Wisconsin Forward Exam

Technical Report 2019

Submitted to
Wisconsin Department of Public Instruction
November 2019

Copyright

Developed and published under contract with the Wisconsin Department of Public Instruction by Data Recognition Corporation, 13490 Bass Lake Road, Maple Grove, MN 55311.
Copyright © 2019 by the Wisconsin Department of Public Instruction. All rights reserved. Only State of Wisconsin educators and citizens may copy, download and/or print the document, located online at http://dpi.wi.gov. Any other use or reproduction of this document, in whole or in part, requires written permission of the Wisconsin Department of Public Instruction.

Foreword

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

TABLE OF CONTENTS

Copyright **2**

Foreword **3**

Part 1: Overview **12**

1.1 Historical Background 12
1.2 Uses of Test Scores 14
1.2.1 Test-Level Scores 15
1.2.2 Standard-Level Subscores and Performance Levels 16
1.3 Technical Report Structure 17

Part 2: Test Blueprint and Item Development **21**

2.1 Test Blueprints 22
2.2 Reading Passage and Item Selection for Spring 2018 Field Test 23
2.3 Field-Testing 23
2.4 Statistical Analysis of Spring 2018 Field Test Data 24
2.5 Review of Items with Data 25
2.6 Summary 25

Part 3: Test Form Development **32**

3.1 Design of the Wisconsin Forward Exam 32
3.1.1 English Language Arts 32
3.1.2 Mathematics 33
3.1.3 Science 33
3.1.4 Social Studies 33
3.2 Test Development Process 33
3.2.1 Wisconsin Forward Test Form Creation 34
3.2.2 Item Selection 34
3.2.3 Item and Form Quality Reviews 36
3.3 DPI Approvals 37
3.4 Summary 38

Part 4: Test Administration **42**

4.1 Accessibility Resources 43
4.1.1 Universal Tools 43
4.1.2 Designated Supports 44
4.1.3 Accommodations 45
4.1.4 Translation 45
4.2 Reporting Results of Assessments Taken with Accommodations 46
4.3 Test Security 46
4.3.1 Secure Student Access 47
4.3.2 Test Security during Breaks 48
4.4 Test Administration Training 48
4.5 Summary 51

Part 5: Scoring **62**

5.1 Multiple-Choice and Multi-Select Item Scoring Process 62
5.2 Technology-Enhanced, Short-Answer, and Evidence-Based Selected Response Item Scoring Process 62
5.3 Scoring of Text-Dependent Analysis Items 63
5.3.1 Description of Scoring Rubrics and Non-score Codes 63
5.3.2 Artificial Intelligence Scoring 64
5.3.3 Handscoring Process 65
5.3.4 Handscoring System 65
5.3.5 Anchor Papers and Training Papers 65

5.3.6 Scoring Personnel and Qualifications 66
5.3.7 Scorer Training 67
5.3.8 Monitoring the Scoring Process 67
5.3.9 Final Scores 68
5.4 Inter-rater Reliability 68
5.4.1 Distribution of TDA Item Scores 68
5.5 Summary 69
Part 6: Calibration, Equating, and Deriving Scale Scores 75
6.1 Item Calibration 75
6.1.1 Calibration Models 75
6.1.2 Calibration Sample 76
6.1.3 Calibration Procedure 77
6.1.4 Calibration Software 77
6.1.5 Calibration Results 78
6.2 Test Equating: English Language Arts, Mathematics, and Social Studies 80
6.2.1 Evaluation of Anchor Items 81
6.2.2 Removal of Anchor Items 82
6.2.3 Evaluation of Equating Results 83
6.2.4 Test Scales 84
6.3 Deriving Scale Scores in the Wisconsin Forward Exam 88
6.3.1 Conditional Standard Error of Measurement 93
6.3.2 LOSS and HOSS 93
6.4 Summary 94
Part 7: Standard Setting 146
7.1 Background Information 146
7.2 Standard Setting Methodology and Process 146
7.3 Performance Level Descriptors 149
7.4 Cut Scores 149
7.5 Summary 149
Part 8: Test Results 151
8.1 Classical Item Analysis: Item Level Statistics 151
8.1.1 Flagging for a Positive Distractor Correlation 154
8.1.2 Flagging for the Item-Total Correlation 154
8.1.3 Flagging for p-Value 154
8.1.4 Flagging for Omit Rate 154
8.1.5 Speededness 154
8.1.6 Supplemental Tables on Classical Item Analysis 155
8.2 Raw Score Results 155
8.2.1 Subgroup Performance Patterns in Raw Score Results 157
8.3 Summary Statistics for Scale Scores 159
8.3.1 Subgroup Performance Patterns in Scale Score Results 160
8.4 Cut Scores and Performance Level Classifications 162
8.5 Standard Performance Index for Content Standards 164
8.6 Longitudinal Comparisons of Test Scores 167
8.7 Summary 169
Part 9: Reliability 247
9.1 Measures of Internal Consistency and Standard Error of Measurement 249
9.1.1 Conditional Standard Error of Measurement 251
9.2 Classification Consistency and Accuracy 252
9.2.1 Kolen and Kim's Method for Pattern Scoring 253
9.3 Inter-rater Reliability for TDA Items 256
9.4 Summary 259
Part 10: Validity 274

Copyright © 2019 by the Wisconsin Department of Public Instruction

10.1 Differential Item Functioning....278
10.2 Validity Evidence Based on Internal Test Structure....282
10.2.1 Correlations between Content Standards....282
10.2.2 Principal Component Analysis....283
10.3 Validity Evidence Based on Relationship with Other Variables....284
10.3.1 Correlations between Content Area Test Scores....284
10.3.2 Comparison of the Wisconsin Forward Exam and Wisconsin NAEP Impact Data....285
10.4 Test Integrity: Data Forensic Analyses....287
10.5 Standardized Test Administration....288
10.6 Summary....288
Part 11: Summary Recommendations....305
References....306

APPENDICES

Appendix A: Spring 2017 Item Review Training Slides....310
Appendix B: Spring 2018 Field Test Data Review Training Slides....337
Appendix C: Spring 2019 English Language Arts Operational Test Maps....357
Appendix D: Spring 2019 Mathematics Operational Test Maps....370
Appendix E: Spring 2019 Science Operational Test Maps....383
Appendix F: Spring 2019 Social Studies Operational Test Maps....390
Appendix G: Classical Item Analysis Results....397
Appendix H: Wisconsin Standard Performance Index Score Computation....433
Appendix I: Conditional Standard Error of Measurement with Cut Scores....441
Appendix J: Classification Consistency and Accuracy Indices by Subgroup....459
Appendix K: Glossary....489

Copyright © 2019 by the Wisconsin Department of Public Instruction

LIST OF TABLES

PART 2

Table 2-1 College- and Career-Ready Item Bank Development Activities....26
Table 2-2 Item Type Descriptions for Items on the Wisconsin Forward Exam....27
Table 2-3 English Language Arts Test Blueprints for Grades 3–8....28
Table 2-4 Mathematics Test Blueprints for Grades 3–8....29
Table 2-5 Science Test Blueprints for Grades 4 and 8....30
Table 2-6 Social Studies Test Blueprints for Grades 4, 8, and 10....30
Table 2-7 Items Reviewed at Summer 2016 Item Review....30
Table 2-8 Items Reviewed at Summer 2017 Data Review....31

PART 3

Table 3-1 English Language Arts Test Design....38
Table 3-2 Mathematics Test Design....40
Table 3-3 Science Test Design....41
Table 3-4 Social Studies Test Design....41

PART 4

Table 4-1 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 3....52
Table 4-2 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 4....53
Table 4-3 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 5....54
Table 4-4 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 6....55
Table 4-5 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 7....56
Table 4-6 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 8....57
Table 4-7 Number and Percentage of Students Using Accommodations or Designated Supports: Grade 10....58
Table 4-8 Summary Table of Manual Materials....59

PART 5

Table 5-1 TDA Item Scoring Guidelines, Grades 3–8....70
Table 5-2 TDA Item Non-scorable Codes, Grades 3–8....72
Table 5-3 TDA Item Score Distribution....73
Table 5-4 TDA Item Score Distribution: AI Engine vs. Human Scorer....73
Table 5-5 TDA Item Percentage Score Distribution: AI Engine vs. Human Scorer....74

PART 6

Table 6-A Example of Item Parameters for a Test....89
Table 6-B Example of Item Response Pattern....90
Table 6-1 English Language Arts Calibration Sample Demographics Compared to Population....95
Table 6-2 Mathematics Calibration Sample Demographics Compared to Population....98
Table 6-3 Science Calibration Sample Demographics Compared to Population....101
Table 6-4 Social Studies Calibration Sample Demographics Compared to Population....102
Table 6-5 Item Flagged Based on Yen's Q1....104
Table 6-6 Equating Evaluation Results, Stocking and Lord Method....105
Table 6-7 Statistics Comparing IRT Item-Ability Regression Curves for Flagged Anchor Items....105
Table 6-8 Scale Transformation Constants....106
Table 6-9 Scoring Table for English Language Arts Grade 3....107
Table 6-10 Scoring Table for English Language Arts Grade 4....108
Table 6-11 Scoring Table for English Language Arts Grade 5....109
Table 6-12 Scoring Table for English Language Arts Grade 6....110

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-13 Scoring Table for English Language Arts Grade 7 111
Table 6-14 Scoring Table for English Language Arts Grade 8 112
Table 6-15 Scoring Table for Mathematics Grade 3....113
Table 6-16 Scoring Table for Mathematics Grade 4....114
Table 6-17 Scoring Table for Mathematics Grade 5....115
Table 6-18 Scoring Table for Mathematics Grade 6....116
Table 6-19 Scoring Table for Mathematics Grade 7....117
Table 6-20 Scoring Table for Mathematics Grade 8....118
Table 6-21 Scoring Table for Science Grade 4....119
Table 6-22 Scoring Table for Science Grade 8....120
Table 6-23 Scoring Table for Social Studies Grade 4 121
Table 6-24 Scoring Table for Social Studies Grade 8 122
Table 6-25 Scoring Table for Social Studies Grade 10 123
Table 6-26 The Number and Percentage of Students at LOSS and HOSS 124

PART 7

Table 7-1 Policy Performance Level Descriptors for the Wisconsin Forward Exam 150
Table 7-2 Wisconsin Forward Exam Cut Scores....150

PART 8

Table 8-A Summary of Flagged Operational Items on the Wisconsin Forward Exam 170
Table 8-B English Language Arts Items Flagged for Classical Item Analysis Statistics 171
Table 8-C Mathematics Items Flagged for Classical Item Analysis Statistics 172
Table 8-D Science Items Flagged for Classical Item Analysis Statistics 173
Table 8-E Percentage of Students Attempting Last Operational Item in Test 173
Table 8-1 Item Analysis, Grade 3 English Language Arts 174
Table 8-2 Item Analysis, Grade 4 English Language Arts 175
Table 8-3 Item Analysis, Grade 5 English Language Arts 176
Table 8-4 Item Analysis, Grade 6 English Language Arts 177
Table 8-5 Item Analysis, Grade 7 English Language Arts 178
Table 8-6 Item Analysis, Grade 8 English Language Arts 179
Table 8-7 Item Analysis, Grade 3 Mathematics 180
Table 8-8 Item Analysis, Grade 4 Mathematics 182
Table 8-9 Item Analysis, Grade 5 Mathematics 184
Table 8-10 Item Analysis, Grade 6 Mathematics 186
Table 8-11 Item Analysis, Grade 7 Mathematics 188
Table 8-12 Item Analysis, Grade 8 Mathematics 190
Table 8-13 Item Analysis, Grade 4 Science....192
Table 8-14 Item Analysis, Grade 8 Science....194
Table 8-15 Item Analysis, Grade 4 Social Studies 196
Table 8-16 Item Analysis, Grade 8 Social Studies 197
Table 8-17 Item Analysis, Grade 10 Social Studies 198
Table 8-18 Raw Score Descriptive Statistics for Total Population....200
Table 8-19 Raw Score Descriptive Statistics by Gender 201
Table 8-20 Raw Score Descriptive Statistics for English Language Arts by Race/Ethnicity 202
Table 8-21 Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity 203
Table 8-22 Raw Score Descriptive Statistics for Science by Race/Ethnicity 204
Table 8-23 Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity 204
Table 8-24 Raw Score Descriptive Statistics by Socioeconomic Status....205
Table 8-25 Raw Score Descriptive Statistics by Disability 206
Table 8-26 Raw Score Descriptive Statistics by English Language Proficiency....207
Table 8-27 Raw Score Descriptive Statistics by Accommodation Use 208
Table 8-28 Scale Score Descriptive Statistics for Total Population 209
Table 8-29 Scale Score Descriptive Statistics by Gender....210

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-30 Scale Score Descriptive Statistics for English Language Arts by Race/Ethnicity....211
Table 8-31 Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity....212
Table 8-32 Scale Score Descriptive Statistics for Science by Race/Ethnicity....213
Table 8-33 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity....213
Table 8-34 Scale Score Descriptive Statistics by Socioeconomic Status....214
Table 8-35 Scale Score Descriptive Statistics by Disability....215
Table 8-36 Scale Score Descriptive Statistics by English Language Proficiency....216
Table 8-37 Scale Score Descriptive Statistics by Accommodation Use....217
Table 8-38 Performance Level Cut Scores for All Content Areas....218
Table 8-39 Cut Scores and Associated Impact Data, English Language Arts....218
Table 8-40 Cut Scores and Associated Impact Data, Mathematics....219
Table 8-41 Cut Scores and Associated Impact Data, Science....219
Table 8-42 Cut Scores and Associated Impact Data, Social Studies....220
Table 8-43 Percentage of Students in Each Performance Level by Subgroup, English Language Arts....221
Table 8-44 Percentage of Students in Each Performance Level by Subgroup, Mathematics....223
Table 8-45 Percentage of Students in Each Performance Level by Subgroup, Science....225
Table 8-46 Percentage of Students in Each Performance Level by Subgroup, Social Studies....226
Table 8-47a Summary Statistics for Content Standards Raw and SPI Scores, English Language Arts....227
Table 8-47b Summary Statistics for Domain Raw and SPI Scores, English Language Arts....230
Table 8-48 Summary Statistics for Content Standards Raw and SPI Scores, Mathematics....231
Table 8-49 Summary Statistics for Content Standards Raw and SPI Scores, Science....233
Table 8-50 Summary Statistics for Content Standards Raw and SPI Scores, Social Studies....234
Table 8-51 SPI Cut Scores, English Language Arts....235
Table 8-52 SPI Cut Scores, Mathematics....237
Table 8-53 SPI Cut Scores, Science....239
Table 8-54 SPI Cut Scores, Social Studies....240
Table 8-55 Longitudinal Comparison of State-Level Scale Score Means: ELA....241
Table 8-56 Longitudinal Comparison of State-Level Scale Score Means: Mathematics....242
Table 8-57 Baseline Year State-Level Scale Score Means: Science....243
Table 8-58 Longitudinal Comparison of State-Level Scale Score Means: Social Studies....243
Table 8-59 Longitudinal Comparison of State-Level Impact Data: ELA....244
Table 8-60 Longitudinal Comparison of State-Level Impact Data: Mathematics....245
Table 8-61 Longitudinal Comparison of State-Level Impact Data: Science....246
Table 8-62 Longitudinal Comparison of State-Level Impact Data: Social Studies....246

PART 9

Table 9-A Example Contingency Table with Three Cut Scores....253
Table 9-B Example Classification Table for One Cut Point....254
Table 9-C. Data Structure 1: Enumeration by Response....257
Table 9-D. Data Structure 2: Cross-Tabulation of Score 1 and Score 2....257
Table 9-1 Reliability for Total Group and Subgroups Using Cronbach's Alpha....260
Table 9-2 Standard Error of Measurement for Total Group and Subgroups....261
Table 9-3 Cronbach's Alpha Reliability Coefficients for Content Standard and Domain....262
Table 9-4 Standard Error of Measurement per Content Standard and Domain....263
Table 9-5 Classification Consistency and Classification Accuracy for English Language Arts Grade 3....264
Table 9-6 Classification Consistency and Classification Accuracy for English Language Arts Grade 4....264
Table 9-7 Classification Consistency and Classification Accuracy for English Language Arts Grade 5....265
Table 9-8 Classification Consistency and Classification Accuracy for English Language Arts Grade 6....265
Table 9-9 Classification Consistency and Classification Accuracy for English Language Arts Grade 7....266
Table 9-10 Classification Consistency and Classification Accuracy for English Language Arts Grade 8....266
Table 9-11 Classification Consistency and Classification Accuracy for Mathematics Grade 3....267
Table 9-12 Classification Consistency and Classification Accuracy for Mathematics Grade 4....267
Table 9-13 Classification Consistency and Classification Accuracy for Mathematics Grade 5....268
Table 9-14 Classification Consistency and Classification Accuracy for Mathematics Grade 6....268
Table 9-15 Classification Consistency and Classification Accuracy for Mathematics Grade 7....269

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 9-16 Classification Consistency and Classification Accuracy for Mathematics Grade 8 269
Table 9-17 Classification Consistency and Classification Accuracy for Science Grade 4 270
Table 9-18 Classification Consistency and Classification Accuracy for Science Grade 8 270
Table 9-19 Classification Consistency and Classification Accuracy for Social Studies Grade 4 271
Table 9-20 Classification Consistency and Classification Accuracy for Social Studies Grade 8 271
Table 9-21 Classification Consistency and Classification Accuracy for Social Studies Grade 10 272
Table 9-22 Inter-Rater Reliability, English Language Arts 273

PART 10

Table 10-1 Items Flagged for DIF by Gender, Focal Group: Female 289
Table 10-2 Items Flagged for DIF by Race/Ethnicity, Focal Group: African-American 290
Table 10-3 Items Flagged for DIF by Race/Ethnicity, Focal Group: Hispanic 291
Table 10-4 Items Flagged for DIF by Race/Ethnicity, Focal Group: Asian 291
Table 10-5 Items Flagged for DIF by Race/Ethnicity, Focal Group: American Indian 292
Table 10-6 Items Flagged for DIF by English Language Proficiency, Focal Group: Students Not English Language Proficient 292
Table 10-7 Items Flagged for DIF by Socioeconomic Status, Focal Group: Socioeconomically Disadvantaged Students 292
Table 10-8 Items Flagged for DIF by Disability Status, Focal Group: Students with One or More Disabilities 293
Table 10-9 Items Flagged for DIF by Accommodation Use, Focal Group: Students Using Testing Accommodations 293
Table 10-10 Correlations among English Language Arts Test Domains 294
Table 10-11 Correlations among English Language Arts Standards 295
Table 10-12 Correlations among Mathematics Standards 296
Table 10-13 Correlations among Science Standards 297
Table 10-14 Correlations among Social Studies Standards 297
Table 10-15 Principal Components Analysis 298
Table 10-16 Correlations between Content Area Scale Scores 298
Table 10-17 Correlations between Content Area Scale Scores by Gender 299
Table 10-18 Correlations between Content Area Scale Scores by Ethnicity/Race 300
Table 10-19 Correlations between Content Area Scale Scores by English Proficiency Status 301
Table 10-20 Correlations between Content Area Scale Scores by SES Status 302
Table 10-21 Correlations between Content Area Scale Scores by Disability Status 303
Table 10-22 Partial Correlations between Content Area Scale Scores 303
Table 10-23 Comparison of Most Recent Wisconsin NAEP and Spring 2019 Wisconsin Forward Exam Impact Data 304

Copyright © 2019 by the Wisconsin Department of Public Instruction

TABLE OF FIGURES

PART 6

Figure 6-A Examples of Likelihood Functions, or the Probability of Each Ability Level Estimate (or Scale Score).92
Figure 6-1 Anchor Set TCCs: ELA Grade 3....125
Figure 6-2 Anchor Set TCCs: ELA Grade 4....125
Figure 6-3 Anchor Set TCCs: ELA Grade 5....126
Figure 6-4 Anchor Set TCCs: ELA Grade 6....126
Figure 6-5 Anchor Set TCCs: ELA Grade 7....127
Figure 6-6 Anchor Set TCCs: ELA Grade 8....127
Figure 6-7 Anchor Set TCCs: Mathematics Grade 3....128
Figure 6-8 Anchor Set TCCs: Mathematics Grade 4....128
Figure 6-9 Anchor Set TCCs: Mathematics Grade 5....129
Figure 6-10 Anchor Set TCCs: Mathematics Grade 6....129
Figure 6-11 Anchor Set TCCs: Mathematics Grade 7....130
Figure 6-12 Anchor Set TCCs: Mathematics Grade 8....130
Figure 6-13 Anchor Set TCCs: Social Studies Grade 4....131
Figure 6-14 Anchor Set TCCs: Social Studies Grade 8....131
Figure 6-15 Anchor Set TCCs: Social Studies Grade 10....132
Figure 6-16 Item Characteristic Curves for the Flagged ELA Grade 5 Anchor....133
Figure 6-17 Item Characteristic Curves for the Flagged ELA Grade 8 Anchor....133
Figure 6-18 English Language Arts Test Characteristic Curves....134
Figure 6-19 English Language Arts Standard Error Curves....135
Figure 6-20 English Language Arts Growth at Quartiles....136
Figure 6-21 Mathematics Test Characteristic Curves....137
Figure 6-22 Mathematics Standard Error Curves....138
Figure 6-23 Mathematics Growth at Quartiles....139
Figure 6-24 Science Test Characteristic Curves....140
Figure 6-25 Science Standard Error Curves....141
Figure 6-26 Science Growth at Quartiles....142
Figure 6-27 Social Studies Test Characteristic Curves....143
Figure 6-28 Social Studies Standard Error Curves....144
Figure 6-29 Social Studies Growth at Quartiles....145

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 1: Overview

The *Wisconsin Forward Exam Spring 2019 Technical Report* documents the processes and procedures applied in the Spring 2019 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 2019 Wisconsin Forward Exam adhered to the appropriate standards and practices of educational assessment. Ultimately, this report provides evidence that valid inferences about Wisconsin student performance can be derived from this assessment.

1.1 Historical Background

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

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

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

It was within this policy context that the WKCE was constructed, as a criterion-referenced test, for the Fall 2005 administration, replacing the previously existing norm-referenced WKCE in Reading and Mathematics. The criterion-referenced WKCE was designed specifically for Wisconsin students to measure their performance on the WMAS. These assessments were designed to evaluate students' knowledge and to measure achievement in the

Copyright © 2019 by the Wisconsin Department of Public Instruction

basic skills taught in schools at grades 3–8 and 10. The Fall 2013 WKCE was the ninth administration of these assessments and the last administration of Reading, ELA, and Mathematics. The assessments in Science and Social Studies under the existing WKCE model continued to be administered until Fall 2014.

A major change in the Wisconsin assessments occurred for the 2014–15 test administration. First, the ELA and Mathematics assessments were moved from the Fall testing window to the Spring testing window. Second, the new ELA and Mathematics tests for grades 3–8 developed for the Spring 2015 administration consisted of new Smarter Balanced Assessment Consortium (SBAC) items aligned to the Common Core State Standards (CCSS). Thus, the 2014–15 ELA and Mathematics assessments were not comparable content- and construct-wise to the assessments administered in prior years. Third, while the prior years' assessments included CTB's *TerraNova* items that yielded norm-referenced scores, the 2014–15 assessments did not include such items. Fourth, the regular versions of the 2014–15 assessments were administered as fixed forms in the online mode, in contrast to the previous assessments, which were all administered in the paper-and-pencil mode. Fifth, 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) can be found at https://dpi.wi.gov/assessment/historical/smarter.

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

Science (grades 4 and 8) and Social Studies (grades 4, 8, and 10) assessments have been on a different trajectory, 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 the Science and Social Studies tests were mainly drawn from the pool of previously administered items, but new items were also included. Several of the previously administered items were edited to improve item quality and reflect test content changes over time. Despite the fact that many Science and Social Studies items in the Spring 2016 administration came from the previous item pool, statistically linking the Spring 2016 forms to the previous forms was not recommended due to the change of the testing window and the numerous changes to the items themselves. Instead, similar to what was done for the ELA and Mathematics assessments, new scales were developed for the Science and

Copyright © 2019 by the Wisconsin Department of Public Instruction

Social Studies tests under the new Wisconsin Forward Exam program. Following the new scale development, the new performance level cut scores were set for Science and Social Studies in Summer 2016.

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

Spring 2019 was the fourth administration year for the Wisconsin Forward Exam in ELA, Mathematics, and Social Studies, using the test blueprint and test design developed for the Spring 2016 test administration. The ELA, Mathematics, and Social Studies tests were developed based on the input of Wisconsin educators and with adherence to Wisconsin's standards and, with a few exceptions, consisted of items administered to Wisconsin students in Spring 2017 and Spring 2018 as part of the operational test or a field test. Previously administered operational test items served as linking items between the Spring 2018 and Spring 2019 administrations, allowing the Spring 2019 ELA, Mathematics, and Social Studies assessments to be placed on the Wisconsin Forward Exam scales using statistical equating procedures. Test equating, in turn, allows for direct comparison of student scores within a content area and for evaluation of the year-to-year student performance change.

A change to the Science (grades 4 and 8) test blueprint and design was made for the Spring 2019 operational test administration. New Science tests, aligned to the new Wisconsin Standards for Science (WSS) and the Next Generation Science Standards (NGSS) were developed and administered to Wisconsin students for the first time in Spring 2019. The new assessments focus on content understanding linked to work with science and engineering practices and crosscutting concepts as detailed in the National Research Council Framework for K–12 Science Education (https://www.nap.edu/read/13165/chapter/1). The items contained in the Science tests were drawn from the pool of items aligning to the new WSS and NGSS, field-tested in Spring 2017 and 2018. No previously administered operational test items were included in the Spring 2019 Science tests. Due to the change of standards, the statistical linking of the Spring 2019 forms to the previous forms was not performed. Instead, new scales were developed for the new Science tests, and new performance level cut scores were set after the Spring 2019 test administration.

This Technical Report documents all aspects of the 2018–19 testing cycle. The structure of this report mirrors the testing cycle. A brief content summary of the report is provided later in this part of the report.

1.2 Uses of Test Scores

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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, and the public) about the status of the progress toward meeting academic achievement standards of the state
- Meeting the requirements of the state's accountability program

This Technical Report refers to the use of the test-level scores (scale scores and performance levels) and standard-level (objective) scores (Standard Performance Index [SPI] scores and performance levels).

1.2.1 Test-Level Scores

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Scale Scores

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

Levels of Performance

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

Use of Test-Level Scores

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

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*).

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 students' strengths and needs within a particular content area and grade level. Caution should be exercised when comparing standard-level subscores across years because different items will contribute to these subscores and these items may vary in difficulty between test forms or test administrations.

1.3 Technical Report Structure

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

Part 2: Test Blueprint and Item Development

Part 2 of this report describes the test blueprint, the item development 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 were the basis of the test blueprints and item specifications for their respective content areas. Wisconsin educators were involved in reviewing the items in all content areas to ensure the appropriateness of the test to the standards. The first item review for grades 3–8 in ELA and Mathematics, and grades 4, 8, and 10 in Social Studies occurred in December 2015. The first item review for new assessments in Science grades 4 and 8 occurred in August 2017. Each year after that, new items were reviewed and added to the Wisconsin pool of items for future field-testing. The item reviews served to establish the accessibility of the items and reading passages. Simultaneously, DRC created the test specifications documents that were later approved by DPI and will continue to serve as a foundation for item and test development. Additional item reviews, supported by the item data, occurred after each field test administration and were conducted by DPI content

Copyright © 2019 by the Wisconsin Department of Public Instruction

experts. The purpose of these reviews was to refine the pool of items from which the subsequent operational test forms were selected.

Part 3: Test Form Development

Part 3 presents the Wisconsin Forward Exam design and discusses key development tasks related to creating the Spring 2019 Wisconsin Forward Exam forms. Item selection was based on the approved test blueprints. DRC's CCR item bank contained a sufficient number of items to fulfill the test design needs for the ELA, Mathematics, and new Science assessments. Social Studies test forms consisted of Wisconsin-owned items. Part 3 also discusses the process of selecting operational test items and the process of obtaining DPI approvals. As detailed in Part 3, in addition to the operational test items, there were numerous unique field test items on each form. Selection of the Spring 2019 test forms was done using the approved test blueprints, test designs, and psychometric specifications as guides.

Part 4: Test Administration

Part 4 describes test administration and accommodations. The Wisconsin Forward Exam is a component of the Wisconsin Student Assessment System (WSAS), which is considered to be a comprehensive statewide program of assessments. In the 2015–16 school year, this assessment replaced the Wisconsin Badger Exam (SBAC) in the areas of ELA and Mathematics in grades 3–8 and the WKCE in the areas of Science (grades 4 and 8) and Social Studies (grades 4, 8, and 10). In the 2018–19 school year, the Wisconsin Forward Exam was administered to Wisconsin students for the fourth time.

The Spring 2019 Wisconsin Forward Exam was an online assessment with a single print-on-demand form at each grade level. Student responses to the print-on-demand form were transcribed by a proctor into the online assessment system. Other variations of the forms included stacked Spanish translation forms, video sign language, and closed captioning. These were provided in an online format at each grade level.

Test administration was conducted during a seven-week window from March 18 to May 3, 2019. All testing was conducted online, administered via DRC's INSIGHT platform.

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

Part 5: Scoring

Part 5 documents the scoring process for different item types: scanning of multiple-choice (MC) items and multi-select (MS) items; autoscoring of technology-enhanced (TE) items, short-answer (SA) items, and evidence-based selected response (EBSR) items; and artificial intelligence (AI) scoring and handscoring of text-dependent analysis (TDA) items. The description of the handscoring process includes the development and review of the scoring rubrics, anchor (sample) paper selection, training of scoring personnel, ongoing quality

Copyright © 2019 by the Wisconsin Department of Public Instruction

assurance, and a systematic review of the resulting score distributions supporting reliable and valid reported test scores. The scoring rubric used in the handscoring of the TDA writing items is presented in detail.

Part 6: Calibration, Equating, and Deriving Scale Scores

The Spring 2019 administration year is the fourth administration year for the Wisconsin Forward Exam in all grades and content areas. Part 6 discusses characteristics of the sample of student data used for data analysis and describes the calibration, equating, and scoring methods implemented for the Wisconsin Forward Exam after the Spring 2019 test administration. The data were calibrated using two different item response theory (IRT) models, one for constructed-response items and one for MC 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 includes model-to-data fit and the standard error of measurement (SEM). The equating of Spring 2019 ELA, Mathematics, and Social Studies test forms to the scales established after the Spring 2016 administration was performed using the Stocking and Lord procedure. New scales were developed for Science grades 4 and 8. Item-pattern scoring was applied to the Spring 2019 Wisconsin Forward Exam. As discussed in Part 6, item-pattern scoring is generally recommended over number-correct scoring because it produces more accurate scores for individual students. Part 6 also explains how a student's scale score is derived from the raw score using item-pattern scoring.

Part 7: Standard Setting

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

Part 8: Test Results

Part 8 summarizes the results of item analysis and the test reliability reported using Cronbach's alpha and SEM. Summary descriptive statistics for all scores (i.e., raw scores, scale scores, SPI scores, and performance levels) are reported for the total population and for subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, accommodation use, and English language proficiency. In addition, the longitudinal test results are presented in Part 8.

Part 9: Reliability

Part 9 elaborates on the reliability of the test based on results presented in previous parts of the report. SEM 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,

Copyright © 2019 by the Wisconsin Department of Public Instruction

accommodation use, 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, correlations among content standards, and a relationship between the Wisconsin Forward Exam scores and external variables 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 2019 Wisconsin Forward Exam administration are presented in the body of the report. However, some issues 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 2019 Wisconsin Forward Exam administration cover different phases of the testing cycle: item development; scoring; and psychometric, or measurement-based, research and evaluation.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 2: Test Blueprint 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 blueprint and item development process for the Spring 2019 administration.

This part of the Technical Report is particularly relevant to American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (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 section of the report. AERA, APA, & NCME (2014) Standard 4.0 states the following:

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

English Language Arts (ELA), Mathematics, and Science test items included in the Spring 2019 Wisconsin Forward Exam were selected from DRC's College- and Career-Ready (CCR) item bank. DRC's CCR item bank contains nationally field-tested CCR items that support the next generation of standards and assessments. It is aligned to the College and Career Readiness standards in ELA and Mathematics grades 3–8. Science items are aligned to Wisconsin's Standards for Science and enhanced by the Next Generation Science Standards (NGSS) based on the National Research Council's Framework for K–12 Science Education. The item bank is designed to support states like Wisconsin that have adopted, or are preparing to adopt, more rigorous content standards, curricula, and assessments that better prepare students for college and careers.

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Therefore, DRC employs the Principles of Universal Design throughout all stages of both the item development process and the test development process.

All DRC's ELA, Mathematics, and Science items that appear on the Wisconsin Forward Exam were reviewed for content and for fairness not only by DRC's content experts but also by a panel of external experts and more recently by Wisconsin educators. The external reviewers have a broad range of experience in the educational field. All the reviewers have bachelor's-level, master's-level, or doctoral-level degrees and teaching experience in their specific area of expertise. Table 2-1 provides a high-level sequence of the activities that occurred in the development of the DRC CCR item bank.

Wisconsin-owned Social Studies items were developed by DRC content specialists. These items are aligned to Wisconsin's Model Academic Standards for Social Studies. Social Studies items underwent reviews by DRC content experts as well as DRC bias and sensitivity experts. All Social Studies items were also reviewed and approved by committees of Wisconsin educators.

Various item types were developed and included in the Wisconsin Forward Exam in order to best assess students' understandings of the standards. Descriptions of each item type used in the Wisconsin Forward Exam are provided in Table 2-2.

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. (p. 63)
>
> **Standard 3.2** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (p. 64)

As stated earlier, Wisconsin licensed ELA, Mathematics, and Science items from DRC's CCR item bank for the Spring 2019 test administration. Due to the state-specific nature of the Social Studies standards, DPI owns the items for that content area. Details regarding the development of the items in the CCR bank created prior to their field-testing on the Forward Exam are provided in the *Wisconsin Forward Exam Spring 2016 Technical Report*, available on the DPI website at https://dpi.wi.gov.

2.1 Test Blueprints

The test blueprints specify the number of items for each reporting category and subskill as well as the allowable DOK levels for the respective reporting categories. The process used for

Copyright © 2019 by the Wisconsin Department of Public Instruction

developing the blueprints for Wisconsin Forward Exam 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 for Spring 2018 Field Test

The test items typically begin their life cycle two years prior to their operational administration. New ELA, Mathematics, Science, and Social Studies passages and items were first reviewed and approved for placement on the Wisconsin Forward Exam by both DPI and Wisconsin educators. For these reviews, educators from across the state convened in Madison, Wisconsin, to review items in an online format so that items could be evaluated in the same testing engine and style in which items are presented to students during the actual administration. ELA and Mathematics item reviews were held from July 31 to August 2, 2017; Social Studies and Science item reviews were held on August 3 and 4, 2017. An example of the training PowerPoint presentation used at the reviews can be found in Appendix A of this report.

Table 2-7 shows the number of items taken to the item review by grade and content area. Using the approved test blueprints as a guide, DRC content specialists determined the focus of the items that would be taken to item review. Using an electronic tally sheet, Wisconsin educators made the determinations of standard alignment, DOK levels, and key(s). They noted any bias and sensitivity concerns and had the opportunity to determine whether items were accepted as is or accepted with revisions. They also had the opportunity to register a "dissenting view" in which the committee preferred the item not be selected to appear on the Wisconsin Forward Exam in a field test position.

Items and passages that were approved by the Wisconsin educators were then included in the next field test administration in Spring 2017. The purpose of the Spring 2017 field test was to expand the pool of items eligible for inclusion in the subsequent operational test forms, such as the Spring 2018 Forward Exam.

2.3 Field-Testing

Items approved for the field test administration during the Summer 2017 item review were field-tested in Spring 2018 during the operational test administration. Field test items were fully embedded in the operational forms, and students were not able to distinguish between the operational and field test items. The field test items were embedded in several test forms administered in each grade and content area. Each test form contained the same operational test items and unique field test items. The test forms were spiraled at the student level within a grade and a content area. A total of 377 new items were field-tested for ELA. A total of 189 items were field-tested for Mathematics. A total of 240 items were field-tested for Science, and a total of 104 items were field-tested for Social Studies in the Spring 2018 test administration.

Copyright © 2019 by the Wisconsin Department of Public Instruction

2.4 Statistical Analysis of Spring 2018 Field Test Data

Following the field test data acquisition, the field test data analyses were conducted. The analyses included classical item analysis, differential functioning item (DIF) analysis, and item response theory (IRT). The classical item analysis included computation and evaluation of the following statistics: item *p*-values (difficulty), item-total test correlation, percentage of students selecting incorrect responses, point-biserial correlation for incorrect responses for the multiple-choice (MC) items, score point distribution for items worth more than 1 point, and omit rates for all items. More details on classical item analysis methodology are provided in Part 8 of this report.

DIF was conducted for all field test items to examine potential item bias and to determine whether item performance differences between identifiable subgroups were due to factors other than student ability, making the items unfairly difficult for a particular subgroup in the student population. DIF analyses were conducted based on gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency (ELP) groups. More details on the DIF methodology are provided in Part 10 of this report.

As the last step of the field test data analyses, the field test items were calibrated and equated to operational test scales using the IRT methodology (explained in detail in Part 6 of this report). Note that ELA, Mathematics, and Social Studies field test items were equated to their respective operational test scales using a common student design. All operational test items contained in the Spring 2018 operational test forms served as anchor items to place the field test items on the operational test scores using the Stocking and Lord procedure. Science field test items were calibrated after the Spring 2018 test administration and were back-equated to the new operational Science test scales using a common item design, after the new scales were established in Spring 2019.

The field test item statistics are used as a means of detecting items that deserve closer scrutiny, rather than being a mechanism for automatic retention or rejection. Toward this end, a set of criteria was used as a screening tool to identify items that needed a closer review. For an item to be flagged for an additional review, the criteria included any of the following:

- *p*-value <0.20 or >0.90
- item-total test correlation (point biserial for MC items) <0.15
- positive point biserial on a distractor for an MC item
- omit rate >5%
- large DIF

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

2.5 Review of Items with Data

In the preceding section, it was stated that test development content area specialists used certain statistics from item and DIF analyses of the 2018 field test to identify items for further review. Specific flagging criteria for this purpose were specified in the previous section. Items without statistical flags were regarded as statistically acceptable and were not included in the data review. Likewise, items of extremely poor statistical quality were regarded as unacceptable and needed no further review. Such items were excluded from the Wisconsin item pool prior to the data review with DPI. The remaining flagged items were regarded by DRC content area test development specialists and DRC psychometric specialists as needing further review. The intent was to capture all items that needed an additional review based on their statistical properties; thus, the criteria employed for item flagging tended to overidentify rather than under-identify potential item issues.

The review of the items with data was conducted by DPI staff and DRC content specialists, who were broken out into content area and/or grade-level groups. The data review took place in Madison, Wisconsin, in August 2018. In these sessions, reviewers were first trained by a representative from DRC's staff with regard to the statistical indices used in item evaluation. This was followed by a discussion with examples concerning reasons that an item might be retained regardless of the statistics. The review process involved a brief exploration of possible reasons for the statistical profile of an item (e.g., possible bias, grade appropriateness, instructional issues) and a decision regarding acceptance. DRC content area test development specialists facilitated the review of the items. Each group reviewed the pool of field test items and made recommendations on each item and/or scenario/passage. The training presentation used at the data review meeting may be found in Appendix B. A summary of the data review results, including the number of items that were field-tested, the number and percentage of items with statistical flags, and the number and percentage of items rejected by DPI during the data review, is presented in Table 2-8. Items accepted for subsequent use in the Wisconsin Forward Exam were included in the pool of items for Spring 2019 operational test form selection.

2.6 Summary

In summary, the items included in the Spring 2019 Wisconsin Forward Exam were reviewed by DRC, DPI, and Wisconsin educators for issues regarding accessibility, bias, sensitivity, and content. During the reviews, experts identified (1) issues that could negatively affect a student's ability to access stimuli and items, (2) content in stimuli and items that could unfairly affect a student's response because of his or her background, (3) developmental appropriateness, and (4) alignment of stimuli and items to the content specifications. Item content was checked for the accuracy of the content, answer keys, and scoring rules. Following Spring 2018 field-testing, items flagged for accessibility, bias and sensitivity, and/or other content concerns were further reviewed by DRC and DPI to determine if these flagged items should be removed from the Wisconsin item pool prior to the form construction of the Wisconsin Forward Exam. In addition, item statistics from the Spring 2018 operational and field test administration were used to refine the item pool used in the selection of Spring 2019 Wisconsin Forward Exam forms.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 2-2 Item Type Descriptions for Items on the Wisconsin Forward Exam

Item Type	Name	Description
EBSR	Evidence-Based Selected Response	Each evidence-based selected response item has two parts, and each two-part item is designed to elicit an evidence-based response from a student who has read a literature text passage, an informational text passage, or a writing concept. In part one, which is similar to a multiple-choice item, the student analyzes a passage or writing concept and chooses the best answer from four response options. In part two, the student uses evidence from the passage or writing concept to select one or more answers based on the response to part one. Each of these items is worth one point.
MC	Multiple Choice	Each multiple-choice item has four response options, only one of which is correct. Multiple-choice items are used to assess a variety of skill levels, from short-term recall of information to inference and problem solving. Each of these items is worth one point.
MS	Multiple Select	Each multiple-select item requires a student to evaluate information presented and respond by choosing two or more correct responses. Multiple-select items can be used to assess multiple skills and concepts in a given content area.
SA	Short Answer	Each short-answer item requires a student to enter a short numeric or algebraic response. These items are designed to assess a student's ability to formulate a solution to a pure or applied math problem without the assistance of response options. The short-answer items are scored on a 0–1-point scale using item-specific autoscoring rules.
TE	Technology Enhanced	Each technology-enhanced item is designed to elicit evidence of a broad range of student understanding. A student interacts with the enhanced features of these computer-delivered, auto-scorable 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 item 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. A weight of 2 is applied to the item scores in computation of the student total test raw scores and scale scores. That is, the TDA prompts contribute up to 8 raw score points toward the student total test raw score. This item type is supported by all Wisconsin ELA standards across all grades for both Reading Literature and Reading Informational Texts and by 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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

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

Table 2-7 Items Reviewed at Summer 2017 Item Review

Grade	Number of Items			
	English Language Arts	Mathematics	Science	Social Studies
3	87	37		
4	87	37	153	49
5	87	37		
6	87	37		
7	88	37		
8	87	37	155	48
10				64
TOTAL	**523**	**222**	**308**	**161**

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 2-8 Items Reviewed at Summer 2018 Data Review

Content Area	Grade	No. of Items in 2018 Field Test	Flagged Items in 2018 Field Test Examined at 2018 Data Review				Flagged Items in 2018 Field Test Rejected at 2018 Data Review	
			Number of Flagged Items	Items Flagged for DIF only	Total	Total (% of FT)	No. of Field Test Items	% of FT
English Language Arts	**3**	63	7	1	8	13%	5	8%
	4	63	12	4	16	25%	3	5%
	5	63	10	1	11	17%	4	6%
	6	63	11	2	13	21%	3	5%
	7	63	10	3	13	21%	3	5%
	8	62	9	2	11	18%	3	5%
Mathematics	**3**	32	5	1	6	19%	0	0%
	4	32	7	1	8	25%	5	16%
	5	32	5	0	5	16%	2	6%
	6	29	8	2	10	34%	3	10%
	7	32	13	1	14	44%	8	25%
	8	32	12	0	12	38%	3	9%
Social Studies	**4**	32	4	1	5	16%	2	6%
	8	32	1	3	4	13%	0	0%
	10	40	4	2	6	15%	2	5%
Science	**4**	120	56	1	57	48%	41	34%
	8	120	43	1	44	37%	30	25%

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 3: Test Form Development

Part 3 of this report focuses on key development tasks related to creating the Spring 2018 Wisconsin Forward Exam operational forms. The test blueprint 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:

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

3.1 Design of the Wisconsin Forward Exam

The following sections provide detailed information about the test design of the content areas assessed on the Spring 2019 Wisconsin Forward Exam assessments.

3.1.1 English Language Arts

Table 3-1 shows the ELA test design, including the number of passages, items, and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the eight field test forms at each grade level. There were five or six additional forms per grade that contained items from MetaMetrics. These multiple-choice items were inserted into the same positions on the forms as the field test items and served the purpose of linking the Wisconsin ELA scale to the MetaMetrics Lexile scale. The item data for the Lexile items acquired during the Spring 2019 test administration were delivered to MetaMetrics for the linking analyses. The MetaMetrics item data were not processed by DRC and are not discussed in this report. Instead, a separate report concerning Lexile data analysis and linking of Wisconsin ELA assessments to the Lexile scale was developed by MetaMetrics and delivered to DPI. Student performance on these items did not count toward their ELA scores. Table 3-1 also identifies the various item types that appeared on the ELA forms, including the points for item scoring. Detailed descriptions of the item types are provided in Table 2-2 of this report.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

3.1.2 Mathematics

Table 3-2 shows the Mathematics test design, including the number of items and points at each grade level that were used in the core and embedded field test positions. There was one common set of core operational items in each of the four field test forms at each grade level.

The Mathematics section of the exam was divided into two testing sessions, with students able to take the sessions in either order. In grades 3–5, no calculator was allowed for any of the Mathematics items. In grades 6–8, no calculator was allowed for the first session, and the second session allowed students to use an embedded calculator. There were an additional four forms per grade that contained items from MetaMetrics. These multiple-choice items were inserted into the same positions on the forms as the field test items and served the purpose of linking Wisconsin Mathematics scale to the MetaMetrics Quantile scale. The item data for the Quantile items acquired during the Spring 2019 test administration were delivered to MetaMetrics for the linking analyses. The MetaMetrics item data were not processed by DRC and are not discussed in this report. Instead, a separate report concerning Quantile data analysis and linking of Wisconsin Mathematics assessments to the Quantile scale was developed by MetaMetrics and delivered to DPI. Student performance on these items did not count toward their Mathematics scores. Recommended testing times for both sessions were included in the test design document as well as in the test administration manual.

3.1.3 Science

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

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

3.1.4 Social Studies

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

3.2 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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

The Wisconsin Forward Exam test development process complied with the following AERA, APA, & NCME (2014) standards:

> **Standard 4.1** Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s). (p. 85)
>
> **Standard 4.7** The procedures used to develop, review, and try out items and to select items from the item pool should be documented. (p. 87)
>
> **Standard 4.12** Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications. (p. 89)

3.2.1 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 used an integrated team approach to test development, which included content area specialists, psychometricians, and scoring specialists working as a unit in collaboration with DPI content experts.

3.2.2 Item Selection

New operational test forms were developed for all content areas for the Spring 2019 test administration. 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, which is called IDEAS. The form, format, extent, and organization of items in their respective test sessions were 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 within DRC's integrated development team of assessment specialists, psychometric services specialists, and scoring specialists.

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

- Using the pool of vendor-owned items for ELA, Mathematics, and Science and Wisconsin-owned Social Studies 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 each item, with available item statistics, met psychometric guidelines for inclusion in the test.

Copyright © 2019 by the Wisconsin Department of Public Instruction

- DRC test development specialists verified that each item met technical quality for well-crafted items, including that each item
  - had one clearly correct answer (or answers if the item was multi-select);
  - used clear and concise wording;
  - was grammatically correct;
  - had an appropriate range of difficulty;
  - was free of any offensive, inappropriate, or biased content; and
  - met the Principles of Universal Design and maximum accessibility.

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

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

The statistical properties of the Spring 2018 test forms were used as targets for selection of the Spring 2019 ELA, Mathematics, and Social Studies test forms. The item selection for these content areas was conducted in two phases.

In the first phase, the anchor (linking) items were selected. The anchor items are used for statistical linking of the new forms to the previous test forms on already established test scales. The anchor items on the Spring 2019 test forms were selected mainly from the Spring 2018 operational item pool (with a small number of anchor items being selected from the Spring 2017 operational tests). The anchor set was selected as a “mini” version of the full operational test for each grade level and content area in regard to its length, content coverage, and psychometric properties.

The length of the anchor sets was at least one-third of the length of the total test. The items included in the anchor sets met the same blueprint specification as the full test in regard to the percentage of score points measuring each content standard. In addition, the psychometric properties of the anchor sets matched the corresponding properties of the target forms as closely as possible. Anchor selections were reviewed and approved by a DRC psychometrician.

No anchor item selection was conducted for Science assessments. Because the Spring 2019 Science assessments measured new Wisconsin Science Standards, these assessments were not linked to the previous Science scales. As such, there was no need to include items from the

Copyright © 2019 by the Wisconsin Department of Public Instruction

previous operational Science tests in the new Science assessments or match the new Science assessments to the previous ones in terms of the test statistical properties.

In the second phase of the item selection process, non-anchor operational items were selected for all content areas. With the exception of ELA TDA items, the non-anchor operational items came from the Spring 2017 and 2018 Wisconsin Forward Exam operational and field test item pool for ELA, Mathematics, and Social Studies and from the Spring 2017 and 2018 Wisconsin Forward Exam field test item pool for Science. Of the six TDA items selected for the Spring 2019 ELA test administration, two items were previously administered in Wisconsin (in grades 3 and 4 in the Spring 2016 operational test administration) and the remaining four TDA items (for grades 5 through 8) were not previously field-tested in Wisconsin.

The non-anchor operational items were selected using the item selection guidelines presented earlier in this section. Full form selections were reviewed and approved by a DRC psychometrician.

After selection of all operational items, the new field test items were added to each form in each grade and content area. In ELA and Mathematics, the MetaMetrics items were also inserted into the forms to match the field test positions of the other forms. In constructing the final forms, the DRC content area test development specialists followed the guidelines provided below:

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

The test maps in Appendices C, D, E, and F provide details on the operational items placed on the Spring 2019 Wisconsin Forward Exam per grade and content area. The test maps include the session number, item sequence, item type, item usage, item maximum score, depth-of-knowledge level, standard code, and domain name. The ELA test map is included in Appendix C, the Mathematics test map is contained in Appendix D, the Science test map is provided in Appendix E, and the Social Studies test map is given in Appendix F.

3.2.3 Item and Form Quality Reviews

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

by DPI prior to submission, for review by Wisconsin educators to ensure that all items align with Wisconsin's standards and adhere to DPI's standards for high-quality items.

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

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

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

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

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

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

3.3 DPI Approvals

DPI had the opportunity to review passages and items to be placed on the Spring 2019 Wisconsin Forward Exam during the following phases:

- prior to item content review in Summer 2017
- at item content review in Summer 2017
- during review of flagged field test data in Summer 2018
- during the Spring 2019 form 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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

3.4 Summary

In summary, the efforts and procedures used in the development of the Spring 2019 Wisconsin Forward Exam balanced the content and psychometric requirements for the form development. The content of the Spring 2019 test forms adhered to the test blueprint requirements. The psychometric properties of the new test forms were comparable to the psychometric properties of the Spring 2018 forms for ELA, Mathematics, and Social Studies. New Science test forms were built for the Spring 2019 test administration. Overall, the process implemented in the Spring 2019 operational form development was in alignment with multiple best practices of the test industry.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 3-1 English Language Arts Test Design

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Passage Sets	**Literature**	4	4	4	4	3	4
	Informational	1	2	3	2	3	2
	Listening	3	3	3	3	3	3
Number of Core (OP) Items	**Item Types: MC/TE (1 pt)**	27	28	30	24	22	28
	Item Types: MS/TE/EBSR (2 pts)	9	10	9	12	13	10
	Item Type: TDA (4 pts x 2)	1	1	1	1	1	1
	Total Core Items	37	39	40	37	36	39
Total Core Points		53	56	56	56	56	56
Embedded Field Test (FT)	**Number of Forms**	8	8	8	8	8	8
	Passages (Reading + Listening)	2	2	2	2	2	2
	FT Items per Form	10	10	8	8	8	8
	Total Items Field-Tested	80	80	64	64	64	64
Total Items (Core + FT) per Form		47	49	48	45	44	47
MetaMetrics Forms							
Embedded in FT Positions	**Number of Forms**	5	5	5	6	6	6
	Number of Items per Form	8	8	6	6	6	6
Total Estimated Testing Time (minutes)		130	130	130	130	130	130

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 3-2 Mathematics Test Design

Test Design		Grade					
		3	**4**	**5**	**6**	**7**	**8**
Number of Core (OP) Items	**Item Types: MC/SA (1 pt)**	37	41	39	40	41	40
	Item Type: TE (1 pt)	5	5	7	6	5	6
	Total Core Items	42	46	46	46	46	46
Total Core Points		42	42	46	46	46	46
Embedded Field Test (FT)	**Number of Forms**	4	4	4	4	4	4
	FT Items per Form	8	8	8	8	8	8
	Total Items Field-Tested	32	32	32	32	32	32
Total Items (Core + FT) per Form		50	50	54	54	54	54
MetaMetrics Forms							
Embedded in FT Positions	**Number of Forms**	4	4	4	4	4	4
	Number of Items per Form	8	8	8	8	8	8
Total Estimated Testing Time (minutes)		90	90	90	105	105	115

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 3-3 Science Test Design

Test Design		Grade	
		4	**8**
Number of Core (OP) Items	**Item Types: MC/MS/TE/EBSR (1 pt)**	40	40
Total Core Points		40	40
Embedded Field Test (FT)	**Number of Forms**	20	20
	Scenarios/Tasks	10	10
	FT Items per Form	5	5
	Total Items Field-Tested	100	100
Total Items (Core + FT) per Form		45	45
Total Estimated Testing Time (minutes)		120	120

Table 3-4 Social Studies Test Design

Test Design		Grade		
		4	**8**	**10**
Number of Core (OP) Items	**Item Types: MC/TE/MS (1 pt)**	38	40	50
Total Core Points		38	40	50
Embedded Field Test (FT)	**Number of Forms**	4	4	4
	FT Items per Form	4	4	5
	Total Items Field-Tested	16	16	20
Total Items (Core + FT) per Form		46	42	44
Total Estimated Testing Time (minutes)		70	70	70

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 4: Test Administration

In the Spring of 2019, Wisconsin administered assessments in ELA and Mathematics for grades 3–8. Science was administered in grades 4 and 8, and Social Studies was administered in grades 4, 8, and 10. The test administration window was March 18–May 3, 2019. 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, which has important implications for the integrity of the results and, thus, the validity of Wisconsin Forward Exam scores, is also discussed. Documentation citing the written procedures provided to test administrators and school personnel in order to standardize the administration of the test is 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 will be held accountable to a common set of high academic content standards, the Wisconsin Academic Standards. As an alternate assessment for students being instructed using alternate academic achievement standards, the Wisconsin Essential Elements, the Dynamic Learning Maps assessment measures the academic progress of students with the most significant cognitive disabilities in the subject areas of ELA and Mathematics at grades 3–11 and Science at grades 4 and 8–11. A teacher rater form is used to assess these students in Social Studies at grades 4, 8, and 10.

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

accommodations and supports intended to provide equitable access to grade-level content and assessments.

Test administrators indicated which accommodations and supports were to be available for use by each student within the student learning profile in DRC's eDIRECT system. All student accommodations and supports are managed and can be monitored 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 Accessibility Resources

Accommodations were allowed for eligible individual students participating in the Wisconsin Forward Exam. Accommodations provided to a student must be documented in a current IEP and used during routine instruction. IEP teams were directed to refer to the Wisconsin Forward Exam accommodations policy and guidance at 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 students' needs. Those supports 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 accessibility features of the assessment are either provided as digitally delivered components of the test administration system (embedded) or separate from it (non-embedded).

Embedded Universal Tools ("Online")

- Calculators
- Click to Enlarge
- Cross-off Tools
- Flag/Mark for Review
- Help/What's This?
- Highlighter
- Go to Question
- Keyboard Navigation
- Line Guide

Copyright © 2019 by the Wisconsin Department of Public Instruction

- Magnifier Tool (Zoom)
- Measuring Tools
- Pause (Breaks)
- Review Page
- Sticky Notes (Digital Notepad)
- Test Directions
- Tool Tips

Non-embedded Universal Tools ("Standard")

- Scratch Paper

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) and are part of the student's classroom instruction. They are either provided as part of the online test administration system or separate from it (i.e., embedded or non-embedded). All embedded and non-embedded designated supports must be entered into eDIRECT prior to test administration. Embedded and non-embedded supports will appear on student test tickets.

Embedded Designated Supports ("Online")

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

Non-embedded Designated Supports ("Standard")

- Amplification Device
- Word-to-Word Bilingual Dictionary
- Color Overlay
- Magnification
- Noise Buffers
- Read Aloud in English
- Read Aloud in Spanish (new in 2019)
- Scribe
- Separate Setting
- Small Group Translation
- Translation
- Translator/Interpreter

Copyright © 2019 by the Wisconsin Department of Public Instruction

4.1.3 Accommodations

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

Embedded Accommodations ("Online")

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

Non-embedded Accommodations ("Standard")

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

4.1.4 Translation

For the Spring 2019 Wisconsin Forward Exam administration, the State of Wisconsin used an embedded stacked Spanish translation for Mathematics, Science, and Social Studies items. For ELA assessments, only the test directions are available in stacked translation. The stacked Spanish translation is a designated support for students who are native Spanish speakers and are limited English proficient to demonstrate their knowledge on the Wisconsin Forward Exam. In addition to the embedded stacked translation, bilingual word lists and translation of the test directions are allowable designated supports

DPI recognizes that approximately 5 percent of the Wisconsin limited English proficient population speaks a language other than Spanish, and specific guidelines are provided for these students. Districts that serve students who speak languages other than Spanish may have used qualified translators to provide oral translation support to students. However, the use of translation support was restricted to Mathematics, Science, and Social Studies tests, given that the test constructs are not specific to the English language. DPI recommended that educators consult the list of allowable accommodations and supports (referenced above) to create the most appropriate testing situation for their students.

Copyright © 2019 by the Wisconsin Department of Public Instruction

4.1.5 Additional Accessibility Resources

Additional accessibility resources and guidance included the following:

- **Multiplication Table:** This resource is a non-embedded accommodation available for students who have it in their IEP or 504 plan for grades 4–8 Mathematics.
- **Read Aloud Guidelines:** This document outlines the qualifications, guidelines, and procedures required for a test reader. The test reader must sign the Read Aloud Agreement to Maintain Security and Confidentiality prior to test administration. Completed agreement forms should be retained by the Site Assessment Coordinator.
- **Scribing Guidelines:** This document outlines the qualifications, guidelines, and procedures required when using a scribe.
- **Interpreter Guidelines:** This document outlines the qualifications, guidelines, and procedures required when using an interpreter.

Tables 4-1 through 4-7 provide the list of accommodations or designated supports made available for the Spring 2019 Wisconsin Forward Exam along with the number and percentage of students provided these accommodations or supports. The counts are based on the accommodations and designated supports selected via the eDIRECT portal.

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 have been implemented for the Wisconsin Forward Exam with compliance to the following AERA, APA, & NCME (2014) standards:

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

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

The primary goal of test security is to protect the integrity of the assessments and ensure that scores retain their interpretability. To ensure that trends in achievement results can be calculated across years and to provide longitudinal data, a certain number of test questions must

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

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

4.3.1 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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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, after a student logs 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.

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

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

4.4 Test Administration Training

DRC provided online webinars for District Assessment Coordinators (DACs) and Test Administrators (TAs) for the Spring 2019 administration of the Wisconsin Forward Exam. The webinars were recorded by DPI and DRC. The purpose of the webinars and the ancillary materials was to keep districts and schools informed about policies and procedures related to the Wisconsin Forward Exam administration. The information covered in the webinars included standardizing the administration of the Wisconsin Forward Exam, maintaining the security of the assessments, allowing access to the assessments for special populations by providing appropriate designated supports or accommodations, and providing guidance on appropriate interpretations of the test results. These communication efforts by DPI and the ancillary information developed by DRC are in alignment with multiple best practices of the testing industry and, in particular, support the following *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

Copyright © 2019 by the Wisconsin Department of Public Instruction

process for reviewing requests for additional testing variations should also be documented. (p. 90)

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

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

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

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

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

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

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

In addition, Test Administration Manuals, made available to all test administrators, included the following topics:

- Key Dates
- TA Responsibilities
- Test Times

Copyright © 2019 by the Wisconsin Department of Public Instruction

- Test Security
- Accessibility Information
- Prior to Testing Instructions
- Test Tickets
- During Testing Information
- Test Administration Script

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

Student Preparation for Online Testing

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

Student and Administrator Tutorial Videos

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

Online Tools Training

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

Item Samplers

Item samplers were developed for use by both educators and students to gain familiarity with the types of items and their functionality. The format appears as a “guided practice test” in the online, PDF, and Braille versions of the tests.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Administration Supports before and after Testing

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

Table 4-8 displays a list of electronic 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 Administration Manual that is available online at https://dpi.wi.gov/assessment/forward/resources#manuals.

4.5 Summary

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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**
Print-on-Demand (POD)	3	0.00	2	0.00
Bilingual Dictionary			178	0.29
Magnification	91	0.15	90	0.15
Noise Buffers	877	1.44	876	1.43
Read Aloud in English	704	1.15	767	1.25
Scribe	569	0.93	525	0.86
Separate Setting	7849	12.85	7889	12.89
Alternate Response Options	3	0.00	2	0.00
Read Aloud (Reading Passages)	3	0.00		
Provided Color Choices (CC)	112	0.18	110	0.18
Contrasting Color (CTC)	70	0.11	68	0.11
Reverse Contrast (RC)	29	0.05	29	0.05
Masking (MSK)	662	1.08	661	1.08
Text-to-Speech (TTS)	11962	19.58	12388	20.24
Spanish Translation (ST)	584	0.96	834	1.36
Video Sign Language (VSL [ASL])	16	0.03	16	0.03
Color Overlay	27	0.04	27	0.04
Amplification Device	53	0.09	53	0.09
Small Group Translation	95	0.16	143	0.23
Translator/Interpreter	45	0.07	61	0.10
Read Aloud in Spanish	77	0.13	172	0.28
Closed Captioning (C CAP) ELA	40	0.07		
Listening Scripts (LS) ELA	8	0.01		
Abacus Math			32	0.05
Non-embedded Calculator Math			129	0.21
Multiplication Table Math			633	1.03

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 4	English Language Arts		Mathematics		Science		Social Studies	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**	**N Count**	**Percent**	**N Count**	**Percent**
Braille (BRL)	3	0.00	3	0.00	3	0.00	3	0.00
Print-on-Demand (POD)	1	0.00	1	0.00	1	0.00	1	0.00
Bilingual Dictionary			248	0.39	258	0.41	258	0.41
Magnification	146	0.23	146	0.23	146	0.23	146	0.23
Noise Buffers	878	1.38	878	1.38	876	1.38	875	1.38
Read Aloud in English	668	1.05	710	1.12	702	1.10	700	1.10
Scribe	546	0.86	519	0.82	514	0.81	511	0.80
Separate Setting	8487	13.36	8574	13.47	8447	13.28	8438	13.27
Alternate Response Options	13	0.02	12	0.02	14	0.02	14	0.02
Read Aloud (Reading Passages)	3	0.00						
Provided Color Choices (CC)	129	0.20	127	0.20	120	0.19	120	0.19
Contrasting Color (CTC)	148	0.23	147	0.23	147	0.23	147	0.23
Reverse Contrast (RC)	36	0.06	36	0.06	37	0.06	37	0.06
Masking (MSK)	808	1.27	806	1.27	802	1.26	803	1.26
Text-to-Speech (TTS)	12428	19.56	12740	20.02	12498	19.65	12502	19.66
Spanish Translation (ST)	518	0.82	747	1.17	769	1.21	735	1.16
Video Sign Language (VSL [ASL])	13	0.02	13	0.02	12	0.02	13	0.02
Color Overlay	45	0.07	45	0.07	45	0.07	45	0.07
Amplification Device	44	0.07	45	0.07	44	0.07	44	0.07
Small Group Translation	70	0.11	102	0.16	100	0.16	97	0.15
Translator/Interpreter	27	0.04	39	0.06	39	0.06	37	0.06
Read Aloud in Spanish	44	0.07	92	0.14	89	0.14	89	0.14
Closed Captioning (C CAP) ELA	53	0.08						
Listening Scripts (LS) ELA	5	0.01						
Abacus Math			36	0.06				
Non-embedded Calculator Math			263	0.41				
Multiplication Table Math			2074	3.26				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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**
Braille (BRL)	8	0.01	8	0.01
Bilingual Dictionary			173	0.27
Magnification	112	0.17	110	0.17
Noise Buffers	834	1.29	825	1.27
Read Aloud in English	699	1.08	745	1.15
Scribe	508	0.79	465	0.72
Separate Setting	8373	12.95	8435	13.03
Alternate Response Options	10	0.02	10	0.02
Read Aloud (Reading Passages)	5	0.01		
Provided Color Choices (CC)	119	0.18	110	0.17
Contrasting Color (CTC)	88	0.14	88	0.14
Reverse Contrast (RC)	48	0.07	47	0.07
Masking (MSK)	657	1.02	646	1.00
Text-to-Speech (TTS)	11387	17.61	11657	18.01
Spanish Translation (ST)	565	0.87	718	1.11
Video Sign Language (VSL [ASL])	18	0.03	16	0.02
Color Overlay	35	0.05	35	0.05
Amplification Device	44	0.07	42	0.06
Small Group Translation	62	0.10	78	0.12
Translator/Interpreter	30	0.05	40	0.06
Read Aloud in Spanish	55	0.09	94	0.15
Closed Captioning (C CAP) ELA	50	0.08		
Listening Scripts (LS) ELA	15	0.02		
Abacus Math			17	0.03
Non-embedded Calculator Math			279	0.43
Multiplication Table Math			2415	3.73

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 6	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Braille (BRL)	5	0.01	5	0.01
Print-on-Demand (POD)	1	0.00		
Bilingual Dictionary			225	0.34
Magnification	68	0.10	69	0.11
Noise Buffers	461	0.71	458	0.70
Read Aloud in English	446	0.68	470	0.72
Scribe	388	0.59	367	0.56
Separate Setting	7675	11.74	7718	11.79
Alternate Response Options	6	0.01	5	0.01
Read Aloud (Reading Passages)	1	0.00		
Provided Color Choices (CC)	82	0.13	80	0.12
Contrasting Color (CTC)	63	0.10	63	0.10
Reverse Contrast (RC)	31	0.05	31	0.05
Masking (MSK)	715	1.09	703	1.07
Text-to-Speech (TTS)	9478	14.50	9674	14.78
Spanish Translation (ST)	296	0.45	407	0.62
Video Sign Language (VSL [ASL])	21	0.03	20	0.03
Color Overlay	29	0.04	29	0.04
Amplification Device	48	0.07	48	0.07
Small Group Translation	56	0.09	72	0.11
Translator/Interpreter	24	0.04	33	0.05
Read Aloud in Spanish	22	0.03	50	0.08
Closed Captioning (C CAP) ELA	61	0.09		
Listening Scripts (LS) ELA	13	0.02		
Abacus Math			5	0.01
Non-embedded Calculator Math			426	0.65
Multiplication Table Math			2683	4.10

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 7	English Language Arts		Mathematics	
Accommodation or Support	**N Count**	**Percent**	**N Count**	**Percent**
Braille (BRL)	4	0.01	4	0.01
Print-on-Demand (POD)	4	0.01	2	0.00
Bilingual Dictionary			281	0.44
Magnification	66	0.10	68	0.11
Noise Buffers	554	0.87	548	0.86
Read Aloud in English	397	0.62	441	0.69
Scribe	269	0.42	259	0.40
Separate Setting	7173	11.23	7224	11.29
Alternate Response Options	8	0.01	8	0.01
Read Aloud (Reading Passages)	3	0.00		
Provided Color Choices (CC)	148	0.23	144	0.23
Contrasting Color (CTC)	106	0.17	104	0.16
Reverse Contrast (RC)	75	0.12	72	0.11
Masking (MSK)	721	1.13	717	1.12
Text-to-Speech (TTS)	8570	13.42	8701	13.60
Spanish Translation (ST)	317	0.50	425	0.66
Video Sign Language (VSL [ASL])	18	0.03	18	0.03
Color Overlay	31	0.05	31	0.05
Amplification Device	24	0.04	23	0.04
Small Group Translation	30	0.05	42	0.07
Translator/Interpreter	16	0.03	30	0.05
Read Aloud in Spanish	21	0.03	45	0.07
Closed Captioning (C CAP) ELA	61	0.10		
Listening Scripts (LS) ELA	25	0.04		
Abacus Math			6	0.01
Non-embedded Calculator Math			499	0.78
Multiplication Table Math			2544	3.98

Copyright © 2019 by the Wisconsin Department of Public Instruction

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**
Braille (BRL)	4	0.01	4	0.01	4	0.01	4	0.01
Print-on-Demand (POD)	2	0.00	1	0.00	1	0.00	1	0.00
Bilingual Dictionary			176	0.28	177	0.28	174	0.28
Magnification	58	0.09	59	0.09	58	0.09	58	0.09
Noise Buffers	386	0.61	380	0.60	377	0.60	376	0.60
Read Aloud in English	367	0.58	370	0.59	371	0.59	365	0.58
Scribe	207	0.33	203	0.32	197	0.31	196	0.31
Separate Setting	6899	10.94	6969	11.04	6882	10.91	6868	10.89
Alternate Response Options	8	0.01	7	0.01	7	0.01	7	0.01
Read Aloud (Reading Passages)	1	0.00						
Provided Color Choices (CC)	179	0.28	177	0.28	168	0.27	169	0.27
Contrasting Color (CTC)	118	0.19	117	0.19	117	0.19	117	0.19
Reverse Contrast (RC)	86	0.14	85	0.13	84	0.13	85	0.13
Masking (MSK)	566	0.90	562	0.89	552	0.88	551	0.87
Text-to-Speech (TTS)	7990	12.67	8115	12.86	8182	12.97	7974	12.65
Spanish Translation (ST)	296	0.47	384	0.61	401	0.64	393	0.62
Video Sign Language (VSL [ASL])	21	0.03	21	0.03	21	0.03	21	0.03
Color Overlay	32	0.05	32	0.05	32	0.05	32	0.05
Amplification Device	36	0.06	36	0.06	36	0.06	36	0.06
Small Group Translation	45	0.07	59	0.09	58	0.09	57	0.09
Translator/Interpreter	24	0.04	35	0.06	30	0.05	30	0.05
Read Aloud in Spanish	21	0.03	44	0.07	44	0.07	45	0.07
Closed Captioning (C CAP) ELA	61	0.10						
Listening Scripts (LS) ELA	11	0.02						
Abacus Math			8	0.01				
Non-embedded Calculator Math			714	1.13				
Multiplication Table Math			2269	3.60				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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**
Braille (BRL)	2	0.00
Print-on-Demand (POD)	2	0.00
Bilingual Dictionary	216	0.34
Magnification	38	0.06
Noise Buffers	41	0.06
Read Aloud in English	183	0.29
Scribe	86	0.14
Separate Setting	4312	6.79
Alternate Response Options	1	0.00
Provided Color Choices (CC)	22	0.03
Contrasting Color (CTC)	10	0.02
Reverse Contrast (RC)	9	0.01
Masking (MSK)	186	0.29
Text-to-Speech (TTS)	3035	4.78
Spanish Translation (ST)	259	0.41
Video Sign Language (VSL [ASL])	21	0.03
Color Overlay	6	0.01
Amplification Device	19	0.03
Small Group Translation	102	0.16
Translator/Interpreter	12	0.02
Read Aloud in Spanish	39	0.06

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 4-8 Summary Table of Manual Materials

Material	Configuration
eDIRECT User Guide: User Management, Students, and Testing	The eDIRECT Users Guide is a 51-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 • 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 31-page document that outlines the various accessibility options available to students taking the Wisconsin Forward Exam. Guidelines for using the various accessibility features are also included.
Student/Administrator Tutorials	The Student Tutorial includes 12 videos intended for students in grades 4–10 and 7 videos for students in grade 3. It is designed to show students the interface of the online testing system and familiarize them with the tools and features available. It is intended to accompany the Online Tools Training (OTT). The 2019 tutorial also includes ten videos for test administrators to familiarize them with the administrative features and functionality of eDIRECT as well as the accessibility features of the Wisconsin Forward Exam.
Item Samplers	The item samplers can be used by both educators and students to gain familiarity with the types of items and functionality. The format appears as a "guided practice test" in the online, PDF, and Braille versions. Accommodations, universal tools, and supports are available in the test engine for the item samplers. Item samplers are accessible through the OTT menu page. They include the answer key and scoring guide for each item.
Online Tools Training (OTT)	The OTT is a hands-on opportunity for students to become familiar with logging in, navigating, using tools, using accessibility features, reviewing, and submitting the test prior to signing in to an actual test. It is designed to be a second step after viewing the student tutorials.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Material	Configuration
TAM (Test Administration Manual) and Test Directions	The TAM is a 79-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 The TAM also includes information previously provided in the DAC/SAC Guide (District Assessment Coordinator/School Assessment Coordinator Guide), which includes: • 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 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 362-page document, split into five volumes, intended for Technology Coordinators. It includes detailed instructions on the installation and configuration of INSIGHT and the Testing Site Manager or Central Office Services for all supported platforms.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Material	Configuration
Interpretive Guide	The Interpretive Guide is a 28-page document that includes the following information: • Interpreting Wisconsin Forward Exam scores • Accessing Individual Student Reports (ISRs) and summary reports via the eDIRECT Portal
Technology Readiness Package	The Technology Readiness Package is a suite of documents and tools for Technology Coordinators to prepare for the Wisconsin Forward Exam that includes the following: • Capacity estimator • System requirements • Technology overview presentation • Technology Coordinator Checklist • Tech FAQ
Technical Report	The Technical Report is a manual that covers all grades and all psychometric details associated with administering the Wisconsin Forward Exam. The Technical Report provided by DRC presents thorough documentation to demonstrate the assessment validity. The document contains the following information: • Description of the item pool used in the Wisconsin form-development process • Description of the test administration process and test security • Scoring of various types of items • Summary information of student performance (including means and standard deviations of scale scores, percentage of examinees within each performance level for each content area and grade level, and scale score distribution tables) • Item- and test-level analysis information for each content area and grade level, test scaling procedure, and student scoring process • Measures of scoring reliability for text-dependent analysis items • Evidence of test validity
Data Forensic Report	A separate Data Forensic Report includes analyses of the following: • Evaluation of response changes • Evaluation of student response time to items

Copyright © 2019 by the Wisconsin Department of Public Instruction

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. (p. 91)

Part 5 describes

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

5.1 Multiple-Choice and Multi-Select Item Scoring Process

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

- A scoring rubric was created for each TE, SA, and EBSR item. It was similar to describing the one correct answer for dichotomously scored items (scored as either right or wrong). For ELA EBSR items worth 2 points, the rubric described in detail the type of response that could receive partial credit for 1 score point.
- The information from the scoring rubric was entered into the scoring system within the item banking system so that all information about the item resided in one place, along with the item image and other metadata. This scoring information designated specific information that varied by item type. For example, for a drag-and-drop item, the information included which objects are to be placed into which drop region to receive credit.
- The information was then verified by another autoscoring expert.
- After testing started, reports were generated that showed every response, how many students gave each response, and the score the scoring system provided.
- The scoring was then checked against the scoring rubric using two levels of verification.
- If any discrepancies were found, the scoring information was modified and verified again. Scoring was then rerun. This checking and modification process continued until no other issues were found.
- As a final check, a final report was run that showed all student responses, along with frequencies and received scores.

In the case of the Braille or paper-and-pencil form administrations, student responses to paper-and-pencil TE, SA, EBSR, or TE-equivalent items were transcribed (entered) into the online system by a 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 2019 test administration, the ELA forms in grades 3–8 contained one TDA item at each grade level. As stated in Part 2, Table 2-2, of this report, the TDA prompts are scored using a holistic scoring guideline on a 1–4-point scale. A weight of 2 is later applied to the item scores in computation of the student total test raw scores and scale scores. That is, the TDA prompts will contribute up to 8 raw score points toward the student total test raw score.

The TDA responses were scored using an AI engine, and then validation scoring was performed by human scorers on approximately 10 percent of the AI scored responses. Table 5-1 presents the scoring rubric. In cases where student responses could not be scored, a non-score

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 Scoring

DRC partnered with Measurement Incorporated (MI) to score the TDA tasks. MI employed its essay scoring engine (called Project Essay Grade or PEG) to score all student responses. The AI model for scoring the Wisconsin student responses was built by first having DRC expert scorers score a representative sample of Wisconsin responses twice, independently, and resolving any scores that did not agree. While the engine only requires one score per response to build a model, the second score provides necessary information about how well two humans are able to agree on a score, which is then used as a benchmark for how well the engine's predictions should agree with the human scores. TDA items administered in grades 3 and 4 in Spring 2019 were previously administered as part of the operational test in Wisconsin in Spring 2016. As such, approximately 2,000 Spring 2016 student responses per grade that were already scored in Spring 2016 and supplemented by an additional sample of about 500 responses per grade from the Spring 2019 test administration were used in the AI scoring model building for grades 3 and 4. For grades 5 through 8, approximately 3,000 student responses per grade from the Spring 2019 test administration were selected, handscored independently twice, and used in the AI model building.

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

scoring. Quadratic weighted kappa as a metric can detect changes in mean difference and variance between scorers and is, therefore, well suited for comparing the accuracy of AI scoring with that of human scoring as well as measuring the agreement of two independent human scorers.

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

5.3.3 Handscoring Process

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

5.3.4 Handscoring System

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

5.3.5 Anchor Papers and Training Papers

DRC's project managers and scoring directors began preparations for rangefinding by using the scoring guidelines, or rubrics, to select a representative sampling of student responses for each score point. The sample reflects the various, common response types produced for a specific item. The responses were then assembled into sample sets and duplicated for all rangefinding participants, including project managers and scoring directors. This rangefinding committee read the passage(s) for the first grade/item, read and analyzed the writing prompt, and discussed the holistic scoring guideline. When an understanding of the scoring guideline had

Copyright © 2019 by the Wisconsin Department of Public Instruction

been established, participants read, scored, and discussed each response until consensus was reached. The scoring director for the specific grade took detailed notes, capturing scores and specific rationales for each score. Each grade and TDA item progressed in the same manner, using the same process. Once all sets were reviewed and scored, each grade-level scoring director selected responses to create a set of anchor papers, training papers, and qualifying papers. These anchor, training, and qualifying papers were then used to train a select group of scorers who scored the student responses that were used to train the AI engine in a process called model building. For this model-building activity, each student response was independently scored by two separate scorers. If there was any disagreement between the two readers, the scores were adjudicated to 100 percent agreement. The 2,500–3,000 responses per grade were then delivered to the AI vendor to build the AI engine model. Once the model was built, the AI engine scored the remaining Wisconsin student responses. Upon completion of the AI scoring, a random sample consisting of approximately 10 percent of the student responses scored by the AI engine was sent to DRC for a human read. DRC then scored the 10 percent read-behind sample using the original AI engine scoring group and additional scorers, trained to the same qualification standards, to ensure consistency. The 10 percent read-behind with human scorers served as a validation check of the AI engine scoring data.

5.3.6 Scoring Personnel and Qualifications

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

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

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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. (p. 118)

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

5.3.8 Monitoring the Scoring Process

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

5.3.9 Final Scores

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

5.4 Inter-rater Reliability

A random 10 percent of the AI-scored responses were sent to human scorers for second reads to validate the AI scores. The statistics for the inter-rater reliability were calculated for all TDA items. To determine the reliability of scoring, the score distribution and percentage of agreement of the two readers were examined. In this section, the distribution of TDA item scores is presented. Additional inter-rater reliability measures including intra-class correlation and weighted kappa statistics are presented in Part 9 of the Technical Report.

5.4.1 Distribution of TDA Item Scores

Table 5-3 shows the score and non-scorable code distributions for TDA items based on the census data. The presented scores, on a 1–4 point scale, are from the AI engine supplemented by non-scorable responses resolved by human readers.

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

As shown in Tables 5-4 and 5-5, there was a generally acceptable degree of agreement between the AI engine and the human scorers, with the differences being approximately 2 percent or less for most score points. Greater differences between the AI engine and the human scorers were found at score points 1 and 2 compared to the differences at points 3 and 4 at all grade levels. These discrepancies ranged from approximately 2.5 to 9 percent at score points 1 and 2, and from approximately 0 to 3 percent at score points 3 and 4. It was observed that the human scorers had higher percentages of scores 1 and 2 (combined) compared to the AI engine, while the AI engine tended to have higher percentages of scores 3 and 4 (combined).

Copyright © 2019 by the Wisconsin Department of Public Instruction

5.5 Summary

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Score Value	Score Description	Scoring Rubrics
4	Demonstrates effective analysis of text and skillful writing	• Effective addressing of all parts of the task to demonstrate an in-depth understanding of the text(s) • Strong organizational structure and focus on the task with logically grouped and related ideas, including an effective introduction, development, and conclusion • Thorough analysis based on explicit and implicit meanings from the text(s) to support claims, opinions, and ideas • 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) • Appropriate organizational structure and focus on the task with logically grouped and related ideas, including a clear introduction, development, and conclusion • Clear analysis based on explicit and implicit meanings from the text(s) to support claims, opinions, and ideas • 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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 5-1 TDA Item 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) • Weak organizational structure and focus on the task with ineffectively grouped ideas, including a weak introduction, development, and/or conclusion • Inconsistent analysis based on explicit and/or implicit meanings from the text(s) that ineffectively supports claims, opinions, and ideas • 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 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 • Minimal analysis based on the text(s) that may or may not support claims, opinions, and ideas • 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 in sentence formation, grammar, usage, spelling, capitalization, and punctuation that often interfere with meaning

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Non-scorable Code	Definition/Example/Notes
B – Blank	A response that is completely blank. This includes responses that • 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 that 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-scorable	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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 5-3 TDA Item Score Distribution

Grade	Item Number	Total Count	Item Score				Non-scorable Code				
			1	2	3	4	B	C	N	R	T
3	1	61091	19868	23826	6624	584	248	72	9424	355	90
4	1	63528	21628	23657	5593	403	245	58	11448	405	91
5	1	64654	30687	23446	5180	378	154	18	4580	189	22
6	1	65386	30002	24115	5918	580	149	10	4255	327	30
7	1	63878	27434	25229	6832	957	166	66	2795	343	56
8	1	63056	27601	20736	10836	731	166	11	2592	373	10

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

Grade	Scorer	Total Count	Score Count				Non-scorable Code Count				
			1	2	3	4	B	C	N	R	T
3	Scorer 1 (AI Engine)	9162	2643	2988	678	48			2805		
	Scorer 2 (Human)	9162	2875	2761	650	71	5	14	2703	53	30
4	Scorer 1 (AI Engine)	8057	2440	2184	526	32			2875		
	Scorer 2 (Human)	8057	3194	1558	376	54		4	2812	10	49
5	Scorer 1 (AI Engine)	6200	2957	2035	371	20			817		
	Scorer 2 (Human)	6200	3336	1765	271	11		1	800	13	3
6	Scorer 1 (AI Engine)	6512	3534	2151	454	37			336		
	Scorer 2 (Human)	6512	3349	2449	337	41		3	295	28	10
7	Scorer 1 (AI Engine)	6146	3057	2268	561	84			176		
	Scorer 2 (Human)	6146	3512	2039	370	49			145	16	15
8	Scorer 1 (AI Engine)	6019	2901	1887	911	44			276		
	Scorer 2 (Human)	6019	2474	2409	754	106	8	1	253	12	2

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Scorer	Total Count	Score Percentage				Non-scorable Code Percentage				
			1	2	3	4	B	C	N	R	T
3	Scorer 1 (AI Engine)	9162	28.85	32.61	7.40	0.52			30.62		
	Scorer 2 (Human)	9162	31.38	30.14	7.09	0.77	0.05	0.15	29.50	0.58	0.33
4	Scorer 1 (AI Engine)	8057	30.28	27.11	6.53	0.40			35.68		
	Scorer 2 (Human)	8057	39.64	19.34	4.67	0.67		0.05	34.90	0.12	0.61
5	Scorer 1 (AI Engine)	6200	47.69	32.82	5.98	0.32			13.18		
	Scorer 2 (Human)	6200	53.81	28.47	4.37	0.18		0.02	12.90	0.21	0.05
6	Scorer 1 (AI Engine)	6512	54.27	33.03	6.97	0.57			5.16		
	Scorer 2 (Human)	6512	51.43	37.61	5.18	0.63		0.05	4.53	0.43	0.15
7	Scorer 1 (AI Engine)	6146	49.74	36.90	9.13	1.37			2.86		
	Scorer 2 (Human)	6146	57.14	33.18	6.02	0.80			2.36	0.26	0.24
8	Scorer 1 (AI Engine)	6019	48.20	31.35	15.14	0.73			4.59		
	Scorer 2 (Human)	6019	41.10	40.02	12.53	1.76	0.13	0.02	4.20	0.20	0.03

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 6: Calibration, Equating, and Deriving Scale Scores

This part of the Technical Report describes the analyses involving test calibrating, equating, and student scoring that occurred for the Wisconsin Forward Exam after the 2019 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, 5.13, 5.15, and 7.2. Each standard will be explicated within the appropriate section of this part. Standard 7.2 provides general guidance that is relevant to this part:

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

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 methods that were applied to the Spring 2019 Wisconsin Forward Exam, followed by a presentation of a calibration sample, and a discussion of the calibration models and the software used. The results of the calibration process, using model-to-data fit statistics, and the outcomes of test scaling are also discussed in Section 6.1. Section 6.2 describes test equating procedures and results for ELA, Mathematics, and Social Studies. Section 6.3 addresses the process for deriving scale scores from raw scores.

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

6.1 Item Calibration

This section of the report outlines the calibration procedures and results for the Spring 2019 Wisconsin Forward Exam.

6.1.1 Calibration Models

The three-parameter logistic (3PL) model and the two-parameter partial credit (2PPC) IRT model (Bock & Aitkin, 1981; Thissen, 1982) were used to estimate parameters for

Copyright © 2019 by the Wisconsin Department of Public Instruction

multiple-choice (MC) items and constructed-response (CR) items, respectively. All non-MC items, including technology-enhanced (TE) items, evidence-based selected response (EBSR) items, short-answer (SA) items, and text-dependent analysis (TDA) items, were treated as CR items in calibrations. Item parameters for items contained in all Wisconsin assessments were estimated using a marginal maximum-likelihood procedure.

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

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

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

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

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

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

Using the 3PL/2PPC models for estimation of ELA, Mathematics, Science, and Social Studies, item parameters were consistent with the past methodology (except for the 2014–15 administration for ELA and Mathematics) implemented for Wisconsin assessments. Item parameters estimated after the 2018–19 test administration were used to score the responses of Wisconsin students who took these tests.

6.1.2 Calibration Sample

The calibration of the Wisconsin Forward Exam occurred after the Spring 2019 test administration and was based on the student data acquired during the entire testing window. This

Copyright © 2019 by the Wisconsin Department of Public Instruction

section provides information on the comparability of the calibration sample to the census data in terms of demographic characteristics in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) *Standards*:

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

The calibration sample consisted of the student data acquired during the entire testing window and included students from public, choice, and private schools. The characteristics of the calibration sample compared to the total population of students are presented in Tables 6-1 through 6-4 for ELA, Mathematics, Science, and Social Studies, respectively. The 2019 calibration samples consisted of over 99 percent of the student population and, as such, were comparable to the Wisconsin student population.

6.1.3 Calibration Procedure

The calibrations were conducted separately for each grade level and content area using the marginal maximum-likelihood procedures implemented with the expected maximum algorithm (Bock & Aitkin, 1981; Thissen, 1982). In a process of item calibration, the number of estimation cycles was set to 99 with the convergence criterion of 0.001 for all content areas. The maximum value of *a*-parameter was set to 5.0, and the range for *b*-parameter was set between -7.5 and 7.5. For all items, the estimated *a*- and *b*-parameters were within the prescribed parameter ranges. The *c*-parameters for anchor items were fixed to their Spring 2018 values for ELA, Mathematics, and Social Studies. It should be noted that there was a small number of items with the default value for the *c*-parameter on all tests. When the PARDUX (Burket, 2002) program, which is used to calibrate the items, encounters difficulty estimating the *c*-parameter, it assigns a default *c*-parameter value of 0.20.

6.1.4 Calibration Software

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

PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. Extensive simulation studies and comparisons between PARDUX and MULTILOG (Thissen, 1990), PARSCALE (Muraki & Bock, 1991), and BIGSTEPS (Wright & Linacre, 1992) have shown that PARDUX provides precise parameter and ability estimates and performs as well or more efficiently than these programs (Fitzpatrick, 1991; Fitzpatrick and Julian, 1996). Extensive research with simulation data has also shown that the IRT procedures used for calibration and scaling of Wisconsin assessments produce accurate vertical scaling (Yen & Burket, 1997).

Copyright © 2019 by the Wisconsin Department of Public Instruction

6.1.5 Calibration Results

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

IRT Item Parameters

During calibration, items may not converge, meaning the characteristics of the items will not be determined. When this occurs, items may be suppressed from student scoring and future assessments. In Spring 2019, no non-convergence issues occurred for any item on the operational tests.

IRT Item Fit

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

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

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

estimated student ability and item parameters reached a certain threshold, the item was flagged for "misfit."

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

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

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

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

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

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

Table 6-5 presents items that were flagged for less-than-optimal fit when the obtained *Z*-statistic exceeded the critical *Z*-statistic value. This table specifies the content area, grade level, item number in the calibration, item type (MC or CR), *N* size (i.e., the number of students who took this item), *Z*, and critical *Z*, as described previously. Fifteen items were flagged for poor fit for ELA, seven items were flagged for Mathematics, three items were flagged for Science, and one item was flagged for Social Studies. Most of the flagged items were CR items (TE and EBSR). For example, item #3 for ELA grade 3 was flagged because the observed *Z* of 361.06 was larger than the critical *Z* value of 162.62 based on a sample size of 60,981. For many of the flagged items, the observed *Z* and the critical *Z* were not very far apart, indicating small misfit; however, it was observed that for some items, the misfit was moderate (e.g., item #3 for ELA grade 3, item #23 for ELA grade 4, or item #38 for Mathematics grade 7). Somewhat larger item misfit was observed for item #6 for ELA grade 5.

