WISCONSIN FORWARD EXAM

SPRING 2016 TECHNICAL REPORT

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 *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 2

PART 3

PART 4

PART 5

PART 6

PART 7

PART 8

PART 9

PART 10

TABLE OF FIGURES

PART 2

PART 6

Part 1: Overview

The *Wisconsin Forward Exam Spring 2016 Technical Report* documents the processes and procedures applied in the Spring 2016 to the test development, administration, and scoring, as well as the assessment results. This report also provides evidence in support of validity and reliability of the testing program in adherence to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). This report demonstrates that the Spring 2016 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 (WKCE) norm-referenced assessments in the 1997 school year. The assessments used at that time were *TerraNova*™ tests developed by CTB/McGraw-Hill (1997, 2000, 2009). 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 school year 2007–08, 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. At that time Wisconsin students in grades 4, 8, and 10 continued to be assessed in Language Arts, Science, and Social Studies as required by state law.

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 adopted by the state. 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, Language Arts, and Mathematics. The assessments in Science and Social Studies under the existing WKCE model continued to be administered until Fall 2014.

A major change of the Wisconsin assessments occurred for the 2014–15 test administration. First, the English Language Arts (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 through 8 developed for the Spring 2015 administration consisted of new Smarter Balanced Assessment Consortium (SBAC) items aligned to the Common Core State Standards (CCSS). As a result, 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 year assessments included CTB's *TerraNova* items yielding 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, technology-enhanced 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) assessments can be found at http://dpi.wi.gov/sites/default/files/imce/assessment/pdf/TOMS%20Reporting%20and%20Interpretative%20Guide.pdf.

The ELA and Mathematics assessments have undergone yet another change in the 2015–16 administration year. The Wisconsin DPI partnered with Data Recognition Corporation (DRC) to develop new ELA and Mathematics grades 3 through 8 assessments for the Spring 2016 administration. The items contained in these assessments were drawn from DRC's nationally field-tested College- and Career-Ready 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 Badger Exam tests, they 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 the Summer of 2016.

Science (grades 4 and 8) and Social Studies (grades 4, 8, and 10) assessments have been on a different trajectory, and 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 Science and Social Studies tests were mainly drawn from the pool of previously administered items and also included some new items. 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, the statistical linking of 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 the Summer of 2016.

In summary, the ELA, Mathematics, Science, and Social Studies assessments administered in Spring 2016 were developed based on the input of Wisconsin educators and with adherence to the Wisconsin's standards. The new test scales are considered to be the new baseline for the year-to-year student performance comparison and tracking. This Technical Report documents all aspects of the 2015–16 testing cycle. The structure of this Technical 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. (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. In this section, we examine some possible uses of the Wisconsin Forward Exam scores.

The following parts (Parts 2 through 10) of this Technical Report provide additional evidence for these uses as well as technical support for some of the interpretations and uses of test scores. The information in Parts 2 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, 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 (objective)-level 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 that 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 all content areas were recommended by Wisconsin educators at the standard setting workshop in June 2016. 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.

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 10 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 their 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 comprise 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: Test Design and Item Development

Part 2 of this report describes the test design, the item development and selection process, and some aspects of the content-related validity of the Wisconsin Forward Exam. More specifically, it describes how DRC, DPI, and Wisconsin educators 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 for ELA and Mathematics were the basis of the test blueprints and item specifications for their respective content areas. For Science and Social Studies, the Wisconsin Model Academic Standards formed the basis for test blueprints and item specifications. Wisconsin educators were involved in reviewing the items in all contents to ensure the appropriateness of the test to the standards. Item review occurred in December 2015 with the convention of approximately 74 educators for grades 3–8 ELA and Mathematics, grades 4 and 8 Science, and grades 4, 8, and 10 Social Studies. This item review 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.

Part 3: Test Form Development

Part 3 discusses key development tasks related to creating the Spring 2016 Wisconsin Forward Exam forms. The Spring 2016 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.

Item selection was based upon the approved test blueprints. DRC's College- and Career-Ready (CCR) item bank contained a sufficient number of items to fulfill the test design needs for the ELA and Mathematics exams. Science and Social Studies forms were supplemented through the use of *TerraNova* items (CTB/McGraw-Hill, 2009). Part 3 also discusses the process of selecting operational test items and the process of obtaining DPI approvals. As detailed in Part 3, there were numerous unique items on each form. In addition to the unique items, the ELA and Mathematics forms contained vertical linking items aligned to the grade above and the grade below. The purpose of this test design was to develop a vertical scale for comparing students' progress from year to year. Selection of the Spring 2016 test forms was done using the approved test blueprints and test designs as guides.

Part 4: Test Administration

Part 4 briefly describes test administration and accommodations. The Wisconsin Forward Exam is a component of the Wisconsin Student Assessment System (WSAS), considered to be a comprehensive statewide program of assessments. In the 2015–16 school year, this assessment replaced the Badger Exam (SBAC) in the areas of ELA and Mathematics in grades 3–8 and

replaced the WKCE in the areas of Science (grades 4 and 8) and Social Studies (grades 4, 8, and 10).

Test administration was conducted over an eight-week window: March 28–May 20, 2016. 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 items, auto-scoring of technology-enhanced items, and artificial intelligence (AI) scoring and handscoring of text dependent analysis (TDA) items. The description of the handscoring process includes the development and review of the scoring rubrics, anchor (sample) paper selection, training of scoring personnel, ongoing quality assurance, and a systematic review of the resulting score distributions supporting reliable and valid reported test scores. The scoring rubric used in handscoring of the TDA writing items is presented in detail.

Part 6: Calibration, Scaling, and Deriving Scale Scores

The Spring 2016 administration year is the new baseline for the Wisconsin Forward Exam in all grades and content areas. Part 6 discusses characteristics of the sample of student data used for data analysis and describes the calibration, scaling, and scoring methods implemented for the Wisconsin Forward Exam after the Spring 2016 test administration. The data were calibrated and scaled using two different item response theory (IRT) models, one for constructed-response items and one for multiple-choice items, which are the item types used for most large-scale standardized testing programs in education. Evaluation of the sufficiency of the IRT model results include model-to-data fit and the standard error of measurement. Item-pattern scoring was applied to the Spring 2016 Wisconsin Forward Exam. As discussed in Part 6, item-pattern scoring is generally recommended over number-correct scoring because it produces more accurate scores for individual students. Part 6 also explains how a student's scale score is derived from the raw score using item-pattern scoring.

Part 7: Standard Setting

Part 7 provides a brief overview of the standard setting process during which the performance level cut scores were set for the Wisconsin Forward Exam. The standard setting methodology and results, including performance level descriptors and cut scores, are presented.

Part 8: Test Results

Part 8 summarizes results of item analyses as well as test reliability reported using Cronbach's alpha and standard error of measurement. Summary descriptive statistics for all scores (i.e., raw scores, scale scores, SPI scores, performance levels) are reported for all public

school students and for subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency.

Part 9: Reliability

Part 9 elaborates on the reliability of the test based on results presented in previous parts of the report. Standard error of measurement was assessed for raw scores and scale scores. Inter-rater reliability was computed for TDA items on ELA tests that were scored using the AI scoring engine with human scorer verification. Internal consistency was evaluated for all tests for the total student population and for subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Classification consistency and accuracy were estimated for performance classification.

Part 10: Validity

Part 10 reviews the validity evidence presented in all previous parts of the report and provides additional validity evidence supporting the Wisconsin Forward Exam. Factor analysis and correlations among content standards are presented in the context of construct validity. An analysis of differential item functioning is presented. Forensic analysis procedures, implemented to detect possible aberrant testing behavior, are also discussed.

Part 11: Summary Recommendations

Key findings of the Spring 2016 Wisconsin Forward Exam administration are presented in the body of the report. However, some items of a more technical nature, which 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 2016 Wisconsin Forward Exam administration cover three different phases of the testing cycle: item development; scoring; and psychometric, or measurement-based, research and evaluation.

Part 2: Test Design and Item 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 2 documents the test design and item development process for the Spring 2016 administration.

DRC's College- and Career-Ready (CCR) item bank contains nationally field-tested college- and career-ready items that support the next generation of standards and assessments. It is aligned to the college- and career-ready (CCR) standards in Mathematics and English Language Arts in grades 3–8, and is designed to support states like Wisconsin that have adopted more rigorous content standards, curricula, and assessments that better prepare students for college and careers.

Alignment to the CCR standards, grade-level appropriateness, depth of knowledge (DOK), item/task level of complexity, estimated difficulty level, relevancy of context, rationale for distractors, style, accuracy, and correct terminology were major considerations in the item development process. DRC's item development processes for the CCR item bank followed the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). DRC's item development work was and continues to be designed to produce reliable and instructionally valid tests that reflect the complete range of performance articulated in the AERA, APA, and NCME *Standards*.

This chapter is particularly relevant to AERA, APA, & NCME (2014) Standards 3.1, 3.2, and 4.0. Each of these Standards and the way each Standard is addressed will be presented in this chapter. 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. (85)

Furthermore, DRC's item development work for the CCR item bank 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 items were reviewed for content and for fairness not only by DRC's content experts but also by a panel of external experts. The external reviewers have a broad range of experience in the educational field. All of the reviewers have bachelor's-level, master's-level, or doctoral-level degrees and teaching experience in their specific area of expertise. Table 2-1 provides a

high-level sequence of the activities that occurred in the development of the DRC CCR item bank for ELA and Mathematics items.

Various item types were developed in order to best assess students' understandings of the standards. Descriptions of each item type used in the CCR item bank are included in Table 2-2.

It was determined that the State of Wisconsin would license ELA and Mathematics items from DRC's CCR item bank for the Spring 2016 test administration. Since Wisconsin students had not participated in DRC's national field-test, the test design incorporated an operational/field-test model. The flowchart in Figure 2-1 outlines the steps that were used to develop the Wisconsin Forward Exam administered in Spring 2016.

For Science and Social Studies, it was determined that due to the short window of time between contract award and forms construction, a combination of *TerraNova Third Edition* (CTB/McGraw-Hill, 2009) items that had not appeared on previous Wisconsin assessments and revised items that had appeared previously on other Wisconsin assessments would make up the core of the operational/field test. Details regarding the development and process for the review of the *TerraNova Third Edition*, may be found in the *2010 Wisconsin Knowledge and Concept (WKCE) Technical Report*. The efforts by DRC in developing items are in alignment with multiple best practices of the test 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. (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. (64)

2.1 Test Blueprints

The test blueprints specify the number of items for each reporting category and subskill. The process used for developing the blueprints was a collaborative effort between DRC and DPI. The DPI-approved blueprints can be found in Tables 2-3 through 2-6.

2.2 Reading Passage and Item Selection

Reading passages and items on the 2016 operational[1] field-test forms were selected, reviewed, and approved for placement on the Wisconsin Forward Exam in December 2015 by

[1] Operational items are those items that contribute to student scores. Operational items are abbreviated in this report as OP.

both DPI and Wisconsin educators at the passage and item review. The training Power Point presentation used at the review can be found in Appendix A.

2.3 Item Editing

Due to DPI leasing DRC's ELA and Mathematics CCR items, Wisconsin educators at the December 2015 item review could recommend item placement on the Wisconsin Forward Exam or recommend non placement. Item editing was not allowed for the ELA and Mathematics items. Committees were allowed to make recommendations for item edits to the Science and Social Studies items. Appendix B identifies the number of items for each standard per content area that were reviewed at the December 2015 item review.

Table 2-1 College- and Career-Ready Item Bank Development Activities

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

Table 2-2 CCR Item Bank Item Type Descriptions, by Item Type

Item Type	Name	Description
ESR	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 both mathematics and English language arts. Each of these items is worth one point.
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.
SCR	Short Constructed Response	Each short-constructed response item is designed to address writing through a short response as opposed to an essay. It assesses writing skills in ways a multiple-choice item cannot. The short-constructed response items are scored on a 0–2 point scale using item-specific scoring rubric.
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, auto-scoreable 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.
TDA	Text-Dependent Analysis	Each text-dependent analysis item is a text-based analysis based on a passage or a multiple-passage set that each student has read during the assessment. Both literature and informational texts are addressed through this item type. Students must draw on basic writing skills while inferring and synthesizing information from the passage in order to develop a comprehensive, holistic essay response. The demand required of a student's reading and writing skills in response to a TDA coincides with the similar demands required for a student to be college and career ready. The TDA prompts are scored using a holistic scoring guideline on a 1–4-point scale. This item type is supported by all Wisconsin English Language Arts standards across all grades for both Reading Literature and Reading Informational Texts and by the Writing standards 1, 2, 3, 4, and 9 across all grades. The TDA items were scored using artificial intelligence (AI) scoring, with an appropriate level of human scoring to validate the AI algorithms for all TDA items used in the Wisconsin ELA grades 3–8 assessments.

Table 2-3 English Language Arts Test Blueprints for Grades 3–8

Domain (Reporting Category)	Points Total by Grade					
	3	**4**	**5**	**6**	**7**	**8**
Reading	20	20	20	20	20	20
Key Ideas and Details	6–10	6–10	6–10	6–10	6–10	6–10
Craft and Structure/Integration of Knowledge and Ideas	4–10	4–10	4–10	4–10	4–10	4–10
Vocabulary Use Includes Language Standards 4 and 5	2–6	2–6	2–6	2–6	2–6	2–6
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	14	16	16	16	16	16
Text Types and Purposes	3–8	3–8	3–8	3–8	3–8	3–8
Research	3–8	3–8	3–8	3–8	3–8	3–8
Language Conventions	3–8	3–8	3–8	3–8	3–8	3–8
Text-Dependent Writing	12	12	12	12	12	12
Text-Dependent Analysis	12	12	12	12	12	12
Listening	7	8	8	8	8	8
ELA Points Total	53	56	56	56	56	56

Table 2-4 Mathematics Test Blueprints for Grades 3–8

Reporting Category	Total Points by Grade					
	3	**4**	**5**	**6**	**7**	**8**
Operations and Algebraic Thinking	8–10	9–11	8–10			
Number and Operations in Base Ten	7–9	8–10	8–10			
Number and Operations–Fractions	7–9	9–11	8–10			
Measurement and Data	9–11	9–11	9–11			
Geometry	6–8	6–8	8–10	6–8	9–11	9–11
Ratios and Proportional Relationships				6–8	7–9	
The Number System				10–12	6–8	7–9
Expressions and Equations				10–12	9–11	9–11
Statistics and Probability				9–11	10–12	7–9
Functions						9–11
Mathematics Points Total	**42**	**46**	**46**	**46**	**46**	**46**

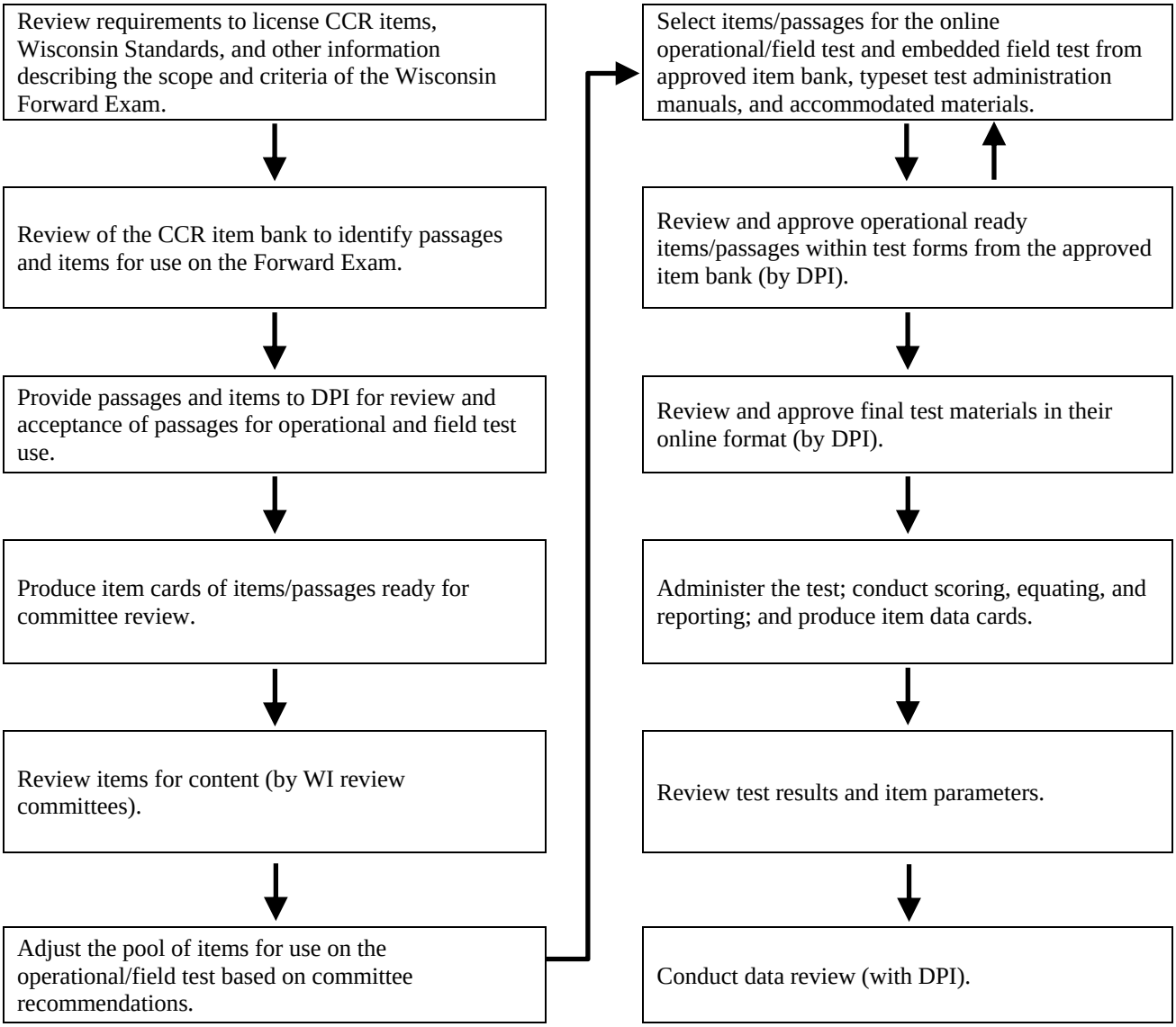
Table 2-5 Science Test Blueprints for Grades 4 and 8

Reporting Category	Total Points by Grade	
	4	**8**
Science Connections & Nature of Science	7–10	6–9
Science Inquiry	6–9	7–10
Physical Science	5–7	5–7
Earth and Space Science	5–7	5–7
Life & Environmental Science	5–7	5–7
Science Applications & Science in Social and Personal Perspectives	6–9	6–9
Science Total Points	**40**	**40**

Table 2-6 Social Studies Test Blueprints for Grades 4, 8, and 10

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

Figure 2-1 Operational/Field Test Development



Part 3: Test Form Development

Part 3 of this report focuses on key development tasks and issues related to creating the Spring 2016 Wisconsin Forward Exam operational/field-test forms. The test specifications and item development activities described in Part 2 explain how specific development processes provided evidence to support test validity, primarily content validity, 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 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 the test's use as an indicator of student achievement of state standards. Information is provided in Part 3 relating to the following topics:

- A general discussion of DRC's test creation and form review process
- The process of selecting operational/field-test items
- The process of obtaining DPI approvals

3.1 Overall Test Development Process

The creation of test forms involved the expertise of multiple DRC departments and DPI. The activities that contributed to the creation of the test forms are described below. The Wisconsin Forward Exam test development 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). (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. (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. (89)

3.1.1 Wisconsin Forward Test Form Creation

The DRC team worked cooperatively with DPI content and assessment specialists to select passages and prompts with associated content-specific items for the online assessments. The DRC team constructed forms that complied with the approved test blueprints and form construction guidelines. DRC content specialists used their extensive test design experience throughout forms construction, item development, test administration, scoring, and reporting processes. DRC successfully used an integrated team approach to test development, including content area specialists, psychometricians, and scoring specialists working as a unit in collaboration with DPI content experts.

3.1.2 Item Selection

As a first step in building the online assessments, the DRC team prepared all items that could be considered in the process in DRC's item banking system called IDEAS. The form, format, extent, and organization of items in their respective test sessions was determined in consultation with DPI.

Following preparation of all necessary materials and resources, forms construction began. Construction of the test forms themselves was a collaborative effort between DRC's integrated development team of assessment specialists, psychometric services specialists, and scoring specialists.

Before test forms were created, passages, items/performance tasks, and artwork were carefully selected. Below, we have described the process used for item selection:

- Using the pool of vendor-owned items, DRC test development specialists first selected items to match the approved test blueprints.
- DRC test development specialists checked to see that each item clearly aligned with the standards where applicable and that item, with available item statistics, met psychometric guidelines for inclusion in the test.
- DRC test development specialists verified that each item met technical quality for well-crafted items, including the following criteria:
    - One clearly correct answer (or answers if multi select)
    - Clear and concise wording
    - Grammatical correctness
    - Appropriate range of difficulty
    - Free of any offensive, inappropriate, or biased content
    - Met the Principles of Universal Design and maximum accessibility

In constructing the forms, the DRC content area test development specialists followed the guidelines provided below:

- Forms included adequate standards coverage, as required by test blueprints.
- No item in a form "clued" another item on that same form.
- "Clang" was avoided (i.e., distractors were unique from one another).
- Forms were ethnically diverse as needed, in terms of artwork and graphics.
- Forms included a wide range of topics and a variety of questions.
- Correct answer distributions were psychometrically sound.
- Forms did not contain any items that had been released to the public.
- DPI reviewed and gave final approval of all online test forms.

The Tables of Test Specifications in Appendices C, D, E, and F provide details on the number of items placed on the Spring 2016 Wisconsin Forward Exam per grade and content area for Reporting Category, Item Type, and Depth of Knowledge level. The ELA Table of Test Specifications is included in Appendix C; the Mathematics Table of Test Specifications is

contained in Appendix D; the Science Table of Test Specifications is provided in Appendix E; and the Social Studies Table of Test Specifications is given in Appendix F.

3.1.3 Quality Reviews

Content-area test development specialists and content-editorial specialists reviewed items and passages for technical quality; match to standard; bias, fairness, and sensitivity; depth of knowledge; estimated difficulty; estimated performance level descriptors (PLD); adherence to the Principles of Universal Design in all steps of the forms creation and forms review process. The aim for this team approach was to conduct a multi-tiered internal review of all passages and items prior to submission for review by DPI and then, with approval by DPI prior to submission, for external committees to ensure that all items align with Wisconsin's standards and adhere to DPI's standards for high-quality items.

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

- Content alignment or congruence with the knowledge and skills specified in the standards
- A range of estimated difficulty levels
- A range of estimated PLD
- Appropriate grade-level vocabulary, subject matter, and assumed student knowledge
- Freedom from issues or concerns regarding bias, sensitivity, or fairness
- Accessibility, following the Principles of Universal Design
- Correct grammar, usage, and structure/format

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

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

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

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

3.2 Description of the Wisconsin Forward Exam Assessments

The following sections detail each of the content areas assessed on the Spring 2016 Wisconsin Forward Exam assessments.

3.2.1 English Language Arts

Table 3-1 highlights the details of the ELA forms, including the number of passages and items at each grade level that were used in the core, vertical link, and embedded field-testing positions. In grades 3 and 8, there were two vertical linking forms, and at grades 4–7 there were three, each containing items from above- and below-grade level. Table 3-1 also identifies the various item types that appeared on the ELA forms, including the points for item scoring. Detailed description of the item types is provided in Part 2 of this report.

The ELA section of the online Wisconsin Forward Exam was divided into three sessions: reading, writing, 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 3-2 shows the operational Mathematics test structure. The Mathematics tests for grades 3–8 was administered in two testing sessions, with students able to take the sessions in any given order. Table 3-2 also illustrates the vertical link plan and the embedded field-test item count. Grades 3 and 8 had three forms: one contained vertical linking items and the other two had embedded field test items. Grades 4 through 7 had four forms: two contained vertical linking items and the other two had embedded field-test items.

In grades 6–8, the first session included both a non-calculator part and a calculator part in which the use of an embedded online calculator was allowed. Once students had completed the non-calculator part of the session, they were not allowed to return to those specific items and continued on with the remainder of that 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 3-3 presents the operational/field-test Science test structure. The Science test at grades 4 and 8 consisted almost entirely of *TerraNova* items but also included a few custom items developed specifically for placement on the Wisconsin Forward Exam. The embedded field-test design included the use of scenarios or tasks for students to respond to.

The Science test design detailed the number of points and recommended testing times for each grade level. These recommended testing times were also made available in the test administration manual.

3.2.4 Social Studies

Table 3-4 represents the operational/field-test structure of the Social Studies test. Each grade-level exam was administered in two testing sessions, with students able to complete the sessions in any order. The Social Studies test at grades 4, 8, and 10 consisted almost entirely of *TerraNova* items but also included a few custom items developed specifically for the Wisconsin Forward Exam.

The Social Studies test design detailed the number of points and recommended testing times for each grade level. These recommended testing times were also made available in the test administration manual.

3.3 DPI Approvals

The phases during which DPI had the opportunity to review passages and items to be placed on the Spring 2016 Wisconsin Forward Exam included the following:

- Prior to item content review
- At item content review
- During forms construction

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.

Table 3-1 English Language Arts Test Structure

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Passage Sets	**Literature**	2–3	2–3	2–3	2–3	2–3	2-3
	Informational	1–2	1–2	1–2	2–3	2–3	2-3
	Listening	2–3	2–3	2–3	2–3	2–3	2–3
Number of Core (OP) Items	**Item Type: SR/TE (1 pt.)**	24–26	25–27	25–27	25–27	25–27	25-27
	Item Type: SR/TE/EBSR (2 pts)	8–12	8–12	8–12	8–12	8–12	8-12
	Item Type TDA (12 pts)	1	1	1	1	1	1
	Total Core Items	34	35	35	36	36	36
Total Core Points		53	56	56	56	56	56
Vertical Linking (VL)	**Number of Forms**	2	3	3	3	3	2
	Passages	1–2*	1–2*	1–2*	1–2*	1–2*	1-2*
	Linking Items per Form	8	8	8	8	8	8
	Total Linking Items	16	24	24	24	24	16
Embedded Field Test (FT)	**Number of Forms**	8	8	8	8	8	8
	Passages	1	1	1	1	1	1
	FT Items per Form	8	8	8	8	8	8
	Total Items Field Tested	64	64	64	64	64	64
Total Items (Core + FT or VL) per Form		42	43	43	44	44	44
Total Estimated Testing Time (minutes)		133.5	134.5	133.5	140	139	140

*There are no more than two passages, but of those two, one is a listening passage.

Table 3-2 Mathematics Test Structure

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Core (OP) Items	**Item Type: MC/MS/SA (1 pt.)**	37	41	41	41	41	41
	Item Type: TE (1 pt.)	5	5	5	5	5	5
	Total Core Items	42	46	46	46	46	46
Total Core Points		42	46	46	46	46	46
Vertical Linking (VL)	**Number of Forms**	1	2	2	2	2	1
	Linking Items per Form	8	8	8	8	8	8
	Total Linking Items	8	16	16	16	16	8
Embedded Field Test (FT)	**Number of Forms**	2	2	2	2	2	2
	Field Test Items per Form	8	8	8	8	8	8
	Total Items Field Tested	16	16	16	16	16	16
Total Items per Form (Core + VL or FT)		50	54	54	54	54	54
Total Estimated Testing Time (minutes)		100	108	108	108	108	108

Table 3-3 Science Test Structure

Test Design		Grade	
		4	**8**
Number of Core (OP) Items	**Item Type: SR (1 pt.)**	40	40
Total Core Points		40	40
Embedded Field Test (FT)	**Number of Forms**	2	2
	Scenarios/Tasks	2	2
	Field Test Items per Form	8	8
	Total Items Field Tested	16	16
Total Items (Core + FT) per Form		48	48
Total Estimated Testing Time (minutes)		105	105

Table 3-4 Social Studies Test Structure

Test Design		Grade		
		4	**8**	**10**
Number of Core (OP) Items	**Item Type: SR (1 pt.)**	38	40	50
Total Core Points		38	40	50
Embedded Field Test (FT)	**Number of Forms**	2	2	2
	Field Test Items per Form	8	8	10
	Total Items Field Tested	16	16	20
Total Items (Core + FT) per Form		46	48	60
Total Estimated Testing Time (minutes)		69	72	90

Part 4: Test Administration

In the spring of 2016, Wisconsin administered assessments in ELA and Mathematics for grades 3–8. Science was administered in grades 4 and 8 and Social Studies in grades 4, 8, and 10. The test administration window was March 28–May 20, 2016. Part 4 of the Technical Report describes a set of standardized procedures and policies applied to administer the Wisconsin Forward Exam. The issue of test security in test administration that 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 in order to standardize the administration of the test are provided in this part as well. The following American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) standards are addressed in Part 4: 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 within schools will be held accountable to a common set of high academic content standards, the Wisconsin Academic Standards. As an alternate assessment, for students primarily being instructed using the Wisconsin Essential Elements as content standards, the DLM™ assessment measures the academic progress of students with significant cognitive disabilities in the subject areas of ELA and Mathematics at grades 3–11, Science at grades 4 and 8–11, and in Social Studies at grades 4, 8, and 10.

All other students are accountable to the knowledge and skills outlined in the Wisconsin Academic Standards. Those students who have an Individualized Education Program (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. Accommodations are changes in the routine conditions under which a student takes an assessment in order to provide the student an equal opportunity to demonstrate his or her knowledge. Accommodations provided to a student must be documented in his or her current IEP and used as a component of his or her regular instructional setting. DPI guidance makes it clear that the accommodations or supports provided to a student must be consistent for 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. (115)

An overview of the types of accommodations and 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.

Test administrators indicated which accommodations were to be available for use by each student within the student learning profile in DRC's eDIRECT system. All student accommodations are managed through DRC's eDIRECT system. 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 eDIRECT were District Assessment Coordinators and School Assessment Coordinators who were approved by DPI and assigned permissions accordingly for security purposes. The major functions are those of managing users and managing students. As such, eDIRECT was used to manage and update student information including demographic and accommodations/accessibilities information. All eDIRECT user roles and permission levels were approved by DPI.

4.1 Accommodations

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 (http://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 their needs. Those are defined as follows.

4.1.1 Universal Tools

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

4.1.2 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). They are either provided as digitally-delivered components of the test administration system or separate from it (embedded or non-embedded). All designated supports (embedded and non-embedded) must be entered into eDIRECT prior to test administration. Embedded supports will appear on student test tickets. Non-embedded supports will not appear on student test tickets; therefore, it is important to note which students have these supports to ensure they have access to them during testing.

4.1.3 Accommodations

Accommodations are those changes in procedures or materials that increase equitable access but do not compromise the grade-level standard or intended outcome of the assessment and are available for students for whom there is documentation of the need in the Individualized Education Program (IEP) or 504 accommodation plan. Accommodations are either provided as digitally delivered components of the test administration system or separate from it (embedded or non-embedded). All accommodations must be entered into eDIRECT prior to test administration. Embedded accommodations will appear on student test tickets.

Embedded Tools (online)

- Pause
- Breaks
- Sticky Notes
- Highlighter
- Keyboard Navigation
- Flag/Mark for Review
- Review Page
- Measuring Tools (Math)
- Cross-off Tool (Strikethrough)
- Magnifier Tool (Zoom)
- Help/What's This?
- Click to Enlarge
- Go to Question
- Tool Tips
- Test Directions

Embedded Designated Supports (online)

- Color Choices
- Contrasting Color
- Reverse Contrast
- Masking
- Text-to-Speech
- Spanish Translations (Stacked)

Embedded Accommodations (online)

- Visual Sign Language (online VSL delivery)
- Braille
- Text-to-Speech (reading passages)
- Print on Demand

Non-Embedded Tools, Supports, and Accommodations

- Pause (Breaks)
- Scratch Paper
- Word-to-Word Bilingual Dictionary
- Color Overlay
- Magnification
- Noise Buffers
- Read Aloud
- Scribe
- Separate Setting
- Abacus
- Alternate Response Options
- Multiplication Table
- Used translation
- Signed test questions and content to student
- Used Braille
- Used assistive device (e.g., text-talker, adaptive keyboard, picture symbols)
- Used a print-on-demand, paper-based version of the Wisconsin Forward Exam
- Used another DPI-approved accommodation
- Used a non-allowed accommodation resulting in the invalidation of test results

4.1.4 Translation

For the Spring 2016 Wisconsin Forward Exam administration, the State of Wisconsin used Spanish translation scripts. The aim of these scripts is to better help students demonstrate their knowledge on the Wisconsin Forward Exam when English language is part of the test construct. Students whose native language is Spanish were given the choice to use all or parts of the translation accommodation, which included a bilingual word list of commonly used content area vocabulary, translation of the test directions, and a written translation script of Mathematics, Science, and Social Studies test items. DPI recommended that educators also consult the list of allowable accommodations (referenced above) to create the most appropriate testing situation for their students.

DPI recognizes that approximately five percent of the Wisconsin limited English proficient population speaks a language other than Spanish. Districts who 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.

4.1.5 Additional Accessibility Resources

Additional accessibility resources guidance available at the testing sites 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 2016 Wisconsin Forward Exam along with the number and percentage of students provided these accommodations or supports.

4.2 Reporting Results of Assessments Taken with Accommodations

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.

4.3 Test Security

Maintaining the security of all test materials is crucial to preventing the possibility of random or systematic errors, such as unauthorized exposure of test items that would affect the valid interpretation of test scores. Several test security measures are 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. (116)

> **Standard 6.7** Test users have the responsibility of protecting the security of test materials at all times. (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 virtually administered 100% 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 (Braille and Print-on-Demand), so the assessment exposure is limited to those educators who required 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 2016 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 test session.
- Device lockdown during testing prevents students from copying, pasting, printing, and accessing other applications.
- If 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 SQA 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 demanded by DPI 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.

4.3.1 Secure Student Access

Students are required to provide a valid username and password to access the online testing system. The test administrator 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 eDIRECT secure administrative system, which is pre-

populated with student records. As an additional security measure, upon logging in, a Student Verification Page prompts the student to verify his or her 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 test administrator.

Because login tickets are secure material, it is recommended they be printed as close to the date of testing as possible and kept secure until given to the test administrator for distribution.

Test tickets and rosters are considered secure materials. As such, sites are instructed that they should be kept in a secure location until the session is scheduled to begin. Test tickets are distributed just prior to student login 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 sign 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.

4.3.2 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 test examiner. In addition, students must not be allowed to use any form of wireless communication during these breaks.

4.4 Test Administration

The purpose of each of the test administration workshops and the ancillary materials is to keep districts informed about policies and procedures related to testing in general and the Wisconsin Forward Exam program in particular. The information imparted is clearly related to standardizing the administration of the Wisconsin Forward Exam, maintaining the security of the assessment, allowing access to the assessments for special populations by clearly delineating appropriate designated supports or accommodations, and providing guidance on appropriate interpretations of the test results. These communication and training 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 *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

> **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. (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. (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. (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. (115)

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

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

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

- Responsibilities of District Assessment Coordinators (DACs)
- Responsibilities of School Assessment Coordinators (SACs)
- Responsibilities of District Technology Coordinators
- Responsibilities of Test Administrators (TA)/Proctors
- Test Times and Schedules
- Test Security
- Testing Procedures
- Accessibility Information
- Before Online Testing
- Technology Resources
- Additional Materials
- After Online Testing
- Packaging the Test Materials
- Procedures for Returning Materials
- Test Results
- Checklists for Responsible Parties (DACs, SACs, TAs)

In addition, Test Administration Manuals were made available to all test administrators. The manuals included the following:

- Test Administrator (TA)/Proctor Responsibilities
- Test Times/Schedules
- Test Security
- Accessibility Information
- Before Testing
- Test Tickets
- Testing Materials
- Setting Up Testing Environment
- During Online Testing
- After Testing

Student Preparation for Online Testing

Prior to testing, sites were encouraged to provide students with time to complete both a tutorial video series and an online tools training.

Student Tutorial Video

The Student Tutorial video was available for students (and TAs) to become familiar with the online testing environment. The video is broken into multiple chapters. A table was provided to help educators determine which videos students should view and the time required for each video chapter. 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 Trainings

The Online Tools Training (OTTs) are provided for students to allow them a hands-on opportunity to practice the types of items and tools available in the online testing system. OTTs are 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. 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 Video Sign Language (VSL), Text-to-Speech (TTS), Spanish translation, Masking, and Color Choice tools. VSL and Spanish OTTs were available by grade band (3–5, 6–8, and 10). The OTT was not scored and was not intended for content practice

Administration Supports Before and Following Testing

This administration was the first year that all testing was conducted 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 a final signoff for each document prior to delivery and public posting.

Table 4-8 displays a list of electronic 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 or specific information related to test administration, refer to the Test Coordinator's Guide and/or the Test Administration Manuals that are available online at: http://oea.dpi.wi.gov/oea_publications.

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

Grade 3	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Used Braille [BRL]	5	0.01%	4	0.01%
Used Print on Demand [POD]	6	0.01%	7	0.01%
Used Bilingual Dictionary			439	0.72%
Used Magnification	163	0.27%	152	0.25%
Used Noise Buffers	890	1.46%	881	1.44%
Used Read Aloud	2046	3.35%	2486	4.06%
Used Scribe	812	1.33%	796	1.30%
Used Separate Setting	6105	9.99%	6165	10.10%
Used Alternate Response Options	15	0.02%	15	0.02%
Used Read Aloud (Reading Passages)	431	0.71%	432	0.71%
Provided Color Choices [CC]	102	0.17%	104	0.17%
Used Contrasting Color [CTC]	109	0.18%	106	0.17%
Used Reverse Contrast [RC]	20	0.03%	20	0.03%
Used Masking [MSK]	723	1.18%	716	1.17%
Used Text-to-Speech [TTS]	7918	13.00%	9082	14.80%
Used Spanish Translation [ST]	634	1.04%	964	1.57%
Used Video Sign Language [VSL (ASL)]	27	0.04%	25	0.04%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	435	0.71%		
Used Abacus Math			36	0.06%
Used Non-embedded Calculator Math			171	0.28%
Used Multiplication Table Math			1051	1.72%

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**
Used Braille [BRL]	3	0.01%	3	0.01%	3	0.01%	3	0.01%
Used Print on Demand [POD]	7	0.01%	5	0.01%	5	0.01%	5	0.01%
Used Bilingual Dictionary			328	0.55%	323	0.54%	324	0.54%
Used Magnification	121	0.20%	121	0.20%	117	0.20%	117	0.20%
Used Noise Buffers	792	1.32%	786	1.31%	774	1.29%	769	1.29%
Used Read Aloud	1992	3.33%	2404	4.02%	2315	3.87%	2308	3.86%
Used Scribe	754	1.26%	758	1.27%	725	1.21%	724	1.21%
Used Separate Setting	6393	10.70%	6441	10.80%	6305	10.50%	6293	10.50%
Used Alternate Response Options	9	0.02%	9	0.02%	9	0.02%	9	0.02%
Used Read Aloud (Reading Passages)	457	0.76%	458	0.77%	458	0.77%	457	0.76%
Provided Color Choices [CC]	172	0.29%	164	0.27%	165	0.28%	165	0.28%
Used Contrasting Color [CTC]	148	0.25%	147	0.25%	147	0.25%	147	0.25%
Used Reverse Contrast [RC]	43	0.07%	43	0.07%	43	0.07%	42	0.07%
Used Masking [MSK]	796	1.33%	795	1.33%	783	1.31%	784	1.31%
Used Text-to-Speech [TTS]	7621	12.70%	8829	14.80%	8744	14.60%	8696	14.50%
Used Spanish Translation [ST]	650	1.09%	871	1.46%	849	1.42%	830	1.39%
Used Video Sign Language [VSL (ASL)]	36	0.06%	32	0.05%	28	0.05%	28	0.05%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	469	0.78%						
Used Abacus Math			19	0.03%				
Used Non-embedded Calculator Math			249	0.42%				
Used Multiplication Table Math			2035	3.40%				

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

Grade 5	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Used Braille [BRL]	2	0.00%	3	0.01%
Used Print on Demand [POD]	8	0.01%	7	0.01%
Used Bilingual Dictionary			261	0.44%
Used Magnification	224	0.38%	225	0.38%
Used Noise Buffers	826	1.38%	818	1.37%
Used Read Aloud	2056	3.45%	2373	3.97%
Used Scribe	763	1.28%	740	1.24%
Used Separate Setting	6442	10.80%	6478	10.80%
Used Alternate Response Options	15	0.03%	15	0.03%
Used Read Aloud (Reading Passages)	484	0.81%	487	0.82%
Provided Color Choices [CC]	109	0.18%	109	0.18%
Used Contrasting Color [CTC]	133	0.22%	139	0.23%
Used Reverse Contrast [RC]	31	0.05%	32	0.05%
Used Masking [MSK]	777	1.30%	760	1.27%
Used Text-to-Speech [TTS]	6966	11.70%	8025	13.40%
Used Spanish Translation [ST]	427	0.72%	601	1.01%
Used Video Sign Language [VSL (ASL)]	47	0.08%	41	0.07%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	545	0.91%		
Used Abacus Math			30	0.05%
Used Non-embedded Calculator Math			406	0.68%
Used Multiplication Table Math			2420	4.05%

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

Grade 6	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Used Braille [BRL]	3	0.00%	4	0.01%
Used Print on Demand [POD]	5	0.01%	5	0.01%
Used Bilingual Dictionary			216	0.36%
Used Magnification	79	0.13%	79	0.13%
Used Noise Buffers	371	0.62%	370	0.61%
Used Read Aloud	1442	2.40%	1721	2.86%
Used Scribe	402	0.67%	379	0.63%
Used Separate Setting	5821	9.68%	5866	9.74%
Used Alternate Response Options	9	0.01%	7	0.01%
Used Read Aloud (Reading Passages)	459	0.76%	460	0.76%
Provided Color Choices [CC]	99	0.16%	97	0.16%
Used Contrasting Color [CTC]	125	0.21%	125	0.21%
Used Reverse Contrast [RC]	45	0.07%	44	0.07%
Used Masking [MSK]	964	1.60%	906	1.50%
Used Text-to-Speech [TTS]	6356	10.60%	7290	12.10%
Used Spanish Translation [ST]	180	0.30%	260	0.43%
Used Video Sign Language [VSL (ASL)]	31	0.05%	25	0.04%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	517	0.86%		
Used Abacus Math			5	0.01%
Used Non-embedded Calculator Math			563	0.93%
Used Multiplication Table Math			2478	4.11%

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

Grade 7	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Used Braille [BRL]	3	0.01%	4	0.01%
Used Print on Demand [POD]	7	0.01%	6	0.01%
Used Bilingual Dictionary			235	0.39%
Used Magnification	121	0.20%	124	0.21%
Used Noise Buffers	265	0.45%	262	0.44%
Used Read Aloud	1160	1.95%	1337	2.24%
Used Scribe	271	0.46%	254	0.43%
Used Separate Setting	5673	9.53%	5695	9.56%
Used Alternate Response Options	11	0.02%	12	0.02%
Used Read Aloud (Reading Passages)	396	0.67%	399	0.67%
Provided Color Choices [CC]	200	0.34%	198	0.33%
Used Contrasting Color [CTC]	116	0.19%	117	0.20%
Used Reverse Contrast [RC]	32	0.05%	33	0.06%
Used Masking [MSK]	1239	2.08%	1231	2.07%
Used Text-to-Speech [TTS]	6109	10.30%	7019	11.80%
Used Spanish Translation [ST]	203	0.34%	256	0.43%
Used Video Sign Language [VSL (ASL)]	41	0.07%	32	0.05%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	502	0.84%		
Used Abacus Math			16	0.03%
Used Non-embedded Calculator Math			705	1.18%
Used Multiplication Table Math			2310	3.88%

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

Grade 8	English Language Arts		Mathematics		Science		Social Studies	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**	**N Count**	**Percent**	**N Count**	**Percent**
Used Braille [BRL]	5	0.01%	5	0.01%	5	0.01%	5	0.01%
Used Print on Demand [POD]	6	0.01%	6	0.01%	6	0.01%	6	0.01%
Used Bilingual Dictionary			238	0.40%	237	0.40%	238	0.40%
Used Magnification	95	0.16%	95	0.16%	94	0.16%	93	0.16%
Used Noise Buffers	275	0.47%	276	0.47%	269	0.46%	268	0.45%
Used Read Aloud	1014	1.72%	1231	2.08%	1183	2.01%	1178	2.00%
Used Scribe	239	0.41%	235	0.40%	231	0.39%	230	0.39%
Used Separate Setting	5686	9.64%	5709	9.66%	5586	9.47%	5576	9.45%
Used Alternate Response Options	5	0.01%	6	0.01%	5	0.01%	5	0.01%
Used Read Aloud (Reading Passages)	306	0.52%	309	0.52%	307	0.52%	308	0.52%
Provided Color Choices [CC]	204	0.35%	203	0.34%	201	0.34%	199	0.34%
Used Contrasting Color [CTC]	128	0.22%	129	0.22%	127	0.22%	127	0.22%
Used Reverse Contrast [RC]	28	0.05%	28	0.05%	26	0.04%	26	0.04%
Used Masking [MSK]	952	1.61%	950	1.61%	948	1.61%	947	1.60%
Used Text-to-Speech [TTS]	5699	9.66%	6606	11.20%	6412	10.90%	6383	10.80%
Used Spanish Translation [ST]	199	0.34%	253	0.43%	248	0.42%	249	0.42%
Used Video Sign Language [VSL (ASL)]	34	0.06%	25	0.04%	25	0.04%	25	0.04%
Used Text-to-Speech for Reading Passages [TTS (PSGS)] ELA	456	0.77%						
Used Abacus Math			13	0.02%				
Used Non-embedded Calculator Math			774	1.31%				
Used Multiplication Table Math			2011	3.40%				

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

Grade 10	Social Studies	
Accommodation or Support	**N Count**	**Percent**
Used Braille [BRL]	7	0.01%
Used Print on Demand [POD]	14	0.02%
Used Bilingual Dictionary	127	0.20%
Used Magnification	58	0.09%
Used Noise Buffers	40	0.06%
Used Read Aloud	609	0.98%
Used Scribe	96	0.15%
Used Separate Setting	3759	6.04%
Used Alternate Response Options	12	0.02%
Provided Color Choices [CC]	31	0.05%
Used Contrasting Color [CTC]	24	0.04%
Used Reverse Contrast [RC]	7	0.01%
Used Masking [MSK]	319	0.51%
Used Text-to-Speech [TTS]	2979	4.79%
Used Spanish Translation [ST]	134	0.22%
Used Video Sign Language [VSL (ASL)]	13	0.02%

Table 4-8 Summary Table of Manual Materials

Material	Configuration
DAC/SAC Guide (District Assessment Coordinator/School Assessment Coordinator Guide)	The DAC/SAC Guide is a 30-page handbook that includes the following information: • Key dates • Roles and responsibilities • Test security • Accessibility information • Procedures before testing begins • Technology resources • Testing times and schedules • Braille ordering • Overview of testing and test management software • Procedures for once testing is finished • Transferring students • Coordinator checklists • Guidelines and procedures for documenting a test security incident • Multiplication chart (for use with some tests) • Sample test schedules
eDIRECT Guide: Manage Users	The Manage Users Guide is a 38-page guide that includes the following information: • Managing user's own eDIRECT account • Adding and editing other eDIRECT users • Adding and removing eDIRECT user permissions
eDIRECT Guide: Students and Testing	The Students and Testing Guide is a 52-page guide that includes the following information: • 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
Accessibility Guide	The Accessibility Guide is a 19-page document that outlines the various accessibility options available to students taking the Wisconsin Forward Exam. Guidelines for using the various accessibility features were also included.
Student Tutorial	The Student Tutorial includes 11 video "chapters" intended for students. 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 2016 tutorial also includes four chapters for test coordinators and proctors to familiarize them with the functionality of the accessibility features of the Wisconsin Forward Exam.

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

Material	Configuration
TAM (Test Administration Manual) and Test Directions	The TAMs was a 29-page document intended for test proctors. It includes the following information: • Key dates • Test times and schedules • Test security • Accessibility information • Procedures for before testing • Test ticket management • Test material management • Setting up the testing environment • Procedures for during testing • Procedures for after testing • Proctor checklist and guidelines • Read-aloud protocol • Scribe guidelines Test Directions are presented in seven documents, one per grade. Each set of test directions includes a script for test proctors as they guide students through logging in to the INSIGHT test software and through the online test directions screens.
Technology User Guide (TUG)	The TUG is an approximately 248-page document intended for Technology Coordinators. It includes detailed instructions on the installation and configuration of INSIGHT and the TSM for all supported platforms.
Interpretive Guide	The Interpretive Guide is a 30-page document that includes the following information: • Interpreting Wisconsin Forward Exam scores • Accessing Individual Student Reports (ISRs) and summary reports in eDIRECT and PRISM
Technology Readiness Package	The Technology Readiness Package is a suite of documents and tools for Technology Coordinators to prepare for the Wisconsin Forward Exams that includes the following: • Capacity Estimator • System requirements • Technology overview presentation • Technology Coordinator Checklist • Tech FAQ
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
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 scaled 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 • Measures of scoring reliability for text-dependent analysis items • Evidence of test validity
Data Forensic Report	A separate Data Forensic Report will include analyses of the following: • Evaluation of wrong-to-right response changes • Evaluation of student response time to items • Examination of possible copying of written responses from another student

Part 5: Scoring

The purpose of Part 5 is to demonstrate adherence to the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) Standards 4.18, 4.20, 6.8, and 6.9. Standard 4.18 provides some general guidance for Part 5:

> 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. (91)

Part 5 describes the following:

- The scoring process of multiple-choice (MC) and multi-select items
- The auto-scoring process of technology-enhanced (TE) items
- The scoring of text-dependent analysis (TDA) items
  - scoring rubrics
  - Artificial Intelligence (AI) scoring process
  - handscoring process
  - electronic handscoring system
  - scoring personnel selection
  - anchor papers selection
  - TDA item scores distribution

5.1 Multiple-Choice and Multi-Select Item Scoring Process

Responses to MC and multi-select 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 test administrator. All MC and multi-select items had one and only one correct item response for each item.

5.2 Technology-Enhanced Item Scoring Process

All TE 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 TE items. DRC established an adjudication process for TE items and any gridded responses to verify that correct answers were identified. The quality process for DRC's TE item scoring included the following:

- A scoring rubric was created for each TE item. It was similar to describing the one and only correct answer for dichotomously scored items (scored as either right or

wrong). If partial credit was possible, the rubric described in detail the type of response that could receive credit for each score point.

- The information from the scoring rubric was entered into the scoring system within the item banking system so that the truth 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 that 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 re-run. 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 their frequencies and received scores.

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

5.3 Scoring of Text Dependent Analysis Items

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

5.3.1 Description of Scoring Rubrics and Non-Score Codes

In the 2016 administration, the ELA forms in grades 3–8 contained one TDA item at each grade level. The TDA item responses were scored using a 4-point holistic rubric. The responses were scored using AI engine and then validation scoring was performed, using human scorers, on approximately twenty percent of the AI scored responses. Table 5-1 presents the scoring rubric. In cases where student responses could not be scored, a non-score code was used. The non-score codes are presented in Table 5-2. All non-score codes were converted to a score of “0” in derivation of student total test scores.

5.3.2 Artificial Intelligence (AI) Scoring

DRC partnered with Measurement Incorporated (MI) to score the TDA tasks. MI is a recognized leader in the field of automated essay scoring. MI employed its essay scoring engine (PEG) to score all student responses. The AI model for scoring the Wisconsin student responses was built by first having DRC expert scorers score a representative sample of Wisconsin responses twice, independently. Once the sample was scored, responses and corresponding scores were delivered to the AI team at MI for model development. MI's linguistics, software developers, psychometricians, and human-computer interactions specialists created task-specific algorithms that were then used to accurately predict how humans would score these responses.

MI's AI scoring software flagged a small percentage of student responses that could not be AI scored. The software has various triggers for identifying alert responses and responses in which it has low confidence. These responses lack proper development, lack enough content to be scored, are written in an unsupported language, or contain inappropriate language or represent a bad faith effort to complete the test (e.g., repeated text, off-topic text). The limited number of responses that could not be scored by AI were routed to DRC for human scoring with a condition code indicating why the response could not be AI scored.

5.3.3 Handscoring Process

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

5.3.4 Handscoring System

Scoreboard, a DRC's handscoring system, was used to score TDA items. Scoreboard presented images of rendered online responses to trained scorers who assigned scores for the TDA items. The rendered student responses were viewed on high-quality workstation monitors. Images of each student's responses were automatically routed to designated groups of scorers trained to score these items.

5.3.5 Anchor Papers and Training Papers

All training materials, including scoring guides and rubrics, anchor papers, training papers, and qualification papers, were selected from live student work. Prior to actual scoring, a

selected group of papers written by Wisconsin students were selected as models to train scorers for scoring. These papers, referred to as anchor papers, played an important role in deciding which level of writing should receive which score. The range finding committee, made up of six scoring directors (one from each grade) then chose those papers that had a high level of agreement to create a set of anchor papers and a set of training papers for each grade. These anchor and training papers were then used to train a select group of scorers who scored approximately 2,000 student responses used to train the AI engine (model building). For this model-building activity, each student response was independently scored twice, by two separate scorers. Only those responses that had two identical scores were used to train the AI engine. Once trained, the AI engine scored the remaining Wisconsin student responses. Upon completion of the AI scoring, a random sample consisting of approximately twenty percent of the student responses scored by the AI engine was sent to DRC for a read-behind. DRC then scored the twenty percent read-behind sample using the original AI engine scoring group to ensure consistency. The twenty percent read-behind with human scorers served as a validation check of the AI engine scoring data.

5.3.6 Scoring Personnel and Qualifications

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

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

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

Scorers—Many DRC scorers had years of classroom teaching experience. The DRC scorer pool included many retired and current educators, as well as engineers, editors, published authors, and individuals with advanced degrees. The minimum qualification for all scorers was a Bachelor's degree. Scorers were required to participate in training and successfully pass a qualification round. Once qualified, scorers could start scoring, but throughout the scoring process, scorer performance was assessed by a scoring director, a team leader, and the project manager through read-behinds and reviews of inter-rater reliability statistics, as described in Sections 5.3.8, 5.4, and Part 9.

Team Leaders—Team leaders were selected on the basis of their ability to maintain a high degree of scoring accuracy and consistency, often across multiple content areas and grades. Team leaders were also required to possess good interpersonal and leadership skills in order to be effective when training and counseling scorers. Team leaders were each responsible for a small team of scorers. In addition to performing read-behinds on scorers, team leaders also

coached scorers when needs were identified through data review or otherwise by supervisory staff.

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

5.3.7 Scorer Training

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

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

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

5.3.8 Monitoring the Scoring Process

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

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

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

5.3.9 Final Scores

All TDA responses were sent to the AI engine for scoring. The AI scores were the final scores (i.e., scores of record). In all cases where the AI engine returned a non-scoreable condition code, the student responses were reviewed and scored by humans and a resolution was reached. Those scores, then, became scores of record.

5.4 Inter-Rater Reliability

A random twenty percent of the AI-scored responses were sent to human scorers for the second reads and used to validate (assess the accuracy of) the AI score. The statistics for the inter-rater reliability were calculated for all TDA items. To determine the reliability of scoring, the score distribution and percentage of agreement of the two readers were examined. In this section, the distribution of TDA item scores is presented. Additional inter-rater reliability measures including intra-class correlation and weighted kappa statistics are presented in Part 9 of the Technical Report.

5.4.1 Distribution of TDA Item Scores

Table 5-3 shows the score and non-scoreable code distributions for TDA items. The presented scores are from the AI engine. It should be noted that a large number of records displayed condition code "B" (blank/omit) or "N" (insufficient to score). Such an outcome may be influenced by the fact that the TDA item type was administered to Wisconsin students for the first time in the 2016 administration, and many students might not have been familiar with that item type. It is expected that the number of students in these two non-scoreable code categories will decrease over time.

Table 5-4 shows the score and non-scorable code distribution for TDA items for responses selected for the second read (hand-scoring). Table 5-5 shows the associated percentage of scores and non-scorable code for TDA items for responses selected for the second read. In both tables, Scorer 1 is the AI engine and Scorer 2 is a human scorer. As shown in Tables 5-4 and 5-5, there was a high degree of agreement between the AI engine and the human scorer for all grades except grade 4 where a human scorer scored approximately nine percent of the responses as non-scoreable code N while the AI engine scored less than one percent of the responses as non-scoreable code N.

Taken together, the information presented in this part of the Technical Report summarizes the scoring procedures for different types of items and the steps taken by DRC to ensure accuracy in the TE item scoring, AI scoring, and handscoring processes. The inter-rater reliability statistics presented in Section 5.4 demonstrate that the items are scored reliably during the scoring process. These efforts by DRC follow multiple best practices of the testing industry and support AERA, APA, & NCME (2014) Standards 4.18, 4.20, 6.8, and 6.9 as presented in Part 5.

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

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

Table 5-1 TDA Scoring Guidelines, Grades 3–8 (cont.)

Score Value	Score Description	Scoring Rubrics
2	Demonstrates limited analysis of text and inconsistent writing	• Inconsistent addressing of some parts of the task to demonstrate a partial understanding of the text(s) • Inconsistent analysis based on explicit and/or implicit meanings from the text(s) that ineffectively supports claims, opinions, and ideas • Weak organizational structure and focus on the task with ineffectively grouped ideas, including a weak introduction, development, and/or conclusion • Limited and/or vague reference to the text(s) using some details, examples, quotes, and/or facts • Limited reference to the main ideas and relevant details of the text(s) • Limited use of transitions to link ideas within categories of textual and supporting information • Inconsistent use of precise language and domain-specific vocabulary drawn from the text(s) • Errors in sentence formation, grammar, usage, spelling, capitalization, and punctuation that may interfere with meaning
1	Demonstrates minimal analysis of text and inadequate writing	• Minimal addressing of part(s) of the task to demonstrate an inadequate understanding of the text(s) • Minimal analysis based on the text(s) that may or may not support claims, opinions, and ideas • Minimal evidence of an organizational structure and focus on the task with arbitrarily grouped ideas that may or may not include an introduction, development, and/or conclusion • Insufficient reference to the text(s) using few details, examples, quotes, and/or facts • Minimal reference to the main ideas and relevant details of the text(s) • Few, if any, transitions to link ideas • Little or no use of precise language or domain-specific vocabulary drawn from the text(s) • Many errors may in sentence formation, grammar, usage, spelling, capitalization, and punctuation that often interfere with meaning

Table 5-2 TDA Non-Scoreable Codes, Grades 3–8

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

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

Table 5-3 TDA Score Distribution

Grade	Item	N	1	2	3	4	B+N	C	R	T
3	20	61120	28301	13038	2044		17670	6	41	20
4	21	59776	22676	6748	1139	54	28994	17	93	55
5	20	59662	27065	12101	2323	20	18103		40	10
6	22	60164	29119	10897	1302		18785	3	33	25
7	21	59539	29869	12479	2961	429	13743	3	39	16
8	21	59006	26596	13626	3415	387	14933	1	43	5

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

Grade	Scorer	Total Count	Score Count				Non-Scoreable Code Count				
			1	2	3	4	B	C	N	R	T
3	Scorer 1 (AI Engine)	8982	5820	2727	435						
3	Scorer 2 (Human)	8982	5866	2708	139	1		11	253		4
4	Scorer 1 (AI Engine)	6353	4675	1393	232	10			43		
4	Scorer 2 (Human)	6353	4591	1039	125	5			589		4
5	Scorer 1 (AI Engine)	8595	5666	2470	455	3			1		
5	Scorer 2 (Human)	8595	5960	2136	298	6			192		3
6	Scorer 1 (AI Engine)	8547	5947	2309	290				1		
6	Scorer 2 (Human)	8547	6039	2119	263	15			105	3	3
7	Scorer 1 (AI Engine)	9506	6208	2602	624	72					
7	Scorer 2 (Human)	9506	6148	2694	324	36			303		1
8	Scorer 1 (AI Engine)	9093	5499	2829	668	97					
8	Scorer 2 (Human)	9093	5601	2677	511	91			211	2	

Note: This table does not include records for when the AI engine returned a non-scoreable code. Such cases were resolved by a human scorer and are not included in computation of rater-agreement statistics.

