Cut Scores, English Language Arts, Grades 3–5**

| Content Standard/ Domain | Performance Level | Grade 3 | | Grade 4 | | Grade 5 | |
|---|---|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| Reading—Key Ideas and Details | 1 | 0 | 32 | 0 | 34 | 0 | 34 |
| | 2 | 33 | 66 | 35 | 56 | 35 | 57 |
| | 3 | 67 | 89 | 57 | 79 | 58 | 86 |
| | 4 | 90 | 100 | 80 | 100 | 87 | 100 |
| Reading—Craft & Structure | 1 | 0 | 36 | 0 | 31 | 0 | 34 |
| | 2 | 37 | 56 | 32 | 53 | 35 | 55 |
| | 3 | 57 | 81 | 54 | 80 | 56 | 85 |
| | 4 | 82 | 100 | 81 | 100 | 86 | 100 |
| Reading—Vocabulary Use | 1 | 0 | 42 | 0 | 50 | 0 | 49 |
| | 2 | 43 | 69 | 51 | 77 | 50 | 63 |
| | 3 | 70 | 92 | 78 | 94 | 64 | 81 |
| | 4 | 93 | 100 | 95 | 100 | 82 | 100 |
| Writing/Language—Text Types and Purposes | 1 | 0 | 33 | 0 | 33 | 0 | 33 |
| | 2 | 34 | 44 | 34 | 42 | 34 | 42 |
| | 3 | 45 | 57 | 43 | 58 | 43 | 60 |
| | 4 | 58 | 100 | 59 | 100 | 61 | 100 |
| Writing/Language— Research | 1 | 0 | 46 | 0 | 35 | 0 | 51 |
| | 2 | 47 | 68 | 36 | 56 | 52 | 75 |
| | 3 | 69 | 84 | 57 | 81 | 76 | 92 |
| | 4 | 85 | 100 | 82 | 100 | 93 | 100 |
| Writing/Language— Language Conventions | 1 | 0 | 41 | 0 | 49 | 0 | 49 |
| | 2 | 42 | 62 | 50 | 66 | 50 | 68 |
| | 3 | 63 | 84 | 67 | 83 | 69 | 85 |
| | 4 | 85 | 100 | 84 | 100 | 86 | 100 |
| Listening | 1 | 0 | 44 | 0 | 38 | 0 | 53 |
| | 2 | 45 | 65 | 39 | 55 | 54 | 73 |
| | 3 | 66 | 82 | 56 | 75 | 74 | 89 |
| | 4 | 83 | 100 | 76 | 100 | 90 | 100 |
| Reading | 1 | 0 | 35 | 0 | 37 | 0 | 37 |
| | 2 | 36 | 63 | 38 | 60 | 38 | 57 |
| | 3 | 64 | 87 | 61 | 83 | 58 | 85 |
| | 4 | 88 | 100 | 84 | 100 | 86 | 100 |
| Writing | 1 | 0 | 38 | 0 | 38 | 0 | 42 |
| | 2 | 39 | 55 | 39 | 52 | 43 | 57 |
| | 3 | 56 | 71 | 53 | 70 | 58 | 74 |
| | 4 | 72 | 100 | 71 | 100 | 75 | 100 |

**Table 10-24B. SPI Cut Scores, English Language Arts, Grades 6–8**

| Content Standard/ Domain | Performance Level | Grade 6 | | Grade 7 | | Grade 8 | |
|---|---|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| Reading—Key Ideas and Details | 1 | 0 | 42 | 0 | 28 | 0 | 36 |
| | 2 | 43 | 70 | 29 | 50 | 37 | 58 |
| | 3 | 71 | 87 | 51 | 74 | 59 | 76 |
| | 4 | 88 | 100 | 75 | 100 | 77 | 100 |
| Reading—Craft & Structure | 1 | 0 | 41 | 0 | 28 | 0 | 40 |
| | 2 | 42 | 71 | 29 | 52 | 41 | 63 |
| | 3 | 72 | 90 | 53 | 80 | 64 | 80 |
| | 4 | 91 | 100 | 81 | 100 | 81 | 100 |
| Reading—Vocabulary Use | 1 | 0 | 48 | 0 | 54 | 0 | 43 |
| | 2 | 49 | 75 | 55 | 82 | 44 | 77 |
| | 3 | 76 | 92 | 83 | 94 | 78 | 93 |
| | 4 | 93 | 100 | 95 | 100 | 94 | 100 |
| Writing/Language—Text Types and Purposes | 1 | 0 | 33 | 0 | 36 | 0 | 42 |
| | 2 | 34 | 47 | 37 | 51 | 43 | 58 |
| | 3 | 48 | 63 | 52 | 68 | 59 | 72 |
| | 4 | 64 | 100 | 69 | 100 | 73 | 100 |
| Writing/Language— Research | 1 | 0 | 32 | 0 | 38 | 0 | 36 |
| | 2 | 33 | 47 | 39 | 59 | 37 | 64 |
| | 3 | 48 | 65 | 60 | 77 | 65 | 85 |
| | 4 | 66 | 100 | 78 | 100 | 86 | 100 |
| Writing/Language— Language Conventions | 1 | 0 | 41 | 0 | 56 | 0 | 52 |
| | 2 | 42 | 57 | 57 | 71 | 53 | 66 |
| | 3 | 58 | 77 | 72 | 84 | 67 | 80 |
| | 4 | 78 | 100 | 85 | 100 | 81 | 100 |
| Listening | 1 | 0 | 47 | 0 | 51 | 0 | 43 |
| | 2 | 48 | 71 | 52 | 76 | 44 | 68 |
| | 3 | 72 | 86 | 77 | 91 | 69 | 84 |
| | 4 | 87 | 100 | 92 | 100 | 85 | 100 |
| Reading | 1 | 0 | 43 | 0 | 34 | 0 | 39 |
| | 2 | 44 | 71 | 35 | 57 | 40 | 64 |
| | 3 | 72 | 89 | 58 | 80 | 65 | 81 |
| | 4 | 90 | 100 | 81 | 100 | 82 | 100 |
| Writing | 1 | 0 | 35 | 0 | 42 | 0 | 43 |
| | 2 | 36 | 50 | 43 | 59 | 44 | 61 |
| | 3 | 51 | 67 | 60 | 75 | 62 | 77 |
| | 4 | 68 | 100 | 76 | 100 | 78 | 100 |

**Table 10-25A. SPI Cut Scores, Mathematics, Grades 3–5**

| Content Standard/ Domain | Performance Level | Grade 3 | | Grade 4 | | Grade 5 | |
|---|---|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| **Operations and Algebraic Thinking** | 1 | 0 | 35 | 0 | 28 | 0 | 26 |
| | 2 | 36 | 63 | 29 | 55 | 27 | 45 |
| | 3 | 64 | 87 | 56 | 79 | 46 | 75 |
| | 4 | 88 | 100 | 80 | 100 | 76 | 100 |
| **Number and Operations in Base Ten** | 1 | 0 | 26 | 0 | 30 | 0 | 30 |
| | 2 | 27 | 56 | 31 | 57 | 31 | 54 |
| | 3 | 57 | 86 | 58 | 82 | 55 | 81 |
| | 4 | 87 | 100 | 83 | 100 | 82 | 100 |
| **Number and Operations—Fractions** | 1 | 0 | 32 | 0 | 22 | 0 | 22 |
| | 2 | 33 | 57 | 23 | 53 | 23 | 39 |
| | 3 | 58 | 87 | 54 | 85 | 40 | 75 |
| | 4 | 88 | 100 | 86 | 100 | 76 | 100 |
| **Measurement and Data** | 1 | 0 | 27 | 0 | 25 | 0 | 29 |
| | 2 | 28 | 51 | 26 | 44 | 30 | 46 |
| | 3 | 52 | 78 | 45 | 68 | 47 | 73 |
| | 4 | 79 | 100 | 69 | 100 | 74 | 100 |
| **Geometry** | 1 | 0 | 36 | 0 | 31 | 0 | 28 |
| | 2 | 37 | 59 | 32 | 54 | 29 | 49 |
| | 3 | 60 | 83 | 55 | 78 | 50 | 78 |
| | 4 | 84 | 100 | 79 | 100 | 79 | 100 |

**Table 10-25B. SPI Cut Scores, Mathematics, Grades 6–8**

| Content Standard/ Domain | Performance Level | Grade 6 | | Grade 7 | | Grade 8 | |
|---|---|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| Geometry | 1 | 0 | 23 | 0 | 33 | 0 | 27 |
| | 2 | 24 | 45 | 34 | 46 | 28 | 46 |
| | 3 | 46 | 86 | 47 | 76 | 47 | 78 |
| | 4 | 87 | 100 | 77 | 100 | 79 | 100 |
| Ratios and Proportional Relationships* | 1 | 0 | 26 | 0 | 26 | | |
| | 2 | 27 | 46 | 27 | 53 | | |
| | 3 | 47 | 82 | 54 | 88 | | |
| | 4 | 83 | 100 | 89 | 100 | | |
| The Number System | 1 | 0 | 29 | 0 | 26 | 0 | 21 |
| | 2 | 30 | 51 | 27 | 51 | 22 | 53 |
| | 3 | 52 | 84 | 52 | 85 | 54 | 83 |
| | 4 | 85 | 100 | 86 | 100 | 84 | 100 |
| Expressions and Equations | 1 | 0 | 22 | 0 | 32 | 0 | 27 |
| | 2 | 23 | 46 | 33 | 50 | 28 | 48 |
| | 3 | 47 | 87 | 51 | 85 | 49 | 80 |
| | 4 | 88 | 100 | 86 | 100 | 81 | 100 |
| Statistics and Probability | 1 | 0 | 32 | 0 | 29 | 0 | 37 |
| | 2 | 33 | 50 | 30 | 51 | 38 | 59 |
| | 3 | 51 | 83 | 52 | 85 | 60 | 82 |
| | 4 | 84 | 100 | 86 | 100 | 83 | 100 |
| Functions** | 1 | | | | | 0 | 31 |
| | 2 | | | | | 32 | 57 |
| | 3 | | | | | 58 | 84 |
| | 4 | | | | | 85 | 100 |

* Content standard in grades 6 and 7 only.
** Content standard in grade 8 only.

**Table 10-26. SPI Cut Scores, Science**

| Content Standard/ Domain | Performance Level | Grade 4 | | Grade 8 | |
|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| Life Science | 1 | 0 | 34 | 0 | 32 |
| | 2 | 35 | 56 | 33 | 51 |
| | 3 | 57 | 74 | 52 | 72 |
| | 4 | 75 | 100 | 73 | 100 |
| Physical Science | 1 | 0 | 41 | 0 | 28 |
| | 2 | 42 | 63 | 29 | 47 |
| | 3 | 64 | 81 | 48 | 68 |
| | 4 | 82 | 100 | 69 | 100 |
| Earth and Space Science | 1 | 0 | 22 | 0 | 32 |
| | 2 | 23 | 38 | 33 | 47 |
| | 3 | 39 | 61 | 48 | 64 |
| | 4 | 62 | 100 | 65 | 100 |
| Engineering | 1 | 0 | 21 | 0 | 39 |
| | 2 | 22 | 50 | 40 | 63 |
| | 3 | 51 | 78 | 64 | 84 |
| | 4 | 79 | 100 | 85 | 100 |

**Table 10-27. SPI Cut Scores, Social Studies**

| Content Standard/ Domain | Performance Level | Grade 4 | | Grade 8 | | Grade 10 | |
|---|---|---|---|---|---|---|---|
| | | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound | Score Lower Bound | Score Upper Bound |
| Geography | 1 | 0 | 36 | 0 | 34 | 0 | 38 |
| | 2 | 37 | 57 | 35 | 53 | 39 | 59 |
| | 3 | 58 | 84 | 54 | 79 | 60 | 77 |
| | 4 | 85 | 100 | 80 | 100 | 78 | 100 |
| History | 1 | 0 | 34 | 0 | 33 | 0 | 34 |
| | 2 | 35 | 50 | 34 | 48 | 35 | 54 |
| | 3 | 51 | 75 | 49 | 71 | 55 | 72 |
| | 4 | 76 | 100 | 72 | 100 | 73 | 100 |
| Political Science | 1 | 0 | 35 | 0 | 33 | 0 | 34 |
| | 2 | 36 | 50 | 34 | 56 | 35 | 53 |
| | 3 | 51 | 68 | 57 | 83 | 54 | 75 |
| | 4 | 69 | 100 | 84 | 100 | 76 | 100 |
| Economics | 1 | 0 | 33 | 0 | 37 | 0 | 30 |
| | 2 | 34 | 45 | 38 | 53 | 31 | 49 |
| | 3 | 46 | 69 | 54 | 76 | 50 | 70 |
| | 4 | 70 | 100 | 77 | 100 | 71 | 100 |
| Behavioral Sciences | 1 | 0 | 36 | 0 | 37 | 0 | 32 |
| | 2 | 37 | 56 | 38 | 58 | 33 | 45 |
| | 3 | 57 | 81 | 59 | 82 | 46 | 62 |
| | 4 | 82 | 100 | 83 | 100 | 63 | 100 |

**Table 10-28. Longitudinal Comparison of State-Level Participation Rates and Scale Score Means, English Language Arts**

| Grade | Year | Enrolled | Number Tested | Percent Tested | Scale Score Mean | Scale Score SD |
|---|---|---|---|---|---|---|
| 3 | 2016 | 65793 | 64107 | 97.44 | 560.57 | 47.31 |
| | 2017 | 65340 | 63946 | 97.87 | 559.12 | 46.93 |
| | 2018 | 64693 | 63194 | 97.68 | 556.70 | 46.66 |
| | 2019 | 62646 | 61091 | 97.52 | 554.59 | 45.54 |
| | 2021 | 60785 | 52930 | 87.08 | 550.13 | 46.61 |
| | 2022 | 60956 | 58275 | 95.60 | 549.75 | 47.41 |
| 4 | 2016 | 64361 | 62609 | 97.28 | 582.71 | 49.41 |
| | 2017 | 66001 | 64423 | 97.61 | 585.26 | 52.44 |
| | 2018 | 65885 | 64354 | 97.68 | 580.90 | 51.81 |
| | 2019 | 65222 | 63528 | 97.40 | 582.01 | 51.05 |
| | 2021 | 61127 | 52706 | 86.22 | 577.65 | 51.31 |
| | 2022 | 61465 | 58773 | 95.62 | 578.32 | 52.20 |
| 5 | 2016 | 64045 | 62300 | 97.28 | 599.62 | 51.11 |
| | 2017 | 64624 | 62995 | 97.48 | 603.24 | 51.00 |
| | 2018 | 66542 | 64903 | 97.54 | 600.78 | 48.35 |
| | 2019 | 66250 | 64654 | 97.59 | 595.58 | 48.77 |
| | 2021 | 62405 | 54010 | 86.55 | 593.12 | 49.01 |
| | 2022 | 61735 | 59044 | 95.64 | 593.86 | 52.07 |
| 6 | 2016 | 64594 | 62728 | 97.11 | 610.36 | 52.16 |
| | 2017 | 64446 | 62754 | 97.37 | 614.59 | 49.82 |
| | 2018 | 65363 | 63600 | 97.30 | 609.61 | 50.18 |
| | 2019 | 67243 | 65386 | 97.24 | 607.00 | 50.15 |
| | 2021 | 64925 | 55511 | 85.50 | 604.02 | 50.12 |
| | 2022 | 63048 | 60112 | 95.34 | 602.26 | 51.02 |
| 7 | 2016 | 64044 | 62084 | 96.94 | 623.84 | 54.85 |
| | 2017 | 65049 | 63091 | 96.99 | 626.80 | 59.14 |
| | 2018 | 64975 | 63140 | 97.18 | 627.43 | 56.56 |
| | 2019 | 65904 | 63878 | 96.93 | 627.70 | 54.88 |
| | 2021 | 66361 | 56295 | 84.83 | 625.36 | 55.21 |
| | 2022 | 65386 | 61871 | 94.62 | 618.80 | 56.30 |
| 8 | 2016 | 63861 | 61486 | 96.28 | 637.23 | 57.27 |
| | 2017 | 64265 | 62109 | 96.65 | 637.69 | 61.61 |
| | 2018 | 65638 | 63248 | 96.36 | 630.98 | 59.94 |
| | 2019 | 65355 | 63056 | 96.48 | 629.06 | 59.84 |
| | 2021 | 67572 | 56756 | 83.99 | 628.22 | 58.76 |
| | 2022 | 66789 | 62684 | 93.85 | 622.78 | 61.00 |

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates lower than in a typical administration year, resulting in overrepresentation of some subgroups and underrepresentation of other subgroups.

**Table 10-29. Longitudinal Comparison of State-Level Participation Rates and Scale Score Means, Mathematics**

| Grade | Year | Enrolled | Number Tested | Percent Tested | Scale Score Mean | Scale Score SD |
|-------|------|----------|---------------|----------------|------------------|----------------|
| 3 | 2016 | 65793 | 64194 | 97.57 | 554.28 | 46.47 |
| | 2017 | 65340 | 64066 | 98.05 | 555.03 | 48.63 |
| | 2018 | 64693 | 63314 | 97.87 | 555.94 | 50.87 |
| | 2019 | 62646 | 61210 | 97.71 | 555.78 | 53.50 |
| | 2021 | 60785 | 52892 | 87.01 | 548.97 | 56.39 |
| | 2022 | 60956 | 58449 | 95.89 | 552.25 | 56.04 |
| 4 | 2016 | 64361 | 62674 | 97.38 | 573.45 | 56.15 |
| | 2017 | 66001 | 64533 | 97.78 | 574.33 | 54.92 |
| | 2018 | 65885 | 64462 | 97.84 | 576.76 | 52.99 |
| | 2019 | 65222 | 63630 | 97.56 | 577.09 | 51.78 |
| | 2021 | 61127 | 52658 | 86.15 | 571.50 | 53.33 |
| | 2022 | 61465 | 58931 | 95.88 | 573.54 | 57.00 |
| 5 | 2016 | 64045 | 62368 | 97.38 | 599.57 | 50.19 |
| | 2017 | 64624 | 63152 | 97.72 | 599.73 | 51.00 |
| | 2018 | 66542 | 65021 | 97.71 | 598.82 | 56.65 |
| | 2019 | 66250 | 64728 | 97.70 | 601.48 | 53.14 |
| | 2021 | 62405 | 53932 | 86.42 | 594.26 | 56.01 |
| | 2022 | 61735 | 59187 | 95.87 | 599.38 | 52.67 |
| 6 | 2016 | 64594 | 62772 | 97.18 | 612.67 | 53.00 |
| | 2017 | 64446 | 62847 | 97.52 | 612.93 | 54.81 |
| | 2018 | 65363 | 63669 | 97.41 | 611.97 | 57.64 |
| | 2019 | 67243 | 65470 | 97.36 | 610.77 | 58.31 |
| | 2021 | 64925 | 55462 | 85.42 | 602.06 | 57.74 |
| | 2022 | 63048 | 60234 | 95.54 | 604.64 | 59.36 |
| 7 | 2016 | 64044 | 62144 | 97.03 | 627.49 | 57.40 |
| | 2017 | 65049 | 63200 | 97.16 | 627.48 | 58.65 |
| | 2018 | 64975 | 63218 | 97.30 | 622.82 | 65.55 |
| | 2019 | 65904 | 63973 | 97.07 | 625.25 | 60.69 |
| | 2021 | 66361 | 56247 | 84.76 | 620.00 | 60.06 |
| | 2022 | 65386 | 61969 | 94.77 | 617.55 | 62.13 |
| 8 | 2016 | 63861 | 61551 | 96.38 | 640.79 | 57.54 |
| | 2017 | 64265 | 62175 | 96.75 | 641.11 | 59.36 |
| | 2018 | 65638 | 63318 | 96.47 | 644.24 | 60.78 |
| | 2019 | 65355 | 63108 | 96.56 | 644.53 | 57.85 |
| | 2021 | 67572 | 56726 | 83.95 | 638.33 | 56.94 |
| | 2022 | 66789 | 62762 | 93.97 | 634.95 | 60.04 |

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation lower than in a typical administration year, resulting in overrepresentation of some subgroups and underrepresentation of other subgroups.

**Table 10-30. Longitudinal Comparison of State-Level Participation Rates and Scale Score Means, Science**

| Grade | Year | Enrolled | Number Tested | Percent Tested | Scale Score Mean | Scale Score SD |
|---|---|---|---|---|---|---|
| 4 | 2019 | 65222 | 63611 | 97.53 | 499.88 | 50.24 |
| | 2021 | 61127 | 52417 | 85.75 | 497.39 | 50.30 |
| | 2022 | 61465 | 58880 | 95.79 | 495.81 | 52.11 |
| 8 | 2019 | 65355 | 63062 | 96.49 | 699.70 | 50.55 |
| | 2021 | 67572 | 56485 | 83.59 | 697.31 | 49.86 |
| | 2022 | 66789 | 62644 | 93.79 | 693.86 | 51.11 |

Note 1: New reporting scales were established for Science in Spring 2019.
Note 2: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates lower than in a typical administration year, resulting in overrepresentation of some subgroups and underrepresentation of other subgroups.

**Table 10-31. Baseline Year State-Level Participation Rates Scale Score Means, Social Studies**

| Grade | Year | Enrolled | Number Tested | Percent Tested | Scale Score Mean | Scale Score SD |
|---|---|---|---|---|---|---|
| 4 | 2022 | 61465 | 58833 | 95.72 | 499.85 | 50.27 |
| 8 | 2022 | 66789 | 62606 | 93.74 | 699.24 | 50.80 |
| 10 | 2022 | 67427 | 59391 | 88.08 | 798.49 | 50.78 |

Note: New reporting scales were established for Social Studies in Spring 2022.

**Table 10-32. Longitudinal Comparison of State-Level Impact Data, English Language Arts**

| Grade | Year | N | Below Basic | Basic | Proficient | Advanced | Prof. & Adv. |
|-------|------|------|------|------|------|------|------|
| 3 | 2016 | 64107 | 21.99 | 34.88 | 34.29 | 8.84 | 43.13 |
| | 2017 | 63946 | 21.45 | 36.72 | 33.81 | 8.02 | 41.83 |
| | 2018 | 63194 | 22.78 | 37.47 | 32.58 | 7.17 | 39.75 |
| | 2019 | 61091 | 23.28 | 38.04 | 33.21 | 5.48 | 38.69 |
| | 2021 | 52930 | 27.71 | 37.74 | 29.48 | 5.08 | 34.56 |
| | 2022 | 58275 | 27.55 | 37.52 | 29.61 | 5.32 | 34.93 |
| 4 | 2016 | 62609 | 22.81 | 33.88 | 34.77 | 8.54 | 43.30 |
| | 2017 | 64423 | 21.14 | 32.14 | 37.00 | 9.71 | 46.72 |
| | 2018 | 64354 | 24.04 | 32.06 | 35.72 | 8.19 | 43.91 |
| | 2019 | 63528 | 23.88 | 33.14 | 34.10 | 8.89 | 42.98 |
| | 2021 | 52706 | 27.17 | 32.71 | 32.67 | 7.45 | 40.12 |
| | 2022 | 58773 | 26.66 | 31.74 | 33.63 | 7.96 | 41.60 |
| 5 | 2016 | 62300 | 23.17 | 34.37 | 34.55 | 7.91 | 42.47 |
| | 2017 | 62995 | 20.36 | 33.22 | 37.88 | 8.54 | 46.42 |
| | 2018 | 64903 | 21.53 | 34.30 | 37.40 | 6.77 | 44.17 |
| | 2019 | 64654 | 26.11 | 33.83 | 34.34 | 5.72 | 40.06 |
| | 2021 | 54010 | 27.67 | 34.80 | 32.41 | 5.12 | 37.52 |
| | 2022 | 59044 | 27.99 | 31.73 | 34.21 | 6.06 | 40.28 |
| 6 | 2016 | 62728 | 21.12 | 36.30 | 31.67 | 10.91 | 42.58 |
| | 2017 | 62754 | 18.23 | 36.52 | 33.51 | 11.75 | 45.26 |
| | 2018 | 63600 | 22.06 | 35.08 | 32.73 | 10.12 | 42.86 |
| | 2019 | 65386 | 23.56 | 35.48 | 31.87 | 9.09 | 40.96 |
| | 2021 | 55511 | 25.18 | 36.36 | 30.65 | 7.80 | 38.45 |
| | 2022 | 60112 | 26.19 | 35.83 | 30.71 | 7.27 | 37.98 |
| 7 | 2016 | 62084 | 23.11 | 34.91 | 34.09 | 7.89 | 41.98 |
| | 2017 | 63091 | 22.27 | 34.10 | 33.52 | 10.11 | 43.63 |
| | 2018 | 63140 | 21.29 | 33.57 | 35.72 | 9.43 | 45.15 |
| | 2019 | 63878 | 21.88 | 33.25 | 35.36 | 9.51 | 44.87 |
| | 2021 | 56295 | 23.08 | 33.99 | 34.12 | 8.81 | 42.92 |
| | 2022 | 61871 | 26.71 | 35.03 | 30.70 | 7.55 | 38.26 |
| 8 | 2016 | 61486 | 21.24 | 37.21 | 31.26 | 10.30 | 41.56 |
| | 2017 | 62109 | 21.66 | 37.22 | 29.19 | 11.93 | 41.12 |
| | 2018 | 63248 | 24.66 | 38.01 | 27.93 | 9.40 | 37.33 |
| | 2019 | 63056 | 25.94 | 37.04 | 28.80 | 8.23 | 37.03 |
| | 2021 | 56756 | 26.05 | 38.29 | 28.02 | 7.64 | 35.66 |
| | 2022 | 62684 | 29.81 | 36.72 | 26.45 | 7.02 | 33.46 |

**Table 10-33. Longitudinal Comparison of State-Level Impact Data, Mathematics**

| Grade | Year | N | Below Basic | Basic | Proficient | Advanced | Prof. & Adv. |
|---|---|---|---|---|---|---|---|
| 3 | 2016 | 64194 | 18.59 | 33.41 | 38.90 | 9.10 | 48.00 |
| | 2017 | 64066 | 18.90 | 33.06 | 37.84 | 10.20 | 48.03 |
| | 2018 | 63314 | 18.68 | 31.48 | 38.47 | 11.37 | 49.83 |
| | 2019 | 61210 | 19.28 | 31.28 | 37.17 | 12.27 | 49.44 |
| | 2021 | 52892 | 23.78 | 31.23 | 34.46 | 10.54 | 44.99 |
| | 2022 | 58449 | 22.76 | 30.04 | 35.02 | 12.18 | 47.20 |
| 4 | 2016 | 62674 | 19.59 | 36.22 | 33.33 | 10.86 | 44.20 |
| | 2017 | 64533 | 19.13 | 37.37 | 32.67 | 10.83 | 43.50 |
| | 2018 | 64462 | 18.37 | 37.17 | 32.71 | 11.74 | 44.46 |
| | 2019 | 63630 | 18.87 | 36.09 | 32.82 | 12.23 | 45.05 |
| | 2021 | 52658 | 22.32 | 36.61 | 30.84 | 10.23 | 41.07 |
| | 2022 | 58931 | 22.24 | 34.03 | 30.95 | 12.79 | 43.73 |
| 5 | 2016 | 62368 | 25.94 | 29.98 | 34.14 | 9.94 | 44.08 |
| | 2017 | 63152 | 24.97 | 30.57 | 34.58 | 9.88 | 44.46 |
| | 2018 | 65021 | 24.73 | 29.32 | 35.05 | 10.90 | 45.95 |
| | 2019 | 64728 | 24.22 | 29.20 | 35.09 | 11.49 | 46.58 |
| | 2021 | 53932 | 29.46 | 28.94 | 32.12 | 9.47 | 41.59 |
| | 2022 | 59187 | 26.82 | 28.38 | 34.28 | 10.52 | 44.80 |
| 6 | 2016 | 62772 | 25.51 | 31.66 | 36.78 | 6.05 | 42.84 |
| | 2017 | 62847 | 24.70 | 31.68 | 37.50 | 6.11 | 43.61 |
| | 2018 | 63669 | 24.78 | 31.27 | 37.78 | 6.18 | 43.96 |
| | 2019 | 65470 | 26.72 | 30.79 | 35.80 | 6.69 | 42.49 |
| | 2021 | 55462 | 32.61 | 31.82 | 30.90 | 4.66 | 35.57 |
| | 2022 | 60234 | 30.96 | 30.28 | 33.27 | 5.49 | 38.76 |
| 7 | 2016 | 62144 | 30.45 | 30.28 | 34.81 | 4.45 | 39.26 |
| | 2017 | 63200 | 30.80 | 29.92 | 34.53 | 4.75 | 39.29 |
| | 2018 | 63218 | 31.36 | 29.67 | 34.33 | 4.64 | 38.97 |
| | 2019 | 63973 | 32.18 | 28.99 | 34.05 | 4.78 | 38.83 |
| | 2021 | 56247 | 34.99 | 30.17 | 31.38 | 3.47 | 34.84 |
| | 2022 | 61969 | 38.11 | 28.16 | 29.58 | 4.15 | 33.73 |
| 8 | 2016 | 61551 | 28.66 | 37.48 | 28.12 | 5.74 | 33.86 |
| | 2017 | 62175 | 28.43 | 36.95 | 28.33 | 6.29 | 34.62 |
| | 2018 | 63318 | 27.95 | 35.44 | 28.71 | 7.90 | 36.61 |
| | 2019 | 63108 | 28.55 | 35.60 | 27.83 | 8.01 | 35.85 |
| | 2021 | 56726 | 32.47 | 37.53 | 23.62 | 6.38 | 30.00 |
| | 2022 | 62762 | 35.96 | 33.73 | 24.05 | 6.26 | 30.31 |

**Table 10-34. Longitudinal Comparison of State-Level Impact Data, Science**

| Grade | Year | N | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Prof. & Adv.* |
|-------|------|-------|-------|-------|-------|-------|-------|
| 4 | 2019 | 63611 | 14.98 | 32.25 | 33.29 | 19.49 | 52.78 |
|   | 2021 | 52417 | 16.18 | 32.67 | 33.03 | 18.13 | 51.16 |
|   | 2022 | 58880 | 18.20 | 30.95 | 32.32 | 18.53 | 50.85 |
| 8 | 2019 | 63062 | 17.76 | 28.29 | 31.50 | 22.45 | 53.95 |
|   | 2021 | 56485 | 18.56 | 29.97 | 30.65 | 20.82 | 51.47 |
|   | 2022 | 62644 | 20.77 | 30.33 | 28.99 | 19.91 | 48.90 |

Note: New cut scores were used to classify students into performance levels after the Spring 2019 Science test administration.

**Table 10-35. Baseline Year State-Level Impact Data, Social Studies**

| Grade | Year | N | *Below Basic* | *Basic* | *Proficient* | *Advanced* | *Prof. & Adv.* |
|-------|------|-------|-------|-------|-------|-------|-------|
| 4 | 2022 | 58833 | 21.02 | 20.09 | 36.79 | 22.10 | 58.89 |
| 8 | 2022 | 62606 | 19.53 | 21.74 | 35.08 | 23.64 | 58.73 |
| 10 | 2022 | 59391 | 25.49 | 26.84 | 25.73 | 21.94 | 47.67 |

Note: New cut scores were used to classify students into performance levels after the Spring 2022 Social Studies test administration.

# Part 11: Summary and Recommendations

Results and key findings of the Spring 2022 Wisconsin Forward Exam administration are presented throughout the body of this report. This last section of the report presents some recommendations for DPI consideration.

The 2022 test administration was the sixth administration of the Forward assessments. Since Spring 2016, the assessment results have been reported on the same scales and students have been classified into performance levels using the same cut scores, allowing for longitudinal tracking of student performance in ELA and Mathematics. New test scales were established, and new performance level cut scores were set for Science assessments after the Spring 2019 test administration. The 2022 Wisconsin Forward Exam administration was the third administration of the Science assessments that measured the new Science standards and reported on the new scales. For Social Studies, new test scales were established and new performance level cut scores were set after the Spring 2022 test administration. The Spring 2022 assessment results serve as the new baseline for monitoring student performance in Social Studies across years.

In keeping with the field-testing of new test items in all content areas in the past administrations, DRC recommends that, in the future, all new items continue to be field-tested in Wisconsin prior to their operational test administration to provide accurate information on how students may perform on these items once they are administered operationally. DRC also recommends continuing to embed field test items in each operational test administration for all content areas in order to build a high-quality Wisconsin item bank for future form development.

DRC recommends continuing to use an artificial intelligence (AI) engine in the scoring of text-dependent analysis items for efficiency and accuracy. As indicated in Part 5 and Part 8 of this report, the AI scores were in good or very good agreement with scores by trained human scorers.

In Mathematics grades 4 through 8, more than 2% of students received the lowest obtainable scale score (LOSS). The Mathematics assessments continue to be difficult for some students. The response patterns of students at the LOSS in Mathematics indicated that these students typically answered very few MC items and none of the non-MC items. As explained in Part 6 of this report, for these students to receive a scale score above the LOSS, they would need to correctly answer more items, including some non-MC items. Therefore, DRC continues to recommend that some easier non-MC items be included in the future forms of Mathematics tests. In addition, more than 2% of students scored at LOSS in Social Studies grade 8. Similar to Mathematics, students who scored at LOSS in Social Studies grade 8 did not answer any of the non-MC items correctly. It is recommended that more non-MC items, including easier items, be included in the Social Studied grade 8 assessment in the future.

From the psychometric perspective and per the peer review recommendation for Science, the magnitude of the conditional standard error, particularly at the lower and upper end of the scale, will be monitored for Science and other content areas. Efforts will be made to include more items of higher discrimination that span the entire range of the difficulty scale with a goal to improve the precision of measurement, particularly for students of lower and higher ability.

# References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bejar, I. I., & Wingersky, M.S. (1982). A study of pre-equating based on item response theory. *Applied Psychological Measurement, 6*(3), 309–325.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, 46, 443–459.

Burket, G. R. (2002). PARDUX, Version 1.54 [Computer program]. CTB/McGraw-Hill.

Camilli, G., & Shepard, A. L. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications.

Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing bias in item response theory. *Applied Psychological Measurement, 12*(3), 253–260.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

Council of Chief State School Officers (2020). Restart and recovery: Assessments in Spring 2021. https://www.nciea.org/sites/default/files/publications/Assessments-Spring-2021.pdf

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Belmont, CA: Wadsworth Group/Thomson Learning.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

CTB/McGraw-Hill. (1997). *TerraNova* (1st ed.). Monterey, CA: Author.

CTB/McGraw-Hill. (2000). *TerraNova* (2nd ed.). Monterey, CA: Author.

CTB/McGraw-Hill. (2009). *TerraNova 3rd Edition Technical Addendum: Forms E and F*. Monterey, CA: Author.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. Princeton, NJ: Educational Testing Service.

Fitzpatrick, A. R. (1991). *Status report on the results of preliminary analyses of dichotomous and multi-level items using the PARMATE program*. Monterey, CA: CTB/McGraw-Hill.

Fitzpatrick, A. R., & Julian, M. W. (1996). *Two studies comparing the parameter estimates produced by PARDUX and PARSCALE*. Unpublished manuscript.

Green, D. R. (December 1975). *Procedures for assessing bias in achievement tests*. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*(3), 159–170.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.

Holland, P. W., & Thayer, D. T. (1985). *An alternate definition of the ETS delta scale of item difficulty* (Research Report No. RR-85-43). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (April 1986). *Differential item performance and the Mantel-Haenszel procedure*. Paper presented at the annual meeting of the American Educational Research Association Annual Meeting, San Francisco, CA.

Karantonis, A., & Sireci, S. G. (2006). The Bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice, 25*(1), 4–12.

Kim, D. (2005). KKCLASS [Computer program]. Unpublished.

Kim, D., Barton, K., & Kim, J. (April 2007). *Estimating classification consistency and classification accuracy with pattern scoring*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kim, D., Choi, S., Um, K., & Kim, J. (April 2006). *A comparison of methods for estimating classification consistency*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag.

Kolen, M., & Kim, D. (2004). [Personal correspondence].

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Linn, R. L. (Ed.) (1989). *Educational measurement* (3rd ed.). New York, NY: Macmillan.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*(2), 179–197.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719–748.

McCallin, R. C. (2006). Test administration. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 625–652). Mahwah, NJ: Lawrence Erlbaum Associates.

Michaelides, M. P., & Haertel, E. H. (2004). *Sampling of common items: An unrecognized source of error in test equating*. Los Angeles, CA: Center for the Study of Evaluation.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*(1), 59–71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

Muraki, E., & Bock, R. D. (1991). PARSCALE: Parameter scaling of rating data [Computer program]. Chicago, IL: Scientific Software, Inc.

National Research Council (2012). *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 17*(1), pp. 41–55.

Shu, L. (2020). PARDUX, Version 1.69 [Computer program]. Data Recognition Corporation.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*(2), 201–210.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion referenced tests: A decision theoretic formulation. *Journal of Educational Measurement, 11*(4), 263–267.

Swineford, F. (1956). *Technical manual for users of test analysis* (Statistical Report No. 56 42). Princeton, NJ: Educational Testing Service.

Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. Psychometrika, 47(2), 175–186.

Thissen, D. (1990). MULTILOG: Multiple categorical item analysis and test scoring (Version 6) [Computer program]. Chicago, IL: Scientific Software, Inc.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessment* (Synthesis Report No. 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Tong, Y., Wu, S., & Xu, M. (2008). *A comparison of pre-equating and post-equating using large-scale assessment data*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

van der Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory*. New York, NY: Springer.

Wright, B. D., & Linacre, J. M. (1992). BIGSTEPS Rasch analysis [Computer program]. Chicago, IL: MESA Press.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*(2), 245–262.

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement, 21*(2), 93–111.

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

Yen, W. M., & Burket, G. R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement, 34*(4), 293–313.

Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4*(3), 209–228.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*(3), 233–251.

# Meeting Overview

- Brief overview of the Forward Exam
- Item review process and training
- Break into workgroups
- Review items for placement on exam

# Roles & Responsibilities

**Participants**

- Item Review

**DRC Facilitators**

- Lead the group through the agenda

- Encourage interaction

- Lead discussions

- Collect secure materials

**DPI and DRC**

- Answer questions

# Wisconsin's Definition of College and Career Readiness

# College and Career Readiness Vision

- Wisconsin's Guiding Principles for Teaching and Learning inform the design and implementation of all academic standards.

  https://dpi.wi.gov/standards/guiding-principles

- Wisconsin's Academic Standards specify what students should know and be able to do in the classroom.

  https://dpi.wi.gov/standards

# Wisconsin Forward Exam

- Provides a measure of whether students are proficient in the skills and abilities identified in the Wisconsin Academic Standards
- All exam items are aligned to the standards:
  - English Language Arts and Mathematics tested in grades 3-8
  - Science tested in grades 4 and 8
  - Social Studies tested in grades 4, 8, and 10

# Purpose of Meeting

- Provide overview of the Wisconsin Forward Exam

- Unpack the Social Studies standards

- Provide item review training

- Review items for potential placement on Wisconsin Forward Exam

# Critical Importance of Security and Confidentiality

- All item review participants complete a security/nondisclosure agreement

- Security of passage and item content

- Note-taking policy

- Cellphone and personal computer use - phones not allowed to be out on tables during review

- Communication following the meeting

# Understanding the Standards

- 6 Strands (Inquiry Practices and Processes, Behavioral Sciences, Economics, Geography, History, Political Science)

- 27 Standards across those strands, K-12

- Each standard has a K-12 progression

- Grade Bands (K-2, 3-5, 6-8, 9-12)

# New Language and Terms

- <u>Standard</u>: Broad statement that tells what students are expected to know or be able to do

- <u>Learning Priority</u>: Breaks down the broad statement into manageable learning pieces

- <u>Performance Indicator</u>: Measurable degree to which a standard has been developed and/or met

**Discipline: Social Studies**

**Content Area: Economics (Econ)**

**Standard SS.Econ1: Wisconsin students use economic reasoning to understand issues.**

| Learning Priority | Performance Indicators (By Grade Band) | | | |
|---|---|---|---|---|
| | **K-2 (e)** | **3-5 (i)** | **6-8 (m)** | **9-12 (h)** |
| **Econ1.a: Choices and Decision -Making** | SS.Econ1.a.1 Differentiate between a "want" and a "need". Describe resources that are important or useful to you, your family, community, and country. | SS.Econ1.a.3 Use economic reasoning to compare and contrast the costs and benefits of a decision. Categorize different limited resources (e.g., money, materials, time, labor/workers, land, natural resources, renewable or non-renewable). | SS.Econ1.a.m Predict the opportunity costs of various decisions, and explain why the opportunity cost might differ from person to person or in different situations. Assess how limited resources (e.g., money, land, natural resources, workers, time) impact the choices of individuals, households, communities, businesses, and countries. | SS.Econ1.a.h Perform a cost-benefit analysis on a real-world situation, using economic thinking to describe the marginal costs and benefits of a particular decision. |
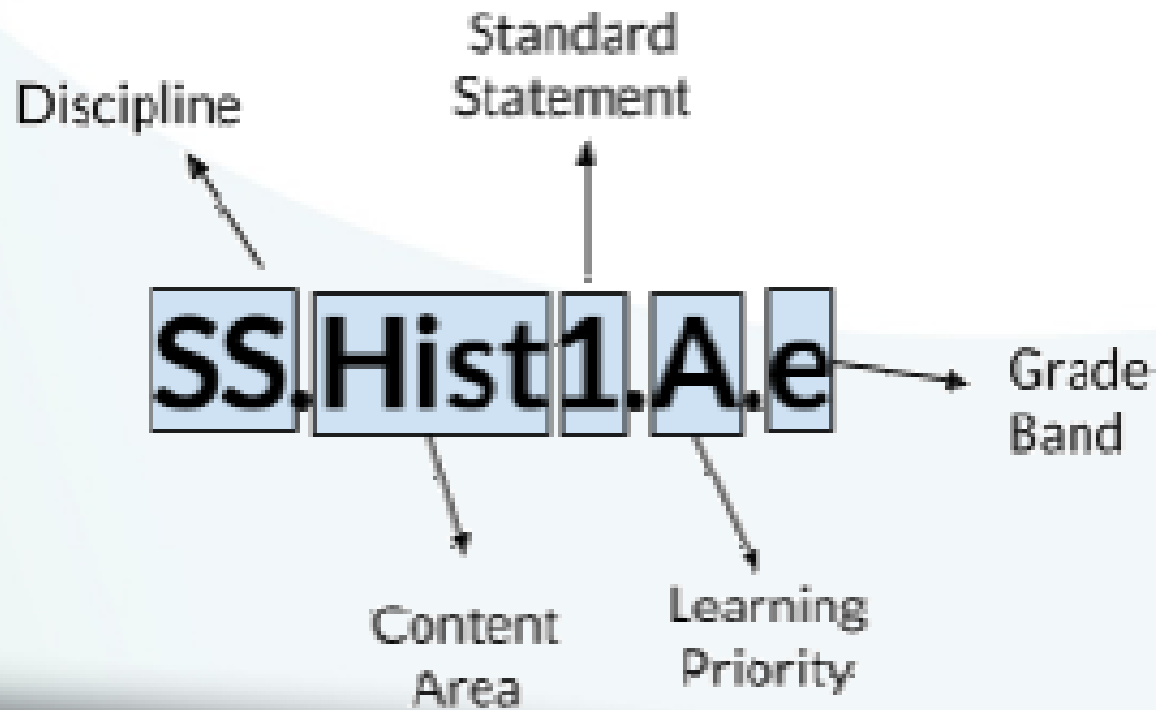
# What about this?

| BH1.a: Individual cognition, perception, and behavior | SS.BH1.a.2 Understand we are individuals influenced by our relationships and environments. | SS.BH1.a.4 Describe how a person's understanding, perceptions, and behaviors are affected by relationships and environments. |
|---|---|---|

- Guidance and assistance to districts

- Consistency across the state

- Support state assessment preparation

- Support students who transfer between districts

- Not required, recommended

- Will be finalized with the K-5 Suggested Scope & Sequence Appendix

# Wisconsin Standards for Social Studies

## At-A-Glance: Wisconsin Standards for Social Studies

### Wisconsin students will…

#### Social Studies Inquiry Practices and Processes

1. Construct meaningful questions that initiate an inquiry.
2. Gather and evaluate sources.
3. Develop claims using evidence to support reasoning.
4. Communicate and critique conclusions.
5. Be civically engaged.

#### Behavioral Science

1. Examine individual cognition, perception, behavior, and identity (Psychology).
2. Investigate interactions between individuals and groups (Sociology).
3. Assess the role that human behavior and cultures play in the development of social endeavors (Anthropology).
4. Examine the progression of specific forms of technology and their influence within various societies.

#### Economics

1. Use economic reasoning to understand issues.
2. Analyze how decisions are made and interactions occur among individuals, households, and firms/businesses (Microeconomics).
3. Analyze how an economy functions as a whole (Macroeconomics).
4. Evaluate government decisions and their impact on individuals, businesses, markets, and resources (Role of Government).
5. Assess economic interdependence of regions and countries through trade.

#### Geography

1. Use geographic tools and ways of thinking to analyze the world.
2. Analyze human movement and population patterns.
3. Examine the impacts of global interconnections and relationships.
4. Evaluate the relationship between identity and place.
5. Evaluate the relationship between humans and the environment.

#### History

1. Use historical evidence for determining cause and effect.
2. Analyze, recognize, and evaluate patterns of continuity and change over time and contextualization of historical events.
3. Connect past events, people, and ideas to the present, use different perspectives to draw conclusions, and suggest current implications.
4. Evaluate a variety of primary and secondary sources to interpret the historical context, intended audience, purpose, and/or author's point of view (Historical Methodology).

#### Political Science

1. Identify and analyze democratic principles and ideals.
2. Examine and interpret rights, privileges, and responsibilities in society.
3. Analyze and evaluate the powers and purposes of political and civic institutions.
4. Develop and employ skills for civic literacy.

DATA RECOGNITION
DRC
CORPORATION

WISCONSIN
DEPARTMENT OF
PUBLIC
INSTRUCTION

# 5 Main Shifts

- Inquiry as an Instructional Framework
- Integration Across All Strands
- Dynamic Balance of Skills and Content
- K-12 is a Progression
- Civic Engagement and Authentic Audiences

# Unpacking & Repacking Standards & Indicators

There is a training presentation available at:

http://bit.ly/UnpackWISSS

# Forward Exam Item Types

- Selected Response
    - Multiple Choice (MC)
    - Enhanced Selected Response (ESR)
- Technology Enhanced (TE)

# Multiple Choice (MC)

- All MC items have 4 answer choices
  - 3 distractors and 1 correct answer
- Used in all content areas
- Can be linked to a passage or stimuli or used as a "stand-alone MC"
- May have graphs, tables, or other information to support the stem

# MC Sample

An increase in smartphone sales would <u>most</u> likely result in

(a) lower salaries for workers at smartphone companies

(b) fewer resources being used by the smartphone industry

(c) higher earnings for stockholders in smartphone companies

(d) higher unemployment rates for workers in the smartphone industry

# Enhanced Selected Response (ESR)

- Varying combinations of multiple choice, multiple response, completion or short answer

- Explores authentic problem-solving skills

- Multiple correct answers

# ESR Sample

Cooperation between individuals and groups is important. Select <u>two</u> examples of cooperation.

(a) Alexis and Ryan take turns using the computer.

(b) Ethan and Christopher sit next to each other at lunchtime.

(c) Ms. Chen's students work as a group to complete an assignment.

(d) Jack and Natalie argue over who ran faster across the playground.

(e) Mr. Pierce allows his students to read quietly after finishing a test.

(f) Mr. Smith allows his students to choose where they sit in the classroom.

# Technology Enhanced (TE)

- TE items present in all content areas

- Interactive

- Wide Variety: drop down list, matching, graphing, highlighting text, drag and drop

# TE Sample Item

It is important that the powers of the three branches of government are balanced. The federal government ensures the balance of power by giving each branch separate powers. Drag each item listed into the box for the branch that has that power.

| Executive Branch | Legislative Branch | Judicial Branch |
|---|---|---|
|  |  |  |

enforces laws

interprets laws

can override presidential vetoes

determines whether laws are constitutional

can make treaties with other countries

votes on whether bills should be turned into laws

# Life Cycle of an Item
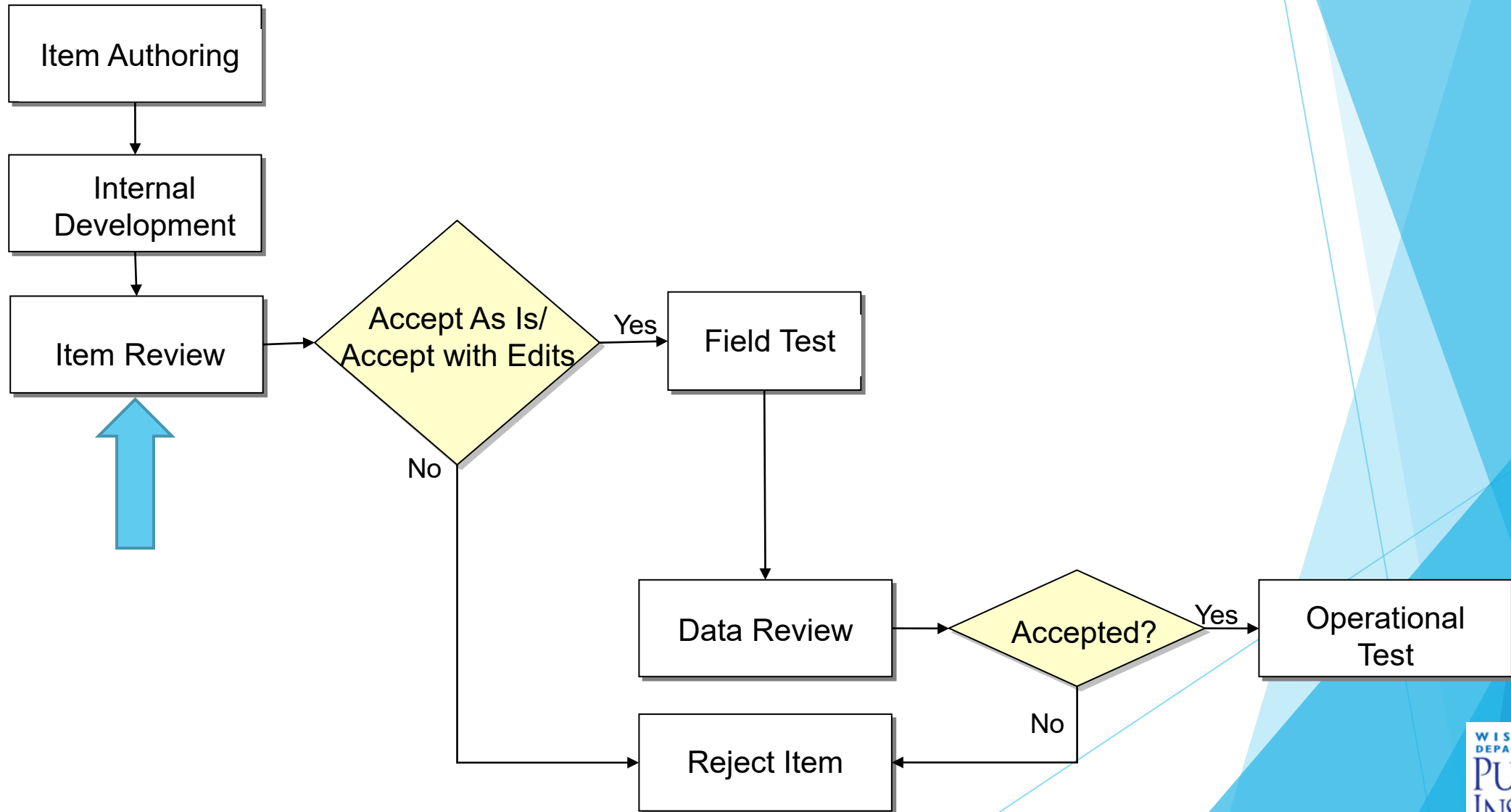
# Item Review Process

- Participants will view items online using INSIGHT—the same testing engine students use

- Allows interaction with item functionality, particularly useful for technology-enhanced items

- Facilitator will provide specific directions for logging in to begin reviews

# Item Review Process – (cont.)

Reviews will be completed in groups and individually. Items will be reviewed for:

- Standard alignment

- Grade-level appropriateness

- Correct answer key(s)

- Depth of Knowledge (DOK) level

- Bias and sensitivity concerns

- Is the wording and technical requirements of the item clear and easy to understand?

# Item Review Tally Sheet

# Evaluating an Item: Grade 8 Social Studies

# Step 1: Standard Alignment

After reading item ask yourself:

**Does the standard listed match the previous state standard? What NEW standard does it align to?**

- Each member will have copy of standards
- Match item to NEW appropriate standard as noted on item rating sheet

# Step 2: Check the Answer(s)

Is the answer (or answers) listed correct?

- If yes, move on to step 3
- If no, discuss with committee and note new answer(s)

# Step 3: Confirm the Depth of Knowledge Level

Is the DOK level listed correct?

- If yes, move on to step 4.
- If no, mark your thinking and discuss with committee.

We will go into detail about DOK a little later in this presentation.

# Step 4: Check for Bias and Sensitivity

- Stereotyping
- Gender
- Regional or geographical
- Ethnic or cultural
- Socioeconomic class
- Persons with a disability
- Ageism
- Religious

# Steps 5 and 6: Mark Comments

In spreadsheet, mark column noting the following:

- Accept - "A"
  - Item is OK as is
- Accept with Revisions - "AR"
  - Accept but apply recommended edits
- Dissenting View - "DV"
  - If you disagree with the committee
- Reject -
  - Item contains major flaws; do not recommend placement on assessment; note this in the comments column.
- Additional comments as needed

# What if I Disagree with the Committee?

- Speak up! It's possible that another committee member has the same concern, or you may have noticed something that other committee members have not.

- Record your dissenting view on the item review tracking sheet. Discussion by all is encouraged; however, if you choose not to share your opinion, your facilitator can voice your concern for you.

- DRC and DPI will reconcile any major disagreements/concerns noted on tracking sheet following the meeting. **A consensus is not always needed.**

# Step 7: Indicate Item Preference

- Rank item on a scale of 1-5 (with 5 being highest), your preference for having this item appear on the Wisconsin Forward Exam.

NOTE: This ranking will be used internally and not necessarily discussed as a committee for consensus.

# Things to Keep in Mind...

- Items need to measure what students should know and be able to do at their grade level, based on the academic standards. This may be different than what your personal experience is with students.

- Questions to ask during review:

  - Does the item provide for an optimal standard assessment of all students?

  - Are there items written to ALL ability levels? It is OK to have easy items.

# Things to Keep in Mind…Technical Design

- Items should not be confusing or tricky

- Does the item meet requirements for technical quality?

- Do graphics/visuals complement and support item?

- Does the stem provide a complete, clear and concise question/problem and directions?

- Does the stem not clue the correct answer(s)?

- Are correct answer(s) clear and accurate?

