Wisconsin Forward Exam

Spring 2021
Technical Report

WISCONSIN DEPARTMENT OF
Public Instruction

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

PART 7

PART 8

PART 9

TABLE OF FIGURES

PART 6

Executive Summary

This report is a technical summary of the 2021 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 are aligned with Wisconsin Academic Standards. The ELA, Mathematics, and Science test forms administered in Spring 2021 were developed by Data Recognition Corporation (DRC) using DRC's college- and career-ready item bank. The Spring 2021 Social Studies assessments contained Wisconsin-owned items. For the Spring 2021 administration, operational test forms from the Spring 2019 administration were reused or reused with modifications. The Wisconsin Department of Public Instruction (DPI) made a decision to reuse previously administered forms following test cancellation in Spring 2020 due to the COVID-19 pandemic. 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 later 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. The new reporting scales for the ELA and Mathematics tests were established after the Spring 2016 test administration, and the new performance level cut scores were set for these assessments in Summer 2016. The ELA and Mathematics 2015–16 results are considered a new baseline for year-to-year student performance comparisons. The 2020–21 ELA and Mathematics assessments are statistically linked to the established scales allowing for test score comparability between Spring 2019 and Spring 2021.

Science (grades 4 and 8) assessments 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) 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 2020–21 Science assessments are statistically linked to the scales established in Spring 2019, allowing for test score comparability between the two administrations.

Social Studies (grades 4, 8, and 10) assessments continue to be aligned with the Wisconsin Model Academic Standards (WMAS). New scales were developed for the Social

Studies tests under the new Wisconsin Forward Exam program in Spring 2016. Following the new scale development, the new performance level cut scores were set for Social Studies in Summer 2016. The 2020–21 Social Studies assessments are statistically linked to the scales established in Spring 2016, allowing for score comparability across years.

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) items. Braille, Large Print, 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 2021, 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 22 to May 14, 2021. Test administration is discussed in Part 4 of this report.

A total of 447 public school districts and 360 choice schools had students expected to take 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 2021. 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 2021. 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 8 ranged from approximately 84% to 87% across all content areas. The participation rate for Social Studies grade 10 was approximately 75%. These participation rates were considerably lower than the participation rates of over 95% for grades 3 through 8 or the participation rate of approximately 94% for grade 10 observed in a typical administration year (presented in Part 10 of the report). Further analysis of the Spring 2021 participation rates is provided in Part 4 of this report.

E.3 Student Performance

This is the fifth year of the ELA, Mathematics, and Social Studies scores being reported on the scales established in Spring 2016. Spring 2021 also marks the second year of the Science assessments measuring the new Wisconsin Standards for Science. Caution should be used when making statewide data comparisons over time as the participation rates for Spring 2021 are considerably lower than previous years. In addition, the makeup of the Spring 2021 tested population is not representative of the Spring 2021 enrolled population. Tables E-2 and E-3 present the percentages of students classified as *Proficient* or *Advanced* in 2016 through 2021 in ELA and Mathematics, respectively. Table E-4 shows the percentages of students classified as *Proficient* or *Advanced* in 2019 and 2021 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 and previous administrations and the previous data are not reported in this table. Student results are comparable between the Spring 2019 and Spring

2021 administrations for Science. Table E-5 shows the percentages of students classified as *Proficient* or *Advanced* in 2016 through 2021 in Social Studies.

The percentages of students classified as *Proficient* or *Advanced* in ELA in 2021 ranged from approximately 35% for grade 3 to approximately 43% for grade 7. The percentages of students classified as *Proficient* or *Advanced* in Mathematics in 2021 ranged from 30% for grade 8 to 45% for grade 3. Approximately 51% of students were classified as *Proficient* or *Advanced* in both Science grades in 2021. The percentages of students classified as *Proficient* or *Advanced* in Social Studies in 2021 ranged from 47% for grade 10 to 49% for grade 4. 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 2021 compared to Spring 2019 for all grades and content areas, except for Social Studies, grade 10. This observed change in performance between Spring 2019 and Spring 2021 should be interpreted in the context of circumstances related to the COVID-19 pandemic, including school closures, nonstandard instruction delivery modes in the 2020–21 school year, potential diminished opportunity to learn for students, and other unknown effects of the pandemic on students and their families.

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, Large Print, 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 MC items was conducted using item keys. All MC items had only one correct response. TE, SA, MS, and EBSR items followed predefined scoring criteria and rubrics. All items on the Wisconsin Forward Exam were autoscored (refer to Part 5 of this report for details).

Because the Spring 2019 test forms were reused, or reused with modifications, for the Spring 2021 administration, the pre-equated design was implemented for the purpose of student scoring. That is, students' scale scores were derived using item parameters estimated after the Spring 2019 test administration (or after the most recent past test administration). The post-equating verification was conducted to evaluate student score and parameter stability. The item response theory (IRT) models used for Wisconsin Forward Exam scaling were appropriate for the test data supporting the operational data analysis and ensured that the test items, as well as the overall tests, were functioning appropriately. For details on the IRT models used for Wisconsin Forward Exam test calibration, scaling, and post-equating verification, refer to Part 6 of this report. The cut scores used for classification of students into different performance levels and associated performance level descriptors were established during the Summer 2016 standard setting for ELA, Mathematics, and Social Studies and in the Spring 2019 standard setting for Science in a collaborative and participatory process, further supporting the validity and interpretation of the Wisconsin Forward Exam scores (refer to Part 7 of this report 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 was evaluated through differential item functioning (DIF) 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. 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 2021

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	60785	52930	87.08	52892	87.01				
4	61127	52706	86.22	52658	86.15	52417	85.75	52392	85.71
5	62405	54010	86.55	53932	86.42				
6	64925	55511	85.50	55462	85.42				
7	66361	56295	84.83	56247	84.76				
8	67572	56756	83.99	56726	83.95	56485	83.59	56409	83.48
10	68221							51433	75.39

Table E-2 Percentage of Students Classified as *Proficient* or *Advanced* in 2016 through 2021, English Language Arts

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

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 95%.

Table E-3 Percentage of Students Classified as *Proficient* or *Advanced* in 2016 through 2021, Mathematics

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

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 95%.

Table E-4 Percentage of Students Classified as *Proficient* or *Advanced* in 2019 and 2021, Science

Grade	Science	
	2019	**2021**
4	52.78	51.16
8	53.95	51.47

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 95%.

Table E-5 Percentage of Students Classified as *Proficient* or *Advanced* in 2016 through 2021, Social Studies

Grade	Social Studies				
	2016	**2017**	**2018**	**2019**	**2021**
4	52.93	52.04	53.66	52.49	49.19
8	49.78	50.03	52.21	51.60	48.41
10	48.50	48.17	48.20	45.50	47.21

Note: Caution should be exercised when interpreting the Spring 2021 statewide data due to participation rates below 95%.

Part 1: Overview

The *Wisconsin Forward Exam Spring 2021 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 2021 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 2021 was the fifth administration year for the Wisconsin Forward Exam in ELA, Mathematics, and Social Studies. Following the test cancellation in Spring 2020, the operational test forms from Spring 2019 were reused, or reused with modifications, in Spring 2021 for ELA, Mathematics, and Social Studies. In order to shorten test time for students, text-dependent analysis writing items were removed from ELA test forms. These items were replaced by two or three autoscored items measuring the same content sub-domain. The ELA test blueprints were not affected by this change. In addition, embedded field test items were removed from ELA and Mathematics test forms in Spring 2021. Three items in the Social Studies grade 8 form and one item in the Social Studies grade 10 form were replaced with items measuring the same content standards due to potential COVID-19 sensitivity. Social Studies tests included embedded field test items needed for development of the Spring 2022 Social Studies assessments based on the new Wisconsin Standards for Social Studies. The Spring 2021 ELA, Mathematics, and Social Studies assessments were statistically linked to their respective Spring 2019 scales, allowing for student score comparisons across the last five years of these assessments.

Spring 2021 is also the second 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* (https://www.nap.edu/read/13165/chapter/1). The Science operational test forms administered in Spring 2019 were reused for the Spring 2021 administration. In order to shorten the test time for students, the embedded field test items were removed from the tests. The Spring 2021 Science assessments were statistically linked to the Spring 2019 Science scales, allowing for student score comparisons across the last two years of Science assessment administration.

This Technical Report documents all aspects of the 2020–21 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, Mathematics, and Social Studies 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. 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 2021 test results should be considered in the context of factors related to the COVID-19 pandemic, including school closures, nonstandard instruction delivery modes in the 2020–21 school year, potential diminished opportunity to learn for students, and other unknown effects of the 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 and for grades 4, 8, and 10 in Social Studies occurred in December 2015. The first item review for new assessments for grades 4 and 8 in Science occurred in August 2017. Each year after that, 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 2021 Wisconsin Forward Exam forms, including their origination in the Spring 2019 test forms. The Spring 2021 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 2021. In the 2020–21 school year, the Wisconsin Forward Exam was administered to Wisconsin students for the fifth time.

The Spring 2021 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 an eight-week window from March 22 to May 14, 2021. 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 and autoscoring of technology-enhanced (TE) items, short-answer (SA) items, and evidence-based selected response (EBSR) items.

Part 6: Psychometric Analyses

The Spring 2021 administration year is the fifth administration year for the Wisconsin Forward Exam in all grades and content areas. Part 6 describes the classical and IRT procedure implemented to analyze the Wisconsin Forward Exam test data. The results of item analysis and statistical properties are presented in Part 6. The Wisconsin Forward Exam assessments were pre-equated in Spring 2021, and student responses were scored using item parameters that were obtained in the most recent (prior to Spring 2021) test administration. Item-pattern scoring was applied to the Spring 2021 Wisconsin Forward Exam. Part 6 explains that 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. Post-equating verification procedures and results, including presentation of the calibration sample, item calibration, test equating, and comparison of pre- and post-equated student scores, are presented in Appendix G of this report.

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, Mathematics, and Social Studies tests in Summer 2016 and for the Science tests in Spring 2019. 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.

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 2021 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.

Part 11: Summary and Recommendations

Key findings of the Spring 2021 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 2021 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 2021 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 are autoscored. 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 for all 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, Mathematics, and Social Studies, conducted in June 2016, and for Science, conducted in May 2019, 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 2021 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. 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 2021 administration and longitudinal data trends for all content areas. 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 2021 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 at each content area and grade level. Specifications for a range of items by standard and item type demonstrated the desired proportions within the summative assessment. In summary, the Wisconsin Forward Exam assessment blueprint at 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 2021 operational assessments included the use of the Spring 2019 operational test forms and additional 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 2021 Wisconsin Forward Exam assessments.

3.2.1 English Language Arts

Table 3-6 shows the ELA test design, including the number of passage sets, items, and points at each grade level that were used in the operational test forms. No field test items were

embedded in the Spring 2021 forms. 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 Wisconsin Forward Exam was divided into three sessions: Reading, Writing/Language, and Listening. 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 operational test forms. No field test items were embedded in the Spring 2021 forms.

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, but students were allowed to use a calculator in the second session. 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 operational test forms. No field test items were embedded in the Spring 2021 forms.

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 all 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 administered at each grade level. Fifteen field test item sets (called field test forms) were administered in grade 4, and thirteen field test forms were administered in grades 8 and 10. Each field test item set consisted of 8 items. Each student was administered only one set of field test items embedded in the operational test form. The Social Studies exam included two test sessions that could be administered in either order. The Social Studies exam 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 2021 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 and Large Print test versions were constructed for each grade and content area to enable students who are blind or visually impaired to participate in the Wisconsin Forward Exam

testing. Braille and Large Print 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, Science, and Social Studies test forms from the Spring 2019 administration were reused, or reused with modifications, in Spring 2021.

Test items included in the ELA, Mathematics, and Science assessments were selected from DRC's CCR item bank. DRC's CCR item bank contains nationally field-tested CCR items that support the next generation of standards and assessments. It is 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 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, or are preparing to adopt, more rigorous content standards, curricula, and assessments that better prepare students for college and careers.

Alignment to standards, grade-level appropriateness, 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* (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 *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 more recently by 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 content specialists. These items are aligned to Wisconsin's Model Academic Standards for Social Studies. 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 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. Because the Spring 2019 operational forms were reused, or reused with modifications, for the Spring 2021 administration, the review of most of the items included in these forms occurred in Summer 2017. As part of the passage and item review process, Wisconsin educators made the determinations of standard alignment, DOK levels, and key(s). They noted any bias and sensitivity concerns and had the opportunity to determine whether items were accepted as is or accepted with revisions. They also had the opportunity to register a "dissenting view" in which the committee preferred the item not be selected to appear on the Wisconsin Forward Exam in a field test position. More information on the item and passage review for items later included in the Spring 2019 operational test forms can be found in the *Wisconsin Forward Exam Spring 2019 Technical Report*, available on the DPI website at https://dpi.wi.gov.

3.5 Field-Testing of Items Included in Spring 2021 Forms

With the exception of two or three items per grade on the ELA tests, all items included in the Spring 2021 operational test forms were previously administered operationally and no additional field-testing of these items was needed. ELA autoscored items that were selected as

replacements for text-dependent analysis (TDA) items that were field-tested in Spring 2018 or Spring 2019 during the operational test administration. These field test items were fully embedded in the operational forms, and students were not able to distinguish between the operational and field test items. Following the field test data analysis and item review with data, accepted items were included in the Wisconsin item bank for future use. For detailed information on item field-testing, field test data analysis, and review of items with data, refer to Part 2 "Test Blueprint and Item Development" of the *Wisconsin Forward Exam Spring 2019 Technical Report*, available on the DPI website at https://dpi.wi.gov.

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 complied 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. For details of the form development and item selection process in a typical administration cycle, refer to Part 3 "Test Form Development" of the *Wisconsin Forward Exam Spring 2019 Technical Report*, available on the DPI website at https://dpi.wi.gov.

3.6.1 Form Reuse and Modifications

Because of testing cancellation in Spring 2020 due to the COVID-19 pandemic, DPI decided to reuse the Spring 2019 test forms. A benefit of reusing the 2019 forms in Spring 2021 is that not only total test scores can be directly compared across administrations but also domain-level scores and item-level statistics, provided that no items are added or removed from scoring. This offers a unique opportunity to compare pre- and post-COVID-19 student performance on the same sets of items (CCSSO, 2020).

At the same time, DPI decided to shorten the test time in order to reduce the burden on students and educators in Spring 2021. The test time was shortened by removing embedded field

test items from ELA, Mathematics, and Science assessments and removing operational TDA items from the ELA assessments in all grades in Spring 2021. The TDA items were subsequently replaced by two or three autoscored items to enable reporting for the Text Types and Purposes subscore. The TDA replacement items were previously field-tested items with acceptable statistics. This change did not negatively impact the ELA test blueprint. When the TDA items were replaced with autoscored items, the difference between the Spring 2021 and target test blueprints was approximately 4% for the Reading domain, about 6% for the Writing/Language domain, and about 1% for the Listening domain. Typically, blueprint differences up to 10% between administrations are considered to be acceptable.

All operational items and passages on the Spring 2019 operational forms were reviewed for potential COVID-19 sensitivity issues. No items with problematic content were identified on the ELA, Mathematics, and Science forms. Three items on the Social Studies grade 8 test and one item on the Social Studies grade 10 test were found to be potentially sensitive to issues related to COVID-19 and were replaced by items with neutral content. The replacement items were operational items with acceptable statistics, previously administered operationally in Spring 2016 or 2017. This change did not impact the Social Studies test blueprint.

After ELA grades 3 through 8 and Social Studies grades 8 and 10 forms were updated, these forms were reviewed by DPI staff and DRC content experts. In addition, new field test items were added to the Social Studies forms in each grade.

The test maps in Appendices A, B, C, and D provide details on the operational items on the Spring 2021 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 A, the Mathematics test maps are contained in Appendix B, the Science test maps are provided in Appendix C, and the Social Studies test maps are given in Appendix D.

3.7 DPI Approvals

DPI had the opportunity to review passages and items placed on the Spring 2021 Wisconsin Forward Exam during the following phases:

- prior to item content review in Summer 2017 and 2018
- at item content review in Summer 2017, 2018, and 2019
- during review of flagged field test data in Summer 2018 and 2019
- during the Spring 2019 form construction
- after the Spring 2021 form update

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 2021 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 2021 Wisconsin Forward Exam were reviewed by DRC, DPI, and Wisconsin educators for issues regarding accessibility, bias, sensitivity, and content. During the reviews, item content was checked for the accuracy of the content, fairness, accessibility, developmental appropriateness, alignment to content specifications, answer keys, and scoring rules. The efforts and procedures used in the development of the Spring 2021 Wisconsin Forward Exam forms balanced the content and psychometric requirements for the form development. The psychometric properties of the updated test forms (ELA all grades, and Social Studies grades 8 and 10) were comparable to the psychometric properties of the Spring 2019 forms for all content areas. Overall, the process implemented in the Spring 2021 operational form development was in alignment with multiple best practices of the testing industry.

Table 3-1 English Language Arts Test Blueprints for 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		19	19	19	19	19	19
Text Types and Purposes/Text-Dependent Analysis	all grades: 2–3	6–10	6–10	6–10	6–10	6–10	6–10
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		**48**	**51**	**51**	**51**	**51**	**51**

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

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 Science Test Blueprints for Grades 4 and 8

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 Social Studies Test Blueprints for Grades 4, 8, and 10

Reporting Category	Depth of Knowledge	Total Points by Grade		
		4	8	10
Geography: People, Places, and Environments	all grades: 1–3	7–11	8–12	9–11
History: Time, Continuity, and Change	all grades: 1–3	6–10	10–15	11–14
Political Science and Citizenship: Power, Authority, Governance, and Responsibility	grade 4: 2–3 grades 8, 10: 1–3	5–9	5–7	11–14
Economics: Production, Distribution, Exchange, and Consumption	all grades: 1–3	5–9	5–7	7–10
The Behavioral Sciences: Individuals, Institutions, and Cultures	all grades: 2–3	5–9	4–6	7–10
Social Studies Total Points		**38**	**40**	**50**

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. Each of these items is worth one point.
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.
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.
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. The technology-enhanced items are scored on a 0–2-point scale using item-specific scoring rules.

Table 3-6 English Language Arts Test Design

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Passage Sets	**Literature**	2	2	2	2	2	2
	Informational	1	2	2	2	2	2
	Listening	2	2	2	2	2	2
Number of Core (OP) Items	**Item Types: MC/TE (1 pt)**	28	31	33	27	23	29
	Item Types: MS/TE/EBSR (2 pts)	10	10	9	12	14	11
	Total Core Items	38	41	42	39	37	40
Total Core Points		48	51	51	51	51	51
Total Estimated Testing Time (minutes)		70	70	70	70	70	70

Note: TDA items were replaced by autoscored items on the Spring 2021 ELA test forms.

Table 3-7 Mathematics Test Design

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Core (OP) Items	**Item Types: MC/SA (1 pt)**	37	41	39	40	41	40
	Item Type: TE (1 pt)	5	5	7	6	5	6
	Total Core Items	42	46	46	46	46	46
Total Core Points		42	42	46	46	46	46
Total Estimated Testing Time (minutes)		80	80	80	95	95	105

Table 3-8 Science Test Design

Test Design		Grade	
		4	8
Number of Core (OP) Items	Item Types: MC/MS/TE/EBSR (1 pt)	40	40
Total Core Items and Points		40	40
Total Estimated Testing Time (minutes)		90	90

Table 3-9 Social Studies Test Design

Test Design		Grade		
		4	8	10
Number of Core (OP) Items	Item Types: MC/TE/MS (1 pt)	38	40	50
Total Core Points		38	40	50
Embedded Field Test (FT)	Number of Forms	15	13	13
	FT Items per Form	8	8	8
	Total Items Field-Tested	120	102	101
Total Items (Core + FT) per Form		46	48	58
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.

Part 4: Test Administration

In the Spring of 2021, 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 22–May 14, 2021. Part 4 of the Technical Report provides information on student participation rates in the Spring 2021 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 limited English proficient (FLEP) 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. These test participation rates for the students in grades 3 through 8 ranged from approximately 84% to 87% across content areas. The participation rate in Social Studies grade 10 was approximately 75%. The Spring 2021 test participation rates were lower than participation rates observed in a typical administration year. Typically, at least 95% of eligible students take the Wisconsin Forward Exam in grades 3 through 8 and approximately 94% of students take the Social Studies test in grade 10. However due to circumstances related to the Covid-19 pandemic, participation in the Spring 2021 administration was lower. When test participation rates by demographic groups were considered, it was observed that students from ethnic minority groups: African American, Hispanic, American Indian and Asian, and students from historically disadvantaged groups: limited English proficiency, students with disabilities or economically disadvantaged students participated in the assessment at lower rates compared to White students, fully English proficient students, students without disabilities or students who were not economically disadvantaged. Detailed information on the test participation rates in Spring 2021 for all students and disaggregated by demographic characteristics is provided in Appendix E.

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 2021 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 2021 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 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 online webinars DACs, SACs, and TAs for the Spring 2021 administration of the Wisconsin Forward Exam. The webinars were recorded by DPI and 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 webinars 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.

Accommodation 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
Braille [BRL]	2	0.00	1	0.00
Print on Demand [POD]	2	0.00	2	0.00
Bilingual Dictionary			124	0.23
Magnification	59	0.11	60	0.11
Noise Buffers	622	1.18	618	1.17
Read Aloud	351	0.66	358	0.68
Scribe	308	0.58	295	0.56
Separate Setting	6400	12.09	6386	12.07
Alternate Response Options	15	0.03	15	0.03
Read Aloud (Reading Passages)	4	0.01	4	0.01
Provided Color Choices [CC]	53	0.10	53	0.10
Contrasting Color [CTC]	45	0.09	44	0.08
Reverse Contrast [RC]	21	0.04	21	0.04
Masking [MSK]	523	0.99	526	0.99
Text-to-Speech [TTS]	10302	19.46	10384	19.63
Spanish Translation [ST]	267	0.50	412	0.78
Video Sign Language [VSL (ASL)]	10	0.02	9	0.02
Color Overlay	24	0.05	24	0.05
Amplification Device	61	0.12	61	0.12
Small Group Translation	59	0.11	78	0.15
Translator/Interpreter	18	0.03	23	0.04
Read Aloud in Spanish	62	0.12	87	0.16
Closed Captioning [C CAP] ELA	51	0.10		
Listening Scripts [LS] ELA	7	0.01		
Abacus Math			31	0.06
Non-embedded Calculator Math			135	0.26
Multiplication Table Math			2	0.00

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

Grade 4	English Language Arts		Mathematics		Science		Social Studies	
Accommodation or Support	N Count	Percent	N Count	Percent	N Count	Percent	N Count	Percent
Braille [BRL]	2	0.00	2	0.00	2	0.00	2	0.00
Bilingual Dictionary			141	0.27	139	0.27	139	0.27
Magnification	49	0.09	49	0.09	48	0.09	48	0.09
Noise Buffers	588	1.12	588	1.12	587	1.12	583	1.11
Read Aloud	352	0.67	348	0.66	340	0.65	338	0.65
Scribe	315	0.60	303	0.58	301	0.57	302	0.58
Separate Setting	6763	12.83	6763	12.84	6679	12.74	6652	12.70
Alternate Response Options	14	0.03	13	0.02	14	0.03	14	0.03
Read Aloud (Reading Passages)	3	0.01	3	0.01	3	0.01	3	0.01
Provided Color Choices [CC]	66	0.13	66	0.13	67	0.13	65	0.12
Contrasting Color [CTC]	64	0.12	64	0.12	64	0.12	64	0.12
Reverse Contrast [RC]	34	0.06	34	0.06	34	0.06	34	0.06
Masking [MSK]	506	0.96	506	0.96	504	0.96	502	0.96
Text-to-Speech [TTS]	9966	18.91	10012	19.01	9914	18.91	9904	18.90
Spanish Translation [ST]	322	0.61	407	0.77	377	0.72	367	0.70
Video Sign Language [VSL (ASL)]	7	0.01	6	0.01	7	0.01	6	0.01
Color Overlay	21	0.04	21	0.04	21	0.04	21	0.04
Amplification Device	37	0.07	37	0.07	36	0.07	36	0.07
Small Group Translation	64	0.12	75	0.14	64	0.12	62	0.12
Translator/Interpreter	20	0.04	22	0.04	21	0.04	22	0.04
Read Aloud in Spanish	46	0.09	74	0.14	66	0.13	65	0.12
Closed Captioning [C CAP] ELA	42	0.08						
Listening Scripts [LS] ELA	3	0.01						
Abacus Math			17	0.03				
Non-embedded Calculator Math			191	0.36				
Multiplication Table Math			1587	3.01				

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
Print on Demand [POD]	2	0.00	2	0.00
Bilingual Dictionary			71	0.13
Magnification	69	0.13	68	0.13
Noise Buffers	741	1.37	739	1.37
Read Aloud	333	0.62	334	0.62
Scribe	303	0.56	285	0.53
Separate Setting	6635	12.28	6613	12.26
Alternate Response Options	8	0.01	6	0.01
Read Aloud (Reading Passages)	3	0.01	3	0.01
Provided Color Choices [CC]	78	0.14	78	0.14
Contrasting Color [CTC]	64	0.12	64	0.12
Reverse Contrast [RC]	36	0.07	36	0.07
Masking [MSK]	537	0.99	538	1.00
Text-to-Speech [TTS]	8736	16.17	8807	16.33
Spanish Translation [ST]	319	0.59	415	0.77
Video Sign Language [VSL (ASL)]	14	0.03	14	0.03
Color Overlay	30	0.06	29	0.05
Amplification Device	56	0.10	56	0.10
Small Group Translation	49	0.09	56	0.10
Translator/Interpreter	21	0.04	23	0.04
Read Aloud in Spanish	27	0.05	33	0.06
Closed Captioning [C CAP] ELA	52	0.10		
Listening Scripts [LS] ELA	8	0.01		
Abacus Math			7	0.01
Non-embedded Calculator Math			283	0.52
Multiplication Table Math			2030	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
Braille [BRL]	6	0.01	6	0.01
Bilingual Dictionary			92	0.17
Magnification	90	0.16	91	0.16
Noise Buffers	425	0.77	426	0.77
Read Aloud	291	0.52	277	0.50
Scribe	206	0.37	187	0.34
Separate Setting	6245	11.25	6220	11.21
Alternate Response Options	9	0.02	8	0.01
Read Aloud (Reading Passages)	4	0.01	4	0.01
Provided Color Choices [CC]	94	0.17	92	0.17
Contrasting Color [CTC]	97	0.17	97	0.17
Reverse Contrast [RC]	22	0.04	21	0.04
Masking [MSK]	427	0.77	423	0.76
Text-to-Speech [TTS]	7398	13.33	7511	13.54
Spanish Translation [ST]	284	0.51	297	0.54
Video Sign Language [VSL (ASL)]	6	0.01	5	0.01
Color Overlay	24	0.04	24	0.04
Amplification Device	47	0.08	46	0.08
Small Group Translation	64	0.12	44	0.08
Translator/Interpreter	8	0.01	9	0.02
Read Aloud in Spanish	52	0.09	38	0.07
Closed Captioning [C CAP] ELA	59	0.11		
Listening Scripts [LS] ELA	5	0.01		
Abacus Math			7	0.01
Non-embedded Calculator Math			368	0.66
Multiplication Table Math			2233	4.03

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
Braille [BRL]	5	0.01	5	0.01
Bilingual Dictionary			132	0.23
Magnification	71	0.13	72	0.13
Noise Buffers	314	0.56	313	0.56
Read Aloud	310	0.55	318	0.57
Scribe	167	0.30	157	0.28
Separate Setting	6127	10.88	6120	10.88
Alternate Response Options	8	0.01	6	0.01
Read Aloud (Reading Passages)	4	0.01	4	0.01
Provided Color Choices [CC]	87	0.15	88	0.16
Contrasting Color [CTC]	71	0.13	71	0.13
Reverse Contrast [RC]	37	0.07	36	0.06
Masking [MSK]	537	0.95	529	0.94
Text-to-Speech [TTS]	6945	12.34	6989	12.43
Spanish Translation [ST]	280	0.50	253	0.45
Video Sign Language [VSL (ASL)]	12	0.02	12	0.02
Color Overlay	29	0.05	30	0.05
Amplification Device	46	0.08	46	0.08
Small Group Translation	29	0.05	39	0.07
Translator/Interpreter	10	0.02	10	0.02
Read Aloud in Spanish	25	0.04	31	0.06
Closed Captioning [C CAP] ELA	54	0.10		
Listening Scripts [LS] ELA	3	0.01		
Abacus Math			10	0.02
Non-embedded Calculator Math			517	0.92
Multiplication Table Math			2166	3.85

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

Grade 8 Accommodation or Support	English Language Arts		Mathematics		Science		Social Studies	
	N Count	Percent	N Count	Percent	N Count	Percent	N Count	Percent
Braille [BRL]	3	0.01	3	0.01	3	0.01	3	0.01
Print on Demand [POD]	4	0.01	4	0.01	4	0.01	4	0.01
Bilingual Dictionary			140	0.25	139	0.25	134	0.24
Magnification	75	0.13	75	0.13	75	0.13	75	0.13
Noise Buffers	249	0.44	249	0.44	247	0.44	240	0.43
Read Aloud	232	0.41	230	0.41	227	0.40	226	0.40
Scribe	148	0.26	134	0.24	137	0.24	138	0.24
Separate Setting	6275	11.06	6272	11.06	6197	10.97	6178	10.95
Alternate Response Options	4	0.01	3	0.01	3	0.01	3	0.01
Read Aloud (Reading Passages)	2	0.00	2	0.00	2	0.00	2	0.00
Provided Color Choices [CC]	84	0.15	84	0.15	84	0.15	84	0.15
Contrasting Color [CTC]	81	0.14	82	0.14	81	0.14	81	0.14
Reverse Contrast [RC]	51	0.09	51	0.09	51	0.09	50	0.09
Masking [MSK]	323	0.57	329	0.58	322	0.57	318	0.56
Text-to-Speech [TTS]	6812	12.00	6857	12.09	6737	11.93	6714	11.90
Spanish Translation [ST]	222	0.39	255	0.45	249	0.44	256	0.45
Video Sign Language [VSL (ASL)]	11	0.02	9	0.02	8	0.01	8	0.01
Color Overlay	17	0.03	17	0.03	17	0.03	17	0.03
Amplification Device	44	0.08	42	0.07	42	0.07	41	0.07
Small Group Translation	31	0.05	30	0.05	30	0.05	32	0.06
Translator/Interpreter	16	0.03	16	0.03	17	0.03	17	0.03
Read Aloud in Spanish	18	0.03	22	0.04	22	0.04	23	0.04
Closed Captioning [C CAP] ELA	64	0.11						
Listening Scripts [LS] ELA	14	0.02						
Non-embedded Calculator Math			602	1.06				
Multiplication Table Math			2037	3.59				

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
Print on Demand [POD]	4	0.01
Bilingual Dictionary	87	0.17
Magnification	52	0.10
Noise Buffers	59	0.11
Read Aloud	161	0.31
Scribe	61	0.12
Separate Setting	3756	7.30
Provided Color Choices [CC]	17	0.03
Contrasting Color [CTC]	13	0.03
Reverse Contrast [RC]	5	0.01
Masking [MSK]	145	0.28
Text-to-Speech [TTS]	2696	5.24
Spanish Translation [ST]	137	0.27
Video Sign Language [VSL (ASL)]	4	0.01
Color Overlay	9	0.02
Amplification Device	17	0.03
Small Group Translation	69	0.13
Translator/Interpreter	11	0.02
Read Aloud in Spanish	10	0.02

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: • Managing user's own DRC INSIGHT Portal account • Managing other DRC INSIGHT Portal users • Adding and editing students and student demographics, accommodations, and testing codes • Viewing, adding, and editing student test session information • Printing and managing student test tickets • Transferring students between schools and districts • Entering Not-Tested or Invalidation Codes • Unlocking or purging a student test • Managing test sessions • 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 12 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). The 2021 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. Item samplers are accessible through the OTT menu page. They 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 Summary Table of Manual Materials (cont.)

Material	Configuration
TAM (Test Administration Manual)	The TAM is a document intended for test administrators (TAs) and proctors. It includes the following information: • Key dates • TA and proctor responsibilities • Test times • Test security • Accessibility information • Procedures for before testing • Test ticket management • Test administration scripts The TAM also includes information previously provided in the DAC/SAC Guide (District Assessment Coordinator/School Assessment Coordinator Guide), which includes the following: • Roles and responsibilities • Test security • Resources and training materials • Test schedules • DRC INSIGHT Portal and DRC INSIGHT secure browser • Accessibility • Student transfers • Procedures to be completed prior to the close of the testing window • 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: • Interpreting Wisconsin Forward Exam scores • Accessing Individual Student Reports (ISRs) and interactive summary reports via the DRC INSIGHT Portal

Table 4-8 Summary Table of Manual Materials (cont.)

Material	Configuration
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: • What is new and changing • Assessing online testing readiness • Capacity estimator • COS decision guide • Extended retries • Keyboard settings • Installation of COS and INSIGHT • Installation of INSIGHT App • Evaluation and troubleshooting • System requirements • Technology overview presentation • Technology Coordinator Checklist • Technology FAQ
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: • Description of the item pool used in the Wisconsin form-development process • Description of the test administration process and test security • Scoring of various types of items • 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) • Item- and test-level analysis information for each content area and grade level, test scaling procedure, and student scoring process • Evidence of test validity
Data Forensic Report	A separate Data Forensic Report includes analyses of the following: • Evaluation of response changes • 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, as well as the autoscoring process of technology-enhanced (TE), short-answer (SA), and evidence-based selected response (EBSR) items. In order to reduce testing time in Spring 2021, the text-dependent analysis (TDA) items were not administered as part of ELA tests.

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 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 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 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 item scoring of all item types. The scoring process presented in this part of the report shows that the items were scored reliably and accurately. 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.

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 post-equating verification. The classical analyses were conducted using the census data. Post-equating verification was conducted using the calibration samples.

6.1 Overview of the Operational Test Data Analysis

In 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 IRT models and rationale for the pre-equated design for the Wisconsin Forward Exam in Spring 2021 are presented, followed by the summary of the post-equating verification results and description of methodology of student scoring that occurred for the Wisconsin Forward Exam after the Spring 2021 test administration. 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 polytomously scored 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)

---

[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.

For the Spring 2021 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%.

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, 53 operational items across all Wisconsin assessments were flagged on the Spring 2021 operational tests as meeting the investigational criteria bulleted above. While the numbers of flagged items on ELA, Science, and Social Studies tests were comparable to the

numbers of items flagged on these tests in Spring 2019, more items were flagged for Mathematics in Spring 2021 compared to Spring 2019.

Table 6-1 shows the number of scored items in the Spring 2021 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 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. All flagged items were judged to be acceptable based on their other statistics and were retained in order to meet the Wisconsin Forward Exam test blueprints.

6.2.2 Flagging for the Item-Total Correlation

Between one and three items per grade in Mathematics grades 4 through 7 were flagged for item-total test correlation <0.15. In addition, one item in Science grade 8 was flagged for item-total test correlation <0.15. The item-total test correlations for the flagged items ranged from 0.06 to 0.14. These flagged items also tended to have close to or just above the flagging threshold item-total test correlations in Spring 2019.

6.2.3 Flagging for *p*-Value

Fourteen items in Mathematics assessments were flagged for *p*-values <0.20. The flagged items had *p*-values between 0.10 and 0.19. While these statistics indicate items that were very difficult, the number of items flagged for difficulty was found to be reasonable, given that COVID-19 pandemic–related learning disruptions likely occurred in the 2020–21 school year. 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 an omit rate 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% or more 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 distribution for non-multiple choice items are included in Appendix F.

6.3 Test-Level Statistics

Test-level statistics, including test reliability, were computed on the Spring 2021 Wisconsin Forward Exam data for students with available item responses. 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, 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 48

in grade 3 and 51 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. In Social Studies, the maximum raw scores are 38, 40, and 50 in grades 4, 8, and 10, respectively.

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 22.34 and the maximum possible score is 42, resulting in the test mean p-value of approximately 0.53.

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 Academic Standards for ELA, Science, and Social Studies. The Mathematics assessments in grades 4 through 8 tend 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 fourteen out of seventeen tests was zero, meaning that at least one student failed all items for each of those tests. The table also shows that the maximum obtained scores are 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 is at least one student who failed all items and at least one student who obtained a perfect 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 influence 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 0.61.
- Standard deviations ranged from 9.47 to 10.16 raw score points.
- Reliability coefficient alpha was relatively high in every grade (0.88 to 0.90).
- SEM ranged from 3.16 to 3.28.

Mathematics

- Test difficulty ranged from 0.41 to 0.53, with generally lower difficulty in lower grades and higher difficulty in higher grades.
- Standard deviations ranged from 9.25 to 10.15 raw score points.
- Reliability coefficient Alpha was high in every grade (0.90 to 0.93).
- SEM ranged from 2.68 to 2.90.

Science

- Test difficulty was 0.55 in grade 4 and 0.53 in grade 8.
- Standard deviations were 8.20 and 8.49 raw score points for grades 4 and 8, respectively.
- Reliability coefficient Alpha was 0.89 in both grades.
- SEM was 2.71 and 2.77 for grades 4 and 8, respectively.

Social Studies

- Test difficulty was 0.64 in grades 4 and 10, and 0.66 in grade 8.
- Standard deviations ranged from 8.10 to 10.28 raw score points.
- Reliability coefficient Alpha was 0.90 in grade 4 and 0.91 in grades 8 and 10.
- SEM ranged from 2.60 to 3.01.

6.4 Item Response Theory Methodology

This section of the report outlines 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.

Item parameters for items contained in Wisconsin Forward Exam tests are estimated using a marginal maximum-likelihood procedure to simultaneously estimate the item parameters for MC and non-MC items using the 3-parameter logistic (3PL) model and 2-parameter partial credit (2PPC) IRT model (Bock & Aitkin, 1981; Thissen, 1982). All non-MC items are treated as constructed-response (CR) items in the calibration. Under the 3PL model, the probability that a student with the trait or scale score θ 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 trait or scale score θ will respond in category k to partial-credit item j is

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

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

The summary output of the 3PL and 2PPC models is in two different metrics. The location and discrimination parameters for the MC items are in the traditional 3PL metric and are labeled b and a, respectively. In the 2PPC model, f (alpha) and g (gamma) are analogous to b and a, where alpha is the discrimination parameter and gamma over alpha (g/f) is the location where adjacent trace lines cross on the ability scale. Because of the different metrics used, the 3PL parameters b and a are not directly comparable to the 2PPC parameters f and g; however, they can be converted to a common metric. The two metrics are related by $b = g/f$ and $a = f/1.7$ (Burket, 2002). As a result of this procedure, the MC and non-MC 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.

Item parameters estimated after the Spring 2019 test administration (or in some cases older item parameters) were used to score Wisconsin students who took ELA, Mathematics, Science, and Social Studies test forms in Spring 2021. A summary of the scaling and equating procedures implemented for the Wisconsin Forward Exam in Spring 2019 is presented in the next section. A detailed description and the results of this process can be found in the *Wisconsin Forward Exam Spring 2019 Technical Report*, available on the DPI website at https://dpi.wi.gov.

6.4.1 Test Scaling and Equating in Spring 2019

Because the Spring 2019 (or older) item parameter estimates were used to score Wisconsin students after the Spring 2021 test administration, this section provides a short overview of the scaling and equating procedures implemented for ELA, Mathematics, Science, and Social Studies assessments after the Spring 2019 test administration.

The purpose of scaling a test is to enhance its validity by increasing the comparability of test takers' scores. The Wisconsin Forward Exam scales were established for ELA, Mathematics, and Social Studies after the Spring 2016 test administration. A common-item design was used to link the Spring 2019 tests to the established Wisconsin Forward Exam scales. Sets of previously administered operational items—representative of the test content and psychometric properties—were administered to Wisconsin students in Spring 2019 and served as the anchor (linking) items on each test form. After the initial IRT item calibration, the Spring 2019 item parameters were linked to the Wisconsin Forward Exam scales using the Stocking & Lord equating procedure (1983). Equating of new forms after each test administration allowed for longitudinal comparison of student performance on the ELA, Mathematics, and Social Studies assessments from Spring 2016 to Spring 2019.

New scales were established for Science assessments after the Spring 2019 test administration. The new Science test forms were not statistically linked to the previous Science scales. Instead, the student performance on the Science tests in Spring 2019 became a new baseline for the longitudinal score comparison.

A detailed description of the calibration, scaling, and equating procedures and results of the Wisconsin Forward Exam assessments can be found in the *Wisconsin Forward Exam Spring 2019 Technical Report*, in compliance with AERA, APA, & NCME (2014) Standard 5.2, 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)

6.4.2 Pre-equated Design

The term "pre-equating" refers to equating prior to the operational testing. Post-equating, similarly, refers to conducting equating after the operational testing. Both equating processes are widely used in K–12 large-scale assessment programs. With post-equating, statistical adjustment of test difficulty is applied to operational test data after the operational test administration, so the resultant parameters are sometimes considered to yield more accurate scores.

In the Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies pre-equated design, statistical procedures (e.g., calibration, scaling, equating) were applied to the assessments administered in Spring 2019. The item parameter estimates obtained after the Spring 2019 test administration were then used to score students who participated in the Spring 2021 Wisconsin Forward Exam. The exceptions were the following: one item in ELA grade 4, one

item in ELA grade 5, two items in ELA grade 6, three items in Social Studies grade 8, and one item in Social Studies grade 10, for which item parameters were estimated in Spring 2018, Spring 2017, or Spring 2016. Because the majority of the pre-equated item parameters were obtained in the operational test administration and the Wisconsin student populations participating in the ELA, Mathematics, Science, and Social Studies assessments are comparable between the administration years, the use of these older item parameters in student scoring was appropriate. In fact, research shows that item parameters and resultant student scores have been found to be acceptably accurate under pre-equating (Bejar & Wingersky, 1982; Tong, Wu, & Xu, 2008).

In addition, the use of pre-equating was particularly appropriate in Spring 2021 given the unknown effects of the COVID-19 pandemic on student learning in the 2020–21 school year and on the performance on the Spring 2021 assessments.

6.4.3 Post-equating Verification

As stated in the previous section, the Spring 2019 (or older) item parameters were used to score students who were administered the Wisconsin Forward Exam in 2021. Following the test administration and reporting, DRC conducted post-equating verification and item parameter evaluation. The 2021 operational test data were calibrated and equated to the 2019 scales using the item parameters from the past administrations as anchors. All test items served as anchors in post-equating. The item parameters estimated in post-equating were found to be stable and comparable to the pre-equated item parameters. The post-equating procedure and results are presented in Appendix G of this report. These procedures demonstrate compliance with AERA, APA, & NCME (2014) Standards 1.8, 5.13, and 5.15.

Student scores were computed using the pre-equated and post-equated item parameters. The mean scale scores, scale score standard deviations, and percentages of students classified in the four performance levels were found to be comparable when students were scored using the pre-equated versus post-equated item parameters (refer to Appendix G). These results further support the use of pre-equated parameters in ELA, Mathematics, Science, and Social Studies student scoring.

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). Pre-equated item parameters were used to derive student scale scores in the Spring 2021 test administration.

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.

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

In this model, θ 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 $P(u_i = 1 \mid \theta)$. That is, P is an estimation of the probability that a student with an ability value θ 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 ($u_1 u_2 \Lambda u_n$) on a test with some number of items (n items), the "likelihood function," or the probability that a student with a given ability value (θ) would produce this particular response pattern, is given by

$$P(u_1 u_2 \Lambda u_n \mid \theta) = \prod_{i=1}^{n} P(u_i \mid \theta). \qquad (2)$$

The formula indicates that the "estimated maximum likelihood" IRT item-pattern scoring method searches for the ability estimate (θ_0) that maximizes the probability function in the equation shown above (2) and assigns an ability estimate (θ_0) as the test score for the student with the response pattern ($u_1 u_2 \Lambda u_n$). 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 ($u_1 u_2 \Lambda\ u_n$)	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 2021 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 6-24 through 6-40. 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

Estimated Theta = 140.0

b) Craig

Estimated Theta = 246.0

c) Vicki

Estimated Theta = 266.0

d) Tom

Estimated Theta = 259.0

e) Evan

Estimated Theta = 265.0

f) Dan

Estimated Theta = 379.0
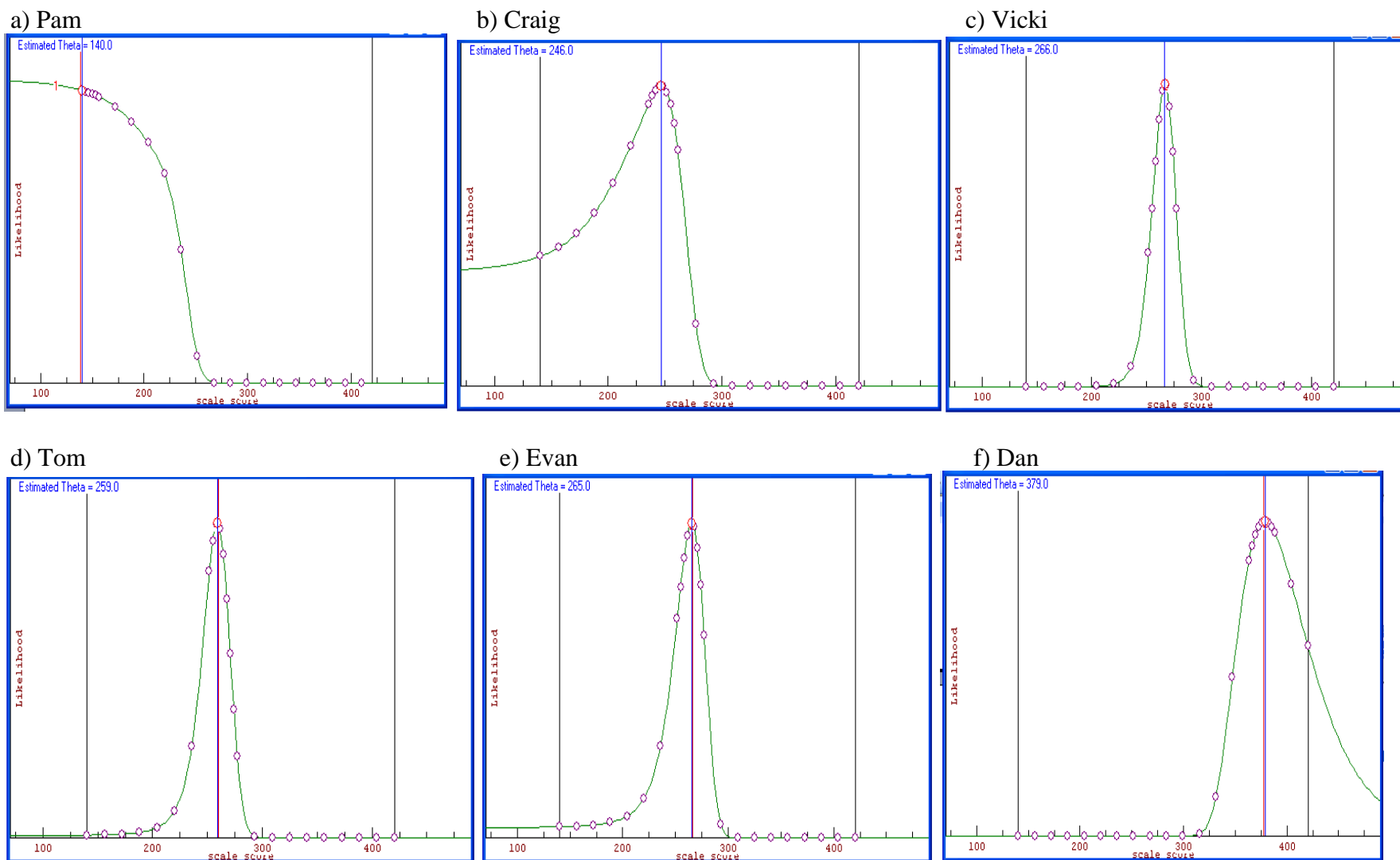
Note: The circular dots in the likelihood functions indicate that the software program used is searching for a maximum likelihood estimate (scale score) for the student.

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, Mathematics, and Social Studies were established after the Spring 2016 test administration and remained unchanged through the Spring 2021 test administration. New LOSS and HOSS values for Science were established after the Spring 2019 test administration.

Table 6-41 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.

Half of a percent or less of students in all grades of ELA and Science received the lowest obtainable scale scores. For Social Studies, the percentages of students at the LOSS were approximately one and a quarter percent or less. However, in Mathematics, all grades had more than two percent of students at the LOSS, ranging from 2.06% in grade 3 to 4.47% in grade 5. These percentages of students at the LOSS in Mathematics in Spring 2021 were higher than the corresponding percentages of students at the LOSS in Mathematics in Spring 2019, indicating that the Mathematics assessments were more difficult for some students in Spring 2021 than the same assessments were for the Spring 2019 student cohorts. The increase of students at LOSS in Mathematics may to some degree reflect disrupted instruction and learning in Spring 2020 and the 2020–21 school year, which could have affected Mathematics more than other subjects.

The response patterns of students at the LOSS in Mathematics were investigated after the Spring 2021 test administration. It was found that these students 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. The percentages of students scoring at the HOSS were less than one percent in all grades and content areas.

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	1	0	1	0	0
	4	2	0	2	0	0
	5	2	0	2	0	0
	6	2	0	2	0	0
	7	0	0	0	0	0
	8	2	0	2	0	0
Mathematics	3	3	0	2	0	1
	4	6	3	4	0	1
	5	7	1	6	0	2
	6	11	2	7	0	4
	7	8	1	5	0	3
	8	6	0	3	0	3
Science	4	1	0	1	0	0
	8	1	1	0	0	0
Social Studies	4	0	0	0	0	0
	8	1	0	1	0	0
	10	0	0	0	0	0
Total		53	8	38	0	14

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

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

Grade	Content	Item	Item Type	p-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	p-Value
3	ELA	25	MC	0.54	0.21	0.34		+	0.05		
4	ELA	12	MC	0.55	0.18	0.18		+	0.04		
4	ELA	31	MC	0.41	0.21	0.24		+	0.07		
5	ELA	23	MC	0.57	0.18	0.20		+	0.03		
5	ELA	29	MC	0.54	0.22	0.21		+	0.01		
6	ELA	4	MC	0.34	0.19	0.14		+	0.03		
6	ELA	16	MC	0.54	0.35	0.06		+	0.02		
8	ELA	22	MC	0.46	0.18	0.21		+	0.02		
8	ELA	36	MC	0.38	0.21	0.37		+	0.06		

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

Grade	Content	Item	Item Type	p-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	p-Value
3	Math	7	MC	0.40	0.33	0.15		+	0.04		
3	Math	12	TE	0.19	0.37	1.06					+
3	Math	16	MC	0.32	0.39	0.24		+	0.03		
4	Math	14	MC	0.43	0.27	0.17		+	0.01		
4	Math	15	MC	0.26	0.11	0.15	+	+	0.04		
4	Math	22	MC	0.28	0.16	0.17		+	0.13		
4	Math	31	MC	0.38	0.14	0.54	+				
4	Math	35	SA	0.16	0.46	0.37					+
4	Math	42	MC	0.31	0.09	0.21	+	+	0.09		
5	Math	3	MC	0.56	0.15	0.16		+	0.05		
5	Math	11	MC	0.32	0.22	0.21		+	0.04		
5	Math	20	MC	0.39	0.25	0.25		+	0.00		
5	Math	21	SA	0.17	0.43	0.41					+
5	Math	33	MC	0.39	0.25	0.15		+	0.11		
5	Math	41	MC	0.12	0.24	0.28		+	0.00		+
5	Math	45	MC	0.27	0.14	0.20	+	+	0.10		
6	Math	1	MC	0.38	0.34	0.07		+	0.06		
6	Math	9	SA	0.13	0.48	0.16					+
6	Math	14	SA	0.18	0.51	0.56					+
6	Math	25	MC	0.28	0.31	0.36		+	0.06		
6	Math	28	MC	0.37	0.13	0.29	+	+	0.06		
6	Math	31	MC	0.33	0.16	0.27		+	0.02		
6	Math	32	MC	0.29	0.06	0.33	+	+	0.02		
6	Math	37	SA	0.17	0.55	0.84					+
6	Math	38	MC	0.40	0.20	0.40		+	0.01		
6	Math	43	MC	0.31	0.22	0.34		+	0.05		
6	Math	44	TE	0.13	0.15	0.43					+
7	Math	8	TE	0.15	0.41	0.11					+
7	Math	9	MC	0.33	0.24	0.11		+	0.08		
7	Math	17	MC	0.22	0.13	0.28	+	+	0.01		
7	Math	20	MC	0.37	0.22	0.36		+	0.01		
7	Math	25	SA	0.17	0.49	0.94					+
7	Math	33	TE	0.10	0.50	0.50					+
7	Math	34	MC	0.55	0.30	0.50		+	0.03		
7	Math	40	MC	0.28	0.20	0.49		+	0.05		

Table 6-3 Mathematics Items Flagged for Classical Item Analysis Statistics (cont.)

Grade	Content	Item	Item Type	p-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	p-Value
8	Math	2	MC	0.21	0.15	0.11		+	0.02		
8	Math	5	SA	0.16	0.49	0.27					+
8	Math	16	MC	0.47	0.26	0.19		+	0.04		
8	Math	25	TE	0.18	0.47	0.50					+
8	Math	27	SA	0.17	0.46	0.87					+
8	Math	34	MC	0.37	0.22	0.52		+	0.02		

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

Grade	Content	Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
							Corr	Distractor	Omit	*p*-Value
4	Science	35	MC	0.35	0.23	0.18		+	0.03	
8	Science	24	TE	0.48	0.08	0.14	+			
8	Social Studies	37	MC	0.47	0.30	0.27		+	0.00	

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

Content	Grade						
	3	4	5	6	7	8	10
English Language Arts	99.64	99.68	99.70	99.63	99.65	99.66	
Mathematics	99.77	99.74	99.54	99.45	99.07	99.43	
Science		99.71				99.69	
Social Studies		99.86				99.66	99.47

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

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.67	0.32	0.07				
2	TE	0.68	0.38	0.59				
3	MC	0.54	0.40	0.14				
4	MC	0.42	0.29	0.16				
5	TE	0.52	0.45	0.78				
6	MC	0.53	0.38	0.17				
7	TE	0.54	0.42	0.31				
8	MC	0.61	0.44	0.18				
9	MC	0.69	0.44	0.17				
10	MC	0.58	0.39	0.17				
11	TE	0.30	0.39	0.92				
12	MC	0.56	0.47	0.22				
13	MC	0.43	0.36	0.17				
14	MC	0.69	0.54	0.21				
15	TE	0.62	0.55	0.30				
16	MC	0.70	0.36	0.10				
17	TE	0.62	0.54	0.13				
18	MC	0.54	0.37	0.12				
19	MS	0.71	0.54	0.11				
20	MC	0.44	0.33	0.21				
21	MC	0.52	0.22	0.16				
22	MC	0.62	0.41	0.29				
23	MC	0.40	0.28	0.30				
24	MC	0.51	0.31	0.42				
25	MC	0.54	0.21	0.34		+		
26	TE	0.64	0.64	0.34				
27	MC	0.54	0.32	0.34				
28	MS	0.54	0.49	0.32				
29	MC	0.55	0.44	0.30				
30	MC	0.70	0.48	0.34				
31	MC	0.27	0.27	0.34				
32	EBSR	0.43	0.46	0.20				
33	MC	0.42	0.33	0.34				
34	MC	0.43	0.37	0.37				
35	MC	0.59	0.53	0.37				
36	MC	0.46	0.32	0.38				
37	MC	0.41	0.29	0.32				
38	TE	0.50	0.50	0.36				

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TE	0.52	0.46	0.63				
2	TE	0.59	0.40	0.27				
3	MC	0.68	0.34	0.16				
4	TE	0.45	0.42	0.55				
5	MC	0.54	0.39	0.13				
6	MC	0.47	0.31	0.16				
7	MC	0.63	0.45	0.20				
8	TE	0.63	0.47	0.55				
9	MC	0.54	0.45	0.21				
10	MC	0.55	0.36	0.20				
11	MC	0.56	0.36	0.19				
12	MC	0.55	0.18	0.18		+		
13	TE	0.68	0.41	1.02				
14	MC	0.66	0.47	0.18				
15	TE	0.79	0.34	0.19				
16	MC	0.41	0.23	0.21				
17	MC	0.57	0.27	0.07				
18	MC	0.74	0.34	0.12				
19	EBSR	0.31	0.35	0.06				
20	MC	0.55	0.37	0.11				
21	MC	0.63	0.37	0.13				
22	EBSR	0.35	0.33	0.08				
23	MC	0.71	0.45	0.14				
24	TE	0.55	0.42	0.57				
25	MC	0.49	0.40	0.27				
26	MC	0.57	0.52	0.25				
27	MS	0.43	0.44	0.28				
28	EBSR	0.55	0.53	0.12				
29	MC	0.58	0.34	0.21				
30	MC	0.61	0.50	0.22				
31	MC	0.41	0.21	0.24		+		
32	MC	0.51	0.37	0.25				
33	MC	0.52	0.37	0.32				
34	TE	0.44	0.42	0.26				
35	MC	0.58	0.39	0.25				
36	MC	0.60	0.50	0.33				
37	MC	0.53	0.39	0.36				
38	MC	0.65	0.37	0.34				
39	MC	0.58	0.44	0.31				
40	TE	0.47	0.44	0.33				
41	MC	0.60	0.45	0.32				

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.78	0.41	0.04				
2	MC	0.66	0.36	0.09				
3	TE	0.57	0.44	0.09				
4	MC	0.42	0.24	0.09				
5	TE	0.67	0.36	0.10				
6	MC	0.80	0.33	0.09				
7	MC	0.75	0.27	0.10				
8	TE	0.49	0.49	0.40				
9	MC	0.66	0.34	0.10				
10	TE	0.71	0.54	0.28				
11	MC	0.61	0.44	0.16				
12	MC	0.53	0.36	0.14				
13	TE	0.72	0.41	0.13				
14	MC	0.57	0.32	0.15				
15	TE	0.58	0.35	0.30				
16	TE	0.60	0.43	0.04				
17	MC	0.61	0.34	0.11				
18	MC	0.77	0.44	0.10				
19	MC	0.61	0.41	0.09				
20	MC	0.61	0.37	0.10				
21	EBSR	0.58	0.53	0.05				
22	MC	0.69	0.48	0.10				
23	MC	0.57	0.18	0.20		+		
24	MC	0.67	0.42	0.24				
25	TE	0.84	0.49	0.44				
26	MC	0.53	0.32	0.18				
27	TE	0.67	0.56	0.48				
28	MC	0.56	0.43	0.23				
29	MC	0.54	0.22	0.21		+		
30	MC	0.52	0.26	0.23				
31	MS	0.58	0.50	0.27				
32	MC	0.61	0.40	0.25				
33	MC	0.55	0.48	0.28				
34	MC	0.65	0.54	0.22				
35	MC	0.55	0.45	0.23				
36	MC	0.42	0.28	0.28				
37	MC	0.60	0.45	0.33				
38	TE	0.62	0.53	0.28				
39	TE	0.64	0.29	0.63				
40	MC	0.43	0.45	0.27				
41	MC	0.46	0.34	0.30				
42	MC	0.58	0.44	0.30				

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

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.66	0.21	0.05				
2	MC	0.65	0.27	0.09				
3	MC	0.70	0.45	0.11				
4	MC	0.34	0.19	0.14		+		
5	MC	0.43	0.37	0.11				
6	MC	0.39	0.22	0.12				
7	TE	0.42	0.47	0.34				
8	TE	0.43	0.27	0.31				
9	TE	0.67	0.42	0.13				
10	MC	0.44	0.31	0.13				
11	TE	0.56	0.51	0.31				
12	TE	0.65	0.37	0.33				
13	TE	0.62	0.43	0.19				
14	MC	0.46	0.27	0.22				
15	MC	0.47	0.34	0.20				
16	MC	0.54	0.35	0.06		+		
17	MC	0.53	0.32	0.12				
18	TE	0.59	0.41	0.11				
19	EBSR	0.74	0.51	0.06				
20	MC	0.61	0.39	0.15				
21	MC	0.69	0.33	0.18				
22	MC	0.44	0.18	0.10				
23	TE	0.61	0.46	0.23				
24	MC	0.51	0.35	0.28				
25	TE	0.57	0.44	0.30				
26	MC	0.55	0.32	0.31				
27	MC	0.80	0.50	0.29				
28	TE	0.54	0.41	0.32				
29	MC	0.62	0.44	0.29				
30	TE	0.63	0.42	0.28				
31	TE	0.66	0.41	0.58				
32	MC	0.43	0.25	0.37				
33	MC	0.53	0.43	0.33				
34	TE	0.56	0.22	0.36				
35	EBSR	0.53	0.62	0.26				
36	TE	0.59	0.48	0.55				
37	MC	0.43	0.40	0.36				
38	MC	0.57	0.41	0.39				
39	MC	0.67	0.50	0.37				

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.68	0.43	0.07				
2	TE	0.51	0.39	0.09				
3	TE	0.70	0.47	0.18				
4	MC	0.71	0.38	0.10				
5	MS	0.53	0.24	0.11				
6	MC	0.39	0.28	0.13				
7	TE	0.64	0.31	0.19				
8	MC	0.60	0.46	0.13				
9	MC	0.70	0.34	0.10				
10	MC	0.59	0.38	0.11				
11	TE	0.44	0.20	0.16				
12	MC	0.45	0.29	0.20				
13	MC	0.69	0.37	0.14				
14	MC	0.63	0.42	0.15				
15	EBSR	0.75	0.43	0.03				
16	MS	0.75	0.50	0.15				
17	MC	0.58	0.34	0.10				
18	MC	0.57	0.29	0.11				
19	EBSR	0.38	0.39	0.06				
20	TE	0.73	0.37	0.34				
21	MC	0.60	0.53	0.19				
22	EBSR	0.51	0.39	0.08				
23	MC	0.51	0.34	0.25				
24	MC	0.50	0.43	0.28				
25	TE	0.54	0.47	1.68				
26	TE	0.65	0.56	0.32				
27	TE	0.53	0.35	0.29				
28	MC	0.40	0.29	0.37				
29	EBSR	0.46	0.44	0.19				
30	TE	0.34	0.44	0.61				
31	MC	0.71	0.55	0.27				
32	MC	0.43	0.30	0.34				
33	MC	0.54	0.38	0.37				
34	MC	0.49	0.46	0.41				
35	MS	0.58	0.49	0.38				
36	MS	0.73	0.46	0.34				
37	MC	0.57	0.44	0.35				

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

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.71	0.41	0.05				
2	MC	0.60	0.44	0.12				
3	MC	0.68	0.43	0.15				
4	MC	0.55	0.35	0.14				
5	MC	0.58	0.42	0.14				
6	MS	0.60	0.46	0.18				
7	MC	0.54	0.28	0.19				
8	TE	0.67	0.47	0.25				
9	MC	0.51	0.51	0.19				
10	MC	0.78	0.34	0.18				
11	TE	0.45	0.29	0.31				
12	MC	0.40	0.31	0.23				
13	TE	0.86	0.37	0.22				
14	MC	0.55	0.44	0.26				
15	MC	0.53	0.38	0.29				
16	MS	0.66	0.46	0.05				
17	EBSR	0.69	0.50	0.04				
18	MC	0.71	0.48	0.10				
19	MC	0.51	0.37	0.12				
20	EBSR	0.45	0.39	0.08				
21	MC	0.38	0.34	0.11				
22	MC	0.46	0.18	0.21		+		
23	MC	0.70	0.43	0.24				
24	MC	0.53	0.33	0.30				
25	MS	0.60	0.46	0.27				
26	MC	0.37	0.44	0.30				
27	MC	0.60	0.49	0.23				
28	MC	0.50	0.27	0.42				
29	MC	0.74	0.54	0.32				
30	MC	0.45	0.21	0.30				
31	EBSR	0.59	0.52	0.19				
32	MC	0.55	0.40	0.35				
33	EBSR	0.58	0.61	0.23				
34	MC	0.78	0.53	0.44				
35	MC	0.69	0.43	0.35				
36	MC	0.38	0.21	0.37		+		
37	MC	0.69	0.46	0.33				
38	MC	0.53	0.42	0.35				
39	MS	0.68	0.59	0.35				
40	MC	0.60	0.39	0.34				

Table 6-12 Item Analysis, Grade 3 Mathematics

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	TE	0.68	0.55	0.12				
2	SA	0.51	0.57	0.16				
3	TE	0.40	0.61	0.10				
4	MC	0.77	0.44	0.12				
5	MC	0.50	0.49	0.16				
6	SA	0.78	0.45	0.20				
7	MC	0.40	0.33	0.15		+		
8	SA	0.41	0.59	0.30				
9	MC	0.66	0.41	0.21				
10	SA	0.63	0.62	0.23				
11	MC	0.39	0.35	0.19				
12	TE	0.19	0.37	1.06				+
13	MC	0.57	0.44	0.17				
14	SA	0.60	0.57	0.30				
15	SA	0.32	0.52	0.35				
16	MC	0.32	0.39	0.24		+		
17	MC	0.77	0.51	0.19				
18	TE	0.83	0.43	1.45				
19	MC	0.71	0.49	0.17				
20	MC	0.57	0.43	0.26				
21	SA	0.57	0.57	0.30				
22	MC	0.46	0.36	0.10				
23	MC	0.75	0.51	0.11				
24	SA	0.58	0.61	0.13				
25	TE	0.59	0.35	0.13				
26	TE	0.41	0.58	0.53				
27	MC	0.58	0.45	0.25				
28	MC	0.39	0.35	0.24				
29	MC	0.56	0.37	0.16				
30	MC	0.40	0.23	0.17				
31	MC	0.43	0.35	0.25				
32	MC	0.41	0.37	0.22				
33	SA	0.47	0.59	0.28				
34	MC	0.66	0.47	0.18				
35	SA	0.60	0.65	0.24				

Table 6-12 Item Analysis, Grade 3 Mathematics (cont.)

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	TE	0.23	0.49	0.81				
37	MC	0.47	0.48	0.25				
38	MC	0.69	0.40	0.21				
39	SA	0.54	0.64	0.28				
40	MC	0.65	0.50	0.21				
41	TE	0.31	0.46	2.90				
42	MC	0.68	0.48	0.23				

Table 6-13 Item Analysis, Grade 4 Mathematics

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.67	0.43	0.07				
2	MC	0.44	0.28	0.14				
3	SA	0.41	0.42	0.11				
4	MC	0.82	0.42	0.11				
5	TE	0.39	0.54	0.39				
6	MC	0.50	0.51	0.13				
7	MC	0.34	0.50	0.08				
8	SA	0.35	0.62	0.81				
9	MC	0.69	0.50	0.14				
10	MC	0.85	0.41	0.17				
11	MC	0.28	0.33	0.17				
12	MC	0.34	0.31	0.19				
13	MC	0.49	0.60	0.14				
14	MC	0.43	0.27	0.17		+		
15	MC	0.26	0.11	0.15	+	+		
16	SA	0.33	0.52	0.82				
17	MC	0.81	0.45	0.16				
18	MC	0.42	0.48	0.23				
19	TE	0.33	0.59	0.20				
20	MC	0.45	0.38	0.20				
21	MC	0.88	0.32	0.16				
22	MC	0.28	0.16	0.17		+		
23	SA	0.21	0.54	0.28				
24	TE	0.69	0.37	0.51				
25	MC	0.84	0.36	0.14				
26	MC	0.48	0.49	0.16				
27	MC	0.29	0.57	0.15				
28	MC	0.46	0.41	0.12				
29	SA	0.32	0.54	0.32				
30	MC	0.30	0.40	0.15				
31	MC	0.38	0.14	0.54	+			
32	TE	0.20	0.48	0.96				
33	MC	0.50	0.59	0.16				
34	MC	0.39	0.43	0.17				
35	SA	0.16	0.46	0.37				+

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

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	MC	0.53	0.42	0.23				
37	MC	0.83	0.35	0.19				
38	MC	0.33	0.54	0.17				
39	TE	0.55	0.63	0.50				
40	SA	0.53	0.45	0.24				
41	MC	0.57	0.40	0.22				
42	MC	0.31	0.09	0.21	+	+		
43	SA	0.45	0.47	0.36				
44	MC	0.27	0.42	0.26				
45	SA	0.42	0.56	0.35				
46	MC	0.49	0.40	0.26				

Table 6-14 Item Analysis, Grade 5 Mathematics

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	SA	0.61	0.50	0.10				
2	MC	0.42	0.49	0.11				
3	MC	0.56	0.15	0.16		+		
4	MC	0.31	0.39	0.16				
5	MC	0.42	0.34	0.17				
6	TE	0.33	0.47	0.87				
7	MC	0.58	0.51	0.09				
8	MC	0.42	0.34	0.41				
9	TE	0.28	0.54	1.06				
10	SA	0.42	0.54	0.32				
11	MC	0.32	0.22	0.21		+		
12	SA	0.28	0.48	0.63				
13	MC	0.58	0.59	0.14				
14	SA	0.37	0.57	0.35				
15	MC	0.41	0.51	0.18				
16	SA	0.31	0.33	0.57				
17	MC	0.47	0.33	0.22				
18	SA	0.44	0.45	0.30				
19	MC	0.63	0.53	0.23				
20	MC	0.39	0.25	0.25		+		
21	SA	0.17	0.43	0.41				+
22	TE	0.51	0.52	0.39				
23	MC	0.40	0.49	0.33				
24	MC	0.51	0.30	0.08				
25	SA	0.40	0.52	0.29				
26	MC	0.37	0.36	0.13				
27	SA	0.24	0.45	0.24				
28	MC	0.47	0.42	0.13				
29	TE	0.41	0.55	0.45				
30	MC	0.56	0.31	0.09				
31	TE	0.49	0.58	0.45				
32	MC	0.73	0.36	0.16				
33	MC	0.39	0.25	0.15		+		
34	MC	0.47	0.40	0.36				
35	TE	0.41	0.59	0.34				

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

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	MC	0.31	0.44	0.18				
37	SA	0.41	0.54	0.58				
38	MC	0.44	0.40	0.14				
39	MC	0.55	0.44	0.39				
40	TE	0.27	0.44	0.31				
41	MC	0.12	0.24	0.28		+		+
42	MC	0.82	0.44	0.25				
43	MC	0.31	0.40	0.35				
44	SA	0.57	0.50	0.34				
45	MC	0.27	0.14	0.20	+	+		
46	SA	0.28	0.52	0.46				

Table 6-15 Item Analysis, Grade 6 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.38	0.34	0.07		+		
2	MC	0.35	0.49	0.08				
3	TE	0.48	0.54	0.62				
4	MC	0.64	0.49	0.09				
5	SA	0.43	0.45	0.37				
6	SA	0.41	0.53	0.48				
7	MC	0.24	0.39	0.08				
8	MC	0.73	0.45	0.12				
9	SA	0.13	0.48	0.16				+
10	MC	0.44	0.44	0.13				
11	TE	0.40	0.54	0.70				
12	MC	0.90	0.33	0.11				
13	SA	0.42	0.60	0.34				
14	SA	0.18	0.51	0.56				+
15	MC	0.36	0.59	0.20				
16	MC	0.45	0.46	0.17				
17	MC	0.34	0.48	0.08				
18	MC	0.53	0.51	0.18				
19	MC	0.70	0.37	0.15				
20	SA	0.61	0.51	0.23				
21	TE	0.54	0.53	0.28				
22	MC	0.47	0.42	0.24				
23	TE	0.67	0.32	1.36				
24	MC	0.52	0.22	0.48				
25	MC	0.28	0.31	0.36		+		
26	MC	0.41	0.35	0.41				
27	SA	0.27	0.45	0.56				
28	MC	0.37	0.13	0.29	+	+		
29	MC	0.55	0.30	0.32				
30	MC	0.57	0.33	0.26				
31	MC	0.33	0.16	0.27		+		
32	MC	0.29	0.06	0.33	+	+		
33	MC	0.62	0.44	0.25				
34	MC	0.59	0.47	0.41				
35	MC	0.41	0.47	0.40				

Table 6-15 Item Analysis, Grade 6 Mathematics (cont.)