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

Grade	Scorer	Total Count	Score Percentage				Non-Scoreable Code Percentage				
			1	2	3	4	B	C	N	R	T
3	Scorer 1 (AI Engine)	8982	64.80	30.36	4.84						
	Scorer 2 (Human)	8982	65.31	30.15	1.55	0.01		0.12	2.82		0.04
4	Scorer 1 (AI Engine)	6353	73.59	21.93	3.65	0.16			0.68		
	Scorer 2 (Human)	6353	72.27	16.35	1.97	0.08			9.27		0.06
5	Scorer 1 (AI Engine)	8595	65.92	28.74	5.29	0.03			0.01		
	Scorer 2 (Human)	8595	69.34	24.85	3.47	0.07			2.23		0.03
6	Scorer 1 (AI Engine)	8547	69.58	27.02	3.39				0.01		
	Scorer 2 (Human)	8547	70.66	24.79	3.08	0.18			1.23	0.04	0.04
7	Scorer 1 (AI Engine)	9506	65.31	27.37	6.56	0.76					
	Scorer 2 (Human)	9506	64.67	28.34	3.41	0.38			3.19		0.01
8	Scorer 1 (AI Engine)	9093	60.48	31.11	7.35	1.07					
	Scorer 2 (Human)	9093	61.60	29.44	5.62	1.00			2.32	0.02	

Note: This table does not include records for when the AI engine returned a non-scoreable code. Such cases were resolved by a human scorer and are not included in computation of rater-agreement statistics.

Part 6: Calibration, Scaling, and Deriving Scale Scores

This part of the Technical Report describes the analyses involving test calibrating, scaling, and student scoring that occurred for the Wisconsin Forward Exam after the 2016 test administration. Part 6 demonstrates adherence in the Wisconsin Forward Exam program data analysis to AERA, APA, & NCME (2014) Standards 1.8, 2.13, 5.2, and 7.2. Each standard will be explicated within the appropriate section of this chapter. Standard 7.2 provides general guidance that is relevant to this chapter:

> 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. (126)

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

Calibration is the mathematical process of estimating characteristics of individual items. These characteristics are termed "item parameters." Section 6.1 serves to explain this process, beginning with a description of the calibration and scaling design and methods that were applied to the Spring 2016 Wisconsin Forward Exam, followed by a presentation of a calibration sample, and a discussion of the calibration models and the software used. The results of the calibration process, using model-to-data fit statistics, and the outcomes of test scaling are discussed in Section 6.2. Section 6.3 addresses the process for derivation of scale scores from raw scores.

Readers should note that calibration, scaling, and scoring using IRT are mathematically complex and computationally intensive processes. A full understanding of these topics requires a background in psychometrics. However, in order to make these processes more accessible and transparent to a wider range of audiences, a brief, nontechnical explanation of how scale scores are derived from raw scores is provided in Section 6.3. Additional references are also provided.

6.1 Test Scaling Design

This section of the report outlines the scaling design for the Spring 2016 Wisconsin Forward Exam. A historical background of the Wisconsin ELA, Mathematics, Science, and Social Studies reporting scales is provided first.

ELA and Mathematics —The 2014–15 ELA and Mathematics Wisconsin Badger Exam for grades 3–8 was a customized version of the Smarter Balanced Assessment Consortium (SBAC) assessment. In the 2014–15 administration year, the ELA and Mathematics grades 3–8 test scores were reported on SBAC vertical scales. These scales ranged from ~2100 to ~2700 scale

score points for ELA and from ~2200 to ~2800 scale score points for Mathematics across grades 3 through 8. The reported test scores were four-digit scale scores.

In prior administration years (up to the 2013–14 administration), the Reading and Mathematics Wisconsin Knowledge and Concept Examination (WKCE) scores were reported on the custom vertical scales. These scales were developed after the Fall 2005 Reading and Mathematics test administration. Reading and Mathematics assessments spanned grades 3–8 and grade 10. Language Arts, Science, and Social Studies assessments were administered in grades 4, 8, and 10, and these assessments were reported on grade-unique scales. The reported scores for all assessments were three-digit scores. Science and Social Studies continued to be reported on WKCE scales in 2014–15 test administration.

The Spring 2016 Wisconsin Forward Exam for ELA and Mathematics were developed to the same content standards as the Spring 2015 assessments but did not include any items from the 2015 assessments. Also, DPI has expressed a wish to move away from SBAC scales and revert to reporting total test scores as 3-digit scale scores. Therefore, the 2016 ELA and Mathematics assessments cannot be directly equated to the previous year's scales. In addition, due to changes in the test content and the test administration window, linking of the Spring 2016 assessments to the Fall 2013 assessments was not recommended. Therefore, no items from the prior administration years were included in the 2016 ELA and Mathematics assessments.

Given the constraints described above, new ELA and Mathematics vertical scales were established after the 2016 test administration. To accomplish this goal, sets of off-grade-level items (that is, items from the grade above, from the grade below, or from both grades above and below) were administered to samples of Wisconsin students taking on-grade-level operational tests to facilitate between-grade assessment linking. For example, a sample of grade 3 students took a set of grade 4 items in addition to a full grade 3 operational test; a sample of grade 4 students took a set of grade 3 items in addition to a full grade 4 operational test; and, another sample of grade 4 students took a set of grade 5 items in addition to a full grade 4 operational test. The off-grade-level items (also called vertical linking items) did not count toward the student test score. The vertical linking items represented the test content of the grade from which they were selected and the grade in which they were administered as closely as possible. There were approximately 12 items in each vertical linking set for ELA and approximately 8 items in each vertical linking set for Mathematics.

Vertical scaling is a useful tool to measure student growth from one year to the next. A vertical scale, which can be viewed as a developmental continuum, permits educators to make inferences about student achievement across grades. As students develop new capabilities or demonstrate new skills, they move up the continuum, as demonstrated by their scale scores. These scale scores represent units on a single, equal-interval scale applied across all grade levels.

Science and Social Studies —The Spring 2016 Science and Social Studies tests were developed to the same content standards as the Fall 2014 assessments. However, due to the test window changes (from Fall to Spring) and the inclusion of new or revised items on the test, the new tests could not be equated to previous scales. Instead, new scales for Science grades 4 and 8 and Social Studies grades 4, 8, and 10 were established after the Spring 2016 administration.

The scaling of the Science and Social Studies tests was conducted separately for each grade in each content area. Because Science and Social Studies assessments are not administered in adjacent grades, vertical scaling of these assessments could not be performed. Instead, the scales were established in such a way to show "vertical relationship" (i.e., an increase in scale score means) across grades. This approach is similar to the method used to establish original WKCE scales for Science and Social Studies after the 2005 test administration (refer to Part 8 and Part 11 of the *WKCE Technical Report* from the Fall 2005 WKCE administration, which can be found in Appendix 3 of the *WKCE 2010 Technical Report* available at http://dpi.wi.gov/sites/default/files/imce/assessment/pdf/td-2010-techman.pdf).

These new scales for all content areas become new baselines for monitoring student performance in these content areas moving forward.

6.1.1 Calibration Models

Item parameters for items contained in ELA and Mathematics tests were estimated using a marginal maximum-likelihood procedure to simultaneously estimate the item parameters for MC and CR items using the three-parameter logistic (3PL) model and the two-parameter partial credit (2PPC) IRT model (Bock & Aitkin, 1981; Thissen, 1982). All non-MC items, including TE, ESR, MS, SA, and TDA items, were treated as CR items in calibrations. Item parameters for items contained in Science and Social Studies were estimated using a marginal maximum-likelihood procedure and the 3PL model (all items in Science and Social Studies tests were MC items).

Under the 3PL model, the probability that a student with a trait or scale score θ will respond correctly to a multiple-choice item *j* is

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

In the equation, a_j is the item discrimination, b_j is the item difficulty, and c_j is the probability of a correct response by a very low-ability student. Under the 2PPC model, the probability that a student with a trait or scale score 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 discrimination and location parameters for the MC items are in the traditional 3PL metric and are labeled *a* and *b*, respectively. In the 2PPC model, *f* (alpha) and *g* (gamma) are analogous to *a* and *b*, where alpha is the discrimination parameter and gamma over alpha (*g*/*f*) is the location where adjacent trace lines cross on the ability scale. Because of the different metrics used, the 3PL parameters *a* and *b* are not directly comparable to the 2PPC parameters *g* and *f*; however, they

can be converted to a common metric. The two metrics are related by $a = f / 1.7$ and $b = g/f$ (Burket, 2002). As a result of this procedure, the MC and CR items are placed on the same scale. Note that for the 2PPC model, there are m_j–1 (where m_j is a score level j) independent g's and one f, for a total of m_j independent parameters estimated for each item, while there is one a and one b per item in the 3PL model.

Using the 3PL/2PPC model for estimation of ELA and Mathematics item parameters and 3PL model for estimation of Science and Social Studies item parameters was consistent with the past methodology (except for administration year 2014–15 for ELA and Mathematics) implemented for these content areas in the Wisconsin testing program. Item parameters estimated after the 2015–16 test administration were used to score Wisconsin students who took these tests.

6.1.2 Calibration Sample

The calibration and scaling of the Wisconsin Forward Exam occurred after the Spring 2016 test administration and was based on student data from an early return sample of the state test data. This arrangement was chosen in order to expedite the data analysis in preparation for the standard setting which occurred in June 2016. This section provides information on the comparability of the calibration sample to the census data in terms of demographic characteristics in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) *Standards*:

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

The calibration sample consisted of the student data acquired before the testing window ended and included students from public and charter schools. The characteristics of the calibration sample compared to the population of students in public schools are presented in Tables 6-1 through 6-4 for ELA, Mathematics, Science, and Social Studies, respectively. The 2016 calibration sample was comparable to the Wisconsin public school student population.

6.1.3 Test Calibration and Scaling

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

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

The Wisconsin Forward Exam scores are produced using the IRT models which assume that each of the items and tasks is an independent indicator of the underlying ability governing the propensity for students to answer an item correctly (or with greater correctness in the case of the multilevel constructed-response items).

Calibrating and scaling of the Wisconsin Forward Exam data were 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.

ELA and Mathematics

In a process of ELA and Mathematics item calibration, the number of estimation cycles was set to 200 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. It should be noted that there was a small number of items with the default value for the *c*-parameter on the ELA and Mathematics tests. When the PARDUX (Burket, 2002) program, used to calibrate the items, encounters difficulty estimating the *c*-parameter, it assigns a default *c*-parameter value of 0.20.

As stated in the previous section, new scales were established for ELA and Mathematics after the 2015–16 test administration. The test forms in adjacent grade levels of each content area shared common items and were calibrated concurrently at that grade level.

Concurrent calibration is a method that allows for establishing the common scale in a single step—the calibration phase—by simultaneously estimating parameters for all items at all grades. The estimated parameters in the theta metric are on the same scale. In addition, population ability estimates are obtained for multiple groups. The population mean and standard deviation for the base grade are then used to compute the M1 and M2 transformation parameters to convert the parameter estimates of the other grades onto the common scale score metric. Tables 6-5 and 6-6 present the calibration sample mean and standard deviation ability estimates for multiple groups, as obtained from the concurrent calibration for ELA and Mathematics, respectively.

After placing item parameters on common scales for ELA and Mathematics, the grade 5 theta means were re-estimated using only item parameters for on-grade-level items. These estimates were then used to identify transformation constants that would allow transformation of item parameter estimates in a theta metric into a scale score metric and produce a scale with a target mean of 600 and a standard deviation of 50 for grade 5 of both ELA and Mathematics assessments.

Science and Social Studies

In the process of Science and Social Studies 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.

Similar to results obtained for ELA and Mathematics, there was a small number of items with the default value of 0.20 for the *c*-parameter on the Science and Social Studies tests.

Science and Social Studies test data were calibrated separately for each grade level and content area. As stated in the previous section, new scales were established for these content areas after the 2015–16 test administration. The mean and standard deviation of ability estimates for each grade were estimated and used to identify transformation constants that allowed transformation of item parameter estimates in a theta metric (from calibration) into a scale score metric and to produce student scale score distribution with a target mean and standard deviation for each grade. In order to differentiate the new Science scales from the previous ones, a scale score mean of 400 and a standard deviation of 50 were set for grade 4, and a scale score mean of 600 and a standard deviation of 50 were set for grade 8 Science. Similarly, in order to differentiate the new Social Studies scales from the previous ones, a scale score mean of 400 and a standard deviation of 50 were specified for grade 4, a scale score mean of 600 and a standard deviation of 50 were specified for grade 8, and a scale score mean of 700 and a standard deviation of 50 were set for grade 10. The resulting grade level scale score means show "vertical relationship" (increasing scale score means across grades) but are not true vertical scales.

All Content Areas

The following formulae were used to compute transformation constants for the transformation of the base-grade item parameter estimates for ELA and Mathematics and each grade-level parameter estimates for Science and Social Studies from the theta metric to the scale score metric:

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

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

where:
M1 and *M2* are the transformation constants,
$SD_{ss,\,G}$ is the target standard deviation in the scale score metric for the base grade for ELA and Mathematics, and for each grade of Science and Social Studies,
$SD_{\theta,\,G}$ is the estimated standard deviation in the theta metric for the base grade for ELA and Mathematics, and for each grade of Science and Social Studies,
$\bar{\theta}_G$ is the estimated population mean in the theta metric for the base grade for ELA and Mathematics, and for each grade of Science and Social Studies,
$\bar{X}_G$ is the target mean in the scale score metric for the base grade for ELA and Mathematics, and for each grade of Science and Social Studies.

The *M1* and *M2* transformation constants were then applied to item parameter estimates in the theta metric to transform them into scale score metric using the following formulas:

$$A_{ss} = a_\theta / M1$$
$$B_{ss} = M1 * b_\theta + M2$$
$$F_{ss} = f_\theta / M1$$
$$G_{ss} = g_\theta + (f_\theta / M1) * M2$$
$$C_{ss} = c_\theta$$

where:
A_{ss} is a discrimination parameter in scale score metric for MC items,
B_{ss} is a difficulty parameter in scale score metric for MC items,
F_{ss} is a discrimination parameter in scale score metric for CR items,
G_{ss} is a difficulty level (gamma) for category m_j in scale score metric for CR items,
a_θ is a discrimination parameter in the original theta metric for MC items,
b_θ is a difficulty parameter in the original theta metric for MC items,
f_θ is a discrimination parameter in the original theta metric for CR items,
g_θ is a difficulty level (gamma) for category m_j in the original theta metric for CR items,
C_{ss} and c_θ is a guessing parameter in the original theta metric.

Table 6-7 presents the initial population mean and standard deviation estimates and the transformation constants used for scale transformation of the base grade (5) for ELA and Mathematics and each grade for Science and Social Studies.

Because the parameter estimates in theta metric were estimated for all grades (within ELA and Mathematics content areas) and were already on the same scale the same *M1* and *M2* transformation parameter constants were applied to all (grades 3–8) item parameter estimates.

6.1.4 Calibration Software

The IRT models and the student response data from the Spring 2016 test administration were used to estimate item parameters for each test. The IRT models were implemented using DRC's PARDUX software (Burket, 2002). Using marginal maximum likelihood procedures implemented with the expected maximum algorithm, PARDUX estimates parameters simultaneously for MC and CR items (Bock & Aitkin, 1981; Thissen, 1982).

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)—a program widely used for research purposes—have shown that PARDUX provides precise parameter and ability estimates and performs more efficiently than MULTILOG (Fitzpatrick, 1991). Simulation studies have also compared PARDUX with PARSCALE (Muraki & Bock, 1991) and with BIGSTEPS (Wright & Linacre, 1992). Fitzpatrick and Julian (1996) found that PARDUX provided precise parameter and ability estimates and performed more efficiently than the other programs. Extensive research with simulation data has also shown that the IRT procedures used here produce accurate vertical scaling (Yen & Burket, 1997).

6.2 Calibration and Scaling Results

The following sections describe the calibration results in terms of the estimation of item parameters, model-to-data fit, evaluation of new scales, and the standard error of measurement of the scale scores across content areas and grades.

6.2.1 IRT Item Parameters

At times when calibrating items, items may not converge, meaning the characteristics of the item are not able to be determined. When this occurs, items are suppressed from student scoring and future assessments. In Spring 2016, no convergence issues occurred for any item on the operational tests.

6.2.2 IRT Item Fit

The calibration process produces ability and item parameter estimates that can be used to predict student response patterns to each item. For example, based on the item parameter estimates for item difficulty and item discrimination, we may expect that low-ability students are less likely to answer a difficult and highly discriminating item correctly than higher-ability students. After parameters are produced, we can compare the predicted scoring patterns 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.

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

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

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

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, with other things being equal, 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, along with the associated chi-squares (Q_1), are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984).

Table 6-8 presents items that were flagged for less than optimal fit when the obtained *Z*-statistic exceeded the critical *Z*-statistic value. This table specifies the content area, grade level, item number in the calibration, item type (MC or CR), *N* size (the number of students who took this item), *Z*, and critical *Z*, as described previously. Eighteen items were flagged for poor fit for ELA and six items were flagged for Mathematics. Most of the flagged items were constructed-response items (technology-enhanced). For example, ELA item #38 in calibration was flagged because the observed *Z* of 328.95 is larger than the critical *Z* value of 191.69 based on a sample size of 71,883. This item is an operational item in grade 4 and was also administered as an off-grade level item to a sample of grade 5 students. While for many of the flagged items the observed *Z* and the critical *Z* are not very far apart indicating small misfit, it was observed that for some items the misfit was moderate (for example, item #62 in ELA) or large (for example, item #127 in Mathematics). No items were flagged for poor fit for Science or Social Studies in any grades.

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

The two 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 of the distribution. Similarly, for CR items, there are, in general, fewer students at the lower and higher score levels, which provides less information at the tails of the student distribution. Items that only show misfit at the tails 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 test.

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

6.2.3 Scale Evaluation

In this section, the results of vertical scaling of ELA and Mathematics and grade-level scaling of Science and Social Studies are described and evaluated. The vertical scale evaluation includes the following:

- examination of the student performance on the items common between adjacent grade levels,
- evaluation of the pattern of grade-to-grade growth (means),
- evaluation of grade-to-grade variability (standard deviations),
- examination of separation of scale score distributions across grades,
- setting ordinal highest and lowest obtainable scale scores, and
- evaluation of the test characteristic curves (TCC's) and standard error (SE) curves.

Only on-grade-level operational test items were used in computation of statistics used in scale evaluation.

The Science and Social Studies scale evaluation includes examination of: test characteristic curves, test standard error curves, grade-level scale score means, grade-level variability (standard deviations), reasonability of highest and lowest obtainable scale scores, and separation of grade score distributions.

The scale evaluation results are presented separately for each content area.

6.2.3.1 ELA Scale

Evaluation of Student Performance on Linking Items—Classical item analysis was performed on the data used for vertical scale development. Tables 6-9 to 6-13 present the item analysis results for on-grade level operational items and the same items administered off-grade level for ELA. The following information is provided in Tables 6-9 to 6-13: item type, item classification by test domain (or content category), item difficulty (*p*-value) on- and off-grade level, item-total test correlation on- and off-grade level, omit rates on- and off-grade level, and the number of students who took each item on-and off-grade level. The table headers are labeled as follows: PvalGx is the item *p*-value, RttGx is the item-total test correlation, OmitGx is the proportion of students who omitted the item, and NobsGx is the total number of students who took the items (x is a grade level 3, 4, 5, 6, 7, or 8 in which the item was administered).

As expected, and as demonstrated by average *p*-values of the ELA linking sets, when items from adjacent grades were administered to students in a given grade level, the students performed, on average, better on the items from the lower grade level than on the items coming from the higher grade level. When looking at the average mean item-total test correlations, the items displayed, on average, higher discrimination when administered on-grade level compared to the administration of the same items in adjacent grades. The omit rates for all items were very small.

Grade-to-Grade Growth and Variability—Table 6-14 shows the calibration sample student count, scale score means, standard deviations, and change in mean from previous grade

for ELA. As seen in Table 6-14, the ELA scale score means increase as grade level increases. The mean difference between grades is not uniform across grade levels. Most growth across grades is observed between grades 3 and 4, followed by growth between grades 4 and 5, and between grades 6 and 7, and grades 7 and 8. Least growth is observed between grades 5 and 6. The standard deviations range from 47.12 for grade 3 to 56.82 for grade 8 and show an increasing pattern across grades.

Scale Score Distribution—In addition to the evaluation of grade-to-grade growth using scale score mean changes across grades, the pattern of scale scores at the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles was examined across grades. Table 6-15 summarizes this information for ELA. Ideally, the scale score associated with each percentile will increase from grade to grade. The data in Table 6-15 show that the scale scores increase as the percentile and grade level increase, showing continuous progress upward from grades 3 through 8 at all selected percentiles.

Test Characteristic Curves—Figure 6-1 shows the TCCs for ELA tests. As shown in Figure 6-1, the ELA test TCCs for grades 3 and 4 are ordinal indicating higher difficulty of the grade 4 assessment compared to the grade 3 assessment. The grade 5 test is more difficult than the grade 3 and 4 assessments, but it also appears to be more difficult than the grade 7 assessment for students at all ability levels, and more difficult than the grade 6 and 8 assessments for higher ability students. This pattern of test difficulty is demonstrated by the grade 5 TCC being on the right side of the grade 7 TCC along the entire ability scale and by the grade 5 TCC being to the right of the grade 6 and 8 TCCs at the upper end of ability scale. The grade 6 and 8 TCCs are very close to each other or overlapping at the lower and middle part of the ability scale, indicating comparable test difficulty for students of lower and medium ability in grade 6 and grade 8. Grades 6 and 8 TCCs are crossing at the upper end of the ability scale, indicating that the grade 6 assessment may be more difficult than the grade 8 assessment for the highest ability students in both grades. The grade 7 TCC is located to the left of the grade 5, 6, and 8 TCCs indicating that ELA grade 7 assessment may be easier than the grade 5, 6, and 8 assessments for students at all ability levels.

It should be noted that while TCC ordinality is a desirable property of a vertical scale, the lack of it does not necessarily affect student scores or grade-to-grade growth interpretation. As demonstrated by the grade 3–8 mean scale scores in Table 6-14 and the pattern of scale scores at different percentiles in Table 6-15, student ELA ability increases as grade level increases at all grade levels indicating reasonable grade-to-grade growth.

Last but not least, the lack of clear ordinality of ELA TCCs may indicate that the grade 5 assessment would benefit from the addition of some easier items, while grade 7 and 8 assessments may benefit from the addition of some more difficult items. This consideration, however, must be balanced by the need to keep form difficulty comparable each year to meet the assumptions for alternate parallel forms.

Standard Error Curves—The standard error curves presented in Figure 6-2 are generally U-shaped indicating smaller errors around ability estimates roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale,

where fewer items measuring very high- and very low-achieving students are found. Overall, the standard errors around the scale score were found to be reasonable for ELA assessments (for more details see Section 6.3.1 of this report).

6.2.3.2 Mathematics Scale

Evaluation of Student Performance on Linking Items—A similar pattern to student performance on ELA linking items was observed for Mathematics (Tables 6-16 to 6-20). As expected, students in a given grade level tended to perform better, on average, on the below-grade-level Mathematics items compared to the above-grade-level items. Evaluation of the average item-total test correlations of the linking sets revealed that the items were more discriminating when administered on-grade level compared to being administered off-grade level. The exception was the average item-total test correlation set of grade 3 items administered to grade 4 students, which was higher for grade 4 students. Again, the omit rates for all items were very small.

Grade-to-Grade Growth and Variability—Table 6-21 shows the calibration sample student counts, scale score means, standard deviations, and change in mean from previous grade for Mathematics. As observed in Table 6-21, the Mathematics scale score means increase as grade level increases. The mean difference between grades is not uniform across grade levels. Most growth across grades is observed between grades 4 and 5, followed by growth between grades 3 and 4. Less growth was observed between grades 5 and 6, grades 6 and 7, and grades 7 and 8. The standard deviations range from 46.23 for grade 3 to 56.80 for grade 7 and do not show any consistent pattern across grades.

Scale Score Distribution—The pattern of scale scores at the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles was also examined across grades for Mathematics. The data in Table 6-22, which summarizes this information for Mathematics, show that the scale scores increase as the percentile and grade level increase, showing continuous progress upward from grades 3 through 8 at all selected percentiles except for the 5th percentile (and below) for grades 3 and 4. Higher scale scores for grade 3 at the very low-ability end indicate that very low-ability grade 3 students may perform better on the Mathematics assessment compared to very low-ability grade 4 students.

Test Characteristic Curves—Figure 6-3 shows the TCCs for Mathematics assessments. As observed in Figure 6-3, the TCCs for Mathematics, with the exception of grade 5 and 6 TCCs, are ordinal indicating increasing difficulty of the assessment as the grade level increases. The crossing of the grade 5 and 6 TCCs at the lower end of the ability scale indicates that the grade 5 assessment may be more difficult for lower-ability students compared to the grade 6 assessment.

Standard Error Curves—The standard error curves presented in Figure 6-4 are U-shaped (as expected), indicating smaller errors around ability estimates roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the standard errors around

the scale score were found to be reasonable for Mathematics assessments (for more details on SE magnitude refer to Section 6.3.1 of this report).

6.2.3.3 Science Scales

The Science assessments are not on vertical scales. Instead, they were established in such a way that the scale score means for higher grades are higher than the scale score means for lower grades.

Scale Score Means and Standard Deviations—Table 6-23 shows the calibration sample student count, scale score means and standard deviations for Science grades 4 and 8. The grade 4 scale score mean is approximately 400, and the grade 8 scale score mean is approximately 600. The standard deviation is about 51 scale score points for each grade.

Scale Score Distributions—Table 6-24 presents Science scale score distribution at selected percentiles. As expected, the scale scores increase as the percentile rank increases, showing increasing student ability along the scale for both grade levels.

Test Characteristic Curves—Although the Science assessments are not vertically scaled, the TCCs for grades 4 and 8 are presented together in Figure 6-5 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a student's ability increases. The grade 4 and 8 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

Standard Error Curves—Figure 6-6 shows Science test SE curves for grades 4 and 8. The SE curves are U-shaped, indicating smaller errors around ability estimates approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the standard errors around the scale score were found to be reasonable for Science assessments (for more details on SE magnitude refer to Section 6.3.1 of this report).

6.2.3.4 Social Studies Scales

The Social Studies assessments are not on vertical scales. Instead, they were established in such a way that the scale score means for higher grades are higher than the scale score means for lower grades.

Scale Score Means and Standard Deviations—Table 6-25 shows the calibration sample student count, scale score means, and standard deviations for Social Studies grades 4, 8, and 10. The grade 4 scale score mean is approximately 400; the grade 8 scale score mean is approximately 600; and the grade 10 scale score mean is approximately 700. The standard deviations range from approximately 50 to approximately 51 scale score points for Social Studies assessments.

Scale Score Distributions—Table 6-26 presents Social Studies scale score distribution at selected percentiles. As expected, the scale scores increase as the percentile rank increases, showing increasing student ability along the scale for the three grade levels.

Test Characteristic Curves—As with Science, the Social Studies assessments are not vertically scaled. However, the TCCs for grades 4, 8, and 10 are presented together in Figure 6-7 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a student's ability increases. The grade 4, 8, and 10 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

Standard Error Curves—Figure 6-8 shows Science test SE curves for grades 4, 8, and 10. The SE curves are U-shaped, showing smaller errors around ability estimates approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the standard errors around the scale score were found to be reasonable for Social Studies assessments (for more details on SE magnitude refer to Section 6.3.1 of this report).

6.3 Deriving Scale Scores in the Wisconsin Forward Exam

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 in the test (such as item difficulty).

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. IRT models expect 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; CR items are handled slightly differently but follow 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 his or her 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 correct response, respectively. For an MC item, the

probability model can be denoted as $P(u_i=1|\theta)$. That is, *P* is an estimation of the probability that a student with an ability value 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 ($u_1u_2\cdots u_n$) on a test with some number of items (*n* items), the "likelihood function," or the probability that a student with a given ability value (θ) would produce this particular response pattern, is given by

$$P(u_1u_2\cdots u_n \mid \theta) = \prod_{i=1}^{n} P(u_i \mid \theta). \qquad (2)$$

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 (2) and it assigns an ability estimate (θ_0) as the test score for the student with the response pattern $u_1u_2\cdots u_n$. In other words, the scale score is the most likely, or most probable, estimate of student ability produced in a context where item parameters are known and based on all of 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:

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:

Table 6-B Example of Item Response Pattern

Student	Response Pattern ($u_1 u_2 \cdots u_n$)	Raw Score	Item-Pattern Score
Pam	10000110010100000000000000101	7	140
Craig	10101010101010101010101010101	15	246
Vicki	01010101010101010101010101010	15	266
Tom	00110011001100110011001101101	15	259
Evan	11001100110011001100110010010	15	265
Dan	11111111111111111111111011111	29	379

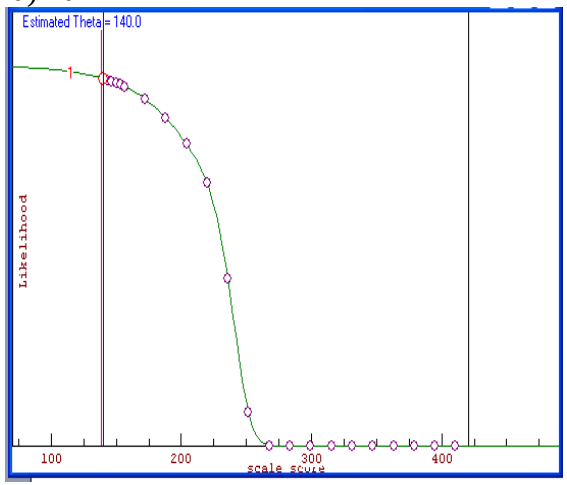
The first student, Pam, answered seven of the items correctly and obtained a scale score of 140, which is equal to the lowest point on the score range, called the "lowest obtainable scale score," or LOSS. The next four students each answered 15 out of 30 items correctly, but the response pattern of each 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 together 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 below shows the probability of each ability estimate (or scale score) for the six examinees. The total scale score range for the ELA 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 that the ability estimate actually reflects the examinee's ability level.

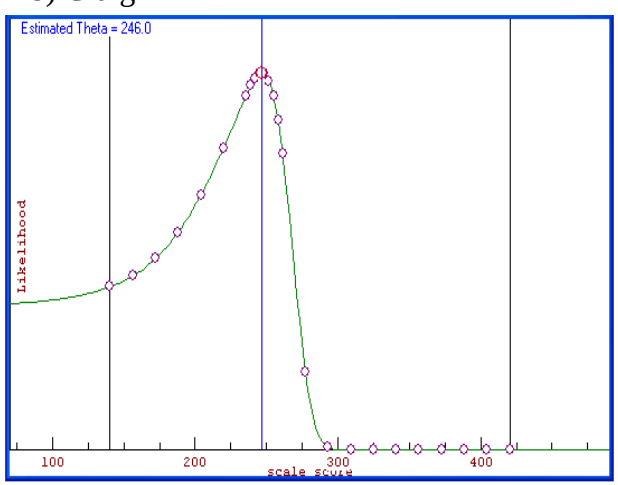
As indicated above, scale scores are the most likely, or the 'maximum likelihood,' estimates of examinee ability. As can be observed for Vicki, Tom, and Evan, scores that are plus or minus only a few scale score points are markedly less likely estimates of their ability. The same is true for Craig and 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).

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

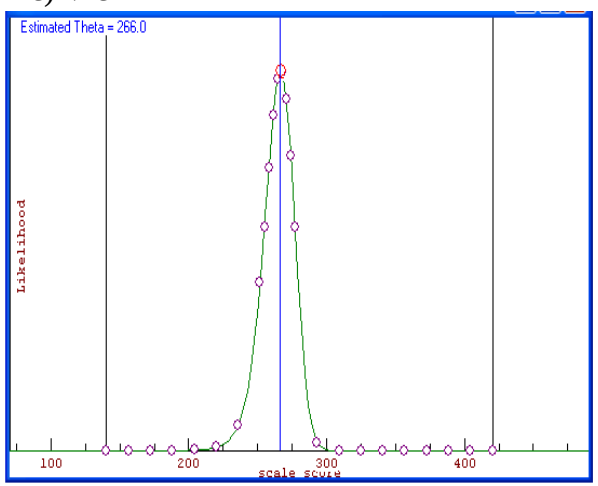
a) Pam



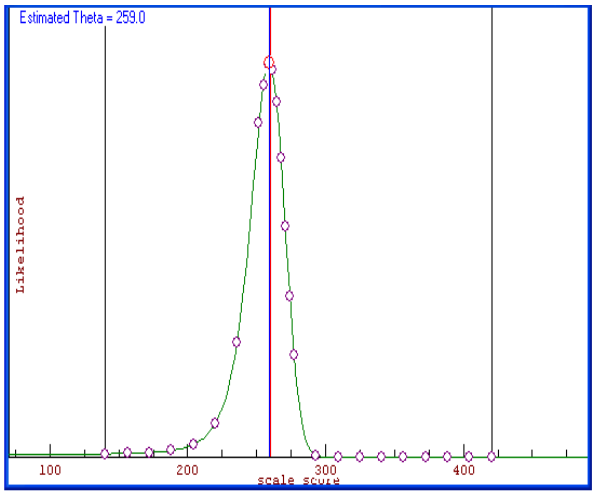
b) Craig



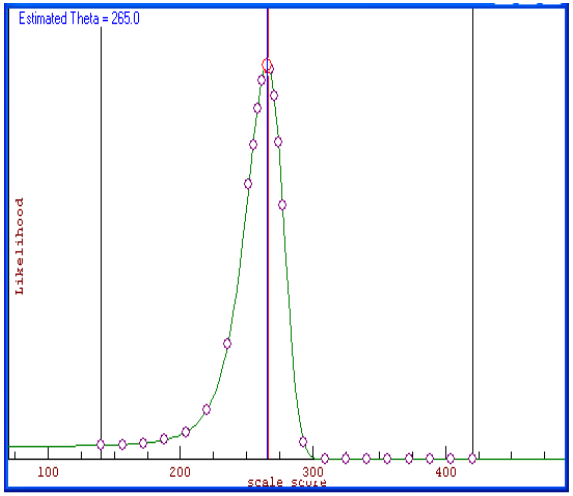
c) Vicki



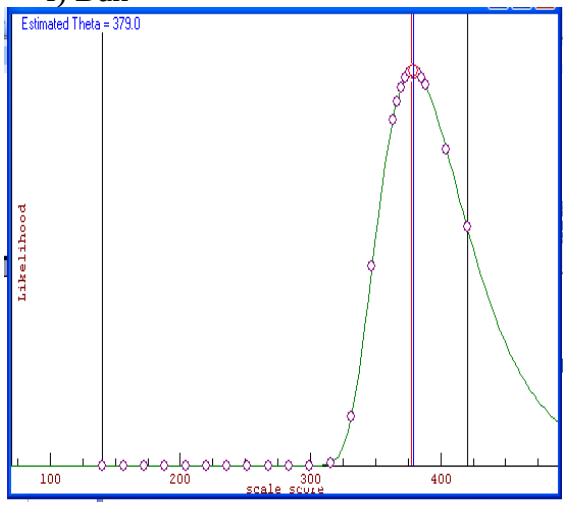
d) Tom



e) Evan



f) Dan



*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.

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 present purposes.

Item-pattern scoring generally produces more accurate scores for individual students. Specifically, it produces a smaller standard error of measurement (SEM) across the scale score range for a given test compared to number-correct scoring. The smaller the SEM, 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 approximately a 15% to 20% increase 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 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.

The procedures applied here are consistent with student scoring in prior Wisconsin Knowledge and Concepts Examinations. 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 and Hambleton (1997). For additional information on the technical details of item-pattern scoring, readers can also refer to Yen & Candell (1991).

6.3.1 Standard Error of Measurement

One way of characterizing the reliability of a reported test score is by examining the standard error associated with the score. An observed score should not be regarded as an absolute value but as a point within a range that with a certain degree of probability includes a student's true score. The SEM 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). The SEM 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 68% of the time a student's score obtained from a single testing will fall within one SEM of that student's true score and that 95% of the time the obtained score will fall within two standard errors of the true score.

Standard 2.13 of the AERA, APA, & NCME (2014) *Standards* that states the following:

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

The SEM of the scale scores in the Spring 2016 Wisconsin Forward Exam is displayed graphically for each grade and content area in Figures 6-2 (for ELA), 6-4 (for Mathematics), 6-6 (for Science), and 6-8 (for Social Studies). The SEM provided is based on item-pattern scoring. Each SEM curve is plotted as a function of the scale scores. These figures show the scale score range within which measurement is most accurate. The figures also show that extreme scale scores have more measurement error than scores in the middle of the distribution. Scale scores in the high or low extremes of the student distribution are less precise than those in the middle of the distribution because there tend to be fewer test items in these score areas and fewer students. The lower and upper limits of the scale, referred to as the lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS), are the starting scale score and the last scale score in these figures. LOSS and HOSS are further discussed in the next section.

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 2016 Wisconsin Forward Exam. However, scoring tables showing a rough relationship between raw score, scale score, and SEM can be produced, and they are provided in Tables 6-27 through 6-43.

6.3.2 LOSS and HOSS

As has been 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 called the LOSS and the HOSS.

Table 6-44 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. An accumulation of a high proportion of students in 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 that they increase as the grade level increases. Setting increasing LOSS 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 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.

In most grades and content areas, the percentage of students at the LOSS and HOSS was small: less than 1%. However, in some grades and content areas the LOSS percentages were larger. In Mathematics, all grades, except grade 3 had more than 1% of students at the LOSS (grade 4–3.35%, grade 5–1.13%, grade 6–1.33%, grade 7–2.98%, and grade 8–3.90%). These percentages at the LOSS indicate that the Mathematics assessments were difficult for some

students and that they can be considered as a point of reference when developing future forms. The percentage at the LOSS in these grades may be reduced in future years by including some additional items that are less difficult. The percentage of students scoring at the HOSS is similar: in most grades and content areas, the percentage was small, although in two cases the percentage was larger. In particular, more than 1% of students obtained the HOSS in Science grade 8 (1.25%) and Social Studies grade 4 (1.54%). The percentage scoring at the HOSS may be reduced by including some additional difficult items in these grades or by including more items on the test.

Table 6-1 English Language Arts Calibration Sample Demographics Compared to Public School Population

Grade 3	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	63842		61120		
Gender					
Male	32645	51.10%	31319	51.20%	-0.10%
Female	31197	48.90%	29801	48.80%	0.10%
Race/Ethnicity					
White	43000	67.40%	42388	69.40%	-2.00%
Black	7017	11.00%	5877	9.62%	1.38%
Hispanic	8264	12.90%	7451	12.20%	0.70%
Asian/Pacific Islander	2487	3.90%	2420	3.96%	-0.06%
American Indian	760	1.19%	757	1.24%	-0.05%
Other	2314	3.62%	2227	3.64%	-0.02%
LEP					
No	58009	91.30%	55702	91.10%	0.20%
Yes	5528	8.70%	5418	8.86%	-0.16%
Disability					
No	56100	88.30%	53659	87.80%	0.50%
Yes	7435	11.70%	7461	12.20%	-0.50%
SES Disadvantaged					
No	34741	54.80%	34790	56.90%	-2.10%
Yes	28678	45.20%	26330	43.10%	2.10%
Grade 4	**N**	**%**	**N**	**%**	**%**
All Students	62341		59776		
Gender					
Male	31895	51.20%	30651	51.30%	-0.10%
Female	30446	48.80%	29125	48.70%	0.10%
Race/Ethnicity					
White	42006	67.40%	41474	69.40%	-2.00%
Black	6815	10.90%	5760	9.64%	1.26%
Hispanic	8184	13.10%	7361	12.30%	0.80%
Asian/Pacific Islander	2496	4.00%	2411	4.03%	-0.03%
American Indian	756	1.21%	755	1.26%	-0.05%
Other	2084	3.34%	2015	3.37%	-0.03%
LEP					
No	57291	92.40%	55192	92.30%	0.10%
Yes	4716	7.61%	4584	7.67%	-0.06%
Disability					
No	54498	87.90%	52244	87.40%	0.50%
Yes	7517	12.10%	7532	12.60%	-0.50%
SES Disadvantaged					
No	34085	55.10%	34144	57.10%	-2.00%
Yes	27822	44.90%	25632	42.90%	2.00%

Table 6-1 English Language Arts Calibration Sample Demographics Compared to Public School Population (cont.)

Grade 5	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	62081		59662		
Gender					
Male	31670	51.00%	30543	51.20%	-0.20%
Female	30411	49.00%	29119	48.80%	0.20%
Race/Ethnicity					
White	42814	69.00%	42259	70.80%	-1.80%
Black	6564	10.60%	5533	9.27%	1.33%
Hispanic	7618	12.30%	6917	11.60%	0.70%
Asian/Pacific Islander	2431	3.92%	2363	3.96%	-0.04%
American Indian	761	1.23%	764	1.28%	-0.05%
Other	1892	3.05%	1826	3.06%	-0.01%
LEP					
No	58548	94.70%	56468	94.60%	0.10%
Yes	3262	5.28%	3194	5.35%	-0.07%
Disability					
No	54374	88.00%	52200	87.50%	0.50%
Yes	7436	12.00%	7462	12.50%	-0.50%
SES Disadvantaged					
No	35322	57.30%	35350	59.30%	-2.00%
Yes	26370	42.70%	24312	40.70%	2.00%
Grade 6	**N**	**%**	**N**	**%**	**%**
All Students	62432		60164		
Gender					
Male	32019	51.30%	30884	51.30%	0.00%
Female	30413	48.70%	29280	48.70%	0.00%
Race/Ethnicity					
White	43718	70.00%	43225	71.80%	-1.80%
Black	6290	10.10%	5384	8.95%	1.15%
Hispanic	7474	12.00%	6738	11.20%	0.80%
Asian/Pacific Islander	2401	3.85%	2331	3.87%	-0.02%
American Indian	735	1.18%	734	1.22%	-0.04%
Other	1813	2.90%	1752	2.91%	-0.01%
LEP					
No	59532	95.70%	57577	95.70%	0.00%
Yes	2653	4.27%	2587	4.30%	-0.03%
Disability					
No	54799	88.10%	52733	87.60%	0.50%
Yes	7388	11.90%	7431	12.40%	-0.50%
SES Disadvantaged					
No	36471	58.70%	36438	60.60%	-1.90%
Yes	25615	41.30%	23726	39.40%	1.90%

Table 6-1 English Language Arts Calibration Sample Demographics Compared to Public School Population (cont.)

Grade 7	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	61739		59539		
Gender					
Male	31652	51.30%	30623	51.40%	-0.10%
Female	30087	48.70%	28916	48.60%	0.10%
Race/Ethnicity					
White	43654	70.70%	43153	72.50%	-1.80%
Black	6157	9.97%	5281	8.87%	1.10%
Hispanic	7314	11.80%	6624	11.10%	0.70%
Asian/Pacific Islander	2282	3.70%	2219	3.73%	-0.03%
American Indian	729	1.18%	729	1.22%	-0.04%
Other	1602	2.59%	1533	2.57%	0.02%
LEP					
No	58948	95.90%	57081	95.90%	0.00%
Yes	2510	4.08%	2458	4.13%	-0.05%
Disability					
No	54217	88.20%	52218	87.70%	0.50%
Yes	7246	11.80%	7321	12.30%	-0.50%
SES Disadvantaged					
No	36912	60.20%	36902	62.00%	-1.80%
Yes	24453	39.80%	22637	38.00%	1.80%
Grade 8	**N**	**%**	**N**	**%**	**%**
All Students	60648		59006		
Gender					
Male	30945	51.00%	30219	51.20%	-0.20%
Female	29703	49.00%	28787	48.80%	0.20%
Race/Ethnicity					
White	43080	71.00%	42945	72.80%	-1.80%
Black	6130	10.10%	5322	9.02%	1.08%
Hispanic	6957	11.50%	6331	10.70%	0.80%
Asian/Pacific Islander	2196	3.62%	2160	3.66%	-0.04%
American Indian	746	1.23%	751	1.27%	-0.04%
Other	1538	2.54%	1497	2.54%	0.00%
LEP					
No	57976	95.90%	56578	95.90%	0.00%
Yes	2457	4.07%	2428	4.11%	-0.04%
Disability					
No	53257	88.10%	51726	87.70%	0.40%
Yes	7178	11.90%	7280	12.30%	-0.40%
SES Disadvantaged					
No	36592	60.60%	36875	62.50%	-1.90%
Yes	23747	39.40%	22131	37.50%	1.90%

Table 6-2 Mathematics Calibration Sample Demographics Compared to Public School Population

Grade 3	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	62648		61220		
Gender					
Male	32051	51.20%	31375	51.20%	0.00%
Female	30597	48.80%	29845	48.80%	0.00%
Race/Ethnicity					
White	42247	67.40%	42369	69.20%	-1.80%
Black	6774	10.80%	5882	9.61%	1.19%
Hispanic	8187	13.10%	7530	12.30%	0.80%
Asian/Pacific Islander	2443	3.90%	2456	4.01%	-0.11%
American Indian	752	1.20%	756	1.23%	-0.03%
Other	2245	3.58%	2227	3.64%	-0.06%
LEP					
No	56785	91.10%	55672	90.90%	0.20%
Yes	5542	8.89%	5548	9.06%	-0.17%
Disability					
No	55037	88.30%	53759	87.80%	0.50%
Yes	7285	11.70%	7461	12.20%	-0.50%
SES Disadvantaged					
No	34127	54.90%	34836	56.90%	-2.00%
Yes	28072	45.10%	26384	43.10%	2.00%
Grade 4	**N**	**%**	**N**	**%**	**%**
All Students	61702		59855		
Gender					
Male	31553	51.10%	30687	51.30%	-0.20%
Female	30149	48.90%	29168	48.70%	0.20%
Race/Ethnicity					
White	41609	67.40%	41479	69.30%	-1.90%
Black	6683	10.80%	5765	9.63%	1.17%
Hispanic	8101	13.10%	7406	12.40%	0.70%
Asian/Pacific Islander	2500	4.05%	2439	4.07%	-0.02%
American Indian	753	1.22%	753	1.26%	-0.04%
Other	2056	3.33%	2013	3.36%	-0.03%
LEP					
No	56617	92.30%	55178	92.20%	0.10%
Yes	4748	7.74%	4677	7.81%	-0.07%
Disability					
No	53936	87.90%	52322	87.40%	0.50%
Yes	7429	12.10%	7533	12.60%	-0.50%
SES Disadvantaged					
No	33766	55.10%	34175	57.10%	-2.00%
Yes	27492	44.90%	25680	42.90%	2.00%

Table 6-2 Mathematics Calibration Sample Demographics Compared to Public School Population (cont.)

Grade 5	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	60884		59733		
Gender					
Male	31043	51.00%	30579	51.20%	-0.20%
Female	29841	49.00%	29154	48.80%	0.20%
Race/Ethnicity					
White	42004	69.00%	42255	70.70%	-1.70%
Black	6324	10.40%	5537	9.27%	1.13%
Hispanic	7520	12.40%	6961	11.70%	0.70%
Asian/Pacific Islander	2427	3.99%	2395	4.01%	-0.02%
American Indian	749	1.23%	761	1.27%	-0.04%
Other	1860	3.05%	1824	3.05%	0.00%
LEP					
No	57313	94.60%	56454	94.50%	0.10%
Yes	3288	5.43%	3279	5.49%	-0.06%
Disability					
No	53350	88.00%	52270	87.50%	0.50%
Yes	7247	12.00%	7463	12.50%	-0.50%
SES Disadvantaged					
No	34679	57.30%	35381	59.20%	-1.90%
Yes	25799	42.70%	24352	40.80%	1.90%
Grade 6	**N**	**%**	**N**	**%**	**%**
All Students	61295		60220		
Gender					
Male	31420	51.30%	30911	51.30%	0.00%
Female	29875	48.70%	29309	48.70%	0.00%
Race/Ethnicity					
White	43044	70.20%	43220	71.80%	-1.60%
Black	6031	9.84%	5385	8.94%	0.90%
Hispanic	7346	12.00%	6770	11.20%	0.80%
Asian/Pacific Islander	2385	3.89%	2358	3.92%	-0.03%
American Indian	729	1.19%	736	1.22%	-0.03%
Other	1759	2.87%	1751	2.91%	-0.04%
LEP					
No	58387	95.60%	57554	95.60%	0.00%
Yes	2656	4.35%	2666	4.43%	-0.08%
Disability					
No	53851	88.20%	52801	87.70%	0.50%
Yes	7192	11.80%	7419	12.30%	-0.50%
SES Disadvantaged					
No	35939	59.00%	36456	60.50%	-1.50%
Yes	24993	41.00%	23764	39.50%	1.50%

Table 6-2 Mathematics Calibration Sample Demographics Compared to Public School Population (cont.)

Grade 7	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	60838		59600		
Gender					
Male	31221	51.30%	30658	51.40%	-0.10%
Female	29617	48.70%	28942	48.60%	0.10%
Race/Ethnicity					
White	43132	70.90%	43138	72.40%	-1.50%
Black	5958	9.79%	5289	8.87%	0.92%
Hispanic	7176	11.80%	6672	11.20%	0.60%
Asian/Pacific Islander	2278	3.74%	2247	3.77%	-0.03%
American Indian	721	1.19%	728	1.22%	-0.03%
Other	1571	2.58%	1526	2.56%	0.02%
LEP					
No	58032	95.90%	57055	95.70%	0.20%
Yes	2510	4.15%	2545	4.27%	-0.12%
Disability					
No	53452	88.30%	52290	87.70%	0.60%
Yes	7093	11.70%	7310	12.30%	-0.60%
SES Disadvantaged					
No	36532	60.40%	36931	62.00%	-1.60%
Yes	23914	39.60%	22669	38.00%	1.60%
Grade 8	**N**	**%**	**N**	**%**	**%**
All Students	60620		59076		
Gender					
Male	30958	51.10%	30248	51.20%	-0.10%
Female	29662	48.90%	28828	48.80%	0.10%
Race/Ethnicity					
White	43170	71.20%	42935	72.70%	-1.50%
Black	6007	9.91%	5339	9.04%	0.87%
Hispanic	6948	11.50%	6368	10.80%	0.70%
Asian/Pacific Islander	2212	3.65%	2186	3.70%	-0.05%
American Indian	747	1.23%	750	1.27%	-0.04%
Other	1535	2.53%	1498	2.54%	-0.01%
LEP					
No	57869	95.80%	56572	95.80%	0.00%
Yes	2528	4.19%	2504	4.24%	-0.05%
Disability					
No	53279	88.20%	51785	87.70%	0.50%
Yes	7119	11.80%	7291	12.30%	-0.50%
SES Disadvantaged					
No	36666	60.80%	36898	62.50%	-1.70%
Yes	23632	39.20%	22178	37.50%	1.70%

Table 6-3 Science Calibration Sample Demographics Compared to Public School Population

Grade 4	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	52631		59832		
Gender					
Male	26932	51.20%	30668	51.30%	-0.10%
Female	25699	48.80%	29164	48.70%	0.10%
Race/Ethnicity					
White	36192	68.80%	41478	69.30%	-0.50%
Black	5081	9.65%	5753	9.62%	0.03%
Hispanic	6877	13.10%	7395	12.40%	0.70%
Asian/Pacific Islander	2079	3.95%	2438	4.07%	-0.12%
American Indian	686	1.30%	755	1.26%	0.04%
Other	1716	3.26%	2013	3.36%	-0.10%
LEP					
No	48463	92.60%	55164	92.20%	0.40%
Yes	3890	7.43%	4668	7.80%	-0.37%
Disability					
No	46054	88.00%	52300	87.40%	0.60%
Yes	6295	12.00%	7532	12.60%	-0.60%
SES Disadvantaged					
No	28683	54.90%	34167	57.10%	-2.20%
Yes	23588	45.10%	25665	42.90%	2.20%
Grade 8	**N**	**%**	**N**	**%**	**%**
All Students	53203		59000		
Gender					
Male	27182	51.10%	30216	51.20%	-0.10%
Female	26021	48.90%	28784	48.80%	0.10%
Race/Ethnicity					
White	38540	72.40%	42911	72.70%	-0.30%
Black	4840	9.10%	5304	8.99%	0.11%
Hispanic	5901	11.10%	6355	10.80%	0.30%
Asian/Pacific Islander	1948	3.66%	2186	3.71%	-0.05%
American Indian	651	1.22%	750	1.27%	-0.05%
Other	1319	2.48%	1494	2.53%	-0.05%
LEP					
No	50925	96.10%	56497	95.80%	0.30%
Yes	2085	3.93%	2503	4.24%	-0.31%
Disability					
No	46827	88.30%	51735	87.70%	0.60%
Yes	6185	11.70%	7265	12.30%	-0.60%
SES Disadvantaged					
No	32237	60.90%	36861	62.50%	-1.60%
Yes	20700	39.10%	22139	37.50%	1.60%

Table 6-4 Social Studies Calibration Sample Demographics Compared to Public School Population

Grade 4	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	48902		59817		
Gender					
Male	25079	51.30%	30654	51.20%	0.10%
Female	23823	48.70%	29163	48.80%	-0.10%
Race/Ethnicity					
White	33918	69.40%	41472	69.30%	0.10%
Black	4649	9.51%	5740	9.60%	-0.09%
Hispanic	6219	12.70%	7400	12.40%	0.30%
Asian/Pacific Islander	1901	3.89%	2436	4.07%	-0.18%
American Indian	644	1.32%	754	1.26%	0.06%
Other	1571	3.21%	2015	3.37%	-0.16%
LEP					
No	45206	92.90%	55149	92.20%	0.70%
Yes	3446	7.08%	4668	7.80%	-0.72%
Disability					
No	42751	87.90%	52305	87.40%	0.50%
Yes	5895	12.10%	7512	12.60%	-0.50%
SES Disadvantaged					
No	26944	55.50%	34171	57.10%	-1.60%
Yes	21629	44.50%	25646	42.90%	1.60%
Grade 8	**N**	**%**	**N**	**%**	**%**
All Students	51809		59024		
Gender					
Male	26526	51.20%	30230	51.20%	0.00%
Female	25283	48.80%	28794	48.80%	0.00%
Race/Ethnicity					
White	37487	72.40%	42927	72.70%	-0.30%
Black	4662	9.00%	5307	8.99%	0.01%
Hispanic	5810	11.20%	6358	10.80%	0.40%
Asian/Pacific Islander	1906	3.68%	2187	3.71%	-0.03%
American Indian	639	1.23%	751	1.27%	-0.04%
Other	1302	2.51%	1494	2.53%	-0.02%
LEP					
No	49564	96.00%	56519	95.80%	0.20%
Yes	2056	3.98%	2505	4.24%	-0.26%
Disability					
No	45658	88.40%	51758	87.70%	0.70%
Yes	5966	11.60%	7266	12.30%	-0.70%
SES Disadvantaged					
No	31685	61.50%	36880	62.50%	-1.00%
Yes	19866	38.50%	22144	37.50%	1.00%

Table 6-4 Social Studies Calibration Sample Demographics Compared to Public School Population (cont.)

Grade 10	Calibration Sample		Public School Population		Difference
	N	%	N	%	%
All Students	43997		62203		
Gender					
Male	22330	50.80%	31744	51.00%	-0.20%
Female	21667	49.20%	30459	49.00%	0.20%
Race/Ethnicity					
White	32449	73.80%	46882	75.40%	-1.60%
Black	4105	9.33%	5103	8.20%	1.13%
Hispanic	4596	10.40%	5863	9.43%	0.97%
Asian/Pacific Islander	1436	3.26%	2287	3.68%	-0.42%
American Indian	451	1.03%	705	1.13%	-0.10%
Other	960	2.18%	1363	2.19%	-0.01%
LEP					
No	42657	97.50%	60471	97.20%	0.30%
Yes	1115	2.55%	1732	2.78%	-0.23%
Disability					
No	38955	89.00%	55185	88.70%	0.30%
Yes	4817	11.00%	7018	11.30%	-0.30%
SES Disadvantaged					
No	28632	65.50%	41631	66.90%	-1.40%
Yes	15068	34.50%	20572	33.10%	1.40%

Table 6-5 English Language Arts Population Ability Estimates across Multiple Groups on All Items

Estimates	Grade					
	3	4	5 (base)	6	7	8
N-count	63842	62341	62081	62432	61739	60648
Mean theta	-1.13	-0.64	-0.24	0.01	0.32	0.63
Theta SD	1.06	1.13	1.13	1.15	1.22	1.28

Table 6-6 Mathematics Population Ability Estimates across Multiple Groups on All Items

Estimates	Grade					
	3	4	5 (base)	6	7	8
N-count	62648	61702	60884	61295	60838	60620
Mean theta	-1.24	-0.87	-0.27	0.00	0.27	0.54
Theta SD	1.00	1.36	1.12	1.22	1.50	1.58

Table 6-7 Scale Transformation Constants

Content Area and Grade	Target Scale Properties in Scale Score Metric		Estimated Population Ability in Theta Metric		Transformation Constants	
	Mean	SD	Mean	SD	M1	M2
English Language Arts 5	600	50	-0.24	1.14	43.74453	610.49869
Math 5	600	50	-0.26	1.08	46.46840	612.08178
Science 4	400	50	-0.04	1.18	42.55319	401.70213
Science 8	600	50	-0.09	1.26	39.55696	603.56013
Social Studies 4	400	50	-0.13	1.24	40.19293	405.22508
Social Studies 8	600	50	-0.02	1.18	42.22973	600.84459
Social Studies 10	700	50	-0.09	1.17	42.88165	703.85935

Table 6-8 Item Flagged Based on Yen's Q1

Content	Grade	Item Number in Calibration	Type	N	Z	Critical Z
ELA	4 (5)	38*	CR	71883	328.95	191.69
	4 (3)	62*	CR	74078	607.96	197.54
	4 (3,5)	69*	CR	83656	620.01	223.08
	5	71	CR	62080	177.40	165.55
	5 (4)	84*	CR	71550	530.13	190.80
	5	91	CR	62080	536.17	165.55
	5 (6)	92*	CR	71837	207.72	191.57
	6 (7)	105*	CR	72131	335.69	192.35
	6 (5)	121*	CR	72010	353.11	192.03
	6	127	CR	62432	241.70	166.49
	6	131	CR	62432	174.86	166.49
	6 (7)	132*	CR	72137	664.78	192.37
	7 (6)	145*	CR	71438	280.17	190.50
	7 (8)	154*	MC	73005	310.76	194.68
	7 (8)	156*	CR	73005	511.74	194.68
	8	183	CR	60640	405.86	161.71
	8 (7)	189*	CR	70274	411.89	187.40
	8	191	CR	60640	183.10	161.71
Math	3	33	CR	62624	303.98	167.00
	3	34	CR	62624	297.86	167.00
	5 (6)	127*	CR	81049	1569.17	216.13
	6 (7)	136*	MC	81411	1926.82	217.10
	6	156	CR	61288	169.75	163.43
	7	207	CR	60827	180.64	162.21

Note: An asterisk (*) indicates a linking item that was administered on- and off-grade level. In the column "Grade" the off-grade level is indicated in parentheses.

