WISCONSIN KNOWLEDGE AND CONCEPTS EXAMINATIONS

FALL 2012 WKCE TECHNICAL REPORT

CTB/MCGRAW-HILL
20 RYAN RANCH ROAD
MONTEREY, CALIFORNIA 93940

Copyright

Developed and published under contract with the Wisconsin Department of Public Instruction by CTB/McGraw-Hill LLC, 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2013 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 *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Copyright © 2013 by the Wisconsin Department of Public Instruction

Acknowledgments

The following CTB/McGraw-Hill (CTB) and Wisconsin Department of Public Instruction (DPI) staff members are primarily responsible for the content and statistical quality of this report:

CTB:

Christina Schneider, Research Manager

Xinhui Xiong, Research Scientist

Soe Aung, Manager, Statistical Analyst

Kristina Wilson, Senior Research Associate

Anne Woods, Research Associate

Russ Wegfehrt, Project Manager, Publishing

Patricia Hildreth, Manager, Content Development

Tracy Keith, Scoring Project Manager

Tammy Bullock, Program Manager

Wisconsin DPI—Office of Student Assessment:

Lynette Russell, Assistant State Superintendent

Troy Couillard, Director

Phil Cranley, Education Consultant

Duane Dorn, Education Consultant

Viji Somasundaram, Education Consultant

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF CONTENTS

Copyright ii

Foreword iii

Acknowledgments iv

Part 1: Overview 1

Part 2: Test Design and Item Development 6

2.1 Content Framework and Assessment Limits 7
2.2 Test Blueprint 7
2.3 Reading Passage Selection 8
2.4 Item Development and Editing 8
2.5 Content/Bias Review and Item Alignment 8

Part 3: Test Form Development 9

3.1 Overall Test Book Development Process 9
3.1.1 WKCE Fall 2012 Form Selection 9
3.1.2 WKCE Field Test Item Selection 11
3.1.3 Quality Reviews 11
3.2 Description of the WKCE 2012 Tests 12
3.2.1 Reading 12
3.2.2 Mathematics 13
3.2.3 Language Arts 13
3.2.4 Social Studies 13
3.2.5 Science 13
3.3 Customer Approvals 13
3.3.1 Item Content and Bias Review 14
3.3.2 Item Selection Approval 14
3.3.3 First Pages Approvals 14
3.3.4 Second Pages Approvals 14
3.3.5 Final Pages Sign-Off 15

Part 4: Test Administration 16

4.1 Accommodations 16
4.2 Reporting Results of Assessments Taken with Accommodations 17
4.3 Test Security 17
4.4 Test Administration 19

Part 5: Scoring 21

5.1 Scoring of Multiple-Choice Items 21
5.1.1 Scanning and Calibration of Scanners 22
5.2 Scoring of Open-Ended or Constructed-Response Items 24
5.2.1 Description of Scoring Rubrics 24
5.2.2 Handscoring Process 24
5.2.3 Electronic Handscoring System 25
5.2.4 Anchor Papers and Training Papers 25
5.2.5 Scoring Personnel and Qualifications 25
5.2.6 Reader Training 26
5.2.7 Inter-Rater Reliability 26
5.3 Distribution of Constructed-Response Item Scores 27

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 6: Characteristics of the Calibration Sample **29**

Part 7: Calibration, Equating, and Deriving Scale Scores **30**

7.1 Calibration and Equating Methods 30
7.1.1 Calibration Models 31
7.1.2 Calibration Software 32
7.2 Calibration Results 33
7.2.1 IRT Item Parameters 33
7.2.2 IRT Item Fit 33
7.2.3 Evaluating Anchor Items 35
7.3 Deriving Scale Scores in the WKCE 36
7.3.1 Standard Error of Measurement 40
7.3.2 LOSS and HOSS 41
7.3.3 Test Characteristic Curves 42

Part 8: Test Results **43**

8.1 Classical Item Analysis: Item Level Statistics 43
8.1.1 Flagging for a Positive Distractor Correlation 49
8.1.2 Flagging for the Item-Total Correlation 50
8.1.3 Flagging for p-Value 50
8.1.4 Flagging for Omit Rate 50
8.1.5 Supplemental Tables on Classical Item Analysis 50
8.1.6 Speededness 51
8.2 Raw Score Results 51
Reading 53
Mathematics 53
Language Arts 53
Social Studies 53
Science 53
Subgroup Performance Patterns in Raw Score Results 54
8.3 Summary Statistics for Scale Scores 55
Reading 55
Mathematics 56
Language Arts 56
Social Studies 56
Science 56
Subgroup Performance Patterns in Scale Score Results 56
Gender 57
Race/Ethnicity 57
Socioeconomic Status 57
Disability Status 58
English Language Proficiency 58
8.4 Cut Scores and Performance Level Classifications 58
Reading 60
Mathematics 60
Language Arts 60
Social Studies 60
Science 61
Subgroup Patterns in Performance Level Results 61
8.5 Standard Performance Indicators for Content Standards 62
Reading 63
Mathematics 63
Language Arts 64
Social Studies 64
Science 64

Copyright © 2013 by the Wisconsin Department of Public Instruction

Summary of Student Performance Indicator Results 64
Summary of Student Achievement Results 65

Part 9: Reliability 66

9.1 Measures of Internal Consistency and SEM 67
9.2 Classification Consistency and Accuracy 69
9.2.1 Kolen and Kim's Method for Pattern Scoring 71
9.3 Inter-Rater Reliability for CR Items and Writing Prompts 73
Reading 76
Mathematics 77
Writing 77
Summary 78

Part 10: Validity 79

10.1 Differential Item Functioning 81
Results 86
10.2 Construct Validity 88
10.3 Test Integrity: Erasure Analysis 89
10.4 Standardized Test Administration 90

Part 11: Summary Recommendations 91

Fall 2012 WKCE Technical Report: References 92

Fall 2012 Tables and Figures 96

Appendices

Appendix 1: Fall 2010 Item Selection Check-Off Form 321

Appendix 2: Fall 2012 WKCE Assessment Accommodations Matrix 334

Appendix 3: WKCE Glossary 340

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF TABLES

PART 2
Table 2-1 Target Reading Test Blueprint: Grades 3–8, 10* 97
Table 2-2 Target Mathematics Test Blueprint: Grades 3–8, 10* 98
Table 2-3 Target Language Arts Test Blueprint: Grades 4, 8, 10 99
Table 2-4 Target Social Studies Test Blueprint: Grades 4, 8, 10 99
Table 2-5 Target Science Test Blueprint: Grades 4, 8, 10 99
Table 2-6 Actual Reading Test Blueprint: Grades 3–8, 10* 100
Table 2-7 Actual Mathematics Test Blueprint: Grades 3–8, 10* 101
Table 2-8 Actual Language Arts Test Blueprint: Grades 4, 8, 10 102
Table 2-9 Actual Social Studies Test Blueprint: Grades 4, 8, 10 102
Table 2-10 Actual Science Test Blueprint: Grades 4, 8, 10 102
Table 2-11 Item Development Each Year and Total to Date* 103

PART 3
Table 3-1 Fall 2012 Test Configuration 104
Table 3-2 Unique Items Field Tested Each Year and Total to Date 105

PART 4
Table 4-1 Test Accommodations 106

PART 5
Table 5-1 Reading Rubric, Grades 3–8 and 10 117
Table 5-2 Mathematics Rubric, Grades 3–8 and 10 118
Table 5-3 Writing Rubric, Conventions of Written English, Grade 4 119
Table 5-4 Writing Rubric, Composing, Grade 4 120
Table 5-5 Writing Rubric, Conventions of Written English, Grade 8 121
Table 5-6 Writing Rubric, Composing, Grade 8 122
Table 5-7 Writing Rubric, Conventions of Written English, Grade 10 124
Table 5-8 Writing Rubric, Composing, Grade 10 125
Table 5-9 Score Distribution for Reading CR Items* 127
Table 5-10 Score Distribution for Mathematics CR Items* 128
Table 5-11 Score Distribution for Grades 4, 8, and 10 Writing: Composing Rubric 130
Table 5-12 Percentage Distribution of Scores, Grades 4, 8, and 10 Writing: Composing Rubric 130
Table 5-13 Score Distribution, Grades 4, 8, and 10 Writing: Conventions Rubric 131
Table 5-14 Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Conventions Rubric 131
Table 5-15 Score Distribution for Grades 4, 8, and 10 Writing 132
Table 5-16 Percentage Distribution of Scores for Grades 4, 8, and 10 Writing 132

PART 7
Table 7-1 Item Flagged Based on Yen's Q_1 133
Table 7-2 Scoring Table for Reading Grade 3 134
Table 7-3 Scoring Table for Reading Grade 4 135
Table 7-4 Scoring Table for Reading Grade 5 136
Table 7-5 Scoring Table for Reading Grade 6 137
Table 7-6 Scoring Table for Reading Grade 7 138
Table 7-7 Scoring Table for Reading Grade 8 139
Table 7-8 Scoring Table for Reading Grade 10 140
Table 7-9 Scoring Table for Mathematics Grade 3 141
Table 7-10 Scoring Table for Mathematics Grade 4 142

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

Table 7-11 Scoring Table for Mathematics Grade 5....143
Table 7-12 Scoring Table for Mathematics Grade 6....144
Table 7-13 Scoring Table for Mathematics Grade 7....145
Table 7-14 Scoring Table for Mathematics Grade 8....146
Table 7-15 Scoring Table for Mathematics Grade 10....147
Table 7-16 Scoring Table for Language Arts Grade 4....148
Table 7-17 Scoring Table for Language Arts Grade 8....149
Table 7-18 Scoring Table for Language Arts Grade 10....150
Table 7-19 Scoring Table for Social Studies Grade 4....151
Table 7-20 Scoring Table for Social Studies Grade 8....152
Table 7-21 Scoring Table for Social Studies Grade 10....153
Table 7-22 Scoring Table for Science Grade 4....154
Table 7-23 Scoring Table for Science Grade 8....155
Table 7-24 Scoring Table for Science Grade 10....156
Table 7-25 The Number of Students and Percents at LOSS and HOSS....157

PART 8
Table 8-1 Item Analysis Grade 3 Reading....158
Table 8-2 Item Analysis Grade 4 Reading....161
Table 8-3 Item Analysis Grade 5 Reading....164
Table 8-4 Item Analysis Grade 6 Reading....167
Table 8-5 Item Analysis Grade 7 Reading....170
Table 8-6 Item Analysis Grade 8 Reading....173
Table 8-7 Item Analysis Grade 10 Reading....176
Table 8-8 Item Analysis Grade 3 Mathematics....179
Table 8-9 Item Analysis Grade 4 Mathematics....182
Table 8-10 Item Analysis Grade 5 Mathematics....185
Table 8-11 Item Analysis Grade 6 Mathematics....188
Table 8-12 Item Analysis Grade 7 Mathematics....191
Table 8-13 Item Analysis Grade 8 Mathematics....194
Table 8-14 Item Analysis Grade 10 Mathematics....197
Table 8-15 Item Analysis Grade 4 Language Arts....200
Table 8-16 Item Analysis Grade 8 Language Arts....202
Table 8-17 Item Analysis Grade 10 Language Arts....204
Table 8-18 Item Analysis Grade 4 Social Studies....206
Table 8-19 Item Analysis Grade 8 Social Studies....208
Table 8-20 Item Analysis Grade 10 Social Studies....210
Table 8-21 Item Analysis Grade 4 Science....212
Table 8-22 Item Analysis Grade 8 Science....214
Table 8-23 Item Analysis Grade 10 Science....216
Table 8-24 The Number of Items Flagged....218
Table 8-25 Raw Score Descriptive Statistics based on Census Data....219
Table 8-26 Raw Score Descriptive Statistics by Gender....220
Table 8-27 Raw Score Descriptive Statistics for Reading by Race/Ethnicity....221
Table 8-28 Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity....222
Table 8-29 Raw Score Descriptive Statistics for Language Arts by Race/Ethnicity....223
Table 8-30 Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity....223
Table 8-31 Raw Score Descriptive Statistics for Science by Race/Ethnicity....224
Table 8-32 Raw Score Descriptive Statistics by Socioeconomic Status....225
Table 8-33 Raw Score Descriptive Statistics by Disability....226
Table 8-34 Raw Score Descriptive Statistics by English Language Proficiency....227
Table 8-35 2012 and 2010* Scale Score Mean and Standard Deviation....228

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

Table 8-36 Scale Score Descriptive Statistics 229
Table 8-37 Scale Score Descriptive Statistics by Gender 230
Table 8-38 Scale Score Descriptive Statistics for Reading by Race/Ethnicity 231
Table 8-39 Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity 232
Table 8-40 Scale Score Descriptive Statistics for Language Arts by Race/Ethnicity 233
Table 8-41 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity 234
Table 8-42 Scale Score Descriptive Statistics for Science by Race/Ethnicity 234
Table 8-43 Scale Score Descriptive Statistics by Socioeconomic Status 235
Table 8-44 Scale Score Descriptive Statistics by Disability 236
Table 8-45 Scale Score Descriptive Statistics by English Language Proficiency 237
Table 8-46 Performance Level Cut Scores for all Contents* 238
Table 8-47 Percentage of Students in Each Performance Level by Sub-Group (Reading) 239
Table 8-48 Percentage of Students in Each Performance Level by Sub-Group (Mathematics) 242
Table 8-49 Percentage of Students in Each Performance Level by Sub-Group (Language Arts) 245
Table 8-50 Percentage of Students in Each Performance Level by Sub-Group (Social Studies) 246
Table 8-51 Percentage of Students in Each Performance Level by Sub-Group (Science) 247
Table 8-52 Cut Scores and Associated Impact Data for WKCE Reading 248
Table 8-53 Cut Scores and Associated Impact Data for WKCE Mathematics 248
Table 8-54 Cut Scores and Associated Impact Data for WKCE Language Arts 249
Table 8-55 Cut Scores and Associated Impact Data for WKCE Social Studies 249
Table 8-56 Cut Scores and Associated Impact Data for WKCE Science 250
Table 8-57 Summary Statistics for Reading Content Standards Raw and SPI Scores 251
Table 8-58 Summary Statistics for Mathematics Content Standards Raw and SPI Scores 253
Table 8-59 Summary Statistics for Language Arts Content Standards Raw and SPI Scores 255
Table 8-60 Summary Statistics for Social Studies Content Standards Raw and SPI Scores 256
Table 8-61 Summary Statistics for Science Content Standards Raw and SPI Scores 257
Table 8-62 SPI Cut Scores 258

PART 9

Table 9-1 Reliability for Total Group and Subgroups Using Cronbach's Alpha 261
Table 9-2 Standard Error of Measurement for Total Group and Subgroups 262
Table 9-3 Cronbach's Alpha Reliability Coefficients for Content Standards 263
Table 9-4 Standard Error of Measurement per Content Standard 264
Table 9-5 Classification Consistency and Classification Accuracy for Reading Grade 3 265
Table 9-6 Classification Consistency and Classification Accuracy for Reading Grade 4 266
Table 9-7 Classification Consistency and Classification Accuracy for Reading Grade 5 267
Table 9-8 Classification Consistency and Classification Accuracy for Reading Grade 6 268
Table 9-9 Classification Consistency and Classification Accuracy for Reading Grade 7 269
Table 9-10 Classification Consistency and Classification Accuracy for Reading Grade 8 270
Table 9-11 Classification Consistency and Classification Accuracy for Reading Grade 10 271
Table 9-12 Classification Consistency and Classification Accuracy for Mathematics Grade 3 272
Table 9-13 Classification Consistency and Classification Accuracy for Mathematics Grade 4 273
Table 9-14 Classification Consistency and Classification Accuracy for Mathematics Grade 5 274
Table 9-15 Classification Consistency and Classification Accuracy for Mathematics Grade 6 275
Table 9-16 Classification Consistency and Classification Accuracy for Mathematics Grade 7 276
Table 9-17 Classification Consistency and Classification Accuracy for Mathematics Grade 8 277
Table 9-18 Classification Consistency and Classification Accuracy for Mathematics Grade 10 278
Table 9-19 Classification Consistency and Classification Accuracy for Language Arts Grade 4 279
Table 9-20 Classification Consistency and Classification Accuracy for Language Arts Grade 8 280
Table 9-21 Classification Consistency and Classification Accuracy for Language Arts Grade 10 281
Table 9-22 Classification Consistency and Classification Accuracy for Social Studies Grade 4 282
Table 9-23 Classification Consistency and Classification Accuracy for Social Studies Grade 8 283

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

Table 9-24 Classification Consistency and Classification Accuracy for Social Studies Grade 10 284
Table 9-25 Classification Consistency and Classification Accuracy for Science Grade 4 285
Table 9-26 Classification Consistency and Classification Accuracy for Science Grade 8 286
Table 9-27 Classification Consistency and Classification Accuracy for Science Grade 10 287
Table 9-28 Inter-Rater Reliability, Reading* 288
Table 9-29 Inter-Rater Reliability, Mathematics* 289
Table 9-30 Inter-Rater Reliability, Writing Prompts* 291

PART 10
Table 10-1 Items Flagged for DIF, By Gender 292
Table 10-2 Items Flagged for DIF, By Race/Ethnicity, African American 293
Table 10-3 Items Flagged for DIF, By Race/Ethnicity, Hispanic 294
Table 10-4 Items Flagged for DIF, By Race/Ethnicity, Asian 295
Table 10-4 Cont'd Items Flagged for DIF, By Race/Ethnicity, Asian 296
Table 10-5 Items Flagged for DIF, By Race/Ethnicity, American Indian* 297
Table 10-6 Items Flagged for DIF, By English Language Proficiency 298
Table 10-6 Cont'd Items Flagged for DIF, By English Language Proficiency 299
Table 10-7 Items Flagged for DIF, By Disability Status 300
Table 10-8 Correlations among Reading Objectives 301
Table 10-9 Correlations among Mathematics Objectives 302
Table 10-10 Correlations among Language Arts Objectives 303
Table 10-11 Correlations among Social Studies Objectives 303
Table 10-12 Correlations among Science Objectives 304
Table 10-13 Principal Components Analysis 305

Copyright © 2013 by the Wisconsin Department of Public Instruction

TABLE OF FIGURES

PART 7

Figure 7-1 SEM Curves, Reading Grades 3-6 306
Figure 7-2 SEM Curves, Mathematics Grades 3-6 308
Figure 7-3 SEM Curves, Language Arts Grades 4, 8, 10 310
Figure 7-4 SEM Curves, Social Studies Grades 4, 8, 10 311
Figure 7-5 SEM Curves, Science Grades 4, 8, 10 312
Figure 7-6 TCC Curve for Reading Grades 3-8, 10 313
Figure 7-7 TCC Curve for Mathematics Grades 3-8, 10 314
Figure 7-8 TCC Curve for Language Arts Grades 4, 8, 10 315
Figure 7-9 TCC Curve for Social Studies Grades 4, 8, 10 316
Figure 7-10 TCC Curve for Science Grades 4, 8, 10 317

PART 9

Figure 9-1 Reading Indices for Classification Consistency and Classification Accuracy 318
Figure 9-2 Mathematics Indices for Classification Consistency and Classification Accuracy 318
Figure 9-3 Language Arts Indices for Classification Consistency and Classification Accuracy 319
Figure 9-4 Social Studies Indices for Classification Consistency and Classification Accuracy 319
Figure 9-5 Science Indices for Classification Consistency and Classification Accuracy 320

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 1: Overview

The *Fall 2012 Wisconsin Knowledge and Concepts Examinations (WKCE) Technical Report* documents the processes and procedures applied in the Fall 2012 WKCE as well as the results. This report also documents processes, procedures, and results of this administration to support validity and reliability evidence for 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], 1999). This report demonstrates that the Fall 2012 WKCE adhered to the appropriate standards and practices of educational assessment. Ultimately, this report serves to document evidence that valid inferences about Wisconsin student performance can be derived from this assessment.

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* spell 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 norm-referenced assessments in the 1997 school year. The assessments used at that time were *TerraNova*™ tests developed by CTB. 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 were adopted. These new standards were the work of the Governor's Commission on Wisconsin Model Academic Standards, chaired by then-current Lieutenant Governor McCallum and the Wisconsin DPI. The Wisconsin Model Academic Standards 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, has been used to determine the Adequate Yearly Progress (AYP) of students at the school, district, and state levels.

Beginning with school year 2007–08, states were also required to administer Science assessments at least once during grades 3–5, grades 6–9, and grades 10–12. Wisconsin students in grades 4, 8, and 10 are, and will continue to be, assessed in Language Arts, Science, and Social Studies as required by state law (§ 118.30 Wisconsin Statutes).

It is 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 Reading and Mathematics tests. The criterion-referenced WKCE is designed specifically for Wisconsin students, and specifically to measure their performance on the Wisconsin Model Academic Standards adopted by the state. These assessments are designed to evaluate students'

Copyright © 2013 by the Wisconsin Department of Public Instruction

knowledge and to measure achievement in the basic skills taught in schools at grades 3–8 and 10. The Fall 2012 WKCE is the seventh administration of these assessments.

The Wisconsin Model Academic Standards describe what students should know and be able to do in grades 4, 8, and 12. To determine what should be tested in grades 3, 5–7, and 10, committees of Wisconsin educators carefully considered what knowledge and skills students should have by the fall of each school year by extrapolating and interpolating the standards for grades 4, 8, and 12. The committees then defined the eligible test content and assessment limits, ensuring that the test framework they designed incorporated the content and performance standards enumerated in the Wisconsin Model Academic Standards. Therefore, the assessment framework, used to define what is tested on the WKCE, reflects what students should have learned by the beginning of the school year in order to be successful in that grade. As a result, the grade 6 test, for example, assesses what students should have learned by the end of grade 5.

The WKCE tests consist of criterion-referenced items, written by CTB and edited and reviewed by Wisconsin teachers, and items from CTB's norm-referenced test, *TerraNova*, The Second Edition (*TerraNova*, CTB/McGraw-Hill, 2001). The Fall 2012 WKCE tests include Reading and Mathematics at grades 3–8 and 10 and Science, Social Studies, and Language Arts (including Writing) at grades 4, 8, and 10.

Based on the input of Wisconsin educators and the Wisconsin Model Academic Standards, a design was derived for the development, administration, and scoring of the WKCE. The present Technical Report documents all aspects of the testing cycle in the subsequent chapters. The structure of the present Technical Report mirrors the testing cycle. A brief content summary of the report is provided below.

Test Design and Item Development

- Part 2 of this report describes test design, the item development process, and some aspects of the content-related validity of the WKCE tests.
- More specifically, Part 2 describes how CTB, DPI, and Wisconsin educators collaborated through a series of test development processes to ensure that the appropriate content was included in the WKCE and to ensure that the test items adequately sampled the domain of content knowledge necessary to make legitimate inferences about student performance.
- Wisconsin Model Academic Standards were translated into grade-level content frameworks, which in turn formed the basis for test blueprints and item specifications.
- Wisconsin educators were involved in design at every step to ensure the appropriateness of the test to the standards.
- Test design started in August 2003 with the convention of approximately 35 educators per content area for grades 3–8 and 10 to establish the grade-level content frameworks based on the Wisconsin Model Academic Standards, establish assessment limits, create the test blueprint, and review reading passage and page specifications. The test specifications documents created and later approved by DPI continue to serve as a foundation for item and test development.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Test Form Development

- Part 3 discusses key development tasks and issues related to creating the Fall 2012 WKCE test forms.
- Item development was based on the approved test blueprints, with a sufficient quantity of items written across years to develop multiple operational test forms.
- Part 3 discusses the process of selecting operational test items and the process of obtaining DPI approvals.
- As detailed in Part 3, there have been 5,025 unique multiple-choice (MC) items and 499 unique constructed-response (CR) items field tested to date, that is, through the Fall of 2012, totaling 5,524 unique items.
- Selection of the Fall 2012 operational forms was done using the ITEMWIN (Burket, 2000) software utilizing methods similar to previous administrations for all grades and content areas.

Test Administration

- Part 4 briefly describes test administration and accommodations.
- The test administration window was October 22–November 23, 2012.
- Delivery of materials was handled through the district and school assessment coordinators.
- In 2012, all content area tests in a grade were administered to students using a single test book.

Scoring

- Part 5 documents the scanning and scoring process for the MC and CR items.
- The machine-scanning process and the handscoring process, including the development and review of the scoring rubrics, anchor (sample) papers, and writing prompts, as well as the training of scoring personnel, ongoing quality assurance, the application of an inter-rater reliability assessment, and a systematic review of the resulting score distributions supported reporting reliable and valid test scores.
- The scoring rubrics used in handscoring are presented in detail for all content areas with handscored items.

Characteristics of the Calibration Sample

- The calibration and equating of the Fall 2012 WKCE tests occurred during the 2010 WKCE administration for all grades and content areas except Reading grade 10. The calibration and equating for Reading grade 10 occurred during the 2008 WKCE administration. The calibration was based on a sample of student response data termed the calibration sample. The calibration sample roughly approximates the

Copyright © 2013 by the Wisconsin Department of Public Instruction

census population for minority students and under-represents majority students, as has been the historical practice.

Calibration, Equating, and Deriving Scale Scores

- Part 7 reviews calibration, equating, and scoring methods implemented for WKCE.
- The Fall 2012 WKCE was calibrated and scaled using two different item response theory (IRT) models, one for CR 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 include model-to-data fit and the standard error of measurement (SEM).
- Item-pattern scoring was applied to the Fall 2012 WKCE. As discussed in Part 7, item-pattern scoring is generally recommended over number-correct scoring because it produces more accurate scores for individual students.
- Part 7 also explains how a student's scale score is derived from the raw score using item-pattern scoring. Examples of a very low-performing student, a very high-performing student, and several students with a 50% correct raw score are provided. Several students with the same 50% correct raw score are provided in order to illustrate how students with the same raw score can have different scale scores.

Test Results

- Part 8 summarizes item analyses, raw scores, scale scores, performance levels, and a standard performance indicator score for content standards.
- Reliability of the WKCE tests are reported using Cronbach's alpha and SEM.
- Summary descriptive statistics for all scores (raw scores, scale scores, standard performance indicator scores, and performance levels) are reported for all students and for subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency.

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 estimated for all CR items.
- Internal consistency was assessed for all MC and CR items using Cronbach's alpha.
- Classification consistency and accuracy were estimated for performance classification.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Validity

- Part 10 reviews the validity evidence presented in all prior parts and provides additional validity evidence supporting the WKCE tests.
- Factor analysis and correlations among content standards are presented in the context of construct validity.
- An analysis of differential item functioning (DIF) is presented.
- Erasure analysis, a procedure used to identify high erasure rates, is also discussed.

Summary Recommendations

- Key findings of the Fall 2012 WKCE administration are presented in the body of the report. However, some items of a more technical nature, which stand out as key recommendations and summary statements that should be considered in subsequent administrations, are presented in Part 11.
- Recommendations based on the Fall 2012 WKCE administration cover three different phases of the testing cycle: item development, scoring, and psychometric, or measurement-based, research and evaluation.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 2: Test Design and Item Development

The purpose of this section is to describe how CTB, DPI, and Wisconsin educators collaborated through a series of test development processes to ensure that appropriate content was included in the WKCE, 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 development process for the Fall 2012 WKCE administration.

As described below, the Wisconsin Model Academic Standards were central to the entire test design process. Part 2 of the Technical Report demonstrates the adherence of the WKCE program to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) and specifically to standards 1.2, 1.6, 3.1, 3.2, 3.3, 3.11, 6.4, 6.15, 13.3, and 13.5.

Operational items administered in 2012 adhered to the test specifications documents developed in previous years. The Fall 2006 Technical Report (Parts 2, 3, and 4) provides a detailed account of the development of the test specifications documents during previous years. Interested readers can find these sections of the Fall 2006 Technical Report in Appendix 2 of the 2010 WKCE Technical Report available online at: http://oea.dpi.wi.gov/files/oea/pdf/td-2010-techman.pdf. The assessment frameworks, test design, test blueprints, reading passage specifications, item specifications, art specifications, and style guide were all developed in 2003, the first year of the WKCE program. The role of Wisconsin educators was an essential component of the development of the WKCE. Their professional expertise and judgment were central to approving content that was appropriately rigorous for the grade and content area in which it was presented and that was expected to have been taught to students.

During the first year of the contract, August 2003 to August 2004, the test specifications documents were developed through an extended, collaborative process with DPI and based on the contributions of Wisconsin educators during meetings conducted in 2003 (see the Fall 2006 Technical Report, p. 6, which is provided in Appendix 2 of the 2010 WKCE Technical Report available online at: http://oea.dpi.wi.gov/files/oea/pdf/td-2010-techman.pdf). Test specification documents include the test blueprints, passage specifications, item specifications, page specifications, and style guide.

According to the most recent edition of the standards (AERA, APA, NCME, 1999), "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of test scores" (p. 9). Much of the content-related validity evidence is produced during the test development process. The content-related evidence supports inferences from a sample of observations (the test) to a domain of observations (the content area). A substantial source of content-related validity evidence is the expert judgment that the test items are an adequate and representative sample of the domain being measured. Content-related validity evidence can support interpretations of test scores in terms of performance over a performance domain. If the content domain is specified clearly and a representative sample of performance tasks is drawn from the domain, then inferences about expected performance over the domain based on observed performances should be legitimate.

Copyright © 2013 by the Wisconsin Department of Public Instruction

2.1 Content Framework and Assessment Limits

The Assessment Framework documents created by DPI provide information about the content measured at each grade level and explain the relationships among the Wisconsin Model Academic Standards, the Assessment Framework, and classroom instruction. The Framework documents are located on DPI's website at http://oea.dpi.wi.gov/oea_wkce-crt. The Fall 2006 Technical Report, Section 3.1.1, explains the structure and development of the Assessment Frameworks (see Appendix 2 of the 2010 WKCE Technical Report available online at: http://oea.dpi.wi.gov/files/oea/pdf/td-2010-techman.pdf).

The Assessment Frameworks specify the broad categories within the content area at which test sub-scores may be reported. For example, "Number Operations and Relations" or "Measurement" for Mathematics and "Understands Text" or "Analyzes Text" for Reading. These broad categories are further delineated into subskills. For example, "Number Operations and Relations" is further subdivided into "Reading, Writing, and Representing Numbers" and "Ordering and Comparing Numbers" and so forth. Assessment limits are bulleted statements that identify the specific content that is eligible for testing for each subskill and may clarify how the content could be assessed. For example, in Mathematics, the size of numbers or the types of plane and solid geometric figures that are appropriate at each grade level would be specified in the assessment limits. For Reading, the assessment limits clarify which prefixes or suffixes or which literary devices are appropriate to assess at each grade level. For the grade 4, 8, and 10 Science assessments, the Wisconsin Model Academic Standards served as the foundation for the creation of the Science Assessment Frameworks. Similarly, the Wisconsin Model Academic Standards for Language Arts and Social Studies provide the content framework for these content area tests at grades 4, 8, and 10.

2.2 Test Blueprint

The test blueprints specify the number of multiple-choice (MC) and constructed-response (CR) items for each reporting category and subskill. The process used for developing the blueprints was described in detail in Parts 2 and 3 of the Fall 2006 Technical Report (see Appendix 2 of the 2010 WKCE Technical Report available online at: http://oea.dpi.wi.gov/files/oea/pdf/td-2010-techman.pdf). Tables 2-1 through 2-5 present the target blueprints for the Fall 2012 WKCE. Tables 2-6 through 2-10 present the actual test blueprints showing how the items selected for the Fall 2012 WKCE forms were distributed by reporting category and subskill for each item type.

In 2007, some changes were made to the blueprints for Mathematics, Science, and Language Arts grade 8. The Mathematics blueprints were modified to reflect the inclusion of a 2-point CR item and the subsequent reduction of the 3-point CR items from four to three. In addition, the number of MC items for each reporting category was adjusted to reflect the use of MC items for reporting category A. The Science blueprints were modified slightly to show a shift in emphasis among reporting categories A and B and among reporting categories G and H. The Language Arts grade 8 blueprint changes involved shifting two MC items from reporting category D to reporting category B. This change was made in response to Wisconsin educators'

Copyright © 2013 by the Wisconsin Department of Public Instruction

concerns expressed at the 2005 content review that the language test should not require excessive reading. When selecting test forms for 2005 and 2006, CTB made an effort to minimize the number of item sets that use a common stimulus, such as a brief essay or letter. However, when selecting the 2008 forms, the use of two lengthy stimuli would have been necessary in order to meet the blueprint. CTB brought this concern to the attention of DPI and suggested that two items be shifted from category D to category B. DPI approved this change to the blueprint on March 9, 2007.

In addition to the changes above, Depth of Knowledge (DOK) requirements were incorporated into the Reading and Mathematics blueprints to indicate the number of items needed at each DOK level for each reporting category.

2.3 Reading Passage Selection

Reading passages on the 2012 operational [1] forms were selected, reviewed, and approved between 2001 and 2008. The processes used for selecting, reviewing, and approving WKCE Reading passages were detailed in Section 3.1.3 of the Fall 2006 Technical Report (see Appendix 2 of the 2010 WKCE Technical Report available online at: http://oea.dpi.wi.gov/files/oea/pdf/td-2010-techman.pdf).

2.4 Item Development and Editing

While historically new items have been developed each year for the WKCE, in 2012 new items were not developed. Table 2-11 shows the number of MC, CR, and total items that have been written up to 2012 for the WKCE.

2.5 Content/Bias Review and Item Alignment

Because there were no field test items on the 2012 test forms, content and bias reviews for field test items did not occur.

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 3: Test Form Development

Part 3 of the Technical Report focuses on key development tasks and issues related to creating the Fall 2012 WKCE operational test forms. The test specifications and item development activities described in Part 2 explain how specific development processes provided evidence to support test validity, primarily content validity, through the use of expert professional judgment from Wisconsin educators and from CTB test development specialists. The foundation test specifications documents—assessment frameworks, assessment limits, passage specifications, item specifications, test blueprints, art and page specifications, and style guide—developed and approved during the initial phases of the project served as critical guides throughout development and field testing of items. These documents contributed to ensuring that each form of the test accurately measured the content in consistent and stable ways, thus providing evidence supporting the test's use as an indicator of student achievement of state standards. Information is provided in Part 3 relating to the following topics:

- a general discussion of CTB's test book creation and editing process
- the process of selecting operational test items
- the process of obtaining DPI approvals

A comprehensive, multi-segment development process guides the development of assessment materials. The following section outlines this process in general terms. The remainder of Part 3 provides details of how these processes were implemented in Wisconsin. This section of the Technical Report addresses the following AERA/APA/NCME (1999) standards: 1.6, 3.1, 3.5, 3.6, 3.7, 3.9, 3.11, 3.16, 6.4, 6.15, 7.3, 7.4, 7.7, 13.3, and 13.5.

3.1 Overall Test Book Development Process

The creation of test book materials involved the expertise of multiple CTB departments, DPI, and Wisconsin educators. The activities that contributed to the creation of the test book materials are described below.

3.1.1 WKCE Fall 2012 Form Selection

The WKCE operational test forms for all content areas and grade levels use the common item non-equivalent groups design in order to equate parallel test forms from year to year. The minimum number of common items (also called anchor items) per content area follows:

- Reading: 14 items
- Mathematics: 18 items
- Language Arts: 15 items
- Social Studies: 15 items
- Science: 15 items

Copyright © 2013 by the Wisconsin Department of Public Instruction

CTB assessment editors selected items for the 2012 operational forms while considering a variety of criteria, including the following:

- Selected items must fully cover the reporting areas of the test blueprint.
- Selected items must represent the diversity of content.
- MC items with *p*-values below 0.30 should be avoided when possible.
- CR items with *p*-values below 0.20 should be avoided when possible.
- Items with positive point-biserials on distractors should be avoided when possible.
- Items should represent a range of scale score values.
- Items with differential item functioning (DIF) flags (C flags) should be avoided when possible.
- Items with poor fit flags should be avoided when possible.

CTB content editors used CTB's proprietary software, called ITEMWIN (Burket, 2000), to select items for the Fall 2007 (and 2008) WKCE operational test forms for all content areas and grade levels. These 2007 (2008) test forms were re-administered in 2010 and again in 2012 with no modifications made to the operational selection. The form was, however, recalibrated in 2010. (For Reading grade 10, the 2008 test form was re-administered.) ITEMWIN has two phases. In the first phase, CTB uses ITEMWIN to select a working item pool of manageable size from the larger tryout pool; items clearly inappropriate to the target grade range are eliminated. There is information about each item in the pool, including the item format to which the item is assigned, a descriptive phrase about the item, the association of the item with a stimulus, the item parameters, a fit rating indicating how well the item fits the expectations based on the IRT model used, and a DIF rating indicating whether the probability of answering the item correctly by students of equal ability differed by a particular group or category, such as gender, race/ethnicity, socioeconomic status, disability status, or English language proficiency. DIF is discussed further in Part 10.

ITEMWIN shows tables with both the expected number correct and standard error of measurement (SEM) as functions of scale score, as well as statistical and graphical summaries of DIF, fit, and the average standard error of the test as selected. Any fault in the selection, whether the test is too easy or too difficult for the target grade, contains items showing DIF, or does not adequately cover part of the scale score range, becomes apparent as the final statistics are generated. CTB assessment editors and the CTB Research team examined these statistics for each of the WKCE selections against those of the previous operational form to confirm that each new form was parallel in difficulty to the previous operational form. In addition, the vertical properties of tests were assessed by CTB and DPI through a visual inspection of the test characteristic curves for all grades when they are plotted side-by-side, where appropriate. Finally, CTB assessment editors reviewed each selection for content diversity to ensure that no two items were similar in content.

CTB assessment editors prepared a detailed document for each selected form that summarized the test and item characteristics, submitted their selections to a content supervisor for review, and in some cases to the Content Development Lead. Appendix 1 shows the Form Selection Summary Document. The supervisor and/or manager requested changes to the

Copyright © 2013 by the Wisconsin Department of Public Instruction

selections, as necessary, in order to improve the test characteristic curve or standard error curve. Form selections were then submitted to the CTB Research team for review. Additional revisions may have been requested at this stage. For the Reading and Mathematics selections, it was important to ensure the test characteristic curves for all grade levels formed a progression. The CTB Research team reviewed the form selections to ensure the test characteristic curves for the 2007 forms (and the 2008 form for Reading grade 10) that were also used in 2012 were as similar as possible to the 2006 (and 2007) forms and that curves for the anchor items were aligned closely to the test forms.

Upon approval of the selections by the CTB Research team in 2007 (and 2008), the CTB assessment editors submitted the selections to DPI for review. For some selections, DPI requested revisions for content, difficulty, or statistical reasons. Upon making the requested changes and submitting revised selection summary forms, all operational forms were approved by DPI. For 2012, DPI reviewed the 2007 (and 2008) forms and accompanying statistics and approved the re-use of these forms. Table 3-1 shows the structure of operational test forms in the Fall 2012 WKCE.

3.1.2 WKCE Field Test Item Selection

No items were field tested in 2012. Table 3-2 shows the number of items that were field tested up to 2009.

3.1.3 Quality Reviews

A smooth test administration requires that all test materials, including test books, manipulatives, and test administration manuals, align with each other. All items, page numbers, and administration times must be accurate in all components of the test program. When materials are not in alignment, not only can rework and additional costs be incurred, but there is also the possibility of jeopardizing the validity of test results and creating poor publicity. Therefore, to help ensure all documents required for the administration of a test are in alignment with each other, a Materials Integration Review (MIR) is conducted prior to moving the materials on to the Quality Assurance (QA) Department within CTB.

During the MIR, a proctor simulated the test experience by administering the test to two test takers for each grade and content area using the WKCE examiner's manual. The purpose of this review is twofold: to ensure the test materials are in alignment with each other and to verify the answer keys are correct.

In addition, a QA review was conducted on each test book and all ancillary materials. The purpose of the QA review is to ensure all publishable products meet the standards and expectations of DPI. The QA review includes, but is not limited to, the review for page number location/order, header/footer information, "go on" and "stop" signs, item sequence numbering, accuracy of directions, vertical and horizontal alignment, conventions of written English, clarity/accuracy of art, accuracy of cross-references, and that there is only one correct answer to

Copyright © 2013 by the Wisconsin Department of Public Instruction

each item. This QA review occurred at the end of the page production cycle and prior to releasing the materials to CTB's Manufacturing Department.

In addition to the MIR and QA review steps, the WKCE test books were reviewed by CTB's Technology Department to verify the scannable test books were constructed to meet CTB's scanning and scoring specifications. With each round of page production, CTB's Production Department staff viewed the position of answer choice bubbles to confirm they were "on grid" and readable by CTB scanners.

3.2 Description of the WKCE 2012 Tests

The 2012 test books contained Reading and Mathematics in a single test book at each grade for grades 3, 5, 6, and 7. The single test books for grades 4, 8, and 10 contained Reading, Mathematics, Science, Language Arts, Writing, and Social Studies. The use of a single test book, rather than multiple test books per student, was first implemented in 2009. This was done to improve data quality because the use of two booklets created problems with matching student records.

The Reading and Mathematics tests for grades 3–8 and 10 consist of custom items developed specifically for the WKCE. Language Arts, Science, and Social Studies at grades 4 and 8 consist primarily of *TerraNova* items. A few custom MC items were added to address content standards not adequately covered by the *TerraNova* items. The grade 10 Language Arts, Science, and Social Studies tests consist of custom items previously developed for Wisconsin.

3.2.1 Reading

Table 3-1 presents the configuration of the operational tests. The Reading tests for grades 3–8 had one operational passage for each of the six types of passages: short literary, long literary, short informational, long informational, poetry, and everyday text.

For grades 3–8 and 10, there was one test form given in three test sessions. Each grade had at least one set of paired reading passages with a few items that required analyzing or synthesizing ideas from the passages. Each of the three sessions had approximately 18 MC items. Two of the three operational sessions included a CR item, with the exception of Reading grade 3 which had one CR item in one session. In grades 4-10, one of the CR items was for the reporting category "Analyzing Text," while the other was for the reporting category "Evaluate and Extend Text." In grade 3, the CR item was for the reporting category "Analyzing Text." Each session was allotted 40 minutes of testing time. The grade 10 test consisted of three sessions: Sessions 1 and 2 were 35 minutes and Session 3 was 40 minutes.

Copyright © 2013 by the Wisconsin Department of Public Instruction

3.2.2 Mathematics

Table 3-1 also shows the operational Mathematics test structure. The Mathematics tests for grades 3, 4, and 5 each had three sessions. Grades 6, 7, 8, and 10 had four sessions.

In each grade, the first session was a "non-calculator" session. Grades 3 and 4 do not permit the use of calculators for any session. For these grades, if a student is provided an accommodation that allows the use of a calculator, the calculator may not be used to answer the items in Session 1.

3.2.3 Language Arts

The operational test configurations of Language Arts tests for grades 4, 8, and 10 are presented in Table 3-1 as well. The grades 4 and 8 Language Arts tests consisted of 24 *TerraNova* MC items and 6 custom MC items that measure content standard F, "Research and Inquiry." The session was allotted 30 minutes of testing time. There was a writing session in grades 4 and 8 that presented an operational writing prompt. This session was allotted 30 minutes. The grade 10 test consisted entirely of custom items developed for Wisconsin. The test was administered in two sessions: the first session contained 30 MC items, and the second session contained the writing prompt.

3.2.4 Social Studies

Table 3-1 also presents the operational Social Studies test structure. The Social Studies test at grades 4 and 8 consisted almost entirely of *TerraNova* items, but also included a few custom items previously developed for the WKCE. There was one test session at these grades, which was allotted 40 minutes. The grade 10 test consisted of 50 custom MC items developed for Wisconsin. The test was administered in two sessions; each session was timed at 25 minutes.

3.2.5 Science

Table 3-1 presents the operational Science test structure as well. The Science test at grades 4 and 8 consisted almost entirely of *TerraNova* items, but also included a few custom items previously developed for the WKCE. There was one test session at these grades, which was allotted 40 minutes. The grade 10 test consisted entirely of custom items developed for Wisconsin. The test was administered in two sessions; each session was allotted 25 minutes.

3.3 Customer Approvals

The development phases where DPI approval was obtained included the following:

- pre-content and bias review of new items
- item content and bias review

Copyright © 2013 by the Wisconsin Department of Public Instruction

- item selection for the Fall 2010 WKCE form that was reused in 2012 (with the exception of the Reading Grade 10 which reused the Fall 2008 form)
- first pages in 2012
- final pages (prior to release to Manufacturing)

More specific information describing DPI's role during the development phases is overviewed in the following sections.

3.3.1 Item Content and Bias Review

Following the review of items, CTB and DPI staff reviewed the edits recommended by the educator committees. DPI gave final approval of educator recommendations. DPI and CTB each kept a copy of the item review book with the edits marked.

3.3.2 Item Selection Approval

In 2010, CTB submitted item selection summaries to DPI for the 2010 test forms, which were subsequently re-administered in 2012 (for Reading Grade 10, the 2008 test form was reused). Item selection summaries included test characteristic curves and standard error plots, lists of the items selected and summary test statistics. DPI approval was obtained using a sign-off form.

3.3.3 First Pages Approvals

CTB assessment editors submitted copies of the test book manuscripts to the CTB Production team. The manuscripts show the items as sequenced within test sessions. The manuscripts for the test administration manuals were also submitted to DPI for review, and content changes were addressed at this stage. DPI approval was obtained using a sign-off form.

The Production team returned the test book pages to CTB style editors as first pages. CTB style editors reviewed first pages to ensure pages followed the proper format. CTB assessment editors reviewed first pages for format and content issues. Assessment editors marked first pages to indicate content changes requested by DPI on the manuscript sign-off form. CTB assessment editors submitted a copy of first pages with correction markup to the Production team, and the edits were incorporated in the InDesign files. CTB editors reviewed the corrected pages before submitting them to DPI for review. If an edit was not incorporated correctly, it was re-marked for correction.

3.3.4 Second Pages Approvals

Because of the re-administration of the 2010 (and 2008 for Reading Grade 10) test forms, it was determined that second pages approvals were not needed for this administration.

Copyright © 2013 by the Wisconsin Department of Public Instruction

3.3.5 Final Pages Sign-Off

The final pages represent DPI's last opportunity to review test book and test administration manual pages prior to releasing the materials to CTB's Manufacturing team. At this stage, the materials had been through CTB's quality assurance process and all queries had been resolved. The focus of this review was to verify that previously requested edits had been made and that there were no errors in content or conventions of standard written English. DPI approval was obtained using a sign-off form.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 4: Test Administration

In the fall of 2012, Wisconsin administered assessments in Reading and Mathematics for grades 3–8 and 10 and Language Arts, Social Studies, and Science for grades 4, 8, and 10. The test administration window was October 22–November 23, 2012. Part 4 of the Technical Report describes a set of standardized procedures and policies applied to administer WKCE assessments. The issue of test security in test administration has important implications for the integrity of the results and thus the validity of WKCE scores. Documentation citing the written procedures provided to test administrators and school personnel in order to standardize the administration of the test are also provided in this part. The following AERA, APA, & NCME (1999) standards are addressed in Part 4: 1.13, 3.3, 3.19, 3.20, 3.21, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 6.11, 6.15, 9.1, 10.1, and 10.2.

DPI is committed to the proposition that all schools, and all students within schools, will be held accountable to a common set of high academic content standards. Students who have an Individualized Education Plan (IEP)—a 504 plan (under Section 504 of the Rehabilitation Act of 1973)—or are identified as limited English proficient (LEP) or formerly limited English proficient (FLEP) may be eligible to receive testing accommodations. Accommodations are changes in the routine conditions under which a student takes an assessment in order to provide the student an equal opportunity to demonstrate his or her knowledge. The types of accommodations and guidelines for test administration conditions are described below.

4.1 Accommodations

Accommodations were allowed for eligible individual students participating in the WKCE. 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 WKCE accommodations policy (Appendix 2 and http://oea.dpi.wi.gov/files/oea/pdf/accomswd.pdf). Test administrators indicated which accommodations were used by each student by completing the Student Assessment Report, which is located on the back cover of the student answer document. The following accommodations information was collected from the Student Assessment Report:

Type of Accommodation:

- Used translation
- Signed test questions and content to student
- Used Braille
- Used assistive device (e.g., text-talker, adaptive keyboard, picture symbols)
- Used objects or manipulatives
- Used another DPI-approved accommodation
- Used a non-allowed accommodation resulting in the invalidation of test results

Copyright © 2013 by the Wisconsin Department of Public Instruction

For the Fall 2012 WKCE administration, the State of Wisconsin developed Spanish and Hmong translation scripts for the WKCE. The aim of these scripts is to better help students demonstrate their knowledge on the WKCE without the interference of language. Students whose native language is Spanish or Hmong were given the choice to use all or parts of the translation accommodation, which included a bilingual word list of commonly used content area vocabulary, translation of the test directions, and a written translation script of Mathematics, Science, and Social Studies test items. DPI recommended that educators also consult the list of allowable accommodations in order to create the most appropriate testing situation for their students.

DPI recognizes that approximately 5% of the Wisconsin limited English proficient population speaks a language other than Spanish or Hmong. Districts who serve students who speak languages other than Spanish or Hmong were allowed to use qualified translators to provide oral translation support to students. However, the use of translation support was restricted to Mathematics, Science, and Social Studies tests.

Table 4-1 provides the list of standard accommodations made available for the Fall 2012 WKCE assessments and the number and percent of students provided these accommodations. Table 4-1 also provides a summary view of the accommodations provided, based on all students. The table is split across pages by accommodation, with one accommodation per page. Additional accommodation tables were also delivered to DPI from CTB, which detailed the accommodations provided for subgroup populations of interest, including gender, race/ethnicity, socioeconomic status, disability status, English language proficiency, and migrant status.

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

The primary goal of test security is to protect the integrity of the assessments. To ensure that trends in achievement results can be calculated across years and in order to provide longitudinal data, a certain number of test questions must be repeated from year to year. If any of these questions are made public, the validity of the test may be compromised. Access to test materials was limited to those educators who required access. DPI ensured that all who had access to test materials understood the critical need for test security. They presented security requirements during the 2012 Pre-Test Workshops and outlined the acceptable and unacceptable test preparation and administration practices (Do’s/Don’ts sheet provided in the Test Coordinator Kits). All WKCE tests were administered under secure testing conditions established by DPI.

The following Wisconsin Student Assessment Security Warning Statement was directed by DPI to appear on every test book beginning with the 2004–05 school year:

Copyright © 2013 by the Wisconsin Department of Public Instruction

Test Security

All passages, stimuli, and questions used in the *Wisconsin Knowledge and Concepts Examinations—Criterion-Referenced Test* are CONFIDENTIAL and must be kept SECURE at all times. Unauthorized use, duplication, or reproduction of ANY or ALL portions of the test materials is prohibited. Violation of security can result in district disciplinary action, prosecution, and/or penalties by the Department of Public Instruction or CTB/McGraw-Hill.

Other security measures for WKCE test administrations are described below.

Limited English proficient students and students with disabilities were allowed to use highlighters. Test administrators were instructed to carefully supervise the use of highlighters because they may cause smudging of pencil marks and bubbles, which could affect reliability of scanning and scoring. If highlighters were used, the following guidelines were provided:

Guidelines for Highlighters:

1. Do not allow the highlighting of track marks, litho codes, skunk lines, barcodes, preslugged bubbles, or any carbon black printing. The highlighters cause these black inks to blur and bleed.

2. Do not allow the highlighting of pencil marks of any kind, whether bubbles or handwriting. The highlighters cause pencil marks to blur and bleed.

3. Use only highlighters from the following list, which were tested and found to have minimal problems:

   - Avery Hi-liter
   - Avery Hi-liter, thin-tipped
   - Bic Brite-Liner
   - Sanford Major Accent
   - Sanford Pocket Accent, thin-tipped

Test Security during Breaks:

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Parameters for marking test books with a No. 2 pencil:

- Do not mark in the bubble answer positions.
- Do not mark in the student Pre-ID Barcode on the barcode label.
- Do not mark in the timing tracks (the parallel lines along the side of the test book).
- Do not mark in the skunk lines (the little squares and rectangles across the bottom of each page of the test book).
- Do not mark in the litho codes (the squares and numbers across the bottom of the document on the first and last pages of the test book).
- Do not mark more than one answer bubble as the scanner cannot determine a response.

4.4 Test Administration

In order to ensure standardized testing administration for all students, a Guide for District Assessment Coordinators and School Assessment Coordinators was made available to all assessment coordinators (DPI, 2012–2013). The guide included the following topics:

- Test Security
- Test Materials and Procedures
- Packaging the Test Materials
- Procedures for Returning Materials
- Test Results
- Responsibilities of District Assessment Coordinators (DACs)
- Responsibilities of School Assessment Coordinators (SACs)
- Checklist for School Assessment Coordinators
- WSAS Policy and Procedure Manual

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

- Test Materials
- Test Security
- Testing Schedules
- Organizing the Classroom
- Preparing Students to Take the Test
- Use of Appropriate Test Procedures
- Filling in the Student Information Page
- Administering the WKCE
- Filling in the Student Assessment Report
- Assembling Materials for Return

Copyright © 2013 by the Wisconsin Department of Public Instruction

For specific information related to test administration, refer to the Test Coordinator's Manual and/or the Test Administration Manuals that are available online at: http://oea.dpi.wi.gov/oea_publications.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 5: Scoring

The purpose of Part 5 is to demonstrate adherence to AERA, APA, & NCME (1999) standards for scoring, including 3.22, 3.23, 3.24, 5.8, and 5.9. Part 5 describes the following:

- The scoring process of MC items
  - The scanning process
  - The calibration of scanners and other quality-control measures

- The scoring of CR items
  - The scoring rubrics
  - The handscoring process
  - The electronic handscoring system
  - The selection of Scoring personnel
  - The selection of anchor papers
  - The distribution of CR item scores

5.1 Scoring of Multiple-Choice Items

At the conclusion of the Fall 2012 WKCE administration window, student test documents were returned to CTB's scoring facility by the districts. Test materials were tracked through the entire scoring process, from the initial retrieval of the student test documents, through all scoring processes, and on to the final document retention period.

CTB's Scoring Operations processes were organized into Lean Processing Scanning Cells. Each cell was a self-contained, cross-functional team made up of the stations, equipment, and personnel skill-sets necessary to efficiently and accurately complete the operational processing cycle for student test documents.

Student answer documents were handled in a series of distinct processes. In order, those processes were as follows:

Receiving—Answer documents were tracked from retrieval to receipt at CTB, checked for damage in shipping, verified for full box counts, registered into an internal tracking system called the On-Hold Tracking System (OHTS), and then passed along to Login.

Login—Answer documents were then removed from the boxes, the pre-work was verified for district accuracy, and stacks of answer documents were aligned and cut for scanning.

Scanning—Stacks of answer documents were fed through optical scanners (see the following section for details) and any scanning problems were monitored and rectified (also detailed below).

Copyright © 2013 by the Wisconsin Department of Public Instruction

Updates—The raw scoring and editing of scanned student data were performed using a system of edits to verify the integrity of each batch of scanned answer documents. The raw scoring and editing of the scanned student data also yielded an error list. Errors were resolved by trained editors using pre-defined guidelines in the Winscore editing system.

Documents were moved directly from process to process or sat momentarily in mini-queues. Once this stepwise process was complete, the student test documents were prepared for secure document retention.

Document Retention—Student test documents were then moved to a staging area where they were caged, warehoused, and ultimately retained for retrieval during the specified retention period. At the end of the 365-day retention period established in the WKCE contract, and upon customer approval, these documents will be loaded into containers provided by a designated NAID-certified[2] secure destruction company following strict national guidelines. The documents will then be picked up and shredded within 24 hours. Until shredded, the documents are caged and locked in a secure environment.

5.1.1 Scanning and Calibration of Scanners

This section provides a description of the scanning process and quality control processes applied in the scoring process.

Optical scanners captured all MC, ancillary, and student demographic data. An optical scanning technology called Optical Mark Recognition (OMR) detected all pencil marks in the answer section of the scanned document. The student test data was processed through CTB's proprietary Winscore editing system. The Winscore scanning program evaluated detectable marks on both sides of each page, recording the intensity and coordinates of solid marks for resolution in the raw scoring step. The scanner reported intensities in the range 0 (lightest) to 15 (darkest). Winscore scored the darkest mark for each question as the intended response. In this way, completed bubbles were turned into characters of data representing test-item responses or other information.

The scanning production systems separated the MC item data from the CR item data. The CR data was handled in a "handscoring" process, as described in Section 5.2. The MC data and the handscoring data were later merged for correction, analysis, and reporting.

CTB's scanning software captured student response data in images called TIFFs. The scanning process also captured data in barcodes and in identification marks (i.e., "skunk marks"), which were used to determine the type of document. Document headers provided customer identification and district, school, and class information. All images were captured during scanning using high-resolution technology, also called "grayscale." Any item determined to be "unclear" was electronically retrieved in grayscale in the Electronic Handscoring System (EHS).

[2] NAID is the National Association for Information Destruction.

Copyright © 2013 by the Wisconsin Department of Public Instruction

The optical scanners were able to run at a rated speed without any interruptions except for problems with the physical documents. At the beginning of each shift, and after scanning every 5,000 sheets, a diagnostic sheet was used to assess the camera functionality. CTB cell leads also cleaned the scanners at the end of their shifts and ran a "quick check utility" to confirm that the equipment was ready for the next shift. If the scanner did not pass the quick check, or a diagnostic check, a field engineer was then called in to address the problem. If the scanning camera was adjusted in any way, the scanner was recalibrated and the quick check utility was run again. When readied, the scanner was then released for scoring. All scanners were calibrated as scheduled.

No recalibration was necessary during the WKCE Fall 2012 administration. The following processing metrics were obtained:

- Number of sheets scanned: 22,740,272
- Number of books scanned/processed: 453,894

The following checks were used to ensure the integrity of the student response data:

Reliability check—When there were low scores, either among groups or at the individual student level, the reasonability of the low-score ranges was verified.

Biographical data—During the Winscore process, a series of checks were completed on critical Wisconsin fields, such as student name, gender, and date of birth. The system flagged missing, double marked, or invalidly marked data. When a record was flagged for any critical Wisconsin field errors, the document was pulled and the bubbled data was verified and corrected accordingly.

Duplicate barcode and litho code checks—Additional checks were completed in Winscore to ensure that each document was scanned only once. A duplicate checker in Winscore flagged duplicate barcodes and litho codes. If either was flagged, the book was pulled and the barcode or litho code was verified to ensure that it had been accurately scanned, that no document was scanned twice, and that no barcode labels had been incorrectly applied. In addition to checks carried out in Winscore, further checks were carried out in Monarch, a back-end data system that flagged duplications and matched district and school data.

Student counts—The actual book counts generated by the scanners were compared to the book counts provided by the school districts on the School Group List and School Header Sheet. In 2012, 124 discrepancies were identified and resolved by emails and telephone calls placed to the districts. These completeness checks occurred from November 20, 2012 to January 14, 2013.

Copyright © 2013 by the Wisconsin Department of Public Instruction

School name/number—Pre-assigned school numbers and names were verified against data provided by DPI.

The scored student response data were later retrieved by the CTB Research and Technology teams for statistical analyses and for producing reports.

5.2 Scoring of Open-Ended or Constructed-Response Items

Sections 5.2.1 through 5.3 document the scoring processes used for CR items. This documentation forms part of the validity evidence supporting the scoring process used for CR items. Sections 5.2.1 through 5.3 describe the scoring rubrics, the scoring process, the selection of sample (anchor) papers used to train scoring personnel, the process of selecting personnel, inter-rater reliability, and the distributions of scores from CR items.

5.2.1 Description of Scoring Rubrics

In the 2012 administration, the Reading and Mathematics forms in grades 3–8 and 10 contained CR items. A Writing prompt was also administered at grades 4, 8, and 10. The Writing prompts were scored using two holistic rubrics: a 3-point Conventions Rubric and a 6-point Composing Rubric. Tables 5-1 through 5-8 present the scoring rubrics.

5.2.2 Handscoring Process

The Scoring personnel who score CR items are referred to as “readers.” As indicated previously, the process of scoring CR items is referred to as “handscoring.” The handscoring readers were trained using customer-approved training materials, such as the anchor papers described in Section 5.2.4. Once qualified, readers were required to maintain accuracy standards throughout the project. These requirements were assessed at the item level primarily through each reader’s daily “checkset” performance (described below), as well as agreement rates with other readers on the second reads (described below), and targeted read-behinds with team leaders (described below). Data monitors generated reports daily that flagged any readers falling below the established quality standards for any item, providing insight on reader scoring trends (such as difficulty with any particular score point). These reports were shared with handscoring supervisors. Those readers identified in the reports received additional coaching, training, reviews, targeted read-behinds, or additional checksets. Readers who did not meet standards with these initial corrective actions were administered another validation (recalibration) round. Failure to recalibrate resulted in dismissal from the scoring assignment. This process was in place throughout the entire handscoring window.

Copyright © 2013 by the Wisconsin Department of Public Instruction

5.2.3 Electronic Handscoring System

The Electronic Handscoring System (EHS) was used to score CR items. EHS presented images of scanned test books to trained readers who assigned scores for the CR items. The scanned student responses were viewed on high-quality, 19-inch workstation monitors. Images of each student's responses were automatically routed to two or more readers when required, and images of specific subsets of test items were routed to designated groups of readers trained to score these items.

5.2.4 Anchor Papers and Training Papers

In 2012, all training materials, including scoring guides and Reading and Mathematics rubrics, anchor papers, training papers, qualification round papers, and checksets, were from the 2007 (2008) operational administration. Prior to the actual scoring in 2007 (2008), the CTB Scoring Center created training materials. A selected group of papers written by WKCE students were selected as models to train raters for scoring. These papers, referred to as "anchor papers," played an important role in deciding which level of writing should receive which score. Range-finding meetings were held with DPI staff and educators to select sample papers for each score point. CTB randomly sampled student answer documents to ensure a representative sample of the possible responses. The sample papers were used to construct scoring guides and training papers. CTB's scoring team collaborated with DPI to make necessary revisions to the rubrics and in the selection of scoring guides and training papers. This process included several pre-sorting steps and subsequent iterative/consensus processes in order to achieve agreement and precision through a "round robin" scoring process. Once approved by DPI, the scoring guides (consisting of rubrics, anchors, and annotations) served as a constant guide, setting the course for all subsequent training and scoring.

5.2.5 Scoring Personnel and Qualifications

CTB recruited, trained, and managed personnel to complete all of the handscoring operations within the timelines of the contract. This involved extensive consultation between CTB's Scoring and Publishing Departments, Wisconsin educators, and DPI in order to review scoring rubrics, develop the anchor papers and other reader training materials, and to provide analyses of student responses to tryout forms. The characteristics of the readers, team leaders, and scoring supervisors are described in the following sections.

Readers—Many CTB readers had years of classroom teaching experience. The CTB reader pool included many retired and current educators, as well as engineers, editors, published authors, and individuals with advanced degrees. The minimum qualification for all readers was a Bachelor's degree. Readers were required to participate in training and successfully pass at least one of two qualification rounds. Once qualified, readers could start scoring, but throughout the scoring process, reader performance was assessed by a supervisor and data-monitoring staff through the use of checksets, read-behinds, and the review of inter-rater reliability statistics, as described in Sections 5.2.7, 5.3, and Part 9.