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	TE	0.34	0.50	0.94				
37	SA	0.17	0.55	0.84				+
38	MC	0.40	0.20	0.40		+		
39	MC	0.33	0.38	0.40				
40	MC	0.31	0.52	0.43				
41	TE	0.33	0.55	0.82				
42	MC	0.38	0.24	0.34				
43	MC	0.31	0.22	0.34		+		
44	TE	0.13	0.15	0.43				+
45	MC	0.41	0.55	0.46				
46	MC	0.85	0.33	0.55				

Table 6-16 Item Analysis, Grade 7 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	SA	0.43	0.60	0.19				
2	MC	0.44	0.51	0.07				
3	MC	0.46	0.32	0.07				
4	SA	0.21	0.46	0.36				
5	MC	0.47	0.41	0.11				
6	MC	0.37	0.42	0.11				
7	MC	0.48	0.41	0.06				
8	TE	0.15	0.41	0.11				+
9	MC	0.33	0.24	0.11		+		
10	MC	0.49	0.39	0.09				
11	MC	0.64	0.20	0.12				
12	MC	0.73	0.23	0.06				
13	MC	0.40	0.38	0.19				
14	MC	0.46	0.40	0.13				
15	SA	0.28	0.41	0.78				
16	MC	0.26	0.47	0.17				
17	MC	0.22	0.13	0.28	+	+		
18	MC	0.39	0.19	0.17				
19	TE	0.33	0.60	1.10				
20	MC	0.37	0.22	0.36		+		
21	MC	0.61	0.27	0.44				
22	MC	0.25	0.19	0.42				
23	MC	0.27	0.19	0.47				
24	SA	0.36	0.51	0.69				
25	SA	0.17	0.49	0.94				+
26	MC	0.40	0.29	0.34				
27	MC	0.50	0.47	0.36				
28	MC	0.78	0.45	0.30				
29	MC	0.30	0.26	0.33				
30	MC	0.40	0.44	0.35				
31	MC	0.60	0.55	0.39				
32	TE	0.66	0.20	1.14				
33	TE	0.10	0.50	0.50				+
34	MC	0.55	0.30	0.50		+		
35	SA	0.35	0.65	1.12				

Table 6-16 Item Analysis, Grade 7 Mathematics (cont.)

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	TE	0.67	0.48	0.55				
37	MC	0.49	0.50	0.47				
38	TE	0.43	0.26	0.90				
39	MC	0.53	0.20	0.54				
40	MC	0.28	0.20	0.49		+		
41	TE	0.24	0.60	0.60				
42	TE	0.20	0.57	1.46				
43	MC	0.47	0.52	0.63				
44	MC	0.48	0.46	0.70				
45	MC	0.65	0.39	0.66				
46	SA	0.61	0.57	0.93				

Table 6-17 Item Analysis, Grade 8 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.59	0.47	0.05				
2	MC	0.21	0.15	0.11		+		
3	MC	0.52	0.36	0.11				
4	MC	0.44	0.20	0.11				
5	SA	0.16	0.49	0.27				+
6	MC	0.30	0.42	0.12				
7	MC	0.42	0.60	0.07				
8	MC	0.45	0.46	0.08				
9	TE	0.22	0.55	0.15				
10	MC	0.36	0.26	0.08				
11	TE	0.21	0.49	0.34				
12	MC	0.46	0.31	0.14				
13	SA	0.39	0.63	0.29				
14	MC	0.46	0.24	0.10				
15	TE	0.42	0.50	0.22				
16	MC	0.47	0.26	0.19		+		
17	SA	0.32	0.52	0.68				
18	MC	0.53	0.40	0.27				
19	MC	0.44	0.35	0.21				
20	MC	0.48	0.41	0.21				
21	SA	0.22	0.57	0.95				
22	MC	0.71	0.38	0.35				
23	TE	0.55	0.50	1.51				
24	MC	0.52	0.37	0.39				
25	TE	0.18	0.47	0.50				+
26	MC	0.25	0.21	0.27				
27	SA	0.17	0.46	0.87				+
28	MC	0.55	0.53	0.31				
29	MC	0.39	0.42	0.35				
30	MC	0.28	0.34	0.36				
31	MC	0.31	0.42	0.40				
32	TE	0.31	0.48	0.69				
33	MC	0.53	0.38	0.48				
34	MC	0.37	0.22	0.52		+		
35	TE	0.22	0.66	0.65				

Table 6-17 Item Analysis, Grade 8 Mathematics (cont.)

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	MC	0.49	0.47	0.45				
37	TE	0.32	0.29	1.15				
38	MC	0.74	0.40	0.43				
39	MC	0.41	0.38	0.50				
40	MC	0.20	0.27	0.51				
41	TE	0.35	0.37	0.64				
42	MC	0.58	0.48	0.47				
43	MC	0.66	0.43	0.58				
44	MC	0.53	0.32	0.57				
45	MC	0.48	0.37	0.54				
46	MC	0.70	0.40	0.57				

Table 6-18 Item Analysis, Grade 4 Science

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.52	0.33	0.13				
2	TE	0.75	0.51	0.13				
3	TE	0.33	0.32	0.12				
4	TE	0.43	0.42	0.32				
5	TE	0.52	0.38	0.60				
6	TE	0.60	0.51	0.25				
7	MC	0.36	0.29	0.19				
8	TE	0.58	0.42	0.27				
9	EBSR	0.41	0.44	0.14				
10	MC	0.66	0.38	0.14				
11	MC	0.50	0.40	0.16				
12	MC	0.53	0.34	0.21				
13	TE	0.69	0.40	0.43				
14	TE	0.63	0.45	0.71				
15	TE	0.39	0.58	0.37				
16	MC	0.61	0.34	0.11				
17	TE	0.84	0.40	0.12				
18	TE	0.68	0.40	0.25				
19	TE	0.50	0.22	0.45				
20	TE	0.88	0.32	0.15				
21	TE	0.52	0.40	0.91				
22	TE	0.53	0.57	0.30				
23	MC	0.57	0.41	0.14				
24	TE	0.38	0.43	0.17				
25	TE	0.54	0.51	0.53				
26	TE	0.76	0.42	0.08				
27	TE	0.63	0.20	0.16				
28	TE	0.34	0.29	0.12				
29	TE	0.45	0.46	0.21				
30	TE	0.64	0.46	0.26				
31	TE	0.63	0.31	0.41				
32	MC	0.57	0.35	0.21				
33	TE	0.74	0.41	0.14				
34	TE	0.90	0.36	0.16				
35	MC	0.35	0.23	0.18		+		

Table 6-18 Item Analysis, Grade 4 Science (cont.)

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	EBSR	0.43	0.26	0.13				
37	MC	0.51	0.40	0.30				
38	TE	0.78	0.47	0.22				
39	MS	0.21	0.38	0.24				
40	TE	0.20	0.35	0.29				

Table 6-19 Item Analysis, Grade 8 Science

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	TE	0.63	0.43	0.19				
2	TE	0.53	0.47	0.15				
3	TE	0.52	0.58	0.21				
4	TE	0.58	0.51	0.18				
5	TE	0.83	0.33	0.18				
6	MC	0.56	0.33	0.21				
7	TE	0.52	0.38	0.16				
8	MC	0.61	0.39	0.23				
9	TE	0.79	0.30	0.12				
10	TE	0.47	0.32	0.12				
11	TE	0.60	0.41	0.15				
12	MC	0.52	0.30	0.16				
13	TE	0.32	0.39	0.26				
14	TE	0.72	0.36	0.23				
15	TE	0.53	0.29	0.17				
16	TE	0.38	0.45	0.09				
17	TE	0.42	0.34	0.06				
18	TE	0.51	0.55	0.09				
19	TE	0.29	0.37	0.15				
20	TE	0.51	0.33	0.21				
21	TE	0.83	0.41	0.31				
22	TE	0.71	0.42	0.13				
23	TE	0.57	0.45	0.17				
24	TE	0.48	0.08	0.14	+			
25	TE	0.51	0.44	0.13				
26	TE	0.34	0.52	0.04				
27	TE	0.48	0.54	0.14				
28	TE	0.58	0.18	0.24				
29	TE	0.64	0.42	0.17				
30	MC	0.51	0.43	0.23				
31	EBSR	0.46	0.47	0.15				
32	TE	0.74	0.38	0.21				
33	TE	0.42	0.36	0.52				
34	TE	0.45	0.35	0.29				
35	TE	0.28	0.36	0.24				

Table 6-19 Item Analysis, Grade 8 Science (cont.)

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	MS	0.30	0.45	0.31				
37	TE	0.42	0.50	0.22				
38	MC	0.69	0.44	0.28				
39	TE	0.55	0.43	0.35				
40	MC	0.45	0.28	0.31				

Table 6-20 Item Analysis, Grade 4 Social Studies

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.77	0.52	0.06				
2	MC	0.81	0.44	0.11				
3	MC	0.79	0.36	0.12				
4	MC	0.76	0.47	0.12				
5	MC	0.72	0.43	0.21				
6	MC	0.43	0.28	0.22				
7	TE	0.69	0.37	0.18				
8	MC	0.51	0.34	0.19				
9	MC	0.76	0.39	0.23				
10	MC	0.65	0.46	0.18				
11	MC	0.51	0.19	0.15				
12	MC	0.78	0.45	0.42				
13	TE	0.42	0.34	0.36				
14	MC	0.37	0.39	0.19				
15	MC	0.65	0.37	0.17				
16	MC	0.58	0.33	0.20				
17	MC	0.65	0.47	0.16				
18	TE	0.70	0.36	0.31				
19	MC	0.72	0.47	0.17				
20	MC	0.63	0.33	0.09				
21	MC	0.70	0.45	0.12				
22	MC	0.64	0.21	0.12				
23	MC	0.81	0.55	0.14				
24	MC	0.52	0.41	0.17				
25	MC	0.63	0.52	0.17				
26	MC	0.56	0.29	0.12				
27	MC	0.61	0.35	0.40				
28	MC	0.60	0.44	0.16				
29	MC	0.50	0.37	0.16				
30	MC	0.73	0.46	0.16				
31	TE	0.49	0.43	0.16				
32	MC	0.64	0.50	0.13				
33	MC	0.72	0.51	0.42				
34	MC	0.69	0.51	0.19				
35	MC	0.67	0.44	0.15				
36	MC	0.74	0.54	0.16				
37	MC	0.63	0.48	0.19				
38	MC	0.69	0.47	0.14				

Table 6-21 Item Analysis, Grade 8 Social Studies

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.82	0.47	0.02				
2	MC	0.80	0.48	0.09				
3	MC	0.85	0.53	0.11				
4	MC	0.80	0.38	0.09				
5	MC	0.82	0.52	0.11				
6	MC	0.75	0.35	0.14				
7	MS	0.58	0.60	0.08				
8	MC	0.66	0.54	0.24				
9	MC	0.53	0.39	0.12				
10	TE	0.60	0.53	0.33				
11	MC	0.80	0.48	0.14				
12	MC	0.73	0.58	0.16				
13	MC	0.76	0.34	0.12				
14	MC	0.50	0.35	0.21				
15	MC	0.64	0.38	0.28				
16	MC	0.62	0.37	0.37				
17	MC	0.69	0.41	0.22				
18	MC	0.62	0.44	0.25				
19	MC	0.66	0.35	0.21				
20	MC	0.66	0.41	0.31				
21	MC	0.92	0.37	0.05				
22	MC	0.58	0.49	0.10				
23	TE	0.57	0.40	0.21				
24	MC	0.74	0.39	0.18				
25	MC	0.74	0.36	0.33				
26	MC	0.69	0.45	0.16				
27	MC	0.54	0.37	0.20				
28	MC	0.77	0.38	0.19				
29	MC	0.59	0.25	0.20				
30	TE	0.51	0.52	0.27				
31	MC	0.68	0.44	0.21				
32	TE	0.56	0.44	0.27				
33	MC	0.50	0.34	0.27				
34	MC	0.50	0.32	0.31				
35	MC	0.60	0.45	0.30				
36	MC	0.54	0.38	0.28				
37	MC	0.47	0.30	0.27		+		
38	MC	0.42	0.35	0.26				
39	MC	0.82	0.48	0.26				
40	MC	0.61	0.51	0.34				

Table 6-22 Item Analysis, Grade 10 Social Studies

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
1	MC	0.81	0.42	0.04				
2	MC	0.74	0.25	0.07				
3	MC	0.72	0.39	0.09				
4	MC	0.69	0.40	0.11				
5	MC	0.75	0.45	0.10				
6	MC	0.75	0.43	0.33				
7	MC	0.82	0.46	0.09				
8	MC	0.76	0.41	0.16				
9	MC	0.58	0.42	0.11				
10	MC	0.69	0.38	0.18				
11	MC	0.59	0.41	0.18				
12	TE	0.33	0.34	0.45				
13	MC	0.56	0.29	0.20				
14	MC	0.59	0.42	0.20				
15	MC	0.68	0.43	0.19				
16	MC	0.65	0.33	0.35				
17	MC	0.58	0.25	0.31				
18	MC	0.73	0.34	0.42				
19	MC	0.67	0.47	0.23				
20	MC	0.63	0.44	0.26				
21	MC	0.60	0.33	0.29				
22	MC	0.55	0.46	0.31				
23	MC	0.58	0.51	0.30				
24	TE	0.31	0.45	0.86				
25	MC	0.62	0.58	0.46				
26	MC	0.74	0.35	0.08				
27	MC	0.73	0.31	0.11				
28	MC	0.44	0.47	0.21				
29	MC	0.42	0.38	0.23				
30	MC	0.57	0.53	0.19				
31	MC	0.74	0.44	0.27				
32	MC	0.84	0.39	0.17				
33	MC	0.62	0.50	0.32				
34	TE	0.70	0.20	0.44				
35	MC	0.63	0.48	0.35				

Table 6-22 Item Analysis, Grade 10 Social Studies (cont.)

Item	Item Type	p-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	p-Value
36	MC	0.73	0.49	0.36				
37	MC	0.47	0.37	0.32				
38	MC	0.56	0.35	0.34				
39	MC	0.72	0.42	0.30				
40	MC	0.71	0.46	0.33				
41	MC	0.82	0.43	0.44				
42	MC	0.64	0.38	0.42				
43	MC	0.70	0.43	0.39				
44	MC	0.43	0.34	0.41				
45	TE	0.31	0.39	0.56				
46	MC	0.54	0.41	0.46				
47	MC	0.46	0.25	0.42				
48	MC	0.69	0.51	0.44				
49	MC	0.70	0.41	0.47				
50	MC	0.86	0.31	0.53				

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	52865	26.17	0.54	9.75	-0.01	-0.97	1	48	48	0.89	3.16
	4	52626	27.82	0.55	9.83	0.04	-0.96	0	51	51	0.89	3.22
	5	53924	30.88	0.61	9.95	-0.21	-0.89	2	51	51	0.90	3.17
	6	55421	28.95	0.56	9.47	-0.17	-0.85	0	51	51	0.88	3.23
	7	56193	29.10	0.57	9.63	-0.16	-0.86	1	51	51	0.88	3.28
	8	56619	30.13	0.59	10.16	-0.22	-0.93	0	51	51	0.90	3.21
Mathematics	3	52851	22.34	0.53	10.15	-0.05	-1.09	0	42	42	0.93	2.68
	4	52594	21.17	0.46	9.90	0.36	-0.84	0	46	46	0.92	2.81
	5	53861	19.32	0.42	10.13	0.41	-0.78	0	46	46	0.92	2.89
	6	55375	19.84	0.43	9.56	0.54	-0.58	0	46	46	0.91	2.87
	7	56161	19.16	0.42	9.25	0.53	-0.51	0	46	46	0.90	2.90
	8	56617	18.73	0.41	9.59	0.72	-0.30	0	46	46	0.91	2.88
Science	4	52330	21.98	0.55	8.20	-0.08	-0.91	0	40	40	0.89	2.71
	8	56338	21.22	0.53	8.49	0.04	-0.94	0	40	40	0.89	2.77
Social Studies	4	52342	24.39	0.64	8.10	-0.43	-0.82	0	38	38	0.90	2.60
	8	56278	26.15	0.66	8.54	-0.47	-0.75	0	40	40	0.91	2.63
	10	51132	31.58	0.64	10.28	-0.30	-0.86	0	50	50	0.91	3.01

Table 6-24 Scoring Table for English Language Arts Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	101	31	569	12
1	330	101	**32**	**573**	**12**
2	330	101	33	578	12
3	330	101	34	582	13
4	330	101	35	586	13
5	339	92	36	591	13
6	403	43	37	596	13
7	429	32	38	601	14
8	445	27	39	607	14
9	458	23	40	613	15
10	468	21	41	619	16
11	477	19	**42**	**627**	**17**
12	484	18	43	635	18
13	491	17	44	646	21
14	497	16	45	659	24
15	502	15	46	677	30
16	508	15	47	709	44
17	513	14	48	900	220
18	517	14			
19	**522**	**13**			
20	526	13			
21	530	13			
22	534	12			
23	538	12			
24	542	12			
25	546	12			
26	549	12			
27	553	12			
28	557	12			
29	561	12			
30	565	12			

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

Table 6-25 Scoring Table for English Language Arts Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	340	74	**31**	**592**	**13**
1	340	74	32	596	13
2	340	74	33	600	13
3	340	74	34	604	13
4	340	74	35	609	13
5	340	74	36	614	14
6	364	62	37	618	14
7	406	46	38	623	15
8	432	38	39	629	15
9	451	33	40	635	16
10	466	30	41	641	17
11	479	27	42	648	18
12	490	25	**43**	**655**	**19**
13	499	23	44	664	20
14	507	21	45	674	22
15	515	20	46	685	24
16	521	19	47	699	28
17	528	18	48	717	33
18	534	17	49	743	42
19	539	17	50	788	63
20	544	16	51	930	187
21	**549**	**15**			
22	554	15			
23	558	15			
24	563	14			
25	567	14			
26	571	14			
27	575	13			
28	580	13			
29	584	13			
30	588	13			

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

Table 6-26 Scoring Table for English Language Arts Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	85	31	592	12
1	350	85	32	596	12
2	350	85	33	600	12
3	350	85	34	604	12
4	350	85	35	608	12
5	350	85	**36**	**612**	**12**
6	350	85	37	616	13
7	387	64	38	621	13
8	428	46	39	626	13
9	452	36	40	631	13
10	470	30	41	636	14
11	484	26	42	642	15
12	494	23	43	648	16
13	504	20	44	656	17
14	512	19	45	664	18
15	519	17	**46**	**674**	**21**
16	525	17	47	686	24
17	531	16	48	703	29
18	536	15	49	727	38
19	541	15	50	772	61
20	546	14	51	940	211
21	551	14			
22	555	14			
23	560	13			
24	**564**	**13**			
25	568	13			
26	572	13			
27	576	13			
28	580	13			
29	584	12			
30	588	12			

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

Table 6-27 Scoring Table for English Language Arts Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	360	68	31	613	14
1	360	68	32	617	14
2	360	68	**33**	**622**	**14**
3	360	68	34	627	14
4	360	68	35	631	14
5	360	68	36	636	14
6	406	48	37	641	15
7	436	39	38	646	15
8	457	34	39	652	16
9	473	30	40	658	16
10	487	28	41	664	17
11	499	25	**42**	**671**	**17**
12	509	24	43	678	18
13	518	22	44	686	19
14	526	21	45	696	21
15	533	20	46	707	23
16	540	19	47	721	27
17	546	18	48	739	32
18	552	17	49	765	41
19	558	16	50	813	65
20	563	16	51	950	183
21	568	16			
22	**573**	**15**			
23	578	15			
24	582	15			
25	587	14			
26	591	14			
27	596	14			
28	600	14			
29	604	14			
30	609	14			

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

Table 6-28 Scoring Table for English Language Arts Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	370	73	31	634	15
1	370	73	**32**	**639**	**15**
2	370	73	33	644	15
3	370	73	34	649	15
4	370	73	35	654	15
5	379	68	36	659	16
6	424	50	37	664	16
7	453	41	38	670	16
8	474	35	39	676	16
9	490	31	40	682	17
10	503	29	41	689	18
11	515	26	42	696	18
12	525	25	**43**	**705**	**20**
13	533	23	44	714	21
14	542	22	45	725	24
15	549	21	46	738	27
16	556	20	47	755	32
17	562	19	48	779	41
18	569	18	49	815	57
19	574	18	50	884	94
20	580	17	51	960	146
21	**585**	**17**			
22	591	16			
23	596	16			
24	601	16			
25	605	16			
26	610	15			
27	615	15			
28	620	15			
29	624	15			
30	629	15			

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

Table 6-29 Scoring Table for English Language Arts Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	380	71	31	630	15
1	380	71	32	635	15
2	380	71	33	640	15
3	380	71	34	645	15
4	380	71	35	650	15
5	380	71	**36**	**655**	**15**
6	380	71	37	660	16
7	429	49	38	665	16
8	457	40	39	671	16
9	478	35	40	677	17
10	494	31	41	684	18
11	507	29	42	691	19
12	519	27	43	699	20
13	529	25	**44**	**708**	**21**
14	538	23	45	718	23
15	546	22	46	730	25
16	553	21	47	745	29
17	560	19	48	765	35
18	566	19	49	792	44
19	572	18	50	841	67
20	578	17	51	970	175
21	583	17			
22	588	16			
23	**593**	**16**			
24	598	15			
25	603	15			
26	608	15			
27	612	15			
28	617	15			
29	621	15			
30	626	14			

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

Table 6-30 Scoring Table for Mathematics Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	360	106	26	565	10
1	360	106	27	569	10
2	360	106	28	573	10
3	360	106	29	577	11
4	360	106	30	581	11
5	423	46	31	585	11
6	453	29	32	590	11
7	470	23	33	595	12
8	482	19	34	600	12
9	491	17	35	606	13
10	498	15	**36**	**612**	**14**
11	504	14	37	619	15
12	510	13	38	628	17
13	515	12	39	639	19
14	**520**	**12**	40	655	24
15	524	11	41	682	36
16	529	11	42	760	103
17	533	11			
18	536	10			
19	540	10			
20	544	10			
21	547	10			
22	551	10			
23	554	10			
24	558	10			
25	**562**	**10**			

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

Table 6-31 Scoring Table for Mathematics Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	405	102	26	599	10
1	405	102	27	602	10
2	405	102	28	606	10
3	405	102	29	609	10
4	405	102	30	613	10
5	405	102	31	616	10
6	442	65	32	620	10
7	478	33	33	623	10
8	496	25	34	627	10
9	509	21	35	631	10
10	519	18	**36**	**635**	**10**
11	527	17	37	639	11
12	535	16	38	644	11
13	**541**	**15**	39	649	12
14	548	14	40	655	13
15	553	13	41	662	14
16	558	13	42	670	15
17	563	12	43	680	18
18	568	12	44	694	22
19	572	11	45	717	31
20	576	11	46	800	101
21	580	11			
22	584	10			
23	**588**	**10**			
24	592	10			
25	595	10			

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

Table 6-32 Scoring Table for Mathematics Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	430	114	26	628	10
1	430	114	27	631	10
2	430	114	28	635	10
3	430	114	29	638	10
4	430	114	30	641	10
5	430	114	31	645	10
6	511	37	32	648	10
7	532	25	33	652	10
8	545	20	34	656	10
9	555	17	**35**	**660**	**10**
10	562	16	36	664	11
11	569	14	37	668	11
12	**575**	**13**	38	673	11
13	580	13	39	678	12
14	585	12	40	684	13
15	589	12	41	690	14
16	593	11	42	698	15
17	597	11	43	708	18
18	601	11	44	722	22
19	604	10	45	746	33
20	608	10	46	830	105
21	**611**	**10**			
22	615	10			
23	618	10			
24	621	10			
25	625	10			

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 Mathematics Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	440	105	26	638	10
1	440	105	27	642	10
2	440	105	28	645	10
3	440	105	29	649	10
4	440	105	30	653	10
5	440	105	31	656	10
6	440	105	32	660	10
7	499	48	33	664	10
8	525	29	34	668	11
9	541	22	35	672	11
10	552	19	36	677	11
11	562	18	37	682	12
12	570	17	38	687	12
13	578	16	**39**	**693**	**13**
14	**584**	**15**	40	699	13
15	590	15	41	706	14
16	596	14	42	715	16
17	601	13	43	726	19
18	606	13	44	743	25
19	611	12	45	778	46
20	615	12	46	870	134
21	619	11			
22	623	11			
23	**627**	**11**			
24	631	11			
25	635	10			

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 Mathematics Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	450	127	26	662	10
1	450	127	27	665	10
2	450	127	28	669	10
3	450	127	29	673	10
4	450	127	30	677	10
5	450	127	31	681	11
6	450	127	32	685	11
7	492	85	33	689	11
8	540	40	34	693	11
9	562	27	35	698	11
10	576	22	36	702	12
11	586	19	37	707	12
12	595	17	**38**	**713**	**13**
13	602	16	39	719	14
14	**609**	**14**	40	726	15
15	615	14	41	734	16
16	620	13	42	744	18
17	625	12	43	756	21
18	630	12	44	775	28
19	634	12	45	810	46
20	639	11	46	880	107
21	643	11			
22	**647**	**11**			
23	650	11			
24	654	10			
25	658	10			

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 Mathematics Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	470	121	26	678	10
1	470	121	27	682	10
2	470	121	28	685	10
3	470	121	29	689	10
4	470	121	30	692	10
5	470	121	31	696	10
6	470	121	32	700	10
7	534	57	33	703	10
8	567	34	34	707	10
9	585	26	35	711	11
10	597	22	36	716	11
11	607	19	**37**	**720**	**11**
12	616	17	38	725	11
13	**623**	**16**	39	730	12
14	629	15	40	736	13
15	634	14	41	743	14
16	639	13	42	752	16
17	644	12	43	762	18
18	648	12	44	778	24
19	653	11	45	806	38
20	657	11	46	890	113
21	660	11			
22	664	10			
23	**668**	**10**			
24	671	10			
25	675	10			

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 Science Grade 4

Raw Score	Scale Score	SEM
0	300	63
1	300	63
2	300	63
3	320	50
4	356	32
5	376	26
6	390	23
7	402	20
8	412	19
9	420	18
10	428	17
11	435	16
12	442	16
13	**448**	**15**
14	454	15
15	459	15
16	465	14
17	470	14
18	475	14
19	480	14
20	486	14
21	491	14
22	**496**	**14**
23	501	14
24	506	14
25	511	14
26	516	14
27	522	14
28	527	15
29	533	15
30	540	16
31	**547**	**17**
32	554	18
33	563	19
34	572	20
35	583	22
36	597	25
37	614	30
38	638	38
39	680	56
40	725	86

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 Science Grade 8

Raw Score	Scale Score	SEM
0	480	79
1	480	79
2	480	79
3	540	42
4	568	31
5	587	26
6	600	23
7	612	21
8	621	19
9	630	18
10	637	17
11	644	16
12	650	16
13	**656**	**15**
14	662	14
15	667	14
16	672	14
17	677	14
18	682	13
19	686	13
20	691	13
21	**696**	**13**
22	700	13
23	705	13
24	709	13
25	714	13
26	719	14
27	724	14
28	729	14
29	735	15
30	**741**	**15**
31	747	16
32	754	17
33	762	18
34	771	19
35	781	21
36	793	24
37	810	29
38	833	36
39	876	58
40	945	115

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 Social Studies Grade 4

Raw Score	Scale Score	SEM
0	200	108
1	200	108
2	200	108
3	200	108
4	200	108
5	200	108
6	200	108
7	213	95
8	261	51
9	284	37
10	299	29
11	310	25
12	319	21
13	327	19
14	334	17
15	340	16
16	346	15
17	352	14
18	357	14
19	362	13
20	**366**	**13**
21	371	13
22	376	13
23	381	13
24	385	13
25	390	13
26	395	13
27	**401**	**13**
28	406	14
29	412	14
30	419	15
31	426	16
32	434	17
33	**443**	**19**
34	454	21
35	469	25
36	489	31
37	525	48
38	570	79

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 Social Studies Grade 8

Raw Score	Scale Score	SEM
0	420	87
1	420	87
2	420	87
3	420	87
4	420	87
5	420	87
6	420	87
7	445	62
8	475	36
9	492	27
10	504	23
11	513	20
12	521	18
13	528	17
14	535	15
15	540	15
16	546	14
17	551	13
18	555	13
19	560	12
20	**564**	**12**
21	568	12
22	573	12
23	577	11
24	581	11
25	585	11
26	590	12
27	594	12
28	**599**	**12**
29	603	12
30	608	13
31	614	13
32	620	14
33	626	15
34	633	16
35	**642**	**17**
36	651	19
37	664	22
38	680	26
39	707	37
40	780	95

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 Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	490	119	26	674	13
1	490	119	27	678	12
2	490	119	28	682	12
3	490	119	29	686	12
4	490	119	30	690	12
5	490	119	31	693	12
6	490	119	32	697	12
7	490	119	33	701	12
8	490	119	**34**	**705**	**12**
9	490	119	35	709	12
10	532	77	36	713	12
11	563	49	37	718	12
12	581	37	38	722	12
13	595	30	39	727	13
14	606	26	40	732	13
15	615	23	41	738	14
16	623	21	**42**	**744**	**14**
17	630	19	43	750	15
18	636	18	44	758	16
19	642	17	45	766	18
20	647	16	46	777	20
21	652	15	47	790	24
22	657	15	48	809	30
23	662	14	49	841	44
24	666	13	50	890	79
25	**670**	**13**			

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 Numbers and Percentages of Students at LOSS and HOSS

Content	Grade	LOSS	N	Percentage	HOSS	N	Percentage
ELA	3	330	26	0.05	900	20	0.04
	4	340	10	0.02	930	9	0.02
	5	350	9	0.02	940	22	0.04
	6	360	10	0.02	950	3	0.01
	7	370	6	0.01	960	6	0.01
	8	380	15	0.03	970	22	0.04
Mathematics	3	360	1088	2.06	760	163	0.31
	4	405	1246	2.37	800	40	0.08
	5	430	2412	4.47	830	27	0.05
	6	440	2171	3.91	870	39	0.07
	7	450	2143	3.81	880	26	0.05
	8	470	1600	2.82	890	69	0.12
Science	4	300	13	0.02	725	35	0.07
	8	480	12	0.02	945	53	0.09
Social Studies	4	200	667	1.27	570	259	0.49
	8	420	292	0.52	780	516	0.91
	10	490	362	0.70	890	154	0.30

Part 7: Standard Setting

In this part of the report, the Wisconsin Forward Exam standard settings that were conducted for ELA, Mathematics, Science, and Social Studies in Spring 2016 and the standard setting conducted for Science in Spring 2019 are briefly described. 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.

7.2 Standard Setting Methodology and Process

Prior to the standard setting workshops in Spring 2016 and 2019, 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. Specifically, the committee performed the following tasks:

1. Participants discussed the state content standards for science and the draft performance level descriptors (PLDs) for their test. The PLDs described the content-based expectations for students in each performance level. Participants refined the PLDs based on their study of the content standards.

2. Participants each examined an ordered item booklet (OIB) which presented test items in order of difficulty. Difficulty was calculated from Wisconsin students' performance.
3. For each item, participants considered whether a student just entering each performance level (e.g., a just *Proficient* student) would have command of the knowledge and skills measured by the item, defined as having at least a 50% chance of answering the item correctly. Participants indicated the set of items in the OIB that measured the content expected of students entering each performance level; they represented these judgments with bookmarks.
4. Participants discussed their bookmarks in three rounds of discussions and decisions. After each round, participants worked individually to revise their bookmark placements.
5. After the second and third rounds, participants examined the impact data for both grades. After the second round, participants also reviewed the impact data associated with their recommended cut scores, as well as the impact data for the Wisconsin 2018 Forward Exam Science assessments and 2015 NAEP Science. The NAEP impact data served as benchmarks in the OIB and were shown to participants to provide contextual information for consideration. Participants were given instructions on how to use these OIB benchmarks as points of reference as they considered their Round 3 judgments, and they were asked to consider how similar or different their Round 2 bookmarks were from the OIB benchmarks.
6. After the second and third rounds, participants reviewed the PLDs. Participants refined them to reflect the content-based expectations for students in each performance level.
7. Participants' cut score recommendations were recorded in terms of scale score. Each group's recommendation was the median of participants' recommendations.

After Round 3 of the Bookmark Procedure, participants reviewed their recommendations and associated impact data. Educators expressed satisfaction in the content-based judgments they made during the process. However, participants also voiced an expectation that the percentages of students classified in each performance level would be more consistent across grades 4 and 8.

To promote consistency in the performance standards across grades and testing programs, the Round 3 cut scores for grade 8 *Proficient* and *Advanced* were adjusted using the CSEM. The CSEM quantifies the amount of statistical error associated with any point on the test scale. Participants examined the adjusted cut scores and considered their reasonableness. Participants indicated the CSEM-adjusted cut scores were consistent with their content-based expectations from the Bookmark Procedure as well as their expectations for the impact data across grades. The committee made the CSEM-adjusted cut scores (shown in Table 7-2) their final recommendations for the Wisconsin Science assessments for grades 4 and 8. The cut scores recommended by the committee were approved by the State Superintendent of Public Instruction on June 5, 2019.

The process of both 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 and 2019 standard settings. The brief policy performance level descriptors, shown in Table 7-1, described 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. (For detailed standards-based PLDs, refer to the *Wisconsin Standard Setting 2019 Final Technical Report*.) At the most recent standard setting for Science, 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, Mathematics, and Social Studies, established in Spring 2016, remained unchanged for the 2021 assessments. New cut scores, reflecting Wisconsin student performance on the new Science assessments, were established for Science after the Spring 2019 test administration, and these cut scores were used for student classification into performance levels in Spring 2021.

7.5 Summary

Part 7 presented a brief overview of the standard setting process used for establishing the Wisconsin Forward Exam cut scores for all content areas after the Spring 2016 test administration and for Science after the Spring 2019 test administration. Both standard setting workshops are described in detail in their respective technical reports: *Wisconsin Standard Setting 2016 Final Technical Report* and *Wisconsin Standard Setting 2019 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 **adequate** understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.
Advanced	Student demonstrates **thorough** 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
ELA	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	363	396	436
	8	563	599	640
	10	670	703	741

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 2021 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 2021 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 (α) 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, σ_X^2 = the total score variance, and σ_i^2 = the variance of item i (Crocker & Algina, 1986). SEM is defined as

$$\text{SEM} = SD\sqrt{1 - reliability},$$

where SD represents the standard deviation of the raw score distribution and *reliability* 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 0.90 to 0.93 across grades for Mathematics. The reliability coefficient was 0.89 for Science grades 4 and 8. The reliability coefficient was 0.90 for Social Studies grade 4 and 0.91 for Social Studies grades 8 and 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, 6, and 7 and the Science assessments 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, 6, and 7 would need to be increased by approximately 3, 4, 10, 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 3 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 for male and female 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 or Asian students and the lowest reliability was observed for African American students in grades 3, 4, 5, 6, and 8 and for African American students and American Indian students in grade 7.

The differences between test reliability coefficients for economically disadvantaged and not economically disadvantaged students were within 0.02 of one another for all grades and content areas, except for Mathematics grades 4 and 6 through 8 where the differences ranged from 0.03 to 0.05. 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 of one another for all grades and content areas, except for Mathematics grades 6 through 8, where the differences were at or greater than 0.07. 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 were lower than for other subgroups for most grades in ELA, Science, and

Social Studies. The reliability coefficients for students using testing accommodations in Mathematics were the lowest of all subgroups. 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, among other factors, by the variability of students' scores. The higher the variability of scores, the higher the reliability coefficient will tend to be. While the patterns of differences in test reliability for different subgroups are similar to the patterns observed in Spring 2019 (for the same or very similar test forms), the magnitude of these differences increased for African American students when compared to White students, for students with disabilities compared to students without disabilities, for economically disadvantaged compared to not economically disadvantaged students, for limited English proficiency students compared to English proficient students, and for students using testing accommodations compared to students not using testing accommodations. These differences in test reliability were most prominent in Mathematics grades 6 through 8. It is possible that the circumstances related to the COVID-19 pandemic, including disrupted learning and nonstandard instruction delivery, negatively affected the test score stability for the historically disadvantaged student groups.

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 within one-half of a score point. The SEMs for ELA assessments and Social Studies grade 10 assessment 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 and the Social Studies grade 10 test 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 6-24 through 6-40).

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 19 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 12 per standard for Science, and from 6 to 13 per standard for Social Studies. A lower level of reliability per content standard or domain is therefore expected. 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 or domain ranged from 0.36 (for the Reading–Vocabulary Use standard in grade 5 and for the Writing/Language–Text Types and Purposes reporting category in grade 7) to 0.83 (for the Reading domain in grades 4, 5, and 8).
- For Mathematics, reliability indices by content standard ranged from 0.49 (for the Geometry standard in grade 7) to 0.82 (for the Number and Operations–Fractions standard in grade 4).
- For Science, reliability indices by content standard ranged from 0.56 (for the Earth and Space Science standard in grade 4) to 0.74 (for the Life Science standard in grade 8).
- For Social Studies, reliability indices by content standard ranged from 0.51 (for the Political Science and Citizenship standard in grade 4) to 0.77 (for the History standard in grade 10).

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, as a sum of item information function 2, 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 6-24 through 6-40) 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 H-1 through H-17 of Appendix H for all grades and content areas. As shown in Appendix H, 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 H-1 through H-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

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

Landis and Koch (1977) suggest that values of kappa equal to or greater than 0.75 indicate "excellent agreement," values between 0.40 and 0.74 represent "good agreement" beyond chance, and values below 0.40 denote "poor agreement."

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, θ_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 θ_j.

$$\theta_j \nearrow (1,1,0,0, \ldots: \text{Item response from the first administration, or Form 1}) \longrightarrow \hat{\theta}_{j1}$$

$$\searrow (0,1,1,0, \ldots: \text{Item response from the second administration, or Form 2}) \longrightarrow \hat{\theta}_{j2}$$
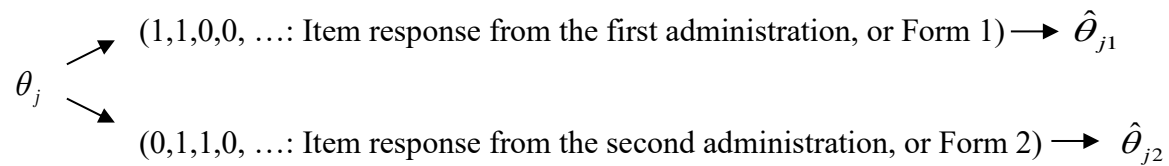
If two parallel (or alternative) forms (e.g., Form 1 and Form 2) are available, the two response patterns can be generated based on the item parameters from the two forms.

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 (C_1)

	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. So, 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 to *Below Basic* in both the first and second administrations. The 0.05 value represents the probability of belonging to *Basic* in the first administration and *Below Basic* in the second administration. The probability of belonging to *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 0.63, and classification accuracy is 0.82. 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 0.75, 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 0.73 for all tests. According to Landis and Koch's criteria for k (presented previously in this report in the discussion of classification consistency), all tests, with the exception of ELA grades 6 and 8 showed excellent agreement based on the cut for the *Proficient* performance level. The ELA grades 6 and 8 tests showed good agreement at the *Proficient* cut.

In addition, the indices for classification consistency and classification accuracy were computed for the subgroups of students. These data are presented in Appendix I. As seen in

Tables I-1 through I-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 0.86 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 0.88 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 0.86 and the classification accuracy was greater than or equal to 0.88 for all subgroups across all grades. The kappa values were greater than or equal to 0.55 for all subgroups in ELA, greater than or equal to 0.66 for all subgroups in Mathematics, greater than or equal to 0.64 for all subgroups in Science, and greater than or equal to 0.62 for all subgroups in Social Studies. The lowest kappa values were observed for the limited English proficiency subgroups in all content areas and for students using testing accommodations in Mathematics. 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 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 performance level classifications, and the SEM around the *Proficient* cut score was low in every grade and content area.

Table 8-1 Reliability for Total Group and Subgroups, Using Cronbach's Alpha

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Students with Accommodations	Students without Accommodations
English Language Arts	3	0.89	0.90	0.89	0.88	0.85	0.87	0.90	0.87	0.89	0.89	0.84	0.87	0.89	0.87	0.89	0.90	0.89
	4	0.89	0.89	0.89	0.88	0.84	0.87	0.89	0.86	0.89	0.89	0.82	0.86	0.89	0.87	0.89	0.87	0.89
	5	0.90	0.90	0.90	0.89	0.87	0.88	0.90	0.86	0.90	0.90	0.82	0.87	0.89	0.88	0.89	0.91	0.90
	6	0.88	0.88	0.89	0.88	0.85	0.87	0.88	0.86	0.88	0.88	0.78	0.85	0.87	0.87	0.87	0.88	0.88
	7	0.88	0.88	0.89	0.87	0.86	0.87	0.89	0.86	0.89	0.88	0.78	0.84	0.87	0.87	0.87	0.89	0.88
	8	0.90	0.90	0.90	0.89	0.88	0.89	0.90	0.88	0.90	0.90	0.81	0.86	0.89	0.89	0.89	0.91	0.90
Mathematics	3	0.93	0.93	0.93	0.92	0.89	0.91	0.94	0.90	0.93	0.93	0.90	0.92	0.93	0.92	0.92	0.86	0.93
	4	0.92	0.92	0.92	0.91	0.83	0.89	0.93	0.89	0.91	0.92	0.85	0.90	0.92	0.89	0.92	0.75	0.92
	5	0.92	0.91	0.92	0.91	0.85	0.89	0.93	0.87	0.92	0.92	0.83	0.89	0.92	0.89	0.91	0.77	0.92
	6	0.91	0.90	0.92	0.91	0.79	0.86	0.93	0.85	0.90	0.91	0.72	0.84	0.91	0.87	0.91	0.62	0.91
	7	0.90	0.89	0.91	0.90	0.81	0.86	0.93	0.81	0.90	0.90	0.73	0.83	0.90	0.86	0.90	0.59	0.90
	8	0.91	0.90	0.92	0.91	0.80	0.87	0.93	0.86	0.91	0.91	0.71	0.80	0.91	0.86	0.91	0.63	0.91
Science	4	0.89	0.89	0.89	0.88	0.83	0.86	0.89	0.86	0.88	0.89	0.82	0.88	0.88	0.87	0.88	-	0.89
	8	0.89	0.88	0.90	0.89	0.84	0.87	0.89	0.87	0.89	0.89	0.76	0.87	0.89	0.88	0.89	-	0.89
Social Studies	4	0.90	0.89	0.90	0.88	0.86	0.88	0.89	0.87	0.90	0.90	0.85	0.89	0.89	0.88	0.88	-	0.90
	8	0.91	0.90	0.91	0.90	0.89	0.89	0.90	0.88	0.90	0.90	0.83	0.88	0.90	0.90	0.90	-	0.91
	10	0.91	0.91	0.92	0.91	0.89	0.90	0.92	0.89	0.92	0.91	0.79	0.89	0.91	0.90	0.91	-	0.91

Note: 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

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Students with Accommodations	Students without Accommodations
English Language Arts	3	3.16	3.14	3.18	3.13	3.17	3.20	3.14	3.19	3.18	3.15	3.19	3.20	3.15	3.21	3.11	3.15	3.16
	4	3.22	3.20	3.23	3.20	3.24	3.25	3.22	3.24	3.23	3.21	3.26	3.22	3.21	3.26	3.18	3.24	3.22
	5	3.17	3.14	3.20	3.12	3.30	3.27	3.18	3.29	3.20	3.16	3.33	3.28	3.15	3.28	3.09	3.22	3.17
	6	3.23	3.22	3.24	3.19	3.30	3.30	3.21	3.28	3.26	3.23	3.30	3.26	3.22	3.30	3.18	3.33	3.23
	7	3.28	3.27	3.29	3.25	3.36	3.36	3.27	3.34	3.31	3.28	3.35	3.29	3.28	3.35	3.23	3.26	3.28
	8	3.21	3.18	3.24	3.17	3.31	3.29	3.18	3.27	3.25	3.20	3.34	3.28	3.19	3.30	3.15	3.40	3.21
Mathematics	3	2.68	2.70	2.66	2.67	2.62	2.71	2.66	2.70	2.70	2.68	2.70	2.68	2.68	2.72	2.65	2.66	2.68
	4	2.81	2.81	2.80	2.83	2.65	2.76	2.76	2.71	2.79	2.81	2.73	2.74	2.81	2.77	2.82	2.65	2.81
	5	2.89	2.89	2.88	2.92	2.63	2.80	2.86	2.76	2.85	2.90	2.70	2.69	2.91	2.81	2.92	2.54	2.90
	6	2.87	2.87	2.86	2.88	2.77	2.82	2.83	2.79	2.84	2.87	2.74	2.75	2.87	2.84	2.87	2.66	2.87
	7	2.90	2.89	2.89	2.92	2.70	2.83	2.84	2.80	2.86	2.90	2.70	2.70	2.91	2.83	2.92	2.59	2.90
	8	2.88	2.89	2.87	2.90	2.73	2.84	2.85	2.79	2.85	2.89	2.76	2.72	2.90	2.84	2.90	2.64	2.89
Science	4	2.71	2.72	2.71	2.69	2.74	2.76	2.72	2.76	2.75	2.71	2.76	2.73	2.71	2.77	2.68	-	2.72
	8	2.77	2.79	2.74	2.76	2.73	2.80	2.78	2.78	2.78	2.77	2.77	2.70	2.77	2.79	2.75	-	2.77
Social Studies	4	2.60	2.60	2.60	2.53	2.79	2.74	2.65	2.77	2.65	2.58	2.80	2.76	2.58	2.74	2.50	-	2.60
	8	2.63	2.63	2.62	2.56	2.82	2.77	2.61	2.79	2.68	2.61	2.89	2.82	2.60	2.77	2.54	-	2.63
	10	3.01	3.01	3.00	2.95	3.27	3.16	2.97	3.19	3.05	3.00	3.29	3.20	2.98	3.16	2.94	-	3.01

Note: 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 Standard and Domain

English Language Arts

Grade	Alpha per Content Standard and Domain									
	A	B	C	D	E	F	Listening	Reading	Writing	Total Test
3	0.73	0.37	0.41	0.45	0.52	0.67	0.61	0.79	0.78	0.89
4	0.66	0.55	0.64	0.58	0.57	0.46	0.48	0.83	0.77	0.89
5	0.69	0.70	0.36	0.46	0.57	0.51	0.61	0.83	0.74	0.90
6	0.66	0.48	0.57	0.48	0.40	0.48	0.57	0.81	0.71	0.88
7	0.67	0.59	0.57	0.36	0.50	0.47	0.53	0.82	0.70	0.88
8	0.71	0.56	0.51	0.60	0.58	0.38	0.58	0.83	0.76	0.90

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										
	A	B	C	D	E	F	G	H	I	J	Total Test
3	0.73	0.78	0.71	0.74	0.71						0.93
4	0.55	0.75	0.82	0.69	0.69						0.92
5	0.72	0.73	0.69	0.70	0.71						0.92
6					0.60	0.68	0.78	0.74	0.58		0.91
7					0.49	0.72	0.69	0.66	0.71		0.90
8					0.70		0.73	0.70	0.55	0.72	0.91

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				
	A	B	C	D	Total Test
4	0.71	0.70	0.56	0.70	0.89
8	0.74	0.68	0.60	0.68	0.89

Science standards: A=Life Science; B=Physical Science; C=Earth and Space Science; D=Engineering.

Social Studies

Grade	Alpha per Content Standard					
	A	B	C	D	E	Total Test
4	0.65	0.73	0.51	0.58	0.70	0.90
8	0.74	0.76	0.55	0.59	0.54	0.91
10	0.64	0.77	0.71	0.64	0.61	0.91

Social Studies standards: A=Geography; B=History; C=Political Science and Citizenship; D=Economics; E=The Behavioral Sciences

Table 8-4 Standard Error of Measurement per Content Standard and Domain

English Language Arts

Grade	SEM per Content Standard and Domain									
	A	B	C	D	E	F	Listening	Reading	Writing	Total Test
3	1.60	1.14	0.91	1.16	1.10	1.13	1.11	2.18	1.97	3.16
4	1.53	1.10	1.05	1.08	1.07	1.10	1.38	2.17	1.88	3.22
5	1.39	1.21	1.06	1.05	1.11	1.15	1.26	2.13	1.94	3.17
6	1.54	1.18	0.94	1.10	1.18	1.14	1.28	2.17	1.98	3.23
7	1.31	1.58	0.89	1.13	1.16	1.02	1.37	2.24	1.92	3.28
8	1.49	1.26	1.04	1.04	1.16	1.00	1.32	2.21	1.86	3.21

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										
	A	B	C	D	E	F	G	H	I	J	Total Test
3	1.23	1.12	1.15	1.34	1.11						2.68
4	1.36	1.24	1.23	1.29	1.07						2.81
5	1.23	1.28	1.26	1.35	1.28						2.89
6					1.12	1.00	1.29	1.43	1.44		2.87
7					1.34	1.18	1.08	1.36	1.43		2.90
8					1.36		1.14	1.35	1.23	1.29	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				
	A	B	C	D	Total Test
4	1.43	1.46	1.28	1.23	2.71
8	1.46	1.37	1.41	1.26	2.77

Science standards: A=Life Science; B=Physical Science; C=Earth and Space Science; D=Engineering

Social Studies

Grade	SEM per Content Standard				
	A	B	C	D	E	Total Test
4	1.32	1.19	1.09	1.04	1.12	2.60
8	1.26	1.40	1.08	1.07	1.01	2.63
10	1.36	1.50	1.33	1.19	1.30	3.01

Social Studies standards: A=Geography; B=History; C=Political Science and Citizenship; D=Economics; E=The 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	0.05	0.26	0.06	0.00	0.37
Proficient	0.00	0.05	0.21	0.02	0.29
Advanced	0.00	0.00	0.02	0.04	0.06
Sum	0.28	0.37	0.29	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	0.89	0.95	0.74
Probability of Chance	0.60	0.54	0.88	0.30
Kappa (*k*)	0.74	0.75	0.61	0.63
Classification Accuracy	0.93	0.92	0.97	0.82

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	0.06	0.00	0.00	0.28
Basic	0.05	0.22	0.05	0.00	0.32
Proficient	0.00	0.06	0.23	0.03	0.32
Advanced	0.00	0.00	0.03	0.06	0.09
Sum	0.27	0.33	0.31	0.09	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	0.94	0.72
Probability of Chance	0.60	0.52	0.84	0.29
Kappa (*k*)	0.73	0.77	0.61	0.61
Classification Accuracy	0.93	0.92	0.96	0.81

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	0.04	0.24	0.05	0.00	0.33
Proficient	0.00	0.05	0.24	0.03	0.32
Advanced	0.00	0.00	0.03	0.04	0.07
Sum	0.27	0.34	0.32	0.07	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	0.95	0.75
Probability of Chance	0.60	0.53	0.87	0.30
Kappa (k)	0.77	0.77	0.57	0.64
Classification Accuracy	0.93	0.92	0.96	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	0.20	0.05	0.00	0.00	0.25
Basic	0.04	0.26	0.07	0.00	0.37
Proficient	0.00	0.06	0.20	0.03	0.29
Advanced	0.00	0.00	0.04	0.06	0.10
Sum	0.25	0.36	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	0.87	0.93	0.72
Probability of Chance	0.62	0.53	0.83	0.29
Kappa (k)	0.76	0.73	0.60	0.60
Classification Accuracy	0.93	0.91	0.95	0.79

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	0.19	0.05	0.00	0.00	0.24
Basic	0.04	0.22	0.06	0.00	0.32
Proficient	0.00	0.06	0.23	0.04	0.33
Advanced	0.00	0.00	0.03	0.07	0.10
Sum	0.23	0.33	0.33	0.11	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	0.93	0.71
Probability of Chance	0.64	0.51	0.81	0.28
Kappa (k)	0.74	0.75	0.63	0.60
Classification Accuracy	0.93	0.91	0.95	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	0.22	0.05	0.00	0.00	0.27
Basic	0.04	0.26	0.06	0.00	0.36
Proficient	0.00	0.06	0.18	0.03	0.27
Advanced	0.00	0.00	0.03	0.06	0.10
Sum	0.26	0.37	0.27	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	0.93	0.72
Probability of Chance	0.61	0.54	0.83	0.29
Kappa (k)	0.76	0.74	0.61	0.61
Classification Accuracy	0.94	0.92	0.95	0.80

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	0.21	0.03	0.00	0.00	0.24
Basic	0.03	0.23	0.04	0.00	0.31
Proficient	0.00	0.05	0.26	0.03	0.33
Advanced	0.00	0.00	0.03	0.09	0.12
Sum	0.24	0.31	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)	0.93	0.91	0.94	0.78
Probability of Chance	0.63	0.51	0.79	0.28
Kappa (k)	0.82	0.82	0.70	0.70
Classification Accuracy	0.95	0.93	0.95	0.84

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	0.18	0.05	0.00	0.00	0.23
Basic	0.05	0.27	0.04	0.00	0.36
Proficient	0.00	0.05	0.23	0.02	0.31
Advanced	0.00	0.00	0.03	0.08	0.11
Sum	0.23	0.36	0.30	0.11	1.00

Indices for Classification Consistency and Classification Accuracy

Indices	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.91	0.95	0.77
Probability of Chance	0.65	0.52	0.81	0.29
Kappa (k)	0.74	0.82	0.74	0.68
Classification Accuracy	0.94	0.93	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	0.24	0.06	0.00	0.00	0.30
Basic	0.05	0.19	0.04	0.00	0.28
Proficient	0.00	0.05	0.24	0.03	0.31
Advanced	0.00	0.00	0.03	0.08	0.10
Sum	0.29	0.29	0.31	0.11	1.00

Indices for Classification Consistency and Classification Accuracy

Indices	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.89	0.91	0.95	0.75
Probability of Chance	0.58	0.51	0.81	0.28
Kappa (k)	0.74	0.82	0.71	0.65
Classification Accuracy	0.92	0.94	0.96	0.82

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	0.27	0.05	0.00	0.00	0.32
Basic	0.05	0.21	0.05	0.00	0.31
Proficient	0.00	0.05	0.25	0.02	0.32
Advanced	0.00	0.00	0.01	0.04	0.05
Sum	0.32	0.31	0.31	0.05	1.00

Indices for Classification Consistency and Classification Accuracy

Indices	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.89	0.91	0.97	0.77
Probability of Chance	0.56	0.54	0.90	0.30
Kappa (k)	0.76	0.80	0.69	0.67
Classification Accuracy	0.92	0.93	0.98	0.83

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	0.30	0.05	0.00	0.00	0.35
Basic	0.05	0.20	0.05	0.00	0.30
Proficient	0.00	0.05	0.24	0.01	0.31
Advanced	0.00	0.00	0.01	0.03	0.04
Sum	0.35	0.30	0.31	0.04	1.00

Indices for Classification Consistency and Classification Accuracy

Indices	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.90	0.90	0.98	0.77
Probability of Chance	0.54	0.55	0.92	0.31
Kappa (k)	0.77	0.77	0.68	0.67
Classification Accuracy	0.93	0.93	0.98	0.84

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	0.27	0.06	0.00	0.00	0.33
Basic	0.07	0.26	0.04	0.00	0.36
Proficient	0.00	0.04	0.18	0.02	0.24
Advanced	0.00	0.00	0.02	0.05	0.07
Sum	0.33	0.36	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)	0.87	0.92	0.97	0.76
Probability of Chance	0.56	0.57	0.87	0.30
Kappa (k)	0.72	0.81	0.73	0.66
Classification Accuracy	0.91	0.95	0.98	0.83

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	0.04	0.00	0.00	0.18
Basic	0.04	0.23	0.05	0.00	0.32
Proficient	0.00	0.06	0.18	0.04	0.28
Advanced	0.00	0.00	0.04	0.18	0.22
Sum	0.18	0.33	0.27	0.22	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	0.92	0.73
Probability of Chance	0.70	0.50	0.65	0.26
Kappa (k)	0.71	0.78	0.76	0.63
Classification Accuracy	0.94	0.92	0.93	0.79

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	0.15	0.04	0.00	0.00	0.19
Basic	0.05	0.18	0.06	0.00	0.29
Proficient	0.00	0.05	0.20	0.05	0.30
Advanced	0.00	0.00	0.04	0.17	0.22
Sum	0.20	0.28	0.30	0.22	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	0.91	0.70
Probability of Chance	0.69	0.50	0.66	0.26
Kappa (k)	0.70	0.76	0.73	0.59
Classification Accuracy	0.94	0.92	0.93	0.79

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	0.24	0.04	0.00	0.00	0.28
Basic	0.05	0.12	0.06	0.00	0.23
Proficient	0.00	0.06	0.16	0.05	0.27
Advanced	0.00	0.00	0.05	0.16	0.22
Sum	0.29	0.22	0.28	0.22	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	0.89	0.69
Probability of Chance	0.60	0.50	0.66	0.25
Kappa (k)	0.78	0.75	0.68	0.58
Classification Accuracy	0.94	0.92	0.92	0.77

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	0.20	0.04	0.00	0.00	0.25
Basic	0.04	0.18	0.05	0.00	0.27
Proficient	0.00	0.06	0.19	0.05	0.30
Advanced	0.00	0.00	0.04	0.14	0.19
Sum	0.24	0.28	0.29	0.19	1.00

Indices for Classification Consistency and Classification Accuracy

Indices	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.88	0.90	0.71
Probability of Chance	0.63	0.50	0.69	0.26
Kappa (k)	0.77	0.77	0.69	0.61
Classification Accuracy	0.94	0.92	0.93	0.79

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	0.23	0.04	0.00	0.00	0.28
Basic	0.04	0.15	0.06	0.00	0.25
Proficient	0.00	0.05	0.16	0.04	0.25
Advanced	0.00	0.00	0.05	0.17	0.22
Sum	0.28	0.25	0.26	0.21	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	0.91	0.72
Probability of Chance	0.60	0.50	0.66	0.25
Kappa (k)	0.78	0.78	0.73	0.62
Classification Accuracy	0.93	0.92	0.94	0.79

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 test 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. Since the Spring 2019 forms were reused, or reused with modifications, for Spring 2021, DIF analyses were already performed on all test items after the Spring 2019 test administration. The DIF analyses were repeated in Spring 2021. 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):

$$\text{Mantel } \chi^2 = \frac{\left(\sum_k F_k - \sum_k E(F_k)\right)^2}{\sum_k \text{Var}(F_k)},$$

where F_k is the sum of scores for the focal group at the k 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:

$$\alpha_{MH} = \frac{\sum\limits_{k=1}^{K} N_{r1k}N_{f0k} / N_k}{\sum\limits_{k=1}^{K} N_{f1k}N_{r0k} / N_k} \, ,$$

where N_{r1k} is the number of correct responses in the reference group at ability level k, N_{f0k} is the number of incorrect responses in the focal group at ability level k, N_k is the total number of responses, N_{f1k} is the number of correct responses in the focal group at ability level k, and N_{r0k} is the number of incorrect responses in the reference group at ability level k. MH-D DIF is then computed:

$$\text{MH-D DIF} = -2.35\ln(\alpha_{MH}) \, .$$

For selected response items, the MH (χ^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 α is assumed to be constant across the k matched levels. Then the χ^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 (Δ_{MH}) for interpretive purposes. Items were flagged for DIF using the following criteria:

- Moderate DIF: Significant MH chi-square statistic ($p < 0.05$) and $1.0 \le |\text{MH D-DIF}| < 1.5$
- Large DIF: Significant MH chi-square statistic ($p < 0.05$) and $|\text{MH D-DIF}| \ge 1.5$

For CR (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):

$$SMD = p_{Fk}(\sum_k m_{Fk} - \sum_k m_{Rk}) \, ,$$

where p_{Fk} = the proportion of the focal group members at the kth level of the matching variable, $m_{Fk} = 1/N_{F1k}$, and $m_{Rk} = 1/N_{R1k}$. 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 $|\text{ES}|$ is between 0.17 and 0.25
- Large DIF: If the MH statistic is significant ($p < 0.05$) and $|\text{ES}| \ge 0.25$

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 9-9 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 9-9 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 9-9.

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 economic status are shown in Table 9-7, results for disability status are presented in Table 9-8, and results for accommodation use are presented in Table 9-9.

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 (Δ_{MH}). After specifying these statistics for each item, the final column provides a flag status. The flag is based on SMD statistics for CR items and on MH (Δ_{MH}) statistics.

In Table 9-1, looking at all items and all grades and content areas, six items were flagged for moderate (B flag) gender DIF in the ELA tests (grades 5, 7, and 8). Of these items, two were flagged in favor of the focal group (females) and four were flagged against the focal group. Nine items were flagged for moderate DIF and one item was flagged for large DIF in the Mathematics tests (grades 3, 4, 6, and 7). For the nine moderate DIF items, three items were flagged in favor of female students and six items were flagged against female students; for the one large DIF item, the item was flagged against female students. Two items were flagged for moderate DIF and one item was flagged for large DIF in the Science tests (grades 4 and 8). All three items were flagged against female students. In addition, eight items were flagged for moderate DIF and four items were flagged for large DIF in the Social Studies tests (grades 4, 8, and 10). A total of five items were flagged in favor of female students and seven items were flagged against female students. Overall, ten items were flagged in favor of the focal group (females) and twenty-one items were flagged against the focal group across all grades and content areas. Of all items flagged for gender DIF, six displayed large DIF (either in favor of or against female students) and twenty-five items displayed moderate DIF.

The other DIF results in Tables 9-2 through 9-9 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 items across all content areas and subgroups. Fewer items were flagged for DIF in Spring 2021 compared to Spring 2019. At the same time, most of the items flagged in Spring 2021 were also flagged in Spring 2019.

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 2021 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 and 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 9-10 shows the correlations between the main test domains for ELA, and Tables 9-11 through 9-14 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 9-10, the correlations between the ELA main test domains of Reading, Writing, and Listening are moderate to high and range from 0.56 to 0.75 across all grades. The lowest correlations were observed between the Listening and Writing domains, while the highest correlations were observed between the Reading and Writing domains. The correlations between ELA content standards (see Table 9-11) are typically moderate for all grades and all standard pairs and range from 0.37 to 0.69. 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 9-12, the correlations between Mathematics content standards are moderate to high and range from 0.51 to 0.76. The correlations between Science content standards range from 0.61 to 0.70 (see Table 9-13), and the correlations between Social Studies content standards range from 0.51 to 0.75 (as shown in Table 9-14). 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 9-15 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 9-15, the ratios of the first two eigenvalues range from 5.60 to 7.66. 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 approximately six to eight times greater than the variance explained by the second component. The eigenvalue ratios range from 6.38 to 7.39 in ELA, from 5.60 to 7.66 in Mathematics, from 6.10 to 7.07 in Science, and from 6.65 to 7.30 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 relationship 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 2021 were computed and examined for the total student population and by subgroup. Table 9-16 shows the correlations between the content area scores for the total population of Wisconsin students. These correlations ranged from 0.71 (between ELA and Mathematics in grades 5 and 8 and between Mathematics and Social Studies in grade 8) to 0.82 (between ELA and Science in grade 4).

Tables 9-17 through 9-21 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 9-17, the correlations between the content area scores for male or female groups ranged from 0.70 to 0.83 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 0.58 to 0.84 (see Table 9-18). 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 9-19, the correlations between the content area scores by English proficiency status ranged from 0.51 to 0.72 for limited English proficiency students and from 0.70 to 0.82 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 9-20. These correlations ranged from 0.69 to 0.80 for students who were not economically disadvantaged and from 0.65 to 0.79 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 9-21. These correlations ranged from 0.69 to 0.81 for students without disabilities and from 0.57 to 0.78 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 learner, economically disadvantaged, or disabled compared to the groups of students classified as fully English proficient, not economically disadvantaged, or not disabled. 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. The correlations between the content area scores for the subgroups of students were found to be moderately to highly related. Despite 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 of 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 9-22. These correlations ranged from 0.62 (between Mathematics and Social Studies in grade 8) to 0.78 (between ELA and Science in grade 4). 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 and corresponding partial correlation coefficients were, however, relatively small, indicating that the effect of the student demographic characteristics on the relationship between the ELA, Mathematics, Science, and Social Studies test scores was small.

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 and the corresponding NAEP assessments for Reading and Mathematics are presented in Table 9-23. 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 9-23). With two 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 *Advanced* performance level on the Wisconsin Forward Science test for both grades 4 and 8 compared to the percentages of students classified in the *Advanced* performance level on the corresponding NAEP assessments.

Looking at the percentages of students classified as at or above *Proficient*, higher proportions of students were classified in this category of combined performance levels on the Wisconsin Forward Exam in ELA grade 4 and Science grades 4 and 8 in Spring 2021 compared to the corresponding NAEP Reading and Science assessments in Spring 2019 (Reading) or Spring 2015 (Science). Higher proportions of students were classified as 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 2021. As stated earlier in this report, such comparisons should be made with caution because the student performance on the Wisconsin Forward Exam in Spring 2021 may not reflect student performance in a typical administration year.

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 2021 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
ELA	5	26	MC	-0.10	-1.05	B-
	7	9	MC	0.08	1.11	B
	7	10	MC	-0.11	-1.27	B-
	8	4	MC	-0.09	-1.04	B-
	8	10	MC	0.08	1.43	B
	8	22	MC	-0.15	-1.44	B-
Mathematics	3	4	MC	-0.06	-1.05	B-
	3	19	MC	-0.07	-1.11	B-
	3	23	MC	-0.08	-1.46	B-
	4	7	MC	-0.14	-2.15	C-
	4	37	MC	0.05	1.01	B
	6	12	MC	-0.03	-1.26	B-
	6	16	MC	-0.08	-1.01	B-
	6	23	TE	0.08	1.03	B
	6	46	MC	0.05	1.03	B
	7	2	MC	-0.08	-1.14	B-
Science	4	28	TE	-0.08	-1.02	B-
	8	20	TE	-0.13	-1.42	C-
	8	34	TE	-0.09	-1.02	B-
Social Studies	4	23	MC	0.05	1.25	B
	4	31	TE	-0.13	-1.63	C-
	4	33	MC	0.09	1.57	C
	8	1	MC	-0.09	-2.00	C-
	8	14	MC	-0.11	-1.22	B-
	8	21	MC	0.03	1.67	C
	8	22	MC	-0.08	-1.07	B-
	8	25	MC	-0.10	-1.47	B-
	10	9	MC	-0.12	-1.48	B-
	10	24	TE	-0.12	-1.75	B-
	10	25	MC	0.07	1.08	B
	10	33	MC	0.11	1.48	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
ELA	5	27	TE	-0.08	-1.15	B-
Mathematics	6	12	MC	-0.09	-1.39	B-
	6	20	SA	-0.09	-1.32	B-
	6	23	TE	-0.08	-0.79	B-
	8	15	TE	-0.10	-2.05	B-
Science	8	21	TE	-0.08	-1.03	B-
Social Studies	4	7	TE	-0.12	-1.35	C-
	8	1	MC	-0.22	-3.02	C-
	8	6	MC	0.18	2.33	C
	8	11	MC	-0.12	-1.59	C-
	8	22	MC	-0.09	-1.24	B-
	8	25	MC	-0.16	-1.85	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	1	MC	-0.12	-1.97	C-

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
ELA	3	12	MC	-0.08	-1.05	B-
	5	14	MC	0.10	1.17	B
Mathematics	5	23	MC	-0.07	-1.03	B-
	6	7	MC	0.07	1.12	B
	6	35	MC	-0.11	-1.54	C-
	6	45	MC	-0.07	-1.10	B-
	6	46	MC	0.04	1.09	B
	7	31	MC	-0.07	-1.21	B-
	8	9	TE	0.09	1.72	B
Social Studies	4	23	MC	0.05	1.12	B
	8	1	MC	-0.09	-1.94	C-
	8	21	MC	0.03	1.84	C
	8	25	MC	-0.08	-1.11	B-
	10	4	MC	-0.08	-1.04	B-
	10	22	MC	0.08	1.07	B
	10	25	MC	0.08	1.46	B
	10	34	TE	-0.08	-0.90	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
Mathematics	6	20	SA	-0.08	-1.15	B-
	6	23	TE	-0.09	-1.00	B-
Social Studies	8	11	MC	-0.07	-1.03	B-
	10	14	MC	0.11	1.27	B

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
ELA	3	12	MC	-0.08	-1.01	B-
Mathematics	6	20	SA	-0.11	-1.61	B-
	7	31	MC	-0.08	-1.20	B-
Social Studies	8	1	MC	-0.13	-1.60	C-
	10	1	MC	-0.10	-1.11	B-

Table 9-7 Items Flagged for DIF by Economic Status, Focal Group: Economically Disadvantaged Students

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
Social Studies	8	1	MC	-0.06	-1.01	B-

Table 9-8 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
ELA	4	15	TE	-0.10		B-
	5	1	MC	-0.09	-1.14	B-
	5	13	TE	-0.19		C-
	5	25	TE	-0.07	-1.04	B-
	6	12	TE	-0.13		B-
	8	10	MC	-0.11	-1.19	B-
Mathematics	4	10	MC	-0.08	-1.27	B-
	6	12	MC	-0.16	-2.48	C-
	6	45	MC	0.08	1.18	B
Science	4	34	TE	0.06	1.28	B
Social Studies	8	21	MC	-0.07	-1.33	B-

Table 9-9 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	4	10	MC	-0.11	-1.39	B-
	4	30	MC	0.07	1.27	B
	4	31	MC	0.10	1.01	B
	5	1	SA	-0.11	-1.71	B-
	5	41	MC	0.05	1.38	B
	5	42	MC	-0.11	-1.29	B-
	6	12	MC	-0.23	-2.92	C-
	6	20	SA	-0.11	-1.88	B-
	6	45	MC	0.09	1.37	B
	8	42	MC	-0.07	-1.06	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 9-10 Correlations between English Language Arts Test Domains

Grade	ELA Domain	Listening	Reading
3	Reading	0.65	
	Writing	0.63	0.75
4	Reading	0.57	
	Writing	0.56	0.75
5	Reading	0.66	
	Writing	0.64	0.74
6	Reading	0.64	
	Writing	0.61	0.72
7	Reading	0.62	
	Writing	0.58	0.71
8	Reading	0.65	
	Writing	0.63	0.75

Table 9-11 Correlations between English Language Arts Content Standards

Grade	Standard Code	A	B	C	D	E	F
3	B	0.52					
	C	0.56	0.39				
	D	0.55	0.38	0.42			
	E	0.59	0.41	0.45	0.50		
	F	0.64	0.45	0.50	0.52	0.56	
	G	0.62	0.42	0.49	0.48	0.53	0.56
4	B	0.61					
	C	0.63	0.59				
	D	0.56	0.53	0.55			
	E	0.56	0.52	0.55	0.55		
	F	0.49	0.46	0.49	0.47	0.47	
	G	0.51	0.47	0.49	0.49	0.48	0.40
5	B	0.69					
	C	0.49	0.50				
	D	0.50	0.51	0.37			
	E	0.60	0.59	0.42	0.49		
	F	0.51	0.52	0.39	0.44	0.49	
	G	0.60	0.60	0.45	0.49	0.57	0.49
6	B	0.60					
	C	0.63	0.54				
	D	0.53	0.46	0.50			
	E	0.52	0.45	0.48	0.46		
	F	0.52	0.44	0.48	0.45	0.44	
	G	0.58	0.50	0.55	0.50	0.49	0.46
7	B	0.64					
	C	0.63	0.56				
	D	0.47	0.44	0.43			
	E	0.56	0.52	0.51	0.43		
	F	0.50	0.46	0.47	0.41	0.45	
	G	0.57	0.52	0.52	0.43	0.50	0.43
8	B	0.66					
	C	0.64	0.57				
	D	0.58	0.52	0.53			
	E	0.62	0.55	0.54	0.56		
	F	0.48	0.42	0.41	0.44	0.46	
	G	0.60	0.53	0.55	0.53	0.56	0.42

Note: Standard Codes are as follows: A = Reading - Key Ideas and Details; B = Reading - Craft & Structure/ Integration of Knowledge & Ideas; C = Reading - Vocabulary Use; D = Writing/Language - Text Types and Purpose; E = Writing/Language - Research; F = Writing/Language - Language Conventions; G = Listening

Table 9-12 Correlations between Mathematics Content Standards

Grade	Standard Code	A	B	C	D	E	F	G	H	I
3	B	0.74								
	C	0.68	0.68							
	D	0.74	0.75	0.70						
	E	0.68	0.69	0.70	0.71					
4	B	0.66								
	C	0.64	0.76							
	D	0.61	0.71	0.73						
	E	0.51	0.60	0.63	0.62					
5	B	0.69								
	C	0.64	0.68							
	D	0.62	0.65	0.66						
	E	0.66	0.67	0.65	0.65					
6	F					0.58				
	G					0.63	0.72			
	H					0.61	0.68	0.74		
	I					0.54	0.58	0.62	0.61	
7	F					0.54				
	G					0.54	0.68			
	H					0.55	0.67	0.67		
	I					0.54	0.70	0.66	0.67	
8	G					0.62				
	H					0.63		0.67		
	I					0.58		0.55	0.59	
	J					0.65		0.63	0.70	0.64

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 9-13 Correlations between Science Content Standards

Grade	Standard Code	A	B	C
4	B	0.70		
	C	0.63	0.61	
	D	0.70	0.68	0.62
8	B	0.69		
	C	0.66	0.64	
	D	0.69	0.67	0.64

Note: Standard Codes are as follows: A = Life Science; B = Physical Science; C = Earth and Space Science; D = Engineering

Table 9-14 Correlations between Social Studies Content Standards

Grade	Standard Code	A	B	C	D
4	B	0.65			
	C	0.51	0.61		
	D	0.58	0.65	0.54	
	E	0.60	0.71	0.59	0.63
8	B	0.72			
	C	0.61	0.66		
	D	0.65	0.67	0.58	
	E	0.63	0.65	0.56	0.58
10	B	0.70			
	C	0.67	0.75		
	D	0.64	0.70	0.67	
	E	0.62	0.70	0.67	0.64

Note: Standard Codes are as follows: A = Geography; B = History; C = Political Science and Citizenship; D = Economics; E = The Behavioral Sciences

Table 9-15 Principal Components Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
ELA	3	8.063	1.191	6.773
	4	8.174	1.106	7.390
	5	8.616	1.188	7.256
	6	7.535	1.181	6.378
	7	7.650	1.155	6.623
	8	8.686	1.177	7.377
Mathematics	3	11.286	1.531	7.373
	4	10.720	1.783	6.014
	5	10.476	1.368	7.656
	6	9.986	1.784	5.596
	7	9.510	1.561	6.094
	8	9.801	1.678	5.840
Science	4	7.945	1.303	6.097
	8	8.160	1.155	7.066
Social Studies	4	8.308	1.250	6.645
	8	8.931	1.262	7.078
	10	9.978	1.367	7.298

Table 9-16 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	0.74	0.82	0.80	0.77	0.73	0.81
5	0.71					
6	0.74					
7	0.74					
8	0.71	0.80	0.81	0.74	0.71	0.81

Table 9-17 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	0.76					
4	Female	0.75	0.83	0.80	0.77	0.73	0.81
	Male	0.75	0.82	0.80	0.78	0.73	0.81
5	Female	0.72					
	Male	0.73					
6	Female	0.74					
	Male	0.76					
7	Female	0.75					
	Male	0.74					
8	Female	0.72	0.81	0.82	0.74	0.71	0.81
	Male	0.72	0.81	0.81	0.74	0.70	0.81

Table 9-18 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	0.74					
	African American	0.62					
	Hispanic	0.70					
	Asian	0.76					
	American Indian	0.66					
	Two or More	0.75					
4	White	0.71	0.80	0.79	0.75	0.70	0.79
	African American	0.62	0.73	0.68	0.66	0.61	0.71
	Hispanic	0.69	0.79	0.76	0.73	0.69	0.78
	Asian	0.76	0.82	0.81	0.78	0.75	0.82
	American Indian	0.68	0.78	0.75	0.72	0.67	0.77
	Two or More	0.72	0.80	0.77	0.77	0.73	0.79
5	White	0.69					
	African American	0.59					
	Hispanic	0.66					
	Asian	0.73					
	American Indian	0.60					
	Two or More	0.70					
6	White	0.73					
	African American	0.62					
	Hispanic	0.68					
	Asian	0.76					
	American Indian	0.69					
	Two or More	0.73					
7	White	0.72					
	African American	0.64					
	Hispanic	0.69					
	Asian	0.78					
	American Indian	0.67					
	Two or More	0.73					
8	White	0.70	0.79	0.79	0.72	0.68	0.79
	African American	0.59	0.74	0.76	0.60	0.58	0.73
	Hispanic	0.66	0.78	0.79	0.68	0.66	0.79
	Asian	0.74	0.81	0.84	0.77	0.73	0.83
	American Indian	0.62	0.73	0.79	0.65	0.64	0.79
	Two or More	0.70	0.79	0.82	0.73	0.69	0.80

Table 9-19 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	0.75					
	Limited English Proficient	0.66					
4	Fully English Proficient	0.74	0.82	0.80	0.77	0.73	0.81
	Limited English Proficient	0.63	0.72	0.68	0.67	0.63	0.71
5	Fully English Proficient	0.71					
	Limited English Proficient	0.55					
6	Fully English Proficient	0.74					
	Limited English Proficient	0.54					
7	Fully English Proficient	0.74					
	Limited English Proficient	0.54					
8	Fully English Proficient	0.71	0.79	0.81	0.74	0.70	0.81
	Limited English Proficient	0.51	0.67	0.67	0.51	0.51	0.67

Table 9-20 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	0.72	0.80	0.79	0.75	0.70	0.79
	Economically Disadvantaged	0.69	0.79	0.76	0.73	0.69	0.78
5	Not Economically Disadvantaged	0.69					
	Economically Disadvantaged	0.66					
6	Not Economically Disadvantaged	0.73					
	Economically Disadvantaged	0.69					
7	Not Economically Disadvantaged	0.73					
	Economically Disadvantaged	0.69					
8	Not Economically Disadvantaged	0.70	0.78	0.79	0.73	0.69	0.79
	Economically Disadvantaged	0.66	0.77	0.79	0.68	0.65	0.79

Table 9-21 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	0.71					
4	Not Disabled	0.73	0.81	0.80	0.77	0.72	0.80
	Disabled	0.68	0.78	0.73	0.73	0.68	0.78
5	Not Disabled	0.70					
	Disabled	0.61					
6	Not Disabled	0.73					
	Disabled	0.64					
7	Not Disabled	0.72					
	Disabled	0.63					
8	Not Disabled	0.70	0.78	0.79	0.73	0.69	0.80
	Disabled	0.58	0.74	0.74	0.60	0.57	0.75

Table 9-22 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	0.67	0.78	0.75	0.70	0.65	0.75
5	0.63					
6	0.66					
7	0.66					
8	0.64	0.76	0.76	0.67	0.62	0.76

Table 9-23 Comparison of Most Recent Wisconsin NAEP and Spring 2021 Wisconsin Forward Exam Impact Data

Content	Grade	Wisconsin NAEP Percentages of Students							Wisconsin Forward Exam Spring 2021 Percentages of Students					
		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	27.17	32.71	32.67	7.45	40.12	72.83
Reading/ELA	8	2019	24	38	34	4	39	76	26.05	38.29	28.02	7.64	35.66	73.95
Math	4	2019	20	35	34	11	45	80	22.32	36.61	30.84	10.23	41.07	77.68
Math	8	2019	24	34	29	12	41	76	32.47	37.53	23.62	6.38	30.00	67.53
Science	4	2015	21	38	40	1	41	79	16.18	32.67	33.03	18.13	51.16	83.82
Science	8	2015	25	35	38	2	40	75	18.56	29.97	30.65	20.82	51.47	81.44

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 2021 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, 2021).