Table 6-9 English Language Arts Grade 3 vs. Grade 4 Vertical Linking Item Statistics

Item Grade	Item Type	Domain	Item Statistics in Administration Grade							
			PvalG3	PvalG4	RttG3	RttG4	OmitG3	OmitG4	NobsG3	NobsG4
3	MC	Writing	0.66	0.75	0.31	0.24	0.00	0.00	63786	9421
3	TE	Reading	0.47	0.58	0.57	0.53	0.00	0.00	63685	9411
3	ESR	Reading	0.64	0.75	0.54	0.47	0.00	0.00	63739	9414
3	MC	Reading	0.47	0.56	0.33	0.27	0.00	0.00	63671	9410
3	MC	Reading	0.57	0.67	0.31	0.23	0.00	0.00	63689	9405
3	MC	Reading	0.41	0.55	0.32	0.28	0.00	0.00	63602	9393
3	MC	Reading	0.39	0.43	0.24	0.16	0.00	0.00	63584	9391
3	ESR	Reading	0.41	0.61	0.54	0.49	0.00	0.00	63657	9393
3	TE	Reading	0.70	0.81	0.53	0.43	0.01	0.00	63552	9380
3	ESR	Listening	0.73	0.81	0.46	0.43	0.00	0.00	63778	9420
3	MC	Listening	0.69	0.81	0.49	0.43	0.00	0.00	63734	9416
3	MC	Listening	0.71	0.77	0.42	0.37	0.00	0.00	63730	9419
3		AVERAGE	0.58	0.68	0.42	0.36	0.00	0.00		
4	TE	Writing	0.94	0.94	0.37	0.42	0.00	0.00	11716	62250
4	MC	Reading	0.69	0.80	0.32	0.41	0.00	0.00	11701	62235
4	MC	Reading	0.51	0.62	0.29	0.39	0.00	0.00	11700	62197
4	MC	Reading	0.77	0.87	0.45	0.43	0.00	0.00	11703	62253
4	TE	Reading	0.46	0.57	0.39	0.49	0.00	0.00	11701	62240
4	MC	Reading	0.44	0.51	0.30	0.37	0.00	0.00	11672	62171
4	TE	Reading	0.68	0.72	0.46	0.49	0.00	0.00	11680	62187
4	TE	Reading	0.19	0.32	0.33	0.49	0.02	0.01	11441	61632
4	MC	Reading	0.47	0.57	0.40	0.50	0.00	0.00	11664	62177
4	TE	Writing	0.82	0.83	0.32	0.35	0.00	0.00	11692	62266
4	TE	Listening	0.57	0.66	0.39	0.48	0.00	0.00	11690	62292
4	ESR	Listening	0.66	0.72	0.30	0.41	0.00	0.00	11720	62261
4	MC	Listening	0.54	0.60	0.33	0.38	0.00	0.00	11712	62246
4	MC	Listening	0.55	0.60	0.34	0.44	0.00	0.00	11718	62261
4	MC	Listening	0.50	0.55	0.28	0.32	0.00	0.00	11692	62269
4	MC	Listening	0.49	0.58	0.23	0.35	0.00	0.00	11686	62281
4		AVERAGE	0.61	0.68	0.34	0.42	0.00	0.00		

Table 6-10 English Language Arts Grade 4 vs. Grade 5 Vertical Linking Item Statistics

Item Grade	Item Type	Domain	Item Statistics in Administration Grade							
			PvalG 4	PvalG 5	RttG 4	RttG 5	OmitG 4	OmitG 5	NobsG 4	NobsG 5
4	MC	Writing	0.60	0.66	0.20	0.10	0.00	0.00	62258	9565
4	MC	Reading	0.48	0.57	0.28	0.25	0.00	0.00	62267	9526
4	MC	Reading	0.30	0.32	-0.03	0.00	0.00	0.00	62239	9529
4	TE	Reading	0.62	0.70	0.52	0.44	0.00	0.00	62281	9526
4	ESR	Reading	0.56	0.65	0.57	0.42	0.00	0.00	62280	9532
4	TE	Writing	0.62	0.76	0.44	0.42	0.00	0.00	62200	9528
4	TE	Listening	0.66	0.73	0.48	0.37	0.00	0.00	62292	9528
4	ESR	Listening	0.72	0.79	0.41	0.28	0.00	0.00	62261	9562
4	MC	Listening	0.60	0.72	0.38	0.31	0.00	0.00	62246	9560
4	MC	Listening	0.60	0.73	0.44	0.40	0.00	0.00	62261	9561
4	MC	Listening	0.55	0.64	0.32	0.24	0.00	0.00	62269	9527
4	MC	Listening	0.58	0.67	0.35	0.34	0.00	0.00	62281	9523
4		AVERAGE	0.59	0.68	0.36	0.30	0.00	0.00		
5	MC	Writing	0.72	0.75	0.38	0.44	0.00	0.00	9396	61962
5	MC	Writing	0.31	0.39	0.06	0.19	0.00	0.00	9456	61950
5	ESR	Reading	0.18	0.22	0.18	0.28	0.00	0.00	9458	61970
5	MC	Reading	0.38	0.44	0.13	0.20	0.00	0.00	9447	61936
5	MC	Reading	0.56	0.61	0.28	0.40	0.00	0.00	9449	61935
5	ESR	Reading	0.39	0.42	0.28	0.33	0.00	0.00	9453	61954
5	ESR	Listening	0.30	0.41	0.28	0.44	0.00	0.00	9458	62020
5	MC	Listening	0.68	0.70	0.37	0.45	0.00	0.00	9395	61960
5	MC	Listening	0.64	0.67	0.40	0.48	0.00	0.00	9395	61968
5	MC	Listening	0.47	0.54	0.20	0.31	0.00	0.00	9453	61993
5	MC	Listening	0.29	0.32	0.14	0.25	0.00	0.00	9457	62024
5	MS	Listening	0.53	0.57	0.42	0.49	0.00	0.00	9401	61958
5		AVERAGE	0.43	0.48	0.26	0.36	0.00	0.00		

Table 6-11 English Language Arts Grade 5 vs. Grade 6 Vertical Linking Item Statistics

Item Grade	Item Type	Domain	Item Statistics in Administration Grade							
			PvalG5	PvalG6	RttG5	RttG6	OmitG5	OmitG6	NobsG5	NobsG6
5	ESR	Writing	0.16	0.14	0.17	0.15	0.00	0.00	62009	9743
5	MC	Writing	0.70	0.72	0.29	0.24	0.00	0.00	62016	9775
5	MC	Reading	0.80	0.83	0.38	0.32	0.00	0.00	62010	9762
5	TE	Reading	0.28	0.29	0.18	0.08	0.00	0.00	61945	9751
5	MC	Reading	0.77	0.80	0.39	0.32	0.00	0.00	62004	9762
5	ESR	Reading	0.82	0.84	0.48	0.42	0.00	0.00	62030	9768
5	ESR	Listening	0.41	0.50	0.44	0.37	0.00	0.00	62020	9746
5	MC	Listening	0.70	0.79	0.45	0.37	0.00	0.00	61960	9770
5	MC	Listening	0.67	0.77	0.48	0.41	0.00	0.00	61968	9773
5	MC	Listening	0.54	0.61	0.31	0.26	0.00	0.00	61993	9741
5	MC	Listening	0.32	0.36	0.25	0.23	0.00	0.00	62024	9743
5	MS	Listening	0.57	0.63	0.49	0.44	0.00	0.00	61958	9772
5		AVERAGE	0.53	0.57	0.36	0.30	0.00	0.00		
6	MC	Writing	0.64	0.67	0.21	0.27	0.00	0.00	9517	62348
6	TE	Reading	0.47	0.46	0.26	0.38	0.00	0.01	9551	61993
6	MC	Reading	0.76	0.77	0.34	0.44	0.00	0.00	9562	62236
6	MC	Reading	0.74	0.75	0.36	0.42	0.00	0.00	9562	62241
6	MC	Reading	0.35	0.39	0.21	0.32	0.00	0.00	9562	62233
6	MC	Reading	0.53	0.58	0.31	0.39	0.00	0.00	9514	62317
6	MC	Reading	0.51	0.53	0.24	0.33	0.00	0.00	9520	62337
6	MC	Reading	0.55	0.66	0.41	0.52	0.00	0.00	9517	62332
6	MC	Reading	0.39	0.41	0.08	0.09	0.00	0.00	9503	62279
6	ESR	Listening	0.67	0.72	0.43	0.50	0.00	0.00	9518	62389
6	MC	Listening	0.74	0.79	0.45	0.45	0.00	0.00	9507	62335
6	MC	Listening	0.20	0.26	0.05	0.20	0.00	0.00	9512	62355
6		AVERAGE	0.55	0.58	0.28	0.36	0.00	0.00		

Table 6-12 English Language Arts Grade 6 vs. Grade 7 Vertical Linking Item Statistics

Item Grade	Item Type	Domain	Item Statistics in Administration Grade							
			PvalG6	PvalG7	RttG6	RttG7	OmitG6	OmitG7	NobsG6	NobsG7
6	ESR	Writing	0.45	0.53	0.37	0.28	0.00	0.00	62303	9693
6	MC	Writing	0.73	0.76	0.41	0.42	0.00	0.00	62370	9684
6	ESR	Reading	0.65	0.72	0.50	0.40	0.00	0.00	62400	9688
6	MC	Reading	0.60	0.65	0.37	0.29	0.00	0.00	62350	9679
6	MC	Reading	0.52	0.64	0.41	0.37	0.00	0.00	62351	9681
6	MC	Reading	0.55	0.64	0.34	0.26	0.00	0.00	62332	9684
6	ESR	Listening	0.72	0.78	0.50	0.44	0.00	0.00	62389	9684
6	ESR	Listening	0.36	0.40	0.28	0.19	0.00	0.00	62353	9693
6	MC	Listening	0.48	0.50	0.31	0.23	0.00	0.00	62315	9692
6	MC	Listening	0.63	0.68	0.40	0.34	0.00	0.00	62268	9688
6	MC	Listening	0.79	0.83	0.45	0.43	0.00	0.00	62335	9680
6	MC	Listening	0.26	0.30	0.20	0.17	0.00	0.00	62355	9677
6		AVERAGE	0.56	0.62	0.38	0.32	0.00	0.00		
7	MC	Writing	0.79	0.80	0.36	0.48	0.00	0.00	9699	61632
7	MC	Reading	0.63	0.63	0.27	0.32	0.00	0.00	9734	61689
7	ESR	Reading	0.62	0.68	0.35	0.46	0.00	0.00	9740	61695
7	MC	Reading	0.60	0.64	0.29	0.37	0.00	0.00	9735	61653
7	TE	Reading	0.91	0.91	0.35	0.41	0.00	0.00	9689	61602
7	MC	Reading	0.71	0.77	0.41	0.46	0.00	0.00	9692	61644
7	MC	Reading	0.72	0.76	0.38	0.44	0.00	0.00	9683	61636
7	MC	Reading	0.72	0.73	0.30	0.36	0.00	0.00	9689	61646
7	MC	Reading	0.75	0.83	0.41	0.46	0.00	0.00	9739	61667
7	TE	Listening	0.73	0.73	0.33	0.48	0.00	0.00	9698	61632
7	MC	Listening	0.93	0.90	0.26	0.38	0.00	0.00	9696	61625
7	MC	Listening	0.65	0.66	0.37	0.43	0.00	0.00	9697	61600
7		AVERAGE	0.73	0.76	0.34	0.42	0.00	0.00		

Table 6-13 English Language Arts Grade 7 vs. Grade 8 Vertical Linking Item Statistics

Item Grade	Item Type	Domain	Item Statistics in Administration Grade							
			PvalG7	PvalG8	RttG7	RttG8	OmitG7	OmitG8	NobsG7	NobsG8
7	MC	Writing	0.65	0.75	0.37	0.31	0.00	0.00	61621	11259
7	MC	Writing	0.57	0.67	0.28	0.27	0.00	0.00	61601	11715
7	MC	Reading	0.35	0.40	0.22	0.19	0.00	0.00	61567	11713
7	MC	Reading	0.67	0.74	0.38	0.25	0.00	0.00	61593	11708
7	MC	Reading	0.89	0.94	0.43	0.34	0.00	0.00	61579	11716
7	MC	Reading	0.69	0.75	0.25	0.15	0.00	0.00	61542	11252
7	TE	Reading	0.63	0.69	0.50	0.44	0.00	0.00	61540	11254
7	ESR	Reading	0.54	0.66	0.53	0.44	0.00	0.00	61574	11257
7	TE	Reading	0.72	0.80	0.48	0.42	0.05	0.03	58916	10940
7	MC	Reading	0.68	0.78	0.48	0.43	0.00	0.00	61573	11706
7	ESR	Listening	0.66	0.70	0.56	0.47	0.00	0.00	61665	11698
7	TE	Listening	0.73	0.80	0.48	0.36	0.00	0.00	61632	11258
7	MC	Listening	0.73	0.82	0.39	0.36	0.00	0.00	61675	11700
7	MC	Listening	0.40	0.44	0.10	0.11	0.00	0.00	61658	11699
7	MC	Listening	0.90	0.94	0.38	0.28	0.00	0.00	61625	11255
7	MC	Listening	0.66	0.75	0.43	0.40	0.00	0.00	61600	11260
7		AVERAGE	0.65	0.72	0.39	0.33	0.00	0.00		
8	MC	Writing	0.75	0.76	0.38	0.47	0.00	0.00	9624	60543
8	TE	Reading	0.74	0.73	0.49	0.58	0.00	0.01	9610	60357
8	MC	Reading	0.64	0.63	0.42	0.49	0.00	0.00	9609	60497
8	TE	Reading	0.49	0.49	0.30	0.42	0.00	0.00	9617	60454
8	TE	Reading	0.52	0.55	0.29	0.37	0.02	0.02	9459	59442
8	MC	Reading	0.42	0.50	0.10	0.22	0.00	0.00	9689	60559
8	MC	Reading	0.65	0.69	0.39	0.40	0.00	0.00	9689	60566
8	MC	Reading	0.72	0.75	0.47	0.50	0.00	0.00	9686	60595
8	MC	Reading	0.66	0.71	0.45	0.48	0.00	0.00	9693	60559
8	ESR	Listening	0.65	0.64	0.36	0.48	0.00	0.00	9628	60521
8	MC	Listening	0.41	0.47	0.28	0.40	0.00	0.00	9626	60520
8	MC	Listening	0.69	0.72	0.31	0.43	0.00	0.00	9621	60515
8		AVERAGE	0.61	0.63	0.35	0.44	0.00	0.00		

Table 6-14 English Language Arts Scale Score Means and Standard Deviations

Grade Level	Scale Statistics			Mean Difference between Grades (in scale score points)
	N Count	Mean	SD	
3	63822	560.85	47.12	
4	62341	582.33	50.00	21.48
5	62081	599.85	50.94	17.52
6	62412	610.78	51.71	10.93
7	61739	624.33	54.38	13.55
8	60648	637.75	56.82	13.43

Table 6-15 English Language Arts Scale Scores at Different Percentiles across Grades

Grade	Percentile						
	5th	10th	25th	50th	75th	90th	95th
3	486	500	527	561	594	621	637
4	499	516	549	583	616	645	663
5	516	535	567	600	633	663	682
6	524	546	579	613	645	673	691
7	531	552	589	627	662	690	709
8	542	562	599	639	677	709	727

Table 6-16 Mathematics Grade 3 vs. Grade 4 Vertical Linking Item Statistics

Item Grade	Item Type	Standard	Item Statistics in Administration Grade							
			PvalG3	PvalG4	RttG3	RttG4	OmitG3	OmitG4	NobsG3	NobsG4
3	MC	NF	0.41	0.62	0.30	0.42	0.00	0.00	62436	19340
3	MC	MD	0.16	0.23	0.25	0.40	0.00	0.00	62510	19350
3	MC	NBT	0.44	0.46	0.37	0.38	0.00	0.00	62560	19351
3	MC	GE	0.52	0.62	0.44	0.40	0.00	0.00	62564	19337
3	MC	OA	0.41	0.59	0.45	0.39	0.00	0.00	62514	19343
3	MC	MD	0.47	0.54	0.37	0.39	0.00	0.00	62566	19334
3	MC	OA	0.45	0.54	0.36	0.39	0.00	0.00	62539	19339
3	MC	MD	0.18	0.26	0.19	0.37	0.00	0.00	62409	19343
3		AVERAGE	0.38	0.48	0.34	0.39	0.00	0.00		
4	MC	OA	0.70	0.75	0.34	0.37	0.00	0.00	31511	61637
4	MC	OA	0.40	0.45	0.18	0.33	0.00	0.00	31508	61649
4	MC	NBT	0.70	0.70	0.42	0.45	0.00	0.00	31486	61651
4	MC	NF	0.35	0.70	0.15	0.39	0.00	0.00	31485	61582
4	MC	NF	0.34	0.53	0.39	0.59	0.00	0.00	31502	61603
4	MC	MD	0.20	0.31	0.10	0.36	0.00	0.00	31460	61574
4	MC	MD	0.59	0.63	0.48	0.49	0.00	0.00	31495	61623
4	MC	GE	0.25	0.47	0.13	0.39	0.00	0.00	31488	61620
4		AVERAGE	0.44	0.57	0.27	0.42	0.00	0.00		

Note: Content categories are as follows: OA = Operations and Algebraic Thinking; NBT = Numbers and Operations in Base Ten; NF = Numbers and Operations - Fractions; MD = Measurement and Data; GE = Geometry; RP = Ratios and Proportional Relationships; NS = The Number System; EE = Expressions and Equations; SP = Statistics and Probability; and F = Functions.

Table 6-17 Mathematics Grade 4 vs. Grade 5 Vertical Linking Item Statistics

Item Grade	Item Type	Standard	Item Statistics in Administration Grade							
			PvalG4	PvalG5	RttG4	RttG5	OmitG4	OmitG5	NobsG4	NobsG5
4	MC	OA	0.59	0.70	0.37	0.32	0.00	0.00	61621	19560
4	MC	NBT	0.38	0.51	0.31	0.35	0.00	0.00	61579	19540
4	MC	MD	0.49	0.64	0.49	0.44	0.00	0.00	61597	19538
4	MC	NF	0.46	0.66	0.56	0.48	0.00	0.00	61613	19551
4	MC	MD	0.53	0.63	0.38	0.28	0.00	0.00	61553	19544
4	MC	OA	0.42	0.51	0.26	0.41	0.00	0.00	61602	19556
4	MC	NF	0.36	0.57	0.58	0.52	0.00	0.00	61623	19542
4	MC	OA	0.25	0.33	0.16	0.21	0.00	0.00	61564	19533
4		AVERAGE	0.44	0.57	0.39	0.37	0.00	0.00		
5	MC	OA	0.39	0.41	0.21	0.26	0.00	0.00	19468	60724
5	MC	MD	0.61	0.58	0.18	0.19	0.00	0.00	19449	60800
5	MC	MD	0.33	0.43	0.16	0.35	0.00	0.00	19440	60766
5	MC	NF	0.53	0.57	0.17	0.30	0.00	0.00	19470	60749
5	MC	NF	0.34	0.52	0.49	0.57	0.00	0.00	19454	60756
5	MC	OA	0.37	0.50	0.30	0.45	0.00	0.00	19472	60746
5	MC	NBT	0.58	0.67	0.36	0.40	0.00	0.00	19460	60830
5	MC	NBT	0.74	0.80	0.30	0.32	0.00	0.00	19466	60823
5		AVERAGE	0.49	0.56	0.27	0.36	0.00	0.00		

Note: Content categories are as follows: OA = Operations and Algebraic Thinking; NBT = Numbers and Operations in Base Ten; NF = Numbers and Operations - Fractions; MD = Measurement and Data; GE = Geometry; RP = Ratios and Proportional Relationships; NS = The Number System; EE = Expressions and Equations; SP = Statistics and Probability; and F = Functions.

Table 6-18 Mathematics Grade 5 vs. Grade 6 Vertical Linking Item Statistics

Item Grade	Item Type	Standard	Item Statistics in Administration Grade							
			PvalG5	PvalG6	RttG5	RttG6	OmitG5	OmitG6	NobsG5	NobsG6
5	ESR	OA	0.32	0.41	0.57	0.50	0.00	0.01	60807	20009
5	MC	OA	0.77	0.86	0.45	0.36	0.00	0.01	60772	20047
5	MC	MD	0.27	0.38	0.38	0.44	0.00	0.01	60687	20009
5	MC	NBT	0.62	0.73	0.45	0.38	0.00	0.01	60696	20001
5	SA	NF	0.45	0.53	0.19	0.20	0.00	0.02	60648	19777
5	SA	NBT	0.44	0.56	0.52	0.40	0.00	0.02	60662	19771
5	ESR	NF	0.15	0.22	0.51	0.52	0.00	0.01	60831	20013
5	ESR	GE	0.11	0.13	0.28	0.28	0.00	0.01	60747	19982
5		AVERAGE	0.39	0.48	0.42	0.38	0.00	0.01		
6	MC	EE	0.27	0.39	0.04	0.29	0.00	0.01	19543	60459
6	MC	NS	0.79	0.84	0.38	0.36	0.00	0.01	19537	60964
6	MC	EE	0.69	0.65	0.19	0.31	0.00	0.01	19551	60630
6	MC	RP	0.46	0.55	0.24	0.34	0.00	0.01	19552	60858
6	MC	NS	0.97	0.94	0.18	0.29	0.00	0.01	19578	60871
6	MC	NS	0.66	0.69	0.39	0.46	0.00	0.01	19570	60578
6	MC	RP	0.91	0.93	0.32	0.34	0.00	0.01	19569	60938
6	MC	EE	0.37	0.53	0.29	0.45	0.00	0.01	19566	60963
6		AVERAGE	0.64	0.69	0.25	0.36	0.00	0.01		

Note: Content categories are as follows: OA = Operations and Algebraic Thinking; NBT = Numbers and Operations in Base Ten; NF = Numbers and Operations - Fractions; MD = Measurement and Data; GE = Geometry; RP = Ratios and Proportional Relationships; NS = The Number System; EE = Expressions and Equations; SP = Statistics and Probability; and F = Functions.

Table 6-19 Mathematics Grade 6 vs. Grade 7 Vertical Linking Item Statistics

Item Grade	Item Type	Standard	Item Statistics in Administration Grade							
			PvalG6	PvalG7	RttG6	RttG7	OmitG6	OmitG7	NobsG6	NobsG7
6	MC	NS	0.47	0.60	0.39	0.36	0.01	0.00	60912	20068
6	MC	EE	0.38	0.42	0.26	0.31	0.00	0.00	61134	20047
6	MC	RP	0.16	0.21	0.44	0.46	0.01	0.00	60949	20068
6	MC	RP	0.34	0.33	0.16	0.06	0.01	0.01	60661	19961
6	MC	EE	0.33	0.45	0.34	0.35	0.00	0.00	61205	20099
6	MC	NS	0.62	0.73	0.45	0.37	0.00	0.00	61182	20116
6	SA	GE	0.12	0.20	0.50	0.55	0.01	0.01	60826	19908
6	ESR	SP	0.07	0.08	0.23	0.24	0.00	0.00	61121	20043
6		AVERAGE	0.31	0.38	0.35	0.34	0.00	0.00		
7	MC	RP	0.57	0.61	0.38	0.47	0.00	0.00	20124	60636
7	MC	NS	0.27	0.38	0.27	0.42	0.01	0.01	19961	60381
7	MC	NS	0.52	0.49	0.40	0.36	0.00	0.01	20092	60468
7	MC	SP	0.51	0.54	0.14	0.23	0.00	0.00	20097	60627
7	MC	RP	0.64	0.66	0.34	0.38	0.00	0.00	20087	60730
7	MC	EE	0.41	0.51	0.32	0.42	0.00	0.00	20114	60666
7	MC	EE	0.34	0.32	0.14	0.24	0.01	0.01	19884	60168
7	MC	EE	0.46	0.55	0.14	0.31	0.01	0.01	19951	60335
7		AVERAGE	0.47	0.51	0.27	0.35	0.01	0.01		

Note: Content categories are as follows: OA = Operations and Algebraic Thinking; NBT = Numbers and Operations in Base Ten; NF = Numbers and Operations - Fractions; MD = Measurement and Data; GE = Geometry; RP = Ratios and Proportional Relationships; NS = The Number System; EE = Expressions and Equations; SP = Statistics and Probability; and F = Functions.

Table 6-20 Mathematics Grade 7 vs. Grade 8 Vertical Linking Item Statistics

Item Grade	Item Type	Standard	Item Statistics in Administration Grade							
			PvalG7	PvalG8	RttG7	RttG8	OmitG7	OmitG8	NobsG7	NobsG8
7	MC	SP	0.56	0.65	0.50	0.46	0.00	0.00	60569	32055
7	MC	RP	0.42	0.46	0.21	0.17	0.01	0.00	60551	32068
7	MC	EE	0.21	0.26	0.41	0.40	0.01	0.00	60504	32102
7	MC	EE	0.37	0.46	0.44	0.44	0.00	0.00	60575	32097
7	MC	GE	0.33	0.49	0.46	0.44	0.00	0.00	60565	32045
7	MC	GE	0.58	0.64	0.33	0.28	0.00	0.00	60656	32086
7	SA	RP	0.33	0.45	0.62	0.52	0.01	0.01	60296	31978
7	ESR	SP	0.11	0.14	0.48	0.44	0.00	0.01	60574	32003
7		AVERAGE	0.36	0.44	0.43	0.39	0.00	0.00		
8	MC	NS	0.39	0.36	0.18	0.25	0.01	0.01	19913	60234
8	MC	EE	0.60	0.63	0.22	0.27	0.00	0.00	20014	60408
8	MC	EE	0.41	0.49	0.39	0.49	0.01	0.01	19922	60318
8	MC	GE	0.50	0.53	0.19	0.31	0.00	0.00	19951	60432
8	MC	SP	0.52	0.59	0.37	0.43	0.00	0.00	19969	60407
8	MC	F	0.47	0.49	0.11	0.22	0.00	0.00	19963	60457
8	MC	F	0.74	0.74	0.42	0.44	0.00	0.00	19998	60412
8	MC	GE	0.29	0.32	0.10	0.17	0.00	0.00	19991	60357
8		AVERAGE	0.49	0.52	0.25	0.32	0.00	0.00		

Note: Content categories are as follows: OA = Operations and Algebraic Thinking; NBT = Numbers and Operations in Base Ten; NF = Numbers and Operations - Fractions; MD = Measurement and Data; GE = Geometry; RP = Ratios and Proportional Relationships; NS = The Number System; EE = Expressions and Equations; SP = Statistics and Probability; and F = Functions.

Table 6-21 Mathematics Scale Score Means and Standard Deviations

Grade Level	Scale Statistics			Mean Difference between Grades (in scale score points)
	N Count	Mean	SD	
3	62648	554.70	46.23	
4	61702	573.83	55.95	19.13
5	60884	599.92	49.92	26.09
6	61295	613.45	52.57	13.53
7	60838	628.30	56.80	14.85
8	60620	641.83	56.54	13.53

Table 6-22 Mathematics Scale Scores at Different Percentiles across Grades

Grade	Percentile						
	5th	10th	25th	50th	75th	90th	95th
3	478	497	528	558	585	609	624
4	466	505	547	581	610	635	650
5	512	538	572	604	633	657	672
6	525	547	582	617	649	676	692
7	530	562	597	633	666	694	709
8	543	575	614	648	678	705	720

Table 6-23 Science Scale Score Means and Standard Deviations

Grade Level	Scale Statistics		
	N Count	Mean	SD
4	52631	400.15	50.55
8	53203	599.99	50.73

Table 6-24 Science Scale Scores at Different Percentiles across Grades

Grade	Percentile						
	5th	10th	25th	50th	75th	90th	95th
4	317	337	369	401	432	460	480
8	518	539	570	600	629	658	679

Table 6-25 Social Studies Scale Score Means and Standard Deviations

Grade Level	Scale Statistics		
	N Count	Mean	SD
4	48902	399.82	50.50
8	51809	600.08	50.39
10	43997	699.73	51.10

Table 6-26 Social Studies Scale Scores at Different Percentiles across Grades

Grade	Percentile						
	5th	10th	25th	50th	75th	90th	95th
4	320	338	369	400	430	457	476
8	521	539	569	600	631	660	679
10	618	638	669	701	732	759	777

Table 6-27 Scoring Table for English Language Arts Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	93	31	579	13
1	330	93	32	584	14
2	330	93	33	589	14
3	330	93	34	593	14
4	396	42	35	599	14
5	423	30	36	604	15
6	439	24	37	609	15
7	451	21	38	615	16
8	461	19	39	622	16
9	470	18	**40**	**629**	**17**
10	477	17	41	636	18
11	484	16	42	644	19
12	490	16	43	653	19
13	496	15	44	663	21
14	502	15	45	674	22
15	507	15	46	687	24
16	512	14	47	702	27
17	517	14	48	722	30
18	**522**	**14**	49	747	34
19	527	14	50	776	36
20	531	13	51	808	38
21	536	13	52	851	48
22	540	13	53	900	78
23	544	13			
24	549	13			
25	553	13			
26	557	13			
27	562	13			
28	566	13			
29	**570**	**13**			
30	575	13			

Note: **Bold** represents SEM 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 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	340	73	31	585	14
1	340	73	32	590	14
2	340	73	**33**	**594**	**14**
3	340	73	34	599	14
4	340	73	35	605	14
5	357	62	36	610	14
6	402	39	37	616	15
7	426	29	38	622	16
8	442	25	39	628	16
9	455	22	40	635	17
10	465	20	41	642	18
11	474	19	**42**	**650**	**19**
12	482	18	43	659	20
13	489	17	44	668	21
14	496	17	45	678	22
15	502	17	46	688	21
16	508	16	47	699	20
17	514	16	48	710	21
18	520	16	49	724	26
19	526	16	50	741	32
20	531	16	51	761	36
21	536	15	52	785	40
22	541	15	53	813	45
23	**546**	**15**	54	851	53
24	551	15	55	908	71
25	556	14	56	930	81
26	561	14			
27	566	14			
28	571	14			
29	575	14			
30	580	14			

Note: **Bold** represents SEM 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 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	93	31	619	15
1	350	93	32	624	16
2	350	93	33	630	16
3	350	93	34	636	17
4	350	93	35	642	17
5	417	47	36	649	18
6	446	34	37	656	19
7	465	28	38	663	19
8	480	25	**39**	**671**	**20**
9	491	22	40	679	21
10	501	21	41	688	22
11	509	20	42	697	23
12	517	19	43	707	24
13	524	18	44	719	26
14	531	17	45	731	28
15	537	17	46	745	30
16	543	16	47	761	33
17	549	16	48	779	36
18	554	16	49	801	39
19	559	15	50	826	42
20	**564**	**15**	51	853	44
21	569	15	52	884	45
22	574	14	53	917	48
23	579	14	54	940	52
24	584	14	55	940	52
25	589	14	56	940	52
26	593	14			
27	598	14			
28	603	14			
29	608	15			
30	**613**	**15**			

Note: **Bold** represents SEM 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 English Language Arts Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	360	92	31	627	14
1	360	92	32	632	15
2	360	92	33	637	15
3	360	92	34	642	15
4	360	92	35	647	15
5	360	92	36	653	16
6	424	57	37	659	16
7	458	43	38	665	17
8	480	34	**39**	**672**	**18**
9	496	29	40	678	18
10	508	26	41	686	19
11	518	23	42	694	20
12	527	22	43	703	21
13	535	20	44	712	22
14	542	19	45	723	24
15	549	18	46	734	25
16	555	17	47	747	27
17	560	17	48	762	30
18	566	16	49	778	32
19	571	16	50	798	35
20	**576**	**15**	51	821	39
21	581	15	52	848	42
22	586	15	53	880	45
23	590	15	54	919	51
24	595	14	55	950	59
25	599	14	56	950	59
26	604	14			
27	608	14			
28	613	14			
29	617	14			
30	**622**	**14**			

Note: **Bold** represents SEM 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 English Language Arts Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	370	72	31	613	15
1	370	72	32	617	15
2	370	72	33	622	15
3	370	72	34	627	15
4	370	72	35	632	15
5	370	72	36	637	15
6	381	65	**37**	**643**	**15**
7	428	42	38	648	16
8	454	33	39	654	16
9	472	28	40	660	16
10	486	25	41	666	16
11	497	23	42	672	17
12	507	21	43	679	17
13	515	20	44	687	18
14	523	19	45	695	19
15	530	18	**46**	**703**	**20**
16	537	18	47	712	21
17	543	17	48	722	22
18	549	17	49	733	23
19	555	16	50	744	24
20	560	16	51	757	25
21	565	15	52	772	28
22	570	15	53	789	31
23	575	15	54	813	37
24	580	15	55	850	51
25	**585**	**15**	56	960	141
26	590	15			
27	594	14			
28	599	14			
29	603	14			
30	608	14			

Note: **Bold** represents SEM 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 English Language Arts Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	380	87	31	632	15
1	380	87	32	637	15
2	380	87	33	642	15
3	380	87	34	647	15
4	380	87	**35**	**652**	**15**
5	416	59	36	658	16
6	457	38	37	663	16
7	479	29	38	669	16
8	494	25	39	675	16
9	505	22	40	681	16
10	515	20	41	687	16
11	523	19	42	694	17
12	530	18	43	701	17
13	537	17	**44**	**708**	**17**
14	544	17	45	715	18
15	550	17	46	723	18
16	556	16	47	731	19
17	561	16	48	741	20
18	567	16	49	751	21
19	572	16	50	762	23
20	577	15	51	775	25
21	583	15	52	789	27
22	588	15	53	807	30
23	**593**	**15**	54	831	36
24	598	15	55	868	50
25	602	15	56	970	132
26	607	15			
27	612	15			
28	617	15			
29	622	15			
30	627	15			

Note: **Bold** represents SEM 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 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	360	104	26	570	11
1	360	104	27	575	11
2	360	104	28	579	11
3	360	104	29	584	11
4	360	104	30	589	12
5	360	104	31	594	12
6	406	60	32	599	12
7	447	34	33	605	12
8	467	26	**34**	**612**	**13**
9	480	22	35	618	14
10	490	19	36	626	14
11	498	17	37	634	15
12	505	16	38	644	17
13	512	14	39	657	20
14	**517**	**14**	40	674	25
15	523	13	41	705	40
16	528	12	42	760	82
17	532	12			
18	537	12			
19	541	11			
20	545	11			
21	550	11			
22	554	11			
23	558	11			
24	**562**	**11**			
25	566	11			

Note: **Bold** represents SEM 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 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	405	116	26	600	9
1	405	116	27	603	9
2	405	116	28	607	9
3	405	116	29	610	9
4	405	116	30	613	9
5	405	116	31	617	9
6	405	116	32	620	9
7	405	116	33	624	10
8	464	59	34	628	10
9	496	37	35	632	10
10	514	27	**36**	**636**	**10**
11	526	23	37	641	11
12	**536**	**19**	38	646	11
13	544	17	39	651	12
14	550	16	40	658	13
15	556	14	41	665	14
16	562	13	42	674	16
17	566	12	43	685	18
18	571	12	44	701	24
19	575	11	45	732	39
20	579	11	46	800	101
21	583	10			
22	586	10			
23	**590**	**10**			
24	593	10			
25	597	9			

Note: **Bold** represents SEM 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 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	430	111	26	630	9
1	430	111	27	633	9
2	430	111	28	637	9
3	430	111	29	640	9
4	430	111	30	644	9
5	430	111	31	647	10
6	430	111	32	651	10
7	498	47	33	655	10
8	527	30	**34**	**659**	**10**
9	543	24	35	663	10
10	555	20	36	667	11
11	564	17	37	672	11
12	571	16	38	677	12
13	**577**	**14**	39	683	12
14	583	13	40	689	13
15	588	12	41	697	15
16	593	12	42	706	17
17	597	11	43	718	20
18	601	11	44	736	26
19	605	11	45	766	39
20	609	10	46	830	89
21	**613**	**10**			
22	616	10			
23	620	10			
24	623	10			
25	627	9			

Note: **Bold** represents SEM 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 Mathematics Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	440	78	26	645	11
1	440	78	27	649	10
2	440	78	28	653	10
3	440	78	29	657	10
4	440	78	30	661	10
5	440	78	31	665	10
6	454	64	32	669	10
7	499	29	33	673	10
8	518	23	34	677	10
9	532	21	35	682	11
10	544	20	36	687	11
11	555	19	**37**	**692**	**11**
12	565	18	38	697	12
13	573	18	39	703	12
14	581	17	40	710	13
15	**589**	**16**	41	719	15
16	595	15	42	730	19
17	601	14	43	747	25
18	607	14	44	776	39
19	613	13	45	841	74
20	618	13	46	870	90
21	623	12			
22	**627**	**12**			
23	632	11			
24	636	11			
25	640	11			

Note: **Bold** represents SEM 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 Mathematics Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	450	127	26	673	10
1	450	127	27	676	10
2	450	127	28	680	10
3	450	127	29	684	10
4	450	127	30	687	10
5	450	127	31	691	10
6	465	112	32	695	10
7	537	43	33	699	10
8	562	29	34	703	10
9	578	23	35	707	11
10	589	20	36	711	11
11	599	17	**37**	**716**	**11**
12	**607**	**16**	38	721	12
13	614	15	39	727	13
14	620	14	40	734	14
15	626	13	41	741	15
16	631	13	42	750	16
17	636	12	43	761	18
18	641	12	44	775	22
19	645	12	45	801	34
20	**649**	**11**	46	880	104
21	654	11			
22	658	11			
23	661	11			
24	665	10			
25	669	10			

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

Table 6-38 Scoring Table for Mathematics Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	470	117	26	686	11
1	470	117	27	690	11
2	470	117	28	694	11
3	470	117	29	698	11
4	470	117	30	702	11
5	470	117	31	706	11
6	501	86	32	711	11
7	555	40	33	715	11
8	577	30	**34**	**719**	**11**
9	592	24	35	724	11
10	603	21	36	729	12
11	612	18	37	734	12
12	**620**	**17**	38	740	12
13	627	15	39	746	13
14	633	14	40	752	14
15	638	13	41	760	15
16	644	13	42	769	17
17	648	12	43	781	20
18	653	12	44	796	24
19	657	11	45	823	36
20	662	11	46	890	90
21	666	11			
22	**670**	**11**			
23	674	11			
24	678	11			
25	682	11			

Note: **Bold** represents SEM 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 Science Grade 4

Raw Score	Scale Score	SEM
0	190	104
1	190	104
2	190	104
3	190	104
4	190	104
5	190	104
6	190	104
7	190	104
8	220	74
9	252	43
10	271	31
11	284	25
12	295	22
13	304	19
14	312	18
15	319	17
16	326	16
17	333	16
18	339	15
19	345	15
20	**351**	**15**
21	356	15
22	362	15
23	368	15
24	373	15
25	379	15
26	385	15
27	391	15
28	397	15
29	**403**	**15**
30	409	15
31	416	16
32	424	16
33	432	17
34	441	18
35	**451**	**20**
36	463	22
37	479	26
38	501	33
39	539	50
40	600	95

Note: **Bold** represents SEM 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 Science Grade 8

Raw Score	Scale Score	SEM
0	390	102
1	390	102
2	390	102
3	390	102
4	390	102
5	390	102
6	390	102
7	390	102
8	390	102
9	435	57
10	458	34
11	473	25
12	483	21
13	492	18
14	499	17
15	506	16
16	513	15
17	519	15
18	524	15
19	530	14
20	536	14
21	541	14
22	546	14
23	551	14
24	**556**	**13**
25	561	13
26	566	13
27	571	13
28	577	13
29	582	14
30	588	14
31	594	14
32	**600**	**15**
33	607	15
34	615	17
35	624	18
36	635	20
37	**650**	**24**
38	669	30
39	702	43
40	770	93

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

Table 6-41 Scoring Table for Social Studies Grade 4

Raw Score	Scale Score	SEM
0	200	110
1	200	110
2	200	110
3	200	110
4	200	110
5	200	110
6	200	110
7	235	75
8	269	41
9	286	30
10	299	24
11	308	20
12	316	18
13	323	17
14	330	16
15	336	15
16	341	14
17	346	14
18	351	13
19	356	13
20	361	13
21	**366**	**13**
22	371	13
23	375	13
24	380	13
25	385	13
26	390	13
27	**396**	**13**
28	401	14
29	407	14
30	413	14
31	419	15
32	426	15
33	434	16
34	**443**	**17**
35	454	20
36	469	23
37	494	33
38	570	97

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

Table 6-42 Scoring Table for Social Studies Grade 8

Raw Score	Scale Score	SEM
0	420	100
1	420	100
2	420	100
3	420	100
4	420	100
5	420	100
6	420	100
7	420	100
8	466	54
9	487	34
10	500	26
11	510	21
12	518	19
13	525	17
14	531	15
15	537	14
16	542	14
17	547	13
18	552	13
19	557	12
20	561	12
21	**565**	**12**
22	570	12
23	574	12
24	578	12
25	582	12
26	587	12
27	591	12
28	596	12
29	**601**	**12**
30	606	12
31	611	13
32	617	14
33	624	14
34	631	15
35	639	17
36	**649**	**19**
37	662	22
38	680	27
39	710	40
40	780	95

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

Table 6-43 Scoring Table for Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	490	126	**26**	**672**	**12**
1	490	126	27	676	12
2	490	126	28	680	12
3	490	126	29	684	11
4	490	126	30	688	11
5	490	126	31	692	11
6	490	126	32	695	11
7	490	126	33	699	12
8	490	126	**34**	**703**	**12**
9	490	126	35	707	12
10	490	126	36	712	12
11	531	85	37	716	12
12	568	48	38	720	12
13	588	34	39	725	12
14	601	27	40	730	12
15	612	24	41	735	12
16	621	21	42	740	13
17	628	19	**43**	**746**	**13**
18	635	18	44	752	14
19	641	16	45	759	15
20	646	15	46	768	17
21	651	14	47	779	19
22	656	14	48	794	24
23	660	13	49	820	36
24	664	13	50	890	94
25	668	12			

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

Table 6-44 The Number and Percent of Students at LOSS and HOSS

Content	Grade	LOSS	N	Percent	HOSS	N	Percent
ELA	3	330	9	0.01	900	0	0.00
	4	340	3	0.01	930	2	0.00
	5	350	4	0.01	940	0	0.00
	6	360	52	0.09	950	0	0.00
	7	370	3	0.01	960	8	0.01
	8	380	17	0.03	970	9	0.02
Math	3	360	385	0.63	760	35	0.06
	4	405	2005	3.35	800	44	0.07
	5	430	672	1.13	830	29	0.05
	6	440	801	1.33	870	8	0.01
	7	450	1775	2.98	880	12	0.02
	8	470	2302	3.90	890	8	0.01
Science	4	190	112	0.19	600	179	0.30
	8	390	201	0.34	770	737	1.25
Social Studies	4	200	198	0.33	570	924	1.54
	8	420	356	0.60	780	503	0.85
	10	490	419	0.67	890	364	0.59

Figure 6-1 English Language Arts Test Characteristic Curves

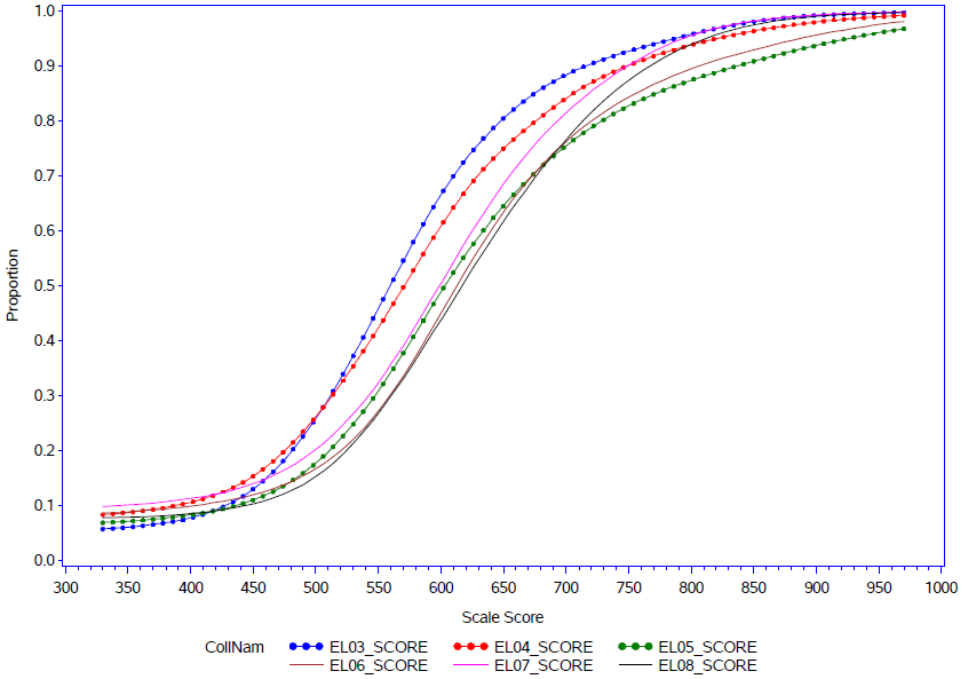


Figure 6-2 English Language Arts Standard Error Curves

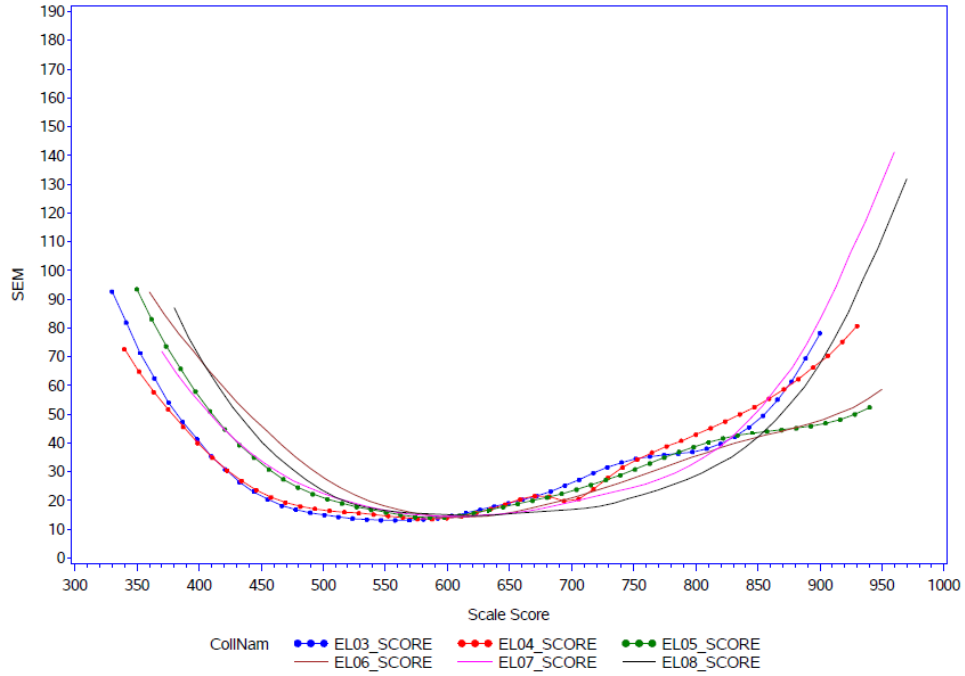


Figure 6-3 Mathematics Test Characteristic Curves

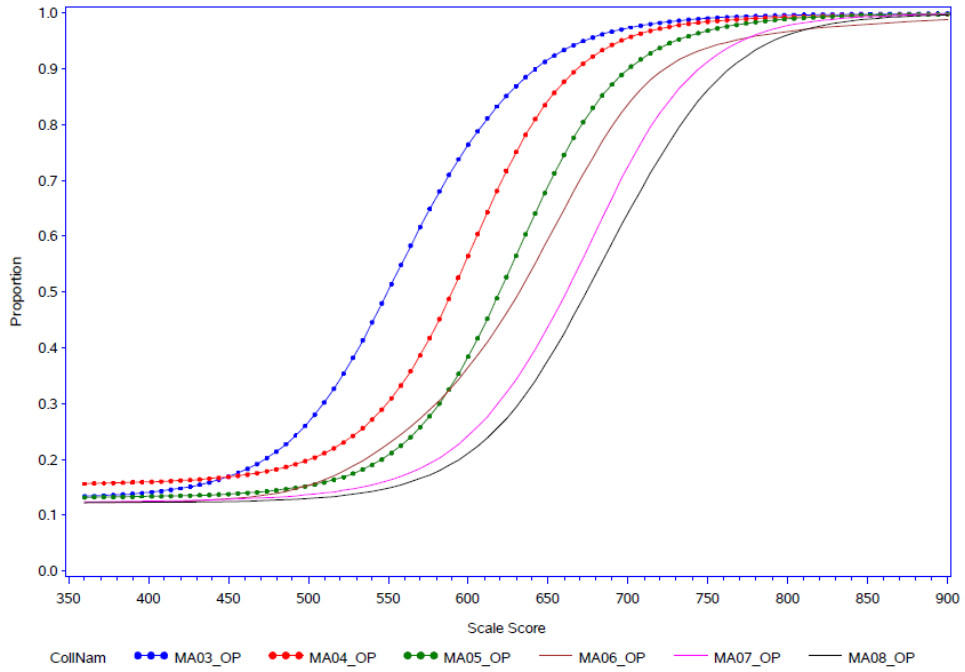


Figure 6-4 Mathematics Standard Error Curves

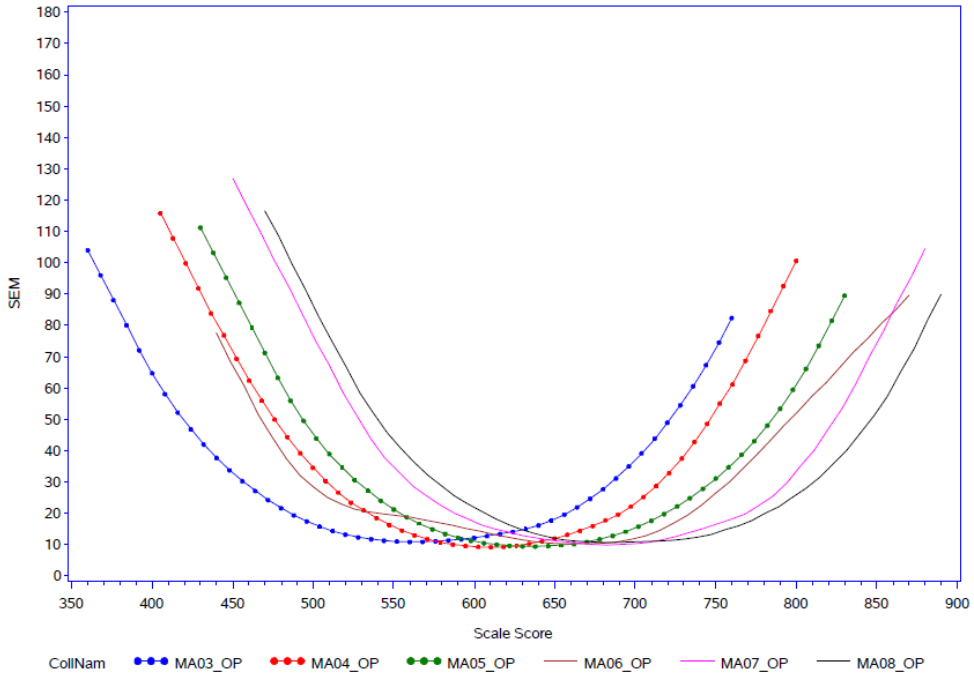


Figure 6-5 Science Test Characteristic Curves

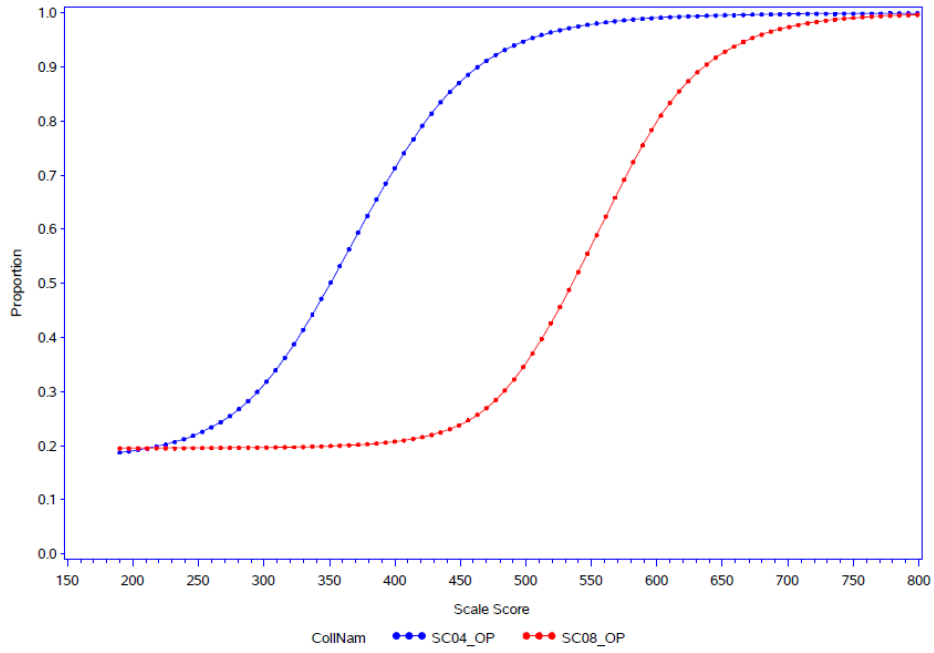


Figure 6-6 Science Standard Error Curves

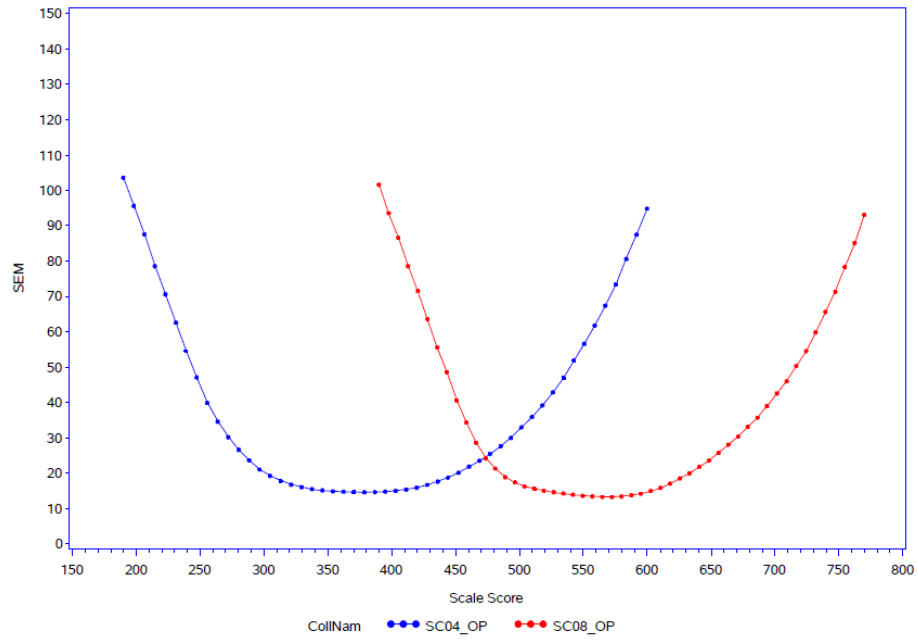


Figure 6-7 Social Studies Test Characteristic Curves

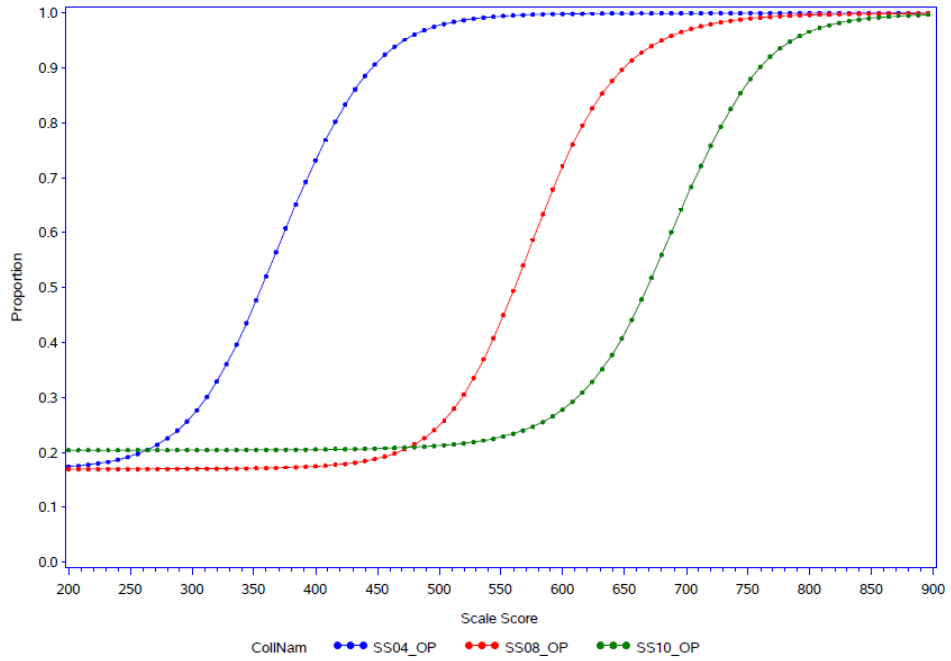
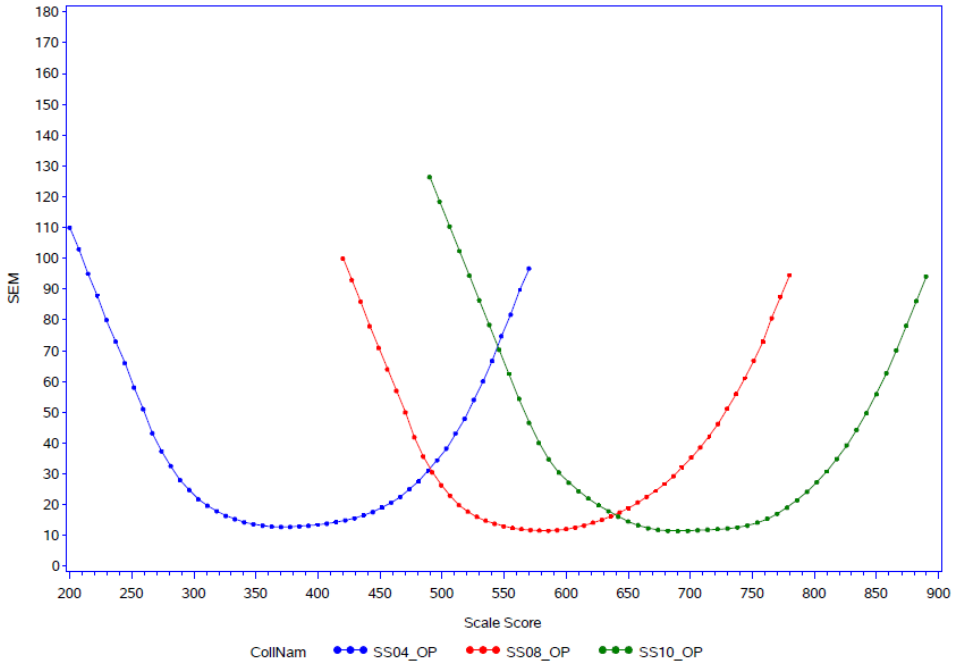


Figure 6-8 Social Studies Standard Error Curves



Part 7: Standard Setting

In this chapter, we briefly describe the Wisconsin Forward Exam standard setting, and we present the cut scores established and the performance-level descriptors derived from the standard setting. The information in this chapter comes from the *Wisconsin Standard Setting 2016 Final Technical Report* submitted to DPI and available at http://dpi.wi.gov/assessment/forward/resources.

7.1 Background Information

Wisconsin's statewide tests have recently gone through several changes, especially for English Language Arts (ELA) and Mathematics. In the 2014–15school 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 were established for all four content areas using data from the Spring 2016 administration of the Wisconsin Forward Exam.

On June 14–17, 2016, DPI and DRC conducted the Wisconsin Forward Exam Standard Setting for grades 3–8 in ELA and Mathematics; for grades 4 and 8 Science; and for grades 4, 8, and 10 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 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.

The process of the standard setting adhered to the 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. (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. (108)

7.2 Standard Setting Methodology

Prior to the standard setting workshop, 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, DPI selected the Bookmark Standard Setting Procedure (BSSP) for the Wisconsin Forward Exam. The BSSP is well suited for standard setting for these assessments because (a) the tests are composed of both multiple-choice and constructed-response 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 the 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).

7.3 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. These brief descriptors, shown in Table 7-1, described DPI's vision for each performance level. At the standard setting, Wisconsin used the policy PLDs in conjunction with the content standards to consider the content-based expectations for students in each performance level on each test in the Wisconsin Forward Exam program.

7.4 Cut Scores

In this section, the cut scores for each grade/content area of the Wisconsin Forward Exam program are presented. 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).

For details on the Wisconsin Forward Exam standard setting procedure and results, please refer to the *Wisconsin Standard Setting 2016 Final Technical Report*.

Table 7-1 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 his or her grade level that are associated with college content–readiness.
Basic	Student demonstrates **partial understanding** of and ability to apply the knowledge and skills for his or her grade level that are associated with college content–readiness.
Proficient	Student demonstrates **adequate understanding** of and ability to apply the knowledge and skills for his or her grade level that are associated with college content–readiness.
Advanced	Student demonstrates **thorough understanding** of and ability to apply the knowledge and skills for his or her 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	348	399	447
	8	552	600	645
Social Studies	4	363	396	436
	8	563	599	640
	10	670	703	741

Part 8: Test Results

Part 8 presents a classical item analysis and summary of student results for the Spring 2016 Wisconsin Forward Exam. The summary results are presented for public school students and cover four types of scores: raw scores; scale scores; performance level results; and scores based on each of the content standards within each content area which are called standard performance index (SPI) scores. Combined, the classical item analysis and the four forms of scores offer the reader several vantage points from which to understand and evaluate the Wisconsin Forward Exam testing program. The AERA, APA, & NCME (2014) standards addressed in Part 8 include 1.8, 4.14, 5.1, 5.21, 7.0, and 7.1. These standards are cited below:

> **Standard 1.8** 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)
>
> **Standard 4.14** 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)
>
> **Standard 5.1** 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)
>
> **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 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)

8.1 Classical Item Analysis: Item Level Statistics

Three statistics are frequently used in item analysis at the item level: 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 represents the proportion of students who answered the item correctly. If all students answered a given MC 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. 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 a CR item represents the mean proportion of possible raw score points that students actually obtained for the item. A *p*-value of 0.33 for a given CR item 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. As such, the *p*-value indicates difficulty for CR 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 MC 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 CR 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.[2] 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.

For the Spring 2016 Wisconsin Forward Exam, items were flagged for further investigation according to the following rules:

[2] 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 CR items worth several points.

- The *p*-value was less than 0.30 for MC items. Such a *p*-value indicates a difficult item, where fewer than 30% 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 5%.

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 (especially easy items) or very low *p*-values (especially 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 5% of the students (the Wisconsin Forward Exam flagging criterion) is not necessarily problematic. Omit rates are typically higher for CR 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 in order 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 that of 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 measurement and lower the reliability of the test. Such items should be removed from scoring.