Copyright © 2013 by the Wisconsin Department of Public Instruction

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 readers. Team leaders were each responsible for a small team of readers. In addition to performing read-behinds on readers, team leaders also coached readers when needs were identified through data monitoring or otherwise by supervisory staff. Team leaders working on the writing component also resolved discrepant scores.

Scoring Supervisors—Scoring supervisors were the core group at CTB who directed and organized the assessment process, and trained team leaders and readers. Scoring supervisors had extensive experience as team leaders prior to their qualification and selection. Scoring supervisors were content area experts in the content areas they supervised and trained. They oversaw all team leaders and readers.

5.2.6 Reader Training

Validation was a critical task in the training process and the final determinant of reader readiness. All readers, 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 dependent on the score point range of an item. For example, where scores were either zero or one point, the level of agreement required was 95%, but where scores could range from zero to two points, the level of agreement required was 90%. Those readers not validating on the first attempt received further training prior to taking an additional qualifying round. Only those who were successfully validated were qualified as readers to score tests. Team leaders were required to complete two validation rounds with at least 80% exact agreement in each round.

5.2.7 Inter-Rater Reliability

Checksets—Throughout the course of the handscoring process, sets of pre-scored papers called “checksets” were administered daily to the team leaders as well as to the readers. The checksets were used to monitor scoring accuracy and to maintain a consistent focus on the established rubric and guidelines. This kind of monitoring occurred without reader knowledge. Readers, whose checkset scores fell below the qualifying level, were flagged for additional coaching (training review, targeted read-behinds, etc.). Those readers who remained below standard were given another validation (recalibration) round. Readers unable to recalibrate were dismissed.

Read Behinds—The “read-behind” was another valuable monitoring technique used. Each team leader was able to read a random selection of a reader’s scored items. This reading could be targeted at the item and score point level. The scores were compared, and if they agreed, the team leader was able to offer feedback, which enhanced the reader’s confidence and ability to score quickly and accurately. However, if a reader strayed from the standards established in the training and validation samples, the aberrant scoring was detected, and the

Copyright © 2013 by the Wisconsin Department of Public Instruction

team leader was able to offer guidance necessary to refocus the reader's effort. Readers, whose scoring was inconsistent, were read behind more frequently by their team leaders, thus correcting any scoring variations.

Final Score—In Writing, Reading, and Mathematics, the first score assigned for each CR item was the final score; however, 5% of the responses per item were double read (in "second reads") to obtain indices regarding the consistency and accuracy of raters. Inter-rater reliability was monitored throughout the scoring process, as described in Part 9.

5.3 Distribution of Constructed-Response Item Scores

Tables 5-9 through 5-16 show distributions of CR item scores across each score point level (one point, two points, etc.) for each CR item and the Writing prompts. The scoring distributions shown for Reading and Mathematics are the scoring distributions of the first read. As described previously, 5% of the responses to the CR items in Reading, Mathematics, and Writing were double read (in "second reads") for statistical purposes. These distributions were examined for quality assurance purposes in the scoring process.

These tables use four condition codes. Condition code "A" denotes items with no response or no attempt, code "B" represents an illegible response, code "C" indicates that another language was used in the response, and code "D" denotes a response that was off topic.[3]

All Reading items had one part and a maximum score of three points. In Mathematics, many CR items in grades 3–8 had two parts: a Part A worth one point and a Part B worth two points. The CR items in grades 3–8 with only one part were worth two points. In grade 10, all Mathematics CR items had one part and were worth two points.

As can be seen in Table 5-9 for Reading, in most cases, most students scored one or two points, and fewer students scored either three points or zero points. Scoring three points was not common in Reading; however, this result may be expected because CR items are often more difficult than MC items.

In Mathematics, although many students scored at the maximum score level for the CR items, many students also obtained a score of zero. This occurred on both Part A and Part B of the two-part CR items.

Students in grades 4, 8, and 10 were administered one Writing prompt. Tables 5-11 through 5-16 present the score distributions for the student responses to the Writing prompts. These tables are split between counts and percentages, and separate tables are provided for the 6-point Composing Rubric and the 3-point Conventions Rubric. The first score assigned for each Writing response on each rubric was the final score; however, 5% of the responses per prompt were double read ("second reads") to obtain indices regarding the consistency and accuracy of

[3] When calculating students' scores on operational items, CR items receiving these condition codes were given zero score points.

Copyright © 2013 by the Wisconsin Department of Public Instruction

raters. Scores from the first read, the second read, and the difference between the two reads are presented in Tables 5-11 through 5-16. As can be observed in Tables 5-11 through 5-16, the rater scores were very similar. As described previously, inter-rater reliability was also monitored in other ways throughout the scoring process. The full results for inter-rater reliability are presented in Part 9.

As can be seen in Tables 5-11 and 5-12, most scores in the Composing Rubric were in the middle of the 6-point range, and relatively few students were at the low and the high extremes. The Conventions Rubric showed similar results. As can be seen in Tables 5-13 and 5-14, a large proportion of students scored in the middle level of the 3-point range for the Conventions Rubric, and relatively few students scored either 1 point or 3 points.

Tables 5-15 and 5-16 show the total score on the Writing prompt, combining scores from the Composing Rubric and the Conventions Rubric. The combined scores for most students were in the middle or upper-middle range of the 9-point total, from 4 points to 6 points. The highest and lowest levels of scoring were less common, but in every grade, a small proportion of students obtained zero score points, and a small proportion obtained the highest possible score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 6: Characteristics of the Calibration Sample

The calibration and equating of the Fall 2012 WKCE occurred in 2010 (or in 2008 for Reading grade 10) based on student data from a preselected sample of districts in the state. This arrangement was chosen in order to expedite the return of score reports to districts. In accordance with AERA, APA, & NCME (1999) standards 1.5, 1.13, 2.4, 4.7, and 6.1, this section provides a description of how the 2010 (or 2008 for Reading grade 10) calibration sample was selected and how the calibration sample and census data compare in terms of demographic characteristics. Part 6 serves to demonstrate that the 2010 (or 2008 for Reading grade 10) calibration sample was sufficiently representative of the Wisconsin student population for the purposes of calibration. This documentation also serves as validity evidence supporting the WKCE program. Information about the calibration sample can be found in the 2010 (or 2008 for Reading grade 10) WKCE Technical Report available from the DPI at: http://oea.dpi.wi.gov/oea_publications.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 7: Calibration, Equating, and Deriving Scale Scores

Student responses on the WKCE are input into complex mathematic 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). WKCE scores are established through the processes of calibration, equating, and item-pattern scoring. Part 7 of the Technical Report describes these processes as they were applied to the Fall 2012 WKCE administration, as well as the results. This portion of the Technical Report addresses AERA, APA, & NCME (1999) standards 1.13, 4.1, 4.2, 4.3, 4.10, 4.11, 7.1, 7.2, and 7.10.

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 7.3. Additional references are also provided.

Calibration is the mathematical process of estimating characteristics of individual items. These characteristics are termed "item parameters." Sections 7.1, 7.1.1, and 7.1.2 serve to explain this process beginning with a description of the calibration and equating methods used in 2010 (or 2008 for Reading grade 10) that were applied to the Fall 2012 WKCE, followed by a discussion of the calibration models and the software used. The derivation of scale scores from raw scores is then addressed, with a focus on nontechnical audiences. The results of the calibration process, using model-to-data fit statistics and the standard error of measurement (SEM), are also discussed.

7.1 Calibration and Equating Methods

In the Fall 2010 (or 2008 for Reading grade 10) WKCE, the three-parameter logistic (3PL) IRT model (Lord & Novick, 1968; Lord, 1980) was used for MC items, and the two-parameter partial credit (2PPC) model (Muraki, 1992; Yen, 1993) was used for CR items. Language Arts, Social Studies, and Science items were calibrated using the 3PL model because these three content area tests are comprised of only MC items. Because the Reading and Mathematics tests consisted of both MC and CR items, a simultaneous calibration with the 3PL and 2PPC models was implemented. A simultaneous calibration was also applied to the Language Arts test in grade 10 because a Writing prompt was included as a component of a student's scale score at this grade level. The 3PL and 2PPC models are described in detail in the next section.

Simultaneous calibration is used for the mixed format tests in part because a single scale communicates that the measured skills relate to the same underlying qualities and characteristics, and that they can be taught and measured using a variety of assessment modes. In considering the simultaneous calibration process, Thissen, Wainer, and Wang (1992) stated that items of diverse types can be scaled together provided that the different types of items assess the same primary characteristics of the content area.

Copyright © 2013 by the Wisconsin Department of Public Instruction

By design, there was a special set of items in each content area and grade level test that was common to both the current (2010) administration and a prior administration. This arrangement is called a *common item non-equivalent group* design. The purpose of this design is to place current operational items on a base scale using the common items. Horizontally equating the current test forms to the previously established scales is necessary in order to obtain results that are comparable across administration years. The equating process also mitigates differences in test difficulty between forms from the current and the previous year, which are built to be similar in difficulty and content (Kolen & Brennan, 1995). The items that were used for equating are called anchor items. In each grade and content area, each set of anchor items was a miniature version of the total test, which adequately represented the test content coverage in terms of item difficulty and the test specifications. The Stocking and Lord (1983) procedure was used to equate the estimated parameters to the scale from which the anchor items were drawn. This procedure estimates the linear transformation constants by minimizing the distance between the test characteristic curves for the calibrated anchor items and the values for the anchor items already on the test scale.

The Reading and Mathematics vertical scales were established in Fall 2005 using a similar plan termed an *adjacent grade common item design*. Based on Fall 2004 data, scores for adjacent grades were linked so that student scores in grades 3–8 and 10 could be expressed on a single scale. Vertical scales were not developed for Language Arts, Social Studies, and Science because these tests were administered only in grades 4, 8, and 10. Instead, the scales for grades 4, 8, and 10 were constructed in such a way to show a vertical relationship (i.e., an increase in scale score means) across grades. For additional information on the scaling methods used to establish the WKCE scales, readers can refer to Part 8 and Part 11 of the WKCE Technical Report from the Fall 2005 WKCE administration, which can be found in Appendix 3 of the 2010 Technical Report, available at: http://oea.dpi.wi.gov/oea_publications. The 2005 Technical Report includes a fairly extensive discussion of the scaling methods.

7.1.1 Calibration Models

The 3PL model defines a MC item in terms of three characteristics, or *item parameters*: (a) item difficulty (or its location on a scale of difficulty/ability), (b) item discrimination (or how well the item differentiates between the low- and high-ability students in relation to its location), and (c) the level of guessing. The 2PPC model defines a CR item in terms of item discrimination and item difficulty for each score point.

In the 3PL model, the probability that a student with scale score θ responds correctly to item *i* is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-1.7 a_i (\theta - b_i)}},$$

where a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response by a very low-scoring student.

Copyright © 2013 by the Wisconsin Department of Public Instruction

The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with ability θ having a score at the *k*th level of the *j*th item is

$$P_{jk}(\theta) = P(x_j = k-1|\theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \; k = 1,...m_j,$$

where $Z_{jk} = A_{jk}\theta + C_{jk}$.

For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1) \text{ and } C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji}, \text{ where } \gamma_{j0} = 0,$$

where α_j and γ_{ji} are parameters freely estimated from the data. The first constraint implies that higher item scores reflect higher-ability levels and items can vary in their discriminations. The 2PPC model estimates a total of m_j independent item parameters; for each item, there are $m_j - 1$ independent γ_{ji} parameters and one α_j parameter.

The item calibration process is a process of estimating item parameters. Parameters are estimated in an iterative process using a computer software program called PARDUX (discussed below). The PARDUX program operates by estimating person parameters (ability) and item parameters (e.g., difficulty) through a series of iterations until the change in parameter estimates between iterations is reduced to a given threshold.

7.1.2 Calibration Software

The IRT models and the student response data from the Fall 2010 (or 2008 for Reading grade 10) WKCE administration were used to estimate item parameters for each test. The IRT models were implemented using CTB's PARDUX software (Burket, 1991). Using marginal maximum likelihood procedures implemented with the expected maximum algorithm, PARDUX estimates parameters simultaneously for MC and CR items (Bock & Aitkin, 1981; Thissen, 1982).

PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. Extensive simulation studies and comparisons between PARDUX and MULTILOG (Thissen, 1990)—a program widely used for research purposes—have shown that PARDUX provides precise parameter and ability estimates and it performs more efficiently than MULTILOG (Fitzpatrick, 1991). Simulation studies have also compared PARDUX with PARSCALE (Muraki & Bock, 1991) and with BIGSTEPS (Wright & Linacre, 1992). Fitzpatrick and Julian (1996) found that PARDUX provided precise parameter and ability estimates and performed more efficiently than the other programs. Extensive research with simulation data has

Copyright © 2013 by the Wisconsin Department of Public Instruction

also shown that the IRT procedures used here produce accurate vertical scaling (Yen & Burket, 1997).

7.2 Calibration Results

The following sections describe the calibration results in terms of the estimation of item parameters, model-to-data fit, and the SEM of the scale scores across content areas and grades.

7.2.1 IRT Item Parameters

At times when calibrating items, items do not converge, meaning the characteristics of the item are not able to be determined. When this occurs, items are suppressed from student scoring and future assessments. In 2010 (or 2008 for Reading grade 10), no convergence issues occurred for any item on the operational tests.

7.2.2 IRT Item Fit

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

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

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

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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Because the value of Z increases as the sample size increases, with other things being equal, the critical values for Z were established using the following equation (Yen, 1991a)

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

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

Table 7-1 presents items that were flagged for less than optimal fit when the obtained Z-statistic exceeded the critical Z-statistic value. To take an example from the table, in Reading grade 6, item 44 was flagged because the observed Z of 17.12 is larger than the critical Z value of 15.76 based on a sample size of 5,911.

Table 7-1 specifies the item status, content area, grade level, test book form, item number, item type (MC or CR), N size (the number of students), Z, and critical Z, as described previously. For many of the flagged items, the observed Z and the critical Z are not very far apart. For example, in the case of the first item in the table, Language Arts grade 8 item 2 was flagged because the observed Z of 14.97 is larger than the critical Z value of 13.89. The misfit in this case may be considered small. Although many items in the table show a moderate degree of difference between the obtained Z and the critical Z statistic, others, such as the Mathematics grade 6 item 39B, show much larger differences.

In order to evaluate item-to-model fit further, CTB 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, the 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.

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

In a large-scale assessment such as the WKCE, with 23 grades and content areas, it is expected that some items will be flagged for misfit. The number of items flagged for misfit in the Fall 2010 (or 2008 for Reading grade 10) WKCE is consistent with the number flagged in the year prior. 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 CTB Development

Copyright © 2013 by the Wisconsin Department of Public Instruction

team and DPI. As noted in Section 3.1.1, such items are avoided in future selections unless there is a compelling reason that they should be included, such as meeting the test blueprint.

7.2.3 Evaluating Anchor Items

To evaluate whether anchor items are performing differently in the current administration versus a previous administration, differences between the ICCs were computed for each anchor item. Differences between the curves were evaluated using the following statistics:

- Average Signed Difference
- Average Absolute (Unsigned) Difference
- Root Mean Squared Difference

Both unweighted and weighted versions of these statistics were calculated. Unweighted differences gave equal weight to differences across the ability spectrum. Weighted differences assigned weights according to the number of test-takers that are impacted by differences in the curves. For both weighted and unweighted versions of the three statistics listed above, differences greater than + 0.10 were considered large, and differences between + 0.07 and 0.10 were considered moderate. In addition, the Maximum Absolute Difference was identified for each item. Large Maximum Absolute Differences were those greater than + 0.15, and moderate differences were all differences between + 0.125 and 0.15.

Although dropping an anchor item flagged based solely on statistical criteria has its simplicity, this option may change the content coverage and equating constants, shift scale score distributions, and affect the performance level classification of students by moving them into different proficiency levels. Before an anchor item may be dropped from an anchor set, the adequacy of the content coverage must be evaluated and a reason for the anchor item differential performance must be identified. As stated above, an item is removed from the anchor set only if it adversely affects quality of scaling, not desirability of results. As such, CTB does not consider how the removal of an item affects the overall mean scale score or the impact data (percent of students in each performance level) when recommending items for removal.

Items removed from the anchor set are still scored as part of the whole test. Anchor items were considered for exclusion from the WKCE under the following conditions:

1. An item is flagged for large differences on the Average Signed Difference, Average Absolute (Unsigned) Difference, or Root Mean Squared Difference and for moderate or large differences on the Maximum Absolute Difference when examining the differences between the previous versus current ICCs.
2. 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 © 2013 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 must remain within 10% of the test blueprint.

7.3 Deriving Scale Scores in the WKCE

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 scale scores account for the characteristics of the items that are in the test (such as item difficulty).

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

Whether looking at an individual item or at a group of items that make up a complete test, IRT uses probability models to describe the relationship between a student's ability and his/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., Language Arts ability) and u_i represents an observed score on a particular item. For MC items, the observed score u_i is either 0 or 1, indicating either an incorrect or correct response, respectively. For a MC item, the probability model can be denoted as $P(u_i = 1|\theta)$. That is, P is an estimation of the probability that a student with an ability value θ would answer item i correctly.

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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

.

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

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

Consider the following example. Suppose six examinees in the fourth grade take a MC test in Language Arts with 30 items. Suppose further that the properties, or parameters, of the items on that test are as follows:

Table 7-A. Item Parameters for a Test

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Now suppose the student response patterns for these six examinees are as follows, where 0 represents an incorrect response, and 1 represents a correct response:

Table 7-B. Item Response Pattern

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

The first student, Pam, answered seven of the items correctly and obtained a scale score of 140, which is equal to the lowest point on the score range, called the "lowest obtainable scale score," or LOSS. The next four students each answered 15 out of 30 items correctly, but the response pattern of each of these students is different. The *raw score* of each of these students is 15. However, the maximum likelihood item-pattern scoring method produced a different scale score for each examinee. Scale scores were 246 for Craig, 266 for Vicki, 259 for Tom, and 265 for Evan. These scores can be accounted for by considering the pattern of the student responses on the test together with the properties (or parameters) of the items, as shown in Table 7-A. By referring to Table 7-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.

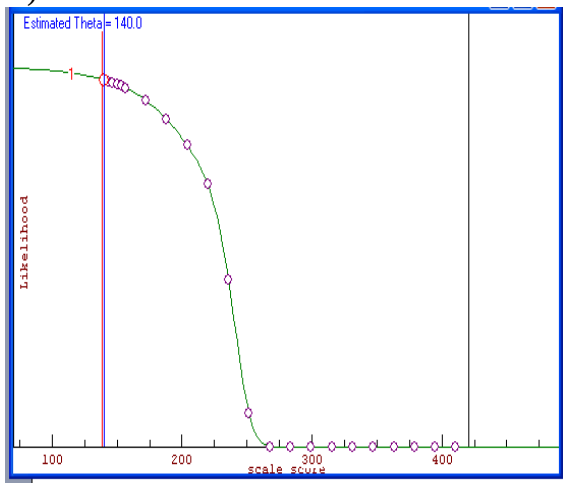
Figure 7-A below shows the probability of each ability estimate (or scale score) for the six examinees. The total scale score range for Language Arts is plotted on the horizontal axis. As indicated by the two vertical lines in the plot, the lower and upper limits of the scale score range are 140 and 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 the ability estimate actually reflects the examinee's ability level.

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

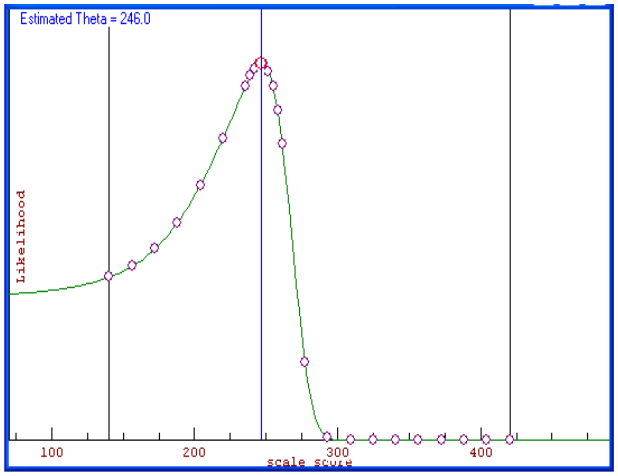
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-A. Likelihood Functions, or the Probability of Each Ability Level Estimate (or Scale Score)*

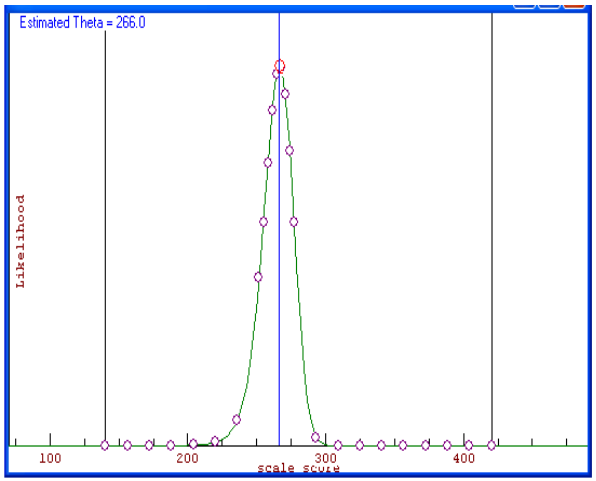
a) Pam



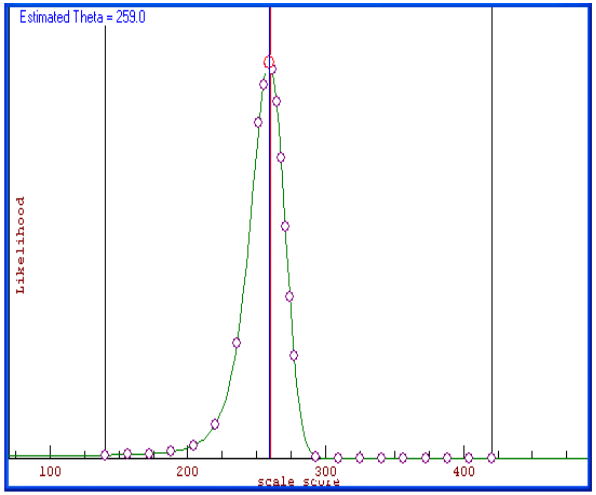
b) Craig



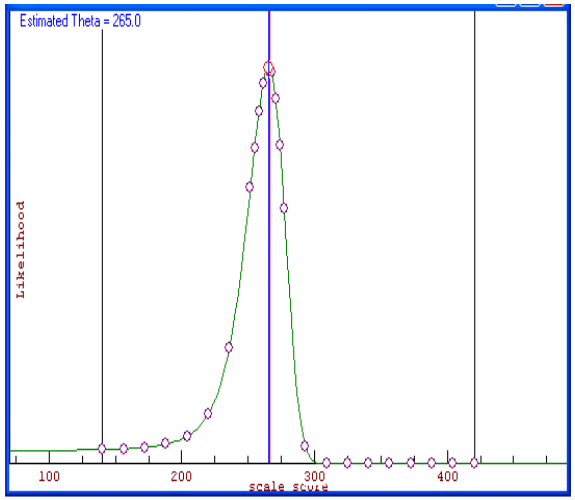
c) Vicki



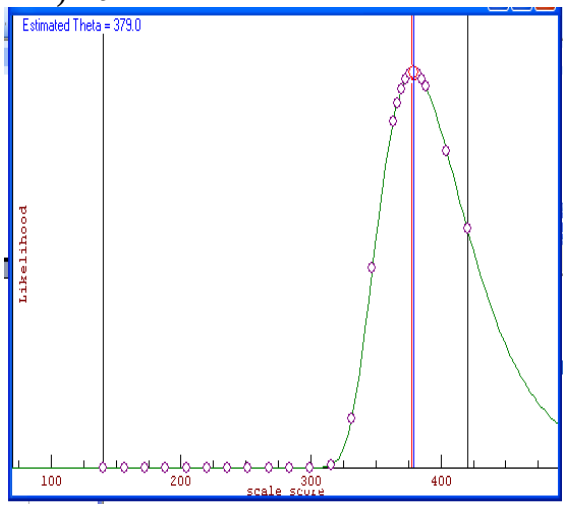
d) Tom



e) Evan



f) Dan



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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

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

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

The procedures applied here are similar to those followed in the development of the *TerraNova* test (CTB/McGraw-Hill, 1997), *TerraNova* 2nd Edition (CTB/McGraw-Hill, 2000), and the prior Wisconsin Knowledge and Concepts Examinations developed in conjunction with CTB (1997–2004). Several supplements to this simplified outline of IRT are available. Introductory discussions of IRT can be found in *Educational Measurement* (Linn, 1989) or Chapter 11 in *Introduction to Measurement Theory* (Allen & Yen, 1979). More advanced discussions of partial credit models may be found in Muraki (1990, 1992), Yen (1993), and van der Linden and Hambleton (1997). For additional information on the technical details of the item-pattern scoring, readers can also refer to Yen & Candell (1991) and to *TerraNova* 2nd Edition (CTB/McGraw-Hill, 2000).

7.3.1 Standard Error of Measurement

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

The SEM of the scale scores in the Fall 2012 WKCE, based on the Fall 2010 WKCE (based on the Fall 2008 WKCE for Reading Grade 10), is displayed graphically for each grade and content area in Figures 7-1 through 7-5. The SEM provided is based on item-pattern scoring. Each SEM curve is plotted as a function of the scale scores. These figures show the scale score range within which measurement is most accurate. The figures also show that extreme scale

Copyright © 2013 by the Wisconsin Department of Public Instruction

scores have more measurement error than scores in the middle of the distribution. Scale scores in the high or low extremes of the student distribution are less precise than those in the middle of the distribution because there 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 lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS), are the starting scale score and the last scale score in these figures. LOSS and HOSS are further discussed in the next section.

Because of the nature of item-pattern scoring, a scoring table showing a simple, direct conversion of raw score to scale score cannot be generated for the Fall 2012 WKCE. However, scoring tables showing a rough relationship between raw score, scale score, and SEM can be produced, and they are provided in Tables 7-2 through 7-24.

7.3.2 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, which are set separately by grade, are called the LOSS and the HOSS.

Table 7-25 shows the number and percent of students at the LOSS and the HOSS. In general, there should not be many students clustered at the LOSS or HOSS. An accumulation of a high proportion of students in the LOSS or HOSS may indicate a floor or ceiling effect.

In most grades and content areas, the percentage of students at the LOSS and HOSS was small: less than one percent. However, in some grades and content areas the percentages were larger. In Reading, three grades had more than one percent of students at the LOSS (Grade 3 – 1.27%, grade 5 – 1.52%, grade 10 – 1.06%) whereas no other subject areas had any grades that had more than one percent of students at the LOSS. These percentages at the LOSS can be considered to fall within an acceptable range, although they can still be considered as a point of reference when developing future forms. The percentage at the LOSS in these grades may be reduced in future years by including some additional items that are not difficult. The percentage of students scoring at the HOSS is similar: In most grades and content areas, the percentage was small, although in a few grades and content areas, the percentage was larger. In particular, more than one percent of students obtained the HOSS in Science grade 8 (1.43%), Social Studies grade 4 (2.09%), and Language Arts grade 8 (3.07). The percentage scoring at the HOSS may be reduced by including some additional difficult items in these grades and content areas, or by including more items on the test.

Copyright © 2013 by the Wisconsin Department of Public Instruction

7.3.3 Test Characteristic Curves

Test characteristic curves (TCCs) are provided in Figures 7-6 through 7-10. These curves model the relationship between student ability and expected scoring outcomes at the test level. By following the plotted line for any grade level and content area, one can observe the estimated scoring outcome (the estimated proportion of the maximum correct score) plotted as a function of examinee ability. These curves are based on the IRT models, methods, and scaling processes described above. The vertical relationship across grade levels that can be observed in the TCCs reflects the typical growth pattern: as grade level increases, ability level is also expected to increase across the ability range.

Although the TCCs, overall, show the expected separation across grades, the separation is somewhat less for Reading than for the other content areas. In addition, the Reading curves overlap in grades 4 and 5 and in grades 7 and 8. Although scale overlap is generally not considered the optimal pattern for a vertically scaled assessment, on Reading assessments this is not uncommon. On the WKCE Reading scales, the cut scores for these grades are closer together than the cut scores across grades for the other content areas. For example, the *Proficient* cut score in Reading is only 4 scale score points higher for grade 5 than grade 4, whereas the *Proficient* cut score for Mathematics is 25 points higher for grade 5 than grade 4. Because the item difficulties in the WKCE tests were chosen, in part, to minimize the standard error around the critical *Proficient* cut score, the proximity of the cut scores in grades 4 and 5 would be expected to yield curves with relatively little separation. The proximity of the curves for grades 7 and 8, however, is less easily explained. Given the greater separation between the scales at these two grades, the observed overlap of the TCCs may indicate that the grade 8 assessment would benefit from the addition of some more difficult items. This consideration, however, must be balanced by the need to keep form difficulty comparable each year to meet the assumptions for alternate parallel forms.

Copyright © 2013 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 Fall 2012 WKCE administration. The summary results 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 called "standardized performance indicator" (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 WKCE testing program. The AERA, APA, & NCME (1999) standards addressed in Part 8 include: 1.5, 3.18, 4.3, 4.5, 4.6, 4.7, 4.19, 7.1, 7.10, 13.15, and 13.19.

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

The *p*-value for a CR item represents the mean proportion of possible raw score points that students actually obtained for the item. A *p*-value of 0.33 for a given CR item would indicate that, on average, students obtained one-third of the possible points for the item. If the *p*-value were 0.75, this would indicate a much easier item where, on average, students scored 75% of the maximum possible points for the item. As such, the *p*-value indicates difficulty for CR items as well, with lower *p*-values indicating more difficult items.

The item-total correlation indicates the extent to which individual test items provide reliable measurement of the construct being measured by the total test, and it is an index of the item's ability to discriminate between high-ability and low-ability students. For dichotomously scored MC items, the item-total correlations are computed as point-biserial correlations between the score on the item and the score on the remaining items in the test. For CR items, the item-total correlations are computed as Pearson product-moment correlations between the score on the item and the score on the remaining items in the test.[4] 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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

the test tending to score higher on the item than students with low scores on the total test. A low positive value (such as 0.10) indicates a weak relationship between scores on the item and the total score, while a negative value indicates that students who do well on the total test tend to score lower on the item than students who do poorly on the total test.

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

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

For the Fall 2012 WKCE administration, items were flagged for further investigation according to the following rules:

- The *p*-value was less than 0.30 for MC items. Such a *p*-value indicates a difficult item, where fewer than 30% of students obtained the correct answer.
- The item-total correlation was less than 0.15 for the correct answer. A low value may indicate that the item is not providing a high degree of discrimination between high-ability and low-ability students, and, in addition, it may be an indication that the correct answer is in question.
- A distractor had a positive correlation with the total test score.
- The omit rate was greater than 5%.

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

Omit rates may reflect a number of different properties, and an item that is omitted by more than 5% of the students (the WKCE flagging criterion) is not necessarily problematic. Omit rates are typically higher for CR items than for MC items because students who are fairly certain they do not know the answer may be inclined to simply skip the item altogether rather than taking the time to form a response. Items with high omit rates are referred to content specialists

Copyright © 2013 by the Wisconsin Department of Public Instruction

for further review in order to ensure there is no unintended ambiguity in the items. If these flagged items are judged to be clear and provide a valid measurement of the intended knowledge, skill, or ability, then they are retained on the test.

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

Overall, 75 items were flagged on the 23 WKCE 2012 operational tests as meeting the investigational criteria bulleted on the previous page. Of the 75 flagged items that were scored, the number flagged for each of the four criteria is consistent with previous administrations.

Table 8-A shows the number of scored items in the Fall 2012 WKCE 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.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-A. Summary of Flagged Operational Items on the Fall 2012 WKCE

Content	Grade	# of Items Flagged	Number of Flags*			
			Corr <0.15	Distractor Corr >0	Omit >5%	*p*-value <0.30
RD	3	0				
	4	2		2		
	5	5	1	4		
	6	4	2	3		
	7	3	1	3		
	8	5	3	4		
	10	4		4		
MA	3	4		1	2	1
	4	3	2			1
	5	3	1	2		1
	6	2		1		1
	7	4		3	1	
	8	9		3	2	4
	10	3		1	2	
LA	4	4	2	4		
	8	2	2			
	10	5		5		1
SS	4	1		1		
	8	1		1		
	10	3	1	3		1
SC	4	3	3	2		
	8	0				
	10	5	2	3		
Total		75	20	50	7	10

*Note that number of flags may be greater than number of flagged items.

The flagged items were referred to CTB's content specialists for further review to ensure that the items were unambiguous and the answer keys correct. As part of this review, CTB's content experts also evaluated each flagged item against the WKCE depth-of-knowledge (DOK) 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.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-B. Fall 2012 WKCE Reading Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
4	RD	12	MC	0.57	0.31	1.11%		+	0.03		
	RD	28	MC	0.42	0.21	0.62%		+	0.02		
5	RD	9	MC	0.33	0.17	0.18%		+	0.17		
	RD	19	MC	0.47	0.17	2.06%		+	0.05		
	RD	24	MC	0.42	0.17	1.02%		+	0.00		
	RD	37	MC	0.45	0.12	0.61%	+				
	RD	46	MC	0.49	0.16	2.22%		+	0.07		
6	RD	10	MC	0.35	0.20	0.15%		+	0.01		
	RD	19	MC	0.39	0.13	0.14%	+	+	0.11		
	RD	20	MC	0.86	0.14	0.19%	+				
	RD	26	MC	0.44	0.19	0.61%		+	0.04		
7	RD	13	MC	0.34	0.18	0.48%		+	0.02		
	RD	45	MC	0.44	0.09	1.18%	+	+	0.02		
	RD	51	MC	0.32	0.25	0.37%		+	0.07		
8	RD	7	MC	0.54	0.15	0.90%	+				
	RD	34	MC	0.61	0.22	0.52%		+	0.02		
	RD	44	MC	0.44	0.11	0.34%	+	+	0.08		
	RD	46	MC	0.67	0.14	0.28%	+	+	0.04		
	RD	55	MC	0.55	0.24	0.53%		+	0.01		
10	RD	17	MC	0.61	0.35	0.22%		+	0.01		
	RD	27	MC	0.50	0.33	0.33%		+	0.08		
	RD	37	MC	0.72	0.25	0.35%		+	0.03		
	RD	45	MC	0.54	0.15	0.56%		+	0.03		

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-C. Fall 2012 WKCE Mathematics Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
3	MA	17	MC	0.58	0.26	0.24%		+	0.01		
	MA	20A	CR	0.49	0.39	8.90%				+	
	MA	20B	CR	0.26	0.43	3.03%					+
	MA	39A	CR	0.41	0.49	5.68%				+	
4	MA	1	MC	0.97	0.15	0.08%	+				
	MA	11A	CR	0.21	0.32	0.78%					+
	MA	36	MC	0.98	0.14	1.35%	+				
5	MA	20B	CR	0.30	0.45	1.37%					+
	MA	38	MC	0.41	0.24	0.33%		+	0.04		
	MA	41	MC	0.35	0.13	0.66%	+	+	0.03		
6	MA	4	MC	0.48	0.22	0.25%		+	0.05		
	MA	22B	CR	0.26	0.43	0.96%			.		+
7	MA	4	MC	0.68	0.21	0.12%		+	0.04		
	MA	12	MC	0.58	0.26	0.62%		+	0.02		
	MA	38A	CR	0.59	0.35	9.04%				+	
	MA	54	MC	0.54	0.41	0.35%		+	0.03		
8	MA	5	MC	0.48	0.32	0.32%		+	0.04		
	MA	10	CR	0.21	0.50	2.49%					+
	MA	33A	CR	0.31	0.48	6.48%				+	
	MA	33B	CR	0.35	0.60	5.07%				+	
	MA	40	MC	0.66	0.17	1.09%		+	0.03		
	MA	45	MC	0.68	0.50	0.28%		+	0.01		
	MA	49A	CR	0.27	0.44	1.46%					+
	MA	49B	CR	0.21	0.44	2.26%					+
	MA	51	MC	0.25	0.30	0.47%					+
10	MA	33	CR	0.34	0.58	9.98%				+	
	MA	36	CR	0.51	0.65	7.17%				+	
	MA	45	MC	0.58	0.43	0.46%		+	0.03		

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-D. Fall 2012 WKCE Language Arts, Science, & Social Studies Items Flagged for Classical Item Analysis Statistics

Grade	Content	Item	Item Type	*p*-Value	Corr	Omit Rate	Flags				
							Corr	Distractor		Omit	*p*-Value
4	LA	5	MC	0.40	0.13	0.52%	+	+	0.03		
	LA	25	MC	0.44	0.20	1.17%		+	0.10		
	LA	28	MC	0.35	0.26	2.19%		+	0.01		
	LA	30	MC	0.43	0.08	2.61%	+	+	0.10		
8	LA	1	MC	0.64	0.12	0.08%	+				
	LA	2	MC	0.91	0.11	0.07%	+				
10	LA	12	MC	0.65	0.21	0.21%		+	0.01		
	LA	13	MC	0.47	0.16	0.23%		+	0.04		
	LA	21	MC	0.39	0.15	0.44%		+	0.07		
	LA	23	MC	0.49	0.33	0.47%		+	0.02		
	LA	27	MC	0.28	0.18	0.78%		+	0.07		+
4	SC	5	MC	0.93	0.14	0.03%	+				
	SC	24	MC	0.45	0.14	0.46%	+	+	0.07		
	SC	32	MC	0.35	0.13	0.53%	+	+	0.16		
10	SC	3	MC	0.45	0.17	0.10%		+	0.02		
	SC	4	MC	0.42	0.13	3.06%	+				
	SC	31	MC	0.40	0.11	0.53%	+				
	SC	37	MC	0.40	0.35	0.79%		+	0.07		
	SC	45	MC	0.36	0.25	1.50%		+	0.02		
4	SS	6	MC	0.58	0.20	0.26%		+	0.07		
8	SS	4	MC	0.42	0.19	0.25%		+	0.02		
10	SS	23	MC	0.48	0.42	0.75%		+	0.02		
	SS	40	MC	0.19	0.12	0.31%	+	+	0.20		+
	SS	43	MC	0.39	0.41	1.01%		+	0.03		

8.1.1 Flagging for a Positive Distractor Correlation

The distractor correlation coefficients are provided in these tables for items that were flagged because of positive distractor correlations. The distractor correlations tend to be very small and are generally much smaller than the item-total correlations for the correct answer key. All items flagged for a positive distractor had a distractor less than 0.20, and only 4 items flagged for a positive distractor had a distractor greater than 0.10. These items were judged to be acceptable on the basis of their other statistics and were retained in order to meet the WKCE test blueprints.

Copyright © 2013 by the Wisconsin Department of Public Instruction

8.1.2 Flagging for the Item-Total Correlation

Twenty items were flagged for item-total correlations <0.15, and all of the flagged items were 0.12 or above except for five items (Reading grades 7 and 8 (0.09 and 0.11), Language Arts grades 4 and 8 (0.08 and 0.11), and Science grade 10 (0.11)). Although these items, with correlation coefficients ranging from 0.08 to 0.15, are fairly low, the fact that they are positive indicates that the items are contributing information about student ability. These items, therefore, were retained in order to meet the WKCE blueprints.

8.1.3 Flagging for p-Value

Ten items were flagged for *p*-values <0.30, and all six of these items had *p*-values between 0.19 and 0.30. While these statistics indicate items that were very difficult, the number of items flagged for difficulty was very small. Only one of the test forms had more than one item flagged for difficulty.

8.1.4 Flagging for Omit Rate

Seven items were flagged for omit rates greater than 5%. All of the items flagged for omit rates were highly discriminating items. All of the other items flagged for high omit rates had consistently good statistics. All were retained to meet the WKCE blueprints.

8.1.5 Supplemental Tables on Classical Item Analysis

Tables 8-1 through 8-23 present more comprehensive results from the classical item analysis for all of the items retained in each grade and content area. Readers may note that the results presented in these tables may differ slightly from testing results presented on DPI's website due to slight differences in the decision rules defining which students are included or excluded from summary results. Official final results are based on the application of detailed inclusion rules, such as whether the student moved into a school and how long he/she was in one school or another over the course of the year.

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

Table 8-24 summarizes the number of flagged items across grade and content areas. As indicated above, relatively few items were flagged. The item analysis indicated that the *p*-values of the items in the operational tests were well distributed throughout the range of difficulty levels, with point-biserial correlations reasonably high for most items.

Copyright © 2013 by the Wisconsin Department of Public Instruction

8.1.6 Speededness

The degree to which a test is speeded can be evaluated by examining the percentage of students who fail to respond to the final items on a test or the last items in a timed section. One criterion of test speededness currently in use in the testing industry is a rule introduced by Educational Testing Services, which formulates that at least 80% of the test takers should be able to answer all items and all test takers should be able to answer at least 75% of the items (Swineford, 1956). However, a more stringent requirement is often applied, considering tests to be unspeeded only if at least 95% of the examinees attempt the final item. As shown in Table 8-E, WKCE tests satisfy this more stringent requirement, with more than 95% of the examinees attempting the final item in each of the five WKCE content areas.

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

Content	Grade						
	3	4	5	6	7	8	10
Reading	99.33%	98.72%	98.95%	99.23%	99.55%	98.38%	98.98%
Mathematics	99.18%	99.18%	99.70%	99.48%	99.62%	99.49%	99.25%
Language Arts		97.39%				98.25%	98.95%
Social Studies		99.22%				99.01%	99.17%
Science		98.77%				99.41%	98.50%

8.2 Raw Score Results

Raw score results based on all students who took the Fall 2012 WKCE assessment are presented in Table 8-25. In order to facilitate interpretation of the raw score results, Table 8-25 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 observed score, the maximum observed score, reliability (Cronbach's alpha), and the SEM for raw scores. These measurements are further explained below. Readers can refer to Table 3-1 for a count of the number of items in the test and the number of raw score points corresponding to each item.

The mean raw score should be understood by grade and content area and specifically in the context of the maximum possible score points. In Reading, for example, the maximum possible raw score ranges from 56 to 60, and in Mathematics it ranges from 57 to 62.

Test difficulty is computed as the mean raw score divided by the maximum possible score points. Test difficulty ranges from 0 to 1.0. A larger test difficulty value indicates a mean raw score that is closer to the maximum possible score and therefore indicates an easier test. A smaller test difficulty value indicates a mean raw score that is further from the maximum possible score and, therefore, indicates a more difficult test. Consider an example: the test difficulty statistic would be 0.90 if a mean score of 45 were obtained on a test with a maximum

Copyright © 2013 by the Wisconsin Department of Public Instruction

possible score of 50. This would be considered an easier test. On the other hand, test difficulty would be 0.50 if a mean raw score of 25 were obtained on the same test. This would then be considered a more difficult test. In Reading grade 5, the test difficulty statistic (0.65) was obtained by taking the mean raw score of 39.26 and dividing it by 60.

Table 8-25 also shows the skewness and kurtosis statistics for each distribution of raw scores. Skewness and kurtosis describe the shape of a distribution. When a distribution is perfectly normal, skewness is zero. A negative skew indicates a long tail on the left side of the distribution because of the presence of some low scores and (because the mean is sensitive to extreme scores) that most student scores are clustered on the high end of the scale. A positive skew indicates a distribution with some extreme high scores and a corresponding increase in the number of scores below the mean. Kurtosis describes a distribution in terms of its shape relative to a perfectly normal distribution. When a distribution is perfectly normal, kurtosis is zero. A negative kurtosis statistic indicates a distribution that is flatter than a perfectly normal curve, and a positive kurtosis statistic indicates a distribution that has more scores in the center of the score distribution (making it peaked) than a perfectly normal curve. Table 8-25 reveals that in most cases the WKCE 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 WKCE 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 Model Academic Standards.

In addition, Table 8-25 shows the minimum observed score is zero where any student failed all items for each test. The maximum observed score is equal to the maximum number of points possible on the test where any student obtained the full scores for all items. For example, as displayed in Table 8-25, in Mathematics grade 8, there is at least one student who failed all items and at least one student who obtained a perfect raw score of 60.

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: Reliability. For present purposes, the reader should note that measurement error is associated with every test score. A student's true score is the hypothetical average score that would result if the test could be administered repeatedly without the effects of practice or fatigue. Obtained scores should not be regarded as absolute, but as one point within a range that, with a certain degree of probability, includes a student's true score.

The raw score results for each content area are summarized and discussed below using the measurements described above. The raw score results are discussed with reference to the total student population and in terms of subgroup comparisons based on gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. These subgroup comparisons draw from Tables 8-26 through 8-34.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Reading

- Test difficulty ranged from 0.64 to 0.69.
- Standard deviations ranged from 9.84 to 11.51 raw score points.
- Alpha was relatively high in every grade (0.89 to 0.93).
- SEM ranged from 2.97 to 3.20.

Mathematics

- Test difficulty ranged from 0.58 to 0.72, with generally lower difficulty in lower grades and higher difficulty in higher grades.
- Standard deviations ranged from 8.89 to 12.91 raw score points.
- Alpha was relatively high in every grade (0.90 to 0.94).
- SEM ranged from 2.88 to 3.36.

Language Arts

- Test difficulty ranged from 0.64 to 0.72.
- Standard deviations ranged from 5.36 to 6.30 raw score points.
- Alpha ranged from 0.82 to 0.85. As discussed in Part 9, alpha is influenced by test length. All else being equal, shorter tests will tend to have lower reliability than longer tests. The reliability levels are consistent with prior years and are within the expected range given the length of the tests.
- SEM ranged from 2.12 to 2.55.

Social Studies

- Test difficulty ranged from 0.66 to 0.76.
- Standard deviations ranged from 6.34 to 9.01 raw score points.
- Alpha ranged from 0.87 to 0.90. This is consistent with prior years and within the expected range for the length of the tests.
- SEM ranged from 2.27 to 2.85.

Science

- Test difficulty ranged from 0.63 to 0.75.
- Standard deviations ranged from 6.31 to 8.78 raw score points.

Copyright © 2013 by the Wisconsin Department of Public Instruction

- Alpha ranged from 0.86 to 0.88. Alpha was lower in grades 4 and 8 and slightly higher in grade 10. As noted previously, alpha is influenced by test length. Grade 10 has more items than grades 4 and 8 so higher reliability is expected. The alpha levels are consistent with prior years and within expected ranges given the lengths of the tests.
- SEM ranged from 2.40 to 3.00.

Subgroup Performance Patterns in Raw Score Results

Overall, the raw score results show some consistent performance patterns by subgroups, that is, in terms of gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Results can be seen in Tables 8-26 through 8-34.

- In Reading, female students, as a group, had a slightly higher mean raw score than male students in each grade level, with differences ranging from 1.17 points in grade 7 to 2.70 points in grade 8.
- In Mathematics, the raw score differences between genders were very small, ranging from 0.04 point in grade 3 to 0.87 point in grade 10. Although in some grades male students showed the higher raw score and in other grades female students showed the higher raw score, small differences like these suggest that the two groups may be best understood as showing similar performance in each grade.
- In Language Arts, female students, as a group, had a slightly higher mean raw score than male students in each grade level, with differences ranging from 1.42 points in grade 4 to 2.17 points in grade 10.
- In Social Studies, the raw score differences between genders were very small, ranging from 0.02 point in grade 8 to 0.70 point in grade 4. Small differences like these suggest that the two groups may be best understood as showing similar performance in each grade.
- In Science, male students had a slightly higher mean raw score than female students in grades 8 and 10, with differences ranging from 0.07 point in grade 4 to 1.51 points in grade 10.

In all grades and content areas, the raw score results showed consistent performance patterns by ethnicity. In every grade and content area, White students, as a group, had the highest mean raw score, followed by Asian students, American Indian students, Hispanic students, and African American students. American Indian students had a slightly higher mean raw score than Hispanic students. Differences between the mean raw scores of American Indian and Hispanic students were all equal to or less than 0.97 points in Language Arts, 1.89 points in Social Studies, 2.41 points in Mathematics, 2.58 points in Science, and 2.81 in Reading.

In every grade and content area, the mean raw score was higher among those students who were not economically disadvantaged than among those who were economically disadvantaged. The mean raw score difference between the two groups ranged from 3.53 points in Language Arts grade 4 to 9.89 points in Mathematics grade 10.

Copyright © 2013 by the Wisconsin Department of Public Instruction

There were also differences in mean raw scores between students with disabilities and those without disabilities in all grades and content areas. The mean raw score of students without disabilities was consistently higher than the mean raw score of students with disabilities, with differences ranging from 4.16 points in Science grade 4 to 15.06 points in Mathematics grade 10.

In every grade and content area, students who were fully English proficient showed a markedly higher mean raw score than students who were limited English proficient. As might be expected, these differences were generally largest in Reading, where fully English proficient students scored 9.37 to 13.81 points higher (in grades 3 and 10, respectively) than students who were limited English proficient. Mean raw score differences ranged from 4.67 to 14.53 points in Mathematics, 3.42 to 7.72 points in Language Arts, 4.46 to 11.22 points in Social Studies, and 4.65 to 10.85 points in Science.

8.3 Summary Statistics for Scale Scores

The WKCE 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 achievement in that content area. Higher scale scores indicate higher levels of achievement, and lower scale scores indicate lower levels of achievement. 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. Results for all students are described, as are results based on gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Table 8-36 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 observed scale scores, and LOSS and HOSS for all content areas and grades based on the census data. The LOSS and HOSS, as discussed in Part 7, 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. The results for gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency are drawn from Tables 8-37 through 8-45.

Reading

- Mean scale score increased by grade level, ranging from 457.99 to 545.76.
- Standard deviations ranged from 42.17 to 61.82 scale score points.
- In each grade level, except Reading grade 7, student scores spanned the full-scale score range from the LOSS to the HOSS.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Mathematics

- Mean scale score increased by grade level, ranging from 433.95 to 565.19.
- Standard deviations ranged from 44.83 to 52.36 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Language Arts

- Mean scale score increased by grade level, ranging from 294.59 to 450.82.
- Standard deviations ranged from 30.56 to 44.51 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Social Studies

- Mean scale score increased by grade level, ranging from 295.28 to 454.79.
- Standard deviations ranged from 27.21 to 43.74 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Science

- Mean scale score increased by grade level, ranging from 298.34 to 456.65.
- Standard deviations ranged from 29.99 to 43.95 scale score points.
- In each grade level, student scores spanned the full-scale score range from the LOSS to the HOSS.

Subgroup Performance Patterns in Scale Score Results

The scale score results, like the raw score results, showed some consistent performance patterns in terms of subgroups (gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency).

Copyright © 2013 by the Wisconsin Department of Public Instruction

Gender

- In terms of gender, male students, as a group, showed a slightly lower mean scale score in Reading than female students in each grade level. The difference ranged from 5.43 to 12.63 scale score points.
- In Mathematics, the differences between genders were very small, from 0.52 scale score point to 3.31 scale score points, with male students scoring slightly higher than female students in all grades except grade 10.
- In Language Arts, female students scored from 8.33 to 16.28 scale score points higher than male students.
- There were small differences between scale scores by gender in Social Studies, from 0.01 scale score point to 3.03 scale score points, and male and female students alternated between the higher and lower score groups.
- In Science, female students, as a group, showed a slightly lower mean scale score than male students in grades 8 and 10. The difference ranged from 0.62 to 6.84 scale score points.

Race/Ethnicity

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

Socioeconomic Status

- Economically disadvantaged students, as a group, scored lower than students who were not economically disadvantaged across all grades and content areas. Differences ranged from 17.64 scale score points in Social Studies grade 4 to 41.84 scale score points in Reading grade 10.
- For every grade and content area, the mean scale score of students who were economically disadvantaged was more than one-half standard deviation lower than the mean scale score of students who were not economically disadvantaged.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Disability Status

- Students with disabilities and students without disabilities showed consistent and large differences in mean scale score by group. Differences ranged from 42.49 scale score points in Social Studies grade 4 to 47.16 scale score points in Reading grade 10.
- For every grade and content area, the mean scale score of students with disabilities was more than one-half standard deviation lower than the mean scale score of students without disabilities.

English Language Proficiency

- Students who were fully English proficient and students who were limited English proficient showed consistent and large differences in mean scale score by group. Differences ranged from 17.05 scale score points in Social Studies grade 4 to 80.38 scale score points in Reading grade 10.
- For every grade and content area, the mean scale score of limited English proficient students was more than one-half standard deviation lower than the mean scale score of fully English proficient students. These differences increased to approximately one standard deviation at the upper grade levels.

8.4 Cut Scores and Performance Level Classifications

Student performance on the WKCE is reported in terms of four performance categories: *Minimal, Basic, Proficient,* and *Advanced.* These performance categories are established through "cut scores."

Standard 4.19 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) indicates that "[w]hen proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented" (p. 59). In terms of the validity of the WKCE, it is essential to understand that the cut scores were established in a collaborative, participatory process. The two key activities in that process were *standard setting* and *descriptor writing*. Simply speaking, standard setting is a collaborative process of setting cut scores, and descriptor writing is a collaborative process of establishing a plain-language description of what students must know in order to be classified within each of the performance levels established though cut scores.

Performance level descriptors describe the content-based expectations regarding what Wisconsin students should know and be able to do in each grade/content area. Descriptors and cut scores together define, in qualitative and quantitative terms, the differences between a student who is *Proficient* and a student who is not. The Wisconsin Model Academic Standards guided the standard setting and descriptor writing process. These guided participatory processes served to ensure that the performance levels reported for the WKCE reflect the achievement standards and abilities intended by the Wisconsin legislature, teachers, citizens, and DPI.

Copyright © 2013 by the Wisconsin Department of Public Instruction

CTB performed a special study in which the previous WKCE assessments (those that existed until the Fall 2005 WKCE administration) were linked to the current WKCE assessments (those that began with the Fall 2005 WKCE administration) as an important part of setting the cut scores. For details of the linking study, the standard setting activities, and the descriptor writing process, please refer to the Fall 2005 Technical Report (Part 11) and the Fall 2006 Technical Report (Parts 8 and 12), which can be found in Appendices 3 and 2, respectively, of the Fall 2010 WKCE Technical Manual available from the DPI at: http://oea.dpi.wi.gov/oea_publications. Interested readers can also refer to the WKCE 2005 Cut Score Development Report, which can be located at http://oea.dpi.wi.gov/oea_publications.

In 2012 the DPI adjusted cut scores on the WKCE for Reading and Mathematics to better meet the state's policy goals for its students. Therefore, comparisons to the 2011 percentages in each performance level for these subject areas are not appropriate. The adjusted cut scores were established based on the decision to adopt cut-scores for the Reading and Mathematics sections of the WKCE that provide comparable proficiency rates with the National Assessment of Educational Progress (NAEP) Reading and Mathematics assessments. First, NAEP percentile ranks were identified for 4th and 8th grade Reading and Mathematics using Wisconsin student data on NAEP for the Spring 2011 administration. Second, corresponding percentile ranks for grades 5, 6, and 7 were identified using linear interpolation (the grade 4 NAEP percentile rank serve as the basis for the WKCE grade 3 rank and the grade 8 NAEP percentile rank serve as the basis for the WKCE grade 10 rank). Then the WKCE scale scores for the last five administrations that corresponded to the Spring 2011 NAEP percentile ranks were identified by using the WKCE percentile rank tables from the last five administrations. Finally, the rounded average of the derived scale scores from the five administrations was used as the adjusted new NAEP-based cut-score. After applying the new cuts, for Reading, in 2012 on average across grade levels 24.73% of students were classified in the Minimal performance level, 39.34% of students were classified in the Basic performance level, 30.77% of students were classified in the Proficient performance level, and 5.16% of students were classified in the Advanced performance level. After applying the new cuts, for Mathematics, in 2012 on average across grade levels 15.79% of students were classified in the Minimal performance level, 36.22% of students were classified in the Basic performance level, 37.18% of students were classified in the Proficient performance level, and 10.81% of students were classified in the Advanced performance level.

Table 8-46 shows the cut scores for each content and grade level. Tables 8-47 through 8-51 show the percentage of all students in each performance category, as well as subgroup comparisons by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. The results for each content area and grade are summarized below. For ease of reference, Tables 8-52 through 8-56 provide the scale score ranges that define performance levels together with the percentage of students in each performance level.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Reading

- Across all grade levels, approximately 36% of students were either *Proficient* or *Advanced*.
- Approximately 5% of the total student population was classified as *Advanced* in Reading.
- Across all grade levels, approximately 64% of students were below *Proficient*. The difference ranged from 60% below *Proficient* in grade 8 to 67% below *Proficient* in grade 4.

Mathematics

- Across grade levels, over 45% of the student population was either *Proficient* or *Advanced* in Mathematics.
- The proportion of students who were *Advanced* climbed by grade, from 8% in grade 3 to 15% in grade 6, and then declined to 11%, 12%, and 10% in grades 7, 8, and 10, respectively.
- Across all grade levels, approximately 52% of students were below *Proficient*. The difference ranged from 48% below *Proficient* in grade 6 to 55% below *Proficient* in grades 8 and 10.

Language Arts

- At least 61% of the student population was either *Proficient* or *Advanced* in Language Arts.
- In grades 4 and 10, over 74% of students were either *Proficient* or *Advanced*, and in grade 8, 61% of students were either *Proficient* or *Advanced*.
- In grades 4 and 10, 24% and 26% of students, respectively, were below *Proficient*, but in grade 8, 39% of students were below *Proficient*.

Social Studies

- Most of the total student population was either *Proficient* or *Advanced* in Social Studies. The proportion of *Proficient* or *Advanced* students was 92% in grade 4 and 82% in grades 8 and 10.
- A large proportion of students were *Advanced*, especially in grade 4: 65% in grade 4, 45% in grade 8, and 53% in grade 10.
- The proportion of students classified as below *Proficient* was 8% in grade 4 and 18% in grades 8 and 10.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Science

- Approximately 78% of students were either *Proficient* or *Advanced* in Science.
- The percentage of students classified as *Advanced* increased from 20% in grade 4 and 32% in grade 8 to 43% in grade 10.
- Approximately 22% of students in each tested grade level were below *Proficient*.

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 8-49 through 8-53.

In terms of gender, a higher percentage of female students, as a group, were classified as *Proficient* or above in Reading than male students. Conversely, there were a higher percentage of male students than female students in the lowest performance level category in Reading. In Mathematics, the percentage of both genders was approximately equal in the *Proficient* reporting category and in the lowest performance category. However, the percentage of male students was slightly higher than the percentage of female students in the *Advanced* category for all. In Language Arts, there was a markedly higher percentage of female students than male students who were *Proficient* or above, and a much lower percentage of female students than male students in the lowest performance category. In Social Studies, the percentage of female students who were *Proficient* or above was slightly higher than the percentage of male students. There were also more male than female students in the lowest performance category. In Science, the percentages of both genders were approximately equal in *Proficient* or above, although in every grade there were higher percentages of female students who were classified as *Proficient* and there were higher percentages of male students who were classified as *Advanced*.

There were some consistent patterns in performance by ethnicity across grades and content areas. First, in terms of the *Proficient* or above category, the prevailing tendency was that there were higher percentages of White students, as a group, to be classified as *Proficient* or *Advanced*, followed by Asian students, American Indian students, Hispanic students, and African American students. Exceptions to this occurred in Reading grades 6-10 and in all grades of Mathematics, where more Asian students, as a group, were classified as *Proficient* or *Advanced* than White students. Also, in Mathematics grade 7 there were slightly more Hispanic students classified as *Proficient* or *Advanced* than there were American Indian students. The inverse sequence was found at the *Minimal* performance level.

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

Performance level results showed that there were higher percentages of students without disabilities who were classified as *Proficient* or above, and there were much higher percentages of students without disabilities in the reporting category *Advanced.* There were also much lower percentages of students without disabilities in the lowest performance level than students with disabilities. This pattern was evident in all grades and all content areas.

8.5 Standard Performance Indicators 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 weaknesses. This kind of information can be derived from scores on subsets of test items that estimate how much a student knows in a clearly defined skill domain. These skill domains are called content standards (or standards or objectives). Scores on subsets of test items at the content standard level are called standard performance indicator (SPI) scores. The purpose of reporting SPI scores on the WKCE assessments 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 of 77 for a given reporting category means that if the student was given 100 similar items, the student would be expected to answer 77 of them correctly. These are criterion-referenced scores, in that they estimate how much a student knows in a clearly defined skill domain (i.e., the criterion). Technical readers can refer to *TerraNova* 2nd Edition Technical Report (CTB/McGraw-Hill, 2000) for details of the estimation procedures for SPI.

This approach, identifying student proficiency on each content standard, relates to the Wisconsin Model Academic Standards. The SPI provides a more reliable estimate of student achievement on each content standard than is possible by simply reporting percent correct. However, *the SPI information should be used for low-stakes purposes because the SPI cannot be considered stable for any content standard with a small number of items.*

Copyright © 2013 by the Wisconsin Department of Public Instruction

Readers should note that the average difficulty of items will vary across content standards and grades. Content standards vary in their complexity, level of abstraction, and cognitive demand. Some standards may be intrinsically more difficult than others, and the difficulty of individual items is determined, in part, by the difficulty of the content domain being measured. The current test blueprints do not specify the average difficulty level of items for each content standard within grades or across grades. If the difficulty of the items varies across years, grades, and content standards, the mean SPI scores will be affected by differences in item difficulty as well as differences in student ability. *Thus, differences in SPI scores across years, grades, or content standards should not be seen as reliable indicators of differences in student ability since these differences may be explained in whole or in part by differences in the difficulty of the items themselves.* However, comparisons across years, grades, or content standards are appropriate for assessing the relative difficulty of the items, and comparisons of individual student scores or of group mean scores on a single SPI can provide useful information about the *relative* strengths and weaknesses of individual students or groups on these standards.

Tables 8-57 through 8-61 identify the content standards, the number of MC and CR items within each standard, the total number of possible points per standard, 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. Table 8-62 identifies the SPI cut scores for each content area reporting category and grade level. The results from Tables 8-57 through 8-61 are summarized below.

Reading

Table 8-57 presents mean *p*-values and SPI scores for Reading across content standards and grades. The mean of the mean Reading SPI scores across grades and content standards was 66.78%, indicating that the items were moderately difficult for examinees. Results show that the mean *p*-values and SPI scores varied across standards in all grades. Mean SPI scores ranged from 53.49% to 78.12%. In general, the difference between the lowest and highest mean SPI scores was greatest in grade 5 (21.28%). The difference was smallest in grade 6 (5.35%), while content standard 4 (Evaluates/Extends Text) and content standard 3 (Analyzes Text) were the most difficult standards at all grades.

Mathematics

Table 8-58 presents Mathematics *p*-values and SPI scores across grades and content standards. The mean of the mean Mathematics SPI scores across grades and content standards was 66.13%, indicating a moderate degree of difficulty. Results show that the mean *p*-values and SPI scores varied across standards in all grades. Mean SPI scores ranged from 39.81% to 82.83%, with the largest difference observed in grade 3 (where SPI scores ranged from 39.81% to 81.07%). Differences between the highest and lowest mean SPI scores ranged from 12.97% (grade 10) to 41.26% (grade 3). Content standard A (Mathematical Processes) was the most difficult standard in grades 3 through 8. In grade 10, standard D (Measurement) showed lower mean SPI scores, indicating relatively difficult items in the standards.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Language Arts

Table 8-59 presents Language Arts *p*-values and SPI scores across grades and content standards. The mean of the mean Language Arts SPI scores across grades and content standards was 64.95%, indicating a moderate degree of difficulty. Mean SPI scores ranged from 53.20% to 76.61%, with differences between the highest and lowest mean SPI scores of 10.57% in grade 10, 16.54% in grade 8, and 18.65% in grade 4. The mean *p*-values and SPI scores indicated that content standard F (Research and Inquiry) was the most difficult standard in grade 4, content standard D (Language) was the most difficult standard in grade 8, and content standard B (Writing) was the most difficult standard in grade 10.