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 2020–21 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 2021.

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 administration.

Social Studies scale scores range from 200 to 570 for grade 4, from 420 to 780 for grade 8, and from 490 to 890 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. Social Studies scale scores can be compared within a grade level across the administrations from Spring 2016 to Spring 2021.

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 2020–21 administration of the Wisconsin Forward Exam. The *Guide* was developed collaboratively by DRC and DPI staff.

10.1.3 Note about Spring 2021 Test Results

It should be noted that the Spring 2021 test takers were not fully representative of the Wisconsin student population. Due to the circumstances related to the COVID-19 pandemic, participation in the Spring 2021 Forward Exam was recommended but not mandated by DPI. As stated in Part 4 of this report, the test participation rates for grades 3 through 8 ranged from approximately 84% to 87% and the test participation rate for grade 10 was approximately 75%. Therefore, aggregated test performance results in Spring 2021 should be interpreted in the context of several factors including possible disrupted learning in Spring 2020 and the subsequent school year, differences in instruction delivery (e.g., online, in-person, hybrid model), lower than typical test participation rates, overrepresentation of certain demographic groups and underrepresentation of other groups in the tested population. As such, any comparison of the group test results from Spring 2021 with the test results from Spring 2019 should be done with caution.

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 2021 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. 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 550.13 for grade 3 to 628.22 for grade 8. This mean scale score pattern supports the ELA vertical scale properties.

- Standard deviations ranged from 46.61 to 58.76 scale score points.
- In each grade level, student scores spanned the full scale score range from the LOSS to the HOSS.

Mathematics

- Mean scale score increased as grade level increased, ranging from 548.97 for grade 3 to 638.33 for grade 8. This mean scale score pattern supports the Mathematics vertical scale properties.
- Standard deviations ranged from 53.33 to 60.06 scale score points.
- In each grade level, student scores spanned the full scale score range from the LOSS to the HOSS.

Science

- Mean scale scores were 497.39 and 697.31 for grades 4 and 8, respectively.
- Standard deviations were 50.30 and 49.86 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 391.30, 595.46, and 696.64 for grades 4, 8, and 10, respectively.
- Standard deviations ranged from 52.04 to 56.94 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 6.50 scale score points in grade 3 to 14.11 scale score points in grade 8.
- In Mathematics, male students as a group showed higher mean scale scores in grades 3 through 6 (differences between 1.63 and 5.73 scale score points) and a lower mean scale score in grade 8 (difference of 2.69 scale score points) than female students.

There was no practical difference between male and female student performance in grade 7.

- In Science, the mean scale scores were lower for female students, with a difference of 3.76 scale score points in grade 4 and 1.20 scale score points in grade 8.
- In Social Studies, female students performed better than male students in all grades. There was a difference of less than 1 scale score point between mean scale scores in grade 4. The differences between mean scale scores in grades 8 and 10 were 2.93 and 2.91, respectively.

Race/Ethnicity

- The scale score results showed some consistent performance differences by ethnicity.
- In almost every grade and content area, White students as a group had the highest mean scale scores, followed by Asian students, Hispanic students, American Indian students, and African American students. Few exceptions included Mathematics grade 8, where Asian students slightly outperformed White students (difference between the two groups was 1.62 scale score points), Social Studies grade 10, where Asian students performed as well as White students (difference was less than half a scale score point between the two groups), Science grade 4 and Social Studies grade 10, where American Indian students performed as well as Hispanic students (difference between the two groups was less than one scale score point).
- 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 of a standard deviation lower than the mean scale scores of White students for most grades and content areas. The mean scale scores of Asian students were less than 6 scale score points lower than the mean scale scores of White students in grades 6, 7, 8, and 10 (any content area). The differences in scale score means between Asian and White students were larger in grades 3 through 5 of any content area and ranged from 10.84 scale score points in Mathematics grade 5 to 22.12 scale score points in Science grade 4.

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 31.15 scale score points in Science grade 8 to over 40 scale score points in Mathematics grades 3, 4, 6, and 7.
- For every grade and content area, the mean scale scores of students who were economically disadvantaged were typically more than two-thirds 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. Differences ranged from 32.29 scale score points in ELA grade 3 to over 60 scale score points in ELA grade 8 and Mathematics grade 7.
- 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 at least three quarters of a standard deviation to over one standard deviation.

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. Differences ranged from 33.36 scale score points in ELA grade 3 to 64.52 scale score points in Social Studies grade 10.
- 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 about three quarters of a standard deviation to over 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 12.44 to 34.16 scale score points for ELA grades and from 64.70 to 78.30 scale score points for Mathematics grades. 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 fewer than 100 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 half of a standard deviation for ELA to over one standard deviation for Mathematics.
- Science and Social Studies student performance for students using testing accommodations was 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 35% (grade 3) and 43% (grade 7) of students were either *Proficient* or *Advanced*.
- Between about 5% and 9% 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 57% below *Proficient* in grade 7 to approximately 65% below *Proficient* in grade 3.

Mathematics

- Between approximately 30% (grade 8) and 45% (grade 3) of students were either *Proficient* or *Advanced*.
- The proportion of students who were *Advanced* ranged from approximately 3% (grade 7) to over 10% (grades 3 and 4).
- Across all grade levels, the percentages of students below *Proficient* ranged from approximately 55% in grade 3 to 70% in grade 8.

Science

- Approximately 51% of students were either *Proficient* or *Advanced* in both grades 4 and 8.
- The percentage of students classified as *Advanced* was approximately 18% in grade 4 and 21% in grade 8.
- The proportion of students classified as below *Proficient* was approximately 49% in both grades.

Social Studies

- Approximately 49% of students in grade 4, about 48% of students in grade 8, and about 47% of students in grade 10 were classified as *Proficient* or *Advanced*.
- Between 17% and 20% of students were classified as *Advanced* across the three grades.
- The percentage of students classified as below *Proficient* ranged from approximately 51% in grade 4 to 53% 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 5% (grade 3) to over 8% (grade 8) for ELA. A reversed trend was observed for Mathematics. Between approximately 3% (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 less than 1%. In Science, more male students than female students were classified as *Proficient* or above in grade 4 (difference of approximately 3%). The difference between the gender groups in Science grade 8 was less than 1%. In Social Studies, the differences in percentages of male and female classified as *Proficient* or above were about 2% or less across the three grades, with slightly more female students being classified in the two highest performance categories.

There were some consistent patterns in performance by ethnicity across grades and content areas. 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, there were higher percentages of students who were fully English proficient classified as *Proficient* or *Advanced* compared to students with limited English proficiency, with the differences ranging from approximately 27% in ELA grade 3 to over 40% in Science grade 8 and Social Studies grades 8 and 10. There were much lower percentages of fully English proficient students who were classified in the lowest performance level in all grades and content areas.

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 37% (Science grade 8 and Social Studies grade 8), depending on grade level and content area. There were much higher percentages of students without disabilities in the reporting category *Advanced.* There were also much lower percentages of students without disabilities in the lowest performance level than students with disabilities. This pattern was evident in all grades and all content areas.

There were consistent differences in performance between economically disadvantaged students and not economically disadvantaged students. In every grade and content area, between approximately 23% (ELA grade 8) and 32% (Mathematics grade 3) more students who were not economically disadvantaged were classified as *Proficient* or above compared to their economically disadvantaged peers. There were much higher percentages of students who were economically disadvantaged who were classified in the two lowest performance categories.

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 20% across ELA grades and from 30% to 39% 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 fewer than 100 students per grade using 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 fewer than 50 students per grade using 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 J 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's Model Academic 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.* The exception to this recommendation in Spring 2021 is SPI scores in Mathematics assessments (all grades), Science assessments (both grades), and the Social Studies assessment in grade 4. These forms were reused without any modification to operational test items from Spring 2019, and the SPI scores are comparable for the same standard across the two administrations.

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 43.36 to 67.92 for content standards (including Listening) and from 50.98 to 62.41 for domains (Reading or Writing), indicating that the items were moderately difficult to easy for examinees. No specific pattern of standard difficulty was found across ELA grades and standards. Content standard C (Reading—Vocabulary Use) was the easiest in grade 7. According

to mean SPI scores, content standard D (Writing/Language—Text Types and Purposes) was the easiest in grade 5. Content standard F (Writing/Language—Language Conventions) was easier than other standards in grades 4 and 8. Listening domain (content standard G) was the easiest in grades 3 and 6. Among the most difficult standards were content standard B (Reading—Craft and Structure) in grades 3 and 7 and content standard G (Listening) in grade 4. Content standard F (Writing/Language—Language Conventions) was the easiest in grades 4 and 8. The Reading domain was more difficult than the Listening and Writing domains for students in grades 3, 5, 7, and 8.

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 35.26 for content standard E (Geometry) in grade 6 to 58.48 for content standard B (Number and Operations in Base Ten) in grade 3. The Mathematics items were more challenging in higher grades than lower grades. There was no consistent pattern in regard to the content standard difficulty across grade levels. Content standard E (Geometry) was most difficult in grades 6 and 7, and content standard G (The Number System) was the most difficult in grade 8. Content standard B (Number and Operations in Base Ten) was easier than other standards in grades 3, 4, and 5.

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 45.91 to 62.77, indicating that the test items were of medium difficulty. Mean SPI scores indicate that content standard C (Earth and Space Science) was the most difficult in grade 4 and that content standard B (Physical Science) was 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 56.68 to 71.06, indicating that the test items ranged from somewhat difficult to relatively easy. The mean SPI scores indicate that the most difficult content standard varied between the three Social Studies grades. In grades 4 and 8, the most difficult standard was content standard D (Economics). In grade 10, the most difficult standard was content standard C (Political Science and Citizenship).

Summary of Student Performance Indicator Results

Overall, the mean SPI scores across grades and content standards ranged in difficulty. The content standards with SPI mean scores greater than 70 were content standard B (History) and content standard E (The Behavioral Sciences) in Social Studies grade 4. There were no SPI mean scores lower than 30 in the Wisconsin Forward Exam Spring 2021 test administration.

It is important to note that some variation in 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.

For the Wisconsin Forward Exam assessments, four years of test scores on the same reporting scales are available, and the state-level mean scale scores and standard deviations for the 2016, 2017, 2018, 2019, and 2021 administrations are presented for ELA, Mathematics, and Social Studies in Tables 10-28, 10-29, and 10-31, 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 2021 scale scores are comparable only with the Spring 2019 scale scores and not with the scores prior to the Spring 2019 administration. Therefore, only two years of scale score data are presented for Science in Table 10-30. The scale score statistics, along with the test participation rates, presented in Tables 10-28 through 10-31 are based on the census data, including students attending public, choice, and private schools. The test 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.

As stated earlier in this report, students who participated in the Spring 2021 test administration were not fully representative of the Wisconsin student population. As such, the longitudinal comparisons of student performance should be considered with caution. The Spring 2021 impact data reflect the performance of students who participated in the assessment while the impact data in administrations prior to Sprig 2021 reflect the performance of the full student population. Therefore, the differences in performance between the Spring 2019 (total population) and Spring 2021 (not a full population) may not accurately represent performance changes between the two administration years or the impact of the COVID-19 pandemic on student learning.

It was observed that the mean scale score for ELA decreased consistently for all grades between Spring 2019 and Spring 2021. The mean scale score decrease ranged from about 1 scale score point for grade 8 to over 4 scale score points for grades 3 and 4 (see Table 10-28).

Similar to ELA, the Mathematics mean scale scores decreased for all grades between Spring 2019 and Spring 2021. The mean scale score decrease ranged from approximately 5 scale score points for grade 7 to approximately 9 scale score points for grade 6 (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 just over 2 points in each grade (refer to Table 10-30).

Students in grades 4 and 8 performed less well in the Social Studies assessments in Spring 2021 compared to their peers in Spring 2019. The mean scale score decrease was approximately 5 scale score points in grade 4 and approximately 3 scale score points in grade 8. Contrary to the mean scale score change pattern observed for ELA, Mathematics, Science, and Social Studies grades 4 and 8, an increase in the mean scale score (approximately 4 scale score points) was observed for Social Studies grade 10 between 2019 and 2021 (see Table 10-31).

Tables 10-32 through 10-35 show the percentages of students in each achievement level in the Spring 2016, 2017, 2018, 2019, and 2021 test administrations for ELA, Mathematics, and Social Studies and in the Spring 2019 and 2021 test administrations for Science. The results presented in these tables are based on all Wisconsin students who participated in assessment in a given year, including students attending public, choice, and private schools. The pattern of student performance classification change between Spring 2019 and Spring 2021 was consistent with the pattern of the mean scale score change between the two test administrations in all grades and content areas.

For ELA, a decrease in the percentage of students at or above *Proficient* was observed for all grades, ranging from approximately 1% for grade 8 to about 4% for grade 3 (see Table 10-32). For Mathematics, the decrease in the percentage of students classified as *Proficient* or above ranged from approximately 4% for grades 4 and 7 to approximately 7% for grade 6 (see Table 10-33).

For Science, approximately 2% fewer students were classified as *Proficient* or above in Spring 2021 compared to Spring 2019 for both grades (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. The Spring 2019 and 2021 Science impact data are comparable with each other but should not be directly compared with the previous years' impact data. The Spring 2019 impact data in Science are a new baseline for longitudinal comparisons, and 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 Technical Report 2018* available at DPI's website: https://dpi.wi.gov/assessment/forward/resources#documentation.

For Social Studies, a decrease of approximately 3% in the percentage of students at or above *Proficient* was observed for grades 4 and 8 and an increase of approximately 2% in the percentage of students at or above *Proficient* was observed for grade 10 (refer to Table 10-35).

Overall, the percentages of students classified in the *Proficient* or above categories were found to be lower in Spring 2021 compared to Spring 2019 for each grade and content area

except for Social Studies grade 10, where a higher percentage of students were classified into the *Proficient* or above categories in Spring 2021 compared to Spring 2019.

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 2021. 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. With the exception of Social Studies grade 10, students who participated in the Spring 2021 ELA, Mathematics, Science, and Social Studies assessments performed less well on the assessments compared to the entire population of students in Spring 2019. However, because the Spring 2021 test participants were not fully representative of the Wisconsin student population, the true difference in performance between the Spring 2019 and Spring 2021 cannot be assessed without further studies.

Classroom teachers may still use the Spring 2021 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. If the Spring 2021 results are used for accountability purposes, steps should be taken to account for lower than 95% participation in the assessment..

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	52930	550.13	46.61	0.11	1.18	330	900	330	900
	4	52706	577.65	51.31	0.04	0.53	340	930	340	930
	5	54010	593.12	49.01	0.13	1.22	350	940	350	940
	6	55511	604.02	50.12	-0.27	0.36	360	950	360	950
	7	56295	625.36	55.21	-0.05	0.42	370	960	370	960
	8	56756	628.22	58.76	-0.06	0.53	380	970	380	970
Mathematics	3	52892	548.97	56.39	-0.61	1.87	360	760	360	760
	4	52658	571.50	53.33	-0.61	1.11	405	800	405	800
	5	53932	594.26	56.01	-0.91	1.62	430	830	430	830
	6	55462	602.06	57.74	-0.56	1.13	440	870	440	870
	7	56247	620.00	60.06	-0.74	1.14	450	880	450	880
	8	56726	638.33	56.94	-0.47	1.41	470	890	470	890
Science	4	52417	497.39	50.30	0.18	0.32	300	725	300	725
	8	56485	697.31	49.86	0.27	0.72	480	945	480	945
Social Studies	4	52392	391.30	56.94	-0.36	1.23	200	570	200	570
	8	56409	595.46	52.04	0.03	1.27	420	780	420	780
	10	51433	696.64	56.00	-0.34	1.25	490	890	490	890

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	27044	546.95	46.38	330	900	25886	553.45	46.61	330	900
	4	26877	573.49	51.31	340	930	25829	581.98	50.95	340	930
	5	27494	588.31	49.04	350	940	26516	598.11	48.48	350	940
	6	28342	599.04	50.99	360	950	27169	609.21	48.66	360	950
	7	28847	619.58	55.53	370	960	27448	631.44	54.22	370	960
	8	29223	621.38	59.23	380	970	27533	635.49	57.38	380	970
Mathematics	3	27027	551.42	57.57	360	760	25865	546.40	55.02	360	760
	4	26853	574.31	54.38	405	800	25805	568.58	52.06	405	800
	5	27444	596.33	57.48	430	830	26488	592.13	54.37	430	830
	6	28313	602.86	59.78	440	870	27149	601.23	55.52	440	870
	7	28832	619.93	63.37	450	880	27415	620.08	56.36	450	880
	8	29197	637.03	60.24	470	890	27529	639.72	53.19	470	890
Science	4	26737	499.23	51.36	300	725	25680	495.47	49.09	300	725
	8	29078	697.89	52.19	480	945	27407	696.69	47.26	480	945
Social Studies	4	26712	391.03	58.05	200	570	25680	391.59	55.77	200	570
	8	29055	594.04	54.51	420	780	27354	596.97	49.24	420	780
	10	26475	695.23	59.23	490	890	24958	698.14	52.32	490	890

Table 10-3 Scale Score Descriptive Statistics for English Language Arts by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	36547	559.74	43.49	330	900
	4	36592	587.76	48.41	340	930
	5	37720	601.98	46.55	350	940
	6	38850	612.58	47.19	360	950
	7	39754	634.19	52.34	370	960
	8	40254	636.92	56.35	380	970
African American	3	4101	509.62	42.18	330	900
	4	4079	535.00	46.14	340	739
	5	4255	553.54	44.96	350	761
	6	4215	564.34	49.83	360	950
	7	4251	582.59	52.69	370	827
	8	4362	585.12	55.65	380	866
Hispanic	3	6787	528.56	42.77	330	900
	4	6789	555.89	48.16	340	930
	5	6900	573.64	44.97	350	741
	6	7117	583.49	47.99	360	751
	7	7199	603.96	53.47	370	908
	8	7112	608.24	56.61	380	866
Asian	3	2237	544.68	49.08	330	900
	4	2166	571.13	51.41	374	930
	5	2047	591.10	50.49	350	940
	6	2256	608.45	49.22	360	783
	7	2129	631.88	56.32	429	870
	8	2098	635.77	58.04	397	970
American Indian	3	563	525.74	43.06	330	678
	4	599	549.14	45.68	352	742
	5	595	566.49	41.81	445	725
	6	584	581.37	46.69	451	712
	7	568	596.28	49.85	445	748
	8	635	597.61	55.96	380	770
Two or More	3	2695	545.28	47.16	330	900
	4	2481	570.80	51.47	340	930
	5	2493	588.45	50.91	390	940
	6	2489	597.52	50.84	360	792
	7	2394	620.21	55.53	436	908
	8	2295	621.16	59.10	380	970

Table 10-4 Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	36533	562.14	49.20	360	760
	4	36579	583.88	46.87	405	800
	5	37709	606.09	49.42	430	830
	6	38823	613.76	52.16	440	870
	7	39723	631.29	54.24	450	880
	8	40214	648.53	52.39	470	890
African American	3	4058	492.43	59.00	360	700
	4	4051	515.84	54.00	405	685
	5	4187	540.17	60.50	430	725
	6	4182	548.35	58.68	440	711
	7	4225	564.36	62.12	450	844
	8	4351	588.23	56.41	470	793
Hispanic	3	6803	520.30	56.09	360	760
	4	6794	545.91	51.98	405	800
	5	6902	568.51	56.39	430	830
	6	7132	574.84	55.57	440	747
	7	7203	594.05	59.91	450	880
	8	7128	614.46	54.50	470	890
Asian	3	2241	544.16	59.32	360	760
	4	2164	568.14	57.03	405	800
	5	2055	595.25	58.09	430	830
	6	2252	608.93	60.46	440	870
	7	2127	628.50	65.03	450	880
	8	2103	650.15	61.61	470	890
American Indian	3	562	514.70	55.11	360	760
	4	595	537.14	55.48	405	723
	5	597	559.46	57.48	430	685
	6	584	567.20	56.46	440	707
	7	565	585.31	57.59	450	768
	8	636	605.83	56.65	470	762
Two or More	3	2695	539.04	57.17	360	760
	4	2475	561.12	52.40	405	717
	5	2482	584.99	56.45	430	830
	6	2489	589.80	59.82	440	870
	7	2404	609.67	60.94	450	880
	8	2294	626.87	59.48	470	890

214

Table 10-5 Scale Score Descriptive Statistics for Science by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	36548	508.85	46.99	300	725
	8	40179	706.15	47.76	480	945
African American	4	3941	448.64	42.12	300	709
	8	4230	654.49	41.68	480	847
Hispanic	4	6720	472.23	45.01	300	708
	8	7082	676.20	45.56	480	945
Asian	4	2143	486.73	49.49	300	709
	8	2087	700.36	49.89	556	945
American Indian	4	598	471.58	44.53	311	634
	8	626	671.43	45.84	521	933
Two or More	4	2467	489.42	47.83	345	725
	8	2281	690.85	49.87	480	945

Table 10-6 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	36546	403.05	51.77	200	570
	8	40138	604.12	49.25	420	780
	10	39414	703.08	53.51	490	890
African American	4	3926	337.31	59.59	200	570
	8	4218	552.49	50.10	420	780
	10	2421	650.45	58.15	490	841
Hispanic	4	6718	367.64	54.61	200	570
	8	7061	575.00	49.89	420	780
	10	5531	673.05	55.74	490	890
Asian	4	2143	382.26	57.58	200	570
	8	2081	599.08	52.17	420	780
	10	1854	702.69	57.09	490	890
American Indian	4	595	364.51	51.19	200	570
	8	625	571.80	44.92	420	780
	10	526	673.50	51.08	490	809
Two or More	4	2464	382.00	57.87	200	570
	8	2286	589.11	51.84	420	780
	10	1687	690.48	56.51	490	890

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	20537	530.50	43.49	330	900	32393	562.57	44.17	330	900
	4	20068	555.89	47.72	340	784	32638	591.03	48.80	349	930
	5	20742	572.75	45.83	350	940	33268	605.82	46.59	350	940
	6	20699	583.30	48.88	360	792	34812	616.34	46.68	360	950
	7	20598	602.60	52.83	370	960	35697	638.50	52.21	370	960
	8	20225	605.30	56.43	380	970	36531	640.92	56.12	380	970
Mathematics	3	20480	523.52	57.01	360	760	32412	565.04	49.69	360	760
	4	20032	546.36	53.17	405	800	32626	586.94	47.21	405	800
	5	20667	569.66	57.78	430	830	33265	609.55	49.02	430	830
	6	20667	576.12	57.01	440	806	34795	617.46	52.40	440	870
	7	20543	593.65	60.17	450	826	35704	635.16	54.49	450	880
	8	20212	613.64	55.27	470	890	36514	652.00	53.13	470	890
Science	4	19832	475.07	46.13	300	725	32585	510.97	47.82	300	725
	8	20012	677.19	46.37	480	945	36473	708.34	48.24	480	945
Social Studies	4	19815	366.75	56.10	200	570	32577	406.24	52.07	200	570
	8	19957	574.17	49.98	420	780	36452	607.12	49.39	420	780
	10	14578	673.04	55.72	490	890	36855	705.98	53.30	490	890

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	6921	522.06	43.39	330	704	46009	554.35	45.60	330	900
	4	6831	543.56	48.08	340	743	45875	582.73	49.82	340	930
	5	6710	554.06	45.06	350	750	47300	598.66	46.99	350	940
	6	6538	559.29	47.94	360	766	48973	609.99	47.31	360	950
	7	6308	575.28	49.62	370	908	49987	631.68	52.59	370	960
	8	6455	574.05	53.20	380	825	50301	635.18	55.75	380	970
Mathematics	3	6907	513.11	64.66	360	760	45985	554.35	52.99	360	760
	4	6808	533.33	59.51	405	720	45850	577.17	49.92	405	800
	5	6683	550.53	63.98	430	830	47249	600.45	51.90	430	830
	6	6523	551.65	60.77	440	870	48939	608.78	53.87	440	870
	7	6305	565.22	63.09	450	796	49942	626.92	55.97	450	880
	8	6431	589.32	55.95	470	890	50295	644.60	53.95	470	890
Science	4	6766	466.36	48.30	300	656	45651	501.99	48.94	300	725
	8	6371	658.66	45.21	480	945	50114	702.22	48.25	480	945
Social Studies	4	6746	354.64	60.72	200	570	45646	396.72	54.31	200	570
	8	6366	549.95	50.66	420	780	50043	601.25	49.29	420	780
	10	5126	644.83	59.57	490	890	46307	702.38	52.54	490	890

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	4333	519.50	39.00	330	900	48597	552.86	46.25	330	900
	4	4146	542.42	41.54	354	735	48560	580.66	50.95	340	930
	5	3460	553.81	37.10	368	718	50550	595.81	48.57	350	940
	6	2988	558.02	40.85	360	738	52523	606.63	49.33	360	950
	7	2945	574.60	42.30	387	729	53350	628.16	54.48	370	960
	8	2863	578.03	46.35	380	707	53893	630.89	58.15	380	970
Mathematics	3	4364	513.13	54.86	360	760	48528	552.19	55.41	360	760
	4	4157	535.35	49.87	405	687	48501	574.60	52.47	405	800
	5	3488	551.31	56.01	430	830	50444	597.23	54.78	430	830
	6	3009	550.82	52.28	440	755	52453	605.00	56.65	440	870
	7	2956	567.47	55.61	450	722	53291	622.92	58.94	450	880
	8	2882	593.96	50.13	470	751	53844	640.71	56.31	470	890
Science	4	4112	458.82	38.84	300	688	48305	500.67	49.79	300	725
	8	2856	652.59	35.80	480	789	53629	699.69	49.38	480	945
Social Studies	4	4110	354.24	51.74	200	535	48282	394.46	56.25	200	570
	8	2847	548.77	42.68	420	711	53562	597.94	51.32	420	780
	10	1634	634.17	49.47	490	832	49799	698.69	55.01	490	890

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	75	516.01	51.84	379	635	52855	550.17	46.58	330	900
	4	66	549.00	49.58	373	657	52640	577.69	51.31	340	930
	5	69	576.96	51.06	468	677	53941	593.14	49.00	350	940
	6	76	578.66	54.18	433	669	55435	604.05	50.11	360	950
	7	70	612.94	67.75	511	908	56225	625.38	55.19	370	960
	8	87	595.66	68.59	434	763	56669	628.27	58.73	380	970
Mathematics	3	185	484.49	59.12	360	592	52707	549.19	56.26	360	760
	4	1631	507.39	51.87	405	648	51027	573.55	52.09	405	800
	5	2079	526.10	58.73	430	680	51853	597.00	54.14	430	830
	6	2318	530.76	53.09	440	681	53144	605.17	55.90	440	870
	7	2292	544.89	56.00	450	702	53955	623.19	58.11	450	880
	8	2224	575.01	51.73	470	890	54502	640.92	55.64	470	890
Science	4	23	446.00	37.75	403	537	52394	497.41	50.29	300	725
	8	18	674.56	58.57	561	796	56467	697.31	49.86	480	945
Social Studies	4	22	329.86	59.66	200	446	52370	391.33	56.93	200	570
	8	18	568.44	67.71	420	685	56391	595.47	52.04	420	780
	10	8	689.13	42.59	642	771	51425	696.64	56.00	490	890

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	27.71	37.74	29.48	5.08	34.56
4	340–545	546–591	592–649	650–930	27.17	32.71	32.67	7.45	40.12
5	350–563	564–609	610–669	670–940	27.67	34.80	32.41	5.12	37.52
6	360–571	572–621	622–670	671–950	25.18	36.36	30.65	7.80	38.45
7	370–584	585–637	638–696	697–960	23.08	33.99	34.12	8.81	42.92
8	380–591	592–651	652–707	708–970	26.05	38.29	28.02	7.64	35.66

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	23.78	31.23	34.46	10.54	44.99
4	405–535	536–587	588–632	633–800	22.32	36.61	30.84	10.23	41.07
5	430–573	574–610	611–657	658–830	29.46	28.94	32.12	9.47	41.59
6	440–581	582–625	626–687	688–870	32.61	31.82	30.90	4.66	35.57
7	450–605	606–646	647–711	712–880	34.99	30.17	31.38	3.47	34.84
8	470–619	620–666	667–717	718–890	32.47	37.53	23.62	6.38	30.00

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	16.18	32.67	33.03	18.13	51.16
8	480–652	653–694	695–736	737–945	18.56	29.97	30.65	20.82	51.47

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	200–362	363–395	396–435	436–570	27.63	23.18	29.38	19.81	49.19
8	420–562	563–598	599–639	640–780	24.10	27.49	31.01	17.40	48.41
10	490–669	670–702	703–740	741–890	28.23	24.56	27.12	20.09	47.21

Table 10-15 Percentage of Students in Each Performance Level by Subgroup, English Language Arts

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
3	BB	14666	27.7	25.5	29.8	19.2	64.6	45.8	33.7	49.7	32.0	25.3	54.3	53.3	23.9	43.7	17.6	58.7	27.7
	B	19974	37.7	37.3	38.2	39.1	26.8	37.4	37.0	35.2	37.3	37.9	35.8	32.6	38.5	37.5	37.9	25.3	37.8
	P	15602	29.5	31.3	27.7	35.3	7.9	15.2	24.2	13.5	27.0	31.3	9.3	12.4	32.0	17.1	37.3	13.3	29.5
	A	2688	5.1	5.9	4.3	6.4	0.7	1.6	5.1	1.6	3.7	5.5	0.6	1.7	5.6	1.6	7.3	2.7	5.1
	Total	52930	100.0	25886	27044	36547	4101	6787	2237	563	2695	48597	4333	6921	46009	20537	32392	75	52855
4	BB	14320	27.2	24.3	30.0	19.2	62.6	43.3	32.7	48.9	32.2	24.9	54.0	55.0	23.0	42.8	17.5	51.5	27.1
	B	17242	32.7	32.6	32.9	33.1	25.9	34.1	32.8	33.6	33.5	32.6	33.6	29.0	33.3	34.3	31.8	28.8	32.7
	P	17218	32.7	34.6	30.8	38.4	10.4	19.6	28.5	15.9	28.4	34.4	12.0	14.1	35.4	20.4	40.2	18.2	32.7
	A	3926	7.4	8.6	6.4	9.3	1.0	2.9	6.0	1.7	5.9	8.0	0.4	2.0	8.3	2.5	10.5	1.5	7.5
	Total	52706	100.0	25829	26877	36592	4079	6789	2166	599	2481	48560	4146	6831	45875	20068	32639	66	52640
5	BB	14946	27.7	23.8	31.4	20.3	61.6	42.8	30.4	49.2	32.7	25.4	60.9	62.3	22.8	43.3	17.9	37.7	27.7
	B	18798	34.8	35.1	34.5	35.7	27.0	34.9	35.0	37.6	34.0	35.0	32.4	26.2	36.0	35.4	34.4	36.2	34.8
	P	17502	32.4	35.0	29.9	37.8	10.5	20.7	28.2	11.6	28.6	34.2	6.6	10.6	35.5	19.5	40.4	24.6	32.4
	A	2764	5.1	6.0	4.2	6.3	0.8	1.6	6.3	1.5	4.7	5.5	0.1	1.0	5.7	1.8	7.2	1.4	5.1
	Total	54010	100.0	26516	27494	37720	4255	6900	2047	595	2493	50550	3460	6710	47300	20742	33268	69	53941

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 10-15 Percentage of Students in Each Performance Level by Subgroup, English Language Arts (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
6	BB	13980	25.2	21.2	29.0	18.9	56.1	39.0	21.2	41.4	31.4	23.1	61.6	62.8	20.2	39.9	16.4	43.4	25.2
	B	20186	36.4	36.7	36.1	36.4	31.4	38.6	38.3	39.2	35.0	36.5	33.4	27.2	37.6	38.0	35.4	36.8	36.4
	P	17016	30.7	33.3	28.1	35.1	11.4	19.9	31.2	16.8	27.4	32.1	4.9	8.6	33.6	19.1	37.5	19.7	30.7
	A	4329	7.8	8.8	6.8	9.6	1.1	2.5	9.3	2.6	6.2	8.2	0.2	1.3	8.7	3.0	10.7	0.0	7.8
	Total	55511	100.0	27169	28342	38850	4215	7117	2256	584	2489	52523	2988	6538	48973	20699	34812	76	55435
7	BB	12995	23.1	19.3	26.7	17.1	53.5	36.3	20.9	41.7	26.1	21.1	59.3	61.2	18.3	37.1	15.0	38.6	23.1
	B	19136	34.0	33.8	34.1	33.9	30.4	36.3	33.9	36.8	34.5	34.0	34.0	27.6	34.8	36.3	32.7	30.0	34.0
	P	19206	34.1	36.4	31.9	38.4	14.6	24.0	32.9	20.1	31.6	35.6	6.4	10.0	37.2	23.5	40.2	21.4	34.1
	A	4958	8.8	10.5	7.2	10.5	1.5	3.4	12.4	1.4	7.9	9.3	0.2	1.2	9.8	3.1	12.1	10.0	8.8
	Total	56295	100.0	27448	28847	39754	4251	7199	2129	568	2394	53350	2945	6308	49987	20598	35697	70	56225
8	BB	14786	26.1	21.9	30.0	20.3	56.3	38.0	21.1	45.7	31.3	24.2	60.3	65.5	21.0	40.1	18.3	51.7	26.0
	B	21730	38.3	38.0	38.5	38.7	31.3	39.6	40.8	37.5	37.4	38.5	35.0	26.4	39.8	39.0	37.9	26.4	38.3
	P	15903	28.0	30.8	25.4	31.9	10.9	18.9	28.4	14.6	24.8	29.3	4.7	7.0	30.7	17.9	33.6	17.2	28.0
	A	4337	7.6	9.3	6.1	9.1	1.4	3.5	9.8	2.2	6.4	8.0	0.0	1.1	8.5	3.0	10.2	4.6	7.6
	Total	56756	100.0	27533	29223	40254	4362	7112	2098	635	2295	53893	2863	6455	50301	20225	36531	87	56669

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 10-16 Percentage of Students in Each Performance Level by Subgroup, Mathematics

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
3	BB	12576	23.8	25.3	22.4	14.4	65.0	42.9	29.8	45.7	31.0	21.6	48.3	48.1	20.1	40.2	13.4	70.3	23.6
	B	16518	31.2	31.9	30.6	31.1	25.1	34.7	31.4	35.9	31.9	30.9	34.8	28.6	31.6	34.2	29.3	22.2	31.3
	P	18225	34.5	33.7	35.2	41.3	8.9	19.2	26.6	16.0	28.6	36.2	14.6	19.4	36.7	21.7	42.5	7.6	34.6
	A	5573	10.5	9.2	11.9	13.1	1.0	3.2	12.1	2.3	8.5	11.3	2.3	4.0	11.5	3.8	14.8	0.0	10.6
	Total	52892	100.0	25865	27027	36533	4058	6803	2241	562	2695	48528	4364	6907	45985	20480	32412	185	52707
4	BB	11754	22.3	23.3	21.4	13.6	62.8	39.1	27.0	45.7	28.5	20.2	47.5	49.8	18.2	38.2	12.6	70.3	20.8
	B	19278	36.6	38.8	34.5	36.5	29.4	40.5	36.4	39.0	39.7	36.4	39.4	32.2	37.3	40.0	34.5	25.4	37.0
	P	16240	30.8	29.2	32.4	37.1	7.2	17.1	24.6	11.6	24.6	32.5	11.9	14.4	33.3	18.6	38.3	3.8	31.7
	A	5386	10.2	8.7	11.7	12.8	0.7	3.3	11.9	3.7	7.2	11.0	1.3	3.6	11.2	3.2	14.5	0.4	10.5
	Total	52658	100.0	25805	26853	36579	4051	6794	2164	595	2475	48501	4157	6808	45850	20032	32626	1631	51027
5	BB	15889	29.5	30.2	28.8	20.3	70.5	49.0	29.9	55.4	38.2	27.2	62.8	62.1	24.8	47.4	18.3	81.6	27.4
	B	15610	28.9	30.7	27.2	29.6	20.5	30.0	31.2	29.3	27.7	29.1	27.0	22.4	29.9	29.6	28.5	14.1	29.5
	P	17324	32.1	31.3	32.9	38.3	8.0	18.2	26.7	13.2	27.0	33.7	9.5	13.0	34.8	19.7	39.8	4.0	33.2
	A	5109	9.5	7.8	11.1	11.8	1.0	2.8	12.2	2.0	7.1	10.1	0.7	2.4	10.5	3.2	13.4	0.2	9.8
	Total	53932	100.0	26488	27444	37709	4187	6902	2055	597	2482	50444	3488	6683	47249	20667	33265	2079	51853

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 10-16 Percentage of Students in Each Performance Level by Subgroup, Mathematics (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
6	BB	18088	32.6	32.8	32.4	23.9	71.3	52.5	30.8	59.8	41.9	30.3	72.6	69.5	27.7	50.8	21.8	84.9	30.3
	B	17647	31.8	33.2	30.5	33.1	22.1	31.2	30.7	27.7	31.9	32.3	22.9	20.9	33.3	31.2	32.2	13.3	32.6
	P	17140	30.9	30.0	31.8	37.3	6.4	15.3	30.0	11.6	22.5	32.4	4.4	8.7	33.9	16.8	39.3	1.7	32.2
	A	2587	4.7	4.0	5.3	5.7	0.2	1.0	8.5	0.9	3.7	4.9	0.1	1.0	5.2	1.1	6.8	0.0	4.9
Total		55462	100.0	27149	28313	38823	4182	7132	2252	584	2489	52453	3009	6523	48939	20667	34795	2318	53144
7	BB	19680	35.0	34.8	35.2	26.4	74.7	53.6	35.4	59.5	44.8	32.7	76.7	75.4	29.9	53.8	24.2	89.4	32.7
	B	16969	30.2	32.4	28.1	32.0	18.8	28.7	25.8	30.3	27.2	30.8	19.0	16.1	31.9	28.7	31.0	8.8	31.1
	P	17648	31.4	30.0	32.7	37.5	6.1	16.7	30.0	9.9	25.3	32.9	4.2	7.9	34.3	16.6	39.9	1.8	32.6
	A	1950	3.5	2.9	4.0	4.1	0.4	0.9	8.8	0.4	2.7	3.7	0.1	0.6	3.8	0.9	5.0	0.0	3.6
Total		56247	100.0	27415	28832	39723	4225	7203	2127	565	2404	53291	2956	6305	49942	20543	35704	2292	53955
8	BB	18418	32.5	30.6	34.2	24.3	71.5	51.1	26.9	57.9	41.7	30.6	68.0	72.0	27.4	50.6	22.4	83.2	30.4
	B	21288	37.5	39.7	35.5	39.8	22.5	34.8	38.8	30.8	36.0	38.0	28.3	22.3	39.5	35.3	38.8	15.2	38.4
	P	13400	23.6	24.2	23.1	28.4	5.4	11.9	21.5	9.4	16.9	24.7	3.4	4.8	26.0	12.1	30.0	1.5	24.5
	A	3620	6.4	5.5	7.2	7.5	0.7	2.2	12.7	1.9	5.4	6.7	0.2	0.9	7.1	2.0	8.8	0.1	6.6
Total		56726	100.0	27529	29197	40214	4351	7128	2103	636	2294	53844	2882	6431	50295	20212	36514	2224	54502

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

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
4	BB	8479	16.2	16.2	16.1	9.1	51.5	30.3	21.7	29.3	18.5	14.2	39.3	38.4	12.9	28.4	8.7	56.5	16.2
	B	17124	32.7	34.4	31.0	30.1	35.1	41.2	36.8	43.0	37.6	31.7	43.6	35.4	32.3	39.5	28.5	30.4	32.7
	P	17311	33.0	32.8	33.2	38.0	11.3	21.8	29.1	22.1	30.8	34.5	15.3	19.2	35.1	24.5	38.2	13.0	33.0
	A	9503	18.1	16.6	19.6	22.8	2.1	6.8	12.5	5.7	13.1	19.5	1.7	6.9	19.8	7.6	24.5	0.0	18.1
	Total	52417	100.0	25680	26737	36548	3941	6720	2143	598	2467	48305	4112	6766	45651	19832	32585	23	52394
8	BB	10483	18.6	17.2	19.8	12.5	51.8	31.5	15.6	35.0	22.5	16.9	50.1	49.3	14.6	31.0	11.7	38.9	18.6
	B	16927	30.0	31.6	28.4	28.2	31.9	36.2	33.2	38.8	32.3	29.5	38.0	31.7	29.7	35.5	26.9	22.2	30.0
	P	17313	30.7	32.1	29.2	34.3	12.9	22.8	29.2	19.2	28.2	31.7	10.6	13.3	32.9	23.3	34.7	27.8	30.7
	A	11762	20.8	19.0	22.5	25.0	3.5	9.6	22.0	7.0	17.0	21.9	1.2	5.7	22.7	10.3	26.6	11.1	20.8
	Total	56485	100.0	27407	29078	40179	4230	7082	2087	626	2281	53629	2856	6371	50114	20012	36473	18	56467

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

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		Economic Status		Accommodations	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
4	BB	14476	27.6	26.5	28.7	19.6	65.5	43.5	33.5	48.6	33.5	25.5	52.8	54.1	23.7	43.9	17.7	86.4	27.6
	B	12146	23.2	24.0	22.4	22.9	19.2	26.1	25.3	25.5	23.1	22.8	27.9	21.4	23.4	25.5	21.8	0.0	23.2
	P	15392	29.4	30.5	28.3	33.0	12.1	22.1	26.3	18.8	28.6	30.5	15.9	17.3	31.2	22.1	33.8	4.5	29.4
	A	10378	19.8	19.0	20.6	24.5	3.2	8.3	14.8	7.1	14.8	21.2	3.4	7.2	21.7	8.5	26.7	9.1	19.8
	Total	52392	100.0	25680	26712	36546	3926	6718	2143	595	2464	48282	4110	6746	45646	19815	32577	22	52370
8	BB	13595	24.1	22.0	26.1	17.5	58.6	38.1	22.4	43.0	28.6	22.2	60.6	61.2	19.4	39.0	15.9	44.4	24.1
	B	15505	27.5	28.5	26.5	26.9	24.2	31.3	28.1	32.3	29.6	27.4	29.4	23.1	28.0	30.4	25.9	16.7	27.5
	P	17494	31.0	32.7	29.4	34.8	13.5	22.3	29.8	19.0	27.6	32.2	9.0	11.7	33.5	22.6	35.6	33.3	31.0
	A	9815	17.4	16.8	18.0	20.7	3.8	8.3	19.7	5.6	14.3	18.3	1.0	4.0	19.1	8.0	22.5	5.6	17.4
	Total	56409	100.0	27354	29055	40138	4218	7061	2081	625	2286	53562	2847	6366	50043	19957	36452	18	56391
10	BB	14520	28.2	26.1	30.2	23.6	61.0	44.5	26.2	43.7	33.7	26.7	75.9	68.2	23.8	44.7	21.7	25.0	28.2
	B	12632	24.6	25.9	23.3	24.6	21.6	26.3	23.2	29.1	22.9	24.7	19.4	17.4	25.4	26.1	24.0	50.0	24.6
	P	13948	27.1	28.8	25.6	29.1	12.5	20.0	27.3	20.0	26.3	27.9	4.0	9.2	29.1	19.7	30.1	12.5	27.1
	A	10333	20.1	19.2	20.9	22.7	4.9	9.2	23.4	7.2	17.1	20.7	0.6	5.3	21.7	9.5	24.3	12.5	20.1
	Total	51433	100.0	24958	26475	39414	2421	5531	1854	526	1687	49799	1634	5126	46307	14578	36855	8	51425

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	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
				MC	CR					Mean	SD
3	52930	A	Reading—Key Ideas and Details	6	3	12	6.47	0.55	3.06	54.13	23.16
	52930	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	6	0	6	2.55	0.43	1.44	43.36	17.26
	52930	C	Reading—Vocabulary Use	2	1	4	2.12	0.53	1.18	53.74	21.39
	52930	D	Writing/Language—Text Types and Purposes	2	2	6	3.56	0.60	1.57	59.13	19.60
	52930	E	Writing/Language—Research	4	1	6	3.33	0.56	1.60	54.88	21.24
	52930	F	Writing/Language—Language Conventions	4	2	7	3.80	0.53	1.96	54.73	24.12
	52930	G	Listening	3	2	7	4.33	0.60	1.78	60.89	20.98
4	52706	A	Reading—Key Ideas and Details	2	5	12	6.02	0.52	2.65	50.72	19.59
	52706	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	6	0	6	3.17	0.53	1.65	52.70	22.34
	52706	C	Reading—Vocabulary Use	6	0	6	3.55	0.59	1.75	59.37	24.45
	52706	D	Writing/Language—Text Types and Purposes	4	2	6	3.41	0.57	1.68	56.69	22.37
	52706	E	Writing/Language—Research	3	2	6	3.34	0.55	1.62	55.18	21.76
	52706	F	Writing/Language—Language Conventions	3	2	7	4.52	0.62	1.49	64.25	15.69
	52706	G	Listening	4	2	8	3.79	0.52	1.92	47.50	18.39

Table 10-19 Summary Statistics for Content Standards Raw and SPI Scores, English Language Arts (cont.)

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
				MC	CR					Mean	SD
5	54010	A	Reading—Key Ideas and Details	6	2	10	5.63	0.56	2.51	56.21	22.40
	54010	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	7	1	8	4.70	0.59	2.20	59.58	24.12
	54010	C	Reading—Vocabulary Use	3	2	6	3.73	0.62	1.33	62.43	14.82
	54010	D	Writing/Language—Text Types and Purposes	5	1	6	3.96	0.66	1.43	65.04	18.46
	54010	E	Writing/Language—Research	1	3	6	3.44	0.60	1.70	58.45	22.93
	54010	F	Writing/Language—Language Conventions	3	2	7	4.47	0.64	1.65	63.18	17.52
	54010	G	Listening	4	2	8	4.93	0.63	2.04	61.07	20.87
6	55511	A	Reading—Key Ideas and Details	2	5	12	6.73	0.56	2.68	56.62	19.88
	55511	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	5	1	7	3.88	0.56	1.65	54.94	18.33
	55511	C	Reading—Vocabulary Use	3	2	5	2.97	0.60	1.44	58.79	23.15
	55511	D	Writing/Language—Text Types and Purposes	5	1	6	3.25	0.54	1.52	55.03	19.44
	55511	E	Writing/Language—Research	1	3	6	3.12	0.49	1.53	52.65	18.60
	55511	F	Writing/Language—Language Conventions	3	2	7	3.98	0.54	1.58	56.86	16.74
	55511	G	Listening	4	2	8	5.00	0.62	1.97	62.13	20.25

Table 10-19 Summary Statistics for Content Standards Raw and SPI Scores, English Language Arts (cont.)

Grade	N	Content Standard	Standard	No. of Items MC	No. of Items CR	Total Score Points	Mean	Mean p-Value	SD	SPI Mean	SPI SD
7	56295	A	Reading—Key Ideas and Details	4	3	9	4.81	0.52	2.28	53.72	22.77
	56295	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	4	3	10	4.97	0.50	2.47	49.21	20.46
	56295	C	Reading—Vocabulary Use	1	3	5	3.40	0.67	1.37	67.92	22.44
	56295	D	Writing/Language—Text Types and Purposes	2	2	6	3.14	0.53	1.42	51.88	16.35
	56295	E	Writing/Language—Research	3	2	7	4.07	0.59	1.65	59.11	18.92
	56295	F	Writing/Language—Language Conventions	4	1	6	3.81	0.64	1.41	63.60	17.21
	56295	G	Listening	2	3	8	4.88	0.60	2.02	60.51	20.59
8	56756	A	Reading—Key Ideas and Details	5	3	11	6.52	0.59	2.78	59.26	22.94
	56756	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	8	0	8	4.33	0.55	1.92	54.41	19.67
	56756	C	Reading—Vocabulary Use	3	1	5	2.90	0.58	1.49	56.95	24.19
	56756	D	Writing/Language—Text Types and Purposes	4	1	6	3.76	0.63	1.65	63.57	21.52
	56756	E	Writing/Language—Research	5	1	7	4.00	0.56	1.80	57.62	20.90
	56756	F	Writing/Language—Language Conventions	2	2	6	3.80	0.63	1.27	64.40	14.99
	56756	G	Listening	2	3	8	4.78	0.60	2.05	59.33	20.51

Table 10-20 Summary Statistics for Domain Raw and SPI Scores, English Language Arts

Grade	N	Domain	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
			MC	CR					Mean	SD
3	52930	Listening	3	2	7	4.33	0.60	1.78	60.89	20.98
	52930	Reading	14	4	22	11.24	0.50	4.76	50.98	20.56
	52930	Writing	10	5	19	10.77	0.56	4.24	56.25	21.21
4	52706	Listening	4	2	8	3.79	0.52	1.92	47.50	18.39
	52706	Reading	14	5	24	12.84	0.55	5.22	53.27	21.08
	52706	Writing	10	6	19	11.33	0.58	3.90	59.07	19.14
5	54010	Listening	4	2	8	4.93	0.63	2.04	61.07	20.87
	54010	Reading	16	5	24	14.15	0.59	5.15	58.80	20.63
	54010	Writing	9	6	19	11.91	0.64	3.84	62.41	18.88
6	55511	Listening	4	2	8	5.00	0.62	1.97	62.13	20.25
	55511	Reading	10	8	24	13.66	0.57	4.97	56.61	19.87
	55511	Writing	9	6	19	10.39	0.53	3.69	54.89	17.77
7	56295	Listening	2	3	8	4.88	0.60	2.02	60.51	20.59
	56295	Reading	9	9	24	13.24	0.55	5.31	54.81	21.27
	56295	Writing	9	5	19	11.06	0.59	3.52	58.27	16.91
8	56756	Listening	2	3	8	4.78	0.60	2.05	59.33	20.51
	56756	Reading	16	4	24	13.80	0.57	5.40	57.20	21.83
	56756	Writing	11	4	19	11.58	0.60	3.85	61.54	18.69

Table 10-21 Summary Statistics for Content Standards Raw and SPI Scores, Mathematics

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-Value	SD	SPI	
				MC	CR					Mean	SD
3	52892	A	Operations and Algebraic Thinking	5	4	9	4.60	0.51	2.39	51.07	23.83
	52892	B	Number and Operations in Base Ten	5	3	8	4.72	0.59	2.38	58.48	27.42
	52892	C	Number and Operations— Fractions	4	4	8	4.04	0.51	2.12	50.56	23.01
	52892	D	Measurement and Data	5	5	10	5.05	0.51	2.63	50.92	23.95
	52892	E	Geometry	4	3	7	3.93	0.56	2.06	56.67	25.55
4	52658	A	Operations and Algebraic Thinking	8	2	10	4.49	0.45	2.02	44.91	17.09
	52658	B	Number and Operations in Base Ten	5	4	9	4.29	0.48	2.48	47.79	24.75
	52658	C	Number and Operations— Fractions	7	3	10	4.39	0.44	2.87	43.70	26.07
	52658	D	Measurement and Data	8	2	10	4.73	0.47	2.33	46.94	20.95
	52658	E	Geometry	4	3	7	3.25	0.47	1.92	47.62	21.42
5	53932	A	Operations and Algebraic Thinking	4	5	9	3.84	0.43	2.31	43.41	23.04
	53932	B	Number and Operations in Base Ten	4	5	9	4.00	0.45	2.46	45.02	24.31
	53932	C	Number and Operations— Fractions	7	2	9	3.70	0.41	2.27	40.80	22.07
	53932	D	Measurement and Data	7	3	10	4.12	0.41	2.47	40.33	21.27
	53932	E	Geometry	5	4	9	3.66	0.41	2.37	41.10	23.13

Table 10-21 Summary Statistics for Content Standards Raw and SPI Scores, Mathematics (cont.)

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
				MC	CR					Mean	SD
6	55462	E	Geometry	5	2	7	2.43	0.35	1.77	35.26	20.35
	55462	F	Ratios and Proportional Relationships	3	4	7	2.91	0.42	1.78	41.54	22.26
	55462	G	The Number System	7	3	10	4.81	0.48	2.72	48.16	24.65
	55462	H	Expressions and Equations	8	3	11	4.74	0.43	2.82	43.05	22.92
	55462	I	Statistics and Probability	8	3	11	4.94	0.45	2.23	44.97	17.04
7	56247	E	Geometry	7	3	10	3.67	0.37	1.88	36.21	14.46
	56247	F	Ratios and Proportional Relationships	6	2	8	3.97	0.50	2.23	49.36	25.01
	56247	G	The Number System	4	3	7	2.57	0.37	1.93	37.73	23.79
	56247	H	Expressions and Equations	7	3	10	3.95	0.40	2.33	39.41	19.90
	56247	I	Statistics and Probability	7	4	11	4.98	0.46	2.65	44.74	21.26
8	56726	E	Geometry	7	3	10	4.16	0.42	2.50	41.60	21.71
	56726	G	The Number System	5	3	8	2.91	0.36	2.20	37.98	23.41
	56726	H	Expressions and Equations	7	3	10	3.91	0.39	2.44	39.89	21.53
	56726	I	Statistics and Probability	6	2	8	3.39	0.43	1.84	41.54	18.21
	56726	J	Functions	7	3	10	4.33	0.44	2.44	42.50	21.89

Table 10-22 Summary Statistics for Content Standards Raw and SPI Scores, Science

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
				MC	CR					Mean	SD
4	52417	A	Life Science	2	10	12	6.69	0.56	2.67	55.66	20.20
	52417	B	Physical Science	4	8	12	7.15	0.60	2.65	59.90	19.79
	52417	C	Earth and Space Science	3	5	8	3.83	0.48	1.94	47.79	19.64
	52417	D	Engineering	1	7	8	4.30	0.54	2.24	53.96	24.79
8	56485	A	Life Science	1	10	11	5.38	0.49	2.88	49.37	23.93
	56485	B	Physical Science	2	8	10	4.62	0.46	2.43	45.91	21.67
	56485	C	Earth and Space Science	2	8	10	5.50	0.55	2.25	55.16	19.36
	56485	D	Engineering	1	8	9	5.69	0.64	2.25	62.77	22.32

Table 10-23 Summary Statistics for Content Standards Raw and SPI Scores, Social Studies

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean p-Value	SD	SPI	
				MC	CR					Mean	SD
4	52392	A	Geography	8	1	9	5.35	0.60	2.22	60.46	20.61
	52392	B	History	8	1	9	6.34	0.71	2.29	70.41	23.09
	52392	C	Political Science and Citizenship	6	0	6	3.58	0.60	1.57	59.89	20.65
	52392	D	Economics	5	1	6	3.41	0.57	1.61	56.68	22.24
	52392	E	The Behavioral Sciences	7	1	8	5.69	0.71	2.04	71.06	23.10
8	56409	A	Geography	9	1	10	6.93	0.70	2.47	69.19	22.07
	56409	B	History	11	1	12	8.11	0.68	2.84	67.86	21.82
	56409	C	Political Science and Citizenship	5	1	6	3.63	0.61	1.62	60.65	21.16
	56409	D	Economics	4	2	6	3.56	0.60	1.67	59.15	22.93
	56409	E	The Behavioral Sciences	6	0	6	3.89	0.65	1.49	65.61	20.25
10	51433	A	Geography	10	0	10	6.71	0.68	2.26	66.90	19.47
	51433	B	History	13	0	13	8.43	0.65	3.15	65.17	22.37
	51433	C	Political Science and Citizenship	9	1	10	5.82	0.59	2.48	58.12	21.84
	51433	D	Economics	7	1	8	4.94	0.62	1.99	62.24	20.90
	51433	E	The Behavioral Sciences	7	2	9	5.58	0.63	2.10	61.63	19.58

Table 10-24 SPI Cut Scores, English Language Arts

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	36	0	37	0	38
	2	37	66	38	56	39	64
	3	67	87	57	77	65	90
	4	88	100	78	100	91	100
Reading—Craft & Structure	1	0	30	0	33	0	42
	2	31	45	34	57	43	71
	3	46	75	58	85	72	93
	4	76	100	86	100	94	100
Reading—Vocabulary Use	1	0	39	0	39	0	54
	2	40	62	40	68	55	67
	3	63	87	69	92	68	81
	4	88	100	93	100	82	100
Writing/Language—Text Types and Purposes	1	0	47	0	40	0	53
	2	48	68	41	64	54	72
	3	69	86	65	86	73	89
	4	87	100	87	100	90	100
Writing/Language—Research	1	0	39	0	38	0	43
	2	40	65	39	60	44	69
	3	66	85	61	86	70	89
	4	86	100	87	100	90	100
Writing/Language—Language Conventions	1	0	36	0	54	0	52
	2	37	67	55	69	53	70
	3	68	89	70	84	71	86
	4	90	100	85	100	87	100
Listening	1	0	47	0	34	0	46
	2	48	71	35	51	47	69
	3	72	90	52	72	70	89
	4	91	100	73	100	90	100
Reading	1	0	35	0	36	0	43
	2	36	60	37	59	44	67
	3	61	84	60	83	68	89
	4	85	100	84	100	90	100
Writing	1	0	40	0	44	0	49
	2	41	67	45	64	50	70
	3	68	87	65	85	71	88
	4	88	100	86	100	89	100

Table 10-24 SPI Cut Scores, English Language Arts (cont.)

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	41	0	33	0	42
	2	42	64	34	58	43	70
	3	65	83	59	84	71	89
	4	84	100	85	100	90	100
Reading—Craft & Structure	1	0	39	0	30	0	39
	2	40	62	31	50	40	63
	3	63	78	51	77	64	80
	4	79	100	78	100	81	100
Reading— Vocabulary Use	1	0	41	0	50	0	37
	2	42	68	51	77	38	69
	3	69	88	78	92	70	86
	4	89	100	93	100	87	100
Writing/Language— Text Types and Purposes	1	0	39	0	38	0	49
	2	40	61	39	54	50	75
	3	62	81	55	72	76	90
	4	82	100	73	100	91	100
Writing/Language— Research	1	0	39	0	43	0	41
	2	40	58	44	65	42	67
	3	59	76	66	82	68	85
	4	77	100	83	100	86	100
Writing/Language— Language Conventions	1	0	44	0	50	0	56
	2	45	61	51	68	57	70
	3	62	79	69	83	71	82
	4	80	100	84	100	83	100
Listening	1	0	48	0	45	0	44
	2	49	71	46	67	45	68
	3	72	85	68	84	69	85
	4	86	100	85	100	86	100
Reading	1	0	40	0	36	0	40
	2	41	64	37	59	41	67
	3	65	82	60	83	68	85
	4	83	100	84	100	86	100
Writing	1	0	41	0	44	0	48
	2	42	60	45	62	49	70
	3	61	79	63	79	71	86
	4	80	100	80	100	87	100

Table 10-25 SPI Cut Scores, Mathematics

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	30	0	30	0	28
	2	31	53	31	47	29	48
	3	54	82	48	66	49	75
	4	83	100	67	100	76	100
Number and Operations in Base Ten	1	0	32	0	24	0	27
	2	33	67	25	52	28	50
	3	68	91	53	82	51	79
	4	92	100	83	100	80	100
Number and Operations— Fractions	1	0	30	0	18	0	23
	2	31	51	19	47	24	41
	3	52	81	48	80	42	73
	4	82	100	81	100	74	100
Measurement and Data	1	0	30	0	27	0	23
	2	31	55	28	49	24	41
	3	56	81	50	76	42	70
	4	82	100	77	100	71	100
Geometry	1	0	33	0	29	0	24
	2	34	63	30	51	25	43
	3	64	88	52	77	44	75
	4	89	100	78	100	76	100

Table 10-25 SPI Cut Scores, Mathematics (cont.)

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	20	0	27	0	28
	2	21	34	28	36	29	50
	3	35	78	37	68	51	79
	4	79	100	69	100	80	100
Ratios and Proportional Relationships*	1	0	28	0	35		
	2	29	47	36	62		
	3	48	82	63	89		
	4	83	100	90	100		
The Number System	1	0	32	0	21	0	21
	2	33	57	22	45	22	49
	3	58	90	46	84	50	79
	4	91	100	85	100	80	100
Expressions and Equations	1	0	26	0	26	0	24
	2	27	49	27	43	25	46
	3	50	85	44	82	47	80
	4	86	100	83	100	81	100
Statistics and Probability	1	0	35	0	32	0	30
	2	36	50	33	52	31	47
	3	51	74	53	85	48	72
	4	75	100	86	100	73	100
Functions**	1					0	28
	2					29	51
	3					52	80
	4					81	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	33	0	24
	2	34	55	25	47
	3	56	75	48	72
	4	76	100	73	100
Physical Science	1	0	38	0	23
	2	39	60	24	42
	3	61	79	43	64
	4	80	100	65	100
Earth and Space Science	1	0	26	0	35
	2	27	44	36	54
	3	45	67	55	72
	4	68	100	73	100
Engineering	1	0	25	0	40
	2	26	54	41	64
	3	55	79	65	83
	4	80	100	84	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	47	0	53	0	55
	2	48	62	54	75	56	70
	3	63	79	76	89	71	84
	4	80	100	90	100	85	100
History	1	0	54	0	51	0	49
	2	55	76	52	73	50	70
	3	77	92	74	88	71	87
	4	93	100	89	100	88	100
Political Science and Citizenship	1	0	46	0	42	0	42
	2	47	62	43	62	43	59
	3	63	78	63	81	60	78
	4	79	100	82	100	79	100
Economics	1	0	41	0	40	0	49
	2	42	58	41	61	50	66
	3	59	76	62	81	67	81
	4	77	100	82	100	82	100
The Behavioral Sciences	1	0	58	0	49	0	49
	2	59	78	50	68	50	64
	3	79	91	69	86	65	79
	4	92	100	87	100	80	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
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
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
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
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
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

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
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
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
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
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
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

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
8	2019	65355	63062	96.49	699.70	50.55
	2021	67572	56485	83.59	697.31	49.86

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 Longitudinal Comparison of State-Level Participation Rates Scale Score Means, Social Studies

Grade	Year	Enrolled	Number Tested	Percent Tested	Scale Score Mean	Scale Score SD
4	2016	64361	62630	97.31	398.02	51.49
	2017	66001	64512	97.74	397.05	51.71
	2018	65885	64456	97.83	398.23	53.72
	2019	65222	63603	97.52	396.68	55.69
	2021	61127	52392	85.71	391.30	56.94
8	2016	63861	61496	96.30	598.06	51.68
	2017	64265	62079	96.60	597.60	54.26
	2018	65638	63230	96.33	599.17	53.25
	2019	65355	63045	96.47	598.81	52.30
	2021	67572	56409	83.48	595.46	52.04
10	2016	67935	63991	94.19	698.51	53.74
	2017	67765	63764	94.10	696.92	56.56
	2018	66894	62630	93.63	695.70	58.24
	2019	67852	63476	93.55	692.82	58.30
	2021	68221	51433	75.39	696.64	56.00

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 10-32 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
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
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
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
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
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

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
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
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
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
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
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

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
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

Note: New cut scores were used to classify students into performance levels after the Spring 2019 Science test administration.

Table 10-35 Longitudinal Comparison of State-Level Impact Data, Social Studies

Grade	Year	N	Below Basic	Basic	Proficient	Advanced	Prof. & Adv.
4	2016	62630	22.55	24.52	32.26	20.66	52.93
	2017	64512	23.02	24.93	31.84	20.20	52.04
	2018	64456	22.14	24.20	31.69	21.97	53.66
	2019	63603	24.04	23.47	30.38	22.11	52.49
	2021	52392	27.63	23.18	29.38	19.81	49.19
8	2016	61496	22.74	27.47	30.82	18.96	49.78
	2017	62079	23.47	26.50	31.04	18.98	50.03
	2018	63230	22.84	24.95	31.85	20.36	52.21
	2019	63045	22.24	26.16	32.25	19.35	51.60
	2021	56409	24.10	27.49	31.01	17.40	48.41
10	2016	63991	26.32	25.18	28.80	19.70	48.50
	2017	63764	27.72	24.12	27.83	20.33	48.17
	2018	62630	28.19	23.61	28.01	20.18	48.20
	2019	63476	30.97	23.53	26.12	19.38	45.50
	2021	51433	28.23	24.56	27.12	20.09	47.21

Part 11: Summary and Recommendations

Results and key findings of the Spring 2021 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 2020 Wisconsin Forward Exam administration was cancelled, and the 2021 test administration was the fifth administration of the Forward assessment. Since Spring 2016, the assessment results were reported on the same scales and students were classified into the performance levels using the same cut scores, allowing for longitudinal tracking of student performance in ELA, Mathematics, and Social Studies. New test scales were established and new performance level cut scores were set for Science assessments after the Spring 2019 test administration. The Spring 2019 assessment results serve as the new baseline for monitoring student performance in Science across years. The 2021 Wisconsin Forward Exam administration was the second administration of the Science assessments measuring the new Science standards and reported on the new scales.

In keeping with the field-testing of new test items in all content areas in the Spring 2016 through 2019 test administrations and the field-testing of new Social Studies items in the Spring 2021 test administration, 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.

While text-dependent analysis (TDA) items were not included in the Spring 2021 ELA tests, these items will again be included in the assessments planned for Spring 2022 and beyond. DRC recommends continuing to use an artificial intelligence (AI) engine in the scoring of TDA items for its efficiency and accuracy.

The Spring 2021 test forms were pre-equated, and students were scored using Spring 2019 (or older) item parameters. This approach to student scoring was appropriate given an unknown impact of the COVID-19 pandemic on student learning in the 2020–21 school year. While the post-equating verification process yielded comparable pre- and post-equated item parameters and resulting impact data, the Wisconsin Technical Advisory Committee did not recommend using the Spring 2021 post-equated parameters in future form linking. DRC will follow this recommendation and will use pre-pandemic item parameters in linking the 2022 test forms to the existing scales.

The test forms originally developed for the Spring 2020 test administration of ELA, Mathematics, and Science will be used, with possible small modifications, in Spring 2022 testing. These forms will be post-equated in Spring 2022, and the pre-pandemic item parameters for the anchor items included in these forms will be used to link the ELA, Mathematics, and Science tests to their respective operational scales.

New assessments measuring new standards will be developed for Social Studies grades 4, 8, and 10. New reporting scales and new performance level cut scores will be set for the Social Studies tests in Spring 2022.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

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 [Computer program]. Unpublished.

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 Center for Education Statistics (2021). School District Characteristic 2018-2019 [Data file]. Retrieved from https://nces.ed.gov/ccd/districtsearch/

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.

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.