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.

All cases of MC items flagged for misfit had empirical (observed) information that differed from the model in the lower-ability range, where there are fewer students to provide information at the tail end of the distribution. Similarly, for CR items, there were, in general,

Copyright © 2019 by the Wisconsin Department of Public Instruction

fewer students at the lower score levels, which provides less information at the tail ends of the student distribution. Items that only show misfit at the tail ends of the distribution provide stable information about the majority of the students—those in the middle range of the distribution. However, if the misfit happens around the middle of the ability range, where there are many students, this may be a concern and may lead to the item being dropped from the item pool.

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

6.2 Test Equating: English Language Arts, Mathematics, and Social Studies

Test equating is a statistical process of placing scores from two or more parallel assessments onto a common scale, resulting in direct comparability of scores from two different test forms. A common-item design was used to link the assessments from 2019 to the established ELA, Mathematics, and Social Studies scales for the Wisconsin Forward Exam. Sets of items that were administered to Wisconsin students in previous operational test administrations and that were included in the Spring 2019 assessments served as the anchor sets in each ELA, Mathematics, and Social Studies grade. All anchor items were selected from Spring 2018 operational test assessments. The anchor sets constituted at least 25 percent of the Spring 2019 assessments and were representative of the Spring 2019 test content. After the item calibration, item parameters were linked to the Wisconsin Forward Exam scales using the Stocking & Lord (1983) equating procedure.

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

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

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

6.2.1 Evaluation of Anchor Items

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

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

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

Test Characteristic Curve Method

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

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

Item Response Theory Item-Ability Regression Curves

Differences between the item-ability regression curves of the anchor items in the Spring 2019 Wisconsin Forward Exam administration were also compared to previous calibrations from Spring 2018. The differences between the item curves were evaluated using the following statistics:

- UnWtd Mean = Average signed difference in estimated probability

Copyright © 2019 by the Wisconsin Department of Public Instruction

- UnWtd Mean Abs = Average absolute (unsigned) difference in estimated probability
- UnWtd RMSD = Root mean squared difference
- Wtd Mean = Weighted average signed difference in estimated probability
- Wtd Mean Abs = Weighted average absolute (unsigned) difference in estimated probability
- Wtd RMSD = Weighted root mean squared difference

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

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

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

6.2.2 Removal of Anchor Items

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

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

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

1. An item may be a candidate for removal if it is flagged for moderate or large differences on at least four of the seven statistics (listed in Section 6.2.1) considered when examining the differences between the IRT item-ability regression curves.
2. Removal of the item will only be considered after alternative explanations have been considered that may explain shifts in performance. For example, performance on the anchor item may improve because of a statewide initiative emphasizing instruction on a particular set of skills. In this case, improved performance on the item represents true growth in that area. Removing the anchor item may artificially lower test scores.

Copyright © 2019 by the Wisconsin Department of Public Instruction

3. Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of the anchor items across the content standards should remain within 10 percent of the Wisconsin Forward Exam test blueprint.
4. The number of remaining items will remain at an acceptable level of anchor set reliability. Operationally, this means the anchor set will still be representative of the total test blueprint and the anchor set may not be less than 20 percent of the total test length.

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

6.2.3 Evaluation of Equating Results

Table 6-6 provides equating results for the TCC method for ELA, Mathematics, and Social Studies. This table summarizes the following information for each grade content area: number of anchors, number of iterations, quadratic loss function (*F*), correlation between the *a*-parameter input and estimates, correlation between the *b*-parameter input and estimates, number of *a*- and *b*-parameter outliers as indicated by the root mean square deviation method, and equating constants (*A* and *B*). Note that two sets of equating results are included for ELA grade 5 and ELA grade 8 due to exclusion of one flagged anchor item from equating in each grade.

The overall alignment of the anchor TCCs was very good for all grades and content areas. Figures 6-1 through 6-15 show the TCC alignment of the anchor set before and after equating for all grades and content areas. In these figures, the input anchor set TCC (before equating) is indicated by the dashed red line and the new anchor estimate TCC (after equating) is indicated by the solid blue line. The correlations between the *a*-parameter input and estimates and between the *b*-parameter input and estimates were 0.90 or higher for all grades and content areas, except for ELA grade 5, where the correlations between the *a*-parameter input and estimates was 0.70 (in the final equating run with a reduced anchor set). One anchor item was flagged as an *a*-parameter outlier in each of the following: ELA grades 3 and 8, Mathematics grade 5, and Social Studies grades 4 and 8. Two anchor items were flagged as *a*-parameter outliers in Mathematics grades 4, 6, and 7, and in Social Studies grade 10. Three anchor items were flagged as *a*-parameter outliers in Mathematics grade 8. One anchor item was flagged as a *b*-parameter outlier in each of the following: ELA grades 3, 4, 6, and 8; Mathematics grades 3 and 5; and Social Studies grades 8 and 10. Two anchor items were flagged as *b*-parameter outliers in Mathematics grades 4, 6, and 8. Overall, the number of anchor items flagged using the TCC method was small.

Table 6-7 presents the item-ability regression statistics for the ELA grade 5 anchor item (anchor position 39; test question 24 in Session 4 of the test) and the ELA grade 8 anchor item (anchor position 19; test question 7 in Session 3 of the test) flagged using the item-ability regression curve criteria described in an earlier section of this report. These items were flagged by at least four of the statistics used to examine ICC differences using the IRT item-ability

Copyright © 2019 by the Wisconsin Department of Public Instruction

regression curve method. Figures 6-16 and 6-17 show the ICCs before and after equating for the flagged items in ELA grades 5 and 8, respectively. In these figures, the dashed red line is the ICC before equating (based on input parameters) and the solid blue line is the ICC after equating (based on new parameter estimates). Examination of statistical properties of the flagged anchor items revealed that students performed less well on these items during the Spring 2019 test administration compared to the Spring 2018 test administration. No other anchor items in any other grades or content areas were flagged using the IRT item-ability regression curve method.

The flagged anchor items were reviewed by DRC test development experts who verified that no changes to item content or format occurred between the Spring 2018 and Spring 2019 administrations. One factor of the student performance change on the ELA grade 5 flagged anchor item might have been the item position change in the test session. The passage including this item was placed at the beginning of Session 4 (Reading) in the Spring 2018 test administration and was placed toward the end of Session 4 in the Spring 2019 test administration. No plausible explanation was found for differential item performance between the two administration years for the ELA grade 8 flagged item. Both flagged anchor items were excluded from the equating process for the ELA grade 5 and 8 tests. Exclusion of the flagged anchor items from the anchor set did not significantly affect the anchor set content coverage or the equating results for these grades.

6.2.4 Test Scales

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

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

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

The scaling constants, *M1* and *M2*, used to transform equated item parameters and ability estimates in the theta metric into the scale score metric are the same as the scaling constants used in the Spring 2016 scale development. They are presented in Table 6-8.

As stated in the previous section, new scales were established for Science grades 4 and 8 after the Spring 2019 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 ability estimates in a theta metric into a scale score metric and to produce a student scale score distribution with a target mean and standard deviation for each grade. In order to differentiate the new Science scales from the previous ones, a scale score mean of 500 and a standard deviation of 50 were set for grade 4 and a scale score mean of 700 and a standard deviation of 50 were set for grade 8. The resulting grade level scale score means show "vertical relationships" (increasing scale score means across grades), but they are not true vertical scales.

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

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

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

where

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

The scale transformation formulae used to transform item parameters in the theta metric to the scale score metric for all content areas are presented below:

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

where

- A_{ss} is a discrimination parameter in the scale score metric for MC items,
- B_{ss} is a difficulty parameter in the scale score metric for MC items,
- F_{ss} is a discrimination parameter in the scale score metric for CR items,
- G_{ss} is a difficulty level (gamma) for category m_j in the scale score metric for CR items,
- a_θ is a discrimination parameter in the original theta metric for MC items,

Copyright © 2019 by the Wisconsin Department of Public Instruction

- 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, and
- C_{ss} and c_θ is a guessing parameter in the original theta metric.

Future Science test forms will be equted to the Spring 2019 baseline scale using the same equating methodology as described for ELA, Mathematics, and Social Studies in Section 6.2.

ELA Scale

Test Characteristic Curves—Figure 6-18 shows the TCCs for ELA tests. As shown in Figure 6-18, the ELA TCCs for grades 3 and 4, and grades 5 through 7 are ordinal, indicating increasing difficulty of these assessments as the grade level increases. The grade 4 TCC overlaps with the grade 5 TCC, indicating comparable difficulty of ELA grade 4 and grade 5 assessments for students of all ability levels. The grade 7 and grade 8 TCCs also overlap at all ability levels, indicating that the grade 7 and 8 ELA assessments are of comparable difficulty 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 pattern of scale scores at quartiles (see the "Growth at Quartiles" paragraph below) for grades 3–8, student ability on ELA assessments increases as grade level increases at all grade levels, indicating grade-to-grade growth.

Standard Error Curves—The SE curves for ELA presented in Figure 6-19 are U-shaped, indicating smaller errors around ability estimates that are roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring very high- and very low-achieving students are found. Overall, the SEs around the scale score were found to be reasonable for ELA assessments (for more details, see Section 6.3.1 of this report).

Growth at Quartiles—The estimated scale scores for the ELA calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-20. It can be observed that the scale scores increase as the percentile increases within each grade. Consistent with the properties of a vertical scale, the scale scores also increase at the same percentile across grade levels, indicating growth on the ELA ability scale as students move from one grade to the next (except for the grade 8 scale score at the 25th percentile, where the grade 8 scale score is slightly lower than the grade 7 scale score, indicating that the lower-ability grade 7 students scored higher on the ELA test for grade 7 compared to the performance of the lower-ability grade 8 students on the ELA test for grade 8).

Copyright © 2019 by the Wisconsin Department of Public Instruction

Mathematics Scale

Test Characteristic Curves—Figure 6-21 shows the TCCs for Mathematics assessments, which are on a vertical scale. As observed in Figure 6-21, the TCCs for Mathematics are ordinal, indicating increasing difficulty of the assessment as the grade level increases.

Standard Error Curves—The SE curves for Mathematics presented in Figure 6-22 are U-shaped (as expected), indicating smaller errors around ability estimates that are roughly in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Mathematics assessments (for more details, see Section 6.3.1 of this report).

Growth at Quartiles—The estimated scale scores for the calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-23. It can be observed that the scale scores increase as the percentile increases within each grade level. Consistent with the properties of a vertical scale, the scale scores also increase at the same percentile across grade levels, indicating growth on the Mathematics ability scale as students move from one grade to the next.

Science Scale

Test Characteristic Curves—Although the Science assessments are not vertically scaled, the TCCs for grades 4 and 8 are presented together in Figure 6-24 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a student's ability increases. The grade 4 and grade 8 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

Standard Error Curves—Figure 6-25 shows the SE curves for Science grades 4 and 8. The SE curves are U-shaped, indicating smaller errors around ability estimates that are approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Science assessments (for more details, see Section 6.3.1 of this report).

Growth at Quartiles—The estimated scale scores for the Science calibration sample at the 25th, 50th, and 75th percentiles for both grade levels are presented in Figure 6-26. The data pattern presented in this figure indicates that the scale scores increase as the percentile increases within each grade level. Because the Science assessments are not on a vertical scale, it is not appropriate to compare scale scores between grades.

Social Studies Scale

Test Characteristic Curves—Although the Social Studies assessments are not vertically scaled, the TCCs for grades 4, 8, and 10 are presented together in Figure 6-27 for comparison purposes. The TCCs are S-shaped, indicating increasing probability of a higher test score as a

Copyright © 2019 by the Wisconsin Department of Public Instruction

student's ability increases. The grade 4 and grade 8 TCCs are parallel to each other, indicating similar overall test discrimination of the two assessments.

Standard Error Curves—Figure 6-28 shows Social Studies SE curves for grades 4, 8, and 10. The SE curves are U-shaped, indicating smaller errors around ability estimates that are approximately in the middle of the scale score distribution. The SE is expected to be higher at the top and bottom ends of the ability scale, where fewer items measuring these students are found. Overall, the SEs around the scale score were found to be reasonable for Social Studies assessments (for more details, see Section 6.3.1 of this report).

Growth at Quartiles—The estimated scale scores for the Social Studies calibration sample at the 25th, 50th, and 75th percentiles for all grade levels are presented in Figure 6-29. The data pattern presented in this figure indicates that the scale scores increase as the percentile increases within each grade level. Because the Social Studies assessments are not on a vertical scale, it is not appropriate to compare scale scores between grades.

6.3 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 on 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. It is expected that as examinee ability increases, the probability of a correct response to a given item also increases, given certain conditions and assumptions. This description applies specifically to MC items; non-MC items are treated as CR items and are handled slightly differently, but they follow a logic that is essentially the same.

Whether looking at an individual item or at a group of items that make up a complete test, IRT uses probability models to describe the relationship between a student's ability and 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 a correct response, respectively. For an MC item, the probability model can be denoted as $P(u_i = 1|\theta)$. That is, P is an estimation of the probability that a student

Copyright © 2019 by the Wisconsin Department of Public Instruction

with an ability value θ would answer item *i* correctly.

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

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

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

The formula indicates that the "estimated maximum likelihood" IRT item-pattern scoring method searches for the ability estimate (θ_0) that maximizes the probability function in the equation shown above (2) and assigns an ability estimate (θ_0) as the test score for the student with the response pattern ($u_1 u_2 \Lambda\ u_n$). In other words, the scale score is the most likely, or most probable, estimate of student ability, produced in a context in which item parameters are known and based on all the items in a given test.

As indicated, the item-pattern scoring method takes into account not only a student's total raw score but also the psychometric characteristics of all items the student responded to, including the items the student responded to incorrectly. It should be noted that a weight of two was applied to ELA TDA item scores in estimation of the student total test scale scores.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-A Example of Item Parameters for a Test (cont.)

Item	Discrimination (*a*)	Location (*b*)	Guessing (*c*)	Item	Discrimination (*a*)	Location (*b*)	Guessing (*c*)
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 ($u_1 u_2 \Lambda\ u_n$)	Raw Score	Item-Pattern Score
Pam	100001100101000000000000000101	7	140
Craig	101010101010101010101010101010	15	246
Vicki	010101010101010101010101010101	15	266
Tom	001100110011001100110011001101	15	259
Evan	110011001100110011001100110010	15	265
Dan	111111111111111111111111011111	29	379

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

Figure 6-A shows the probability of each ability estimate (or scale score) for the six examinees. The total scale score range for the test is plotted on the horizontal axis. As indicated by the two vertical lines in the plot, the lower and upper limits of the scale score range are 140 and 420, respectively. The likelihood, or probability, of all possible ability estimates for each examinee is plotted on the vertical axis and ranges from 0 to 1.0. The higher the likelihood, the more probable it is that the ability estimate accurately reflects the examinee's ability level.

As indicated above, scale scores are the most likely, or the maximum likelihood, estimates of examinee ability. As can be observed for Vicki, Tom, and Evan, scores that are plus or minus only a few scale score points are markedly less likely estimates of the students' abilities. The same is true for Craig and Dan, though to a slightly lesser extent. In the case of

Copyright © 2019 by the Wisconsin Department of Public Instruction

Pam, a few scores were almost as likely as the maximum likelihood estimate reported. Those scores that appear to be more likely than the reported score are outside of the scale score range of the test (below the LOSS).

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

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

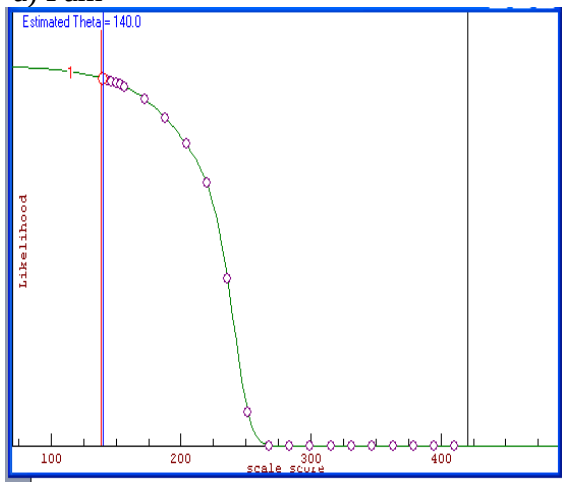
Second, reliability tends to be higher using item-pattern scoring, which means (a) fewer items are needed to achieve a given level of reliability and (b) a given test with a given number of items will have higher reliability than it would when using number-correct scoring. Yen (1984) has demonstrated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

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

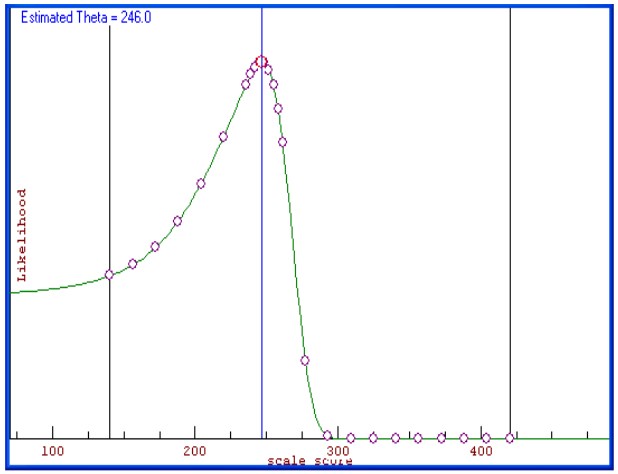
Copyright © 2019 by the Wisconsin Department of Public Instruction

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

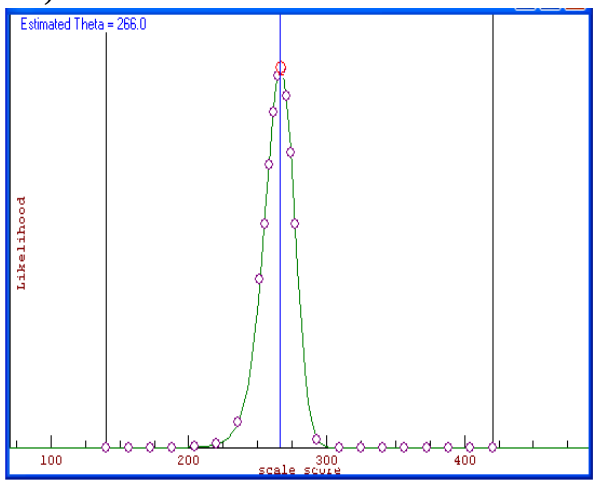
a) Pam



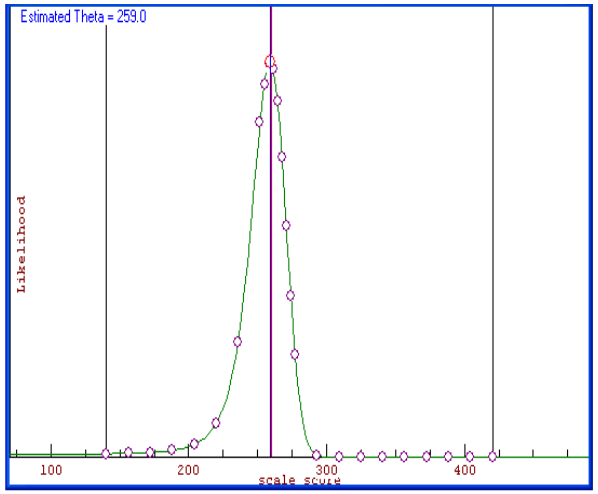
b) Craig



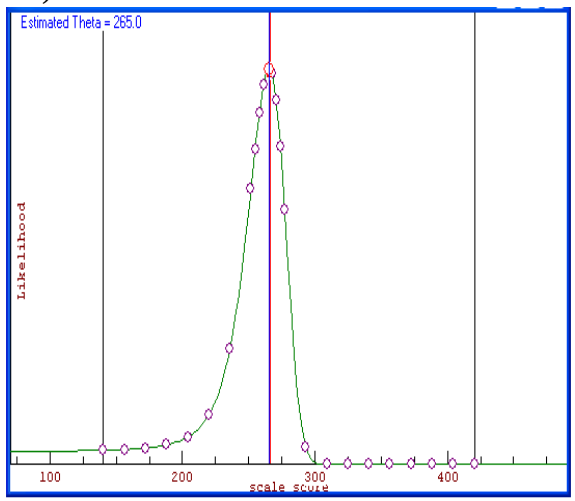
c) Vicki



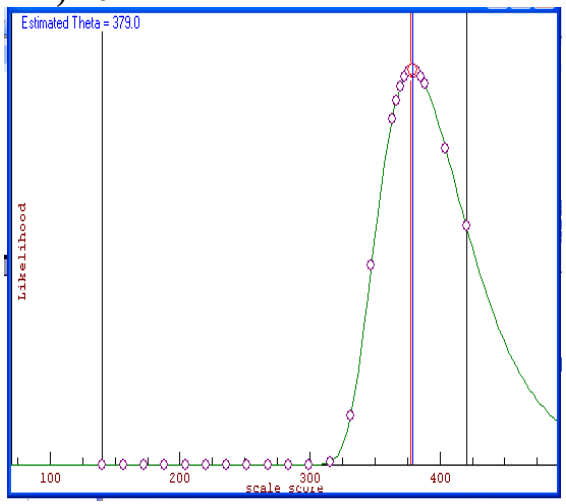
d) Tom



e) Evan



f) Dan



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

Copyright © 2019 by the Wisconsin Department of Public Instruction

6.3.1 Conditional 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 CSEM is defined as the reciprocal of the square root of the test information function and can be estimated across all points of the ability continuum (Hambleton & Swaminathan, 1985). The CSEM can be used to obtain the range within which a student's true score is likely to fall (that is, with a certain degree of probability). It is expected that a student's score obtained from a single testing will fall within one CSEM of that student's true score 68 percent of the time and that the obtained score will fall within two CSEMs of the true score 95 percent of the time.

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

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

The CSEM of the scale scores in the Spring 2019 Wisconsin Forward Exam is displayed graphically for each grade and content area in Figures 6-19 (for ELA), 6-22 (for Mathematics), 6-25 (for Science), and 6-28 (for Social Studies). The CSEM provided is based on item-pattern scoring. Each CSEM 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 higher 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 tends to be fewer test items in these score areas and fewer students. The lower and upper limits of the scale, referred to as the LOSS and HOSS, are the first scale score and the last scale score in these figures. The 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 2019 Wisconsin Forward Exam. However, scoring tables showing an approximate raw score-to-scale score relationship and the associated CSEM can be produced, and these are provided in Tables 6-9 through 6-25. These tables are provided to illustrate the approximate raw score-to-scale score relationship for each unique raw score and do not include all combinations of raw score-to-scale score associations.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

are set separately by grade and are called the LOSS and the HOSS. The LOSS and HOSS values for ELA, Mathematics, and Social Studies were established after the Spring 2016 test administration and remained unchanged through the Spring 2019 test administration. New LOSS and HOSS values for Science were established after the Spring 2019 test administration.

Table 6-26 shows the number and percentage of students at the LOSS and the HOSS. In general, there should not be many students clustered at the LOSS or HOSS. A high proportion of students at the LOSS or HOSS may indicate a floor or ceiling effect.

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

In all grades of ELA, Science, and Social Studies, the percentages of students at the LOSS and HOSS were small: approximately 1 percent or less. However, in Mathematics, all grades had more than 1 percent of students at the LOSS, ranging from 1.36perce t in Grade 3 to 3.40 percent in Grade 7. 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 response patterns of students at the LOSS in Mathematics were investigated after the Spring 2019 test administration. It was found that these students typically answered fewer than ten MC items and none of the non-MC items, which resulted in their LOSS values. For these students to receive a scale score above the LOSS, they would need to correctly answer more items, including some non-MC items. Non-MC items do not assume guessing, so the correct responses tend to represent student ability more accurately. The percentage of students at the LOSS in Mathematics may be reduced in future years by including some additional items, particularly non-MC items, that are less difficult. The percentages of students scoring at the HOSS were less than 1percent in all grades and content areas.

6.4 Summary

In summary, the overall purpose of the test scaling and equating is to ensure that the test items, as well as the overall test, are functioning appropriately. It also helps maintain the test scale so that test results may be appropriately compared across years. The data analyses undertaken by DRC are in alignment with multiple best practices of the testing industry and, in particular, support the following AERA, APA, & NCME (2014) *Standards*: 1.8, 2.13, 5.2, and 7.2.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 3	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	60986		61091		
Gender					
Male	31064	50.94	31117	50.94	0.00
Female	29922	49.06	29974	49.06	0.00
Race/Ethnicity					
White	40170	65.87	40204	65.81	-0.06
African-American	6530	10.71	6565	10.75	0.04
Hispanic	8268	13.56	8295	13.58	0.02
Asian/Pacific Islander	2552	4.18	2556	4.18	0.00
American Indian	726	1.19	726	1.19	0.00
Other	2740	4.49	2745	4.49	0.00
LEP					
No	55390	90.82	55479	90.81	-0.01
Yes	5596	9.18	5612	9.19	0.01
Disability					
No	52981	86.87	53064	86.86	-0.01
Yes	8005	13.13	8027	13.14	0.01
SES Disadvantaged					
No	33643	55.17	33672	55.12	-0.05
Yes	27343	44.83	27419	44.88	0.05
Grade 4	**N**	**%**	**N**	**%**	
All Students	63407		63528		
Gender					
Male	32446	51.17	32515	51.18	0.01
Female	30961	48.83	31013	48.82	-0.01
Race/Ethnicity					
White	41476	65.41	41517	65.35	-0.06
African-American	6954	10.97	6998	11.02	0.05
Hispanic	8659	13.66	8682	13.67	0.01
Asian/Pacific Islander	2747	4.33	2748	4.33	-0.01
American Indian	756	1.19	761	1.20	0.01
Other	2815	4.44	2822	4.44	0.00
LEP					
No	57787	91.14	57896	91.13	0.00
Yes	5620	8.86	5632	8.87	0.00
Disability					
No	55306	87.22	55388	87.19	-0.04
Yes	8101	12.78	8140	12.81	0.04
SES Disadvantaged					
No	35037	55.26	35067	55.20	-0.06
Yes	28370	44.74	28461	44.80	0.06

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 5	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	64531		64654		
Gender					
Male	32836	50.88	32901	50.89	0.00
Female	31695	49.12	31753	49.11	0.00
Race/Ethnicity					
White	42601	66.02	42646	65.96	-0.06
African-American	6977	10.81	7023	10.86	0.05
Hispanic	8847	13.71	8868	13.72	0.01
Asian/Pacific Islander	2589	4.01	2592	4.01	0.00
American Indian	750	1.16	750	1.16	0.00
Other	2767	4.29	2775	4.29	0.00
LEP					
No	59633	92.41	59741	92.40	-0.01
Yes	4898	7.59	4913	7.60	0.01
Disability					
No	56565	87.66	56653	87.62	-0.03
Yes	7966	12.34	8001	12.38	0.03
SES Disadvantaged					
No	35857	55.57	35888	55.51	-0.06
Yes	28674	44.43	28766	44.49	0.06
Grade 6	**N**	**%**	**N**	**%**	
All Students	65212		65386		
Gender					
Male	33409	51.23	33503	51.24	0.01
Female	31803	48.77	31883	48.76	-0.01
Race/Ethnicity					
White	43509	66.72	43569	66.63	-0.09
African-American	6849	10.50	6906	10.56	0.06
Hispanic	8782	13.47	8814	13.48	0.01
Asian/Pacific Islander	2596	3.98	2598	3.97	-0.01
American Indian	806	1.24	811	1.24	0.00
Other	2670	4.09	2688	4.11	0.02
LEP					
No	61231	93.90	61388	93.89	-0.01
Yes	3981	6.10	3998	6.11	0.01
Disability					
No	57277	87.83	57398	87.78	-0.05
Yes	7935	12.17	7988	12.22	0.05
SES Disadvantaged					
No	37069	56.84	37111	56.76	-0.09
Yes	28143	43.16	28275	43.24	0.09

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 7	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	63683		63878		
Gender					
Male	32678	51.31	32786	51.33	0.01
Female	31005	48.69	31092	48.67	-0.01
Race/Ethnicity					
White	42777	67.17	42845	67.07	-0.10
African-American	6493	10.20	6573	10.29	0.09
Hispanic	8644	13.57	8672	13.58	0.00
Asian/Pacific Islander	2536	3.98	2539	3.97	-0.01
American Indian	800	1.26	805	1.26	0.00
Other	2433	3.82	2444	3.83	0.01
LEP					
No	60095	94.37	60272	94.35	-0.01
Yes	3588	5.63	3606	5.65	0.01
Disability					
No	56030	87.98	56166	87.93	-0.06
Yes	7653	12.02	7712	12.07	0.06
SES Disadvantaged					
No	36940	58.01	36985	57.90	-0.11
Yes	26743	41.99	26893	42.10	0.11
Grade 8	**N**	**%**	**N**	**%**	
All Students	62817		63056		
Gender					
Male	32112	51.12	32235	51.12	0.00
Female	30705	48.88	30821	48.88	0.00
Race/Ethnicity					
White	43135	68.67	43222	68.55	-0.12
African-American	6226	9.91	6310	10.01	0.10
Hispanic	8016	12.76	8064	12.79	0.03
Asian/Pacific Islander	2445	3.89	2448	3.88	-0.01
American Indian	751	1.20	754	1.20	0.00
Other	2244	3.57	2258	3.58	0.01
LEP					
No	59841	95.26	60059	95.25	-0.02
Yes	2976	4.74	2997	4.75	0.02
Disability					
No	55451	88.27	55614	88.20	-0.08
Yes	7366	11.73	7442	11.80	0.08
SES Disadvantaged					
No	37915	60.36	37969	60.21	-0.14
Yes	24902	39.64	25087	39.79	0.14

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-2 Mathematics Calibration Sample Demographics Compared to Population

Grade 3	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	61128		61210		
Gender					
Male	31131	50.93	31178	50.94	0.01
Female	29997	49.07	30032	49.06	-0.01
Race/Ethnicity					
White	40198	65.76	40220	65.71	-0.05
African-American	6541	10.70	6568	10.73	0.03
Hispanic	8350	13.66	8375	13.68	0.02
Asian/Pacific Islander	2576	4.21	2578	4.21	0.00
American Indian	725	1.19	725	1.18	0.00
Other	2738	4.48	2744	4.48	0.00
LEP					
No	55416	90.66	55486	90.65	-0.01
Yes	5712	9.34	5724	9.35	0.01
Disability					
No	53110	86.88	53177	86.88	-0.01
Yes	8018	13.12	8033	13.12	0.01
SES Disadvantaged					
No	33708	55.14	33731	55.11	-0.04
Yes	27420	44.86	27479	44.89	0.04
Grade 4	**N**	**%**	**N**	**%**	
All Students	63533		63630		
Gender					
Male	32517	51.18	32575	51.19	0.01
Female	31016	48.82	31055	48.81	-0.01
Race/Ethnicity					
White	41488	65.30	41527	65.26	-0.04
African-American	6964	10.96	6994	10.99	0.03
Hispanic	8746	13.77	8759	13.77	0.00
Asian/Pacific Islander	2764	4.35	2767	4.35	0.00
American Indian	755	1.19	760	1.19	0.01
Other	2816	4.43	2823	4.44	0.00
LEP					
No	57804	90.98	57895	90.99	0.00
Yes	5729	9.02	5735	9.01	0.00
Disability					
No	55425	87.24	55493	87.21	-0.03
Yes	8108	12.76	8137	12.79	0.03
SES Disadvantaged					
No	35094	55.24	35118	55.19	-0.05
Yes	28439	44.76	28512	44.81	0.05

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 5	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	64631		64728		
Gender					
Male	32892	50.89	32945	50.90	0.01
Female	31739	49.11	31783	49.10	-0.01
Race/Ethnicity					
White	42609	65.93	42645	65.88	-0.04
African-American	6995	10.82	7023	10.85	0.03
Hispanic	8905	13.78	8929	13.79	0.02
Asian/Pacific Islander	2608	4.04	2608	4.03	-0.01
American Indian	749	1.16	750	1.16	0.00
Other	2765	4.28	2773	4.28	0.01
LEP					
No	59642	92.28	59732	92.28	0.00
Yes	4989	7.72	4996	7.72	0.00
Disability					
No	56654	87.66	56723	87.63	-0.02
Yes	7977	12.34	8005	12.37	0.02
SES Disadvantaged					
No	35907	55.56	35925	55.50	-0.06
Yes	28724	44.44	28803	44.50	0.06
Grade 6	**N**	**%**	**N**	**%**	
All Students	65356		65470		
Gender					
Male	33489	51.24	33540	51.23	-0.01
Female	31867	48.76	31930	48.77	0.01
Race/Ethnicity					
White	43518	66.59	43555	66.53	-0.06
African-American	6867	10.51	6907	10.55	0.04
Hispanic	8865	13.56	8887	13.57	0.01
Asian/Pacific Islander	2618	4.01	2622	4.00	0.00
American Indian	807	1.23	809	1.24	0.00
Other	2681	4.10	2690	4.11	0.01
LEP					
No	61277	93.76	61383	93.76	0.00
Yes	4079	6.24	4087	6.24	0.00
Disability					
No	57402	87.83	57484	87.80	-0.03
Yes	7954	12.17	7986	12.20	0.03
SES Disadvantaged					
No	37117	56.79	37143	56.73	-0.06
Yes	28239	43.21	28327	43.27	0.06

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 7	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	63832		63973		
Gender					
Male	32777	51.35	32850	51.35	0.00
Female	31055	48.65	31123	48.65	0.00
Race/Ethnicity					
White	42826	67.09	42875	67.02	-0.07
African-American	6506	10.19	6562	10.26	0.07
Hispanic	8707	13.64	8729	13.64	0.00
Asian/Pacific Islander	2553	4.00	2554	3.99	-0.01
American Indian	800	1.25	802	1.25	0.00
Other	2440	3.82	2451	3.83	0.01
LEP					
No	60148	94.23	60280	94.23	0.00
Yes	3684	5.77	3693	5.77	0.00
Disability					
No	56156	87.97	56262	87.95	-0.03
Yes	7676	12.03	7711	12.05	0.03
SES Disadvantaged					
No	36997	57.96	37032	57.89	-0.07
Yes	26835	42.04	26941	42.11	0.07
Grade 8	**N**	**%**	**N**	**%**	
All Students	62966		63108		
Gender					
Male	32194	51.13	32270	51.13	0.01
Female	30772	48.87	30838	48.87	-0.01
Race/Ethnicity					
White	43162	68.55	43213	68.47	-0.07
African-American	6246	9.92	6303	9.99	0.07
Hispanic	8102	12.87	8122	12.87	0.00
Asian/Pacific Islander	2459	3.91	2461	3.90	-0.01
American Indian	753	1.20	756	1.20	0.00
Other	2244	3.56	2253	3.57	0.01
LEP					
No	59898	95.13	60031	95.12	0.00
Yes	3068	4.87	3077	4.88	0.00
Disability					
No	55578	88.27	55676	88.22	-0.04
Yes	7388	11.73	7432	11.78	0.04
SES Disadvantaged					
No	37969	60.30	37992	60.20	-0.10
Yes	24997	39.70	25116	39.80	0.10

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-3 Science Calibration Sample Demographics Compared to Population

Grade 4	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	63513		63611		
Gender					
Male	32520	51.20	32565	51.19	-0.01
Female	30993	48.80	31046	48.81	0.01
Race/Ethnicity					
White	41486	65.32	41521	65.27	-0.05
African-American	6954	10.95	6982	10.98	0.03
Hispanic	8740	13.76	8762	13.77	0.01
Asian/Pacific Islander	2765	4.35	2767	4.35	0.00
American Indian	752	1.18	759	1.19	0.01
Other	2816	4.43	2820	4.43	0.00
LEP					
No	57783	90.98	57875	90.98	0.00
Yes	5730	9.02	5736	9.02	0.00
Disability					
No	55417	87.25	55484	87.22	-0.03
Yes	8096	12.75	8127	12.78	0.03
SES Disadvantaged					
No	35087	55.24	35113	55.20	-0.04
Yes	28426	44.76	28498	44.80	0.04
Grade 8	**N**	**%**	**N**	**%**	
All Students	62875		63062		
Gender					
Male	32144	51.12	32244	51.13	0.01
Female	30731	48.88	30818	48.87	-0.01
Race/Ethnicity					
White	43140	68.61	43208	68.52	-0.10
African-American	6206	9.87	6275	9.95	0.08
Hispanic	8085	12.86	8113	12.87	0.01
Asian/Pacific Islander	2459	3.91	2462	3.90	-0.01
American Indian	749	1.19	754	1.20	0.00
Other	2236	3.56	2250	3.57	0.01
LEP					
No	59818	95.14	59991	95.13	-0.01
Yes	3057	4.86	3071	4.87	0.01
Disability					
No	55533	88.32	55664	88.27	-0.05
Yes	7342	11.68	7398	11.73	0.05
SES Disadvantaged					
No	37931	60.33	37973	60.22	-0.11
Yes	24944	39.67	25089	39.78	0.11

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 4	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	63519		63603		
Gender					
Male	32517	51.19	32562	51.20	0.00
Female	31002	48.81	31041	48.80	0.00
Race/Ethnicity					
White	41491	65.32	41522	65.28	-0.04
African-American	6951	10.94	6978	10.97	0.03
Hispanic	8743	13.76	8761	13.77	0.01
Asian/Pacific Islander	2764	4.35	2766	4.35	0.00
American Indian	752	1.18	756	1.19	0.00
Other	2818	4.44	2820	4.43	0.00
LEP					
No	57788	90.98	57867	90.98	0.00
Yes	5731	9.02	5736	9.02	0.00
Disability					
No	55416	87.24	55472	87.22	-0.03
Yes	8103	12.76	8131	12.78	0.03
SES Disadvantaged					
No	35085	55.24	35109	55.20	-0.04
Yes	28434	44.76	28494	44.80	0.04
Grade 8	**N**	**%**	**N**	**%**	
All Students	62898		63045		
Gender					
Male	32157	51.13	32234	51.13	0.00
Female	30741	48.87	30811	48.87	0.00
Race/Ethnicity					
White	43141	68.59	43202	68.53	-0.06
African-American	6218	9.89	6273	9.95	0.06
Hispanic	8091	12.86	8112	12.87	0.00
Asian/Pacific Islander	2458	3.91	2459	3.90	-0.01
American Indian	752	1.20	753	1.19	0.00
Other	2238	3.56	2246	3.56	0.00
LEP					
No	59837	95.13	59974	95.13	0.00
Yes	3061	4.87	3071	4.87	0.00
Disability					
No	55538	88.30	55643	88.26	-0.04
Yes	7360	11.70	7402	11.74	0.04
SES Disadvantaged					
No	37947	60.33	37973	60.23	-0.10
Yes	24951	39.67	25072	39.77	0.10

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade 10	Calibration Sample		Population		Difference
	N	%	N	%	%
All Students	62768		63476		
Gender					
Male	32055	51.07	32423	51.08	0.01
Female	30713	48.93	31053	48.92	-0.01
Race/Ethnicity					
White	45154	71.94	45451	71.60	-0.33
African-American	5052	8.05	5261	8.29	0.24
Hispanic	7536	12.01	7668	12.08	0.07
Asian/Pacific Islander	2421	3.86	2439	3.84	-0.01
American Indian	703	1.12	713	1.12	0.00
Other	1902	3.03	1944	3.06	0.03
LEP					
No	60336	96.13	60984	96.07	-0.05
Yes	2432	3.87	2492	3.93	0.05
Disability					
No	56120	89.41	56652	89.25	-0.16
Yes	6648	10.59	6824	10.75	0.16
SES Disadvantaged					
No	41096	65.47	41371	65.18	-0.30
Yes	21672	34.53	22105	34.82	0.30

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-5 Items Flagged Based on Yen's Q1

Content	Grade	Item Number in Calibration	Type	N	*Z*	Critical *Z*
ELA	3	3	CR	60981	361.06	162.62
	3	37	CR	60981	170.44	162.62
	4	32	CR	63404	397.37	169.08
	5	6*	CR	64527	624.31	172.07
	5	13	CR	64527	268.24	172.07
	5	37*	CR	64527	270.18	172.07
	6	32	CR	65209	203.15	173.89
	7	4	CR	63681	253.84	169.82
	7	18*	CR	63681	223.78	169.82
	7	21*	CR	63681	286.68	169.82
	7	26*	CR	63681	259.54	169.82
	7	28	CR	63681	213.50	169.82
	8	19	CR	62810	200.58	167.49
	8	30*	CR	62810	240.23	167.49
	8	32*	CR	62810	196.46	167.49
Math	3	12	CR	60891	203.33	162.38
	5	25*	CR	64588	197.25	172.23
	5	40	CR	64588	292.02	172.23
	7	15	CR	63786	180.26	170.10
	7	38	CR	63786	378.65	170.10
	8	26	MC	62891	242.45	167.71
	8	42*	MC	62891	237.75	167.71
Science	4	24	CR	63466	251.19	169.24
	8	6	MC	62803	199.87	167.47
	8	28	CR	62803	177.90	167.47
Social Studies	10	39	MC	62611	231.23	166.96

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-6 Equating Evaluation Results, Stocking and Lord Method

Content Area	Grade	Number of Anchors	Stocking and Lord TCC Method Results						Equating Constants	
			TCC Results		Parameter Comparison Statistics					
					a-Parameter		*b*-Parameter			
			# of Iterations	*F* Value	Corr	# of RMSD Outliers	Corr	# of RMSD Outliers	*A*	*B*
ELA	3	13	6	0.16847	0.99	1	0.99	1	0.931	-1.243
	4	14	4	0.15880	0.98	0	1.00	1	1.043	-0.622
	5	12	8	0.40212	0.69	0	0.96	1	0.991	-0.304
	5*	11	9	0.40050	0.70	0	0.99	0	1.000	-0.273
	6	14	6	0.09086	1.00	0	1.00	1	1.022	-0.035
	7	14	7	0.02275	0.94	0	0.90	0	1.124	0.437
	8	15	8	0.06063	0.99	1	0.99	1	1.159	0.456
	8*	14	8	0.05087	0.99	1	0.99	1	1.200	0.529
Math	3	27	3	0.12743	0.98	0	0.99	1	0.961	-1.193
	4	23	4	0.02733	0.99	2	1.00	2	0.955	-0.705
	5	23	32	0.01662	0.94	1	0.98	1	0.903	-0.158
	6	25	18	0.01724	0.98	2	0.99	2	1.045	0.023
	7	29	30	0.01323	0.98	2	0.99	0	1.048	0.381
	8	26	20	0.03684	0.96	3	0.99	2	1.015	0.773
Social Studies	4	14	16	0.06069	0.98	1	0.98	0	1.121	-0.092
	8	14	15	0.08449	0.99	1	0.99	1	1.023	0.059
	10	18	7	0.10498	0.96	2	0.98	1	1.122	-0.078

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

Table 6-7 Statistics Comparing IRT Item-Ability Regression Curves for Flagged Anchor Items

Content Area	Grade	Anchor Item Position	UnWtd RMSD	UnWtd Mean Abs	Max Abs	UnWtd Mean	Wtd RMSD	Wtd Mean Abs	Wtd Mean
ELA	5	39	0.0567	0.0371	**0.1277**	-0.0369	**0.0887**	**0.0799**	**-0.0799**
ELA	8	19	**0.0747**	0.0588	**0.1339**	-0.0575	**0.1025**	**0.0944**	**-0.0943**

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-8 Scale Transformation Constants

Content Area	Grade	Scale Transformation Constants	
		M1	M2
ELA	3–8	43.7445	610.4987
Mathematics	3–8	46.4684	612.0818
Science	4	45.0450	500.4505
	8	45.0450	699.5496
Social Studies	4	40.1929	405.2251
	8	42.2297	600.8446
	10	42.8817	703.8594

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-9 Scoring Table for English Language Arts Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	100	31	563	12
1	330	100	32	567	12
2	330	100	**33**	**571**	**12**
3	330	100	34	575	12
4	330	100	35	579	12
5	362	69	36	583	12
6	410	38	37	587	12
7	432	29	38	591	12
8	447	25	39	596	13
9	459	22	40	601	13
10	468	20	41	606	13
11	476	18	42	611	14
12	483	17	43	617	14
13	489	16	44	623	15
14	495	15	**45**	**630**	**16**
15	500	15	46	638	17
16	505	14	47	647	18
17	510	13	48	657	20
18	514	13	49	669	22
19	519	13	50	684	25
20	**523**	**12**	51	703	29
21	526	12	52	735	40
22	530	12	53	900	188
23	534	12			
24	538	12			
25	541	12			
26	545	12			
27	549	11			
28	552	11			
29	556	12			
30	559	12			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-10 Scoring Table for English Language Arts Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	340	70	31	587	13
1	340	70	32	591	13
2	340	70	**33**	**595**	**13**
3	340	70	34	599	13
4	340	70	35	603	13
5	340	70	36	607	13
6	380	52	37	612	14
7	413	41	38	616	14
8	435	35	39	621	14
9	452	31	40	626	15
10	466	28	41	631	15
11	478	25	42	637	16
12	488	24	43	643	16
13	497	22	**44**	**650**	**17**
14	505	21	45	657	18
15	512	20	46	665	19
16	518	19	47	673	21
17	524	18	48	683	22
18	530	17	49	694	24
19	535	17	50	706	26
20	541	16	51	721	29
21	545	16	52	739	33
22	**550**	**15**	53	761	37
23	555	15	54	790	44
24	559	14	55	837	61
25	563	14	56	930	129
26	567	14			
27	571	13			
28	575	13			
29	579	13			
30	583	13			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-11 Scoring Table for English Language Arts Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	76	31	590	12
1	350	76	32	594	12
2	350	76	33	598	12
3	350	76	34	601	12
4	350	76	35	605	12
5	350	76	36	609	12
6	350	76	**37**	**613**	**12**
7	400	50	38	617	12
8	432	38	39	622	12
9	453	32	40	626	13
10	469	28	41	631	13
11	482	25	42	636	13
12	492	22	43	641	14
13	501	20	44	647	15
14	509	19	45	653	15
15	517	17	46	660	16
16	523	17	47	667	18
17	529	16	**48**	**676**	**19**
18	534	15	49	686	21
19	540	15	50	698	23
20	545	14	51	711	25
21	549	14	52	728	28
22	554	14	53	748	32
23	558	13	54	775	38
24	562	13	55	816	53
25	**566**	**13**	56	940	156
26	570	13			
27	574	13			
28	578	12			
29	582	12			
30	586	12			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-12 Scoring Table for English Language Arts Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	360	63	31	606	14
1	360	63	32	610	14
2	360	63	33	614	14
3	360	63	34	619	14
4	360	63	**35**	**623**	**14**
5	374	56	36	628	14
6	412	41	37	632	14
7	436	35	38	637	14
8	454	30	39	642	15
9	469	28	40	647	15
10	481	26	41	652	15
11	492	24	42	658	16
12	502	23	43	664	16
13	510	22	44	670	17
14	518	20	**45**	**677**	**17**
15	526	19	46	684	18
16	532	19	47	692	19
17	539	18	48	701	20
18	545	17	49	710	21
19	550	17	50	721	23
20	556	16	51	734	25
21	561	16	52	749	27
22	566	15	53	767	31
23	571	15	54	793	39
24	**575**	**15**	55	836	57
25	580	14	56	950	155
26	584	14			
27	589	14			
28	593	14			
29	597	14			
30	602	14			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-13 Scoring Table for English Language Arts Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	370	65	31	626	15
1	370	65	32	630	14
2	370	65	33	635	14
3	370	65	**34**	**639**	**15**
4	370	65	35	644	15
5	396	54	36	648	15
6	430	42	37	653	15
7	454	35	38	658	15
8	471	31	39	663	15
9	486	29	40	668	15
10	498	27	41	674	16
11	509	25	42	680	16
12	519	24	43	686	17
13	527	22	44	692	17
14	535	21	**45**	**699**	**18**
15	543	20	46	707	19
16	550	20	47	715	20
17	556	19	48	724	21
18	562	18	49	734	23
19	568	18	50	746	25
20	573	17	51	760	28
21	579	17	52	777	31
22	584	16	53	798	36
23	**589**	**16**	54	827	45
24	594	16	55	880	70
25	598	15	56	960	133
26	603	15			
27	608	15			
28	612	15			
29	617	15			
30	621	15			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-14 Scoring Table for English Language Arts Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	380	62	31	625	14
1	380	62	32	629	14
2	380	62	33	633	14
3	380	62	34	638	14
4	380	62	35	642	14
5	380	62	36	647	14
6	383	60	**37**	**652**	**15**
7	425	44	38	656	15
8	451	37	39	661	15
9	470	33	40	667	15
10	486	30	41	672	16
11	499	28	42	678	16
12	511	26	43	684	17
13	521	25	44	690	17
14	531	23	45	697	18
15	539	22	46	704	19
16	547	21	**47**	**713**	**20**
17	554	20	48	722	22
18	560	19	49	733	24
19	567	18	50	745	26
20	572	17	51	760	29
21	578	17	52	778	32
22	583	16	53	801	37
23	588	16	54	832	44
24	**593**	**15**	55	880	60
25	598	15	56	970	126
26	602	15			
27	607	15			
28	611	14			
29	616	14			
30	620	14			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-15 Scoring Table for Mathematics Grade 3

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-16 Scoring Table for Mathematics Grade 4

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-17 Scoring Table for Mathematics Grade 5

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-18 Scoring Table for Mathematics Grade 6

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-19 Scoring Table for Mathematics Grade 7

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-20 Scoring Table for Mathematics Grade 8

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-21 Scoring Table for Science Grade 4

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-22 Scoring Table for Science Grade 8

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-23 Scoring Table for Social Studies Grade 4

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-24 Scoring Table for Social Studies Grade 8

Raw Score	Scale Score	SEM
0	420	86
1	420	86
2	420	86
3	420	86
4	420	86
5	420	86
6	420	86
7	420	86
8	461	46
9	483	32
10	497	25
11	508	22
12	517	19
13	525	17
14	532	16
15	538	15
16	543	14
17	549	14
18	554	13
19	558	13
20	**563**	**12**
21	567	12
22	572	12
23	576	12
24	580	12
25	585	12
26	589	12
27	594	12
28	**599**	**12**
29	604	13
30	609	13
31	614	14
32	620	14
33	627	15
34	634	16
35	**642**	**17**
36	652	19
37	665	22
38	681	27
39	709	38
40	780	95

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-25 Scoring Table for Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	490	119	26	674	13
1	490	119	27	678	12
2	490	119	28	682	12
3	490	119	29	685	12
4	490	119	30	689	12
5	490	119	31	693	11
6	490	119	32	697	11
7	490	119	33	701	11
8	490	119	**34**	**705**	**12**
9	490	119	35	709	12
10	538	71	36	713	12
11	566	46	37	717	12
12	583	36	38	722	12
13	597	29	39	726	13
14	607	25	40	731	13
15	616	23	41	737	14
16	623	21	**42**	**743**	**14**
17	630	19	43	749	15
18	636	18	44	757	16
19	642	17	45	765	18
20	647	16	46	776	20
21	652	15	47	789	24
22	657	14	48	808	30
23	661	14	49	841	45
24	666	13	50	890	79
25	**670**	**13**			

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 6-26 Numbers and Percentages of Students at LOSS and HOSS

Content	Grade	LOSS	N	Percentage	HOSS	N	Percentage
ELA	3	330	42	.07	900	5	.01
	4	340	15	.02	930	3	.00
	5	350	5	.01	940	4	.01
	6	360	8	.01	950	3	.00
	7	370	9	.01	960	2	.00
	8	380	17	.03	970	7	.01
Math	3	360	831	1.36	760	235	.38
	4	405	1119	1.76	800	42	.07
	5	430	2131	3.29	830	42	.06
	6	440	2108	3.22	870	63	.10
	7	450	2178	3.40	880	43	.07
	8	470	1597	2.53	890	71	.11
Science	4	300	12	.02	725	47	.07
	8	480	10	.02	945	69	.11
Social Studies	4	200	659	1.04	570	393	.62
	8	420	301	.48	780	570	.90
	10	490	606	.95	890	156	.25

Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-1 Anchor Set TCCs: ELA Grade 3

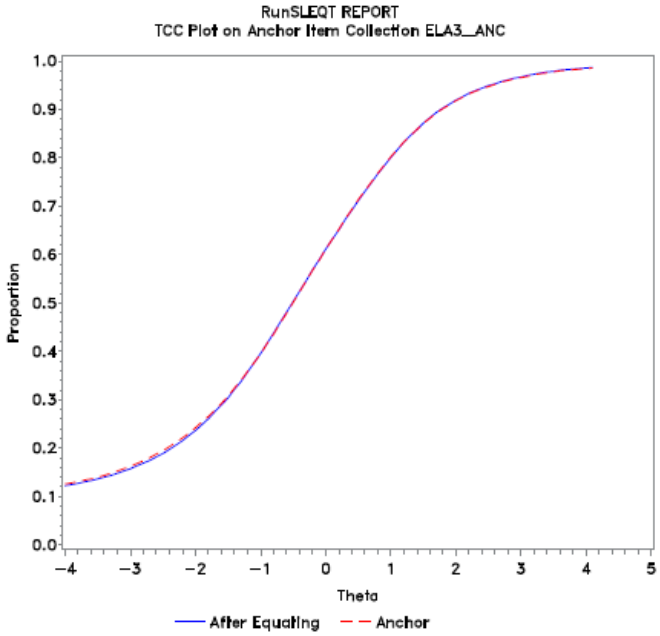


Figure 6-2 Anchor Set TCCs: ELA Grade 4

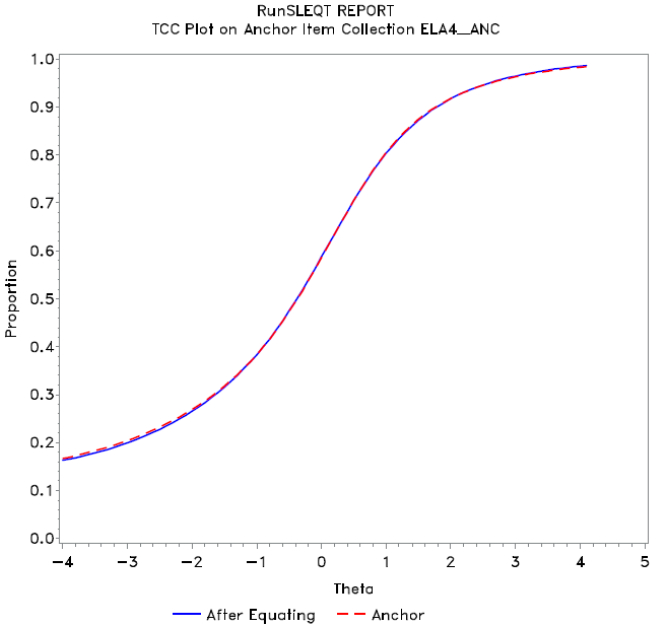


Figure 6-3 Anchor Set TCCs: ELA Grade 5

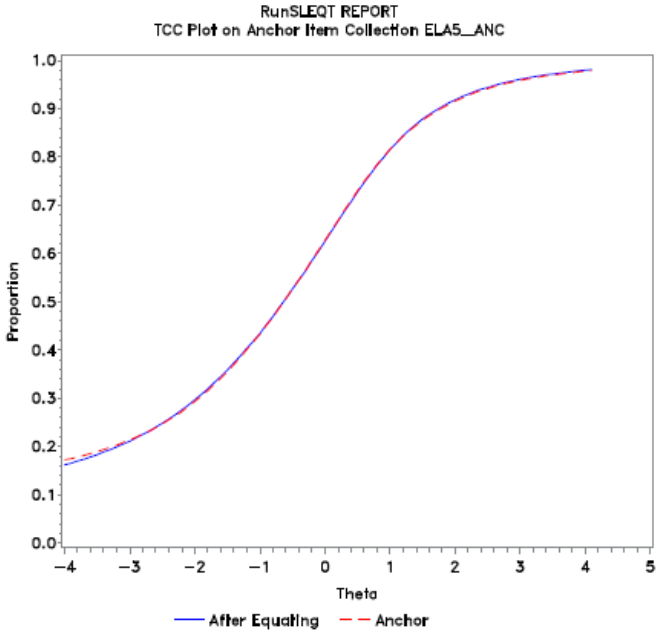


Figure 6-4 Anchor Set TCCs: ELA Grade 6

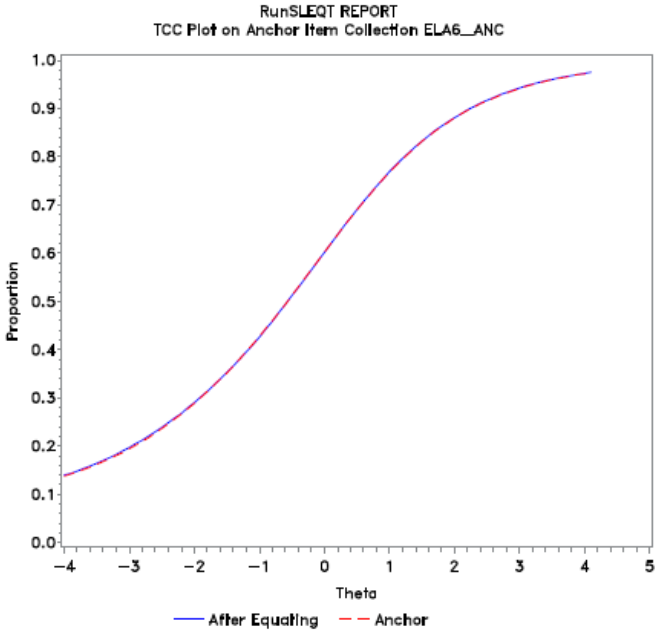


Figure 6-5 Anchor Set TCCs: ELA Grade 7

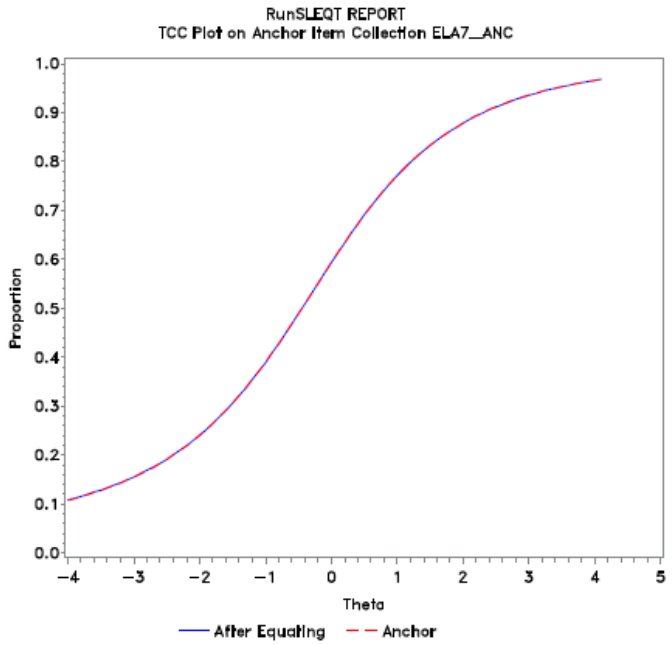
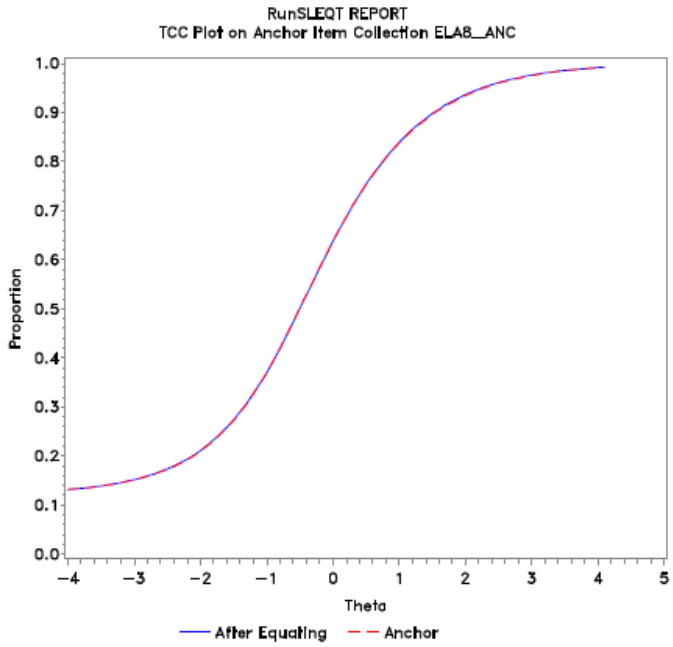


Figure 6-6 Anchor Set TCCs: ELA Grade 8



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-7 Anchor Set TCCs: Mathematics Grade 3

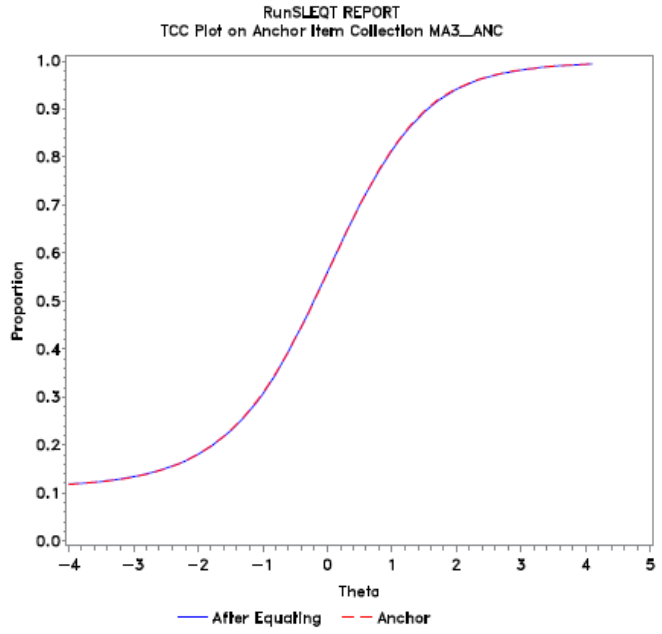
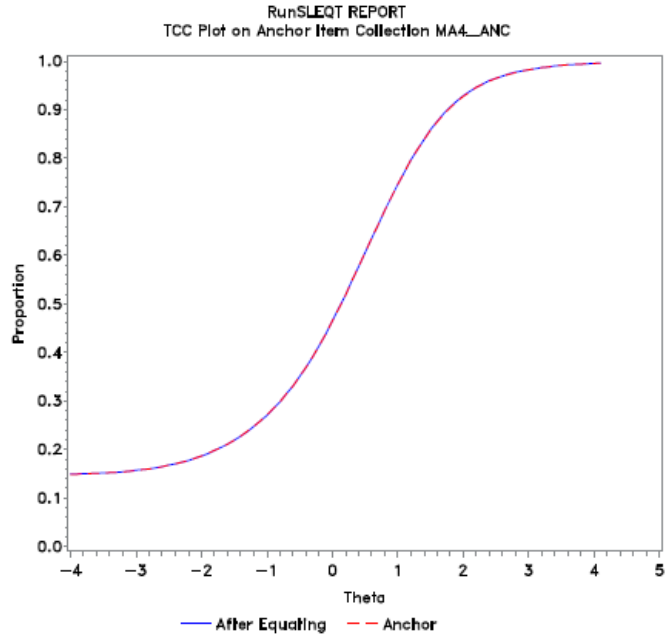


Figure 6-8 Anchor Set TCCs: Mathematics Grade 4



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-9 Anchor Set TCCs: Mathematics Grade 5

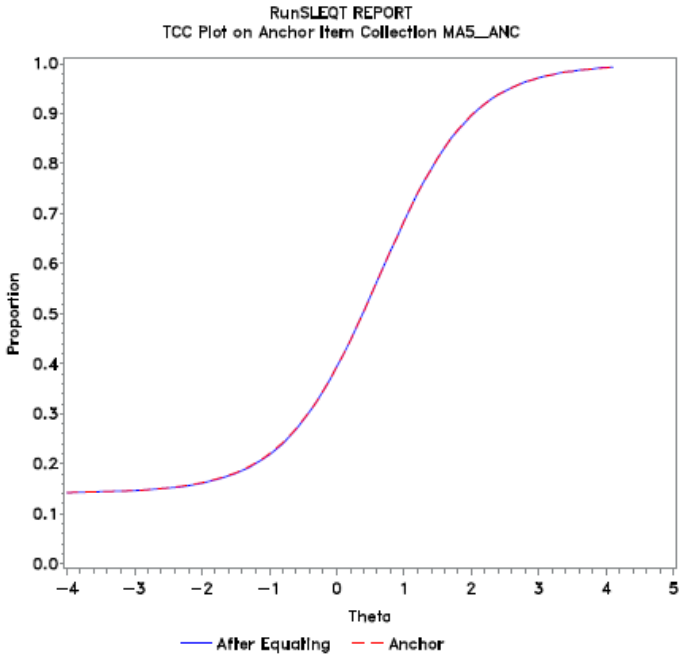
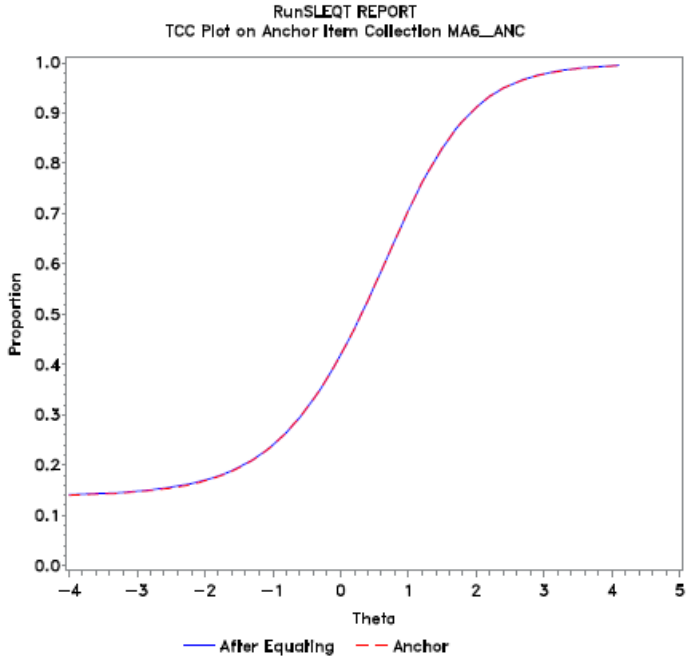


Figure 6-10 Anchor Set TCCs: Mathematics Grade 6



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-11 Anchor Set TCCs: Mathematics Grade 7

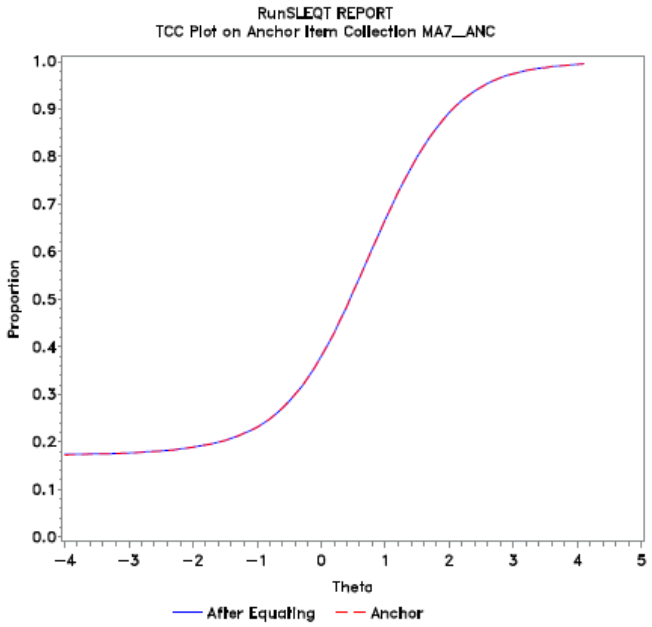
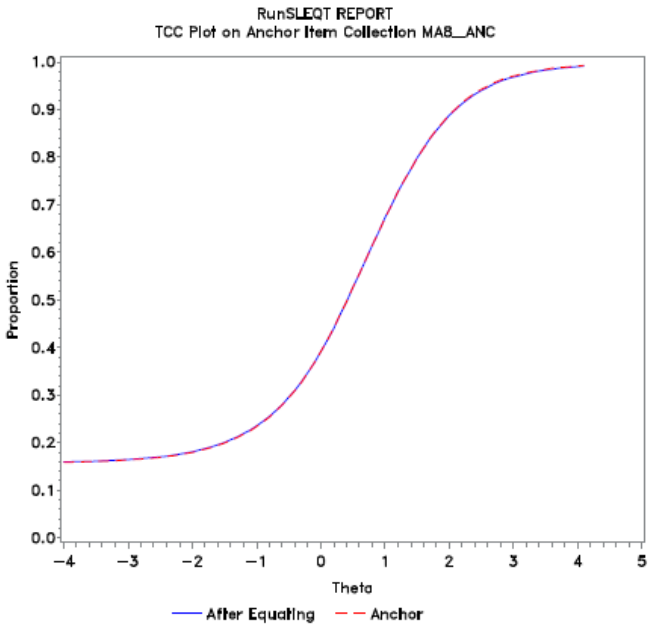


Figure 6-12 Anchor Set TCCs: Mathematics Grade 8



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-13 Anchor Set TCCs: Social Studies Grade 4

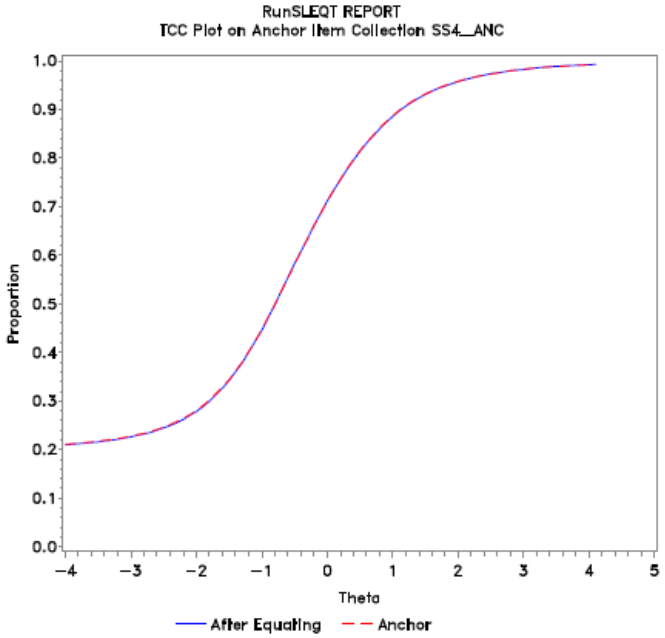


Figure 6-14 Anchor Set TCCs: Social Studies Grade 8

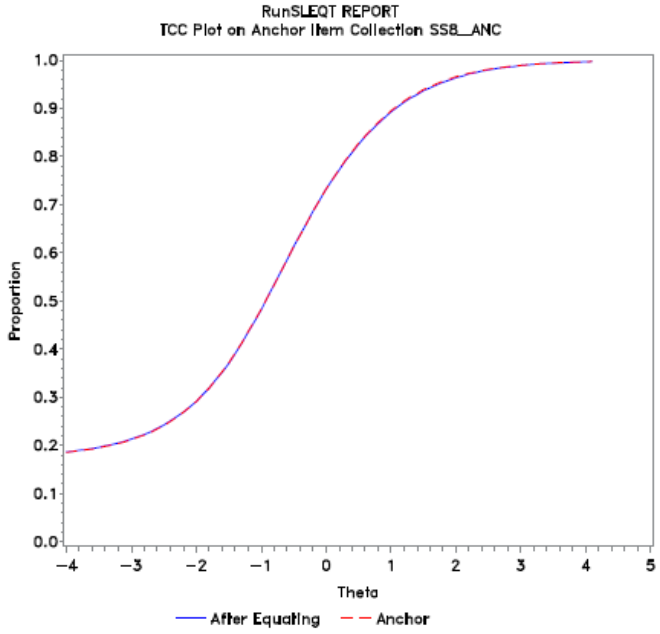


Figure 6-15 Anchor Set TCCs: Social Studies Grade 10

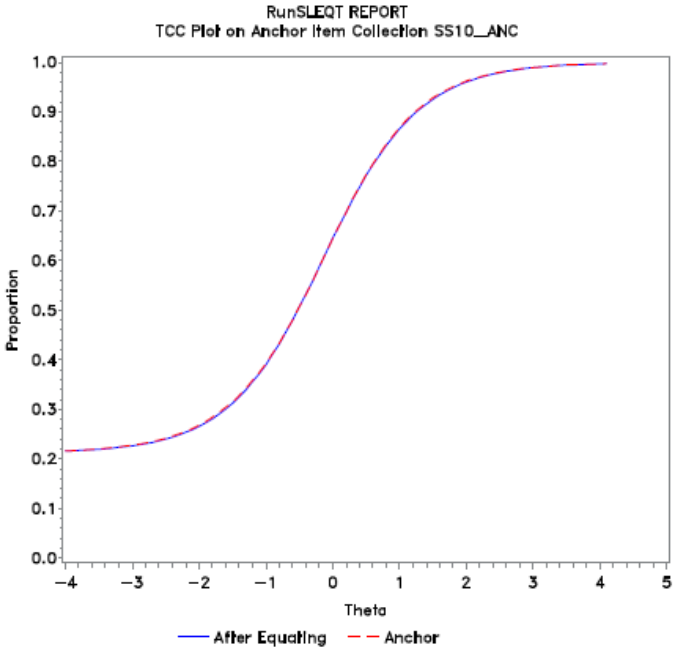


Figure 6-16 Item Characteristic Curves for the Flagged ELA Grade 5 Anchor

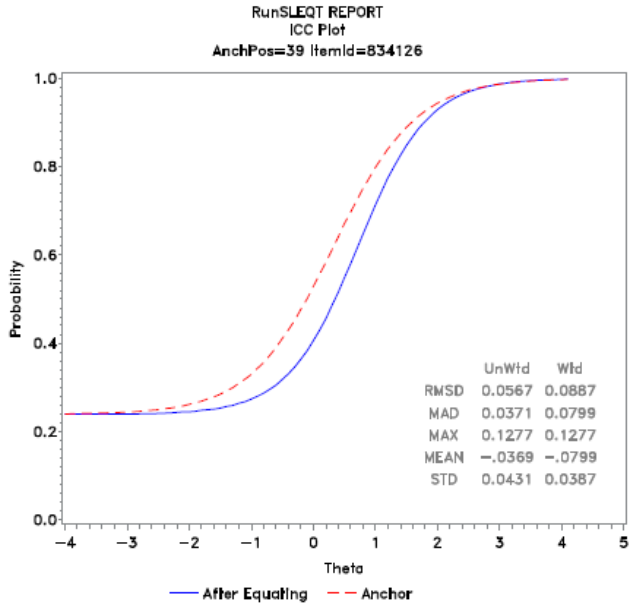


Figure 6-17 Item Characteristic Curves for the Flagged ELA Grade 8 Anchor

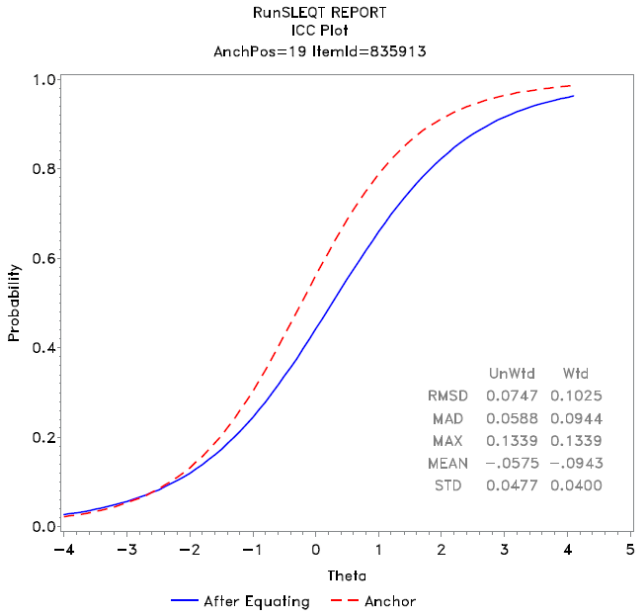


Figure 6-18 English Language Arts Test Characteristic Curves

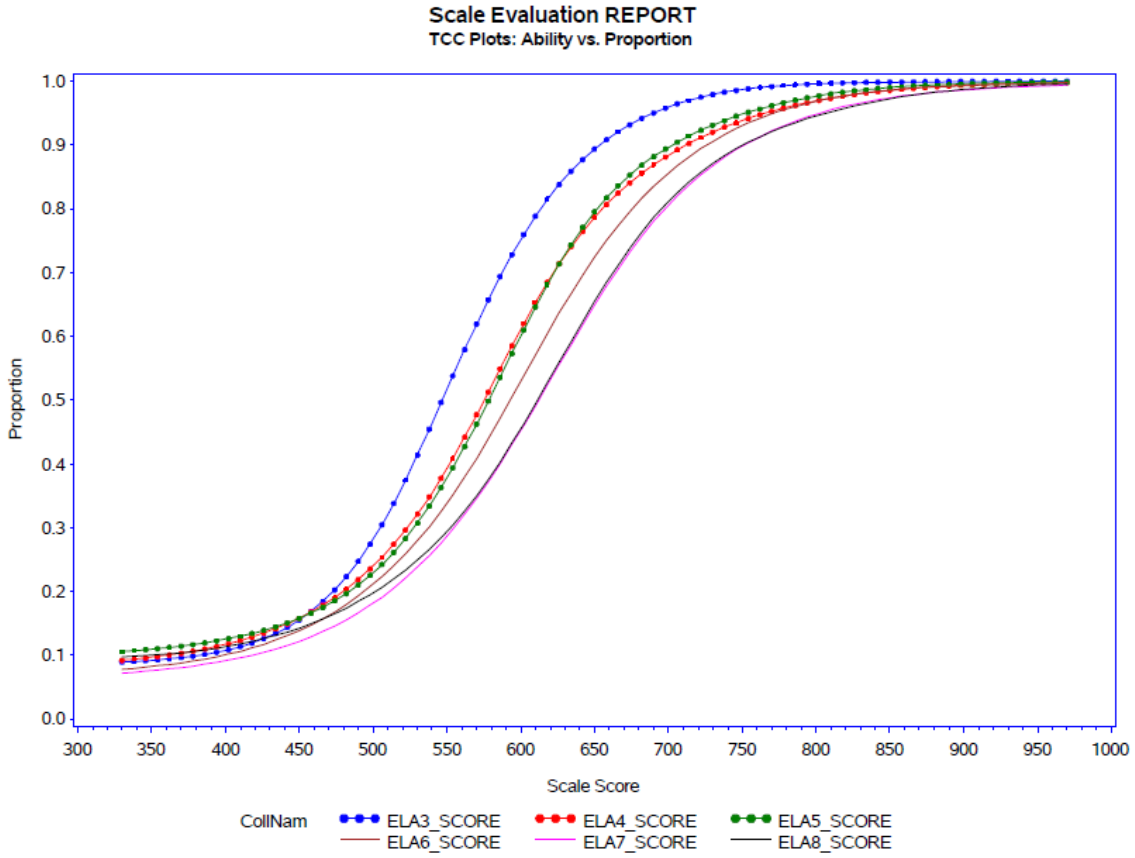


Figure 6-19 English Language Arts Standard Error Curves

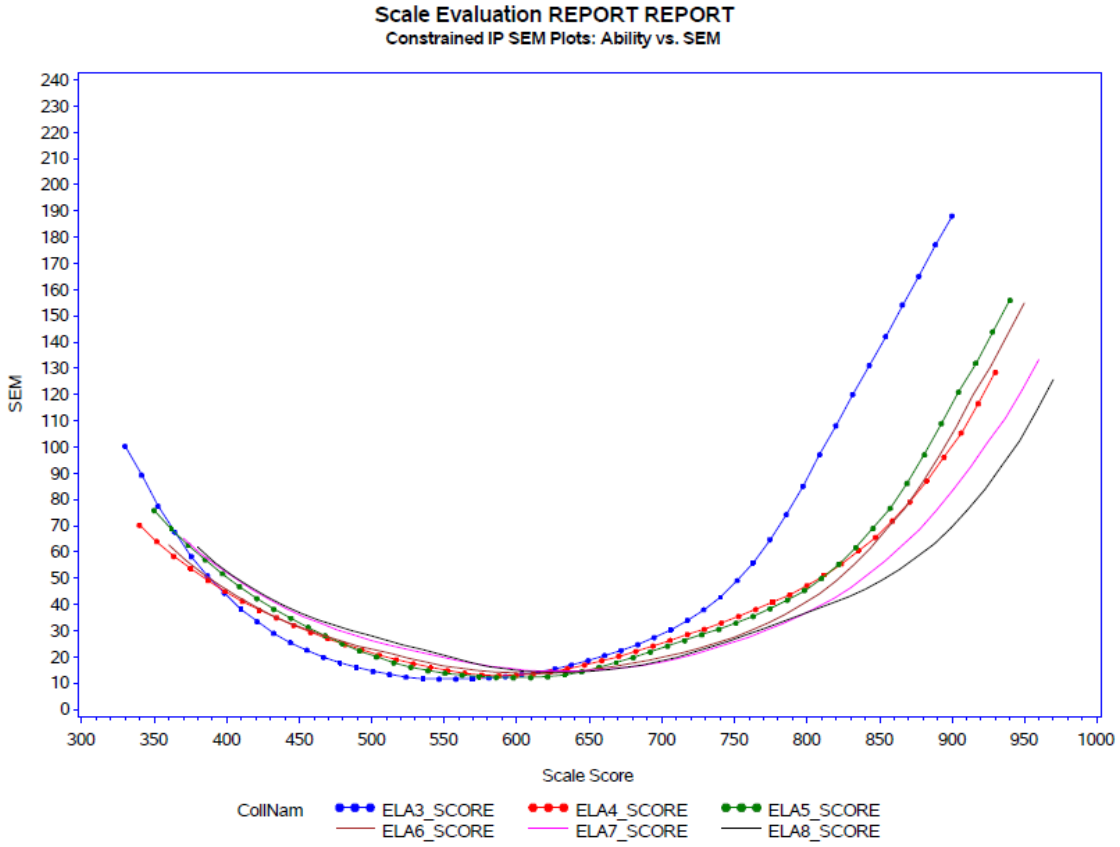
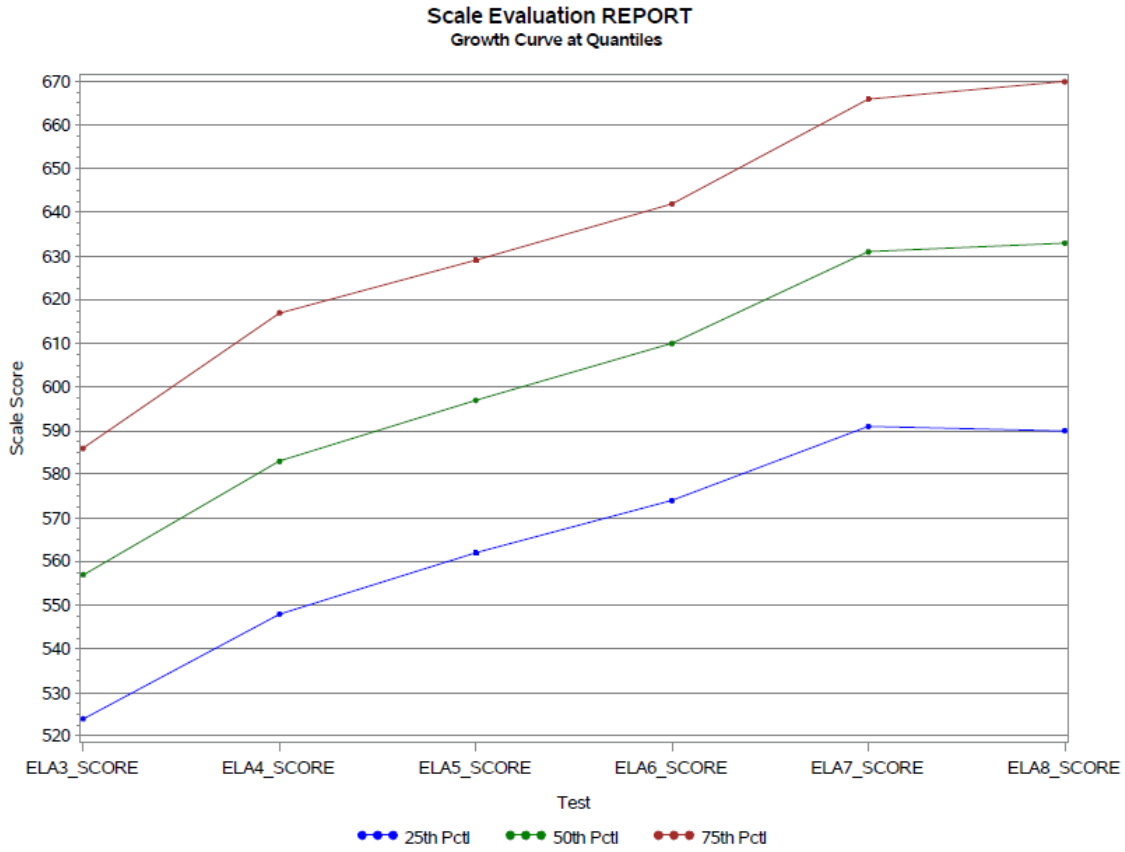
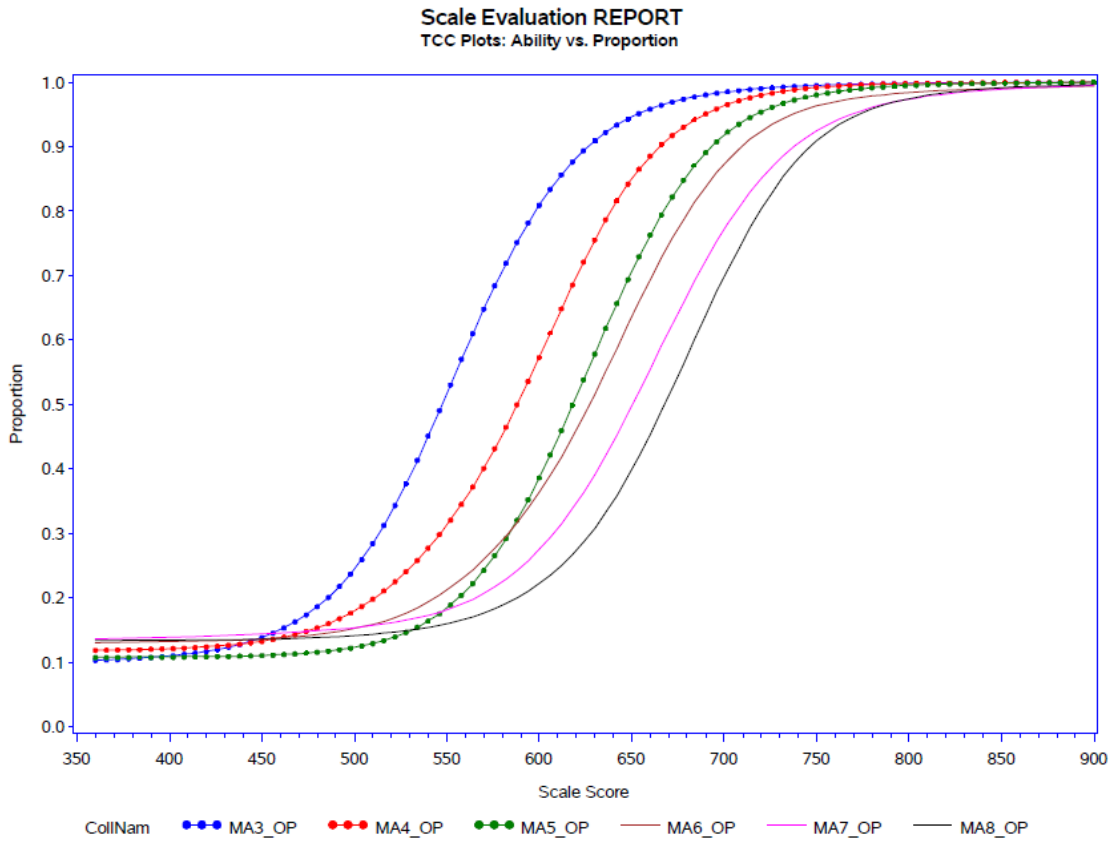


Figure 6-20 English Language Arts Growth at Quartiles



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-21 Mathematics Test Characteristic Curves



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure 6-22 Mathematics Standard Error Curves