Social Studies

Table 8-60 presents Social Studies *p*-values and SPI scores across grades and content standards. The mean of the mean Social Studies SPI scores across grades and content standards was 70.56%. While this number is somewhat higher than the mean for the other content areas, this is largely the result of the relatively low difficulty of the grade 4 items, with most of the other grades exhibiting more moderate difficulty. Mean SPI scores ranged from 61.47% to 82.90%, with differences between the highest and lowest mean SPI scores of 16.52% in grade 4, 20.89% in grade 8, and 15.59% in grade 10. The mean *p*-values and SPI scores indicated that the most difficult content standard varied between the three Social Studies grades. In grades 4 and 10, the most difficult standard was C (Political Science) and in grade 8 the most difficult standard was E (Behavioral Science).

Science

Table 8-61 presents Science *p*-values and SPI scores across grades and content standards. The mean of the mean Science SPI scores across grades and content standards was 69.80%. The results indicate that the content standards in grade 10 were considerably more difficult than in grades 4 and 8. Across all grades and content standards, mean SPI scores ranged from 57.79% to 82.11%, with differences between the highest and lowest mean SPI scores of 7.45% in grade 4, 16.21% in grade 8, and 7.93% in grade 10. The mean *p*-values and SPI scores indicated that content standard E (Earth and Space) was the most difficult standard across all grades.

Summary of Student Performance Indicator Results

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

- Grades 4 and 5 Reading standard 4 (Evaluates/Extends Text)
- Grades 3, 4, 6, and 8 Mathematics standard A (Mathematical Processes)
- Grade 4 Language Arts standards F (Research and Inquiry) and D (Language)
- Grade 10 Science standards E (Earth and Space) and F (Life and Environment)

Copyright © 2013 by the Wisconsin Department of Public Instruction

The mean SPI scores are consistent with those found in previous years, suggesting that some of the differences in mean SPI scores across content standards may reflect the differential difficulty of the standards themselves and not merely variations in the difficulty of the particular items that were selected for the test forms. Nevertheless, 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. For this reason, the SPI scores should be interpreted with caution and should not be used to make comparisons of student performance across testing years or grade levels.

Summary of Student Achievement Results

In the WKCE, the purpose of the Reading, Mathematics, Language Arts, 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 validity evidence, indicate that the scale scores and performance levels reported in the WKCE program are valid and reliable evidence of student achievement in the tested content areas and grades. As such, these test scores can be used to classify students, schools, districts, and the state with respect to how much achievement is shown for each content area. Classroom teachers may use these scores as evidence of student achievement in these content areas. District and school administrators may use this information for activities such as planning curricula. At the state level, the overall results can be drawn upon for accountability and reporting purposes associated with *No Child Left Behind* or school improvement initiatives.

Copyright © 2013 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 Fall 2012 WKCE was estimated in four ways:

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

The present chapter addresses AERA, APA, & NCME (1999) standards 2.1, 2.2, 2.10, 2.11, 2.14, and 2.15.

Standard 2.1 advises providing reliability estimates and the SEM for all total scores and subscores reported, standard 2.2 advises reporting SEM in both raw score and scale score units, and standard 2.11 advises that reliability and SEM should be assessed for all population subgroups. To meet these standards, this chapter of the report presents raw score reliability coefficients and SEMs for the five WKCE content areas, for each reported content standard for the total group of examinees, and for subgroups identified by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. The scale score conditional SEMs are provided in Section 7.3.1.

Standard 2.15 advises that when testing measures are used to make categorical decisions, the reliability of those decisions should be estimated. In the present context, standard 2.15 applies specifically to performance level determinations, such as who is *Proficient* or *Advanced*. As described below, the Fall 2012 WKCE 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 WKCE program. This analysis also addresses standard 2.14 by providing a conditional SEM for the cut scores that separate the performance levels.

Standard 2.10 advises reporting measures of inter-rater consistency where subjective judgment is involved in scoring. As we saw in Part 5, CR items were scored by human raters; the process thus involved subjective judgment. As this section will show, a detailed assessment of inter-rater consistency was applied to the WKCE. The assessment conducted is termed inter-rater reliability; it measures the reliability of human raters as they score CR items.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Combined, Cronbach's alpha, SEM, classification consistency, classification accuracy, and inter-rater reliability provide several forms of evidence bearing on the reliability of the WKCE. Cronbach's alpha and the SEM operate at the content level: they provide estimates of reliability for student scores in Reading or Mathematics, for example. Classification consistency and classification accuracy operate on the associated performance level classifications. These are of particular interest in the context of NCLB and the associated AYP requirements. Inter-rater reliability probes further, looking at individual items and evaluating the reliability of the human raters as they assign scores, item by item.

9.1 Measures of Internal Consistency and SEM

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

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

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

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

where SD represents the standard deviation of the raw score distribution and reliability represents Cronbach's alpha.

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

As indicated in Table 9-1, reliability was highest in Reading and Mathematics. Looking at all examinees together in the "Total" column, reliability ranges from 0.89 to 0.93 across grades for Reading, from 0.90 to 0.94 for Mathematics, from 0.82 to 0.85 for Language Arts, from 0.87 to 0.90 for Social Studies, and from 0.86 to 0.88 for Science. Ideally, we would like all reliability coefficients to be 0.90 or above. However, for relatively short tests that are designed to measure a fairly broad range of content, this is not always a realistic expectation. If 0.90 is considered a conservative criterion for an acceptable level of reliability, as measured by Cronbach's alpha, then the grade 3 Reading, grades 4, 8, and 10 Language Arts, grades 4 and 8 Social Studies, and grades 4, 8, and 10 Science tests would not meet this criterion. The reliability coefficients for these tests are consistent with the small number of items (and score points) and the diversity of the content being assessed. Applying the Spearman-Brown prophecy formula to these results indicates that to achieve the 0.90 reliability threshold, the current 56-item test in Reading grade 7 would need to be increased to 63 items, the current 30-item test in Language Arts in grade 4, the 29-item test in grade 8, and the 32-item test in grade 10, would need to be

Copyright © 2013 by the Wisconsin Department of Public Instruction

increased in length to 60, 47, and 55 items, respectively; the current 39-item test in Science grade 4, the 40-item test in grade 8, and the 50-item test in grade 10 would need to be increased to 58, 59, and 62 items, respectively; and the current 38-item test in Social Studies grades 4 and the 40-item test in grade 8 would need to be increased to 52 and 54 items, respectively.

Table 9-1 shows that many of the subgroup reliability coefficients were similar to, albeit slightly lower than, the total reliability coefficients. Reliability coefficients are particularly sensitive to the score distribution and variance, so this result is consistent with the generally larger standard deviations (as previously discussed in Part 8 of this report and summarized in Tables 8-27 through 8-35) among many of these subgroups.

The differences in reliability among most subgroups on most tests were quite small. Differences between male and female students were within 0.04 of one another for all grades and content areas.

The difference between disabled and not disabled and economically disadvantaged and not disadvantaged students was within 0.04 of one another. Most differences among the five racial/ethnic groups also were quite small, within 0.03 of one another for all grades and content areas except Science grade 4, where the reliability for Asian students was 0.04 higher than the reliability for Hispanic students. The greatest differences were between fully English proficient and limited English proficient students, with consistently lower reliability among limited English proficient students.

Table 9-2 presents the raw score SEM for the total population and for the subgroups described above. These values provide important information for raw score interpretation since we can expect that an individual's obtained score will fall within two standard errors of his or her true score approximately 95% of the time. Although there were some observable differences in SEM for the different subgroups, all differences were within one-half of a score point. The SEMs for Reading and Mathematics were larger than those for the other content areas. Because these SEMs are on the raw score scale, this result is consistent with the fact that the Reading and Mathematics tests have more raw score points and larger raw score standard deviations than the other content areas. For every grade and content area, the conditional SEM for individual scale scores are provided in the scoring tables previously discussed in Part 7 (Tables 7-2 through 7-24). The SEM at the *Proficient* cut score was low in all grades and content areas. The SEMs are also plotted in Figures 7-1 through 7-5, with the locations of the cut scores shown in each plot so that the associated SEMs can be easily located.

Reliability, as measured by Cronbach's alpha, was also computed for each content standard within each content area. Table 9-3 shows these reliability coefficients by content standard. The last column presents the reliability for the total content area (with all content standards) for all examinees. It is clear that the reliability per content standard is lower than that for the total test per content area. As discussed above, 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. As discussed in Part 2 of this report, and summarized in Tables 2-1 through 2-5, the targeted number of items per content standard ranged from 5 to 23 items for

Copyright © 2013 by the Wisconsin Department of Public Instruction

Reading,[5] 6 to 15 items for Mathematics, 5 to 20 items for Language Arts, 5 to 13 items for Social Studies, and 6 to 10 items for Science. A lower level of reliability statistics per content standard is therefore expected. The generally lower level of reliability per standard is one of the reasons why the information based on the content standards should be used for low-stakes purposes only (this issue was previously discussed in the context of SPI).

By content standard, the reliability ranges were as follows:

- For Reading, reliability indices by content standard ranged from 0.42 (for standard 4 in grade 4 with 6 items) to 0.85 (for standard 3 in grade 3, with 22 items).

- For Mathematics, reliability indices by content standard ranged from 0.49 (for standard C in grade 5 with 10 items) to 0.77 (for standard D in grade 10, with 10 items).

- For Language Arts, reliability indices by content standard ranged from 0.34 (for standard F in grade 8 with 6 items) to 0.80 (for standard B in grade 8, with 19 items).

- For Social Studies, reliability indices by content standard ranged from 0.47 (for standard C in grade 4, with 7 items, and content standard E in grade 8, with 5 items) to 0.72 (for standard C in grade 10 with 12 items).

- For Science, reliability indices by content standard ranged from 0.22 (for standard D in grade 4, with 6 items) to 0.72 (for standards G/H in grade 10 with 10 items).

The SEM associated with each content standard is presented in Table 9-4 by content area and grade level. Some differences in SEM by content standard can be observed. As indicated by the discussion above, these SEMs were smaller than those for the total test and are generally consistent with the number of items within each content standard.

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

9.2 Classification Consistency and Accuracy

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

[5] Note that content standard 4 at grade 3 contains 5 items but is worth 7 points because it includes four MC items and one 3-point CR item. Therefore, the point values for Reading range from 7 to 25 points.

Copyright © 2013 by the Wisconsin Department of Public Instruction

consistency and accuracy of these classifications were evaluated, and provides evidence supporting the validity of these classifications.

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

A contingency table representing the probability of particular classification outcomes under specific scenarios is a convenient way to measure classification consistency. The table below is a contingency table of (H+1) × (H+1), where H is the number of cut scores. Three cut scores yield a 4 × 4 contingency table, as can be seen below in Table 9-A.

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

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

Table 9-A
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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

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

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

9.2.1 Kolen and Kim's Method for Pattern Scoring

As stated in Part 7, when item response theory (IRT) is applied to score examinees' responses, two types of scoring are available: number-correct scoring and item-pattern scoring. WKCE uses item-pattern scoring. Many methods of estimating the consistency and accuracy of classification based on number-correct scoring have been suggested in the psychometric literature. However, there have been relatively few studies dealing with item-pattern scoring based on IRT. Kolen and Kim (2004) suggest a simple procedure for pattern scoring (KKM) based on IRT and simulated item responses. KKM requires a simulation of item responses as follows:

Step 1: Obtain item parameters (**I**) and the ability distribution weight ($\hat{g}(\theta)$) at each quadrature point.

Step 2: Compute two ability estimates at each quadrature point. At a given quadrature point, θ_j, generate two sets of item responses using the item parameters from a test form, assuming that the same test form was administered twice to an examinee with the true ability θ_j.

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

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

If two parallel (or alternative) forms (e.g., Form 1 and Form 2) are available, the two response patterns can be generated based on the item parameters from the two forms.

Step 3: Construct a classification matrix at each quadrature point. Determine the joint event for the cells in Table 9-B using the two ability estimates obtained from Step 2.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-B
Classification Table for One Cut Point (C_1)[6]

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

Step 4: Repeat Steps 2 and 3 *R* times and get average values over *R* replications. *R* should be a large number (e.g., 500) to obtain stable results.

Step 5: Multiply distribution weight ($\hat{g}(\theta)$) by the average values in Step 4 for each quadrature point and sum across all quadrature points. From this, a final contingency table and classification consistency indices, such as kappa, can be computed.

Because examinees' abilities are estimated at each quadrature point, this quadrature point can be considered the true score. Therefore, classification accuracy is computed using both examinees' estimated abilities (observed scores) and quadrature point (true score). Just as 0.90 is generally considered the criterion for acceptable test score reliability, the criterion value of 0.90 is considered to be an acceptably high level of classification accuracy.

As can be seen in Tables 9-5 through 9-27, there are two tables for each grade and content area. The first table is a contingency table with all three cut scores, which was prepared based on the KKM procedure. The rows represent the first administration of an assessment, and the columns represent the second administration of the same assessment to the same students. As mentioned above, in the KKM procedure the score distributions for the first administration and the second administration are estimated using a simulation. So, the value in each cell represents the probability of belonging to a particular pair of performance levels in the first administration and the second administration. For example, in Reading grade 3, 0.28 represents the probability of belonging to *Minimal Performance* in both the first and second administrations. The 0.03 represents the probability of belonging to *Proficient* in the first administration and *Advanced* in the second administration. "Sum" is obtained simply by adding the four row values or the four column values. This "Sum" is not always identical to the sum of the values shown in the table because the values displayed have been rounded to two decimal places.

The second table shows indices for classification consistency and classification accuracy. Because there are four performance levels for the WKCE, there are three cut scores. The values in "All Cuts" were obtained by applying all three cuts together. In Table 9-5 for Reading grade 3, when all three cuts were used for the computation, classification consistency (P) is 0.75, chance probability is 0.28, kappa (*k*) is 0.66, 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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

applied (i.e., performance levels above and below the cut). It is clear that the values for P, kappa, 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 AYP reports, the reliability values for this second cut need to be considered carefully. In Table 9-5, for example, the P for the second cut, which establishes the *Proficient* performance level, was 0.89, kappa was 0.77, and classification accuracy was 0.93. The interpretation of the values illustrated for Table 9-5 is the same for Tables 9-6 through 9-27.

When only the *Proficient* cut score was applied, P was greater than or equal to 0.86 and kappa was greater than or equal to 0.60 for all Reading and Mathematics tests. For Language Arts, the average P associated with the *Proficient* cut was 0.88 and the lowest kappa was 0.66. In Social Studies, the lowest P associated with the *Proficient* cut was 0.91 and the lowest kappa was 0.71. For Science, the lowest P was 0.90 and the lowest kappa was 0.71. According to Landis and Koch's criteria for kappa (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.

Figures 9-1 through 9-5 also show P, kappa, and classification accuracy when students were classified based on "All Cuts." These values are provided in Tables 9-5 through 9-27, but the results are also provided in the plots for ease of understanding. As can be seen in the plots, all grades and content areas indicated classification consistency (P) based on all cuts over 0.71 for all grades in Reading, Mathematics, Social Studies, and Science. In Language Arts, P was 0.67, 0.66, and 0.71 in grades 4, 8, and 10, respectively. The values of kappa were greater than 0.50 for all grades and content areas. In summary, based on the Landis and Koch criteria, all test forms showed good agreement.

9.3 Inter-Rater Reliability for CR Items and Writing Prompts

The reliability of handscoring may be measured in a variety of ways. Two of the most effective ways are 1) tabulations of exact and adjacent agreement and 2) reliability coefficients. Reliability for CR items is typically examined by calculating indices of inter-rater agreement, the degree of reliability with which different human raters assign scores to a given student response. Two indices for inter-rater reliability, intraclass correlation and weighted kappa, are presented here.

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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

where

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

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

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

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

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

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

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

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

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

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

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

If agreement is perfect, ρ_{IC} =1. Always, $0 \le \rho_{IC} \le 1$.

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. Always, $0 \le k \le 1$.

Ordinal rating scales (e.g., 0, 1, 2) used in scoring CR items contain a certain level of chance agreement that is expected. Although the intraclass correlation is reported in this report, it does not take into account the possibility of chance agreement between the two raters, but Cohen's kappa does take this into consideration. In general, kappa will have values equal to or smaller than the intraclass correlation. If agreement is perfect, then the value of kappa is 1.0. If agreement is at chance levels, the value of kappa is zero. As noted in Section 9.2.1, Landis and Koch (1977) suggest that values of kappa greater than 0.75 indicate "excellent agreement," values between 0.40 and 0.74 represent "good agreement" beyond chance, and values below 0.40 denote "poor agreement." Specific criteria for intraclass correlation or weighted kappa are not established.

Tables 9-28 through 9-30 present the rater agreement statistics for CR items and the Writing prompt. The evidence supporting inter-rater reliability is presented in terms of the percentage of agreement between raters, two indices of inter-reliability, and the distributions of

Copyright © 2013 by the Wisconsin Department of Public Instruction

scores across score levels. In the table, “Perfect” agreement is defined as scores that are exactly the same. “Adjacent” agreement is defined as scores differing by one point. “Discrepant” cases are those cases where the scores of the two raters differed by more than one raw score point. The column for “Codes” reflects the number of students who received the condition codes A, B, C, or D, which indicate illegible responses, responses that are off-topic, blank responses, or in another language, respectively. “Mean” reflects mean score. “Number of Reads” is exactly two times the number of papers submitted for the purpose of computing inter-rater reliability, as each paper submitted for that purpose is scored twice. The “Frequency” column represents the scoring outcomes for the student responses based on the raw scores given by each of the two raters. For example, as shown in Table 9-28, for Reading grade 4, item 13, the perfect agreement, adjacent agreement, discrepant agreement, and codes are 71%, 26%, 1%, and 2%, respectively.

For Reading, Mathematics, and Writing, student responses were scored by a single rater. To calculate inter-rater reliability, 5% of the responses were scored by a second rater.

The inter-rater reliability results for Reading, Mathematics, and Writing are discussed separately in the following sections. Overall, the results indicate a high degree of reliability for scores on the handscored items in all three content areas.

Reading

Inter-rater reliability results for Reading CR items are shown in Table 9-28. Overall, the rater agreement was very high. The mean percentage of non-discrepant ratings (i.e., perfect agreement plus adjacent scores), averaged across all items, was approximately 96%. As noted in Section 9.2.1, Landis and Koch (1977) suggest that values of kappa greater than 0.75 indicate “excellent agreement,” values between 0.40 and 0.74 represent “good agreement” beyond chance, and values below 0.40 denote “poor agreement.” The mean kappa across all items was approximately 0.74.

Each of the Reading CR items had a maximum possible score of 3. The percentage of discrepant (i.e., nonadjacent) ratings was 4% or less for each of the operational CR items.

The percentages of discrepant ratings for the Reading CR items are summarized below. For these operational CR items, the results were as follows:

- 1% discrepant ratings—8 items (57%)
- 2% discrepant ratings—4 items (29%)
- 3% discrepant ratings—1 items (7%)
- 4% discrepant ratings—1 items (7%)

The percentage of responses with condition codes ranged from 1% to 8% across all items; the percentage exceeded 4% for only two items (14% of the 14 items). The mean intraclass correlation, averaged across all items, was 0.87. Intraclass correlations ranged from 0.79 to 0.93, and weighted kappa ranged from 0.54 to 0.86.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Mathematics

Table 9-29 provides the inter-rater reliability results for the Mathematics CR items. Overall, the rater agreement was high. The mean percentage of non-discrepant ratings (i.e., perfect agreement plus adjacent scores), averaged across all items, was approximately 96% for operational items. The mean kappa across all items was approximately 0.93.

Treating the two-part CR items as separate items, the maximum possible points per CR item ranges from one to two points. The percentage of discrepant (i.e., nonadjacent) ratings was 4% or less for all 46 operational CR items.

The percentages of discrepant ratings for the Mathematics CR items are summarized below. For these operational CR items, the results were as follows:

- No discrepant ratings—22 items (48%)
- 1% discrepant ratings—17 items (37%)
- 2% discrepant ratings—4 items (9%)
- 3% discrepant ratings—1 item (2%)
- 4% discrepant ratings—2 item (4%)

The percentage of responses with condition codes ranged from 1% to 10% across all items; the percentage exceeded 4% for 9 items (20% of the 46 items). The mean intraclass correlation, averaged across all items, was 0.97. Intraclass correlations ranged from 0.88 to 1.0, and weighted kappa ranged from 0.77 to 0.99.

Writing

Table 9-30 shows inter-rater reliability results for the Writing prompts. As indicated previously, the Writing prompts were scored on two rubrics, the Composing Rubric (six points) and the Conventions Rubric (three points). Table 9-30 shows that the rate of perfect agreement was lower on the 6-point Composing Rubric than on the 3-point Conventions Rubric. The difference is due to the difference in score points. Perfect agreement is, as discussed previously, less likely with a higher number of possible score points than with a lower number of possible score points. Adjacent and discrepant modes of agreement were, as may also be expected, more common where there were more possible score points. Perfect agreement ranged from 60% to 67% on the Composing Rubric and from 87% to 96% on the Conventions Rubric. Adjacent agreement ranged from 30% to 33% on the Composing Rubric and from 3% to 10% on the Conventions Rubric. The percentage of discrepant (i.e., nonadjacent) ratings for the Writing prompt in grades 4, 8, and 10 ranged from 2% to 3% for the Composing Rubric and was 0% for the Conventions Rubric. Codes were generated in 1% to 4% of the cases. Intraclass correlation ranged from 0.84 to 0.92, and weighted kappa ranged from 0.69 to 0.84.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Summary

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 10: Validity

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) defines validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests" (p. 9). The purpose of test score validation is not to validate the test itself, but to validate interpretations of the test scores for particular purposes or actions. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process. Every aspect of an assessment provides evidence in support of (or a challenge to) its validity, including design, content specifications, item development, psychometric quality, and inferences made from the results.

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

Part 2 of the Technical Report described the involvement of Wisconsin educators, DPI, and CTB in the test development process. As indicated in Part 2, the test development process and the involvement of Wisconsin educators in that process formed an important part of the validity of the entire WKCE. The knowledge, expertise, and professional judgment offered by Wisconsin educators ultimately ensured that the content of the WKCE formed an adequate and representative sample of appropriate content, and that the content formed a legitimate basis upon which to derive valid conclusions about student achievement.

Part 3 of the Technical Report addressed the issue of test form development. Part 3 provided a general discussion of CTB's test book creation and editing process, the process of selecting operational test items, and the process of obtaining DPI approvals. The test design process and the participation of Wisconsin educators in the process of test selection, including item content and bias reviews, provide a solid rationale for having confidence in the content and design of the WKCE and using it as a tool from which to derive valid inferences about Wisconsin student performance. Parts 2 and 3 together provided evidence to support the content validity of the WKCE and addressed AERA/APA/NCME (1999) standards 1.2, 1.6, 3.1, 3.2, 3.3, 3.5, 3.6, 3.7, 3.9, 3.11, 3.16, 6.4, 6.15, 7.3, 7.4, 7.7, 13.3, and 13.5.

Part 4 of the Technical Report described the process, procedures, and policies that guided the administration of the WKCE, including accommodations, security, and the written procedures provided to test administrators and school personnel. The following AERA, APA, & NCME (1999) standards were addressed: 1.13, 3.3, 3.19, 3.20, 3.21, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 6.11, 6.15, 9.1, 10.1, and 10.2. The process, procedures, and policies detailed in that section contributed to the validity of the WKCE assessments by reducing the impact of construct-

Copyright © 2013 by the Wisconsin Department of Public Instruction

irrelevant variables (such as non-standardized administration methods, limitations associated with student disabilities, security breaches, etc.) on test performance.

Part 5 of the Technical Report demonstrated adherence to AERA/APA/NCME AERA, APA, & NCME (1999) standards 3.22, 3.23, 3.24, 5.8, and 5.9. It described how MC items and CR items were scored, the handscoring process, the training and selection of readers, the scoring rubrics used for scoring CR items, and the resulting score distributions. The procedures described in that section contributed to the validity of the WKCE assessments by preventing hardware- or software-related errors in machine scoring, and reducing construct-irrelevant score variance associated with variations in raters' interpretation and application of scoring rubrics.

Part 6 described the sample data used for calibration and scaling, referring the reader to information found in the 2010 WKCE Technical Report (and the 2008 WKCE Technical Report for Reading Grade 10).

Part 7 of the Technical Report described the calibration and equating methods, as well as processes and procedures for deriving scale scores from response patterns. Some references to introductory and advanced discussions of IRT were provided. Several axes upon which to evaluate the calibration and equating 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, were all discussed. Part 7 of this report thereby addressed AERA, APA, & NCME (1999) standards 1.13, 4.1, 4.2, 4.3, 4.10, 4.11, 7.1, 7.2, and 7.10. These processes and procedures contributed to the validity of the WKCE by providing the opportunity to identify and eliminate items that were not contributing to the accurate and reliable measurement of the intended constructs, and by ensuring that valid comparisons of the WKCE test scores can be made within and across years.

Part 8 presented 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 NCLB policy environment, as well as interests of parents, students, and educators. The SPI scores then probed further, assessing specific skills and abilities. Combined, scale scores, performance levels, and SPI scores provided a comprehensive set of tools to assess Wisconsin student performance by content and grade level and by gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Part 8 thus addressed AERA, APA, & NCME (1999) standards 1.5, 3.18, 4.3, 4.5, 4.6, 4.7, 4.19, 7.1, 7.10, 13.15, and 13.19. The analyses addressed in Part 8 contributed to the validity of the WKCE by providing further opportunity to identify and eliminate items that were not contributing to the accurate and reliable measurement of the intended constructs.

Part 9 demonstrated adherence to AERA, APA, & NCME (1999) standards through several analyses of the reliability of the Fall 2012 WKCE. It presented a reliability analysis using Cronbach's alpha, SEM results, a detailed analysis of classification consistency and classification accuracy, and a full analysis of inter-rater reliability. The Fall 2012 WKCE Technical Report thereby addressed AERA, APA, & NCME (1999) standards 2.1, 2.2, 2.10, 2.11, 2.14, and 2.15. Reliability is a prerequisite to score validity, and the analyses in that section

Copyright © 2013 by the Wisconsin Department of Public Instruction

contributed to the WKCE validity evidence by establishing the reliability of the WKCE test scores and proficiency classifications.

In the subsequent pages, Part 10 will, as stated, present additional metrics with which to evaluate the validity of the WKCE program. As described below, the WKCE program formally assessed the issue of test bias through an analysis of differential item functioning (DIF). It is possible for items to function differently among 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 if that possibility occurred and to what degree, item by item, for each of the categories of gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. This analysis specifically addresses standards 7.1, 7.2, and 7.3.

The present chapter also provides estimations of construct validity. Two measures are provided: correlations among 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 Reading ability or Mathematics ability. The presence of a single, underlying trait is a fundamental issue when scaling and analyzing results through IRT models. As such, these analyses are essential elements in assessing the validity of the WKCE. Finally, this chapter outlines the erasure analysis procedures that were employed to ensure the integrity of test scores by identifying test papers that may have been fraudulently altered.

10.1 Differential Item Functioning

An empirical differential item functioning (DIF) approach was used to examine potential item bias, and to determine if 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, 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.

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Three kinds of DIF statistics were used: Linn-Harnisch, Mantel (or Mantel-Haenszel), and standardized mean difference. Each of these DIF methods can be used to determine if identified groups of examinees with the same underlying level of ability had the same probability of correctly responding to the item. The Mantel-Haenszel method is applied to MC items only. The Linn-Harnisch method is used for both MC and CR items. The Mantel statistic and standardized mean difference are applied to CR items. These DIF statistics and the flagging criteria are described in detail below.

(1) Linn-Harnisch (L-H)

Because the WKCE was built using item response theory (IRT), an appropriate procedure for examining item bias should reflect the IRT model. Several IRT-based procedures are available, such as a procedure that tests the equality of item parameters across groups (Lord, 1980) or any of the procedures that assess the differences in the area between the item characteristic curves (e.g., Linn, Levine, Hastings, & Wardrop, 1981). However, these procedures require a minimum of 800 to 1,000 cases in each group to make reliable comparisons. A procedure that still relies on the predictions of the three-parameter model but does not require as many cases has been suggested by Linn and Harnisch (1981).

To take an example, in the case of gender DIF analyses, item parameters (e.g., discrimination, location, and guessing) and the scale score (θ) for each examinee were estimated using the three-parameter logistic model for MC items and the two-parameter partial credit model for CR items. The sample was then divided into male and female gender subgroups. The members in each group were sorted into ten equal score categories (deciles) based upon their location in the scale score (θ) range. The expected proportion correct for each group based on the model prediction was compared to the observed (actual) proportion correct obtained by the group. The proportion of students in decile g who are expected to answer item i correctly is:

$$P_{ig} = \frac{1}{n_g} \sum_{j \varepsilon g} P_{ij},$$

where n_g is the number of examinees in decile g. To compute the proportion of students expected to answer item i correctly (over all deciles) for a specific subgroup, the following statistic was computed

$$P_{i\cdot} = \frac{\sum_{g=1}^{10} n_g P_{ig}}{\sum_{g=1}^{10} n_g}.$$

Copyright © 2013 by the Wisconsin Department of Public Instruction

The corresponding observed proportion correct for examinees in a decile (O_{ig}) is the number of examinees in decile g who answered item i correctly divided by the number of students in the decile (n_g). That is,

$$O_{ig} = \frac{\sum_{j \varepsilon g} u_{ij}}{n_g},$$

where u_{ij} is the dichotomous score for item i for examinee j. The corresponding formula to compute the observed proportion answering each item correctly (over all deciles) for a subgroup is given by

$$O_{i\cdot} = \frac{\sum_{g=1}^{10} n_g O_{ig}}{\sum_{g=1}^{10} n_g}.$$

After the values are calculated for these variables, the difference between the subgroup's observed proportion correct and expected proportion correct can be computed. The decile group difference (D_{ig}) for the observed and expected proportions correctly answering item i in decile g is

$$D_{ig} = O_{ig} - P_{ig},$$

and the overall group difference (D_i) between the observed and expected proportions correct for item *i* in the complete group (over all deciles) is

$$D_{i\cdot} = O_{i\cdot} - P_{i\cdot}.$$

These indices are indicators of the degree to which subgroup members performed better or worse than expected on each item based on the parameter estimates from all subgroups. Differences for decile groups provide an index for each of the ten regions on the scale score (θ) range. The decile group difference (D_{ig}) can be either positive or negative. Use of the decile group differences, as well as the overall group difference, allows one to detect items that give a large positive difference in one range of θ and a large negative difference in another range of θ, yet have a small overall difference.

DIF is defined in terms of the decile group and total target subsample differences, the D_{i-} (sum of the negative group differences) and D_{i+} (sum of the positive group differences) values, and the corresponding standardized difference score for the subsample (Linn & Harnisch, 1981, p. 112). The standardized difference score ($Z_{i\,g}$) for ability group *g* is computed as

Copyright © 2013 by the Wisconsin Department of Public Instruction

$$Z_{ig} = \frac{1}{n_g} \sum_{j \in g} \left[\frac{U_{ij} - P_{ij}}{\sqrt{P_{ij}(1 - P_{ij})}} \right],$$

where U_{ij} = 1 when student j answers item i correctly, and U_{ij} = 0 otherwise. The standardized difference over all the ability groups is

$$Z_{i.} = \frac{\sum_g n_g Z_{ig}}{\sqrt{\sum_g n_g^2}}.$$

Items for which $|D_i| \geq 0.10$ and $|Z_i| \geq 2.58$ are flagged for DIF. If D_i is positive, the item is biased in favor of the focal group. If D_i is negative, the item is biased against the focal group.

(2) Mantel and Mantel-Haenszel (M-H)

The Mantel (1963) and Mantel-Haenszel (1959) chi-square statistics are used to evaluate potential bias in individual items by examining item-level differences between different groups of students (e.g., students classified by gender, ethnicity, disability, or other variables of interest), controlling for differences in the relevant ability or abilities measured by the test. In this procedure, subgroups are matched by their raw total test score using a contingency table with K levels. The Mantel statistic is computed by first dividing students into K levels of ability on the total test, then comparing the performance of these matched groups using the formula

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

where F_k is the sum of scores for the focal group at the kth level of the matching variable, and $E(F_k)$ is the expected sum of scores for the focal group at the kth level of the matching variable.

For dichotomous items, the Mantel statistic is equivalent to the Mantel-Haenszel statistic without the continuity correction (Zwick, Donoghue, & Grima, 1993). With the continuity correction added (Holland & Thayer, 1986), the Mantel-Haenszel statistic has the form

$$Mantel - Haenszel \ \chi^2 = \frac{(|\sum_k F_k - \sum_k E(F_k)| - 1/2)^2}{\sum_k Var(F_k)},$$

Copyright © 2013 by the Wisconsin Department of Public Instruction

with all terms defined as in the prior equation.

In addition to the Mantel-Haenszel chi-square statistic, the delta statistic (Δ_{MH}) was computed for all MC items (Holland & Thayer, 1985). To compute delta, the odds ratio α is first computed as

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

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.

The Δ_{MH} statistic is then computed as

$$\Delta_{MH} = -2.35 \ln(\alpha_{MH}).$$

Positive values of Δ_{MH} indicate items that favor the focal group, whereas negative values of Δ_{MH} indicate items that favor the reference group. WKCE MC items were flagged for DIF using the following criteria (Zwick, Donoghue, & Grima, 1993):

- A= No DIF: Non-significant Mantel-Haenszel χ^2 or $|\Delta_{MH}|<1.0$
- B= Weak to moderate DIF: Mantel-Haenszel χ^2 is significantly greater than zero ($p < 0.05$) and $1.0 < |\Delta_{MH}| < 1.5$
- C= Large DIF: Mantel-Haenszel χ^2 is significantly greater than zero ($p < 0.05$) and $|\Delta_{MH}|$ exceeds 1.5.

For CR items, an effect size (ES) statistic based on the Mantel Haenszel χ^2 was used. ES is obtained by dividing the standardized mean difference (SMD) statistics by the standard deviation of the item (detailed description of these procedures can be found in Zwick, et al., 1993). WKCE items are flagged using the same rules that are used in The National Assessment of Educational Progress (NAEP):

Copyright © 2013 by the Wisconsin Department of Public Instruction

- No DIF: Non-significant Mantel χ^2 or |ES| < 0.17
- Weak to moderate DIF: Mantel χ^2 is significant ($p < 0.05$) and $0.17 \leq |ES| < 0.25$
- Large DIF: Mantel χ^2 is significant ($p < 0.05$) and $|ES| \geq 0.25$

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group.

(3) Standardized Mean Difference (SMD)

A standardized mean difference statistic (SMD) was also computed for CR items. The SMD is an effect size index of DIF which is relatively easy to interpret (Zwick, et al., 1993). The SMD compares the means of the reference and focal groups, adjusting for the distribution of reference and focal group members on the conditioning (i.e., matching) variable (Zwick, et al., 1993). SMD is computed as (Zwick, et al., 1993)

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

where

p_{Fk} = proportion of the focal group members at the kth level of the matching variable,

$m_{Fk} = 1/N_{F1k}$, where N_{F1k} is the number of correct responses in the focal group at ability level k, and

$m_{Rk} = 1/N_{R1k}$, where N_{R1k} is the number of correct responses in the reference group at ability level k.

A negative SMD value indicates an item on which the focal group has a lower mean than the reference group. A positive SMD value indicates an item on which the reference group has a lower mean than the focal group. An item is flagged when

$$|ES - SMD| \geq 0.25.$$

Results

Tables 10-1 through 10-7 show items flagged based on the criteria described previously. Readers may note that some items are flagged by both Linn-Harnisch and Mantel-Haenszel methods and some only by one of the methods. For the Linn-Harnisch, Mantel, and Mantel-Haenszel methods, the summary flag information in the DIF tables is always expressed with reference to the focal group. That means that negative flags (such as -B or -C, as described above) indicate that an item disadvantages the focal group, such as female students, African American students, or economically disadvantaged students. A positive flag indicates that the item favors the focal group. The B flag represents a lower threshold for DIF. Only items that were flagged with a C flag were included in the tables described below. Readers can see

Copyright © 2013 by the Wisconsin Department of Public Instruction

B-flagged items in the tables, but that occurs because those items were also flagged with a C flag.

The DIF results for gender are presented in Table 10-1; results for race/ethnicity are presented in Tables 10-2 through 10-5; English language proficiency (ELP) results are presented in Table 10-6; and results based on disability status are presented in Table 10-7. No items were flagged for DIF for socioeconomic status.

Each DIF table references the grade and content area of the items flagged for DIF, as well as the test form, the item number, and the item type. The tables present Linn-Harnisch statistics (D+, D-, and *Z*) first, then the SMD, and finally the Mantel or Mantel-Haenszel statistic (Δ_{MH}). MH is only computed for the focal group. After specifying these statistics for each item, two final columns provide a summary flag status. There is a column "LH Flag" to indicate where any of the Linn-Harnisch statistics produced a flag and a "MH Flag" column to indicate where either Δ_{MH} or the SMD produced a flag.

In Table 10-1, looking at all items and all grades and content areas, 12 items were flagged for gender DIF. More items were flagged in the Reading tests than in the other content area. This is expected given that more grades are tested in Reading. The number of flagged items in Reading relative to the other content areas should be understood in this context. Note that eight of the 12 items flagged by Linn-Harnisch indicate that the DIF favors (rather than disadvantages) female students.

The other DIF results in Tables 10-2 through 10-7 can be understood in the same fashion. Note that a single item can be flagged for multiple subgroup categories, such as for African American students, Hispanic students, and socioeconomically disadvantaged students. Readers should also note that Linn-Harnisch DIF statistics cannot be computed unless the sample sizes are at least 50, with at least five students per group in each decile. In some cases (as noted in the DIF table for American Indian students) the size of the tested population was too small to include valid Linn-Harnisch DIF statistics. DIF results for focal groups containing fewer than 100 students may be unstable and should be interpreted with caution.

The Fall 2012 WKCE tests were developed using procedures to minimize item and test bias. Expertise in this area is not, however, a substitute for statistical analyses of the items. Combined, the DIF statistical analyses discussed above and the expert reviews provide an appropriate set of tools with which to minimize the extraneous or construct-irrelevant information associated with item bias, or DIF, in the WKCE. However, in large-scale assessments, such as the WKCE, it is expected that some items will show DIF. All of the items in the Fall 2012 WKCE flagged for DIF were notated as such in the classical item analyses and in the item pool so that content experts would be able to reevaluate these items in future item selection activities. Items with DIF (particularly items flagged for strong DIF) are to be avoided in future selections.

Copyright © 2013 by the Wisconsin Department of Public Instruction

10.2 Construct Validity

Construct validity can be defined as the extent to which tests measure the skills or constructs they intend to measure, and it is the central concept underlying the Fall 2012 WKCE assessment validation process. Evidence for construct validity is comprehensive and integrates evidence from both content- and criterion-related validity. The WKCE test development process included specifications, item writing, review, and test construction.

Threats to construct validity include the unintended measurement of variables unrelated to the desired constructs and multidimensionality of the tests. To ensure that the test items are focused on the desired constructs, standardized procedures are employed to select items with sound statistical properties to align the items to content standards, and to ensure that each test form meets the WKCE blueprint. A test can be said to be unidimensional when all of the items in the test measure the same underlying ability or trait.

Analyses of the internal structure of a test can indicate the extent to which the relationships among 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 among content standards within a test can be expected to be relatively high. Tables 10-8 through 10-12 show the correlations among content standards for each WKCE content area. The correlation coefficients here reflect the degree of linear relationship and direction between any two given content standards. The correlation can range from +1 to -1. A correlation of +1 indicates a perfect positive linear relationship, and a correlation of -1 indicates a perfect negative linear relationship between two content standards. A correlation of zero means there is no linear relationship. In general, the size of the correlation coefficient is influenced by the number of items or score points and by the score variance. Readers are cautioned not to confuse correlation with causation. The presence of a high correlation between two content standards should not be taken as an indication that there is a causal relationship between them.

As may be observed in Tables 10-8 through 10-12, correlations among content standards were generally higher in Reading than in the other content areas. The correlations among content standards ranged from 0.50 to 0.84 in Reading, from 0.47 to 0.75 in Mathematics, from 0.38 to 0.69 in Language Arts, from 0.47 to 0.69 in Social Studies, and from 0.32 to 0.67 in Science. 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 among 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.

WKCE test items are calibrated using unidimensional IRT models, which posit that the test items are measuring an essentially unidimensional construct. To assess the dimensionality of the WKCE assessments, a principal components analysis was conducted for each content area and grade. Principal components analysis is a statistical technique commonly used to evaluate dimensionality by detecting patterns of relationships among items. This method is useful in determining whether the observed scores on a test can be explained largely or entirely in terms of a much smaller number of components. To take an example, if answering the Mathematics items

Copyright © 2013 by the Wisconsin Department of Public Instruction

in a Mathematics test required a lot of reading ability, the Mathematics test would not be only a measure of mathematics ability, it would be a measure of reading ability as well. Such a test would be said to be multidimensional rather than essentially unidimensional. One way of evaluating the dimensions detected in the analysis is by examining the eigenvectors and eigenvalues. In principal components analysis, the eigenvectors correspond to factors, and the eigenvalues correspond to the variance explained by these factors. The sum of the eigenvalues is equal to the number of items in the test. The eigenvalues can be ordered from first to last in terms of the amount of the common variance that each explains. Data are generally considered to be unidimensional if the second eigenvalue is less than or equal to 1.0. Previous research shows that the examination of the ratio of the first two (i.e., the two largest) eigenvalues can be useful in determining the existence of dominant factors. Specifically, where large ratios exist between the first and second eigenvalues, a single dominant factor can be said to exist. Although the definition of large in the present context is subjective, the results in Table 10-13 show that the eigenvalue of the first factor, in almost every case, is at least five times as large as the eigenvalue of the second factor.

As may be seen in Table 10-13, the ratios of the first two eigenvalues range from 4.05 to 7.74. The eigenvalues are proportional to the amount of common variance explained by each component, so these ratios indicate that the variance explained by the first component alone is approximately 4 to 8 times greater than the variance explained by the second component. The eigenvalue ratios ranged from 5.31 to 7.74 in Reading, from 4.90 to 7.44 in Mathematics, from 4.05 to 4.97 in Language Arts, from 4.38 to 5.26 in Social Studies, and from 4.23 to 5.70 in Science. These ratios suggest that the unidimensionality of each of the WKCE 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 WKCE assessments. The correlations among 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 Test Integrity: Erasure Analysis

The Fall 2012 WKCE test results were subjected to a special program that analyzed erasures on MC items. The focus of the analysis was on those cases where an incorrect answer choice was erased and replaced with the correct choice. These answer changes are referred to as wrong-to-right (WTR) answer changes. The WTR answer change analyses included a statistical test of the null hypothesis (H0) that the mean WTR answer changes for a test administration group constituted a random sample from the state distribution of WTR answer changes. The hypothesis was tested against the right-sided alternative (H1), that the mean number was too high to be explained by random sampling. Test administration groups (schools) for which the null hypothesis was rejected were flagged. To be flagged, the test administration group (school) erasure mean had to be 4 standard deviations above the state mean for erasures. The standard normal table shows that under random sampling the asymptotic probability of observing a sample mean more than four standard deviations above the population mean is around 0.0001, or one in ten thousand. A high rate of erasures can identify situations in which test integrity needs

Copyright © 2013 by the Wisconsin Department of Public Instruction

to be examined further. Separate erasure analyses were performed by grade and content area within schools. A summary erasure report was provided to DPI for evaluation.

10.4 Standardized Test Administration

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

The standardized WKCE 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.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Part 11: Summary Recommendations

Results and key findings of the Fall 2012 WKCE test administration are presented throughout the body of this report. Test difficulty in comparison to the student ability may warrant further attention in subsequent administrations as explained below.

Table 8-25 reveals that in most grades and content areas the WKCE students are not normally distributed along the raw score scale. This is indicated by the negative values of the skewness statistics, and this occurs because many students are answering most of the test items correctly. From a criterion-referenced testing perspective, the clustering of student scores on the high end of the raw score scale indicates that students on the whole are tending to demonstrate the knowledge, skills, and abilities specified in the Wisconsin Model Academic Standards.

From a measurement perspective, the WKCE may provide limited growth information for students in the highest performance level because large numbers of students are scoring in regions of the scale with the most amount of error. To measure these students with precision, more difficult items need to be added to the test. That is, for these students the test serves as a general measure of student skill; however, DPI would expect to see less fluctuation in score for individual students in this highest performing group from year to year if more difficult items were added to the assessment.

For these reasons, the DPI may wish to consider increasing the difficulty of the WKCE tests. This will likely provide more specific information about the higher ability students and allow the opportunity for the students to show growth. Because equating requires that tests maintain a similar level of difficulty from year to year, increasing the test rigor would likely require a cut score review and an examination regarding whether or not a new test scale should be set.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Fall 2012 WKCE Technical Report: 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. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association, Inc.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46*, 443–459.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41*, 687–699.

Burket, G. R. (1991). *PARDUX* [Computer program]. Unpublished.

Burket, G. R. (2000). *ITEMWIN* [Computer program]. Unpublished.

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/Thompson Learning.

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement, 25*, 291–312.

CTB/McGraw-Hill. (1997). *TerraNova* 1st Edition. Monterey, CA: Author.

CTB/McGraw-Hill. (2000). *TerraNova* 2nd Edition. Monterey, CA: Author.

CTB/McGraw-Hill. (1997–2001). WKCE. Monterey, CA: Author.

CTB/McGraw-Hill. (2005). *Wisconsin Knowledge and Concepts Examinations 2005 Standard Setting Technical Manual.* Monterey, CA: Author.

CTB/McGraw-Hill. (2006). *Wisconsin Knowledge and Concepts Examinations Fall 2005 WKCE-CRT Technical Manual.* Monterey, CA: Author.

CTB/McGraw-Hill. (2007). *Wisconsin Knowledge and Concepts Examinations Fall 2006 WKCE Technical Manual.* Monterey, CA: Author.

Copyright © 2013 by the Wisconsin Department of Public Instruction

CTB/McGraw-Hill. (2008). *Wisconsin Knowledge and Concepts Examinations Fall 2007 WKCE Technical Manual.* Monterey, CA: Author.

Fitzpatrick, A. R. (1991). *Status report on the results of preliminary analyses of dichotomous and multi-level items using the PARMATE program*. Monterey, CA: CTB/McGraw-Hill.

Fitzpatrick, A. R., & Julian, M.W. (1996). *Two studies comparing the parameter estimates produced by PARDUX and PARSCALE*. Unpublished manuscript.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*, 159–96.

Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty*. Princeton, NJ: Educational Testing Service, Research Report RR-85-43.

Holland, P. W., & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, California, April 1986.

Kolen, M., & Kim, D. (2004). Personal Correspondence.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and Practices.* New York: Springer-Verlag.

Linn, R. L. (1989). *Educational measurement* (3rd ed). New York: Macmillan.

Linn, R. L., & Harnisch, D. (1981). Interactions between item content and group membership in achievement test items. *Journal of Educational Measurement, 18*, 109–118.

Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement, 5*, 159–173.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179–197.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems* (71, 179–181). Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–748.

McCallin, R. C. (2006). Test Administration. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 625-652). Mahwah, NJ: Lawrence Erlbaum Associates.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*, 59–71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.

Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.

Patz, R. (1998). *Calculating Handscoring Reliability Coefficients.* Unpublished manuscript.

Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201–210.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion referenced tests: A decision theoretic formulation. *Journal of Educational Measurement, 11*, 263–268.

Swineford, F. (1956). Technical manual for users of test analysis. *Statistical Report 56–42.* Princeton, NJ: Educational Testing Service.

Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika, 47*, 175–186.

Thissen, D. (1990). *MULTILOG: Multiple categorical item analysis and test scoring*. Version 6 [Computer program]. Mooresville, IN: Scientific Software.

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

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

Yen, W. M. (1981). Using Simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*, 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*, 93–111.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 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*, 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*, 209–228.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*, 233–251.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-1
Target Reading Test Blueprint: Grades 3–8, 10*

	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
1	Determines meaning of words or phrases in context	12	0	11	0	11	0	11	0	11	0	11	0	7	0
1.1	Uses context clues to determine meaning of words or phrases	8		8		7		7		7		7		5	
1.2	Uses knowledge of word structure to determine meaning of words	2		2		2		2		2		2		1	
1.3	Uses word reference materials to determine meaning of words and phrases	2		1		2		2		2		2		1	
2	Understands Text	17	0	17	0	15	0	14	0	14	0	14	0	7	0
2.1	Demonstrates understanding of literal meaning by identifying stated information in literary text	7		7		6		6		6		6		2	
2.2	Demonstrates understanding of literal meaning by identifying stated information in informational text	7		8		6		6		6		6		3	
2.3	Demonstrates understanding of explicitly stated sequence of events in literary and informational text	3		2		3		2		2		2		2	
3	Analyzes Text	21	1	21	1	20	1	18	1	18	1	18	1	22	1
3.1	Analyzes literary text	10		10		9		8		8		8		7	
3.2	Analyzes informational text.	8		8		8		6		6		6		7	
33	Analyzes author's use of language in literary and informational text.	3		3		3		4		4		4		8	
4	Evaluates and Extends Text	4	1	5	1	8	1	11	1	11	1	11	1	14	1
4.1	Evaluates and extends literary text	2		2		3		3		3		3		4	
4.2	Evaluates and extends informational text	1		2		3		5		5		5		5	
4.3	Evaluates and extends author's use of language in literary and informational text	1		1		2		3		3		3		5	
	Number of Items	54	2	54	2	54	2	54	2	54	2	54	2	50	2
	Total Score Points for Test	60		60		60		60		60		60		56	

*Note: The CR items do not report out to any single subskill.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-2
Target Mathematics Test Blueprint: Grades 3–8, 10*

	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
A	Mathematical Processes	3	3	3	3	3	3	3	3	3	3	3	3	7	1
Aa	Reasoning														
Ab	Communication														
Ac	Connections														
Ad	Representation														
Ae	Problem Solving														
B	Number Operations and Relationships	11	1	11	0	11	0	12	0	12	0	7	0	7	0
Ba	Number Concepts	6		5		5		6		6		4		4	
Bb	Number Computation	5		6		6		6		6		3		3	
C	Geometry	9	1	8	1	9	1	9	1	10	2	8	1	8	1
Ca	Describing Figures	4		3		3		2		3		2		4	
Cb	Spatial Relationships and Transformations	4		4		4		4		4		4		2	
Cc	Coordinate System	1		1		2		3		3		2		2	
D	Measurement	8	0	8	1	9	1	9	1	9	0	11	1	9	1
Da	Measurable Attributes	3		3		4		2		3		2		1	
Db	Direct Measurement	4		4		3		3		3		3		2	
Dc	Indirect Measurement	1		1		2		4		3		6		6	
E	Statistics and Probability	7	1	7	1	9	1	8	1	8	1	8	1	9	0
Ea	Data Analysis and Statistics	5		4		6		5		5		5		4	
Eb	Probability	2		3		3		3		3		3		5	
F	Algebraic Relationships	8	1	9	1	10	1	10	1	9	1	14	1	10	1
Fa	Patterns, Relations, and Functions	4		5		5		5		2		7		5	
Fb	Expressions, Equations, and Inequalities	2		2		3		2		3		6		4	
Fc	Properties	2		2		2		3		4		1		1	
	Number of Items	46	4	46	4	51	4	51	4	51	4	51	4	50	4
	Total Score Points for Test	57		57		62		62		62		62		58	

*Note: The CR items do not report out to any single subskill. The items in “A: Mathematical Processes” also do not report out to any single subskill.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-3
Target Language Arts Test Blueprint: Grades 4, 8, 10

Content Standard		Grade 4		Grade 8		Grade 10	
		MC	Prompt	MC	Prompt	MC	Prompt
B	Writing	19	1	18	1	15	1
D	Language	5	0	6	0	9	0
F	Research and Inquiry	6	0	6	0	6	0
	Number of Items	30	1	30	1	30	1
	Total Number of Points	30	9	30	9	30	9

Table 2-4
Target Social Studies Test Blueprint: Grades 4, 8, 10

Content Standard		Grade 4	Grade 8	Grade 10
A	Geography	9	10	10
B	History	8	13	12
C	Political Science	7	6	12
D	Economics	7	6	8
E	Behavioral Science	7	5	8
	Total Number of MC Items	38	40	50

Table 2-5
Target Science Test Blueprint: Grades 4, 8, 10

Content Standard		Grade 4	Grade 8	Grade 10
A	Science Connections	4	4	5
B	Nature of Science	4	3	5
C	Science Inquiry	7	8	10
D	Physical Science	6	6	7
E	Earth and Space	6	6	6
F	Life and Environment	6	6	7
G	Science Applications	4	4	5
H	Personal/Social Perspectives	3	3	5
	Total Number of MC Items	40	40	50

*Note: Standard A, Science Connections, and Standard B, Nature of Science, are combined to form a reporting category; Standard G, Science Applications, and Standard H, Personal/Social Perspectives, are combined to form a reporting category.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-6
Actual Reading Test Blueprint: Grades 3–8, 10*

	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade10	
		MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
1	Determines meaning of words or phrases in context	12	0	11	0	11	0	11	0	11	0	11	0	7	0
1.1	Uses context clues to determine meaning of words or phrases	6		7		8		5		7		8		5	
1.2	Uses knowledge of word structure to determine meaning of words	4		2		1		4		2		0		0	
1.3	Uses word reference materials to determine meaning of words and phrases	2		2		2		2		2		3		2	
2	Understands Text	17	0	17	0	15	0	14	0	14	0	14	0	7	0
2.1	Demonstrates understanding of literal meaning by identifying stated information in literary text	6		7		6		6		5		3		1	
2.2	Demonstrates understanding of literal meaning by identifying stated information in informational text	10		6		7		5		5		8		6	
2.3	Demonstrates understanding of explicitly stated sequence of events in literary and informational text	1		4		2		3		4		3		0	
3	Analyzes Text	21	1	21	1	20	1	18	1	18	1	18	1	22	1
3.1	Analyzes literary text	8		9		5		8		6		9		8	
3.2	Analyzes informational text.	8		7		11		6		8		6		9	
3.3	Analyzes author's use of language in literary and informational text.	5		5		4		4		4		3		5	1
4	Evaluates and Extends Text	4	1	5	1	8	1	11	1	11	1	11	1	14	1
4.1	Evaluates and extends literary text	1		1		2		5		2		4		5	
4.2	Evaluates and extends informational text	1		2		4		1		5		4		7	1
4.3	Evaluates and extends author's use of language in literary and informational text	2		2		2		5		4		3		2	
	Number of Items	54	2	54	2	54	2	54	2	54	2	54	2	50	2
	Total Score Points for Test	60		60		60		60		60		60		56	

* Note: The CR items do not report out to any single subskill.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-7
Actual Mathematics Test Blueprint: Grades 3–8, 10*

	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
A	Mathematical Processes	3	3	3	3	3	3	3	3	3	3	3	3	7	1
Aa	Reasoning														
Ab	Communication														
Ac	Connections														
Ad	Representation														
Ae	Problem Solving														
B	Number Operations and Relationships	11	1	11	0	11	0	12	0	12	0	7	0	7	0
Ba	Number Concepts	6		5		5		6		6		4		4	
Bb	Number Computation	5		6		6		6		6		3		3	
C	Geometry	9	1	8	1	9	1	9	1	10	2	8	1	8	1
Ca	Describing Figures	4		3		3		2		3		2		4	
Cb	Spatial Relationships and Transformations	4		4		4		4		4		4		2	
Cc	Coordinate System	1		1		2		3		3		2		2	
D	Measurement	8	0	8	1	9	1	9	1	9	0	11	1	9	1
Da	Measurable Attributes	3		3		4		2		3		2		1	
Db	Direct Measurement	4		4		3		3		3		3		2	
Dc	Indirect Measurement	1		1		2		4		3		6		6	
E	Statistics and Probability	7	1	7	1	9	1	8	1	8	1	8	1	9	0
Ea	Data Analysis and Statistics	5		4		6		5		5		5		4	
Eb	Probability	2		3		3		3		3		3		5	
F	Algebraic Relationships	8	1	9	1	10	1	10	1	9	1	14	1	10	1
Fa	Patterns, Relations, and Functions	4		5		5		5		2		7		5	
Fb	Expressions, Equations, and Inequalities	2		2		3		2		3		6		4	
Fc	Properties	2		2		2		3		4		1		1	
	Number of Items	46	4	46	4	51	4	51	4	51	4	51	4	50	4
	Total Score Points for Test	57		57		62		62		62		62		58	

*The items in "A: Mathematical Processes" do not report out to any single subskill. Note also that some CR items in Grades 3–8 report out to more than one standard. The total number of CR items is 4 per grade even though some items are associated with more than one standard.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-8
Actual Language Arts Test Blueprint: Grades 4, 8, 10

Content Standard		Grade 4		Grade 8		Grade 10	
		MC	**Prompt**	**MC**	**Prompt**	**MC**	**Prompt**
B	Writing	19	1	18	1	15	1
D	Language	5	0	6	0	9	0
F	Research and Inquiry	6	0	6	0	6	0
	Total Number of Items	30	1	30	1	30	1
	Total Number of Points	30	9	30	9	30	9

Table 2-9
Actual Social Studies Test Blueprint: Grades 4, 8, 10

Content Standard		Grade 4	Grade 8	Grade 10
A	Geography	9	10	10
B	History	8	11	12
C	Political Science	7	6	12
D	Economics	7	8	8
E	Behavioral Science	7	5	8
	Total Number of MC Items	38	40	50

Table 2-10
Actual Science Test Blueprint: Grades 4, 8, 10

Content Standard*		Grade 4	Grade 8	Grade 10
A	Science Connections	4	3	5
B	Nature of Science	4	4	5
C	Science Inquiry	7	8	9
D	Physical Science	6	6	8
E	Earth and Space	6	6	6
F	Life and Environment	6	6	6
G	Science Applications	4	4	5
H	Personal/Social Perspectives	3	3	6
	Total Number of MC Items	40	40	50

*Note: Standard A, Science Connections, and Standard B, Nature of Science, are combined to form a reporting category; Standard G, Science Applications, and Standard H, Personal/Social Perspectives, are combined to form a reporting category.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 2-11
Item Development Each Year and Total to Date*

	MC items for 2004	CR items for 2004	MC items for 2005	CR items for 2005	MC items for 2006	CR items for 2006	MC items for 2007	CR items for 2007	MC items for 2008	CR items for 2008	MC items for 2009	CR items for 2009	Total MC to date	Total CR to date
Grade 3														
Reading	411	52	23	2	30	4	40	3	52	4	51	7	607	72
Math	317	36	33	14	18	2	30	4	28	11	52	6	478	73
Total	728	88	56	16	48	6	70	7	80	15	103	13	1085	145
Grade 4														
Reading	380	56	32	3	34	3	25	4	54	4	52	7	577	77
Math	265	35	45	9	29	1	26	4	28	13	54	11	447	73
Language Arts	0	0	0	10	0	0	0	0	0	0	0	0	0	10
Science	0	0	0	0	123	34	0	0	0	0	0	0	123	34
Total	645	91	77	22	186	38	51	8	82	17	106	18	1147	194
Grade 5														
Reading	433	59	36	6	29	5	29	7	44	4	52	7	623	88
Math	305	49	38	11	26	3	30	5	28	13	53	8	480	89
Total	738	108	74	17	55	8	59	12	72	17	105	15	1103	177
Grade 6														
Reading	511	56	32	5	42	5	37	6	46	5	50	7	718	84
Math	310	41	53	16	7	2	28	4	30	12	41	8	469	83
Total	821	97	85	21	49	7	65	10	76	17	91	15	1187	167
Grade 7														
Reading	359	44	35	4	38	4	25	5	50	4	50	7	557	68
Math	305	34	32	23	20	0	28	4	31	10	40	6	456	77
Total	664	78	67	27	58	4	53	9	81	14	90	13	1013	145
Grade 8														
Reading	365	44	30	4	34	4	25	4	44	4	50	7	548	67
Math	289	51	47	25	20	2	28	4	32	17	40	8	456	107
Language Arts	0	0	0	10	0	0	0	0	0	0	0	0	0	10
Science	0	0	0	0	125	34	0	0	0	0	0	0	125	34
Total	654	95	77	39	179	40	53	8	76	21	90	15	1129	218
Grade 10														
Reading	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Math	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Language Arts	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Science	0	0	0	0	18	8	0	0	0	0	0	0	18	8
Total	0	0	0	0	18	8	0	0	0	0	0	0	18	8
TOTALS														
Reading	2,459	311	188	24	207	25	181	29	290	25	305	42	3,630	456
Mathematics	1,791	246	248	98	120	10	170	25	177	76	280	47	2,786	502
Language Arts	0	0	0	20	0	0	0	0	0	0	0	0	0	20
Science	0	0	0	0	266	76	0	0	0	0	0	0	266	76
Grand Total	4,250	557	436	142	593	111	351	54	467	101	585	89	6,682	1,054

*Note: This table includes 17 Fall 2009 Math items rejected by DPI prior to the Content and Bias Review.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 3-1
Fall 2012 Test Configuration

Content	Grade	No. of OP MC Items	No. of OP CR Items					Total Score Point	Total OP (MC + CR) Items
			1 point	2 point	3 point	4 point	6 point		
Reading	3	54			1*			57	55
	4	54			2			60	56
	5	54			2			60	56
	6	54			2			60	56
	7	54			2			60	56
	8	54			2			60	56
	10	50			2			56	52
Mathematics**	3	46	3	4				57	53
	4	46	3	4				57	53
	5	51	3	4				62	58
	6	51	3	4				62	58
	7	51	3	4				62	58
	8	51	3	4				62	58
	10	50	0	4				58	54
Language Arts***	4	30						30	30
	8	29*						29	29
	10	30			1		1	39	32
Social Studies	4	38						38	38
	8	40						40	40
	10	49*						49	49
Science	4	39*						39	39
	8	40						40	40
	10	49*						50	50

* An item was dropped in Reading grade 3, Language Arts grade 8, Social Studies grade 10, and Science grade 4. See Part 8 for information.
** Some Mathematics items include two parts, Part A and Part B. Each part is counted as an item above.
*** For Language Arts grade 10, the two CR items are from the grade 10 Writing prompt. The Writing prompt in grade 10 is part of the scale score for Language Arts in grade 10.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 3-2
Unique Items Field Tested Each Year and Total to Date