  - Distractors (or incorrect options) may contain common misperceptions or processes

# Things to Keep in Mind…
# Principles of Universal Design

- Items should respect the diversity of the assessment population.
  - Every student must be able to access the information.
  - Items must measure what is intended.
- Items should have:
  - A clear format for text
  - Clear pictures and graphics
  - Concise and readable text

# Everything in Moderation
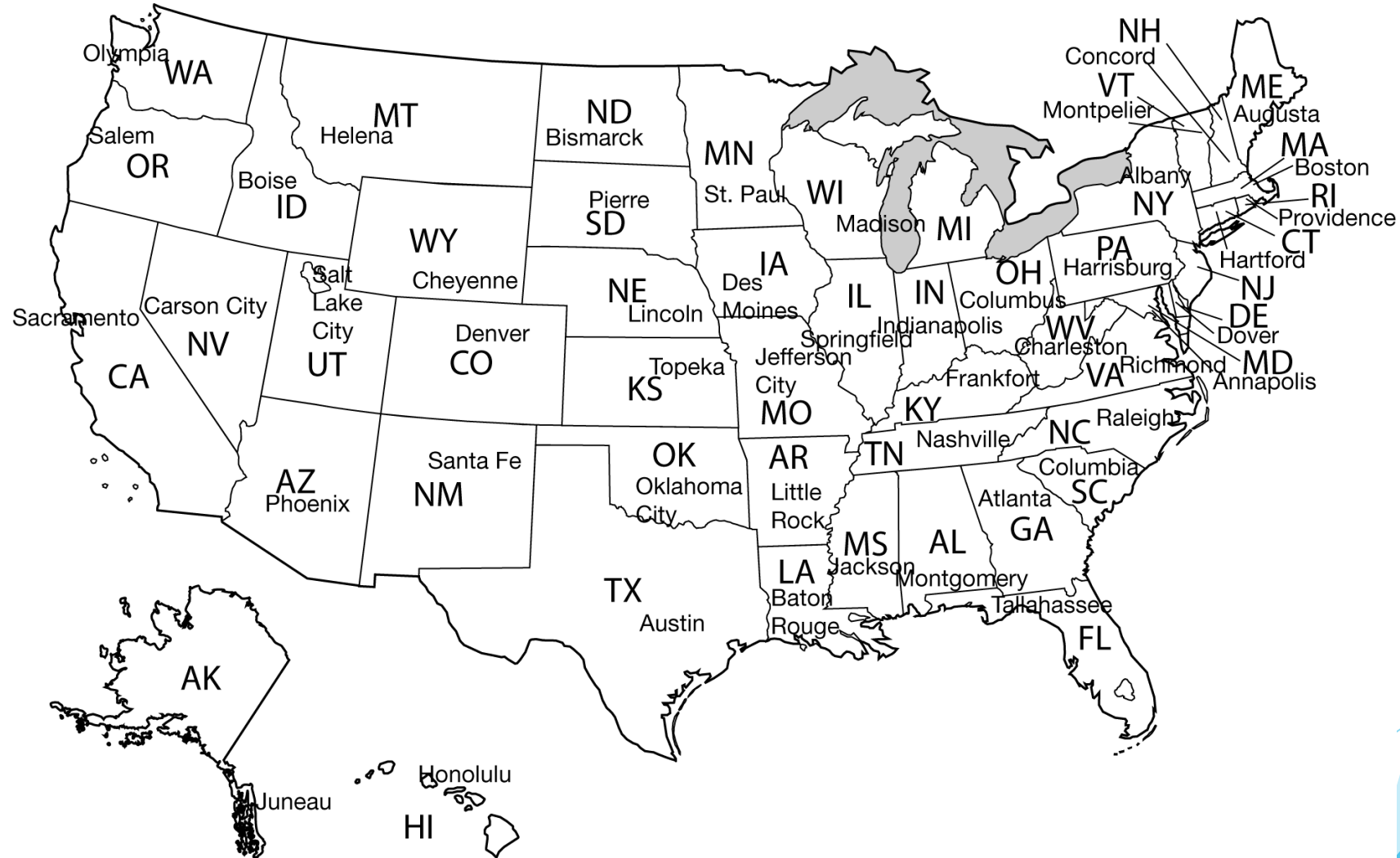
# When to Edit an Item

Reasons to edit an item include, but are not limited to the following:

- If the subject matter is above grade level or out of scope for the standard.

- If assigned DOK is not appropriate.

- If there is an opportunity to make the item/passage/stimulus easier for students to understand.

- If the topic or language is inappropriate, controversial, or inflammatory.

# Definition of DOK

**The degree or complexity of knowledge that the content curriculum standards and expectations require.**

- Includes four levels, from lowest (basic recall) to highest (extended thinking)

- Focuses on how well the students need to know the content before they can respond to a given item

- Used by item writers to gauge the *cognitive level* of item, **does not** correlate to the *difficulty* of the item

# DOK Levels

**DOK 1**   Recall and Reproduction

**DOK 2**   Skills and Concepts

**DOK 3**   Strategic Thinking and Reasoning

**DOK 4**   Extended Thinking

(rarely on standardized assessments — more "project-like" or on performance assessments)

# DOK 1: Recall and Reproduction

- Students demonstrate a rote response, use a well-known formula, or follow a simple procedure.

- A "simple" procedure is well defined and typically involves only **one** step.

*Key words: identify, recall, recognize facts, use, measure, solve a one-step problem*

# DOK 2: Skills and Concepts

- Students make some decisions regarding how to approach the question or problem.

- Requires deeper knowledge than just giving a definition, such as explaining _how_ or _why_.

- It _may_ involve two or more steps, however two steps does not automatically make a DOK 2.

_Key words: explain, categorize, use context clues, select a procedure, compare/contrast_

# DOK 2 - (cont.)

Activities may include:

- Making observations/collecting information
- Classifying/comparing information
- Organizing/displaying data or information in tables and graphs

*Note: Some action verbs, such as "explain," "describe," or "interpret," could be classified at different DOK levels, depending on the complexity of the action.*

# DOK 3: Strategic Thinking and Reasoning

- Students demonstrate deep understanding through planning, using evidence, and exhibiting higher levels of cognitive reasoning.

*Key words: connect ideas, explain thinking, cite evidence, analyze, apply a concept*

# DOK 3 - (cont.)

Activities may include the following:

- Use concepts to solve non-routine problems

- Describe how word choice, point of view or bias, may help the readers' interpretation of text

- Apply a concept in a new context

- Cite evidence and develop a logical argument for concepts

- Compare information within or across data sets

# Cognitive Level vs. Difficulty

DOK is used by item writers to gauge the *cognitive level* of item, it **does not** correlate to the *difficulty* of the item.

# Sample of Difficult DOK 1 Item

Which of these events occurred in the twentieth century?

(a) Germans began migrating to Wisconsin.

(b) Madison was chosen to be the capital of Wisconsin.

(c) Wisconsin senator Joseph McCarthy became a controversial national figure.

(d) Father Jacques Marquette and Louis Joliet traced the Wisconsin and Mississippi Rivers.

# Sample of an Easy DOK 3 Item

Read the news headline below.

**Daily Herald**

Red Cross Opens Emergency Shelter

Why did the event in the news headline take place?

(a) to raise money

(b) to teach fire safety

(c) to help people in a disaster

(d) to share different customs

DOK used by item writers to gauge the *cognitive level* of item, **does not** correlate to the *difficulty* of the item.

# Item Review Process: Summary

- Standard Alignment

- Key(s)

- DOK Levels

- Grade-level Appropriateness

- Bias and Sensitivity

| Session # | Seq # | Item ID | Passage Title | Standard Code | Item Type | Points | Key(s) | Proposed DOK | New Standard | Bias/Sensitivity Comments | Accept (A) Accept with Revisions (AR) Dissenting View (DV) | Comments | Item Preference |
|-----------|-------|---------|---------------|---------------|-----------|--------|--------|--------------|-------------|---------------------------|----------------------------------------------------------|----------|-----------------|
| | | | | | | | | | | | | | |

# Roles & Responsibilities: Summary

- Invest yourself in the process

- Share your opinions

- Listen to your colleagues

# Questions?

# Wisconsin Forward Exam
# Item Data Review
# Social Studies

August 10, 2021

# Purpose

- Establish a robust pool of items for use in new Social Studies test development to ensure proper representation:
  - New content standards
  - Test design
- General statistical guidelines are presented
  - Item flags are not created equal
  - Items with statistical flags are not necessarily poor items
  - Item content needs to be considered as well
  - Approving an item does not guarantee its appearance on a future test, but rather maximizes the size of the pool for item selection during test development.

# Key Objectives

- Review and understand item card layout

- Understand and interpret item statistics

- Review item cards for a few Social Studies field test items with different statistics

- Apply knowledge of item statistics to evaluate the remaining field test items

# Special Considerations

- New Wisconsin Standards for Social Studies
  - Field test items on average more difficult than operational items
- Impact of the pandemic on item statistics
  - Disrupted learning
  - Participation rates
    - 85% in grade 4; 83% in grade 8; and 75% in grade 10
  - Representativeness
    - Overrepresented groups: White students; Rural/Town/Suburban districts
    - Underrepresented groups: Black students; City districts; students with disabilities, LEP, SES-disadvantaged
  - Year-to-year performance change (operational tests)
- Field test statistics are reliable for the purpose of form selection

# Sample Item Card

Standard: Analyze how a person's local actions can have global consequences, and how global patterns and processes can affect seemingly unrelated local actions.

**Assessed standard**

1. **Read the information in the box.**

In 1988, a biologist secretly made a video that showed large numbers of dolphins dying in nets that fishermen used to catch tuna. The release of the video resulted in a consumer boycott of tuna, led in part by school children who urged their parents to not buy tuna. Two years later, one of the largest tuna companies announced it would no longer buy tuna caught by methods that threatened dolphins. Other major producers of canned tuna followed suit. Tuna fishermen changed their fishing methods so that dolphins would no longer accidentally get caught in their nets.

**Stem**

Which conclusion is best supported by the information in the box?

A. Fishing nets are dangerous and should be outlawed.
B. Individuals acting locally can make an impact far away.
C. Large companies do not care about fish or mammals.
D. Evidence made secretly cannot be used in a court of law.

**Answer choices**

## WI - Data Card

| Field | Value |
|---|---|
| Item ID | 1035047 |
| Content Area | Social Studies |
| Course | |
| Passage ID | |
| Passage Title | |
| Grade | 8 |
| Standards | WMAS: BH.3.a.r |
| Item Type | Multiple Choice |
| Points | 1 |
| Depth of Knowledge | 3 |
| Est Difficulty | |
| Key | B |

**Item ID**

**Content Area**

**Grade**

**Standard**

**Item Type**

**Key(s)**

# Sample Item Card (cont.)

**Administration(s)**

| Form Name | Use Function | Seq | Period | Year | Session | Calc | Model/Ext | Grade |
|---|---|---|---|---|---|---|---|---|
| FT1 | FT | 13 | Spring | 2021 | 1 | No | 3PL/3PL | 8 |

Administration information

**Traditional Statistics**

| N | P-Val | Mean | Item Total Corr |
|---|---|---|---|
| 10723 | 0.52 | | 0.42 |

Classical statistics

**Fit Statistics**

| Outfit t | Infit t | Outfit MnSq | Infit MnSq | Chi-sq | Deg Free | Item Fit | Fit |
|---|---|---|---|---|---|---|---|
| | | | | | | 1.32 | |

Item Fit

**IRT Statistics**

| Label | Final | Final S.E. | Preliminary | Preliminary S.E. | Displ |
|---|---|---|---|---|---|
| Slope | 0.98 | | | | |
| Location | -0.19 | | | | |
| Asymptote | 0.25 | | | | |

IRT statistics

**Distractor/Step Specific**

| Part | Label | Freq | Proportion | Corr | Avg Meas | Step Meas |
|---|---|---|---|---|---|---|
| | A | 3069 | 0.29 | -0.14 | | |
| | B | 5618 | 0.52 | 0.42 | | |
| | C | 1371 | 0.13 | -0.27 | | |
| | D | 665 | 0.06 | -0.23 | | |
| | OMITS | 28 | 0.00 | | | |

Distractor analysis

**DIF Analysis**

| Category | Bias Code | Num Value | N - Ref | N - Focal |
|---|---|---|---|---|
| ACC | | | 3802 | 18 |
| DISAB | A- | -0.34 | 6300 | 4380 |
| ECODISAD | A | -0.01 | 5830 | 4919 |
| LEP | A | -0.16 | 9056 | 1417 |
| MALEFEMALE | A | 0.35 | 6181 | 4568 |
| WHITEAMIN | | | 6238 | 179 |
| WHITEASIAN | A | 0.45 | 6785 | 444 |
| WHITEBLACK | A | 0.18 | 6746 | 998 |
| WHITEHISPANIC | A | -0.13 | 6799 | 1899 |
| WHITEMULTI | A | 0.38 | 6768 | 431 |

DIF Index

# Classical Statistics: Item Difficulty

**Difficulty**

- **"P-Value"** : proportion of students who answered an item correctly (or a percent of maximum points possible for polytomously scored items)
  - 0.0 means all students answered incorrectly
  - 1.0 means all students answered correctly
  - The higher the p-value, the easier the item

## Traditional Statistics

| N | P-Val | Mean | Item Total Corr |
|------|-------|------|------------------|
| 4349 | 0.73  |      | 0.49             |

# Item Difficulty: Considerations

**Targeted Range**

- P-Value: 0.20 to 0.90
- Items outside of target range may be approved if content is appropriate

**Content Consideration**

- We need to build tests with a wide range of p-values in order to effectively place students into the four performance categories
    - Hard items to distinguish between Proficient/Advanced
    - Easy items to distinguish between Below Basic/Basic
- Why did most students answer this item correctly or incorrectly?
- Are there any reasons other than item difficulty to support a decision to ACCEPT or REJECT this item?

# Classical Statistics: Item Discrimination

**Discrimination**

- Measures item's ability to differentiate between high and low performers
- Item-Total Test Correlation (or point biserial for dichotomously scored items) is the correlation of the examinees' raw scores on a single item with their raw scores on all remaining test items (-1.0 to +1.0)
  - Positive—high achievers outperformed low achievers (targeted).
  - Negative—low achievers outperformed high achievers (unexpected).
  - Around zero—high and low achievers performed about the same on an item (not desired).

## Traditional Statistics

| N | P-Val | Mean | Item Total Corr |
|---|-------|------|-----------------|
| 4349 | 0.73 | | 0.49 |

9

# Item Discrimination: Considerations

**Targeted Range**

- at or above 0.15
  - Smaller sometimes is okay, depending on difficulty
  - Items with negative or around 0.0 item discrimination are poorly discriminating and often should be rejected

**Content Consideration**

- Why is this item less able to differentiate between high and low achievers?
- Is the low discrimination associated with extreme low or high P-Values (item difficulty)?
- Are there any other reasons other than item discrimination to support your decision on ACCEPTING or REJECTING this item?

10

# Classical Stats on Item Card (MC Items)

- (*) indicates key
- **P-value**
  - Proportion of students who got the item right
- **Item-Total Correlation (point biserial)**
  - Discrimination power
- **Proportion**
  - Proportion of students selecting different options

## Traditional Statistics

| N | P-Val | Mean | Item Total Corr |
|---|---|---|---|
| 14016 | 0.78 | | 0.48 |

## Distractor/Step Specific

| Label | Proportion | Corr | Avg Meas | Step Meas |
|---|---|---|---|---|
| A* | 0.78 | 0.48 | | |
| B | 0.06 | -0.29 | | |
| C | 0.11 | -0.25 | | |
| D | 0.04 | -0.23 | | |
| MULTS | 0.00 | | | |
| OMITS | 0.00 | | | |

# Distractor Specifc Analysis (MC Items)

## Distractor/Step Specific

| Label | Proportion | Corr | Avg Meas | Threshold |
|-------|-----------|-------|----------|-----------|
| A | 0.05 | -0.22 | | |
| B | 0.10 | -0.26 | | |
| C | 0.12 | -0.28 | | |
| D* | 0.73 | 0.49 | | |
| MULTS | 0.00 | | | |
| OMITS | 0.00 | | | |

**Guideline**

•MC items:
  ➢Correlations for the distractors should be negative.
  ➢Correlations for the distractors should never be higher than correlation for the correct answer
  ➢Proportion of distractor < proportion of key

**Content Consideraton**

•Is the correlation of selecting any incorrect option greater than 0? If yes, why does this option distract more high achievers than low achievers?

•Is the proportion of selecting any incorrect option greater than the proportion of selecting the key? If yes, why?

# Score Point-Specific Analysis

## Distractor/Step Specific

| Part | Label | Freq | Proportion | Corr |
|------|-------|------|-----------|------|
|  | 0 | 408 | 0.16 | -0.37 |
|  | 1 | 2107 | 0.84 | 0.37 |
|  | BL | 3 | 0.00 |  |

**Guideline**

- Non-MC items:
  - All items are worth 1 point
  - Correlations for the score 0 expected to be negative
  - Correlation for the score 1 should be positive

**Content Consideration**

Non-MC items

- Is the proportion of students at each a score point reasonable?
- Is the pattern of item score correlation as expected?

13

# Option Analysis for MS items

## Distractor/Step Specific

| Part | Label | Freq | Proportion | Corr | Avg Meas | Step Meas |
|------|-------|------|-----------|------|----------|-----------|
| A | A | 1767 | 0.45 | -0.07 | | |
| A | B | 398 | 0.10 | -0.29 | | |
| A | C | 1790 | 0.45 | 0.00 | | |
| A | D | 2195 | 0.56 | 0.28 | | |
| A | E | 1183 | 0.30 | 0.20 | | |
| | 0 | 3356 | 0.85 | -0.27 | | |
| | 1 | 580 | 0.15 | 0.28 | | |
| | BL | 10 | 0.00 | | | |

### Guidelines

- Correlations for correct options should be positive; for incorrect options – negative
- Proportions of students at correct options expected to be higher than for incorrect options
- Is the pattern of option proportions and correlations as expected?

# IRT Statistics: Item Fit and Non-Convergence

**IRT Statistics**

**Item Fit**

- IRT statistic obtained after item calibration

- Measures how well the student responses to each item fit the test data (by comparing parameter estimation prediction relative to the observed data)

- Item is flagged when the observed data pattern differs from the predicted probability of responding to the item.

- There is no specific criterion value for the fit flag: criterion is dependent on the number of students taking the item

- Typically, not a serious flag by itself.

**Item Non-Convergence**

- Item parameters cannot be estimated, and the item is not eligible for future use (two items in grade 4 and one item in grade 10)

# IRT Statistics on Item Cards

Fit Statistics

| Outfit t | Infit t | Outfit MnSq | Infit MnSq | Chi-sq | Deg Free | Item Fit | Fit |
|----------|---------|-------------|------------|--------|----------|----------|-----|
|          |         |             |            |        |          | 11.58    | MISFIT |

| Outfit t | Infit t | Outfit MnSq | Infit MnSq | Chi-sq | Deg Free | Item Fit | Fit |
|----------|---------|-------------|------------|--------|----------|----------|-----|
|          |         |             |            |        |          | 2.58     |     |

Non-Convergent Items (no Item Fit or IRT Stats)

| Outfit t | Infit t | Outfit MnSq | Infit MnSq | Chi-sq | Deg Free | Item Fit | Fit |
|----------|---------|-------------|------------|--------|----------|----------|-----|
|          |         |             |            |        |          |          |     |

# Differential Item Functioning

**DIF**

- Procedure used to identify items that function differently for particular groups of students (e.g., gender, ethnicity, and disability status, SES status, and LEP status).
- Hypothesis is that test takers with similar knowledge or ability should perform in similar ways on a test item.
- Items are flagged if they do not behave the same in different groups of students, after controlling for student ability.

**Procedure**

- Compares "focal" vs. "reference" groups.
- Reference groups: Males, Whites, students w/out disabilites, students not SES-disadvantaged, English proficient students, students not using accommodations.
- Focal groups: Females, non-White ethnic groups, students with disabilities, SES-disadvantaged students, LEP students, and students using accommodations

# Visualizing DIF (Gender)
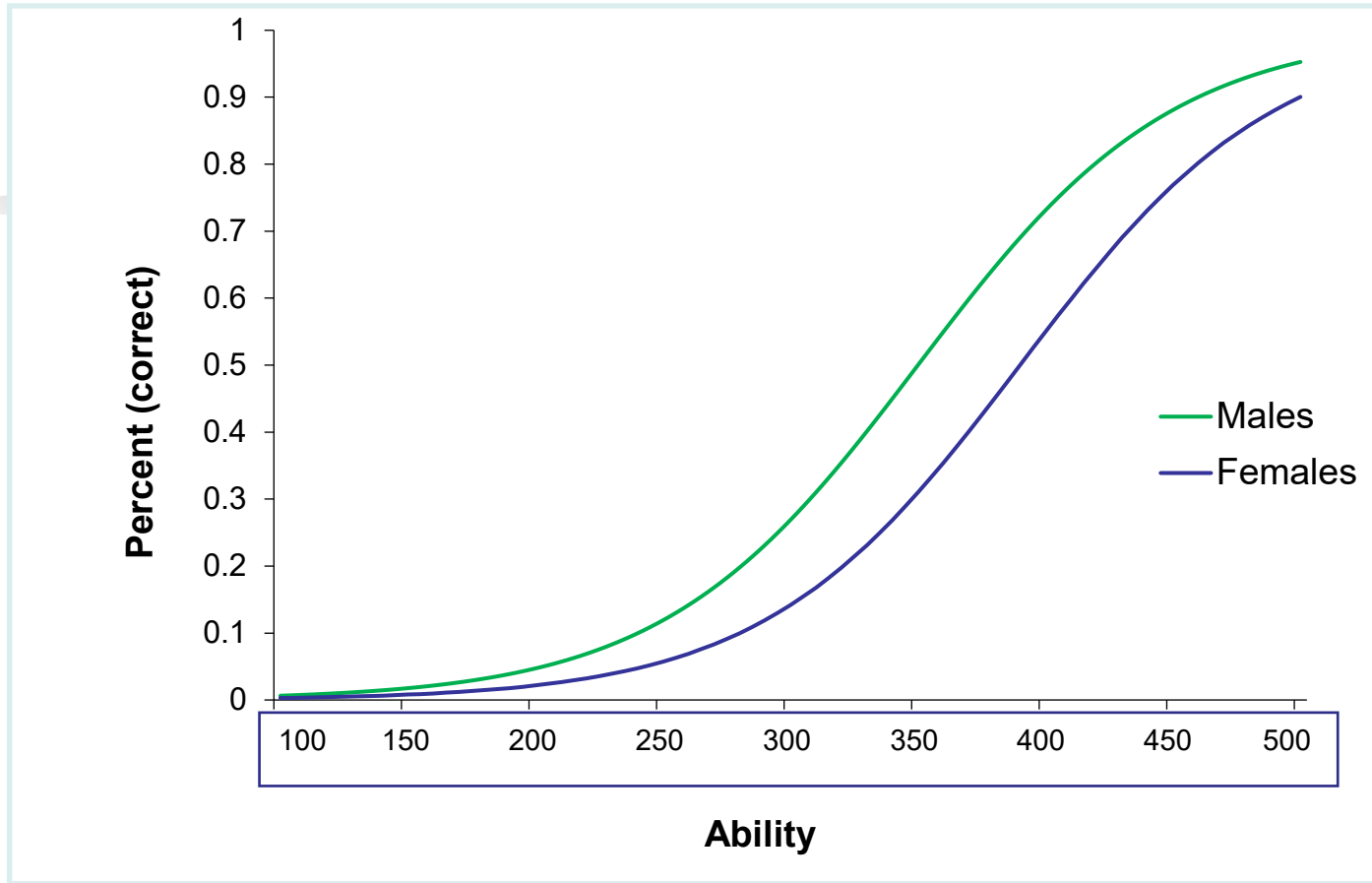
# Differential Item Functioning

**Guideline**

- Each item is assigned a bias code of A, B, or C.
  - A – minor DIF (no DIF)
  - B – moderate DIF
  - C – Large DIF

DIF signs: "-" favors Reference group; '+' favors Focal group.

- Only items with C (i.e., large) DIF require review. Items with C DIF may be acceptable if no potential bias causes the differential item functioning.

**Content Consideration**

- Is there anything in the content or format of the item that may interfere with, or advantage, one group of students over another based on:
  - Gender?
  - Ethnicity?
  - Disability status, SES status, LEP status, accommodation use?

# DIF Statistics and Codes on Item Cards

## DIF Analysis

| Category | Bias Code | Num Value | N - Ref | N - Focal |
|---|---|---|---|---|
| ACC | | | 0 | 0 |
| DISAB | | | 2270 | 136 |
| ECODISAD | A | -0.01 | 1524 | 991 |
| LEP | | | 2173 | 101 |
| MALEFEMALE | A | 0.02 | 1231 | 1284 |
| WHITEAMIN | | | 1092 | 35 |
| WHITEASIAN | | | 1701 | 105 |
| WHITEBLACK | C- | -0.12 | 1524 | 237 |
| WHITEHISPANIC | A | 0.00 | 1742 | 249 |
| WHITEMULTI | | | 1754 | 121 |
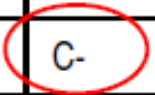
Reference Group/Focal Group

# DIF: Summary

- All biased items should show DIF, but **Not** all items with DIF will be biased.
  - The smaller sample sizes of the minority ethnicity groups causes many false positives.
  - DIF not computed if focal group N <200 or reference group N <400
  - You **must** be able to provide a reason for the bias to call the item biased.



ALL ITEMS WITH DIF

ALL BIASED ITEMS

# Summary of Item Flags

- P-value less than 0.20 or higher than 0.90
- Item-total test correlation < 0.15
    - Negative or close to 0 item-total test correlation is a very serious flag, especially when combined with a positive correlation for a distractor for MC items
- Positive pt. biserial correlation for a distractor
    - Especially if pt. biserial for a distractor is higher than pt. biserial for the correct option
- Poor Fit
- Non-Convergence (kills the item)
- Large DIF (C +/-)
- Omit rates > 3% (not used in this data review)

# Roles, Responsibilities, Questions

- DPI
  - Review Spring 2021 Social Studies field test item data
  - Accept or reject items

- DRC
  - Facilitate Data Review
  - Answer DPI questions

- Questions?

# Thank you!

# Appendix C
# Spring 2022 English Language Arts Operational Test Maps

# List of Tables

**Table C-1. English Language Arts, Grade 3 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 1 | TDA | OP | 4 | 3 | 3.W.2 | Writing |
| 3 | 2 | 2 | MC | OP | 1 | 2 | 3.W.8 | Writing |
| 3 | 2 | 3 | MC | OP | 1 | 1 | 3.L.1.b | Writing |
| 3 | 2 | 4 | TE | OP | 2 | 2 | 3.W.1.a | Writing |
| 3 | 2 | 5 | MC | OP | 1 | 1 | 3.W.1.c | Writing |
| 3 | 2 | 6 | TE | OP | 2 | 2 | 3.W.8 | Writing |
| 3 | 2 | 7 | MC | OP | 1 | 2 | 3.L.1.i | Writing |
| 3 | 2 | 8 | MC | OP | 1 | 3 | 3.W.8 | Writing |
| 3 | 2 | 9 | MC | OP | 1 | 1 | 3.L.1.i | Writing |
| 3 | 2 | 10 | MC | OP | 1 | 2 | 3.L.2.d | Writing |
| 3 | 2 | 11 | TE | OP | 2 | 2 | 3.W.8 | Writing |
| 3 | 2 | 12 | MC | OP | 1 | 1 | 3.L.1.d | Writing |
| 3 | 2 | 13 | MC | OP | 1 | 2 | 3.L.1.g | Writing |
| 3 | 2 | 14 | TE | OP | 1 | 2 | 3.W.2.c | Writing |
| 3 | 3 | 15 | EBSR | OP | 2 | 3 | 3.SL.2 | Listening |
| 3 | 3 | 16 | MC | OP | 1 | 1 | 3.SL.3 | Listening |
| 3 | 3 | 17 | MC | OP | 1 | 2 | 3.SL.2 | Listening |
| 3 | 3 | 18 | MC | OP | 1 | 2 | 3.SL.2 | Listening |
| 3 | 3 | 19 | MS | OP | 2 | 2 | 3.SL.3 | Listening |
| 3 | 4 | 20 | MC | OP | 1 | 3 | 3.RL.7 | Reading |
| 3 | 4 | 21 | MC | OP | 1 | 2 | 3.RL.4 | Reading |
| 3 | 4 | 22 | MC | OP | 1 | 3 | 3.RL.5 | Reading |
| 3 | 4 | 23 | EBSR | OP | 2 | 3 | 3.RL.2 | Reading |
| 3 | 4 | 24 | MC | OP | 1 | 3 | 3.RL.6 | Reading |
| 3 | 4 | 25 | TE | OP | 2 | 2 | 3.RL.2 | Reading |
| 3 | 4 | 26 | MC | OP | 1 | 2 | 3.RL.4 | Reading |
| 3 | 4 | 27 | MC | OP | 1 | 2 | 3.L.4.a | Reading |
| 3 | 4 | 28 | MC | OP | 1 | 2 | 3.RL.1 | Reading |
| 3 | 4 | 29 | EBSR | OP | 2 | 3 | 3.RL.2 | Reading |
| 3 | 4 | 30 | MC | OP | 1 | 2 | 3.RL.5 | Reading |
| 3 | 4 | 31 | MC | OP | 1 | 2 | 3.RL.6 | Reading |
| 3 | 4 | 32 | MC | OP | 1 | 2 | 3.L.4.RI | Reading |
| 3 | 4 | 33 | MC | OP | 1 | 2 | 3.RI.5 | Reading |
| 3 | 4 | 34 | MC | OP | 1 | 2 | 3.RI.1 | Reading |
| 3 | 4 | 35 | EBSR | OP | 2 | 2 | 3.RI.2 | Reading |
| 3 | 4 | 36 | MC | OP | 1 | 2 | 3.RI.1 | Reading |
| 3 | 4 | 37 | MC | OP | 1 | 2 | 3.RI.7 | Reading |

**Table C-2. English Language Arts, Grade 4 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 1 | TDA | OP | 4 | 3 | 4.W.2 | Writing |
| 4 | 2 | 2 | MC | OP | 1 | 2 | 4.W.1.c | Writing |
| 4 | 2 | 3 | TE | OP | 1 | 3 | 4.W.3.e | Writing |
| 4 | 2 | 4 | MC | OP | 1 | 1 | 4.L.3.a | Writing |
| 4 | 2 | 5 | TE | OP | 1 | 2 | 4.W.8 | Writing |
| 4 | 2 | 6 | TE | OP | 2 | 2 | 4.W.8 | Writing |
| 4 | 2 | 7 | MC | OP | 1 | 2 | 4.W.8 | Writing |
| 4 | 2 | 8 | MC | OP | 1 | 3 | 4.W.8 | Writing |
| 4 | 2 | 9 | MC | OP | 1 | 2 | 4.W.8 | Writing |
| 4 | 2 | 10 | MC | OP | 1 | 1 | 4.L.2.c | Writing |
| 4 | 2 | 11 | TE | OP | 2 | 2 | 4.L.1.b | Writing |
| 4 | 2 | 12 | MC | OP | 1 | 1 | 4.L.2.a | Writing |
| 4 | 2 | 13 | TE | OP | 2 | 1 | 4.W.2.c | Writing |
| 4 | 2 | 14 | MC | OP | 1 | 1 | 4.L.2.a | Writing |
| 4 | 3 | 15 | MC | OP | 1 | 2 | 4.SL.3 | Listening |
| 4 | 3 | 16 | MC | OP | 1 | 1 | 4.SL.2 | Listening |
| 4 | 3 | 17 | EBSR | OP | 2 | 2 | 4.SL.3 | Listening |
| 4 | 3 | 18 | MC | OP | 1 | 1 | 4.SL.3 | Listening |
| 4 | 3 | 19 | MC | OP | 1 | 2 | 4.SL.2 | Listening |
| 4 | 3 | 20 | EBSR | OP | 2 | 3 | 4.SL.3 | Listening |
| 4 | 4 | 21 | MC | OP | 1 | 2 | 4.L.4 | Reading |
| 4 | 4 | 22 | TE | OP | 2 | 2 | 4.RI.2 | Reading |
| 4 | 4 | 23 | MC | OP | 1 | 2 | 4.RI.7 | Reading |
| 4 | 4 | 24 | MC | OP | 1 | 2 | 4.L.4 | Reading |
| 4 | 4 | 25 | MS | OP | 2 | 2 | 4.RI.1 | Reading |
| 4 | 4 | 26 | MC | OP | 1 | 2 | 4.L.4.a | Reading |
| 4 | 4 | 27 | MC | OP | 1 | 1 | 4.L.5.c | Reading |
| 4 | 4 | 28 | EBSR | OP | 2 | 3 | 4.RL.2 | Reading |
| 4 | 4 | 29 | MC | OP | 1 | 2 | 4.RL.9 | Reading |
| 4 | 4 | 30 | MC | OP | 1 | 2 | 4.RL.6 | Reading |
| 4 | 4 | 31 | MC | OP | 1 | 2 | 4.RI.5 | Reading |
| 4 | 4 | 32 | EBSR | OP | 2 | 3 | 4.RI.8 | Reading |
| 4 | 4 | 33 | MC | OP | 1 | 2 | 4.L.4.RI | Reading |
| 4 | 4 | 34 | MC | OP | 1 | 2 | 4.RI.2 | Reading |
| 4 | 4 | 35 | MC | OP | 1 | 2 | 4.RL.2 | Reading |
| 4 | 4 | 36 | MC | OP | 1 | 2 | 4.RL.3 | Reading |
| 4 | 4 | 37 | MC | OP | 1 | 1 | 4.L.5.RL | Reading |
| 4 | 4 | 38 | TE | OP | 2 | 2 | 4.RL.1 | Reading |
| 4 | 4 | 39 | MC | OP | 1 | 2 | 4.RL.6 | Reading |

**Table C-3. English Language Arts, Grade 5 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 5 | 1 | 1 | TDA | OP | 4 | 3 | 5.W.9 | Writing |
| 5 | 2 | 2 | MC | OP | 1 | 1 | 5.L.2.b | Writing |
| 5 | 2 | 3 | MC | OP | 1 | 3 | 5.W.3.d | Writing |
| 5 | 2 | 4 | TE | OP | 2 | 2 | 5.W.8 | Writing |
| 5 | 2 | 5 | MC | OP | 1 | 1 | 5.W.2.d | Writing |
| 5 | 2 | 6 | MC | OP | 1 | 2 | 5.W.8 | Writing |
| 5 | 2 | 7 | TE | OP | 2 | 2 | 5.W.1.c | Writing |
| 5 | 2 | 8 | TE | OP | 1 | 2 | 5.W.8 | Writing |
| 5 | 2 | 9 | MC | OP | 1 | 2 | 5.L.2 | Writing |
| 5 | 2 | 10 | MC | OP | 1 | 1 | 5.L.2.c | Writing |
| 5 | 2 | 11 | TE | OP | 2 | 2 | 5.W.8 | Writing |
| 5 | 2 | 12 | MC | OP | 1 | 2 | 5.L.2.b | Writing |
| 5 | 2 | 13 | TE | OP | 2 | 1 | 5.L.2.b | Writing |
| 5 | 3 | 14 | MC | OP | 1 | 2 | 5.SL.2 | Listening |
| 5 | 3 | 15 | MC | OP | 1 | 2 | 5.SL.2 | Listening |
| 5 | 3 | 16 | EBSR | OP | 2 | 2 | 5.SL.3 | Listening |
| 5 | 3 | 17 | MC | OP | 1 | 2 | 5.SL.3 | Listening |
| 5 | 3 | 18 | MC | OP | 1 | 2 | 5.SL.2 | Listening |
| 5 | 3 | 19 | EBSR | OP | 2 | 3 | 5.SL.3 | Listening |
| 5 | 4 | 20 | MC | OP | 1 | 1 | 5.RL.1 | Reading |
| 5 | 4 | 21 | MC | OP | 1 | 3 | 5.RL.5 | Reading |
| 5 | 4 | 22 | EBSR | OP | 2 | 3 | 5.RL.2 | Reading |
| 5 | 4 | 23 | MC | OP | 1 | 2 | 5.RL.6 | Reading |
| 5 | 4 | 24 | MC | OP | 1 | 2 | 5.RI.1 | Reading |
| 5 | 4 | 25 | MC | OP | 1 | 2 | 5.L.5 | Reading |
| 5 | 4 | 26 | MC | OP | 1 | 2 | 5.RI.8 | Reading |
| 5 | 4 | 27 | MC | OP | 1 | 2 | 5.RI.2 | Reading |
| 5 | 4 | 28 | MC | OP | 1 | 2 | 5.L.5 | Reading |
| 5 | 4 | 29 | MS | OP | 2 | 2 | 5.RL.2 | Reading |
| 5 | 4 | 30 | MC | OP | 1 | 2 | 5.RL.6 | Reading |
| 5 | 4 | 31 | MC | OP | 1 | 2 | 5.RL.9 | Reading |
| 5 | 4 | 32 | MC | OP | 1 | 2 | 5.RI.5 | Reading |
| 5 | 4 | 33 | MC | OP | 1 | 2 | 5.RI.8 | Reading |
| 5 | 4 | 34 | MC | OP | 1 | 2 | 5.RI.3 | Reading |
| 5 | 4 | 35 | MC | OP | 1 | 2 | 5.RI.8 | Reading |
| 5 | 4 | 36 | MC | OP | 1 | 2 | 5.RI.2 | Reading |
| 5 | 4 | 37 | TE | OP | 2 | 2 | 5.L.4.RI | Reading |
| 5 | 4 | 38 | MC | OP | 1 | 2 | 5.RI.1 | Reading |
| 5 | 4 | 39 | MC | OP | 1 | 1 | 5.RI.3 | Reading |
| 5 | 4 | 40 | MC | OP | 1 | 2 | 5.RI.1 | Reading |

**Table C-4. English Language Arts, Grade 6 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 6 | 1 | 1 | TDA | OP | 4 | 3 | 6.W.9 | Writing |
| 6 | 2 | 2 | MC | OP | 1 | 2 | 6.W.2.a | Writing |
| 6 | 2 | 3 | MC | OP | 1 | 1 | 6.W.3.b | Writing |
| 6 | 2 | 4 | MC | OP | 1 | 2 | 6.W.1 | Writing |
| 6 | 2 | 5 | MC | OP | 1 | 2 | 6.W.8 | Writing |
| 6 | 2 | 6 | TE | OP | 2 | 2 | 6.W.8 | Writing |
| 6 | 2 | 7 | MC | OP | 1 | 1 | 6.L.1.a | Writing |
| 6 | 2 | 8 | MC | OP | 1 | 2 | 6.L.2.a | Writing |
| 6 | 2 | 9 | TE | OP | 2 | 2 | 6.L.1.d | Writing |
| 6 | 2 | 10 | TE | OP | 2 | 1 | 6.L.2.b | Writing |
| 6 | 2 | 11 | TE | OP | 2 | 2 | 6.W.8 | Writing |
| 6 | 2 | 12 | MC | OP | 1 | 2 | 6.W.8 | Writing |
| 6 | 2 | 13 | MC | OP | 1 | 1 | 6.L.2.a | Writing |
| 6 | 3 | 14 | MC | OP | 1 | 2 | 6.SL.3 | Listening |
| 6 | 3 | 15 | MC | OP | 1 | 2 | 6.SL.2 | Listening |
| 6 | 3 | 16 | TE | OP | 2 | 1 | 6.SL.2 | Listening |
| 6 | 3 | 17 | MC | OP | 1 | 2 | 6.SL.3 | Listening |
| 6 | 3 | 18 | MC | OP | 1 | 1 | 6.SL.2 | Listening |
| 6 | 3 | 19 | EBSR | OP | 2 | 3 | 6.SL.3 | Listening |
| 6 | 4 | 20 | MC | OP | 1 | 2 | 6.RL.1 | Reading |
| 6 | 4 | 21 | MC | OP | 1 | 2 | 6.RL.4 | Reading |
| 6 | 4 | 22 | MC | OP | 1 | 2 | 6.RL.5 | Reading |
| 6 | 4 | 23 | MS | OP | 2 | 2 | 6.RL.2 | Reading |
| 6 | 4 | 24 | MC | OP | 1 | 2 | 6.RI.8 | Reading |
| 6 | 4 | 25 | MC | OP | 1 | 2 | 6.L.4.RI | Reading |
| 6 | 4 | 26 | TE | OP | 2 | 2 | 6.RI.1 | Reading |
| 6 | 4 | 27 | MC | OP | 1 | 3 | 6.RI.7 | Reading |
| 6 | 4 | 28 | TE | OP | 2 | 2 | 6.RI.2 | Reading |
| 6 | 4 | 29 | MC | OP | 1 | 2 | 6.RI.1 | Reading |
| 6 | 4 | 30 | MC | OP | 1 | 2 | 6.RI.5 | Reading |
| 6 | 4 | 31 | EBSR | OP | 2 | 3 | 6.RI.6 | Reading |
| 6 | 4 | 32 | TE | OP | 2 | 2 | 6.RI.6 | Reading |
| 6 | 4 | 33 | MC | OP | 1 | 2 | 6.RI.4 | Reading |
| 6 | 4 | 34 | EBSR | OP | 2 | 3 | 6.RL.3 | Reading |
| 6 | 4 | 35 | MC | OP | 1 | 2 | 6.RL.4 | Reading |
| 6 | 4 | 36 | MC | OP | 1 | 2 | 6.RL.2 | Reading |
| 6 | 4 | 37 | MC | OP | 1 | 2 | 6.RL.6 | Reading |

**Table C-5. English Language Arts, Grade 7 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 7 | 1 | 1 | TDA | OP | 4 | 3 | 7.W.9 | Writing |
| 7 | 2 | 2 | MC | OP | 1 | 3 | 7.W.2.a | Writing |
| 7 | 2 | 3 | TE | OP | 2 | 2 | 7.W.3.c | Writing |
| 7 | 2 | 4 | EBSR | OP | 2 | 3 | 7.W.8 | Writing |
| 7 | 2 | 5 | MC | OP | 1 | 2 | 7.L.2 | Writing |
| 7 | 2 | 6 | TE | OP | 2 | 2 | 7.L.3.a | Writing |
| 7 | 2 | 7 | MC | OP | 1 | 2 | 7.W.8 | Writing |
| 7 | 2 | 8 | MC | OP | 1 | 2 | 7.W.8 | Writing |
| 7 | 2 | 9 | MC | OP | 1 | 2 | 7.L.1.a | Writing |
| 7 | 2 | 10 | TE | OP | 2 | 3 | 7.W.8 | Writing |
| 7 | 2 | 11 | MC | OP | 1 | 3 | 7.L.1.b | Writing |
| 7 | 2 | 12 | TE | OP | 1 | 1 | 7.L.2.b | Writing |
| 7 | 2 | 13 | MC | OP | 1 | 2 | 7.W.8 | Writing |
| 7 | 3 | 14 | EBSR | OP | 2 | 3 | 7.SL.2 | Listening |
| 7 | 3 | 15 | MC | OP | 1 | 2 | 7.SL.2 | Listening |
| 7 | 3 | 16 | MS | OP | 2 | 2 | 7.SL.2 | Listening |
| 7 | 3 | 17 | MC | OP | 1 | 2 | 7.SL.2 | Listening |
| 7 | 3 | 18 | EBSR | OP | 2 | 3 | 7.SL.3 | Listening |
| 7 | 4 | 19 | MC | OP | 1 | 1 | 7.RI.1 | Reading |
| 7 | 4 | 20 | MS | OP | 2 | 3 | 7.RI.2 | Reading |
| 7 | 4 | 21 | MC | OP | 1 | 2 | 7.L.4.a | Reading |
| 7 | 4 | 22 | MC | OP | 1 | 2 | 7.RI.5 | Reading |
| 7 | 4 | 23 | EBSR | OP | 2 | 3 | 7.RI.3 | Reading |
| 7 | 4 | 24 | MC | OP | 1 | 3 | 7.RI.3 | Reading |
| 7 | 4 | 25 | TE | OP | 1 | 2 | 7.RL.4 | Reading |
| 7 | 4 | 26 | MC | OP | 1 | 2 | 7.RL.3 | Reading |
| 7 | 4 | 27 | EBSR | OP | 2 | 2 | 7.RL.6 | Reading |
| 7 | 4 | 28 | MC | OP | 1 | 3 | 7.RL.2 | Reading |
| 7 | 4 | 29 | TE | OP | 1 | 2 | 7.RL.1 | Reading |
| 7 | 4 | 30 | MC | OP | 1 | 2 | 7.RL.4 | Reading |
| 7 | 4 | 31 | EBSR | OP | 2 | 3 | 7.RI.8 | Reading |
| 7 | 4 | 32 | MC | OP | 1 | 3 | 7.RL.9 | Reading |
| 7 | 4 | 33 | MC | OP | 1 | 3 | 7.RI.6 | Reading |
| 7 | 4 | 34 | MS | OP | 2 | 2 | 7.RI.2 | Reading |
| 7 | 4 | 35 | MS | OP | 2 | 2 | 7.RI.4 | Reading |
| 7 | 4 | 36 | MC | OP | 1 | 3 | 7.RI.8 | Reading |

**Table C-6. English Language Arts, Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 1 | TDA | OP | 4 | 3 | 8.W.2 | Writing |
| 8 | 2 | 2 | MC | OP | 1 | 2 | 8.W.1.e | Writing |
| 8 | 2 | 3 | MC | OP | 1 | 2 | 8.W.2.e | Writing |
| 8 | 2 | 4 | MC | OP | 1 | 2 | 8.W.8 | Writing |
| 8 | 2 | 5 | MC | OP | 1 | 2 | 8.W.8 | Writing |
| 8 | 2 | 6 | MC | OP | 1 | 2 | 8.W.1.e | Writing |
| 8 | 2 | 7 | MC | OP | 1 | 2 | 8.W.3.c | Writing |
| 8 | 2 | 8 | MC | OP | 1 | 2 | 8.L.2.b | Writing |
| 8 | 2 | 9 | EBSR | OP | 2 | 3 | 8.W.8 | Writing |
| 8 | 2 | 10 | TE | OP | 2 | 2 | 8.L.1.d | Writing |
| 8 | 2 | 11 | MC | OP | 1 | 2 | 8.L.2.a | Writing |
| 8 | 2 | 12 | TE | OP | 2 | 2 | 8.L.1.c | Writing |
| 8 | 2 | 13 | MC | OP | 1 | 3 | 8.W.8 | Writing |
| 8 | 2 | 14 | MC | OP | 1 | 2 | 8.W.8 | Writing |
| 8 | 3 | 15 | MS | OP | 2 | 2 | 8.SL.3 | Listening |
| 8 | 3 | 16 | EBSR | OP | 2 | 3 | 8.SL.3 | Listening |
| 8 | 3 | 17 | MC | OP | 1 | 2 | 8.SL.3 | Listening |
| 8 | 3 | 18 | MC | OP | 1 | 2 | 8.SL.2 | Listening |
| 8 | 3 | 19 | EBSR | OP | 2 | 3 | 8.SL.3 | Listening |
| 8 | 4 | 20 | MC | OP | 1 | 2 | 8.RI.8 | Reading |
| 8 | 4 | 21 | MC | OP | 1 | 2 | 8.RI.6 | Reading |
| 8 | 4 | 22 | MC | OP | 1 | 2 | 8.RI.5 | Reading |
| 8 | 4 | 23 | MS | OP | 2 | 2 | 8.RI.3 | Reading |
| 8 | 4 | 24 | MC | OP | 1 | 2 | 8.RI.2 | Reading |
| 8 | 4 | 25 | MC | OP | 1 | 2 | 8.RL.4 | Reading |
| 8 | 4 | 26 | MC | OP | 1 | 3 | 8.RL.3 | Reading |
| 8 | 4 | 27 | TE | OP | 2 | 2 | 8.RL.4 | Reading |
| 8 | 4 | 28 | MC | OP | 1 | 2 | 8.RL.3 | Reading |
| 8 | 4 | 29 | MC | OP | 1 | 3 | 8.RL.6 | Reading |
| 8 | 4 | 30 | MC | OP | 1 | 2 | 8.RI.3 | Reading |
| 8 | 4 | 31 | MC | OP | 1 | 2 | 8.RI.1 | Reading |
| 8 | 4 | 32 | MC | OP | 1 | 3 | 8.RI.5 | Reading |
| 8 | 4 | 33 | EBSR | OP | 2 | 2 | 8.L.4.RI | Reading |
| 8 | 4 | 34 | MC | OP | 1 | 2 | 8.RI.5 | Reading |
| 8 | 4 | 35 | MC | OP | 1 | 2 | 8.RI.6 | Reading |
| 8 | 4 | 36 | MC | OP | 1 | 2 | 8.RL.1 | Reading |
| 8 | 4 | 37 | MC | OP | 1 | 2 | 8.RL.1 | Reading |
| 8 | 4 | 38 | EBSR | OP | 2 | 3 | 8.RL.3 | Reading |
| 8 | 4 | 39 | MC | OP | 1 | 2 | 8.RL.6 | Reading |

# Appendix D
# Spring 2022 Mathematics
# Operational Test Maps

# List of Tables

**Table D-1. Mathematics Grade 3 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 1 | TE | OP | 1 | 2 | 3.OA.1 | OA |
| 3 | 1 | 2 | SA | OP | 1 | 1 | 3.NF.1 | NF |
| 3 | 1 | 3 | MC | OP | 1 | 2 | 3.MD.1 | MD |
| 3 | 1 | 4 | TE | OP | 1 | 2 | 3.NBT.1 | NBT |
| 3 | 1 | 5 | MC | OP | 1 | 2 | 3.OA.2 | OA |
| 3 | 1 | 6 | SA | OP | 1 | 1 | 3.G.2 | G |
| 3 | 1 | 7 | MC | OP | 1 | 2 | 3.MD.4 | MD |
| 3 | 1 | 8 | SA | OP | 1 | 1 | 3.NBT.1 | NBT |
| 3 | 1 | 9 | TE | OP | 1 | 2 | 3.MD.3 | MD |
| 3 | 1 | 10 | MC | OP | 1 | 2 | 3.NF.3.d | NF |
| 3 | 1 | 11 | MC | OP | 1 | 2 | 3.OA.6 | OA |
| 3 | 1 | 12 | MC | OP | 1 | 2 | 3.MD.7.d | MD |
| 3 | 1 | 13 | MC | OP | 1 | 1 | 3.NBT.2 | NBT |
| 3 | 1 | 14 | SA | OP | 1 | 1 | 3.G.1 | G |
| 3 | 1 | 15 | MC | OP | 1 | 1 | 3.NF.2.a | NF |
| 3 | 1 | 16 | MC | OP | 1 | 3 | 3.OA.8 | OA |
| 3 | 1 | 17 | SA | OP | 1 | 1 | 3.MD.6 | MD |
| 3 | 1 | 18 | TE | OP | 1 | 3 | 3.NF.3 | NF |
| 3 | 1 | 19 | MC | OP | 1 | 2 | 3.NBT.3 | NBT |
| 3 | 1 | 20 | MC | OP | 1 | 2 | 3.G.2 | G |
| 3 | 1 | 21 | SA | OP | 1 | 2 | 3.OA.7 | OA |
| 3 | 2 | 22 | MC | OP | 1 | 2 | 3.MD.8 | MD |
| 3 | 2 | 23 | TE | OP | 1 | 1 | 3.OA.5 | OA |
| 3 | 2 | 24 | SA | OP | 1 | 1 | 3.NBT.1 | NBT |
| 3 | 2 | 25 | TE | OP | 1 | 2 | 3.G.1 | G |
| 3 | 2 | 26 | SA | OP | 1 | 1 | 3.NF.2.b | NF |
| 3 | 2 | 27 | SA | OP | 1 | 1 | 3.OA.3 | OA |
| 3 | 2 | 28 | MC | OP | 1 | 3 | 3.NBT.2 | NBT |
| 3 | 2 | 29 | MC | OP | 1 | 1 | 3.G.2 | G |
| 3 | 2 | 30 | MC | OP | 1 | 1 | 3.MD.2 | MD |
| 3 | 2 | 31 | MC | OP | 1 | 2 | 3.NF.3.a | NF |
| 3 | 2 | 32 | SA | OP | 1 | 1 | 3.OA.4 | OA |
| 3 | 2 | 33 | SA | OP | 1 | 1 | 3.MD.7.b | MD |
| 3 | 2 | 34 | MC | OP | 1 | 1 | 3.NF.2.b | NF |
| 3 | 2 | 35 | SA | OP | 1 | 1 | 3.NBT.2 | NBT |
| 3 | 2 | 36 | MC | OP | 1 | 1 | 3.G.1 | G |
| 3 | 2 | 37 | MC | OP | 1 | 1 | 3.MD.2 | MD |
| 3 | 2 | 38 | MC | OP | 1 | 1 | 3.NBT.3 | NBT |
| 3 | 2 | 39 | MC | OP | 1 | 2 | 3.NF.1 | NF |
| 3 | 2 | 40 | MC | OP | 1 | 2 | 3.OA.9 | OA |
| 3 | 2 | 41 | SA | OP | 1 | 1 | 3.MD.5.a | MD |
| 3 | 2 | 42 | TE | OP | 1 | 2 | 3.G.2 | G |

Domain Names: OA = Operations and Algebraic Thinking; NBT = Number and Operations in Base Ten; NF = Number and Operations – Fractions; MD = Measurement and Data; G = Geometry

## Table D-2. Mathematics Grade 4 Test Map

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|-------|---------|---------------|-----------|------------|------------------|--------------------|----------|--------|
| 4 | 1 | 1 | MC | OP | 1 | 2 | 4.NF.7 | NF |
| 4 | 1 | 2 | SA | OP | 1 | 1 | 4.OA.1 | OA |
| 4 | 1 | 3 | TE | OP | 1 | 2 | 4.NBT.1 | NBT |
| 4 | 1 | 4 | MC | OP | 1 | 3 | 4.G.3 | G |
| 4 | 1 | 5 | MC | OP | 1 | 1 | 4.MD.6 | MD |
| 4 | 1 | 6 | MC | OP | 1 | 2 | 4.NBT.5 | NBT |
| 4 | 1 | 7 | MC | OP | 1 | 2 | 4.OA.2 | OA |
| 4 | 1 | 8 | SA | OP | 1 | 2 | 4.MD.7 | MD |
| 4 | 1 | 9 | MC | OP | 1 | 2 | 4.NBT.3 | NBT |
| 4 | 1 | 10 | TE | OP | 1 | 1 | 4.NF.2 | NF |
| 4 | 1 | 11 | MC | OP | 1 | 2 | 4.G.2 | G |
| 4 | 1 | 12 | MC | OP | 1 | 2 | 4.NBT.6 | NBT |
| 4 | 1 | 13 | MC | OP | 1 | 2 | 4.NF.3 | NF |
| 4 | 1 | 14 | MC | OP | 1 | 2 | 4.MD.4 | MD |
| 4 | 1 | 15 | MC | OP | 1 | 3 | 4.OA.4 | OA |
| 4 | 1 | 16 | SA | OP | 1 | 1 | 4.NBT.3 | NBT |
| 4 | 1 | 17 | MC | OP | 1 | 1 | 4.MD.3 | MD |
| 4 | 1 | 18 | MC | OP | 1 | 2 | 4.NF.4 | NF |
| 4 | 1 | 19 | TE | OP | 1 | 2 | 4.OA.3 | OA |
| 4 | 1 | 20 | SA | OP | 1 | 1 | 4.G.1 | G |
| 4 | 1 | 21 | MC | OP | 1 | 1 | 4.NF.6 | NF |
| 4 | 1 | 22 | MC | OP | 1 | 3 | 4.OA.5 | OA |
| 4 | 1 | 23 | MC | OP | 1 | 2 | 4.MD.2 | MD |
| 4 | 2 | 24 | TE | OP | 1 | 1 | 4.NBT.2 | NBT |
| 4 | 2 | 25 | MC | OP | 1 | 2 | 4.OA.1 | OA |
| 4 | 2 | 26 | MC | OP | 1 | 2 | 4.NF.5 | NF |
| 4 | 2 | 27 | MC | OP | 1 | 1 | 4.NF.1 | NF |
| 4 | 2 | 28 | MC | OP | 1 | 3 | 4.MD.1 | MD |
| 4 | 2 | 29 | MC | OP | 1 | 2 | 4.OA.5 | OA |
| 4 | 2 | 30 | MC | OP | 1 | 2 | 4.NBT.5 | NBT |
| 4 | 2 | 31 | SA | OP | 1 | 2 | 4.NF.4.c | NF |
| 4 | 2 | 32 | TE | OP | 1 | 2 | 4.G.3 | G |
| 4 | 2 | 33 | MC | OP | 1 | 1 | 4.MD.5.b | MD |
| 4 | 2 | 34 | MC | OP | 1 | 2 | 4.OA.2 | OA |
| 4 | 2 | 35 | SA | OP | 1 | 2 | 4.MD.3 | MD |
| 4 | 2 | 36 | MC | OP | 1 | 2 | 4.NF.7 | NF |
| 4 | 2 | 37 | MC | OP | 1 | 1 | 4.G.3 | G |
| 4 | 2 | 38 | MC | OP | 1 | 2 | 4.NBT.6 | NBT |
| 4 | 2 | 39 | TE | OP | 1 | 1 | 4.NF.3.a | NF |
| 4 | 2 | 40 | SA | OP | 1 | 1 | 4.G.2 | G |
| 4 | 2 | 41 | MC | OP | 1 | 2 | 4.MD.7 | MD |
| 4 | 2 | 42 | MC | OP | 1 | 2 | 4.OA.3 | OA |
| 4 | 2 | 43 | SA | OP | 1 | 1 | 4.NBT.4 | NBT |