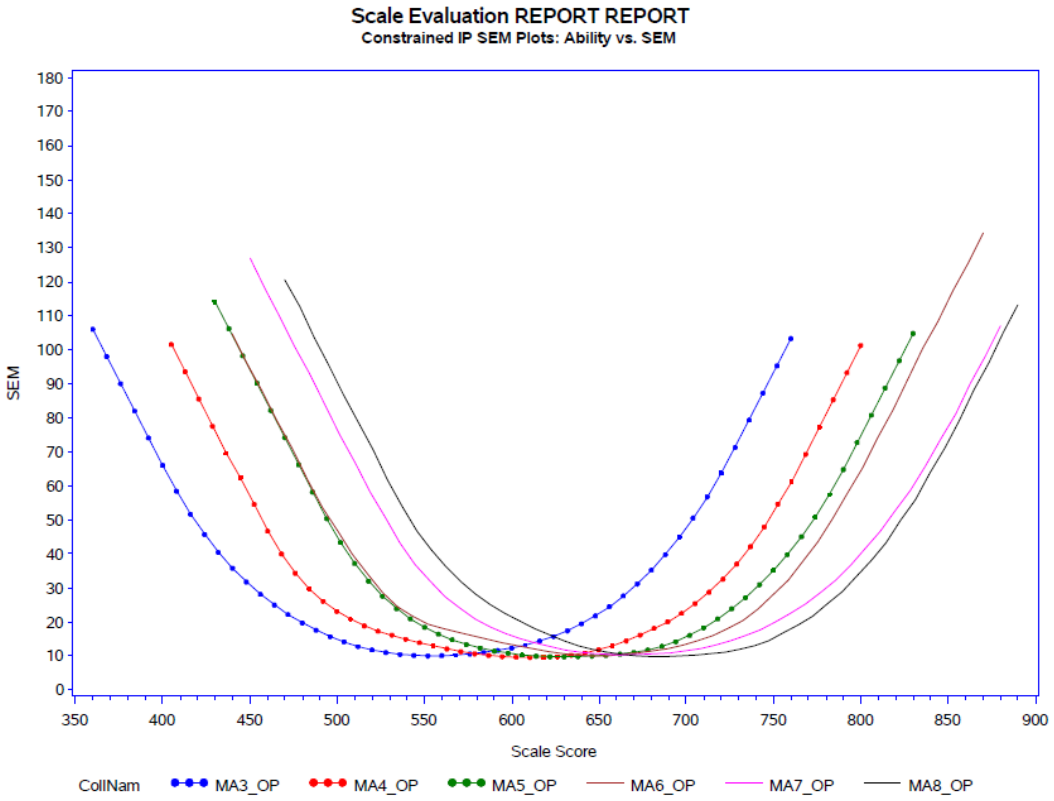


Figure 6-23 Mathematics Growth at Quartiles

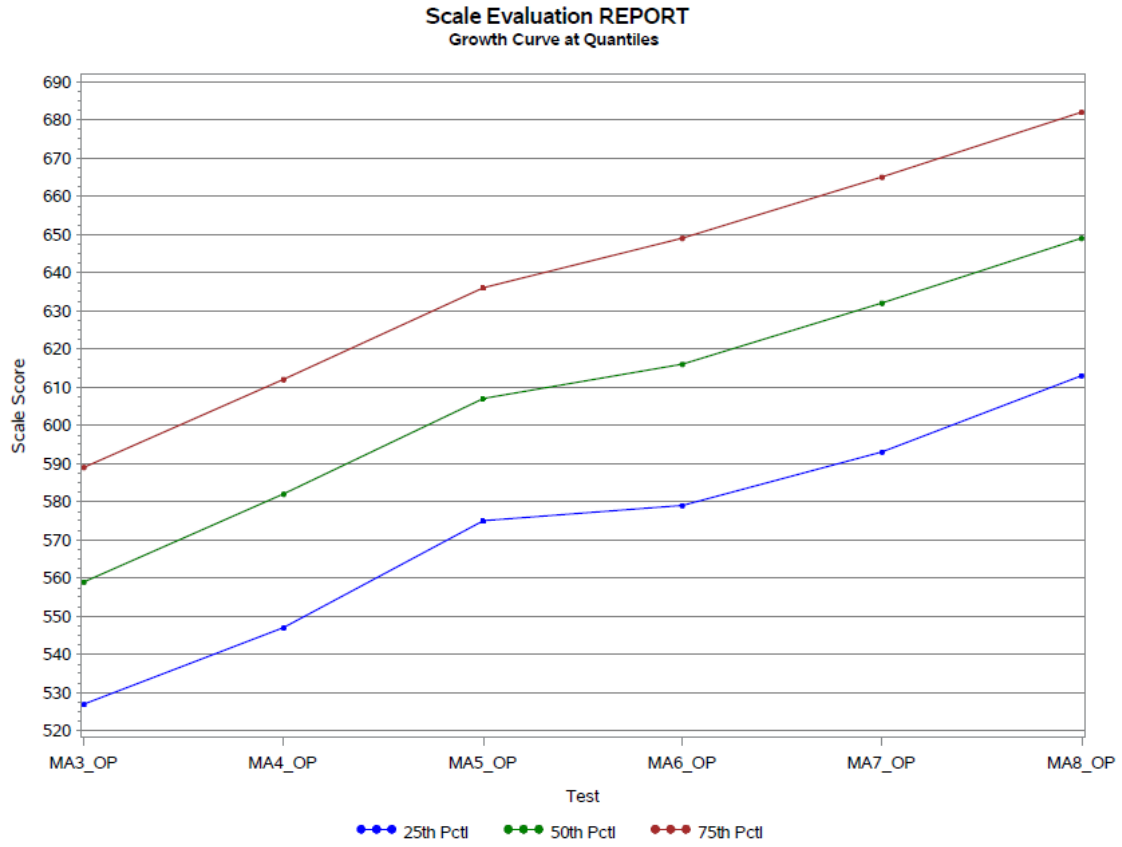


Figure 6-24 Science Test Characteristic Curves

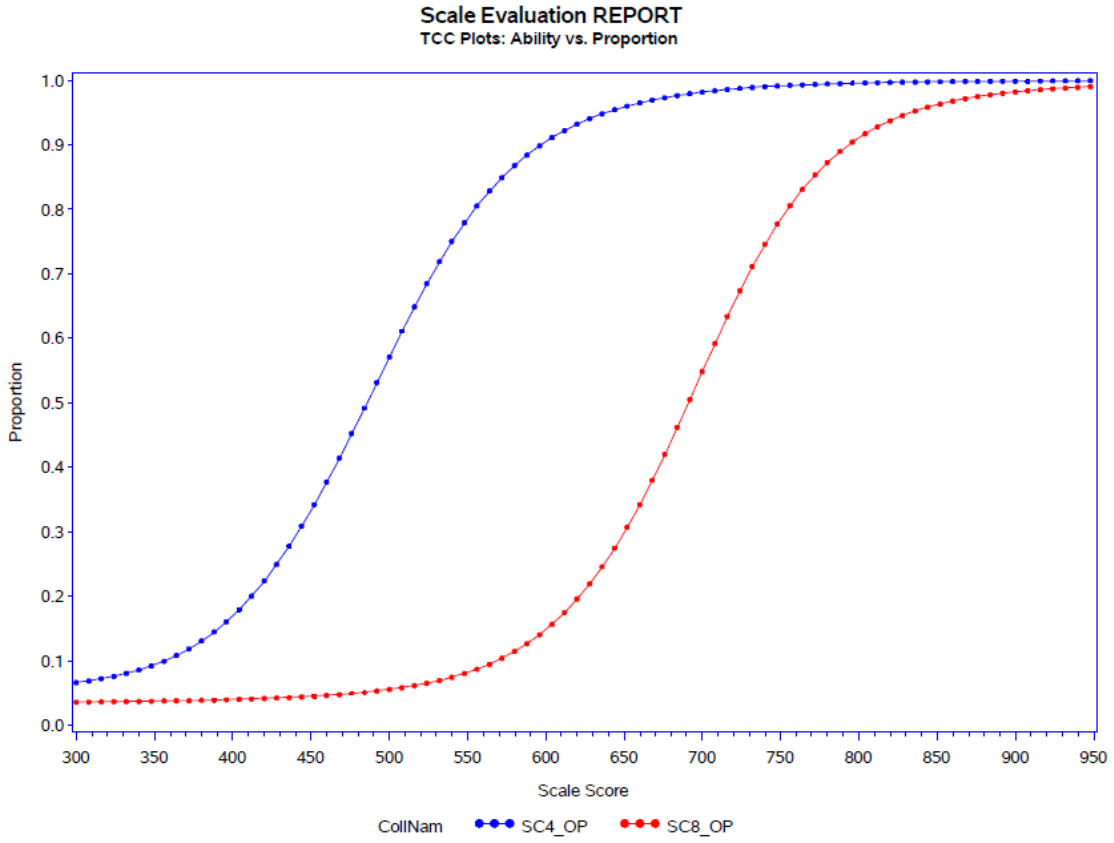


Figure 6-25 Science Standard Error Curves

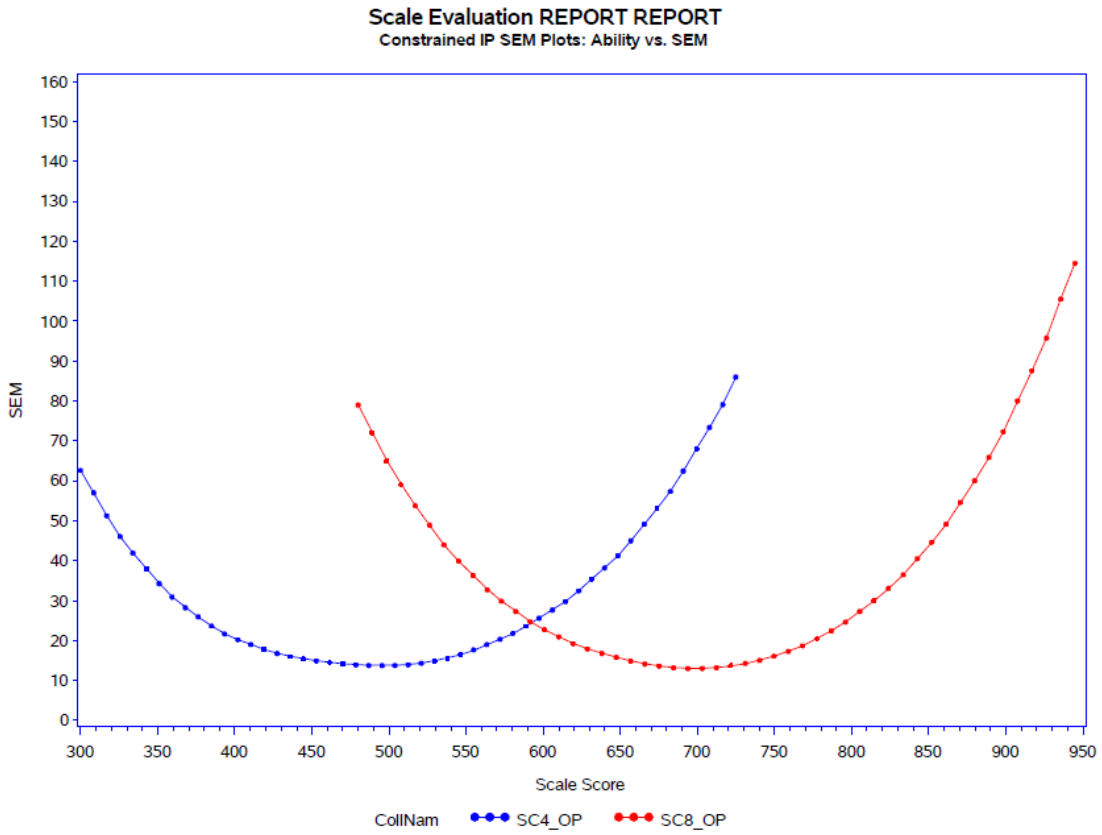


Figure 6-26 Science Growth at Quartiles

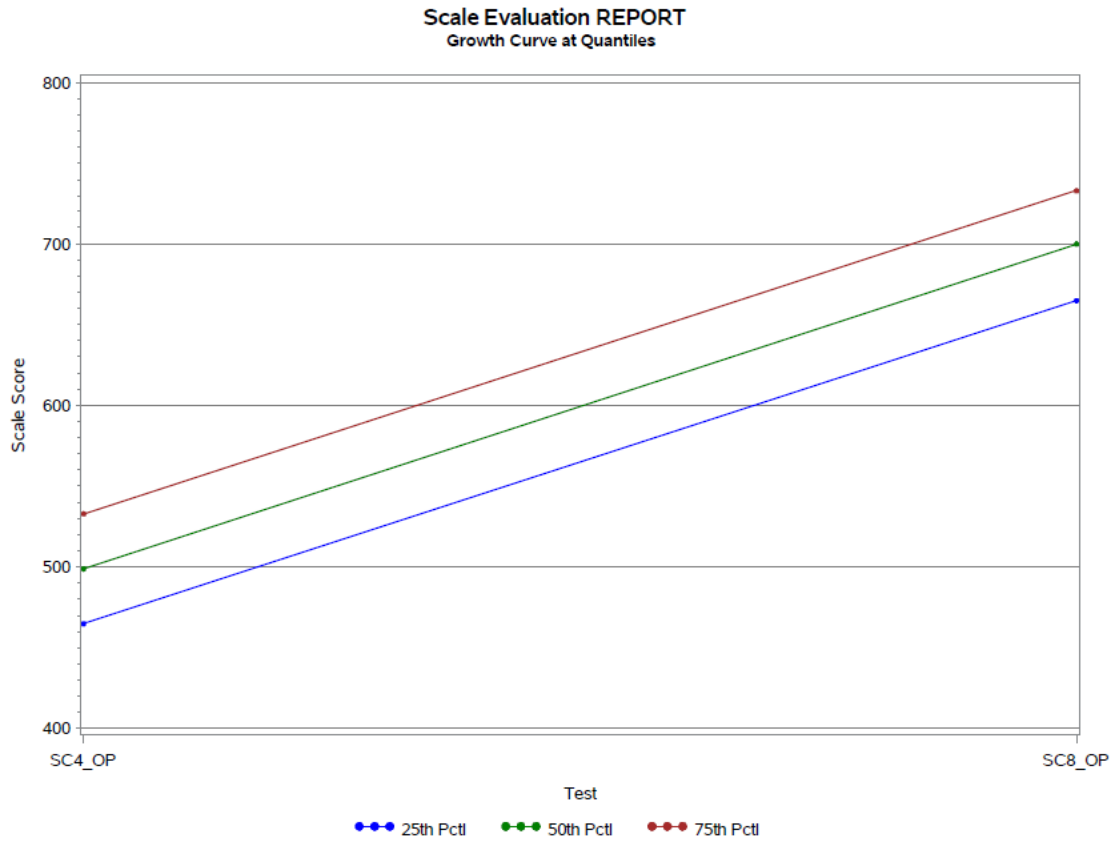


Figure 6-27 Social Studies Test Characteristic Curves

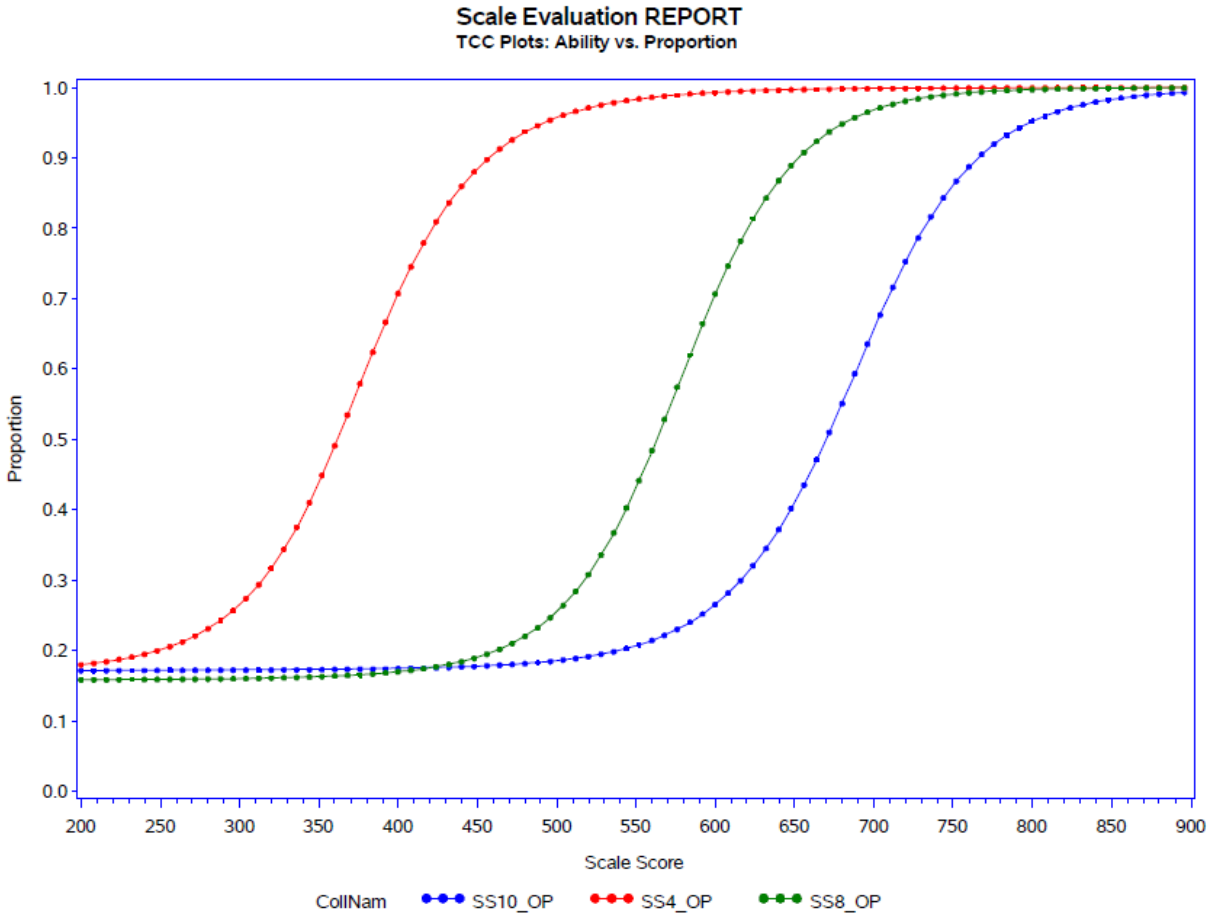


Figure 6-28 Social Studies Standard Error Curves

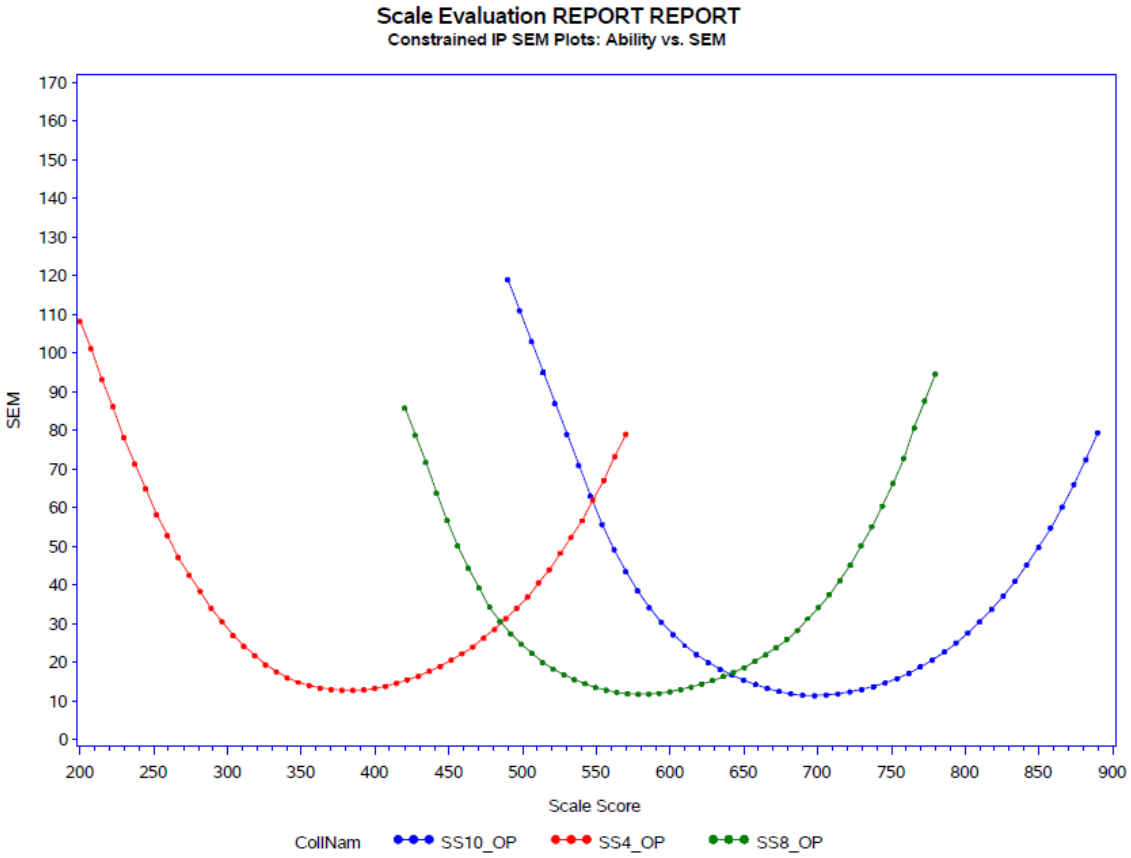
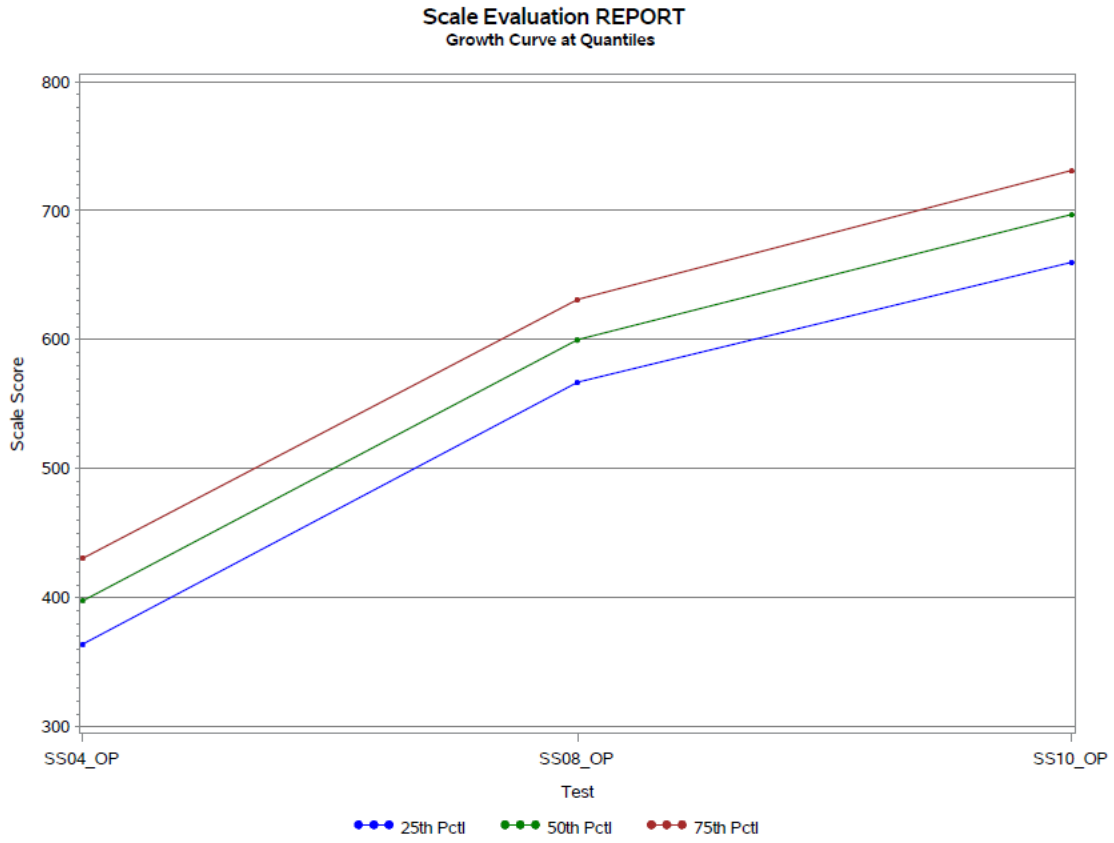


Figure 6-29 Social Studies Growth at Quartiles



Part 7: Standard Setting

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

7.1 Background Information

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

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

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

7.2 Standard Setting Methodology and Process

Prior to the standard setting workshops in Spring 2016 and 2019, DPI worked in collaboration with DRC and its other technical advisors to select the methodology to be used at the standard setting. In recognition of its use in Wisconsin and widespread use across the country, DPI selected the Bookmark Standard Setting Procedure (BSSP) for the Wisconsin Forward Exam. The BSSP was 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

Copyright © 2019 by the Wisconsin Department of Public Instruction

and by other major testing firms, making it the most widely used standard setting procedure in K–12 education (Karantonis & Sireci, 2006; Cizek & Bunch, 2007).

7.2.1 Spring 2016 Standard Setting for All Content Areas

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

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

7.2.2 Spring 2019 Standard Setting for Science

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

A total of 27 Wisconsin educators, 13 for grade 4 and 14 for grade 8, working individually and in grade-specific committees, recommended performance standards associated with four performance levels for the two Science assessments: *Below Basic*, *Basic*, *Proficient*, and *Advanced*. Participants engaged in three rounds of discussions and judgments to make their cut score recommendations. Specifically, the committee performed the following tasks:

1. Participants discussed the state content standards for science and the draft performance level descriptors (PLDs) for their test. The PLDs described the content-based expectations for students in each performance level. Participants refined the PLDs based on their study of the content standards.
2. Participants each examined an ordered item booklet (OIB) which presented test items in order of difficulty. Difficulty was calculated from Wisconsin students' performance.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

After Round 3 of the Bookmark Procedure, participants reviewed their recommendations and associated impact data, as shown in Table 7-2. Educators expressed satisfaction in the content-based judgments they made during the process. However, participants also voiced an expectation that the percentages of students classified in each performance level would be more consistent across grades 4 and 8.

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

The process of both standard settings 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. (p. 107)

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

7.3 Performance Level Descriptors

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

7.4 Cut Scores

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

7.5 Summary

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Level	Performance Level Descriptor
Below Basic	Student demonstrates **minimal understanding** of and ability to apply the knowledge and skills for 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	447	496	543
	8	653	695	737
Social Studies	4	363	396	436
	8	563	599	640
	10	670	703	741

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 8: Test Results

Part 8 presents a classical item analysis and summary of student results for the Spring 2019 Wisconsin Forward Exam. The summary results are presented for all Wisconsin 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 American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (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 or any item with a maximum score of 1 represents the proportion of students who answered the

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

For the Spring 2019 Wisconsin Forward Exam, items were flagged for further investigation in the following situations:

- The *p*-value was less than 0.20. Such a *p*-value indicates a difficult item, where fewer than 20% of students obtained the correct answer.
- The item-total correlation was less than 0.15 for the correct answer. A low value may indicate that the item is not providing a high degree of discrimination between high-ability and low-ability students, and, in addition, it may be an indication that the correct answer is in question.
- A distractor had a positive correlation with the total test score.
- The omit rate was greater than 3%.

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

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

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

Overall, 41 operational items across the ELA, Mathematics, and Science assessments were flagged on the Spring 2019 Wisconsin Forward Exam operational tests as meeting the investigational criteria bulleted above. No items were flagged in Social Studies assessments.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-A shows the number of scored items in the Spring 2019 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 through 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. The majority of items flagged for a positive distractor-total test correlation had a distractor-total test correlation close to 0 and an acceptable item-total test correlation for the correct answer. All flagged items were judged to be acceptable based on their other statistics and were retained in order to meet the Wisconsin Forward Exam test blueprints.

8.1.2 Flagging for the Item-Total Correlation

One item per grade was flagged for item-total test correlation <0.15 in the following assessments: Mathematics grades 4, 6, and 7, and Science grade 8. All of the flagged items had item-total test correlations of at least 0.12.

8.1.3 Flagging for *p*-Value

Six items were flagged for *p*-values <0.20 in Mathematics assessments, and all flagged items had *p*-values between 0.12 and 0.18. While these statistics indicate items that were very difficult, the number of items flagged for difficulty was very small. No operational items were flagged for difficulty in ELA, Science, or Social Studies.

8.1.4 Flagging for Omit Rate

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

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 stipulates that at least 80% of test takers should be able to

Copyright © 2019 by the Wisconsin Department of Public Instruction

answer all of the items and all test takers should be able to answer at least 75% of the items (Swineford, 1956). However, a more stringent requirement is often applied, considering tests to be non-speeded only if at least 95% of examinees attempt the final item. As shown in Table 8-E, the Wisconsin Forward Exam satisfies this more stringent requirement, with approximately 99% or more of the examinees attempting the final item in each of the four content areas.

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 3 percent, and the *p*-value is flagged when it is below 0.20.

Tables 8-1 through 8-17 also show the item numbers, which can be used to understand the location of test items as students actually encountered them on the test. The item analysis tables also indicate item type (e.g., MC, EBSR).

The numbers of flagged items across grade and content areas are summarized in Table 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 reasonably high point-biserial correlations for most items. Detailed item analysis results including distractor statistics for MC items and score point distribution for non-multiple choice items are included in Appendix G.

8.2 Raw Score Results

Raw score and test reliability statistics were computed on the Spring 2019 Wisconsin Forward Exam data for students with available item responses. These statistics are presented in Table 8-18. 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 score points corresponding to each test.

The mean raw score varies by grade and content area and, specifically, in the context of the maximum possible score points. In ELA, for example, the maximum possible raw score is 53 or 56. In Mathematics, the maximum possible raw score is 42 or 46.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

possible score and, therefore, indicates a more difficult test. Consider an example: A test difficulty statistic would be 0.90 if a mean score of 45 were obtained on a test with a maximum possible score of 50. This would be considered an easier test. On the other hand, test difficulty would be 0.50 if a mean raw score of 25 were obtained on the same test. This would then be considered a more difficult test. For example, the Mathematics grade 3 test mean raw score is 23.56 and the maximum possible score is 42, resulting in the test mean *p*-value of approximately 0.56. Note that this computation formula will not apply to ELA results. The mean *p*-value for ELA was computed using unweighted item scores, while the mean raw score was computed with weighted TDA items. (The *p*-value reflects the overall test difficulty to which each test item contributes only once, while the mean raw score reflects the student performance on the test and was computed based on the student test scores that included weighted TDA item scores.)

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 has a long tail on the left side of the distribution because of the presence of some low scores, and, because the mean is sensitive to extreme scores, it indicates that most student scores are clustered on the high end of the scale. A positive skew indicates a distribution with some 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 Standards for ELA, Science, and Social Studies. The Mathematics assessments in grades 4 through 8 tend to be more difficult, however, showing most of the scores clustered below the mean (as indicated by positively skewed score distributions).

In addition, Table 8-18 shows that the minimum obtained scores in eleven out of seventeen tests are zero, meaning that at least one student failed all items for each of those tests. The table also shows that the maximum obtained scores are equal to the maximum number of points possible on the test in all grades, meaning that at least one student obtained the full score for all items on each of those tests. For example, as displayed in Table 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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 0.55 to 0.60.
- Standard deviations ranged from 9.84 to 10.79 raw score points.
- Alpha was relatively high in every grade (0.88 to 0.90).
- SEM ranged from 3.29 to 3.41.

Mathematics

- Test difficulty ranged from 0.44 to 0.56, with generally lower difficulty in lower grades and higher difficulty in higher grades.
- Standard deviations ranged from 9.64 to 10.17 raw score points.
- Alpha was relatively high in every grade (0.91 to 0.93).
- SEM ranged from 2.69 to 2.90.

Science

- Test difficulty was 0.56 in grade 4 and 0.54 in grade 8.
- Standard deviations were 8.14 and 8.60 raw score points for grades 4 and 8, respectively.
- Alpha was 0.89 in grade 4 and 0.90 in grade 8.
- SEM was 2.71 and 2.75 for grades 4 and 8, respectively.

Social Studies

- Test difficulty was 0.66, 0.67, and 0.62 in grades 4, 8, and 10, respectively.
- Standard deviations ranged from 7.86 to 10.68 raw score points.
- Alpha ranged from 0.89 to 0.92.
- SEM ranged from 2.56 to 3.01.

8.2.1 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, use of testing accommodations, and English language proficiency. These subgroup comparisons draw from Tables 8-19 through 8-27 and show some consistent performance patterns by subgroups.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Regarding scores by gender, in ELA, 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 grade 4 to 0.06 in grade 8. In Mathematics, the test difficulties were very similar between male and female students in all grades except grade 4 (differences of 0.1 or 0.0 in test *p*-value). In grade 4, the test was slightly easier for male students than for female students, with a difference of 0.02. In Science, the test difficulties were very similar (0.0) between male and female students in grades 4 and 8. In Social Studies, the differences in test difficulty between genders were 0.01 for grade 4 and 0.02 for grades 8 and 10, with female students performing slightly better than male students in all 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 and Asian students, followed by Hispanic students, American Indian students, and African-American students. American Indian students had similar or slightly lower mean raw scores than Hispanic students. Differences in test difficulty between American Indian and Hispanic students were between 0.0 and 0.04 in all grades and content areas. Differences in test difficulty between the highest achieving group (White students) and the lowest achieving group (African-American students) were between 0.20 and 0.25 across all grades and content areas.

In every grade and content area, the test was easier for students who were not economically disadvantaged than for those who were economically disadvantaged. The difference in test difficulty between the two groups ranged from 0.14 to 0.17 across all grades and content areas, with the largest differences observed in Mathematics.

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 0.15 in Science grade 4 to 0.25 in ELA grade 8. Larger differences in student performance were observed for higher grade levels compared to lower grade levels.

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 0.13 in ELA grades 3 and 4, Mathematics grade 4, and Social Studies grade 4 to 0.23 in Social Studies grade 10. Larger differences in student performance were observed for higher grade levels compared to lower grade levels.

When looking at the test difficulty for students using testing accommodations, it should be noted that only approximately 100 or fewer students per grade used ELA testing accommodations and fewer than 40 students per grade used Science or Social Studies testing accommodations. While it was observed that the tests were more difficult for students using testing accommodations, given small numbers of students using testing accommodations in ELA, Science, and Social Studies, comparisons of the test difficulty for these students with the test difficulty for students not using testing accommodations for the corresponding grades and content areas should be made with caution. The number of students using testing accommodations in Mathematics ranged from 706 in Grade 3 to 2,782 in Grade 6. In all

Copyright © 2019 by the Wisconsin Department of Public Instruction

Mathematics grades, the test was much easier for students not using testing accommodations, with differences in test difficulty ranging from 0.23 to 0.28.

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 8-28 is the summary scale score table based on census data. The table shows the mean scale score, the standard deviation of the scale scores, skewness and kurtosis, the minimum and maximum obtained scale scores, and the lowest and highest obtainable scale scores (LOSS and HOSS, respectively) for all content areas and grades based on the census data. The LOSS and HOSS, as discussed in Part 6, identify the lower and upper limits of the scale score range. These values were established when the current scales were developed and do not change from one administration to another.

English Language Arts

- Mean scale score increased as grade level increased, ranging from 554.59 for grade 3 to 629.06 for grade 8. This mean scale score pattern supports the ELA vertical scale properties.
- Standard deviations ranged from 45.54 to 59.84 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Mathematics

- Mean scale score increased as grade level increased, ranging from 555.78 for grade 3 to 644.53 for grade 8. This mean scale score pattern supports the Mathematics vertical scale properties.
- Standard deviations ranged from 51.78 to 60.69 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 499.88 and 699.70 for grades 4 and 8, respectively.
- Standard deviations were 50.24 and 50.55 scale score points for grades 4 and 8, respectively.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Social Studies

- Mean scale scores were 396.68, 598.81, and 692.82 for grades 4, 8, and 10, respectively.
- Standard deviations ranged from 52.30 to 58.30 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

8.3.1 Subgroup Performance Patterns in Scale Score Results

The scale score results, like the raw score results, showed some consistent performance patterns in terms of subgroups. The results for gender, race/ethnicity, socioeconomic status, disability status, English language proficiency, and accommodation use are presented in Tables 8-29 through 8-37. The scale score statistics were computed based on the census data.

Gender

- In terms of gender, male students as a group showed lower mean scale scores in ELA than female students as a group in each grade level. The difference ranged from 8.83 scale score points in grade 3 to 18.89 scale score points in grade 8.
- In Mathematics, male students as a group showed slightly higher mean scale scores in grades 3 through 5 (differences between 1.13 and 3.75 points) and lower mean scale scores in grades 6 through 8 (differences between 2.35 and 5.41 scale score points) than female students.
- In Science, the mean scale scores were lower for female students in grade 4, with a difference of 2.79 score points. A difference of less than one scale score point was observed in grade 8, with female students performing just slightly better than male students.
- There were small differences between mean scale scores by gender in Social Studies, from 2.25 scale score points in grade 4 to 5.94 scale score points in grade 8. Female students performed better than male students in all grades.

Race/Ethnicity

- The scale score results showed some consistent performance differences by ethnicity.
- In almost every grade and content area, White students as a group had the highest mean scale scores, followed by Asian students, Hispanic students, American Indian students, and African-American students. The two exceptions were ELA grade 8, where Asian students performed as well as White students (difference between the two groups was less than half scale score point), and Mathematics grade 8, where Asian students slightly outperformed White students (difference was 3.49 scale score points)
- The mean scale scores of African-American students were typically more than one standard deviation lower than the mean scale scores of White students. The mean scale scores of Hispanic and American Indian students were approximately two-thirds of a standard deviation lower than the mean scale scores of White students for most

Copyright © 2019 by the Wisconsin Department of Public Instruction

grades and content areas. The mean scale scores of Asian students were typically less than 10 scale score points lower than the mean scale scores of White students, except for ELA grades 3 and 4, Mathematics grade 3, Science grade 4, and Social Studies grade 4, where the differences were larger than 10 scale score points.

- 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.

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 32.15 scale score points in ELA grade 3 to 44.23 scale score points in Mathematics grade 7.
- For every grade and content area, the mean scale scores of students who were economically disadvantaged were typically more than two-thirds of a standard deviation lower than the mean scale scores of students who were not economically disadvantaged.

Disability Status

- Students with disabilities and students without disabilities showed consistent and large differences in mean scale scores by group. Differences ranged from 37.17 scale score points in ELA grade 3 to 67.09 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 at least three quarters of a standard deviation to over one 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 26.07 scale score points in ELA grade 3 to 61.44 scale score points in Social Studies grade 10.
- For every grade and content area, the mean scale scores of limited English proficient students were lower than the mean scale scores of fully English proficient students by about half of a standard deviation to just over one standard deviation.

Accommodation Use

- Students using testing accommodations (listed in section 4.1.3 of this report) performed less well on the tests compared to their peers not using testing accommodations. The differences ranged from 23.01 scale score points for ELA grade 5 to 80.46 points for Mathematics grade 7.
- For ELA and Mathematics, the mean scale scores of students using testing accommodations were lower than the mean scale scores of students not using testing accommodations by about half of a standard deviation to over one standard deviation.

Copyright © 2019 by the Wisconsin Department of Public Instruction

- Science and Social Studies student performance for students using testing accommodations was not compared with the performance of their peers not using testing accommodations. The differences in mean scale scores for Science and Social Studies between the two groups of students should be interpreted with caution due to fewer than 50 students per grade using testing accommodations in these content areas.

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 (PLDs) are established in a collaborative and participatory process. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, particularly cut scores. PLDs summarize the knowledge, skills, and abilities expected of students in each achievement level. As stated in Part 7, DPI provided policy PLDs for the Wisconsin Forward Exam assessments. At the standard setting, 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 8-38 shows the cut scores for each content and grade level. For ease of reference, Tables 8-39 through 8-42 provide the scale score ranges that define performance levels together with the percentage of students in each performance level. The results for each content area and grade are summarized below.

English Language Arts

- Between approximately 37% (grade 8) and 45% (grade 7) of students were either *Proficient* or *Advanced* in ELA.
- Between 5% and 10% of students were classified as *Advanced*, depending on the grade level.
- Across all grade levels, more than 50% of students were below *Proficient*. These percentages ranged from approximately 55% below *Proficient* in grade 7 to 63% below *Proficient* in grade 8.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Mathematics

- Between approximately 36% (grade 8) and 49% (grade 3) of students were either *Proficient* or *Advanced* in Mathematics.
- The proportion of students who were *Advanced* was between approximately 5% (grade 7) and 12% (grades 3 and 4).
- Across all grade levels, 50% or more students were below *Proficient*. These percentages ranged from approximately 51% below *Proficient* in grade 3 to 64% below *Proficient* in grade 8.

Science

- Approximately 53% of students were either *Proficient* or *Advanced* in grade 4 and approximately 54% of students were either *Proficient* or *Advanced* in grade 8.
- The percentage of students classified as *Advanced* was approximately 19% in grade 4 and just above 22% in grade 8.
- The proportion of students classified as below *Proficient* was approximately 47% in grade 4 and 46% in grade 8.

Social Studies

- More than half of students in grades 4 and 8 were either *Proficient* or *Advanced* in Social Studies. The percentage of *Proficient* or *Advanced* students was approximately 52% in these grades. Close to 46% of students in grade 10 were classified as either *Proficient* or *Advanced*.
- Between 19% and 22% of students were *Advanced* in all three grades.
- The percentage of students classified as below *Proficient* was approximately 48% in grades 4 and 8, and was close to 55% 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 8-43 through 8-46.

In terms of gender, the percentages of both genders were generally similar in *Proficient* or above performance levels for Mathematics, Science, and Social Studies across all grades. The differences in the percentages of male and female students in *Proficient* or above categories for these content areas were, on average, less than 5%. For ELA, more female students than male students were classified as *Proficient* or above (with the differences between genders ranging from approximately 8% to 11%) across all grades.

There were some consistent patterns in performance by ethnicity across grades and content areas. In terms of the *Proficient* or above categories, the prevailing tendency was that there were higher percentages of White students as a group, followed by Asian students,

Copyright © 2019 by the Wisconsin Department of Public Instruction

American Indian students and Hispanic students, and African-American students. The inverse sequence was found at the *Below Basic* performance level.

Performance level results showed that there were higher percentages of fully English proficient students who were classified as *Proficient* or above compared to students who were limited English proficient. In every grade and content area, there were higher percentages of students who were fully English proficient classified as *Proficient* or *Advanced* compared to students with limited English proficiency, with the differences ranging from approximately 25% to over 40%, depending on grade level and content area. There were much lower percentages of fully English proficient students who were classified in the lowest performance level in all grades and content areas.

Performance level results showed a similar pattern in comparisons for students without disabilities who were classified as *Proficient* or above compared to students with disabilities, with the differences ranging from approximately 28% to 40%, depending on grade level and content area. There were much higher percentages of students without disabilities in the reporting category *Advanced.* There were also much lower percentages of students without disabilities in the lowest performance level than students with disabilities. This pattern was evident in all grades and all content areas.

There were consistent differences in performance between economically disadvantaged students and not economically disadvantaged students. In every grade and content area, between approximately 26% and 32% more students who were not economically disadvantaged were classified as *Proficient* or above compared to their economically disadvantaged peers. There were much higher percentages of students who were economically disadvantaged who were classified in the lowest performance category.

Performance level results showed that there were higher percentages of students not using testing accommodations who were classified as *Proficient* or above compared to students using testing accommodations. The differences ranged from approximately 10% to 36% for ELA, Science, and Social Studies, depending on the grade level. For Mathematics, these differences ranged from approximately 36% to 44%, across all grade levels. The differences in the percentages of students in different performance levels between groups of students using and not using testing accommodations should be interpreted with caution for ELA, Science, and Social Studies due to a very low number of students using testing accommodations in these 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, standards, or objectives. Scores on subsets of test items at the content standard level are called standard performance

Copyright © 2019 by the Wisconsin Department of Public Instruction

index (SPI) scores. The purpose of reporting SPI scores on the Wisconsin Forward Exam is to show the relationship between the overall achievement being measured (represented by the test score) and the skills within each of the content standards associated with the overall content area. Teachers may use the SPI scores for individual students as indicators of strengths and weaknesses, but the SPI scores are best corroborated by other evidence, such as homework, class participation, diagnostic test scores, or observation. District and school administrators may compare their results by content standard and grade level with the state mean percentage to better understand their strengths and weaknesses within a particular content area and grade level.

An SPI score can be interpreted as an estimate of the number of items a student would be expected to answer correctly if there had been 100 similar items for a given reporting category. For example, an SPI score of 77 for a given reporting category means that, if the student were given 100 similar items, the student would be expected to answer 77 of them correctly. This is a criterion-referenced score because it estimates how much a student knows in a clearly defined skill domain (i.e., the criterion). Technical readers can refer to Appendix H of this report for more details.

This approach—identifying student proficiency on each content standard—relates to the ELA and Mathematics Wisconsin Academic Standards and Wisconsin's Model Academic Standards for Science and Social Studies. SPI scores provide a more reliable estimate of student achievement on each content standard than is possible by simply reporting the percentage correct. However, *SPI scores should be used for low-stakes purposes because these scores cannot be considered stable for any content standard with a small number of items*.

Readers should note that the average difficulty of items will vary across content standards and grades. Content standards vary in their complexity, level of abstraction, and cognitive demand. Some standards may be intrinsically more difficult than others, and the difficulty of individual items is determined, in part, by the difficulty of the content domain being measured. The current test blueprints do not specify the average difficulty level of items for each content standard within grades or across grades. If the difficulty of the items varies across years, grades, or content standards, the mean SPI scores will be affected by differences in item difficulty as well as differences in student ability. *Thus, differences in SPI scores across years, grades, or content standards should not be seen as reliable indicators of differences in student ability, since these differences may be explained in whole or in part by differences in the difficulty of the items themselves.* 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 8-47 through 8-50 identify the content standards or domains, the number of MC and CR items within each standard or domain, the total number of possible points per standard or domain, the mean raw score, the mean *p*-value, the standard deviation of the raw scores, the mean SPI score, and the standard deviation of SPI scores for all content areas across grades. The results from Tables 8-47 through 8-50 are summarized below. Tables 8-51 through 8-54 identify the SPI cut scores for each content area reporting category and grade level.

Copyright © 2019 by the Wisconsin Department of Public Instruction

English Language Arts

Tables 8-47a and 8-47b present mean *p*-values and SPI scores for ELA across content standards or domains and grades. Results show that the mean ELA SPI scores across grades ranged from 42.00 to 69.03 for content standards and from 48.70 to 63.16 for domains, indicating that the items were moderately difficult to easy for examinees. In general, content standard D (Writing/Language—Text Types and Purposes) was the most difficult in all grades. These content standards contained the TDA item, which was generally difficult for students. Content standard C (Reading—Vocabulary Use) was the easiest in grade 7. The Writing/Language—Language Convention standard (F) was the easiest in grades 4, 5, and 8. The Listening domain was easier than other domains for students in grades 3 and 5 through 8. The Writing domain was the most difficult domain for students in all grades except for grade 4 students, for whom the Listening domain was more difficult than other domains.

Mathematics

Table 8-48 presents Mathematics *p*-values and SPI scores across grades and content standards. Results show that the mean *p*-values and SPI scores varied across standards in all grades. Mean SPI scores, across all content standards, ranged from 37.64 (Geometry in grade 7) to 61.54 (Number and Operations in Base Ten in grade 3). The Mathematics items were more challenging in higher grades than lower grades. There was no consistent pattern in regard to the content standard difficulty across grade levels except for grades 3 and 4, where content standards C (Number and Operations—Fractions) and A (Operations and Algebraic Thinking) were the most difficult in both grades and content standards E (Geometry) and B (Number and Operations in Base Ten) were the easiest in both grades.

Content standard D (Measurement and Data) was the most difficult in grade 5, and content standard E (Geometry) was the most difficult in grades 6 and 7. Content standard G (The Number System) was the most difficult in grade 8.

Science

Table 8-49 presents Science *p*-values and SPI scores across grades and content standards. The mean Science SPI scores across both grades and all content standards ranged from 46.78 to 63.73, indicating that the test items were of medium difficulty. SPI scores indicated that content standard C (Earth and Space Science) was the most difficult in grade 4 and content standard B (Physical Science) was more difficult than other content standards in grade 8.

Social Studies

Table 8-50 presents Social Studies *p*-values and SPI scores across grades and content standards. The mean Social Studies SPI scores across all grades and content standards ranged from 57.37 to 73.23, indicating that the test items ranged from somewhat difficult to relatively easy. The mean SPI scores indicated that the most difficult content standard varied between the three Social Studies grades. In grades 4 and 8, the most difficult standard was content standard D

Copyright © 2019 by the Wisconsin Department of Public Instruction

(Economics). In grade 10, the most difficult standard was content standard C (Political Science and Citizenship).

Summary of Student Performance Indicator Results

Overall, the mean SPI scores across grades and content standards ranged in difficulty. The content standards with SPI mean scores of >70 were the following:

- Grade 4 Social Studies content standards B (History)
- Grade 4 Social Studies content standards E (The Behavioral Sciences)
- Grade 8 Social Studies content standard A (Geography)

There were no SPI mean scores of <30 in the Wisconsin Spring 2019 test administration.

It is important to note that some variation in difficulty of the items across content standards within and across grades and test forms is inevitable and that some of that variation is independent of any intrinsic differences in the difficulty of the standards themselves (e.g., variations in the difficulty of the particular items that were selected for the test forms). For this reason, SPI scores should be interpreted with caution and should not be used to make comparisons of student performance across testing years or grade levels.

8.6 Longitudinal Comparisons of Test Scores

It is often desirable to examine the scores of students across time and monitor group performance. This is possible if the test content and the construct measured by the test are comparable from year to year and if the scores are reported on the same scale in multiple years.

For the Wisconsin Forward Exam assessments, four years of test scores on the same reporting scales are available, and the state-level mean scale scores and standard deviations for the 2016, 2017, 2018, and 2019 administrations are presented for ELA, Mathematics, and Social Studies in Tables 8-55, 8-56, and 8-58, respectively. New scales were established for the Science assessments after the Spring 2019 test administration. Because the new Science assessments were not linked to the previous scales, the Spring 2019 scale scores are not comparable with the previous administration scores. Therefore, only one year of scale score data is presented for Science in Table 8-57. The Spring 2019 student performance in Science is a new baseline for longitudinal comparisons. The statistics presented in Tables 8-55 through 8-58 are based on the total population of Wisconsin students, including students attending public, choice, and private schools.

It was observed that the mean scale score for ELA increased by approximately 1 scale score point for grade 4. The mean scale scores decreased for grade 3 (by approximately 2 points), grade 5 (by approximately 5 points), grade 6 (by less than 3 points), and grade 8 (by approximately 2 points). The year-to-year difference was less than half of 1 scale score point, showing no practical change for grade 7.

Copyright © 2019 by the Wisconsin Department of Public Instruction

For Mathematics, the mean scale score increased for grade 5 (by less than 3 points) and for grade 7 (by approximately 2 scale score points) between the Spring 2018 and Spring 2019 administrations. The mean scale score decreased by approximately 1 point for Mathematics grade 6. For grades 3, 4, and 8, the year-to-year differences were less than half of 1 scale score point, showing no practical change.

For Social Studies, the mean scale score decreased by less than 2 scale score points between the 2018 and 2019 test administrations for grades 4 and by less than 3 scale score points for grade 10. No practical change in the scale score mean was observed between the last two years for grade 8 (a difference of less than half a point).

Tables 8-59 through 8-62 show the percentages of students in each achievement level in the Spring 2016, 2017, 2018, and 2019 test administrations for ELA, Mathematics, Science, and Social Studies. The results presented in these tables are based on the total population of Wisconsin students, including students attending public, choice, and private schools.

For ELA, a decrease in the percentage of students at or above *Proficient* was observed for grades 3 through 6, ranging from approximately 1% for grades 3 and 4 to 4% for grade 5. There was no practical change in the percentage of students at or above *Proficient* for grades 7 and 8 (differences of less than half of 1 percent) between Spring 2018 and Spring 2019.

For Mathematics, a small decrease in the percentage of students at or above *Proficient* was observed for grade 6 (a difference of approximately 1.5%). For all other grades, the change in the percentage of students at or above *Proficient* was less than 1% (either increase or decrease).

As stated earlier in the report, new performance level cut scores were established for Science after the Spring 2019 test administration. Therefore, the Spring 2019 Science impact data should not be directly compared with the previous years' impact data. The Spring 2019 impact data in Science is a new baseline for longitudinal comparisons. The Spring 2019 impact data are separated from the previous years' impact data with a grey bar in Table 8-61. The historical data are included in Table 8-61 for illustration of student performance on the past Science assessments.

For Social Studies, a decrease of approximately 1% in the percentage of students at or above *Proficient* was observed for grade 4 and a decrease of approximately 3% in the percentage of students at or above *Proficient* was observed for grade 10. There was no practical change in the percentage of students at or above *Proficient* for grade 8 (a difference of less than 1% between the last two years).

Overall, the percentages of students classified in each of the four performance level categories were found to be comparable between the Spring 2018 and 2019 test administrations across all grade levels for ELA, Mathematics, and Social Studies. With the exception of ELA grade 5 (*Below Basic* and *Proficient*) and Social Studies grade 10 (*Below Basic*), the change between the percentage of students in Spring 2018 and in Spring 2019 in any performance level category, grade, or content area was less than 2%.

Copyright © 2019 by the Wisconsin Department of Public Instruction

8.7 Summary

In the Wisconsin Forward Exam, the purpose of the ELA, Mathematics, Science, and Social Studies assessments is to demonstrate student achievement through test scores in the respective content areas. The results presented in Part 8, together with the reliability and validity evidence presented in Parts 9 and 10, 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. Therefore, test scores and performance levels 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, including the longitudinal test results, can be drawn upon for accountability and reporting purposes.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Content	Grade	# of Items Flagged	Number of Flags			
			Correlation <0.15	Distractor Correlation >0	Omit >3%	*p*-Value <0.20
ELA	3	2	0	2	0	0
	4	2	0	2	0	0
	5	2	0	2	0	0
	6	2	0	2	0	0
	7	1	0	1	0	0
	8	2	0	2	0	0
MA	3	2	0	2	0	0
	4	5	1	3	0	1
	5	5	0	4	0	1
	6	7	1	4	0	2
	7	5	1	3	0	1
	8	4	0	3	0	1
SC	4	1	0	1	0	0
	8	1	1	0	0	0
SS	4	0	0	0	0	0
	8	0	0	0	0	0
	10	0	0	0	0	0
Total		**41**	4	31	0	6

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Content	Item	Item Type	*p*-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	*p*-Value
3	ELA	19	MC	0.44	0.32	0.20		+	0.00		
3	ELA	24	MC	0.56	0.21	0.26		+	0.03		
4	ELA	10	MC	0.56	0.19	0.16		+	0.04		
4	ELA	29	MC	0.41	0.22	0.29		+	0.06		
5	ELA	21	MC	0.59	0.17	0.19		+	0.03		
5	ELA	27	MC	0.54	0.20	0.44		+	0.03		
6	ELA	5	MC	0.36	0.20	0.18		+	0.03		
6	ELA	14	MC	0.56	0.33	0.10		+	0.02		
7	ELA	27	MC	0.38	0.25	0.31		+	0.01		
8	ELA	21	MC	0.45	0.17	0.27		+	0.01		
8	ELA	35	MC	0.38	0.22	0.33		+	0.07		

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Content	Item	Item Type	*p*-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	*p*-Value
3	MA	7	MC	0.44	0.28	0.60		+	0.02		
3	MA	16	MC	0.33	0.39	0.18		+	0.04		
4	MA	15	MC	0.28	0.15	0.34		+	0.04		
4	MA	22	MC	0.29	0.17	0.17		+	0.12		
4	MA	35	SA	0.18	0.46	0.16					+
4	MA	42	MC	0.28	0.13	0.15	+	+	0.10		
5	MA	3	MC	0.57	0.18	0.15		+	0.02		
5	MA	11	MC	0.33	0.24	0.17		+	0.04		
5	MA	33	MC	0.40	0.28	0.15		+	0.12		
5	MA	41	MC	0.13	0.28	0.22					+
5	MA	45	MC	0.27	0.17	0.18		+	0.09		
6	MA	1	MC	0.39	0.35	0.14		+	0.04		
6	MA	9	SA	0.18	0.50	0.18					+
6	MA	25	MC	0.32	0.34	0.20		+	0.02		
6	MA	28	MC	0.38	0.16	0.39		+	0.04		
6	MA	32	MC	0.33	0.12	0.28	+				
6	MA	43	MC	0.33	0.22	0.24		+	0.03		
6	MA	44	TE	0.14	0.17	0.27					+
7	MA	9	MC	0.31	0.25	0.14		+	0.07		
7	MA	17	MC	0.22	0.14	0.29	+	+	0.00		
7	MA	33	TE	0.12	0.52	0.41					+
7	MA	40	MC	0.30	0.22	0.59		+	0.04		
8	MA	2	MC	0.24	0.16	0.12		+	0.01		
8	MA	16	MC	0.47	0.26	0.35		+	0.04		
8	MA	27	SA	0.17	0.44	1.10					+
8	MA	34	MC	0.37	0.22	0.50		+	0.02		

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Content	Item	Item Type	*p*-Value	Corr	Percent Omit	Flags				
							Corr	Distractor		Omit	*p*-Value
4	SC	35	MC	0.37	0.23	0.11		+	0.01		
8	SC	24	TE	0.50	0.12	0.15	+				

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

Content	Grade						
	3	4	5	6	7	8	10
English Language Arts	99.68	99.76	99.77	99.79	99.68	99.71	
Mathematics	99.83	99.79	99.78	99.74	99.40	99.62	
Science		99.90				99.82	
Social Studies		99.94				99.84	99.21

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.37	0.60	0.36				
2	MC	0.67	0.31	0.06				
3	TE	0.71	0.39	0.47				
4	MC	0.58	0.40	0.16				
5	MC	0.41	0.27	0.17				
6	TE	0.56	0.43	0.27				
7	MC	0.62	0.42	0.18				
8	MC	0.71	0.44	0.19				
9	MC	0.61	0.39	0.15				
10	TE	0.35	0.37	0.59				
11	MC	0.60	0.46	0.19				
12	MC	0.45	0.36	0.17				
13	MC	0.71	0.54	0.17				
14	TE	0.65	0.53	0.20				
15	MC	0.71	0.36	0.09				
16	TE	0.64	0.52	0.17				
17	MC	0.55	0.35	0.15				
18	MS	0.72	0.53	0.20				
19	MC	0.44	0.32	0.20		+		
20	MC	0.54	0.21	0.36				
21	MC	0.66	0.39	0.21				
22	MC	0.44	0.27	0.23				
23	MC	0.52	0.29	0.27				
24	MC	0.56	0.21	0.26		+		
25	TE	0.70	0.62	0.24				
26	MC	0.55	0.32	0.32				
27	MS	0.57	0.47	0.47				
28	MC	0.59	0.45	0.73				
29	MC	0.73	0.44	0.22				
30	MC	0.26	0.25	0.24				
31	EBSR	0.46	0.44	0.16				
32	MC	0.45	0.34	0.40				
33	MC	0.43	0.36	0.40				
34	MC	0.61	0.50	0.30				
35	MC	0.46	0.31	0.31				
36	MC	0.44	0.30	0.29				
37	TE	0.52	0.47	0.32				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.34	0.52	0.33				
2	TE	0.56	0.46	0.40				
3	TE	0.63	0.40	0.18				
4	MC	0.68	0.32	0.13				
5	TE	0.45	0.41	0.37				
6	TE	0.64	0.45	0.34				
7	MC	0.58	0.45	0.19				
8	MC	0.56	0.37	0.18				
9	MC	0.55	0.39	0.14				
10	MC	0.56	0.19	0.16		+		
11	TE	0.70	0.40	0.54				
12	MC	0.68	0.46	0.16				
13	TE	0.78	0.34	0.14				
14	MC	0.41	0.21	0.17				
15	MC	0.57	0.27	0.06				
16	MC	0.76	0.34	0.13				
17	EBSR	0.32	0.32	0.05				
18	MC	0.56	0.38	0.10				
19	MC	0.64	0.38	0.14				
20	EBSR	0.36	0.33	0.07				
21	MC	0.73	0.44	0.15				
22	TE	0.57	0.39	0.35				
23	MC	0.53	0.40	0.21				
24	MC	0.59	0.51	0.20				
25	MS	0.47	0.44	0.19				
26	EBSR	0.58	0.51	0.08				
27	MC	0.60	0.33	0.31				
28	MC	0.68	0.48	0.43				
29	MC	0.41	0.22	0.29		+		
30	MC	0.50	0.36	0.24				
31	MC	0.53	0.38	0.64				
32	TE	0.47	0.45	0.56				
33	MC	0.60	0.40	0.28				
34	MC	0.61	0.49	0.27				
35	MC	0.56	0.38	0.26				
36	MC	0.64	0.37	0.27				
37	MC	0.58	0.44	0.24				
38	TE	0.47	0.43	0.23				
39	MC	0.61	0.43	0.24				

.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.37	0.52	0.19				
2	MC	0.80	0.39	0.03				
3	MC	0.68	0.40	0.08				
4	TE	0.59	0.43	0.09				
5	MC	0.41	0.26	0.11				
6	TE	0.54	0.50	0.22				
7	MC	0.66	0.34	0.11				
8	TE	0.72	0.53	0.19				
9	MC	0.64	0.43	0.10				
10	MC	0.53	0.36	0.13				
11	TE	0.71	0.38	0.10				
12	MC	0.59	0.33	0.13				
13	TE	0.58	0.34	0.16				
14	TE	0.62	0.40	0.03				
15	MC	0.64	0.35	0.11				
16	MC	0.78	0.43	0.10				
17	MC	0.59	0.41	0.08				
18	MC	0.59	0.37	0.13				
19	EBSR	0.57	0.50	0.03				
20	MC	0.73	0.48	0.12				
21	MC	0.59	0.17	0.19		+		
22	MC	0.69	0.42	0.22				
23	TE	0.86	0.46	0.29				
24	MC	0.55	0.33	0.24				
25	TE	0.69	0.54	0.32				
26	MC	0.60	0.44	0.21				
27	MC	0.54	0.20	0.44		+		
28	MC	0.51	0.25	0.25				
29	MS	0.60	0.51	0.29				
30	MC	0.61	0.38	0.24				
31	MC	0.56	0.49	0.21				
32	MC	0.65	0.54	0.42				
33	MC	0.57	0.46	0.59				
34	MC	0.44	0.29	0.32				
35	MC	0.60	0.44	0.32				
36	TE	0.62	0.55	0.21				
37	TE	0.65	0.25	0.41				
38	MC	0.43	0.45	0.25				
39	MC	0.46	0.33	0.25				
40	MC	0.58	0.44	0.23				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.38	0.53	0.16				
2	MC	0.64	0.22	0.07				
3	MC	0.66	0.26	0.14				
4	MC	0.71	0.45	0.16				
5	MC	0.36	0.20	0.18		+		
6	TE	0.42	0.23	0.24				
7	TE	0.68	0.41	0.17				
8	MC	0.44	0.31	0.14				
9	TE	0.59	0.51	0.19				
10	TE	0.65	0.35	0.23				
11	TE	0.63	0.42	0.19				
12	MC	0.48	0.28	0.18				
13	MC	0.46	0.33	0.16				
14	MC	0.56	0.33	0.10		+		
15	MC	0.56	0.31	0.18				
16	TE	0.59	0.38	0.18				
17	EBSR	0.74	0.52	0.07				
18	MC	0.61	0.39	0.18				
19	MC	0.70	0.33	0.16				
20	MC	0.45	0.20	0.11				
21	TE	0.66	0.45	0.19				
22	MC	0.51	0.32	0.24				
23	TE	0.59	0.44	0.29				
24	MC	0.55	0.32	0.30				
25	MC	0.80	0.49	0.32				
26	TE	0.58	0.41	0.25				
27	MC	0.62	0.45	0.43				
28	TE	0.63	0.40	0.30				
29	TE	0.67	0.41	0.41				
30	MC	0.43	0.26	0.25				
31	MC	0.53	0.43	0.44				
32	TE	0.56	0.22	0.28				
33	EBSR	0.52	0.60	0.17				
34	TE	0.59	0.45	0.37				
35	MC	0.45	0.39	0.26				
36	MC	0.58	0.41	0.25				
37	MC	0.64	0.50	0.21				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.40	0.53	0.18				
2	MC	0.71	0.41	0.07				
3	TE	0.52	0.38	0.08				
4	TE	0.70	0.46	0.15				
5	MC	0.73	0.37	0.11				
6	TE	0.65	0.30	0.19				
7	MC	0.61	0.46	0.17				
8	MC	0.71	0.35	0.11				
9	MC	0.63	0.37	0.11				
10	TE	0.47	0.23	0.16				
11	MC	0.47	0.32	0.19				
12	MC	0.68	0.36	0.13				
13	MC	0.66	0.45	0.14				
14	EBSR	0.75	0.43	0.03				
15	MS	0.77	0.48	0.21				
16	MC	0.59	0.36	0.16				
17	MC	0.56	0.30	0.16				
18	EBSR	0.38	0.41	0.05				
19	TE	0.73	0.34	0.43				
20	MC	0.61	0.52	0.20				
21	EBSR	0.51	0.38	0.08				
22	MC	0.51	0.36	0.22				
23	MC	0.54	0.44	0.32				
24	TE	0.61	0.43	1.47				
25	TE	0.67	0.54	0.41				
26	TE	0.53	0.33	0.44				
27	MC	0.38	0.25	0.31		+		
28	EBSR	0.44	0.43	0.18				
29	TE	0.34	0.44	1.17				
30	MC	0.71	0.55	0.32				
31	MC	0.44	0.32	0.34				
32	MC	0.54	0.37	0.35				
33	MC	0.48	0.45	0.38				
34	MS	0.59	0.48	0.34				
35	MS	0.72	0.45	0.30				
36	MC	0.57	0.42	0.32				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TDA	0.42	0.57	0.17				
2	MC	0.74	0.39	0.05				
3	MC	0.62	0.45	0.12				
4	MC	0.69	0.45	0.14				
5	MC	0.51	0.33	0.12				
6	MC	0.52	0.25	0.17				
7	TE	0.69	0.47	0.21				
8	MC	0.55	0.52	0.14				
9	MC	0.80	0.33	0.12				
10	TE	0.48	0.30	0.19				
11	MC	0.45	0.31	0.20				
12	TE	0.86	0.37	0.14				
13	MC	0.57	0.45	0.15				
14	MC	0.52	0.37	0.15				
15	MS	0.67	0.45	0.07				
16	EBSR	0.69	0.49	0.05				
17	MC	0.70	0.47	0.14				
18	MC	0.49	0.37	0.15				
19	EBSR	0.44	0.39	0.09				
20	MC	0.40	0.34	0.19				
21	MC	0.45	0.17	0.27		+		
22	MC	0.71	0.43	0.24				
23	MC	0.54	0.33	0.30				
24	MS	0.62	0.46	0.28				
25	MC	0.41	0.44	0.31				
26	MC	0.60	0.47	0.23				
27	MC	0.50	0.27	0.30				
28	MC	0.74	0.54	0.38				
29	MC	0.45	0.22	0.42				
30	EBSR	0.60	0.52	0.16				
31	MC	0.53	0.38	0.28				
32	EBSR	0.56	0.62	0.20				
33	MC	0.76	0.52	0.30				
34	MC	0.67	0.43	0.32				
35	MC	0.38	0.22	0.33		+		
36	MC	0.68	0.47	0.30				
37	MC	0.53	0.40	0.31				
38	MS	0.66	0.59	0.28				
39	MC	0.59	0.40	0.29				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-7 Item Analysis, Grade 3 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TE	0.71	0.52	0.10				
2	SA	0.58	0.53	0.10				
3	TE	0.45	0.59	0.08				
4	MC	0.76	0.45	0.12				
5	MC	0.53	0.46	0.13				
6	SA	0.81	0.39	0.46				
7	MC	0.44	0.28	0.60		+		
8	SA	0.46	0.56	0.17				
9	MC	0.67	0.39	0.17				
10	SA	0.64	0.59	0.15				
11	MC	0.41	0.35	0.14				
12	TE	0.22	0.39	0.59				
13	MC	0.62	0.42	0.43				
14	SA	0.66	0.51	0.74				
15	SA	0.34	0.51	0.24				
16	MC	0.33	0.39	0.18		+		
17	MC	0.79	0.48	0.17				
18	TE	0.86	0.39	0.99				
19	MC	0.71	0.50	0.16				
20	MC	0.59	0.41	0.22				
21	SA	0.60	0.52	0.20				
22	MC	0.47	0.36	0.16				
23	MC	0.77	0.49	0.10				
24	SA	0.61	0.57	0.14				
25	TE	0.62	0.34	0.13				
26	TE	0.48	0.56	0.26				
27	MC	0.61	0.44	0.73				
28	MC	0.42	0.35	0.84				
29	MC	0.61	0.39	0.15				
30	MC	0.41	0.22	0.19				
31	MC	0.44	0.34	0.20				
32	MC	0.42	0.36	0.21				
33	SA	0.51	0.57	0.18				
34	MC	0.67	0.45	0.46				
35	SA	0.62	0.63	0.56				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	TE	0.25	0.48	0.56				
37	MC	0.48	0.46	0.24				
38	MC	0.71	0.37	0.17				
39	SA	0.59	0.61	0.16				
40	MC	0.67	0.48	0.13				
41	TE	0.35	0.47	1.57				
42	MC	0.72	0.45	0.17				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-8 Item Analysis, Grade 4 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.68	0.45	0.23				
2	MC	0.45	0.30	0.08				
3	SA	0.42	0.38	0.09				
4	MC	0.84	0.40	0.11				
5	TE	0.41	0.50	0.15				
6	MC	0.56	0.48	0.11				
7	MC	0.34	0.49	0.11				
8	SA	0.37	0.60	0.34				
9	MC	0.71	0.46	0.10				
10	MC	0.87	0.40	0.10				
11	MC	0.31	0.38	0.15				
12	MC	0.37	0.33	0.13				
13	MC	0.50	0.58	0.14				
14	MC	0.42	0.28	0.14				
15	MC	0.28	0.15	0.34		+		
16	SA	0.34	0.51	0.38				
17	MC	0.80	0.46	0.12				
18	MC	0.43	0.46	0.17				
19	TE	0.35	0.59	0.12				
20	MC	0.45	0.39	0.15				
21	MC	0.88	0.32	0.11				
22	MC	0.29	0.17	0.17		+		
23	SA	0.22	0.54	0.16				
24	TE	0.70	0.36	0.81				
25	MC	0.86	0.34	0.11				
26	MC	0.50	0.50	0.13				
27	MC	0.34	0.56	0.16				
28	MC	0.51	0.40	0.11				
29	SA	0.38	0.52	0.20				
30	MC	0.32	0.42	0.17				
31	MC	0.40	0.20	0.30				
32	TE	0.24	0.49	0.38				
33	MC	0.52	0.58	0.12				
34	MC	0.41	0.42	0.15				
35	SA	0.18	0.46	0.16				+

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.53	0.41	0.19				
37	MC	0.85	0.32	0.15				
38	MC	0.35	0.54	0.31				
39	TE	0.59	0.61	0.25				
40	SA	0.57	0.45	0.15				
41	MC	0.58	0.41	0.17				
42	MC	0.28	0.13	0.15	+	+		
43	SA	0.51	0.44	0.18				
44	MC	0.28	0.40	0.16				
45	SA	0.49	0.54	0.18				
46	MC	0.53	0.38	0.21				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-9 Item Analysis, Grade 5 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	SA	0.66	0.50	0.20				
2	MC	0.46	0.50	0.08				
3	MC	0.57	0.18	0.15		+		
4	MC	0.34	0.39	0.12				
5	MC	0.42	0.34	0.14				
6	TE	0.39	0.47	0.47				
7	MC	0.64	0.50	0.10				
8	MC	0.43	0.35	0.25				
9	TE	0.31	0.52	0.59				
10	SA	0.42	0.51	0.15				
11	MC	0.33	0.24	0.17		+		
12	SA	0.30	0.49	0.32				
13	MC	0.59	0.56	0.16				
14	SA	0.37	0.57	0.21				
15	MC	0.46	0.51	0.39				
16	SA	0.32	0.35	0.31				
17	MC	0.48	0.32	0.17				
18	SA	0.47	0.44	0.21				
19	MC	0.63	0.52	0.17				
20	MC	0.39	0.26	0.21				
21	SA	0.21	0.42	0.22				
22	TE	0.58	0.46	0.22				
23	MC	0.45	0.49	0.23				
24	MC	0.55	0.32	0.22				
25	SA	0.46	0.47	0.13				
26	MC	0.39	0.37	0.13				
27	SA	0.25	0.44	0.13				
28	MC	0.50	0.40	0.14				
29	TE	0.46	0.53	0.27				
30	MC	0.58	0.29	0.13				
31	TE	0.53	0.56	0.24				
32	MC	0.75	0.36	0.17				
33	MC	0.40	0.28	0.15		+		
34	MC	0.51	0.41	0.24				
35	TE	0.44	0.58	0.16				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.30	0.42	0.14				
37	SA	0.45	0.53	0.28				
38	MC	0.46	0.39	0.29				
39	MC	0.60	0.43	0.33				
40	TE	0.29	0.41	0.16				
41	MC	0.13	0.28	0.22				+
42	MC	0.84	0.41	0.20				
43	MC	0.32	0.38	0.23				
44	SA	0.61	0.48	0.18				
45	MC	0.27	0.17	0.18		+		
46	SA	0.31	0.51	0.22				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-10 Item Analysis, Grade 6 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.39	0.35	0.14		+		
2	MC	0.39	0.50	0.10				
3	TE	0.56	0.53	0.39				
4	MC	0.69	0.49	0.14				
5	SA	0.49	0.48	0.31				
6	SA	0.49	0.51	0.37				
7	MC	0.28	0.40	0.15				
8	MC	0.79	0.43	0.17				
9	SA	0.18	0.50	0.18				+
10	MC	0.48	0.46	0.18				
11	TE	0.44	0.52	0.41				
12	MC	0.90	0.36	0.22				
13	SA	0.49	0.56	0.29				
14	SA	0.23	0.50	0.42				
15	MC	0.44	0.59	0.19				
16	MC	0.46	0.47	0.17				
17	MC	0.36	0.49	0.19				
18	MC	0.57	0.52	0.31				
19	MC	0.70	0.38	0.32				
20	SA	0.62	0.51	0.16				
21	TE	0.57	0.53	0.20				
22	MC	0.50	0.42	0.21				
23	TE	0.70	0.32	0.93				
24	MC	0.52	0.26	0.22				
25	MC	0.32	0.34	0.20		+		
26	MC	0.45	0.34	0.32				
27	SA	0.31	0.43	0.58				
28	MC	0.38	0.16	0.39		+		
29	MC	0.56	0.31	0.39				
30	MC	0.58	0.34	0.21				
31	MC	0.37	0.22	0.20				
32	MC	0.33	0.12	0.28	+			
33	MC	0.63	0.45	0.21				
34	MC	0.61	0.48	0.24				
35	MC	0.44	0.48	0.19				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	TE	0.36	0.49	0.36				
37	SA	0.24	0.58	0.67				
38	MC	0.40	0.23	0.45				
39	MC	0.42	0.43	0.40				
40	MC	0.35	0.53	0.47				
41	TE	0.41	0.54	0.77				
42	MC	0.39	0.28	0.25				
43	MC	0.33	0.22	0.24		+		
44	TE	0.14	0.17	0.27				+
45	MC	0.43	0.56	0.24				
46	MC	0.85	0.34	0.26				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-11 Item Analysis, Grade 7 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	SA	0.47	0.62	0.20				
2	MC	0.45	0.53	0.09				
3	MC	0.50	0.34	0.12				
4	SA	0.27	0.46	0.30				
5	MC	0.49	0.42	0.15				
6	MC	0.42	0.44	0.16				
7	MC	0.50	0.44	0.11				
8	TE	0.20	0.43	0.14				
9	MC	0.31	0.25	0.14		+		
10	MC	0.49	0.39	0.16				
11	MC	0.62	0.20	0.15				
12	MC	0.72	0.25	0.21				
13	MC	0.42	0.38	0.30				
14	MC	0.46	0.41	0.22				
15	SA	0.29	0.43	0.82				
16	MC	0.28	0.49	0.27				
17	MC	0.22	0.14	0.29	+	+		
18	MC	0.42	0.21	0.24				
19	TE	0.38	0.61	0.58				
20	MC	0.37	0.28	0.28				
21	MC	0.60	0.26	0.29				
22	MC	0.24	0.19	0.28				
23	MC	0.29	0.22	0.46				
24	SA	0.38	0.48	0.75				
25	SA	0.20	0.51	0.97				
26	MC	0.40	0.30	0.44				
27	MC	0.51	0.48	0.51				
28	MC	0.78	0.45	0.37				
29	MC	0.33	0.27	0.37				
30	MC	0.40	0.44	0.49				
31	MC	0.61	0.56	0.35				
32	TE	0.66	0.17	0.67				
33	TE	0.12	0.52	0.41				+
34	MC	0.56	0.34	0.49				
35	SA	0.40	0.66	0.82				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	TE	0.71	0.48	0.59				
37	MC	0.51	0.50	0.54				
38	TE	0.45	0.30	0.93				
39	MC	0.55	0.27	0.61				
40	MC	0.30	0.22	0.59		+		
41	TE	0.27	0.61	0.55				
42	TE	0.23	0.59	1.19				
43	MC	0.50	0.53	0.51				
44	MC	0.50	0.46	0.50				
45	MC	0.65	0.39	0.47				
46	SA	0.67	0.56	0.60				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-12 Item Analysis, Grade 8 Mathematics

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.66	0.46	0.08				
2	MC	0.24	0.16	0.12		+		
3	MC	0.55	0.40	0.14				
4	MC	0.45	0.21	0.12				
5	SA	0.20	0.51	0.28				
6	MC	0.34	0.42	0.16				
7	MC	0.49	0.58	0.10				
8	MC	0.49	0.48	0.10				
9	TE	0.26	0.55	0.20				
10	MC	0.41	0.30	0.17				
11	TE	0.24	0.50	0.36				
12	MC	0.47	0.30	0.15				
13	SA	0.46	0.61	0.28				
14	MC	0.48	0.25	0.25				
15	TE	0.46	0.49	0.29				
16	MC	0.47	0.26	0.35		+		
17	SA	0.37	0.53	0.95				
18	MC	0.56	0.41	0.21				
19	MC	0.47	0.39	0.27				
20	MC	0.48	0.42	0.23				
21	SA	0.29	0.59	0.73				
22	MC	0.73	0.39	0.27				
23	TE	0.57	0.51	0.98				
24	MC	0.54	0.38	0.38				
25	TE	0.20	0.47	0.77				
26	MC	0.25	0.21	0.39				
27	SA	0.17	0.44	1.10				+
28	MC	0.57	0.54	0.44				
29	MC	0.42	0.42	0.35				
30	MC	0.30	0.35	0.43				
31	MC	0.35	0.44	0.40				
32	TE	0.35	0.52	0.59				
33	MC	0.57	0.39	0.34				
34	MC	0.37	0.22	0.50		+		
35	TE	0.24	0.65	0.65				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MC	0.53	0.48	0.49				
37	TE	0.33	0.27	1.39				
38	MC	0.74	0.43	0.50				
39	MC	0.44	0.42	0.50				
40	MC	0.21	0.30	0.44				
41	TE	0.36	0.37	0.56				
42	MC	0.57	0.49	0.41				
43	MC	0.67	0.43	0.45				
44	MC	0.54	0.34	0.43				
45	MC	0.49	0.38	0.41				
46	MC	0.70	0.40	0.38				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-13 Item Analysis, Grade 4 Science

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.52	0.31	0.07				
2	TE	0.74	0.51	0.06				
3	TE	0.32	0.28	0.12				
4	TE	0.44	0.42	0.15				
5	TE	0.52	0.38	0.15				
6	TE	0.64	0.50	0.10				
7	MC	0.37	0.28	0.14				
8	TE	0.59	0.43	0.16				
9	EBSR	0.42	0.43	0.10				
10	MC	0.66	0.39	0.08				
11	MC	0.51	0.40	0.08				
12	MC	0.54	0.34	0.13				
13	TE	0.70	0.39	0.12				
14	TE	0.64	0.45	0.25				
15	TE	0.40	0.58	0.12				
16	MC	0.63	0.33	0.06				
17	TE	0.84	0.38	0.09				
18	TE	0.69	0.40	0.10				
19	TE	0.51	0.21	0.31				
20	TE	0.88	0.31	0.13				
21	TE	0.53	0.43	0.22				
22	TE	0.50	0.56	0.13				
23	MC	0.61	0.43	0.20				
24	TE	0.43	0.45	0.13				
25	TE	0.57	0.54	0.20				
26	TE	0.75	0.41	0.05				
27	TE	0.63	0.20	0.09				
28	EBSR	0.35	0.30	0.07				
29	EBSR	0.45	0.47	0.14				
30	TE	0.66	0.45	0.11				
31	TE	0.62	0.31	0.18				
32	MC	0.58	0.36	0.12				
33	TE	0.74	0.40	0.09				
34	TE	0.90	0.35	0.10				
35	MC	0.37	0.23	0.11		+		

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	EBSR	0.43	0.29	0.08				
37	MC	0.51	0.40	0.18				
38	TE	0.78	0.46	0.15				
39	MS	0.23	0.39	0.13				
40	TE	0.21	0.36	0.10				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-14 Item Analysis, Grade 8 Science

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	TE	0.65	0.43	0.19				
2	TE	0.54	0.46	0.18				
3	TE	0.55	0.57	0.20				
4	TE	0.61	0.51	0.22				
5	TE	0.83	0.33	0.19				
6	MC	0.57	0.35	0.19				
7	TE	0.53	0.36	0.16				
8	MC	0.62	0.40	0.23				
9	TE	0.79	0.31	0.14				
10	TE	0.49	0.33	0.12				
11	TE	0.63	0.41	0.17				
12	MC	0.53	0.31	0.15				
13	TE	0.36	0.41	0.24				
14	TE	0.70	0.36	0.23				
15	EBSR	0.54	0.30	0.13				
16	TE	0.40	0.45	0.07				
17	TE	0.43	0.38	0.24				
18	TE	0.55	0.57	0.13				
19	TE	0.30	0.37	0.22				
20	TE	0.52	0.35	0.19				
21	TE	0.82	0.41	0.35				
22	TE	0.73	0.44	0.18				
23	TE	0.59	0.46	0.26				
24	TE	0.50	0.12	0.15	+			
25	TE	0.54	0.46	0.13				
26	EBSR	0.36	0.51	0.04				
27	TE	0.51	0.53	0.17				
28	TE	0.59	0.21	0.21				
29	TE	0.64	0.43	0.17				
30	MC	0.52	0.44	0.19				
31	EBSR	0.46	0.48	0.08				
32	TE	0.73	0.39	0.17				
33	TE	0.42	0.37	0.49				
34	TE	0.43	0.37	0.21				
35	TE	0.29	0.37	0.16				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	MS	0.30	0.44	0.24				
37	TE	0.42	0.49	0.17				
38	MC	0.69	0.45	0.19				
39	TE	0.56	0.44	0.24				
40	MC	0.45	0.29	0.18				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-15 Item Analysis, Grade 4 Social Studies

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	1	MC	0.80	0.49	0.03			
2	2	MC	0.84	0.43	0.07			
3	3	MC	0.80	0.38	0.06			
4	4	MC	0.78	0.45	0.08			
5	5	MC	0.73	0.41	0.11			
6	6	MC	0.44	0.28	0.15			
7	7	TE	0.71	0.37	0.06			
8	8	MC	0.59	0.33	0.09			
9	9	MC	0.81	0.37	0.12			
10	10	MC	0.68	0.43	0.10			
11	11	MC	0.52	0.23	0.09			
12	12	MC	0.78	0.45	0.08			
13	13	TE	0.40	0.31	0.14			
14	14	MC	0.39	0.40	0.13			
15	15	MC	0.70	0.38	0.09			
16	16	MC	0.58	0.32	0.10			
17	17	MC	0.68	0.46	0.07			
18	18	TE	0.70	0.36	0.11			
19	19	MC	0.74	0.48	0.07			
20	20	MC	0.64	0.33	0.04			
21	21	MC	0.70	0.43	0.06			
22	22	MC	0.66	0.21	0.06			
23	23	MC	0.83	0.54	0.09			
24	24	MC	0.53	0.38	0.10			
25	25	MC	0.65	0.52	0.10			
26	26	MC	0.62	0.33	0.07			
27	27	MC	0.62	0.34	0.08			
28	28	MC	0.59	0.42	0.09			
29	29	MC	0.52	0.37	0.09			
30	30	MC	0.73	0.48	0.08			
31	31	TE	0.53	0.40	0.06			
32	32	MC	0.67	0.48	0.08			
33	33	MC	0.75	0.51	0.12			
34	34	MC	0.72	0.49	0.10			
35	35	MC	0.69	0.44	0.10			
36	36	MC	0.77	0.53	0.09			
37	37	MC	0.64	0.48	0.09			
38	38	MC	0.70	0.45	0.06			

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-16 Item Analysis, Grade 8 Social Studies

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.82	0.48	0.03				
2	MC	0.79	0.50	0.10				
3	MC	0.85	0.52	0.09				
4	MC	0.81	0.38	0.07				
5	MC	0.83	0.51	0.10				
6	MC	0.77	0.35	0.20				
7	MS	0.59	0.59	0.11				
8	MC	0.66	0.54	0.17				
9	MC	0.53	0.37	0.12				
10	TE	0.60	0.54	0.27				
11	MC	0.81	0.49	0.14				
12	MC	0.72	0.58	0.13				
13	MC	0.76	0.35	0.10				
14	MC	0.52	0.36	0.24				
15	MC	0.64	0.35	0.22				
16	MC	0.62	0.36	0.27				
17	MC	0.69	0.41	0.18				
18	MC	0.62	0.44	0.18				
19	MC	0.83	0.19	0.13				
20	MC	0.71	0.33	0.14				
21	MC	0.92	0.36	0.04				
22	MC	0.61	0.46	0.09				
23	TE	0.62	0.41	0.19				
24	MC	0.76	0.39	0.20				
25	MC	0.74	0.37	0.29				
26	MC	0.71	0.48	0.14				
27	MC	0.57	0.38	0.14				
28	MC	0.77	0.37	0.24				
29	MC	0.57	0.22	0.16				
30	TE	0.52	0.52	0.24				
31	MC	0.69	0.44	0.17				
32	TE	0.55	0.42	0.19				
33	MC	0.50	0.32	0.13				
34	MC	0.52	0.33	0.34				
35	MC	0.62	0.44	0.19				
36	MC	0.56	0.38	0.19				
37	MC	0.51	0.32	0.20				
38	MC	0.44	0.35	0.19				
39	MC	0.83	0.48	0.14				
40	MC	0.68	0.41	0.16				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-17 Item Analysis, Grade 10 Social Studies

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
1	MC	0.79	0.42	0.09				
2	MC	0.73	0.25	0.10				
3	MC	0.68	0.39	0.13				
4	MC	0.69	0.41	0.13				
5	MC	0.74	0.46	0.17				
6	MC	0.74	0.45	0.37				
7	MC	0.80	0.47	0.16				
8	MC	0.75	0.41	0.16				
9	MC	0.63	0.42	0.22				
10	MC	0.69	0.41	0.22				
11	MC	0.59	0.41	0.28				
12	TE	0.34	0.37	0.71				
13	MC	0.53	0.28	0.31				
14	MC	0.56	0.43	0.29				
15	MC	0.65	0.45	0.35				
16	MC	0.63	0.35	0.49				
17	MC	0.59	0.28	0.58				
18	MC	0.72	0.35	0.78				
19	MC	0.69	0.48	0.37				
20	MC	0.62	0.45	0.41				
21	MC	0.59	0.36	0.42				
22	MC	0.51	0.45	0.41				
23	MC	0.58	0.52	0.39				
24	TE	0.29	0.43	1.27				
25	MC	0.59	0.60	0.47				
26	MC	0.72	0.38	0.28				
27	MC	0.71	0.31	0.36				
28	MC	0.45	0.48	0.41				
29	MC	0.41	0.37	0.52				
30	MC	0.57	0.54	0.45				
31	MC	0.72	0.46	0.50				
32	MC	0.83	0.42	0.44				
33	MC	0.60	0.49	0.55				
34	MC	0.61	0.50	0.56				
35	MC	0.71	0.50	0.71				

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item	Item Type	*p*-Value	Corr	Percent Omit	Flags			
					Corr	Distractor	Omit	*p*-Value
36	TE	0.65	0.21	1.04				
37	MC	0.44	0.38	0.68				
38	MC	0.59	0.46	0.69				
39	MC	0.72	0.46	0.62				
40	MC	0.69	0.48	0.69				
41	MC	0.78	0.47	0.71				
42	MC	0.63	0.40	0.79				
43	MC	0.68	0.45	0.81				
44	MC	0.43	0.33	0.75				
45	TE	0.28	0.38	1.17				
46	MC	0.55	0.40	0.81				
47	MC	0.49	0.25	0.78				
48	MC	0.68	0.53	0.76				
49	MC	0.69	0.44	0.75				
50	MC	0.86	0.35	0.79				

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-18 Raw Score Descriptive Statistics for Total Population

Content	Grade	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
English Language Arts	3	61019	28.56	0.55	10.32	-0.14	-0.90	1	53	53	0.90	3.34
	4	63444	29.72	0.56	10.32	-0.01	-0.92	0	56	56	0.89	3.39
	5	64578	32.12	0.60	10.39	-0.18	-0.90	2	56	56	0.90	3.29
	6	65279	31.22	0.57	9.84	-0.22	-0.79	0	56	56	0.88	3.36
	7	63767	31.34	0.58	10.24	-0.18	-0.84	0	56	56	0.89	3.41
	8	62914	31.87	0.58	10.79	-0.19	-0.93	2	56	56	0.90	3.37
Mathematics	3	61151	23.56	0.56	9.90	-0.15	-1.02	0	42	42	0.93	2.69
	4	63561	22.17	0.48	9.98	0.27	-0.91	1	46	46	0.92	2.81
	5	64666	20.60	0.45	10.17	0.31	-0.85	0	46	46	0.92	2.90
	6	65393	21.48	0.47	9.98	0.37	-0.83	0	46	46	0.92	2.87
	7	63870	19.97	0.44	9.64	0.46	-0.68	0	46	46	0.91	2.90
	8	62989	19.99	0.44	9.94	0.54	-0.62	0	46	46	0.92	2.89
Science	4	63544	22.37	0.56	8.14	-0.10	-0.91	0	40	40	0.89	2.71
	8	62954	21.62	0.54	8.60	-0.03	-0.96	0	40	40	0.90	2.75
Social Studies	4	63541	25.19	0.66	7.86	-0.51	-0.68	2	38	38	0.89	2.56
	8	62951	26.83	0.67	8.26	-0.54	-0.63	1	40	40	0.90	2.60
	10	63227	30.94	0.62	10.68	-0.27	-0.95	0	50	50	0.92	3.01