	MC 2004	CR 2004	MC 2005	CR 2005	MC 2006	CR 2006	MC 2007	CR 2007	MC 2008	CR 2008	MC 2009	CR 2009	Total MC to Date	Total CR to Date
Grade 3														
RD	242	12	24	2	27	2	40	4	40	4	40	4	413	28
MA	252	24	15	2	32	4	34	5	31	8	40	4	404	47
Total	494	36	39	4	59	6	74	9	71	12	80	8	817	75
Grade 4														
RD	294	12	24	2	32	3	40	4	40	4	40	4	470	29
MA	231	29	15	2	32	4	34	4	28	8	40	4	380	51
LA	0	0	0	6	0	0	0	0	0	0	0	0	0	6
SS	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	40	0	0	0	0	0	0	0	40	0
Total	525	41	39	10	104	7	74	8	68	12	80	8	890	86
Grade 5														
RD	235	14	24	2	28	2	29	6	40	4	40	4	396	32
MA	257	34	15	2	32	4	40	4	34	6	40	4	418	54
Total	492	48	39	4	60	6	69	10	74	10	80	8	814	86
Grade 6														
RD	259	14	24	1	33	3	35	5	40	4	40	4	431	31
MA	252	33	15	2	32	4	32	4	30	5	32	4	393	52
Total	511	47	39	3	65	7	67	9	70	9	72	8	824	83
Grade 7														
RD	259	14	24	1	17	2	35	4	40	4	40	4	415	29
MA	243	33	15	2	32	4	32	3	33	4	32	4	387	50
Total	502	47	39	3	49	6	67	7	73	8	72	8	802	79
Grade 8														
RD	274	14	24	1	33	4	32	5	40	4	40	4	443	32
MA	234	33	15	2	40	4	32	4	32	5	32	4	385	52
LA	0	0	0	6	0	0	0	0	0	0	0	0	0	6
SS	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	40	0	0	0	0	0	0	0	40	0
Total	508	47	39	9	113	8	64	9	72	9	72	8	868	90
Grade 10														
RD	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SS	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	10	0	0	0	0	0	0	0	10	0
Total	0	0	0	0	10	0	0	0	0	0	0	0	10	0
TOTALS														
Grand Totals	3,032	266	234	33	460	40	415	52	428	60	456	48	5,025	499

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used a Scribe	3	60192	Reading	1018	1.69
		60376	Mathematics	1204	1.99
	4	59664	Reading	836	1.40
		59792	Mathematics	917	1.53
		59736	Science	395	0.66
		59680	Language Arts	349	0.58
		59832	Writing	454	0.76
		59680	Social Studies	394	0.66
	5	59256	Reading	637	1.08
		59296	Mathematics	673	1.14
	6	60376	Reading	494	0.82
		60432	Mathematics	503	0.83
	7	61224	Reading	327	0.53
		61296	Mathematics	322	0.53
	8	60704	Reading	317	0.52
		60752	Mathematics	293	0.48
		60656	Science	171	0.28
		60504	Language Arts	180	0.30
		60776	Writing	324	0.53
		60624	Social Studies	168	0.28
	10	62784	Reading	139	0.22
		62696	Mathematics	129	0.21
		62576	Science	84	0.13
		62472	Language Arts	79	0.13
		63312	Writing	144	0.23
		62512	Social Studies	83	0.13

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Provided Extra Time	3	60192	Reading	7073	11.75
		60376	Mathematics	7121	11.79
	4	59664	Reading	7308	12.25
		59792	Mathematics	7412	12.40
		59736	Science	7191	12.04
		59680	Language Arts	7226	12.11
		59832	Writing	7128	11.91
		59680	Social Studies	7147	11.98
	5	59256	Reading	7005	11.82
		59296	Mathematics	7056	11.90
	6	60376	Reading	6742	11.17
		60432	Mathematics	6859	11.35
	7	61224	Reading	6617	10.81
		61296	Mathematics	6680	10.90
	8	60704	Reading	6473	10.66
		60752	Mathematics	6588	10.84
		60656	Science	6330	10.44
		60504	Language Arts	6376	10.54
		60776	Writing	6297	10.36
		60624	Social Studies	6349	10.47
	10	62784	Reading	4235	6.74
		62696	Mathematics	4314	6.88
		62576	Science	4100	6.55
		62472	Language Arts	4172	6.68
		63312	Writing	4147	6.55
		62512	Social Studies	4053	6.48

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Read Test Questions and Content to Student	3	60192	Reading	0	0.00
		60376	Mathematics	6268	10.38
	4	59664	Reading	0	0.00
		59792	Mathematics	6426	10.75
		59736	Science	6257	10.47
		59680	Language Arts	5798	9.72
		59832	Writing	5895	9.85
		59680	Social Studies	6181	10.36
	5	59256	Reading	0	0.00
		59296	Mathematics	5774	9.74
	6	60376	Reading	0	0.00
		60432	Mathematics	5570	9.22
	7	61224	Reading	0	0.00
		61296	Mathematics	5048	8.24
	8	60704	Reading	0	0.00
		60752	Mathematics	4724	7.78
		60656	Science	4639	7.65
		60504	Language Arts	4253	7.03
		60776	Writing	4470	7.35
		60624	Social Studies	4606	7.60
	10	62784	Reading	0	0.00
		62696	Mathematics	2647	4.22
		62576	Science	2760	4.41
		62472	Language Arts	2628	4.21
		63312	Writing	2786	4.40
		62512	Social Studies	2719	4.35

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used DPI-Provided Test Translation	3	60192	Reading	0	0.00
		60376	Mathematics	1202	1.99
	4	59664	Reading	0	0.00
		59792	Mathematics	976	1.63
		59736	Science	934	1.56
		59680	Language Arts	0	0.00
		59832	Writing	747	1.25
		59680	Social Studies	917	1.54
	5	59256	Reading	0	0.00
		59296	Mathematics	632	1.07
	6	60376	Reading	0	0.00
		60432	Mathematics	213	0.35
	7	61224	Reading	0	0.00
		61296	Mathematics	224	0.37
	8	60704	Reading	0	0.00
		60752	Mathematics	180	0.30
		60656	Science	170	0.28
		60504	Language Arts	0	0.00
		60776	Writing	136	0.22
		60624	Social Studies	170	0.28
	10	62784	Reading	0	0.00
		62696	Mathematics	78	0.12
		62576	Science	71	0.11
		62472	Language Arts	0	0.00
		63312	Writing	54	0.09
		62512	Social Studies	68	0.11

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used Locally Provided Test Translation	3	60192	Reading	0	0.00
		60376	Mathematics	0	0.00
	4	59664	Reading	0	0.00
		59792	Mathematics	146	0.24
		59736	Science	143	0.24
		59680	Language Arts	0	0.00
		59832	Writing	102	0.17
		59680	Social Studies	138	0.23
	5	59256	Reading	0	0.00
		59296	Mathematics	0	0.00
	6	60376	Reading	0	0.00
		60432	Mathematics	0	0.00
	7	61224	Reading	0	0.00
		61296	Mathematics	0	0.00
	8	60704	Reading	0	0.00
		60752	Mathematics	49	0.08
		60656	Science	35	0.06
		60504	Language Arts	0	0.00
		60776	Writing	26	0.04
		60624	Social Studies	34	0.06
	10	62784	Reading	0	0.00
		62696	Mathematics	35	0.06
		62576	Science	47	0.08
		62472	Language Arts	0	0.00
		63312	Writing	29	0.05
		62512	Social Studies	47	0.08

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used DPI-Provided Glossary of Terms	3	60192	Reading	0	0.00
		60376	Mathematics	0	0.00
	4	59664	Reading	0	0.00
		59792	Mathematics	313	0.52
		59736	Science	287	0.48
		59680	Language Arts	0	0.00
		59832	Writing	0	0.00
		59680	Social Studies	276	0.46
	5	59256	Reading	0	0.00
		59296	Mathematics	0	0.00
	6	60376	Reading	0	0.00
		60432	Mathematics	0	0.00
	7	61224	Reading	0	0.00
		61296	Mathematics	0	0.00
	8	60704	Reading	0	0.00
		60752	Mathematics	113	0.19
		60656	Science	118	0.19
		60504	Language Arts	0	0.00
		60776	Writing	0	0.00
		60624	Social Studies	102	0.17
	10	62784	Reading	0	0.00
		62696	Mathematics	44	0.07
		62576	Science	48	0.08
		62472	Language Arts	0	0.00
		63312	Writing	0	0.00
		62512	Social Studies	50	0.08

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used Text Talker	3	60192	Reading	0	0.00
		60376	Mathematics	8	0.01
	4	59664	Reading	0	0.00
		59792	Mathematics	11	0.02
		59736	Science	7	0.01
		59680	Language Arts	7	0.01
		59832	Writing	6	0.01
		59680	Social Studies	7	0.01
	5	59256	Reading	0	0.00
		59296	Mathematics	7	0.01
	6	60376	Reading	0	0.00
		60432	Mathematics	2	0.00
	7	61224	Reading	0	0.00
		61296	Mathematics	4	0.01
	8	60704	Reading	0	0.00
		60752	Mathematics	4	0.01
		60656	Science	1	0.00
		60504	Language Arts	1	0.00
		60776	Writing	3	0.00
		60624	Social Studies	3	0.00
	10	62784	Reading	0	0.00
		62696	Mathematics	26	0.04
		62576	Science	19	0.03
		62472	Language Arts	18	0.03
		63312	Writing	27	0.04
		62512	Social Studies	19	0.03

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Signed Test Questions and Content to Student	3	60192	Reading	0	0.00
		60376	Mathematics	22	0.04
	4	59664	Reading	0	0.00
		59792	Mathematics	29	0.05
		59736	Science	29	0.05
		59680	Language Arts	29	0.05
		59832	Writing	25	0.04
		59680	Social Studies	28	0.05
	5	59256	Reading	0	0.00
		59296	Mathematics	22	0.04
	6	60376	Reading	0	0.00
		60432	Mathematics	19	0.03
	7	61224	Reading	0	0.00
		61296	Mathematics	18	0.03
	8	60704	Reading	0	0.00
		60752	Mathematics	10	0.02
		60656	Science	8	0.01
		60504	Language Arts	7	0.01
		60776	Writing	7	0.01
		60624	Social Studies	8	0.01
	10	62784	Reading	0	0.00
		62696	Mathematics	15	0.02
		62576	Science	13	0.02
		62472	Language Arts	15	0.02
		63312	Writing	13	0.02
		62512	Social Studies	14	0.02

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used Another DPI-Approved Accommodation	3	60192	Reading	1871	3.11
		60376	Mathematics	1980	3.28
	4	59664	Reading	1956	3.28
		59792	Mathematics	2067	3.46
		59736	Science	1912	3.20
		59680	Language Arts	1877	3.15
		59832	Writing	1871	3.13
		59680	Social Studies	1908	3.20
	5	59256	Reading	2024	3.42
		59296	Mathematics	2155	3.63
	6	60376	Reading	1928	3.19
		60432	Mathematics	1979	3.27
	7	61224	Reading	2003	3.27
		61296	Mathematics	2062	3.36
	8	60704	Reading	1612	2.66
		60752	Mathematics	1605	2.64
		60656	Science	1554	2.56
		60504	Language Arts	1547	2.56
		60776	Writing	1557	2.56
		60624	Social Studies	1551	2.56
	10	62784	Reading	539	0.86
		62696	Mathematics	555	0.89
		62576	Science	526	0.84
		62472	Language Arts	520	0.83
		63312	Writing	542	0.86
		62512	Social Studies	526	0.84

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used DPI-Provided Braille Test	3	60192	Reading	0	0.00
		60376	Mathematics	0	0.00
	4	59664	Reading	9	0.02
		59792	Mathematics	9	0.02
		59736	Science	9	0.02
		59680	Language Arts	9	0.02
		59832	Writing	9	0.02
		59680	Social Studies	9	0.02
	5	59256	Reading	13	0.02
		59296	Mathematics	13	0.02
	6	60376	Reading	0	0.00
		60432	Mathematics	0	0.00
	7	61224	Reading	0	0.00
		61296	Mathematics	0	0.00
	8	60704	Reading	3	0.00
		60752	Mathematics	3	0.00
		60656	Science	3	0.00
		60504	Language Arts	3	0.01
		60776	Writing	3	0.00
		60624	Social Studies	3	0.00
	10	62784	Reading	0	0.00
		62696	Mathematics	0	0.00
		62576	Science	0	0.00
		62472	Language Arts	0	0.00
		63312	Writing	0	0.00
		62512	Social Studies	0	0.00

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 4-1
Test Accommodations Cont'd

Accommodation	Grade	N Count	Content Area	Number of Students	Percent of Total
Used a Non-Allowed Accommodation	3	60192	Reading	0	0.00
		60376	Mathematics	0	0.00
	4	59664	Reading	0	0.00
		59792	Mathematics	0	0.00
		59736	Science	0	0.00
		59680	Language Arts	0	0.00
		59832	Writing	0	0.00
		59680	Social Studies	0	0.00
	5	59256	Reading	0	0.00
		59296	Mathematics	0	0.00
	6	60376	Reading	0	0.00
		60432	Mathematics	0	0.00
	7	61224	Reading	0	0.00
		61296	Mathematics	0	0.00
	8	60704	Reading	0	0.00
		60752	Mathematics	0	0.00
		60656	Science	0	0.00
		60504	Language Arts	0	0.00
		60776	Writing	0	0.00
		60624	Social Studies	0	0.00
	10	62784	Reading	0	0.00
		62696	Mathematics	0	0.00
		62576	Science	0	0.00
		62472	Language Arts	0	0.00
		63312	Writing	0	0.00
		62512	Social Studies	0	0.00

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-1
Reading Rubric, Grades 3–8 and 10

Reading items at all grade levels were scored using item-specific scoring guides that are based on a generic, 0–3 holistic rubric.

3 points

- The response demonstrates *thorough understanding* of the reading concept embodied in the task.
- The response is *accurate, complete, insightful, and fulfills all the requirements* of the task.
- Necessary support and/or examples are included.
- Information is clearly *text-based*.

2 points

- The response demonstrates *partial understanding* of the reading concept embodied in the task.
- The response is *accurate* and *fulfills most of the requirements* of the task.
- Necessary support and/or examples may not be complete or clearly text-based.

1 point

- The response demonstrates *an incomplete understanding* of the reading concept embodied in the task.
- The response provides *some information that is text-based*, but does not fulfill the requirements of the task.
- Information provided is *too general* or *too simplistic*.
- Necessary support and/or examples may be incomplete or omitted.

0 points

- The response demonstrates *no understanding* of the reading concept embodied in the task.
- The response is *inaccurate, confused*, or *irrelevant.*
- The student has written a response but *failed to respond to the task.*

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-2
Mathematics Rubric, Grades 3–8 and 10

Generic Rubric for Mathematics for 2-point Constructed Response Items

2 points	The student demonstrates a thorough understanding of the mathematical concepts and/or procedures represented in the problem. The student states appropriate mathematical responses and/or uses procedures and/or concepts to explain or justify a response. The student provides clear and complete explanations and interpretations containing words, calculations, or symbols, when specified in the item stem. The response may contain minor flaws that do not detract from the demonstration of a thorough understanding of the problem.
1 point	The student demonstrates only a partial understanding of the mathematical concepts and/or procedures represented in the problem. The response lacks an appropriate mathematical response or reflects the lack of an essential understanding of the underlying mathematical concepts used in the item. The response contains errors related to the misinterpretation of important aspects of the problem, misuse of mathematical procedures and/or concepts, or misinterpretation of results.
0 points	The student provides completely incorrect responses, explanations, or justifications, or ones that cannot be interpreted, for all responses required in the item.

Generic Rubric for Mathematics for 3-point Constructed Response Items

Mathematics 3-point constructed response items have two parts. Part A is scored as correct/incorrect. Part B is scored using the 2-point holistic rubric below.

2 points	The student demonstrates a thorough understanding of the mathematical concepts and/or procedures represented in the problem. The student uses appropriate mathematical procedures and/or concepts to explain or justify the response to Step A, and provides clear and complete explanations and interpretations containing words, calculations, or symbols, unless otherwise specified in the item stem. The response may contain minor flaws that do not detract from the demonstration of a thorough understanding of the problem.
1 point	The student demonstrates only a partial understanding of the mathematical concepts and/or procedures represented in the problem. The response lacks an essential understanding of the underlying mathematical concepts used to provide the response to Step A. The response contains errors related to the misinterpretation of important aspects of the problem, misuse of mathematical procedures and/or concepts, or misinterpretation of results.
0 points	The student provides a completely incorrect explanation or justification, or one that cannot be interpreted.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-3
Writing Rubric, Conventions of Written English, Grade 4

3 points Advanced Control

The response demonstrates advanced control of a wide range of conventions identified in the 4th grade Wisconsin Model Academic Standards in English Language Arts:

- Uses parts of speech effectively, including nouns, pronouns, and adjectives
- Uses adverbials effectively, including words and phrases
- Employs principles of agreement related to number, gender, and case
- Capitalizes proper nouns, titles, and initial words of sentences
- Uses punctuation marks and conjunctions, as appropriate, to separate sentences and connect independent clauses
- Uses commas correctly to punctuate appositives and lists
- Spells correctly in general and even on difficult words
- Uses word order and punctuation marks to distinguish statements, questions, exclamations, and commands
- Makes errors that are infrequent and minor

2 points Proficient Control

The response demonstrates proficient control of the essential conventions identified in the 4th grade Wisconsin Model Academic Standards in English Language Arts:

- Generally controls grammar and usage (principles of agreement, noun and verb forms, superlative and comparative forms)
- Capitalizes proper nouns, titles, and initial words of sentences
- Uses end-stop punctuation correctly most of the time; internal punctuation (commas, apostrophes) is sometimes missing or wrong.
- Generally uses correct spelling with common words but more difficult words are problematic
- Makes errors typical of those commonly found in a rough draft; errors do not significantly distract the reader

1 point Minimal Control

The response demonstrates minimal control of the essential conventions identified in the 4th grade Wisconsin Model Academic Standards in English Language Arts:

- Contains numerous serious end-stop punctuation errors, resulting in fragments, comma splices, run-ons
- Shows poor control of subject/verb agreement, possessive forms, capitalization, superlatives, and comparatives
- Spelling errors are frequent, even on common words
- Makes errors that are frequent, varied, and distracting

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-4
Writing Rubric, Composing, Grade 4

Wisconsin Writing Grade 4 Rubric 6-Point Scoring Guide					
Elements of Rubric	**Purpose & Focus**	**Organization & Coherence**	**Development of Content**	**Sentence Fluency**	**Word Choice**
Element Description	Consistently focuses on the topic and maintains a unified purpose Demonstrates understanding of the requirements of the assigned task	Uses a logical plan of development with an effective beginning, middle, and end Keeps relationships among ideas clear Paragraphs logically and uses appropriate transitional devices	Expands and supports main ideas with specific details, examples, and/or reasons that are 1) clearly related to the topic and purpose, and 2) effective for audience	Uses varied sentence structures, creating a fluent, effective, and readable style	Controls word choice with respect to both denotation and connotation Demonstrates attention to context (audience, purpose, situation, tone) Evidences some control over figurative language for rhetorical effect (e.g. metaphors, similes)
Positive Descriptors	Focused, unified, controlled, relevant	Well organized, integrated, smooth, controlled, coherent	Thorough, specific, well-developed, well-supported, well-illustrated, insightful, convincing	Fluid, varied, controlled, effective	Vivid, precise, concrete, concise
Negative Descriptors	Rambling, loosely related, redundant, irrelevant, lacks purpose	Disorganized, hard to follow, mechanical, illogical shifts, incoherent	Vague, general, simplistic, superficial, incomplete, illogical, inadequately supported, lacks illustration	Choppy, simple, repetitive, garbled, ineffective, awkward	Awkward, imprecise, vague wordy, repetitive

Rubric Holistic Scoring Scale						
Scores	**6**	**5**	**4**	**3**	**2**	**1**
Description	Exemplary control of the domain	Advanced control of the domain	Proficient control of the domain	Adequate control of the domain	Basic control of the domain	Minimal control of the domain

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-5
Writing Rubric, Conventions of Written English, Grade 8

3 points Advanced Control

The response demonstrates advanced control of a wide range of conventions identified in the 8th grade Wisconsin Model Academic Standards in English Language Arts:

- Uses words, phrases, and clauses effectively, including coordinate and subordinate conjunctions, relative pronouns, and comparative adjectives
- Uses correct tenses to indicate the relative order of events
- Employs principles of agreement, including subject-verb, pronoun-noun, and preposition-pronoun
- Punctuates compound, complex, and compound-complex sentences correctly
- Employs the conventions of capitalization
- Spells frequently used words correctly and uses effective strategies for spelling unfamiliar words
- Makes errors that are infrequent and minor

2 points Proficient Control

The response demonstrates proficient control of the conventions identified in the 8th grade Wisconsin Model Academic Standards in English Language Arts:

- Generally controls grammar and usage (principles of agreement, noun and verb forms, pronoun reference, superlative, and comparative forms)
- Generally uses phrases, dependent, and independent clauses clearly and correctly
- Capitalizes most words correctly; control over more sophisticated capitalization skills may be spotty
- Uses end-stop punctuation correctly most of the time; internal punctuation (commas, apostrophes, semicolons) is sometimes missing or wrong
- Generally uses correct spelling with grade-level words and reasonable phonetic approaches to more difficult words
- Makes errors typical of those commonly found in a rough draft; errors do not seriously distract the reader

1 point Minimal Control

The response demonstrates minimal control of the conventions identified in the 8th grade Wisconsin Model Academic Standards in English Language Arts:

- Contains numerous serious end-stop or internal punctuation errors, resulting in fragments, comma splices, run-ons
- Shows poor control of grammar and usage (principles of agreement; verb and/or noun forms including possessives; pronoun reference; superlative and comparative forms; appropriate use of phrases/independent, dependent clauses, capitalization)
- Frequently misspells words, even those on grade-level
- Makes errors that are frequent, varied, and distracting

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-6
Writing Rubric, Composing, Grade 8

Wisconsin Writing Grade 8 Rubric 6-Point Scoring Guide					
Elements of Rubric	**Purpose & Focus**	**Organization & Coherence**	**Development of Content**	**Sentence Fluency**	**Word Choice**
Element Description	Clearly presents and maintains a unified purpose, focus, and/or thesis Demonstrates understanding of the requirements of the assigned task	Frames the discussion with an effective introduction and conclusion Creates a logical structure of development for the topic, thesis, and purpose Uses transitional strategies (from idea to idea, paragraph to paragraph, and sentence to sentence)	Demonstrates ***quality*** of invented content (e.g. of explanations, arguments, rationale, ideas, details, examples, illustrations) Demonstrates ***thoroughness*** in the elaboration of content	Demonstrates use of varied syntactic structures including simple, compound, complex, and compound/complex sentences Evidences some control over stylistic effects (e.g. variety, readability**)**	Controls word choice with respect to both denotation and connotation Demonstrates attention to context (audience, purpose, situation, tone) Evidences some control over figurative language for rhetorical effect (e.g. similes, metaphors, personification)
Positive Descriptors	Focused, unified, controlled, relevant	Well organized, integrated, smooth, controlled, coherent	Quality: clear, convincing, accurate, effective, well-reasoned, insightful Thoroughness: specific, well-developed, well-supported, well-illustrated	Fluid, varied, controlled, effective	Apt, discriminating, vivid, precise, concrete, concise
Negative Descriptors	Rambling, loosely related, redundant, irrelevant, lacks purpose	Disorganized, hard to follow, mechanical, illogical shifts, incoherent	Quality: vague, imprecise, inaccurate, simplistic, poorly reasoned, superficial Thoroughness: incomplete, general, inadequately developed, inadequately supported, lacks illustration	Choppy, monotonous, garbled, ineffective, awkward	Inappropriate, clichéd, awkward, imprecise, vague, wordy

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-6 Cont'd
Writing Rubric, Composing, Grade 8

Rubric Holistic Scoring Scale						
Scores	**6**	**5**	**4**	**3**	**2**	**1**
Description	Exemplary control of the domain	Advanced control of the domain	Proficient control of the domain	Adequate control of the domain	Basic control of the domain	Minimal control of the domain

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-7
Writing Rubric, Conventions of Written English, Grade 10

3 points Advanced Control

The response demonstrates advanced control of a wide range of conventions identified in the 12th grade Wisconsin Model Academic Standards in English Language Arts:

- Uses words, phrases, and clauses effectively, including interrelated clauses in complex sentences
- Uses correct tenses, including conditionals, to indicate the relative order and relationship of events
- Employs principles of agreement, including subject-verb, pronoun-noun, and preposition-pronoun
- Punctuates compound, complex, and compound-complex sentences correctly, including appropriate use of colons, hyphens, dashes, ellipses, and italics; punctuates dialogue correctly; follows citation conventions
- Employs the conventions of capitalization
- Spells frequently used words correctly and uses effective strategies for spelling unfamiliar words
- Makes errors that are infrequent and minor

2 points Proficient Control

The response demonstrates proficient control of essential conventions identified in the 12th grade Wisconsin Model Academic Standards in English Language Arts:

- Generally controls grammar and usage (principles of agreement, noun and verb forms, pronoun references, superlative, and comparative forms)
- Generally uses phrases, dependent, and independent clauses clearly and correctly
- Uses end-stop punctuation correctly most of the time; internal punctuation (commas, apostrophes, semicolons, colons) is sometimes missing or wrong; sometimes fails to punctuate dialogue correctly or to accurately follow citation conventions
- Employs the conventions of capitalization
- Generally uses correct spelling with grade-level words and reasonable phonetic approaches to more difficult words
- Makes errors typical of those commonly found in a rough draft; errors do not seriously distract the reader

1 point Minimal Control

The response demonstrates minimal control of essential conventions identified in the 12th grade Wisconsin Model Academic Standards in English Language Arts

- Contains numerous serious end-stop or internal punctuation errors, resulting in fragments, comma splices, run-ons
- Shows poor control of grammar and usage (principles of agreement, verb and/or noun forms; pronoun reference; superlative and comparative forms)
- Shows poor control of spelling, even on grade-level words
- Makes errors that are frequent, varied, and distracting

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-8
Writing Rubric, Composing, Grade 10

Wisconsin Writing Grade 10 Rubric 6-Point Scoring Guide					
Elements of Rubric	**Purpose & Focus**	**Organization & Coherence**	**Development of Content**	**Sentence Fluency**	**Word Choice**
Element Description	Explicitly states, or strongly implies, a thesis or unifying purpose which firmly guides the paper Demonstrates understanding of the requirements of the assigned task	Frames the discussion with an effective introduction and conclusion Creates a logical structure of development for the topic, thesis, and purpose Uses effective and varied transitional strategies (from idea to idea, paragraph to paragraph, and sentence to sentence)	Demonstrates ***quality*** of invented content (e.g. of explanations, arguments, rationale, ideas, details, examples, illustrations) Demonstrates ***thoroughness*** in the elaboration of content	Demonstrates syntactic control of simple, compound, complex, and compound/complex sentences Evidences some control over stylistic effects (e.g. flow, cadence, parallelism, variety, readability, judicious use of active and passive voice, effective repetition**)**	Controls word choice with respect to both denotation and connotation Demonstrates attention to context (audience, purpose, situation, tone) Evidences some control over figurative language for rhetorical effect (e.g. metaphors, similes, hyperbole, analogies)
Positive Descriptors	Focused, unified, controlled, relevant	Well organized, integrated, smooth, controlled, coherent	Quality: clear, precise, accurate, effective, well-reasoned, insightful Thoroughness: complete, specific, well-developed, well-supported, well-illustrated	Fluid, varied, controlled, effective, skilled	Apt, discriminating, vivid, precise, concrete, concise
Negative Descriptors	Rambling, loosely related, redundant, irrelevant, lacks purpose	Disorganized, hard to follow, mechanical, illogical shifts, incoherent	Quality: vague, imprecise, inaccurate, simplistic, poorly reasoned, superficial Thoroughness: incomplete, general, inadequately, developed, inadequately supported, lacks illustration	Choppy, monotonous, garbled, ineffective, awkward	Inappropriate, clichéd, awkward, imprecise, vague, wordy

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-8 Cont'd
Writing Rubric, Composing, Grade 10

Rubric Holistic Scoring Scale						
Scores	**6**	**5**	**4**	**3**	**2**	**1**
Description	Exemplary control of the domain	Advanced control of the domain	Proficient control of the domain	Adequate control of the domain	Basic control of the domain	Minimal control of the domain

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-9
Score Distribution for Reading CR Items*

Grade	Test Book Item No.	N	Scores				Condition Codes**			
			0	1	2	3	A	B	C	D
3	19 **	60146								
	31	60146	5972	24518	27071	983	1320	2	30	250
4	13	59596	5910	22616	27144	3327	564	1	6	28
	47	59596	9955	25288	19133	4257	927	3	5	28
5	13	59207	6311	19764	27866	4491	727		6	42
	47	59207	6936	21215	27813	2602	624		5	12
6	18	60346	6257	31198	20890	1499	487		6	9
	24	60346	16177	22278	19045	2310	514		6	16
7	31	61185	3327	23118	30091	3818	673		2	156
	50	61185	20238	18525	18072	3714	614		1	21
8	19	60687	3115	20059	30862	6061	557	1	5	27
	56	60687	7412	12298	31191	8693	984	3	3	103
10	7	62717	11758	19184	21102	7905	2738		2	28
	21	62717	14172	26506	15054	5076	1884	1	2	22

* This is the score distribution of the first read.
** A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-10
Score Distribution for Mathematics CR Items*

Grade	Test Book Item No.	Part	N	Scores			Condition Codes**			
				0	1	2	A	B	C	D
3	5		60335	27644	7776	24497	414	3		1
	20	A	60335	28135	26826		5371	1		2
	20	B	60335	37094	12767	8563	1829	29	44	9
	27	A	60335	20918	38063		1336	10	6	2
	27	B	60335	35364	7809	14568	2538	31	22	3
	39	A	60335	33475	23429		3425	4	1	1
	39	B	60335	32373	12337	14354	1193	34	31	13
4	11	A	59762	47108	12182		466	2	3	1
	11	B	59762	26895	23861	7300	1657	9	26	14
	22		59762	3675	14781	40890	400	3	10	3
	25	A	59762	31588	27861		311	2		
	25	B	59762	30712	18467	9225	1309	11	32	6
	39	A	59762	24553	34850		350	6	3	
	39	B	59762	34801	8078	15784	1040	16	41	2
5	12		59256	17203	4826	36576	641		2	8
	20	A	59256	8173	50781		276	3	5	18
	20	B	59256	34376	13485	10532	813	7	19	24
	27	A	59256	22931	35907		411	4	3	
	27	B	59256	12970	6736	38772	744	10	22	2
	46	A	59256	36671	22261		319	1	4	
	46	B	59256	37780	2420	18533	504	3	14	2
6	7	A	60405	26730	31558		2113	1	2	1
	7	B	60405	31441	2279	24130	2535	7	8	5
	22	A	60405	4727	55424		247	2	2	3
	22	B	60405	30576	27574	1641	578	15	17	4
	39	A	60405	32942	26937		521	2		3
	39	B	60405	27041	14641	17694	998	13	14	4
	51		60405	20784	25330	13915	368		3	5
7	7	A	61266	16923	43940		401	1	1	
	7	B	61266	16821	9071	34506	841	3	18	6
	19	A	61266	28112	32488		631	12	23	
	19	B	61266	30384	20913	8660	1281	3	21	4
	38	A	61266	22624	33099		5538	2	1	2
	38	B	61266	22615	30193	6410	2020	10	11	7
	52		61266	28345	20498	11907	502		5	9

* This is the score distribution of the first read.
** A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-10 Cont'd
Score Distribution for Mathematics CR Items*

Grade	Test Book Item No.	Part	N	Scores			Condition Codes**			
				0	1	2	A	B	C	D
8	10		60729	42328	8958	7931	1510			2
	19	A	60729	17328	42805		596			
	19	B	60729	17812	11458	29678	1764	1	6	10
	33	A	60729	39449	17344		3933	1	1	1
	33	B	60729	25854	22715	9053	3079	4	13	11
	49	A	60729	43949	15892		887		1	
	49	B	60729	44216	5039	10081	1371		10	12
10	23		62644	19969	24005	15916	2719	2	16	17
	33		62644	29593	15415	11353	6252		4	27
	36		62644	21687	13930	22514	4491		9	13
	46		62644	24410	13377	22548	2268	3	16	22

* This is the score distribution of the first read.
** A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-11
Score Distribution for Grades 4, 8, and 10 Writing: Composing Rubric

Grade	Rater	Total N	Scores						Condition Codes**			
			1	2	3	4	5	6	A	B	C	D
4	Rater 1	3070	60	534	1194	1033	166	19	19			43
	Rater 2	3070	63	507	1225	1027	183	7	21			35
	Diff*	0	-3	27	-31	6	-17	12	-2	0	0	8
8	Rater 1	3029	29	449	1117	1088	266	12	35			33
	Rater 2	3029	32	445	1142	1087	256	10	35			22
	Diff*	0	-3	4	-25	1	10	2	0	0		11
10	Rater 1	3148	94	452	838	1279	323	45	84	1	1	31
	Rater 2	3148	99	443	873	1246	331	43	86			26
	Diff*	0	-5	9	-35	33	-8	2	-2	1	1	5

* Diff = N of Rater1 – N of Rater 2.
** A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.

Table 5-12
Percentage Distribution of Scores, Grades 4, 8, and 10 Writing: Composing Rubric

Grade	Rater	Total N	Scores						Condition Codes**			
			1	2	3	4	5	6	A	B	C	D
4	Rater 1	3070	1.96	17.40	38.90	33.64	5.40	0.62	0.62	0.00	0.00	1.40
	Rater 2	3070	2.06	16.52	39.90	33.46	5.96	0.22	0.68	0.00	0.00	1.14
8	Rater 1	3029	0.96	14.82	36.88	35.92	8.78	0.40	1.16	0.00		1.08
	Rater 2	3029	1.06	14.70	37.70	35.88	8.46	0.34	1.16	0.00		0.72
10	Rater 1	3148	2.98	14.36	26.62	40.62	10.26	1.42	2.66	0.04	0.04	0.98
	Rater 2	3148	3.14	14.08	27.74	39.58	10.52	1.36	2.74	0.00	0.00	0.82

** A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-13
Score Distribution, Grades 4, 8, and 10 Writing: Conventions Rubric

Grade	Rater	Total N	Scores			Condition Codes**		
			1	2	3	A	B	C
4	Rater 1	3070	56	2971	22	19		2
	Rater 2	3070	63	2954	30	21		2
	Diff*	0	-7	17	-8	-2	0	0
8	Rater 1	3029	26	2948	20	35		
	Rater 2	3029	29	2940	25	35		
	Diff*	0	-3	8	-5	0	0	0
10	Rater 1	3148	19	2798	245	84	1	1
	Rater 2	3148	20	2783	258	86		1
	Diff*	0	-1	15	-13	-2	1	0

* Diff = N of Rater 1 – N of Rater 2.
** A: No response or no attempt, B: Illegible, C: Another Language.

Table 5-14
Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Conventions Rubric

Grade	Rater	Total N	Scores			Condition Codes**		
			1	2	3	A	B	C
4	Rater 1	3070	1.82	96.78	0.72	0.62	0.00	0.06
	Rater 2	3070	2.06	96.22	0.98	0.68	0.00	0.06
8	Rater 1	3029	0.86	97.32	0.66	1.16		
	Rater 2	3029	0.96	97.06	0.82	1.16		
10	Rater 1	3148	0.60	88.88	7.78	2.66	0.04	0.04
	Rater 2	3148	0.64	88.40	8.20	2.74	0.00	0.04

** A: No response or no attempt, B: Illegible, C: Another Language.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 5-15
Score Distribution for Grades 4, 8, and 10 Writing: Total Score, Composing and Conventions Combined

Grade	Rater	Total N	Scores									
			0	1	2	3	4	5	6	7	8	9
4	Rater 1	3070	21	4	63	55	522	1187	1027	169	11	11
	Rater 2	3070	23	2	61	58	492	1216	1025	165	25	3
	Diff*	0	-2	2	2	-3	30	-29	2	4	-14	8
8	Rater 1	3029	35	2	41	33	435	1115	1085	267	7	9
	Rater 2	3029	35	2	37	24	436	1140	1086	248	13	8
	Diff*	0	0	0	4	9	-1	-25	-1	19	-6	1
10	Rater 1	3148	86	5	38	84	449	832	1268	162	184	40
	Rater 2	3148	87	2	35	94	437	871	1220	172	190	40
	Diff*	0	-1	3	3	-10	12	-39	48	-10	-6	0

* Diff = N of Rater 1 – N of Rater 2.

Table 5-16
Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Total Score, Composing and Conventions Combined

Grade	Rater	Total N	Scores									
			0	1	2	3	4	5	6	7	8	9
4	Rater 1	3070	0.68	0.14	2.06	1.80	17.00	38.66	33.46	5.50	0.36	0.36
	Rater 2	3070	0.74	0.06	1.98	1.88	16.02	39.60	33.38	5.38	0.82	0.10
8	Rater 1	3029	1.16	0.06	1.36	1.08	14.36	36.82	35.82	8.82	0.24	0.30
	Rater 2	3029	1.16	0.06	1.22	0.80	14.40	37.64	35.86	8.18	0.42	0.26
10	Rater 1	3148	2.74	0.16	1.20	2.66	14.26	26.42	40.28	5.14	5.84	1.28
	Rater 2	3148	2.76	0.06	1.12	2.98	13.88	27.66	38.76	5.46	6.04	1.28

Copyright © 2012 by the Wisconsin Department of Public Instruction

Table 7-1
Item Flagged Based on Yen's Q_1

Content	Grade	Item Number	Type	N	Z	Critical Z
RD	6	44	MC	5911	17.12	15.76
RD	7	11	MC	5268	27.32	14.05
RD	7	30	MC	5268	34.95	14.05
RD	7	51	MC	5268	22.49	14.05
RD	8	4	MC	5410	18.30	14.43
RD	8	46	MC	5410	18.99	14.43
MA	6	39B	CR	5941	35.01	15.84
MA	8	51	MC	5421	16.77	14.46
MA	10	46	CR	5714	19.19	15.24
LA	8	2	MC	5207	14.97	13.89
LA	10	32	CR	5672	16.33	15.13
SC	4	23	MC	5264	17.24	14.04
SS	8	20	MC	5344	15.60	14.25
SS	10	2	MC	5654	21.25	15.08
SS	10	42	MC	5654	16.04	15.08

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-2
Scoring Table for Reading Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	119	31	436	7
1	270	119	32	439	7
2	270	119	33	441	7
3	270	119	34	443	7
4	270	119	**35**	**445**	**7**
5	270	119	36	448	7
6	270	119	37	450	7
7	270	119	38	452	7
8	270	119	39	455	7
9	270	119	40	457	7
10	270	119	41	460	7
11	270	119	42	462	8
12	339	50	43	465	8
13	362	31	44	468	8
14	374	24	45	471	8
15	383	19	**46**	**474**	**8**
16	389	16	47	478	9
17	395	14	48	481	9
18	400	13	49	485	9
19	404	11	50	490	10
20	407	11	51	495	11
21	411	10	52	502	13
22	414	9	**53**	**510**	**15**
23	417	9	54	521	18
24	419	9	55	539	25
25	422	8	56	582	47
26	425	8	57**	640	72
27	427	8			
28	429	8			
29	432	7			
30	434	7			

* **Bold** represents SEM around cut score.
** A suppressed item in Reading grade 3 reduced the maximum possible score from 60 to 57. See Part 8 for more information.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-3
Scoring Table for Reading Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	280	110	31	452	9
1	280	110	32	455	9
2	280	110	33	458	9
3	280	110	**34**	**461**	**9**
4	280	110	35	464	9
5	280	110	36	467	9
6	280	110	37	469	9
7	280	110	38	472	9
8	280	110	39	475	9
9	280	110	40	479	9
10	280	110	41	482	9
11	312	78	42	485	10
12	347	47	43	489	10
13	365	34	44	492	10
14	378	27	45	496	11
15	387	23	**46**	**500**	**11**
16	395	19	47	504	11
17	401	17	48	509	12
18	407	15	49	514	12
19	412	14	50	519	13
20	416	13	51	525	14
21	420	12	52	531	14
22	424	11	**53**	**538**	**15**
23	428	11	54	546	16
24	431	11	55	555	18
25	434	10	56	567	21
26	438	10	57	583	26
27	441	10	58	605	33
28	444	10	59	642	49
29	447	9	60	650	53
30	450	9			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-4
Scoring Table for Reading Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	290	110	31	456	10
1	290	110	32	460	10
2	290	110	**33**	**463**	**10**
3	290	110	34	466	9
4	290	110	35	469	9
5	290	110	36	472	9
6	290	110	37	475	9
7	290	110	38	478	9
8	290	110	39	481	9
9	290	110	40	484	9
10	290	110	41	487	10
11	290	110	42	491	10
12	290	110	43	494	10
13	341	62	44	498	10
14	366	39	45	501	10
15	381	28	**46**	**505**	**11**
16	391	22	47	510	11
17	399	19	48	514	12
18	405	17	49	519	12
19	411	15	50	524	13
20	416	14	51	530	13
21	421	13	52	537	14
22	425	12	53	544	15
23	429	12	**54**	**552**	**17**
24	433	11	55	563	19
25	437	11	56	575	22
26	440	11	57	592	27
27	444	11	58	616	35
28	447	10	59	658	53
29	450	10	60	690	72
30	453	10			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-5
Scoring Table for Reading Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	104	31	475	12
1	300	104	**32**	**478**	**12**
2	300	104	33	482	12
3	300	104	34	486	12
4	300	104	35	489	12
5	300	104	36	493	12
6	300	104	37	497	12
7	300	104	38	500	12
8	300	104	39	504	12
9	300	104	40	508	12
10	300	104	41	512	12
11	300	104	42	516	12
12	300	104	43	520	12
13	332	76	**44**	**525**	**12**
14	363	51	45	529	13
15	381	38	46	534	13
16	394	30	47	539	13
17	405	24	48	544	14
18	413	21	49	550	14
19	420	19	50	556	15
20	426	17	51	562	15
21	432	16	52	569	16
22	437	15	**53**	**578**	**18**
23	442	14	54	587	19
24	447	14	55	598	21
25	451	13	56	611	24
26	455	13	57	629	29
27	459	12	58	654	38
28	463	12	59	699	58
29	467	12	60	730	77
30	471	12			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-6
Scoring Table for Reading Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	112	31	484	12
1	310	112	**32**	**487**	**12**
2	310	112	33	491	12
3	310	112	34	495	12
4	310	112	35	499	12
5	310	112	36	503	12
6	310	112	37	507	12
7	310	112	38	511	12
8	310	112	39	514	12
9	310	112	40	518	12
10	310	112	41	522	12
11	310	112	42	526	12
12	328	94	43	531	12
13	366	56	**44**	**535**	**12**
14	386	38	45	539	13
15	399	29	46	544	13
16	409	24	47	549	13
17	418	20	48	554	14
18	425	18	49	560	14
19	431	16	50	566	15
20	437	15	51	572	16
21	442	14	52	580	17
22	447	14	53	588	18
23	451	13	**54**	**597**	**20**
24	456	13	55	609	23
25	460	13	56	623	26
26	464	13	57	641	31
27	468	12	58	668	40
28	472	12	59	716	62
29	476	12	60	780	105
30	480	12			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-7
Scoring Table for Reading Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	74	31	487	12
1	330	74	32	491	12
2	330	74	**33**	**495**	**12**
3	330	74	34	498	12
4	330	74	35	502	12
5	330	74	36	506	12
6	330	74	37	510	12
7	330	74	38	513	12
8	330	74	39	517	12
9	330	74	40	521	12
10	330	74	41	525	12
11	330	74	42	529	12
12	330	74	43	534	12
13	330	74	44	538	13
14	350	59	45	543	13
15	375	45	**46**	**547**	**13**
16	393	37	47	552	14
17	406	32	48	558	14
18	417	27	49	563	14
19	426	24	50	569	15
20	434	22	51	576	16
21	440	20	52	583	17
22	447	18	53	592	18
23	452	17	54	601	20
24	457	16	**55**	**613**	**22**
25	462	15	56	627	26
26	467	14	57	647	32
27	471	14	58	676	44
28	475	13	59	731	71
29	479	13	60	790	112
30	483	13			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-8
Scoring Table for Reading Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	75	29	502	15
1	350	75	30	507	15
2	350	75	31	512	15
3	350	75	32	516	15
4	350	75	33	521	15
5	350	75	34	526	15
6	350	75	35	530	15
7	350	75	36	535	15
8	350	75	37	540	15
9	350	75	38	545	15
10	350	75	39	549	15
11	350	75	40	555	15
12	350	75	41	560	15
13	366	60	**42**	**565**	**15**
14	388	45	43	571	16
15	404	36	44	577	16
16	417	31	45	583	16
17	427	28	46	590	17
18	437	25	47	597	18
19	445	23	48	604	18
20	452	22	49	613	20
21	459	20	50	623	21
22	465	19	51	634	24
23	471	19	**52**	**649**	**27**
24	477	18	53	667	33
25	482	17	54	694	42
26	488	17	55	743	65
27	493	16	56	820	123
28	**498**	**16**			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-9
Scoring Table for Mathematics Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	220	100	29	386	11
1	220	100	**30**	**390**	**11**
2	220	100	31	394	11
3	220	100	32	397	11
4	220	100	33	401	11
5	220	100	34	405	11
6	220	100	35	408	11
7	220	100	36	412	11
8	220	100	37	416	11
9	220	100	38	420	11
10	244	76	39	423	11
11	274	47	40	427	11
12	291	34	41	431	12
13	303	28	42	435	12
14	313	23	**43**	**440**	**12**
15	321	21	44	444	12
16	328	19	45	448	12
17	334	17	46	453	12
18	340	16	47	458	13
19	345	15	48	463	13
20	350	14	49	468	14
21	355	14	50	474	14
22	359	13	51	481	15
23	363	13	52	488	16
24	367	12	**53**	**496**	**18**
25	371	12	54	507	21
26	375	12	55	522	26
27	379	12	56	547	37
28	383	11	57	630	109

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-10
Scoring Table for Mathematics Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	71	29	419	13
1	240	71	30	423	13
2	240	71	**31**	**427**	**12**
3	240	71	32	431	12
4	240	71	33	435	12
5	240	71	34	439	12
6	240	71	35	443	12
7	240	71	36	447	12
8	240	71	37	451	11
9	240	71	38	455	11
10	240	71	39	459	11
11	244	69	40	463	11
12	279	52	41	467	12
13	302	43	42	471	12
14	319	35	**43**	**475**	**12**
15	332	30	44	480	12
16	343	25	45	484	12
17	352	23	46	489	12
18	360	20	47	494	13
19	367	19	48	499	13
20	374	17	49	505	14
21	380	17	50	512	15
22	385	16	51	519	16
23	391	15	**52**	**527**	**17**
24	396	15	53	538	20
25	401	14	54	551	24
26	405	14	55	570	30
27	410	13	56	604	46
28	414	13	57	650	78

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-11
Scoring Table for Mathematics Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	83	32	456	13
1	270	83	33	459	12
2	270	83	34	463	12
3	270	83	35	467	12
4	270	83	36	470	12
5	270	83	37	474	11
6	270	83	38	477	11
7	270	83	39	481	11
8	270	83	40	484	11
9	270	83	41	488	11
10	270	83	42	491	11
11	271	82	43	495	11
12	306	50	44	499	11
13	327	38	**45**	**502**	**11**
14	342	32	46	506	11
15	354	28	47	510	11
16	365	25	48	514	12
17	374	23	49	518	12
18	382	22	50	522	12
19	390	20	51	527	12
20	397	19	52	532	13
21	403	18	53	537	13
22	409	18	54	543	14
23	415	17	55	549	15
24	420	16	**56**	**556**	**16**
25	425	16	57	564	18
26	430	15	58	574	20
27	435	15	59	587	24
28	439	14	60	606	31
29	444	14	61	641	48
30	**448**	**13**	62	680	75
31	452	13			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-12
Scoring Table for Mathematics Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	69	32	489	12
1	310	69	33	492	12
2	310	69	34	496	11
3	310	69	35	499	11
4	310	69	36	503	11
5	310	69	37	506	11
6	310	69	38	509	11
7	310	69	39	513	11
8	310	69	40	516	11
9	310	69	41	519	11
10	310	69	**42**	**523**	**11**
11	310	69	43	526	11
12	335	52	44	530	11
13	361	39	45	533	11
14	378	33	46	537	11
15	392	29	47	541	11
16	403	26	48	545	12
17	413	23	49	549	12
18	421	21	50	553	12
19	428	19	51	558	13
20	435	18	52	563	13
21	441	17	53	568	14
22	446	16	**54**	**574**	**14**
23	451	15	55	580	15
24	456	15	56	588	17
25	461	14	57	596	18
26	465	14	58	607	21
27	469	13	59	622	25
28	**474**	**13**	60	643	32
29	477	13	61	680	47
30	481	12	62	700	56
31	485	12			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-13
Scoring Table for Mathematics Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	90	32	506	10
1	330	90	33	509	10
2	330	90	34	513	10
3	330	90	35	516	10
4	330	90	36	519	10
5	330	90	37	522	10
6	330	90	38	525	10
7	330	90	39	528	10
8	330	90	40	531	10
9	330	90	41	534	10
10	330	90	42	538	10
11	330	90	43	541	10
12	365	55	**44**	**544**	**10**
13	390	35	45	547	10
14	406	28	46	551	10
15	418	24	47	554	10
16	428	22	48	558	11
17	436	20	49	562	11
18	444	18	50	566	11
19	450	17	51	570	11
20	457	16	52	575	12
21	462	15	53	579	12
22	467	14	54	584	13
23	472	13	**55**	**590**	**13**
24	477	13	56	597	14
25	481	12	57	604	15
26	485	12	58	613	17
27	489	11	59	624	20
28	492	11	60	640	25
29	496	11	61	669	38
30	**499**	**11**	62	710	67
31	503	10			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-14
Scoring Table for Mathematics Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	101	32	536	12
1	350	101	33	540	12
2	350	101	34	543	12
3	350	101	35	547	11
4	350	101	36	550	11
5	350	101	37	554	11
6	350	101	**38**	**557**	**11**
7	350	101	39	561	11
8	350	101	40	564	11
9	350	101	41	567	11
10	350	101	42	571	11
11	350	101	43	574	11
12	388	63	44	578	11
13	417	41	45	581	11
14	434	32	46	585	11
15	447	27	47	589	11
16	457	23	48	592	11
17	465	21	49	596	11
18	472	19	50	600	12
19	479	17	**51**	**605**	**12**
20	485	16	52	609	12
21	490	15	53	614	13
22	495	14	54	620	14
23	500	14	55	626	15
24	505	14	56	632	16
25	**509**	**13**	57	640	17
26	513	13	58	650	20
27	517	13	59	662	23
28	521	13	60	681	30
29	525	12	61	715	48
30	529	12	62	730	58
31	533	12			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-15
Scoring Table for Mathematics Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	410	81	30	555	9
1	410	81	31	558	9
2	410	81	32	560	9
3	410	81	33	563	8
4	410	81	34	565	8
5	410	81	35	568	8
6	410	81	36	570	8
7	410	81	**37**	**573**	**8**
8	410	81	38	575	8
9	410	81	39	577	8
10	410	81	40	580	8
11	432	59	41	582	8
12	459	35	42	585	8
13	475	26	43	587	8
14	486	21	44	590	8
15	495	18	45	593	8
16	502	16	46	595	8
17	508	15	47	598	8
18	513	14	48	601	9
19	518	13	49	605	9
20	523	12	50	608	9
21	**527**	**12**	51	612	10
22	531	11	**52**	**617**	**11**
23	534	11	53	622	12
24	538	10	54	629	13
25	541	10	55	637	15
26	544	10	56	649	19
27	547	10	57	669	28
28	550	9	58	750	100
29	553	9			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-16
Scoring Table for Language Arts Grade 4

Raw Score	Scale Score	SEM
0	140	107
1	140	107
2	140	107
3	140	107
4	140	107
5	140	107
6	165	82
7	215	32
8	230	20
9	240	16
10	248	14
11	**254**	**13**
12	260	12
13	265	11
14	270	10
15	274	10
16	**278**	**10**
17	283	9
18	287	9
19	291	9
20	295	9
21	299	9
22	304	10
23	**309**	**10**
24	314	10
25	320	11
26	328	12
27	337	15
28	351	20
29	377	33
30	420	66

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-17
Scoring Table for Language Arts Grade 8

Raw Score	Scale Score	SEM
0	250	83
1	250	83
2	250	83
3	250	83
4	250	83
5	250	83
6	250	83
7	273	60
8	299	35
9	314	25
10	325	20
11	333	17
12	340	15
13	347	13
14	352	13
15	**358**	**12**
16	363	12
17	368	12
18	374	12
19	379	12
20	384	12
21	**390**	**12**
22	396	12
23	402	13
24	409	14
25	417	15
26	**427**	**17**
27	440	20
28	463	30
29**	520	78

* Bold represents SEM around cut score.
** A suppressed item in Language Arts grade 8 reduced the maximum possible score from 30 to 29. See Part 8 for more information.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-18
Scoring Table for Language Arts Grade 10

Raw Score	Scale Score	SEM
0	290	68
1	290	68
2	290	68
3	290	68
4	290	68
5	290	68
6	292	66
7	326	33
8	342	22
9	353	18
10	361	17
11	368	16
12	375	16
13	382	16
14	388	16
15	**394**	**15**
16	399	15
17	405	14
18	410	14
19	416	13
20	421	13
21	426	13
22	**431**	**13**
23	436	13
24	442	13
25	447	13
26	453	14
27	459	14
28	465	14
29	472	15
30	479	16
31	**488**	**17**
32	497	18
33	507	19
34	518	21
35	531	22
36	547	25
37	567	30
38	600	43
39	630	60

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-19
Scoring Table for Social Studies Grade 4

Raw Score	Scale Score	SEM
0	170	69
1	170	69
2	170	69
3	170	69
4	170	69
5	170	69
6	170	69
7	170	69
8	197	42
9	216	23
10	225	17
11	232	13
12	238	11
13	**243**	**10**
14	247	9
15	250	9
16	254	8
17	257	8
18	260	8
19	**263**	**8**
20	266	7
21	269	7
22	272	7
23	275	7
24	277	7
25	280	7
26	283	7
27	285	7
28	**288**	**7**
29	291	7
30	294	7
31	297	7
32	301	7
33	305	8
34	309	8
35	315	10
36	323	12
37	337	19
38	400	77

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-20
Scoring Table for Social Studies Grade 8

Raw Score	Scale Score	SEM
0	230	96
1	230	96
2	230	96
3	230	96
4	230	96
5	230	96
6	230	96
7	230	96
8	230	96
9	230	96
10	261	65
11	294	32
12	308	21
13	318	17
14	326	15
15	333	15
16	**339**	**14**
17	345	13
18	350	13
19	355	12
20	360	12
21	**364**	**12**
22	369	11
23	373	11
24	377	11
25	381	11
26	386	11
27	390	11
28	394	11
29	398	11
30	**403**	**11**
31	407	11
32	412	11
33	417	11
34	423	12
35	429	13
36	436	14
37	446	16
38	458	19
39	481	30
40	530	71

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-21
Scoring Table for Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	126	26	429	11
1	240	126	27	433	11
2	240	126	28	437	11
3	240	126	29	440	11
4	240	126	30	444	11
5	240	126	31	448	10
6	240	126	32	451	10
7	240	126	**33**	**455**	**10**
8	240	126	34	459	10
9	240	126	35	463	10
10	274	92	36	467	10
11	324	42	37	471	10
12	343	29	38	475	11
13	356	23	39	479	11
14	366	20	40	484	11
15	374	18	41	490	12
16	381	17	42	495	13
17	387	15	43	502	14
18	393	14	44	510	15
19	398	14	45	519	17
20	403	13	46	532	21
21	**408**	**13**	47	551	27
22	413	12	48	586	44
23	417	12	49**	620	65
24	**421**	**12**			
25	425	12			

* **Bold** represents SEM around cut score.
** A suppressed item in Social Studies grade 10 reduced the maximum possible score from 50 to 49.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-22
Scoring Table for Science Grade 4

Raw Score	Scale Score	SEM
0	170	60
1	170	60
2	170	60
3	170	60
4	170	60
5	170	60
6	170	60
7	170	60
8	170	60
9	170	60
10	189	41
11	208	25
12	219	19
13	228	16
14	235	14
15	241	13
16	246	12
17	**251**	**11**
18	255	10
19	259	10
20	263	9
21	267	9
22	271	9
23	274	9
24	278	9
25	**281**	**9**
26	285	9
27	288	9
28	292	9
29	296	9
30	299	9
31	303	9
32	307	9
33	311	9
34	316	9
35	**321**	**10**
36	328	11
37	337	14
38	351	20
39	380	36
40	440	90

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-23
Scoring Table for Science Grade 8

Raw Score	Scale Score	SEM
0	230	81
1	230	81
2	230	81
3	230	81
4	230	81
5	230	81
6	230	81
7	230	81
8	230	81
9	230	81
10	260	51
11	282	31
12	295	23
13	305	20
14	314	18
15	321	16
16	328	16
17	334	15
18	340	14
19	346	14
20	**351**	**14**
21	356	13
22	361	13
23	365	12
24	370	12
25	374	12
26	**379**	**12**
27	383	11
28	388	11
29	392	11
30	397	12
31	402	12
32	407	12
33	413	13
34	**419**	**13**
35	426	14
36	435	16
37	445	18
38	459	22
39	483	31
40	560	98

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-24
Scoring Table for Science Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	137	26	434	12
1	240	137	27	438	12
2	240	137	28	442	12
3	240	137	29	446	12
4	240	137	30	450	12
5	240	137	31	454	12
6	240	137	32	458	12
7	240	137	33	462	12
8	240	137	**34**	**466**	**12**
9	240	137	35	471	12
10	290	87	36	475	12
11	327	50	37	479	12
12	346	34	38	484	12
13	359	27	39	489	13
14	370	22	40	494	13
15	378	20	41	499	13
16	385	18	42	505	14
17	392	17	43	511	14
18	398	16	44	518	15
19	403	15	45	527	17
20	408	14	46	536	18
21	**413**	**14**	47	548	21
22	417	13	48	564	26
23	422	13	49	592	38
24	426	13	50	610	48
25	**430**	**12**			

* **Bold** represents SEM around cut score.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 7-25
The Number of Students and Percents at LOSS and HOSS

Content	Grade	LOSS	N	Percent	HOSS	N	Percent
RD	3	270	766	1.27	640	43	0.07
	4	280	568	0.95	650	9	0.02
	5	290	903	1.52	690	2	0.00
	6	300	576	0.95	730	6	0.01
	7	310	314	0.51	780	0	0.00
	8	330	291	0.48	790	21	0.03
	10	350	662	1.06	820	63	0.10
MA	3	220	63	0.10	630	242	0.40
	4	240	62	0.10	650	125	0.21
	5	270	74	0.12	680	136	0.23
	6	310	101	0.17	700	41	0.07
	7	330	88	0.14	710	109	0.18
	8	350	456	0.75	730	108	0.18
	10	410	540	0.86	750	287	0.46
LA	4	140	274	0.46	420	181	0.30
	8	250	556	0.92	520	1854	3.07
	10	290	195	0.31	630	7	0.01
SS	4	170	148	0.25	400	1244	2.09
	8	230	319	0.53	530	563	0.93
	10	240	168	0.27	620	169	0.27
SC	4	170	98	0.16	440	362	0.61
	8	230	126	0.21	560	866	1.43
	10	240	219	0.35	610	75	0.12