**Table D-2 (continued). Mathematics Grade 4 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|-------|---------|---------------|-----------|------------|------------------|--------------------|----------|--------|
| 4 | 2 | 44 | MC | OP | 1 | 1 | 4.G.1 | G |
| 4 | 2 | 45 | SA | OP | 1 | 1 | 4.OA.4 | OA |
| 4 | 2 | 46 | MC | OP | 1 | 1 | 4.MD.6 | MD |

Domain Names: OA = Operations and Algebraic Thinking; NBT = Number and Operations in Base Ten; NF = Number and Operations – Fractions; MD = Measurement and Data; G = Geometry

**Table D-3. Mathematics Grade 5 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 5 | 1 | 1 | SA | OP | 1 | 2 | 5.NF.7.b | NF |
| 5 | 1 | 2 | MC | OP | 1 | 2 | 5.MD.2 | MD |
| 5 | 1 | 3 | TE | OP | 1 | 2 | 5.OA.2 | OA |
| 5 | 1 | 4 | MC | OP | 1 | 2 | 5.G.2 | G |
| 5 | 1 | 5 | MC | OP | 1 | 2 | 5.OA.3 | OA |
| 5 | 1 | 6 | SA | OP | 1 | 1 | 5.NBT.3.a | NBT |
| 5 | 1 | 7 | TE | OP | 1 | 2 | 5.MD.1 | MD |
| 5 | 1 | 8 | MC | OP | 1 | 2 | 5.G.1 | G |
| 5 | 1 | 9 | MC | OP | 1 | 2 | 5.NBT.1 | NBT |
| 5 | 1 | 10 | SA | OP | 1 | 1 | 5.MD.3.b | MD |
| 5 | 1 | 11 | MC | OP | 1 | 3 | 5.OA.3 | OA |
| 5 | 1 | 12 | MC | OP | 1 | 1 | 5.G.3 | G |
| 5 | 1 | 13 | SA | OP | 1 | 2 | 5.NF.4.a | NF |
| 5 | 1 | 14 | MC | OP | 1 | 1 | 5.MD.4 | MD |
| 5 | 1 | 15 | MC | OP | 1 | 2 | 5.OA.1 | OA |
| 5 | 1 | 16 | SA | OP | 1 | 2 | 5.G.4 | G |
| 5 | 1 | 17 | MC | OP | 1 | 1 | 5.NF.6 | NF |
| 5 | 1 | 18 | SA | OP | 1 | 1 | 5.NBT.4 | NBT |
| 5 | 1 | 19 | MC | OP | 1 | 2 | 5.MD.5.a | MD |
| 5 | 1 | 20 | MC | OP | 1 | 3 | 5.NBT.6 | NBT |
| 5 | 1 | 21 | SA | OP | 1 | 2 | 5.OA.1 | OA |
| 5 | 1 | 22 | TE | OP | 1 | 2 | 5.NBT.5 | NBT |
| 5 | 1 | 23 | MC | OP | 1 | 2 | 5.NF.2 | NF |
| 5 | 2 | 24 | MC | OP | 1 | 1 | 5.NBT.7 | NBT |
| 5 | 2 | 25 | TE | OP | 1 | 2 | 5.NF.1 | NF |
| 5 | 2 | 26 | MC | OP | 1 | 1 | 5.G.3 | G |
| 5 | 2 | 27 | SA | OP | 1 | 1 | 5.OA.2 | OA |
| 5 | 2 | 28 | MC | OP | 1 | 2 | 5.NF.7.a | NF |
| 5 | 2 | 29 | MC | OP | 1 | 1 | 5.MD.3.b | MD |
| 5 | 2 | 30 | TE | OP | 1 | 2 | 5.G.2 | G |
| 5 | 2 | 31 | TE | OP | 1 | 2 | 5.NBT.3.b | NBT |
| 5 | 2 | 32 | MC | OP | 1 | 3 | 5.NF.4.a | NF |
| 5 | 2 | 33 | MC | OP | 1 | 2 | 5.OA.3 | OA |
| 5 | 2 | 34 | SA | OP | 1 | 2 | 5.MD.2 | MD |
| 5 | 2 | 35 | MC | OP | 1 | 2 | 5.NBT.3.b | NBT |
| 5 | 2 | 36 | MC | OP | 1 | 2 | 5.MD.5 | MD |
| 5 | 2 | 37 | SA | OP | 1 | 1 | 5.G.1 | G |
| 5 | 2 | 38 | MC | OP | 1 | 2 | 5.NF.5 | NF |
| 5 | 2 | 39 | SA | OP | 1 | 1 | 5.NBT.2 | NBT |
| 5 | 2 | 40 | MC | OP | 1 | 1 | 5.G.4 | G |
| 5 | 2 | 41 | SA | OP | 1 | 2 | 5.MD.4 | MD |
| 5 | 2 | 42 | MC | OP | 1 | 2 | 5.OA.1 | OA |
| 5 | 2 | 43 | MC | OP | 1 | 2 | 5.NF.3 | NF |

**Table D-3 (continued). Mathematics Grade 5 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|-------|---------|---------------|-----------|------------|------------------|--------------------|----------|--------|
| 5 | 2 | 44 | SA | OP | 1 | 2 | 5.G.2 | G |
| 5 | 2 | 45 | MC | OP | 1 | 2 | 5.MD.1 | MD |
| 5 | 2 | 46 | MC | OP | 1 | 2 | 5.OA.2 | OA |

Domain Names: OA = Operations and Algebraic Thinking; NBT = Number and Operations in Base Ten; NF = Number and Operations – Fractions; MD = Measurement and Data; G = Geometry

**Table D-4. Mathematics Grade 6 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 6 | 1 | 1 | MC | OP | 1 | 2 | 6.RP.3.d | RP |
| 6 | 1 | 2 | MC | OP | 1 | 1 | 6.EE.2.a | EE |
| 6 | 1 | 3 | TE | OP | 1 | 3 | 6.RP.3 | RP |
| 6 | 1 | 4 | MC | OP | 1 | 2 | 6.RP.2 | RP |
| 6 | 1 | 5 | MC | OP | 1 | 1 | 6.EE.3 | EE |
| 6 | 1 | 6 | SA | OP | 1 | 1 | 6.RP.3.c | RP |
| 6 | 1 | 7 | SA | OP | 1 | 1 | 6.NS.2 | NS |
| 6 | 1 | 8 | MC | OP | 1 | 1 | 6.NS.3 | NS |
| 6 | 1 | 9 | MC | OP | 1 | 2 | 6.EE.4 | EE |
| 6 | 1 | 10 | SA | OP | 1 | 2 | 6.RP.1 | RP |
| 6 | 1 | 11 | MC | OP | 1 | 2 | 6.NS.1 | NS |
| 6 | 1 | 12 | SA | OP | 1 | 1 | 6.EE.1 | EE |
| 6 | 1 | 13 | MC | OP | 1 | 2 | 6.RP.3.b | RP |
| 6 | 1 | 14 | TE | OP | 1 | 2 | 6.NS.4 | NS |
| 6 | 1 | 15 | MC | OP | 1 | 2 | 6.RP.3.a | RP |
| 6 | 1 | 16 | TE | OP | 1 | 1 | 6.EE.2.c | EE |
| 6 | 2 | 17 | MC | OP | 1 | 2 | 6.G.1 | G |
| 6 | 2 | 18 | MC | OP | 1 | 2 | 6.EE.7 | EE |
| 6 | 2 | 19 | MC | OP | 1 | 1 | 6.NS.5 | NS |
| 6 | 2 | 20 | SA | OP | 1 | 2 | 6.SP.5.a | SP |
| 6 | 2 | 21 | TE | OP | 1 | 2 | 6.EE.7 | EE |
| 6 | 2 | 22 | MC | OP | 1 | 2 | 6.NS.6 | NS |
| 6 | 2 | 23 | TE | OP | 1 | 1 | 6.SP.4 | SP |
| 6 | 2 | 24 | MC | OP | 1 | 2 | 6.G.3 | G |
| 6 | 2 | 25 | MC | OP | 1 | 2 | 6.EE.8 | EE |
| 6 | 2 | 26 | MC | OP | 1 | 1 | 6.SP.1 | SP |
| 6 | 2 | 27 | SA | OP | 1 | 1 | 6.NS.6.b | NS |
| 6 | 2 | 28 | SA | OP | 1 | 2 | 6.EE.9 | EE |
| 6 | 2 | 29 | SA | OP | 1 | 1 | 6.NS.7.c | NS |
| 6 | 2 | 30 | MC | OP | 1 | 2 | 6.SP.5.b | SP |
| 6 | 2 | 31 | TE | OP | 1 | 2 | 6.G.3 | G |
| 6 | 2 | 32 | MC | OP | 1 | 2 | 6.SP.5.b | SP |
| 6 | 2 | 33 | MC | OP | 1 | 2 | 6.NS.8 | NS |
| 6 | 2 | 34 | MC | OP | 1 | 2 | 6.EE.9 | EE |
| 6 | 2 | 35 | MC | OP | 1 | 2 | 6.SP.4 | SP |
| 6 | 2 | 36 | SA | OP | 1 | 2 | 6.G.4 | G |
| 6 | 2 | 37 | MC | OP | 1 | 2 | 6.EE.6 | EE |
| 6 | 2 | 38 | MC | OP | 1 | 1 | 6.SP.3 | SP |
| 6 | 2 | 39 | TE | OP | 1 | 2 | 6.G.1 | G |
| 6 | 2 | 40 | MC | OP | 1 | 2 | 6.NS.8 | NS |
| 6 | 2 | 41 | MC | OP | 1 | 2 | 6.G.2 | G |
| 6 | 2 | 42 | MC | OP | 1 | 2 | 6.SP.2 | SP |
| 6 | 2 | 43 | MC | OP | 1 | 2 | 6.G.4 | G |

**Table D-4 (continued). Mathematics Grade 6 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|-------|---------|---------------|-----------|------------|------------------|--------------------|----------|--------|
| 6 | 2 | 44 | MC | OP | 1 | 1 | 6.SP.1 | SP |
| 6 | 2 | 45 | TE | OP | 1 | 1 | 6.NS.7.a | NS |
| 6 | 2 | 46 | MC | OP | 1 | 2 | 6.SP.5 | SP |

Domain Names: G = Geometry; EE = Expressions and Equations; NS = The Number System; SP = Statistics and Probability; RP = Ratios and Proportional Relationships

**Table D-5. Mathematics Grade 7 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 7 | 1 | 1 | SA | OP | 1 | 2 | 7.NS.3 | NS |
| 7 | 1 | 2 | SA | OP | 1 | 1 | 7.NS.1.c | NS |
| 7 | 1 | 3 | MC | OP | 1 | 2 | 7.EE.2 | EE |
| 7 | 1 | 4 | SA | OP | 1 | 1 | 7.NS.2.c | NS |
| 7 | 1 | 5 | MC | OP | 1 | 2 | 7.NS.1.b | NS |
| 7 | 1 | 6 | MC | OP | 1 | 2 | 7.NS.2 | NS |
| 7 | 1 | 7 | MC | OP | 1 | 2 | 7.EE.1 | EE |
| 7 | 1 | 8 | TE | OP | 1 | 1 | 7.NS.2 | NS |
| 7 | 1 | 9 | MC | OP | 1 | 2 | 7.EE.1 | EE |
| 7 | 1 | 10 | MC | OP | 1 | 2 | 7.NS.3 | NS |
| 7 | 1 | 11 | MC | OP | 1 | 3 | 7.NS.2.b | NS |
| 7 | 2 | 12 | MC | OP | 1 | 2 | 7.G.3 | G |
| 7 | 2 | 13 | MC | OP | 1 | 1 | 7.RP.1 | RP |
| 7 | 2 | 14 | MC | OP | 1 | 2 | 7.EE.3 | EE |
| 7 | 2 | 15 | SA | OP | 1 | 2 | 7.SP.8.c | SP |
| 7 | 2 | 16 | TE | OP | 1 | 2 | 7.EE.4.a | EE |
| 7 | 2 | 17 | MC | OP | 1 | 2 | 7.G.1 | G |
| 7 | 2 | 18 | MC | OP | 1 | 2 | 7.RP.3 | RP |
| 7 | 2 | 19 | TE | OP | 1 | 2 | 7.RP.2.d | RP |
| 7 | 2 | 20 | MC | OP | 1 | 2 | 7.G.4 | G |
| 7 | 2 | 21 | TE | OP | 1 | 2 | 7.SP.5 | SP |
| 7 | 2 | 22 | MC | OP | 1 | 2 | 7.EE.3 | EE |
| 7 | 2 | 23 | MC | OP | 1 | 2 | 7.G.4 | G |
| 7 | 2 | 24 | SA | OP | 1 | 1 | 7.SP.7.a | SP |
| 7 | 2 | 25 | SA | OP | 1 | 2 | 7.RP.3 | RP |
| 7 | 2 | 26 | MC | OP | 1 | 2 | 7.EE.4.a | EE |
| 7 | 2 | 27 | MC | OP | 1 | 2 | 7.SP.1 | SP |
| 7 | 2 | 28 | MC | OP | 1 | 2 | 7.G.3 | G |
| 7 | 2 | 29 | SA | OP | 1 | 1 | 7.RP.2.b | RP |
| 7 | 2 | 30 | MC | OP | 1 | 2 | 7.SP.7.b | SP |
| 7 | 2 | 31 | TE | OP | 1 | 2 | 7.G.2 | G |
| 7 | 2 | 32 | MC | OP | 1 | 3 | 7.SP.2 | SP |
| 7 | 2 | 33 | MC | OP | 1 | 2 | 7.RP.3 | RP |
| 7 | 2 | 34 | MC | OP | 1 | 2 | 7.G.6 | G |
| 7 | 2 | 35 | SA | OP | 1 | 2 | 7.EE.4.a | EE |
| 7 | 2 | 36 | TE | OP | 1 | 1 | 7.SP.8.b | SP |
| 7 | 2 | 37 | MC | OP | 1 | 2 | 7.G.1 | G |
| 7 | 2 | 38 | SA | OP | 1 | 1 | 7.SP.6 | SP |
| 7 | 2 | 39 | MC | OP | 1 | 2 | 7.RP.2 | RP |
| 7 | 2 | 40 | SA | OP | 1 | 2 | 7.G.5 | G |
| 7 | 2 | 41 | MC | OP | 1 | 2 | 7.SP.3 | SP |
| 7 | 2 | 42 | TE | OP | 1 | 3 | 7.EE.4.b | EE |
| 7 | 2 | 43 | MC | OP | 1 | 2 | 7.SP.2 | SP |

**Table D-5 (continued). Mathematics Grade 7 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 7 | 2 | 44 | MC | OP | 1 | 2 | 7.G.2 | G |
| 7 | 2 | 45 | SA | OP | 1 | 1 | 7.RP.1 | RP |
| 7 | 2 | 46 | MC | OP | 1 | 2 | 7.SP.8.b | SP |

Domain Names: G = Geometry; EE = Expressions and Equations; NS = The Number System; SP=Statistics and Probability; RP = Ratios and Proportional Relationships

**Table D-6. Mathematics Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 1 | MC | OP | 1 | 1 | 8.NS.2 | NS |
| 8 | 1 | 2 | MC | OP | 1 | 1 | 8.EE.2 | EE |
| 8 | 1 | 3 | MC | OP | 1 | 1 | 8.NS.1 | NS |
| 8 | 1 | 4 | MC | OP | 1 | 2 | 8.NS.2 | NS |
| 8 | 1 | 5 | SA | OP | 1 | 1 | 8.EE.1 | EE |
| 8 | 1 | 6 | MC | OP | 1 | 3 | 8.NS.1 | NS |
| 8 | 1 | 7 | MC | OP | 1 | 2 | 8.EE.3 | EE |
| 8 | 1 | 8 | SA | OP | 1 | 2 | 8.NS.2 | NS |
| 8 | 1 | 9 | TE | OP | 1 | 1 | 8.NS.1 | NS |
| 8 | 1 | 10 | MC | OP | 1 | 2 | 8.EE.2 | EE |
| 8 | 1 | 11 | SA | OP | 1 | 1 | 8.NS.2 | NS |
| 8 | 1 | 12 | MC | OP | 1 | 2 | 8.EE.3 | EE |
| 8 | 1 | 13 | MC | OP | 1 | 1 | 8.NS.1 | NS |
| 8 | 2 | 14 | MC | OP | 1 | 2 | 8.SP.4 | SP |
| 8 | 2 | 15 | TE | OP | 1 | 3 | 8.G.5 | G |
| 8 | 2 | 16 | MC | OP | 1 | 2 | 8.F.2 | F |
| 8 | 2 | 17 | SA | OP | 1 | 2 | 8.EE.7.a | EE |
| 8 | 2 | 18 | MC | OP | 1 | 2 | 8.F.4 | F |
| 8 | 2 | 19 | MC | OP | 1 | 2 | 8.G.7 | G |
| 8 | 2 | 20 | MC | OP | 1 | 2 | 8.SP.3 | SP |
| 8 | 2 | 21 | MC | OP | 1 | 1 | 8.G.1.b | G |
| 8 | 2 | 22 | MC | OP | 1 | 1 | 8.SP.2 | SP |
| 8 | 2 | 23 | TE | OP | 1 | 2 | 8.EE.5 | EE |
| 8 | 2 | 24 | MC | OP | 1 | 2 | 8.F.5 | F |
| 8 | 2 | 25 | TE | OP | 1 | 2 | 8.F.2 | F |
| 8 | 2 | 26 | MC | OP | 1 | 2 | 8.G.8 | G |
| 8 | 2 | 27 | SA | OP | 1 | 2 | 8.SP.3 | SP |
| 8 | 2 | 28 | MC | OP | 1 | 2 | 8.F.4 | F |
| 8 | 2 | 29 | MC | OP | 1 | 2 | 8.SP.4 | SP |
| 8 | 2 | 30 | MC | OP | 1 | 2 | 8.EE.7 | EE |
| 8 | 2 | 31 | SA | OP | 1 | 2 | 8.G.5 | G |
| 8 | 2 | 32 | TE | OP | 1 | 2 | 8.F.3 | F |
| 8 | 2 | 33 | MC | OP | 1 | 2 | 8.G.4 | G |
| 8 | 2 | 34 | TE | OP | 1 | 2 | 8.F.4 | F |
| 8 | 2 | 35 | MC | OP | 1 | 3 | 8.EE.8.a | EE |
| 8 | 2 | 36 | MC | OP | 1 | 2 | 8.G.2 | G |
| 8 | 2 | 37 | TE | OP | 1 | 2 | 8.SP.2 | SP |
| 8 | 2 | 38 | TE | OP | 1 | 2 | 8.F.1 | F |
| 8 | 2 | 39 | MC | OP | 1 | 2 | 8.G.6 | G |
| 8 | 2 | 40 | MC | OP | 1 | 2 | 8.SP.1 | SP |
| 8 | 2 | 41 | MC | OP | 1 | 3 | 8.EE.6 | EE |
| 8 | 2 | 42 | TE | OP | 1 | 2 | 8.G.3 | G |
| 8 | 2 | 43 | MC | OP | 1 | 1 | 8.F.5 | F |

**Table D-6 (continued). Mathematics Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 8 | 2 | 44 | MC | OP | 1 | 2 | 8.SP.1 | SP |
| 8 | 2 | 45 | MC | OP | 1 | 1 | 8.G.1.a | G |
| 8 | 2 | 46 | MC | OP | 1 | 1 | 8.F.1 | F |

Domain Names: G = Geometry; EE = Expressions and Equations; NS = The Number System; SP = Statistics and Probability; F = Functions

# Appendix E
# Spring 2022 Science
# Operational Test Maps

# List of Tables

**Table E-1. Science Grade 4 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain | Science and Engineering Practices Standard | Crosscutting Concepts Standard |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 1 | TE | OP | 1 | 3 | SCI.PS3.B.4 | Physical Science | SCI.SEP6.B.3-5 | SCI.CC5.3-5 |
| 4 | 1 | 2 | TE | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | SCI.SEP7.A.3-5 | |
| 4 | 1 | 3 | MC | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | | SCI.CC4.3-5 |
| 4 | 1 | 4 | TE | OP | 1 | 3 | SCI.ESS3.B.4 | Earth and Space Science | | SCI.CC2.3-5 |
| 4 | 1 | 5 | TE | OP | 1 | 3 | SCI.ESS1.C.4 | Earth and Space Science | SCI.SEP6.A.3-5 | |
| 4 | 1 | 6 | TE | OP | 1 | 2 | SCI.ETS1.B.3-5 | Engineering | SCI.SEP6.B.3-5 | |
| 4 | 1 | 7 | TE | OP | 1 | 2 | SCI.ETS1.C.3-5 | Engineering | SCI.SEP3.A.3-5 | |
| 4 | 1 | 8 | MC | OP | 1 | 2 | SCI.PS3.B.4 | Physical Science | SCI.SEP1.A.3-5 | SCI.CC5.3-5 |
| 4 | 1 | 9 | TE | OP | 1 | 2 | SCI.ETS1.B.3-5 | Engineering | SCI.SEP6.B.3-5 | |
| 4 | 1 | 10 | EBSR | OP | 1 | 3 | SCI.ETS1.C.3-5 | Engineering | SCI.SEP3.A.3-5 | |
| 4 | 1 | 11 | MC | OP | 1 | 2 | SCI.ESS2.A.4 | Earth and Space Science | SCI.SEP3.A.3-5 | SCI.CC2.3-5 |
| 4 | 1 | 12 | MC | OP | 1 | 2 | SCI.ETS1.B.3-5 | Engineering | SCI.SEP3.A.3-5 | |
| 4 | 1 | 13 | MC | OP | 1 | 3 | SCI.ESS3.A.4 | Earth and Space Science | SCI.SEP8.A.3-5 | |
| 4 | 1 | 14 | TE | OP | 1 | 2 | SCI.ESS1.C.4 | Earth and Space Science | SCI.SEP6.A.3-5 | SCI.CC1.3-5 |
| 4 | 1 | 15 | TE | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | SCI.SEP7.A.3-5 | |
| 4 | 2 | 16 | TE | OP | 1 | 2 | SCI.PS3.A.4 | Physical Science | SCI.SEP6.A.3-5 | SCI.CC5.3-5 |
| 4 | 2 | 17 | TE | OP | 1 | 2 | SCI.PS4.B.4 | Physical Science | SCI.SEP2.A.3-5 | |
| 4 | 2 | 18 | MC | OP | 1 | 3 | SCI.PS3.B.4 | Physical Science | SCI.SEP3.A.3-5 | SCI.CC5.3-5 |
| 4 | 2 | 19 | TE | OP | 1 | 3 | SCI.LS1.A.4 | Life Science | SCI.SEP7.A.3-5 | |
| 4 | 2 | 20 | TE | OP | 1 | 2 | SCI.PS3.A.4 | Physical Science | SCI.SEP6.A.3-5 | SCI.CC5.3-5 |
| 4 | 2 | 21 | TE | OP | 1 | 2 | SCI.PS4.B.4 | Physical Science | SCI.SEP2.A.3-5 | SCI.CC2.3-5 |
| 4 | 2 | 22 | MC | OP | 1 | 2 | SCI.ETS1.A.3-5 | Engineering | SCI.SEP1.B.3-5 | |
| 4 | 2 | 23 | MC | OP | 1 | 2 | SCI.PS3.B.4 | Physical Science | SCI.SEP6.B.3-5 | |
| 4 | 2 | 24 | TE | OP | 1 | 2 | SCI.PS4.C.4 | Physical Science | SCI.SEP8.A.3-5 | |

**Table E-1 (continued). Science Grade 4 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain | Science and Engineering Practices Standard | Crosscutting Concepts Standard |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 25 | MC | OP | 1 | 2 | SCI.ESS2.A.4 | Earth and Space Science | SCI.SEP3.A.3-5 | SCI.CC2.3-5 |
| 4 | 3 | 26 | MC | OP | 1 | 2 | SCI.PS4.A.4 | Physical Science | SCI.SEP2.A.3-5 | SCI.CC2.3-5 |
| 4 | 3 | 27 | MC | OP | 1 | 2 | SCI.PS3.B.4 | Physical Science | SCI.SEP8.A.3-5 | SCI.CC5.3-5 |
| 4 | 3 | 28 | MC | OP | 1 | 2 | SCI.ESS2.B.4 | Earth and Space Science | SCI.SEP6.A.3-5 | SCI.CC1.3-5 |
| 4 | 3 | 29 | TE | OP | 1 | 2 | SCI.ESS2.A.4 | Earth and Space Science | SCI.SEP4.A.3-5 | SCI.CC2.3-5 |
| 4 | 3 | 30 | TE | OP | 1 | 3 | SCI.ESS2.B.4 | Earth and Space Science | SCI.SEP4.A.3-5 | |
| 4 | 3 | 31 | TE | OP | 1 | 3 | SCI.ETS1.A.3-5 | Engineering | SCI.SEP1.A.3-5 | |
| 4 | 3 | 32 | TE | OP | 1 | 2 | SCI.ETS1.B.3-5 | Engineering | SCI.SEP6.B.3-5 | |
| 4 | 3 | 33 | MC | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | SCI.SEP6.A.3-5 | SCI.CC6.3-5 |
| 4 | 3 | 34 | TE | OP | 1 | 3 | SCI.PS3.A.4 | Physical Science | SCI.SEP6.A.3-5 | |
| 4 | 3 | 35 | TE | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | | SCI.CC4.3-5 |
| 4 | 3 | 36 | TE | OP | 1 | 2 | SCI.LS1.A.4 | Life Science | SCI.SEP8.A.3-5 | SCI.CC4.3-5 |
| 4 | 3 | 37 | TE | OP | 1 | 3 | SCI.LS1.A.4 | Life Science | SCI.SEP7.A.3-5 | SCI.CC4.3-5 |
| 4 | 3 | 38 | MC | OP | 1 | 2 | SCI.LS1.D.4 | Life Science | SCI.SEP6.A.3-5 | SCI.CC4.3-5 |
| 4 | 3 | 39 | TE | OP | 1 | 3 | SCI.LS1.A.4 | Life Science | SCI.SEP7.A.3-5 | SCI.CC4.3-5 |
| 4 | 3 | 40 | MC | OP | 1 | 2 | SCI.PS3.C.4 | Physical Science | | SCI.CC2.3-5 |

**Table E-2. Science Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain | Science and Engineering Practices Standard | Crosscutting Concepts Standard |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 1 | MS | OP | 1 | 3 | SCI.LS4.A.m | Life Science | SCI.SEP4.A.m | SCI.CC1.m |
| 8 | 1 | 2 | MC | OP | 1 | 3 | SCI.PS4.B.m | Physical Science | SCI.SEP2.A.m | |
| 8 | 1 | 3 | TE | OP | 1 | 2 | SCI.LS2.C.m | Life Science | SCI.SEP7.A.m | SCI.CC1.m |
| 8 | 1 | 4 | TE | OP | 1 | 2 | SCI.LS2.A.m | Life Science | SCI.SEP4.A.m | SCI.CC1.m |
| 8 | 1 | 5 | TE | OP | 1 | 2 | SCI.PS4.A.m | Physical Science | SCI.SEP2.A.m | SCI.CC6.m |
| 8 | 1 | 6 | TE | OP | 1 | 2 | SCI.ESS2.C.m | Earth and Space Science | SCI.SEP2.A.m | SCI.CC4.m |
| 8 | 1 | 7 | TE | OP | 1 | 2 | SCI.ESS1.B.m | Earth and Space Science | SCI.SEP2.A.m | SCI.CC1.m |
| 8 | 1 | 8 | MC | OP | 1 | 2 | SCI.ESS1.B.m | Earth and Space Science | SCI.SEP2.A.m | SCI.CC4.m |
| 8 | 1 | 9 | TE | OP | 1 | 2 | SCI.ESS1.B.m | Earth and Space Science | SCI.SEP2.A.m | |
| 8 | 1 | 10 | MC | OP | 1 | 2 | SCI.ESS1.C.m | Earth and Space Science | SCI.SEP6.A.m | SCI.CC3.m |
| 8 | 1 | 11 | MC | OP | 1 | 3 | SCI.LS4.C.m | Life Science | SCI.SEP5.A.m | SCI.CC2.m |
| 8 | 1 | 12 | MC | OP | 1 | 2 | SCI.LS3.B.m | Life Science | SCI.SEP2.A.m | |
| 8 | 1 | 13 | TE | OP | 1 | 3 | SCI.LS4.B.m | Life Science | SCI.SEP8.A.m | SCI.CC2.m |
| 8 | 1 | 14 | TE | OP | 1 | 2 | SCI.LS4.B.m | Life Science | SCI.SEP7.A.m | SCI.CC2.m |
| 8 | 1 | 15 | TE | OP | 1 | 3 | SCI.ESS3.A.m | Earth and Space Science | SCI.SEP6.A.m | SCI.CC2.m |
| 8 | 2 | 16 | TE | OP | 1 | 3 | SCI.PS2.A.m | Physical Science | SCI.SEP3.A.m | SCI.CC7.m |
| 8 | 2 | 17 | MC | OP | 1 | 2 | SCI.PS4.B.m | Physical Science | SCI.SEP2.A.m | SCI.CC4.m |
| 8 | 2 | 18 | TE | OP | 1 | 2 | SCI.PS1.A.m | Physical Science | SCI.SEP2.A.m | SCI.CC2.m |
| 8 | 2 | 19 | MC | OP | 1 | 2 | SCI.ETS1.A.m | Engineering | | SCI.CC6.m |
| 8 | 2 | 20 | MC | OP | 1 | 3 | SCI.ESS1.B.m | Earth and Space Science | SCI.SEP6.A.m | |
| 8 | 2 | 21 | TE | OP | 1 | 2 | SCI.ETS1.A.m | Engineering | SCI.SEP6.A.m | |
| 8 | 2 | 22 | TE | OP | 1 | 2 | SCI.ETS1.A.m | Engineering | SCI.SEP1.A.m | |
| 8 | 2 | 23 | TE | OP | 1 | 2 | SCI.ESS1.C.m | Earth and Space Science | SCI.SEP6.A.m | SCI.CC3.m |

**Table E-2 (continued). Science Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain | Science and Engineering Practices Standard | Crosscutting Concepts Standard |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 2 | 24 | TE | OP | 1 | 3 | SCI.ESS3.C.m | Earth and Space Science | SCI.SEP7.A.m | SCI.CC2.m |
| 8 | 2 | 25 | TE | OP | 1 | 2 | SCI.ESS2.A.m | Earth and Space Science | SCI.SEP6.A.m | SCI.CC3.m |
| 8 | 3 | 26 | TE | OP | 1 | 2 | SCI.LS2.B.m | Life Science | SCI.SEP2.A.m | SCI.CC5.m |
| 8 | 3 | 27 | TE | OP | 1 | 2 | SCI.LS2.A.m | Life Science | SCI.SEP4.A.m | |
| 8 | 3 | 28 | TE | OP | 1 | 2 | SCI.LS1.C.m | Life Science | SCI.SEP6.A.m | SCI.CC5.m |
| 8 | 3 | 29 | TE | OP | 1 | 2 | SCI.ETS1.A.m | Engineering | SCI.SEP1.A.m | |
| 8 | 3 | 30 | TE | OP | 1 | 2 | SCI.ETS1.B.m | Engineering | SCI.SEP4.A.m | |
| 8 | 3 | 31 | TE | OP | 1 | 2 | SCI.PS1.B.m | Physical Science | SCI.SEP2.A.m | SCI.CC5.m |
| 8 | 3 | 32 | MC | OP | 1 | 3 | SCI.PS1.B.m | Physical Science | SCI.SEP6.B.m | |
| 8 | 3 | 33 | TE | OP | 1 | 3 | SCI.ETS1.A.m | Engineering | SCI.SEP1.A.m | |
| 8 | 3 | 34 | EBSR | OP | 1 | 3 | SCI.ETS1.A.m | Engineering | SCI.SEP1.A.m | |
| 8 | 3 | 35 | TE | OP | 1 | 2 | SCI.PS1.A.m | Physical Science | SCI.SEP2.A.m | SCI.CC3.m |
| 8 | 3 | 36 | MC | OP | 1 | 2 | SCI.PS4.C.m | Physical Science | SCI.SEP1.A.m | |
| 8 | 3 | 37 | TE | OP | 1 | 3 | SCI.PS2.A.m | Physical Science | SCI.SEP3.A.m | SCI.CC7.m |
| 8 | 3 | 38 | MC | OP | 1 | 2 | SCI.ETS1.B.m | Engineering | SCI.SEP4.A.m | |
| 8 | 3 | 39 | TE | OP | 1 | 2 | SCI.ESS3.D.m | Earth and Space Science | SCI.SEP1.A.m | SCI.CC7.m |
| 8 | 3 | 40 | TE | OP | 1 | 3 | SCI.PS3.A.m | Physical Science | SCI.SEP2.A.m | |

# Appendix F
# Spring 2022 Social Studies
# Operational Test Maps

# List of Tables

## Table F-1. Social Studies Grade 4 Test Map

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 1 | TE | OP | 1 | 2 | BH.2.b.4 | Behavioral Sciences |
| 4 | 1 | 2 | MC | OP | 1 | 2 | Geog.4.a.4 | Geography |
| 4 | 1 | 3 | TE | OP | 1 | 2 | PS.3.a.4-5 | Political Science |
| 4 | 1 | 4 | MC | OP | 1 | 2 | Hist.2.a.i | History |
| 4 | 1 | 5 | MC | OP | 1 | 2 | BH.4.a.i | Behavioral Sciences |
| 4 | 1 | 6 | MS | OP | 1 | 2 | BH.2.b.4 | Behavioral Sciences |
| 4 | 1 | 7 | MC | OP | 1 | 2 | Econ.1.b.4 | Economics |
| 4 | 1 | 8 | MC | OP | 1 | 2 | PS.2.a.i | Political Science |
| 4 | 1 | 9 | TE | OP | 1 | 2 | BH.4.a.i | Behavioral Sciences |
| 4 | 1 | 10 | MC | OP | 1 | 3 | Econ.1.a.3 | Economics |
| 4 | 1 | 11 | MC | OP | 1 | 3 | Econ.3.a.4 | Economics |
| 4 | 1 | 12 | MS | OP | 1 | 3 | BH.2.a.4-5 | Behavioral Sciences |
| 4 | 1 | 13 | MC | OP | 1 | 3 | BH.2.a.4-5 | Behavioral Sciences |
| 4 | 1 | 14 | TE | OP | 1 | 1 | Geog.1.c.4-5 | Geography |
| 4 | 1 | 15 | TE | OP | 1 | 2 | Econ.4.a.3 | Economics |
| 4 | 1 | 16 | MC | OP | 1 | 2 | BH.2.b.4 | Behavioral Sciences |
| 4 | 1 | 17 | TE | OP | 1 | 2 | PS.3.b.3-4 | Political Science |
| 4 | 1 | 18 | MC | OP | 1 | 2 | Geog.5.a.3-4 | Geography |
| 4 | 1 | 19 | MC | OP | 1 | 2 | Econ.2.a.3-4 | Economics |
| 4 | 1 | 20 | MC | OP | 1 | 2 | Geog.4.a.4 | Geography |
| 4 | 2 | 21 | MC | OP | 1 | 3 | Hist.4.c.i | History |
| 4 | 2 | 22 | MC | OP | 1 | 3 | Hist.4.d.i | History |
| 4 | 2 | 23 | TE | OP | 1 | 3 | Geog.3.b.4 | Geography |
| 4 | 2 | 24 | MC | OP | 1 | 2 | Hist.4.c.i | History |
| 4 | 2 | 25 | TE | OP | 1 | 2 | Geog.2.d.4-5 | Geography |
| 4 | 2 | 26 | TE | OP | 1 | 2 | PS.2.a.i | Political Science |
| 4 | 2 | 27 | TE | OP | 1 | 2 | Geog.4.a.4 | Geography |
| 4 | 2 | 28 | MC | OP | 1 | 2 | Econ.1.a.3 | Economics |
| 4 | 2 | 29 | TE | OP | 1 | 2 | PS.1.a.i | Political Science |
| 4 | 2 | 30 | TE | OP | 1 | 3 | Hist.4.a.i | History |
| 4 | 2 | 31 | MC | OP | 1 | 2 | Geog.2.a.3 | Geography |
| 4 | 2 | 32 | MC | OP | 1 | 2 | Geog.1.b.i | Geography |
| 4 | 2 | 33 | TE | OP | 1 | 1 | Geog.1.c.4-5 | Geography |

**Table F-1 (continued). Social Studies Grade 4 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 34 | MC | OP | 1 | 2 | Geog.1.b.i | Geography |
| 4 | 2 | 35 | MC | OP | 1 | 2 | Hist.2.c.i | History |
| 4 | 2 | 36 | MC | OP | 1 | 3 | PS.2.a.i | Political Science |
| 4 | 2 | 37 | MC | OP | 1 | 3 | Hist.4.d.i | History |
| 4 | 2 | 38 | TE | OP | 1 | 2 | PS.1.b.4-5 | Political Science |
| 4 | 2 | 39 | MC | OP | 1 | 2 | Econ.1.b.4 | Economics |
| 4 | 2 | 40 | MC | OP | 1 | 2 | Hist.4.c.i | History |

**Table F-2. Social Studies Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 1 | MC | OP | 1 | 3 | Geog.3.a.m | Geography |
| 8 | 1 | 2 | MC | OP | 1 | 3 | Geog.2.c.m | Geography |
| 8 | 1 | 3 | MC | OP | 1 | 3 | PS.1.a.m | Political Science |
| 8 | 1 | 4 | MS | OP | 1 | 3 | Hist.1.a.m | History |
| 8 | 1 | 5 | MC | OP | 1 | 3 | PS.3.c.m | Political Science |
| 8 | 1 | 6 | MC | OP | 1 | 3 | Hist.4.a.m | History |
| 8 | 1 | 7 | MC | OP | 1 | 2 | Hist.2.c.m | History |
| 8 | 1 | 8 | MC | OP | 1 | 2 | BH.4.a.m | Behavioral Sciences |
| 8 | 1 | 9 | MC | OP | 1 | 2 | BH.2.a.m | Behavioral Sciences |
| 8 | 1 | 10 | MC | OP | 1 | 2 | BH.3.a.m | Behavioral Sciences |
| 8 | 1 | 11 | MC | OP | 1 | 3 | Geog.1.b.m | Geography |
| 8 | 1 | 12 | MC | OP | 1 | 2 | PS.2.c.m | Political Science |
| 8 | 1 | 13 | MC | OP | 1 | 3 | Econ.2.b.m | Economics |
| 8 | 1 | 14 | MC | OP | 1 | 2 | Geog.5.b.m | Geography |
| 8 | 1 | 15 | MC | OP | 1 | 2 | Econ.2.c.m | Economics |
| 8 | 1 | 16 | MC | OP | 1 | 2 | Geog.2.d.m | Geography |
| 8 | 1 | 17 | MC | OP | 1 | 3 | PS.3.a.m | Political Science |
| 8 | 1 | 18 | MC | OP | 1 | 2 | Hist.2.a.m | History |
| 8 | 1 | 19 | TE | OP | 1 | 3 | Geog.4.a.m | Geography |
| 8 | 1 | 20 | MC | OP | 1 | 3 | Econ.4.a.m | Economics |
| 8 | 2 | 21 | MC | OP | 1 | 2 | PS.1.a.m | Political Science |
| 8 | 2 | 22 | MC | OP | 1 | 3 | Econ.2.b.m | Economics |
| 8 | 2 | 23 | MC | OP | 1 | 3 | Hist.3.c.m | History |
| 8 | 2 | 24 | MC | OP | 1 | 2 | Hist.4.c.m | History |
| 8 | 2 | 25 | TE | OP | 1 | 2 | PS.2.b.m | Political Science |
| 8 | 2 | 26 | MC | OP | 1 | 3 | PS.3.a.m | Political Science |
| 8 | 2 | 27 | MC | OP | 1 | 2 | Econ.2.c.m | Economics |
| 8 | 2 | 28 | MC | OP | 1 | 2 | BH.1.b.m | Behavioral Sciences |
| 8 | 2 | 29 | MC | OP | 1 | 3 | Econ.4.c.m | Economics |
| 8 | 2 | 30 | MC | OP | 1 | 3 | BH.3.a.m | Behavioral Sciences |
| 8 | 2 | 31 | MC | OP | 1 | 3 | Econ.3.c.m | Economics |
| 8 | 2 | 32 | TE | OP | 1 | 2 | Econ.4.a.m | Economics |

**Table F-2 (continued). Social Studies Grade 8 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 8 | 2 | 33 | TE | OP | 1 | 2 | BH.1.a.m | Behavioral Sciences |
| 8 | 2 | 34 | TE | OP | 1 | 2 | Geog.5.a.m | Geography |
| 8 | 2 | 35 | MC | OP | 1 | 2 | Geog.1.a.m | Geography |
| 8 | 2 | 36 | MC | OP | 1 | 3 | Hist.2.b.m | History |
| 8 | 2 | 37 | MC | OP | 1 | 2 | Hist.4.d.m | History |
| 8 | 2 | 38 | MC | OP | 1 | 3 | Hist.4.b.m | History |
| 8 | 2 | 39 | MC | OP | 1 | 2 | Geog.2.b.m | Geography |
| 8 | 2 | 40 | TE | OP | 1 | 1 | PS.2.a.m | Political Science |

**Table F-3. Social Studies Grade 10 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|---|---|---|---|---|---|---|---|---|
| 10 | 1 | 1 | TE | OP | 1 | 2 | PS.2.b.h | Political Science |
| 10 | 1 | 2 | MC | OP | 1 | 2 | Geog.2.a.h | Geography |
| 10 | 1 | 3 | MC | OP | 1 | 2 | Geog.1.a.h | Geography |
| 10 | 1 | 4 | MC | OP | 1 | 2 | Geog.5.b.h | Geography |
| 10 | 1 | 5 | MC | OP | 1 | 3 | Hist.4.c.h | History |
| 10 | 1 | 6 | MC | OP | 1 | 2 | Geog.5.a.h | Geography |
| 10 | 1 | 7 | MC | OP | 1 | 2 | Econ.4.e.h | Economics |
| 10 | 1 | 8 | MC | OP | 1 | 3 | Econ.1.b.h | Economics |
| 10 | 1 | 9 | MC | OP | 1 | 3 | Hist.3.a.h | History |
| 10 | 1 | 10 | MC | OP | 1 | 2 | Hist.3.c.h | History |
| 10 | 1 | 11 | MC | OP | 1 | 2 | PS.2.a.h | Political Science |
| 10 | 1 | 12 | MC | OP | 1 | 2 | PS.1.b.h | Political Science |
| 10 | 1 | 13 | MC | OP | 1 | 2 | Econ.4.b.h | Economics |
| 10 | 1 | 14 | MC | OP | 1 | 3 | Hist.3.c.h | History |
| 10 | 1 | 15 | MC | OP | 1 | 2 | Geog.4.a.h | Geography |
| 10 | 1 | 16 | MC | OP | 1 | 2 | PS.2.a.h | Political Science |
| 10 | 1 | 17 | MC | OP | 1 | 2 | BH.2.b.h | Behavioral Sciences |
| 10 | 1 | 18 | MC | OP | 1 | 2 | PS.1.b.h | Political Science |
| 10 | 1 | 19 | MC | OP | 1 | 2 | Geog.3.b.h | Geography |
| 10 | 1 | 20 | MS | OP | 1 | 2 | BH.2.a.h | Behavioral Sciences |
| 10 | 2 | 21 | MC | OP | 1 | 3 | Hist.2.b.h | History |
| 10 | 2 | 22 | MS | OP | 1 | 2 | BH.1.a.h | Behavioral Sciences |
| 10 | 2 | 23 | MC | OP | 1 | 2 | Hist.2.c.h | History |
| 10 | 2 | 24 | MC | OP | 1 | 2 | Econ.3.b.h | Economics |
| 10 | 2 | 25 | ESR | OP | 1 | 3 | Hist.1.b.h | History |
| 10 | 2 | 26 | TE | OP | 1 | 3 | Econ.2.c.h | Economics |
| 10 | 2 | 27 | MC | OP | 1 | 2 | Econ.2.c.h | Economics |
| 10 | 2 | 28 | EBSR | OP | 1 | 3 | Hist.3.b.h | History |
| 10 | 2 | 29 | MC | OP | 1 | 2 | PS.1.b.h | Political Science |
| 10 | 2 | 30 | MC | OP | 1 | 3 | BH.2.a.h | Behavioral Sciences |
| 10 | 2 | 31 | TE | OP | 1 | 2 | Geog.3.b.h | Geography |
| 10 | 2 | 32 | MC | OP | 1 | 2 | Geog.4.a.h | Geography |
| 10 | 2 | 33 | MC | OP | 1 | 2 | Geog.1.c.h | Geography |
| 10 | 2 | 34 | TE | OP | 1 | 2 | Geog.2.c.h | Geography |
| 10 | 2 | 35 | MC | OP | 1 | 2 | Hist.4.d.h | History |

**Table F-3 (continued). Social Studies Grade 10 Test Map**

| Grade | Session | Item Sequence | Item Type | Item Usage | Max Score Points | Depth of Knowledge | Standard | Domain |
|-------|---------|---------------|-----------|------------|------------------|--------------------|----------|--------|
| 10 | 2 | 36 | TE | OP | 1 | 2 | PS.2.a.h | Political Science |
| 10 | 2 | 37 | MC | OP | 1 | 2 | BH.2.a.h | Behavioral Sciences |
| 10 | 2 | 38 | MC | OP | 1 | 2 | BH.1.a.h | Behavioral Sciences |
| 10 | 2 | 39 | MC | OP | 1 | 2 | PS.2.c.h | Political Science |
| 10 | 2 | 40 | TE | OP | 1 | 2 | PS.3.a.h | Political Science |

# Appendix G
# Spring 2022 Test Participation
# Rates by Subgroup

# List of Tables

**Table G-1. Test Participation Rates, Grade 3**

| Group | Category | Enrolled | English Language Arts | | Mathematics | |
|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 60956 | 58275 | 95.60 | 58449 | 95.89 |
| Gender | Male | 31115 | 29613 | 95.17 | 29706 | 95.47 |
| | Female | 29841 | 28662 | 96.05 | 28743 | 96.32 |
| Race/Ethnicity | White | 39550 | 38091 | 96.31 | 38088 | 96.30 |
| | African American | 6053 | 5679 | 93.82 | 5675 | 93.76 |
| | Hispanic | 8521 | 7998 | 93.86 | 8152 | 95.67 |
| | Asian | 2860 | 2744 | 95.94 | 2770 | 96.85 |
| | American Indian | 646 | 599 | 92.72 | 597 | 92.41 |
| | Two or More | 3326 | 3164 | 95.13 | 3167 | 95.22 |
| Limited English Proficiency | No | 55533 | 53189 | 95.78 | 53198 | 95.80 |
| | Yes | 5423 | 5086 | 93.79 | 5251 | 96.83 |
| Disability Status | No | 52153 | 50372 | 96.59 | 50553 | 96.93 |
| | Yes | 8803 | 7903 | 89.78 | 7896 | 89.70 |
| Economically Disadvantaged | No | 35209 | 33821 | 96.06 | 33939 | 96.39 |
| | Yes | 25747 | 24454 | 94.98 | 24510 | 95.20 |

**Table G-2. Test Participation Rates, Grade 4**

| Group | Category | Enrolled | English Language Arts | | Mathematics | | Science | | Social Studies | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 61465 | 58773 | 95.62 | 58931 | 95.88 | 58880 | 95.79 | 58833 | 95.72 |
| Gender | Male | 31558 | 30084 | 95.33 | 30161 | 95.57 | 30129 | 95.47 | 30104 | 95.39 |
| | Female | 29907 | 28689 | 95.93 | 28770 | 96.20 | 28751 | 96.13 | 28729 | 96.06 |
| Race/Ethnicity | White | 39960 | 38419 | 96.14 | 38425 | 96.16 | 38422 | 96.15 | 38403 | 96.10 |
| | African American | 6001 | 5604 | 93.38 | 5606 | 93.42 | 5577 | 92.93 | 5566 | 92.75 |
| | Hispanic | 8792 | 8305 | 94.46 | 8432 | 95.91 | 8422 | 95.79 | 8407 | 95.62 |
| | Asian | 2819 | 2706 | 95.99 | 2725 | 96.67 | 2721 | 96.52 | 2721 | 96.52 |
| | American Indian | 625 | 593 | 94.88 | 591 | 94.56 | 590 | 94.40 | 593 | 94.88 |
| | Two or More | 3268 | 3146 | 96.27 | 3152 | 96.45 | 3148 | 96.33 | 3143 | 96.18 |
| Limited English Proficiency | No | 55701 | 53390 | 95.85 | 53396 | 95.86 | 53359 | 95.80 | 53317 | 95.72 |
| | Yes | 5764 | 5383 | 93.39 | 5535 | 96.03 | 5521 | 95.78 | 5516 | 95.70 |
| Disability Status | No | 52699 | 50901 | 96.59 | 51067 | 96.90 | 51016 | 96.81 | 50983 | 96.74 |
| | Yes | 8766 | 7872 | 89.80 | 7864 | 89.71 | 7864 | 89.71 | 7850 | 89.55 |
| Economically Disadvantaged | No | 35488 | 34131 | 96.18 | 34226 | 96.44 | 34215 | 96.41 | 34201 | 96.37 |
| | Yes | 25977 | 24642 | 94.86 | 24705 | 95.10 | 24665 | 94.95 | 24632 | 94.82 |

**Table G-3. Test Participation Rates, Grade 5**

| Group | Category | Enrolled | English Language Arts | | Mathematics | |
|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 61735 | 59044 | 95.64 | 59187 | 95.87 |
| Gender | Male | 31576 | 30093 | 95.30 | 30171 | 95.55 |
| | Female | 30159 | 28951 | 95.99 | 29016 | 96.21 |
| Race/Ethnicity | White | 40292 | 38782 | 96.25 | 38776 | 96.24 |
| | African American | 6170 | 5774 | 93.58 | 5776 | 93.61 |
| | Hispanic | 8824 | 8320 | 94.29 | 8446 | 95.72 |
| | Asian | 2686 | 2590 | 96.43 | 2612 | 97.24 |
| | American Indian | 676 | 634 | 93.79 | 634 | 93.79 |
| | Two or More | 3087 | 2944 | 95.37 | 2943 | 95.34 |
| Limited English Proficiency | No | 56611 | 54257 | 95.84 | 54257 | 95.84 |
| | Yes | 5124 | 4787 | 93.42 | 4930 | 96.21 |
| Disability Status | No | 53057 | 51294 | 96.68 | 51441 | 96.95 |
| | Yes | 8678 | 7750 | 89.31 | 7746 | 89.26 |
| Economically Disadvantaged | No | 36093 | 34743 | 96.26 | 34842 | 96.53 |
| | Yes | 25642 | 24301 | 94.77 | 24345 | 94.94 |

**Table G-4. Test Participation Rates, Grade 6**

| Group | Category | Enrolled | English Language Arts | | Mathematics | |
|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 63048 | 60112 | 95.34 | 60234 | 95.54 |
| Gender | Male | 32243 | 30649 | 95.06 | 30732 | 95.31 |
| | Female | 30805 | 29463 | 95.64 | 29502 | 95.77 |
| Race/Ethnicity | White | 41457 | 39875 | 96.18 | 39861 | 96.15 |
| | African American | 6393 | 5873 | 91.87 | 5874 | 91.88 |
| | Hispanic | 8975 | 8431 | 93.94 | 8545 | 95.21 |
| | Asian | 2569 | 2470 | 96.15 | 2489 | 96.89 |
| | American Indian | 655 | 620 | 94.66 | 618 | 94.35 |
| | Two or More | 2999 | 2843 | 94.80 | 2847 | 94.93 |
| Limited English Proficiency | No | 58705 | 56089 | 95.54 | 56081 | 95.53 |
| | Yes | 4343 | 4023 | 92.63 | 4153 | 95.63 |
| Disability Status | No | 54476 | 52585 | 96.53 | 52710 | 96.76 |
| | Yes | 8572 | 7527 | 87.81 | 7524 | 87.77 |
| Economically Disadvantaged | No | 37070 | 35632 | 96.12 | 35702 | 96.31 |
| | Yes | 25978 | 24480 | 94.23 | 24532 | 94.43 |

**Table G-5. Test Participation Rates, Grade 7**

| Group | Category | Enrolled | English Language Arts | | Mathematics | |
|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 65386 | 61871 | 94.62 | 61969 | 94.77 |
| Gender | Male | 33554 | 31682 | 94.42 | 31740 | 94.59 |
| | Female | 31832 | 30189 | 94.84 | 30229 | 94.96 |
| Race/Ethnicity | White | 42791 | 40934 | 95.66 | 40900 | 95.58 |
| | African American | 6717 | 6072 | 90.40 | 6067 | 90.32 |
| | Hispanic | 9326 | 8702 | 93.31 | 8832 | 94.70 |
| | Asian | 2782 | 2671 | 96.01 | 2687 | 96.59 |
| | American Indian | 681 | 623 | 91.48 | 620 | 91.04 |
| | Two or More | 3089 | 2869 | 92.88 | 2863 | 92.68 |
| Limited English Proficiency | No | 61152 | 57982 | 94.82 | 57942 | 94.75 |
| | Yes | 4234 | 3889 | 91.85 | 4027 | 95.11 |
| Disability Status | No | 56616 | 54280 | 95.87 | 54389 | 96.07 |
| | Yes | 8770 | 7591 | 86.56 | 7580 | 86.43 |
| Economically Disadvantaged | No | 38738 | 37042 | 95.62 | 37112 | 95.80 |
| | Yes | 26648 | 24829 | 93.17 | 24857 | 93.28 |

**Table G-6. Test Participation Rates, Grade 8**

| Group | Category | Enrolled | English Language Arts | | Mathematics | | Science | | Social Studies | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested | Number Tested | Percent Tested |
| State | All Students | 66789 | 62684 | 93.85 | 62762 | 93.97 | 62644 | 93.79 | 62606 | 93.74 |
| Gender | Male | 34164 | 32133 | 94.06 | 32172 | 94.17 | 32104 | 93.97 | 32088 | 93.92 |
| | Female | 32625 | 30551 | 93.64 | 30590 | 93.76 | 30540 | 93.61 | 30518 | 93.54 |
| Race/Ethnicity | White | 44041 | 41778 | 94.86 | 41754 | 94.81 | 41734 | 94.76 | 41752 | 94.80 |
| | African American | 6848 | 6135 | 89.59 | 6139 | 89.65 | 6068 | 88.61 | 6027 | 88.01 |
| | Hispanic | 9588 | 8846 | 92.26 | 8960 | 93.45 | 8942 | 93.26 | 8930 | 93.14 |
| | Asian | 2628 | 2540 | 96.65 | 2539 | 96.61 | 2540 | 96.65 | 2539 | 96.61 |
| | American Indian | 647 | 590 | 91.19 | 584 | 90.26 | 580 | 89.64 | 577 | 89.18 |
| | Two or More | 3037 | 2795 | 92.03 | 2786 | 91.74 | 2780 | 91.54 | 2781 | 91.57 |
| Limited English Proficiency | No | 62556 | 58797 | 93.99 | 58777 | 93.96 | 58668 | 93.78 | 58629 | 93.72 |
| | Yes | 4233 | 3887 | 91.83 | 3985 | 94.14 | 3976 | 93.93 | 3977 | 93.95 |
| Disability Status | No | 58317 | 55531 | 95.22 | 55616 | 95.37 | 55524 | 95.21 | 55493 | 95.16 |
| | Yes | 8472 | 7153 | 84.43 | 7146 | 84.35 | 7120 | 84.04 | 7113 | 83.96 |
| Economically Disadvantaged | No | 39852 | 37833 | 94.93 | 37881 | 95.05 | 37845 | 94.96 | 37851 | 94.98 |
| | Yes | 26937 | 24851 | 92.26 | 24881 | 92.37 | 24799 | 92.06 | 24755 | 91.90 |

**Table G-7. Test Participation Rates, Grade 10**

| Group | Category | Enrolled | Social Studies | |
|---|---|---|---|---|
| | | | Number Tested | Percent Tested |
| State | All Students | 67427 | 59391 | 88.08 |
| Gender | Male | 34566 | 30546 | 88.37 |
| | Female | 32861 | 28845 | 87.78 |
| Race/Ethnicity | White | 45824 | 41976 | 91.60 |
| | African American | 6174 | 4203 | 68.08 |
| | Hispanic | 9344 | 7932 | 84.89 |
| | Asian | 2641 | 2458 | 93.07 |
| | American Indian | 740 | 570 | 77.03 |
| | Two or More | 2704 | 2252 | 83.28 |
| Limited English Proficiency | No | 64034 | 56567 | 88.34 |
| | Yes | 3393 | 2824 | 83.23 |
| Disability Status | No | 59075 | 53228 | 90.10 |
| | Yes | 8352 | 6163 | 73.79 |
| Economically Disadvantaged | No | 42633 | 39119 | 91.76 |
| | Yes | 24794 | 20272 | 81.76 |