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-19 Raw Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
English Language Arts	3	31077	27.61	0.61	10.29	0.89	29942	29.55	0.65	10.26	0.90
	4	32471	28.74	0.61	10.32	0.89	30973	30.75	0.64	10.21	0.89
	5	32860	30.79	0.64	10.41	0.90	31718	33.50	0.69	10.19	0.90
	6	33444	29.90	0.63	9.93	0.89	31835	32.61	0.67	9.55	0.88
	7	32727	29.98	0.64	10.28	0.89	31040	32.77	0.69	9.99	0.88
	8	32158	30.32	0.64	10.84	0.90	30756	33.49	0.70	10.50	0.90
Mathematics	3	31145	23.84	0.57	10.06	0.93	30006	23.27	0.56	9.72	0.92
	4	32536	22.65	0.49	10.19	0.92	31025	21.65	0.47	9.72	0.92
	5	32910	20.85	0.45	10.49	0.92	31756	20.35	0.44	9.82	0.91
	6	33500	21.46	0.47	10.25	0.92	31893	21.49	0.47	9.69	0.91
	7	32798	19.95	0.44	9.90	0.91	31072	19.99	0.44	9.36	0.90
	8	32203	19.72	0.43	10.21	0.92	30786	20.27	0.44	9.64	0.91
Science	4	32532	22.54	0.56	8.28	0.89	31012	22.19	0.56	7.99	0.88
	8	32193	21.58	0.54	8.91	0.91	30761	21.66	0.54	8.26	0.89
Social Studies	4	32529	25.03	0.66	8.04	0.90	31012	25.36	0.67	7.65	0.89
	8	32185	26.53	0.66	8.54	0.91	30766	27.15	0.68	7.95	0.89
	10	32302	30.42	0.61	11.09	0.93	30925	31.47	0.63	10.20	0.91

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Race/Ethnicity	Grade	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
White	3	40184	30.94	0.68	9.57	0.88
	4	41487	32.05	0.67	9.76	0.88
	5	42617	34.41	0.71	9.68	0.89
	6	43525	33.29	0.69	9.19	0.87
	7	42808	33.41	0.71	9.58	0.88
	8	43164	33.83	0.71	10.22	0.89
African-American	3	6544	20.00	0.46	9.41	0.87
	4	6970	21.55	0.47	8.81	0.85
	5	7001	23.52	0.50	9.27	0.87
	6	6872	23.31	0.49	8.97	0.85
	7	6531	23.12	0.50	9.32	0.86
	8	6263	23.57	0.50	9.80	0.88
Hispanic	3	8272	24.64	0.54	9.58	0.87
	4	8664	25.86	0.55	9.40	0.86
	5	8853	28.34	0.59	9.77	0.88
	6	8795	27.75	0.58	9.38	0.87
	7	8652	27.79	0.59	9.74	0.87
	8	8036	28.17	0.59	10.36	0.89
Asian	3	2553	27.93	0.61	10.34	0.90
	4	2747	29.82	0.62	10.24	0.89
	5	2589	32.67	0.67	10.52	0.90
	6	2597	31.98	0.66	9.64	0.88
	7	2537	32.22	0.68	10.22	0.89
	8	2447	33.65	0.70	10.66	0.90
American Indian	3	726	23.48	0.52	9.04	0.86
	4	756	24.78	0.53	9.10	0.86
	5	750	27.73	0.58	9.22	0.87
	6	809	26.15	0.55	9.20	0.86
	7	801	25.93	0.55	9.37	0.87
	8	754	26.29	0.55	9.89	0.88
Two or More	3	2740	27.78	0.61	10.36	0.89
	4	2820	28.72	0.61	10.38	0.89
	5	2768	31.45	0.66	10.47	0.90
	6	2681	30.18	0.63	9.80	0.88
	7	2438	30.36	0.65	10.49	0.89
	8	2250	30.51	0.64	10.82	0.90

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-21 Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
White	3	40203	25.94	0.62	9.18	0.92
	4	41496	24.60	0.54	9.56	0.91
	5	42616	22.91	0.50	9.83	0.91
	6	43526	23.67	0.52	9.75	0.91
	7	42840	22.09	0.48	9.42	0.90
	8	43167	21.90	0.48	9.79	0.91
African-American	3	6551	15.22	0.37	8.60	0.90
	4	6974	13.99	0.31	7.28	0.86
	5	7009	12.41	0.27	7.22	0.86
	6	6883	13.73	0.30	7.02	0.84
	7	6518	12.13	0.27	6.45	0.82
	8	6255	12.37	0.27	6.87	0.84
Hispanic	3	8354	19.50	0.47	9.13	0.91
	4	8749	17.86	0.39	8.60	0.89
	5	8917	16.34	0.36	8.76	0.89
	6	8873	17.33	0.38	8.46	0.89
	7	8715	16.00	0.35	7.97	0.87
	8	8108	15.80	0.35	8.20	0.88
Asian	3	2577	23.52	0.56	10.10	0.93
	4	2766	22.88	0.50	10.47	0.93
	5	2608	22.22	0.48	10.70	0.93
	6	2619	22.59	0.49	10.39	0.93
	7	2553	20.98	0.46	10.45	0.92
	8	2459	22.47	0.49	10.92	0.93
American Indian	3	725	18.21	0.43	8.64	0.90
	4	755	16.78	0.37	8.12	0.88
	5	750	15.73	0.34	8.56	0.89
	6	808	16.23	0.35	8.34	0.88
	7	801	14.63	0.32	7.77	0.87
	8	754	15.00	0.33	7.48	0.85
Two or More	3	2741	22.43	0.54	9.85	0.92
	4	2821	20.62	0.45	9.60	0.91
	5	2766	19.34	0.42	9.97	0.92
	6	2684	19.95	0.44	9.75	0.91
	7	2443	18.45	0.40	9.39	0.91
	8	2246	18.54	0.41	9.65	0.91

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Race/Ethnicity	Grade	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
White	4	41492	24.59	0.62	7.44	0.87
	8	43164	23.50	0.59	8.09	0.89
African-American	4	6965	14.95	0.38	6.64	0.83
	8	6236	13.97	0.35	6.84	0.84
Hispanic	4	8747	18.64	0.47	7.42	0.86
	8	8098	17.90	0.45	7.85	0.87
Asian	4	2766	21.51	0.54	8.04	0.88
	8	2460	22.36	0.56	8.67	0.90
American Indian	4	755	18.33	0.46	7.18	0.85
	8	751	17.72	0.45	7.67	0.87
Two or More	4	2819	21.46	0.54	8.01	0.88
	8	2245	20.61	0.52	8.66	0.90

Table 8-23 Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
White	4	41495	27.16	0.72	7.01	0.88
	8	43159	28.54	0.71	7.51	0.89
	10	45333	32.91	0.66	9.99	0.91
African-American	4	6964	18.27	0.48	7.63	0.87
	8	6239	19.68	0.50	8.07	0.88
	10	5195	21.11	0.44	9.51	0.89
Hispanic	4	8746	22.08	0.58	7.66	0.87
	8	8100	23.61	0.59	8.22	0.89
	10	7624	26.39	0.53	10.21	0.90
Asian	4	2765	24.80	0.65	7.57	0.88
	8	2458	27.45	0.69	8.17	0.90
	10	2434	31.59	0.64	10.55	0.92
American Indian	4	752	21.50	0.57	7.56	0.87
	8	752	22.98	0.58	8.06	0.88
	10	706	26.62	0.54	9.88	0.90
Two or More	4	2819	24.41	0.64	7.83	0.89
	8	2243	26.10	0.65	8.33	0.90
	10	1935	29.70	0.60	10.94	0.92

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-24 Raw Score Descriptive Statistics by Socioeconomic Status

Content	Grade	Economically Disadvantaged					Not Economically Disadvantaged				
		N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
English Language Arts	3	27365	24.49	0.54	9.89	0.88	33654	31.87	0.69	9.44	0.88
	4	28396	25.69	0.55	9.61	0.87	35048	32.99	0.69	9.69	0.88
	5	28709	27.84	0.58	9.87	0.88	35869	35.55	0.74	9.49	0.89
	6	28197	27.20	0.57	9.45	0.87	37082	34.28	0.71	8.99	0.87
	7	26808	27.12	0.58	9.80	0.88	36959	34.40	0.73	9.43	0.87
	8	24976	27.30	0.57	10.32	0.89	37938	34.88	0.73	10.01	0.89
Mathematics	3	27436	19.52	0.47	9.45	0.92	33715	26.85	0.64	9.01	0.91
	4	28460	18.03	0.39	8.75	0.90	35101	25.52	0.56	9.64	0.91
	5	28755	16.39	0.36	8.89	0.90	35911	23.98	0.52	9.87	0.91
	6	28270	17.21	0.38	8.54	0.89	37123	24.73	0.54	9.77	0.91
	7	26865	15.82	0.35	8.09	0.88	37005	22.98	0.50	9.55	0.91
	8	25021	15.66	0.34	8.26	0.88	37968	22.84	0.50	9.92	0.91
Science	4	28447	19.08	0.48	7.72	0.87	35097	25.04	0.63	7.47	0.87
	8	25005	17.91	0.45	8.10	0.88	37949	24.06	0.60	8.03	0.88
Social Studies	4	28446	22.07	0.58	7.87	0.88	35095	27.72	0.73	6.88	0.87
	8	24999	23.20	0.58	8.35	0.89	37952	29.23	0.73	7.27	0.88
	10	21960	26.13	0.53	10.41	0.91	41267	33.50	0.67	9.91	0.91

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-25 Raw Score Descriptive Statistics by Disability

Content	Grade	Disabled					Not Disabled				
		N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
English Language Arts	3	8014	21.25	0.48	9.65	0.88	53005	29.66	0.65	9.96	0.89
	4	8111	21.99	0.48	9.30	0.87	55333	30.85	0.65	9.96	0.89
	5	7979	23.07	0.49	9.35	0.87	56599	33.40	0.69	9.89	0.89
	6	7959	21.88	0.47	8.63	0.85	57320	32.52	0.68	9.28	0.87
	7	7685	21.34	0.46	8.63	0.85	56082	32.71	0.69	9.66	0.88
	8	7395	21.33	0.45	8.95	0.86	55519	33.27	0.70	10.22	0.89
Mathematics	3	8024	16.82	0.40	9.57	0.92	53127	24.58	0.59	9.55	0.92
	4	8111	15.51	0.34	8.68	0.90	55450	23.14	0.50	9.78	0.92
	5	7987	13.11	0.29	8.41	0.89	56679	21.66	0.47	9.95	0.91
	6	7962	13.55	0.30	7.44	0.86	57431	22.57	0.49	9.79	0.91
	7	7692	12.13	0.27	6.75	0.84	56178	21.04	0.46	9.48	0.91
	8	7396	11.92	0.26	6.56	0.83	55593	21.06	0.46	9.82	0.91
Science	4	8102	17.04	0.43	7.90	0.88	55442	23.15	0.58	7.88	0.88
	8	7372	14.58	0.37	7.48	0.87	55582	22.55	0.57	8.30	0.89
Social Studies	4	8107	19.55	0.52	8.22	0.89	55434	26.02	0.69	7.45	0.88
	8	7376	19.01	0.48	8.15	0.88	55575	27.87	0.70	7.70	0.89
	10	6755	21.20	0.43	9.48	0.89	56472	32.10	0.65	10.21	0.91

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Content	Grade	Limited English Proficient					Fully English Proficient				
		N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
English Language Arts	3	5599	23.01	0.51	8.88	0.85	55420	29.12	0.64	10.29	0.90
	4	5621	23.97	0.51	8.21	0.82	57823	30.28	0.64	10.33	0.89
	5	4904	24.71	0.51	8.06	0.82	59674	32.73	0.68	10.33	0.90
	6	3988	22.94	0.48	7.40	0.79	61291	31.76	0.66	9.74	0.88
	7	3597	22.30	0.47	7.62	0.79	60170	31.88	0.68	10.12	0.89
	8	2983	21.93	0.46	7.95	0.82	59931	32.36	0.68	10.67	0.90
Mathematics	3	5715	18.53	0.44	8.87	0.90	55436	24.08	0.58	9.86	0.93
	4	5733	16.74	0.36	7.93	0.88	57828	22.71	0.49	10.00	0.92
	5	4992	13.95	0.30	7.29	0.85	59674	21.16	0.46	10.18	0.92
	6	4080	13.68	0.30	5.94	0.78	61313	21.99	0.48	9.98	0.92
	7	3689	12.31	0.27	5.56	0.75	60181	20.44	0.45	9.64	0.91
	8	3071	12.17	0.27	5.64	0.76	59918	20.39	0.45	9.94	0.92
Science	4	5731	17.10	0.43	6.62	0.82	57813	22.89	0.57	8.09	0.89
	8	3062	13.50	0.34	5.64	0.76	59892	22.04	0.55	8.52	0.90
Social Studies	4	5733	20.71	0.55	7.13	0.85	57808	25.64	0.68	7.78	0.89
	8	3064	18.53	0.46	6.93	0.83	59887	27.26	0.68	8.10	0.90
	10	2472	19.71	0.40	7.43	0.81	60755	31.39	0.63	10.54	0.92

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-27 Raw Score Descriptive Statistics by Accommodation Use

Content	Grade	Students Using Testing Accommodations					Students Not Using Accommodations				
		N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha	N Count	Mean Raw Score	Test Difficulty	Raw Score SD	Alpha
English Language Arts	3	57	21.74	0.50	10.77	0.90	60962	28.57	0.63	10.32	0.89
	4	75	23.68	0.51	10.10	0.89	63369	29.73	0.62	10.31	0.89
	5	76	27.26	0.57	12.05	0.92	64502	32.13	0.67	10.39	0.90
	6	84	25.38	0.54	10.10	0.89	65195	31.23	0.65	9.84	0.88
	7	101	22.63	0.49	9.87	0.87	63666	31.35	0.67	10.23	0.89
	8	88	23.72	0.50	12.08	0.92	62826	31.88	0.67	10.79	0.90
Mathematics	3	706	11.87	0.29	6.89	0.85	60445	23.70	0.57	9.85	0.93
	4	2156	11.95	0.26	5.78	0.79	61405	22.53	0.49	9.90	0.92
	5	2479	10.01	0.22	5.80	0.80	62187	21.03	0.46	10.08	0.92
	6	2782	10.86	0.24	4.78	0.69	62611	21.95	0.48	9.89	0.92
	7	2724	9.93	0.22	4.26	0.63	61146	20.41	0.45	9.57	0.91
	8	2495	9.95	0.22	4.17	0.60	60494	20.40	0.45	9.89	0.91
Science	4	29	14.97	0.37	8.65	-	63515	22.37	0.56	8.14	0.89
	8	33	13.88	0.35	8.13	-	62921	21.62	0.54	8.60	0.90
Social Studies	4	30	17.40	0.46	8.99	-	63511	25.20	0.66	7.85	0.89
	8	33	17.33	0.43	8.63	-	62918	26.84	0.67	8.26	0.90
	10	26	28.31	0.57	9.73	-	63201	30.94	0.62	10.68	0.92

Note: Test reliability coefficients are not reported for subgroups with less than 50 students.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-28 Scale Score Descriptive Statistics for Total Population

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
English Language Arts	3	61091	554.59	45.54	-0.17	0.52	330	900	330	900
	4	63528	582.01	51.05	0.01	0.31	340	930	340	930
	5	64654	595.58	48.77	-0.02	0.37	350	940	350	940
	6	65386	607.00	50.15	-0.25	0.24	360	950	360	950
	7	63878	627.70	54.88	-0.14	0.24	370	960	370	960
	8	63056	629.06	59.84	-0.16	0.35	380	970	380	970
Mathematics	3	61210	555.78	53.50	-0.50	2.19	360	760	360	760
	4	63630	577.09	51.78	-0.60	1.22	405	800	405	800
	5	64728	601.48	53.14	-0.91	2.12	430	830	430	830
	6	65470	610.77	58.31	-0.58	1.26	440	870	440	870
	7	63973	625.25	60.69	-0.69	1.17	450	880	450	880
	8	63108	644.53	57.85	-0.55	1.30	470	890	470	890
Science	4	63611	499.88	50.24	0.23	0.31	300	725	300	725
	8	63062	699.70	50.55	0.23	0.66	480	945	480	945
Social Studies	4	63603	396.68	55.69	-0.29	1.39	200	570	200	570
	8	63045	598.81	52.30	-0.01	1.16	420	780	420	780
	10	63476	692.82	58.30	-0.42	1.05	490	890	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-29 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	31117	550.26	45.26	330	900	29974	559.09	45.39	330	900
	4	32515	576.84	50.96	340	836	31013	587.43	50.57	340	930
	5	32901	589.31	48.53	350	852	31753	602.09	48.17	350	940
	6	33503	599.74	50.67	360	950	31883	614.64	48.45	360	950
	7	32786	620.38	55.11	370	915	31092	635.42	53.58	370	960
	8	32235	619.83	60.06	380	970	30821	638.72	58.07	380	970
Mathematics	3	31178	557.27	55.55	360	760	30032	554.24	51.25	360	760
	4	32575	578.92	53.18	405	800	31055	575.17	50.19	405	800
	5	32945	602.03	55.28	430	830	31783	600.90	50.83	430	830
	6	33540	609.62	61.39	440	870	31930	611.97	54.87	440	870
	7	32850	623.98	63.94	450	880	31123	626.60	57.04	450	880
	8	32270	641.89	60.81	470	890	30838	647.30	54.44	470	890
Science	4	32565	501.24	51.45	300	725	31046	498.45	48.90	300	725
	8	32244	699.33	52.79	480	945	30818	700.08	48.08	480	945
Social Studies	4	32562	395.58	57.34	200	570	31041	397.83	53.89	200	570
	8	32234	597.23	54.32	420	780	30811	600.46	50.04	420	780
	10	32423	689.92	61.36	490	890	31053	695.86	54.77	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-30 Scale Score Descriptive Statistics for English Language Arts by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	40204	565.11	41.68	330	900
	4	41517	593.46	47.83	340	930
	5	42646	606.20	45.48	350	940
	6	43569	617.51	46.38	360	950
	7	42845	638.79	51.00	382	960
	8	43222	639.82	56.30	380	970
African-American	3	6565	515.97	44.92	330	690
	4	6998	541.22	46.52	340	755
	5	7023	555.47	44.65	350	800
	6	6906	566.18	48.35	360	793
	7	6573	583.14	52.51	370	848
	8	6310	583.07	57.00	380	818
Hispanic	3	8295	537.64	42.07	330	702
	4	8682	563.41	46.13	340	793
	5	8868	578.06	44.75	362	786
	6	8814	589.68	47.60	370	836
	7	8672	609.04	51.78	370	885
	8	8064	609.05	57.10	380	823
Asian	3	2556	553.04	45.02	403	759
	4	2748	583.15	51.21	359	930
	5	2592	598.50	50.53	350	852
	6	2598	611.86	49.20	423	803
	7	2539	633.64	54.96	428	884
	8	2448	640.20	59.88	380	894
American Indian	3	726	533.73	40.55	330	652
	4	761	558.21	44.66	414	695
	5	750	575.82	41.43	387	700
	6	811	581.19	47.73	396	734
	7	805	599.44	49.73	439	747
	8	754	598.86	55.78	390	768
Two or More	3	2745	551.08	45.04	330	717
	4	2822	577.37	52.09	340	761
	5	2775	592.53	49.34	350	779
	6	2688	601.51	50.50	403	798
	7	2444	622.61	56.62	370	840
	8	2258	621.25	60.33	380	838

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	3	40220	568.09	48.04	360	760
	4	41527	589.98	45.41	405	800
	5	42645	613.62	46.32	430	830
	6	43555	624.24	51.80	440	870
	7	42875	639.35	53.64	450	880
	8	43213	656.42	52.13	470	890
African-American	3	6568	510.67	55.49	360	700
	4	6994	530.21	53.32	405	706
	5	7023	555.36	57.42	430	721
	6	6907	558.58	60.42	440	772
	7	6562	568.69	61.83	450	752
	8	6303	593.20	59.31	470	798
Hispanic	3	8375	535.73	50.18	360	760
	4	8759	556.26	49.15	405	715
	5	8929	581.11	52.70	430	830
	6	8887	588.37	55.21	440	870
	7	8729	602.60	57.69	450	880
	8	8122	621.47	55.01	470	890
Asian	3	2578	556.87	56.08	360	760
	4	2767	582.55	52.93	405	800
	5	2608	610.42	53.59	430	830
	6	2622	618.44	58.42	440	870
	7	2554	631.82	63.42	450	880
	8	2461	659.91	57.78	470	890
American Indian	3	725	530.27	48.17	360	698
	4	760	550.51	48.30	405	672
	5	750	577.13	54.05	430	733
	6	809	580.67	58.06	440	780
	7	802	590.31	62.27	450	800
	8	756	617.73	52.06	470	756
Two or More	3	2744	550.37	53.78	360	760
	4	2823	570.04	51.80	405	800
	5	2773	595.30	53.75	430	830
	6	2690	602.07	59.57	440	870
	7	2451	615.32	62.13	450	880
	8	2253	635.40	61.08	470	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-32 Scale Score Descriptive Statistics for Science by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	41521	513.22	46.79	300	725
	8	43208	710.47	47.59	480	945
African-American	4	6982	455.21	40.71	300	663
	8	6275	655.23	42.20	480	945
Hispanic	4	8762	477.46	44.08	300	725
	8	8113	678.78	45.26	480	945
Asian	4	2767	495.71	50.10	334	725
	8	2462	704.99	52.55	521	945
American Indian	4	759	475.49	41.68	372	657
	8	754	677.60	44.10	550	847
Two or More	4	2820	494.27	48.03	300	725
	8	2250	693.87	50.47	524	945

Table 8-33 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity

Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
White	4	41522	409.96	50.81	200	570
	8	43202	609.25	48.68	420	780
	10	45451	703.44	53.51	490	890
African-American	4	6978	349.82	56.34	200	570
	8	6273	554.89	49.77	420	780
	10	5261	638.71	59.40	490	890
Hispanic	4	8761	375.44	51.46	200	570
	8	8112	579.23	48.88	420	780
	10	7668	669.43	56.22	490	890
Asian	4	2766	394.69	52.23	200	570
	8	2459	603.46	53.63	420	780
	10	2439	697.24	58.43	490	890
American Indian	4	756	371.37	51.47	200	514
	8	753	574.84	49.55	420	717
	10	713	670.28	53.31	490	846
Two or More	4	2820	391.85	53.35	200	570
	8	2246	594.20	50.64	420	780
	10	1944	686.12	61.25	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-34 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	27419	536.87	43.98	330	731	33672	569.02	41.51	330	900
	4	28461	562.34	47.79	340	839	35067	597.97	47.94	340	930
	5	28766	575.68	45.47	350	940	35888	611.54	45.34	350	940
	6	28275	586.63	48.53	360	803	37111	622.53	45.61	394	950
	7	26893	605.29	52.46	370	915	36985	644.00	50.70	370	960
	8	25087	604.09	57.53	380	896	37969	645.57	55.46	380	970
Mathematics	3	27479	534.90	53.10	360	760	33731	572.80	47.46	360	760
	4	28512	555.96	51.28	405	800	35118	594.25	45.45	405	800
	5	28803	580.40	54.20	430	830	35925	618.38	45.73	430	830
	6	28327	586.16	57.67	440	870	37143	629.53	51.41	440	870
	7	26941	599.65	60.62	450	880	37032	643.88	53.54	450	880
	8	25116	619.23	57.17	470	890	37992	661.26	51.92	470	890
Science	4	28498	479.86	46.18	300	725	35113	516.12	47.47	330	725
	8	25089	678.36	47.13	480	945	37973	713.80	47.70	480	945
Social Studies	4	28494	375.40	54.05	200	570	35109	413.94	50.82	200	570
	8	25072	576.53	50.02	420	780	37973	613.52	48.43	420	780
	10	22105	667.23	58.08	490	890	41371	706.50	53.63	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-35 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	8027	522.31	44.17	330	695	53064	559.48	43.71	330	900
	4	8140	543.23	49.00	340	930	55388	587.71	48.81	340	930
	5	8001	553.07	45.89	350	814	56653	601.59	46.11	350	940
	6	7988	558.78	47.49	360	828	57398	613.72	46.72	360	950
	7	7712	574.28	49.38	370	822	56166	635.04	51.43	370	960
	8	7442	569.89	55.04	380	896	55614	636.98	55.89	380	970
Mathematics	3	8033	516.62	62.36	360	760	53177	561.70	49.40	360	760
	4	8137	536.43	58.74	405	800	55493	583.05	47.86	405	800
	5	8005	555.65	62.76	430	830	56723	607.95	48.26	430	830
	6	7986	556.40	62.06	440	870	57484	618.32	53.58	440	870
	7	7711	567.98	63.42	450	828	56262	633.10	55.91	450	880
	8	7432	590.16	58.18	470	890	55676	651.79	53.80	470	890
Science	4	8127	467.44	48.37	300	725	55484	504.63	48.73	300	725
	8	7398	658.72	45.17	480	945	55664	705.14	48.69	480	945
Social Studies	4	8131	357.75	60.91	200	570	55472	402.38	52.52	200	570
	8	7402	552.15	50.48	420	780	55643	605.01	49.31	420	780
	10	6824	639.71	59.03	490	890	56652	699.22	54.85	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-36 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	5612	530.92	39.11	330	698	55479	556.99	45.46	330	900
	4	5632	554.40	40.62	340	721	57896	584.70	51.16	340	930
	5	4913	561.80	36.68	350	702	59741	598.36	48.60	350	940
	6	3998	566.28	39.04	378	714	61388	609.66	49.65	360	950
	7	3606	580.99	41.90	370	736	60272	630.50	54.31	370	960
	8	2997	576.03	47.29	380	732	60059	631.71	59.17	380	970
Mathematics	3	5724	530.86	49.91	360	760	55486	558.36	53.20	360	760
	4	5735	550.87	48.26	405	693	57895	579.69	51.39	405	800
	5	4996	569.31	51.46	430	830	59732	604.17	52.40	430	830
	6	4087	565.68	51.13	440	806	61383	613.77	57.52	440	870
	7	3693	576.40	55.24	450	725	60280	628.25	59.73	450	880
	8	3077	598.29	49.96	470	826	60031	646.90	57.23	470	890
Science	4	5736	468.23	38.80	300	688	57875	503.01	50.16	300	725
	8	3071	653.63	34.59	480	933	59991	702.06	50.10	480	945
Social Studies	4	5736	366.64	47.39	200	570	57867	399.66	55.57	200	570
	8	3071	550.15	42.28	420	708	59974	601.30	51.54	420	780
	10	2492	633.80	49.23	490	806	60984	695.24	57.37	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-37 Scale Score Descriptive Statistics by Accommodation Use

Content	Grade	Students Using Testing Accommodations					Students Not Using Accommodations				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
English Language Arts	3	57	521.18	54.97	336	614	61034	554.62	45.52	330	900
	4	75	550.89	50.43	429	662	63453	582.05	51.04	340	930
	5	77	572.60	58.93	439	750	64577	595.61	48.75	350	940
	6	85	576.08	54.04	444	684	65301	607.04	50.14	360	950
	7	101	579.84	55.44	464	718	63777	627.78	54.85	370	960
	8	89	579.73	78.12	380	791	62967	629.13	59.79	380	970
Mathematics	3	707	485.94	57.52	360	631	60503	556.60	52.91	360	760
	4	2164	512.71	51.63	405	661	61466	579.36	50.30	405	800
	5	2482	533.81	59.05	430	692	62246	604.18	51.07	430	830
	6	2792	534.65	54.61	440	722	62678	614.16	56.12	440	870
	7	2731	548.23	56.95	450	828	61242	628.69	58.54	450	880
	8	2512	574.55	51.78	470	742	60596	647.43	56.24	470	890
Science	4	29	452.52	55.99	380	618	63582	499.90	50.23	300	725
	8	33	654.24	48.39	566	768	63029	699.72	50.54	480	945
Social Studies	4	30	339.43	67.46	200	454	63573	396.71	55.67	200	570
	8	33	540.36	64.98	420	737	63012	598.84	52.27	420	780
	10	26	682.12	46.50	610	787	63450	692.83	58.31	490	890

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-38 Performance Level Cut Scores for All Content Areas

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				447	496	543										653	695	737			
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 8-39 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	23.28	38.04	33.21	5.48	38.69
4	340–545	546–591	592–649	650–930	23.88	33.14	34.10	8.89	42.98
5	350–563	564–609	610–669	670–940	26.11	33.83	34.34	5.72	40.06
6	360–571	572–621	622–670	671–950	23.56	35.48	31.87	9.09	40.96
7	370–584	585–637	638–696	697–960	21.88	33.25	35.36	9.51	44.87
8	380–591	592–651	652–707	708–970	25.94	37.04	28.80	8.23	37.03

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-40 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	19.28	31.28	37.17	12.27	49.44
4	405–535	536–587	588–632	633–800	18.87	36.09	32.82	12.23	45.05
5	430–573	574–610	611–657	658–830	24.22	29.20	35.09	11.49	46.58
6	440–581	582–625	626–687	688–870	26.72	30.79	35.80	6.69	42.49
7	450–605	606–646	647–711	712–880	32.18	28.99	34.05	4.78	38.83
8	470–619	620–666	667–717	718–890	28.55	35.60	27.83	8.01	35.85

Table 8-41 Cut Scores and Associated Impact Data, Science

Grade	Score Range				Impact Data				
	Below Basic	**Basic**	**Proficient**	**Advanced**	**Below Basic**	**Basic**	**Proficient**	**Advanced**	**Proficient + Advanced**
4	300-446	447-495	496-542	543-725	14.98	32.25	33.29	19.49	52.78
8	480-652	653-694	695-736	737-945	17.76	28.29	31.50	22.45	53.95

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-42 Cut 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	24.04	23.47	30.38	22.11	52.49
8	420–562	563–598	599–639	640–780	22.24	26.16	32.25	19.35	51.60
10	490–669	670–702	703–740	741–890	30.97	23.53	26.12	19.38	45.50

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-43 Percentage of Students in Each Performance Level by Subgroup, English Language Arts

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
3	BB	14219	23.3	20.2	26.2	14.6	56.8	35.6	24.9	39.8	26.7	21.5	40.6	52.2	18.9	36.5	12.5	49.1	23.3
	B	23236	38.0	37.2	38.8	38.4	30.9	41.5	39.5	43.4	37.2	37.5	42.9	33.1	38.8	40.0	36.4	35.1	38.0
	P	20290	33.2	35.9	30.6	40.0	11.5	21.1	29.0	14.7	31.4	35.0	15.7	13.2	36.2	21.4	42.8	15.8	33.2
	A	3346	5.5	6.7	4.3	7.0	.8	1.9	6.6	2.1	4.7	6.0	.8	1.5	6.1	2.1	8.2	.	5.5
Total		61091	100.0	29974	31117	40204	6565	8295	2556	726	2745	55479	5612	8027	53064	27419	33672	57	61034
4	BB	15169	23.9	20.4	27.2	15.6	56.1	35.5	22.3	39.6	27.2	22.2	40.6	55.2	19.3	36.8	13.4	49.3	23.8
	B	21052	33.1	32.7	33.6	32.3	29.9	37.8	36.4	37.6	34.1	32.3	41.7	28.4	33.8	36.3	30.6	25.3	33.1
	P	21662	34.1	36.4	31.9	40.7	12.5	23.4	31.1	20.5	30.5	35.8	16.7	14.1	37.0	23.5	42.7	22.7	34.1
	A	5645	8.9	10.5	7.3	11.4	1.4	3.3	10.2	2.4	8.2	9.7	1.0	2.2	9.9	3.4	13.3	2.7	8.9
Total		63528	100.0	31013	32515	41517	6998	8682	2748	761	2822	57896	5632	8140	55388	28461	35067	75	63453
5	BB	16883	26.1	21.5	30.6	17.7	60.0	38.7	24.8	38.1	28.1	24.0	52.1	63.0	20.9	40.6	14.5	53.2	26.1
	B	21873	33.8	33.5	34.2	33.9	27.8	37.1	34.4	40.5	35.4	33.5	38.3	25.3	35.0	35.9	32.2	16.9	33.9
	P	22201	34.3	37.8	31.0	41.2	11.6	22.0	32.9	20.5	30.6	36.4	9.3	10.5	37.7	21.7	44.5	26.0	34.3
	A	3697	5.7	7.3	4.2	7.2	.6	2.1	7.9	.8	5.9	6.2	.2	1.2	6.4	1.8	8.8	3.9	5.7
Total		64654	100.0	31753	32901	42646	7023	8868	2592	750	2775	59741	4913	8001	56653	28766	35888	77	64577

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-43 Percentage of Students in Each Performance Level by Subgroup, English Language Arts (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
6	BB	15407	23.6	18.6	28.3	16.1	54.5	34.7	19.7	42.2	26.7	21.6	53.4	62.9	18.1	37.1	13.2	47.1	23.5
	B	23199	35.5	34.9	36.0	34.9	32.7	39.2	37.1	39.6	37.3	35.2	39.5	27.0	36.7	38.3	33.3	35.3	35.5
	P	20839	31.9	35.0	28.9	37.6	11.3	22.3	32.1	15.4	28.2	33.5	6.8	8.9	35.1	21.2	40.0	15.3	31.9
	A	5941	9.1	11.5	6.8	11.4	1.4	3.8	11.0	2.8	7.9	9.7	.3	1.2	10.2	3.4	13.4	2.4	9.1
Total		65386	100.0	31883	33503	43569	6906	8814	2598	811	2688	61388	3998	7988	57398	28275	37111	85	65301
7	BB	13975	21.9	17.4	26.2	14.9	52.3	31.8	18.8	38.5	25.8	20.0	53.0	61.5	16.4	34.9	12.4	57.4	21.8
	B	21238	33.2	32.5	34.0	32.1	32.3	38.6	34.3	39.8	33.3	32.9	38.5	27.9	34.0	37.3	30.3	24.8	33.3
	P	22590	35.4	38.4	32.5	41.2	13.8	25.6	34.8	19.4	31.8	37.0	8.3	9.6	38.9	24.4	43.4	15.8	35.4
	A	6075	9.5	11.7	7.4	11.8	1.7	4.0	12.1	2.4	9.1	10.1	.2	1.0	10.7	3.5	13.9	2.0	9.5
Total		63878	100.0	31092	32786	42845	6573	8672	2539	805	2444	60272	3606	7712	56166	26893	36985	101	63777
8	BB	16355	25.9	20.8	30.8	19.0	57.7	37.7	19.5	43.0	29.9	24.2	61.7	66.3	20.5	40.8	16.1	60.7	25.9
	B	23354	37.0	36.3	37.7	37.7	29.9	38.7	37.9	39.4	37.3	37.2	32.9	26.7	38.4	38.0	36.4	24.7	37.1
	P	18159	28.8	32.3	25.5	33.4	10.8	19.8	30.4	15.8	26.5	30.0	5.3	6.2	31.8	18.1	35.9	7.9	28.8
	A	5188	8.2	10.6	6.0	10.0	1.5	3.8	12.2	1.9	6.2	8.6	.1	.8	9.2	3.0	11.6	6.7	8.2
Total		63056	100.0	30821	32235	43222	6310	8064	2448	754	2258	60059	2997	7442	55614	25087	37969	89	62967

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
3	BB	11802	19.3	19.7	18.9	11.4	50.3	30.4	20.1	34.1	22.3	17.8	33.8	45.3	15.4	31.5	9.4	68.3	18.7
	B	19146	31.3	32.4	30.2	29.0	33.5	38.7	32.0	41.2	32.8	30.4	39.3	31.6	31.2	36.5	27.0	25.7	31.3
	P	22749	37.2	37.1	37.2	43.8	14.3	26.3	34.3	21.4	35.1	38.6	23.6	19.2	39.9	27.0	45.4	5.5	37.5
	A	7513	12.3	10.8	13.7	15.8	1.9	4.5	13.6	3.3	9.8	13.2	3.2	4.0	13.5	5.0	18.2	.4	12.4
Total		61210	100.0	30032	31178	40220	6568	8375	2578	725	2744	55486	5724	8033	53177	27479	33731	707	60503
4	BB	12004	18.9	19.1	18.6	10.6	51.6	30.5	15.8	33.0	21.8	17.5	32.9	47.6	14.6	30.7	9.3	67.1	17.2
	B	22962	36.1	38.2	34.1	34.1	35.9	43.1	37.5	45.8	39.6	35.2	45.4	33.8	36.4	42.0	31.2	27.2	36.4
	P	20884	32.8	32.0	33.6	39.4	11.0	22.1	30.2	18.7	29.1	34.2	18.9	14.4	35.5	22.6	41.1	5.3	33.8
	A	7780	12.2	10.7	13.7	15.8	1.5	4.3	16.5	2.5	9.5	13.2	2.7	4.2	13.4	4.7	18.4	.4	12.6
Total		63630	100.0	31055	32575	41527	6994	8759	2767	760	2823	57895	5735	8137	55493	28512	35118	2164	61466
5	BB	15677	24.2	23.6	24.8	15.5	58.1	37.8	19.3	40.0	28.6	22.4	46.5	58.3	19.4	38.4	12.8	75.1	22.2
	B	18903	29.2	30.9	27.5	28.1	28.5	34.0	29.8	33.7	30.4	28.7	35.4	24.4	29.9	32.5	26.6	18.8	29.6
	P	22713	35.1	35.5	34.7	41.6	12.2	24.3	34.4	23.2	31.2	36.6	16.7	14.3	38.0	24.9	43.3	5.4	36.3
	A	7435	11.5	10.0	12.9	14.7	1.2	3.9	16.5	3.1	9.8	12.3	1.3	3.0	12.7	4.2	17.4	.7	11.9
Total		64728	100.0	31783	32945	42645	7023	8929	2608	750	2773	59732	4996	8005	56723	28803	35925	2482	62246

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-44 Percentage of Students in Each Performance Level by Subgroup, Mathematics (cont.)

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
6	BB	17494	26.7	25.3	28.1	17.6	62.4	40.9	23.3	47.1	32.8	24.6	58.1	65.3	21.4	42.2	14.9	82.1	24.3
	B	20157	30.8	32.2	29.4	30.4	26.5	34.9	32.3	33.0	31.9	30.6	33.1	22.9	31.9	33.4	28.8	15.2	31.5
	P	23438	35.8	36.4	35.2	43.5	10.5	22.0	33.5	17.6	29.6	37.6	8.6	10.5	39.3	22.3	46.1	2.7	37.3
	A	4381	6.7	6.1	7.3	8.4	.7	2.2	10.9	2.3	5.7	7.1	.2	1.2	7.5	2.1	10.2	.0	7.0
Total		65470	100.0	31930	33540	43555	6907	8887	2622	809	2690	61383	4087	7986	57484	28327	37143	2792	62678
7	BB	20585	32.2	30.9	33.4	21.9	72.9	48.0	31.7	58.0	39.0	29.9	69.2	73.6	26.5	49.6	19.5	87.4	29.7
	B	18546	29.0	31.0	27.1	30.4	18.7	30.5	28.3	24.4	28.9	29.3	24.5	17.5	30.6	29.2	28.9	10.7	29.8
	P	21784	34.1	34.0	34.1	41.7	8.1	20.2	30.8	16.3	28.3	35.8	6.1	8.0	37.6	20.0	44.3	1.8	35.5
	A	3058	4.8	4.1	5.4	6.0	.3	1.4	9.2	1.2	3.8	5.1	.1	.9	5.3	1.3	7.3	.0	5.0
Total		63973	100.0	31123	32850	42875	6562	8729	2554	802	2451	60280	3693	7711	56262	26941	37032	2731	61242
8	BB	18018	28.6	25.5	31.5	19.6	66.9	45.2	22.6	48.0	33.5	26.6	65.9	69.7	23.1	45.9	17.1	82.4	26.3
	B	22466	35.6	37.7	33.6	37.3	24.2	36.0	32.6	36.0	37.4	36.0	28.5	23.2	37.3	35.4	35.7	16.2	36.4
	P	17566	27.8	29.5	26.2	33.3	7.9	16.0	30.2	14.7	22.5	29.0	5.2	5.8	30.8	16.1	35.6	1.4	28.9
	A	5058	8.0	7.3	8.7	9.8	1.1	2.8	14.5	1.3	6.7	8.4	.4	1.2	8.9	2.6	11.6	.0	8.3
Total		63108	100.0	30838	32270	43213	6303	8122	2461	756	2253	60031	3077	7432	55676	25116	37992	2512	60596

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-45 Percentage of Students in Each Performance Level by Subgroup, Science

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
4	BB	9528	15.0	14.7	15.2	7.4	44.5	25.5	15.8	25.8	16.7	13.5	30.1	37.9	11.6	24.7	7.0	51.7	15.0
	B	20512	32.2	33.6	30.9	28.0	39.8	42.3	38.1	44.9	36.0	30.8	46.6	36.5	31.6	40.1	25.9	20.7	32.3
	P	21176	33.3	33.8	32.8	39.2	13.1	24.7	29.3	21.9	30.7	34.6	19.9	18.0	35.5	25.9	39.3	24.1	33.3
	A	12395	19.5	17.9	21.0	25.4	2.6	7.6	16.9	7.4	16.6	21.1	3.4	7.6	21.2	9.3	27.8	3.4	19.5
Total		63611	100.0	31046	32565	41521	6982	8762	2767	759	2820	57875	5736	8127	55484	28498	35113	29	63582
8	BB	11198	17.8	16.2	19.3	10.6	50.7	29.2	14.6	28.4	21.7	16.1	49.3	49.3	13.6	30.4	9.4	57.6	17.7
	B	17841	28.3	29.2	27.4	26.0	32.7	35.7	30.3	38.2	28.4	27.7	39.6	31.0	27.9	34.3	24.3	24.2	28.3
	P	19867	31.5	33.3	29.8	35.8	12.9	24.8	28.2	24.3	30.5	32.6	10.0	14.1	33.8	24.4	36.2	9.1	31.5
	A	14156	22.4	21.3	23.5	27.6	3.8	10.2	26.9	9.2	19.3	23.5	1.1	5.6	24.7	10.9	30.1	9.1	22.5
Total		63062	100.0	30818	32244	43208	6275	8113	2462	754	2250	59991	3071	7398	55664	25089	37973	33	63029

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-46 Percentage of Students in Each Performance Level by Subgroup, Social Studies

Grade	Performance Level	Examinees		Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Using Accommodations	No Accommodations
4	BB	15291	24.0	22.2	25.8	15.4	57.0	36.4	24.4	40.9	26.8	22.3	41.2	52.2	19.9	37.2	13.3	63.3	24.0
	B	14927	23.5	24.4	22.6	22.0	23.9	28.1	26.4	28.4	25.4	22.6	31.9	22.9	23.6	27.1	20.5	13.3	23.5
	P	19324	30.4	31.8	29.0	34.2	14.7	26.2	30.4	21.8	28.3	31.2	22.5	16.5	32.4	24.9	34.8	16.7	30.4
	A	14061	22.1	21.6	22.6	28.4	4.4	9.2	18.7	8.9	19.5	23.9	4.4	8.4	24.1	10.8	31.3	6.7	22.1
Total		63603	100.0	31041	32562	41522	6978	8761	2766	756	2820	57867	5736	8131	55472	28494	35109	30	63573
8	BB	14024	22.2	20.3	24.1	14.8	55.6	34.5	19.9	36.0	25.6	20.3	59.7	59.4	17.3	36.9	12.6	66.7	22.2
	B	16491	26.2	26.9	25.4	25.2	26.2	30.8	25.5	33.6	26.0	26.0	29.0	23.9	26.5	29.9	23.7	18.2	26.2
	P	20329	32.2	33.7	30.9	36.2	14.2	26.1	32.9	23.2	32.0	33.3	10.7	12.5	34.9	24.6	37.3	9.1	32.3
	A	12201	19.4	19.1	19.6	23.8	4.0	8.7	21.8	7.2	16.4	20.3	.6	4.3	21.4	8.6	26.5	6.1	19.4
Total		63045	100.0	30811	32234	43202	6273	8112	2459	753	2246	59974	3071	7402	55643	25072	37973	33	63012
10	BB	19657	31.0	27.7	34.1	23.3	70.1	48.0	28.9	45.3	35.5	29.0	77.9	71.0	26.1	48.9	21.4	46.2	31.0
	B	14936	23.5	24.8	22.4	24.1	16.6	24.2	23.5	28.3	23.7	23.8	16.3	16.3	24.4	23.7	23.4	26.9	23.5
	P	16579	26.1	28.3	24.0	29.4	9.8	18.9	26.0	18.8	24.1	27.0	4.9	8.9	28.2	18.8	30.0	19.2	26.1
	A	12304	19.4	19.2	19.5	23.2	3.5	8.9	21.6	7.6	16.7	20.1	.8	3.9	21.3	8.6	25.1	7.7	19.4
Total		63476	100.0	31053	32423	45451	5261	7668	2439	713	1944	60984	2492	6824	56652	22105	41371	26	63450

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-47a 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	61091	A	Reading—Key Ideas and Details	6	3	12	6.79	0.57	2.98	56.60	22.55
	61091	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	6	0	6	2.66	0.44	1.44	44.83	17.25
	61091	C	Reading—Vocabulary Use	2	1	4	2.24	0.56	1.17	56.01	21.15
	61091	D	Writing/Language—Text Types and Purposes	1	2	11	5.02	0.61	2.37	45.69	17.78
	61091	E	Writing/Language—Research	4	1	6	3.42	0.57	1.59	56.82	20.93
	61091	F	Writing/Language—Language Conventions	4	2	7	4.01	0.56	1.96	57.29	23.76
	61091	G	Listening	3	2	7	4.42	0.61	1.76	62.62	20.45
4	63528	A	Reading—Key Ideas and Details	2	5	12	6.30	0.54	2.64	52.54	19.42
	63528	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	6	0	6	3.24	0.54	1.66	54.33	22.37
	63528	C	Reading—Vocabulary Use	6	0	6	3.69	0.62	1.72	61.40	24.18
	63528	D	Writing/Language—Text Types and Purposes	1	3	11	4.61	0.57	2.38	42.00	17.43
	63528	E	Writing/Language—Research	3	2	6	3.40	0.55	1.63	56.66	21.81
	63528	F	Writing/Language—Language Conventions	3	2	7	4.60	0.63	1.48	65.42	15.54
	63528	G	Listening	4	2	8	3.87	0.53	1.93	48.70	18.34

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-47a 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	64654	A	Reading—Key Ideas and Details	6	2	10	5.69	0.56	2.53	57.18	22.47
	64654	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	7	1	8	4.87	0.61	2.19	61.03	23.92
	64654	C	Reading—Vocabulary Use	3	2	6	3.80	0.63	1.30	63.15	14.43
	64654	D	Writing/Language—Text Types and Purposes	3	1	11	4.68	0.54	2.04	42.87	14.71
	64654	E	Writing/Language—Research	1	3	6	3.61	0.62	1.69	59.96	22.81
	64654	F	Writing/Language—Language Conventions	3	2	7	4.49	0.64	1.64	63.87	17.41
	64654	G	Listening	4	2	8	4.97	0.63	2.01	61.93	20.68
6	65386	A	Reading—Key Ideas and Details	2	5	12	6.94	0.57	2.67	57.84	19.76
	65386	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	5	1	7	3.90	0.56	1.66	55.81	18.30
	65386	C	Reading—Vocabulary Use	3	2	5	3.00	0.60	1.42	59.87	22.92
	65386	D	Writing/Language—Text Types and Purposes	3	1	11	5.02	0.60	2.05	45.88	14.77
	65386	E	Writing/Language—Research	1	3	6	3.22	0.50	1.51	53.70	18.50
	65386	F	Writing/Language—Language Conventions	3	2	7	4.03	0.54	1.59	57.70	16.82
	65386	G	Listening	4	2	8	5.09	0.63	1.96	63.16	19.96

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-47a 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	63878	A	Reading—Key Ideas and Details	4	3	9	4.92	0.53	2.30	54.76	22.70
	63878	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	4	3	10	4.93	0.49	2.43	49.59	20.20
	63878	C	Reading—Vocabulary Use	1	3	5	3.48	0.70	1.35	69.03	22.08
	63878	D	Writing/Language—Text Types and Purposes	1	2	11	4.94	0.55	2.10	45.32	15.48
	63878	E	Writing/Language—Research	3	2	7	4.22	0.61	1.68	60.14	18.83
	63878	F	Writing/Language—Language Conventions	4	1	6	3.88	0.65	1.40	64.37	17.06
	63878	G	Listening	2	3	8	4.95	0.61	2.02	61.33	20.40
8	63056	A	Reading—Key Ideas and Details	5	3	11	6.56	0.59	2.80	59.55	23.09
	63056	B	Reading—Craft & Structure/Integration of Knowledge & Ideas	8	0	8	4.35	0.55	1.94	54.62	19.75
	63056	C	Reading—Vocabulary Use	3	1	5	2.84	0.57	1.51	56.80	24.38
	63056	D	Writing/Language—Text Types and Purposes	3	1	11	5.36	0.62	2.30	49.03	17.47
	63056	E	Writing/Language—Research	5	1	7	4.06	0.56	1.80	57.99	20.96
	63056	F	Writing/Language—Language Conventions	2	2	6	3.90	0.64	1.28	64.84	15.01
	63056	G	Listening	2	3	8	4.76	0.60	2.04	59.44	20.53

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-47b 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	61091	Listening	3	2	7	4.42	0.61	1.76	62.62	20.45
	61091	Reading	14	4	22	11.69	0.53	4.68	53.22	20.12
	61091	Writing	9	5	24	12.44	0.58	4.98	51.90	19.78
4	63528	Listening	4	2	8	3.87	0.53	1.93	48.70	18.34
	63528	Reading	14	5	24	13.24	0.56	5.21	55.16	20.92
	63528	Writing	7	7	24	12.60	0.58	4.47	52.50	17.33
5	64654	Listening	4	2	8	4.97	0.63	2.01	61.93	20.68
	64654	Reading	16	5	24	14.36	0.60	5.14	59.94	20.48
	64654	Writing	7	6	24	12.77	0.60	4.38	53.33	16.95
6	65386	Listening	4	2	8	5.09	0.63	1.96	63.16	19.96
	65386	Reading	10	8	24	13.84	0.58	4.94	57.70	19.73
	65386	Writing	7	6	24	12.27	0.55	4.13	51.25	15.87
7	63878	Listening	2	3	8	4.95	0.61	2.02	61.33	20.40
	63878	Reading	9	9	24	13.32	0.55	5.24	55.57	21.02
	63878	Writing	8	5	24	13.04	0.61	4.19	54.44	16.19
8	63056	Listening	2	3	8	4.76	0.60	2.04	59.44	20.53
	63056	Reading	16	4	24	13.75	0.57	5.48	57.32	22.00
	63056	Writing	10	4	24	13.32	0.60	4.45	55.60	17.39

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-48 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	61210	A	Operations and Algebraic Thinking	5	4	9	4.82	0.54	2.34	53.70	23.44
	61210	B	Number and Operations in Base Ten	5	3	8	4.93	0.62	2.35	61.54	26.65
	61210	C	Number and Operations—Fractions	4	4	8	4.23	0.53	2.09	53.02	22.75
	61210	D	Measurement and Data	5	5	10	5.38	0.54	2.58	53.85	23.28
	61210	E	Geometry	4	3	7	4.19	0.60	2.02	59.74	24.91
4	63630	A	Operations and Algebraic Thinking	8	2	10	4.66	0.47	2.06	46.65	17.30
	63630	B	Number and Operations in Base Ten	5	4	9	4.52	0.50	2.47	50.28	24.71
	63630	C	Number and Operations—Fractions	7	3	10	4.59	0.46	2.85	46.10	26.12
	63630	D	Measurement and Data	8	2	10	4.88	0.49	2.39	48.98	21.25
	63630	E	Geometry	4	3	7	3.51	0.50	1.91	50.10	21.53
5	64728	A	Operations and Algebraic Thinking	4	5	9	4.21	0.47	2.34	46.65	23.01
	64728	B	Number and Operations in Base Ten	4	5	9	4.35	0.48	2.44	48.31	24.19
	64728	C	Number and Operations—Fractions	7	2	9	3.87	0.43	2.29	43.18	22.36
	64728	D	Measurement and Data	7	3	10	4.22	0.42	2.48	42.41	21.40
	64728	E	Geometry	5	4	9	3.95	0.44	2.39	44.03	23.23

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-48 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	65470	E	Geometry	5	2	7	2.65	0.38	1.86	38.53	21.91
	65470	F	Ratios and Proportional Relationships	3	4	7	3.16	0.45	1.86	45.21	23.23
	65470	G	The Number System	7	3	10	5.24	0.53	2.73	52.27	25.16
	65470	H	Expressions and Equations	8	3	11	5.15	0.47	2.90	46.88	23.90
	65470	I	Statistics and Probability	8	3	11	5.26	0.48	2.33	47.81	17.73
7	63973	E	Geometry	7	3	10	3.74	0.38	1.96	37.64	15.46
	63973	F	Ratios and Proportional Relationships	6	2	8	4.15	0.52	2.26	51.66	25.51
	63973	G	The Number System	4	3	7	2.79	0.40	2.01	40.32	24.89
	63973	H	Expressions and Equations	7	3	10	4.11	0.41	2.41	41.35	20.94
	63973	I	Statistics and Probability	7	4	11	5.15	0.47	2.72	46.71	22.07
8	63108	E	Geometry	7	3	10	4.45	0.45	2.59	44.35	22.61
	63108	G	The Number System	5	3	8	3.30	0.41	2.26	41.60	24.28
	63108	H	Expressions and Equations	7	3	10	4.25	0.43	2.55	42.86	22.57
	63108	I	Statistics and Probability	6	2	8	3.47	0.44	1.86	43.47	18.85
	63108	J	Functions	7	3	10	4.50	0.45	2.50	45.01	22.65

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-49 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	63611	A	Life Science	2	10	12	6.79	0.57	2.64	56.57	20.00
	63611	B	Physical Science	4	8	12	7.32	0.61	2.63	60.97	19.62
	63611	C	Earth and Space Science	3	5	8	3.85	0.48	1.92	48.51	19.65
	63611	D	Engineering	1	7	8	4.41	0.55	2.23	55.13	24.66
8	63062	A	Life Science	1	10	11	5.57	0.51	2.90	50.79	24.13
	63062	B	Physical Science	2	8	10	4.64	0.47	2.46	46.78	21.87
	63062	C	Earth and Space Science	2	8	10	5.62	0.56	2.28	56.12	19.49
	63062	D	Engineering	1	8	9	5.77	0.64	2.27	63.73	22.50

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-50 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	63603	A	Geography	8	1	9	5.61	0.62	2.14	62.53	20.01
	63603	B	History	8	1	9	6.53	0.73	2.20	72.49	22.26
	63603	C	Political Science and Citizenship	6	0	6	3.69	0.62	1.55	61.76	20.22
	63603	D	Economics	5	1	6	3.48	0.58	1.60	58.47	21.81
	63603	E	The Behavioral Sciences	7	1	8	5.88	0.74	1.97	73.23	22.14
8	63045	A	Geography	9	1	10	7.07	0.71	2.42	70.59	21.81
	63045	B	History	11	1	12	8.29	0.69	2.79	69.17	21.37
	63045	C	Political Science and Citizenship	5	1	6	3.82	0.64	1.51	64.02	19.47
	63045	D	Economics	4	2	6	3.62	0.61	1.66	60.50	22.95
	63045	E	The Behavioral Sciences	6	0	6	4.01	0.67	1.45	67.07	19.23
10	63476	A	Geography	10	0	10	6.58	0.67	2.34	65.63	20.25
	63476	B	History	13	0	13	8.29	0.64	3.24	63.91	23.05
	63476	C	Political Science and Citizenship	9	1	10	5.70	0.58	2.55	57.37	22.86
	63476	D	Economics	7	1	8	4.88	0.62	2.02	61.02	21.60
	63476	E	The Behavioral Sciences	7	2	9	5.42	0.61	2.15	60.31	20.22

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-51 SPI Cut Scores, English Language Arts

Content Standard/Domain	Performance Level	Grade 3		Grade 4		Grade 5	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Reading—Key Ideas and Details	1	0	36	0	37	0	38
	2	37	66	38	56	39	64
	3	67	87	57	77	65	90
	4	88	100	78	100	91	100
Reading—Craft & Structure	1	0	30	0	33	0	42
	2	31	45	34	57	43	71
	3	46	75	58	85	72	93
	4	76	100	86	100	94	100
Reading—Vocabulary Use	1	0	39	0	39	0	54
	2	40	62	40	68	55	67
	3	63	87	69	92	68	81
	4	88	100	93	100	82	100
Writing/Language—Text Types and Purposes	1	0	32	0	28	0	32
	2	33	51	29	45	33	45
	3	52	69	46	62	46	63
	4	70	100	63	100	64	100
Writing/Language—Research	1	0	39	0	38	0	43
	2	40	65	39	60	44	69
	3	66	85	61	86	70	89
	4	86	100	87	100	90	100
Writing/Language—Language Conventions	1	0	36	0	54	0	52
	2	37	67	55	69	53	70
	3	68	89	70	84	71	86
	4	90	100	85	100	87	100
Listening	1	0	47	0	34	0	46
	2	48	71	35	51	47	69
	3	72	90	52	72	70	89
	4	91	100	73	100	90	100
Reading	1	0	35	0	36	0	43
	2	36	60	37	59	44	67
	3	61	84	60	83	68	89
	4	85	100	84	100	90	100
Writing	1	0	35	0	38	0	41
	2	36	59	39	56	42	58
	3	60	79	57	74	59	76
	4	80	100	75	100	77	100

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-51 SPI Cut Scores, English Language Arts (cont.)

Content Standard/Domain	Performance Level	Grade 6		Grade 7		Grade 8	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Reading—Key Ideas and Details	1	0	41	0	33	0	42
	2	42	64	34	58	43	70
	3	65	83	59	84	71	89
	4	84	100	85	100	90	100
Reading—Craft & Structure	1	0	39	0	30	0	39
	2	40	62	31	50	40	63
	3	63	78	51	77	64	80
	4	79	100	78	100	81	100
Reading—Vocabulary Use	1	0	41	0	50	0	37
	2	42	68	51	77	38	69
	3	69	88	78	92	70	86
	4	89	100	93	100	87	100
Writing/Language—Text Types and Purposes	1	0	34	0	32	0	36
	2	35	49	33	46	37	55
	3	50	63	47	63	56	71
	4	64	100	64	100	72	100
Writing/Language—Research	1	0	39	0	43	0	41
	2	40	58	44	65	42	67
	3	59	76	66	82	68	85
	4	77	100	83	100	86	100
Writing/Language—Language Conventions	1	0	44	0	50	0	56
	2	45	61	51	68	57	70
	3	62	79	69	83	71	82
	4	80	100	84	100	83	100
Listening	1	0	48	0	45	0	44
	2	49	71	46	67	45	68
	3	72	85	68	84	69	85
	4	86	100	85	100	86	100
Reading	1	0	40	0	36	0	40
	2	41	64	37	59	41	67
	3	65	82	60	83	68	85
	4	83	100	84	100	86	100
Writing	1	0	38	0	40	0	42
	2	39	55	41	57	43	62
	3	56	71	58	73	63	78
	4	72	100	74	100	79	100

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-52 SPI Cut Scores, Mathematics

Content Standard/Domain	Performance Level	Grade 3		Grade 4		Grade 5	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Operations and Algebraic Thinking	1	0	30	0	30	0	28
	2	31	53	31	47	29	48
	3	54	82	48	66	49	75
	4	83	100	67	100	76	100
Number and Operations in Base Ten	1	0	32	0	24	0	27
	2	33	67	25	52	28	50
	3	68	91	53	82	51	79
	4	92	100	83	100	80	100
Number and Operations—Fractions	1	0	30	0	18	0	23
	2	31	51	19	47	24	41
	3	52	81	48	80	42	73
	4	82	100	81	100	74	100
Measurement and Data	1	0	30	0	27	0	23
	2	31	55	28	49	24	41
	3	56	81	50	76	42	70
	4	82	100	77	100	71	100
Geometry	1	0	33	0	29	0	24
	2	34	63	30	51	25	43
	3	64	88	52	77	44	75
	4	89	100	78	100	76	100

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-52 SPI Cut Scores, Mathematics (cont.)

Content Standard/Domain	Performance Level	Grade 6		Grade 7		Grade 8	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Geometry	1	0	20	0	27	0	28
	2	21	34	28	36	29	50
	3	35	78	37	68	51	79
	4	79	100	69	100	80	100
Ratios and Proportional Relationships*	1	0	28	0	35		
	2	29	47	36	62		
	3	48	82	63	89		
	4	83	100	90	100		
The Number System	1	0	32	0	21	0	21
	2	33	57	22	45	22	49
	3	58	90	46	84	50	79
	4	91	100	85	100	80	100
Expressions and Equations	1	0	26	0	26	0	24
	2	27	49	27	43	25	46
	3	50	85	44	82	47	80
	4	86	100	83	100	81	100
Statistics and Probability	1	0	35	0	32	0	30
	2	36	50	33	52	31	47
	3	51	74	53	85	48	72
	4	75	100	86	100	73	100
Functions**	1					0	28
	2					29	51
	3					52	80
	4					81	100

* Content standard in grades 6 and 7 only.
** Content standard in grade 8 only.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-53 SPI Cut Scores, Science

Content Standard/Domain	Performance Level	Grade 4		Grade 8	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Life Science	1	0	33	0	24
	2	34	55	25	47
	3	56	75	48	72
	4	76	100	73	100
Physical Science	1	0	38	0	23
	2	39	60	24	42
	3	61	79	43	64
	4	80	100	65	100
Earth and Space Science	1	0	26	0	35
	2	27	44	36	54
	3	45	67	55	72
	4	68	100	73	100
Engineering	1	0	25	0	40
	2	26	54	41	64
	3	55	79	65	83
	4	80	100	84	100

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-54 SPI Cut Scores, Social Studies

Content Standard/Domain	Performance Level	Grade 4		Grade 8		Grade 10	
		Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound	Score Lower Bound	Score Upper Bound
Geography	1	0	47	0	53	0	55
	2	48	62	54	75	56	70
	3	63	79	76	89	71	84
	4	80	100	90	100	85	100
History	1	0	54	0	51	0	49
	2	55	76	52	72	50	70
	3	77	92	73	88	71	87
	4	93	100	89	100	88	100
Political Science and Citizenship	1	0	46	0	46	0	43
	2	47	62	47	64	44	60
	3	63	78	65	82	61	80
	4	79	100	83	100	81	100
Economics	1	0	41	0	40	0	49
	2	42	58	41	61	50	66
	3	59	76	62	81	67	81
	4	77	100	82	100	82	100
The Behavioral Sciences	1	0	58	0	50	0	49
	2	59	78	51	66	50	64
	3	79	91	67	85	65	79
	4	92	100	86	100	80	100

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-55 Longitudinal Comparison of State-Level Scale Score Means: ELA

Grade	Year	N	Mean	Stand. Dev
3	2016	64107	560.57	47.31
	2017	63946	559.12	46.93
	2018	63194	556.70	46.66
	2019	61091	554.59	45.54
4	2016	62609	582.71	49.41
	2017	64423	585.26	52.44
	2018	64354	580.90	51.81
	2019	63528	582.01	51.05
5	2016	62300	599.62	51.11
	2017	62995	603.24	51.00
	2018	64903	600.78	48.35
	2019	64654	595.58	48.77
6	2016	62728	610.36	52.16
	2017	62754	614.59	49.82
	2018	63600	609.61	50.18
	2019	65386	607.00	50.15
7	2016	62084	623.84	54.85
	2017	63091	626.80	59.14
	2018	63140	627.43	56.56
	2019	63878	627.70	54.88
8	2016	61486	637.23	57.27
	2017	62109	637.69	61.61
	2018	63248	630.98	59.94
	2019	63056	629.06	59.84

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-56 Longitudinal Comparison of State-Level Scale Score Means: Mathematics

Grade	Year	N	Mean	Stand. Dev
3	2016	64194	554.28	46.47
	2017	64066	555.03	48.63
	2018	63314	555.94	50.87
	2019	61210	555.78	53.50
4	2016	62674	573.45	56.15
	2017	64533	574.33	54.92
	2018	64462	576.76	52.99
	2019	63630	577.09	51.78
5	2016	62368	599.57	50.19
	2017	63152	599.73	51.00
	2018	65021	598.82	56.65
	2019	64728	601.48	53.14
6	2016	62772	612.67	53.00
	2017	62847	612.93	54.81
	2018	63669	611.97	57.64
	2019	65470	610.77	58.31
7	2016	62144	627.49	57.40
	2017	63200	627.48	58.65
	2018	63218	622.82	65.55
	2019	63973	625.25	60.69
8	2016	61551	640.79	57.54
	2017	62175	641.11	59.36
	2018	63318	644.24	60.78
	2019	63108	644.53	57.85

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-57 Baseline Year State-Level Scale Score Means: Science

Grade	Year	N	Mean	Stand. Dev
4	2019	63611	499.88	50.24
8	2019	63062	699.70	50.55

Table 8-58 Longitudinal Comparison of State-Level Scale Score Means: Social Studies

Grade	Year	N	Mean	Stand. Dev
4	2016	62630	398.02	51.49
	2017	64512	397.05	51.71
	2018	64456	398.23	53.72
	2019	63603	396.68	55.69
8	2016	61496	598.06	51.68
	2017	62079	597.60	54.26
	2018	63230	599.17	53.25
	2019	63045	598.81	52.30
10	2016	63991	698.51	53.74
	2017	63764	696.92	56.56
	2018	62630	695.70	58.24
	2019	63476	692.82	58.30

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-59 Longitudinal Comparison of State-Level Impact Data: ELA

Grade	Year	N	Below Basic	Basic	Proficient	Advanced	Prof. & Adv.
3	2016	64107	21.99	34.88	34.29	8.84	43.13
	2017	63946	21.45	36.72	33.81	8.02	41.83
	2018	63194	22.78	37.47	32.58	7.17	39.75
	2019	61091	23.28	38.04	33.21	5.48	38.69
4	2016	62609	22.81	33.88	34.77	8.54	43.30
	2017	64423	21.14	32.14	37.00	9.71	46.72
	2018	64354	24.04	32.06	35.72	8.19	43.91
	2019	63528	23.88	33.14	34.10	8.89	42.98
5	2016	62300	23.17	34.37	34.55	7.91	42.47
	2017	62995	20.36	33.22	37.88	8.54	46.42
	2018	64903	21.53	34.30	37.40	6.77	44.17
	2019	64654	26.11	33.83	34.34	5.72	40.06
6	2016	62728	21.12	36.30	31.67	10.91	42.58
	2017	62754	18.23	36.52	33.51	11.75	45.26
	2018	63600	22.06	35.08	32.73	10.12	42.86
	2019	65386	23.56	35.48	31.87	9.09	40.96
7	2016	62084	23.11	34.91	34.09	7.89	41.98
	2017	63091	22.27	34.10	33.52	10.11	43.63
	2018	63140	21.29	33.57	35.72	9.43	45.15
	2019	63878	21.88	33.25	35.36	9.51	44.87
8	2016	61486	21.24	37.21	31.26	10.30	41.56
	2017	62109	21.66	37.22	29.19	11.93	41.12
	2018	63248	24.66	38.01	27.93	9.40	37.33
	2019	63056	25.94	37.04	28.80	8.23	37.03

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-60 Longitudinal Comparison of State-Level Impact Data: Mathematics

Grade	Year	N	Below Basic	Basic	Proficient	Advanced	Prof. & Adv.
3	2016	64194	18.59	33.41	38.90	9.10	48.00
	2017	64066	18.90	33.06	37.84	10.20	48.03
	2018	63314	18.68	31.48	38.47	11.37	49.83
	2019	61210	19.28	31.28	37.17	12.27	49.44
4	2016	62674	19.59	36.22	33.33	10.86	44.20
	2017	64533	19.13	37.37	32.67	10.83	43.50
	2018	64462	18.37	37.17	32.71	11.74	44.46
	2019	63630	18.87	36.09	32.82	12.23	45.05
5	2016	62368	25.94	29.98	34.14	9.94	44.08
	2017	63152	24.97	30.57	34.58	9.88	44.46
	2018	65021	24.73	29.32	35.05	10.90	45.95
	2019	64728	24.22	29.20	35.09	11.49	46.58
6	2016	62772	25.51	31.66	36.78	6.05	42.84
	2017	62847	24.70	31.68	37.50	6.11	43.61
	2018	63669	24.78	31.27	37.78	6.18	43.96
	2019	65470	26.72	30.79	35.80	6.69	42.49
7	2016	62144	30.45	30.28	34.81	4.45	39.26
	2017	63200	30.80	29.92	34.53	4.75	39.29
	2018	63218	31.36	29.67	34.33	4.64	38.97
	2019	63973	32.18	28.99	34.05	4.78	38.83
8	2016	61551	28.66	37.48	28.12	5.74	33.86
	2017	62175	28.43	36.95	28.33	6.29	34.62
	2018	63318	27.95	35.44	28.71	7.90	36.61
	2019	63108	28.55	35.60	27.83	8.01	35.85

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 8-61 Longitudinal Comparison of State-Level Impact Data: Science

Grade	Year	N	Below Basic	Basic	Proficient	Advanced	Prof. & Adv.
4	2016	62636	14.85	33.73	35.70	15.73	51.42
	2017	64520	15.29	33.63	34.70	16.37	51.07
	2018	64448	15.24	34.07	34.43	16.26	50.69
	2019	63611	14.98	32.25	33.29	19.49	52.78
8	2016	61471	16.31	34.07	34.36	15.27	49.63
	2017	62113	17.61	34.74	34.11	13.54	47.65
	2018	63272	17.18	33.96	34.16	14.70	48.86
	2019	63062	17.76	28.29	31.50	22.45	53.95

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

Table 8-62 Longitudinal Comparison of State-Level Impact Data: Social Studies

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 2019 Wisconsin Forward Exam was estimated in four ways:

- Internal consistency was assessed for all items using Cronbach's alpha (1951).
- Standard error of measurement (SEM) was calculated for raw score and scale score.
- Classification consistency and classification accuracy were estimated for the performance level classifications.
- Inter-rater reliability was estimated for the text-dependent analysis (TDA) items.

This part of the report addresses American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (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)

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Standard 2.3 advises providing reliability estimates and the SEM for all total scores and subscores reported, Standard 2.13 advises reporting SEM in both raw score and scale score units, and Standard 2.11 advises assessing reliability and SEM for all population subgroups. This 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 the subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. The scale score conditional standard errors of measurement (CSEMs) 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 *Proficient* or *Advanced*. As described below, the Spring 2019 Wisconsin Forward Exam adhered to this standard by applying a detailed analysis of classification consistency and classification accuracy—two related measures used to evaluate the reliability of the performance level classifications used in the test program. This analysis also addresses Standard 2.14 by providing a CSEM for the cut scores that separate the performance levels.

Standard 2.7 advises reporting measures of inter-rater consistency in which subjective judgment is involved in scoring. As discussed in Part 5, English Language Arts (ELA) TDA items were scored by the artificial intelligence (AI) engine with second reads performed by human scorers. As this section will show, a study of inter-rater consistency was conducted for the ELA TDA items. This study was conducted to evaluate 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 related to the reliability of the Wisconsin Forward Exam. Cronbach's alpha and the SEM operate at the content level. For example, they provide estimates of reliability for student scores in ELA or Mathematics. Classification consistency and classification accuracy operate on the associated performance level classifications. These are of particular interest in the context of the Elementary and Secondary Education Act and the associated accountability requirements. Inter-rater reliability probes further, looking at individual items and evaluating the reliability of the AI engine scores 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 present 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, which is targeted at each intended use of the Wisconsin Forward Exam scores, addresses Standard 2.0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

9.1 Measures of Internal Consistency and Standard Error of Measurement

Cronbach's alpha is a frequently used measure of internal consistency for tests consisting of multiple-choice (MC) and constructed-response (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 (SD) 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, English language proficiency, and accommodation use.

As indicated in Table 9-1, the test reliability was highest for Mathematics tests. As shown in the "Total" column, reliability ranges from 0.88 to 0.90 across grades for ELA, from 0.91 to 0.93 across grades for Mathematics, from 0.89 to 0.90 for Science, and from 0.89 to 0.92 across grades for Social Studies. All reliability coefficients would ideally be 0.90 or above. However, for relatively short tests that are designed to measure a fairly broad range of content, this is not always a realistic expectation. If 0.90 is considered a conservative criterion for an acceptable level of reliability, as measured by Cronbach's alpha, then the ELA grade 4, 6, and 7 assessments, the Science grade 4 assessment, and the Social Studies grade 4 assessment would not meet this criterion. The reliability coefficients for these tests may be affected by the number of items (and score points) and the diversity of the content being assessed. Applying the Spearman-Brown prophecy formula to these results indicates that to achieve the 0.90 reliability threshold, the current ELA assessments for grades 4, 6, and 7 would need to be increased from 56 points to 61, 66, and 63 score points, respectively. For the current Science assessments in grade 4, the increase would need to be from 40 to 45 score points. For the current Social Studies assessment in grade 4, the increase would need to be from 38 to 41 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 score distribution and variance, so this result is consistent with the generally larger SDs (as previously discussed in Part 8 of this report and summarized in Tables 8-19 through 8-27) among many of these subgroups.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Most differences among the five racial/ethnic groups were also quite small, within 0.05 of one another for all grades in ELA, Science, and Social Studies. In Mathematics, higher test reliabilities were observed for White or Asian students and the lowest reliability was observed for African-American students in grades 4 through 8.

The differences between economically disadvantaged and not economically disadvantaged students were within 0.03 of one another for all grades and content areas. The differences between disabled and not disabled students were within 0.03 of one another for all grades and content areas, except for Mathematics grades 6 through 8, where the differences were at or greater than 0.05. The greatest differences were between fully English proficient and limited English proficient students and between students using and not using testing accommodations, with consistently lower reliability among limited English proficient students and students using testing accommodations. The test reliability coefficients for limited English proficient students were lower than for other subgroups for most grades in ELA, Science, and Social Studies. The reliability coefficients for students using testing accommodations in Mathematics were the lowest of all subgroups. The reliability coefficients for students using testing accommodations in ELA should be interpreted with caution because of the low number of students using those accommodations. The reliability coefficients were not computed for students using testing accommodations in Science or Social Studies, because the number of students using testing accommodations in these subject areas was less than 50. The reliability coefficient is affected, among other factors, by the variability of the students' scores. The higher the variability of scores, the higher the reliability coefficient will tend to be. Based on the evaluation of the distribution of the limited English proficient student test scores and Mathematics scores for students using testing accommodations, it was observed that the variance of these scores was often lower than the variance of the scores for other groups. The limited English proficient student groups and students using testing accommodations in Mathematics appear to be more homogeneous on the ability being measured by the test, leading to lower test reliability for these groups of 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 an individual's obtained score can be expected to fall within two SEMs 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 and Social Studies grade 10 were slightly larger than those for the other content areas. Because these SEMs are in the raw score metric, this result is consistent with the fact that ELA tests and the Social Studies grade 10 test have more raw score points and relatively larger raw score SDs than other content areas. For every grade and content area, the CSEM for individual scale scores is provided in the scoring tables previously discussed in Part 6 (Tables 6-8 through 6-24).

Reliability, as measured by Cronbach's alpha, and SEM were also computed for content standards within each content area as well as for each language domain in ELA.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 9-3 shows these reliability coefficients by content standard and domain. The last column presents the reliability for the total test per grade for each content area (with all content standards or domains) for all examinees. It is clear that the reliability per content standard or domain is lower than the reliability for the total test per content area. The number of items or score points has a close relationship with reliability, and a smaller number of items or score points is generally associated with lower reliability. The number of score points for ELA per domain was 7 or 8 in Listening, 22 or 24 in Reading, and 24 in Writing. The number of score points ranged from 4 to 12 per standard for ELA, from 7 to 11 per standard for Mathematics, from 8 to 12 per standard for Science, and from 6 to 13 per standard for Social Studies. A lower level of reliability per content standard or domain is therefore expected. The lower level of reliability per standard or domain is one of the reasons why the information based on the content standards or domains should be used for low-stakes purposes only (this issue was previously discussed in the context of SPI).

As shown in Table 9-3, the reliability ranges by content standard/domain were as follows:

- For ELA, reliability indices by content standard or domain ranged from 0.34 (for standard C in grade 5) to 0.83 (for the Reading domain in grades 4, 5, and 8).
- For Mathematics, reliability indices by content standard ranged from 0.53 (for standard E in grade 7) to 0.81 (for standard C in grade 4).
- For Science, reliability indices by content standard ranged from 0.56 (for standard C in grade 4) to 0.75 (for standard A in grade 8).
- For Social Studies, reliability indices by content standard ranged from 0.50 (for standard C in grade 8) to 0.78 (for standard B in grade 10).

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 were generally consistent with the number of items within each content standard.

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

9.1.1 Conditional Standard Error of Measurement

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

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

where $I(\theta_i)$ is the test information function, as a sum of item information function 2, obtained as

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

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

Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scale scores) and are smaller in the middle of the score distribution and larger at the tails. This pattern is seen for all Wisconsin Forward Exam CSEMs and is to be expected when item response theory (IRT) methods are used. In compliance with Standard 2.14, the CSEM of each cut score was presented in the raw score-to-scale score tables (Tables 6-8 through 6-24) for all grades and content areas in Part 6 of this report. In addition, graphical representation of the CSEM with the cut scores is presented in Figures I-1 through I-17 of Appendix I for all grades and content areas. As shown in Appendix I, the estimates of CSEM tend to be higher at the low and high ends of the scale score range. The CSEM increases when there are few observations at a particular ability level. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. Figures I-1 through I-17 demonstrate that the CSEM is minimized at the cut scores and in the middle of the scale range, where most students are located.

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 are of particular interest. The following section describes how the consistency and accuracy of these classifications were evaluated and provides evidence that supports the validity of these classifications.

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 equal to or greater than 0.75 indicate "excellent agreement," values between 0.40 and 0.74 represent "good agreement" beyond chance, and values below 0.40 denote "poor agreement."

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

9.2.1 Kolen and Kim's Method for Pattern Scoring

As stated in Part 6, when IRT is applied to score examinees' responses, two types of scoring are available: number-correct scoring and item-pattern scoring. The Wisconsin Forward Exam uses item-pattern scoring. Many methods of estimating the consistency and accuracy of classification based on number-correct scoring have been suggested in psychometric literature.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 by using the two ability estimates obtained in Step 2. Note that this table is constructed for each quadrature point and replication. One, and only one, cell will have a value of one and zeros elsewhere.

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

	First Administration, or Form 1		
	$\hat{\theta}_{j1} \geq C_1$	$\hat{\theta}_{j1} < C_1$	
$\hat{\theta}_{j2} \geq C_1$			Second Administration, 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 the examinees' abilities are estimated at each quadrature point, these quadrature points can be considered the true scores. Therefore, classification accuracy is computed using the

Copyright © 2019 by the Wisconsin Department of Public Instruction

examinees' estimated abilities (observed scores) and quadrature points (true scores). Just as 0.90 is generally considered the criterion for acceptable test score reliability, the criterion value of 0.90 is considered to be an acceptably high level of classification accuracy.

In Tables 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, when considering the first column of data in the ELA grade 3 table, 0.20 represents the probability of belonging to *Below Basic* in both the first and second administrations. The 0.04 value represents the probability of belonging to *Basic* in the first administration and *Below Basic* in the second administration. The probability of belonging to *Proficient* or *Advanced* in the first administration and *Below Basic* in the second administration is 0.00. "Sum" is obtained simply by adding the four row values or the four column values. Because the values displayed have been rounded to two decimal places, this sum is not always identical to the sum of the values shown in the table.

The second table shows indices for classification consistency and classification accuracy. Because there are four performance levels for the Wisconsin Forward Exam, there are three cut scores. The values in "All Cuts" were obtained by applying all three cuts together. In Table 9-5 for ELA grade 3, when all three cuts were used for the computation, classification consistency (P) is 0.75, probability of chance is 0.30, kappa (k) is 0.64, and classification accuracy is 0.83. 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 0.76, 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.

As shown in Tables 9-5 through 9-21, when only the *Proficient* cut score was applied, the classification consistency (P) was greater than or equal to 0.87 and the classification accuracy was greater than or equal to 0.91 for all tests. The kappa value was greater than or equal to 0.75 for all tests. According to Landis and Koch's criteria for k (presented previously in this report in the discussion of classification consistency), all tests showed excellent agreement based on the cut for the *Proficient* performance level.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Tables J-1 through J-17, when the *Proficient* cut is considered, classification consistency, accuracy coefficients, and kappa values were good or very good for all subgroups, grades, and content areas. Specifically for ELA, the classification consistency was greater than or equal to 0.85 and the classification accuracy was greater than or equal to 0.90 for all subgroups across all grades. For Mathematics, the classification consistency was greater than or equal to 0.88 and the classification accuracy was greater than or equal to 0.91 for all subgroups across all grades. For Science, the classification consistency was greater than or equal to 0.87 and the classification accuracy was greater than or equal to 0.91 for all subgroups across both grades. For Social Studies, the classification consistency was greater than or equal to 0.86 and the classification accuracy was greater than or equal to 0.91 for all subgroups across all grades. The kappa values were greater than or equal to 0.57 for all subgroups in ELA, greater than or equal to 0.65 for all subgroups in Mathematics and Science, and greater than or equal to 0.62 for all subgroups in Social Studies. The lowest kappa values were observed for the limited English proficiency subgroups in ELA, Science, and Social Studies and for students using testing accommodations in Mathematics. This is consistent with the trend of the test reliability coefficients, which were found to be lower for these groups of students compared to other subgroups. Because the number of students using testing accommodations in Science and Social Studies was less than 50 per grade, the indices for classification consistency and classification accuracy were not computed for students using testing accommodations in these subject areas. The indices for classification consistency and classification accuracy for students using testing accommodations in ELA should be interpreted with caution because of the low number of students using accommodations in this content area.

9.3 Inter-rater Reliability for TDA Items

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 9-C. 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.

Table 9-D. Data Structure 2: Cross-Tabulation of Score 1 and Score 2. As an alternative, a square table of counts may be created for each Score 1 by Score 2 (i.e., $X_{n1} \times X_{n2}$) combination:

		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_{++}

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

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

If agreement is perfect, 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., 1, 2, 3, 4) used in scoring TDA items contain a certain level of chance agreement that is expected. Although the intraclass correlation is reported in this report, it does not take into account the possibility of chance agreement between the two raters. Cohen's kappa does take this into consideration. In general, k will have values equal to or less than the intraclass correlation. If agreement is perfect, the value of k is 1.0. If agreement is at chance levels, the value of k is 0. As noted in Section 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, "Exact" agreement is defined as scores that are exactly the same. "Adjacent" agreement is defined as scores differing by 1 point.

Copyright © 2019 by the Wisconsin Department of Public Instruction

“Discrepant” cases are those cases in which the scores of the two raters differed by more than one raw score point. For example, as shown in Table 9-22, for the grade 3 TDA item, the exact agreement, adjacent agreement, and discrepant agreement rates are 72.58%, 27.40%, and 0.02%, respectively. “Mean” reflects the item mean score from the second reads, which are done by human scorers. “No. of Second Reads” is the number of student responses selected for the purpose of the second read and computing inter-rater reliability. The “Score Frequency” columns represent the scoring outcomes for the student responses based on the raw scores given by the human scorers. The column for “Codes” reflects the number of students who received the condition codes B, C, N, R, or T (described in detail in Part 5, Table 5-2 of this report).

Overall, the exact rater agreement percentages were good for all TDA items and ranged from 72.58% in grade 3 to 81.61% in grade 6. The combined exact and adjacent agreement percentages were over 99% in all grades. The intraclass correlation coefficients ranged from 0.84 in grade 5 to 0.90 in grade 8. The weighted kappa ranged from 0.69 in grade 5 to 0.79 in grade 8, indicating good or excellent rater agreement for all TDA items.