Overall, 117 items were flagged on the Spring 2016 Wisconsin Forward Exam operational tests as meeting the investigational criteria bulleted above.

Table 8-A shows the number of scored items in the Spring 2016 Wisconsin Forward Exam operational tests flagged for these conditions by grade and content area. Because some items were flagged for more than one condition, the number of flags may be greater than the number of flagged items.

The flagged items were referred to DRC's content specialists for further review to ensure that the items were unambiguous and the answer keys were correct. As part of this review, DRC's content experts also evaluated each flagged item against the Wisconsin Forward Exam depth-of-knowledge criteria to ensure that the cognitive demands of the item reflected the skills and knowledge that the item was designed to measure. Tables 8-B, 8-C, and 8-D provide more information about the flagged items.

8.1.1 Flagging for a Positive Distractor Correlation

In tables 8-B to 8-D, 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 key. All items flagged for a positive distractor-total test correlation had a distractor-total test correlation less than or equal to 0.25. These items were judged to be acceptable based on their other statistics and were retained in order to meet the Wisconsin Forward Exam test blueprints.

8.1.2 Flagging for the Item-Total Correlation

Thirteen items were flagged for item-total correlations <0.15, and all of the flagged items were 0.10 or above except for six items (ELA Grades 3, 4, 5, and 6 [0.08, -0.03, 0.09, and 0.09] and Mathematics grades 4 and 7 [0.03 and -0.01]). Although these items, with correlation coefficients ranging below 0.15, are fairly low, the fact that the majority are positive indicates that the items are contributing information about student ability. These items, therefore, were retained in order to meet the Wisconsin Forward Exam blueprints.

8.1.3 Flagging for *p*-Value

Seventy items were flagged for *p*-values <0.30, and all the items had *p*-values between 0.04 and 0.29. While these statistics indicate items that were very difficult, the number of items flagged for difficulty was very small. Only one of the test forms had more than one item flagged for difficulty.

8.1.4 Flagging for Omit Rate

All ELA TDA items were flagged for omit rates or responses being "insufficient to score" greater than 5%. The "omit" and "insufficient to score" codes were combined in the data analysis. It is hypothesized that the high percentage of "omits" or "insufficient to score" may have resulted from Wisconsin students' unfamiliarity with this type of item. TDA items were reasonably discriminating items and, since they were an essential part of the ELA test blueprint, were retained to meet the Wisconsin Forward Exam blueprints.

8.1.5 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 formulates that at least 80% of the test takers should be able to answer all 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 unspeeded only if at least 95% of the examinees attempt the final item. As shown in Table 8-E, the Wisconsin Forward Exam satisfies this more stringent requirement, with more than 99% of the examinees attempting the final item in each of the four content areas.

8.1.6 Supplemental Tables on Classical Item Analysis

Tables 8-1 through 8-17 present more comprehensive results from the classical item analysis for all of 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 5%, and the *p*-value is flagged when it is below 0.30.

Readers may note that the results presented in these tables may differ slightly from testing results presented on DPI's website due to slight differences in the decision rules defining which students are included or excluded from summary results. Official final results are based on the application of detailed inclusion rules, such as whether the student moved into a school and how long he or she was in one school or another over the course of the year.

The item analysis tables show the item number, which can be used to understand the location of test items as students actually encountered them in test booklets. The item analysis tables also indicate item type (e.g., MC, ESR). Items removed from the scoring of these tests are not included in these tables.

The number of flagged items across grade and content areas are summarized in Table 8-A. 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 point-biserial correlations reasonably high for most items.

8.2 Raw Score Results

Raw score results based on all students who took the Spring 2016 Wisconsin Forward Exam are presented in Table 8-18. In order to facilitate interpretation of the raw score results, Table 8-18 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 3-1 through 3-4 for a count of the number of items in the test and the number of raw score points corresponding to each item.

The mean raw score should be understood by grade and content area and specifically in the context of the maximum possible score points. In English Language Arts, for example, the maximum possible raw score ranges from 53 to 56, and it ranges from 42 to 46 in Mathematics.

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. In English Language Arts grade 5, the test difficulty statistic (0.57) was obtained by taking the mean raw score and dividing it by 56.

Table 8-18 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 indicates 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) that most student scores are clustered on the high end of the scale. A positive skew indicates a distribution with some extreme high scores and a corresponding increase in the 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 8-18 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 English Language Arts and Mathematics and Wisconsin's Model Academic Standards for Science and Social Studies.

In addition, Table 8-18 shows that the minimum observed score in most content areas/grades are zero, meaning that at least one student failed all items for each of those tests. The majority the maximum obtained scores are equal to the maximum number of points possible on the test, meaning that at least one student obtained the full scores for all items on each of those tests. For example, as displayed in Table 8-18, 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 9. 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 raw score results for each content area are summarized and discussed below using the measurements described above.

English Language Arts

- Test difficulty ranged from 0.56 to 0.66.
- Standard deviations ranged from 8.79 to 10.14 raw score points.
- Alpha was relatively high in every grade (0.85 to 0.89).
- SEM ranged from 3.16 to 3.50.

Mathematics

- Test difficulty ranged from 0.40 to 0.54, with generally lower difficulty in lower grades and higher difficulty in higher grades.
- Standard deviations ranged from 8.37 to 9.63 raw score points.
- Alpha was relatively high in every grade (0.90 to 0.91).
- SEM ranged from 2.67 to 2.85.

Science

- Test difficulty were close to or above 0.70.
- Standard deviations were close to 7.0 raw score points.
- Alpha was slightly below 0.90 for both grades.
- SEM was around 2.50 for both grades.

Social Studies

- Test difficulty ranged from 0.65 to 0.70.
- Standard deviations ranged from 7.48 to 10.02 raw score points.
- Alpha ranged from 0.89 to 0.91.
- SEM ranged from 2.45 to 2.98.

Subgroup Performance Patterns in Raw Score Results

In the previous section, the raw score results were discussed with reference to the total student population. In this section, subgroup comparisons are made based on gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. These subgroup comparisons draw from Tables 8-19 through 8-26.

Overall, the raw score results show some consistent performance patterns by subgroups, that is, in terms of gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency.

Regarding scores by gender, in English Language Arts, the tests were slightly easier for female students as a group than for male students as a group in each grade level, with test difficulty differences ranging from 0.03 in grades 3, 4, and 7 to 0.05 in grade 8. In Mathematics, the test difficulties were very similar between male and female students at grades 5 and above. At grades 3 and 4, the tests were slightly easier for male students than for female students, with the differences at 0.02 and 0.03, respectively. In Science, the test difficulties were very similar between male and female students at grades 4 and 8, with the differences at 0.00 and 0.01, respectively. In Social Studies, the differences in test difficulty between genders were, again, very small (at 0.01) across grades.

In all grades and content areas, the raw score results showed consistent performance patterns by ethnicity. In every grade and content area, the test was generally the easiest for White students, followed by Asian students, American Indian students and Hispanic students, and African American students. American Indian students had similar or slightly higher mean raw score than Hispanic students. Differences in test difficulty between American Indian and Hispanic students were all equal to or less than 0.01 across grades and content areas.

In every grade and content area, the test was easier for those students who were not economically disadvantaged than for those who were economically disadvantaged. The difference in test difficulty between the two groups ranged from 0.11 (ELA grade 5) to 0.15 (Mathematics grades 4 and 5, and Social Studies grade 4).

There were also differences in test difficulty between students with disabilities and those without disabilities in all grades and content areas. The test was consistently easier for students without disabilities than for students with disabilities, with differences ranging from 0.14 in English Language Arts grade 3 and Science grade 4, to 0.23 in Social Studies grade 8.

In every grade and content area, the test was markedly easier for students who were fully English proficient than for students who were limited English proficient. Differences in test difficulty ranged from 0.12 to 0.21 in ELA, 0.12 to 0.18 in Mathematics, 0.14 to 0.21 in Science, and 0.16 to 0.23 in Social Studies.

8.3 Summary Statistics for Scale Scores

The Wisconsin Forward Exam program reports scale scores as well as raw 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 8-27 is the summary scale score table based on public school census data. The table shows the mean scale score, the standard deviation (SD) of the scale scores, skewness and kurtosis, the minimum and maximum obtained scale scores, and the minimum and maximum obtainable 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 by grade level, ranging from 561.89 to 638.02.
- Standard deviations ranged from 47.00 to 57.26 scale score points.
- In half of the grade levels, student scores spanned the full-scale score range from the LOSS to the HOSS (grades 4, 7, and 8).

Mathematics

- Mean scale score increased by grade level, ranging from 555.56 to 642.13.
- Standard deviations ranged from 46.18 to 57.10 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 400.16 and 598.93 for grades 4 and 8, respectively.
- Standard deviations ranged from 51.19 to 52.41 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Social Studies

- Mean scale score increased by grade level, ranging from 399.25 to 698.92.
- Standard deviations ranged from 51.18 to 53.76 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

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, socioeconomic status, disability status, and English language proficiency are drawn from Tables 8-28 through 8-35.

Gender

- In terms of gender, male students as a group showed lower mean scale scores in English Language Arts than female students as a group in each grade level. The difference ranged from 10.19 to 19.94 scale score points.
- In Mathematics, the differences between genders were very small, from 0.49 scale score points to 5.20 scale score points, and male and female students alternated between the higher and lower score groups.
- In Science, the mean scale scores between genders were very similar, with the differences ranging from 0.69 scale score points to 1.53 scale score points, and male and female students alternated between the higher and lower score groups.
- There were very small differences between mean scale scores by gender in Social Studies, from 1.25 scale score points to 2.72 scale score points, and male and female students alternated between the higher and lower score groups.

Race/Ethnicity

- The scale score results showed some consistent performance differences by ethnicity.
- In every grade and content area, White students as a group had the highest mean scale scores, followed by Asian students, American Indian students and Hispanic students, and African American students.
- As was noted in the context of the raw score results, the differences in mean scale scores for American Indian students and Hispanic students were often very small. In all grades and content areas, differences were less than four scale score points.

Socioeconomic 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 30.63 scale score points in ELA grade 3 to 40.38 scale score points in Mathematics grade 4.
- For every grade and content area, the mean scale score of students who were economically disadvantaged was more than two-thirds standard deviation lower than the mean scale score 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 36.79 scale

score points in ELA grade 3 and Mathematics grade 3, to 65.95 scale score points in ELA grade 8.
- For every grade and content area, the mean scale scores of students with disabilities were lower than the mean scale scores of students without disabilities by about or more than four-fifths standard deviation.

English Language Proficiency

- Students who were fully English proficient and students who were limited English proficient showed consistent and large differences in mean scale scores by group. Differences ranged from 27.33 scale score points in Mathematics grade 3 to 59.66 scale score points in Social Studies grade 10.
- For every grade and content area, the mean scale scores of limited English proficient students were more than three-fifths standard deviation lower than the mean scale scores of fully English proficient students.

8.4 Cut Scores and 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 for Educational and Psychological Testing* (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 performance level descriptors 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 summarize the knowledge, skills, and abilities expected of students in each achievement level. As stated in Part 7, DPI provided policy performance level descriptors (PLDs) for the Wisconsin Forward Exam assessments. At the standard setting, Wisconsin 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.

Table 8-36 shows the cut scores for each content and grade level. For ease of reference, Tables 8-37 through 8-40 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

- Across all grade levels, over 42% of students were either *Proficient* or *Advanced* in ELA.
- Over 8% of the total student population was classified as *Advanced*.
- Across all grade levels, approximately 57% of students were below *Proficient*. The difference ranged from 56% below *Proficient* in grades 3 and 4 to 58% below *Proficient* in grade 8.

Mathematics

- Across all grade levels, over 34% of students were either *Proficient* or *Advanced* in Mathematics.
- The proportion of students who were *Advanced* was between approximately 5% and 11%.
- Across all grade levels, approximately 56% of students were below *Proficient*. The difference ranged from 51% below *Proficient* in grade 3 to 65% below *Proficient* in grade 8.

Science

- More than 50% of students were either *Proficient* or *Advanced* in Science.
- The percentage of students classified as *Advanced* was approximately 16% in both grades.
- The proportion of students classified as below *Proficient* was 48% in grade 4 and 49% in grade 8.

Social Studies

- About half or more of the total students in each grade level were either *Proficient* or *Advanced* in Social Studies. The proportion of *Proficient* or *Advanced* students was 54% in grade 4, 51% in grade 8, and 49% in grade 10.
- Approximately 20% of students were *Advanced*.
- The proportion of students classified as below *Proficient* was 46% in grade 4, 49% in grade 8, and 51% in grade 10.

Subgroup Patterns in Performance Level Results

The performance level results varied by subgroup: gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. The main subgroup performance patterns are described below. These comparisons are based on Tables 8-41 through 8-44.

In terms of gender, the percentages of both genders were approximately equal in *Proficient* or above, across grades and content areas. Although in every grade and content area except English Language Arts, there were higher percentages of male students who were classified as *Advanced*.

There were some consistent patterns in performance by ethnicity across grades and content areas. In terms of the *Proficient* or above category, the prevailing tendency was that there were higher percentages of White students as a group, followed by Asian students, American Indian students and Hispanic students, and African American students. The inverse sequence was found at the *Below Basic* performance level.

There were consistent differences in performance between economically disadvantaged students and not economically disadvantaged students. In every grade and content area, there were much higher percentages of students who were not economically disadvantaged classified as *Proficient* or above. There were much higher percentages of students who were economically disadvantaged who were classified in the lowest performance category.

Performance level results showed that there were higher percentages of students without disabilities who were classified as *Proficient* or above, and 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.

Performance level results showed a similar pattern in comparisons of students who were fully English proficient with students who were limited English proficient. In every grade and content area, there were generally higher percentages of students who were fully English proficient classified as *Proficient* and much higher percentages of students who were fully English proficient classified as *Advanced.* There were much lower percentages of fully English proficient students who were classified in the lowest performance category in all grades and content areas.

8.5 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 or objectives). 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, in that it estimates how much a student knows in a clearly defined skill domain (i.e., the criterion). Technical readers can refer to Appendix G 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 percent 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, and 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.* However, comparisons across years, grades, or content standards are appropriate for assessing the relative difficulty of the items, and comparisons of individual student scores or of group mean scores on a single SPI score can provide useful information about the *relative* strengths and needs of individual students or groups on these standards.

Tables 8-45 through 8-48 identify the content standards/domain, the number of MC and CR items within each standard/domain, the total number of possible points per standard/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 8-45 through 8-48 are summarized below. Tables 8-49 through 8-52 identify the SPI cut scores for each content area reporting category and grade level.

English Language Arts

Tables 8-45a and 8-45b present mean *p*-values and SPI scores for English Language Arts across content standards/domains and grades. The mean of the mean ELA SPI scores across grades was 58.10 for content standards and 55.64 for domains, indicating that the items were moderately difficult for examinees. Results show that the mean *p*-values and SPI scores varied across standards in all grades. Mean SPI scores ranged from 29.02 to 80.55 for content standards and from 41.92 to 69.09 for domains. In general, the difference between the lowest and highest mean SPI scores was greatest among content standards at grade 6 (46.83). The difference was

smallest among domains at Grade g (12.39). Content standard D (Writing/Language - Text Types and Purposes) was the most difficult in grade 3, and content Standard E (Writing/Language - Research) was the most difficult in grades 4 and above.

Mathematics

Table 8-46 presents Mathematics *p*-values and SPI scores across grades and content standards. The mean of the mean Mathematics SPI scores across grades and content standards was 45.44, indicating that the items were somewhat challenging. Results show that the mean *p*-values and SPI scores varied across standards in all grades. Mean SPI scores ranged from 28.88 to 63.80, with the largest difference observed in grade 8 (where SPI scores ranged from 28.88 to 54.54). Differences between the highest and lowest mean SPI scores ranged from 12.44 (grade 6) to 25.66 (grade 8). Content standard D (Measurement and Data) was the most difficult in grades 3 and 4. Content standard E (Geometry) was the most difficult in grades 5 and 8. Content standard H (Expressions and Equations) was the most difficult in grade 7, and content standard G (The Number System) was the most difficult in grade 8.

Science

Table 8-47 presents Science *p*-values and SPI scores across grades and content standards. The mean of the mean Science SPI scores across grades and content standards was 71.94, indicating that the test items were relatively easy. Across all grades and content standards, mean SPI scores ranged from 55.42 to 82.89, with differences between the highest and lowest mean SPI scores of 27.47 in grade 4 and 18.38 in grade 8. The mean *p*-values and SPI scores indicated that content standard E (Earth and Space Science) was the most difficult in both grades.

Social Studies

Table 8-48 presents Social Studies *p*-values and SPI scores across grades and content standards. The mean of the mean Social Studies SPI scores across grades and content standards was 67.10, indicating that the test items were relatively easy. Mean SPI scores ranged from 55.33 to 76.79, with differences between the highest and lowest mean SPI scores of 21.46 in grade 4, 11.10 in grade 8, and 12.36 in grade 10. The mean *p*-values and SPI scores indicated that the most difficult content standard varied between the three Social Studies grades. In grades 4 and 10, the most difficult standard was content standard D (Economics) and in grade 8 the most difficult standard was content standard E (The Behavioral Sciences).

Summary of Student Performance Indicator Results

Overall, the mean SPI scores across grades and content standards range in difficulty. There are, however, a few instances of high SPI scores (>75):

- Grade 4 ELA content standard F (Writing/Language - Language Conventions)
- Grade 6 ELA content standard D (Writing/Language - Text Types and Purposes).
- Grades 7 and 8 ELA content standard C (Reading - Vocabulary Use)

- Grades 4 and 8 Science content standard G/H (Science Applications & Personal Social Perspectives)
- Grade 8 Science content standard A/B (Science Connections & Nature of Science) and content standard C (Science Inquiry)
- Grade 4 Social Studies content standard C (Political Science and Citizenship) and content standard E (The Behavioral Sciences)

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.

Summary of Student Achievement Results

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 8, together with the reliability and validity evidence, 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. As such, these test scores can be used to classify students, schools, districts, and the state with respect to how much achievement is shown for each content area. Classroom teachers may use these scores as evidence of student achievement in these content areas. District and school administrators may use this information for activities such as planning curricula. At the state level, the overall results can be drawn upon for accountability and reporting purposes.

Table 8-A Summary of Flagged Operational Items on the Spring 2016 Wisconsin Forward Exam

Content	Grade	# of Items Flagged	Number of Flags*			
			Correlation <0.15	Distractor Correlation >0	Omit or Insufficient to Score >5%*	*p*-Value <0.30
ELA	3	4	2	3	1	
	4	6	1	5	1	
	5	8	1	4	1	3
	6	5	1	3	1	2
	7	2	1	1	1	
	8	2		1	1	1
MA	3	8		2		7
	4	8	1	2		7
	5	14	1	4		11
	6	15	2	3		11
	7	18	1	5		16
	8	17	1	6		12
SC	4	4		4		
	8	1	1	1		
SS	4	1		1		
	8	3		3		
	10	1		1		
Total		117	13	49	6	70

Note: The number of flags may be greater than the number of flagged items.
* all flagged items are TDA items.

Table 8-B English Language Arts Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
3	ELA	10	MC	0.32	0.12	0.30%	+	+	0.04		
	ELA	15	MC	0.39	0.24	0.40%		+	0.04		
	ELA	16	TDA	0.35	0.32	29.66%*				+	
	ELA	19	MC	0.35	0.08	0.13%	+	+	0.12		
4	ELA	5	MC	0.30	-0.03	0.16%	+	+	0.10		
	ELA	16	MC	0.44	0.30	0.27%		+	0.01		
	ELA	17	TDA	0.32	0.25	49.14%*				+	
	ELA	18	MC	0.39	0.23	0.09%		+	0.00		
	ELA	22	MC	0.60	0.20	0.13%		+	0.01		
	ELA	27	MC	0.43	0.25	0.15%		+	0.02		
5	ELA	4	MC	0.35	0.09	0.14%	+	+	0.19		
	ELA	7	TE	0.28	0.18	0.22%					+
	ELA	11	MC	0.44	0.20	0.23%		+	0.08		
	ELA	12	ESR	0.22	0.28	0.18%					+
	ELA	16	TDA	0.35	0.31	30.86%*				+	
	ELA	23	ESR	0.16	0.17	0.12%					+
	ELA	28	MC	0.39	0.19	0.21%		+	0.06		
	ELA	30	MC	0.32	0.25	0.09%		+	0.04		
6	ELA	5	MC	0.41	0.09	0.25%	+	+	0.09		
	ELA	18	TDA	0.33	0.26	31.53%*				+	
	ELA	27	TE	0.25	0.25	0.28%					+
	ELA	30	MC	0.39	0.17	0.23%		+	0.11		
	ELA	32	MC	0.26	0.20	0.12%		+	0.05		+
7	ELA	17	TDA	0.36	0.32	23.23%*				+	
	ELA	32	MC	0.40	0.10	0.13%	+	+	0.13		
8	ELA	17	TDA	0.37	0.38	25.52%*				+	
	ELA	26	MC	0.28	0.23	0.20%		+	0.00		+

Note: An asterisk (*) indicates a combined percentage of “omit” and “insufficient to score” for a TDA item.

Table 8-C Mathematics Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
3	MA	12	TE	0.20	0.39	0.14%					+
	MA	16	MC	0.16	0.25	0.22%					+
	MA	18	SA	0.28	0.38	0.21%					+
	MA	30	MC	0.35	0.16	0.20%		+	0.03		
	MA	33	TE	0.08	0.25	0.19%					+
	MA	34	SA	0.26	0.29	0.19%					+
	MA	36	MC	0.18	0.19	0.38%		+	0.07		+
	MA	41	SA	0.08	0.29	0.36%					+
4	MA	6	TE	0.13	0.18	0.07%					+
	MA	16	SA	0.24	0.44	0.19%					+
	MA	20	TE	0.24	0.55	0.70%					+
	MA	29	TE	0.27	0.46	0.17%					+
	MA	34	SA	0.09	0.37	0.33%					+
	MA	35	TE	0.08	0.38	1.32%					+
	MA	43	MC	0.25	0.16	0.22%		+	0.11		+
	MA	46	MC	0.54	0.03	0.20%	+	+	0.07		
5	MA	2	ESR	0.15	0.51	0.09%					+
	MA	3	MC	0.58	0.19	0.14%		+	0.00		
	MA	11	MC	0.64	0.13	0.16%	+				
	MA	12	TE	0.14	0.46	0.56%					+
	MA	16	ESR	0.06	0.25	0.35%					+
	MA	25	MC	0.26	0.20	0.11%		+	0.08		+
	MA	26	ESR	0.16	0.46	0.14%					+
	MA	27	SA	0.13	0.45	0.53%					+
	MA	28	SA	0.15	0.46	0.23%					+
	MA	35	TE	0.24	0.50	1.42%					+
	MA	37	ESR	0.11	0.28	0.23%					+
	MA	38	MC	0.41	0.26	0.26%		+	0.10		
	MA	41	MC	0.27	0.38	0.32%		+	0.01		+
	MA	46	ESR	0.07	0.29	0.30%					+

Table 8-C Mathematics Items Flagged for Classical Item Analysis Statistics (cont.)

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
6	MA	2	MC	0.16	0.44	0.56%					+
	MA	9	TE	0.11	0.43	1.61%					+
	MA	10	MC	0.34	0.15	1.03%		+	0.24		
	MA	13	SA	0.20	0.49	0.99%					+
	MA	14	TE	0.12	0.43	1.46%					+
	MA	22	ESR	0.15	0.41	0.65%					+
	MA	24	SA	0.24	0.60	0.17%					+
	MA	26	MC	0.51	0.24	0.23%		+	0.00		
	MA	27	TE	0.19	0.48	0.22%					+
	MA	28	MC	0.33	0.34	0.15%		+	0.04		
	MA	29	ESR	0.07	0.23	0.28%					+
	MA	33	SA	0.12	0.50	0.77%					+
	MA	41	MC	0.34	0.13	0.32%	+				
	MA	42	SA	0.26	0.52	0.75%					+
	MA	44	ESR	0.04	0.13	0.26%	+				+
7	MA	1	MC	0.21	0.41	0.55%		+	0.01		+
	MA	4	TE	0.20	0.42	0.59%					+
	MA	5	ESR	0.21	0.49	0.51%					+
	MA	7	SA	0.14	0.48	2.94%					+
	MA	11	MC	0.32	0.24	1.10%		+	0.07		
	MA	15	TE	0.14	0.39	0.73%					+
	MA	16	MC	0.29	0.49	0.38%					+
	MA	19	SA	0.23	0.60	1.04%					+
	MA	21	ESR	0.18	0.40	0.58%					+
	MA	22	MC	0.28	0.23	0.56%					+
	MA	27	TE	0.11	0.35	0.28%					+
	MA	29	ESR	0.14	0.33	0.27%					+
	MA	30	SA	0.29	0.58	0.44%					+
	MA	33	MC	0.24	-0.01	0.30%	+	+	0.25		+
	MA	37	TE	0.13	0.51	0.62%					+
	MA	39	MC	0.30	0.23	0.42%		+	0.04		
	MA	40	ESR	0.11	0.48	0.43%					+
	MA	45	MC	0.29	0.34	0.44%		+	0.11		+

Table 8-C Mathematics Items Flagged for Classical Item Analysis Statistics (cont.)

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
8	MA	1	MC	0.22	0.15	0.76%	+				+
	MA	4	ESR	0.05	0.27	0.55%					+
	MA	5	MC	0.47	0.49	0.62%		+	0.03		
	MA	6	SA	0.08	0.36	1.60%					+
	MA	9	MC	0.41	0.29	0.66%		+	0.03		
	MA	10	MC	0.39	0.23	0.74%		+	0.08		
	MA	11	SA	0.05	0.31	2.71%					+
	MA	13	SA	0.17	0.50	2.05%					+
	MA	18	SA	0.08	0.38	0.70%					+
	MA	19	ESR	0.09	0.30	0.37%					+
	MA	21	TE	0.09	0.37	0.71%					+
	MA	25	MC	0.21	0.19	0.14%		+	0.02		+
	MA	26	SA	0.21	0.48	1.45%					+
	MA	40	ESR	0.17	0.42	0.56%					+
	MA	41	MC	0.37	0.28	0.55%		+	0.00		
	MA	42	MC	0.32	0.17	0.43%		+	0.14		
	MA	44	SA	0.24	0.57	1.64%					+

Table 8-D Science & Social Studies Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
4	SC	14	MC	0.39	0.20	0.10%		+	0.01		
	SC	24	MC	0.49	0.17	0.11%		+	0.06		
	SC	25	MC	0.37	0.20	0.10%		+	0.02		
	SC	35	MC	0.33	0.20	0.11%		+	0.12		
8	SC	40	MC	0.43	0.14	0.17%	+	+	0.08		
4	SS	21	MC	0.50	0.35	0.11%		+	0.01		
8	SS	20	MC	0.51	0.34	0.15%		+	0.03		
	SS	24	MC	0.52	0.40	0.15%		+	0.03		
	SS	35	MC	0.38	0.18	0.23%		+	0.06		
10	SS	22	MC	0.40	0.29	0.12%		+	0.02		

Table 8-E Percentage of Students Attempting Last Operational Item in Test

Content	Grade						
	3	4	5	6	7	8	10
English Language Arts	99.75%	99.87%	99.80%	99.74%	99.78%	99.79%	
Mathematics	99.77%	99.80%	99.70%	99.71%	99.56%	99.57%	
Science		99.83%				99.83%	
Social Studies		99.85%				99.77%	99.71%

Table 8-1 Item Analysis, Grade 3 English Language Arts

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.68	0.48	0.12%				
2	MC	0.66	0.42	0.20%				
3	ESR	0.43	0.44	0.12%				
4	MC	0.52	0.43	0.21%				
5	MC	0.66	0.35	0.23%				
6	MC	0.57	0.31	0.24%				
7	ESR	0.64	0.54	0.16%				
8	TE	0.47	0.57	0.25%				
9	MC	0.47	0.33	0.27%				
10	MC	0.32	0.12	0.30%	+	+		
11	TE	0.70	0.53	0.45%				
12	ESR	0.41	0.54	0.29%				
13	MC	0.41	0.32	0.38%				
14	MC	0.54	0.38	0.39%				
15	MC	0.39	0.24	0.40%		+		
16	TDA	0.35	0.32	29.66%*			+	
17	MC	0.66	0.31	0.09%				
18	MC	0.54	0.33	0.16%				
19	MC	0.35	0.08	0.13%	+	+		
20	TE	0.59	0.36	0.53%				
21	TE	0.37	0.22	0.45%				
22	MC	0.64	0.30	0.18%				
23	MC	0.58	0.42	0.18%				
24	MC	0.80	0.45	0.20%				
25	TE	0.94	0.34	0.21%				
26	MC	0.53	0.30	0.22%				
27	MC	0.51	0.29	0.20%				
28	MC	0.47	0.40	0.16%				
29	ESR	0.73	0.46	0.10%				
30	MC	0.69	0.49	0.17%				
31	MC	0.71	0.42	0.18%				
32	MC	0.60	0.34	0.22%				
33	MC	0.62	0.42	0.24%				
34	MC	0.73	0.41	0.25%				

Note: An asterisk (*) indicates a combined percentage of “omit” and “insufficient to score” for a TDA item.

Table 8-2 Item Analysis, Grade 4 English Language Arts

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TE	0.62	0.52	0.10%				
2	TE	0.80	0.45	1.01%				
3	MC	0.48	0.28	0.12%				
4	ESR	0.56	0.57	0.10%				
5	MC	0.30	-0.03	0.16%	+	+		
6	TE	0.57	0.49	0.16%				
7	MC	0.79	0.41	0.17%				
8	MC	0.87	0.43	0.14%				
9	MC	0.62	0.39	0.23%				
10	MC	0.68	0.51	0.18%				
11	MC	0.72	0.50	0.26%				
12	TE	0.32	0.49	1.14%				
13	MC	0.57	0.50	0.26%				
14	TE	0.72	0.49	0.25%				
15	MC	0.51	0.37	0.27%				
16	MC	0.44	0.30	0.27%		+		
17	TDA	0.32	0.25	49.14%*			+	
18	MC	0.39	0.23	0.09%		+		
19	TE	0.62	0.44	0.23%				
20	MC	0.82	0.34	0.10%				
21	TE	0.37	0.27	0.24%				
22	MC	0.60	0.20	0.13%		+		
23	MC	0.75	0.47	0.15%				
24	TE	0.83	0.35	0.12%				
25	TE	0.91	0.38	0.11%				
26	TE	0.52	0.41	0.15%				
27	MC	0.43	0.25	0.15%		+		
28	TE	0.94	0.42	0.15%				
29	MC	0.55	0.42	0.17%				
30	TE	0.66	0.48	0.08%				
31	MC	0.58	0.35	0.10%				
32	MC	0.55	0.32	0.12%				
33	MC	0.60	0.44	0.13%				
34	MC	0.60	0.38	0.15%				
35	ESR	0.72	0.41	0.13%				

Note: An asterisk (*) indicates a combined percentage of "omit" and "insufficient to score" for a TDA item.

Table 8-3 Item Analysis, Grade 5 English Language Arts

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.65	0.32	0.06%				
2	TE	0.35	0.20	0.65%				
3	MC	0.91	0.34	0.08%				
4	MC	0.35	0.09	0.14%	+	+		
5	MC	0.85	0.38	0.10%				
6	MC	0.80	0.38	0.11%				
7	TE	0.28	0.18	0.22%				+
8	MC	0.77	0.39	0.12%				
9	ESR	0.82	0.48	0.08%				
10	MC	0.74	0.48	0.13%				
11	MC	0.44	0.20	0.23%		+		
12	ESR	0.22	0.28	0.18%				+
13	MC	0.61	0.40	0.24%				
14	MC	0.65	0.54	0.27%				
15	ESR	0.41	0.33	0.20%				
16	TDA	0.35	0.31	30.86%*			+	
17	MC	0.66	0.22	0.07%				
18	TE	0.83	0.37	0.14%				
19	MC	0.70	0.29	0.10%				
20	MC	0.82	0.44	0.12%				
21	TE	0.48	0.37	0.15%				
22	TE	0.48	0.46	0.43%				
23	ESR	0.16	0.17	0.12%				+
24	TE	0.56	0.42	0.17%				
25	MC	0.48	0.28	0.17%				
26	MC	0.75	0.44	0.19%				
27	MC	0.70	0.45	0.19%				
28	MC	0.39	0.19	0.21%		+		
29	MC	0.64	0.40	0.22%				
30	MC	0.32	0.25	0.09%		+		
31	MC	0.54	0.31	0.14%				
32	ESR	0.41	0.44	0.10%				
33	MC	0.70	0.44	0.19%				
34	MC	0.67	0.48	0.18%				
35	MS	0.57	0.49	0.20%				

Note: An asterisk (*) indicates a combined percentage of “omit” and “insufficient to score” for a TDA item.

Table 8-4 Item Analysis, Grade 6 English Language Arts

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	ESR	0.65	0.50	0.05%				
2	MC	0.52	0.40	0.13%				
3	MC	0.55	0.34	0.16%				
4	MC	0.60	0.37	0.13%				
5	MC	0.41	0.09	0.25%	+	+		
6	MC	0.66	0.51	0.16%				
7	MC	0.58	0.39	0.18%				
8	MC	0.53	0.33	0.15%				
9	MC	0.69	0.49	0.24%				
10	TE	0.69	0.47	0.96%				
11	MC	0.53	0.31	0.27%				
12	MC	0.48	0.38	0.33%				
13	MC	0.44	0.34	0.31%				
14	MC	0.39	0.32	0.32%				
15	MC	0.77	0.44	0.31%				
16	MC	0.75	0.42	0.31%				
17	TE	0.46	0.38	0.70%				
18	TDA	0.33	0.26	31.53%*			+	
19	MC	0.81	0.20	0.07%				
20	MC	0.87	0.38	0.10%				
21	MC	0.73	0.41	0.10%				
22	MC	0.67	0.27	0.13%				
23	TE	0.62	0.27	0.21%				
24	TE	0.34	0.21	0.27%				
25	TE	0.56	0.48	0.37%				
26	MC	0.81	0.39	0.20%				
27	TE	0.25	0.25	0.28%				+
28	ESR	0.45	0.37	0.21%				
29	MC	0.56	0.40	0.22%				
30	MC	0.39	0.17	0.23%		+		

Table 8-4 Item Analysis, Grade 6 English Language Arts (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
31	ESR	0.72	0.50	0.07%				
32	MC	0.26	0.20	0.12%		+		+
33	MC	0.79	0.45	0.16%				
34	ESR	0.36	0.28	0.13%				
35	MC	0.48	0.31	0.19%				
36	MC	0.63	0.40	0.26%				

Note: An asterisk (*) indicates a combined percentage of “omit” and “insufficient to score” for a TDA item.

Table 8-5 Item Analysis, Grade 7 English Language Arts

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.63	0.32	0.08%				
2	MC	0.64	0.37	0.14%				
3	MC	0.83	0.46	0.12%				
4	ESR	0.68	0.46	0.07%				
5	TE	0.91	0.41	0.22%				
6	MC	0.73	0.35	0.15%				
7	MC	0.76	0.44	0.17%				
8	MC	0.77	0.46	0.15%				
9	MC	0.89	0.43	0.26%				
10	MC	0.67	0.38	0.24%				
11	MC	0.35	0.21	0.28%				
12	MC	0.68	0.48	0.27%				
13	TE	0.72	0.48	4.57%				
14	MC	0.69	0.25	0.32%				
15	TE	0.63	0.50	0.32%				
16	ESR	0.54	0.53	0.27%				
17	TDA	0.36	0.32	23.23%*			+	
18	TE	0.64	0.29	0.36%				
19	MC	0.51	0.25	0.14%				
20	MC	0.45	0.32	0.17%				
21	MC	0.69	0.32	0.18%				
22	MC	0.79	0.37	0.15%				
23	TE	0.66	0.27	0.41%				
24	MC	0.80	0.48	0.17%				
25	MC	0.78	0.38	0.19%				
26	MC	0.51	0.38	0.20%				
27	MC	0.65	0.37	0.19%				
28	TE	0.72	0.59	0.21%				
29	MC	0.57	0.28	0.22%				
30	ESR	0.49	0.46	0.19%				

Table 8-5 Item Analysis, Grade 7 English Language Arts (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
31	MC	0.73	0.39	0.10%				
32	MC	0.40	0.10	0.13%	+	+		
33	ESR	0.66	0.56	0.12%				
34	MC	0.90	0.38	0.18%				
35	TE	0.73	0.48	0.17%				
36	MC	0.66	0.43	0.23%				

Note: An asterisk (*) indicates a combined percentage of “omit” and “insufficient to score” for a TDA item.

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

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.75	0.50	0.09%				
2	MC	0.50	0.22	0.15%				
3	MC	0.69	0.40	0.14%				
4	MC	0.53	0.25	0.19%				
5	MC	0.71	0.48	0.15%				
6	MC	0.88	0.43	0.17%				
7	ESR	0.76	0.51	0.10%				
8	MC	0.90	0.34	0.14%				
9	MC	0.61	0.39	0.21%				
10	MC	0.63	0.49	0.25%				
11	TE	0.55	0.37	1.99%				
12	TE	0.49	0.42	0.32%				
13	TE	0.73	0.58	0.48%				
14	MC	0.88	0.44	0.30%				
15	TE	0.70	0.50	0.30%				
16	MC	0.68	0.49	0.30%				
17	TDA	0.37	0.38	25.52%			+	
18	MC	0.61	0.35	0.07%				
19	MC	0.61	0.40	0.13%				
20	MC	0.49	0.24	0.14%				
21	MC	0.64	0.36	0.15%				
22	MC	0.46	0.42	0.15%				
23	TE	0.53	0.48	0.80%				
24	MC	0.76	0.47	0.17%				
25	MC	0.71	0.38	0.23%				
26	MC	0.28	0.23	0.20%		+		+
27	TE	0.95	0.39	0.39%				
28	TE	0.52	0.37	1.19%				
29	MC	0.53	0.25	0.25%				

Table 8-6 Item Analysis, Grade 8 English Language Arts (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
30	TE	0.54	0.40	0.25%				
31	MC	0.59	0.28	0.09%				
32	MC	0.87	0.41	0.12%				
33	ESR	0.54	0.51	0.10%				
34	MC	0.47	0.40	0.21%				
35	MC	0.72	0.43	0.22%				
36	ESR	0.64	0.48	0.21%				

Note: An asterisk (*) indicates a combined percentage of "omit" and "insufficient to score" for a TDA item.

Table 8-7 Item Analysis, Grade 3 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.64	0.45	0.14%				
2	MC	0.71	0.42	0.09%				
3	MC	0.75	0.48	0.15%				
4	MC	0.76	0.44	0.15%				
5	MC	0.52	0.44	0.13%				
6	TE	0.40	0.56	0.10%				
7	MC	0.82	0.42	0.12%				
8	SA	0.52	0.37	0.19%				
9	MC	0.64	0.37	0.16%				
10	SA	0.60	0.57	0.15%				
11	MC	0.45	0.36	0.17%				
12	TE	0.20	0.39	0.14%				+
13	MC	0.66	0.41	0.17%				
14	MC	0.41	0.27	0.20%				
15	MC	0.41	0.29	0.34%				
16	MC	0.16	0.25	0.22%				+
17	MC	0.41	0.45	0.21%				
18	SA	0.28	0.38	0.21%				+
19	MC	0.72	0.49	0.22%				
20	TE	0.86	0.32	1.13%				
21	MC	0.82	0.40	0.20%				
22	MC	0.47	0.37	0.13%				
23	MC	0.44	0.37	0.14%				
24	MC	0.62	0.44	0.20%				
25	MC	0.58	0.36	0.15%				
26	SA	0.60	0.57	0.17%				
27	TE	0.64	0.30	0.16%				
28	MC	0.64	0.42	0.17%				
29	MC	0.77	0.43	0.23%				
30	MC	0.35	0.16	0.20%		+		
31	MC	0.54	0.41	0.22%				
32	MC	0.41	0.38	0.21%				
33	TE	0.08	0.25	0.19%				+
34	SA	0.26	0.29	0.19%				+
35	MC	0.62	0.48	0.24%				

Table 8-7 Item Analysis, Grade 3 Mathematics (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.18	0.19	0.38%		+		+
37	MC	0.68	0.48	0.25%				
38	SA	0.67	0.55	0.24%				
39	MC	0.79	0.40	0.23%				
40	MC	0.77	0.36	0.25%				
41	SA	0.08	0.29	0.36%				+
42	MC	0.70	0.47	0.23%				

Table 8-8 Item Analysis, Grade 4 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.32	0.49	0.10%				
2	MC	0.45	0.33	0.09%				
3	MC	0.58	0.47	0.09%				
4	MC	0.58	0.54	0.11%				
5	MC	0.70	0.45	0.08%				
6	TE	0.13	0.18	0.07%				+
7	MC	0.63	0.42	0.09%				
8	MC	0.63	0.49	0.13%				
9	MC	0.75	0.37	0.11%				
10	MC	0.35	0.31	0.16%				
11	MC	0.59	0.37	0.13%				
12	TE	0.69	0.36	0.12%				
13	MC	0.46	0.56	0.14%				
14	MC	0.41	0.27	0.16%				
15	MC	0.38	0.31	0.20%				
16	SA	0.24	0.44	0.19%				+
17	SA	0.41	0.56	0.20%				
18	MC	0.37	0.41	0.23%				
19	MC	0.58	0.31	0.20%				
20	TE	0.24	0.55	0.70%				+
21	MC	0.59	0.38	0.19%				
22	MC	0.82	0.28	0.21%				
23	MC	0.45	0.45	0.18%				
24	MC	0.81	0.39	0.13%				
25	MC	0.84	0.34	0.11%				
26	MC	0.47	0.39	0.13%				
27	MC	0.49	0.48	0.17%				
28	MC	0.36	0.58	0.13%				
29	TE	0.27	0.46	0.17%				+
30	SA	0.59	0.47	0.16%				
31	MC	0.31	0.36	0.21%				
32	MC	0.42	0.26	0.16%				
33	MC	0.53	0.59	0.16%				
34	SA	0.09	0.37	0.33%				+
35	TE	0.08	0.38	1.32%				+

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

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.70	0.39	0.19%				
37	SA	0.43	0.49	0.24%				
38	MC	0.65	0.42	0.21%				
39	MC	0.39	0.53	0.27%				
40	MC	0.47	0.46	0.23%				
41	MC	0.30	0.52	0.26%				
42	SA	0.42	0.58	0.28%				
43	MC	0.25	0.16	0.22%		+		+
44	SA	0.42	0.52	0.27%				
45	MC	0.53	0.38	0.24%				
46	MC	0.54	0.03	0.20%	+	+		

Table 8-9 Item Analysis, Grade 5 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.80	0.32	0.10%				
2	ESR	0.15	0.51	0.09%				+
3	MC	0.58	0.19	0.14%		+		
4	MC	0.33	0.38	0.13%				
5	MC	0.66	0.40	0.09%				
6	TE	0.31	0.59	0.21%				
7	ESR	0.32	0.57	0.13%				
8	MC	0.43	0.35	0.19%				
9	MC	0.69	0.45	0.13%				
10	MC	0.77	0.45	0.18%				
11	MC	0.64	0.13	0.16%	+			
12	TE	0.14	0.46	0.56%				+
13	SA	0.37	0.53	0.21%				
14	SA	0.44	0.52	0.36%				
15	MC	0.55	0.43	0.24%				
16	ESR	0.06	0.25	0.35%				+
17	MC	0.43	0.32	0.30%				
18	MC	0.41	0.46	0.38%				
19	MC	0.43	0.50	0.32%				
20	TE	0.61	0.44	0.34%				
21	MC	0.53	0.54	0.30%				
22	SA	0.47	0.56	0.35%				
23	MC	0.67	0.38	0.33%				
24	MC	0.55	0.33	0.13%				
25	MC	0.26	0.20	0.11%		+		+
26	ESR	0.16	0.46	0.14%				+
27	SA	0.13	0.45	0.53%				+
28	SA	0.15	0.46	0.23%				+
29	TE	0.45	0.50	0.32%				
30	SA	0.54	0.51	0.26%				
31	MC	0.57	0.30	0.22%				
32	MC	0.50	0.45	0.23%				
33	MC	0.52	0.57	0.21%				
34	SA	0.34	0.48	0.26%				
35	TE	0.24	0.50	1.42%				+

Table 8-9 Item Analysis, Grade 5 Mathematics (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.52	0.48	0.26%				
37	ESR	0.11	0.28	0.23%				+
38	MC	0.41	0.26	0.26%		+		
39	SA	0.45	0.19	0.39%				
40	MC	0.62	0.45	0.31%				
41	MC	0.27	0.38	0.32%		+		+
42	MC	0.41	0.41	0.30%				
43	MC	0.56	0.45	0.35%				
44	MC	0.79	0.26	0.30%				
45	MC	0.67	0.37	0.32%				
46	ESR	0.07	0.29	0.30%				+

Table 8-10 Item Analysis, Grade 6 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.53	0.45	0.54%				
2	MC	0.16	0.44	0.56%				+
3	MC	0.84	0.36	0.54%				
4	TE	0.58	0.49	2.04%				
5	MC	0.55	0.34	0.71%				
6	SA	0.40	0.61	1.95%				
7	MC	0.93	0.34	0.58%				
8	MC	0.94	0.29	0.69%				
9	TE	0.11	0.43	1.61%				+
10	MC	0.34	0.15	1.03%		+		
11	SA	0.41	0.49	1.17%				
12	MC	0.39	0.29	1.36%				
13	SA	0.20	0.49	0.99%				+
14	TE	0.12	0.43	1.46%				+
15	MC	0.65	0.31	1.08%				
16	MC	0.68	0.46	1.17%				
17	MC	0.73	0.43	0.62%				
18	MC	0.67	0.37	0.64%				
19	SA	0.66	0.51	0.75%				
20	MC	0.34	0.49	0.67%				
21	MC	0.68	0.30	0.71%				
22	ESR	0.15	0.41	0.65%				+
23	MC	0.47	0.39	0.62%				
24	SA	0.24	0.60	0.17%				+
25	MC	0.91	0.31	0.15%				
26	MC	0.51	0.24	0.23%		+		
27	TE	0.19	0.48	0.22%				+
28	MC	0.33	0.34	0.15%		+		
29	ESR	0.07	0.23	0.28%				+
30	ESR	0.43	0.36	0.14%				
31	MC	0.68	0.47	0.21%				
32	MC	0.62	0.45	0.18%				
33	SA	0.12	0.50	0.77%				+
34	MC	0.45	0.47	0.24%				
35	MC	0.65	0.48	0.24%				

Table 8-10 Item Analysis, Grade 6 Mathematics (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.47	0.45	0.23%				
37	TE	0.38	0.48	0.37%				
38	MC	0.38	0.26	0.26%				
39	MC	0.41	0.26	0.34%				
40	MC	0.36	0.54	0.31%				
41	MC	0.34	0.13	0.32%	+			
42	SA	0.26	0.52	0.75%				+
43	MC	0.65	0.39	0.39%				
44	ESR	0.04	0.13	0.26%	+			+
45	MC	0.60	0.46	0.32%				
46	MC	0.36	0.46	0.30%				

Table 8-11 Item Analysis, Grade 7 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.21	0.41	0.55%		+		+
2	MC	0.46	0.53	0.55%				
3	SA	0.58	0.58	0.95%				
4	TE	0.20	0.42	0.59%				+
5	ESR	0.21	0.49	0.51%				+
6	MC	0.38	0.42	0.75%				
7	SA	0.14	0.48	2.94%				+
8	MC	0.49	0.36	0.61%				
9	MC	0.55	0.31	0.83%				
10	MC	0.51	0.39	0.95%				
11	MC	0.32	0.24	1.10%		+		
12	MC	0.51	0.42	0.28%				
13	MC	0.61	0.47	0.33%				
14	SA	0.61	0.52	0.48%				
15	TE	0.14	0.39	0.73%				+
16	MC	0.29	0.49	0.38%				+
17	MC	0.33	0.46	0.45%				
18	MC	0.42	0.21	0.47%				
19	SA	0.23	0.60	1.04%				+
20	TE	0.66	0.23	1.10%				
21	ESR	0.18	0.40	0.58%				+
22	MC	0.28	0.23	0.56%				+
23	MC	0.65	0.29	0.50%				
24	MC	0.82	0.39	0.20%				
25	MC	0.33	0.31	0.18%				
26	MC	0.66	0.38	0.18%				
27	TE	0.11	0.35	0.28%				+
28	MC	0.52	0.43	0.21%				
29	ESR	0.14	0.33	0.27%				+
30	SA	0.29	0.58	0.44%				+
31	MC	0.65	0.52	0.27%				
32	MC	0.56	0.17	0.24%				
33	MC	0.24	-0.01	0.30%	+	+		+
34	MC	0.52	0.50	0.27%				
35	SA	0.39	0.53	0.33%				

Table 8-11 Item Analysis, Grade 7 Mathematics (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.58	0.33	0.30%				
37	TE	0.13	0.51	0.62%				+
38	MC	0.54	0.23	0.35%				
39	MC	0.30	0.23	0.42%		+		
40	ESR	0.11	0.48	0.43%				+
41	SA	0.50	0.23	0.86%				
42	MC	0.37	0.44	0.43%				
43	MC	0.56	0.50	0.44%				
44	SA	0.33	0.62	0.89%				
45	MC	0.29	0.34	0.44%		+		+
46	MC	0.44	0.33	0.44%				

Table 8-12 Item Analysis, Grade 8 Mathematics

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.22	0.15	0.76%	+			+
2	SA	0.36	0.61	2.64%				
3	MC	0.51	0.39	0.42%				
4	ESR	0.05	0.27	0.55%				+
5	MC	0.47	0.49	0.62%		+		
6	SA	0.08	0.36	1.60%				+
7	MC	0.36	0.25	0.64%				
8	MC	0.49	0.49	0.50%				
9	MC	0.41	0.29	0.66%		+		
10	MC	0.39	0.23	0.74%		+		
11	SA	0.05	0.31	2.71%				+
12	MC	0.45	0.34	0.83%				
13	SA	0.17	0.50	2.05%				+
14	MC	0.36	0.39	0.28%				
15	MC	0.63	0.27	0.35%				
16	TE	0.49	0.49	0.57%				
17	MC	0.59	0.43	0.35%				
18	SA	0.08	0.38	0.70%				+
19	ESR	0.09	0.30	0.37%				+
20	MC	0.54	0.34	0.42%				
21	TE	0.09	0.37	0.71%				+
22	MC	0.42	0.44	0.46%				
23	MC	0.39	0.27	0.41%				
24	MC	0.67	0.37	0.24%				
25	MC	0.21	0.19	0.14%		+		+
26	SA	0.21	0.48	1.45%				+
27	TE	0.46	0.47	0.84%				
28	MC	0.55	0.54	0.23%				
29	MC	0.49	0.22	0.27%				
30	ESR	0.31	0.35	0.24%				
31	SA	0.31	0.57	1.40%				
32	TE	0.35	0.47	0.96%				
33	MC	0.65	0.31	0.34%				
34	MC	0.53	0.31	0.31%				
35	MC	0.74	0.44	0.34%				

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

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.56	0.47	0.40%				
37	TE	0.34	0.36	0.76%				
38	MC	0.78	0.41	0.39%				
39	MC	0.40	0.33	0.51%				
40	ESR	0.17	0.42	0.56%				+
41	MC	0.37	0.28	0.55%		+		
42	MC	0.32	0.17	0.43%		+		
43	MC	0.69	0.50	0.48%				
44	SA	0.24	0.57	1.64%				+
45	MC	0.69	0.34	0.46%				
46	MC	0.59	0.44	0.43%				

Table 8-13 Item Analysis, Grade 4 Science

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.77	0.38	0.07%				
2	MC	0.54	0.37	0.13%				
3	MC	0.87	0.46	0.09%				
4	MC	0.92	0.19	0.08%				
5	MC	0.94	0.32	0.08%				
6	MC	0.79	0.43	0.09%				
7	MC	0.76	0.46	0.08%				
8	MC	0.79	0.48	0.09%				
9	MC	0.92	0.43	0.08%				
10	MC	0.63	0.37	0.17%				
11	MC	0.87	0.39	0.08%				
12	MC	0.80	0.36	0.12%				
13	MC	0.82	0.44	0.08%				
14	MC	0.39	0.20	0.10%		+		
15	MC	0.57	0.37	0.09%				
16	MC	0.79	0.28	0.09%				
17	MC	0.72	0.44	0.10%				
18	MC	0.56	0.42	0.16%				
19	MC	0.49	0.41	0.14%				
20	MC	0.86	0.33	0.13%				
21	MC	0.89	0.25	0.05%				
22	MC	0.84	0.41	0.11%				
23	MC	0.65	0.36	0.07%				
24	MC	0.49	0.17	0.11%		+		
25	MC	0.37	0.20	0.10%		+		
26	MC	0.61	0.35	0.14%				
27	MC	0.62	0.38	0.11%				
28	MC	0.94	0.34	0.10%				
29	MC	0.57	0.31	0.08%				
30	MC	0.74	0.32	0.13%				
31	MC	0.80	0.42	0.10%				
32	MC	0.86	0.29	0.12%				
33	MC	0.45	0.24	0.11%				
34	MC	0.71	0.36	0.15%				
35	MC	0.33	0.20	0.11%		+		

Table 8-13 Item Analysis, Grade 4 Science (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.80	0.44	0.11%				
37	MC	0.61	0.42	0.14%				
38	MC	0.46	0.34	0.15%				
39	MC	0.56	0.44	0.13%				
40	MC	0.54	0.37	0.17%				

Table 8-14 Item Analysis, Grade 8 Science

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.87	0.40	0.03%				
2	MC	0.89	0.38	0.08%				
3	MC	0.91	0.42	0.07%				
4	MC	0.93	0.41	0.09%				
5	MC	0.98	0.34	0.08%				
6	MC	0.79	0.38	0.10%				
7	MC	0.96	0.35	0.09%				
8	MC	0.87	0.31	0.08%				
9	MC	0.87	0.39	0.05%				
10	MC	0.80	0.43	0.12%				
11	MC	0.85	0.40	0.08%				
12	MC	0.78	0.28	0.09%				
13	MC	0.93	0.34	0.08%				
14	MC	0.78	0.36	0.09%				
15	MC	0.80	0.33	0.07%				
16	MC	0.64	0.28	0.07%				
17	MC	0.68	0.42	0.10%				
18	MC	0.96	0.34	0.14%				
19	MC	0.88	0.48	0.11%				
20	MC	0.68	0.43	0.12%				
21	MC	0.72	0.36	0.04%				
22	MC	0.56	0.22	0.12%				
23	MC	0.75	0.45	0.11%				
24	MC	0.74	0.47	0.15%				
25	MC	0.57	0.26	0.08%				
26	MC	0.79	0.37	0.11%				
27	MC	0.78	0.49	0.11%				
28	MC	0.60	0.30	0.12%				
29	MC	0.65	0.44	0.11%				
30	MC	0.64	0.32	0.12%				
31	MC	0.72	0.43	0.14%				
32	MC	0.86	0.45	0.16%				
33	MC	0.76	0.40	0.12%				
34	MC	0.67	0.47	0.14%				
35	MC	0.47	0.28	0.11%				

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

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.73	0.44	0.12%				
37	MC	0.76	0.44	0.14%				
38	MC	0.70	0.38	0.20%				
39	MC	0.64	0.21	0.18%				
40	MC	0.43	0.14	0.17%	+	+		

Table 8-15 Item Analysis, Grade 4 Social Studies

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.90	0.33	0.04%				
2	MC	0.83	0.43	0.07%				
3	MC	0.56	0.39	0.07%				
4	MC	0.52	0.39	0.10%				
5	MC	0.84	0.34	0.06%				
6	MC	0.75	0.38	0.09%				
7	MC	0.86	0.40	0.07%				
8	MC	0.85	0.36	0.08%				
9	MC	0.87	0.32	0.07%				
10	MC	0.61	0.46	0.09%				
11	MC	0.82	0.46	0.08%				
12	MC	0.65	0.32	0.09%				
13	MC	0.77	0.42	0.12%				
14	MC	0.79	0.50	0.20%				
15	MC	0.52	0.36	0.08%				
16	MC	0.80	0.46	0.10%				
17	MC	0.88	0.45	0.08%				
18	MC	0.82	0.33	0.09%				
19	MC	0.75	0.43	0.10%				
20	MC	0.45	0.34	0.06%				
21	MC	0.50	0.35	0.11%		+		
22	MC	0.77	0.44	0.07%				
23	MC	0.60	0.48	0.11%				
24	MC	0.70	0.42	0.09%				
25	MC	0.56	0.46	0.16%				
26	MC	0.82	0.43	0.10%				
27	MC	0.86	0.42	0.16%				
28	MC	0.70	0.42	0.08%				
29	MC	0.66	0.36	0.12%				
30	MC	0.59	0.32	0.14%				
31	MC	0.62	0.47	0.15%				
32	MC	0.44	0.30	0.15%				
33	MC	0.47	0.38	0.26%				
34	MC	0.51	0.40	0.14%				
35	MC	0.84	0.44	0.16%				

Table 8-15 Item Analysis, Grade 4 Social Studies (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.51	0.33	0.17%				
37	MC	0.84	0.47	0.16%				
38	MC	0.70	0.37	0.15%				

Table 8-16 Item Analysis, Grade 8 Social Studies

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.83	0.38	0.07%				
2	MC	0.87	0.42	0.09%				
3	MC	0.86	0.40	0.12%				
4	MC	0.83	0.41	0.13%				
5	MC	0.92	0.37	0.10%				
6	MC	0.80	0.42	0.13%				
7	MC	0.60	0.21	0.10%				
8	MC	0.58	0.36	0.10%				
9	MC	0.84	0.38	0.08%				
10	MC	0.76	0.47	0.14%				
11	MC	0.81	0.42	0.12%				
12	MC	0.86	0.49	0.17%				
13	MC	0.78	0.43	0.09%				
14	MC	0.69	0.42	0.14%				
15	MC	0.62	0.29	0.12%				
16	MC	0.69	0.54	0.20%				
17	MC	0.70	0.50	0.13%				
18	MC	0.73	0.41	0.17%				
19	MC	0.61	0.55	0.12%				
20	MC	0.51	0.34	0.15%		+		
21	MC	0.65	0.47	0.08%				
22	MC	0.67	0.42	0.16%				
23	MC	0.45	0.31	0.14%				
24	MC	0.52	0.40	0.15%		+		
25	MC	0.75	0.45	0.12%				
26	MC	0.68	0.56	0.14%				
27	MC	0.57	0.39	0.16%				
28	MC	0.73	0.37	0.16%				
29	MC	0.42	0.38	0.16%				
30	MC	0.85	0.45	0.18%				
31	MC	0.69	0.40	0.19%				
32	MC	0.86	0.47	0.20%				
33	MC	0.59	0.38	0.20%				
34	MC	0.73	0.53	0.23%				
35	MC	0.38	0.18	0.23%		+		

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

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.56	0.45	0.24%				
37	MC	0.72	0.39	0.22%				
38	MC	0.64	0.45	0.23%				
39	MC	0.35	0.34	0.21%				
40	MC	0.72	0.47	0.23%				

Table 8-17 Item Analysis, Grade 10 Social Studies

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.68	0.31	0.06%				
2	MC	0.56	0.34	0.05%				
3	MC	0.83	0.24	0.03%				
4	MC	0.87	0.40	0.04%				
5	MC	0.70	0.28	0.04%				
6	MC	0.65	0.28	0.04%				
7	MC	0.65	0.42	0.07%				
8	MC	0.77	0.36	0.08%				
9	MC	0.83	0.38	0.05%				
10	MC	0.57	0.40	0.06%				
11	MC	0.51	0.28	0.05%				
12	MC	0.63	0.34	0.07%				
13	MC	0.46	0.27	0.05%				
14	MC	0.76	0.31	0.06%				
15	MC	0.83	0.42	0.06%				
16	MC	0.70	0.48	0.10%				
17	MC	0.55	0.45	0.15%				
18	MC	0.72	0.50	0.08%				
19	MC	0.75	0.40	0.11%				
20	MC	0.74	0.49	0.10%				
21	MC	0.66	0.39	0.12%				
22	MC	0.40	0.29	0.12%		+		
23	MC	0.31	0.36	0.10%				
24	MC	0.46	0.24	0.11%				
25	MC	0.55	0.30	0.10%				
26	MC	0.87	0.29	0.06%				
27	MC	0.71	0.45	0.16%				
28	MC	0.65	0.44	0.13%				
29	MC	0.69	0.48	0.13%				
30	MC	0.78	0.48	0.14%				
31	MC	0.66	0.45	0.19%				
32	MC	0.60	0.37	0.10%				
33	MC	0.51	0.46	0.23%				
34	MC	0.38	0.42	0.15%				
35	MC	0.45	0.38	0.14%				

Table 8-17 Item Analysis, Grade 10 Social Studies (cont.)