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-1
Item Analysis Grade 3 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.59	0.18	0.05%				
2	MC	0.64	0.40	0.13%				
3	MC	0.85	0.51	0.17%				
4	MC	0.66	0.52	0.43%				
5	MC	0.62	0.42	0.31%				
6	MC	0.57	0.40	0.45%				
7	MC	0.74	0.53	1.11%				
8	MC	0.85	0.56	0.32%				
9	MC	0.73	0.48	0.82%				
10	MC	0.80	0.40	0.58%				
11	MC	0.68	0.55	0.99%				
12	MC	0.75	0.51	0.97%				
13	MC	0.84	0.49	1.31%				
14	MC	0.68	0.50	1.59%				
15	MC	0.76	0.47	2.26%				
16	MC	0.76	0.47	3.88%				
17	MC	0.43	0.42	1.88%				
18	MC	0.68	0.47	2.17%				
19**	CR	.	.					
20	MC	0.60	0.43	0.14%				
21	MC	0.85	0.42	0.21%				
22	MC	0.79	0.52	0.55%				
23	MC	0.56	0.34	0.28%				
24	MC	0.71	0.38	0.40%				
25	MC	0.60	0.49	1.07%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.
** Item dropped from scoring.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-1 Cont'd
Item Analysis Grade 3 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.74	0.52	0.21%				
27	MC	0.42	0.24	0.39%				
28	MC	0.79	0.49	0.76%				
29	MC	0.73	0.54	0.45%				
30	MC	0.66	0.35	0.98%				
31	CR	0.46	0.45	2.19%				
32	MC	0.97	0.27	1.26%		+		
33	MC	0.81	0.44	1.55%				
34	MC	0.87	0.48	1.96%				
35	MC	0.89	0.46	1.48%				
36	MC	0.70	0.37	1.68%				
37	MC	0.73	0.37	1.66%				
38	MC	0.46	0.38	1.87%				
39	MC	0.75	0.45	0.14%				
40	MC	0.65	0.46	0.21%				
41	MC	0.52	0.34	0.29%				
42	MC	0.77	0.55	0.56%				
43	MC	0.83	0.52	1.63%				
44	MC	0.60	0.49	0.21%				
45	MC	0.79	0.44	0.41%				
46	MC	0.87	0.55	0.30%				
47	MC	0.56	0.26	1.22%				
48	MC	0.83	0.44	0.30%				
49	MC	0.61	0.35	0.46%				
50	MC	0.84	0.48	0.57%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-1 Cont'd
Item Analysis Grade 3 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.77	0.48	0.54%				
52	MC	0.79	0.53	0.67%				
53	MC	0.50	0.24	1.46%				
54	MC	0.64	0.47	0.55%				
55	MC	0.74	0.56	1.23%				
56	MC	0.56	0.44	0.67%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-2
Item Analysis Grade 4 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.80	0.50	0.01%				
2	MC	0.59	0.50	0.07%				
3	MC	0.54	0.26	0.34%				
4	MC	0.85	0.49	0.75%				
5	MC	0.71	0.44	0.06%				
6	MC	0.66	0.47	0.29%				
7	MC	0.80	0.48	0.13%				
8	MC	0.87	0.42	1.06%				
9	MC	0.74	0.56	0.17%				
10	MC	0.72	0.53	0.48%		+		
11	MC	0.88	0.49	0.40%				
12	MC	0.57	0.31	1.11%				
13	CR	0.49	0.45	0.95%				
14	MC	0.77	0.42	1.02%				
15	MC	0.66	0.18	1.34%				
16	MC	0.66	0.32	1.19%				
17	MC	0.83	0.29	1.55%				
18	MC	0.89	0.43	1.32%				
19	MC	0.39	0.35	0.13%				
20	MC	0.80	0.46	0.19%				
21	MC	0.55	0.21	0.17%				
22	MC	0.66	0.22	0.52%				
23	MC	0.86	0.38	0.98%				
24	MC	0.53	0.21	0.29%				
25	MC	0.59	0.40	0.36%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-2 Cont'd
Item Analysis Grade 4 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.88	0.50	0.51%				
27	MC	0.72	0.37	0.28%				
28	MC	0.42	0.21	0.62%				
29	MC	0.39	0.30	0.54%				
30	MC	0.45	0.23	0.46%				
31	MC	0.33	0.22	0.74%				
32	MC	0.56	0.43	0.68%				
33	MC	0.71	0.24	0.79%				
34	MC	0.80	0.51	0.87%				
35	MC	0.52	0.42	0.83%				
36	MC	0.77	0.52	1.33%				
37	MC	0.84	0.43	0.16%				
38	MC	0.58	0.30	0.32%				
39	MC	0.54	0.25	0.50%				
40	MC	0.66	0.57	0.38%				
41	MC	0.89	0.41	0.21%				
42	MC	0.88	0.45	0.30%				
43	MC	0.46	0.31	0.41%				
44	MC	0.78	0.42	0.52%				
45	MC	0.66	0.48	1.01%				
46	MC	0.33	0.19	1.68%				
47	CR	0.43	0.45	1.56%				
48	MC	0.59	0.41	0.61%				
49	MC	0.71	0.58	0.80%				
50	MC	0.68	0.35	0.82%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-2 Cont'd
Item Analysis Grade 4 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.67	0.46	0.72%				
52	MC	0.75	0.50	0.92%				
53	MC	0.59	0.41	0.93%				
54	MC	0.71	0.54	0.98%				
55	MC	0.60	0.42	1.11%				
56	MC	0.66	0.43	1.28%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-3
Item Analysis Grade 5 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.88	0.41	0.02%				
2	MC	0.89	0.40	0.07%				
3	MC	0.78	0.49	0.10%				
4	MC	0.67	0.41	0.32%				
5	MC	0.31	0.35	0.42%				
6	MC	0.66	0.33	0.39%				
7	MC	0.80	0.47	0.56%				
8	MC	0.80	0.44	1.15%				
9	MC	0.33	0.17	0.18%		+		
10	MC	0.64	0.45	0.24%				
11	MC	0.57	0.37	0.41%				
12	MC	0.85	0.51	0.18%				
13	CR	0.51	0.41	1.23%				
14	MC	0.88	0.40	1.49%				
15	MC	0.78	0.39	1.54%				
16	MC	0.51	0.35	1.82%				
17	MC	0.73	0.40	1.83%				
18	MC	0.86	0.43	1.97%				
19	MC	0.47	0.17	2.06%		+		
20	MC	0.69	0.32	0.17%				
21	MC	0.39	0.36	0.22%				
22	MC	0.61	0.29	0.40%				
23	MC	0.71	0.44	0.66%				
24	MC	0.42	0.17	1.02%		+		
25	MC	0.70	0.35	0.22%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-3 Cont'd
Item Analysis Grade 5 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.51	0.22	0.29%				
27	MC	0.68	0.37	0.53%				
28	MC	0.51	0.40	0.17%				
29	MC	0.57	0.49	0.35%				
30	MC	0.75	0.43	0.40%				
31	MC	0.70	0.47	0.25%				
32	MC	0.73	0.33	0.55%				
33	MC	0.70	0.48	0.60%				
34	MC	0.68	0.43	0.96%				
35	MC	0.62	0.46	0.35%				
36	MC	0.46	0.20	0.48%				
37	MC	0.45	0.12	0.61%	+			
38	MC	0.60	0.29	1.18%				
39	MC	0.95	0.36	0.12%				
40	MC	0.83	0.47	0.75%				
41	MC	0.91	0.48	0.22%				
42	MC	0.87	0.43	0.20%				
43	MC	0.86	0.42	0.37%				
44	MC	0.82	0.48	0.51%				
45	MC	0.82	0.36	1.18%				
46	MC	0.49	0.16	2.22%		+		
47	CR	0.48	0.52	1.05%				
48	MC	0.88	0.45	0.35%				
49	MC	0.60	0.46	0.52%				
50	MC	0.59	0.32	0.75%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-3 Cont'd
Item Analysis Grade 5 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.55	0.35	0.57%				
52	MC	0.75	0.52	0.45%				
53	MC	0.67	0.49	0.52%				
54	MC	0.77	0.46	3.07%				
55	MC	0.68	0.58	0.64%				
56	MC	0.64	0.37	1.05%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-4
Item Analysis Grade 6 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.54	0.23	0.09%				
2	MC	0.91	0.41	0.03%				
3	MC	0.62	0.43	0.09%				
4	MC	0.82	0.50	0.09%				
5	MC	0.67	0.22	0.38%				
6	MC	0.67	0.44	0.10%				
7	MC	0.45	0.28	0.09%				
8	MC	0.86	0.40	1.82%				
9	MC	0.60	0.18	2.55%				
10	MC	0.35	0.20	0.15%		+		
11	MC	0.79	0.34	0.20%				
12	MC	0.68	0.32	0.37%				
13	MC	0.49	0.36	0.37%				
14	MC	0.82	0.33	0.14%				
15	MC	0.79	0.34	0.41%				
16	MC	0.72	0.42	0.19%				
17	MC	0.78	0.42	0.28%				
18	CR	0.43	0.40	0.81%				
19	MC	0.39	0.13	0.14%	+	+		
20	MC	0.86	0.14	0.19%	+			
21	MC	0.70	0.45	0.40%				
22	MC	0.34	0.19	0.30%				
23	MC	0.53	0.33	0.63%				
24	CR	0.38	0.52	0.85%				
25	MC	0.57	0.34	0.24%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-4 Cont'd
Item Analysis Grade 6 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.44	0.19	0.61%		+		
27	MC	0.55	0.44	0.43%				
28	MC	0.76	0.26	1.93%				
29	MC	0.52	0.30	0.46%				
30	MC	0.63	0.36	0.85%				
31	MC	0.69	0.41	1.30%				
32	MC	0.45	0.29	2.33%				
33	MC	0.47	0.34	2.82%				
34	MC	0.65	0.41	1.41%				
35	MC	0.89	0.45	2.17%				
36	MC	0.81	0.45	1.76%				
37	MC	0.86	0.49	1.67%				
38	MC	0.84	0.54	0.16%				
39	MC	0.67	0.43	0.24%				
40	MC	0.73	0.28	0.32%				
41	MC	0.71	0.38	0.27%				
42	MC	0.85	0.50	0.44%				
43	MC	0.68	0.35	1.53%				
44	MC	0.68	0.17	0.26%				
45	MC	0.77	0.41	0.33%				
46	MC	0.90	0.37	1.28%				
47	MC	0.76	0.35	1.93%				
48	MC	0.66	0.33	0.26%				
49	MC	0.78	0.49	0.30%				
50	MC	0.42	0.35	0.61%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-4 Cont'd
Item Analysis Grade 6 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.69	0.34	0.84%				
52	MC	0.64	0.42	0.40%				
53	MC	0.64	0.39	0.56%				
54	MC	0.96	0.38	0.36%				
55	MC	0.86	0.44	0.77%				
56	MC	0.49	0.36	0.77%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-5
Item Analysis Grade 7 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.74	0.55	0.13%				
2	MC	0.72	0.32	0.08%				
3	MC	0.74	0.35	0.09%				
4	MC	0.60	0.38	0.05%				
5	MC	0.71	0.27	0.07%				
6	MC	0.90	0.40	0.13%				
7	MC	0.31	0.28	0.17%				
8	MC	0.87	0.48	0.23%				
9	MC	0.64	0.26	0.75%				
10	MC	0.72	0.28	0.83%				
11	MC	0.71	0.24	1.53%				
12	MC	0.79	0.34	0.29%				
13	MC	0.34	0.18	0.48%		+		
14	MC	0.72	0.39	0.87%				
15	MC	0.61	0.33	1.06%				
16	MC	0.63	0.30	0.98%				
17	MC	0.50	0.31	1.04%				
18	MC	0.80	0.40	0.17%				
19	MC	0.63	0.39	0.16%				
20	MC	0.46	0.39	0.42%				
21	MC	0.47	0.33	0.27%				
22	MC	0.80	0.43	0.54%				
23	MC	0.47	0.37	0.65%				
24	MC	0.96	0.29	0.16%				
25	MC	0.75	0.46	0.31%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-5 Cont'd
Item Analysis Grade 7 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.61	0.47	0.63%				
27	MC	0.47	0.33	1.43%				
28	MC	0.77	0.29	1.85%				
29	MC	0.68	0.35	0.17%				
30	MC	0.72	0.34	0.24%				
31	CR	0.52	0.38	1.10%				
32	MC	0.94	0.31	0.42%				
33	MC	0.83	0.40	0.51%				
34	MC	0.46	0.21	0.53%				
35	MC	0.78	0.29	0.55%				
36	MC	0.88	0.41	0.56%				
37	MC	0.67	0.41	0.19%				
38	MC	0.62	0.26	0.25%				
39	MC	0.85	0.48	0.28%				
40	MC	0.91	0.40	1.06%				
41	MC	0.79	0.44	1.19%				
42	MC	0.59	0.38	0.21%				
43	MC	0.66	0.32	0.54%				
44	MC	0.53	0.26	0.32%				
45	MC	0.44	0.09	1.18%	+	+		
46	MC	0.85	0.45	0.24%				
47	MC	0.59	0.37	0.42%				
48	MC	0.70	0.32	0.60%				
49	MC	0.41	0.27	1.11%				
50	CR	0.36	0.41	1.00%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-5 Cont'd
Item Analysis Grade 7 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.32	0.25	0.37%		+		
52	MC	0.72	0.46	0.42%				
53	MC	0.75	0.36	0.53%				
54	MC	0.80	0.44	0.79%				
55	MC	0.94	0.41	0.40%				
56	MC	0.83	0.44	0.45%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-6
Item Analysis Grade 8 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.96	0.25	0.06%				
2	MC	0.85	0.40	0.08%				
3	MC	0.81	0.34	0.05%				
4	MC	0.88	0.28	0.22%				
5	MC	0.82	0.40	0.09%				
6	MC	0.78	0.35	0.10%				
7	MC	0.54	0.15	0.90%	+			
8	MC	0.83	0.42	1.13%				
9	MC	0.66	0.30	0.09%				
10	MC	0.69	0.35	0.79%				
11	MC	0.74	0.47	0.15%				
12	MC	0.92	0.37	0.10%				
13	MC	0.81	0.46	0.14%				
14	MC	0.92	0.31	0.19%				
15	MC	0.53	0.34	0.48%				
16	MC	0.63	0.34	0.24%				
17	MC	0.78	0.40	0.22%				
18	MC	0.75	0.27	0.39%				
19	CR	0.55	0.47	0.92%				
20	MC	0.72	0.27	0.15%				
21	MC	0.86	0.45	0.14%				
22	MC	0.60	0.36	0.19%				
23	MC	0.84	0.44	0.15%				
24	MC	0.80	0.49	0.32%				
25	MC	0.74	0.53	0.63%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-6 Cont'd
Item Analysis Grade 8 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.43	0.31	0.96%				
27	MC	0.55	0.42	0.25%				
28	MC	0.59	0.32	0.61%				
29	MC	0.60	0.18	1.11%				
30	MC	0.64	0.48	0.31%				
31	MC	0.79	0.39	0.39%				
32	MC	0.88	0.51	0.22%				
33	MC	0.47	0.20	0.33%				
34	MC	0.61	0.22	0.52%		+		
35	MC	0.66	0.35	0.26%				
36	MC	0.84	0.48	0.32%				
37	MC	0.78	0.36	0.36%				
38	MC	0.80	0.48	0.32%				
39	MC	0.73	0.38	0.34%				
40	MC	0.71	0.37	0.29%				
41	MC	0.67	0.43	0.21%				
42	MC	0.54	0.37	1.11%				
43	MC	0.80	0.51	0.25%				
44	MC	0.44	0.11	0.34%	+	+		
45	MC	0.78	0.34	0.53%				
46	MC	0.67	0.14	0.28%	+	+		
47	MC	0.64	0.29	0.32%				
48	MC	0.83	0.51	0.55%				
49	MC	0.51	0.29	0.50%				
50	MC	0.89	0.48	0.77%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-6 Cont'd
Item Analysis Grade 8 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.84	0.44	0.26%				
52	MC	0.79	0.31	0.34%				
53	MC	0.57	0.44	0.63%				
54	MC	0.48	0.36	0.41%				
55	MC	0.55	0.24	0.53%		+		
56	CR	0.56	0.47	1.62%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-7
Item Analysis Grade 10 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.87	0.35	0.04%				
2	MC	0.93	0.34	0.06%				
3	MC	0.67	0.29	0.19%				
4	MC	0.82	0.39	0.13%				
5	MC	0.48	0.37	0.24%				
6	MC	0.88	0.34	0.87%				
7	CR	0.47	0.47	4.37%				
8	MC	0.48	0.36	0.13%				
9	MC	0.78	0.50	0.19%				
10	MC	0.58	0.29	0.16%				
11	MC	0.72	0.38	0.30%				
12	MC	0.79	0.46	0.19%				
13	MC	0.78	0.43	0.22%				
14	MC	0.84	0.45	0.33%				
15	MC	0.88	0.48	0.34%				
16	MC	0.39	0.44	0.36%				
17	MC	0.61	0.35	0.22%		+		
18	MC	0.62	0.42	0.22%				
19	MC	0.83	0.43	0.20%				
20	MC	0.50	0.25	0.54%				
21	CR	0.39	0.52	3.00%				
22	MC	0.78	0.44	0.23%				
23	MC	0.66	0.48	0.31%				
24	MC	0.79	0.42	0.25%				
25	MC	0.57	0.47	0.44%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-7 Cont'd
Item Analysis Grade 10 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.85	0.30	0.25%				
27	MC	0.50	0.33	0.33%		+		
28	MC	0.78	0.34	0.55%				
29	MC	0.64	0.27	0.97%				
30	MC	0.77	0.38	0.35%				
31	MC	0.57	0.30	0.42%				
32	MC	0.66	0.23	0.44%				
33	MC	0.75	0.42	0.45%				
34	MC	0.71	0.39	0.52%				
35	MC	0.78	0.36	0.29%				
36	MC	0.84	0.26	0.29%				
37	MC	0.72	0.25	0.35%		+		
38	MC	0.75	0.51	0.63%				
39	MC	0.93	0.38	0.30%				
40	MC	0.83	0.47	0.37%				
41	MC	0.71	0.42	0.56%				
42	MC	0.66	0.42	0.78%				
43	MC	0.77	0.40	0.31%				
44	MC	0.59	0.48	0.36%				
45	MC	0.54	0.15	0.56%		+		
46	MC	0.59	0.26	0.34%				
47	MC	0.47	0.38	0.39%				
48	MC	0.50	0.39	0.67%				
49	MC	0.75	0.41	0.70%				
50	MC	0.58	0.35	0.75%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-7 Cont'd
Item Analysis Grade 10 Reading

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.82	0.39	0.83%				
52	MC	0.70	0.43	1.02%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-8
Item Analysis Grade 3 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.94	0.32	0.13%				
2	MC	0.92	0.43	0.16%				
3	MC	0.82	0.49	0.16%				
4	MC	0.73	0.35	0.17%				
5	CR	0.47	0.52	0.69%				
6	MC	0.86	0.42	0.27%				
7	MC	0.65	0.41	0.41%				
8	MC	0.89	0.36	0.38%				
9	MC	0.90	0.31	0.95%				
10	MC	0.83	0.31	0.77%				
11	MC	0.92	0.32	0.53%				
12	MC	0.63	0.33	0.56%				
13	MC	0.79	0.45	0.93%				
14	MC	0.96	0.26	1.13%				
15	MC	0.91	0.23	0.14%				
16	MC	0.75	0.42	0.22%				
17	MC	0.58	0.26	0.24%		+		
18	MC	0.91	0.37	0.21%				
19	MC	0.73	0.34	0.37%				
20A	CR	0.49	0.39	8.90%			+	
20B	CR	0.26	0.43	3.03%				+
21	MC	0.87	0.46	0.42%				
22	MC	0.93	0.25	0.43%				
23	MC	0.67	0.40	0.44%				
24	MC	0.80	0.22	1.11%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-8 Cont'd
Item Analysis Grade 3 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
25	MC	0.70	0.44	0.55%				
26	MC	0.89	0.38	1.07%				
27A	CR	0.65	0.51	2.21%				
27B	CR	0.32	0.45	4.21%				
28	MC	0.75	0.40	2.00%				
29	MC	0.53	0.35	2.63%				
30	MC	0.77	0.40	2.59%				
31	MC	0.87	0.44	0.23%				
32	MC	0.75	0.46	0.72%				
33	MC	0.91	0.40	0.46%				
34	MC	0.61	0.33	0.32%				
35	MC	0.84	0.34	0.43%				
36	MC	0.32	0.35	2.22%				
37	MC	0.93	0.18	0.45%				
38	MC	0.92	0.23	0.53%				
39A	CR	0.41	0.49	5.68%			+	
39B	CR	0.35	0.53	1.98%				
40	MC	0.92	0.31	0.31%				
41	MC	0.94	0.34	0.38%				
42	MC	0.65	0.38	0.49%				
43	MC	0.80	0.41	0.31%				
44	MC	0.92	0.31	0.37%				
45	MC	0.80	0.45	0.52%				
46	MC	0.70	0.28	0.56%				
47	MC	0.70	0.43	0.69%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-8 Cont'd
Item Analysis Grade 3 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
48	MC	0.74	0.42	0.78%				
49	MC	0.73	0.42	0.98%				
50	MC	0.93	0.31	0.82%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-9
Item Analysis Grade 4 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.97	0.15	0.08%	+			
2	MC	0.90	0.22	0.09%				
3	MC	0.90	0.28	0.19%				
4	MC	0.90	0.42	0.26%				
5	MC	0.87	0.43	0.18%				
6	MC	0.80	0.43	0.54%				
7	MC	0.75	0.38	0.64%				
8	MC	0.91	0.37	0.17%				
9	MC	0.63	0.47	0.35%				
10	MC	0.97	0.31	0.35%				
11A	CR	0.21	0.32	0.78%				+
11B	CR	0.33	0.38	2.77%				
12	MC	0.61	0.54	0.90%				
13	MC	0.74	0.51	0.91%				
14	MC	0.83	0.44	0.94%				
15	MC	0.93	0.33	0.15%				
16	MC	0.69	0.41	0.28%				
17	MC	0.69	0.34	0.31%				
18	MC	0.84	0.38	0.50%				
19	MC	0.74	0.28	0.18%				
20	MC	0.98	0.15	0.26%				
21	MC	0.72	0.42	0.53%				
22	CR	0.81	0.23	0.67%				
23	MC	0.72	0.30	0.31%				
24	MC	0.84	0.33	0.35%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-9 Cont'd
Item Analysis Grade 4 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
25A	CR	0.47	0.54	0.52%				
25B	CR	0.32	0.43	2.19%				
26	MC	0.69	0.31	0.42%				
27	MC	0.91	0.31	0.50%				
28	MC	0.84	0.39	0.57%				
29	MC	0.80	0.41	0.72%				
30	MC	0.50	0.40	0.64%				
31	MC	0.92	0.30	0.29%				
32	MC	0.94	0.32	1.30%				
33	MC	0.95	0.28	0.35%				
34	MC	0.77	0.37	0.36%				
35	MC	0.87	0.27	0.27%				
36	MC	0.98	0.14	1.35%	+			
37	MC	0.54	0.34	0.31%				
38	MC	0.76	0.50	3.29%				
39A	CR	0.59	0.42	0.59%				
39B	CR	0.34	0.45	1.74%				
40	MC	0.57	0.42	0.35%				
41	MC	0.90	0.21	0.46%				
42	MC	0.93	0.22	0.38%				
43	MC	0.81	0.36	0.45%				
44	MC	0.49	0.26	0.57%				
45	MC	0.62	0.45	0.55%				
46	MC	0.64	0.42	0.95%				
47	MC	0.58	0.44	0.73%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-9 Cont'd
Item Analysis Grade 4 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
48	MC	0.78	0.28	0.61%				
49	MC	0.90	0.42	0.70%				
50	MC	0.89	0.33	0.82%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-10
Item Analysis Grade 5 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.89	0.30	0.06%				
2	MC	0.86	0.41	0.15%				
3	MC	0.70	0.37	0.15%				
4	MC	0.60	0.48	0.47%				
5	MC	0.94	0.25	0.14%				
6	MC	0.75	0.55	0.20%				
7	MC	0.46	0.59	0.32%				
8	MC	0.74	0.43	0.33%				
9	MC	0.53	0.29	0.53%				
10	MC	0.67	0.55	0.54%				
11	MC	0.63	0.41	0.47%				
12	CR	0.67	0.59	1.08%				
13	MC	0.51	0.42	0.74%				
14	MC	0.66	0.43	0.95%				
15	MC	0.90	0.30	0.15%				
16	MC	0.61	0.29	0.25%				
17	MC	0.77	0.36	0.17%				
18	MC	0.75	0.37	0.47%				
19	MC	0.66	0.23	0.16%				
20A	CR	0.86	0.24	0.47%				
20B	CR	0.30	0.45	1.37%				+
21	MC	0.35	0.38	0.25%				
22	MC	0.86	0.32	0.22%				
23	MC	0.70	0.55	0.27%				
24	MC	0.96	0.24	0.26%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-10 Cont'd
Item Analysis Grade 5 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
25	MC	0.73	0.43	0.19%				
26	MC	0.52	0.42	0.44%				
27A	CR	0.61	0.46	0.69%				
27B	CR	0.72	0.49	1.26%				
28	MC	0.81	0.34	0.43%				
29	MC	0.88	0.31	0.35%				
30	MC	0.46	0.40	0.68%				
31	MC	0.69	0.35	0.46%				
32	MC	0.93	0.27	0.60%				
33	MC	0.96	0.29	0.65%				
34	MC	0.65	0.47	0.58%				
35	MC	0.93	0.36	0.67%				
36	MC	0.68	0.35	0.22%				
37	MC	0.54	0.36	0.35%				
38	MC	0.41	0.24	0.33%		+		
39	MC	0.91	0.29	0.29%				
40	MC	0.89	0.24	0.30%				
41	MC	0.35	0.13	0.66%	+	+		
42	MC	0.90	0.20	0.34%				
43	MC	0.82	0.31	0.44%				
44	MC	0.74	0.20	0.50%				
45	MC	0.88	0.38	0.85%				
46A	CR	0.38	0.48	0.54%				
46B	CR	0.34	0.47	0.85%				
47	MC	0.88	0.30	0.35%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-10 Cont'd
Item Analysis Grade 5 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
48	MC	0.85	0.48	0.53%				
49	MC	0.62	0.30	0.58%				
50	MC	0.90	0.37	0.39%				
51	MC	0.54	0.44	0.94%				
52	MC	0.84	0.26	0.31%				
53	MC	0.78	0.34	0.24%				
54	MC	0.99	0.19	0.24%				
55	MC	0.81	0.46	0.30%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-11
Item Analysis Grade 6 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.79	0.43	0.25%				
2	MC	0.87	0.42	0.10%				
3	MC	0.70	0.41	0.31%				
4	MC	0.48	0.22	0.25%		+		
5	MC	0.88	0.37	0.33%				
6	MC	0.57	0.43	0.24%				
7A	CR	0.54	0.36	3.50%				
7B	CR	0.44	0.44	4.20%				
8	MC	0.68	0.39	0.64%				
9	MC	0.89	0.42	0.53%				
10	MC	0.67	0.38	1.03%				
11	MC	0.87	0.38	1.02%				
12	MC	0.53	0.35	1.34%				
13	MC	0.53	0.37	1.74%				
14	MC	0.88	0.49	1.58%				
15	MC	0.84	0.47	1.69%				
16	MC	0.52	0.42	0.41%				
17	MC	0.72	0.36	0.16%				
18	MC	0.65	0.40	0.21%				
19	MC	0.92	0.33	0.17%				
20	MC	0.90	0.34	0.26%				
21	MC	0.49	0.40	0.37%				
22A	CR	0.92	0.33	0.41%				
22B	CR	0.26	0.43	0.96%				+
23	MC	0.67	0.36	0.33%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-11 Cont'd
Item Analysis Grade 6 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
24	MC	0.65	0.35	0.33%				
25	MC	0.42	0.38	0.37%				
26	MC	0.83	0.38	0.44%				
27	MC	0.73	0.43	0.42%				
28	MC	0.63	0.51	0.50%				
29	MC	0.80	0.35	0.67%				
30	MC	0.46	0.39	0.28%				
31	MC	0.83	0.39	0.23%				
32	MC	0.59	0.47	0.26%				
33	MC	0.96	0.18	0.21%				
34	MC	0.77	0.36	0.29%				
35	MC	0.97	0.16	0.27%				
36	MC	0.70	0.48	0.40%				
37	MC	0.86	0.22	0.47%				
38	MC	0.93	0.27	0.44%				
39A	CR	0.45	0.59	0.86%				
39B	CR	0.42	0.58	1.65%				
40	MC	0.58	0.52	0.43%				
41	MC	0.69	0.42	0.39%				
42	MC	0.96	0.17	0.24%				
43	MC	0.77	0.35	0.20%				
44	MC	0.71	0.54	0.26%				
45	MC	0.80	0.43	0.26%				
46	MC	0.74	0.51	0.43%				
47	MC	0.84	0.42	0.21%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-11 Cont'd
Item Analysis Grade 6 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
48	MC	0.80	0.45	0.32%				
49	MC	0.61	0.43	0.31%				
50	MC	0.58	0.45	0.43%				
51	CR	0.44	0.51	0.61%				
52	MC	0.45	0.25	0.36%				
53	MC	0.79	0.25	0.44%				
54	MC	0.60	0.35	0.47%				
55	MC	0.60	0.46	0.52%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-12
Item Analysis Grade 7 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.75	0.46	0.16%				
2	MC	0.92	0.38	0.05%				
3	MC	0.62	0.50	0.12%				
4	MC	0.68	0.21	0.12%		+		
5	MC	0.78	0.41	0.13%				
6	MC	0.75	0.39	0.14%				
7A	CR	0.72	0.51	0.65%				
7B	CR	0.65	0.56	1.37%				
8	MC	0.74	0.33	0.28%				
9	MC	0.60	0.48	0.41%				
10	MC	0.68	0.50	0.68%				
11	MC	0.62	0.58	0.86%				
12	MC	0.58	0.26	0.62%		+		
13	MC	0.59	0.57	0.78%				
14	MC	0.89	0.33	0.59%				
15	MC	0.79	0.41	0.72%				
16	MC	0.61	0.23	0.17%				
17	MC	0.59	0.27	0.21%				
18	MC	0.87	0.48	0.14%				
19A	CR	0.54	0.37	1.03%				
19B	CR	0.32	0.48	2.09%				
20	MC	0.98	0.20	0.18%				
21	MC	0.51	0.28	0.31%				
22	MC	0.74	0.32	0.34%				
23	MC	0.57	0.48	0.34%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-12 Cont'd
Item Analysis Grade 7 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
24	MC	0.87	0.47	0.18%				
25	MC	0.41	0.42	0.31%				
26	MC	0.92	0.18	0.23%				
27	MC	0.73	0.42	0.30%				
28	MC	0.85	0.48	0.19%				
29	MC	0.87	0.22	0.19%				
30	MC	0.87	0.32	0.13%				
31	MC	0.54	0.38	0.46%				
32	MC	0.67	0.48	0.30%				
33	MC	0.61	0.36	0.44%				
34	MC	0.41	0.32	0.43%				
35	MC	0.71	0.53	0.37%				
36	MC	0.69	0.37	0.33%				
37	MC	0.92	0.32	0.36%				
38A	CR	0.59	0.35	9.04%			+	
38B	CR	0.36	0.38	3.30%				
39	MC	0.67	0.43	0.24%				
40	MC	0.90	0.33	0.40%				
41	MC	0.76	0.43	0.24%				
42	MC	0.92	0.34	0.34%				
43	MC	0.61	0.45	0.22%				
44	MC	0.98	0.23	0.24%				
45	MC	0.39	0.41	0.27%				
46	MC	0.90	0.39	0.23%				
47	MC	0.84	0.30	0.43%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-12 Cont'd
Item Analysis Grade 7 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
48	MC	0.88	0.31	0.13%				
49	MC	0.53	0.52	0.30%				
50	MC	0.75	0.60	0.66%				
51	MC	0.53	0.37	0.48%				
52	CR	0.36	0.57	0.82%				
53	MC	0.86	0.32	0.25%				
54	MC	0.54	0.41	0.35%		+		
55	MC	0.44	0.17	0.38%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-13
Item Analysis Grade 8 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.45	0.21	0.11%				
2	MC	0.53	0.48	0.11%				
3	MC	0.74	0.47	0.12%				
4	MC	0.58	0.53	0.27%				
5	MC	0.48	0.32	0.32%		+		
6	MC	0.88	0.40	0.22%				
7	MC	0.67	0.50	0.18%				
8	MC	0.53	0.35	0.30%				
9	MC	0.53	0.41	0.25%				
10	CR	0.21	0.50	2.49%				+
11	MC	0.87	0.41	0.33%				
12	MC	0.43	0.36	0.82%				
13	MC	0.69	0.51	0.73%				
14	MC	0.90	0.37	0.68%				
15	MC	0.59	0.30	0.90%				
16	MC	0.75	0.40	0.23%				
17	MC	0.43	0.38	0.37%				
18	MC	0.69	0.26	0.30%				
19A	CR	0.71	0.43	0.98%				
19B	CR	0.60	0.49	2.90%				
20	MC	0.78	0.36	0.29%				
21	MC	0.35	0.36	0.27%				
22	MC	0.39	0.39	0.43%				
23	MC	0.83	0.41	0.39%				
24	MC	0.96	0.23	0.25%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-13 Cont'd
Item Analysis Grade 8 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
25	MC	0.52	0.39	0.47%				
26	MC	0.64	0.35	0.69%				
27	MC	0.60	0.42	0.54%				
28	MC	0.74	0.37	0.77%				
29	MC	0.43	0.39	0.24%				
30	MC	0.53	0.32	0.20%				
31	MC	0.54	0.36	0.30%				
32	MC	0.54	0.37	0.40%				
33A	CR	0.31	0.48	6.48%			+	
33B	CR	0.35	0.60	5.07%			+	
34	MC	0.54	0.36	0.47%				
35	MC	0.65	0.32	0.73%				
36	MC	0.83	0.45	0.73%				
37	MC	0.39	0.45	0.40%				
38	MC	0.65	0.34	0.61%				
39	MC	0.74	0.31	1.08%				
40	MC	0.66	0.17	1.09%		+		
41	MC	0.88	0.37	0.69%				
42	MC	0.67	0.33	0.77%				
43	MC	0.76	0.31	0.36%				
44	MC	0.79	0.37	0.30%				
45	MC	0.68	0.50	0.28%		+		
46	MC	0.41	0.31	0.41%				
47	MC	0.83	0.28	0.46%				
48	MC	0.48	0.47	0.40%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-13 Cont'd
Item Analysis Grade 8 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
49A	CR	0.27	0.44	1.46%				+
49B	CR	0.21	0.44	2.26%				+
50	MC	0.78	0.37	0.37%				
51	MC	0.25	0.30	0.47%				+
52	MC	0.60	0.54	0.49%				
53	MC	0.60	0.51	0.54%				
54	MC	0.86	0.48	0.39%				
55	MC	0.68	0.25	0.51%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-14
Item Analysis Grade 10 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.85	0.43	0.05%				
2	MC	0.69	0.39	0.07%				
3	MC	0.51	0.30	0.26%				
4	MC	0.58	0.49	0.23%				
5	MC	0.54	0.39	0.14%				
6	MC	0.64	0.44	0.14%				
7	MC	0.41	0.43	0.14%				
8	MC	0.70	0.48	0.13%				
9	MC	0.53	0.41	0.12%				
10	MC	0.50	0.38	0.14%				
11	MC	0.54	0.47	0.10%				
12	MC	0.74	0.43	0.38%				
13	MC	0.82	0.35	0.15%				
14	MC	0.49	0.43	0.24%				
15	MC	0.64	0.41	0.29%				
16	MC	0.82	0.49	0.41%				
17	MC	0.72	0.60	0.34%				
18	MC	0.74	0.46	0.32%				
19	MC	0.50	0.40	0.66%				
20	MC	0.59	0.49	0.26%				
21	MC	0.75	0.46	0.30%				
22	MC	0.45	0.55	0.38%				
23	CR	0.47	0.61	4.34%				
24	MC	0.72	0.38	0.90%				
25	MC	0.75	0.42	1.11%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-14 Cont'd
Item Analysis Grade 10 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.58	0.57	1.36%				
27	MC	0.54	0.41	1.68%				
28	MC	0.51	0.48	0.36%				
29	MC	0.85	0.42	0.28%				
30	MC	0.60	0.40	0.41%				
31	MC	0.63	0.23	0.33%				
32	MC	0.48	0.41	0.45%				
33	CR	0.34	0.58	9.98%			+	
34	MC	0.75	0.41	0.49%				
35	MC	0.95	0.22	0.45%				
36	CR	0.51	0.65	7.17%			+	
37	MC	0.68	0.31	0.38%				
38	MC	0.59	0.41	0.53%				
39	MC	0.68	0.49	0.52%				
40	MC	0.45	0.44	0.40%				
41	MC	0.56	0.65	0.46%				
42	MC	0.60	0.47	0.41%				
43	MC	0.50	0.53	0.56%				
44	MC	0.77	0.46	0.54%				
45	MC	0.58	0.43	0.46%		+		
46	CR	0.48	0.52	3.62%				
47	MC	0.57	0.46	0.54%				
48	MC	0.88	0.37	0.50%				
49	MC	0.74	0.39	0.51%				
50	MC	0.79	0.46	0.68%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-14 Cont'd
Item Analysis Grade 10 Mathematics

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
51	MC	0.40	0.40	0.61%				
52	MC	0.44	0.48	0.76%				
53	MC	0.52	0.36	0.87%				
54	MC	0.73	0.55	0.75%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-15
Item Analysis Grade 4 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.41	0.29	0.11%				
2	MC	0.93	0.29	0.11%				
3	MC	0.79	0.29	0.46%				
4	MC	0.93	0.28	0.38%				
5	MC	0.40	0.13	0.52%	+	+		
6	MC	0.84	0.36	0.25%				
7	MC	0.59	0.37	0.28%				
8	MC	0.85	0.26	0.32%				
9	MC	0.85	0.41	0.35%				
10	MC	0.49	0.32	0.87%				
11	MC	0.91	0.36	0.35%				
12	MC	0.95	0.31	0.35%				
13	MC	0.78	0.42	0.32%				
14	MC	0.66	0.40	0.56%				
15	MC	0.54	0.24	0.30%				
16	MC	0.67	0.44	1.65%				
17	MC	0.65	0.46	0.41%				
18	MC	0.69	0.44	2.44%				
19	MC	0.45	0.31	1.01%				
20	MC	0.70	0.40	0.56%				
21	MC	0.73	0.42	0.69%				
22	MC	0.61	0.38	1.60%				
23	MC	0.53	0.32	1.18%				
24	MC	0.58	0.42	0.89%				
25	MC	0.44	0.20	1.17%		+		

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-15 Cont'd
Item Analysis Grade 4 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.56	0.38	1.60%				
27	MC	0.68	0.40	1.44%				
28	MC	0.35	0.26	2.19%		+		
29	MC	0.77	0.41	2.31%				
30	MC	0.43	0.08	2.61%	+	+		

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-16
Item Analysis Grade 8 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.64	0.12	0.08%	+			
2	MC	0.91	0.11	0.07%	+			
3	MC	0.68	0.30	0.49%				
4	MC	0.83	0.37	0.03%				
5	MC	0.72	0.38	0.19%				
6	MC	0.52	0.30	0.24%				
7	MC	0.89	0.37	0.23%				
8	MC	0.66	0.45	0.18%				
9	MC	0.81	0.41	0.31%				
10	MC	0.66	0.34	0.15%				
11	MC	0.84	0.35	0.14%				
12	MC	0.71	0.44	0.66%				
13	MC	0.51	0.35	0.15%				
14	MC	0.81	0.47	0.19%				
15	MC	0.87	0.40	0.21%				
16	MC	0.88	0.35	0.24%				
17	MC	0.65	0.42	0.30%				
18	MC	0.84	0.46	0.90%				
19	MC	0.80	0.41	0.37%				
20	MC	0.59	0.53	2.96%				
21	MC	0.91	0.47	0.52%				
22	MC	0.60	0.41	1.34%				
23	MC	0.68	0.44	1.13%				
24	MC	0.78	0.38	0.66%				
25	MC	0.71	0.40	1.94%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-16 Cont'd
Item Analysis Grade 8 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.45	0.36	1.01%				
27	MC	0.74	0.34	1.59%				
28 **	MC	.	.					
29	MC	0.67	0.35	1.51%				
30	MC	0.66	0.43	1.75%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.
** Item dropped from scoring.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-17
Item Analysis Grade 10 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.72	0.37	0.07%				
1A	CR	0.58	0.54	1.35%				
1B	CR	0.69	0.34	1.35%				
2	MC	0.83	0.41	0.10%				
3	MC	0.61	0.26	0.10%				
4	MC	0.67	0.27	0.92%				
5	MC	0.83	0.32	0.18%				
6	MC	0.71	0.35	0.12%				
7	MC	0.42	0.20	0.28%				
8	MC	0.82	0.34	0.14%				
9	MC	0.60	0.31	0.14%				
10	MC	0.68	0.37	0.18%				
11	MC	0.78	0.40	1.43%				
12	MC	0.65	0.21	0.21%		+		
13	MC	0.47	0.16	0.23%		+		
14	MC	0.80	0.46	0.11%				
15	MC	0.61	0.41	0.13%				
16	MC	0.70	0.32	0.20%				
17	MC	0.88	0.38	0.59%				
18	MC	0.70	0.26	0.47%				
19	MC	0.65	0.45	0.29%				
20	MC	0.54	0.37	0.61%				
21	MC	0.39	0.15	0.44%		+		
22	MC	0.67	0.43	0.49%				
23	MC	0.49	0.33	0.47%		+		

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-17 Cont'd
Item Analysis Grade 10 Language Arts

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
24	MC	0.32	0.20	0.55%				
25	MC	0.77	0.53	0.88%				
26	MC	0.75	0.41	0.62%				
27	MC	0.28	0.18	0.78%		+		+
28	MC	0.86	0.39	0.83%				
29	MC	0.66	0.37	1.03%				
30	MC	0.75	0.47	1.05%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.
** Writing prompt items are included here. The Writing raw score contributes to the scale score for Language Arts in grade 10.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-18
Item Analysis Grade 4 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.50	0.37	0.04%				
2	MC	0.91	0.34	0.06%				
3	MC	0.95	0.33	0.13%				
4	MC	0.52	0.40	0.09%				
5	MC	0.86	0.21	0.10%				
6	MC	0.58	0.20	0.26%		+		
7	MC	0.77	0.40	0.51%				
8	MC	0.89	0.37	0.22%				
9	MC	0.82	0.37	0.83%				
10	MC	0.76	0.26	0.28%				
11	MC	0.81	0.46	0.50%				
12	MC	0.60	0.44	0.59%				
13	MC	0.95	0.29	0.23%				
14	MC	0.96	0.28	1.29%				
15	MC	0.83	0.38	0.28%				
16	MC	0.50	0.27	1.28%				
17	MC	0.81	0.32	0.39%				
18	MC	0.92	0.33	0.18%				
19	MC	0.97	0.29	0.67%				
20	MC	0.68	0.45	1.01%				
21	MC	0.55	0.44	0.45%				
22	MC	0.72	0.46	0.31%				
23	MC	0.86	0.41	0.44%				
24	MC	0.94	0.29	0.28%				
25	MC	0.93	0.37	0.54%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-18 Cont'd
Item Analysis Grade 4 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.75	0.47	1.96%				
27	MC	0.94	0.39	0.30%				
28	MC	0.78	0.40	0.49%				
29	MC	0.65	0.40	0.32%				
30	MC	0.68	0.40	0.55%				
31	MC	0.65	0.38	0.63%				
32	MC	0.70	0.36	0.46%				
33	MC	0.52	0.28	0.70%				
34	MC	0.72	0.39	1.52%				
35	MC	0.88	0.42	0.48%				
36	MC	0.61	0.46	1.08%				
37	MC	0.90	0.46	0.67%				
38	MC	0.72	0.46	0.78%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-19
Item Analysis Grade 8 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.96	0.30	0.01%				
2	MC	0.95	0.29	0.36%				
3	MC	0.69	0.37	0.26%				
4	MC	0.42	0.19	0.25%		+		
5	MC	0.81	0.31	0.26%				
6	MC	0.97	0.32	0.38%				
7	MC	0.72	0.23	0.44%				
8	MC	0.93	0.33	0.15%				
9	MC	0.59	0.38	0.37%				
10	MC	0.84	0.44	0.83%				
11	MC	0.74	0.41	1.53%				
12	MC	0.76	0.38	0.15%				
13	MC	0.61	0.31	0.53%				
14	MC	0.79	0.46	0.73%				
15	MC	0.85	0.39	0.68%				
16	MC	0.77	0.53	0.25%				
17	MC	0.69	0.45	1.27%				
18	MC	0.80	0.36	0.26%				
19	MC	0.52	0.37	0.41%				
20	MC	0.53	0.35	0.22%				
21	MC	0.58	0.40	1.04%				
22	MC	0.94	0.35	0.24%				
23	MC	0.77	0.35	0.30%				
24	MC	0.60	0.39	0.32%				
25	MC	0.82	0.35	0.58%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-19 Cont'd
Item Analysis Grade 8 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.64	0.38	3.26%				
27	MC	0.64	0.41	0.35%				
28	MC	0.66	0.51	0.49%				
29	MC	0.62	0.39	1.01%				
30	MC	0.48	0.27	0.46%				
31	MC	0.67	0.33	1.32%				
32	MC	0.82	0.48	0.44%				
33	MC	0.77	0.43	0.98%				
34	MC	0.84	0.44	0.96%				
35	MC	0.63	0.39	0.69%				
36	MC	0.52	0.31	0.56%				
37	MC	0.53	0.21	0.83%				
38	MC	0.56	0.32	0.74%				
39	MC	0.49	0.38	0.94%				
40	MC	0.65	0.37	0.99%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-20
Item Analysis Grade 10 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.81	0.43	0.08%				
2	MC	0.98	0.26	0.68%				
3	MC	0.68	0.32	0.09%				
4	MC	0.52	0.34	0.22%				
5	MC	0.67	0.45	0.27%				
6	MC	0.44	0.37	0.11%				
7	MC	0.55	0.36	0.27%				
8	MC	0.57	0.24	0.40%				
9	MC	0.61	0.40	0.45%				
10	MC	0.83	0.34	0.49%				
11	MC	0.55	0.38	0.57%				
12	MC	0.48	0.26	0.18%				
13	MC	0.53	0.26	0.20%				
14	MC	0.78	0.35	0.24%				
15	MC	0.33	0.30	0.17%				
16	MC	0.70	0.46	0.64%				
17	MC	0.56	0.36	0.26%				
18	MC	0.42	0.21	0.33%				
19	MC	0.90	0.37	0.16%				
20	MC	0.88	0.32	0.35%				
21	MC	0.73	0.39	0.86%				
22	MC	0.79	0.45	0.26%				
23	MC	0.48	0.42	0.75%		+		
24	MC	0.70	0.47	0.38%				
25	MC	0.98	0.20	0.22%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-20 Cont'd
Item Analysis Grade 10 Social Studies

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.72	0.43	0.27%				
27	MC	0.78	0.42	1.09%				
28	MC	0.81	0.43	0.31%				
29	MC	0.52	0.19	0.28%				
30	MC	0.65	0.39	0.36%				
31	MC	0.82	0.41	0.27%				
32	MC	0.63	0.43	1.29%				
33	MC	0.60	0.40	0.56%				
34	MC	0.80	0.45	0.29%				
35	MC	0.64	0.39	1.35%				
36	MC	0.77	0.41	0.59%				
37	MC	0.60	0.35	0.28%				
38	MC	0.78	0.48	0.31%				
39	MC	0.85	0.40	0.36%				
40	MC	0.19	0.12	0.31%	+	+		+
41 **	MC	.	.					
42	MC	0.66	0.34	0.41%				
43	MC	0.39	0.41	1.01%		+		
44	MC	0.68	0.51	0.46%				
45	MC	0.51	0.50	0.41%				
46	MC	0.81	0.48	0.48%				
47	MC	0.86	0.43	0.65%				
48	MC	0.66	0.44	1.06%				
49	MC	0.49	0.44	0.52%				
50	MC	0.89	0.46	0.83%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.
** Item dropped from scoring.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-21
Item Analysis Grade 4 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.98	0.21	0.06%				
2	MC	0.87	0.23	0.05%				
3	MC	0.88	0.45	0.43%				
4	MC	0.82	0.34	0.22%				
5	MC	0.93	0.14	0.03%	+			
6	MC	0.83	0.28	0.51%				
7	MC	0.94	0.35	0.06%				
8	MC	0.75	0.38	0.16%				
9	MC	0.82	0.32	0.22%				
10	MC	0.94	0.21	0.25%				
11	MC	0.75	0.37	0.19%				
12	MC	0.79	0.31	0.49%				
13	MC	0.67	0.31	0.28%				
14 **	MC	.	.					
15	MC	0.89	0.47	2.66%				
16	MC	0.85	0.39	2.80%				
17	MC	0.82	0.43	0.29%				
18	MC	0.57	0.37	0.42%				
19	MC	0.52	0.43	0.46%				
20	MC	0.66	0.36	0.40%				
21	MC	0.65	0.38	0.62%				
22	MC	0.83	0.49	0.69%				
23	MC	0.88	0.33	0.34%				
24	MC	0.45	0.14	0.46%	+	+		
25	MC	0.86	0.29	0.25%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.
**Item dropped from scoring.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-21 Cont'd
Item Analysis Grade 4 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.54	0.35	0.27%				
27	MC	0.69	0.34	0.39%				
28	MC	0.92	0.40	0.57%				
29	MC	0.73	0.30	0.70%				
30	MC	0.79	0.46	0.52%				
31	MC	0.41	0.31	1.49%				
32	MC	0.35	0.13	0.53%	+	+		
33	MC	0.71	0.42	1.04%				
34	MC	0.96	0.28	1.32%				
35	MC	0.78	0.49	1.27%				
36	MC	0.70	0.36	0.69%				
37	MC	0.53	0.38	0.74%				
38	MC	0.51	0.29	0.84%				
39	MC	0.48	0.40	0.97%				
40	MC	0.50	0.34	1.23%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-22
Item Analysis Grade 8 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.90	0.33	0.12%				
2	MC	0.98	0.25	0.07%				
3	MC	0.97	0.22	0.09%				
4	MC	0.88	0.26	0.62%				
5	MC	0.92	0.35	0.19%				
6	MC	0.55	0.26	0.29%				
7	MC	0.94	0.29	0.20%				
8	MC	0.77	0.33	0.36%				
9	MC	0.61	0.37	0.62%				
10	MC	0.89	0.36	0.31%				
11	MC	0.96	0.30	0.76%				
12	MC	0.96	0.28	0.72%				
13	MC	0.77	0.32	0.21%				
14	MC	0.69	0.37	0.26%				
15	MC	0.72	0.37	0.18%				
16	MC	0.77	0.41	0.47%				
17	MC	0.77	0.32	0.29%				
18	MC	0.89	0.24	0.18%				
19	MC	0.72	0.43	0.38%				
20	MC	0.78	0.41	0.40%				
21	MC	0.74	0.35	0.34%				
22	MC	0.74	0.39	0.25%				
23	MC	0.86	0.37	0.43%				
24	MC	0.71	0.46	0.58%				
25	MC	0.84	0.47	0.43%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-22 Cont'd
Item Analysis Grade 8 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.59	0.33	0.30%				
27	MC	0.66	0.42	4.41%				
28	MC	0.74	0.46	0.25%				
29	MC	0.79	0.39	0.30%				
30	MC	0.75	0.27	0.98%				
31	MC	0.46	0.30	1.48%				
32	MC	0.62	0.40	0.35%				
33	MC	0.84	0.34	0.49%				
34	MC	0.70	0.41	0.48%				
35	MC	0.57	0.29	0.39%				
36	MC	0.48	0.26	4.92%				
37	MC	0.69	0.49	0.71%				
38	MC	0.62	0.33	0.91%				
39	MC	0.63	0.21	0.70%				
40	MC	0.62	0.29	0.59%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-23
Item Analysis Grade 10 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
1	MC	0.77	0.41	0.05%				
2	MC	0.90	0.27	0.05%				
3	MC	0.45	0.17	0.10%		+		
4	MC	0.42	0.13	3.06%	+			
5	MC	0.71	0.37	0.12%				
6	MC	0.86	0.37	0.14%				
7	MC	0.73	0.40	0.20%				
8	MC	0.84	0.36	0.15%				
9	MC	0.90	0.33	0.16%				
10	MC	0.71	0.33	0.37%				
11	MC	0.71	0.41	0.57%				
12	MC	0.34	0.22	0.20%				
13	MC	0.87	0.40	0.27%				
14	MC	0.75	0.26	0.14%				
15	MC	0.52	0.35	0.37%				
16	MC	0.52	0.28	0.12%				
17	MC	0.67	0.33	0.17%				
18	MC	0.58	0.27	0.24%				
19	MC	0.79	0.37	0.20%				
20	MC	0.73	0.39	0.16%				
21	MC	0.87	0.23	0.13%				
22	MC	0.44	0.39	0.49%				
23	MC	0.48	0.42	0.28%				
24	MC	0.44	0.26	0.50%				
25	MC	0.58	0.44	0.17%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-23 Cont'd
Item Analysis Grade 10 Science

	Item Statistic Fields				Flag			
Test Book Item	**Item Type**	***p*-value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distractor**	**Flag Omit**	**Flag *p*-value**
26	MC	0.46	0.32	0.26%				
27	MC	0.68	0.40	0.24%				
28	MC	0.68	0.49	0.31%				
29	MC	0.32	0.22	0.28%				
30	MC	0.71	0.46	0.32%				
31	MC	0.40	0.11	0.53%	+			
32	MC	0.68	0.18	0.28%				
33	MC	0.36	0.30	0.38%				
34	MC	0.66	0.31	0.38%				
35	MC	0.73	0.45	0.46%				
36	MC	0.77	0.40	0.57%				
37	MC	0.40	0.35	0.79%		+		
38	MC	0.67	0.33	0.31%				
39	MC	0.71	0.46	1.13%				
40	MC	0.71	0.36	0.42%				
41	MC	0.49	0.42	0.54%				
42	MC	0.62	0.37	0.31%				
43	MC	0.46	0.48	0.44%				
44	MC	0.75	0.41	0.43%				
45	MC	0.36	0.25	1.50%		+		
46	MC	0.58	0.43	0.48%				
47	MC	0.64	0.39	0.59%				
48	MC	0.54	0.34	0.47%				
49	MC	0.89	0.24	1.38%				
50	MC	0.72	0.44	1.50%				

* Note: The correlation is flagged when it falls below 0.15, the distractor is flagged when it has a positive correlation with the correct answer, the omit rate is flagged when it is above 5%, the *p*-value is flagged when it is below 0.30.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-24
The Number of Items Flagged

Content	Grade	OP Items			
		Flag Corr	Flag Distractor	Flag Omit	Flag *p*-value
RD	3				
	4		2		
	5	1	4		
	6	2	3		
	7	1	3		
	8	3	4		
	10		4		
MA	3		1	2	1
	4	2			1
	5	1	2		1
	6		1		1
	7		3	1	
	8		3	2	4
	10		1	2	
LA	4	2	4		
	8	2			
	10		5		1
SC	4	3	2		
	8				
	10	2	3		
SS	4		1		
	8		1		
	10	1	3		1

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-25
Raw Score Descriptive Statistics based on Census Data

Content	Grade	N Count	Mean	Test Difficulty	SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
Reading	3	60163	39.13	0.69	11.51	-0.68	-0.47	1	57	57	0.93	2.97
	4	59602	38.54	0.64	11.02	-0.61	-0.40	1	60	60	0.92	3.19
	5	59214	39.26	0.65	10.64	-0.64	-0.32	1	60	60	0.91	3.16
	6	60351	38.56	0.64	10.03	-0.60	-0.26	2	60	60	0.90	3.20
	7	61187	39.14	0.65	9.84	-0.63	-0.16	0	59	60	0.89	3.20
	8	60676	41.69	0.69	10.02	-0.76	0.00	1	60	60	0.90	3.14
	10	62736	37.35	0.67	10.13	-0.58	-0.32	0	56	56	0.91	3.10
Mathematics	3	60355	40.54	0.71	9.57	-0.66	-0.03	2	57	57	0.90	2.96
	4	59770	41.07	0.72	8.89	-0.70	0.11	3	57	57	0.90	2.88
	5	59263	43.12	0.70	10.63	-0.53	-0.33	1	62	62	0.91	3.22
	6	60409	41.48	0.67	11.32	-0.43	-0.58	1	62	62	0.92	3.21
	7	61268	41.39	0.67	11.25	-0.44	-0.55	1	62	62	0.92	3.19
	8	60718	36.08	0.58	11.81	-0.10	-0.74	0	62	62	0.92	3.36
	10	62664	34.80	0.60	12.91	-0.10	-1.05	0	58	58	0.94	3.27
Language Arts	4	59592	19.58	0.65	5.36	-0.39	-0.57	1	30	30	0.82	2.26
	8	60430	20.89	0.72	5.51	-0.66	-0.26	0	29	29	0.85	2.12
	10	62465	24.96	0.64	6.30	-0.64	-0.03	0	39	39	0.84	2.55
Social Studies	4	59632	28.96	0.76	6.34	-0.94	0.48	1	38	38	0.87	2.27
	8	60602	28.00	0.70	7.17	-0.52	-0.41	0	40	40	0.87	2.54
	10	62463	32.41	0.66	9.01	-0.43	-0.56	0	49	49	0.90	2.85
Science	4	59651	28.36	0.73	6.31	-0.68	-0.06	0	39	39	0.86	2.40
	8	60654	29.90	0.75	6.51	-0.74	0.05	0	40	40	0.86	2.43
	10	62526	31.43	0.63	8.78	-0.30	-0.60	0	50	50	0.88	3.00

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-26
Raw Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean	Test Difficulty	SD	Alpha	N Count	Mean	Test Difficulty	SD	Alpha
Reading	3	30810	37.95	0.67	11.92	0.94	29351	40.37	0.71	10.93	0.93
	4	30582	37.48	0.62	11.31	0.92	29018	39.67	0.66	10.59	0.91
	5	30193	38.46	0.64	10.89	0.91	29019	40.09	0.67	10.32	0.91
	6	30983	37.68	0.63	10.30	0.90	29367	39.49	0.66	9.66	0.89
	7	31298	38.57	0.64	10.18	0.90	29888	39.74	0.66	9.43	0.89
	8	31101	40.37	0.67	10.36	0.91	29575	43.07	0.72	9.45	0.89
	10	32026	36.60	0.65	10.38	0.91	30706	38.14	0.68	9.80	0.90
Mathematics	3	30911	40.57	0.71	9.62	0.91	29442	40.52	0.71	9.51	0.90
	4	30676	41.10	0.72	8.98	0.90	29091	41.03	0.72	8.80	0.89
	5	30220	43.26	0.70	10.78	0.91	29041	42.97	0.69	10.47	0.90
	6	31012	41.39	0.67	11.52	0.92	29396	41.57	0.67	11.10	0.92
	7	31346	41.41	0.67	11.45	0.92	29921	41.36	0.67	11.03	0.92
	8	31126	36.22	0.58	12.07	0.92	29592	35.94	0.58	11.53	0.91
	10	31995	35.23	0.61	13.15	0.94	30665	34.36	0.59	12.64	0.93
Language Arts	4	30566	18.89	0.63	5.45	0.82	29023	20.31	0.68	5.17	0.82
	8	30952	19.89	0.69	5.73	0.85	29478	21.94	0.76	5.06	0.84
	10	31873	23.90	0.61	6.52	0.84	30588	26.07	0.67	5.85	0.82
Social Studies	4	30589	28.62	0.75	6.49	0.88	29041	29.32	0.77	6.17	0.87
	8	31050	28.01	0.70	7.41	0.88	29552	27.99	0.70	6.92	0.86
	10	31875	32.49	0.66	9.46	0.91	30584	32.32	0.66	8.51	0.89
Science	4	30596	28.33	0.73	6.48	0.86	29052	28.39	0.73	6.13	0.85
	8	31082	29.97	0.75	6.80	0.87	29572	29.82	0.75	6.20	0.85
	10	31904	32.17	0.64	9.23	0.90	30618	30.66	0.61	8.21	0.86

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-27
Raw Score Descriptive Statistics for Reading by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Reading	W	3	43305	41.37	0.73	10.55	0.93
		4	43207	40.64	0.68	10.07	0.90
		5	43193	41.27	0.69	9.73	0.90
		6	44388	40.46	0.67	9.17	0.88
		7	45486	40.90	0.68	9.01	0.88
		8	45341	43.44	0.72	9.16	0.89
		10	48423	39.02	0.70	9.29	0.89
	AA	3	6657	31.45	0.55	11.78	0.92
		4	6461	30.58	0.51	11.45	0.91
		5	6363	31.30	0.52	11.09	0.91
		6	6251	31.04	0.52	10.21	0.89
		7	6470	31.61	0.53	10.29	0.89
		8	6366	33.96	0.57	10.64	0.90
		10	5905	29.00	0.52	10.54	0.90
	H	3	6646	32.92	0.58	11.53	0.92
		4	6479	33.10	0.55	11.04	0.91
		5	6238	34.24	0.57	10.60	0.90
		6	6230	33.47	0.56	10.03	0.89
		7	5784	34.68	0.58	9.80	0.88
		8	5633	37.02	0.62	10.12	0.89
		10	5000	32.34	0.58	10.31	0.90
	A	3	2578	38.68	0.68	11.39	0.93
		4	2468	38.08	0.63	11.00	0.91
		5	2388	38.52	0.64	10.90	0.91
		6	2488	37.46	0.62	10.69	0.91
		7	2443	38.34	0.64	10.22	0.90
		8	2337	41.52	0.69	9.90	0.90
		10	2415	35.80	0.64	10.87	0.92
	AI	3	975	35.73	0.63	11.50	0.93
		4	985	35.57	0.59	10.99	0.91
		5	1029	36.11	0.60	10.12	0.89
		6	993	35.84	0.60	9.77	0.89
		7	1003	35.68	0.59	9.79	0.89
		8	997	38.49	0.64	10.24	0.90
		10	985	34.47	0.62	10.16	0.90

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-28
Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Mathematics	W	3	43426	42.36	0.74	8.63	0.89
		4	43261	42.77	0.75	7.92	0.87
		5	43187	45.07	0.73	9.80	0.90
		6	44397	43.61	0.70	10.39	0.91
		7	45507	43.54	0.70	10.33	0.91
		8	45333	38.27	0.62	11.13	0.91
		10	48378	37.09	0.64	12.22	0.93
	AA	3	6659	32.71	0.57	10.25	0.91
		4	6486	33.40	0.59	9.78	0.90
		5	6370	34.58	0.56	10.82	0.90
		6	6247	31.75	0.51	11.14	0.91
		7	6468	31.56	0.51	10.86	0.90
		8	6374	25.90	0.42	10.27	0.89
		10	5876	22.85	0.39	10.49	0.90
	H	3	6688	36.69	0.64	9.48	0.90
		4	6538	37.46	0.66	8.93	0.89
		5	6268	38.49	0.62	10.21	0.89
		6	6268	36.06	0.58	10.84	0.90
		7	5830	36.00	0.58	10.63	0.90
		8	5660	30.30	0.49	10.70	0.90
		10	5015	27.61	0.48	11.31	0.91
	A	3	2602	41.44	0.73	9.48	0.90
		4	2498	42.00	0.74	8.79	0.90
		5	2406	44.21	0.71	10.62	0.91
		6	2503	43.16	0.70	11.31	0.92
		7	2461	42.21	0.68	11.52	0.93
		8	2352	37.38	0.60	11.83	0.92
		10	2401	34.88	0.60	13.04	0.94
	AI	3	978	37.14	0.65	9.53	0.90
		4	984	38.23	0.67	8.71	0.88
		5	1029	39.37	0.63	9.98	0.89
		6	993	37.39	0.60	10.69	0.90
		7	1001	36.39	0.59	10.62	0.90
		8	997	31.65	0.51	10.69	0.90
		10	986	30.02	0.52	11.85	0.92

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Language Arts	W	4	43174	20.46	0.68	5.06	0.81
		8	45158	21.66	0.75	5.24	0.84
		10	48316	26.00	0.67	5.81	0.82
	AA	4	6438	15.91	0.53	5.42	0.80
		8	6317	17.47	0.60	5.68	0.83
		10	5771	19.42	0.50	6.49	0.82
	H	4	6506	17.60	0.59	5.17	0.79
		8	5614	18.75	0.65	5.44	0.83
		10	4978	21.92	0.56	6.10	0.81
	A	4	2488	19.58	0.65	5.37	0.82
		8	2343	21.14	0.73	5.20	0.84
		10	2415	24.57	0.63	6.46	0.84
	AI	4	983	18.07	0.60	5.23	0.80
		8	996	19.09	0.66	5.53	0.83
		10	977	22.89	0.59	6.24	0.82

Table 8-30
Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Social Studies	W	4	43198	30.16	0.79	5.63	0.85
		8	45285	29.31	0.73	6.62	0.86
		10	48306	33.86	0.69	8.42	0.89
	AA	4	6452	24.04	0.63	7.36	0.88
		8	6332	22.16	0.55	7.20	0.85
		10	5777	24.93	0.51	8.92	0.88
	H	4	6510	26.34	0.69	6.46	0.86
		8	5641	24.54	0.61	6.96	0.85
		10	4978	27.91	0.57	8.64	0.88
	A	4	2487	28.48	0.75	6.39	0.87
		8	2345	27.79	0.69	7.07	0.87
		10	2417	31.58	0.64	9.19	0.90
	AI	4	983	27.13	0.71	6.34	0.86
		8	997	25.77	0.64	6.95	0.85
		10	977	29.79	0.61	8.82	0.89

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Science	W	4	43198	29.63	0.76	5.69	0.83
		8	45312	31.15	0.78	5.88	0.84
		10	48338	33.02	0.66	8.12	0.87
	AA	4	6455	23.20	0.59	6.73	0.84
		8	6353	24.11	0.60	6.80	0.84
		10	5798	23.12	0.46	8.06	0.84
	H	4	6520	25.43	0.65	6.15	0.82
		8	5644	26.74	0.67	6.48	0.83
		10	4985	26.52	0.53	8.25	0.85
	A	4	2492	27.99	0.72	6.35	0.86
		8	2347	29.66	0.74	6.25	0.84
		10	2417	30.62	0.61	8.92	0.88
	AI	4	983	26.79	0.69	6.26	0.84
		8	996	28.12	0.70	6.49	0.84
		10	980	29.10	0.58	8.50	0.87

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-32
Raw Score Descriptive Statistics by Socioeconomic Status