9.4 Summary

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
			Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Students with Accommodations	Students without Accommodations
English Language Arts	3	0.90	0.90	0.89	0.88	0.87	0.87	0.90	0.86	0.89	0.90	0.85	0.88	0.89	0.88	0.88	0.90	0.89
	4	0.89	0.89	0.89	0.88	0.85	0.86	0.89	0.86	0.89	0.89	0.82	0.87	0.89	0.87	0.88	0.89	0.89
	5	0.90	0.90	0.90	0.89	0.87	0.88	0.90	0.87	0.90	0.90	0.82	0.87	0.89	0.88	0.89	0.92	0.90
	6	0.88	0.88	0.89	0.87	0.85	0.87	0.88	0.86	0.88	0.88	0.79	0.85	0.87	0.87	0.87	0.89	0.88
	7	0.89	0.88	0.89	0.88	0.86	0.87	0.89	0.87	0.89	0.89	0.79	0.85	0.88	0.88	0.87	0.87	0.89
	8	0.90	0.90	0.90	0.89	0.88	0.89	0.90	0.88	0.90	0.90	0.82	0.86	0.89	0.89	0.89	0.92	0.90
Mathematics	3	0.93	0.92	0.93	0.92	0.90	0.91	0.93	0.90	0.92	0.93	0.90	0.92	0.92	0.92	0.91	0.85	0.93
	4	0.92	0.92	0.92	0.91	0.86	0.89	0.93	0.88	0.91	0.92	0.88	0.90	0.92	0.90	0.91	0.79	0.92
	5	0.92	0.91	0.92	0.91	0.86	0.89	0.93	0.89	0.92	0.92	0.85	0.89	0.91	0.90	0.91	0.80	0.92
	6	0.92	0.91	0.92	0.91	0.84	0.89	0.93	0.88	0.91	0.92	0.78	0.86	0.91	0.89	0.91	0.69	0.92
	7	0.91	0.90	0.91	0.90	0.82	0.87	0.92	0.87	0.91	0.91	0.75	0.84	0.91	0.88	0.91	0.63	0.91
	8	0.92	0.91	0.92	0.91	0.84	0.88	0.93	0.85	0.91	0.92	0.76	0.83	0.91	0.88	0.91	0.60	0.91
Science	4	0.89	0.88	0.89	0.87	0.83	0.86	0.88	0.85	0.88	0.89	0.82	0.88	0.88	0.87	0.87	-	0.89
	8	0.90	0.89	0.91	0.89	0.84	0.87	0.90	0.87	0.90	0.90	0.76	0.87	0.89	0.88	0.88	-	0.90
Social Studies	4	0.89	0.89	0.90	0.88	0.87	0.87	0.88	0.87	0.89	0.89	0.85	0.89	0.88	0.88	0.87	-	0.89
	8	0.90	0.89	0.91	0.89	0.88	0.89	0.90	0.88	0.90	0.90	0.83	0.88	0.89	0.89	0.88	-	0.90
	10	0.92	0.91	0.93	0.91	0.89	0.90	0.92	0.90	0.92	0.92	0.81	0.89	0.91	0.91	0.91	-	0.92

Note: The reliability coefficients were not computed for students using testing accommodations in Science or Social Studies because the number of students using testing accommodations in these subject areas was less than 50 per grade.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Content	Grade	Total	Gender		Race/Ethnicity						ELP		Disability		SES		Accommodations	
			Female	Male	White	African-American	Hispanic	Asian	American Indian	Two or More	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged	Students with Accommodations	Students without Accommodations
English Language Arts	3	3.34	3.32	3.35	3.29	3.41	3.42	3.34	3.41	3.36	3.33	3.44	3.38	3.32	3.41	3.26	3.43	3.34
	4	3.39	3.38	3.40	3.36	3.44	3.46	3.38	3.41	3.41	3.39	3.48	3.39	3.37	3.44	3.34	3.33	3.39
	5	3.29	3.25	3.30	3.23	3.39	3.36	3.27	3.35	3.29	3.27	3.41	3.37	3.25	3.37	3.19	3.35	3.28
	6	3.36	3.34	3.36	3.31	3.43	3.40	3.34	3.45	3.40	3.35	3.41	3.37	3.33	3.41	3.29	3.39	3.36
	7	3.41	3.39	3.40	3.37	3.47	3.46	3.42	3.42	3.42	3.40	3.47	3.39	3.38	3.46	3.34	3.52	3.40
	8	3.37	3.34	3.37	3.33	3.42	3.42	3.31	3.43	3.40	3.36	3.41	3.36	3.35	3.42	3.31	3.42	3.37
Mathematics	3	2.69	2.71	2.66	2.66	2.70	2.76	2.66	2.77	2.71	2.68	2.77	2.71	2.68	2.75	2.63	2.65	2.69
	4	2.81	2.82	2.80	2.83	2.71	2.80	2.77	2.78	2.82	2.81	2.79	2.75	2.82	2.80	2.81	2.67	2.81
	5	2.90	2.91	2.88	2.93	2.72	2.86	2.86	2.85	2.89	2.91	2.81	2.73	2.92	2.86	2.92	2.60	2.91
	6	2.87	2.88	2.86	2.87	2.80	2.86	2.85	2.83	2.87	2.87	2.81	2.77	2.88	2.87	2.87	2.68	2.88
	7	2.90	2.90	2.89	2.92	2.71	2.86	2.87	2.82	2.89	2.90	2.75	2.71	2.91	2.85	2.92	2.61	2.90
	8	2.89	2.90	2.87	2.91	2.74	2.86	2.84	2.86	2.88	2.89	2.77	2.72	2.90	2.86	2.90	2.64	2.89
Science	4	2.71	2.71	2.70	2.68	2.74	2.76	2.73	2.77	2.74	2.70	2.78	2.72	2.70	2.75	2.66	-	2.71
	8	2.75	2.77	2.72	2.74	2.72	2.79	2.73	2.79	2.76	2.75	2.75	2.70	2.75	2.77	2.73	-	2.75
Social Studies	4	2.56	2.56	2.56	2.48	2.79	2.71	2.59	2.75	2.61	2.54	2.78	2.75	2.53	2.71	2.44	-	2.56
	8	2.60	2.60	2.59	2.52	2.81	2.74	2.55	2.77	2.63	2.58	2.87	2.81	2.56	2.75	2.48	-	2.60
	10	3.01	3.00	3.00	2.94	3.21	3.15	2.98	3.13	3.08	3.00	3.25	3.19	2.98	3.15	2.92	-	3.01

Note: The SEMs were not computed for students using testing accommodations in Science or Social Studies because the number of students using testing accommodations in these subject areas was less than 50 per grade.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 9-3 Cronbach's Alpha Reliability Coefficients for Content Standard and Domain

English Language Arts

Grade	Alpha per Content Standard and Domain									
	A	B	C	D	E	F	G/Listening	Reading	Writing	Total Test
3	0.72	0.37	0.40	0.47	0.53	0.67	0.60	0.78	0.79	0.90
4	0.66	0.56	0.63	0.51	0.58	0.45	0.48	0.83	0.75	0.89
5	0.70	0.70	0.34	0.43	0.58	0.51	0.60	0.83	0.75	0.90
6	0.66	0.50	0.56	0.42	0.40	0.48	0.57	0.81	0.70	0.88
7	0.68	0.58	0.56	0.45	0.52	0.48	0.54	0.82	0.73	0.89
8	0.71	0.57	0.52	0.57	0.58	0.39	0.58	0.83	0.76	0.90

Mathematics

Grade	Alpha per Content Standard										
	A	B	C	D	E	F	G	H	I	J	Total Test
3	0.72	0.77	0.70	0.72	0.70						0.93
4	0.57	0.74	0.81	0.71	0.68						0.92
5	0.72	0.73	0.70	0.70	0.71						0.92
6					0.63	0.69	0.78	0.76	0.62		0.92
7					0.53	0.73	0.71	0.68	0.72		0.91
8					0.73		0.74	0.72	0.56	0.73	0.92

Science

Grade	Alpha per Content Standard				
	A	B	C	D	Total Test
4	0.71	0.70	0.56	0.70	0.89
8	0.75	0.69	0.62	0.70	0.90

Social Studies

Grade	Alpha per Content Standard					
	A	B	C	D	E	Total Test
4	0.63	0.72	0.52	0.57	0.69	0.89
8	0.74	0.75	0.50	0.59	0.52	0.90
10	0.66	0.78	0.73	0.66	0.63	0.92

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 9-4 Standard Error of Measurement per Content Standard and Domain

English Language Arts

Grade	SEM per Content Standard and Domain									
	A	B	C	D	E	F	G/Listening	Reading	Writing	Total Test
3	1.59	1.14	0.90	1.73	1.10	1.13	1.11	2.17	2.27	3.34
4	1.54	1.10	1.04	1.67	1.06	1.09	1.39	2.17	2.21	3.39
5	1.38	1.19	1.05	1.53	1.10	1.15	1.27	2.12	2.17	3.29
6	1.55	1.18	0.95	1.56	1.18	1.14	1.28	2.17	2.24	3.36
7	1.31	1.58	0.89	1.56	1.16	1.01	1.37	2.24	2.17	3.41
8	1.50	1.26	1.05	1.52	1.16	1.01	1.32	2.23	2.16	3.37

Mathematics

Grade	SEM per Content Standard										
	A	B	C	D	E	F	G	H	I	J	Total Test
3	1.23	1.12	1.15	1.35	1.10						2.69
4	1.35	1.25	1.24	1.29	1.08						2.81
5	1.24	1.28	1.26	1.35	1.29						2.90
6					1.14	1.03	1.29	1.43	1.44		2.87
7					1.35	1.18	1.09	1.36	1.43		2.90
8					1.36		1.16	1.35	1.23	1.29	2.89

Science

Grade	SEM per Content Standard				
	A	B	C	D	Total Test
4	1.43	1.45	1.28	1.22	2.71
8	1.45	1.37	1.40	1.25	2.75

Social Studies

Grade	SEM per Content Standard					
	A	B	C	D	E	Total Test
4	1.31	1.17	1.08	1.04	1.09	2.56
8	1.25	1.39	1.06	1.06	1.00	2.60
10	1.37	1.51	1.33	1.18	1.31	3.01

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.20	0.05	0.00	0.00	0.24
Basic	0.04	0.27	0.05	0.00	0.36
Proficient	0.00	0.06	0.24	0.02	0.33
Advanced	0.00	0.00	0.03	0.04	0.07
Sum	0.23	0.38	0.32	0.06	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.89	0.95	0.75
Probability of Chance	0.64	0.52	0.88	0.30
Kappa (*k*)	0.77	0.76	0.59	0.64
Classification Accuracy	0.94	0.92	0.97	0.83

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	0.19	0.05	0.00	0.00	0.24
Basic	0.06	0.22	0.06	0.00	0.33
Proficient	0.00	0.05	0.24	0.03	0.32
Advanced	0.00	0.00	0.03	0.07	0.11
Sum	0.25	0.32	0.33	0.10	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.89	0.89	0.93	0.72
Probability of Chance	0.63	0.51	0.81	0.28
Kappa (k)	0.71	0.77	0.65	0.61
Classification Accuracy	0.93	0.92	0.95	0.81

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.22	0.04	0.00	0.00	0.27
Basic	0.04	0.23	0.05	0.00	0.32
Proficient	0.00	0.06	0.25	0.03	0.34
Advanced	0.00	0.00	0.03	0.05	0.07
Sum	0.27	0.33	0.33	0.07	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.89	0.95	0.75
Probability of Chance	0.61	0.52	0.87	0.29
Kappa (k)	0.77	0.78	0.59	0.65
Classification Accuracy	0.94	0.92	0.96	0.83

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	0.20	0.05	0.00	0.00	0.24
Basic	0.05	0.24	0.06	0.00	0.35
Proficient	0.00	0.06	0.21	0.04	0.30
Advanced	0.00	0.00	0.04	0.07	0.11
Sum	0.24	0.34	0.31	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	0.92	0.71
Probability of Chance	0.63	0.51	0.81	0.28
Kappa (k)	0.74	0.75	0.60	0.60
Classification Accuracy	0.94	0.92	0.95	0.80

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.04	0.00	0.00	0.22
Basic	0.04	0.23	0.06	0.00	0.33
Proficient	0.00	0.06	0.25	0.04	0.34
Advanced	0.00	0.00	0.04	0.07	0.11
Sum	0.22	0.33	0.34	0.11	

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	0.73
Probability of Chance	0.66	0.51	0.81	0.29
Kappa (k)	0.77	0.77	0.62	0.62
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	0.22	0.05	0.00	0.00	0.27
Basic	0.04	0.26	0.05	0.00	0.36
Proficient	0.00	0.06	0.19	0.03	0.28
Advanced	0.00	0.00	0.03	0.06	0.09
Sum	0.26	0.37	0.27	0.10	

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	0.74
Probability of Chance	0.61	0.53	0.83	0.29
Kappa (k)	0.79	0.76	0.63	0.63
Classification Accuracy	0.94	0.92	0.95	0.82

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.17	0.03	0.00	0.00	0.20
Basic	0.03	0.23	0.04	0.00	0.31
Proficient	0.00	0.05	0.28	0.04	0.36
Advanced	0.00	0.00	0.03	0.10	0.13
Sum	0.20	0.31	0.36	0.14	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.94	0.91	0.93	0.78
Probability of Chance	0.68	0.50	0.77	0.28
Kappa (k)	0.80	0.81	0.70	0.69
Classification Accuracy	0.95	0.93	0.95	0.83

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	0.15	0.04	0.00	0.00	0.19
Basic	0.04	0.26	0.04	0.00	0.35
Proficient	0.00	0.05	0.25	0.03	0.33
Advanced	0.00	0.00	0.03	0.10	0.13
Sum	0.19	0.36	0.32	0.13	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.91	0.94	0.77
Probability of Chance	0.69	0.50	0.78	0.28
Kappa (k)	0.73	0.81	0.75	0.67
Classification Accuracy	0.94	0.93	0.96	0.83

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.20	0.05	0.00	0.00	0.25
Basic	0.05	0.19	0.04	0.00	0.28
Proficient	0.00	0.05	0.26	0.03	0.34
Advanced	0.00	0.00	0.03	0.09	0.13
Sum	0.24	0.29	0.34	0.13	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.90	0.91	0.94	0.75
Probability of Chance	0.63	0.50	0.78	0.28
Kappa (k)	0.73	0.81	0.72	0.65
Classification Accuracy	0.93	0.93	0.96	0.82

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	0.22	0.04	0.00	0.00	0.27
Basic	0.05	0.20	0.05	0.00	0.30
Proficient	0.00	0.05	0.29	0.02	0.36
Advanced	0.00	0.00	0.02	0.05	0.07
Sum	0.27	0.30	0.36	0.08	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.90	0.96	0.77
Probability of Chance	0.61	0.51	0.86	0.30
Kappa (k)	0.77	0.80	0.70	0.67
Classification Accuracy	0.93	0.93	0.97	0.84

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.28	0.05	0.00	0.00	0.32
Basic	0.05	0.19	0.05	0.00	0.29
Proficient	0.00	0.05	0.27	0.01	0.33
Advanced	0.00	0.00	0.02	0.04	0.06
Sum	0.33	0.29	0.33	0.05	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.90	0.90	0.97	0.77
Probability of Chance	0.56	0.52	0.90	0.30
Kappa (k)	0.78	0.78	0.69	0.67
Classification Accuracy	0.93	0.93	0.98	0.84

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	0.24	0.05	0.00	0.00	0.29
Basic	0.06	0.25	0.04	0.00	0.35
Proficient	0.00	0.04	0.21	0.02	0.28
Advanced	0.00	0.00	0.02	0.06	0.09
Sum	0.29	0.34	0.28	0.08	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.89	0.91	0.96	0.76
Probability of Chance	0.59	0.54	0.84	0.29
Kappa (k)	0.74	0.81	0.73	0.67
Classification Accuracy	0.92	0.94	0.97	0.84

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.12	0.04	0.00	0.00	0.16
Basic	0.04	0.21	0.06	0.00	0.31
Proficient	0.00	0.07	0.22	0.05	0.33
Advanced	0.00	0.00	0.05	0.16	0.20
Sum	0.16	0.32	0.32	0.20	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.88	0.91	0.70
Probability of Chance	0.74	0.50	0.68	0.27
Kappa (k)	0.69	0.75	0.71	0.59
Classification Accuracy	0.94	0.91	0.93	0.78

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	0.14	0.04	0.00	0.00	0.19
Basic	0.04	0.17	0.05	0.00	0.27
Proficient	0.00	0.05	0.20	0.05	0.31
Advanced	0.00	0.00	0.04	0.19	0.23
Sum	0.19	0.27	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	0.89	0.90	0.71
Probability of Chance	0.70	0.50	0.64	0.26
Kappa (k)	0.72	0.78	0.73	0.61
Classification Accuracy	0.94	0.92	0.93	0.79

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.20	0.04	0.00	0.00	0.24
Basic	0.05	0.12	0.06	0.00	0.23
Proficient	0.00	0.06	0.17	0.06	0.28
Advanced	0.00	0.00	0.06	0.18	0.24
Sum	0.25	0.22	0.29	0.24	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.87	0.88	0.68
Probability of Chance	0.63	0.50	0.63	0.25
Kappa (k)	0.77	0.75	0.68	0.57
Classification Accuracy	0.94	0.91	0.92	0.77

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	0.23
Basic	0.04	0.16	0.06	0.00	0.26
Proficient	0.00	0.06	0.19	0.05	0.30
Advanced	0.00	0.00	0.05	0.16	0.21
Sum	0.23	0.26	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	0.88	0.90	0.70
Probability of Chance	0.65	0.50	0.67	0.25
Kappa (k)	0.76	0.76	0.69	0.59
Classification Accuracy	0.94	0.92	0.92	0.78

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	0.26	0.04	0.00	0.00	0.31
Basic	0.04	0.14	0.05	0.00	0.23
Proficient	0.00	0.05	0.16	0.04	0.25
Advanced	0.00	0.00	0.04	0.16	0.21
Sum	0.31	0.24	0.25	0.21	

Indexes for Classification Consistency and Classification Accuracy

Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.90	0.91	0.73
Probability of Chance	0.57	0.50	0.67	0.26
Kappa (k)	0.80	0.79	0.73	0.63
Classification Accuracy	0.94	0.93	0.94	0.80

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Item No.	Max	Percentage of Agreement			Intra. Corr.	Weighted Kappa	Mean	Score Frequency					
			Exact	Adjacent	Discrepant				No. of Second Reads	1	2	3	4	Codes
3	1	4	72.58	27.40	0.02	0.86	0.71	1.71	9162	2875	2761	650	71	2805
4	1	4	74.18	25.30	0.52	0.85	0.71	1.64	8057	3194	1558	376	54	2875
5	1	4	78.00	21.31	0.69	0.84	0.69	1.53	6200	3336	1765	271	11	817
6	1	4	81.61	18.39	0.00	0.89	0.78	1.51	6512	3349	2449	337	41	336
7	1	4	78.96	21.04	0.00	0.89	0.78	1.61	6146	3512	2039	370	49	176
8	1	4	75.94	24.03	0.03	0.90	0.79	1.67	6019	2474	2409	754	106	276

Note: The sum of the modes of agreement and codes may not equal exactly 100% due to rounding.
Note: TDA item scores presented in this table reflect a 1–4-point scoring rubric (before application of a weight of 2).

Copyright © 2019 by the Wisconsin Department of Public Instruction

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* (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. (p. 22)

As stated by the *Standards*, the validity of a testing program hinges on the use of the test scores. Validity evidence that supports the interpretation and 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, including design, content specifications, item development, psychometric quality, and inferences made from the results, provides evidence that either supports or challenges the validity of an intended interpretation of test scores.

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 four final issues related to the evidence of the validity of an intended interpretation of test scores: test fairness, evidence of validity based on the internal structure of the test, evidence of validity based on the relationship between test scores and other variables, and test integrity. 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 test blueprint and 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 forms a legitimate basis upon which to derive valid conclusions about student achievement.

Part 3 of this report presents the test design and describes the key development tasks related to creating the Spring 2019 Wisconsin Forward Exam operational test forms. The test blueprint and item development activities described in Part 2 explain how specific development

Copyright © 2019 by the Wisconsin Department of Public Instruction

processes provide evidence in support of the validity of an intended interpretation of test scores, primarily based on the test content and through the use of expert professional judgment from Wisconsin educators and from DRC test development specialists. The foundational documents—test blueprints and test designs—developed and approved during the initial phases of test development served as critical guides throughout the development and field-testing of items. These documents contributed to ensuring that each form of the test accurately measured the content in consistent and stable ways, thus providing evidence supporting the use of test scores as an indicator of student achievement of state standards.

Parts 2 and 3 together provide evidence to support the validity of an intended interpretation of test scores based on test content of the Wisconsin Forward Exam and address AERA, APA, & NCME (2014) Standards 3.1, 3.2, 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 an intended interpretation of test scores 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, MS, EBSR, SA, and TE auto-scored items were scored. It also describes how TDA writing items were scored, including 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 evidence of the validity of an intended interpretation of test scores of the Wisconsin Forward Exam by preventing hardware- or software-related errors in machine scoring and reducing construct-irrelevant score variance associated with variations in raters' interpretations and variations in the application of scoring rubrics.

Part 6 describes the sample data used for the item calibration, test equating, and test scaling. The calibration, equating, and scaling methods, and the processes and procedures for deriving scale scores from response patterns are also described in this part of the Technical Report. Some references to introductory and advanced discussions of IRT are provided. Several axes upon which to evaluate the calibration, equating, and scaling procedures, such as the models and data used, the software applied, the vertical relationship across grades, the 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, 5.13, 5.15 and 7.2. These processes and procedures contribute to the validity of an intended interpretation of test scores 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 stability of the Wisconsin Forward Exam in its fourth administration year.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Part 8 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 the interests of parents, students, and educators. The SPI scores then probed further, assessing specific skills and abilities. Together, the scale scores, performance levels, and SPI scores provided a comprehensive set of tools to assess Wisconsin student performance by content area and grade level and by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. In addition, longitudinal evaluation of student performance on the tests is included in this part of the Technical Report. 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 an intended interpretation of test scores of the Wisconsin Forward Exam by providing further evidence of the tests being accurate and reliable measurements of the intended constructs.

Part 9 demonstrates adherence to AERA, APA, & NCME (2014) standards through several analyses of the reliability of the Spring 2019 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 for TDA items. The *Spring 2019 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 evidence of the validity of an intended interpretation of test scores 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 an intended interpretation of test scores of the Wisconsin Forward Exam program. As described below, the Wisconsin Forward Exam program formally assessed the issue of test fairness through an analysis of differential item functioning (DIF). It is possible for items to function differently across different population groups, and it is also possible that results for an item do not reflect student ability but instead reflect irrelevant information influenced by demographic factors. The DIF analysis provided below serves to determine whether that possibility occurred and, if so, to what degree, item by item, for each of the categories of gender, race/ethnicity, socioeconomic status, disability status, accommodation use, 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 and the way in which the standard is addressed will be presented in this part.

Copyright © 2019 by the Wisconsin Department of Public Instruction

> **Standard 3.6** Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (p. 65)

There is no particular research on the Wisconsin Forward Exam showing that the test scores of examinee subgroups differ in meaning; however, this is an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, DRC 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. (p. 64)

DRC conducted DIF studies following the operational administration of the Wisconsin Forward Exam. Items are often evaluated for possible DIF in the field test phase of test development, and items flagged for DIF are typically further examined for possible bias. In the case of the Wisconsin Forward Exam, the DIF analyses are first conducted after the items are field-tested. Items flagged for DIF during the Spring 2018 field test analysis were reviewed again by DRC content experts for potential bias and were avoided during the selection of the Spring 2019 operational test forms. Only items deemed to be free of bias were included in the selection of the Spring 2019 forms. An additional DIF analysis was performed on the Spring 2019 operational test items. Items flagged for DIF were again evaluated by DRC content experts for potential bias. 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 3.2.3 of Part 3 discusses the form quality review conducted for the 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. (p. 63)

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

The present part of the report also provides the evidence of the validity of an intended interpretation of test scores related to test construct. Two measures are provided: correlations

Copyright © 2019 by the Wisconsin Department of Public Instruction

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. Therefore, these analyses are essential elements in assessing the validity of the Wisconsin Forward Exam. Next, the relationship between the Wisconsin Forward Exam scores and other variables is explored in order to support the evidence of the validity of an intended interpretation of test scores. These measures include evaluation of the correlations of the content area scores with other content area scores for the total population and by subgroups. They also include comparison of student performance on the Wisconsin Forward Exam with performance on the National Assessment of Educational Progress (NAEP). In addition, this chapter outlines the forensic analysis procedures that were employed to ensure the integrity of test scores by identifying schools and individual students that might have engaged in inappropriate behaviors during testing. Last but not least, a summary of standardized test administration procedures is provided as additional evidence supporting the validity of an intended interpretation 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, English language proficiency, and accommodation use groups. The reference and focal groups are as follows:

- **Gender**—reference group: male students; focal group: female students
- **Race/Ethnicity**—reference group: White students; focal groups: African-American, Asian, Hispanic, American Indian students
- **Socioeconomic status**—reference group: not economically disadvantaged students; focal group: economically disadvantaged students
- **Disability status**—reference group: students without disabilities; focal group: students with disabilities
- **English language proficiency**—reference group: fully English proficient students; focal group: limited English proficient students
- **Accommodation use**—reference group: students not using testing accommodations; focal group: students using testing accommodations

Copyright © 2019 by the Wisconsin Department of Public Instruction

Two DIF statistics that are commonly used for this purpose are the Mantel-Haenszel (MH) statistic (1959) and the Standardized Mean Difference (SMD) between the reference and focal groups, proposed by Dorans and Schmitt (1991).

The MH statistic is computed as follows (Zwick, Donoghue, & Grima, 1993):

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

where F_k is the sum of scores for the focal group at the k level of the matching variable. Note that the MH statistic is sensitive to N such that larger sample sizes increase the value of the chi-square.

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

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

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. Then the χ^2_{MH} estimates a pooled common-odds ratio. Taking the natural logarithm of the common-odds ratio and its confidence limits and multiplying these with the constant −2.35, the resulting values may then be placed on the MH delta metric (Δ_{MH}) for interpretive purposes. Items were flagged for DIF using the following criteria:

- Moderate DIF: Significant MH chi-square statistic (p <0.05) and 1.0 ≤ |MH D-DIF| <1.5

Copyright © 2019 by the Wisconsin Department of Public Instruction

- Large DIF: Significant MH chi-square statistic (p <0.05) and |MH D-DIF| ≥ 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):

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

where p_{Fk} = the proportion of the focal group members at the kth level of the matching variable, $m_{Fk} = 1/N_{F1k}$, and $m_{Rk} = 1/N_{R1k}$. Items are flagged using the same rules that are used in the NAEP:

- Moderate DIF: If the MH statistic is significant (p <0.05) and |ES| is between 0.17 and 0.25
- Large DIF: If the MH statistic is significant (p <0.05) and |ES| ≥ 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 all subgroups of students.

A negative SMD value implies that the focal group has a lower mean item score than the reference group, whereas a positive value implies that the focal group has a higher mean item score than the reference group, conditioned on the matching test score.

The minimum case count for the focal group was set at 200, and the minimum case count for the reference group was set at 400. The DIF analyses were not performed for subgroups of fewer than 200 students. In these cases, the statistical procedures do not have sufficient power to detect differences should they exist.

Tables 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 Tables 10-1 through 10-9.

The DIF results for gender are presented in Table 10-1, results for race/ethnicity are presented in Tables 10-2 through 10-5, results for English language proficiency are presented in Table 10-6, results for socioeconomic status DIF are shown in Table 10-7, the DIF results for disability status are presented in Table 10-8, and the DIF results for accommodation use are presented in Table 10-9.

Each DIF table references the grade and content area of the items flagged for DIF, the item number on the test and the item type. The tables present the MH SMD statistics and the Mantel-Haenszel statistic (Δ_{MH}). After specifying these statistics for each item, the final

Copyright © 2019 by the Wisconsin Department of Public Instruction

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, seven items were flagged for moderate (B flag) gender DIF and four items were flagged for large (C flag) DIF in ELA across all grade levels. Of these items, seven were flagged in favor of the focal group (females) and four were flagged against the focal group. Eight items were flagged for moderate DIF (two items in favor of and six items against female students) and two items were flagged for large DIF (both against female students) across all grades in Mathematics. One item was flagged for moderate DIF and one item was flagged for large DIF in Science (both against female students). In addition, nine items were flagged for moderate DIF and two items were flagged for large DIF in Social Studies (a total of four items in favor of and seven items against female students). Overall, thirteen items were flagged in favor of the focal group (females), and twenty-one items were flagged against the focal group across all grades and content areas. Of all items flagged for gender DIF, nine displayed large DIF (either in favor of or against female students) and twenty-five items displayed 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.

When looking at DIF results by item type, it was observed that most of the flagged items were MC items across all content areas and subgroups. The exceptions were DIF results for ELA conducted for subgroups of students with and without disabilities. As can be seen in Table 10-8, the item type flagged most often was a TDA item. TDA items were flagged against students with disabilities in all grades.

The Spring 2019 Wisconsin Forward Exam was developed to minimize item and test bias. As stated earlier in this part of the Technical Report, all operational and field test items flagged for DIF in Spring 2018 were reviewed by DRC content experts for potential content-related bias. Only items deemed to be free of bias were included in the selection of the Spring 2019 forms. Items flagged for DIF after the Spring 2019 test administration were again evaluated by DRC content experts for potential bias.

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 2019 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 large DIF) are to be avoided in future selections.

Copyright © 2019 by the Wisconsin Department of Public Instruction

10.2 Validity Evidence Based on Internal Test Structure

Construct-related evidence of the validity of an intended interpretation of test scores can be defined as the extent to which tests measure the skills or constructs they intend to measure and is the central concept underlying the Spring 2019 Wisconsin Forward Exam validation process. Evidence for construct-related validity is comprehensive and integrates evidence from both content- and criterion-related validity. The Wisconsin Forward Exam development process included specifications, item writing, review, and test construction.

Threats to construct-related validity include the unintended measurement of variables unrelated to the desired constructs and multidimensionality of the tests. To ensure that the test items are focused on the desired constructs, standardized procedures are employed to select items with sound statistical properties, to align the items to content standards, and to ensure that each test form meets the Wisconsin Forward Exam blueprint. A test can be said to be unidimensional when all of the items in the test measure the same underlying ability or trait. For example, Mathematics items should measure Mathematics ability and not Reading skills. Standard 1.13 of the AERA, APA, & NCME (2014) *Standards* states the following:

> If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided. (pp. 26 and 27)

10.2.1 Correlations between Content Standards

Analyses of the internal structure of a test can indicate the extent to which the relationships between test items and components conform to the construct the test purports to measure. For educational assessments that are designed to measure a single construct or content domain, the correlations between content standards within a test can be expected to be relatively high. Table 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 between two content standards, and a correlation of -1 indicates a perfect negative linear relationship between two content standards. A correlation of zero means there is no linear relationship. In general, the size of the correlation coefficient is influenced by the number of items or score points and by the score variance. Readers are cautioned not to confuse correlation with causation. The presence of a high correlation between two content standards should not be taken as an indication that there is a causal relationship between them.

As may be observed in Table 10-10, the correlations between the ELA main test domains of Reading, Writing, and Listening are moderate to high and range from 0.55 to 0.76 across all grades. The lowest correlations were observed between Listening and Writing domains, while the highest correlations were observed between Reading and Writing domains. The correlations between ELA content standards (see Table 10-11) are typically moderate for all grades and all

Copyright © 2019 by the Wisconsin Department of Public Instruction

standard pairs and range from 0.37 to 0.69. It should be noted that the number of items in content standards was smaller than the number of items in ELA domains, resulting in lower correlations at the standard level compared to the correlations at the ELA domain level.

As indicated in Table 10-12, the correlations between Mathematics content standards are moderate to high and range from 0.53 to 0.75. The correlations between Science content standards range from 0.61 to 0.70 (see Table 10-13), and the correlations between Social Studies content standards range from 0.52 to 0.76 (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.

10.2.2 Principal Component Analysis

Wisconsin Forward Exam items are calibrated using unidimensional IRT models, which suggests that the test items are measuring an essentially unidimensional construct. To assess the dimensionality of the Wisconsin Forward Exam, a principal components analysis was conducted for each content area and grade. A principal components analysis is a statistical technique commonly used to evaluate dimensionality by detecting patterns of relationships among items. This method is useful in determining whether the observed scores on a test can be explained largely or entirely in terms of a much smaller number of components. For example, if answering the Mathematics items in a Mathematics test required a high level of reading ability, the Mathematics test would be measuring not only mathematics ability but also reading ability. Such a test would be said to be multidimensional rather than essentially unidimensional. One way of evaluating the dimensions detected in the analysis is by examining the eigenvectors and eigenvalues. In a principal component analysis, the eigenvectors correspond to factors, and the eigenvalues correspond to the variance explained by these factors. The sum of the eigenvalues is equal to the number of items in the test. The eigenvalues can be ordered from first to last in terms of the amount of common variance that each explains. Data are generally considered to be unidimensional if the second eigenvalue is less than or equal to 1.0. Previous research shows that 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 is more than five times as large as the eigenvalue of the second factor.

As can be seen in Table 10-15, the ratios of the first two eigenvalues range from 5.86 to 7.83. The eigenvalues are proportional to the amount of common variance explained by each component, indicating that the variance explained by the first component alone is approximately six to eight times greater than the variance explained by the second component. The eigenvalue

Copyright © 2019 by the Wisconsin Department of Public Instruction

ratios range from 6.54 to 7.62 in ELA, from 5.86 to 7.36 in Mathematics, from 6.25 to 7.52 in Science, and from 6.56 to 7.83 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.

10.3 Validity Evidence Based on Relationship with Other Variables

The relationship between the Wisconsin Forward Exam scores and other variables was examined to further support the validity of the intended score interpretation. This was done using two measures: evaluation of correlations between the Wisconsin Forward Exam content area scores and comparisons of the percentages of students classified in different proficiency levels (impact data) on the State assessment and on the NAEP assessment.

10.3.1 Correlations between Content Area Test Scores

The test score relationship with other variables can be assessed by the extent to which measures of constructs that theoretically should not be related to each other are, in fact, observed as not related to each other. Typically, correlation coefficients among measures of unrelated or distantly related constructs are examined in support of the validity evidence based on the relationship of the test scores with other variables.

To assess the relationship between the Wisconsin Forward Exam content area scores, the correlations between the ELA, Mathematics, Science, and Social Studies scale scores for students who took more than one subject area test in 2019 were computed and examined for the total student population and by subgroup. Table 10-16 shows the correlations between the content area scores for the total population of Wisconsin students. These correlations ranged from 0.71 (between Mathematics and Social Studies in grade 8) to 0.82 (between ELA and Science in grade 4).

Tables 10-17 through 10-21 show correlation coefficients between the content area scores by gender, ethnicity, English language proficiency status, socioeconomic status, and disability status, respectively. As seen in Table 10-17, the correlations between the content area scores for male or female groups ranged from 0.71 to 0.83 and were comparable for the two gender groups for each pair of correlated scores. The correlations between the content area scores for different ethnic groups ranged from 0.59 to 0.83 (see Table 10-18). The highest correlations by ethnic group were observed for Asian students. Correlations between the content area scores for the African-American student subgroup were lower than the correlations for other subgroups. As shown in Table 10-19, the correlations between the content area scores by English proficiency status ranged from 0.50 to 0.74 for English language learners and from 0.71 to 0.82 for fully English proficient students across all grade levels and all pairs of correlated scores. The

Copyright © 2019 by the Wisconsin Department of Public Instruction

correlations between the content area scores by student socioeconomic status are presented in Table 10-20. These correlations ranged from 0.66 to 0.80 across all grades and pairs of correlated scores. In all grade levels, the correlations between each pair of scores were, in most cases, comparable for the groups of students considered economically disadvantaged and not economically disadvantaged. The correlations between the content area scores by student disability status are shown in Table 10-21. These correlations ranged from 0.58 to 0.81 across all grades and pairs of correlated scores. The correlations between each pair of scores were fairly comparable for the groups of students with and without disabilities in grades 3 and 4. In higher grades, between each pair of scores, correlations were lower for the group of students with disabilities compared to the group of students without disabilities. The correlation coefficients between the content area scores were not computed by accommodation use, because the accommodation use status is not consistent across content areas for the same students (for example, students who used accommodations in one content area did not necessarily use accommodations in another content area).

Overall, the correlations between the content area scores for the total population of students were found to be highly related. The correlations between the content area scores for the subgroups of students were found to be moderately to highly related. Despite high correlations, the tests are not perfectly related to one another, suggesting that different constructs are being tapped; however, if the test scores are highly related to one another, they may be tapping into a similar knowledge base or general underlying ability.

Partial Correlations

In addition to the simple correlations between the content area scores, partial correlations, which are measures of the strength of the relationship between the content area scores while controlling for the student demographic characteristics (gender, ethnicity, English proficiency status, disability status, and socioeconomic status), were also computed. Partial correlations allow for the evaluation of the relationship of two content area scores with the effect of the student demographic characteristics removed (or held constant). The partial correlations between the ELA, Mathematics, Science, and Social Studies test scores for the total population of students and at each grade level are presented in Table 10-22. These correlations ranged from 0.61 (between Mathematics and Social Studies in grade 8) to 0.78 (between ELA and Science in grade 4). Although the magnitude of these correlations is considered to be strong, as expected, the partial correlations between the content area scores were lower than the corresponding simple correlations, indicating that the student demographic characteristics did contribute to the strength of the relationship between the content area test scores. The differences between the simple correlation and corresponding partial correlation coefficients were, however, relatively small, indicating that the effect of the student demographic characteristics on the relationship between the ELA, Mathematics, Science, and Social Studies test scores was small.

10.3.2 Comparison of the Wisconsin Forward Exam and Wisconsin NAEP Impact Data

The NAEP is the largest nationally representative and continuing assessment of what America’s students know and can do in various content areas. Assessments in several content areas, including Reading, Mathematics, and Science, are administered to students in grades 4, 8,

Copyright © 2019 by the Wisconsin Department of Public Instruction

and 12 and conducted periodically. Representative samples of students from different states, including Wisconsin, participated in the latest NAEP assessment, which occurred in Spring 2019.

The main NAEP assessments are constructed using detailed frameworks that result from a comprehensive national process in which teachers, curriculum experts, policymakers, and members of the general public work to create a unified vision of how a particular subject ought to be assessed. This vision is based on current educational research on achievement and its measurement as well as good educational practices. These frameworks are updated about every decade in order to keep them current (for details, refer to https://nces.ed.gov).

The NAEP results are reported for all assessed content areas and for all participating grades at the national level. At the state level, the results for Reading, Mathematics, Science, and Writing are reported for grades 4 and 8. The results may also be reported at the district level (within a state) for these four content areas. No results are reported at the student level.

Wisconsin students participated in the last two NAEP assessments in Spring 2019 (Reading and Mathematics) and Spring 2015 (Science). The Wisconsin Forward Exam state assessment results are compared to the latest available NAEP results in grades 4 and 8. The percentages of Wisconsin students classified in different proficiency levels on the Wisconsin Forward Exam and the corresponding NAEP assessments are presented in Table 10-23. With two exceptions, the percentages of students classified in different performance levels on the NAEP assessments and on the Wisconsin Forward Exam were comparable within 10% or less for any performance level in both grades and in all three content areas. The exceptions were the percentages of students classified in the *Advanced* level for Science, where the differences were over 18% in grade 4 and over 20% in grade 8, with a larger percentage of students classified as *Advanced* on the Wisconsin Forward Exam compared to the NAEP Science assessment.

Looking at the percentages of students classified as *Proficient* or above, higher proportions of students were classified in these two combined categories on the Wisconsin Forward Exam in ELA grade 4 and Science (both grades) compared to the corresponding NAEP Reading and Science assessments. Comparable proportions of students were classified as *Proficient* or above on the Wisconsin Forward Exam in Mathematics grade 4 and the corresponding NAEP Mathematics assessment. Higher proportions of students were classified in the *Proficient* or above category on the NAEP Reading and Mathematics assessments in grade 8 compared to the Wisconsin Forward Exam in ELA and Mathematics for that grade level. The differences between NAEP and Wisconsin Forward Exam were 7% or less for ELA grades 4 and 8 and Mathematics grade 8. The differences between NAEP and Wisconsin Forward Exam for Science were approximately 12% at grade 4 and approximately 14% at grade 8. A similar pattern of impact data was observed for students classified as *Basic* or above with more students classified in the *Basic* or above categories on the Wisconsin Forward Exam in ELA grade 4, Mathematics grade 4, and both Science grades compared to the corresponding NAEP Reading, Mathematics, and Science assessments. Fewer students were classified in the *Basic* or above categories on the Wisconsin Forward Exam in ELA grade 8 and Mathematics grade 8 compared to the NAEP. All differences in the percentages of students classified as *Basic* or above between the Wisconsin Forward Exam and NAEP were approximately 10% or less.

Copyright © 2019 by the Wisconsin Department of Public Instruction

It should be noted that the Spring 2015 Reading and Mathematics Wisconsin NAEP impact data were used as benchmarks during the Wisconsin Forward Exam standard setting for ELA and Mathematics after the Spring 2016 test administration. The Spring 2015 Science Wisconsin NAEP impact data were also shown to the participants for reference and guidance in performance level setting during the Spring 2019 standard setting. While the standard setting participants were free to deviate from the NAEP impact data while placing their bookmarks in the ordered item booklets in consideration of the Wisconsin performance level descriptors, the final Wisconsin impact data achieved after the standard setting were generally aligned with the Wisconsin state-level NAEP data. When considering the Wisconsin content standards and impact data articulation across grades, the Wisconsin Forward Exam cut scores for ELA, Mathematics, and Science remained in most cases aligned with the *Proficient* benchmarks, further supporting the evidence of the relationship between the state and the national assessments in these content areas.

10.4 Test Integrity: Data Forensic Analyses

With the high-stakes nature of large-scale statewide assessment programs, there can be situations in which student responses, and hence their scores, may not be a true representation of student ability. Various activities may take place, such as a student copying from another student's paper, a student receiving inappropriate assistance before or during testing, or a student's responses being altered during or after testing. To maintain the integrity of the Wisconsin Forward Exam and the validity of the results, it is important that any such instances be discovered.

Two studies were conducted to evaluate the Wisconsin Forward Exam student data for any indicators of possible inappropriate testing behavior. The first study examines incorrect student responses to MC items on the Spring 2019 Wisconsin Forward Exam in ELA, Mathematics, Science, and Social Studies that were changed to correct responses. These answer changes are referred to as wrong-to-right answer changes. Inordinate numbers of wrong-to-right answer changes in a specifically identifiable testing administration group may indicate inappropriate student behavior or intervention by an educator during the testing session.

The second study evaluates the time spent on the test and individual test items by students. These analyses serve to inform of any events in which students (within one school) spent a very short or very long time on the test or specific items. Inordinate numbers of unusual test or item response times may indicate inappropriate pre-knowledge of the items or other interventions during the testing session.

The results of the two studies are provided to DPI for evaluation. We emphasize that the results from these studies may be used in conjunction with other information to investigate whether inappropriate interventions may have taken place. The statistical results by themselves may simply be coincidental and do not necessarily indicate inappropriate behavior.

Copyright © 2019 by the Wisconsin Department of Public Instruction

10.5 Standardized Test Administration

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

Standardized test administration, extensive training of the test scorers and artificial intelligence (AI) engine, and rigorous scoring rules for 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. (p. 65)
>
> **Standard 3.5** Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population. (p. 65)

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.

10.6 Summary

In summary, the overall purpose of Part 10 was to provide additional evidence of the validity of an intended interpretation of test scores related to test construct. Through the measures of correlations between content area objectives and principal components analysis, the existence of a single, underlying trait or ability for each content area was demonstrated. Next, the relationship between the Wisconsin Forward Exam scores and other variables was explored and validated through the evaluation of correlations of the content area scores with other content area scores for the total population and by subgroups. In addition, the student performance on the Wisconsin Forward Exam with the performance on the NAEP was also compared. The forensic analysis procedures that were employed to ensure the integrity of test scores by identifying schools and individual students that might have engaged in inappropriate behaviors during testing were also described in this part of the report. Finally, a summary of standardized test administration procedures was provided as additional evidence supporting the validity of an intended interpretation of test scores.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	4	1	TDA	0.16		B
	5	1	TDA	0.18		B
	5	24	MC	-0.11	-1.16	B-
	6	1	TDA	0.23		C
	7	1	TDA	0.24		C
	7	9	MC	-0.11	-1.36	B-
	8	1	TDA	0.27		C
	8	9	MC	0.08	1.40	B
	8	18	MC	-0.09	-1.02	B-
	8	21	MC	-0.17	-1.70	C-
	8	39	MC	0.09	1.01	B
Math	3	4	MC	-0.07	-1.21	B-
	3	19	MC	-0.06	-1.02	B-
	3	23	MC	-0.08	-1.56	C-
	4	7	MC	-0.12	-1.79	C-
	4	37	MC	0.05	1.06	B
	6	12	MC	-0.03	-1.02	B-
	6	33	MC	-0.08	-1.02	B-
	6	46	MC	0.05	1.22	B
	7	2	MC	-0.08	-1.13	B-
	7	31	MC	-0.06	-1.04	B-
Science	8	20	TE	-0.15	-1.64	C-
	8	34	TE	-0.09	-1.00	B-
Social Studies	4	23	MC	0.04	1.13	B
	4	31	TE	-0.12	-1.43	B-
	4	33	MC	0.08	1.44	B
	8	1	MC	-0.08	-1.88	C-
	8	14	MC	-0.11	-1.19	B-
	8	21	MC	0.03	1.22	B
	8	22	MC	-0.09	-1.24	B-
	8	25	MC	-0.09	-1.38	B-
	10	9	MC	-0.11	-1.38	B-
	10	24	TE	-0.13	-1.98	C-
	10	33	MC	0.09	1.20	B

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	3	11	MC	-0.09	-1.21	B-
	5	25	TE	-0.09	-1.18	B-
Math	4	17	MC	-0.07	-1.04	B-
	6	12	MC	-0.08	-1.33	B-
	6	20	SA	-0.10	-1.34	B-
	6	23	TE	-0.08	-0.75	B-
	7	17	MC	0.06	1.06	B
	7	31	MC	-0.07	-1.02	B-
	8	10	MC	0.08	1.06	B
	8	15	TE	-0.11	-2.08	B-
Science	4	33	TE	-0.07	-0.80	B-
	8	21	TE	-0.09	-1.10	C-
Social Studies	4	2	MC	0.08	1.33	B
	4	7	TE	-0.11	-1.31	C-
	4	11	MC	0.10	1.16	B
	8	1	MC	-0.24	-3.31	C-
	8	6	MC	0.18	2.41	C
	8	11	MC	-0.13	-1.71	C-
	8	20	MC	0.19	2.20	C
	8	22	MC	-0.08	-1.01	B-
	8	25	MC	-0.18	-2.11	C-
	10	49	MC	-0.09	-1.01	B-

Copyright © 2019 by the Wisconsin Department of Public Instruction

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
Social Studies	8	1	MC	-0.12	-2.06	C-

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	TDA	0.18		B
	3	11	MC	-0.12	-1.58	C-
	4	1	TDA	0.18		B
	4	27	MC	-0.11	-1.25	B-
	5	1	TDA	0.13		B
	5	12	MC	0.09	1.02	B
	5	15	MC	-0.10	-1.22	B-
	5	25	TE	-0.07	-1.18	B-
	8	1	TDA	0.15		B
Math	4	2	MC	-0.10	-1.11	B-
	4	17	MC	-0.05	-1.12	B-
	5	22	TE	0.11	1.53	B
	5	23	MC	-0.08	-1.08	B-
	6	7	MC	0.08	1.13	B
	6	19	MC	-0.08	-1.07	B-
	6	20	SA	-0.09	-1.40	B-
	6	35	MC	-0.09	-1.25	B-
	6	45	MC	-0.07	-1.22	B-
	7	31	MC	-0.08	-1.35	B-
	8	9	TE	0.09	1.47	B
Social Studies	4	19	MC	0.06	1.04	B
	4	23	MC	0.04	1.04	B
	4	31	TE	0.09	1.11	B
	8	1	MC	-0.08	-1.97	C-
	8	25	MC	-0.07	-1.09	B-
	8	40	MC	0.09	1.35	B
	10	20	MC	-0.08	-1.05	B-
	10	22	MC	0.09	1.12	B
	10	25	MC	0.07	1.22	B
	10	36	TE	-0.09	-0.93	B-

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	1	TDA	-0.15		B-
	7	1	TDA	-0.15		B-
Social Studies	8	1	MC	-0.06	-1.13	B-

Table 10-6 Items Flagged for DIF by English Language Proficiency, Focal Group: Students Not English Language Proficient

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
ELA	3	11	MC	-0.09	-1.13	B-
	4	27	MC	-0.12	-1.27	B-
	5	11	TE	-0.12		B-
	8	16	EBSR	-0.12		B-
Math	6	7	MC	0.07	1.14	B
	6	20	SA	-0.11	-1.59	B-
Social Studies	8	1	MC	-0.11	-1.46	B-

Table 10-7 Items Flagged for DIF by Socioeconomic Status, Focal Group: Socioeconomically Disadvantaged Students

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
Social Studies	8	1	MC	-0.07	-1.12	B-

Copyright © 2019 by the Wisconsin Department of Public Instruction

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	1	TDA	-0.22		B-
	4	1	TDA	-0.29		C-
	5	1	TDA	-0.27		C-
	5	11	TE	-0.15		B-
	5	23	TE	-0.07	-1.11	B-
	6	1	TDA	-0.28		C-
	7	1	TDA	-0.27		C-
	8	1	TDA	-0.27		C-
	8	9	MC	-0.10	-1.15	B-
Math	4	10	MC	-0.08	-1.33	B-
	6	12	MC	-0.12	-1.93	C-
Science	4	34	TE	0.06	1.14	B
Social Studies	8	21	MC	-0.06	-1.10	B-

Table 10-9 Items Flagged for DIF by Accommodation Use, Focal Group: Students Using Testing Accommodations

Content	Grade	Item Number	Item Type	MH SMD Statistic	MH Delta Statistic	DIF Flag
Math	3	4	MC	-0.08	-1.01	B-
	3	16	MC	0.09	1.39	B
	3	18	TE	-0.06	-0.81	B-
	3	20	MC	-0.08	-1.04	B-
	3	28	MC	0.08	1.09	B
	3	40	MC	-0.09	-1.21	B-
	4	10	MC	-0.11	-1.41	B-
	4	31	MC	0.11	1.17	B
	5	1	SA	-0.09	-1.40	B-
	5	41	MC	0.06	1.55	C
	6	12	MC	-0.17	-2.11	C-
	6	45	MC	0.06	1.05	B
	7	46	SA	-0.10	-2.01	B-

Note: DIF analysis by accommodation use was not performed on ELA, Science, and Social Studies data due to insufficient number of students using testing accommodations in these content areas.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-10 Correlations among English Language Arts Test Domains

Grade	ELA Domain	Listening	Reading
3	Reading	0.64	
	Writing	0.63	0.76
4	Reading	0.58	
	Writing	0.55	0.74
5	Reading	0.67	
	Writing	0.63	0.75
6	Reading	0.64	
	Writing	0.60	0.73
7	Reading	0.63	
	Writing	0.60	0.74
8	Reading	0.66	
	Writing	0.63	0.76

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-11 Correlations among English Language Arts Standards

Grade	Standard Code	A	B	C	D	E	F
3	B	0.51					
	C	0.55	0.38				
	D	0.60	0.41	0.46			
	E	0.59	0.42	0.45	0.52		
	F	0.64	0.45	0.51	0.58	0.57	
	G	0.61	0.42	0.48	0.52	0.52	0.55
4	B	0.61					
	C	0.63	0.59				
	D	0.54	0.51	0.52			
	E	0.57	0.53	0.56	0.52		
	F	0.49	0.45	0.48	0.47	0.46	
	G	0.52	0.48	0.50	0.46	0.48	0.39
5	B	0.69					
	C	0.47	0.48				
	D	0.54	0.55	0.38			
	E	0.61	0.60	0.41	0.52		
	F	0.51	0.52	0.37	0.48	0.49	
	G	0.61	0.61	0.43	0.50	0.57	0.47
6	B	0.60					
	C	0.61	0.53				
	D	0.55	0.47	0.50			
	E	0.52	0.46	0.48	0.45		
	F	0.52	0.45	0.48	0.47	0.45	
	G	0.59	0.51	0.54	0.50	0.48	0.46
7	B	0.64					
	C	0.62	0.54				
	D	0.56	0.51	0.49			
	E	0.57	0.52	0.51	0.49		
	F	0.51	0.45	0.47	0.46	0.47	
	G	0.58	0.52	0.52	0.49	0.51	0.45
8	B	0.66					
	C	0.65	0.58				
	D	0.62	0.56	0.52			
	E	0.62	0.56	0.55	0.56		
	F	0.48	0.43	0.42	0.48	0.46	
	G	0.61	0.54	0.55	0.53	0.56	0.42

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-12 Correlations among Mathematics Standards

Grade	Standard Code	A	B	C	D	E	F	G	H	I
3	B	0.73								
	C	0.67	0.67							
	D	0.72	0.74	0.68						
	E	0.67	0.68	0.68	0.69					
4	B	0.67								
	C	0.65	0.74							
	D	0.63	0.72	0.73						
	E	0.53	0.59	0.62	0.63					
5	B	0.69								
	C	0.64	0.68							
	D	0.61	0.64	0.66						
	E	0.66	0.67	0.65	0.65					
6	F					0.61				
	G					0.64	0.73			
	H					0.64	0.70	0.75		
	I					0.57	0.61	0.65	0.64	
7	F					0.56				
	G					0.57	0.69			
	H					0.58	0.68	0.69		
	I					0.57	0.72	0.68	0.69	
8	G					0.63				
	H					0.65		0.69		
	I					0.59		0.55	0.60	
	J					0.67		0.65	0.71	0.65

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-13 Correlations among Science Standards

Grade	Standard Code	A	B	C
4	B	0.70		
	C	0.63	0.61	
	D	0.70	0.68	0.61
8	B	0.70		
	C	0.67	0.64	
	D	0.70	0.67	0.65

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

Table 10-14 Correlations among Social Studies Standards

Grade	Standard Code	A	B	C	D
4	B	0.64			
	C	0.52	0.61		
	D	0.57	0.64	0.53	
	E	0.60	0.71	0.59	0.62
8	B	0.71			
	C	0.59	0.63		
	D	0.65	0.66	0.56	
	E	0.61	0.63	0.53	0.56
10	B	0.72			
	C	0.69	0.76		
	D	0.66	0.71	0.69	
	E	0.64	0.71	0.69	0.64

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-15 Principal Components Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
ELA	3	8.037	1.180	6.813
	4	8.061	1.090	7.397
	5	8.718	1.180	7.387
	6	7.448	1.139	6.539
	7	7.873	1.142	6.896
	8	8.806	1.155	7.624
Mathematics	3	10.842	1.473	7.363
	4	10.674	1.753	6.088
	5	10.366	1.419	7.305
	6	10.432	1.781	5.857
	7	10.034	1.566	6.408
	8	10.240	1.633	6.272
Science	4	7.905	1.264	6.252
	8	8.365	1.112	7.525
Social Studies	4	8.088	1.233	6.559
	8	8.690	1.234	7.045
	10	10.582	1.352	7.828

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	0.77					
4	0.75	0.82	0.81	0.78	0.73	0.81
5	0.74					
6	0.77					
7	0.76					
8	0.74	0.80	0.81	0.75	0.71	0.81

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-17 Correlations between Content Area Scale Scores by Gender

Grade	Demographic Group	ELA & Math	ELA & Science	ELA & Social Studies	Math & Science	Math & Social Studies	Science & Social Studies
3	Female	0.78					
	Male	0.78					
4	Female	0.76	0.83	0.81	0.78	0.74	0.81
	Male	0.76	0.83	0.81	0.78	0.73	0.81
5	Female	0.74					
	Male	0.75					
6	Female	0.77					
	Male	0.78					
7	Female	0.77					
	Male	0.77					
8	Female	0.74	0.82	0.82	0.75	0.71	0.81
	Male	0.75	0.81	0.81	0.75	0.71	0.81

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-18 Correlations between Content Area Scale Scores by Ethnicity/Race

Grade	Demographic Group	ELA & Math	ELA & Science	ELA & Social Studies	Math & Science	Math & Social Studies	Science & Social Studies
3	White	0.74					
	African-American	0.69					
	Hispanic	0.73					
	Asian	0.78					
	American Indian	0.67					
	Two or More	0.75					
4	White	0.72	0.80	0.78	0.75	0.69	0.79
	African-American	0.67	0.75	0.72	0.68	0.65	0.73
	Hispanic	0.71	0.80	0.77	0.73	0.70	0.78
	Asian	0.77	0.83	0.82	0.79	0.75	0.81
	American Indian	0.71	0.79	0.75	0.71	0.66	0.77
	Two or More	0.74	0.83	0.80	0.77	0.72	0.81
5	White	0.72					
	African-American	0.63					
	Hispanic	0.69					
	Asian	0.76					
	American Indian	0.68					
	Two or More	0.74					
6	White	0.74					
	African-American	0.68					
	Hispanic	0.73					
	Asian	0.77					
	American Indian	0.67					
	Two or More	0.76					
7	White	0.74					
	African-American	0.65					
	Hispanic	0.71					
	Asian	0.80					
	American Indian	0.69					
	Two or More	0.75					
8	White	0.72	0.78	0.78	0.73	0.68	0.78
	African-American	0.64	0.74	0.75	0.62	0.59	0.73
	Hispanic	0.69	0.78	0.80	0.69	0.66	0.80
	Asian	0.77	0.81	0.82	0.78	0.75	0.82
	American Indian	0.66	0.77	0.79	0.72	0.67	0.79
	Two or More	0.73	0.80	0.80	0.74	0.68	0.80

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-19 Correlations between Content Area Scale Scores by English Proficiency Status

Grade	Demographic Group	ELA & Math	ELA & Science	ELA & Social Studies	Math & Science	Math & Social Studies	Science & Social Studies
3	Fully English Proficient	0.77					
	Limited English Proficient	0.71					
4	Fully English Proficient	0.75	0.82	0.80	0.78	0.73	0.81
	Limited English Proficient	0.65	0.74	0.72	0.69	0.66	0.74
5	Fully English Proficient	0.74					
	Limited English Proficient	0.60					
6	Fully English Proficient	0.77					
	Limited English Proficient	0.60					
7	Fully English Proficient	0.76					
	Limited English Proficient	0.58					
8	Fully English Proficient	0.74	0.80	0.80	0.75	0.71	0.81
	Limited English Proficient	0.53	0.64	0.68	0.54	0.50	0.69

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-20 Correlations between Content Area Scale Scores by SES Status

Grade	Demographic Group	ELA & Math	ELA & Science	ELA & Social Studies	Math & Science	Math & Social Studies	Science & Social Studies
3	Not Economically Disadvantaged	0.74					
	Economically Disadvantaged	0.74					
4	Not Economically Disadvantaged	0.72	0.80	0.78	0.75	0.69	0.79
	Economically Disadvantaged	0.72	0.80	0.78	0.74	0.70	0.79
5	Not Economically Disadvantaged	0.72					
	Economically Disadvantaged	0.69					
6	Not Economically Disadvantaged	0.74					
	Economically Disadvantaged	0.73					
7	Not Economically Disadvantaged	0.74					
	Economically Disadvantaged	0.72					
8	Not Economically Disadvantaged	0.72	0.78	0.78	0.73	0.68	0.78
	Economically Disadvantaged	0.69	0.78	0.79	0.70	0.66	0.80

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-21 Correlations between Content Area Scale Scores by Disability Status

Grade	Demographic Group	ELA & Math	ELA & Science	ELA & Social Studies	Math & Science	Math & Social Studies	Science & Social Studies
3	Not Disabled	0.76					
	Disabled	0.73					
4	Not Disabled	0.74	0.81	0.80	0.77	0.71	0.80
	Disabled	0.70	0.79	0.75	0.74	0.69	0.78
5	Not Disabled	0.72					
	Disabled	0.65					
6	Not Disabled	0.75					
	Disabled	0.66					
7	Not Disabled	0.75					
	Disabled	0.63					
8	Not Disabled	0.72	0.79	0.79	0.74	0.69	0.79
	Disabled	0.60	0.73	0.74	0.63	0.58	0.76

Table 10-22 Partial Correlations between Content Area Scale Scores

Grade	ELA & Mathematics	ELA & Science	ELA & Social Studies	Mathematics & Science	Mathematics & Social Studies	Science & Social Studies
3	0.70					
4	0.68	0.78	0.75	0.70	0.64	0.75
5	0.65					
6	0.68					
7	0.68					
8	0.65	0.75	0.75	0.67	0.61	0.75

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table 10-23 Comparison of Most Recent Wisconsin NAEP and Spring 2019 Wisconsin Forward Exam Impact Data

Content	Grade	Wisconsin NAEP Percentages of Students							Wisconsin Forward Exam Spring 2019 Percentages of Students					
		NAEP Year	Below Basic	Basic	Proficient	Advanced	At or Above Proficient	At or Above Basic	Below Basic	Basic	Proficient	Advanced	At or Above Proficient	At or Above Basic
Reading/ ELA	4	2019	34	30	26	9	36	66	23.88	33.14	34.10	8.89	42.98	76.12
Reading/ ELA	8	2019	24	38	34	4	39	76	25.94	37.04	28.80	8.23	37.03	74.06
Math	4	2019	20	35	34	11	45	80	18.87	36.09	32.82	12.23	45.05	81.13
Math	8	2019	24	34	29	12	41	76	28.55	35.60	27.83	8.01	35.85	71.45
Science	4	2015	21	38	40	1	41	79	14.98	32.25	33.29	19.49	52.78	85.02
Science	8	2015	25	35	38	2	40	75	17.76	28.29	31.50	22.45	53.95	82.24

Note: NAEP assessed student knowledge and skills in Reading, while Wisconsin Forward Exam assessed student knowledge and skills in ELA, which included Reading, Listening, and Writing.
Note: NAEP data are from https://nces.ed.gov/nationsreportcard.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Part 11: Summary Recommendations

Results and key findings of the Spring 2019 Wisconsin Forward Exam administration are presented throughout the body of this report. This last section of the report presents some recommendations for DPI consideration.

The 2019 Wisconsin Forward Exam administration was the fourth administration of the assessment. For four consecutive years, the assessment results were reported on the same scales and students were classified into the proficiency levels using the same cut scores, allowing for longitudinal tracking of student performance in ELA, Mathematics, and Social Studies. Using the same scales and the same cut scores for Wisconsin assessments allows for monitoring student growth across administration years. New test scales were established and new performance level cut scores were set for Science assessments after the Spring 2019 test administration. The Spring 2019 assessment results will serve as the new baseline for monitoring student performance in Science across years.

Following the Spring 2016 through 2019 field-testing of new test items in Wisconsin, DRC recommends 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. DRC also recommends 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 development.

DRC recommends continuing to use an artificial intelligence (AI) engine in the 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 good agreement with scores by trained human scorers.

From the psychometric perspective, it was noticed that the ELA grade 4 and 5 tests and ELA grade 7 and 8 tests are of comparable difficulty, as indicated by the test characteristic curves included in Part 6 of this report. It was also noticed that the ELA grade 5 and 8 assessments contained relatively few items accurately measuring high achieving students. In order to achieve better ordinality of the ELA assessments' overall difficulty across grade levels and provide a better measurement at the upper end of the ability scale, a few more difficult items could be added to the grade 5 and 8 tests. However, it should be noted that because equating requires tests to maintain a similar level of difficulty from year to year, large shifts in the test rigor may require a cut score review and an examination whether a new test scale should be set.

In Mathematics grades 5 through 8, more than 2% of students received the lowest obtainable scale score (LOSS). While fewer students received the LOSS in Mathematics in Spring 2019 compared to Spring 2018, the Mathematics assessments continue to be difficult for some students. The response patterns of students at the LOSS in Mathematics indicated that these students typically answered very few MC items and none of the non-MC items. As explained in Part 6 of this report, for these students to receive a scale score above the LOSS, they would need to correctly answer more items, including some non-MC items. Therefore, DRC continues to recommend that some easier non-MC items be included in the future forms of Mathematics tests.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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.

Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing bias in item response theory. *Applied Psychological Measurement*, *12*(3), 253–260.

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

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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Fitzpatrick, A. R., & Julian, M. W. (1996). *Two studies comparing the parameter estimates produced by PARDUX and PARSCALE*. Unpublished manuscript.

Green, D. R. (December 1975). *Procedures for assessing bias in achievement tests*. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, *10*(3), 159–170.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.

Holland, P. W., & Thayer, D. T. (1985). *An alternate definition of the ETS delta scale of item difficulty* (Research Report No. RR-85-43). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (April 1986). *Differential item performance and the Mantel-Haenszel procedure*. Paper presented at the annual meeting of the American Educational Research Association Annual Meeting, San Francisco, CA.

Karantonis, A., & Sireci, S. G. (2006). The Bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, *25*(1), 4–12.

Kim, D. (2005). KKCLASS [Computer program]. Unpublished.

Kim, D., Barton, K., & Kim, J. (April 2007). *Estimating classification consistency and classification accuracy with pattern scoring*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kim, D., Choi, S., Um, K., & Kim, J. (April 2006). *A comparison of methods for estimating classification consistency*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag.

Kolen, M., & Kim, D. (2004). [Personal correspondence].

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Linn, R. L. (Ed.) (1989). *Educational measurement* (3rd ed.). New York, NY: Macmillan.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*(2), 179–197.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719–748.

McCallin, R. C. (2006). Test administration. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 625–652). Mahwah, NJ: Lawrence Erlbaum Associates.

Michaelides, M. P., & Haertel, E. H. (2004). *Sampling of common items: An unrecognized source of error in test equating*. Los Angeles, CA: Center for the Study of Evaluation.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, *14*(1), 59–71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

Muraki, E., & Bock, R. D. (1991). PARSCALE: Parameter scaling of rating data [Computer program]. Chicago, IL: Scientific Software, Inc.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision theoretic formulation. *Journal of Educational Measurement*, *11*(4), 263–267.

Swineford, F. (1956). *Technical manual for users of test analysis* (Statistical Report No. 56-42). Princeton, NJ: Educational Testing Service.

Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*(2), 175–186.

Thissen, D. (1990). MULTILOG: Multiple categorical item analysis and test scoring (Version 6) [Computer program]. Chicago, IL: Scientific Software, Inc.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessment* (Synthesis Report No. 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, *5*(2), 245–262.

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement*, *21*(2), 93–111.

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

Yen, W. M., & Burket, G. R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement*, *34*(4), 293–313.

Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education*, *4*(3), 209–228.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, *30*(3), 233–251.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix A

Summer 2017 Item Review Training Slides

Copyright © 2019 by the Wisconsin Department of Public Instruction

Wisconsin Forward Exam Item Review

Madison, WI
July/August 2017

Purpose of Meeting

- Provide overview of the Wisconsin Forward Exam
- Provide item review training
- Review items for potential placement on Wisconsin Forward Exam

Copyright © 2019 by the Wisconsin Department of Public Instruction

Wisconsin's Definition of College and Career Readiness

Wisconsin Forward Exam

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Security and Confidentiality

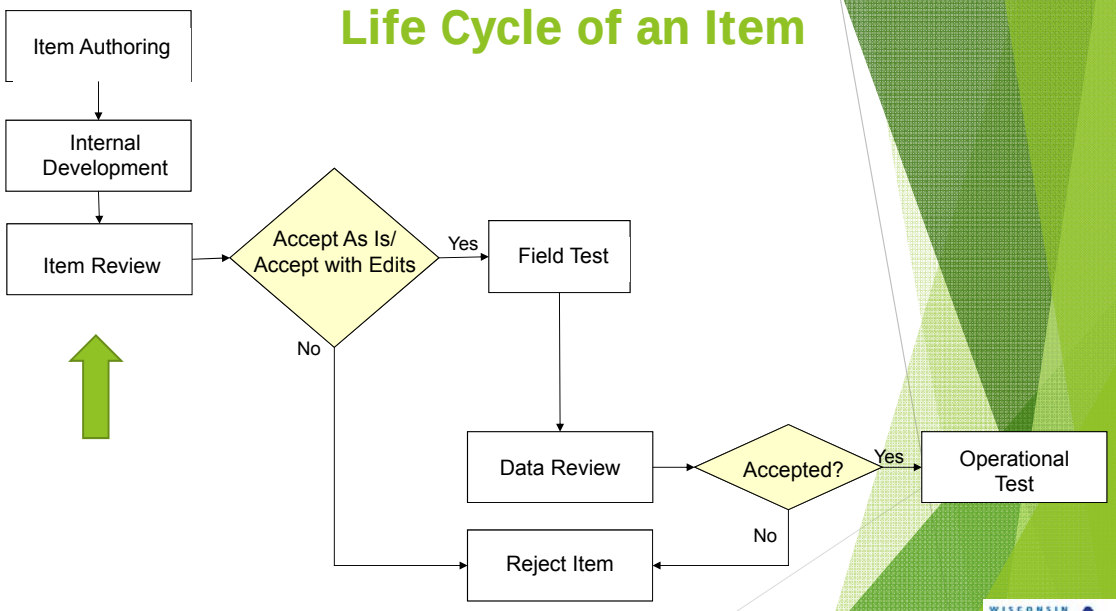
- **Critical Importance of Security**
  - Security/Nondisclosure Agreement
  - Security of passage and item content
  - Note-taking policy
  - Cell phone and personal computer use
  - Communication following the meeting

Item Review Process

- Reviews will be completed online using the same testing engine students use- INSIGHT
  - Allows for reviewer interaction with item functionality, particularly technology-enhanced items
  - Facilitator will provide specific directions for logging in to begin reviews

Copyright © 2019 by the Wisconsin Department of Public Instruction

Life Cycle of an Item



Item Types

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Selected-Response Item Type-Multiple Choice (MC)

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

MC Sample

A student is writing a report about how elevators make modern life easier.

Which sentence would **best** support the topic?

(a) Climbing stairs is a great way to exercise.

(b) Some buildings have only one floor and do not need an elevator.

(c) An elevator saves time, especially if the building it serves is very tall.

(d) The Empire State Building in New York has a viewing deck on the 102nd floor.

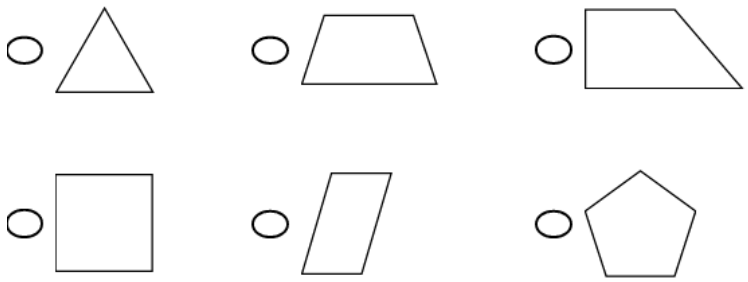
Copyright © 2019 by the Wisconsin Department of Public Instruction

Selected-Response Item Type- ESR

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

ESR Sample

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



Copyright © 2019 by the Wisconsin Department of Public Instruction

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 (If Part A incorrect = 0)

EBSR Sample

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

Part A

What is the main way the passage "Public Transportation, Not for Everyone" supports the claim that taking public transportation may be problematic for some people?

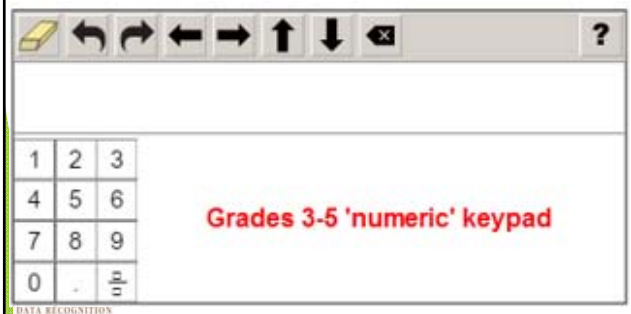
Part B

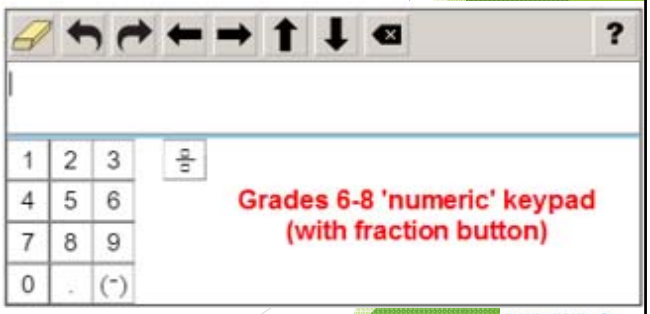
Which sentence from the passage **best** supports your answer in part A?

Copyright © 2019 by the Wisconsin Department of Public Instruction

Scorable Equation/Numeric Item Type-SEQ

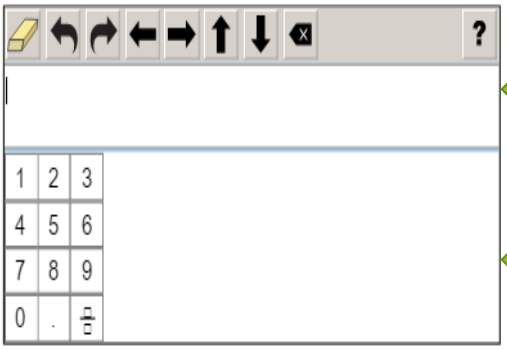
- Used in Mathematics Items
- Grade-level specific keypad that allows for more guided input of student responses





SEQ Sample

A rectangular section of a kitchen wall will be tiled. What is the area, in square feet, of the section of wall that will be tiled?



Student Response Area

Grade 5 Keypad

Copyright © 2019 by the Wisconsin Department of Public Instruction

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
- Character counter feature

TDA Sample

Both passages focus on creatures from two different species helping each other. Write a response explaining how both passages show ways in which people and animals help each other. Use evidence from **both** passages to support your response.

0/5000

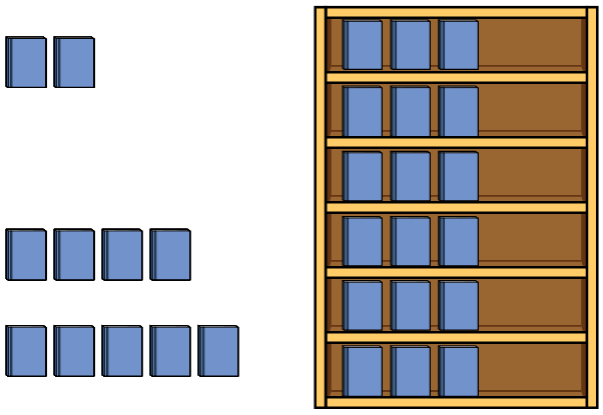
Copyright © 2019 by the Wisconsin Department of Public Instruction

Technology Enhanced (TE)

- TE items present in all content areas
- Interactive
- Wide Variety: clock input, angle draw, drop down list, matching, graphing, highlighting text, drag and drop

TE Sample Item

Clayton puts 18 books in a bookcase. He puts the same number of books on each shelf. Move groups of books to the bookcase to show how Clayton could arrange them.



Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

DOK Levels

DOK 1 Recall and Reproduction

DOK 2 Skills and Concepts

DOK 3 Strategic Thinking and Reasoning

DOK 4 Extended Thinking

(rarely on standardized assessments — more "project-like" or on performance assessments)

DOK 1: Recall and Reproduction

- Students demonstrate a rote response, use a well-known formula, or follow a simple procedure.
- A "simple" procedure is well defined and typically involves only **one** step.

Key Words: identify, recall, recognize facts, use, measure, solve a one-step problem

Copyright © 2019 by the Wisconsin Department of Public Instruction

DOK 2: Skills and Concepts

- 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: explain, categorize, use context clues, select a procedure, compare/contrast

DOK 2-(cont.)

Activities may include:

- Making observations/collecting information
- Classifying/comparing information
- Organizing/displaying data or information in tables and graphs

Note: Some action verbs, such as "explain," "describe," or "interpret," could be classified at different DOK levels, depending on the complexity of the action.

Copyright © 2019 by the Wisconsin Department of Public Instruction

DOK 3: Strategic Thinking and Reasoning

- Students demonstrate deep understanding through planning, using evidence, and exhibiting higher levels of cognitive reasoning.

Key Words: connect ideas, explain thinking, cite evidence, analyze, apply a concept,

DOK 3-(cont.)

Activities may include the following:

- Use concepts to solve non-routine problems
- Describe how word choice, point of view or bias, may help the readers' interpretation of text
- Apply a concept in a new context
- Cite evidence and develop a logical argument for concepts
- Compare information within or across data sets

Copyright © 2019 by the Wisconsin Department of Public Instruction

DOK 4: Extended Thinking

- Students demonstrate an integrated use of higher order thinking processes such as critical and creative and productive thinking, reflection, and adjustment of plans.

Key words: analyze, synthesize, examine and explain, describe and illustrate common themes

DOK 4- cont.

- Higher order thinking skills

Activities may include the following:

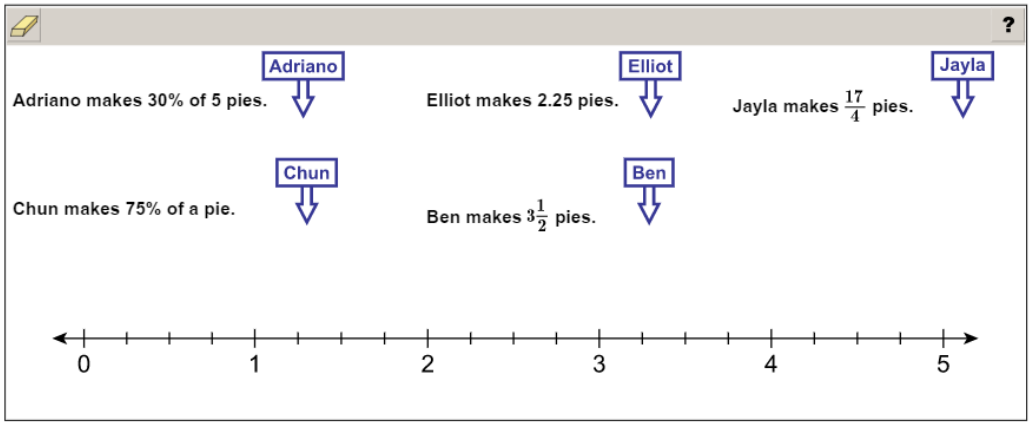
- Developing generalizations
- Analyzing abstract themes
- Evaluating relevancy, accuracy, and completeness of information from multiple sources

Key words: analyze, synthesize, examine and explain, describe and illustrate common themes

Copyright © 2019 by the Wisconsin Department of Public Instruction

Sample of Difficult DOK 1 Item

Five students take part in a pie-making contest. A student wins the contest by making the most pies before time is up. Below are the five students' results. Drag each student's arrow to the point on the number line that corresponds to how many pies he or she makes.



Click the name of the student who wins the pie-making contest.

Adriano Ben Chun Jayla Elliot

Sample of an Easy DOK 3 Item

The area model below is used to solve a multiplication problem.

20 × 10	3 × 10
20 × 7	3 × 7

200 + 30 + 140 + 21 = 391

Fill in the multiplication problem that goes with this area model.

[ ] × [ ] = 391

Copyright © 2019 by the Wisconsin Department of Public Instruction

Special Note:

- DOK is used by item writers to gauge the *cognitive level* of item, it **does not** correlate to the *difficulty* of the item.

Item Review Process

- **Reviews will be completed in groups and individually**
- **Items will be reviewed for:**
  - Standard alignment
  - Grade-level appropriateness
  - Correct Key(s)
  - Rigor-level alignment
    - DOK level
    - Bias and sensitivity concerns

Copyright © 2019 by the Wisconsin Department of Public Instruction

Item Review Tally Sheet

Step 2

Step 4

Session Number	Sequence #	Item ID	Passage Title	Standard	Item Type	Key(s)	DOK	Bias/Sensitivity Comments	Accept (A); /Accept with Revisions A (AR); Dissenting View (DV)	Comments

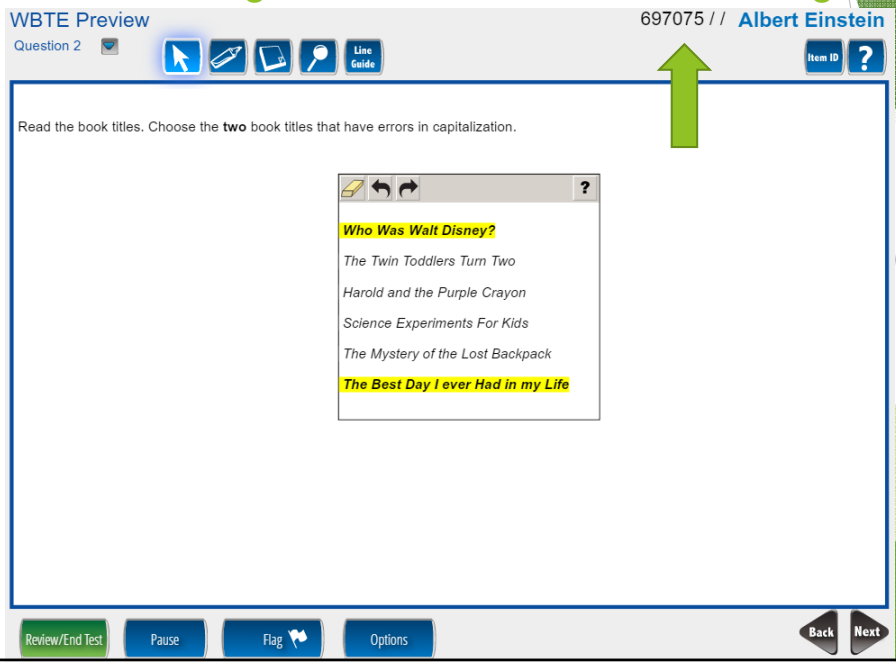
Step 1

Step 3

Step 5

Step 6

Evaluating an Item: Grade 3 Writing



Copyright © 2019 by the Wisconsin Department of Public Instruction

Step 1: Standard Alignment

After reading item ask yourself:

Does the standard listed match the state standard?

- Each member will have copy of standards
- Match item to appropriate standard as noted on item rating sheet
- Indicate agreement of alignment on item rating sheet or recommend new standard

Step 2: Check the Keys

- Is the key (or keys) listed correct?
  - If yes, move on to step 3
  - If no, discuss with committee and note new key(s)

Copyright © 2019 by the Wisconsin Department of Public Instruction

Step 3: Confirm the DOK Levels

- Is the DOK level listed correct?
  - If yes, move on to step 4.
  - If no, mark your thinking and discuss with committee.

Step 4: Check for Bias and Sensitivity

- Stereotyping
- Gender
- Regional or geographical
- Ethnic or cultural
- Socioeconomic class
- Persons with a disability
- Ageism
- Religious

Copyright © 2019 by the Wisconsin Department of Public Instruction

Also Keep in Mind…Technical Design

- Does the item meet requirements for technical quality?
  - Stem: Complete question/problem; does not clue correct answer(s)
  - Correct answer(s): clear and accurate
  - Distractors (or incorrect options): may contain common misperceptions or processes
  - Graphics/visuals: compliment and support item

Be Mindful of 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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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.

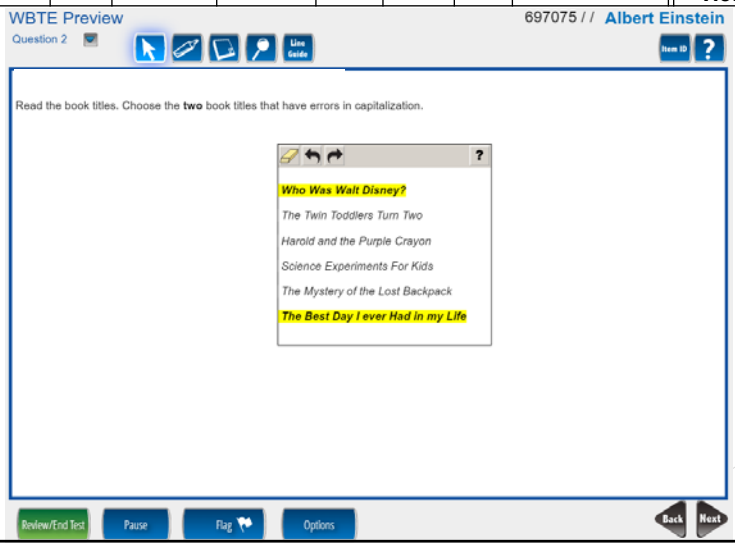
Steps 5 and 6: Mark Comments

In document, mark column noting the following:

- Accept- "A"
  - Item is OK as is
- Accept with Revisions- "AR"
  - Accept but apply recommended edits
- Dissenting View- "DV"
  - Item contains major flaws; do not recommend placement on assessment
- Additional comments as needed

Copyright © 2019 by the Wisconsin Department of Public Instruction

Session Number	Sequence #	Item ID	Passage Title	Standard	Item Type	Key(s)	DOK	Bias/Sensitivity Comments	Accept (A); /Accept with Revisions A (AR); Dissenting View (DV)	Comments



Main Questions to Ask During Review

- Does the item provide for an optimal standard assessment of all students?
- Are there items written to ALL ability levels?
  - OK to have easy items

Copyright © 2019 by the Wisconsin Department of Public Instruction

When to Edit an Item

- If the subject matter is above grade level or out of scope for the standard/course.
- If there is an opportunity to make the item/passage/stimulus easier for students to understand.
- If the topic or language is inappropriate, controversial, or inflammatory.

What if I Disagree with the Committee?

- Speak up! It's possible that another committee member has the same concern, or you may have noticed something that other committee members have not.
- Record your dissenting view on the item review tracking sheet. Discussion by all is encouraged; however, if you choose not to share your opinion, your facilitator can voice your concern for you.
- DRC and DPI will reconcile any major disagreements/concerns noted on tracking sheet following the meeting. A consensus is not always needed.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Item Review Process: Summary

- Standard Alignment
- Key(s)
- DOK Levels
- Grade-level Appropriateness
- Bias and Sensitivity

Roles & Responsibilities

- **Participants**
  - Item Review
- **DRC Facilitators**
  - Lead the group through the agenda
  - Encourage interaction
  - Lead discussions
  - Collect secure materials
- **DPI and DRC**
  - Answer questions

Copyright © 2019 by the Wisconsin Department of Public Instruction

Roles & Responsibilities

- Educators
  - Invest yourself in the process
  - Share your opinions
  - Listen to your colleagues

Questions?

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix B

Spring 2018 Field Test Data Review Training Slides

Copyright © 2019 by the Wisconsin Department of Public Instruction

Wisconsin Forward Exam Item Data Review

Wisconsin Department of Public Instruction
&
Data Recognition Corporation

August 2018

1
1

Purpose

- Establish a robust pool of items for use in new test development to ensure proper representation:
  - Content standards
  - Test design
- General statistical guidelines are presented
  - No item flags are created equal
  - Guidelines vs. hard-and-fast rules
  - Item content needs to be considered as well
  - Approving an item does not guarantee its appearance on a future test, but rather maximizes the size of the pool for item selection during test development.

2

Copyright © 2019 by the Wisconsin Department of Public Instruction

Key Objectives

- Review item development process
- Review and understand item card layout
- Understand and interpret item statistics
- Review item cards for a few Wisconsin field test items with different statistics
- Apply knowledge of item statistics to evaluate the remaining field test items

3

Some Definitions

- **Item pool**: Set of items on a given test scale that are available for operational test construction
- **Item statistics**: Statistical values generated during data analysis after item administration (more detail later)
- **Field tested items**: Items that have been embedded among operational items to gather item statistics before placing them on the operational tests
- **Operational items**: Items that have already been used in an operational test administration

4

Copyright © 2019 by the Wisconsin Department of Public Instruction

Some More Definitions

- **Item type**: refers to the format of the item
  - Multiple-choice (MC)
  - Technology-enhanced (TE)
  - Multi-select (MS)
  - Short answers (SA)
  - Evidence-based selected response (EBSR)
  - Text Dependent Analysis (TDA)
- **Item Scoring**: refers to the score range
  - ***Dichotomous***: Item is scored as 0 or 1 (Math, Science, Social Studies)
  - ***Polytomous***: Item has a range of possible scores from 0 to greater than 1 (ELA only 2-point EBSR, TE, MS, and 4-point TDA)

5

Sample Item Card

1. The distances Esteban runs during the first 7 days form a pattern. The pattern starts with [illegible] miles. The table below shows the mileage he runs each day.

Day	Miles Run
Monday	1.2
Tuesday	1.6
Wednesday	2.0
Thursday	2.4
Friday	2.8
Saturday	3.2
Sunday	3.6

What is the rule of the pattern displayed in the table?

A. The pattern increases by 4 miles every day.
B. The pattern is skip counting by 4.
C. The pattern increases by $\frac{4}{10}$ of a mile every day.
D. The pattern increases by 1.4 miles each day.

Stem

Distractors

Item ID

Content Area

Grade

Standard

Key(s)

Item ID	
851887	
Content Area	
Mathematics	
Passage ID	
122959	
Passage Title	
Training	
Grade	
5	
Standards	
MLS2016: 5.RA.A.2	
Item Type	
Multiple Choice	
Points	
1	
Key	
C	
Calculator	
No	
Previous Use	

6

Copyright © 2019 by the Wisconsin Department of Public Instruction

Sample Item Card (cont.)

Administration(s)

Form Name	Use Function	Seq	Period	Year	Session	Calc	Model/Ext	Grade
Q5 MA1	FT	51	Spring	2017	3	No	3PL/GPL	5

Admin Info

Traditional Statistics

N	P-Val	Mean	Item Total Corr
29080	0.44		0.45

Classical Stats

Fit Statistics

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit
						60.43	

Item Fit

IRT Statistics

Label	Final	Final S.E.	Preliminary	Preliminary S.E.	Displ
Slope	2.26				
Location	0.49				
Asymptote	0.21				

IRT stats

Distractor/Step Specific

Label	Percent	Corr	Avg Meas	>Step Meas
A	0.24	-0.19		
B	0.18	-0.19		
C*	0.44	0.45		
D	0.13	-0.20		
OMITS	0.00			

Distractor or Score Point Stats

DIF Analysis

Category	Bias Code	Num Value	N - Ref	N - Focal
ACC	A	0.18	24589	2392
MALE/FEMALE	B-	-1.22	15233	13863
WHITE/AMIN			17974	102
WHITE/ASIAN	A	-0.36	19525	622
WHITE/BLACK	A-	-0.88	19444	5806
WHITE/HISPANIC	A-	-0.68	19467	2058
WHITE/MULTI	A	-0.35	19444	982

DIF Index

7

Classical Statistics: Item Difficulty

Difficulty

- "**P-Value**" : proportion of students who answered an item correctly (or a percent of maximum points possible for polytomously scored items)
  - 0.0 means all students answered incorrectly
  - 1.0 means all students answered correctly
  - The higher the p-value, the easier the item
- "**Mean**" : Average score obtained by students on polytomously scored items
  - The higher the mean, the easier the item

Dichotomously Scored Item

Traditional Statistics

N	P-Val	Mean	Item Total Corr
4349	0.73		0.49

Polytomously Scored Item

Traditional Statistics

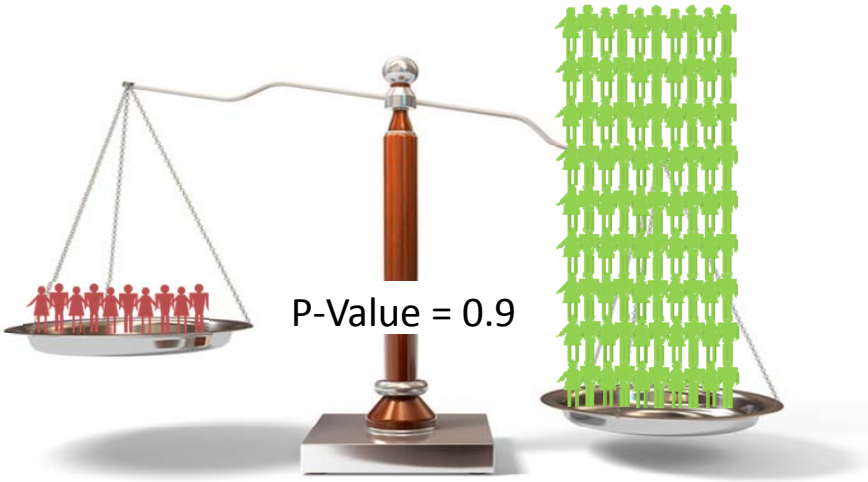
N	P-Val	Mean	Item Total Corr
2008	0.25	1.01	0.65

8

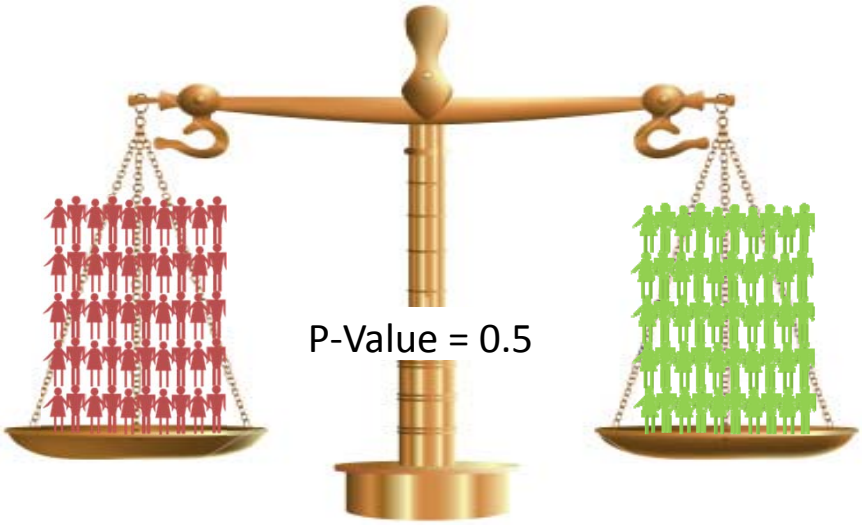
8

Copyright © 2019 by the Wisconsin Department of Public Instruction

Visualizing P-Values



Visualizing P-Values



10

Copyright © 2019 by the Wisconsin Department of Public Instruction

P-Values on Item Card (MC Items)

- (*) indicates key
- Other answer options (distractors)
  - Proportions selecting each distractor are also displayed.
  - MULTS means multiple marks.
  - OMITS means student omitted.

Traditional Statistics

N	P-Val	Mean	Item Total Corr
14016	0.78		0.48

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Step Meas
A*	0.78	0.48		
B	0.06	-0.29		
C	0.11	-0.25		
D	0.04	-0.23		
MULTS	0.00			
OMITS	0.00			

11

Item Stats for Polytomously Scored Items

Traditional Statistics

N	P-Val	Mean	Item Total Corr
1101	0.25	1.00	0.57

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Step Meas
0	0.50	-0.51		
1	0.22	0.04		
2	0.12	0.26		
3	0.11	0.29		
4	0.05	0.28		
BL	0.01			

- **Mean**
  - Average student score on that item
- **Item Total Corr**
  - Item-Total Correlation
- **Proportion**
  - Percent of students receiving a certain score point
- **Score Distribution**
  - Expected to be monotonic

12

Copyright © 2019 by the Wisconsin Department of Public Instruction

Polytomously Scored Items

- The mean gives a general idea of item difficulty but can sometimes be deceptive.

Proportion of students			
Score 0	**Score 1**	**Score 2**	**Item Mean**
0.40	0.20	0.40	1
0.15	0.70	0.15	1
0.33	0.33	0.33	1

- Use the score point proportions to determine if the distribution is reasonable.
- We want some students in all score-point categories.
  - item parameters cannot be estimated for the category with no or very few students.

13

Guidelines for Polytomously Scored Items

- For a 2-point item, a mean of 1.8 and above may be too easy, and a mean of 0.2 and below may be too difficult.
- For a 4-point item, a mean of 3.6 and above may be too easy and a mean of 0.4 and below may be too difficult.

14

Copyright © 2019 by the Wisconsin Department of Public Instruction

Item Difficulty: Summary

Theoretcial Range

P-Value: 0 to 1
- 0 = no students answered item correctly
- 1 = all students answered item correctly

- Lower values = more difficult (hard)
- Higher values = less difficult (easy)

Targeted Range

- P-Value: 0.20 to 0.90
- Items outside of target range may be approved if content is appropriate

Content Consideration

- We need to build tests with a wide range of p-values (generally 0.20 – 0.90) in order to effectively place students into the four performance categories
  - Hard items to distinguish between Proficient/Advanced
  - Easy items to distinguish between Below Basic/Basic
- Why did most students answer this item correctly or incorrectly?
- Are there any reasons other than item difficulty to support a decision to ACCEPT or REJECT this item?