Item	Item Type	*p*-Value	Corr	Omit Rate	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.70	0.48	0.16%				
37	MC	0.81	0.50	0.16%				
38	MC	0.61	0.37	0.15%				
39	MC	0.51	0.39	0.25%				
40	MC	0.60	0.51	0.16%				
41	MC	0.67	0.50	0.17%				
42	MC	0.58	0.43	0.23%				
43	MC	0.80	0.51	0.20%				
44	MC	0.96	0.30	0.20%				
45	MC	0.56	0.36	0.19%				
46	MC	0.60	0.42	0.26%				
47	MC	0.50	0.31	0.23%				
48	MC	0.67	0.46	0.20%				
49	MC	0.69	0.37	0.18%				
50	MC	0.81	0.36	0.29%				

Table 8-18 Raw Score Descriptive Statistics

Content	Grade	N Count	Mean N of Items Correct	Test Difficulty	SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
English Language Arts	3	61120	26.72	0.57	9.39	-0.09	-0.87	1	50	53	0.87	3.37
	4	59776	30.26	0.61	8.94	-0.19	-0.61	0	56	56	0.87	3.16
	5	59662	26.97	0.57	8.79	-0.18	-0.65	0	52	56	0.85	3.40
	6	60164	27.53	0.56	9.30	-0.16	-0.68	0	53	56	0.86	3.50
	7	59539	32.75	0.66	9.83	-0.41	-0.55	0	56	56	0.88	3.37
	8	59006	31.63	0.63	10.14	-0.33	-0.59	0	56	56	0.89	3.36
Mathematics	3	61220	22.76	0.54	8.37	-0.09	-0.83	0	42	42	0.90	2.67
	4	59855	21.91	0.47	9.63	0.31	-0.85	0	46	46	0.91	2.85
	5	59733	20.21	0.44	9.38	0.44	-0.66	1	46	46	0.91	2.80
	6	60220	20.96	0.46	8.93	0.36	-0.62	0	46	46	0.91	2.72
	7	59600	18.32	0.40	9.05	0.62	-0.35	0	46	46	0.90	2.82
	8	59076	18.32	0.40	8.71	0.50	-0.44	0	46	46	0.90	2.81
Science	4	59832	27.60	0.69	6.97	-0.61	-0.24	2	40	40	0.87	2.53
	8	59000	30.18	0.76	6.94	-1.03	0.60	1	40	40	0.88	2.40
Social Studies	4	59817	26.42	0.70	7.48	-0.59	-0.48	1	38	38	0.89	2.45
	8	59024	27.17	0.69	8.20	-0.55	-0.62	0	40	40	0.90	2.55
	10	62203	32.42	0.65	10.02	-0.32	-0.78	1	50	50	0.91	2.98

Table 8-19 Raw Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
English Language Arts	3	31319	25.73	0.56	9.28	0.87	29801	27.75	0.59	9.41	0.87
	4	30651	29.26	0.60	8.93	0.88	29125	31.32	0.63	8.83	0.87
	5	30543	25.62	0.56	8.59	0.85	29119	28.40	0.60	8.77	0.85
	6	30884	26.17	0.55	9.27	0.86	29280	28.97	0.59	9.12	0.85
	7	30623	31.33	0.65	9.92	0.89	28916	34.27	0.68	9.51	0.88
	8	30219	29.97	0.61	10.26	0.89	28787	33.38	0.66	9.71	0.88
Mathematics	3	31375	23.08	0.55	8.57	0.90	29845	22.41	0.53	8.15	0.89
	4	30687	22.51	0.49	9.89	0.92	29168	21.26	0.46	9.31	0.91
	5	30579	20.31	0.44	9.66	0.92	29154	20.10	0.44	9.08	0.90
	6	30911	20.82	0.46	9.21	0.91	29309	21.11	0.46	8.62	0.90
	7	30658	18.34	0.40	9.22	0.91	28942	18.30	0.40	8.87	0.90
	8	30248	18.19	0.40	8.96	0.90	28828	18.46	0.40	8.43	0.89
Science	4	30668	27.74	0.69	7.13	0.88	29164	27.45	0.69	6.79	0.86
	8	30216	30.09	0.75	7.27	0.89	28784	30.29	0.76	6.58	0.87
Social Studies	4	30654	26.30	0.69	7.65	0.90	29163	26.55	0.70	7.30	0.89
	8	30230	27.00	0.68	8.52	0.91	28794	27.34	0.69	7.85	0.90
	10	31744	32.72	0.66	10.37	0.92	30459	32.12	0.65	9.65	0.90

Table 8-20 Raw Score Descriptive Statistics for English Language Arts by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
White	3	42388	28.50	0.61	8.88	0.86
	4	41474	32.05	0.65	8.30	0.86
	5	42259	28.57	0.61	8.20	0.83
	6	43225	29.27	0.60	8.71	0.84
	7	43153	34.42	0.70	9.11	0.87
	8	42945	33.29	0.67	9.51	0.88
African American	3	5877	19.28	0.42	8.46	0.84
	4	5760	22.85	0.47	8.29	0.85
	5	5533	19.52	0.43	8.31	0.83
	6	5384	19.73	0.42	8.54	0.83
	7	5281	24.25	0.51	9.82	0.87
	8	5322	23.42	0.49	9.76	0.87
Hispanic	3	7451	22.99	0.49	8.92	0.85
	4	7361	26.80	0.54	8.64	0.86
	5	6917	23.52	0.50	8.42	0.83
	6	6738	23.24	0.48	8.87	0.84
	7	6624	28.92	0.59	9.68	0.87
	8	6331	27.66	0.56	10.00	0.88
Asian	3	2420	26.86	0.57	9.27	0.86
	4	2411	30.06	0.60	9.21	0.88
	5	2363	27.84	0.58	8.97	0.86
	6	2331	27.74	0.56	9.34	0.86
	7	2219	34.29	0.68	9.70	0.88
	8	2160	32.94	0.64	10.06	0.88
American Indian	3	757	22.97	0.50	8.90	0.85
	4	755	25.95	0.53	8.27	0.85
	5	764	22.91	0.50	7.82	0.81
	6	734	22.80	0.48	8.61	0.83
	7	729	28.28	0.58	9.28	0.86
	8	751	27.45	0.56	9.60	0.88
Two or More	3	2227	26.08	0.56	9.34	0.87
	4	2015	29.27	0.59	9.06	0.87
	5	1826	26.21	0.56	8.73	0.85
	6	1752	26.83	0.55	9.35	0.86
	7	1533	31.69	0.65	10.01	0.88
	8	1497	30.34	0.61	10.29	0.89

Table 8-1 Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
White	3	42369	24.52	0.58	7.88	0.89
	4	41479	24.02	0.52	9.28	0.91
	5	42255	22.01	0.48	9.18	0.91
	6	43220	22.64	0.49	8.68	0.90
	7	43138	19.97	0.44	8.93	0.90
	8	42935	19.85	0.43	8.51	0.89
African American	3	5882	15.77	0.38	7.11	0.85
	4	5765	13.89	0.30	6.78	0.83
	5	5537	12.47	0.27	6.33	0.83
	6	5385	13.31	0.30	6.35	0.83
	7	5289	10.88	0.24	5.57	0.78
	8	5339	11.05	0.25	5.86	0.80
Hispanic	3	7530	18.90	0.45	7.65	0.87
	4	7406	17.20	0.38	8.14	0.88
	5	6961	15.76	0.34	7.66	0.87
	6	6770	16.67	0.37	7.37	0.87
	7	6672	14.01	0.31	7.16	0.85
	8	6368	14.32	0.31	7.17	0.85
Asian	3	2456	22.70	0.54	8.60	0.91
	4	2439	22.22	0.48	10.33	0.93
	5	2395	21.63	0.47	10.10	0.93
	6	2358	22.09	0.48	9.21	0.91
	7	2247	19.53	0.43	10.27	0.93
	8	2186	20.00	0.44	9.79	0.92
American Indian	3	756	19.38	0.46	7.86	0.88
	4	753	16.67	0.36	7.82	0.87
	5	761	15.81	0.35	7.22	0.85
	6	736	16.33	0.36	7.20	0.86
	7	728	13.56	0.30	6.67	0.83
	8	750	14.29	0.31	6.96	0.84
Two or More	3	2227	21.87	0.52	8.34	0.90
	4	2013	20.28	0.44	9.29	0.91
	5	1824	18.71	0.41	9.09	0.91
	6	1751	19.86	0.44	8.83	0.90
	7	1526	16.83	0.37	8.68	0.90
	8	1498	16.86	0.37	8.42	0.89

Table 8-22 Raw Score Descriptive Statistics for Science by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
White	4	41478	29.26	0.73	6.10	0.84
	8	42911	31.69	0.79	5.96	0.85
African American	4	5753	20.77	0.52	7.06	0.84
	8	5304	22.96	0.58	7.70	0.87
Hispanic	4	7395	24.40	0.61	6.88	0.85
	8	6355	26.75	0.67	7.24	0.87
Asian	4	2438	26.70	0.67	7.03	0.86
	8	2186	29.89	0.75	6.53	0.86
American Indian	4	755	24.58	0.62	6.74	0.84
	8	750	27.06	0.68	7.45	0.88
Two or More	4	2013	26.81	0.67	7.02	0.87
	8	1494	29.12	0.73	7.17	0.88

Table 8-2 Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
White	4	41472	28.15	0.74	6.60	0.87
	8	42927	28.76	0.72	7.51	0.89
	10	46882	34.00	0.68	9.49	0.91
African American	4	5740	19.33	0.51	7.49	0.87
	8	5307	19.46	0.49	7.92	0.87
	10	5103	23.56	0.48	9.21	0.87
Hispanic	4	7400	23.17	0.61	7.51	0.88
	8	6358	23.62	0.59	8.00	0.88
	10	5863	28.19	0.57	9.47	0.89
Asian	4	2436	25.66	0.68	7.59	0.89
	8	2187	27.08	0.68	7.98	0.90
	10	2287	32.43	0.65	9.78	0.91
American Indian	4	754	22.84	0.60	7.54	0.88
	8	751	23.26	0.58	7.97	0.88
	10	705	27.49	0.56	9.81	0.89
Two or More	4	2015	25.27	0.67	7.56	0.89
	8	1494	25.93	0.65	8.35	0.90
	10	1363	32.19	0.65	10.16	0.91

Table 8-24 Raw Score Descriptive Statistics by Socioeconomic Status

		Economically Disadvantaged					Not Economically Disadvantaged				
Content	**Grade**	**N Count**	**Mean N of Items Correct**	**Test Difficulty**	**SD**	**Alpha**	**N Count**	**Mean N of Items Correct**	**Test Difficulty**	**SD**	**Alpha**
English Language Arts	3	26330	23.22	0.50	8.97	0.85	34790	29.37	0.62	8.82	0.86
	4	25632	26.72	0.54	8.58	0.86	34144	32.92	0.67	8.26	0.86
	5	24312	23.44	0.51	8.45	0.84	35350	29.40	0.62	8.17	0.83
	6	23726	23.55	0.49	8.81	0.84	36438	30.12	0.62	8.68	0.84
	7	22637	28.79	0.59	9.73	0.87	36902	35.18	0.71	9.07	0.87
	8	22131	27.55	0.56	9.95	0.88	36875	34.08	0.68	9.44	0.88
Mathematics	3	26384	19.39	0.46	7.89	0.88	34836	25.31	0.60	7.81	0.89
	4	25680	17.76	0.39	8.33	0.88	34175	25.02	0.54	9.37	0.91
	5	24352	16.24	0.35	7.97	0.88	35381	22.93	0.50	9.31	0.91
	6	23764	16.99	0.37	7.54	0.87	36456	23.55	0.51	8.81	0.90
	7	22669	14.37	0.32	7.24	0.85	36931	20.75	0.45	9.20	0.90
	8	22178	14.55	0.32	7.26	0.86	36898	20.58	0.45	8.72	0.90
Science	4	25665	24.84	0.62	7.09	0.86	34167	29.67	0.74	6.10	0.84
	8	22139	27.28	0.68	7.42	0.88	36861	31.93	0.80	6.00	0.86
Social Studies	4	25646	23.31	0.61	7.52	0.88	34171	28.76	0.76	6.54	0.87
	8	22144	23.60	0.59	8.23	0.89	36880	29.30	0.73	7.40	0.89
	10	20572	27.87	0.56	9.75	0.90	41631	34.67	0.70	9.37	0.91

Table 8-25 Raw Score Descriptive Statistics by Disability

Content	Grade	Disabled					Not Disabled				
		N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
English Language Arts	3	7461	20.33	0.45	8.84	0.86	53659	27.61	0.59	9.12	0.86
	4	7532	23.27	0.48	8.64	0.86	52244	31.27	0.63	8.52	0.86
	5	7462	19.22	0.43	8.00	0.83	52200	28.08	0.60	8.32	0.83
	6	7431	18.47	0.40	8.19	0.83	52733	28.81	0.59	8.73	0.84
	7	7321	22.77	0.48	9.18	0.86	52218	34.15	0.69	9.08	0.86
	8	7280	21.11	0.45	8.90	0.86	51726	33.11	0.66	9.40	0.87
Mathematics	3	7461	17.28	0.41	8.29	0.89	53759	23.52	0.56	8.10	0.89
	4	7533	15.75	0.34	8.43	0.89	52322	22.79	0.50	9.47	0.91
	5	7463	13.77	0.30	7.54	0.87	52270	21.12	0.46	9.26	0.91
	6	7419	13.45	0.30	7.04	0.86	52801	22.01	0.48	8.65	0.90
	7	7310	11.26	0.25	6.27	0.82	52290	19.31	0.42	8.94	0.90
	8	7291	10.96	0.24	5.86	0.80	51785	19.36	0.42	8.54	0.89
Science	4	7532	22.76	0.57	7.62	0.87	52300	28.29	0.71	6.58	0.86
	8	7265	23.09	0.58	7.87	0.87	51735	31.18	0.78	6.18	0.86
Social Studies	4	7512	21.04	0.56	7.98	0.89	52305	27.20	0.72	7.08	0.88
	8	7266	19.19	0.48	7.94	0.87	51758	28.28	0.71	7.60	0.89
	10	7018	22.83	0.46	9.11	0.87	55185	33.64	0.68	9.46	0.90

Table 8-26 Raw Score Descriptive Statistics by English Language Proficiency

Content	Grade	Limited English Proficient					Fully English Proficient				
		N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha	N Count	Mean N of Items Correct	Test Difficulty	SD	Alpha
English Language Arts	3	5418	21.68	0.46	8.24	0.82	55702	27.21	0.58	9.36	0.87
	4	4584	24.27	0.49	7.61	0.81	55192	30.76	0.63	8.86	0.87
	5	3194	19.13	0.41	6.86	0.74	56468	27.42	0.59	8.67	0.85
	6	2587	17.52	0.37	6.30	0.69	57577	27.98	0.58	9.16	0.85
	7	2458	23.09	0.47	8.10	0.81	57081	33.17	0.67	9.68	0.88
	8	2428	21.62	0.44	7.89	0.80	56578	32.06	0.64	10.00	0.89
Mathematics	3	5548	18.00	0.43	7.34	0.86	55672	23.23	0.55	8.32	0.90
	4	4677	15.31	0.33	7.08	0.84	55178	22.47	0.49	9.61	0.91
	5	3279	12.98	0.28	5.91	0.79	56454	20.63	0.45	9.38	0.91
	6	2666	13.02	0.29	5.43	0.77	57554	21.33	0.47	8.89	0.91
	7	2545	10.77	0.24	5.05	0.73	57055	18.66	0.41	9.04	0.90
	8	2504	11.13	0.24	5.24	0.75	56572	18.64	0.41	8.70	0.90
Science	4	4668	22.35	0.56	6.40	0.81	55164	28.04	0.70	6.83	0.87
	8	2503	22.41	0.56	6.64	0.81	56497	30.53	0.77	6.75	0.88
Social Studies	4	4668	20.94	0.55	6.92	0.85	55149	26.89	0.71	7.34	0.89
	8	2505	18.62	0.47	6.45	0.80	56519	27.54	0.69	8.07	0.90
	10	1732	21.17	0.43	7.44	0.81	60471	32.75	0.66	9.90	0.91

Table 8-27 Scale Score Descriptive Statistics

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
English Language Arts	3	61120	561.89	47.00	0.09	-0.05	330	761	330	900
	4	59776	583.93	49.19	0.06	0.16	340	930	340	930
	5	59662	600.71	50.83	0.02	0.48	350	876	350	940
	6	60164	611.40	52.06	-0.29	1.08	360	863	360	950
	7	59539	624.73	54.81	-0.11	0.37	370	960	370	960
	8	59006	638.02	57.26	-0.10	0.27	380	970	380	970
Mathematics	3	61220	555.56	46.18	-0.52	1.77	360	760	360	760
	4	59855	575.03	55.65	-0.89	1.59	405	800	405	800
	5	59733	600.82	49.95	-0.61	1.27	430	830	430	830
	6	60220	614.02	52.76	-0.46	0.90	440	870	440	870
	7	59600	628.74	57.10	-0.79	1.42	450	880	450	880
	8	59076	642.13	57.05	-0.94	1.57	470	890	470	890
Science	4	59832	400.16	51.19	-0.03	1.15	190	600	190	600
	8	59000	598.93	52.41	-0.05	1.64	390	770	390	770
Social Studies	4	59817	399.25	51.18	0.13	1.55	200	570	200	570
	8	59024	598.87	51.59	0.04	1.28	420	780	420	780
	10	62203	698.92	53.76	-0.32	1.79	490	890	490	890

Table 8-28 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	31319	556.92	45.94	330	761	29801	567.11	47.53	330	761
	4	30651	578.70	48.75	340	877	29125	589.43	49.06	370	930
	5	30543	593.09	48.83	350	818	29119	608.71	51.65	350	876
	6	30884	604.13	52.17	360	863	29280	619.07	50.84	360	863
	7	30623	616.67	54.60	370	960	28916	633.27	53.73	383	960
	8	30219	628.29	57.27	380	970	28787	648.23	55.44	380	970
Mathematics	3	31375	557.37	47.94	360	760	29845	553.64	44.18	360	760
	4	30687	577.56	57.32	405	800	29168	572.36	53.72	405	800
	5	30579	600.58	51.90	430	830	29154	601.07	47.83	430	830
	6	30911	612.63	54.73	440	870	29309	615.49	50.56	440	870
	7	30658	628.00	59.26	450	880	28942	629.53	54.72	450	880
	8	30248	640.08	59.85	470	890	28828	644.28	53.87	470	890
Science	4	30668	400.91	53.23	190	600	29164	399.38	48.94	190	600
	8	30216	598.59	55.75	390	770	28784	599.28	48.66	390	770
Social Studies	4	30654	398.64	52.46	200	570	29163	399.89	49.78	200	570
	8	30230	598.05	54.53	420	780	28794	599.74	48.31	420	780
	10	31744	700.25	56.93	490	890	30459	697.53	50.19	490	890

Table 8-29 Scale Score Descriptive Statistics for English Language Arts by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	42388	570.75	44.55	330	761
	4	41474	593.82	46.16	351	930
	5	42259	610.00	47.47	350	876
	6	43225	621.20	47.97	360	863
	7	43153	633.89	51.25	370	960
	8	42945	647.22	53.98	380	970
African American	3	5877	524.97	42.40	330	700
	4	5760	543.52	43.85	378	723
	5	5533	557.64	49.17	350	790
	6	5384	566.99	52.37	360	771
	7	5281	578.08	53.67	370	817
	8	5322	592.23	54.28	380	789
Hispanic	3	7451	543.25	44.08	330	732
	4	7361	564.07	46.17	340	777
	5	6917	580.30	47.60	350	772
	6	6738	587.37	50.33	360	815
	7	6624	603.39	52.13	384	872
	8	6331	615.85	55.40	380	970
Asian	3	2420	562.27	46.86	399	719
	4	2411	582.27	50.77	442	788
	5	2363	605.64	53.21	423	813
	6	2331	612.79	51.30	378	847
	7	2219	633.67	56.05	398	960
	8	2160	645.87	59.48	465	970
American Indian	3	757	543.30	44.12	399	741
	4	755	560.71	43.79	440	697
	5	764	577.89	43.02	435	719
	6	734	585.45	49.70	360	780
	7	729	599.77	49.46	441	737
	8	751	615.17	53.19	380	874
Two or More	3	2227	558.92	46.68	330	734
	4	2015	579.04	49.96	340	814
	5	1826	596.71	49.55	431	790
	6	1752	607.50	51.60	400	824
	7	1533	618.64	55.70	370	832
	8	1497	630.85	57.48	431	883

Table 8-3 Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	42369	565.01	42.01	360	760
	4	41479	587.68	48.05	405	800
	5	42255	610.83	45.12	430	830
	6	43220	624.54	47.73	440	870
	7	43138	639.94	51.32	450	880
	8	42935	652.93	50.43	470	890
African American	3	5882	516.67	46.64	360	724
	4	5765	522.88	59.94	405	721
	5	5537	555.06	49.93	430	763
	6	5385	563.56	52.58	440	741
	7	5289	575.54	57.88	450	745
	8	5339	586.65	62.60	470	769
Hispanic	3	7530	535.51	44.19	360	724
	4	7406	548.84	56.88	405	721
	5	6961	578.16	48.48	430	830
	6	6770	588.99	49.75	440	858
	7	6672	601.46	56.68	450	880
	8	6368	617.07	57.22	470	778
Asian	3	2456	556.44	49.05	360	760
	4	2439	577.83	58.22	405	800
	5	2395	609.12	51.48	430	830
	6	2358	621.93	55.02	440	870
	7	2247	636.09	59.35	450	880
	8	2186	652.74	59.53	470	890
American Indian	3	756	538.18	45.50	360	760
	4	753	547.25	54.80	405	675
	5	761	579.34	45.38	430	686
	6	736	585.64	50.67	440	727
	7	728	598.83	55.79	450	747
	8	750	617.93	54.68	470	825
Two or More	3	2227	551.01	46.62	360	760
	4	2013	566.99	56.06	405	754
	5	1824	592.48	52.03	430	830
	6	1751	607.62	53.23	440	858
	7	1526	619.37	58.00	450	789
	8	1498	633.36	58.08	470	793

Table 8-31 Scale Score Descriptive Statistics for Science by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	41478	411.80	46.72	190	600
	8	42911	609.28	48.58	390	770
African American	4	5753	353.09	49.13	190	600
	8	5304	550.52	49.84	390	770
Hispanic	4	7395	377.25	47.08	190	600
	8	6355	574.81	48.19	390	770
Asian	4	2438	393.77	51.67	190	600
	8	2186	596.73	50.00	390	770
American Indian	4	755	378.53	45.35	190	529
	8	750	575.37	51.54	390	770
Two or More	4	2013	395.01	50.80	190	600
	8	1494	591.18	52.92	390	770

Table 8-42 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	41472	410.33	47.16	200	570
	8	42927	608.25	48.89	420	780
	10	46882	707.04	50.58	490	890
African American	4	5740	354.18	48.57	200	570
	8	5307	553.90	47.86	420	780
	10	5103	652.35	54.00	490	838
Hispanic	4	7400	377.90	46.89	200	570
	8	6358	577.77	46.04	420	780
	10	5863	677.60	50.58	490	890
Asian	4	2436	394.98	53.83	200	570
	8	2187	598.68	50.23	420	780
	10	2287	699.58	53.87	490	890
American Indian	4	754	375.81	46.15	200	570
	8	751	575.22	47.81	420	780
	10	705	673.81	53.41	490	890
Two or More	4	2015	391.95	51.19	200	570
	8	1494	591.02	50.46	420	780
	10	1363	697.47	54.15	490	890

Table 8-33 Scale Score Descriptive Statistics by Socioeconomic Status

Content	Grade	Economically Disadvantaged					Not Economically Disadvantaged				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
English Language Arts	3	26330	544.45	44.30	330	761	34790	575.08	44.63	330	761
	4	25632	564.31	45.58	340	816	34144	598.66	46.60	340	930
	5	24312	580.29	47.82	350	806	35350	614.75	48.00	386	876
	6	23726	589.35	49.62	360	815	36438	625.76	48.47	360	863
	7	22637	602.74	52.23	370	873	36902	638.22	51.92	370	960
	8	22131	615.32	54.73	380	970	36875	651.65	54.35	380	970
Mathematics	3	26384	537.53	45.58	360	760	34836	569.21	41.74	360	760
	4	25680	551.97	56.72	405	800	34175	592.35	48.04	405	800
	5	24352	580.03	49.46	430	830	35381	615.13	45.02	430	830
	6	23764	590.83	50.66	440	853	36456	629.14	48.43	440	870
	7	22669	604.26	56.47	450	799	36931	643.77	52.07	450	880
	8	22178	617.88	58.20	470	821	36898	656.71	51.07	470	890
Science	4	25665	380.34	48.77	190	600	34167	415.05	47.81	190	600
	8	22139	578.03	50.51	390	770	36861	611.48	49.45	390	770
Social Studies	4	25646	378.68	47.25	200	570	34171	414.69	48.51	200	570
	8	22144	577.51	48.17	420	780	36880	611.70	49.28	420	780
	10	20572	675.59	52.06	490	890	41631	710.44	50.76	490	890

Table 8-34 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	7461	529.59	44.33	330	723	53659	566.38	45.58	330	761
	4	7532	546.20	46.96	340	746	52244	589.37	47.08	340	930
	5	7462	556.42	47.01	350	780	52200	607.04	48.13	350	876
	6	7431	559.82	52.11	360	811	52733	618.67	47.77	360	863
	7	7321	570.49	50.05	370	794	52218	632.34	51.03	370	960
	8	7280	580.21	49.92	380	803	51726	646.16	53.41	380	970
Mathematics	3	7461	523.25	53.26	360	760	53759	560.04	43.25	360	760
	4	7533	532.30	65.86	405	754	52322	581.18	51.17	405	800
	5	7463	561.78	53.43	430	830	52270	606.39	46.86	430	830
	6	7419	562.18	55.87	440	802	52801	621.31	48.02	440	870
	7	7310	574.46	61.43	450	880	52290	636.33	52.15	450	880
	8	7291	585.35	61.42	470	778	51785	650.12	51.61	470	890
Science	4	7532	366.83	52.64	190	600	52300	404.96	49.15	190	600
	8	7265	550.61	52.31	390	770	51735	605.71	48.73	390	770
Social Studies	4	7512	365.29	51.62	200	570	52305	404.13	49.22	200	570
	8	7266	552.33	48.60	420	780	51758	605.41	48.55	420	780
	10	7018	648.86	54.61	490	890	55185	705.28	50.19	490	890

Table 8-35 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	5418	536.59	40.64	330	709	55702	564.35	46.85	330	761
	4	4584	549.91	39.48	340	712	55192	586.75	48.86	340	930
	5	3194	555.44	39.23	350	728	56468	603.27	50.20	350	876
	6	2587	556.29	40.61	360	692	57577	613.88	51.15	360	863
	7	2458	572.18	42.95	398	732	57081	626.99	54.13	370	960
	8	2428	583.11	44.00	380	816	56578	640.38	56.58	380	970
Mathematics	3	5548	530.70	44.16	360	760	55672	558.03	45.64	360	760
	4	4677	537.30	56.31	405	800	55178	578.22	54.41	405	800
	5	3279	561.67	46.54	430	721	56454	603.09	49.20	430	830
	6	2666	562.92	46.82	440	715	57554	616.39	51.81	440	870
	7	2545	574.80	55.49	450	772	57055	631.15	55.97	450	880
	8	2504	591.68	56.11	470	751	56572	644.36	56.05	470	890
Science	4	4668	362.92	42.61	190	518	55164	403.31	50.61	190	600
	8	2503	546.50	42.09	390	700	56497	601.25	51.61	390	770
Social Studies	4	4668	363.88	42.37	200	570	55149	402.24	50.73	200	570
	8	2505	550.78	37.34	420	780	56519	601.00	51.10	420	780
	10	1732	640.92	48.45	490	804	60471	700.58	52.97	490	890

Table 8-5 Performance Level Cut Scores for All Contents

Content	3			4			5			6			7			8			10		
	B	P	A	B	P	A	B	P	A	B	P	A	B	P	A	B	P	A	B	P	A
English Language Arts	522	570	624	546	592	650	564	610	670	572	622	671	585	638	697	592	652	708			
Mathematics	517	560	611	536	588	633	574	611	658	582	626	688	606	647	712	620	667	718			
Science				348	399	447										552	600	645			
Social Studies				363	396	436										563	599	640	670	703	741

Note: The abbreviation “B” is for the Basic performance level, “P” is for the Proficient performance level, and “A” is for the Advanced performance level.

Table 8-37 Cut Scores 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	20.94	34.91	35.00	9.14	44.15
4	340–545	546–591	592–649	650–930	21.95	33.77	35.48	8.81	44.29
5	350–563	564–609	610–669	670–940	22.35	34.41	35.11	8.14	43.25
6	360–571	572–621	622–670	671–950	20.47	36.06	32.23	11.23	43.46
7	370–584	585–637	638–696	697–960	22.52	34.80	34.60	8.08	42.68
8	380–591	592–651	652–707	708–970	20.79	37.06	31.64	10.50	42.14

Table 8-38 Cut Scores 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	17.68	33.14	39.74	9.44	49.18
4	405–535	536–587	588–632	633–800	18.68	35.87	34.17	11.28	45.45
5	430–573	574–610	611–657	658–830	24.97	29.85	34.88	10.30	45.18
6	440–581	582–625	626–687	688–870	24.58	31.49	37.65	6.28	43.94
7	450–605	606–646	647–711	712–880	29.46	30.32	35.62	4.59	40.22
8	470–619	620–666	667–717	718–890	27.64	37.67	28.79	5.90	34.69

Table 8-6Cut Scores and Associated Impact Data, Science

Grade	Score Range				Impact Data				
	Below Basic	Basic	Proficient	Advanced	Below Basic	Basic	Proficient	Advanced	Proficient + Advanced
4	190–347	348–398	399–446	447–600	14.05	33.48	36.32	16.15	52.47
8	390–551	552–599	600–644	645–770	15.62	33.87	34.89	15.63	50.51

Table 8-7Cut Scores 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	21.62	24.47	32.71	21.20	53.91
8	420–562	563–598	599–639	640–780	22.10	27.33	31.25	19.32	50.56
10	490–669	670–702	703–740	741–890	25.99	25.10	28.99	19.91	48.90

Table 8-8Percentage of Students in Each Performance Level by Subgroup, English Language Arts

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
3	BB	12800	20.94	18.05	23.69	14.24	51.68	33.12	20.50	33.82	22.86	19.39	36.86	47.70	17.22	32.49	12.20
	B	21337	34.91	33.43	36.32	34.03	33.20	39.67	36.74	39.89	36.51	34.16	42.58	33.28	35.14	39.14	31.71
	P	21394	35.00	37.13	32.98	40.63	13.46	22.99	32.36	21.80	32.33	36.63	18.29	16.50	37.58	24.26	43.13
	A	5589	9.14	11.39	7.01	11.10	1.67	4.21	10.41	4.49	8.31	9.81	2.27	2.52	10.07	4.11	12.95
Total		61120	100	29801	31319	42388	5877	7451	2420	757	2227	55702	5418	7461	53659	26330	34790
4	BB	13118	21.95	18.40	25.31	14.66	53.92	34.59	24.47	36.69	25.76	20.01	45.22	52.70	17.51	34.75	12.33
	B	20186	33.77	33.31	34.20	32.93	32.07	38.58	34.92	38.68	35.09	33.16	41.10	30.30	34.27	37.91	30.66
	P	21206	35.48	37.65	33.41	41.54	12.80	23.38	30.57	22.52	30.42	37.37	12.63	14.76	38.46	24.10	44.01
	A	5266	8.81	10.64	7.07	10.87	1.22	3.45	10.04	2.12	8.73	9.45	1.05	2.24	9.76	3.23	12.99
Total		59776	100	29125	30651	41474	5760	7361	2411	755	2015	55192	4584	7532	52244	25632	34144
5	BB	13332	22.35	18.49	26.02	15.45	56.68	35.16	20.78	37.17	25.25	20.42	56.32	57.76	17.28	35.64	13.21
	B	20527	34.41	31.80	36.89	34.27	29.32	38.90	33.09	39.40	35.49	34.28	36.54	29.48	35.11	37.74	32.11
	P	20948	35.11	38.56	31.83	40.52	12.33	22.64	34.70	21.86	32.20	36.70	6.95	11.61	38.47	23.57	43.05
	A	4855	8.14	11.16	5.26	9.76	1.68	3.30	11.43	1.57	7.06	8.59	0.19	1.15	9.14	3.05	11.64
Total		59662	100	29119	30543	42259	5533	6917	2363	764	1826	56468	3194	7462	52200	24312	35350

Table 8-9Percentage of Students in Each Performance Level by Subgroup, English Language Arts (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
6	BB	12318	20.47	16.03	24.68	13.78	52.28	35.71	19.61	38.69	22.89	18.56	62.97	59.83	14.93	33.92	11.72
	B	21698	36.07	35.17	36.92	35.24	34.53	40.62	40.11	39.51	36.70	36.15	34.13	29.32	37.01	40.67	33.06
	P	19389	32.23	34.58	30.00	37.36	11.18	19.58	27.71	17.98	30.94	33.55	2.74	9.15	35.48	21.34	39.32
	A	6759	11.23	14.22	8.40	13.62	2.01	4.10	12.57	3.82	9.47	11.73	0.15	1.70	12.58	4.07	15.90
Total		60164	100	29280	30884	43225	5384	6738	2331	734	1752	57577	2587	7431	52733	23726	36438
7	BB	13406	22.52	17.82	26.95	16.23	55.75	35.95	18.34	37.17	26.03	20.86	61.07	62.70	16.88	35.70	14.43
	B	20719	34.80	33.75	35.79	34.80	30.01	38.01	34.29	40.19	35.49	34.89	32.59	27.56	35.81	38.22	32.70
	P	20601	34.60	37.80	31.58	39.42	12.86	22.60	35.06	20.58	31.70	35.82	6.18	8.76	38.22	23.21	41.59
	A	4813	8.08	10.63	5.68	9.55	1.38	3.44	12.30	2.06	6.78	8.42	0.16	0.98	9.08	2.88	11.28
Total		59539	100	28916	30623	43153	5281	6624	2219	729	1533	57081	2458	7321	52218	22637	36902
8	BB	12269	20.79	15.19	26.13	15.19	49.68	33.11	17.82	32.36	25.25	19.22	57.33	61.74	15.03	33.51	13.16
	B	21869	37.06	36.31	37.78	36.41	36.88	40.59	37.50	42.34	38.28	37.05	37.27	29.93	38.07	40.58	34.95
	P	18672	31.64	34.75	28.68	35.93	11.63	21.94	30.19	21.97	27.86	32.78	5.15	7.23	35.08	21.66	37.64
	A	6196	10.50	13.75	7.41	12.47	1.80	4.36	14.49	3.33	8.62	10.94	0.25	1.10	11.82	4.25	14.25
Total		59006	100	28787	30219	42945	5322	6331	2160	751	1497	56578	2428	7280	51726	22131	36875

Table 8-10 Percentage of Students in Each Performance Level by Subgroup, Mathematics

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
3	BB	10821	17.68	18.09	17.28	10.96	47.14	30.52	16.98	29.23	21.10	16.16	32.89	42.49	14.23	28.68	9.34
	B	20288	33.14	34.86	31.50	30.96	36.74	40.68	35.50	39.29	34.85	32.12	43.39	33.16	33.14	39.33	28.45
	P	24329	39.74	39.03	40.41	46.28	14.93	25.62	36.24	27.65	36.51	41.56	21.45	20.80	42.37	28.40	48.33
	A	5782	9.45	8.02	10.80	11.80	1.19	3.19	11.28	3.84	7.54	10.16	2.27	3.55	10.26	3.59	13.88
Total		61220	100	29845	31375	42369	5882	7530	2456	756	2227	55672	5548	7461	53759	26384	34836
4	BB	11183	18.68	19.14	18.25	10.94	53.04	32.58	18.57	34.53	22.95	16.90	39.77	46.09	14.74	31.07	9.37
	B	21468	35.87	38.27	33.58	34.39	34.85	42.91	36.94	42.50	39.54	35.14	44.45	34.02	36.13	41.92	31.32
	P	20453	34.17	33.57	34.74	40.43	10.95	21.15	30.59	20.72	29.06	35.86	14.26	15.88	36.80	23.11	42.48
	A	6751	11.28	9.02	13.43	14.25	1.16	3.36	13.90	2.26	8.45	12.11	1.52	4.01	12.33	3.89	16.83
Total		59855	100	29168	30687	41479	5765	7406	2439	753	2013	55178	4677	7533	52322	25680	34175
5	BB	14915	24.97	23.98	25.91	17.03	63.32	41.11	21.29	39.03	29.88	23.20	55.38	57.28	20.36	39.69	14.84
	B	17832	29.85	30.92	28.83	29.41	25.54	34.29	29.85	36.40	33.55	29.62	33.94	26.08	30.39	33.16	27.58
	P	20832	34.88	35.84	33.95	40.91	9.93	21.43	33.82	22.60	28.51	36.32	9.94	13.88	37.87	23.64	42.61
	A	6154	10.30	9.25	11.31	12.65	1.21	3.16	15.03	1.97	8.06	10.86	0.73	2.76	11.38	3.51	14.98
Total		59733	100	29154	30579	42255	5537	6961	2395	761	1824	56454	3279	7463	52270	24352	35381

Table 8-11 Percentage of Students in Each Performance Level by Subgroup, Mathematics (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
6	BB	14800	24.58	22.19	26.84	16.97	62.12	41.62	19.04	45.79	29.41	22.74	64.33	64.55	18.96	39.91	14.58
	B	18962	31.49	33.15	29.91	31.10	27.39	36.03	34.27	33.83	31.41	31.60	29.14	23.02	32.68	35.32	28.99
	P	22674	37.65	39.06	36.32	44.32	9.92	20.55	36.81	18.89	33.64	39.11	6.19	11.21	41.37	23.07	47.16
	A	3784	6.28	5.61	6.93	7.61	0.58	1.80	9.88	1.50	5.54	6.56	0.34	1.21	7.00	1.70	9.27
Total		60220	100	29309	30911	43220	5385	6770	2358	736	1751	57554	2666	7419	52801	23764	36456
7	BB	17561	29.47	28.64	30.25	20.94	69.96	49.15	27.86	49.73	36.89	27.62	70.92	70.63	23.71	46.39	19.08
	B	18070	30.32	31.61	29.10	31.13	21.91	31.29	30.53	33.52	30.34	30.64	23.03	19.55	31.82	31.88	29.36
	P	21232	35.62	35.47	35.77	42.42	7.90	18.53	31.82	15.93	29.42	36.95	5.85	8.96	39.35	20.70	44.79
	A	2737	4.59	4.29	4.88	5.51	0.23	1.03	9.79	0.82	3.34	4.79	0.20	0.86	5.11	1.03	6.78
Total		59600	100	28942	30658	43138	5289	6672	2247	728	1526	57055	2545	7310	52290	22669	36931
8	BB	16329	27.64	25.52	29.66	19.72	67.86	45.18	24.06	46.53	32.44	25.93	66.33	69.72	21.72	44.03	17.79
	B	22252	37.67	39.05	36.35	39.17	25.15	38.46	34.81	38.67	39.39	38.05	28.99	23.99	39.59	37.99	37.47
	P	17008	28.79	30.17	27.47	34.07	6.56	14.67	29.37	13.07	23.63	29.88	4.27	5.71	32.04	16.44	36.22
	A	3487	5.90	5.27	6.51	7.03	0.43	1.70	11.76	1.73	4.54	6.15	0.40	0.59	6.65	1.55	8.52
Total		59076	100	28828	30248	42935	5339	6368	2186	750	1498	56572	2504	7291	51785	22178	36898

Table 8-43 Percentage of Students in Each Performance Level by Subgroup, Science

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
4	BB	8406	14.05	13.43	14.64	7.55	44.85	24.50	16.49	22.52	15.40	12.47	32.73	35.69	10.93	23.47	6.97
	B	20031	33.48	34.32	32.68	30.22	38.52	43.80	38.15	45.43	38.25	32.19	48.65	38.29	32.79	41.29	27.61
	P	21730	36.32	37.35	35.34	41.93	14.24	25.72	31.79	25.83	32.14	37.97	16.77	19.42	38.75	27.89	42.65
	A	9665	16.15	14.91	17.34	20.30	2.40	5.98	13.58	6.23	14.21	17.36	1.84	6.60	17.53	7.34	22.77
Total		59832	100	29164	30668	41478	5753	7395	2438	755	2013	55164	4668	7532	52300	25665	34167
8	BB	9214	15.62	13.88	17.27	9.56	47.40	28.37	14.27	27.07	18.74	14.03	51.34	50.03	10.78	26.78	8.91
	B	19983	33.87	35.80	32.03	31.26	38.67	43.34	40.39	44.00	36.81	33.55	40.99	34.22	33.82	40.73	29.75
	P	20583	34.89	36.00	33.82	40.02	11.99	22.17	32.07	22.53	33.07	36.12	7.15	12.50	38.03	25.22	40.69
	A	9220	15.63	14.32	16.88	19.16	1.94	6.12	13.27	6.40	11.38	16.30	0.52	3.25	17.37	7.28	20.64
Total		59000	100	28784	30216	42911	5304	6355	2186	750	1494	56497	2503	7265	51735	22139	36861

Table 8-44 Percentage of Students in Each Performance Level by Subgroup, Social Studies

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES	
		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
4	BB	12931	21.62	20.47	22.71	13.49	57.30	35.61	25.00	36.87	26.10	19.56	45.97	48.60	17.74	34.76	11.76
	B	14638	24.47	24.87	24.09	23.04	24.41	29.59	28.82	30.64	27.74	23.73	33.27	24.83	24.42	29.42	20.76
	P	19568	32.71	33.77	31.71	37.22	14.04	25.53	27.05	23.87	29.68	34.01	17.42	19.06	34.67	26.12	37.66
	A	12680	21.20	20.89	21.49	26.25	4.25	9.27	19.13	8.62	16.48	22.71	3.34	7.51	23.16	9.71	29.82
Total		59817	100	29163	30654	41472	5740	7400	2436	754	2015	55149	4668	7512	52305	25646	34171
8	BB	13046	22.10	20.16	23.95	15.46	56.81	35.55	20.16	37.15	27.78	20.38	60.88	60.87	16.66	36.19	13.64
	B	16133	27.33	28.85	25.89	25.94	27.57	34.11	31.69	34.75	27.51	27.11	32.26	24.08	27.79	31.97	24.55
	P	18444	31.25	32.76	29.80	35.20	12.55	22.35	29.26	20.37	30.32	32.36	6.11	11.02	34.09	23.44	35.94
	A	11401	19.32	18.23	20.35	23.40	3.07	7.99	18.88	7.72	14.39	20.14	0.76	4.02	21.46	8.39	25.88
Total		59024	100	28794	30230	42927	5307	6358	2187	751	1494	56519	2505	7266	51758	22144	36880
10	BB	16169	25.99	25.70	26.28	20.10	61.55	40.37	24.75	43.83	26.63	24.62	74.08	66.03	20.90	42.17	18.00
	B	15616	25.11	27.00	23.29	24.74	22.63	29.03	27.55	26.24	25.31	25.29	18.71	20.13	25.74	27.77	23.79
	P	18031	28.99	29.97	28.05	31.79	12.23	22.31	28.51	22.55	28.10	29.65	6.00	9.93	31.41	21.63	32.62
	A	12387	19.91	17.33	22.39	23.37	3.59	8.29	19.20	7.38	19.96	20.45	1.21	3.90	21.95	8.43	25.59
Total		62203	100	30459	31744	46882	5103	5863	2287	705	1363	60471	1732	7018	55185	20572	41631

Table 8-45a 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	61120	A	Reading - Key Ideas and Details	4	3	10	5.29	0.53	2.65	53.10	23.36
	61120	B	Reading - Craft & Structure	4	2	8	4.02	0.48	1.95	50.29	20.28
	61120	C	Reading - Vocabulary Use	2	0	2	1.26	0.62	0.74	n/a*	n/a*
	61120	D	Writing/Language - Text Types and Purposes	3	2	16	5.12	0.50	2.92	32.76	15.15
	61120	E	Writing/Language - Research	3	1	5	2.79	0.60	1.25	55.42	17.54
	61120	F	Writing/Language - Language Conventions	3	1	5	3.40	0.61	1.13	67.51	16.26
	61120	G	Listening	5	1	7	4.84	0.68	1.77	68.19	21.66
4	59776	A	Reading - Key Ideas and Details	2	4	10	5.97	0.58	2.41	59.83	21.22
	59776	B	Reading - Craft & Structure	5	0	5	2.92	0.58	1.27	58.39	18.61
	59776	C	Reading - Vocabulary Use	3	2	5	3.19	0.64	1.31	63.82	21.10
	59776	D	Writing/Language - Text Types and Purposes	2	2	4	2.21	0.55	1.08	55.67	17.50
	59776	E	Writing/Language - Research	2	3	18	5.92	0.55	2.86	33.90	12.17
	59776	F	Writing/Language - Language Conventions	2	2	6	4.94	0.77	1.07	80.55	13.74
	59776	G	Listening	4	2	8	5.11	0.62	2.03	63.77	20.43

* SPI scores are not computed for content standards measured by fewer than four items.

Table 8-45a 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	59662	A	Reading - Key Ideas and Details	1	4	9	4.47	0.53	1.66	49.72	14.09
	59662	B	Reading - Craft & Structure	6	1	8	4.44	0.59	1.75	55.40	17.72
	59662	C	Reading - Vocabulary Use	3	0	3	2.06	0.69	0.74	n/a*	n/a*
	59662	D	Writing/Language - Text Types and Purposes	3	2	5	3.50	0.70	1.23	69.28	18.55
	59662	E	Writing/Language - Research	0	4	18	5.34	0.39	3.24	30.35	14.82
	59662	F	Writing/Language - Language Conventions	5	0	5	2.96	0.59	1.35	59.23	20.21
	59662	G	Listening	4	2	8	4.20	0.53	2.12	52.79	22.15
6	60164	A	Reading - Key Ideas and Details	5	2	9	4.85	0.53	2.28	54.04	21.15
	60164	B	Reading - Craft & Structure	7	1	9	5.61	0.61	2.19	62.28	21.55
	60164	C	Reading - Vocabulary Use	2	0	2	1.06	0.53	0.75	n/a*	n/a*
	60164	D	Writing/Language - Text Types and Purposes	4	0	4	3.09	0.77	0.98	75.85	17.97
	60164	E	Writing/Language - Research	2	3	18	5.10	0.40	3.08	29.02	13.84
	60164	F	Writing/Language - Language Conventions	1	3	6	3.50	0.58	1.34	58.23	16.55
	60164	G	Listening	4	2	8	4.33	0.54	1.87	53.79	18.89

* SPI scores are not computed for content standards measured by fewer than four items.

Table 8-45a 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
7	59539	A	Reading - Key Ideas and Details	5	2	9	5.96	0.68	2.26	66.07	21.99
	59539	B	Reading - Craft & Structure	3	2	7	4.85	0.65	1.54	68.77	16.70
	59539	C	Reading - Vocabulary Use	3	1	4	3.06	0.77	1.08	75.33	21.70
	59539	D	Writing/Language - Text Types and Purposes	4	1	5	3.08	0.61	1.28	61.42	17.76
	59539	E	Writing/Language - Research	2	3	18	6.93	0.56	3.54	39.42	17.37
	59539	F	Writing/Language - Language Conventions	3	1	5	3.40	0.69	1.16	67.74	17.06
	59539	G	Listening	4	2	8	5.47	0.68	1.89	68.21	19.39
8	59006	A	Reading - Key Ideas and Details	5	2	9	5.87	0.63	2.14	64.96	20.28
	59006	B	Reading - Craft & Structure	3	2	7	4.59	0.67	1.75	65.23	21.70
	59006	C	Reading - Vocabulary Use	3	1	4	3.20	0.80	0.93	79.19	18.56
	59006	D	Writing/Language - Text Types and Purposes	5	0	5	2.83	0.56	1.39	56.65	20.29
	59006	E	Writing/Language - Research	1	4	18	7.34	0.58	3.35	41.45	16.22
	59006	F	Writing/Language - Language Conventions	3	1	5	2.78	0.57	1.27	55.83	19.03
	59006	G	Listening	4	2	8	5.02	0.64	2.16	62.41	22.01

Table 8-45b 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	61120	Listening	5	1	7	4.84	0.68	1.77	68.10	21.65
	61120	Reading	10	5	20	10.58	0.52	4.54	52.93	21.56
	61120	Writing	9	4	26	11.31	0.56	4.27	43.76	15.08
4	59776	Listening	4	2	8	5.11	0.62	2.03	63.55	20.62
	59776	Reading	10	6	20	12.08	0.60	4.26	60.50	20.25
	59776	Writing	6	7	28	13.07	0.62	3.93	46.98	12.49
5	59662	Listening	4	2	8	4.20	0.53	2.12	52.78	22.16
	59662	Reading	10	5	20	10.97	0.59	3.28	54.78	14.85
	59662	Writing	8	6	28	11.80	0.57	4.75	42.39	15.64
6	60164	Listening	4	2	8	4.33	0.54	1.87	53.78	18.77
	60164	Reading	14	3	20	11.52	0.57	4.44	57.63	21.07
	60164	Writing	7	6	28	11.68	0.57	4.35	41.92	14.03
7	59539	Listening	4	2	8	5.47	0.68	1.89	68.04	19.48
	59539	Reading	11	5	20	13.87	0.70	4.12	69.09	19.47
	59539	Writing	9	5	28	13.41	0.61	4.99	48.29	16.73
8	59006	Listening	4	2	8	5.02	0.64	2.16	62.23	22.27
	59006	Reading	11	5	20	13.66	0.69	4.16	68.08	19.88
	59006	Writing	9	5	28	12.95	0.57	5.02	46.60	16.93

Table 8-46 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	61220	A	Operations and Algebraic Thinking	8	1	9	5.46	0.60	2.38	60.69	23.62
	61220	B	Number and Operations in Base Ten	6	2	8	5.13	0.64	2.15	63.80	23.64
	61220	C	Number and Operations - Fractions	5	3	8	4.25	0.53	1.87	53.50	19.09
	61220	D	Measurement and Data	7	3	10	4.00	0.40	1.96	40.30	16.72
	61220	E	Geometry	4	3	7	3.92	0.56	1.76	55.87	19.81
4	59855	A	Operations and Algebraic Thinking	9	1	10	5.44	0.54	2.16	54.50	18.07
	59855	B	Number and Operations in Base Ten	4	5	9	4.66	0.51	2.37	51.92	23.56
	59855	C	Number and Operations - Fractions	8	2	10	4.47	0.44	2.95	44.30	27.29
	59855	D	Measurement and Data	7	3	10	4.00	0.40	2.35	40.36	20.60
	59855	E	Geometry	6	1	7	3.34	0.48	1.67	47.59	17.74
5	59733	A	Operations and Algebraic Thinking	5	4	9	3.68	0.41	2.22	41.05	21.54
	59733	B	Number and Operations in Base Ten	6	3	9	5.19	0.57	2.26	57.18	22.32
	59733	C	Number and Operations - Fractions	5	4	9	4.09	0.45	2.28	45.50	21.48
	59733	D	Measurement and Data	7	3	10	4.05	0.40	2.32	40.93	19.80
	59733	E	Geometry	4	5	9	3.20	0.35	2.13	35.72	20.11

Table 8-46 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	60220	E	Geometry	4	3	7	2.90	0.41	1.84	41.82	21.80
	60220	F	Ratios and Proportional Relationships	4	3	7	2.86	0.41	1.53	41.84	16.70
	60220	G	The Number System	7	4	11	5.96	0.54	2.73	54.26	22.69
	60220	H	Expressions and Equations	7	4	11	4.78	0.44	2.53	43.63	20.18
	60220	I	Statistics and Probability	8	2	10	4.46	0.45	1.96	44.70	15.51
7	59600	E	Geometry	6	4	10	3.91	0.40	2.16	39.77	17.99
	59600	F	Ratios and Proportional Relationships	6	2	8	4.52	0.56	2.10	56.00	22.99
	59600	G	The Number System	4	3	7	2.76	0.39	1.91	39.57	23.56
	59600	H	Expressions and Equations	6	4	10	3.14	0.31	2.29	31.59	20.06
	59600	I	Statistics and Probability	7	4	11	3.99	0.36	2.32	36.39	18.08
8	59076	E	Geometry	6	4	10	3.45	0.35	2.15	34.80	17.54
	59076	G	The Number System	4	4	8	2.24	0.28	1.77	28.88	17.79
	59076	H	Expressions and Equations	6	4	10	3.93	0.39	2.39	39.54	21.17
	59076	I	Statistics and Probability	7	1	8	4.41	0.55	2.00	54.54	21.48
	59076	J	Functions	6	4	10	4.28	0.43	2.23	42.76	19.12

Table 8-47 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	59832	A/B	Science Connections & Nature of Science	8	0	8	5.52	0.69	1.77	69.08	18.42
	59832	C	Science Inquiry	8	0	8	5.34	0.67	2.03	67.07	22.04
	59832	D	Physical Science	5	0	5	3.38	0.68	1.11	67.80	14.38
	59832	E	Earth and Space Science	6	0	6	3.27	0.55	1.45	55.42	17.29
	59832	F	Life & Environmental Science	6	0	6	4.24	0.71	1.35	70.85	17.32
	59832	G/H	Science Applications & Personal Social Perspectives	7	0	7	5.84	0.84	1.39	82.89	16.95
8	59000	A/B	Science Connections & Nature of Science	7	0	7	5.37	0.77	1.61	76.88	19.67
	59000	C	Science Inquiry	9	0	9	7.41	0.83	1.77	82.22	17.40
	59000	D	Physical Science	6	0	6	4.35	0.73	1.32	72.77	16.65
	59000	E	Earth and Space Science	6	0	6	3.79	0.64	1.46	63.94	17.80
	59000	F	Life & Environmental Science	6	0	6	4.29	0.72	1.33	72.08	16.17
	59000	G/H	Science Applications & Personal Social Perspectives	6	0	6	4.96	0.83	1.34	82.32	19.12

Table 8-48 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	59817	A	Geography	10	0	10	7.06	0.71	2.20	70.57	19.42
	59817	B	History	8	0	8	5.43	0.68	1.85	68.10	19.56
	59817	C	Political Science and Citizenship	7	0	7	5.27	0.76	1.68	75.06	20.57
	59817	D	Economics	6	0	6	3.27	0.55	1.69	55.33	22.87
	59817	E	The Behavioral Sciences	7	0	7	5.39	0.78	1.65	76.79	20.33
8	59024	A	Geography	10	0	10	7.37	0.74	2.29	73.57	20.43
	59024	B	History	13	0	13	8.50	0.66	2.91	65.63	20.57
	59024	C	Political Science and Citizenship	6	0	6	4.26	0.72	1.58	70.88	22.31
	59024	D	Economics	6	0	6	3.94	0.66	1.47	65.70	19.53
	59024	E	The Behavioral Sciences	5	0	5	3.10	0.63	1.41	62.47	22.60
10	62203	A	Geography	10	0	10	6.76	0.68	2.26	67.42	18.98
	62203	B	History	12	0	12	7.76	0.65	2.73	64.74	20.65
	62203	C	Political Science and Citizenship	12	0	12	8.16	0.68	2.78	67.92	21.04
	62203	D	Economics	8	0	8	4.41	0.55	2.00	55.56	20.96
	62203	E	The Behavioral Sciences	8	0	8	5.34	0.67	1.95	66.83	20.43

Table 8-49 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	30	0	41	0	40
	2	31	56	42	64	41	52
	3	57	85	65	87	53	65
	4	86	100	88	100	66	100
Reading - Craft & Structure	1	0	31	0	40	0	39
	2	32	54	41	63	40	61
	3	55	75	64	81	62	76
	4	76	100	82	100	77	100
Reading - Vocabulary Use*	1	*	*	0	46	*	*
	2	*	*	47	69	*	*
	3	*	*	70	89	*	*
	4	*	*	90	100	*	*
Writing/Language - Text Types and Purposes	1	0	19	0	41	0	57
	2	20	33	42	59	58	75
	3	34	50	60	76	76	89
	4	51	100	77	100	90	100
Writing/Language - Research	1	0	40	0	24	0	18
	2	41	60	25	33	19	31
	3	61	74	34	47	32	49
	4	75	100	48	100	50	100
Writing/Language - Language Conventions	1	0	55	0	72	0	42
	2	56	70	73	85	43	64
	3	71	86	86	92	65	84
	4	87	100	93	100	85	100
Listening	1	0	48	0	46	0	33
	2	49	76	47	69	34	57
	3	77	92	70	89	58	83
	4	93	100	90	100	84	100
Reading	1	0	31	0	42	0	43
	2	32	57	43	65	44	59
	3	58	81	66	86	60	71
	4	82	100	87	100	72	100
Writing	1	0	30	0	37	0	29
	2	31	45	38	48	30	45
	3	46	62	49	61	46	62
	4	63	100	62	100	63	100

* SPI scores are not computed for content standards measured by fewer than four items.

Table 8-49 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	34	0	47	0	47
	2	35	58	48	74	48	72
	3	59	79	75	92	73	87
	4	80	100	93	100	88	100
Reading - Craft & Structure	1	0	42	0	58	0	45
	2	43	68	59	74	46	74
	3	69	87	75	86	75	88
	4	88	100	87	100	89	100
Reading - Vocabulary Use*	1	*	*	0	60	0	69
	2	*	*	61	86	70	87
	3	*	*	87	96	88	94
	4	*	*	97	100	95	100
Writing/Language - Text Types and Purposes	1	0	64	0	48	0	37
	2	65	83	49	64	38	60
	3	84	92	65	84	61	82
	4	93	100	85	100	83	100
Writing/Language - Research	1	0	17	0	25	0	28
	2	18	29	26	41	29	43
	3	30	43	42	60	44	59
	4	44	100	61	100	60	100
Writing/Language - Language Conventions	1	0	45	0	55	0	39
	2	46	62	56	72	40	59
	3	63	75	73	88	60	78
	4	76	100	89	100	79	100
Listening	1	0	37	0	52	0	42
	2	38	58	53	76	43	70
	3	59	73	77	89	71	87
	4	74	100	90	100	88	100
Reading	1	0	37	0	53	0	50
	2	38	63	54	76	51	76
	3	64	83	77	91	77	89
	4	84	100	92	100	90	100
Writing	1	0	30	0	34	0	32
	2	31	44	35	51	33	49
	3	45	57	52	70	50	67
	4	58	100	71	100	68	100

* SPI scores are not computed for content standards measured by fewer than four items.

Table 8-50 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	34	0	36	0	22
	2	35	63	37	56	23	41
	3	64	90	57	76	42	71
	4	91	100	77	100	72	100
Number and Operations in Base Ten	1	0	37	0	27	0	40
	2	38	67	28	55	41	62
	3	68	92	56	81	63	84
	4	93	100	82	100	85	100
Number and Operations - Fractions	1	0	33	0	15	0	26
	2	34	52	16	43	27	44
	3	53	79	44	83	45	76
	4	80	100	84	100	77	100
Measurement and Data	1	0	23	0	19	0	24
	2	24	39	20	40	25	38
	3	40	61	41	67	39	69
	4	62	100	68	100	70	100
Geometry	1	0	35	0	31	0	18
	2	36	56	32	46	19	35
	3	57	81	47	69	36	63
	4	82	100	70	100	64	100

Table 8-50 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	23	0	27	0	22
	2	24	39	28	39	23	38
	3	40	82	40	75	39	66
	4	83	100	76	100	67	100
Ratios and Proportional Relationships*	1	0	30	0	41		
	2	31	41	42	64		
	3	42	69	65	89		
	4	70	100	90	100		
The Number System	1	0	35	0	21	0	15
	2	36	57	22	43	16	31
	3	58	89	44	81	32	60
	4	90	100	82	100	61	100
Expressions and Equations	1	0	26	0	16	0	22
	2	27	44	17	29	23	45
	3	45	78	30	74	46	77
	4	79	100	75	100	78	100
Statistics and Probability	1	0	32	0	22	0	39
	2	33	46	23	36	40	65
	3	47	68	37	73	66	85
	4	69	100	74	100	86	100
Functions**	1					0	29
	2					30	49
	3					50	73
	4					74	100

Table 8-51 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
Science Connections & Nature of Science	1	0	47	0	55
	2	48	71	56	82
	3	72	86	83	94
	4	87	100	95	100
Science Inquiry	1	0	38	0	65
	2	39	68	66	87
	3	69	89	88	96
	4	90	100	97	100
Physical Science	1	0	52	0	56
	2	53	68	57	75
	3	69	80	76	87
	4	81	100	88	100
Earth and Space Science	1	0	35	0	45
	2	36	53	46	63
	3	54	71	64	81
	4	72	100	82	100
Life & Environmental Science	1	0	51	0	54
	2	52	70	55	73
	3	71	88	74	87
	4	89	100	88	100
Science Applications & Social and Personal Perspectives	1	0	66	0	63
	2	67	87	64	89
	3	88	95	90	97
	4	96	100	98	100

Table 8-52 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	56	0	58	0	54
	2	57	72	59	77	55	70
	3	73	86	78	91	71	84
	4	87	100	92	100	85	100
History	1	0	51	0	48	0	50
	2	52	67	49	67	51	67
	3	68	85	68	84	68	84
	4	86	100	85	100	85	100
Political Science and Citizenship	1	0	58	0	51	0	52
	2	59	78	52	75	53	72
	3	79	92	76	92	73	87
	4	93	100	93	100	88	100
Economics	1	0	33	0	49	0	40
	2	34	51	50	68	41	55
	3	52	75	69	83	56	74
	4	76	100	84	100	75	100
The Behavioral Sciences	1	0	62	0	41	0	52
	2	63	81	42	64	53	68
	3	82	93	65	83	69	86
	4	94	100	84	100	87	100

Part 9: Reliability

Part 9 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 2016 Wisconsin Forward Exam was estimated in four ways:

1. Internal consistency was assessed for all items using Cronbach's alpha.
2. Standard error of measurement (SEM) was calculated for raw score and scale score.
3. Classification consistency and classification accuracy were estimated for the performance level classifications.
4. Inter-rater reliability was estimated for all of the CR items.