# Appendix H
# Classical Item Analysis Results

# List of Tables

**Explanation of Data Columns in Tables H-1 through H-17**

| Column Number | Data Description |
|---|---|
| 1 | Item Number |
| 2 | Item Type |
| 3 | Maximum Points |
| 4 | Number of Students |
| 5 | Item *p*-value |
| 6 | Item-Total Test Correlation |
| 7 | Proportion Omit |
| 8 | Proportion of Students at Score Point 0 |
| 9 | Proportion of Students at Score Point 1 or Option 1 |
| 10 | Proportion of Students at Score Point 2 or Option 2 |
| 11 | Proportion of Students at Score Point 3 or Option 3 |
| 12 | Proportion of Students at Score Point 4 or Option 4 |
| 13 | Item-Total Test Correlation for Score Point 0 |
| 14 | Item-Total Test Correlation for Score Point 1 or Option 1 |
| 15 | Item-Total Test Correlation for Score Point 2 or Option 2 |
| 16 | Item-Total Test Correlation for Score Point 3 or Option 3 |
| 17 | Item-Total Test Correlation for Score Point 4 or Option 4 |

## Table H-1. Item Statistics, English Language Arts Grade 3

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 52595 | 0.33 | 0.38 | 0.01 | | 0.66 | 0.23 | 0.02 | 0.00 | | -0.16 | 0.35 | 0.17 | 0.04 |
| 2 | MC | 1 | 57570 | 0.79 | 0.42 | 0.00 | | 0.11 | 0.79 | 0.06 | 0.04 | | -0.28 | 0.42 | -0.20 | -0.17 |
| 3 | MC | 1 | 57653 | 0.63 | 0.23 | 0.00 | | 0.62 | 0.16 | 0.08 | 0.13 | | 0.23 | -0.03 | -0.19 | -0.14 |
| 4 | TE | 2 | 57650 | 0.70 | 0.53 | 0.00 | 0.12 | 0.38 | 0.51 | | | -0.36 | -0.27 | 0.49 | | |
| 5 | MC | 1 | 57661 | 0.67 | 0.20 | 0.00 | | 0.13 | 0.67 | 0.05 | 0.15 | | -0.24 | 0.20 | -0.12 | 0.04 |
| 6 | TE | 2 | 57546 | 0.55 | 0.45 | 0.00 | 0.22 | 0.45 | 0.32 | | | -0.30 | -0.14 | 0.43 | | |
| 7 | MC | 1 | 57647 | 0.72 | 0.39 | 0.00 | | 0.07 | 0.09 | 0.13 | 0.72 | | -0.29 | -0.16 | -0.16 | 0.39 |
| 8 | MC | 1 | 57634 | 0.73 | 0.48 | 0.00 | | 0.15 | 0.04 | 0.08 | 0.73 | | -0.28 | -0.23 | -0.24 | 0.48 |
| 9 | MC | 1 | 57648 | 0.60 | 0.38 | 0.00 | | 0.21 | 0.60 | 0.10 | 0.09 | | -0.22 | 0.39 | -0.19 | -0.14 |
| 10 | MC | 1 | 57640 | 0.44 | 0.28 | 0.00 | | 0.44 | 0.19 | 0.21 | 0.16 | | 0.28 | -0.14 | -0.06 | -0.16 |
| 11 | TE | 2 | 57609 | 0.45 | 0.46 | 0.00 | 0.28 | 0.52 | 0.19 | | | -0.39 | 0.09 | 0.34 | | |
| 12 | MC | 1 | 57634 | 0.42 | 0.36 | 0.00 | | 0.20 | 0.24 | 0.42 | 0.15 | | -0.11 | -0.20 | 0.36 | -0.13 |
| 13 | MC | 1 | 57626 | 0.51 | 0.30 | 0.00 | | 0.14 | 0.51 | 0.17 | 0.18 | | -0.22 | 0.30 | -0.11 | -0.08 |
| 14 | TE | 1 | 57123 | 0.40 | 0.22 | 0.01 | 0.59 | 0.40 | | | | -0.20 | 0.23 | | | |
| 15 | EBSR | 2 | 57717 | 0.53 | 0.34 | 0.00 | 0.29 | 0.37 | 0.34 | | | -0.25 | -0.08 | 0.33 | | |
| 16 | MC | 1 | 57595 | 0.57 | 0.35 | 0.00 | | 0.22 | 0.04 | 0.17 | 0.57 | | -0.16 | -0.21 | -0.18 | 0.36 |
| 17 | MC | 1 | 57635 | 0.44 | 0.27 | 0.00 | | 0.19 | 0.25 | 0.11 | 0.44 | | -0.23 | 0.03 | -0.19 | 0.28 |
| 18 | MC | 1 | 57624 | 0.53 | 0.37 | 0.00 | | 0.27 | 0.10 | 0.09 | 0.53 | | -0.13 | -0.25 | -0.17 | 0.37 |
| 19 | MS | 2 | 57636 | 0.68 | 0.53 | 0.00 | 0.12 | 0.41 | 0.47 | | | -0.36 | -0.26 | 0.49 | | |
| 20 | MC | 1 | 57651 | 0.52 | 0.20 | 0.00 | | 0.24 | 0.04 | 0.21 | 0.51 | | -0.06 | -0.21 | -0.08 | 0.20 |
| 21 | MC | 1 | 57560 | 0.63 | 0.41 | 0.00 | | 0.23 | 0.63 | 0.09 | 0.05 | | -0.24 | 0.41 | -0.16 | -0.22 |
| 22 | MC | 1 | 57553 | 0.42 | 0.29 | 0.00 | | 0.20 | 0.42 | 0.14 | 0.23 | | -0.05 | 0.29 | -0.18 | -0.14 |
| 23 | EBSR | 2 | 57663 | 0.54 | 0.51 | 0.00 | 0.31 | 0.30 | 0.39 | | | -0.43 | -0.04 | 0.46 | | |
| 24 | MC | 1 | 57540 | 0.50 | 0.21 | 0.00 | | 0.50 | 0.15 | 0.21 | 0.13 | | 0.22 | 0.06 | -0.07 | -0.29 |
| 25 | TE | 2 | 57518 | 0.66 | 0.65 | 0.00 | 0.23 | 0.21 | 0.55 | | | -0.51 | -0.25 | 0.65 | | |
| 26 | MC | 1 | 57146 | 0.55 | 0.44 | 0.01 | | 0.54 | 0.20 | 0.11 | 0.13 | | 0.44 | -0.23 | -0.19 | -0.18 |
| 27 | MC | 1 | 57200 | 0.77 | 0.47 | 0.01 | | 0.09 | 0.05 | 0.09 | 0.77 | | -0.25 | -0.25 | -0.24 | 0.47 |
| 28 | MC | 1 | 57585 | 0.50 | 0.22 | 0.00 | | 0.11 | 0.28 | 0.50 | 0.12 | | -0.21 | 0.11 | 0.22 | -0.29 |
| 29 | EBSR | 2 | 57637 | 0.49 | 0.55 | 0.00 | 0.42 | 0.17 | 0.41 | | | -0.46 | -0.10 | 0.55 | | |
| 30 | MC | 1 | 57568 | 0.70 | 0.50 | 0.00 | | 0.69 | 0.07 | 0.10 | 0.13 | | 0.51 | -0.25 | -0.27 | -0.24 |
| 31 | MC | 1 | 57553 | 0.54 | 0.35 | 0.00 | | 0.13 | 0.17 | 0.16 | 0.54 | | -0.23 | -0.11 | -0.15 | 0.35 |

**Table H-1 (continued). Item Statistics, English Language Arts Grade 3**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 57460 | 0.42 | 0.33 | 0.01 | | 0.18 | 0.42 | 0.24 | 0.15 | | -0.17 | 0.33 | -0.04 | -0.21 |
| 33 | MC | 1 | 57483 | 0.43 | 0.37 | 0.00 | | 0.20 | 0.13 | 0.43 | 0.23 | | -0.06 | -0.17 | 0.37 | -0.23 |
| 34 | MC | 1 | 57516 | 0.57 | 0.53 | 0.00 | | 0.57 | 0.18 | 0.16 | 0.09 | | 0.53 | -0.19 | -0.28 | -0.28 |
| 35 | EBSR | 2 | 57573 | 0.43 | 0.54 | 0.00 | 0.51 | 0.11 | 0.37 | | | -0.47 | -0.09 | 0.55 | | |
| 36 | MC | 1 | 57344 | 0.48 | 0.30 | 0.01 | | 0.15 | 0.12 | 0.24 | 0.48 | | -0.10 | -0.15 | -0.14 | 0.30 |
| 37 | MC | 1 | 57157 | 0.42 | 0.31 | 0.01 | | 0.17 | 0.42 | 0.09 | 0.31 | | -0.09 | 0.32 | -0.24 | -0.10 |

# Table H-2. Item Statistics, English Language Arts Grade 4

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 55251 | 0.35 | 0.41 | 0.00 | | 0.62 | 0.27 | 0.06 | 0.00 | | -0.28 | 0.32 | 0.23 | 0.05 |
| 2 | MC | 1 | 58182 | 0.47 | 0.35 | 0.00 | | 0.29 | 0.47 | 0.09 | 0.15 | | -0.18 | 0.36 | -0.21 | -0.10 |
| 3 | TE | 1 | 58124 | 0.60 | 0.42 | 0.00 | 0.40 | 0.60 | | | | -0.41 | 0.42 | | | |
| 4 | MC | 1 | 58162 | 0.83 | 0.44 | 0.00 | | 0.04 | 0.08 | 0.05 | 0.83 | | -0.24 | -0.22 | -0.27 | 0.45 |
| 5 | TE | 1 | 58056 | 0.51 | 0.47 | 0.00 | 0.49 | 0.51 | | | | -0.47 | 0.47 | | | |
| 6 | TE | 2 | 58105 | 0.48 | 0.42 | 0.00 | 0.21 | 0.60 | 0.18 | | | -0.35 | 0.05 | 0.31 | | |
| 7 | MC | 1 | 58105 | 0.52 | 0.44 | 0.00 | | 0.15 | 0.24 | 0.52 | 0.09 | | -0.23 | -0.20 | 0.44 | -0.17 |
| 8 | MC | 1 | 58124 | 0.56 | 0.38 | 0.00 | | 0.23 | 0.05 | 0.15 | 0.56 | | -0.08 | -0.25 | -0.27 | 0.38 |
| 9 | MC | 1 | 58138 | 0.56 | 0.41 | 0.00 | | 0.25 | 0.11 | 0.08 | 0.56 | | -0.11 | -0.29 | -0.25 | 0.41 |
| 10 | MC | 1 | 58132 | 0.42 | 0.25 | 0.00 | | 0.13 | 0.42 | 0.19 | 0.26 | | -0.15 | 0.25 | -0.11 | -0.07 |
| 11 | TE | 2 | 57830 | 0.68 | 0.41 | 0.01 | 0.06 | 0.52 | 0.41 | | | -0.17 | -0.31 | 0.41 | | |
| 12 | MC | 1 | 58122 | 0.65 | 0.50 | 0.00 | | 0.65 | 0.11 | 0.11 | 0.12 | | 0.50 | -0.29 | -0.22 | -0.24 |
| 13 | TE | 2 | 58124 | 0.63 | 0.29 | 0.00 | 0.03 | 0.69 | 0.29 | | | -0.16 | -0.20 | 0.26 | | |
| 14 | MC | 1 | 58123 | 0.42 | 0.23 | 0.00 | | 0.17 | 0.41 | 0.18 | 0.24 | | -0.11 | 0.23 | -0.12 | -0.06 |
| 15 | MC | 1 | 58185 | 0.56 | 0.28 | 0.00 | | 0.11 | 0.20 | 0.56 | 0.13 | | -0.13 | -0.08 | 0.28 | -0.20 |
| 16 | MC | 1 | 58134 | 0.74 | 0.37 | 0.00 | | 0.02 | 0.20 | 0.04 | 0.74 | | -0.19 | -0.23 | -0.22 | 0.37 |
| 17 | EBSR | 2 | 58190 | 0.32 | 0.34 | 0.00 | 0.56 | 0.23 | 0.20 | | | -0.30 | 0.08 | 0.30 | | |
| 18 | MC | 1 | 58143 | 0.49 | 0.22 | 0.00 | | 0.04 | 0.49 | 0.30 | 0.17 | | -0.18 | 0.22 | -0.09 | -0.08 |
| 19 | MC | 1 | 58127 | 0.53 | 0.33 | 0.00 | | 0.17 | 0.17 | 0.14 | 0.53 | | -0.17 | -0.09 | -0.20 | 0.33 |
| 20 | EBSR | 2 | 58172 | 0.58 | 0.46 | 0.00 | 0.31 | 0.22 | 0.46 | | | -0.35 | -0.18 | 0.48 | | |
| 21 | MC | 1 | 58142 | 0.72 | 0.46 | 0.00 | | 0.09 | 0.72 | 0.09 | 0.10 | | -0.28 | 0.46 | -0.27 | -0.17 |
| 22 | TE | 2 | 57936 | 0.56 | 0.43 | 0.00 | 0.18 | 0.51 | 0.30 | | | -0.35 | -0.04 | 0.35 | | |
| 23 | MC | 1 | 58059 | 0.49 | 0.41 | 0.00 | | 0.49 | 0.29 | 0.11 | 0.11 | | 0.41 | -0.13 | -0.25 | -0.21 |
| 24 | MC | 1 | 58027 | 0.59 | 0.54 | 0.00 | | 0.11 | 0.19 | 0.11 | 0.58 | | -0.24 | -0.22 | -0.31 | 0.54 |
| 25 | MS | 2 | 58075 | 0.46 | 0.46 | 0.00 | 0.33 | 0.42 | 0.25 | | | -0.36 | -0.02 | 0.41 | | |
| 26 | MC | 1 | 58027 | 0.87 | 0.46 | 0.00 | | 0.86 | 0.06 | 0.04 | 0.03 | | 0.47 | -0.30 | -0.25 | -0.21 |
| 27 | MC | 1 | 58080 | 0.78 | 0.45 | 0.00 | | 0.06 | 0.08 | 0.78 | 0.07 | | -0.19 | -0.27 | 0.45 | -0.25 |
| 28 | EBSR | 2 | 58072 | 0.41 | 0.46 | 0.00 | 0.54 | 0.10 | 0.36 | | | -0.38 | -0.16 | 0.50 | | |
| 29 | MC | 1 | 57955 | 0.69 | 0.42 | 0.00 | | 0.15 | 0.09 | 0.07 | 0.68 | | -0.27 | -0.08 | -0.28 | 0.42 |
| 30 | MC | 1 | 58057 | 0.63 | 0.41 | 0.00 | | 0.63 | 0.10 | 0.14 | 0.12 | | 0.41 | -0.25 | -0.11 | -0.24 |
| 31 | MC | 1 | 57746 | 0.49 | 0.44 | 0.01 | | 0.48 | 0.15 | 0.18 | 0.18 | | 0.44 | -0.20 | -0.25 | -0.12 |

**Table H-2 (continued). Item Statistics, English Language Arts Grade 4**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | EBSR | 2 | 58088 | 0.27 | 0.42 | 0.00 | 0.62 | 0.20 | 0.17 | | | -0.33 | 0.00 | 0.43 | | |
| 33 | MC | 1 | 58027 | 0.52 | 0.40 | 0.00 | | 0.52 | 0.18 | 0.17 | 0.12 | | 0.40 | -0.06 | -0.24 | -0.26 |
| 34 | MC | 1 | 58027 | 0.64 | 0.51 | 0.00 | | 0.14 | 0.10 | 0.13 | 0.63 | | -0.18 | -0.28 | -0.30 | 0.52 |
| 35 | MC | 1 | 57998 | 0.53 | 0.40 | 0.00 | | 0.53 | 0.23 | 0.13 | 0.11 | | 0.40 | -0.19 | -0.23 | -0.13 |
| 36 | MC | 1 | 58021 | 0.64 | 0.40 | 0.00 | | 0.20 | 0.64 | 0.10 | 0.06 | | -0.15 | 0.40 | -0.25 | -0.23 |
| 37 | MC | 1 | 58039 | 0.56 | 0.43 | 0.00 | | 0.14 | 0.19 | 0.11 | 0.56 | | -0.20 | -0.17 | -0.24 | 0.43 |
| 38 | TE | 2 | 58014 | 0.46 | 0.45 | 0.00 | 0.25 | 0.58 | 0.17 | | | -0.37 | 0.07 | 0.35 | | |
| 39 | MC | 1 | 57874 | 0.59 | 0.44 | 0.01 | | 0.08 | 0.14 | 0.58 | 0.19 | | -0.21 | -0.16 | 0.45 | -0.26 |

# Table H-3. Item Statistics, English Language Arts Grade 5

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 56362 | 0.36 | 0.48 | 0.00 | | 0.60 | 0.30 | 0.07 | 0.00 | | -0.38 | 0.34 | 0.29 | 0.05 |
| 2 | MC | 1 | 58501 | 0.78 | 0.42 | 0.00 | | 0.14 | 0.78 | 0.04 | 0.03 | | -0.33 | 0.43 | -0.17 | -0.14 |
| 3 | MC | 1 | 58470 | 0.29 | 0.22 | 0.00 | | 0.27 | 0.29 | 0.14 | 0.30 | | -0.12 | 0.22 | -0.16 | 0.02 |
| 4 | TE | 2 | 58476 | 0.58 | 0.47 | 0.00 | 0.18 | 0.47 | 0.35 | | | -0.34 | -0.14 | 0.42 | | |
| 5 | MC | 1 | 58464 | 0.41 | 0.28 | 0.00 | | 0.41 | 0.11 | 0.26 | 0.23 | | 0.28 | -0.16 | -0.05 | -0.15 |
| 6 | MC | 1 | 58463 | 0.62 | 0.44 | 0.00 | | 0.23 | 0.07 | 0.62 | 0.08 | | -0.21 | -0.28 | 0.44 | -0.19 |
| 7 | TE | 2 | 58468 | 0.74 | 0.36 | 0.00 | 0.04 | 0.43 | 0.52 | | | -0.24 | -0.22 | 0.32 | | |
| 8 | TE | 1 | 58366 | 0.70 | 0.53 | 0.00 | 0.30 | 0.70 | | | | -0.53 | 0.53 | | | |
| 9 | MC | 1 | 58473 | 0.46 | 0.29 | 0.00 | | 0.17 | 0.23 | 0.46 | 0.15 | | -0.23 | -0.05 | 0.29 | -0.10 |
| 10 | MC | 1 | 58453 | 0.52 | 0.35 | 0.00 | | 0.31 | 0.52 | 0.08 | 0.10 | | -0.15 | 0.35 | -0.20 | -0.18 |
| 11 | TE | 2 | 58466 | 0.72 | 0.57 | 0.00 | 0.10 | 0.37 | 0.53 | | | -0.37 | -0.34 | 0.55 | | |
| 12 | MC | 1 | 58477 | 0.76 | 0.46 | 0.00 | | 0.08 | 0.07 | 0.76 | 0.09 | | -0.22 | -0.26 | 0.46 | -0.24 |
| 13 | TE | 2 | 58371 | 0.59 | 0.39 | 0.00 | 0.15 | 0.52 | 0.33 | | | -0.22 | -0.21 | 0.39 | | |
| 14 | MC | 1 | 58456 | 0.89 | 0.38 | 0.00 | | 0.02 | 0.89 | 0.03 | 0.05 | | -0.18 | 0.38 | -0.21 | -0.24 |
| 15 | MC | 1 | 58368 | 0.86 | 0.41 | 0.00 | | 0.05 | 0.05 | 0.86 | 0.04 | | -0.25 | -0.24 | 0.41 | -0.17 |
| 16 | EBSR | 2 | 58469 | 0.60 | 0.34 | 0.00 | 0.22 | 0.35 | 0.43 | | | -0.35 | 0.05 | 0.25 | | |
| 17 | MC | 1 | 58439 | 0.61 | 0.42 | 0.00 | | 0.14 | 0.18 | 0.61 | 0.08 | | -0.31 | -0.11 | 0.42 | -0.22 |
| 18 | MC | 1 | 58396 | 0.58 | 0.38 | 0.00 | | 0.10 | 0.19 | 0.12 | 0.58 | | -0.15 | -0.18 | -0.23 | 0.39 |
| 19 | EBSR | 2 | 58451 | 0.56 | 0.53 | 0.00 | 0.29 | 0.31 | 0.40 | | | -0.46 | -0.04 | 0.46 | | |
| 20 | MC | 1 | 58463 | 0.79 | 0.38 | 0.00 | | 0.09 | 0.79 | 0.09 | 0.04 | | -0.27 | 0.38 | -0.18 | -0.15 |
| 21 | MC | 1 | 58377 | 0.55 | 0.30 | 0.00 | | 0.20 | 0.11 | 0.14 | 0.55 | | -0.09 | -0.12 | -0.21 | 0.30 |
| 22 | EBSR | 2 | 58482 | 0.37 | 0.42 | 0.00 | 0.56 | 0.13 | 0.31 | | | -0.36 | -0.06 | 0.43 | | |
| 23 | MC | 1 | 58364 | 0.61 | 0.40 | 0.00 | | 0.08 | 0.20 | 0.10 | 0.61 | | -0.21 | -0.16 | -0.24 | 0.40 |
| 24 | MC | 1 | 58324 | 0.55 | 0.43 | 0.00 | | 0.17 | 0.55 | 0.18 | 0.10 | | -0.26 | 0.43 | -0.13 | -0.21 |
| 25 | MC | 1 | 58374 | 0.48 | 0.38 | 0.00 | | 0.10 | 0.25 | 0.17 | 0.48 | | -0.13 | -0.12 | -0.26 | 0.38 |
| 26 | MC | 1 | 58210 | 0.34 | 0.21 | 0.01 | | 0.31 | 0.20 | 0.34 | 0.15 | | -0.09 | -0.03 | 0.22 | -0.13 |
| 27 | MC | 1 | 58275 | 0.50 | 0.25 | 0.00 | | 0.27 | 0.50 | 0.10 | 0.12 | | -0.09 | 0.25 | -0.16 | -0.10 |
| 28 | MC | 1 | 58343 | 0.52 | 0.29 | 0.00 | | 0.52 | 0.18 | 0.08 | 0.22 | | 0.30 | -0.19 | -0.21 | -0.03 |
| 29 | MS | 2 | 58383 | 0.56 | 0.53 | 0.00 | 0.22 | 0.46 | 0.33 | | | -0.41 | -0.10 | 0.47 | | |
| 30 | MC | 1 | 58199 | 0.60 | 0.40 | 0.01 | | 0.22 | 0.59 | 0.12 | 0.06 | | -0.11 | 0.41 | -0.29 | -0.21 |
| 31 | MC | 1 | 58213 | 0.54 | 0.50 | 0.01 | | 0.53 | 0.14 | 0.18 | 0.14 | | 0.50 | -0.27 | -0.19 | -0.22 |

**Table H-3 (continued). Item Statistics, English Language Arts Grade 5**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 58153 | 0.46 | 0.45 | 0.01 | | 0.16 | 0.12 | 0.26 | 0.46 | | -0.17 | -0.27 | -0.16 | 0.45 |
| 33 | MC | 1 | 58308 | 0.49 | 0.39 | 0.00 | | 0.12 | 0.49 | 0.16 | 0.23 | | -0.20 | 0.39 | -0.23 | -0.10 |
| 34 | MC | 1 | 58312 | 0.62 | 0.52 | 0.00 | | 0.62 | 0.12 | 0.13 | 0.13 | | 0.52 | -0.26 | -0.26 | -0.24 |
| 35 | MC | 1 | 58292 | 0.53 | 0.35 | 0.00 | | 0.08 | 0.27 | 0.53 | 0.12 | | -0.24 | -0.06 | 0.35 | -0.24 |
| 36 | MC | 1 | 58334 | 0.49 | 0.38 | 0.00 | | 0.17 | 0.49 | 0.18 | 0.16 | | -0.12 | 0.38 | -0.22 | -0.15 |
| 37 | TE | 2 | 58109 | 0.70 | 0.27 | 0.01 | 0.05 | 0.49 | 0.45 | | | -0.22 | -0.11 | 0.22 | | |
| 38 | MC | 1 | 58327 | 0.43 | 0.45 | 0.00 | | 0.16 | 0.27 | 0.13 | 0.43 | | -0.23 | -0.10 | -0.27 | 0.45 |
| 39 | MC | 1 | 58319 | 0.47 | 0.33 | 0.00 | | 0.18 | 0.15 | 0.47 | 0.19 | | -0.12 | -0.14 | 0.33 | -0.16 |
| 40 | MC | 1 | 58198 | 0.56 | 0.45 | 0.01 | | 0.18 | 0.16 | 0.56 | 0.09 | | -0.21 | -0.23 | 0.46 | -0.19 |

## Table H-4. Item Statistics, English Language Arts Grade 6

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 57072 | 0.39 | 0.49 | 0.00 | | 0.51 | 0.35 | 0.08 | 0.01 | | -0.38 | 0.33 | 0.28 | 0.13 |
| 2 | MC | 1 | 59590 | 0.67 | 0.23 | 0.00 | | 0.66 | 0.20 | 0.10 | 0.04 | | 0.23 | -0.09 | -0.14 | -0.15 |
| 3 | MC | 1 | 59556 | 0.42 | 0.35 | 0.00 | | 0.23 | 0.22 | 0.13 | 0.41 | | -0.20 | -0.06 | -0.19 | 0.35 |
| 4 | MC | 1 | 59511 | 0.69 | 0.44 | 0.00 | | 0.03 | 0.17 | 0.69 | 0.11 | | -0.19 | -0.27 | 0.44 | -0.22 |
| 5 | MC | 1 | 59492 | 0.34 | 0.19 | 0.00 | | 0.15 | 0.34 | 0.24 | 0.27 | | -0.10 | 0.19 | -0.16 | 0.03 |
| 6 | TE | 2 | 59488 | 0.27 | 0.27 | 0.00 | 0.55 | 0.35 | 0.09 | | | -0.22 | 0.10 | 0.23 | | |
| 7 | MC | 1 | 59530 | 0.36 | 0.26 | 0.00 | | 0.04 | 0.14 | 0.36 | 0.46 | | -0.20 | -0.20 | 0.26 | -0.03 |
| 8 | MC | 1 | 59541 | 0.41 | 0.31 | 0.00 | | 0.41 | 0.07 | 0.25 | 0.27 | | 0.31 | -0.24 | -0.13 | -0.07 |
| 9 | TE | 2 | 59553 | 0.63 | 0.45 | 0.00 | 0.14 | 0.47 | 0.39 | | | -0.37 | -0.10 | 0.37 | | |
| 10 | TE | 2 | 59397 | 0.64 | 0.35 | 0.00 | 0.06 | 0.59 | 0.34 | | | -0.29 | -0.12 | 0.28 | | |
| 11 | TE | 2 | 59509 | 0.60 | 0.47 | 0.00 | 0.17 | 0.46 | 0.37 | | | -0.39 | -0.07 | 0.38 | | |
| 12 | MC | 1 | 59510 | 0.54 | 0.34 | 0.00 | | 0.05 | 0.24 | 0.53 | 0.17 | | -0.26 | -0.13 | 0.34 | -0.15 |
| 13 | MC | 1 | 59522 | 0.43 | 0.32 | 0.00 | | 0.15 | 0.29 | 0.12 | 0.43 | | -0.11 | -0.10 | -0.22 | 0.32 |
| 14 | MC | 1 | 59561 | 0.70 | 0.42 | 0.00 | | 0.70 | 0.17 | 0.06 | 0.07 | | 0.42 | -0.21 | -0.26 | -0.20 |
| 15 | MC | 1 | 59479 | 0.54 | 0.34 | 0.00 | | 0.06 | 0.54 | 0.26 | 0.14 | | -0.16 | 0.34 | -0.21 | -0.11 |
| 16 | TE | 2 | 59513 | 0.60 | 0.43 | 0.00 | 0.16 | 0.48 | 0.36 | | | -0.33 | -0.12 | 0.37 | | |
| 17 | MC | 1 | 59498 | 0.42 | 0.27 | 0.00 | | 0.43 | 0.08 | 0.07 | 0.42 | | 0.01 | -0.26 | -0.26 | 0.27 |
| 18 | MC | 1 | 59481 | 0.78 | 0.48 | 0.00 | | 0.04 | 0.07 | 0.78 | 0.10 | | -0.19 | -0.28 | 0.48 | -0.29 |
| 19 | EBSR | 2 | 59545 | 0.65 | 0.54 | 0.00 | 0.27 | 0.14 | 0.58 | | | -0.41 | -0.29 | 0.58 | | |
| 20 | MC | 1 | 59507 | 0.64 | 0.45 | 0.00 | | 0.23 | 0.64 | 0.09 | 0.04 | | -0.27 | 0.45 | -0.23 | -0.17 |
| 21 | MC | 1 | 59409 | 0.71 | 0.36 | 0.00 | | 0.71 | 0.17 | 0.06 | 0.06 | | 0.36 | -0.17 | -0.24 | -0.18 |
| 22 | MC | 1 | 59356 | 0.68 | 0.47 | 0.00 | | 0.07 | 0.09 | 0.68 | 0.16 | | -0.23 | -0.30 | 0.47 | -0.20 |
| 23 | MS | 2 | 59375 | 0.64 | 0.50 | 0.00 | 0.13 | 0.46 | 0.41 | | | -0.40 | -0.15 | 0.42 | | |
| 24 | MC | 1 | 59337 | 0.55 | 0.35 | 0.00 | | 0.25 | 0.09 | 0.55 | 0.11 | | -0.15 | -0.20 | 0.35 | -0.15 |
| 25 | MC | 1 | 59349 | 0.79 | 0.53 | 0.00 | | 0.07 | 0.05 | 0.08 | 0.79 | | -0.28 | -0.29 | -0.28 | 0.53 |
| 26 | TE | 2 | 59323 | 0.55 | 0.42 | 0.00 | 0.18 | 0.54 | 0.27 | | | -0.29 | -0.09 | 0.37 | | |
| 27 | MC | 1 | 59181 | 0.62 | 0.44 | 0.01 | | 0.06 | 0.21 | 0.11 | 0.61 | | -0.19 | -0.14 | -0.33 | 0.44 |
| 28 | TE | 2 | 59278 | 0.64 | 0.45 | 0.01 | 0.13 | 0.46 | 0.41 | | | -0.34 | -0.15 | 0.39 | | |
| 29 | MC | 1 | 59337 | 0.74 | 0.53 | 0.00 | | 0.74 | 0.08 | 0.13 | 0.05 | | 0.53 | -0.28 | -0.30 | -0.24 |
| 30 | MC | 1 | 59359 | 0.48 | 0.37 | 0.00 | | 0.24 | 0.16 | 0.48 | 0.11 | | -0.08 | -0.17 | 0.37 | -0.28 |
| 31 | EBSR | 2 | 59409 | 0.57 | 0.48 | 0.00 | 0.32 | 0.22 | 0.46 | | | -0.41 | -0.07 | 0.45 | | |

**Table H-4 (continued). Item Statistics, English Language Arts Grade 6**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | TE | 2 | 59321 | 0.67 | 0.54 | 0.00 | 0.15 | 0.35 | 0.49 | | | -0.43 | -0.18 | 0.48 | | |
| 33 | MC | 1 | 59328 | 0.67 | 0.47 | 0.00 | | 0.18 | 0.09 | 0.66 | 0.05 | | -0.22 | -0.30 | 0.48 | -0.21 |
| 34 | EBSR | 2 | 59412 | 0.50 | 0.62 | 0.00 | 0.38 | 0.24 | 0.37 | | | -0.56 | 0.02 | 0.55 | | |
| 35 | MC | 1 | 59341 | 0.42 | 0.39 | 0.00 | | 0.42 | 0.24 | 0.20 | 0.14 | | 0.40 | -0.10 | -0.26 | -0.12 |
| 36 | MC | 1 | 59372 | 0.57 | 0.42 | 0.00 | | 0.11 | 0.57 | 0.17 | 0.14 | | -0.21 | 0.42 | -0.23 | -0.16 |
| 37 | MC | 1 | 59348 | 0.64 | 0.55 | 0.00 | | 0.63 | 0.10 | 0.11 | 0.15 | | 0.55 | -0.22 | -0.29 | -0.28 |

## Table H-5. Item Statistics, English Language Arts Grade 7

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 59126 | 0.46 | 0.52 | 0.00 | | 0.34 | 0.44 | 0.15 | 0.03 | | -0.41 | 0.20 | 0.32 | 0.20 |
| 2 | MC | 1 | 61411 | 0.66 | 0.42 | 0.00 | | 0.07 | 0.65 | 0.23 | 0.05 | | -0.22 | 0.42 | -0.27 | -0.14 |
| 3 | TE | 2 | 61386 | 0.49 | 0.40 | 0.00 | 0.28 | 0.46 | 0.26 | | | -0.31 | -0.02 | 0.34 | | |
| 4 | EBSR | 2 | 61426 | 0.54 | 0.46 | 0.00 | 0.35 | 0.22 | 0.43 | | | -0.39 | -0.07 | 0.44 | | |
| 5 | MC | 1 | 61397 | 0.79 | 0.36 | 0.00 | | 0.79 | 0.08 | 0.08 | 0.05 | | 0.36 | -0.22 | -0.26 | -0.06 |
| 6 | TE | 2 | 61247 | 0.62 | 0.30 | 0.00 | 0.08 | 0.59 | 0.33 | | | -0.17 | -0.17 | 0.29 | | |
| 7 | MC | 1 | 61370 | 0.53 | 0.26 | 0.00 | | 0.22 | 0.14 | 0.53 | 0.10 | | -0.06 | -0.23 | 0.26 | -0.07 |
| 8 | MC | 1 | 61372 | 0.58 | 0.48 | 0.00 | | 0.58 | 0.23 | 0.11 | 0.08 | | 0.48 | -0.27 | -0.24 | -0.17 |
| 9 | MC | 1 | 61361 | 0.68 | 0.37 | 0.00 | | 0.17 | 0.09 | 0.68 | 0.07 | | -0.10 | -0.24 | 0.37 | -0.27 |
| 10 | TE | 2 | 61323 | 0.45 | 0.24 | 0.00 | 0.25 | 0.59 | 0.15 | | | -0.19 | 0.03 | 0.20 | | |
| 11 | MC | 1 | 61328 | 0.42 | 0.29 | 0.00 | | 0.42 | 0.25 | 0.21 | 0.12 | | 0.29 | -0.13 | -0.13 | -0.10 |
| 12 | TE | 1 | 61395 | 0.83 | 0.31 | 0.00 | 0.17 | 0.83 | | | | -0.31 | 0.31 | | | |
| 13 | MC | 1 | 61344 | 0.59 | 0.44 | 0.00 | | 0.09 | 0.10 | 0.22 | 0.59 | | -0.20 | -0.31 | -0.15 | 0.44 |
| 14 | EBSR | 2 | 61419 | 0.73 | 0.45 | 0.00 | 0.21 | 0.12 | 0.67 | | | -0.36 | -0.22 | 0.46 | | |
| 15 | MC | 1 | 61334 | 0.56 | 0.36 | 0.00 | | 0.06 | 0.22 | 0.16 | 0.56 | | -0.17 | -0.14 | -0.21 | 0.36 |
| 16 | MS | 2 | 61362 | 0.61 | 0.52 | 0.00 | 0.11 | 0.56 | 0.33 | | | -0.38 | -0.16 | 0.43 | | |
| 17 | MC | 1 | 61333 | 0.78 | 0.45 | 0.00 | | 0.08 | 0.78 | 0.09 | 0.05 | | -0.27 | 0.45 | -0.22 | -0.23 |
| 18 | EBSR | 2 | 61392 | 0.62 | 0.54 | 0.00 | 0.34 | 0.07 | 0.59 | | | -0.47 | -0.22 | 0.57 | | |
| 19 | MC | 1 | 61325 | 0.47 | 0.41 | 0.00 | | 0.47 | 0.14 | 0.34 | 0.05 | | 0.41 | -0.23 | -0.22 | -0.08 |
| 20 | MS | 2 | 61239 | 0.36 | 0.25 | 0.00 | 0.37 | 0.54 | 0.09 | | | -0.20 | 0.07 | 0.22 | | |
| 21 | MC | 1 | 61236 | 0.74 | 0.41 | 0.00 | | 0.09 | 0.09 | 0.74 | 0.07 | | -0.23 | -0.20 | 0.41 | -0.20 |
| 22 | MC | 1 | 61153 | 0.41 | 0.22 | 0.00 | | 0.23 | 0.41 | 0.12 | 0.24 | | 0.04 | 0.22 | -0.26 | -0.08 |
| 23 | EBSR | 2 | 61318 | 0.50 | 0.54 | 0.00 | 0.44 | 0.13 | 0.43 | | | -0.43 | -0.22 | 0.59 | | |
| 24 | MC | 1 | 61238 | 0.28 | 0.18 | 0.00 | | 0.51 | 0.08 | 0.13 | 0.28 | | 0.12 | -0.27 | -0.18 | 0.18 |
| 25 | TE | 1 | 60717 | 0.71 | 0.42 | 0.01 | 0.29 | 0.70 | | | | -0.40 | 0.43 | | | |
| 26 | MC | 1 | 61059 | 0.56 | 0.53 | 0.01 | | 0.13 | 0.21 | 0.09 | 0.56 | | -0.28 | -0.20 | -0.29 | 0.53 |
| 27 | EBSR | 2 | 61304 | 0.48 | 0.42 | 0.00 | 0.45 | 0.16 | 0.40 | | | -0.33 | -0.13 | 0.44 | | |
| 28 | MC | 1 | 61211 | 0.48 | 0.37 | 0.00 | | 0.22 | 0.48 | 0.18 | 0.12 | | -0.08 | 0.38 | -0.15 | -0.29 |
| 29 | TE | 1 | 60679 | 0.30 | 0.42 | 0.01 | 0.69 | 0.30 | | | | -0.39 | 0.42 | | | |
| 30 | MC | 1 | 61098 | 0.67 | 0.57 | 0.01 | | 0.07 | 0.18 | 0.08 | 0.67 | | -0.26 | -0.30 | -0.30 | 0.57 |
| 31 | EBSR | 2 | 61236 | 0.42 | 0.44 | 0.00 | 0.48 | 0.20 | 0.32 | | | -0.37 | -0.01 | 0.42 | | |

**Table H-5 (continued). Item Statistics, English Language Arts Grade 7**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 61154 | 0.54 | 0.40 | 0.00 | | 0.16 | 0.20 | 0.54 | 0.09 | | -0.15 | -0.18 | 0.40 | -0.24 |
| 33 | MC | 1 | 61122 | 0.46 | 0.46 | 0.00 | | 0.46 | 0.21 | 0.16 | 0.16 | | 0.46 | -0.21 | -0.26 | -0.12 |
| 34 | MS | 2 | 61177 | 0.55 | 0.51 | 0.00 | 0.20 | 0.50 | 0.29 | | | -0.41 | -0.03 | 0.41 | | |
| 35 | MS | 2 | 61196 | 0.71 | 0.50 | 0.00 | 0.10 | 0.40 | 0.51 | | | -0.35 | -0.24 | 0.45 | | |
| 36 | MC | 1 | 61122 | 0.53 | 0.43 | 0.00 | | 0.14 | 0.53 | 0.18 | 0.14 | | -0.17 | 0.43 | -0.22 | -0.19 |

# Table H-6. Item Statistics, English Language Arts Grade 8

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TDA | 4 | 60338 | 0.48 | 0.48 | 0.00 | | 0.32 | 0.43 | 0.17 | 0.04 | | -0.39 | 0.14 | 0.30 | 0.20 |
| 2 | MC | 1 | 62275 | 0.47 | 0.21 | 0.00 | | 0.08 | 0.26 | 0.19 | 0.47 | | -0.25 | -0.08 | -0.01 | 0.21 |
| 3 | MC | 1 | 62201 | 0.59 | 0.41 | 0.00 | | 0.18 | 0.58 | 0.14 | 0.09 | | -0.15 | 0.41 | -0.25 | -0.18 |
| 4 | MC | 1 | 62192 | 0.61 | 0.40 | 0.00 | | 0.12 | 0.16 | 0.11 | 0.61 | | -0.23 | -0.15 | -0.20 | 0.40 |
| 5 | MC | 1 | 62187 | 0.56 | 0.36 | 0.00 | | 0.14 | 0.56 | 0.11 | 0.19 | | -0.22 | 0.36 | -0.23 | -0.07 |
| 6 | MC | 1 | 62186 | 0.80 | 0.40 | 0.00 | | 0.11 | 0.06 | 0.80 | 0.03 | | -0.20 | -0.27 | 0.41 | -0.20 |
| 7 | MC | 1 | 62192 | 0.57 | 0.35 | 0.00 | | 0.16 | 0.09 | 0.17 | 0.57 | | -0.20 | -0.20 | -0.10 | 0.35 |
| 8 | MC | 1 | 62196 | 0.65 | 0.37 | 0.00 | | 0.12 | 0.09 | 0.65 | 0.14 | | -0.17 | -0.19 | 0.37 | -0.18 |
| 9 | EBSR | 2 | 62240 | 0.49 | 0.54 | 0.00 | 0.46 | 0.09 | 0.45 | | | -0.50 | -0.06 | 0.54 | | |
| 10 | TE | 2 | 62090 | 0.45 | 0.29 | 0.00 | 0.27 | 0.54 | 0.18 | | | -0.19 | -0.04 | 0.28 | | |
| 11 | MC | 1 | 62193 | 0.39 | 0.29 | 0.00 | | 0.12 | 0.39 | 0.06 | 0.43 | | -0.20 | 0.29 | -0.24 | -0.04 |
| 12 | TE | 2 | 62188 | 0.85 | 0.39 | 0.00 | 0.05 | 0.21 | 0.74 | | | -0.23 | -0.28 | 0.38 | | |
| 13 | MC | 1 | 62152 | 0.51 | 0.25 | 0.00 | | 0.17 | 0.51 | 0.19 | 0.13 | | -0.11 | 0.25 | -0.15 | -0.07 |
| 14 | MC | 1 | 62187 | 0.48 | 0.38 | 0.00 | | 0.14 | 0.11 | 0.28 | 0.47 | | -0.17 | -0.29 | -0.09 | 0.38 |
| 15 | MS | 2 | 62212 | 0.63 | 0.46 | 0.00 | 0.10 | 0.54 | 0.37 | | | -0.27 | -0.25 | 0.43 | | |
| 16 | EBSR | 2 | 62214 | 0.66 | 0.50 | 0.00 | 0.14 | 0.40 | 0.46 | | | -0.41 | -0.14 | 0.42 | | |
| 17 | MC | 1 | 62146 | 0.67 | 0.48 | 0.00 | | 0.10 | 0.15 | 0.67 | 0.08 | | -0.26 | -0.25 | 0.48 | -0.22 |
| 18 | MC | 1 | 62122 | 0.47 | 0.38 | 0.00 | | 0.47 | 0.11 | 0.26 | 0.16 | | 0.39 | -0.15 | -0.19 | -0.16 |
| 19 | EBSR | 2 | 62183 | 0.42 | 0.41 | 0.00 | 0.53 | 0.11 | 0.36 | | | -0.35 | -0.09 | 0.42 | | |
| 20 | MC | 1 | 62163 | 0.36 | 0.34 | 0.00 | | 0.17 | 0.30 | 0.17 | 0.36 | | -0.25 | -0.08 | -0.09 | 0.34 |
| 21 | MC | 1 | 62053 | 0.69 | 0.45 | 0.00 | | 0.69 | 0.11 | 0.06 | 0.14 | | 0.45 | -0.26 | -0.28 | -0.18 |
| 22 | MC | 1 | 62056 | 0.53 | 0.33 | 0.00 | | 0.53 | 0.15 | 0.19 | 0.13 | | 0.34 | -0.06 | -0.20 | -0.20 |
| 23 | MS | 2 | 62028 | 0.59 | 0.48 | 0.00 | 0.15 | 0.51 | 0.34 | | | -0.36 | -0.13 | 0.41 | | |
| 24 | MC | 1 | 62050 | 0.38 | 0.47 | 0.00 | | 0.26 | 0.25 | 0.10 | 0.38 | | -0.12 | -0.22 | -0.25 | 0.47 |
| 25 | MC | 1 | 61985 | 0.59 | 0.49 | 0.00 | | 0.04 | 0.08 | 0.29 | 0.59 | | -0.21 | -0.31 | -0.26 | 0.49 |
| 26 | MC | 1 | 62078 | 0.51 | 0.28 | 0.00 | | 0.14 | 0.51 | 0.19 | 0.16 | | -0.17 | 0.28 | -0.16 | -0.05 |
| 27 | TE | 2 | 61879 | 0.67 | 0.56 | 0.01 | 0.10 | 0.46 | 0.44 | | | -0.31 | -0.35 | 0.55 | | |
| 28 | MC | 1 | 61946 | 0.71 | 0.55 | 0.00 | | 0.71 | 0.12 | 0.10 | 0.07 | | 0.55 | -0.27 | -0.32 | -0.24 |
| 29 | MC | 1 | 62039 | 0.44 | 0.20 | 0.00 | | 0.08 | 0.32 | 0.16 | 0.44 | | -0.15 | -0.07 | -0.06 | 0.20 |
| 30 | MC | 1 | 61971 | 0.67 | 0.43 | 0.00 | | 0.10 | 0.11 | 0.67 | 0.13 | | -0.16 | -0.23 | 0.43 | -0.24 |
| 31 | MC | 1 | 61939 | 0.51 | 0.28 | 0.00 | | 0.18 | 0.13 | 0.18 | 0.50 | | -0.12 | -0.17 | -0.09 | 0.28 |

**Table H-6 (continued). Item Statistics, English Language Arts Grade 8**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 61912 | 0.53 | 0.34 | 0.01 | | 0.52 | 0.14 | 0.26 | 0.07 | | 0.34 | -0.27 | -0.01 | -0.25 |
| 33 | EBSR | 2 | 62058 | 0.56 | 0.64 | 0.00 | 0.30 | 0.28 | 0.42 | | | -0.59 | 0.02 | 0.54 | | |
| 34 | MC | 1 | 61975 | 0.73 | 0.54 | 0.00 | | 0.07 | 0.73 | 0.11 | 0.09 | | -0.23 | 0.55 | -0.32 | -0.28 |
| 35 | MC | 1 | 61982 | 0.39 | 0.20 | 0.00 | | 0.28 | 0.21 | 0.12 | 0.39 | | 0.07 | -0.14 | -0.21 | 0.21 |
| 36 | MC | 1 | 61946 | 0.50 | 0.30 | 0.00 | | 0.29 | 0.49 | 0.11 | 0.10 | | 0.00 | 0.31 | -0.25 | -0.24 |
| 37 | MC | 1 | 61977 | 0.55 | 0.44 | 0.00 | | 0.10 | 0.19 | 0.54 | 0.17 | | -0.19 | -0.29 | 0.45 | -0.12 |
| 38 | EBSR | 2 | 62044 | 0.23 | 0.16 | 0.00 | 0.70 | 0.13 | 0.17 | | | -0.05 | -0.20 | 0.26 | | |
| 39 | MC | 1 | 61885 | 0.61 | 0.43 | 0.01 | | 0.05 | 0.14 | 0.20 | 0.61 | | -0.21 | -0.16 | -0.27 | 0.43 |

## Table H-7. Item Statistics, Mathematics Grade 3

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TE | 1 | 58279 | 0.70 | 0.55 | 0.00 | 0.30 | 0.70 | | | | -0.55 | 0.55 | | | |
| 2 | SA | 1 | 58288 | 0.53 | 0.57 | 0.00 | 0.47 | 0.53 | | | | -0.57 | 0.57 | | | |
| 3 | MC | 1 | 58298 | 0.76 | 0.47 | 0.00 | | 0.04 | 0.08 | 0.76 | 0.12 | | -0.14 | -0.33 | 0.48 | -0.26 |
| 4 | TE | 1 | 58310 | 0.42 | 0.62 | 0.00 | 0.58 | 0.42 | | | | -0.62 | 0.62 | | | |
| 5 | MC | 1 | 58270 | 0.49 | 0.46 | 0.00 | | 0.16 | 0.30 | 0.05 | 0.49 | | -0.39 | -0.08 | -0.23 | 0.46 |
| 6 | SA | 1 | 57747 | 0.45 | 0.59 | 0.01 | 0.55 | 0.44 | | | | -0.57 | 0.59 | | | |
| 7 | MC | 1 | 57894 | 0.35 | 0.42 | 0.01 | | 0.28 | 0.28 | 0.35 | 0.08 | | -0.19 | -0.22 | 0.42 | -0.04 |
| 8 | SA | 1 | 58219 | 0.36 | 0.55 | 0.00 | 0.64 | 0.36 | | | | -0.54 | 0.55 | | | |
| 9 | TE | 1 | 57618 | 0.31 | 0.49 | 0.01 | 0.68 | 0.30 | | | | -0.46 | 0.49 | | | |
| 10 | MC | 1 | 58240 | 0.49 | 0.42 | 0.00 | | 0.20 | 0.49 | 0.23 | 0.08 | | -0.30 | 0.42 | -0.12 | -0.13 |
| 11 | MC | 1 | 58235 | 0.41 | 0.35 | 0.00 | | 0.22 | 0.18 | 0.41 | 0.20 | | -0.10 | -0.11 | 0.35 | -0.21 |
| 12 | MC | 1 | 58211 | 0.44 | 0.32 | 0.00 | | 0.27 | 0.15 | 0.44 | 0.13 | | -0.24 | -0.14 | 0.33 | 0.00 |
| 13 | MC | 1 | 58051 | 0.58 | 0.44 | 0.01 | | 0.58 | 0.11 | 0.20 | 0.10 | | 0.45 | -0.14 | -0.25 | -0.23 |
| 14 | SA | 1 | 57996 | 0.36 | 0.34 | 0.01 | 0.64 | 0.36 | | | | -0.33 | 0.34 | | | |
| 15 | MC | 1 | 58195 | 0.44 | 0.41 | 0.00 | | 0.25 | 0.44 | 0.17 | 0.14 | | -0.09 | 0.41 | -0.25 | -0.20 |
| 16 | MC | 1 | 58210 | 0.32 | 0.40 | 0.00 | | 0.32 | 0.16 | 0.22 | 0.30 | | 0.40 | 0.04 | -0.16 | -0.29 |
| 17 | SA | 1 | 58174 | 0.56 | 0.56 | 0.00 | 0.44 | 0.56 | | | | -0.55 | 0.56 | | | |
| 18 | TE | 1 | 57313 | 0.83 | 0.44 | 0.02 | 0.17 | 0.81 | | | | -0.43 | 0.46 | | | |
| 19 | MC | 1 | 58222 | 0.72 | 0.49 | 0.00 | | 0.05 | 0.09 | 0.15 | 0.71 | | -0.22 | -0.25 | -0.29 | 0.49 |
| 20 | MC | 1 | 58202 | 0.58 | 0.44 | 0.00 | | 0.18 | 0.58 | 0.12 | 0.12 | | -0.22 | 0.44 | -0.15 | -0.25 |
| 21 | SA | 1 | 58222 | 0.70 | 0.59 | 0.00 | 0.30 | 0.70 | | | | -0.59 | 0.59 | | | |
| 22 | MC | 1 | 58246 | 0.45 | 0.39 | 0.00 | | 0.45 | 0.33 | 0.08 | 0.13 | | 0.39 | -0.03 | -0.18 | -0.38 |
| 23 | TE | 1 | 58247 | 0.81 | 0.42 | 0.00 | 0.19 | 0.81 | | | | -0.42 | 0.43 | | | |
| 24 | SA | 1 | 58226 | 0.70 | 0.62 | 0.00 | 0.30 | 0.70 | | | | -0.62 | 0.62 | | | |
| 25 | TE | 1 | 58225 | 0.58 | 0.36 | 0.00 | 0.42 | 0.57 | | | | -0.36 | 0.37 | | | |
| 26 | SA | 1 | 58210 | 0.40 | 0.57 | 0.00 | 0.60 | 0.40 | | | | -0.56 | 0.57 | | | |
| 27 | SA | 1 | 57939 | 0.62 | 0.59 | 0.01 | 0.38 | 0.61 | | | | -0.58 | 0.59 | | | |
| 28 | MC | 1 | 57887 | 0.40 | 0.37 | 0.01 | | 0.14 | 0.15 | 0.30 | 0.39 | | -0.26 | -0.19 | -0.04 | 0.37 |
| 29 | MC | 1 | 58237 | 0.92 | 0.32 | 0.00 | | 0.92 | 0.05 | 0.02 | 0.01 | | 0.32 | -0.22 | -0.19 | -0.12 |
| 30 | MC | 1 | 58232 | 0.44 | 0.27 | 0.00 | | 0.20 | 0.20 | 0.16 | 0.43 | | -0.13 | -0.04 | -0.18 | 0.27 |
| 31 | MC | 1 | 58229 | 0.39 | 0.43 | 0.00 | | 0.40 | 0.13 | 0.07 | 0.39 | | -0.25 | -0.13 | -0.16 | 0.43 |

**Table H-7 (continued). Item Statistics, Mathematics Grade 3**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|----|---|-------|------|------|------|------|------|------|------|------|-------|------|-------|-------|-------|
|   |    |   |       |      |      |      | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | SA | 1 | 58133 | 0.52 | 0.54 | 0.00 | 0.48 | 0.52 |      |      |      | -0.53 | 0.54 |       |       |       |
| 33 | SA | 1 | 58174 | 0.52 | 0.61 | 0.00 | 0.48 | 0.52 |      |      |      | -0.60 | 0.61 |       |       |       |
| 34 | MC | 1 | 58022 | 0.68 | 0.49 | 0.01 |      | 0.67 | 0.08 | 0.09 | 0.15 |       | 0.49 | -0.22 | -0.20 | -0.30 |
| 35 | SA | 1 | 57935 | 0.48 | 0.57 | 0.01 | 0.51 | 0.48 |      |      |      | -0.56 | 0.57 |       |       |       |
| 36 | MC | 1 | 58208 | 0.64 | 0.46 | 0.00 |      | 0.19 | 0.64 | 0.07 | 0.10 |       | -0.34 | 0.47 | -0.20 | -0.13 |
| 37 | MC | 1 | 58220 | 0.69 | 0.59 | 0.00 |      | 0.20 | 0.07 | 0.68 | 0.04 |       | -0.45 | -0.20 | 0.59 | -0.20 |
| 38 | MC | 1 | 58212 | 0.61 | 0.48 | 0.00 |      | 0.06 | 0.19 | 0.61 | 0.14 |       | -0.25 | -0.32 | 0.48 | -0.14 |
| 39 | MC | 1 | 58215 | 0.72 | 0.49 | 0.00 |      | 0.06 | 0.07 | 0.72 | 0.15 |       | -0.22 | -0.23 | 0.50 | -0.31 |
| 40 | MC | 1 | 58230 | 0.66 | 0.51 | 0.00 |      | 0.16 | 0.66 | 0.13 | 0.05 |       | -0.28 | 0.51 | -0.22 | -0.29 |
| 41 | SA | 1 | 58193 | 0.38 | 0.41 | 0.00 | 0.62 | 0.38 |      |      |      | -0.40 | 0.41 |       |       |       |
| 42 | TE | 1 | 57607 | 0.44 | 0.52 | 0.01 | 0.55 | 0.43 |      |      |      | -0.49 | 0.53 |       |       |       |