Appendix A

Spring 2021 English Language Arts Operational Test Maps

Table A-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	MC	OP	1	2	3.W.1.c	Writing
3	1	2	TE	OP	2	3	3.W.1.b	Writing
3	1	3	MC	OP	1	2	3.W.8	Writing
3	1	4	MC	OP	1	2	3.W.8	Writing
3	1	5	TE	OP	2	2	3.W.1.b	Writing
3	1	6	MC	OP	1	2	3.W.2.a	Writing
3	1	7	TE	OP	2	2	3.W.8	Writing
3	1	8	MC	OP	1	2	3.W.8	Writing
3	1	9	MC	OP	1	3	3.W.8	Writing
3	1	10	MC	OP	1	1	3.L.1.i	Writing
3	1	11	TE	OP	1	1	3.L.2.c	Writing
3	1	12	MC	OP	1	1	3.L.1.f	Writing
3	1	13	MC	OP	1	1	3.L.1.d	Writing
3	1	14	MC	OP	1	2	3.L.1.d	Writing
3	1	15	TE	OP	2	2	3.L.1.h	Writing
3	2	16	MC	OP	1	1	3.SL.3	Listening
3	2	17	TE	OP	2	2	3.SL.3	Listening
3	2	18	MC	OP	1	2	3.SL.2	Listening
3	2	19	MS	OP	2	2	3.SL.3	Listening
3	2	20	MC	OP	1	2	3.SL.2	Listening
3	3	21	MC	OP	1	3	3.RL.7	Reading
3	3	22	MC	OP	1	2	3.RL.4	Reading
3	3	23	MC	OP	1	3	3.RL.5	Reading
3	3	24	MC	OP	1	2	3.RL.1	Reading
3	3	25	MC	OP	1	3	3.RL.6	Reading
3	3	26	TE	OP	2	2	3.RL.2	Reading
3	3	27	MC	OP	1	2	3.RL.1	Reading
3	3	28	MS	OP	2	2	3.L.4	Reading
3	3	29	MC	OP	1	3	3.RL.3	Reading
3	3	30	MC	OP	1	2	3.RL.3	Reading
3	3	31	MC	OP	1	2	3.RL.5	Reading
3	3	32	EBSR	OP	2	3	3.RL.2	Reading

Table A-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	3	33	MC	OP	1	2	3.L.4.RI	Reading
3	3	34	MC	OP	1	2	3.RI.5	Reading
3	3	35	MC	OP	1	2	3.RI.1	Reading
3	3	36	MC	OP	1	2	3.RI.1	Reading
3	3	37	MC	OP	1	2	3.RI.7	Reading
3	3	38	TE	OP	2	2	3.RI.1	Reading

Table A-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	TE	OP	1	2	4.W.1.b	Writing
4	1	2	TE	OP	1	3	4.W.3.e	Writing
4	1	3	MC	OP	1	2	4.W.1.d	Writing
4	1	4	TE	OP	1	2	4.W.8	Writing
4	1	5	MC	OP	1	2	4.W.1.a	Writing
4	1	6	MC	OP	1	3	4.W.2.a	Writing
4	1	7	MC	OP	1	2	4.W.2.a	Writing
4	1	8	TE	OP	2	2	4.W.8	Writing
4	1	9	MC	OP	1	2	4.W.8	Writing
4	1	10	MC	OP	1	3	4.W.8	Writing
4	1	11	MC	OP	1	2	4.W.8	Writing
4	1	12	MC	OP	1	2	4.L.2.b	Writing
4	1	13	TE	OP	2	2	4.L.1.b	Writing
4	1	14	MC	OP	1	1	4.L.2.a	Writing
4	1	15	TE	OP	2	2	4.L.1.c	Writing
4	1	16	MC	OP	1	1	4.L.2.a	Writing
4	2	17	MC	OP	1	2	4.SL.3	Listening
4	2	18	MC	OP	1	1	4.SL.2	Listening
4	2	19	EBSR	OP	2	2	4.SL.3	Listening
4	2	20	MC	OP	1	2	4.SL.3	Listening
4	2	21	MC	OP	1	2	4.SL.3	Listening
4	2	22	EBSR	OP	2	3	4.SL.2	Listening
4	3	23	MC	OP	1	2	4.L.4	Reading
4	3	24	TE	OP	2	2	4.RI.2	Reading
4	3	25	MC	OP	1	2	4.RI.7	Reading
4	3	26	MC	OP	1	2	4.L.4	Reading
4	3	27	MS	OP	2	2	4.RI.1	Reading
4	3	28	EBSR	OP	2	3	4.RL.3	Reading
4	3	29	MC	OP	1	2	4.L.5	Reading
4	3	30	MC	OP	1	2	4.L.4	Reading
4	3	31	MC	OP	1	2	4.RL.5	Reading

Table A-2. English Language Arts, Grade 4 Test Map (cont.)

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	3	32	MC	OP	1	2	4.RI.5	Reading
4	3	33	MC	OP	1	2	4.L.5.RI	Reading
4	3	34	TE	OP	2	2	4.RI.3	Reading
4	3	35	MC	OP	1	2	4.RI.7	Reading
4	3	36	MC	OP	1	2	4.RI.8	Reading
4	3	37	MC	OP	1	2	4.RL.2	Reading
4	3	38	MC	OP	1	2	4.RL.3	Reading
4	3	39	MC	OP	1	1	4.L.5.RL	Reading
4	3	40	TE	OP	2	2	4.RL.1	Reading
4	3	41	MC	OP	1	2	4.RL.6	Reading

Table A-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	MC	OP	1	1	5.L.2.b	Writing
5	1	2	MC	OP	1	2	5.W.3.e	Writing
5	1	3	TE	OP	2	2	5.W.8	Writing
5	1	4	MC	OP	1	1	5.W.2.d	Writing
5	1	5	TE	OP	1	1	5.W.1.c	Writing
5	1	6	MC	OP	1	2	5.W.1.d	Writing
5	1	7	MC	OP	1	2	5.W.3.e	Writing
5	1	8	TE	OP	2	2	5.W.8	Writing
5	1	9	MC	OP	1	2	5.W.1.d	Writing
5	1	10	TE	OP	1	2	5.W.8	Writing
5	1	11	MC	OP	1	2	5.W.8	Writing
5	1	12	MC	OP	1	1	5.L.2.c	Writing
5	1	13	TE	OP	2	2	5.L.1.b	Writing
5	1	14	MC	OP	1	2	5.L.3.a	Writing
5	1	15	TE	OP	2	1	5.L.2.b	Writing
5	2	16	TE	OP	2	3	5.SL.3	Listening
5	2	17	MC	OP	1	2	5.SL.2	Listening
5	2	18	MC	OP	1	2	5.SL.3	Listening
5	2	19	MC	OP	1	2	5.SL.3	Listening
5	2	20	MC	OP	1	2	5.SL.2	Listening
5	2	21	EBSR	OP	2	3	5.SL.3	Listening
5	3	22	MC	OP	1	2	5.RL.5	Reading
5	3	23	MC	OP	1	2	5.RL.4	Reading
5	3	24	MC	OP	1	2	5.RL.6	Reading
5	3	25	TE	OP	1	2	5.L.4	Reading
5	3	26	MC	OP	1	2	5.RI.4	Reading
5	3	27	TE	OP	1	2	5.RI.8	Reading
5	3	28	MC	OP	1	2	5.RI.5	Reading
5	3	29	MC	OP	1	2	5.RI.2	Reading

Table A-3. English Language Arts, Grade 5 Test Map (cont.)

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	3	30	MC	OP	1	2	5.L.5	Reading
5	3	31	MS	OP	2	2	5.RL.2	Reading
5	3	32	MC	OP	1	2	5.RL.6	Reading
5	3	33	MC	OP	1	2	5.RL.9	Reading
5	3	34	MC	OP	1	2	5.RL.2	Reading
5	3	35	MC	OP	1	3	5.RL.6	Reading
5	3	36	MC	OP	1	2	5.RL.5	Reading
5	3	37	MC	OP	1	2	5.RL.2	Reading
5	3	38	TE	OP	2	2	5.RL.3	Reading
5	3	39	TE	OP	2	2	5.L.4.RI	Reading
5	3	40	MC	OP	1	2	5.RI.1	Reading
5	3	41	MC	OP	1	1	5.RI.3	Reading
5	3	42	MC	OP	1	2	5.RI.1	Reading

Table A-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	MC	OP	1	2	6.W.2.a	Writing
6	1	2	MC	OP	1	2	6.W.3.b	Writing
6	1	3	MC	OP	1	2	6.W.1	Writing
6	1	4	MC	OP	1	2	6.W.8	Writing
6	1	5	MC	OP	1	1	6.W.3.b	Writing
6	1	6	MC	OP	1	2	6.W.2.b	Writing
6	1	7	TE	OP	1	2	6.W.3.b	Writing
6	1	8	TE	OP	1	2	6.W.8	Writing
6	1	9	TE	OP	2	2	6.L.1.d	Writing
6	1	10	MC	OP	1	2	6.L.2.a	Writing
6	1	11	TE	OP	2	2	6.W.7	Writing
6	1	12	TE	OP	2	1	6.L.2.b	Writing
6	1	13	TE	OP	2	2	6.W.8	Writing
6	1	14	MC	OP	1	1	6.L.1d	Writing
6	1	15	MC	OP	1	1	6.L.2.a	Writing
6	2	16	MC	OP	1	2	6.SL.2	Listening
6	2	17	MC	OP	1	2	6.SL.2	Listening
6	2	18	TE	OP	2	1	6.SL.2	Listening
6	2	19	EBSR	OP	2	3	6.SL.3	Listening
6	2	20	MC	OP	1	2	6.SL.2	Listening
6	2	21	MC	OP	1	2	6.SL.2	Listening
6	3	22	MC	OP	1	3	6.RL.5	Reading
6	3	23	TE	OP	2	2	6.RL.1	Reading
6	3	24	MC	OP	1	2	6.RL.4	Reading
6	3	25	TE	OP	2	2	6.RL.1	Reading
6	3	26	MC	OP	1	2	6.RI.8	Reading
6	3	27	MC	OP	1	2	6.L.4.RI	Reading
6	3	28	TE	OP	2	2	6.RI.1	Reading
6	3	29	MC	OP	1	3	6.RI.7	Reading
6	3	30	TE	OP	2	2	6.RI.2	Reading
6	3	31	TE	OP	1	2	6.L.4	Reading

Table A-4. English Language Arts, Grade 6 Test Map (cont.)

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
6	3	32	MC	OP	1	2	6.RI.3	Reading
6	3	33	MC	OP	1	2	6.RI.6	Reading
6	3	34	TE	OP	2	3	6.RI.8	Reading
6	3	35	EBSR	OP	2	3	6.RL.3	Reading
6	3	36	TE	OP	1	2	6.RL.4	Reading
6	3	37	MC	OP	1	2	6.RL.4	Reading
6	3	38	MC	OP	1	2	6.RL.2	Reading
6	3	39	MC	OP	1	2	6.RL.6	Reading

Table A-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	MC	OP	1	3	7.W.2.a	Writing
7	1	2	TE	OP	2	2	7.W.3.c	Writing
7	1	3	TE	OP	2	2	7.W.7	Writing
7	1	4	MC	OP	1	2	7.L.2.a	Writing
7	1	5	MS	OP	2	2	7.W.2.b	Writing
7	1	6	MC	OP	1	2	7.W.3.d	Writing
7	1	7	TE	OP	2	2	7.L.3.a	Writing
7	1	8	MC	OP	1	2	7.W.8	Writing
7	1	9	MC	OP	1	2	7.L.2.a	Writing
7	1	10	MC	OP	1	2	7.W.8	Writing
7	1	11	TE	OP	2	3	7.W.8	Writing
7	1	12	MC	OP	1	3	7.L.1.b	Writing
7	1	13	MC	OP	1	2	7.L.1.a	Writing
7	1	14	MC	OP	1	2	7.W.8	Writing
7	2	15	EBSR	OP	2	3	7.SL.2	Listening
7	2	16	MS	OP	2	3	7.SL.3	Listening
7	2	17	MC	OP	1	2	7.SL.2	Listening
7	2	18	MC	OP	1	2	7.SL.2	Listening
7	2	19	EBSR	OP	2	3	7.SL.3	Listening
7	3	20	TE	OP	1	2	7.RL.4	Reading
7	3	21	MC	OP	1	2	7.RL.3	Reading
7	3	22	EBSR	OP	2	2	7.RL.6	Reading
7	3	23	MC	OP	1	3	7.RL.2	Reading
7	3	24	MC	OP	1	2	7.RI.1	Reading
7	3	25	TE	OP	1	2	7.RI.4	Reading
7	3	26	TE	OP	2	2	7.RI.1	Reading
7	3	27	TE	OP	2	2	7.RI.5	Reading
7	3	28	MC	OP	1	2	7.RI.5	Reading
7	3	29	EBSR	OP	2	3	7.RI.8	Reading

Table A-5. English Language Arts, Grade 7 Test Map (cont.)

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
7	3	30	TE	OP	1	2	7.RL.1	Reading
7	3	31	MC	OP	1	2	7.RL.4	Reading
7	3	32	MC	OP	1	2	7.RL.2	Reading
7	3	33	MC	OP	1	3	7.RL.9	Reading
7	3	34	MC	OP	1	3	7.RI.6	Reading
7	3	35	MS	OP	2	2	7.RI.2	Reading
7	3	36	MS	OP	2	2	7.RI.4	Reading
7	3	37	MC	OP	1	3	7.RI.8	Reading

Table A-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	MC	OP	1	2	8.W.3.d	Writing
8	1	2	MC	OP	1	2	8.W.2.e	Writing
8	1	3	MC	OP	1	2	8.W.3.d	Writing
8	1	4	MC	OP	1	2	8.W.8	Writing
8	1	5	MC	OP	1	2	8.W.2.b	Writing
8	1	6	MS	OP	2	2	8.W.3.b	Writing
8	1	7	MC	OP	1	2	8.W.7	Writing
8	1	8	TE	OP	2	2	8.W.8	Writing
8	1	9	MC	OP	1	2	8.W.7	Writing
8	1	10	MC	OP	1	2	8.L.2.c	Writing
8	1	11	TE	OP	2	2	8.L.1.d	Writing
8	1	12	MC	OP	1	2	8.L.2.a	Writing
8	1	13	TE	OP	2	2	8.L.1.c	Writing
8	1	14	MC	OP	1	2	8.W.8	Writing
8	1	15	MC	OP	1	2	8.W.8	Writing
8	2	16	MS	OP	2	2	8.SL.3	Listening
8	2	17	EBSR	OP	2	3	8.SL.3	Listening
8	2	18	MC	OP	1	2	8.SL.3	Listening
8	2	19	MC	OP	1	2	8.SL.2	Listening
8	2	20	EBSR	OP	2	3	8.SL.3	Listening
8	3	21	MC	OP	1	2	8.RI.8	Reading
8	3	22	MC	OP	1	2	8.L.4	Reading
8	3	23	MC	OP	1	2	8.RI.6	Reading
8	3	24	MC	OP	1	2	8.RI.5	Reading
8	3	25	MS	OP	2	2	8.RI.3	Reading
8	3	26	MC	OP	1	2	8.RI.2	Reading
8	3	27	MC	OP	1	2	8.RL.4	Reading
8	3	28	MC	OP	1	3	8.RL.3	Reading
8	3	29	MC	OP	1	2	8.RL.3	Reading

Table A-6. English Language Arts, Grade 8 Test Map (cont.)

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	3	30	MC	OP	1	3	8.RL.6	Reading
8	3	31	EBSR	OP	2	3	8.RL.2	Reading
8	3	32	MC	OP	1	3	8.RI.5	Reading
8	3	33	EBSR	OP	2	2	8.L.4.RI	Reading
8	3	34	MC	OP	1	2	8.RI.5	Reading
8	3	35	MC	OP	1	2	8.RI.3	Reading
8	3	36	MC	OP	1	2	8.RI.6	Reading
8	3	37	MC	OP	1	2	8.RL.4	Reading
8	3	38	MC	OP	1	2	8.RL.1	Reading
8	3	39	MS	OP	2	2	8.RL.1	Reading
8	3	40	MC	OP	1	2	8.RL.6	Reading

Appendix B

Spring 2021 Mathematics Operational Test Maps

Table B-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	TE	OP	1	2	3.NBT.1	NBT
3	1	4	MC	OP	1	2	3.MD.1	MD
3	1	5	MC	OP	1	3	3.G.1	G
3	1	6	SA	OP	1	1	3.OA.4	OA
3	1	7	MC	OP	1	2	3.MD.5.b	MD
3	1	8	SA	OP	1	1	3.G.2	G
3	1	9	MC	OP	1	1	3.NF.2	NF
3	1	10	SA	OP	1	1	3.NBT.1	NBT
3	1	11	MC	OP	1	2	3.OA.6	OA
3	1	12	TE	OP	1	1	3.NF.2.b	NF
3	1	13	MC	OP	1	1	3.NBT.2	NBT
3	1	14	SA	OP	1	2	3.MD.7.a	MD
3	1	15	SA	OP	1	1	3.NF.3.b	NF
3	1	16	MC	OP	1	3	3.OA.8	OA
3	1	17	MC	OP	1	2	3.G.2	G
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.MD.2	MD
3	2	22	MC	OP	1	2	3.MD.8	MD
3	2	23	MC	OP	1	1	3.NBT.1	NBT
3	2	24	SA	OP	1	2	3.OA.7	OA
3	2	25	TE	OP	1	2	3.G.1	G
3	2	26	TE	OP	1	2	3.MD.3	MD
3	2	27	MC	OP	1	2	3.OA.2	OA
3	2	28	MC	OP	1	3	3.NBT.2	NBT
3	2	29	MC	OP	1	1	3.G.1	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	MC	OP	1	1	3.OA.5	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	2	3.NBT.3	NBT

Table B-1 Mathematics Grade 3 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
3	2	36	TE	OP	1	3	3.OA.3	OA
3	2	37	MC	OP	1	3	3.NF.3.d	NF
3	2	38	MC	OP	1	2	3.MD.6	MD
3	2	39	SA	OP	1	1	3.G.2	G
3	2	40	MC	OP	1	2	3.OA.9	OA
3	2	41	TE	OP	1	2	3.MD.4	MD
3	2	42	MC	OP	1	2	3.NBT.2	NBT

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 B-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	3	4.NBT.2	NBT
4	1	2	MC	OP	1	1	4.OA.1	OA
4	1	3	SA	OP	1	1	4.G.3	G
4	1	4	MC	OP	1	2	4.NF.3.d	NF
4	1	5	TE	OP	1	2	4.NBT.5	NBT
4	1	6	MC	OP	1	2	4.G.1	G
4	1	7	MC	OP	1	2	4.NBT.1	NBT
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	MC	OP	1	2	4.OA.2	OA
4	1	11	MC	OP	1	3	4.MD.3	MD
4	1	12	MC	OP	1	2	4.G.2	G
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.NF.6	NF
4	1	17	MC	OP	1	1	4.MD.5.b	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	MC	OP	1	2	4.NBT.4	NBT
4	1	21	MC	OP	1	1	4.MD.1	MD
4	1	22	MC	OP	1	3	4.OA.5	OA
4	1	23	SA	OP	1	2	4.NF.5	NF
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.MD.2	MD
4	2	27	MC	OP	1	1	4.NF.1	NF
4	2	28	MC	OP	1	2	4.G.1	G
4	2	29	SA	OP	1	2	4.NBT.5	NBT
4	2	30	MC	OP	1	2	4.MD.3	MD
4	2	31	MC	OP	1	2	4.OA.3	OA
4	2	32	TE	OP	1	2	4.G.3	G
4	2	33	MC	OP	1	1	4.NF.2	NF
4	2	34	MC	OP	1	2	4.OA.2	OA

Table B-2 Mathematics Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	2	35	SA	OP	1	1	4.MD.6	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.b	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.5	OA
4	2	43	SA	OP	1	1	4.NBT.4	NBT
4	2	44	MC	OP	1	2	4.NF.4.c	NF
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 B-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.OA.2	OA
5	1	2	MC	OP	1	2	5.NF.1	NF
5	1	3	MC	OP	1	2	5.MD.1	MD
5	1	4	MC	OP	1	2	5.G.1	G
5	1	5	MC	OP	1	3	5.NF.5	NF
5	1	6	TE	OP	1	2	5.OA.1	OA
5	1	7	MC	OP	1	1	5.NBT.3.a	NBT
5	1	8	MC	OP	1	2	5.MD.2	MD
5	1	9	TE	OP	1	2	5.G.2	G
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	SA	OP	1	1	5.NBT.1	NBT
5	1	13	MC	OP	1	2	5.MD.3	MD
5	1	14	SA	OP	1	2	5.NF.7.c	NF
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.G.1	G
5	2	24	MC	OP	1	1	5.NBT.2	NBT
5	2	25	SA	OP	1	3	5.OA.1	OA
5	2	26	MC	OP	1	1	5.G.3	G
5	2	27	SA	OP	1	1	5.MD.4	MD
5	2	28	MC	OP	1	2	5.NF.7.a	NF
5	2	29	TE	OP	1	2	5.OA.1	OA
5	2	30	MC	OP	1	1	5.G.4	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	MC	OP	1	2	5.G.2	G

Table B-3 Mathematics Grade 5 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	2	35	TE	OP	1	2	5.NF.4	NF
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.MD.3	MD
5	2	39	MC	OP	1	2	5.NBT.7	NBT
5	2	40	TE	OP	1	2	5.MD.5.b	MD
5	2	41	MC	OP	1	2	5.NF.6	NF
5	2	42	MC	OP	1	2	5.OA.1	OA
5	2	43	MC	OP	1	3	5.NF.2	NF
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	SA	OP	1	2	5.NBT.7	NBT

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 B-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.NS.3	NS
6	1	5	SA	OP	1	1	6.EE.2.a	EE
6	1	6	SA	OP	1	1	6.NS.2	NS
6	1	7	MC	OP	1	2	6.EE.3	EE
6	1	8	MC	OP	1	2	6.NS.2	NS
6	1	9	SA	OP	1	2	6.RP.1	RP
6	1	10	MC	OP	1	2	6.EE.2.b	EE
6	1	11	TE	OP	1	2	6.RP.3.a	RP
6	1	12	MC	OP	1	2	6.RP.2	RP
6	1	13	SA	OP	1	1	6.EE.1	EE
6	1	14	SA	OP	1	1	6.RP.2	RP
6	1	15	MC	OP	1	1	6.NS.4	NS
6	1	16	MC	OP	1	2	6.RP.3.a	RP
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	2	6.NS.6.b	NS
6	2	28	MC	OP	1	2	6.G.1	G
6	2	29	MC	OP	1	2	6.EE.9	EE
6	2	30	MC	OP	1	2	6.SP.5.b	SP
6	2	31	MC	OP	1	2	6.SP.5.c	SP
6	2	32	MC	OP	1	2	6.SP.3	SP
6	2	33	MC	OP	1	2	6.NS.8	NS
6	2	34	MC	OP	1	2	6.EE.6	EE

Table B-4 Mathematics Grade 6 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
6	2	35	MC	OP	1	2	6.SP.4	SP
6	2	36	TE	OP	1	1	6.NS.6	NS
6	2	37	SA	OP	1	1	6.G.2	G
6	2	38	MC	OP	1	2	6.EE.5	EE
6	2	39	MC	OP	1	2	6.SP.5	SP
6	2	40	MC	OP	1	2	6.NS.8	NS
6	2	41	TE	OP	1	2	6.G.3	G
6	2	42	MC	OP	1	2	6.SP.2	SP
6	2	43	MC	OP	1	2	6.G.4	G
6	2	44	TE	OP	1	2	6.SP.5.c	SP
6	2	45	MC	OP	1	2	6.G.3	G
6	2	46	MC	OP	1	2	6.SP.4	SP

Domain Names: G=Geometry; EE=Expressions and Equations; NS=The Number System; SP=Statistics and Probability: RP= Ratios and Proportional Relationships

Table B-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	MC	OP	1	2	7.NS.3	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.EE.1	EE
7	1	6	MC	OP	1	2	7.NS.2	NS
7	1	7	MC	OP	1	2	7.NS.1	NS
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	2	7.EE.2	EE
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	MC	OP	1	2	7.EE.4	EE
7	2	17	MC	OP	1	2	7.G.5	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	MC	OP	1	2	7.SP.1	SP
7	2	22	MC	OP	1	1	7.G.2	G
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.G.1	G
7	2	26	MC	OP	1	2	7.SP.7.b	SP
7	2	27	MC	OP	1	2	7.SP.1	SP
7	2	28	MC	OP	1	2	7.RP.2	RP
7	2	29	MC	OP	1	2	7.SP.5	SP
7	2	30	MC	OP	1	3	7.RP.2.a	RP
7	2	31	MC	OP	1	2	7.RP.1	RP
7	2	32	TE	OP	1	2	7.G.2	G
7	2	33	TE	OP	1	2	7.EE.3	EE
7	2	34	MC	OP	1	2	7.G.6	G

Table B-5 Mathematics Grade 7 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
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.RP.3	RP
7	2	38	TE	OP	1	3	7.EE.4.b	EE
7	2	39	MC	OP	1	2	7.SP.3	SP
7	2	40	MC	OP	1	2	7.G.4	G
7	2	41	TE	OP	1	2	7.SP.6	SP
7	2	42	TE	OP	1	2	7.G.1	G
7	2	43	MC	OP	1	2	7.SP.2	SP
7	2	44	MC	OP	1	2	7.EE.4.b	EE
7	2	45	MC	OP	1	2	7.SP.2	SP
7	2	46	SA	OP	1	1	7.RP.1	RP

Domain Names: G=Geometry; EE=Expressions and Equations; NS=The Number System; SP=Statistics and Probability: RP= Ratios and Proportional Relationships

Table B-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	2	8.EE.1	EE
8	1	3	MC	OP	1	1	8.EE.2	EE
8	1	4	MC	OP	1	1	8.NS.1	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.NS.2	NS
8	1	8	MC	OP	1	1	8.EE.2	EE
8	1	9	TE	OP	1	1	8.NS.1	NS
8	1	10	MC	OP	1	2	8.NS.1	NS
8	1	11	TE	OP	1	2	8.NS.2	NS
8	1	12	MC	OP	1	2	8.EE.3	EE
8	1	13	SA	OP	1	1	8.NS.2	NS
8	2	14	MC	OP	1	1	8.F.1	F
8	2	15	TE	OP	1	3	8.G.5	G
8	2	16	MC	OP	1	1	8.SP.1	SP
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	3	8.F.2	F
8	2	21	SA	OP	1	2	8.G.7	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.G.3	G
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.4	SP
8	2	28	MC	OP	1	2	8.F.4	F
8	2	29	MC	OP	1	2	8.G.2	G
8	2	30	MC	OP	1	2	8.SP.3	SP
8	2	31	MC	OP	1	2	8.EE.8.a	EE
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	MC	OP	1	2	8.SP.4	SP

Table B-6 Mathematics Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	2	35	TE	OP	1	2	8.F.4	F
8	2	36	MC	OP	1	2	8.EE.6	EE
8	2	37	TE	OP	1	2	8.SP.2	SP
8	2	38	MC	OP	1	2	8.F.5	F
8	2	39	MC	OP	1	2	8.G.6	G
8	2	40	MC	OP	1	2	8.F.3	F
8	2	41	TE	OP	1	2	8.G.2	G
8	2	42	MC	OP	1	2	8.SP.3	SP
8	2	43	MC	OP	1	1	8.G.1.a	G
8	2	44	MC	OP	1	2	8.SP.1	SP
8	2	45	MC	OP	1	2	8.EE.8	EE
8	2	46	MC	OP	1	1	8.F.5	F

Domain Names: G=Geometry; EE=Expressions and Equations; NS=The Number System; SP=Statistics and Probability: F= Functions

Appendix C

Spring 2021 Science Operational Test Maps

Table C-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	MC	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	1	2	TE	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	
4	1	3	TE	OP	1	3	SCI.ESS3.B.4	Earth and Space Science		SCI.CC2.3-5
4	1	4	TE	OP	1	3	SCI.ESS1.C.4	Earth and Space Science	SCI.SEP6.A.3-5	
4	1	5	TE	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	1	6	TE	OP	1	2	SCI.ETS1.C.3-5	Engineering	SCI.SEP3.A.3-5	
4	1	7	MC	OP	1	2	SCI.PS3.B.4	Physical Science	SCI.SEP1.A.3-5	SCI.CC5.3-5
4	1	8	TE	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	1	9	EBSR	OP	1	3	SCI.ETS1.C.3-5	Engineering	SCI.SEP3.A.3-5	
4	1	10	MC	OP	1	2	SCI.ESS2.A.4	Earth and Space Science	SCI.SEP3.A.3-5	SCI.CC2.3-5
4	1	11	MC	OP	1	2	SCI.ESS2.B.4	Earth and Space Science	SCI.SEP4.A.3-5	SCI.CC1.3-5
4	1	12	MC	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	SCI.CC4.3-5
4	1	13	TE	OP	1	2	SCI.LS1.D.4	Life Science		SCI.CC4.3-5
4	1	14	TE	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP4.A.3-5	
4	1	15	TE	OP	1	2	SCI.LS1.D.4	Life Science	SCI.SEP2.A.3-5	SCI.CC4.3-5
4	2	16	MC	OP	1	3	SCI.PS3.B.4	Physical Science	SCI.SEP3.A.3-5	SCI.CC5.3-5
4	2	17	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5

Table C-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	2	18	TE	OP	1	2	SCI.PS4.B.4	Physical Science	SCI.SEP2.A.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	TE	OP	1	2	SCI.ESS3.A.4	Earth and Space Science		SCI.CC2.3-5
4	2	23	MC	OP	1	2	SCI.PS3.B.4	Physical Science		SCI.CC5.3-5
4	2	24	TE	OP	1	3	SCI.PS3.C.4	Physical Science		SCI.CC5.3-5
4	2	25	TE	OP	1	2	SCI.ETS1.A.3-5	Engineering	SCI.SEP6.A.3-5	
4	3	26	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5
4	3	27	TE	OP	1	3	SCI.ESS3.B.4	Earth and Space Science	SCI.SEP6.B.3-5	SCI.CC2.3-5
4	3	28	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5
4	3	29	TE	OP	1	3	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	3	30	TE	OP	1	3	SCI.ETS1.B.3-5	Engineering	SCI.SEP1.B.3-5	
4	3	31	TE	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	3	32	MC	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	3	33	TE	OP	1	3	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	

Table C-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	3	34	TE	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	3	35	MC	OP	1	2	SCI.ESS2.B.4	Earth and Space Science	SCI.SEP4.A.3-5	SCI.CC1.3-5
4	3	36	EBSR	OP	1	2	SCI.ESS1.C.4	Earth and Space Science	SCI.SEP6.A.3-5	
4	3	37	MC	OP	1	3	SCI.PS4.A.4	Physical Science	SCI.SEP2.A.3-5	
4	3	38	TE	OP	1	3	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	SCI.CC4.3-5
4	3	39	MS	OP	1	3	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	
4	3	40	TE	OP	1	3	SCI.LS1.D.4	Life Science	SCI.SEP2.A.3-5	SCI.CC4.3-5

Table C-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	TE	OP	1	2	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	SCI.CC2.m
8	1	2	TE	OP	1	3	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	SCI.CC2.m
8	1	3	TE	OP	1	3	SCI.LS2.A.m	Life Science		SCI.CC1.m
8	1	4	TE	OP	1	3	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.m	
8	1	5	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP2.A.m	
8	1	6	MC	OP	1	3	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.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.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	1	10	TE	OP	1	2	SCI.LS1.B.m	Life Science	SCI.SEP2.A.m	SCI.CC2.m
8	1	11	TE	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	1	12	MC	OP	1	3	SCI.LS1.B.m	Life Science	SCI.SEP7.A.m	SCI.CC2.m
8	1	13	TE	OP	1	2	SCI.PS1.A.m	Physical Science	SCI.SEP2.A.m	SCI.CC3.m
8	1	14	TE	OP	1	2	SCI.PS3.B.m	Physical Science	SCI.SEP7.A.m	SCI.CC5.m
8	1	15	TE	OP	1	2	SCI.PS4.A.m	Physical Science	SCI.SEP5.A.m	SCI.CC1.m
8	2	16	TE	OP	1	2	SCI.LS1.C.m	Life Science	SCI.SEP2.A.m	
8	2	17	TE	OP	1	2	SCI.LS4.B.m	Life Science		SCI.CC2.m
8	2	18	TE	OP	1	2	SCI.LS4.B.m	Life Science		SCI.CC2.m

Table C-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	2	19	TE	OP	1	3	SCI.ESS1.C.m	Earth and Space Science		SCI.CC3.m
8	2	20	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP6.A.m	
8	2	21	TE	OP	1	2	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.m	
8	2	22	TE	OP	1	2	SCI.ESS1.C.m	Earth and Space Science	SCI.SEP6.A.m	SCI.CC3.m
8	2	23	TE	OP	1	3	SCI.ESS3.C.m	Earth and Space Science	SCI.SEP7.A.m	SCI.CC2.m
8	2	24	TE	OP	1	2	SCI.ESS2.A.m	Earth and Space Science	SCI.SEP6.A.m	SCI.CC3.m
8	2	25	TE	OP	1	2	SCI.ETS1.C.m	Engineering	SCI.SEP2.A.m	
8	3	26	TE	OP	1	2	SCI.LS1.C.m	Life Science	SCI.SEP6.A.m	SCI.CC5.m
8	3	27	TE	OP	1	2	SCI.LS2.B.m	Life Science	SCI.SEP2.A.m	SCI.CC5.m
8	3	28	TE	OP	1	2	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	
8	3	29	TE	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	3	30	MC	OP	1	3	SCI.PS1.B.m	Physical Science	SCI.SEP6.B.m	
8	3	31	EBSR	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	3	32	TE	OP	1	2	SCI.ETS1.B.m	Engineering	SCI.SEP4.A.m	
8	3	33	TE	OP	1	2	SCI.PS1.B.m	Physical Science	SCI.SEP2.A.m	SCI.CC5.m
8	3	34	TE	OP	1	2	SCI.PS2.B.m	Physical Science	SCI.SEP6.A.m	SCI.CC2.m
8	3	35	TE	OP	1	3	SCI.PS2.B.m	Physical Science	SCI.SEP6.A.m	SCI.CC2.m
8	3	36	MS	OP	1	2	SCI.PS2.B.m	Physical Science	SCI.SEP1.A.m	SCI.CC2.m

Table C-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	3	37	TE	OP	1	2	SCI.PS1.A.m	Physical Science	SCI.SEP2.A.m	SCI.CC3.m
8	3	38	MC	OP	1	3	SCI.PS2.B.m	Physical Science	SCI.SEP7.A.m	SCI.CC4.m
8	3	39	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP4.A.m	SCI.CC3.m
8	3	40	MC	OP	1	2	SCI.ESS1.B.m	Earth and Space Science		SCI.CC4.m

Appendix D

Spring 2021 Social Studies Operational Test Maps

Table D-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	MC	OP	1	2	E.4.8	Behavioral Sciences
4	1	2	MC	OP	1	1	B.4.6	History
4	1	3	MC	OP	1	2	E.4.15	Behavioral Sciences
4	1	4	MC	OP	1	2	B.4.2	History
4	1	5	MC	OP	1	3	D.4.3	Economics
4	1	6	MC	OP	1	2	A.4.7	Geography
4	1	7	TE	OP	1	2	B.4.8	History
4	1	8	MC	OP	1	2	A.4.1	Geography
4	1	9	MC	OP	1	2	A.4.1	Geography
4	1	10	MC	OP	1	2	E.4.15	Behavioral Sciences
4	1	11	MC	OP	1	2	C.4.6	Civics
4	1	12	MC	OP	1	2	B.4.7	History
4	1	13	TE	OP	1	2	D.4.6	Economics
4	1	14	MC	OP	1	3	D.4.3	Economics
4	1	15	MC	OP	1	2	E.4.6	Behavioral Sciences
4	1	16	MC	OP	1	1	C.4.2	Civics
4	1	17	MC	OP	1	3	B.4.9	History
4	1	18	TE	OP	1	1	E.4.5	Behavioral Sciences
4	1	19	MC	OP	1	3	E.4.11	Behavioral Sciences
4	2	20	MC	OP	1	2	A.4.7	Geography
4	2	21	MC	OP	1	2	C.4.1	Civics
4	2	22	MC	OP	1	1	A.4.2	Geography
4	2	23	MC	OP	1	2	E.4.9	Behavioral Sciences
4	2	24	MC	OP	1	2	C.4.3	Civics
4	2	25	MC	OP	1	2	B.4.8	History
4	2	26	MC	OP	1	2	C.4.4	Civics
4	2	27	MC	OP	1	3	B.4.4	History
4	2	28	MC	OP	1	2	A.4.9	Geography
4	2	29	MC	OP	1	2	D.4.2	Economics
4	2	30	MC	OP	1	2	D.4.7	Economics
4	2	31	TE	OP	1	1	A.4.2	Geography
4	2	32	MC	OP	1	2	A.4.6	Geography

Table D-1 Social Studies Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	2	33	MC	OP	1	2	C.4.1	Civics
4	2	34	MC	OP	1	2	B.4.7	History
4	2	35	MC	OP	1	2	A.4.4	Geography
4	2	36	MC	OP	1	2	B.4.6	History
4	2	37	MC	OP	1	3	E.4.12	Behavioral Sciences
4	2	38	MC	OP	1	2	D.4.1	Economics

Table D-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	2	A.8.8	Geography
8	1	2	MC	OP	1	3	B.8.5	History
8	1	3	MC	OP	1	3	B.8.9	History
8	1	4	MC	OP	1	3	B.8.4	History
8	1	5	MC	OP	1	2	B.8.2	History
8	1	6	MC	OP	1	2	B.8.7	History
8	1	7	MS	OP	1	2	B.8.7	History
8	1	8	MC	OP	1	3	C.8.9	Civics
8	1	9	MC	OP	1	2	A.8.9	Geography
8	1	10	TE	OP	1	2	D.8.2	Economics
8	1	11	MC	OP	1	2	A.8.5	Geography
8	1	12	MC	OP	1	2	A.8.11	Geography
8	1	13	MC	OP	1	2	D.8.8	Economics
8	1	14	MC	OP	1	3	E.8.14	Behavioral Sciences
8	1	15	MC	OP	1	2	A.8.7	Geography
8	1	16	MC	OP	1	3	D.8.7	Economics
8	1	17	MC	OP	1	3	E.8.9	Behavioral Sciences
8	1	18	MC	OP	1	2	B.8.3	History
8	1	19	MC	OP	1	2	C.8.1	Civics
8	1	20	MC	OP	1	2	B.8.7	History
8	2	21	MC	OP	1	2	E.8.10	Behavioral Sciences
8	2	22	MC	OP	1	2	A.8.11	Geography
8	2	23	TE	OP	1	2	A.8.10	Geography
8	2	24	MC	OP	1	2	A.8.5	Geography
8	2	25	MC	OP	1	3	A.8.5	Geography
8	2	26	MC	OP	1	2	C.8.6	Civics
8	2	27	MC	OP	1	3	D.8.3	Economics
8	2	28	MC	OP	1	2	E.8.11	Behavioral Sciences
8	2	29	MC	OP	1	3	C.8.8	Civics
8	2	30	TE	OP	1	2	D.8.5	Economics
8	2	31	MC	OP	1	3	B.8.9	History
8	2	32	TE	OP	1	2	C.8.1	Civics
8	2	33	MC	OP	1	2	C.8.6	Civics
8	2	34	MC	OP	1	2	B.8.5	History

Table D-2 Social Studies Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	2	35	MC	OP	1	3	B.8.3	History
8	2	36	MC	OP	1	2	D.8.7	Economics
8	2	37	MC	OP	1	3	B.8.10	History
8	2	38	MC	OP	1	2	E.8.9	Behavioral Sciences
8	2	39	MC	OP	1	3	A.8.9	Geography
8	2	40	MC	OP	1	2	E.8.4	Behavioral Sciences

Table D-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	MC	OP	1	2	A.10.7	Geography
10	1	2	MC	OP	1	3	A.10.5	Geography
10	1	3	MC	OP	1	2	B.10.8	History
10	1	4	MC	OP	1	2	E.10.12	Behavioral Sciences
10	1	5	MC	OP	1	2	B.10.5	History
10	1	6	MC	OP	1	2	B.10.9	History
10	1	7	MC	OP	1	2	B.10.7	History
10	1	8	MC	OP	1	2	E.10.14	Behavioral Sciences
10	1	9	MC	OP	1	2	D.10.10	Economics
10	1	10	MC	OP	1	2	C.10.6	Civics
10	1	11	MC	OP	1	2	B.10.16	History
10	1	12	TE	OP	1	2	E.10.17	Behavioral Sciences
10	1	13	MC	OP	1	3	D.10.7	Economics
10	1	14	MC	OP	1	2	E.10.5	Behavioral Sciences
10	1	15	MC	OP	1	2	A.10.6	Geography
10	1	16	MC	OP	1	2	A.10.6	Geography
10	1	17	MC	OP	1	2	A.10.4	Geography
10	1	18	MC	OP	1	2	A.10.1	Geography
10	1	19	MC	OP	1	3	B.10.9	History
10	1	20	MC	OP	1	3	D.10.2	Economics
10	1	21	MC	OP	1	2	B.10.13	History
10	1	22	MC	OP	1	3	C.10.4	Civics
10	1	23	MC	OP	1	2	D.10.14	Economics
10	1	24	TE	OP	1	2	C.10.13	Civics
10	1	25	MC	OP	1	2	B.10.6	History
10	2	26	MC	OP	1	2	E.10.17	Behavioral Sciences
10	2	27	MC	OP	1	2	E.10.6	Behavioral Sciences
10	2	28	MC	OP	1	2	B.10.10	History
10	2	29	MC	OP	1	3	C.10.3	Civics
10	2	30	MC	OP	1	2	C.10.13	Civics
10	2	31	MC	OP	1	2	D.10.3	Economics
10	2	32	MC	OP	1	2	C.10.15	Civics

Table D-3 Social Studies Grade 10 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
10	2	33	MC	OP	1	2	E.10.8	Behavioral Sciences
10	2	34	TE	OP	1	2	E.10.5	Behavioral Sciences
10	2	35	MC	OP	1	3	B.10.18	History
10	2	36	MC	OP	1	3	B.10.6	History
10	2	37	MC	OP	1	3	E.10.4	Behavioral Sciences
10	2	38	MC	OP	1	2	C.10.6	Civics
10	2	39	MC	OP	1	2	D.10.10	Economics
10	2	40	MC	OP	1	2	B.10.6	History
10	2	41	MC	OP	1	3	C.10.14	Civics
10	2	42	MC	OP	1	2	A.10.8	Geography
10	2	43	MC	OP	1	3	A.10.5	Geography
10	2	44	MC	OP	1	2	C.10.1	Civics
10	2	45	TE	OP	1	2	D.10.7	Economics
10	2	46	MC	OP	1	2	A.10.12	Geography
10	2	47	MC	OP	1	2	B.10.3	History
10	2	48	MC	OP	1	2	C.10.11	Civics
10	2	49	MC	OP	1	1	A.10.8	Geography
10	2	50	MC	OP	1	3	D.10.8	Economics

Appendix E

Spring 2021 Test Participation Rates by Subgroup

Table E-1. Test Participation Rates, Grade 3

Group	Category	Enrolled	English Language Arts		Mathematics	
			Number Tested	Percent Tested	Number Tested	Percent Tested
State	All Students	60785	52930	87.08	52892	87.01
Gender	Male	31194	27044	86.70	27027	86.64
	Female	29591	25886	87.48	25865	87.41
Race/ Ethnicity	White	39415	36547	92.72	36533	92.69
	African American	6130	4101	66.90	4058	66.20
	Hispanic	8619	6787	78.74	6803	78.93
	Asian	2787	2237	80.27	2241	80.41
	American Indian	646	563	87.15	562	87.00
	Two or More	3188	2695	84.54	2695	84.54
Limited English Proficiency	No	55183	48597	88.07	48528	87.94
	Yes	5602	4333	77.35	4364	77.90
Disability Status	No	52189	46009	88.16	45985	88.11
	Yes	8596	6921	80.51	6907	80.35
Economically Disadvantaged	No	35338	32393	91.67	32412	91.72
	Yes	25447	20537	80.70	20480	80.48

Table E-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	61127	52706	86.22	52658	86.15	52417	85.75	52392	85.71
Gender	Male	31284	26877	85.91	26853	85.84	26737	85.47	26712	85.39
	Female	29843	25829	86.55	25805	86.47	25680	86.05	25680	86.05
Race/Ethnicity	White	39812	36592	91.91	36579	91.88	36548	91.80	36546	91.80
	African American	6311	4079	64.63	4051	64.19	3941	62.45	3926	62.21
	Hispanic	8639	6789	78.59	6794	78.64	6720	77.79	6718	77.76
	Asian	2702	2166	80.16	2164	80.09	2143	79.31	2143	79.31
	American Indian	683	599	87.70	595	87.12	598	87.55	595	87.12
	Two or More	2980	2481	83.26	2475	83.05	2467	82.79	2464	82.68
Limited English Proficiency	No	55749	48560	87.10	48501	87.00	48305	86.65	48282	86.61
	Yes	5378	4146	77.09	4157	77.30	4112	76.46	4110	76.42
Disability Status	No	52443	45875	87.48	45850	87.43	45651	87.05	45646	87.04
	Yes	8684	6831	78.66	6808	78.40	6766	77.91	6746	77.68
Economically Disadvantaged	No	35822	32638	91.11	32626	91.08	32585	90.96	32577	90.94
	Yes	25305	20068	79.30	20032	79.16	19832	78.37	19815	78.30

Table E-3. Test Participation Rates, Grade 5

Group	Category	Enrolled	English Language Arts		Mathematics	
			Number Tested	Percent Tested	Number Tested	Percent Tested
State	All Students	62405	54010	86.55	53932	86.42
Gender	Male	31905	27494	86.17	27444	86.02
	Female	30500	26516	86.94	26488	86.85
Race/ Ethnicity	White	40858	37720	92.32	37709	92.29
	African American	6567	4255	64.79	4187	63.76
	Hispanic	8800	6900	78.41	6902	78.43
	Asian	2562	2047	79.90	2055	80.21
	American Indian	694	595	85.73	597	86.02
	Two or More	2924	2493	85.26	2482	84.88
Limited English Proficiency	No	57855	50550	87.37	50444	87.19
	Yes	4550	3460	76.04	3488	76.66
Disability Status	No	53826	47300	87.88	47249	87.78
	Yes	8579	6710	78.21	6683	77.90
Economically Disadvantaged	No	36356	33268	91.51	33265	91.50
	Yes	26049	20742	79.63	20667	79.34

Table E-4. Test Participation Rates, Grade 6

Group	Category	Enrolled	English Language Arts		Mathematics	
			Number Tested	Percent Tested	Number Tested	Percent Tested
State	All Students	64925	55511	85.50	55462	85.42
Gender	Male	33295	28342	85.12	28313	85.04
	Female	31630	27169	85.90	27149	85.83
Race/ Ethnicity	White	42380	38850	91.67	38823	91.61
	African American	6938	4215	60.75	4182	60.28
	Hispanic	9128	7117	77.97	7132	78.13
	Asian	2781	2256	81.12	2252	80.98
	American Indian	693	584	84.27	584	84.27
	Two or More	3005	2489	82.83	2489	82.83
Limited English Proficiency	No	60982	52523	86.13	52453	86.01
	Yes	3943	2988	75.78	3009	76.31
Disability Status	No	56222	48973	87.11	48939	87.05
	Yes	8703	6538	75.12	6523	74.95
Economically Disadvantaged	No	38298	34812	90.90	34795	90.85
	Yes	26627	20699	77.74	20667	77.62

Table E-5. Test Participation Rates, Grade 7

Group	Category	Enrolled	English Language Arts		Mathematics	
			Number Tested	Percent Tested	Number Tested	Percent Tested
State	All Students	66361	56295	84.83	56247	84.76
Gender	Male	33945	28847	84.98	28832	84.94
	Female	32416	27448	84.67	27415	84.57
Race/ Ethnicity	White	43669	39754	91.03	39723	90.96
	African American	6943	4251	61.23	4225	60.85
	Hispanic	9422	7199	76.41	7203	76.45
	Asian	2647	2129	80.43	2127	80.36
	American Indian	698	568	81.38	565	80.95
	Two or More	2982	2394	80.28	2404	80.62
Limited English Proficiency	No	62336	53350	85.58	53291	85.49
	Yes	4025	2945	73.17	2956	73.44
Disability Status	No	57849	49987	86.41	49942	86.33
	Yes	8512	6308	74.11	6305	74.07
Economically Disadvantaged	No	39602	35697	90.14	35704	90.16
	Yes	26759	20598	76.98	20543	76.77

Table E-6. Test Participation Rates, ELA 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	67572	56756	83.99	56726	83.95	56485	83.59	56409	83.48
Gender	Male	34640	29223	84.36	29197	84.29	29078	83.94	29055	83.88
	Female	32932	27533	83.61	27529	83.59	27407	83.22	27354	83.06
Race/Ethnicity	White	44872	40254	89.71	40214	89.62	40179	89.54	40138	89.45
	African American	7003	4362	62.29	4351	62.13	4230	60.40	4218	60.23
	Hispanic	9365	7112	75.94	7128	76.11	7082	75.62	7061	75.40
	Asian	2647	2098	79.26	2103	79.45	2087	78.84	2081	78.62
	American Indian	773	635	82.15	636	82.28	626	80.98	625	80.85
	Two or More	2912	2295	78.81	2294	78.78	2281	78.33	2286	78.50
Limited English Proficiency	No	63600	53893	84.74	53844	84.66	53629	84.32	53562	84.22
	Yes	3972	2863	72.08	2882	72.56	2856	71.90	2847	71.68
Disability Status	No	58840	50301	85.49	50295	85.48	50114	85.17	50043	85.05
	Yes	8732	6455	73.92	6431	73.65	6371	72.96	6366	72.90
Economically Disadvantaged	No	40899	36531	89.32	36514	89.28	36473	89.18	36452	89.13
	Yes	26673	20225	75.83	20212	75.78	20012	75.03	19957	74.82

Table E-7. Test Participation Rates, Grade 10

Group	Category	Enrolled	Social Studies	
			Number Tested	Percent Tested
State	All Students	68221	51433	75.39
Gender	Male	34779	26475	76.12
	Female	33442	24958	74.63
Race/ Ethnicity	White	46682	39414	84.43
	African American	6694	2421	36.17
	Hispanic	8955	5531	61.76
	Asian	2582	1854	71.80
	American Indian	785	526	67.01
	Two or More	2523	1687	66.86
Limited English Proficiency	No	65234	49799	76.34
	Yes	2987	1634	54.70
Disability Status	No	59820	46307	77.41
	Yes	8401	5126	61.02
Economically Disadvantaged	No	44385	36855	83.03
	Yes	23836	14578	61.16

Appendix F

Classical Item Analysis Results

Table F-1. Item Statistics, ELA Grade 3

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	52780	0.67	0.32	0.00		0.12	0.16	0.67	0.05		-0.17	-0.14	0.32	-0.18
2	TE	2	52508	0.68	0.38	0.01	0.19	0.27	0.54			-0.22	-0.26	0.42		
3	MC	1	52741	0.54	0.40	0.00		0.54	0.20	0.17	0.09		0.40	-0.22	-0.17	-0.17
4	MC	1	52735	0.42	0.29	0.00		0.25	0.42	0.07	0.27		-0.13	0.29	-0.18	-0.09
5	TE	2	52403	0.52	0.45	0.01	0.25	0.46	0.28			-0.30	-0.11	0.43		
6	MC	1	52728	0.53	0.38	0.00		0.15	0.13	0.19	0.53		-0.20	-0.16	-0.17	0.39
7	TE	2	52654	0.54	0.42	0.00	0.24	0.45	0.31			-0.29	-0.12	0.40		
8	MC	1	52721	0.61	0.44	0.00		0.15	0.13	0.11	0.61		-0.25	-0.16	-0.23	0.44
9	MC	1	52727	0.69	0.44	0.00		0.17	0.04	0.10	0.69		-0.25	-0.21	-0.22	0.44
10	MC	1	52725	0.58	0.39	0.00		0.22	0.58	0.10	0.09		-0.23	0.39	-0.18	-0.14
11	TE	1	52332	0.30	0.39	0.01	0.70	0.29				-0.36	0.40			
12	MC	1	52701	0.56	0.47	0.00		0.56	0.13	0.17	0.14		0.47	-0.22	-0.18	-0.26
13	MC	1	52728	0.43	0.36	0.00		0.18	0.24	0.43	0.15		-0.10	-0.20	0.37	-0.15
14	MC	1	52707	0.69	0.54	0.00		0.10	0.13	0.08	0.69		-0.28	-0.29	-0.25	0.55
15	TE	2	52657	0.62	0.55	0.00	0.18	0.39	0.43			-0.43	-0.15	0.49		
16	MC	1	52742	0.70	0.36	0.00		0.11	0.70	0.10	0.08		-0.16	0.36	-0.18	-0.21
17	TE	2	52730	0.62	0.54	0.00	0.15	0.46	0.39			-0.39	-0.19	0.48		
18	MC	1	52734	0.54	0.37	0.00		0.28	0.10	0.09	0.54		-0.14	-0.24	-0.18	0.37
19	MS	2	52739	0.71	0.54	0.00	0.09	0.38	0.52			-0.34	-0.31	0.51		
20	MC	1	52684	0.44	0.33	0.00		0.12	0.44	0.31	0.13		-0.19	0.33	-0.22	0.00

Table F-1. Item Statistics, ELA Grade 3 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	52703	0.52	0.22	0.00		0.24	0.04	0.20	0.52		-0.06	-0.20	-0.11	0.22
22	MC	1	52637	0.62	0.41	0.00		0.24	0.62	0.09	0.05		-0.25	0.41	-0.15	-0.22
23	MC	1	52632	0.40	0.28	0.00		0.21	0.40	0.14	0.24		-0.05	0.28	-0.16	-0.13
24	MC	1	52568	0.51	0.31	0.00		0.51	0.10	0.22	0.17		0.31	-0.16	-0.17	-0.09
25	MC	1	52606	0.54	0.21	0.00		0.53	0.15	0.19	0.13		0.21	0.05	-0.08	-0.27
26	TE	2	52607	0.64	0.64	0.00	0.24	0.23	0.53			-0.51	-0.22	0.63		
27	MC	1	52611	0.54	0.32	0.00		0.17	0.54	0.13	0.16		-0.21	0.32	-0.18	-0.04
28	MS	2	52618	0.54	0.49	0.00	0.22	0.46	0.31			-0.36	-0.11	0.45		
29	MC	1	52632	0.55	0.44	0.00		0.08	0.21	0.55	0.17		-0.21	-0.17	0.44	-0.25
30	MC	1	52608	0.70	0.48	0.00		0.14	0.09	0.07	0.70		-0.27	-0.24	-0.22	0.48
31	MC	1	52608	0.27	0.27	0.00		0.26	0.23	0.23	0.27		0.27	-0.01	-0.07	-0.19
32	EBSR	2	52685	0.43	0.46	0.00	0.32	0.49	0.19			-0.43	0.18	0.30		
33	MC	1	52609	0.42	0.33	0.00		0.17	0.42	0.25	0.15		-0.15	0.33	-0.05	-0.23
34	MC	1	52594	0.43	0.37	0.00		0.20	0.13	0.43	0.23		-0.07	-0.17	0.37	-0.23
35	MC	1	52593	0.59	0.53	0.00		0.59	0.16	0.16	0.08		0.53	-0.20	-0.29	-0.28
36	MC	1	52587	0.46	0.32	0.00		0.16	0.12	0.26	0.46		-0.11	-0.16	-0.14	0.32
37	MC	1	52617	0.41	0.29	0.00		0.18	0.40	0.09	0.33		-0.09	0.29	-0.22	-0.09
38	TE	2	52599	0.50	0.50	0.00	0.33	0.33	0.33			-0.39	-0.08	0.48		

Table F-2. Item Statistics, ELA Grade 4

| Item Number | Item Type | Maximum Points | Number of Students | Item p-value | Item-Total Test Correlation | Proportion Omit | Proportion of Students | | | | | Item-Total Test Correlation | | | | |
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TE	1	52264	0.52	0.46	0.01	0.48	0.51				-0.45	0.47			
2	TE	1	52454	0.59	0.40	0.00	0.41	0.59				-0.40	0.40			
3	MC	1	52511	0.68	0.34	0.00		0.19	0.68	0.09	0.04		-0.24	0.34	-0.13	-0.13
4	TE	1	52307	0.45	0.42	0.01	0.55	0.45				-0.41	0.42			
5	MC	1	52525	0.54	0.39	0.00		0.30	0.10	0.54	0.06		-0.26	-0.13	0.39	-0.14
6	MC	1	52511	0.47	0.31	0.00		0.19	0.12	0.22	0.47		-0.10	-0.22	-0.10	0.31
7	MC	1	52490	0.63	0.45	0.00		0.63	0.11	0.18	0.08		0.45	-0.27	-0.18	-0.25
8	TE	2	52309	0.63	0.47	0.01	0.12	0.50	0.38			-0.33	-0.17	0.41		
9	MC	1	52483	0.54	0.45	0.00		0.15	0.22	0.54	0.09		-0.24	-0.23	0.45	-0.14
10	MC	1	52490	0.55	0.36	0.00		0.23	0.05	0.16	0.55		-0.06	-0.22	-0.28	0.36
11	MC	1	52497	0.56	0.36	0.00		0.27	0.10	0.07	0.55		-0.08	-0.27	-0.23	0.36
12	MC	1	52501	0.55	0.18	0.00		0.55	0.25	0.10	0.10		0.18	0.04	-0.16	-0.19
13	TE	2	52058	0.68	0.41	0.01	0.07	0.51	0.41			-0.16	-0.31	0.42		
14	MC	1	52499	0.66	0.47	0.00		0.66	0.11	0.11	0.12		0.48	-0.26	-0.21	-0.24
15	TE	2	52497	0.79	0.34	0.00	0.03	0.36	0.61			-0.21	-0.24	0.32		
16	MC	1	52483	0.41	0.23	0.00		0.17	0.41	0.17	0.24		-0.12	0.23	-0.12	-0.05
17	MC	1	52531	0.57	0.27	0.00		0.11	0.20	0.57	0.13		-0.12	-0.08	0.27	-0.19
18	MC	1	52503	0.74	0.34	0.00		0.02	0.20	0.04	0.74		-0.16	-0.23	-0.19	0.34
19	EBSR	2	52537	0.31	0.35	0.00	0.58	0.22	0.20			-0.31	0.09	0.30		
20	MC	1	52506	0.55	0.37	0.00		0.12	0.15	0.55	0.18		-0.28	-0.13	0.37	-0.12

Table F-2. Item Statistics, ELA Grade 4 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	52500	0.63	0.37	0.00		0.14	0.63	0.11	0.12		-0.21	0.37	-0.17	-0.17
22	EBSR	2	52526	0.35	0.33	0.00	0.53	0.24	0.22			-0.27	0.00	0.32		
23	MC	1	52473	0.71	0.45	0.00		0.10	0.71	0.10	0.10		-0.28	0.45	-0.25	-0.17
24	TE	2	52247	0.55	0.42	0.01	0.18	0.53	0.29			-0.33	-0.04	0.34		
25	MC	1	52402	0.49	0.40	0.00		0.49	0.29	0.11	0.12		0.40	-0.14	-0.23	-0.21
26	MC	1	52412	0.57	0.52	0.00		0.12	0.19	0.12	0.57		-0.23	-0.21	-0.30	0.52
27	MS	2	52399	0.43	0.44	0.00	0.36	0.41	0.22			-0.34	0.00	0.40		
28	EBSR	2	52482	0.55	0.53	0.00	0.24	0.42	0.34			-0.46	-0.02	0.44		
29	MC	1	52433	0.58	0.34	0.00		0.10	0.12	0.19	0.58		-0.13	-0.24	-0.12	0.34
30	MC	1	52431	0.61	0.50	0.00		0.09	0.11	0.19	0.61		-0.21	-0.30	-0.23	0.50
31	MC	1	52416	0.41	0.21	0.00		0.41	0.13	0.34	0.12		0.22	-0.20	0.07	-0.22
32	MC	1	52414	0.51	0.37	0.00		0.20	0.16	0.51	0.12		-0.18	-0.15	0.37	-0.17
33	MC	1	52378	0.52	0.37	0.00		0.52	0.20	0.13	0.16		0.37	-0.19	-0.24	-0.08
34	TE	2	52410	0.44	0.42	0.00	0.33	0.46	0.20			-0.27	-0.10	0.44		
35	MC	1	52412	0.58	0.39	0.00		0.11	0.09	0.58	0.22		-0.15	-0.25	0.39	-0.17
36	MC	1	52371	0.60	0.50	0.00		0.60	0.14	0.15	0.10		0.50	-0.22	-0.24	-0.25
37	MC	1	52354	0.53	0.39	0.00		0.53	0.23	0.13	0.10		0.39	-0.18	-0.22	-0.14
38	MC	1	52365	0.65	0.37	0.00		0.20	0.64	0.10	0.05		-0.13	0.37	-0.24	-0.22
39	MC	1	52380	0.58	0.44	0.00		0.14	0.17	0.11	0.58		-0.21	-0.18	-0.24	0.44
40	TE	2	52372	0.47	0.44	0.00	0.24	0.59	0.17			-0.36	0.06	0.34		
41	MC	1	52378	0.60	0.45	0.00		0.08	0.14	0.60	0.18		-0.22	-0.17	0.45	-0.26

Table F-3. Item Statistics, ELA Grade 5

							Proportion of Students					Item-Total Test Correlation				
Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	53862	0.78	0.41	0.00		0.15	0.78	0.04	0.03		-0.32	0.41	-0.17	-0.13
2	MC	1	53840	0.66	0.36	0.00		0.05	0.14	0.15	0.66		-0.18	-0.15	-0.21	0.36
3	TE	2	53840	0.57	0.44	0.00	0.19	0.48	0.33			-0.33	-0.10	0.39		
4	MC	1	53840	0.42	0.24	0.00		0.42	0.10	0.24	0.24		0.24	-0.16	-0.02	-0.14
5	TE	1	53830	0.67	0.36	0.00	0.33	0.67				-0.35	0.36			
6	MC	1	53838	0.80	0.33	0.00		0.10	0.03	0.07	0.80		-0.19	-0.18	-0.17	0.33
7	MC	1	53833	0.75	0.27	0.00		0.06	0.12	0.07	0.75		-0.09	-0.10	-0.24	0.27
8	TE	2	53671	0.49	0.49	0.00	0.36	0.29	0.34			-0.34	-0.18	0.53		
9	MC	1	53832	0.66	0.34	0.00		0.08	0.09	0.66	0.16		-0.17	-0.22	0.34	-0.13
10	TE	1	53733	0.71	0.54	0.00	0.29	0.71				-0.53	0.54			
11	MC	1	53800	0.61	0.44	0.00		0.24	0.07	0.61	0.08		-0.20	-0.28	0.44	-0.20
12	MC	1	53811	0.53	0.36	0.00		0.29	0.53	0.08	0.10		-0.15	0.36	-0.20	-0.18
13	TE	2	53815	0.72	0.41	0.00	0.12	0.32	0.56			-0.36	-0.10	0.33		
14	MC	1	53806	0.57	0.32	0.00		0.04	0.31	0.07	0.57		-0.26	-0.12	-0.19	0.32
15	TE	2	53724	0.58	0.35	0.00	0.15	0.54	0.30			-0.19	-0.19	0.36		
16	TE	2	53835	0.60	0.43	0.00	0.14	0.53	0.33			-0.34	-0.09	0.35		
17	MC	1	53802	0.61	0.34	0.00		0.61	0.30	0.05	0.04		0.34	-0.20	-0.16	-0.19
18	MC	1	53807	0.77	0.44	0.00		0.16	0.04	0.04	0.76		-0.29	-0.22	-0.20	0.44
19	MC	1	53809	0.61	0.41	0.00		0.14	0.17	0.61	0.08		-0.31	-0.09	0.41	-0.21
20	MC	1	53806	0.61	0.37	0.00		0.10	0.17	0.11	0.61		-0.16	-0.17	-0.22	0.37

Table F-3. Item Statistics, ELA Grade 5 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	EBSR	2	53831	0.58	0.53	0.00	0.28	0.29	0.43			-0.45	-0.07	0.47		
22	MC	1	53789	0.69	0.48	0.00		0.11	0.69	0.06	0.14		-0.21	0.48	-0.25	-0.27
23	MC	1	53738	0.57	0.18	0.00		0.10	0.04	0.57	0.28		-0.17	-0.26	0.18	0.03
24	MC	1	53715	0.67	0.42	0.00		0.17	0.06	0.10	0.67		-0.14	-0.26	-0.27	0.42
25	TE	1	53607	0.84	0.49	0.00	0.16	0.83				-0.49	0.50			
26	MC	1	53746	0.53	0.32	0.00		0.11	0.53	0.14	0.21		-0.14	0.32	-0.22	-0.09
27	TE	1	53584	0.67	0.56	0.00	0.33	0.66				-0.55	0.56			
28	MC	1	53721	0.56	0.43	0.00		0.56	0.10	0.21	0.12		0.43	-0.24	-0.21	-0.15
29	MC	1	53730	0.54	0.22	0.00		0.25	0.06	0.14	0.54		0.01	-0.25	-0.15	0.22
30	MC	1	53723	0.52	0.26	0.00		0.52	0.18	0.08	0.22		0.26	-0.16	-0.19	-0.03
31	MS	2	53701	0.58	0.50	0.00	0.19	0.46	0.35			-0.37	-0.13	0.44		
32	MC	1	53711	0.61	0.40	0.00		0.22	0.61	0.11	0.06		-0.13	0.40	-0.29	-0.20
33	MC	1	53696	0.55	0.48	0.00		0.55	0.13	0.19	0.12		0.48	-0.29	-0.18	-0.21
34	MC	1	53729	0.65	0.54	0.00		0.13	0.06	0.15	0.65		-0.32	-0.29	-0.22	0.54
35	MC	1	53722	0.55	0.45	0.00		0.55	0.14	0.11	0.19		0.45	-0.22	-0.28	-0.14
36	MC	1	53693	0.42	0.28	0.00		0.42	0.18	0.32	0.08		0.28	-0.20	-0.05	-0.14
37	MC	1	53670	0.60	0.45	0.00		0.22	0.09	0.10	0.59		-0.21	-0.25	-0.21	0.45
38	TE	2	53694	0.62	0.53	0.00	0.16	0.45	0.39			-0.38	-0.20	0.49		
39	TE	2	53508	0.64	0.29	0.01	0.07	0.56	0.36			-0.28	-0.03	0.20		
40	MC	1	53698	0.43	0.45	0.00		0.17	0.28	0.12	0.43		-0.24	-0.09	-0.27	0.45
41	MC	1	53685	0.46	0.34	0.00		0.18	0.14	0.46	0.22		-0.13	-0.16	0.34	-0.15
42	MC	1	53686	0.58	0.44	0.00		0.17	0.15	0.58	0.10		-0.21	-0.23	0.44	-0.19

Table F-4. Item Statistics, ELA Grade 6

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	55359	0.66	0.21	0.00		0.66	0.20	0.10	0.03		0.21	-0.08	-0.14	-0.14
2	MC	1	55334	0.65	0.27	0.00		0.65	0.06	0.14	0.14		0.27	-0.14	-0.15	-0.12
3	MC	1	55322	0.70	0.45	0.00		0.03	0.16	0.70	0.11		-0.18	-0.29	0.45	-0.21
4	MC	1	55310	0.34	0.19	0.00		0.15	0.34	0.23	0.28		-0.09	0.19	-0.16	0.03
5	MC	1	55326	0.43	0.37	0.00		0.22	0.21	0.14	0.43		-0.23	-0.08	-0.17	0.37
6	MC	1	55316	0.39	0.22	0.00		0.12	0.39	0.16	0.32		-0.07	0.22	-0.17	-0.05
7	TE	1	55194	0.42	0.47	0.00	0.58	0.42				-0.47	0.48			
8	TE	1	55214	0.43	0.27	0.00	0.57	0.43				-0.26	0.27			
9	TE	2	55314	0.67	0.42	0.00	0.11	0.45	0.44			-0.34	-0.14	0.35		
10	MC	1	55313	0.44	0.31	0.00		0.44	0.07	0.24	0.26		0.31	-0.22	-0.12	-0.10
11	TE	2	55211	0.56	0.51	0.00	0.28	0.31	0.41			-0.40	-0.12	0.48		
12	TE	2	55202	0.65	0.37	0.00	0.07	0.57	0.36			-0.31	-0.11	0.28		
13	TE	2	55281	0.62	0.43	0.00	0.15	0.46	0.39			-0.37	-0.08	0.35		
14	MC	1	55261	0.46	0.27	0.00		0.13	0.17	0.23	0.46		-0.14	-0.15	-0.07	0.28
15	MC	1	55273	0.47	0.34	0.00		0.14	0.27	0.13	0.47		-0.13	-0.13	-0.19	0.34
16	MC	1	55246	0.54	0.35	0.00		0.08	0.23	0.53	0.15		-0.16	-0.33	0.35	0.02
17	MC	1	55216	0.53	0.32	0.00		0.07	0.53	0.26	0.14		-0.12	0.32	-0.22	-0.08
18	TE	2	55219	0.59	0.41	0.00	0.18	0.48	0.35			-0.31	-0.11	0.36		
19	EBSR	2	55248	0.74	0.51	0.00	0.15	0.22	0.63			-0.37	-0.26	0.50		
20	MC	1	55199	0.61	0.39	0.00		0.16	0.61	0.08	0.14		-0.12	0.39	-0.25	-0.21

Table F-4. Item Statistics, ELA Grade 6 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	55184	0.69	0.33	0.00		0.11	0.15	0.05	0.69		-0.07	-0.20	-0.27	0.33
22	MC	1	55199	0.44	0.18	0.00		0.25	0.27	0.03	0.44		-0.08	-0.02	-0.26	0.18
23	TE	2	55131	0.61	0.46	0.00	0.14	0.49	0.37			-0.37	-0.10	0.37		
24	MC	1	55102	0.51	0.35	0.00		0.08	0.51	0.21	0.19		-0.21	0.35	-0.23	-0.06
25	TE	2	55090	0.57	0.44	0.00	0.16	0.53	0.30			-0.38	-0.03	0.34		
26	MC	1	55083	0.55	0.32	0.00		0.26	0.09	0.55	0.10		-0.14	-0.19	0.33	-0.15
27	MC	1	55095	0.80	0.50	0.00		0.07	0.04	0.08	0.80		-0.27	-0.27	-0.27	0.50
28	TE	2	55078	0.54	0.41	0.00	0.18	0.54	0.27			-0.29	-0.09	0.36		
29	MC	1	55099	0.62	0.44	0.00		0.06	0.22	0.11	0.61		-0.19	-0.15	-0.34	0.44
30	TE	2	55101	0.63	0.42	0.00	0.13	0.48	0.39			-0.31	-0.14	0.36		
31	TE	1	54939	0.66	0.41	0.01	0.34	0.66				-0.41	0.42			
32	MC	1	55055	0.43	0.25	0.00		0.16	0.43	0.28	0.14		-0.06	0.26	-0.14	-0.11
33	MC	1	55072	0.53	0.43	0.00		0.53	0.24	0.13	0.11		0.44	-0.16	-0.21	-0.25
34	TE	2	55060	0.56	0.22	0.00	0.15	0.59	0.26			-0.11	-0.11	0.23		
35	EBSR	2	55113	0.53	0.62	0.00	0.34	0.25	0.40			-0.54	-0.03	0.56		
36	TE	1	54954	0.59	0.48	0.01	0.41	0.59				-0.47	0.48			
37	MC	1	55060	0.43	0.40	0.00		0.43	0.24	0.21	0.12		0.41	-0.12	-0.28	-0.11
38	MC	1	55040	0.57	0.41	0.00		0.11	0.57	0.18	0.14		-0.18	0.41	-0.24	-0.15
39	MC	1	55050	0.67	0.50	0.00		0.66	0.10	0.10	0.13		0.50	-0.20	-0.28	-0.26

Table F-5. Item Statistics, ELA Grade 7

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	56126	0.68	0.43	0.00		0.07	0.68	0.21	0.04		-0.22	0.43	-0.28	-0.14
2	TE	2	56111	0.51	0.39	0.00	0.26	0.46	0.28			-0.30	-0.04	0.34		
3	TE	2	56063	0.70	0.47	0.00	0.13	0.34	0.53			-0.28	-0.30	0.48		
4	MC	1	56106	0.71	0.38	0.00		0.06	0.71	0.13	0.10		-0.24	0.38	-0.20	-0.16
5	MS	2	56103	0.53	0.24	0.00	0.21	0.54	0.26			-0.16	-0.06	0.23		
6	MC	1	56092	0.39	0.28	0.00		0.09	0.20	0.32	0.39		-0.12	-0.10	-0.13	0.28
7	TE	2	56057	0.64	0.31	0.00	0.07	0.57	0.35			-0.16	-0.20	0.30		
8	MC	1	56092	0.60	0.46	0.00		0.60	0.22	0.10	0.08		0.46	-0.28	-0.22	-0.15
9	MC	1	56106	0.70	0.34	0.00		0.12	0.07	0.11	0.69		-0.17	-0.18	-0.18	0.35
10	MC	1	56101	0.59	0.38	0.00		0.11	0.05	0.59	0.25		-0.25	-0.25	0.38	-0.12
11	TE	2	56073	0.44	0.20	0.00	0.27	0.59	0.14			-0.15	0.02	0.17		
12	MC	1	56052	0.45	0.29	0.00		0.45	0.23	0.20	0.11		0.29	-0.14	-0.13	-0.10
13	MC	1	56084	0.69	0.37	0.00		0.17	0.08	0.69	0.06		-0.12	-0.23	0.37	-0.26
14	MC	1	56079	0.63	0.42	0.00		0.07	0.08	0.21	0.63		-0.19	-0.32	-0.15	0.42
15	EBSR	2	56085	0.75	0.43	0.00	0.19	0.12	0.69			-0.35	-0.22	0.45		
16	MS	2	56016	0.75	0.50	0.00	0.10	0.30	0.60			-0.34	-0.29	0.48		
17	MC	1	56044	0.58	0.34	0.00		0.06	0.20	0.16	0.58		-0.17	-0.16	-0.18	0.34
18	MC	1	56038	0.57	0.29	0.00		0.09	0.17	0.17	0.57		-0.07	-0.21	-0.11	0.29
19	EBSR	2	56067	0.38	0.39	0.00	0.52	0.21	0.27			-0.30	-0.08	0.41		
20	TE	1	55880	0.73	0.37	0.00	0.27	0.73				-0.37	0.38			