The present chapter addresses AERA, APA, & NCME (2014) Standards 2.0, 2.3, 2.7, 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 2.7** When subjective judgment enters into test scoring, evidence should be provided on both interrater consistency in scoring and within-examinee consistency over repeated measurements. A clear distinction should be made among reliability data based on (a) independent panels of raters scoring the same performance or products, (b) a single panel scoring successive performances or new products, and (c) independent panels scoring successive performances or new products. (p. 44)
>
> **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 that reliability and SEM should be assessed for all population subgroups. To meet these standards, this chapter 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 subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. The scale score conditional SEMs are provided in Section 6.3.1.

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 who is *Proficient* or *Advanced*. As described below, the Spring 2016 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 conditional SEM for the cut scores that separate the performance levels.

Standard 2.7 advises reporting measures of inter-rater consistency where subjective judgment is involved in scoring. As discussed in Part 5, ELA TDA items were scored by the AI engine with second reads performed by human scorers. As this section will show, a detailed assessment of inter-rater consistency was applied to the Wisconsin Forward Exam. The assessment conducted is termed inter-rater reliability; it measures the reliability of the AI engine versus human scorers in terms of the scores given to TDA items.

Combined, Cronbach's alpha, SEM, classification consistency, classification accuracy, and inter-rater reliability provide several forms of evidence bearing on the reliability of the Wisconsin Forward Exam. Cronbach's alpha and the SEM operate at the content level: they provide estimates of reliability for student scores in ELA or Mathematics, for example. 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* and the associated accountability requirements. Inter-rater reliability probes further, looking at individual items and evaluating the reliability of the AI engine versus human scorers as the scores are assigned to TDA items. In addition, statistics on Cronbach's alpha and the SEM and the procedure

for setting the standard performance index (SPI) cut scores at the reported content standard level presents reliability and precision evidence in support of the diagnostic use of the Wisconsin Forward Exam subscores. Altogether, the provided evidence in this Part of the Technical Report targeted at each intended use of the Wisconsin Forward Exam scores addresses Standard 2.0.

9.1 Measures of Internal Consistency and SEM

Cronbach's alpha (1951) 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 9-1 and 9-2, respectively. These tables include information for all students and for the subgroup categories of gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency.

As indicated in Table 9-1, reliability was highest in Mathematics and Social Studies. Looking at all examinees together in the "Total" column, reliability ranges from 0.85 to 0.89 across grades for ELA, from 0.9 to 0.91 for Mathematics, from 0.87 to 0.88 for Science, and from 0.89 to 0.91 for Social Studies. Ideally, we would like all reliability coefficients to 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 none of the ELA and Science assessments nor the Social Studies grade 4 assessments would meet this criterion. The reliability coefficients for these tests are consistent with the small 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 through 8 would need to be increased from 53, 56, 56, 56, 56, and 56 points to 71, 75, 89, 82, 69, and 62 score points, respectively. For the current Science assessments at grades 4 and 8, the increase would need to be from 40 points for both grades, to 54 and

49 score points, respectively; for the current Social Studies grade 4, the increase would need to be from 38 to 42 score points.

Table 9-1 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 the score distribution and variance, so this result is consistent with the generally larger standard deviations (as previously discussed in Part 8 of this report and summarized in Tables 8-19 through 8-26) among many of these subgroups.

The differences in reliability among most subgroups on most tests were quite small. Differences between male and female students were within 0.02 of one another for all grades and content areas.

The difference between disabled and not disabled and economically disadvantaged and not disadvantaged students was within 0.09 of one another. Most differences among the five racial/ethnic groups also were quite small, within 0.04 of one another for all grades in ELA, Science, and Social Studies. In Mathematics, the differences ranged from 0.06 (grade 3) to 0.15 (grade 7), where the higher reliabilities were observed for White or Asian students and the lowest ones consistently for African American students. The greatest differences were between fully English proficient and limited English proficient students, with consistently lower reliability among limited English proficient students.

Table 9-2 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 we can expect that an individual's obtained score will fall within two standard errors of his or her 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 were slightly larger than those for the other content areas. Because these SEMs are on the raw score scale, this result is consistent with the fact that the ELA tests have more raw score points and relatively large raw score standard deviations when compared with the other content areas. For every grade and content area, the conditional SEM for individual scale scores are provided in the scoring tables previously discussed in Part 6 (Tables 6-27 through 6-43).

Reliability, as measured by Cronbach's alpha, was also computed for each content standard within each content area, as well as for each language domain in ELA. Table 9-3 shows these reliability coefficients by content standard/domain. The last column presents the reliability for the total content area (with all content standards/domains) for all examinees. It is clear that the reliability per content standard/domain is lower than that 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 items ranged from 6 to 17 per domain and from 2 to 8 per standard for ELA, from 7 to 11 items per standard for Mathematics, from 5 to 8 items per standard for Science, and from 5 to 13 items per standard for Social Studies. A lower level of reliability statistics per content standard or domain is therefore expected. The generally 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 was previously discussed in the context of SPI).

By content standard/domain, the reliability ranges were as follows (Table 9-3):

- For ELA, reliability indices by content standard or domain ranged from 0.36 (for standard D in grade 4, with 4 items) to 0.81 (for the Reading domain in grade 8, with 16 items).
- For Mathematics, reliability indices by content standard ranged from 0.56 (for standard E in grade 4, with 7 items) to 0.82 (for standard C in grade 4, with 10 items).
- For Science, reliability indices by content standard ranged from 0.31 (for standard D in grade 4, with 5 items) to 0.69 (for standard C in grade 8, with 9 items).
- For Social Studies, reliability indices by content standard ranged from 0.54 (for standard D in grade 8, with 6 items to 0.74 (for standard C in grade 10, with 12 items).

The SEM associated with each content standard is presented in Table 9-4 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 are 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 the reliability is influenced by the number of items, lower reliability for the content standards with fewer items is to be expected.

9.2 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 is of particular interest. The following section describes how the consistency and accuracy of these classifications were evaluated and provides evidence supporting the validity of these classifications.

Conceptually, classification consistency is defined as the extent to which two classifications of a single student agree, either based on two independent administrations of the same test or 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) × (H + 1), where H is the number of cut scores. Three cut scores yield a 4 × 4 contingency table, as can be seen below in Table 9-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 9-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 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 those 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 corresponding to observed scores. For the Wisconsin Forward Exam, the method used to estimate classification accuracy and consistency is the Kolen and Kim method (2004), described in the next section of this report (see also Kim, Choi, Um, & Kim, 2006; Kim, Barton, & Kim, 2007).

9.2.1 Kolen and Kim's Method for Pattern Scoring

As stated in Part 6, when item response theory (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 .

θ_j → (1,1,0,0, …: Item response from the first administration, or Form 1) → $\hat{\theta}_{j1}$

θ_j → (0,1,1,0, …: Item response from the second administration, or Form 2) → $\hat{\theta}_{j2}$

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 9-B using the two ability estimates obtained from Step 2.

Table 9-B Example Classification Table for One Cut Point (C_1)[3]

		First administration; or Form 1	
		$\hat{\theta}_{j1} \geq C_1$	$\hat{\theta}_{j1} < C_1$
$\hat{\theta}_{j2} \geq C_1$			Second administration; or Form 2
$\hat{\theta}_{j2} < C_1$			

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 examinees' abilities are estimated at each quadrature point, this quadrature point can be considered the true score. Therefore, classification accuracy is computed using both examinees' estimated abilities (observed scores) and quadrature point (true score). 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 9-5 through 9-21, 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, in ELA grade 3, 0.17 represents the probability of belonging to *Below Basic* in both the first and second administrations. The 0.03 represents the probability of belonging to *Proficient* in the first administration and *Advanced* in the second administration. "Sum" is obtained simply by adding the four row values or the four column values. This Sum is not always identical to the sum of the values shown in the table because the values displayed have been rounded to two decimal places.

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 9-5 for ELA grade 3, when all three cuts were used for the

[3] This table is constructed for each quadrature point and replication. One, and only one, cell will have a value of one and zeros elsewhere.

computation, classification consistency (P) is 0.72, probability of chance is 0.28, kappa (*k*) is 0.61, and classification accuracy is 0.80. 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 9-5, for example, the P for the second cut, which establishes the *Proficient* performance level, was 0.89, kappa was 0.77, and classification accuracy was 0.92. The interpretation of the values illustrated for Table 9-5 is the same for Tables 9-6 through 9-21.

When only the *Proficient* cut score was applied, P was greater than or equal to 0.87 and *k* was greater than or equal to 0.73 for all ELA and Mathematics tests. For Science, the lowest P was 0.85 and the lowest *k* was 0.71. In Social Studies, the lowest P associated with the *Proficient* cut was 0.86 and the lowest *k* was 0.72. According to Landis and Koch's criteria for *k* (presented previously in this report in the discussion of classification consistency), all tests showed good or excellent agreement based on the cut for the *Proficient* performance level.

9.3 Inter-Rater Reliability for TDA Items

The reliability of scoring of TDA was measured in two ways 1) tabulations of exact and adjacent agreement of two scorers and 2) reliability coefficients. Reliability for TDA items was examined by calculating indices of inter-rater agreement, the degree of reliability with which the AI engine and a human scorer assign scores to a given student response. Two indices for inter-rater reliability, intraclass correlation and weighted kappa, are presented here.

Notation: To assess reliability, it is necessary to replicate the scoring process for a subset of papers. This is usually done with "blind double-reads." Suppose that we have *N* responses, each of which is scored twice. We denote the two scores of response *n* by X_{n1} and X_{n2}, where $n=1, 2, \ldots N$. The resulting data may be presented in two ways, enumeration by response and cross-tabulation.

Data Structure 1: Enumeration by Response. Each row represents a single student response:

Response #	Score 1	Score 2	Mean Score
1	X_{11}	X_{12}	$\overline{X}_{1.}$
2	X_{21}	X_{12}	$\overline{X}_{2.}$
.	.	.	.
.	.	.	.
N	X_{N1}	X_{N11}	$\overline{X}_{N.}$
Column Mean	$\overline{X}_{.1}$	$\overline{X}_{.2}$	$\overline{X}_{..}$

where

$$\overline{X}_{1.} = (X_{11} + X_{12})/2$$

is the mean score for response 1 (similarly for responses 2, 3, …N),

$$\overline{X}_{.1} = \frac{1}{N}\sum_{n=1}^{N} X_{.1} = (X_{11} + X_{21} + ... + X_{N1})/N$$

is the mean of Score 1 over all responses (similarly for Score 2), and

$$\overline{X}_{..} = \frac{1}{N}\sum_{n=1}^{N} 1(X_{n1} + X_{n2})/2$$

is the overall mean score across both scores of all responses.

Data Structure 2: Cross-Tabulation of Score 1 and Score 2. As an alternative, we may create a square table of counts for each Score 1 by Score 2 (i.e., $X_{n1} \times X_{n2}$) combination:

		Score 2				Row Total
		0	**1**	**…**	***m***	
Score 1	0	n_{00}	n_{01}	…	n_{0m}	n_{0+}
	1	n_{10}	n_{11}	…	n_{1m}	n_{1+}
	.	.	.	…	.	.
	.	.	.	…	.	.
	m	n_{m0}	n_{m1}	…	n_{mm}	n_{m+}
Column Total		n_{+0}	n_{+1}	…	n_{+m}	n_{++}

where *m* is the maximum score (for a rubric including zero) obtainable for the item, n_{ij} is the number of responses for which Score 1 = *i* and Score 2 = *j*, n_{i+} is the number of responses for which Score 1 = *i*, and n_{+j} is the number of responses for which Score 2 = *j*.

Formulas for the two reliability coefficients of interest are then given:

1. **Intraclass Correlation**, ρ_{IC}, describes the percentage of overall score variance accounted for by the variance of mean response scores:

$$\rho_{IC} = \frac{Var_n(\overline{X}_{n\cdot})}{Var_n(X_{n1}, X_{n2})} = \frac{\frac{1}{N-1}\sum_{n=1}^{N}(\overline{X}_{n\cdot} - \overline{X}_{\cdot\cdot})^2}{\frac{1}{2(N-1)}\sum_{n=1}^{N}[(X_{n1} - \overline{X}_{\cdot\cdot})^2 + (X_{n2} - \overline{X}_{\cdot\cdot})^2]}.$$

If agreement is perfect, ρ_{IC} = 1. The following is always true: $0 \le \rho_{IC} \le 1$.

2. **Weighted Kappa**, *k*, is used in many contexts as a measure of association in square contingency tables:

$$k = \frac{\sum_{i=0}^{m}\sum_{j=0}^{m} w_{ij}\frac{n_{ij}}{n_{++}} - \sum_{i=0}^{m}\sum_{j=0}^{m} w_{ij}\frac{n_{i+}n_{+j}}{n^2{}_{++}}}{1 - \sum_{i=0}^{m}\sum_{j=0}^{m} w_{ij}\frac{n_{i+}n_{+j}}{n^2{}_{++}}}, \text{ where } w_{ij} = 1 - \frac{(i-j)^2}{M^2}.$$

If agreement is perfect, *k* = 1. If agreement is what would be expected by chance, *k* = 0. The following is always true: $0 \le k \le 1$.

Ordinal rating scales (e.g., 0, 1, 2) used in scoring TDA items contain a certain level of chance agreement that is expected. Although the intraclass correlation is reported in this report, it does not take into account the possibility of chance agreement between the two raters, but Cohen's kappa (*k*) does take this into consideration. In general, *k* will have values equal to or smaller than the intraclass correlation. If agreement is perfect, then the value of *k* is 1.0. If agreement is at chance levels, the value of *k* is zero. As noted in Section 9.2, Landis and Koch (1977) suggest that values of *k* 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." Specific criteria for intraclass correlation or weighted *k* are not established.

Table 9-22 presents the rater agreement statistics for TDA items. The evidence supporting inter-rater reliability is presented in terms of the percentage of agreement between raters (the AI engine and a human rater), two indices of inter-rater reliability, and the distributions of scores across score levels. In the table, "Perfect" agreement is defined as scores that are exactly the same. "Adjacent" agreement is defined as scores differing by one point. "Discrepant" cases are those cases where the scores of the two raters differed by

more than one raw score point. For example, as shown in Table 9-22, for grade 4 TDA item, the perfect agreement, adjacent agreement, discrepant agreement rates are 88%, 12%, and close to 0%, respectively. "Mean" reflects the item mean score from the second reads (by human scorers). "Number of Second Reads" is the number of student responses selected for the purpose of the second read and computing inter-rater reliability. The "Score Frequency" column represents the scoring outcomes for the student responses based on the raw scores given by the second (human) scorer. The column for "Codes" reflects the number of students who received the condition codes B, C, N, R, or T (described in detail in Part 5, Table 5-2 of this report).

Overall, the rater agreement was very high. The mean percentage of non-discrepant scores (i.e., perfect agreement plus adjacent scores), averaged across all items, was approximately 99%. The percentage of responses with condition codes ranged from 1.3% to 9.33% across all items; the percentage exceeded 3% for two TDA items. The mean intraclass correlation, averaged across all items, was 0.92. The mean kappa across all items was approximately 0.83. Intraclass correlations ranged from 0.88 to 0.95, and weighted kappa ranged from 0.76 to 0.89.

Summary

Overall, the analyses discussed in this section of the report indicate acceptable levels of reliability for the Wisconsin Forward Exam. The internal consistency reliability estimates, as measured by Cronbach's alpha coefficient, are reasonable given the number of items in each test. The analyses of classification consistency and accuracy indicated acceptable levels of consistency and accuracy of student proficiency level classifications, and the SEM around the *Proficient* cut score was low in every grade and content area. The levels of rater agreement were high and the discrepancy rates low, with acceptably high values for the weighted kappa and intraclass correlations. The results of the inter-rater reliability analyses indicate a high degree of reliability for scores on the ELA TDA items in the Wisconsin Forward Exam.

Table 9-1 Reliability for Total Group and Subgroups Using Cronbach's Alpha

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		SES	
			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
English Language Arts	3	0.87	0.87	0.87	0.86	0.84	0.85	0.86	0.85	0.87	0.87	0.82	0.86	0.86	0.85	0.86
	4	0.87	0.87	0.88	0.86	0.85	0.86	0.88	0.85	0.87	0.87	0.81	0.86	0.86	0.86	0.86
	5	0.85	0.85	0.85	0.83	0.83	0.83	0.86	0.81	0.85	0.85	0.74	0.83	0.83	0.84	0.83
	6	0.86	0.85	0.86	0.84	0.83	0.84	0.86	0.83	0.86	0.85	0.69	0.83	0.84	0.84	0.84
	7	0.88	0.88	0.89	0.87	0.87	0.87	0.88	0.86	0.88	0.88	0.81	0.86	0.86	0.87	0.87
	8	0.89	0.88	0.89	0.88	0.87	0.88	0.88	0.88	0.89	0.89	0.80	0.86	0.87	0.88	0.88
Mathematics	3	0.90	0.89	0.90	0.89	0.85	0.87	0.91	0.88	0.90	0.90	0.86	0.89	0.89	0.88	0.89
	4	0.91	0.91	0.92	0.91	0.83	0.88	0.93	0.87	0.91	0.91	0.84	0.89	0.91	0.88	0.91
	5	0.91	0.90	0.92	0.91	0.83	0.87	0.93	0.85	0.91	0.91	0.79	0.87	0.91	0.88	0.91
	6	0.91	0.90	0.91	0.90	0.83	0.87	0.91	0.86	0.90	0.91	0.77	0.86	0.90	0.87	0.90
	7	0.90	0.90	0.91	0.90	0.78	0.85	0.93	0.83	0.90	0.90	0.73	0.82	0.90	0.85	0.90
	8	0.90	0.89	0.90	0.89	0.80	0.85	0.92	0.84	0.89	0.90	0.75	0.80	0.89	0.86	0.90
Science	4	0.87	0.86	0.88	0.84	0.84	0.85	0.86	0.84	0.87	0.87	0.81	0.87	0.86	0.86	0.84
	8	0.88	0.87	0.89	0.85	0.87	0.87	0.86	0.88	0.88	0.88	0.81	0.87	0.86	0.88	0.86
Social Studies	4	0.89	0.89	0.90	0.87	0.87	0.88	0.89	0.88	0.89	0.89	0.85	0.89	0.88	0.88	0.87
	8	0.90	0.90	0.91	0.89	0.87	0.88	0.90	0.88	0.90	0.90	0.80	0.87	0.89	0.89	0.89
	10	0.91	0.90	0.92	0.91	0.87	0.89	0.91	0.89	0.91	0.91	0.81	0.87	0.90	0.90	0.91

Table 9-2 Standard Error of Measurement for Total Group and Subgroups

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		SES	
			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
English Language Arts	3	3.37	3.38	3.35	3.34	3.42	3.43	3.42	3.40	3.42	3.37	3.45	3.34	3.36	3.42	3.33
	4	3.16	3.17	3.14	3.13	3.26	3.26	3.20	3.21	3.22	3.15	3.29	3.19	3.15	3.23	3.11
	5	3.40	3.40	3.37	3.39	3.45	3.46	3.40	3.42	3.44	3.40	3.49	3.33	3.39	3.43	3.38
	6	3.50	3.49	3.47	3.48	3.53	3.55	3.52	3.54	3.52	3.50	3.51	3.41	3.49	3.53	3.47
	7	3.37	3.35	3.35	3.33	3.50	3.46	3.35	3.48	3.40	3.37	3.52	3.42	3.34	3.46	3.30
	8	3.36	3.33	3.34	3.30	3.48	3.46	3.42	3.38	3.37	3.35	3.55	3.37	3.33	3.45	3.29
Mathematics	3	2.67	2.68	2.66	2.64	2.74	2.74	2.65	2.74	2.69	2.66	2.76	2.74	2.65	2.74	2.61
	4	2.85	2.85	2.84	2.85	2.76	2.84	2.82	2.81	2.85	2.85	2.81	2.80	2.85	2.85	2.84
	5	2.80	2.80	2.79	2.81	2.64	2.75	2.75	2.79	2.79	2.80	2.68	2.68	2.80	2.77	2.81
	6	2.72	2.71	2.73	2.73	2.63	2.69	2.70	2.69	2.73	2.72	2.62	2.64	2.73	2.70	2.73
	7	2.82	2.82	2.82	2.85	2.64	2.74	2.79	2.74	2.80	2.83	2.62	2.63	2.84	2.76	2.85
	8	2.81	2.82	2.79	2.82	2.63	2.74	2.79	2.75	2.79	2.81	2.63	2.61	2.82	2.76	2.83
Science	4	2.53	2.53	2.52	2.44	2.79	2.69	2.58	2.68	2.56	2.50	2.79	2.72	2.49	2.67	2.41
	8	2.40	2.40	2.40	2.29	2.79	2.64	2.44	2.61	2.48	2.38	2.86	2.78	2.34	2.59	2.28
Social Studies	4	2.45	2.45	2.45	2.36	2.74	2.63	2.49	2.63	2.52	2.43	2.72	2.68	2.42	2.62	2.32
	8	2.55	2.54	2.54	2.47	2.82	2.72	2.56	2.74	2.62	2.53	2.89	2.83	2.50	2.72	2.43
	10	2.98	3.01	2.94	2.91	3.26	3.14	3.01	3.19	2.98	2.97	3.28	3.23	2.94	3.15	2.88

Table 9-3 Cronbach's Alpha Reliability Coefficients for Content Standards/Domains

English Language Arts

Grade	Alpha per Content Standard/Domain									
	A	B	C	D	E	F	G/Listening	Reading	Writing	Total
3	0.67	0.52	*	0.41	0.37	0.39	0.64	0.78	0.66	0.87
4	0.69	0.42	0.56	0.36	0.48	0.38	0.57	0.80	0.66	0.87
5	0.41	0.55	*	0.45	0.49	0.48	0.61	0.68	0.71	0.85
6	0.54	0.65	*	0.37	0.43	0.40	0.50	0.78	0.65	0.86
7	0.67	0.42	0.55	0.40	0.59	0.44	0.56	0.79	0.73	0.88
8	0.61	0.65	0.49	0.47	0.53	0.50	0.60	0.81	0.74	0.89

*Results are not reported for the content standards measured by fewer than four items.

Mathematics

Grade	Alpha per Content Standard										
	A	B	C	D	E	F	G	H	I	J	Total
3	0.73	0.72	0.64	0.58	0.58						0.90
4	0.61	0.73	0.82	0.69	0.56						0.91
5	0.71	0.71	0.68	0.68	0.70						0.91
6					0.66	0.57	0.77	0.71	0.59		0.91
7					0.62	0.67	0.68	0.69	0.65		0.90
8					0.61		0.63	0.70	0.66	0.65	0.90

Science

Grade	Alpha per Content Standard						
	A/B	C	D	E	F	G/H	Total
4	0.60	0.67	0.31	0.43	0.51	0.61	0.87
8	0.63	0.69	0.47	0.44	0.44	0.64	0.88

Social Studies

Grade	Alpha per Content Standard					
	A	B	C	D	E	Total
4	0.69	0.62	0.65	0.59	0.66	0.89
8	0.72	0.73	0.63	0.54	0.59	0.90
10	0.65	0.72	0.74	0.63	0.63	0.91

Table 9-4 Standard Error of Measurement per Content Standard/Domains

English Language Arts

Grade	SEM per Content Standard/Domain									
	A	B	C	D	E	F	G/Listening	Reading	Writing	Total
3	1.51	1.35	0.61	2.24	0.99	0.88	1.07	2.12	2.49	3.37
4	1.35	0.97	0.87	0.87	2.07	0.84	1.33	1.90	2.29	3.16
5	1.28	1.17	0.65	0.91	2.31	0.97	1.33	1.86	2.56	3.40
6	1.54	1.30	0.60	0.77	2.33	1.04	1.32	2.11	2.58	3.50
7	1.31	1.17	0.72	0.99	2.26	0.87	1.25	1.90	2.57	3.37
8	1.34	1.03	0.67	1.01	2.29	0.90	1.36	1.82	2.54	3.36

Mathematics

Grade	SEM per Content Standard										
	A	B	C	D	E	F	G	H	I	J	Total
3	1.23	1.13	1.13	1.27	1.14						2.67
4	1.36	1.23	1.26	1.31	1.11						2.85
5	1.19	1.22	1.29	1.31	1.16						2.80
6					1.08	1.00	1.31	1.36	1.26		2.72
7					1.34	1.20	1.08	1.27	1.37		2.82
8					1.33		1.09	1.31	1.17	1.31	2.81

Science

Grade	SEM per Content Standard						
	A/B	C	D	E	F	G/H	Total
4	1.12	1.16	0.92	1.09	0.95	0.87	2.53
8	0.99	0.98	0.96	1.09	1.00	0.80	2.40

Social Studies

Grade	SEM per Content Standard					
	A	B	C	D	E	Total
4	1.23	1.15	1.00	1.09	0.97	2.45
8	1.20	1.50	0.97	0.99	0.90	2.55
10	1.33	1.46	1.41	1.22	1.18	2.98

Table 9-5 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	0.17	0.05	0.00	0.00	0.22
Basic	0.05	0.24	0.06	0.00	0.35
Proficient	0.00	0.06	0.23	0.03	0.32
Advanced	0.00	0.00	0.04	0.08	0.12
Sum	0.22	0.35	0.33	0.11	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.90	0.89	0.93	0.72
Probability of Chance	0.66	0.51	0.80	0.28
Kappa (k)	0.72	0.77	0.64	0.61
Classification Accuracy	0.93	0.92	0.95	0.80

Table 9-6 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	0.18	0.05	0.00	0.00	0.23
Basic	0.05	0.22	0.06	0.00	0.33
Proficient	0.00	0.06	0.24	0.04	0.33
Advanced	0.00	0.00	0.03	0.07	0.10
Sum	0.23	0.33	0.33	0.11	

Indexes for Classification Consistency and Classification Accuracy

Indexes	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	0.29
Kappa (k)	0.74	0.76	0.61	0.60
Classification Accuracy	0.94	0.91	0.95	0.80

Table 9-7 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	0.19	0.04	0.00	0.00	0.23
Basic	0.05	0.22	0.06	0.00	0.33
Proficient	0.00	0.06	0.24	0.03	0.34
Advanced	0.00	0.00	0.04	0.06	0.10
Sum	0.24	0.33	0.34	0.10	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.91	0.87	0.93	0.71
Probability of Chance	0.64	0.51	0.83	0.29
Kappa (k)	0.74	0.74	0.61	0.59
Classification Accuracy	0.94	0.91	0.95	0.80

Table 9-8 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	0.17	0.05	0.00	0.00	0.22
Basic	0.05	0.24	0.06	0.00	0.34
Proficient	0.00	0.07	0.20	0.04	0.31
Advanced	0.00	0.00	0.04	0.09	0.13
Sum	0.22	0.36	0.30	0.13	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.90	0.87	0.92	0.69
Probability of Chance	0.66	0.51	0.77	0.28
Kappa (k)	0.71	0.73	0.63	0.57
Classification Accuracy	0.93	0.91	0.94	0.78

Table 9-9 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	0.18	0.05	0.00	0.00	0.23
Basic	0.04	0.24	0.05	0.00	0.34
Proficient	0.00	0.06	0.23	0.04	0.33
Advanced	0.00	0.00	0.03	0.06	0.10
Sum	0.23	0.35	0.32	0.10	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.91	0.88	0.93	0.72
Probability of Chance	0.65	0.51	0.82	0.29
Kappa (k)	0.75	0.76	0.60	0.61
Classification Accuracy	0.94	0.92	0.95	0.81

Table 9-10 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	0.17	0.04	0.00	0.00	0.21
Basic	0.04	0.27	0.05	0.00	0.37
Proficient	0.00	0.06	0.21	0.03	0.30
Advanced	0.00	0.00	0.04	0.08	0.12
Sum	0.21	0.37	0.30	0.12	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.88	0.93	0.74
Probability of Chance	0.67	0.51	0.79	0.29
Kappa (k)	0.75	0.76	0.67	0.63
Classification Accuracy	0.94	0.92	0.95	0.81

Table 9-11 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	0.15	0.04	0.00	0.00	0.19
Basic	0.04	0.22	0.05	0.00	0.32
Proficient	0.00	0.06	0.29	0.03	0.38
Advanced	0.00	0.00	0.03	0.07	0.11
Sum	0.20	0.32	0.38	0.11	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.89	0.94	0.75
Probability of Chance	0.69	0.50	0.81	0.30
Kappa (k)	0.74	0.78	0.67	0.64
Classification Accuracy	0.94	0.91	0.96	0.81

Table 9-12 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	0.16	0.05	0.00	0.00	0.20
Basic	0.05	0.25	0.05	0.00	0.35
Proficient	0.00	0.05	0.26	0.03	0.33
Advanced	0.00	0.00	0.03	0.09	0.11
Sum	0.21	0.35	0.33	0.12	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.90	0.90	0.95	0.75
Probability of Chance	0.67	0.51	0.80	0.29
Kappa (k)	0.69	0.80	0.73	0.65
Classification Accuracy	0.93	0.93	0.96	0.82

Table 9-13 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	0.22	0.05	0.00	0.00	0.26
Basic	0.06	0.19	0.05	0.00	0.29
Proficient	0.00	0.05	0.26	0.03	0.34
Advanced	0.00	0.00	0.02	0.08	0.11
Sum	0.27	0.28	0.33	0.11	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.89	0.90	0.95	0.74
Probability of Chance	0.61	0.51	0.81	0.28
Kappa (k)	0.73	0.80	0.72	0.64
Classification Accuracy	0.92	0.92	0.96	0.81

Table 9-14 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	0.21	0.05	0.00	0.00	0.26
Basic	0.05	0.20	0.05	0.00	0.30
Proficient	0.00	0.06	0.30	0.01	0.38
Advanced	0.00	0.00	0.02	0.05	0.07
Sum	0.26	0.31	0.37	0.07	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.89	0.89	0.97	0.75
Probability of Chance	0.62	0.51	0.88	0.30
Kappa (k)	0.73	0.77	0.75	0.65
Classification Accuracy	0.93	0.92	0.98	0.82

Table 9-15 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	0.25	0.05	0.00	0.00	0.30
Basic	0.06	0.19	0.05	0.00	0.30
Proficient	0.00	0.05	0.28	0.02	0.35
Advanced	0.00	0.00	0.01	0.04	0.05
Sum	0.31	0.30	0.34	0.05	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.89	0.90	0.97	0.76
Probability of Chance	0.57	0.52	0.90	0.30
Kappa (k)	0.73	0.79	0.69	0.65
Classification Accuracy	0.92	0.93	0.98	0.83

Table 9-16 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	0.23	0.06	0.00	0.00	0.29
Basic	0.06	0.26	0.05	0.00	0.37
Proficient	0.00	0.05	0.21	0.02	0.28
Advanced	0.00	0.00	0.02	0.05	0.07
Sum	0.29	0.37	0.28	0.06	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.88	0.90	0.96	0.74
Probability of Chance	0.59	0.55	0.88	0.30
Kappa (k)	0.71	0.77	0.71	0.63
Classification Accuracy	0.92	0.93	0.97	0.82

Table 9-17 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	0.11	0.04	0.00	0.00	0.15
Basic	0.03	0.23	0.08	0.00	0.34
Proficient	0.00	0.06	0.21	0.06	0.33
Advanced	0.00	0.00	0.05	0.13	0.18
Sum	0.14	0.33	0.34	0.19	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.85	0.89	0.67
Probability of Chance	0.75	0.50	0.70	0.28
Kappa (k)	0.71	0.71	0.62	0.55
Classification Accuracy	0.95	0.90	0.92	0.77

Table 9-18 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	0.12	0.03	0.00	0.00	0.15
Basic	0.03	0.23	0.07	0.00	0.34
Proficient	0.00	0.07	0.19	0.07	0.33
Advanced	0.00	0.00	0.06	0.12	0.18
Sum	0.15	0.34	0.32	0.19	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.86	0.87	0.66
Probability of Chance	0.74	0.50	0.70	0.28
Kappa (k)	0.74	0.71	0.57	0.54
Classification Accuracy	0.95	0.90	0.91	0.76

Table 9-19 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	0.17	0.04	0.00	0.00	0.21
Basic	0.04	0.14	0.07	0.00	0.25
Proficient	0.00	0.06	0.18	0.06	0.31
Advanced	0.00	0.00	0.05	0.17	0.23
Sum	0.22	0.24	0.30	0.24	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.91	0.86	0.88	0.67
Probability of Chance	0.66	0.50	0.64	0.25
Kappa (k)	0.75	0.72	0.67	0.55
Classification Accuracy	0.94	0.91	0.91	0.76

Table 9-20 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	0.18	0.04	0.00	0.00	0.22
Basic	0.03	0.18	0.05	0.00	0.27
Proficient	0.00	0.06	0.19	0.06	0.31
Advanced	0.00	0.00	0.05	0.16	0.21
Sum	0.21	0.28	0.29	0.22	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.89	0.89	0.70
Probability of Chance	0.66	0.50	0.66	0.26
Kappa (k)	0.77	0.78	0.66	0.60
Classification Accuracy	0.94	0.92	0.92	0.79

Table 9-21 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	0.21	0.04	0.00	0.00	0.25
Basic	0.04	0.15	0.06	0.00	0.25
Proficient	0.00	0.06	0.19	0.05	0.30
Advanced	0.00	0.00	0.04	0.16	0.20
Sum	0.25	0.25	0.29	0.20	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.91	0.88	0.91	0.71
Probability of Chance	0.62	0.50	0.68	0.25
Kappa (k)	0.77	0.75	0.72	0.60
Classification Accuracy	0.94	0.91	0.94	0.79

Table 9-22 Inter-Rater Reliability, English Language Arts

Grade	Item No.	Max	Percentage Absolute Difference			Intra. Corr.	Weighted Kappa	Mean	Score Frequency					
			Perfect	Adjacent	Discrepant				No. of Second Reads	1	2	3	4	Codes
3	16	4	89.45	10.51	0.03	0.91	0.82	1.41	8,982	5,866	2,708	139	1	268
4	17	4	88.07	11.68	0.24	0.88	0.76	1.32	6,353	4,591	1,039	125	5	593
5	16	4	86.93	13.04	0.04	0.90	0.80	1.40	8,595	5,960	2,136	298	6	195
6	18	4	93.11	6.82	0.07	0.94	0.88	1.34	8,547	6,039	2,119	263	15	111
7	17	4	87.98	11.99	0.03	0.92	0.84	1.44	9,506	6,148	2,694	324	36	304
8	17	4	90.69	9.19	0.12	0.95	0.89	1.50	9,093	5,601	2,677	511	91	213

Note: The sum of the modes of agreement and codes may not equal exactly 100% due to rounding.

Part 10: Validity

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*; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [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. (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. 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 a challenge to) its validity, including design, content specifications, item development, psychometric quality, and inferences made from the results.

As the Technical Report has progressed part by part, it has moved 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 also highlights the meaning and significance of the procedures, processes, and results in terms of validity or a relationship to the *Standards*. Part 10 addresses three final issues in validity: the issues of bias, construct validity, and test integrity. The analyses presented here add to the perspectives provided in Parts 2 through 9. Below is a brief review.

Part 2 of the Technical Report describes the involvement of Wisconsin educators, DPI, and DRC in the test development process. As indicated in Part 2, 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 formed a legitimate basis upon which to derive valid conclusions about student achievement.

Part 3 of this report focuses on key development tasks related to creating the Spring 2016 Wisconsin Forward Exam operational field test forms. The test specifications and item development activities described in Part 2 explain how specific development processes provide evidence to support test validity, primarily content validity, 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 the project served as critical guides throughout development and field testing of items. These documents contributed to ensuring that each form of the test accurately measured the content in consistent and stable ways, thus providing evidence supporting the test's use as an indicator of student achievement of state standards.

Parts 2 and 3 together provide evidence to support the content validity of the Wisconsin Forward Exam and address AERA, APA, & NCME (2014) Standards 3.1, 3.2, 4.0, 4.1, 4.7, and 4.12.

Part 4 of the Technical Report describes the process, procedures, and policies that guided the administration of the Wisconsin Forward Exam, including accommodations, security, and the written procedures provided to test administrators and school personnel. The following AERA, APA, & NCME (2014) Standards are addressed: 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 the Wisconsin Forward Exam by reducing the impact of construct-irrelevant variables (e.g., nonstandardized 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 items, TE items, and TDA writing items were scored: the handscoring process, the training and selection of scorers, the scoring rubrics used for scoring TDA items, and the resulting score distributions. The procedures described in this section contribute to the validity of the Wisconsin Forward Exam by preventing hardware- or software-related errors in machine scoring and reducing construct-irrelevant score variance associated with variations in raters' interpretation and application of scoring rubrics.

Part 6 describes the sample data used for calibration and scaling, the calibration and scaling methods as well as processes and procedures for deriving scale scores from response patterns. Some references to introductory and advanced discussions of IRT are provided. Several axes upon which to evaluate the calibration and scaling procedures, such as the models and data used, the software applied, the vertical relationship across grades, the successful 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, 5.2, and 7.2. These processes and procedures contribute to the validity of the Wisconsin Forward Exam by providing the opportunity to evaluate items contributing to the accurate and reliable measurement of the intended constructs and by ensuring that a valid baseline year for the Wisconsin Forward Exam is set.

Part 7 of the Technical Report provides a brief summary of the Wisconsin Forward Exam standard setting, conducted in June 2016, during which the cut scores were set for all content areas. The process of the standard setting adhered to the 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 presents classical item analysis data, raw score results, scale score results, performance level information, and SPI scores. Scale score results provided a basic quantitative

reference to student performance as derived through the IRT models applied. The performance level information reflected the performance level requirements of the DPI policy environment, as well as interests of parents, students, and educators. The SPI scores then probed further, assessing specific skills and abilities. Combined, scale scores, performance levels, and SPI scores provided a comprehensive set of tools to assess Wisconsin student performance by content and grade level and by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Part 8 thus addresses AERA, APA, & NCME (2014) Standards 1.8, 4.14, 5.1, 5.2, 5.21, 7.0, and 7.1. The analyses addressed in Part 8 contribute to the validity of the Wisconsin Forward Exam by providing further opportunity to identify and eliminate items that are not contributing to the accurate and reliable measurement of the intended constructs.

Part 9 demonstrates adherence to AERA, APA, & NCME (2014) standards through several analyses of the reliability of the Spring 2016 Wisconsin Forward Exam. It presents a reliability analysis using Cronbach's alpha, SEM results, a detailed analysis of classification consistency and classification accuracy, and a full analysis of inter-rater reliability. The Spring 2016 Wisconsin Forward Exam Technical Report thereby addresses AERA, APA, & NCME (2014) Standards 2.0, 2.3, 2.7, 2.11, 2.13, 2.14, and 2.16. Reliability is a prerequisite to score validity, and the analyses in that section contribute to the Wisconsin Forward Exam validity evidence by establishing the reliability of the Wisconsin Forward Exam scores and proficiency classifications.

In the subsequent pages, Part 10 will, as stated, present additional metrics with which to evaluate the validity of the Wisconsin Forward Exam program. As described below, the Wisconsin Forward Exam program formally assessed the issue of test bias 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, socioeconomic status, disability status, and English language proficiency.

This part is particularly relevant to AERA, APA, & NCME (2014) Standards 3.1 through 3.6. These standards are from Chapter 3 of the AERA, APA, & NCME (2014) *Standards* "Fairness in Testing." Each of these standards will be presented as will be the way the standard is addressed 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. (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

has several steps that are followed in item development and selection as is explained in Part 3. These practices adhere to Standard 3.3:

> **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. (64)

DRC conducted DIF studies following the operational administration of the Wisconsin Forward Exam. Often items are evaluated for possible DIF in the field test phase of the test development, and items flagged for DIF are typically further examined for possible bias. In case of the Wisconsin Forward Exam, the DIF analyses were conducted after the operational/field-test administration. Section 10.1 of this part of the Technical Report explains the steps taken to evaluate the Wisconsin Forward Exam items through the use of DIF.

Section 3.1.3 of Part 3 discusses the form quality review conducted for Wisconsin Forward Exam and the steps taken by DRC to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. This review is also critical in fulfilling AERA, APA, & NCME (2014) Standards 3.1 and 3.2.

> **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. (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. (64)

The present part of the report also provides estimations of construct validity. Two measures are provided: correlations between content area objectives and principal components 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. As such, these analyses are essential elements in assessing the validity of the Wisconsin Forward Exam. In addition, this chapter outlines the forensic analysis procedures that were employed to ensure the integrity of test scores by identifying test papers that may have been fraudulently altered. Last but not least, a summary of standardized test administration procedures is provided as an additional evidence supporting validity of test scores.

10.1 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, socioeconomic status, disability status, and English language proficiency (ELP) groups. For the DIF analysis by gender, the reference group is male, meaning that the results for female students are considered with reference to male student performance. In the DIF analysis for race/ethnicity, the reference group is White. This means that the performance of students of each race/ethnicity is considered with reference to the performance of White students. The DIF analysis on socioeconomic status defines students identified as not economically disadvantaged as the reference group and students identified as economically disadvantaged as the focal group. The DIF analysis for disability status uses students identified as not disabled as a reference group to assess DIF within the student population identified as disabled. The DIF analysis for ELP compares item functioning among students identified as fully English proficient to those identified as limited English proficient. Students identified as fully English proficient comprise the reference group, and those identified as limited English proficient comprise the focal group.

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 chi square.

In addition to the MH chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. Educational Testing Service (ETS) first developed the MH-D DIF statistic (Holland & Thayer, 1985, 1986). To compute delta, alpha (the odds ratio) is first computed:

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

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. The χ^2_{MH}, then, 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 $1.0 \leq |\text{MH D-DIF}| < 1.5$
- Large DIF: Significant MH chi-square statistic ($p < 0.05$) and $|\text{MH D-DIF}| \geq 1.5$

For constructed-response 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 of the item. The SMD is an effect size index of DIF, which is relatively easy to interpret (Zwick et al., 1993). The SMD compares the mean of the reference and focal group, 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} = proportion of the focal group members at the *k*th 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 NAEP:

- Moderate DIF: If the MH statistic is significant ($p < .05$) and |ES| is between 0.17 and 0.25
- Large DIF: If the MH statistic is significant ($p < .05$) and $|\text{ES}| \geq 0.25$

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group. Tables 10.1 through 10.9 show the DIF results for the following subgroups:

- ***Gender*:** Focal group is females; reference group is males.

- ***Race/Ethnicity*:** Focal groups are students whose race/ethnicity is reported as African-American, Hispanic, Asian, American Indian, or Two or More Ethnicities; reference group is students whose race/ethnicity is reported as White.
- ***English Language Proficiency*:** Focal group is students who are classified as not fully English language proficient; reference group is all others.
- ***Disability Status*:** Focal group is students with one or more disabilities; reference group is all others.
- ***SES status*:** Focal group is students who are socioeconomically disadvantaged; reference group is all others.

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 10-1 through 10-9 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 the tables described below.

The DIF results for gender are presented in Table 10-1; results for race/ethnicity are presented in Tables 10-2 through 10-6; English language proficiency (ELP) results are presented in Table 10-7; results based on disability status are presented in Table 10-8; and results based on SES status are shown in Table 10-9.

Each DIF table references the grade and content area of the items flagged for DIF, as well as the item number on the test and the item type. The tables present the SMD statistics and the Mantel-Haenszel statistic (Δ_{MH}). After specifying these statistics for each item, the final column provides a flag status. The flag is based on SMD statistics for constructed-response items and on MH (Δ_{MH}) statistics.

In Table 10-1, looking at all items and all grades and content areas, 13 items were flagged for gender DIF in ELA; 7 items were flagged in Mathematics; 6 items were flagged in Science; and 9 items were flagged in Social Studies. Overall, 12 items were flagged in favor of the focal group (Females) and 23 items were flagged against the focal group. Of all items flagged for gender DIF, 3 displayed large DIF (1 in ELA, 1 in Science, and 1 is Social Studies) and 32 items display moderate DIF.

The other DIF results in Tables 10-2 through 10-9 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.

The Spring 2016 Wisconsin Forward Exam was developed to minimize item and test bias. Expertise in this area is not, however, a substitute for statistical analyses of the items. 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 of the items in the Spring 2016 Wisconsin Forward Exam flagged for DIF were notated as such in the classical item analyses and 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 strong DIF) are to be avoided in future selections.

10.2 Construct Validity

Construct validity can be defined as the extent to which tests measure the skills or constructs they intend to measure, and it is the central concept underlying the Spring 2016 Wisconsin Forward Exam validation process. Evidence for construct validity is comprehensive and integrates evidence from both content- and criterion-related validity. The Wisconsin Forward Exam test development process included specifications, item writing, review, and test construction.

Threats to construct 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.

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 10-10 shows the correlations between main test domains for ELA, and Tables 10-11 through 10-14 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, 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 10-10, the correlations between the ELA main test domains of Reading, Writing, and Listening are moderate to high and ranging from 0.55 to 0.72 across all grades. The correlations between ELA content standards (see Table 10-11) are typically

moderate for all grades and all standard pairs and range from 0.24 to 0.65. It should be noted however, that the number of items in several content standards was small, which was very likely a contributing factor to the lower correlations at the standard level compared to the correlations at the ELA domain level.

As indicated in Table 10-12, correlations between Mathematics content standards are also moderate to high and range from 0.47 to 0.73. The correlations between Science content standards range from 0.37 to 0.68 (see Table 10-13), and the correlations between Social Studies range from 0.55 to 0.74 (as shown in Table 10-14). 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.

Wisconsin Forward Exam items are calibrated using unidimensional IRT models, which posit 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. 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 lot of reading ability, the Mathematics test would not be only a measure of mathematics ability, it would be a measure of reading ability as well. 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 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 the 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 the 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 10-15 show that the eigenvalue of the first factor, in most cases, is at least five times as large as the eigenvalue of the second factor.

As may be seen in Table 10-15, the ratios of the first two eigenvalues range from 4.46 to 9.93. The eigenvalues are proportional to the amount of common variance explained by each component, so these ratios indicate that the variance explained by the first component alone is approximately 4 to 8 times greater than the variance explained by the second component. The eigenvalue ratios ranges from 5.02 to 6.14 in ELA, from 4.46 to 6.50 in Mathematics, from 4.67

to 4.83 in Science, and from 5.99 to 6.63 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.

10.2.1 Divergent Validity

Divergent (discriminant) validity is a subtype of construct validity that 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 among measures of unrelated or distantly related constructs are examined in support of divergent validity.

To assess the divergent validity of the Wisconsin Forward Exam, correlations were computed between the ELA, Mathematics, Science, and Social Studies scale scores for students who took more than one subject area test in 2016. These correlations are shown in Table 10-16. The correlation coefficients ranged from 0.71 (between Mathematics and Social Studies in grade 8) to 0.81 (between Social Studies and Science in grade 4). The correlation coefficients suggest that individual student scores for ELA, Mathematics, Science, and Social Studies are highly related. Despite high correlations, the tests are not perfectly related to each other, suggesting that different constructs are being tapped; however, the test scores do appear at highly related to one another, suggesting they may be tapping into a similar knowledge base or general underlying ability.

10.3 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 students' own abilities. Various activities may take place, such as a student copying from another student's paper, students receiving inappropriate assistance before or during testing, or students' 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.

Three 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 multiple choice items on the Spring 2016 Wisconsin Forward Exam ELA, Mathematics, Science, and Social Studies tests that were changed to correct responses. We refer to these answer changes as wrong-to-right (WTR) answer changes. Inordinate numbers of WTR answer changes in a specifically identifiable testing administration group may indicate inappropriate intervention on students' answer documents by an educator.

The second study evaluates the student's irregular response behavior related to giving a correct response to an item in too short or too long time period compared to a typical response time for similar items. Inordinate numbers of unusual response times may indicate inappropriate pre-knowledge of the items or other interventions during the testing session.

Lastly, in order to identify the students who copied TDA responses from their peers, student responses were compared with other student responses to the same TDA item within each school. Comparisons of a unique set of all words that were used by the students in their responses to a TDA item allows for mathematical computation of a degree of response similarity. Once responses are flagged, they are manually reviewed to confirm a high degree of similarity and potential copying issue.

The results of the three 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.

10.4 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 ability.

Standardized test administration, extensive training of the test scorers and AI engine, and rigorous scoring rules for auto-scored 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. (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. (65)

Taken together, the standardized Wisconsin Forward Exam test administration procedures described in Part 4 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 District Assessment Coordinators, School Assessment Coordinators, and test administrators.

Table 10-1 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	16	TDA	0.18		B
	5	20	MC	-0.06	-1.46	B-
	5	33	MC	-0.09	-1.33	B-
	5	35	MS	-0.16		B-
	6	6	MC	-0.09	-1.36	B-
	6	18	TDA	0.22		C
	6	21	MC	-0.07	-1.17	B-
	7	17	TDA	0.20		B
	7	27	MC	-0.09	-1.10	B-
	8	17	TDA	0.21		B
	8	22	MC	-0.09	-1.08	B-
	8	33	ESR	-0.19		B-
	8	35	MC	-0.07	-1.05	B-
Math	3	4	MC	-0.07	-1.17	B-
	3	31	MC	0.08	1.01	B
	4	1	MC	-0.09	-1.40	B-
	4	7	MC	-0.09	-1.16	B-
	5	5	MC	-0.08	-1.11	B-
	6	2	MC	-0.04	-1.04	B-
	6	32	MC	-0.09	-1.17	B-
Science	4	21	MC	-0.05	-1.29	B-
	8	7	MC	0.01	1.01	B
	8	18	MC	0.02	1.46	B
	8	19	MC	0.03	1.07	B
	8	24	MC	0.06	1.02	B
	8	30	MC	-0.13	-1.51	C-
Social Studies	8	21	MC	-0.11	-1.52	C-
	8	24	MC	-0.09	-1.03	B-
	8	40	MC	0.07	1.11	B
	10	10	MC	-0.09	-1.03	B-
	10	18	MC	-0.07	-1.16	B-
	10	37	MC	-0.06	-1.42	B-
	10	40	MC	0.10	1.41	B
	10	41	MC	-0.09	-1.29	B-
	10	44	MC	0.02	1.08	B

Table 10-2 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	4	23	MC	-0.08	-1.09	B-
	4	28	TE	-0.06		B-
	5	20	MC	-0.12	-1.67	C-
	5	34	MC	-0.09	-1.10	B-
	5	35	MS	-0.13		B-
	6	6	MC	-0.08	-1.07	B-
	8	3	MC	0.08	1.01	B
Math	8	7	MC	0.08	1.04	B
	8	16	TE	-0.11	-1.87	B-
Science	4	19	MC	-0.06	-1.01	B-
	8	7	MC	0.04	1.32	B
	8	9	MC	-0.07	-1.04	B-
Social Studies	4	3	MC	0.10	1.20	B
	4	4	MC	0.09	1.20	B
	4	8	MC	0.07	1.06	B
	4	17	MC	-0.09	-1.42	B-
	8	4	MC	0.08	1.09	B
	8	24	MC	-0.10	-1.38	B-
	8	34	MC	-0.13	-1.86	C-
	10	24	MC	0.09	1.02	B

Table 10-3 Items Flagged for DIF, by Race/Ethnicity, Focal Group: Hispanic

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
ELA	3	1	MC	-0.09	-1.15	B-
	5	16	TDA	0.16		B
	5	20	MC	-0.09	-1.56	C-
	5	34	MC	-0.10	-1.24	B-
	6	6	MC	-0.10	-1.30	B-
	6	31	ESR	-0.15		B-
	7	9	MC	-0.05	-1.07	B-
Science	8	11	MC	-0.07	-1.09	B-
Social Studies	4	7	MC	-0.06	-1.10	B-
	8	14	MC	0.12	1.49	B
	10	9	MC	0.20	4.14	C
	10	50	MC	-0.08	-1.15	B-

Table 10-4 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	1	MC	-0.07	-1.02	B-
	4	13	MC	-0.08	-1.10	B-
	4	17	TDA	0.15		B
	4	24	TE	-0.10		B-
	4	25	TE	-0.06		B-
	5	16	TDA	0.19		B
	5	20	MC	-0.13	-2.59	C-
	5	34	MC	-0.09	-1.40	B-
	6	6	MC	-0.11	-1.60	C-
	6	18	TDA	0.18		B
	6	25	TE	0.16		B
	6	31	ESR	-0.15		B-
	7	9	MC	-0.05	-1.58	C-
	7	18	TE	-0.08	-0.96	B-
	7	22	MC	-0.06	-1.02	B-
	7	24	MC	-0.06	-1.31	B-
	8	1	MC	-0.12	-2.02	C-
	8	14	MC	-0.05	-1.58	C-
	8	17	TDA	0.20		B
	8	32	MC	-0.06	-1.64	C-
Math	3	4	MC	-0.07	-1.09	B-
	3	7	MC	0.06	1.51	C
	3	37	MC	0.08	1.29	B
	4	2	MC	-0.10	-1.14	B-
	4	29	TE	0.10	1.52	B
	5	5	MC	-0.08	-1.06	B-
	5	20	TE	0.10	1.40	B
	6	7	MC	-0.02	-1.28	B-
	6	31	MC	0.07	1.16	B
	6	34	MC	-0.10	-1.28	B-
	6	35	MC	0.07	1.03	B
	7	31	MC	-0.06	-1.09	B-
	8	35	MC	-0.07	-1.21	B-

Table 10-4 Items Flagged for DIF, by Race/Ethnicity, Focal Group: Asian (cont.)

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
Science	4	9	MC	-0.03	-1.04	B-
	8	7	MC	0.03	2.48	C
	8	11	MC	-0.06	-1.27	B-
	8	19	MC	0.03	1.08	B
	8	36	MC	0.07	1.05	B
Social Studies	4	7	MC	-0.09	-1.82	C-
	4	25	MC	-0.09	-1.26	B-
	4	29	MC	0.08	1.07	B
	8	26	MC	-0.09	-1.45	B-
	8	32	MC	0.04	1.30	B
	8	34	MC	-0.07	-1.24	B-
	10	3	MC	-0.24	-3.33	C-
	10	13	MC	0.12	1.27	B
	10	15	MC	-0.09	-1.74	C-
	10	40	MC	0.14	2.06	C
	10	42	MC	0.09	1.10	B
	10	50	MC	-0.06	-1.09	B-

Table 10-5 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
ELA	6	31	ESR	-0.13		B-
Social Studies	10	9	MC	0.09	1.40	B

Table 10-6 Items Flagged for DIF, by Race/Ethnicity, Focal Group: Two or More Ethnicities

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
Social Studies	10	9	MC	0.05	1.07	B

Table 10-7 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	1	MC	-0.11	-1.37	B-
	4	17	TDA	0.15		B
	4	24	TE	-0.09		B-
	4	25	TE	-0.08		B-
	5	16	TDA	0.21		C
	5	20	MC	-0.15	-1.79	C-
	5	34	MC	-0.14	-1.68	C-
	6	6	MC	-0.12	-1.48	B-
	6	18	TDA	0.15		B
	6	31	ESR	-0.22		C-
	7	9	MC	-0.11	-1.51	C-
	7	13	TE	-0.08	-0.98	B-
	7	17	TDA	0.15		B
	7	24	MC	-0.09	-1.11	B-
	8	1	MC	-0.12	-1.36	B-
	8	32	MC	-0.11	-1.37	B-
Science	8	11	MC	-0.10	-1.21	B-
Social Studies	4	7	MC	-0.12	-1.62	C-
	8	14	MC	0.12	1.35	B
	8	22	MC	0.10	1.03	B
	10	3	MC	-0.19	-2.01	C-
	10	7	MC	0.09	1.00	B
	10	9	MC	0.22	2.93	C
	10	15	MC	-0.10	-1.11	B-
	10	50	MC	-0.13	-1.43	B-

Table 10-8 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	3	25	TE	-0.08		B-
	6	25	TE	-0.20		C-
	8	23	TE	-0.18		C-
Math	4	41	MC	0.05	1.04	B
	5	1	MC	-0.08	-1.01	B-
	5	39	SA	0.11	1.09	B
	6	7	MC	-0.10	-1.83	C-
	7	1	MC	0.05	1.11	B
	7	24	MC	-0.09	-1.08	B-
Science	4	13	MC	-0.09	-1.24	B-
	8	5	MC	-0.04	-1.36	B-

Table 10-9 Items Flagged for DIF, by SES Status, Focal Group: Socioeconomically Disadvantaged Students

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
ELA	5	20	MC	-0.06	-1.04	B-

Table 10-10 Correlations among English Language Arts Test Domains

Grade	ELA Domain	Listening	Reading
3	Reading	0.65	
	Writing	0.59	0.69
4	Reading	0.66	
	Writing	0.55	0.66
5	Reading	0.57	
	Writing	0.58	0.64
6	Reading	0.60	
	Writing	0.54	0.67
7	Reading	0.64	
	Writing	0.61	0.72
8	Reading	0.65	
	Writing	0.62	0.73

Table 10-11 Correlations among English Language Arts Test Standards

Grade	Standard Code	A	B	C	D	E	F
3	B	0.62					
	C	0.44	0.40				
	D	0.55	0.52	0.36			
	E	0.45	0.42	0.31	0.41		
	F	0.46	0.44	0.32	0.43	0.36	
	G	0.59	0.56	0.42	0.50	0.45	0.45
4	B	0.54					
	C	0.64	0.49				
	D	0.46	0.36	0.41			
	E	0.50	0.39	0.44	0.36		
	F	0.45	0.34	0.42	0.32	0.34	
	G	0.61	0.47	0.55	0.44	0.44	0.41
5	B	0.50					
	C	0.28	0.3				
	D	0.41	0.45	0.28			
	E	0.43	0.48	0.24	0.41		
	F	0.40	0.47	0.27	0.43	0.45	
	G	0.43	0.52	0.32	0.46	0.47	0.49
6	B	0.61					
	C	0.43	0.48				
	D	0.41	0.45	0.32			
	E	0.49	0.51	0.35	0.36		
	F	0.44	0.48	0.33	0.37	0.43	
	G	0.51	0.55	0.40	0.41	0.45	0.43
7	B	0.55					
	C	0.58	0.50				
	D	0.48	0.37	0.42			
	E	0.61	0.47	0.50	0.46		
	F	0.52	0.42	0.46	0.41	0.49	
	G	0.59	0.48	0.53	0.45	0.55	0.48
8	B	0.65					
	C	0.55	0.55				
	D	0.50	0.53	0.41			
	E	0.57	0.58	0.48	0.47		
	F	0.49	0.50	0.42	0.48	0.51	
	G	0.58	0.59	0.47	0.51	0.53	0.49

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 10-12 Correlations Among Mathematics Standards

Grade	Standard Code	A	B	C	D	E	F	G	H	I
3	B	0.73								
	C	0.58	0.57							
	D	0.64	0.63	0.55						
	E	0.59	0.58	0.55	0.56					
4	B	0.67								
	C	0.64	0.71							
	D	0.62	0.67	0.71						
	E	0.47	0.50	0.54	0.57					
5	B	0.66								
	C	0.65	0.61							
	D	0.62	0.59	0.64						
	E	0.66	0.60	0.61	0.61					
6	F					0.56				
	G					0.67	0.65			
	H					0.66	0.62	0.72		
	I					0.57	0.54	0.63	0.62	
7	F					0.58				
	G					0.61	0.65			
	H					0.63	0.63	0.67		
	I					0.61	0.62	0.64	0.67	
8	G					0.54				
	H					0.61		0.62		
	I					0.57		0.52	0.64	
	J					0.59		0.56	0.67	0.67

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 10-13 Correlations among Science Standards

Grade	Standard Code	A/B	C	D	E	F
4	C	0.64				
	D	0.42	0.45			
	E	0.45	0.51	0.37		
	F	0.51	0.55	0.41	0.46	
	G/H	0.58	0.58	0.42	0.43	0.50
8	C	0.66				
	D	0.55	0.54			
	E	0.52	0.49	0.45		
	F	0.53	0.52	0.45	0.44	
	G/H	0.65	0.68	0.53	0.49	0.52

Note: Standard Codes are as follows: A/B = Science Connections & Nature of Science; C =Science Inquiry; D = Physical Science; E = Earth and Space Science; F = Life & Environmental Science; G/H = Science Applications & Social and Personal Perspectives

Table 10-14 Correlations among Social Studies Standards

Grade	Standard Code	A	B	C	D
4	B	0.63			
	C	0.64	0.61		
	D	0.55	0.56	0.58	
	E	0.63	0.59	0.65	0.56
8	B	0.71			
	C	0.66	0.69		
	D	0.60	0.63	0.59	
	E	0.60	0.65	0.6	0.56
10	B	0.66			
	C	0.66	0.74		
	D	0.59	0.65	0.66	
	E	0.62	0.69	0.69	0.62

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

Table 10-15 Principal Components Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
ELA	3	6.90	1.15	6.00
	4	7.24	1.27	5.71
	5	6.48	1.29	5.02
	6	6.64	1.19	5.58
	7	7.61	1.24	6.14
	8	7.91	1.35	5.84
Mathematics	3	8.48	1.70	5.01
	4	9.93	1.53	6.50
	5	9.84	1.67	5.88
	6	9.54	2.14	4.46
	7	9.51	1.69	5.62
	8	8.79	1.61	5.45
Science	4	7.08	1.47	4.83
	8	7.85	1.68	4.67
Social Studies	4	7.94	1.33	5.99
	8	8.87	1.34	6.60
	10	9.80	1.48	6.63

Table 10-16 Correlations Between Content Area Scale Scores

Grade	ELA & Mathematics	ELA & Science	ELA & Social Studies	Mathematics & Science	Mathematics & Social Studies	Science & Social Studies
3	0.73					
4	0.72	0.79	0.8	0.72	0.72	0.81
5	0.72					
6	0.76					
7	0.73					
8	0.73	0.75	0.78	0.72	0.71	0.78

Part 11: Summary Recommendations

Results and key findings of the Spring 2016 Wisconsin Forward Exam test administration are presented throughout the body of this report. This last section of the report presents some recommendations for DPI consideration.