## Table H-8. Item Statistics, Mathematics Grade 4

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | MC | 1 | 58780 | 0.71 | 0.43 | 0.00 | | 0.71 | 0.10 | 0.08 | 0.11 | | 0.43 | -0.23 | -0.22 | -0.21 |
| 2 | SA | 1 | 58736 | 0.55 | 0.59 | 0.00 | 0.45 | 0.55 | | | | -0.59 | 0.59 | | | |
| 3 | TE | 1 | 58800 | 0.69 | 0.40 | 0.00 | 0.31 | 0.69 | | | | -0.40 | 0.40 | | | |
| 4 | MC | 1 | 58754 | 0.48 | 0.31 | 0.00 | | 0.19 | 0.48 | 0.16 | 0.17 | | -0.21 | 0.31 | -0.21 | 0.00 |
| 5 | MC | 1 | 58808 | 0.61 | 0.45 | 0.00 | | 0.60 | 0.13 | 0.17 | 0.09 | | 0.45 | -0.18 | -0.23 | -0.25 |
| 6 | MC | 1 | 58619 | 0.63 | 0.48 | 0.00 | | 0.09 | 0.17 | 0.63 | 0.12 | | -0.26 | -0.26 | 0.48 | -0.19 |
| 7 | MC | 1 | 58639 | 0.61 | 0.41 | 0.00 | | 0.31 | 0.61 | 0.05 | 0.03 | | -0.28 | 0.41 | -0.22 | -0.14 |
| 8 | SA | 1 | 58718 | 0.35 | 0.63 | 0.00 | 0.65 | 0.35 | | | | -0.63 | 0.63 | | | |
| 9 | MC | 1 | 58766 | 0.70 | 0.52 | 0.00 | | 0.70 | 0.09 | 0.14 | 0.07 | | 0.52 | -0.28 | -0.25 | -0.28 |
| 10 | TE | 1 | 58768 | 0.55 | 0.62 | 0.00 | 0.45 | 0.55 | | | | -0.61 | 0.62 | | | |
| 11 | MC | 1 | 58762 | 0.54 | 0.44 | 0.00 | | 0.13 | 0.17 | 0.54 | 0.16 | | -0.17 | -0.22 | 0.44 | -0.21 |
| 12 | MC | 1 | 58767 | 0.58 | 0.42 | 0.00 | | 0.10 | 0.14 | 0.58 | 0.18 | | -0.17 | -0.18 | 0.42 | -0.24 |
| 13 | MC | 1 | 58660 | 0.50 | 0.60 | 0.00 | | 0.19 | 0.12 | 0.19 | 0.50 | | -0.26 | -0.32 | -0.24 | 0.60 |
| 14 | MC | 1 | 58593 | 0.41 | 0.29 | 0.00 | | 0.13 | 0.34 | 0.11 | 0.41 | | 0.01 | -0.25 | -0.07 | 0.29 |
| 15 | MC | 1 | 58746 | 0.40 | 0.29 | 0.00 | | 0.17 | 0.40 | 0.31 | 0.11 | | -0.24 | 0.29 | 0.03 | -0.19 |
| 16 | SA | 1 | 58714 | 0.29 | 0.50 | 0.00 | 0.70 | 0.29 | | | | -0.49 | 0.50 | | | |
| 17 | MC | 1 | 58732 | 0.26 | 0.06 | 0.00 | | 0.35 | 0.15 | 0.23 | 0.26 | | 0.28 | -0.17 | -0.23 | 0.06 |
| 18 | MC | 1 | 58723 | 0.40 | 0.49 | 0.00 | | 0.16 | 0.27 | 0.17 | 0.40 | | -0.26 | -0.14 | -0.22 | 0.49 |
| 19 | TE | 1 | 58745 | 0.32 | 0.60 | 0.00 | 0.68 | 0.32 | | | | -0.60 | 0.60 | | | |
| 20 | SA | 1 | 58744 | 0.41 | 0.50 | 0.00 | 0.59 | 0.41 | | | | -0.49 | 0.50 | | | |
| 21 | MC | 1 | 58752 | 0.50 | 0.50 | 0.00 | | 0.12 | 0.50 | 0.09 | 0.30 | | -0.06 | 0.50 | -0.30 | -0.31 |
| 22 | MC | 1 | 58575 | 0.30 | 0.17 | 0.00 | | 0.20 | 0.29 | 0.18 | 0.32 | | -0.18 | 0.17 | -0.16 | 0.13 |
| 23 | MC | 1 | 58607 | 0.51 | 0.50 | 0.00 | | 0.28 | 0.09 | 0.51 | 0.11 | | -0.33 | -0.15 | 0.50 | -0.18 |
| 24 | TE | 1 | 58549 | 0.69 | 0.39 | 0.00 | 0.31 | 0.69 | | | | -0.39 | 0.40 | | | |
| 25 | MC | 1 | 58743 | 0.82 | 0.39 | 0.00 | | 0.07 | 0.06 | 0.82 | 0.05 | | -0.26 | -0.19 | 0.39 | -0.17 |
| 26 | MC | 1 | 58728 | 0.37 | 0.45 | 0.00 | | 0.34 | 0.37 | 0.24 | 0.05 | | -0.13 | 0.46 | -0.28 | -0.17 |
| 27 | MC | 1 | 58713 | 0.33 | 0.58 | 0.00 | | 0.33 | 0.07 | 0.46 | 0.14 | | 0.58 | -0.10 | -0.25 | -0.35 |
| 28 | MC | 1 | 58741 | 0.34 | 0.19 | 0.00 | | 0.44 | 0.34 | 0.14 | 0.08 | | 0.10 | 0.19 | -0.24 | -0.19 |
| 29 | MC | 1 | 58621 | 0.54 | 0.49 | 0.00 | | 0.12 | 0.54 | 0.15 | 0.19 | | -0.21 | 0.49 | -0.22 | -0.24 |
| 30 | MC | 1 | 58530 | 0.40 | 0.33 | 0.01 | | 0.40 | 0.22 | 0.21 | 0.17 | | 0.34 | -0.11 | -0.15 | -0.14 |
| 31 | SA | 1 | 58704 | 0.42 | 0.57 | 0.00 | 0.58 | 0.42 | | | | -0.56 | 0.57 | | | |

**Table H-8 (continued). Item Statistics, Mathematics Grade 4**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | TE | 1 | 58398 | 0.20 | 0.50 | 0.01 | 0.80 | 0.19 | | | | -0.47 | 0.50 | | | |
| 33 | MC | 1 | 58726 | 0.49 | 0.19 | 0.00 | | 0.03 | 0.08 | 0.49 | 0.40 | | -0.10 | -0.18 | 0.19 | -0.05 |
| 34 | MC | 1 | 58711 | 0.39 | 0.42 | 0.00 | | 0.07 | 0.47 | 0.07 | 0.39 | | -0.26 | -0.17 | -0.23 | 0.43 |
| 35 | SA | 1 | 58673 | 0.18 | 0.54 | 0.00 | 0.81 | 0.18 | | | | -0.53 | 0.54 | | | |
| 36 | MC | 1 | 58585 | 0.52 | 0.43 | 0.00 | | 0.52 | 0.17 | 0.17 | 0.13 | | 0.44 | -0.27 | -0.17 | -0.14 |
| 37 | MC | 1 | 58609 | 0.84 | 0.37 | 0.00 | | 0.04 | 0.07 | 0.05 | 0.84 | | -0.20 | -0.24 | -0.16 | 0.37 |
| 38 | MC | 1 | 58700 | 0.35 | 0.55 | 0.00 | | 0.35 | 0.18 | 0.26 | 0.20 | | 0.55 | -0.08 | -0.20 | -0.35 |
| 39 | TE | 1 | 58705 | 0.67 | 0.45 | 0.00 | 0.33 | 0.67 | | | | -0.45 | 0.45 | | | |
| 40 | SA | 1 | 58708 | 0.61 | 0.50 | 0.00 | 0.39 | 0.61 | | | | -0.50 | 0.50 | | | |
| 41 | MC | 1 | 58719 | 0.64 | 0.53 | 0.00 | | 0.64 | 0.13 | 0.12 | 0.11 | | 0.53 | -0.24 | -0.28 | -0.26 |
| 42 | MC | 1 | 58720 | 0.68 | 0.45 | 0.00 | | 0.18 | 0.07 | 0.06 | 0.68 | | -0.26 | -0.22 | -0.23 | 0.46 |
| 43 | SA | 1 | 58668 | 0.45 | 0.48 | 0.00 | 0.55 | 0.45 | | | | -0.48 | 0.48 | | | |
| 44 | MC | 1 | 58713 | 0.56 | 0.37 | 0.00 | | 0.19 | 0.14 | 0.56 | 0.10 | | -0.17 | -0.18 | 0.38 | -0.17 |
| 45 | SA | 1 | 58607 | 0.51 | 0.60 | 0.00 | 0.49 | 0.50 | | | | -0.60 | 0.61 | | | |
| 46 | MC | 1 | 58571 | 0.58 | 0.43 | 0.00 | | 0.58 | 0.15 | 0.19 | 0.07 | | 0.44 | -0.21 | -0.22 | -0.19 |

**Table H-9. Item Statistics, Mathematics Grade 5**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | SA | 1 | 58995 | 0.30 | 0.59 | 0.00 | 0.69 | 0.30 | | | | -0.59 | 0.59 | | | |
| 2 | MC | 1 | 59002 | 0.34 | 0.20 | 0.00 | | 0.21 | 0.34 | 0.26 | 0.20 | | -0.10 | 0.20 | -0.04 | -0.08 |
| 3 | TE | 1 | 59027 | 0.52 | 0.55 | 0.00 | 0.48 | 0.52 | | | | -0.55 | 0.55 | | | |
| 4 | MC | 1 | 59032 | 0.52 | 0.46 | 0.00 | | 0.06 | 0.18 | 0.52 | 0.23 | | -0.21 | -0.20 | 0.46 | -0.24 |
| 5 | MC | 1 | 59011 | 0.39 | 0.45 | 0.00 | | 0.29 | 0.16 | 0.17 | 0.38 | | -0.09 | -0.21 | -0.26 | 0.45 |
| 6 | SA | 1 | 58842 | 0.35 | 0.44 | 0.00 | 0.65 | 0.35 | | | | -0.43 | 0.45 | | | |
| 7 | TE | 1 | 58863 | 0.20 | 0.53 | 0.00 | 0.79 | 0.20 | | | | -0.51 | 0.53 | | | |
| 8 | MC | 1 | 59024 | 0.66 | 0.52 | 0.00 | | 0.08 | 0.19 | 0.66 | 0.07 | | -0.25 | -0.33 | 0.52 | -0.18 |
| 9 | MC | 1 | 59010 | 0.75 | 0.41 | 0.00 | | 0.75 | 0.14 | 0.08 | 0.03 | | 0.41 | -0.21 | -0.27 | -0.18 |
| 10 | SA | 1 | 58989 | 0.48 | 0.57 | 0.00 | 0.52 | 0.47 | | | | -0.56 | 0.57 | | | |
| 11 | MC | 1 | 58967 | 0.33 | 0.21 | 0.00 | | 0.17 | 0.33 | 0.18 | 0.32 | | -0.17 | 0.21 | -0.14 | 0.04 |
| 12 | MC | 1 | 59015 | 0.53 | 0.31 | 0.00 | | 0.15 | 0.14 | 0.18 | 0.53 | | -0.11 | -0.13 | -0.18 | 0.31 |
| 13 | SA | 1 | 58872 | 0.48 | 0.55 | 0.00 | 0.52 | 0.48 | | | | -0.54 | 0.55 | | | |
| 14 | MC | 1 | 58874 | 0.65 | 0.35 | 0.00 | | 0.11 | 0.65 | 0.11 | 0.13 | | -0.19 | 0.35 | -0.19 | -0.13 |
| 15 | MC | 1 | 58979 | 0.38 | 0.44 | 0.00 | | 0.35 | 0.16 | 0.10 | 0.38 | | -0.19 | -0.17 | -0.19 | 0.44 |
| 16 | SA | 1 | 58978 | 0.30 | 0.35 | 0.00 | 0.70 | 0.30 | | | | -0.35 | 0.35 | | | |
| 17 | MC | 1 | 58987 | 0.49 | 0.32 | 0.00 | | 0.49 | 0.22 | 0.15 | 0.14 | | 0.32 | -0.04 | -0.25 | -0.15 |
| 18 | SA | 1 | 58956 | 0.43 | 0.47 | 0.00 | 0.57 | 0.43 | | | | -0.47 | 0.47 | | | |
| 19 | MC | 1 | 59008 | 0.63 | 0.54 | 0.00 | | 0.07 | 0.15 | 0.15 | 0.63 | | -0.24 | -0.25 | -0.31 | 0.54 |
| 20 | MC | 1 | 58980 | 0.38 | 0.26 | 0.00 | | 0.09 | 0.38 | 0.33 | 0.19 | | 0.01 | 0.26 | -0.14 | -0.16 |
| 21 | SA | 1 | 58936 | 0.19 | 0.43 | 0.00 | 0.81 | 0.19 | | | | -0.42 | 0.43 | | | |
| 22 | TE | 1 | 58728 | 0.52 | 0.50 | 0.01 | 0.47 | 0.52 | | | | -0.49 | 0.50 | | | |
| 23 | MC | 1 | 58808 | 0.34 | 0.43 | 0.01 | | 0.26 | 0.17 | 0.22 | 0.34 | | -0.21 | -0.30 | 0.01 | 0.43 |
| 24 | MC | 1 | 59019 | 0.70 | 0.38 | 0.00 | | 0.05 | 0.10 | 0.15 | 0.70 | | -0.20 | -0.22 | -0.19 | 0.38 |
| 25 | TE | 1 | 58876 | 0.21 | 0.58 | 0.00 | 0.79 | 0.21 | | | | -0.57 | 0.58 | | | |
| 26 | MC | 1 | 58992 | 0.38 | 0.37 | 0.00 | | 0.22 | 0.25 | 0.38 | 0.15 | | -0.06 | -0.09 | 0.37 | -0.32 |
| 27 | SA | 1 | 58915 | 0.43 | 0.60 | 0.00 | 0.57 | 0.43 | | | | -0.59 | 0.60 | | | |
| 28 | MC | 1 | 58991 | 0.47 | 0.42 | 0.00 | | 0.47 | 0.12 | 0.14 | 0.27 | | 0.42 | -0.24 | -0.23 | -0.11 |
| 29 | MC | 1 | 58950 | 0.76 | 0.38 | 0.00 | | 0.04 | 0.76 | 0.13 | 0.07 | | -0.15 | 0.38 | -0.29 | -0.13 |
| 30 | TE | 1 | 57570 | 0.29 | 0.58 | 0.03 | 0.69 | 0.29 | | | | -0.51 | 0.58 | | | |
| 31 | TE | 1 | 58939 | 0.53 | 0.59 | 0.00 | 0.47 | 0.53 | | | | -0.59 | 0.59 | | | |

**Table H-9 (continued). Item Statistics, Mathematics Grade 5**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 58945 | 0.71 | 0.39 | 0.00 | | 0.11 | 0.14 | 0.71 | 0.05 | | -0.18 | -0.26 | 0.40 | -0.16 |
| 33 | MC | 1 | 58976 | 0.42 | 0.30 | 0.00 | | 0.28 | 0.42 | 0.14 | 0.16 | | -0.23 | 0.30 | -0.24 | 0.11 |
| 34 | SA | 1 | 58872 | 0.26 | 0.57 | 0.00 | 0.73 | 0.26 | | | | -0.56 | 0.57 | | | |
| 35 | MC | 1 | 58976 | 0.61 | 0.57 | 0.00 | | 0.61 | 0.19 | 0.10 | 0.09 | | 0.57 | -0.34 | -0.27 | -0.20 |
| 36 | MC | 1 | 58931 | 0.32 | 0.46 | 0.00 | | 0.40 | 0.17 | 0.32 | 0.10 | | -0.37 | -0.03 | 0.46 | -0.07 |
| 37 | SA | 1 | 58682 | 0.47 | 0.55 | 0.01 | 0.52 | 0.47 | | | | -0.54 | 0.55 | | | |
| 38 | MC | 1 | 58947 | 0.48 | 0.50 | 0.00 | | 0.48 | 0.13 | 0.27 | 0.12 | | 0.50 | -0.31 | -0.15 | -0.23 |
| 39 | SA | 1 | 58863 | 0.24 | 0.55 | 0.00 | 0.76 | 0.24 | | | | -0.54 | 0.55 | | | |
| 40 | MC | 1 | 58987 | 0.47 | 0.33 | 0.00 | | 0.20 | 0.16 | 0.17 | 0.47 | | -0.12 | -0.12 | -0.20 | 0.33 |
| 41 | SA | 1 | 58950 | 0.68 | 0.36 | 0.00 | 0.32 | 0.68 | | | | -0.36 | 0.36 | | | |
| 42 | MC | 1 | 58964 | 0.83 | 0.45 | 0.00 | | 0.07 | 0.08 | 0.83 | 0.03 | | -0.27 | -0.28 | 0.45 | -0.16 |
| 43 | MC | 1 | 58929 | 0.22 | 0.25 | 0.00 | | 0.27 | 0.24 | 0.26 | 0.22 | | -0.01 | -0.15 | -0.07 | 0.25 |
| 44 | SA | 1 | 58924 | 0.61 | 0.52 | 0.00 | 0.39 | 0.61 | | | | -0.52 | 0.53 | | | |
| 45 | MC | 1 | 58904 | 0.24 | 0.14 | 0.00 | | 0.24 | 0.49 | 0.12 | 0.15 | | 0.14 | 0.13 | -0.15 | -0.22 |
| 46 | MC | 1 | 58851 | 0.55 | 0.47 | 0.00 | | 0.21 | 0.55 | 0.11 | 0.13 | | -0.16 | 0.47 | -0.26 | -0.26 |

## Table H-10. Item Statistics, Mathematics Grade 6

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | MC | 1 | 60019 | 0.36 | 0.38 | 0.00 | | 0.36 | 0.26 | 0.30 | 0.07 | | 0.38 | -0.20 | -0.22 | 0.04 |
| 2 | MC | 1 | 60069 | 0.36 | 0.48 | 0.00 | | 0.14 | 0.41 | 0.09 | 0.36 | | -0.26 | -0.22 | -0.12 | 0.48 |
| 3 | TE | 1 | 59550 | 0.51 | 0.56 | 0.01 | 0.49 | 0.50 | | | | -0.54 | 0.57 | | | |
| 4 | MC | 1 | 59978 | 0.51 | 0.38 | 0.00 | | 0.16 | 0.51 | 0.17 | 0.15 | | -0.29 | 0.38 | -0.11 | -0.11 |
| 5 | MC | 1 | 59999 | 0.51 | 0.47 | 0.00 | | 0.25 | 0.15 | 0.50 | 0.10 | | -0.26 | -0.23 | 0.47 | -0.14 |
| 6 | SA | 1 | 59845 | 0.20 | 0.56 | 0.00 | 0.80 | 0.20 | | | | -0.54 | 0.56 | | | |
| 7 | SA | 1 | 59741 | 0.41 | 0.55 | 0.01 | 0.59 | 0.41 | | | | -0.54 | 0.55 | | | |
| 8 | MC | 1 | 59983 | 0.79 | 0.39 | 0.00 | | 0.03 | 0.09 | 0.09 | 0.79 | | -0.17 | -0.22 | -0.23 | 0.39 |
| 9 | MC | 1 | 59984 | 0.41 | 0.44 | 0.00 | | 0.39 | 0.12 | 0.41 | 0.08 | | -0.14 | -0.29 | 0.44 | -0.20 |
| 10 | SA | 1 | 59986 | 0.15 | 0.48 | 0.00 | 0.85 | 0.15 | | | | -0.46 | 0.48 | | | |
| 11 | MC | 1 | 59992 | 0.34 | 0.35 | 0.00 | | 0.34 | 0.31 | 0.20 | 0.15 | | 0.35 | -0.16 | -0.13 | -0.11 |
| 12 | SA | 1 | 59870 | 0.46 | 0.60 | 0.00 | 0.54 | 0.46 | | | | -0.59 | 0.60 | | | |
| 13 | MC | 1 | 59989 | 0.75 | 0.39 | 0.00 | | 0.75 | 0.10 | 0.07 | 0.08 | | 0.39 | -0.26 | -0.26 | -0.09 |
| 14 | TE | 1 | 59828 | 0.35 | 0.66 | 0.00 | 0.65 | 0.35 | | | | -0.64 | 0.66 | | | |
| 15 | MC | 1 | 59972 | 0.48 | 0.49 | 0.00 | | 0.24 | 0.16 | 0.11 | 0.48 | | -0.27 | -0.20 | -0.17 | 0.49 |
| 16 | TE | 1 | 59527 | 0.32 | 0.57 | 0.01 | 0.67 | 0.32 | | | | -0.55 | 0.58 | | | |
| 17 | MC | 1 | 59967 | 0.35 | 0.49 | 0.00 | | 0.19 | 0.35 | 0.31 | 0.16 | | -0.31 | 0.49 | -0.07 | -0.22 |
| 18 | MC | 1 | 59894 | 0.53 | 0.52 | 0.00 | | 0.53 | 0.24 | 0.12 | 0.11 | | 0.52 | -0.36 | -0.21 | -0.13 |
| 19 | MC | 1 | 59906 | 0.69 | 0.39 | 0.00 | | 0.07 | 0.69 | 0.13 | 0.12 | | -0.17 | 0.39 | -0.20 | -0.23 |
| 20 | SA | 1 | 59902 | 0.58 | 0.55 | 0.00 | 0.42 | 0.58 | | | | -0.55 | 0.55 | | | |
| 21 | TE | 1 | 59836 | 0.53 | 0.55 | 0.00 | 0.46 | 0.53 | | | | -0.54 | 0.55 | | | |
| 22 | MC | 1 | 59794 | 0.49 | 0.42 | 0.00 | | 0.31 | 0.49 | 0.16 | 0.04 | | -0.24 | 0.42 | -0.20 | -0.11 |
| 23 | TE | 1 | 58428 | 0.68 | 0.34 | 0.03 | 0.31 | 0.66 | | | | -0.31 | 0.36 | | | |
| 24 | MC | 1 | 59760 | 0.62 | 0.40 | 0.01 | | 0.62 | 0.12 | 0.16 | 0.09 | | 0.41 | -0.22 | -0.15 | -0.22 |
| 25 | MC | 1 | 59768 | 0.30 | 0.31 | 0.01 | | 0.49 | 0.30 | 0.12 | 0.08 | | -0.17 | 0.32 | -0.21 | 0.05 |
| 26 | MC | 1 | 59890 | 0.43 | 0.37 | 0.00 | | 0.26 | 0.17 | 0.14 | 0.43 | | -0.12 | -0.22 | -0.13 | 0.37 |
| 27 | SA | 1 | 59712 | 0.36 | 0.23 | 0.01 | 0.64 | 0.36 | | | | -0.22 | 0.24 | | | |
| 28 | SA | 1 | 59707 | 0.33 | 0.54 | 0.01 | 0.66 | 0.33 | | | | -0.52 | 0.54 | | | |
| 29 | SA | 1 | 59620 | 0.29 | 0.41 | 0.01 | 0.70 | 0.29 | | | | -0.39 | 0.41 | | | |
| 30 | MC | 1 | 59885 | 0.57 | 0.34 | 0.00 | | 0.17 | 0.20 | 0.57 | 0.05 | | -0.22 | -0.14 | 0.34 | -0.10 |
| 31 | TE | 1 | 59719 | 0.44 | 0.60 | 0.01 | 0.56 | 0.44 | | | | -0.58 | 0.60 | | | |

**Table H-10 (continued). Item Statistics, Mathematics Grade 6**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 59748 | 0.28 | 0.39 | 0.01 | | 0.23 | 0.25 | 0.24 | 0.28 | | -0.14 | -0.05 | -0.21 | 0.39 |
| 33 | MC | 1 | 59785 | 0.64 | 0.47 | 0.00 | | 0.14 | 0.16 | 0.63 | 0.07 | | -0.25 | -0.25 | 0.47 | -0.18 |
| 34 | MC | 1 | 59733 | 0.29 | 0.46 | 0.01 | | 0.28 | 0.35 | 0.17 | 0.20 | | 0.46 | -0.15 | -0.17 | -0.17 |
| 35 | MC | 1 | 59713 | 0.42 | 0.47 | 0.01 | | 0.38 | 0.08 | 0.12 | 0.41 | | -0.13 | -0.27 | -0.27 | 0.47 |
| 36 | SA | 1 | 59632 | 0.17 | 0.56 | 0.01 | 0.83 | 0.16 | | | | -0.53 | 0.56 | | | |
| 37 | MC | 1 | 59804 | 0.59 | 0.37 | 0.00 | | 0.16 | 0.14 | 0.59 | 0.10 | | -0.15 | -0.24 | 0.38 | -0.13 |
| 38 | MC | 1 | 59807 | 0.63 | 0.31 | 0.00 | | 0.11 | 0.62 | 0.16 | 0.10 | | 0.04 | 0.31 | -0.23 | -0.25 |
| 39 | TE | 1 | 59051 | 0.50 | 0.55 | 0.02 | 0.49 | 0.49 | | | | -0.52 | 0.56 | | | |
| 40 | MC | 1 | 59798 | 0.29 | 0.48 | 0.00 | | 0.30 | 0.29 | 0.09 | 0.32 | | -0.23 | 0.48 | -0.12 | -0.15 |
| 41 | MC | 1 | 59827 | 0.51 | 0.48 | 0.00 | | 0.15 | 0.17 | 0.17 | 0.51 | | -0.21 | -0.23 | -0.20 | 0.48 |
| 42 | MC | 1 | 59826 | 0.38 | 0.26 | 0.00 | | 0.13 | 0.17 | 0.32 | 0.38 | | -0.17 | -0.18 | 0.00 | 0.26 |
| 43 | MC | 1 | 59825 | 0.35 | 0.23 | 0.00 | | 0.22 | 0.27 | 0.34 | 0.17 | | -0.21 | -0.06 | 0.23 | 0.02 |
| 44 | MC | 1 | 59759 | 0.45 | 0.41 | 0.01 | | 0.15 | 0.19 | 0.20 | 0.45 | | -0.16 | -0.20 | -0.17 | 0.42 |
| 45 | TE | 1 | 59682 | 0.36 | 0.58 | 0.01 | 0.64 | 0.36 | | | | -0.56 | 0.58 | | | |
| 46 | MC | 1 | 59758 | 0.26 | 0.49 | 0.01 | | 0.26 | 0.11 | 0.11 | 0.51 | | 0.50 | -0.01 | -0.13 | -0.34 |

**Table H-11. Item Statistics, Mathematics Grade 7**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | SA | 1 | 61649 | 0.42 | 0.65 | 0.00 | 0.57 | 0.42 | | | | -0.64 | 0.65 | | | |
| 2 | SA | 1 | 61658 | 0.43 | 0.61 | 0.00 | 0.57 | 0.42 | | | | -0.61 | 0.62 | | | |
| 3 | MC | 1 | 61723 | 0.47 | 0.33 | 0.00 | | 0.47 | 0.14 | 0.24 | 0.16 | | 0.33 | -0.08 | -0.22 | -0.12 |
| 4 | SA | 1 | 61560 | 0.22 | 0.48 | 0.00 | 0.77 | 0.22 | | | | -0.47 | 0.48 | | | |
| 5 | MC | 1 | 61724 | 0.62 | 0.49 | 0.00 | | 0.62 | 0.17 | 0.13 | 0.08 | | 0.49 | -0.27 | -0.27 | -0.16 |
| 6 | MC | 1 | 61711 | 0.42 | 0.44 | 0.00 | | 0.42 | 0.09 | 0.30 | 0.18 | | 0.44 | -0.26 | -0.19 | -0.14 |
| 7 | MC | 1 | 61733 | 0.49 | 0.43 | 0.00 | | 0.07 | 0.48 | 0.16 | 0.28 | | -0.16 | 0.43 | -0.28 | -0.16 |
| 8 | TE | 1 | 61732 | 0.17 | 0.42 | 0.00 | 0.83 | 0.17 | | | | -0.42 | 0.42 | | | |
| 9 | MC | 1 | 61728 | 0.33 | 0.28 | 0.00 | | 0.28 | 0.13 | 0.26 | 0.33 | | 0.07 | -0.25 | -0.19 | 0.28 |
| 10 | MC | 1 | 61721 | 0.47 | 0.41 | 0.00 | | 0.47 | 0.22 | 0.24 | 0.07 | | 0.42 | -0.32 | -0.06 | -0.19 |
| 11 | MC | 1 | 61742 | 0.53 | 0.34 | 0.00 | | 0.13 | 0.53 | 0.13 | 0.21 | | -0.16 | 0.34 | -0.27 | -0.06 |
| 12 | MC | 1 | 61695 | 0.72 | 0.26 | 0.00 | | 0.16 | 0.06 | 0.07 | 0.71 | | -0.09 | -0.16 | -0.18 | 0.26 |
| 13 | MC | 1 | 61663 | 0.39 | 0.38 | 0.00 | | 0.25 | 0.15 | 0.39 | 0.21 | | -0.17 | -0.18 | 0.38 | -0.12 |
| 14 | MC | 1 | 61673 | 0.43 | 0.41 | 0.00 | | 0.42 | 0.20 | 0.20 | 0.18 | | 0.41 | -0.14 | -0.07 | -0.31 |
| 15 | SA | 1 | 61374 | 0.28 | 0.42 | 0.01 | 0.71 | 0.28 | | | | -0.41 | 0.42 | | | |
| 16 | TE | 1 | 61593 | 0.72 | 0.22 | 0.00 | 0.28 | 0.71 | | | | -0.22 | 0.22 | | | |
| 17 | MC | 1 | 61480 | 0.47 | 0.55 | 0.01 | | 0.13 | 0.19 | 0.20 | 0.47 | | -0.30 | -0.33 | -0.10 | 0.55 |
| 18 | MC | 1 | 61486 | 0.40 | 0.21 | 0.01 | | 0.04 | 0.19 | 0.40 | 0.36 | | -0.12 | -0.06 | 0.21 | -0.11 |
| 19 | TE | 1 | 60996 | 0.32 | 0.63 | 0.01 | 0.67 | 0.32 | | | | -0.60 | 0.63 | | | |
| 20 | MC | 1 | 61532 | 0.37 | 0.25 | 0.00 | | 0.27 | 0.25 | 0.36 | 0.11 | | -0.14 | -0.14 | 0.25 | 0.02 |
| 21 | TE | 1 | 61513 | 0.58 | 0.48 | 0.00 | 0.42 | 0.57 | | | | -0.47 | 0.48 | | | |
| 22 | MC | 1 | 61481 | 0.34 | 0.38 | 0.01 | | 0.34 | 0.21 | 0.22 | 0.23 | | 0.39 | -0.15 | -0.24 | -0.04 |
| 23 | MC | 1 | 61516 | 0.28 | 0.20 | 0.00 | | 0.36 | 0.29 | 0.28 | 0.06 | | -0.05 | -0.11 | 0.20 | -0.06 |
| 24 | SA | 1 | 61359 | 0.35 | 0.55 | 0.01 | 0.65 | 0.35 | | | | -0.53 | 0.55 | | | |
| 25 | SA | 1 | 61257 | 0.17 | 0.57 | 0.01 | 0.83 | 0.17 | | | | -0.54 | 0.57 | | | |
| 26 | MC | 1 | 61523 | 0.49 | 0.48 | 0.00 | | 0.08 | 0.49 | 0.18 | 0.24 | | -0.13 | 0.49 | -0.22 | -0.27 |
| 27 | MC | 1 | 61530 | 0.43 | 0.47 | 0.00 | | 0.43 | 0.16 | 0.17 | 0.24 | | 0.47 | -0.30 | -0.25 | -0.07 |
| 28 | MC | 1 | 61501 | 0.52 | 0.16 | 0.00 | | 0.08 | 0.14 | 0.26 | 0.51 | | -0.02 | -0.07 | -0.11 | 0.16 |
| 29 | SA | 1 | 61165 | 0.51 | 0.56 | 0.01 | 0.49 | 0.50 | | | | -0.54 | 0.56 | | | |
| 30 | MC | 1 | 61429 | 0.36 | 0.35 | 0.01 | | 0.12 | 0.36 | 0.28 | 0.23 | | -0.05 | 0.36 | -0.14 | -0.20 |
| 31 | TE | 1 | 60823 | 0.63 | 0.24 | 0.02 | 0.37 | 0.62 | | | | -0.22 | 0.25 | | | |

**Table H-11 (continued). Item Statistics, Mathematics Grade 7**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 61438 | 0.30 | 0.19 | 0.01 | | 0.35 | 0.22 | 0.29 | 0.12 | | -0.08 | -0.13 | 0.19 | 0.03 |
| 33 | MC | 1 | 61438 | 0.47 | 0.52 | 0.01 | | 0.47 | 0.22 | 0.21 | 0.10 | | 0.52 | -0.28 | -0.15 | -0.26 |
| 34 | MC | 1 | 61479 | 0.54 | 0.33 | 0.01 | | 0.13 | 0.54 | 0.20 | 0.13 | | -0.22 | 0.33 | -0.22 | 0.00 |
| 35 | SA | 1 | 61227 | 0.36 | 0.68 | 0.01 | 0.63 | 0.36 | | | | -0.66 | 0.68 | | | |
| 36 | TE | 1 | 61405 | 0.64 | 0.51 | 0.01 | 0.36 | 0.64 | | | | -0.50 | 0.52 | | | |
| 37 | MC | 1 | 61450 | 0.26 | 0.48 | 0.01 | | 0.33 | 0.27 | 0.14 | 0.26 | | -0.11 | -0.17 | -0.21 | 0.48 |
| 38 | SA | 1 | 61241 | 0.29 | 0.57 | 0.01 | 0.70 | 0.29 | | | | -0.54 | 0.57 | | | |
| 39 | MC | 1 | 61427 | 0.48 | 0.59 | 0.01 | | 0.48 | 0.14 | 0.23 | 0.14 | | 0.59 | -0.19 | -0.32 | -0.25 |
| 40 | SA | 1 | 61204 | 0.27 | 0.53 | 0.01 | 0.72 | 0.27 | | | | -0.51 | 0.53 | | | |
| 41 | MC | 1 | 61374 | 0.42 | 0.33 | 0.01 | | 0.12 | 0.20 | 0.26 | 0.42 | | -0.06 | -0.10 | -0.22 | 0.33 |
| 42 | TE | 1 | 60971 | 0.44 | 0.25 | 0.01 | 0.56 | 0.43 | | | | -0.23 | 0.26 | | | |
| 43 | MC | 1 | 61356 | 0.44 | 0.52 | 0.01 | | 0.13 | 0.19 | 0.24 | 0.43 | | -0.16 | -0.26 | -0.23 | 0.52 |
| 44 | MC | 1 | 61330 | 0.24 | 0.18 | 0.01 | | 0.13 | 0.33 | 0.29 | 0.24 | | -0.07 | -0.06 | -0.04 | 0.18 |
| 45 | SA | 1 | 61217 | 0.62 | 0.60 | 0.01 | 0.37 | 0.62 | | | | -0.58 | 0.60 | | | |
| 46 | MC | 1 | 61357 | 0.59 | 0.48 | 0.01 | | 0.06 | 0.14 | 0.59 | 0.20 | | -0.18 | -0.24 | 0.48 | -0.25 |

## Table H-12. Item Statistics, Mathematics Grade 8

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | MC | 1 | 62544 | 0.59 | 0.49 | 0.00 | | 0.17 | 0.59 | 0.08 | 0.16 | | -0.20 | 0.49 | -0.21 | -0.29 |
| 2 | MC | 1 | 62528 | 0.51 | 0.39 | 0.00 | | 0.24 | 0.10 | 0.51 | 0.15 | | -0.14 | -0.16 | 0.39 | -0.24 |
| 3 | MC | 1 | 62518 | 0.36 | 0.31 | 0.00 | | 0.36 | 0.26 | 0.24 | 0.14 | | 0.31 | 0.05 | -0.26 | -0.16 |
| 4 | MC | 1 | 62530 | 0.44 | 0.50 | 0.00 | | 0.16 | 0.44 | 0.19 | 0.21 | | 0.04 | 0.50 | -0.30 | -0.35 |
| 5 | SA | 1 | 62406 | 0.14 | 0.50 | 0.00 | 0.86 | 0.14 | | | | -0.49 | 0.50 | | | |
| 6 | MC | 1 | 62515 | 0.31 | 0.44 | 0.00 | | 0.35 | 0.17 | 0.16 | 0.31 | | -0.14 | -0.15 | -0.21 | 0.44 |
| 7 | MC | 1 | 62523 | 0.38 | 0.34 | 0.00 | | 0.38 | 0.38 | 0.16 | 0.08 | | 0.34 | -0.12 | -0.16 | -0.17 |
| 8 | SA | 1 | 62345 | 0.39 | 0.67 | 0.00 | 0.61 | 0.39 | | | | -0.66 | 0.67 | | | |
| 9 | TE | 1 | 62455 | 0.18 | 0.54 | 0.00 | 0.82 | 0.18 | | | | -0.53 | 0.54 | | | |
| 10 | MC | 1 | 62489 | 0.51 | 0.39 | 0.00 | | 0.07 | 0.29 | 0.12 | 0.51 | | -0.20 | -0.20 | -0.15 | 0.39 |
| 11 | SA | 1 | 62392 | 0.42 | 0.65 | 0.00 | 0.58 | 0.42 | | | | -0.64 | 0.65 | | | |
| 12 | MC | 1 | 62496 | 0.46 | 0.29 | 0.00 | | 0.19 | 0.20 | 0.46 | 0.15 | | -0.07 | -0.26 | 0.29 | -0.04 |
| 13 | MC | 1 | 62522 | 0.45 | 0.34 | 0.00 | | 0.09 | 0.21 | 0.25 | 0.45 | | -0.02 | -0.21 | -0.18 | 0.34 |
| 14 | MC | 1 | 62470 | 0.74 | 0.33 | 0.00 | | 0.74 | 0.13 | 0.07 | 0.05 | | 0.33 | -0.14 | -0.24 | -0.16 |
| 15 | TE | 1 | 62434 | 0.43 | 0.52 | 0.00 | 0.57 | 0.43 | | | | -0.51 | 0.52 | | | |
| 16 | MC | 1 | 62447 | 0.74 | 0.43 | 0.00 | | 0.06 | 0.74 | 0.13 | 0.06 | | -0.22 | 0.43 | -0.24 | -0.21 |
| 17 | SA | 1 | 62082 | 0.33 | 0.54 | 0.01 | 0.66 | 0.33 | | | | -0.52 | 0.54 | | | |
| 18 | MC | 1 | 62389 | 0.53 | 0.41 | 0.00 | | 0.12 | 0.20 | 0.53 | 0.14 | | -0.15 | -0.27 | 0.41 | -0.12 |
| 19 | MC | 1 | 62297 | 0.44 | 0.34 | 0.00 | | 0.11 | 0.44 | 0.38 | 0.07 | | -0.14 | 0.34 | -0.20 | -0.10 |
| 20 | MC | 1 | 62290 | 0.57 | 0.36 | 0.00 | | 0.19 | 0.56 | 0.13 | 0.11 | | -0.11 | 0.36 | -0.28 | -0.13 |
| 21 | MC | 1 | 62362 | 0.39 | 0.24 | 0.00 | | 0.25 | 0.20 | 0.39 | 0.16 | | 0.01 | -0.19 | 0.24 | -0.11 |
| 22 | MC | 1 | 62354 | 0.68 | 0.42 | 0.00 | | 0.13 | 0.68 | 0.13 | 0.06 | | -0.24 | 0.43 | -0.21 | -0.18 |
| 23 | TE | 1 | 61476 | 0.50 | 0.52 | 0.02 | 0.49 | 0.50 | | | | -0.50 | 0.53 | | | |
| 24 | MC | 1 | 62376 | 0.54 | 0.37 | 0.00 | | 0.06 | 0.53 | 0.21 | 0.19 | | -0.15 | 0.37 | -0.22 | -0.13 |
| 25 | TE | 1 | 62119 | 0.20 | 0.49 | 0.01 | 0.80 | 0.19 | | | | -0.47 | 0.49 | | | |
| 26 | MC | 1 | 62327 | 0.27 | 0.19 | 0.00 | | 0.12 | 0.35 | 0.26 | 0.26 | | -0.19 | -0.02 | 0.19 | -0.02 |
| 27 | SA | 1 | 62025 | 0.30 | 0.67 | 0.01 | 0.69 | 0.30 | | | | -0.64 | 0.67 | | | |
| 28 | MC | 1 | 62337 | 0.53 | 0.57 | 0.00 | | 0.06 | 0.18 | 0.23 | 0.52 | | -0.21 | -0.24 | -0.32 | 0.57 |
| 29 | MC | 1 | 62209 | 0.30 | 0.39 | 0.01 | | 0.16 | 0.24 | 0.29 | 0.30 | | -0.11 | -0.19 | -0.11 | 0.39 |
| 30 | MC | 1 | 62280 | 0.53 | 0.51 | 0.00 | | 0.15 | 0.09 | 0.23 | 0.53 | | -0.15 | -0.21 | -0.33 | 0.51 |
| 31 | SA | 1 | 61331 | 0.25 | 0.61 | 0.02 | 0.74 | 0.24 | | | | -0.56 | 0.61 | | | |

**Table H-12 (continued). Item Statistics, Mathematics Grade 8**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | TE | 1 | 62076 | 0.32 | 0.51 | 0.01 | 0.67 | 0.32 | | | | -0.49 | 0.51 | | | |
| 33 | MC | 1 | 62187 | 0.52 | 0.38 | 0.01 | | 0.16 | 0.52 | 0.18 | 0.13 | | -0.10 | 0.38 | -0.24 | -0.17 |
| 34 | TE | 1 | 62145 | 0.22 | 0.66 | 0.01 | 0.78 | 0.22 | | | | -0.64 | 0.66 | | | |
| 35 | MC | 1 | 62184 | 0.43 | 0.26 | 0.01 | | 0.09 | 0.22 | 0.43 | 0.25 | | -0.13 | -0.26 | 0.26 | 0.05 |
| 36 | MC | 1 | 62251 | 0.43 | 0.43 | 0.01 | | 0.43 | 0.19 | 0.22 | 0.15 | | 0.43 | -0.18 | -0.21 | -0.14 |
| 37 | TE | 1 | 61686 | 0.28 | 0.25 | 0.01 | 0.71 | 0.28 | | | | -0.22 | 0.26 | | | |
| 38 | TE | 1 | 61863 | 0.44 | 0.38 | 0.01 | 0.55 | 0.44 | | | | -0.36 | 0.39 | | | |
| 39 | MC | 1 | 62205 | 0.38 | 0.37 | 0.01 | | 0.38 | 0.20 | 0.24 | 0.18 | | 0.38 | -0.27 | -0.12 | -0.04 |
| 40 | MC | 1 | 62245 | 0.45 | 0.30 | 0.01 | | 0.44 | 0.18 | 0.29 | 0.08 | | 0.30 | -0.26 | 0.00 | -0.16 |
| 41 | MC | 1 | 62150 | 0.37 | 0.23 | 0.01 | | 0.15 | 0.26 | 0.22 | 0.37 | | -0.08 | 0.02 | -0.20 | 0.23 |
| 42 | TE | 1 | 61002 | 0.23 | 0.50 | 0.03 | 0.75 | 0.23 | | | | -0.44 | 0.50 | | | |
| 43 | MC | 1 | 62174 | 0.66 | 0.42 | 0.01 | | 0.08 | 0.12 | 0.66 | 0.14 | | -0.20 | -0.22 | 0.42 | -0.20 |
| 44 | MC | 1 | 62174 | 0.60 | 0.39 | 0.01 | | 0.07 | 0.15 | 0.59 | 0.17 | | -0.13 | -0.19 | 0.40 | -0.23 |
| 45 | MC | 1 | 62181 | 0.65 | 0.43 | 0.01 | | 0.09 | 0.65 | 0.18 | 0.08 | | -0.25 | 0.44 | -0.19 | -0.21 |
| 46 | MC | 1 | 62186 | 0.37 | 0.35 | 0.01 | | 0.37 | 0.15 | 0.23 | 0.24 | | 0.36 | -0.14 | -0.23 | -0.04 |

## Table H-13. Item Statistics, Science Grade 4

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TE | 1 | 58647 | 0.77 | 0.36 | 0.00 | 0.23 | 0.77 | | | | -0.36 | 0.36 | | | |
| 2 | TE | 1 | 58705 | 0.73 | 0.52 | 0.00 | 0.27 | 0.73 | | | | -0.52 | 0.52 | | | |
| 3 | MC | 1 | 58692 | 0.50 | 0.32 | 0.00 | | 0.31 | 0.11 | 0.50 | 0.08 | | -0.06 | -0.29 | 0.32 | -0.14 |
| 4 | TE | 1 | 58691 | 0.32 | 0.32 | 0.00 | 0.68 | 0.32 | | | | -0.32 | 0.32 | | | |
| 5 | TE | 1 | 58678 | 0.44 | 0.44 | 0.00 | 0.56 | 0.44 | | | | -0.44 | 0.44 | | | |
| 6 | TE | 1 | 58566 | 0.52 | 0.39 | 0.00 | 0.48 | 0.52 | | | | -0.38 | 0.39 | | | |
| 7 | TE | 1 | 58661 | 0.60 | 0.52 | 0.00 | 0.40 | 0.60 | | | | -0.51 | 0.52 | | | |
| 8 | MC | 1 | 58688 | 0.35 | 0.30 | 0.00 | | 0.35 | 0.27 | 0.32 | 0.05 | | 0.30 | -0.06 | -0.16 | -0.19 |
| 9 | TE | 1 | 58669 | 0.59 | 0.43 | 0.00 | 0.41 | 0.58 | | | | -0.43 | 0.43 | | | |
| 10 | EBSR | 1 | 58719 | 0.41 | 0.45 | 0.00 | 0.59 | 0.40 | | | | -0.44 | 0.45 | | | |
| 11 | MC | 1 | 58706 | 0.65 | 0.38 | 0.00 | | 0.18 | 0.65 | 0.08 | 0.09 | | -0.19 | 0.39 | -0.25 | -0.14 |
| 12 | MC | 1 | 58700 | 0.66 | 0.49 | 0.00 | | 0.11 | 0.07 | 0.15 | 0.66 | | -0.19 | -0.25 | -0.30 | 0.49 |
| 13 | MC | 1 | 58624 | 0.38 | 0.12 | 0.00 | | 0.38 | 0.34 | 0.16 | 0.11 | | 0.12 | -0.04 | -0.10 | -0.02 |
| 14 | TE | 1 | 58616 | 0.42 | 0.57 | 0.00 | 0.58 | 0.41 | | | | -0.56 | 0.57 | | | |
| 15 | TE | 1 | 58678 | 0.42 | 0.37 | 0.00 | 0.58 | 0.42 | | | | -0.36 | 0.37 | | | |
| 16 | TE | 1 | 58727 | 0.82 | 0.41 | 0.00 | 0.18 | 0.82 | | | | -0.41 | 0.41 | | | |
| 17 | TE | 1 | 58649 | 0.68 | 0.43 | 0.00 | 0.32 | 0.68 | | | | -0.42 | 0.43 | | | |
| 18 | MC | 1 | 58674 | 0.59 | 0.34 | 0.00 | | 0.12 | 0.17 | 0.12 | 0.59 | | -0.08 | -0.26 | -0.13 | 0.34 |
| 19 | TE | 1 | 58551 | 0.50 | 0.24 | 0.00 | 0.49 | 0.50 | | | | -0.24 | 0.24 | | | |
| 20 | TE | 1 | 58663 | 0.87 | 0.32 | 0.00 | 0.13 | 0.87 | | | | -0.32 | 0.32 | | | |
| 21 | TE | 1 | 58565 | 0.51 | 0.42 | 0.00 | 0.48 | 0.51 | | | | -0.42 | 0.42 | | | |
| 22 | MC | 1 | 58652 | 0.50 | 0.55 | 0.00 | | 0.13 | 0.18 | 0.19 | 0.50 | | -0.26 | -0.28 | -0.21 | 0.55 |
| 23 | MC | 1 | 58648 | 0.57 | 0.43 | 0.00 | | 0.08 | 0.19 | 0.16 | 0.57 | | -0.18 | -0.19 | -0.24 | 0.43 |
| 24 | TE | 1 | 58693 | 0.23 | 0.40 | 0.00 | 0.77 | 0.23 | | | | -0.39 | 0.40 | | | |
| 25 | MC | 1 | 58682 | 0.42 | 0.37 | 0.00 | | 0.13 | 0.42 | 0.20 | 0.25 | | -0.13 | 0.37 | -0.19 | -0.14 |
| 26 | MC | 1 | 58657 | 0.64 | 0.35 | 0.00 | | 0.14 | 0.11 | 0.11 | 0.64 | | -0.25 | -0.20 | -0.06 | 0.35 |
| 27 | MC | 1 | 58648 | 0.83 | 0.33 | 0.00 | | 0.03 | 0.83 | 0.06 | 0.08 | | -0.22 | 0.33 | -0.22 | -0.12 |
| 28 | MC | 1 | 58606 | 0.38 | 0.35 | 0.00 | | 0.15 | 0.30 | 0.17 | 0.38 | | -0.12 | -0.08 | -0.24 | 0.35 |
| 29 | TE | 1 | 58364 | 0.39 | 0.34 | 0.01 | 0.61 | 0.39 | | | | -0.33 | 0.34 | | | |
| 30 | TE | 1 | 58607 | 0.36 | 0.16 | 0.00 | 0.64 | 0.36 | | | | -0.16 | 0.17 | | | |
| 31 | TE | 1 | 58571 | 0.45 | 0.61 | 0.00 | 0.55 | 0.45 | | | | -0.60 | 0.61 | | | |

**Table H-13 (continued). Item Statistics, Science Grade 4**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | TE | 1 | 58522 | 0.34 | 0.36 | 0.00 | 0.65 | 0.34 | | | | -0.36 | 0.37 | | | |
| 33 | MC | 1 | 58623 | 0.66 | 0.44 | 0.00 | | 0.66 | 0.17 | 0.09 | 0.08 | | 0.44 | -0.29 | -0.19 | -0.15 |
| 34 | TE | 1 | 58636 | 0.71 | 0.41 | 0.00 | 0.29 | 0.71 | | | | -0.41 | 0.41 | | | |
| 35 | TE | 1 | 58643 | 0.89 | 0.37 | 0.00 | 0.11 | 0.89 | | | | -0.37 | 0.37 | | | |
| 36 | TE | 1 | 58594 | 0.55 | 0.34 | 0.00 | 0.45 | 0.55 | | | | -0.34 | 0.34 | | | |
| 37 | TE | 1 | 58602 | 0.35 | 0.30 | 0.00 | 0.64 | 0.35 | | | | -0.29 | 0.30 | | | |
| 38 | MC | 1 | 58605 | 0.44 | 0.42 | 0.00 | | 0.22 | 0.18 | 0.16 | 0.44 | | -0.15 | -0.23 | -0.15 | 0.42 |
| 39 | TE | 1 | 58629 | 0.53 | 0.38 | 0.00 | 0.47 | 0.53 | | | | -0.37 | 0.38 | | | |
| 40 | MC | 1 | 58608 | 0.51 | 0.52 | 0.00 | | 0.15 | 0.21 | 0.12 | 0.51 | | -0.23 | -0.23 | -0.26 | 0.52 |

# Table H-14. Item Statistics, Science Grade 8

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | MS | 1 | 62457 | 0.43 | 0.44 | 0.00 | 0.57 | 0.43 | | | | -0.44 | 0.44 | | | |
| 2 | MC | 1 | 62413 | 0.47 | 0.22 | 0.00 | | 0.16 | 0.25 | 0.47 | 0.11 | | 0.04 | -0.19 | 0.22 | -0.12 |
| 3 | TE | 1 | 62361 | 0.73 | 0.31 | 0.00 | 0.27 | 0.72 | | | | -0.31 | 0.32 | | | |
| 4 | TE | 1 | 62033 | 0.66 | 0.34 | 0.01 | 0.34 | 0.66 | | | | -0.33 | 0.34 | | | |
| 5 | TE | 1 | 62135 | 0.43 | 0.40 | 0.01 | 0.57 | 0.42 | | | | -0.39 | 0.40 | | | |
| 6 | TE | 1 | 62460 | 0.22 | 0.24 | 0.00 | 0.77 | 0.22 | | | | -0.24 | 0.24 | | | |
| 7 | TE | 1 | 62328 | 0.47 | 0.39 | 0.00 | 0.53 | 0.47 | | | | -0.38 | 0.39 | | | |
| 8 | MC | 1 | 62333 | 0.61 | 0.39 | 0.00 | | 0.09 | 0.61 | 0.24 | 0.05 | | -0.08 | 0.40 | -0.30 | -0.17 |
| 9 | TE | 1 | 62364 | 0.78 | 0.32 | 0.00 | 0.22 | 0.78 | | | | -0.31 | 0.32 | | | |
| 10 | MC | 1 | 62411 | 0.41 | 0.16 | 0.00 | | 0.17 | 0.41 | 0.21 | 0.21 | | -0.09 | 0.16 | -0.17 | 0.07 |
| 11 | MC | 1 | 62387 | 0.54 | 0.41 | 0.00 | | 0.54 | 0.18 | 0.19 | 0.09 | | 0.41 | -0.21 | -0.22 | -0.13 |
| 12 | MC | 1 | 62322 | 0.44 | 0.23 | 0.00 | | 0.22 | 0.44 | 0.23 | 0.11 | | -0.09 | 0.23 | -0.17 | 0.00 |
| 13 | TE | 1 | 62413 | 0.40 | 0.49 | 0.00 | 0.60 | 0.40 | | | | -0.49 | 0.49 | | | |
| 14 | TE | 1 | 62392 | 0.68 | 0.36 | 0.00 | 0.32 | 0.68 | | | | -0.35 | 0.36 | | | |
| 15 | TE | 1 | 62396 | 0.33 | 0.41 | 0.00 | 0.67 | 0.33 | | | | -0.41 | 0.41 | | | |
| 16 | TE | 1 | 62386 | 0.60 | 0.37 | 0.00 | 0.40 | 0.60 | | | | -0.37 | 0.37 | | | |
| 17 | MC | 1 | 62278 | 0.49 | 0.24 | 0.00 | | 0.26 | 0.09 | 0.49 | 0.17 | | -0.04 | -0.12 | 0.24 | -0.18 |
| 18 | TE | 1 | 62326 | 0.49 | 0.31 | 0.00 | 0.51 | 0.49 | | | | -0.31 | 0.31 | | | |
| 19 | MC | 1 | 62226 | 0.53 | 0.42 | 0.00 | | 0.19 | 0.10 | 0.53 | 0.18 | | -0.12 | -0.22 | 0.42 | -0.26 |
| 20 | MC | 1 | 62282 | 0.48 | 0.36 | 0.00 | | 0.48 | 0.14 | 0.19 | 0.18 | | 0.36 | -0.22 | -0.14 | -0.12 |
| 21 | TE | 1 | 62309 | 0.79 | 0.38 | 0.00 | 0.21 | 0.78 | | | | -0.38 | 0.38 | | | |
| 22 | TE | 1 | 62306 | 0.53 | 0.43 | 0.00 | 0.47 | 0.53 | | | | -0.43 | 0.43 | | | |
| 23 | TE | 1 | 62256 | 0.69 | 0.43 | 0.00 | 0.31 | 0.69 | | | | -0.42 | 0.43 | | | |
| 24 | TE | 1 | 62283 | 0.61 | 0.45 | 0.00 | 0.38 | 0.61 | | | | -0.45 | 0.45 | | | |
| 25 | TE | 1 | 62298 | 0.46 | 0.04 | 0.00 | 0.54 | 0.45 | | | | -0.03 | 0.04 | | | |
| 26 | TE | 1 | 62247 | 0.48 | 0.55 | 0.00 | 0.52 | 0.48 | | | | -0.55 | 0.55 | | | |
| 27 | TE | 1 | 62160 | 0.57 | 0.17 | 0.00 | 0.43 | 0.57 | | | | -0.17 | 0.17 | | | |
| 28 | TE | 1 | 62259 | 0.32 | 0.51 | 0.00 | 0.68 | 0.32 | | | | -0.50 | 0.51 | | | |
| 29 | TE | 1 | 62221 | 0.75 | 0.33 | 0.00 | 0.25 | 0.75 | | | | -0.32 | 0.33 | | | |
| 30 | TE | 1 | 62204 | 0.75 | 0.37 | 0.00 | 0.25 | 0.75 | | | | -0.37 | 0.37 | | | |
| 31 | TE | 1 | 61944 | 0.40 | 0.37 | 0.01 | 0.59 | 0.40 | | | | -0.36 | 0.37 | | | |