15

15

Classical Statistics: Item Discrimination

Discrimination

- Measures item's ability to differentiate between high and low performers
- Item-Total Test Correlation: Correlation of examinee raw scores on a single item with their raw scores on all remaining test items (-1.0 to +1.0)
  - Positive—high achievers outperformed low achievers (targeted).
  - Negative—low achievers outperformed high achievers (unexpected).
  - Around zero—high and low achievers performed about the same on an item (not desired).

Traditional Statistics

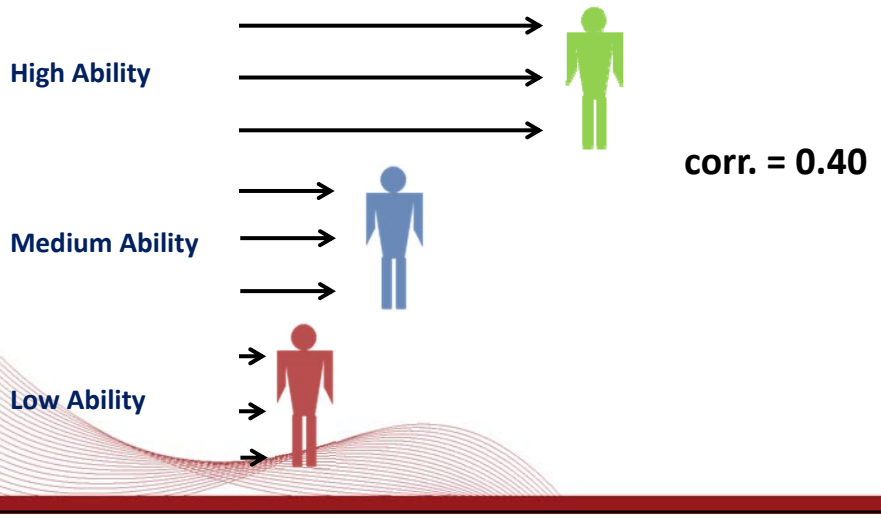
N	P-Val	Mean	Item Total Corr
4349	0.73		0.49

16

16

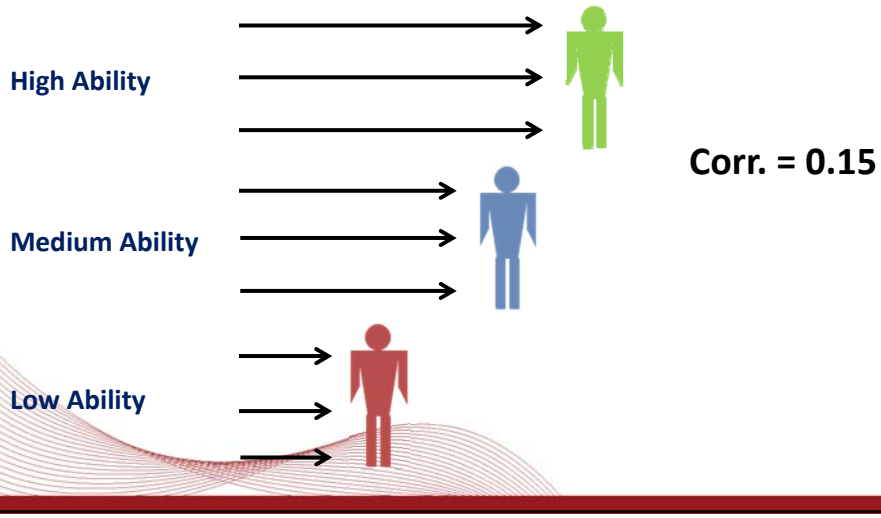
Copyright © 2019 by the Wisconsin Department of Public Instruction

Visualizing Item-Total Test Correlation



17

Visualizing Item-Total Test Correlation



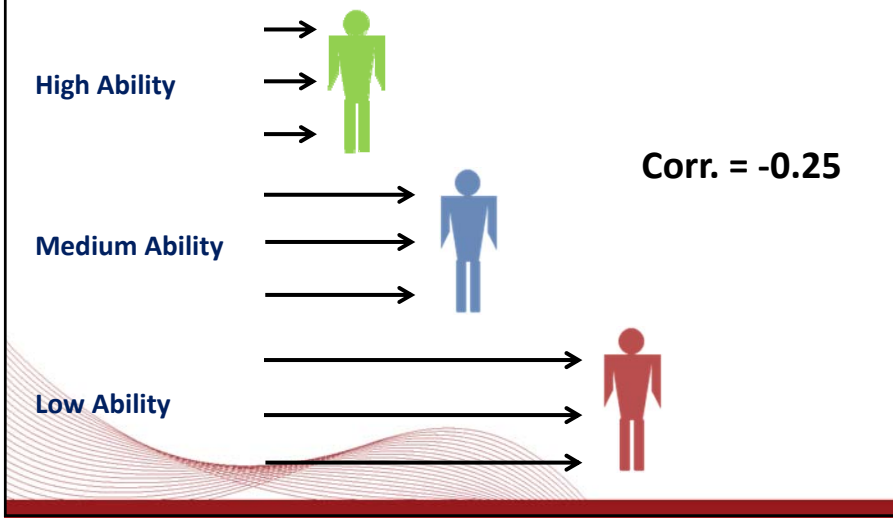
18

Copyright © 2019 by the Wisconsin Department of Public Instruction

Visualizing Item-Total Test Correlation

19

Visualizing Item-Total Test Correlation



20

Copyright © 2019 by the Wisconsin Department of Public Instruction

Item-Total Test Correlation on Item Card

- (*) indicates key

Traditional Statistics

N	P-Val	Mean	Item Total Corr
14016	0.78		0.48

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Step Meas
A*	0.78	0.48		
B	0.06	-0.29		
C	0.11	-0.25		
D	0.04	-0.23		
MULTS	0.00			
OMITS	0.00			

21

Item Discrimination: Summary

Theoretical Range

- Ranges from -1 to +1
- -1 = "perfect" negative relationship
- 0 = no linear relationship
- +1 = "perfect" positive relationship

Targeted Range

- at or above 0.15
  - Smaller sometimes is okay, depending on difficulty
  - Items with negative or around 0.0 item discrimination are poorly discriminating and often should be rejected

Content Consideration

- Why is this item less able to differentiate between high and low achievers?
- Is the low discrimination associated with extreme low or high P-Values (item difficulty)?
- Are there any other reasons other than item discrimination to support your decision on ACCEPTING or REJECTING this item?

22

22

Copyright © 2019 by the Wisconsin Department of Public Instruction

Distractor Specifc Analysis (MC Items)

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Threshold
A	0.05	-0.22		
B	0.10	-0.26		
C	0.12	-0.28		
D*	0.73	0.49		
MULTS	0.00			
OMITS	0.00			

Guideline

- **MC items:**
  - **Correlations for the distractors should be negative.**
  - **Correlations for the distractors should never be higher than correlation for the correct answer**
  - **Proportion of distractor < proportion of key**

Content Consideraton

- **Is the correlation of selecting any incorrect option greater than 0? If yes, why does this option distract more high achievers than low achievers?**
- **Is the proportion of selecting any incorrect option greater than the proportion of selecting the key? If yes, why?**

23

23

Score Point-Specific Analysis

Distractor/Step Specific

Label	Percent	Corr	Avg Meas	>Step Meas
0	0.39	-0.46		
1	0.09	-0.09		
2	0.52	0.51		
BL	0.00			

Guideline

- **Non-MC items:**
  - **Correlations for the score 0 expected to be negative**
  - **Correlation for higher scores should be positive**
- **Proportion for each each score point>=0.05 – desirable property**

Content Consideration

Non-MC items

- **Is the proportion to a score point <0.05? If yes, is there a reason that explains why so few students received this score point?**
- **Is the pattern of item score correlation as expected?**

24

24

Copyright © 2019 by the Wisconsin Department of Public Instruction

IRT: Item Fit and Non-Convergence

IRT Statistics

Item Fit

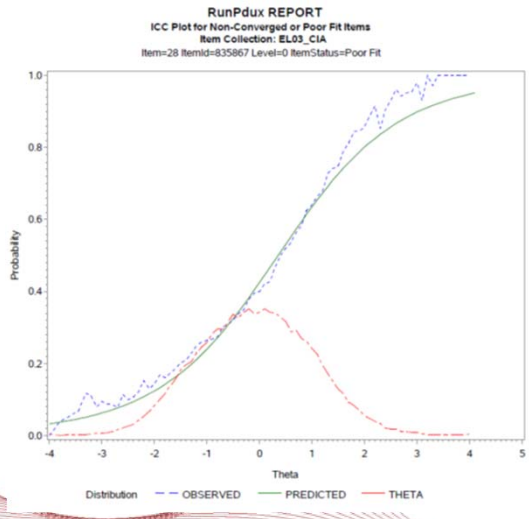
- IRT statistic obtained after item calibration
- Measures how well the student responses to each item fit the test data (by comparing parameter estimation prediction relative to the observed data)
- Item is flagged when the observed data pattern differs from the predicted probability of responding to the item.
- There is no specific criterion value for the fit flag: criterion is dependent on the number of students taking the item

Item Non-Convergence

- Item parameters cannot be estimated and the item is not eligible for future use (5 FT ELA items, 1 FT Science item, and 1 FT Social Studies item)

25

Item Misfit (Graphical Representation)



26

Copyright © 2019 by the Wisconsin Department of Public Instruction

Item Fit on Item Cards

Fit Statistics

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit
						11.58	MISFIT

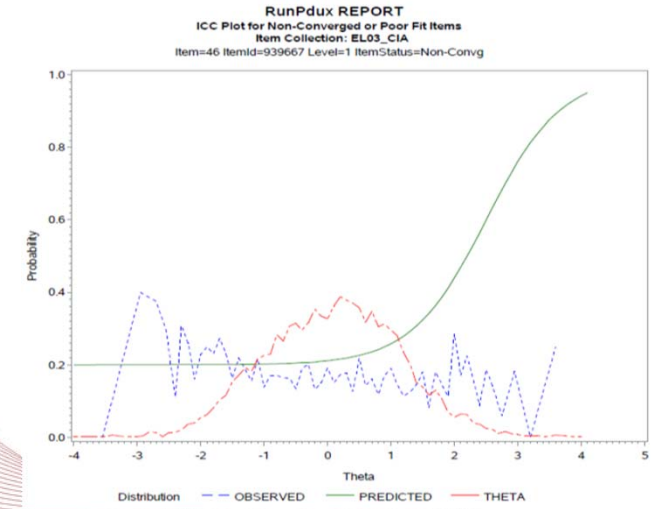
Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit
						2.58	

Non-Convergent Items (no Item Fit or IRT Stats)

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit

27

Item Non-Convergence (Graphical Representation)



28

Copyright © 2019 by the Wisconsin Department of Public Instruction

Non-Convergent Items on Item Cards

- Classical statistics and DIF index present
- No item fit statistics available
- No IRT statistics (parameters) available

Fit Statistics

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit

IRT Statistics

Label	Final	Final S.E.
Slope		
Location		
Asymptote		

29

Item Fit and Non-Convergence: Summary

Item Fit

- **Item misfit is not a serious flag by itself if the misfit happens at the ends of ability scale where there are few students**
- **If the misfit is in the middle of ability scale then the IRT model used to calibrate the data does not fit the item well**
- **It is best to avoid selecting misfitting items to be anchor items**

Non-convergence

- **A fatal flag (item parameters are not estimated and item classical statistics are typically poor)**

30

Copyright © 2019 by the Wisconsin Department of Public Instruction

Differential Item Functioning

DIF

- Procedure used to identify items that function differently for particular groups of students (e.g., gender, ethnicity, and disability status, SES status, and LEP status).
- Hypothesis is that test takers with similar knowledge or ability should perform in similar ways on a test item.
- Items are flagged if they do not behave the same in different groups of students, after controlling for student ability.

Procedure

- Compares "focal" vs. "reference" groups.
- Reference groups: Males, Whites, students w/out disabilites, students not SES-disadvantaged, English proficient students , students not using accommodations.
- Focal groups: Females, non-White ethnic groups, students with disabilities, SES-disadvantaged students, LEP students, and students using accommodations

31 31

Differential Item Functioning

Guideline

- Each item is assigned a bias code of A, B, or C.
  - A – minor DIF (no DIF)
  - B – moderate DIF
  - C – Large DIF

DIF signs: "−" favors Reference group; '+' favors Focal group.

- Only items with C (i.e., large) DIF require review. Items with C DIF may be acceptable if no potential bias causes the differential item functioning.

Content Consideration

- Is there anything in the content or format of the item that may interfere with, or advantage, one group of students over another based on:
  - Gender?
  - Ethnicity?
  - Disability status, SES status, LEP status, accommodation use?

32 32

Copyright © 2019 by the Wisconsin Department of Public Instruction

An Index for DIF

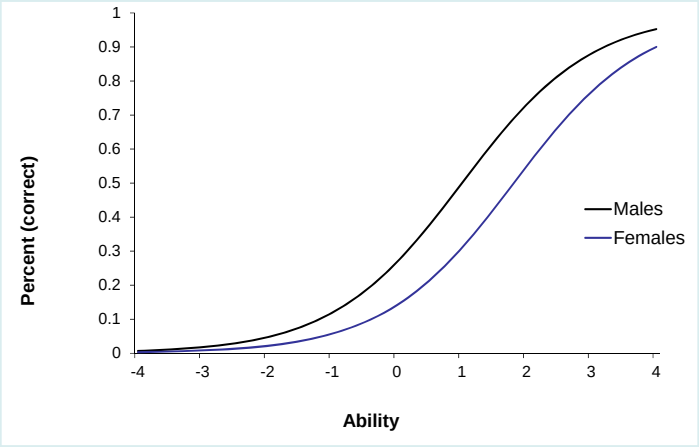
DIF Analysis

Category	Bias Code	Num Value	N - Ref	N - Focal
DISAB	A	-0.02	56644	7107
ECODISAD	A-	-0.04	37424	26416
LEP	A-	-0.05	58028	5718
MALEFEMALE	A-	-0.05	32508	31334
WHITEAMIN	A	0.01	41397	766
WHITEASIAN	A-	-0.09	42293	2557
WHITEBLACK	A	-0.01	42225	7116
WHITEHISPANIC	A	-0.02	42290	8531
WHITEMULTI	A	-0.02	42282	2566

Reference Group/Focal Group

33

Visualizing DIF (Gender)



34

34

Copyright © 2019 by the Wisconsin Department of Public Instruction

DIF: Summary

- All biased items should show DIF, but **Not** all items with DIF will be biased.
  - The smaller sample sizes of the minority ethnicity groups causes many false positives.
  - DIF not computed if focal group N <200.
  - You **must** be able to provide a reason for the bias to call the item biased.

35

Summary of Item Flags

- P-value less than 0.20 or higher than 0.90
- Item-total test correlation < 0.15
  - Negative or close to 0 item-total test correlation is a very serious flag, especially when combined with a positive correlation for a distractor for MC items
- Positive pt. biserial correlation for a distractor
  - Especially if pt. biserial for a distractor is higher than pt. biserial for the correct option
- Fewer than 5% of students at each score point for non-MC items
  - No students at any of the score points leads to collapsed levels
- Poor Fit
- Non-Convergence (kills the item)
- Large DIF (C +/-)
- Omit rates > 3% (not used in this data review)

36

Copyright © 2019 by the Wisconsin Department of Public Instruction

Unique for ELA

- DPI will be reviewing a selection of TDA items
  - **Set one**: TDA items that WI educators have previously reviewed; not placed on Forward
  - **Set two**: TDA items that WI educators saw this week; data are from another state
  - **Set three**: TDA items appearing on 2018 Forward; Wisconsin student data
- Review the data and determine which item at each grade level will be placed on 2019 Forward Exam

37

Roles, Responsibilities, Questions

- DPI
  - Review Spring 2018 Wisconsin field test item data
  - Accept or reject items
- DRC
  - Facilitate Data Review
  - Answer DPI questions
- Questions?

38

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix C

Spring 2019 English Language Arts Operational Test Maps

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-1. English Language Arts, Grade 3 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
3	1	1	TDA	OP	4	3	3.W.3	Writing
3	2	2	MC	OP	1	2	3.W.1.c	Writing
3	2	3	TE	OP	2	3	3.W.1.b	Writing
3	2	4	MC	OP	1	2	3.W.8	Writing
3	2	5	MC	OP	1	2	3.W.8	Writing
3	2	6	TE	OP	2	2	3.W.8	Writing
3	2	7	MC	OP	1	2	3.W.8	Writing
3	2	8	MC	OP	1	3	3.W.8	Writing
3	2	9	MC	OP	1	1	3.L.1.i	Writing
3	2	10	TE	OP	1	1	3.L.2.c	Writing
3	2	11	MC	OP	1	1	3.L.1.f	Writing
3	2	12	MC	OP	1	1	3.L.1.d	Writing
3	2	13	MC	OP	1	2	3.L.1.d	Writing
3	2	14	TE	OP	2	2	3.L.1.h	Writing
3	3	15	MC	OP	1	1	3.SL.3	Listening
3	3	16	TE	OP	2	2	3.SL.3	Listening
3	3	17	MC	OP	1	2	3.SL.2	Listening
3	3	18	MS	OP	2	2	3.SL.3	Listening
3	3	19	MC	OP	1	2	3.SL.2	Listening
3	4	20	MC	OP	1	3	3.RL.7	Reading
3	4	21	MC	OP	1	2	3.RL.4	Reading
3	4	22	MC	OP	1	3	3.RL.5	Reading
3	4	23	MC	OP	1	2	3.RL.1	Reading
3	4	24	MC	OP	1	3	3.RL.6	Reading
3	4	25	TE	OP	2	2	3.RL.2	Reading
3	4	26	MC	OP	1	2	3.RL.1	Reading
3	4	27	MS	OP	2	2	3.L.4	Reading
3	4	28	MC	OP	1	3	3.RL.3	Reading
3	4	29	MC	OP	1	2	3.RL.3	Reading
3	4	30	MC	OP	1	2	3.RL.5	Reading
3	4	31	EBSR	OP	2	3	3.RL.2	Reading
3	4	32	MC	OP	1	2	3.L.4.RI	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-1. English Language Arts, Grade 3 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
3	4	33	MC	OP	1	2	3.RI.5	Reading
3	4	34	MC	OP	1	2	3.RI.1	Reading
3	4	35	MC	OP	1	2	3.RI.1	Reading
3	4	36	MC	OP	1	2	3.RI.7	Reading
3	4	37	TE	OP	2	2	3.RI.1	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-2. English Language Arts, Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	1	1	TDA	OP	4	3	4.W.9	Writing
4	2	2	TE	OP	1	2	4.W.1.b	Writing
4	2	3	TE	OP	1	3	4.W.3.e	Writing
4	2	4	MC	OP	1	2	4.W.1.d	Writing
4	2	5	TE	OP	1	2	4.W.8	Writing
4	2	6	TE	OP	2	2	4.W.8	Writing
4	2	7	MC	OP	1	2	4.W.8	Writing
4	2	8	MC	OP	1	3	4.W.8	Writing
4	2	9	MC	OP	1	2	4.W.8	Writing
4	2	10	MC	OP	1	2	4.L.2.b	Writing
4	2	11	TE	OP	2	2	4.L.1.b	Writing
4	2	12	MC	OP	1	1	4.L.2.a	Writing
4	2	13	TE	OP	2	2	4.L.1.c	Writing
4	2	14	MC	OP	1	1	4.L.2.a	Writing
4	3	15	MC	OP	1	2	4.SL.3	Listening
4	3	16	MC	OP	1	1	4.SL.2	Listening
4	3	17	EBSR	OP	2	2	4.SL.3	Listening
4	3	18	MC	OP	1	2	4.SL.3	Listening
4	3	19	MC	OP	1	2	4.SL.3	Listening
4	3	20	EBSR	OP	2	3	4.SL.2	Listening
4	4	21	MC	OP	1	2	4.L.4	Reading
4	4	22	TE	OP	2	2	4.RI.2	Reading
4	4	23	MC	OP	1	2	4.RI.7	Reading
4	4	24	MC	OP	1	2	4.L.4	Reading
4	4	25	MS	OP	2	2	4.RI.1	Reading
4	4	26	EBSR	OP	2	3	4.RL.3	Reading
4	4	27	MC	OP	1	2	4.L.5	Reading
4	4	28	MC	OP	1	2	4.L.4	Reading
4	4	29	MC	OP	1	2	4.RL.5	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	4	30	MC	OP	1	2	4.RI.5	Reading
4	4	31	MC	OP	1	2	4.L.5.RI	Reading
4	4	32	TE	OP	2	2	4.RI.3	Reading
4	4	33	MC	OP	1	2	4.RI.7	Reading
4	4	34	MC	OP	1	2	4.RI.8	Reading
4	4	35	MC	OP	1	2	4.RL.2	Reading
4	4	36	MC	OP	1	2	4.RL.3	Reading
4	4	37	MC	OP	1	1	4.L.5.RL	Reading
4	4	38	TE	OP	2	2	4.RL.1	Reading
4	4	39	MC	OP	1	2	4.RL.6	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-3. English Language Arts, Grade 5 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	1	1	TDA	OP	4	3	5.W.2	Writing
5	2	2	MC	OP	1	1	5.L.2.b	Writing
5	2	3	MC	OP	1	2	5.W.3.e	Writing
5	2	4	TE	OP	2	2	5.W.8	Writing
5	2	5	MC	OP	1	1	5.W.2.d	Writing
5	2	6	TE	OP	2	2	5.W.8	Writing
5	2	7	MC	OP	1	2	5.W.1.d	Writing
5	2	8	TE	OP	1	2	5.W.8	Writing
5	2	9	MC	OP	1	2	5.W.8	Writing
5	2	10	MC	OP	1	1	5.L.2.c	Writing
5	2	11	TE	OP	2	2	5.L.1.b	Writing
5	2	12	MC	OP	1	2	5.L.3.a	Writing
5	2	13	TE	OP	2	1	5.L.2.b	Writing
5	3	14	TE	OP	2	3	5.SL.3	Listening
5	3	15	MC	OP	1	2	5.SL.2	Listening
5	3	16	MC	OP	1	2	5.SL.3	Listening
5	3	17	MC	OP	1	2	5.SL.3	Listening
5	3	18	MC	OP	1	2	5.SL.2	Listening
5	3	19	EBSR	OP	2	3	5.SL.3	Listening
5	4	20	MC	OP	1	2	5.RL.5	Reading
5	4	21	MC	OP	1	2	5.RL.4	Reading
5	4	22	MC	OP	1	2	5.RL.6	Reading
5	4	23	TE	OP	1	2	5.L.4	Reading
5	4	24	MC	OP	1	2	5.RI.4	Reading
5	4	25	TE	OP	1	2	5.RI.8	Reading
5	4	26	MC	OP	1	2	5.RI.5	Reading
5	4	27	MC	OP	1	2	5.RI.2	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	4	28	MC	OP	1	2	5.L.5	Reading
5	4	29	MS	OP	2	2	5.RL.2	Reading
5	4	30	MC	OP	1	2	5.RL.6	Reading
5	4	31	MC	OP	1	2	5.RL.9	Reading
5	4	32	MC	OP	1	2	5.RL.2	Reading
5	4	33	MC	OP	1	3	5.RL.6	Reading
5	4	34	MC	OP	1	2	5.RL.5	Reading
5	4	35	MC	OP	1	2	5.RL.2	Reading
5	4	36	TE	OP	2	2	5.RL.3	Reading
5	4	37	TE	OP	2	2	5.L.4.RI	Reading
5	4	38	MC	OP	1	2	5.RI.1	Reading
5	4	39	MC	OP	1	1	5.RI.3	Reading
5	4	40	MC	OP	1	2	5.RI.1	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-4. English Language Arts, Grade 6 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
6	1	1	TDA	OP	4	3	6.W.9	Writing
6	2	2	MC	OP	1	2	6.W.2.a	Writing
6	2	3	MC	OP	1	2	6.W.3.b	Writing
6	2	4	MC	OP	1	2	6.W.1	Writing
6	2	5	MC	OP	1	2	6.W.8	Writing
6	2	6	TE	OP	1	2	6.W.8	Writing
6	2	7	TE	OP	2	2	6.L.1.d	Writing
6	2	8	MC	OP	1	2	6.L.2.a	Writing
6	2	9	TE	OP	2	2	6.W.7	Writing
6	2	10	TE	OP	2	1	6.L.2.b	Writing
6	2	11	TE	OP	2	2	6.W.8	Writing
6	2	12	MC	OP	1	1	6.L.1.d	Writing
6	2	13	MC	OP	1	1	6.L.2.a	Writing
6	3	14	MC	OP	1	2	6.SL.2	Listening
6	3	15	MC	OP	1	2	6.SL.2	Listening
6	3	16	TE	OP	2	1	6.SL.2	Listening
6	3	17	EBSR	OP	2	3	6.SL.3	Listening
6	3	18	MC	OP	1	2	6.SL.2	Listening
6	3	19	MC	OP	1	2	6.SL.2	Listening
6	4	20	MC	OP	1	3	6.RL.5	Reading
6	4	21	TE	OP	2	2	6.RL.1	Reading
6	4	22	MC	OP	1	2	6.RL.4	Reading
6	4	23	TE	OP	2	2	6.RL.1	Reading
6	4	24	MC	OP	1	2	6.RI.8	Reading
6	4	25	MC	OP	1	2	6.L.4.RI	Reading
6	4	26	TE	OP	2	2	6.RI.1	Reading
6	4	27	MC	OP	1	3	6.RI.7	Reading
6	4	28	TE	OP	2	2	6.RI.2	Reading
6	4	29	TE	OP	1	2	6.L.4	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-5. English Language Arts, Grade 7 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
7	1	1	TDA	OP	4	3	7.W.9	Writing
7	2	2	MC	OP	1	3	7.W.2.a	Writing
7	2	3	TE	OP	2	2	7.W.3.c	Writing
7	2	4	TE	OP	2	2	7.W.7	Writing
7	2	5	MC	OP	1	2	7.L.2.a	Writing
7	2	6	TE	OP	2	2	7.L.3.a	Writing
7	2	7	MC	OP	1	2	7.W.8	Writing
7	2	8	MC	OP	1	2	7.L.2.a	Writing
7	2	9	MC	OP	1	2	7.W.8	Writing
7	2	10	TE	OP	2	3	7.W.8	Writing
7	2	11	MC	OP	1	3	7.L.1.b	Writing
7	2	12	MC	OP	1	2	7.L.1.a	Writing
7	2	13	MC	OP	1	2	7.W.8	Writing
7	3	14	EBSR	OP	2	3	7.SL.2	Listening
7	3	15	MS	OP	2	3	7.SL.3	Listening
7	3	16	MC	OP	1	2	7.SL.2	Listening
7	3	17	MC	OP	1	2	7.SL.2	Listening
7	3	18	EBSR	OP	2	3	7.SL.3	Listening
7	4	19	TE	OP	1	2	7.RL.4	Reading
7	4	20	MC	OP	1	2	7.RL.3	Reading
7	4	21	EBSR	OP	2	2	7.RL.6	Reading
7	4	22	MC	OP	1	3	7.RL.2	Reading
7	4	23	MC	OP	1	2	7.RI.1	Reading
7	4	24	TE	OP	1	2	7.RI.4	Reading
7	4	25	TE	OP	2	2	7.RI.1	Reading
7	4	26	TE	OP	2	2	7.RI.5	Reading
7	4	27	MC	OP	1	2	7.RI.5	Reading
7	4	28	EBSR	OP	2	3	7.RI.8	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
7	4	29	TE	OP	1	2	7.RL.1	Reading
7	4	30	MC	OP	1	2	7.RL.4	Reading
7	4	31	MC	OP	1	2	7.RL.2	Reading
7	4	32	MC	OP	1	3	7.RL.9	Reading
7	4	33	MC	OP	1	3	7.RI.6	Reading
7	4	34	MS	OP	2	2	7.RI.2	Reading
7	4	35	MS	OP	2	2	7.RI.4	Reading
7	4	36	MC	OP	1	3	7.RI.8	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table C-6. English Language Arts, Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	1	1	TDA	OP	4	3	8.W.2	Writing
8	2	2	MC	OP	1	2	8.W.3.d	Writing
8	2	3	MC	OP	1	2	8.W.2.e	Writing
8	2	4	MC	OP	1	2	8.W.3.d	Writing
8	2	5	MC	OP	1	2	8.W.8	Writing
8	2	6	MC	OP	1	2	8.W.7	Writing
8	2	7	TE	OP	2	2	8.W.8	Writing
8	2	8	MC	OP	1	2	8.W.7	Writing
8	2	9	MC	OP	1	2	8.L.2.c	Writing
8	2	10	TE	OP	2	2	8.L.1.d	Writing
8	2	11	MC	OP	1	2	8.L.2.a	Writing
8	2	12	TE	OP	2	2	8.L.1.c	Writing
8	2	13	MC	OP	1	2	8.W.8	Writing
8	2	14	MC	OP	1	2	8.W.8	Writing
8	3	15	MS	OP	2	2	8.SL.3	Listening
8	3	16	EBSR	OP	2	3	8.SL.3	Listening
8	3	17	MC	OP	1	2	8.SL.3	Listening
8	3	18	MC	OP	1	2	8.SL.2	Listening
8	3	19	EBSR	OP	2	3	8.SL.3	Listening
8	4	20	MC	OP	1	2	8.RI.8	Reading
8	4	21	MC	OP	1	2	8.L.4	Reading
8	4	22	MC	OP	1	2	8.RI.6	Reading
8	4	23	MC	OP	1	2	8.RI.5	Reading
8	4	24	MS	OP	2	2	8.RI.3	Reading
8	4	25	MC	OP	1	2	8.RI.2	Reading
8	4	26	MC	OP	1	2	8.RL.4	Reading
8	4	27	MC	OP	1	3	8.RL.3	Reading
8	4	28	MC	OP	1	2	8.RL.3	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	4	29	MC	OP	1	3	8.RL.6	Reading
8	4	30	EBSR	OP	2	3	8.RL.2	Reading
8	4	31	MC	OP	1	3	8.RI.5	Reading
8	4	32	EBSR	OP	2	2	8.L.4.RI	Reading
8	4	33	MC	OP	1	2	8.RI.5	Reading
8	4	34	MC	OP	1	2	8.RI.3	Reading
8	4	35	MC	OP	1	2	8.RI.6	Reading
8	4	36	MC	OP	1	2	8.RL.4	Reading
8	4	37	MC	OP	1	2	8.RL.1	Reading
8	4	38	MS	OP	2	2	8.RL.1	Reading
8	4	39	MC	OP	1	2	8.RL.6	Reading

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix D

Spring 2019 Mathematics Operational Test Maps

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-1 Mathematics Grade 3 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
3	1	1	TE	OP	1	2	3.OA.1	OA
3	1	2	SA	OP	1	1	3.NF.1	NF
3	1	3	TE	OP	1	2	3.NBT.1	NBT
3	1	4	MC	OP	1	2	3.MD.1	MD
3	1	5	MC	OP	1	3	3.G.1	G
3	1	6	SA	OP	1	1	3.OA.4	OA
3	1	7	MC	OP	1	2	3.MD.5.b	MD
3	1	8	SA	OP	1	1	3.G.2	G
3	1	9	MC	OP	1	1	3.NF.2	NF
3	1	10	SA	OP	1	1	3.NBT.1	NBT
3	1	11	MC	OP	1	2	3.OA.6	OA
3	1	12	TE	OP	1	1	3.NF.2.b	NF
3	1	13	MC	OP	1	1	3.NBT.2	NBT
3	1	14	SA	OP	1	2	3.MD.7.a	MD
3	1	15	SA	OP	1	1	3.NF.3.b	NF
3	1	16	MC	OP	1	3	3.OA.8	OA
3	1	17	MC	OP	1	2	3.G.2	G
3	1	18	TE	OP	1	3	3.NF.3	NF
3	1	19	MC	OP	1	2	3.NBT.3	NBT
3	1	20	MC	OP	1	2	3.G.2	G
3	1	21	SA	OP	1	2	3.MD.2	MD
3	2	22	MC	OP	1	2	3.MD.8	MD
3	2	23	MC	OP	1	1	3.NBT.1	NBT
3	2	24	SA	OP	1	2	3.OA.7	OA
3	2	25	TE	OP	1	2	3.G.1	G
3	2	26	TE	OP	1	2	3.MD.3	MD
3	2	27	MC	OP	1	2	3.OA.2	OA
3	2	28	MC	OP	1	3	3.NBT.2	NBT
3	2	29	MC	OP	1	1	3.G.1	G
3	2	30	MC	OP	1	1	3.MD.2	MD
3	2	31	MC	OP	1	2	3.NF.3.a	NF
3	2	32	MC	OP	1	1	3.OA.5	OA
3	2	33	SA	OP	1	1	3.MD.7.b	MD
3	2	34	MC	OP	1	1	3.NF.2.b	NF
3	2	35	SA	OP	1	2	3.NBT.3	NBT

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-1 Mathematics Grade 3 Test Map

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

Domain Names: OA= Operations and Algebraic Thinking; NBT= Number and Operations in Base Ten; NF= Number and Operations – Fractions; MD= Measurement and Data; G=Geometry

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-2 Mathematics Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	1	1	MC	OP	1	3	4.NBT.2	NBT
4	1	2	MC	OP	1	1	4.OA.1	OA
4	1	3	SA	OP	1	1	4.G.3	G
4	1	4	MC	OP	1	2	4.NF.3.d	NF
4	1	5	TE	OP	1	2	4.NBT.5	NBT
4	1	6	MC	OP	1	2	4.G.1	G
4	1	7	MC	OP	1	2	4.NBT.1	NBT
4	1	8	SA	OP	1	2	4.MD.7	MD
4	1	9	MC	OP	1	2	4.NBT.3	NBT
4	1	10	MC	OP	1	2	4.OA.2	OA
4	1	11	MC	OP	1	3	4.MD.3	MD
4	1	12	MC	OP	1	2	4.G.2	G
4	1	13	MC	OP	1	2	4.NF.3	NF
4	1	14	MC	OP	1	2	4.MD.4	MD
4	1	15	MC	OP	1	3	4.OA.4	OA
4	1	16	SA	OP	1	1	4.NF.6	NF
4	1	17	MC	OP	1	1	4.MD.5.b	MD
4	1	18	MC	OP	1	2	4.NF.4	NF
4	1	19	TE	OP	1	2	4.OA.3	OA
4	1	20	MC	OP	1	2	4.NBT.4	NBT
4	1	21	MC	OP	1	1	4.MD.1	MD
4	1	22	MC	OP	1	3	4.OA.5	OA
4	1	23	SA	OP	1	2	4.NF.5	NF
4	2	24	TE	OP	1	1	4.NBT.2	NBT
4	2	25	MC	OP	1	2	4.OA.1	OA
4	2	26	MC	OP	1	2	4.MD.2	MD
4	2	27	MC	OP	1	1	4.NF.1	NF
4	2	28	MC	OP	1	2	4.G.1	G
4	2	29	SA	OP	1	2	4.NBT.5	NBT
4	2	30	MC	OP	1	2	4.MD.3	MD
4	2	31	MC	OP	1	2	4.OA.3	OA
4	2	32	TE	OP	1	2	4.G.3	G
4	2	33	MC	OP	1	1	4.NF.2	NF
4	2	34	MC	OP	1	2	4.OA.2	OA

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-2 Mathematics Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	2	35	SA	OP	1	1	4.MD.6	MD
4	2	36	MC	OP	1	2	4.NF.7	NF
4	2	37	MC	OP	1	1	4.G.3	G
4	2	38	MC	OP	1	2	4.NBT.6	NBT
4	2	39	TE	OP	1	1	4.NF.3.b	NF
4	2	40	SA	OP	1	1	4.G.2	G
4	2	41	MC	OP	1	2	4.MD.7	MD
4	2	42	MC	OP	1	2	4.OA.5	OA
4	2	43	SA	OP	1	1	4.NBT.4	NBT
4	2	44	MC	OP	1	2	4.NF.4.c	NF
4	2	45	SA	OP	1	1	4.OA.4	OA
4	2	46	MC	OP	1	1	4.MD.6	MD

Domain Names: OA= Operations and Algebraic Thinking; NBT= Number and Operations in Base Ten; NF= Number and Operations – Fractions; MD= Measurement and Data; G=Geometry

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-3 Mathematics Grade 5 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	1	1	SA	OP	1	2	5.OA.2	OA
5	1	2	MC	OP	1	2	5.NF.1	NF
5	1	3	MC	OP	1	2	5.MD.1	MD
5	1	4	MC	OP	1	2	5.G.1	G
5	1	5	MC	OP	1	3	5.NF.5	NF
5	1	6	TE	OP	1	2	5.OA.1	OA
5	1	7	MC	OP	1	1	5.NBT.3.a	NBT
5	1	8	MC	OP	1	2	5.MD.2	MD
5	1	9	TE	OP	1	2	5.G.2	G
5	1	10	SA	OP	1	1	5.MD.3.b	MD
5	1	11	MC	OP	1	3	5.OA.3	OA
5	1	12	SA	OP	1	1	5.NBT.1	NBT
5	1	13	MC	OP	1	2	5.MD.3	MD
5	1	14	SA	OP	1	2	5.NF.7.c	NF
5	1	15	MC	OP	1	2	5.OA.1	OA
5	1	16	SA	OP	1	2	5.G.4	G
5	1	17	MC	OP	1	1	5.NF.6	NF
5	1	18	SA	OP	1	1	5.NBT.4	NBT
5	1	19	MC	OP	1	2	5.MD.5.a	MD
5	1	20	MC	OP	1	3	5.NBT.6	NBT
5	1	21	SA	OP	1	2	5.OA.1	OA
5	1	22	TE	OP	1	2	5.NBT.5	NBT
5	1	23	MC	OP	1	2	5.G.1	G
5	2	24	MC	OP	1	1	5.NBT.2	NBT
5	2	25	SA	OP	1	3	5.OA.1	OA
5	2	26	MC	OP	1	1	5.G.3	G
5	2	27	SA	OP	1	1	5.MD.4	MD
5	2	28	MC	OP	1	2	5.NF.7.a	NF
5	2	29	TE	OP	1	2	5.OA.1	OA
5	2	30	MC	OP	1	1	5.G.4	G
5	2	31	TE	OP	1	2	5.NBT.3.b	NBT
5	2	32	MC	OP	1	3	5.NF.4.a	NF
5	2	33	MC	OP	1	2	5.OA.3	OA
5	2	34	MC	OP	1	2	5.G.2	G

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-3 Mathematics Grade 5 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
5	2	35	TE	OP	1	2	5.NF.4	NF
5	2	36	MC	OP	1	2	5.MD.5	MD
5	2	37	SA	OP	1	1	5.G.1	G
5	2	38	MC	OP	1	2	5.MD.3	MD
5	2	39	MC	OP	1	2	5.NBT.7	NBT
5	2	40	TE	OP	1	2	5.MD.5.b	MD
5	2	41	MC	OP	1	2	5.NF.6	NF
5	2	42	MC	OP	1	2	5.OA.1	OA
5	2	43	MC	OP	1	3	5.NF.2	NF
5	2	44	SA	OP	1	2	5.G.2	G
5	2	45	MC	OP	1	2	5.MD.1	MD
5	2	46	SA	OP	1	2	5.NBT.7	NBT

Domain Names: OA= Operations and Algebraic Thinking; NBT= Number and Operations in Base Ten; NF= Number and Operations – Fractions; MD= Measurement and Data; G=Geometry

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-4 Mathematics Grade 6 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
6	1	1	MC	OP	1	2	6.RP.3.d	RP
6	1	2	MC	OP	1	1	6.EE.2.a	EE
6	1	3	TE	OP	1	3	6.RP.3	RP
6	1	4	MC	OP	1	2	6.NS.3	NS
6	1	5	SA	OP	1	1	6.EE.2.a	EE
6	1	6	SA	OP	1	1	6.NS.2	NS
6	1	7	MC	OP	1	2	6.EE.3	EE
6	1	8	MC	OP	1	2	6.NS.2	NS
6	1	9	SA	OP	1	2	6.RP.1	RP
6	1	10	MC	OP	1	2	6.EE.2.b	EE
6	1	11	TE	OP	1	2	6.RP.3.a	RP
6	1	12	MC	OP	1	2	6.RP.2	RP
6	1	13	SA	OP	1	1	6.EE.1	EE
6	1	14	SA	OP	1	1	6.RP.2	RP
6	1	15	MC	OP	1	1	6.NS.4	NS
6	1	16	MC	OP	1	2	6.RP.3.a	RP
6	2	17	MC	OP	1	2	6.G.1	G
6	2	18	MC	OP	1	2	6.EE.7	EE
6	2	19	MC	OP	1	1	6.NS.5	NS
6	2	20	SA	OP	1	2	6.SP.5.a	SP
6	2	21	TE	OP	1	2	6.EE.7	EE
6	2	22	MC	OP	1	2	6.NS.6	NS
6	2	23	TE	OP	1	1	6.SP.4	SP
6	2	24	MC	OP	1	2	6.G.3	G
6	2	25	MC	OP	1	2	6.EE.8	EE
6	2	26	MC	OP	1	1	6.SP.1	SP
6	2	27	SA	OP	1	2	6.NS.6.b	NS
6	2	28	MC	OP	1	2	6.G.1	G
6	2	29	MC	OP	1	2	6.EE.9	EE
6	2	30	MC	OP	1	2	6.SP.5.b	SP
6	2	31	MC	OP	1	2	6.SP.5.c	SP
6	2	32	MC	OP	1	2	6.SP.3	SP
6	2	33	MC	OP	1	2	6.NS.8	NS
6	2	34	MC	OP	1	2	6.EE.6	EE

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-4 Mathematics Grade 6 Test Map

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-5 Mathematics Grade 7 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
7	1	1	SA	OP	1	2	7.NS.3	NS
7	1	2	MC	OP	1	2	7.NS.3	NS
7	1	3	MC	OP	1	2	7.EE.2	EE
7	1	4	SA	OP	1	1	7.NS.2.c	NS
7	1	5	MC	OP	1	2	7.EE.1	EE
7	1	6	MC	OP	1	2	7.NS.2	NS
7	1	7	MC	OP	1	2	7.NS.1	NS
7	1	8	TE	OP	1	1	7.NS.2	NS
7	1	9	MC	OP	1	2	7.EE.1	EE
7	1	10	MC	OP	1	2	7.NS.3	NS
7	1	11	MC	OP	1	2	7.EE.2	EE
7	2	12	MC	OP	1	2	7.G.3	G
7	2	13	MC	OP	1	1	7.RP.1	RP
7	2	14	MC	OP	1	2	7.EE.3	EE
7	2	15	SA	OP	1	2	7.SP.8.c	SP
7	2	16	MC	OP	1	2	7.EE.4	EE
7	2	17	MC	OP	1	2	7.G.5	G
7	2	18	MC	OP	1	2	7.RP.3	RP
7	2	19	TE	OP	1	2	7.RP.2.d	RP
7	2	20	MC	OP	1	2	7.G.4	G
7	2	21	MC	OP	1	2	7.SP.1	SP
7	2	22	MC	OP	1	1	7.G.2	G
7	2	23	MC	OP	1	2	7.G.4	G
7	2	24	SA	OP	1	1	7.SP.7.a	SP
7	2	25	SA	OP	1	2	7.G.1	G
7	2	26	MC	OP	1	2	7.SP.7.b	SP
7	2	27	MC	OP	1	2	7.SP.1	SP
7	2	28	MC	OP	1	2	7.RP.2	RP
7	2	29	MC	OP	1	2	7.SP.5	SP
7	2	30	MC	OP	1	3	7.RP.2.a	RP
7	2	31	MC	OP	1	2	7.RP.1	RP
7	2	32	TE	OP	1	2	7.G.2	G
7	2	33	TE	OP	1	2	7.EE.3	EE
7	2	34	MC	OP	1	2	7.G.6	G

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-5 Mathematics Grade 7 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
7	2	35	SA	OP	1	2	7.EE.4.a	EE
7	2	36	TE	OP	1	1	7.SP.8.b	SP
7	2	37	MC	OP	1	2	7.RP.3	RP
7	2	38	TE	OP	1	3	7.EE.4.b	EE
7	2	39	MC	OP	1	2	7.SP.3	SP
7	2	40	MC	OP	1	2	7.G.4	G
7	2	41	TE	OP	1	2	7.SP.6	SP
7	2	42	TE	OP	1	2	7.G.1	G
7	2	43	MC	OP	1	2	7.SP.2	SP
7	2	44	MC	OP	1	2	7.EE.4.b	EE
7	2	45	MC	OP	1	2	7.SP.2	SP
7	2	46	SA	OP	1	1	7.RP.1	RP

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-6 Mathematics Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	1	1	MC	OP	1	1	8.NS.2	NS
8	1	2	MC	OP	1	2	8.EE.1	EE
8	1	3	MC	OP	1	1	8.EE.2	EE
8	1	4	MC	OP	1	1	8.NS.1	NS
8	1	5	SA	OP	1	1	8.EE.1	EE
8	1	6	MC	OP	1	3	8.NS.1	NS
8	1	7	MC	OP	1	2	8.NS.2	NS
8	1	8	MC	OP	1	1	8.EE.2	EE
8	1	9	TE	OP	1	1	8.NS.1	NS
8	1	10	MC	OP	1	2	8.NS.1	NS
8	1	11	TE	OP	1	2	8.NS.2	NS
8	1	12	MC	OP	1	2	8.EE.3	EE
8	1	13	SA	OP	1	1	8.NS.2	NS
8	2	14	MC	OP	1	1	8.F.1	F
8	2	15	TE	OP	1	3	8.G.5	G
8	2	16	MC	OP	1	1	8.SP.1	SP
8	2	17	SA	OP	1	2	8.EE.7.a	EE
8	2	18	MC	OP	1	2	8.F.4	F
8	2	19	MC	OP	1	2	8.G.7	G
8	2	20	MC	OP	1	3	8.F.2	F
8	2	21	SA	OP	1	2	8.G.7	G
8	2	22	MC	OP	1	1	8.SP.2	SP
8	2	23	TE	OP	1	2	8.EE.5	EE
8	2	24	MC	OP	1	2	8.G.3	G
8	2	25	TE	OP	1	2	8.F.2	F
8	2	26	MC	OP	1	2	8.G.8	G
8	2	27	SA	OP	1	2	8.SP.4	SP
8	2	28	MC	OP	1	2	8.F.4	F
8	2	29	MC	OP	1	2	8.G.2	G
8	2	30	MC	OP	1	2	8.SP.3	SP
8	2	31	MC	OP	1	2	8.EE.8.a	EE
8	2	32	TE	OP	1	2	8.F.3	F
8	2	33	MC	OP	1	2	8.G.4	G
8	2	34	MC	OP	1	2	8.SP.4	SP

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table D-6 Mathematics Grade 8 Test Map

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix E

Spring 2019 Science Operational Test Maps

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-1 Science Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
4	1	1	MC	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	1	2	TE	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	
4	1	3	TE	OP	1	3	SCI.ESS3.B.4	Earth and Space Science		SCI.CC2.3-5
4	1	4	TE	OP	1	3	SCI.ESS1.C.4	Earth and Space Science	SCI.SEP6.A.3-5	
4	1	5	TE	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	1	6	TE	OP	1	2	SCI.ETS1.C.3-5	Engineering	SCI.SEP3.A.3-5	
4	1	7	MC	OP	1	2	SCI.PS3.B.4	Physical Science	SCI.SEP1.A.3-5	SCI.CC5.3-5
4	1	8	TE	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	1	9	EBSR	OP	1	3	SCI.ETS1.C.3-5	Engineering	SCI.SEP3.A.3-5	
4	1	10	MC	OP	1	2	SCI.ESS2.A.4	Earth and Space Science	SCI.SEP3.A.3-5	SCI.CC2.3-5
4	1	11	MC	OP	1	2	SCI.ESS2.B.4	Earth and Space Science	SCI.SEP4.A.3-5	SCI.CC1.3-5
4	1	12	MC	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	SCI.CC4.3-5
4	1	13	TE	OP	1	2	SCI.LS1.D.4	Life Science		SCI.CC4.3-5
4	1	14	TE	OP	1	2	SCI.LS1.A.4	Life Science	SCI.SEP4.A.3-5	
4	1	15	TE	OP	1	2	SCI.LS1.D.4	Life Science	SCI.SEP2.A.3-5	SCI.CC4.3-5
4	2	16	MC	OP	1	3	SCI.PS3.B.4	Physical Science	SCI.SEP3.A.3-5	SCI.CC5.3-5
4	2	17	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-1 Science Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
4	2	18	TE	OP	1	2	SCI.PS4.B.4	Physical Science	SCI.SEP2.A.3-5	
4	2	19	TE	OP	1	3	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	
4	2	20	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5
4	2	21	TE	OP	1	2	SCI.PS4.B.4	Physical Science	SCI.SEP2.A.3-5	SCI.CC2.3-5
4	2	22	TE	OP	1	2	SCI.ESS3.A.4	Earth and Space Science		SCI.CC2.3-5
4	2	23	MC	OP	1	2	SCI.PS3.B.4	Physical Science		SCI.CC5.3-5
4	2	24	TE	OP	1	3	SCI.PS3.C.4	Physical Science		SCI.CC5.3-5
4	2	25	TE	OP	1	2	SCI.ETS1.A.3-5	Engineering	SCI.SEP6.A.3-5	
4	3	26	TE	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5
4	3	27	TE	OP	1	3	SCI.ESS3.B.4	Earth and Space Science	SCI.SEP6.B.3-5	SCI.CC2.3-5
4	3	28	EBSR	OP	1	2	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	SCI.CC5.3-5
4	3	29	EBSR	OP	1	3	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	3	30	TE	OP	1	3	SCI.ETS1.B.3-5	Engineering	SCI.SEP1.B.3-5	
4	3	31	TE	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	3	32	MC	OP	1	2	SCI.ETS1.B.3-5	Engineering	SCI.SEP6.B.3-5	
4	3	33	TE	OP	1	3	SCI.PS3.A.4	Physical Science	SCI.SEP6.A.3-5	

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-1 Science Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
4	3	34	TE	OP	1	2	SCI.LS1.A.4	Life Science		SCI.CC4.3-5
4	3	35	MC	OP	1	2	SCI.ESS2.B.4	Earth and Space Science	SCI.SEP4.A.3-5	SCI.CC1.3-5
4	3	36	EBSR	OP	1	2	SCI.ESS1.C.4	Earth and Space Science	SCI.SEP6.A.3-5	
4	3	37	MC	OP	1	3	SCI.PS4.A.4	Physical Science	SCI.SEP2.A.3-5	
4	3	38	TE	OP	1	3	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	SCI.CC4.3-5
4	3	39	MS	OP	1	3	SCI.LS1.A.4	Life Science	SCI.SEP7.A.3-5	
4	3	40	TE	OP	1	3	SCI.LS1.D.4	Life Science	SCI.SEP2.A.3-5	SCI.CC4.3-5

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-2 Science Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
8	1	1	TE	OP	1	2	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	SCI.CC2.m
8	1	2	TE	OP	1	3	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	SCI.CC2.m
8	1	3	TE	OP	1	3	SCI.LS2.A.m	Life Science		SCI.CC1.m
8	1	4	TE	OP	1	3	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.m	
8	1	5	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP2.A.m	
8	1	6	MC	OP	1	3	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.m	
8	1	7	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP2.A.m	SCI.CC1.m
8	1	8	MC	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP2.A.m	SCI.CC4.m
8	1	9	TE	OP	1	2	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	1	10	TE	OP	1	2	SCI.LS1.B.m	Life Science	SCI.SEP2.A.m	SCI.CC2.m
8	1	11	TE	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	1	12	MC	OP	1	3	SCI.LS1.B.m	Life Science	SCI.SEP7.A.m	SCI.CC2.m
8	1	13	TE	OP	1	2	SCI.PS1.A.m	Physical Science	SCI.SEP2.A.m	SCI.CC3.m
8	1	14	TE	OP	1	2	SCI.PS3.B.m	Physical Science	SCI.SEP7.A.m	SCI.CC5.m
8	1	15	EBSR	OP	1	2	SCI.PS4.A.m	Physical Science	SCI.SEP5.A.m	SCI.CC1.m
8	2	16	TE	OP	1	2	SCI.LS1.C.m	Life Science	SCI.SEP2.A.m	
8	2	17	TE	OP	1	2	SCI.LS4.B.m	Life Science		SCI.CC2.m
8	2	18	TE	OP	1	2	SCI.LS4.B.m	Life Science		SCI.CC2.m

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-2 Science Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
8	2	19	TE	OP	1	3	SCI.ESS1.C.m	Earth and Space Science		SCI.CC3.m
8	2	20	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP6.A.m	
8	2	21	TE	OP	1	2	SCI.ETS1.B.m	Engineering	SCI.SEP7.A.m	
8	2	22	TE	OP	1	2	SCI.ESS1.C.m	Earth and Space Science	SCI.SEP6.A.m	SCI.CC3.m
8	2	23	TE	OP	1	3	SCI.ESS3.C.m	Earth and Space Science	SCI.SEP7.A.m	SCI.CC2.m
8	2	24	TE	OP	1	2	SCI.ESS2.A.m	Earth and Space Science	SCI.SEP6.A.m	SCI.CC3.m
8	2	25	TE	OP	1	2	SCI.ETS1.C.m	Engineering	SCI.SEP2.A.m	
8	3	26	EBSR	OP	1	2	SCI.LS1.C.m	Life Science	SCI.SEP6.A.m	SCI.CC5.m
8	3	27	TE	OP	1	2	SCI.LS2.B.m	Life Science	SCI.SEP2.A.m	SCI.CC5.m
8	3	28	TE	OP	1	2	SCI.LS2.A.m	Life Science	SCI.SEP4.A.m	
8	3	29	TE	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	3	30	MC	OP	1	3	SCI.PS1.B.m	Physical Science	SCI.SEP6.B.m	
8	3	31	EBSR	OP	1	3	SCI.ETS1.A.m	Engineering	SCI.SEP1.A.m	
8	3	32	TE	OP	1	2	SCI.ETS1.B.m	Engineering	SCI.SEP4.A.m	
8	3	33	TE	OP	1	2	SCI.PS1.B.m	Physical Science	SCI.SEP2.A.m	SCI.CC5.m
8	3	34	TE	OP	1	2	SCI.PS2.B.m	Physical Science	SCI.SEP6.A.m	SCI.CC2.m
8	3	35	TE	OP	1	3	SCI.PS2.B.m	Physical Science	SCI.SEP6.A.m	SCI.CC2.m
8	3	36	MS	OP	1	2	SCI.PS2.B.m	Physical Science	SCI.SEP1.A.m	SCI.CC2.m

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table E-2 Science Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain	Science and Engineering Practices Standard	Crosscutting Concepts Standard
8	3	37	TE	OP	1	2	SCI.PS1.A.m	Physical Science	SCI.SEP2.A.m	SCI.CC3.m
8	3	38	MC	OP	1	3	SCI.PS2.B.m	Physical Science	SCI.SEP7.A.m	SCI.CC4.m
8	3	39	TE	OP	1	2	SCI.ESS1.B.m	Earth and Space Science	SCI.SEP4.A.m	SCI.CC3.m
8	3	40	MC	OP	1	2	SCI.ESS1.B.m	Earth and Space Science		SCI.CC4.m

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix F

Spring 2019 Social Studies Operational Test Maps

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-1 Social Studies Grade 4 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
4	1	1	MC	OP	1	2	E.4.8	Behavioral Sciences
4	1	2	MC	OP	1	1	B.4.6	History
4	1	3	MC	OP	1	2	E.4.15	Behavioral Sciences
4	1	4	MC	OP	1	2	B.4.2	History
4	1	5	MC	OP	1	3	D.4.3	Economics
4	1	6	MC	OP	1	2	A.4.7	Geography
4	1	7	TE	OP	1	2	B.4.8	History
4	1	8	MC	OP	1	2	A.4.1	Geography
4	1	9	MC	OP	1	2	A.4.1	Geography
4	1	10	MC	OP	1	2	E.4.15	Behavioral Sciences
4	1	11	MC	OP	1	2	C.4.6	Civics
4	1	12	MC	OP	1	2	B.4.7	History
4	1	13	TE	OP	1	2	D.4.6	Economics
4	1	14	MC	OP	1	3	D.4.3	Economics
4	1	15	MC	OP	1	2	E.4.6	Behavioral Sciences
4	1	16	MC	OP	1	1	C.4.2	Civics
4	1	17	MC	OP	1	3	B.4.9	History
4	1	18	TE	OP	1	1	E.4.5	Behavioral Sciences
4	1	19	MC	OP	1	3	E.4.11	Behavioral Sciences
4	2	20	MC	OP	1	2	A.4.7	Geography
4	2	21	MC	OP	1	2	C.4.1	Civics
4	2	22	MC	OP	1	1	A.4.2	Geography
4	2	23	MC	OP	1	2	E.4.9	Behavioral Sciences
4	2	24	MC	OP	1	2	C.4.3	Civics
4	2	25	MC	OP	1	2	B.4.8	History
4	2	26	MC	OP	1	2	C.4.4	Civics
4	2	27	MC	OP	1	3	B.4.4	History
4	2	28	MC	OP	1	2	A.4.9	Geography
4	2	29	MC	OP	1	2	D.4.2	Economics
4	2	30	MC	OP	1	2	D.4.7	Economics
4	2	31	TE	OP	1	1	A.4.2	Geography
4	2	32	MC	OP	1	2	A.4.6	Geography

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-1 Social Studies Grade 4 Test Map

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-2 Social Studies Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	1	1	MC	OP	1	2	A.8.8	Geography
8	1	2	MC	OP	1	3	B.8.5	History
8	1	3	MC	OP	1	3	B.8.9	History
8	1	4	MC	OP	1	3	B.8.4	History
8	1	5	MC	OP	1	2	B.8.2	History
8	1	6	MC	OP	1	2	B.8.7	History
8	1	7	MS	OP	1	2	B.8.7	History
8	1	8	MC	OP	1	3	C.8.9	Civics
8	1	9	MC	OP	1	2	A.8.9	Geography
8	1	10	TE	OP	1	2	D.8.2	Economics
8	1	11	MC	OP	1	2	A.8.5	Geography
8	1	12	MC	OP	1	2	A.8.11	Geography
8	1	13	MC	OP	1	2	D.8.8	Economics
8	1	14	MC	OP	1	3	E.8.14	Behavioral Sciences
8	1	15	MC	OP	1	2	A.8.7	Geography
8	1	16	MC	OP	1	3	D.8.7	Economics
8	1	17	MC	OP	1	3	E.8.9	Behavioral Sciences
8	1	18	MC	OP	1	2	B.8.3	History
8	1	19	MC	OP	1	2	C.8.1	Civics
8	1	20	MC	OP	1	2	B.8.7	History
8	2	21	MC	OP	1	2	E.8.10	Behavioral Sciences
8	2	22	MC	OP	1	2	A.8.11	Geography
8	2	23	TE	OP	1	2	A.8.10	Geography
8	2	24	MC	OP	1	2	A.8.5	Geography
8	2	25	MC	OP	1	3	A.8.5	Geography
8	2	26	MC	OP	1	2	C.8.6	Civics
8	2	27	MC	OP	1	3	D.8.3	Economics
8	2	28	MC	OP	1	2	E.8.11	Behavioral Sciences
8	2	29	MC	OP	1	3	C.8.8	Civics
8	2	30	TE	OP	1	2	D.8.5	Economics
8	2	31	MC	OP	1	3	B.8.9	History
8	2	32	TE	OP	1	2	C.8.1	Civics
8	2	33	MC	OP	1	2	C.8.6	Civics

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-2 Social Studies Grade 8 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
8	2	34	MC	OP	1	2	B.8.5	History
8	2	35	MC	OP	1	3	B.8.3	History
8	2	36	MC	OP	1	2	D.8.7	Economics
8	2	37	MC	OP	1	3	B.8.10	History
8	2	38	MC	OP	1	2	E.8.9	Behavioral Sciences
8	2	39	MC	OP	1	3	A.8.9	Geography
8	2	40	MC	OP	1	3	E.8.4	Behavioral Sciences

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-3 Social Studies Grade 10 Test Map

Grade	Session	Item Sequence	Item Type	Item Usage	Max Score Points	Depth of Knowledge	Standard	Domain
10	1	1	MC	OP	1	2	A.10.7	Geography
10	1	2	MC	OP	1	3	A.10.5	Geography
10	1	3	MC	OP	1	2	B.10.8	History
10	1	4	MC	OP	1	2	E.10.12	Behavioral Sciences
10	1	5	MC	OP	1	2	B.10.5	History
10	1	6	MC	OP	1	2	B.10.9	History
10	1	7	MC	OP	1	2	B.10.7	History
10	1	8	MC	OP	1	2	E.10.14	Behavioral Sciences
10	1	9	MC	OP	1	2	D.10.10	Economics
10	1	10	MC	OP	1	2	C.10.6	Civics
10	1	11	MC	OP	1	2	B.10.16	History
10	1	12	TE	OP	1	2	E.10.17	Behavioral Sciences
10	1	13	MC	OP	1	3	D.10.7	Economics
10	1	14	MC	OP	1	2	E.10.5	Behavioral Sciences
10	1	15	MC	OP	1	2	A.10.6	Geography
10	1	16	MC	OP	1	2	A.10.6	Geography
10	1	17	MC	OP	1	2	A.10.4	Geography
10	1	18	MC	OP	1	2	A.10.1	Geography
10	1	19	MC	OP	1	3	B.10.9	History
10	1	20	MC	OP	1	3	D.10.2	Economics
10	1	21	MC	OP	1	2	B.10.13	History
10	1	22	MC	OP	1	3	C.10.4	Civics
10	1	23	MC	OP	1	2	D.10.14	Economics
10	1	24	TE	OP	1	2	C.10.13	Civics
10	1	25	MC	OP	1	2	B.10.6	History
10	2	26	MC	OP	1	2	E.10.17	Behavioral Sciences
10	2	27	MC	OP	1	2	E.10.6	Behavioral Sciences
10	2	28	MC	OP	1	2	B.10.10	History
10	2	29	MC	OP	1	3	C.10.3	Civics
10	2	30	MC	OP	1	2	C.10.13	Civics
10	2	31	MC	OP	1	2	D.10.3	Economics
10	2	32	MC	OP	1	2	C.10.15	Civics
10	2	33	MC	OP	1	2	E.10.8	Behavioral Sciences

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table F-3 Social Studies Grade 10 Test Map

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix G

Classical Item Analysis Results

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-1. Item Statistics, ELA Grade 3

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	61003	0.37	0.60	0.00	0.16	0.33	0.39	0.11	0.01	-0.44	-0.21	0.32	0.31	0.12
2	MC	1	60975	0.67	0.31	0.00		0.11	0.17	0.67	0.04		-0.18	-0.14	0.31	-0.18
3	TE	2	60726	0.71	0.39	0.00	0.17	0.25	0.58			-0.23	-0.28	0.43		
4	MC	1	60911	0.58	0.40	0.00		0.58	0.17	0.16	0.09		0.40	-0.23	-0.16	-0.18
5	MC	1	60909	0.41	0.27	0.00		0.22	0.41	0.06	0.31		-0.11	0.27	-0.17	-0.09
6	TE	2	60849	0.56	0.43	0.00	0.22	0.45	0.33			-0.29	-0.13	0.41		
7	MC	1	60899	0.62	0.42	0.00		0.15	0.13	0.11	0.61		-0.23	-0.15	-0.23	0.42
8	MC	1	60898	0.71	0.44	0.00		0.16	0.04	0.09	0.71		-0.25	-0.21	-0.22	0.44
9	MC	1	60921	0.61	0.39	0.00		0.21	0.61	0.10	0.08		-0.23	0.39	-0.19	-0.14
10	TE	1	60654	0.35	0.37	0.01	0.65	0.35				-0.36	0.38			
11	MC	1	60894	0.60	0.46	0.00		0.60	0.11	0.17	0.11		0.47	-0.22	-0.20	-0.25
12	MC	1	60908	0.45	0.36	0.00		0.18	0.22	0.45	0.15		-0.12	-0.19	0.36	-0.15
13	MC	1	60906	0.71	0.54	0.00		0.10	0.12	0.07	0.71		-0.28	-0.29	-0.25	0.54
14	TE	2	60889	0.65	0.53	0.00	0.16	0.38	0.46			-0.41	-0.17	0.47		
15	MC	1	60956	0.71	0.36	0.00		0.11	0.71	0.10	0.08		-0.16	0.36	-0.19	-0.20
16	TE	2	60906	0.64	0.52	0.00	0.13	0.45	0.42			-0.37	-0.20	0.46		
17	MC	1	60918	0.55	0.35	0.00		0.28	0.08	0.09	0.55		-0.14	-0.22	-0.17	0.35
18	MS	2	60892	0.72	0.53	0.00	0.09	0.37	0.54			-0.34	-0.30	0.50		
19	MC	1	60887	0.44	0.32	0.00		0.12	0.44	0.30	0.14		-0.18	0.32	-0.22	0.00
20	MC	1	60785	0.54	0.21	0.00		0.24	0.03	0.20	0.53		-0.07	-0.18	-0.10	0.21

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	60879	0.66	0.39	0.00		0.22	0.65	0.08	0.05		-0.24	0.39	-0.13	-0.22
22	MC	1	60864	0.44	0.27	0.00		0.21	0.44	0.14	0.21		-0.06	0.28	-0.16	-0.14
23	MC	1	60845	0.52	0.29	0.00		0.52	0.09	0.21	0.17		0.29	-0.17	-0.18	-0.07
24	MC	1	60847	0.56	0.21	0.00		0.55	0.15	0.19	0.11		0.21	0.03	-0.09	-0.25
25	TE	2	60863	0.70	0.62	0.00	0.19	0.21	0.59			-0.49	-0.25	0.61		
26	MC	1	60810	0.55	0.32	0.00		0.16	0.55	0.13	0.16		-0.23	0.32	-0.17	-0.05
27	MS	2	60720	0.57	0.47	0.00	0.20	0.45	0.34			-0.35	-0.12	0.42		
28	MC	1	60561	0.59	0.45	0.01		0.06	0.19	0.59	0.16		-0.21	-0.19	0.45	-0.25
29	MC	1	60870	0.73	0.44	0.00		0.11	0.09	0.07	0.73		-0.26	-0.22	-0.19	0.44
30	MC	1	60862	0.26	0.25	0.00		0.26	0.26	0.23	0.25		0.25	-0.01	-0.05	-0.19
31	EBSR	2	60911	0.46	0.44	0.00	0.29	0.50	0.20			-0.42	0.15	0.29		
32	MC	1	60766	0.45	0.34	0.00		0.17	0.44	0.24	0.14		-0.18	0.34	-0.07	-0.20
33	MC	1	60763	0.43	0.36	0.00		0.21	0.13	0.43	0.23		-0.08	-0.16	0.36	-0.22
34	MC	1	60823	0.61	0.50	0.00		0.61	0.17	0.15	0.07		0.50	-0.19	-0.28	-0.27
35	MC	1	60818	0.46	0.31	0.00		0.17	0.11	0.26	0.46		-0.11	-0.15	-0.14	0.31
36	MC	1	60830	0.44	0.30	0.00		0.17	0.44	0.08	0.31		-0.10	0.30	-0.21	-0.12
37	TE	2	60813	0.52	0.47	0.00	0.31	0.33	0.36			-0.37	-0.09	0.45		

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-2. Item Statistics, ELA Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	63426	0.34	0.52	0.00	0.19	0.34	0.37	0.09	0.01	-0.38	-0.15	0.27	0.29	0.11
2	TE	1	63170	0.56	0.46	0.00	0.44	0.56				-0.46	0.46			
3	TE	1	63309	0.63	0.40	0.00	0.37	0.63				-0.40	0.40			
4	MC	1	63342	0.68	0.32	0.00		0.20	0.67	0.09	0.04		-0.23	0.32	-0.12	-0.13
5	TE	1	63192	0.45	0.41	0.00	0.55	0.45				-0.40	0.41			
6	TE	2	63213	0.64	0.45	0.00	0.12	0.49	0.39			-0.31	-0.18	0.39		
7	MC	1	63307	0.58	0.45	0.00		0.14	0.20	0.58	0.08		-0.24	-0.23	0.45	-0.16
8	MC	1	63313	0.56	0.37	0.00		0.24	0.05	0.15	0.55		-0.08	-0.23	-0.27	0.37
9	MC	1	63336	0.55	0.39	0.00		0.28	0.09	0.08	0.55		-0.12	-0.25	-0.24	0.39
10	MC	1	63327	0.56	0.19	0.00		0.56	0.24	0.09	0.11		0.20	0.04	-0.17	-0.20
11	TE	2	63083	0.70	0.40	0.01	0.06	0.48	0.46			-0.17	-0.31	0.40		
12	MC	1	63322	0.68	0.46	0.00		0.68	0.10	0.10	0.12		0.46	-0.26	-0.21	-0.23
13	TE	2	63338	0.78	0.34	0.00	0.03	0.37	0.60			-0.21	-0.23	0.31		
14	MC	1	63319	0.41	0.21	0.00		0.17	0.41	0.18	0.25		-0.10	0.21	-0.11	-0.05
15	MC	1	63394	0.57	0.27	0.00		0.11	0.20	0.57	0.13		-0.12	-0.08	0.27	-0.19
16	MC	1	63354	0.76	0.34	0.00		0.02	0.20	0.03	0.75		-0.17	-0.22	-0.19	0.34
17	EBSR	2	63404	0.32	0.32	0.00	0.57	0.23	0.20			-0.29	0.08	0.27		
18	MC	1	63374	0.56	0.38	0.00		0.11	0.14	0.56	0.18		-0.27	-0.15	0.38	-0.13
19	MC	1	63344	0.64	0.38	0.00		0.14	0.63	0.11	0.12		-0.20	0.38	-0.18	-0.18
20	EBSR	2	63392	0.36	0.33	0.00	0.52	0.24	0.24			-0.27	-0.01	0.33		

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	63328	0.73	0.44	0.00		0.09	0.73	0.09	0.10		-0.26	0.44	-0.25	-0.17
22	TE	2	63206	0.57	0.39	0.00	0.17	0.53	0.30			-0.32	-0.04	0.31		
23	MC	1	63293	0.53	0.40	0.00		0.53	0.27	0.09	0.10		0.40	-0.16	-0.23	-0.20
24	MC	1	63298	0.59	0.51	0.00		0.12	0.19	0.11	0.59		-0.23	-0.22	-0.30	0.51
25	MS	2	63306	0.47	0.44	0.00	0.32	0.41	0.27			-0.34	-0.03	0.40		
26	EBSR	2	63373	0.58	0.51	0.00	0.21	0.42	0.37			-0.43	-0.05	0.42		
27	MC	1	63232	0.60	0.33	0.00		0.09	0.11	0.20	0.60		-0.13	-0.23	-0.12	0.33
28	MC	1	63153	0.68	0.48	0.00		0.07	0.08	0.17	0.68		-0.22	-0.28	-0.23	0.48
29	MC	1	63244	0.41	0.22	0.00		0.41	0.12	0.34	0.13		0.23	-0.19	0.06	-0.22
30	MC	1	63273	0.50	0.36	0.00		0.21	0.17	0.50	0.12		-0.17	-0.15	0.36	-0.17
31	MC	1	63018	0.53	0.38	0.01		0.53	0.19	0.12	0.16		0.38	-0.20	-0.24	-0.09
32	TE	2	63072	0.47	0.45	0.01	0.30	0.46	0.24			-0.29	-0.12	0.46		
33	MC	1	63250	0.60	0.40	0.00		0.11	0.08	0.59	0.21		-0.16	-0.24	0.40	-0.19
34	MC	1	63255	0.61	0.49	0.00		0.61	0.14	0.15	0.10		0.49	-0.21	-0.24	-0.25
35	MC	1	63261	0.56	0.38	0.00		0.56	0.21	0.13	0.10		0.38	-0.19	-0.21	-0.13
36	MC	1	63257	0.64	0.37	0.00		0.20	0.64	0.10	0.06		-0.13	0.37	-0.25	-0.22
37	MC	1	63272	0.58	0.44	0.00		0.14	0.17	0.11	0.58		-0.20	-0.19	-0.24	0.44
38	TE	2	63281	0.47	0.43	0.00	0.23	0.59	0.17			-0.36	0.07	0.32		
39	MC	1	63275	0.61	0.43	0.00		0.08	0.14	0.61	0.17		-0.21	-0.17	0.43	-0.24

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-3. Item Statistics, ELA Grade 5

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	64566	0.37	0.52	0.00	0.07	0.47	0.36	0.08	0.01	-0.31	-0.30	0.32	0.27	0.09
2	MC	1	64542	0.80	0.39	0.00		0.13	0.80	0.04	0.03		-0.30	0.39	-0.17	-0.12
3	MC	1	64511	0.68	0.40	0.00		0.05	0.13	0.14	0.68		-0.18	-0.20	-0.22	0.40
4	TE	2	64502	0.59	0.43	0.00	0.16	0.48	0.35			-0.32	-0.12	0.37		
5	MC	1	64495	0.41	0.26	0.00		0.41	0.10	0.25	0.24		0.26	-0.17	-0.04	-0.14
6	TE	2	64423	0.54	0.50	0.00	0.31	0.29	0.39			-0.35	-0.21	0.53		
7	MC	1	64494	0.66	0.34	0.00		0.09	0.09	0.65	0.16		-0.14	-0.23	0.34	-0.16
8	TE	1	64443	0.72	0.53	0.00	0.28	0.71				-0.53	0.53			
9	MC	1	64496	0.64	0.43	0.00		0.23	0.07	0.63	0.07		-0.20	-0.29	0.43	-0.19
10	MC	1	64482	0.53	0.36	0.00		0.29	0.53	0.08	0.10		-0.17	0.36	-0.20	-0.17
11	TE	2	64499	0.71	0.38	0.00	0.12	0.34	0.54			-0.34	-0.09	0.31		
12	MC	1	64482	0.59	0.33	0.00		0.03	0.29	0.08	0.59		-0.23	-0.14	-0.20	0.33
13	TE	2	64458	0.58	0.34	0.00	0.16	0.53	0.31			-0.16	-0.21	0.36		
14	TE	2	64541	0.62	0.40	0.00	0.13	0.51	0.36			-0.32	-0.11	0.33		
15	MC	1	64492	0.64	0.35	0.00		0.64	0.28	0.05	0.04		0.35	-0.23	-0.15	-0.19
16	MC	1	64497	0.78	0.43	0.00		0.15	0.04	0.04	0.78		-0.29	-0.21	-0.19	0.43
17	MC	1	64513	0.59	0.41	0.00		0.14	0.18	0.59	0.08		-0.30	-0.11	0.41	-0.20
18	MC	1	64479	0.59	0.37	0.00		0.13	0.17	0.11	0.58		-0.15	-0.18	-0.20	0.37
19	EBSR	2	64543	0.57	0.50	0.00	0.28	0.28	0.43			-0.43	-0.06	0.45		
20	MC	1	64483	0.73	0.48	0.00		0.09	0.73	0.06	0.12		-0.20	0.48	-0.24	-0.29

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	64436	0.59	0.17	0.00		0.10	0.03	0.59	0.28		-0.17	-0.25	0.17	0.03
22	MC	1	64414	0.69	0.42	0.00		0.16	0.06	0.09	0.69		-0.14	-0.26	-0.28	0.42
23	TE	1	64371	0.86	0.46	0.00	0.14	0.86				-0.46	0.46			
24	MC	1	64406	0.55	0.33	0.00		0.12	0.55	0.12	0.21		-0.14	0.33	-0.22	-0.11
25	TE	1	64351	0.69	0.54	0.00	0.31	0.69				-0.54	0.54			
26	MC	1	64422	0.60	0.44	0.00		0.60	0.10	0.19	0.11		0.44	-0.25	-0.22	-0.16
27	MC	1	64272	0.54	0.20	0.00		0.27	0.06	0.13	0.54		0.03	-0.24	-0.16	0.20
28	MC	1	64396	0.51	0.25	0.00		0.51	0.19	0.10	0.20		0.25	-0.17	-0.18	-0.01
29	MS	2	64371	0.60	0.51	0.00	0.18	0.45	0.37			-0.38	-0.15	0.45		
30	MC	1	64404	0.61	0.38	0.00		0.23	0.61	0.10	0.06		-0.13	0.39	-0.29	-0.19
31	MC	1	64422	0.56	0.49	0.00		0.55	0.13	0.18	0.13		0.49	-0.30	-0.18	-0.21
32	MC	1	64288	0.65	0.54	0.00		0.13	0.07	0.15	0.65		-0.32	-0.28	-0.22	0.54
33	MC	1	64174	0.57	0.46	0.01		0.56	0.14	0.11	0.18		0.46	-0.23	-0.29	-0.14
34	MC	1	64353	0.44	0.29	0.00		0.44	0.18	0.30	0.08		0.29	-0.20	-0.05	-0.14
35	MC	1	64353	0.60	0.44	0.00		0.21	0.09	0.09	0.60		-0.20	-0.24	-0.21	0.44
36	TE	2	64422	0.62	0.55	0.00	0.15	0.44	0.40			-0.39	-0.20	0.50		
37	TE	2	64294	0.65	0.25	0.00	0.07	0.56	0.37			-0.26	-0.03	0.18		
38	MC	1	64396	0.43	0.45	0.00		0.16	0.28	0.13	0.43		-0.22	-0.11	-0.27	0.45
39	MC	1	64394	0.46	0.33	0.00		0.18	0.15	0.46	0.21		-0.13	-0.15	0.33	-0.14
40	MC	1	64407	0.58	0.44	0.00		0.16	0.16	0.58	0.10		-0.21	-0.23	0.44	-0.18

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-4. Item Statistics, ELA Grade 6

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	65260	0.38	0.53	0.00	0.07	0.46	0.37	0.09	0.01	-0.32	-0.31	0.32	0.27	0.11
2	MC	1	65217	0.64	0.22	0.00		0.64	0.21	0.11	0.04		0.22	-0.08	-0.14	-0.15
3	MC	1	65174	0.66	0.26	0.00		0.66	0.07	0.14	0.13		0.26	-0.14	-0.14	-0.11
4	MC	1	65158	0.71	0.45	0.00		0.04	0.14	0.71	0.10		-0.20	-0.28	0.45	-0.22
5	MC	1	65147	0.36	0.20	0.00		0.15	0.36	0.21	0.27		-0.10	0.20	-0.17	0.03
6	TE	1	65104	0.42	0.23	0.00	0.58	0.42				-0.22	0.23			
7	TE	2	65149	0.68	0.41	0.00	0.10	0.43	0.47			-0.34	-0.14	0.34		
8	MC	1	65174	0.44	0.31	0.00		0.44	0.06	0.22	0.28		0.31	-0.21	-0.14	-0.10
9	TE	2	65141	0.59	0.51	0.00	0.25	0.31	0.44			-0.40	-0.13	0.48		
10	TE	2	65115	0.65	0.35	0.00	0.07	0.55	0.37			-0.30	-0.11	0.28		
11	TE	2	65136	0.63	0.42	0.00	0.14	0.46	0.40			-0.36	-0.07	0.33		
12	MC	1	65148	0.48	0.28	0.00		0.12	0.17	0.24	0.47		-0.14	-0.17	-0.08	0.29
13	MC	1	65160	0.46	0.33	0.00		0.13	0.27	0.13	0.46		-0.13	-0.14	-0.18	0.33
14	MC	1	65185	0.56	0.33	0.00		0.08	0.21	0.56	0.16		-0.16	-0.32	0.34	0.02
15	MC	1	65127	0.56	0.31	0.00		0.06	0.56	0.23	0.14		-0.12	0.31	-0.22	-0.09
16	TE	2	65128	0.59	0.38	0.00	0.17	0.48	0.35			-0.27	-0.12	0.34		
17	EBSR	2	65202	0.74	0.52	0.00	0.15	0.21	0.64			-0.39	-0.25	0.51		
18	MC	1	65132	0.61	0.39	0.00		0.17	0.61	0.08	0.14		-0.11	0.39	-0.26	-0.22
19	MC	1	65140	0.70	0.33	0.00		0.11	0.14	0.05	0.70		-0.08	-0.20	-0.28	0.33
20	MC	1	65165	0.45	0.20	0.00		0.25	0.27	0.03	0.45		-0.08	-0.04	-0.25	0.20

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	TE	2	65114	0.66	0.45	0.00	0.12	0.45	0.43			-0.34	-0.15	0.38		
22	MC	1	65081	0.51	0.32	0.00		0.08	0.51	0.20	0.22		-0.20	0.32	-0.23	-0.03
23	TE	2	65047	0.59	0.44	0.00	0.15	0.52	0.33			-0.35	-0.08	0.35		
24	MC	1	65043	0.55	0.32	0.00		0.26	0.09	0.55	0.10		-0.14	-0.20	0.32	-0.15
25	MC	1	65031	0.80	0.49	0.00		0.07	0.04	0.08	0.80		-0.26	-0.27	-0.26	0.49
26	TE	2	65076	0.58	0.41	0.00	0.15	0.54	0.31			-0.29	-0.13	0.37		
27	MC	1	64960	0.62	0.45	0.00		0.05	0.21	0.11	0.62		-0.19	-0.16	-0.34	0.45
28	TE	2	65045	0.63	0.40	0.00	0.13	0.47	0.40			-0.31	-0.11	0.33		
29	TE	1	64968	0.67	0.41	0.00	0.33	0.66				-0.40	0.41			
30	MC	1	65076	0.43	0.26	0.00		0.17	0.43	0.27	0.13		-0.06	0.26	-0.15	-0.12
31	MC	1	64953	0.53	0.43	0.00		0.53	0.24	0.13	0.11		0.43	-0.15	-0.21	-0.25
32	TE	2	65055	0.56	0.22	0.00	0.14	0.59	0.27			-0.11	-0.11	0.22		
33	EBSR	2	65125	0.52	0.60	0.00	0.36	0.24	0.40			-0.54	-0.01	0.54		
34	TE	1	64997	0.59	0.45	0.00	0.41	0.59				-0.44	0.45			
35	MC	1	65070	0.45	0.39	0.00		0.45	0.24	0.20	0.12		0.39	-0.13	-0.27	-0.10
36	MC	1	65076	0.58	0.41	0.00		0.11	0.58	0.18	0.14		-0.18	0.41	-0.24	-0.15
37	MC	1	65099	0.64	0.50	0.00		0.64	0.11	0.11	0.14		0.50	-0.19	-0.28	-0.26

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-5. Item Statistics, ELA Grade 7

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	63738	0.40	0.53	0.00	0.05	0.43	0.40	0.11	0.02	-0.28	-0.37	0.29	0.28	0.14
2	MC	1	63703	0.71	0.41	0.00		0.06	0.71	0.19	0.04		-0.21	0.42	-0.28	-0.14
3	TE	2	63693	0.52	0.38	0.00	0.25	0.46	0.29			-0.29	-0.05	0.34		
4	TE	2	63648	0.70	0.46	0.00	0.14	0.32	0.54			-0.27	-0.30	0.48		
5	MC	1	63676	0.73	0.37	0.00		0.06	0.73	0.12	0.09		-0.23	0.38	-0.20	-0.16
6	TE	2	63627	0.65	0.30	0.00	0.06	0.57	0.36			-0.16	-0.19	0.28		
7	MC	1	63639	0.61	0.46	0.00		0.61	0.22	0.08	0.08		0.46	-0.29	-0.23	-0.14
8	MC	1	63675	0.71	0.35	0.00		0.12	0.07	0.10	0.71		-0.17	-0.18	-0.19	0.35
9	MC	1	63677	0.63	0.37	0.00		0.10	0.04	0.63	0.23		-0.24	-0.25	0.38	-0.14
10	TE	2	63645	0.47	0.23	0.00	0.24	0.58	0.17			-0.17	0.00	0.19		
11	MC	1	63624	0.47	0.32	0.00		0.47	0.22	0.20	0.11		0.32	-0.17	-0.13	-0.10
12	MC	1	63666	0.68	0.36	0.00		0.18	0.08	0.68	0.06		-0.10	-0.24	0.36	-0.26
13	MC	1	63655	0.66	0.45	0.00		0.06	0.09	0.19	0.66		-0.20	-0.34	-0.17	0.45
14	EBSR	2	63710	0.75	0.43	0.00	0.19	0.12	0.70			-0.35	-0.21	0.44		
15	MS	2	63598	0.77	0.48	0.00	0.09	0.28	0.63			-0.32	-0.29	0.46		
16	MC	1	63629	0.59	0.36	0.00		0.05	0.21	0.15	0.59		-0.17	-0.17	-0.20	0.36
17	MC	1	63627	0.56	0.30	0.00		0.09	0.17	0.17	0.56		-0.07	-0.22	-0.12	0.30
18	EBSR	2	63697	0.38	0.41	0.00	0.52	0.20	0.28			-0.33	-0.06	0.41		
19	TE	1	63457	0.73	0.34	0.00	0.27	0.73				-0.33	0.35			
20	MC	1	63602	0.61	0.52	0.00		0.11	0.21	0.07	0.60		-0.29	-0.21	-0.27	0.52

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	EBSR	2	63674	0.51	0.38	0.00	0.41	0.16	0.43			-0.30	-0.13	0.40		
22	MC	1	63586	0.51	0.36	0.00		0.22	0.51	0.17	0.10		-0.09	0.36	-0.18	-0.26
23	MC	1	63527	0.54	0.44	0.00		0.07	0.19	0.20	0.53		-0.25	-0.13	-0.25	0.44
24	TE	1	62791	0.61	0.43	0.01	0.38	0.60				-0.41	0.44			
25	TE	2	63467	0.67	0.54	0.00	0.15	0.37	0.48			-0.43	-0.17	0.48		
26	TE	2	63446	0.53	0.33	0.00	0.17	0.58	0.24			-0.33	0.09	0.20		
27	MC	1	63531	0.38	0.25	0.00		0.27	0.38	0.13	0.22		-0.10	0.26	-0.25	0.01
28	EBSR	2	63611	0.44	0.43	0.00	0.44	0.24	0.32			-0.36	-0.02	0.41		
29	TE	1	62983	0.34	0.44	0.01	0.66	0.33				-0.41	0.44			
30	MC	1	63522	0.71	0.55	0.00		0.07	0.14	0.07	0.71		-0.26	-0.29	-0.30	0.55
31	MC	1	63513	0.44	0.32	0.00		0.23	0.20	0.44	0.12		-0.17	-0.14	0.32	-0.10
32	MC	1	63505	0.54	0.37	0.00		0.15	0.21	0.54	0.10		-0.14	-0.17	0.38	-0.21
33	MC	1	63484	0.48	0.45	0.00		0.47	0.19	0.16	0.17		0.45	-0.24	-0.25	-0.09
34	MS	2	63511	0.59	0.48	0.00	0.17	0.49	0.34			-0.39	-0.07	0.39		
35	MS	2	63538	0.72	0.45	0.00	0.08	0.40	0.52			-0.31	-0.24	0.41		
36	MC	1	63525	0.57	0.42	0.00		0.13	0.57	0.16	0.14		-0.17	0.42	-0.24	-0.18

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-6. Item Statistics, ELA Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TDA	4	62891	0.42	0.57	0.00	0.05	0.44	0.33	0.17	0.01	-0.28	-0.40	0.22	0.37	0.13
2	MC	1	62867	0.74	0.39	0.00		0.07	0.02	0.16	0.74		-0.20	-0.20	-0.25	0.39
3	MC	1	62823	0.62	0.45	0.00		0.18	0.62	0.11	0.09		-0.19	0.45	-0.27	-0.20
4	MC	1	62808	0.69	0.45	0.00		0.09	0.11	0.11	0.69		-0.18	-0.25	-0.24	0.45
5	MC	1	62821	0.51	0.33	0.00		0.15	0.51	0.12	0.22		-0.21	0.33	-0.21	-0.04
6	MC	1	62793	0.52	0.25	0.00		0.29	0.52	0.10	0.08		-0.09	0.26	-0.20	-0.09
7	TE	2	62767	0.69	0.47	0.00	0.09	0.43	0.48			-0.33	-0.23	0.42		
8	MC	1	62812	0.55	0.52	0.00		0.55	0.11	0.20	0.14		0.52	-0.22	-0.28	-0.22
9	MC	1	62825	0.80	0.33	0.00		0.08	0.80	0.03	0.09		-0.16	0.34	-0.23	-0.18
10	TE	2	62779	0.48	0.30	0.00	0.25	0.54	0.21			-0.20	-0.06	0.29		
11	MC	1	62773	0.45	0.31	0.00		0.11	0.44	0.06	0.38		-0.20	0.31	-0.24	-0.07
12	TE	2	62807	0.86	0.37	0.00	0.04	0.20	0.75			-0.23	-0.27	0.36		
13	MC	1	62806	0.57	0.45	0.00		0.13	0.12	0.18	0.57		-0.27	-0.24	-0.14	0.45
14	MC	1	62802	0.52	0.37	0.00		0.13	0.09	0.26	0.52		-0.13	-0.29	-0.13	0.38
15	MS	2	62826	0.67	0.45	0.00	0.08	0.50	0.42			-0.26	-0.28	0.43		
16	EBSR	2	62835	0.69	0.49	0.00	0.12	0.38	0.50			-0.39	-0.18	0.43		
17	MC	1	62781	0.70	0.47	0.00		0.09	0.14	0.70	0.07		-0.26	-0.25	0.47	-0.21
18	MC	1	62774	0.49	0.37	0.00		0.49	0.11	0.24	0.16		0.37	-0.16	-0.20	-0.14
19	EBSR	2	62809	0.44	0.39	0.00	0.51	0.11	0.38			-0.33	-0.09	0.40		
20	MC	1	62743	0.40	0.34	0.00		0.16	0.28	0.16	0.40		-0.24	-0.10	-0.10	0.35