Content	Grade	Economically Disadvantaged					Not Economically Disadvantaged				
		N Count	Mean	Test Difficulty	SD	Alpha	N Count	Mean	Test Difficulty	SD	Alpha
Reading	3	26827	34.64	0.61	11.84	0.93	33336	42.74	0.75	9.86	0.92
	4	26216	34.12	0.57	11.26	0.91	33386	42.01	0.70	9.47	0.90
	5	25788	34.89	0.58	10.88	0.91	33426	42.64	0.71	9.12	0.89
	6	25685	34.16	0.57	10.13	0.89	34666	41.82	0.70	8.61	0.87
	7	25204	34.88	0.58	10.01	0.89	35983	42.12	0.70	8.53	0.87
	8	24403	37.32	0.62	10.34	0.90	36273	44.63	0.74	8.63	0.88
	10	21993	32.79	0.59	10.43	0.90	40743	39.81	0.71	9.06	0.89
Mathematics	3	26912	36.84	0.65	9.76	0.90	33443	43.53	0.76	8.28	0.88
	4	26337	37.46	0.66	9.20	0.89	33433	43.91	0.77	7.50	0.87
	5	25835	38.69	0.62	10.67	0.90	33428	46.54	0.75	9.26	0.89
	6	25735	36.45	0.59	11.17	0.91	34674	45.21	0.73	9.89	0.90
	7	25265	36.25	0.58	11.01	0.91	36003	44.99	0.73	9.93	0.91
	8	24452	30.70	0.50	11.02	0.90	36266	39.71	0.64	10.90	0.91
	10	21964	28.38	0.49	11.94	0.92	40700	38.27	0.66	12.05	0.93
Language Arts	4	26224	17.61	0.59	5.31	0.80	33368	21.13	0.70	4.87	0.80
	8	24301	18.77	0.65	5.59	0.83	36129	22.32	0.77	4.97	0.84
	10	21809	21.99	0.56	6.42	0.83	40656	26.55	0.68	5.61	0.81
Social Studies	4	26251	26.50	0.70	6.72	0.87	33381	30.89	0.81	5.29	0.84
	8	24370	24.84	0.62	7.19	0.86	36232	30.13	0.75	6.32	0.85
	10	21818	28.15	0.57	8.94	0.89	40645	34.69	0.71	8.18	0.89
Science	4	26274	25.85	0.66	6.49	0.85	33377	30.33	0.78	5.41	0.82
	8	24408	27.10	0.68	6.78	0.85	36246	31.78	0.79	5.59	0.83
	10	21863	27.34	0.55	8.68	0.87	40663	33.63	0.67	8.01	0.87

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-33
Raw Score Descriptive Statistics by Disability

Content	Grade	Disabled					Not Disabled				
		N Count	Mean	Test Difficulty	SD	Alpha	N Count	Mean	Test Difficulty	SD	Alpha
Reading	3	7474	28.99	0.51	12.59	0.93	52689	40.57	0.71	10.59	0.92
	4	7686	28.15	0.47	11.93	0.92	51916	40.08	0.67	10.00	0.90
	5	7847	28.58	0.48	11.38	0.91	51367	40.89	0.68	9.53	0.89
	6	7979	27.61	0.46	10.45	0.89	52372	40.23	0.67	8.85	0.87
	7	7915	28.19	0.47	10.21	0.89	53272	40.77	0.68	8.67	0.87
	8	7745	29.79	0.50	10.33	0.89	52931	43.43	0.72	8.70	0.88
	10	7226	26.00	0.46	9.95	0.89	55510	38.83	0.69	9.18	0.89
Mathematics	3	7500	34.52	0.61	10.56	0.91	52855	41.40	0.73	9.10	0.90
	4	7706	34.24	0.60	10.23	0.91	52064	42.08	0.74	8.21	0.88
	5	7847	33.83	0.55	11.45	0.91	51416	44.53	0.72	9.75	0.89
	6	7986	30.48	0.49	11.41	0.91	52423	43.15	0.70	10.32	0.91
	7	7927	29.50	0.48	10.83	0.90	53341	43.15	0.70	10.19	0.90
	8	7746	24.15	0.39	10.12	0.89	52972	37.83	0.61	11.00	0.91
	10	7205	21.47	0.37	10.12	0.89	55459	36.53	0.63	12.20	0.93
Language Arts	4	7679	15.87	0.53	5.30	0.79	51913	20.13	0.67	5.15	0.81
	8	7689	15.56	0.54	5.57	0.81	52741	21.67	0.75	5.05	0.83
	10	7104	18.11	0.46	6.12	0.79	55361	25.84	0.66	5.76	0.81
Social Studies	4	7679	24.84	0.65	6.94	0.87	51953	29.57	0.78	6.02	0.86
	8	7702	21.42	0.54	7.24	0.85	52900	28.96	0.72	6.64	0.86
	10	7109	23.47	0.48	8.88	0.88	55354	33.56	0.68	8.36	0.89
Science	4	7692	24.73	0.63	6.81	0.85	51959	28.90	0.74	6.06	0.85
	8	7718	23.97	0.60	7.11	0.85	52936	30.76	0.77	5.95	0.84
	10	7140	23.18	0.46	8.49	0.86	55386	32.49	0.65	8.23	0.87

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-34
Raw Score Descriptive Statistics by English Language Proficiency

Content	Grade	Limited English Proficient					Fully English Proficient				
		N Count	Mean	Test Difficulty	SD	Alpha	N Count	Mean	Test Difficulty	SD	Alpha
Reading	3	4858	30.52	0.54	10.75	0.90	55305	39.89	0.70	11.27	0.93
	4	4238	29.78	0.50	9.86	0.88	55364	39.21	0.65	10.81	0.91
	5	3241	28.61	0.48	9.51	0.87	55973	39.88	0.66	10.38	0.91
	6	3046	27.28	0.45	8.50	0.83	57305	39.16	0.65	9.75	0.89
	7	3131	29.55	0.49	8.56	0.83	58056	39.66	0.66	9.63	0.89
	8	3081	32.06	0.53	8.76	0.84	57595	42.21	0.70	9.82	0.90
	10	1759	23.93	0.43	8.32	0.83	60977	37.74	0.67	9.91	0.90
Mathematics	3	4930	36.25	0.64	9.22	0.89	55425	40.93	0.72	9.50	0.90
	4	4342	36.21	0.64	8.83	0.88	55428	41.45	0.73	8.78	0.89
	5	3298	35.02	0.56	9.64	0.87	55965	43.59	0.70	10.50	0.91
	6	3115	31.82	0.51	9.82	0.88	57294	42.00	0.68	11.16	0.92
	7	3202	31.82	0.51	9.38	0.87	58066	41.91	0.68	11.11	0.92
	8	3134	26.42	0.43	9.06	0.86	57584	36.61	0.59	11.71	0.92
	10	1789	20.68	0.36	8.48	0.84	60875	35.22	0.61	12.78	0.93
Language Arts	4	4299	16.40	0.55	4.79	0.75	55293	19.83	0.66	5.33	0.82
	8	3099	16.66	0.57	4.87	0.76	57331	21.12	0.73	5.45	0.85
	10	1780	17.46	0.45	5.39	0.73	60685	25.18	0.65	6.19	0.83
Social Studies	4	4304	24.82	0.65	6.27	0.84	55328	29.28	0.77	6.24	0.87
	8	3110	21.36	0.53	5.99	0.77	57492	28.36	0.71	7.06	0.87
	10	1774	21.51	0.44	6.97	0.79	60689	32.73	0.67	8.86	0.90
Science	4	4329	24.05	0.62	5.88	0.80	55322	28.70	0.74	6.22	0.85
	8	3116	24.06	0.60	5.80	0.77	57538	30.21	0.76	6.40	0.86
	10	1787	20.89	0.42	6.49	0.75	60739	31.74	0.63	8.64	0.88

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-35
2012 and 2010* Scale Score Mean and Standard Deviation

Content	Grade	2012 Mean	2010 Mean	Diff = 2012 – 2010 Mean	2012 Standard Deviation	2010 Standard Deviation	Diff = 2012 - 2010 Standard Deviation
Reading	3	457.99	457.90	0.09	42.17	40.83	1.34
	4	476.56	477.62	-1.06	44.91	44.91	0.00
	5	483.45	484.07	-0.62	45.96	45.71	0.25
	6	503.85	504.40	-0.55	48.34	47.17	1.17
	7	517.15	516.48	0.67	45.67	45.85	-0.18
	8	533.42	532.78	0.64	51.32	51.10	0.22
	10*	545.76	539.97	5.79	61.82	63.17	-1.35
Mathematics	3	433.95	431.95	2.00	45.22	43.49	1.73
	4	470.45	469.93	0.52	44.83	44.14	0.69
	5	499.27	498.20	1.07	49.19	48.55	0.64
	6	524.24	522.54	1.70	49.73	49.29	0.44
	7	538.69	537.25	1.44	45.02	44.04	0.98
	8	547.64	547.32	0.32	52.36	51.09	1.27
	10	565.19	562.28	2.91	46.03	46.46	-0.43
Language Arts	4	294.59	295.23	-0.64	30.56	29.93	0.63
	8	396.04	397.55	-1.51	44.51	44.58	-0.07
	10	450.82	450.15	0.67	40.32	42.75	-2.43
Social Studies	4	295.28	296.16	-0.88	27.21	27.52	-0.31
	8	396.99	397.30	-0.31	39.85	39.92	-0.07
	10	454.79	452.77	2.02	43.74	44.24	-0.50
Science	4	298.34	298.37	-0.03	29.99	29.78	0.21
	8	402.53	402.37	0.16	41.97	42.08	-0.11
	10	456.65	454.45	2.20	43.95	44.66	-0.71

* Note: The prior administration data for Reading Grade 10 was taken from 2008 rather than 2010.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-36
Scale Score Descriptive Statistics

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
Reading	3	60164	457.99	42.17	-1.03	4.29	270	640	270	640
	4	59602	476.56	44.91	-1.01	3.12	280	650	280	650
	5	59214	483.45	45.96	-1.14	3.62	290	690	290	690
	6	60351	503.85	48.34	-0.97	2.78	300	730	300	730
	7	61187	517.15	45.67	-0.70	1.88	310	716	310	780
	8	60676	533.42	51.32	-0.43	1.78	330	790	330	790
	10	62736	545.76	61.82	-0.33	1.14	350	820	350	820
Mathematics	3	60355	433.95	45.22	-0.01	1.75	220	630	220	630
	4	59770	470.45	44.83	-0.29	1.79	240	650	240	650
	5	59263	499.27	49.19	-0.28	1.59	270	680	270	680
	6	60409	524.24	49.73	-0.25	1.02	310	700	310	700
	7	61268	538.69	45.02	-0.25	1.16	330	710	330	710
	8	60718	547.64	52.36	-0.58	1.44	350	730	350	730
	10	62664	565.19	46.03	-0.23	1.48	410	750	410	750
Language Arts	4	59592	294.59	30.56	-0.44	3.63	140	420	140	420
	8	60430	396.04	44.51	0.14	1.45	250	520	250	520
	10	62465	450.82	40.32	-0.24	0.89	290	630	290	630
Social Studies	4	59632	295.28	27.21	0.58	4.42	170	400	170	400
	8	60602	396.99	39.85	-0.17	1.91	230	530	230	530
	10	62463	454.79	43.74	-0.38	2.09	240	620	240	620
Science	4	59651	298.34	29.99	0.32	3.15	170	440	170	440
	8	60654	402.53	41.97	0.36	2.49	230	560	230	560
	10	62526	456.65	43.95	-0.54	2.16	240	610	240	610

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-37
Scale Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
Reading	3	30811	453.65	43.91	270	640	29351	462.56	39.73	270	640
	4	30582	472.15	46.92	280	650	29018	481.21	42.19	280	650
	5	30193	480.04	48.07	290	678	29019	487.00	43.37	290	690
	6	30983	499.67	50.37	300	730	29367	508.28	45.66	300	730
	7	31298	514.50	47.88	310	716	29888	519.93	43.07	310	711
	8	31101	527.27	52.64	330	790	29575	539.90	49.08	330	790
	10	32026	541.24	63.70	350	820	30706	550.49	59.42	350	820
Mathematics	3	30911	434.21	45.53	220	630	29442	433.69	44.88	220	630
	4	30676	471.11	45.67	240	650	29091	469.76	43.91	240	650
	5	30220	500.37	50.68	270	680	29041	498.13	47.56	270	680
	6	31012	523.89	51.07	310	700	29396	524.61	48.26	310	700
	7	31346	539.07	46.41	330	710	29921	538.30	43.52	330	710
	8	31126	547.95	54.49	350	730	29592	547.30	50.03	350	730
	10	31995	566.82	47.83	410	750	30665	563.51	43.98	410	750
Language Arts	4	30566	290.53	31.28	140	420	29023	298.86	29.18	140	420
	8	30952	388.10	44.90	250	520	29478	404.38	42.53	250	520
	10	31873	444.32	41.19	290	618	30588	457.61	38.23	290	630
Social Studies	4	30589	293.81	27.37	170	400	29041	296.84	26.95	170	400
	8	31050	397.00	41.97	230	530	29552	396.99	37.50	230	530
	10	31875	455.13	47.17	240	620	30584	454.44	39.84	240	620
Science	4	30596	298.04	31.07	170	440	29052	298.66	28.80	170	440
	8	31082	403.46	44.94	230	560	29572	401.55	38.57	230	560
	10	31904	460.00	47.69	240	610	30618	453.16	39.35	240	610

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-38
Scale Score Descriptive Statistics for Reading by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Reading	W	3	43305	465.51	38.45	270	640
		4	43207	484.72	40.54	280	650
		5	43193	491.70	41.38	290	690
		6	44388	512.70	43.64	300	730
		7	45486	525.20	41.71	310	716
		8	45341	542.14	47.87	330	790
		10	48423	555.63	57.06	350	820
	AA	3	6657	431.49	46.20	270	584
		4	6461	444.76	51.33	280	604
		5	6363	450.30	52.89	290	607
		6	6251	468.11	53.91	300	676
		7	6470	482.55	49.79	310	649
		8	6366	494.89	52.66	330	776
		10	5905	495.90	64.65	350	820
	H	3	6646	437.09	43.25	270	640
		4	6479	455.91	45.14	280	608
		5	6238	463.16	46.74	290	612
		6	6230	480.37	48.29	300	655
		7	5784	496.69	44.61	310	674
		8	5633	509.99	48.24	330	756
		10	5000	516.14	60.93	350	820
	A	3	2579	458.00	42.25	270	640
		4	2468	475.59	44.40	280	641
		5	2388	480.75	46.82	290	621
		6	2488	499.60	51.31	300	685
		7	2443	513.77	47.42	310	685
		8	2337	532.90	51.34	330	790
		10	2415	538.33	67.27	350	820
	AI	3	975	447.45	40.57	270	640
		4	985	465.69	43.06	280	627
		5	1029	471.64	41.88	290	602
		6	993	491.81	46.25	300	719
		7	1003	501.65	45.58	310	627
		8	997	517.04	52.46	330	756
		10	985	528.78	61.23	350	820

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-39
Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Mathematics	W	3	43426	442.19	42.14	220	630
		4	43261	478.72	40.94	240	650
		5	43187	507.88	45.89	270	680
		6	44397	533.27	45.85	310	700
		7	45507	546.96	41.56	330	710
		8	45333	557.47	47.17	350	730
		10	48378	573.23	42.40	410	750
	AA	3	6659	398.83	45.05	220	630
		4	6486	433.51	47.97	240	650
		5	6370	461.03	50.57	270	656
		6	6247	482.47	50.35	310	679
		7	6468	500.40	44.71	330	710
		8	6374	500.42	55.59	350	730
		10	5876	521.45	45.48	410	681
	H	3	6688	415.80	41.32	220	630
		4	6538	452.47	42.09	240	650
		5	6268	478.97	44.21	270	680
		6	6268	501.09	45.47	310	684
		7	5830	517.99	40.79	330	710
		8	5660	522.69	50.00	350	697
		10	5015	541.40	41.57	410	750
	A	3	2602	438.77	48.17	220	630
		4	2498	475.73	45.93	240	650
		5	2406	505.95	51.71	270	680
		6	2503	533.04	52.52	310	700
		7	2461	543.51	48.50	330	710
		8	2352	554.31	53.34	350	730
		10	2401	566.34	48.59	410	750
	AI	3	978	418.54	41.37	220	630
		4	984	456.67	41.16	240	600
		5	1029	482.82	43.82	270	680
		6	993	507.49	44.92	310	678
		7	1001	518.94	40.43	330	637
		8	997	528.43	49.89	350	682
		10	986	549.96	42.77	410	750

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-40
Scale Score Descriptive Statistics for Language Arts by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Language Arts	W	4	43174	299.41	28.61	140	420
		8	45158	402.00	43.62	250	520
		10	48316	457.13	38.00	290	630
	AA	4	6438	274.38	33.10	140	420
		8	6317	369.65	41.92	250	520
		10	5771	417.26	39.63	290	576
	H	4	6506	283.69	29.18	140	420
		8	5614	379.39	39.76	250	520
		10	4978	432.27	36.52	290	581
	A	4	2488	294.94	31.63	140	420
		8	2343	398.22	42.68	250	520
		10	2415	448.61	42.51	290	630
	AI	4	983	286.45	27.81	140	420
		8	996	382.15	42.74	250	520
		10	977	437.49	37.72	290	537

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-41
Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Social Studies	W	4	43198	299.90	26.10	170	400
		8	45285	403.93	37.40	230	530
		10	48306	461.62	40.68	240	620
	AA	4	6452	276.59	27.27	170	400
		8	6332	365.94	39.77	230	530
		10	5777	418.86	46.99	240	620
	H	4	6510	284.94	23.54	170	400
		8	5641	378.54	36.41	230	530
		10	4978	434.43	40.05	240	584
	A	4	2487	293.62	26.95	170	400
		8	2345	396.22	40.06	230	530
		10	2417	451.09	46.16	240	620
	AI	4	983	287.65	23.79	170	400
		8	997	385.42	36.37	230	530
		10	977	442.74	41.44	240	584

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

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Science	W	4	43198	303.91	28.36	170	440
		8	45312	409.81	40.18	230	560
		10	48338	464.38	40.00	240	610
	AA	4	6455	275.91	28.99	170	440
		8	6353	369.27	37.97	230	560
		10	5798	415.38	46.61	240	583
	H	4	6520	285.25	26.62	170	440
		8	5644	383.71	36.26	230	560
		10	4985	433.57	41.27	240	610
	A	4	2492	297.05	30.62	170	440
		8	2347	401.98	41.57	230	560
		10	2417	452.98	46.12	240	610
	AI	4	983	291.21	27.62	199	440
		8	996	391.42	38.72	230	560
		10	980	445.69	40.62	240	576

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-43
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
Reading	3	26828	442.57	43.68	270	640	33336	470.40	36.46	270	640
	4	26216	459.32	46.93	280	650	33386	490.10	38.15	280	650
	5	25788	465.38	48.38	290	690	33426	497.40	38.60	290	690
	6	25685	483.31	50.01	300	719	34666	519.08	40.87	300	730
	7	25204	497.73	46.27	310	674	35983	530.75	39.98	310	716
	8	24403	511.44	50.49	330	790	36273	548.21	46.34	330	790
	10	21993	518.59	61.98	350	820	40743	560.43	56.55	350	820
Mathematics	3	26912	416.63	42.97	220	630	33443	447.89	42.06	220	630
	4	26337	452.61	44.16	240	650	33433	484.50	40.11	240	650
	5	25835	479.29	47.64	270	680	33428	514.72	44.58	270	680
	6	25735	502.51	47.71	310	700	34674	540.36	44.82	310	700
	7	25265	518.54	42.87	330	710	36003	552.83	40.92	330	710
	8	24452	524.14	52.08	350	730	36266	563.48	46.25	350	730
	10	21964	542.97	44.67	410	750	40700	577.19	42.12	410	750
Language Arts	4	26224	283.75	30.12	140	420	33368	303.11	28.10	140	420
	8	24301	379.43	41.39	250	520	36129	407.21	43.02	250	520
	10	21809	432.30	39.05	290	630	40656	460.76	37.38	290	630
Social Studies	4	26251	285.41	25.07	170	400	33381	303.05	26.29	170	400
	8	24370	380.05	38.25	230	530	36232	408.39	36.73	230	530
	10	21818	434.82	43.38	240	620	40645	465.51	40.00	240	620
Science	4	26274	287.01	28.24	170	440	33377	307.26	28.26	170	440
	8	24408	385.78	39.33	230	560	36246	413.81	39.87	230	560
	10	21863	436.77	44.60	240	610	40663	467.33	39.67	240	610

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-44
Scale Score Descriptive Statistics by Disability

Content	Grade	Disabled					Not Disabled				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
Reading	3	7474	420.78	53.83	270	640	52690	463.27	37.36	270	640
	4	7686	433.06	57.61	280	638	51916	483.00	38.76	280	650
	5	7847	436.94	59.43	290	661	51367	490.56	38.92	290	690
	6	7979	450.08	60.11	300	637	52372	512.05	40.43	300	730
	7	7915	467.21	52.09	310	638	53272	524.57	39.59	310	716
	8	7745	475.34	54.91	330	680	52931	541.92	44.86	330	790
	10	7226	478.62	63.32	350	723	55510	554.50	55.99	350	820
Mathematics	3	7500	406.71	47.49	220	630	52855	437.82	43.53	220	630
	4	7706	436.97	51.05	240	650	52064	475.40	41.61	240	650
	5	7847	456.73	55.38	270	680	51416	505.77	44.75	270	680
	6	7986	476.30	53.27	310	700	52423	531.54	44.88	310	700
	7	7927	491.73	46.37	330	710	53341	545.67	40.40	330	710
	8	7746	491.72	56.22	350	730	52972	555.81	46.44	350	730
	10	7205	514.81	46.06	410	750	55459	571.74	41.78	410	750
Language Arts	4	7679	273.44	33.52	140	420	51913	297.72	28.80	140	420
	8	7689	356.72	42.47	250	520	52741	401.77	41.82	250	520
	10	7104	409.51	37.05	290	584	55361	456.13	37.57	290	630
Social Studies	4	7679	279.96	26.00	170	400	51953	297.55	26.64	170	400
	8	7702	361.72	41.43	230	530	52900	402.13	36.91	230	530
	10	7109	411.98	47.91	240	620	55354	460.29	39.98	240	620
Science	4	7692	282.90	30.05	170	440	51959	300.63	29.29	170	440
	8	7718	368.09	42.08	230	560	52936	407.55	39.52	230	560
	10	7140	414.87	49.28	240	610	55386	462.03	40.16	240	610

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-45
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
Reading	3	4858	429.69	41.31	270	640	55306	460.48	41.33	270	640
	4	4238	443.62	42.39	280	575	55364	479.09	44.10	280	650
	5	3241	440.59	47.40	290	578	55973	485.94	44.63	290	690
	6	3046	452.52	46.22	300	611	57305	506.58	46.90	300	730
	7	3131	473.75	40.85	310	611	58056	519.49	44.74	310	716
	8	3081	487.13	41.37	330	659	57595	535.90	50.62	330	790
	10	1759	467.64	53.87	350	663	60977	548.02	60.56	350	820
Mathematics	3	4930	413.55	39.99	220	630	55425	435.77	45.21	220	630
	4	4342	446.43	41.54	240	650	55428	472.33	44.53	240	650
	5	3298	464.28	42.44	270	680	55965	501.34	48.78	270	680
	6	3115	483.67	42.83	310	678	57294	526.44	49.12	310	700
	7	3202	502.43	36.76	330	620	58066	540.69	44.58	330	710
	8	3134	505.26	47.48	350	705	57584	549.94	51.63	350	730
	10	1789	514.29	39.40	410	750	60875	566.69	45.35	410	750
Language Arts	4	4299	276.99	28.38	140	420	55293	295.96	30.29	140	420
	8	3099	364.51	33.99	250	520	57331	397.75	44.38	250	520
	10	1780	406.11	32.69	290	508	60685	452.14	39.77	290	630
Social Studies	4	4304	279.46	21.07	170	400	55328	296.51	27.24	170	400
	8	3110	362.46	32.74	230	465	57492	398.86	39.35	230	530
	10	1774	404.35	38.46	240	513	60689	456.26	43.00	240	620
Science	4	4329	279.15	25.31	170	440	55322	299.85	29.81	170	440
	8	3116	368.95	30.92	230	485	57538	404.35	41.72	230	560
	10	1787	404.48	39.50	240	576	60739	458.18	43.13	240	610

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-46
Performance Level Cut Scores for all Contents*

Content	3			4			5			6			7			8			10		
	B	**P**	**A**	**B**	**P**	**A**	**B**	**P**	**A**	**B**	**P**	**A**	**B**	**P**	**A**	**B**	**P**	**A**	**B**	**P**	**A**
Reading	445	475	507	462	498	536	464	503	546	479	525	573	486	535	590	493	548	613	496	566	644
Mathematics	388	438	492	425	474	526	449	501	553	475	524	573	500	544	591	510	558	605	528	574	618
Language Arts				252	277	308										358	385	418	393	428	484
Social Studies				242	263	288										334	364	403	408	420	455
Science				249	279	320										349	375	419	411	429	466

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-47
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
3	M	18958	31.51	27.27	35.55	23.79	57.46	53.30	33.50	43.49	28.80	62.33	65.87	26.64	46.81	19.20
	B	20322	33.78	34.52	33.07	34.57	29.94	31.99	34.86	34.15	34.10	30.16	21.90	35.46	33.13	34.30
	P	16030	26.64	28.94	24.46	31.61	11.09	12.58	22.88	18.15	28.39	6.81	9.87	29.02	16.83	34.54
	A	4854	8.07	9.26	6.93	10.03	1.52	2.14	8.76	4.21	8.72	0.70	2.35	8.88	3.23	11.96
Total		60164	100	29351	30811	43305	6657	6646	2579	975	55306	4858	7474	52690	26828	33336
4	M	18749	31.46	28.02	34.72	23.72	60.32	51.37	34.56	42.74	28.89	64.94	67.72	26.09	47.16	19.12
	B	21413	35.93	36.24	35.62	37.55	28.32	33.29	34.56	35.13	36.39	29.92	22.16	37.97	34.28	37.22
	P	15843	26.58	28.57	24.70	31.32	9.97	13.61	24.03	19.29	28.25	4.74	8.67	29.23	16.24	34.70
	A	3597	6.04	7.17	4.96	7.40	1.39	1.73	6.85	2.84	6.47	0.40	1.46	6.71	2.32	8.96
Total		59602	100	29018	30582	43207	6461	6479	2468	985	55364	4238	7686	51916	26216	33386
5	M	16242	27.43	24.96	29.80	20.15	56.58	45.08	31.62	36.15	25.02	69.08	66.38	21.48	42.50	15.81
	B	22529	38.05	38.34	37.76	38.96	31.20	38.33	36.39	44.12	38.71	26.57	23.64	40.25	37.86	38.19
	P	17279	29.18	30.68	27.74	34.25	11.25	15.33	26.34	17.98	30.63	4.23	8.86	32.29	17.93	37.86
	A	3164	5.34	6.02	4.69	6.64	0.97	1.27	5.65	1.75	5.65	0.12	1.12	5.99	1.71	8.15
Total		59214	100	29019	30193	43193	6363	6238	2388	1029	55973	3241	7847	51367	25788	33426

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-47 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
6	M	14977	24.82	21.96	27.52	17.61	52.65	44.29	31.43	32.93	22.26	72.95	67.20	18.36	40.45	13.23
	B	24615	40.79	41.07	40.52	41.56	36.94	39.81	37.30	45.52	41.66	24.33	24.80	43.22	41.34	40.38
	P	17927	29.71	31.41	28.09	35.14	9.66	14.62	24.84	19.94	31.15	2.43	7.18	33.14	16.84	39.23
	A	2832	4.69	5.56	3.87	5.70	0.75	1.28	6.43	1.61	4.93	0.30	0.81	5.28	1.37	7.15
Total		60351	100.00	29367	30983	44388	6251	6230	2488	993	57305	3046	7979	52372	25685	34666
7	M	13250	21.66	19.21	23.99	15.49	48.38	36.70	25.34	32.90	19.66	58.58	64.46	15.30	35.68	11.83
	B	25418	41.54	43.25	39.91	41.30	39.12	45.49	42.08	44.17	41.74	37.85	27.05	43.69	44.16	39.70
	P	20326	33.22	33.67	32.79	38.93	12.01	16.44	27.88	20.94	34.82	3.51	8.05	36.96	19.04	43.15
	A	2193	3.58	3.87	3.31	4.28	0.49	1.37	4.71	1.99	3.77	0.06	0.44	4.05	1.12	5.31
Total		61187	100.00	29888	31298	45486	6470	5784	2443	1003	58056	3131	7915	53272	25204	35983
8	M	11069	18.24	14.60	21.70	12.60	44.02	32.65	19.21	26.78	16.50	50.73	60.80	12.02	30.93	9.71
	B	25331	41.75	41.49	41.99	40.87	42.27	46.26	44.20	47.04	41.55	45.44	31.70	43.22	46.24	38.73
	P	21628	35.65	38.33	33.10	41.28	13.02	19.94	30.64	24.47	37.35	3.70	7.15	39.81	21.62	45.08
	A	2648	4.36	5.58	3.21	5.26	0.69	1.15	5.95	1.71	4.59	0.13	0.35	4.95	1.21	6.48
Total		60676	100.00	29575	31101	45341	6366	5633	2337	997	57595	3081	7745	52931	24403	36273

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-47 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
10	M	11292	18.00	15.40	20.48	12.51	45.86	34.08	23.52	25.48	16.47	70.95	59.29	12.62	31.53	10.70
	B	27311	43.53	44.35	42.75	43.42	41.61	45.64	44.02	48.83	44.03	26.21	32.76	44.94	46.83	41.75
	P	21596	34.42	35.59	33.31	39.37	11.87	19.00	27.08	23.15	35.34	2.79	7.69	37.90	20.40	41.99
	A	2537	4.04	4.66	3.46	4.71	0.66	1.28	5.38	2.54	4.16	0.06	0.26	4.54	1.24	5.56
Total		62736	100	30706	32026	48423	5905	5000	2415	985	60977	1759	7226	55510	21993	40743

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-48
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
3	M	8318	13.78	13.58	13.97	8.67	37.39	23.12	11.72	21.57	12.86	24.14	32.88	11.07	22.86	6.48
	B	23193	38.43	38.77	38.10	35.92	43.58	48.42	38.55	45.81	37.35	50.57	42.65	37.83	45.99	32.34
	P	23930	39.65	39.81	39.50	45.53	17.57	25.54	38.24	29.24	41.13	22.96	21.07	42.29	28.01	49.02
	A	4914	8.14	7.84	8.43	9.88	1.46	2.93	11.49	3.37	8.66	2.33	3.40	8.81	3.14	12.16
Total		60355	100	29442	30911	43426	6659	6688	2602	978	55425	4930	7500	52855	26912	33443
4	M	7937	13.28	13.36	13.20	7.99	39.16	22.53	10.89	19.82	12.28	25.98	37.31	9.72	22.90	5.70
	B	22687	37.96	38.87	37.08	35.80	41.81	47.29	37.59	46.14	37.03	49.84	40.07	37.64	45.33	32.15
	P	24051	40.24	39.74	40.72	45.90	17.45	27.00	39.83	30.59	41.66	22.13	19.88	43.25	28.40	49.56
	A	5095	8.52	8.02	9.00	10.31	1.57	3.18	11.69	3.46	9.03	2.05	2.74	9.38	3.36	12.59
Total		59770	100	29091	30676	43261	6486	6538	2498	984	55428	4342	7706	52064	26337	33433
5	M	7728	13.04	12.86	13.21	8.41	35.95	21.60	10.85	18.76	11.96	31.32	40.81	8.80	22.39	5.82
	B	21429	36.16	37.90	34.49	33.29	42.92	47.50	35.04	48.10	35.32	50.36	39.09	35.71	44.31	29.86
	P	23267	39.26	38.81	39.69	44.31	19.09	27.12	38.20	28.57	40.57	17.10	16.97	42.66	29.02	47.18
	A	6839	11.54	10.43	12.61	13.99	2.04	3.78	15.92	4.57	12.15	1.21	3.13	12.82	4.28	17.15
Total		59263	100	29041	30220	43187	6370	6268	2406	1029	55965	3298	7847	51416	25835	33428

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-48 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
6	M	8700	14.40	13.77	15.00	9.05	41.41	25.61	10.91	21.55	13.13	37.75	47.83	9.31	25.16	6.42
	B	20270	33.56	34.19	32.95	31.20	39.07	43.32	33.00	43.81	32.82	47.09	34.80	33.37	41.64	27.55
	P	22406	37.09	37.56	36.64	41.85	16.82	25.94	34.88	27.69	38.38	13.45	14.17	40.58	27.53	44.19
	A	9033	14.95	14.47	15.41	17.89	2.69	5.14	21.21	6.95	15.67	1.70	3.19	16.74	5.67	21.84
Total		60409	100	29396	31012	44397	6247	6268	2503	993	57294	3115	7986	52423	25735	34674
7	M	10706	17.47	17.10	17.83	11.29	47.19	30.74	17.31	29.87	15.94	45.32	56.38	11.69	30.15	8.58
	B	21722	35.45	36.69	34.28	34.01	37.65	43.57	33.73	43.76	35.03	43.13	31.92	35.98	42.03	30.84
	P	22292	36.38	36.31	36.46	41.79	13.57	22.45	34.66	23.28	37.77	11.21	10.16	40.28	24.36	44.82
	A	6548	10.69	9.91	11.43	12.91	1.59	3.24	14.30	3.10	11.26	0.34	1.55	12.05	3.46	15.76
Total		61268	100	29921	31346	45507	6468	5830	2461	1001	58066	3202	7927	53341	25265	36003
8	M	12238	20.16	19.40	20.87	13.53	52.87	35.94	16.71	30.69	18.55	49.59	61.75	14.07	34.43	10.53
	B	21272	35.03	36.27	33.86	34.28	34.08	40.12	36.99	42.13	34.76	40.01	28.08	36.05	39.74	31.86
	P	20145	33.18	33.90	32.49	38.19	11.37	20.44	30.78	22.67	34.47	9.48	8.44	36.79	21.88	40.80
	A	7063	11.63	10.42	12.78	14.00	1.68	3.50	15.52	4.51	12.22	0.93	1.73	13.08	3.95	16.81
Total		60718	100	29592	31126	45333	6374	5660	2352	997	57584	3134	7746	52972	24452	36266

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-48 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
10	M	11522	18.39	18.55	18.22	12.32	53.40	34.20	18.37	26.57	17.10	62.21	61.35	12.81	33.16	10.42
	B	23157	36.95	38.64	35.34	36.07	35.87	44.77	37.23	46.25	37.07	32.87	29.60	37.91	42.56	33.93
	P	21592	34.46	34.30	34.61	39.54	9.41	18.56	31.53	22.21	35.34	4.36	7.83	37.92	21.08	41.67
	A	6393	10.20	8.51	11.83	12.06	1.31	2.47	12.87	4.97	10.49	0.56	1.22	11.37	3.20	13.98
Total		62664	100	30665	31995	48378	5876	5015	2401	986	60875	1789	7205	55459	21964	40700

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-49
Percentage of Students in Each Performance Level by Sub-Group (Language Arts)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	3623	6.08	4.14	7.92	3.65	17.65	10.31	5.95	9.46	5.50	13.49	18.21	4.29	10.43	2.66
	B	10815	18.15	15.41	20.75	14.76	31.95	26.04	18.17	24.21	17.23	29.94	32.67	16.00	26.11	11.89
	P	26091	43.78	43.67	43.89	44.20	38.06	46.11	44.65	45.47	43.48	47.66	37.67	44.69	44.71	43.06
	A	19063	31.99	36.78	27.44	37.39	12.35	17.54	31.23	20.86	33.78	8.91	11.45	35.03	18.75	42.39
Total		59592	100.00	29023	30566	43174	6438	6506	2488	983	55293	4299	7679	51913	26224	33368
8	M	9535	15.78	10.42	20.88	12.24	33.94	24.03	12.38	22.39	14.83	33.40	47.64	11.13	25.39	9.31
	B	14033	23.22	20.69	25.64	20.73	30.76	32.21	25.57	32.13	22.27	40.88	29.85	22.26	29.85	18.77
	P	20213	33.45	35.47	31.52	34.95	25.84	30.48	34.14	28.72	34.02	22.94	17.06	35.84	30.36	35.53
	A	16649	27.55	33.42	21.96	32.07	9.47	13.29	27.91	16.77	28.89	2.78	5.45	30.77	14.40	36.40
Total		60430	100.00	29478	30952	45158	6317	5614	2343	996	57331	3099	7689	52741	24301	36129
10	M	4979	7.97	5.07	10.75	5.19	26.06	13.34	7.54	12.08	7.32	30.28	32.08	4.88	15.17	4.11
	B	11120	17.80	14.49	20.98	14.37	33.81	29.21	21.70	24.97	17.00	45.06	37.78	15.24	27.71	12.49
	P	34066	54.54	56.76	52.40	57.39	35.70	50.36	52.22	52.00	55.43	23.99	27.65	57.99	48.67	57.68
	A	12300	19.69	23.67	15.87	23.05	4.44	7.09	18.55	10.95	20.25	0.67	2.49	21.90	8.45	25.72
Total		62465	100.00	30588	31873	48316	5771	4978	2415	977	60685	1780	7104	55361	21809	40656

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-50
Percentage of Students in Each Performance Level by Sub-Group (Social Studies)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	1301	2.18	1.75	2.59	1.03	8.66	3.30	2.25	2.44	2.01	4.44	5.91	1.63	4.06	0.70
	B	3476	5.83	5.17	6.46	3.59	16.58	9.69	5.79	8.14	5.38	11.57	13.84	4.64	9.89	2.63
	P	16392	27.49	26.36	28.56	22.97	40.56	40.97	32.09	39.17	25.85	48.51	42.22	25.31	37.67	19.48
	A	38463	64.50	66.72	62.40	72.40	34.19	46.04	59.87	50.25	66.76	35.48	38.03	68.41	48.38	77.18
Total		59632	100.00	29041	30589	43198	6452	6510	2487	983	55328	4304	7679	51953	26251	33381
8	M	3016	4.98	4.07	5.84	2.86	16.61	8.78	4.35	7.02	4.46	14.50	20.29	2.75	9.30	2.07
	B	7593	12.53	12.08	12.96	8.97	28.24	22.23	13.56	17.15	11.46	32.22	29.77	10.02	20.67	7.05
	P	22953	37.88	40.04	35.81	36.42	39.34	45.31	40.81	45.54	37.42	46.27	36.21	38.12	43.48	34.11
	A	27040	44.62	43.80	45.39	51.75	15.81	23.68	41.28	30.29	46.65	7.01	13.72	49.12	26.55	56.77
Total		60602	100.00	29552	31050	45285	6332	5641	2345	997	57492	3110	7702	52900	24370	36232
10	M	7749	12.41	10.51	14.22	8.35	36.16	22.76	13.41	17.09	11.33	49.32	43.97	8.35	22.83	6.81
	B	3668	5.87	5.95	5.80	4.84	10.97	9.32	6.50	7.68	5.64	13.75	12.73	4.99	9.29	4.04
	P	18224	29.18	32.10	26.37	27.61	32.44	37.59	33.10	34.80	29.13	30.78	26.83	29.48	35.46	25.80
	A	32822	52.55	51.44	53.61	59.20	20.43	30.33	47.00	40.43	53.90	6.14	16.47	57.18	32.43	63.35
Total		62463	100.00	30584	31875	48306	5777	4978	2417	977	60689	1774	7109	55354	21818	40645

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-51
Percentage of Students in Each Performance Level by Sub-Group (Science)

Grade	Proficiency Level	Examinees		Gender		Race/Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	2743	4.60	3.97	5.20	2.36	16.16	7.75	4.94	5.19	4.15	10.37	10.26	3.76	8.20	1.76
	B	10535	17.66	16.98	18.31	13.13	35.68	28.33	18.78	24.72	16.44	33.24	32.37	15.48	26.94	10.36
	P	34503	57.84	59.58	56.18	60.16	43.07	57.04	57.99	57.88	58.16	53.71	49.09	59.14	55.29	59.85
	A	11870	19.90	19.47	20.31	24.34	5.10	6.89	18.30	12.21	21.25	2.68	8.28	21.62	9.57	28.03
Total		59651	100.00	29052	30596	43198	6455	6520	2492	983	55322	4329	7692	51959	26274	33377
8	M	4849	8.00	7.01	8.93	4.81	24.93	14.60	6.43	11.25	7.25	21.82	28.80	4.96	14.53	3.59
	B	8680	14.31	14.53	14.10	10.77	29.26	24.06	16.32	19.88	13.28	33.34	27.87	12.33	22.05	9.10
	P	27773	45.79	48.76	42.97	46.53	38.36	46.90	48.10	47.69	46.05	40.92	34.19	47.48	46.26	45.47
	A	19352	31.91	29.70	34.00	37.89	7.45	14.44	29.14	21.19	33.42	3.92	9.13	35.23	17.16	41.84
Total		60654	100.00	29572	31082	45312	6353	5644	2347	996	57538	3116	7718	52936	24408	36246
10	M	8310	13.29	12.94	13.62	8.22	42.69	26.84	14.23	18.16	12.09	54.11	45.07	9.19	25.09	6.95
	B	6326	10.12	11.17	9.11	8.21	18.49	16.93	12.74	13.57	9.75	22.44	17.09	9.22	15.23	7.37
	P	20827	33.31	37.59	29.21	33.63	27.32	35.39	35.62	36.63	33.71	19.59	25.03	34.38	34.42	32.71
	A	27063	43.28	38.30	48.07	49.94	11.50	20.84	37.40	31.63	44.44	3.86	12.82	47.21	25.26	52.97
Total		62526	100.00	30618	31904	48338	5798	4985	2417	980	60739	1787	7140	55386	21863	40663

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-52
Cut Scores and Associated Impact Data for WKCE Reading

Grade	Score Range				Impact Data				
	Minimal	**Basic**	**Proficient**	**Advanced**	**Minimal**	**Basic**	**Proficient**	**Advanced**	**Proficient +Advanced**
3	270-444	445-474	475-506	507-640	31.51	33.78	26.64	8.07	34.71
4	280-461	462-497	498-535	536-650	31.46	35.93	26.58	6.04	32.62
5	290-463	464-502	503-545	546-690	27.43	38.05	29.18	5.34	34.52
6	300-478	479-524	525-572	573-730	24.82	40.79	29.70	4.69	34.40
7	310-485	486-534	535-589	590-780	21.65	41.54	33.22	3.58	36.80
8	330-492	493-547	548-612	613-790	18.24	41.75	35.65	4.36	40.01
10	350-495	496-565	566-643	644-820	18.00	43.53	34.42	4.04	38.47

Table 8-53
Cut Scores and Associated Impact Data for WKCE Mathematics

Grade	Score Range				Impact Data				
	Minimal	**Basic**	**Proficient**	**Advanced**	**Minimal**	**Basic**	**Proficient**	**Advanced**	**Proficient +Advanced**
3	220-387	388-437	438-491	492-630	13.78	38.43	39.65	8.14	47.79
4	240-424	425-473	474-525	526-650	13.28	37.96	40.24	8.52	48.76
5	270-448	449-500	501-552	553-680	13.04	36.16	39.26	11.54	50.80
6	310-474	475-523	524-572	573-700	14.40	33.55	37.09	14.95	52.04
7	330-499	500-543	544-590	591-710	17.47	35.45	36.38	10.69	47.07
8	350-509	510-557	558-604	605-730	20.16	35.03	33.18	11.63	44.81
10	410-527	528-573	574-617	618-750	18.39	36.95	34.46	10.20	44.66

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-54
Cut Scores and Associated Impact Data for WKCE Language Arts

Grade	Score Range				Impact Data				
	Minimal	**Basic**	**Proficient**	**Advanced**	**Minimal**	**Basic**	**Proficient**	**Advanced**	**Proficient +Advanced**
4	140-251	252-276	277-307	308-420	6.08	18.15	43.78	31.99	75.77
8	250-357	358-384	385-417	418-520	15.78	23.22	33.45	27.55	61.00
10	290-392	393-427	428-483	484-630	7.97	17.80	54.54	19.69	74.23

Table 8-55
Cut Scores and Associated Impact Data for WKCE Social Studies

Grade	Score Range				Impact Data				
	Minimal	**Basic**	**Proficient**	**Advanced**	**Minimal**	**Basic**	**Proficient**	**Advanced**	**Proficient +Advanced**
4	170-241	242-262	263-287	288-400	2.18	5.83	27.49	64.50	91.99
8	230-333	334-363	364-402	403-530	4.98	12.53	37.88	44.62	82.49
10	240-407	408-419	420-454	455-620	12.41	5.87	29.18	52.55	81.72

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-56
Cut Scores and Associated Impact Data for WKCE Science

Grade	Score Range				Impact Data				
	Minimal	**Basic**	**Proficient**	**Advanced**	**Minimal**	**Basic**	**Proficient**	**Advanced**	**Proficient +Advanced**
4	170-248	249-278	279-319	320-440	4.60	17.66	57.84	19.90	77.74
8	230-348	349-374	375-418	419-560	7.99	14.31	45.79	31.91	77.69
10	240-410	411-428	429-465	466-610	13.29	10.12	33.31	43.28	76.59

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-57
Summary Statistics for Reading Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
3	60163	1	Determines Meaning	12	0	12	8.00	0.67	2.73	66.53	20.69
	60163	2	Understands Text	17	0	17	12.63	0.74	3.78	74.54	21.05
	60163	3	Analyzes Text	21	1	24	15.86	0.66	4.96	66.20	19.96
	60163	4	Evaluates/Extends Text	4	0	4	2.63	0.66	1.19	64.22	21.97
4	59602	1	Determines Meaning	11	0	11	7.90	0.72	2.45	71.83	20.10
	59602	2	Understands Text	17	0	17	11.78	0.69	3.61	69.07	19.31
	59602	3	Analyzes Text	21	1	24	14.54	0.61	4.68	61.16	18.66
	59602	4	Evaluates/Extends Text	5	1	8	4.32	0.54	1.64	53.49	15.54
5	59214	1	Determines Meaning	11	0	11	7.53	0.68	2.61	68.13	21.53
	59214	2	Understands Text	15	0	15	11.74	0.78	2.96	78.12	18.36
	59214	3	Analyzes Text	20	1	23	13.86	0.60	4.17	60.23	16.96
	59214	4	Evaluates/Extends Text	8	1	11	6.14	0.56	2.23	56.84	16.49
6	60351	1	Determines Meaning	11	0	11	7.55	0.69	2.31	68.14	18.32
	60351	2	Understands Text	14	0	14	8.91	0.64	2.83	63.92	17.42
	60351	3	Analyzes Text	18	1	21	13.15	0.63	3.72	62.79	16.32
	60351	4	Evaluates/Extends Text	11	1	14	8.96	0.64	2.61	63.97	16.10
7	61187	1	Determines Meaning	11	0	11	8.04	0.73	2.15	72.57	16.85
	61187	2	Understands Text	14	0	14	9.47	0.68	2.69	67.66	16.68
	61187	3	Analyzes Text	18	1	21	12.54	0.60	3.82	60.17	16.70
	61187	4	Evaluates/Extends Text	11	1	14	9.09	0.65	2.62	65.05	15.94
8	60676	1	Determines Meaning	11	0	11	8.18	0.74	2.30	73.99	18.16
	60676	2	Understands Text	14	0	14	10.25	0.73	2.76	73.51	17.67
	60676	3	Analyzes Text	18	1	21	13.89	0.66	3.54	66.10	15.08
	60676	4	Evaluates/Extends Text	11	1	14	9.37	0.67	2.82	67.54	18.24

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-57 Cont'd
Summary Statistics for Reading Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
10	62736	1	Determines Meaning	7	0	7	5.16	0.74	1.51	73.32	16.92
	62736	2	Understands Text	7	0	7	4.91	0.70	1.48	70.21	15.78
	62736	3	Analyzes Text	22	1	25	16.26	0.65	4.85	65.36	18.32
	62736	4	Evaluates/Extends Text	14	1	17	11.03	0.65	3.56	65.05	19.77

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-58
Summary Statistics for Mathematics Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
3	60355	A	Mathematical Processes	3	3	9	3.55	0.39	2.32	39.81	23.40
	60355	B	Number Operations	11	1	12	9.15	0.76	2.43	76.45	18.14
	60355	C	Geometry	9	1	10	8.14	0.81	1.66	81.07	13.88
	60355	D	Measurement	8	0	8	6.19	0.77	1.53	74.74	15.17
	60355	E	Statistics/Probability	7	1	8	5.75	0.72	1.87	72.35	19.94
	60355	F	Algebraic Relationships	8	1	10	7.76	0.78	1.95	78.60	16.08
4	59770	A	Mathematical Processes	3	3	9	4.18	0.46	2.05	47.77	19.16
	59770	B	Number Operations	11	0	11	9.18	0.83	1.84	82.83	14.46
	59770	C	Geometry	8	1	10	7.72	0.77	1.78	77.13	13.34
	59770	D	Measurement	8	1	9	7.02	0.78	1.71	76.11	16.66
	59770	E	Statistics/Probability	7	1	8	5.16	0.64	1.70	63.44	17.55
	59770	F	Algebraic Relationships	9	1	10	7.80	0.78	2.07	79.04	18.09
5	59263	A	Mathematical Processes	3	3	9	5.28	0.59	2.11	59.67	19.39
	59263	B	Number Operations	11	0	11	8.76	0.80	1.92	79.04	14.54
	59263	C	Geometry	9	1	10	7.96	0.80	1.59	79.15	12.12
	59263	D	Measurement	9	1	11	7.28	0.66	2.49	66.76	19.40
	59263	E	Statistics/Probability	9	1	10	6.19	0.62	2.25	62.57	19.22
	59263	F	Algebraic Relationships	10	1	11	7.64	0.69	2.63	69.36	21.78

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-58 Cont'd
Summary Statistics for Mathematics Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
6	60409	A	Mathematical Processes	3	3	9	4.34	0.48	2.20	48.89	20.61
	60409	B	Number Operations	12	0	12	7.84	0.65	2.61	65.72	18.87
	60409	C	Geometry	9	1	10	8.00	0.80	1.86	78.30	15.44
	60409	D	Measurement	9	1	11	7.11	0.65	2.46	64.66	19.56
	60409	E	Statistics/Probability	8	1	9	6.10	0.68	2.17	68.38	20.41
	60409	F	Algebraic Relationships	10	1	11	8.09	0.74	2.27	74.31	17.94
7	61268	A	Mathematical Processes	3	3	9	4.53	0.50	2.19	50.87	20.99
	61268	B	Number Operations	12	0	12	8.32	0.69	2.66	69.28	19.46
	61268	C	Geometry	10	2	12	8.42	0.70	2.41	70.56	16.99
	61268	D	Measurement	9	0	9	6.67	0.74	1.78	73.98	16.21
	61268	E	Statistics/Probability	8	1	10	6.63	0.66	2.05	65.91	17.78
	61268	F	Algebraic Relationships	9	1	10	6.81	0.68	2.27	69.31	19.24
8	60718	A	Mathematical Processes	3	3	9	4.24	0.47	2.33	48.33	22.07
	60718	B	Number Operations	7	0	7	3.85	0.55	1.93	55.41	22.86
	60718	C	Geometry	8	1	9	6.28	0.70	1.94	69.44	17.19
	60718	D	Measurement	11	1	12	6.04	0.50	2.71	49.88	19.89
	60718	E	Statistics/Probability	8	1	9	5.50	0.61	2.03	60.14	18.84
	60718	F	Algebraic Relationships	14	1	16	10.19	0.64	3.20	63.86	18.47
10	60409	A	Mathematical Processes	7	1	9	5.41	0.60	2.37	60.57	23.82
	60409	B	Number Operations	7	0	7	4.70	0.67	1.87	67.25	22.91
	60409	C	Geometry	8	1	10	6.14	0.61	2.58	61.24	22.36
	60409	D	Measurement	9	1	11	5.92	0.54	3.04	54.28	25.08
	60409	E	Statistics/Probability	9	0	9	5.72	0.64	2.08	63.58	19.25
	60409	F	Algebraic Relationships	10	1	12	6.90	0.58	2.96	57.28	22.19

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-59
Summary Statistics for Language Arts Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
4	59592	B	Writing	19	0	19	13.62	0.72	3.59	71.85	18.34
	59592	D	Language	5	0	5	2.79	0.56	1.31	55.32	18.39
	59592	F	Research and Inquiry	6	0	6	3.16	0.53	1.47	53.20	18.29
8	60430	B	Writing	18	0	18	13.79	0.77	3.52	76.61	18.85
	60430	D	Language	6	0	6	3.57	0.59	1.68	60.07	23.76
	60430	F	Research and Inquiry	5	0	5	3.53	0.71	1.16	70.44	16.22
10	62465	B	Writing	15	2	24	14.72	0.61	3.68	61.34	14.70
	62465	D	Language	9	0	9	6.42	0.71	2.11	71.91	20.34
	62465	F	Research and Inquiry	6	0	6	3.82	0.64	1.42	63.77	17.04

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-60
Summary Statistics for Social Studies Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
4	59632	A	Geography	9	0	9	7.08	0.79	1.90	79.03	18.33
	59632	B	History	8	0	8	6.68	0.84	1.44	82.90	15.57
	59632	C	Political Science	7	0	7	4.66	0.67	1.58	66.38	17.10
	59632	D	Economics	7	0	7	5.29	0.76	1.29	76.08	14.62
	59632	E	Behavioral Science	7	0	7	5.25	0.75	1.66	75.12	20.20
8	60602	A	Geography	10	0	10	8.29	0.83	1.86	82.36	16.44
	60602	B	History	11	0	11	7.21	0.66	2.33	66.50	18.07
	60602	C	Political Science	6	0	6	3.93	0.66	1.55	65.38	19.82
	60602	D	Economics	8	0	8	5.55	0.69	1.79	69.02	18.64
	60602	E	Behavioral Science	5	0	5	3.03	0.61	1.35	61.47	20.47
10	62463	A	Geography	9	0	9	5.73	0.64	1.79	62.99	15.04
	62463	B	History	12	0	12	7.74	0.64	2.66	65.52	19.49
	62463	C	Political Science	12	0	12	7.41	0.62	2.75	61.82	20.36
	62463	D	Economics	8	0	8	5.29	0.66	1.89	66.41	20.17
	62463	E	Behavioral Science	8	0	8	6.24	0.78	1.63	77.41	17.08

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-61
Summary Statistics for Science Content Standards Raw and SPI Scores

Grade	N	Content Standard	Standard	No. of Items		Total Score Points	Mean	Mean *p*-value	SD	SPI	
				MC	CR					Mean	SD
4	59651	A/B	Connections & Nature of Sci	8	0	8	5.63	0.70	1.69	70.95	17.41
	59651	C	Science Inquiry	7	0	7	4.89	0.70	1.57	70.53	18.17
	59651	D	Physical Science	6	0	6	4.56	0.76	0.96	77.17	9.87
	59651	E	Earth and Space	5	0	5	3.56	0.71	1.15	69.72	16.37
	59651	F	Life and Environment	6	0	6	4.39	0.73	1.38	71.62	17.79
	59651	G/H	Appl & Social Perspectives	7	0	7	5.33	0.76	1.61	76.18	19.09
8	60654	A/B	Connections & Nature of Sci	7	0	7	4.96	0.71	1.64	70.57	19.37
	60654	C	Science Inquiry	8	0	8	6.44	0.81	1.49	80.39	15.19
	60654	D	Physical Science	6	0	6	4.57	0.76	1.16	75.83	14.05
	60654	E	Earth and Space	6	0	6	3.96	0.66	1.46	65.90	18.69
	60654	F	Life and Environment	6	0	6	4.22	0.70	1.37	71.31	16.77
	60654	G/H	Appl & Social Perspectives	7	0	7	5.74	0.82	1.40	82.11	16.58
10	62526	A/B	Connections & Nature of Sci	10	0	10	6.56	0.66	2.20	65.72	19.19
	62526	C	Science Inquiry	9	0	9	5.56	0.62	2.01	61.71	18.17
	62526	D	Physical Science	8	0	8	5.13	0.64	1.85	63.08	18.32
	62526	E	Earth and Space	6	0	6	3.52	0.59	1.30	57.79	13.88
	62526	F	Life and Environment	6	0	6	3.47	0.58	1.28	60.40	15.56
	62526	G/H	Appl & Social Perspectives	11	0	11	7.21	0.66	2.59	65.40	20.69

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-62
SPI Cut Scores

Content	Content Standard	Performance Level	SPI Cut Score Ranges						
			Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 10
RD	Standard 1 Determines Meaning	1	0-56	0-66	0-55	0-55	0-59	0-56	0-58
		2	57-78	67-84	56-80	56-78	60-80	57-82	59-81
		3	79-91	85-93	81-94	79-91	81-95	83-95	82-94
		4	92-100	94-100	95-100	92-100	96-100	96-100	95-100
	Standard 2 Understands Text	1	0-68	0-61	0-70	0-52	0-54	0-57	0-55
		2	69-87	62-81	71-89	53-72	55-75	58-81	56-76
		3	88-95	82-91	90-97	73-88	76-90	82-95	77-91
		4	96-100	92-100	98-100	89-100	91-100	96-100	92-100
	Standard 3 Analyzes Text	1	0-58	0-52	0-49	0-52	0-46	0-52	0-47
		2	59-77	53-72	50-68	53-70	47-67	53-71	48-72
		3	78-89	73-85	69-83	71-84	68-84	72-86	73-91
		4	90-100	86-100	84-100	85-100	85-100	87-100	92-100
	Standard 4 Evaluates/Extends Text	1	0-51	0-46	0-46	0-53	0-52	0-50	0-45
		2	52-75	47-60	47-64	54-72	53-72	51-74	46-74
		3	76-93	61-74	65-80	73-84	73-86	75-91	75-91
		4	94-100	75-100	81-100	85-100	87-100	92-100	92-100
MA	Standard A Mathematical Processes	1	0-14	0-25	0-37	0-24	0-28	0-27	0-35
		2	15-35	26-47	38-60	25-47	29-53	28-51	36-66
		3	36-76	48-73	61-82	48-71	54-76	52-75	67-90
		4	77-100	74-100	83-100	72-100	77-100	76-100	91-100
	Standard B Number Operations	1	0-54	0-67	0-63	0-43	0-48	0-31	0-43
		2	55-82	68-86	64-81	44-65	49-73	32-58	44-74
		3	83-95	87-96	82-93	66-86	74-92	59-85	75-93
		4	96-100	97-100	94-100	87-100	93-100	86-100	94-100
	Standard C Geometry	1	0-66	0-62	0-66	0-60	0-54	0-55	0-36
		2	67-85	63-79	67-80	61-80	55-73	56-74	37-67
		3	86-94	80-91	81-91	81-92	74-90	75-87	68-87
		4	95-100	92-100	92-100	93-100	91-100	88-100	88-100
	Standard D Measurement	1	0-57	0-55	0-42	0-41	0-57	0-30	0-28
		2	58-77	56-79	43-69	42-64	58-77	31-49	29-54
		3	78-91	80-94	70-88	65-85	78-92	50-76	55-90
		4	92-100	95-100	89-100	86-100	93-100	77-100	91-100
	Standard E Statistics/Probability	1	0-48	0-43	0-38	0-44	0-48	0-42	0-44
		2	49-77	44-62	39-64	45-69	49-67	43-62	45-66
		3	78-94	63-87	65-84	70-89	68-87	63-82	67-89
		4	95-100	88-100	85-100	90-100	88-100	83-100	90-100
	Standard F Algebraic Relationships	1	0-61	0-57	0-40	0-54	0-49	0-48	0-33
		2	62-83	58-85	41-72	55-76	50-75	49-68	34-60
		3	84-95	86-96	73-93	77-92	76-90	69-84	61-86
		4	96-100	97-100	94-100	93-100	91-100	85-100	87-100

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-62 Cont'd
SPI Cut Scores

Content	Content Standard	Performance Level	SPI Cut Score Ranges						
			Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 10
LA	Standard B Writing	1		0-39				0-56	0-38
		2		40-58				57-75	39-52
		3		59-83				76-90	53-73
		4		84-100				91-100	74-100
	Standard D Language	1		0-27				0-32	0-37
		2		28-41				33-52	38-59
		3		42-63				53-77	60-90
		4		64-100				78-100	91-100
	Standard F Research and Inquiry	1		0-27				0-54	0-36
		2		28-37				55-67	37-53
		3		38-61				68-81	54-78
		4		62-100				82-100	79-100
SS	Standard A Geography	1		0-30				0-47	0-44
		2		31-49				48-68	45-49
		3		50-77				69-89	50-63
		4		78-100				90-100	64-100
	Standard B History	1		0-38				0-36	0-40
		2		39-58				37-48	41-45
		3		59-81				49-70	46-65
		4		82-100				71-100	66-100
	Standard C Political Science	1		0-31				0-31	0-35
		2		32-40				32-45	36-41
		3		41-60				46-69	42-60
		4		61-100				70-100	61-100
	Standard D Economics	1		0-40				0-37	0-39
		2		41-55				38-49	40-46
		3		56-71				50-72	47-67
		4		72-100				73-100	68-100
	Standard E Behavioral Science	1		0-26				0-26	0-56
		2		27-40				27-39	57-62
		3		41-71				40-66	63-80
		4		72-100				67-100	81-100

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 8-62 Cont'd
SPI Cut Scores

Content	Content Standard	Performance Level	SPI Cut Score Ranges						
			Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 10
SC	Standard A/B Connections & Nature of Science	1		0-35				0-38	0-41
		2		36-58				39-55	42-51
		3		59-85				56-82	52-71
		4		86-100				83-100	72-100
	Standard C Science Inquiry	1		0-37				0-56	0-39
		2		38-56				57-70	40-47
		3		57-87				71-89	48-66
		4		88-100				90-100	67-100
	Standard D Physical Science	1		0-59				0-54	0-41
		2		60-71				55-66	42-48
		3		72-84				67-82	49-66
		4		85-100				83-100	67-100
	Standard E Earth and Space	1		0-39				0-38	0-42
		2		40-57				39-50	43-47
		3		58-83				51-75	48-59
		4		84-100				76-100	60-100
	Standard F Life and Environment	1		0-37				0-46	0-42
		2		38-57				47-57	43-50
		3		58-87				58-81	51-63
		4		88-100				82-100	64-100
	Standard G/H Science Applications & Social Perspectives	1		0-36				0-55	0-39
		2		37-63				56-71	40-48
		3		64-92				72-92	49-71
		4		93-100				93-100	72-100