**Table H-14 (continued). Item Statistics, Science Grade 8**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 62121 | 0.50 | 0.43 | 0.00 | | 0.15 | 0.21 | 0.50 | 0.13 | | -0.26 | -0.27 | 0.43 | -0.03 |
| 33 | TE | 1 | 62161 | 0.60 | 0.42 | 0.00 | 0.40 | 0.60 | | | | -0.42 | 0.42 | | | |
| 34 | EBSR | 1 | 62216 | 0.42 | 0.48 | 0.00 | 0.58 | 0.42 | | | | -0.48 | 0.48 | | | |
| 35 | TE | 1 | 62208 | 0.41 | 0.50 | 0.00 | 0.59 | 0.41 | | | | -0.49 | 0.50 | | | |
| 36 | MC | 1 | 62168 | 0.55 | 0.47 | 0.00 | | 0.08 | 0.23 | 0.14 | 0.55 | | -0.16 | -0.20 | -0.31 | 0.47 |
| 37 | TE | 1 | 62141 | 0.39 | 0.40 | 0.00 | 0.60 | 0.39 | | | | -0.40 | 0.40 | | | |
| 38 | MC | 1 | 62144 | 0.61 | 0.55 | 0.00 | | 0.60 | 0.12 | 0.21 | 0.06 | | 0.56 | -0.23 | -0.36 | -0.20 |
| 39 | TE | 1 | 62163 | 0.27 | 0.36 | 0.00 | 0.73 | 0.27 | | | | -0.36 | 0.36 | | | |
| 40 | TE | 1 | 62130 | 0.64 | 0.38 | 0.00 | 0.36 | 0.64 | | | | -0.37 | 0.38 | | | |

# Table H-15. Item Statistics, Social Studies Grade 4

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TE | 1 | 58739 | 0.54 | 0.47 | 0.00 | 0.46 | 0.54 | | | | -0.47 | 0.47 | | | |
| 2 | MC | 1 | 58711 | 0.72 | 0.41 | 0.00 | | 0.07 | 0.15 | 0.06 | 0.72 | | -0.28 | -0.19 | -0.21 | 0.42 |
| 3 | TE | 1 | 58685 | 0.25 | 0.20 | 0.00 | 0.75 | 0.25 | | | | -0.20 | 0.20 | | | |
| 4 | MC | 1 | 58687 | 0.72 | 0.48 | 0.00 | | 0.07 | 0.11 | 0.11 | 0.72 | | -0.22 | -0.32 | -0.20 | 0.48 |
| 5 | MC | 1 | 58719 | 0.72 | 0.44 | 0.00 | | 0.72 | 0.05 | 0.12 | 0.11 | | 0.45 | -0.25 | -0.21 | -0.25 |
| 6 | MS | 1 | 58712 | 0.36 | 0.50 | 0.00 | 0.64 | 0.36 | | | | -0.50 | 0.50 | | | |
| 7 | MC | 1 | 58685 | 0.73 | 0.49 | 0.00 | | 0.06 | 0.07 | 0.14 | 0.72 | | -0.26 | -0.26 | -0.25 | 0.49 |
| 8 | MC | 1 | 58680 | 0.51 | 0.29 | 0.00 | | 0.12 | 0.25 | 0.12 | 0.50 | | -0.27 | 0.05 | -0.24 | 0.29 |
| 9 | TE | 1 | 58688 | 0.81 | 0.50 | 0.00 | 0.19 | 0.81 | | | | -0.50 | 0.50 | | | |
| 10 | MC | 1 | 58672 | 0.35 | 0.23 | 0.00 | | 0.35 | 0.23 | 0.26 | 0.17 | | 0.23 | -0.07 | -0.08 | -0.11 |
| 11 | MC | 1 | 58684 | 0.53 | 0.34 | 0.00 | | 0.16 | 0.14 | 0.53 | 0.17 | | -0.14 | -0.21 | 0.34 | -0.12 |
| 12 | MS | 1 | 58421 | 0.37 | 0.50 | 0.01 | 0.62 | 0.37 | | | | -0.49 | 0.50 | | | |
| 13 | MC | 1 | 58636 | 0.65 | 0.47 | 0.00 | | 0.16 | 0.10 | 0.65 | 0.10 | | -0.20 | -0.22 | 0.47 | -0.30 |
| 14 | TE | 1 | 58691 | 0.61 | 0.45 | 0.00 | 0.39 | 0.61 | | | | -0.44 | 0.45 | | | |
| 15 | TE | 1 | 58693 | 0.64 | 0.46 | 0.00 | 0.36 | 0.64 | | | | -0.46 | 0.47 | | | |
| 16 | MC | 1 | 58667 | 0.76 | 0.49 | 0.00 | | 0.76 | 0.09 | 0.11 | 0.04 | | 0.49 | -0.31 | -0.25 | -0.21 |
| 17 | TE | 1 | 58575 | 0.76 | 0.51 | 0.00 | 0.24 | 0.75 | | | | -0.50 | 0.51 | | | |
| 18 | MC | 1 | 58680 | 0.59 | 0.50 | 0.00 | | 0.59 | 0.12 | 0.09 | 0.20 | | 0.50 | -0.28 | -0.22 | -0.23 |
| 19 | MC | 1 | 58678 | 0.40 | 0.41 | 0.00 | | 0.22 | 0.21 | 0.17 | 0.40 | | -0.05 | -0.22 | -0.23 | 0.41 |
| 20 | MC | 1 | 58510 | 0.53 | 0.40 | 0.00 | | 0.52 | 0.15 | 0.21 | 0.12 | | 0.40 | -0.27 | -0.15 | -0.12 |
| 21 | MC | 1 | 58698 | 0.44 | 0.31 | 0.00 | | 0.44 | 0.29 | 0.21 | 0.06 | | 0.31 | -0.11 | -0.12 | -0.23 |
| 22 | MC | 1 | 58651 | 0.71 | 0.53 | 0.00 | | 0.05 | 0.71 | 0.14 | 0.10 | | -0.19 | 0.53 | -0.32 | -0.29 |
| 23 | TE | 1 | 58668 | 0.53 | 0.52 | 0.00 | 0.47 | 0.53 | | | | -0.52 | 0.52 | | | |
| 24 | MC | 1 | 58615 | 0.60 | 0.49 | 0.00 | | 0.13 | 0.60 | 0.13 | 0.14 | | -0.16 | 0.49 | -0.33 | -0.21 |
| 25 | TE | 1 | 58646 | 0.68 | 0.61 | 0.00 | 0.31 | 0.68 | | | | -0.61 | 0.61 | | | |
| 26 | TE | 1 | 58581 | 0.50 | 0.37 | 0.00 | 0.49 | 0.50 | | | | -0.37 | 0.38 | | | |
| 27 | TE | 1 | 58606 | 0.51 | 0.63 | 0.00 | 0.49 | 0.51 | | | | -0.62 | 0.63 | | | |
| 28 | MC | 1 | 58528 | 0.46 | 0.30 | 0.00 | | 0.10 | 0.19 | 0.45 | 0.25 | | -0.16 | -0.14 | 0.30 | -0.09 |
| 29 | TE | 1 | 58637 | 0.53 | 0.36 | 0.00 | 0.47 | 0.53 | | | | -0.36 | 0.36 | | | |
| 30 | TE | 1 | 58644 | 0.28 | 0.32 | 0.00 | 0.72 | 0.28 | | | | -0.32 | 0.33 | | | |
| 31 | MC | 1 | 58643 | 0.71 | 0.38 | 0.00 | | 0.13 | 0.06 | 0.09 | 0.71 | | -0.27 | -0.24 | -0.07 | 0.38 |

**Table H-15 (continued). Item Statistics, Social Studies Grade 4**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | MC | 1 | 58620 | 0.71 | 0.45 | 0.00 | | 0.09 | 0.10 | 0.11 | 0.71 | | -0.18 | -0.20 | -0.29 | 0.45 |
| 33 | TE | 1 | 58636 | 0.43 | 0.43 | 0.00 | 0.57 | 0.43 | | | | -0.43 | 0.43 | | | |
| 34 | MC | 1 | 58570 | 0.74 | 0.49 | 0.00 | | 0.74 | 0.10 | 0.12 | 0.04 | | 0.49 | -0.31 | -0.26 | -0.19 |
| 35 | MC | 1 | 58566 | 0.51 | 0.45 | 0.00 | | 0.51 | 0.18 | 0.15 | 0.15 | | 0.46 | -0.21 | -0.25 | -0.15 |
| 36 | MC | 1 | 58603 | 0.79 | 0.49 | 0.00 | | 0.06 | 0.07 | 0.79 | 0.07 | | -0.25 | -0.26 | 0.49 | -0.27 |
| 37 | MC | 1 | 58630 | 0.70 | 0.43 | 0.00 | | 0.12 | 0.69 | 0.08 | 0.10 | | -0.17 | 0.43 | -0.30 | -0.20 |
| 38 | TE | 1 | 58603 | 0.38 | 0.30 | 0.00 | 0.62 | 0.38 | | | | -0.29 | 0.30 | | | |
| 39 | MC | 1 | 58635 | 0.57 | 0.36 | 0.00 | | 0.16 | 0.56 | 0.21 | 0.06 | | -0.17 | 0.36 | -0.17 | -0.19 |
| 40 | MC | 1 | 58526 | 0.55 | 0.27 | 0.00 | | 0.15 | 0.55 | 0.19 | 0.11 | | -0.11 | 0.27 | -0.17 | -0.09 |

**Table H-16. Item Statistics, Social Studies Grade 8**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | MC | 1 | 62485 | 0.75 | 0.51 | 0.00 | | 0.04 | 0.75 | 0.07 | 0.14 | | -0.19 | 0.51 | -0.29 | -0.31 |
| 2 | MC | 1 | 62437 | 0.41 | 0.32 | 0.00 | | 0.16 | 0.41 | 0.14 | 0.29 | | -0.05 | 0.32 | -0.22 | -0.14 |
| 3 | MC | 1 | 62413 | 0.56 | 0.56 | 0.00 | | 0.18 | 0.17 | 0.10 | 0.56 | | -0.22 | -0.32 | -0.25 | 0.56 |
| 4 | MS | 1 | 62418 | 0.25 | 0.26 | 0.00 | 0.75 | 0.25 | | | | -0.25 | 0.26 | | | |
| 5 | MC | 1 | 62418 | 0.64 | 0.49 | 0.00 | | 0.09 | 0.64 | 0.13 | 0.14 | | -0.18 | 0.49 | -0.29 | -0.25 |
| 6 | MC | 1 | 62372 | 0.78 | 0.51 | 0.00 | | 0.05 | 0.77 | 0.10 | 0.07 | | -0.24 | 0.51 | -0.32 | -0.23 |
| 7 | MC | 1 | 62385 | 0.52 | 0.40 | 0.00 | | 0.52 | 0.13 | 0.21 | 0.14 | | 0.40 | -0.25 | -0.18 | -0.11 |
| 8 | MC | 1 | 62344 | 0.39 | 0.38 | 0.00 | | 0.39 | 0.17 | 0.27 | 0.17 | | 0.38 | -0.22 | -0.12 | -0.12 |
| 9 | MC | 1 | 62428 | 0.67 | 0.54 | 0.00 | | 0.67 | 0.14 | 0.09 | 0.11 | | 0.54 | -0.29 | -0.25 | -0.26 |
| 10 | MC | 1 | 62405 | 0.80 | 0.41 | 0.00 | | 0.04 | 0.80 | 0.13 | 0.03 | | -0.18 | 0.41 | -0.26 | -0.23 |
| 11 | MC | 1 | 62384 | 0.53 | 0.42 | 0.00 | | 0.53 | 0.20 | 0.13 | 0.14 | | 0.43 | -0.17 | -0.26 | -0.16 |
| 12 | MC | 1 | 62363 | 0.52 | 0.52 | 0.00 | | 0.10 | 0.52 | 0.20 | 0.18 | | -0.06 | 0.52 | -0.31 | -0.29 |
| 13 | MC | 1 | 62388 | 0.71 | 0.54 | 0.00 | | 0.06 | 0.13 | 0.10 | 0.71 | | -0.21 | -0.29 | -0.33 | 0.55 |
| 14 | MC | 1 | 62308 | 0.75 | 0.47 | 0.00 | | 0.10 | 0.08 | 0.75 | 0.07 | | -0.22 | -0.31 | 0.47 | -0.20 |
| 15 | MC | 1 | 62341 | 0.72 | 0.40 | 0.00 | | 0.71 | 0.13 | 0.07 | 0.08 | | 0.40 | -0.16 | -0.27 | -0.19 |
| 16 | MC | 1 | 62323 | 0.75 | 0.39 | 0.00 | | 0.75 | 0.03 | 0.16 | 0.06 | | 0.39 | -0.21 | -0.17 | -0.28 |
| 17 | MC | 1 | 62324 | 0.46 | 0.44 | 0.00 | | 0.46 | 0.18 | 0.19 | 0.16 | | 0.44 | -0.11 | -0.29 | -0.15 |
| 18 | MC | 1 | 62346 | 0.72 | 0.38 | 0.00 | | 0.04 | 0.18 | 0.05 | 0.72 | | -0.24 | -0.14 | -0.29 | 0.38 |
| 19 | TE | 1 | 62331 | 0.50 | 0.60 | 0.00 | 0.50 | 0.50 | | | | -0.59 | 0.60 | | | |
| 20 | MC | 1 | 62293 | 0.50 | 0.43 | 0.00 | | 0.50 | 0.17 | 0.25 | 0.07 | | 0.44 | -0.30 | -0.14 | -0.16 |
| 21 | MC | 1 | 62322 | 0.79 | 0.54 | 0.00 | | 0.06 | 0.11 | 0.05 | 0.79 | | -0.27 | -0.34 | -0.26 | 0.54 |
| 22 | MC | 1 | 62263 | 0.83 | 0.45 | 0.00 | | 0.05 | 0.08 | 0.83 | 0.04 | | -0.22 | -0.31 | 0.45 | -0.20 |
| 23 | MC | 1 | 62247 | 0.36 | 0.41 | 0.00 | | 0.17 | 0.17 | 0.29 | 0.36 | | -0.13 | -0.29 | -0.08 | 0.41 |
| 24 | MC | 1 | 62271 | 0.55 | 0.35 | 0.00 | | 0.55 | 0.06 | 0.22 | 0.16 | | 0.35 | -0.19 | -0.08 | -0.25 |
| 25 | TE | 1 | 62279 | 0.77 | 0.46 | 0.00 | 0.23 | 0.77 | | | | -0.46 | 0.46 | | | |
| 26 | MC | 1 | 62218 | 0.65 | 0.56 | 0.00 | | 0.65 | 0.09 | 0.13 | 0.12 | | 0.56 | -0.33 | -0.32 | -0.19 |
| 27 | MC | 1 | 62281 | 0.55 | 0.28 | 0.00 | | 0.22 | 0.18 | 0.55 | 0.04 | | 0.01 | -0.24 | 0.28 | -0.25 |
| 28 | MC | 1 | 62222 | 0.63 | 0.49 | 0.00 | | 0.63 | 0.11 | 0.12 | 0.14 | | 0.50 | -0.24 | -0.30 | -0.19 |
| 29 | MC | 1 | 62195 | 0.48 | 0.39 | 0.00 | | 0.18 | 0.48 | 0.21 | 0.13 | | -0.13 | 0.39 | -0.23 | -0.15 |
| 30 | MC | 1 | 62144 | 0.63 | 0.44 | 0.00 | | 0.21 | 0.63 | 0.11 | 0.06 | | -0.17 | 0.44 | -0.27 | -0.25 |
| 31 | MC | 1 | 62206 | 0.44 | 0.27 | 0.00 | | 0.15 | 0.44 | 0.14 | 0.27 | | -0.07 | 0.27 | -0.24 | -0.06 |

**Table H-16 (continued). Item Statistics, Social Studies Grade 8**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 32 | TE | 1 | 62172 | 0.48 | 0.45 | 0.00 | 0.52 | 0.48 | | | | -0.44 | 0.45 | | | |
| 33 | TE | 1 | 62159 | 0.63 | 0.47 | 0.00 | 0.37 | 0.63 | | | | -0.47 | 0.48 | | | |
| 34 | TE | 1 | 62190 | 0.57 | 0.42 | 0.00 | 0.43 | 0.57 | | | | -0.42 | 0.42 | | | |
| 35 | MC | 1 | 62169 | 0.48 | 0.37 | 0.00 | | 0.21 | 0.48 | 0.21 | 0.09 | | -0.01 | 0.38 | -0.28 | -0.23 |
| 36 | MC | 1 | 62184 | 0.69 | 0.45 | 0.00 | | 0.08 | 0.15 | 0.68 | 0.08 | | -0.17 | -0.25 | 0.45 | -0.26 |
| 37 | MC | 1 | 62212 | 0.48 | 0.33 | 0.00 | | 0.23 | 0.15 | 0.48 | 0.14 | | -0.05 | -0.24 | 0.33 | -0.16 |
| 38 | MC | 1 | 62171 | 0.57 | 0.43 | 0.00 | | 0.10 | 0.57 | 0.18 | 0.14 | | -0.27 | 0.43 | -0.22 | -0.13 |
| 39 | MC | 1 | 62178 | 0.59 | 0.45 | 0.00 | | 0.11 | 0.18 | 0.58 | 0.12 | | -0.23 | -0.21 | 0.45 | -0.21 |
| 40 | TE | 1 | 61980 | 0.46 | 0.37 | 0.01 | 0.54 | 0.46 | | | | -0.36 | 0.37 | | | |

## Table H-17. Item Statistics, Social Studies Grade 10

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | TE | 1 | 58876 | 0.70 | 0.42 | 0.00 | 0.30 | 0.70 | | | | -0.42 | 0.43 | | | |
| 2 | MC | 1 | 58998 | 0.63 | 0.44 | 0.00 | | 0.11 | 0.63 | 0.06 | 0.20 | | -0.15 | 0.45 | -0.22 | -0.28 |
| 3 | MC | 1 | 58991 | 0.77 | 0.32 | 0.00 | | 0.10 | 0.07 | 0.77 | 0.06 | | -0.14 | -0.15 | 0.32 | -0.22 |
| 4 | MC | 1 | 58943 | 0.63 | 0.39 | 0.00 | | 0.13 | 0.20 | 0.63 | 0.04 | | -0.19 | -0.24 | 0.39 | -0.13 |
| 5 | MC | 1 | 58939 | 0.41 | 0.28 | 0.00 | | 0.12 | 0.41 | 0.19 | 0.27 | | -0.04 | 0.29 | -0.34 | 0.01 |
| 6 | MC | 1 | 58971 | 0.70 | 0.35 | 0.00 | | 0.07 | 0.17 | 0.70 | 0.06 | | -0.18 | -0.17 | 0.36 | -0.22 |
| 7 | MC | 1 | 58931 | 0.62 | 0.41 | 0.00 | | 0.12 | 0.62 | 0.18 | 0.08 | | -0.06 | 0.41 | -0.30 | -0.23 |
| 8 | MC | 1 | 58838 | 0.48 | 0.44 | 0.00 | | 0.20 | 0.13 | 0.19 | 0.48 | | -0.17 | -0.23 | -0.18 | 0.44 |
| 9 | MC | 1 | 58832 | 0.70 | 0.46 | 0.00 | | 0.10 | 0.70 | 0.15 | 0.06 | | -0.22 | 0.47 | -0.31 | -0.14 |
| 10 | MC | 1 | 58835 | 0.76 | 0.27 | 0.00 | | 0.02 | 0.11 | 0.11 | 0.75 | | -0.12 | -0.20 | -0.11 | 0.28 |
| 11 | MC | 1 | 58835 | 0.72 | 0.41 | 0.00 | | 0.07 | 0.10 | 0.11 | 0.72 | | -0.17 | -0.18 | -0.27 | 0.42 |
| 12 | MC | 1 | 58819 | 0.54 | 0.55 | 0.00 | | 0.11 | 0.21 | 0.14 | 0.54 | | -0.22 | -0.27 | -0.25 | 0.55 |
| 13 | MC | 1 | 58780 | 0.37 | 0.34 | 0.01 | | 0.37 | 0.13 | 0.30 | 0.19 | | 0.34 | -0.19 | -0.06 | -0.17 |
| 14 | MC | 1 | 58707 | 0.57 | 0.38 | 0.01 | | 0.56 | 0.13 | 0.19 | 0.10 | | 0.39 | -0.16 | -0.13 | -0.26 |
| 15 | MC | 1 | 58712 | 0.72 | 0.54 | 0.01 | | 0.09 | 0.09 | 0.10 | 0.72 | | -0.25 | -0.28 | -0.28 | 0.54 |
| 16 | MC | 1 | 58626 | 0.48 | 0.48 | 0.01 | | 0.48 | 0.21 | 0.17 | 0.14 | | 0.48 | -0.21 | -0.27 | -0.13 |
| 17 | MC | 1 | 58662 | 0.33 | 0.27 | 0.01 | | 0.33 | 0.16 | 0.43 | 0.08 | | 0.27 | -0.22 | 0.04 | -0.22 |
| 18 | MC | 1 | 58713 | 0.66 | 0.43 | 0.01 | | 0.09 | 0.15 | 0.65 | 0.10 | | -0.23 | -0.21 | 0.43 | -0.19 |
| 19 | MC | 1 | 58692 | 0.52 | 0.29 | 0.01 | | 0.13 | 0.26 | 0.52 | 0.09 | | -0.08 | -0.12 | 0.29 | -0.20 |
| 20 | MS | 1 | 58702 | 0.24 | 0.27 | 0.01 | 0.76 | 0.24 | | | | -0.25 | 0.27 | | | |
| 21 | MC | 1 | 58130 | 0.59 | 0.39 | 0.00 | | 0.27 | 0.13 | 0.58 | 0.02 | | -0.26 | -0.17 | 0.39 | -0.14 |
| 22 | MS | 1 | 58047 | 0.42 | 0.50 | 0.00 | 0.58 | 0.42 | | | | -0.49 | 0.50 | | | |
| 23 | MC | 1 | 58031 | 0.69 | 0.53 | 0.00 | | 0.07 | 0.09 | 0.15 | 0.68 | | -0.14 | -0.28 | -0.35 | 0.53 |
| 24 | MC | 1 | 58012 | 0.62 | 0.38 | 0.00 | | 0.07 | 0.21 | 0.62 | 0.09 | | -0.18 | -0.20 | 0.39 | -0.18 |
| 25 | ESR | 1 | 58055 | 0.23 | 0.37 | 0.00 | 0.76 | 0.23 | | | | -0.36 | 0.37 | | | |
| 26 | TE | 1 | 57833 | 0.37 | 0.50 | 0.01 | 0.63 | 0.37 | | | | -0.49 | 0.50 | | | |
| 27 | MC | 1 | 57861 | 0.55 | 0.42 | 0.01 | | 0.26 | 0.11 | 0.08 | 0.55 | | -0.10 | -0.27 | -0.27 | 0.42 |
| 28 | EBSR | 1 | 57958 | 0.35 | 0.50 | 0.00 | 0.65 | 0.34 | | | | -0.49 | 0.50 | | | |
| 29 | MC | 1 | 57849 | 0.42 | 0.35 | 0.01 | | 0.15 | 0.19 | 0.41 | 0.23 | | -0.14 | -0.24 | 0.36 | -0.06 |
| 30 | MC | 1 | 57823 | 0.55 | 0.38 | 0.01 | | 0.20 | 0.12 | 0.55 | 0.12 | | -0.08 | -0.20 | 0.38 | -0.26 |
| 31 | TE | 1 | 57738 | 0.32 | 0.26 | 0.01 | 0.67 | 0.32 | | | | -0.24 | 0.26 | | | |

**Table H-17 (continued). Item Statistics, Social Studies Grade 10**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
|---|----|---|-------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|
|   |   |   |       |      |      |      | 8    | 9    | 10   | 11   | 12   | 13    | 14    | 15    | 16    | 17    |
| 32 | MC | 1 | 57701 | 0.58 | 0.57 | 0.01 |      | 0.06 | 0.23 | 0.13 | 0.57 |       | -0.24 | -0.28 | -0.30 | 0.57  |
| 33 | MC | 1 | 57679 | 0.49 | 0.43 | 0.01 |      | 0.48 | 0.14 | 0.25 | 0.11 |       | 0.44  | -0.20 | -0.16 | -0.23 |
| 34 | TE | 1 | 57712 | 0.39 | 0.44 | 0.01 | 0.60 | 0.39 |      |      |      | -0.42 | 0.44  |       |       |       |
| 35 | MC | 1 | 57731 | 0.53 | 0.49 | 0.01 |      | 0.09 | 0.19 | 0.18 | 0.53 |       | -0.17 | -0.25 | -0.24 | 0.49  |
| 36 | TE | 1 | 57672 | 0.37 | 0.47 | 0.01 | 0.62 | 0.37 |      |      |      | -0.45 | 0.47  |       |       |       |
| 37 | MC | 1 | 57712 | 0.68 | 0.36 | 0.01 |      | 0.16 | 0.68 | 0.09 | 0.07 |       | -0.11 | 0.37  | -0.30 | -0.17 |
| 38 | MC | 1 | 57725 | 0.65 | 0.28 | 0.01 |      | 0.03 | 0.08 | 0.24 | 0.64 |       | -0.17 | -0.15 | -0.13 | 0.28  |
| 39 | MC | 1 | 57702 | 0.47 | 0.31 | 0.01 |      | 0.16 | 0.47 | 0.25 | 0.11 |       | -0.06 | 0.32  | -0.20 | -0.13 |
| 40 | TE | 1 | 57549 | 0.56 | 0.43 | 0.01 | 0.43 | 0.56 |      |      |      | -0.42 | 0.44  |       |       |       |

# Appendix I
# Conditional Standard Error of Measurement with Cut Scores

# List of Figures

**Figure I-1. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 3**
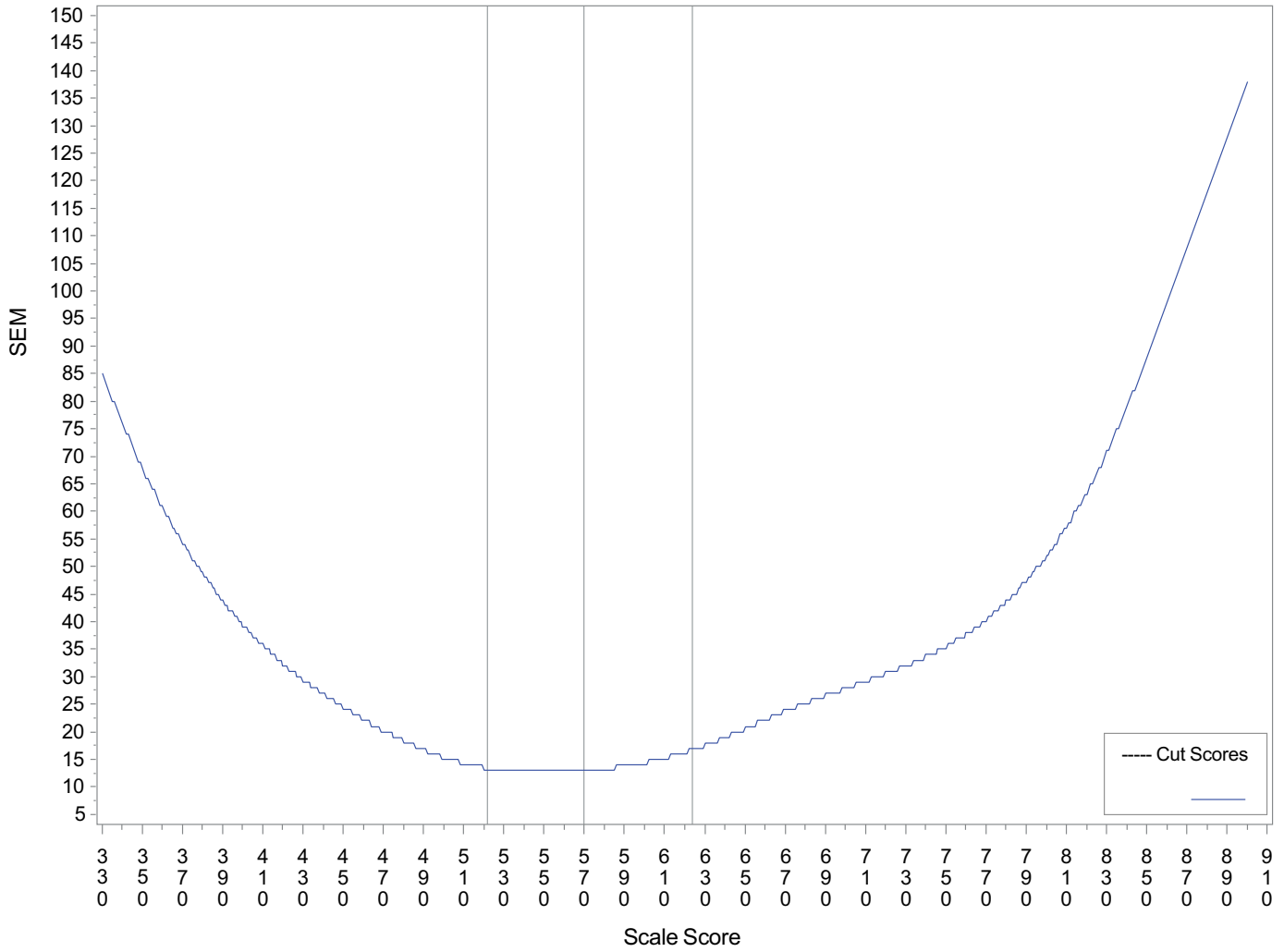
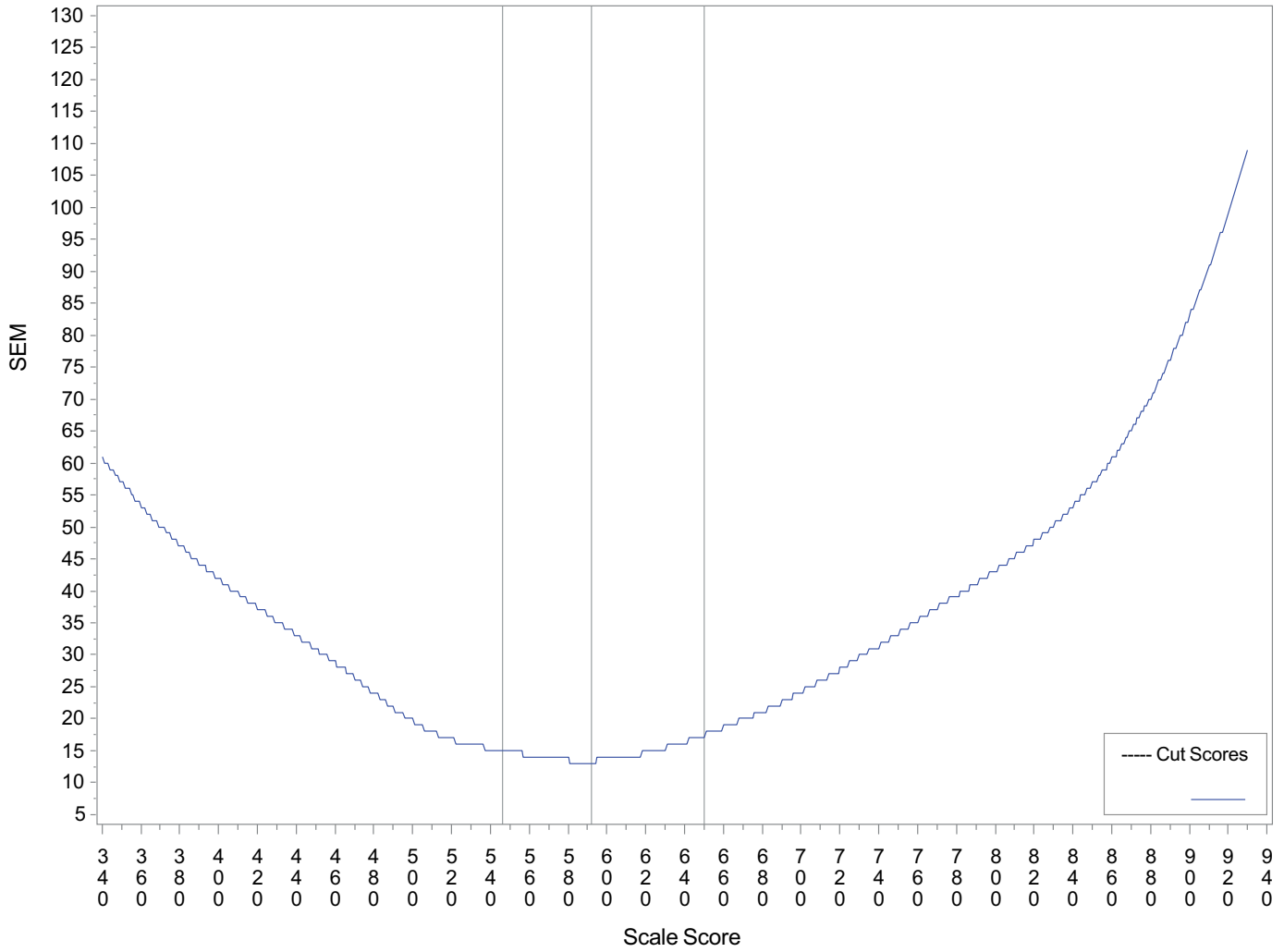**Figure I-2. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 4**

**Figure I-3. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 5**
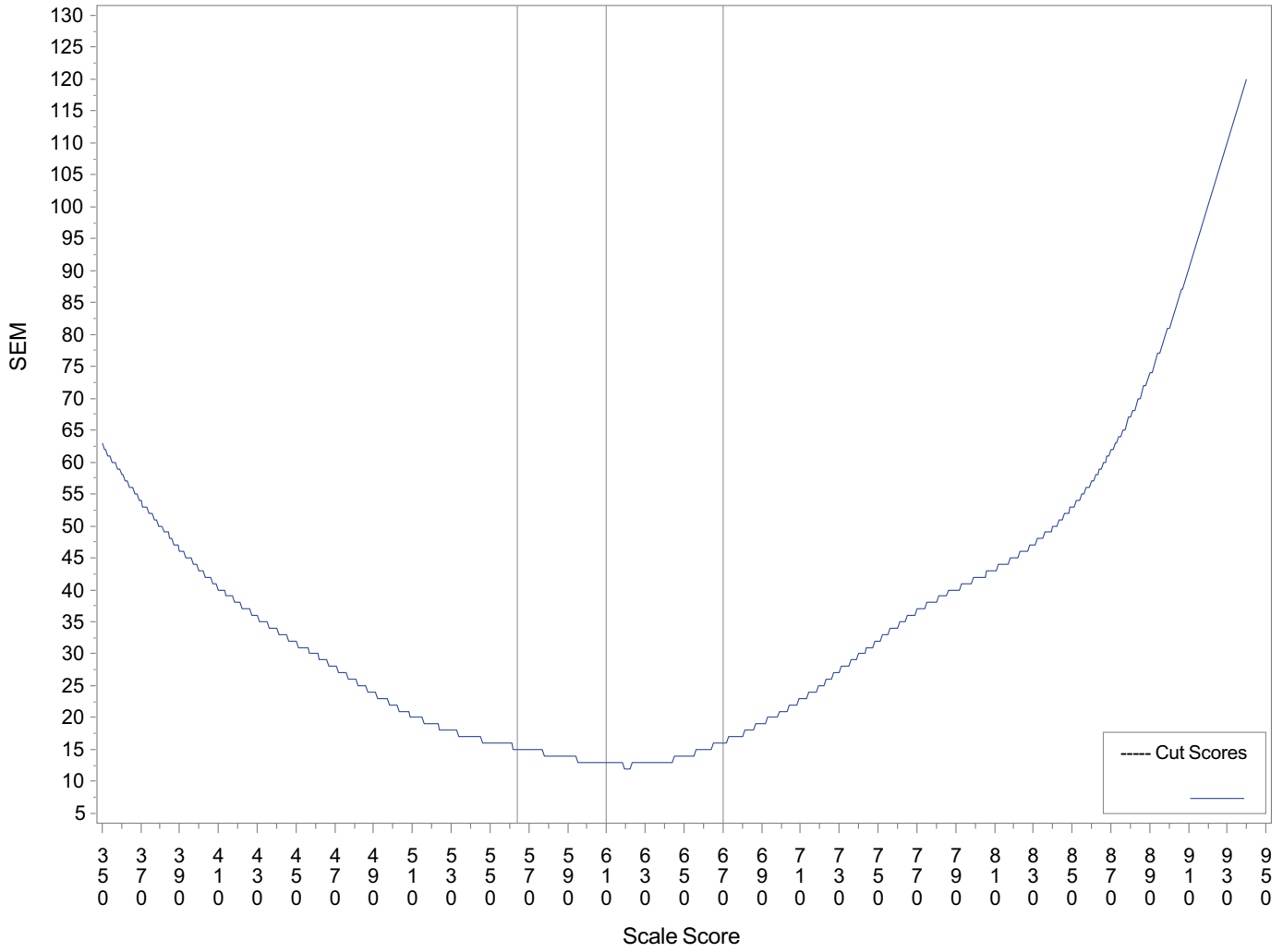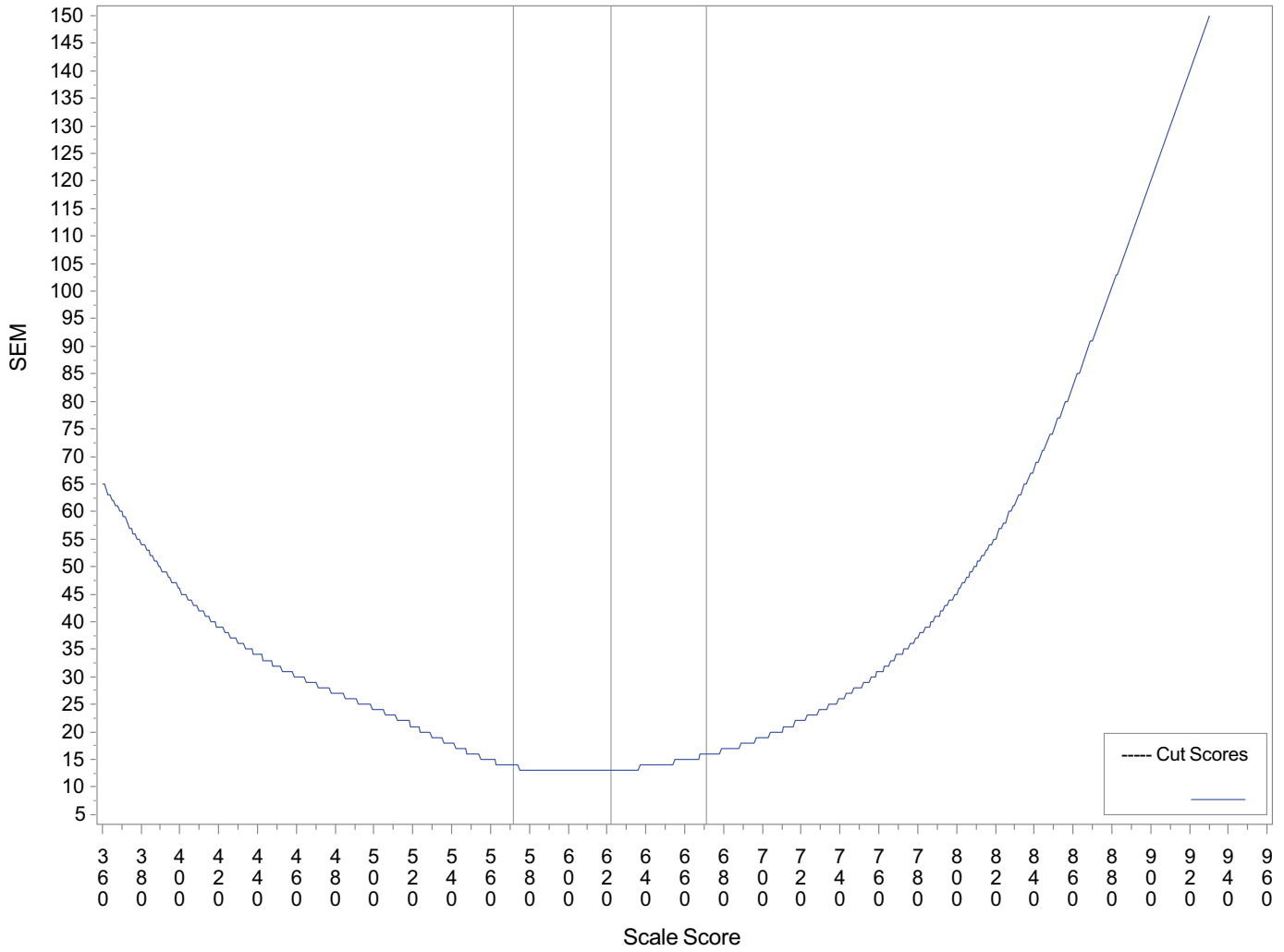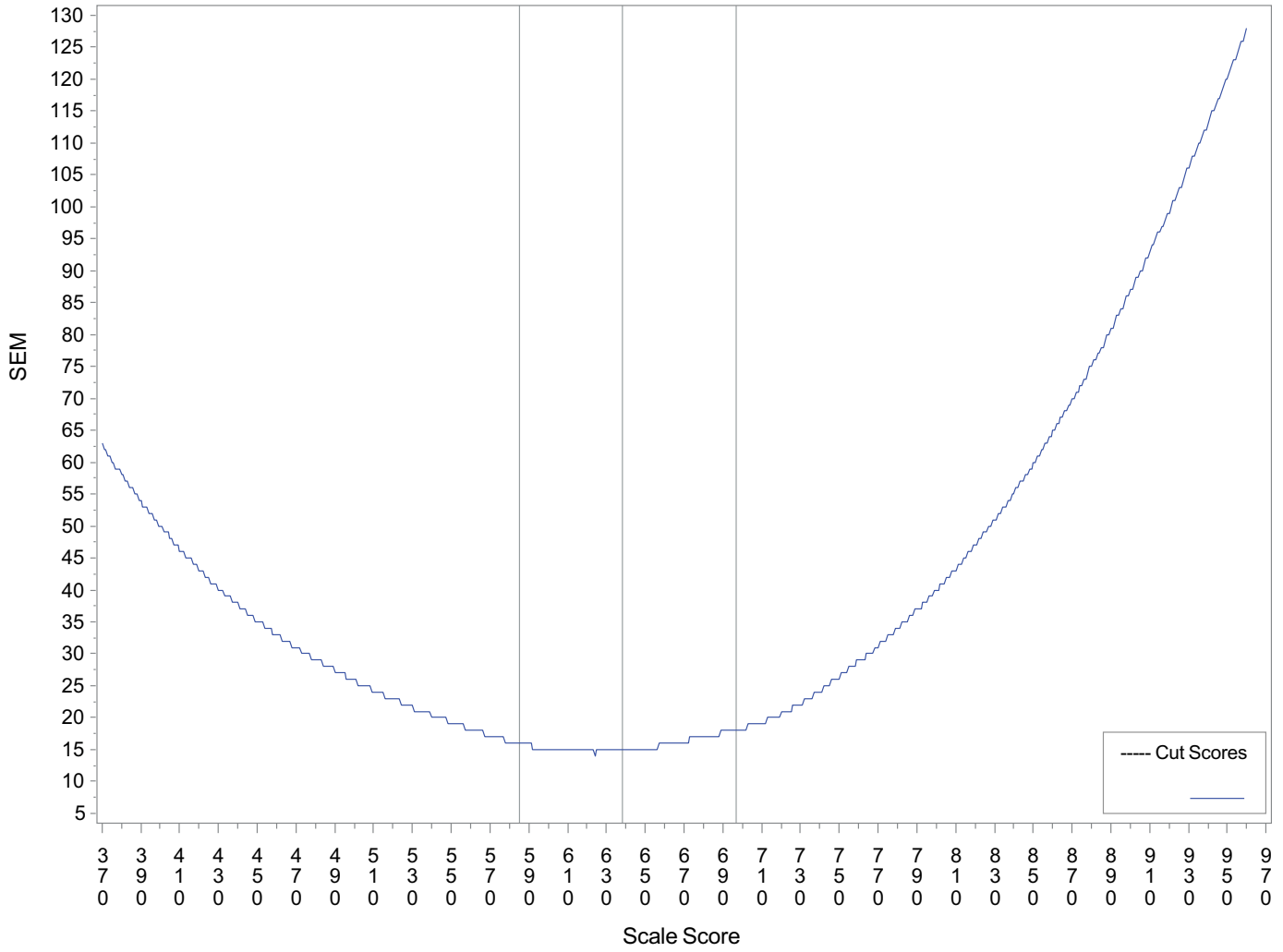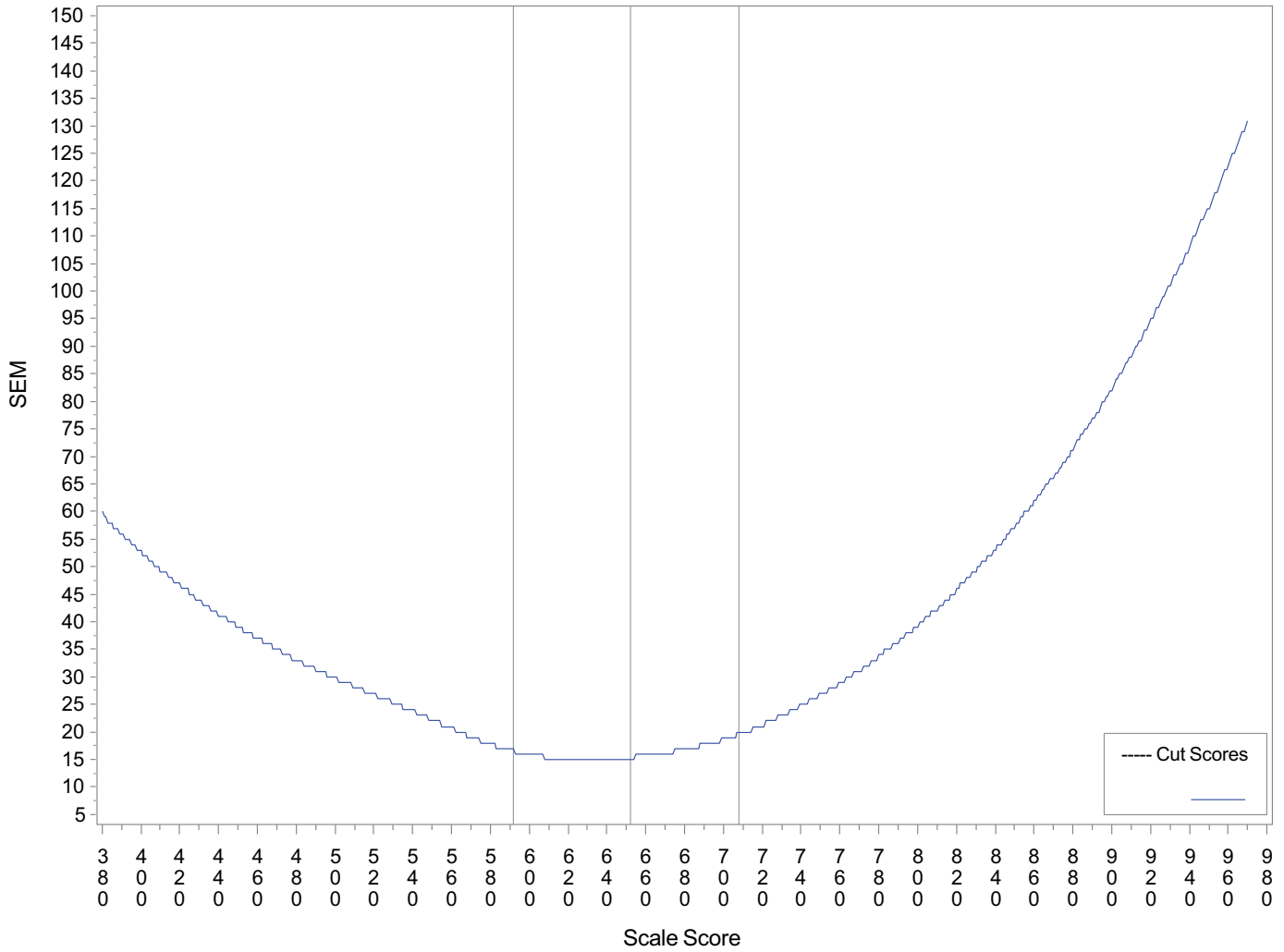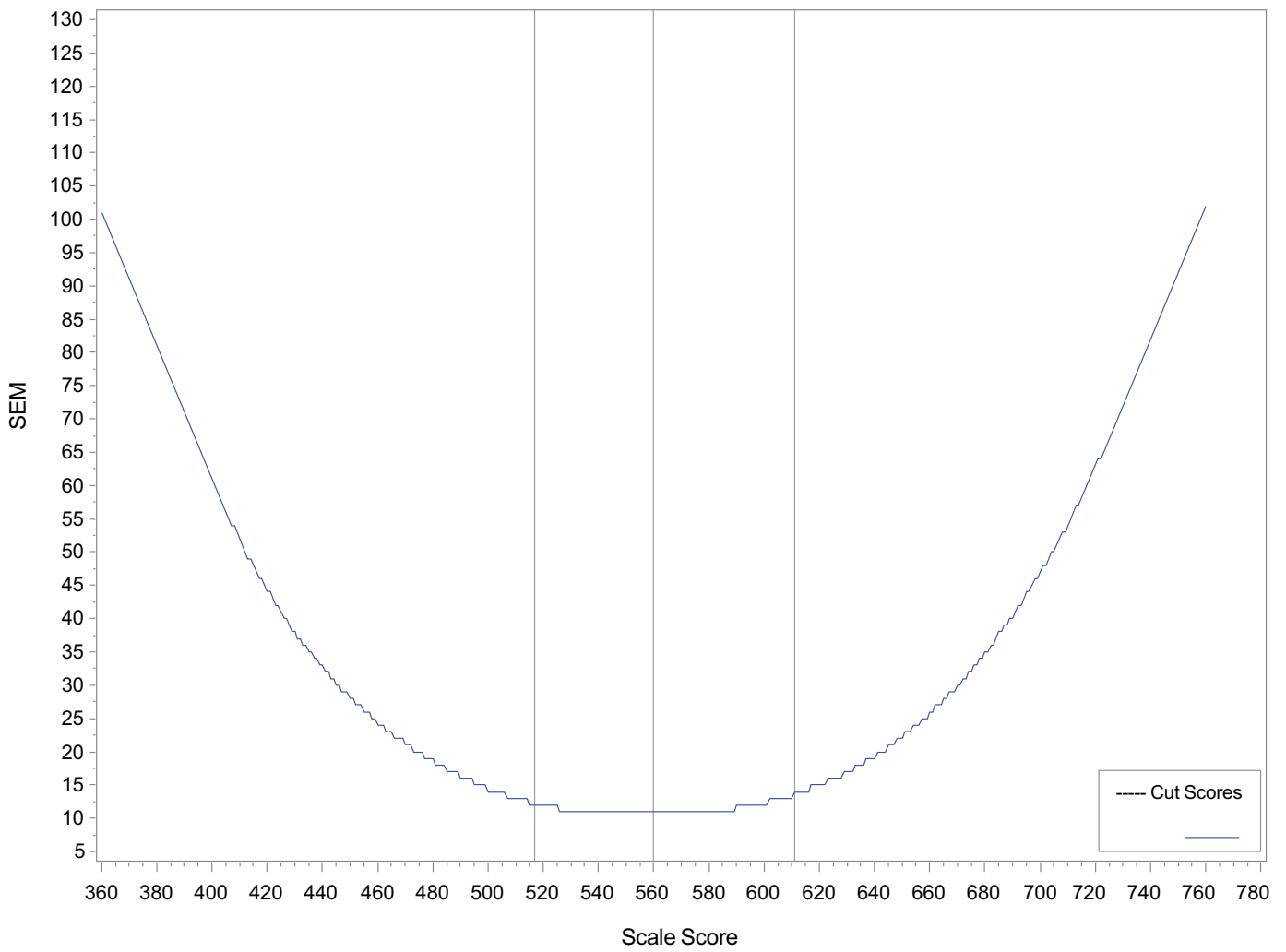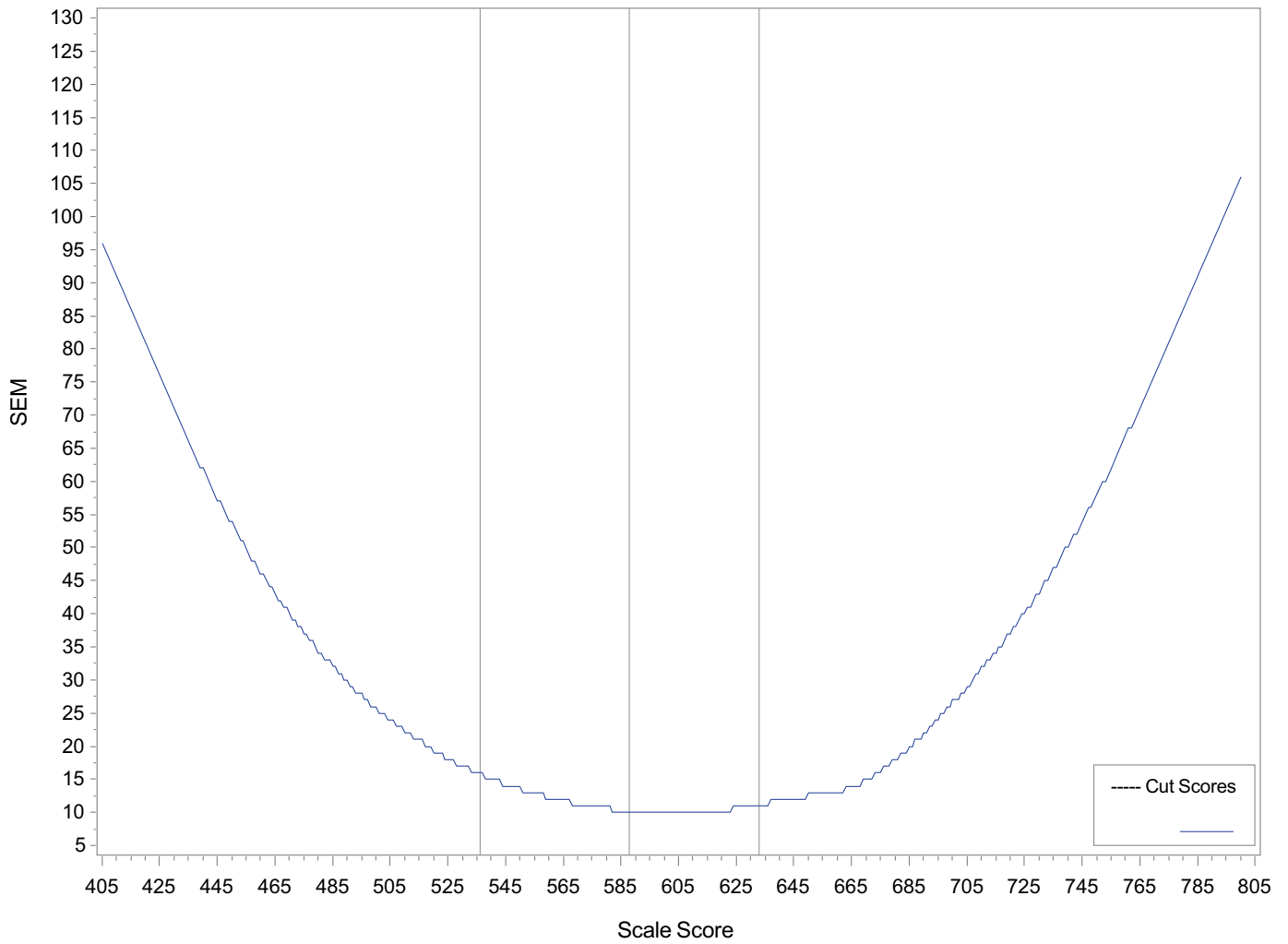
**Figure I-4. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 6**