Table F-5. Item Statistics, ELA Grade 7 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	55963	0.60	0.53	0.00		0.13	0.20	0.07	0.60		-0.30	-0.22	-0.27	0.53
22	EBSR	2	56029	0.51	0.39	0.00	0.41	0.16	0.43			-0.31	-0.13	0.41		
23	MC	1	55934	0.51	0.34	0.00		0.22	0.51	0.17	0.10		-0.10	0.34	-0.14	-0.25
24	MC	1	55915	0.50	0.43	0.00		0.08	0.19	0.23	0.50		-0.27	-0.09	-0.26	0.43
25	TE	1	55129	0.54	0.47	0.02	0.46	0.53				-0.44	0.47			
26	TE	2	55893	0.65	0.56	0.00	0.16	0.38	0.46			-0.45	-0.15	0.48		
27	TE	2	55907	0.53	0.35	0.00	0.18	0.57	0.24			-0.35	0.08	0.22		
28	MC	1	55865	0.40	0.29	0.00		0.27	0.40	0.13	0.21		-0.12	0.29	-0.25	0.00
29	EBSR	2	55964	0.46	0.44	0.00	0.42	0.24	0.34			-0.37	-0.02	0.41		
30	TE	1	55729	0.34	0.44	0.01	0.66	0.33				-0.42	0.44			
31	MC	1	55921	0.71	0.55	0.00		0.06	0.16	0.07	0.70		-0.25	-0.30	-0.30	0.55
32	MC	1	55879	0.43	0.30	0.00		0.25	0.20	0.43	0.12		-0.16	-0.11	0.30	-0.11
33	MC	1	55864	0.54	0.38	0.00		0.16	0.21	0.54	0.09		-0.15	-0.17	0.38	-0.22
34	MC	1	55843	0.49	0.46	0.00		0.49	0.19	0.15	0.16		0.46	-0.25	-0.26	-0.09
35	MS	2	55861	0.58	0.49	0.00	0.16	0.50	0.33			-0.39	-0.08	0.40		
36	MS	2	55883	0.73	0.46	0.00	0.08	0.39	0.53			-0.32	-0.24	0.41		
37	MC	1	55873	0.57	0.44	0.00		0.13	0.57	0.16	0.14		-0.18	0.44	-0.25	-0.18

Table F-6. Item Statistics, ELA Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	56546	0.71	0.41	0.00		0.08	0.02	0.18	0.71		-0.20	-0.21	-0.25	0.41
2	MC	1	56508	0.60	0.44	0.00		0.20	0.60	0.11	0.09		-0.20	0.44	-0.25	-0.21
3	MC	1	56491	0.68	0.43	0.00		0.09	0.12	0.11	0.68		-0.18	-0.23	-0.23	0.43
4	MC	1	56498	0.55	0.35	0.00		0.15	0.55	0.11	0.19		-0.22	0.35	-0.23	-0.06
5	MC	1	56494	0.58	0.42	0.00		0.58	0.20	0.04	0.18		0.42	-0.19	-0.22	-0.22
6	MS	2	56471	0.60	0.46	0.00	0.16	0.47	0.36			-0.37	-0.09	0.38		
7	MC	1	56466	0.54	0.28	0.00		0.29	0.53	0.10	0.08		-0.11	0.28	-0.21	-0.09
8	TE	2	56435	0.67	0.47	0.00	0.10	0.47	0.43			-0.34	-0.20	0.41		
9	MC	1	56465	0.51	0.51	0.00		0.51	0.12	0.22	0.15		0.51	-0.19	-0.27	-0.22
10	MC	1	56472	0.78	0.34	0.00		0.08	0.78	0.03	0.11		-0.16	0.34	-0.22	-0.19
11	TE	2	56399	0.45	0.29	0.00	0.27	0.54	0.18			-0.19	-0.03	0.28		
12	MC	1	56445	0.40	0.31	0.00		0.10	0.40	0.05	0.44		-0.21	0.31	-0.23	-0.07
13	TE	2	56449	0.86	0.37	0.00	0.04	0.20	0.75			-0.23	-0.27	0.36		
14	MC	1	56426	0.55	0.44	0.00		0.15	0.11	0.18	0.55		-0.26	-0.22	-0.13	0.44
15	MC	1	56412	0.53	0.38	0.00		0.13	0.09	0.25	0.52		-0.13	-0.28	-0.14	0.38
16	MS	2	56464	0.66	0.46	0.00	0.09	0.50	0.41			-0.26	-0.28	0.44		
17	EBSR	2	56468	0.69	0.50	0.00	0.13	0.37	0.50			-0.40	-0.18	0.44		
18	MC	1	56433	0.71	0.48	0.00		0.09	0.13	0.71	0.07		-0.26	-0.25	0.48	-0.22
19	MC	1	56426	0.51	0.37	0.00		0.51	0.09	0.24	0.16		0.38	-0.15	-0.20	-0.16
20	EBSR	2	56448	0.45	0.39	0.00	0.50	0.11	0.39			-0.33	-0.10	0.41		

Table F-6. Item Statistics, ELA Grade 8 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	56420	0.38	0.34	0.00		0.16	0.29	0.17	0.38		-0.26	-0.08	-0.09	0.34
22	MC	1	56363	0.46	0.18	0.00		0.07	0.46	0.24	0.23		-0.27	0.19	0.02	-0.07
23	MC	1	56349	0.70	0.43	0.00		0.69	0.09	0.05	0.16		0.43	-0.26	-0.27	-0.16
24	MC	1	56316	0.53	0.33	0.00		0.53	0.15	0.18	0.13		0.33	-0.04	-0.22	-0.19
25	MS	2	56331	0.60	0.46	0.00	0.15	0.50	0.35			-0.34	-0.13	0.40		
26	MC	1	56313	0.37	0.44	0.00		0.27	0.26	0.10	0.37		-0.10	-0.21	-0.25	0.44
27	MC	1	56354	0.60	0.49	0.00		0.03	0.06	0.30	0.60		-0.20	-0.30	-0.28	0.49
28	MC	1	56248	0.50	0.27	0.00		0.15	0.50	0.19	0.16		-0.14	0.27	-0.16	-0.05
29	MC	1	56301	0.74	0.54	0.00		0.74	0.10	0.09	0.07		0.54	-0.24	-0.33	-0.27
30	MC	1	56313	0.45	0.21	0.00		0.07	0.33	0.15	0.44		-0.16	-0.10	-0.03	0.21
31	EBSR	2	56376	0.59	0.52	0.00	0.31	0.20	0.49			-0.42	-0.17	0.52		
32	MC	1	56285	0.55	0.40	0.00		0.55	0.18	0.21	0.06		0.40	-0.25	-0.10	-0.25
33	EBSR	2	56352	0.58	0.61	0.00	0.26	0.32	0.42			-0.57	0.02	0.50		
34	MC	1	56236	0.78	0.53	0.00		0.06	0.78	0.09	0.07		-0.23	0.53	-0.32	-0.27
35	MC	1	56283	0.69	0.43	0.00		0.09	0.09	0.68	0.14		-0.16	-0.19	0.43	-0.28
36	MC	1	56274	0.38	0.21	0.00		0.30	0.22	0.10	0.38		0.06	-0.16	-0.20	0.21
37	MC	1	56294	0.69	0.46	0.00		0.69	0.19	0.08	0.03		0.47	-0.26	-0.28	-0.18
38	MC	1	56288	0.53	0.42	0.00		0.15	0.17	0.52	0.15		-0.17	-0.26	0.42	-0.14
39	MS	2	56285	0.68	0.59	0.00	0.14	0.36	0.50			-0.40	-0.28	0.56		
40	MC	1	56292	0.60	0.39	0.00		0.04	0.15	0.20	0.60		-0.22	-0.12	-0.25	0.39

Table F-7. Item Statistics, Mathematics Grade 3

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TE	1	52778	0.68	0.55	0.00	0.32	0.68				-0.55	0.55			
2	SA	1	52755	0.51	0.57	0.00	0.49	0.51				-0.56	0.57			
3	TE	1	52787	0.40	0.61	0.00	0.60	0.40				-0.60	0.61			
4	MC	1	52779	0.77	0.44	0.00		0.04	0.07	0.77	0.11		-0.12	-0.31	0.45	-0.26
5	MC	1	52758	0.50	0.49	0.00		0.26	0.05	0.19	0.50		-0.29	-0.06	-0.26	0.49
6	SA	1	52733	0.78	0.45	0.00	0.22	0.78				-0.45	0.45			
7	MC	1	52762	0.40	0.33	0.00		0.13	0.28	0.19	0.40		-0.28	0.04	-0.21	0.33
8	SA	1	52679	0.41	0.59	0.00	0.59	0.41				-0.58	0.59			
9	MC	1	52731	0.66	0.41	0.00		0.66	0.12	0.13	0.09		0.41	-0.26	-0.14	-0.22
10	SA	1	52716	0.63	0.62	0.00	0.37	0.63				-0.62	0.62			
11	MC	1	52742	0.39	0.35	0.00		0.22	0.21	0.39	0.18		-0.09	-0.13	0.36	-0.21
12	TE	1	52281	0.19	0.37	0.01	0.80	0.19				-0.34	0.37			
13	MC	1	52748	0.57	0.44	0.00		0.57	0.11	0.22	0.10		0.44	-0.13	-0.26	-0.23
14	SA	1	52682	0.60	0.57	0.00	0.39	0.60				-0.57	0.58			
15	SA	1	52653	0.32	0.52	0.00	0.68	0.32				-0.51	0.52			
16	MC	1	52713	0.32	0.39	0.00		0.32	0.15	0.22	0.31		0.39	0.03	-0.16	-0.27
17	MC	1	52737	0.77	0.51	0.00		0.77	0.14	0.06	0.03		0.51	-0.36	-0.22	-0.20
18	TE	1	52076	0.83	0.43	0.01	0.17	0.82				-0.42	0.44			
19	MC	1	52752	0.71	0.49	0.00		0.05	0.09	0.15	0.71		-0.22	-0.25	-0.28	0.49
20	MC	1	52702	0.57	0.43	0.00		0.20	0.57	0.10	0.13		-0.21	0.43	-0.15	-0.24

Table F-7. Item Statistics, Mathematics Grade 3 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	SA	1	52681	0.57	0.57	0.00	0.43	0.57				-0.56	0.57			
22	MC	1	52702	0.46	0.36	0.00		0.46	0.34	0.08	0.13		0.36	-0.02	-0.18	-0.37
23	MC	1	52698	0.75	0.51	0.00		0.12	0.75	0.09	0.05		-0.26	0.51	-0.32	-0.23
24	SA	1	52684	0.58	0.61	0.00	0.42	0.58				-0.61	0.61			
25	TE	1	52683	0.59	0.35	0.00	0.40	0.59				-0.35	0.35			
26	TE	1	52475	0.41	0.58	0.01	0.58	0.41				-0.57	0.58			
27	MC	1	52622	0.58	0.45	0.00		0.17	0.13	0.57	0.12		-0.24	-0.24	0.45	-0.15
28	MC	1	52626	0.39	0.35	0.00		0.15	0.16	0.30	0.39		-0.25	-0.18	-0.04	0.35
29	MC	1	52670	0.56	0.37	0.00		0.15	0.16	0.56	0.13		-0.20	-0.17	0.37	-0.14
30	MC	1	52664	0.40	0.23	0.00		0.21	0.22	0.18	0.40		-0.09	-0.03	-0.17	0.23
31	MC	1	52621	0.43	0.35	0.00		0.18	0.21	0.18	0.42		-0.12	-0.08	-0.24	0.35
32	MC	1	52638	0.41	0.37	0.00		0.29	0.15	0.15	0.41		-0.10	-0.20	-0.18	0.37
33	SA	1	52606	0.47	0.59	0.00	0.53	0.47				-0.59	0.59			
34	MC	1	52657	0.66	0.47	0.00		0.66	0.09	0.10	0.16		0.47	-0.21	-0.19	-0.29
35	SA	1	52625	0.60	0.65	0.00	0.39	0.60				-0.65	0.65			
36	TE	1	52327	0.23	0.49	0.01	0.77	0.23				-0.47	0.49			
37	MC	1	52620	0.47	0.48	0.00		0.19	0.15	0.19	0.47		-0.25	-0.23	-0.14	0.48
38	MC	1	52643	0.69	0.40	0.00		0.11	0.03	0.17	0.69		-0.11	-0.10	-0.36	0.40
39	SA	1	52608	0.54	0.64	0.00	0.46	0.54				-0.64	0.64			
40	MC	1	52643	0.65	0.50	0.00		0.17	0.65	0.12	0.05		-0.29	0.51	-0.22	-0.27
41	TE	1	51226	0.31	0.46	0.03	0.67	0.30				-0.39	0.46			
42	MC	1	52635	0.68	0.48	0.00		0.09	0.10	0.14	0.67		-0.29	-0.22	-0.22	0.48

Table F-8. Item Statistics, Mathematics Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	52541	0.67	0.43	0.00		0.67	0.17	0.07	0.09		0.43	-0.27	-0.19	-0.18
2	MC	1	52507	0.44	0.28	0.00		0.06	0.43	0.07	0.44		-0.21	-0.09	-0.18	0.28
3	SA	1	52521	0.41	0.42	0.00	0.59	0.41				-0.42	0.43			
4	MC	1	52522	0.82	0.42	0.00		0.12	0.82	0.05	0.02		-0.30	0.42	-0.23	-0.14
5	TE	1	52373	0.39	0.54	0.00	0.60	0.39				-0.53	0.54			
6	MC	1	52512	0.50	0.51	0.00		0.50	0.14	0.18	0.18		0.51	-0.15	-0.26	-0.26
7	MC	1	52536	0.34	0.50	0.00		0.02	0.34	0.16	0.48		-0.06	0.50	-0.09	-0.39
8	SA	1	52153	0.35	0.62	0.01	0.65	0.35				-0.60	0.62			
9	MC	1	52507	0.69	0.50	0.00		0.69	0.10	0.15	0.06		0.50	-0.27	-0.24	-0.26
10	MC	1	52491	0.85	0.41	0.00		0.02	0.08	0.85	0.05		-0.16	-0.31	0.41	-0.18
11	MC	1	52491	0.28	0.33	0.00		0.11	0.28	0.35	0.25		-0.03	0.33	-0.26	-0.03
12	MC	1	52481	0.34	0.31	0.00		0.27	0.27	0.34	0.11		-0.14	-0.12	0.31	-0.09
13	MC	1	52506	0.49	0.60	0.00		0.20	0.12	0.19	0.49		-0.26	-0.31	-0.24	0.60
14	MC	1	52487	0.43	0.27	0.00		0.13	0.34	0.10	0.42		0.01	-0.25	-0.06	0.27
15	MC	1	52501	0.26	0.11	0.00		0.26	0.31	0.25	0.17		0.11	-0.14	0.00	0.04
16	SA	1	52149	0.33	0.52	0.01	0.67	0.33				-0.50	0.52			
17	MC	1	52493	0.81	0.45	0.00		0.05	0.81	0.10	0.04		-0.21	0.45	-0.30	-0.21
18	MC	1	52460	0.42	0.48	0.00		0.15	0.25	0.18	0.42		-0.27	-0.12	-0.23	0.48
19	TE	1	52472	0.33	0.59	0.00	0.67	0.32				-0.58	0.59			
20	MC	1	52475	0.45	0.38	0.00		0.18	0.44	0.16	0.22		-0.14	0.38	-0.21	-0.14
21	MC	1	52497	0.88	0.32	0.00		0.01	0.08	0.88	0.03		-0.11	-0.24	0.33	-0.16
22	MC	1	52490	0.28	0.16	0.00		0.21	0.28	0.17	0.34		-0.18	0.16	-0.16	0.13
23	SA	1	52434	0.21	0.54	0.00	0.79	0.21				-0.53	0.54			

Table F-8. Item Statistics, Mathematics Grade 4 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	TE	1	52223	0.69	0.37	0.01	0.31	0.69				-0.37	0.38			
25	MC	1	52421	0.84	0.36	0.00		0.07	0.05	0.84	0.04		-0.26	-0.17	0.36	-0.15
26	MC	1	52410	0.48	0.49	0.00		0.26	0.48	0.15	0.11		-0.35	0.49	-0.17	-0.09
27	MC	1	52411	0.29	0.57	0.00		0.29	0.07	0.49	0.14		0.57	-0.08	-0.25	-0.33
28	MC	1	52430	0.46	0.41	0.00		0.12	0.20	0.21	0.46		-0.18	-0.14	-0.21	0.41
29	SA	1	52322	0.32	0.54	0.00	0.68	0.31				-0.53	0.54			
30	MC	1	52413	0.30	0.40	0.00		0.30	0.13	0.15	0.42		0.40	-0.13	-0.07	-0.23
31	MC	1	52211	0.38	0.14	0.01		0.10	0.22	0.30	0.38		-0.10	-0.06	-0.02	0.15
32	TE	1	51990	0.20	0.48	0.01	0.80	0.19				-0.45	0.48			
33	MC	1	52409	0.50	0.59	0.00		0.50	0.36	0.08	0.07		0.59	-0.51	-0.09	-0.10
34	MC	1	52401	0.39	0.43	0.00		0.07	0.47	0.06	0.39		-0.25	-0.19	-0.22	0.43
35	SA	1	52297	0.16	0.46	0.00	0.83	0.16				-0.45	0.46			
36	MC	1	52369	0.53	0.42	0.00		0.53	0.17	0.15	0.15		0.42	-0.25	-0.19	-0.12
37	MC	1	52393	0.83	0.35	0.00		0.04	0.06	0.06	0.83		-0.19	-0.22	-0.15	0.35
38	MC	1	52402	0.33	0.54	0.00		0.32	0.19	0.28	0.20		0.54	-0.07	-0.21	-0.32
39	TE	1	52230	0.55	0.63	0.00	0.44	0.55				-0.62	0.63			
40	SA	1	52365	0.53	0.45	0.00	0.47	0.53				-0.45	0.45			
41	MC	1	52375	0.57	0.40	0.00		0.24	0.57	0.13	0.06		-0.27	0.40	-0.17	-0.10
42	MC	1	52384	0.31	0.09	0.00		0.43	0.16	0.09	0.31		-0.11	0.09	-0.08	0.10
43	SA	1	52303	0.45	0.47	0.00	0.55	0.45				-0.46	0.47			
44	MC	1	52354	0.27	0.42	0.00		0.27	0.21	0.22	0.29		0.42	-0.10	-0.14	-0.19
45	SA	1	52310	0.42	0.56	0.00	0.58	0.42				-0.55	0.56			
46	MC	1	52355	0.49	0.40	0.00		0.49	0.20	0.24	0.08		0.40	-0.20	-0.18	-0.15

Table F-9. Item Statistics, Mathematics Grade 5

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	SA	1	53796	0.61	0.50	0.00	0.39	0.61				-0.50	0.50			
2	MC	1	53788	0.42	0.49	0.00		0.42	0.51	0.05	0.02		0.49	-0.35	-0.24	-0.09
3	MC	1	53761	0.56	0.15	0.00		0.56	0.25	0.15	0.04		0.15	0.05	-0.21	-0.09
4	MC	1	53762	0.31	0.39	0.00		0.12	0.37	0.20	0.31		-0.22	-0.16	-0.08	0.39
5	MC	1	53756	0.42	0.34	0.00		0.15	0.26	0.41	0.17		-0.19	-0.17	0.34	-0.06
6	TE	1	53380	0.33	0.47	0.01	0.67	0.32				-0.45	0.47			
7	MC	1	53801	0.58	0.51	0.00		0.13	0.58	0.05	0.24		-0.10	0.51	-0.15	-0.43
8	MC	1	53626	0.42	0.34	0.00		0.42	0.23	0.11	0.24		0.34	-0.15	-0.20	-0.09
9	TE	1	53279	0.28	0.54	0.01	0.72	0.27				-0.52	0.54			
10	SA	1	53673	0.42	0.54	0.00	0.58	0.42				-0.53	0.54			
11	MC	1	53735	0.32	0.22	0.00		0.17	0.32	0.20	0.31		-0.17	0.22	-0.15	0.04
12	SA	1	53508	0.28	0.48	0.01	0.72	0.28				-0.46	0.48			
13	MC	1	53770	0.58	0.59	0.00		0.20	0.11	0.10	0.58		-0.36	-0.26	-0.19	0.59
14	SA	1	53659	0.37	0.57	0.00	0.63	0.37				-0.56	0.57			
15	MC	1	53749	0.41	0.51	0.00		0.34	0.15	0.10	0.41		-0.27	-0.19	-0.19	0.51
16	SA	1	53540	0.31	0.33	0.01	0.69	0.31				-0.32	0.33			
17	MC	1	53729	0.47	0.33	0.00		0.47	0.22	0.15	0.15		0.33	-0.03	-0.25	-0.17
18	SA	1	53689	0.44	0.45	0.00	0.56	0.44				-0.45	0.45			
19	MC	1	53723	0.63	0.53	0.00		0.07	0.14	0.15	0.63		-0.24	-0.24	-0.30	0.53
20	MC	1	53716	0.39	0.25	0.00		0.09	0.39	0.33	0.19		0.00	0.25	-0.14	-0.14
21	SA	1	53629	0.17	0.43	0.00	0.83	0.17				-0.41	0.43			
22	TE	1	53639	0.51	0.52	0.00	0.49	0.50				-0.51	0.52			
23	MC	1	53669	0.40	0.49	0.00		0.40	0.13	0.21	0.25		0.49	-0.17	-0.28	-0.15

Table F-9. Item Statistics, Mathematics Grade 5 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	53681	0.51	0.30	0.00		0.15	0.20	0.51	0.14		-0.15	-0.08	0.30	-0.19
25	SA	1	53566	0.40	0.52	0.00	0.59	0.40				-0.52	0.52			
26	MC	1	53656	0.37	0.36	0.00		0.25	0.24	0.37	0.13		-0.07	-0.11	0.36	-0.29
27	SA	1	53593	0.24	0.45	0.00	0.76	0.24				-0.45	0.46			
28	MC	1	53652	0.47	0.42	0.00		0.47	0.12	0.15	0.26		0.42	-0.22	-0.23	-0.13
29	TE	1	53483	0.41	0.55	0.00	0.59	0.41				-0.54	0.55			
30	MC	1	53677	0.56	0.31	0.00		0.08	0.26	0.10	0.56		-0.19	-0.03	-0.30	0.31
31	TE	1	53482	0.49	0.58	0.00	0.51	0.48				-0.57	0.58			
32	MC	1	53636	0.73	0.36	0.00		0.10	0.12	0.73	0.05		-0.17	-0.24	0.36	-0.15
33	MC	1	53642	0.39	0.25	0.00		0.31	0.39	0.16	0.14		-0.19	0.25	-0.20	0.11
34	MC	1	53529	0.47	0.40	0.00		0.17	0.20	0.47	0.16		-0.26	-0.08	0.40	-0.19
35	TE	1	53539	0.41	0.59	0.00	0.59	0.41				-0.58	0.59			
36	MC	1	53626	0.31	0.44	0.00		0.41	0.18	0.31	0.10		-0.37	-0.03	0.44	-0.03
37	SA	1	53413	0.41	0.54	0.01	0.59	0.41				-0.53	0.54			
38	MC	1	53651	0.44	0.40	0.00		0.34	0.12	0.10	0.43		-0.13	-0.24	-0.20	0.41
39	MC	1	53513	0.55	0.44	0.00		0.08	0.18	0.19	0.54		-0.21	-0.19	-0.23	0.44
40	TE	1	53558	0.27	0.44	0.00	0.73	0.27				-0.43	0.44			
41	MC	1	53572	0.12	0.24	0.00		0.47	0.25	0.16	0.12		-0.09	-0.07	0.00	0.24
42	MC	1	53589	0.82	0.44	0.00		0.07	0.08	0.82	0.03		-0.26	-0.27	0.44	-0.16
43	MC	1	53534	0.31	0.40	0.00		0.21	0.28	0.20	0.31		-0.18	-0.07	-0.19	0.40
44	SA	1	53543	0.57	0.50	0.00	0.43	0.57				-0.50	0.50			
45	MC	1	53617	0.27	0.14	0.00		0.27	0.46	0.12	0.15		0.15	0.10	-0.14	-0.19
46	SA	1	53479	0.28	0.52	0.00	0.71	0.28				-0.51	0.52			

Table F-10. Item Statistics, Mathematics Grade 6

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	55316	0.38	0.34	0.00		0.38	0.28	0.28	0.06		0.34	-0.22	-0.17	0.06
2	MC	1	55307	0.35	0.49	0.00		0.15	0.41	0.09	0.35		-0.26	-0.22	-0.11	0.49
3	TE	1	55009	0.48	0.54	0.01	0.51	0.48				-0.53	0.54			
4	MC	1	55302	0.64	0.49	0.00		0.64	0.17	0.10	0.09		0.49	-0.25	-0.23	-0.26
5	SA	1	55148	0.43	0.45	0.00	0.57	0.43				-0.45	0.45			
6	SA	1	55089	0.41	0.53	0.00	0.59	0.41				-0.52	0.53			
7	MC	1	55306	0.24	0.39	0.00		0.22	0.24	0.07	0.47		-0.06	0.39	-0.07	-0.25
8	MC	1	55285	0.73	0.45	0.00		0.03	0.06	0.18	0.73		-0.16	-0.18	-0.33	0.45
9	SA	1	55263	0.13	0.48	0.00	0.87	0.13				-0.47	0.48			
10	MC	1	55279	0.44	0.44	0.00		0.19	0.21	0.15	0.44		-0.18	-0.19	-0.20	0.44
11	TE	1	54966	0.40	0.54	0.01	0.60	0.39				-0.52	0.54			
12	MC	1	55291	0.90	0.33	0.00		0.90	0.02	0.02	0.05		0.33	-0.15	-0.16	-0.24
13	SA	1	55162	0.42	0.60	0.00	0.57	0.42				-0.59	0.60			
14	SA	1	55040	0.18	0.51	0.01	0.81	0.18				-0.49	0.51			
15	MC	1	55241	0.36	0.59	0.00		0.36	0.17	0.41	0.05		0.59	-0.16	-0.38	-0.14
16	MC	1	55258	0.45	0.46	0.00		0.27	0.16	0.12	0.45		-0.26	-0.18	-0.14	0.46
17	MC	1	55162	0.34	0.48	0.00		0.19	0.34	0.31	0.15		-0.32	0.48	-0.06	-0.21
18	MC	1	55104	0.53	0.51	0.00		0.53	0.25	0.11	0.11		0.51	-0.35	-0.18	-0.14
19	MC	1	55119	0.70	0.37	0.00		0.07	0.70	0.12	0.11		-0.16	0.37	-0.19	-0.20
20	SA	1	55079	0.61	0.51	0.00	0.39	0.60				-0.50	0.51			
21	TE	1	55049	0.54	0.53	0.00	0.46	0.54				-0.52	0.53			
22	MC	1	55072	0.47	0.42	0.00		0.32	0.46	0.17	0.04		-0.25	0.42	-0.18	-0.10
23	TE	1	54454	0.67	0.32	0.01	0.33	0.66				-0.31	0.33			

Table F-10. Item Statistics, Mathematics Grade 6 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	54940	0.52	0.22	0.00		0.12	0.24	0.52	0.12		-0.07	-0.18	0.22	-0.02
25	MC	1	55005	0.28	0.31	0.00		0.51	0.28	0.12	0.09		-0.18	0.31	-0.20	0.06
26	MC	1	54976	0.41	0.35	0.00		0.29	0.17	0.12	0.41		-0.12	-0.21	-0.11	0.35
27	SA	1	54897	0.27	0.45	0.01	0.73	0.27				-0.44	0.45			
28	MC	1	55044	0.37	0.13	0.00		0.18	0.37	0.24	0.20		-0.20	0.13	-0.02	0.06
29	MC	1	55030	0.55	0.30	0.00		0.12	0.55	0.27	0.05		-0.09	0.30	-0.20	-0.13
30	MC	1	55062	0.57	0.33	0.00		0.18	0.19	0.57	0.05		-0.22	-0.13	0.33	-0.11
31	MC	1	55053	0.33	0.16	0.00		0.13	0.33	0.21	0.33		-0.06	0.16	-0.14	0.02
32	MC	1	55020	0.29	0.06	0.00		0.32	0.21	0.29	0.18		-0.02	0.02	0.06	-0.06
33	MC	1	55066	0.62	0.44	0.00		0.15	0.16	0.61	0.07		-0.24	-0.22	0.45	-0.17
34	MC	1	54980	0.59	0.47	0.00		0.15	0.09	0.18	0.58		-0.16	-0.23	-0.28	0.47
35	MC	1	54984	0.41	0.47	0.00		0.41	0.07	0.11	0.41		-0.17	-0.24	-0.26	0.47
36	TE	1	54686	0.34	0.50	0.01	0.65	0.34				-0.47	0.50			
37	SA	1	54743	0.17	0.55	0.01	0.82	0.17				-0.53	0.55			
38	MC	1	54983	0.40	0.20	0.00		0.19	0.35	0.39	0.06		-0.16	0.01	0.20	-0.15
39	MC	1	54984	0.33	0.38	0.00		0.34	0.20	0.33	0.13		-0.16	-0.06	0.38	-0.23
40	MC	1	54968	0.31	0.52	0.00		0.27	0.31	0.10	0.31		-0.24	0.52	-0.12	-0.20
41	TE	1	54751	0.33	0.55	0.01	0.66	0.33				-0.53	0.55			
42	MC	1	55014	0.38	0.24	0.00		0.13	0.17	0.31	0.38		-0.15	-0.15	-0.01	0.25
43	MC	1	55014	0.31	0.22	0.00		0.25	0.30	0.31	0.13		-0.23	-0.03	0.22	0.05
44	TE	1	54966	0.13	0.15	0.00	0.87	0.13				-0.13	0.15			
45	MC	1	54951	0.41	0.55	0.00		0.10	0.15	0.34	0.41		-0.24	-0.21	-0.24	0.55
46	MC	1	54898	0.85	0.33	0.01		0.06	0.85	0.04	0.06		-0.18	0.34	-0.15	-0.20

Table F-11. Item Statistics, Mathematics Grade 7

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	SA	1	56038	0.43	0.60	0.00	0.56	0.43				-0.60	0.60			
2	MC	1	56103	0.44	0.51	0.00		0.33	0.44	0.20	0.02		-0.23	0.51	-0.33	-0.13
3	MC	1	56105	0.46	0.32	0.00		0.46	0.13	0.26	0.15		0.32	-0.07	-0.20	-0.12
4	SA	1	55943	0.21	0.46	0.00	0.78	0.21				-0.45	0.46			
5	MC	1	56081	0.47	0.41	0.00		0.06	0.47	0.15	0.32		-0.13	0.41	-0.28	-0.15
6	MC	1	56084	0.37	0.42	0.00		0.37	0.08	0.31	0.24		0.42	-0.24	-0.18	-0.13
7	MC	1	56111	0.48	0.41	0.00		0.41	0.05	0.07	0.47		-0.27	-0.15	-0.17	0.41
8	TE	1	56084	0.15	0.41	0.00	0.84	0.15				-0.41	0.41			
9	MC	1	56081	0.33	0.24	0.00		0.30	0.12	0.25	0.33		0.08	-0.24	-0.16	0.24
10	MC	1	56092	0.49	0.39	0.00		0.49	0.20	0.25	0.06		0.39	-0.32	-0.04	-0.19
11	MC	1	56079	0.64	0.20	0.00		0.17	0.64	0.09	0.10		-0.08	0.20	-0.19	-0.04
12	MC	1	56039	0.73	0.23	0.00		0.16	0.05	0.06	0.73		-0.10	-0.14	-0.16	0.23
13	MC	1	55965	0.40	0.38	0.00		0.25	0.13	0.40	0.22		-0.18	-0.16	0.38	-0.13
14	MC	1	55995	0.46	0.40	0.00		0.46	0.18	0.20	0.16		0.40	-0.15	-0.07	-0.31
15	SA	1	55634	0.28	0.41	0.01	0.72	0.28				-0.39	0.41			
16	MC	1	55974	0.26	0.47	0.00		0.26	0.15	0.24	0.35		0.47	-0.12	-0.04	-0.31
17	MC	1	55914	0.22	0.13	0.00		0.26	0.24	0.27	0.22		0.01	-0.10	-0.03	0.13
18	MC	1	55976	0.39	0.19	0.00		0.04	0.18	0.39	0.39		-0.10	-0.05	0.19	-0.11
19	TE	1	55451	0.33	0.60	0.01	0.66	0.33				-0.57	0.60			
20	MC	1	55870	0.37	0.22	0.00		0.27	0.25	0.37	0.12		-0.12	-0.13	0.22	0.01
21	MC	1	55822	0.61	0.27	0.00		0.23	0.11	0.61	0.05		-0.03	-0.27	0.27	-0.13
22	MC	1	55835	0.25	0.19	0.00		0.26	0.33	0.16	0.25		-0.10	-0.05	-0.02	0.19
23	MC	1	55804	0.27	0.19	0.00		0.39	0.29	0.27	0.05		-0.06	-0.10	0.20	-0.05

Table F-11. Item Statistics, Mathematics Grade 7 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	SA	1	55685	0.36	0.51	0.01	0.64	0.35				-0.50	0.52			
25	SA	1	55545	0.17	0.49	0.01	0.82	0.17				-0.46	0.49			
26	MC	1	55880	0.40	0.29	0.00		0.17	0.27	0.16	0.40		-0.15	-0.15	-0.06	0.29
27	MC	1	55866	0.50	0.47	0.00		0.49	0.13	0.15	0.22		0.47	-0.29	-0.24	-0.10
28	MC	1	55900	0.78	0.45	0.00		0.05	0.08	0.10	0.77		-0.18	-0.25	-0.28	0.46
29	MC	1	55887	0.30	0.26	0.00		0.29	0.24	0.30	0.16		-0.09	-0.07	0.26	-0.12
30	MC	1	55873	0.40	0.44	0.00		0.14	0.11	0.36	0.39		-0.16	-0.18	-0.21	0.44
31	MC	1	55853	0.60	0.55	0.00		0.09	0.16	0.15	0.60		-0.13	-0.35	-0.28	0.56
32	TE	1	55428	0.66	0.20	0.01	0.34	0.65				-0.19	0.21			
33	TE	1	55790	0.10	0.50	0.00	0.90	0.10				-0.47	0.50			
34	MC	1	55792	0.55	0.30	0.00		0.13	0.55	0.18	0.14		-0.22	0.31	-0.21	0.03
35	SA	1	55441	0.35	0.65	0.01	0.64	0.35				-0.62	0.65			
36	TE	1	55762	0.67	0.48	0.01	0.33	0.67				-0.47	0.48			
37	MC	1	55808	0.49	0.50	0.00		0.49	0.21	0.21	0.09		0.50	-0.29	-0.14	-0.25
38	TE	1	55566	0.43	0.26	0.01	0.56	0.43				-0.25	0.27			
39	MC	1	55766	0.53	0.20	0.01		0.03	0.53	0.39	0.04		-0.15	0.20	-0.06	-0.19
40	MC	1	55795	0.28	0.20	0.00		0.15	0.28	0.29	0.28		0.00	0.20	0.05	-0.25
41	TE	1	55732	0.24	0.60	0.01	0.75	0.24				-0.58	0.60			
42	TE	1	55254	0.20	0.57	0.01	0.79	0.20				-0.52	0.57			
43	MC	1	55717	0.47	0.52	0.01		0.12	0.17	0.24	0.46		-0.18	-0.26	-0.23	0.52
44	MC	1	55677	0.48	0.46	0.01		0.47	0.20	0.24	0.07		0.46	-0.27	-0.17	-0.17
45	MC	1	55702	0.65	0.39	0.01		0.11	0.65	0.14	0.09		-0.18	0.39	-0.21	-0.17
46	SA	1	55547	0.61	0.57	0.01	0.39	0.61				-0.56	0.57			

Table F-12. Item Statistics, Mathematics Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	56578	0.59	0.47	0.00		0.17	0.58	0.08	0.17		-0.21	0.47	-0.20	-0.26
2	MC	1	56539	0.21	0.15	0.00		0.13	0.21	0.21	0.45		0.02	-0.12	0.15	-0.03
3	MC	1	56543	0.52	0.36	0.00		0.23	0.10	0.52	0.15		-0.13	-0.14	0.36	-0.23
4	MC	1	56543	0.44	0.20	0.00		0.44	0.17	0.25	0.14		0.20	-0.02	-0.14	-0.09
5	SA	1	56449	0.16	0.49	0.00	0.84	0.16				-0.48	0.49			
6	MC	1	56538	0.30	0.42	0.00		0.41	0.14	0.15	0.30		-0.13	-0.16	-0.20	0.42
7	MC	1	56562	0.42	0.60	0.00		0.35	0.17	0.42	0.07		-0.48	-0.17	0.60	-0.02
8	MC	1	56561	0.45	0.46	0.00		0.11	0.32	0.45	0.12		-0.24	-0.14	0.46	-0.27
9	TE	1	56519	0.22	0.55	0.00	0.78	0.22				-0.55	0.55			
10	MC	1	56559	0.36	0.26	0.00		0.36	0.30	0.15	0.19		0.26	-0.14	-0.11	-0.04
11	TE	1	56411	0.21	0.49	0.00	0.79	0.21				-0.48	0.49			
12	MC	1	56523	0.46	0.31	0.00		0.18	0.19	0.46	0.17		-0.10	-0.25	0.31	-0.05
13	SA	1	56441	0.39	0.63	0.00	0.61	0.39				-0.63	0.63			
14	MC	1	56411	0.46	0.24	0.00		0.27	0.46	0.16	0.11		-0.12	0.24	-0.04	-0.15
15	TE	1	56346	0.42	0.50	0.00	0.57	0.42				-0.49	0.50			
16	MC	1	56363	0.47	0.26	0.00		0.10	0.47	0.22	0.21		-0.12	0.27	-0.26	0.04
17	SA	1	56084	0.32	0.52	0.01	0.68	0.32				-0.51	0.53			
18	MC	1	56318	0.53	0.40	0.00		0.12	0.18	0.53	0.16		-0.16	-0.26	0.41	-0.13
19	MC	1	56347	0.44	0.35	0.00		0.11	0.44	0.38	0.06		-0.14	0.35	-0.22	-0.09
20	MC	1	56350	0.48	0.41	0.00		0.47	0.18	0.21	0.14		0.41	-0.16	-0.17	-0.20
21	SA	1	55931	0.22	0.57	0.01	0.77	0.22				-0.54	0.57			
22	MC	1	56272	0.71	0.38	0.00		0.12	0.71	0.13	0.05		-0.21	0.38	-0.21	-0.15
23	TE	1	55614	0.55	0.50	0.02	0.44	0.54				-0.48	0.51			

Table F-12. Item Statistics, Mathematics Grade 8 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	56248	0.52	0.37	0.00		0.52	0.24	0.16	0.07		0.37	-0.15	-0.22	-0.14
25	TE	1	56187	0.18	0.47	0.00	0.81	0.18				-0.45	0.47			
26	MC	1	56313	0.25	0.21	0.00		0.11	0.37	0.25	0.28		-0.20	-0.02	0.21	-0.03
27	SA	1	55977	0.17	0.46	0.01	0.82	0.17				-0.43	0.46			
28	MC	1	56293	0.55	0.53	0.00		0.07	0.14	0.24	0.55		-0.21	-0.22	-0.31	0.53
29	MC	1	56268	0.39	0.42	0.00		0.13	0.27	0.21	0.39		-0.16	-0.16	-0.19	0.42
30	MC	1	56267	0.28	0.34	0.00		0.28	0.37	0.26	0.09		0.34	-0.06	-0.23	-0.07
31	MC	1	56242	0.31	0.42	0.00		0.14	0.20	0.35	0.31		-0.11	-0.14	-0.20	0.43
32	TE	1	56078	0.31	0.48	0.01	0.68	0.31				-0.47	0.49			
33	MC	1	56195	0.53	0.38	0.00		0.16	0.53	0.18	0.14		-0.12	0.38	-0.21	-0.18
34	MC	1	56172	0.37	0.22	0.01		0.23	0.09	0.37	0.31		-0.12	-0.22	0.23	0.02
35	TE	1	56100	0.22	0.66	0.01	0.78	0.21				-0.64	0.66			
36	MC	1	56214	0.49	0.47	0.00		0.15	0.49	0.26	0.10		-0.12	0.47	-0.32	-0.16
37	TE	1	55816	0.32	0.29	0.01	0.67	0.32				-0.27	0.30			
38	MC	1	56224	0.74	0.40	0.00		0.13	0.05	0.74	0.07		-0.22	-0.17	0.40	-0.23
39	MC	1	56184	0.41	0.38	0.01		0.41	0.19	0.22	0.18		0.38	-0.26	-0.13	-0.06
40	MC	1	56180	0.20	0.27	0.01		0.22	0.36	0.21	0.20		-0.08	-0.04	-0.13	0.28
41	TE	1	56108	0.35	0.37	0.01	0.64	0.35				-0.36	0.38			
42	MC	1	56203	0.58	0.48	0.00		0.57	0.18	0.16	0.07		0.48	-0.16	-0.31	-0.20
43	MC	1	56138	0.66	0.43	0.01		0.07	0.66	0.19	0.08		-0.22	0.43	-0.22	-0.21
44	MC	1	56147	0.53	0.32	0.01		0.08	0.19	0.53	0.19		-0.13	-0.10	0.32	-0.21
45	MC	1	56163	0.48	0.37	0.01		0.09	0.48	0.32	0.11		-0.10	0.37	-0.23	-0.14
46	MC	1	56147	0.70	0.40	0.01		0.07	0.11	0.69	0.12		-0.20	-0.21	0.40	-0.20

Table F-13. Item Statistics, Science Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	52248	0.52	0.33	0.00		0.29	0.11	0.51	0.08		-0.09	-0.27	0.33	-0.14
2	TE	1	52248	0.75	0.51	0.00	0.25	0.75				-0.51	0.51			
3	TE	1	52252	0.33	0.32	0.00	0.67	0.33				-0.32	0.32			
4	TE	1	52151	0.43	0.42	0.00	0.57	0.43				-0.41	0.42			
5	TE	1	52004	0.52	0.38	0.01	0.48	0.52				-0.37	0.39			
6	TE	1	52187	0.60	0.51	0.00	0.40	0.60				-0.50	0.51			
7	MC	1	52214	0.36	0.29	0.00		0.36	0.28	0.31	0.05		0.29	-0.07	-0.13	-0.20
8	TE	1	52175	0.58	0.42	0.00	0.42	0.58				-0.41	0.42			
9	EBSR	1	52242	0.41	0.44	0.00	0.59	0.41				-0.43	0.44			
10	MC	1	52243	0.66	0.38	0.00		0.18	0.65	0.07	0.09		-0.20	0.38	-0.24	-0.14
11	MC	1	52232	0.50	0.40	0.00		0.18	0.14	0.50	0.17		-0.23	-0.24	0.40	-0.07
12	MC	1	52204	0.53	0.34	0.00		0.17	0.16	0.53	0.14		-0.10	-0.22	0.34	-0.14
13	TE	1	52091	0.69	0.40	0.00	0.31	0.69				-0.39	0.40			
14	TE	1	51944	0.63	0.45	0.01	0.36	0.63				-0.44	0.46			
15	TE	1	52122	0.39	0.58	0.00	0.61	0.39				-0.57	0.58			
16	MC	1	52202	0.61	0.34	0.00		0.11	0.17	0.11	0.61		-0.10	-0.25	-0.13	0.34
17	TE	1	52194	0.84	0.40	0.00	0.16	0.84				-0.40	0.40			
18	TE	1	52129	0.68	0.40	0.00	0.32	0.68				-0.40	0.40			
19	TE	1	52025	0.50	0.22	0.00	0.50	0.50				-0.21	0.22			
20	TE	1	52181	0.88	0.32	0.00	0.12	0.87				-0.32	0.32			

Table F-13. Item Statistics, Science Grade 4 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	TE	1	51782	0.52	0.40	0.01	0.48	0.51				-0.39	0.40			
22	TE	1	52100	0.53	0.57	0.00	0.47	0.52				-0.56	0.57			
23	MC	1	52183	0.57	0.41	0.00		0.13	0.57	0.24	0.05		-0.20	0.41	-0.22	-0.19
24	TE	1	52170	0.38	0.43	0.00	0.62	0.38				-0.42	0.43			
25	TE	1	51980	0.54	0.51	0.01	0.46	0.54				-0.50	0.52			
26	TE	1	52154	0.76	0.42	0.00	0.24	0.76				-0.42	0.43			
27	TE	1	52113	0.63	0.20	0.00	0.37	0.63				-0.19	0.20			
28	TE	1	52130	0.34	0.29	0.00	0.65	0.34				-0.29	0.29			
29	TE	1	52087	0.45	0.46	0.00	0.55	0.45				-0.46	0.46			
30	TE	1	52059	0.64	0.46	0.00	0.36	0.64				-0.46	0.46			
31	TE	1	51981	0.63	0.31	0.00	0.37	0.62				-0.30	0.31			
32	MC	1	52084	0.57	0.35	0.00		0.57	0.24	0.11	0.09		0.35	-0.13	-0.25	-0.14
33	TE	1	52123	0.74	0.41	0.00	0.26	0.74				-0.41	0.41			
34	TE	1	52113	0.90	0.36	0.00	0.10	0.90				-0.35	0.36			
35	MC	1	52099	0.35	0.23	0.00		0.32	0.18	0.35	0.15		-0.05	-0.25	0.23	0.03
36	EBSR	1	52124	0.43	0.26	0.00	0.57	0.43				-0.26	0.26			
37	MC	1	52037	0.51	0.40	0.00		0.51	0.16	0.21	0.12		0.40	-0.25	-0.17	-0.12
38	TE	1	52080	0.78	0.47	0.00	0.22	0.78				-0.47	0.47			
39	MS	1	52069	0.21	0.38	0.00	0.79	0.21				-0.37	0.38			
40	TE	1	52041	0.20	0.35	0.00	0.80	0.20				-0.34	0.35			

Table F-14. Item Statistics, Science Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TE	1	56208	0.63	0.43	0.00	0.37	0.63				-0.43	0.43			
2	TE	1	56229	0.53	0.47	0.00	0.47	0.53				-0.47	0.47			
3	TE	1	56198	0.52	0.58	0.00	0.48	0.52				-0.57	0.58			
4	TE	1	56214	0.58	0.51	0.00	0.41	0.58				-0.51	0.51			
5	TE	1	56213	0.83	0.33	0.00	0.17	0.83				-0.32	0.33			
6	MC	1	56198	0.56	0.33	0.00		0.08	0.25	0.10	0.56		-0.21	-0.09	-0.22	0.33
7	TE	1	56228	0.52	0.38	0.00	0.48	0.52				-0.37	0.38			
8	MC	1	56188	0.61	0.39	0.00		0.13	0.61	0.21	0.06		-0.07	0.39	-0.31	-0.18
9	TE	1	56251	0.79	0.30	0.00	0.21	0.79				-0.30	0.31			
10	TE	1	56250	0.47	0.32	0.00	0.53	0.47				-0.32	0.32			
11	TE	1	56230	0.60	0.41	0.00	0.40	0.60				-0.41	0.41			
12	MC	1	56225	0.52	0.30	0.00		0.09	0.12	0.52	0.26		-0.12	-0.27	0.30	-0.06
13	TE	1	56170	0.32	0.39	0.00	0.68	0.32				-0.38	0.39			
14	TE	1	56184	0.72	0.36	0.00	0.28	0.71				-0.35	0.36			
15	TE	1	56223	0.53	0.29	0.00	0.47	0.53				-0.28	0.29			
16	TE	1	56132	0.38	0.45	0.00	0.62	0.38				-0.45	0.45			
17	TE	1	56152	0.42	0.34	0.00	0.58	0.42				-0.34	0.34			
18	TE	1	56135	0.51	0.55	0.00	0.49	0.51				-0.54	0.55			
19	TE	1	56101	0.29	0.37	0.00	0.71	0.29				-0.36	0.37			
20	TE	1	56065	0.51	0.33	0.00	0.48	0.51				-0.33	0.33			

Table F-14. Item Statistics, Science Grade 8 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	TE	1	56012	0.83	0.41	0.00	0.17	0.83				-0.41	0.41			
22	TE	1	56113	0.71	0.42	0.00	0.29	0.71				-0.42	0.42			
23	TE	1	56088	0.57	0.45	0.00	0.43	0.57				-0.45	0.45			
24	TE	1	56107	0.48	0.08	0.00	0.52	0.48				-0.07	0.08			
25	TE	1	56112	0.51	0.44	0.00	0.48	0.51				-0.44	0.44			
26	TE	1	56066	0.34	0.52	0.00	0.66	0.34				-0.52	0.52			
27	TE	1	56007	0.48	0.54	0.00	0.51	0.48				-0.54	0.54			
28	TE	1	55954	0.58	0.18	0.00	0.42	0.58				-0.18	0.18			
29	TE	1	55991	0.64	0.42	0.00	0.36	0.64				-0.41	0.42			
30	MC	1	55959	0.51	0.43	0.00		0.14	0.19	0.51	0.15		-0.26	-0.27	0.43	-0.04
31	EBSR	1	56005	0.46	0.47	0.00	0.54	0.46				-0.47	0.47			
32	TE	1	55972	0.74	0.38	0.00	0.26	0.74				-0.37	0.38			
33	TE	1	55799	0.42	0.36	0.01	0.58	0.42				-0.35	0.36			
34	TE	1	55927	0.45	0.35	0.00	0.55	0.45				-0.35	0.36			
35	TE	1	55956	0.28	0.36	0.00	0.72	0.28				-0.35	0.36			
36	MS	1	55914	0.30	0.45	0.00	0.69	0.30				-0.44	0.45			
37	TE	1	55962	0.42	0.50	0.00	0.58	0.42				-0.49	0.50			
38	MC	1	55930	0.69	0.44	0.00		0.09	0.69	0.14	0.08		-0.20	0.44	-0.29	-0.18
39	TE	1	55889	0.55	0.43	0.00	0.45	0.55				-0.43	0.43			
40	MC	1	55914	0.45	0.28	0.00		0.15	0.25	0.45	0.14		-0.03	-0.16	0.28	-0.17

Table F-15. Item Statistics, Social Studies Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	52306	0.77	0.52	0.00		0.10	0.05	0.07	0.77		-0.27	-0.27	-0.28	0.52
2	MC	1	52278	0.81	0.44	0.00		0.11	0.81	0.03	0.05		-0.27	0.44	-0.23	-0.21
3	MC	1	52273	0.79	0.36	0.00		0.79	0.05	0.11	0.06		0.36	-0.26	-0.16	-0.18
4	MC	1	52273	0.76	0.47	0.00		0.07	0.11	0.76	0.06		-0.27	-0.26	0.47	-0.20
5	MC	1	52224	0.72	0.43	0.00		0.05	0.08	0.14	0.72		-0.23	-0.24	-0.22	0.43
6	MC	1	52219	0.43	0.28	0.00		0.42	0.19	0.23	0.16		0.28	-0.11	-0.13	-0.11
7	TE	1	52241	0.69	0.37	0.00	0.30	0.69				-0.37	0.38			
8	MC	1	52235	0.51	0.34	0.00		0.22	0.22	0.05	0.51		-0.19	-0.10	-0.21	0.34
9	MC	1	52217	0.76	0.39	0.00		0.07	0.76	0.10	0.07		-0.17	0.39	-0.23	-0.20
10	MC	1	52243	0.65	0.46	0.00		0.11	0.11	0.14	0.65		-0.19	-0.27	-0.22	0.46
11	MC	1	52254	0.51	0.19	0.00		0.17	0.51	0.14	0.18		-0.05	0.19	-0.10	-0.10
12	MC	1	52117	0.78	0.45	0.00		0.08	0.77	0.08	0.06		-0.25	0.46	-0.30	-0.16
13	TE	1	52149	0.42	0.34	0.00	0.58	0.41				-0.33	0.34			
14	MC	1	52237	0.37	0.39	0.00		0.22	0.28	0.13	0.37		-0.18	-0.06	-0.25	0.39
15	MC	1	52248	0.65	0.37	0.00		0.18	0.10	0.08	0.65		-0.24	-0.18	-0.12	0.37
16	MC	1	52230	0.58	0.33	0.00		0.14	0.15	0.58	0.14		-0.06	-0.30	0.33	-0.10
17	MC	1	52249	0.65	0.47	0.00		0.07	0.08	0.21	0.65		-0.21	-0.22	-0.28	0.47
18	TE	1	52175	0.70	0.36	0.00	0.30	0.70				-0.35	0.36			
19	MC	1	52244	0.72	0.47	0.00		0.09	0.72	0.10	0.09		-0.30	0.48	-0.29	-0.14
20	MC	1	52207	0.63	0.33	0.00		0.63	0.22	0.08	0.06		0.34	-0.13	-0.23	-0.17

Table F-15. Item Statistics, Social Studies Grade 4 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	52193	0.70	0.45	0.00		0.08	0.05	0.70	0.17		-0.24	-0.27	0.45	-0.22
22	MC	1	52189	0.64	0.21	0.00		0.64	0.21	0.07	0.08		0.21	-0.05	-0.17	-0.12
23	MC	1	52180	0.81	0.55	0.00		0.05	0.06	0.08	0.80		-0.25	-0.29	-0.34	0.55
24	MC	1	52164	0.52	0.41	0.00		0.52	0.16	0.18	0.14		0.42	-0.26	-0.20	-0.10
25	MC	1	52164	0.63	0.52	0.00		0.08	0.19	0.63	0.10		-0.23	-0.28	0.52	-0.26
26	MC	1	52190	0.56	0.29	0.00		0.09	0.30	0.56	0.05		-0.29	-0.06	0.29	-0.16
27	MC	1	52044	0.61	0.35	0.00		0.19	0.10	0.61	0.10		-0.10	-0.21	0.35	-0.21
28	MC	1	52172	0.60	0.44	0.00		0.11	0.60	0.12	0.17		-0.19	0.44	-0.26	-0.19
29	MC	1	52169	0.50	0.37	0.00		0.10	0.30	0.10	0.50		-0.21	-0.13	-0.21	0.37
30	MC	1	52173	0.73	0.46	0.00		0.09	0.09	0.73	0.08		-0.25	-0.26	0.46	-0.20
31	TE	1	52172	0.49	0.43	0.00	0.50	0.49				-0.42	0.43			
32	MC	1	52185	0.64	0.50	0.00		0.64	0.13	0.13	0.10		0.50	-0.30	-0.23	-0.20
33	MC	1	52034	0.72	0.51	0.00		0.15	0.06	0.07	0.72		-0.20	-0.30	-0.31	0.51
34	MC	1	52153	0.69	0.51	0.00		0.69	0.18	0.09	0.04		0.52	-0.28	-0.31	-0.21
35	MC	1	52175	0.67	0.44	0.00		0.15	0.06	0.67	0.12		-0.27	-0.28	0.44	-0.13
36	MC	1	52173	0.74	0.54	0.00		0.07	0.74	0.12	0.06		-0.28	0.54	-0.32	-0.24
37	MC	1	52157	0.63	0.48	0.00		0.19	0.12	0.63	0.06		-0.21	-0.27	0.48	-0.24
38	MC	1	52182	0.69	0.47	0.00		0.69	0.05	0.08	0.18		0.47	-0.26	-0.26	-0.24

Table F-16. Item Statistics, Social Studies Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	56258	0.82	0.47	0.00		0.09	0.04	0.05	0.82		-0.33	-0.22	-0.19	0.47
2	MC	1	56218	0.80	0.48	0.00		0.80	0.11	0.04	0.06		0.48	-0.35	-0.23	-0.17
3	MC	1	56207	0.85	0.53	0.00		0.85	0.04	0.06	0.05		0.53	-0.27	-0.32	-0.27
4	MC	1	56223	0.80	0.38	0.00		0.07	0.06	0.80	0.07		-0.20	-0.21	0.38	-0.20
5	MC	1	56209	0.82	0.52	0.00		0.04	0.82	0.07	0.07		-0.24	0.52	-0.32	-0.28
6	MC	1	56190	0.75	0.35	0.00		0.03	0.03	0.18	0.75		-0.18	-0.25	-0.20	0.35
7	MS	1	56224	0.58	0.60	0.00	0.41	0.58				-0.60	0.60			
8	MC	1	56136	0.66	0.54	0.00		0.66	0.10	0.13	0.11		0.54	-0.29	-0.32	-0.19
9	MC	1	56206	0.53	0.39	0.00		0.14	0.28	0.06	0.53		-0.23	-0.11	-0.27	0.39
10	TE	1	56087	0.60	0.53	0.00	0.40	0.60				-0.53	0.54			
11	MC	1	56195	0.80	0.48	0.00		0.79	0.03	0.12	0.05		0.49	-0.21	-0.32	-0.25
12	MC	1	56179	0.73	0.58	0.00		0.08	0.09	0.11	0.73		-0.21	-0.28	-0.39	0.58
13	MC	1	56203	0.76	0.34	0.00		0.09	0.07	0.76	0.07		-0.10	-0.19	0.34	-0.25
14	MC	1	56153	0.50	0.35	0.00		0.15	0.25	0.10	0.50		-0.13	-0.10	-0.29	0.36
15	MC	1	56112	0.64	0.38	0.00		0.11	0.64	0.12	0.12		-0.10	0.38	-0.31	-0.14
16	MC	1	56064	0.62	0.37	0.00		0.05	0.62	0.22	0.11		-0.19	0.37	-0.14	-0.25
17	MC	1	56146	0.69	0.41	0.00		0.13	0.11	0.69	0.06		-0.15	-0.26	0.41	-0.23
18	MC	1	56133	0.62	0.44	0.00		0.14	0.15	0.62	0.09		-0.15	-0.28	0.45	-0.22
19	MC	1	56154	0.66	0.35	0.00		0.07	0.12	0.66	0.16		-0.20	-0.22	0.36	-0.12
20	MC	1	56099	0.66	0.41	0.00		0.10	0.12	0.11	0.66		-0.17	-0.21	-0.22	0.41

Table F-16. Item Statistics, Social Studies Grade 8 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	56068	0.92	0.37	0.00		0.04	0.02	0.01	0.92		-0.20	-0.23	-0.20	0.37
22	MC	1	56040	0.58	0.49	0.00		0.31	0.06	0.05	0.58		-0.27	-0.24	-0.26	0.49
23	TE	1	55978	0.57	0.40	0.00	0.43	0.57				-0.40	0.41			
24	MC	1	55991	0.74	0.39	0.00		0.11	0.74	0.08	0.07		-0.18	0.39	-0.28	-0.14
25	MC	1	55909	0.74	0.36	0.00		0.74	0.05	0.03	0.19		0.36	-0.26	-0.24	-0.16
26	MC	1	56004	0.69	0.45	0.00		0.17	0.69	0.07	0.08		-0.21	0.46	-0.33	-0.18
27	MC	1	55984	0.54	0.37	0.00		0.54	0.09	0.19	0.17		0.37	-0.27	-0.20	-0.07
28	MC	1	55990	0.77	0.38	0.00		0.08	0.07	0.77	0.07		-0.14	-0.18	0.38	-0.27
29	MC	1	55983	0.59	0.25	0.00		0.09	0.59	0.09	0.24		-0.16	0.25	-0.26	-0.01
30	TE	1	55940	0.51	0.52	0.00	0.49	0.51				-0.52	0.53			
31	MC	1	55974	0.68	0.44	0.00		0.06	0.68	0.14	0.12		-0.20	0.45	-0.28	-0.19
32	TE	1	55943	0.56	0.44	0.00	0.44	0.56				-0.43	0.44			
33	MC	1	55942	0.50	0.34	0.00		0.50	0.08	0.20	0.22		0.34	-0.30	-0.15	-0.06
34	MC	1	55918	0.50	0.32	0.00		0.22	0.50	0.09	0.19		-0.10	0.33	-0.33	-0.06
35	MC	1	55928	0.60	0.45	0.00		0.60	0.12	0.16	0.12		0.45	-0.26	-0.25	-0.14
36	MC	1	55936	0.54	0.38	0.00		0.10	0.54	0.17	0.18		-0.11	0.38	-0.28	-0.13
37	MC	1	55944	0.47	0.30	0.00		0.30	0.47	0.13	0.10		0.00	0.30	-0.28	-0.17
38	MC	1	55946	0.42	0.35	0.00		0.42	0.21	0.20	0.17		0.35	-0.11	-0.22	-0.10
39	MC	1	55949	0.82	0.48	0.00		0.03	0.82	0.08	0.06		-0.22	0.48	-0.28	-0.28
40	MC	1	55906	0.61	0.51	0.00		0.07	0.18	0.15	0.60		-0.15	-0.25	-0.31	0.51

Table F-17. Item Statistics, Social Studies Grade 10

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	51091	0.81	0.42	0.00		0.81	0.09	0.04	0.06		0.42	-0.24	-0.26	-0.18
2	MC	1	51073	0.74	0.25	0.00		0.05	0.02	0.74	0.19		-0.14	-0.18	0.25	-0.13
3	MC	1	51065	0.72	0.39	0.00		0.72	0.07	0.11	0.10		0.39	-0.26	-0.20	-0.16
4	MC	1	51054	0.69	0.40	0.00		0.69	0.12	0.14	0.05		0.40	-0.29	-0.14	-0.18
5	MC	1	51058	0.75	0.45	0.00		0.03	0.75	0.06	0.16		-0.20	0.45	-0.26	-0.26
6	MC	1	50941	0.75	0.43	0.00		0.07	0.07	0.11	0.75		-0.18	-0.19	-0.29	0.43
7	MC	1	51065	0.82	0.46	0.00		0.05	0.05	0.82	0.08		-0.22	-0.27	0.46	-0.26
8	MC	1	51027	0.76	0.41	0.00		0.02	0.16	0.76	0.06		-0.17	-0.25	0.41	-0.23
9	MC	1	51055	0.58	0.42	0.00		0.58	0.23	0.12	0.07		0.42	-0.23	-0.23	-0.13
10	MC	1	51020	0.69	0.38	0.00		0.13	0.15	0.69	0.03		-0.13	-0.28	0.39	-0.19
11	MC	1	51017	0.59	0.41	0.00		0.11	0.59	0.25	0.04		-0.19	0.41	-0.20	-0.25
12	TE	1	50882	0.33	0.34	0.00	0.66	0.33				-0.33	0.35			
13	MC	1	51010	0.56	0.29	0.00		0.16	0.56	0.10	0.17		-0.04	0.30	-0.26	-0.13
14	MC	1	51010	0.59	0.42	0.00		0.59	0.12	0.19	0.10		0.42	-0.17	-0.26	-0.15
15	MC	1	51016	0.68	0.43	0.00		0.14	0.68	0.09	0.09		-0.14	0.43	-0.29	-0.22
16	MC	1	50933	0.65	0.33	0.00		0.10	0.07	0.64	0.19		-0.12	-0.24	0.34	-0.16
17	MC	1	50955	0.58	0.25	0.00		0.04	0.28	0.58	0.10		-0.18	-0.05	0.25	-0.21
18	MC	1	50897	0.73	0.34	0.00		0.05	0.11	0.73	0.11		-0.10	-0.21	0.34	-0.19
19	MC	1	50992	0.67	0.47	0.00		0.18	0.67	0.07	0.08		-0.27	0.48	-0.28	-0.16
20	MC	1	50978	0.63	0.44	0.00		0.63	0.14	0.15	0.07		0.44	-0.14	-0.29	-0.22

Table F-17. Item Statistics, Social Studies Grade 10 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	50961	0.60	0.33	0.00		0.05	0.16	0.60	0.18		-0.25	-0.11	0.33	-0.16
22	MC	1	50954	0.55	0.46	0.00		0.10	0.25	0.10	0.55		-0.22	-0.17	-0.29	0.46
23	MC	1	50959	0.58	0.51	0.00		0.58	0.17	0.14	0.11		0.51	-0.18	-0.30	-0.24
24	TE	1	50671	0.31	0.45	0.01	0.68	0.31				-0.42	0.45			
25	MC	1	50874	0.62	0.58	0.00		0.11	0.13	0.14	0.61		-0.22	-0.32	-0.30	0.58
26	MC	1	50751	0.74	0.35	0.00		0.07	0.74	0.15	0.04		-0.21	0.36	-0.22	-0.13
27	MC	1	50733	0.73	0.31	0.00		0.05	0.06	0.73	0.16		-0.19	-0.22	0.31	-0.11
28	MC	1	50684	0.44	0.47	0.00		0.22	0.15	0.19	0.44		-0.11	-0.25	-0.24	0.47
29	MC	1	50673	0.42	0.38	0.00		0.16	0.22	0.20	0.42		-0.18	-0.05	-0.25	0.38
30	MC	1	50697	0.57	0.53	0.00		0.57	0.12	0.17	0.13		0.53	-0.28	-0.22	-0.26
31	MC	1	50655	0.74	0.44	0.00		0.73	0.07	0.15	0.04		0.44	-0.23	-0.27	-0.19
32	MC	1	50706	0.84	0.39	0.00		0.02	0.06	0.84	0.08		-0.16	-0.18	0.40	-0.29
33	MC	1	50628	0.62	0.50	0.00		0.61	0.18	0.12	0.09		0.50	-0.28	-0.27	-0.16
34	TE	1	50569	0.70	0.20	0.00	0.30	0.69				-0.19	0.20			
35	MC	1	50615	0.63	0.48	0.00		0.12	0.62	0.12	0.13		-0.19	0.48	-0.33	-0.17
36	MC	1	50606	0.73	0.49	0.00		0.08	0.13	0.73	0.06		-0.22	-0.27	0.49	-0.27
37	MC	1	50626	0.47	0.37	0.00		0.47	0.23	0.10	0.19		0.37	-0.15	-0.30	-0.07
38	MC	1	50616	0.56	0.35	0.00		0.13	0.18	0.56	0.13		-0.09	-0.15	0.35	-0.24
39	MC	1	50639	0.72	0.42	0.00		0.72	0.06	0.09	0.12		0.42	-0.30	-0.22	-0.15
40	MC	1	50625	0.71	0.46	0.00		0.11	0.71	0.09	0.09		-0.12	0.46	-0.34	-0.26

Table F-17. Item Statistics, Social Studies Grade 10 (cont.)

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
41	MC	1	50570	0.82	0.43	0.00		0.09	0.05	0.81	0.05		-0.18	-0.26	0.43	-0.27
42	MC	1	50578	0.64	0.38	0.00		0.64	0.07	0.25	0.04		0.38	-0.31	-0.13	-0.23
43	MC	1	50591	0.70	0.43	0.00		0.03	0.08	0.19	0.70		-0.20	-0.16	-0.30	0.43
44	MC	1	50585	0.43	0.34	0.00		0.16	0.43	0.26	0.14		-0.19	0.34	-0.10	-0.14
45	TE	1	50507	0.31	0.39	0.01	0.68	0.31				-0.37	0.39			
46	MC	1	50559	0.54	0.41	0.00		0.16	0.54	0.21	0.09		-0.12	0.41	-0.22	-0.24
47	MC	1	50576	0.46	0.25	0.00		0.07	0.46	0.37	0.09		-0.15	0.25	-0.03	-0.24
48	MC	1	50568	0.69	0.51	0.00		0.08	0.69	0.11	0.12		-0.22	0.51	-0.30	-0.24
49	MC	1	50551	0.70	0.41	0.00		0.10	0.13	0.07	0.70		-0.29	-0.14	-0.20	0.41
50	MC	1	50524	0.86	0.31	0.01		0.08	0.04	0.86	0.01		-0.17	-0.22	0.32	-0.13

Appendix G

English Language Arts, Mathematics, Science, and Social Studies

Post-equating Verification

This Appendix consists of two sections: a description of the post-equating verification process and results and a comparison of the impact data obtained using pre-equated and post-equated item parameters.

1. Post-equating Verification

Following student scoring using pre-equated item parameters (as described in Part 6 of the Technical Report), post-equating verification was performed. The purpose of the post-equating verification procedure was twofold: (1) examine stability of the operational item parameter estimates after the 2021 Spring administration and (2) evaluate student scores estimated using pre-equated versus post-equated item parameters. Calibration samples were used to conduct post-equating.

1.1 Item Calibration

This section outlines the calibration procedures and results for the Spring 2021 Wisconsin Forward Exam. As described in Part 6, Section 6.4 " Item Response Theory Methodology" of this Technical Report, the three-parameter logistic (3PL) model and the two-parameter partial credit (2PPC) IRT model 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, and short-answer (SA) items, were treated as CR items in calibrations. The calibration sample description and results of the calibration process, including model-to-data fit statistics are presented in this section. Section 1.2 describes test equating procedures and results for English Language Arts (ELA), Mathematics, Science, and Social Studies.

Calibration Sample

The Spring calibration of the Wisconsin Forward Exam was based on samples of student data acquired during the entire testing window of Spring 2021. Student participation in the Spring 2021 administration ranged from approximately 75% in grade 10 to about 87% in grade 3. The Spring 2021 tested population was found to be different from the Spring 2019 tested population in regard to several demographic variables. Specifically, Black students, socioeconomically disadvantaged students, and students from "City" districts were underrepresented in the Spring 2021 test participants, while White and not socioeconomically disadvantaged students were overrepresented. To align the student demographics characteristics between the Spring 2021 and Spring 2019 test administrations, sampling of the Spring 2021 data was performed using the propensity score matching (PSM) method (Rosenbaum & Rubin, 1983). The covariates in the PSM model included student gender, ethnicity, socioeconomic status, English language proficiency (ELP) status, disability status, and district locale as indicated in the National Center for Education Statistics database (https://nces.ed.gov/ccd/districtsearch).

The comparability of the Spring 2021 calibration sample to the Spring 2019 census data in terms of demographic characteristics was examined 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 characteristics of the Spring 2019 census data, Spring 2021 tested population data, and Spring 2021 calibration samples are presented in Tables G-1 to G-7. The 2021 calibration samples consisted of approximately 60% of tested students who had complete operational test data and had valid pre-equated test scores. These samples were comparable to the Spring 2019 Wisconsin student population for the same grade. As shown in Tables G-1 through G-6, the differences between the demographic characteristics of the Spring 2021 calibration samples and the Spring 2019 census were less than 1% for any demographic category for grades 3 through 8. The differences between the demographic characteristics of the Spring 2021 calibration sample and the Spring 2019 census were no larger than approximately 2% for any demographic category in grade 10. The calibration samples were considered to be comparable to the Spring 2019 census in terms of the student demographic characteristics.

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 the 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 MC anchor items were fixed to their Spring 2019 (or most recent) values.

Calibration Software

Calibration of the Wisconsin Forward Exam data was performed using PARDUX software (Burket, 2002). 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).

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% 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, $DF = 10(m_j - 1) - m_j = 9m_j - 10$.

We 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 G-8 lists 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 in the calibration sample who took this item), Z, and critical Z, as described previously. Sixteen items were flagged for poor fit for ELA, eight items were flagged for Mathematics, four items were flagged for Science, and one item was flagged for Social Studies. Most of the flagged items were TE and EBSR items. For example, item #2 for ELA grade 3 was flagged because the observed Z of 171.24 was larger than the critical Z value of 81.24 based on a sample size of 30,465. For many of the flagged items, the observed Z and the critical Z were not very far apart, indicating small misfit; however, it was observed that for some items, the misfit was moderate to large (e.g., item #2 for ELA grade 3, item #34 for ELA grade 4, item #8 for ELA grade 5, item #38 in Mathematics grade 7).

To evaluate item-to-model fit further, we 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 predications and the observed student responses. The ICCs show 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.

1.2 Test Equating

Test equating is a 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. In a typical administration year, Wisconsin Forward Exam assessments are post-equated. However, due to a disrupted school year and potentially compromised student opportunity to learn, the pre-equated design with post-equating verification was implemented in Spring 2021. A common-item design was used for post-equating of the assessments from 2021 to the established ELA, Mathematics, Science, and Social Studies scales for the Wisconsin Forward Exam. All items included in the Spring 2021 assessments were used as anchor items in post-equating verification. 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. (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 $\widehat{\Psi}_j$ be the TCC based on estimates from a previous calibration and $\widehat{\Psi}_j^*$ 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 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.