Note: TDA responses that received a condition code were converted to 0 and are reported as a score of 0.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	62690	0.45	0.17	0.00		0.06	0.45	0.26	0.23		-0.26	0.18	0.01	-0.06
22	MC	1	62712	0.71	0.43	0.00		0.71	0.09	0.05	0.15		0.44	-0.25	-0.28	-0.18
23	MC	1	62673	0.54	0.33	0.00		0.54	0.16	0.18	0.12		0.33	-0.05	-0.22	-0.19
24	MS	2	62688	0.62	0.46	0.00	0.14	0.49	0.37			-0.35	-0.13	0.39		
25	MC	1	62663	0.41	0.44	0.00		0.25	0.24	0.10	0.40		-0.10	-0.21	-0.26	0.44
26	MC	1	62717	0.60	0.47	0.00		0.03	0.06	0.31	0.60		-0.19	-0.29	-0.28	0.47
27	MC	1	62672	0.50	0.27	0.00		0.16	0.50	0.18	0.16		-0.14	0.27	-0.17	-0.05
28	MC	1	62623	0.74	0.54	0.00		0.74	0.10	0.10	0.06		0.54	-0.26	-0.33	-0.24
29	MC	1	62597	0.45	0.22	0.00		0.07	0.30	0.17	0.45		-0.16	-0.12	-0.03	0.22
30	EBSR	2	62762	0.60	0.52	0.00	0.30	0.20	0.49			-0.41	-0.17	0.52		
31	MC	1	62682	0.53	0.38	0.00		0.53	0.20	0.21	0.06		0.38	-0.24	-0.09	-0.22
32	EBSR	2	62735	0.56	0.62	0.00	0.29	0.30	0.41			-0.58	0.03	0.51		
33	MC	1	62673	0.76	0.52	0.00		0.07	0.76	0.09	0.08		-0.22	0.52	-0.32	-0.27
34	MC	1	62662	0.67	0.43	0.00		0.09	0.10	0.66	0.14		-0.16	-0.20	0.43	-0.27
35	MC	1	62651	0.38	0.22	0.00		0.27	0.24	0.11	0.38		0.07	-0.16	-0.22	0.23
36	MC	1	62670	0.68	0.47	0.00		0.68	0.19	0.09	0.04		0.47	-0.25	-0.27	-0.21
37	MC	1	62666	0.53	0.40	0.00		0.13	0.17	0.53	0.16		-0.16	-0.25	0.41	-0.13
38	MS	2	62684	0.66	0.59	0.00	0.16	0.36	0.48			-0.41	-0.26	0.56		
39	MC	1	62681	0.59	0.40	0.00		0.05	0.15	0.22	0.59		-0.22	-0.13	-0.25	0.40

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-7. Item Statistics, Mathematics Grade 3

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TE	1	61087	0.71	0.52	0.00	0.29	0.71				-0.52	0.52			
2	SA	1	61085	0.58	0.53	0.00	0.42	0.58				-0.53	0.53			
3	TE	1	61101	0.45	0.59	0.00	0.55	0.45				-0.58	0.59			
4	MC	1	61077	0.76	0.45	0.00		0.04	0.08	0.76	0.12		-0.13	-0.32	0.45	-0.25
5	MC	1	61070	0.53	0.46	0.00		0.25	0.05	0.17	0.53		-0.29	-0.07	-0.24	0.46
6	SA	1	60865	0.81	0.39	0.00	0.19	0.81				-0.39	0.40			
7	MC	1	60781	0.44	0.28	0.01		0.09	0.33	0.14	0.43		-0.25	0.02	-0.20	0.28
8	SA	1	61042	0.46	0.56	0.00	0.54	0.46				-0.55	0.56			
9	MC	1	61043	0.67	0.39	0.00		0.67	0.12	0.12	0.10		0.39	-0.25	-0.14	-0.20
10	SA	1	61059	0.64	0.59	0.00	0.36	0.64				-0.58	0.59			
11	MC	1	61062	0.41	0.35	0.00		0.21	0.20	0.41	0.18		-0.09	-0.14	0.35	-0.21
12	TE	1	60786	0.22	0.39	0.01	0.78	0.22				-0.37	0.39			
13	MC	1	60884	0.62	0.42	0.00		0.62	0.10	0.19	0.09		0.42	-0.12	-0.27	-0.21
14	SA	1	60698	0.66	0.51	0.01	0.33	0.66				-0.50	0.51			
15	SA	1	61001	0.34	0.51	0.00	0.66	0.34				-0.50	0.51			
16	MC	1	61040	0.33	0.39	0.00		0.33	0.15	0.21	0.31		0.39	0.04	-0.16	-0.29
17	MC	1	61048	0.79	0.48	0.00		0.79	0.13	0.05	0.03		0.48	-0.34	-0.21	-0.20
18	TE	1	60542	0.86	0.39	0.01	0.14	0.85				-0.38	0.40			
19	MC	1	61049	0.71	0.50	0.00		0.05	0.10	0.14	0.71		-0.23	-0.28	-0.27	0.50
20	MC	1	61014	0.59	0.41	0.00		0.19	0.59	0.10	0.12		-0.20	0.41	-0.15	-0.24

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	SA	1	61029	0.60	0.52	0.00	0.40	0.60				-0.52	0.52			
22	MC	1	61032	0.47	0.36	0.00		0.47	0.34	0.08	0.12		0.36	-0.04	-0.17	-0.35
23	MC	1	61067	0.77	0.49	0.00		0.11	0.77	0.07	0.05		-0.25	0.49	-0.30	-0.23
24	SA	1	61045	0.61	0.57	0.00	0.39	0.61				-0.57	0.57			
25	TE	1	61050	0.62	0.34	0.00	0.38	0.62				-0.34	0.34			
26	TE	1	60970	0.48	0.56	0.00	0.52	0.47				-0.55	0.56			
27	MC	1	60681	0.61	0.44	0.01		0.15	0.12	0.61	0.11		-0.23	-0.23	0.44	-0.17
28	MC	1	60617	0.42	0.35	0.01		0.13	0.14	0.30	0.42		-0.25	-0.19	-0.04	0.36
29	MC	1	61035	0.61	0.39	0.00		0.13	0.15	0.61	0.11		-0.22	-0.19	0.39	-0.15
30	MC	1	61014	0.41	0.22	0.00		0.20	0.23	0.16	0.41		-0.10	-0.02	-0.16	0.22
31	MC	1	61006	0.44	0.34	0.00		0.18	0.21	0.17	0.44		-0.13	-0.08	-0.23	0.35
32	MC	1	60998	0.42	0.36	0.00		0.29	0.14	0.15	0.42		-0.11	-0.19	-0.17	0.36
33	SA	1	61018	0.51	0.57	0.00	0.48	0.51				-0.56	0.57			
34	MC	1	60849	0.67	0.45	0.00		0.66	0.09	0.10	0.15		0.45	-0.19	-0.19	-0.27
35	SA	1	60786	0.62	0.63	0.01	0.38	0.62				-0.63	0.64			
36	TE	1	60786	0.25	0.48	0.01	0.74	0.25				-0.46	0.48			
37	MC	1	60984	0.48	0.46	0.00		0.19	0.15	0.18	0.48		-0.24	-0.22	-0.14	0.46
38	MC	1	61024	0.71	0.37	0.00		0.11	0.03	0.15	0.71		-0.11	-0.09	-0.33	0.37
39	SA	1	61031	0.59	0.61	0.00	0.40	0.59				-0.61	0.61			
40	MC	1	61049	0.67	0.48	0.00		0.16	0.67	0.12	0.04		-0.28	0.48	-0.21	-0.26
41	TE	1	60167	0.35	0.47	0.02	0.64	0.35				-0.43	0.47			
42	MC	1	61027	0.72	0.45	0.00		0.07	0.09	0.12	0.72		-0.27	-0.22	-0.21	0.45

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-8. Item Statistics, Mathematics Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	63407	0.68	0.45	0.00		0.67	0.16	0.08	0.09		0.45	-0.29	-0.21	-0.18
2	MC	1	63506	0.45	0.30	0.00		0.05	0.42	0.07	0.45		-0.20	-0.12	-0.17	0.30
3	SA	1	63497	0.42	0.38	0.00	0.58	0.42				-0.37	0.38			
4	MC	1	63485	0.84	0.40	0.00		0.10	0.84	0.04	0.02		-0.29	0.40	-0.21	-0.14
5	TE	1	63462	0.41	0.50	0.00	0.59	0.41				-0.50	0.50			
6	MC	1	63485	0.56	0.48	0.00		0.56	0.12	0.17	0.16		0.48	-0.15	-0.26	-0.26
7	MC	1	63484	0.34	0.49	0.00		0.02	0.34	0.16	0.48		-0.08	0.49	-0.09	-0.37
8	SA	1	63336	0.37	0.60	0.00	0.63	0.37				-0.60	0.60			
9	MC	1	63489	0.71	0.46	0.00		0.71	0.09	0.14	0.06		0.46	-0.26	-0.22	-0.25
10	MC	1	63493	0.87	0.40	0.00		0.02	0.07	0.87	0.04		-0.16	-0.31	0.40	-0.16
11	MC	1	63462	0.31	0.38	0.00		0.10	0.31	0.33	0.26		-0.03	0.38	-0.28	-0.08
12	MC	1	63474	0.37	0.33	0.00		0.26	0.26	0.37	0.11		-0.15	-0.14	0.33	-0.10
13	MC	1	63463	0.50	0.58	0.00		0.20	0.10	0.19	0.50		-0.27	-0.29	-0.23	0.58
14	MC	1	63465	0.42	0.28	0.00		0.15	0.33	0.11	0.41		-0.01	-0.24	-0.05	0.28
15	MC	1	63342	0.28	0.15	0.00		0.28	0.29	0.25	0.18		0.15	-0.16	-0.02	0.04
16	SA	1	63313	0.34	0.51	0.00	0.66	0.34				-0.50	0.51			
17	MC	1	63479	0.80	0.46	0.00		0.05	0.80	0.10	0.04		-0.22	0.46	-0.30	-0.21
18	MC	1	63450	0.43	0.46	0.00		0.15	0.25	0.18	0.43		-0.27	-0.09	-0.24	0.46
19	TE	1	63480	0.35	0.59	0.00	0.65	0.35				-0.59	0.59			
20	MC	1	63457	0.45	0.39	0.00		0.18	0.45	0.15	0.21		-0.16	0.39	-0.21	-0.13
21	MC	1	63484	0.88	0.32	0.00		0.01	0.08	0.88	0.03		-0.12	-0.24	0.33	-0.16
22	MC	1	63444	0.29	0.17	0.00		0.21	0.29	0.17	0.33		-0.18	0.17	-0.15	0.12
23	SA	1	63454	0.22	0.54	0.00	0.77	0.22				-0.53	0.54			

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	TE	1	63023	0.70	0.36	0.01	0.29	0.70				-0.35	0.36			
25	MC	1	63466	0.86	0.34	0.00		0.06	0.05	0.86	0.04		-0.23	-0.17	0.34	-0.16
26	MC	1	63454	0.50	0.50	0.00		0.26	0.50	0.14	0.10		-0.36	0.50	-0.17	-0.10
27	MC	1	63437	0.34	0.56	0.00		0.33	0.07	0.47	0.13		0.56	-0.08	-0.28	-0.31
28	MC	1	63470	0.51	0.40	0.00		0.12	0.18	0.19	0.51		-0.21	-0.13	-0.21	0.41
29	SA	1	63415	0.38	0.52	0.00	0.62	0.38				-0.51	0.52			
30	MC	1	63429	0.32	0.42	0.00		0.32	0.13	0.14	0.40		0.42	-0.13	-0.07	-0.27
31	MC	1	63351	0.40	0.20	0.00		0.10	0.21	0.28	0.40		-0.13	-0.09	-0.04	0.21
32	TE	1	63299	0.24	0.49	0.00	0.76	0.24				-0.47	0.49			
33	MC	1	63461	0.52	0.58	0.00		0.52	0.32	0.08	0.08		0.58	-0.50	-0.09	-0.12
34	MC	1	63442	0.41	0.42	0.00		0.06	0.47	0.05	0.41		-0.25	-0.20	-0.20	0.42
35	SA	1	63436	0.18	0.46	0.00	0.82	0.18				-0.46	0.46			
36	MC	1	63420	0.53	0.41	0.00		0.53	0.16	0.15	0.15		0.41	-0.24	-0.18	-0.14
37	MC	1	63445	0.85	0.32	0.00		0.04	0.06	0.05	0.85		-0.18	-0.21	-0.14	0.32
38	MC	1	63342	0.35	0.54	0.00		0.35	0.17	0.27	0.21		0.54	-0.07	-0.21	-0.34
39	TE	1	63377	0.59	0.61	0.00	0.41	0.59				-0.61	0.61			
40	SA	1	63445	0.57	0.45	0.00	0.43	0.57				-0.45	0.45			
41	MC	1	63433	0.58	0.41	0.00		0.23	0.58	0.12	0.06		-0.29	0.41	-0.17	-0.11
42	MC	1	63443	0.28	0.13	0.00		0.48	0.16	0.09	0.27		-0.14	0.10	-0.08	0.13
43	SA	1	63425	0.51	0.44	0.00	0.49	0.51				-0.44	0.44			
44	MC	1	63439	0.28	0.40	0.00		0.28	0.22	0.22	0.28		0.40	-0.10	-0.13	-0.19
45	SA	1	63427	0.49	0.54	0.00	0.51	0.49				-0.54	0.54			
46	MC	1	63408	0.53	0.38	0.00		0.53	0.19	0.21	0.07		0.38	-0.20	-0.17	-0.16

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-9. Item Statistics, Mathematics Grade 5

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	SA	1	64537	0.66	0.50	0.00	0.34	0.66				-0.50	0.51			
2	MC	1	64616	0.46	0.50	0.00		0.46	0.47	0.04	0.02		0.50	-0.38	-0.23	-0.10
3	MC	1	64566	0.57	0.18	0.00		0.57	0.24	0.15	0.04		0.18	0.02	-0.22	-0.10
4	MC	1	64585	0.34	0.39	0.00		0.11	0.35	0.21	0.34		-0.23	-0.14	-0.10	0.39
5	MC	1	64577	0.42	0.34	0.00		0.15	0.26	0.42	0.17		-0.20	-0.18	0.34	-0.05
6	TE	1	64363	0.39	0.47	0.00	0.61	0.39				-0.46	0.47			
7	MC	1	64603	0.64	0.50	0.00		0.13	0.64	0.05	0.18		-0.14	0.50	-0.16	-0.40
8	MC	1	64502	0.43	0.35	0.00		0.43	0.22	0.11	0.23		0.35	-0.14	-0.21	-0.10
9	TE	1	64281	0.31	0.52	0.01	0.69	0.31				-0.51	0.53			
10	SA	1	64569	0.42	0.51	0.00	0.58	0.42				-0.50	0.51			
11	MC	1	64553	0.33	0.24	0.00		0.17	0.33	0.19	0.31		-0.18	0.24	-0.16	0.04
12	SA	1	64458	0.30	0.49	0.00	0.70	0.30				-0.48	0.49			
13	MC	1	64563	0.59	0.56	0.00		0.21	0.11	0.09	0.59		-0.36	-0.25	-0.17	0.56
14	SA	1	64531	0.37	0.57	0.00	0.63	0.37				-0.56	0.57			
15	MC	1	64413	0.46	0.51	0.00		0.29	0.16	0.09	0.45		-0.27	-0.20	-0.20	0.51
16	SA	1	64466	0.32	0.35	0.00	0.67	0.32				-0.34	0.35			
17	MC	1	64554	0.48	0.32	0.00		0.48	0.24	0.13	0.14		0.32	-0.05	-0.23	-0.17
18	SA	1	64532	0.47	0.44	0.00	0.53	0.47				-0.44	0.45			
19	MC	1	64552	0.63	0.52	0.00		0.07	0.14	0.16	0.63		-0.23	-0.23	-0.30	0.52
20	MC	1	64530	0.39	0.26	0.00		0.10	0.39	0.31	0.20		-0.01	0.26	-0.14	-0.14
21	SA	1	64525	0.21	0.42	0.00	0.78	0.21				-0.41	0.42			
22	TE	1	64523	0.58	0.46	0.00	0.42	0.58				-0.46	0.46			
23	MC	1	64516	0.45	0.49	0.00		0.44	0.12	0.20	0.23		0.49	-0.18	-0.29	-0.17

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	64488	0.55	0.32	0.00		0.12	0.19	0.55	0.14		-0.14	-0.09	0.32	-0.22
25	SA	1	64545	0.46	0.47	0.00	0.54	0.46				-0.47	0.47			
26	MC	1	64547	0.39	0.37	0.00		0.24	0.23	0.39	0.13		-0.07	-0.11	0.37	-0.30
27	SA	1	64545	0.25	0.44	0.00	0.74	0.25				-0.44	0.44			
28	MC	1	64543	0.50	0.40	0.00		0.50	0.10	0.14	0.26		0.40	-0.20	-0.23	-0.13
29	TE	1	64458	0.46	0.53	0.00	0.54	0.46				-0.52	0.53			
30	MC	1	64548	0.58	0.29	0.00		0.07	0.26	0.09	0.58		-0.19	-0.02	-0.30	0.29
31	TE	1	64475	0.53	0.56	0.00	0.47	0.53				-0.55	0.56			
32	MC	1	64519	0.75	0.36	0.00		0.09	0.11	0.75	0.05		-0.17	-0.22	0.36	-0.16
33	MC	1	64532	0.40	0.28	0.00		0.30	0.40	0.14	0.15		-0.24	0.28	-0.19	0.12
34	MC	1	64479	0.51	0.41	0.00		0.15	0.19	0.51	0.15		-0.26	-0.10	0.41	-0.21
35	TE	1	64527	0.44	0.58	0.00	0.56	0.44				-0.58	0.58			
36	MC	1	64539	0.30	0.42	0.00		0.41	0.19	0.30	0.10		-0.37	0.00	0.42	-0.03
37	SA	1	64450	0.45	0.53	0.00	0.55	0.45				-0.52	0.53			
38	MC	1	64443	0.46	0.39	0.00		0.33	0.11	0.10	0.46		-0.12	-0.25	-0.20	0.40
39	MC	1	64421	0.60	0.43	0.00		0.07	0.15	0.18	0.60		-0.20	-0.19	-0.23	0.43
40	TE	1	64530	0.29	0.41	0.00	0.71	0.29				-0.41	0.41			
41	MC	1	64493	0.13	0.28	0.00		0.47	0.24	0.15	0.13		-0.11	-0.06	-0.03	0.28
42	MC	1	64503	0.84	0.41	0.00		0.06	0.06	0.84	0.03		-0.25	-0.25	0.41	-0.16
43	MC	1	64482	0.32	0.38	0.00		0.19	0.29	0.20	0.32		-0.19	-0.06	-0.20	0.38
44	SA	1	64515	0.61	0.48	0.00	0.39	0.61				-0.47	0.48			
45	MC	1	64516	0.27	0.17	0.00		0.27	0.46	0.12	0.14		0.17	0.09	-0.13	-0.22
46	SA	1	64493	0.31	0.51	0.00	0.69	0.31				-0.50	0.51			

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-10. Item Statistics, Mathematics Grade 6

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	65292	0.39	0.35	0.00		0.39	0.26	0.27	0.07		0.35	-0.21	-0.20	0.04
2	MC	1	65318	0.39	0.50	0.00		0.13	0.40	0.08	0.39		-0.26	-0.26	-0.11	0.50
3	TE	1	65130	0.56	0.53	0.00	0.43	0.56				-0.52	0.53			
4	MC	1	65292	0.69	0.49	0.00		0.69	0.15	0.09	0.08		0.49	-0.24	-0.24	-0.27
5	SA	1	65180	0.49	0.48	0.00	0.51	0.49				-0.47	0.48			
6	SA	1	65144	0.49	0.51	0.00	0.51	0.49				-0.51	0.51			
7	MC	1	65283	0.28	0.40	0.00		0.23	0.28	0.07	0.42		-0.07	0.40	-0.09	-0.25
8	MC	1	65270	0.79	0.43	0.00		0.03	0.05	0.14	0.79		-0.17	-0.18	-0.32	0.43
9	SA	1	65266	0.18	0.50	0.00	0.82	0.18				-0.50	0.50			
10	MC	1	65264	0.48	0.46	0.00		0.18	0.20	0.14	0.48		-0.19	-0.20	-0.22	0.46
11	TE	1	65113	0.44	0.52	0.00	0.55	0.44				-0.51	0.52			
12	MC	1	65243	0.90	0.36	0.00		0.90	0.02	0.02	0.05		0.36	-0.15	-0.17	-0.26
13	SA	1	65197	0.49	0.56	0.00	0.51	0.48				-0.55	0.56			
14	SA	1	65111	0.23	0.50	0.00	0.77	0.22				-0.49	0.50			
15	MC	1	65257	0.44	0.59	0.00		0.44	0.18	0.33	0.05		0.59	-0.19	-0.38	-0.15
16	MC	1	65275	0.46	0.47	0.00		0.26	0.15	0.12	0.46		-0.27	-0.18	-0.15	0.47
17	MC	1	65238	0.36	0.49	0.00		0.16	0.36	0.31	0.17		-0.28	0.49	-0.08	-0.25
18	MC	1	65162	0.57	0.52	0.00		0.57	0.22	0.10	0.10		0.52	-0.35	-0.20	-0.16
19	MC	1	65157	0.70	0.38	0.00		0.06	0.70	0.12	0.11		-0.16	0.38	-0.18	-0.23
20	SA	1	65261	0.62	0.51	0.00	0.38	0.62				-0.51	0.51			
21	TE	1	65236	0.57	0.53	0.00	0.43	0.57				-0.52	0.53			
22	MC	1	65226	0.50	0.42	0.00		0.30	0.50	0.15	0.04		-0.27	0.42	-0.18	-0.10
23	TE	1	64756	0.70	0.32	0.01	0.29	0.70				-0.31	0.33			

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	65220	0.52	0.26	0.00		0.13	0.23	0.52	0.12		-0.12	-0.18	0.26	-0.05
25	MC	1	65237	0.32	0.34	0.00		0.47	0.32	0.12	0.09		-0.19	0.34	-0.21	0.02
26	MC	1	65156	0.45	0.34	0.00		0.26	0.16	0.14	0.44		-0.13	-0.21	-0.11	0.34
27	SA	1	64987	0.31	0.43	0.01	0.69	0.31				-0.42	0.43			
28	MC	1	65107	0.38	0.16	0.00		0.17	0.37	0.25	0.20		-0.20	0.16	-0.04	0.04
29	MC	1	65112	0.56	0.31	0.00		0.12	0.56	0.26	0.05		-0.11	0.31	-0.19	-0.13
30	MC	1	65231	0.58	0.34	0.00		0.17	0.20	0.58	0.06		-0.22	-0.14	0.34	-0.12
31	MC	1	65234	0.37	0.22	0.00		0.13	0.37	0.19	0.31		-0.10	0.23	-0.17	-0.02
32	MC	1	65184	0.33	0.12	0.00		0.30	0.22	0.33	0.15		-0.01	-0.02	0.12	-0.11
33	MC	1	65230	0.63	0.45	0.00		0.14	0.16	0.63	0.08		-0.24	-0.23	0.45	-0.19
34	MC	1	65209	0.61	0.48	0.00		0.14	0.08	0.17	0.61		-0.18	-0.24	-0.28	0.48
35	MC	1	65242	0.44	0.48	0.00		0.38	0.07	0.11	0.44		-0.18	-0.25	-0.26	0.48
36	TE	1	65131	0.36	0.49	0.00	0.64	0.36				-0.48	0.49			
37	SA	1	64924	0.24	0.58	0.01	0.75	0.24				-0.56	0.58			
38	MC	1	65073	0.40	0.23	0.00		0.18	0.35	0.40	0.07		-0.15	-0.02	0.23	-0.16
39	MC	1	65104	0.42	0.43	0.00		0.27	0.21	0.42	0.10		-0.20	-0.13	0.43	-0.23
40	MC	1	65061	0.35	0.53	0.00		0.27	0.35	0.08	0.30		-0.26	0.53	-0.11	-0.22
41	TE	1	64861	0.41	0.54	0.01	0.59	0.40				-0.52	0.54			
42	MC	1	65199	0.39	0.28	0.00		0.13	0.17	0.31	0.38		-0.16	-0.17	-0.04	0.28
43	MC	1	65209	0.33	0.22	0.00		0.23	0.29	0.33	0.15		-0.22	-0.05	0.22	0.03
44	TE	1	65189	0.14	0.17	0.00	0.86	0.14				-0.17	0.18			
45	MC	1	65208	0.43	0.56	0.00		0.10	0.15	0.32	0.43		-0.24	-0.23	-0.26	0.56
46	MC	1	65196	0.85	0.34	0.00		0.06	0.85	0.04	0.06		-0.19	0.34	-0.15	-0.21

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-11. Item Statistics, Mathematics Grade 7

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	SA	1	63727	0.47	0.62	0.00	0.53	0.46				-0.61	0.62			
2	MC	1	63797	0.45	0.53	0.00		0.31	0.45	0.22	0.02		-0.22	0.53	-0.35	-0.13
3	MC	1	63779	0.50	0.34	0.00		0.50	0.14	0.22	0.14		0.34	-0.11	-0.21	-0.13
4	SA	1	63666	0.27	0.46	0.00	0.73	0.27				-0.45	0.46			
5	MC	1	63762	0.49	0.42	0.00		0.07	0.49	0.14	0.30		-0.14	0.42	-0.28	-0.16
6	MC	1	63756	0.42	0.44	0.00		0.42	0.08	0.29	0.21		0.44	-0.25	-0.20	-0.14
7	MC	1	63784	0.50	0.44	0.00		0.38	0.05	0.07	0.50		-0.29	-0.15	-0.18	0.44
8	TE	1	63765	0.20	0.43	0.00	0.80	0.20				-0.42	0.43			
9	MC	1	63765	0.31	0.25	0.00		0.32	0.12	0.24	0.31		0.07	-0.24	-0.17	0.25
10	MC	1	63756	0.49	0.39	0.00		0.49	0.20	0.25	0.06		0.39	-0.33	-0.05	-0.19
11	MC	1	63761	0.62	0.20	0.00		0.19	0.61	0.10	0.10		-0.09	0.20	-0.19	0.00
12	MC	1	63712	0.72	0.25	0.00		0.17	0.05	0.06	0.71		-0.10	-0.14	-0.17	0.25
13	MC	1	63651	0.42	0.38	0.00		0.26	0.14	0.42	0.18		-0.20	-0.18	0.39	-0.10
14	MC	1	63702	0.46	0.41	0.00		0.46	0.17	0.20	0.16		0.41	-0.14	-0.08	-0.32
15	SA	1	63323	0.29	0.43	0.01	0.70	0.29				-0.41	0.43			
16	MC	1	63670	0.28	0.49	0.00		0.28	0.13	0.25	0.34		0.49	-0.11	-0.06	-0.33
17	MC	1	63662	0.22	0.14	0.00		0.28	0.24	0.26	0.22		0.00	-0.11	-0.03	0.14
18	MC	1	63694	0.42	0.21	0.00		0.04	0.18	0.42	0.36		-0.09	-0.05	0.21	-0.13
19	TE	1	63474	0.38	0.61	0.01	0.62	0.38				-0.59	0.61			
20	MC	1	63667	0.37	0.28	0.00		0.27	0.26	0.36	0.10		-0.14	-0.13	0.28	-0.03
21	MC	1	63658	0.60	0.26	0.00		0.24	0.11	0.60	0.06		-0.03	-0.28	0.26	-0.13
22	MC	1	63669	0.24	0.19	0.00		0.27	0.33	0.16	0.24		-0.09	-0.05	-0.03	0.19
23	MC	1	63549	0.29	0.22	0.00		0.37	0.28	0.29	0.06		-0.06	-0.13	0.23	-0.05

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	SA	1	63369	0.38	0.48	0.01	0.61	0.38				-0.47	0.49			
25	SA	1	63225	0.20	0.51	0.01	0.79	0.20				-0.48	0.51			
26	MC	1	63565	0.40	0.30	0.00		0.16	0.27	0.16	0.40		-0.16	-0.13	-0.08	0.31
27	MC	1	63521	0.51	0.48	0.01		0.51	0.14	0.15	0.20		0.48	-0.29	-0.24	-0.12
28	MC	1	63611	0.78	0.45	0.00		0.05	0.07	0.10	0.77		-0.18	-0.24	-0.28	0.46
29	MC	1	63610	0.33	0.27	0.00		0.28	0.22	0.33	0.17		-0.07	-0.08	0.27	-0.15
30	MC	1	63529	0.40	0.44	0.00		0.14	0.12	0.33	0.40		-0.18	-0.18	-0.19	0.44
31	MC	1	63619	0.61	0.56	0.00		0.09	0.15	0.15	0.61		-0.14	-0.36	-0.29	0.56
32	TE	1	63418	0.66	0.17	0.01	0.34	0.65				-0.17	0.18			
33	TE	1	63584	0.12	0.52	0.00	0.88	0.12				-0.50	0.52			
34	MC	1	63533	0.56	0.34	0.00		0.12	0.56	0.18	0.14		-0.22	0.34	-0.24	-0.01
35	SA	1	63323	0.40	0.66	0.01	0.59	0.40				-0.64	0.66			
36	TE	1	63470	0.71	0.48	0.01	0.29	0.70				-0.47	0.48			
37	MC	1	63503	0.51	0.50	0.01		0.50	0.21	0.20	0.08		0.50	-0.29	-0.15	-0.25
38	TE	1	63254	0.45	0.30	0.01	0.55	0.44				-0.28	0.30			
39	MC	1	63457	0.55	0.27	0.01		0.03	0.54	0.37	0.05		-0.15	0.27	-0.12	-0.20
40	MC	1	63469	0.30	0.22	0.01		0.15	0.30	0.29	0.26		0.00	0.22	0.04	-0.25
41	TE	1	63491	0.27	0.61	0.01	0.72	0.27				-0.60	0.62			
42	TE	1	63087	0.23	0.59	0.01	0.77	0.22				-0.55	0.59			
43	MC	1	63519	0.50	0.53	0.01		0.12	0.16	0.22	0.49		-0.20	-0.26	-0.24	0.53
44	MC	1	63524	0.50	0.46	0.01		0.50	0.19	0.24	0.07		0.46	-0.27	-0.17	-0.18
45	MC	1	63544	0.65	0.39	0.00		0.11	0.64	0.15	0.09		-0.18	0.39	-0.22	-0.16
46	SA	1	63463	0.67	0.56	0.01	0.33	0.66				-0.55	0.56			

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-12. Item Statistics, Mathematics Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	62931	0.66	0.46	0.00		0.15	0.66	0.07	0.12		-0.23	0.46	-0.21	-0.25
2	MC	1	62908	0.24	0.16	0.00		0.13	0.21	0.24	0.42		0.01	-0.11	0.17	-0.06
3	MC	1	62891	0.55	0.40	0.00		0.22	0.10	0.55	0.14		-0.15	-0.17	0.40	-0.25
4	MC	1	62904	0.45	0.21	0.00		0.45	0.17	0.25	0.13		0.21	-0.02	-0.15	-0.09
5	SA	1	62807	0.20	0.51	0.00	0.80	0.20				-0.51	0.51			
6	MC	1	62881	0.34	0.42	0.00		0.37	0.15	0.15	0.34		-0.14	-0.17	-0.20	0.42
7	MC	1	62922	0.49	0.58	0.00		0.28	0.16	0.49	0.07		-0.46	-0.20	0.58	-0.04
8	MC	1	62922	0.49	0.48	0.00		0.10	0.29	0.49	0.12		-0.26	-0.15	0.48	-0.28
9	TE	1	62853	0.26	0.55	0.00	0.73	0.26				-0.54	0.55			
10	MC	1	62877	0.41	0.30	0.00		0.41	0.29	0.14	0.16		0.30	-0.17	-0.13	-0.06
11	TE	1	62755	0.24	0.50	0.00	0.76	0.24				-0.49	0.50			
12	MC	1	62890	0.47	0.30	0.00		0.17	0.18	0.47	0.17		-0.08	-0.25	0.30	-0.06
13	SA	1	62805	0.46	0.61	0.00	0.54	0.46				-0.60	0.61			
14	MC	1	62814	0.48	0.25	0.00		0.26	0.48	0.16	0.11		-0.12	0.25	-0.06	-0.16
15	TE	1	62788	0.46	0.49	0.00	0.54	0.46				-0.48	0.49			
16	MC	1	62751	0.47	0.26	0.00		0.09	0.47	0.22	0.21		-0.12	0.27	-0.27	0.04
17	SA	1	62373	0.37	0.53	0.01	0.63	0.36				-0.51	0.53			
18	MC	1	62841	0.56	0.41	0.00		0.12	0.17	0.56	0.15		-0.16	-0.26	0.41	-0.14
19	MC	1	62804	0.47	0.39	0.00		0.10	0.47	0.37	0.06		-0.14	0.40	-0.26	-0.12
20	MC	1	62827	0.48	0.42	0.00		0.48	0.18	0.21	0.13		0.42	-0.18	-0.17	-0.20
21	SA	1	62509	0.29	0.59	0.01	0.71	0.29				-0.58	0.59			
22	MC	1	62802	0.73	0.39	0.00		0.11	0.73	0.11	0.04		-0.23	0.39	-0.21	-0.15
23	TE	1	62354	0.57	0.51	0.01	0.42	0.57				-0.50	0.52			

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
24	MC	1	62731	0.54	0.38	0.00		0.54	0.24	0.15	0.07		0.38	-0.15	-0.23	-0.15
25	TE	1	62484	0.20	0.47	0.01	0.80	0.19				-0.44	0.47			
26	MC	1	62723	0.25	0.21	0.00		0.10	0.36	0.25	0.29		-0.18	-0.03	0.21	-0.05
27	SA	1	62277	0.17	0.44	0.01	0.82	0.17				-0.41	0.44			
28	MC	1	62696	0.57	0.54	0.00		0.06	0.14	0.23	0.57		-0.21	-0.23	-0.32	0.54
29	MC	1	62749	0.42	0.42	0.00		0.12	0.26	0.21	0.42		-0.16	-0.16	-0.20	0.42
30	MC	1	62700	0.30	0.35	0.00		0.30	0.34	0.26	0.09		0.36	-0.08	-0.22	-0.09
31	MC	1	62716	0.35	0.44	0.00		0.13	0.18	0.33	0.34		-0.11	-0.15	-0.23	0.44
32	TE	1	62600	0.35	0.52	0.01	0.65	0.34				-0.50	0.52			
33	MC	1	62755	0.57	0.39	0.00		0.13	0.57	0.17	0.13		-0.12	0.39	-0.23	-0.19
34	MC	1	62657	0.37	0.22	0.00		0.23	0.09	0.37	0.31		-0.11	-0.23	0.23	0.02
35	TE	1	62560	0.24	0.65	0.01	0.76	0.23				-0.63	0.65			
36	MC	1	62665	0.53	0.48	0.00		0.14	0.53	0.23	0.09		-0.13	0.48	-0.34	-0.17
37	TE	1	62094	0.33	0.27	0.01	0.66	0.32				-0.25	0.28			
38	MC	1	62656	0.74	0.43	0.01		0.13	0.05	0.74	0.08		-0.23	-0.19	0.43	-0.24
39	MC	1	62655	0.44	0.42	0.01		0.44	0.17	0.22	0.16		0.42	-0.27	-0.17	-0.09
40	MC	1	62691	0.21	0.30	0.00		0.21	0.35	0.22	0.21		-0.10	-0.06	-0.12	0.30
41	TE	1	62617	0.36	0.37	0.01	0.63	0.36				-0.35	0.37			
42	MC	1	62715	0.57	0.49	0.00		0.57	0.19	0.17	0.07		0.49	-0.16	-0.33	-0.20
43	MC	1	62688	0.67	0.43	0.00		0.08	0.67	0.17	0.08		-0.25	0.44	-0.20	-0.21
44	MC	1	62702	0.54	0.34	0.00		0.09	0.19	0.54	0.18		-0.14	-0.13	0.34	-0.20
45	MC	1	62714	0.49	0.38	0.00		0.08	0.49	0.32	0.10		-0.11	0.38	-0.24	-0.15
46	MC	1	62731	0.70	0.40	0.00		0.08	0.10	0.70	0.11		-0.21	-0.20	0.41	-0.20

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-13. Item Statistics, Science Grade 4

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	63466	0.52	0.31	0.00		0.30	0.11	0.52	0.08		-0.08	-0.26	0.31	-0.14
2	TE	1	63473	0.74	0.51	0.00	0.26	0.74				-0.51	0.51			
3	TE	1	63438	0.32	0.28	0.00	0.68	0.32				-0.27	0.28			
4	TE	1	63418	0.44	0.42	0.00	0.56	0.44				-0.41	0.42			
5	TE	1	63417	0.52	0.38	0.00	0.48	0.52				-0.37	0.38			
6	TE	1	63447	0.64	0.50	0.00	0.36	0.64				-0.49	0.50			
7	MC	1	63427	0.37	0.28	0.00		0.37	0.26	0.32	0.05		0.28	-0.07	-0.13	-0.19
8	TE	1	63409	0.59	0.43	0.00	0.41	0.59				-0.42	0.43			
9	EBSR	1	63451	0.42	0.43	0.00	0.58	0.42				-0.43	0.43			
10	MC	1	63464	0.66	0.39	0.00		0.18	0.66	0.07	0.09		-0.20	0.39	-0.24	-0.16
11	MC	1	63463	0.51	0.40	0.00		0.18	0.14	0.51	0.17		-0.23	-0.23	0.40	-0.08
12	MC	1	63432	0.54	0.34	0.00		0.18	0.16	0.54	0.13		-0.11	-0.23	0.34	-0.13
13	TE	1	63438	0.70	0.39	0.00	0.30	0.70				-0.39	0.39			
14	TE	1	63355	0.64	0.45	0.00	0.35	0.64				-0.45	0.45			
15	TE	1	63438	0.40	0.58	0.00	0.60	0.40				-0.57	0.58			
16	MC	1	63476	0.63	0.33	0.00		0.11	0.15	0.11	0.63		-0.10	-0.25	-0.12	0.33
17	TE	1	63455	0.84	0.38	0.00	0.16	0.84				-0.38	0.38			
18	TE	1	63449	0.69	0.40	0.00	0.31	0.69				-0.40	0.40			
19	TE	1	63317	0.51	0.21	0.00	0.49	0.51				-0.21	0.21			
20	TE	1	63433	0.88	0.31	0.00	0.12	0.88				-0.31	0.31			

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	TE	1	63371	0.53	0.43	0.00	0.47	0.52				-0.43	0.43			
22	TE	1	63428	0.50	0.56	0.00	0.50	0.50				-0.56	0.56			
23	MC	1	63386	0.61	0.43	0.00		0.11	0.61	0.23	0.05		-0.21	0.43	-0.25	-0.19
24	TE	1	63430	0.43	0.45	0.00	0.57	0.42				-0.44	0.45			
25	TE	1	63386	0.57	0.54	0.00	0.43	0.57				-0.54	0.54			
26	TE	1	63484	0.75	0.41	0.00	0.25	0.75				-0.41	0.41			
27	TE	1	63454	0.63	0.20	0.00	0.37	0.63				-0.20	0.21			
28	EBSR	1	63469	0.35	0.30	0.00	0.65	0.35				-0.30	0.30			
29	EBSR	1	63424	0.45	0.47	0.00	0.55	0.45				-0.47	0.47			
30	TE	1	63442	0.66	0.45	0.00	0.34	0.66				-0.45	0.45			
31	TE	1	63397	0.62	0.31	0.00	0.38	0.62				-0.31	0.32			
32	MC	1	63438	0.58	0.36	0.00		0.58	0.23	0.11	0.08		0.36	-0.14	-0.25	-0.15
33	TE	1	63459	0.74	0.40	0.00	0.26	0.74				-0.40	0.40			
34	TE	1	63448	0.90	0.35	0.00	0.10	0.90				-0.35	0.35			
35	MC	1	63445	0.37	0.23	0.00		0.30	0.17	0.37	0.16		-0.05	-0.24	0.23	0.01
36	EBSR	1	63463	0.43	0.29	0.00	0.57	0.43				-0.28	0.29			
37	MC	1	63398	0.51	0.40	0.00		0.51	0.15	0.20	0.13		0.40	-0.24	-0.17	-0.12
38	TE	1	63417	0.78	0.46	0.00	0.22	0.78				-0.46	0.46			
39	MS	1	63428	0.23	0.39	0.00	0.77	0.23				-0.39	0.39			
40	TE	1	63451	0.21	0.36	0.00	0.79	0.21				-0.36	0.36			

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table G-14. Item Statistics, Science Grade 8

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	TE	1	62756	0.65	0.43	0.00	0.35	0.65				-0.43	0.43			
2	TE	1	62760	0.54	0.46	0.00	0.46	0.54				-0.46	0.46			
3	TE	1	62752	0.55	0.57	0.00	0.45	0.54				-0.57	0.57			
4	TE	1	62736	0.61	0.51	0.00	0.39	0.61				-0.51	0.51			
5	TE	1	62753	0.83	0.33	0.00	0.17	0.83				-0.33	0.33			
6	MC	1	62753	0.57	0.35	0.00		0.09	0.24	0.10	0.57		-0.22	-0.11	-0.20	0.35
7	TE	1	62772	0.53	0.36	0.00	0.47	0.53				-0.36	0.36			
8	MC	1	62728	0.62	0.40	0.00		0.13	0.62	0.19	0.06		-0.09	0.40	-0.31	-0.18
9	TE	1	62788	0.79	0.31	0.00	0.21	0.79				-0.31	0.31			
10	TE	1	62799	0.49	0.33	0.00	0.51	0.49				-0.33	0.33			
11	TE	1	62769	0.63	0.41	0.00	0.37	0.63				-0.41	0.41			
12	MC	1	62783	0.53	0.31	0.00		0.10	0.12	0.53	0.25		-0.11	-0.28	0.31	-0.07
13	TE	1	62722	0.36	0.41	0.00	0.64	0.36				-0.40	0.41			
14	TE	1	62731	0.70	0.36	0.00	0.30	0.70				-0.36	0.36			
15	EBSR	1	62792	0.54	0.30	0.00	0.46	0.54				-0.29	0.30			
16	TE	1	62828	0.40	0.45	0.00	0.60	0.40				-0.45	0.46			
17	TE	1	62722	0.43	0.38	0.00	0.57	0.43				-0.37	0.38			
18	TE	1	62792	0.55	0.57	0.00	0.45	0.55				-0.57	0.57			
19	TE	1	62738	0.30	0.37	0.00	0.70	0.30				-0.37	0.37			
20	TE	1	62753	0.52	0.35	0.00	0.48	0.52				-0.34	0.35			

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	TE	1	62658	0.82	0.41	0.00	0.18	0.82				-0.41	0.42			
22	TE	1	62763	0.73	0.44	0.00	0.26	0.73				-0.44	0.44			
23	TE	1	62714	0.59	0.46	0.00	0.41	0.59				-0.46	0.46			
24	TE	1	62779	0.50	0.12	0.00	0.50	0.50				-0.12	0.13			
25	TE	1	62794	0.54	0.46	0.00	0.46	0.54				-0.46	0.46			
26	EBSR	1	62851	0.36	0.51	0.00	0.63	0.36				-0.51	0.51			
27	TE	1	62771	0.51	0.53	0.00	0.49	0.51				-0.53	0.53			
28	TE	1	62743	0.59	0.21	0.00	0.41	0.59				-0.20	0.21			
29	TE	1	62768	0.64	0.43	0.00	0.36	0.64				-0.43	0.43			
30	MC	1	62753	0.52	0.44	0.00		0.14	0.19	0.52	0.15		-0.26	-0.28	0.44	-0.04
31	EBSR	1	62827	0.46	0.48	0.00	0.54	0.46				-0.48	0.48			
32	TE	1	62766	0.73	0.39	0.00	0.27	0.73				-0.38	0.39			
33	TE	1	62565	0.42	0.37	0.00	0.58	0.41				-0.36	0.37			
34	TE	1	62743	0.43	0.37	0.00	0.57	0.43				-0.36	0.37			
35	TE	1	62774	0.29	0.37	0.00	0.71	0.29				-0.37	0.38			
36	MS	1	62724	0.30	0.44	0.00	0.70	0.29				-0.43	0.44			
37	TE	1	62765	0.42	0.49	0.00	0.58	0.42				-0.49	0.49			
38	MC	1	62754	0.69	0.45	0.00		0.09	0.69	0.14	0.08		-0.20	0.45	-0.29	-0.18
39	TE	1	62725	0.56	0.44	0.00	0.44	0.55				-0.43	0.44			
40	MC	1	62760	0.45	0.29	0.00		0.15	0.25	0.45	0.14		-0.04	-0.16	0.29	-0.16

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	63519	0.80	0.49	0.00		0.08	0.05	0.07	0.80		-0.27	-0.26	-0.28	0.49
2	MC	1	63493	0.84	0.43	0.00		0.09	0.83	0.02	0.05		-0.28	0.43	-0.22	-0.21
3	MC	1	63497	0.80	0.38	0.00		0.80	0.04	0.10	0.05		0.38	-0.26	-0.18	-0.20
4	MC	1	63488	0.78	0.45	0.00		0.06	0.10	0.78	0.06		-0.26	-0.25	0.45	-0.21
5	MC	1	63468	0.73	0.41	0.00		0.05	0.07	0.14	0.73		-0.24	-0.23	-0.20	0.41
6	MC	1	63444	0.44	0.28	0.00		0.44	0.17	0.23	0.15		0.28	-0.10	-0.13	-0.12
7	TE	1	63499	0.71	0.37	0.00	0.29	0.71				-0.37	0.37			
8	MC	1	63481	0.59	0.33	0.00		0.19	0.18	0.04	0.59		-0.20	-0.11	-0.21	0.33
9	MC	1	63461	0.81	0.37	0.00		0.05	0.81	0.07	0.06		-0.18	0.37	-0.22	-0.19
10	MC	1	63472	0.68	0.43	0.00		0.10	0.10	0.11	0.68		-0.19	-0.24	-0.22	0.43
11	MC	1	63480	0.52	0.23	0.00		0.17	0.52	0.14	0.17		-0.08	0.23	-0.10	-0.13
12	MC	1	63486	0.78	0.45	0.00		0.08	0.78	0.07	0.06		-0.26	0.45	-0.29	-0.16
13	TE	1	63452	0.40	0.31	0.00	0.60	0.40				-0.31	0.32			
14	MC	1	63455	0.39	0.40	0.00		0.21	0.26	0.13	0.39		-0.19	-0.08	-0.24	0.40
15	MC	1	63483	0.70	0.38	0.00		0.17	0.07	0.06	0.70		-0.26	-0.18	-0.14	0.38
16	MC	1	63476	0.58	0.32	0.00		0.14	0.14	0.58	0.13		-0.06	-0.29	0.32	-0.10
17	MC	1	63494	0.68	0.46	0.00		0.06	0.07	0.19	0.68		-0.22	-0.22	-0.27	0.46
18	TE	1	63471	0.70	0.36	0.00	0.30	0.70				-0.36	0.36			
19	MC	1	63495	0.74	0.48	0.00		0.08	0.74	0.10	0.09		-0.30	0.48	-0.29	-0.16
20	MC	1	63496	0.64	0.33	0.00		0.64	0.22	0.08	0.06		0.33	-0.13	-0.23	-0.18

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	63482	0.70	0.43	0.00		0.07	0.05	0.69	0.18		-0.23	-0.25	0.43	-0.22
22	MC	1	63480	0.66	0.21	0.00		0.66	0.19	0.07	0.08		0.21	-0.05	-0.18	-0.13
23	MC	1	63465	0.83	0.54	0.00		0.04	0.05	0.07	0.83		-0.24	-0.28	-0.33	0.54
24	MC	1	63455	0.53	0.38	0.00		0.53	0.16	0.17	0.14		0.38	-0.25	-0.18	-0.09
25	MC	1	63455	0.65	0.52	0.00		0.08	0.18	0.65	0.10		-0.23	-0.28	0.52	-0.27
26	MC	1	63475	0.62	0.33	0.00		0.08	0.25	0.62	0.05		-0.29	-0.10	0.33	-0.17
27	MC	1	63472	0.62	0.34	0.00		0.20	0.09	0.62	0.09		-0.10	-0.22	0.34	-0.21
28	MC	1	63464	0.59	0.42	0.00		0.11	0.59	0.12	0.19		-0.17	0.42	-0.26	-0.17
29	MC	1	63465	0.52	0.37	0.00		0.09	0.29	0.09	0.52		-0.20	-0.13	-0.22	0.37
30	MC	1	63472	0.73	0.48	0.00		0.10	0.08	0.73	0.08		-0.25	-0.28	0.48	-0.22
31	TE	1	63482	0.53	0.40	0.00	0.47	0.53				-0.40	0.40			
32	MC	1	63471	0.67	0.48	0.00		0.67	0.12	0.12	0.09		0.48	-0.30	-0.21	-0.21
33	MC	1	63443	0.75	0.51	0.00		0.14	0.05	0.06	0.75		-0.23	-0.30	-0.31	0.51
34	MC	1	63458	0.72	0.49	0.00		0.72	0.17	0.07	0.04		0.49	-0.27	-0.30	-0.20
35	MC	1	63456	0.69	0.44	0.00		0.14	0.06	0.69	0.12		-0.26	-0.27	0.44	-0.15
36	MC	1	63464	0.77	0.53	0.00		0.06	0.77	0.12	0.05		-0.28	0.54	-0.33	-0.24
37	MC	1	63465	0.64	0.48	0.00		0.19	0.12	0.64	0.06		-0.22	-0.27	0.48	-0.24
38	MC	1	63481	0.70	0.45	0.00		0.70	0.04	0.07	0.19		0.45	-0.26	-0.26	-0.22

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	62930	0.82	0.48	0.00		0.09	0.04	0.04	0.82		-0.34	-0.23	-0.19	0.48
2	MC	1	62887	0.79	0.50	0.00		0.79	0.11	0.04	0.06		0.50	-0.37	-0.22	-0.19
3	MC	1	62891	0.85	0.52	0.00		0.85	0.04	0.06	0.05		0.52	-0.26	-0.32	-0.26
4	MC	1	62905	0.81	0.38	0.00		0.07	0.05	0.81	0.07		-0.20	-0.22	0.39	-0.20
5	MC	1	62886	0.83	0.51	0.00		0.04	0.83	0.06	0.07		-0.24	0.51	-0.31	-0.28
6	MC	1	62821	0.77	0.35	0.00		0.03	0.03	0.17	0.77		-0.18	-0.25	-0.19	0.35
7	MS	1	62876	0.59	0.59	0.00	0.41	0.59				-0.59	0.59			
8	MC	1	62842	0.66	0.54	0.00		0.66	0.10	0.13	0.11		0.54	-0.29	-0.33	-0.19
9	MC	1	62875	0.53	0.37	0.00		0.14	0.28	0.05	0.53		-0.22	-0.11	-0.27	0.37
10	TE	1	62779	0.60	0.54	0.00	0.40	0.60				-0.53	0.54			
11	MC	1	62863	0.81	0.49	0.00		0.80	0.03	0.12	0.04		0.49	-0.21	-0.32	-0.25
12	MC	1	62865	0.72	0.58	0.00		0.08	0.09	0.10	0.72		-0.20	-0.29	-0.40	0.58
13	MC	1	62884	0.76	0.35	0.00		0.10	0.08	0.76	0.07		-0.10	-0.20	0.35	-0.25
14	MC	1	62794	0.52	0.36	0.00		0.15	0.24	0.10	0.52		-0.15	-0.09	-0.29	0.36
15	MC	1	62809	0.64	0.35	0.00		0.12	0.64	0.12	0.12		-0.09	0.35	-0.30	-0.12
16	MC	1	62781	0.62	0.36	0.00		0.05	0.62	0.23	0.10		-0.20	0.36	-0.13	-0.24
17	MC	1	62834	0.69	0.41	0.00		0.13	0.12	0.69	0.06		-0.16	-0.25	0.42	-0.23
18	MC	1	62836	0.62	0.44	0.00		0.14	0.15	0.62	0.08		-0.15	-0.27	0.44	-0.22
19	MC	1	62867	0.83	0.19	0.00		0.06	0.06	0.83	0.06		-0.12	-0.10	0.19	-0.08
20	MC	1	62859	0.71	0.33	0.00		0.04	0.71	0.04	0.21		-0.05	0.33	-0.23	-0.23

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	62873	0.92	0.36	0.00		0.04	0.02	0.01	0.92		-0.21	-0.23	-0.19	0.36
22	MC	1	62842	0.61	0.46	0.00		0.29	0.05	0.05	0.61		-0.26	-0.24	-0.24	0.46
23	TE	1	62779	0.62	0.41	0.00	0.38	0.62				-0.41	0.42			
24	MC	1	62775	0.76	0.39	0.00		0.11	0.76	0.07	0.06		-0.19	0.39	-0.29	-0.14
25	MC	1	62718	0.74	0.37	0.00		0.74	0.04	0.02	0.19		0.37	-0.27	-0.24	-0.18
26	MC	1	62810	0.71	0.48	0.00		0.14	0.71	0.07	0.08		-0.23	0.48	-0.33	-0.19
27	MC	1	62813	0.57	0.38	0.00		0.57	0.09	0.18	0.16		0.38	-0.28	-0.20	-0.09
28	MC	1	62751	0.77	0.37	0.00		0.09	0.07	0.77	0.08		-0.13	-0.18	0.37	-0.28
29	MC	1	62800	0.57	0.22	0.00		0.08	0.57	0.08	0.26		-0.16	0.23	-0.24	0.00
30	TE	1	62750	0.52	0.52	0.00	0.48	0.52				-0.51	0.52			
31	MC	1	62796	0.69	0.44	0.00		0.06	0.69	0.13	0.11		-0.20	0.45	-0.27	-0.21
32	TE	1	62778	0.55	0.42	0.00	0.45	0.55				-0.42	0.42			
33	MC	1	62816	0.50	0.32	0.00		0.50	0.08	0.21	0.21		0.33	-0.31	-0.15	-0.04
34	MC	1	62685	0.52	0.33	0.00		0.20	0.51	0.09	0.19		-0.12	0.33	-0.33	-0.05
35	MC	1	62780	0.62	0.44	0.00		0.62	0.11	0.15	0.12		0.44	-0.25	-0.25	-0.14
36	MC	1	62779	0.56	0.38	0.00		0.10	0.56	0.16	0.17		-0.09	0.38	-0.28	-0.15
37	MC	1	62774	0.51	0.32	0.00		0.27	0.51	0.12	0.10		-0.04	0.32	-0.28	-0.17
38	MC	1	62782	0.44	0.35	0.00		0.44	0.21	0.19	0.16		0.35	-0.10	-0.24	-0.10
39	MC	1	62809	0.83	0.48	0.00		0.03	0.83	0.08	0.06		-0.22	0.48	-0.28	-0.28
40	MC	1	62800	0.68	0.41	0.00		0.12	0.10	0.68	0.10		-0.13	-0.28	0.41	-0.22

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
1	MC	1	63124	0.79	0.42	0.00		0.79	0.10	0.05	0.07		0.42	-0.24	-0.27	-0.18
2	MC	1	63122	0.73	0.25	0.00		0.05	0.02	0.73	0.20		-0.15	-0.19	0.25	-0.12
3	MC	1	63102	0.68	0.39	0.00		0.68	0.08	0.13	0.11		0.39	-0.27	-0.19	-0.14
4	MC	1	63101	0.69	0.41	0.00		0.69	0.11	0.13	0.06		0.41	-0.29	-0.14	-0.20
5	MC	1	63079	0.74	0.46	0.00		0.03	0.74	0.06	0.16		-0.21	0.47	-0.28	-0.26
6	MC	1	62948	0.74	0.45	0.00		0.07	0.07	0.12	0.73		-0.18	-0.20	-0.30	0.45
7	MC	1	63083	0.80	0.47	0.00		0.05	0.06	0.80	0.08		-0.23	-0.29	0.48	-0.25
8	MC	1	63086	0.75	0.41	0.00		0.02	0.16	0.75	0.06		-0.18	-0.25	0.42	-0.24
9	MC	1	63045	0.63	0.42	0.00		0.62	0.18	0.13	0.07		0.42	-0.20	-0.24	-0.16
10	MC	1	63042	0.69	0.41	0.00		0.13	0.15	0.68	0.03		-0.16	-0.28	0.41	-0.19
11	MC	1	63009	0.59	0.41	0.00		0.11	0.59	0.25	0.05		-0.20	0.41	-0.18	-0.26
12	TE	1	62733	0.34	0.37	0.01	0.66	0.33				-0.35	0.37			
13	MC	1	62986	0.53	0.28	0.00		0.18	0.52	0.12	0.17		-0.03	0.29	-0.26	-0.11
14	MC	1	63000	0.56	0.43	0.00		0.56	0.14	0.20	0.10		0.44	-0.19	-0.24	-0.17
15	MC	1	62966	0.65	0.45	0.00		0.15	0.65	0.10	0.10		-0.15	0.45	-0.30	-0.23
16	MC	1	62874	0.63	0.35	0.00		0.11	0.08	0.63	0.18		-0.12	-0.25	0.36	-0.16
17	MC	1	62819	0.59	0.28	0.01		0.04	0.27	0.58	0.10		-0.19	-0.08	0.28	-0.20
18	MC	1	62693	0.72	0.35	0.01		0.05	0.12	0.72	0.10		-0.12	-0.21	0.36	-0.20
19	MC	1	62953	0.69	0.48	0.00		0.17	0.69	0.07	0.07		-0.27	0.48	-0.28	-0.17
20	MC	1	62924	0.62	0.45	0.00		0.62	0.14	0.16	0.08		0.45	-0.14	-0.30	-0.22

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
21	MC	1	62920	0.59	0.36	0.00		0.06	0.17	0.58	0.18		-0.28	-0.13	0.36	-0.15
22	MC	1	62927	0.51	0.45	0.00		0.11	0.27	0.11	0.51		-0.22	-0.14	-0.29	0.45
23	MC	1	62938	0.58	0.52	0.00		0.58	0.16	0.15	0.11		0.52	-0.18	-0.30	-0.25
24	TE	1	62380	0.29	0.43	0.01	0.70	0.28				-0.39	0.44			
25	MC	1	62890	0.59	0.60	0.00		0.12	0.14	0.15	0.59		-0.22	-0.32	-0.29	0.60
26	MC	1	62636	0.72	0.38	0.00		0.08	0.72	0.15	0.05		-0.22	0.38	-0.23	-0.12
27	MC	1	62585	0.71	0.31	0.00		0.05	0.07	0.71	0.16		-0.20	-0.23	0.32	-0.09
28	MC	1	62555	0.45	0.48	0.00		0.20	0.16	0.19	0.44		-0.12	-0.26	-0.23	0.48
29	MC	1	62483	0.41	0.37	0.01		0.16	0.23	0.20	0.40		-0.18	-0.02	-0.25	0.37
30	MC	1	62531	0.57	0.54	0.00		0.57	0.12	0.17	0.13		0.54	-0.27	-0.23	-0.26
31	MC	1	62496	0.72	0.46	0.01		0.72	0.08	0.15	0.05		0.47	-0.24	-0.27	-0.20
32	MC	1	62532	0.83	0.42	0.00		0.02	0.06	0.83	0.08		-0.18	-0.20	0.42	-0.29
33	MC	1	62466	0.60	0.49	0.01		0.60	0.17	0.13	0.09		0.49	-0.26	-0.27	-0.15
34	MC	1	62460	0.61	0.50	0.01		0.13	0.61	0.12	0.13		-0.18	0.50	-0.34	-0.19
35	MC	1	62367	0.71	0.50	0.01		0.08	0.13	0.71	0.07		-0.22	-0.27	0.50	-0.28
36	TE	1	62156	0.65	0.21	0.01	0.34	0.65				-0.19	0.22			
37	MC	1	62383	0.44	0.38	0.01		0.44	0.24	0.11	0.20		0.38	-0.12	-0.31	-0.08
38	MC	1	62379	0.59	0.46	0.01		0.13	0.13	0.14	0.59		-0.15	-0.23	-0.27	0.46
39	MC	1	62419	0.72	0.46	0.01		0.72	0.06	0.09	0.12		0.47	-0.30	-0.25	-0.17
40	MC	1	62379	0.69	0.48	0.01		0.11	0.69	0.11	0.09		-0.12	0.49	-0.34	-0.25

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Item Number	Item Type	Maximum Points	Number of Students	Item p-value	Item-Total Test Correlation	Proportion Omit	Proportion of Students					Item-Total Test Correlation				
							Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4	Score Point 0	Score Point 1 or Option 1	Score Point 2 or Option 2	Score Point 3 or Option 3	Score Point 4 or Option 4
41	MC	1	62363	0.78	0.47	0.01		0.10	0.06	0.77	0.06		-0.19	-0.28	0.48	-0.28
42	MC	1	62316	0.63	0.40	0.01		0.63	0.08	0.24	0.04		0.41	-0.32	-0.13	-0.23
43	MC	1	62300	0.68	0.45	0.01		0.03	0.09	0.19	0.67		-0.20	-0.18	-0.29	0.45
44	MC	1	62339	0.43	0.33	0.01		0.16	0.43	0.26	0.14		-0.19	0.34	-0.09	-0.14
45	TE	1	62079	0.28	0.38	0.01	0.71	0.28				-0.34	0.38			
46	MC	1	62304	0.55	0.40	0.01		0.15	0.54	0.22	0.08		-0.09	0.40	-0.24	-0.22
47	MC	1	62324	0.49	0.25	0.01		0.07	0.49	0.36	0.08		-0.17	0.26	-0.03	-0.23
48	MC	1	62332	0.68	0.53	0.01		0.09	0.67	0.12	0.12		-0.22	0.53	-0.31	-0.24
49	MC	1	62340	0.69	0.44	0.01		0.11	0.12	0.07	0.69		-0.30	-0.15	-0.20	0.44
50	MC	1	62316	0.86	0.35	0.01		0.08	0.05	0.85	0.02		-0.19	-0.24	0.36	-0.13

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix H

Wisconsin Standard Performance Index Score Computation

Copyright © 2019 by the Wisconsin Department of Public Instruction

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:

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

where

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

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

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

with

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

and

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

The SPI is defined to be the mean of this posterior distribution:

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

Following Novick and Jackson (1974, p. 119), a mastery band is created to be the *C*% central credibility interval for T_j. It is obtained by identifying the values that place $\frac{1}{2}(100 - C)\%$ of the $\beta(p_j, q_j)$ density in each tail of the distribution.

Estimation of the Prior Distribution of T_j

The *k* items in each test are scaled together using a generalized IRT model (3PL/2PPC) that fits a three-parameter logistic model (3PL) to the MC items and a generalized partial-credit model (2PPC) to the CR items (Yen, 1993).

The 3PL model is

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

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,

Copyright © 2019 by the Wisconsin Department of Public Instruction

& 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

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

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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

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:

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

and

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

The SPI is

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

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

The SPI can also be written in terms of the relative contribution of the prior estimate $\hat{T}_j$ and the observed proportion of maximum raw (correct score) (OPM), x_j / n_j, as

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

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

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

The term n_j^* may be interpreted as the contribution of the prior in terms of theoretical numbers of items.

<u>Check on Consistency and Adjustment of Weight Given to Prior Estimate</u>

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

difficult addition but not easy subtraction. Under these circumstances, it is not appropriate to pool the prior estimate, $\hat{T}_j$, with x_j / n_j. In calculating the SPI, the following statistic was used to identify examinees with unexpected performance on the strands in a test:

$$Q = \sum_{j=1}^{J} n_j (\frac{x_j}{n_j} - \hat{T}_j)^2 / (\hat{T}_j (1 - \hat{T}_j)) . \qquad (24)$$

If $Q \le \chi^2(J, .10)$, 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,

Copyright © 2019 by the Wisconsin Department of Public Instruction

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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix I

Conditional Standard Error of Measurement with Cut Scores

Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-1 CSEM with cut scores, ELA Grade 3

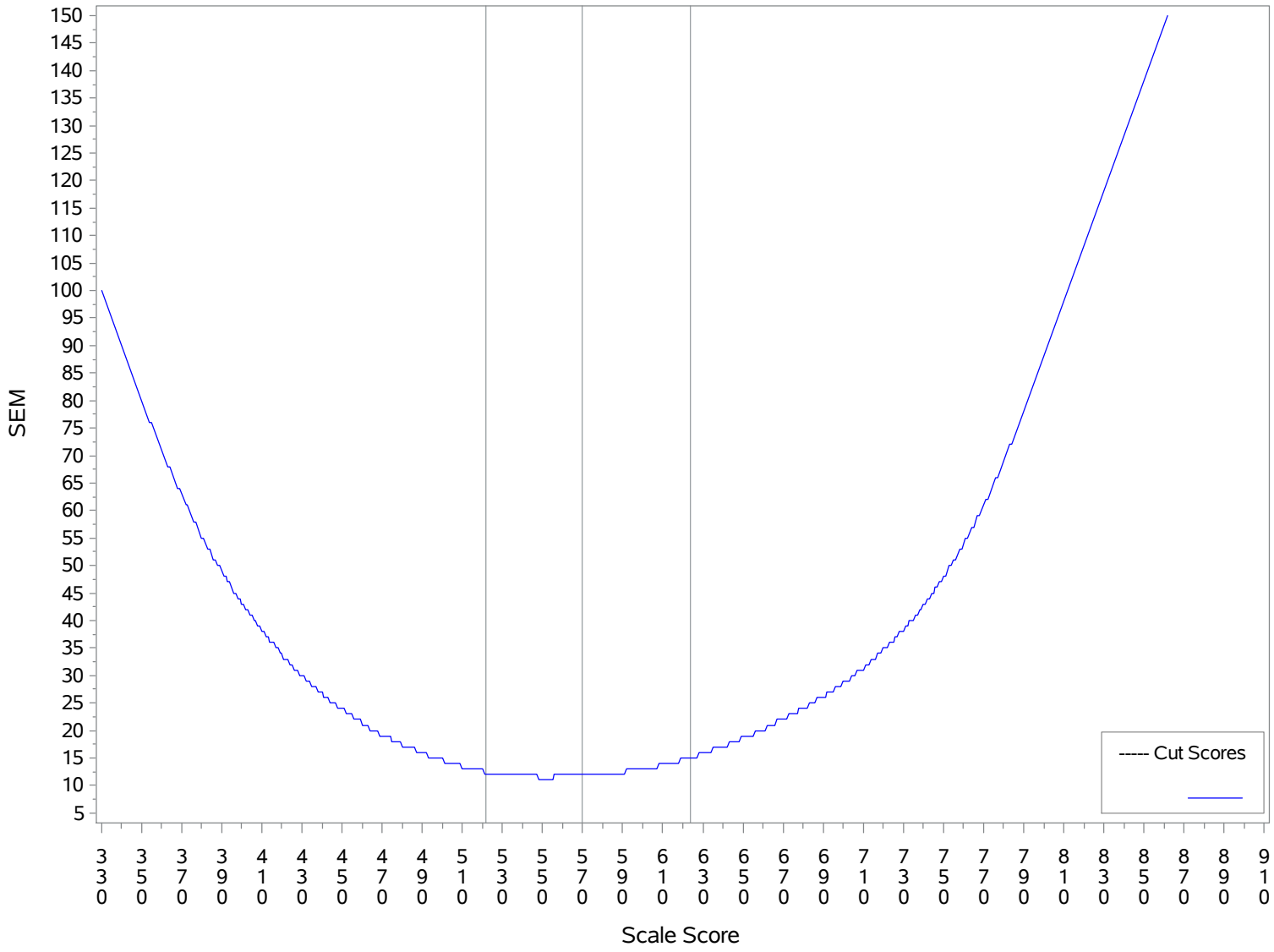


Figure I-2 CSEM with cut scores, ELA Grade 4

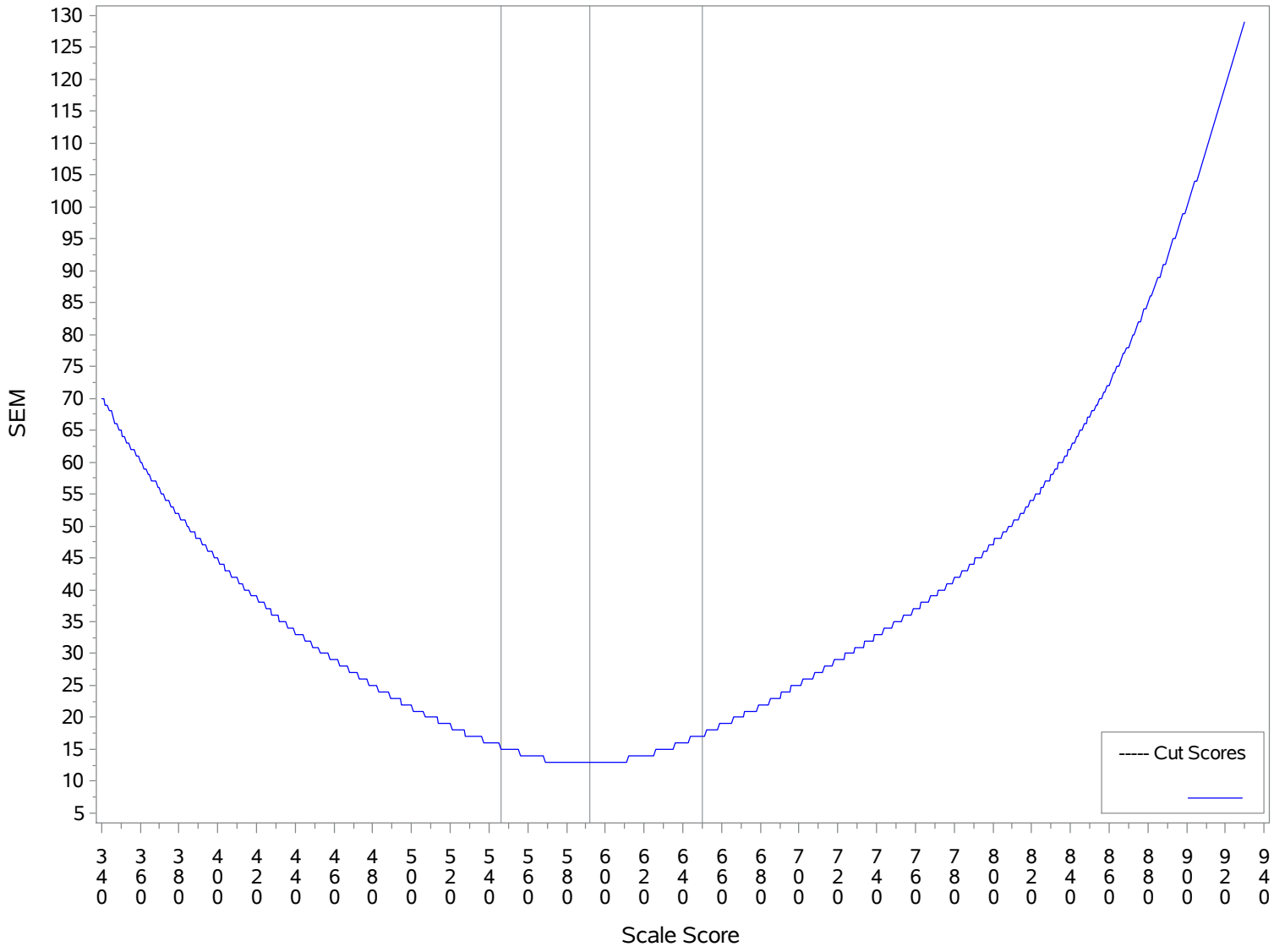
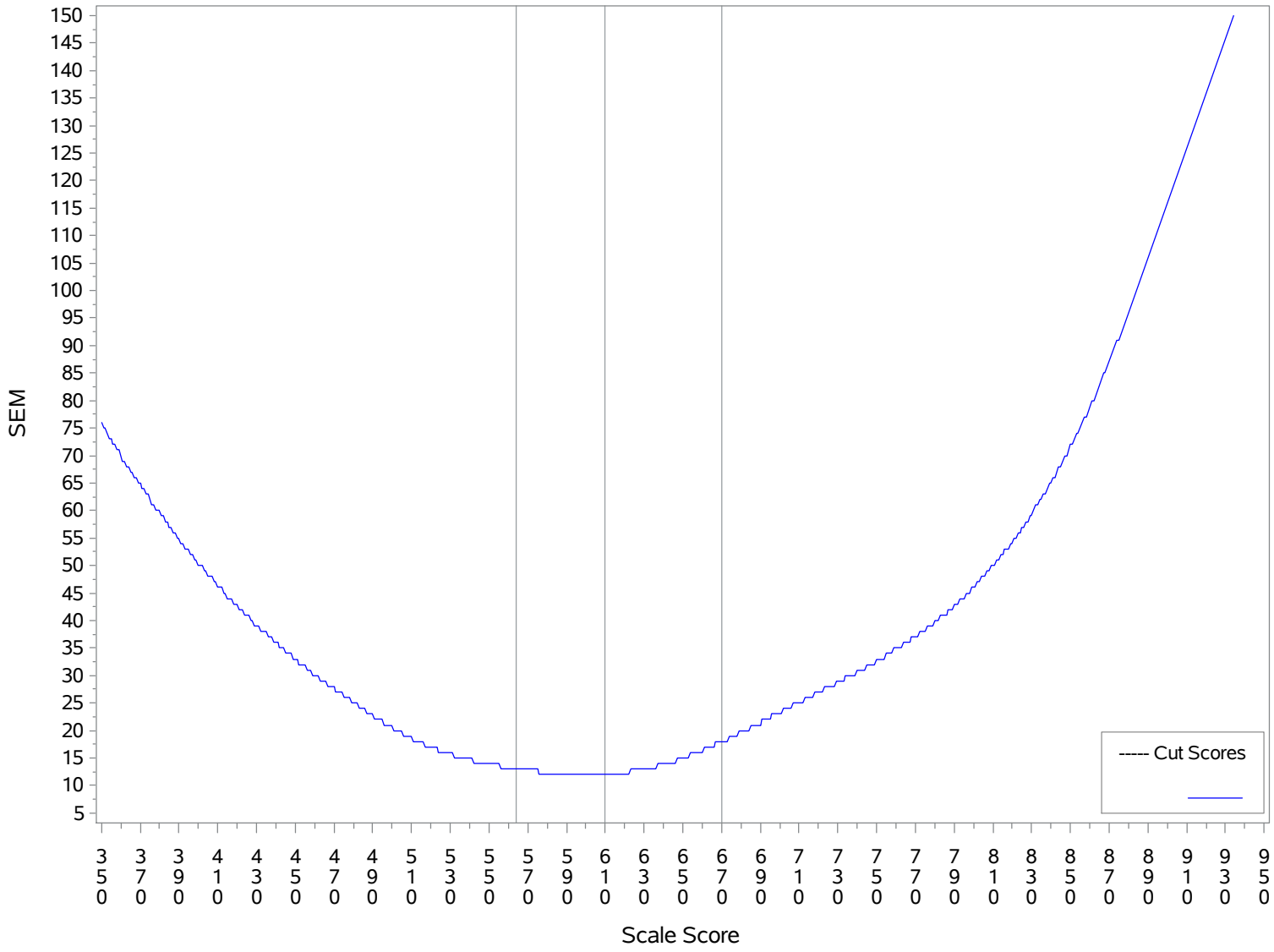


Figure I-3 CSEM with cut scores, ELA Grade 5



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-4 CSEM with cut scores, ELA Grade 6

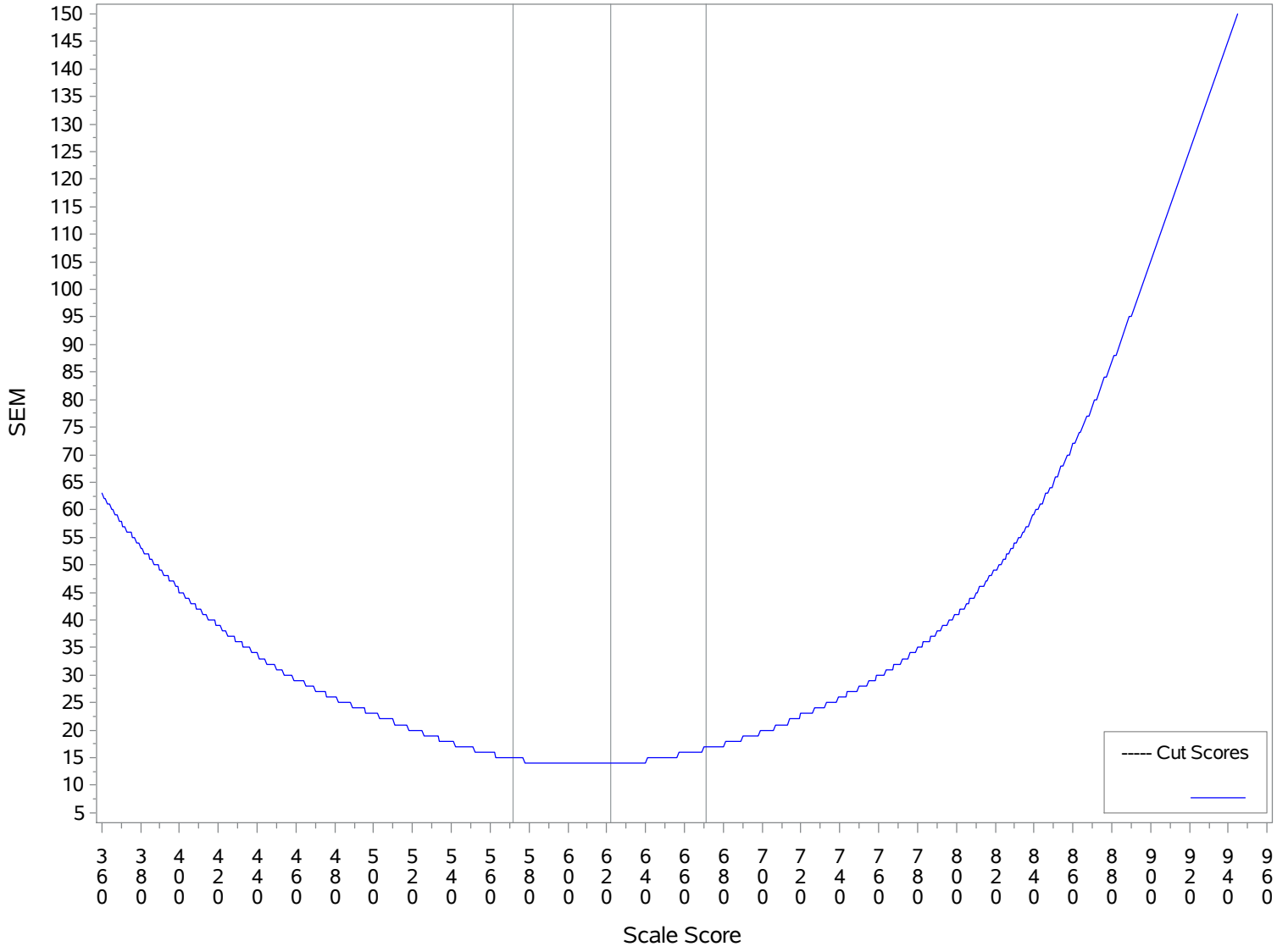


Figure I-5 CSEM with cut scores, ELA Grade 7

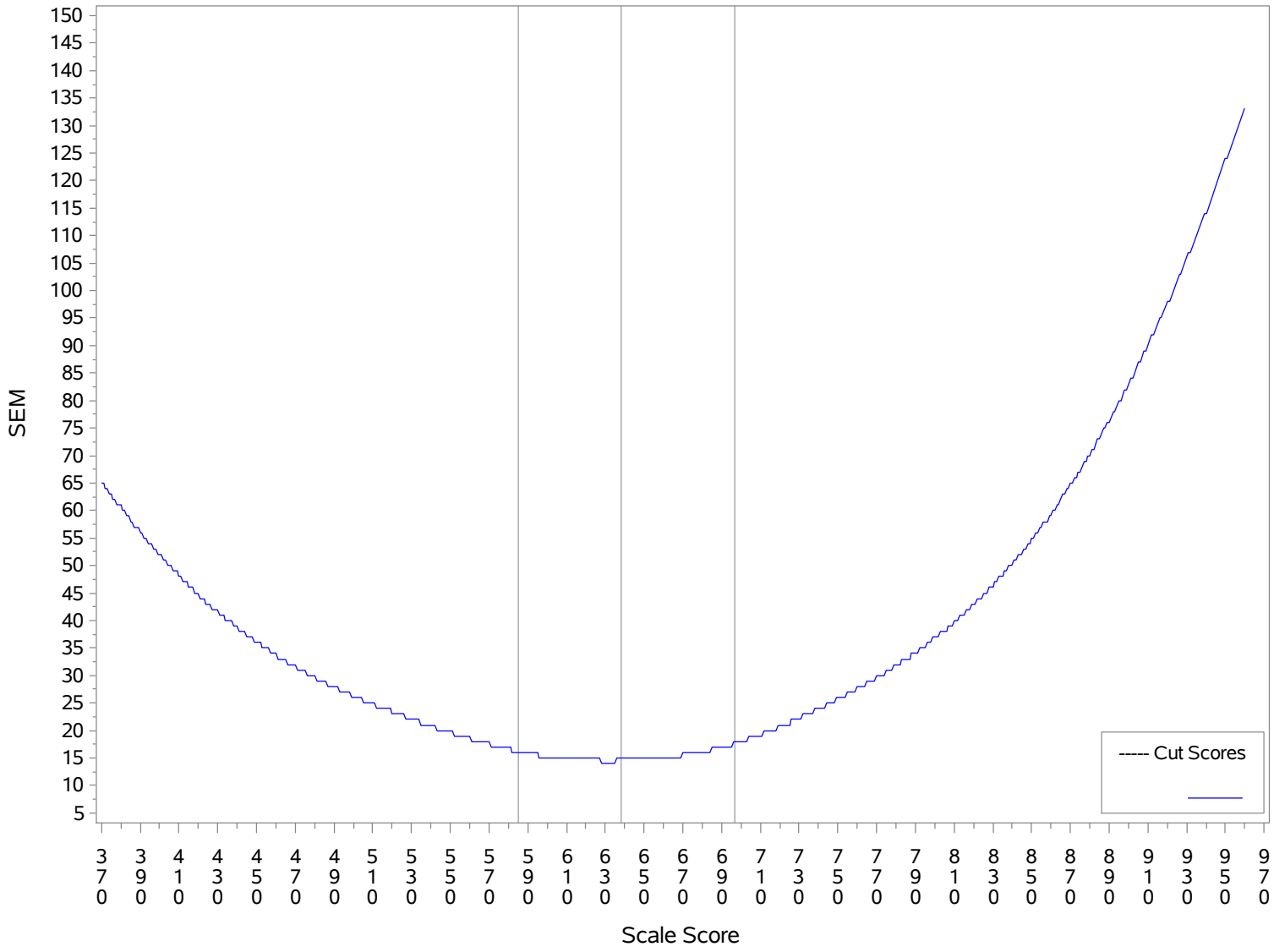
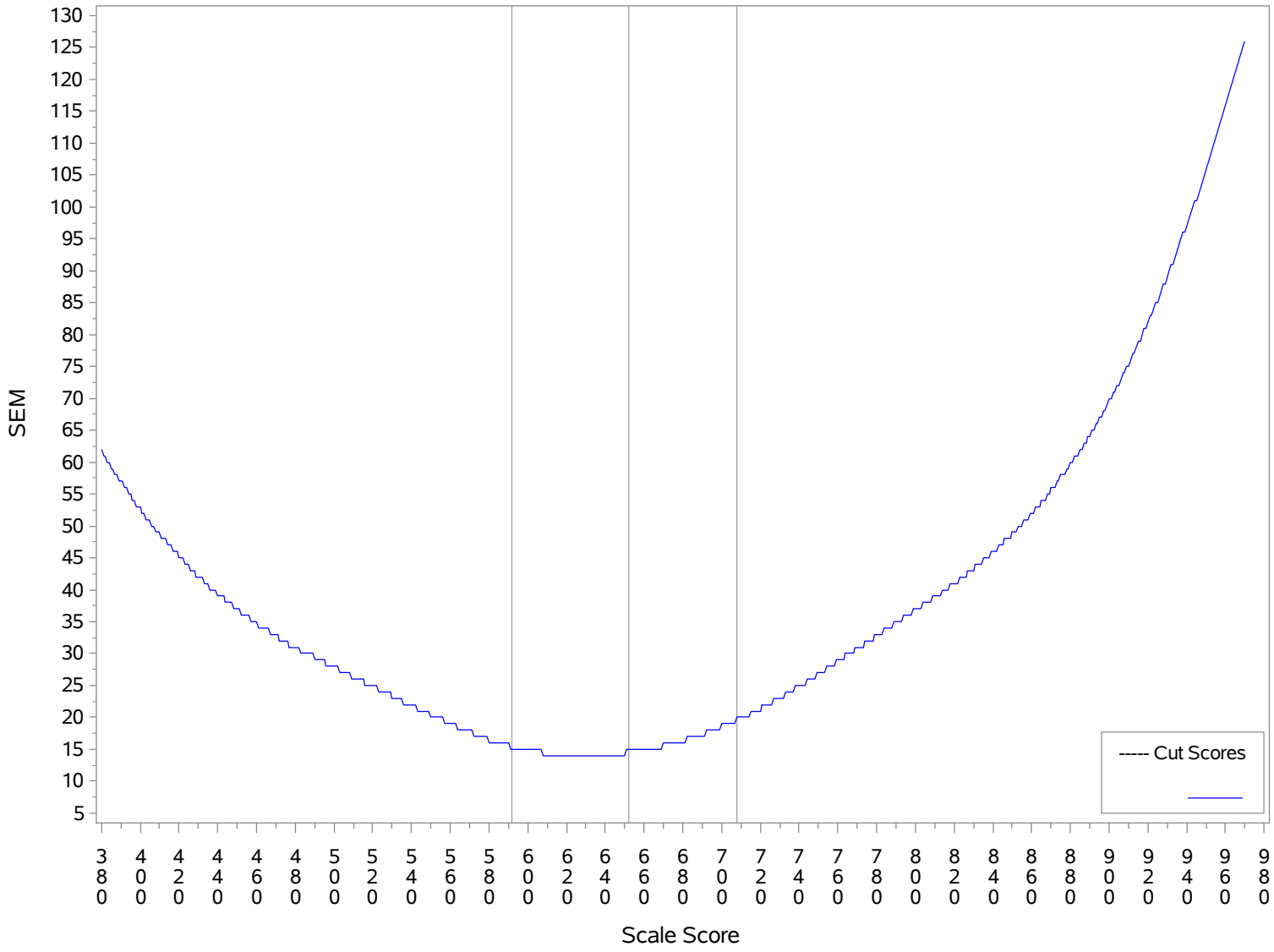
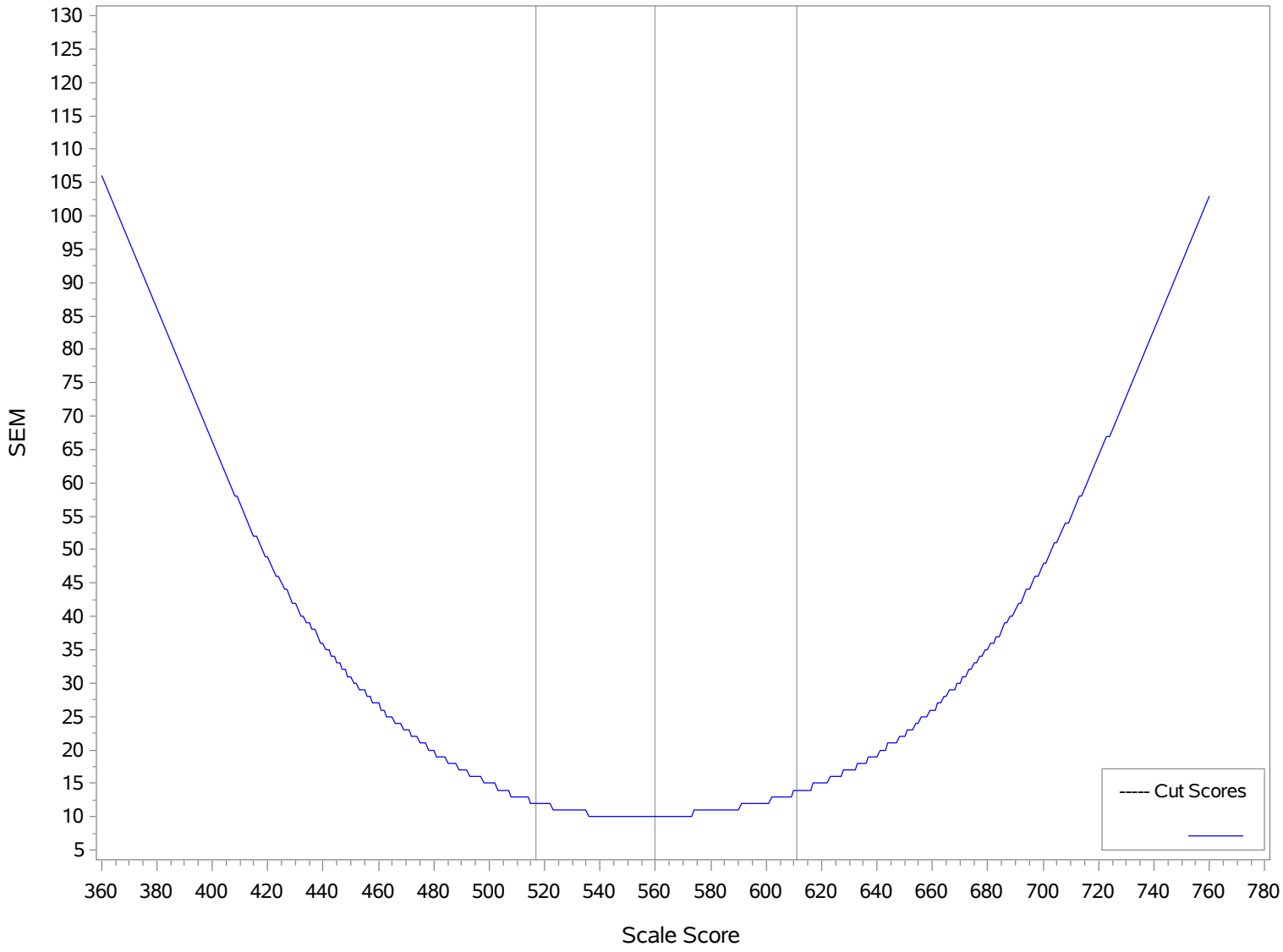