**Figure I-5. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 7**

**Figure I-6. Conditional Standard Error of Measurement with Cut Scores, English Language Arts Grade 8**

**Figure I-7. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 3**

**Figure I-8. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 4**
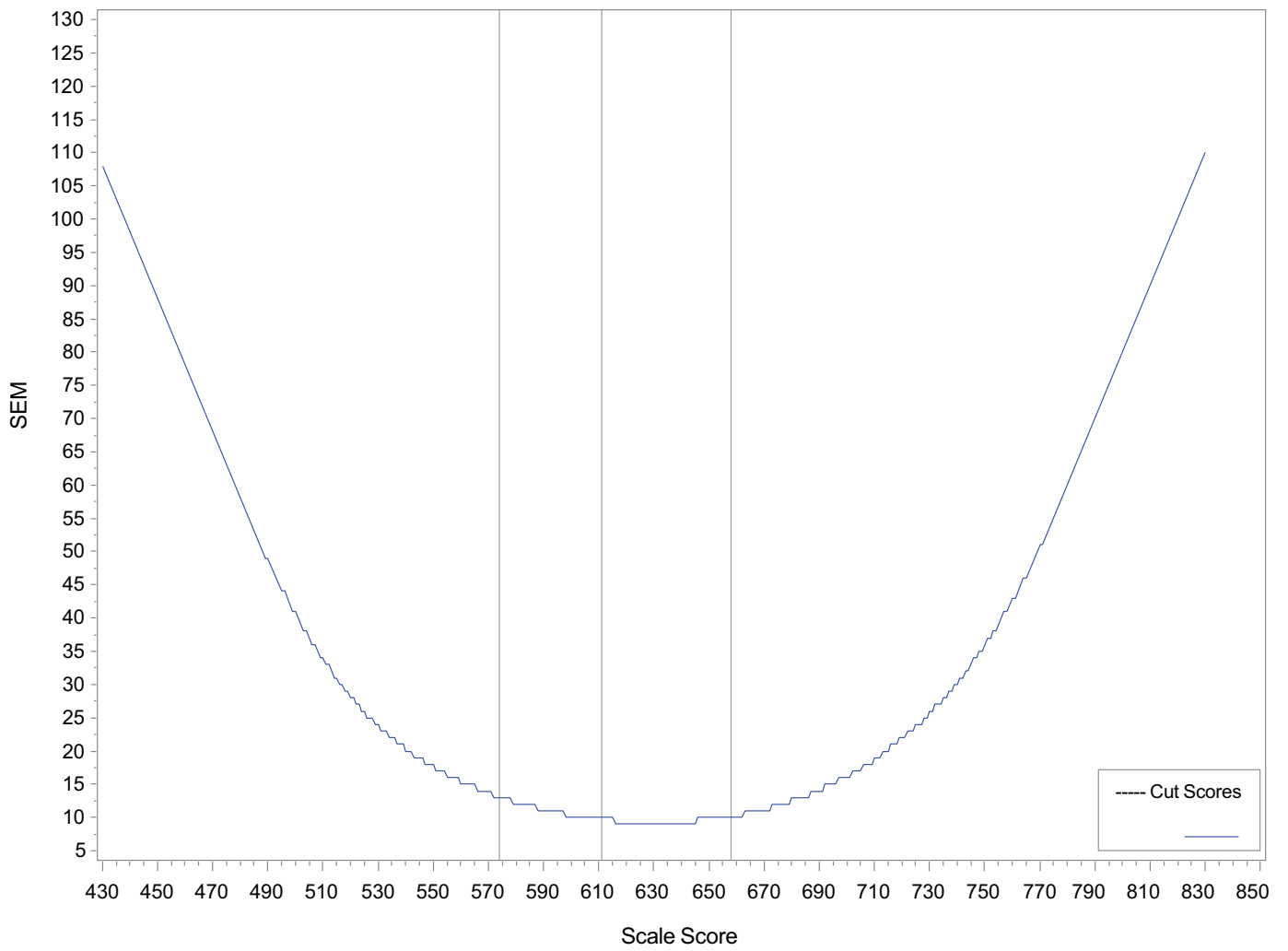
**Figure I-9. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 5**

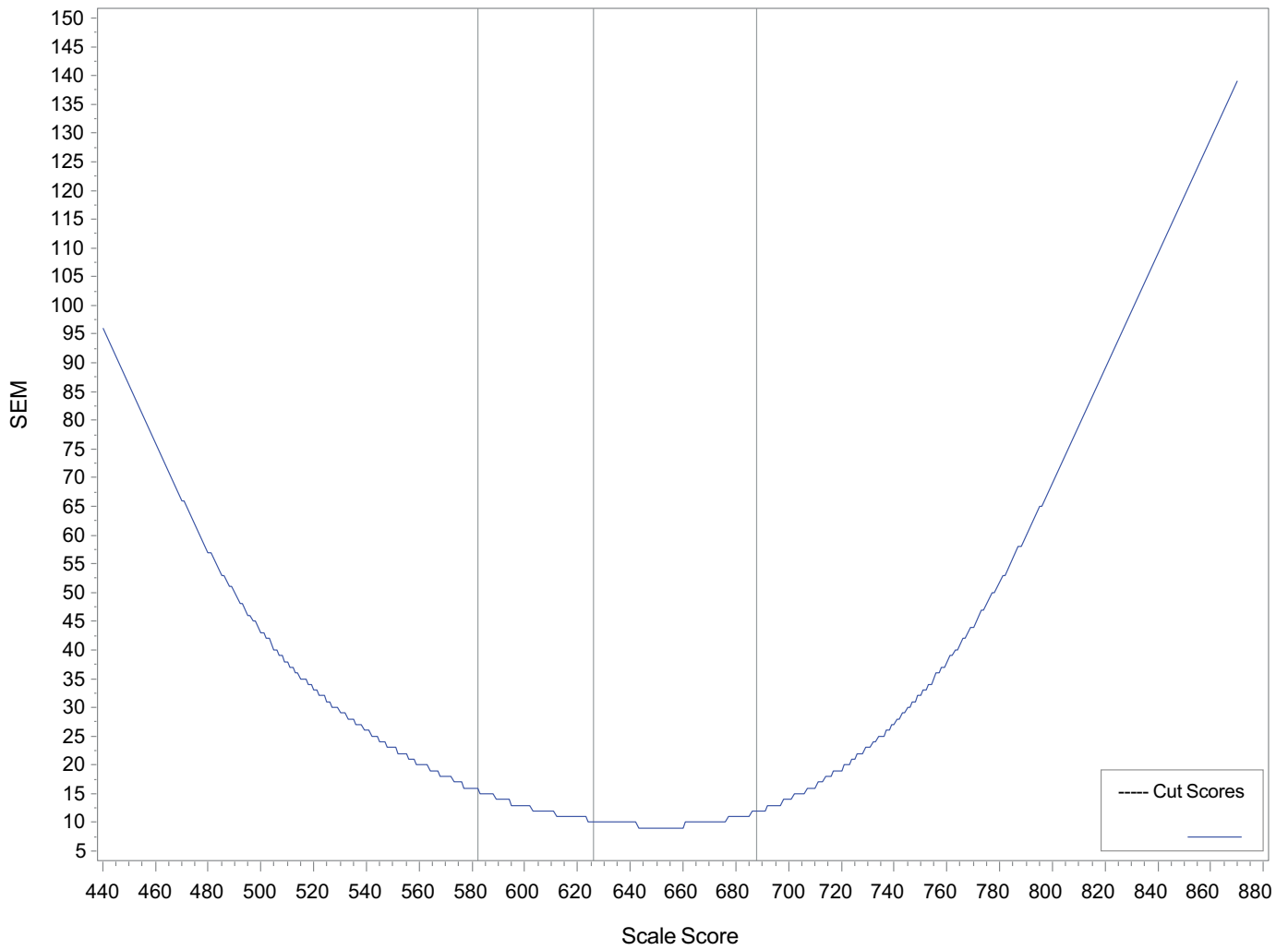**Figure I-10. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 6**

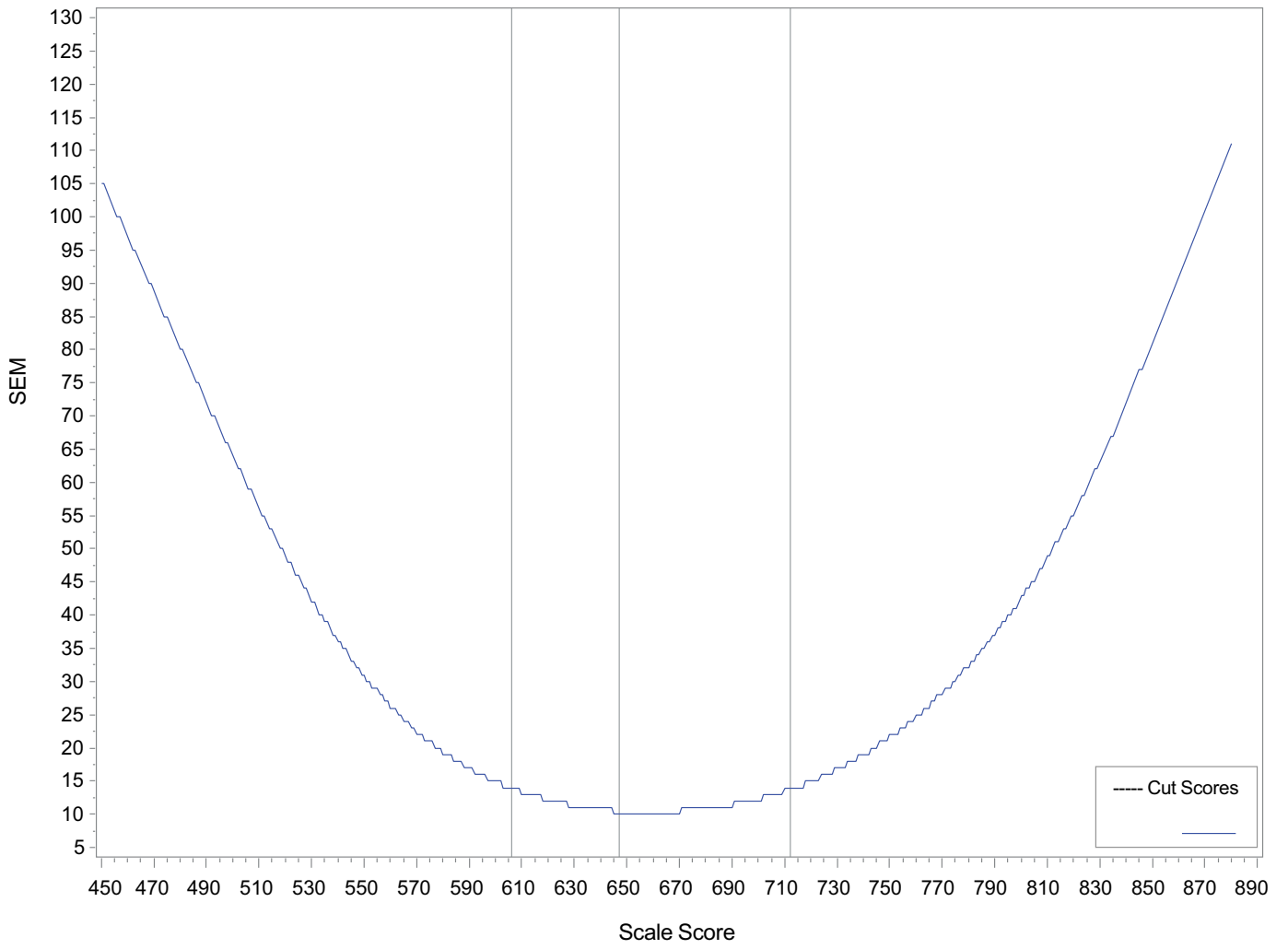**Figure I-11. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 7**

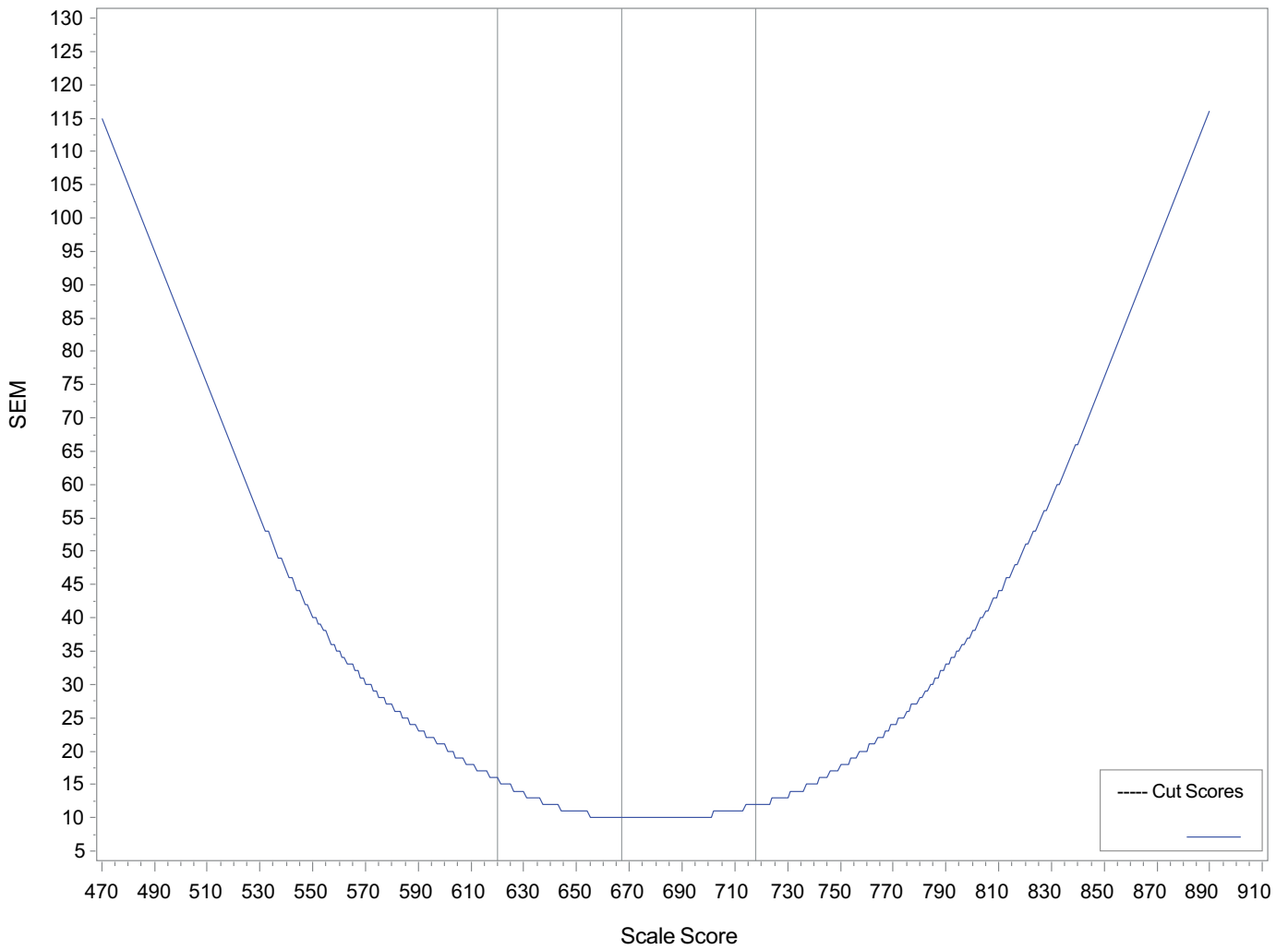**Figure I-12. Conditional Standard Error of Measurement with Cut Scores, Mathematics Grade 8**

**Figure I-13. Conditional Standard Error of Measurement with Cut Scores, Science Grade 4**
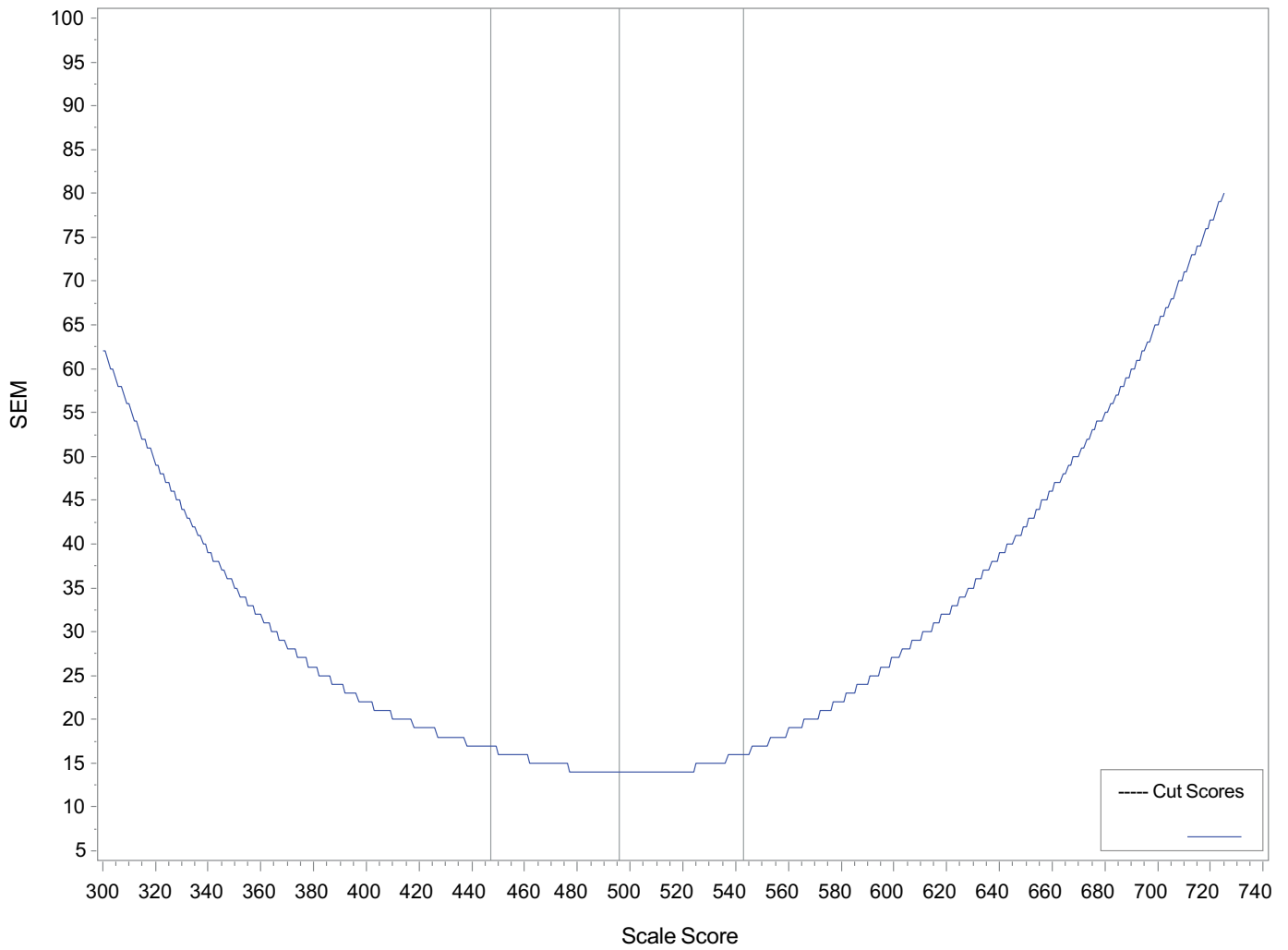
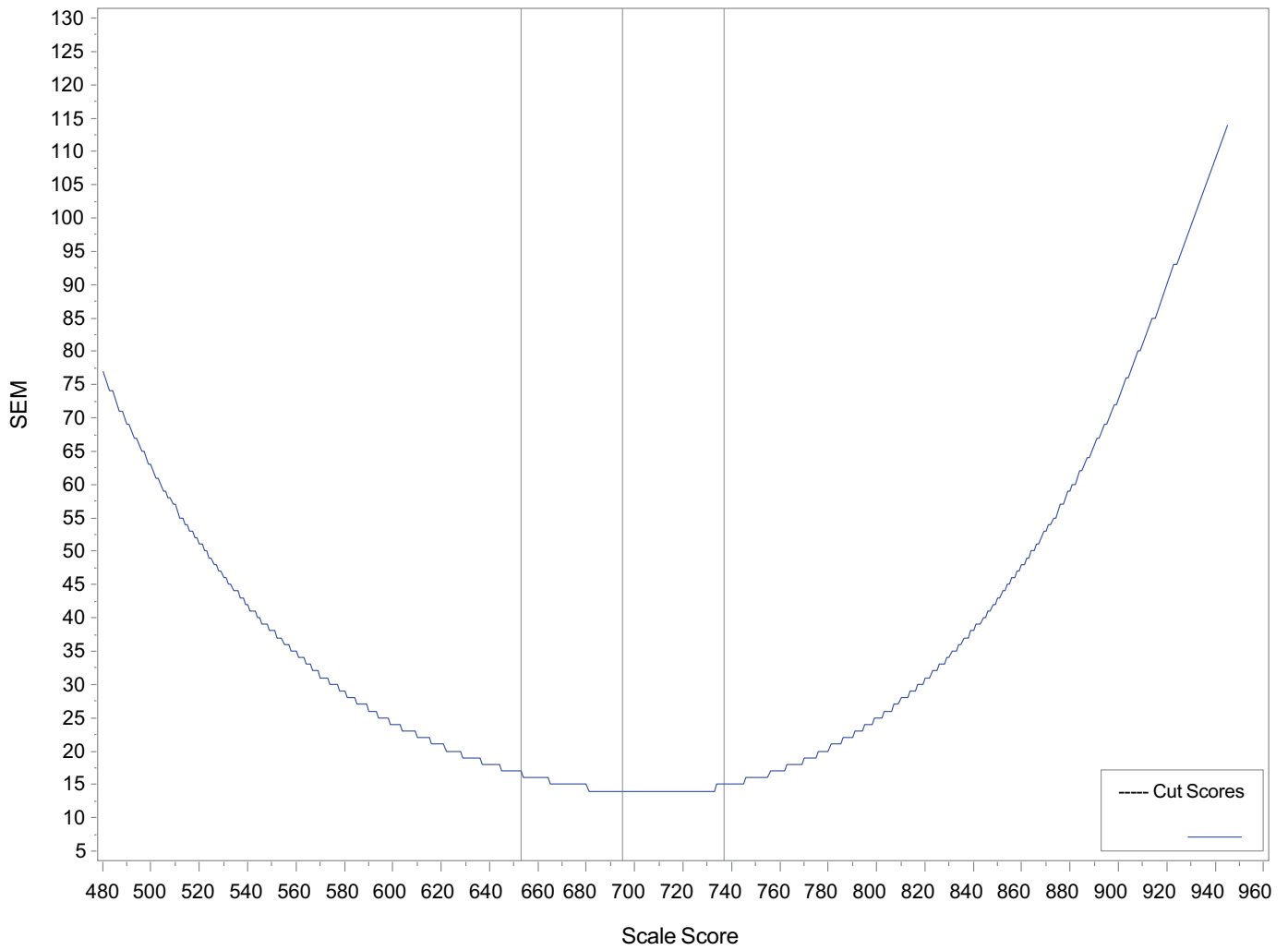**Figure I-14. Conditional Standard Error of Measurement with Cut Scores, Science Grade 8**

**Figure I-15. Conditional Standard Error of Measurement with Cut Scores, Social Studies Grade 4**
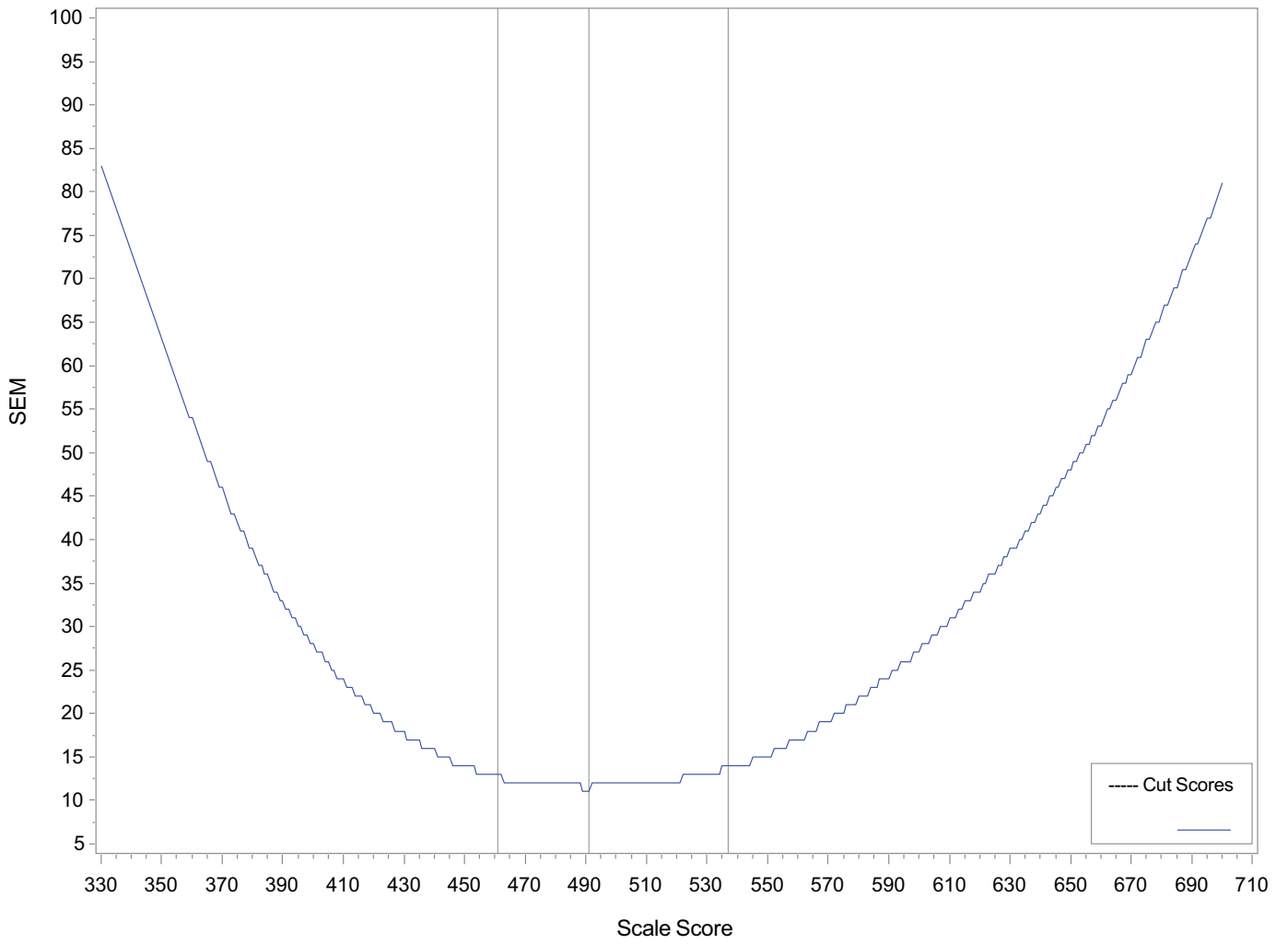
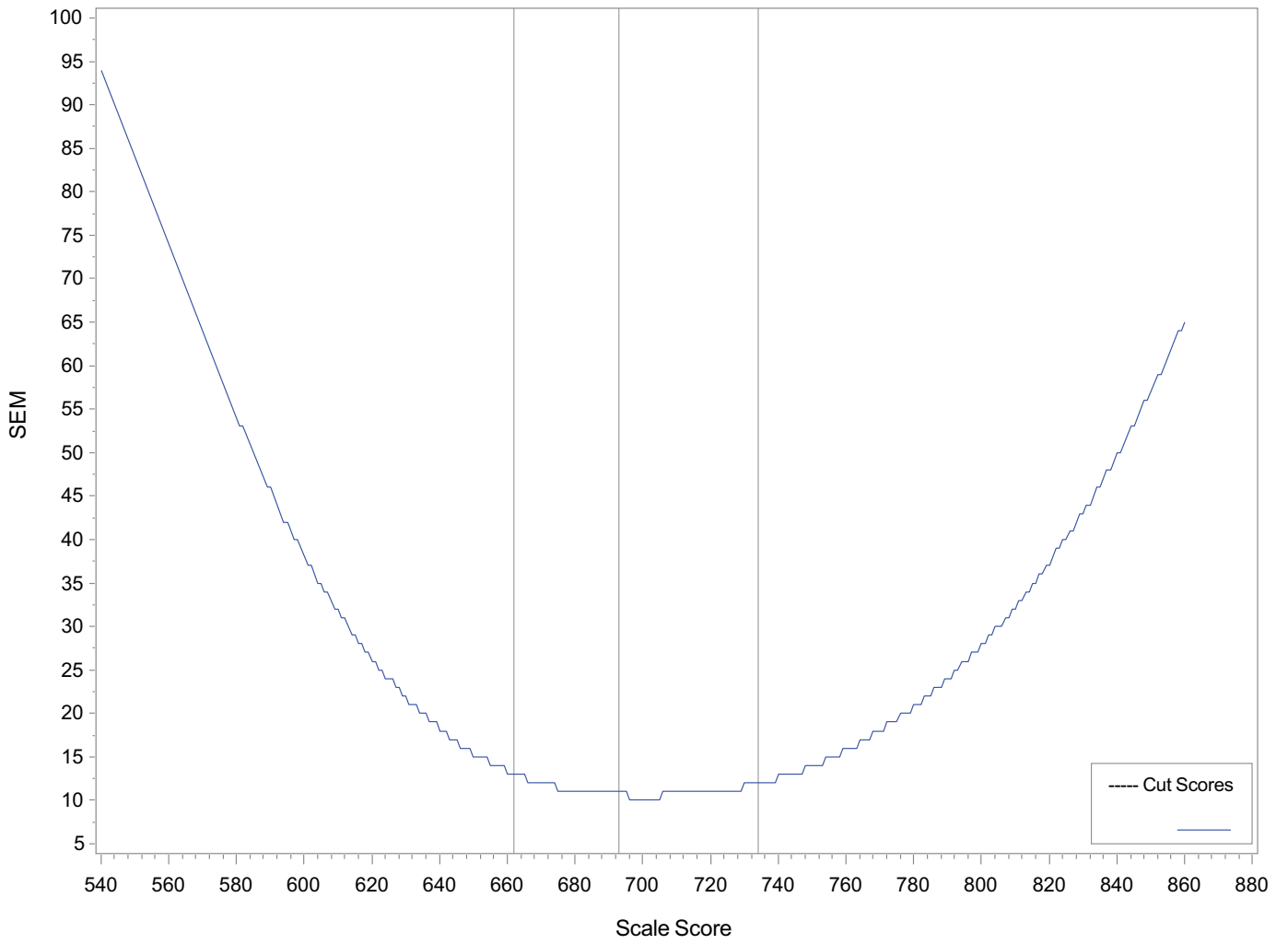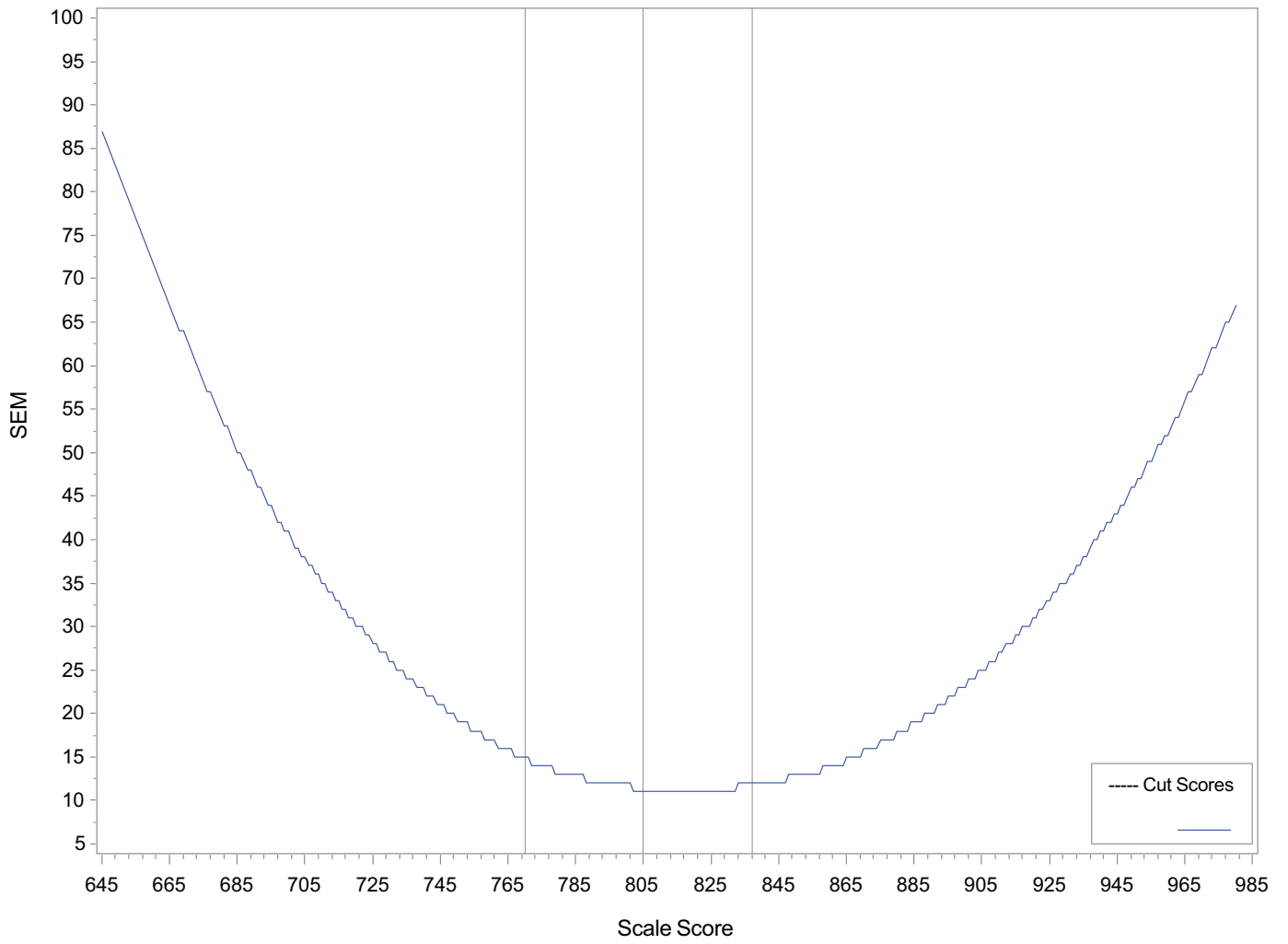**Figure I-16. Conditional Standard Error of Measurement with Cut Scores, Social Studies Grade 8**

**Figure I-17. Conditional Standard Error of Measurement with Cut Scores, Social Studies Grade 10**

# Appendix J
# Classification Consistency and Accuracy Analysis by Subgroup

# List of Tables

**Table J-1. Indices for Classification Consistency and Accuracy, ELA Grade 3**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.90 | 0.89 | 0.94 | 0.73 |
| | | Probability of Chance | 0.62 | 0.53 | 0.86 | 0.30 |
| | | Kappa (k) | 0.73 | 0.77 | 0.59 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.81 |
| | Male | Classification Consistency (P) | 0.89 | 0.90 | 0.95 | 0.74 |
| | | Probability of Chance | 0.57 | 0.57 | 0.89 | 0.31 |
| | | Kappa (k) | 0.74 | 0.76 | 0.58 | 0.63 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.97 | 0.82 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.88 | 0.93 | 0.72 |
| | | Probability of Chance | 0.69 | 0.51 | 0.84 | 0.31 |
| | | Kappa (k) | 0.70 | 0.75 | 0.58 | 0.59 |
| | | Classification Accuracy | 0.94 | 0.91 | 0.95 | 0.80 |
| | African American | Classification Consistency (P) | 0.87 | 0.96 | 0.99 | 0.82 |
| | | Probability of Chance | 0.54 | 0.85 | 0.98 | 0.50 |
| | | Kappa (k) | 0.71 | 0.73 | 0.52 | 0.64 |
| | | Classification Accuracy | 0.91 | 0.97 | 0.99 | 0.87 |
| | Hispanic | Classification Consistency (P) | 0.86 | 0.92 | 0.98 | 0.76 |
| | | Probability of Chance | 0.51 | 0.69 | 0.95 | 0.36 |
| | | Kappa (k) | 0.71 | 0.73 | 0.56 | 0.62 |
| | | Classification Accuracy | 0.90 | 0.94 | 0.99 | 0.83 |
| | Asian | Classification Consistency (P) | 0.87 | 0.91 | 0.95 | 0.73 |
| | | Probability of Chance | 0.57 | 0.57 | 0.87 | 0.30 |
| | | Kappa (k) | 0.70 | 0.78 | 0.63 | 0.61 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.97 | 0.81 |
| | American Indian | Classification Consistency (P) | 0.86 | 0.92 | 0.99 | 0.77 |
| | | Probability of Chance | 0.51 | 0.74 | 0.97 | 0.38 |
| | | Kappa (k) | 0.72 | 0.70 | 0.62 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.99 | 0.84 |
| | Two or More | Classification Consistency (P) | 0.88 | 0.90 | 0.96 | 0.74 |
| | | Probability of Chance | 0.57 | 0.58 | 0.89 | 0.31 |
| | | Kappa (k) | 0.73 | 0.77 | 0.62 | 0.63 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.97 | 0.82 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.84 | 0.93 | 0.99 | 0.76 |
| | | Probability of Chance | 0.50 | 0.77 | 0.98 | 0.40 |
| | | Kappa (k) | 0.68 | 0.70 | 0.48 | 0.61 |
| | | Classification Accuracy | 0.89 | 0.95 | 0.99 | 0.83 |
| Disability Status | Yes | Classification Consistency (P) | 0.86 | 0.94 | 0.98 | 0.78 |
| | | Probability of Chance | 0.50 | 0.75 | 0.96 | 0.40 |
| | | Kappa (k) | 0.72 | 0.75 | 0.56 | 0.63 |
| | | Classification Accuracy | 0.90 | 0.96 | 0.99 | 0.85 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.86 | 0.92 | 0.98 | 0.76 |
| | | Probability of Chance | 0.51 | 0.68 | 0.95 | 0.35 |
| | | Kappa (k) | 0.72 | 0.74 | 0.53 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.98 | 0.83 |

**Table J-1 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 3**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.91 | 0.93 | 0.99 | 0.82 |
| | | Probability of Chance | 0.51 | 0.77 | 0.99 | 0.43 |
| | | Kappa (k) | 0.81 | 0.68 | 0.13 | 0.69 |
| | | Classification Accuracy | 0.94 | 0.95 | 0.99 | 0.88 |

**Table J-2. Indices for Classification Consistency and Accuracy, ELA Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.72 |
| | | Probability of Chance | 0.63 | 0.51 | 0.81 | 0.28 |
| | | Kappa (k) | 0.73 | 0.78 | 0.64 | 0.61 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.95 | 0.80 |
| | Male | Classification Consistency (P) | 0.92 | 0.93 | 0.93 | 0.79 |
| | | Probability of Chance | 0.56 | 0.51 | 0.76 | 0.27 |
| | | Kappa (k) | 0.82 | 0.86 | 0.71 | 0.71 |
| | | Classification Accuracy | 0.94 | 0.95 | 0.91 | 0.80 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.88 | 0.92 | 0.71 |
| | | Probability of Chance | 0.70 | 0.50 | 0.79 | 0.29 |
| | | Kappa (k) | 0.70 | 0.75 | 0.63 | 0.59 |
| | | Classification Accuracy | 0.94 | 0.91 | 0.94 | 0.79 |
| | African American | Classification Consistency (P) | 0.86 | 0.95 | 0.99 | 0.80 |
| | | Probability of Chance | 0.53 | 0.78 | 0.97 | 0.46 |
| | | Kappa (k) | 0.71 | 0.75 | 0.56 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |
| | Hispanic | Classification Consistency (P) | 0.86 | 0.91 | 0.97 | 0.75 |
| | | Probability of Chance | 0.51 | 0.64 | 0.93 | 0.34 |
| | | Kappa (k) | 0.72 | 0.75 | 0.61 | 0.62 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.98 | 0.82 |
| | Asian | Classification Consistency (P) | 0.90 | 0.89 | 0.94 | 0.73 |
| | | Probability of Chance | 0.60 | 0.52 | 0.82 | 0.28 |
| | | Kappa (k) | 0.75 | 0.77 | 0.68 | 0.63 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.81 |
| | American Indian | Classification Consistency (P) | 0.87 | 0.91 | 0.98 | 0.75 |
| | | Probability of Chance | 0.51 | 0.65 | 0.94 | 0.34 |
| | | Kappa (k) | 0.73 | 0.74 | 0.59 | 0.62 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.98 | 0.83 |
| | Two or More | Classification Consistency (P) | 0.89 | 0.89 | 0.94 | 0.73 |
| | | Probability of Chance | 0.57 | 0.53 | 0.85 | 0.29 |
| | | Kappa (k) | 0.74 | 0.77 | 0.62 | 0.62 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.96 | 0.81 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.85 | 0.91 | 0.99 | 0.75 |
| | | Probability of Chance | 0.50 | 0.72 | 0.97 | 0.38 |
| | | Kappa (k) | 0.70 | 0.70 | 0.51 | 0.60 |
| | | Classification Accuracy | 0.89 | 0.94 | 0.99 | 0.82 |
| Disability Status | Yes | Classification Consistency (P) | 0.87 | 0.93 | 0.98 | 0.78 |
| | | Probability of Chance | 0.50 | 0.71 | 0.95 | 0.40 |
| | | Kappa (k) | 0.73 | 0.77 | 0.60 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.85 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.87 | 0.91 | 0.97 | 0.75 |
| | | Probability of Chance | 0.51 | 0.62 | 0.93 | 0.33 |
| | | Kappa (k) | 0.73 | 0.75 | 0.59 | 0.62 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.98 | 0.82 |

**Table J-2 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.87 | 0.95 | 0.96 | 0.79 |
| | | Probability of Chance | 0.50 | 0.61 | 0.86 | 0.35 |
| | | Kappa (k) | 0.75 | 0.87 | 0.72 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.97 | 0.97 | 0.85 |

**Table J-3. Indices for Classification Consistency and Accuracy, ELA Grade 5**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.91 | 0.89 | 0.94 | 0.74 |
| | | Probability of Chance | 0.63 | 0.51 | 0.84 | 0.29 |
| | | Kappa (k) | 0.75 | 0.78 | 0.62 | 0.63 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.82 |
| | Male | Classification Consistency (P) | 0.89 | 0.90 | 0.95 | 0.75 |
| | | Probability of Chance | 0.57 | 0.53 | 0.88 | 0.30 |
| | | Kappa (k) | 0.76 | 0.78 | 0.59 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.97 | 0.82 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.88 | 0.93 | 0.73 |
| | | Probability of Chance | 0.68 | 0.50 | 0.83 | 0.30 |
| | | Kappa (k) | 0.72 | 0.77 | 0.60 | 0.61 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.96 | 0.81 |
| | African American | Classification Consistency (P) | 0.87 | 0.95 | 0.99 | 0.81 |
| | | Probability of Chance | 0.54 | 0.80 | 0.98 | 0.48 |
| | | Kappa (k) | 0.72 | 0.73 | 0.54 | 0.64 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.87 |
| | Hispanic | Classification Consistency (P) | 0.87 | 0.90 | 0.98 | 0.75 |
| | | Probability of Chance | 0.52 | 0.62 | 0.94 | 0.33 |
| | | Kappa (k) | 0.74 | 0.75 | 0.58 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.98 | 0.83 |
| | Asian | Classification Consistency (P) | 0.89 | 0.90 | 0.95 | 0.73 |
| | | Probability of Chance | 0.58 | 0.52 | 0.85 | 0.29 |
| | | Kappa (k) | 0.74 | 0.78 | 0.64 | 0.62 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.96 | 0.81 |
| | American Indian | Classification Consistency (P) | 0.86 | 0.90 | 0.98 | 0.74 |
| | | Probability of Chance | 0.51 | 0.65 | 0.95 | 0.35 |
| | | Kappa (k) | 0.72 | 0.71 | 0.58 | 0.61 |
| | | Classification Accuracy | 0.90 | 0.93 | 0.99 | 0.82 |
| | Two or More | Classification Consistency (P) | 0.89 | 0.90 | 0.95 | 0.74 |
| | | Probability of Chance | 0.57 | 0.53 | 0.88 | 0.30 |
| | | Kappa (k) | 0.75 | 0.78 | 0.60 | 0.63 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.97 | 0.82 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.85 | 0.92 | 0.99 | 0.77 |
| | | Probability of Chance | 0.50 | 0.76 | 0.99 | 0.41 |
| | | Kappa (k) | 0.71 | 0.68 | 0.48 | 0.61 |
| | | Classification Accuracy | 0.90 | 0.95 | 1.00 | 0.84 |
| Disability Status | Yes | Classification Consistency (P) | 0.87 | 0.95 | 0.99 | 0.81 |
| | | Probability of Chance | 0.52 | 0.76 | 0.97 | 0.45 |
| | | Kappa (k) | 0.74 | 0.77 | 0.56 | 0.65 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.87 | 0.91 | 0.98 | 0.76 |
| | | Probability of Chance | 0.51 | 0.63 | 0.95 | 0.34 |
| | | Kappa (k) | 0.74 | 0.75 | 0.53 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.99 | 0.83 |

**Table J-3 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 5**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.88 | 0.92 | 0.99 | 0.79 |
| | | Probability of Chance | 0.50 | 0.76 | 0.99 | 0.42 |
| | | Kappa (k) | 0.76 | 0.66 | 0.09 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.99 | 0.85 |

**Table J-4. Indices for Classification Consistency and Accuracy, ELA Grade 6**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.92 | 0.88 | 0.93 | 0.73 |
| | | Probability of Chance | 0.65 | 0.51 | 0.80 | 0.28 |
| | | Kappa (k) | 0.76 | 0.76 | 0.63 | 0.62 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.94 | 0.80 |
| | Male | Classification Consistency (P) | 0.91 | 0.89 | 0.95 | 0.74 |
| | | Probability of Chance | 0.58 | 0.55 | 0.87 | 0.30 |
| | | Kappa (k) | 0.78 | 0.75 | 0.60 | 0.64 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.81 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.94 | 0.91 | 0.92 | 0.77 |
| | | Probability of Chance | 0.66 | 0.50 | 0.70 | 0.26 |
| | | Kappa (k) | 0.83 | 0.82 | 0.74 | 0.69 |
| | | Classification Accuracy | 0.96 | 0.92 | 0.91 | 0.79 |
| | African American | Classification Consistency (P) | 0.88 | 0.94 | 0.99 | 0.81 |
| | | Probability of Chance | 0.51 | 0.79 | 0.97 | 0.44 |
| | | Kappa (k) | 0.76 | 0.72 | 0.55 | 0.66 |
| | | Classification Accuracy | 0.92 | 0.96 | 0.99 | 0.86 |
| | Hispanic | Classification Consistency (P) | 0.91 | 0.92 | 0.96 | 0.80 |
| | | Probability of Chance | 0.51 | 0.61 | 0.89 | 0.32 |
| | | Kappa (k) | 0.82 | 0.80 | 0.67 | 0.70 |
| | | Classification Accuracy | 0.94 | 0.94 | 0.96 | 0.84 |
| | Asian | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.73 |
| | | Probability of Chance | 0.64 | 0.52 | 0.81 | 0.29 |
| | | Kappa (k) | 0.73 | 0.77 | 0.64 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.95 | 0.80 |
| | American Indian | Classification Consistency (P) | 0.88 | 0.91 | 0.98 | 0.78 |
| | | Probability of Chance | 0.51 | 0.71 | 0.96 | 0.37 |
| | | Kappa (k) | 0.76 | 0.70 | 0.51 | 0.65 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.99 | 0.84 |
| | Two or More | Classification Consistency (P) | 0.90 | 0.90 | 0.95 | 0.75 |
| | | Probability of Chance | 0.58 | 0.55 | 0.85 | 0.29 |
| | | Kappa (k) | 0.77 | 0.77 | 0.63 | 0.64 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.96 | 0.81 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.86 | 0.95 | 0.99 | 0.79 |
| | | Probability of Chance | 0.51 | 0.81 | 0.98 | 0.43 |
| | | Kappa (k) | 0.72 | 0.71 | 0.49 | 0.64 |
| | | Classification Accuracy | 0.90 | 0.96 | 0.99 | 0.85 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.95 | 0.99 | 0.82 |
| | | Probability of Chance | 0.53 | 0.82 | 0.97 | 0.48 |
| | | Kappa (k) | 0.74 | 0.74 | 0.60 | 0.66 |
| | | Classification Accuracy | 0.91 | 0.97 | 0.99 | 0.87 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.89 | 0.91 | 0.97 | 0.76 |
| | | Probability of Chance | 0.52 | 0.65 | 0.93 | 0.34 |
| | | Kappa (k) | 0.76 | 0.74 | 0.56 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.98 | 0.83 |

**Table J-4 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 6**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.89 | 0.90 | 0.97 | 0.76 |
| | | Probability of Chance | 0.51 | 0.65 | 0.94 | 0.35 |
| | | Kappa (k) | 0.78 | 0.70 | 0.52 | 0.63 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.98 | 0.83 |

**Table J-5. Indices for Classification Consistency and Accuracy, ELA Grade 7**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.91 | 0.88 | 0.93 | 0.72 |
| | | Probability of Chance | 0.64 | 0.51 | 0.81 | 0.28 |
| | | Kappa (k) | 0.74 | 0.75 | 0.62 | 0.61 |
| | | Classification Accuracy | 0.93 | 0.91 | 0.95 | 0.79 |
| | Male | Classification Consistency (P) | 0.90 | 0.89 | 0.94 | 0.73 |
| | | Probability of Chance | 0.58 | 0.54 | 0.86 | 0.29 |
| | | Kappa (k) | 0.77 | 0.76 | 0.59 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.96 | 0.80 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.93 | 0.90 | 0.92 | 0.75 |
| | | Probability of Chance | 0.67 | 0.50 | 0.75 | 0.27 |
| | | Kappa (k) | 0.78 | 0.80 | 0.68 | 0.66 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.94 | 0.79 |
| | African American | Classification Consistency (P) | 0.88 | 0.93 | 0.99 | 0.80 |
| | | Probability of Chance | 0.51 | 0.77 | 0.96 | 0.43 |
| | | Kappa (k) | 0.75 | 0.72 | 0.58 | 0.65 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.99 | 0.85 |
| | Hispanic | Classification Consistency (P) | 0.88 | 0.91 | 0.97 | 0.76 |
| | | Probability of Chance | 0.52 | 0.63 | 0.92 | 0.33 |
| | | Kappa (k) | 0.75 | 0.74 | 0.59 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.98 | 0.82 |
| | Asian | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.72 |
| | | Probability of Chance | 0.65 | 0.51 | 0.79 | 0.28 |
| | | Kappa (k) | 0.71 | 0.78 | 0.66 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.95 | 0.80 |
| | American Indian | Classification Consistency (P) | 0.88 | 0.93 | 0.98 | 0.78 |
| | | Probability of Chance | 0.51 | 0.69 | 0.94 | 0.36 |
| | | Kappa (k) | 0.75 | 0.76 | 0.62 | 0.66 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.85 |
| | Two or More | Classification Consistency (P) | 0.89 | 0.90 | 0.95 | 0.74 |
| | | Probability of Chance | 0.56 | 0.55 | 0.86 | 0.29 |
| | | Kappa (k) | 0.76 | 0.78 | 0.62 | 0.63 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.96 | 0.81 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.85 | 0.94 | 0.99 | 0.79 |
| | | Probability of Chance | 0.51 | 0.83 | 0.98 | 0.45 |
| | | Kappa (k) | 0.70 | 0.65 | 0.51 | 0.62 |
| | | Classification Accuracy | 0.89 | 0.96 | 1.00 | 0.85 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.95 | 0.99 | 0.82 |
| | | Probability of Chance | 0.54 | 0.82 | 0.98 | 0.48 |
| | | Kappa (k) | 0.74 | 0.72 | 0.58 | 0.65 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.88 | 0.90 | 0.97 | 0.75 |
| | | Probability of Chance | 0.52 | 0.64 | 0.93 | 0.33 |
| | | Kappa (k) | 0.75 | 0.73 | 0.57 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.93 | 0.98 | 0.82 |

**Table J-5 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 7**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.90 | 0.91 | 0.97 | 0.78 |
| | | Probability of Chance | 0.50 | 0.69 | 0.95 | 0.38 |
| | | Kappa (k) | 0.79 | 0.70 | 0.50 | 0.64 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.98 | 0.84 |

**Table J-6. Indices for Classification Consistency and Accuracy, ELA Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.91 | 0.89 | 0.93 | 0.73 |
| | | Probability of Chance | 0.63 | 0.53 | 0.82 | 0.29 |
| | | Kappa (k) | 0.76 | 0.76 | 0.62 | 0.62 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.95 | 0.81 |
| | Male | Classification Consistency (P) | 0.90 | 0.89 | 0.95 | 0.75 |
| | | Probability of Chance | 0.55 | 0.58 | 0.88 | 0.31 |
| | | Kappa (k) | 0.78 | 0.75 | 0.59 | 0.64 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.97 | 0.82 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.88 | 0.93 | 0.72 |
| | | Probability of Chance | 0.65 | 0.52 | 0.82 | 0.29 |
| | | Kappa (k) | 0.75 | 0.75 | 0.60 | 0.61 |
| | | Classification Accuracy | 0.94 | 0.91 | 0.95 | 0.80 |
| | African American | Classification Consistency (P) | 0.88 | 0.94 | 0.99 | 0.81 |
| | | Probability of Chance | 0.52 | 0.79 | 0.97 | 0.44 |
| | | Kappa (k) | 0.75 | 0.72 | 0.52 | 0.65 |
| | | Classification Accuracy | 0.92 | 0.96 | 0.99 | 0.87 |
| | Hispanic | Classification Consistency (P) | 0.89 | 0.91 | 0.97 | 0.76 |
| | | Probability of Chance | 0.51 | 0.66 | 0.92 | 0.34 |
| | | Kappa (k) | 0.77 | 0.73 | 0.58 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.98 | 0.83 |
| | Asian | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.73 |
| | | Probability of Chance | 0.64 | 0.52 | 0.79 | 0.28 |
| | | Kappa (k) | 0.73 | 0.77 | 0.68 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.95 | 0.80 |
| | American Indian | Classification Consistency (P) | 0.89 | 0.92 | 0.98 | 0.79 |
| | | Probability of Chance | 0.50 | 0.73 | 0.96 | 0.39 |
| | | Kappa (k) | 0.78 | 0.70 | 0.53 | 0.66 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.99 | 0.85 |
| | Two or More | Classification Consistency (P) | 0.90 | 0.90 | 0.95 | 0.75 |
| | | Probability of Chance | 0.56 | 0.58 | 0.87 | 0.30 |
| | | Kappa (k) | 0.77 | 0.76 | 0.59 | 0.64 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.96 | 0.82 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.86 | 0.95 | 0.99 | 0.80 |
| | | Probability of Chance | 0.52 | 0.85 | 0.98 | 0.46 |
| | | Kappa (k) | 0.71 | 0.67 | 0.53 | 0.63 |
| | | Classification Accuracy | 0.90 | 0.96 | 0.99 | 0.86 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.96 | 0.99 | 0.84 |
| | | Probability of Chance | 0.56 | 0.86 | 0.98 | 0.52 |
| | | Kappa (k) | 0.74 | 0.72 | 0.54 | 0.66 |
| | | Classification Accuracy | 0.92 | 0.97 | 0.99 | 0.89 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.88 | 0.91 | 0.97 | 0.77 |
| | | Probability of Chance | 0.51 | 0.68 | 0.94 | 0.35 |
| | | Kappa (k) | 0.76 | 0.73 | 0.55 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.98 | 0.84 |

**Table J-6 (continued). Indices for Classification Consistency and Accuracy, ELA Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.86 | 0.91 | 0.98 | 0.75 |
| | | Probability of Chance | 0.50 | 0.66 | 0.89 | 0.34 |
| | | Kappa (k) | 0.73 | 0.74 | 0.77 | 0.62 |
| | | Classification Accuracy | 0.90 | 0.94 | 0.98 | 0.82 |