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)

Three statistical methods were used to evaluate anchor items: (1) iterative linking (Candell & Drasgow, 1988) using Stocking & Lord's (1983) TCC method, (2) differences between the item-ability regression curves, and (3) delta-plot method.

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 2021 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.

Delta-Plot Method

As an added measure in Spring 2021 post-equating, the evaluation of item statistics differences using the delta-plot method was performed. The delta-plot method relies on the differences in the standardized proportion correct value (*p*-value). *P*-values of the items based on the previous (2019) and current year's (2021) population were converted to z-scores that corresponded to the (1-p)[th] percentiles. For example, for a *p*-value of 0.90, the corresponding *z*-score is the (1-0.90)[th] percentile, which is -1.2816. A simple rule to identify outlier items that are functioning differentially between the two groups with respect to the level of difficulty is to draw the perpendicular distance to the line of best fit. The fitted line is chosen so as to minimize the sum of squared perpendicular distances of the points to the line. The perpendicular distance is given by:

$$D = \frac{AZ_{old} - Z_{new} + B}{\sqrt{A^2 + 1}}$$

Where $A = \dfrac{(SD_{Znew}^2 - SD_{Zold}^2) + \sqrt{(SD_{Znew}^2 - SD_{Zold}^2)^2 + 4r_{(Zold)(Znew)}^2 SD_{Zold}^2 SD_{Znew}^2}}{2r_{(Znew)(Zold)} SD_{Zold} SD_{Znew}}$

and B = Mean (Z_{new}) – A * Mean (Z_{old}). The standard deviation (SD) of the perpendicular distance is given by:

$$SD_D = [(SD_{Znew} + SD_{Zold})/2] * \sqrt{1 - r_{(Zold)(Znew)}}$$

As a rule of thumb, any items lying more than two standard deviations of the distance away from the fitted line are flagged as outliers for further review.

Rules for 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 performance level) when recommending items for removal.

Items removed from the anchor set are still scored as part of the whole test. In a typical administration year, 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 subsection "Item Response Theory Item-Ability Regression Curves") 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.

These rules were followed in the post-equating verification in Spring 2021. Items flagged by any of the three methods were 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 last and the current administration. For the flagged non-MC anchor items, verification that no changes to scoring rubrics occurred between the two administrations is performed. It should be noted that anchor items are not considered for removal from equating solely due to a change in student performance on these items between administrations.

Evaluation of Equating Results

Table G-9 provides equating results for the TCC method for ELA, Mathematics, Science, and Social Studies. This table summarizes the following information for each content area and grade: 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).

The overall alignment of the anchor TCCs was very good for all grades and content areas. Figures G-1 through G-4 show the TCC alignment of the anchor set before and after equating for ELA, Mathematics, Science, and Social Studies, respectively. 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 0.93 or higher for all grades and content areas. The correlations between the b-parameter input and estimates were 0.98 or higher for all grades and content areas.

One anchor item was flagged as an a-parameter outlier in each of the following: ELA grades 6 and 7; Mathematics grades 3, 5, and 6; and Social Studies grade 4. Two anchor items

were flagged as *a*-parameter outliers in ELA grades 3, 4, 5, and 8; and Mathematics grades 4, 7 and 8; and in Social Studies grade 8. Four anchor items were flagged in Social Studies grade 10.

One anchor item was flagged as a *b*-parameter outlier in each of the following: ELA grades 3, 4, 5, and 8; Mathematics grades 4, 5, and 8; Science grade 4; and Social Studies grade 10. Two anchor items were flagged as *b*-parameter outliers in ELA grade 6 and Mathematics grades 3, 6, and 7. Three anchor items were flagged as *b*-parameter outliers in Social Studies grades 4 and 8. Overall, the number of anchor items flagged using the TCC method was small.

The list of items flagged using the delta-plot method is provided in Table G-10. These results include item *p*-values in the Spring 2021 test administration and the most recent previous administration, *D* statistics, and the critical *D* values. Two items were flagged per grade in ELA grades 3, 4, and 5, and one item was flagged per grade in ELA grades 6, 7, and 8. Three items were flagged in Mathematics grade 4, two items were flagged per grade in Mathematics grades 7 and 8, and one item was flagged per grade in Mathematics grades 5 and 6. Two items were flagged per grade in all the Science and Social Studies grades. The ELA *p*-value plots are shown in Figures G-5 through G-10. The Mathematics *p*-value plots are shown in Figures G-11 through G-16. The Science *p*-value plots are shown in Figures G-17 and G-18. The Social Studies *p*-value plots are shown in Figures G-19 through G-21.

No anchor items in any grade or content area were flagged using the IRT item-ability regression curve method.

The anchor items flagged as *a*- and *b*-parameter outliers using the TCC method or flagged using the delta-plot method were reviewed by DRC test development experts who verified that no changes to item content, format, or scoring rule occurred between the most recent and current administrations. In addition, none of the flagged items moved by more than 5 positions within a test session after removal of field test items from the ELA, Mathematics, and Science tests in Spring 2021. Flagged items were also reviewed for content-related patterns in an attempt to determine whether certain standards might not have been taught or not taught to the full extent in the 2020–21 academic year. No such patterns were identified for any of the grades or contents. Therefore, no anchor items were removed from the anchor sets used in equating.

2. Student Scoring

To evaluate the comparability of student scores estimated using pre-equated versus post-equated item parameters, the item pattern method and both sets of item parameters were used to score students who took ELA, Mathematics, Science, and Social Studies tests. The means and standard deviations of the scale scores were computed, and both the pre-equated and post-equated results are presented in Table G-11 for ELA, Mathematics, Science, and Social Studies. In addition, students were classified into performance levels using the post-equated scale scores. The percentages of students in each performance level based on pre-equated and post-equated results are presented in Table G-12.

As shown in Table G-11, the differences between the scale score means derived using the pre- and post-equated item parameters were within half of a scale score point for all grades in all

content areas, except for Mathematics grade 8. The mean scale score based on the pre-equated parameters was larger by 0.7 scale score points than the mean scale score based on post-equated item parameters for Mathematics grade 8. The scale score standard deviations of the scale score distributions based on the pre- and post-equated parameters were comparable within half of a scale score point for all grades and content areas, except for Mathematics grades 3 and 8. The differences between pre- and post-equated scale score standard deviations were within 1 scale score point for Mathematics grades 3 and 8.

Regarding the percentages of students in different performance levels, it was observed that the pre- and post-equating methods produced similar results (see Table G-12). The differences between the pre- and post-equating scoring approaches resulted in less than half of a percent difference in student classification for any performance category in all ELA grades, Mathematics grades 3 through 6, both Science grades, and all Social Studies grades. A difference of approximately 0.6% was observed at the *Proficient* level for Mathematics grade 7 (more students were classified as *Proficient* using post-equated scoring). In addition, differences of approximately 0.7% and 1% at the *Below Basic* and *Basic* levels, respectively, were observed for Mathematics grade 8 (more students were classified as *Below Basic* and fewer students were classified as *Basic* using post-equated scoring).

When the percentages of students in the *Proficient* or *Advanced* categories were considered, the differences between the pre- and post-equating scoring methods were less than half of a percent for any grade and content, except for Mathematics grade 7, where the difference was approximately 0.6% (slightly more students were classified as *Proficient* or *Advanced* using post-equated scoring).

It is expected that the pre- and post-equated scale scores may be different for the same student resulting from differences between pre- and post-equated item parameters. It is also possible that the difference in the scale scores based on the pre- and post-equated item parameters may lead to differences in performance level classification for some students. As the last step of pre- and post-equated results comparison, changes in the performance level classification were examined. Tables G-13 through G-16 show the numbers and percentages of students who were classified in either the same or a different performance level based on pre- and post-equated item parameters. More than 98% of students were classified in the same performance levels when scored using the pre- and post-equated item parameters across all grades and content areas. Less than 2% of students were classified in a different performance level when scored using the pre- and post-equated item parameters. Changes in the performance level based on the scoring approach were either one level up or one level down.

3. Summary

In summary, the pre-equated item parameters that were obtained in the Spring 2019 (or earlier) operational or field test administrations were used in student scoring after the Spring 2021 test administration. The post-equating verification was conducted to evaluate the parameter stability and the comparability of scale scores estimated using the pre- and post-equated parameters. The calibration results were satisfactory, with very few items flagged for poor fit across all grade levels and content areas. The equating results showed very high correlations

between the input and estimates of *a*- and *b*-parameters and very good alignment of anchor and estimate TCCs. The number of items flagged using the delta-plot method was small. No anchor items were flagged using the item-ability regression method. No anchor items were removed from the anchor sets. In addition, the comparison of mean scale scores, scale score standard deviations, and percentages of students classified in the four performance levels revealed only very small differences in student classification based on the use of the pre-equated and post-equated item parameters. In conclusion, the use of pre-equated parameters in Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies student scoring was appropriate.

Table G-1. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 3

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
3	All Students	All Students	61091	100.0	52930	100.0		30474	100.0	
	Gender	Female	29974	49.1	25886	48.9	-0.2	14891	48.9	-0.2
		Male	31117	50.9	27044	51.1	0.2	15583	51.1	0.2
	Race/ Ethnicity	American Indian	726	1.2	563	1.1	-0.1	297	1.0	-0.2
		Asian	2512	4.1	2237	4.2	0.0	1304	4.3	0.2
		African American	6565	10.7	4101	7.7	-3.0	3235	10.6	-0.1
		Hispanic	8295	13.6	6787	12.8	-0.8	4108	13.5	-0.1
		White	40204	65.8	36547	69.0	3.2	20141	66.1	0.3
		Two or More	2745	4.5	2695	5.1	0.6	1389	4.6	0.1
	Disability	No	53064	86.9	46009	86.9	0.1	26558	87.1	0.3
		Yes	8027	13.1	6921	13.1	-0.1	3916	12.9	-0.3
	LEP	No	55479	90.8	48597	91.8	1.0	27717	91.0	0.1
		Yes	5612	9.2	4333	8.2	-1.0	2757	9.0	-0.1
	Economically Disadv.	No	33672	55.1	32393	61.2	6.1	16870	55.4	0.2
		Yes	27419	44.9	20537	38.8	-6.1	13604	44.6	-0.2
	Locale	Non-Public	2989	4.9	3143	5.9	1.0	1479	4.9	0.0
		City	18707	30.6	12798	24.2	-6.4	9280	30.5	-0.2
		Suburban	16667	27.3	15705	29.7	2.4	8330	27.3	0.1
		Town	11901	19.5	10799	20.4	0.9	5969	19.6	0.1
		Rural	10827	17.7	10275	19.4	1.7	5416	17.8	0.0

Table G-2. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 4

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
4	All Students	All Students	63528	100.0	52706	100.0		31563	100.0	
	Gender	Female	31013	48.8	25829	49.0	0.2	15424	48.9	0.0
		Male	32515	51.2	26877	51.0	-0.2	16139	51.1	0.0
	Race/ Ethnicity	American Indian	761	1.2	599	1.1	-0.1	338	1.1	-0.1
		Asian	2704	4.3	2166	4.1	-0.2	1376	4.4	0.1
		African American	6998	11.0	4079	7.7	-3.3	3281	10.4	-0.6
		Hispanic	8682	13.7	6789	12.9	-0.8	4340	13.8	0.1
		White	41517	65.4	36592	69.4	4.1	20858	66.1	0.7
		Two or More	2822	4.4	2481	4.7	0.3	1370	4.3	-0.1
	Disability	No	55388	87.2	45875	87.0	-0.1	27586	87.4	0.2
		Yes	8140	12.8	6831	13.0	0.1	3977	12.6	-0.2
	LEP	No	57896	91.1	48560	92.1	1.0	28742	91.1	-0.1
		Yes	5632	8.9	4146	7.9	-1.0	2821	8.9	0.1
	Economically Disadv.	No	35067	55.2	32638	61.9	6.7	17528	55.5	0.3
		Yes	28461	44.8	20068	38.1	-6.7	14035	44.5	-0.3
	Locale	Non-Public	3277	5.2	3031	5.8	0.6	1665	5.3	0.1
		City	19131	30.1	12521	23.8	-6.4	9293	29.4	-0.7
		Suburban	17550	27.6	15643	29.7	2.1	8820	27.9	0.3
		Town	12269	19.3	11072	21.0	1.7	6126	19.4	0.1
		Rural	11301	17.8	10234	19.4	1.6	5659	17.9	0.1

Table G-3. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 5

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
5	All Students	All Students	64654	100.0	54010	100.0		32198	100.0	
	Gender	Female	31753	49.1	26516	49.1	0.0	15813	49.1	0.0
		Male	32901	50.9	27494	50.9	0.0	16385	50.9	0.0
	Race/ Ethnicity	American Indian	750	1.2	595	1.1	-0.1	331	1.0	-0.1
		Asian	2545	3.9	2047	3.7	-0.2	1326	4.1	0.2
		African American	7023	10.9	4255	7.9	-3.0	3401	10.6	-0.3
		Hispanic	8868	13.7	6900	12.8	-0.9	4532	14.1	0.4
		White	42646	66.0	37720	69.8	3.9	21235	66.0	0.0
		Two or More	2775	4.3	2493	4.6	0.3	1373	4.3	0.0
	Disability	No	56653	87.6	47300	87.6	0.0	28202	87.6	0.0
		Yes	8001	12.4	6710	12.4	0.0	3996	12.4	0.0
	LEP	No	59741	92.4	50550	93.6	1.2	29740	92.4	0.0
		Yes	4913	7.6	3460	6.4	-1.2	2458	7.6	0.0
	Economically Disadv.	No	35888	55.5	33268	61.6	6.1	17862	55.5	0.0
		Yes	28766	44.5	20742	38.4	-6.1	14336	44.5	0.0
	Locale	Non-Public	3114	4.8	2955	5.5	0.7	1592	4.9	0.1
		City	19339	29.9	12853	23.8	-6.1	9633	29.9	0.0
		Suburban	17932	27.7	15924	29.5	1.7	8901	27.6	-0.1
		Town	12769	19.7	11411	21.1	1.4	6349	19.7	0.0
		Rural	11500	17.8	10650	19.7	1.9	5723	17.8	0.0

Table G-4. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 6

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
6	All Students	All Students	65386	100.0	55511	100.0		32492	100.0	
	Gender	Female	31883	48.8	27169	48.9	0.2	15811	48.7	-0.1
		Male	33503	51.2	28342	51.1	-0.2	16681	51.3	0.1
	Race/ Ethnicity	American Indian	811	1.2	584	1.1	-0.2	342	1.1	-0.2
		Asian	2558	3.9	2256	4.0	0.1	1319	4.1	0.1
		African American	6906	10.6	4215	7.6	-3.0	3277	10.1	-0.5
		Hispanic	8814	13.5	7117	12.8	-0.7	4373	13.5	0.0
		White	43569	66.6	38850	70.0	3.4	21864	67.3	0.7
		Two or More	2688	4.1	2489	4.5	0.4	1317	4.1	-0.1
	Disability	No	57398	87.8	48973	88.2	0.4	28720	88.4	0.6
		Yes	7988	12.2	6538	11.8	-0.4	3772	11.6	-0.6
	LEP	No	61388	93.9	52523	94.6	0.7	30576	94.1	0.2
		Yes	3998	6.1	2988	5.4	-0.7	1916	5.9	-0.2
	Economically Disadv.	No	37111	56.8	34812	62.7	6.0	18599	57.2	0.5
		Yes	28275	43.2	20699	37.3	-6.0	13893	42.8	-0.5
	Locale	Non-Public	3138	4.8	3132	5.6	0.8	1588	4.9	0.1
		City	18674	28.6	12541	22.6	-6.0	9048	27.8	-0.7
		Suburban	18347	28.1	16612	29.9	1.9	9130	28.1	0.0
		Town	13290	20.3	11767	21.2	0.9	6718	20.7	0.4
		Rural	11937	18.3	11146	20.1	1.8	6008	18.5	0.2

Table G-5. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 7

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
7	All Students	All Students	63878	100.0	56295	100.0		31812	100.0	
	Gender	Female	31092	48.7	27448	48.8	0.1	15287	48.1	-0.6
		Male	32786	51.3	28847	51.2	-0.1	16525	51.9	0.6
	Race/ Ethnicity	American Indian	805	1.3	568	1.0	-0.3	357	1.1	-0.1
		Asian	2493	3.9	2129	3.7	-0.2	1313	4.1	0.2
		African American	6573	10.3	4251	7.6	-2.7	3153	9.9	-0.4
		Hispanic	8672	13.6	7199	12.8	-0.8	4349	13.7	0.1
		White	42845	67.1	39754	70.6	3.5	21399	67.3	0.2
		Two or More	2444	3.8	2394	4.3	0.4	1241	3.9	0.1
	Disability	No	56166	87.9	49987	88.8	0.9	28050	88.2	0.2
		Yes	7712	12.1	6308	11.2	-0.9	3762	11.8	-0.2
	LEP	No	60272	94.4	53350	94.8	0.4	29990	94.3	-0.1
		Yes	3606	5.6	2945	5.2	-0.4	1822	5.7	0.1
	Economically Disadv.	No	36985	57.9	35697	63.4	5.5	18438	58.0	0.1
		Yes	26893	42.1	20598	36.6	-5.5	13374	42.0	-0.1
	Locale	Non-Public	2993	4.7	3122	5.5	0.9	1484	4.7	0.0
		City	18131	28.4	12543	22.3	-6.1	8946	28.1	-0.3
		Suburban	18146	28.4	16785	29.8	1.4	9121	28.7	0.3
		Town	12796	20.0	12143	21.6	1.5	6385	20.1	0.0
		Rural	11812	18.5	11420	20.3	1.8	5876	18.5	0.0

Table G-6. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 8

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
8	All Students	All Students	63056	100.0	56756	100.0		31528	100.0	
	Gender	Female	30821	48.9	27533	48.5	-0.4	15507	49.2	0.3
		Male	32235	51.1	29223	51.5	0.4	16021	50.8	-0.3
	Race/ Ethnicity	American Indian	754	1.2	635	1.1	-0.1	386	1.2	0.0
		Asian	2391	3.8	2098	3.7	-0.1	1267	4.0	0.2
		African American	6310	10.0	4362	7.7	-2.3	3121	9.9	-0.1
		Hispanic	8064	12.8	7112	12.5	-0.3	4019	12.7	0.0
		White	43222	68.5	40254	70.9	2.4	21594	68.5	-0.1
		Two or More	2258	3.6	2295	4.0	0.5	1141	3.6	0.0
	Disability	No	55614	88.2	50301	88.6	0.4	27920	88.6	0.4
		Yes	7442	11.8	6455	11.4	-0.4	3608	11.4	-0.4
	LEP	No	60059	95.2	53893	95.0	-0.3	30029	95.2	0.0
		Yes	2997	4.8	2863	5.0	0.3	1499	4.8	0.0
	Economically Disadv.	No	37969	60.2	36531	64.4	4.2	18919	60.0	-0.2
		Yes	25087	39.8	20225	35.6	-4.2	12609	40.0	0.2
	Locale	Non-Public	2698	4.3	2990	5.3	1.0	1334	4.2	0.0
		City	17475	27.7	12548	22.1	-5.6	8678	27.5	-0.2
		Suburban	18146	28.8	16671	29.4	0.6	9010	28.6	-0.2
		Town	12878	20.4	12487	22.0	1.6	6516	20.7	0.2
		Rural	11859	18.8	11715	20.6	1.8	5990	19.0	0.2

Table G-7. Spring 2019 Census, Spring 2021 Tested Population, and Spring 2021 Calibration Sample Characteristics, Grade 10

Grade	Demo. variable	Student Group	Census 2019 N-count	Census 2019 %	Participants 2021 N-count	Participants 2021 %	Difference % Participants 2021-Census 2019	Calib. Sample 2021 N-count	Calib. Sample 2021 %	Difference % Calib. Sample 2021-Census 2019
10	All Students	All Students	63476	100.0	51433	100.0		30985	100.0	
	Gender	Female	31053	48.9	24958	48.5	-0.4	15207	49.1	0.2
		Male	32423	51.1	26475	51.5	0.4	15778	50.9	-0.2
	Race/ Ethnicity	American Indian	713	1.1	526	1.0	-0.1	313	1.0	-0.1
		Asian	2396	3.8	1805	3.5	-0.3	1226	4.0	0.2
		African American	5261	8.3	2421	4.7	-3.6	1976	6.4	-1.9
		Hispanic	7668	12.1	5531	10.8	-1.3	3636	11.7	-0.3
		White	45451	71.6	39414	76.6	5.0	22877	73.8	2.2
		Two or More	1944	3.1	1687	3.3	0.2	957	3.1	0.0
	Disability	No	56652	89.2	46307	90.0	0.8	27766	89.6	0.4
		Yes	6824	10.8	5126	10.0	-0.8	3219	10.4	-0.4
	LEP	No	60984	96.1	49799	96.8	0.7	29853	96.3	0.3
		Yes	2492	3.9	1634	3.2	-0.7	1132	3.7	-0.3
	Economically Disadv.	No	41371	65.2	36855	71.7	6.5	20622	66.6	1.4
		Yes	22105	34.8	14578	28.3	-6.5	10363	33.4	-1.4
	Locale	Non-Public	2216	3.5	1691	3.3	-0.2	1069	3.5	0.0
		City	16616	26.2	9103	17.7	-8.5	7473	24.1	-2.1
		Suburban	18935	29.8	16007	31.1	1.3	9647	31.1	1.3
		Town	13444	21.2	12333	24.0	2.8	6648	21.5	0.3
		Rural	12265	19.3	11636	22.6	3.3	6148	19.8	0.5

Table G-8. Items Flagged for Poor Fit based on Yen's Q1 Statistic

Content	Grade	Item Number in Calibration	Item Type	Total N	Z Score	Critical
ELA	3	2	CR	30465	171.24	81.24
		25	MC	30465	84.09	81.24
	4	34	CR	31558	204.45	84.15
	5	8	CR	32184	337.65	85.82
		39	CR	32184	155.12	85.82
	6	12	CR	32490	103.18	86.64
		34	CR	32490	98.99	86.64
	7	3	CR	31809	98.26	84.82
		15	CR	31809	99.27	84.82
		19	CR	31809	159.46	84.82
		22	CR	31809	131.10	84.82
		27	CR	31809	116.24	84.82
		29	CR	31809	90.43	84.82
	8	20	CR	31518	118.39	84.05
		31	CR	31518	103.04	84.05
		33	CR	31518	104.83	84.05
Mathematics	3	15	CR	30223	93.92	80.59
	5	6	CR	31953	103.77	85.21
		25	CR	31953	95.57	85.21
		40	CR	31953	105.59	85.21
	7	33	CR	31588	88.74	84.23
		38	CR	31588	195.58	84.23
	8	26	MC	31260	127.97	83.36
		42	MC	31260	132.49	83.36
Science	4	21	CR	31142	91.71	83.05
		24	CR	31142	133.99	83.05
	8	6	MC	30987	107.32	82.63
		28	CR	30987	115.13	82.63
Social Studies	10	39	MC	30898	89.70	82.39

Table G-9. Equating Results, Stocking and Lord Method

Content Area	Grade	Number of Anchors	Stocking and Lord TCC Method Results							Equating Constants	
			TCC Results		Parameter Comparison Statistics						
					a-Parameter		b-Parameter				
			# of Iterations	F Value	Corr.	# of RMSD Outliers*	Corr.	# of RMSD Outliers*	A	B
ELA	3	38	4	0.0595	0.98	2 (23, 30)	0.99	1 (4)	0.9698	-1.4475
	4	41	4	0.0526	0.98	2 (7, 11)	0.99	1 (30)	1.0587	-0.7934
	5	42	6	0.0186	0.99	2 (2, 7)	0.99	1 (7)	1.0192	-0.4690
	6	39	3	0.0936	0.99	1 (27)	0.99	2 (1,6)	1.0459	-0.1845
	7	37	4	0.0678	0.98	1 (6)	0.99	0	1.1461	0.2837
	8	40	4	0.0621	0.98	2 (12, 26)	0.99	1 (4)	1.2234	0.3691
Math	3	42	6	0.0611	0.98	1 (19)	0.99	2 (4, 19)	0.9925	-1.3901
	4	46	18	0.0273	0.93	2 (15, 42)	0.99	1 (42)	0.9835	-0.8837
	5	46	28	0.1013	0.98	1 (20)	0.99	1 (3)	0.9555	-0.3422
	6	46	26	0.0840	0.95	1 (32)	0.99	2 (12, 39)	1.0510	-0.2001
	7	46	23	0.1069	0.96	3 (3, 9, 20)	0.99	2 (11, 12)	1.0459	0.2288
	8	46	33	0.1005	0.97	2 (1,2)	0.99	1 (10)	1.0284	0.5898
Science	4	40	3	0.0271	1.00	0	1.00	1 (23)	1.0166	-0.1392
	8	40	3	0.0474	1.00	0	1.00	0	0.9854	-0.0792
Social Studies	4	38	7	0.0250	0.99	1 (34)	0.98	3 (8, 9, 26)	1.2093	-0.4170
	8	40	8	0.0387	0.99	2 (22, 40)	0.99	3 (20, 29, 40)	1.0714	-0.1864
	10	50	6	0.0769	0.98	4 (19, 37, 44, 47)	0.99	1 (9)	1.1243	-0.1819

*Item number is provided in parenthesis.

Table G-10. Items Flagged using Delta-Plot Method

Content Area	Grade	Number of Items	Item Number	*p*-value Spring 2021	*p*-value Spring 2019 (or most recent)	D	Critical D
ELA	3	2	11	0.28	0.35	0.1423	0.1111
			26	0.62	0.70	0.1297	0.1111
	4	2	7	0.62	0.58	-0.1892	0.1296
			30	0.60	0.68	0.1935	0.1296
	5	2	8	0.48	0.54	0.1230	0.1028
			20	0.60	0.59	-0.1093	0.1028
	6	1	7	0.41	0.53	0.2571	0.1316
	7	1	25	0.52	0.61	0.2035	0.1089
	8	1	6	0.59	0.65	0.1526	0.1204
Mathematics	4	3	29	0.30	0.38	0.1693	0.1424
			42	0.31	0.28	-0.2230	0.1424
			45	0.40	0.49	0.1545	0.1424
	5	1	21	0.16	0.21	0.1665	0.1389
	6	1	8	0.72	0.79	0.1827	0.1685
	7	2	4	0.20	0.27	0.1692	0.1436
			46	0.59	0.67	0.1707	0.1436
	8	2	1	0.58	0.66	0.1619	0.1443
			21	0.21	0.29	0.1817	0.1443
Science	4	2	22	0.51	0.50	-0.0985	0.0782
			24	0.37	0.43	0.1075	0.0782
	8	2	18	0.50	0.55	0.0927	0.0826
			34	0.45	0.43	-0.1074	0.0826
Social Studies	4	2	8	0.50	0.59	0.1783	0.1225
			9	0.75	0.81	0.1314	0.1225
	8	2	20	0.66	0.73	0.1762	0.1261
			40	0.60	0.68	0.1851	0.1261
	10	2	9	0.57	0.63	0.1801	0.1081
			19	0.66	0.69	0.1083	0.1081

Table G-11. Pre- and Post-equated Results: Scale Score Summary

Content	Grade	N-count	Scale Score Summary Based on Pre-equated Parameters		Scale Score Summary Based on Post-equated Parameters		Difference Pre-Post	
			Scale Score Mean	Scale Score SD	Scale Score Mean	Scale Score SD	Scale Score Mean Difference	Scale Score SD Difference
ELA	3	52865	550.21	46.54	550.01	46.57	0.20	-0.03
	4	52626	577.76	51.20	578.12	50.90	-0.36	0.30
	5	53924	593.23	48.92	593.42	48.81	-0.19	0.11
	6	55421	604.15	49.99	604.05	49.89	0.10	0.10
	7	56193	625.50	55.08	625.91	55.06	-0.41	0.02
	8	56619	628.42	58.61	628.23	58.52	0.19	0.09
Mathematics	3	52851	549.01	56.36	549.39	55.51	-0.38	0.85
	4	52594	571.56	53.29	571.75	52.87	-0.19	0.42
	5	53861	594.35	55.95	594.55	55.73	-0.20	0.22
	6	55375	602.20	57.61	602.26	57.27	-0.06	0.34
	7	56161	620.10	59.97	620.59	59.96	-0.49	0.01
	8	56617	638.47	56.84	637.75	57.54	0.72	-0.70
Science	4	52330	497.51	50.19	497.29	50.20	0.22	-0.01
	8	56338	697.48	49.74	697.46	49.53	0.02	0.21
Social Studies	4	52342	391.36	56.90	391.18	57.15	0.18	-0.25
	8	56278	595.60	51.95	595.40	51.85	0.20	0.10
	10	51132	697.24	55.38	697.03	55.39	0.21	-0.01

Table G-12. Pre- and Post-equated Results: Proficiency Level Summary

Content	Grade	N-count	Proficiency Level Summary Based on Pre-equated Parameters					Proficiency Level Summary Based on Post-equated Parameters					Difference Pre-Post
			% Below Basic	% Basic	% Prof.	% Adv.	% Prof. + Adv.	% Below Basic	% Basic	% Prof.	% Adv.	% Prof. + Adv.	% Prof. + Adv.
ELA	3	52865	27.64	37.77	29.51	5.08	34.59	27.64	37.77	29.66	4.93	34.59	0.00
	4	52626	27.10	32.74	32.71	7.46	40.17	26.71	33.09	32.71	7.49	40.20	-0.03
	5	53924	27.59	34.84	32.44	5.13	37.57	27.32	35.02	32.49	5.17	37.66	-0.09
	6	55421	25.10	36.39	30.70	7.81	38.51	25.31	36.44	30.33	7.92	38.25	0.25
	7	56193	23.00	34.02	34.17	8.82	42.99	22.75	34.05	34.30	8.90	43.20	-0.22
	8	56619	25.94	38.33	28.07	7.66	35.73	26.28	37.90	28.25	7.56	35.81	-0.09
Math	3	52851	23.74	31.23	34.48	10.54	45.02	23.72	31.48	34.25	10.56	44.81	0.22
	4	52594	22.28	36.61	30.87	10.24	41.11	22.34	36.49	30.90	10.27	41.16	-0.06
	5	53861	29.40	28.96	32.15	9.49	41.64	29.15	29.18	32.29	9.38	41.67	-0.03
	6	55375	32.54	31.84	30.95	4.67	35.62	32.26	31.93	31.25	4.56	35.81	-0.19
	7	56161	34.93	30.19	31.41	3.47	34.88	34.60	29.92	32.03	3.45	35.48	-0.61
	8	56617	32.39	37.56	23.66	6.39	30.05	33.05	36.58	24.09	6.28	30.37	-0.32
Science	4	52330	16.10	32.67	33.07	18.16	51.23	16.29	32.57	32.99	18.15	51.14	0.09
	8	56338	18.43	30.00	30.70	20.87	51.57	18.27	29.88	31.15	20.70	51.85	-0.28
Social Studies	4	52342	27.59	23.19	29.40	19.82	49.22	27.72	23.42	29.24	19.62	48.86	0.36
	8	56278	24.00	27.52	31.06	17.43	48.49	24.25	27.17	31.24	17.33	48.58	-0.09
	10	51132	27.84	24.69	27.27	20.19	47.47	27.78	24.52	27.67	20.03	47.69	-0.23

Table G-13. Pre- vs. Post-equated Proficiency Level Classification, English Language Arts

Pre-equated PL	ELA Grade 3 Post-equated PL					ELA Grade 4 Post-equated PL					ELA Grade 5 Post-equated PL				
	Below Basic	Basic	Prof.	Adv.	Total	Below Basic	Basic	Prof.	Adv.	Total	Below Basic	Basic	Prof.	Adv.	Total
Below Basic	14490 27.41	123 0.23	0 0.00	0 0.00	14613 27.64	14013 26.63	247 0.47	0 0.00	0 0.00	14260 27.10	14656 27.18	222 0.41	0 0.00	0 0.00	14878 27.59
Basic	123 0.23	19721 37.30	122 0.23	0 0.00	19966 37.77	46 0.09	17053 32.40	129 0.25	0 0.00	17228 32.74	78 0.14	18541 34.38	168 0.31	0 0.00	18787 34.84
Proficient	0 0.00	122 0.23	15462 29.25	14 0.03	15598 29.51	0 0.00	112 0.21	17025 32.35	77 0.15	17214 32.71	0 0.00	119 0.22	17299 32.08	77 0.14	17495 32.44
Advanced	0 0.00	0 0.00	97 0.18	2591 4.90	2688 5.08	0 0.00	0 0.00	59 0.11	3865 7.34	3924 7.46	0 0.00	0 0.00	53 0.10	2711 5.03	2764 5.13
Total	14613 27.64	19966 37.77	15681 29.66	2605 4.93	52865 100.00	14059 26.71	17412 33.09	17213 32.71	3942 7.49	52626 100.00	14734 27.32	18882 35.02	17520 32.49	2788 5.17	53924 100.00

Pre-equated PL	ELA Grade 6 Post-equated PL					ELA Grade 7 Post-equated PL					ELA Grade 8 Post-equated PL				
	Below Basic	Basic	Prof.	Adv.	Total	Below Basic	Basic	Prof.	Adv.	Total	Below Basic	Basic	Prof.	Adv.	Total
Below Basic	13873 25.03	38 0.07	0 0.00	0 0.00	13911 25.10	12731 22.66	191 0.34	0 0.00	0 0.00	12922 23.00	14657 25.89	30 0.05	0 0.00	0 0.00	14687 25.94
Basic	152 0.27	19926 35.95	91 0.16	0 0.00	20169 36.39	51 0.09	18847 33.54	217 0.39	0 0.00	19115 34.02	225 0.40	21336 37.68	142 0.25	0 0.00	21703 38.33
Proficient	0 0.00	232 0.42	16680 30.10	102 0.18	17014 30.70	0 0.00	95 0.17	18995 33.80	110 0.20	19200 34.17	0 0.00	93 0.16	15745 27.81	54 0.10	15892 28.07
Advanced	0 0.00	0 0.00	37 0.07	4290 7.74	4327 7.81	0 0.00	0 0.00	65 0.12	4891 8.70	4956 8.82	0 0.00	0 0.00	108 0.19	4229 7.47	4337 7.66
Total	14025 25.31	20196 36.44	16808 30.33	4392 7.92	55421 100.00	12782 22.75	19133 34.05	19277 34.30	5001 8.90	56193 100.00	14882 26.28	21459 37.90	15995 28.25	4283 7.56	56619 100.00

Note: Student count is in the top row of each cell. Percentage of students in each performance level is in the bottom row of each cell.

Table G-14. Pre- vs. Post-equated Proficiency Level Classification, Mathematics

Mathematics Grade 3

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	12463 23.58	85 0.16	0 0.00	0 0.00	12548 23.74
Basic	73 0.14	16379 30.99	55 0.10	0 0.00	16507 31.23
Proficient	0 0.00	171 0.32	17979 34.02	74 0.14	18224 34.48
Advanced	0 0.00	0 0.00	65 0.12	5507 10.42	5572 10.54
Total	12536 23.72	16635 31.48	18099 34.25	5581 10.56	52851 100.00

Mathematics Grade 4

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	11623 22.10	95 0.18	0 0.00	0 0.00	11718 22.28
Basic	128 0.24	19021 36.17	108 0.21	0 0.00	19257 36.61
Proficient	0 0.00	78 0.15	16093 30.60	64 0.12	16235 30.87
Advanced	0 0.00	0 0.00	49 0.09	5335 10.14	5384 10.24
Total	11751 22.34	19194 36.49	16250 30.90	5399 10.27	52594 100.00

Mathematics Grade 5

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	15591 28.95	244 0.45	0 0.00	0 0.00	15835 29.40
Basic	111 0.21	15325 28.45	162 0.30	0 0.00	15598 28.96
Proficient	0 0.00	146 0.27	17135 31.81	38 0.07	17319 32.15
Advanced	0 0.00	0 0.00	95 0.18	5014 9.31	5109 9.49
Total	15702 29.15	15715 29.18	17392 32.29	5052 9.38	53861 100.00

Mathematics Grade 6

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	17771 32.09	247 0.45	0 0.00	0 0.00	18018 32.54
Basic	92 0.17	17368 31.36	173 0.31	0 0.00	17633 31.84
Proficient	0 0.00	68 0.12	17039 30.77	30 0.05	17137 30.95
Advanced	0 0.00	0 0.00	90 0.16	2497 4.51	2587 4.67
Total	17863 32.26	17683 31.93	17302 31.25	2527 4.56	55375 100.00

Mathematics Grade 7

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	19374 34.50	245 0.44	0 0.00	0 0.00	19619 34.93
Basic	56 0.10	16550 29.47	349 0.62	0 0.00	16955 30.19
Proficient	0 0.00	8 0.01	17593 31.33	37 0.07	17638 31.41
Advanced	0 0.00	0 0.00	47 0.08	1902 3.39	1949 3.47
Total	19430 34.60	16803 29.92	17989 32.03	1939 3.45	56161 100.00

Mathematics Grade 8

Pre-equated PL	Post-equated PL Below Basic	Basic	Prof.	Adv.	Total
Below Basic	18254 32.24	82 0.14	0 0.00	0 0.00	18336 32.39
Basic	456 0.81	20581 36.35	228 0.40	0 0.00	21265 37.56
Proficient	0 0.00	47 0.08	13336 23.55	13 0.02	13396 23.66
Advanced	0 0.00	0 0.00	77 0.14	3543 6.26	3620 6.39
Total	18710 33.05	20710 36.58	13641 24.09	3556 6.28	56617 100.00

Note: Student count is in the top row of each cell. Percentage of students in each performance level is in the bottom row of each cell.

Table G-15. Pre- vs. Post-equated Proficiency Level Classification, Science

Pre-equated PL	Science Grade 4					Science Grade 8				
	Post-equated PL					Post-equated PL				
	Below Basic	Basic	Proficient	Advanced	Total	Below Basic	Basic	Proficient	Advanced	Total
Below Basic	8393 16.04	32 0.06	0 0.00	0 0.00	8425 16.10	10230 18.16	155 0.28	0 0.00	0 0.00	10385 18.43
Basic	134 0.26	16869 32.24	95 0.18	0 0.00	17098 32.67	64 0.11	16623 29.51	212 0.38	0 0.00	16899 30.00
Proficient	0 0.00	141 0.27	17070 32.62	94 0.18	17305 33.07	0 0.00	57 0.10	17195 30.52	44 0.08	17296 30.70
Advanced	0 0.00	0 0.00	99 0.19	9403 17.97	9502 18.16	0 0.00	0 0.00	140 0.25	11618 20.62	11758 20.87
Total	8527 16.29	17042 32.57	17264 32.99	9497 18.15	52330 100.00	10294 18.27	16835 29.88	17547 31.15	11662 20.70	56338 100.00

Note: Student count is in the top row of each cell. Percentage of students in each performance level is in the bottom row of each cell.

Table G-16. Pre- vs. Post-equated Proficiency Level Classification, Social Studies

Pre-equated PL	Social Studies Grade 4					Social Studies Grade 8					Social Studies Grade 10				
	Post-equated PL					Post-equated PL					Post-equated PL				
	Below Basic	Basic	Proficient	Advanced	Total	Below Basic	Basic	Proficient	Advanced	Total	Below Basic	Basic	Proficient	Advanced	Total
Below Basic	14373 27.46	70 0.13	0 0.00	0 0.00	14443 27.59	13448 23.90	56 0.10	0 0.00	0 0.00	13504 24.00	14108 27.59	129 0.25	0 0.00	0 0.00	14237 27.84
Basic	136 0.26	11939 22.81	62 0.12	0 0.00	12137 23.19	201 0.36	15097 26.83	187 0.33	0 0.00	15485 27.52	98 0.19	12377 24.21	150 0.29	0 0.00	12625 24.69
Proficient	0 0.00	251 0.48	15036 28.73	99 0.19	15386 29.40	0 0.00	138 0.25	17234 30.62	106 0.19	17478 31.06	0 0.00	33 0.06	13842 27.07	69 0.13	13944 27.27
Advanced	0 0.00	0 0.00	207 0.40	10169 19.43	10376 19.82	0 0.00	0 0.00	163 0.29	9648 17.14	9811 17.43	0 0.00	0 0.00	155 0.30	10171 19.89	10326 20.19
Total	14509 27.72	12260 23.42	15305 29.24	10268 19.62	52342 100.00	13649 24.25	15291 27.17	17584 31.24	9754 17.33	56278 100.00	14206 27.78	12539 24.52	14147 27.67	10240 20.03	51132 100.00

Note: Student count is in the top row of each cell. Percentage of students in each performance level is in the bottom row of each cell.

Figure G-1. Anchor Set Input and Estimate Test Characteristic Curves, English Language Arts
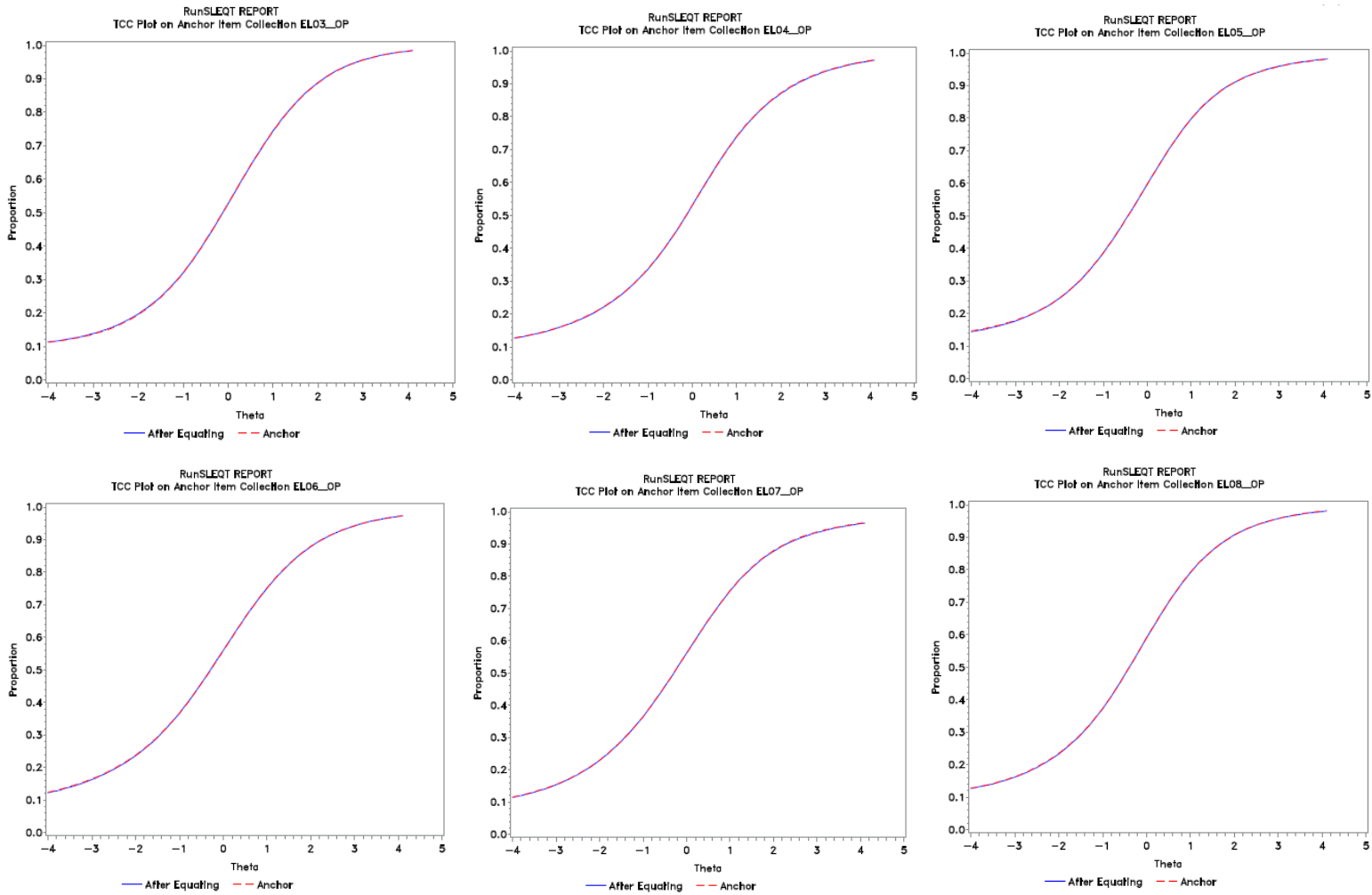
Figure G-2. Anchor Set Input and Estimate Test Characteristic Curves, Mathematics
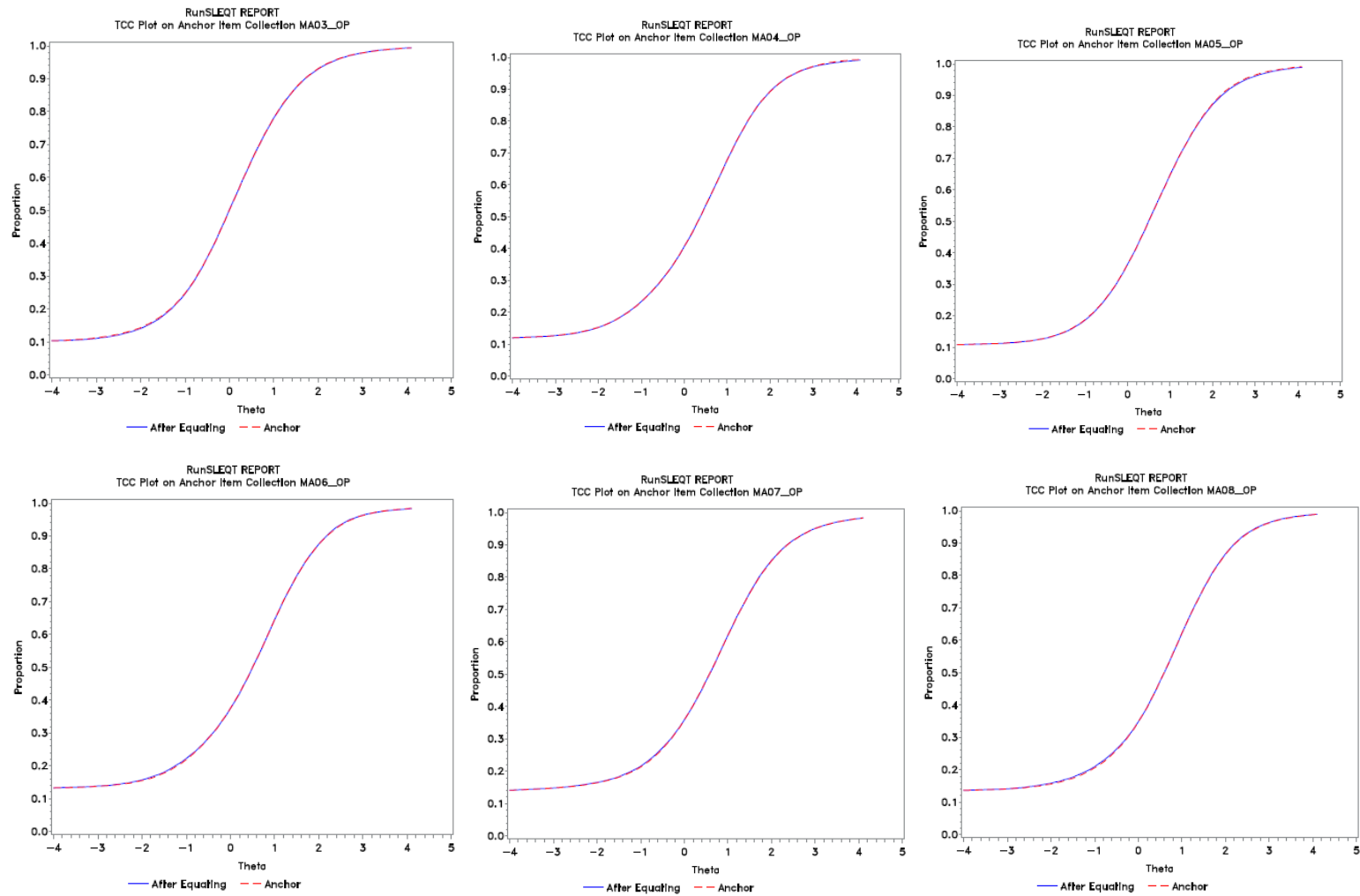
Figure G-3. Anchor Set Input and Estimate Test Characteristic Curves, Science
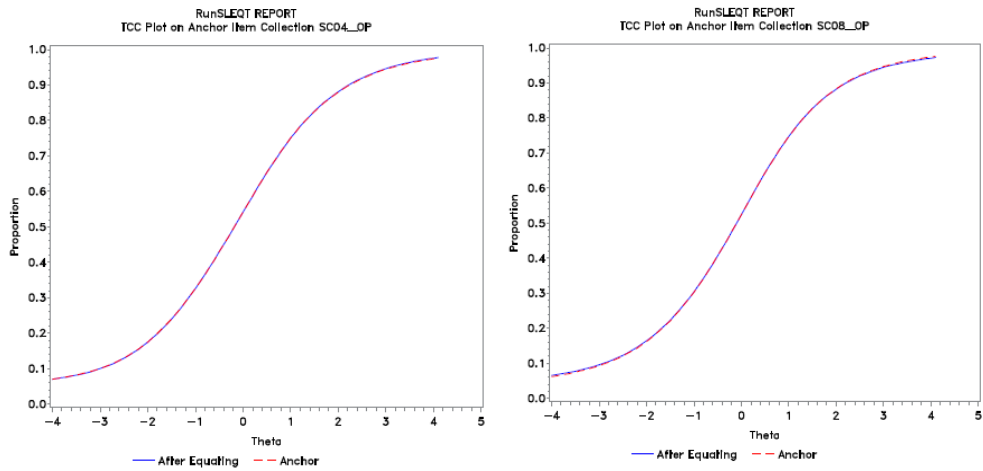
Figure G-4. Anchor Set Input and Estimate Test Characteristic Curves, Social Studies
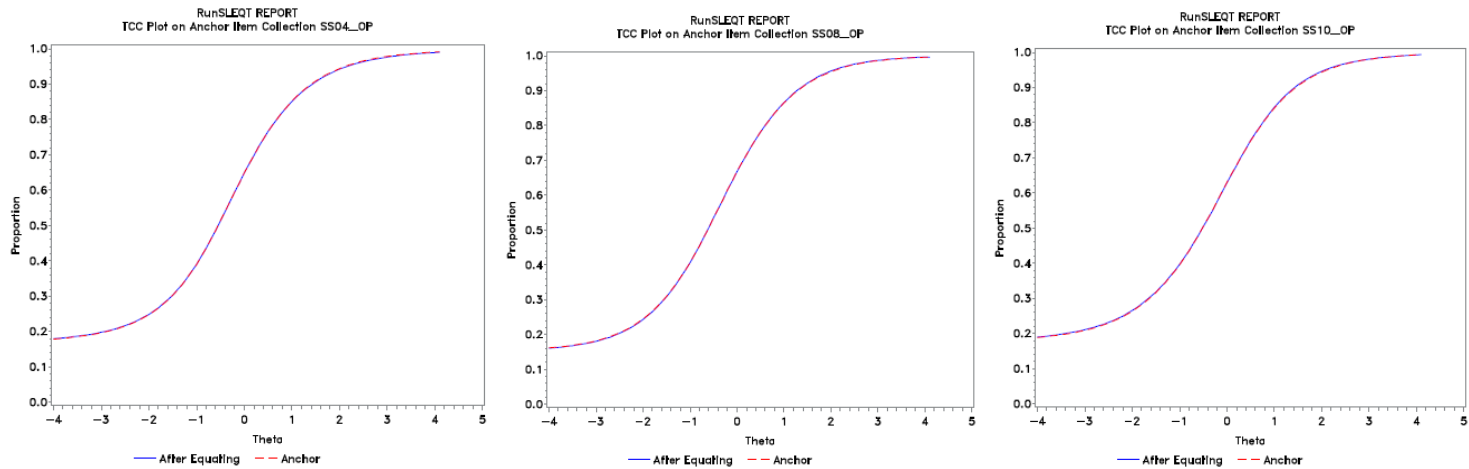
Figure G-5. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 3
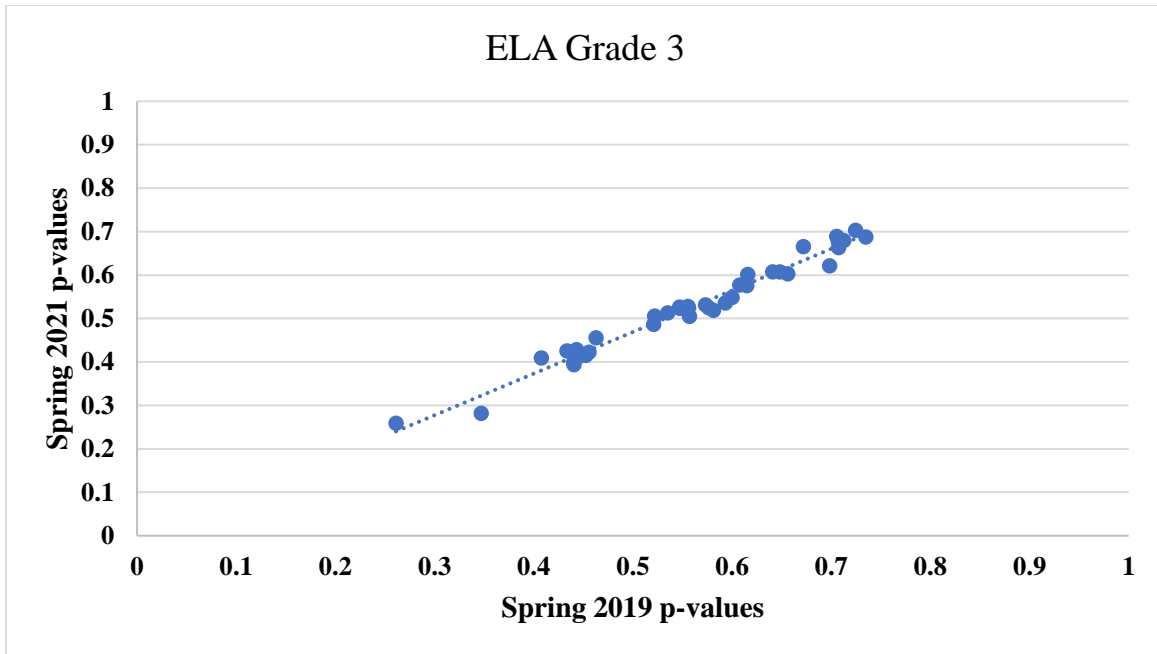
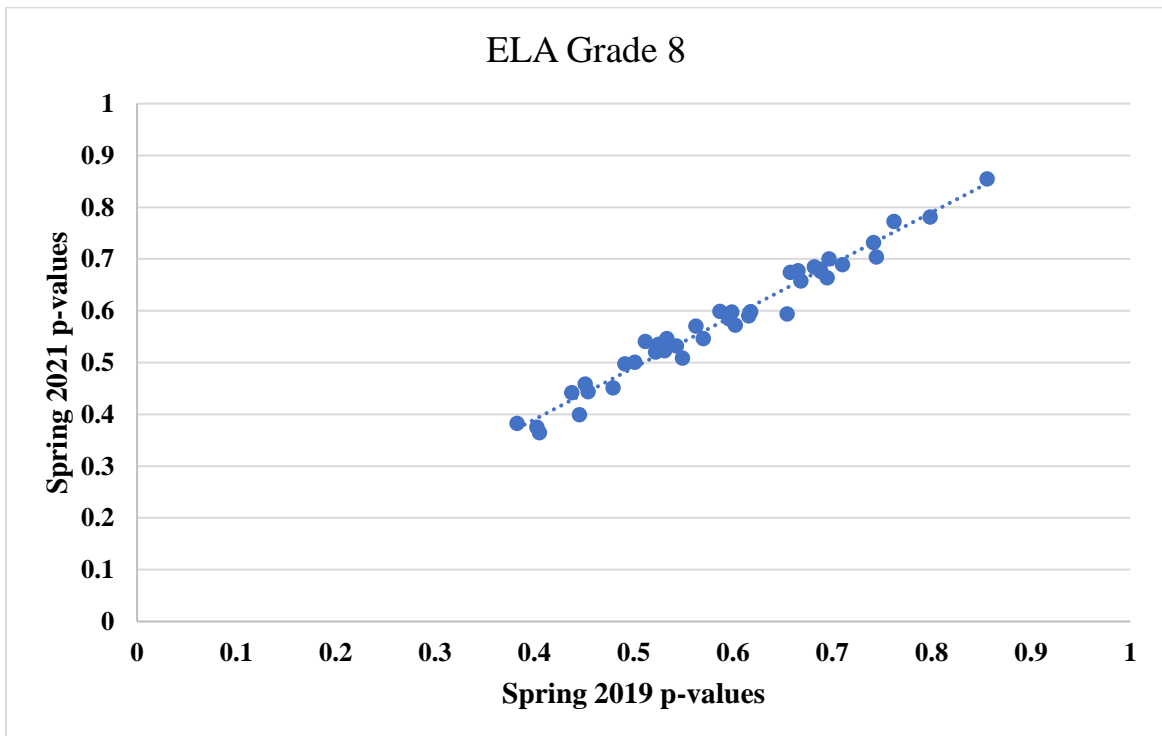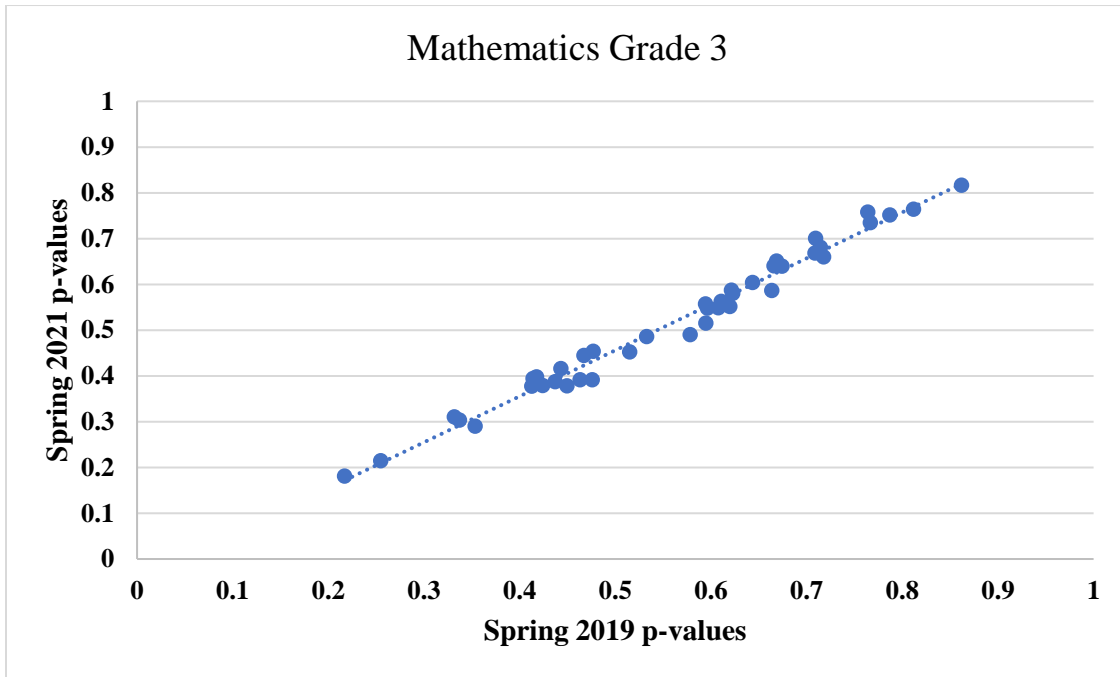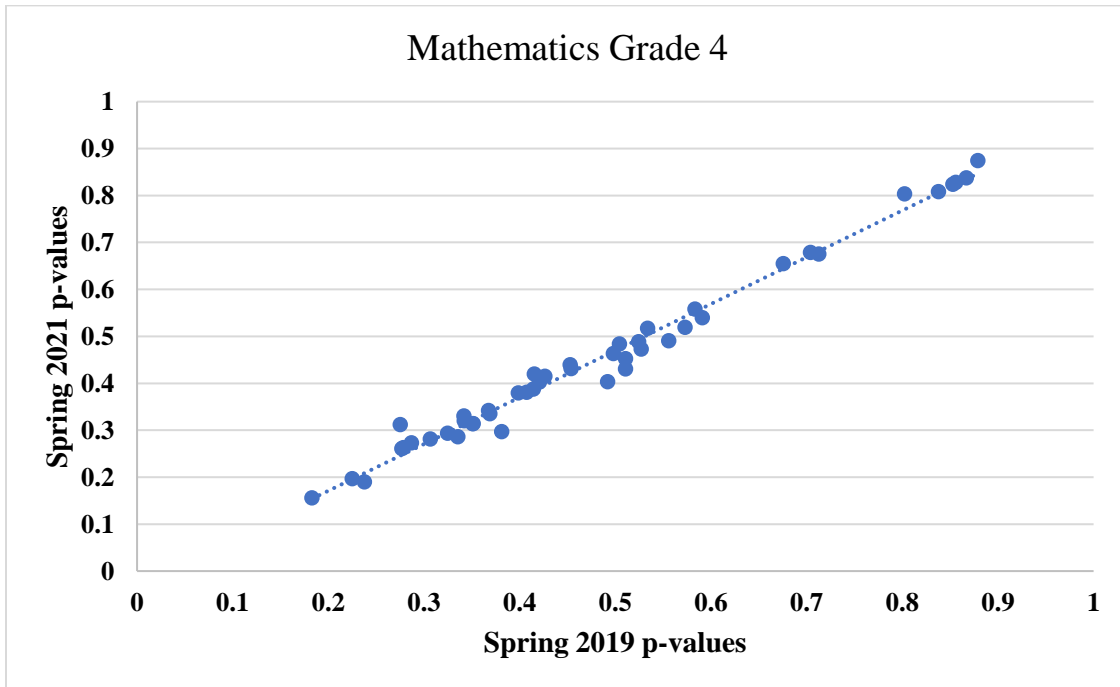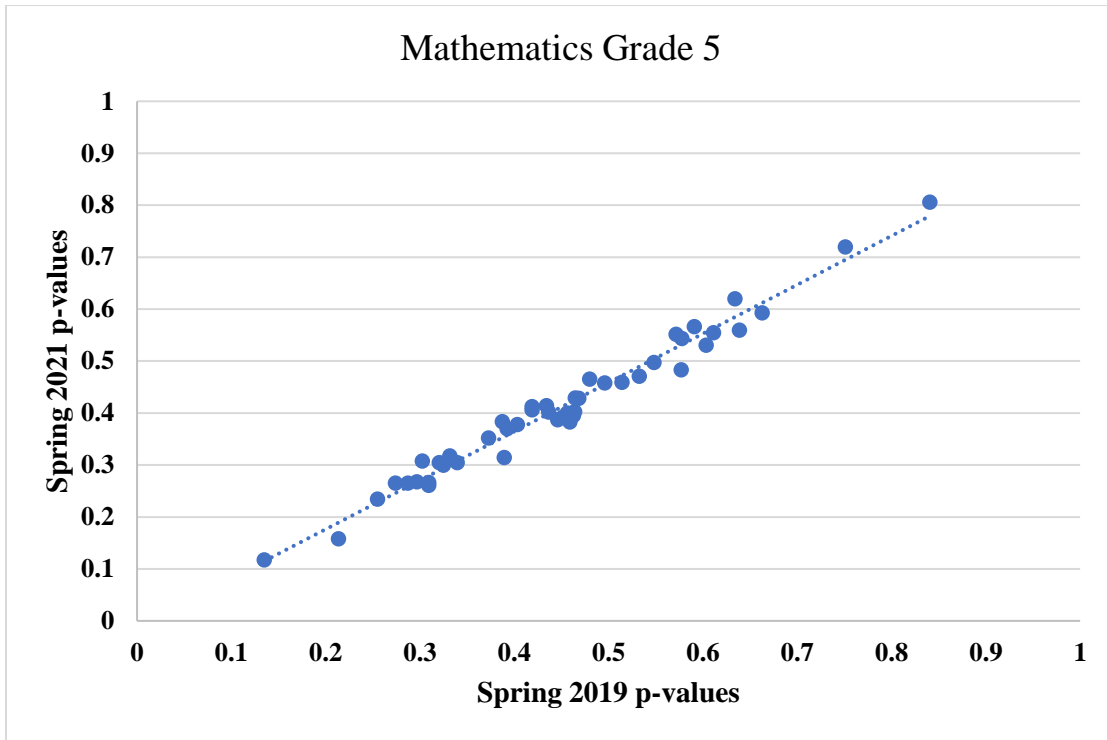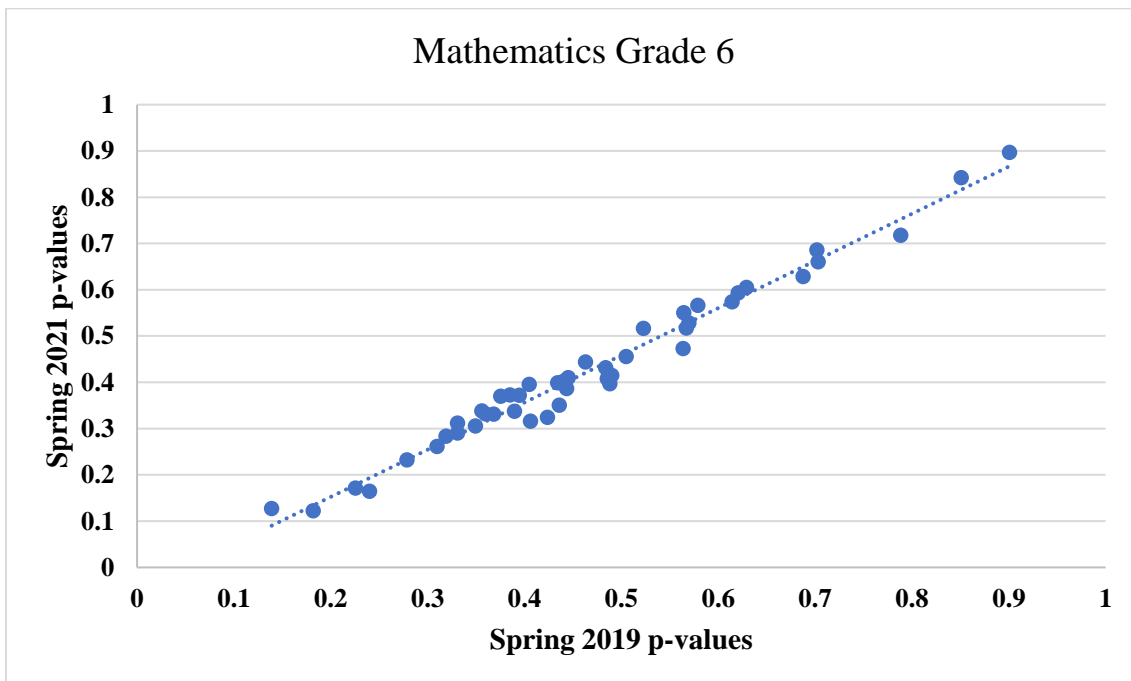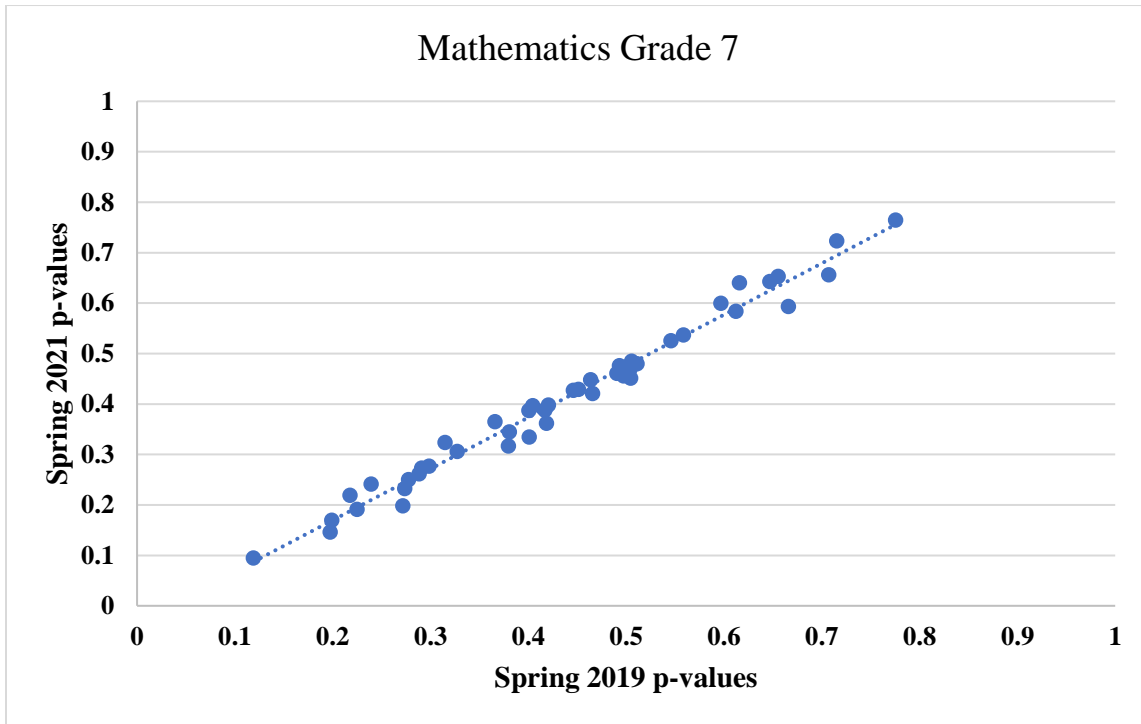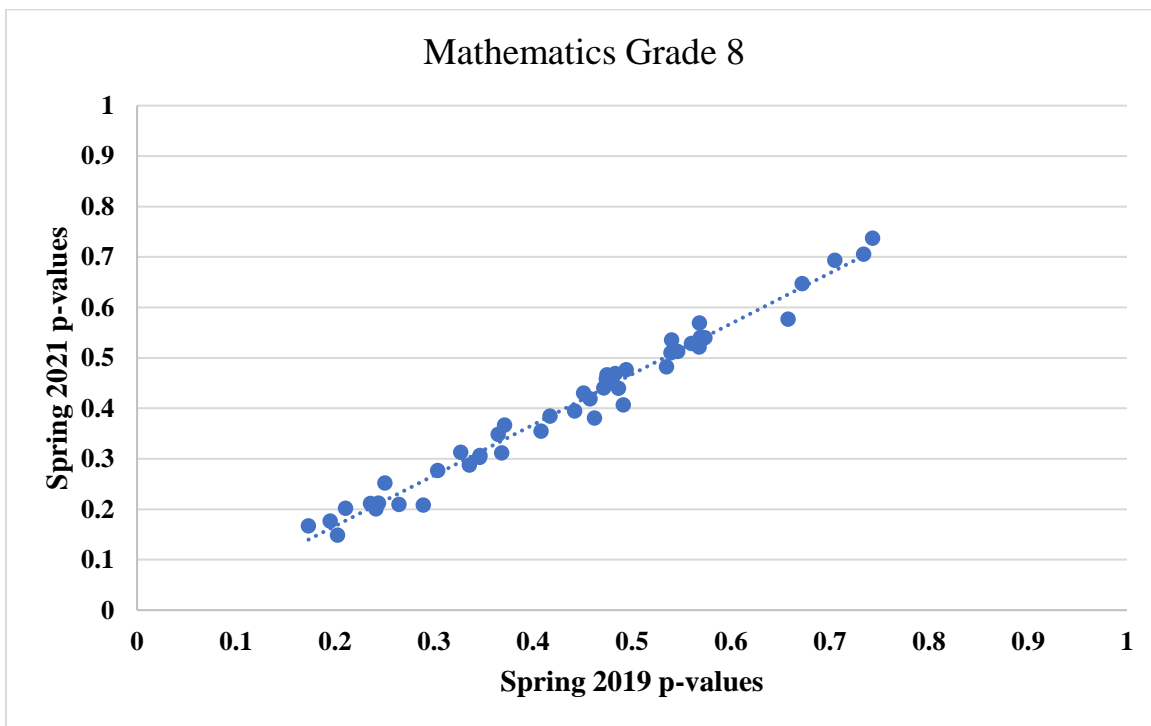Figure G-6. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 4
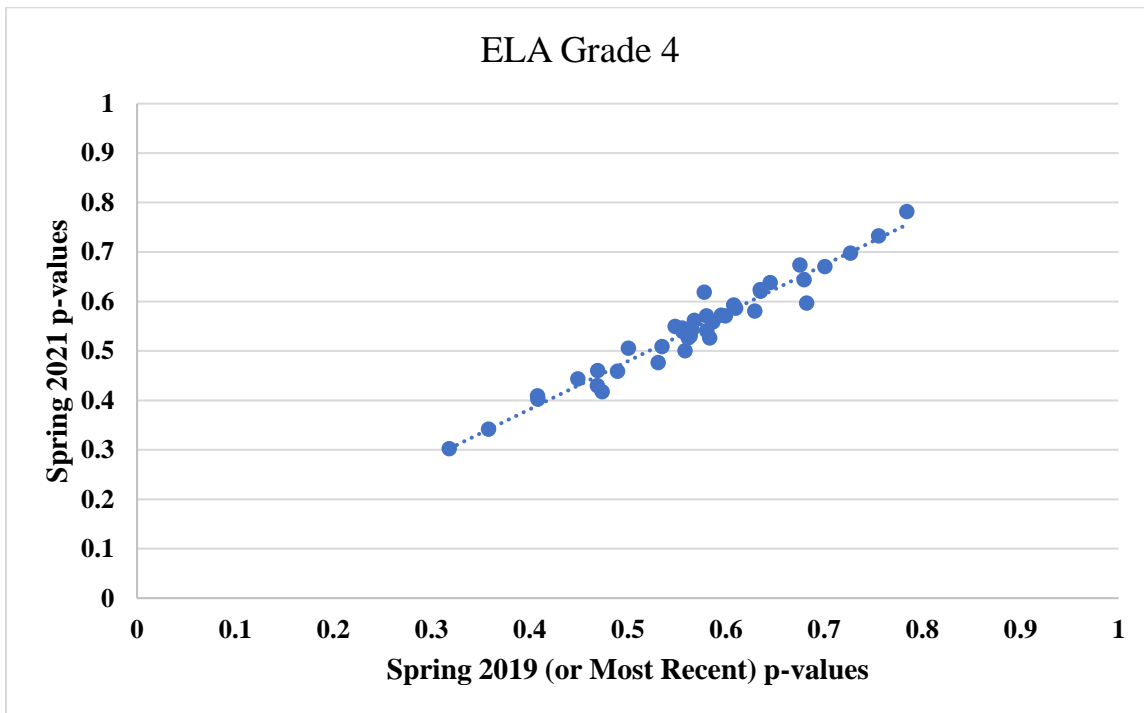
Figure G-7. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 5
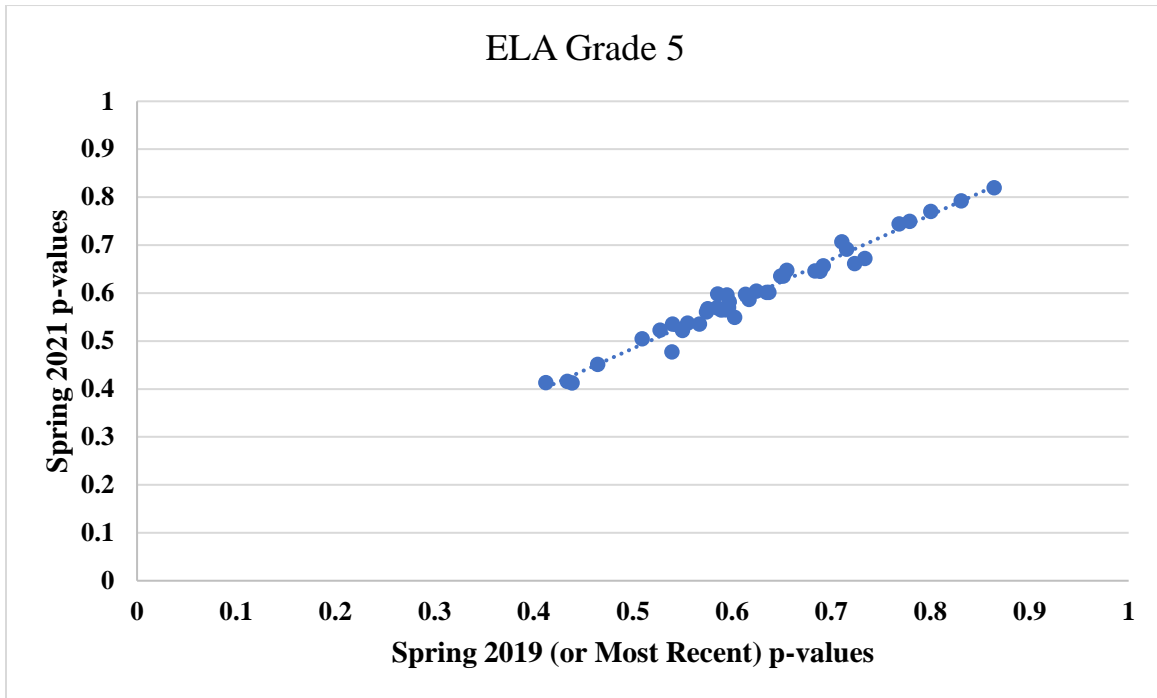
Figure G-8. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 6
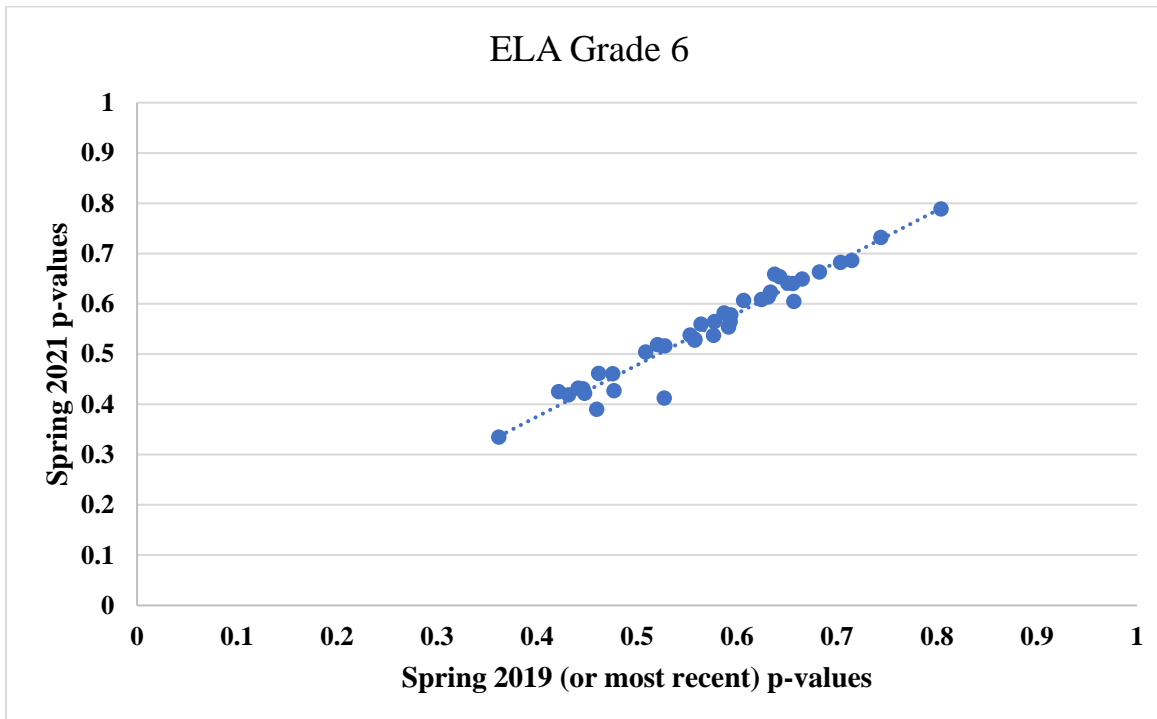
Figure G-9. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 7
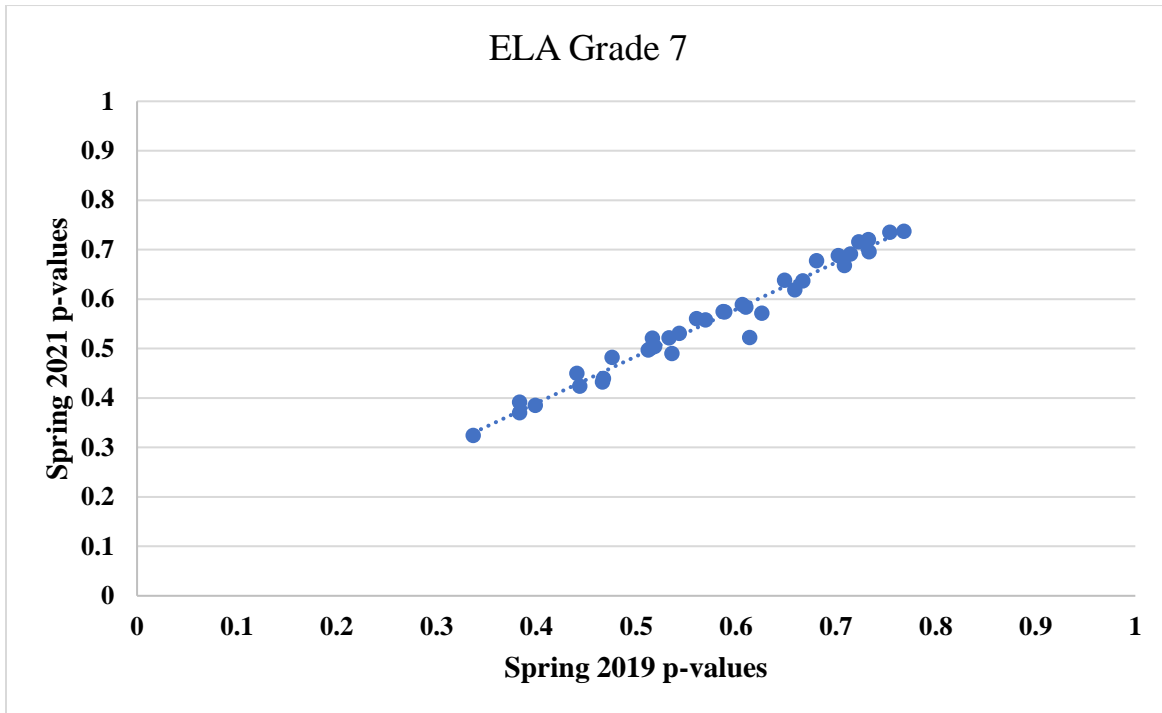
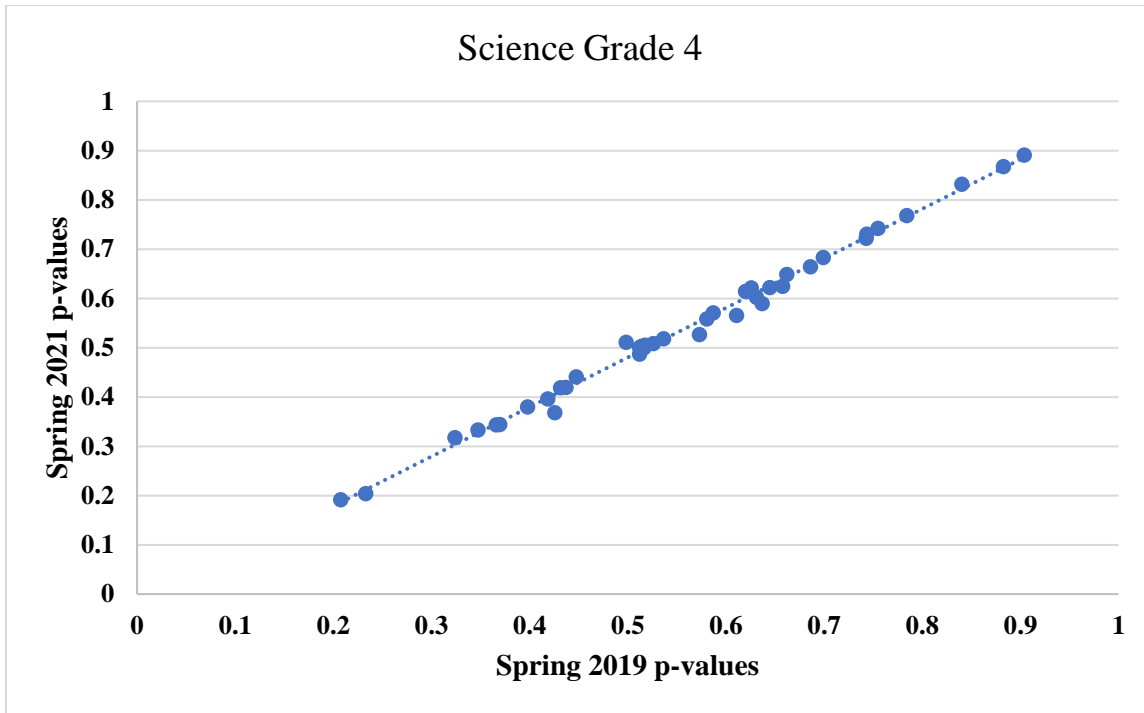Figure G-10. P-value Plot Spring 2019 vs. Spring 2021, ELA Grade 8