Copyright © 2013 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	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
Reading	3	0.93	0.93	0.94	0.93	0.92	0.92	0.93	0.93	0.93	0.90	0.93	0.92	0.93	0.92
	4	0.92	0.91	0.92	0.90	0.91	0.91	0.91	0.91	0.91	0.88	0.92	0.90	0.91	0.90
	5	0.91	0.91	0.91	0.90	0.91	0.90	0.91	0.89	0.91	0.87	0.91	0.89	0.91	0.89
	6	0.90	0.89	0.90	0.88	0.89	0.89	0.91	0.89	0.89	0.83	0.89	0.87	0.89	0.87
	7	0.89	0.89	0.90	0.88	0.89	0.88	0.90	0.89	0.89	0.83	0.89	0.87	0.89	0.87
	8	0.90	0.89	0.91	0.89	0.90	0.89	0.90	0.90	0.90	0.84	0.89	0.88	0.90	0.88
	10	0.91	0.90	0.91	0.89	0.90	0.90	0.92	0.90	0.90	0.83	0.89	0.89	0.90	0.89
Mathematics	3	0.90	0.90	0.91	0.89	0.91	0.90	0.90	0.90	0.90	0.89	0.91	0.90	0.90	0.88
	4	0.90	0.89	0.90	0.87	0.90	0.89	0.90	0.88	0.89	0.88	0.91	0.88	0.89	0.87
	5	0.91	0.90	0.91	0.90	0.90	0.89	0.91	0.89	0.91	0.87	0.91	0.89	0.90	0.89
	6	0.92	0.92	0.92	0.91	0.91	0.90	0.92	0.90	0.92	0.88	0.91	0.91	0.91	0.90
	7	0.92	0.92	0.92	0.91	0.90	0.90	0.93	0.90	0.92	0.87	0.90	0.90	0.91	0.91
	8	0.92	0.91	0.92	0.91	0.89	0.90	0.92	0.90	0.92	0.86	0.89	0.91	0.90	0.91
	10	0.94	0.93	0.94	0.93	0.90	0.91	0.94	0.92	0.93	0.84	0.89	0.93	0.92	0.93
Language Arts	4	0.82	0.82	0.82	0.81	0.80	0.79	0.82	0.80	0.82	0.75	0.79	0.81	0.80	0.80
	8	0.85	0.84	0.85	0.84	0.83	0.83	0.84	0.83	0.85	0.76	0.81	0.83	0.83	0.84
	10	0.84	0.82	0.84	0.82	0.82	0.81	0.84	0.82	0.83	0.73	0.79	0.81	0.83	0.81
Social Studies	4	0.87	0.87	0.88	0.85	0.88	0.86	0.87	0.86	0.87	0.84	0.87	0.86	0.87	0.84
	8	0.87	0.86	0.88	0.86	0.85	0.85	0.87	0.85	0.87	0.77	0.85	0.86	0.86	0.85
	10	0.90	0.89	0.91	0.89	0.88	0.88	0.90	0.89	0.90	0.79	0.88	0.89	0.89	0.89
Science	4	0.86	0.85	0.86	0.83	0.84	0.82	0.86	0.84	0.85	0.80	0.85	0.85	0.85	0.82
	8	0.86	0.85	0.87	0.84	0.84	0.83	0.84	0.84	0.86	0.77	0.85	0.84	0.85	0.83
	10	0.88	0.86	0.90	0.87	0.84	0.85	0.88	0.87	0.88	0.75	0.86	0.87	0.87	0.87

Copyright © 2013 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	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Fully English Proficient	Limited English Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
Reading	3	2.97	2.91	3.02	2.87	3.26	3.23	2.99	3.14	2.94	3.31	3.28	2.92	3.17	2.79
	4	3.19	3.16	3.21	3.11	3.42	3.38	3.24	3.30	3.16	3.48	3.41	3.15	3.34	3.06
	5	3.16	3.13	3.18	3.08	3.41	3.36	3.20	3.29	3.14	3.49	3.43	3.11	3.33	3.02
	6	3.20	3.18	3.21	3.14	3.40	3.36	3.25	3.30	3.18	3.48	3.44	3.16	3.35	3.08
	7	3.20	3.17	3.21	3.14	3.39	3.35	3.24	3.31	3.18	3.49	3.44	3.15	3.33	3.09
	8	3.14	3.07	3.18	3.05	3.45	3.35	3.18	3.28	3.11	3.53	3.50	3.06	3.33	2.98
	10	3.10	3.08	3.10	3.04	3.32	3.28	3.17	3.19	3.09	3.41	3.33	3.06	3.24	3.01
Mathematics	3	2.96	2.96	2.96	2.91	3.12	3.07	2.94	3.03	2.95	3.10	3.13	2.93	3.06	2.86
	4	2.88	2.88	2.87	2.81	3.09	3.02	2.85	2.98	2.86	3.06	3.08	2.84	3.00	2.76
	5	3.22	3.24	3.20	3.15	3.41	3.37	3.18	3.35	3.21	3.44	3.42	3.18	3.36	3.09
	6	3.21	3.22	3.20	3.15	3.39	3.37	3.15	3.33	3.19	3.45	3.39	3.17	3.35	3.08
	7	3.19	3.20	3.18	3.12	3.41	3.35	3.15	3.34	3.17	3.43	3.39	3.15	3.34	3.06
	8	3.36	3.37	3.35	3.34	3.40	3.44	3.36	3.42	3.36	3.45	3.39	3.35	3.42	3.31
	10	3.27	3.30	3.23	3.23	3.33	3.38	3.27	3.36	3.26	3.34	3.32	3.25	3.36	3.20
Language Arts	4	2.26	2.21	2.30	2.22	2.40	2.36	2.25	2.35	2.25	2.41	2.43	2.23	2.36	2.18
	8	2.12	2.04	2.20	2.07	2.33	2.27	2.11	2.26	2.11	2.38	2.41	2.08	2.27	2.01
	10	2.55	2.47	2.61	2.49	2.76	2.67	2.59	2.68	2.54	2.79	2.77	2.50	2.69	2.46
Social Studies	4	2.27	2.25	2.29	2.19	2.54	2.44	2.30	2.41	2.25	2.51	2.50	2.23	2.44	2.13
	8	2.54	2.54	2.52	2.47	2.80	2.72	2.55	2.67	2.52	2.84	2.81	2.49	2.70	2.41
	10	2.85	2.86	2.83	2.79	3.08	3.03	2.88	2.97	2.84	3.16	3.10	2.81	3.01	2.75
Science	4	2.40	2.39	2.41	2.32	2.65	2.58	2.42	2.51	2.38	2.65	2.60	2.37	2.55	2.27
	8	2.43	2.44	2.41	2.34	2.74	2.64	2.47	2.56	2.41	2.79	2.75	2.38	2.61	2.29
	10	3.00	3.03	2.95	2.95	3.18	3.16	3.03	3.09	2.99	3.23	3.18	2.97	3.13	2.92

Copyright © 2013 by the Wisconsin Department of Public Instruction

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

Content Area	Grade	Alpha Per Content Standard								
		A/1	A/B	B/2	C/3	D/4	E	F	G/H	Total
Reading	3	0.73		0.84	0.85	0.51				0.93
	4	0.73		0.79	0.81	0.42				0.92
	5	0.73		0.78	0.77	0.56				0.91
	6	0.66		0.69	0.73	0.67				0.90
	7	0.66		0.68	0.73	0.65				0.89
	8	0.69		0.74	0.71	0.68				0.90
	10	0.52		0.48	0.81	0.78				0.91
Mathematics	3	0.63		0.73	0.57	0.54	0.66	0.59		0.90
	4	0.55		0.67	0.52	0.62	0.59	0.70		0.90
	5	0.53		0.63	0.49	0.67	0.65	0.76		0.91
	6	0.57		0.69	0.64	0.70	0.71	0.70		0.92
	7	0.60		0.72	0.68	0.62	0.64	0.68		0.92
	8	0.64		0.64	0.62	0.69	0.61	0.75		0.92
	10	0.72		0.69	0.71	0.77	0.63	0.75		0.94
Language Arts	4			0.78		0.44		0.44		0.82
	8			0.80		0.61		0.34		0.85
	10			0.71		0.68		0.45		0.84
Social Studies	4	0.68		0.60	0.47	0.48	0.64			0.87
	8	0.68		0.64	0.56	0.60	0.47			0.87
	10	0.53		0.70	0.72	0.65	0.61			0.90
Science	4		0.57		0.56	0.22	0.40	0.50	0.64	0.86
	8		0.56		0.56	0.40	0.48	0.46	0.59	0.86
	10		0.64		0.59	0.57	0.33	0.37	0.72	0.88

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-4
Standard Error of Measurement per Content Standard

Content Area	Grade	SEM Per Content Standard								
		A/1	A/B	B/2	C/3	D/4	E	F	G/H	Total
Reading	3	1.42		1.51	1.92	0.83				1.42
	4	1.27		1.65	2.04	1.25				1.27
	5	1.36		1.39	2.00	1.48				1.36
	6	1.35		1.58	1.93	1.50				1.35
	7	1.25		1.52	1.98	1.55				1.25
	8	1.28		1.41	1.91	1.60				1.28
	10	1.05		1.07	2.11	1.67				1.05
Mathematics	3	1.41		1.26	1.09	1.04	1.09	1.25		1.41
	4	1.38		1.06	1.23	1.05	1.09	1.13		1.38
	5	1.45		1.17	1.14	1.43	1.33	1.29		1.45
	6	1.44		1.45	1.12	1.35	1.17	1.24		1.44
	7	1.39		1.41	1.36	1.10	1.23	1.28		1.39
	8	1.40		1.16	1.20	1.51	1.27	1.60		1.40
	10	1.25		1.04	1.39	1.46	1.27	1.48		1.25
Language Arts	4			1.68		0.98		1.10		
	8			1.57		1.05		0.94		
	10			1.98		1.19		1.05		
Social Studies	4	1.07		0.91	1.15	0.93	1.00			1.07
	8	1.05		1.40	1.03	1.13	0.98			1.05
	10	1.23		1.46	1.46	1.12	1.02			1.23
Science	4		1.11		1.04	0.85	0.89	0.98	0.97	
	8		1.09		0.99	0.90	1.05	1.01	0.90	
	10		1.32		1.29	1.21	1.06	1.02	1.37	

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-5
Classification Consistency and Classification Accuracy for Reading Grade 3

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.28	0.04	0.00	0.00	0.31
Basic Proficient	0.04	0.24	0.05	0.00	0.33
Proficient	0.00	0.06	0.17	0.03	0.25
Advanced	0.00	0.00	0.03	0.07	0.10
Sum	0.32	0.34	0.24	0.10	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.89	0.93	0.75
Probability of Chance	0.57	0.54	0.82	0.28
Kappa (k)	0.83	0.77	0.64	0.66
Classification Accuracy	0.95	0.93	0.95	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-6
Classification Consistency and Classification Accuracy for Reading Grade 4

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.27	0.05	0.00	0.00	0.32
Basic Proficient	0.04	0.25	0.05	0.00	0.35
Proficient	0.00	0.06	0.16	0.03	0.26
Advanced	0.00	0.00	0.03	0.05	0.08
Sum	0.31	0.36	0.24	0.08	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.91	0.88	0.94	0.73
Probability of Chance	0.57	0.56	0.85	0.29
Kappa (k)	0.79	0.74	0.58	0.62
Classification Accuracy	0.94	0.92	0.96	0.81

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-7
Classification Consistency and Classification Accuracy for Reading Grade 5

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.23	0.06	0.00	0.00	0.29
Basic Proficient	0.07	0.35	0.06	0.00	0.48
Proficient	0.00	0.07	0.12	0.01	0.21
Advanced	0.00	0.00	0.01	0.01	0.02
Sum	0.30	0.49	0.19	0.02	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.87	0.86	0.97	0.71
Probability of Chance	0.58	0.65	0.96	0.36
Kappa (k)	0.69	0.60	0.38	0.54
Classification Accuracy	0.91	0.85	0.96	0.72

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-8
Classification Consistency and Classification Accuracy for Reading Grade 6

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.21	0.04	0.00	0.00	0.25
Basic Proficient	0.04	0.29	0.06	0.00	0.39
Proficient	0.00	0.07	0.20	0.03	0.29
Advanced	0.00	0.00	0.03	0.04	0.06
Sum	0.25	0.39	0.29	0.06	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.87	0.94	0.73
Probability of Chance	0.62	0.54	0.88	0.31
Kappa (k)	0.79	0.71	0.53	0.62
Classification Accuracy	0.94	0.91	0.96	0.81

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-9
Classification Consistency and Classification Accuracy for Reading Grade 7

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.18	0.04	0.00	0.00	0.22
Basic Proficient	0.04	0.30	0.06	0.00	0.40
Proficient	0.00	0.06	0.24	0.03	0.33
Advanced	0.00	0.00	0.02	0.03	0.05
Sum	0.22	0.40	0.32	0.06	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.88	0.95	0.75
Probability of Chance	0.65	0.53	0.90	0.32
Kappa (k)	0.78	0.75	0.51	0.64
Classification Accuracy	0.95	0.91	0.97	0.82

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-10
Classification Consistency and Classification Accuracy for Reading Grade 8

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.15	0.03	0.00	0.00	0.19
Basic Proficient	0.04	0.30	0.06	0.00	0.40
Proficient	0.00	0.07	0.25	0.03	0.35
Advanced	0.00	0.00	0.03	0.04	0.06
Sum	0.19	0.41	0.34	0.06	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.87	0.94	0.74
Probability of Chance	0.70	0.52	0.88	0.32
Kappa (k)	0.76	0.72	0.52	0.61
Classification Accuracy	0.95	0.91	0.96	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-11
Classification Consistency and Classification Accuracy for Reading Grade 10

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.15	0.03	0.00	0.00	0.19
Basic Proficient	0.04	0.34	0.06	0.00	0.44
Proficient	0.00	0.05	0.24	0.02	0.32
Advanced	0.00	0.00	0.02	0.03	0.05
Sum	0.19	0.42	0.33	0.05	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.88	0.95	0.76
Probability of Chance	0.69	0.53	0.90	0.33
Kappa (k)	0.76	0.75	0.53	0.65
Classification Accuracy	0.95	0.92	0.97	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-12
Classification Consistency and Classification Accuracy for Mathematics Grade 3

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.12	0.03	0.00	0.00	0.14
Basic Proficient	0.03	0.29	0.06	0.00	0.38
Proficient	0.00	0.06	0.28	0.03	0.37
Advanced	0.00	0.00	0.04	0.07	0.10
Sum	0.15	0.38	0.38	0.10	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.88	0.93	0.76
Probability of Chance	0.75	0.50	0.82	0.32
Kappa (k)	0.77	0.77	0.62	0.65
Classification Accuracy	0.96	0.91	0.95	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-13
Classification Consistency and Classification Accuracy for Mathematics Grade 4

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.11	0.03	0.00	0.00	0.14
Basic Proficient	0.03	0.27	0.07	0.00	0.37
Proficient	0.00	0.06	0.28	0.04	0.38
Advanced	0.00	0.00	0.04	0.07	0.11
Sum	0.15	0.36	0.39	0.11	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.87	0.93	0.73
Probability of Chance	0.75	0.50	0.81	0.31
Kappa (k)	0.74	0.74	0.62	0.61
Classification Accuracy	0.96	0.91	0.95	0.82

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-14
Classification Consistency and Classification Accuracy for Mathematics Grade 5

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.11	0.03	0.00	0.00	0.14
Basic Proficient	0.03	0.27	0.05	0.00	0.35
Proficient	0.00	0.05	0.28	0.04	0.38
Advanced	0.00	0.00	0.04	0.09	0.13
Sum	0.14	0.35	0.38	0.13	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.89	0.92	0.75
Probability of Chance	0.76	0.50	0.77	0.30
Kappa (k)	0.74	0.78	0.65	0.64
Classification Accuracy	0.96	0.92	0.95	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-15
Classification Consistency and Classification Accuracy for Mathematics Grade 6

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.12	0.03	0.00	0.00	0.15
Basic Proficient	0.03	0.25	0.05	0.00	0.33
Proficient	0.00	0.05	0.26	0.04	0.35
Advanced	0.00	0.00	0.05	0.12	0.17
Sum	0.15	0.34	0.36	0.16	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.89	0.92	0.75
Probability of Chance	0.75	0.50	0.73	0.29
Kappa (k)	0.75	0.78	0.69	0.64
Classification Accuracy	0.96	0.93	0.94	0.82

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-16
Classification Consistency and Classification Accuracy for Mathematics Grade 7

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.15	0.03	0.00	0.00	0.18
Basic Proficient	0.03	0.27	0.05	0.00	0.36
Proficient	0.00	0.06	0.26	0.03	0.35
Advanced	0.00	0.00	0.04	0.09	0.12
Sum	0.18	0.35	0.35	0.12	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.89	0.93	0.76
Probability of Chance	0.71	0.50	0.79	0.29
Kappa (k)	0.79	0.78	0.68	0.66
Classification Accuracy	0.96	0.92	0.95	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-17
Classification Consistency and Classification Accuracy for Mathematics Grade 8

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.17	0.04	0.00	0.00	0.21
Basic Proficient	0.04	0.26	0.05	0.00	0.34
Proficient	0.00	0.05	0.24	0.03	0.32
Advanced	0.00	0.00	0.03	0.10	0.13
Sum	0.21	0.35	0.32	0.13	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.90	0.94	0.76
Probability of Chance	0.67	0.51	0.78	0.28
Kappa (k)	0.78	0.79	0.72	0.67
Classification Accuracy	0.95	0.92	0.96	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-18
Classification Consistency and Classification Accuracy for Mathematics Grade 10

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.15	0.03	0.00	0.00	0.19
Basic Proficient	0.04	0.29	0.04	0.00	0.36
Proficient	0.00	0.04	0.26	0.03	0.34
Advanced	0.00	0.00	0.03	0.08	0.11
Sum	0.19	0.36	0.33	0.11	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.91	0.94	0.79
Probability of Chance	0.69	0.51	0.80	0.29
Kappa (k)	0.77	0.83	0.71	0.70
Classification Accuracy	0.95	0.94	0.96	0.86

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-19
Classification Consistency and Classification Accuracy for Language Arts Grade 4

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.04	0.02	0.00	0.00	0.07
Basic Proficient	0.03	0.09	0.06	0.00	0.17
Proficient	0.00	0.06	0.28	0.08	0.42
Advanced	0.00	0.00	0.08	0.26	0.34
Sum	0.07	0.18	0.41	0.33	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.87	0.85	0.67
Probability of Chance	0.86	0.63	0.55	0.32
Kappa (k)	0.58	0.66	0.65	0.52
Classification Accuracy	0.96	0.91	0.89	0.77

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-20
Classification Consistency and Classification Accuracy for Language Arts Grade 8

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.12	0.04	0.00	0.00	0.16
Basic Proficient	0.04	0.12	0.06	0.00	0.22
Proficient	0.00	0.06	0.19	0.06	0.32
Advanced	0.00	0.00	0.07	0.22	0.30
Sum	0.16	0.22	0.33	0.29	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.87	0.86	0.66
Probability of Chance	0.73	0.53	0.59	0.27
Kappa (k)	0.69	0.72	0.66	0.53
Classification Accuracy	0.94	0.90	0.90	0.74

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-21
Classification Consistency and Classification Accuracy for Language Arts Grade 10

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.06	0.03	0.00	0.00	0.09
Basic Proficient	0.03	0.10	0.05	0.00	0.19
Proficient	0.00	0.05	0.39	0.07	0.51
Advanced	0.00	0.00	0.06	0.15	0.22
Sum	0.09	0.18	0.51	0.22	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.89	0.87	0.71
Probability of Chance	0.84	0.61	0.66	0.35
Kappa (k)	0.64	0.73	0.61	0.55
Classification Accuracy	0.96	0.92	0.91	0.79

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-22
Classification Consistency and Classification Accuracy for Social Studies Grade 4

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.02	0.01	0.00	0.00	0.02
Basic Proficient	0.01	0.03	0.02	0.00	0.06
Proficient	0.00	0.02	0.19	0.06	0.27
Advanced	0.00	0.00	0.07	0.58	0.64
Sum	0.03	0.06	0.28	0.64	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.98	0.95	0.87	0.81
Probability of Chance	0.95	0.84	0.54	0.49
Kappa (k)	0.68	0.71	0.73	0.64
Classification Accuracy	0.99	0.97	0.91	0.86

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-23
Classification Consistency and Classification Accuracy for Social Studies Grade 8

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.04	0.02	0.00	0.00	0.06
Basic Proficient	0.02	0.07	0.04	0.00	0.13
Proficient	0.00	0.04	0.25	0.07	0.36
Advanced	0.00	0.00	0.07	0.38	0.46
Sum	0.05	0.13	0.37	0.45	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.96	0.92	0.86	0.74
Probability of Chance	0.90	0.70	0.50	0.36
Kappa (k)	0.61	0.72	0.71	0.59
Classification Accuracy	0.97	0.94	0.90	0.81

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-24
Classification Consistency and Classification Accuracy for Social Studies Grade 10

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.10	0.02	0.02	0.00	0.13
Basic Proficient	0.02	0.02	0.03	0.00	0.06
Proficient	0.01	0.03	0.17	0.06	0.27
Advanced	0.00	0.00	0.06	0.47	0.53
Sum	0.13	0.06	0.28	0.53	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.91	0.88	0.76
Probability of Chance	0.77	0.69	0.50	0.38
Kappa (k)	0.73	0.72	0.75	0.61
Classification Accuracy	0.96	0.94	0.91	0.83

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-25
Classification Consistency and Classification Accuracy for Science Grade 4

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.03	0.02	0.00	0.00	0.05
Basic Proficient	0.02	0.11	0.05	0.00	0.17
Proficient	0.00	0.05	0.43	0.07	0.55
Advanced	0.00	0.00	0.06	0.17	0.22
Sum	0.05	0.18	0.54	0.23	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.96	0.90	0.88	0.74
Probability of Chance	0.90	0.65	0.65	0.38
Kappa (k)	0.62	0.72	0.65	0.58
Classification Accuracy	0.98	0.93	0.91	0.82

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-26
Classification Consistency and Classification Accuracy for Science Grade 8

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.06	0.02	0.00	0.00	0.09
Basic Proficient	0.02	0.07	0.05	0.00	0.14
Proficient	0.00	0.05	0.31	0.08	0.44
Advanced	0.00	0.00	0.07	0.26	0.33
Sum	0.09	0.14	0.43	0.34	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.95	0.90	0.86	0.71
Probability of Chance	0.84	0.65	0.55	0.33
Kappa (k)	0.66	0.71	0.68	0.56
Classification Accuracy	0.96	0.93	0.90	0.79

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-27
Classification Consistency and Classification Accuracy for Science Grade 10

Contingency Table with All Cut Scores

	Minimal Performance	Basic Proficient	Proficient	Advanced	Sum
Minimal Performance	0.10	0.03	0.01	0.00	0.14
Basic Proficient	0.03	0.04	0.04	0.00	0.10
Proficient	0.01	0.03	0.21	0.07	0.33
Advanced	0.00	0.00	0.06	0.37	0.43
Sum	0.14	0.10	0.31	0.45	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.91	0.87	0.72
Probability of Chance	0.76	0.63	0.51	0.32
Kappa (k)	0.68	0.74	0.73	0.59
Classification Accuracy	0.94	0.93	0.90	0.79

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-28
Inter-Rater Reliability, Reading*

			Percentage Absolute Difference								Frequency			
Grade	Item No.	Max	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	No. of Reads	0	1	2	3
3	19	3	0.76	0.14	0.01	0.08	0.93	0.87	0.65	6,250	3,122	2,196	915	17
3	31	3	0.76	0.20	0.01	0.03	0.91	0.81	1.33	6,250	947	2,421	2,775	107
4	13	3	0.71	0.26	0.01	0.02	0.88	0.75	1.48	6,372	697	2,290	3,020	365
4	47	3	0.62	0.32	0.04	0.02	0.83	0.66	1.28	6,272	1,166	2,715	2,048	443
5	13	3	0.65	0.31	0.02	0.02	0.84	0.69	1.49	6,180	759	2,071	2,918	432
5	47	3	0.68	0.29	0.01	0.01	0.86	0.73	1.42	6,180	817	2,199	2,902	262
6	18	3	0.64	0.34	0.02	0.01	0.79	0.58	1.29	6,278	686	3,243	2,197	152
6	24	3	0.77	0.21	0.01	0.01	0.92	0.84	1.10	6,278	1,825	2,250	1,972	231
7	31	3	0.65	0.32	0.01	0.02	0.82	0.63	1.54	6,420	444	2,457	3,136	383
7	50	3	0.67	0.29	0.02	0.02	0.89	0.78	1.05	6,420	2,277	1,910	1,879	354
8	19	3	0.66	0.32	0.01	0.01	0.85	0.69	1.63	6,284	403	2,097	3,180	604
8	56	3	0.63	0.32	0.02	0.03	0.88	0.75	1.63	6,284	965	1,237	3,220	862
10	7	3	0.70	0.23	0.01	0.06	0.93	0.86	1.33	6,436	1,626	1,860	2,157	793
10	21	3	0.57	0.36	0.03	0.04	0.86	0.71	1.14	6,436	1,772	2,588	1,509	567

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-29
Inter-Rater Reliability, Mathematics*

			Percentage Absolute Difference								Frequency		
Grade	Item No.	Max	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	No. of Reads	0	1	2
3	5	2	0.90	0.09	0.01	0.01	0.97	0.94	0.92	6,250	2,941	847	2,462
3	20A	1	0.87	0.02	0.03	0.08	0.97	0.94	0.43	6,250	3,550	2,700	
3	20B	2	0.84	0.11	0.01	0.03	0.94	0.88	0.49	6,250	4,083	1,299	868
3	27A	1	0.95	0.02	0.00	0.02	0.98	0.96	0.62	6,250	2,358	3,892	
3	27B	2	0.80	0.12	0.04	0.04	0.91	0.82	0.60	6,250	3,963	807	1,480
3	39A	1	0.92	0.01	0.01	0.06	0.99	0.98	0.37	6,250	3,911	2,339	
3	39B	2	0.87	0.09	0.02	0.00	0.95	0.90	0.66	6,250	3,581	1,223	1,446
4	11A	1	0.98	0.01	0.00	0.01	0.98	0.96	0.21	6,372	5,049	1,323	
4	11B	2	0.83	0.13	0.01	0.03	0.92	0.95	0.65	6,372	3,006	2,608	758
4	22	2	0.95	0.04	0.00	0.01	0.98	0.95	1.60	6,372	491	1,582	4,299
4	25A	1	0.99	0.00	0.00	0.01	1.00	0.99	0.47	6,372	3,400	2,972	
4	25B	2	0.80	0.15	0.02	0.02	0.90	0.79	0.63	6,372	3,379	1,982	1,011
4	39A	1	0.98	0.01	0.00	0.01	0.99	0.98	0.58	6,372	2,690	3,682	
4	39B	2	0.89	0.08	0.01	0.02	0.96	0.92	0.67	6,372	3,826	841	1,705
5	12	2	0.96	0.02	0.00	0.02	0.99	0.99	1.32	6,180	1,865	494	3,821
5	20A	1	0.99	0.00	0.00	0.01	0.99	0.98	0.84	6,180	972	5,208	
5	20B	2	0.87	0.10	0.01	0.02	0.95	0.90	0.57	6,180	3,742	1,338	1,100
5	27A	1	0.99	0.00	0.00	0.01	1.00	0.99	0.59	6,180	2,509	3,671	
5	27B	2	0.84	0.12	0.02	0.01	0.93	0.87	1.41	6,180	1,457	751	3,972
5	46A	1	0.99	0.00	0.00	0.01	1.00	1.00	0.36	6,180	3,957	2,223	
5	46B	2	0.93	0.04	0.02	0.01	0.97	0.94	0.63	6,180	4,097	253	1,830
6	7A	1	0.95	0.01	0.00	0.03	0.99	0.98	0.52	6,278	3,003	3,275	
6	7B	2	0.86	0.06	0.04	0.04	0.94	0.89	0.85	6,278	3,486	281	2,511
6	22A	1	0.99	0.00	0.00	0.01	0.99	0.98	0.90	6,278	607	5,671	
6	22B	2	0.91	0.07	0.00	0.02	0.94	0.87	0.49	6,278	3,390	2,692	196

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-29 Cont'd
Inter-Rater Reliability, Mathematics*

			Percentage Absolute Difference								Frequency		
Grade	Item No.	Max	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	No. of Reads	0	1	2
6	39A	1	0.97	0.01	0.00	0.01	0.99	0.97	0.44	6,278	3,535	2,743	
6	39B	2	0.92	0.05	0.01	0.02	0.98	0.95	0.82	6,278	3,005	1,390	1,883
6	51	2	0.94	0.00	0.00	0.01	0.98	0.95	0.86	6,278	2,286	2,559	1,433
7	7A	1	0.98	0.00	0.00	0.01	1.00	0.99	0.69	6,420	1,979	4,441	
7	7B	2	0.86	0.10	0.01	0.02	0.96	0.92	1.22	6,420	2,053	925	3,442
7	19A	1	0.98	0.01	0.00	0.02	0.99	0.99	0.52	6,420	3,088	3,332	
7	19B	2	0.91	0.06	0.01	0.03	0.97	0.93	0.61	6,420	3,379	2,159	882
7	38A	1	0.88	0.02	0.01	0.09	0.98	0.95	0.53	6,420	3,040	3,380	
7	38B	2	0.92	0.04	0.01	0.03	0.97	0.95	0.71	6,420	2,558	3,193	669
7	52	2	0.97	0.02	0.00	0.01	0.99	0.99	0.70	6,420	3,113	2,134	1,173
8	10	2	0.89	0.07	0.00	0.03	0.96	0.92	0.39	6,284	4,645	837	802
8	19A	1	0.98	0.01	0.00	0.01	1.00	0.98	0.70	6,284	1,876	4,408	
8	19B	2	0.86	0.10	0.01	0.03	0.97	0.93	1.13	6,284	2,149	1,175	2,960
8	33A	1	0.84	0.09	0.01	0.07	0.88	0.77	0.27	6,284	4,578	1,706	
8	33B	2	0.81	0.13	0.01	0.06	0.93	0.86	0.66	6,284	3,037	2,341	906
8	49A	1	0.98	0.00	0.00	0.02	1.00	0.99	0.26	6,284	4,622	1,662	
8	49B	2	0.88	0.08	0.01	0.03	0.96	0.91	0.42	6,284	4,697	523	1,064
10	23	2	0.86	0.09	0.00	0.05	0.96	0.92	0.87	6,436	2,448	2,354	1,634
10	33	2	0.82	0.07	0.01	0.10	0.97	0.94	0.61	6,436	3,700	1,551	1,185
10	36	2	0.81	0.11	0.00	0.08	0.96	0.93	0.92	6,436	2,795	1,387	2,254
10	46	2	0.84	0.10	0.01	0.05	0.96	0.91	0.91	6,436	2,852	1,342	2,242

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

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 9-30
Inter-Rater Reliability, Writing Prompts*

			Percentage Absolute Difference								Frequency						
Grade	Item No.	Max Score	P	A	D	Codes	Intra. Corr.	Weighted Kappa	Mean	No. of Reads	0	1	2	3	4	5	6
4	1A	6	0.67	0.30	0.02	0.02	0.91	0.81	3.18	6,372	134	133	1,088	2,501	2,133	357	26
4	1B	3	0.95	0.04	0.00	0.01	0.84	0.69	1.97	6,372	54	123	6,142	63			
8	1A	6	0.64	0.32	0.02	0.02	0.90	0.80	3.32	6,284	131	61	923	2,341	2,258	548	22
8	1B	3	0.96	0.03	0.00	0.01	0.89	0.79	1.98	6,284	76	55	6,106	47			
10	1A	6	0.60	0.33	0.03	0.04	0.92	0.84	3.32	6,436	258	197	921	1,761	2,553	658	88
10	1B	3	0.87	0.10	0.00	0.03	0.87	0.75	2.01	6,436	197	40	5,695	504			

* Note that P is percent perfect agreement, A is percent adjacent agreement, and D is percent discrepant. Also, note that the sum of the modes of agreement and codes may not equal exactly 100% due to rounding.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-1
Items Flagged for DIF, By Gender

Content	Grade	Test Book Number	Item Type	Female			Male			SMD	Delta	LH Flag Female	LH Flag Male	Flag MH
				D+	D-	Z	D+	D-	Z					
RD	5	13	CR	0.17	-0.08	6.97	0.02	-0.13	-7.01	0.21		CC		BB
RD	6	18	CR	0.09	0.00	7.55	0.00	-0.09	-7.53	0.27				CC
RD	7	31	CR	0.10	0.00	7.61	0.01	-0.11	-7.68	0.30		CC		CC
RD	8	19	CR	0.10	0.00	7.96	0.03	-0.12	-8.42	0.30		CC	-CC	CC
RD	8	21	MC	0.01	-0.05	-3.89	0.03	0.00	5.53	-0.16	-1.60			-C
RD	8	23	MC	0.00	-0.06	-5.15	0.05	0.00	6.13	-0.22	-2.13			-C
RD	8	31	MC	0.00	-0.06	-6.67	0.05	0.00	6.77	-0.21	-1.61			-C
RD	8	56	CR	0.14	0.00	8.81	0.00	-0.14	-9.23	0.25		CC	-CC	CC
RD	10	7A	CR	0.11	0.00	7.03	0.02	-0.14	-8.23	0.22		CC	-CC	BB
RD	10	21A	CR	0.12	-0.08	7.71	0.03	-0.15	-10.28	0.31		CC	-CC	CC
LA	4	1A	CR	0.15	0.00	8.87	0.00	-0.16	-9.41	0.27		CC	-CC	CC
LA	8	1A	CR	0.12	-0.02	5.94	0.05	-0.15	-7.92	0.22		CC	-CC	BB

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-2
Items Flagged for DIF, By Race/Ethnicity, African American

Content	Grade	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	10	39A	MC	0.01	-0.03	-2.62	-0.41	-1.79		-C
MA	5	35	MC	0.04	-0.06	-1.78	-0.36	-1.57		-C
MA	8	54	MC	0.05	-0.03	-0.48	-0.44	-1.97		-C
LA	8	1A	CR	0.08	-0.18	-3.60	0.03		-CC	
LA	10	26	MC	0.04	-0.07	-3.07	-0.43	-2.01		-C
SC	4	1	MC	0.03	-0.01	1.87	-0.24	-1.60		-C
SS	10	2	MC	0.03	-0.03	0.58	-0.32	-1.64		-C

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-3
Items Flagged for DIF, By Race/Ethnicity, Hispanic

Content	Grade	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	6	3	MC	0.00	-0.08	-4.64	-0.28	-1.54		-C
RD	6	37	MC	0.01	-0.05	-2.77	-0.33	-1.68		-C
RD	8	21	MC	0.00	-0.06	-3.72	-0.36	-1.83		-C
MA	6	7B	CR	0.15	0.00	5.17	0.13		CC	
MA	8	54	MC	0.02	-0.06	-2.06	-0.34	-1.84		-C
LA	4	11	MC	0.00	-0.03	-2.06	-0.26	-1.53		-C
LA	10	26	MC	0.03	-0.11	-5.28	-0.31	-1.54		-C
SC	4	1	MC	0.01	-0.01	-0.28	-0.41	-3.17		-C

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-4
Items Flagged for DIF, By Race/Ethnicity, Asian

Content	Grade	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	3	31	CR	0.15	-0.03	4.60	0.10		CC	
RD	4	13	CR	0.15	0.00	4.95	0.16		CC	
RD	5	47	CR	0.13	0.00	4.06	0.23		CC	BB
RD	6	18	CR	0.15	0.00	5.02	0.22		CC	BB
RD	6	24	CR	0.14	0.00	4.40	0.21		CC	BB
RD	6	29	MC	0.04	-0.15	-6.08	-0.25	-1.36	-C	-B
RD	6	37	MC	0.01	-0.05	-2.31	-0.23	-1.54		-C
RD	6	43	MC	0.08	-0.13	-5.83	-0.19	-1.02	-C	-B
RD	6	44	MC	0.15	-0.10	4.91	0.24	1.33	C	B
RD	7	31	CR	0.17	-0.01	4.88	0.24		CC	BB
RD	8	19	CR	0.21	-0.07	6.00	0.24		CC	BB
RD	8	21	MC	0.01	-0.07	-3.16	-0.32	-2.10		-C
RD	8	56	CR	0.26	-0.09	5.65	0.30		CC	CC
RD	10	2A	MC	0.05	0.00	3.65	0.18	1.83		C
RD	10	7A	CR	0.24	0.00	6.77	0.29		CC	CC
MA	3	16	MC	0.00	-0.13	-7.73	-0.27	-1.68	-C	-C
MA	5	18	MC	0.00	-0.09	-4.32	-0.28	-1.60		-C
MA	6	24	MC	0.00	-0.12	-5.85	-0.13	-0.72	-C	
MA	6	39B	CR	0.17	-0.07	3.71	0.08		CC	
MA	6	47	MC	0.00	-0.08	-5.16	-0.21	-1.56		-C
MA	7	18	MC	0.01	-0.06	-1.08	-0.19	-1.66		-C
MA	7	21	MC	0.04	-0.15	-4.30	-0.06	-0.33	-C	
MA	8	54	MC	0.01	-0.13	-2.97	-0.28	-2.27		-C
MA	10	4	MC	0.08	-0.14	-6.01	-0.17	-1.07	-C	-B
MA	10	33	CR	0.15	-0.02	5.40	0.11		CC	
LA	4	13	MC	0.00	-0.15	-8.51	-0.41	-2.47	-C	-C
LA	4	1A	CR	0.14	0.00	3.60	0.11		CC	

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-5 Cont'd
Items Flagged for DIF, By Race/Ethnicity, Asian

Content	Grade	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
LA	8	17	MC	0.00	-0.08	-3.60	-0.27	-1.60		-C
LA	8	1A	CR	0.29	-0.03	6.18	0.16	.	CC	
LA	10	26	MC	0.00	-0.21	-12.53	-0.60	-3.19	-C	-C
LA	10	1A	CR	0.24	0.00	5.92	0.23	.	CC	BB
SC	4	1	MC	0.00	-0.06	-3.12	-0.47	-4.32		-C
SC	4	34	MC	0.04	-0.01	2.53	0.14	1.78		C
SC	10	34	MC	0.00	-0.12	-6.43	-0.32	-1.65	-C	-C
SS	8	3	MC	0.00	-0.13	-6.97	-0.39	-2.07	-C	-C
SS	8	37	MC	0.12	0.00	5.53	0.29	1.55	C	C
SS	10	25	MC	0.01	-0.14	-1.13	-0.15	-1.65		-C

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-6
Items Flagged for DIF, By Race/Ethnicity, American Indian*

Content	Grade	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	7	29	MC	0.13	-0.24	-2.60	0.01	0.08	-C	
RD	8	56	CR	0.11	-0.44	-3.18	-0.03		-CC	
MA	6	12	MC	0.06	-0.26	-3.58	-0.02	-0.13	-C	
MA	7	13	MC	0.22	-0.19	-2.84	-0.04	-0.29	-C	
MA	7	41	MC	0.19	-0.19	-3.10	-0.02	-0.10	-C	
MA	7	53	MC	0.08	-0.25	-2.79	-0.03	-0.15	-C	
SS	10	6	MC	0.21	-0.11	2.74	0.03	0.18	C	
SS	10	10	MC	0.21	-0.02	2.74	0.02	0.15	C	
SS	10	44	MC	0.02	-0.16	-2.59	-0.01	-0.07	-C	

* Note: DIF statistics can only be calculated for items with sufficient student N counts. In some cases here, the size of the tested population was too small to include valid DIF statistics.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-7
Items Flagged for DIF, By English Language Proficiency

Content	Grade	Test Book Number	Item Type	Limited English Proficient			Fully English Proficient			SMD	Delta	LH Flag Limited English Proficient	LH Flag Fully English Proficient	MH Flag
				D+	D-	Z	D+	D-	Z					
RD	3	4	MC	0.00	-0.11	-5.67	0.02	-0.01	1.65	-0.19	-1.08	-C		-B
RD	4	13	CR	0.16	-0.10	4.95	0.04	-0.05	-2.38	0.16		CC		
RD	5	13	CR	0.17	-0.16	3.42	0.06	-0.10	-1.62	0.09		CC		
RD	5	47	CR	0.13	-0.06	3.63	0.03	-0.03	-2.17	0.17		CC		
RD	6	3	MC	0.00	-0.08	-4.69	0.02	-0.01	1.60	-0.27	-1.63			-C
RD	6	4	MC	0.01	-0.10	-5.15	0.02	0.00	2.64	-0.33	-1.50			-C
RD	6	18	CR	0.13	-0.06	4.53	0.02	-0.04	-1.89	0.22		CC		BB
RD	6	24	CR	0.13	-0.03	4.30	0.03	-0.04	-2.31	0.16		CC		
RD	6	37	MC	0.00	-0.07	-4.65	0.01	0.00	3.23	-0.36	-1.61			-C
RD	6	44	MC	0.15	-0.17	6.18	0.02	-0.05	-2.52	0.23	1.21	C		B
RD	7	31	CR	0.15	0.00	5.31	0.03	-0.05	-2.16	0.25		CC		CC
RD	8	19	CR	0.13	0.00	4.78	0.07	-0.03	-2.09	0.20		CC		BB
RD	8	21	MC	0.02	-0.12	-6.01	0.02	0.00	3.39	-0.40	-1.78	-C		-C
RD	8	56	CR	0.23	-0.09	5.32	0.02	-0.05	-2.28	0.32		CC		CC
RD	10	19A	MC	0.00	-0.13	-6.46	0.02	-0.01	2.43	-0.29	-1.25	-C		-B
MA	3	16	MC	0.00	-0.13	-7.81	0.02	-0.01	3.20	-0.28	-1.46	-C		-B
MA	5	18	MC	0.00	-0.11	-5.71	0.03	-0.03	1.80	-0.27	-1.29	-C		-B
MA	6	7B	CR	0.20	-0.03	5.19	0.04	-0.04	-2.53	0.16		CC		
MA	6	47	MC	0.00	-0.07	-4.55	0.03	-0.01	1.79	-0.32	-1.59			-C
MA	8	37	MC	0.00	-0.12	-6.88	0.02	-0.01	1.75	-0.16	-1.42	-C		-B
MA	8	54	MC	0.02	-0.08	-3.14	0.02	-0.01	3.11	-0.35	-1.79			-C
MA	10	4	MC	0.04	-0.16	-5.99	0.02	-0.03	1.48	-0.05	-0.31	-C		
LA	4	11	MC	0.01	-0.03	-2.15	0.01	0.00	3.59	-0.33	-1.87			-C
LA	4	1A	CR	0.13	0.00	4.08	0.05	-0.07	-2.22	0.08		CC		
LA	8	1A	CR	0.18	-0.07	4.11	0.11	-0.05	-3.13	0.17		CC		

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-6 Cont'd
Items Flagged for DIF, By English Language Proficiency

Content	Grade	Test Book Number	Item Type	Limited English Proficient			Fully English Proficient			SMD	Delta	LH Flag Limited English Proficient	LH Flag Fully English Proficient	MH Flag
				D+	D-	Z	D+	D-	Z					
LA	10	22	MC	0.03	-0.12	-4.92	0.02	-0.02	1.53	-0.08	-0.45	-C		
LA	10	26	MC	0.00	-0.21	-10.30	0.02	0.00	3.29	-0.42	-2.03	-C		-C
LA	10	1A	CR	0.21	0.00	4.83	0.07	-0.06	-2.50	0.14		CC		
SC	4	1	MC	0.01	-0.05	-3.05	0.01	0.00	2.79	-0.46	-3.53			-C
SS	8	3	MC	0.01	-0.14	-6.27	0.02	-0.01	1.97	-0.24	-1.15	-C		-B
SS	10	28	MC	0.01	-0.13	-5.67	0.03	-0.01	1.89	-0.21	-0.98	-C		

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-8
Items Flagged for DIF, By Disability Status

Content	Grade	Test Book Number	Item Type	Not Disabled			Disabled			SMD	Delta	LH Flag Not Disabled	LH Flag Disabled	MH Flag
				D+	D-	Z	D+	D-	Z					
RD	5	13	CR	0.09	-0.09	1.75	0.01	-0.16	-5.27	-0.17			-CC	
RD	7	31	CR	0.05	-0.03	1.68	0.15	-0.14	-5.01	-0.13			-CC	
RD	8	19	CR	0.06	-0.02	2.02	0.03	-0.21	-6.13	-0.14			-CC	
RD	8	56	CR	0.06	-0.10	2.31	0.00	-0.19	-6.67	-0.15			-CC	
RD	10	21A	CR	0.01	-0.02	-0.31	0.01	-0.12	-4.33	-0.11			-CC	
MA	3	29	MC	0.01	-0.04	-2.87	0.15	0.00	6.95	0.25	1.34		C	B
LA	4	11	MC	0.02	0.00	4.78	0.00	-0.05	-5.37	-0.29	-1.59			-C
LA	4	1A	CR	0.10	-0.07	2.77	0.00	-0.30	-8.56	-0.34			-CC	-CC
LA	4	1B	CR	0.02	-0.01	0.73	0.04	-0.02	-1.73	-0.27				-CC
LA	8	1A	CR	0.09	-0.04	1.06	0.09	-0.28	-6.30	-0.22			-CC	-BB
LA	10	1A	CR	0.11	-0.05	0.34	0.20	-0.27	-3.38	-0.15			-CC	
SC	8	2	MC	0.01	0.00	1.62	0.00	-0.02	-1.58	-0.27	-1.84			-C

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-9
Correlations among Reading Objectives

Grade	CS	1	2	3
3	2	0.76	-	-
	3	0.77	0.84	
	4	0.58	0.60	0.63
4	2	0.73	-	-
	3	0.75	0.79	-
	4	0.57	0.59	0.61
5	2	0.74	-	-
	3	0.74	0.77	-
	4	0.64	0.67	0.67
6	2	0.65	-	-
	3	0.69	0.71	-
	4	0.65	0.67	0.72
7	2	0.65	-	-
	3	0.69	0.71	-
	4	0.62	0.65	0.69
8	2	0.64	-	-
	3	0.67	0.71	-
	4	0.66	0.70	0.73
10	2	0.50	-	-
	3	0.63	0.63	-
	4	0.64	0.62	0.81

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-10
Correlations among Mathematics Objectives

Grade	CS	A	B	C	D	E
3	B	0.62	-	-	-	-
	C	0.53	0.60	-	-	-
	D	0.50	0.60	0.54	-	-
	E	0.64	0.66	0.59	0.58	-
	F	0.56	0.68	0.55	0.54	0.60
4	B	0.54	-	-	-	-
	C	0.50	0.48	-	-	-
	D	0.60	0.63	0.52	-	-
	E	0.55	0.53	0.47	0.56	-
	F	0.62	0.67	0.53	0.65	0.55
5	B	0.56	-	-	-	-
	C	0.52	0.50	-	-	-
	D	0.57	0.62	0.52	-	-
	E	0.65	0.57	0.53	0.61	-
	F	0.66	0.68	0.53	0.69	0.65
6	B	0.63	-	-	-	-
	C	0.57	0.58	-	-	-
	D	0.64	0.69	0.62	-	-
	E	0.69	0.64	0.60	0.65	-
	F	0.67	0.66	0.56	0.65	0.61
7	B	0.65	-	-	-	-
	C	0.65	0.64	-	-	-
	D	0.61	0.69	0.62	-	-
	E	0.64	0.68	0.64	0.65	-
	F	0.66	0.70	0.62	0.64	0.65
8	B	0.60	-	-	-	-
	C	0.63	0.55	-	-	-
	D	0.65	0.66	0.58	-	-
	E	0.66	0.59	0.57	0.63	-
	F	0.67	0.66	0.64	0.67	0.66
10	B	0.70	-	-	-	-
	C	0.71	0.66	-	-	-
	D	0.72	0.68	0.71	-	-
	E	0.69	0.65	0.65	0.68	-
	F	0.72	0.69	0.72	0.75	0.68

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-11
Correlations among Language Arts Objectives

Grade	CS	B	D
4	D	0.52	-
	F	0.53	0.38
8	D	0.68	-
	F	0.50	0.44
10	D	0.69	-
	F	0.55	0.53

Table 10-12
Correlations among Social Studies Objectives

Grade	CS	A	B	C	D
4	B	0.65	-	-	-
	C	0.53	0.51	-	-
	D	0.54	0.52	0.47	-
	E	0.65	0.61	0.54	0.54
8	B	0.62	-	-	-
	C	0.55	0.57	-	-
	D	0.59	0.60	0.54	-
	E	0.52	0.54	0.50	0.52
10	B	0.60	-	-	-
	C	0.60	0.69	-	-
	D	0.58	0.66	0.69	-
	E	0.54	0.60	0.62	0.61

Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-13
Correlations among Science Objectives

Grade	CS	A/B	C	D	E	F
4	C	0.58	-	-	-	-
	D	0.35	0.34	-	-	-
	E	0.45	0.44	0.32	-	-
	F	0.52	0.52	0.34	0.47	-
	G/H	0.59	0.58	0.36	0.46	0.54
8	C	0.57	-	-	-	-
	D	0.48	0.46	-	-	-
	E	0.51	0.47	0.44	-	-
	F	0.51	0.48	0.44	0.47	-
	G/H	0.58	0.59	0.46	0.47	0.50
10	C	0.60	-	-	-	-
	D	0.60	0.58	-	-	-
	E	0.45	0.41	0.42	-	-
	F	0.46	0.43	0.42	0.33	-
	G/H	0.67	0.62	0.59	0.45	0.47

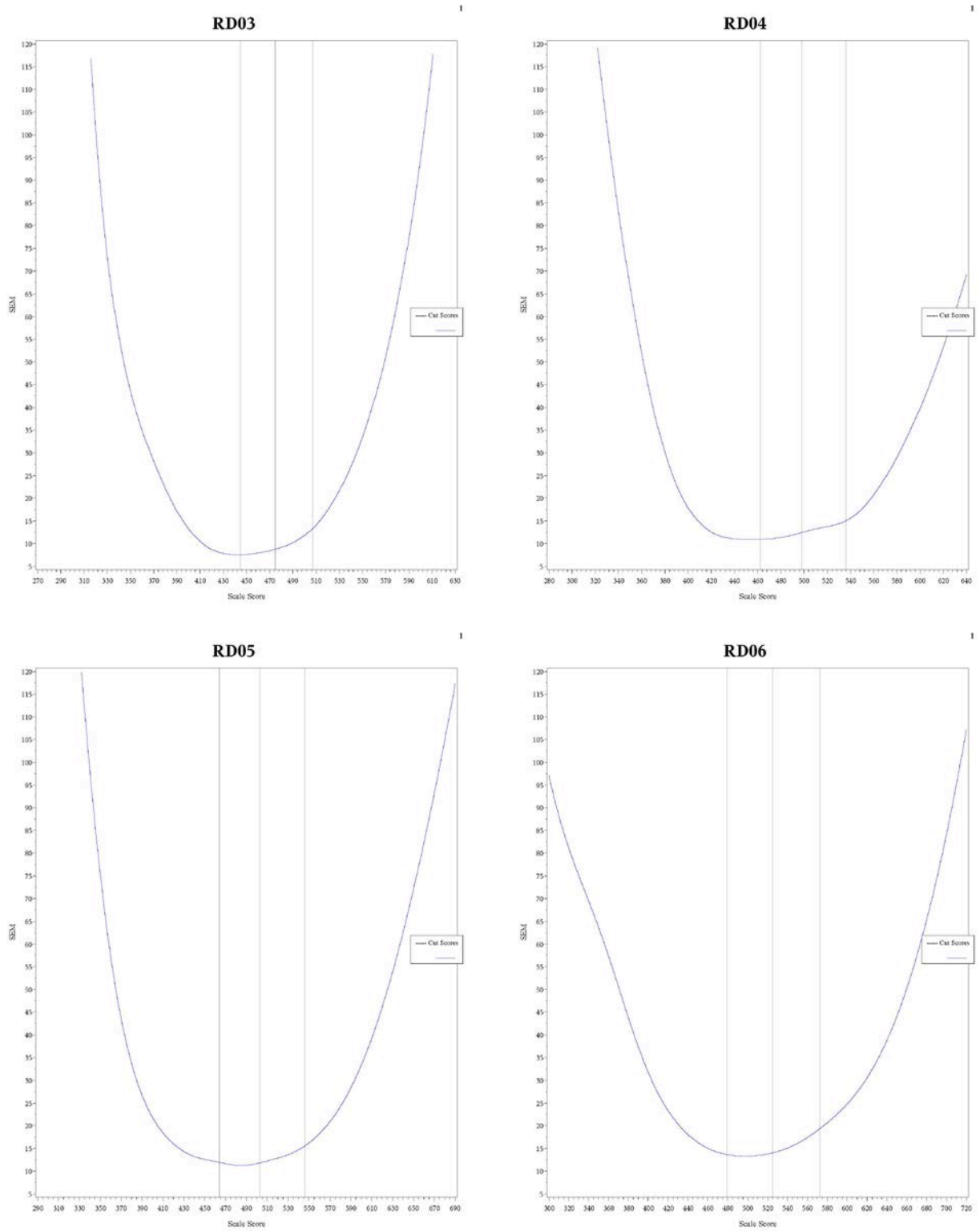
Copyright © 2013 by the Wisconsin Department of Public Instruction

Table 10-14
Principal Components Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
Reading	3	12.78	1.65	7.74
	4	11.18	1.55	7.23
	5	10.68	1.65	6.46
	6	9.43	1.63	5.78
	7	9.17	1.73	5.31
	8	9.85	1.50	6.59
	10	9.72	1.51	6.43
Mathematics	3	9.60	1.86	5.16
	4	8.86	1.81	4.90
	5	10.05	1.93	5.20
	6	11.04	1.78	6.21
	7	11.16	1.79	6.24
	8	10.89	1.88	5.78
	10	12.75	1.71	7.44
Language Arts	4	5.29	1.31	4.05
	8	5.95	1.20	4.97
	10	5.80	1.19	4.89
Social Studies	4	7.04	1.61	4.38
	8	7.28	1.58	4.60
	10	9.03	1.72	5.26
Science	4	6.49	1.54	4.23
	8	6.60	1.53	4.31
	10	7.99	1.40	5.70

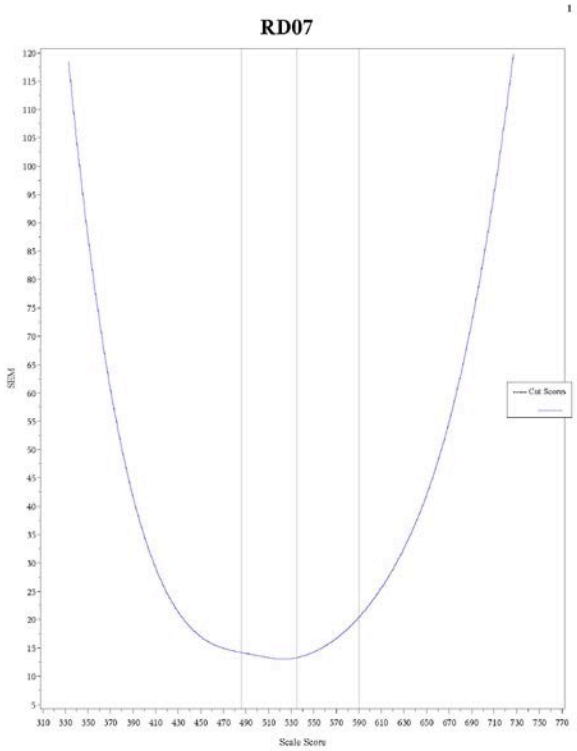
Copyright © 2013 by the Wisconsin Department of Public Instruction

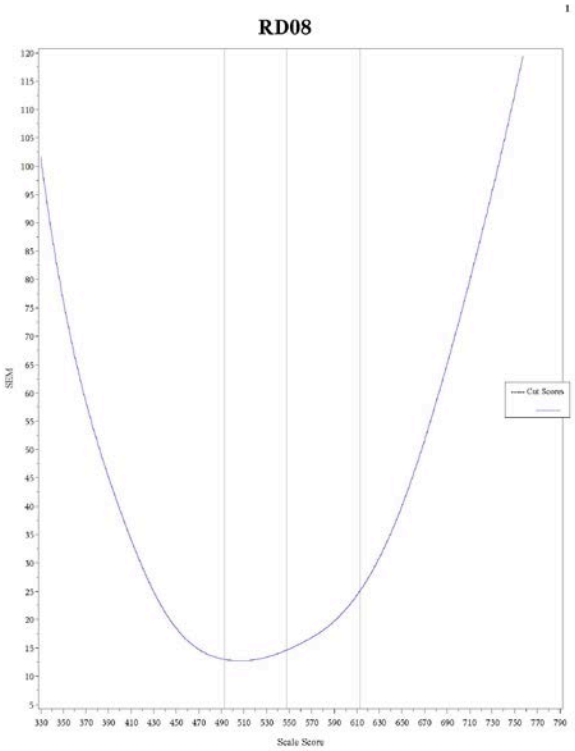
Figure 7-1
SEM Curves, Reading Grades 3-6

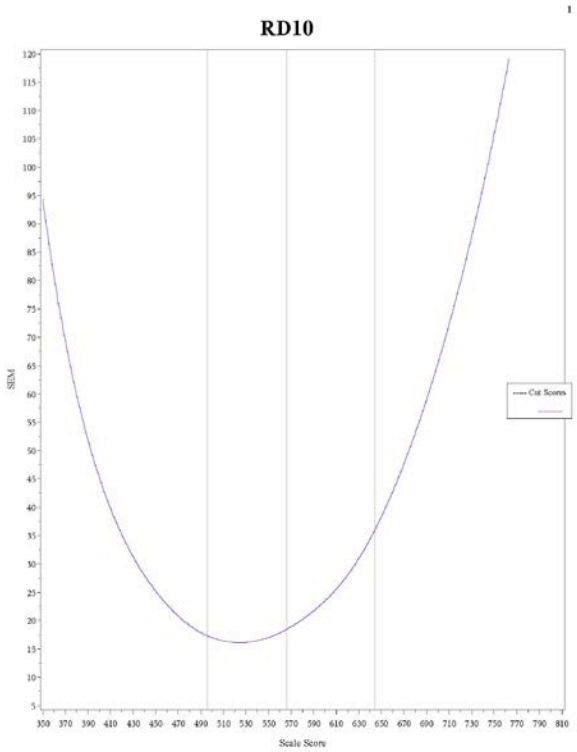


Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-1 Cont'd
SEM Curves, Reading Grades 7, 8, 10

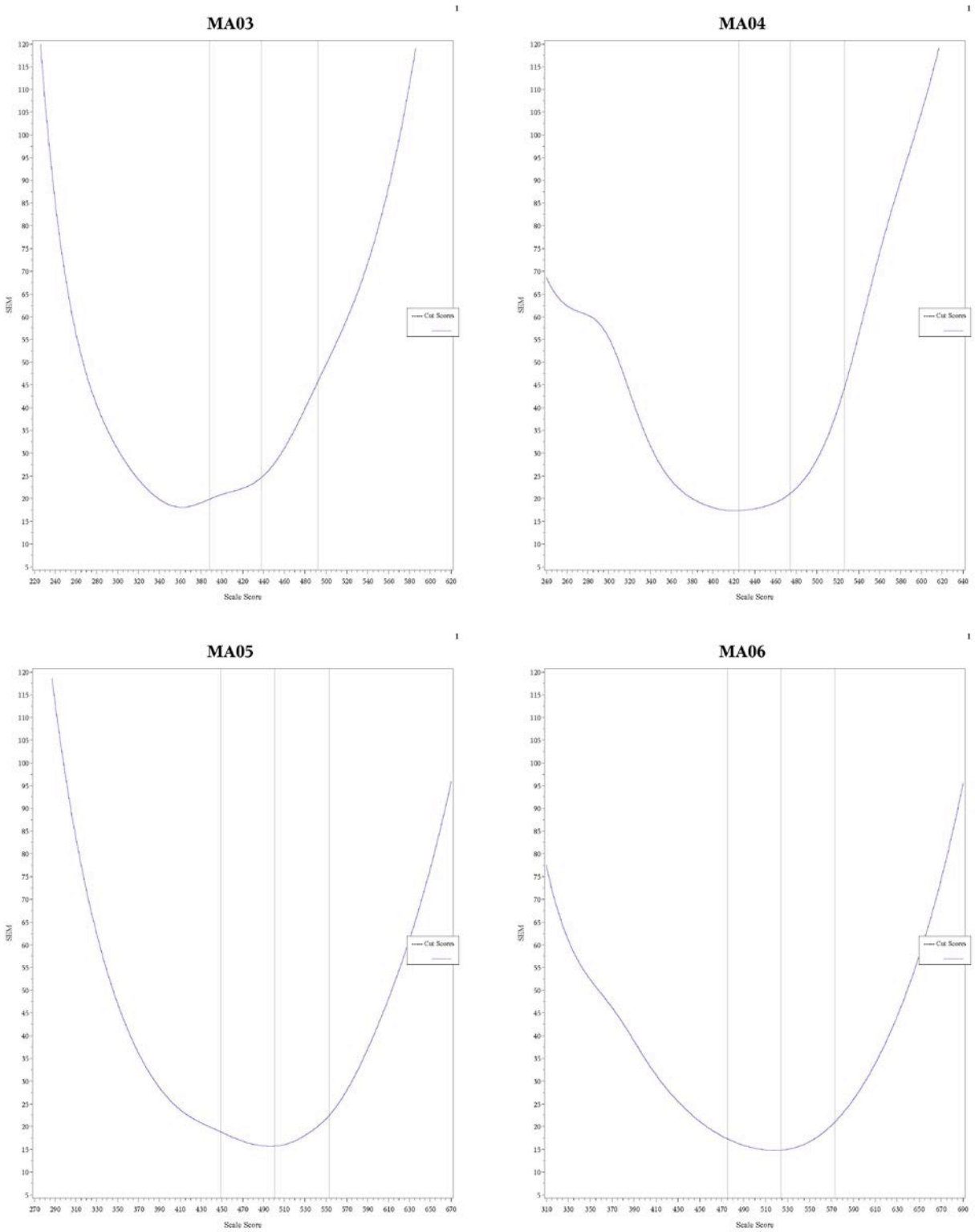






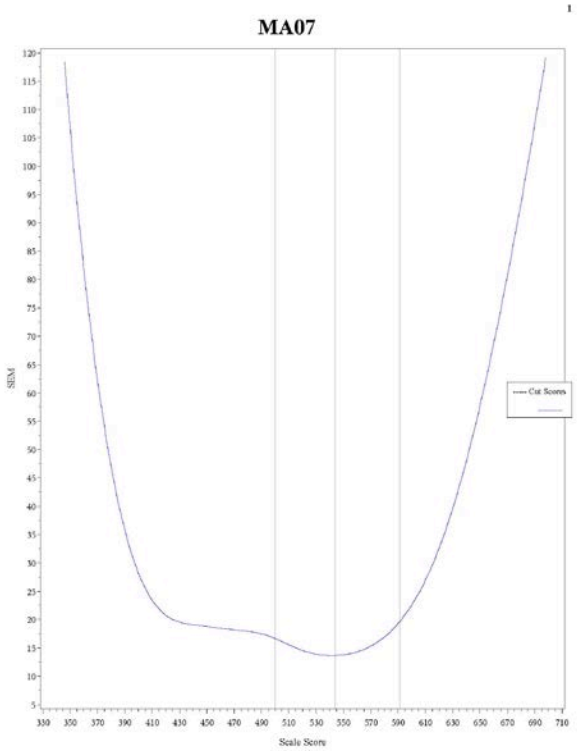
Copyright © 2013 by the Wisconsin Department of Public Instruction