Figure I-6 CSEM with cut scores, ELA Grade 8



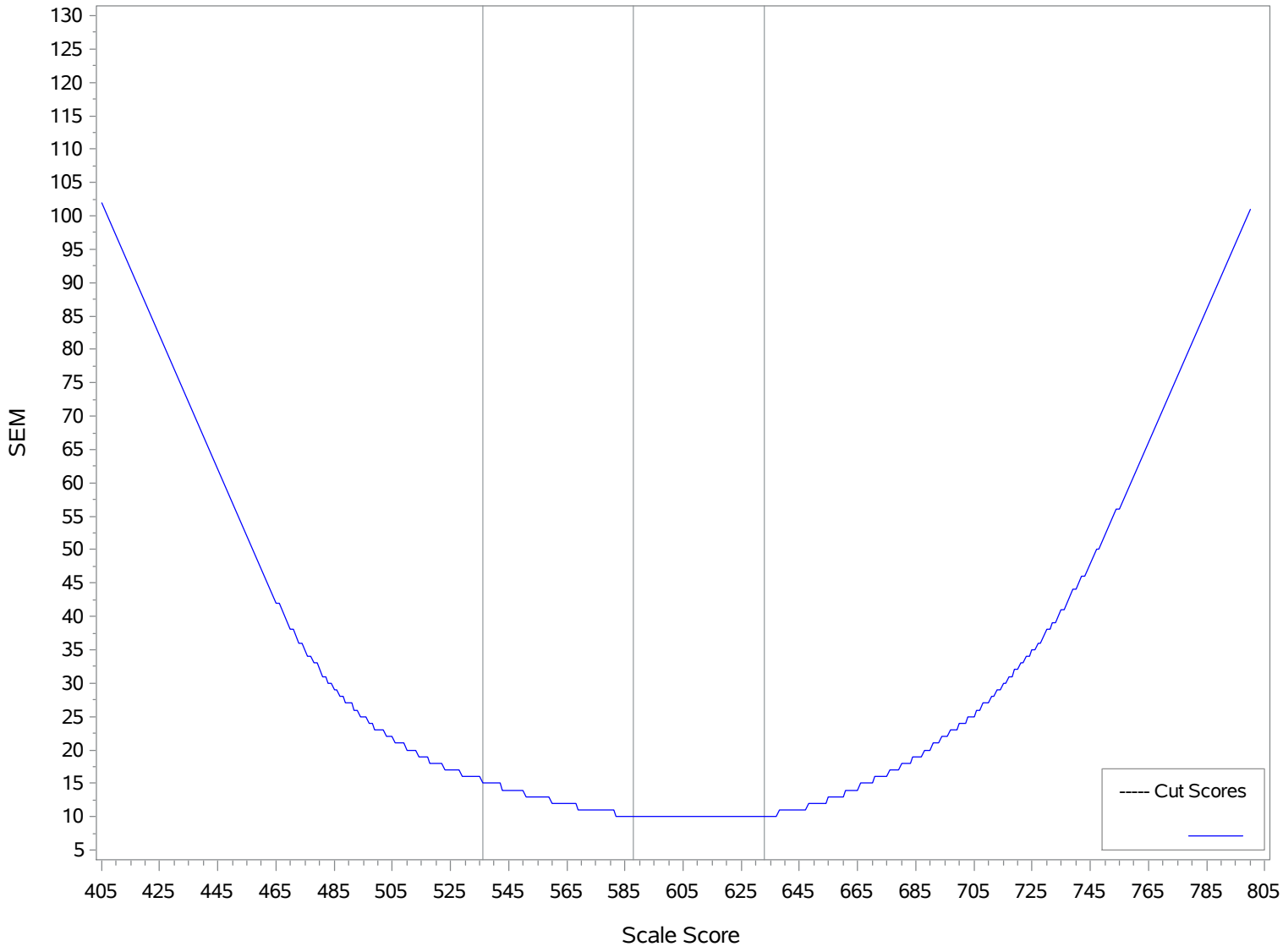
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-7 CSEM with cut scores, Mathematics Grade 3



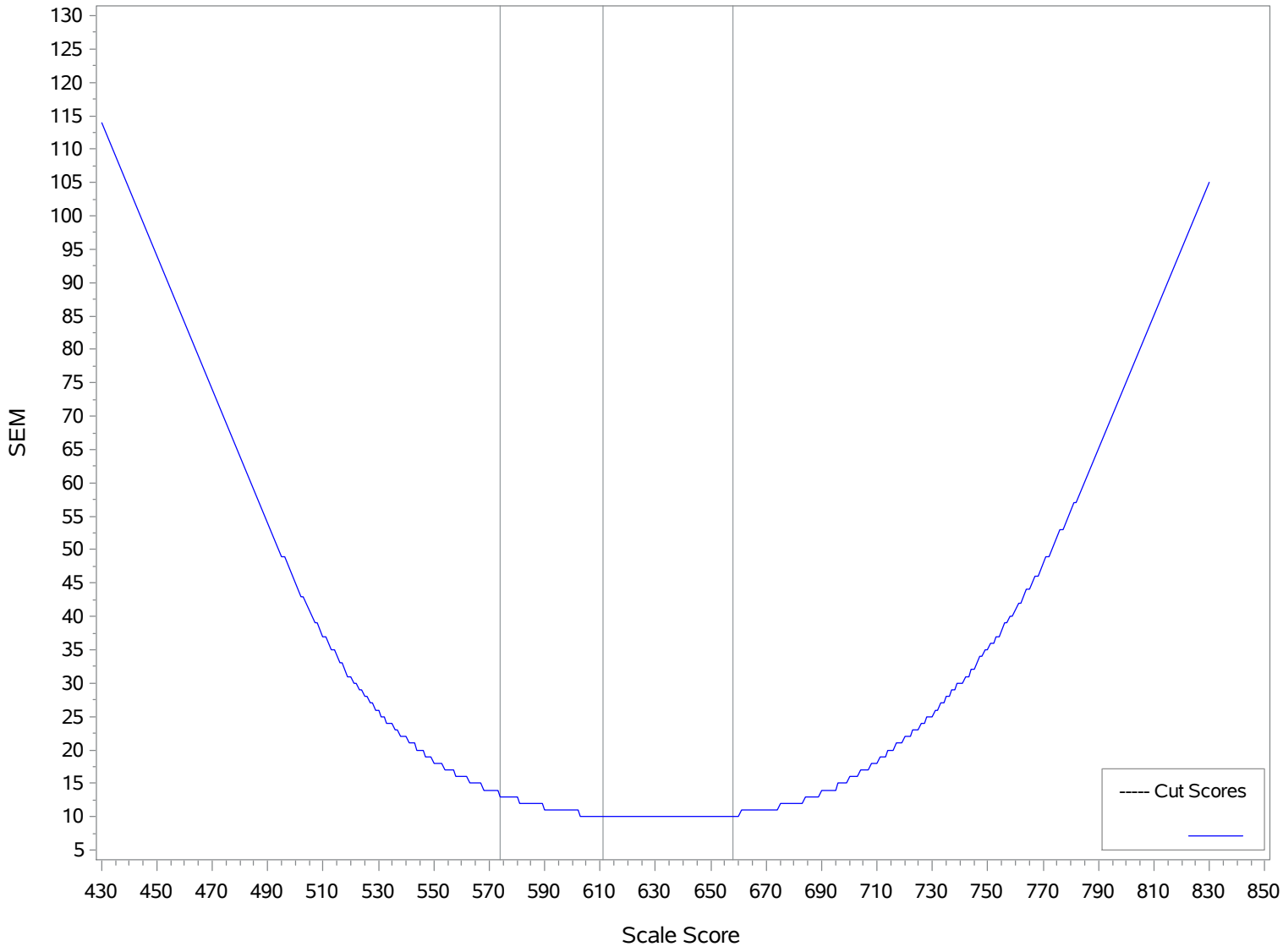
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-8 CSEM with cut scores, Mathematics Grade 4



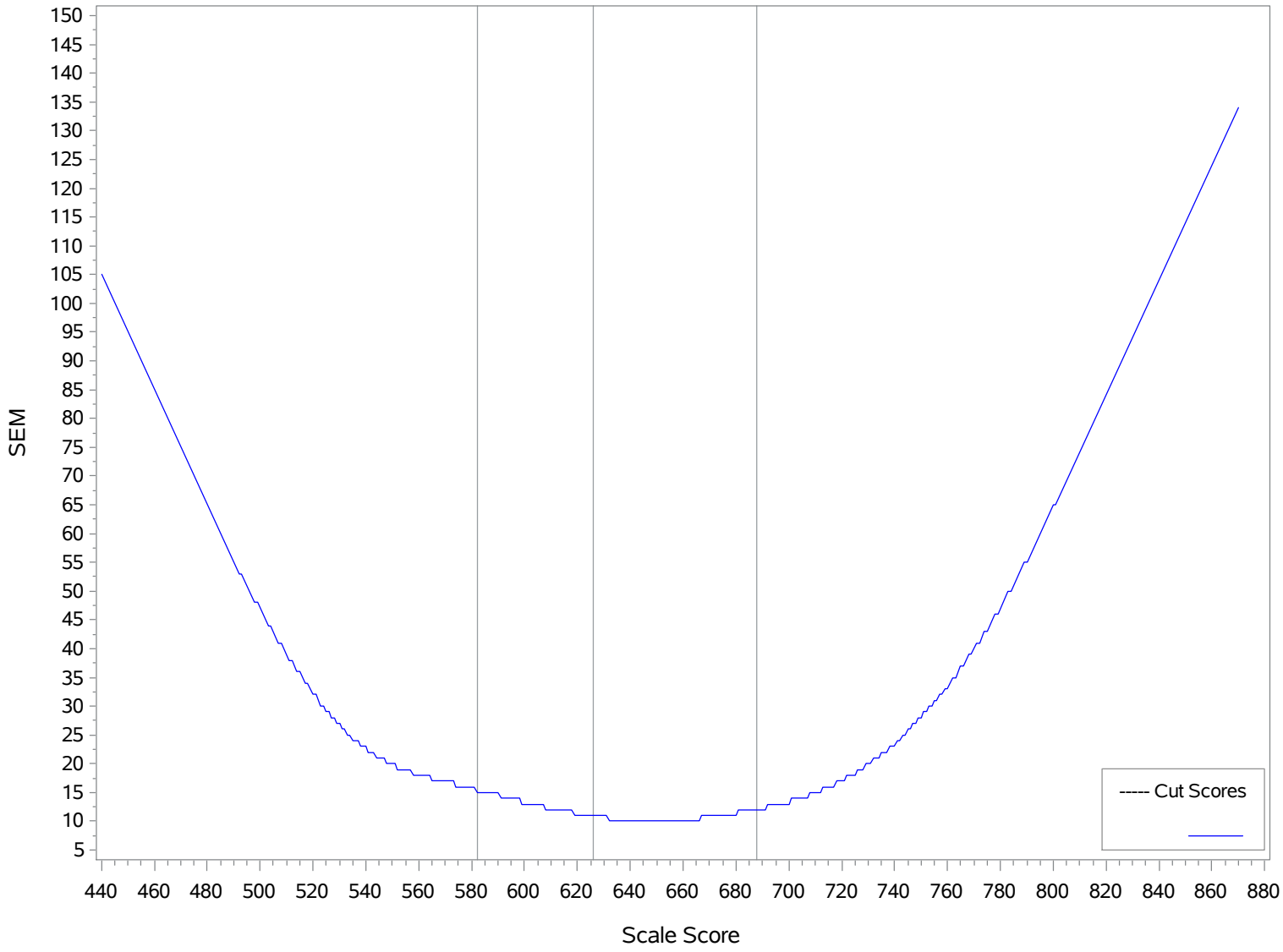
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-9 CSEM with cut scores, Mathematics Grade 5



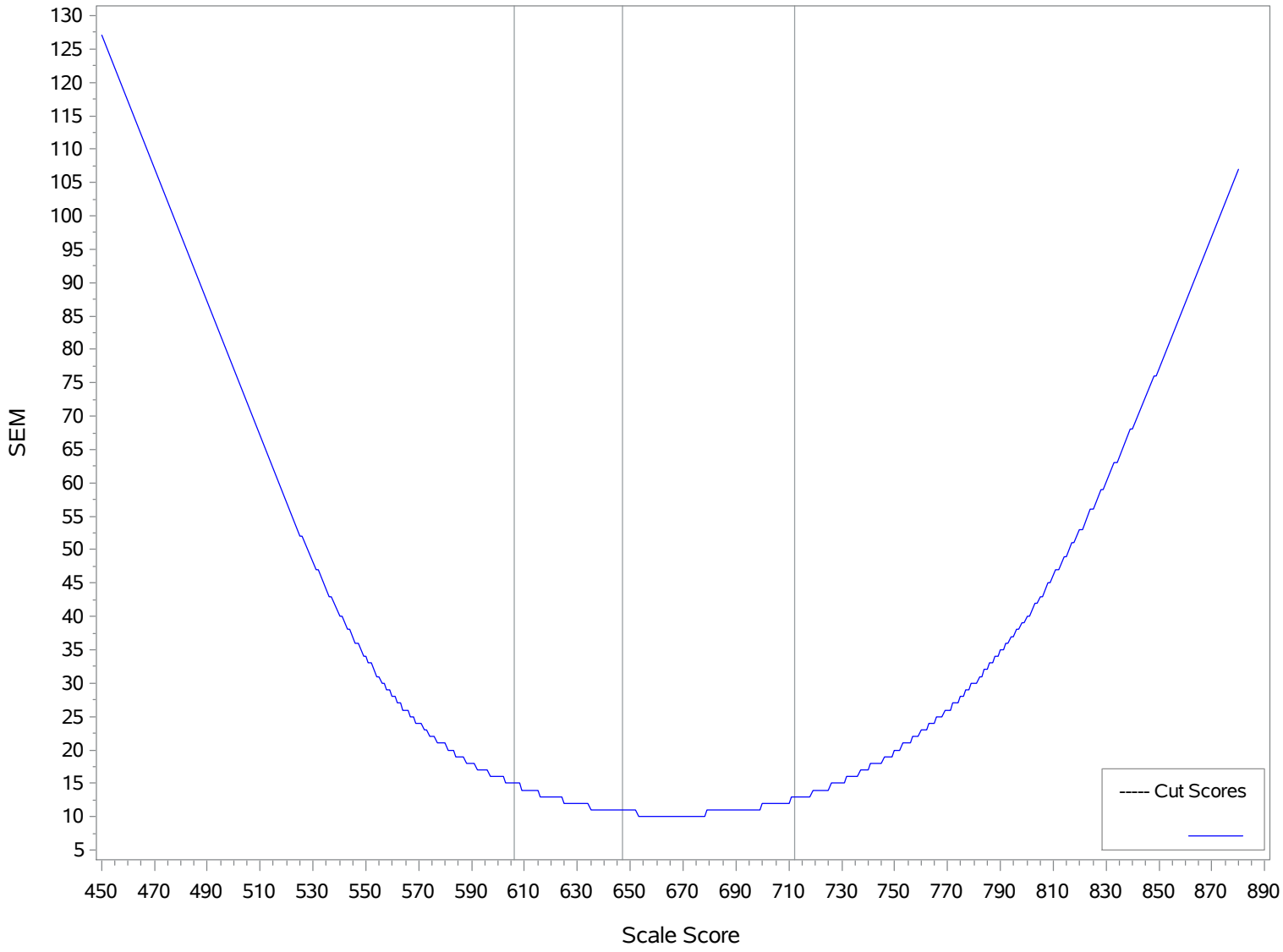
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-10 CSEM with cut scores, Mathematics Grade 6



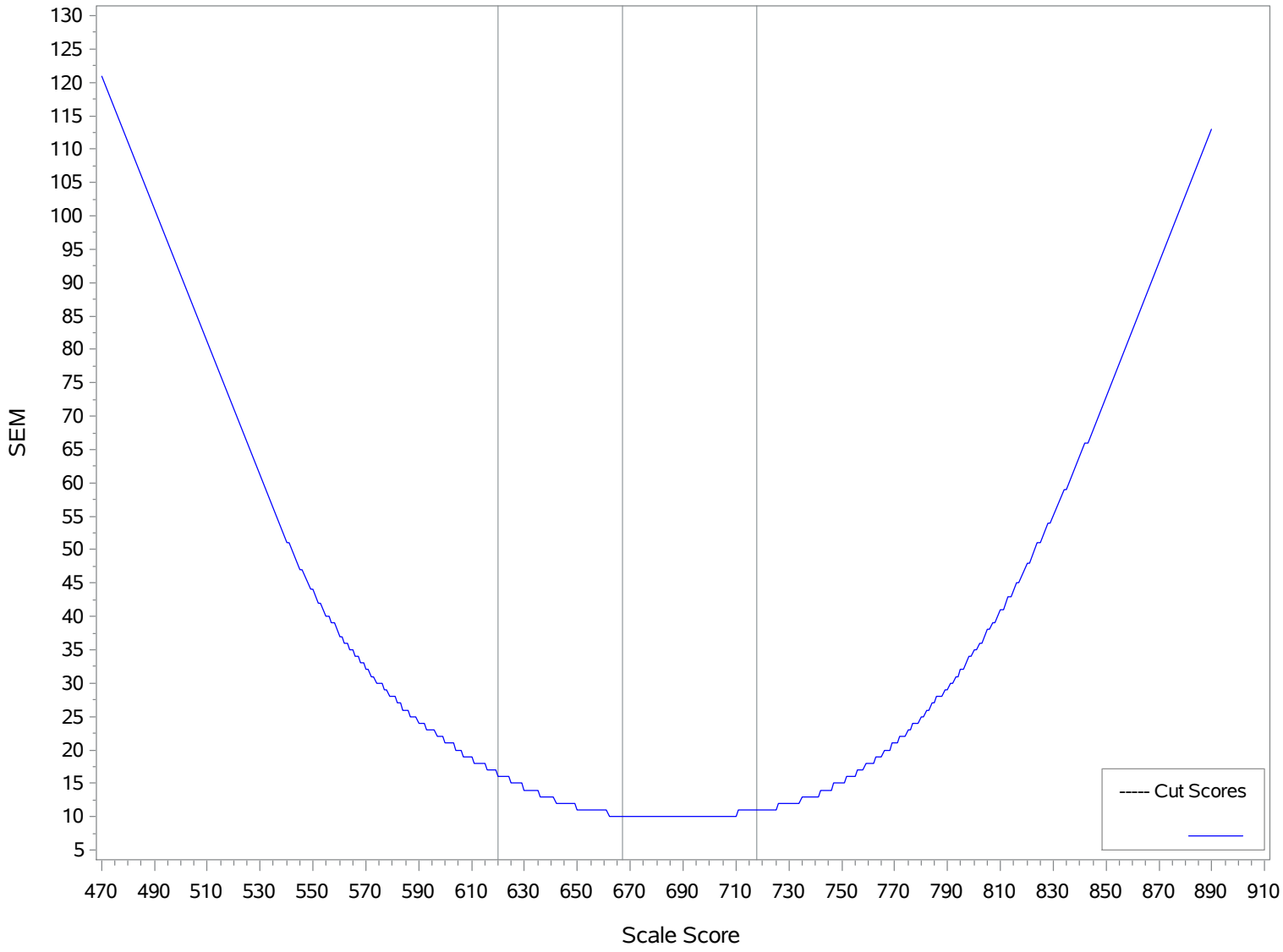
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-11 CSEM with cut scores, Mathematics Grade 7



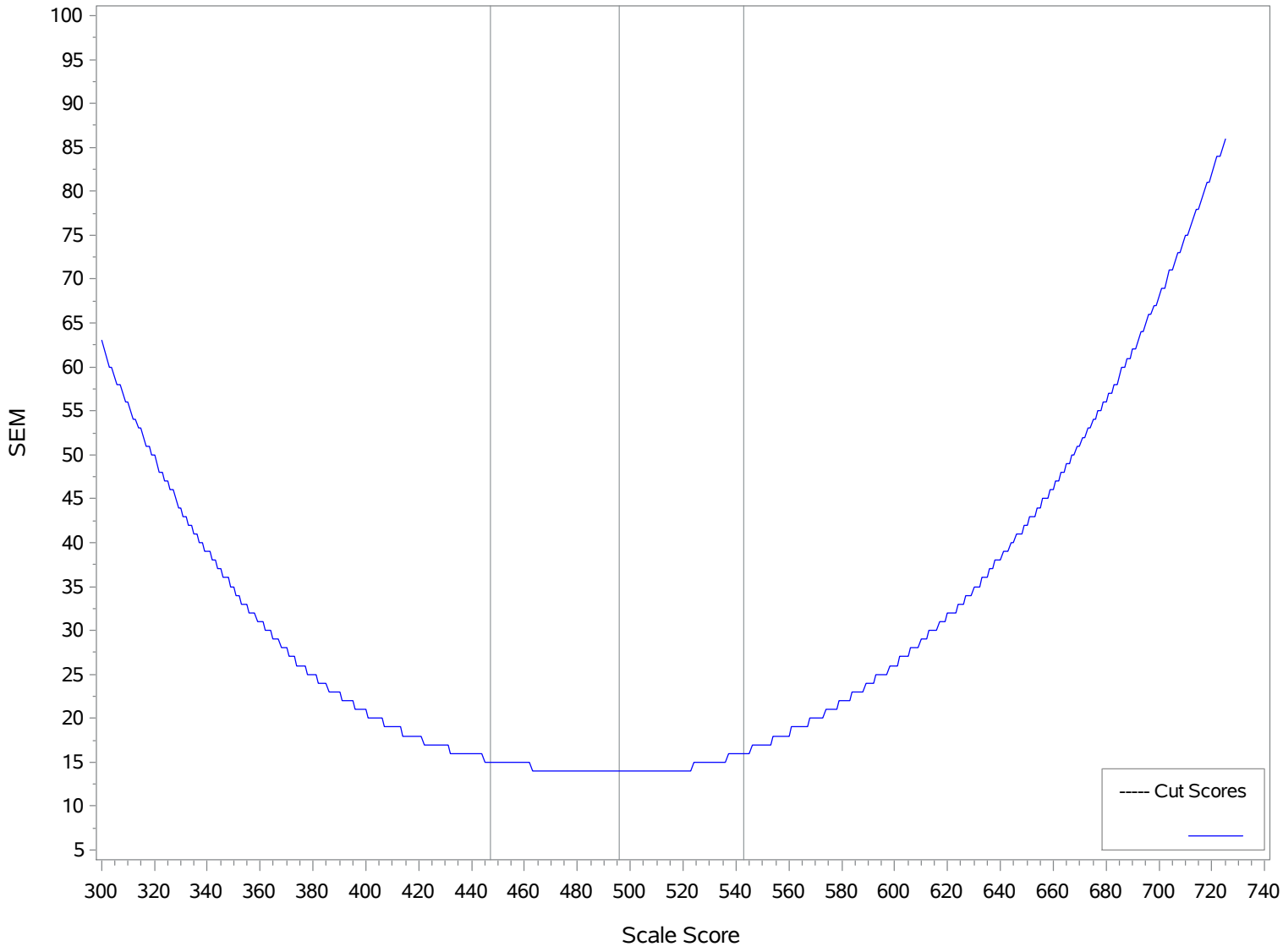
Copyright © 2019 by the Wisconsin Department of Public Instruction

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



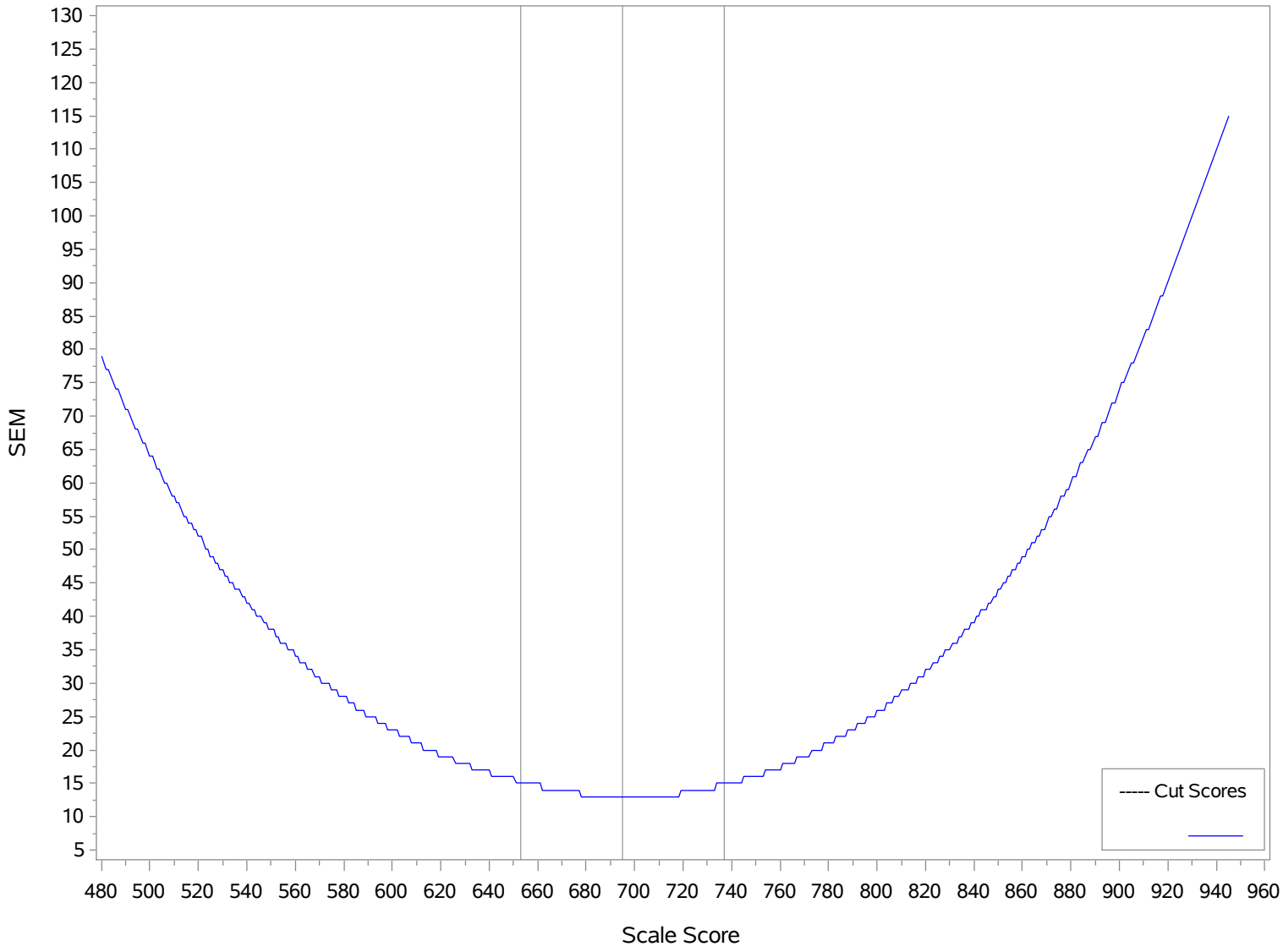
Copyright © 2019 by the Wisconsin Department of Public Instruction

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



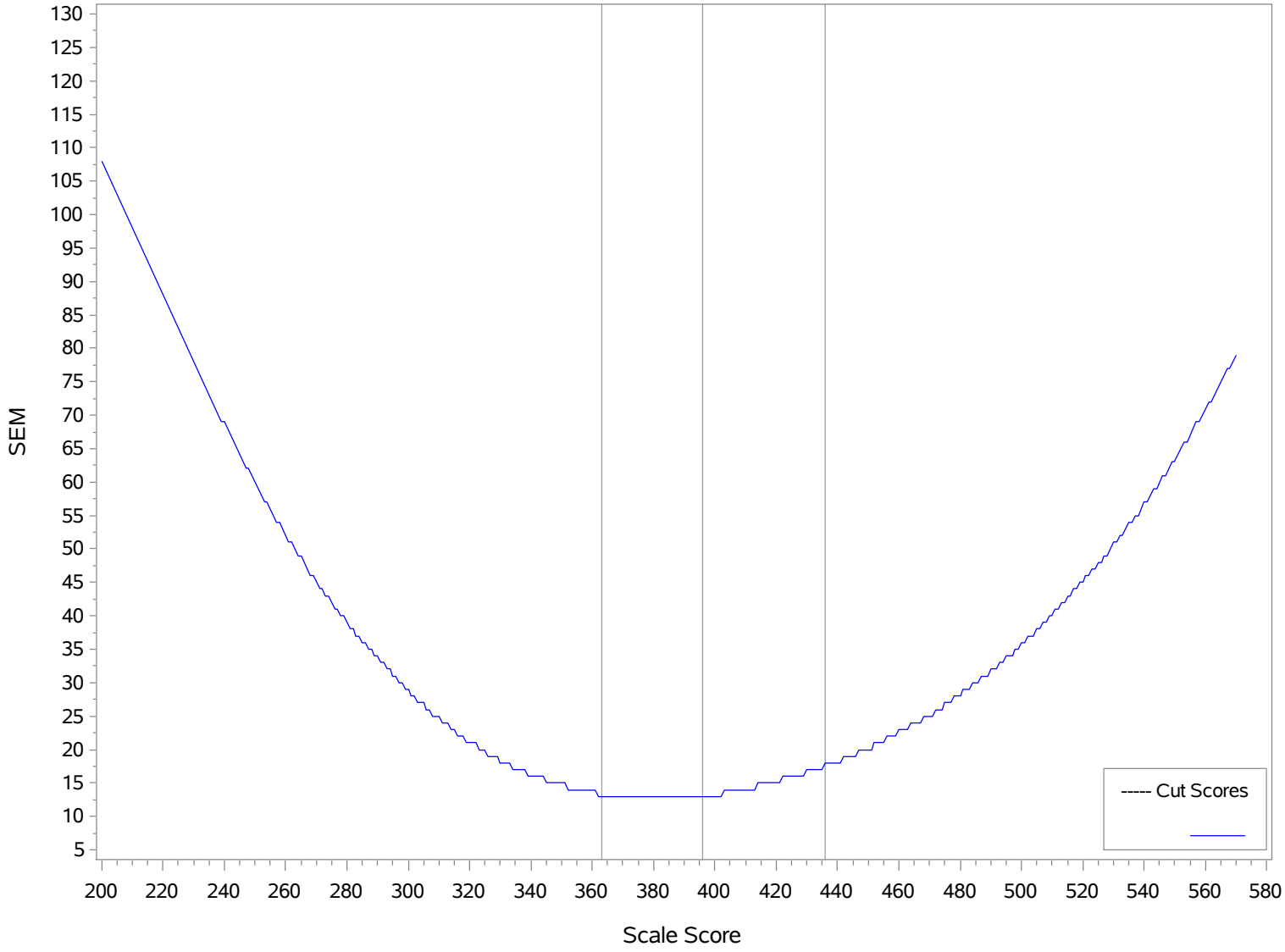
Copyright © 2019 by the Wisconsin Department of Public Instruction

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



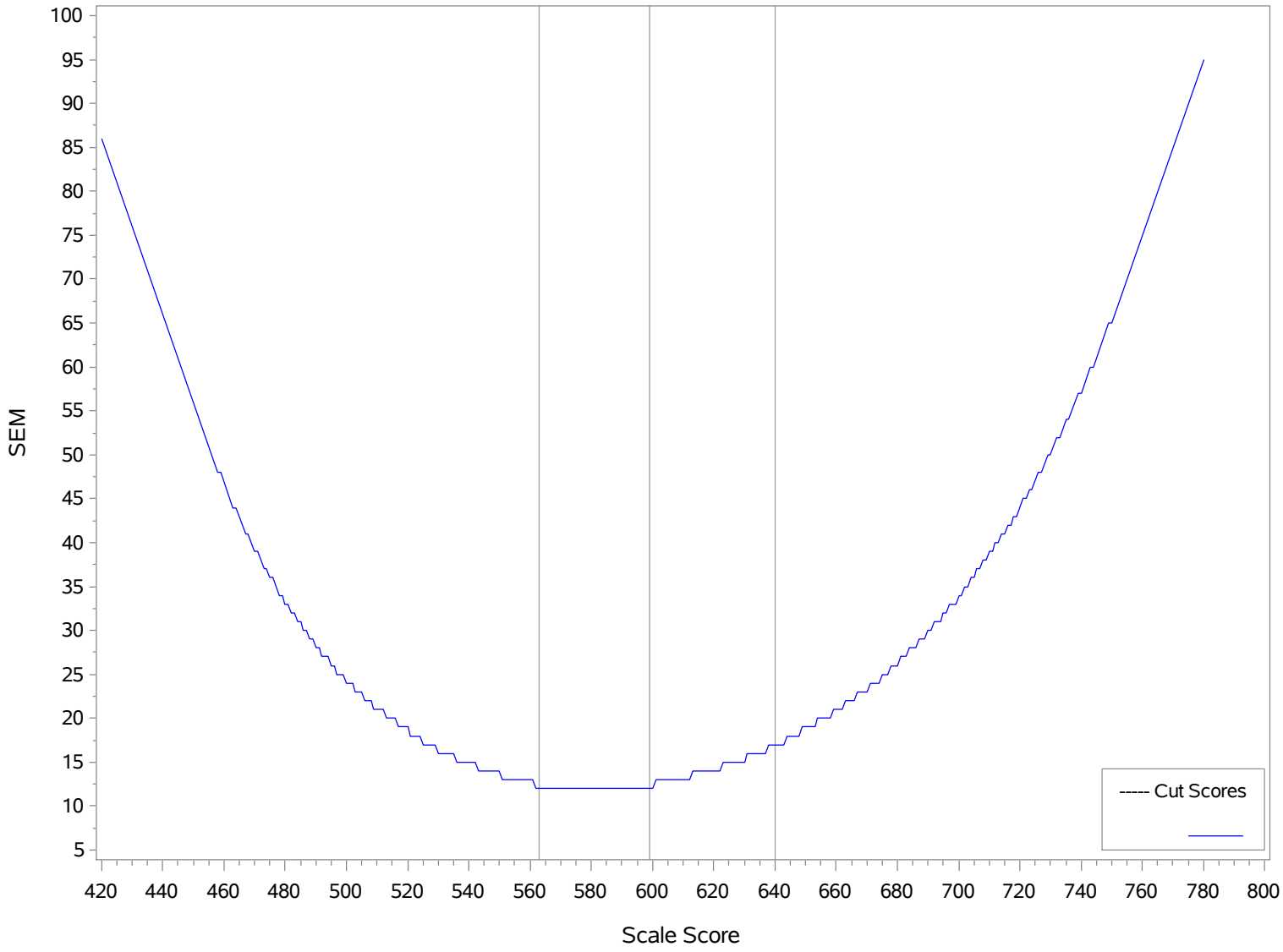
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-15 CSEM with cut scores, Social Studies Grade 4



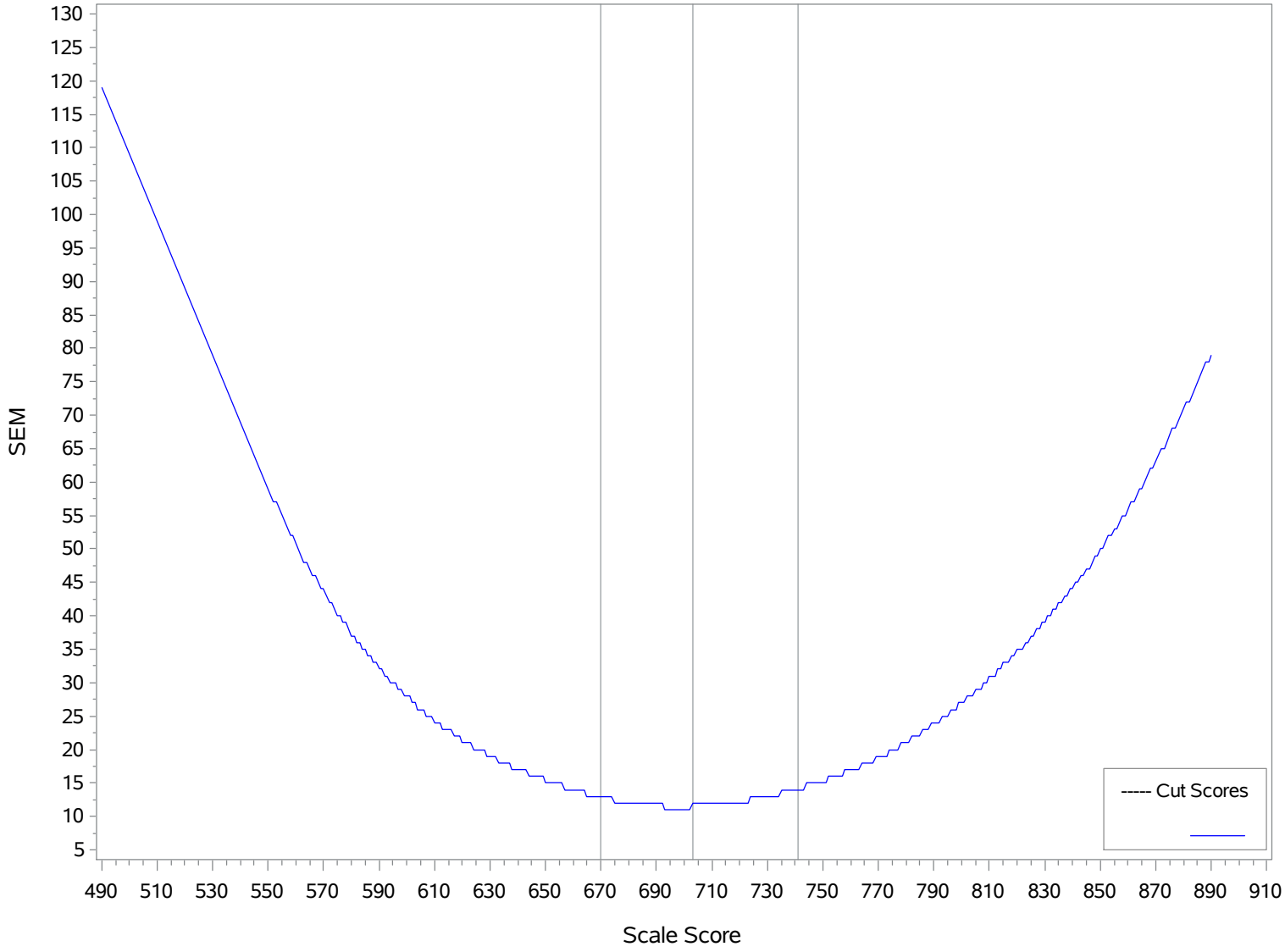
Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-16 CSEM with cut scores, Social Studies Grade 8



Copyright © 2019 by the Wisconsin Department of Public Instruction

Figure I-17 CSEM with cut scores, Social Studies Grade 10



Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix J

Classification Consistency and Accuracy Analysis by Subgroup

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-1 Indexes for Classification Consistency and Accuracy, ELA Grade 3

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.93	0.88	0.94	0.75
		Probability of Chance	0.67	0.51	0.85	0.30
		Kappa (k)	0.77	0.76	0.60	0.64
		Classification Accuracy	0.95	0.92	0.96	0.83
	Male	Classification Consistency (P)	0.91	0.89	0.96	0.76
		Probability of Chance	0.61	0.54	0.90	0.31
		Kappa (k)	0.77	0.76	0.57	0.65
		Classification Accuracy	0.94	0.92	0.97	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.94	0.90	0.93	0.77
		Probability of Chance	0.70	0.50	0.75	0.28
		Kappa (k)	0.79	0.80	0.72	0.68
		Classification Accuracy	0.95	0.93	0.92	0.79
	African-American	Classification Consistency (P)	0.89	0.94	0.99	0.82
		Probability of Chance	0.51	0.77	0.98	0.42
		Kappa (k)	0.77	0.75	0.51	0.69
		Classification Accuracy	0.92	0.96	0.99	0.87
	Hispanic	Classification Consistency (P)	0.89	0.90	0.98	0.77
		Probability of Chance	0.54	0.64	0.95	0.34
		Kappa (k)	0.75	0.74	0.54	0.65
		Classification Accuracy	0.92	0.93	0.98	0.84
	Asian	Classification Consistency (P)	0.90	0.89	0.95	0.74
		Probability of Chance	0.62	0.54	0.86	0.30
		Kappa (k)	0.74	0.77	0.61	0.63
		Classification Accuracy	0.93	0.93	0.96	0.82
	American Indian	Classification Consistency (P)	0.88	0.91	0.98	0.77
		Probability of Chance	0.52	0.69	0.96	0.36
		Kappa (k)	0.74	0.72	0.58	0.64
		Classification Accuracy	0.91	0.94	0.99	0.84
	Two or More	Classification Consistency (P)	0.91	0.89	0.95	0.76
		Probability of Chance	0.61	0.54	0.89	0.30
		Kappa (k)	0.78	0.77	0.58	0.65
		Classification Accuracy	0.94	0.93	0.97	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.87	0.91	0.99	0.77
		Probability of Chance	0.52	0.71	0.98	0.37
		Kappa (k)	0.73	0.69	0.48	0.64
		Classification Accuracy	0.91	0.94	0.99	0.84
Disability Status	Yes	Classification Consistency (P)	0.88	0.94	0.99	0.80
		Probability of Chance	0.50	0.74	0.96	0.39
		Kappa (k)	0.77	0.76	0.60	0.68
		Classification Accuracy	0.91	0.96	0.99	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.90	0.98	0.77
		Probability of Chance	0.54	0.63	0.95	0.33
		Kappa (k)	0.77	0.74	0.55	0.66
		Classification Accuracy	0.92	0.93	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.94	0.93	0.98	0.86
		Probability of Chance	0.50	0.68	0.98	0.37
		Kappa (k)	0.89	0.79	0.26	0.77
		Classification Accuracy	0.96	0.95	0.99	0.90

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-2 Indexes for Classification Consistency and Accuracy, ELA Grade 4

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.90	0.89	0.93	0.71
		Probability of Chance	0.67	0.50	0.79	0.28
		Kappa (k)	0.70	0.77	0.66	0.60
		Classification Accuracy	0.93	0.92	0.95	0.80
	Male	Classification Consistency (P)	0.89	0.89	0.94	0.72
		Probability of Chance	0.60	0.52	0.84	0.29
		Kappa (k)	0.72	0.77	0.63	0.61
		Classification Accuracy	0.92	0.93	0.96	0.81
Race/Ethnicity	White	Classification Consistency (P)	0.91	0.88	0.92	0.71
		Probability of Chance	0.72	0.50	0.77	0.29
		Kappa (k)	0.68	0.75	0.64	0.59
		Classification Accuracy	0.94	0.92	0.94	0.80
	African-American	Classification Consistency (P)	0.85	0.94	0.99	0.77
		Probability of Chance	0.50	0.75	0.96	0.41
		Kappa (k)	0.69	0.75	0.61	0.61
		Classification Accuracy	0.90	0.96	0.99	0.85
	Hispanic	Classification Consistency (P)	0.85	0.90	0.97	0.72
		Probability of Chance	0.54	0.60	0.92	0.32
		Kappa (k)	0.68	0.75	0.60	0.59
		Classification Accuracy	0.90	0.93	0.98	0.81
	Asian	Classification Consistency (P)	0.89	0.88	0.94	0.71
		Probability of Chance	0.64	0.51	0.80	0.28
		Kappa (k)	0.69	0.76	0.71	0.60
		Classification Accuracy	0.92	0.92	0.96	0.79
	American Indian	Classification Consistency (P)	0.85	0.90	0.97	0.73
		Probability of Chance	0.52	0.64	0.94	0.33
		Kappa (k)	0.69	0.73	0.59	0.59
		Classification Accuracy	0.89	0.93	0.98	0.81
	Two or More	Classification Consistency (P)	0.89	0.89	0.94	0.72
		Probability of Chance	0.60	0.52	0.83	0.28
		Kappa (k)	0.72	0.77	0.66	0.61
		Classification Accuracy	0.92	0.92	0.96	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.83	0.91	0.99	0.72
		Probability of Chance	0.52	0.70	0.97	0.36
		Kappa (k)	0.65	0.69	0.51	0.57
		Classification Accuracy	0.89	0.94	0.99	0.81
Disability Status	Yes	Classification Consistency (P)	0.86	0.94	0.98	0.77
		Probability of Chance	0.50	0.72	0.95	0.40
		Kappa (k)	0.71	0.77	0.63	0.62
		Classification Accuracy	0.90	0.96	0.99	0.85

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.86	0.90	0.97	0.73
		Probability of Chance	0.53	0.60	0.91	0.32
		Kappa (k)	0.70	0.75	0.60	0.60
		Classification Accuracy	0.91	0.93	0.98	0.81
Accommodation Use	Yes	Classification Consistency (P)	0.88	0.91	0.98	0.76
		Probability of Chance	0.50	0.63	0.94	0.35
		Kappa (k)	0.76	0.76	0.58	0.64
		Classification Accuracy	0.92	0.93	0.98	0.83

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-3 Indexes for Classification Consistency and Accuracy, ELA Grade 5

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.92	0.89	0.94	0.74
		Probability of Chance	0.66	0.50	0.84	0.30
		Kappa (k)	0.77	0.77	0.61	0.63
		Classification Accuracy	0.95	0.92	0.96	0.82
	Male	Classification Consistency (P)	0.90	0.90	0.95	0.76
		Probability of Chance	0.57	0.54	0.89	0.30
		Kappa (k)	0.78	0.78	0.56	0.65
		Classification Accuracy	0.94	0.93	0.97	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.88	0.93	0.74
		Probability of Chance	0.70	0.50	0.83	0.31
		Kappa (k)	0.74	0.76	0.59	0.62
		Classification Accuracy	0.95	0.91	0.95	0.82
	African-American	Classification Consistency (P)	0.89	0.94	0.99	0.82
		Probability of Chance	0.52	0.77	0.98	0.44
		Kappa (k)	0.76	0.75	0.46	0.67
		Classification Accuracy	0.92	0.96	0.99	0.87
	Hispanic	Classification Consistency (P)	0.89	0.90	0.97	0.76
		Probability of Chance	0.52	0.62	0.94	0.33
		Kappa (k)	0.76	0.75	0.56	0.65
		Classification Accuracy	0.92	0.93	0.98	0.84
	Asian	Classification Consistency (P)	0.91	0.90	0.94	0.75
		Probability of Chance	0.62	0.52	0.83	0.29
		Kappa (k)	0.77	0.79	0.66	0.66
		Classification Accuracy	0.94	0.93	0.96	0.83
	American Indian	Classification Consistency (P)	0.88	0.91	0.98	0.77
		Probability of Chance	0.53	0.65	0.97	0.34
		Kappa (k)	0.74	0.74	0.47	0.65
		Classification Accuracy	0.91	0.93	0.99	0.84
	Two or More	Classification Consistency (P)	0.91	0.90	0.95	0.75
		Probability of Chance	0.59	0.53	0.87	0.29
		Kappa (k)	0.78	0.78	0.59	0.65
		Classification Accuracy	0.94	0.93	0.96	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.93	1.00	0.78
		Probability of Chance	0.50	0.80	0.99	0.42
		Kappa (k)	0.71	0.64	0.38	0.62
		Classification Accuracy	0.90	0.95	1.00	0.85
Disability Status	Yes	Classification Consistency (P)	0.89	0.95	0.99	0.83
		Probability of Chance	0.53	0.78	0.97	0.46
		Kappa (k)	0.77	0.77	0.59	0.68
		Classification Accuracy	0.93	0.96	0.99	0.88

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.91	0.98	0.77
		Probability of Chance	0.52	0.63	0.95	0.33
		Kappa (k)	0.77	0.75	0.52	0.66
		Classification Accuracy	0.92	0.93	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.91	0.94	0.98	0.83
		Probability of Chance	0.50	0.59	0.90	0.36
		Kappa (k)	0.82	0.84	0.78	0.73
		Classification Accuracy	0.94	0.95	0.99	0.88

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-4 Indexes for Classification Consistency and Accuracy, ELA Grade 6

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.88	0.91	0.70
		Probability of Chance	0.69	0.50	0.77	0.28
		Kappa (k)	0.73	0.75	0.61	0.58
		Classification Accuracy	0.94	0.91	0.94	0.79
	Male	Classification Consistency (P)	0.90	0.88	0.94	0.72
		Probability of Chance	0.59	0.54	0.85	0.29
		Kappa (k)	0.75	0.75	0.58	0.60
		Classification Accuracy	0.93	0.92	0.96	0.81
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.87	0.90	0.69
		Probability of Chance	0.72	0.50	0.76	0.29
		Kappa (k)	0.72	0.74	0.59	0.57
		Classification Accuracy	0.94	0.91	0.93	0.79
	African-American	Classification Consistency (P)	0.87	0.93	0.98	0.78
		Probability of Chance	0.50	0.76	0.96	0.41
		Kappa (k)	0.73	0.72	0.54	0.63
		Classification Accuracy	0.91	0.95	0.99	0.85
	Hispanic	Classification Consistency (P)	0.88	0.89	0.96	0.73
		Probability of Chance	0.54	0.61	0.90	0.32
		Kappa (k)	0.73	0.73	0.58	0.60
		Classification Accuracy	0.91	0.93	0.97	0.81
	Asian	Classification Consistency (P)	0.91	0.87	0.92	0.70
		Probability of Chance	0.67	0.51	0.78	0.28
		Kappa (k)	0.72	0.74	0.64	0.58
		Classification Accuracy	0.93	0.92	0.94	0.80
	American Indian	Classification Consistency (P)	0.87	0.89	0.97	0.74
		Probability of Chance	0.52	0.67	0.93	0.34
		Kappa (k)	0.74	0.68	0.56	0.60
		Classification Accuracy	0.91	0.92	0.98	0.81
	Two or More	Classification Consistency (P)	0.90	0.88	0.94	0.72
		Probability of Chance	0.61	0.54	0.84	0.29
		Kappa (k)	0.75	0.75	0.62	0.61
		Classification Accuracy	0.93	0.92	0.96	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.83	0.93	0.99	0.76
		Probability of Chance	0.50	0.85	0.99	0.44
		Kappa (k)	0.67	0.57	0.45	0.58
		Classification Accuracy	0.89	0.96	1.00	0.84
Disability Status	Yes	Classification Consistency (P)	0.86	0.95	0.99	0.80
		Probability of Chance	0.53	0.81	0.97	0.47
		Kappa (k)	0.71	0.74	0.57	0.62
		Classification Accuracy	0.91	0.97	0.99	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.88	0.90	0.96	0.73
		Probability of Chance	0.53	0.62	0.91	0.32
		Kappa (k)	0.73	0.73	0.56	0.61
		Classification Accuracy	0.91	0.93	0.97	0.82
Accommodation Use	Yes	Classification Consistency (P)	0.90	0.91	0.96	0.76
		Probability of Chance	0.51	0.66	0.93	0.35
		Kappa (k)	0.79	0.72	0.45	0.64
		Classification Accuracy	0.93	0.93	0.97	0.82

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-5 Indexes for Classification Consistency and Accuracy, ELA Grade 7

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.93	0.88	0.91	0.72
		Probability of Chance	0.71	0.50	0.77	0.29
		Kappa (k)	0.76	0.76	0.62	0.61
		Classification Accuracy	0.95	0.91	0.94	0.80
	Male	Classification Consistency (P)	0.93	0.93	0.93	0.79
		Probability of Chance	0.57	0.51	0.75	0.27
		Kappa (k)	0.84	0.86	0.72	0.72
		Classification Accuracy	0.94	0.94	0.92	0.80
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.87	0.91	0.72
		Probability of Chance	0.74	0.50	0.77	0.30
		Kappa (k)	0.75	0.75	0.61	0.60
		Classification Accuracy	0.95	0.91	0.94	0.80
	African-American	Classification Consistency (P)	0.88	0.93	0.98	0.79
		Probability of Chance	0.50	0.73	0.96	0.39
		Kappa (k)	0.76	0.73	0.55	0.66
		Classification Accuracy	0.91	0.95	0.99	0.84
	Hispanic	Classification Consistency (P)	0.89	0.89	0.96	0.75
		Probability of Chance	0.57	0.58	0.91	0.31
		Kappa (k)	0.76	0.74	0.58	0.63
		Classification Accuracy	0.92	0.92	0.97	0.81
	Asian	Classification Consistency (P)	0.91	0.85	0.93	0.70
		Probability of Chance	0.72	0.51	0.84	0.31
		Kappa (k)	0.69	0.71	0.59	0.57
		Classification Accuracy	0.93	0.90	0.94	0.76
	American Indian	Classification Consistency (P)	0.87	0.91	0.97	0.75
		Probability of Chance	0.53	0.65	0.93	0.34
		Kappa (k)	0.72	0.74	0.53	0.62
		Classification Accuracy	0.91	0.94	0.98	0.82
	Two or More	Classification Consistency (P)	0.90	0.90	0.94	0.74
		Probability of Chance	0.62	0.52	0.81	0.28
		Kappa (k)	0.75	0.78	0.70	0.64
		Classification Accuracy	0.93	0.92	0.96	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.93	1.00	0.78
		Probability of Chance	0.50	0.82	0.99	0.43
		Kappa (k)	0.70	0.61	0.39	0.62
		Classification Accuracy	0.88	0.95	1.00	0.83
Disability Status	Yes	Classification Consistency (P)	0.88	0.95	0.99	0.81
		Probability of Chance	0.52	0.80	0.97	0.45
		Kappa (k)	0.74	0.74	0.54	0.66
		Classification Accuracy	0.90	0.96	0.99	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.89	0.96	0.75
		Probability of Chance	0.55	0.59	0.92	0.32
		Kappa (k)	0.76	0.74	0.56	0.64
		Classification Accuracy	0.92	0.92	0.97	0.82
Accommodation Use	Yes	Classification Consistency (P)	0.89	0.92	0.97	0.78
		Probability of Chance	0.51	0.71	0.94	0.41
		Kappa (k)	0.77	0.73	0.52	0.63
		Classification Accuracy	0.91	0.94	0.98	0.84

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-6 Indexes for Classification Consistency and Accuracy, ELA Grade 8

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.92	0.88	0.92	0.73
		Probability of Chance	0.66	0.51	0.79	0.28
		Kappa (k)	0.77	0.76	0.64	0.62
		Classification Accuracy	0.95	0.92	0.94	0.81
	Male	Classification Consistency (P)	0.91	0.89	0.95	0.75
		Probability of Chance	0.57	0.57	0.87	0.30
		Kappa (k)	0.79	0.75	0.61	0.64
		Classification Accuracy	0.94	0.92	0.96	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.87	0.92	0.72
		Probability of Chance	0.68	0.51	0.80	0.29
		Kappa (k)	0.76	0.74	0.63	0.61
		Classification Accuracy	0.95	0.91	0.94	0.81
	African-American	Classification Consistency (P)	0.89	0.94	0.98	0.81
		Probability of Chance	0.51	0.77	0.96	0.43
		Kappa (k)	0.77	0.73	0.56	0.67
		Classification Accuracy	0.92	0.96	0.99	0.87
	Hispanic	Classification Consistency (P)	0.89	0.90	0.96	0.76
		Probability of Chance	0.53	0.63	0.91	0.33
		Kappa (k)	0.78	0.74	0.61	0.65
		Classification Accuracy	0.93	0.93	0.97	0.83
	Asian	Classification Consistency (P)	0.92	0.88	0.92	0.72
		Probability of Chance	0.68	0.51	0.76	0.28
		Kappa (k)	0.76	0.76	0.66	0.62
		Classification Accuracy	0.94	0.92	0.94	0.81
	American Indian	Classification Consistency (P)	0.89	0.92	0.98	0.79
		Probability of Chance	0.51	0.70	0.94	0.36
		Kappa (k)	0.78	0.72	0.61	0.67
		Classification Accuracy	0.92	0.94	0.98	0.84
	Two or More	Classification Consistency (P)	0.91	0.90	0.94	0.75
		Probability of Chance	0.58	0.55	0.85	0.29
		Kappa (k)	0.78	0.77	0.60	0.64
		Classification Accuracy	0.93	0.93	0.96	0.82
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.95	1.00	0.81
		Probability of Chance	0.53	0.89	0.99	0.49
		Kappa (k)	0.71	0.59	0.43	0.64
		Classification Accuracy	0.90	0.97	1.00	0.87
Disability Status	Yes	Classification Consistency (P)	0.89	0.96	0.99	0.84
		Probability of Chance	0.55	0.86	0.98	0.51
		Kappa (k)	0.75	0.71	0.59	0.67
		Classification Accuracy	0.92	0.97	0.99	0.89

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.91	0.97	0.77
		Probability of Chance	0.52	0.66	0.93	0.34
		Kappa (k)	0.78	0.73	0.58	0.65
		Classification Accuracy	0.92	0.94	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.94	0.95	0.97	0.86
		Probability of Chance	0.52	0.72	0.89	0.42
		Kappa (k)	0.87	0.81	0.77	0.76
		Classification Accuracy	0.95	0.97	0.98	0.90

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-7 Indexes for Classification Consistency and Accuracy, Mathematics Grade 3

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.93	0.91	0.93	0.77
		Probability of Chance	0.67	0.50	0.79	0.28
		Kappa (k)	0.80	0.81	0.69	0.68
		Classification Accuracy	0.95	0.93	0.95	0.83
	Male	Classification Consistency (P)	0.94	0.91	0.93	0.78
		Probability of Chance	0.69	0.50	0.75	0.28
		Kappa (k)	0.81	0.82	0.71	0.69
		Classification Accuracy	0.96	0.93	0.95	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.95	0.90	0.91	0.77
		Probability of Chance	0.79	0.52	0.71	0.30
		Kappa (k)	0.78	0.79	0.69	0.66
		Classification Accuracy	0.96	0.93	0.94	0.83
	African-American	Classification Consistency (P)	0.90	0.94	0.98	0.82
		Probability of Chance	0.50	0.72	0.95	0.38
		Kappa (k)	0.79	0.79	0.64	0.71
		Classification Accuracy	0.92	0.96	0.99	0.87
	Hispanic	Classification Consistency (P)	0.89	0.88	0.97	0.74
		Probability of Chance	0.59	0.57	0.93	0.32
		Kappa (k)	0.73	0.71	0.57	0.61
		Classification Accuracy	0.92	0.91	0.97	0.81
	Asian	Classification Consistency (P)	0.93	0.91	0.94	0.78
		Probability of Chance	0.67	0.50	0.75	0.27
		Kappa (k)	0.79	0.82	0.75	0.70
		Classification Accuracy	0.95	0.93	0.95	0.84
	American Indian	Classification Consistency (P)	0.89	0.91	0.98	0.78
		Probability of Chance	0.54	0.62	0.93	0.33
		Kappa (k)	0.77	0.77	0.67	0.68
		Classification Accuracy	0.92	0.94	0.98	0.84
	Two or More	Classification Consistency (P)	0.93	0.91	0.94	0.78
		Probability of Chance	0.65	0.51	0.80	0.28
		Kappa (k)	0.79	0.81	0.71	0.69
		Classification Accuracy	0.95	0.93	0.96	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.90	0.91	0.97	0.79
		Probability of Chance	0.55	0.60	0.92	0.32
		Kappa (k)	0.78	0.78	0.66	0.68
		Classification Accuracy	0.93	0.94	0.98	0.84
Disability Status	Yes	Classification Consistency (P)	0.91	0.93	0.97	0.82
		Probability of Chance	0.50	0.64	0.91	0.34
		Kappa (k)	0.82	0.82	0.68	0.72
		Classification Accuracy	0.93	0.95	0.98	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.91	0.91	0.96	0.79
		Probability of Chance	0.56	0.56	0.89	0.30
		Kappa (k)	0.80	0.80	0.67	0.70
		Classification Accuracy	0.94	0.93	0.97	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.91	0.96	1.00	0.87
		Probability of Chance	0.56	0.87	0.99	0.53
		Kappa (k)	0.79	0.71	0.60	0.72
		Classification Accuracy	0.93	0.97	1.00	0.90

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-8 Indexes for Classification Consistency and Accuracy, Mathematics Grade 4

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.90	0.95	0.76
		Probability of Chance	0.68	0.51	0.80	0.29
		Kappa (k)	0.72	0.80	0.74	0.67
		Classification Accuracy	0.94	0.93	0.96	0.83
	Male	Classification Consistency (P)	0.92	0.91	0.94	0.77
		Probability of Chance	0.69	0.50	0.75	0.28
		Kappa (k)	0.74	0.82	0.75	0.68
		Classification Accuracy	0.94	0.93	0.96	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.94	0.90	0.93	0.76
		Probability of Chance	0.80	0.51	0.72	0.30
		Kappa (k)	0.69	0.79	0.74	0.66
		Classification Accuracy	0.96	0.92	0.95	0.83
	African-American	Classification Consistency (P)	0.85	0.95	0.99	0.79
		Probability of Chance	0.50	0.77	0.97	0.40
		Kappa (k)	0.70	0.77	0.66	0.65
		Classification Accuracy	0.90	0.96	0.99	0.85
	Hispanic	Classification Consistency (P)	0.87	0.92	0.97	0.76
		Probability of Chance	0.57	0.60	0.91	0.32
		Kappa (k)	0.70	0.79	0.70	0.65
		Classification Accuracy	0.91	0.94	0.98	0.83
	Asian	Classification Consistency (P)	0.91	0.91	0.95	0.78
		Probability of Chance	0.71	0.50	0.72	0.28
		Kappa (k)	0.70	0.82	0.82	0.69
		Classification Accuracy	0.94	0.94	0.97	0.84
	American Indian	Classification Consistency (P)	0.86	0.92	0.98	0.77
		Probability of Chance	0.55	0.66	0.93	0.35
		Kappa (k)	0.70	0.77	0.70	0.64
		Classification Accuracy	0.90	0.94	0.99	0.83
	Two or More	Classification Consistency (P)	0.90	0.91	0.96	0.77
		Probability of Chance	0.65	0.53	0.82	0.29
		Kappa (k)	0.71	0.81	0.76	0.67
		Classification Accuracy	0.93	0.93	0.97	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.92	0.98	0.76
		Probability of Chance	0.55	0.65	0.94	0.35
		Kappa (k)	0.69	0.77	0.67	0.64
		Classification Accuracy	0.91	0.94	0.99	0.83
Disability Status	Yes	Classification Consistency (P)	0.87	0.95	0.98	0.80
		Probability of Chance	0.50	0.69	0.92	0.36
		Kappa (k)	0.74	0.82	0.75	0.68
		Classification Accuracy	0.91	0.96	0.99	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.88	0.92	0.97	0.77
		Probability of Chance	0.57	0.60	0.90	0.32
		Kappa (k)	0.72	0.79	0.70	0.66
		Classification Accuracy	0.92	0.94	0.98	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.85	0.97	1.00	0.82
		Probability of Chance	0.55	0.88	0.99	0.51
		Kappa (k)	0.67	0.76	0.62	0.63
		Classification Accuracy	0.90	0.98	1.00	0.87

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-9 Indexes for Classification Consistency and Accuracy, Mathematics Grade 5

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.90	0.90	0.94	0.74
		Probability of Chance	0.63	0.50	0.80	0.28
		Kappa (k)	0.71	0.80	0.70	0.64
		Classification Accuracy	0.93	0.93	0.96	0.82
	Male	Classification Consistency (P)	0.90	0.91	0.94	0.75
		Probability of Chance	0.62	0.50	0.76	0.27
		Kappa (k)	0.74	0.82	0.73	0.66
		Classification Accuracy	0.93	0.94	0.96	0.83
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.90	0.92	0.74
		Probability of Chance	0.73	0.51	0.73	0.29
		Kappa (k)	0.70	0.80	0.71	0.64
		Classification Accuracy	0.94	0.93	0.95	0.82
	African-American	Classification Consistency (P)	0.85	0.94	0.99	0.78
		Probability of Chance	0.51	0.76	0.97	0.43
		Kappa (k)	0.69	0.75	0.64	0.62
		Classification Accuracy	0.89	0.96	0.99	0.84
	Hispanic	Classification Consistency (P)	0.87	0.91	0.97	0.75
		Probability of Chance	0.53	0.59	0.91	0.32
		Kappa (k)	0.72	0.78	0.68	0.64
		Classification Accuracy	0.91	0.93	0.98	0.82
	Asian	Classification Consistency (P)	0.91	0.89	0.94	0.75
		Probability of Chance	0.68	0.50	0.71	0.27
		Kappa (k)	0.73	0.77	0.80	0.65
		Classification Accuracy	0.94	0.92	0.96	0.83
	American Indian	Classification Consistency (P)	0.87	0.92	0.98	0.76
		Probability of Chance	0.52	0.61	0.93	0.33
		Kappa (k)	0.72	0.79	0.68	0.64
		Classification Accuracy	0.91	0.94	0.99	0.83
	Two or More	Classification Consistency (P)	0.89	0.90	0.96	0.75
		Probability of Chance	0.59	0.52	0.82	0.28
		Kappa (k)	0.74	0.80	0.77	0.66
		Classification Accuracy	0.93	0.93	0.97	0.83
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.92	0.99	0.76
		Probability of Chance	0.50	0.70	0.97	0.37
		Kappa (k)	0.70	0.72	0.67	0.62
		Classification Accuracy	0.90	0.94	0.99	0.83
Disability Status	Yes	Classification Consistency (P)	0.87	0.95	0.98	0.80
		Probability of Chance	0.51	0.71	0.93	0.41
		Kappa (k)	0.73	0.82	0.71	0.66
		Classification Accuracy	0.91	0.96	0.99	0.86

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.87	0.91	0.97	0.75
		Probability of Chance	0.53	0.58	0.91	0.31
		Kappa (k)	0.72	0.79	0.68	0.64
		Classification Accuracy	0.90	0.94	0.98	0.82
Accommodation Use	Yes	Classification Consistency (P)	0.87	0.97	1.00	0.84
		Probability of Chance	0.61	0.88	0.99	0.58
		Kappa (k)	0.67	0.74	0.67	0.61
		Classification Accuracy	0.91	0.98	1.00	0.88

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-10 Indexes for Classification Consistency and Accuracy, Mathematics Grade 6

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.90	0.96	0.77
		Probability of Chance	0.62	0.51	0.87	0.30
		Kappa (k)	0.76	0.79	0.70	0.67
		Classification Accuracy	0.93	0.93	0.97	0.83
	Male	Classification Consistency (P)	0.91	0.91	0.96	0.78
		Probability of Chance	0.60	0.51	0.85	0.29
		Kappa (k)	0.78	0.81	0.71	0.68
		Classification Accuracy	0.94	0.93	0.97	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.89	0.95	0.77
		Probability of Chance	0.70	0.50	0.83	0.31
		Kappa (k)	0.75	0.78	0.69	0.66
		Classification Accuracy	0.94	0.92	0.96	0.83
	African-American	Classification Consistency (P)	0.88	0.95	0.99	0.82
		Probability of Chance	0.52	0.78	0.98	0.46
		Kappa (k)	0.75	0.75	0.65	0.67
		Classification Accuracy	0.91	0.96	1.00	0.87
	Hispanic	Classification Consistency (P)	0.89	0.95	0.98	0.83
		Probability of Chance	0.51	0.66	0.92	0.34
		Kappa (k)	0.78	0.85	0.80	0.73
		Classification Accuracy	0.92	0.95	0.98	0.85
	Asian	Classification Consistency (P)	0.90	0.90	0.95	0.76
		Probability of Chance	0.63	0.50	0.79	0.28
		Kappa (k)	0.73	0.80	0.78	0.66
		Classification Accuracy	0.93	0.93	0.97	0.83
	American Indian	Classification Consistency (P)	0.85	0.93	0.99	0.77
		Probability of Chance	0.50	0.67	0.95	0.36
		Kappa (k)	0.70	0.79	0.73	0.64
		Classification Accuracy	0.90	0.95	0.99	0.84
	Two or More	Classification Consistency (P)	0.89	0.91	0.97	0.77
		Probability of Chance	0.55	0.54	0.88	0.30
		Kappa (k)	0.75	0.80	0.74	0.67
		Classification Accuracy	0.92	0.93	0.98	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.85	0.94	1.00	0.79
		Probability of Chance	0.51	0.82	0.99	0.44
		Kappa (k)	0.69	0.66	0.64	0.62
		Classification Accuracy	0.89	0.96	1.00	0.85
Disability Status	Yes	Classification Consistency (P)	0.89	0.95	0.99	0.83
		Probability of Chance	0.54	0.78	0.97	0.48
		Kappa (k)	0.76	0.79	0.70	0.68
		Classification Accuracy	0.92	0.97	0.99	0.88

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.91	0.98	0.78
		Probability of Chance	0.51	0.62	0.95	0.33
		Kappa (k)	0.77	0.77	0.64	0.67
		Classification Accuracy	0.91	0.94	0.99	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.89	0.98	1.00	0.87
		Probability of Chance	0.68	0.94	1.00	0.67
		Kappa (k)	0.66	0.68	0.62	0.61
		Classification Accuracy	0.92	0.99	1.00	0.91

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-11 Indexes for Classification Consistency and Accuracy, Mathematics Grade 7

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.89	0.90	0.97	0.76
		Probability of Chance	0.57	0.53	0.91	0.30
		Kappa (k)	0.75	0.78	0.69	0.66
		Classification Accuracy	0.93	0.93	0.98	0.84
	Male	Classification Consistency (P)	0.91	0.90	0.97	0.78
		Probability of Chance	0.55	0.52	0.89	0.30
		Kappa (k)	0.80	0.79	0.70	0.68
		Classification Accuracy	0.94	0.93	0.98	0.85
Race/Ethnicity	White	Classification Consistency (P)	0.91	0.90	0.96	0.77
		Probability of Chance	0.65	0.50	0.87	0.31
		Kappa (k)	0.75	0.80	0.67	0.67
		Classification Accuracy	0.94	0.93	0.97	0.84
	African-American	Classification Consistency (P)	0.89	0.96	1.00	0.85
		Probability of Chance	0.59	0.84	0.99	0.55
		Kappa (k)	0.73	0.74	0.62	0.65
		Classification Accuracy	0.92	0.97	1.00	0.89
	Hispanic	Classification Consistency (P)	0.87	0.93	0.99	0.79
		Probability of Chance	0.50	0.66	0.97	0.36
		Kappa (k)	0.74	0.79	0.64	0.67
		Classification Accuracy	0.91	0.95	0.99	0.85
	Asian	Classification Consistency (P)	0.89	0.92	0.97	0.78
		Probability of Chance	0.57	0.52	0.83	0.28
		Kappa (k)	0.75	0.83	0.80	0.69
		Classification Accuracy	0.92	0.94	0.98	0.83
	American Indian	Classification Consistency (P)	0.88	0.95	0.99	0.82
		Probability of Chance	0.51	0.71	0.98	0.41
		Kappa (k)	0.76	0.82	0.69	0.70
		Classification Accuracy	0.91	0.96	0.99	0.86
	Two or More	Classification Consistency (P)	0.89	0.92	0.97	0.79
		Probability of Chance	0.53	0.56	0.91	0.31
		Kappa (k)	0.77	0.82	0.70	0.69
		Classification Accuracy	0.92	0.94	0.98	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.96	1.00	0.81
		Probability of Chance	0.56	0.87	1.00	0.52
		Kappa (k)	0.65	0.70	0.62	0.59
		Classification Accuracy	0.89	0.97	1.00	0.86
Disability Status	Yes	Classification Consistency (P)	0.89	0.96	0.99	0.85
		Probability of Chance	0.60	0.83	0.98	0.56
		Kappa (k)	0.74	0.78	0.71	0.66
		Classification Accuracy	0.93	0.97	1.00	0.90

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-11 Indexes for Classification Consistency and Accuracy, Mathematics Grade 7

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.88	0.92	0.99	0.79
		Probability of Chance	0.50	0.66	0.97	0.37
		Kappa (k)	0.76	0.76	0.63	0.66
		Classification Accuracy	0.91	0.94	0.99	0.85
Accommodation Use	Yes	Classification Consistency (P)	0.91	0.98	1.00	0.89
		Probability of Chance	0.75	0.95	1.00	0.74
		Kappa (k)	0.63	0.66	0.77	0.58
		Classification Accuracy	0.94	0.99	1.00	0.93

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-12 Indexes for Classification Consistency and Accuracy, Mathematics Grade 8

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.89	0.91	0.96	0.76
		Probability of Chance	0.61	0.53	0.85	0.29
		Kappa (k)	0.72	0.80	0.71	0.66
		Classification Accuracy	0.92	0.94	0.97	0.83
	Male	Classification Consistency (P)	0.89	0.92	0.96	0.77
		Probability of Chance	0.57	0.54	0.83	0.29
		Kappa (k)	0.75	0.82	0.74	0.68
		Classification Accuracy	0.92	0.95	0.97	0.84
Race/Ethnicity	White	Classification Consistency (P)	0.90	0.90	0.95	0.75
		Probability of Chance	0.67	0.51	0.81	0.29
		Kappa (k)	0.71	0.80	0.72	0.65
		Classification Accuracy	0.93	0.94	0.96	0.83
	African-American	Classification Consistency (P)	0.86	0.96	0.99	0.82
		Probability of Chance	0.54	0.83	0.98	0.49
		Kappa (k)	0.70	0.79	0.67	0.64
		Classification Accuracy	0.90	0.98	1.00	0.87
	Hispanic	Classification Consistency (P)	0.85	0.93	0.98	0.77
		Probability of Chance	0.51	0.68	0.94	0.36
		Kappa (k)	0.70	0.79	0.69	0.64
		Classification Accuracy	0.89	0.96	0.99	0.84
	Asian	Classification Consistency (P)	0.89	0.91	0.95	0.76
		Probability of Chance	0.65	0.51	0.74	0.27
		Kappa (k)	0.70	0.82	0.81	0.67
		Classification Accuracy	0.92	0.94	0.97	0.83
	American Indian	Classification Consistency (P)	0.84	0.93	0.99	0.77
		Probability of Chance	0.50	0.72	0.97	0.37
		Kappa (k)	0.68	0.77	0.71	0.63
		Classification Accuracy	0.89	0.96	0.99	0.83
	Two or More	Classification Consistency (P)	0.88	0.92	0.97	0.77
		Probability of Chance	0.55	0.58	0.87	0.30
		Kappa (k)	0.73	0.81	0.80	0.67
		Classification Accuracy	0.91	0.94	0.98	0.84
Limited English Proficiency	Yes	Classification Consistency (P)	0.83	0.97	1.00	0.80
		Probability of Chance	0.54	0.88	0.99	0.50
		Kappa (k)	0.64	0.73	0.63	0.59
		Classification Accuracy	0.88	0.98	1.00	0.85
Disability Status	Yes	Classification Consistency (P)	0.88	0.98	0.99	0.86
		Probability of Chance	0.60	0.86	0.97	0.56
		Kappa (k)	0.72	0.84	0.82	0.67
		Classification Accuracy	0.91	0.99	1.00	0.89

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
SES Disadvantaged	Yes	Classification Consistency (P)	0.86	0.93	0.98	0.78
		Probability of Chance	0.50	0.69	0.94	0.36
		Kappa (k)	0.72	0.79	0.69	0.65
		Classification Accuracy	0.90	0.96	0.99	0.84
Accommodation Use	Yes	Classification Consistency (P)	0.86	0.99	1.00	0.85
		Probability of Chance	0.67	0.97	1.00	0.67
		Kappa (k)	0.57	0.65	0.65	0.55
		Classification Accuracy	0.90	0.99	1.00	0.89

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-13 Indexes for Classification Consistency and Accuracy, Science Grade 4

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.89	0.92	0.72
		Probability of Chance	0.71	0.50	0.65	0.26
		Kappa (k)	0.69	0.78	0.77	0.62
		Classification Accuracy	0.93	0.92	0.93	0.78
	Male	Classification Consistency (P)	0.92	0.88	0.90	0.70
		Probability of Chance	0.73	0.50	0.65	0.27
		Kappa (k)	0.69	0.75	0.71	0.59
		Classification Accuracy	0.94	0.91	0.93	0.78
Race/Ethnicity	White	Classification Consistency (P)	0.95	0.88	0.88	0.71
		Probability of Chance	0.85	0.54	0.61	0.29
		Kappa (k)	0.65	0.73	0.70	0.59
		Classification Accuracy	0.96	0.91	0.91	0.78
	African-American	Classification Consistency (P)	0.83	0.91	0.98	0.72
		Probability of Chance	0.51	0.72	0.94	0.37
		Kappa (k)	0.65	0.70	0.64	0.56
		Classification Accuracy	0.88	0.94	0.98	0.80
	Hispanic	Classification Consistency (P)	0.87	0.87	0.95	0.69
		Probability of Chance	0.62	0.55	0.84	0.30
		Kappa (k)	0.65	0.71	0.67	0.55
		Classification Accuracy	0.90	0.91	0.96	0.77
	Asian	Classification Consistency (P)	0.90	0.88	0.92	0.71
		Probability of Chance	0.72	0.50	0.71	0.28
		Kappa (k)	0.64	0.77	0.73	0.59
		Classification Accuracy	0.93	0.92	0.94	0.79
	American Indian	Classification Consistency (P)	0.85	0.88	0.95	0.68
		Probability of Chance	0.61	0.57	0.85	0.31
		Kappa (k)	0.62	0.71	0.66	0.53
		Classification Accuracy	0.89	0.91	0.96	0.76
	Two or More	Classification Consistency (P)	0.91	0.89	0.92	0.71
		Probability of Chance	0.71	0.50	0.71	0.27
		Kappa (k)	0.69	0.77	0.71	0.61
		Classification Accuracy	0.94	0.91	0.94	0.79
Limited English Proficiency	Yes	Classification Consistency (P)	0.84	0.88	0.97	0.69
		Probability of Chance	0.58	0.63	0.92	0.34
		Kappa (k)	0.63	0.67	0.61	0.53
		Classification Accuracy	0.89	0.91	0.98	0.78
Disability Status	Yes	Classification Consistency (P)	0.85	0.91	0.96	0.72
		Probability of Chance	0.53	0.61	0.85	0.31
		Kappa (k)	0.68	0.76	0.72	0.59
		Classification Accuracy	0.89	0.93	0.97	0.80
SES Disadvantaged	Yes	Classification Consistency (P)	0.88	0.88	0.94	0.70
		Probability of Chance	0.62	0.54	0.82	0.29
		Kappa (k)	0.67	0.73	0.69	0.57
		Classification Accuracy	0.91	0.91	0.96	0.78

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-14 Indexes for Classification Consistency and Accuracy, Science Grade 8

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.89	0.90	0.70
		Probability of Chance	0.72	0.50	0.65	0.26
		Kappa (k)	0.70	0.77	0.71	0.60
		Classification Accuracy	0.94	0.92	0.93	0.78
	Male	Classification Consistency (P)	0.91	0.90	0.90	0.72
		Probability of Chance	0.68	0.50	0.63	0.25
		Kappa (k)	0.73	0.79	0.74	0.62
		Classification Accuracy	0.94	0.93	0.93	0.80
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.89	0.88	0.71
		Probability of Chance	0.79	0.53	0.59	0.28
		Kappa (k)	0.68	0.76	0.72	0.60
		Classification Accuracy	0.95	0.92	0.92	0.79
	African-American	Classification Consistency (P)	0.84	0.92	0.97	0.73
		Probability of Chance	0.50	0.70	0.92	0.37
		Kappa (k)	0.67	0.73	0.69	0.57
		Classification Accuracy	0.89	0.95	0.98	0.82
	Hispanic	Classification Consistency (P)	0.87	0.88	0.94	0.69
		Probability of Chance	0.58	0.54	0.80	0.28
		Kappa (k)	0.69	0.74	0.70	0.57
		Classification Accuracy	0.91	0.92	0.96	0.79
	Asian	Classification Consistency (P)	0.91	0.89	0.90	0.70
		Probability of Chance	0.73	0.51	0.60	0.26
		Kappa (k)	0.66	0.77	0.74	0.59
		Classification Accuracy	0.93	0.92	0.93	0.78
	American Indian	Classification Consistency (P)	0.86	0.88	0.94	0.68
		Probability of Chance	0.59	0.55	0.81	0.29
		Kappa (k)	0.66	0.73	0.69	0.56
		Classification Accuracy	0.90	0.91	0.96	0.77
	Two or More	Classification Consistency (P)	0.90	0.88	0.91	0.70
		Probability of Chance	0.66	0.50	0.67	0.26
		Kappa (k)	0.71	0.77	0.71	0.59
		Classification Accuracy	0.93	0.92	0.94	0.78
Limited English Proficiency	Yes	Classification Consistency (P)	0.80	0.92	0.98	0.71
		Probability of Chance	0.50	0.78	0.97	0.40
		Kappa (k)	0.60	0.65	0.55	0.52
		Classification Accuracy	0.86	0.94	0.99	0.79
Disability Status	Yes	Classification Consistency (P)	0.83	0.92	0.97	0.72
		Probability of Chance	0.50	0.67	0.88	0.36
		Kappa (k)	0.66	0.75	0.75	0.57
		Classification Accuracy	0.88	0.95	0.98	0.81
SES Disadvantaged	Yes	Classification Consistency (P)	0.87	0.89	0.94	0.71
		Probability of Chance	0.57	0.54	0.79	0.28
		Kappa (k)	0.71	0.77	0.71	0.60
		Classification Accuracy	0.91	0.92	0.96	0.79

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-15 Indexes for Classification Consistency and Accuracy, Social Studies Grade 4

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.87	0.88	0.67
		Probability of Chance	0.65	0.50	0.64	0.25
		Kappa (k)	0.75	0.73	0.67	0.56
		Classification Accuracy	0.94	0.91	0.91	0.76
	Male	Classification Consistency (P)	0.91	0.88	0.89	0.69
		Probability of Chance	0.61	0.50	0.63	0.25
		Kappa (k)	0.78	0.76	0.69	0.58
		Classification Accuracy	0.94	0.92	0.92	0.78
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.87	0.86	0.67
		Probability of Chance	0.73	0.53	0.58	0.27
		Kappa (k)	0.73	0.72	0.67	0.55
		Classification Accuracy	0.95	0.91	0.90	0.76
	African-American	Classification Consistency (P)	0.88	0.91	0.96	0.75
		Probability of Chance	0.51	0.68	0.89	0.40
		Kappa (k)	0.75	0.72	0.63	0.58
		Classification Accuracy	0.91	0.94	0.97	0.82
	Hispanic	Classification Consistency (P)	0.88	0.87	0.92	0.68
		Probability of Chance	0.53	0.54	0.79	0.28
		Kappa (k)	0.75	0.71	0.63	0.56
		Classification Accuracy	0.92	0.91	0.94	0.77
	Asian	Classification Consistency (P)	0.90	0.88	0.90	0.69
		Probability of Chance	0.62	0.50	0.67	0.25
		Kappa (k)	0.73	0.75	0.71	0.58
		Classification Accuracy	0.93	0.91	0.93	0.77
	American Indian	Classification Consistency (P)	0.87	0.88	0.93	0.69
		Probability of Chance	0.52	0.56	0.81	0.29
		Kappa (k)	0.73	0.72	0.64	0.55
		Classification Accuracy	0.91	0.91	0.95	0.77
	Two or More	Classification Consistency (P)	0.90	0.87	0.89	0.68
		Probability of Chance	0.60	0.50	0.67	0.25
		Kappa (k)	0.75	0.75	0.67	0.57
		Classification Accuracy	0.93	0.91	0.92	0.77
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.86	0.95	0.68
		Probability of Chance	0.51	0.60	0.88	0.32
		Kappa (k)	0.72	0.66	0.55	0.53
		Classification Accuracy	0.90	0.91	0.96	0.77
Disability Status	Yes	Classification Consistency (P)	0.89	0.91	0.95	0.75
		Probability of Chance	0.50	0.61	0.83	0.35
		Kappa (k)	0.77	0.76	0.69	0.61
		Classification Accuracy	0.92	0.94	0.96	0.82
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.87	0.92	0.69
		Probability of Chance	0.53	0.54	0.78	0.28
		Kappa (k)	0.76	0.73	0.65	0.57
		Classification Accuracy	0.92	0.91	0.94	0.78

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-16 Indexes for Classification Consistency and Accuracy, Social Studies Grade 8

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.92	0.87	0.89	0.69
		Probability of Chance	0.67	0.50	0.67	0.26
		Kappa (k)	0.75	0.75	0.68	0.58
		Classification Accuracy	0.94	0.91	0.92	0.77
	Male	Classification Consistency (P)	0.92	0.88	0.90	0.70
		Probability of Chance	0.63	0.50	0.66	0.25
		Kappa (k)	0.77	0.77	0.70	0.60
		Classification Accuracy	0.94	0.92	0.92	0.78
Race/Ethnicity	White	Classification Consistency (P)	0.93	0.87	0.88	0.68
		Probability of Chance	0.74	0.52	0.62	0.27
		Kappa (k)	0.72	0.74	0.68	0.57
		Classification Accuracy	0.95	0.91	0.90	0.77
	African-American	Classification Consistency (P)	0.88	0.91	0.97	0.76
		Probability of Chance	0.50	0.69	0.91	0.39
		Kappa (k)	0.75	0.72	0.65	0.61
		Classification Accuracy	0.91	0.94	0.98	0.83
	Hispanic	Classification Consistency (P)	0.89	0.88	0.93	0.71
		Probability of Chance	0.55	0.55	0.81	0.28
		Kappa (k)	0.75	0.74	0.65	0.59
		Classification Accuracy	0.92	0.92	0.95	0.79
	Asian	Classification Consistency (P)	0.92	0.89	0.88	0.70
		Probability of Chance	0.68	0.50	0.64	0.26
		Kappa (k)	0.77	0.77	0.67	0.60
		Classification Accuracy	0.95	0.92	0.92	0.78
	American Indian	Classification Consistency (P)	0.88	0.87	0.95	0.71
		Probability of Chance	0.54	0.57	0.84	0.29
		Kappa (k)	0.75	0.71	0.66	0.59
		Classification Accuracy	0.92	0.91	0.96	0.79
	Two or More	Classification Consistency (P)	0.91	0.89	0.90	0.70
		Probability of Chance	0.62	0.50	0.70	0.26
		Kappa (k)	0.76	0.78	0.66	0.60
		Classification Accuracy	0.93	0.92	0.93	0.78
Limited English Proficiency	Yes	Classification Consistency (P)	0.86	0.92	0.99	0.76
		Probability of Chance	0.52	0.78	0.97	0.44
		Kappa (k)	0.70	0.62	0.47	0.57
		Classification Accuracy	0.90	0.94	0.99	0.83
Disability Status	Yes	Classification Consistency (P)	0.88	0.93	0.97	0.78
		Probability of Chance	0.51	0.71	0.90	0.42
		Kappa (k)	0.75	0.76	0.68	0.62
		Classification Accuracy	0.91	0.95	0.98	0.84
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.88	0.94	0.71
		Probability of Chance	0.53	0.56	0.82	0.29
		Kappa (k)	0.76	0.74	0.65	0.59
		Classification Accuracy	0.92	0.92	0.95	0.79

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Table J-17 Indexes for Classification Consistency and Accuracy, Social Studies Grade 10

Category	Group	Indexes	Cut 1	Cut 2	Cut 3	All Cuts
Gender	Female	Classification Consistency (P)	0.91	0.89	0.91	0.71
		Probability of Chance	0.60	0.50	0.67	0.25
		Kappa (k)	0.78	0.78	0.71	0.61
		Classification Accuracy	0.94	0.92	0.93	0.79
	Male	Classification Consistency (P)	0.91	0.90	0.92	0.74
		Probability of Chance	0.55	0.51	0.67	0.26
		Kappa (k)	0.81	0.80	0.75	0.65
		Classification Accuracy	0.94	0.93	0.94	0.81
Race/Ethnicity	White	Classification Consistency (P)	0.92	0.89	0.90	0.71
		Probability of Chance	0.64	0.50	0.63	0.25
		Kappa (k)	0.77	0.77	0.72	0.61
		Classification Accuracy	0.94	0.92	0.93	0.79
	African-American	Classification Consistency (P)	0.91	0.94	0.98	0.83
		Probability of Chance	0.57	0.76	0.92	0.51
		Kappa (k)	0.79	0.77	0.69	0.65
		Classification Accuracy	0.93	0.96	0.98	0.88
	Hispanic	Classification Consistency (P)	0.89	0.91	0.95	0.75
		Probability of Chance	0.50	0.59	0.82	0.32
		Kappa (k)	0.78	0.77	0.70	0.63
		Classification Accuracy	0.92	0.94	0.96	0.82
	Asian	Classification Consistency (P)	0.91	0.89	0.91	0.73
		Probability of Chance	0.59	0.50	0.65	0.25
		Kappa (k)	0.79	0.79	0.76	0.64
		Classification Accuracy	0.94	0.93	0.94	0.81
	American Indian	Classification Consistency (P)	0.89	0.89	0.95	0.74
		Probability of Chance	0.51	0.60	0.85	0.32
		Kappa (k)	0.78	0.73	0.69	0.62
		Classification Accuracy	0.92	0.93	0.96	0.81
	Two or More	Classification Consistency (P)	0.90	0.90	0.92	0.73
		Probability of Chance	0.55	0.51	0.70	0.26
		Kappa (k)	0.79	0.80	0.73	0.63
		Classification Accuracy	0.93	0.93	0.94	0.80
Limited English Proficiency	Yes	Classification Consistency (P)	0.89	0.96	0.99	0.84
		Probability of Chance	0.63	0.88	0.98	0.61
		Kappa (k)	0.69	0.68	0.60	0.60
		Classification Accuracy	0.92	0.97	0.99	0.89
Disability Status	Yes	Classification Consistency (P)	0.90	0.95	0.98	0.83
		Probability of Chance	0.57	0.77	0.91	0.52
		Kappa (k)	0.77	0.79	0.72	0.65
		Classification Accuracy	0.93	0.97	0.98	0.88
SES Disadvantaged	Yes	Classification Consistency (P)	0.89	0.91	0.95	0.76
		Probability of Chance	0.50	0.60	0.82	0.33
		Kappa (k)	0.79	0.78	0.70	0.63
		Classification Accuracy	0.92	0.94	0.96	0.82

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

Appendix K

Glossary

Copyright © 2019 by the Wisconsin Department of Public Instruction

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

Copyright © 2019 by the Wisconsin Department of Public Instruction

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 $_{2\,MH}\chi$) 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.

Copyright © 2019 by the Wisconsin Department of Public Instruction

WKCE: Wisconsin Knowledge and Concepts Examination. Previous Wisconsin assessment program.

Copyright © 2019 by the Wisconsin Department of Public Instruction