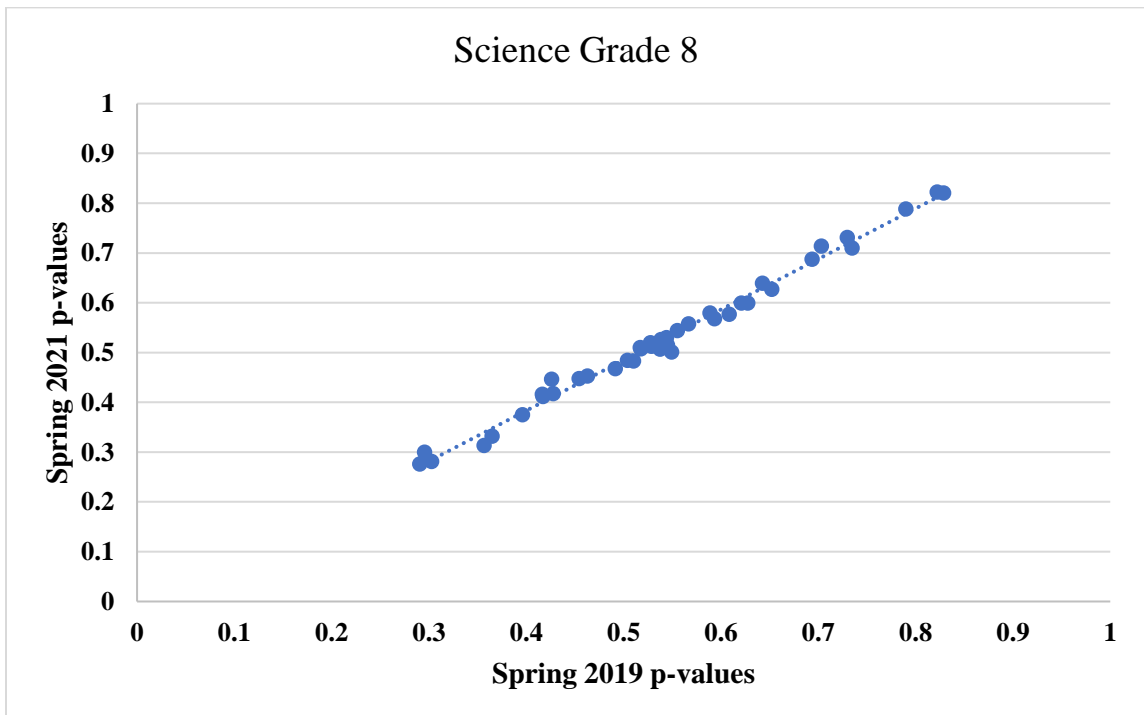Figure G-11. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 3

Mathematics Grade 3

Figure G-12. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 4

Mathematics Grade 4

Figure G-13. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 5

Mathematics Grade 5

Figure G-14. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 6

Mathematics Grade 6

Figure G-15. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 7

Figure G-16. P-value Plot Spring 2019 vs. Spring 2021, Mathematics Grade 8

Figure G-17. P-value Plot Spring 2019 vs. Spring 2021, Science Grade 4

Figure G-18. P-value Plot Spring 2019 vs. Spring 2021, Science Grade 8

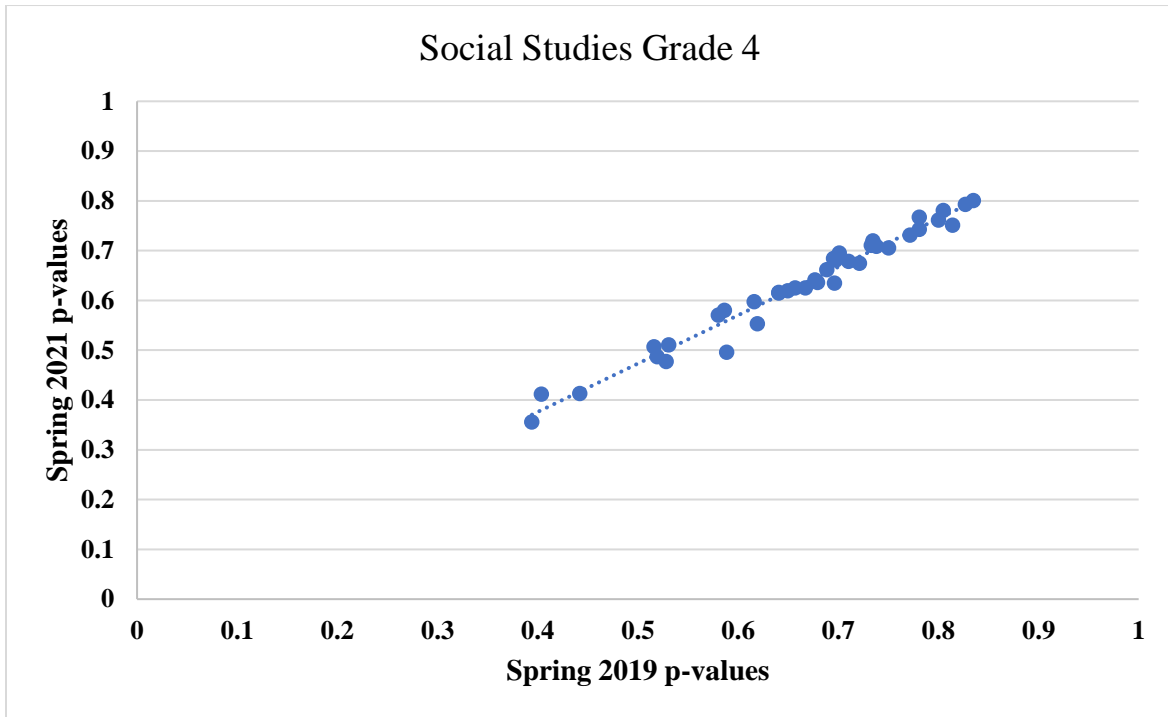Figure G-19. P-value Plot Spring 2019 vs. Spring 2021, Social Studies Grade 4

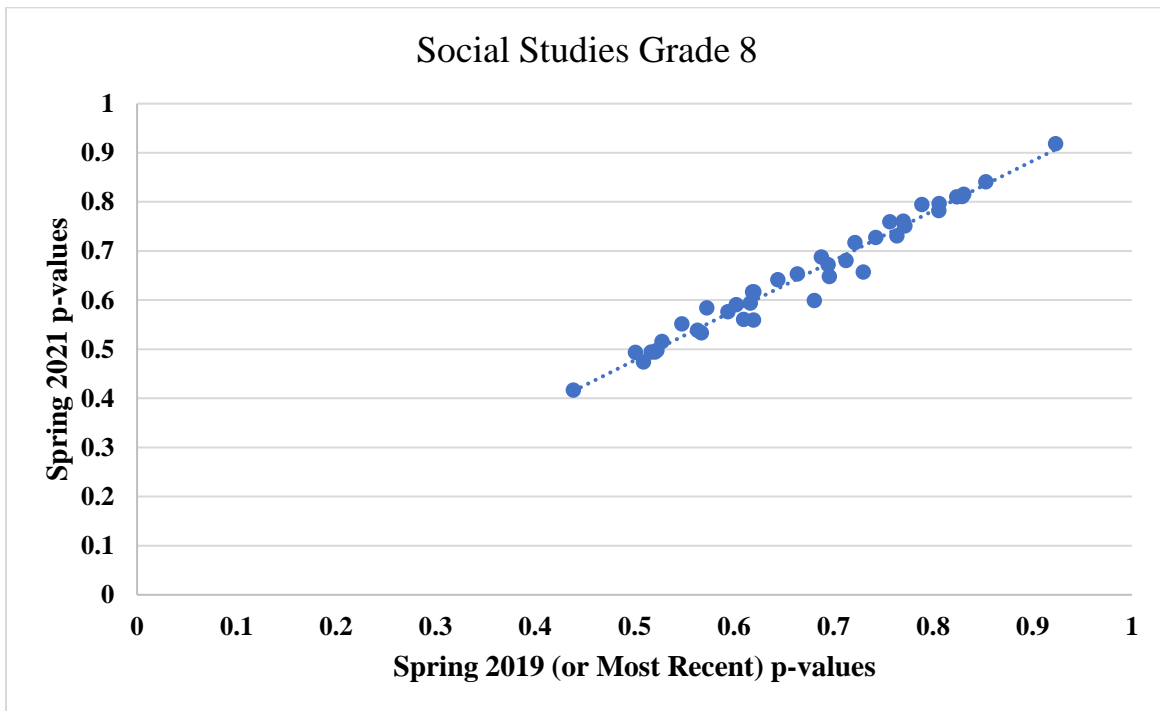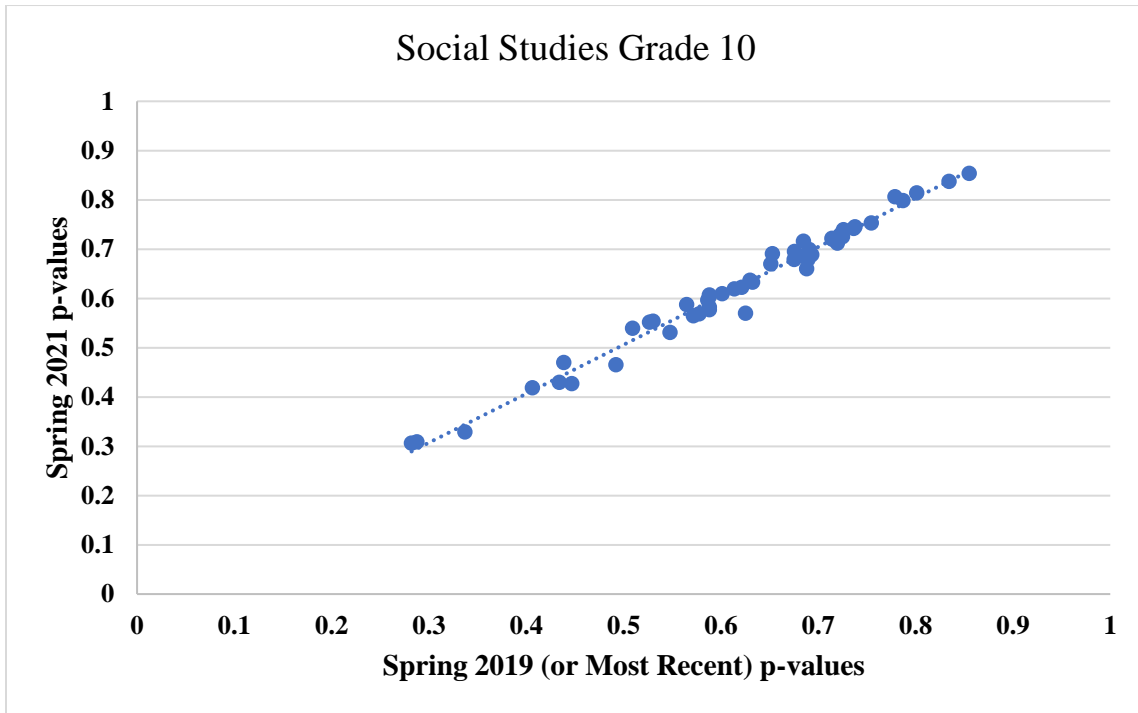Figure G-20. P-value Plot Spring 2019 vs. Spring 2021, Social Studies Grade 8

Figure G-21. P-value Plot Spring 2019 vs. Spring 2021, Social Studies Grade 10

Social Studies Grade 10

Appendix H

Conditional Standard Error of Measurement with Cut Scores

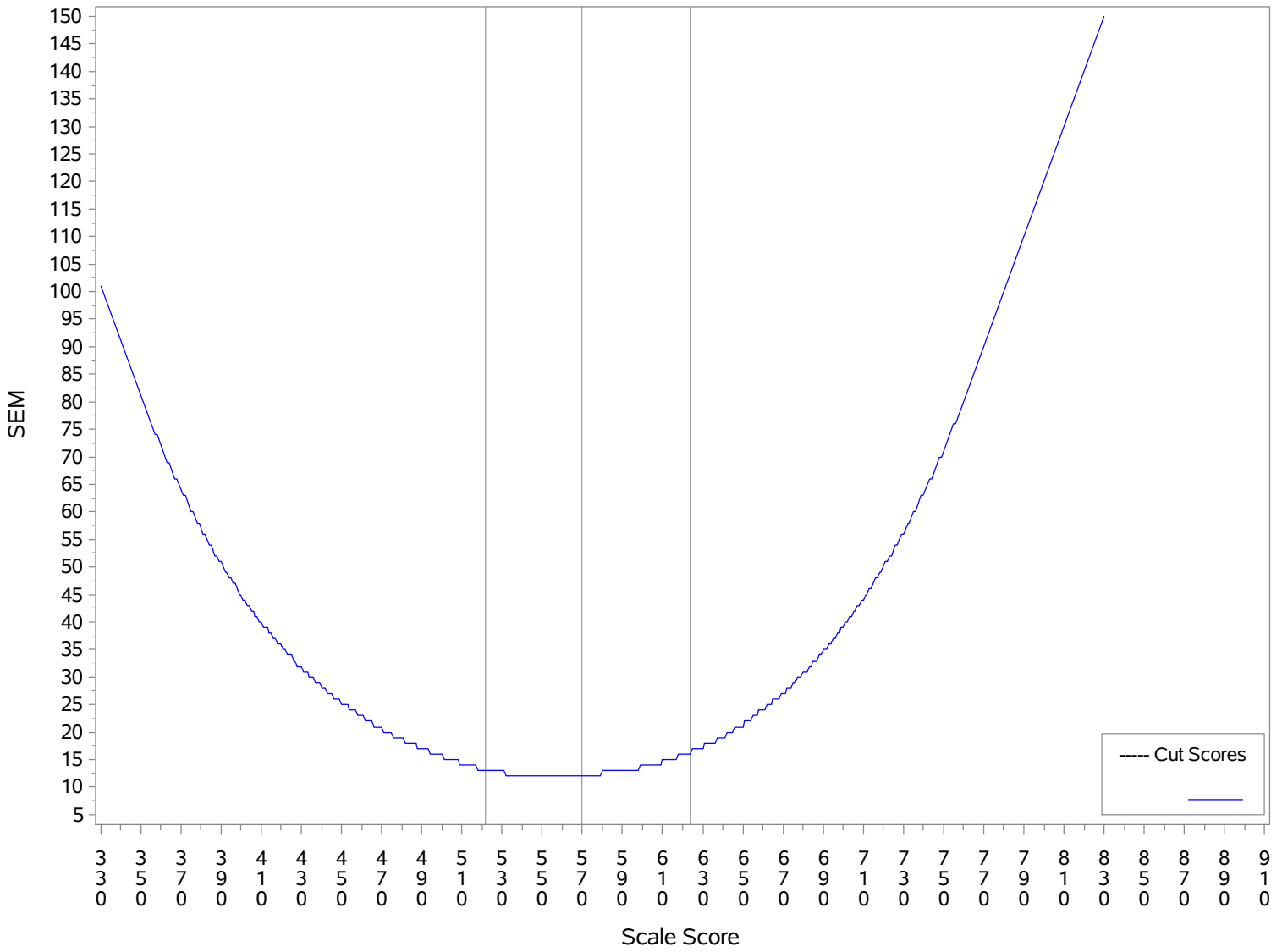**F**igure H-1 CSEM with cut scores, ELA Grade 3
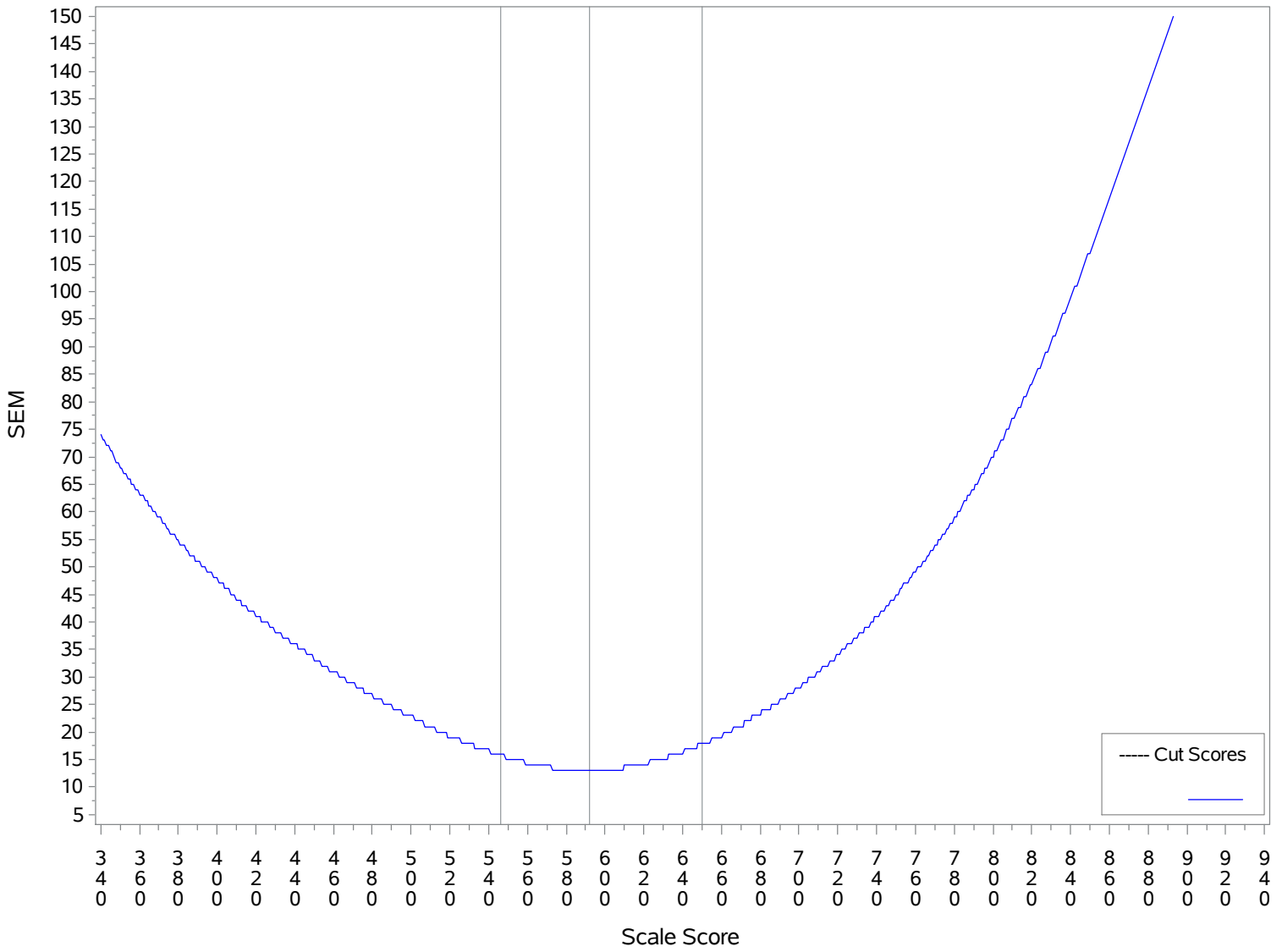
Figure H-2 CSEM with cut scores, ELA Grade 4

Figure H-3 CSEM with cut scores, ELA Grade 5
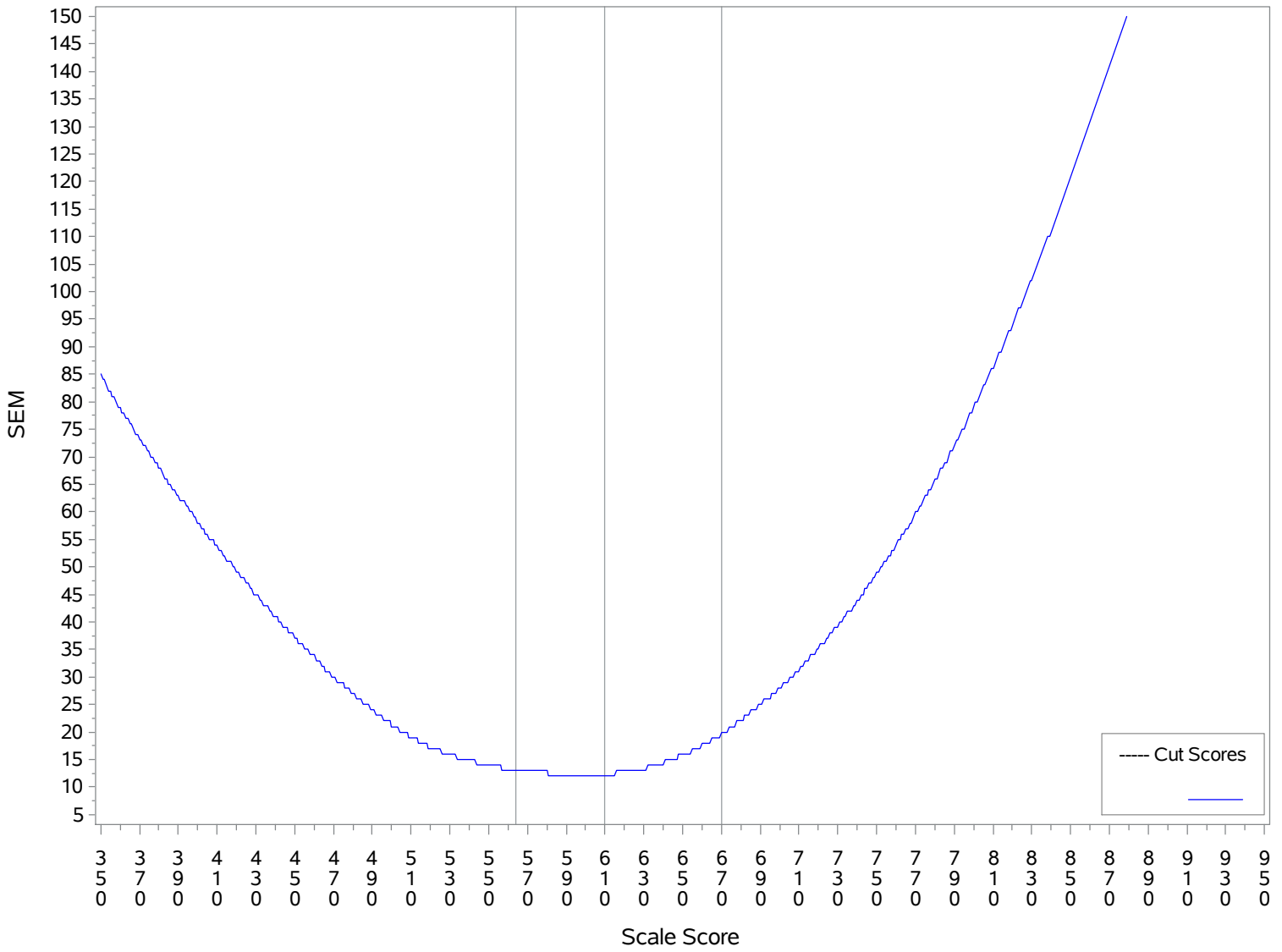
Figure H-4 CSEM with cut scores, ELA Grade 6
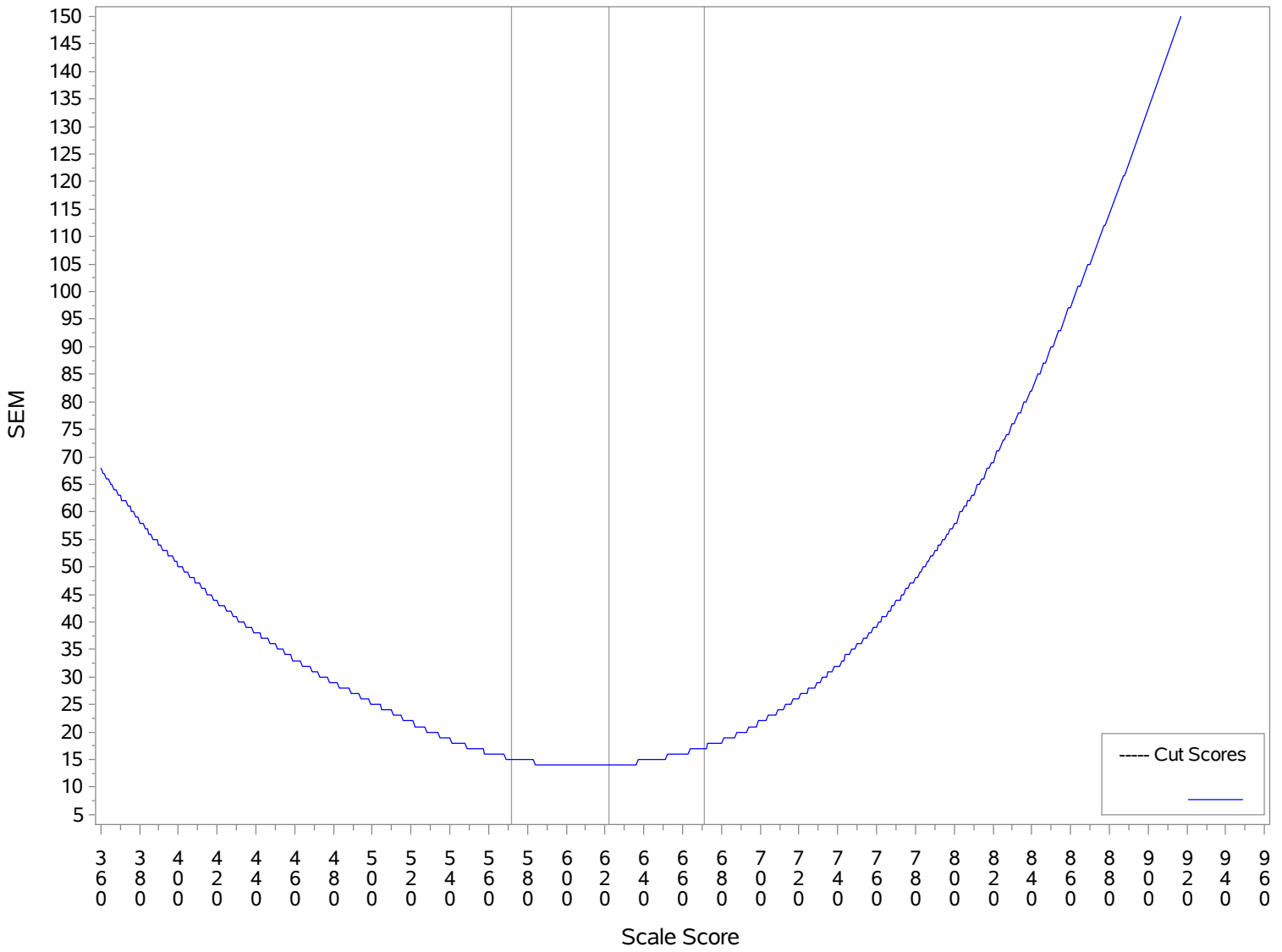
Figure H-5 CSEM with cut scores, ELA Grade 7
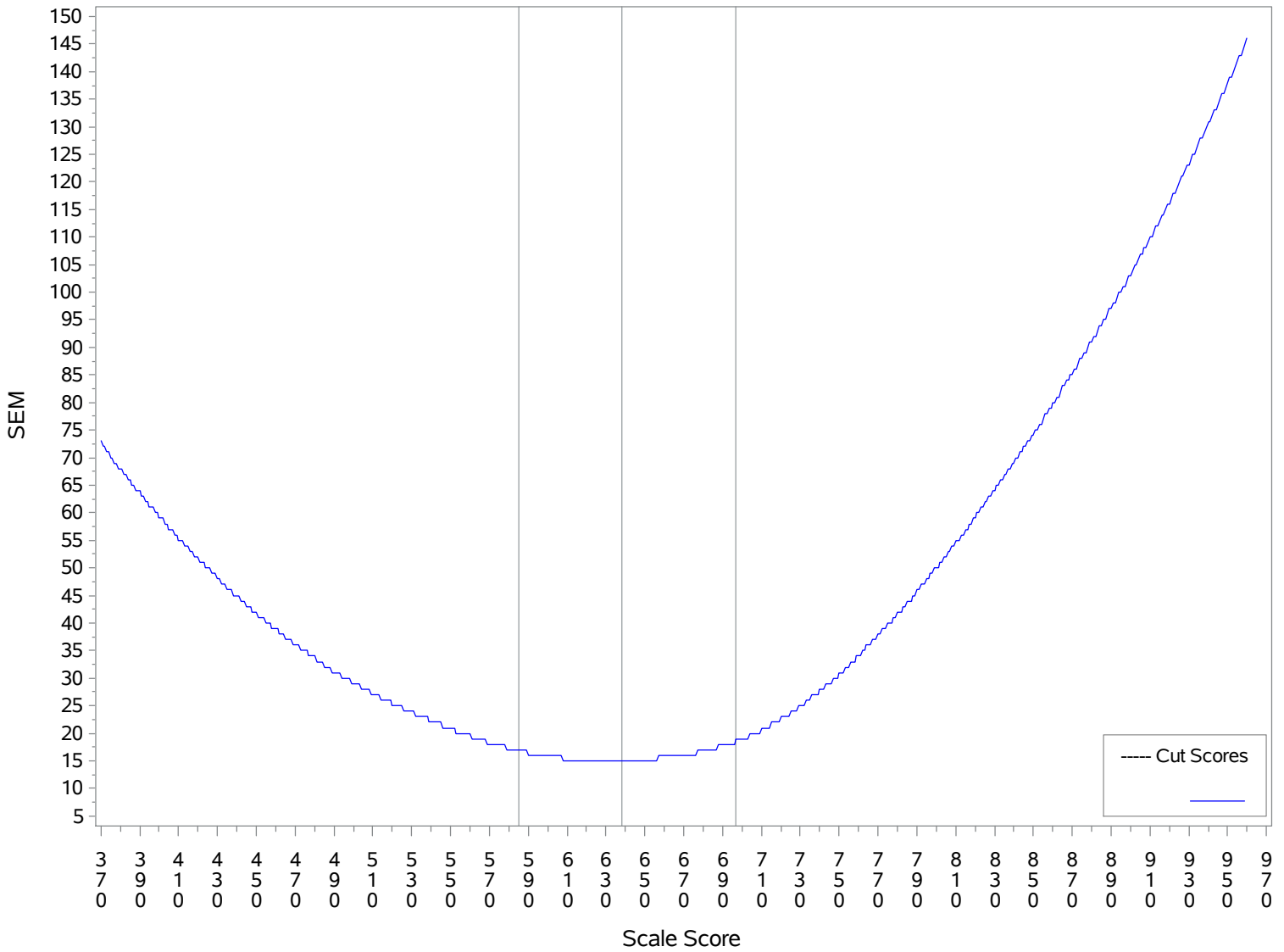
Figure H-6 CSEM with cut scores, ELA Grade 8
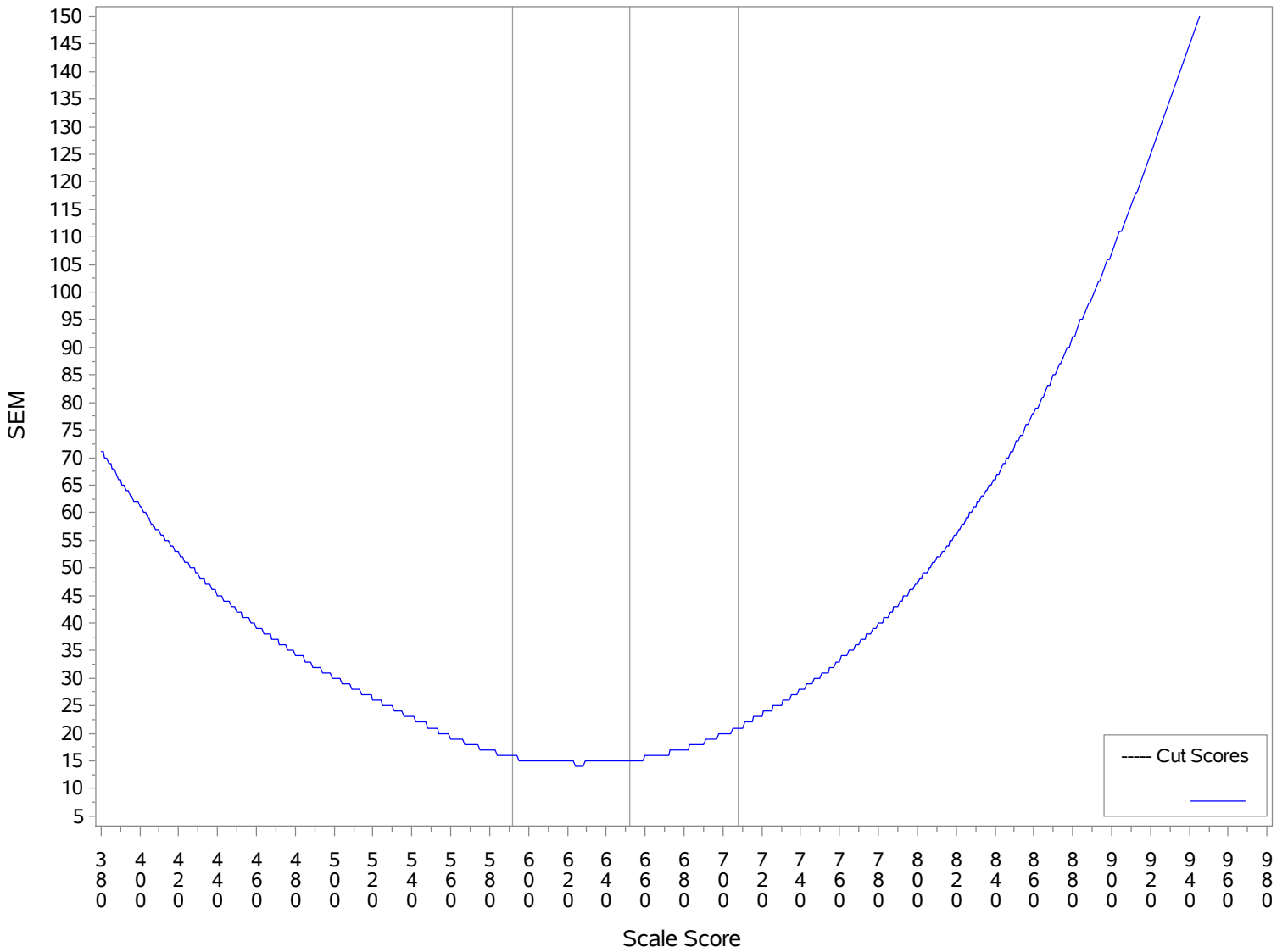
Figure H-7 CSEM with cut scores, Mathematics Grade 3
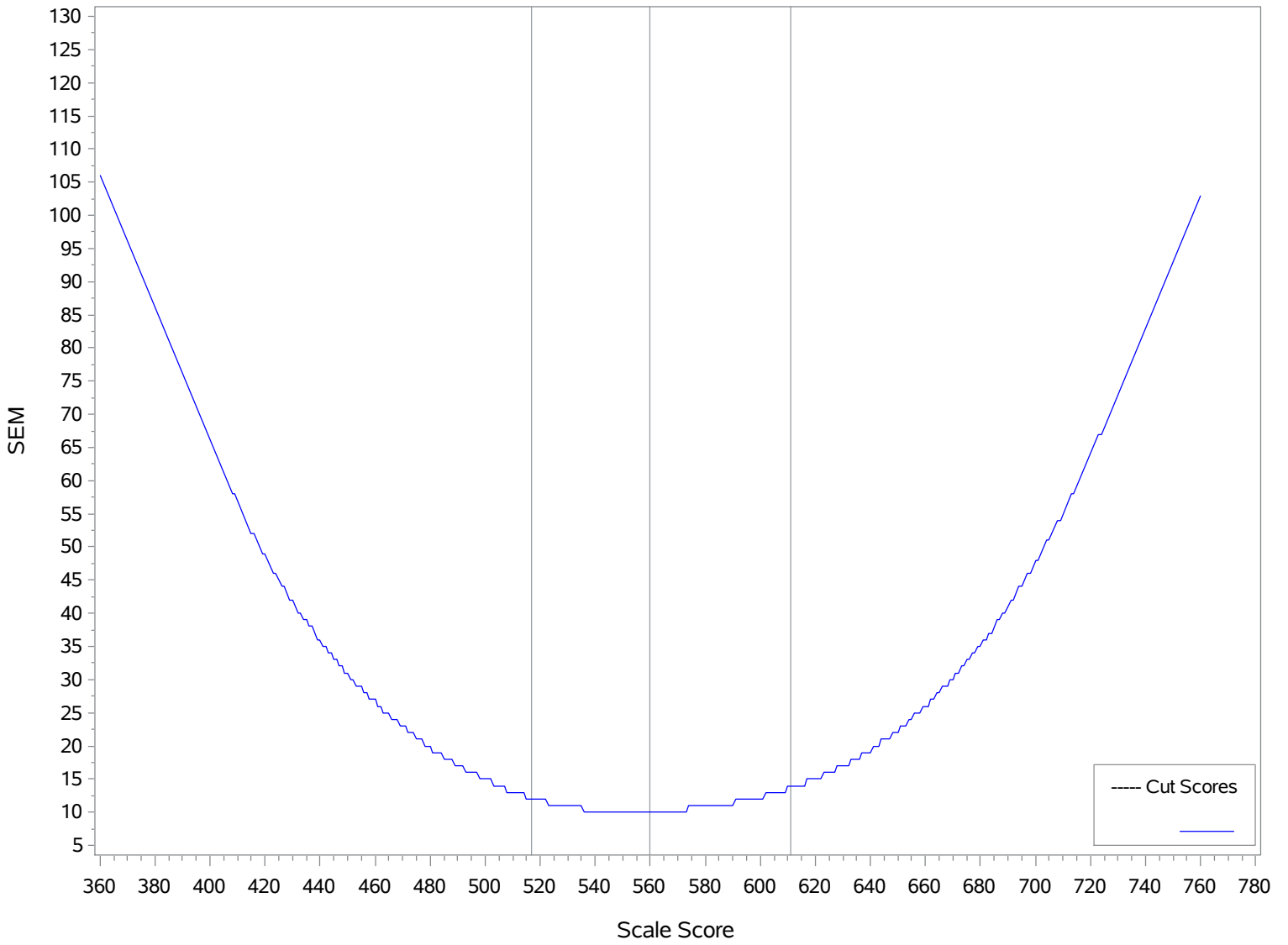
Figure H-8 CSEM with cut scores, Mathematics Grade 4
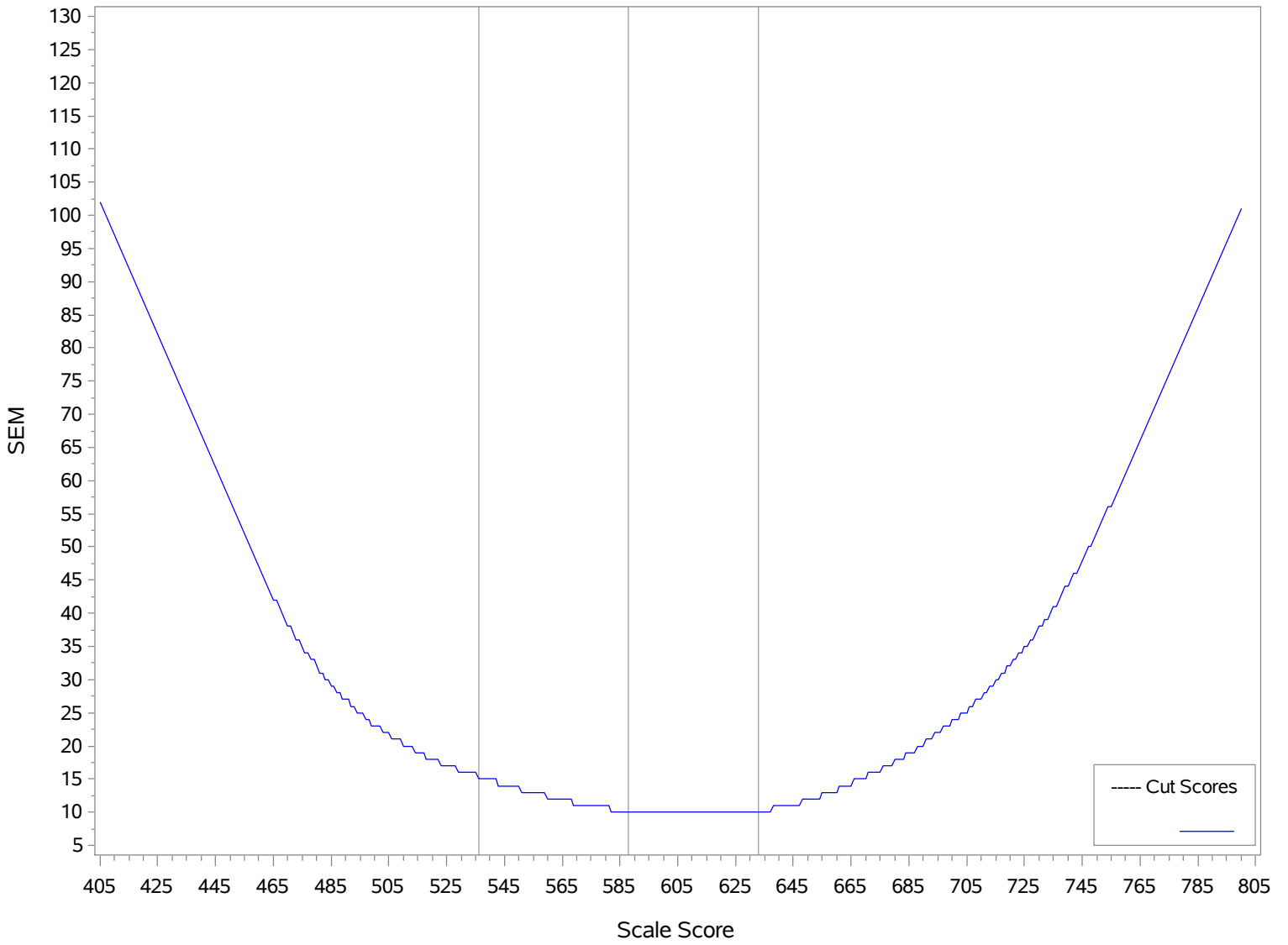
Figure H-9 CSEM with cut scores, Mathematics Grade 5
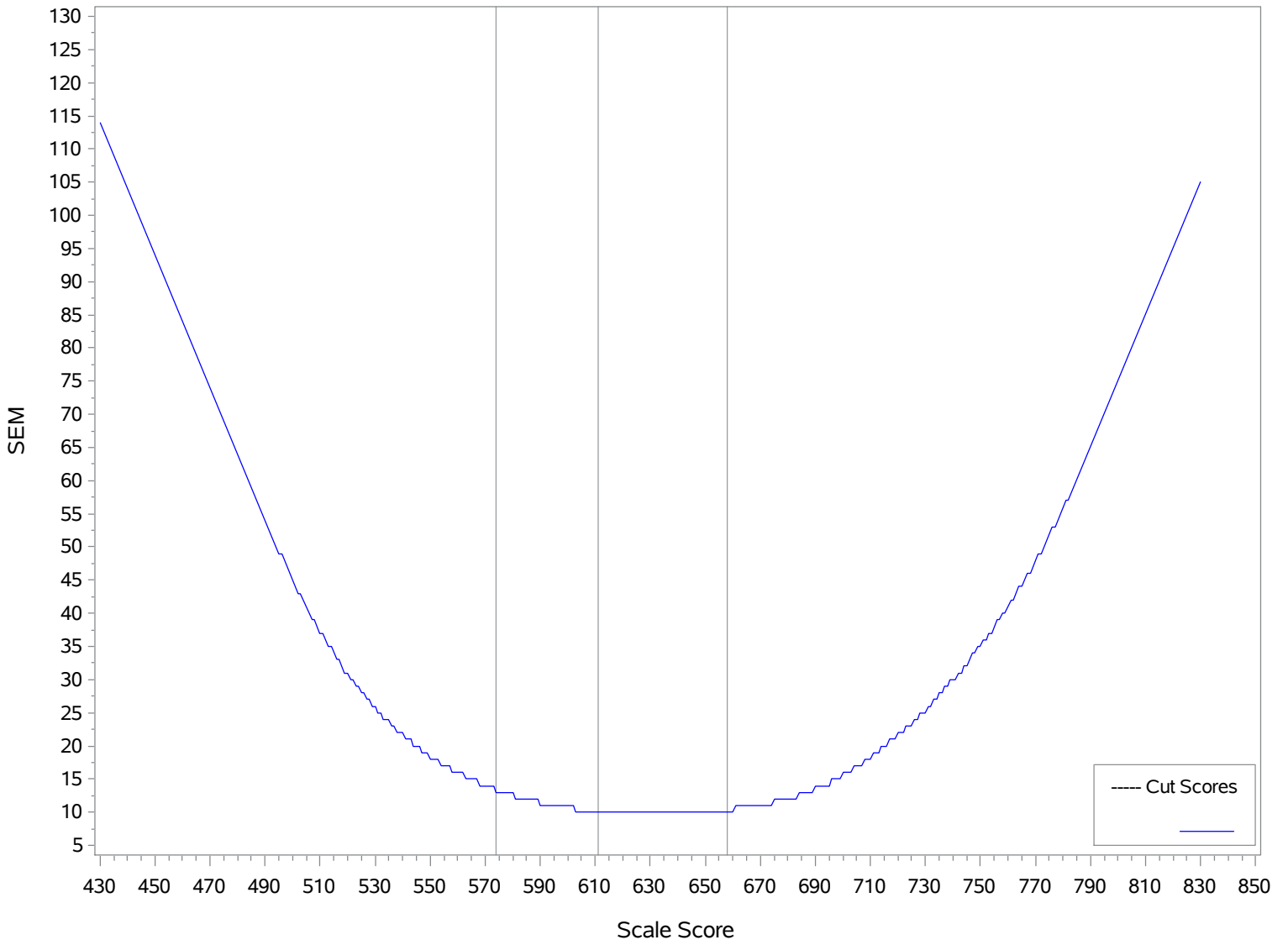
Figure H-10 CSEM with cut scores, Mathematics Grade 6
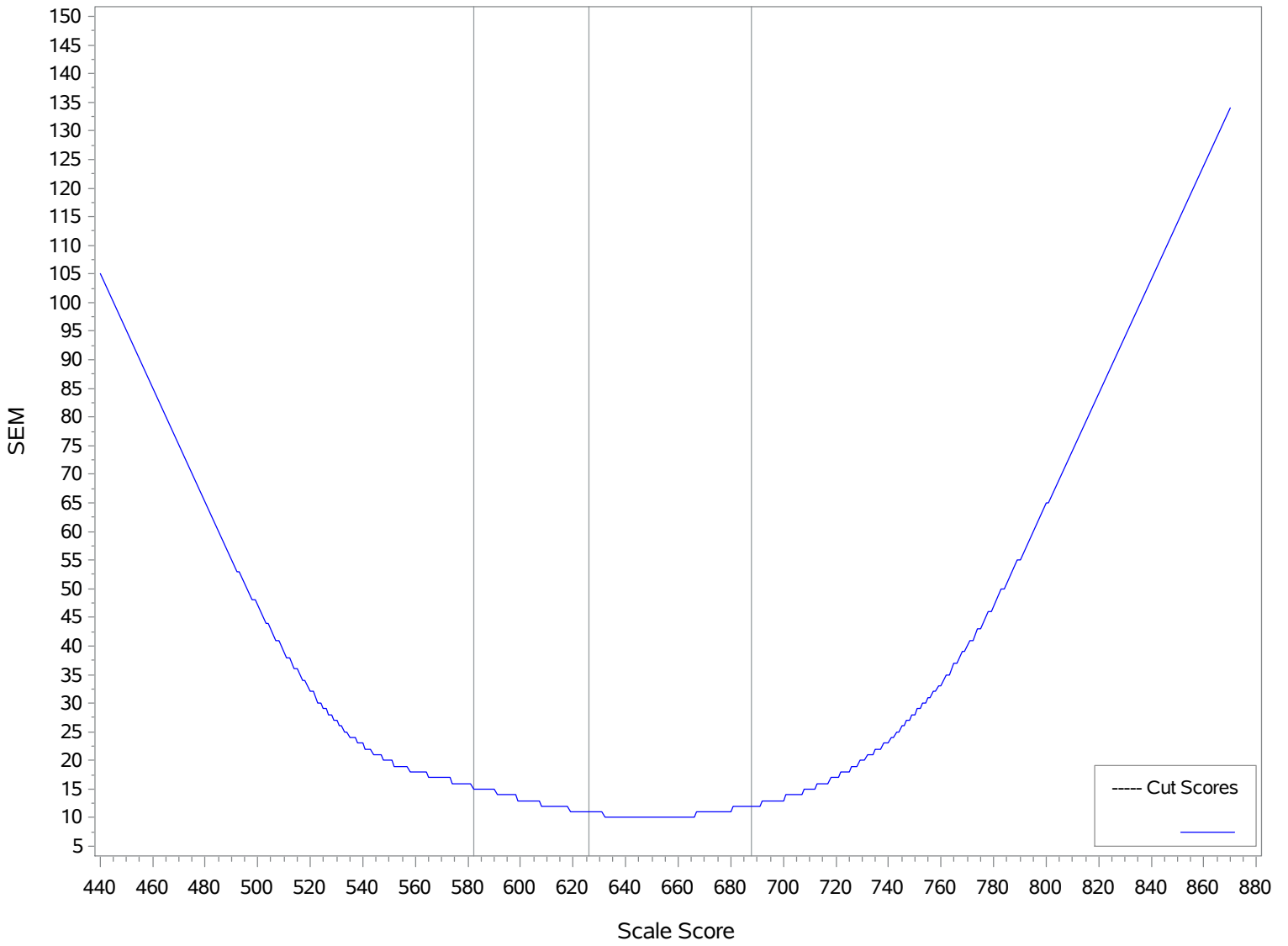
Figure H-11 CSEM with cut scores, Mathematics Grade 7