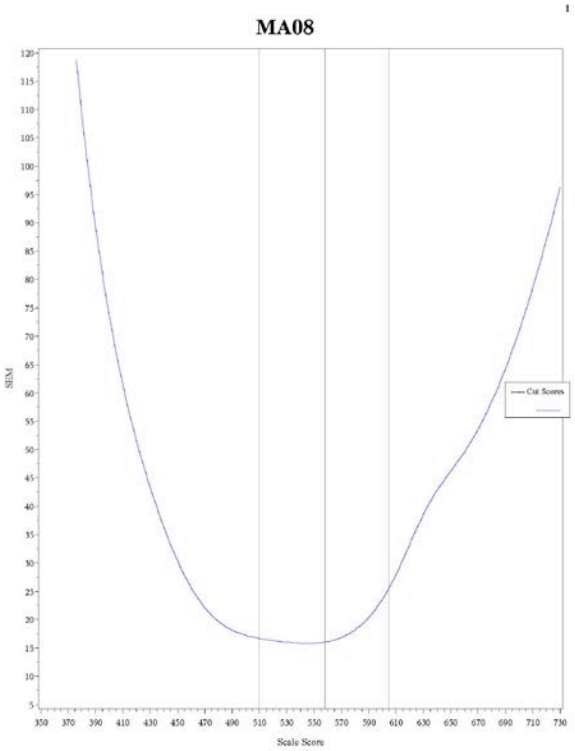
Figure 7-2
SEM Curves, Mathematics Grades 3-6

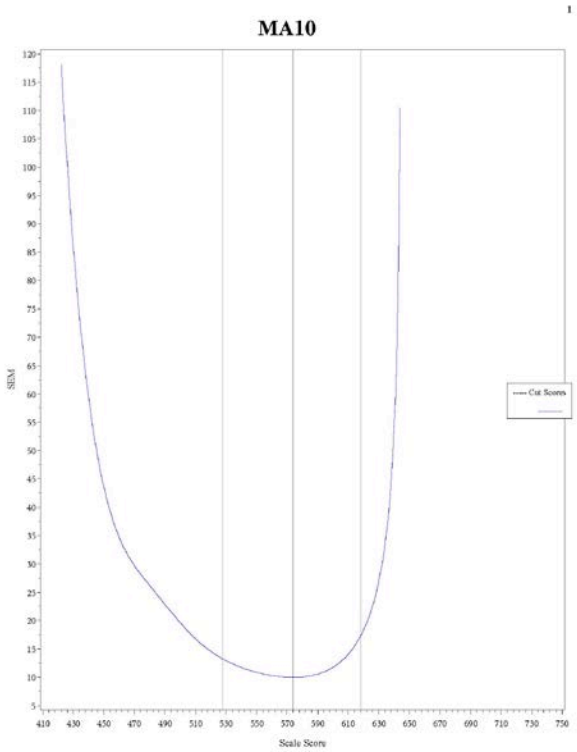


Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-2 Cont'd
SEM Curves, Mathematics Grades 7, 8, 10

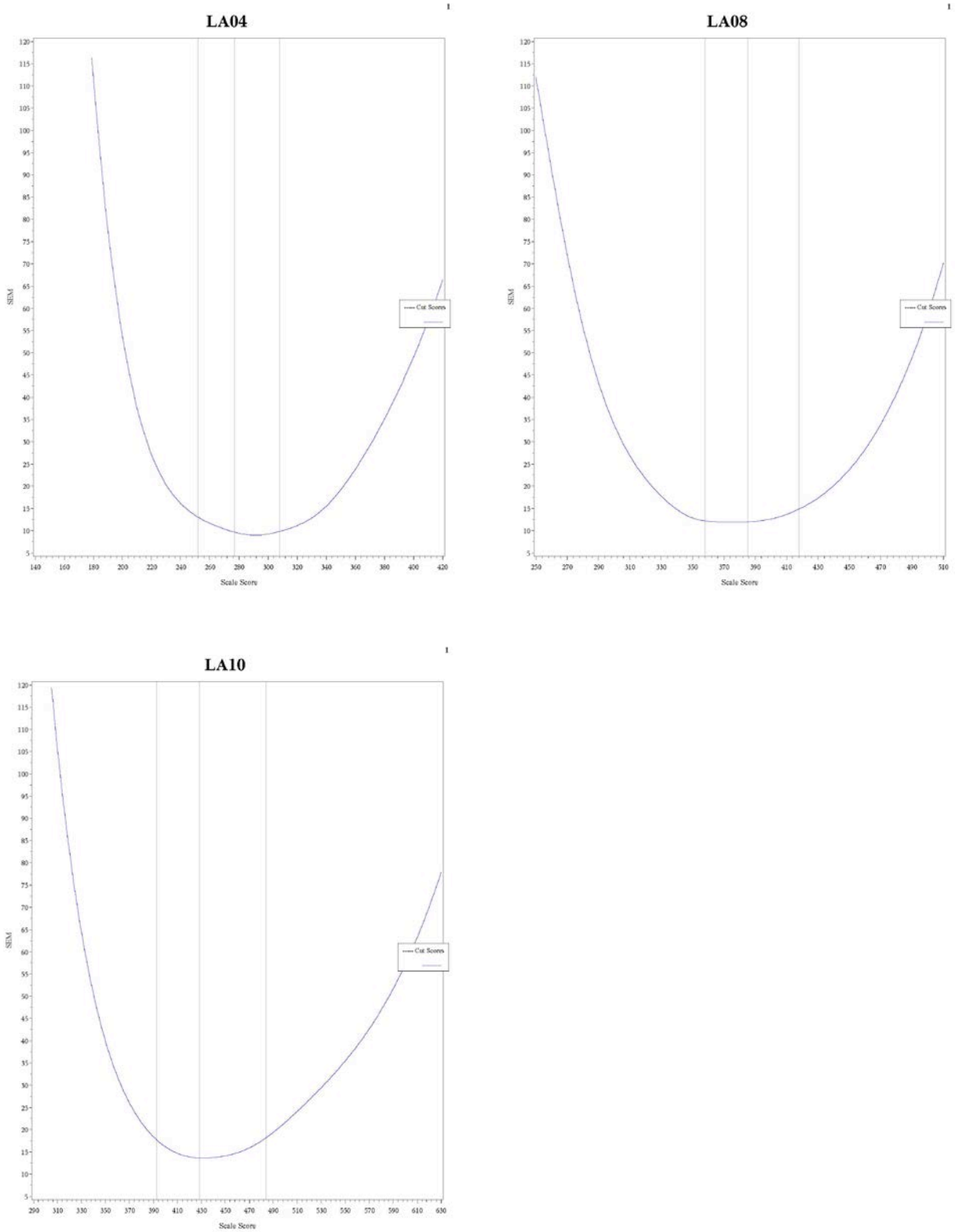






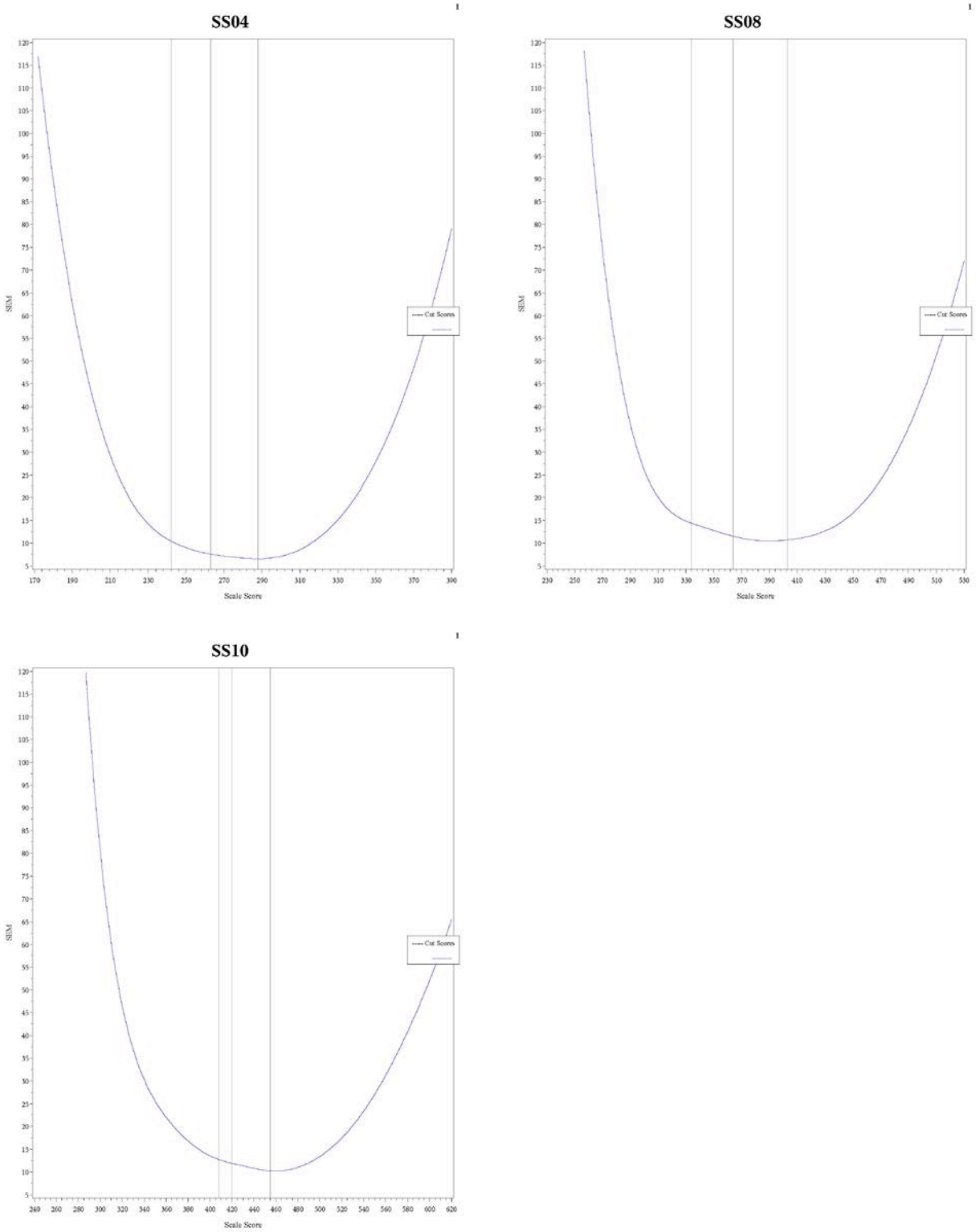
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-3
SEM Curves, Language Arts Grades 4, 8, 10



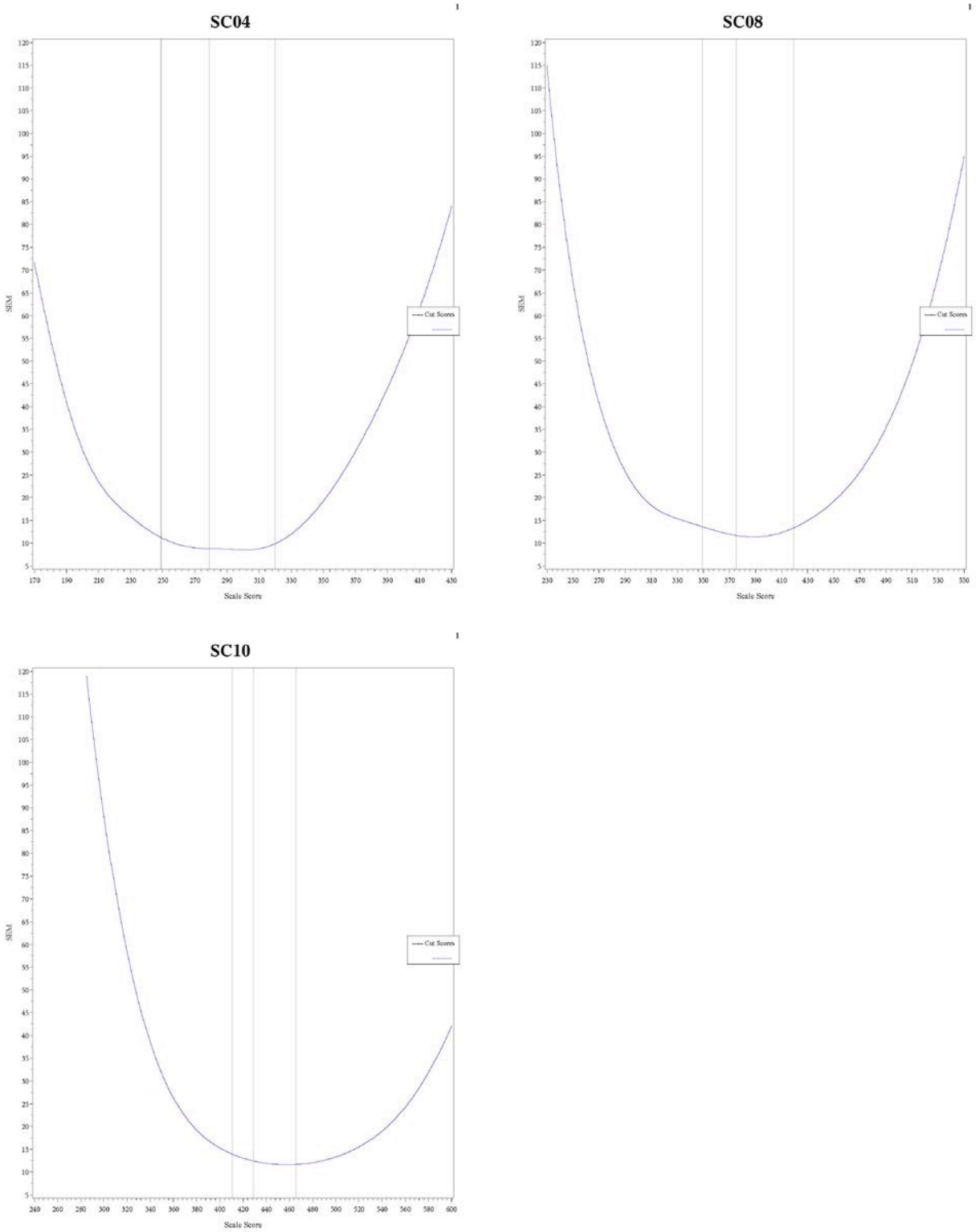
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-4
SEM Curves, Social Studies Grades 4, 8, 10



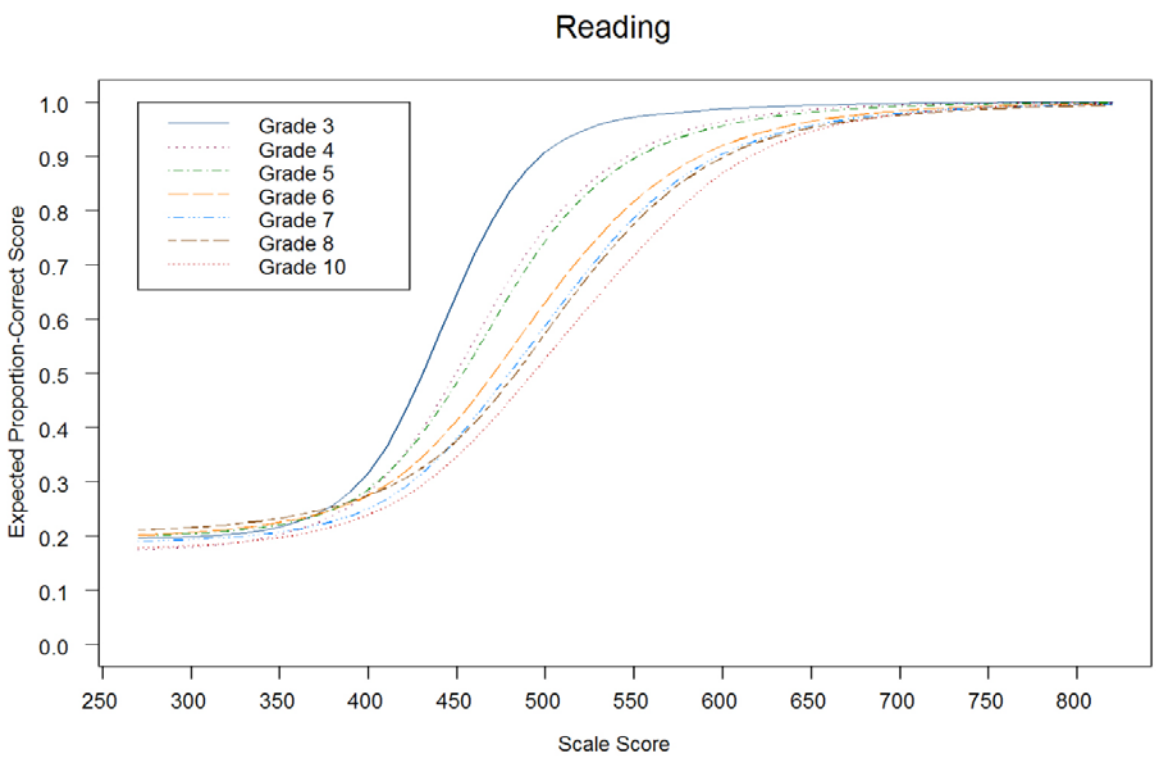
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-5
SEM Curves, Science Grades 4, 8, 10



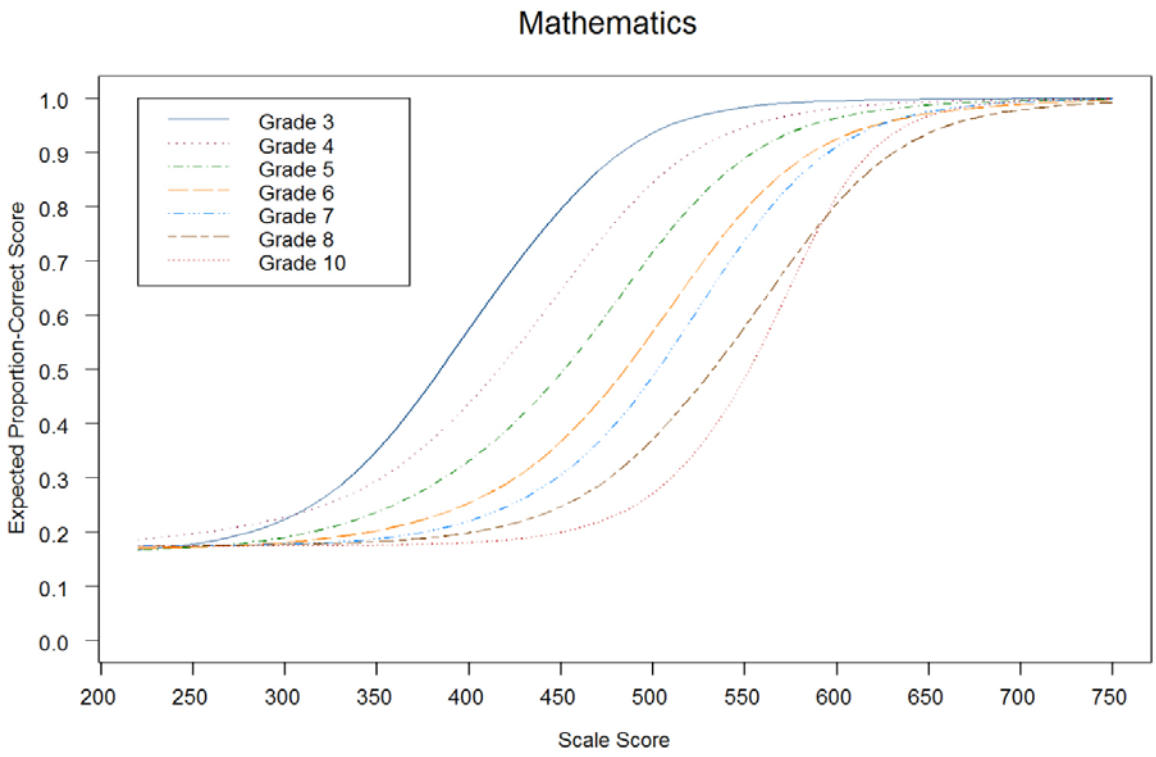
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-6
TCC Curve for Reading Grades 3-8, 10



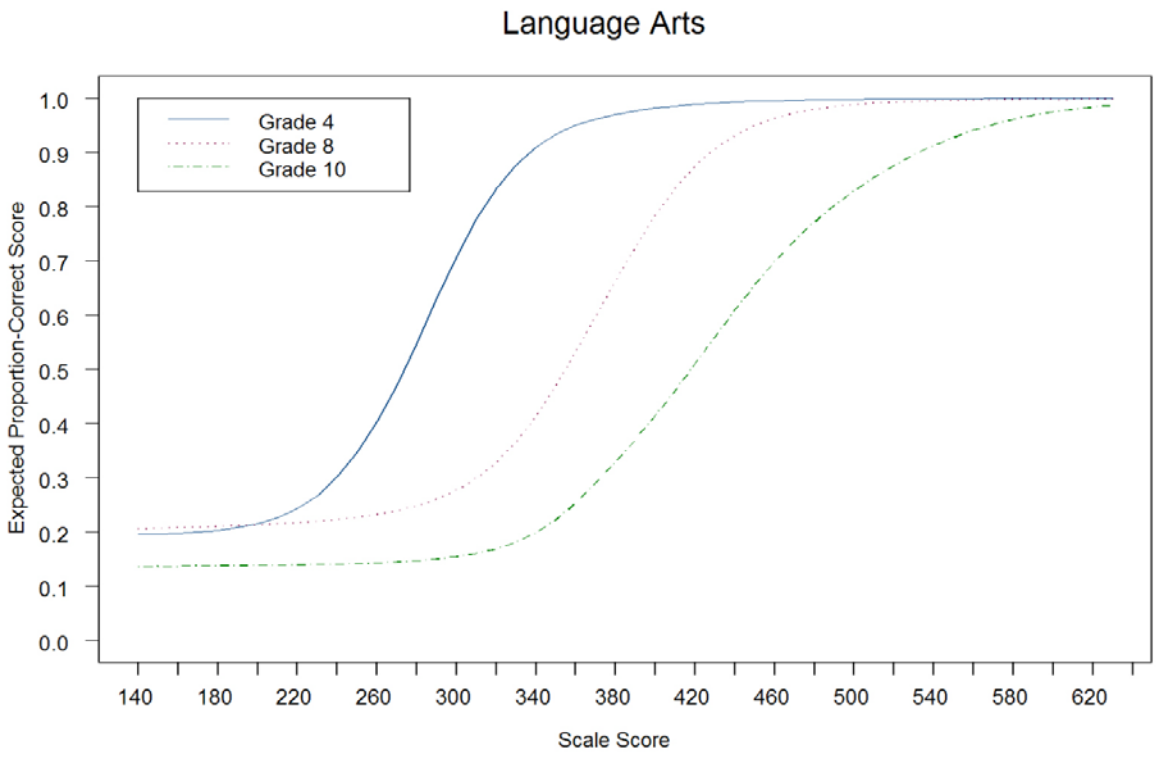
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-7
TCC Curve for Mathematics Grades 3-8, 10



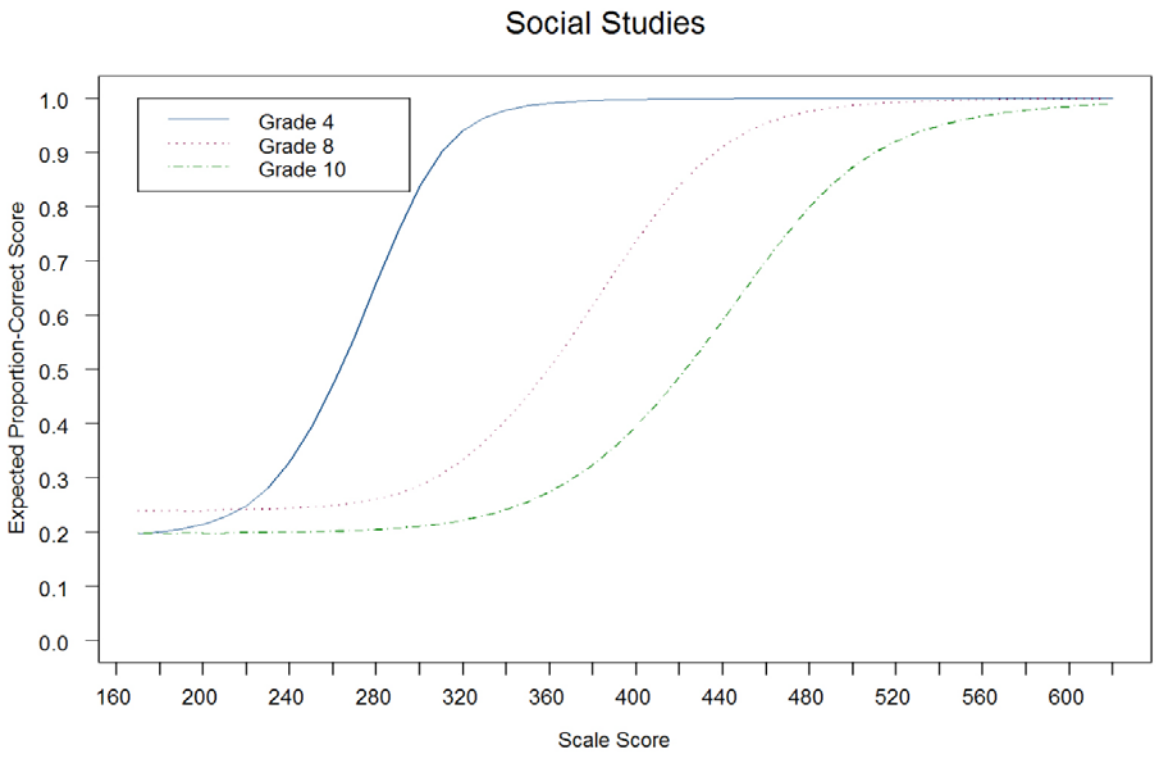
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-8
TCC Curve for Language Arts Grades 4, 8, 10



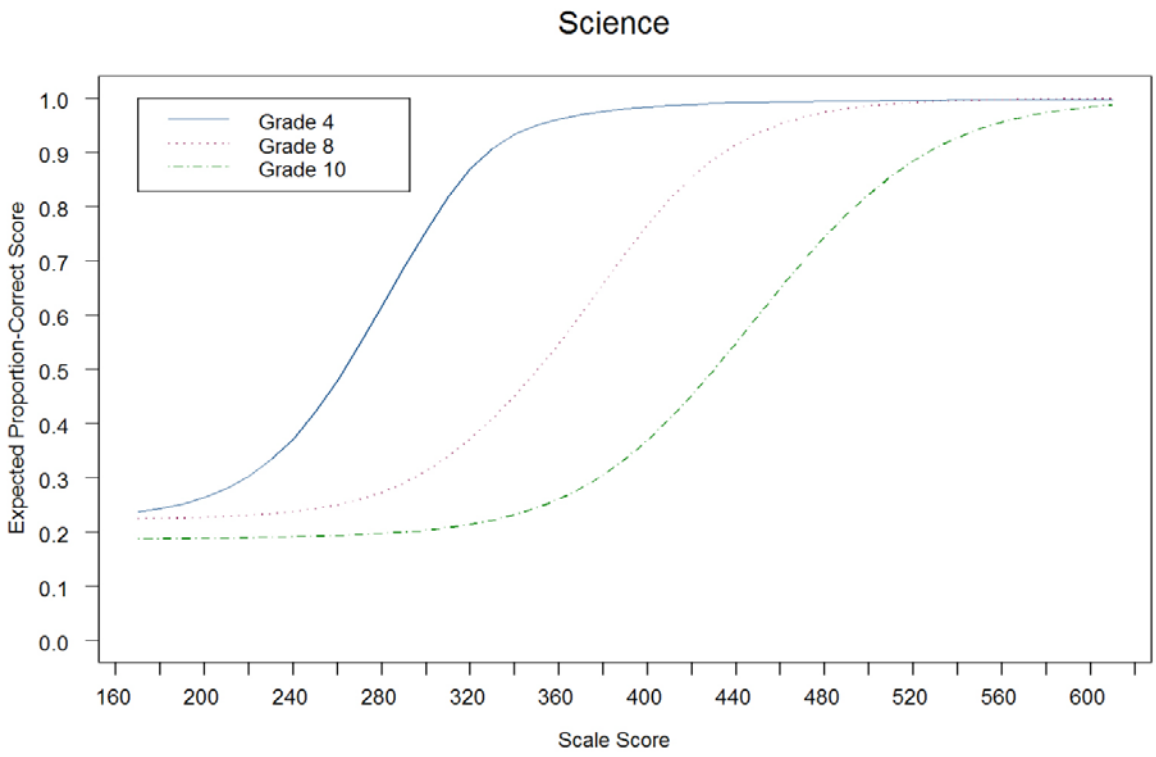
Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-9
TCC Curve for Social Studies Grades 4, 8, 10



Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 7-10
TCC Curve for Science Grades 4, 8, 10



Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 9-1
Reading Indices for Classification Consistency and Classification Accuracy

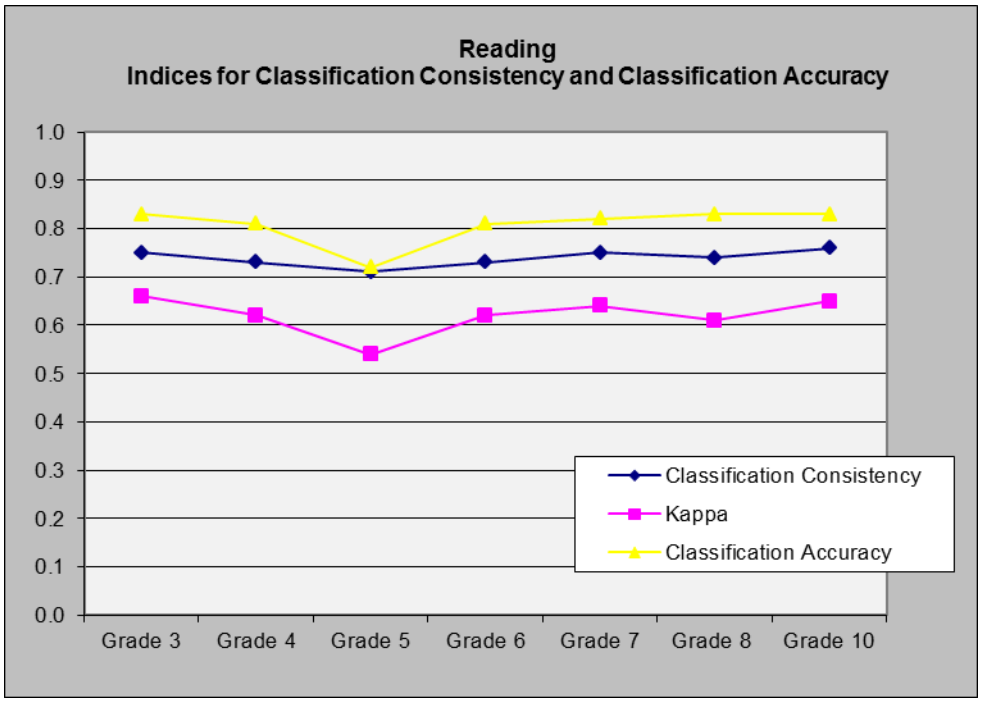
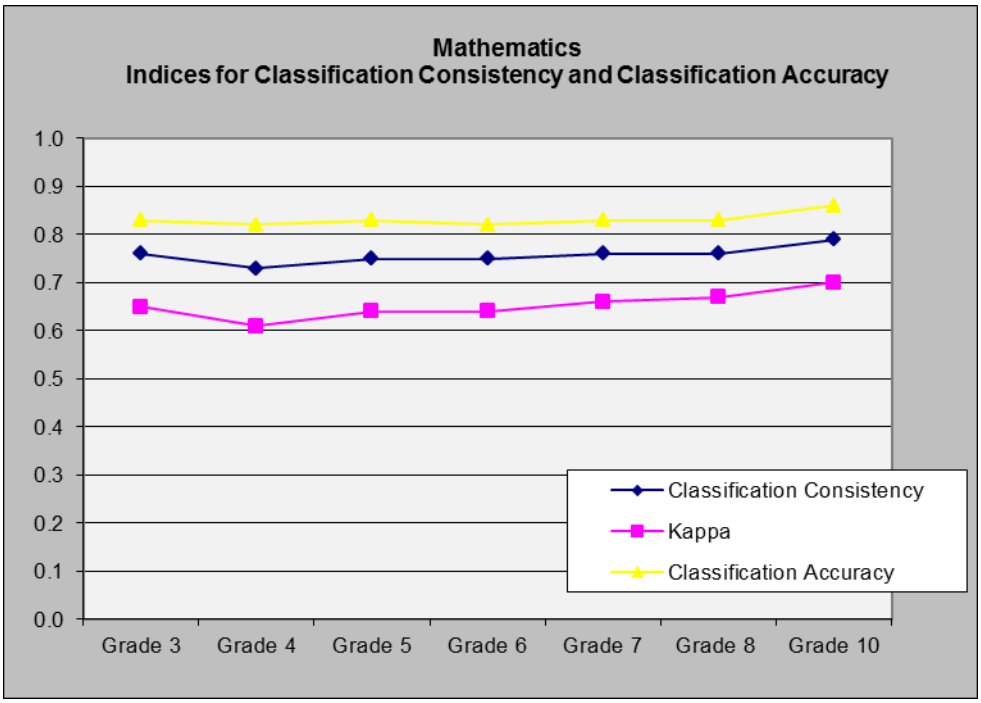


Figure 9-2
Mathematics Indices for Classification Consistency and Classification Accuracy



Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 9-3
Language Arts Indices for Classification Consistency and Classification Accuracy

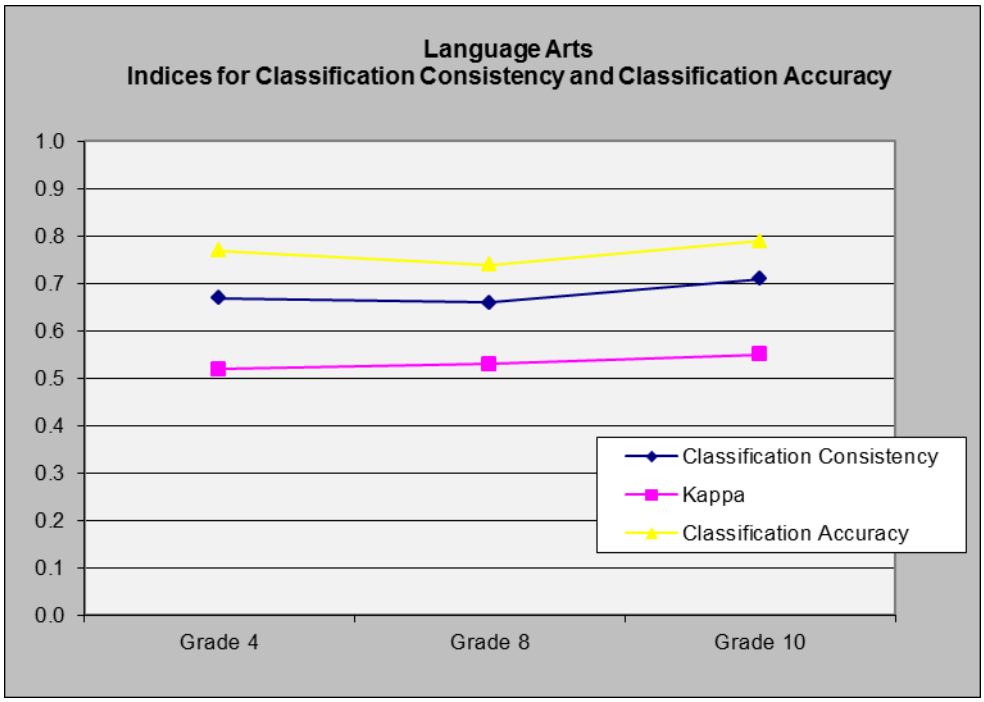
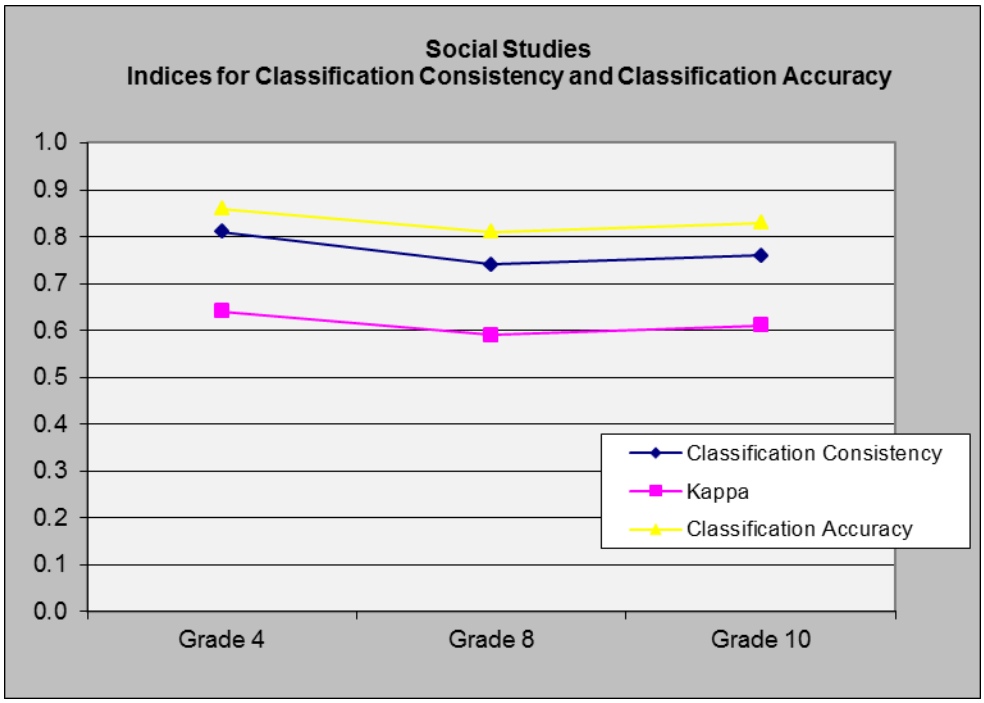
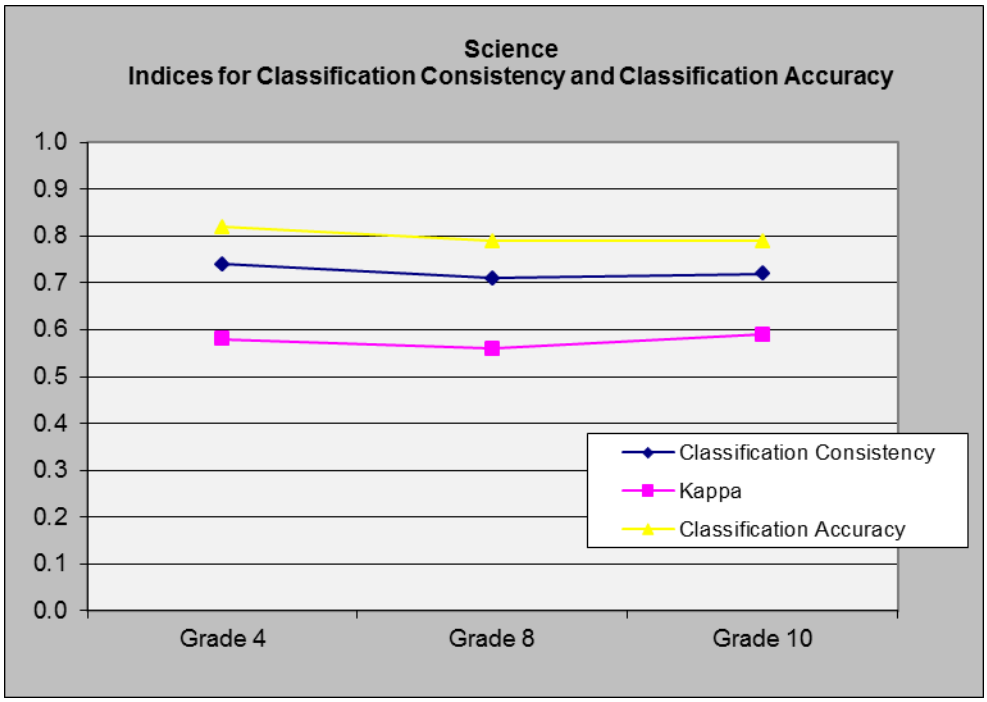


Figure 9-4
Social Studies Indices for Classification Consistency and Classification Accuracy



Copyright © 2013 by the Wisconsin Department of Public Instruction

Figure 9-5
Science Indices for Classification Consistency and Classification Accuracy



Copyright © 2013 by the Wisconsin Department of Public Instruction

Appendix 1: Fall 2011 Item Selection Check-Off Form

Copyright © 2013 by the Wisconsin Department of Public Instruction

Fall 2012 Item Selection Check-Off Form

Program Name:	Wisconsin Knowledge and Concepts Examinations (WKCE)
Administration Year:	Fall 2012
Content Area:	
Grade Level:	

	Fall 2011				**Anchor Items: Fall 2012**				**Total Form: Fall 2012**			
	No. Items	% No. Items	No. Points	% No. Points	No. Items	% No. Items	No. Points	% No. Points	No. Items	% No. Items	No. Points	% No. Points
SR												
CR												
Prompt												
Total												

Copyright © 2013 by the Wisconsin Department of Public Instruction

Blueprint Comparison (Number of items)

	Fall 2011 Blueprint Requirement			Fall 2012 Blueprint Requirement			Fall 2011 Actual Content Distribution			Fall 2012 Anchors			Fall 2012 Anchors			Fall 2012 Complete Form		
Reporting Category	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt
A																		
B																		
C																		
D																		
E																		
F																		
G																		
Total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Blueprint Comparison (% of items)

	Fall 2011 Blueprint Requirement			Fall 2012 Blueprint Requirement			Fall 2011 Actual Content Distribution			Fall 2011 Anchors			Fall 2012 Anchors			Fall 2012 Complete Form		
Reporting Category	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt	SR	CR	Prompt
A																		
B																		
C																		
D																		
E																		
F																		
Total																		

Copyright © 2013 by the Wisconsin Department of Public Instruction

Fall 2012 Form Distribution of Items by DOK & Objective (number of items)

Objective	Obj DOK	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	50% ≥ Obj DOK?	Comments
A							
B							
C							
D							
E							
F							
G							

*Combine SR & CR items

Answer Key Distribution

	A	**B**	**C**	**D**
Selected Items Session 1				
Session 2				
Session 3				
Session 4				
Session 5				
Total Test				

- The "Selected Items" entry should be the same as the sum of the 5 sessions on the total test.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Number of Items on DPI Watch List

	Anchor Items	Full Form	Item PEID IDs	Reasons for Use of Watch Items
Number of items				

Number of easy and difficult items for preventing ceiling and floor effect

	Previous Year's Form			Current Year's Anchors			Current Year's Full Form		
	SR	CR	ER	SR	CR	ER	SR	CR	ER
Mean P-value									
No. of items: P < .30									
No. of items: .30 < P < .40									
No. of items: .80 < P < .90									
No. of items: P > .90									

Copyright © 2013 by the Wisconsin Department of Public Instruction

Number of items flagged for point biserials (Pbis) indicating poor discrimination

	Fall 2011 Form			Fall 2012 Anchors			Fall 2012 Full Form		
	SR	CR	ER	SR	CR	ER	SR	CR	ER
No. of items: Pbis < .15									
No. of items: Pbis for distracter > 0									
No. of items: Pbis for correct choice is negative									
PEID ID of Flagged Items in Current Form:									
Reasons for Using Flagged Items in Current Form:									

Number of items near the Proficient Cut Score

	Fall 2011 Form		Anchors 2012 (SR only)		Fall 2012 Form	
	SR	CR	SR	CR	SR	CR
Proficient cut score = _____						
No. of items +/- 8 points around cut score						

Copyright © 2013 by the Wisconsin Department of Public Instruction

TCCs overlay each other closely?

	Fall 2011 Form and Fall 2012 Anchors	Fall 2011 Form and Fall 2012 Form	Fall 2011 Anchor and Fall 2012 Form
TCCs of Selected Form			

SE curves are smoothly bow-shaped without dips, bumps, and twists?

	Fall 2011 Form and Fall 2012 Anchors	Fall 2011 Form and Fall 2012 Form	Fall 2011 Anchor and Fall 2012 Form
SE curves of Selected Form			

Expected % Max. RS Difference between any two Selected Forms ≤ 0.05:

	Fall 2011 Form and Fall 2012 Anchors	Fall 2011 Form and Fall 2012 Form	Fall 2011 Anchor and Fall 2012 Form
Max Raw Score Difference			

Copyright © 2013 by the Wisconsin Department of Public Instruction

Number of Items with DIF

Group	Statistic	Fall 2011 Form		Fall 2012 Anchors		Fall 2012 Full Form		PEID ID of Items Flagged using one or more DIF method
		Against	Favor	Against	Favor	Against	Favor	
Gender	Female (Linn-Harnisch)							
	Male (Linn-Harnisch)							
	Mantel-Haenszel							
Ethnicity	White (Linn-Harnisch)							
	African American (Linn-Harnisch)							
	African American (Mantel-Haenszel)							
	Hispanic (Linn-Harnisch)							
	Hispanic (Mantel-Haenszel)							
	Asian (Linn-Harnisch)							
	Asian (Mantel-Haenszel)							
	American Indian (Linn-Harnisch)							
	American Indian (Mantel-Haenszel)							

Copyright © 2013 by the Wisconsin Department of Public Instruction

Number of Items with DIF cont'd

ELL	Proficient (Linn-Harnisch)							
	Not Proficient (Linn-Harnisch)							
	Mantel-Haenszel							
SES	Disadvantaged (Linn-Harnisch)							
	Not Disadvantaged (Linn-Harnisch)							
	Mantel-Haenszel							
Disability	Disabled (Linn-Harnisch)							
	Not Disabled (Linn-Harnisch)							
	Mantel-Haenszel							

Copyright © 2013 by the Wisconsin Department of Public Instruction

Number of Items with Less Than Optimal Fit

	Fall 2011 Form	Fall 2012 Anchor	Fall 2012 Full Form	PEID ID of Items Flagged using one or more DIF method	Reasons for Using Flagged Items
Fit					

Items Dropped in Fall 2011 Test to Not Use in Future Tests

Grade	Subject	Item#	PEID-Item	Form

Copyright © 2013 by the Wisconsin Department of Public Instruction

The Complete 2012 Operational Selection

2011 Book Item#	PEID#	2012 Book Item#	Anchor Status	P-value	Pbis	Is this item on a list of items to be avoided? Please explain.

Copyright © 2013 by the Wisconsin Department of Public Instruction

The Complete 2012 Operational Selection

2011 Book Item#	PEID#	2012 Book Item#	Anchor Status	P-value	Pbis	Is this item on a list of items to be avoided? Please explain.

Copyright © 2013 by the Wisconsin Department of Public Instruction

Approvals: Two independent reviews and approvals within Development are required prior to submitting to Research

	Name	Digital Signature	Date
Assessment Editor I			
Assessment Editor II			
Project Mgr/Development Lead			
Research Scientist			
WDPI			

Copyright © 2013 by the Wisconsin Department of Public Instruction

Appendix 2: 2012 WKCE Assessment Accommodations Matrix

Copyright © 2013 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2012

WISCONSIN DEPARTMENT OF PUBLIC INSTRUCTION

Accommodations for Students with Disabilities
on the *Wisconsin Knowledge and Concepts Examination (WKCE)* and *Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD)*

- All accommodations for a student with a disability must be documented on an IEP or Section 504 plan in the section for statewide assessment.* Refer to page 3
- All *Allowable Test Practices for All Students* must be documented in an IEP or Section 504 plan in the section for statewide assessment.
- Accommodations should be consistent with day-to-day instructional methods and should not be first introduced during testing.
- Accommodations should enhance access without changing the skill or construct measured.
- Districts should monitor the use of accommodations by comparing assessment accommodations received with those stated in IEP or Section 504 plans.

Accommodation Description For *Students with Disabilities* (D)		WKCE	WAA-SwD
Test Directions			
D 1	Sign language for directions.[1, 11]	✓	✓
D 2	Mark or highlight directions.[1, 2, 3]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 3	NEW Provide printed copy of teacher directions (i.e. bold text following the SAY icon) from the WKCE Test Administration Manual.[1]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 4	Explain or clarify directions.[1]	✓	✓
D 5	Student rereads and/or restates directions.[1]	✓	✓
Content Presentation			
D 6	Turn pages for student.	✓	✓
D 7	Braille; student responses must be transcribed into scorable test book by a licensed teacher of the visually impaired or a certified transcriber.[6, 14]	✓	✓
D 8	DPI-provided WAA-SwD Picture Descriptions; appropriate only for a student who cannot access the printed WAA-SwD, even with magnification, or the Braille WAA-SwD.[13]	*N/A*	✓
D 9	Large-print; student responses must be transcribed into scorable test book.[6, 14]	✓	*N/A: WAA-SwD is 18 pt. font, no separate large print edition.*
D 10	Extra test book; answers must be recorded in one scorable test book.[14]	✓	*N/A: All items are presented to the student so that they view one entire item at a time.*
D 11	Sign language for test passages and questions **(Not allowed on Reading tests)**.[11]	✓	✓
D 12	Text talker for test passages and questions **(Not allowed on Reading tests)**.[4]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 13	Student reads aloud to self.	✓	✓
D 14	Test administrator reads test passages and questions aloud **(Not allowed on WKCE Reading test or WAA-SwD "Read-by-Student" items)**.[9]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 15	Student records him/herself reading aloud and plays back recording.[4]	✓	✓

Copyright © 2012 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2012

Accommodation Description for *Students with Disabilities* (D)		WKCE	WAA-SwD
Content Presentation (cont.)			
D 16	Audio recording of test passages and questions in English **(Not allowed on WKCE Reading test or WAA-SwD)**.[4, 9]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 17	Read the Reading test ONLY in the following scenarios as described in *Form I-7-B*:[8, 9] a) For a student who is blind or visually impaired who is not yet proficient in contracted Braille, the WKCE Reading test passages and questions may be read aloud. b) For a student who is blind or visually impaired who is not yet proficient in un-contracted Braille, the WAA-SwD "Read-by-Student" Reading test items may be read aloud.	✓	✓
Response			
D 18	Manipulatives, base-ten blocks, 3-D shapes, 100's chart (not multiplication table), whole integer number lines, number boards, etc. are allowed as long as they do not provide a definition or description.	✓	✓ Follow guidelines in WAA-SwD Manipulatives Guide. http://dpi.wi.gov/oea/pdf/maniguide.pdf
D 19	Calculator and/or multiplication table (**Not allowed on sections of the Mathematics test measuring computation skills** -refer to each appropriate grade's Test Administrator's Manual at http://dpi.wi.gov/oea/publications.html).	✓	*N/A: A calculator is not allowed on the WAA-SwD.*
D 20	Braille output device; transcribe student responses into scorable test book.[4, 6, 14]	✓	✓
D 21	Student indicates responses orally to scribe.[5]	✓	*N/A: Test administrator records all student responses.*
D 22	Student signs responses to interpreter/scribe. For the Writing test, translation from American Sign Language (ASL) is not allowed; student must use English-based sign.[5, 11]	✓	✓
D 23	Student records responses using an audio or video device: a) Test administrator transcribes student's responses into scorable test book.[6, 14] b) Student watches or listens to his/her recorded responses and transcribes into scorable test book.[4, 6, 14]	✓	*N/A: Student is allowed to communicate responses in whichever mode is best for the student. Test administrator records student responses.*
D 24	Computer or word processor; responses must be transcribed into the scorable test book. For the Language Arts and Writing tests, all spell- and grammar-checking devices must be turned off; for the Mathematics test, the calculator function must be turned off for non-calculator sessions.[4, 6, 14]	✓	*N/A: Student is allowed to communicate responses in whichever mode is best for the student. Test administrator records student responses.*
D 25	NEW Speech-to-text devices; responses must be transcribed into the scorable test book. For the Mathematics test, the calculator function must be turned off for non-calculator sessions **(Not allowed on Language Arts or Writing tests).**[4, 6, 14]	✓	*N/A*
D 26	Provide spelling assistance or a spell-check device, where appropriate **(Not allowed on Language Arts or Writing tests).**	✓	*N/A: Student is not required to spell responses.*
Setting			

Copyright © 2012 by the Wisconsin Department of Public Instruction

D 27	Student moves, stands, or paces during individual administration.	✓	✓
Timing/Scheduling			
D 28	Extra time; test session must be completed within the same day the student started the session.[7]	✓	✓

Other Accommodations for Students with Disabilities	
D 29	Any accommodation not on this list must be submitted to DPI for approval, as it may represent a modification which changes the skill being measured. o All requests for an additional accommodation must be made to DPI **at least two weeks before the test administration window** begins, by completing and submitting the Request for Accommodation Form located at http://dpi.wi.gov/oea/accommtrx.html. o Requests will be reviewed by a committee to determine whether the request can be approved; approval or non-approval will be returned via fax or email.

*Allowable Accommodations for Students in Unique Circumstances

Some students who do *not* have an IEP or 504 plan, due to unique circumstances at the time of testing, may be able to demonstrate their learning more accurately through the use of accommodations on an ***as needed basis only***. In these unique cases, please follow the guidelines outlined in the matrix for Students with Disabilities; call DPI's Office of Educational Accountability with any questions at (608) 267-1072. Examples of unique circumstances:

- A student with a broken arm may need a scribe or be able to use a word processor to record responses.
- A student who forgot to wear eyeglasses may need a visual magnification device.

Copyright © 2012 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2012

Explanation of Footnotes - Only footnotes 1-9, 11, 13 and 14 apply to students with disabilities.

[1] **Test directions:**

- Any portion of the WKCE test book where the word "Directions" appears in a shaded/colored box, typically at the top of a page preceding a particular section of test content. In addition, test directions refer to anything that the test administrator reads aloud to the class from the WKCE Test Administration Manual (i.e. bold text following the SAY icon).
- WKCE item stems and test questions should not be considered directions.
- Test Directions for the WAA-SwD are incorporated into the teacher test book and are read aloud to the student. These directions must be read verbatim but may be reread if a student needs further clarification.
- Directions may not be expanded.

[2] **Marking test book with #2 pencil:** Student should not make pencil marks near answer bubbles, other than to mark one correct answer. Student should not mark in any of the following areas in the test book:

- the student Pre-ID Barcode on barcode label,
- the timing tracks (the parallel lines along the side of the test book),
- the skunk lines (the little squares and rectangles across the bottom of each page of the test book), or
- the Litho codes (the squares and numbers across the bottom of the first and last page of the test book).

[3] **Highlighters:**

- Carefully supervise the use of highlighters as they may cause smudging of pencil marks and bubbles and, therefore, could affect scoring.
- Do not allow the highlighting of track marks, litho codes, skunk lines, barcodes, pre-slugged bubbles or any carbon black printing. The highlighters cause these black inks to blur and bleed, which could affect scoring.
- Use only a highlighter from the following list, which were tested and found to have minimal problems:
  - Avery Hi-liter (regular or thin-tipped), Bic Brite-Liner, Sanford Major Accent, or Sanford Pocket Accent (thin-tipped)

[4] **Using audio/video or electronic (e.g., word processor or text talker) recordings:** when using accommodations that involve audio, video or electronic recordings or saved files, the test administrator must ensure that the recording or file is deleted upon completion of testing for security purposes.

[5] **Use of a scribe** (student dictates orally to scribe):

- A scribe may be provided when a student's documented disability, ELL status, or injury prevents them from writing their answer.
- When a student dictates responses orally to a scribe, the test must be administered in a separate, individual setting so as not to disturb other students.
- The WKCE Writing prompts measure composition, grammar, punctuation, capitalization, and spelling; therefore, a student must dictate these exactly as they are to be written.
- A scribe must be impartial and should allow the student adequate time to review and approve the response, if desired.
- All scribing should be done with a #2 pencil; responses scribed in ink will not be scored.

[6] **Transcribing student responses** (student's answers are documented in a manner other than in the scorable test book [e.g., large-print, Braille version, computer response, etc]):

- A translator who scribes student responses from native language to English should translate word-for-word to the extent possible for all content areas except Writing. For the Writing test, student must dictate or write responses in English (translation not allowed) exactly as they are to be written.
- The answers must be transcribed into the regular WKCE test book or WAA-SwD student Answer Document with a #2 pencil to be scored.
- Transcription of the student's responses must be verbatim, including spelling, formatting, punctuation, etc.
- Test security must be maintained. After answers are transcribed, destroy all electronically-saved student responses, including audio tapes. All paper copies of student work (e.g., Braille tests, large-print tests, graph/lined/grid paper, printed copies of computer responses, etc.) must be returned with non-scorable test materials.

[7] **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, a student requiring the use of restroom facilities should be escorted by either a test administrator or other school staff. In addition, a student must not be allowed to use any form of wireless communication during these breaks.

Copyright © 2012 by the Wisconsin Department of Public Instruction

[8] **Student who is blind or visually impaired and is not proficient in Braille** may have the Reading portion of the WKCE and the "Read by Student" Reading items of the WAA-SwD read aloud by a test administrator.

- The WKCE is available in contracted Braille; if a student designated by his/her IEP Team, by use of *Form I-7-B* (available at http://dpi.wi.gov/oea/dacforms.html), to take the WKCE is not proficient in contracted Braille and is receiving instruction in reading contracted Braille, the student may have the Reading test passages and items read by a test administrator.
- The WAA-SwD is available in uncontracted Braille; if a student designated by his/her IEP Team, by use of *Form I-7-B*, to take the WAA-SwD is not proficient in uncontracted Braille, the student may have the "Read by Student" items in the Reading test read by a test administrator.

[9] **Test Administrator Read Aloud Accommodation (not allowed on Reading test except for students qualifying for accommodation D17):**

- Test administrator must read in a pace and tone that is appropriate for each individual student. Careful attention must be given such that no changes in tone or inflection are detectable which might indicate a correct answer.
- Students may direct test administrator to reread a portion of a passage, test question, or answer choice as needed.

[10] **For students who have test items and/or directions translated into native language:**

- A qualified translator and interpreter (see http://dpi.wi.gov/oea/pdf/translator_guidelines.pdf) should have a Bachelor's Degree in Modern Languages or a certification in interpretation or translation. When this is not possible, be sure that a translator or interpreter has the following qualifications:
  1. Mastery of the target language and dialect
  2. Familiarity with both cultures
  3. Extensive general and academic vocabulary in both languages
  4. Ability to express thoughts clearly and concisely in both languages
- *Translators* work with the written word, transferring meaning from a source language into a target language. *Interpreters* work with the spoken word, transferring meaning from a source language into a target language.
- Translators and interpreters should participate in all aspects of staff training related to test administration and test security.
- For more information about state provided scripts available in Spanish and bilingual word lists in Spanish and Hmong for the WKCE, please see http://dpi.wi.gov/oea/ells.html.
- In order for this accommodation to be most effective, a student should have content-area knowledge in their native language.

[11] **Sign Language and Oral Interpreters**

- An interpreter needs to be able to translate in the same method of sign language typically used by the student (e.g., American Sign Language [ASL] or English-based Sign Language. The interpreters must not clarify, elaborate, or provide assistance with the meaning of words, intent of test questions, or responses to test items.
  - E.g. The sign for many math symbols often defines for the student what the item is intending to measure and would therefore invalidate the item.

[12] **Simplified English:** The test administrator providing an accommodation in which English is simplified for words not related to content or vocabulary should be familiar with the content area being tested. The WAA-SwD is already in simplified language.

Example (Grade 5 WKCE Released Item) of a simplified English test item:

The sales receipt below shows the groceries that José purchased from the supermarket. What is the estimated cost of José's groceries?
Simplified English: The receipt below shows the food that José bought from the store. Estimate how much money José spent on the food.
Note: It is important that "estimate" remain in this test item because it is part of the standard which is being tested.

[13] **DPI-provided Picture Descriptions** are descriptions of the graphic found within an item. Picture descriptions are intended to replace, *not* supplement graphics for a student who is blind or is visually impaired who is not able to access the printed WAA-SwD, even with magnification, or the Braille WAA-SwD. Ordering information can be found at: http://dpi.wi.gov/oea/dacforms.html.

[14] **Scorable Test Books** are the documents that are returned to the test vendor for scoring. For the WKCE, this is the test book itself. For the WAA-SwD, this is the student Answer Document. All student responses must be recorded on these documents in order to be scored.

Copyright © 2012 by the Wisconsin Department of Public Instruction

Appendix 3: WKCE Glossary

Copyright © 2013 by the Wisconsin Department of Public Instruction

Glossary

Abbreviations used in the WKCE Technical Report

2PPC: The two-parameter partial credit (2PPC) 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 (3PL) item response theory model. A mathematical model that shows the relationship between student achievement on a test and a single MC 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.

AYP: Adequate Yearly Progress. A state-defined criteria of educational accountability required as an outcome of the Federal NCLB law.

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. DIF is the degree to which an item performs differently for one group of examinees than it performs for another group of equally able examinees. DIF 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, and 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.

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.

FT: Field test item. A field test item is a newly developed item in a content area that is being administered to students for the first time. It does not contribute to student's score in a content area on WKCE.

Copyright © 2013 by the Wisconsin Department of Public Instruction

HOSS: Highest obtainable scale score. The highest possible scale score on a test.

ICC: Item characteristic curve. ICCs show the mathematical probabilities of students of varying degrees of achievement answering an item correctly as well as the characteristics of the item (e.g., item difficulty, item discrimination, item guessing).

IRT: Item response theory. IRT is a mathematic model that shows the relationship between student achievement on a test and the performance on a test item.

LA: Language Arts. A content area in the WKCE.

LH: Linn-Harnisch. A DIF statistic that utilizes information provided by the three-parameter IRT model for multiple-choice and constructed-response items.

LOSS: Lowest obtainable scale score. The lowest possible scale score on a test.

MA: Mathematics. A content area in the WKCE.

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. The Mantel-Haenszel (MH χ^2_{MH}) statistic is a commonly used DIF statistic for multiple-choice items.

NCLB: No Child Left Behind Act. The name of Federal Public Law No. 107-110, 115 Stat. 1425

NCME: National Council on Measurement in Education. A professional organization centered in assessment, evaluation, testing, and educational measurement.

OP: Operational item. An operational item is one that has previously undergone field testing so it contributes to a student's score in a specific content area on the WKCE.

RD: Reading. A content area in the WKCE.

SC: Science. A content area in the WKCE.

SD: Standard deviation. The SD is a measure of the variability of observations from the mean.

SEM: Standard error of measurement. The SEM is an estimated average standard deviation of the observed score.

SES: Socioeconomic status. A student population subgroup category describing students as economically disadvantaged or not economically disadvantaged.

Copyright © 2013 by the Wisconsin Department of Public Instruction

SMD: Standardized mean difference. SMD is commonly used DIF statistic for constructed-choice items.

SPI: Standardized performance indicator score. A subcontent area reporting score based on the items from a single content standard within given content area.

SS: Social Studies. A content area in the WKCE.

TCC: Test characteristic curve. TCCs show the mathematical relationship between students with varying degrees of achievement and their estimated overall test performance.

WKCE: Wisconsin Knowledge and Concepts Examinations. A criterion-referenced test designed to measure student achievement on the Wisconsin Model Academic Standards

WR: Writing. A content area in the WKCE.

Copyright © 2013 by the Wisconsin Department of Public Instruction