The 2016 ELA and Mathematics test administration was considered to be operational/field test because the items contained in these assessments had not been previously field-tested in Wisconsin. We recommend that in the future all items 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. We recommend continuing to develop and 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 developments.

DRC also recommends continuing to use an artificial intelligence (AI) engine in scoring of text dependent analysis items for its efficiency and accuracy. As indicated in Part 5 and Part 9 of this report, the AI scores were in very high agreement with scores by trained human scorers.

From the psychometric perspective, it was noticed that the ELA test difficulty for some grades may warrant further attention in subsequent administrations. The vertical scaling results described in Part 6 of this report indicate that the ELA grade 5 and grade 6 tests were relatively difficult, while the ELA grade 7 test was relatively easy, as indicated by the test characteristic curves. In order to achieve better ordinality of the ELA assessments' overall difficulty across grade levels, more difficult items could be added to the grade 7 test and/or more easy items could be added to the grade 5 and 6 tests. However, it should be noted that because equating requires that tests maintain a similar level of difficulty from year to year, increasing or decreasing the test rigor would likely require a cut score review and an examination regarding whether a new test scale should be set.

In addition, several items, particularly in higher grades of Mathematics assessments, were found to be very difficult for Wisconsin students. While use of some difficult items may be necessary to fulfill the test content specifications, both DPI and DRC recommend careful review of these items and determination whether they should be included in the Wisconsin Forward Exam item bank for future use or be removed and replaced with other items measuring the same content standards.

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.

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.

Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

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. (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*. Princeton, NJ: Educational Testing Service, Research Report RR-85-43.

Holland, P. W., & Thayer, D. T. (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. (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. (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., & 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.

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 rata* [Computer program]. Chicago, IL: Scientific Software, Inc.

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 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 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

van der Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer.

Wright, B. D., & Linacre, J. M. (1992). *BIGSTEPS Rasch Analysis* [Computer program]. Chicago, IL: MESA Press.

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

Wisconsin Forward Exam Item Review Training

Wisconsin Forward Exam Item Review

December 2015

Purpose of Meeting

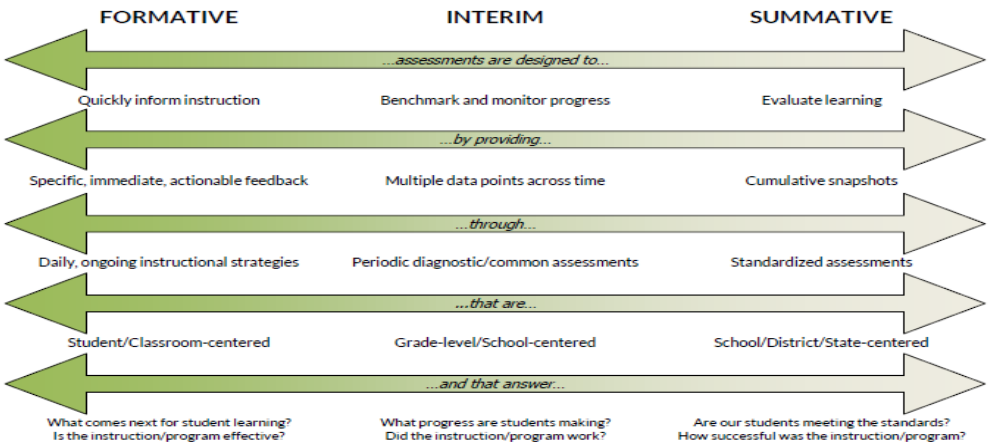
- Provide overview of the Wisconsin Forward Exam
- Provide specific training for reviewing items for **content**

Wisconsin's Definition of College and Career Readiness

Part of a Strategic Assessment System

Strategic Assessment System

By Type: *What are the differences between assessment types within a strategic system?*



Wisconsin Department of Public Instruction - August 2015

Wisconsin Forward Exam

- Grades 3–8 for English Language Arts and Mathematics
- Grades 4, 8, and 10 for Social Studies
- Grades 4 and 8 for Science
- All items written and aligned to Wisconsin State Standards

Security and Confidentiality

- **Critical Importance of Security**
  - Security/Nondisclosure Agreement
  - Security of passage and item content
  - Note-taking policy
  - Cell phone and personal computer use
  - Communication following the meeting

Item Types

- Selected Response (SR)
  - Multiple Choice (MC)
  - Enhanced Selected Response (ESR)
  - Evidence-Based Selected Response (EBSR)
- Text Dependent Analysis (TDA)
- Technology Enhanced (TE)

Selected Response Item Type- Multiple Choice (MC)

- All MC items have 4 answers choices
  - 3 distractors and 1 correct answer
- Used in all content areas
- Can be linked to a passage or stimuli or used as a "stand-alone MC"
- May have graphs, tables, or other information to support the stem

MC Sample

Which title should be enclosed in quotation marks instead of being written in italics?

(a) a novel entitled [redacted]

(b) a play entitled *I* [redacted]

(c) a movie entitled [redacted]

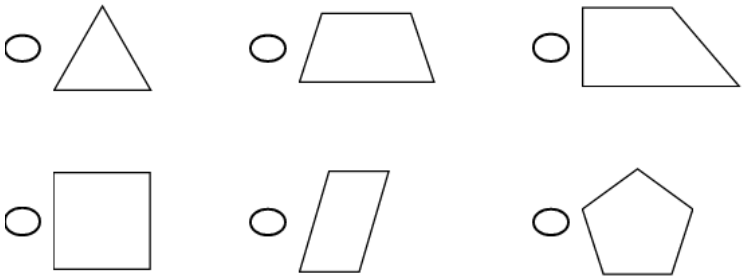
(d) a newspaper ar[redacted]

Selected Response Item Type- ESR

- Varying combinations of multiple choice, multiple response, completion of short answer
- Explores authentic problem-solving skills
- Multi-part, auto scored

ESR Sample

Select all the shapes that are quadrilaterals but **not** rectangles.



Selected Response Item Type-EBSR

- 2-part item
  - Part A-Accuracy portion; single correct answer
  - Part B-Evidence portion; one or more correct answers based upon Part A
- 2 point item; student may get 0, 1, or 2 points

EBSR Sample

This question has two parts. First, answer part A. Then, answer part B.

Part A

What is the main way the passage …

Part B

Which sentence from the passage **best** supports your answer in part A?

Text Dependent Analysis (TDA)

- Used in ELA assessment
- Based on a passage
- Used for both Literature and Informational texts
- Basic writing skills used while inferring and synthesizing information from the passage
- Scored using a holistic scoring guide

TDA Sample

… about public transportation. Use evidence from **both** passages to support your essay.

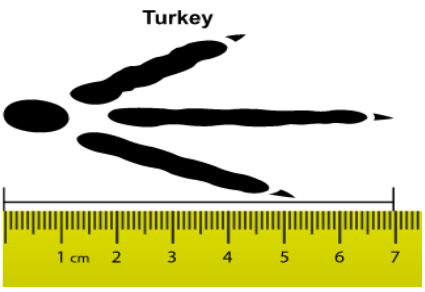
0/5000

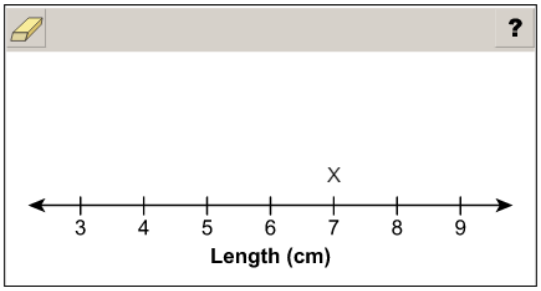
Technology Enhanced (TE)

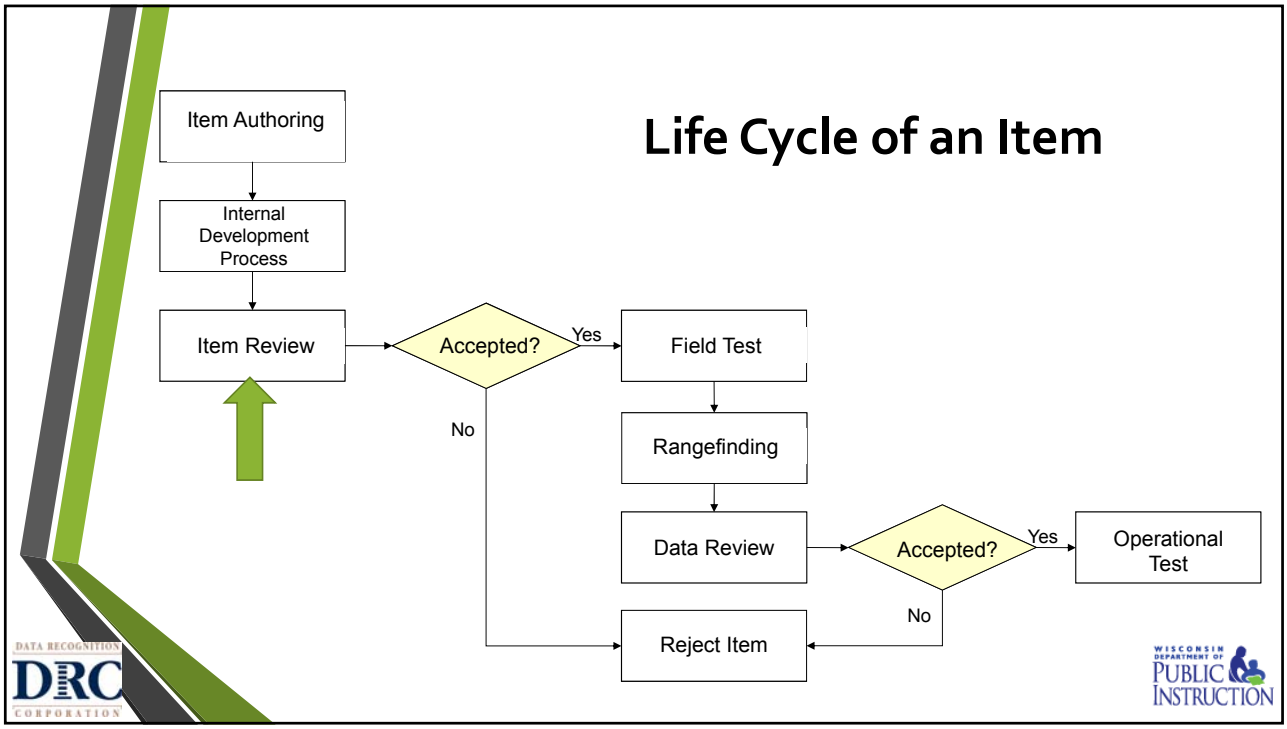
- Presently for ELA and Mathematics
- Interactive
- Wide Variety: clock input, angle draw, drop down list, matching, graphing, highlighting text, drag and drop

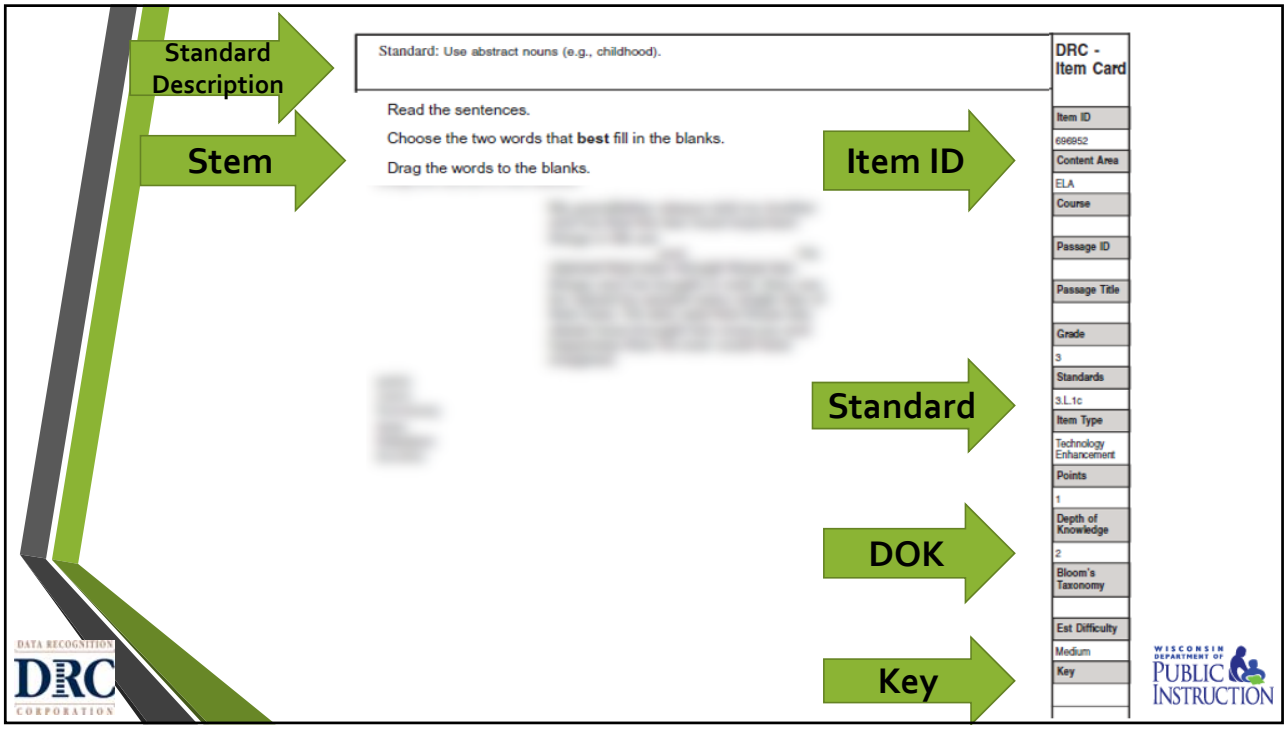
TE Sample Item

Use the ruler tool to measure the lengths of the dog, raccoon, and turkey tracks Rob sees.






Life Cycle of an Item
Item Authoring
Internal Development Process
Item Review
Accepted?
Yes
No
Field Test
Rangefinding
Data Review
Accepted?
Yes
No
Operational Test
Reject Item


Standard Description
Stem
Standard: Use abstract nouns (e.g., childhood).
Read the sentences.
Choose the two words that best fill in the blanks.
Drag the words to the blanks.
Item ID
Standard
DOK
Key
DRC - Item Card
Item ID
696852
Content Area
ELA
Course
Passage ID
Passage Title
Grade
3
Standards
3.L.1c
Item Type
Technology Enhancement
Points
1
Depth of Knowledge
2
Bloom's Taxonomy
Est Difficulty
Medium
Key

Item Review Process

- **Reviews will be completed in groups and individually**
- **Items will be reviewed for**
  - Content alignment
  - Rigor level alignment
    - Grade-level appropriateness, Difficulty, DOK
  - Technical design
  - Universal Design

Item Review Rating Sheet

	Content Alignment	Rigor Level Alignment			Technical Design			Universal Design		STATUS
	Standards	Grade	Difficulty	Depth of Knowledge	Correct Answer	Distractors	Graphics	Language Demand	Bias	Acceptance Status
Unique ID number	—Higher —Lower —None	—Above —At —Below	—High —Medium —Low	—Recall —Application —Strategic Thinking	—Yes —No	—Yes —No —N/A	—Yes —No —N/A	—Yes —No	—Yes —No	— Approved as is — Dissenting View
PASSAGE NAME										

Content Alignment

Does the standard listed, match the state standard?

- Each member will have copy of standards
- Match item to appropriate standard as noted on item card
- Indicate alignment on Item Rating Sheet

Rigor Level Alignment

- Is the grade level of the item appropriate?
  - Conceptual load; vocabulary that is widely accessible to students; syntactic patterns; clear, straight forward language; level of student interest
- Is the difficulty of the item appropriate?
  - Keeping in mind the general population of classroom
- Is the Depth of Knowledge of the item appropriate?
- Does the technology (if TE item) enhance the content knowledge?

Webb's Depth-of-Knowledge (DOK) Levels

Definition of DOK

the degree or complexity of knowledge that the content curriculum standards and expectations require

- Includes four levels, from lowest (basic recall) to highest (extended thinking)
- Focuses on how well the students need to know the content before they can respond to a given item
- Used by item writers to gauge the *cognitive level* of item, does not correlate to the *difficulty* of the item

DOK Levels

DOK 1 Basic Recall

DOK 2 Basic Application of Skill/Concept

DOK 3 Strategic Thinking

DOK 4 Extended Thinking

(rarely on standardized assessments—more "project-like" or on performance assessments)

DOK 1

- Students demonstrate a rote response, use a well-known formula, or follow a simple procedure.
- A "simple" procedure is well defined and typically involves only **one** step.

Key Words: identify, recall, recognize, use, measure

DOK 2

- Students make some decisions regarding how to approach the question or problem.
- This level requires deeper knowledge than just giving a definition, such as explaining *how* or *why*; it may involve two or more steps.

Key Words: classify, organize, estimate, observe, interpret, describe, calculate

DOK 2-(cont.)

Activities may include the following:

- Making observations/collecting information
- Classifying/comparing information
- Organizing/displaying data or information in tables and graphs

Note: Some action verbs, such as "explain," "describe," or "interpret," could be classified at different DOK levels, depending on the complexity of the action.

DOK 3

- Students demonstrate deep understanding through planning, using evidence, and exhibiting higher levels of cognitive reasoning.

Key Words: analyze, order, explain, evaluate, predict, infer, compare

DOK 3-(cont.)

Activities may include the following:

- Drawing conclusions from observations
- Explaining phenomena in terms of concepts
- Using concepts to solve nonroutine problems
- Citing evidence and developing a logical argument for concepts
- Analyzing experimental designs that involve more than one dependent variable

DOK 4

- Higher order thinking skills

Activities may include the following:

- Creating an experiment and recording data
- Processing multiple conditions of a problem
- Developing hypotheses

Key words: analyze, synthesize, examine and explain, describe and illustrate common themes

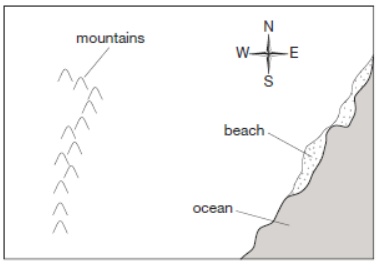
Item Review Rating Sheet

	Content Alignment	Rigor Level Alignment			Technical Design			Universal Design		STATUS
	Standards	Grade	Difficulty	Depth of Knowledge	Correct Answer	Distractors	Graphics	Language Demand	Bias	Acceptance Status
Unique ID number	—Higher —Lower —None	—Above —At —Below	—High —Medium —Low	—Recall —Application —Strategic Thinking	—Yes —No	—Yes —No —N/A	—Yes —No —N/A	—Yes —No	—Yes —No	— Approved as is — Dissenting View
PASSAGE NAME										

Item Card

Standard

Use the drawing below to answer the question.



1. A student is at the beach. Which direction are the mountains from the student?

Item ID

DOK

KEY

Item ID
6
Content Area
Science
Scenario ID
Scenario Title
Grade
4
Standards
6
Item Type
Multiple Choice
Points
1
Depth of Knowledge
2
Bloom's Taxonomy
Est Difficulty
Key
D

Technical Design

- Does the item meet requirements for technical quality?
  - Stem: Complete question/problem; does not clue correct answer
  - Correct answer: clear and accurate
  - Distractors (or incorrect options): may contain common misperceptions or processes
  - Relationship to stimulus or passage
  - Graphics/visuals: compliment the item

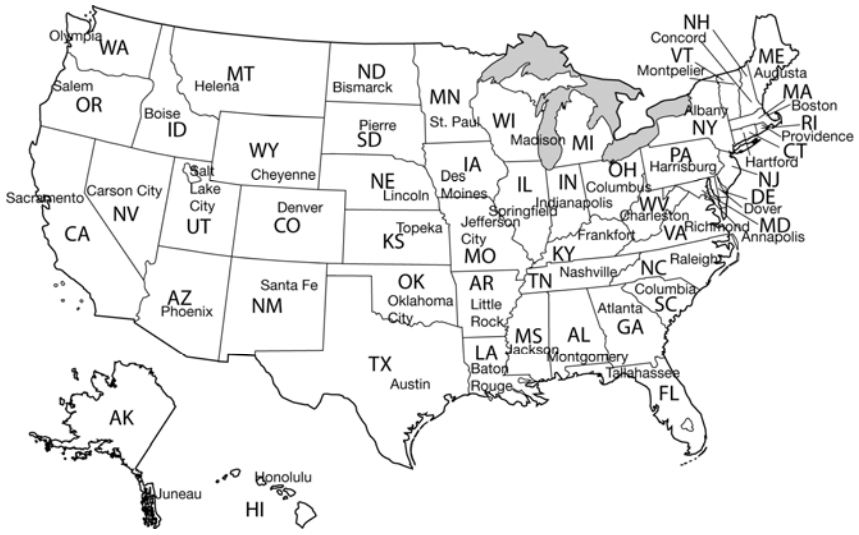
Guidelines for the Principles of Universal Design

- Items should respect the diversity of the assessment population.
- Items should have a clear format for text.
- Items should measure what is intended.
- Stimuli and items should have clear pictures and graphics.

Guidelines for the Principles of Universal Design (cont.)

- Items should have concise and readable text.
- Items should be written to provide for a test that will have an overall appearance that is clean and organized.

Everything in Moderation



Item Review Rating Sheet

	Content Alignment	Rigor Level Alignment			Technical Design			Universal Design		STATUS
	Standards	Grade	Difficulty	Depth of Knowledge	Correct Answer	Distractors	Graphics	Language Demand	Bias	Acceptance Status
Unique ID number	—Higher —Lower —None	—Above —At —Below	—High —Medium —Low	—Recall —Application —Strategic Thinking	—Yes —No	—Yes —No —N/A	—Yes —No —N/A	—Yes —No	—Yes —No	— Approved as is — Dissenting View
PASSAGE NAME										

Evaluating an Item: Grade 8 Science

Use the data table below to answer the question.

Densities of Substances

Substance	Density (grams/cm³)
plastic X	1.38
plastic Y	0.90
water	1.00

A student is given a mixture of 200 plastic beads that all look alike. Each bead is made from one of two types of plastic: plastic X or plastic Y. Which statement describes what will happen when the mixture of beads is placed in water?

(a) nc
(b) ea
(c) so
(d) we

Item Card

Standard

Item ID

DOK

KEY

Use the data table below to answer the question.

Densities of Substances

Substance	Density (grams/cm³)
plastic X	1.38
plastic Y	0.90
water	1.00

A student is given a mixture of 200 plastic beads that all look alike. Each bead is made from one of two types of plastic: plastic X or plastic Y. Which statement describes what will happen when the mixture of beads is placed in water?

(a) nc
(b) ea
(c) so
(d) we

Item ID
Content Area
Science
Scenario ID
Scenario Title
Grade
Standards
Item Type
Multiple Choice
Points
1
Depth of Knowledge
2
Bloom's Taxonomy
2
Est Difficulty
Key
D

Main Question to Ask During Review

- **Does the item provide for an optimal standard assessment of all students?**

Item Review Process: Summary

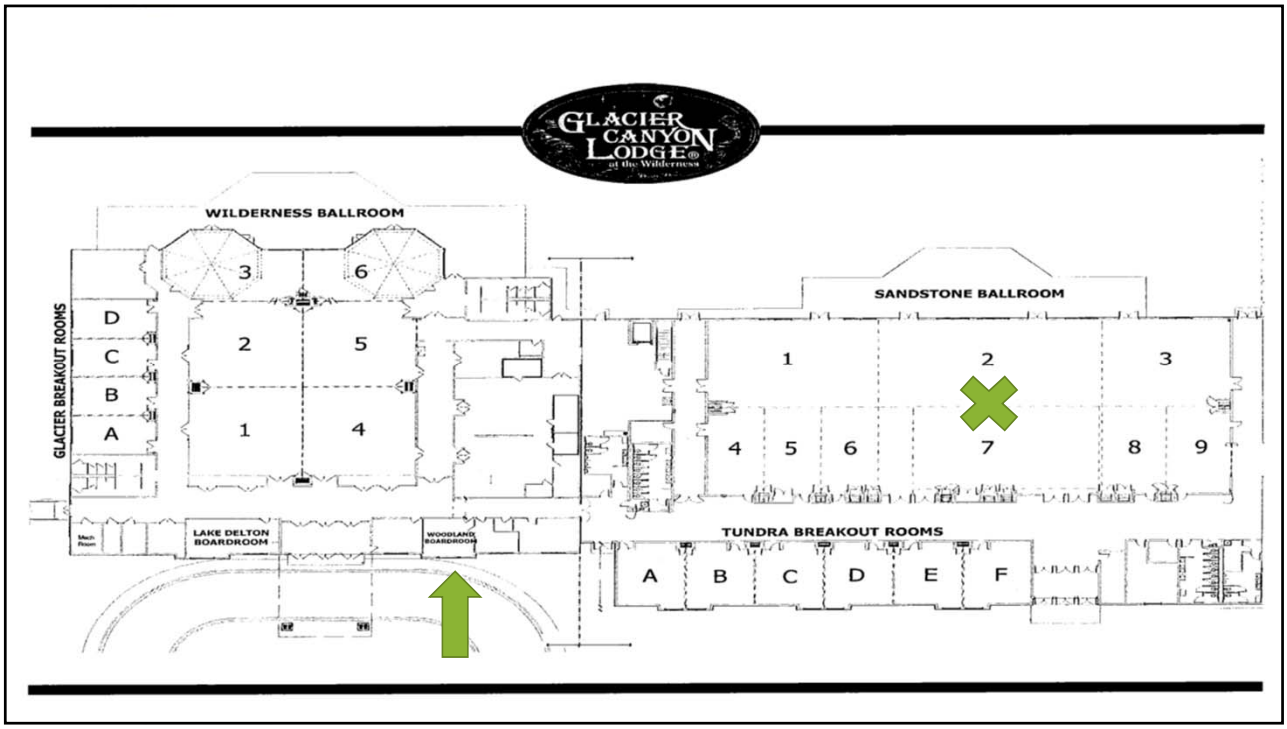
- Content Alignment
- Grade-level Appropriateness
- Rigor level Alignment
- Technical Design
- Universal Design

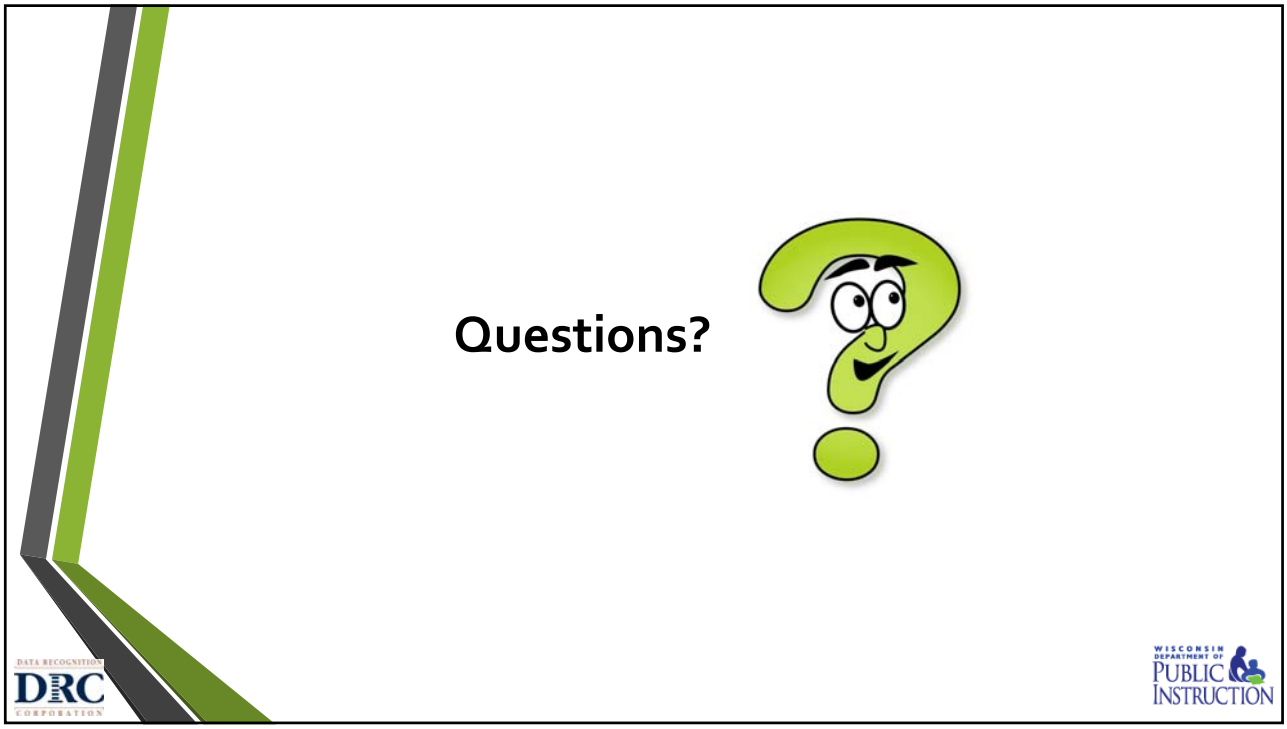
Roles & Responsibilities

- DRC & Wisconsin Department Staff
  - Facilitate discussion
  - Monitor time
  - Answer any questions

Roles & Responsibilities

- Educators
  - Invest yourself in the process
  - Share your opinions
  - Listen to your colleagues


GLACIER CANYON LODGE at the Wilderness
WILDERNESS BALLROOM
3
6
2
5
1
4
GLACIER BREAKOUT ROOMS
D
C
B
A
SANDSTONE BALLROOM
1
2
3
4
5
6
7
8
9
LAKE DELTON BOARDROOM
WOODLAND BOARDROOM
TUNDRA BREAKOUT ROOMS
A
B
C
D
E
F


Questions?
DRC
WISCONSIN DEPARTMENT OF PUBLIC INSTRUCTION

Appendix B

Number of Items Taken to Wisconsin Forward Exam Item Review Meetings

Number of English Language Arts Items Taken to Item Review

English Language Arts Grade 3

Grade	Standard Coding	Number of Items Taken to Review
3	**L**	**26**
3	1b	1
3	1c	1
3	1d	1
3	1g	1
3	1h	1
3	1i	2
3	2a	3
3	2c	1
3	2d	1
3	3a	3
3	3b	1
3	4	8
3	5	2
3	**RI**	**41**
3	1	11
3	2	5
3	3	10
3	5	3
3	6	3
3	7	2
3	8	5
3	9	2
3	**RL**	**45**
3	1	17
3	2	8
3	3	12
3	5	5

English Language Arts Grade 3 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
3	6	2
3	9	1
3	SL	24
3	2	12
3	3	12
3	**W**	**41**
3	1a	1
3	1c	2
3	2	3
3	2a	1
3	2b	1
3	2c	1
3	2d	3
3	3	4
3	3a	1
3	3b	3
3	3c	1
3	3d	2
3	3e	1
3	8	17

English Language Arts Grade 4

Grade	Standard Coding	Number of Items Taken to Review
4	1b	1
4	1c	1
4	1e	1
4	2a	1
4	2b	1
4	3a	1
4	3b	1
4	4	13
4	5	4
4	**RI**	**35**
4	1	12
4	2	8
4	3	4
4	5	6
4	8	5
4	**RL**	**44**
4	1	8
4	2	9
4	3	20
4	5	3
4	6	3
4	9	1
4	**SL**	**20**
4	2	7
4	3	13
4	**W**	**36**
4	1a	2
4	1b	1
4	1c	1
4	1d	1
4	2a	1

English Language Arts Grade 4 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
4	2c	1
4	2d	1
4	2e	1
4	3a	1
4	3b	1
4	3c	1
4	3d	2
4	3e	1
4	5	2
4	8	13
4	9	6

English Language Arts Grade 5

Grade	Standard Coding	Number of Items Taken to Review
5	1b	1
5	1c	1
5	1e	1
5	2	1
5	2a	1
5	2b	1
5	3a	2
5	4	6
5	5	7
5	**RI**	**35**
5	1	10
5	2	5
5	3	9
5	5	1
5	6	1
5	8	8
5	9	1
5	**RL**	**46**
5	1	12
5	2	8
5	3	9
5	4	3
5	5	7
5	6	5
5	9	2
5	**SL**	**20**
5	2	9
5	3	11
5	**W**	**38**
5	1a	1
5	1c	2

English Language Arts Grade 5 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
5	1d	1
5	2a	1
5	2b	1
5	2c	1
5	2d	1
5	2e	1
5	3a	1
5	3b	1
5	3c	1
5	3e	1
5	5	6
5	8	13
5	9	6

English Language Arts Grade 6

Grade	Standard Coding	Number of Items Taken to Review
6	**L**	**19**
6	1	1
6	1b	1
6	1c	2
6	1d	1
6	2a	3
6	2b	1
6	3a	1
6	3b	1
6	4	6
6	5	2
6	**RI**	**37**
6	1	6
6	2	3
6	3	5
6	4	3
6	5	5
6	6	6
6	8	5
6	9	4
6	**RL**	**47**
6	1	11
6	2	6
6	3	10
6	4	4
6	5	11
6	6	5

English Language Arts Grade 6 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
6	**SL**	**20**
6	2	11
6	3	9
6	**W**	**31**
6	1a	2
6	1b	1
6	1c	2
6	2a	1
6	2c	1
6	2d	1
6	2e	1
6	3b	1
6	3d	1
6	3e	2
6	5	2
6	8	12
6	9	4

English Language Arts Grade 7

Grade	Standard Coding	Number of Items Taken to Review
7	1b	2
7	1c	1
7	2	1
7	2a	1
7	3a	2
7	4	7
7	5	3
7	**RI**	**38**
7	1	8
7	2	4
7	3	6
7	4	1
7	5	7
7	6	7
7	8	5
7	**RL**	**43**
7	1	10
7	2	7
7	3	16
7	4	3
7	5	2
7	6	5
7	**SL**	**20**
7	2	12
7	3	8
7	**W**	**30**
7	1a	1
7	1b	1
7	1c	1
7	1e	1
7	2a	1

English Language Arts Grade 7 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
7	2b	1
7	2d	1
7	2e	1
7	3c	1
7	3e	1
7	5	5
7	8	12
7	9	3

English Language Arts Grade 8

Grade	Standard Coding	Number of Items Taken to Review
8	2a	1
8	2b	2
8	2c	2
8	4	6
8	5	3
8	**RI**	**34**
8	1	8
8	2	4
8	3	6
8	5	5
8	6	2
8	8	6
8	9	3
8	**RL**	**38**
8	1	9
8	2	5
8	3	14
8	4	4
8	6	4
8	9	2

English Language Arts Grade 8 (cont.)

Grade	Standard Coding	Number of Items Taken to Review
8	**SL**	**20**
8	2	7
8	3	13
8	**W**	**32**
8	1b	1
8	1c	2
8	1e	1
8	2b	2
8	2e	1
8	3b	1
8	3c	1
8	3d	1
8	5	4
8	8	12
8	9	6

Number of Mathematics Items Taken to Item Review

Mathematics Grade 3

Grade	Standard Coding	Number of Items Taken to Review
3	3.G.2	6
3	3.MD.1	4
3	3.MD.2	2
3	3.MD.3	3
3	3.MD.4	2
3	3.MD.5	1
3	3.MD.6	1
3	3.MD.7	2
3	3.MD.8	2
3	3.NBT.1	5
3	3.NBT.2	5
3	3.NBT.3	4
3	3.NF.1	4
3	3.NF.2	3
3	3.NF.3	5
3	3.OA.1	1
3	3.OA.2	2
3	3.OA.3	1
3	3.OA.4	2
3	3.OA.5	3
3	3.OA.6	1
3	3.OA.7	3
3	3.OA.8	3
3	3.OA.9	2

Mathematics Grade 4

Grade	Standard Coding	Number of Items Taken to Review
4	4.G.2	5
4	4.G.3	3
4	4.MD.1	3
4	4.MD.2	2
4	4.MD.3	2
4	4.MD.4	2
4	4.MD.5	2
4	4.MD.6	2
4	4.MD.7	2
4	4.NBT.1	3
4	4.NBT.2	2
4	4.NBT.3	1
4	4.NBT.4	2
4	4.NBT.5	4
4	4.NBT.6	2
4	4.NF.1	2
4	4.NF.2	3
4	4.NF.3	3
4	4.NF.4	3
4	4.NF.5	2
4	4.NF.6	3
4	4.NF.7	2
4	4.OA.1	2
4	4.OA.2	2
4	4.OA.3	3
4	4.OA.4	4
4	4.OA.5	4

Mathematics Grade 5

Grade	Standard Coding	Number of Items Taken to Review
5	5.G.2	4
5	5.G.3	4
5	5.G.4	3
5	5.MD.1	3
5	5.MD.2	3
5	5.MD.3	3
5	5.MD.4	3
5	5.MD.5	4
5	5.NBT.1	2
5	5.NBT.2	2
5	5.NBT.3	2
5	5.NBT.4	2
5	5.NBT.5	2
5	5.NBT.6	2
5	5.NBT.7	3
5	5.NF.1	3
5	5.NF.2	2
5	5.NF.3	2
5	5.NF.4	2
5	5.NF.5	2
5	5.NF.6	2
5	5.NF.7	2
5	5.OA.1	6
5	5.OA.2	4
5	5.OA.3	5

Mathematics Grade 6

Grade	Standard Coding	Number of Items Taken to Review
6	6.EE.2	3
6	6.EE.3	1
6	6.EE.4	1
6	6.EE.5	2
6	6.EE.6	2
6	6.EE.7	2
6	6.EE.8	2
6	6.EE.9	3
6	6.G.1	3
6	6.G.2	4
6	6.G.3	3
6	6.G.4	3
6	6.NS.1	3
6	6.NS.2	3
6	6.NS.3	4
6	6.NS.4	1
6	6.NS.5	1
6	6.NS.6	3
6	6.NS.7	2
6	6.NS.8	2
6	6.RP.1	5
6	6.RP.2	4
6	6.RP.3	3
6	6.SP.1	3
6	6.SP.2	3
6	6.SP.3	4
6	6.SP.4	4
6	6.SP.5	3

Mathematics Grade 7

Grade	Standard Coding	Number of Items Taken to Review
7	7.EE.2	4
7	7.EE.3	5
7	7.EE.4	5
7	7.G.1	2
7	7.G.2	3
7	7.G.3	2
7	7.G.4	2
7	7.G.5	2
7	7.G.6	6
7	7.NS.1	3
7	7.NS.2	7
7	7.NS.3	3
7	7.RP.1	3
7	7.RP.2	6
7	7.RP.3	4
7	7.SP.1	3
7	7.SP.2	4
7	7.SP.3	3
7	7.SP.4	2
7	7.SP.5	1
7	7.SP.6	2
7	7.SP.7	3
7	7.SP.8	1

Mathematics Grade 8

Grade	Standard Coding	Number of Items Taken to Review
8	8.EE.2	2
8	8.EE.3	2
8	8.EE.4	1
8	8.EE.5	3
8	8.EE.6	2
8	8.EE.7	3
8	8.EE.8	3
8	8.F.1	3
8	8.F.2	4
8	8.F.3	3
8	8.F.4	4
8	8.F.5	3
8	8.G.1	2
8	8.G.2	2
8	8.G.3	3
8	8.G.5	4
8	8.G.6	1
8	8.G.7	1
8	8.G.8	1
8	8.G.9	4
8	8.NS.1	7
8	8.NS.2	6
8	8.SP.1	4
8	8.SP.2	3
8	8.SP.3	4
8	8.SP.4	3

Number of Social Studies Items Taken to Item Review		
Social Studies Grade 4		
Grade	Standard Coding	Number of Items Taken to Review
4	A.4.2	1
4	A.4.3	1
4	A.4.5	1
4	A.4.6	1
4	A.4.7	3
4	A.4.9	2
4	B.4.1	1
4	B.4.2	4
4	B.4.3	1
4	B.4.8	1
4	B.4.9	1
4	B.4.10	1
4	C.4.1	2
4	C.4.2	4
4	C.4.3	1
4	C.4.4	1
4	C.4.5	1
4	C.4.6	1
4	D.4.1	1
4	D.4.2	1
4	D.4.4	2
4	D.4.7	1
4	E.4.6	1
4	E.4.12	1
4	E.4.15	1

Social Studies Grade 8		
Grade	**Standard Coding**	**Number of Items Taken to Review**
8	A.8.2	4
8	A.8.5	1
8	A.8.6	1
8	A.8.9	2
8	A.8.11	1
8	B.8.1	4
8	B.8.3	1
8	B.8.6	1
8	B.8.7	4
8	B.8.9	1
8	B.8.10	1
8	B.8.11	1
8	C.8.6	1
8	C.8.8	1
8	C.8.9	2
8	D.8.2	4
8	D.8.7	1
8	D.8.8	2
8	D.8.10	2
8	E.8.3	1
8	E.8.12	2

Social Studies Grade 10		
Grade	**Standard Coding**	**Number of Items Taken to Review**
10	A.10.3	5
10	A.10.4	1
10	A.10.5	2
10	A.10.7	1
10	A.10.8	1
10	A.10.12	1
10	A.10.13	1
10	B.10.3	2
10	B.10.6	1
10	B.10.7	1
10	B.10.8	2
10	B.10.9	1
10	B.10.14	3
10	B.10.16	2
10	C.10.1	2
10	C.10.2	1
10	C.10.6	1
10	C.10.9	1
10	C.10.11	1
Social Studies Grade 10		
Grade	**Standard Coding**	**Number of Items Taken to Review**
10	C.10.13	3
10	D.10.1	1
10	D.10.2	2
10	D.10.5	1
10	D.10.7	2
10	D.10.8	1
10	D.10.10	1
10	E.10.5	1
10	E.10.6	1
10	E.10.12	1
10	E.10.17	2

Number of Science Items Taken to Item Review		
Science Grade 4		
Grade	Standard Coding	Number of Items Taken to Review
4	A.4.3	2
4	A.4.4	1
4	C.4.1	1
4	C.4.2	3
4	C.4.4	2
4	C.4.5	2
4	C.4.6	2
4	C.4.8	1
4	D.4.5	1
4	E.4.4	1
4	E.4.6	1
4	E.4.8	1
4	F.4.1	4
4	F.4.2	2
4	F.4.3	3
4	F.4.4	3
4	G.4.1	1
4	H.4.1	1

Science Grade 8		
Grade	**Standard Coding**	**Number of Items Taken to Review**
8	C.8.1	1
8	C.8.2	1
8	C.8.4	1
8	C.8.6	4
8	C.8.7	1
8	D.8.1	1
8	D.8.2	1
8	D.8.3	1
8	D.8.4	2
8	D.8.5	1
8	D.8.6	2
8	D.8.8	1
8	D.8.9	1
8	E.8.1	2
8	E.8.3	3
8	E.8.4	1
8	E.8.7	1
8	F.8.1	1
8	F.8.8	1
8	F.8.9	1
8	G.8.3	1
8	G.8.4	3
8	G.8.5	1
8	H.8.3	1

Appendix C

Wisconsin Forward Exam English Language Arts Grades 3-8 Table of Specifications

Wisconsin Forward Exam English Language Arts Grade 3 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	3	Reading	3.RL.1	L	SR, TE	2-3	3
ELA	3	Reading	3.RL.2	L	SR	3	1
ELA	3	Reading	3.RL.3	L	SR	2	2
ELA	3	Reading	3.RL.4	NA	NA	NA	0
ELA	3	Reading	3.RL.5	L	SR	3	1
ELA	3	Reading	3.RL.6	L	SR	1	1
ELA	3	Reading	3.RL.7	NA	NA	NA	0
ELA	3	Reading	3.RL.9	L	SR	3	1
ELA	3	Reading	3.RL.10	NA	NA	NA	0
ELA	3	Reading	3.RI.1	NA	NA	NA	0
ELA	3	Reading	3.RI.2	NA	NA	NA	0
ELA	3	Reading	3.RI.3	I	SR	2	1
ELA	3	Reading	3.RI.4	NA	NA	NA	0
ELA	3	Reading	3.RI.5	NA	NA	NA	0
ELA	3	Reading	3.RI.6	I	SR	2	1
ELA	3	Reading	3.RI.7	NA	NA	NA	0
ELA	3	Reading	3.RI.8	I	SR	2	1
ELA	3	Reading	3.RI.9	I	SR	2	1
ELA	3	Reading	3.RI.10	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1a	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1b	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1c	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1d	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1e	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1f	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1g	NA	SR	3	1
ELA	3	Writing/Language	3.L.1h	NA	NA	NA	0
ELA	3	Writing/Language	3.L.1i	NA	TE	2	1
ELA	3	Writing/Language	3.L.2	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 3 Table of Specifications Spring 2016							
ELA	3	Writing/Language	3.L.2a	NA	SR	2	1
ELA	3	Writing/Language	3.L.2b	NA	NA	NA	0
ELA	3	Writing/Language	3.L.2c	NA	NA	NA	0
ELA	3	Writing/Language	3.L.2d	NA	SR	2	1
ELA	3	Writing/Language	3.L.2e	NA	NA	NA	0
ELA	3	Writing/Language	3.L.2f	NA	NA	NA	0
ELA	3	Writing/Language	3.L.2g	NA	NA	NA	0
ELA	3	Writing/Language	3.L.3	NA	NA	NA	0
ELA	3	Writing/Language	3.L.3a	NA	NA	NA	0
ELA	3	Writing/Language	3.L.3b	NA	NA	NA	0
ELA	3	Reading	3.L.4	I	SR	2	1
ELA	3	Reading	3.L.4a	NA	NA	NA	0
ELA	3	Reading	3.L.4b	NA	NA	NA	0
ELA	3	Reading	3.L.4c	NA	NA	NA	0
ELA	3	Reading	3.L.4d	NA	NA	NA	0
ELA	3	Reading	3.L.5	L	SR	2	1
ELA	3	Reading	3.L.5a	NA	NA	NA	0
ELA	3	Reading	3.L.5b	NA	NA	NA	0
ELA	3	Reading	3.L.5c	NA	NA	NA	0
ELA	3	Writing/Language	3.L.6	NA	NA	NA	0
ELA	3	Writing/Language	3.W.1	NA	NA	NA	0
ELA	3	Writing/Language	3.W.1a	NA	NA	NA	0
ELA	3	Writing/Language	3.W.1b	NA	NA	NA	0
ELA	3	Writing/Language	3.W.1c	NA	SR	2	1
ELA	3	Writing/Language	3.W.1d	NA	NA	NA	0
ELA	3	Writing/Language	3.W.2	NA	NA	NA	0
ELA	3	Writing/Language	3.W.2a	NA	SR	2	1
ELA	3	Writing/Language	3.W.2b	NA	NA	NA	0
ELA	3	Writing/Language	3.W.2c	NA	NA	NA	0
ELA	3	Writing/Language	3.W.2d	NA	SR	2	1
ELA	3	Research: Writing/Language	3.W.3	L	TDA	3	1
ELA	3	Writing/Language	3.W.3a	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 3 Table of Specifications Spring 2016							
ELA	3	Writing/Language	3.W.3b	NA	TE	2	1
ELA	3	Writing/Language	3.W.3c	NA	NA	NA	0
ELA	3	Writing/Language	3.W.3d	NA	NA	NA	0
ELA	3	Writing/Language	3.W.4	NA	NA	NA	0
ELA	3	Writing/Language	3.W.5	NA	NA	NA	0
ELA	3	Writing/Language	3.W.6	NA	NA	NA	0
ELA	3	Writing/Language	3.W.7	NA	NA	NA	0
ELA	3	Writing/Language	3.W.8	NA	TE, SR	2	4
ELA	3	Writing/Language	3.W.10	NA	NA	NA	0
ELA	3	Listening	3.SL.1	NA	NA	NA	0
ELA	3	Listening	3.SL.1a	NA	NA	NA	0
ELA	3	Listening	3.SL.1b	NA	NA	NA	0
ELA	3	Listening	3.SL.1c	NA	NA	NA	0
ELA	3	Listening	3.SL.1d	NA	NA	NA	0
ELA	3	Listening	3.SL.2	I	SR	2	3
ELA	3	Listening	3.SL.3	I	SR	1-3	3
ELA	3	Listening	3.SL.4	NA	NA	NA	0
ELA	3	Listening	3.SL.5	NA	NA	NA	0
ELA	3	Listening	3.SL.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 3 Table of Specifications Spring 2016

Item Type		Passage Type	
SR	Multiple-choice, multiple selected response and evidence-based selected response	**L**	**Literary** (poems, narratives, realistic fiction, historical fiction, fantasy, legends/myths, etc.)
TE	highlighting text, drop-down pull list, drag and drop, drag and paste, list input, matching	**I**	**Informational** (biographies, instructional/how-tos, articles, essays, science and social studies topics, etc.)
TDA	Text Dependent Analysis - One constructed response item aligned to a reading comprehension indicator as well as the indicators designated on the TDA rubric		

Reporting Categories	Total Items	Total Points
Reading	15	20
Writing/Language	13	26
Listening	6	7
	34	**53**

Wisconsin Forward Exam English Language Arts Grade 4 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	4	Reading	4.RL.1	L	SR	2	1
ELA	4	Reading	4.RL.2	L	SR	2	1
ELA	4	Reading	4.RL.3	L	SR	2-3	3
ELA	4	Reading	4.RL.4	NA	NA	NA	0
ELA	4	Reading	4.RL.5	L	SR	2	1
ELA	4	Reading	4.RL.6	L	SR	2	1
ELA	4	Reading	4.RL.7	NA	NA	NA	0
ELA	4	Reading	4.RL.9	NA	NA	NA	0
ELA	4	Reading	4.RL.10	NA	NA	NA	0
ELA	4	Reading	4.RI.1	I	SR	2	1
ELA	4	Reading	4.RI.2	NA	NA	NA	0
ELA	4	Reading	4.RI.3	NA	NA	NA	0
ELA	4	Reading	4.RI.4	NA	NA	NA	0
ELA	4	Reading	4.RI.5	I	SR	2	3
ELA	4	Reading	4.RI.6	NA	NA	NA	0
ELA	4	Reading	4.RI.7	NA	NA	NA	0
ELA	4	Reading	4.RI.8	NA	NA	NA	0
ELA	4	Reading	4.RI.9	NA	NA	NA	0
ELA	4	Reading	4.RI.10	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1a	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1b	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1c	NA	TE	2	1
ELA	4	Writing/Language	4.L.1d	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1e	NA	TE	2	1
ELA	4	Writing/Language	4.L.1f	NA	NA	NA	0
ELA	4	Writing/Language	4.L.1g	NA	NA	NA	0
ELA	4	Writing/Language	4.L.2	NA	NA	NA	0
ELA	4	Writing/Language	4.L.2a	NA	SR	1	1
ELA	4	Writing/Language	4.L.2b	NA	SR	2	1

Wisconsin Forward Exam English Language Arts Grade 4 Table of Specifications Spring 2016							
ELA	4	Writing/Language	4.L.2c	NA	NA	NA	0
ELA	4	Writing/Language	4.L.2d	NA	NA	NA	0
ELA	4	Writing/Language	4.L.3	NA	NA	NA	0
ELA	4	Writing/Language	4.L.3a	NA	NA	NA	0
ELA	4	Writing/Language	4.L.3b	NA	NA	NA	0
ELA	4	Writing/Language	4.L.3c	NA	NA	NA	0
ELA	4	Reading	4.L.4	I and L	SR, TE	2	4
ELA	4	Reading	4.L.4a	NA	NA	NA	0
ELA	4	Reading	4.L.4b	NA	NA	NA	0
ELA	4	Reading	4.L.4c	NA	NA	NA	0
ELA	4	Reading	4.L.5	L	SR	2	1
ELA	4	Reading	4.L.5a	NA	NA	NA	0
ELA	4	Reading	4.L.5b	NA	NA	NA	0
ELA	4	Reading	4.L.5c	NA	NA	NA	0
ELA	4	Writing/Language	4.L.6	NA	NA	NA	0
ELA	4	Writing/Language	4.W.1	NA	NA	NA	0
ELA	4	Writing/Language	4.W.1a	NA	SR	2	1
ELA	4	Writing/Language	4.W.1b	NA	NA	NA	0
ELA	4	Writing/Language	4.W.1c	NA	NA	NA	0
ELA	4	Writing/Language	4.W.1d	NA	NA	NA	0
ELA	4	Writing/Language	4.W.2	NA	NA	NA	0
ELA	4	Writing/Language	4.W.2a	NA	SR	2	1
ELA	4	Writing/Language	4.W.2b	NA	NA	NA	0
ELA	4	Writing/Language	4.W.2c	NA	NA	NA	0
ELA	4	Writing/Language	4.W.2d	NA	NA	NA	0
ELA	4	Writing/Language	4.W.2e	NA	TE	3	1
ELA	4	Writing/Language	4.W.3	NA	NA	NA	0
ELA	4	Writing/Language	4.W.3a	NA	NA	NA	0
ELA	4	Writing/Language	4.W.3b	NA	NA	NA	0
ELA	4	Writing/Language	4.W.3c	NA	NA	NA	0
ELA	4	Writing/Language	4.W.3d	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 4 Table of Specifications Spring 2016							
ELA	4	Writing/Language	4.W.3e	NA	TE	3	1
ELA	4	Writing/Language	4.W.4	NA	NA	NA	0
ELA	4	Writing/Language	4.W.5	NA	NA	NA	0
ELA	4	Writing/Language	4.W.6	NA	NA	NA	0
ELA	4	Writing/Language	4.W.7	NA	NA	NA	0
ELA	4	Writing/Language	4.W.8	NA	SR, TE	2	4
ELA	4	Research: Writing/Language	4.W.9	L	TDA	3	1
ELA	4	Writing/Language	4.W.9a	NA	NA	NA	0
ELA	4	Writing/Language	4.W.9b	NA	NA	NA	0
ELA	4	Writing/Language	4.W.10	NA	NA	NA	0
ELA	4	Listening	4.SL.1	NA	NA	NA	0
ELA	4	Listening	4.SL.1a	NA	NA	NA	0
ELA	4	Listening	4.SL.1b	NA	NA	NA	0
ELA	4	Listening	4.SL.1c	NA	NA	NA	0
ELA	4	Listening	4.SL.1d	NA	NA	NA	0
ELA	4	Listening	4.SL.2	I	SR	2-3	2
ELA	4	Listening	4.SL.3	I	SR	1-3	4
ELA	4	Listening	4.SL.4	NA	NA	NA	0
ELA	4	Listening	4.SL.5	NA	NA	NA	0
ELA	4	Listening	4.SL.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 4 Table of Specifications Spring 2016

Item Type		Passage Type	
SR	Multiple-choice, multiple selected response and evidence-based selected response	**L**	**Literary** (poems, narratives, realistic fiction, historical fiction, fantasy, legends/myths, etc.)
TE	highlighting text, drop-down pull list, drag and drop, drag and paste, list input, matching	**I**	**Informational** (biographies, instructional/how-tos, articles, essays, science and social studies topics, etc.)
TDA	Text Dependent Analysis - One constructed response item aligned to a reading comprehension indicator as well as the indicators designated on the TDA rubric		

Reporting Categories	Total Items	Total Points
Reading	16	20
Writing/Language	13	28
Listening	6	8
	35	**56**

Wisconsin Forward Exam English Language Arts Grade 5 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	5	Reading	5.RL.1	L	SR, TE	2	2
ELA	5	Reading	5.RL.2	L	SR	2	2
ELA	5	Reading	5.RL.3	NA	NA	NA	0
ELA	5	Reading	5.RL.4	L	SR	2	1
ELA	5	Reading	5.RL.5	NA	NA	NA	0
ELA	5	Reading	5.RL.6	L	SR	2	1
ELA	5	Reading	5.RL.7	NA	NA	NA	0
ELA	5	Reading	5.RL.9	L	SR	3	2
ELA	5	Reading	5.RL.10	NA	NA	NA	0
ELA	5	Reading	5.RI.1	I	SR	3	1
ELA	5	Reading	5.RI.2	NA	NA	NA	0
ELA	5	Reading	5.RI.3	NA	NA	NA	0
ELA	5	Reading	5.RI.4	NA	NA	NA	0
ELA	5	Reading	5.RI.5	I	TE	2	1
ELA	5	Reading	5.RI.6	I	SR	3	1
ELA	5	Reading	5.RI.7	NA	NA	NA	0
ELA	5	Reading	5.RI.8	NA	NA	NA	0
ELA	5	Reading	5.RI.9	I	SR	3	1
ELA	5	Reading	5.RI.10	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1a	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1b	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1c	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1d	NA	NA	NA	0
ELA	5	Writing/Language	5.L.1e	NA	NA	NA	0
ELA	5	Writing/Language	5.L.2	NA	SR	2	1
ELA	5	Writing/Language	5.L.2a	NA	NA	NA	0
ELA	5	Writing/Language	5.L.2b	NA	SR	2	1
ELA	5	Writing/Language	5.L.2c	NA	NA	NA	0
ELA	5	Writing/Language	5.L.2d	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 5 Table of Specifications Spring 2016							
ELA	5	Writing/Language	5.L.2e	NA	NA	NA	0
ELA	5	Writing/Language	5.L.3	NA	NA	NA	0
ELA	5	Writing/Language	5.L.3a	NA	SR	2	1
ELA	5	Writing/Language	5.L.3b	NA	NA	NA	0
ELA	5	Reading	5.L.4	I	SR	2	1
ELA	5	Reading	5.L.4a	NA	NA	NA	0
ELA	5	Reading	5.L.4b	NA	NA	NA	0
ELA	5	Reading	5.L.4c	NA	NA	NA	0
ELA	5	Reading	5.L.5	L	SR	2	2
ELA	5	Reading	5.L.5a	NA	NA	NA	0
ELA	5	Reading	5.L.5b	NA	NA	NA	0
ELA	5	Reading	5.L.5c	NA	NA	NA	0
ELA	5	Writing/Language	5.L.6	NA	NA	NA	0
ELA	5	Writing/Language	5.W.1	NA	NA	NA	0
ELA	5	Writing/Language	5.W.1a	NA	TE	2	1
ELA	5	Writing/Language	5.W.1b	NA	NA	NA	0
ELA	5	Writing/Language	5.W.1c	NA	SR	2	1
ELA	5	Writing/Language	5.W.1d	NA	NA	NA	0
ELA	5	Writing/Language	5.W.2	NA	NA	NA	0
ELA	5	Writing/Language	5.W.2a	NA	NA	NA	0
ELA	5	Writing/Language	5.W.2b	NA	SR	2	1
ELA	5	Writing/Language	5.W.2c	NA	NA	NA	0
ELA	5	Writing/Language	5.W.2d	NA	SR	2	1
ELA	5	Writing/Language	5.W.2e	NA	NA	NA	0
ELA	5	Writing/Language	5.W.3	NA	NA	NA	0
ELA	5	Writing/Language	5.W.3a	NA	TE	2	1
ELA	5	Writing/Language	5.W.3b	NA	NA	NA	0
ELA	5	Writing/Language	5.W.3c	NA	NA	NA	0
ELA	5	Writing/Language	5.W.3d	NA	NA	NA	0
ELA	5	Writing/Language	5.W.3e	NA	NA	NA	0
ELA	5	Writing/Language	5.W.4	NA	NA	NA	0
ELA	5	Writing/Language	5.W.5	NA	SR	2-3	2
ELA	5	Writing/Language	5.W.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 5 Table of Specifications Spring 2016							
ELA	5	Writing/Language	5.W.7	NA	NA	NA	0
ELA	5	Writing/Language	5.W.8	NA	SR, TE	2-3	3
ELA	5	Research: Writing/Language	5.W.9	L	TDA	3	1
ELA	5	Writing/Language	5.W.9a	NA	NA	NA	0
ELA	5	Writing/Language	5.W.9b	NA	NA	NA	0
ELA	5	Writing/Language	5.W.10	NA	NA	NA	0
ELA	5	Listening	5.SL.1	NA	NA	NA	0
ELA	5	Listening	5.SL.1a	NA	NA	NA	0
ELA	5	Listening	5.SL.1b	NA	NA	NA	0
ELA	5	Listening	5.SL.1c	NA	NA	NA	0
ELA	5	Listening	5.SL.1d	NA	NA	NA	0
ELA	5	Listening	5.SL.2	I	SR	2-3	3
ELA	5	Listening	5.SL.3	I	SR	1-3	3
ELA	5	Listening	5.SL.4	NA	NA	NA	0
ELA	5	Listening	5.SL.5	NA	NA	NA	0
ELA	5	Listening	5.SL.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 5 Table of Specifications Spring 2016

Item Type

SR	Multiple-choice, multiple selected response and evidence-based selected response
TE	highlighting text, drop-down pull list, drag and drop, drag and paste, list input, matching
TDA	Text Dependent Analysis - One constructed response item aligned to a reading comprehension indicator as well as the indicators designated on the TDA rubric

Passage Type

L	**Literary** (poems, narratives, realistic fiction, historical fiction, fantasy, legends/myths, etc.)
I	**Informational** (biographies, instructional/how-tos, articles, essays, science and social studies topics, etc.)