**Table J-7. Indices for Classification Consistency and Accuracy, Mathematics Grade 3**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.92 | 0.90 | 0.94 | 0.76 |
| | | Probability of Chance | 0.63 | 0.51 | 0.80 | 0.28 |
| | | Kappa (k) | 0.80 | 0.79 | 0.70 | 0.67 |
| | | Classification Accuracy | 0.95 | 0.93 | 0.96 | 0.83 |
| | Male | Classification Consistency (P) | 0.94 | 0.90 | 0.93 | 0.77 |
| | | Probability of Chance | 0.66 | 0.50 | 0.74 | 0.27 |
| | | Kappa (k) | 0.81 | 0.80 | 0.73 | 0.68 |
| | | Classification Accuracy | 0.96 | 0.93 | 0.95 | 0.84 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.95 | 0.89 | 0.92 | 0.75 |
| | | Probability of Chance | 0.76 | 0.51 | 0.71 | 0.29 |
| | | Kappa (k) | 0.77 | 0.77 | 0.71 | 0.65 |
| | | Classification Accuracy | 0.96 | 0.92 | 0.94 | 0.82 |
| | African American | Classification Consistency (P) | 0.89 | 0.95 | 0.99 | 0.83 |
| | | Probability of Chance | 0.52 | 0.79 | 0.97 | 0.46 |
| | | Kappa (k) | 0.78 | 0.75 | 0.62 | 0.69 |
| | | Classification Accuracy | 0.93 | 0.96 | 0.99 | 0.88 |
| | Hispanic | Classification Consistency (P) | 0.90 | 0.91 | 0.97 | 0.78 |
| | | Probability of Chance | 0.53 | 0.60 | 0.91 | 0.32 |
| | | Kappa (k) | 0.79 | 0.78 | 0.67 | 0.68 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.98 | 0.85 |
| | Asian | Classification Consistency (P) | 0.92 | 0.91 | 0.94 | 0.77 |
| | | Probability of Chance | 0.63 | 0.51 | 0.74 | 0.26 |
| | | Kappa (k) | 0.79 | 0.81 | 0.76 | 0.68 |
| | | Classification Accuracy | 0.95 | 0.93 | 0.96 | 0.83 |
| | American Indian | Classification Consistency (P) | 0.90 | 0.92 | 0.98 | 0.79 |
| | | Probability of Chance | 0.52 | 0.65 | 0.94 | 0.34 |
| | | Kappa (k) | 0.79 | 0.77 | 0.60 | 0.68 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.98 | 0.85 |
| | Two or More | Classification Consistency (P) | 0.91 | 0.90 | 0.95 | 0.77 |
| | | Probability of Chance | 0.60 | 0.53 | 0.82 | 0.28 |
| | | Kappa (k) | 0.78 | 0.79 | 0.72 | 0.67 |
| | | Classification Accuracy | 0.94 | 0.93 | 0.97 | 0.83 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.89 | 0.92 | 0.98 | 0.78 |
| | | Probability of Chance | 0.51 | 0.65 | 0.93 | 0.34 |
| | | Kappa (k) | 0.78 | 0.76 | 0.66 | 0.67 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.98 | 0.85 |
| Disability Status | Yes | Classification Consistency (P) | 0.91 | 0.93 | 0.97 | 0.81 |
| | | Probability of Chance | 0.50 | 0.64 | 0.89 | 0.34 |
| | | Kappa (k) | 0.82 | 0.81 | 0.73 | 0.71 |
| | | Classification Accuracy | 0.94 | 0.95 | 0.98 | 0.87 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.90 | 0.91 | 0.97 | 0.78 |
| | | Probability of Chance | 0.53 | 0.59 | 0.90 | 0.31 |
| | | Kappa (k) | 0.80 | 0.78 | 0.68 | 0.68 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.98 | 0.85 |

**Table J-7 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 3**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|-------|----------|---------|-------|-------|-------|----------|
| Accommodation Use | Yes | Classification Consistency (P) | 0.91 | 0.96 | 1.00 | 0.87 |
| | | Probability of Chance | 0.63 | 0.88 | 0.99 | 0.60 |
| | | Kappa (k) | 0.76 | 0.67 | 0.83 | 0.67 |
| | | Classification Accuracy | 0.94 | 0.97 | 1.00 | 0.91 |

**Table J-8. Indices for Classification Consistency and Accuracy, Mathematics Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.91 | 0.91 | 0.95 | 0.77 |
| | | Probability of Chance | 0.64 | 0.52 | 0.79 | 0.28 |
| | | Kappa (k) | 0.75 | 0.81 | 0.74 | 0.68 |
| | | Classification Accuracy | 0.94 | 0.94 | 0.96 | 0.83 |
| | Male | Classification Consistency (P) | 0.93 | 0.91 | 0.94 | 0.78 |
| | | Probability of Chance | 0.66 | 0.50 | 0.74 | 0.27 |
| | | Kappa (k) | 0.78 | 0.83 | 0.75 | 0.69 |
| | | Classification Accuracy | 0.95 | 0.94 | 0.95 | 0.84 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.94 | 0.90 | 0.92 | 0.76 |
| | | Probability of Chance | 0.78 | 0.50 | 0.71 | 0.29 |
| | | Kappa (k) | 0.73 | 0.80 | 0.74 | 0.67 |
| | | Classification Accuracy | 0.96 | 0.93 | 0.95 | 0.83 |
| | African American | Classification Consistency (P) | 0.87 | 0.96 | 0.99 | 0.82 |
| | | Probability of Chance | 0.52 | 0.84 | 0.97 | 0.47 |
| | | Kappa (k) | 0.72 | 0.76 | 0.69 | 0.66 |
| | | Classification Accuracy | 0.90 | 0.97 | 0.99 | 0.87 |
| | Hispanic | Classification Consistency (P) | 0.88 | 0.93 | 0.98 | 0.78 |
| | | Probability of Chance | 0.52 | 0.65 | 0.91 | 0.33 |
| | | Kappa (k) | 0.74 | 0.80 | 0.71 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.84 |
| | Asian | Classification Consistency (P) | 0.89 | 0.92 | 0.95 | 0.76 |
| | | Probability of Chance | 0.65 | 0.51 | 0.74 | 0.27 |
| | | Kappa (k) | 0.70 | 0.84 | 0.80 | 0.68 |
| | | Classification Accuracy | 0.93 | 0.95 | 0.97 | 0.84 |
| | American Indian | Classification Consistency (P) | 0.85 | 0.92 | 0.98 | 0.75 |
| | | Probability of Chance | 0.52 | 0.68 | 0.94 | 0.35 |
| | | Kappa (k) | 0.69 | 0.75 | 0.62 | 0.62 |
| | | Classification Accuracy | 0.90 | 0.94 | 0.98 | 0.82 |
| | Two or More | Classification Consistency (P) | 0.90 | 0.92 | 0.95 | 0.76 |
| | | Probability of Chance | 0.60 | 0.54 | 0.81 | 0.28 |
| | | Kappa (k) | 0.74 | 0.82 | 0.74 | 0.67 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.97 | 0.83 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.87 | 0.93 | 0.98 | 0.78 |
| | | Probability of Chance | 0.51 | 0.70 | 0.94 | 0.36 |
| | | Kappa (k) | 0.73 | 0.77 | 0.69 | 0.66 |
| | | Classification Accuracy | 0.90 | 0.95 | 0.99 | 0.84 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.94 | 0.98 | 0.80 |
| | | Probability of Chance | 0.50 | 0.68 | 0.91 | 0.36 |
| | | Kappa (k) | 0.76 | 0.82 | 0.75 | 0.69 |
| | | Classification Accuracy | 0.92 | 0.96 | 0.98 | 0.86 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.88 | 0.93 | 0.97 | 0.78 |
| | | Probability of Chance | 0.53 | 0.62 | 0.90 | 0.32 |
| | | Kappa (k) | 0.75 | 0.80 | 0.71 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.95 | 0.98 | 0.84 |

**Table J-8 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.86 | 0.97 | 1.00 | 0.83 |
| | | Probability of Chance | 0.56 | 0.90 | 0.99 | 0.53 |
| | | Kappa (k) | 0.68 | 0.70 | 0.71 | 0.63 |
| | | Classification Accuracy | 0.90 | 0.98 | 1.00 | 0.88 |

**Table J-9. Indices for Classification Consistency and Accuracy, Mathematics Grade 5**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.90 | 0.90 | 0.95 | 0.75 |
| | | Probability of Chance | 0.60 | 0.51 | 0.82 | 0.28 |
| | | Kappa (k) | 0.76 | 0.79 | 0.72 | 0.66 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.96 | 0.82 |
| | Male | Classification Consistency (P) | 0.92 | 0.92 | 0.94 | 0.78 |
| | | Probability of Chance | 0.61 | 0.50 | 0.77 | 0.27 |
| | | Kappa (k) | 0.79 | 0.84 | 0.74 | 0.69 |
| | | Classification Accuracy | 0.94 | 0.94 | 0.96 | 0.84 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.93 | 0.90 | 0.93 | 0.76 |
| | | Probability of Chance | 0.72 | 0.51 | 0.74 | 0.29 |
| | | Kappa (k) | 0.75 | 0.79 | 0.75 | 0.66 |
| | | Classification Accuracy | 0.95 | 0.93 | 0.95 | 0.83 |
| | African American | Classification Consistency (P) | 0.88 | 0.96 | 0.99 | 0.83 |
| | | Probability of Chance | 0.56 | 0.82 | 0.98 | 0.51 |
| | | Kappa (k) | 0.73 | 0.76 | 0.66 | 0.66 |
| | | Classification Accuracy | 0.91 | 0.97 | 1.00 | 0.88 |
| | Hispanic | Classification Consistency (P) | 0.88 | 0.91 | 0.98 | 0.77 |
| | | Probability of Chance | 0.51 | 0.62 | 0.92 | 0.33 |
| | | Kappa (k) | 0.76 | 0.77 | 0.73 | 0.66 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.98 | 0.83 |
| | Asian | Classification Consistency (P) | 0.91 | 0.91 | 0.95 | 0.77 |
| | | Probability of Chance | 0.61 | 0.50 | 0.77 | 0.27 |
| | | Kappa (k) | 0.76 | 0.82 | 0.79 | 0.68 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.97 | 0.84 |
| | American Indian | Classification Consistency (P) | 0.86 | 0.92 | 0.98 | 0.76 |
| | | Probability of Chance | 0.51 | 0.63 | 0.92 | 0.34 |
| | | Kappa (k) | 0.73 | 0.78 | 0.74 | 0.64 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.99 | 0.84 |
| | Two or More | Classification Consistency (P) | 0.92 | 0.92 | 0.96 | 0.81 |
| | | Probability of Chance | 0.54 | 0.54 | 0.80 | 0.28 |
| | | Kappa (k) | 0.83 | 0.83 | 0.81 | 0.73 |
| | | Classification Accuracy | 0.94 | 0.94 | 0.96 | 0.85 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.86 | 0.93 | 0.99 | 0.78 |
| | | Probability of Chance | 0.50 | 0.73 | 0.97 | 0.39 |
| | | Kappa (k) | 0.72 | 0.73 | 0.65 | 0.64 |
| | | Classification Accuracy | 0.90 | 0.95 | 0.99 | 0.84 |
| Disability Status | Yes | Classification Consistency (P) | 0.89 | 0.95 | 0.98 | 0.82 |
| | | Probability of Chance | 0.51 | 0.71 | 0.93 | 0.42 |
| | | Kappa (k) | 0.77 | 0.83 | 0.73 | 0.69 |
| | | Classification Accuracy | 0.92 | 0.96 | 0.99 | 0.87 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.88 | 0.93 | 0.98 | 0.78 |
| | | Probability of Chance | 0.51 | 0.61 | 0.92 | 0.33 |
| | | Kappa (k) | 0.76 | 0.81 | 0.69 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.95 | 0.98 | 0.85 |

**Table J-9 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 5**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.89 | 0.97 | 1.00 | 0.86 |
| | | Probability of Chance | 0.65 | 0.90 | 0.99 | 0.63 |
| | | Kappa (k) | 0.68 | 0.71 | 0.67 | 0.61 |
| | | Classification Accuracy | 0.92 | 0.98 | 1.00 | 0.90 |

**Table J-10. Indices for Classification Consistency and Accuracy, Mathematics Grade 6**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.89 | 0.91 | 0.97 | 0.78 |
| | | Probability of Chance | 0.57 | 0.53 | 0.90 | 0.30 |
| | | Kappa (k) | 0.75 | 0.81 | 0.72 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.98 | 0.84 |
| | Male | Classification Consistency (P) | 0.90 | 0.91 | 0.97 | 0.78 |
| | | Probability of Chance | 0.57 | 0.52 | 0.88 | 0.30 |
| | | Kappa (k) | 0.78 | 0.82 | 0.72 | 0.69 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.98 | 0.84 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.90 | 0.96 | 0.77 |
| | | Probability of Chance | 0.67 | 0.50 | 0.86 | 0.31 |
| | | Kappa (k) | 0.73 | 0.80 | 0.71 | 0.66 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.97 | 0.83 |
| | African American | Classification Consistency (P) | 0.88 | 0.96 | 1.00 | 0.84 |
| | | Probability of Chance | 0.58 | 0.85 | 0.99 | 0.55 |
| | | Kappa (k) | 0.71 | 0.76 | 0.54 | 0.65 |
| | | Classification Accuracy | 0.92 | 0.97 | 1.00 | 0.89 |
| | Hispanic | Classification Consistency (P) | 0.87 | 0.93 | 0.99 | 0.79 |
| | | Probability of Chance | 0.50 | 0.68 | 0.97 | 0.37 |
| | | Kappa (k) | 0.74 | 0.79 | 0.67 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.99 | 0.85 |
| | Asian | Classification Consistency (P) | 0.89 | 0.92 | 0.96 | 0.77 |
| | | Probability of Chance | 0.58 | 0.52 | 0.82 | 0.28 |
| | | Kappa (k) | 0.75 | 0.83 | 0.79 | 0.69 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.97 | 0.84 |
| | American Indian | Classification Consistency (P) | 0.87 | 0.94 | 1.00 | 0.81 |
| | | Probability of Chance | 0.51 | 0.75 | 0.99 | 0.43 |
| | | Kappa (k) | 0.74 | 0.77 | 0.64 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.96 | 1.00 | 0.86 |
| | Two or More | Classification Consistency (P) | 0.89 | 0.92 | 0.98 | 0.78 |
| | | Probability of Chance | 0.53 | 0.57 | 0.91 | 0.31 |
| | | Kappa (k) | 0.76 | 0.81 | 0.74 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.98 | 0.84 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.86 | 0.95 | 1.00 | 0.81 |
| | | Probability of Chance | 0.54 | 0.82 | 0.99 | 0.48 |
| | | Kappa (k) | 0.70 | 0.74 | 0.67 | 0.64 |
| | | Classification Accuracy | 0.90 | 0.97 | 1.00 | 0.87 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.96 | 0.99 | 0.84 |
| | | Probability of Chance | 0.56 | 0.80 | 0.98 | 0.50 |
| | | Kappa (k) | 0.74 | 0.80 | 0.72 | 0.67 |
| | | Classification Accuracy | 0.92 | 0.97 | 1.00 | 0.89 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.87 | 0.93 | 0.99 | 0.79 |
| | | Probability of Chance | 0.50 | 0.67 | 0.97 | 0.37 |
| | | Kappa (k) | 0.74 | 0.79 | 0.67 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.99 | 0.85 |

**Table J-10 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 7**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.90 | 0.99 | 1.00 | 0.89 |
| | | Probability of Chance | 0.72 | 0.96 | 1.00 | 0.71 |
| | | Kappa (k) | 0.64 | 0.67 | 0.65 | 0.60 |
| | | Classification Accuracy | 0.93 | 0.99 | 1.00 | 0.92 |

**Table J-11. Indices for Classification Consistency and Accuracy, Mathematics Grade 7**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.89 | 0.92 | 0.98 | 0.78 |
| | | Probability of Chance | 0.52 | 0.57 | 0.92 | 0.32 |
| | | Kappa (k) | 0.77 | 0.81 | 0.71 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.98 | 0.84 |
| | Male | Classification Consistency (P) | 0.91 | 0.92 | 0.97 | 0.79 |
| | | Probability of Chance | 0.53 | 0.54 | 0.90 | 0.31 |
| | | Kappa (k) | 0.80 | 0.82 | 0.71 | 0.70 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.98 | 0.85 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.90 | 0.90 | 0.97 | 0.77 |
| | | Probability of Chance | 0.60 | 0.51 | 0.89 | 0.30 |
| | | Kappa (k) | 0.76 | 0.80 | 0.70 | 0.67 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.98 | 0.84 |
| | African American | Classification Consistency (P) | 0.90 | 0.97 | 1.00 | 0.87 |
| | | Probability of Chance | 0.65 | 0.88 | 0.99 | 0.62 |
| | | Kappa (k) | 0.73 | 0.77 | 0.62 | 0.67 |
| | | Classification Accuracy | 0.93 | 0.98 | 1.00 | 0.91 |
| | Hispanic | Classification Consistency (P) | 0.87 | 0.95 | 0.99 | 0.81 |
| | | Probability of Chance | 0.51 | 0.73 | 0.97 | 0.42 |
| | | Kappa (k) | 0.74 | 0.80 | 0.69 | 0.68 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |
| | Asian | Classification Consistency (P) | 0.89 | 0.93 | 0.96 | 0.78 |
| | | Probability of Chance | 0.54 | 0.53 | 0.84 | 0.29 |
| | | Kappa (k) | 0.77 | 0.84 | 0.77 | 0.70 |
| | | Classification Accuracy | 0.92 | 0.95 | 0.97 | 0.84 |
| | American Indian | Classification Consistency (P) | 0.92 | 0.97 | 0.99 | 0.88 |
| | | Probability of Chance | 0.57 | 0.80 | 0.98 | 0.53 |
| | | Kappa (k) | 0.80 | 0.83 | 0.72 | 0.74 |
| | | Classification Accuracy | 0.92 | 0.97 | 0.99 | 0.89 |
| | Two or More | Classification Consistency (P) | 0.89 | 0.93 | 0.98 | 0.80 |
| | | Probability of Chance | 0.50 | 0.61 | 0.92 | 0.34 |
| | | Kappa (k) | 0.77 | 0.82 | 0.73 | 0.69 |
| | | Classification Accuracy | 0.92 | 0.95 | 0.98 | 0.85 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.88 | 0.97 | 1.00 | 0.85 |
| | | Probability of Chance | 0.61 | 0.88 | 0.99 | 0.58 |
| | | Kappa (k) | 0.69 | 0.76 | 0.70 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.98 | 1.00 | 0.89 |
| Disability Status | Yes | Classification Consistency (P) | 0.90 | 0.97 | 1.00 | 0.87 |
| | | Probability of Chance | 0.62 | 0.85 | 0.98 | 0.59 |
| | | Kappa (k) | 0.74 | 0.80 | 0.71 | 0.67 |
| | | Classification Accuracy | 0.93 | 0.98 | 1.00 | 0.90 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.88 | 0.94 | 0.99 | 0.81 |
| | | Probability of Chance | 0.51 | 0.72 | 0.97 | 0.41 |
| | | Kappa (k) | 0.76 | 0.79 | 0.66 | 0.68 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.87 |

**Table J-11 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 7**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.92 | 0.99 | 1.00 | 0.91 |
| | | Probability of Chance | 0.79 | 0.96 | 1.00 | 0.79 |
| | | Kappa (k) | 0.62 | 0.72 | 0.63 | 0.58 |
| | | Classification Accuracy | 0.94 | 0.99 | 1.00 | 0.94 |

**Table J-12. Indices for Classification Consistency and Accuracy, Mathematics Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.88 | 0.92 | 0.97 | 0.78 |
| | | Probability of Chance | 0.55 | 0.57 | 0.88 | 0.30 |
| | | Kappa (k) | 0.75 | 0.82 | 0.72 | 0.68 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.98 | 0.84 |
| | Male | Classification Consistency (P) | 0.89 | 0.93 | 0.96 | 0.79 |
| | | Probability of Chance | 0.53 | 0.57 | 0.86 | 0.30 |
| | | Kappa (k) | 0.77 | 0.85 | 0.74 | 0.70 |
| | | Classification Accuracy | 0.93 | 0.95 | 0.97 | 0.85 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.89 | 0.91 | 0.95 | 0.75 |
| | | Probability of Chance | 0.61 | 0.53 | 0.84 | 0.29 |
| | | Kappa (k) | 0.72 | 0.81 | 0.71 | 0.65 |
| | | Classification Accuracy | 0.93 | 0.94 | 0.97 | 0.83 |
| | African American | Classification Consistency (P) | 0.88 | 0.97 | 1.00 | 0.85 |
| | | Probability of Chance | 0.60 | 0.89 | 0.99 | 0.58 |
| | | Kappa (k) | 0.70 | 0.77 | 0.72 | 0.65 |
| | | Classification Accuracy | 0.92 | 0.98 | 1.00 | 0.90 |
| | Hispanic | Classification Consistency (P) | 0.86 | 0.95 | 0.99 | 0.80 |
| | | Probability of Chance | 0.50 | 0.75 | 0.95 | 0.40 |
| | | Kappa (k) | 0.73 | 0.81 | 0.72 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |
| | Asian | Classification Consistency (P) | 0.89 | 0.93 | 0.96 | 0.78 |
| | | Probability of Chance | 0.57 | 0.53 | 0.77 | 0.27 |
| | | Kappa (k) | 0.75 | 0.84 | 0.81 | 0.70 |
| | | Classification Accuracy | 0.93 | 0.95 | 0.97 | 0.84 |
| | American Indian | Classification Consistency (P) | 0.86 | 0.95 | 0.99 | 0.80 |
| | | Probability of Chance | 0.51 | 0.80 | 0.97 | 0.43 |
| | | Kappa (k) | 0.71 | 0.75 | 0.78 | 0.65 |
| | | Classification Accuracy | 0.90 | 0.96 | 1.00 | 0.86 |
| | Two or More | Classification Consistency (P) | 0.87 | 0.93 | 0.97 | 0.78 |
| | | Probability of Chance | 0.51 | 0.62 | 0.90 | 0.33 |
| | | Kappa (k) | 0.74 | 0.83 | 0.72 | 0.67 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.85 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.85 | 0.97 | 1.00 | 0.82 |
| | | Probability of Chance | 0.57 | 0.89 | 0.99 | 0.54 |
| | | Kappa (k) | 0.65 | 0.73 | 0.73 | 0.61 |
| | | Classification Accuracy | 0.90 | 0.98 | 1.00 | 0.88 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.98 | 0.99 | 0.85 |
| | | Probability of Chance | 0.61 | 0.87 | 0.98 | 0.58 |
| | | Kappa (k) | 0.69 | 0.81 | 0.70 | 0.64 |
| | | Classification Accuracy | 0.92 | 0.98 | 1.00 | 0.90 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.87 | 0.95 | 0.99 | 0.81 |
| | | Probability of Chance | 0.50 | 0.74 | 0.96 | 0.40 |
| | | Kappa (k) | 0.73 | 0.81 | 0.70 | 0.68 |
| | | Classification Accuracy | 0.91 | 0.96 | 0.99 | 0.86 |

**Table J-12 (continued). Indices for Classification Consistency and Accuracy, Mathematics Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Accommodation Use | Yes | Classification Consistency (P) | 0.90 | 0.99 | 1.00 | 0.89 |
| | | Probability of Chance | 0.74 | 0.98 | 1.00 | 0.74 |
| | | Kappa (k) | 0.60 | 0.62 | 0.36 | 0.57 |
| | | Classification Accuracy | 0.93 | 0.99 | 1.00 | 0.92 |

**Table J-13. Indices for Classification Consistency and Accuracy, Science Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.91 | 0.88 | 0.90 | 0.69 |
| | | Probability of Chance | 0.69 | 0.50 | 0.69 | 0.26 |
| | | Kappa (k) | 0.70 | 0.77 | 0.67 | 0.58 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.93 | 0.78 |
| | Male | Classification Consistency (P) | 0.91 | 0.89 | 0.90 | 0.70 |
| | | Probability of Chance | 0.69 | 0.50 | 0.67 | 0.26 |
| | | Kappa (k) | 0.70 | 0.78 | 0.70 | 0.59 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.93 | 0.79 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.93 | 0.88 | 0.87 | 0.69 |
| | | Probability of Chance | 0.81 | 0.53 | 0.62 | 0.28 |
| | | Kappa (k) | 0.65 | 0.74 | 0.67 | 0.57 |
| | | Classification Accuracy | 0.95 | 0.91 | 0.91 | 0.78 |
| | African American | Classification Consistency (P) | 0.83 | 0.93 | 0.98 | 0.74 |
| | | Probability of Chance | 0.50 | 0.75 | 0.95 | 0.40 |
| | | Kappa (k) | 0.66 | 0.72 | 0.59 | 0.57 |
| | | Classification Accuracy | 0.88 | 0.95 | 0.99 | 0.82 |
| | Hispanic | Classification Consistency (P) | 0.86 | 0.90 | 0.95 | 0.70 |
| | | Probability of Chance | 0.56 | 0.58 | 0.85 | 0.30 |
| | | Kappa (k) | 0.67 | 0.75 | 0.65 | 0.57 |
| | | Classification Accuracy | 0.90 | 0.92 | 0.96 | 0.79 |
| | Asian | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.71 |
| | | Probability of Chance | 0.65 | 0.51 | 0.73 | 0.27 |
| | | Kappa (k) | 0.70 | 0.77 | 0.73 | 0.61 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.95 | 0.79 |
| | American Indian | Classification Consistency (P) | 0.85 | 0.89 | 0.94 | 0.68 |
| | | Probability of Chance | 0.58 | 0.58 | 0.86 | 0.31 |
| | | Kappa (k) | 0.63 | 0.74 | 0.58 | 0.54 |
| | | Classification Accuracy | 0.89 | 0.92 | 0.96 | 0.77 |
| | Two or More | Classification Consistency (P) | 0.90 | 0.88 | 0.92 | 0.71 |
| | | Probability of Chance | 0.67 | 0.50 | 0.72 | 0.27 |
| | | Kappa (k) | 0.71 | 0.76 | 0.71 | 0.60 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.94 | 0.78 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.83 | 0.90 | 0.97 | 0.70 |
| | | Probability of Chance | 0.52 | 0.66 | 0.92 | 0.34 |
| | | Kappa (k) | 0.65 | 0.70 | 0.61 | 0.55 |
| | | Classification Accuracy | 0.88 | 0.93 | 0.98 | 0.79 |
| Disability Status | Yes | Classification Consistency (P) | 0.85 | 0.91 | 0.95 | 0.72 |
| | | Probability of Chance | 0.53 | 0.60 | 0.85 | 0.31 |
| | | Kappa (k) | 0.69 | 0.77 | 0.69 | 0.59 |
| | | Classification Accuracy | 0.90 | 0.93 | 0.97 | 0.80 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.87 | 0.89 | 0.94 | 0.70 |
| | | Probability of Chance | 0.57 | 0.55 | 0.82 | 0.29 |
| | | Kappa (k) | 0.69 | 0.76 | 0.66 | 0.58 |
| | | Classification Accuracy | 0.91 | 0.92 | 0.96 | 0.78 |

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50

**Table J-14. Indices for Classification Consistency and Accuracy, Science Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.89 | 0.86 | 0.91 | 0.66 |
| | | Probability of Chance | 0.67 | 0.50 | 0.70 | 0.26 |
| | | Kappa (k) | 0.67 | 0.72 | 0.69 | 0.55 |
| | | Classification Accuracy | 0.92 | 0.90 | 0.93 | 0.76 |
| | Male | Classification Consistency (P) | 0.89 | 0.88 | 0.91 | 0.69 |
| | | Probability of Chance | 0.65 | 0.50 | 0.65 | 0.25 |
| | | Kappa (k) | 0.70 | 0.76 | 0.73 | 0.58 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.93 | 0.77 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.86 | 0.89 | 0.67 |
| | | Probability of Chance | 0.75 | 0.52 | 0.62 | 0.27 |
| | | Kappa (k) | 0.65 | 0.72 | 0.71 | 0.55 |
| | | Classification Accuracy | 0.94 | 0.90 | 0.92 | 0.76 |
| | African American | Classification Consistency (P) | 0.81 | 0.92 | 0.97 | 0.71 |
| | | Probability of Chance | 0.50 | 0.73 | 0.94 | 0.39 |
| | | Kappa (k) | 0.63 | 0.68 | 0.58 | 0.52 |
| | | Classification Accuracy | 0.86 | 0.94 | 0.98 | 0.79 |
| | Hispanic | Classification Consistency (P) | 0.85 | 0.88 | 0.94 | 0.67 |
| | | Probability of Chance | 0.55 | 0.56 | 0.83 | 0.29 |
| | | Kappa (k) | 0.66 | 0.72 | 0.67 | 0.54 |
| | | Classification Accuracy | 0.89 | 0.91 | 0.96 | 0.76 |
| | Asian | Classification Consistency (P) | 0.90 | 0.88 | 0.92 | 0.70 |
| | | Probability of Chance | 0.70 | 0.50 | 0.65 | 0.26 |
| | | Kappa (k) | 0.65 | 0.76 | 0.76 | 0.59 |
| | | Classification Accuracy | 0.93 | 0.91 | 0.94 | 0.78 |
| | American Indian | Classification Consistency (P) | 0.85 | 0.89 | 0.95 | 0.69 |
| | | Probability of Chance | 0.55 | 0.57 | 0.85 | 0.30 |
| | | Kappa (k) | 0.66 | 0.74 | 0.67 | 0.56 |
| | | Classification Accuracy | 0.89 | 0.92 | 0.96 | 0.77 |
| | Two or More | Classification Consistency (P) | 0.87 | 0.89 | 0.93 | 0.69 |
| | | Probability of Chance | 0.62 | 0.51 | 0.71 | 0.26 |
| | | Kappa (k) | 0.66 | 0.77 | 0.74 | 0.58 |
| | | Classification Accuracy | 0.91 | 0.92 | 0.95 | 0.78 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.80 | 0.90 | 0.98 | 0.69 |
| | | Probability of Chance | 0.50 | 0.74 | 0.95 | 0.38 |
| | | Kappa (k) | 0.60 | 0.62 | 0.61 | 0.50 |
| | | Classification Accuracy | 0.85 | 0.93 | 0.99 | 0.77 |
| Disability Status | Yes | Classification Consistency (P) | 0.83 | 0.92 | 0.97 | 0.71 |
| | | Probability of Chance | 0.50 | 0.68 | 0.88 | 0.36 |
| | | Kappa (k) | 0.65 | 0.74 | 0.72 | 0.55 |
| | | Classification Accuracy | 0.87 | 0.94 | 0.98 | 0.79 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.85 | 0.88 | 0.94 | 0.68 |
| | | Probability of Chance | 0.55 | 0.56 | 0.81 | 0.29 |
| | | Kappa (k) | 0.67 | 0.72 | 0.69 | 0.55 |
| | | Classification Accuracy | 0.89 | 0.91 | 0.96 | 0.76 |

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

**Table J-15. Indices for Classification Consistency and Accuracy, Social Studies Grade 4**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.92 | 0.90 | 0.90 | 0.72 |
| | | Probability of Chance | 0.67 | 0.51 | 0.65 | 0.27 |
| | | Kappa (k) | 0.75 | 0.80 | 0.70 | 0.62 |
| | | Classification Accuracy | 0.94 | 0.93 | 0.93 | 0.80 |
| | Male | Classification Consistency (P) | 0.91 | 0.88 | 0.89 | 0.69 |
| | | Probability of Chance | 0.68 | 0.51 | 0.71 | 0.28 |
| | | Kappa (k) | 0.72 | 0.76 | 0.63 | 0.57 |
| | | Classification Accuracy | 0.93 | 0.92 | 0.90 | 0.75 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.94 | 0.90 | 0.87 | 0.72 |
| | | Probability of Chance | 0.77 | 0.57 | 0.58 | 0.29 |
| | | Kappa (k) | 0.72 | 0.77 | 0.70 | 0.61 |
| | | Classification Accuracy | 0.96 | 0.93 | 0.91 | 0.80 |
| | African American | Classification Consistency (P) | 0.88 | 0.93 | 0.97 | 0.78 |
| | | Probability of Chance | 0.51 | 0.65 | 0.90 | 0.39 |
| | | Kappa (k) | 0.75 | 0.79 | 0.66 | 0.63 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.84 |
| | Hispanic | Classification Consistency (P) | 0.88 | 0.90 | 0.94 | 0.73 |
| | | Probability of Chance | 0.54 | 0.52 | 0.80 | 0.28 |
| | | Kappa (k) | 0.74 | 0.79 | 0.68 | 0.62 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.96 | 0.80 |
| | Asian | Classification Consistency (P) | 0.91 | 0.90 | 0.91 | 0.72 |
| | | Probability of Chance | 0.63 | 0.50 | 0.67 | 0.26 |
| | | Kappa (k) | 0.74 | 0.80 | 0.72 | 0.62 |
| | | Classification Accuracy | 0.94 | 0.93 | 0.94 | 0.80 |
| | American Indian | Classification Consistency (P) | 0.88 | 0.89 | 0.94 | 0.71 |
| | | Probability of Chance | 0.56 | 0.52 | 0.81 | 0.28 |
| | | Kappa (k) | 0.71 | 0.77 | 0.68 | 0.59 |
| | | Classification Accuracy | 0.91 | 0.92 | 0.96 | 0.78 |
| | Two or More | Classification Consistency (P) | 0.90 | 0.91 | 0.90 | 0.72 |
| | | Probability of Chance | 0.63 | 0.50 | 0.66 | 0.26 |
| | | Kappa (k) | 0.74 | 0.81 | 0.71 | 0.62 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.93 | 0.80 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.87 | 0.90 | 0.95 | 0.73 |
| | | Probability of Chance | 0.51 | 0.57 | 0.88 | 0.31 |
| | | Kappa (k) | 0.73 | 0.76 | 0.59 | 0.61 |
| | | Classification Accuracy | 0.91 | 0.92 | 0.97 | 0.80 |
| Disability Status | Yes | Classification Consistency (P) | 0.89 | 0.92 | 0.95 | 0.76 |
| | | Probability of Chance | 0.50 | 0.56 | 0.82 | 0.32 |
| | | Kappa (k) | 0.77 | 0.82 | 0.71 | 0.65 |
| | | Classification Accuracy | 0.92 | 0.94 | 0.96 | 0.82 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.89 | 0.90 | 0.93 | 0.73 |
| | | Probability of Chance | 0.55 | 0.52 | 0.79 | 0.28 |
| | | Kappa (k) | 0.76 | 0.80 | 0.68 | 0.62 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.95 | 0.80 |

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

**Table J-16. Indices for Classification Consistency and Accuracy, Social Studies Grade 8**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.93 | 0.89 | 0.90 | 0.72 |
| | | Probability of Chance | 0.71 | 0.52 | 0.64 | 0.27 |
| | | Kappa (k) | 0.76 | 0.77 | 0.72 | 0.62 |
| | | Classification Accuracy | 0.95 | 0.93 | 0.93 | 0.80 |
| | Male | Classification Consistency (P) | 0.92 | 0.91 | 0.90 | 0.74 |
| | | Probability of Chance | 0.66 | 0.51 | 0.62 | 0.26 |
| | | Kappa (k) | 0.78 | 0.81 | 0.75 | 0.65 |
| | | Classification Accuracy | 0.94 | 0.94 | 0.93 | 0.81 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.94 | 0.90 | 0.88 | 0.73 |
| | | Probability of Chance | 0.77 | 0.56 | 0.58 | 0.28 |
| | | Kappa (k) | 0.75 | 0.78 | 0.72 | 0.62 |
| | | Classification Accuracy | 0.96 | 0.93 | 0.92 | 0.81 |
| | African American | Classification Consistency (P) | 0.88 | 0.91 | 0.96 | 0.75 |
| | | Probability of Chance | 0.50 | 0.62 | 0.89 | 0.34 |
| | | Kappa (k) | 0.75 | 0.76 | 0.66 | 0.62 |
| | | Classification Accuracy | 0.91 | 0.94 | 0.97 | 0.82 |
| | Hispanic | Classification Consistency (P) | 0.90 | 0.89 | 0.94 | 0.72 |
| | | Probability of Chance | 0.57 | 0.51 | 0.77 | 0.27 |
| | | Kappa (k) | 0.76 | 0.77 | 0.72 | 0.62 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.96 | 0.80 |
| | Asian | Classification Consistency (P) | 0.92 | 0.89 | 0.91 | 0.72 |
| | | Probability of Chance | 0.72 | 0.53 | 0.60 | 0.27 |
| | | Kappa (k) | 0.71 | 0.77 | 0.77 | 0.62 |
| | | Classification Accuracy | 0.94 | 0.93 | 0.94 | 0.81 |
| | American Indian | Classification Consistency (P) | 0.88 | 0.88 | 0.95 | 0.71 |
| | | Probability of Chance | 0.56 | 0.53 | 0.82 | 0.28 |
| | | Kappa (k) | 0.72 | 0.74 | 0.72 | 0.59 |
| | | Classification Accuracy | 0.92 | 0.91 | 0.97 | 0.79 |
| | Two or More | Classification Consistency (P) | 0.92 | 0.89 | 0.91 | 0.73 |
| | | Probability of Chance | 0.64 | 0.50 | 0.66 | 0.26 |
| | | Kappa (k) | 0.78 | 0.78 | 0.74 | 0.63 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.94 | 0.81 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.86 | 0.89 | 0.97 | 0.73 |
| | | Probability of Chance | 0.50 | 0.64 | 0.93 | 0.34 |
| | | Kappa (k) | 0.71 | 0.70 | 0.65 | 0.59 |
| | | Classification Accuracy | 0.89 | 0.93 | 0.98 | 0.80 |
| Disability Status | Yes | Classification Consistency (P) | 0.88 | 0.92 | 0.97 | 0.77 |
| | | Probability of Chance | 0.50 | 0.64 | 0.86 | 0.36 |
| | | Kappa (k) | 0.75 | 0.79 | 0.75 | 0.64 |
| | | Classification Accuracy | 0.91 | 0.95 | 0.98 | 0.83 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.90 | 0.89 | 0.93 | 0.73 |
| | | Probability of Chance | 0.57 | 0.51 | 0.77 | 0.27 |
| | | Kappa (k) | 0.76 | 0.78 | 0.71 | 0.63 |
| | | Classification Accuracy | 0.92 | 0.93 | 0.95 | 0.80 |

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

**Table J-17. Indices for Classification Consistency and Accuracy, Social Studies Grade 10**

| Group | Category | Indices | Cut 1 | Cut 2 | Cut 3 | All Cuts |
|---|---|---|---|---|---|---|
| Gender | Female | Classification Consistency (P) | 0.90 | 0.88 | 0.91 | 0.69 |
| | | Probability of Chance | 0.65 | 0.50 | 0.64 | 0.25 |
| | | Kappa (k) | 0.71 | 0.76 | 0.74 | 0.59 |
| | | Classification Accuracy | 0.93 | 0.91 | 0.93 | 0.78 |
| | Male | Classification Consistency (P) | 0.89 | 0.90 | 0.92 | 0.71 |
| | | Probability of Chance | 0.59 | 0.50 | 0.64 | 0.25 |
| | | Kappa (k) | 0.73 | 0.79 | 0.77 | 0.61 |
| | | Classification Accuracy | 0.93 | 0.93 | 0.94 | 0.79 |
| Race/Ethnicity | White | Classification Consistency (P) | 0.91 | 0.89 | 0.90 | 0.70 |
| | | Probability of Chance | 0.68 | 0.51 | 0.60 | 0.25 |
| | | Kappa (k) | 0.70 | 0.77 | 0.75 | 0.59 |
| | | Classification Accuracy | 0.94 | 0.92 | 0.93 | 0.78 |
| | African American | Classification Consistency (P) | 0.85 | 0.92 | 0.97 | 0.74 |
| | | Probability of Chance | 0.50 | 0.68 | 0.88 | 0.38 |
| | | Kappa (k) | 0.70 | 0.76 | 0.71 | 0.58 |
| | | Classification Accuracy | 0.90 | 0.94 | 0.98 | 0.82 |
| | Hispanic | Classification Consistency (P) | 0.86 | 0.89 | 0.94 | 0.70 |
| | | Probability of Chance | 0.53 | 0.56 | 0.80 | 0.29 |
| | | Kappa (k) | 0.70 | 0.75 | 0.72 | 0.58 |
| | | Classification Accuracy | 0.90 | 0.92 | 0.96 | 0.79 |
| | Asian | Classification Consistency (P) | 0.90 | 0.87 | 0.91 | 0.69 |
| | | Probability of Chance | 0.67 | 0.50 | 0.63 | 0.25 |
| | | Kappa (k) | 0.71 | 0.75 | 0.76 | 0.59 |
| | | Classification Accuracy | 0.93 | 0.91 | 0.94 | 0.78 |
| | American Indian | Classification Consistency (P) | 0.86 | 0.90 | 0.96 | 0.71 |
| | | Probability of Chance | 0.51 | 0.60 | 0.86 | 0.32 |
| | | Kappa (k) | 0.71 | 0.74 | 0.68 | 0.58 |
| | | Classification Accuracy | 0.90 | 0.93 | 0.97 | 0.80 |
| | Two or More | Classification Consistency (P) | 0.88 | 0.89 | 0.93 | 0.71 |
| | | Probability of Chance | 0.59 | 0.51 | 0.66 | 0.25 |
| | | Kappa (k) | 0.72 | 0.78 | 0.78 | 0.61 |
| | | Classification Accuracy | 0.92 | 0.92 | 0.95 | 0.80 |
| Limited English Proficiency | Yes | Classification Consistency (P) | 0.82 | 0.93 | 0.99 | 0.75 |
| | | Probability of Chance | 0.53 | 0.81 | 0.96 | 0.46 |
| | | Kappa (k) | 0.62 | 0.65 | 0.69 | 0.53 |
| | | Classification Accuracy | 0.88 | 0.95 | 0.99 | 0.82 |
| Disability Status | Yes | Classification Consistency (P) | 0.85 | 0.94 | 0.98 | 0.76 |
| | | Probability of Chance | 0.52 | 0.73 | 0.90 | 0.43 |
| | | Kappa (k) | 0.68 | 0.76 | 0.76 | 0.58 |
| | | Classification Accuracy | 0.90 | 0.95 | 0.98 | 0.83 |
| Economically Disadvantaged | Yes | Classification Consistency (P) | 0.86 | 0.89 | 0.94 | 0.70 |
| | | Probability of Chance | 0.52 | 0.56 | 0.79 | 0.29 |
| | | Kappa (k) | 0.71 | 0.76 | 0.72 | 0.58 |
| | | Classification Accuracy | 0.90 | 0.92 | 0.96 | 0.79 |

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

# Appendix K
# Wisconsin Standard Performance Index Score Computation

# Table of Contents

Technical details of the SPI estimation procedure described in this section are based on description of the SPI computation methodology included in the *TerraNova* 2nd Edition Technical Report (CTB/McGraw-Hill, 2000).

The Standard Performance Index (SPI) is an estimate of the true score (estimated proportion of total, or maximum, points possible) for a content standard based on the performance of a given student. Because most standards are measured by a relatively small number of items, a Bayesian procedure that takes into account the overall test performance is used to improve the reliability of the standard scores. Given a student's scale score on the test, item response theory (IRT) is used, via the 3-paremeter logistic (3PL) model for MC items and the 2-paremeter-partial credit (2PPC) model for CR items, to compute the estimated proportion of the maximum points obtained for that standard.

The estimated proportion of the maximum points obtained for the standard provides the initial (Bayesian prior) estimate of the student's mastery score. If this initial estimate is consistent with the student's observed proportion, as indicated by a chi-square test, the two scores are combined as a weighted average to obtain the SPI score (the estimated true score). The appropriate weight for the Bayesian prior estimate is computed as a function of the standard error (SE) of the scale score on which it is based: the smaller the SE, the larger the weight. If the prior estimate and the observed proportion differ significantly, the observed proportion of the maximum score is used without the prior estimate to compute the student's score on that objective.

## Standard Performance Index Computation

The standard performance index (SPI) is an estimated true score (estimated proportion of total or maximum points obtained) based on the performance of a given examinee for the items in a given learning strand. Assume a $k$-item test is composed of $j$ strands with a maximum possible raw score of $n$. Also assume that each item contributes to, at most, one strand, and the $k_j$ items in strand $j$ contribute a maximum of $n_j$ points. Define $X_j$ as the observed raw score on strand $j$. The true score is

$$T_j \equiv E(X_j / n_j).$$

It is assumed that there is information available about the examinee in addition to the strand score, and this information provides a prior distribution for $T_j$. This prior distribution of $T_j$ for a given examinee is assumed to be $\beta(r_j, s_j)$:

$$g(T_j) = \frac{(r_j + s_j - 1)! \, T_j^{r_j - 1} (1 - T_j)^{s_j - 1}}{(r_j - 1)! \, (s_j - 1)!} \quad (1)$$

for $0 \leq T_j \leq 1$; $r_j, s_j > 0$. Estimates of $r_j$ and $s_j$ are derived from IRT (Lord, 1980).

It is assumed that $X_j$ follows a binomial distribution, given $T_j$:

$$p(X_j = x_j | T_j) = Binomial\ (n_j,\ T_j = \sum_{i=1}^{k_j} T_i / n_j),$$

where

$T_i$ is the expected value of the score for item $i$ in strand $j$ for a given $\theta$.

Given these assumptions, the posterior distribution of $T_j$, given $x_j$, is

$$g(T_j | X_j = x_j) = \beta(p_j, q_j), \quad (2)$$

with

$$p_j = r_j + x_j \quad (3)$$

and

$$q_j = s_j + n_j - x_j. \quad (4)$$

The SPI is defined to be the mean of this posterior distribution:

$$\tilde{T}_j = \frac{p_j}{p_j + q_j}.$$

Following Novick and Jackson (1974, p. 119), a mastery band is created to be the $C\%$ central credibility interval for $T_j$. It is obtained by identifying the values that place $\frac{1}{2}(100 - C)\%$ of the $\beta(p_j, q_j)$ density in each tail of the distribution.

## Estimation of the Prior Distribution of $T_j$

The $k$ items in each test are scaled together using a generalized IRT model (3PL/2PPC) that fits a three-parameter logistic model (3PL) to the MC items and a generalized partial-credit model (2PPC) to the CR items (Yen, 1993).

The 3PL model is

$$P_i(\theta) = P(X_i = 1|\theta) = c_i + \frac{1 - c_i}{1 + exp[-1.7A_i(\theta - B_i)]}, \quad (5)$$

where

$A_i$ is the discrimination, $B_i$ is the location, and $c_i$ is the guessing parameter for item $i$.

A generalization of Master's (1982) partial credit (2PPC) model was used for the CR items. The 2PPC model, the same as Muraki's (1992) "generalized partial credit model," has been shown to fit response data obtained from a wide variety of mixed-item type achievement tests (Fitzpatrick, Link, Yen, Burket, Ito, & Sykes, 1996). For a CR item with $1_i$ score levels, integer scores were assigned that ranged from 0 to $1_i - 1$:

$$P_{im}(\theta) = P(X_i = m - 1|\theta) = \frac{exp(z_{im})}{\sum_{g=1}^{1_i} exp(z_{ig})}, m = 1 \dots 1_i, \quad (6)$$

where

$$z_{ig} = \alpha_i(m - 1)\theta - \sum_{h=0}^{m-1} \gamma_{ih}, \quad (7)$$

and

$$\gamma_{i0} = 0.$$

Alpha ($\alpha_i$) is the item discrimination, and gamma ($\gamma_{ih}$) is related to the difficulty of the item levels; the trace lines for adjacent score levels intersect at $\gamma_{ih}/\alpha_i$.

Item parameters estimated from the national standardization sample are used to obtain SPI values. $T_{ij}(\theta)$ is the expected score for item $i$ in strand $j$, and ($\theta$) is the common trait value to which the items are scaled:

$$T_{ij}(\theta) = \sum_{m=1}^{1_i} (m-1) P_{ijm}(\theta),$$

where

$1_i$ is the number of score levels in item $i$, including 0.

$T_j$, the expected proportion of maximum score for strand $j$, is

$$T_j = \frac{1}{n_j} \left[ \sum_{i=1}^{k_j} T_{ij}(\theta) \right]. \quad (8)$$

The expected score for item $i$ and estimated proportion-correct of maximum score for strand $j$ are obtained by substituting the estimate of the trait ($\hat{\theta}$) for the actual trait value.

The theoretical random variation in item response vectors and resulting ($\hat{\theta}$) values for a given examinee produces the distribution $g(\hat{T}_j|\hat{\theta})$ with mean $\mu(\hat{T}_j|\theta)$ and variance $\sigma(\hat{T}_j|\theta)$. This distribution is used to estimate a prior distribution of $T_j$. Given that $T_j$ is assumed to be distributed as a beta distribution (equation 1), the mean $[\mu(\hat{T}_j|\theta)]$ and variance $[\sigma(\hat{T}_j|\theta)]$ of this distribution can be expressed in terms of its parameters, $r_j$ and $s_j$.

Expressing the mean and variance of the prior distribution in terms of the parameters of the beta distribution (Novick & Jackson, 1974, p. 113) produces

$$\mu(\hat{T}_j|\theta) = \frac{r_j}{r_j + s_j} \quad (9)$$

and

$$\sigma(\hat{T}_j|\theta) = \frac{r_j s_j}{(r_j + s_j)^2 (r_j + s_j + 1)}. \quad (10)$$

Solving these equations for $r_j$ and $s_j$ produces

$$r_j = \mu(\hat{T}_j|\theta) n_j^* \quad (11)$$

and

$$s_j = [1 - \mu(\hat{T}_j|\theta)] n_j^*, \quad (12)$$

where

$$n_j^* = \frac{\mu(\hat{T}_j|\theta)\left[1 - \mu(\hat{T}_j|\theta)\right]}{\sigma^2(\hat{T}_j|\theta)} - 1. \quad (13)$$

Using IRT, $\sigma(\hat{T}_j|\theta)$ can be expressed in terms of item parameters (Lord, 1983):

$$\mu(\hat{T}_j|\theta) \approx \frac{1}{n_j}\sum_{i=1}^{k_j} \hat{T}_{ij}(\theta). \quad (14)$$

Because $T_j$ is a monotonic transformation of $\theta$ (Lord, 1980, p.71),

$$\sigma^2(\hat{T}_j|\theta) = \sigma^2(\hat{T}_j|T_j) \approx I(T_j, \hat{T}_j)^{-1} \quad (15)$$

where

$I(T_j, \hat{T}_j)$ is the information that $\hat{T}_j$ contributes about $T_j$.

Given these results, Lord (1980, p. 79 and 85) produces

$$I(T_j, \hat{T}_j) = \frac{I(\theta, \hat{T}_j)}{\left(\partial T_j/\partial\theta\right)^2}, \quad (16)$$

and

$$I(\theta, \hat{T}_j) \approx I(\theta, \hat{\theta}). \quad (17)$$

Thus,

$$\sigma^2(\hat{T}_j|\theta) \approx \frac{\left[\frac{1}{n_j}\Sigma_{i=1}^{k_j} \hat{T}_{ij}(\theta)\right]^2}{I(\theta, \hat{\theta})}$$

and the parameters of the prior beta distribution for $T_j$ can be expressed in terms of the parameters of the 3PL IRT and 2PPC models. Furthermore, the parameters of the posterior distribution of $T_j$ also can be expressed in terms of the IRT parameters:

$$p_j = \hat{T}_j n_j^* + x_j, \quad (18)$$

and

$$q_j = \left[1 - \hat{T}_j\right]n_j^* + n_j + x_j. \quad (19)$$

The SPI is

$$\tilde{T}_j = \frac{p_j}{p_j + q_j} \quad (20)$$

$$= \frac{\hat{T}_j n^*_j{}_+ + x_j}{n^*_j + n_j}. \quad (21)$$

The SPI can also be written in terms of the relative contribution of the prior estimate $\hat{T}_j$ and the observed proportion of maximum raw (correct score) (OPM), $x_j/n_j$, as

$$\tilde{T}_j = w_j \hat{T}_j + (1 - w_j) [x_j/n_j]. \quad (22)$$

$w_j$, a function of the mean and variance of the prior distribution, is the relative weight given to the prior estimate:

$$w_j = \frac{n^*_j}{n^*_j + n_j}. \quad (23)$$

The term $n^*_j$ may be interpreted as the contribution of the prior in terms of theoretical numbers of items.

**Check on Consistency and Adjustment of Weight Given to Prior Estimate**

The item responses are assumed to be described by $P_i(\hat{\theta})$ or $P_{im}(\hat{\theta})$, depending on the type of item. Even if the IRT model accurately described item performance over examinees, their item responses grouped by strand may be multidimensional. For example, a particular examinee may be able to perform difficult addition but not easy subtraction. Under these circumstances, it is not appropriate to pool the prior estimate, $\hat{T}_j$, with $x_j/n_j$. In calculating the SPI, the following statistic was used to identify examinees with unexpected performance on the strands in a test:

$$Q = \sum_{j=1}^{J} n_j \left(\frac{x_j}{n_j} - \hat{T}_j\right)^2 / (\hat{T}_j(1 - \hat{T}_j)). \quad (24)$$

If $Q \leq \chi^2 (J, .10)$, the weight, $w_j$, is computed and the SPI is produced. If $Q > \chi^2 (J, .10)$, $n^*_j$ and subsequently $w_j$ is set equal to 0 and the OPM is used as the estimate of strand performance.

As previously noted, the prior is estimated using an ability estimate based on responses to all the items (including the items of strand $j$) and hence is not independent of $X_j$. An adjustment for the overlapping information that requires minimal computation is to multiply the test information in equation 5 by the factor $(n - n_j)/n$. The application of this factor produces an "adjusted" SPI estimate that can be compared to the "unadjusted" estimate.

## Possible Violations of the Assumptions

Even if the IRT model fits the test items, the responses for a given examinee, grouped by strand, may be multidimensional. In these cases, it would not be appropriate to pool the prior estimate, $\widehat{T}_j$, with $x_j/n_j$. A chi-square fit statistic is used to evaluate the observed proportion of maximum raw score (OPM) relative to that predicted for the items in the strand on the basis of the student's overall trait estimate. If the chi-square is significant, the prior estimate is not used and the OPM obtained becomes the student's strand score.

If the items in the strand do not permit guessing, it is reasonable to assume $\widehat{T}_j$, the expected proportion correct of the maximum score for a strand, will be greater or equal to zero. If correct guessing is possible, as it is with MC items, there will be a non-zero lower limit to $\widehat{T}_j$, and a three-parameter beta distribution, in which $\widehat{T}_j$ is greater than or equal to this lower limit (Johnson & Kotz, 1979, p. 37), would be more appropriate. The use of the two-parameter beta distribution would tend to underestimate $T_j$ among very low-performing examinees. While working with tests containing exclusively MC items, Yen found that there does not appear to be a practical importance to this underestimation (Yen, 1997). The impact of any such effect would be reduced as the proportion of CR items in the test increases. The size of this effect, nonetheless, was evaluated using simulations (Yen, Sykes, Ito, & Julian, 1997).

The SPI procedure assumes that $p(X_j T_j)$ is a binomial distribution. This assumption is appropriate only when all the items in a strand have the same Bernoulli item response function. Not only do real items differ in difficulty, but when there are mixed-item types, $X_j$ is not the sum of $n_j$ independent Bernoulli variables. It is instead the total raw score. In essence, the simplifying assumption has been made that each CR item with a maximum score of $1_j - 1$ is the sum of $1_j - 1$ independent Bernoulli variables. Thus, a complex compound distribution is theoretically more applicable than the binomial. Given the complexity of working with such a model, it appears valuable to determine if the simpler model described here is sufficiently accurate to be useful.

Finally, because the prior estimate of $T_j$, $\widehat{T}_j$ is based on performance on the entire test, including strand $j$, the prior estimate is not independent of $X_j$. The smaller the ratio $n_j/n$, the less impact this dependence will have. The effect of the overlapping information would be to understate the width of the credibility interval. The extent to which the size of the credibility interval is too small was examined (Yen et al, 1997) by simulating strands that contained varying proportions of the total test points.

# References

CTB/McGraw-Hill. (2000). *TerraNova* 2nd Edition. Monterey, CA.

Fitzpatrick, A. R., V. Link, W. M. Yen, G. Burket, K. Ito & R. Sykes (1996). Scaling performance assessments: A comparison between one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, 33, 291–314.

Johnson, N. L. & S. Kotz (1970). *Distributions in statistics: Continuous univariate distributions* (Vol. 2). New York: John Wiley.

Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159–176.

Novick, M. R. & P. H. Jackson (1974). *Statistical methods for educational and psychological research*. New York: McGraw-Hill.

Yen, W. M. (1997). The technical quality of performance assessments: Standard errors of percents of students reaching standards. *Educational Measurement: Issues and Practice*, 5–15.

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

Yen, W. M., R. C. Sykes, K. Ito & M. Julian (1997). *A Bayesian/IRT index of objective performance for tests with mixed-item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.