Figure H-12 CSEM with cut scores, Mathematics Grade 8

Figure H-13 CSEM with cut scores, Science Grade 4

Figure H-14 CSEM with cut scores, Science Grade 8

Figure H-15 CSEM with cut scores, Social Studies Grade 4
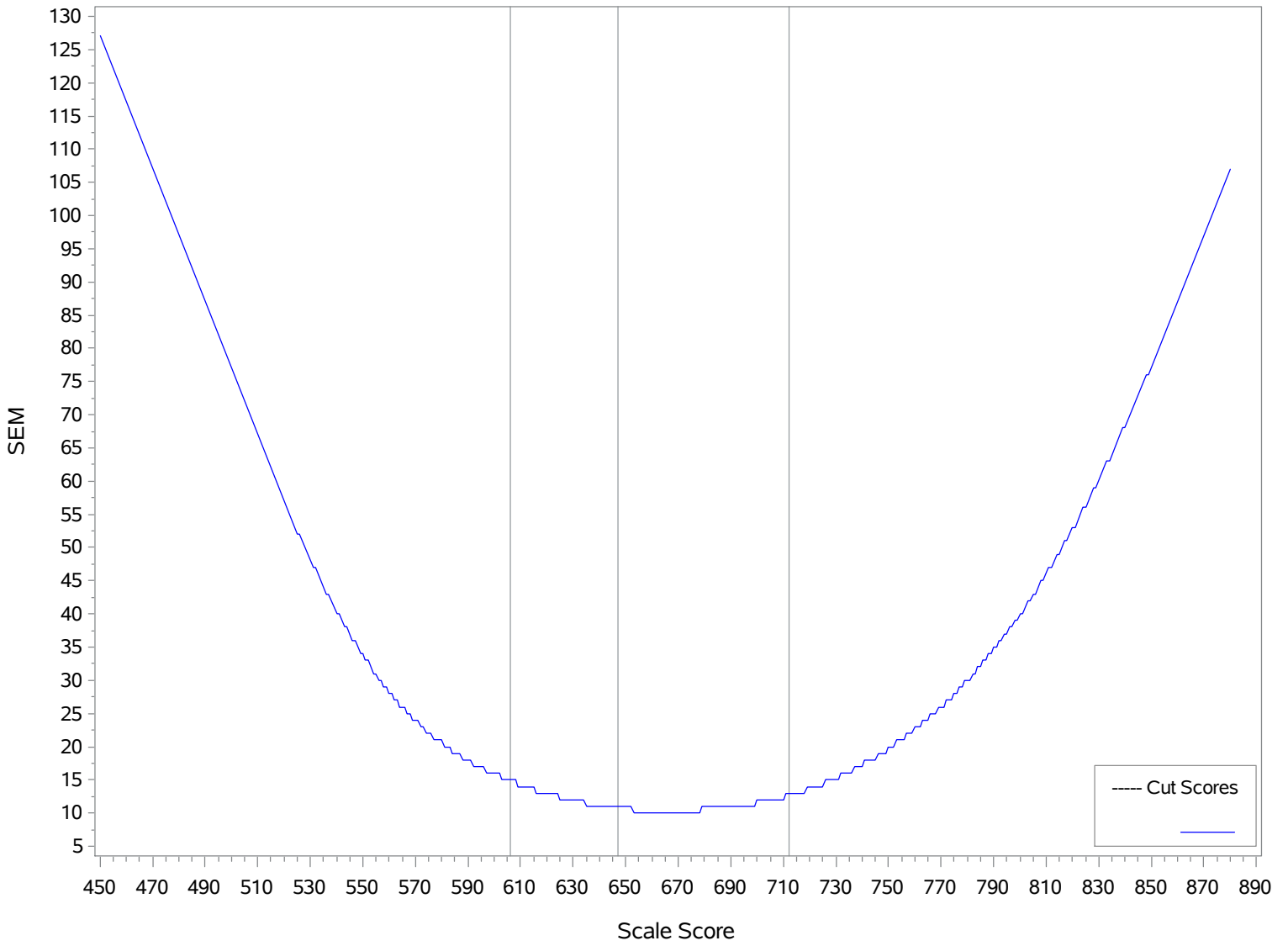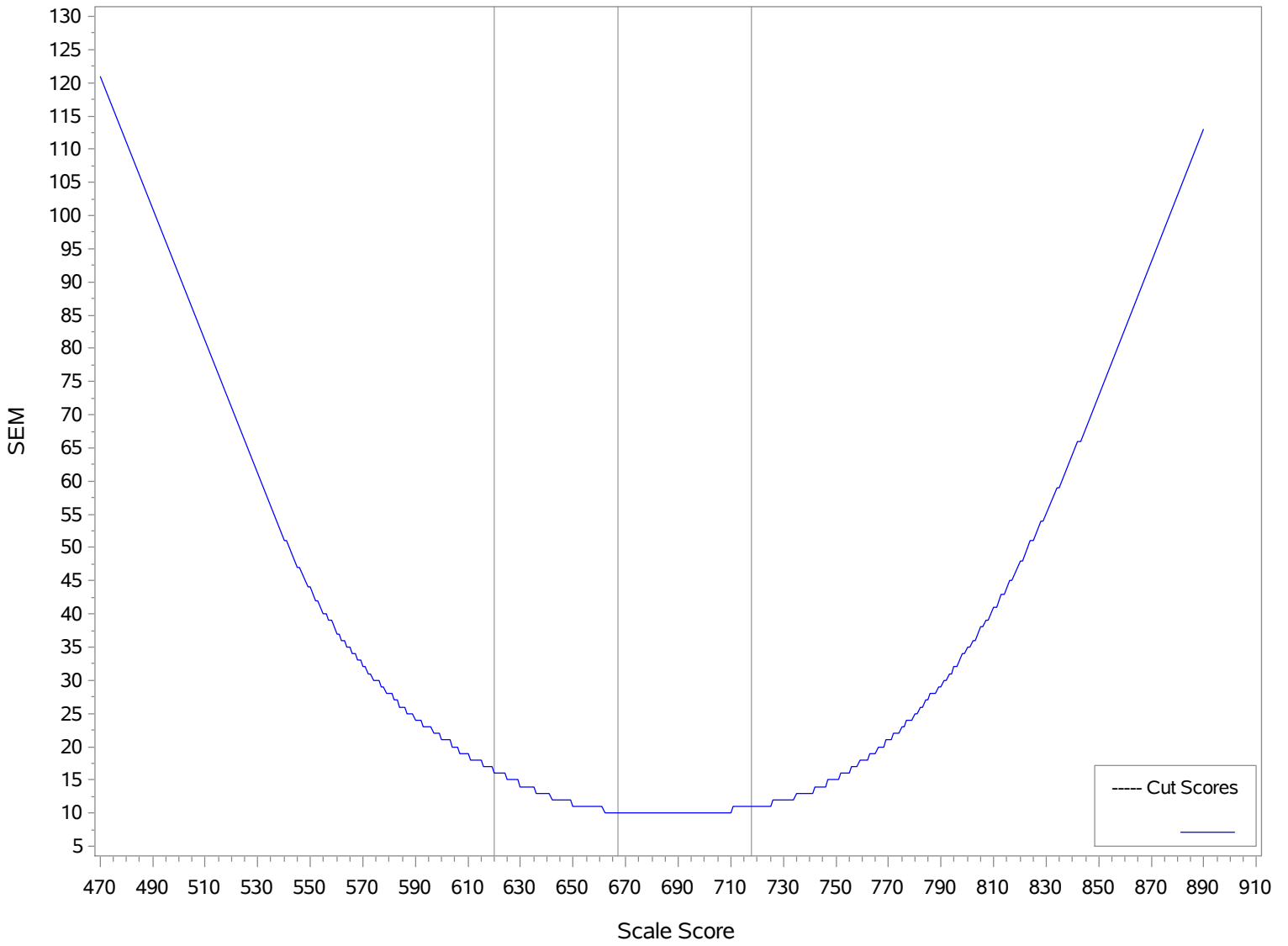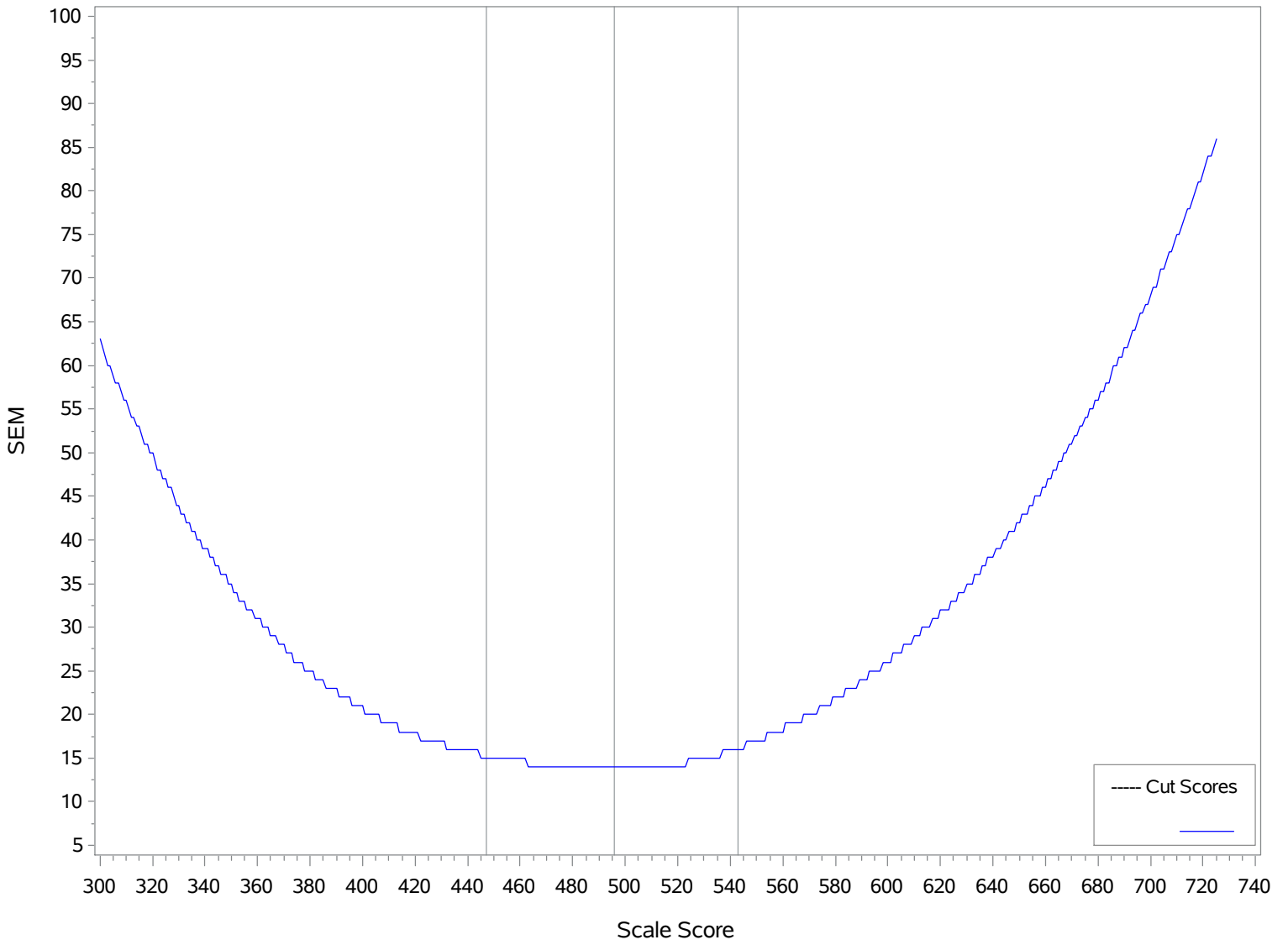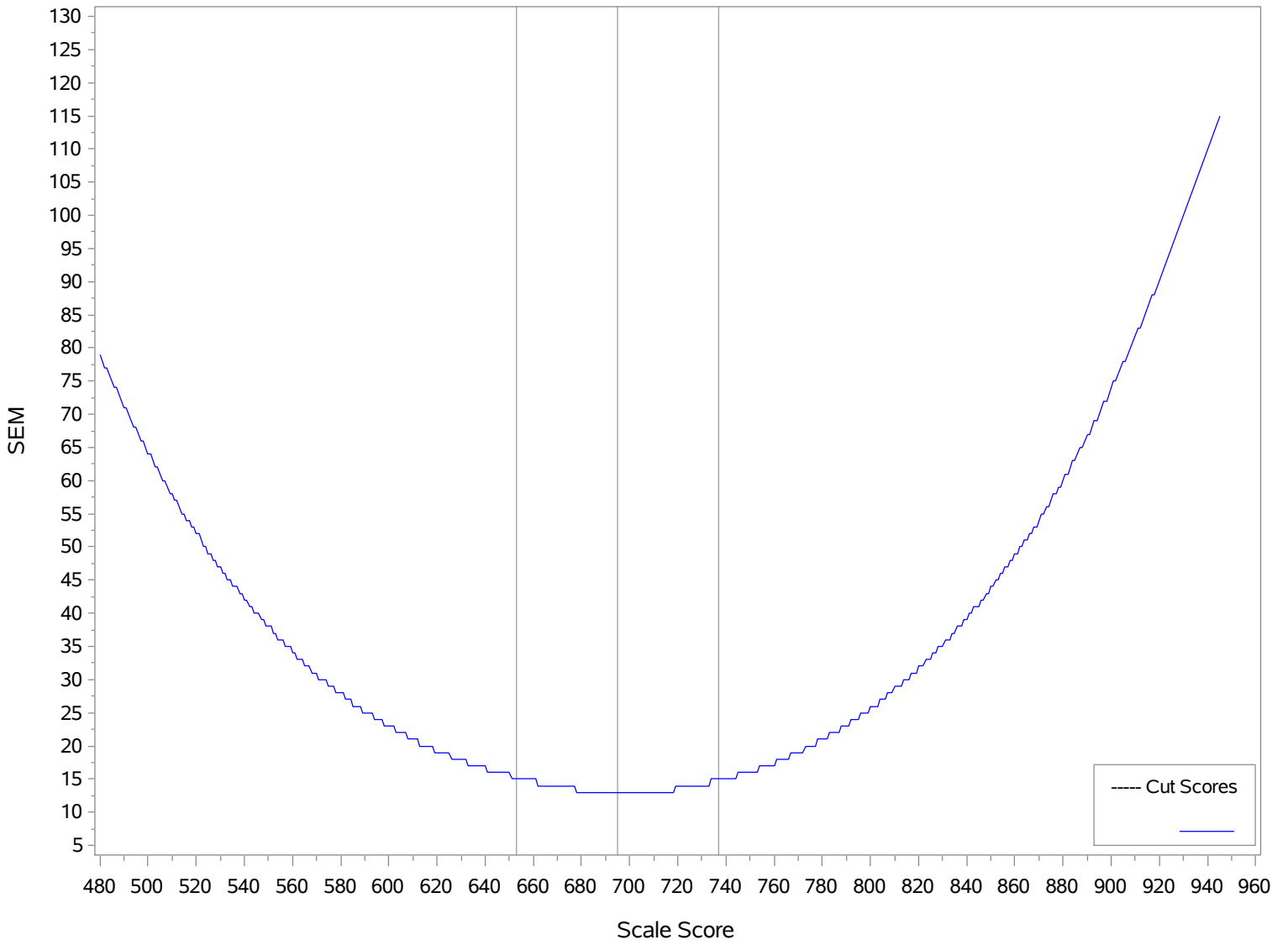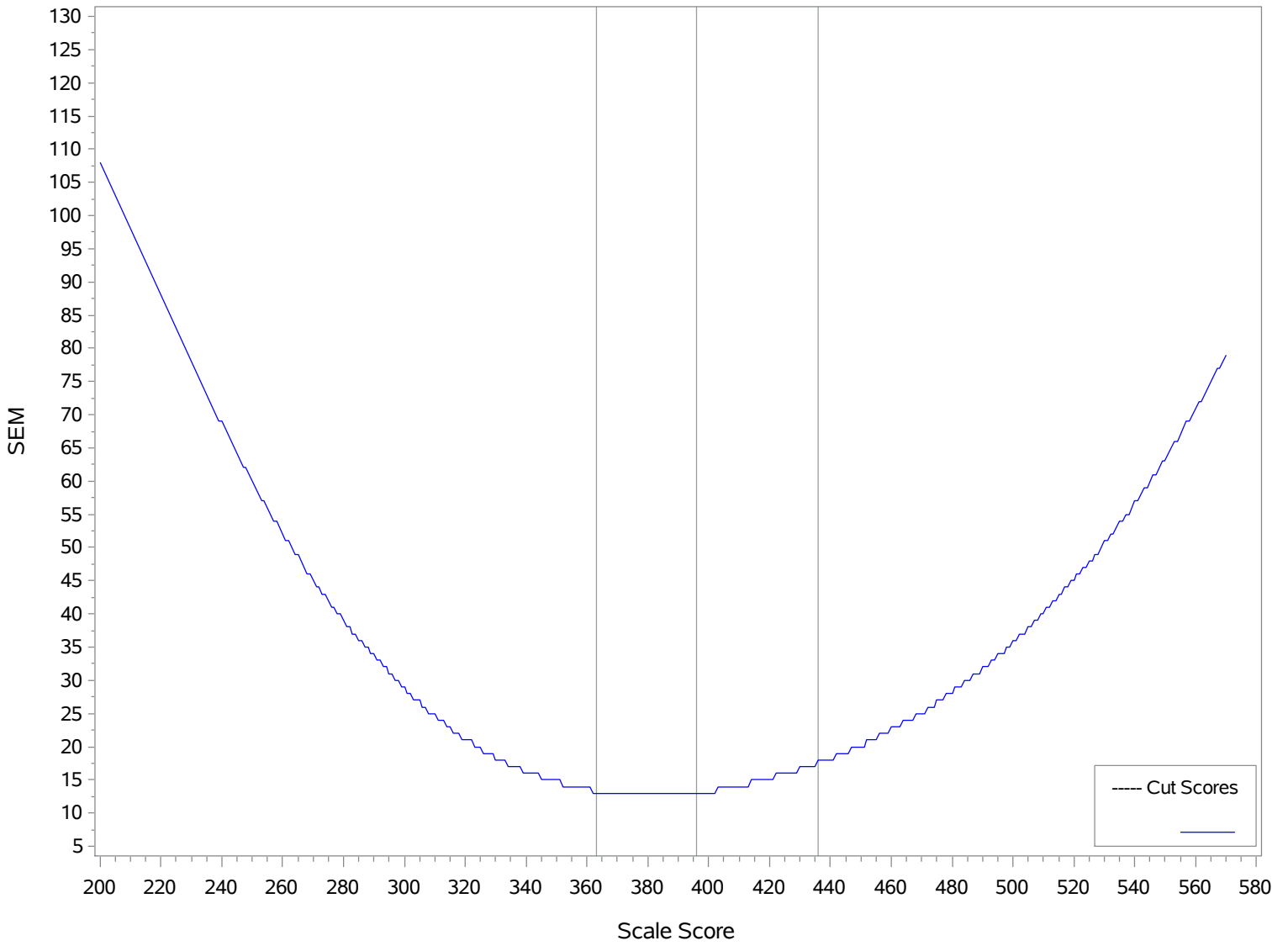
Figure H-16 CSEM with cut scores, Social Studies Grade 8
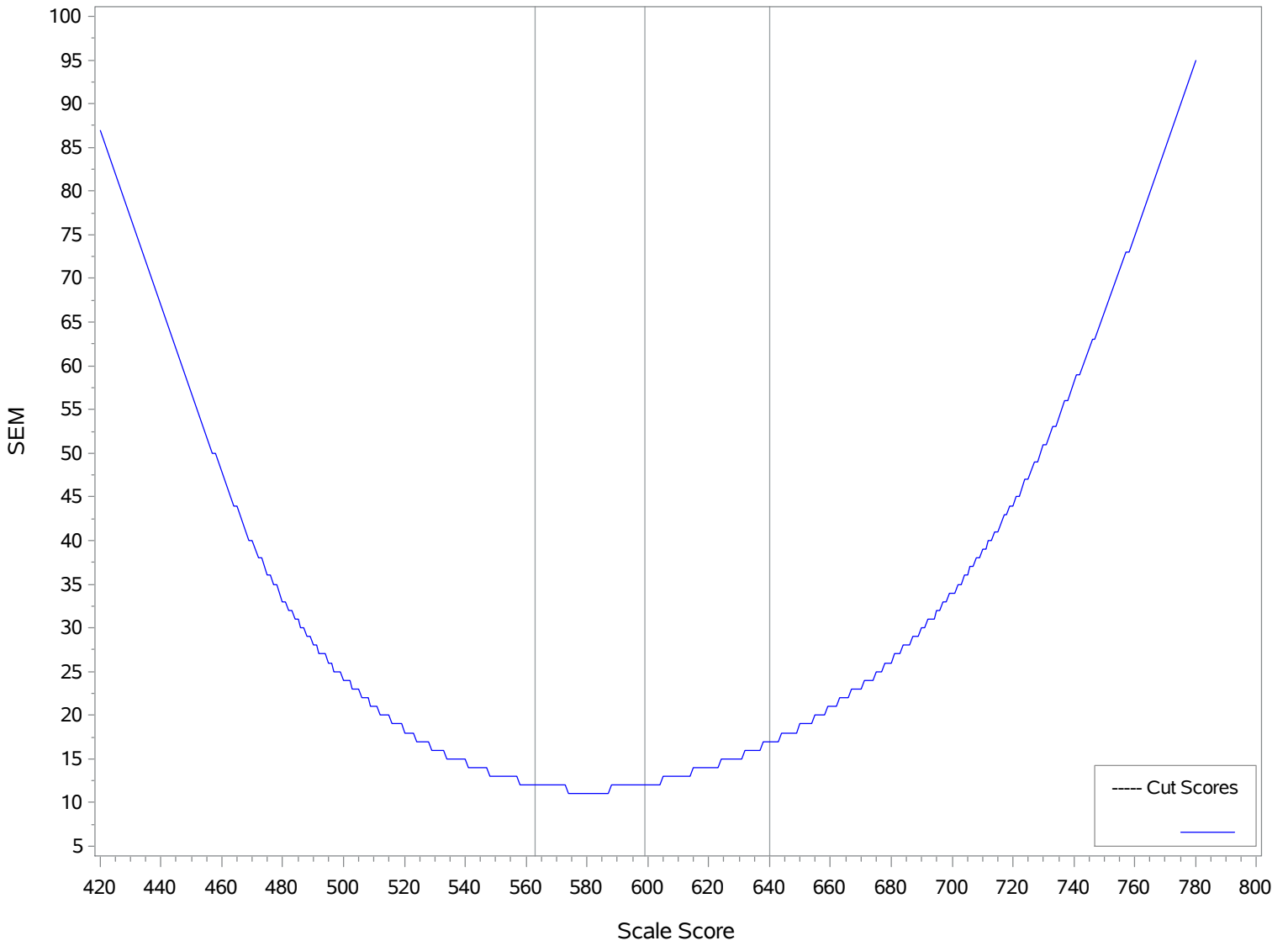
Figure H-17 CSEM with cut scores, Social Studies Grade 10
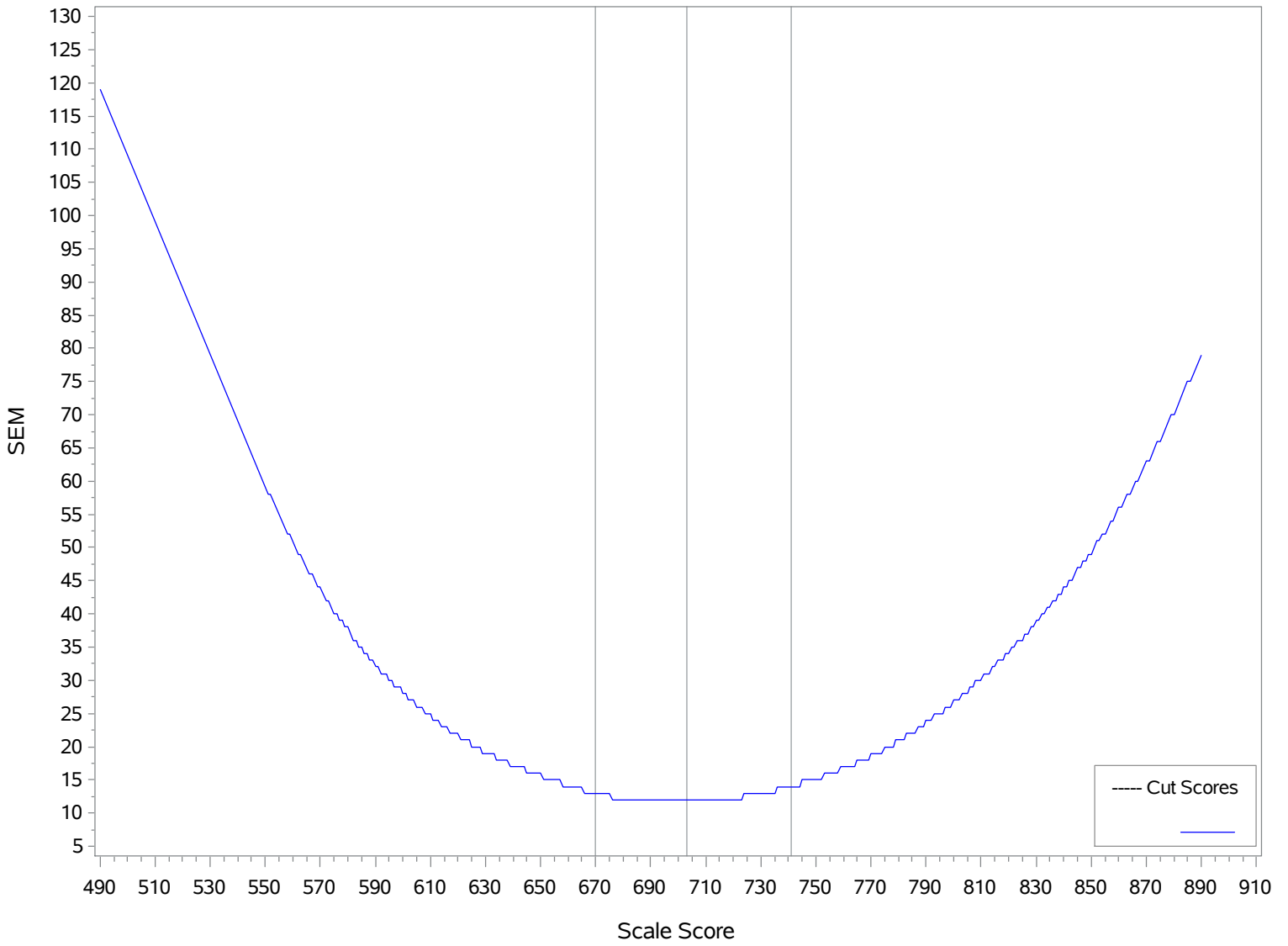
Appendix I

Classification Consistency and Accuracy Analysis by Subgroup

Table I-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	0.95	0.74
		Probability of Chance	0.62	0.53	0.87	0.30
		Kappa (k)	0.74	0.76	0.62	0.62
		Classification Accuracy	0.93	0.92	0.96	0.82
	Male	Classification Consistency (P)	0.89	0.89	0.96	0.74
		Probability of Chance	0.58	0.56	0.90	0.31
		Kappa (k)	0.74	0.75	0.59	0.63
		Classification Accuracy	0.92	0.93	0.97	0.82
Race/Ethnicity	White	Classification Consistency (P)	0.91	0.87	0.94	0.73
		Probability of Chance	0.68	0.51	0.86	0.31
		Kappa (k)	0.71	0.74	0.61	0.60
		Classification Accuracy	0.94	0.91	0.96	0.81
	African American	Classification Consistency (P)	0.86	0.95	0.99	0.81
		Probability of Chance	0.53	0.83	0.98	0.48
		Kappa (k)	0.71	0.72	0.58	0.63
		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	0.50	0.71	0.96	0.37
		Kappa (k)	0.71	0.73	0.60	0.62
		Classification Accuracy	0.90	0.95	0.99	0.83
	Asian	Classification Consistency (P)	0.88	0.91	0.96	0.75
		Probability of Chance	0.56	0.58	0.88	0.31
		Kappa (k)	0.73	0.78	0.64	0.64
		Classification Accuracy	0.92	0.94	0.97	0.83
	American Indian	Classification Consistency (P)	0.86	0.91	0.99	0.76
		Probability of Chance	0.50	0.72	0.97	0.38
		Kappa (k)	0.73	0.68	0.61	0.62
		Classification Accuracy	0.90	0.94	0.99	0.83
	Two or More	Classification Consistency (P)	0.89	0.89	0.96	0.74
		Probability of Chance	0.57	0.57	0.90	0.31
		Kappa (k)	0.75	0.75	0.58	0.63
		Classification Accuracy	0.92	0.93	0.97	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.94	0.99	0.77
		Probability of Chance	0.50	0.81	0.98	0.42
		Kappa (k)	0.68	0.68	0.58	0.60
		Classification Accuracy	0.89	0.96	0.99	0.84
Disability Status	Yes	Classification Consistency (P)	0.91	0.96	0.98	0.85
		Probability of Chance	0.52	0.72	0.91	0.43
		Kappa (k)	0.82	0.85	0.81	0.74
		Classification Accuracy	0.91	0.97	0.97	0.84

Table I-1 Indices for Classification Consistency and Accuracy, ELA Grade 3 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.91	0.98	0.76
		Probability of Chance	0.51	0.68	0.96	0.35
		Kappa (k)	0.73	0.73	0.55	0.63
		Classification Accuracy	0.91	0.94	0.99	0.83
Accommodation Use	Yes	Classification Consistency (P)	0.90	0.95	0.97	0.82
		Probability of Chance	0.51	0.71	0.94	0.42
		Kappa (k)	0.80	0.82	0.58	0.70
		Classification Accuracy	0.93	0.97	0.98	0.87

Table I-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	0.82	0.28
		Kappa (k)	0.73	0.77	0.62	0.61
		Classification Accuracy	0.93	0.92	0.95	0.80
	Male	Classification Consistency (P)	0.89	0.89	0.94	0.73
		Probability of Chance	0.58	0.53	0.86	0.29
		Kappa (k)	0.74	0.77	0.60	0.61
		Classification Accuracy	0.92	0.92	0.96	0.81
Race/Ethnicity	White	Classification Consistency (P)	0.91	0.88	0.92	0.71
		Probability of Chance	0.68	0.50	0.80	0.29
		Kappa (k)	0.71	0.76	0.61	0.59
		Classification Accuracy	0.93	0.91	0.95	0.80
	African American	Classification Consistency (P)	0.86	0.94	0.99	0.79
		Probability of Chance	0.52	0.78	0.97	0.45
		Kappa (k)	0.70	0.73	0.59	0.61
		Classification Accuracy	0.90	0.96	0.99	0.85
	Hispanic	Classification Consistency (P)	0.86	0.91	0.97	0.74
		Probability of Chance	0.51	0.64	0.93	0.34
		Kappa (k)	0.71	0.75	0.59	0.61
		Classification Accuracy	0.90	0.94	0.98	0.82
	Asian	Classification Consistency (P)	0.88	0.89	0.95	0.72
		Probability of Chance	0.56	0.55	0.87	0.29
		Kappa (k)	0.73	0.76	0.62	0.61
		Classification Accuracy	0.92	0.92	0.97	0.80
	American Indian	Classification Consistency (P)	0.85	0.92	0.98	0.75
		Probability of Chance	0.50	0.70	0.96	0.37
		Kappa (k)	0.69	0.75	0.46	0.60
		Classification Accuracy	0.89	0.94	0.98	0.82
	Two or More	Classification Consistency (P)	0.88	0.89	0.95	0.73
		Probability of Chance	0.57	0.55	0.87	0.30
		Kappa (k)	0.73	0.77	0.62	0.61
		Classification Accuracy	0.92	0.92	0.97	0.81
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.93	0.99	0.76
		Probability of Chance	0.50	0.77	0.99	0.41
		Kappa (k)	0.67	0.69	0.40	0.59
		Classification Accuracy	0.89	0.95	0.99	0.83
Disability Status	Yes	Classification Consistency (P)	0.85	0.94	0.98	0.77
		Probability of Chance	0.50	0.72	0.95	0.40
		Kappa (k)	0.71	0.77	0.59	0.62
		Classification Accuracy	0.90	0.95	0.99	0.84

Table I-2 Indices for Classification Consistency and Accuracy, ELA Grade 4 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.86	0.91	0.97	0.74
		Probability of Chance	0.51	0.64	0.94	0.34
		Kappa (k)	0.71	0.75	0.55	0.61
		Classification Accuracy	0.90	0.93	0.98	0.82
Accommodation Use	Yes	Classification Consistency (P)	0.88	0.93	0.97	0.78
		Probability of Chance	0.50	0.66	0.95	0.37
		Kappa (k)	0.75	0.81	0.43	0.65
		Classification Accuracy	0.91	0.96	0.98	0.85

Table I-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	0.86	0.30
		Kappa (k)	0.76	0.77	0.58	0.63
		Classification Accuracy	0.94	0.92	0.96	0.82
	Male	Classification Consistency (P)	0.90	0.90	0.95	0.75
		Probability of Chance	0.57	0.54	0.89	0.30
		Kappa (k)	0.77	0.77	0.56	0.65
		Classification Accuracy	0.93	0.93	0.97	0.82
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.88	0.94	0.73
		Probability of Chance	0.67	0.50	0.85	0.30
		Kappa (k)	0.74	0.76	0.57	0.62
		Classification Accuracy	0.94	0.92	0.96	0.81
	African American	Classification Consistency (P)	0.88	0.94	0.99	0.82
		Probability of Chance	0.52	0.78	0.98	0.45
		Kappa (k)	0.75	0.74	0.50	0.67
		Classification Accuracy	0.91	0.96	0.99	0.87
	Hispanic	Classification Consistency (P)	0.88	0.92	0.97	0.77
		Probability of Chance	0.51	0.65	0.95	0.34
		Kappa (k)	0.76	0.76	0.48	0.65
		Classification Accuracy	0.92	0.94	0.98	0.84
	Asian	Classification Consistency (P)	0.90	0.90	0.94	0.75
		Probability of Chance	0.58	0.54	0.87	0.29
		Kappa (k)	0.77	0.79	0.58	0.65
		Classification Accuracy	0.93	0.93	0.96	0.82
	American Indian	Classification Consistency (P)	0.86	0.92	0.98	0.77
		Probability of Chance	0.50	0.74	0.96	0.38
		Kappa (k)	0.73	0.71	0.58	0.63
		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	0.56	0.55	0.88	0.30
		Kappa (k)	0.76	0.79	0.58	0.65
		Classification Accuracy	0.92	0.93	0.97	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.95	1.00	0.80
		Probability of Chance	0.52	0.85	0.99	0.47
		Kappa (k)	0.71	0.63	0.38	0.63
		Classification Accuracy	0.90	0.96	1.00	0.86
Disability Status	Yes	Classification Consistency (P)	0.88	0.95	0.99	0.82
		Probability of Chance	0.52	0.78	0.97	0.45
		Kappa (k)	0.74	0.77	0.54	0.66
		Classification Accuracy	0.91	0.97	0.99	0.87

Table I-3 Indices for Classification Consistency and Accuracy, ELA Grade 5 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.88	0.91	0.98	0.77
		Probability of Chance	0.51	0.65	0.95	0.34
		Kappa (k)	0.76	0.75	0.53	0.65
		Classification Accuracy	0.91	0.94	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.92	0.89	0.96	0.78
		Probability of Chance	0.52	0.59	0.93	0.32
		Kappa (k)	0.84	0.74	0.39	0.67
		Classification Accuracy	0.95	0.93	0.97	0.85

Table I-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)	0.93	0.92	0.93	0.77
		Probability of Chance	0.62	0.51	0.69	0.26
		Kappa (k)	0.82	0.83	0.77	0.70
		Classification Accuracy	0.94	0.94	0.90	0.78
	Male	Classification Consistency (P)	0.90	0.88	0.94	0.72
		Probability of Chance	0.59	0.54	0.85	0.29
		Kappa (k)	0.76	0.74	0.60	0.61
		Classification Accuracy	0.92	0.92	0.96	0.80
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.86	0.92	0.70
		Probability of Chance	0.69	0.51	0.79	0.29
		Kappa (k)	0.74	0.73	0.59	0.58
		Classification Accuracy	0.94	0.91	0.94	0.78
	African American	Classification Consistency (P)	0.91	0.95	0.98	0.84
		Probability of Chance	0.52	0.75	0.93	0.44
		Kappa (k)	0.81	0.80	0.72	0.72
		Classification Accuracy	0.92	0.97	0.98	0.86
	Hispanic	Classification Consistency (P)	0.88	0.89	0.97	0.74
		Probability of Chance	0.53	0.64	0.93	0.33
		Kappa (k)	0.74	0.70	0.52	0.61
		Classification Accuracy	0.91	0.93	0.98	0.81
	Asian	Classification Consistency (P)	0.91	0.87	0.92	0.71
		Probability of Chance	0.66	0.52	0.81	0.29
		Kappa (k)	0.74	0.72	0.61	0.59
		Classification Accuracy	0.94	0.91	0.95	0.79
	American Indian	Classification Consistency (P)	0.88	0.90	0.97	0.75
		Probability of Chance	0.52	0.66	0.93	0.34
		Kappa (k)	0.75	0.69	0.59	0.61
		Classification Accuracy	0.91	0.93	0.98	0.82
	Two or More	Classification Consistency (P)	0.90	0.88	0.94	0.72
		Probability of Chance	0.58	0.55	0.85	0.29
		Kappa (k)	0.76	0.73	0.59	0.60
		Classification Accuracy	0.92	0.92	0.96	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.95	1.00	0.79
		Probability of Chance	0.52	0.88	0.99	0.48
		Kappa (k)	0.68	0.56	0.38	0.60
		Classification Accuracy	0.88	0.97	1.00	0.84
Disability Status	Yes	Classification Consistency (P)	0.87	0.95	0.99	0.80
		Probability of Chance	0.52	0.81	0.96	0.46
		Kappa (k)	0.73	0.72	0.62	0.64
		Classification Accuracy	0.90	0.96	0.99	0.85

Table I-4 Indices for Classification Consistency and Accuracy, ELA Grade 6 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.88	0.90	0.97	0.74
		Probability of Chance	0.52	0.65	0.92	0.33
		Kappa (k)	0.75	0.71	0.56	0.62
		Classification Accuracy	0.91	0.93	0.98	0.81
Accommodation Use	Yes	Classification Consistency (P)	0.87	0.93	0.95	0.75
		Probability of Chance	0.51	0.65	0.93	0.34
		Kappa (k)	0.74	0.79	0.25	0.62
		Classification Accuracy	0.91	0.95	0.97	0.83

Table I-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)	0.94	0.93	0.92	0.79
		Probability of Chance	0.62	0.50	0.68	0.26
		Kappa (k)	0.83	0.86	0.76	0.72
		Classification Accuracy	0.94	0.94	0.91	0.78
	Male	Classification Consistency (P)	0.90	0.88	0.94	0.72
		Probability of Chance	0.60	0.52	0.84	0.29
		Kappa (k)	0.75	0.75	0.62	0.60
		Classification Accuracy	0.93	0.91	0.96	0.80
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.87	0.92	0.70
		Probability of Chance	0.70	0.50	0.78	0.29
		Kappa (k)	0.72	0.73	0.63	0.58
		Classification Accuracy	0.94	0.91	0.95	0.79
	African American	Classification Consistency (P)	0.87	0.92	0.98	0.77
		Probability of Chance	0.50	0.71	0.96	0.39
		Kappa (k)	0.73	0.73	0.52	0.62
		Classification Accuracy	0.91	0.94	0.99	0.84
	Hispanic	Classification Consistency (P)	0.87	0.89	0.96	0.73
		Probability of Chance	0.54	0.59	0.91	0.31
		Kappa (k)	0.73	0.73	0.60	0.60
		Classification Accuracy	0.91	0.92	0.98	0.81
	Asian	Classification Consistency (P)	0.89	0.88	0.92	0.69
		Probability of Chance	0.67	0.50	0.76	0.28
		Kappa (k)	0.68	0.76	0.65	0.57
		Classification Accuracy	0.93	0.91	0.94	0.78
	American Indian	Classification Consistency (P)	0.86	0.90	0.97	0.73
		Probability of Chance	0.52	0.65	0.95	0.34
		Kappa (k)	0.71	0.70	0.43	0.59
		Classification Accuracy	0.91	0.93	0.98	0.81
	Two or More	Classification Consistency (P)	0.90	0.89	0.94	0.72
		Probability of Chance	0.61	0.52	0.83	0.28
		Kappa (k)	0.73	0.77	0.64	0.61
		Classification Accuracy	0.93	0.92	0.96	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.81	0.95	1.00	0.76
		Probability of Chance	0.51	0.85	0.99	0.46
		Kappa (k)	0.62	0.64	0.48	0.55
		Classification Accuracy	0.87	0.96	1.00	0.83
Disability Status	Yes	Classification Consistency (P)	0.86	0.94	0.99	0.79
		Probability of Chance	0.52	0.78	0.97	0.45
		Kappa (k)	0.70	0.74	0.58	0.61
		Classification Accuracy	0.90	0.96	0.99	0.85

Table I-5 Indices for Classification Consistency and Accuracy, ELA Grade 7 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.89	0.96	0.73
		Probability of Chance	0.53	0.60	0.92	0.32
		Kappa (k)	0.73	0.73	0.57	0.60
		Classification Accuracy	0.91	0.92	0.98	0.81
Accommodation Use	Yes	Classification Consistency (P)	0.85	0.92	0.96	0.74
		Probability of Chance	0.52	0.58	0.84	0.30
		Kappa (k)	0.68	0.82	0.78	0.62
		Classification Accuracy	0.88	0.95	0.97	0.81

Table I-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	0.87	0.92	0.71
		Probability of Chance	0.65	0.52	0.80	0.29
		Kappa (k)	0.75	0.74	0.62	0.60
		Classification Accuracy	0.94	0.91	0.94	0.80
	Male	Classification Consistency (P)	0.90	0.88	0.94	0.73
		Probability of Chance	0.58	0.56	0.86	0.30
		Kappa (k)	0.77	0.73	0.60	0.62
		Classification Accuracy	0.93	0.92	0.96	0.81
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.87	0.92	0.71
		Probability of Chance	0.67	0.51	0.80	0.29
		Kappa (k)	0.75	0.73	0.61	0.59
		Classification Accuracy	0.94	0.91	0.94	0.79
	African American	Classification Consistency (P)	0.88	0.93	0.98	0.79
		Probability of Chance	0.51	0.76	0.96	0.41
		Kappa (k)	0.76	0.71	0.55	0.65
		Classification Accuracy	0.91	0.95	0.99	0.85
	Hispanic	Classification Consistency (P)	0.88	0.90	0.96	0.75
		Probability of Chance	0.53	0.64	0.91	0.33
		Kappa (k)	0.76	0.72	0.58	0.62
		Classification Accuracy	0.92	0.93	0.97	0.82
	Asian	Classification Consistency (P)	0.91	0.89	0.93	0.74
		Probability of Chance	0.66	0.52	0.79	0.29
		Kappa (k)	0.74	0.77	0.67	0.63
		Classification Accuracy	0.94	0.92	0.95	0.81
	American Indian	Classification Consistency (P)	0.93	0.95	0.96	0.85
		Probability of Chance	0.50	0.67	0.85	0.37
		Kappa (k)	0.87	0.85	0.76	0.76
		Classification Accuracy	0.91	0.96	0.94	0.81
	Two or More	Classification Consistency (P)	0.90	0.90	0.95	0.75
		Probability of Chance	0.57	0.56	0.86	0.30
		Kappa (k)	0.77	0.76	0.64	0.64
		Classification Accuracy	0.93	0.92	0.97	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.95	1.00	0.80
		Probability of Chance	0.52	0.88	0.99	0.48
		Kappa (k)	0.69	0.57	0.29	0.61
		Classification Accuracy	0.90	0.97	1.00	0.86
Disability Status	Yes	Classification Consistency (P)	0.88	0.95	0.99	0.82
		Probability of Chance	0.54	0.84	0.97	0.49
		Kappa (k)	0.74	0.71	0.58	0.65
		Classification Accuracy	0.92	0.97	0.99	0.88

Table I-6 Indices for Classification Consistency and Accuracy, ELA Grade 8 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.88	0.90	0.97	0.75
		Probability of Chance	0.52	0.65	0.92	0.34
		Kappa (k)	0.76	0.72	0.58	0.63
		Classification Accuracy	0.92	0.93	0.98	0.83
Accommodation Use	Yes	Classification Consistency (P)	0.90	0.93	0.95	0.78
		Probability of Chance	0.50	0.65	0.89	0.35
		Kappa (k)	0.80	0.79	0.60	0.67
		Classification Accuracy	0.93	0.94	0.97	0.84

Table I-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)	0.93	0.91	0.94	0.78
		Probability of Chance	0.62	0.51	0.82	0.28
		Kappa (k)	0.82	0.82	0.69	0.70
		Classification Accuracy	0.95	0.93	0.96	0.84
	Male	Classification Consistency (P)	0.94	0.91	0.93	0.78
		Probability of Chance	0.65	0.50	0.77	0.27
		Kappa (k)	0.82	0.82	0.71	0.70
		Classification Accuracy	0.95	0.93	0.95	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.95	0.90	0.92	0.77
		Probability of Chance	0.74	0.50	0.75	0.29
		Kappa (k)	0.79	0.80	0.69	0.67
		Classification Accuracy	0.96	0.93	0.94	0.83
	African American	Classification Consistency (P)	0.90	0.96	0.99	0.85
		Probability of Chance	0.54	0.81	0.97	0.48
		Kappa (k)	0.79	0.79	0.62	0.72
		Classification Accuracy	0.93	0.97	0.99	0.89
	Hispanic	Classification Consistency (P)	0.90	0.93	0.97	0.81
		Probability of Chance	0.51	0.65	0.92	0.34
		Kappa (k)	0.80	0.80	0.67	0.71
		Classification Accuracy	0.93	0.95	0.98	0.86
	Asian	Classification Consistency (P)	0.92	0.92	0.95	0.79
		Probability of Chance	0.58	0.52	0.78	0.27
		Kappa (k)	0.81	0.83	0.77	0.71
		Classification Accuracy	0.94	0.94	0.96	0.84
	American Indian	Classification Consistency (P)	0.90	0.93	0.98	0.81
		Probability of Chance	0.50	0.70	0.95	0.37
		Kappa (k)	0.80	0.76	0.60	0.70
		Classification Accuracy	0.93	0.95	0.99	0.86
	Two or More	Classification Consistency (P)	0.92	0.92	0.95	0.79
		Probability of Chance	0.58	0.53	0.84	0.28
		Kappa (k)	0.82	0.82	0.70	0.71
		Classification Accuracy	0.95	0.94	0.96	0.85
Limited English Proficiency	Yes	Classification Consistency (P)	0.89	0.93	0.98	0.81
		Probability of Chance	0.50	0.71	0.95	0.37
		Kappa (k)	0.79	0.78	0.67	0.70
		Classification Accuracy	0.92	0.95	0.99	0.86
Disability Status	Yes	Classification Consistency (P)	0.92	0.94	0.97	0.83
		Probability of Chance	0.50	0.64	0.91	0.35
		Kappa (k)	0.83	0.83	0.69	0.73
		Classification Accuracy	0.94	0.95	0.98	0.87

Table I-7 Indices for Classification Consistency and Accuracy, Mathematics Grade 3 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.91	0.92	0.97	0.80
		Probability of Chance	0.52	0.62	0.91	0.32
		Kappa (k)	0.81	0.80	0.66	0.71
		Classification Accuracy	0.93	0.94	0.98	0.86
Accommodation Use	Yes	Classification Consistency (P)	0.92	0.95	1.00	0.86
		Probability of Chance	0.58	0.83	1.00	0.55
		Kappa (k)	0.80	0.67	0.17	0.70
		Classification Accuracy	0.94	0.96	1.00	0.90

Table I-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)	0.90	0.91	0.95	0.77
		Probability of Chance	0.64	0.53	0.83	0.29
		Kappa (k)	0.73	0.81	0.73	0.67
		Classification Accuracy	0.93	0.93	0.97	0.84
	Male	Classification Consistency (P)	0.91	0.91	0.94	0.77
		Probability of Chance	0.66	0.51	0.78	0.28
		Kappa (k)	0.75	0.82	0.74	0.68
		Classification Accuracy	0.94	0.93	0.96	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.90	0.94	0.76
		Probability of Chance	0.75	0.50	0.76	0.30
		Kappa (k)	0.70	0.80	0.73	0.66
		Classification Accuracy	0.95	0.93	0.96	0.83
	African American	Classification Consistency (P)	0.85	0.96	1.00	0.81
		Probability of Chance	0.53	0.85	0.98	0.48
		Kappa (k)	0.69	0.76	0.70	0.64
		Classification Accuracy	0.90	0.97	1.00	0.87
	Hispanic	Classification Consistency (P)	0.86	0.93	0.98	0.77
		Probability of Chance	0.52	0.67	0.93	0.34
		Kappa (k)	0.71	0.79	0.71	0.66
		Classification Accuracy	0.91	0.95	0.99	0.84
	Asian	Classification Consistency (P)	0.89	0.92	0.96	0.77
		Probability of Chance	0.60	0.53	0.78	0.28
		Kappa (k)	0.73	0.83	0.80	0.68
		Classification Accuracy	0.92	0.95	0.97	0.84
	American Indian	Classification Consistency (P)	0.85	0.94	0.99	0.77
		Probability of Chance	0.50	0.73	0.93	0.37
		Kappa (k)	0.70	0.76	0.80	0.64
		Classification Accuracy	0.89	0.95	0.99	0.83
	Two or More	Classification Consistency (P)	0.89	0.92	0.96	0.77
		Probability of Chance	0.59	0.56	0.86	0.30
		Kappa (k)	0.73	0.81	0.72	0.67
		Classification Accuracy	0.92	0.95	0.97	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.94	0.99	0.78
		Probability of Chance	0.50	0.76	0.97	0.39
		Kappa (k)	0.69	0.77	0.65	0.64
		Classification Accuracy	0.90	0.96	0.99	0.85
Disability Status	Yes	Classification Consistency (P)	0.87	0.95	0.98	0.80
		Probability of Chance	0.50	0.70	0.92	0.37
		Kappa (k)	0.74	0.82	0.73	0.68
		Classification Accuracy	0.91	0.96	0.99	0.86

Table I-8 Indices for Classification Consistency and Accuracy, Mathematics Grade 4 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.93	0.98	0.77
		Probability of Chance	0.53	0.65	0.93	0.34
		Kappa (k)	0.72	0.79	0.68	0.66
		Classification Accuracy	0.91	0.95	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.86	0.97	1.00	0.83
		Probability of Chance	0.57	0.91	0.99	0.54
		Kappa (k)	0.67	0.70	0.69	0.62
		Classification Accuracy	0.90	0.98	1.00	0.88

Table I-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)	0.89	0.91	0.95	0.75
		Probability of Chance	0.58	0.52	0.84	0.28
		Kappa (k)	0.73	0.81	0.69	0.65
		Classification Accuracy	0.92	0.93	0.97	0.82
	Male	Classification Consistency (P)	0.90	0.91	0.94	0.75
		Probability of Chance	0.59	0.51	0.79	0.27
		Kappa (k)	0.75	0.82	0.73	0.66
		Classification Accuracy	0.93	0.94	0.96	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.90	0.90	0.93	0.74
		Probability of Chance	0.67	0.50	0.77	0.28
		Kappa (k)	0.71	0.80	0.71	0.64
		Classification Accuracy	0.93	0.93	0.96	0.82
	African American	Classification Consistency (P)	0.89	0.97	0.99	0.85
		Probability of Chance	0.61	0.85	0.97	0.57
		Kappa (k)	0.72	0.80	0.78	0.66
		Classification Accuracy	0.91	0.98	0.99	0.88
	Hispanic	Classification Consistency (P)	0.86	0.93	0.98	0.77
		Probability of Chance	0.50	0.66	0.94	0.36
		Kappa (k)	0.72	0.79	0.65	0.64
		Classification Accuracy	0.90	0.95	0.99	0.84
	Asian	Classification Consistency (P)	0.90	0.90	0.96	0.75
		Probability of Chance	0.57	0.52	0.78	0.27
		Kappa (k)	0.75	0.79	0.81	0.66
		Classification Accuracy	0.93	0.93	0.97	0.83
	American Indian	Classification Consistency (P)	0.87	0.93	0.98	0.79
		Probability of Chance	0.50	0.72	0.96	0.40
		Kappa (k)	0.74	0.76	0.56	0.65
		Classification Accuracy	0.91	0.95	0.99	0.85
	Two or More	Classification Consistency (P)	0.89	0.91	0.96	0.77
		Probability of Chance	0.53	0.55	0.86	0.30
		Kappa (k)	0.77	0.80	0.74	0.67
		Classification Accuracy	0.92	0.94	0.97	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.95	0.99	0.79
		Probability of Chance	0.52	0.80	0.98	0.46
		Kappa (k)	0.67	0.74	0.64	0.60
		Classification Accuracy	0.89	0.96	1.00	0.85
Disability Status	Yes	Classification Consistency (P)	0.87	0.95	0.98	0.81
		Probability of Chance	0.52	0.73	0.94	0.44
		Kappa (k)	0.73	0.82	0.71	0.66
		Classification Accuracy	0.91	0.97	0.99	0.87

Table I-9 Indices for Classification Consistency and Accuracy, Mathematics Grade 5 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.86	0.91	0.97	0.75
		Probability of Chance	0.51	0.63	0.94	0.34
		Kappa (k)	0.72	0.77	0.57	0.63
		Classification Accuracy	0.89	0.94	0.98	0.81
Accommodation Use	Yes	Classification Consistency (P)	0.89	0.98	1.00	0.86
		Probability of Chance	0.67	0.91	0.99	0.65
		Kappa (k)	0.65	0.75	0.53	0.60
		Classification Accuracy	0.92	0.98	1.00	0.90

Table I-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	0.90	0.98	0.77
		Probability of Chance	0.56	0.54	0.91	0.31
		Kappa (k)	0.76	0.77	0.72	0.66
		Classification Accuracy	0.92	0.93	0.98	0.83
	Male	Classification Consistency (P)	0.90	0.91	0.97	0.77
		Probability of Chance	0.56	0.53	0.89	0.30
		Kappa (k)	0.77	0.81	0.70	0.68
		Classification Accuracy	0.93	0.94	0.98	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.91	0.89	0.96	0.76
		Probability of Chance	0.64	0.51	0.88	0.31
		Kappa (k)	0.75	0.78	0.67	0.66
		Classification Accuracy	0.93	0.92	0.97	0.82
	African American	Classification Consistency (P)	0.88	0.95	1.00	0.83
		Probability of Chance	0.57	0.86	0.99	0.54
		Kappa (k)	0.72	0.68	0.62	0.64
		Classification Accuracy	0.92	0.97	1.00	0.88
	Hispanic	Classification Consistency (P)	0.86	0.93	0.99	0.78
		Probability of Chance	0.50	0.71	0.97	0.39
		Kappa (k)	0.72	0.77	0.62	0.65
		Classification Accuracy	0.90	0.95	0.99	0.84
	Asian	Classification Consistency (P)	0.89	0.91	0.96	0.77
		Probability of Chance	0.57	0.52	0.83	0.28
		Kappa (k)	0.75	0.82	0.77	0.68
		Classification Accuracy	0.92	0.94	0.97	0.84
	American Indian	Classification Consistency (P)	0.86	0.95	0.99	0.80
		Probability of Chance	0.51	0.76	0.98	0.43
		Kappa (k)	0.71	0.78	0.54	0.64
		Classification Accuracy	0.90	0.96	0.99	0.85
	Two or More	Classification Consistency (P)	0.87	0.92	0.98	0.77
		Probability of Chance	0.51	0.60	0.92	0.33
		Kappa (k)	0.74	0.80	0.74	0.66
		Classification Accuracy	0.91	0.94	0.98	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.97	1.00	0.82
		Probability of Chance	0.58	0.90	1.00	0.56
		Kappa (k)	0.66	0.68	0.56	0.60
		Classification Accuracy	0.90	0.98	1.00	0.88
Disability Status	Yes	Classification Consistency (P)	0.89	0.96	0.99	0.84
		Probability of Chance	0.56	0.81	0.98	0.51
		Kappa (k)	0.74	0.78	0.72	0.67
		Classification Accuracy	0.92	0.97	1.00	0.88

Table I-10 Indices for Classification Consistency and Accuracy, Mathematics Grade 6 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.93	0.99	0.79
		Probability of Chance	0.50	0.69	0.97	0.37
		Kappa (k)	0.74	0.76	0.62	0.66
		Classification Accuracy	0.90	0.95	0.99	0.85
Accommodation Use	Yes	Classification Consistency (P)	0.93	0.99	1.00	0.92
		Probability of Chance	0.76	0.97	1.00	0.75
		Kappa (k)	0.71	0.78	0.57	0.68
		Classification Accuracy	0.95	0.99	1.00	0.94

Table I-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)	0.88	0.90	0.98	0.76
		Probability of Chance	0.54	0.56	0.93	0.31
		Kappa (k)	0.75	0.77	0.69	0.66
		Classification Accuracy	0.92	0.93	0.99	0.84
	Male	Classification Consistency (P)	0.90	0.90	0.97	0.78
		Probability of Chance	0.54	0.53	0.91	0.31
		Kappa (k)	0.79	0.79	0.69	0.68
		Classification Accuracy	0.93	0.94	0.98	0.85
Race/Ethnicity	White	Classification Consistency (P)	0.90	0.88	0.97	0.75
		Probability of Chance	0.61	0.51	0.91	0.31
		Kappa (k)	0.75	0.75	0.67	0.65
		Classification Accuracy	0.93	0.92	0.98	0.83
	African American	Classification Consistency (P)	0.88	0.97	1.00	0.85
		Probability of Chance	0.61	0.87	0.99	0.58
		Kappa (k)	0.71	0.75	0.75	0.64
		Classification Accuracy	0.92	0.98	1.00	0.89
	Hispanic	Classification Consistency (P)	0.86	0.94	0.99	0.79
		Probability of Chance	0.50	0.71	0.98	0.39
		Kappa (k)	0.72	0.78	0.66	0.66
		Classification Accuracy	0.91	0.95	0.99	0.85
	Asian	Classification Consistency (P)	0.91	0.94	0.97	0.82
		Probability of Chance	0.52	0.53	0.79	0.29
		Kappa (k)	0.81	0.88	0.85	0.75
		Classification Accuracy	0.92	0.96	0.97	0.85
	American Indian	Classification Consistency (P)	0.87	0.94	1.00	0.80
		Probability of Chance	0.51	0.80	0.99	0.44
		Kappa (k)	0.72	0.71	0.63	0.65
		Classification Accuracy	0.90	0.96	1.00	0.86
	Two or More	Classification Consistency (P)	0.87	0.92	0.98	0.77
		Probability of Chance	0.51	0.60	0.94	0.33
		Kappa (k)	0.74	0.80	0.70	0.66
		Classification Accuracy	0.91	0.94	0.99	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.87	0.97	1.00	0.84
		Probability of Chance	0.62	0.90	1.00	0.60
		Kappa (k)	0.66	0.72	0.60	0.60
		Classification Accuracy	0.91	0.98	1.00	0.89
Disability Status	Yes	Classification Consistency (P)	0.89	0.96	1.00	0.86
		Probability of Chance	0.61	0.84	0.99	0.58
		Kappa (k)	0.73	0.78	0.73	0.66
		Classification Accuracy	0.93	0.98	1.00	0.90

Table I-11 Indices for Classification Consistency and Accuracy, Mathematics Grade 7

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.88	0.92	0.99	0.79
		Probability of Chance	0.50	0.70	0.98	0.39
		Kappa (k)	0.75	0.74	0.66	0.66
		Classification Accuracy	0.91	0.95	1.00	0.86
Accommodation Use	Yes	Classification Consistency (P)	0.91	0.99	1.00	0.89
		Probability of Chance	0.78	0.96	1.00	0.77
		Kappa (k)	0.59	0.66	0.17	0.54
		Classification Accuracy	0.94	0.99	1.00	0.93

Table I-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)	0.87	0.91	0.97	0.75
		Probability of Chance	0.57	0.58	0.89	0.31
		Kappa (k)	0.70	0.80	0.71	0.64
		Classification Accuracy	0.91	0.94	0.98	0.83
	Male	Classification Consistency (P)	0.88	0.92	0.96	0.76
		Probability of Chance	0.55	0.57	0.86	0.30
		Kappa (k)	0.73	0.82	0.74	0.67
		Classification Accuracy	0.91	0.95	0.98	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.88	0.91	0.96	0.75
		Probability of Chance	0.62	0.54	0.85	0.30
		Kappa (k)	0.69	0.80	0.72	0.64
		Classification Accuracy	0.91	0.94	0.97	0.83
	African American	Classification Consistency (P)	0.86	0.97	1.00	0.82
		Probability of Chance	0.57	0.88	0.98	0.54
		Kappa (k)	0.66	0.78	0.75	0.62
		Classification Accuracy	0.90	0.98	1.00	0.88
	Hispanic	Classification Consistency (P)	0.84	0.95	0.99	0.77
		Probability of Chance	0.50	0.75	0.95	0.39
		Kappa (k)	0.68	0.78	0.69	0.63
		Classification Accuracy	0.88	0.96	0.99	0.84
	Asian	Classification Consistency (P)	0.86	0.92	0.96	0.74
		Probability of Chance	0.59	0.54	0.77	0.28
		Kappa (k)	0.66	0.83	0.83	0.65
		Classification Accuracy	0.90	0.95	0.97	0.82
	American Indian	Classification Consistency (P)	0.84	0.96	0.99	0.79
		Probability of Chance	0.51	0.78	0.96	0.43
		Kappa (k)	0.67	0.80	0.78	0.63
		Classification Accuracy	0.88	0.97	0.99	0.85
	Two or More	Classification Consistency (P)	0.85	0.93	0.98	0.76
		Probability of Chance	0.51	0.64	0.89	0.33
		Kappa (k)	0.69	0.81	0.79	0.64
		Classification Accuracy	0.89	0.95	0.98	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.82	0.97	1.00	0.79
		Probability of Chance	0.55	0.92	0.99	0.53
		Kappa (k)	0.59	0.69	0.67	0.55
		Classification Accuracy	0.87	0.98	1.00	0.85
Disability Status	Yes	Classification Consistency (P)	0.85	0.98	1.00	0.82
		Probability of Chance	0.57	0.89	0.98	0.54
		Kappa (k)	0.64	0.78	0.79	0.60
		Classification Accuracy	0.89	0.98	1.00	0.87

Table I-12 Indices for Classification Consistency and Accuracy, Mathematics Grade 8 (cont.)

Group	Category	Indices	Cut 1	Cut 2	Cut 3	All Cuts
Economically Disadvantaged	Yes	Classification Consistency (P)	0.84	0.94	0.99	0.77
		Probability of Chance	0.50	0.75	0.96	0.39
		Kappa (k)	0.69	0.78	0.68	0.63
		Classification Accuracy	0.89	0.96	0.99	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.86	0.99	1.00	0.85
		Probability of Chance	0.68	0.96	1.00	0.67
		Kappa (k)	0.57	0.72	0.81	0.54
		Classification Accuracy	0.90	0.99	1.00	0.90

Table I-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	0.87	0.91	0.70
		Probability of Chance	0.72	0.50	0.71	0.27
		Kappa (k)	0.69	0.74	0.70	0.58
		Classification Accuracy	0.94	0.91	0.94	0.78
	Male	Classification Consistency (P)	0.92	0.87	0.90	0.69
		Probability of Chance	0.72	0.50	0.67	0.27
		Kappa (k)	0.70	0.75	0.70	0.58
		Classification Accuracy	0.94	0.91	0.93	0.78
Race/Ethnicity	White	Classification Consistency (P)	0.94	0.86	0.89	0.69
		Probability of Chance	0.82	0.52	0.63	0.29
		Kappa (k)	0.64	0.71	0.69	0.56
		Classification Accuracy	0.95	0.90	0.92	0.77
	African American	Classification Consistency (P)	0.83	0.92	0.98	0.73
		Probability of Chance	0.50	0.75	0.95	0.39
		Kappa (k)	0.66	0.68	0.66	0.56
		Classification Accuracy	0.87	0.94	0.99	0.81
	Hispanic	Classification Consistency (P)	0.86	0.88	0.95	0.70
		Probability of Chance	0.58	0.58	0.86	0.30
		Kappa (k)	0.67	0.71	0.68	0.56
		Classification Accuracy	0.90	0.91	0.97	0.78
	Asian	Classification Consistency (P)	0.89	0.89	0.93	0.71
		Probability of Chance	0.66	0.51	0.76	0.28
		Kappa (k)	0.69	0.77	0.72	0.60
		Classification Accuracy	0.92	0.91	0.95	0.79
	American Indian	Classification Consistency (P)	0.87	0.89	0.95	0.72
		Probability of Chance	0.57	0.59	0.86	0.31
		Kappa (k)	0.69	0.74	0.66	0.59
		Classification Accuracy	0.90	0.93	0.97	0.80
	Two or More	Classification Consistency (P)	0.90	0.88	0.92	0.70
		Probability of Chance	0.69	0.51	0.75	0.28
		Kappa (k)	0.69	0.75	0.69	0.59
		Classification Accuracy	0.93	0.91	0.94	0.78
Limited English Proficiency	Yes	Classification Consistency (P)	0.83	0.90	0.98	0.71
		Probability of Chance	0.52	0.69	0.95	0.36
		Kappa (k)	0.65	0.67	0.55	0.55
		Classification Accuracy	0.88	0.93	0.98	0.79
Disability Status	Yes	Classification Consistency (P)	0.86	0.91	0.96	0.73
		Probability of Chance	0.53	0.60	0.85	0.31
		Kappa (k)	0.70	0.77	0.71	0.61
		Classification Accuracy	0.90	0.93	0.97	0.80
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.88	0.95	0.70
		Probability of Chance	0.59	0.56	0.84	0.30
		Kappa (k)	0.69	0.73	0.67	0.58
		Classification Accuracy	0.91	0.92	0.96	0.79

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

Table I-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)	0.90	0.87	0.91	0.68
		Probability of Chance	0.70	0.50	0.67	0.26
		Kappa (k)	0.66	0.74	0.72	0.57
		Classification Accuracy	0.93	0.91	0.93	0.78
	Male	Classification Consistency (P)	0.91	0.89	0.91	0.71
		Probability of Chance	0.67	0.50	0.64	0.25
		Kappa (k)	0.72	0.77	0.74	0.61
		Classification Accuracy	0.94	0.92	0.93	0.80
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.87	0.90	0.69
		Probability of Chance	0.76	0.52	0.61	0.27
		Kappa (k)	0.66	0.74	0.73	0.58
		Classification Accuracy	0.94	0.91	0.92	0.78
	African American	Classification Consistency (P)	0.84	0.92	0.97	0.73
		Probability of Chance	0.50	0.71	0.92	0.38
		Kappa (k)	0.67	0.73	0.65	0.57
		Classification Accuracy	0.89	0.95	0.98	0.82
	Hispanic	Classification Consistency (P)	0.86	0.89	0.94	0.69
		Probability of Chance	0.56	0.55	0.81	0.28
		Kappa (k)	0.68	0.74	0.71	0.57
		Classification Accuracy	0.90	0.92	0.96	0.79
	Asian	Classification Consistency (P)	0.89	0.88	0.91	0.69
		Probability of Chance	0.72	0.50	0.64	0.26
		Kappa (k)	0.62	0.76	0.76	0.58
		Classification Accuracy	0.92	0.91	0.94	0.78
	American Indian	Classification Consistency (P)	0.84	0.89	0.95	0.69
		Probability of Chance	0.55	0.59	0.84	0.30
		Kappa (k)	0.65	0.73	0.71	0.56
		Classification Accuracy	0.89	0.92	0.97	0.78
	Two or More	Classification Consistency (P)	0.88	0.87	0.92	0.68
		Probability of Chance	0.65	0.50	0.71	0.26
		Kappa (k)	0.66	0.74	0.73	0.57
		Classification Accuracy	0.92	0.92	0.95	0.78
Limited English Proficiency	Yes	Classification Consistency (P)	0.81	0.92	0.99	0.71
		Probability of Chance	0.50	0.77	0.97	0.40
		Kappa (k)	0.61	0.64	0.58	0.52
		Classification Accuracy	0.87	0.95	0.99	0.80
Disability Status	Yes	Classification Consistency (P)	0.84	0.92	0.97	0.73
		Probability of Chance	0.50	0.68	0.88	0.36
		Kappa (k)	0.68	0.75	0.74	0.58
		Classification Accuracy	0.89	0.95	0.98	0.82
Economically Disadvantaged	Yes	Classification Consistency (P)	0.87	0.89	0.94	0.70
		Probability of Chance	0.57	0.55	0.80	0.28
		Kappa (k)	0.69	0.75	0.71	0.58
		Classification Accuracy	0.91	0.92	0.96	0.79

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

Table I-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)	0.91	0.87	0.89	0.68
		Probability of Chance	0.60	0.50	0.67	0.25
		Kappa (k)	0.77	0.74	0.67	0.57
		Classification Accuracy	0.94	0.91	0.92	0.77
	Male	Classification Consistency (P)	0.91	0.88	0.89	0.69
		Probability of Chance	0.59	0.50	0.65	0.25
		Kappa (k)	0.78	0.77	0.69	0.59
		Classification Accuracy	0.94	0.92	0.92	0.78
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.87	0.87	0.67
		Probability of Chance	0.68	0.51	0.61	0.26
		Kappa (k)	0.75	0.74	0.67	0.55
		Classification Accuracy	0.94	0.91	0.91	0.76
	African American	Classification Consistency (P)	0.89	0.92	0.97	0.79
		Probability of Chance	0.54	0.73	0.92	0.47
		Kappa (k)	0.76	0.72	0.62	0.60
		Classification Accuracy	0.92	0.95	0.98	0.85
	Hispanic	Classification Consistency (P)	0.88	0.88	0.94	0.71
		Probability of Chance	0.51	0.57	0.82	0.31
		Kappa (k)	0.76	0.73	0.65	0.58
		Classification Accuracy	0.92	0.92	0.96	0.79
	Asian	Classification Consistency (P)	0.89	0.88	0.91	0.69
		Probability of Chance	0.55	0.51	0.71	0.26
		Kappa (k)	0.75	0.75	0.69	0.58
		Classification Accuracy	0.92	0.92	0.94	0.78
	American Indian	Classification Consistency (P)	0.86	0.90	0.94	0.71
		Probability of Chance	0.50	0.60	0.84	0.33
		Kappa (k)	0.71	0.75	0.64	0.56
		Classification Accuracy	0.90	0.93	0.96	0.79
	Two or More	Classification Consistency (P)	0.90	0.88	0.90	0.69
		Probability of Chance	0.56	0.51	0.72	0.26
		Kappa (k)	0.78	0.75	0.65	0.58
		Classification Accuracy	0.93	0.92	0.93	0.78
Limited English Proficiency	Yes	Classification Consistency (P)	0.87	0.89	0.96	0.73
		Probability of Chance	0.50	0.67	0.91	0.38
		Kappa (k)	0.73	0.67	0.58	0.56
		Classification Accuracy	0.91	0.93	0.97	0.81
Disability Status	Yes	Classification Consistency (P)	0.89	0.91	0.95	0.75
		Probability of Chance	0.50	0.62	0.84	0.37
		Kappa (k)	0.77	0.76	0.66	0.60
		Classification Accuracy	0.92	0.94	0.96	0.82
Economically Disadvantaged	Yes	Classification Consistency (P)	0.88	0.88	0.93	0.71
		Probability of Chance	0.51	0.57	0.81	0.31
		Kappa (k)	0.76	0.73	0.65	0.58
		Classification Accuracy	0.92	0.92	0.95	0.79

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

Table I-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)	0.92	0.88	0.90	0.70
		Probability of Chance	0.65	0.50	0.70	0.26
		Kappa (k)	0.76	0.75	0.67	0.59
		Classification Accuracy	0.94	0.92	0.92	0.78
	Male	Classification Consistency (P)	0.92	0.89	0.91	0.71
		Probability of Chance	0.61	0.50	0.69	0.25
		Kappa (k)	0.78	0.78	0.70	0.62
		Classification Accuracy	0.94	0.92	0.93	0.79
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.88	0.89	0.69
		Probability of Chance	0.70	0.50	0.65	0.26
		Kappa (k)	0.75	0.75	0.68	0.59
		Classification Accuracy	0.95	0.92	0.91	0.78
	African American	Classification Consistency (P)	0.88	0.93	0.97	0.78
		Probability of Chance	0.51	0.71	0.91	0.41
		Kappa (k)	0.76	0.75	0.65	0.63
		Classification Accuracy	0.92	0.95	0.98	0.84
	Hispanic	Classification Consistency (P)	0.89	0.89	0.94	0.73
		Probability of Chance	0.53	0.57	0.83	0.30
		Kappa (k)	0.76	0.75	0.67	0.61
		Classification Accuracy	0.92	0.93	0.96	0.80
	Asian	Classification Consistency (P)	0.92	0.89	0.90	0.70
		Probability of Chance	0.65	0.50	0.66	0.25
		Kappa (k)	0.76	0.77	0.70	0.60
		Classification Accuracy	0.94	0.92	0.93	0.78
	American Indian	Classification Consistency (P)	0.87	0.90	0.95	0.72
		Probability of Chance	0.51	0.62	0.86	0.32
		Kappa (k)	0.73	0.73	0.64	0.59
		Classification Accuracy	0.91	0.93	0.97	0.81
	Two or More	Classification Consistency (P)	0.90	0.90	0.92	0.72
		Probability of Chance	0.59	0.51	0.73	0.26
		Kappa (k)	0.76	0.79	0.70	0.62
		Classification Accuracy	0.93	0.92	0.94	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.93	0.99	0.78
		Probability of Chance	0.52	0.80	0.97	0.45
		Kappa (k)	0.70	0.67	0.51	0.59
		Classification Accuracy	0.90	0.95	0.99	0.84
Disability Status	Yes	Classification Consistency (P)	0.89	0.94	0.97	0.80
		Probability of Chance	0.52	0.73	0.91	0.43
		Kappa (k)	0.76	0.77	0.68	0.64
		Classification Accuracy	0.92	0.96	0.98	0.85
Economically Disadvantaged	Yes	Classification Consistency (P)	0.89	0.89	0.94	0.73
		Probability of Chance	0.52	0.57	0.83	0.30
		Kappa (k)	0.77	0.75	0.66	0.61
		Classification Accuracy	0.92	0.93	0.96	0.80

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

Table I-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)	0.91	0.88	0.91	0.71
		Probability of Chance	0.62	0.50	0.67	0.25
		Kappa (k)	0.77	0.77	0.72	0.61
		Classification Accuracy	0.93	0.92	0.94	0.79
	Male	Classification Consistency (P)	0.88	0.86	0.92	0.66
		Probability of Chance	0.60	0.52	0.80	0.28
		Kappa (k)	0.69	0.70	0.59	0.53
		Classification Accuracy	0.90	0.88	0.89	0.68
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.89	0.90	0.71
		Probability of Chance	0.64	0.50	0.63	0.25
		Kappa (k)	0.77	0.77	0.73	0.61
		Classification Accuracy	0.94	0.92	0.93	0.79
	African American	Classification Consistency (P)	0.94	0.97	0.96	0.88
		Probability of Chance	0.53	0.64	0.79	0.43
		Kappa (k)	0.88	0.93	0.82	0.78
		Classification Accuracy	0.96	0.95	0.93	0.83
	Hispanic	Classification Consistency (P)	0.89	0.90	0.94	0.74
		Probability of Chance	0.51	0.58	0.81	0.31
		Kappa (k)	0.78	0.77	0.71	0.63
		Classification Accuracy	0.92	0.93	0.96	0.81
	Asian	Classification Consistency (P)	0.91	0.91	0.91	0.73
		Probability of Chance	0.62	0.50	0.63	0.25
		Kappa (k)	0.76	0.81	0.76	0.64
		Classification Accuracy	0.94	0.93	0.94	0.81
	American Indian	Classification Consistency (P)	0.87	0.89	0.95	0.72
		Probability of Chance	0.52	0.59	0.84	0.31
		Kappa (k)	0.74	0.74	0.70	0.60
		Classification Accuracy	0.90	0.92	0.97	0.80
	Two or More	Classification Consistency (P)	0.91	0.90	0.92	0.73
		Probability of Chance	0.56	0.51	0.70	0.26
		Kappa (k)	0.79	0.80	0.73	0.64
		Classification Accuracy	0.93	0.93	0.94	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.88	0.96	0.99	0.83
		Probability of Chance	0.61	0.89	0.98	0.59
		Kappa (k)	0.69	0.62	0.62	0.60
		Classification Accuracy	0.91	0.97	1.00	0.87
Disability Status	Yes	Classification Consistency (P)	0.90	0.95	0.98	0.83
		Probability of Chance	0.55	0.74	0.89	0.48
		Kappa (k)	0.78	0.80	0.78	0.66
		Classification Accuracy	0.92	0.96	0.98	0.87
Economically Disadvantaged	Yes	Classification Consistency (P)	0.89	0.90	0.94	0.74
		Probability of Chance	0.51	0.58	0.81	0.31
		Kappa (k)	0.78	0.77	0.71	0.63
		Classification Accuracy	0.91	0.93	0.96	0.81

Note: Classification consistency and accuracy not computed for students with accommodations due to N-count < 50.

Appendix J

Wisconsin Standard Performance Index Score Computation

Technical Details of Wisconsin Standard Performance Index Score Computation

Technical details of the Standard Performance Index (SPI) estimation procedure described in this Appendix 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 $\beta(r_j, s_j)$:

$$g(T_j) = \frac{(r_j + s_j - 1)!\, T_j^{\,r_j - 1}(1 - T_j)^{s_j - 1}}{(r_j - 1)!(s_j - 1)!} \qquad (1)$$

for $0 \le T_j \le 1$; $r_j, s_j > 0$. 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:

1

$$p(X_j = x_j \mid T_j) = Binomial\ (n_j, T_j = \sum_{i=1}^{k_j} T_i / n_j),$$

where

T_i is the expected value of the score for item i in strand j for a given θ.

Given these assumptions, the posterior distribution of T_j, given x_j, is

$$g(T_j \mid X_j = x_j) = \beta(p_j, q_j), \tag{2}$$

with

$$p_j = r_j + x_j \tag{3}$$

and

$$q_j = s_j + n_j - x_j. \tag{4}$$

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

$$P_i(\theta) = P(X_i = 1 \mid \theta) = c_i + \frac{1 - c_i}{1 + \exp\left[-1.7 A_i (\theta - B_i)\right]}, \tag{5}$$

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,

2

& Sykes, 1996). For a CR item with 1_i score levels, integer scores were assigned that ranged from 0 to $1_i - 1$:

$$P_{im}(\theta) = P(X_i = m - 1|\theta) = \frac{\exp(z_{im})}{\sum_{g=1}^{1_i} \exp(z_{ig})}, \qquad m = 1, \ldots 1_i \qquad (6)$$

where

$$z_{ig} = \alpha_i (m - 1)\theta - \sum_{h=0}^{m-1} \gamma_{ih}' \qquad (7)$$

and

$$\gamma_{i0} = 0.$$

Alpha (α_i) is the item discrimination, and gamma (γ_{ih}) is related to the difficulty of the item levels; the trace lines for adjacent score levels intersect at γ_{ih}/α_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 θ 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

$$T_j = \frac{1}{n_j}\left[\sum_{i=1}^{k_j} T_{ij}(\theta)\right]. \qquad (8)$$

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 $\sigma^2(\hat{T}_j|\theta)$. 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 $[\sigma^2(\hat{T}_j|\theta)]$ of this distribution can be expressed in terms of its parameters, r_j and s_j.

3

Expressing the mean and variance of the prior distribution in terms of the parameters of the beta distribution (Novick & Jackson, 1974, p. 113) produces

$$\mu(\hat{T}_j \mid \theta) = \frac{r_j}{r_j + s_j} \qquad (9)$$

and

$$\sigma^2(\hat{T}_j \mid \theta) = \frac{r_j s_j}{(r_j + s_j)^2 (r_j + s_j + 1)} \quad . \qquad (10)$$

Solving these equations for r_j and s_j produces

$$r_j = \mu(\hat{T}_j \mid \theta) n_j^* \qquad (11)$$

and

$$s_j = [1 - \mu(\hat{T}_j \mid \theta)] n_j^*, \qquad (12)$$

where

$$n_j^* = \frac{\mu(\hat{T}_j \mid \theta)\left[1 - \mu(\hat{T}_j \mid \theta)\right]}{\sigma^2(\hat{T}_j \mid \theta)} - 1. \qquad (13)$$

Using IRT, $\sigma^2(\hat{T}_j \mid \theta)$ can be expressed in terms of item parameters (Lord, 1983):

$$\mu(\hat{T}_j \mid \theta) \approx \frac{1}{n_j} \sum_{i=1}^{k_j} \hat{T}_{ij}(\theta). \qquad (14)$$

Because T_j is a monotonic transformation of θ (Lord, 1980, p.71),

$$\sigma^2(\hat{T}_j \mid \theta) = \sigma^2(\hat{T}_j \mid T_j) \approx I(T_j, \hat{T}_j)^{-1} \qquad (15)$$

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

$$I(T_j, \hat{T}_j) = \frac{I(\theta, \hat{T}_j)}{\left(\partial T_j / \partial \theta\right)^2}, \qquad (16)$$

and

4

$$I(\theta, \hat{T}_j) \approx I(\theta, \hat{\theta}) . \tag{17}$$

Thus,

$$\sigma^2(\hat{T}_j \mid \theta) \approx \frac{\left[\dfrac{1}{n_j} \displaystyle\sum_{i=1}^{k_j} \hat{T}_{ij}(\theta)\right]^2}{I(\theta, \hat{\theta})}$$

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:

$$p_j = \hat{T}_j n_j^* + x_j , \tag{18}$$

and

$$q_j = \left[1 - \hat{T}_j\right] n_j^* + n_j - x_j . \tag{19}$$

The SPI is

$$\tilde{T}_j = \frac{p_j}{p_j + q_j} \tag{20}$$

$$= \frac{\hat{T}_j n_j^* + x_j}{n_j^* + n_j} . \tag{21}$$

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

$$\tilde{T}_j = w_j \hat{T}_j + (1 - w_j)\left[x_j / n_j\right]. \tag{22}$$

w_j, a function of the mean and variance of the prior distribution, is the relative weight given to the prior estimate:

$$w_j = \frac{n_j^*}{n_j^* + n_j} . \tag{23}$$

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

5

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:

$$Q = \sum_{j=1}^{J} n_j (\frac{x_j}{n_j} - \hat{T}_j)^2 / (\hat{T}_j (1 - \hat{T}_j)) .$$
(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, $\hat{T}_j$, 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 $\hat{T}_j$, 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 $\hat{T}_j$, and a three-parameter beta distribution, in which $\hat{T}_j$ 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,

6

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, $\hat{T}_j$, 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.