Reporting Categories	Total Items	Total Points
Reading	15	20
Writing/Language	14	28
Listening	6	8
	35	56

Wisconsin Forward Exam English Language Arts Grade 6 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	6	Reading	6.RL.1	L	SR	3	1
ELA	6	Reading	6.RL.2	L	SR, TE	3	2
ELA	6	Reading	6.RL.3	L	SR	3	1
ELA	6	Reading	6.RL.4	L	SR	3	1
ELA	6	Reading	6.RL.5	L	SR	2	1
ELA	6	Reading	6.RL.6	L	SR	2	1
ELA	6	Reading	6.RL.7	NA	NA	NA	0
ELA	6	Reading	6.RL.9	NA	NA	NA	0
ELA	6	Reading	6.RL.10	NA	NA	NA	0
ELA	6	Reading	6.RI.1	I	SR	2	1
ELA	6	Reading	6.RI.2	NA	NA	NA	0
ELA	6	Reading	6.RI.3	I	SR, TE	2	2
ELA	6	Reading	6.RI.4	I	TE	2	1
ELA	6	Reading	6.RI.5	I	SR	2	1
ELA	6	Reading	6.RI.6	NA	NA	NA	0
ELA	6	Reading	6.RI.7	NA	NA	NA	0
ELA	6	Reading	6.RI.8	I	SR	3	1
ELA	6	Reading	6.RI.9	I	SR	3	2
ELA	6	Reading	6.RI.10	NA	NA	NA	0
ELA	6	Writing/Language	6.L.1	NA	NA	NA	0
ELA	6	Writing/Language	6.L.1a	NA	NA	NA	0
ELA	6	Writing/Language	6.L.1b	NA	NA	NA	0
ELA	6	Writing/Language	6.L.1c	NA	TE	2	1
ELA	6	Writing/Language	6.L.1d	NA	NA	NA	0
ELA	6	Writing/Language	6.L.1e	NA	NA	NA	0
ELA	6	Writing/Language	6.L.2	NA	NA	NA	0
ELA	6	Writing/Language	6.L.2a	NA	SR	2	1
ELA	6	Writing/Language	6.L.2b	NA	TE	2	1
ELA	6	Writing/Language	6.L.3	NA	NA	NA	0
ELA	6	Writing/Language	6.L.3a	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 6 Table of Specifications Spring 2016							
ELA	6	Writing/Language	6.L.3b	NA	TE	2	1
ELA	6	Reading	6.L.4	NA	NA	NA	0
ELA	6	Reading	6.L.4a	NA	NA	NA	0
ELA	6	Reading	6.L.4b	NA	NA	NA	0
ELA	6	Reading	6.L.4c	NA	NA	NA	0
ELA	6	Reading	6.L.4d	NA	NA	NA	0
ELA	6	Reading	6.L.5	I and L	SR	1-2	2
ELA	6	Reading	6.L.5a	NA	NA	NA	0
ELA	6	Reading	6.L.5b	NA	NA	NA	0
ELA	6	Reading	6.L.5c	NA	NA	NA	0
ELA	6	Writing/Language	6.L.6	NA	NA	NA	0
ELA	6	Writing/Language	6.W.1	NA	NA	NA	0
ELA	6	Writing/Language	6.W.1a	NA	NA	NA	0
ELA	6	Writing/Language	6.W.1b	NA	NA	NA	0
ELA	6	Writing/Language	6.W.1c	NA	SR	2	1
ELA	6	Writing/Language	6.W.1d	NA	NA	NA	0
ELA	6	Writing/Language	6.W.1e	NA	NA	NA	0
ELA	6	Writing/Language	6.W.2	NA	NA	NA	0
ELA	6	Writing/Language	6.W.2a	NA	NA	NA	0
ELA	6	Writing/Language	6.W.2b	NA	NA	NA	0
ELA	6	Writing/Language	6.W.2c	NA	NA	NA	0
ELA	6	Writing/Language	6.W.2d	NA	SR	2	1
ELA	6	Writing/Language	6.W.2e	NA	SR	2	1
ELA	6	Writing/Language	6.W.2f	NA	NA	NA	0
ELA	6	Writing/Language	6.W.3	NA	NA	NA	0
ELA	6	Writing/Language	6.W.3a	NA	NA	NA	0
ELA	6	Writing/Language	6.W.3b	NA	SR	2	1
ELA	6	Writing/Language	6.W.3c	NA	NA	NA	0
ELA	6	Writing/Language	6.W.3d	NA	NA	NA	0
ELA	6	Writing/Language	6.W.3e	NA	NA	NA	0
ELA	6	Writing/Language	6.W.4	NA	NA	NA	0
ELA	6	Writing/Language	6.W.5	NA	NA	NA	0
ELA	6	Writing/Language	6.W.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 6 Table of Specifications Spring 2016							
ELA	6	Writing/Language	6.W.7	NA	NA	NA	0
ELA	6	Writing/Language	6.W.8	NA	SR, TE	2	4
ELA	6	Research: Writing/Language	6.W.9	L	TDA	3	1
ELA	6	Writing/Language	6.W.9a	NA	NA	NA	0
ELA	6	Writing/Language	6.W.9b	NA	NA	NA	0
ELA	6	Writing/Language	6.W.10	NA	NA	NA	0
ELA	6	Listening	6.SL.1	NA	NA	NA	0
ELA	6	Listening	6.SL.1a	NA	NA	NA	0
ELA	6	Listening	6.SL.1b	NA	NA	NA	0
ELA	6	Listening	6.SL.1c	NA	NA	NA	0
ELA	6	Listening	6.SL.1d	NA	NA	NA	0
ELA	6	Listening	6.SL.2	I	SR	2-3	2
ELA	6	Listening	6.SL.3	I	SR	2-3	4
ELA	6	Listening	6.SL.4	NA	NA	NA	0
ELA	6	Listening	6.SL.5	NA	NA	NA	0
ELA	6	Listening	6.SL.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 6 Table of Specifications Spring 2016

Item Type		Passage Type	
SR	Multiple-choice, multiple selected response and evidence-based selected response	**L**	**Literary** (poems, narratives, realistic fiction, historical fiction, fantasy, legends/myths, etc.)
TE	highlighting text, drop-down pull list, drag and drop, drag and paste, list input, matching	**I**	**Informational** (biographies, instructional/how-tos, articles, essays, science and social studies topics, etc.)
TDA	Text Dependent Analysis - One constructed response item aligned to a reading comprehension indicator as well as the indicators designated on the TDA rubric		

Reporting Categories	Total Items	Total Points
Reading	17	20
Writing/Language	13	28
Listening	6	8
	36	56

Wisconsin Forward Exam English Language Arts Grade 7 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	7	Reading	7.RL.1	L	SR	2	1
ELA	7	Reading	7.RL.2	L	SR	2-3	3
ELA	7	Reading	7.RL.3	L	SR	3	1
ELA	7	Reading	7.RL.4	L	SR	2	1
ELA	7	Reading	7.RL.5	NA	NA	NA	0
ELA	7	Reading	7.RL.6	L	TE	3	1
ELA	7	Reading	7.RL.7	NA	NA	NA	0
ELA	7	Reading	7.RL.9	NA	NA	NA	0
ELA	7	Reading	7.RL.10	NA	NA	NA	0
ELA	7	Reading	7.RI.1	I	TE	1	1
ELA	7	Reading	7.RI.2	NA	NA	NA	0
ELA	7	Reading	7.RI.3	I	SR	2	1
ELA	7	Reading	7.RI.4	NA	NA	NA	0
ELA	7	Reading	7.RI.5	I	SR	2	1
ELA	7	Reading	7.RI.6	I	SR	2	1
ELA	7	Reading	7.RI.7	NA	NA	NA	0
ELA	7	Reading	7.RI.8	I	SR	2	1
ELA	7	Reading	7.RI.9	NA	NA	NA	0
ELA	7	Reading	7.RI.10	NA	NA	NA	0
ELA	7	Writing/Language	7.L.1	NA	NA	NA	0
ELA	7	Writing/Language	7.L.1a	NA	NA	NA	0
ELA	7	Writing/Language	7.L.1b	NA	SR	2	1
ELA	7	Writing/Language	7.L.1c	NA	SR	2	1
ELA	7	Writing/Language	7.L.2	NA	SR	2	1
ELA	7	Writing/Language	7.L.2a	NA	NA	NA	0
ELA	7	Writing/Language	7.L.2b	NA	NA	NA	0
ELA	7	Writing/Language	7.L.3	NA	NA	NA	0
ELA	7	Writing/Language	7.L.3a	NA	TE	2	1
ELA	7	Reading	7.L.4	I and L	SR, TE	2	3
ELA	7	Reading	7.L.4a	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 7 Table of Specifications Spring 2016							
ELA	7	Reading	7.L.4b	NA	NA	NA	0
ELA	7	Reading	7.L.4c	NA	NA	NA	0
ELA	7	Reading	7.L.4d	NA	NA	NA	0
ELA	7	Reading	7.L.5	I	SR	2	1
ELA	7	Reading	7.L.5a	NA	NA	NA	0
ELA	7	Reading	7.L.5b	NA	NA	NA	0
ELA	7	Reading	7.L.5c	NA	NA	NA	0
ELA	7	Writing/Language	7.L.6	NA	NA	NA	0
ELA	7	Writing/Language	7.W.1	NA	NA	NA	0
ELA	7	Writing/Language	7.W.1a	NA	NA	NA	0
ELA	7	Writing/Language	7.W.1b	NA	SR	2	1
ELA	7	Writing/Language	7.W.1c	NA	NA	NA	0
ELA	7	Writing/Language	7.W.1d	NA	NA	NA	0
ELA	7	Writing/Language	7.W.1e	NA	SR	2	1
ELA	7	Writing/Language	7.W.2	NA	NA	NA	0
ELA	7	Writing/Language	7.W.2a	NA	NA	NA	0
ELA	7	Writing/Language	7.W.2b	NA	NA	NA	0
ELA	7	Writing/Language	7.W.2c	NA	NA	NA	0
ELA	7	Writing/Language	7.W.2d	NA	SR	2	1
ELA	7	Writing/Language	7.W.2e	NA	SR	2	1
ELA	7	Writing/Language	7.W.2f	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3a	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3b	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3c	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3d	NA	NA	NA	0
ELA	7	Writing/Language	7.W.3e	NA	TE	3	1
ELA	7	Writing/Language	7.W.4	NA	NA	NA	0
ELA	7	Writing/Language	7.W.5	NA	NA	NA	0
ELA	7	Writing/Language	7.W.6	NA	NA	NA	0
ELA	7	Writing/Language	7.W.7	NA	NA	NA	0
ELA	7	Writing/Language	7.W.8	NA	SR	2-3	4

Wisconsin Forward Exam English Language Arts Grade 7 Table of Specifications Spring 2016							
ELA	7	Research: Writing/Language	7.W.9	L	TDA	3	1
ELA	7	Writing/Language	7.W.9a	NA	NA	NA	0
ELA	7	Writing/Language	7.W.9b	NA	NA	NA	0
ELA	7	Writing/Language	7.W.10	NA	NA	NA	0
ELA	7	Listening	7.SL.1	NA	NA	NA	0
ELA	7	Listening	7.SL.1a	NA	NA	NA	0
ELA	7	Listening	7.SL.1b	NA	NA	NA	0
ELA	7	Listening	7.SL.1c	NA	NA	NA	0
ELA	7	Listening	7.SL.1d	NA	NA	NA	0
ELA	7	Listening	7.SL.2	I	SR	1-3	5
ELA	7	Listening	7.SL.3	I	SR	3	1
ELA	7	Listening	7.SL.4	NA	NA	NA	0
ELA	7	Listening	7.SL.5	NA	NA	NA	0
ELA	7	Listening	7.SL.6	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 7 Table of Specifications Spring 2016

Item Type		Passage Type	
SR	Multiple-choice, multiple selected response and evidence-based selected response	**L**	**Literary** (poems, narratives, realistic fiction, historical fiction, fantasy, legends/myths, etc.)
TE	highlighting text, drop-down pull list, drag and drop, drag and paste, list input, matching	**I**	**Informational** (biographies, instructional/how-tos, articles, essays, science and social studies topics, etc.)
TDA	Text Dependent Analysis - One constructed response item aligned to a reading comprehension indicator as well as the indicators designated on the TDA rubric		

Reporting Categories	Total Items	Total Points
Reading	16	20
Writing/Language	14	28
Listening	6	8
	36	56

Wisconsin Forward Exam English Language Arts Grade 8 Table of Specifications Spring 2016

Content Area	Grade	Reporting Category	Standard Code	Passage Type	Eligible Item Types	DOK Levels Tested	Total Items
ELA	8	Reading	8.RL.1	L	SR	2	2
ELA	8	Reading	8.RL.2	NA	NA	NA	0
ELA	8	Reading	8.RL.3	L	SR	2	2
ELA	8	Reading	8.RL.4	L	SR	2	1
ELA	8	Reading	8.RL.5	NA	NA	NA	0
ELA	8	Reading	8.RL.6	L	SR	2-3	2
ELA	8	Reading	8.RL.7	NA	NA	NA	0
ELA	8	Reading	8.RL.9	NA	NA	NA	0
ELA	8	Reading	8.RL.10	NA	NA	NA	0
ELA	8	Reading	8.RI.1	I	SR	2	1
ELA	8	Reading	8.RI.2	I	SR	2	1
ELA	8	Reading	8.RI.3	I	SR	2	1
ELA	8	Reading	8.RI.4	NA	NA	NA	0
ELA	8	Reading	8.RI.5	NA	NA	NA	0
ELA	8	Reading	8.RI.6	NA	NA	NA	0
ELA	8	Reading	8.RI.7	NA	NA	NA	0
ELA	8	Reading	8.RI.8	I	SR	2	1
ELA	8	Reading	8.RI.9	I	TE	2	1
ELA	8	Reading	8.RI.10	NA	NA	NA	0
ELA	8	Writing/Language	8.L.1	NA	NA	NA	0
ELA	8	Writing/Language	8.L.1a	NA	NA	NA	0
ELA	8	Writing/Language	8.L.1b	NA	NA	NA	0
ELA	8	Writing/Language	8.L.1c	NA	NA	NA	0
ELA	8	Writing/Language	8.L.1d	NA	NA	NA	0
ELA	8	Writing/Language	8.L.2	NA	NA	NA	0
ELA	8	Writing/Language	8.L.2a	NA	SR	2	1
ELA	8	Writing/Language	8.L.2b	NA	SR	2	2
ELA	8	Writing/Language	8.L.2c	NA	TE	2	1
ELA	8	Writing/Language	8.L.3	NA	NA	NA	0
ELA	8	Writing/Language	8.L.3a	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 8 Table of Specifications Spring 2016							
ELA	8	Reading	8.L.4	I and L	SR	2	3
ELA	8	Reading	8.L.4a	NA	NA	NA	0
ELA	8	Reading	8.L.4b	NA	NA	NA	0
ELA	8	Reading	8.L.4c	NA	NA	NA	0
ELA	8	Reading	8.L.4d	NA	NA	NA	0
ELA	8	Reading	8.L.5	I	TE	2	1
ELA	8	Reading	8.L.5a	NA	NA	NA	0
ELA	8	Reading	8.L.5b	NA	NA	NA	0
ELA	8	Reading	8.L.5c	NA	NA	NA	0
ELA	8	Writing/Language	8.L.6	NA	NA	NA	0
ELA	8	Writing/Language	8.W.1	NA	NA	NA	0
ELA	8	Writing/Language	8.W.1a	NA	NA	NA	0
ELA	8	Writing/Language	8.W.1b	NA	NA	NA	0
ELA	8	Writing/Language	8.W.1c	NA	SR	2	1
ELA	8	Writing/Language	8.W.1d	NA	NA	NA	0
ELA	8	Writing/Language	8.W.1e	NA	SR	2	1
ELA	8	Writing/Language	8.W.2	NA	NA	NA	0
ELA	8	Writing/Language	8.W.2a	NA	NA	NA	0
ELA	8	Writing/Language	8.W.2b	NA	SR	2	1
ELA	8	Writing/Language	8.W.2c	NA	NA	NA	0
ELA	8	Writing/Language	8.W.2d	NA	NA	NA	0
ELA	8	Writing/Language	8.W.2e	NA	NA	NA	0
ELA	8	Writing/Language	8.W.2f	NA	NA	NA	0
ELA	8	Writing/Language	8.W.3	NA	NA	NA	0
ELA	8	Writing/Language	8.W.3a	NA	NA	NA	0
ELA	8	Writing/Language	8.W.3b	NA	NA	NA	0
ELA	8	Writing/Language	8.W.3c	NA	SR	2	1
ELA	8	Writing/Language	8.W.3d	NA	SR	2	1
ELA	8	Writing/Language	8.W.3e	NA	NA	NA	0
ELA	8	Writing/Language	8.W.4	NA	NA	NA	0
ELA	8	Writing/Language	8.W.5	NA	NA	NA	0
ELA	8	Writing/Language	8.W.6	NA	NA	NA	0
ELA	8	Writing/Language	8.W.7	NA	NA	NA	0

Wisconsin Forward Exam English Language Arts Grade 8 Table of Specifications Spring 2016							
ELA	8	Writing/Language	8.W.8	NA	SR, TE	2-3	4
ELA	8	Research: Writing/Language	8.W.9	L	TDA	3	1
ELA	8	Writing/Language	8.W.9a	NA	NA	NA	0
ELA	8	Writing/Language	8.W.9b	NA	NA	NA	0
ELA	8	Writing/Language	8.W.10	NA	NA	NA	0
ELA	8	Listening	8.SL.1	NA	NA	NA	0
ELA	8	Listening	8.SL.1a	NA	NA	NA	0
ELA	8	Listening	8.SL.1b	NA	NA	NA	0
ELA	8	Listening	8.SL.1c	NA	NA	NA	0
ELA	8	Listening	8.SL.1d	NA	NA	NA	0
ELA	8	Listening	8.SL.2	I	SR	2-3	4
ELA	8	Listening	8.SL.3	I	SR	2	2
ELA	8	Listening	8.SL.4	NA	NA	NA	0
ELA	8	Listening	8.SL.5	NA	NA	NA	0
ELA	8	Listening	8.SL.6	NA	NA	NA	0

Item Type

SR	Multiple-choice, multiple selected response and
TE	highlighting text, drop-down pull list, drag
TDA	Text Dependent Analysis - One constructed

Passage Type

L	**Literary** (poems, narratives, realistic
I	**Informational** (biographies,

Reporting Categories	Total Items	Total Points
Reading	16	20
Writing/Language	14	28
Listening	6	8
	36	56

Appendix D

Wisconsin Forward Exam Mathematics Grades 3-8 Table of Specifications

Wisconsin Forward Exam Grade 3 Mathematics Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Math	3	Number and Operations–Fractions	3.NF.1	MC	2	1
Math	3	Number and Operations–Fractions	3.NF.2a-b	MC	1-2	3
Math	3	Number and Operations–Fractions	3.NF.3a-d	MC,SA,TE	1-3	4
Math	3	Number and Operations in Base Ten	3.NBT.1	MC,TE	2	3
Math	3	Number and Operations in Base Ten	3.NBT.2	MC	1-2	3
Math	3	Number and Operations in Base Ten	3.NBT.3	MC,SA	1-2	2
Math	3	Operations and Algebraic Thinking	3.OA.1	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.2	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.3	SA	2	1
Math	3	Operations and Algebraic Thinking	3.OA.4	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.5	MC	1	1
Math	3	Operations and Algebraic Thinking	3.OA.6	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.7	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.8	MC	2	1
Math	3	Operations and Algebraic Thinking	3.OA.9	MC	2	1
Math	3	Geometry	3.G.1	MC,SA,TE	1-3	4
Math	3	Geometry	3.G.2	MC,SA	2	3
Math	3	Measurement and Data	3.MD.1	MC,TE	2	2
Math	3	Measurement and Data	3.MD.2	MC	1	1
Math	3	Measurement and Data	3.MD.3	SA	2	1
Math	3	Measurement and Data	3.MD.4	MC	2	1
Math	3	Measurement and Data	3.MD.5a-b	MC	1	1
Math	3	Measurement and Data	3.MD.6	MC	2	1
Math	3	Measurement and Data	3.MD.7a-d	MC	2	1
Math	3	Measurement and Data	3.MD.8	MC,SA	2-3	2

Wisconsin Forward Exam Grade 3 Mathematics Table of Specifications Spring 2016

Item Types	
MC	multiple-choice
SA	short answer numeric response
TE	analog clock, drag and drop, hot spot, matching

Reporting Categories	Total Items	Total Points
Geometry	7	7
Measurement and Data	10	10
Number and Operations in Base Ten	8	8
Number and Operations–Fractions	8	8
Operations and Algebraic Thinking	9	9
	42	42

Wisconsin Forward Exam Grade 4 Mathematics Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Math	4	Number and Operations–Fractions	4.NF.1	MC	1-2	2
Math	4	Number and Operations–Fractions	4.NF.2	MC	1	1
Math	4	Number and Operations–Fractions	4.NF.3a-d	MC	2	2
Math	4	Number and Operations–Fractions	4.NF.4a-c	MC	2	1
Math	4	Number and Operations–Fractions	4.NF.5	MC	2	1
Math	4	Number and Operations–Fractions	4.NF.6	SA,TE	2	2
Math	4	Number and Operations–Fractions	4.NF.7	MC	2	1
Math	4	Number and Operations in Base Ten	4.NBT.1	MC,TE	2	2
Math	4	Number and Operations in Base Ten	4.NBT.2	MC,SA	1-2	2
Math	4	Number and Operations in Base Ten	4.NBT.3	MC	2	1
Math	4	Number and Operations in Base Ten	4.NBT.4	TE	2	1
Math	4	Number and Operations in Base Ten	4.NBT.5	MC,SA	2	2
Math	4	Number and Operations in Base Ten	4.NBT.6	SA	2	1
Math	4	Operations and Algebraic Thinking	4.OA.1	MC	1-2	2
Math	4	Operations and Algebraic Thinking	4.OA.2	MC	2	1
Math	4	Operations and Algebraic Thinking	4.OA.3	MC	2	2
Math	4	Operations and Algebraic Thinking	4.OA.4	MC,SA	1	2
Math	4	Operations and Algebraic Thinking	4.OA.5	MC	2	3
Math	4	Geometry	4.G.1	MC	2	2
Math	4	Geometry	4.G.2	MC,SA	1-2	3
Math	4	Geometry	4.G.3	MC	2	2
Math	4	Measurement and Data	4.MD.1	TE	2	1
Math	4	Measurement and Data	4.MD.2	MC	2	2
Math	4	Measurement and Data	4.MD.3	MC	2	1

Wisconsin Forward Exam Grade 4 Mathematics Table of Specifications Spring 2016						
Math	4	Measurement and Data	4.MD.4	MC,TE	2	2
Math	4	Measurement and Data	4.MD.5a-b	MC,SA	1-2	2
Math	4	Measurement and Data	4.MD.6	MC	1	1
Math	4	Measurement and Data	4.MD.7	MC	2	1

Item Types	
MC	multiple-choice
SA	short answer numeric response
TE	drag and drop, drop-down pull list, line plot, number line graph

Reporting Categories	Total Items	Total Points
Geometry	7	7
Measurement and Data	10	10
Number and Operations in Base Ten	9	9
Number and Operations–Fractions	10	10
Operations and Algebraic Thinking	10	10
	46	46

Wisconsin Forward Exam Grade 5 Mathematics Table of Specifications Spring 2016						
Content Area	Grade	Reporting Category	Standard Code	Eligible Item Types	DOK Levels Tested	Total Items
Math	5	Number and Operations–Fractions	5.NF.1	MS	2	1
Math	5	Number and Operations–Fractions	5.NF.2	TE	2	1
Math	5	Number and Operations–Fractions	5.NF.3	MC	1-2	2
Math	5	Number and Operations–Fractions	5.NF.4a-b	MC	2	1
Math	5	Number and Operations–Fractions	5.NF.5a-b	MC,SA	2	2
Math	5	Number and Operations–Fractions	5.NF.6	MC	1	1
Math	5	Number and Operations–Fractions	5.NF.7a-c	SA	2	1
Math	5	Number and Operations in Base Ten	5.NBT.1	MC	2	1
Math	5	Number and Operations in Base Ten	5.NBT.2	MC	1	1
Math	5	Number and Operations in Base Ten	5.NBT.3a-b	SA	2	1
Math	5	Number and Operations in Base Ten	5.NBT.4	MC	1	1
Math	5	Number and Operations in Base Ten	5.NBT.5	MC,TE	2	2
Math	5	Number and Operations in Base Ten	5.NBT.6	SA	2	1
Math	5	Number and Operations in Base Ten	5.NBT.7	MC	1-2	2
Math	5	Operations and Algebraic Thinking	5.OA.1	MC,MS,TE	2	3
Math	5	Operations and Algebraic Thinking	5.OA.2	MC	1-2	3
Math	5	Operations and Algebraic Thinking	5.OA.3	MC,MS,TE	2	3

Wisconsin Forward Exam Grade 5 Mathematics Table of Specifications Spring 2016						
Math	5	Geometry	5.G.1	MC, MS, SA	1-2	4
Math	5	Geometry	5.G.2	MC,SA	2	3
Math	5	Geometry	5.G.3	MS	2	1
Math	5	Geometry	5.G.4	MS	2	1
Math	5	Measurement and Data	5.MD.1	MC,SA	2	3
Math	5	Measurement and Data	5.MD.2	MC,TE	2	2
Math	5	Measurement and Data	5.MD.3a-b	MC	2	2
Math	5	Measurement and Data	5.MD.4	MC	2	1
Math	5	Measurement and Data	5.MD.5a-c	MC,SA	2	2

Item Types	
MC	multiple-choice
MS	multiple selected response
SA	short answer numeric response
TE	coordinate graph, drag and drop, line plot

Reporting Categories	Total Items	Total Points
Geometry	9	9
Measurement and Data	10	10
Number and Operations in Base Ten	9	9
Number and Operations–Fractions	9	9
Operations and Algebraic Thinking	9	9
	46	46

Wisconsin Forward Exam Grade 6 Mathematics Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Math	6	The Number System	6.NS.1	SA	2	1
Math	6	The Number System	6.NS.2	MC	2	1
Math	6	The Number System	6.NS.3	MC	2	2
Math	6	The Number System	6.NS.4	SA	2	1
Math	6	The Number System	6.NS.5	MC	1	1
Math	6	The Number System	6.NS.6a-c	MC,TE	1-2	2
Math	6	The Number System	6.NS.7a-d	TE	2	1
Math	6	The Number System	6.NS.8	MC	2	2
Math	6	Ratios and Proportional Relationships	6.RP.1	MC,SA,TE	1-2	3
Math	6	Ratios and Proportional Relationships	6.RP.2	MC	2	1
Math	6	Ratios and Proportional Relationships	6.RP.3a-d	MC,TE	1-3	3
Math	6	Expressions and Equations	6.EE.1	MC	2	1
Math	6	Expressions and Equations	6.EE.2a-c	MC	2	2
Math	6	Expressions and Equations	6.EE.3	NA	NA	NA
Math	6	Expressions and Equations	6.EE.4	TE	2	1
Math	6	Expressions and Equations	6.EE.5	MC	1-2	2
Math	6	Expressions and Equations	6.EE.6	MC,MS	2-3	2
Math	6	Expressions and Equations	6.EE.7	SA	2	1
Math	6	Expressions and Equations	6.EE.8	MC	2	1
Math	6	Expressions and Equations	6.EE.9	SA	2	1
Math	6	Geometry	6.G.1	MC,SA	2	2
Math	6	Geometry	6.G.2	MC,SA	2	2
Math	6	Geometry	6.G.3	MC	2	2
Math	6	Geometry	6.G.4	MS	2	1
Math	6	Statistics and Probability	6.SP.1	MC,MS	2	2
Math	6	Statistics and Probability	6.SP.2	MC	2	1

Wisconsin Forward Exam Grade 6 Mathematics Table of Specifications Spring 2016						
Math	6	Statistics and Probability	6.SP.3	MC	1-2	2
Math	6	Statistics and Probability	6.SP.4	MC	2	2
Math	6	Statistics and Probability	6.SP.5a-d	MC,MS	2	3

Item Types	
MC	multiple-choice
MS	multiple selected response
SA	short answer numeric response
TE	coordinate graph, drag and drop, matching

Reporting Categories	Total Items	Total Points
Expressions and Equations	11	11
Geometry	7	7
The Number System	11	11
Ratios and Proportional Relationships	7	7
Statistics and Probability	10	10
	46	46

Wisconsin Forward Exam Grade 7 Mathematics Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Math	7	The Number System	7.NS.1a-d	MC,SA	2	2
Math	7	The Number System	7.NS.2a-d	MC,SA,TE	1-2	3
Math	7	The Number System	7.NS.3	MC	2	2
Math	7	Ratios and Proportional Relationships	7.RP.1	MC,SA	2	2
Math	7	Ratios and Proportional Relationships	7.RP.2a-d	MC,SA	2	3
Math	7	Ratios and Proportional Relationships	7.RP.3	MC	2	3
Math	7	Expressions and Equations	7.EE.1	MC,MS	2	2
Math	7	Expressions and Equations	7.EE.2	MC	2	2
Math	7	Expressions and Equations	7.EE.3	MC,MS	2	3
Math	7	Expressions and Equations	7.EE.4a-b	MC,SA,TE	2	3
Math	7	Geometry	7.G.1	MC	2	1
Math	7	Geometry	7.G.2	TE	2	1
Math	7	Geometry	7.G.3	MC	2	1
Math	7	Geometry	7.G.4	MC	2	2
Math	7	Geometry	7.G.5	SA	2	2
Math	7	Geometry	7.G.6	MC,TE	2	3
Math	7	Statistics and Probability	7.SP.1	MC	2	2
Math	7	Statistics and Probability	7.SP.2	MC	2	1
Math	7	Statistics and Probability	7.SP.3	MC	2	2
Math	7	Statistics and Probability	7.SP.4	MC	2	1
Math	7	Statistics and Probability	7.SP.5	MC	2	1
Math	7	Statistics and Probability	7.SP.6	SA	2	1
Math	7	Statistics and Probability	7.SP.7a-c	MS,TE	2	2
Math	7	Statistics and Probability	7.SP.7a-b	MS	2	1

Wisconsin Forward Exam Grade 7 Mathematics Table of Specifications Spring 2016

Item Types	
MC	multiple-choice
MS	multiple selected response
SA	short answer numeric response
TE	coordinate graph, drag and drop, hot spot

Reporting Categories	Total Items	Total Points
Expressions and Equations	10	10
Geometry	10	10
The Number System	7	7
Ratios and Proportional Relationships	8	8
Statistics and Probability	11	11
	46	46

Wisconsin Forward Exam Grade 8 Mathematics Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Math	8	The Number System	8.NS.1	MC,MS,SA	1-2	5
Math	8	The Number System	8.NS.2	MC,SA	2	3
Math	8	Expressions and Equations	8.EE.1	MC	2	1
Math	8	Expressions and Equations	8.EE.2	MC	1-2	2
Math	8	Expressions and Equations	8.EE.3	MC	2	1
Math	8	Expressions and Equations	8.EE.4	SA	2	1
Math	8	Expressions and Equations	8.EE.5	MC,TE	2	2
Math	8	Expressions and Equations	8.EE.6	MC	2	1
Math	8	Expressions and Equations	8.EE.7a-b	SA	2	1
Math	8	Expressions and Equations	8.EE.8a-c	SA	2	1
Math	8	Functions	8.F.1	MC,MS	1,3	2
Math	8	Functions	8.F.2	MC	2	2
Math	8	Functions	8.F.3	TE	2	1
Math	8	Functions	8.F.4	MC,SA	2	2
Math	8	Functions	8.F.5	MC,MS	2	3
Math	8	Geometry	8.G.1a-c	MC	2	1
Math	8	Geometry	8.G.2	MS,TE	2	2
Math	8	Geometry	8.G.3	MC	2	1
Math	8	Geometry	8.G.4	NA	NA	NA
Math	8	Geometry	8.G.5	MC,TE	2-3	3
Math	8	Geometry	8.G.6	MC	2	1
Math	8	Geometry	8.G.7	NA	NA	NA
Math	8	Geometry	8.G.8	MC	2	1
Math	8	Geometry	8.G.9	SA	2	1
Math	8	Statistics and Probability	8.SP.1	MC	2	2
Math	8	Statistics and Probability	8.SP.2	MC	2	1
Math	8	Statistics and Probability	8.SP.3	MC	2	3
Math	8	Statistics and Probability	8.SP.4	MC,TE	2	2

Wisconsin Forward Exam Grade 8 Mathematics Table of Specifications Spring 2016

Item Types	
MC	multiple-choice
MS	multiple selected response
SA	short answer numeric response
TE	coordinate graph, drag and drop, hot spot

Reporting Categories	Total Items	Total Points
Expressions and Equations	10	10
Functions	10	10
Geometry	10	10
The Number System	8	8
Statistics and Probability	8	8
	46	46

Appendix E

Wisconsin Forward Exam Science Grades 4 and 8 Table of Specifications

Wisconsin Forward Exam Grade 4 Science Table of Specifications Spring 2016							
Content Area	**Grade**	**Reporting Category**	**Anchor Standard**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Science	4	Science Connections	Standard A: Science Connections	A.4.1	NA	NA	0
Science	4	Science Connections	Standard A: Science Connections	A.4.2	MC	3	1
Science	4	Science Connections	Standard A: Science Connections	A.4.3	MC	2	1
Science	4	Science Connections	Standard A: Science Connections	A.4.4	MC	2	2
Science	4	Science Connections	Standard A: Science Connections	A.4.5	NA	NA	0
Science	4	Nature of Science	Standard B: Nature of Science	B.4.1	MC	2	2
Science	4	Nature of Science	Standard B: Nature of Science	B.4.2	MC	2	1
Science	4	Nature of Science	Standard B: Nature of Science	B.4.3	MC	2	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.1	MC	2	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.2	MC	2, 3	2
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.3	MC	2	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.4	MC	1	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.5	MC	1	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.6	MC	2	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.7	MC	3	1
Science	4	Science Inquiry	Standard C: Science Inquiry	C.4.8	NA	NA	0
Science	4	Physical Science	Standard D: Physical Science Properties of Earth Materials	D.4.1	NA	NA	0
Science	4	Physical Science	Standard D: Physical Science Properties of Earth Materials	D.4.2	NA	NA	0
Science	4	Physical Science	Standard D: Physical Science Properties of Earth Materials	D.4.3	MC	1	1
Science	4	Physical Science	Standard D: Physical Science Properties of Earth Materials	D.4.4	MC	1	2
Science	4	Physical Science	Standard D: Physical Science Properties of Earth Materials	D.4.5	NA	NA	0
Science	4	Physical Science	Standard D: Physical Science Position and Motion of Objects	D.4.6	NA	NA	0

Wisconsin Forward Exam Grade 4 Science Table of Specifications Spring 2016							
Science	4	Physical Science	Standard D: Physical Science Light, Heat, Electricity, and Magnetism	D.4.8	MC	1, 2	2
Science	4	Earth and Space Science	Standard E: Earth and Space Science Properties of Earth Materials	E.4.1	MC	1	1
Science	4	Earth and Space Science	Standard E: Earth and Space Science Properties of Earth Materials	E.4.2	NA	NA	0
Science	4	Earth and Space Science	Standard E: Earth and Space Science Properties of Earth Materials	E.4.3	NA	NA	0
Science	4	Earth and Space Science	Standard E: Earth and Space Science Objects in the Sky	E.4.4	MC	2	1
Science	4	Earth and Space Science	Standard E: Earth and Space Science Changes in the Earth and Sky	E.4.5	MC	1	1
Science	4	Earth and Space Science	Standard E: Earth and Space Science Changes in the Earth and Sky	E.4.6	MC	1, 2	2
Science	4	Earth and Space Science	Standard E: Earth and Space Science Changes in the Earth and Sky	E.4.7	NA	NA	0
Science	4	Earth and Space Science	Standard E: Earth and Space Science Changes in the Earth and Sky	E.4.8	MC	1	1
Science	4	Life & Environmental Science	Standard F: Life and Environmental	F.4.1	MC	1, 2	2
Science	4	Life & Environmental Science	Standard F: Life and Environmental	F.4.2	MC	1, 3	2
Science	4	Life & Environmental Science	Standard F: Life and Environmental	F.4.3	MC	2	1
Science	4	Life & Environmental Science	Standard F: Life and Environmental	F.4.4	MC	2	1
Science	4	Science Applications	Standard G: Science Applications	G.4.1	MC	2	1
Science	4	Science Applications	Standard G: Science Applications	G.4.2	NA	NA	0
Science	4	Science Applications	Standard G: Science Applications	G.4.3	MC	2	1
Science	4	Science Applications	Standard G: Science Applications	G.4.4	MC	2	1
Science	4	Science Applications	Standard G: Science Applications	G.4.5	MC	1	1

Wisconsin Forward Exam Grade 4 Science Table of Specifications Spring 2016							
Science	4	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.4.1	NA	NA	0
Science	4	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.4.2	MC	2	1
Science	4	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.4.3	MC	1	1
Science	4	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.4.4	MC	2	1
		Item Type					
		MC	Multiple-choice			**Total**	**40**

Wisconsin Forward Exam Grade 8 Science Table of Specifications Spring 2016							
Content Area	**Grade**	**Reporting Category**	**Anchor Standard**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Science	8	Science Connections	Standard A: Science Connections	A.8.1	NA	NA	0
Science	8	Science Connections	Standard A: Science Connections	A.8.2	NA	NA	0
Science	8	Science Connections	Standard A: Science Connections	A.8.3	MC	2	1
Science	8	Science Connections	Standard A: Science Connections	A.8.4	NA	NA	0
Science	8	Science Connections	Standard A: Science Connections	A.8.5	MC	1	1
Science	8	Science Connections	Standard A: Science Connections	A.8.6	MC	2	1
Science	8	Science Connections	Standard A: Science Connections	A.8.7	NA	NA	0
Science	8	Science Connections	Standard A: Science Connections	A.8.8	NA	NA	0
Science	8	Nature of Science	Standard B: Nature of Science	B.8.1	MC	1	1
Science	8	Nature of Science	Standard B: Nature of Science	B.8.2	NA	NA	0
Science	8	Nature of Science	Standard B: Nature of Science	B.8.3	MC	2	1
Science	8	Nature of Science	Standard B: Nature of Science	B.8.4	MC	2	1
Science	8	Nature of Science	Standard B: Nature of Science	B.8.5	NA	NA	0
Science	8	Nature of Science	Standard B: Nature of Science	B.8.6	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.1	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.2	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.3	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.4	MC	2	2
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.5	NA	NA	0
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.6	MC	2	2

Wisconsin Forward Exam Grade 8 Science Table of Specifications Spring 2016							
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.7	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.8	NA	NA	0
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.9	NA	NA	0
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.10	MC	2	1
Science	8	Science Inquiry	Standard C: Science Inquiry Performance Standards	C.8.11	NA	NA	0
Science	8	Physical Science	Standard D: Physical Science Properties and Changes of Properties in Matter	D.8.1	NA	NA	0
Science	8	Physical Science	Standard D: Physical Science Properties and Changes of Properties in Matter	D.8.2	MC	2	1
Science	8	Physical Science	Standard D: Physical Science Properties and Changes of Properties in Matter	D.8.3	MC	2	1
Science	8	Physical Science	Standard D: Physical Science Properties and Changes of Properties in Matter	D.8.4	NA	NA	0
Science	8	Physical Science	Standard D: Physical Science Motions and Forces	D.8.5	NA	NA	0
Science	8	Physical Science	Standard D: Physical Science Motions and Forces	D.8.6	MC	2	2
Science	8	Physical Science	Standard D: Physical Science Motions and Forces	D.8.7	NA	NA	0
Science	8	Physical Science	Standard D: Physical Science Transfer of Energy	D.8.8	MC	2	2
Science	8	Physical Science	Standard D: Physical Science Transfer of Energy	D.8.9	NA	NA	0

Wisconsin Forward Exam Science Table of Specifications Spring 2016							
Science	8	Physical Science	Standard D: Physical Science Transfer of Energy	D.8.10	NA	NA	0
Science	8	Earth and Space Science	Standard E: Earth and Space Science Structure of Earth System	E.8.1	MC	2	1
Science	8	Earth and Space Science	Standard E: Earth and Space Science Structure of Earth System	E.8.2	MC	2	1
Science	8	Earth and Space Science	Standard E: Earth and Space Science Structure of Earth System	E.8.3	MC	2	2
Science	8	Earth and Space Science	Standard E: Earth and Space Science Structure of Earth System	E.8.4	NA	NA	0
Science	8	Earth and Space Science	Standard E: Earth and Space Science Earth's History	E.8.5	MC	2	1
Science	8	Earth and Space Science	Standard E: Earth and Space Science Earth's History	E.8.6	NA	NA	0
Science	8	Earth and Space Science	Standard E: Earth and Space Science Earth in the Solar System	E.8.7	MC	1	1
Science	8	Earth and Space Science	Standard E: Earth and Space Science Earth in the Solar System	E.8.8	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental	F.8.1	MC	1	2
Science	8	Life & Environmental Science	Standard F: Life and Environmental	F.8.2	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental	F.8.3	NA	NA	0

Wisconsin Forward Exam Science Table of Specifications Spring 2016							
Science	8	Life & Environmental Science	Standard F: Life and Environmental Reproduction and Heredity	F.8.4	MC	1	1
Science	8	Life & Environmental Science	Standard F: Life and Environmental Reproduction and Heredity	F.8.5	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental Regulation and Behavior	F.8.6	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental Regulation and Behavior	F.8.7	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental Populations and Ecosystems	F.8.8	MC	1 , 2	3
Science	8	Life & Environmental Science	Standard F: Life and Environmental Diversity and Adaptations of Organisms	F.8.9	NA	NA	0
Science	8	Life & Environmental Science	Standard F: Life and Environmental Diversity and Adaptations of Organisms	F.8.10	NA	NA	0
Science	8	Science Applications	Standard G: Science Applications	G.8.1	MC	2	1
Science	8	Science Applications	Standard G: Science Applications	G.8.2	NA	NA	0
Science	8	Science Applications	Standard G: Science Applications	G.8.3	MC	2	1
Science	8	Science Applications	Standard G: Science Applications	G.8.4	NA	NA	0
Science	8	Science Applications	Standard G: Science Applications	G.8.5	NA	NA	0
Science	8	Science Applications	Standard G: Science Applications	G.8.6	MC	2	1
Science	8	Science Applications	Standard G: Science Applications	G.8.7	MC	2	1

Wisconsin Forward Exam Science Table of Specifications Spring 2016

Science	8	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.8.1	MC	2	1
Science	8	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.8.2	NA	NA	0
Science	8	Science in Personal and Social Perspectives	Standard H: Science in Personal and Social Perspectives	H.8.3	MC	2	1

Item Type	
MC	Multiple-choice

Total **40**

Appendix F

Wisconsin Forward Exam Social Studies Grades 4, 8, and 10 Table of Specifications

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Social Studies	4	Geography: People, Places, and Environments	A.4.1	SR	2	1
Social Studies	4	Geography: People, Places, and Environments	A.4.2	SR	1	1
Social Studies	4	Geography: People, Places, and Environments	A.4.3	SR	1, 2	3
Social Studies	4	Geography: People, Places, and Environments	A.4.4	SR	2	2
Social Studies	4	Geography: People, Places, and Environments	A.4.5	SR	1, 2	2
Social Studies	4	Geography: People, Places, and Environments	A.4.6	SR		0
Social Studies	4	Geography: People, Places, and Environments	A.4.7	SR		0
Social Studies	4	Geography: People, Places, and Environments	A.4.8	SR		0
Social Studies	4	Geography: People, Places, and Environments	A.4.9	SR	2	1

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	History: Time, Continuity, And Change	B.4.1	SR	3	1
Social Studies	4	History: Time, Continuity, And Change	B.4.2	SR	2	3
Social Studies	4	History: Time, Continuity, And Change	B.4.3	SR	2	1
Social Studies	4	History: Time, Continuity, And Change	B.4.4	SR		0
Social Studies	4	History: Time, Continuity, And Change	B.4.5	SR		0
Social Studies	4	History: Time, Continuity, And Change	B.4.6	SR	2	1
Social Studies	4	History: Time, Continuity, And Change	B.4.7	SR	2	1
Social Studies	4	History: Time, Continuity, And Change	B.4.8	SR	2	1
Social Studies	4	History: Time, Continuity, And Change	B.4.9	SR		0
Social Studies	4	History: Time, Continuity, And Change	B.4.10	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.1	SR	2	2
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.2	SR	2	1
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.3	SR	2	1
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.4	SR	2	1
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.5	SR	2	1
Social Studies	4	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.4.6	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.1	SR	2	1
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.2	SR	2	1
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.3	SR		0
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.4	SR	2, 3	2
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.5	SR	2	1
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.6	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	Economics: Production, Distribution, Exchange, Consumption	D.4.7	SR	3	1
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.1	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.2	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.3	SR	2	1
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.4	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.5	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.6	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.7	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.8	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.9	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.11	SR	2	2
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.12	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.13	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.14	SR		0
Social Studies	4	The Behavioral Sciences: Individuals, Institutions, And Culture	E.4.15	SR	2	3

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016

SR	Multiple-choice, multiple selected response and evidence-based selected response

Reporting Categories	**Total Items**
Geography: People, Places, and Environments	10
History: Time, Continuity, And Change	8
Political Science And Citizenship: Power, Authority, Governance, And Responsibility	7
Economics: Production, Distribution, Exchange, Consumption	6
The Behavioral Sciences: Individuals, Institutions, And Culture	7
Total	38

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Social Studies	8	Geography: People, Places, and Environments	A.8.1	SR	2	1
Social Studies	8	Geography: People, Places, and Environments	A.8.2	SR	2	4
Social Studies	8	Geography: People, Places, and Environments	A.8.3	SR		0
Social Studies	8	Geography: People, Places, and Environments	A.8.4	SR		0
Social Studies	8	Geography: People, Places, and Environments	A.8.5	SR		0
Social Studies	8	Geography: People, Places, and Environments	A.8.6	SR		0
Social Studies	8	Geography: People, Places, and Environments	A.8.8	SR	2	2
Social Studies	8	Geography: People, Places, and Environments	A.8.9	SR	2	1
Social Studies	8	Geography: People, Places, and Environments	A.8.10	SR	2	1
Social Studies	8	Geography: People, Places, and Environments	A.8.11	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	History: Time, Continuity, And Change	B.8.1	SR	1,2,3	4
Social Studies	8	History: Time, Continuity, And Change	B.8.2	SR	3	1
Social Studies	8	History: Time, Continuity, And Change	B.8.3	SR	2	1
Social Studies	8	History: Time, Continuity, And Change	B.8.4	SR	3	1
Social Studies	8	History: Time, Continuity, And Change	B.8.5	SR	3	1
Social Studies	8	History: Time, Continuity, And Change	B.8.6	SR	0	0
Social Studies	8	History: Time, Continuity, And Change	B.8.7	SR	2, 3	4
Social Studies	8	History: Time, Continuity, And Change	B.8.8	SR		0
Social Studies	8	History: Time, Continuity, And Change	B.8.9	SR		0
Social Studies	8	History: Time, Continuity, And Change	B.8.10	SR		0
Social Studies	8	History: Time, Continuity, And Change	B.8.11	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	History: Time, Continuity, And Change	B.8.12	SR	2	1
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.1	SR	3	1
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.2	SR	3	1
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.3	SR	3	1
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.4	SR	2	1
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.5	SR		0
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.6	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.7	SR		0
Social Studies	8	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.8.9	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.1	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.2	SR	2, 3	5
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.3	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.4	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.5	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.7	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.8	SR	2	1
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.9	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.10	SR		0
Social Studies	8	Economics: Production, Distribution, Exchange, Consumption	D.8.11	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.1	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.3	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.4	SR	2	2
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.5	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.7	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.8	SR	3	1

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.9	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.10	SR	2	1
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.11	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.12	SR		0
Social Studies	8	The Behavioral Sciences: Individuals, Institutions, And Culture	E.8.14	SR	2	1

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016

Item Type	
SR	Multiple-choice, multiple selected response and evidence-based selected response

Reporting Categories	Total Items
Geography: People, Places, and Environments	10
History: Time, Continuity, And Change	13
Political Science And Citizenship: Power, Authority, Governance, And Responsibility	6
Economics: Production, Distribution, Exchange, Consumption	6
The Behavioral Sciences: Individuals, Institutions, And Culture	5
Total	40

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Content Area	**Grade**	**Reporting Category**	**Standard Code**	**Eligible Item Types**	**DOK Levels Tested**	**Total Items**
Social Studies	10	Geography: People, Places, and Environments	A.10.1	SR	1, 2	2
Social Studies	10	Geography: People, Places, and Environments	A.10.2	SR	0	0
Social Studies	10	Geography: People, Places, and Environments	A.10.3	SR	2	3
Social Studies	10	Geography: People, Places, and Environments	A.10.4	SR	3	1
Social Studies	10	Geography: People, Places, and Environments	A.10.5	SR		0
Social Studies	10	Geography: People, Places, and Environments	A.10.6	SR	2	1
Social Studies	10	Geography: People, Places, and Environments	A.10.7	SR	2	1
Social Studies	10	Geography: People, Places, and Environments	A.10.8	SR	2	1
Social Studies	10	Geography: People, Places, and Environments	A.10.9	SR	2	1
Social Studies	10	Geography: People, Places, and Environments	A.10.10	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Geography: People, Places, and Environments	A.10.11	SR		0
Social Studies	10	Geography: People, Places, and Environments	A.10.12	SR		0
Social Studies	10	Geography: People, Places, and Environments	A.10.13	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.1	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.2	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.3	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.4	SR	2, 3	2
Social Studies	10	History: Time, Continuity, And Change	B.10.5	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.6	SR	2	2
Social Studies	10	History: Time, Continuity, And Change	B.10.7	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.8	SR	1, 2	2

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	History: Time, Continuity, And Change	B.10.10	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.11	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.12	SR	1, 2	2
Social Studies	10	History: Time, Continuity, And Change	B.10.13	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.14	SR	2	2
Social Studies	10	History: Time, Continuity, And Change	B.10.15	SR	2	1
Social Studies	10	History: Time, Continuity, And Change	B.10.16	SR	2	1
Social Studies	10	History: Time, Continuity, And Change	B.10.17	SR		0
Social Studies	10	History: Time, Continuity, And Change	B.10.18	SR		0
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.1	SR	1, 2	2

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.2	SR	2,3	2
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.3	SR	3	2
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.4	SR		0
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.5	SR		0
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.6	SR	3	1
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.7	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.9	SR		0
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.10	SR	2	1
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.11	SR		0
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.12	SR	2	1
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.13	SR	2	2
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.14	SR	2	1

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Political Science And Citizenship: Power, Authority, Governance, And Responsibility	C.10.15	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.1	SR	2, 3	2
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.2	SR	2, 3	2
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.3	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.4	SR	2	1
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.5	SR	2	0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.6	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.7	SR		1
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.8	SR	2, 3	2
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.10	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.11	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.12	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.13	SR		0
Social Studies	10	Economics: Production, Distribution, Exchange, Consumption	D.10.14	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.1	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.3	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.4	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.5	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.6	SR	2, 3	2
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.7	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.8	SR	2	2
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.10	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.11	SR	2	1
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.12	SR	2	1

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016						
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.13	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.14	SR	2, 3	2
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.15	SR		0
Social Studies	10	The Behavioral Sciences: Individuals, Institutions, And Cultures	E.10.17	SR		0

Wisconsin Forward Exam Social Studies Table of Specifications Spring 2016

SR	Multiple-choice, multiple selected response and evidence-based selected response

Reporting Categories	**Total Items**
Geography: People, Places, and Environments	10
History: Time, Continuity, And Change	12
Political Science And Citizenship: Power, Authority, Governance, And Responsibility	12
Economics: Production, Distribution, Exchange, Consumption	8
The Behavioral Sciences: Individuals, Institutions, And Cultures	8
Total	50

Appendix G

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:

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

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

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

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,

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

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

where

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

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

$$T_j = \frac{1}{n_j}\left\lfloor \sum_{i=1}^{k_j} T_{ij}(\theta) \right\rfloor. \tag{8}$$

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.

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

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

and

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

Solving these equations for r_j and s_j produces

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

and

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

where

$$n_j^* = \frac{\mu(\hat{T}_j \mid \theta)\lfloor 1 - \mu(\hat{T}_j \mid \theta) \rfloor}{\sigma^2(\hat{T}_j \mid \theta)} - 1. \tag{13}$$

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

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

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

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

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

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

and

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

Thus,

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

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:

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

and

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

The SPI is

$$\tilde{T}_j = \frac{p_j}{p_j + q_j} \quad (20)$$

$$= \frac{\hat{T}_j n_j^* + x_j}{n^*_j + n_j} . \quad (21)$$

The SPI can also be written in terms of the relative contribution of the prior estimate $\hat{T}_j$ and the observed proportion of maximum raw (correct score) (OPM), x_j / n_j, as

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

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

$$w_j = \frac{n_j^*}{n_j^* + n_j} . \quad (23)$$

The term n_j^* may be interpreted as the contribution of the prior in terms of theoretical numbers of items.

Check on Consistency and Adjustment of Weight Given to Prior Estimate

The item responses are assumed to be described by $P_i(\hat{\theta})$ or $P_{im}(\hat{\theta})$, depending on the type of item. Even if the IRT model accurately described item performance over examinees, their item responses grouped by strand may be multidimensional. For example, a particular examinee may be able to perform

difficult addition but not easy subtraction. Under these circumstances, it is not appropriate to pool the prior estimate, $\hat{T}_j$, with x_j / n_j. In calculating the SPI, the following statistic was used to identify examinees with unexpected performance on the strands in a test:

$$Q = \sum_{j=1}^{J} n_j \left(\frac{x_j}{n_j} - \hat{T}_j\right)^2 / (\hat{T}_j(1-\hat{T}_j)) . \quad (24)$$

If $Q \leq \chi^2(J, .10)$, the weight, w_j, is computed and the SPI is produced. If $Q > \chi^2(J, .10)$, n_j^* and subsequently w_j is set equal to 0 and the OPM is used as the estimate of strand performance.

As previously noted, the prior is estimated using an ability estimate based on responses to all the items (including the items of strand *j*) and hence is not independent of X_j. An adjustment for the overlapping information that requires minimal computation is to multiply the test information in equation 5 by the factor $(n - n_j)/n$. The application of this factor produces an "adjusted" SPI estimate that can be compared to the "unadjusted" estimate.

Possible Violations of the Assumptions

Even if the IRT model fits the test items, the responses for a given examinee, grouped by strand, may be multidimensional. In these cases, it would not be appropriate to pool the prior estimate, $\hat{T}_j$, 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,

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.

Appendix H

Glossary

Glossary: Abbreviations most commonly used in the Wisconsin Forward Exam Technical Report

2PPC: Two-parameter partial-credit item response theory model. A mathematical model that shows the relationship between student achievement on a test and the discrimination and difficulty of score points for a constructed-response item.

3PL: Three-parameter logistic item response theory model. A mathematical model that shows the relationship between student achievement on a test and a single multiple-choice item by decomposing the item into three components: difficulty, discrimination, and guessing.

AERA: American Education Research Association. A professional organization whose purpose is to advance the science of educational research and its application.

APA: American Psychological Association. A professional organization centered in psychology.

CCR: College- and Career Ready item bank. Items measuring knowledge and skills in English Language Arts and Mathematics necessary to prepare students for college and the workplace.

CR: Constructed-response item. A type of question, designed to elicit student knowledge of content, that typically comprises a question for which students create (write) a response.

DIF: Differential item functioning. The degree to which an item performs differently for one group of examinees than it performs for another group of equally able examinees. Refers to differential statistical properties of an item in two equally able groups.

DOK: Depth of knowledge. A system of describing the cognitive level a test item elicits from a student. Items are coded such that level 1 indicates students use lower cognitive levels, such as recall, to answer the item correctly; level 4 indicates students use higher cognitive levels, such as analysis skills, to answer the item correctly.

DPI: Wisconsin Department of Public Instruction. The state agency overseeing the implementation of federal and state laws related to public education in Wisconsin.

DRC: Data Recognition Corporation. A testing company partnering with DPI for delivery, scoring, and reporting of Wisconsin Forward Exam assessments.

ELA: English Language Arts. A content area in the Wisconsin Forward Exam.

ELP: English language proficiency. A student population subgroup category describing students for whom English is a second language. Students are described as fully English proficient or limited English proficient.

HOSS: Highest obtainable scale score. The highest possible scale score on a test.

IRT: Item response theory. A mathematic model that shows the relationship between

student achievement on a test and the performance on a test item.

LOSS: Lowest obtainable scale score. The lowest possible scale score on a test.

MA: Mathematics. A content area in the Wisconsin Forward Exam.

MC: Multiple-choice item. A type of question, designed to elicit student knowledge of content, that typically comprises a stem and four options. Students must select the correct option.

MH: Mantel-Haenszel ($MH\ \chi^2_{MH}$) statistic. A commonly used DIF statistic for multiple-choice items.

NCME: National Council on Measurement in Education. A professional organization centered in assessment, evaluation, testing, and educational measurement.

OP: Operational item. An item that has previously undergone field testing and contributes to a student's score in a specific content area on the Wisconsin Forward Exam.

OTTs: Online Training Tools. Provided for students to allow them a hands-on opportunity to practice answering the types of items and using the tools available in the online testing system.

SC: Science. A content area in the Wisconsin Forward Exam.

SD: Standard deviation. A measure of the variability of observations from the mean.

SEM: Standard error of measurement. An estimate of how repeated measures of a person on the same test tend to be distributed around his or her "true" score.

SES: Socioeconomic status. A student population subgroup category describing students as economically disadvantaged or not economically disadvantaged.

SMD: Standardized mean difference. A commonly used DIF statistic for constructed-response items.

SPI: Standard performance index. A content category reporting score based on items from a single content standard or domain within a given content area.

SS: Social Studies. A content area in the Wisconsin Forward Exam.

TDA: Text-dependent analysis. An item based on a passage or a multiple-passage set that each student has read during the assessment. Students must draw on basic writing skills while inferring and synthesizing information from the passage in order to develop a comprehensive, holistic essay response.

TCC: Test characteristic curve. Shows the mathematical relationship between students with varying degrees of achievement and their estimated overall test performance.

WKCE: Wisconsin Knowledge and Concepts Examination. Previous Wisconsin assessment program.